Influence of selected polymorphisms on the expression of breast cancer in Afrikaner *BRCA2* carriers

By

Sue-Rica Schneider

Submitted in accordance with the requirements for the degree of

Magister Scientiae in Medical Science (M.Med.Sc)

In the Faculty of Health Sciences, Division of Human Genetics

University of the Free State, Bloemfontein, South Africa

Supervisor: Dr NC van der Merwe

Co-supervisor: Dr B Visser

December 2011

DECLARATION

I certify that the dissertation hereby submitted for the degree *M.Med.Sc*. at the University of the Free State is my independent effort and had not previously been submitted for a degree at another University/Faculty. I furthermore waive copyright of the dissertation in favor of the University of the Free State.

____

S-R Schneider

To my beloved family

In the beginning, you have your genes

In the end, it’s what you did with them

that makes the difference.

Aubrey Milunsky.

Acknowledgements

The success of this study would not have been possible without the guidance, support and contributions of several individuals and institutions.

I would like to express my deepest gratitude to my study-leaders, Dr NC van der Merwe and Dr B Visser, for their patience, motivation and support through this project, whose guidance and immense knowledge, have been invaluable to me.

Sincere appreciation goes to the breast cancer patients and control individuals for their participation in this study. Without your contribution none of this would have been possible.

I acknowledge my gratitude to the the Division of Human Genetics (UFS) for resources and facilities and the National Health Laboratory Services (NHLS) for financial assistance in the study.

I am thankful to my colleagues at the Department of Genetics for their assistance and understanding.

A special thank to Prof G Joubert for the statistical analysis of the study.

My sincere appreciation to my family, for their care, endless love, dedication and the many years of support. I am deeply appreciative of my beloved husband, Jurgens, my pillar, for his patience, love and unlimited support who stood beside me and encouraged me constantly.

To the Creator for all my blessings.

Tables of contents

Chapter 4 Influence of selected polymorphisms on the expression of breast cancer in Afrikaner *BRCA2* carriers 103

List of Figures

List of Tables

Abbreviations

AF-2	Activation factor 2
APRS	Apert syndrome
Arg	Arginine (amino acid)
Asn	Asparagine (amino acid)
ATP	Adenosine-5'-triphosphate
ATR	Ataxia telangiectasia and RAD3 –related gene
ATM	Ataxia telangiectasia (mutated)
BARD1	BRCA1-associated RING domain
BC	Breast cancer
BCLC	Breast Cancer Linkage Consortium
BIC	Breast Cancer Information Core database
BMI	Body mass index
bp	Base pair
BRAF-35	BRCA2 - associated factor
BRC	BRCA2 repeat motif
BRCA1	Breast cancer susceptibility gene 1
BRCA2	Breast cancer susceptibility gene 2
BRCT	BRCA1 carboxy-terminus
BRIP	BRCA1-interacting protein
BSA	Bovine Serum Albumin
CASP8	Caspase 8 gene
CBP	CREB-binding protein
CHEK2	Checkpoint kinase 2 gene
CI	Confidence interval
CIMBA	Consortium of Investigators of Modifiers of *BRCA1* and *BRCA2*
Cq	Cycle threshold
CREB	cAMP response element-binding

CS	Crouzon syndrome
C-terminus	Carboxy terminus
CtIP	C-terminal-binding –protein-interacting protein
del	Deletion
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleic triphosphates
DSBs	Double strand breaks
DSS1	Deleted in split-hand/split-foot 1 region
DTT	1,4-Dithiothreitol
dup	Duplication
dx	Ages at onset
E_1	Estrone
E_2	Estradiol
E_3	Estriol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
ER	Estrogen receptor
ERα	Estrogen receptor alpha gene
ERβ	Estrogen receptor beta gene
ERE	Estrogen response elements
ERK	Extracellular regulated kinase
*ESR*1	Estrogen receptor 1 gene
ESR2	Estrogen receptor 2 gene
EtBR	Ethidium bromide
F-actin	Actin filament
FAM	6- carboxyfluorescein
FANCD2	FA complementation group D2
FGFs	Fibroblast growth factors
FGFR2	Fibroblast growth factor receptor 2 gene
FRET	Fluorescence resonance energy transfer
FRS	FGF receptor substrates

FTP	Full-term pregnancy
g/l	Grams per litre
Gln	Glutamine (amino acid)
Glu	glutamate (amino acid)
Gly	Glycine (amino acid)
Grb2	Growth factor receptor-bound protein 2
GWAS	Genome-Wide Association Studies
HAPMAP	Haplotype Map
HER	Human epidermal growth factor receptor
HEX	6-Hexachlorofluorescein
HMG	High mobility group
HR	Hazard ration
HR	Homologous recombination
HRT	Hormone replacement therapy
HWE	Hardy-Weinberg equilibrium
IDT	Integrated DNA Technologies
Ig	Immunoglobulin
Ins	Insertion
IPTG	Isopropyl-β-D-thiogalactopyranoside
JNK	c-Jun N-terminal kinase
JWS	Jackson-Weiss syndrome
kb	Kilobases
KCL	Potassium chloride
kDa	Kilodalton
LB	Luria-Bertani
LD	Linkage disequilibrium
LSP1	Lymphocyte-specific protein 1 gene
M	Molar (moles per liter)
$MgCl_2$	Magnesium chloride
MAPK	Mitogen-activated protein kinases
MAP2K	Mitogen-activated protein kinases kinases

MAP3K1	Mitogen-activated protein kinases kinases kinases 1 gene
MAPKAP	MAPK-activated protein
MDM2	Mouse double minute 2 gene
MGB	Minor groove binder
$MgSO_4$	Magnesium sulfate
mM	Millimolar
mRNA	Messenger ribonucleic acid
MyoD	Muscle determination factor
NaCl	Sodium chloride
NCBI	National Center for Biotechnology Information
NCR	National Cancer Registry
ng	Nanograms
ng/µl	Nanograms per microliter
NHEJ	Non-homologous end joining
NHLS	National Health and Laboratory Services
NLS	Nuclear localization sequence
N-terminus	Amino terminus
NTC	No template control
OC	Oral contraceptives
OMIM	Online Mendelian Inheritance in Man
OR	Odds ratio
OCCR	Ovarian cancer cluster region
ORIGO	Dutch hospital-based cohort of breast cancer patients
OVC	Ovarian cancer
PALB2	Partner and localizer of *BRCA2* gene
P/CAF	p300/CBP – associated factor
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
pmol	Pico moles
Pro	Proline (amino acid)

PS	Pfeiffer syndrome
PTEN	Phosphatase and tension homolog gene
qPCR	Quantitative polymerase chain reaction
RAD51	Homolog of RecA of *E. coli*
RFLP	Restriction fragment length polymorphism
RFU	Relative fluorescence units
RING	Zinc-chelating domain
Rnb	Fluorescence emission of the baseline
Rnf	Fluorescence emission intensity of the reporter
RR	Relative risk
SA	South Africa
SDS	Sodium dodecyl sulphate
Ser	Serine (amino acid)
SET	Sodium chloride EDTA-Tris HCl
SM	Second messenger
SNPs	Single nucleotide polymorphisms
SOP	Standard operating procedure
Sp1	Specificity Protein 1
SRC	Steroid receptor co-activator
SSCP	Single-strand conformation polymorphism
STK11	Serine/threonine kinase gene
T_a	Annealing temperature
TAMRA	Tetramethylrhodamine
Taq	*Thermus aquaticus*
TBE	Tris Borate EDTA buffer
TET	Tetrachlorofluorescin
TOX3	Tox high mobility group box family member 3
Tp53	Tumour suppressor p53 gene
TNRC9	Trinucleotide repeat-containing 9 gene
Tris	2-Amino-2-(hydromethyl)-1,3-propanediol
Triton X-100	t-octylphenoxypolyethoxyethanol

U	Units
$V.cm^{-1}$	Volts per centimetre
v/v	Volume per volume
Val	Valine
VIC	2'-chloro-7'-phenyl-1,4-dichloro-6-carboxyfluorescein
w/v	Weight per volume
WHO	World Health Organization
X-GAL	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside
µg	Microgram
µl	Microlitre
µM	Micromolar

Chapter 1

Introduction

The *BRCA1* (Breast cancer susceptibility gene number 1) and *BRCA2* (Breast cancer susceptibility gene number 2) mutation frequencies in populations of Western European descent ranges from 1/190 to 1/900, placing breast cancer (BC) amongst the most prevalent high-risk hereditary disorders (Hughes, 2008). Mutation frequencies are much higher in certain ethnic groups such as the Ashkenazi Jews and the Caucasian South African Afrikaans speaking population due to the presence of founder effects. The prevalence rates of *BRCA1* and *BRCA2* mutations and their highly associated cancer risks make *BRCA1* and *BRCA2* a significant health concern. Despite the high overall lifetime risk of breast and ovarian cancer conferred by these germline mutations, various differences were observed between mutation positive individuals within families relating to the age at onset and the type of cancer present within the index case (Antonio *et al.*, 2003; Simchoni *et al.*, 2006). It is important that these inter-individual phenotypic differences amongst *BRCA* mutation carriers be investigated so that it can be taken into account when deciding upon risk reduction strategies.

The expression of the two familial BC genes *BRCA1* and *BRCA2* are influenced by polymorphisms within various low penetrant genes or environmental factors (Lynch *et al.*, 1989). Mutation detection and single nucleotide polymorphism (SNP) genotyping techniques have become areas of intensive research to identify genetic modifiers of cancer risk conferred by the *BRCA* genes. Various SNPs in other genes have been associated with an increased risk for the development of BC (Tryggvadottir *et al.*, 2003; Haile *et al.*, 2006).

Over the past decade candidate SNPs were selected on the basis of an understanding of relevant biochemical and physiological pathways of carcinogenesis. Current technologies however, allowed for the identification of candidate SNPs by genome-wide association studies (GWAS) (Hirschhorn and Daly, 2005) without prior knowledge of the relevant pathways. Current GWAS proved that polymorphisms in *TNRC9* (trinucleotide repeat-containing 9 gene)*, LSP1* (lymphocyte-specific protein 1 gene)*, MAP3K1* (mitogen-activated protein kinases kinases kinases 1gene) and *FGFR2* (fibroblast growth factor receptor 2 gene) play a significant role in BC (Easton *et al*., 2007; Hunter *et al.*, 2007; Stacey *et al.*, 2007; Gaudet *et al*., 2010). Despite considerable interest in the influence of these candidate gene mutations on BC risk, only a small number of published studies targeted *BRCA* mutation carriers specifically (Foulkes *et al.*, 2003; Campa *et al.*, 2011). Furthermore, no studies on the influence of these candidate genes on *BRCA2* 8162delG (c.7934del, p.Arg2645AsnfsX3) mutation carriers within the Afrikaner population have been published to date.

Another candidate low penetrance gene possibly involved in BC risk is the estrogen receptor 1 gene (*ESR1*) (Siddig *et al*., 2008). Estrogen is an important epidemiological risk factor and its effects are mediated through the estrogen receptor (ER) in breast tissue (Heldring *et al*., 2007). It is reported that estrogen plays a crucial role in breast growth, differentiation and the development of cancer (Liehr, 2000; Noruzinia *et al*., 2005).

The aim of this study was thus to genotype previously identified polymorphisms that have been proven by CIMBA (Consortium of Investigators of Modifiers of *BRCA1* and *BRCA2*) consortium to be associated with an increased BC risk in the general population (Easton *et al*., 2007) and in *BRCA2* mutation carriers specifically (Antoniou *et al.*, 2008). The prevalence of each of these modifying SNPs would be determined for the Caucasian Afrikaner population in order to evaluate whether these polymorphisms play a role in the phenotypic variance seen amongst *BRCA*2 8162delG (c.7934del, p.Arg2645AsnfsX3) mutation carriers.

Chapter 2

Literature Review

2.1 Incidence of breast cancer in South Africa

Breast cancer (BC) is the most common malignancy amongst women in industrialized countries. In South Africa (SA), it was second only to cervical cancer between 1986 and 1992, but became the leading cancer in women between 1993 and 1995 (Vorobiof *et al.*, 2001). The most recent figures published by the South African National Cancer Registry (NCR) in 2001 indicated that BC is currently the most common diagnosed cancer among SA women (http://www.cansa.org.za). Breast cancer accounts for 19.4% of all cancers in SA compared to 10% worldwide, with an overall incidence rate of 1 in 26 (Loubser *et al*., 2008). The risk varies amongst ethnic groups with 45% of all cases reported for this period being Caucasian women, resulting in a 1 in 12 life time risk to develop the disease. Compared to other countries, the incidence rate in SA Caucasian women was fourth highest, with an incidence of 76.5 per 100 000 (Parkin *et al*., 2002). Apart from the Caucasian population, BC is also the leading cancer type in Mixed Ancestry and Asian women, with similar incidence rates of 49 per 100 000 in 1999. The life time risk for these two population groups is 1 in 18. The life time risk was the lowest for Black African women, namely 1 in 49 (Loubser *et al*., 2008). According to Kruger and co-workers (2007), the incidence of the disease is rising for urban Black SA women compared to the rural population. This can possibly be due to the adoption of a Western lifestyle.

2.2 Hereditary Breast Cancer

The hereditary nature of BC was recognized more than a century ago by Broca, (1866). It was not until the mid 1990s that the hereditary aspects of cancer susceptibility became clear when the two highly penetrant *BRCA1* (Online Mendelian Inheritance in Man [OMIM 113705]) and *BRCA2* (OMIM 600185) genes were identified (Miki *et al*., 1994; Wooster *et al*., 1995). During and after the discovery of these genes, several studies focused their research on families with an early age at diagnosis (younger than 35 years) and the presence of bilateral female and male BC, which indicate the presence of a potential genetic predisposition (Lux *et al*., 2006).

The portion of BC cases that can be directly attributed to an inherited predisposition are 5 to 10%, with 15 to 20% being explained by germline mutations in the highly penetrant *BRCA1* and *BRCA2* genes (Silla *et al.*, 1995; Claus *et al.*, 1996). The individual risk associated with hereditary BC depends on the gene implicated, the specific mutation involved, the extent of the family history and ethnicity (Loubser *et al*., 2008). *BRCA* mutation carriers have a lifetime risk of 60-85% to develop BC and a 15-40% for ovarian cancer (OVC) (Thompson and Easton, 2002). The risk of BC further increases for women with the presence of an affected first degree relative and the number of affected relatives present within the family. One of the largest population-based family studies in Sweden reported that of 9,371 daughters with BC below age 54 years, 8.7% had mothers with BC resulting in a familial risk of approximately 1.8% (Hemminki and Vaittinen, 1998).

2.3 Familial breast cancer genes

2.3.1 The breast cancer susceptibility gene number 1 (*BRCA1* OMIM 113705)

The first BC predisposing gene, *BRCA1* that was identified in 1990, is localized on chromosome 17q12 (Hall *et al.*, 1990). It was cloned in 1994 using linkage analysis studies in families with multiple cases of breast and ovarian cancer (Miki *et al.*, 1994). The gene is large and consists of 24 exons spread over 80 kilobases (kb). The 22 coding exons are transcribed into a 7.8 kb mRNA that encodes a nuclear protein with 1863 amino acids and a molecular weight of approximately 220 kDa (Chen *et al*., 1995). Exon 11 is the largest and codes for 60% of the protein (Miki *et al.*, 1994, Chen *et al.*, 1995).

BRCA1 shows no sequence homology to any other genes, but has a RING finger motif (Zinc-chelating domain) near the amino terminus (N-terminus) (Miki *et al.*, 1994) (Fig. 2.1). The N-terminus includes a conserved pattern of seven cysteines and one histidine (Miki *et al.*, 1994). This RING finger motif facilitates protein-protein and protein-DNA interactions (Miki *et al*., 1994, Saurin *et al*., 1996) and enables BRCA1 to interact through this domain with another ring finger motif containing protein called the *BRCA1*-assosiated RING domain protein (BARD1), creating a hetero-dimer (Fig. 2.1) (Wu *et al.*, 1996). According to Simons and co-workers (2006), BRCA1 and BARD1 act together in promoting tumor suppression.

Ring finger motifs are characteristic of proteins that are involved in macro-molecular complexes which facilitate ubiquitination (Lorick *et al*., 1999). The RING finger motif can function as an ubiquitin-protein ligase which targets proteins for degradation by proteasomes. The loss of such a RING finger motif can result in an increase in proteins that could stimulate proliferation (Welcsh *et al.*, 2000).

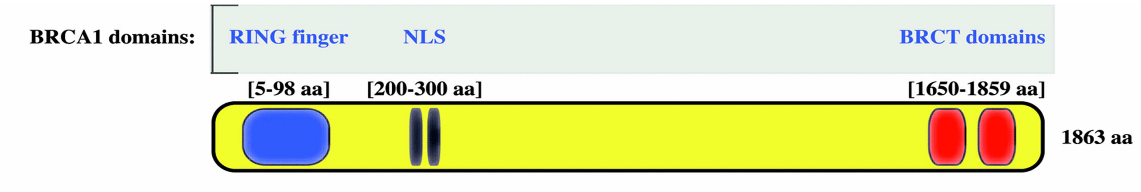


Figure 2.1 A schematic presentation of the primary structure of BRCA1 indicating the RING finger, NLS and BRCT domains as well as the interacting proteins (Boulton *et al.*, 2006).

The carboxy terminus (C-terminus) of BRCA1 also contains two tandem repeats of the poorly conserved BRCA1 carboxyl terminal (BRCT) motif, a RAD51 binding domain as well as a nuclear localizing signal (NLS) that permits entry into the nucleus (Bertwistle *et al*., 1997; Welcsh *et al.*, 2000; Lee *et al.*, 2001; Shiozaki *et al*., 2004). The BRCT domain is involved in DNA damage response, tumor expression, DNA repair, transcription co-activation and cell-cycle regulation (Scully *et al.*, 1997; Williams *et al.*, 2003; Shiozaki *et al*., 2004). This domain facilitates protein-protein interactions (Wu *et al.*, 1996).

BRCA1 maintains chromosomal stability by interacting with and regulating the RAD51 protein (Fig. 2.1) (Scully *et al*., 2004). This protein is a homologue of RecA of *Escherichia coli* (*E. coli*) and is involved in DNA break repair and recombination. DNA breaks are caused by radiation, environmental exposure or chromosomal exchange of genetic material during meiosis. According to Miki *et al*. (1994), the human *BRCA1* mRNA is expressed at high level in the testis, thymus, breast and ovaries (Miki *et al.*, 1994).

2.3.2 Breast cancer susceptibility gene number 2 (*BRCA2* OMIM 600185)

The second familial BC gene *BRCA2* was localized to chromosome 13q12-13 by Wooster and co-workers during 1994. They performed a linkage search in 15 high-risk BC families that were not associated with the *BRCA1* locus on chromosome 17q21. Their analysis uncovered a second BC susceptibility locus, where after it was cloned and sequenced by both Wooster and his team (1995) and Tavtigian and co-workers (1996). *BRCA2* is longer than *BRCA1* with 26 of the 27 exons collectively encoding a nuclear protein of 3418 amino acids with a molecular mass of 382 kDa (Wooster *et al*., 1995; Tavtigian *et al*., 1996). The transcript is approximately 12 kb in length and contained within 70 kb of genomic sequence (Wooster *et al*., 1995).

BRCA2 contains eight conserved BRC (BRCA2 repeat motif) repeats that have been termed the ovarian cancer cluster region (OCCR) and is coded for by exon 11 (Bork *et al.*, 1996). The BRC motif is ~70 amino acids in length with a core sequence of 26 amino acids (Wooster *et al*., 1995). This region mediates direct protein-protein interaction with RAD51 which also plays a role in DNA repair and recombination (Wong *et al.*, 1997; Welcsh *et al*., 2000). Two NLS motifs near the C-terminus are responsible for the nuclear localization and function of the BRCA2 protein (Fig. 2.2).

The binding of BRCA2 to the Deleted in split-hand/split-foot 1 (DSS1) region is essential for DNA repair (Yang *et al.*, 2002). DSS1 is a highly conserved 70 amino acid protein that interacts with the C-terminus DNA/DSS1-binding domain of BRCA2, distal to the BRC region (Marston *et al*., 1999; Yang *et al.*, 2002). DSS1 plays an important role in the maintenance of genomic stability and BRCA2-dependent recombination. DSS1 and BRCA2 target RAD51 to sites of double strand breaks (DSBs) (Venkitaraman, 2002). BRCA2 is expressed in several tissues including the mammary gland, spleen, ovary, lung, testis and thymus (Tavtigian *et al*., 1996).

2.3.3 Function of the BRCA proteins

Since the discovery of the *BRCA* genes, researchers needed to determine the role and function of these proteins. These multifunctional proteins interact with other proteins that are involved in many fundamental cellular processes (Table 2.1), while their gene expressions are regulated by the cell cycle (Bertwistle *et al*., 1997; Ruffner and Verma, 1997). Their main function is to maintain genomic integrity. Both BRCA proteins are involved in the biological response to DNA damage.

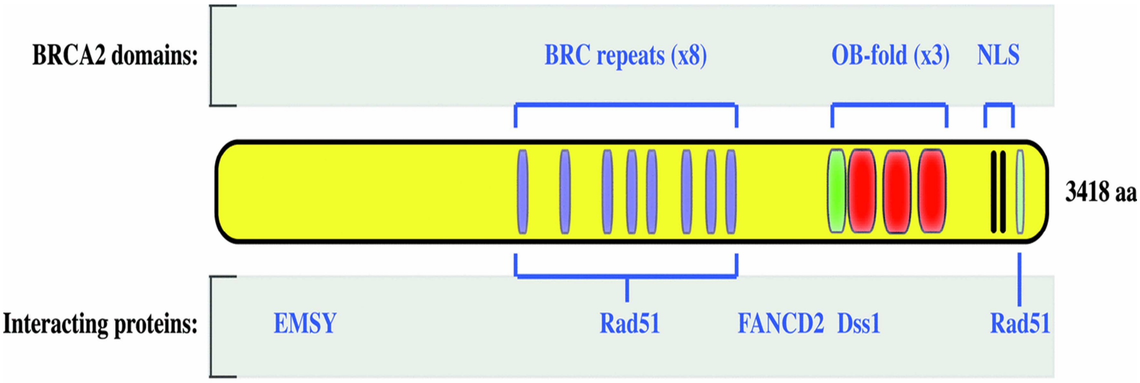


Figure 2.2 A schematic presentation of BRCA2 indicating the domains and interacting proteins (Boulton *et al.*, 2006).

Table 2.1 Proteins that interact with the BRCA proteins. Adapted from Welcsh and King (2001).

BRCA1 interacting protein or complex	Function of protein	Interacting Domain
RAD51	DSB repair	Exon 11
BRCA2	DSB repair	BRCT domain
p53	Transcription factor, tumor suppressor	Exon 11 and BRCT domain
Estrogen receptor	Ligand responsive transcription factor	N-terminus
BARD1	Ubiquitination	RING
CHEK2	Checkpoint Kinase	Ser988
ATM	Checkpoint Kinase	Ser1423 - 1524
ATR	Checkpoint Kinase	Ser1423
CtIP	Binds CtBP; transcriptional co-repressor	BRCT domain
p300/CBP	Transcriptional co-activator	RING and BRCT domain

BRCA2 interacting protein or complex	Function of protein	Interacting Domain
RAD51	DSB repair	Exon 11 BRC repeats and C-terminus
BRCA1	DSB repair	?
DSS1	Deleted in split hand/split foot	Exon 11
P/CAF	Histone acetylation; chromatin remodeling	N-terminus
BRAF-35	Cell cycle progression	Exon 11 BRC repeats 6-8

This involves the repair of DSBs by homologous recombination (HR), activation of cell cycle checkpoints in the DSB repair pathway and repair of damage by transcription-coupled repair (Chen *et al*., 1999; Venkitaraman, 2002). There is also a possible role for BRCA1 in non-homologous end-joining (NHEJ) (Bau *et al*., 2006). BRCA1 and BRCA2 have also been implicated to play a role in regulating centrosome amplification (Tutt *et al*., 1999; Xu *et al*., 1999), chromatin remodeling and protein ubiquitination (Welcsh *et al.*, 2000).

BRCA1 and BRCA2 function as tumor suppressors, gatekeepers and caretakers by inhibiting growth or promoting cell death (Kinzler and Vogelstein, 1997). They ensure that the cell is not compromised by loss, duplication or rearrangement of DNA. The loss of function of these genes needs two mutations according to the Knudson double hit hypothesis to lead to tumor development (Knudson, 1971). Tumor formation is not directly the result of mutations in the caretakers genes, but instead can be caused by genetic instability which increases the inactivation of the gatekeepers and activation of proto-oncogenes (Thompson *et al.*, 1995; Kinzler and Vogelstein, 1997).

2.3.4 DNA double-strand break repair (DSB)

There are two pathways that repair DSB, namely HR and NHEJ (Khanna and Jackson, 2001). NHEJ is error-prone and can occur between breaks on different chromosomes which can lead to translocations and deletions. HR is more accurate, for it uses the complimentary sister chromatid as a template for the repair (Shrivastav *et al.*, 2008).

According to Welcsh *et al.* (2000), BRCA1 and BRCA2 are involved in a macromolecular complex with BARD1 and RAD51 to repair DSB through HR (Fig. 2.3). This complex relocates to the chromosomal regions marked by the proliferating cell nuclear antigen (PCNA).

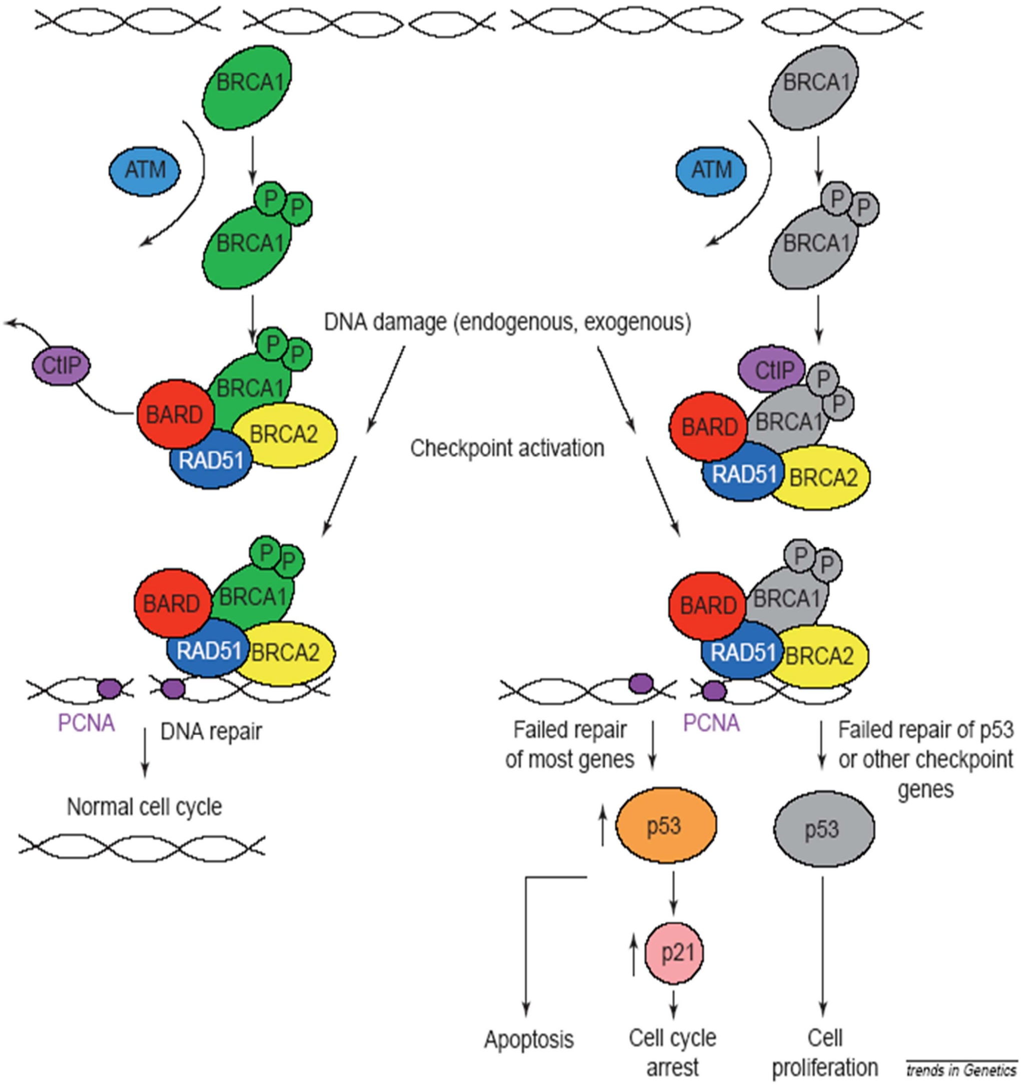


Figure 2.3 Schematic presentation of the macromolecular complex involved in DSB repair (Welcsh *et al.*, 2000).

BRCA1 is phosphorylated by the ATM protein kinase in response to DNA damage and dissociates from the C-terminal-binding-protein-interacting protein CtIP. BRCA1 binds to RAD51 during the S phase of the cell cycle and then re-locates to the damaged DNA site (Venkitaraman, 2001). *BRCA2* is more directly involved in DSB repair than *BRCA1*, for RAD51 binds directly to the BRC motif on the *BRCA2* gene. The BRCA2/RAD51 complex has two states *in vivo*: an inactive state which prevents the binding of single-strand DNA by RAD51 and an active state where RAD51 can form nucleoprotein filaments and deliver it to the DSB repair site (Venkitaraman, 2001). The activation is due to the phosphorylation of BRCA2 by the ATM protein kinase in response to DNA damage that releases the RAD51 (Venkitaraman, 2001).

2.4 Germline mutations in *BRCA1* and *BRCA2*

Deleterious germline mutations within *BRCA1* and *BRCA2* result in a genetic predisposition to develop BC. Women carrying these mutations normally develop BC at a younger age compared to sporadic cases (Lux *et al*., 2006). Thousands of *BRCA* mutations have been identified for BC families. These mutations are found throughout the entire coding region of the genes with no mutational "hot spots" (Cipollini *et al.*, 2004; Thompson and Easton, 2004). According to the Breast Cancer Information Core (BIC) (www.nhgri.nih.gov/bic) database, 1871 different mutations have been identified in *BRCA1* and 2109 in *BRCA2*. The majority of mutations include frame shift mutations caused by small insertions and deletions, nonsense mutations or alterations affecting splice-sites (BIC). These mutations result in a premature stop codon which leads to the truncation of the resultant protein (Ellisen and Haber, 1998; BIC). The significance of the majority missense mutations recorded thus far is still unknown.

Germline mutation in the *BRCA1* and *BRCA2* genes are also associated with an increased risk of other cancers. Brose and co-workers (2002) recorded an increased risk of colon, gastric, male breast, fallopian tube and pancreatic cancer in *BRCA1* mutation carriers. Other studies also suggested an increased risk of prostate cancer in *BRCA1* (Ford *et al*., 1994; Struewing *et al*., 1997). The Breast Cancer Linkage Consortium (BCLC) confirmed an increased risk in prostate and pancreatic cancer in *BRCA2* mutation carriers as well as cancer of the pharynx, stomach, melanoma of the skin, gallbladder and bile duct (BCLC, 1999).

2.5 Penetrance

The penetrance of a specific mutation refers to the life-time probability of a mutation positive individual to develop BC. It usually depends on age, sex, environment, lifestyle and hormonal factors (Newman *et al*., 2001). Initial penetrance estimates for *BRCA1* and *BRCA2* mutations were derived from multiple-case families with germline mutations from the BCLC. These estimates indicated a cumulative lifetime risk of BC at the age of 70 years of 85–87% and 77–84% respectively (Ford *et al*., 1998). Later results suggest that these studies may have overestimated the effect of the *BRCA1/2* mutation within a family.

The average risk of developing BC and OVC in *BRCA1* mutation carriers is currently 65% and 39% respectively by the age of 70 (Antoniou *et al*., 2003). For *BRCA2* mutation carriers, the risk seems lower namely 45% for BC and 11% for OVC (Antoniou *et al.*, 2003). A meta-analysis, utilizing the results of 10 studies, indicated a mean cumulative risk of 57% at age 70 for BC (95% confidence interval (CI), 47% to 66%) in *BRCA1* and 49% (95% CI, 40% to 57%) in *BRCA2* mutation carriers (Chen and Parmigiani, 2007).

The data revealed heterogeneity among the reported risks. Chen and his team (2007) reported that different populations may segregate different mutations and different risk factors. This is further complicated by variability in cancer risk among *BRCA* mutation carriers which can be attributable to risk modifying genes and / or other risk factors (Rebbeck, 2002).

The penetrance of a *BRCA* mutation may also be influenced by the position of the mutation. Mutations in the central region of *BRCA1* are associated with a lower risk for BC, whereas mutations towards the 3' end of the gene have a lower risk for OVC (Thompson and Easton, 2002). Mutations in the OCCR between nucleotides 3035 and 6629 in the central part of *BRCA2* are also associated with a higher OVC risk and a lower risk for BC (Gayther *et al*., 1997; Thompson and Easton, 2001).

2.6 Prevalence and founder effects

The prevalence of *BRCA* mutations in breast and ovarian cancer families has been extensively studied in different populations and ethnic groups and is the highest in populations with founder effects (Neuhausen, 2000). Founder mutations are normally detected or present in certain ethnic populations that have a relatively homogenous ancestry such as the Ashkenazi Jews (Roa *et al*., 1996; Struewing *et al*., 1997) and the Finnish population (Vehmanen *et al*., 1997). Haplotype analysis of families representing these populations carrying a specific *BRCA* mutation can reveal whether these high frequency alleles are derived from a single mutational event or whether they have arisen independently more than once (Newman *et al.*, 2001). Many *BRCA* founder mutations have been described for a large number of populations including the Ashkenazi Jewish (Roa *et al*., 1996; Struewing *et al*., 1997), Dutch (Hartmann *et al.*, 2004; Zeegers *et al.*, 2004), Icelandic (Roa *et al.*, 1996), African American (Olopade *et al*., 2003; Pal *et al.*, 2004) and South African Afrikaner populations (Reeves *et al*., 2004).

The prevalence of mutations can be easily determined in population groups with a restricted number of founder mutations. One such study was done by Struewing (1995) on the Ashkenazi Jewish population. The three most common *BRCA* mutations are 185delAG (c.68_69del, p.Glu23ValfsX17) and 5382insC (c.5266dup, p.Gln1756ProfsX74) in *BRCA1* and 6174delT (c.5946del, p.Ser1982ArgfsX22) in *BRCA2*. *BRCA1* 185delAG is found in 20% of the

Ashkenazi Jewish population with BC diagnosed before the age of 42 whereas *BRCA2* 6174delT accounts for 8% of BC cases (Peto *et al.*, 1999).

2.7 Founder mutations in the South African Afrikaner

Founder mutations within a population are the result of years of geographical or religious isolation with subsequent inbreeding. This results in rare mutations becoming more common over the years (Ferla *et al*., 2007). This is also the case for the SA Afrikaner population. The Afrikaner is considered a unique homogeneous white population that originated from Dutch, French and German ancestors more than three hundred years ago. These European immigrants fled from Europe and settled at the Cape of Good Hope in 1652 and later, and due to geographical isolation, established a relatively isolated community (Tipping *et al*., 2001; Greeff, 2007; van der Merwe *et al*., 2011).

Screening of the two familial BC genes resulted in the identification of the first three founder mutations within the Afrikaner (Reeves *et al.*, 2004). They performed a study on 90 Afrikaner breast and ovarian cancer families containing three or more affected individuals. Two founder mutations were detected within exon 11 of *BRCA1*, namely 1493delC (c.1374del, p.Asp458GlufsX17) and E881X [c.2641G>T, p.Glu881X (2760G>T)]. Haplotype analysis revealed that these mutations originated from a single mutational event. Both mutations were only recorded for the Caucasian Afrikaner population and are internationally unique to SA (Reeves *et al*., 2004). The founding couple for 1493delC was Pieter Louw and Elisabeth Wendels. Pieter Louw's father, Jan Pietersz came to the Cape of Good Hope as a soldier of the Dutch East Indian Company from the Netherlands.

He married Beatrix Weijman who was an orphan from Holland. They had three sons and two daughters of which two sons married, one of them being Pieter

Louw. Pieter and Elizabeth had 10 children of which only two can be linked to the *BRCA1* 1493delC mutation (NC van der Merwe, personal communication).

The founding father of the E881X mutation was Hercules des Prez (du Preez) who was born in France and is believed to be the forefather of the SA du Preez family. He married Cecilia d'Athis. Both left France for Holland after the ruling of King Louis XIV to give partial religious freedom to the Protestants (Edict of Nantes). They then fled from Holland to the Cape of Good Hope in 1688 in fear of a war breaking out. They had six children of which four can be linked to the *BRCA1* E881X mutation (NC van der Merwe, personal communication). The third and only *BRCA2* founder mutation, 8162delG (c.7934del, p.Arg2645AsnfsX3) was observed in exon 17. This mutation is the most common within this population group and accounts for the majority of families (42%).

Although these founder mutations explain the majority of all the high risk Afrikaner BC families, there are still families for which a specific mutation has not yet been identified. Furthermore, when the mutation carriers within the various families carrying an identical mutation are compared, pertinent differences in the age at onset have been observed. For this reason it can be hypothesized that both genetic and environmental factors modify the penetrance of the disease-causing mutations in these genes.

2.8 Breast cancer predisposing genes and genetic modifiers

A significant portion of familial BC is not associated with the highly penetrant *BRCA1* and *BRCA2* mutations or other known BC predisposing genes. This suggests the remaining risk could be attributed to other less penetrant genes or a polygenetic model where the risk is conferred by a large number of low penetrance alleles, each contributing a small risk or interacting with other genetic and/or environmental factors (Antoniou *et al.*, 2002; Pharoah *et al.*, 2002; Wooster and Weber, 2003). Such a model might explain the phenotypical differences seen amongst mutation carriers, all carrying an identical *BRCA* mutation.

Breast cancer susceptibility genes can be divided into three groups namely, high, moderate and low penetrance genes. Genes that are considered to have a high penetrance, in addition to *BRCA1* and *BRCA2*, include the tumor suppressor p53 gene (*Tp53*) (OMIM 191170), the phosphatase and tensin homolog gene (*PTEN* [OMIM 601728]) and a serine/threonine kinase gene (*STK11* [OMIM 602216]) (Garcia-Closas and Chanock, 2008; Stratton *et al*., 2008) (Fig. 2.4). Inherited mutations in *Tp53* causes Li-Fraumeni syndrome which is associated with early onset and often bilateral breast tumours (Malkin *et al.*, 1990). Cowden syndrome is caused by mutations in the *PTEN* gene, which increases the chances of developing tumours in the thyroid, breast, skin and gastro-intestinal tract (Tsou *et al.*, 1997). Peutz-Jeghers syndrome is caused by mutations in the *STK11* gene (Boardman *et al.*, 1998).

Various moderate penetrance genes have been shown to increase the risk for developing BC and OVC (Fig. 2.4). Mutations in the ataxia-telangiectasia mutated gene (*ATM* [OMIM 208900]) lead to ataxia-telangiectasia (Khanna, 2000; Olsen *et al.*, 2001). The checkpoint kinase 2 (*CHEK2* [OMIM 604373]) (Lee *et al.*, 2000), partner and localizer of BRCA2 (*PALB2* [OMIM 610355]) and BRCA1 interacting protein (*BRIP* [OMIM 605882]) genes are all moderate penetrance genes involved in BC risk and together play a role in the cellular response to DNA damage (Dapic *et al.*, 2005).

The altered function of low penetrance genes due to the presence of polymorphisms may affect the gene-environment and gene-gene interactions, thereby increasing or decreasing the risk for BC development (Peto, 2002).

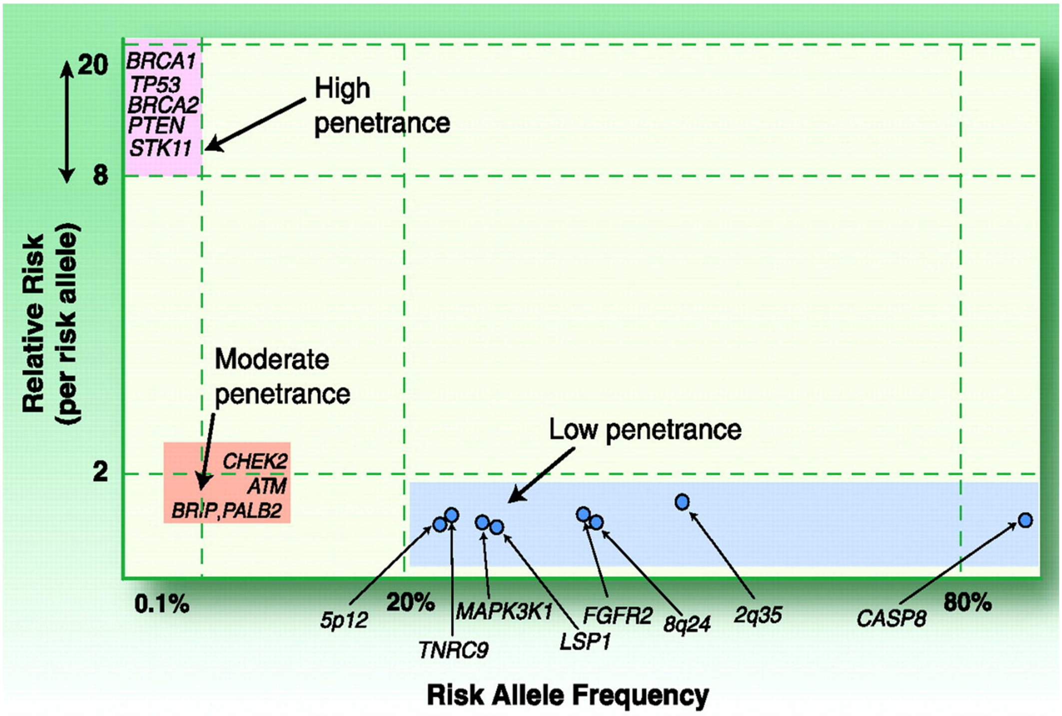


Figure 2.4 A schematic presentation indicating significant low, moderate and high penetrance BC susceptibility genes (Garcia-Closas and Chanock, 2008).

Low penetrance genes that play an important role in BC risk include *FGFR2* (OMIM 176943), *2q35*, caspase 8 (*CASP8* [*OMIM* 601763]), *MAP3K1* (OMIM 600982), *TNRC9* (OMIM 611416), *8q24, 5p12* and *LSP1* (OMIM 153432) (Stratton *et al.*, 2008) (Fig. 2.4).

2.9 The search for low penetrance genes

The search for low penetrance genes utilizes different research approaches. Candidate gene studies make a selection of low penetrance genes that are involved in the biochemical and physiological pathways of carcinogenesis. Candidate genes in these pathways range from the detoxification of environmental carcinogens to steroid hormone metabolism, DNA damage repair and cell cycle checkpoints (Rebbeck, 2002; Garcia-Closas and Chanock, 2008). The *ESR1* (OMIM 133430) gene is one of the possible low penetrance candidate genes that are involved in BC risk (Siddig *et al*., 2008).

Association studies, which compare frequencies of genetic polymorphisms, are based on selected candidate genes suspected to be important in carcinogenesis. Early association studies involved a limited number of polymorphisms and have largely been unsuccessful in identifying robust associations (Varghese and Easton, 2010). Recent advances in whole-genome SNP analysis have led to a number of GWAS in BC (Easton *et al*., 2007; Eeles *et al*., 2008). Unlike candidate gene studies, GWAS studies do not depend on prior knowledge regarding the genes. Some genes that the GWAS studies have indicated as potential role players in the development of BC include the *FGFR2* (OMIM 176943), *LSP1* (OMIM 153432), *MAP3K1* (OMIM 600982) and *TNRC9* (OMIM 611416) (Easton *et al*., 2007).

2.10 Epidemiology of breast cancer

Several risk factors and modifiers can increase the incidence of BC in *BRCA* mutation carriers and the general population. The strongest risk factors include a family history, age and reproductive history. Antoniou and co-workers (2003) suggested an increase risk for BC and OVC due to changing patterns involving reproductive factors, such as age at first pregnancy, oral contraceptive use and breast feeding whereas Dumitrescu and Cotarla (2005) suggested environmental influences and lifestyle habits.

Factors that relate to reproductive history, including age of menarche, age at menopause, parity, age at first full-time pregnancy (FTP) and breastfeeding have been shown to influence BC risk (Kelsey *et al.*, 1993; Key *et al*., 2001). Henderson *et al*. (1985) proposed that BC risk is associated with the number of ovulatory cycles. Therefore women with early-onset menarche (<12) or late menopause (> 55 years of age) have an increased risk of developing BC. An early age at menarche expose breast epithelium to higher levels of estrogens for a longer time. According to Hunter *et al*. (1997), there is a 5% reduction in BC risk for every one-year delay in the onset of menarche. A late age at menopause will result in more ovulatory cycles, thus increasing the risk for BC (Welcsh *et al.*, 2001). A collaborative group on hormonal factors in BC reported that every one-year delay in the onset of menopause increase BC risk with 3% (Lacey *et al*., 2009). The surgical removal of the ovaries in order to induce menopause before the age of 45, is an attempt to reduce BC risk by the removal of estrogen.

Parity and an early age of first FTP are associated with a reduced BC risk. MacMahon and co-workers (1970) were the first to demonstrate the importance of age at FTP. Women who have their FTP before the age of 25 have a lower risk of developing BC compared to women who had their FTP after the age of 30 or nulliparous women. A dual effect was observed for *BRCA* mutation carriers. A reduced risk for BC was observed for *BRCA1* mutation carriers with the first pregnancy over the age of 30 years compared to first pregnancies before the age of 20 (Andrieu *et al*., 2006). However, BC risk increases for *BRCA2* mutation carriers with a later age of FTP. Andrieu *et al.* (2006) also indicated that multiple pregnancies reduce the risk in *BRCA* mutation carriers with 14% for every additional birth. The protective effect of pregnancy is still not fully understood. During pregnancy the breast parenchyma cells are in a stable state thereby lessening proliferation in the second half of the menstruation cycle. The expression of the *BRCA1* gene is also upregulated, limiting proliferation and promoting differentiation (Mueller and Roskelley, 2003). Differentiation of mammary gland cells at an early age further more render the cells less susceptible to BC development (Russo *et al.*, 1982).

Women representing the general population who breastfed, have a decreased risk of developing BC (Russo *et al.*, 2001). It has also been suggested that the longer a women breastfeed, the more they are protected against BC (Collaborative group on hormonal factors in breast cancer, 2002). In one study, breast feeding reduced the risk in *BRCA1* mutation carriers but not in *BRCA2* mutation carriers (Jernström *et al.*, 1998) where as another study indicated a reduction in the risk for all *BRCA* mutation carriers (Andrieu *et al.*, 2006).

Exposure to exogenous hormones such as oral contraceptives (OC) and hormone replacement therapy (HRT) can also increase BC risk. A meta-analysis study using data from 54 epidemiological studies reported that the current use of OC increases BC risk with 24% (Collaborative group of hormonal factors in breast cancer, 1996). Furthermore they reported that 10 years after OC usage was stopped, the risk returned to the same level as if it was never used. For patients with a family history of BC, OC use increases the risk three–fold (Grabrick *et al.*, 2000). Among *BRCA1* and *BRCA2* mutation carriers, the risk is higher compared to the general population (Ursin *et al*., 1997). The use of HRT for more than five years increases BC risk but the risk disappears five years after termination of use (Vecchia *et al.*, 2001). The use of HRT have a 2.3% increase in the relative risk

(RR) of BC for each year it is used (Collaborative group of hormonal factors in breast cancer, 1996).

Lifestyle factors have also been linked to BC. These include smoking, alcohol consumption, diet and obesity. Evidence on the association between smoking and BC risk remains controversial. Various studies have suggested a positive link between smoking and BC, while other studies have found no association or even a protective effect (London *et al*., 1989; MacMahon *et al*., 1990; Collaborative group on hormonal factors in breast cancer, 1996; Brunet *et al*., 1998). In 2004, Reynolds *et al*. reported that BC risk increases significantly among active smokers compared to non-active smokers. The risk increases among women who started to smoke at a young age or started smoking at least 5 years before the FTP, and among women who smoked for longer or with a greater intensity.

It has also been suggested that smoking could decrease BC risk. According to Brunet (1998), cigarette smoking has a protective effect among women with germline *BRCA1* and *BRCA2* mutations who smoked one package of cigarettes (20 cigarettes) daily for one year (pack-year) [odds ratio (OR) = 0.46 for four or more pack-years versus non-smoker]. However, it has been shown that the carcinogenic effect of smoking increases BC risk as tobacco smoke can pass through the alveolar membrane into the blood stream (Yamasaki and Ames, 1977). The tobacco smoke is then transported to the breast where it is metabolized and activated by mammary epithelium cells (Plant *et al.*, 1985). It has also been suggested that carcinogens induce DNA mutations that can later manifest as cancer, but the anti-estrogenic effect of smoking inhibits the growth of existing tumours. The explanation for the discrepancies in previous studies includes the postulated anti-estrogenic effect of cigarette smoking (Baron *et al.*, 1990). This may lead to an early menopause and thus fewer years of menstruation, low levels of urinary estrogens and lower body weight (Kaufman *et al*., 1980; Willett *et al*., 1983; Istvan *et al*., 1992).

In most studies alcohol consumption is associated with an increased BC risk. The Collaborative group on hormonal factors in breast cancer (2002) found a 7.1% increase in the RR of BC for each additional 10 g intake of alcohol per day. The most likely mechanism by which alcohol increases BC risk is by increasing estrogen and androgen levels (Singletary and Gapstur, 2001). Feron *et al*. (1991) suggested that alcohol consumption increases the exposure to carcinogenic metabolites.

Diet, sedentariness and obesity have been proposed as risk factors for BC. A study in Norway found that women who followed a diet with a high fish intake and low meat consumption were at a reduced risk for BC (Vatten *et al*., 1990). Animal fat intake that is rich in polyunsaturated fatty acids can cause mutagenic free radicals (Bartsch and Nair, 2004). Diets rich in fruits and vegetables decrease BC risk as they contain antioxidants. Increased physical activity and the absence of obesity during adolescence are associated with BC onset delay in *BRCA1* and *BRCA2* carriers according to King (2003). This is consistent with the fact that physical activity from a young age delays puberty and reduces ovulation.

Overweight and obesity is regarded as a body mass index (BMI) ≥ 25 by the World Health Organization (WHO). A study performed by Calle *et al*. (2003) searched for associations between overweight and obese patients and the mortality rate over a 16 year period. They found that the mortality rate of obese women was double that of normal women. Another study found that obese women with a family history of BC have an increased risk compared to slimmer women with a family history of BC (Carpenter *et al*., 2003). The correlation between obesity and BC in obese women who have an increased risk for postmenopausal but not pre-menopausal BC (Lahmann *et al*., 2004), is that there is an elevated circulation of estrogens from peripheral aromatization of androgens in adipose tissue. In obese postmenopausal women, estrogen biosynthesis is mainly from the breast, abdomen, thighs and buttocks with levels of aromatase increasing with age and BMI (Grodin *et al*., 1973; Lorincz and Sukumar, 2006).

2.11 Estrogen

Estrogens are steroid hormones that have an effect on the female reproductive system. There are three types of estrogen: estrone (E_1), estradiol (E_2) and estriol (E_3). Estradiol is the most active endogenous estrogen in breast tissue (Sowers *et al*., 2006). These hormones are synthesized mainly by the ovaries and testes where after it is released into the blood. In premenopausal women, the ovaries are the main source of estrogen whereas in postmenopausal women estrogen is synthesized in smaller amounts by the ovaries and adipose tissue and converted from circulating androgens to estrogens (Simpson, 2003).

The link between estrogen and BC has long been studied as it plays a key role in reproductive factors. It also has physiological functions in the cardiovascular, immune and other systems (Liu *et al*., 2003). These effects are all mediated by a ligand-activated transcription factor, namely the estrogen receptor (ER). This biological effect was first described by Jensen and Jacobsen (1962). One of the earliest studies to indicate estrogen as a risk factor for BC was performed in 1896 by Sir George Beatson. He discovered that an oophorectomy may lead to breast tumor regression (Beatson, 1896). Interaction between estrogen and the ER increases the proliferation of target cells. Due to this reason, estrogen deprivation is still the best treatment for endocrine responsive tumors (Pearce and Jordan, 2004). More than 70% of *BRCA1* breast tumours are ER-negative whereas *BRCA2* and sporadic BC tumours mostly are ER-positive (Noruzinia *et al.*, 2005).

2.11.1 Estrogen receptors (ER)

The effects of estrogen are mediated through the ER. The ER is found in the nucleus and has a molecular mass of ~65,000 daltons. These receptors belong to the super family of ligand inducible transcription factors (Hall *et al*., 2001; Boyapati *et al*., 2005) that regulate transcription with co-regulators through binding to the DNA enhancer element which is located in the promoter of the target gene (Heldring *et al*., 2007).

There are two estrogen receptor subtypes, namely ERα (ESR1) and ERβ (ESR2). They both bind estrogen and other agonists and antagonists. *ERα* was the first estrogen receptor gene to be cloned by Walter and co-workers (1985) and sequenced by Greene and his team (1986). The *ERα* - gene is located on chromosome 6q24-q27 and has a coding region of 1785 nucleotides. It is translated into a 595 amino acid protein (Greene *et al*., 1986). *ERβ* is located on chromosome 14q22-q24 and was cloned by Kuiper and co-workers (1996) and characterized by Mosselman *et al*. (1996).

ERs are distributed in various parts of the body (Fig. 2.5). ERα and ERβ are found in the breast, brain, cardiovascular system and bone. ERα is predominantly found in the endometrium, breast cells, ovarian stroma cells and in the hypothalamus, whereas ERβ is found in kidney, brain, lung, bone, heart, intestinal mucosa, prostate and endothelial cells (Pearce and Jordan, 2004). The action of steroid hormones is mediated through binding to the estrogen receptors, ERα and ERβ.

As cellular environments change, ERs can bind different co-factors depending on the cellular environment and binding affinities. Estrogen receptors act through the cell membrane and cytoplasm, which is involved in the transduction of the non-genomic actions of estrogen (Simoncini *et al*., 2003). Co-activators activate target gene transcription by not binding to the DNA. They are recruited to target gene promoters through protein-protein interactions with the ER. Co-activators include steroid receptor co-activator 1 (SRC-1), SRC-2, SRC-3, p300 and the CREB-binding protein (CBP). Negative co-regulators and co-repressors inhibit gene activation and turn off activated target genes.

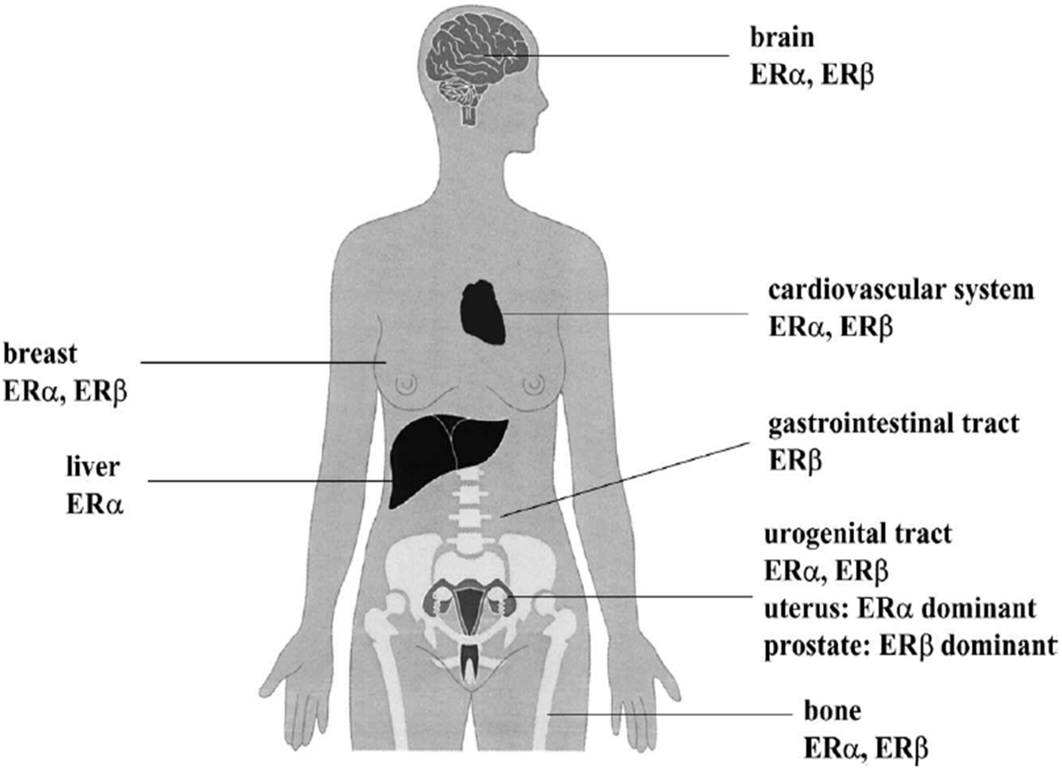


Figure 2.5 Distribution for ERα and ERβ in the body (Pearce and Jordan, 2004).

Heldring *et al*. (2007) suggested several molecular pathways where estrogen and ER play a role in biological processes (Fig. 2.6). The direct pathway entails ligand activation and direct DNA binding to estrogen response elements (ERE) before the modulation of gene regulation. The tethered pathway affects gene regulation by indirect binding through the interaction with other transcription factors.

The non-genomic pathway has a rapid cascading effect which is activated through the second messenger (SM) after the ligand activates a receptor on the membrane or a signal activates the ER located in the cytoplasm. The SM affects ion channels or increase nitric oxide levels in the cytoplasm which leads to a response without involving gene regulation (Heldring *et al*., 2007). The ligand-independent pathway (Fig. 2.6) involves activation through other signaling pathways such as growth factor signaling. In this case, activated kinases phosphorylate ERs thereby activating them to dimerize to bind DNA with gene regulation as the end result.

As mentioned above, the hormones bind to the C-terminus of the receptors in the cytoplasm or nucleus. Where after the middle domain of the receptor in turn binds to the DNA. After binding, the hormone-receptor complex undergoes an activation process. The complex can enter the nucleus or if it is already in the nucleus it can bind to the ERE which is upstream from the basal promoter site at the 5' end of the gene (Fig. 2.7). This may be modulated by cofactors. Transcription of the primary gene messenger by RNA polymerase is induced or repressed which leads to different concentrations of proteins being formed, depending if the levels of the RNA molecules were raised or lowered which will affect the rate of translation of the mRNA (Fig. 2.7) (Levy *et al*., 2006).

2.11.2 Interaction between *BRCA* and *ESR1* genes

Estrogen has a dual role in the development of BC. It indirectly stimulates cell

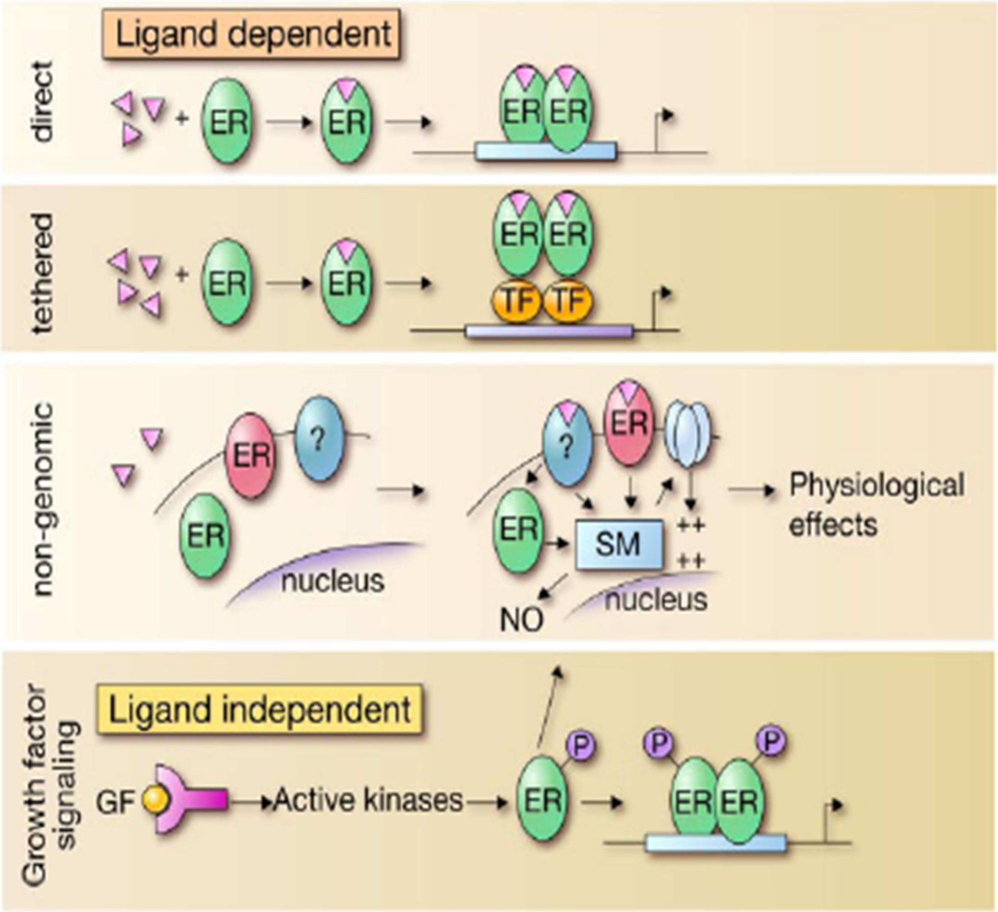


Figure 2.6 A schematic representation of the four molecular pathways of the estrogen receptors (Heldring *et al*., 2007).

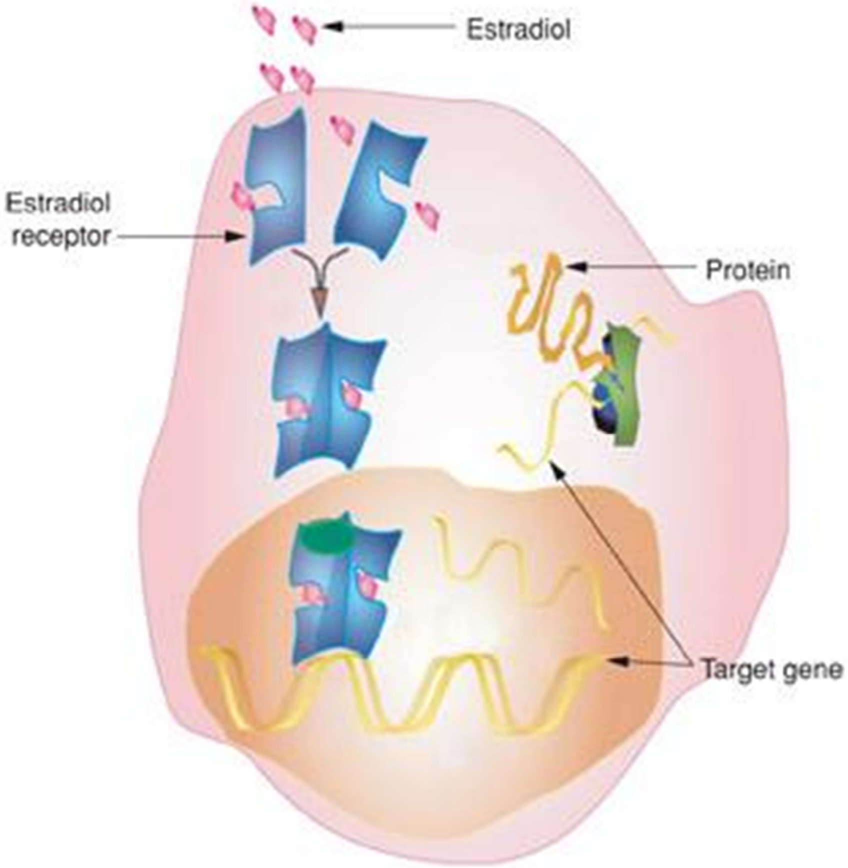


Figure 2.7 Binding of the hormone receptor complex to the ERE (Levy *et al*., 2006).

proliferation, which in turn could lead to the replication of DNA errors obtained during the process (Liehr, 2000; Noruzinia *et al*., 2005). The presence of DNA replication errors will induce activation of certain DNA repair enzymes such as BRCA1. It will form a complex to repair the DNA damage that occurred during cell division (Fig. 2.8).

If the DNA damage is not repaired, it could lead to uncontrollable cell proliferation and tumor formation. Fan and co-workers (1999) demonstrated that BRCA1 binds to ESR1 *in vitro* and *in vivo* and inhibited signaling by ESR. This inhibition is done by the estrogen-responsive enhancer element which blocks the transcriptional activation function AF-2.

Estrogen can promote malignancies in the breast by inducing direct and indirect free radical-mediated DNA damage, genetic instability and mutations in cells (Hilakivi-Clarke, 2000). Mote *et al*. (2004) proposed that disruption of the caretaker genes through mutations may increase the risk for cancer in mutation carriers. This is especially the case for hormone-sensitive tissue such as the breast, the ovaries as well as the prostate in men. With the loss of function of the caretaker genes, *ERα* transcription activity is uncontrolled and could result in continuous proliferation of genetically damaged mammary epithelial cells. This situation contributes to genetic instability which leads to uncontrolled proliferation and tumor progression.

The interaction between *BRCA2* and *ESR1* are not fully understood. The main effects of E_2 are mediated through ESR1 and ESR2, which is essential for the regulation of *BRCA2* transcription (Giguere *et al*., 1998; Pettersson and Gustafsson, 2001; Jin *et al*., 2008). According to a study done by Jin and colleagues (2008), *BRCA2* transcription becomes responsive when the promoter is stimulated with E_2. They reported that during E_2 treatment, ESR1 forms an activating transcriptional complex with CBP/p300, p68/p72 and MyoD.

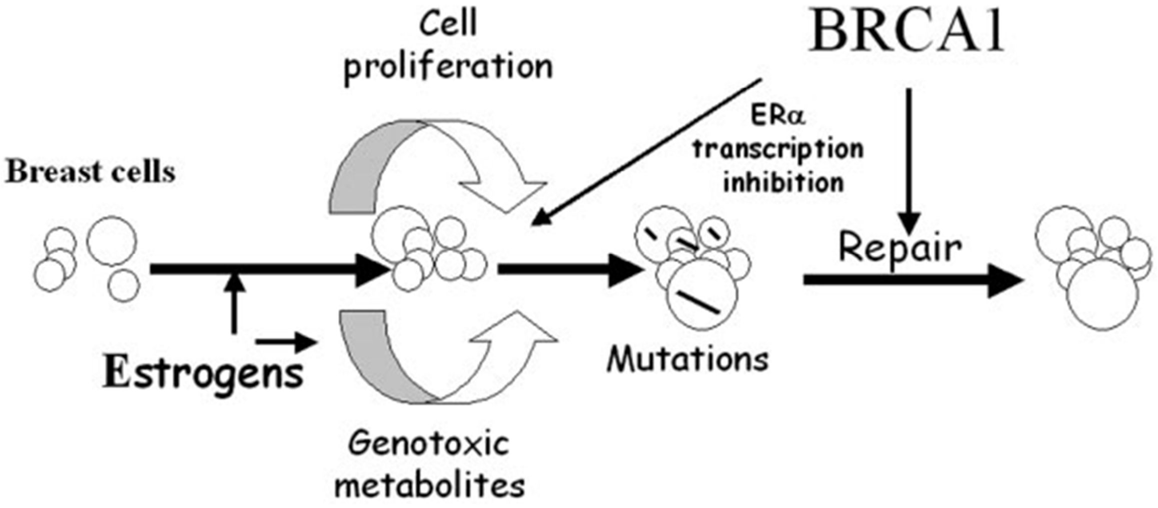


Figure 2.8 BRCA1 controlling cellular proliferation that is induced by E_2 and ER (Noruzinia *et al*., 2005).

This complex binds to the specificity protein 1 (Sp1) site on the *BRCA2* promoter and activates transcription by inducing histone acetylations (Fig. 2.9).

2.11.3 Polymorphisms in *ERα* and *ERβ*

Several studies have been performed on the relationship between polymorphisms in the *ERα* and *ERβ* genes and BC (Andersen *et al*., 1994; Cai *et al*., 2003; Wang *et al*., 2007). It has been proposed that SNPs in these two genes could be responsible for phenotypic variation in the *BRCA* mutation carriers.

Two specific polymorphisms in the *ERα* gene, namely *PvuII* (rs2234693) and *XbaI* (rs9340799) are currently the best studied, although the results are contradictory (Andersen *et al*., 1994; Cai *et al*., 2003; Wang *et al*., 2007). This may be explained by the ethnicity of the populations that were studied. The *PvuII* (rs2234693) polymorphism in intron 1 is represented by a C to T transition and is located 0.4 kb upstream of exon 2 (Heimdal *et al*., 1995; Kobayashi *et al*., 1996). There is an important correlation between *ER* expression in BC and age of diagnosis (Yaich *et al*., 1992). The homozygote genotype (TT) of *PvuII* was shown to be associated with a younger age of BC diagnosis (Cai *et al*., 2003). This study indicated an age-adjusted odds ratio (OD) for heterozygote (CT) and homozygote (TT) being 1.3 [95% CI, 1.0 – 1.7] and 1.4 (95% CI, 1.1 – 1.8) respectively compared to the homozygote (CC). However, a study of 360 BC patients from Norway found that allele frequencies of the *PvuII* polymorphism did not differ between cases and controls (Andersen *et al*., 1994).

The *XbaI* (rs9340799) polymorphism, also in intron 1, is located 50 bp from the *PvuII* restriction site. It is represented by an A to G transition and is located in the A/B region of the ligand-dependent trans-activation domain and could have an effect on the function of ERα (Wang *et al*., 2007). Wang (2007) reported an allelic protective effect for *XbaI* with an OR of 0.82 and 95% CI=0.68-1.00; (*P*=0.04).

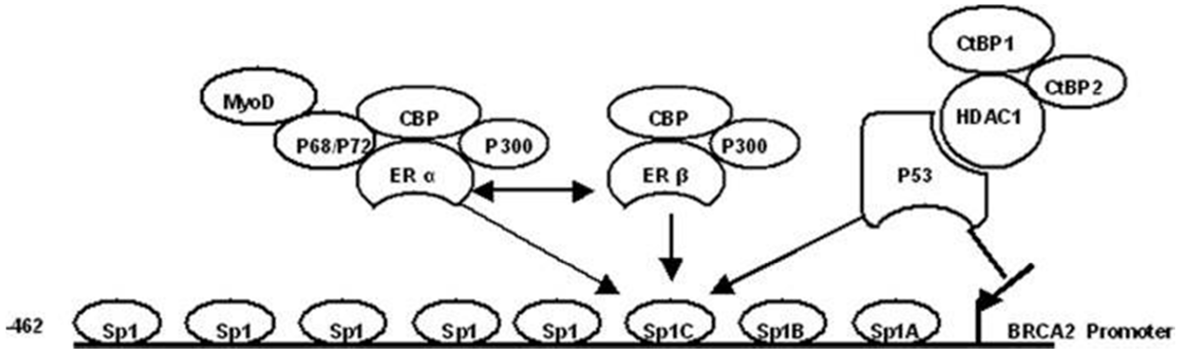


Figure 2.9 Illustration showing the binding of ESR1 complex to the Sp1 sites on the *BRCA2* promoter adapted from Jin *et al*. (2008).

A Korean study (Shin *et al*., 2003) reported a decrease of BC risk associated with the G-allele (OR=0.4; 95%CI: 0.3-0.6). These results were corroborated by a Norwegian study reporting the A allele of *XbaI* as the risk allele (Weiderpass *et al*., 2000). A Shanghai BC study including 1069 cases and 1169 age-matched controls reported that the *XbaI* polymorphism was associated with a non-significant elevated risk in postmenopausal women (Cai *et al*., 2003).

The OR for genotype AG and AA were 1.2 (95% CI: 0.7-1.9) and 1.3 (95% CI: 0.8-2.0), respectively. Both *PvuII* and *XbaI* are associated with Alzheimer disease, obesity, multiple sclerosis, endometriosis, adenomyosis, leiomyomas and bone mineral density (Cai *et al*., 2003).

The Breast and Prostate Cancer Cohort Consortium reported on four polymorphisms (rs4986938, rs1256049, rs1256031, rs3020450) in the *ERβ* gene of which none showed a significant association with BC. According to Gold and coworkers (2004), three SNPs (ESR1002, rs2077647, rs827421) showed significant association with disease in the Ashkenazi Jewish population specifically.

2.12 Trinucleotide repeat-containing 9 (*TNRC9*) gene or OMIM 611416

The *TNRC9* gene is located at chromosome 16q12 and also known as tox high mobility group box family member 3 (*TOX3*). The function of TNRC9 remains unclear. It contains a putative high mobility group (HMG) box motif, suggesting it may act as a transcription factor that was implicated in BC metastasis to the bone (Smid *et al*., 2006).

Transcription, replication and DNA strand repair requires the bending and unwinding of compacted chromatin structures. According to O'Flaherty *et al*.

(2003), TNRC9 may also be involved in bending and unwinding of DNA and thereby alteration of chromatin structure. The C-terminus of TNRC9 contains trinucleotide repeats that codes for glutamine. Polyglutamine repeats have been associated with transcription factors (Gerber *et al*., 1994).

2.12.1 Polymorphisms in *TNRC9*

The rs3803662 SNP is located near the 5' end of *TNRC9* and has been proven to be associated with BC (Easton *et al*., 2007). The SNP entails a single nucleotide change from a C to a T allele. The variant T allele has been implicated in an elevated risk of BC and a younger age at onset (Huijts *et al*., 2007). A GWAS of 1,600 Icelandic BC cases and 11,563 controls followed by a replication of 4,554 affected individuals and 17,577 controls, showed that rs3803662 is associated with an increased ER positive BC risk (Stacey *et al*., 2007). Seven percent of individuals of European descent are homozygous for the variant T allele and had a 1.64-fold greater risk.

A study by Garcia-Closas and co-workers (2008) on patients of European, North American, South-East Asian and Australian descent concluded that the rs3803662 SNP was associated with BC risk. This was confirmed by a meta-analysis study by Chen *et al*. (2010), comprising of a total of 25,828 cases and 36,177 controls, including Caucasian women from British (Latif *et al*., 2010), Russian (Gorodnova *et al*., 2010) and German descent (Hemminki *et al*., 2010) as well as Han Chinese (Liang *et al*., 2010), Caucasian (Antoniou *et al*., 2008), Asian (Li *et al*., 2009) and Western European descent (Tapper *et al*., 2008).

The analysis concluded that the variant T allele is a low-penetrant BC risk factor. A case control study found no association with BC among the Chinese population with a *P*-value of 0.151 (Liang *et al*., 2010). Huijts *et al*. (2007) reported that rs3803662 was associated with a younger age at diagnosis with the mean age of 54.3, 53.4 and 52.5 years for the patients homozygous for the ancestral allele, heterozygous patients and patients homozygous for the variant allele, respectively. However, the difference between these values was not significant (*P* = 0.199).

2.13 Lymphocyte-specific protein 1 *(LSP1)* gene (OMIM 153432)

LSP1 found in both mice and humans codes for an intracellular F-actin binding cytoskeletal protein (Jongstra-Bilen and Jongstra, 2006). It was thought that *LSP1* expression was restricted to B cells, functional T cells and thymocytes, but more recently it has been documented to be expressed in monocytes, macrophages, neutrophils, lymphocytes and endothelium (Jongstra *et al*., 1994; Li *et al*., 1995; Liu *et al*., 2005; Petri *et al*., 2011). LSP1 may regulate neutrophil motility, adhesion to fibrinogen matrix proteins and trans-endothelial migration (Liu *et al*., 2005). It has also been shown to be a major substrate of the (MAPK)-activated protein (MAPKAP) kinase-2 in the p38 MAPK pathway, suggesting that *LSP1* might be important in chemotaxis (Huang *et al*., 1997).

2.13.1 Polymorphisms in *LSP1*

The rs3817198 SNP located in intron 10 of *LSP1* entails a single nucleotide change from a T to a C allele. A study by Antoniou and co-workers (2009) evaluated the association between rs3817198 and BC risk in 9442 *BRCA1* and 5665 *BRCA2* mutation carriers from 33 study centers. The variant C allele was associated with an increased BC risk for *BRCA2* mutation carriers specifically (95% CI: 1.07-1.25, *P*-value = 2.8×10^{-4}). However, a study from the west of Ireland with mostly European ancestry patients showed no association with BC risk for this population (*P*-value = 5.4×10^{-1}) (Mcinerney *et al*., 2009).

2.14 Mitogen-activated protein kinase kinase kinase 1 (*MAP3K1*) gene (OMIM 600982)

The mitogen-activated protein kinases (MAPKs) are serine-threonine kinases that transduce a large variety of external signals, leading to a wide range of cellular responses including growth, differentiation, inflammation and apoptosis (Zhang and Liu, 2002). The *MAPK* signaling pathway has been found to play a role in the initiation and pathogenesis of BC (Sivaraman *et al*., 1997; Coutts and Murphy, 1998). Four MAPK signaling cascades are implicated in breast disease in mammalian cells. These include the extracellular regulated kinase (ERK) 1/2 pathway, the c-Jun N-terminal kinase (JNK) pathway, the p38 pathway and the ERK5 pathway (Wang and Tournier, 2006; Krishna and Narang, 2008). Each MAPK pathway consists of a three-tiered kinase cascade, a MAPK kinase kinase (MAP3K), a MAPK kinase (MAP2K), and a MAPK (Ferrell, 1996). *MAP3K1* encodes the MAPK kinase kinase that phosphorylates and activates MAP2K. This process phosphorylates and activates the MAPK/ERK which produces downstream signaling effects on various genes (Rebbeck *et al*., 2009).

Malignant epithelial cells in the breast and metastatic cells in lymph nodes demonstrate an activated and hyper-expressed *MAPK* pathway (Sivaraman *et al*., 1997*)*. The multiple approaches to show constitutive phosphorylation of MAPK that correlates with its activation and immunochemistry showing its hyper expression, strongly support the belief that activated and amplified *MAPK* expression can contribute to initiation and the metastatic potential of BC.

2.14.1 Polymorphisms is *MAP3K1*

The SNP rs889312 is located in a linkage disequilibrium (LD) block of approximately 280 kb and entails a single nucleotide change from an A to a C allele. According to a study by Antoniou *et al*. (2008), the variant C allele showed a positive association between *BRCA2* mutation carriers and an increased risk of BC. The MAPK pathway has also been strongly linked to human epidermal growth factor receptor (HER) receptor activity and associated with HER2 positive breast tumours (Bild *et al*., 2006; Creighton *et al*., 2006). A meta-analysis of seven studies including 26,015 cases and 33,962 controls was performed to determine an association between BC risk and rs889312. They observed a significant correlation between BC risk and the variant C allele when all data of the various studies were pooled into the meta-analysis (OR 1.09, 95% CI 1.07-1.12) (Lu *et al*., 2010).

2.15 Fibroblast growth factor receptor 2 (*FGFR2*) gene (OMIM 176943)

The fibroblast growth factor receptor 2 (FGFR2) is a tyrosine kinase receptor located on chromosome 10q12 and consists of at least 22 exons (Ingersoll *et al*., 2001). *FGFR2* is a member of the FGFR family of distinct fibroblast growth factor receptor genes that encodes the fibroblast growth factor receptor protein. This*FGFR* family consists of 5 members *FGFR1, FGFR2, FGFR3, FGFR4* and *FGFR5*, some of which generate multiple products via alternative splicing (Klint and Claesson-Welsh, 1999; Kim *et al*., 2001).

It encodes two isoforms, FGFR2b and FGFR2c, due to alternative splicing of exons. *FGFR2b* is predominantly expressed in epithelial cells whereas *FGFR2c* is expressed in mesenchymal cells (Katoh, 2008). FGFR2 can transform human mammary epithelial cells and inhibition of FGFR2 signaling can inhibit BC tumor cell proliferation (Koziczak *et al*., 2004). Expression of *FGFR2* isoforms transforms BC cells by sustained signal transduction (Raskin *et al*., 2008).

Docking proteins FRS2 and FRS3 are tyrosine phosphorylated by FGFRs to recruit and activate the Grb2/Sos1 complex, which then interacts with Ras to activate the MAPK pathway (Kouhara *et al*., 1997; Ong *et al*., 2000). This signaling pathway has been shown to contribute to FGFR-mediated cell proliferation and migration (Klint and Claesson-Welsh, 1999; Boilly *et al*., 2000).

FGFRs mediate their biological effects through the four receptor tyrosine kinases, FGFR1 – FGFR4. These four kinases are compassed of an extracellular ligandbinding domain, a transmembrane segment, tyrosine kinase domain in the cytoplasm and regulatory sequences (Klint and Claesson-Welsh, 1999; Eswarakumar *et al*., 2005). The extracellular ligand binding domain consists of three immunoglobulin-like (Ig-like) domains (D1, D2 and D3), seven to eight acidic residues between D1 and D2 (acidic box) and a conserved region in D2 that is a binding site for heparin. The Ig-like domains and cytoplasmic tyrosine kinase domain form a high affinity complex with fibroblast growth factors (FGFs) (Klint and Claesson-Welsh, 1999; Eswarakumar *et al*., 2005). This complex plays an important role in cellular signaling by activating intracellular tyrosine kinase. This in turn activates signal transduction through direct phosphorylation of adaptor proteins (Nan *et al*., 2009).

FGFRs are involved in many biological processes and play important roles in cell growth, invasiveness, motility and tumor genesis (Liang *et al.,* 2008). FGFs are involved in embryogenic development and in the control of the nervous system, tissue repair, wound healing and in tumor angiogenesis (Givol *et al*., 2003). In human BC, expression of *FGFR2* is estrogen receptor (ER)-dependent and correlates with a lower rate of apoptosis. *FGFR2* is amplified and over expressed in 5-10% of breast tumours (Adnane *et al.,* 1991). This correlates with a poor prognosis and survival rate. Studies by Bane *et al*. (2009) showed that *FGFR2* was highly expressed in *BRCA2*-associated cancers compared to *BRCA1*-associated BC. Defects in *FGFR2* are also the cause of craniosynostosis syndromes like Crouzon syndrome (CS) (Reardon *et al*., 1994), Jackson-Weiss syndrome (JWS) (Jabs *et al*., 1994), Apert syndrome (APRS) (Wilkie *et al*., 1995) and Pfeiffer syndrome (PS) (Lajeunie *et al*., 1995; Rutland *et al.*, 1995).

2.15.1 Polymorphisms in *FGFR2*

The most significant SNP to be associated with BC in *FGFR2* is rs2981582 according to a GWAS study done by Easton and co-workers (2007). This SNP entails a single nucleotide change from a C to T. Two recent (GWAS) studies by Easton *et al*. (2007) and Hunter *et al*. (2007) demonstrated association between SNPs in *FGFR2* and an increased risk for BC. The association was confirmed by Huijts *et al*. (2007) using a candidate-gene approach. The two-stage genome wide study by Easton *et al*. (2007) was conducted using 4,398 BC cases and 4,316 controls. A third stage followed in which 30 SNPs were tested for confirmation in 21,860 cases and 22,578 controls from 22 studies. They used 227,876 SNPs that correlated with 77% of known SNPs in Europeans and only five independent loci showed strong association with BC. The association between *FGFR2* (rs2981582) and BC risk was demonstrated in a study by Antoniou and co-workers (2008). The association was restricted to *BRCA2* mutation carriers with a *P*-value of 2×10^{-8}.

2.16 Genotyping

When selecting mutation detection techniques, it is important to consider the cost-effectiveness, reliability and accuracy of the technique. Various PCR based techniques have been developed for the detection of single base substitutions, small deletions or insertions. These include older techniques such as restriction fragment length polymorphism (RFLP) and DNA sequencing to the latest technology such as Taqman® analysis with real-time PCR (qPCR).

2.16.1 Restriction fragment length polymorphism (RFLP)

RFLP is widely used for the detection of variations in DNA sequences (Botstein *et al*., 1980). The DNA sequences are digested into restriction fragments using suitable endonucleases that are able to detect the specific polymorphism. The endonucleases only cut the DNA molecule where the DNA sequences are recognized by the enzyme (Botstein *et al*., 1980). After digestion, the DNA fragments are separated using horizontal gel electrophoresis based on the size of the fragments. Genotype calling is made by comparing band profiles as different mutations will produce fragments of various lengths which will indicate ancestral and variant alleles (Strachan and Read, 2003).

2.16.2 TaqMan® analysis

The TaqMan® oligonucleotide probes are hydrolysis probes that was first reported by Holland *et al*. (1991) after which Roche Molecular Diagnostics developed it for Applied Biosystems as diagnostic research assays.

This technology is based on the 5’-3’ exonuclease activity of *Taq* (*Thermus aquaticus*) polymerase. TaqMan® oligonucleotide probes are dual-labeled with a fluorescence reporter at the 5’ end and a non-fluorescent quencher at the 3’ end (Fig. 2.10). When the fluorophore is excited by the thermal cycler’s light source through fluorescence resonance energy transfer (FRET), the quencher will absorb the fluorescence as long as they are in close proximity. Different fluorophores that can be used include: 6- carboxyfluorescein (FAM), tetrachlorofluorescin (TET), 6-hexachlorofluorescein (HEX) and 2'-chloro-7'-phenyl-1,4-dichloro-6-carboxyfluorescein (VIC). The most common quenchers include: tetramethylrhodamine (TAMRA) and dihydrocyclopyrroloindole tripeptide (minor groove binder [MGB]).

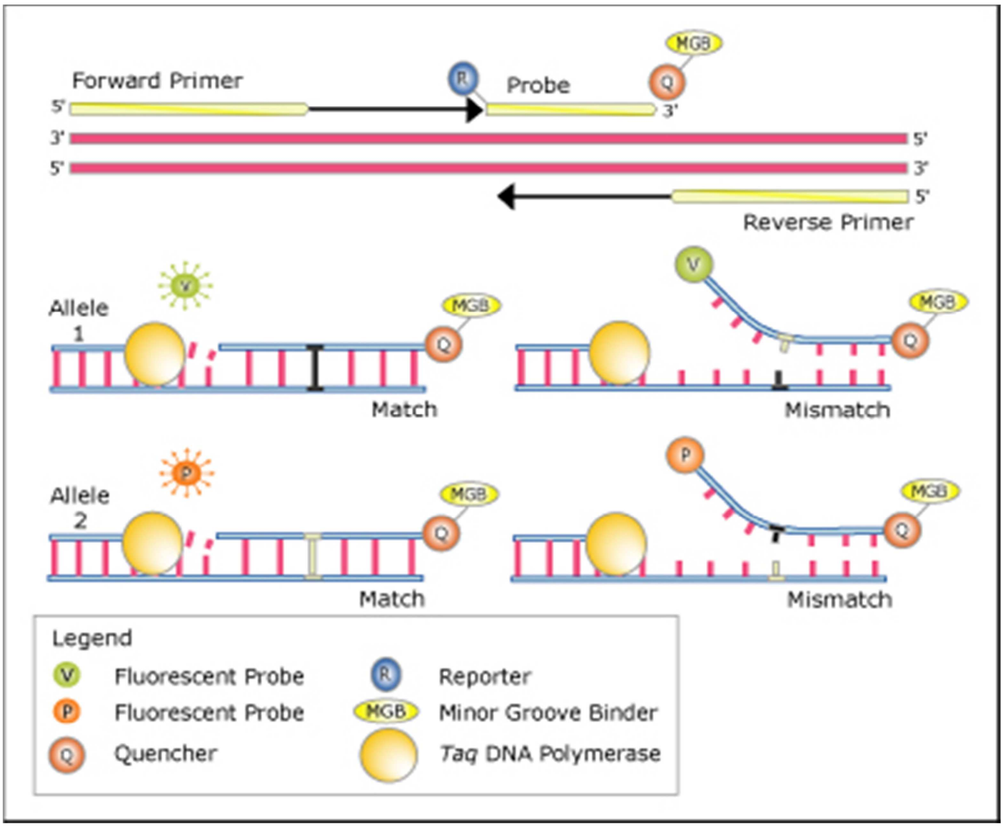


Figure 2.10 The principles of TaqMan® probe technology indicating the fluorescence reporter, MGB quencher and PCR extension phase (http://servicexs.com).

The TaqMan® probe binds to the specific target sequence during the extension phase of PCR, where after the 5' reporter is displaced by the 5-3'' exonuclease activity of Taq polymerase. This cleaves the reporter from the quencher allowing the reporter to emit fluorescence when excited by FRET which can then be detected in real time by qPCR thermal cycler (Fig. 2.10).

This technology enables discrimination between two alleles present within an individual. This is done by using two differently labeled probes, one representing the ancestral allele and one the variant allele. The quantification is done by the computer software which constructs an amplification plot using the fluorescence emitted by the reporter (Fig. 2.11). This plot indicates the fluorescence signal versus cycle number. A baseline is set for the non-significant changes in fluorescence that is detected. This is done automatically by the software for cycles 3 to 15 but it can be done manually. An increase above the baseline indicates a real signal for the PCR product. The Cq or Ct (cycle threshold) value represents the cycle number at which the fluorescence passes the threshold. The Cq value is determined by the amount of product, a small amount of template will generate a high Cq value because more amplification cycles are needed for the fluorescence signal to rise above the threshold. ΔRn is determined by the software. It is calculated as Rn = Rnf – Rnb, where Rnf is the fluorescence emission intensity of the reporter and Rnb is the fluorescence emission of the baseline (fluorescence emission of the passive reference dye) (Arya *et al*., 2005).

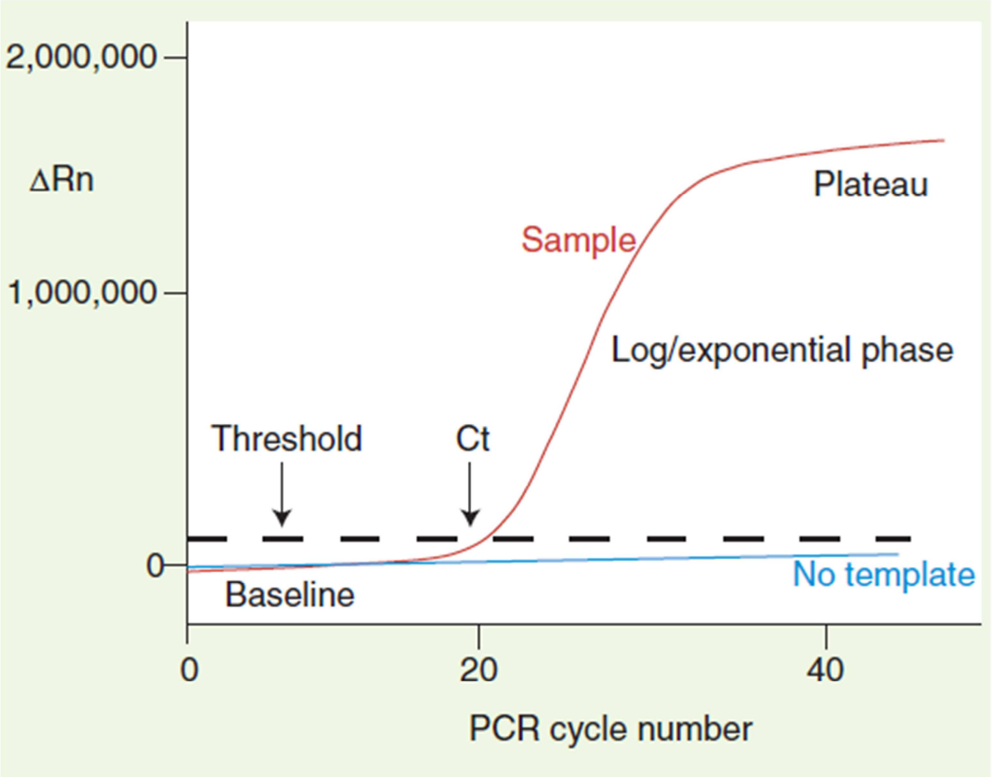


Figure 2.11 An example of an amplification plot reflecting the baseline, Ct (Cq) value, threshold and ΔRn (Arya *et al.*, 2005).

Chapter 3

Analysis of different allelic discrimination approaches for the PCR TaqMan® assay

3.1 Introduction

Breast cancer is the most common malignancy currently diagnosed in SA (Loubser *et al*., 2008). The portion of BC cases that can be directly attributed to an inherited predisposition range between 5 and 10%, of which 15 to 20% is due to germline mutations in the highly penetrant *BRCA1* and *BRCA2* genes (Silla and King, 1995; Claus *et al.,* 1996). Variability in the onset of the disease has been observed amongst mutation carriers worldwide. This can be due to the influence of various low penetrance genes which led to the search for modifying genes. Recent GWAS have searched extensively for low penetrance genes that may play a role in BC development. This search resulted in the identification of polymorphisms in the *TNRC9, LSP1, MAP3K1* and *FGFR2* genes that prove to play a significant role in BC onset (Easton *et al*., 2007).

The demand for faster cost effective mutation detection techniques within the laboratory resulted in the development of new mutation detection methods and SNP genotyping assays. At the forefront of these developments are qPCR applications that seem to have become the gold standard for genotyping. Of these applications, TaqMan® genotyping assays, which utilize FRET probes, are considered the most specific. This technology exploits the 5'-3' exonuclease activity of *Taq* DNA polymerase to degrade allele specific DNA probes (Holland *et al*., 1991). Two probes are utilized within each SNP assay. The first is identical to the ancestral allele, whereas the second corresponds to the variant allele. As the TaqMan® probes are identical except for the base of interest, non-specific binding is reduced. The probes are dual-labeled with a fluorescent reporter at the 5' end and a non-fluorescent quencher at the 3' end. Once bound to the specific target sequence, *Taq* DNA polymerase will cleave the 5' reporter from the probe during the primer extension step. This allows the reporter to emit fluorescence when excited by FRET, which is detected in real time by the thermal cycler (Arya *et al*., 2005).

Genotype calling is based on the intensity of one or both of the released fluorophores at the end of amplification and is performed by the genotyping software of the PCR machine. An amplification plot is drawn using the fluorescence emitted by the reporter fluorophore. Allelic discrimination is automatically performed by the software based on a clustering algorithm. However, manual intervention is required to adjust the fluorescent signal thresholds and select genotypes where the variant allele is rare. This could pose a problem as experience is needed to exclude biased genotyping.

The aim of this chapter is to evaluate three different methods for allelic discrimination and genotype calling to compare the accuracy of the manual and automatic allelic discrimination methods.

3.2 Patients

3.2.1 Patient index and grouping

Sixty females carrying the *BRCA2* 8162delG (c.7934del, p.Arg2645AsnfsX3) founder mutation present in the Caucasian Afrikaner population, were selected for the study. Mutation carriers represented families who had a positive family history of BC, witha minimum of three affected first or second degree relatives. A family was defined as a pedigree consisting of three generations with the *BRCA2* mutation positive patient representing the second generation. Only a single mutation positive BC patient and one unaffected carrier (termed Case) were selected per family to exclude bias. No genetic counseling was required as all the carriers received counseling prior to diagnostic testing.

Group 1 contained mutation positive patients affected with BC, whereas Group 3 included all the unaffected mutation carriers or Cases. Each BC patient carrying the mutation was case- and age-matched with an unaffected mutation carrier and two normal controls representing the average population. The controls were case matched within a five year interval. The age at diagnosis of the BC patient was used as the reference age for the group.

The control participants were of Caucasian Afrikaner decent. They were divided into two groups of 30 individuals each (Groups 2 and 4, Table 3.1). These individuals were recruited with the aid of the South African National Blood Services in Bloemfontein. All control participants were subjected to a baseline screen for the *BRCA2* 8162delG (c.7934del, p.Arg2645AsnfsX3) mutation as previously described (Dajee, 2007; Delport, 2009) to confirm the absence of the Afrikaner founder mutation within each control individual. As the majority of selected SNPs included in this study influence BC risk in *BRCA2* mutation carriers specifically (Easton *et al*., 2007), it was decided not to include *BRCA1* mutation carriers. Each case-matched group (30 in total) thus consisted of four individuals, namely a *BRCA2* mutation positive affected patient (BC), an age specific control (C1), a *BRCA2* mutation positive unaffected individual (Case) and its control (C2) (Table 3.1)

3.2.2 Ethical considerations

The project was approved by the Ethics Committee of the Health Sciences Faculty of the University of the Free State in Bloemfontein (ETOVS 49/06).

Table 3.1 Compilation of groups used in the study.

	Group 1		**Group 2**		**Group 3**		**Group 4**	
	BRCA2 $^{+}$ **Affected (BC)**	**Age at diagnosis**	**Control 1 (C1)**	**Current Age**	***BRCA2*** $^{+}$ **Unaffected (Case)**	**Current Age**	**Control 2 (C2)**	**Current Age**
1	BC18	24	C54	22	Case30	24	C59	28
2	BC14	29	C21	30	Case27	26	C7	31
3	BC23	30	C30	31	Case4	33	C33	32
4	BC11	35	C13	33	Case28	33	C34	33
5	BC17	35	C19	34	Case23	35	C11	35
6	BC5	36	C36	35	Case24	35	C16	35
7	BC29	37	C14	37	Case1	36	C35	37
8	BC12	38	C29	37	Case16	39	C4	38
9	BC27	40	C5	38	Case7	39	C23	38
10	BC3	40	C53	39	Case13	39	C25	39
11	BC22	41	C6	40	Case26	41	C8	41
12	BC19	41	C32	41	Case5	42	C37	41
13	BC15	42	C31	42	Case15	42	C28	43
14	BC16	42	C49	43	Case14	43	C52	44
15	BC25	42	C38	45	Case10	44	C62	44
16	BC6	43	C64	43	Case25	46	C55	46
17	BC30	44	C39	41	Case18	47	C65	51
18	BC10	44	C57	41	Case19	48	C61	51
19	BC13	45	C58	42	Case12	50	C63	54
20	BC7	47	C56	46	Case6	51	C20	48
21	BC8	50	C44	49	Case2	51	C40	51
22	BC20	50	C41	50	Case9	51	C18	51
23	BC28	51	C45	51	Case29	53	C2	52
24	BC4	52	C10	52	Case21	53	C15	53
25	BC2	54	C27	54	Case17	55	C17	55
26	BC24	54	C51	55	Case11	56	C46	55
27	BC9	55	C26	56	Case3	58	C47	57
28	BC26	55	C1	58	Case22	59	C12	58
29	BC21	57	C9	61	Case20	62	C22	61
30	BC1	60	C24	61	Case8	65	C42	62

Permission was obtained from the Head of Clinical Services of Universitas Hospital and the Business Manager of the National Health Laboratory Services (NHLS) to approach and involve possible patients attending clinics at the respective institutions (Appendix A and B). Permission was also obtained from the Head of the Division of Human Genetics for use of the laboratory space and equipment (Appendix C).

Each individual was given unique sample numbers to ensure patient confidentiality. An introductory letter (Appendix D) was given to the participants explaining the aim and protocol of the study. All participants gave written informed consent (Appendix E).

3.3 Methods

3.3.1 DNA extraction

Peripheral blood (10 – 20 ml) was collected in ethylenediaminetetraacetic acid (EDTA) Vacutainer tubes. Genomic DNA was extracted from lymphocytes according to a phenol:chloroform procedure adapted from Sambrook *et al*. (1989). Thawed cells were ruptured in a cold lysis buffer containing 0.3 M sucrose, 10 mM 2-amino-2-(hydroxymethyl)-1,3-propanediol (Tris) pH 7.8, 5 mM $MgCl_2$ and 1% (v/v) t-octylphenoxypolyethoxyethanol (Triton X-100). The solution was centrifuged for 20 min at 4000 *g* at 4°C where after the pellet was resuspended in 4.5 ml SET buffer (10 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM EDTA) containing 100 µl of 10 µg/µl proteinase K and 1% (w/v) sodium dodecyl sulphate (SDS). The solution was incubated overnight at 37°C.

Equal volumes phenol pH 8 (USB Corporation) and chloroform:isoamyl alcohol (24:1) were added and the solution gently mixed on an orbital shaker for 60 min at room temperature. After centrifugation for 10 min at 4000 *g* at 15°C, the supernatant was transferred to a new tube. An equal volume of chloroform:isoamyl alcohol (24:1) was again added, mixed for 60 min at room temperature where after it was centrifuged as described above. The DNA was precipitated from the supernatant by adding two volumes ice-cold 100% (v/v) ethanol and sodium acetate pH 5.4 to a final concentration of 0.3 M. After incubation 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 for a minimum of 60 min. After centrifugation for 5 min at 4000 *g* at 15°C to pellet the DNA, the DNA was air-dried and dissolved in $T._1E$ (10 mM Tris-HCl pH 7.6, 0.1 mM EDTA).

If no DNA was visible after precipitation, the tubes were incubated at -20°C for 24 h. The tubes were then centrifuged for 20 min (4000 *g* at 4°C) to pellet the DNA, where after the pellet was washed with 70% (v/v) ethanol and finally dissolved in $T._1E$. DNA concentration and purity ($A_{260/280}$) were determined using spectrophotometry (NanoDrop® ND-100 Spectrophotometer v3.01, NanoDrop Technologies Inc.) according to the manufacturer's specifications. The DNA concentration was expressed as ng/µl. DNA aliquots of all samples were diluted to 50 ng/µl.

3.3.4 qPCR amplification

3.3.4.1 Molecular analysis of SNPs

Polymorphisms occurring within *TNRC9* (rs3803662), *LSP1* (rs3817198), *MAP3K1* (rs889312) and *FGFR2* (rs2981582) were screened using a modified TaqMan® SNP genotyping assay described by Easton and co-workers (2007). TaqMan® fluorescent probes labeled with FAM and HEX respectively, as well as the specific primer sets, were synthesized for each of the four polymorphisms by Integrated DNA Technologies (IDT; Table 3.2) (Easton *et al.*, 2007).

Table 3.2 Primer (A) and probe (B) sequences used for the molecular analysis of SNPs in *TNRC9*, *LSP1*, *MAP3K1* and *FGFR2*. T_a represents the optimal annealing temperature for each primer set.

A

Polymorphisms	Forward Primer	Reverse Primer	T_a (℃)	Fragment length (bp)
rs3803662 (*TNRC* 9)	5'-GCATTAAGGAGAGAAAATCATTAGGCAGA-3'	5'-CCCAGTACTTTCTTCGCAAATGG-3'	60	101
rs3817198 (*LSP* 1)	5'-CCTCTCTCACCTGATACCAGATTCA-3'	5'-CTGAGCCGGGCTGACT-3'	60	63
rs889312 (*MAP* 3*K* 1)	5'-AGGCCCCATTACTTGAGATGATCT-3'	5'-GGGAAGGAGTCGTTGAGTTTTCA-3'	60	103
rs2981582 (*FGFR* 2)	5'-CAGCACTCATCGCCACTTAATG-3'	5'-GCTGCGGGTTCCTAAAGC-3'	60	76

B

Polymorphisms	Reporter1 Sequence (HEX)	Genotype	Reporter2 Sequence (FAM)	Genotype	Ta (℃)
rs3803662 (*TNRC* 9)	5'-/5HEX/CTTCGCTAAGGGACAGC/3IABkFQ/-3'	Ancestral (C/C)	5'-/56-FAM/TTCGCTAAGAGACAGC/3IABkFQ/-3'	Variant (T/T)	60
rs3817198 (*LSP* 1)	5'-/5HEX/CTAGTGAAATGAGCGGAGAG/3IABkFQ/-3'	Variant (C/C)	5'-/56-FAM/CTAGTGAAATGAGCAGAGAG/3IABkFQ/-3'	Ancestral (T/T)	60
rs889312 (*MAP* 3*K* 1)	5'-/5HEX/CTTAATTTGCACATTCCTTT/3IABkFQ/-3'	Ancestral (A/A)	5'-/56-FAM/ATTTGCACATGCCTTT/3IABkFQ/-3'	Variant (C/C)	60
rs2981582 (*FGFR* 2)	5'-/5HEX/TCTCCGCAAACAGG/3IABkFQ/-3'	Ancestral (C/C)	5'-/56-FAM/CTCTCCACAAACAGG/3IABkFQ/-3'	Variant (T/T)	60

3.3.4.2 TaqMan® assay and amplification

Each 10 µl qPCR reaction contained 100 ng template DNA, 100 nM of each primer and probe and 5 µl IQ Supermix [50 mM KCl, 20 mM Tris-HCl pH 8.4, 0.2 mM deoxyribonucleic acid triphosphates (dNTPs), 25 U/ml i*Taq* DNA polymerase, 3 mM $MgCl_2$]. Thermal cycling conditions were as follows: one cycle at 95℃ for 3 min, followed by 40 cycles at 95℃ for 30 sec, 60℃ for 30 sec and 72℃ for 30 sec. Each plate included 60 participants with four no-template controls (NTC). All amplification reactions were performed in duplicate to confirm each genotype assignment, where after inconclusive genotypes and inconclusive results were repeated separately.

3.3.5 DNA sequencing of heterozygotes

3.3.5.1 DNA cloning

The amplicons of putative heterozygotes for each SNP were cloned in order to be sequenced and confirmed as a true heterozygote. The amplicons were cloned into the pGEM®-T Easy cloning vector (Promega, USA) according to the manufacturer's instructions. The 10 µl ligation reaction mixture contained 30 mM Tris-HCl pH 7.8, 10 mM $MgCl_2$, 10 mM 1,4-Dithiothreitol (DTT), 1 mM Adenosine-5'-triphosphate (ATP), 5% (v/v) polyethylene glycol, 50 ng pGEM®-T Easy vector, 5 U T4 DNA ligase and 3 µl PCR product. The reactions were incubated overnight at 4℃.

The ligated PCR products were transferred to chemical-competent *E. coli* JM109 cells according to Inoue *et al*. (1990) as follows: 50 µl competent cells were mixed with each ligation reaction, incubated on ice for 20 min, heat-shocked for 45-50 sec at 42℃, where after it was placed on ice for 2 min. The reaction was then diluted in 950 µl SOC medium [0.5% (w/v) yeast extract, 2% (w/v) tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose] and incubated on a shaker for 90 min at 37℃. Afterwards 100 µl of each transformed culture was plated on Luria-Bertani (LB) plates [1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl, 1.5% (w/v) agar pH 7.0] supplemented with 50 µg/ml ampicillin, 15 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and 5 mg/ml 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-Gal) (Sambrook *et al*., 1989). The plates were finally incubated overnight at 37℃.

White bacterial colonies containing the cloned qPCR fragments were separately inoculated into 5 ml LB medium containing 50 µg/ml ampicillin and incubated overnight on a shaker at 37℃. Plasmid DNA was isolated from the culture medium using the GeneJET™ Plasmid Miniprep Kit (Fermentas Life Sciences, UK) according to the manufacturer’s instructions. The plasmid DNA was finally eluted with 50 µl elution buffer. DNA quantification was performed as previously described (3.2.2.1).

3.3.5.2 Direct plasmid DNA sequencing

Thirty cloned inserts were sequenced to confirm the presence of the ancestral and variant alleles within each of the selected heterozygotes. These individuals were to serve as positive controls during data analysis. Plasmid DNA (200 ng) was used as template and the insert sequenced with the standard M13 forward sequencing primer (5’-GTAAAACGACGGCCAGT-3’) using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, US). The cycle sequencing reaction consisted of 1 µl of PCR template, 4 µl terminator ready reaction mix, 3.2 pmol primer for each of the polymorphisms and 2 µl BigDye® sequencing buffer to a final volume of 20 µl. The sequencing regime was as follows: one cycle at 96℃ for 1 min, followed by 25 cycles at 96℃ for 10 sec, 53℃ for 5 sec and 60℃ for 4 min, with a final holding temperature of 15℃. The sequenced products were precipitated with 5 µl 125 mM EDTA and 60 µl 100% (v/v) ethanol, followed by a 15 min incubation at room temperature in the dark. After centrifugation at 12 000 *g* for 30 min, the pellet was washed once with 60 µl 70% (v/v) ethanol and finally air-dried in the dark.

Hi-Di™ formamide (Applied Biosystems, US) buffer (30 µl) was added to the pellets. Denaturation followed for 5 min at 96℃, where after the tubes were snap cooled on ice. The DNA sequences of these fragments were resolved using the ABI Prism 310 Genetic Analyzer (Applied Biosystems, US) and the electropherograms analyzed using proprietary sequence analysis software (Chromas version 2.31, www.technelysium.com.au).

3.3.6 Data analysis utilizing different allelic discrimination approaches

Three different approaches were used for data analysis. The first method (Method 1) made use of automated allelic discrimination as detected by the BioRad CFX Manager v1.1.308.1111 software (Bio-Rad, US), which made use of a scatter plot for discrimination (Fig. 3.1). Samples with the same genotype were clustered into groups. The FAM and HEX baselines introduced by the software on the scatter plot were manually adjusted to assign samples located on the baselines groups. Once the data was normalized, it was exported to Excel for further analysis.

Method 2 was based on the theory that any increase of fluorescence above the baseline indicated true amplification regardless of the relative fluorescence units (RFU). Heterozygosity was thus identified based on the detection of both FAM and HEX amplification above the threshold. This method was supported by the fact that TaqMan® analyses are proclaimed to be very specific and that both probes have an equal opportunity to anneal. Therefore, if amplification occurs, the specific allele must be present. For this method, the amplification plot for each sample was used for allelic discrimination. Two potential heterozygotes are illustrated in Fig. 3.2 A and B. A homozygote was identified when only one allele labeled with a specific fluorophore amplified above the baseline (Fig. 3.2 C and D).

Method 3 utilized both the Cq and RFU values for allelic discrimination. Heterozygotes were identified in Cases where the Cq values of the amplification products differed by not more than one (Fig. 3.3 A). A Cq value of more than one represented homozygosity for either the ancestral or variant allele (Fig. 3.3 B). The homozygous genotypes were determined by also taking the respective RFU values into account.

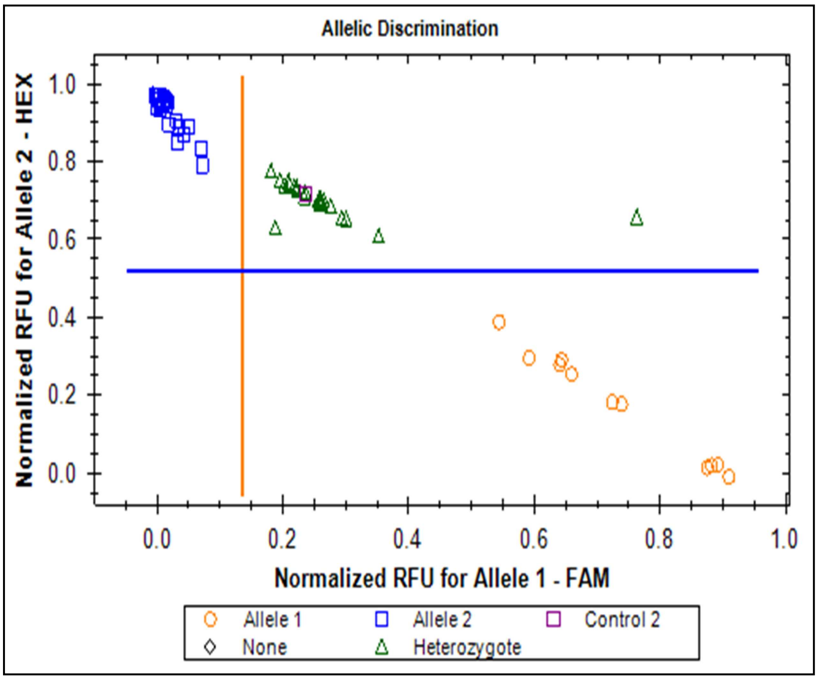


Figure 3.1 Genotype calling based on scatter plot analysis according to Method 1. A blue square represents homozygosity for allele 2 labeled with HEX, whereas an orange circle represents homozygosity for allele 1, labeled with FAM. Heterozygotes are represented by a green triangle, with the positive heterozygote control indicated as a purple square.

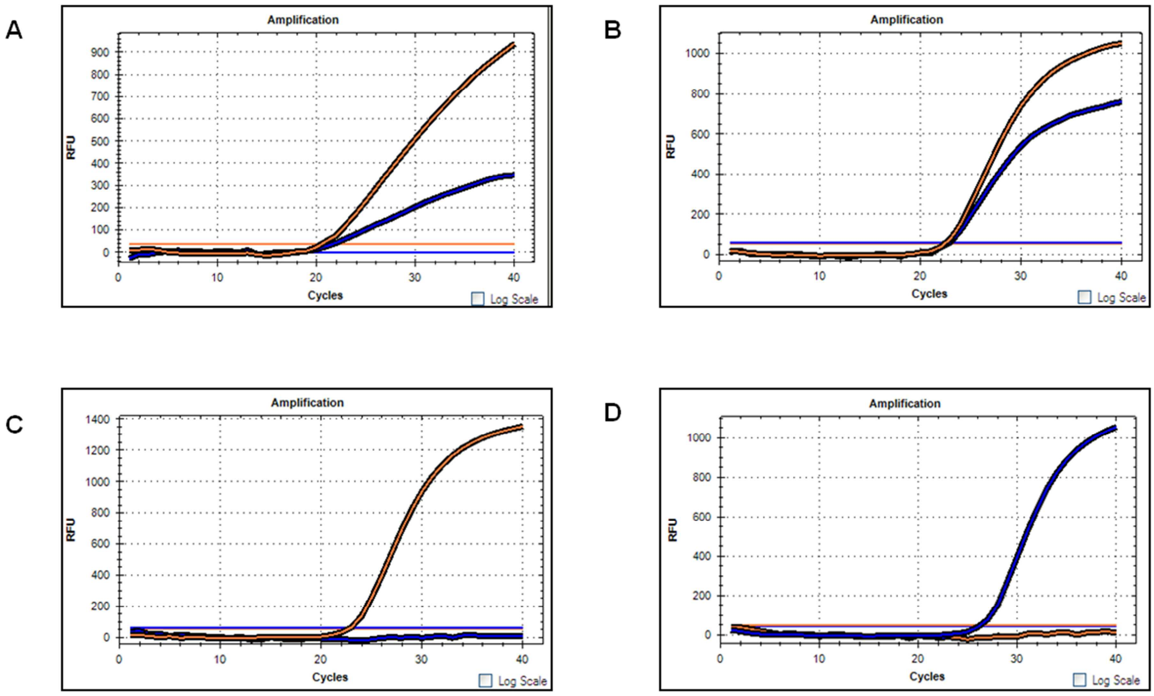
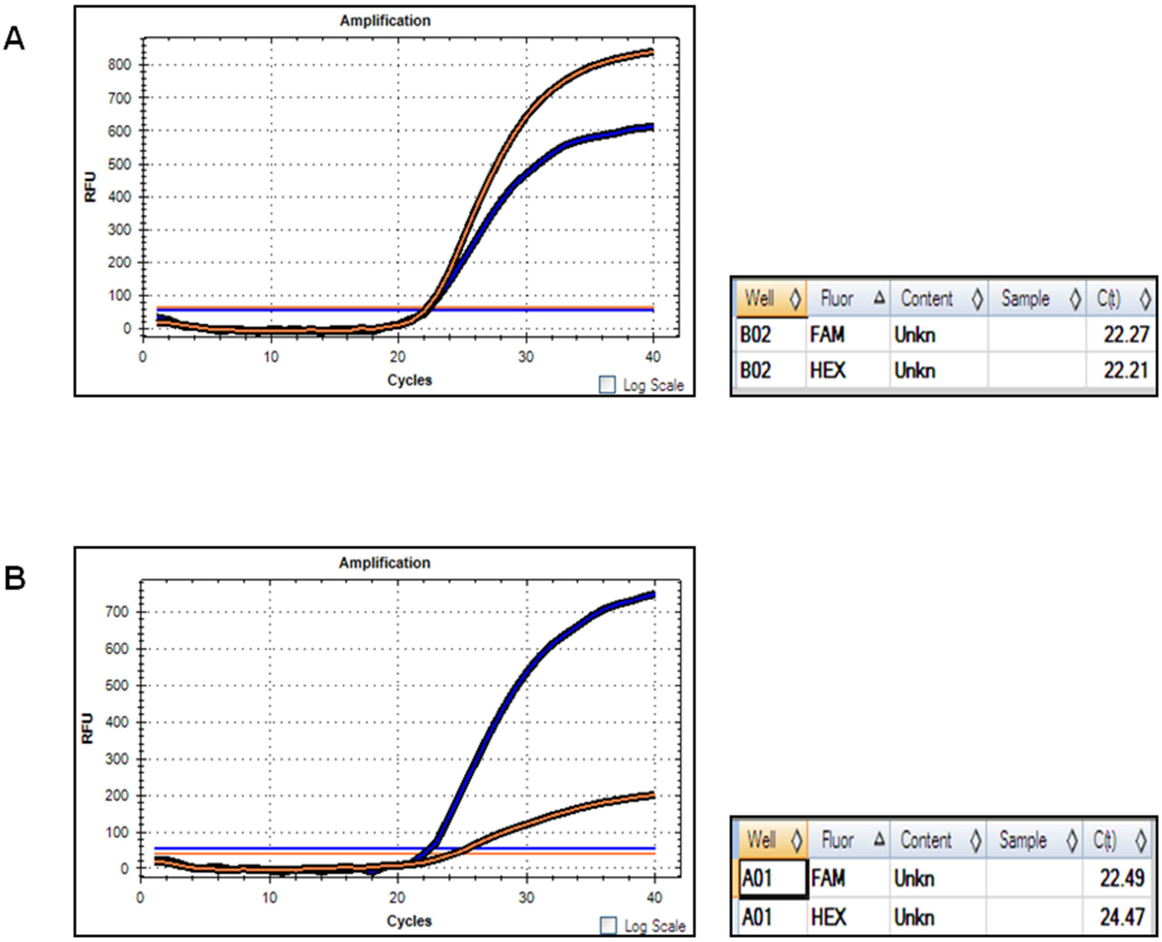


Figure 3.2 Allelic discrimination according to Method 2. While **A** and **B** represent heterozygotes, **C** and **D** represent homozygosity for the respective alleles.



Well	Fluor	Content	Sample	C(t)
B02	FAM	Unkn		22.27
B02	HEX	Unkn		22.21

Well	Fluor	Content	Sample	C(t)
A01	FAM	Unkn		22.49
A01	HEX	Unkn		24.47

Figure 3.3 Examples of allelic discrimination according to Method 3. Based on the Cq values, figure **A** represents a heterozygote (Cq values deviate with no more than 1) whereas **B** is homozygous for the FAM (blue) labeled allele (Cq value differs with more than one).

3.3.7 Statistical analysis

The statistical analyses were performed by the Department of Biostatistics at the University of the Free State. Results were summarized by frequencies and percentages. The analysis was performed between the controls and the *BRCA2* mutation carriers (Comparison 1) only. Odds ratios (OR) and confidence intervals (CI) were calculated using McNemar's test (Fleiss *et al*., 2003).

Although McNemar's test bears a resemblance to the chi-square or Fisher exact probability test, it is more appropriate when using a matched case-control study design. McNemar's test assesses the significance of the difference between two correlated dependent proportions, where the two proportions are based on the same sample of subjects (http://faculty.vassar.edu/lowry/propcorr.html). Cohen's kappa measurement of agreement (Cohen, 1960) was used to determine the agreement between the different allelic discrimination approaches.

3.4 Results

3.4.1 *BRCA2* 8162delG baseline screen

All 120 participants were previously screened using single strand conformation polymorphism (SSCP) analysis for the presence of the Afrikaner *BRCA2* 8162delG (c.7934del,p.Arg2645AsnfsX3) founder mutation in order to be included in the study (Dajee, 2007; Delport, 2009). Six participants included in the initial group had to be replaced due to depleted DNA stocks. These included three BC patients, two Cases and a single control. The mutation status of the six new participants was confirmed prior to enrollment into the study.

3.4.2 Evaluation of qPCR conditions

Before commencing with qPCR, all four primer sets were tested by conventional PCR analysis using six randomly selected participants according to the conditions described by Easton *et al*. (2007). Single distinct amplicons of 103, 101, 76 and 63 bps respectively, were obtained for all six participants for all four primer sets indicating optimal amplification conditions (Fig. 3.4).

Similarly, the qPCR conditions were evaluated using the Bio-Rad IQ Supermix according to Easton *et al*. (2007). qPCR analysis of the rs3803662 SNP located in *TNRC9* was tested for five individuals, together with a no template control (NTC) (Fig. 3.5). Similar amplification efficiencies were obtained, with Cq values ranging between 20 and 30 cycles for all five samples. The NTC did not display any amplification, indicating optimal conditions.

The polymorphism in *TNRC9*, namely rs3803662, entails a single base change of a C to a T. The probe representing the ancestral C allele was fluorescently labeled with HEX (displayed on the amplification graphs as orange) whereas the probe representing the variant T allele was labeled with FAM (displayed on the graphs as blue). When analyzed, all 60 individuals were homozygous for the ancestral C allele (Fig. 3.6 A). In order to confirm the low frequency of the variant T allele and to confirm the optimized qPCR conditions for this Taqman® assay, individuals of African and Mixed ancestry were genotyped for this SNP (Fig. 3.6 B). Several individuals displayed the variant alleles thus confirming the ability of the assay to accurately detect the presence of the variant T allele.

A single putative heterozygote was sequenced to confirm the genotype and to use it as a positive control. The presence of both ancestral and variant sequenced

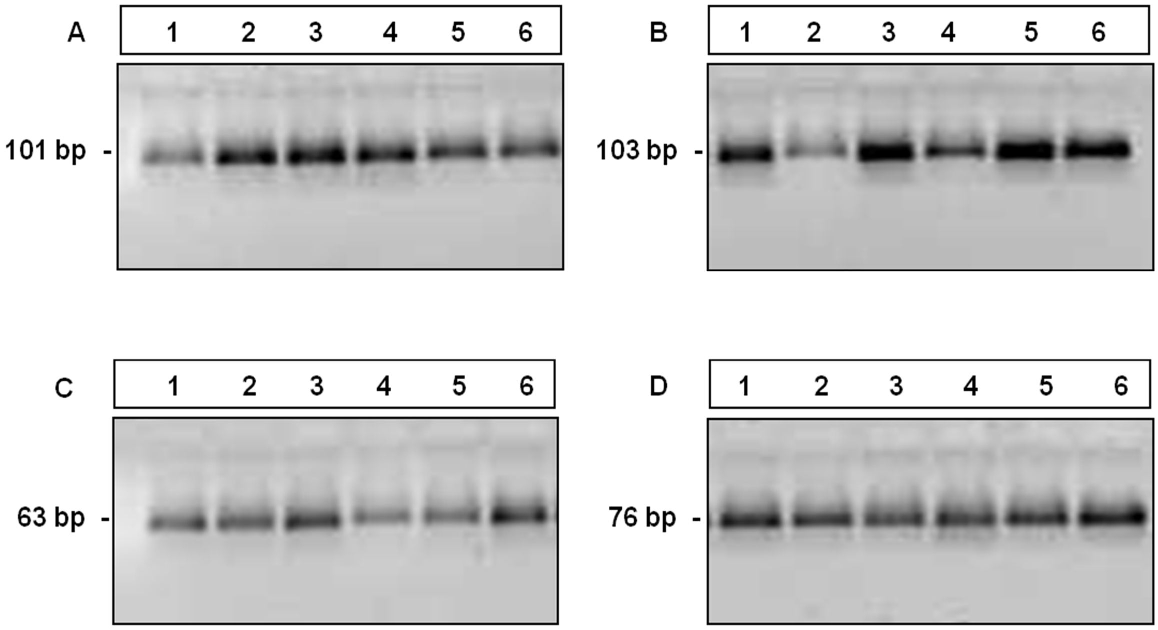


Figure 3.4 Conventional PCR amplification of four selected amplicons according to Easton *et al*. (2007). The indicated amplicons are for **A** rs3803662 in *TNRC9*, **B** rs889312 in *MAP3K1*, **C** rs3817198 in *LSP1* and **D** rs2981582 present in *FGFR*2. The size of each amplicon is as indicated.

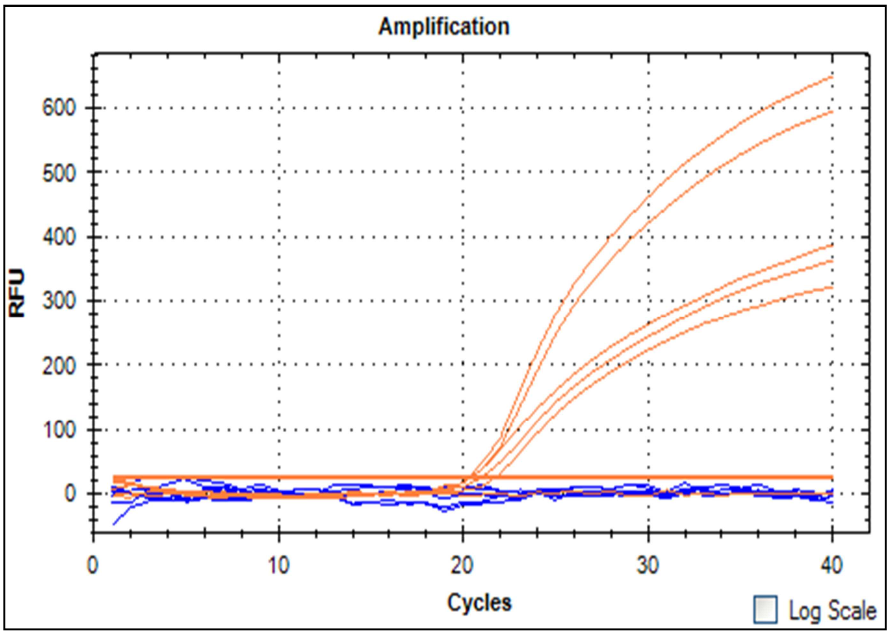


Figure 3.5 Testing of optimal qPCR conditions for the rs3803662 SNP in the *TNRC9* gene. Positive amplification for HEX as indicated in orange.

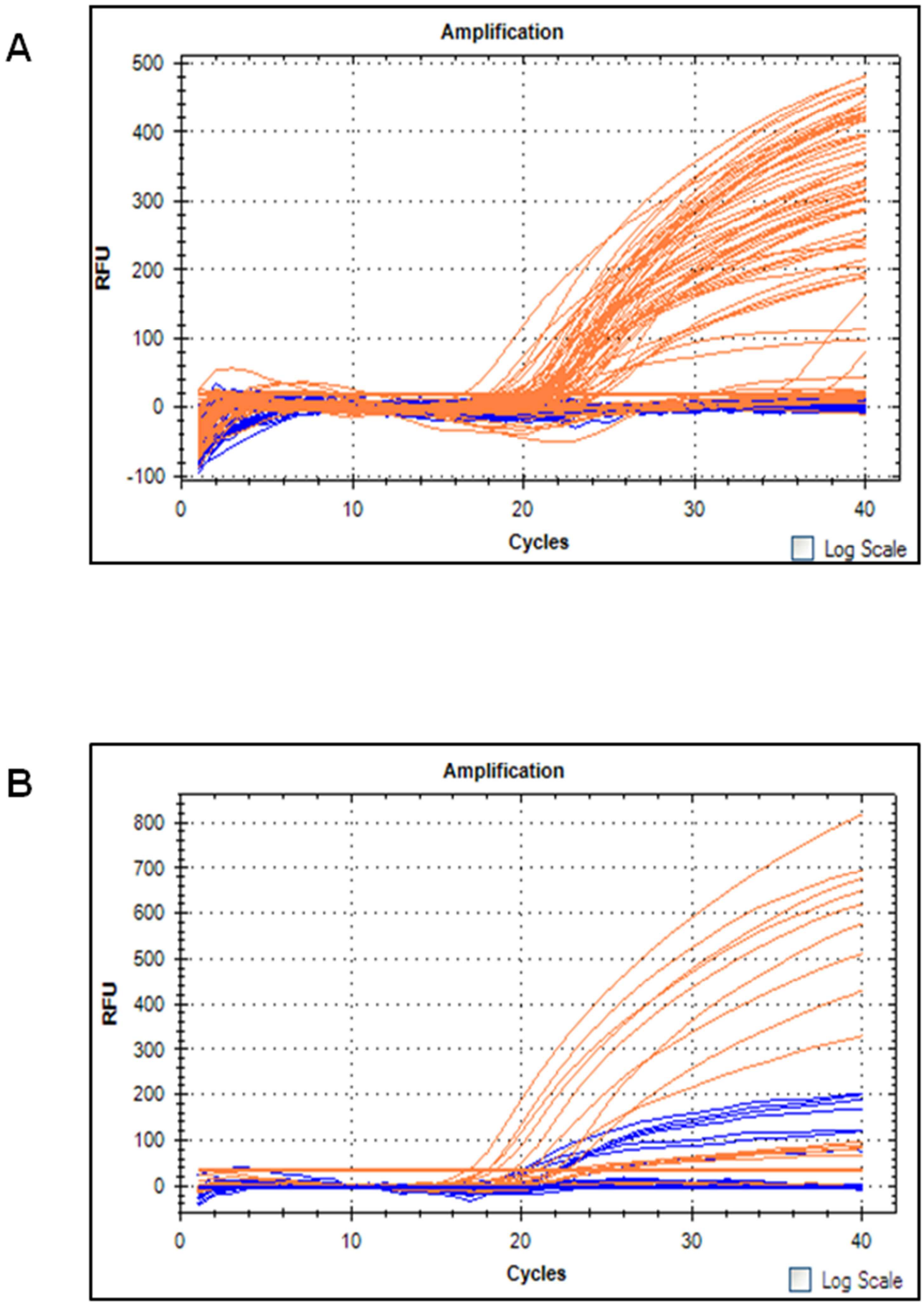


Figure 3.6 Initial qPCR analysis of rs3803662 in *TNRC9*. **A** Amplification plots for 60 Afrikaner participants revealing the absence of the variant T allele. **B** Amplification profiles of several participants of African and Mixed ancestry descent indicating the presence of the variant alleles.

alleles confirmed the selected individual as a true heterozygote (Fig. 3.7). Once the positive and no template controls were assigned to each data set, the results were analyzed. Once the optimization of rs3803662 in *TNRC9* was complete, the analysis of the SNPs in *LSP1* (rs3817198), *MAP3K1* (rs889312) and *FGFR2* (rs2981582) commenced.

3.4.3 Analysis of rs3803662 in *TNRC9*

All 120 study participants were screened for the presence of a single base change (C>T) within *TNRC9*. All three possible genotypes were observed for each method of analysis. For Method 1 (automated allelic discrimination), manual intervention was required to adjust the FAM and HEX baselines using the RFU mode (Fig. 3.8).

For Method 2, the homozygous ancestral genotype (C/C) was displayed as a HEX (orange) signal only with no FAM (blue) signal (Fig. 3.9). The homozygous variant genotype (T/T) was displayed as a FAM signal with no visible HEX signal. A heterozygote (C/T) was recognized when both the FAM and HEX signals were observed regardless of the RFUs of the signal. Method 3 entailed the identification of a homozygous ancestral (C/C) genotype as a HEX signal observed with a low or no RFU signal for FAM (Fig. 3.10). The heterozygote exhibited both FAM and HEX signals with Cq values differing with less than one. Samples, for which inconclusive results were obtained for all three methods, were re-analyzed.

The allele and genotype frequencies for rs3803622 in *TNRC9* according to Methods 1, 2 and 3 are presented in Table 3.3 and 3.4. Method 1 resulted in the identification of 102 participants homozygous for the ancestral allele (52 Controls and 50 *BRCA2* mutation carriers), with only 12 being heterozygous (five Controls and seven *BRCA2* mutation carriers) and six homozygous for the variant T allele

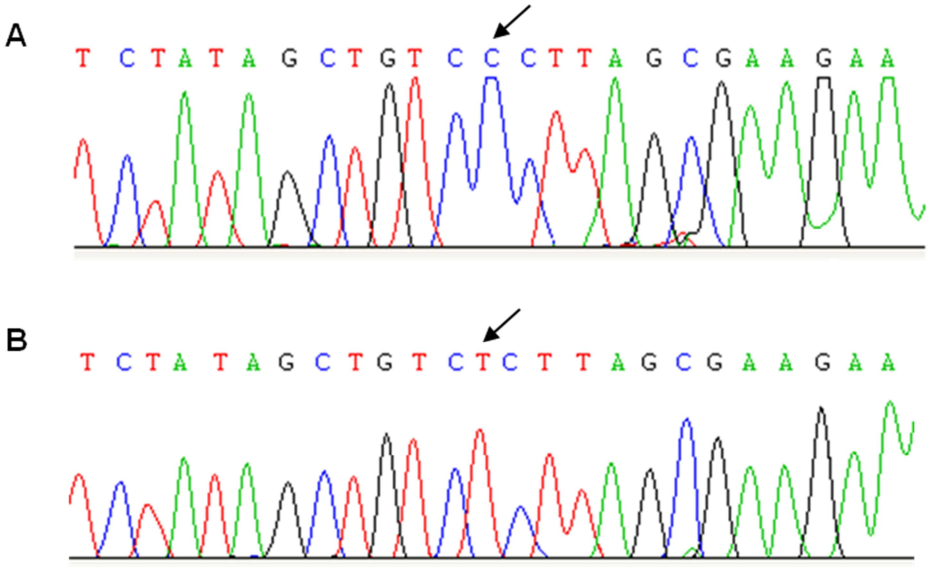


Figure 3.7 Sequence analysis of BC patient 6–1, heterozygous for rs3803662 in *TNRC9*. The position of the SNP is indicated by an arrow. **A** Presence of the ancestral C allele. **B** Presence of the variant T allele.

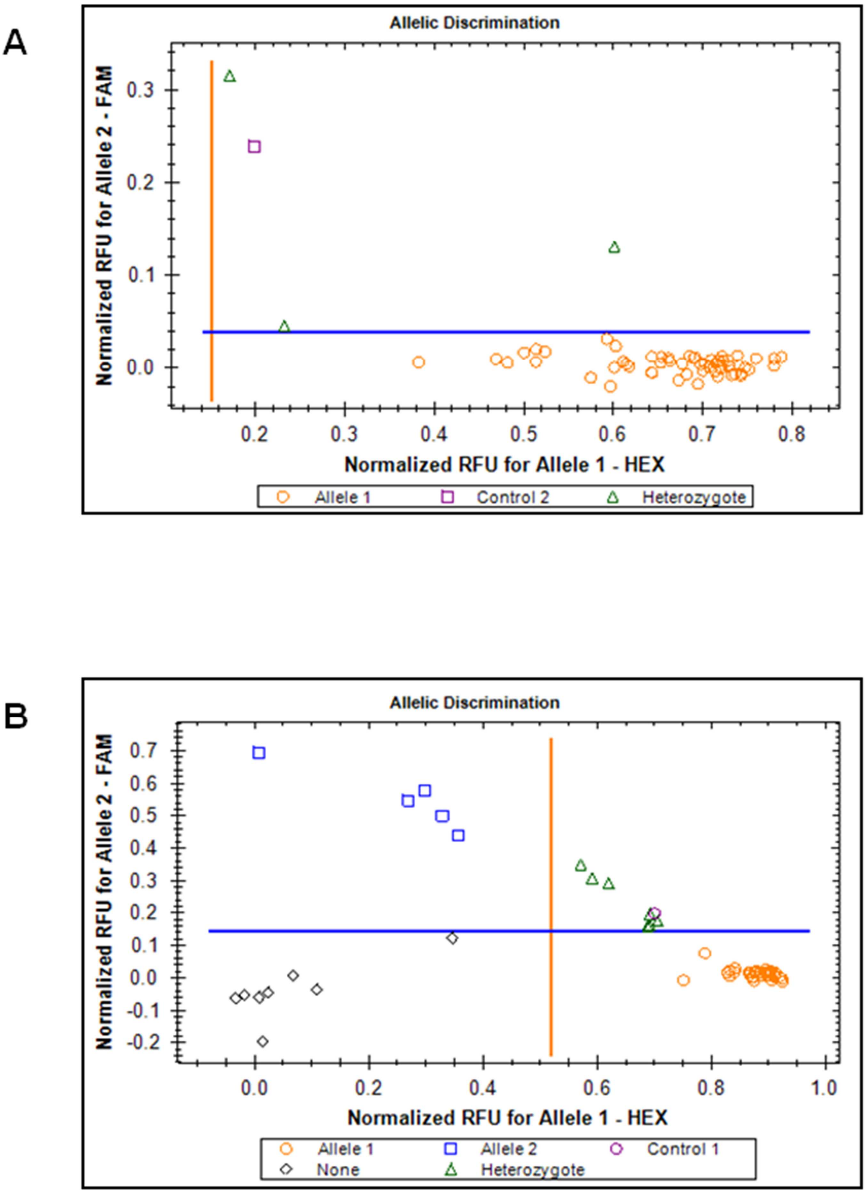


Figure 3.8 Genotyping results of 120 participants for rs3803662 in *TNRC9* according to Method 1, presented in two scatter plots **A** and **B** orange circles represent participants homozygous for the ancestral (C/C) HEX labeled allele 1. The blue squares are individuals homozygous for the FAM labeled variant (T/T) allele 2. Heterozygotes carrying both the ancestral and variant alleles are indicated by a green triangle. The positive control is represented by a purple circle. Samples for which an inconclusive result obtained, are indicated by a black diamond.

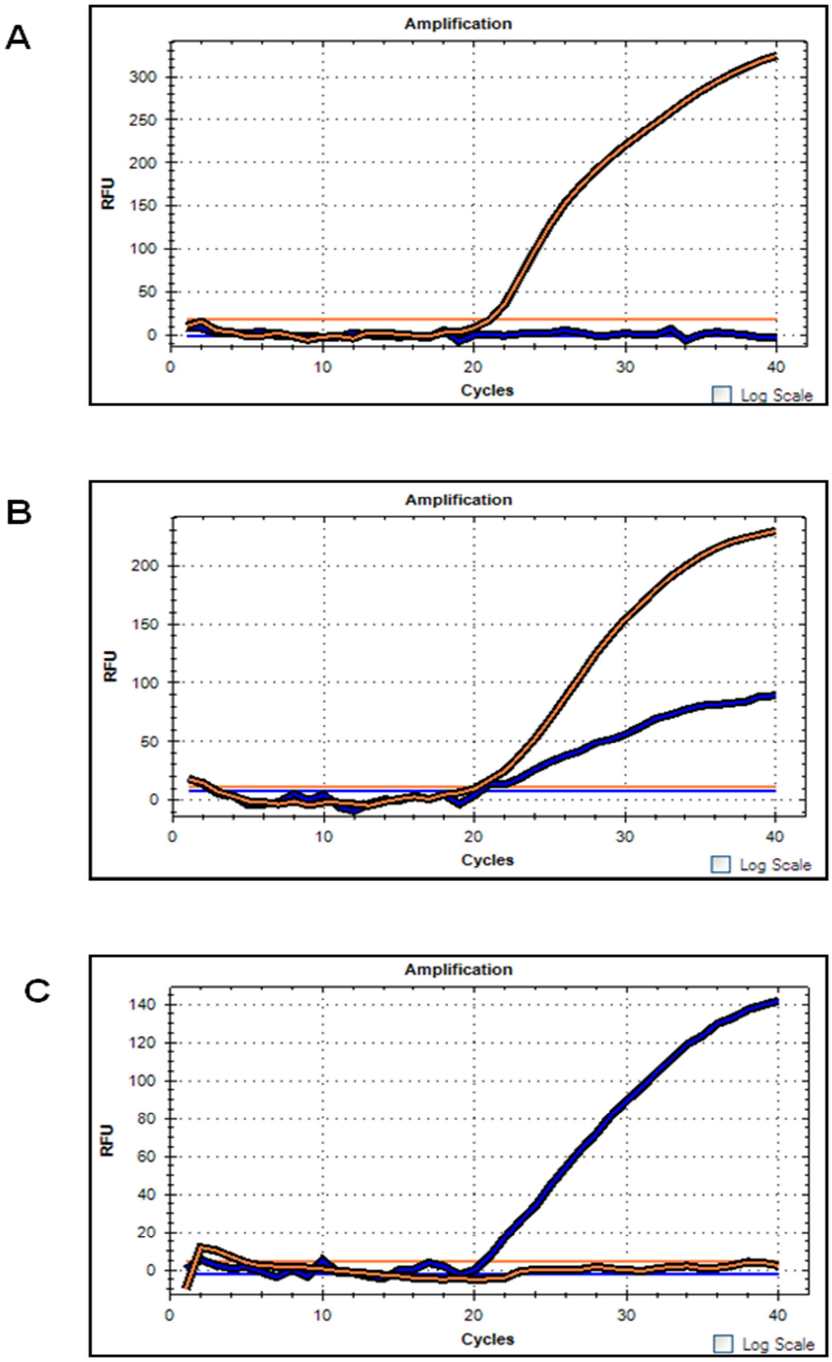
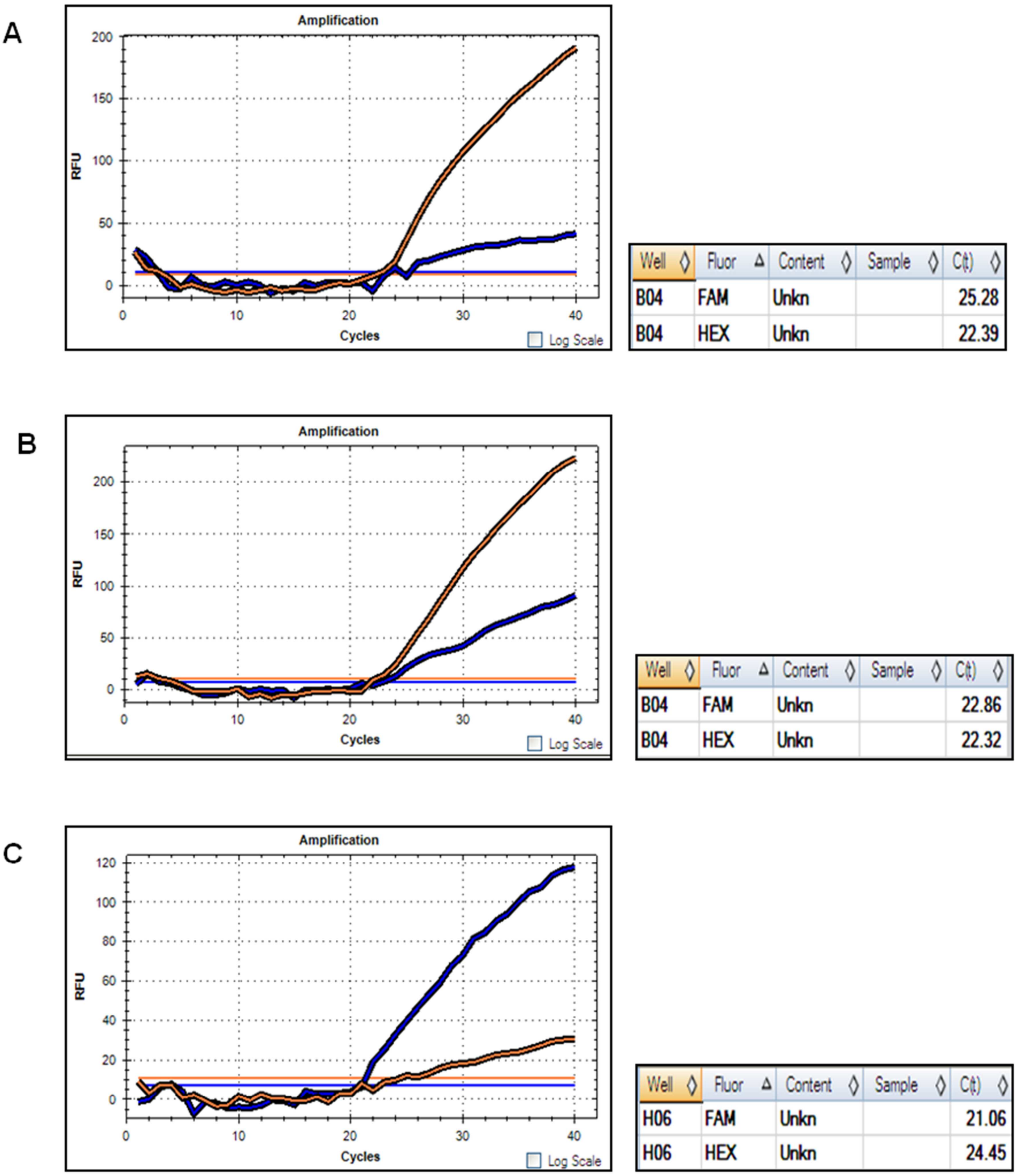


Figure 3.9 Genotype calling of rs3803662 in *TNRC9* according to Method 2. **A** Amplification of the ancestral C allele only, represented by the HEX signal. **B** Heterozygote (C/T) recognized by the amplification of both alleles presented by the FAM and HEX signals. **C** Individual homozygous for the variant (T/T), displayed as a FAM signal only.



Well	Fluor	Content	Sample	C(t)
B04	FAM	Unkn		25.28
B04	HEX	Unkn		22.39

Well	Fluor	Content	Sample	C(t)
B04	FAM	Unkn		22.86
B04	HEX	Unkn		22.32

Well	Fluor	Content	Sample	C(t)
H06	FAM	Unkn		21.06
H06	HEX	Unkn		24.45

Figure 3.10 Genotype calling of rs3803662 in *TNRC9* according to Method 3. **A** Amplification of the homozygous ancestral (C/C) allele represented by a HEX signal with a low or no RFU signal for FAM and a Cq value differing with more than 1. **B** Heterozygote (C/T) recognized by amplification of both alleles represented by the FAM and HEX signals, with Cq values deviating with less than one. **C** Individual homozygous for the variant T allele (T/T), displayed as a FAM signal with a low or no RFU for the HEX signal and a Cq value of >1.

(Table 3.3). In total, the variant allele was only present in 9.17% of the Controls and 10.83% of the *BRCA2* mutation carriers.

The mode of analysis of Method 2 resulted in 107 (54 and 53 respectively) participants being homozygous for the ancestral C allele, with 12 individuals heterozygous (C/T) and a single individual homozygous for the variant T allele (Table 3.3). Method 3 revealed the highest number of homozygous ancestral (C/C) participants namely 111, with only four and five being heterozygous and homozygous for the variant T allele respectively (Table 3.3). The variant T allele was limited to 5.83% of the Controls and *BRCA2* mutation carriers respectively (Table 3.3).

Comparison of the homozygous participants indicated that the results for this genotype of Method 1 and 3 were more similar (individuals 19–2, 27–3, 28–4, 29–2 and 29–4) (Table 3.4). Of the six participants homozygous for the variant T allele in Method 1, five corresponded and were also called homozygous by Method 3 (Table 3.4). The sixth participant (20–3) proved to be heterozygous for the variant T allele when scored using Method 3. Interestingly, only one (0.83%) participant was homozygous for the variant T allele in Method 2 (Table 3.3).

In order to determine which mode of analysis was the most accurate for rs3803662, the genotypes of each hetero- and homozygotic participant exhibiting the variant T allele were compared (Table 3.4). Although Methods 1 and 2 called an equal number of participants heterozygous for the variant T allele, only six corresponded between the two methods (Table 3.4).

Table 3.3 Allele and genotype frequencies of rs3803662 in *TNRC9* according to Methods 1, 2 and 3.

Methods	Frequency distribution	Comparison 1 Control vs *BRCA2* mutation carriers		
		Controls n=60 n (%)	Carriers n=60 n (%)	OR (95% CI)
Method 1	Allele (n/%)			
	C	109 (90.83)	107 (89.17)	0.83 (0.36-1.94)
	T	11 (9.17)	13 (10.83)	1.20 (0.51-2.78)
	Genotype (n/%)			
	C/C	52 (86.67)	50 (83.33)	0.75 (0.21-2.47)
	C/T	5 (8.33)	7 (11.67)	1.67 (0.32-10.73)
	T/T	3 (5.00)	3 (5.00)	1.00 (0.13-7.47)
		P-value = 0.9189		
	C/T + T/T	8 (13.33)	10 (16.67)	1.33 (0.41-4.66)
		P-value = 0.5930		
Method 2	Allele (n/%)			
	C	114 (95.00)	112 (93.33)	0.74 (0.25-2.19)
	T	6 (5.00)	8 (6.67)	1.35 (0.46-4.00)
	Genotype (n/%)			
	C/C	54 (90.00)	53 (88.33)	0.86 (0.24-2.98)
	C/T	6 (10.00)	6 (10.00)	1.00 (0.27-3.74)
	T/T	0 (0.00)	1 (1.67)	
		P-value = 0.8415		
	C/T + T/T	6 (10.00)	7 (11.67)	1.17 (0.34-4.20)
		P-value = 0.7815		
Method 3	Allele (n/%)			
	C	113 (94.17)	113 (94.17)	1.00 (0.34-2.94)
	T	7 (5.83)	7 (5.83)	1.00 (0.34-2.94)
	Genotype (n/%)			
	C/C	56 (93.33)	55 (91.67)	0.80 (0.16-3.72)
	C/T	1 (1.67)	3 (5.00)	3.00 (0.24-157.5)
	T/T	3 (5.00)	2 (3.33)	0.67 (0.06-5.82)
		P-value = 0.7530		
	C/T + T/T	4 (6.67)	5 (8.33)	1.25 (0.27-6.30)
		P-value = 0.7389		

Table 3.4 Discrepancies observed in the genotype analysis of rs3803662 in *TNRC9* between Methods 1, 2 and 3.

Patient No	Method 1	Method 2	Method 3
1–3	CT	CT	CC
15–4	CT	CC	CC
16–3	CT	CC	CC
17–2	CT	CT	CC
17–3	CT	CC	CC
17–4	CT	CT	CC
19–2	TT	CT	TT
20–1	TT	TT	TT
20–3	TT	CT	CT
24–4	CT	CT	CC
25–1	CT	CT	CC
27–3	TT	CC	TT
28–3	CT	CT	CC
28–4	CT	CC	CT
29–2	TT	CT	TT
29–4	TT	CT	TT

3.4.4 Analysis of rs3817198 in *LSP1*

The polymorphism present in *LSP1* (rs3817198) entailed a single nucleotide change from a T to a C (Fig. 3.11). The genotyping analyses of all 120 individuals were performed in duplicate. Samples with inconclusive results for any one of the modes of analyses were repeated. For qPCR analysis, the ancestral T allele was labeled with FAM while the variant C allele was labeled with HEX. All three possible genotypes were observed for all three genotyping methods. Within all the experiments performed, the NTCs displayed no RFU values.

The automated allelic discrimination method (Method 1) clustered the genotypes in groups after manual adjustment of the baselines to assign individuals located on the baselines (Fig. 3.12). For analysis according to Method 2, the homozygous ancestral genotype (T/T) was displayed as a FAM signal with no HEX signal present (Fig. 3.13 A), whereas the homozygous variant genotype (C/C) was displayed only as a HEX signal (Fig. 3.13 C). Heterozygous individuals (T/C) displayed both HEX and FAM signals (Fig. 3.13 B). Similarly, Method 3 also revealed the presence of all three genotypes (Fig. 3.14). Participants heterozygous for this SNP were genotyped when both FAM and HEX signals were detected and the Cq value deviated by no more than one.

The allele and genotype frequencies for rs3817198 in *LSP1* are presented in Table 3.5. The frequencies for Method 1 indicated that 50 participants (24 Controls and 26 *BRCA2* mutation carriers) were homozygous for the ancestral T allele whereas 53 (30 Controls and 23 *BRCA2* mutation carriers) were heterozygous (T/C). Only 17 (six Controls and 11 *BRCA2* mutation carriers) participants were homozygous for the variant C allele (Table 3.5). The frequency distribution of the variant C allele was on average 36.25% of all 240 alleles present (35.00% for the Controls and 37.50% for the *BRCA2* mutation carriers) (Table 3.5).

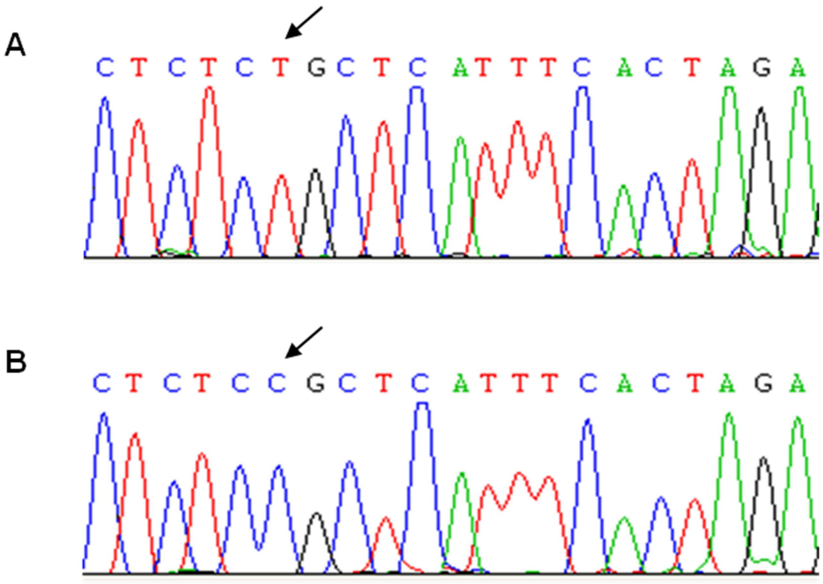


Figure 3.11 Sequencing analysis of Case 28–3, heterozygous for rs3817198 in *LSP1*. The position of the SNP is indicated by an arrow. **A** Presence of the ancestral T allele. **B** Presence of the variant C allele.

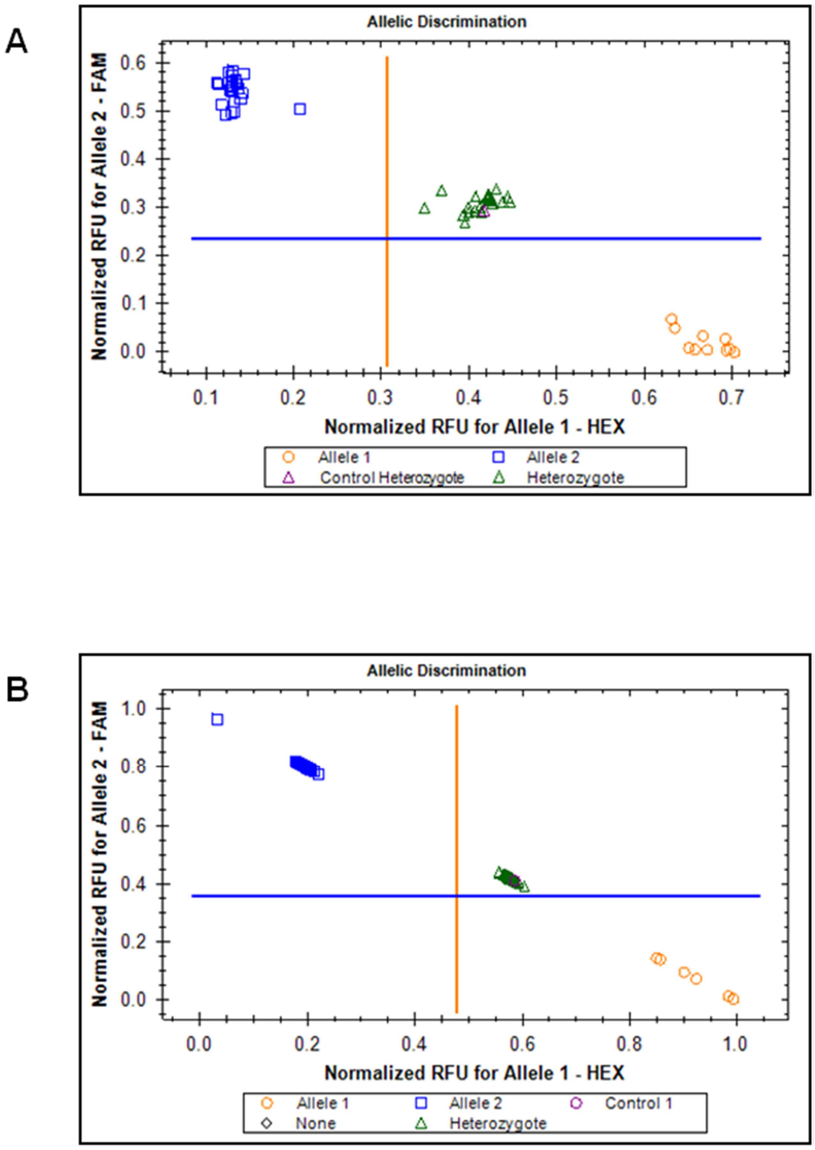


Figure 3.12 Genotyping results for rs3817198 in *LSP1* according to Method 1, presented in scatter plots **A** and **B**. Blue squares represent participants homozygous for the ancestral (T/T) FAM labeled Allele 1. Orange circles represent participants homozygous for the variant (C/C) HEX labeled Allele 2. Heterozygotes carrying both the ancestral and variant alleles are indicated by a green triangle. The positive control is represented by a purple circle, whereas individuals for which inconclusive results were obtained, are indicated by a black diamond.

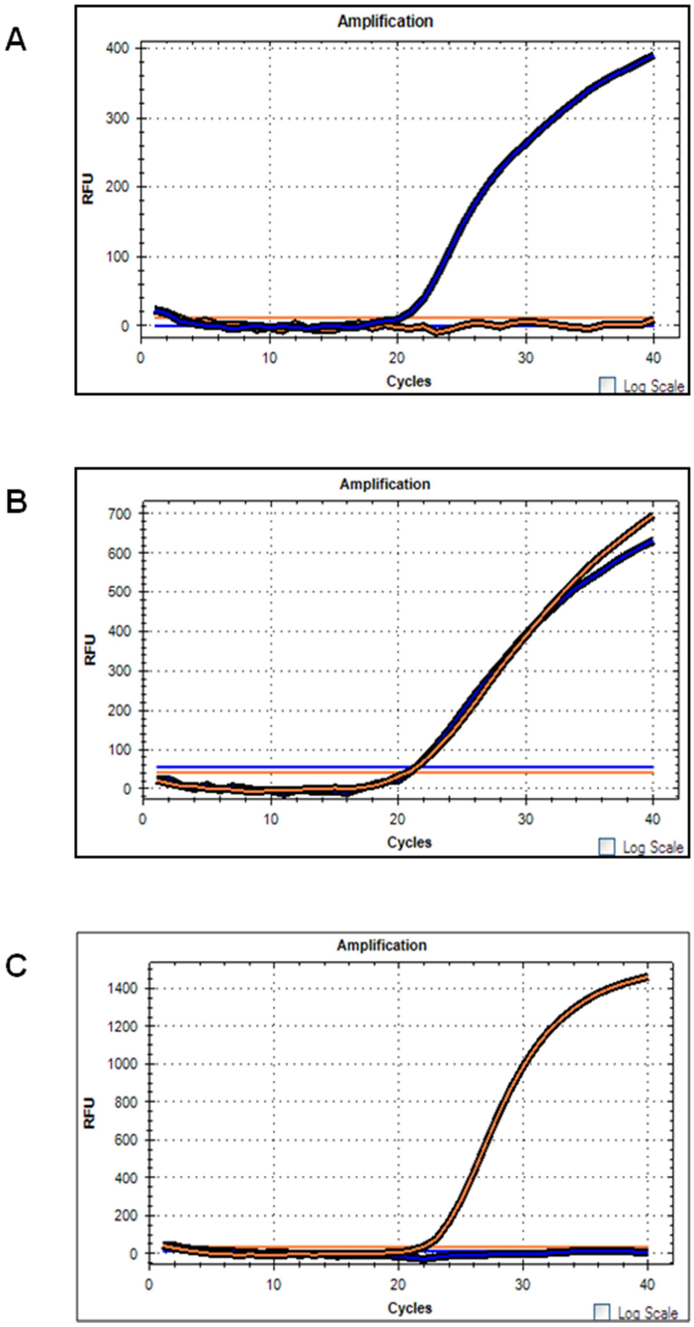
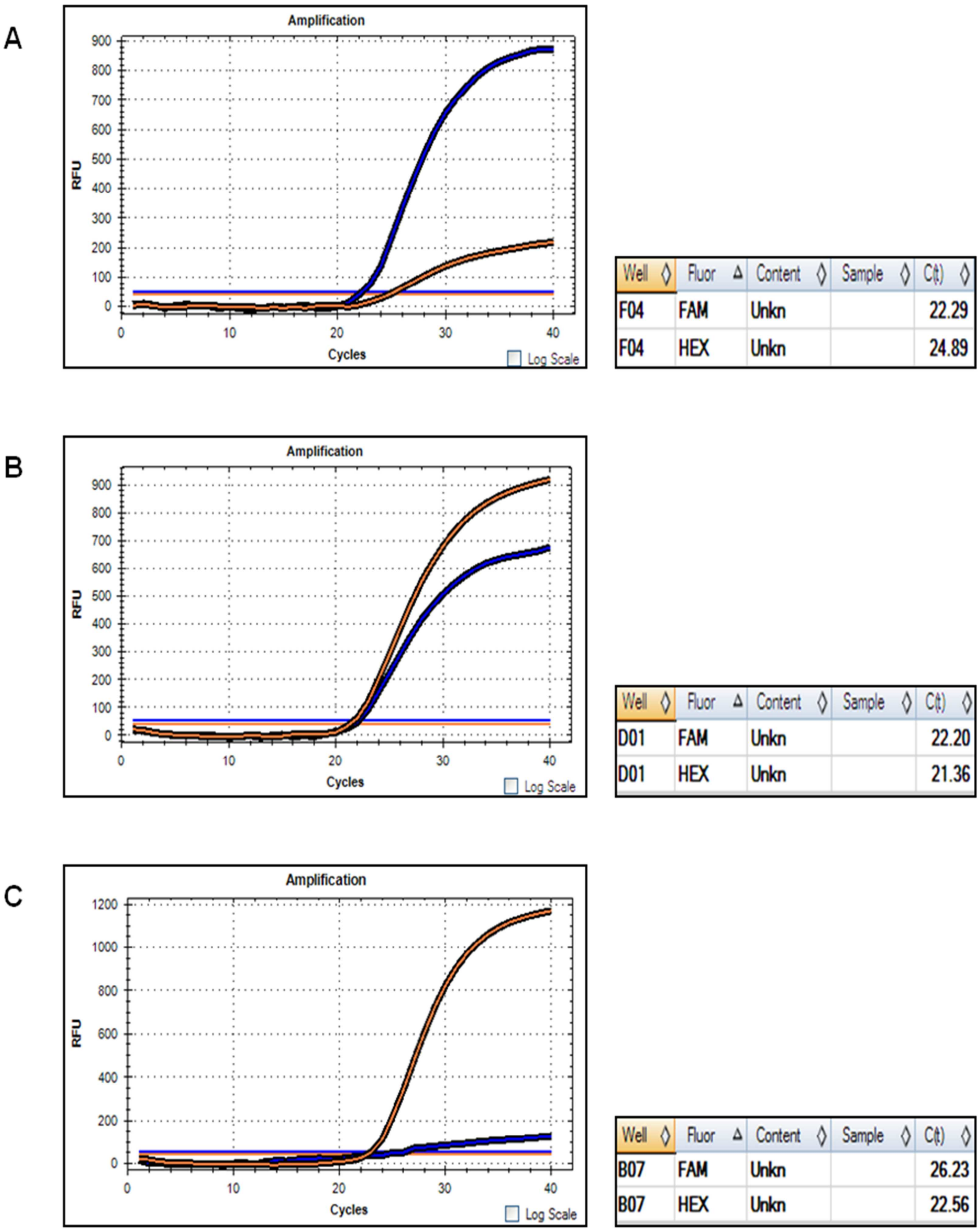


Figure 3.13 Genotyping analysis of rs3817198 in *LSP1* according to Method 2. **A** Amplification of the ancestral T allele only, represented by the FAM signal. **B** Heterozygous individual (T/C) recognition by the amplification of both alleles represented by both the FAM and HEX signals. **C** Participants homozygous for the variant (C/C), displayed as a HEX signal only.



A

Well	Fluor	Content	Sample	C(t)
F04	FAM	Unkn		22.29
F04	HEX	Unkn		24.89

B

Well	Fluor	Content	Sample	C(t)
D01	FAM	Unkn		22.20
D01	HEX	Unkn		21.36

C

Well	Fluor	Content	Sample	C(t)
B07	FAM	Unkn		26.23
B07	HEX	Unkn		22.56

Figure 3.14 Genotyping analysis of rs3817198 in *LSP1* according to Method 3. **A** Amplification of the ancestral T allele represented by a FAM signal with a low or no RFU signal for HEX and a Cq value differing with more than 1. **B** Heterozygous individual recognition by amplification of both alleles represented by the FAM and HEX signals, with a Cq value deviating with less than 1. **C** Individual homozygous for the variant C allele (C/C), displayed as a HEX signal with a low or no RFU for FAM and a Cq value of more than 1.

Table 3.5 Allele and genotype frequencies of rs3817198 in *LSP1* according to Methods 1, 2 and 3.

Methods	Frequency distribution	Comparison 1 Control vs *BRCA2* mutation carriers		
		Controls n=60 n (%)	Carriers n=60 n (%)	OR (95% CI)
Method 1	Allele (n/%)			
	T	78 (65.00)	75 (62.50)	0.90 (0.53-1.52)
	C	42 (35.00)	45 (37.50)	1.11 (0.66-1.89)
	Genotype (n/%)			
	T/T	24 (40.00)	26 (43.33)	1.14 (0.52-2.53)
	T/C	30 (50.00)	23 (38.33)	0.61 (0.26-1.37)
	C/C	6 (10.00)	11 (18.33)	2.00 (0.62-7.46)
		P -value = 0.2863		
	T/C + C/C	36 (60.00)	34 (56.67)	0.88 (0.40-1.91)
		P -value = 0.7150		
Method 2	Allele (n/%)			
	T	55 (45.83)	54 (45.00)	0.97 (0.58-1.61)
	C	65 (54.17)	66 (55.00)	1.03 (0.62-1.72)
	Genotype (n/%)			
	T/T	1 (1.67)	0 (0.00)	
	T/C	53 (88.33)	54 (90.00)	1.12 (0.34-4.20)
	C/C	6 (10.00)	6 (10.00)	1.00 (0.27-3.74)
		P -value = 0.8415		
	T/C + C/C	59 (98.33)	60 (100.00)	2.00 (0.10-118.0)
		P -value = 0.3613		
Method 3	Allele (n/%)			
	T	78 (65.00)	76 (63.33)	0.93 (0.55-1.58)
	C	42 (35.00)	44 (36.67)	1.07 (0.63-1.82)
	Genotype (n/%)			
	T/T	24 (40.00)	26 (43.33)	1.14 (0.52-2.53)
	T/C	30 (50.00)	24 (40.00)	0.67 (0.29-1.46)
	C/C	6 (10.00)	10 (16.67)	1.67 (0.55-5.58)
		P -value = 0.3035		
	T/C + C/C	36 (60.00)	34 (56.67)	0.88 (0.40-1.19)
		P -value = 0.7150		

In sharp contrast, Method 2 resulted in the identification of a single homozygote carrying the ancestral T allele (Table 3.5). The majority of participants (53 Controls and 54 *BRCA2* mutation carriers) were heterozygous (T/C), with 12 (six Controls and six *BRCA2* mutation carriers) being homozygous for the variant C allele. The variant C allele represented on average 54.58% of all the tested alleles. Method 3 revealed similar results to Method 1, with 50 (24 Controls and 26 *BRCA2* mutation carriers) homozygous ancestral (T/T), 54 (30 Controls and 24 *BRCA2* mutation carriers) heterozygous (T/C) and 16 homozygous variant (C/C) individuals (six Controls and 10 *BRCA2* mutation carriers).

The variant C allele represented on average 35.84% of all possible alleles (Table 3.5). In total, 53 discrepancies were observed when comparing the genotype calls for each individual using the three methods (Table 3.6). The heterozygotic participants in Methods 1 and 3 were the same individuals with the exception of individual 22:1, while those homozygous for the variant allele were identical. For this SNP, Methods 1 and 3 were thus equally sensitive to identify both heterozygotes and homozygotes for the variant C allele. It was clear that a secondary factor contributed to the large discrepancy when Method 2 was compared to Methods 1 and 3. During routine sequencing to confirm the heterozygotic phenotype, a second unknown SNP (T to C) was detected 3 bp downstream from the rs3817198 SNP in *LSP1* (Fig. 3.15). This SNP was truly a fortuitous discovery as the heterozygotic sequence implicates the presence of two different plasmid vectors in the *E. coli* host, one carrying the cloned ancestral T allele and one carrying the homozygous C allele.

3.4.5 Analysis of rs889312 in *MAP3K1*

Polymorphism rs889312 entails a single nucleotide change from an A to a C within the *MAP3K1* gene. Genotyping of all participants resulted in the identification of

Table 3.6 Discrepancies between Methods 1, 2 and 3 in the genotype analysis of the rs3817198 SNP in *LSP1*.

Patient No	Method 1	Method 2	Method 3
1–1	TT	CT	TT
1–2	TT	CT	TT
1–3	TT	CT	TT
3–1	TT	CT	TT
4–1	TT	CT	TT
4–2	TT	CT	TT
4–3	CC	CT	CC
4–4	TT	CT	TT
5–1	TT	CT	TT
5–4	TT	CT	TT
6–2	TT	CT	TT
6–4	TT	CT	TT
7–2	TT	CT	TT
7–3	TT	CT	TT
9–1	TT	CT	TT
9–3	TT	CT	TT
10–1	TT	CT	TT
10–2	TT	CT	TT
10–3	TT	CT	TT
10–4	TT	CT	TT
11–1	TT	CT	TT
11–2	TT	CT	TT
12–1	TT	CT	TT
13–2	TT	CT	TT
14–2	TT	CT	TT
14–3	TT	CT	TT

Patient No	Method 1	Method 2	Method 3
15–4	TT	CT	TT
16–3	TT	CT	TT
16–4	TT	CT	TT
18–3	TT	CT	TT
19–3	CC	CT	CC
20–4	TT	CT	TT
21–1	TT	CT	TT
21–3	TT	CT	TT
22–1	CC	CT	CT
22–3	TT	CT	TT
23–4	TT	CT	TT
24–1	TT	CT	TT
24–3	TT	CT	TT
25–1	TT	CT	TT
25–4	TT	CT	TT
26–2	TT	CT	TT
26–3	TT	CT	TT
26–4	TT	CT	TT
27–1	TT	CT	TT
27–2	TT	CT	TT
28–1	TT	CT	TT
28–2	TT	CT	TT
29–1	TT	CT	TT
29–3	CC	CT	CC
29–4	TT	CT	TT
30–2	TT	CT	TT
30–3	TT	CT	TT

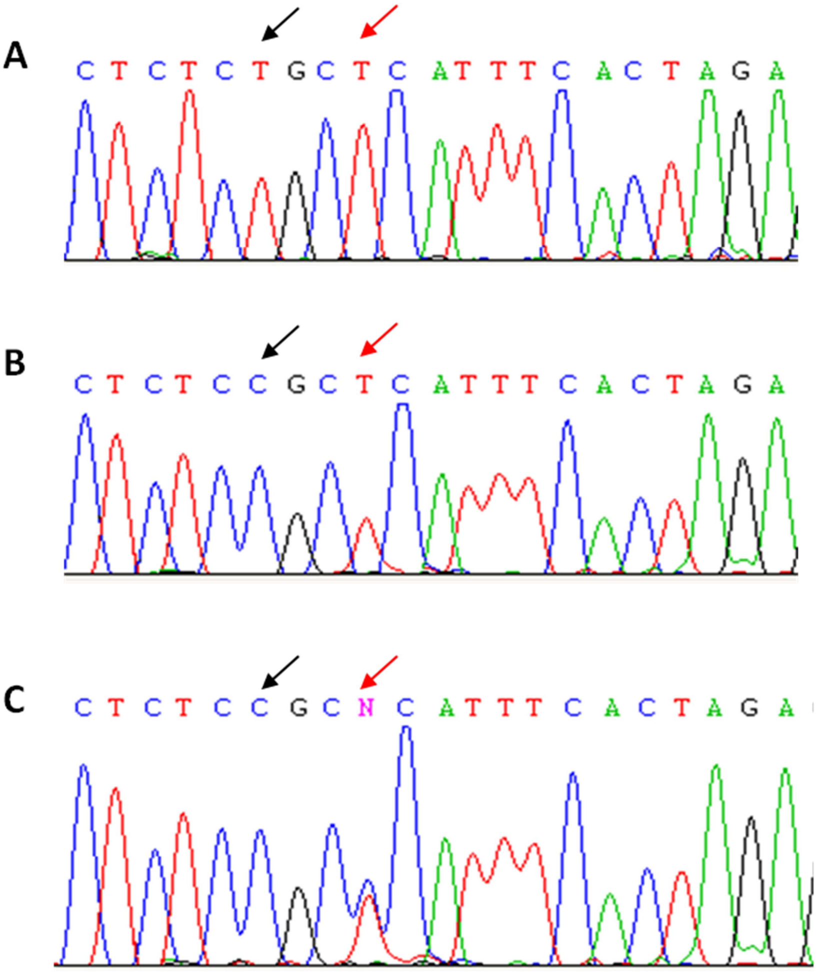


Figure 3.15 Sequencing analysis of BC patient 23–1 for a new SNP in *LSP1*. The position of the new putative SNP is indicated by the red arrow. The position of the rs3817198 SNP in *LSP1* is indicated by the black arrow. **A** Presence of the ancestral T allele for the rs3817198 SNP and the ancestral T allele for the new SNP. **B** Presence of the variant C allele for rs3817198 and the ancestral T allele for the new SNP. **C** Presence of the variant C allele for the rs3817198 SNP and the variant C allele for the new SNP.

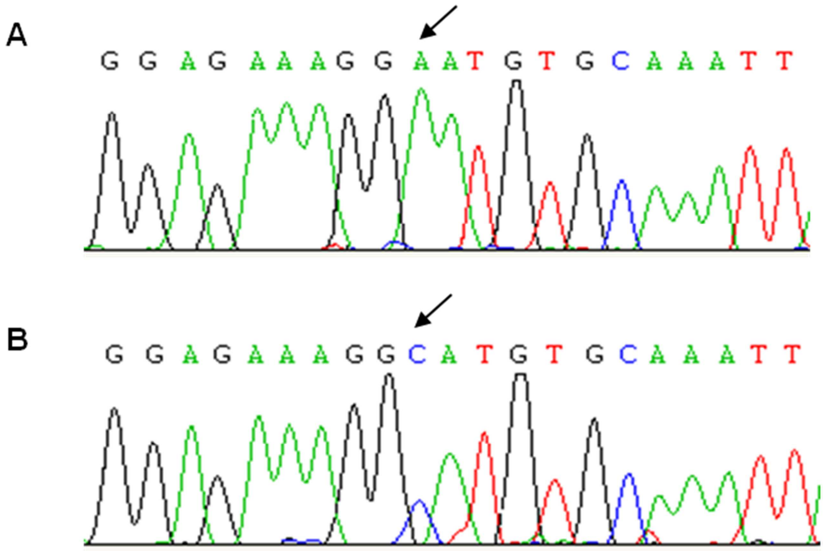


Figure 3.16 Sequencing analysis of Control 22–4, heterozygous for rs889312 in *MAP3K1*. The position of the SNP is indicated by an arrow. **A** Presence of the ancestral A allele. **B** Presence of the variant C allele.

all three possible genotypes. Sequencing of several heterozygotes confirmed the presence of both the A and C alleles (Fig. 3.16). The probes specific for the ancestral A allele and variant C alleles were labeled with HEX and FAM respectively.

For Method 1, the FAM and HEX baseline were manually adjusted to group the individuals (Fig. 3.17 A and B). For Method 2, homozygous ancestral (A/A) individuals were assigned when the amplification curves displayed only a HEX signal (Fig. 3.18 A) whereas a homozygous variant (C/C) genotype was displayed as only a FAM signal (Fig. 3.18 C). A heterozygous (A/C) genotype was recorded when both signals were present (Fig. 3.18 B). The third allelic discrimination method indicated a homozygous ancestral (A/A) individual as a HEX signal with no or a low FAM signal (Fig. 3.19 A), whereas a homozygous variant (C/C) genotype was displayed as a FAM signal with no or a low HEX signal (Fig. 3.19 C). The presence of both signals with Cq values differing with less than one were indicative of true heterozygote (T/C) (Fig. 3.19 B).

The allele and genotype frequencies for all three methods are presented in Table 3.7. According to Method 1, 84 participants (43 Controls and 41 *BRCA2* mutation carriers) were homozygous for the ancestral allele (A/A), 24 (11 Controls and 13 *BRCA2* mutation carriers) were heterozygous (A/C), with only 12 individuals (six Controls and six *BRCA2* mutation carriers) exhibited the variant C allele on both chromosomes. The variant C allele was limited to 19.17% of Controls and 20.83% of the *BRCA2* mutation carriers (Table 3.7).

Genotype analysis utilizing Method 2 scored 73 (38 Controls and 35 *BRCA2* mutation carriers) participants homozygous for the ancestral C allele, 36 (17 Controls and 19 *BRCA2* mutation carriers) heterozygous (A/C) and 11 (five Controls and six *BRCA2* mutation carriers) homozygous for the variant allele (C/C). The variant C allele was identified at an average of 24.16% of all possible alleles (Table 3.7).

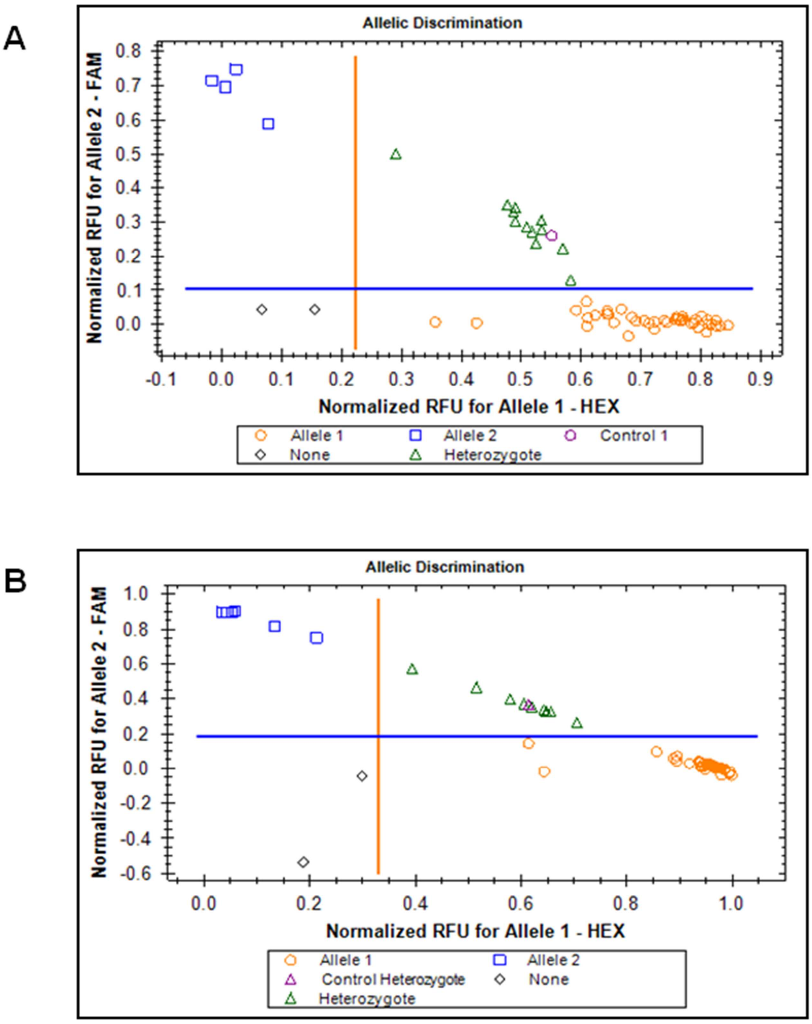


Figure 3.17 Genotyping results for rs889312 in *MAP3K1* according to Method 1 presented in two scatter plots **A** and **B.** Allele 1 represents the homozygous ancestral (A/A) genotype (HEX) and is indicated as an orange circle. Allele 2 represents a homozygous variant (C/C) (FAM) which is indicated as a blue square. Heterozygotes are represented by a green triangle whereas the positive control is indicated as a purple circle. Samples that were inconclusive are indicated by a black diamond.

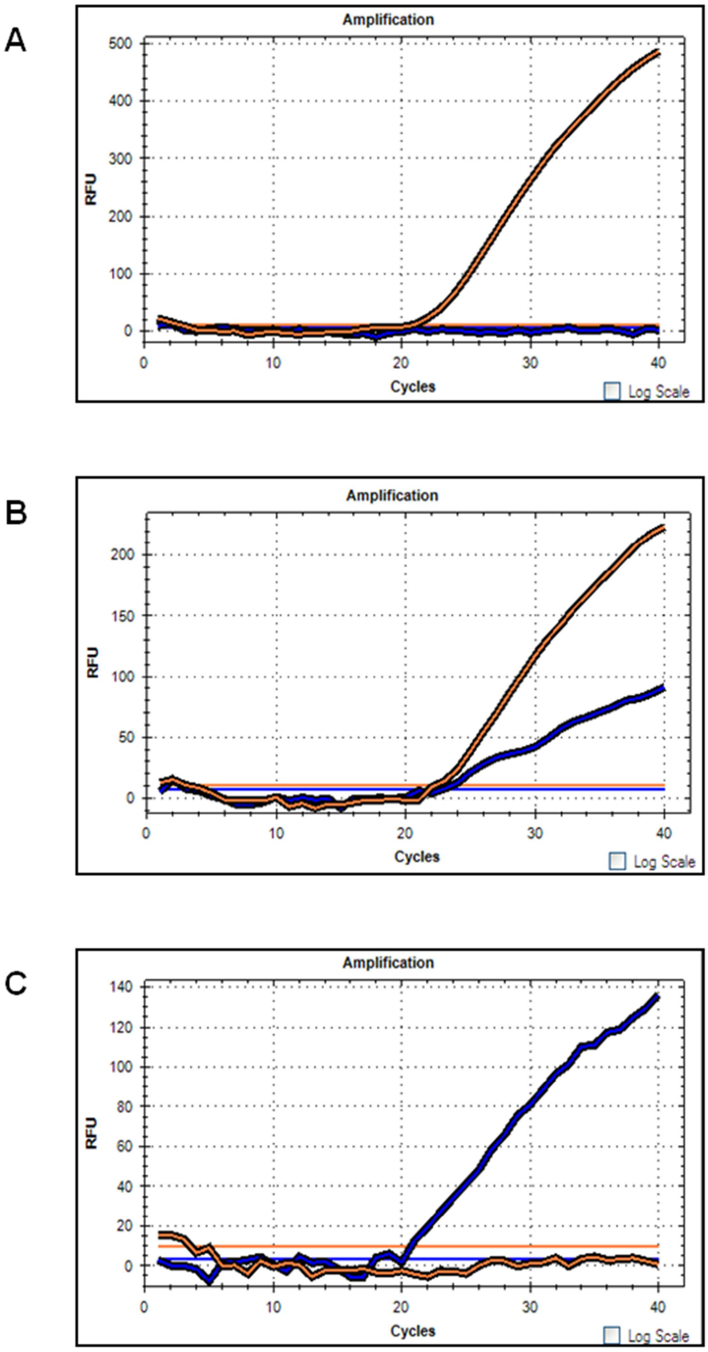
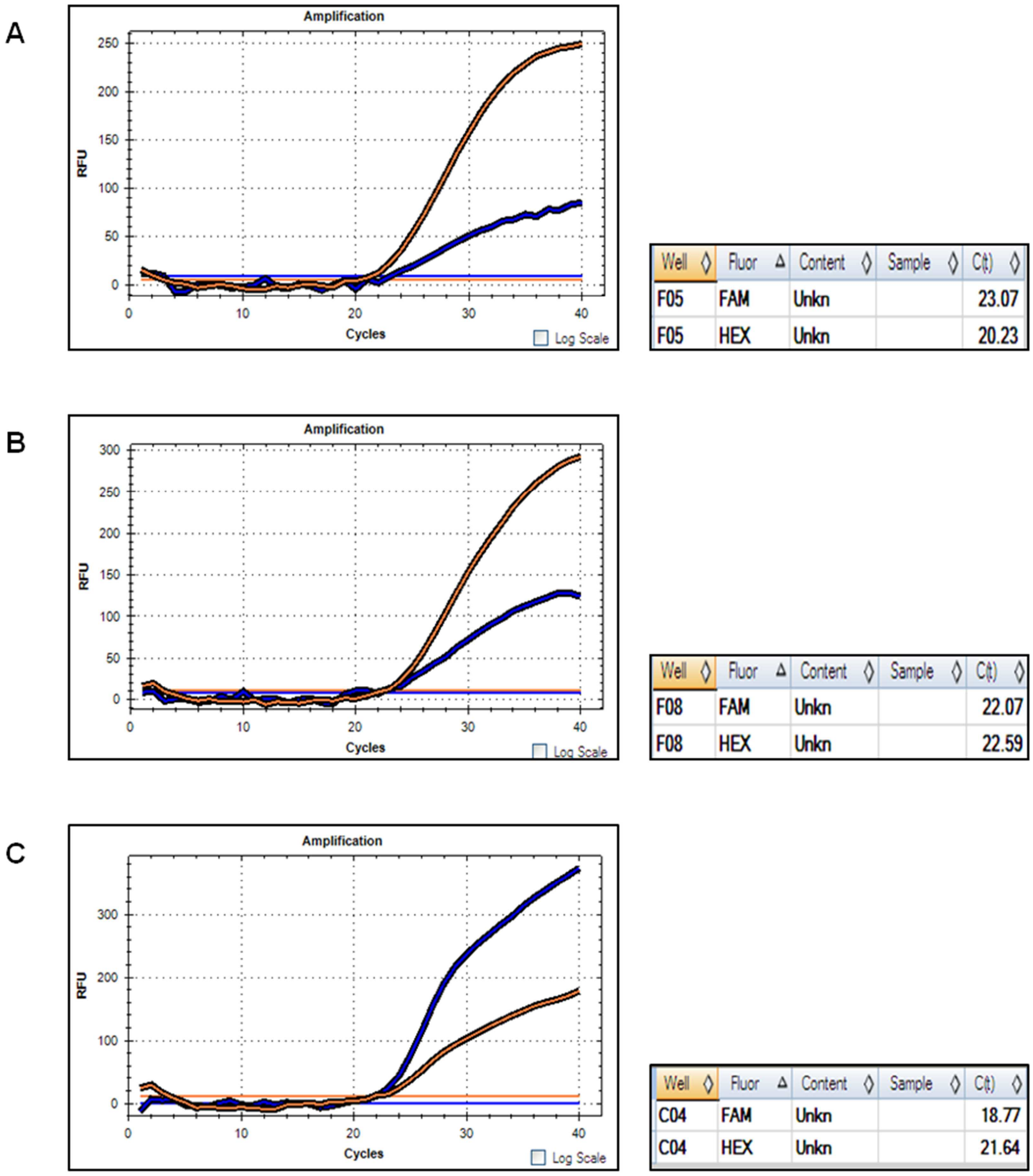


Figure 3.18 Genotype analysis of rs889312 in *MAP3K1* analyzed according to Method 2. **A** Amplification of the ancestral A allele only represented by a HEX signal. **B** Heterozygote (A/C) recognized by the amplification of both alleles represented by both the FAM and the HEX signals. **C** Participants homozygous for the variant (C/C) displayed as a FAM signal only.



A

Well	Fluor	Content	Sample	C(t)
F05	FAM	Unkn		23.07
F05	HEX	Unkn		20.23

B

Well	Fluor	Content	Sample	C(t)
F08	FAM	Unkn		22.07
F08	HEX	Unkn		22.59

C

Well	Fluor	Content	Sample	C(t)
C04	FAM	Unkn		18.77
C04	HEX	Unkn		21.64

Figure 3.19 Genotype analysis of rs889312 in *MAP3K1* according to Method 3. **A** Amplification of the ancestral A allele represented by a HEX signal with a low or no RFU signal for FAM and a Cq value differing with more than 1. **B** Heterozygous individual recognized by amplification of both alleles represented by the FAM and the HEX signals, with a Cq value deviating with less than 1. **C** Individual homozygous for the variant C allele (C/C) displayed as a FAM signal with a low or no RFU HEX signal and a Cq value of >1.

Table 3.7 Allele and genotype frequencies of rs889312 in *MAP3K1* according to Methods 1, 2 and 3.

Methods	Frequency distribution	Comparison 1 Controls vs *BRCA2* positive individuals		
		Controls n=60 n (%)	Carriers n=60 n (%)	OR (95% CI)
Method 1	Allele (n/%)			
	A	97 (80.83)	95 (79.17)	0.90 (0.48-1.70)
	C	23 (19.17)	25 (20.83)	1.11 (0.59-2.08)
	Genotype (n/%)			
	A/A	43 (71.67)	41 (68.33)	0.78 (0.25-2.35)
	A/C	11 (18.33)	13 (21.67)	1.25 (0.44-3.65)
	C/C	6 (10.00)	6 (10.00)	1.00 (0.13-7.47)
		P-value = 0.3916		
	A/C+C/C	17 (28.33)	19 (31.67)	1.29 (0.43-4.06)
		P-value = 0.6171		
Method 2	Allele (n/%)			
	A	93 (77.50)	89 (74.17)	0.83 (0.46-1.51)
	C	27 (22.50)	31 (25.83)	1.20 (0.66-2.17)
	Genotype (n/%)			
	A/A	38 (63.33)	35 (58.33)	0.69 (0.26-1.75)
	A/C	17 (28.33)	19 (31.67)	1.18(0.49-2.91)
	C/C	5 (8.33)	6 (10.00)	1.25 (0.27-6.30)
		P-value = 0.5958		
	A/C+C/C	22 (36.67)	25 (41.67)	1.33 (0.52-3.58)
		P-value = 0.5127		
Method 3	Allele (n/%)			
	A	97 (80.83)	97 (80.83)	1.00 (0.53-1.90)
	C	23 (19.17)	23 (19.17)	1.00 (0.53-1.89)
	Genotype (n/%)			
	A/A	43 (71.67)	45 (75.00)	1.22 (0.46-3.34)
	A/C	11 (18.33)	7 (11.67)	0.60 (0.18-1.82)
	C/C	6 (10.00)	8 (13.33)	2.00 (0.29-22.11)
		P-value = 0.6149		
	A/C+C/C	17 (28.33)	15 (25.00)	0.81 (0.30-2.17)
		P-value = 0.6547		

Method 3 resulted in 88 (43 Controls and 45 *BRCA2* mutation carriers) participants being homozygous for the ancestral allele, with 18 (11 Controls and seven *BRCA2* mutation carriers) heterozygous (A/C) and 14 (six Controls and eight *BRCA2* mutation carriers) homozygous for the variant C allele (Table 3.7). The variant C allele frequency was 19.17% in both the Controls and *BRCA2* mutation carriers.

In order to determine which mode of analysis was the most accurate for this SNP within *MAP3K1*, the genotype of each heterozygous and homozygous participant exhibiting the variant C allele was compared. Twenty one discrepancies were observed (Table 3.8). The main discrepancy involved homozygous individuals such as Case 7–3 genotyped by Methods 1 and 3 as A/A, which according to Method 2 was heterozygous (A/C). This was found for 12 of the individuals that indicated a difference. Another difference was five individuals genotyped by Methods 1 and 2 as heterozygous (Case 2–3, Case 4–3, Control 19–4, BC patient 23–1 and Case 28–3), that was called as homozygous for the variant allele by Method 3 (Table 3.8).

3.4.6 Genotype analysis of rs2981582 in *FGFR2*

rs2981582 in *FGFR2* entails a single nucleotide change from a C to T with the HEX fluorescently labeled probe representing the ancestral C allele and FAM the variant T allele. DNA sequencing confirmed the presence of both the T and C alleles within Control 19–4 (Fig. 3.20).

All three genotypes were observed for all three methods of genotype analysis. The automated allelic discrimination Method 1 grouped the individuals into their

Table 3.8 Discrepancies between Methods 1, 2 and 3 in the genotype analysis of the rs889312 SNP in *MAP3K1*.

Patient No	Method 1	Method 2	Method 3
2–3	AC	AC	AA
4–3	AC	AC	AA
7–3	AA	AC	AA
8–3	CC	AC	CC
10–1	AC	AA	AA
13–3	AA	AC	AA
14–1	AC	CC	CC
14–2	AA	AC	AA
14–4	AA	CC	AA
17–2	CC	AC	CC
19–4	AC	AC	AA
20–1	AA	AC	AA
20–2	AA	AC	AA
21–2	AA	AC	AC
23–1	AC	AC	CC
23–2	CC	AC	CC
24–3	AA	AC	AA
26–1	AA	AC	AA
26–3	AA	AC	AA
28–3	AC	AC	AA
30–1	AA	AC	AA

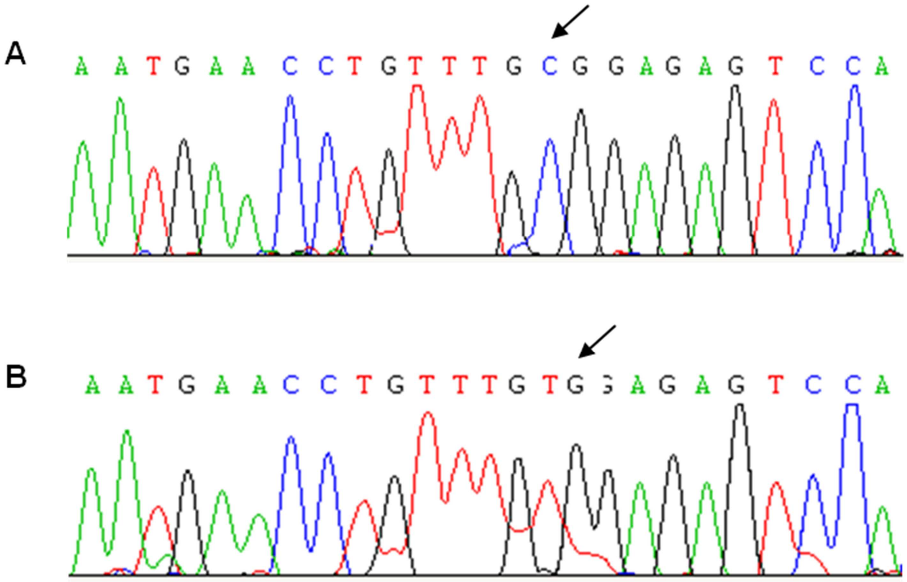


Figure 3.20 Sequence analysis of rs2981582 in *FGFR2*. The position of the SNP is indicated by an arrow. The ancestral C allele for Control 19-4 is indicated in **A** and the variant T allele in **B**.

respective clusters with a few individuals being placed on the FAM and HEX baselines (Fig. 3.21). Manual intervention was required to assign these individuals to their respective clusters. The second method displayed the homozygous ancestral (C/C) individuals as an amplified HEX signal only (Fig. 3.22 A), whereas a homozygous variant (T/T) was displayed as only a FAM signal (Fig. 3.22 C). A heterozygote (C/T) was called due to the presence of both HEX and FAM signals. The third allelic discrimination method characterized an ancestral homozygote (C/C) when only a HEX signal with no or low RFU for FAM (Fig. 3.23 A) was observed, while a homozygous variant (T/T) was displayed as a FAM signal only with no or a low HEX signal (Fig. 3.23 C). Amplification curves for both alleles with Cq values deviating with less than one was indicative of a heterozygote (C/T) (Fig. 3.23 B).

The allele and genotype frequencies for the rs2981582 SNP in *FGFR2* are presented in Table 3.9. For Method 1, 57 participants (29 Controls and 28 *BRCA2* mutation carriers) were homozygous for the ancestral C allele, whereas 46 (25 Controls and 21 *BRCA2* mutation carriers) were heterozygous (C/T) participants and 17 (six Controls and 11 *BRCA2* mutation carriers) homozygous (T/T) for the variant allele. The variant T allele was observed on average in 33.33% of participants (Table 3.9). Method 2 resulted in the identification of 51 (26 Controls and 25 *BRCA2* mutation carriers) homozygous ancestral (C/C) participants, with 61 (33 Controls and 28 *BRCA2* mutation carriers) participants being heterozygous (C/T) and only 8 (one Control and seven *BRCA2* mutation carriers) homozygous for the variant T allele. The variant T allele was observed in 29.17% and 35.00% of Controls and *BRCA2* mutation carriers respectively (Table 3.9). The allele and genotype frequencies for Method 3 revealed 66 (35 Controls and 31 *BRCA2* mutation carriers) participants being homozygous for the ancestral allele (C/C). Thirty seven (19 Controls and 18 *BRCA2* mutation carriers) individuals were heterozygous (C/T) and 17 (six Controls and 11 *BRCA2* mutation carriers) proved homozygous for the variant allele (T/T).

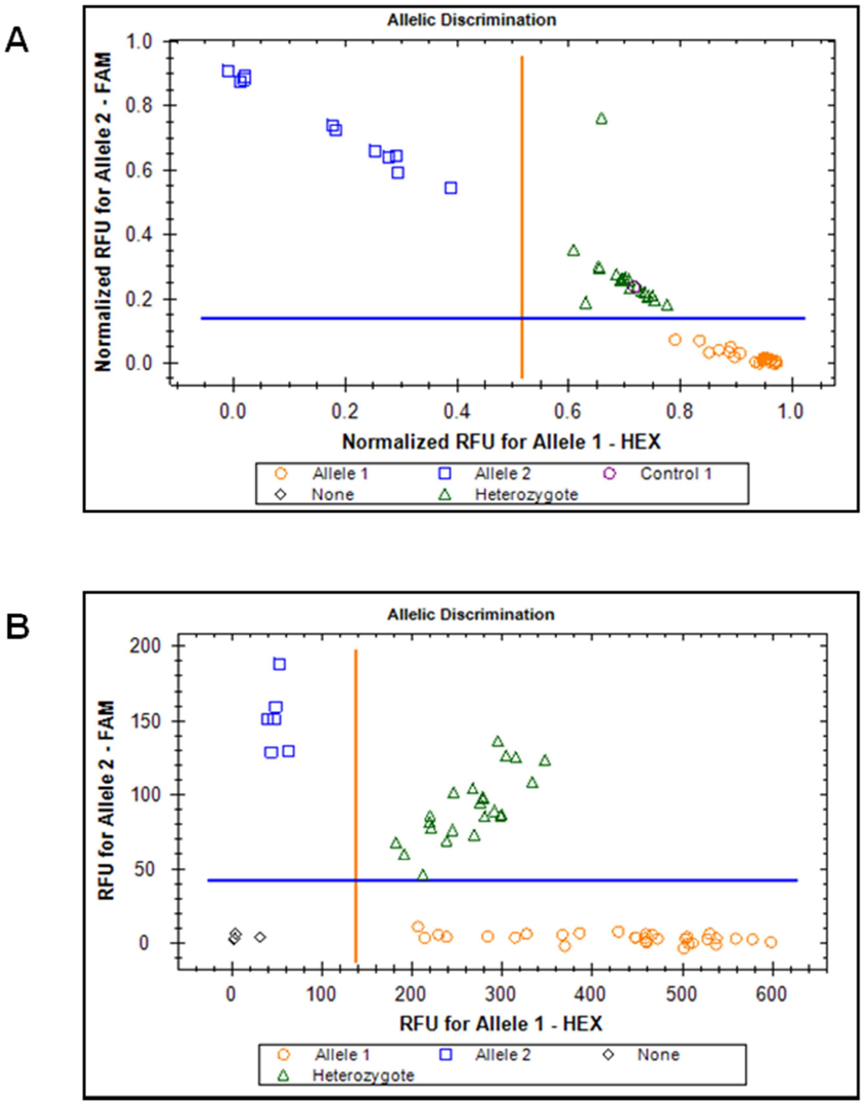


Figure 3.21 Genotyping results for rs2981582 in *FGFR2* presented in two scatter plots **A** and **B**. Allele 1 represents the homozygotic ancestral (C/C) genotype and is indicated as an orange circle. Allele 2 represents the homozygotic variant (T/T) genotype and is indicated as a blue square. Heterozygotes for the ancestral and variant alleles are represented by a green triangle whereas the positive control is indicated as a purple circle. Samples for which an inconclusive result was obtained, are indicated by a black diamond.

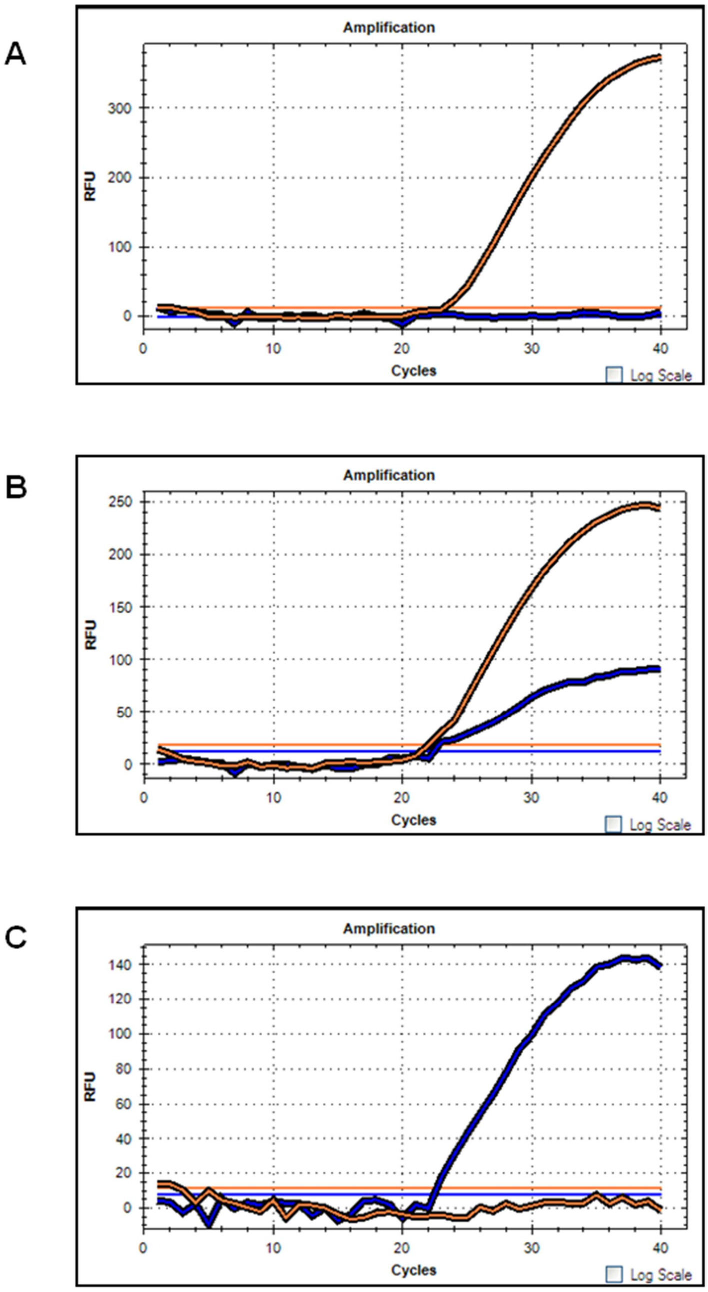
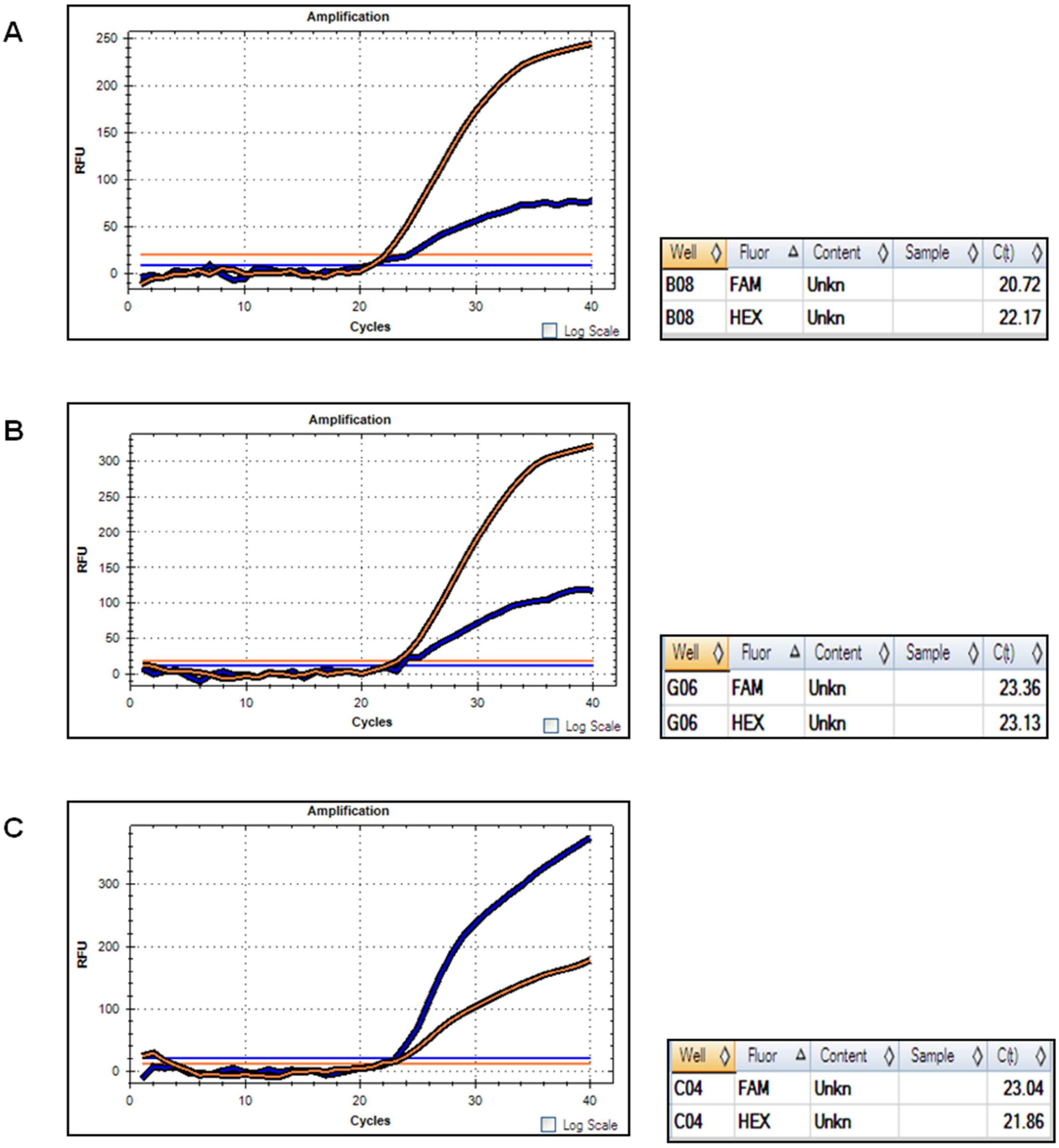


Figure 3.22 Genotype analysis of rs2981582 in *FGFR2* analyzed according to Method 2. **A** Homozygous ancestral genotype (C/C) indicated by a HEX signal with no FAM signal. **B** Heterozygote (C/T) recognized by both FAM and HEX signals. **C** Homozygous variant genotype (T/T) displayed as a FAM signal with no HEX signal.



Well	Fluor	Content	Sample	C(t)
B08	FAM	Unkn		20.72
B08	HEX	Unkn		22.17

Well	Fluor	Content	Sample	C(t)
G06	FAM	Unkn		23.36
G06	HEX	Unkn		23.13

Well	Fluor	Content	Sample	C(t)
C04	FAM	Unkn		23.04
C04	HEX	Unkn		21.86

Figure 3.23 Genotype analysis of rs2981582 in *FGFR2* according to Method 3. **A** Amplification of the ancestral C allele represented by a HEX signal with a low or no RFU signal for FAM. **B** Heterozygote displayed with both FAM and HEX signals with a Cq value deviating by less than one. **C** Homozygous variant allele (T/T) displayed as a FAM signal with a low or no RFU HEX signal.

Table 3.9 Allele and genotype frequencies of rs2981582 in *FGFR2* according to Methods 1, 2 and 3.

Methods	Frequency distribution	Comparison 1 Control vs *BRCA2* mutation carriers		
		Controls n=60 n (%)	Carriers n=60 n (%)	OR (95% CI)
Method 1	Allele (n/%)			
	C	83 (69.17)	77 (64.17)	0.80 (0.47-1.37)
	T	37 (30.83)	43 (35.83)	1.25 (0.73-2.13)
	Genotype (n/%)			
	C/C	29 (48.33)	28 (46.67)	0.94 (0.43-2.03)
	C/T	25 (41.67)	21 (35.00)	0.77 (0.34-1.67)
	T/T	6 (10.00)	11 (18.33)	2.00 (0.62-7.46)
		P-value = 0.3062		
	C/T + T/T	31 (51.67)	32 (53.33)	1.07 (0.49-2.32)
		P-value = 0.8575		
Method 2	Allele (n/%)			
	C	85 (70.83)	78 (65.00)	0.76 (0.44-1.32)
	T	35 (29.17)	42 (35.00)	1.31 (0.76-2.27)
	Genotype (n/%)			
	C/C	26 (43.33)	25 (41.67)	0.94 (0.43-2.03)
	C/T	33 (55.00)	28 (46.67)	0.72 (0.33-1.56)
	T/T	1 (1.67)	7 (11.67)	7.00 (0.90-315.48)
		P-value = 0.1691		
	C/T + T/T	34 (56.67)	35 (58.33)	1.07 (0.49-2.32)
		P-value = 0.8575		
Method 3	Allele (n/%)			
	C	89 (74.17)	80(66.67)	0.70 (0.40-1.22)
	T	31 (25.83)	40(33.33)	1.43 (0.82-2.50)
	Genotype (n/%)			
	C/C	35 (58.33)	31 (51.67)	0.78 (0.36-1.66)
	C/T	19 (31.67)	18 (30.00)	0.92 (0.37-2.27)
	T/T	6 (10.00)	11 (18.33)	2.00 (0.62-7.46)
		P-value = 0.6444		
	C/T + T/T	25 (41.67)	29 (48.33)	1.29 (0.60-2.79)
		P-value = 0.4795		

The variant T allele was present in an average of 29.58% of all possible alleles. When comparing both the heterozygous and homozygous variant allele frequencies between the three methods, it was revealed that the results of Methods 1 and 3 were more similar compared to Method 2.

Discrepancies were observed in the results for 27 participants between the three modes of analysis (Table 3.10). The majority of the discrepancies observed were seen amongst the heterozygous and homozygous variant genotypes. Nine patients were typed as homozygous ancestral by both Methods 1 and 3 (for example Case 2–3), but heterozygous when analyzed with Method 2.

3.4.7 Cohen's Kappa chance of agreement between allelic discrimination methods.

Cohen's kappa (Cohen, 1960) is a measurement of agreement between different modes of analysis that is actually present, compared to the agreement that would be expected by chance alone. The kappa is standardized on a -1 to 1 scale. Perfect agreement would be equal to a kappa value of 1, while chance agreement would be equal to 0. A negative value show worse than chance agreement. The interpretation of kappa values was according to the categorization by Fleiss. (1981). He proposed that a kappa value of > 0.75 represents excellent agreement, values between 0.4 and 0.74 fair to good and values < 0.4 poor.

The Kappa chance of agreement between the different methods of allelic discrimination was measured for the Controls and *BRCA2* mutation positive individuals for all four SNPs (Table 3.11).

Table 3.10 Discrepancies between Methods 1, 2 and 3 in the genotype analysis of the SNP rs2981582 in *FGFR2*.

Patient No	Method 1	Method 2	Method 3
2–3	TT	CT	TT
3–4	TT	CT	TT
5–1	CT	CC	CC
5–2	CT	CT	CC
5–3	CT	CT	CC
5–4	CC	CT	CT
6–3	CC	CT	CC
7–2	CT	CT	CC
7–4	CC	CT	CC
8–2	CT	CT	CC
8–3	TT	CT	TT
8–4	CT	CT	CC
9–2	TT	CT	TT
9–3	CT	CT	CC
11–2	CT	CT	CC
13–4	TT	CT	TT
16–3	CT	CT	CC
18–1	CC	CT	CC
18–4	CT	CT	CC
19–2	TT	CT	TT
20–2	CC	CT	CC
23–1	CC	CT	CT
24–2	CT	CT	CC
26–3	CC	CT	CC
27–3	TT	CT	TT
28–2	TT	CT	TT
30–1	TT	CT	TT

Table 3.11 Kappa chance of agreement analysis of the three employed allelic discrimination methods.

Comparisons	**SNP in gene**	**Comparison 1 Controls vs *BRCA2* positive individuals**			
		Controls		***BRCA2* positive individuals**	
		Kappa value	**OR (95% CI)**	**Kappa value**	**OR (95% CI)**
Method 1 vs Method 2	***TNRC9***	0.66	(0.47-0.84)	0.73	(0.49-0.97)
	LSP1	0.34	(0.16-0.52)	0.24	(0.09-0.39)
	MAP3K1	0.78	(0.63-0.94)	0.74	(0.59-0.89)
	FGFR2	0.78	(0.66-0.91)	0.80	(0.68-0.92)
Method 1 vs Method 3	***TNRC9***	0.76	(0.51-1.00)	0.67	(0.41-0.92)
	LSP1	1.00		0.98	(0.94-1.00)
	MAP3K1	0.94	(0.86-1.00)	0.84	(0.71-0.96)
	FGFR2	0.80	(0.67-0.94)	0.90	(0.81-0.99)
Method 2 vs Method 3	***TNRC9***	0.43	(0.20-0.66)	0.64	(0.29-0.98)
	LSP1	0.34	(0.16-0.52)	0.25	(0.09-0.40)
	MAP3K1	0.78	(0.63-0.94)	0.70	(0.54-086)
	FGFR2	0.63	(0.47-0.78)	0.78	(0.66-0.91)

The results revealed good to excellent agreement between Methods 1 and 3. The most perfect agreement was seen for *LSP1* with a Kappa value of 1.00 and 0.98 for the Controls and *BRCA2* positive individuals respectively. *MAP3K1* and *FGFR2* also indicated excellent agreement between the Controls and *BRCA2* mutation positive individuals with Kappa values of 0.94 and 0.84 and 0.80 and 0.90, respectively. The agreement for *TNRC9* was excellent to fairly good with values of 0.76 and 0.67 for the Controls and *BRCA2* mutation positive individuals, respectively.

The comparison between Methods 1 and 2 deliver excellent to good results with *MAP3K1* (0.78 and 0.74) and *FGFR2* (0.78 and 0.80) and fairly good agreement for *TNRC9* (0.66 and 0.73). However, the agreement was poor for *LSP1* with Kappa values of 0.34 and 0.24. The Kappa values for the comparison between Method 2 and Method 3 revealed a good to excellent agreement for *MAP3K1* (0.78 and 0.70), *FGFR2* (0.63 and 0.78) and *TNRC9* (0.43 and 0.64) with *LSP1* (0.34 and 0.25) having a poor agreement.

3.5 Discussion

The primary aim of this chapter was to compare manual and automatic allelic discrimination methods and study the different outcomes. Sixty *BRCA2* 8162delG (c.7934del, p.Arg2645AsnfsX3) mutation carriers of Caucasian Afrikaner descent, both affected and unaffected with BC, were selected for the study. These patients were screened for SNPs occurring within the *TNRC9* (rs3803662), *LSP1* (rs3817198), *MAP3K1* (rs889312) and *FGFR2* (rs2981582) genes using theTaqMan® SNP genotyping assay according to Easton *et al*. (2007). The genotype and allele frequencies were compared and the Kappa chance of agreement was measured.

3.5.1 Allelic discrimination methods

Analysis of the TaqMan® SNP genotyping assay proved to be more complicated than initially anticipated. The majority of literature describing TaqMan® genotyping assays do not specify genotype calling techniques but mostly use the specific real-time PCR detection system software to analyze the data (Antoniou *et al*., 2009; Reeves *et al*., 2010; Travis *et al*., 2010). In this study the obtained SNP data was analyzed using the BioRad CFX Manager v1.1.308.1111 software (Method 1). For accurate allelic discrimination the software requires manual intervention to adjust the fluorescent signal thresholds and assign genotypes when rare variant alleles are present. This posed a problem as an expert is needed to adjust the software for accurate analyses and to exclude biased genotyping. According to various specialists such as Callegaro *et al*. (2006), most software programs use auto scaling for genotype calling. This causes a problem when analyzing rare alleles. With this in mind two additional methods for manual allelic discrimination were explored.

No conclusive description of a manual genotype calling method was found even after an extensive literature search was done. After personal communication with several representatives of different real-time PCR detection system supplier companies (Roche, Applied Biosystems and Bio-Rad), the first manual allelic discrimination method was implemented. This method (Method 2) was based on the theory that TaqMan® probes are highly specific and mismatch binding will therefore be greatly reduced. The discrimination was done with the concept that any increase above the baseline indicated true amplification regardless the RFU value. This method utilized the amplification plots for allelic discrimination. Problems with the manual genotyping were experienced when some duplicate samples revealed different Cq and RFU values between the different runs for one specific probe. In these cases the sample was run for a third time.

Personal communication with Dr. K Livak (Livak *et al.*, 1995; Livak *et al.*, 1999) supported the theory that TaqMan® probes are not always specific and mismatch pairing of the probes can show a low fluorescent signal. Therefore Method 3 used both the Cq and RFU values for discrimination. A heterozygote was only scored when the Cq values of the FAM and HEX amplification plot differed with less than one Cq value. A Cq difference value of more than one thus represented homozygosity for either the ancestral or variant allele. The homozygous genotype was determined by taking the respective RFU value into account.

3.5.2 Comparison of the manual and automatic allelic discrimination methods

When comparing the allele frequencies for the rs3803662 SNP in *TNRC9* (Table 3.3), Method 2 and Method 3 gave the same number of variant T alleles (14 of a possible 240). It does however seem as if Method 3 was most stringent for the identification of the variant T allele, since the largest number of homozygous individuals for the ancestral C allele was identified with this method. This could imply that the algorithm used by automated calling is based on a higher Cq cut-off value above one which is why more heterozygotes and homozygotes were called. However, when contacted, Bio-Rad confirmed that the algorithm uses a Cq value of one.

The genotype and allele frequencies for rs3817198 SNP in *LSP1* (Table 3.5) indicated that Methods 1 and 3 were equally sensitive regarding the identification of heterozygotes (T/C). Method 2 on the other hand identified almost twice as many heterozygotes (89.00%). This was due to the fact that the manual genotyping procedure utilized for Method 2 was based on any amplification above the baseline, irrespective of the RFU value. Many of the participants showed weak amplification above the baseline for the C allele indicated by the HEX signal. That resulted in many individuals being genotyped as heterozygous instead of homozygous.

It can be speculated that the presence of the new unknown SNP, located three base pairs downstream of the rs3817198 SNP, influenced the binding of both probes, thus lowering the T_a values. This caused weaker amplification above the baseline with lower RFU values and higher Cq values that affected the genotype calls for Method 1 and Method 3. A participant originally genotyped as homozygous ancestral (T/T) or homozygous variant (C/C) according to these methods would have shown a strong amplification (high RFU and low Cq value) for either the ancestral or variant probe with no or weak amplification for the other probe which would be considered as non-specific binding that will be discarded. However, genotyping using Method 2 would correctly indicate this as a heterozygote (T/C) even though the Cq values differed with more than one Cq value. For *LSP1*, Method 2 thus represents the more correct method of genotyping, due to the putative presence of the second SNP that can skew the results. This SNP must however first be confirmed in more white Afrikaner individuals through sequencing. This result however could question the validity of published *LSP1* genotyping calls using Method 1 since no information on the frequency of the second putative SNP in white European individuals is available.

Although all three methods revealed similar frequencies for the rs889312 SNP in *MAP*3*K*1, Method 3 was the most stringent in identifying heterozygotes. Variantion was observed between the three methods for the rs2981582 SNP in *FGFR*2 among the homozygous ancestral, heterozygous and homozygous variant genotypes. The detected discrepancies indicated that the variant C allele was amplified, but with lower RFU values and a Cq difference > 1. That could be the reason why a genotype of CT detected by Method 2, did not correspond to the results obtained by Methods 1 or 3 indicating a degree of non-specific amplification by the variant probe. This is also illustrated by various homozygous variant individuals (Case 6–3, Control 7–4) that were heterozygous when analyzed by Method 2. Method 3 proved to be the most stringent in identifying heterozygotes.

Cohen’s kappa analysis suggested that Methods 1 and 3 were the closest matched with an excellent (> 0.75) agreement for *LSP1* (rs3817198), *FGFR2* (rs2981582) and *MAP3K1* (rs889312) for both the Controls and the *BRCA2* positive individuals. An excellent agreement (0.76) was observed for *TNRC9* (rs3803662) between the Controls and a good (0.67) agreement for the *BRCA2* positive individuals.

These statistics were expected as both Method 1 and Method 3 take the Cq and RFU values into account unlike Method 2 that only focused on the RFU values. From these results it can be concluded that Method 1 and Method 3 (except for *LSP1*) are best suited for accurate allelic discrimination. Most literature support allelic discrimination by the specific real-time PCR detection system software therefore Method 1 will be used for genotype calling of the four SNPs. The accuracy of the data does depend on careful probe design, optimization of PCR conditions and the inclusion of positive controls. As long as these requirements are met and three distinct scatter plot clusters are observed, robust genotyping can be performed.

Chapter 4

Influence of selected polymorphisms on the expression of breast cancer in Afrikaner *BRCA2* carriers

4.1 Introduction

Breast cancer is the most common malignancy in women in the western world (Dumitresu and Cotarla, 2005*)*. In SA, BC is currently the most commonly diagnosed cancer, with a life-time risk of 1 in 12 for Caucasian South African women and 1 in 49 for black South African women (Loubser, 2008). Although most BC cases are sporadic, a small but significant percentage (5 -10%) accounts for a hereditary predisposition (Claus *et al*., 1996). These familial cases are due to mutations in several genes, with 15 to 20% being explained by germline mutations in the two highly penetrant *BRCA1* and *BRCA2* genes (Silla *et al.*, 1995; Claus *et al*., 1996). During the last five years, various differences between mutation positive individuals within the same family have been internationally observed and reported. This includes the age at onset and the type of cancer present within the affected cases (Antonio *et al.*, 2003; Simchoni *et al*., 2006). Evidence suggests that *BRCA* mutations are co-determined by environmental and genetic factors which may act in an additive fashion to increase BC risk in woman (Easton, 1999; Peto, 2002; Antoniou *et al*., 2003; Dapic *et al*., 2005). Various studies explained that such a polygenic model can contribute to these inter-individual phenotypic differences among *BRCA* mutation carriers (Antoniou *et al.*, 2002; Pharoah *et al*., 2002; Wooster and Weber, 2003).

Segregation analysis studies demonstrated that models which allow genes to have a modifying effect on BC risks conferred by *BRCA1* and *BRCA2* mutations, fit significantly better than models without a modifying component (Antonio *et al.*, 2005). This finding resulted in an international search for genetic modifiers of cancer risk conferred by the *BRCA* genes. Candidate gene studies were based on a selection of low penetrance genes that are involved in the biochemical and physiological pathways of carcinogenesis.

The selection of appropriate candidate polymorphisms is complicated as it depends on a basic knowledge of the biochemical and physiological pathways of carcinogenesis. This problem was overcome through the identification of candidate genes with adequately powered GWAS (Hirschhorn and Daly, 2005) and the publication of validated SNPs associated with BC in the general population (Cox *et al.* 2007).

Some genes that the GWAS studies have indicated as potential role players in the development of BC include *FGFR2* (OMIM 176943), *LSP1* (OMIM 153432), *MAP3K1* (OMIM 600982) and *TNRC9* (OMIM 611416) (Easton *et al*., 2007). The *ESR1* (OMIM 133430) gene is another one of these low penetrance candidate genes that could be involved in BC risk (Siddig *et al*., 2008). Estrogen is an important epidemiologic risk factor and its effects are mediated through the ER in breast tissue. It is reported that estrogen plays a crucial role in breast growth, differentiation and the development of cancer.

Despite considerable interest in the SNPs in these candidate genes and the influence on BC risk, only a modest amount of studies were published on *BRCA2* mutation carriers specifically. Apart from the fact that conflicting results exist regarding these SNPs in the five genes and their role on BC risk and penetrance, it was proposed that because the South African Afrikaner is such a universally unique population group (founder effects have been proven for various diseases such as porphyria variegata and familial hypercholesterolemia),the search and effort on the *BRCA2* 8162delG (c.7934del, p.Arg2645AsnfsX3) Afrikaner population would be justified. Therefore the aim of this study was to investigate the presence and the effect of six selected SNPs that have been previously proven to be associated with an increased BC risk. Furthermore no studies have been published on the influence of these candidate genes on carriers of the *BRCA2* 8162delG (c.7934del, p.Arg2645AsnfsX3) mutation in the Afrikaner population to date.

4.2 Methods

4.2.1 Subjects

Sixty *BRCA2* 8162delG (c.7934del, p.Arg2645AsnfsX3) mutation carriers of Caucasian Afrikaner decent, both affected and unaffected with BC, were selected for the study. These mutation carriers were case-matched with control participants representing Afrikaner women who do not have a family history of BC. The selection criteria for mutation carriers, Controls as well as the criteria for the grouping are described in 3.2.1. Ethical permission was obtained and each individual was given a unique number to ensure confidentiality as described in 3.2.2.

4.2.2 DNA extraction

Peripheral blood (10 – 20 ml) was collected in EDTA vacutainer tubes and stored at -20°C. Genomic DNA was extracted from lymphocytes using the phenol chloroform procedure as described in 3.3.1.

4.2.3 Molecular analysis of two SNPs in the *ESR1* gene

The presence or absence of two SNPs within the *ESR1* gene (rs2234693 and rs9340799) was analyzed by restriction digestion of a 1.3 kb PCR amplicon. The two SNPs are 50 bp apart on the 1.3 kb PCR amplicon (Cai *et al*., 2003; Hsieh *et al.*, 2007).

4.2.3.1 PCR amplification of the 1.3 kb amplicon

The PCR conditions used for the amplification of the 1.3 kb fragment were adapted from the standard operating procedure (SOP) (UNIG0032) from the Human Genetics Laboratory at Universitas Hospital in Bloemfontein. To optimize the annealing temperatures (T_a) for each primer set, a series of PCR reactions were performed on the MyCycler™ (Biorad) utilizing a T_a gradient. The annealing temperatures for the gradient were adapted from the oligonucleotide specification sheets provided by IDT and the annealing temperature as described by Hsieh *et al*. (2007). Primer sequences and expected amplicon sizes are indicated in Table 4.1.

Each 50 µl PCR reaction contained 300 ng template DNA, 250 µM dNTPs, 10 pmol of each primer, 100 mM Tris-HCl pH 8.3, 1.5 mM $MgCl_2$, 500 mM KCl and 1 U *Taq* DNA polymerase (Roche). The negative controls contained all the PCR components except genomic DNA. The amplification regime included one cycle at 94℃ for 2 min, followed by 35 cycles of 94℃ for 1 min, optimal annealing temperature for 1 min and 72℃ for 1 min with a final elongation step at 72℃ for 10 min.

Table 4.1 Oligonucleotides used for the molecular analysis of rs2234693 (*PvuII*) and rs9340799 (*XbaI*) indicating the primer sequence, annealing temperature and fragment lengths. T_a represents the optimal annealing temperature for each primer set (Cai *et al*., 2003; Hsieh *et al.*, 2007).

Polymorphisms	**Forward Primer**	**Reverse Primer**	**T_a (ºC)**	**Allelic variation**	**Fragment size (kb)**
rs2234693 Intron 1 C/T	5'-CTGCCACCCTATCTGTATCTTTTCCTATTCTCC-3'	5'-TCTTTCTCTGCCACCCTGGCGTCGATTATCTGA-3'	56℃	C: Ancestral T: Variant	C: 1.3 kb T: 0.85 kb + 0.45 kb
rs9340799 Intron 1 A/G	5'-CTGCCACCCTATCTGTATCTTTTCCTATTCTCC-3'	5'- TCTTTCTCTGCCACCCTGGCGTCGATTATCTGA-3'	56℃	A: Ancestral G: Variant	A: 0.9 kb + 0.4 kb G: 1.3 kb

Amplification products (10 µl) were separated on a 2% (w/v) agarose gel to confirm successful amplification (Sambrook *et al*., 1989). The agarose gel was prepared in 1x Tris-Borate-EDTA (TBE) buffer (0.089 M Tris pH 8, 0.089 M boric acid, 2 mM EDTA) and contained ethidium bromide (EtBr) to a final concentration of 0.5 µg/ml. The gel was run at 20 $V.cm^{-1}$ using 1x TBE as running buffer. All subsequent PCR reactions for these amplicons used 56℃ as the optimal annealing temperature.

4.2.3.2 Restriction fragment length polymorphism (RFLP) analysis

A 10 µl aliquot of each amplified fragment was digested with 10 U PvuII and XbaI restriction endonucleases respectively (Fermentas). The PvuII digestion was performed in the presence of 10 mM Tris-HCl pH 7.5, 10 mM $MgCl_{2,}$ 50 mM NaCl and 0.1 mg/ml Bovine Serum Albumin (BSA), where as the regime for XbaI digestion entailed using 33 mM Tris-acetate pH 7.9, 10 mM magnesium acetate, 66 mM potassium acetate and 0.1 mg/ml BSA. Digestion was done at 37℃ for 1 h.

Twenty microliters of the digested products were mixed with 5 µl loading buffer [0.25% (w/v) orange G, 40% (w/v) sucrose] and separated on a 3% (w/v) NuSieve® 3:1 agarose gel (Lonza group Ltd) (4.2.3.1).

Digested fragments of 0.85 kb and 0.45 kb in size for rs2234693 (*PvuII*) depicted the variant T-allele, whereas fragments of 0.9 kb and 0.4 kb for rs9340799 (*XbaI*) indicated the presence of the ancestral A-allele. Separated DNA fragments were analyzed with a Gel Doc™ XR gel documentation system using the Quantity One® 1-D analysis software (BioRad).

4.2.3.3 DNA cycle sequencing

Bi-directional DNA sequencing of PCR products was performed for 10% of all participants to confirm the genotyping calls. Each sample was re-amplified, where after 10 µl of the PCR products were separated on a 2% (w/v) agarose gel to confirm successful amplification (4.2.3.1). The remaining 40 µl PCR product was purified using SigmaSpin Post-Reaction clean-up columns (Sigma) according to the manufacturer's conditions. DNA fragments were bi-directionally sequenced using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). Each sequencing reaction contained 1 µl (300 ng/µl) PCR template, 4 µl terminator ready reaction mix, 3.2 pmol primer and 2 µl BigDye® sequencing buffer to a final volume of 20 µl. The cycle sequencing regime was as follows: one cycle at 96℃ for 1 min, followed by 25 cycles at 96℃ for 10 sec, 56℃ for 5 sec and 60℃ for 4 min with a final holding temperature of 4℃. The products were precipitated by adding 10 µl water, 5 µl 125 mM EDTA and 60 µl 100% (v/v) ethanol, followed by 15 min incubation at room temperature in the dark. After centrifugation at 14 000 *g* for 30 min at 4°C, the supernatant was aspirated and the pellet washed with 60 µl 70% (v/v) ethanol. The pellet was finally air-dried in the dark.

Hi-Di™ Formamide (Applied Biosystems) buffer (30 µl) was added to the pellet, followed by denaturation for 5 min at 96℃ where after the tube was snap cooled. Products were then mixed by pipetting and loaded on a microtiter plate for analysis on the ABI Prism 310 Genetic Analyzer (Applied Biosystems). Electropherograms were analyzed using proprietary sequence analysis software (Chromas version 2.31, www.technelysium.com.au) followed by visual inspection and confirmation. The sequences were aligned with the ancestral reference sequence (NG_008493.1) for each gene using LALIGN (www.ch.embnet.org/software/LALIGN).

4.2.4 Molecular analysis of SNPs presented in ***TNRC9*****,** ***LSP1*****,** ***MAP3K1*** **and** ***FGFR2***

SNP polymorphisms occurring within *TNRC9* (rs3803662), *LSP1* (rs3817198), *MAP3K1* (rs889312) and *FGFR2* (rs2981582) were respectively screened for using a modified TaqMan® SNP genotyping assay as described in Chapter 3 (Easton *et al*., 2007). Allelic discrimination was performed on the BioRad CFX96^{TM} Real-Time PCR Detection System using the BioRad CFX Manager v1.1.308.1111 software. TaqMan® fluorescent probes labeled with FAM and HEX respectively, as well as the different PCR primers, were synthesized (Table 3.2) (Easton *et al*., 2007).

The TaqMan® assay, DNA cloning and direct plasmid DNA sequencing were performed as described in Chapter 3. The amplicons of putative heterozygotes for each SNP were sequenced to confirm the genotyping calls.

4.2.5 Statistical analysis

Statistical analysis of data was performed by the Department of Biostatistics at the University of the Free State. Results were summarized by frequencies and percentages for categorical variables. The analysis was initially performed between the Controls and the *BRCA2* mutation carriers (Comparison 1). Odds ratios and CI using McNemar's test were calculated (Fleiss *et al*., 2003). Subsequently the mutation carrier group was divided into the BC affected patients and unaffected cases (Comparison 2). These two groups were analyzed separately to evaluate the independent effect of each genotype with regards to BC risk. McNemar's test was used to assess the difference between the two correlated dependent proportions (http://faculty.vassar.edu/lowry/propcorr.html). Fisher's exact test of Hardy-Weinberg equilibrium (HWE) was used to determine the *P*-value for the Control and Case groups separately as well as the *BRCA2* mutation carriers and BC patients. Allelic distributions with a *P*-value of <0.05 is considered to deviate from the HWE. Genotype calling was re-analyzed in both control and case groups where deviations from the HWE was observed in order to exclude possible errors made in genotype assignment. Statistical analyses were performed with Arlequin 3.1 software (http://cmpg.unibe.ch/software/arlequin3) (Excoffier *et al*., 2005).

4.3 Results

4.3.1 Optimization of PCR conditions for the 1.3 kb *ESR1* amplicon

The amplicon surrounding the two polymorphisms were PCR amplified using a single pair of primers (Table 4.1). PCR gradient analysis resulted in amplicons of high quality for all the temperatures tested, so a final T_a of 56℃ was selected (Fig. 4.1). The amplified products were of the correct size (1300 bp) with no secondary amplicons being evident.

4.3.2 Analysis of rs2234693 (*PvuII*) in *ESR1*

Analysis of the rs2234693 SNP (C>T) was done according to Cai *et al*. (2003) and Hsieh *et al.* (2007) using PCR based RFLP. The original 1300 bp PCR amplicon was digested with PvuII to detect the variant T allele while an undigested amplicon was included to assist with genotyping calls (Fig. 4.2). PvuII digestion of the 1300 bp amplicon resulted in three possible banding patterns representing the three genotypes. The presence of a single 1300 bp fragment represented homozygosity for the ancestral allele (C/C, lane 6), while three fragments (1300 bp, 850 bp and 450 bp) indicated heterozygosity for the ancestral allele (C/T, lane 3). Two fragments of 850 and 450 bp represented homozygosity for the variant T allele (T/T, lane 2). All three possible genotypes were observed within the 120 tested individuals, with heterozygotes being the most common.

DNA sequencing confirmed the genotypes of Case 5–3 as being homozygous ancestral (C/C), BC patient 6–1 as heterozygous (C/T) and Case 2–3 as homozygous variant (T/T) (Fig. 4.3). Alignment of the sequence for BC patient 6–

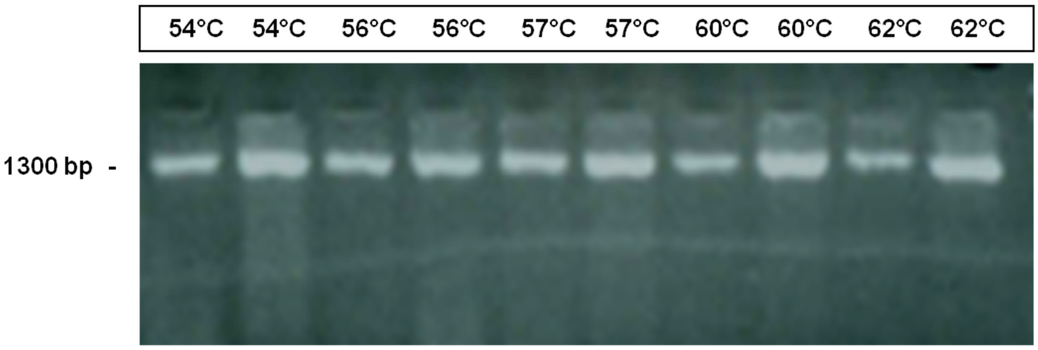


Figure 4.1 Optimization of the T_a value for the PCR amplification of the 1300 bp product of the rs2234693 (*PvuII*) and rs9340799 (*XbaI*) SNP in *ESR1*. A temperature gradient ranging from 54 to 62℃ was used.

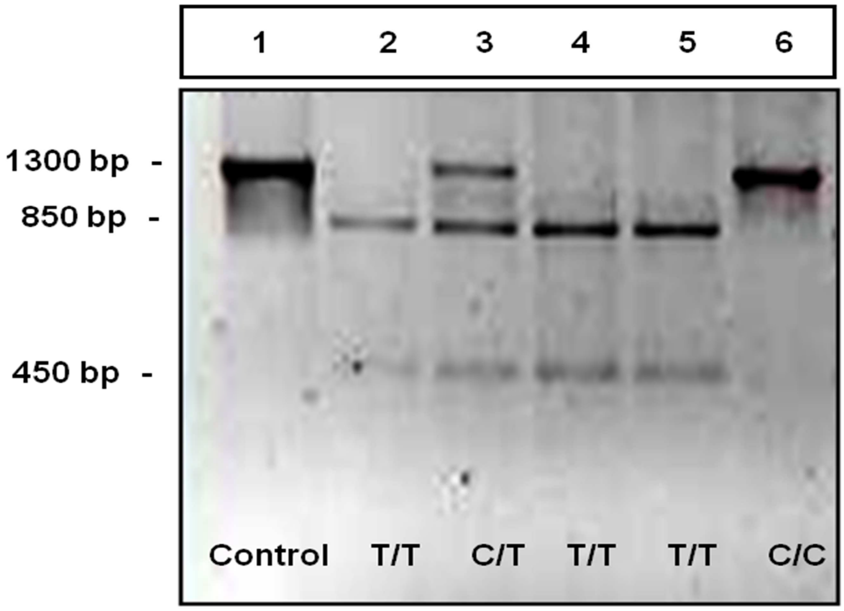


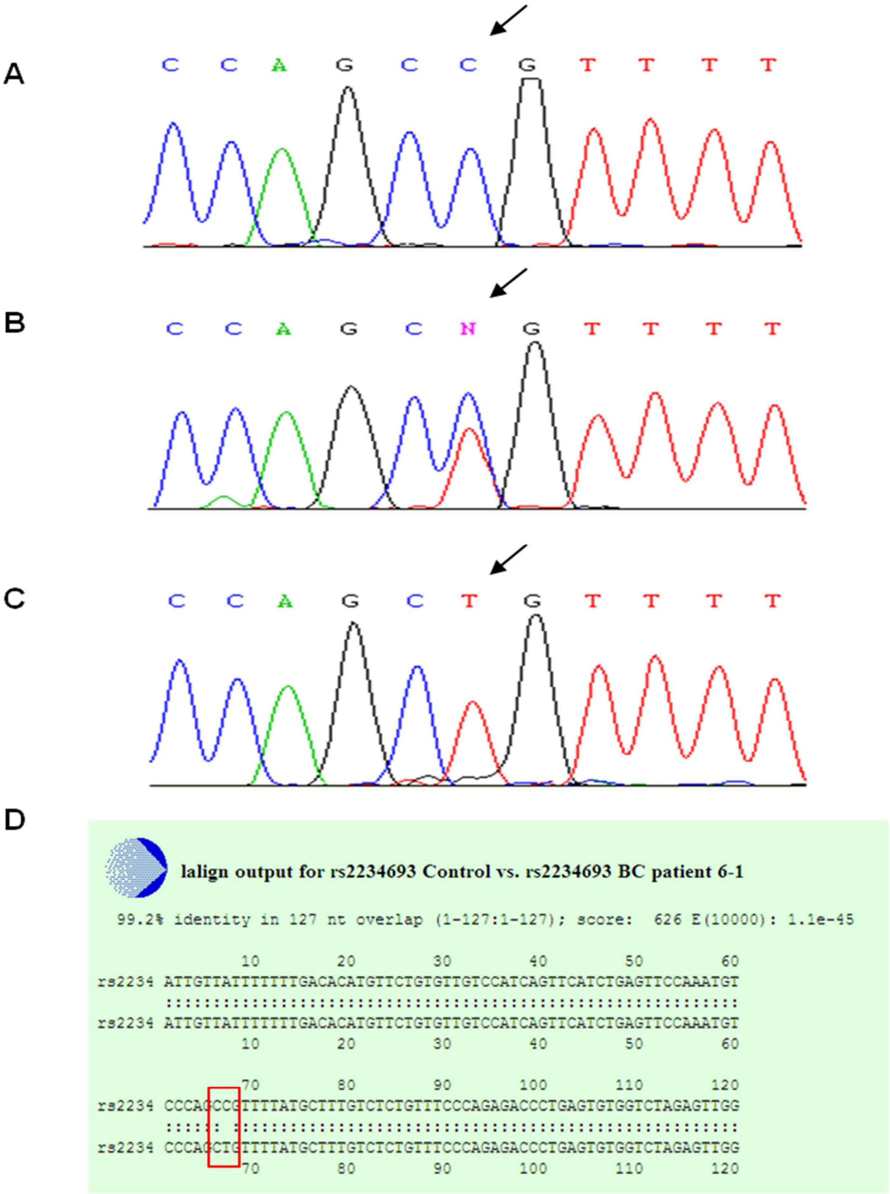
Figure 4.2 RFLP analysis of the 1300 bp amplification product of the rs2234693 (*PvuII*) SNP. Lane 1 - undigested PCR product, lane 2 - Case 2–3 (T/T), lane 3 - BC patient 6–1 (C/T), lane 4 - BC patient 5–1(T/T), lane 5 - Control 5–2 (T/T) and lane 6 - Case 5–3 (C/C). Fragment sizes are as indicated.

1 with the control for rs2234693 in *ESR1* indicated the mismatch caused by the presence of the SNP (Fig. 4.3 D).

4.3.2.1 Allele and genotype frequencies of rs2234693 (*PvuII*) in *ESR1*

The genotype and allele frequencies for rs2234693 (*PvuII*) are presented in Table 4.2. For Comparison 1, thirteen (21.67%) Controls were homozygous for the ancestral C allele, compared to only six (10.00%) of the Carriers. The majority of the 120 participants were heterozygous (C/T) for the SNP (46.67% of the Controls compared to 48.33% of the Carriers). Homozygosity for the variant T allele (T/T) was observed more frequently amongst the *BRCA2* mutation carriers (31.67% of the Controls, compared to 41.67% of the Carriers). The combined genotype (C/T and T/T) frequencies indicated that the variant T allele was more common in the mutation carriers (78.33% of the Controls and 90.00% of the Carriers). This proved to be almost statistical significant (P<0.05) with a *P*-value of 0.0896. The allele frequencies indicated that the majority of Controls (55.00%) and Carriers (65.83%) were positive for the variant T allele.

The data for the mutation carriers were examined more closely in Comparison 2. No distinct differences were observed in both the genotype and allele frequencies for the BC patients compared to the unaffected carriers (Table 4.2). This was substantiated by the high *P*-values of 0.8810 and 1.0, respectively. The allelic distributions for the Control, Cases, *BRCA2* mutation carriers and BC patient group did not deviate from HWE (Table 4.3).



lalign output for rs2234693 Control vs. rs2234693 BC patient 6-1

```
 99.2% identity in 127 nt overlap (1-127:1-127); score:  626 E(10000): 1.1e-45

               10        20        30        40        50        60
rs2234 ATTGTTATTTTTTTGACACATGTTCTGTGTTGTCCATCAGTTCATCTGAGTTCCAAATGT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
rs2234 ATTGTTATTTTTTTGACACATGTTCTGTGTTGTCCATCAGTTCATCTGAGTTCCAAATGT
               10        20        30        40        50        60

               70        80        90       100       110       120
rs2234 CCCAGCCGTTTTATGCTTTGTCTCTGTTTCCCAGAGACCCTGAGTGTGGTCTAGAGTTGG
       :::::: :::::::::::::::::::::::::::::::::::::::::::::::::::::
rs2234 CCCAGCTGTTTTATGCTTTGTCTCTGTTTCCCAGAGACCCTGAGTGTGGTCTAGAGTTGG
               70        80        90       100       110       120
```

Figure 4.3 Sequence analysis of the rs2234693 (*PvuII*) SNP in *ESR1*. **A** Sequencing results for Case 5–3, indicating homozygosity for the ancestral allele (C/C) as indicated by an arrow. **B** Sequence results for BC patient 6–1, indicating heterozygosity (C/T). **C** Sequence results for Case 2–3, indicating homozygosity for the variant T allele (T/T). **D** Alignment of the nucleotide sequence for BC patient 6–1 with the Fasta sequence of rs2234693 (*PvuII*). The nucleotide mismatch is highlighted by the red box.

Table 4.2 Allele and genotype distributions for rs2234693 (*PvuII*) and rs9340799 (*XbaI*) in *ESR1*.

Genes	Frequency distribution	Comparison 1 Controls vs *BRCA2* mutation carriers			Comparison 2 *BRCA2* BC patients vs Unaffected cases		
		Controls n=60 n (%)	Carriers n=60 n (%)	OR (95% CI)	BC patients n=30 n (%)	Cases n=30 n (%)	OR (95% CI)
rs2234693 (PvuII)	**Allele (n/%)**						
	C	54 (45.00)	41 (34.17)	0.63 (0.38-1.07)	20 (33.33)	21 (35.00)	1.08 (0.51-2.29)
	T	66 (55.00)	79 (65.83)	1.59 (0.93-2.63)	40 (66.67)	39 (65.00)	0.92 (0.44-1.96)
	Genotype (n/%)						
	C/C	13 (21.67)	6 (10.00)	0.42 (0.12-1.27)	3 (10.00)	3 (10.00)	1.00 (0.13-7.47)
	C/T	28 (46.67)	29 (48.33)	1.09 (0.44-2.73)	14 (46.67)	15 (50.00)	1.20 (0.31-4.97)
	T/T	19 (31.67)	25 (41.67)	1.60 (0.68-3.94)	13 (43.33)	12 (40.00)	0.83 (0.20-3.28)
		P-value = 0.3712			*P*-value = 0.8810		
	C/T+T/T	47 (78.33)	54 (90.00)	2.40 (0.79-8.70)	27 (90.00)	27 (90.00)	1.00 (0.13-7.47)
		P-value = 0.0896			*P*-value = 1.0000		
rs9340799 (XbaI)	**Allele (n/%)**						
	A	76 (63.33)	83 (69.17)	1.30 (0.76-2.22)	40 (66.67)	43 (71.67)	1.26 (0.58-2.75)
	G	44 (36.67)	37 (30.83)	0.77 (0.45-1.31)	20 (33.33)	17 (28.33)	0.79 (0.36-1.72)
	Genotype (n/%)						
	A/A	26 (43.33)	29 (48.33)	1.23 (0.56-2.78)	14 (46.67)	15 (50.00)	1.20 (0.31-4.97)
	A/G	24 (40.00)	25 (41.67)	1.08 (0.47-2.49)	12 (40.00)	13 (43.33)	1.25 (0.27-6.30)
	G/G	10 (16.67)	6 (10.00)	0.60 (0.18-1.82)	4 (13.33)	2 (6.67)	0.50 (0.05-3.49)
		P-value = 0.3734			*P*-value = 0.5433		
	A/G+G/G	34 (56.67)	31 (51.67)	0.81 (0.36-1.80)	16 (53.33)	15 (50.00)	0.83 (0.20-3.28)
		P-value = 0.5775			*P*-value = 0.7630		

Table 4.3 Exact tests of Hardy–Weinberg equilibrium (HWE) for *ESR1*, *TNRC9*, *LSP1*, *MAP3K1* and *FGFR2* for each of the groups studied stratified by age. Indicated are the respective *P*-values for each group.

rs number	Controls	*BRCA2* Cases	*BRCA2* mutation carriers	*BRCA2* BC patients
rs2234693 (PvuII)	0.79471	0.70826	0.77593	1.00000
rs9340799 (XbaI)	0.27582	1.00000	1.00000	0.6818
rs3803662 (TNRC9)	0.00451	0.10019	0.01494	0.09976
rs3817198 (LSP1)	0.57429	0.13346	0.17158	1.00000
rs889312 (MAP3K1)	0.00328	0.03199	0.01063	0.15836
rs2981582 (FGFR2)	1.00000	1.00000	0.09056	0.02136

4.3.3 Analysis of rs9340799 (*XbaI*) in *ESR1*

This SNP (A>G) was analyzed using the same methods as described in 4.2.3 (Cai *et al*., 2003; Hsieh *et al.*, 2007). PCR amplification yielded a 1300 bp fragment (Fig. 4.4). RFLP analysis of the homozygous ancestral A allele produced two fragments of 900 and 400 bp respectively (lanes 2, 4 and 5). The homozygous variant (G/G) allele produced a single fragment of 1300 bp (lane 6). A heterozygote displayed three fragments of 1300, 900 and 400 bp respectively (lane 3). DNA sequencing confirmed the genotyping of case 2–3 as homozygous ancestral (A/A), BC patient 6–1 as heterozygous (A/G) and case 5–3 as homozygous variant (G/G) (Fig. 4.5). Alignment of the sequence for BC patient 6–1 with the ancestral sequence of *ESR1* indicated the mismatch caused by the SNP (Fig. 4.5 D).

4.3.3.1 Allele and genotype frequencies of rs9340799 (*XbaI*) in *ESR1*

The genotype and allele frequencies for rs9340799 (*XbaI*) are indicated in Table 4.2. The genotype frequencies did not deliver any major differences between the Controls and Carriers for a similar genotype distribution was observed between the homozygous ancestral (A/A) (43.33% for the Controls versus 48.33% for the Carriers) participants and the heterozygous (A/G) (40.00% for the Controls versus 41.67% for the Carriers) participants with only a slight difference between the homozygous variant (G/G) individuals. The combined (A/G and G/G) genotype distribution for the presence of the variant G allele proved to be similar (56.67% for the Controls versus 51.67% for the Carriers). The *P*-value proved to be statistical insignificant (*P* = 0.5775). The majority of Controls (63.33%) and Carriers (69.17%) in Comparison 1 were positive for the ancestral A allele.

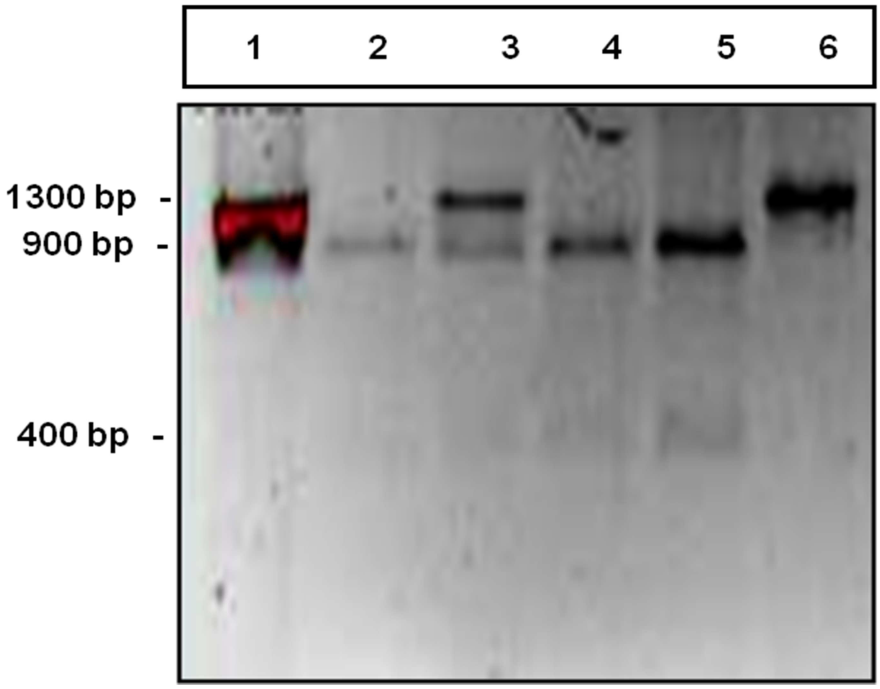
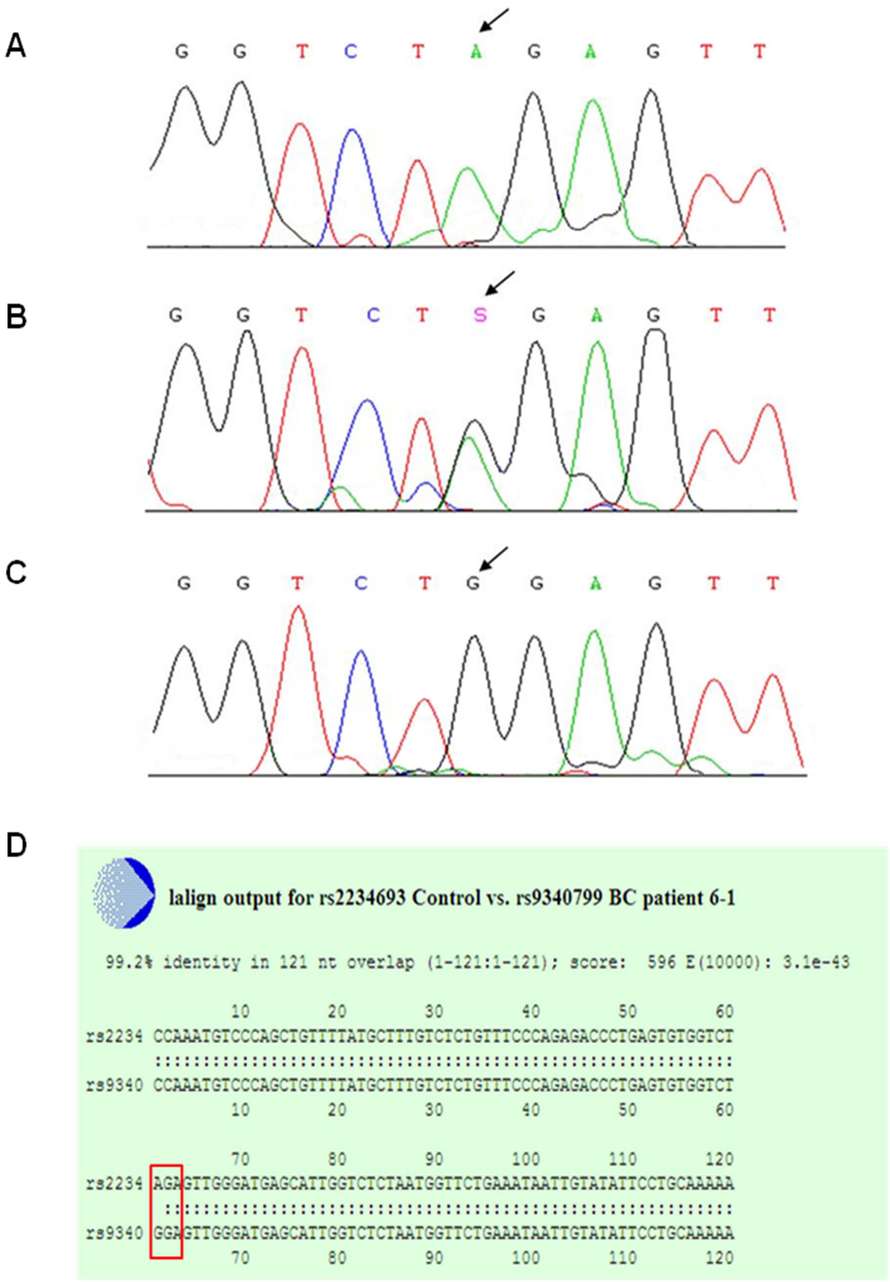


Figure 4.4 RFLP analysis of the 1300 bp amplification product for the rs9340799 (*XbaI*) SNP. Lane 1 - undigested PCR product, lane 2 - Case 2–3 (A/A), lane 3 - BC patient 6–1 (A/G), lane 4 - BC patient 5–1 (A/A), lane 5 - Control 5–2 (A/A), lane 6 - Case 5–3 (G/G). Fragment sizes are indicated.



lalign output for rs2234693 Control vs. rs9340799 BC patient 6-1

```
 99.2% identity in 121 nt overlap (1-121:1-121); score:  596 E(10000): 3.1e-43

               10        20        30        40        50        60
rs2234 CCAAATGTCCCAGCTGTTTTATGCTTTGTCTCTGTTTCCCAGAGACCCTGAGTGTGGTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
rs9340 CCAAATGTCCCAGCTGTTTTATGCTTTGTCTCTGTTTCCCAGAGACCCTGAGTGTGGTCT
               10        20        30        40        50        60

               70        80        90       100       110       120
rs2234 AGAGTTGGGATGAGCATTGGTCTCTAATGGTTCTGAAATAATTGTATATTCCTGCAAAAA
        :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
rs9340 GGAGTTGGGATGAGCATTGGTCTCTAATGGTTCTGAAATAATTGTATATTCCTGCAAAAA
               70        80        90       100       110       120
```

Figure 4.5 Sequencing analysis of the rs9340799 (*XbaI*) SNP in *ESR1*. **A** Sequencing results for Case 2–3, indicating homozygosity for the ancestral A allele (A/A) as indicted by the arrow. **B** Sequencing results for BC patient 6–1, indicating heterozygosity (A/G). **C** Sequencing results for Case 5–3, indicating homozygosity for the variant G allele (G/G). **D** Nucleotide alignment of the obtained sequence for BC patient 6–1 compared to the Fasta sequence of rs9340799 (*XbaI*). The nucleotide mismatch is highlighted by the red box.

Only slight differences were observed in Comparison 2. The genotype frequencies in Comparison 2 were similar for the homozygous ancestral (A/A) (46.67% BC patients versus 50.00% Cases) as well as for the heterozygote (A/G) (40.00% BC patients and 43.33% Cases) participants. The only noticeable difference was the observed homozygous variant (G/G) participants (13.33% for the BC patients versus 6.67% for the Cases). The combined (A/G and G/G) genotype frequencies were also similar between the BC patients (53.33%) and the Cases (50.00%). The *P*-value proved to be statistical insignificant (*P* = 0.7630). The allele frequencies were similar between the BC patients (66.67%) and Cases (71.67%). The allelic distributions for the Control, Cases, *BRCA2* mutation carriers and BC patient group did not deviate from the HWE (Table 4.3).

4.3.4 Construction and analysis of an *ESR1* haplotype

It is important to calculate linkage between the particular SNPs of interest to establish whether the SNPs have a cumulative risk according to the International Haplotype Map (HapMap) Consortium (2003). The alleles of each locus are designated as 0 for the ancestral allele and 1 for the variant allele in either the hetero- or homozygous state. The order of the polymorphisms was as follows: first rs2234693 (*PvuII*) then rs9340799 (*XbaI*). The results are indicated in Table 4.4. A haplotype of 00 was observed in both Comparison 1 and Comparison 2, indicating the presence of homozygosity for the ancestral alleles in rs2234693 and rs9340799 with a percentage of 35.00% for the Controls and 41.67% for the Carriers in Comparison 1 and 43.33% for the BC patients compared to 40.00% for the Cases in Comparison 2. The haplotype 11 representing the presence of only the variant alleles were not observed for both Comparison 1 and 2.

In Comparison 1, a frequency difference was observed between the Controls and Carriers for the 10 haplotype (21.67% and 10.00% for the control and mutation carriers respectively), however it proved to be statistically insignificant (*P* = 0.3674).

Table 4.4 Haplotype frequencies of rs2234693 (*PvuII*) and rs9340799 (*XbaI*) in *ESR1*.

Haplotypes	Comparison 1 Controls vs *BRCA2* mutation carriers			Comparison 2 *BRCA2* BC patients vs Unaffected cases		
	Controls n=60 n (%)	Carriers n=60 n (%)	OR (95% CI)	BC patients n=30 n (%)	Cases n=30 n (%)	OR (95% CI)
00	21 (35.00)	25 (41.67)	1.40 (0.58-3.52)	13 (43.33)	12 (40.00)	0.83 (0.20-3.28)
01	26 (43.33)	29 (48.33)	1.23 (0.56-2.78)	14 (46.67)	15 (50.00)	1.20 (0.31-4.97)
10	13 (21.67)	6 (10.00)	0.42 (0.12-1.27)	3 (10.00)	3 (10.00)	1.00 (0.13-7.47)
	P-value = 0.3674			*P*-value = 1.0000		

In Comparison 2, the majority of BC patients and Cases exhibited a haplotype of 00 and 01 with almost equal frequency distribution between them. The 10 haplotype was equally present in both the BC patients (10.00%) and the Cases (10.00%) with no statistical significance (*P* = 1.0000).

4.3.5 Analysis of four selected SNPs in the *TNRC9*, *LSP1*, *MAP3K1* and *FGFR2* genes

Single nucleotide polymorphisms occurring within four selected genes (*TNRC9*, *LSP1*, *MAP3K1* and *FGFR2*) were optimized and screened for as described (Chapter 3) using a TaqMan® SNP genotyping assay according to Easton *et al*. (2007). The amplification product of each individual was used to determine its potential effect on BC risk in *BRCA2* mutation positive Afrikaner women. As described in Chapter 3, genotype assignment was done using the BioRad CFX Manager V1.1.308.1111 software.

4.3.5.1 Allele and genotype frequencies for rs3803662 in *TNRC9*

The genotype and allele frequencies for rs3803662 in *TNRC9* are presented in Table 4.5. No major differences were observed for both the allele and genotype frequencies between the Controls and the Carriers. The majority of participants in Comparison 1 proved to be homozygous for the ancestral C allele (C/C) (86.67% of the Controls and 83.33% of the *BRCA2* carriers). The heterozygous (C/T) and homozygous variants (T/T) were extremely rare with only 12 participants being heterozygous (C/T) (8.33% Controls versus 11.67% Carriers) and 6 being homozygous for the variant (T/T) (5.00% Controls versus 5.00% Carriers). The combined hetero- and homozygous (C/T and T/T) frequencies were very similar indicating statistical insignificance (*P* = 0.5930).

No distinct differences were observed for Comparison 2 (Table 4.5). The majority of BC patients (93.33%) and Cases (85.00%) carried the ancestral C allele. Homozygosity for the ancestral C allele (C/C) was observed in 90.00% of the BC patients compared to 76.67% for the Cases. Heterozygosity was observed in 6.67% of the BC patients compared to the 16.67% of the Cases, whereas homozygosity for the variant T allele was observed in 3.33% of the BC patients compared to 6.67% of the Cases. The combined genotype (C/T and T/T) frequencies between the BC patients (10.00%) and Cases (23.33%) was shown to be statistically insignificant (*P* = 0.1573). The Controls and *BRCA2* mutation carriers deviated from HWE with *P*-values of 0.00451 and 0.01494 respectively however the Cases and BC patient group was in HWE (Table 4.3).

4.3.5.2 Allele and genotype frequencies for rs3817198 in *LSP1*

The genotype and allele frequencies for this SNP are presented in Table 4.5. This SNP was more common within the Afrikaner population than rs3803662 in *TNRC9*, as a lower majority of participants (65.00% of the Controls versus 62.50% of the Carriers) carried the ancestral T allele. Homozygosity for the ancestral T allele (T/T) was observed in 40.00% of the Controls and 43.33% of the Carriers. Fifty percent of the Controls were heterozygous (T/C) compared to only 38.33% of the Carriers. Homozygosity for the variant C allele (C/C) was observed in only 10.00% of the Controls compared to 18.33% of the Carriers. A combined (C/T and C/C) variant genotype frequency of 60.00% was observed in the Controls compared to 56.67% of the Carriers. Although differences were observed, it proved to be statistically insignificant (*P* = 0.7150).

The ancestral T allele in Comparison 2 was the most abundant in both BC patients (68.33%) and Cases (56.67%) (Table 4.5). The genotype frequencies indicated a

Table 4.5 Allele and genotype frequencies of selected polymorphisms in the *TNRC9*, *LSP1*, *MAP3K1*, *FGFR2* genes.

Genes	Frequency distribution	Comparison 1 Control vs *BRCA2* mutation carriers			Comparison 2 *BRCA2* BC patients vs Unaffected cases		
		Controls n=60 n (%)	Carriers n=60 n (%)	OR (95% CI)	BC patients n=30 n (%)	Cases n=30 n (%)	OR (95% CI)
TNRC9	Allele (n/%)						
	C	109 (90.83)	107 (89.17)	0.83 (0.36-1.94)	56 (93.33)	51 (85.00)	0.41 (0.12-1.40)
	T	11 (9.17)	13 (10.83)	1.20 (0.51-2.78)	4 (6.67)	9 (15.00)	2.44 (0.71-8.33)
	Genotype (n/%)						
	C/C	52 (86.67)	50 (83.33)	0.75 (0.21-2.47)	27 (90.00)	23 (76.67)	0.33 (0.03-1.86)
	C/T	5 (8.33)	7 (11.67)	1.67 (0.32-10.73)	2 (6.67)	5 (16.67)	2.50 (0.41-26.25)
	T/T	3 (5.00)	3 (5.00)	1.00 (0.13-7.47)	1 (3.33)	2 (6.67)	2.00 (0.10-118.0)
		P-value = 0.9189			*P*-value = 0.5153		
	C/T + T/T	8 (13.33)	10 (16.67)	1.33 (0.41-4.66)	3 (10.00)	7 (23.33)	3.00 (0.54-30.39)
		P-value = 0.5930			*P*-value = 0.1573		
LSP1	Allele (n/%)						
	T	78 (65.00)	75 (62.50)	0.90 (0.53-1.52)	41 (68.33)	34 (56.67)	0.61 (0.29-1.28)
	C	42 (35.00)	45 (37.50)	1.11 (0.66-1.89)	19 (31.67)	26 (43.33)	1.64 (0.78-3.45)
	Genotype (n/%)						
	T/T	24 (40.00)	26 (43.33)	1.14 (0.52-2.53)	14 (46.67)	12 (40.00)	0.79 (0.25-2.35)
	T/C	30 (50.00)	23 (38.33)	0.61 (0.26-1.37)	13 (43.33)	10 (33.33)	0.70 (0.23-2.04)
	C/C	6 (10.00)	11 (18.33)	2.00 (0.62-7.46)	3 (10.00)	8 (26.67)	3.50 (0.67-34.53)
		P-value = 0.2863			*P*-value = 0.4235		
	T/C + C/C	36 (60.00)	34 (56.67)	0.88 (0.40-1.91)	16 (53.33)	18 (60.00)	1.29 (0.43-4.06)
		P-value = 0.7150			*P*-value = 0.6171		
MAP3K1	Allele (n/%)						
	A	97 (80.83)	95 (79.17)	0.90 (0.48-1.70)	46 (76.67)	49 (81.67)	1.36 (0.56-3.29)
	C	23 (19.17)	25 (20.83)	1.11 (0.59-2.08)	14 (23.33)	11 (18.33)	0.73 (0.30-1.78)
	Genotype (n/%)						
	A/A	43 (71.67)	41 (68.33)	0.78 (0.25-2.35)	19 (63.33)	22 (73.33)	2.00 (0.43-12.36)
	A/C	11 (18.33)	13 (21.67)	1.25 (0.44-3.65)	8 (26.67)	5 (16.67)	0.57 (0.12-2.25)
	C/C	6 (10.00)	6 (10.00)	1.00 (0.13-7.47)	3 (10.00)	3 (10.00)	1.00 (0.13-7.47)
		P-value = 0.3916			*P*-value = 0.3701		
	A/C+C/C	17 (28.33)	19 (31.67)	1.29 (0.43-4.06)	11 (36.67)	8 (26.67)	0.50 (0.08-2.34)
		P-value = 0.6171			*P*-value = 0.3173		
FGFR2	Allele (n/%)						
	C	83 (69.17)	77 (64.17)	0.80 (0.47-1.37)	38 (63.33)	39 (65.00)	1.08 (0.51-2.27)
	T	37 (30.83)	43 (35.83)	1.25 (0.73-2.13)	22 (36.67)	21 (35.00)	0.92 (0.44-1.96)
	Genotype (n/%)						
	C/C	29 (48.33)	28 (46.67)	0.94 (0.43-2.03)	15 (50.00)	13 (43.33)	0.78 (0.25-2.35)
	C/T	25 (41.67)	21 (35.00)	0.77 (0.34-1.67)	8 (26.67)	13 (43.33)	2.00 (0.62-7.46)
	T/T	6 (10.00)	11 (18.33)	2.00 (0.62-7.46)	7 (23.33)	4 (13.33)	0.50 (0.08-2.34)
		P-value = 0.3062			*P*-value = 0.5686		
	C/T + T/T	31 (51.67)	32 (53.33)	1.07 (0.49-2.32)	15 (50.00)	17 (56.67)	1.29 (0.43-4.06)
		P-value = 0.8575			*P*-value = 0.6171		

similar distribution for the homozygous (T/T) (46.67% of the BC patients versus 40.00% of the Cases) and heterozygous participants (T/C) (43.33% of the BC patients and 33.33% of the Cases). A difference was observed for the homozygous variant genotype (C/C) with 10.00% of the BC patients compared to 26.67% of the Cases. The combined (T/C and C/C) genotype frequency for the variant allele was 53.33% for the BC patients compared to 60.00% for the Carriers. This SNP did not prove to be significant within the sample population (*P* = 0.6171). The allelic distribution for the Control, Cases, *BRCA2* mutation carriers and BC patient group agreed with the HWE (Table 4.3).

4.3.5.3 Allele and genotype frequencies for rs889312 in *MAP3K1*

The allele and genotype frequencies for rs889312 in *MAP3K1* are presented in Table 4.5. In Comparison 1, the majority of Controls (80.83%) and Carriers (79.17%) were positive for the ancestral A allele. The most abundant genotype within Comparison 1 was homozygosity for the ancestral A allele with 71.67% for the Controls and 68.33% for the Carriers. Heterozygosity (A/C) was observed in 18.33% of the Controls compared to 21.67% of the Carriers. Homozygosity for the variant C allele (C/C) was identical for the Carriers (10.00%) and Controls (10.00%). No major difference was observed between the Controls (28.33%) and Carriers (31.67%) regarding the combined (A/C and C/C) genotype frequencies. No statistical significance was reflected in the high *P*-value (*P* = 0.6171).

In Comparison 2, the ancestral A allele was observed for the majority of participants (76.67% in the BC patients versus 81.67% in the Cases). Similar genotype distributions were observed for the homozygous ancestral (A/A) (63.33% BC patients versus 73.33% Cases) participants. Heterozygosity was observed in 26.67% of the BC patients compared to 16.67% of the Cases. An identical genotype distribution was observed for the homozygous variant participants

(10.00% for the BC patients versus 10.00% for the Cases). Although the combined (A/C and C/C) genotype frequency between the BC patients (36.67%) and the Cases (26.67%) was different, it proved to be statistically insignificant ($P = 0.3173$). Deviation from the HWE was observed for the Controls ($P = 0.00328$), Cases ($P = 0.03199$) and *BRCA2* mutation carriers ($P = 0.01063$) whereas the BC patient group was in HWE (Table 4.3).

4.3.5.4 Allele and genotype frequencies for rs2981582 in *FGFR2*

The genotype and allele frequencies for the rs2981582 SNP in *FGFR2* are presented in Table 4.5. The allele frequency revealed that the ancestral C allele was the most abundant and the variant T allele present in only 30.83% of Controls and 35.83% of the Carriers in Comparison 1. The genotype frequencies revealed that the majority of Controls (48.33%) and Carriers (46.67%) were homozygous for the ancestral C allele. Heterozygosity for the variant allele (C/T) was slightly different between the Controls (41.67%) and Carriers (35.00%). A small difference was observed between the homozygous variant Controls (10.00%) and Carriers (18.33%). The combined genotype frequencies (C/T and T/T) were very similar between the Controls (51.67%) and the Carriers (53.33%). The *P*-value proved to be statistical insignificant ($P = 0.8575$).

Similar allele and genotype distributions were observed in Comparison 2. The allele distribution indicated that the majority of BC patients (63.33%) and Cases (65.00%) were positive for the ancestral C allele. Homozygosity for the ancestral C allele was the most abundant genotype for both BC patients (50.00%) and Cases (43.33%). A difference in heterozygous frequency was observed between the BC patients (26.67%) and the Cases (43.33%). Homozygosity for the variant T allele also revealed a difference (23.33% BC patient versus 13.33% Cases). The combined genotype frequencies (C/T and T/T) were almost similar between the Controls (50.00%) and the Carriers (56.67%). The *P*-value proved to be statistical insignificant (*P* = 0.6171). Only the *BRCA2* mutation carriers deviated from the HWE with a *P*-value of 0.09056 (Table 4.3).

4.3.6 Analysis of cumulative risk on BC by compiling a multi-locus haplotype for four polymorphisms

The four polymorphisms located in *TNRC9*, *LSP1*, *MAP3K1* and *FGFR2* respectively were used in a multi-locus haplotype to study their cumulative risk on BC. The alleles of each locus are designated as 0 for the ancestral allele and 1 for the variant allele in either the hetero- or homozygous state. The order of the polymorphisms is as follow: rs3803662 in *TNRC9*, rs3817198 in *LSP1*, rs889312 in *MAP3K1* and rs2981582 in *FGFR2*. The data are presented in Table 4.6.

The multi-locus haplotype of Comparison 1 revealed that 3.33% of the Controls and 1.67% of the *BRCA*2 mutation positive individuals were homozygous for all the ancestral alleles (haplotype 0000). The majority of Controls and *BRCA2* mutation carriers exhibited the haplotypes 1010 (18.33% and 15.00%), 1011 (18.33% and 10.00%) and 1111 (18.33% and 16.67%). Minor differences observed between the Controls and Carriers included the haplotypes 0010 (1.67% versus 5.00%), 1001 (3.33% versus 6.67%), 1100 (6.67% versus 3.33%) and 1110 (6.67% versus 13.33%). No statistical significance was found (*P* = 0.8735).

No significant differences were seen in Comparison 2. The haplotype representing all the ancestral alleles (0000) was only seen in one unaffected case. Some minor differences between the BC patients and unaffected Cases were

Table 4.6 Haplotype analysis of SNPs rs3803662 in *TNRC9*, rs3817198 in *LSP1*, rs889312 in *MAP3K1* and rs2981582 in *FGFR2*.

Haplotypes	Comparison 1 Controls vs *BRCA2* positive individuals		Comparison 2 *BRCA2* BC patients vs Unaffected cases	
	Controls n=60 n (%)	**Carriers n=60 n (%)**	**BC patients n=30 n (%)**	**Cases n=30 n (%)**
0000	**2 (3.33)**	**1 (1.67)**	**0 (0.00)**	**1 (3.33)**
0010	**1 (1.67)**	**3 (5.00)**	**0 (0.00)**	**3 (10.00)**
0011	**2 (3.33)**	**3 (5.00)**	**2 (6.67)**	**1 (3.33)**
0101	**0 (0.00)**	**1 (1.67)**	**1 (3.33)**	**0 (0.00)**
0110	**2 (3.33)**	**1 (1.67)**	**0 (0.00)**	**1 (3.33)**
0111	**1 (1.67)**	**1 (1.67)**	**0 (0.00)**	**1 (3.33)**
1000	**7 (11.67)**	**8 (13.33)**	**4 (13.33)**	**4 (13.33)**
1001	**2 (3.33)**	**4 (6.67)**	**2 (6.67)**	**2 (6.67)**
1010	**11 (18.33)**	**9 (15.00)**	**5 (16.67)**	**4 (13.33)**
1011	**11 (18.33)**	**6 (10.00)**	**3 (10.00)**	**3 (10.00)**
1100	**4 (6.67)**	**2 (3.33)**	**2 (6.67)**	**0 (0.00)**
1101	**2 (3.33)**	**3 (5.00)**	**2 (6.67)**	**1 (3.33)**
1110	**4 (6.67)**	**8 (13.33)**	**4 (13.33)**	**4 (13.33)**
1111	**11 (18.33)**	**10 (16.67)**	**5 (16.67)**	**5 (16.67)**
	***P*-value = 0.8735**		***P*-value = 0.7120**	

seen in the haplotypes 0010 (0.00% and 10.00%) and 0011 (6.67% and 3.33%). The minor differences seen proved to be statistically insignificant ($P = 0.7120$).

4.4 Discussion

Breast cancer is a complex disease for a combination of genetic and environmental factors influences penetrance of disease-causing familial BC mutations. For the two familial breast cancer genes, the variation in penetrance is most striking for *BRCA2* mutation carriers (Tryggvadottir *et al*., 2006; Begg *et al*., 2008), for women with the same mutation may develop breast, ovarian or other cancers at different ages or not at all (Offit, 2006). As this affects the management and genetic counseling of family members, studies on polymorphisms within other genes that may influence the penetrance of these mutations within the white Afrikaner population was launched in 2006 within the department of Human Genetics at the Free State University.

Initially, the studies focused on the identification of potential genetic modifiers of BC risk in the Afrikaner founder mutation *BRCA2* 8162delG (c.7934del, p.Arg2645AsnfsX3) mutation carriers, as it was the most frequent BC associated mutation detected within this population group. Patients from several mutation positive families used during the familial BC screening phase were thus included in the current study. Thirty groups consisting of four individuals each were compiled. The number of groups was limited by the number of mutation positive affected and unaffected carriers who could be contacted to obtain informed consent for the additional study. While the sample size of this pilot study was in fact too small to confirm true modifiers of cancer risk within this population, the data might indicate possible trends.

4.4.1 Genetic modifiers of breast cancer risk in *ESR1*

The world-wide search for genetic modifiers within low-penetrance genes associated with disease-causing mutations within the *BRCA* genes started with genes known to be involved in carcinogenesis (the DNA repair pathway) such as *RAD51* (OMIN 179617) (Wang *et al.*, 2001). The evolution of high throughput genetic techniques and technology however resulted in an explosion of GWAS such as those by Cox *et al.* (2007), Easton *et al.* (2007), Hunter *et al.* (2007), Stacey *et al.* (2007), Antonio *et al.* (2008), Garcia-Closas *et al.* (2008), , Ahmed *et al.* (2009), Antoniou *et al.* (2009), Hemminki *et al.* (2010), Mcinerney *et al.* (2009), Thomas *et al.* (2009) and Zheng *et al.* (2009).

The modifying effect of 11 SNPs specifically within *BRCA2* mutation carriers have since been confirmed (Antonio *et al.*, 2008). Included are polymorphisms within the *FGFR2*, *TNRC9*, *MAP3K* and *LSP1* genes which will be discussed separately from rs2234693 (*PvuII*) and rs9340799 (*XbaI*) present within the *ESR1* gene. The rationale for this is the fact that the latter two SNPs have not yet been validated by GWAS. Their influence on BC risk within the Afrikaner therefore remained unclear.

The link between estrogen and BC has long been investigated as estrogen plays a key role in many reproductive factors such as age at menarche, age at first pregnancy, number of pregnancies and many others (Jernström *et al*., 1999; Andrieu *et al.*, 2006). The estrogen receptors regulate the expression of proteins involved in the development and proliferation of breast tissue. The two polymorphisms in *ESR1* were selected based on the fact that estrogen is an important epidemiologic risk factor and its effects are mediated through the ER in breast tissue (Heldring *et al*., 2007). These two SNPs are further more important in many other diseases such as endometrial cancer risk (Wedrén *et al.*, 2008), tendency to obesity (Nilsson *et al*., 2007), late-life depression (Ryan *et al*., 2011) and cardiovascular disease (Casazza *et al*., 2010).

The variant T allele for the rs2234693 (*PvuII*) SNP was common within the Afrikaner population, as the majority of participants (46.67% of the Controls and 48.33% in Carriers) were heterozygous (Table 4.2, Comparison 1). The variant allele frequency for the two groups was 0.55 and 0.65 respectively. The variant allele was therefore detected within the majority of participants in either a heterozygous or homozygous state (Table 4.2). These findings are in accordance with the data from the International HapMap project as published on the National Center for Biotechnology Information (NCBI) website, where the variant T allele was also detected within the majority of Europeans (http://www.ncbi.nlm.nih.gov).

The results however showed no statistical significance between the genotype frequencies for the Controls and Carriers (Table 4.2, Comparison 1). A possible trend towards statistical significance was observed between the combined minor allele frequencies for the Controls and the Carriers in Comparison 1 (*P*-value = 0.0896). Based on this cut-off ($P < 0.05$), we propose that this association should be explored further in a larger Afrikaner study group. This trend was unfortunately not reflected for Comparison 2, as the genotype and combined minor allele frequencies for the BC patients and Cases indicated no differences (Table 4.2). The data of the Afrikaner is supported by HapMap for white Europeans, which indicates genotype frequencies of 0.336 for homozygous and 0.513 for heterozygous individuals carrying the variant T allele (http://www.ncbi.nlm.nih.gov).

The presence of the variant for rs2234693 (*PvuII*) has been associated with BC risk in other populations, although in smaller studies. Not only did Cai *et al*. (2003) report an association between rs2234693 (*PvuII*) and BC in Shanghai women, he also reported an association between a younger age of BC diagnosis in patients with the variant T allele however did not specify the mean age. Yaich *et al*. (1992) reported that BC patients homozygous for the variant T allele was significantly younger (mean age of 50.4) compared to those homozygous (mean age of 64.6) and heterozygous (mean age of 64.4) for the ancestral C allele. This was not observed in the current study as the mean age of the homozygous variant (T/T) participants were 44.6 years compared to 40.5 years for the homozygous ancestral (C/C) and 45.2 years for the heterozygotic (C/T) participants. The current data agrees with a Norwegian and Swedish case-control study which failed to confirm an association between rs2234693 (*PvuII*) and an increased BC risk (Andersen *et al*., 1994; Weiderpass *et al*., 2000).

The ancestral A allele of rs9340799 (*XbaI*) was common within the Afrikaner population (43.33% for the Controls and 48.33% for the Carriers) (Table 4.2, Comparison 1). The minor allele frequency was 0.36 for the Controls compared to 0.30 for the Carriers. Heterozygosity for the variant allele was observed in 40.00% of the Controls and 41.67% of the Carriers (Table 4.2). These results are in accordance with the findings reported for white Europeans which indicated that the majority of participants are homozygous (0.481) or heterozygous (0.426) for the ancestral A allele (http://www.ncbi.nlm.nih.gov). No significant association was observed for Comparison 1.

Comparison 2 also did not reveal significant differences between the BC patients and the unaffected Cases (*P* = >0.05, Table 4.2). These results are consistent with a large Swedish BC case/control study which failed to confirm any association between the presence of the variant allele and an increased BC risk (Weiderpass *et al*., 2000). However, several other studies did indicate a significant association. Andersen *et al*. (1994) and a Norwegian study (Weiderpass *et al*., 2000) suggested that the ancestral A allele is the risk allele and that it was positively associated with BC risk. This was corroborated by Wang *et al*. (2007) who reported an allelic protective effect for rs9340799 in Caucasians whereas Shin *et al*. (2003) reported a decrease of BC risk associated with the variant G allele for South Korea participants.

The contradictory results of the *ESR1* SNP could be due to ethnicity or the mean age of the study groups (Andersen *et al*., 1994; Cai *et al.*, 2003; Wang *et al.*, 2007). The mean age of the participants in the study by Andersen *et al.* (1994) was 56.5 years and between 50 and 74 in the Weiderpass *et al*. (2000) study. The age distribution of the current study ranged from 24 to 62 years with a mean age of 44.5. As age plays a role in risk factors such as age of menarche, age at menopause, surgical menopause, parity, body mass index and bone mineral density, it should be taken into account when compiling study groups and comparing results (Kelsey *et al.*, 1993; Hunter *et al*., 1997; Key *et al*., 2001; Welcsh *et al.*, 2001; Carpenter *et al*., 2003; Lahmann *et al*., 2004; Andrieu *et al*., 2006). Most studies also focused on sporadic BC cases and did not specify the SNPs studied with regards to BC (Andersen *et al*., 1994; Weiderpass *et al*., 2000; Cai *et al.*, 2003), whereas the current study focused only on Caucasian female familial *BRCA*2 8162delG (c.7934del, p.Arg2645AsnfsX3) mutation carriers affected and unaffected with BC. This factor could be a major contributor to the differences observed between the published results for the white European and that for the Afrikaner.

The two polymorphisms in the *ESR1* gene were incorporated within a haplotype to determine whether in combination, they might have an effect on the penetrance of the *BRCA2* disease-causing mutation. The majority of Controls (43.33%) and Carriers (48.33%) in Comparison 1 displayed a haplotype of 01 representing the ancestral allele for rs2234693 (*PvuII*) and the variant G allele for rs9340799 (*XbaI*) (Table 4.4). In Comparison 2, similar results were observed. The majority of BC patients (46.67%) and Cases (50.00%) also exhibited the 01 haplotype. In Comparison 1 a difference was observed between the Controls and the Carriers. The 10 haplotype was observed in only 10.00% of the Carriers compared to 21.67% of the Controls (Table 4.4). Interestingly, the haplotype 11 representing both variant alleles were not observed in Comparison 1 or 2. Unfortunately, the observations proved to be statistically insignificant for both Comparison 1 (*P* = 0.3674) and 2 (*P* = 1.0000) (Table 4.4). These findings could not be compared to studies on BC specifically and analysis of a haplotype with only rs2234693 (*PvuII*) and rs9340799

(*XbaI*). However, Goulart *et al*. (2009) reported on haplotypes composing of these SNPs and obesity. The 11 haplotype was observed at a frequency of 5% with the most common haplotype 10 showing a frequency of 52%. The 01 haplotype had a frequency of 33% which is considerably lower than that observed for the current study. Since the two SNPs within *ESR1* have not been confirmed by GWAS, it is assumed that they do not act as major contributors to modifying BC risk.

4.4.2 GWAS SNPs in the Afrikaner

The recent multistage GWAS identified 11 low penetrance loci that play a role in BC risk (Easton *et al*., 2007; Hunter *et al.,* 2007; Stacey *et al.*, 2007; Gaudet *et al*., 2010). These include SNPs in the *MAP3K1*, *LSP1* and *FGFR2* genes and a SNP in a region close to the *TNRC9* gene. Since *BRCA2* associated breast tumours differ pathologically from *BRCA1* tumours, it was expected that SNPs influencing *BRCA2* penetrance might differ from those affecting *BRCA1*. This was based on the fact that the majority of *BRCA2* associated breast tumours are ER positive, similar to the majority of sporadic BCs. This is in sharp contrast to *BRCA1*-related ER negative tumours. This was supported by Foulkes *et al.* (2003) who speculated that risk alleles for *BRCA1* could be markedly different from *BRCA2* related- and sporadic tumours due to their ER negativity. That was indeed the case as only SNPs in *FGFR2* (rs2981575), *TNRC9* (rs3803662), *MAP3K1* (rs889312) and *LSP1* (rs3817198) were shown to modify *BRCA2* penetrance specifically (Foulkes *et al.*, 2003; Campa *et al.*, 2011).

At the time the current study commenced, very little information regarding genetic modifiers of BC risk was available. In the five years since then, the situation changed considerably with multiple laboratories cooperating towards the goal of identifying genetic modifiers of BC risk in mutation carriers (CIMBA). This combined project has not only identified useful SNPs, but also validated and calculated the risk for each one.

The international success of GWAS outclassed the current study, as the international search for genetic modifiers is now nearly completed. This resulted in a shift in the focus of this part of the current study, by concentrating on allele frequencies for each SNP and testing of various hypotheses.

The specific function of the majority of the identified genes for which SNPs have been identified, still needs to be unraveled. It has been suggested that *TNRC9* may act as a transcription factor and was implicated in metastasis of the bone (Smid *et al.*, 2006), whereas O'Flaherty and Kaye, (2003) suggested a role in the bending and unwinding of DNA, thereby altering chromatin structure. The rs3803662 SNP in *TNRC9* has been implicated in BC risk by numerous investigators such as Easton *et al*. (2007), Antoniou *et al*. (2008) and Chen *et al*. (2010). This SNP is associated with an increased risk of BC in both *BRCA1* and *BRCA2* mutation carriers. The per-allele hazard ration (HR) was estimated to be 1.15 for *BRCA2* carriers (Antoniou *et al*., 2008).

This SNP was relatively rare within the Afrikaner population, as the majority of Controls and *BRCA2* mutation carriers were homozygous for the ancestral C allele (86.67% Controls and 83.33% Carriers). The variant T allele was observed in the heterozygous form in only 8.33% of the Controls and 11.67% of the Carriers with only six participants being homozygous for the variant allele (T/T).

The minor allele frequency was 0.09 for the Controls and 0.10 for the Carriers (Table 4.5, Comparison 1) which was considerably lower than the 0.25 and 0.26 for *BRCA2* unaffected and affected *BRCA2* mutation carriers observed for 3557 white European *BRCA2* mutation carriers genotyped by Antoniou *et al.* (2008). A similar frequency was reported by Latif *et al*. (2009), although their study population included both *BRCA1* (n = 120) and *BRCA2* (n = 107) mutation carriers. A Dutch study which included the Dutch hospital-based cohort of breast cancer patients (ORIGO) also reported a minor allele frequency of 0.26 (n = 1263), although the study was performed on BC patients only and not necessarily *BRCA* mutation carriers (Huijts *et al*., 2007).

LSP1 is an intracellular F-actin binding cytoskeletal protein that is expressed in B cells, functional T cells, thymocytes, monocytes, macrophages, neutrophils, lymphocytes and endothelium (Jongstra *et al*., 1994; Li *et al*., 1995; Liu *et al*., 2005; Petri *et al*., 2011). The GWAS of Easton *et al.* (2007) and Stacey *et al.* (2007) associated rs3817198 with an increased BC risk in thousands of BC patients. It was only afterwards that Antoniou *et al.* (2009) confirmed the tentative association within *BRCA2* mutation carriers specifically. Although no significant evidence of an association was observed for *BRCA1* mutation carriers, a multiplicative effect was observed for *BRCA2* in which each copy of the minor allele was estimated to confer a HR of 1.16 (95% CI: 1.07–1.25) (Antoniou *et al.*, 2009).

This SNP was common within the Afrikaner population as the majority of control participants were heterozygous for the variant C allele (50% compared to 38.33% for the *BRCA2* mutation carriers, Table 4.5). More *BRCA2* mutation carriers were homozygous for the variant C allele compared to the Controls (18.33% versus 10%). During a closer analysis of the data in Comparison 2 (Table 4.5), more of the unaffected Cases were homozygous for the risk allele, compared to the BC patients (26.67% versus 10.00%). This was contradictory to all GWAS results obtained for this SNP.

Before a hypothesis can be made for the Afrikaner population, it is imperative to remember that evidence of an additional SNP situated three bases downstream from rs3817198 (Chapter 3, Fig. 3.15) has been recorded. As the results of this study were based on the genotype calling of Method 1 (Chapter 3, Table 3.5), the genotype and allele frequencies presented in Table 4.5, might be skewed due to the presence of this unknown second SNP within *LSP1*. In order to determine more accurate genotype and allele frequencies for the Afrikaner population, new probes have to be designed to detect the new SNP specifically to determine its frequency within the same population. Once the dataset for both SNPs can be successfully separated, the true genotype and allele frequencies for rs3817198 can be compared with that of the GWAS studies.

Malignant epithelial cells in the breast and metastatic cells in the lymph nodes demonstrate an activated and hyper-expressed *MAPK* pathway (Sivaraman *et al.*, 1997*)*. This resulted in the hypothesis that the *MAPK* signaling pathway might play a role in the initiation and pathogenesis of BC (Sivaraman *et al*., 1997; Coutts *et al*., 1998). The initial association with BC was correct, as GWAS identified rs889312, a SNP present within *MAP3K1* that acts as a modifier of BC risk (Easton *et al.*, 2007; Antoniou *et al.*, 2008; Gates *et al.*, 2009; Latif *et al.*, 2009). The association has been validated for *BRCA2* mutation carriers and implicates a per-allele HR of 1.12 (Easton *et al*., 2007; Antonio *et al*., 2008).

The SNP in *MAP3K1* was common within the Afrikaner population, although the majority of Controls and *BRCA2* mutation carriers were homozygous for the ancestral A allele (71.67% for Controls and 68.33% for Carriers, Table 4.5). The variant C allele was observed in a heterozygous form in 18.33% of the Controls and 21.67% of the Carriers, with 12 individuals being homozygous for the variant C allele (10.00% each for the Controls and Carriers, Table 4.5, Comparison 1). The NCBI (http://www.ncbi.nlm.nih.gov) reported that the homozygous ancestral (A/A) genotype frequency amongst white Europeans are 0.484, 0.419 for heterozygotes (A/C) and 0.097 for the homozygous variant (C/C) genotype. The white European frequencies indicated by Hapmap are 0.476 for the homozygous ancestral genotype (A/A), 0.450 for the heterozygous (A/C) genotype and 0.083 for the homozygous (C/C) genotype. When the international data and that of the current study are compared, it is evident that the variant is less frequent within the Afrikaner population.

The minor allele frequency within the Afrikaner varied from 0.19 for the Controls to 0.20 for the Carriers (Table 4.5, Comparison 1). As expected from a validated modifier of cancer risk, the risk allele was more common within the BC patients compared to the unaffected Cases (0.23 compared to 0.18, Comparison 2). The frequencies observed within the current study were lower than that reported for other studies. The minor allele frequencies for Controls representing white Europeans varied from 0.26 to 0.31 (Huijts *et al*., 2007; Antoniou *et al.*, 2008; Latif *et al.*, 2009). The minor allele frequency increased to 0.30 within their *BRCA2* mutation carriers (Antoniou *et al.*, 2008) and to 0.32 in British BC affecteds, negative for *BRCA1/2* (Latif *et al.*, 2009). The latter frequency again confirms the association between an increase in BC risk and the presence of the variant allele.

The *FGFR2* gene is involved in several biological processes and plays a role in mammary gland development, cell growth, invasiveness, motility and tumor genesis (Liang *et al.*, 2008). Adnane *et al*. (1991) reported amplification and over expression of *FGFR2* in 5 – 10% of breast tumours. The rs2981582 SNP in *FGFR2* was the first modifier that was identified by GWAS (Easton *et al.*, 2007; Gates *et al.*, 2009; Latif *et al.*, 2009). It was also the first validated SNP to be associated with an increased risk for BC specifically in *BRCA2* mutation carriers (Antoniou *et al.*, 2008).

The results of this study revealed that the ancestral C allele of rs2981582 in *FGFR2* was the most common among participants (48.33% for the Controls compared to 46.67% for the Carriers, Table 4.5). The genotype frequencies for both heterozygotes and homozygotes carrying the variant allele differed between these two groups (Comparison 1). Although the Controls had more heterozygous individuals (41.67% versus 35%), the variant allele present in homozygous form was more common amongst the Carriers. In Comparison 2 (Table 4.5), a difference in the heterozygous and homozygous frequencies for the variant allele was observed between the BC patients and Cases. Although more Cases exhibited one copy of the variant allele (43.33%) compared to the BC patients (26.67%), more of the affected patients were homozygous for the risk allele (23.33% versus 13.33%). This was also expected as each copy of the risk allele confers a HR of 1.32 (95%CI:

1.20–1.45) (Antoniou *et al.*, 2008). A homozygous carrier for the variant T allele will thus have an additional risk of 2.64 added to her *BRCA2* 8162delG risk.

When the *FGFR2* data for the Afrikaner is compared to that of white Europeans, the frequency of the variant allele within the Afrikaner appeared lower than expected as various European countries including France, Brittan, the Netherlands and Belgium are the main ancestors of the Afrikaner population. The genotype frequencies of the white Europeans are listed by the NCBI as 0.387 for the homozygous ancestral (C/C) genotype, 0.419 for heterozygotes (C/T) and 0.194 for the homozygous variant (T/T) genotypes. The International HapMap project indicates approximately similar frequencies of 0.336 for the homozygous ancestral (C/C) genotype, 0.416 for the heterozygotes (C/T) and 0.248 for the homozygous variant (T/T) genotype. These percentages indicated that the variant rs2981582 allele is as common within the Afrikaner population when compared to the Europeans (Huijts *et al*., 2007; Antoniou *et al.*, 2008; Gates *et al.*, 2009; Gorodnova *et al.*, 2010).

Each of the GWAS validated SNPs has also been tested for associations with risk factors such as age at menarche, parity, age at menopause, family history, body mass index, ER status (positive or negative tumours), tumor grade and node positivity (Huijts *et al*., 2007; Antonio *et al.*, 2008; Garcia-Closas *et al*., 2008; Stacey *et al.*, 2008; Gates *et al.*, 2009; Gorodnova *et al.*, 2010). The majority of these studies recorded various associations which included a stronger association with ER-positive than ER-negative tumors (P = 2.10^{-13}), a better association with lower than higher grade tumours (P = 2.10^{-28}) and a better association with node positive than negative tumours (P = 0.013). The factors for which positive associations were obtained are all factors that normally characterize *BRCA2* tumours, for *BRCA2* tumours are in most cases ER^+, have a lower grade and generally have a better outcome (Loman *et al*., 1998; Noguchi *et al*., 1999).

The stronger association of rs2981582 with ER^+ tumours is supported by higher expression levels of *FGFR2* in ER^+ compared to ER^- cell lines and tumours confirming that *FGFR2* is involved in estrogen-related carcinogenesis of the breast

(Hishikawa *et al.*, 2004; Tamaru *et al.*, 2004). It is hypothesized that the causative variant of rs2981582 is situated in the region of intron 2 which contains multiple transcription factor binding sites. These sites may mediate its association with BC risk through differential levels of *FGFR2* expression (Easton *et al.*, 2007). The presence of the variant allele, especially in a homozygous state, could be clinically relevant for a small subset of tumours that express high levels of the FGFR2 receptor. Although the minor allele frequency within the Afrikaner population was lower than expected (Table 4.5), this SNP could still play a role in SA diagnostic setting as the *BRCA2* 8162delG (c.7934del, p.Arg2645AsnfsX3) mutation is the most frequent *BRCA* mutation observed within this population.

4.4.3 Multiplicative combined genotype

Antoniou *et al.* (2008) investigated the multiplicative effect of the risk alleles of *FGFR2* and *TNRC9* on BC risk in *BRCA2* mutation carriers. The HR estimates for the six possible genotypes varied from a low HR of 1.00 for the homozygous ancestral alleles (GG/CC) to 2.26 for carriers of the homozygous risk alleles at both loci (AA/TT).

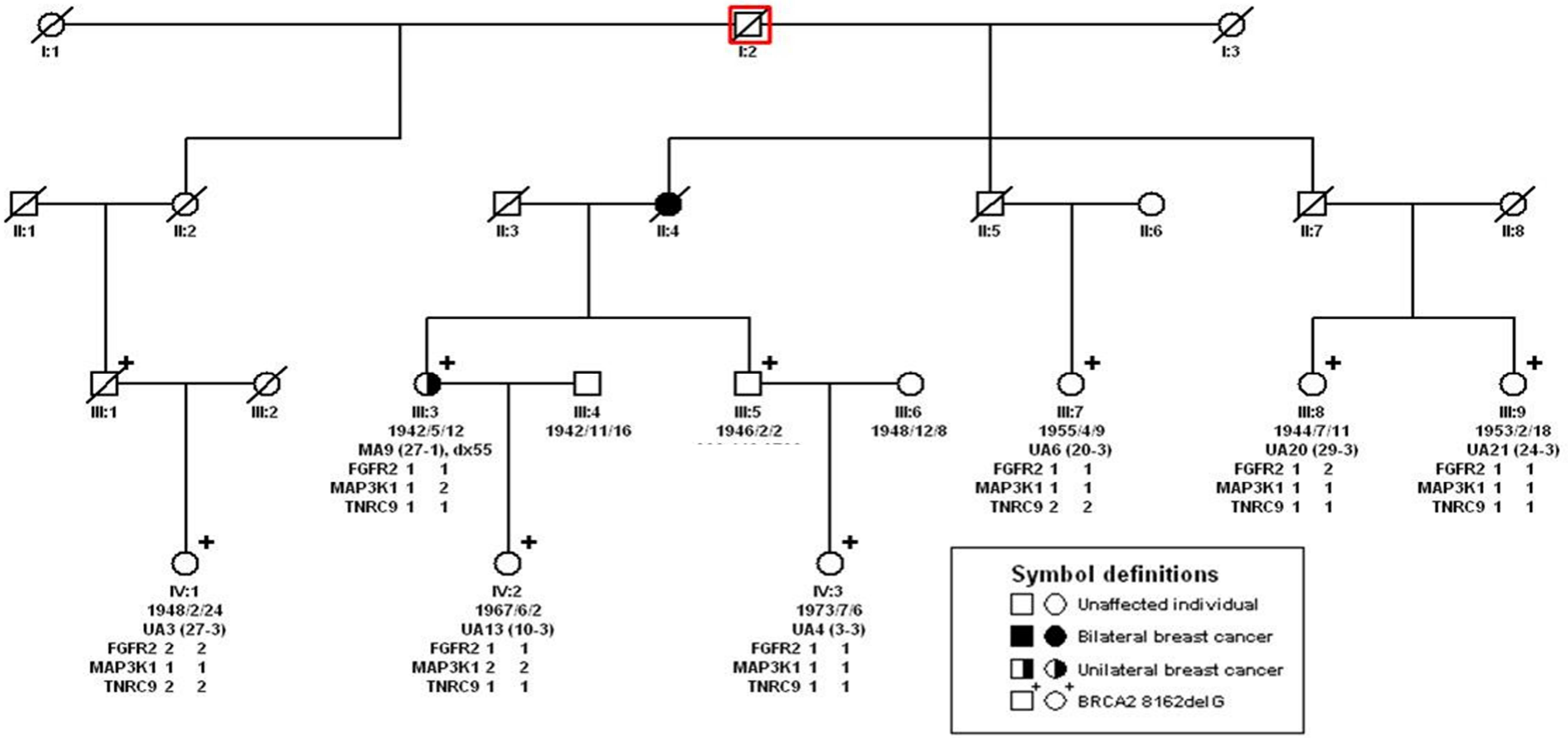
A multi-locus recombinant haplotype was compiled for the *FGRF2*, *TNRC9*, *MAP3K1* and *LSP1* SNPs to determine their combined influence on BC risk (Table 4.6). For Comparison 1, the most common haplotypes for the Controls were 1010, 1011 and 1111 (each 18.33%). A haplotype of 1111 (16.67%) was the most common for the *BRCA2* mutation carriers, followed by 1010 (15.0%) and 1000 (13.33%). From the data, it is clear that the risk allele for *TNRC9* was the most common SNP present (Table 4.6). The data obtained for Comparison 2 indicated that the majority of Carriers exhibited a minimum of one of the variant alleles (Table 4.6). Some minor differences regarding genotype frequencies were observed between the BC patients and Cases, although none was statistically significant.

The possibility of the BC recombinant haplotype influencing the etiology of BC in some of the Afrikaner families was tested by applying the haplotype data of *FGFR2*, *MAP3K1* and *TNRC9* (refer to 4.5). The fourth SNP namely *LSP1* was omitted due the potential of skewed data. The presented pedigrees are families in which the *BRCA2* disease-causing mutation was detected. Family 6 was included in order to illustrate the variability of penetrance in families (Fig. 4.6). The other two (Families 11 and 14) were selected based on an equal number of BC affected patients and unaffected Cases tested within each family (Fig. 4.7 and 4.8).

Incomplete penetrance was most obvious within Family 6 (Fig. 4.6). Although many family members have been affected with other cancer types, only two individuals were affected with BC, namely II:4 and MA9 – III:3. The age at diagnosis of the only living BC patient (MA9 – III:3) was 55. As this patient was heterozygous for the risk allele in *MAP3K1* only, she had a limited additional risk of 1.12 (Antoniou *et al.*, 2008) added to her initial *BRCA2* 8162delG risk. The absence of the risk alleles for the other two variants could have contributed to her post-menopausal age at diagnosis. The ages of the unaffected carriers (UA6 – III:7, UA20 – III:8, UA21 – III:9 and UA3 – IV:1) varied at the time of collection from 51 to 62 (Fig. 4.6). The fact that these Carriers remains unaffected could perhaps be attributed to the presence of mostly the ancestral alleles (1 1 1) (Fig. 4.6). Case UA3 – IV:1 was the only Case who exhibited two of the variants (*FGFR2* and *TNRC9*) in a homozygous state, yet she remained unaffected at a post-menopausal stage despite her multiplicative risk of 2.64 for *FGFR2* (two variant alleles) and 2.3 for *TNRC9* (two minor alleles) (Antoniou *et al.*, 2008). A large part of the variation in penetrance for Family 6 remains unclear and is not explained by the presence of any of the investigated modifying SNPs.

BC was the most common cancer type occurring within Family 11 (Fig. 4.7). The ages at onset varied from as young as 36 years (MA5 – III:6) to age 57

Figure 4.6 Pedigree for Family 6. Indicated are sample and group numbers, ages at onset (dx), mutation status, date of death (where applicable) and symbol descriptions.



(MA21 – III:1) at the time of sample collection. When the haplotype of each of the affected patients are compared, two of the patients (MA5 – III:6 and MA7 – III:2) shared the risk allele of *TNRC9*, one in a homozygous state (MA7 – III:2) and one in a heterozygous state MA5 – III:6) (Fig. 4.7). The other BC patient (III:1) was diagnosed at a later stage (dx = 52) and exhibited the ancestral alleles for each of the BC risk modifying SNPs.

The ages of the three unaffected Cases were 68 (UA8 – III:3) and 59 (UA9 – III:4 and UA11 – III:5) (Fig. 4.7). Noteworthy is the fact that they too did not exhibit any of the risk alleles, therefore do not have a multiplicative risk added to their initial *BRCA2* mutation risk. Similar to Family 6, the presence of these confirmed SNPs did not explain the variation in penetrance within this family.

The female representatives of Family 14 were also mainly affected with unilateral and bilateral BC (Fig. 4.8). The ages at onset varied from 40 (MA3 – III:1) to 54 years (MA2 – III:2). Interestingly, this family included female twins (MA2 – III:2 and UA17 – III:3) that both tested positive for the *BRCA2* disease-causing mutation. The twins are currently 58 years old and differed to some extent with regards to their BC recombinant haplotypes. The affected sister (MA2 – III:2) was heterozygous for two of the three risk alleles, compared to the unaffected sister who was homozygous for the risk allele of *MAP3K1* only (Fig. 4.8).

Regarding the other two patients (MA3 – III:1 and MA22 – III:6) (Fig. 4.8), the results of the BC haplotype did not elucidate the role of these SNPs in the etiology of the disease, for they both exhibited a different risk allele. These BC patients, as well as the three unaffected cases, mostly displayed the ancestral alleles. The presence of BC within this family could not be attributed to a specific part of the BC recombinant haplotype, for there were no universal features amongst the three BC patients.

Figure 4.7 Pedigree for Family 11. Indicated are sample and group numbers, ages at onset (dx), mutation status, date of death (where applicable) and symbol descriptions.

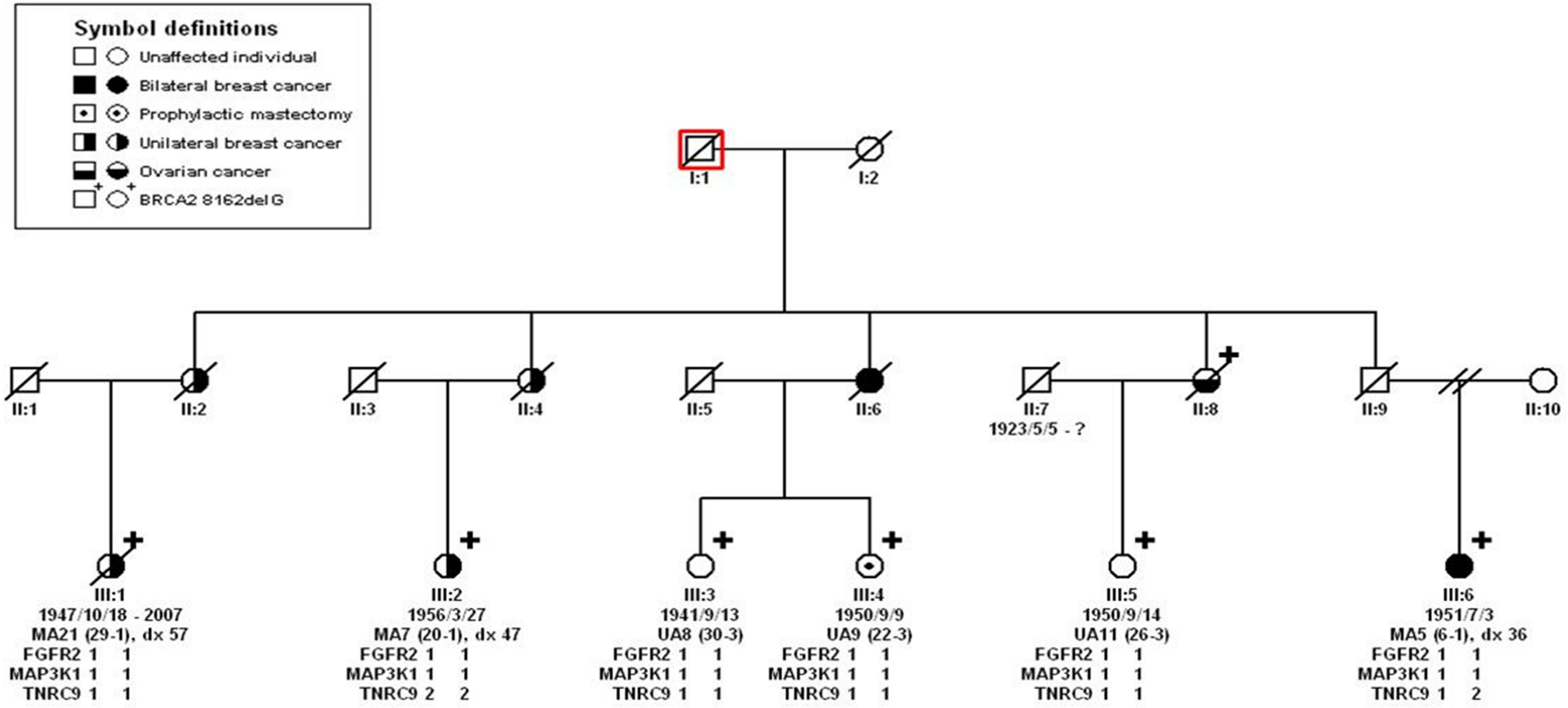
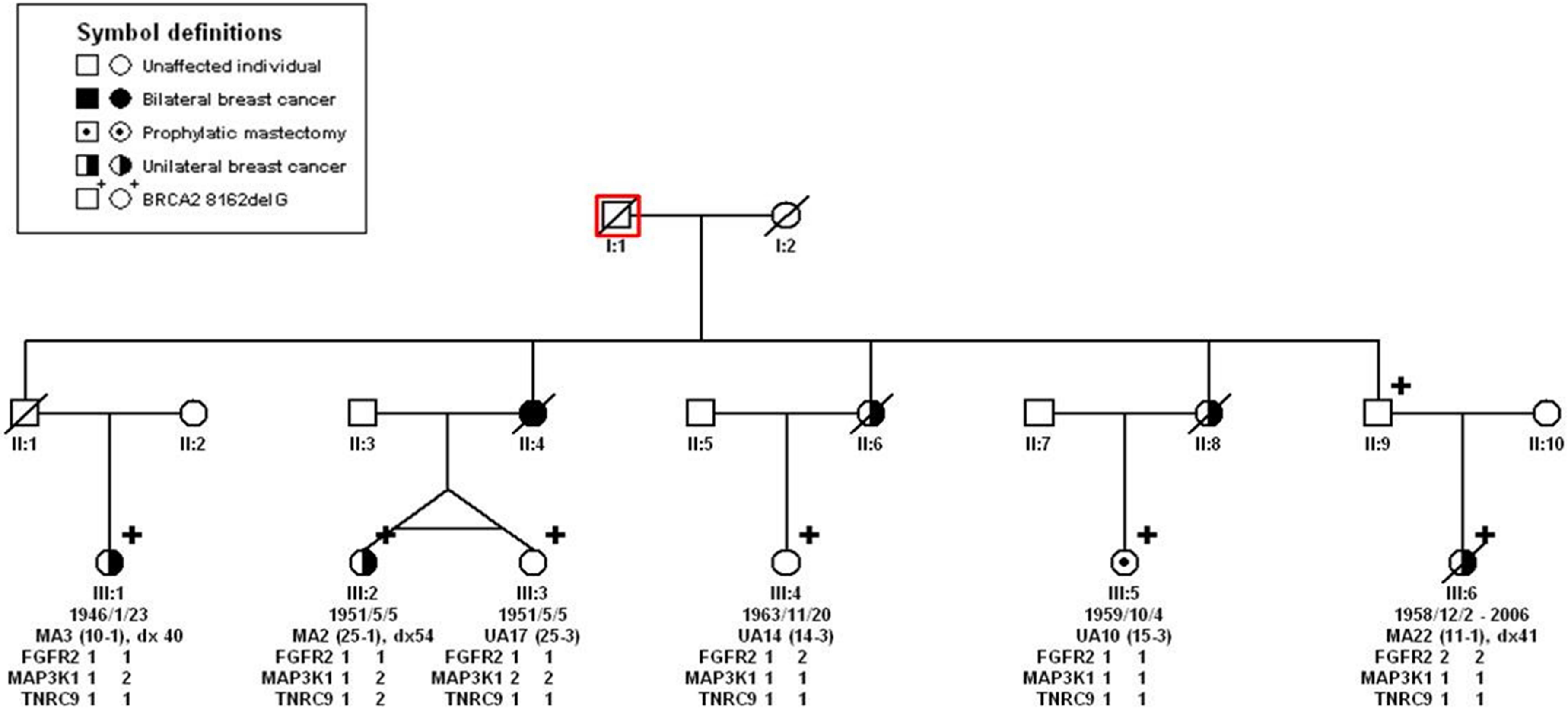


Figure 4.8 Pedigree for Family 14. Indicated are sample and group numbers, ages at onset (dx), mutation status and symbol descriptions.



4.5 Hardy-Weinberg equilibrium

Although the results of the HWE were determined for each of the studied polymorphisms, it did not prove to be critical due to the change in the focus of the study (Table 4.3). In general, deviations from HWE for the Control group only may indicate genotype errors, a too small sample size and population stratification. To confirm that deviations were not due to genotyping errors, genotype calls were duplicated and verified with no change to the data. Normally deviations in the Control group would lead to the exclusion of the SNP, however when both the Control and Case groups deviate from HWE (such as *TNRC9* and *MAP3K1*) the effects should cancel out and not indicate population stratification (Lam *et al*., 2003).

The HWE is based on the assumption that a population is static, therefore acquires no new mutations, is not subjected to any form of selection and that random mating occurs (Trikalinos *et al*. 2006; Ziegler *et al*. 2011). Deviations may therefore indicate failure in these assumptions. With non-random mating (inbreeding) as well as genetic drift which reduce the genetic diversity in a small population, an increase in homozygotes and subsequent decrease of heterozygotes is found. This was found for both *TNRC9* and *MAP3K1*. A founder event may also be considered a cause for deviation from HWE as human colonization of small migrating populations will lower the genetic diversity. These might all be true of the Afrikaner population, since it is a unique homogeneous population due to geographical and religious isolation. It is possible that these SNPs have been more recently introduced into the Afrikaner population by our European ancestors and still has to reach HWE. Deviations in the Case group indicate possible association between the genotype and the disease (Wigginton *et al*., 2005). As only the *BRCA2* mutation BC group of *FGFR2* deviates from the HWE, it can be implied that *FGFR2* is possibly associated with BC and needs to be further investigated (Table 4.3).

4.6 Closing remarks

Antoniou *et al.* (2008) estimated that the *TNRC9* SNP could account for 0.5% of the familial or genetic variance of BC risk for *BRCA1* mutation carriers and that the three SNPs in *FGFR2*, *TNRC9* and *MAP3K1* account for 2.8% of the familial variance in *BRCA2* mutation carriers. With their work on *LSP1*, Antoniou *et al.* (2009) concluded that in total, the four BC susceptibility variants identified by GWAS together with 2q35, account for 3.7% of the *BRCA2* genetic-modifying variance. However for the three presented Afrikaner pedigrees, it was clear that none of the SNPs could explain the variation of penetrance within the Afrikaner families affected with BC.

Since the Afrikaner is of mixed Dutch, French, Belgian and German descent, similar results to European studies were expected (Tipping *et al*., 2001; Greeff, 2007; van der Merwe *et al*., 2011). This was not the case, as for the majority of tested SNPs, the minor allele frequency observed for the Afrikaner was lower compared to that of the white European population. Several factors could have caused this. Allelic heterogeneity was high in the Afrikaner population which is a potential problem as it may affect the interpretation of results and complicate association studies of complex diseases such as BC.

Another major limitation was the small study sample size which contributed to the lack of statistical power to detect modest BC risk ratios associated with the six studied polymorphisms. This was mostly due to the fact that the study only included participants that carried the specific *BRCA2* founder mutation in order for it to serve as a base line. This approach was similar to that of Gaudet *et al.* (2010) that selected participants carrying the Ashkenazi Jewish *BRCA2* 6174delT (c.5946delT) mutation. Only three other studies included *BRCA1* and *BRCA2* participants specifically, even though the mutation carriers exhibited a variety of *BRCA* disease-caution mutations (Antoniou *et al*., 2008; Antoniou *et al*., 2009; Latif *et al*., 2010). The other international studies presented in literature included sporadic BC participants and sometimes did not mention the BC status of the participants. Most studies also included both male and female participants.

The current study focused only on Caucasian female *BRCA2* 8162delG (c.7934del, p.Arg2645AsnfsX3) mutation carriers affected with BC. As the focus was only on genetic or familial variation, this study did not take into account any of the tumor characteristics or epidemiological factors such as age at menarche. The fact that this study focused on a specific homogeneous ethnic group all carrying the same founder mutation provided strength to the study. These results, although currently insignificant, provided preliminary results to be explored in further studies. Sequencing of the second putative SNP in all of the indicated heterozygotes of *LSP1* still has to be confirmed. Large scale genotyping of BC patients positive for the *BRCA2* founder mutation that share the same environment could help elucidate the relationship between the GWAS polymorphisms and BC risk within the Afrikaner.

Chapter 5

Conclusion

The aim of the current study was to investigate possible modifiers of BC risk within a selected group of Caucasian Afrikaner mutation carriers, all carrying the *BRCA*2 8162delG (c.7934del, p.Arg2645AsnfsX3) founder mutation. As the results of various GWAS studies became available during the duration of this study, there was a shift in the research question of the study. The focus shifted from the initial search for potential BC risk modifiers within the Afrikaner to the genotype distribution and allele frequencies of the validated GWAS polymorphisms within this selected group. Apart from the fact that conflicting results exist regarding the six selected polymorphisms and their role in BC risk and penetrance, it was proposed that because the South African Afrikaner is such a universally unique population group (founder effects for various diseases), the search effort within the Afrikaner population will be justified.

The effect of the rs2234693 (*PvuII*) and rs9340799 (*XbaI*) SNPs in *ESR1* on penetrance within the mutation positive Afrikaner delivered no results with statistical significance. However, a possible association was observed between the combined minor allele frequencies of rs2234693 (*PvuII*) for the Controls versus the *BRCA2* mutation carriers. This potential association needs to be further investigated within a larger study group. A haplotype compiled using rs2234693 (*PvuII*) and rs9340799 (*XbaI*) proved to be uninformative as it revealed no significant differences between the BC patients and Carriers (Comparison 2).

The Taqman® genotyping assays for rs3803662 (*TNRC9*), rs3817198 (*LSP1*), rs889312 (*MAP3K1*) and rs2981582 (*FGFR2*) were done using the BioRad CFX Manager v1.1.308.1111 software for allelic discrimination. However, when manual and automated discrimination methods were compared, Cohen's kappa analysis suggested that Methods 1 (automatic allelic discrimination) and 3 (manual allelic discrimination taking both RF and Cq values into account) were the closest matched. The accuracy of the data does however depend on accurate probe design, optimized PCR conditions and the inclusion of positive controls. As long as these requirements are met and three distinct scatter plot clusters are observed, robust genotyping can be performed. Since this project has possible diagnostic application in the future, Method 1 (automated calling) was selected as the method of choice based on the Kappa results. It is impractical and not cost effective to manually call genotypes run on a real-time instrument within a diagnostic laboratory, where turn-around times of samples are critical.

Sequencing of each of the SNP amplicons proved to be critical, as the genotyping data for *LSP1* was influenced by the presence of an additional SNP within the probe binding region. Caution should therefore be taken when incorporating new modifiers of BC risk within *BRCA* mutation carriers into a diagnostic platform. Each fragment analyzed within a new population should be sequenced for the presence of potential new SNPs not previously described or expected. The presence of an additional SNP within this region could have gone unnoticed and influenced the dataset without the investigators knowing, was it not for the comparison of the datasets. The alarming discrepancy amongst the heterozygous individuals was highlighted by Method 2 (any amplification above the baseline irrespective of the Cq value), which resulted in a stricter analysis of the sequencing data in order to determine the possible cause. If the various datasets were not compared, this SNP could have gone undetected, as both Methods 1 and 3 discarded amplification of the minor allele in all samples that had a Cq > 1, which in fact was due to the weaker binding of the probe to the template containing the additional SNP. The validity of the *LSP1* results should be treated cautiously as no information on the frequency of the second putative SNP in white European individuals is available. The frequency of the SNP must first be confirmed in a larger Afrikaner study group through sequencing before the data can be confidently used.

Although several noteworthy differences were observed for all four SNPs between the BC patients and Carriers, no statistical significance was observed. A BC multi-locus recombinant haplotype was compiled for polymorphisms in the *FGRF2, TNRC9*, *MAP3K1* and *LSP1* genes to determine their combined influence on BC risk. This haplotype was uninformative as it revealed no differences between the BC patients and Carriers (Comparison 2). Deviations from the HWE for the *BRCA2* mutation carrier group of *FGFR2* implied a possible association with BC within the Afrikaner. Since this SNP is already confirmed to have a modifying effect on BC and included in diagnostic tests, its role within the Afrikaner population needs to be further studied within an expanded Afrikaner population.

To conclude, no significant associations were observed between the six BC susceptibility alleles and BC risk in the *BRCA*2 8162delG (c.7934del, p.Arg2645AsnfsX3) Afrikaner population. Allelic heterogeneity is high in the Afrikaner population which could cause a potential problem as it may affect the interpretation of the results. A major limitation of this study was the small sample size which contributed to the lack of statistical power to detect modest BC risk ratios associated with the six polymorphisms. The fact that this study focused on a specific homogeneous ethnic group all carrying the same founder mutation provided strength to the study. These results provided new ideas for further association studies. Large scale genotyping of BC patients positive for *BRCA* mutations that share the same environment could help elucidate the relationship between the selected polymorphisms and BC risk.

Chapter 6

References

6.1 General references

Adnane J, Gaudray P, Dionne CA, Crumley G, Jaye M, Schlessinger J, *et al*. (1991). BEK and FLG, two receptors to members of the FGF family, are amplified in subsets of human breast cancers. Oncogene 6: 659-663.

Ahmed S, Thomas G, Ghoussaini M, Healey CS, Humphreys MK, Platte R, *et al*. (2009). Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. Nature Genetics 41: 585-590.

Andersen TI, Heimdal KR, Skrede M, Tveit K, Berg K, and Borresen AL. (1994). Oestrogen receptor (ESR) polymorphisms and breast cancer susceptibility. Human Genetics 94: 665-670.

Andrieu N, Goldgar DE, Easton DF, Rookus M, Brohet R, Antoniou AC, *et al*. (2006). Pregnancies, breast-feeding and breast cancer risk in the International *BRCA1/2* Carrier Cohort Study (IBCCS). Journal of the National Cancer Institute 98: 535-544.

Antoniou AC, Goldgar DE, Andrieu N, Chang-Claude J, Brohet R, Rookus MA, *et al*. (2005) A weighted cohort approach for analysing factors modifying disease risks in carriers of high-risk susceptibility genes. Genet Epidemiol 29: 1-11.

Antoniou AC, Pharoah PD, McMullan G, Day NE, Stratton MR, Peto J, *et al*. (2002). A comprehensive model for familial breast cancer incorporating *BRCA1*, *BRCA2* and other genes. British Journal of Cancer 86: 76-83.

Antoniou AC, Pharoah PDP, Narod S, Risch HA, Eyfjord JE, Hopper JL, *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, Sinilnikova OM, McGuffog L, Healey S, Nevanlinna H, Heikkinen T, *et al*. (2009). Common variants in *LSP1*, 2q35 and 8q24 and breast cancer risk for *BRCA1* and *BRCA2* mutation carriers. Human Molecular Genetics 18(22): 4442-4456.

Antoniou AC, Spurdle AB, Sinilnikova OM, Healey S, Pooley KA, Schmutzler RK, *et al.* (2008). Common breast cancer-predisposition alleles are associated with breast cancer, risk in *BRCA1* and *BRCA2* mutation carriers. The American Journal of Human Genetics 82: 937-948.

Arya M, Shergill IS, Williamson M, Gommersall L, Arya N and Patel HR. (2005). Basic principles of real-time quantitative PCR. Expert Review of Molecular Diagnostics 5: 209-219.

Bane AL, Pinnaduwage D, Colby S, Bull SB, O'Malley FP and Andrulis IL. (2009). Expression profiling of familial breast cancers demonstrates higher expression of FGFR2 in *BRCA2*-associated tumours. Breast Cancer Research and Treatment 117(1): 183-191.

Baron JA, La Vecchia C and Levi F. (1990). The antiestrogenic effect of cigarette smoking in women. American Journal of Obstetrics and Gynecology 162: 502-514.

Bartsch H and Nair J. (2004). Oxidative stress and lipid peroxidation-derived DNA-lesions in inflammation driven carcinogenesis. Cancer Detection and Prevention 28(6): 385-391.

Bau DT, Mau YC and Shen CY. (2006). The role of BRCA1 in non-homologous end joining*. Cancer Letters* 240*:* 1-8*.*

Beatson CT. (1896). On the treatment of inoperable cases of carcinoma of the mamma: suggestions for a new method of treatment with illustrative cases. Lancet 2: 104-107.

Begg C, Haile R, Borg A, Malone K, Concannon P, Thomas D, *et* al. (2008). Variation of breast cancer risk among *BRCA1/2* Carriers. Journal of the American Medical Association 299(2):194-201.

Bertwistle D, Swift S, Marston NJ, Jackson LE, Crossland S, Crompton MR, *et al*. (1997). Nuclear location and cell cycle regulation of the BRCA2 protein. Cancer Research 57(24): 5485-5488.

Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D, *et al*. (2006). Oncogenic pathway signatures in human cancers as a guide to targeted therapies. Nature 439: 353-357.

Boardman LA, Thibodeau SN, Schaid DJ, Lindor NM, McDonnell SK, Burgart LJ, *et al*. (1998). Increased risk for cancer in patients with the Peutz-Jeghers syndrome. Annals of Internal Medicine 128: 896-899.

Boilly B, Vercoutter-Edouart AS, Hondermarck H, Nurcombe V and Le Bourhis X. (2000). FGF signals for cell proliferation and migration through different pathways. Cytokine Growth Factor Reviews 11: 295-302.

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

Botstein D, White RL, Skolnick M and Davis RW. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics 32: 314-331.

Boulton SJ. (2006). Cellular functions of the BRCA tumour-suppressor proteins. Biochemical Society Colworth Medal Lecture 34: 633-645.

Boyapati SM, Shu XO, Ruan ZX, Cai Q, Smith JR, Wen W, *et al*. (2005). Polymorphisms in *ER*-α gene interact with estrogen receptor status in breast cancer survival. Clinical Cancer Research 11(3): 1093-1098.

Broca P. (1866). *Traite des tumeurs*. Paris ; *Asselin* 151-155.

Brose MS, Rebbeck TR, Calzone KA, Stopfer JE, Nathanson KL and Weber BL. (2002). Cancer risk estimates for *BRCA1* mutation carriers identified in a risk evaluation program. Journal of the National Cancer Institute 94:1365-1372.

Brunet JS, Ghadirian P, Rebbeck TR, Lerman C, Garber JE, Tonin PN, *et al*. (1998). Effect of smoking on breast cancer in carriers of mutant *BRCA*1 or *BRCA*2 genes. Journal of the National Cancer Institute 90: 761-766.

Cai Q, Shu XO, Jin F, Dia Q, Wen W, Cheng JR, *et al*. (2003). Genetic polymorphisms in the estrogen receptor alpha gene and risk of breast cancer: results from the Shanghai Breast Cancer Study. Cancer Epidemiology Biomarkers and Prevention 12(9): 853-859.

Calle EE, Rodriquez C, Walker-Thurmond K and Thun MJ. (2003). Overweight, obesity, and morality from cancer in a prospectively studied cohort of U.S. adults. New England Journal of Medicine 348: 1625-1638.

Callegaro A, Spinelli R, B L, Bicciato S, Caristina L, Censuales S, *et al.* (2006). Algorithm for automatic genotype calling of single nucleotide polymorphisms using the full course of TaqMan real-time data. *Nucleic Acids Research* **34**:e56.

Campa D, Kaaks R, LeMarchand L, Haiman CA, Travis RC, Berg CD, *et al*. (2011). Interaction between genetic variants and breast cancer risk factors in the breast and prostate cancer cohort consortium. Journal of the National Cancer Institute 103 (16): 1252-1263.

Carpenter CL, Ross RK, Paganini-Hill A and Bernstein L. (2003). Effect of family history, obesity and exercise on breast cancer risk among postmenopausal women. International Journal of Cancer 106: 96-102.

Casazza K, Page GP and Fernandez JR. (2010). The association between the rs2234693 and rs9340799 estrogen receptor a gene polymorphisms and risk factors for cardiovascular disease: A review. Biological Research for Nursing 12: 84-97.

Chen MB, Wu XY, Shen W, Wei MX, Li C, Cai B, *et al*. (2010). Association between polymorphisms of trinucleotide repeat containing 9 gene and breast cancer risk: evidence from 62,005 subjects. Breast Cancer Research and Treatment 124(1): 237-241.

Chen S and Parmigiani G. (2007). Meta-Analysis of *BRCA*1 and *BRCA*2 penetrance. Journal of Clinical Oncology 25(11): 1329-1333.

Chen Y, Chen CF, Riley DJ, Allred DC, Chen P-L, Von Hoff D, *et al*. (1995). Aberant subnuclear localization of *BRCA*1 in breast cancer. Science 270: 789-791.

Chen Y, Lee WH and Chew HK. (1999). Emerging roles of BRCA1 in transcriptional regulation and DNA repair. Journal of Cellular Physiology 181(3): 385-392.

Cipollini G, Tommasi S, Paradiso A, Aretini P, Bonatti F, Brunetti I, *et al*. (2004). Genetic alteration in hereditary breast cancer. Annals of Oncology 15: i7-i13.

Claus EB, Schildkraut JM, Thompson WD and Risch NJ. (1996). The genetic attributable risk of breast and ovarian cancer. Cancer 77: 2318-2324.

Cohen J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20: 37-46.

Collaborative Group on Hormonal Factors in Breast Cancer. (1996). Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53,297 women with breast cancer and 100,239 women without breast cancer from 54 epidemiological studies. Collaborative Group on Hormonal Factors in Breast Cancer. Lancet 347: 1713-1727.

Collaborative Group on Hormonal Factors in Breast Cancer. (2002). Breast cancer and breastfeeding: Collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50,302 women with breast cancers and 96,973 women without the disease. Lancet 360: 187-195.

Coutts AS and Murphy LC. (1998). Elevated mitogen-activated protein kinase activity in estrogen-nonresponsive human breast cancer cells. Cancer Research 58 4071-4074.

Cox A, Dunning AM, Garcia-Closas M, Balasubramanian S, Reed MWR, Pooley KA *et al*. (2007). A common coding variant in *CASP8* is associated with breast cancer risk. Nature Genetics 39: 352- 358

Creighton CJ, Hilger AM, Murthy S, Rae JM, Chinnaiyan AM and El-Ashry D. (2006) Activation of mitogen-activated protein kinase in estrogen receptor alpha-positive breast cancer cells in vitro induces an in vivo molecular phenotype of estrogen receptor alpha negative human breast tumours. Cancer Research 66(7): 3903-3911.

Dapic V, Carvalho MA and Monteiro AN. (2005). Breast cancer susceptibility and the DNA damage response. Cancer Control 12: 127-136.

Dumitrescu RG and Cotarla I. (2005). Understanding breast cancer risk: where do we stand in 2005? Journal of Cell and Molecular Medicine 9: 208-221.

Easton DF. (1999). How many more breast cancer predisposition genes are there? Breast Cancer 1: 14-17.

Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, *et al*. (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447: 1087-1093.

Eeles RA, Kote-Jarai Z, Giles GG, Olama AA, Guy M, Jugurnauth SK, *et al*. (2008). Multiple newly identified loci associated with prostate cancer susceptibility. Nature Genetics 40: 316-321.

Ellisen LW and Haber DA. (1998). Hereditary breast cancer. Annual Review of Medicine 49: 425-436.

Eswarakumar VP, Lax I and Schlessinger J. (2005). Cellular signaling by fibroblast growth factor receptors. Cytokine Growth Factor Reviews 16: 139-149.

Excoffier L. Laval G and Schneider S. (2005). Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1:47-50.

Fan S, Wang JA, Yuan R, Ma Y, Meng Q, Erdos MR, *et al*. (1999). BRCA1 inhibition of estrogen receptor signaling in transfected cells. Science 284: 1354-1356.

Ferla R, Calò V, Cascio S, Rinaldi G, Badalamenti G, Carreca I, *et al*. (2007). Founder mutations in *BRCA*1 and *BRCA*2 genes. Annals of Oncology 18(6): 93-98.

Feron VJ, Til HP, de Vrijer F, Woutersen RA, Cassee FR and van Bladeren PJ. (1991). Aldehydes: occurrence, carcinogenic potential, mechanism of action and risk assessment. Mutation Research 259: 363-385.

Ferrell JE. (1996). Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. Trends in Biochemical Sciences 21: 460-466.

Fleiss JL. (1981). Statistical Methods for Rates and Proportions. New York. John Wiley & Sons, Inc.

Fleiss JL, Levin B and Paik MC. (2003). Statistical methods for rates and proportions. Third edition. Hoboken, NJ: John Wiley & Sons, Inc.

Ford D, Easton DF, Bishop DT, Narod SA and Goldgar DE. (1994). The Breast Cancer Linkage Consortium. Risks of cancer in *BRCA1*-mutation carriers. Cancer Research 54: 1791-1794.

Ford D, Easton DF, Stratton M, Narod S, Goldgar D, Devilee P, *et al*. (1998). The Breast Cancer Linkage Consortium: Genetic heterogeneity and penetrance analysis of the *BRCA1* and *BRCA2* genes in breast cancer families. American Journal of Human Genetics 62:676-689.

Foulkes WD, Stefansson IM, Chappuis PO, Bégin LR, Goffin JR, Wong N, *et al.* (2003). Germline *BRCA1* mutations and a basal epithelial phenotype in breast cancer. Journal of the National Cancer Institute 95(19):1482-1485.

Garcia-Closas M and Chanock S. (2008). Genetic susceptibility loci for breast cancer by estrogen receptor status. Clinical Cancer Research 14(24): 8000-8009.

Garcia-Closas M, Hall P, Nevanlinna H, Pooley K, Morrison J, Richesson DA, *et al*. (2008). Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. PLoS Genetics 4: e1000054.

Gates MA, Tworoger SS, Terry KL, De Vivo I, Hunter DJ, Hankinson SE, *et al.* (2009). Breast cancer susceptibility alleles and ovarian cancer risk in 2 study populations. International Journal of Cancer 124(3): 729-733.

Gaudet M, Kirchhoff T, Green T, Vijai J, Korn JM, Guiducci C, *et al.* (2010). Common genetic variants and modification of penetrance of *BRCA2*-associated breast cancer. PLoS Genet 6(10):e1001183.

Gayther SA, Harrington P, Russell P, Kharkevich G, Garkavtseva RF and Ponder BAJ. (1997). Frequently occurring germline mutations of the *BRCA1* gene in ovarian cancer families from Russia. American Journal of Human Genetics 60: 1239-1242.

Gerber HP, Seipel K, Georgiev O, Hofferer M, Hug M, Rusconi S, *et al*. (1994). Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science* 263: 808-811.

Giguere V, Tremblay A and Tremblay GB. (1998). Estrogen receptor β: re-evaluation of estrogen and antiestrogen signaling. Steroids 63: 335-339.

Givol D, Eswarakumar VP and Lonai P. (2003). Molecular and cellular biology of FGF signaling. In: Epstein CJ, Erickson RP, Wynshaw-Boris A, editors. Inborn errors of development – the molecular basis of clinical disorders of morphogenesis. Oxford: Oxford University Press: 367-379.

Gold B, Kalush F, Bergeron J, Scott K, Mitra N, Wilson K, *et al*. (2004). Estrogen receptor genotypes and haplotypes associated with breast cancer risk. Cancer Research 64: 8891-8900.

Gorodnova TV, Esh K, Yanus GA, Katanugina AS, Abysheva SN, Togo AV, *et al*. (2010). Distribution of *FGFR2*, *TNRC9*, *MAP3K1*, *LSP1*, and 8q24 alleles in genetically enriched breast cancer patients versus elderly tumour-free women. Cancer Genetics and Cytogenetics 199: 69-72.

Goulart AC, Zee RYL and Rexrode KM. (2009). Estrogen receptor 1 gene polymorphisms and decreased risk of obesity in women. Metabolism 58(6): 759-764.

Grabrick DM, Hartmann LC, Cerhan JR, Vierkant RA, Therneau TM, Vachon CM, *et al.* (2000). Risk of breast cancer with oral contraceptive use in women with a family history of breast cancer. Journal of the American Medical Association 284(14): 1791-1798.

Greeff JM. (2007). Deconstructing Jaco: Genetic Heritage of an Afrikaner. Annals of Human Genetics 71: 674-688.

Greene GL, Gilna P, Waterfield M, Baker A, Hort Y and Shine J. (1986). Sequence and expression of human estrogen receptor complementary DNA. Science 231: 1150-1154.

Grodin JM, Siiteri PK and MacDonald PC. (1973). Source of estrogen production in postmenopausal women. Journal of Clinical Endocrinology and Metabolism 36: 207-214.

Haile RW, Thomas DC, McGuire V, Felberg A, John EM, Milne RL, *et al*. (2006). *BRCA1* and *BRCA2* mutation carriers, oral contraceptive use, and breast cancer before age 50. Cancer Epidemiology, Biomarkers and Prevention 15: 1863-1870.

Hall J, Couse J and Korach K. (2001). The multifaceted mechanisms of estradiol and estrogen receptor signaling. Journal of Biological Chemistry 276:36869-36872.

Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, *et al*. (1990). Linkage of early-onset familial breast cancer to chromosome 17q21. Science 250: 1684-1689.

Hartmann C, John AL, Klaes R, Hofmann W, Bielenu R, Koehleri R, *et al*. (2004). Large *BRCA1* gene deletions are found in 3% of German high-risk breast cancer families. Human Mutation 24: 534.

Heimdal K, Andersen TI, Skrede M, Fossa SD, Berg K and Borresen AL. (1995). Association studies of estrogen receptor polymorphisms in a Norwegian testicular cancer population. Cancer Epidemiology, Biomarkers and Prevention 4:123-126.

Heldring N, Pike A, Andersson S, Matthews J, Cheng G, Hartman J, *et al*. (2007). Estrogen receptors: How do they signal and what are their targets. Physiological Reviews 87: 905-931.

Hemminki K and Vaittinen P. (1998). Familial breast cancer in the family-cancer database. International Journal of Cancer 77(3): 386-391.

Hemminki K, Muller-Myhsok B, Lichtner P, Engel C, Chen B, Burwinkel B, *et al*. (2010). Low-risk variants *FGFR2*, *TNRC9* and *LSP1* in German familial breast cancer patients. International Journal of Cancer 126: 2858-2862.

Henderson BE, Ross RK, Judd HL, Krailo MD and Pike MC. (1985). Do regular ovulatory cycles increase breast cancer risk? Cancer 56: 1206-1208.

Hilakivi-Clarke L. (2000). Estrogens, BRCA1, and breast cancer. Cancer Research 60(18): 4993-5001.

Hirschhorn JN and Daly MJ. (2005). Genome-wide association studies for common diseases and complex traits. Nature Reviews Genetics 6: 95-108

Hishikawa Y, Tamaru N, Ejima K, Hayashi T and Koji T. (2004). Expression of keratinocyte growth factor and its receptor in human breast cancer: its inhibitory role in the induction of apoptosis possibly through the overexpression of Bcl-2. Archives of Histology and Cytology 67:455-464.

Holland PM, Abramson RD, Watson R and Gelfand DH. (1991). Detection of specific polymerase chain reaction product by utilizing the 5′–3′ exonuclease activity of *Thermus aquaticus* DNA polymerase. Proceedings of the National Academy of Sciences U.S.A 88: 7276-7280.

Hsieh YY, Wang YK, Chang CC and Lin CS. (2007). Estrogen receptor {alpha}-351 XbaI*G and -397 PvuII*C-related genotypes and alleles are associated with higher susceptibilities of endometriosis and leiomyoma. Molecular Human Reproduction 13:117-122.

Huang C, Zhan L, Ai Y and Jongstra J. (1997). LSP1 is the major substrate for Mitogen-activated Protein Kinase-activated Protein Kinase 2 in human neutrophils. The Journal of Biological Chemistry 272: 17-19.

Hughes DJ. (2008). Use of association studies to define genetic modifiers of breast cancer risk in *BRCA1* and *BRCA2* mutation carriers. Familial Cancer 7: 233-244.

Huijts PE, Vreeswijk MP, Kroeze-Jansema KH, Jacobi CE, Seynaeve C, Krol-Warmerdam, *et al*. (2007). Clinical correlates of low-risk variants in *FGFR2*, *TNRC9*, *MAP3K1*, *LSP1* and 8q24 in a Dutch cohort of incident breast cancer cases. Breast Cancer Research 9: R78.

Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, *et al*. (2007). A genome-wide association study identifies alleles in *FGFR2* associated with risk of sporadic postmenopausal breast cancer. Nature Genetics 39: 870-874.

Hunter DJ, Spiegelman D, Adami HO, van den Brandt PA, Folsom AR, Goldbohm RA, *et al*. (1997). Non-dietary factors as risk factors for breast cancer, and as effect modifiers of the association of fat intake and risk of breast cancer. Cancer Causes Control 8: 49-56.

Ingersoll RG, Paznekas WA, Tran AK, Scott AF, Jiang G and Jabs EW. (2001). Fibroblast growth factor receptor 2 (FGFR2): genomic sequence and variations. Cytogenetics and Cell Genetics 94(3-4): 121-126.

Inoue H, Nojima H and Okayama H. (1990). High efficiency transformation of *Escherichia coli* with plasmids. Gene 96: 23-28.

Istvan J, Zavela K and Weidner G. (1992). Body weight and psychological distress in NHANES I. International Journal of Obesity 16: 999–1003.

Jabs EW, Li X, Scott AF, Meyers G, Chen W, Eccles M, *et al*. (1994). Jackson-Weiss and Crouzon syndromes are allelic with mutations in fibroblast growth factor receptor 2. Nature Genetics 8: 275-279.

Jensen EV and Jacobsen HI (1962). Basic guides to the mechanism of estrogen action. Recent Progress in Hormone Research 18:387-414.

Jernström H, Johannsson O, Borg A and Olsson H. (1998). Do *BRCA*1 mutations affect the ability to breast-feed? Significantly shorter length of breast feeding among *BRCA*1 mutation carriers compared with their unaffected relatives. Breast 7: 320-324.

Jernström H, Lerman C, Ghadirian P, Lynch HT, Weber B, Garber M, *et al*. (1999). Pregnancy and risk of early breast cancer in carriers of *BRCA1* and *BRCA2*. Lancet 354:1846-1850.

Jin W, Chen U, Di G, Miron P, Hou Y, Gao H, *et al*. (2008). Estrogen Receptor (ER)β or p53 attenuates ERα-mediated transcriptional activation on the *BRCA2* promoter. The Journal of Biological Chemistry 283(44): 29671-29680.

Jongstra J, Ittel ME, Iscove NN and Brady G. (1994). The *LSP1* gene is expressed in cultured normal and transformed mouse macrophages. Molecular Immunology 31(15):1125-1131.

Jongstra-Bilen J and Jongstra J. (2006). Leukocyte-Specific Protein 1 (LSP1) A Regulator of Leukocyte Emigration in Inflammation. Immunologic Research 35(1-2): 65-73.

Katoh M. (2008). Cancer genomics and genetics of FGFR2 [review]. International Journal of Oncology 33(2): 233-237.

Kaufman DW, Slone D, Rosenberg L, Miettinen OS and Shapiro S, *et al*. (1980). Cigarette smoking and age at natural menopause. American Journal of Public Health 70: 420-422.

Kelsey JL, Gammon MD and John EM. (1993). Reproductive factors and breast cancer. Epidemiology Reviews 15: 36-47.

Key TJ, Verkasalo PK and Banks E. (2001). Epidemiology of breast cancer. Lancet 2: 133-140.

Khanna KK (2000). Cancer risk and the *ATM* gene: a continuing debate. Journal of the National Cancer Institute 92: 792-802.

Khanna KK and Jackson SP. (2001). DNA double-strand breaks: signaling, repair and the cancer connection. National Genetics 27: 247-254.

Kim I, Moon S, Yu K, Kim U and Koh GY. (2001). A novel fibroblast growth factor receptor-5 preferentially expressed in the pancreas. Biochimica et Biophysica *Acta* 1518: 152-156.

King MC, Marks JH and Mandell JB. (2003). Breast and ovarian cancer risks due to inherited mutations in *BRCA1* and *BRCA2*. Science 302: 643-646.

Kinzler KW and Vogelstein B. (1997). Gatekeepers and caretakers. Nature 386:761-763.

Klint P and Claesson-Welsh L. (1999). Signal transduction by fibroblast growth factor receptors. Frontiers in Bioscience 4: 165-177.

Knudson AG. (1971). Mutation and cancer: statistical study of retinoblastoma. Proceedings of the National Academy of Sciences of U.S.A 8: 820-823.

Kobayashi S, Inoue S, Hosoi T, Ouchi Y, Shiraki M and Orimo H. (1996). Association of bone mineral density with polymorphism of the estrogen receptor gene. Journal of Bone and Mineral Research 11: 306-311.

Kouhara H, Hadari YR, Spivak-Kroizman T, Schilling J, Bar-Sagi D, Lax I, *et al*. (1997). A lipid-anchored Grb2-binding protein that links FGF receptor activation to the Ras/MAPK signaling pathway. Cell 89: 693-702.

Koziczak M, Holbro T, and Hynes NE. (2004). Blocking of FGFR signaling inhibits breast cancer cell proliferation through downregulation of D-type cyclins. Oncogene 23:3501-3508.

Krishna M and Narang H. (2008). The complexity of mitogen-activated protein kinases (MAPKs) made simple*.* Cellular and Molecular Life Sciences 65*:* 3525-3544*.*

Kruger WM and Apffelstaedt JP. (2007). Young breast cancer patients in the developing world: incidence, choice of surgical treatment and genetic factors. South African Family Practice 49(9): 18-24.

Kuiper GG, Enmark E, Pelto-Huikko M, Nilsson S and Gustafsson JA. (1996). Cloning of a novel receptor expressed in rat prostate and ovary. Proceedings of the National Academy of Science U.S.A 93: 5925-5930.

Lacey JV, Kreimer AR, Buys SS, Marcus PM, Chang SC, Leitzmann MF, *et al*. (2009). Breast cancer epidemiology according to recognized breast cancer risk factors in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial Cohort. BioMed Central Cancer 9(84): 1-8.

Lahmann PH, Hoffmann K, Allen N, van Gils CH, Khaw KT, Tehard B, *et al*. (2004). Body size and breast cancer risk: findings from the European Prospective Investigation into Cancer and Nutrition (EPIC). International Journal of Cancer 111: 762-771.

Lajeunie E, Ma HW, Bonaventure J, Munnich A, Le Merrer M and Renier D. (1995). *FGFR2* mutations in Pfeiffer syndrome. Nature Genetics. 9: 108.

Lam DS, Lee WS, Leung YF, Tam PO, Fan DS, Fan BJ, *et al.* (2003). TGF beta-induced factor: a candidate gene for high myopia. Investigative Ophthalmology and Visual Science 44(3): 1012-1015.

Latif A, Hadfield KD, Roberts SA, Shenton A, Lalloo F, Black GC, *et al*. (2010). Breast cancer susceptibility variants alter risks in familial disease. Journal of Medical Genetics 47: 126-131.

Lee JS, Collins KM, Brown AL, Lee CH and Chung JH. (2000). hCds1-mediated phosphorylation of BRCA1 regulates the DNA damage response. Nature 404: 201-204.

Lee WH and Boyer TG. (2001). *BRCA*1 and *BRCA*2 in breast cancer. The Lancet 358: 5.

Levy NM, Koeppen BM and Stanton BA. (2006). Berne & Levy Principles of Physiology. 4th Edition. St. Louis, MO, USA: Mosby Inc.

Li L, Zhou X, Huang Z, Liu Z, Song M and Guo Z. (2009). *TNRC9/LOC643714* polymorphisms are not associated with breast cancer risk in Chinese women. European Journal of Cancer Prevention 18: 285-290.

Li Y, Guerrero A and Howard TH. (1995). The actin-binding protein, lymphocyte-specific protein 1, is expressed in human leukocytes and human myeloid and lymphoid cell lines. Journal of Immunology 155(7): 3563-3569.

Liang J, Chen P, Hu Z, Shen H, Wang F, Chen L, *et al*. (2010). Genetic variants in trinucleotide repeatcontaining 9 (*TNRC9*) are associated with risk of estrogen receptor positive breast cancer in a Chinese population. Breast Cancer Research and Treatment 124: 237-241.

Liang J, Chen P, Hu Z, Zhou X, Chen L, Li M, *et al*. (2008). Genetic variants in fibroblast growth factor receptor 2 (*FGFR2*) contribute to susceptibility of breast cancer in Chinese women. Carcinogenesis 29(12): 2341-2346.

Liehr JG. (2000). Is estradiol a genotoxic mutagenic carcinogen? Endocrine Review 21(1): 40-54.

Liu H, Loo KK, Palaszynski K,' Ashouri J, Lubahn DB and Voskuhl RR. (2003). Estrogen Receptor α mediates estrogen's immune protection in autoimmune disease. The Journal of Immunology 171: 6936-6940.

Liu L, Cara DC, Kaur J, Raharjo E, Mullaly SC, Jongstra-Bilen J, *et al*. (2005). LSP1 is an endothelial gatekeeper of leukocyte transendothelial migration. Journal of Experimental Medicine. 201(3):409-418.

Livak KJ. (1999). Allelic discrimination using fluorogenic probes and the 5′ nuclease assay. Genetic Analysis 14: 143-149.

Livak KJ, Flood SJ, Marmaro J, Giusti W and Deetz K. (1995). Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods and Applications* 4:357-362.

Loman N, Johannsson O, Bendahl P-O, Borg Å, Fernö M and Olsson H. (1998). Steroid receptors in hereditary breast carcinomas associated with *BRCA1* or *BRCA2* mutations or unknown susceptibility genes. Cancer 83:310-319.

London SJ, Colditz GA, Stampfer MJ, Willett, WC, Rosner BA, Speizer FE, *et al*. (1989). Prospective study of smoking and the risk of breast cancer. Journal of the National Cancer Institute 81: 1625-1631.

Lorick KL, Jensen JP, Fang S, Ong AM, Hatakeyama S and Weissman AM. (1999). RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. Proceedings of the National Academy of Science U.S.A 96: 11364-11369.

Lorincz AM and Sukumar S. (2006). Molecular links between obesity and breast cancer. Endocrine-Related Cancer 13: 279-292.

Loubser F, Edge J and Fieggen K. (2008). Epidemiology, risk factors and genetics of breast cancer. Continuing Medical Education 26(10): 497-501.

Lu PH, Yang J, Li C, Wei MX, Shen W, Shi LP, *et al*. (2011). Association between mitogen-activated protein kinase kinase kinase 1 rs889312 polymorphism and breast cancer risk: evidence from 59,977 subjects. Breast Cancer Research and Treatment 126(3): 663-670

Lux MP, Fasching PA and Beckmann M. (2006). Hereditary breast and ovarian cancer: Review and future perspective. Journal of Molecular Medicine 84: 16-28.

Lynch HT, Marcus JM, Watson P, Conway T, Fitzsimmons ML and Lynch JF. (1989). Genetic epidemiology of breast cancer. In: Lynch HT, Hirayama T (eds) Genetic epidemiology of cancer. CRC Press, Boca Raton: 289-332.

MacMahon B, Cole P, Lin TM, Lowe CR, Mirra AP, Ravnihar B, *et al*. (1970). Age at first birth and breast cancer risk. Bulletin World Health Organization 43(2):209–221.

MacMahon B. Cigarette smoking and cancer of the breast. In: Wald N, Baron J, editors. Smoking and hormone-related disorders. Oxford: Oxford University Press, 1990: 154-166.

Malkin D, Li FP, Strong LC, Fraumeni JF, Nelson CE, Kim DH, *et al*. (1990). Germline p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. Science 250: 1233-1238.

Marston NJ, Richards WJ, Hughes D, Bertwistle D, Marshall CJ and Ashworth A. (1999). Interaction between the product of the breast cancer susceptibility gene *BRCA2* and DSS1, a protein functionally conserved from Yeast to mammals. Molecular Cell Biology 19(7):4633-4642.

Mcinerney N, Colleran G, Rowan A, Walther A, Barclay E, Spain S, *et al*. (2009). Low penetrance breast cancer predisposition SNPs are site specific. Breast Cancer Research and Treatment. 117(1): 151-159.

Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, *et al*. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*. Science 266: 66-71.

Mosselman S, Polman J and Dijkema R. (1996). ER beta: identification and characterization of a novel human estrogen receptor. Federation of European Biochemical Societies Letters 392:49-53.

Mote PA, Leary JA, Avery KA, Sandelin K, Chenevix-Trench G, Kirk JA, *et al*. (2004). Germ-line mutations in *BRCA*1 or *BRCA*2 in the normal breast are associated with altered expression of estrogen-responsive proteins and the predominance of progesterone receptor A. Genes, Chromosomes and Cancer 39: 236-248.

Mueller CR and Roskelley CD. (2003). Regulation of *BRCA1* expression and its relationship to sporadic breast cancer. Breast Cancer Research 5: 45-52.

Nan N, Qureshi AA, Hunter DJ and Han J. (2009). Genetic variants in *FGFR2* and *FGFR4* genes and skin cancer risk in the Nurses' Health Study. BioMed Central Cancer 2009, 9:172.

Neuhausen SL. (2000). Founder populations and their uses for breast cancer genetics. Breast Cancer Research 2: 77-81.

Newman LA, Kuerer HM, Hunt KK, Vlastos G, Ames FC, Ross MI, *et al*. (2001). Educational review: Role of the surgeon in hereditary breast cancer. Annals of Surgical Oncology 8(4): 368-378.

Nilsson M, Dahlman I, Ryden M, Nordstrom EA, Gustafsson JA, Arner P, *et al*. (2007). Oestrogen receptor alpha gene expression levels are reduced in obese compared to normal weight females. International Journal of Obesity 31: 900-907.

Noguchi S, Kasugai T, Miki Y, Fukutomi T, Emi M and Nomizu T. (1999). Clinicopathologic analysis of BRCA1- or BRCA2-associated hereditary breast carcinoma in Japanese women. Cancer. 85:2200-2205

Noruzinia M, Coupier I and Pujol P. (2005). Is *BRCA*1/*BRCA*2-related breast carcinogenesis estrogen dependent? Cancer 104: 1567-1574.

Offit K. (2006). BRCA mutation frequency and penetrance: new data, old debate. Journal of the National Cancer Institute 98: 1675-1677.

O'Flaherty E and Kaye J. (2003). TOX defines a conserved subfamily of HMG-box proteins. BioMed Central Genomics 4:13.

Olopade OI, Fackenthal JD, Dunston G, Tainsky MA, Collins F and Whitfield-Broome C. (2003). Breast cancer genetics in African Americans. Cancer 97: 236-245.

Olsen JH, Hahnemann JM, Borresen-Dale AL, Brondum-Nielsen K, Hammarstrom L, Kleinerman R, *et al.* (2001). Cancer in patients with ataxia-telangiectasia and in their relatives in the Nordic countries. Journal of the National Cancer Institute 93: 121-127.

Ong SH, Guy GR, Hadari YR, Laks S, Gotoh N, Schlessinger J, *et al.* (2000). FRS2 proteins recruit intracellular signaling pathways by binding to diverse targets on fibroblast growth factor and nerve growth factor receptors. Molecular and Cellular Biology 20: 979-989.

Pal T, Permuth-Wey J, Holtje T and Sutphen R. (2004). *BRCA1* and *BRCA2* mutations in a study of African American breast cancer patients. Cancer Epidemiology, Biomarkers and Prevention 13: 1794-1799.

Parkin DM, Whelan S, Ferlay J, Teppo L and Thomas DB. (2002). Cancer incidence in five continents. International Agency of Research on Cancer Scientific Publication 155 (8).

Pearce ST and Jordan VC. (2004). The biological role of estrogen receptors α and β in cancer. Critical Reviews in Oncology/Hematology 50: 3-22.

Peto J, Collins N, Barfoot R, Seal S, Warren W, Rahman N, *et a*l. (1999). Prevalence of *BRCA1* and *BRCA2* gene mutations in patients with early-onset breast cancer. Journal of National Cancer Institute 91:943-949.

Peto J. (2002). Breast cancer susceptibility - a new look at an old model. Cancer Cell 1: 411-412

Petri B, Kaur J, Long EM, Li H, Parsons SA, Butz S, *et al.* (2011). Endothelial LSP1 is involved in endothelial dome formation, minimizing vascular permeability changes during neutrophil transmigration in vivo. Blood 117: 942-952

Pettersson K and Gustafsson JA. (2001). Role of estrogen receptor beta in estrogen action. Annual Reviews of Physiology 63: 165-169.

Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, Easton DF and Ponder BA. (2002). Polygenic susceptibility to breast cancer and implications for prevention. Nature Genetics 31: 33-26.

Plant AL, Benson DM and Smith LC. (1985). Cellular uptake and intracellular localization of benzo(a)pyrene by digital fluorescence imaging microscopy. Journal of Cell Biology 100: 1295-1308.

Raskin L, Pinchev M, Arad C, Lejbkowicz F, Tamir A, Rennert HS, *et al*. (2008). *FGFR2* is a breast cancer susceptibility gene in Jewish and Arab Israeli populations. Cancer Epidemiology Biomarkers and Prevention 17: 1060-1065.

Reardon W, Winter RM, Rutland P, Pulleyn LJ, Jones BM and Malcolm S. (1994). Mutations in the fibroblast growth factor receptor 2 gene cause Crouzon syndrome. Nature Genetics. 8: 98-103.

Rebbeck TR, DeMichele A, Tran TV, Panossian S, Bunin GR, Troxel AB, *et al*. (2009). Hormone-dependent effects of *FGFR2* and *MAP3K1* in breast cancer susceptibility in a population-based sample of post-menopausal African-American and European-American women. Carcinogenesis 30(2): 269-274.

Rebbeck TR. (2002). Inherited predisposition and breast cancer: modifiers of *BRCA*1/2-associated breast cancer risk. Environmental and Molecular Mutagenesis 39: 228-234.

Reeves MD, Yawitch TM, van der Merwe NC, van den Berg HJ, Dreyer G and van Rensburg EJ. (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.

Reynolds P, Hurley S, Goldberg D E, Anton-Culver H, Bernstein L, Deapen D, *et al*. (2004). Active smoking, household passive smoking, and breast cancer: evidence from the California Teachers Study. Journal of the National Cancer Institute 96: 29-37.

Roa BB, Boyd AA, Volcik K and Richards S. (1996). Ashkenazi Jewish population frequencies for common mutations in *BRCA1* and *BRCA2*. Nature Genetics 14: 185-187.

Ruffner H, and Verma I. (1997). BRCA1 is a cell-cycle regulated nuclear phosphoprotein. Proceedings of the National Academy of Science U.S.A 94: 7138-7143.

Russo J, Lynch H and Russo IH. (2001). Mammary gland architecture as a determining factor in the susceptibility of the human breast to cancer. Breast Journal 7: 278-291.

Russo J, Tay LK and Russo IH. (1982). Differentiation of the mammary gland and susceptibility to carcinogenesis. Breast Cancer Research and Treatment 2: 5-73.

Rutland P, Pulleyn LJ, Reardon W, Baraitser M, Hayward R, Jones B, *et al*. (1995). Identical mutations in the *FGFR2* gene cause both Pfeiffer and Crouzon syndrome phenotypes. Nature Genetics. 9: 173-176.

Ryan J, Scali J, Carrière I, Peres K, Rouaud O, Scarabin PY, *et al*. (2011). Oestrogen receptor polymorphisms and late-life depression. The British Journal of Psychiatry 199:126-131.

Sambrook J, Fritsch EF and Maniatis T. (1989). Molecular Cloning: A Laboratory Manual, vol. I. 2nd edition. Cold Spring Harbor Laboratory Press ISBN 0-87969-309-6.

Saurin AJ, Borden KLB, Boddy MN and Freemont PS. (1996). Does this have a familiar RING? Trends in Biochemical Sciences 21: 208-214.

Scully R, Chen J, Ochs RL, Keegan K, Hoekstra M, Feunsteun J, *et al*. (1997). Dynamic changes of *BRCA*1 subnuclear location and phosphorylation state are initiated by DNA damage. Cell 90: 425-435.

Scully R, Xie A and Nagaraju G. (2004). Molecular functions of BRCA1 in the DNA damage response. Cancer Biology & Therapy 3: 521-527.

Shin A, Kang D, Nishio H, Lee MJ, Park SK, Kim SU, *et al.* (2003). Estrogen receptor alpha gene polymorphisms and breast cancer risk. Breast Cancer Research and Treatment 80: 127-131.

Shiozaki EN, Gu L, Yan N, and Shi Y. (2004). Structure of the BRCT repeats of BRCA1 bound to a BACH1 phosphopeptide: implications for signaling. Molecular Cell 14: 405–412.

Shrivastav M,De Haro LP and Nickolof FJA. (2008). Regulation of DNA double-strand break repair pathway choice. Cell Research 18:134-147

Siddig A, Mohamed AO, Awad S, Hassan AH, Zilahi E, Al-Haj M, *et al*. (2008). Estrogen receptor α gene polymorphism and breast cancer. Annals of the New York Academic Sciences 1138: 95-107.

Silla CS and King MC. (1995). Inherited breast and ovarian cancer. Human Molecular Genetics 4: 1811-1817.

Simchoni S, Friedman E, Kaufman B, Gershoni-Baruch R, Orr-Urtreger A, Kedar-Barnes I, *et al.* (2006). Familial clustering of site-specific cancer risks associated with *BRCA1* and *BRCA2* mutations in the Ashkenazi Jewish population. Proceedings of the National Academy of Sciences U.S.A 103: 3770-3774.

Simoncini T, Rabkin E and Liao JK. (2003). Molecular basis of cell membrane estrogen receptor interaction with phosphatidylinositol 3-kinase in endothelial cells. Arteriosclerosis, Thrombosis and Vascular Biology 23(2): 198-203.

Simons AM, Horwitz AA, Starita LM, Griffin K, Williams RS, Glover JNM, *et al*. (2006). BRCA1 DNA-Binding activity is stimulated by BARD1. Cancer Research 66(4): 2012-2018.

Simpson ER. (2003). Sources of estrogen and their importance. Journal of the Steroid Biochemistry and Molecular Biology 86: 225-230.

Singletary KW and Gapstur SM. (2001). Alcohol and breast cancer: Review of epidemiologic and experimental evidence and potential mechanisms. Journal of the American Medical Association 286: 2143-2151.

Sivaraman VS, Wang H-y, Nuovo GJ and Malbon CC. (1997). Hyperexpression of mitogen-activated protein kinase in human breast cancer. Journal of Clinical Investigation 99: 1478-1483.

Smid M, Wang Y, Klijn JG, Sieuwerts AM, Zhang Y, Atkins D, *et al*. (2006). Genes associated with breast cancer metastatic to bone. Journal of Clinical Oncology 24: 2261-2267.

Sowers MR, Jannausch ML, McConnell DS, Kardia SR and Randolph JF. (2006). Endogenous estradiol and its association with estrogen receptor gene polymorphisms. American Journal of Medicine 119(9): S16-S22.

Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, *et al*. (2007). Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nature Genetics 39: 865-869.

Strachan T and Read A. (2003). Human Molecular Genetics. Taylor and Francis Group.

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

Streuwing JP, Hartge P, Wacholder S, Baker SM, Berlin M, McAdams M, *et al*. (1997). The risk of cancer associated with specific mutations of *BRCA*1 and *BRCA*2 among Ashkenazi Jews. New England Journal of Medicine 336(20): 1401-1408.

Struewing JP, Abeliovich D, Peretz T, Avishai N, Kaback MM, Collins S, *et al*. (1995). The carrier frequency of the *BRCA1* 185delAG mutation is approximately 1 percent in Ashkenazi Jewish individuals. Nature Genetics 11: 198-200.

Tamaru N, Hishikawa Y, Ejima K, Nagasue N, Inoue S, Muramatsu M, *et al*. (2004). Estrogen receptor-associated expression of keratinocyte growth factor and its possible role in the inhibition of apoptosis in human breast cancer. Laboratory Investigations 84: 1460-1471.

Tapper W, Hammond V, Gerty S, Ennis S, Simmonds P, Collins A, *et al*. (2008). The influence of genetic variation in 30 selected genes on the clinical characteristics of early onset breast cancer. Breast Cancer Research 10:R108.

Tavtigian SV, Simard J, Rommens J, Couch F, Shattuck-Eidens D, Neuhausen S, *et al.* (1996). The complete *BRCA*2 gene and mutations in chromosome 13q-linked kindreds. National Genetics 12: 333-337.

Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, *et al*. (2009). A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). Nature Genetics 41: 579–584.

Thompson D and Easton D. (2001). Variation in cancer risks, by mutation position, in *BRCA2* mutation carriers. American Journal of Human Genetics 68: 410-419.

Thompson D and Easton DF. (2002). Cancer incidence in *BRCA*1 mutation carriers. Journal of the National Cancer Institute 94: 1358-1365.

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

Thompson LH and Schild D. (2002). Recombinational DNA repair and human disease. Mutation Research 509(1-2): 49-78.

Thompson ME, Jensen RA, Obermiller PS, Page DL, and Holt JT. (1995). Decreased expression of *BRCA1* accelerates growth and is often present during sporadic breast cancer progression. Nature Genetics 9: 444-450.

Tipping AJ, Pearson T, Morgan NV, Gibson RA, Kuyt LP, Havenga C, *et al*. (2001). Molecular and genealogical evidence for a founder effect in Fanconi anemia families of the Afrikaner population of South Africa. Proceedings of the National Academy of Sciences U.S.A 98(10):5734-5739.

Travis RC, Reeves GK, Green J, Bull D, Tipper SJ, Baker K, *et al*. (2010). Gene-environment interactions in 7610 women with breast cancer: prospective evidence from the Million Women Study. Lancet 375(9732): 2143-2151.

Trikalinos TA, Salanti G, Khoury MJ and Ioannidis JP. (2006). Impact of violations and deviations in Hardy-Weinberg equilibrium on postulated gene-disease associations. American Journal of Epidemiology 163(4): 300-309.

Tryggvadottir L, Olafsdottir EJ, Gudlaugsdottir S, Thorlacius S, Jonasson JG, Tulinius H, *et al*. (2003). *BRCA2* mutation carriers, reproductive factors and breast cancer risk. Breast Cancer Research 5: R121-R128.

Tryggvadottir L, Sigvaldason H, Olafsdottir GH, Jonasson JG, Jonsson T, Tulinius H *et al*. (2006). Population-based study of changing breast cancer risk in Icelandic *BRCA2* mutation carriers, 1920–2000. Journal of the National Cancer Institute 98: 116-122.

Tsou HC, Teng DH, Ping XL, Brancolini V, Davis T, Hu R, *et al*. (1997). The role of *MMAC1* mutations in early-onset breast cancer: causative in association with Cowden syndrome and excluded in *BRCA*1-negative cases. American Journal of Human Genetics 61: 1036-1043.

Tutt A, Gabriel A, Bertwistle D, Connor F, Paterson H, Peacock J, *et al*. (1999). Absence of Brca2 causes genome instability by chromosome breakage and loss associated with centrosome amplification. Current Biology 9(19): 1107-1110.

Ursin G, Henderson BE, Haile RW, Pike MC, Zhou N, Diep A, *et al*. (1997). Does oral contraceptive use increase the risk of breast cancer in women with *BRCA1*/*BRCA2* mutations more than in other women? Cancer Research 57: 3678-3681.

van der Merwe NC, Hamel N, Schneider S-R, Apffelstaedt JP, Wijnen JT and Foulkes WD. (2011). A founder *BRCA2* mutation in non-Afrikaner breast cancer patients of the Western Cape of South Africa. Clinical Genetics. doi: 10.1111/j.1399-0004.2010.01617.x

Varghese JS and Easton DF. (2010). Genome-wide association studies in common cancers--what have we learnt? Current Opinion in Genetics & Development 20: 201-209.

Vatten LJ, Solvoll K and Løken EB. (1990). Coffee consumption and the risk of breast cancer. A prospective study of 14,593 Norwegian women. British Journal of Cancer 62(2): 267-270.

Vecchia C, Brinton LA and Mctiernan A. (2001). Hormone replacement therapy and breast cancer risk: epidemiology. Journal Für Menopause 8(2): 5-7.

Vehmanen P, Friedman LS, Eerola H, McClure M, Ward B, Sarantaus L, *et al.* (1997). Low proportion of BRCA1 and BRCA2 mutations in Finnish breast cancer families: evidence for additional susceptibility genes. Human Molecular Genetics 6: 2309-2315.

Venkitaraman AR. (2001). Functions of BRCA1 and BRCA2 in the biological response to DNA damage. Journal of Cell Science 114: 3591-3598.

Venkitaraman AR. (2002). Cancer susceptibility and their functions of BRCA1 and BRCA2. Cell 108: 171-182.

Vorobiof DA, Sitas F and Vorobiof G. (2001). Breast Cancer incidence in south africa. Journal of Clinical Oncology 19: 125s-127s.

Walter P, Green S, Greene G, Krust A, Bornert J-M, Jeltsch J-M, *et al*. (1985). Cloning of the human estrogen receptor cDNA. Proceedings of the National Academy of Sciences U.S.A 82: 7889-7893.

Wang J, Higuchi R, Modugno F, Li J, Umblas N, Lee J, *et al.* (2007). Estrogen receptor alpha haplotypes and breast cancer risk in older Caucasian women. Breast Cancer Research and Treatment 106: 273-280.

Wang W, Spurdle AB, Kolachana P, Bove B, Modan B, Ebbers SM, *et al*. (2001). A single nucleotide polymorphism in the 5′ untranslated region of *RAD51* risk of cancer among *BRCA1/2* mutation carriers. Cancer Epidemiology, Biomarkers and Prevention 10:421.

Wang X and Tournier C.(2006). Regulation of cellular functions by the ERK5 signalling pathway. Cell Signal 18(6): 753-760.

Wedrén S, Lovmar L, Humphreys K, Magnusson C, Melhus H, Syvanen A-C, *et al.* (2008). Estrogen receptor alpha gene polymorphism and endometrial cancer risk - a case-control study. BioMed Central Cancer 8: 322.

Weiderpass E, Pearson I, Melhus H, Wedren S, Kindmark A and Baron JA. (2000). Estrogen receptor alpha gene polymorphisms and endometrial cancer risk. Carcinogenesis 21: 623-627.

Welcsh P and King MC. (2001). BRCA1 and BRCA2 and the genetics of breast and ovarian cancer. Human Molecular Genetics 10: 705-713.

Welcsh PL, King MC and Owens KN. (2000). Insights into the functions of *BRCA*1 and *BRCA*2. Trends in Genetics 16(2): 69-74.

Wigginton JE, Cutler DJ, and Abecasis GR. (2005). A Note on Exact Tests of Hardy-Weinberg Equilibrium. American Journal of Human Genetics 76: 887-883.

Wilkie AOM, Slaney SF, Oldridge M, Poole MD, Ashworth GJ, Hockley A D, *et al*. (1995). Apert syndrome results from localized mutations of *FGFR2* and is allelic with Crouzon syndrome. Nature Genetics 9: 165-172.

Willett WC, Stampfer MJ, Brian C, Lipnick R, Speizer FE, Rosner B, *et al*. (1983). Cigarette smoking, relative weight and menopause. American Journal of Epidemiology 117: 651-658.

Williams RS, Chasman DI, Hau DD, Hui B, Lau AY and Glover JNM. (2003). Detection of protein folding defects caused by BRCA1-BRCT truncation and missense mutations. *Journal of Biological Chemistry* 278(52):53007-53016.

Wong AKC, Pero R, Ormonde PA, Tavtigian SV and Bartel PL. (1997). RAD51 interacts with the evolutionarily conserved BRC motifs in the human breast cancer susceptibility gene brca2*.* Journal of Biological Chemistry 272: 31941-31944.

Wooster R and Weber BL. (2003). Breast and ovarian cancer. New England Journal of Medicine 348(23): 2339-2347.

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

Wooster R, Neuhausen SL, Mangion J, Quirk Y, Ford D, Collins N, *et al.* (1994). Localization of a breast susceptibility gene, *BRCA*2, to chromosome 13q12-13. Science 265: 2088-2090.

Wu LC, Wang ZW, Tsan JT, Spillman MA, Phung A, Xu XL, *et al*. (1996). Identification of a RING protein that can interact in vivo with the *BRCA1* gene product. National Genetics 14: 430-440.

Xu X, Weaver Z, Linke SP, Li C, Gotay J, Wang XW, *et al*. (1999). Centrosome amplification and a defective G2-M cell cycle checkpoint induce genetic instability in BRCA1 exon 11 isoform-deficient cells. Molecular Cell*.* 3: 389-395.

Yaich L, Dupont WD, Cavener DR and Parl FF. (1992). Analysis of the *Pvu*II restriction fragment length polymorphism and exon structure of the estrogen receptor gene in breast cancer and peripheral blood. Cancer Research 52: 77-83.

Yamasaki E and Ames BN. (1977). Concentration of mutagens from urine by absorption with the nonpolar resin XAD-2: cigarette smokers have mutagenic urine. Proceedings of the National Academy Science U.S.A 74: 3555-3559.

Yang H, Jeffrey PD, Miller J, Kinnucan E, Sun Y, Thoma NH, *et al*. (2002). *BRCA*2 function in DNA binding and recombination from a *BRCA2*-DSS1-ssDNA structure. Science 297: 1837-1848.

Zeegers MPA, van Poppel F, Vlietinck R, Spruijt and Ostrer H. (2004). Founder mutations among the Dutch. European Journal of Human Genetics 12: 591-600.

Zhang W and Liu HT. (2002). MAPK signal pathways in the regulation of cell proliferation in mammalian cells. Cell Research 12: 9-18

Zheng W, Long J, Gao YT, Li C, Zheng Y, Xiang YB, *et al*. (2009). Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. Nature Genetics 41: 324–328.

Ziegler A, van Steen K and Wellek S. (2011). Investigating Hardy-Weinberg equilibrium in case-control or cohort studies or meta-analysis. Breast Cancer Research and Treatment 128:197-201.

6.2 Electronic references

- Arlequin [online]
  http://cmpg.unibe.ch/software/arlequin3, 2011.

- Breast cancer Information Core database (BIC) [online].
  Available at www.nhgri.nih.gov/bic, 2011

- Breast cancer linkage consortium (BCLC) [online].
  Available at http://www.humgen.nl/lab-devilee/bclchome.htm, 2010

- Chromas version 2.31 [online].
  Available at www.technelysium.com.au, 2011.

- South African National Cancer Registry [online].
  Available at
  http://www.cansa.org.za/cgibin/giga.cgi?cmd=cause_dir_news&cat=821&cause_id=1056, 2011.

- Online Mendelian Inheritance of Men (OMIM) [online].
  Available at www.ncbi.nlm.nih.gov/sites/entrez?db=omim, 2011.

  http://www.medicine.tcd.ie/neuropsychiatric-genetics/functional-genetics-genomics/genotyping.php

- LALIGN [online].
  Available at www.ch.embnet.org/software/LALIGNl, 2010.

- Vassar [online].

  Available at (http://faculty.vassar.edu/lowry/propcorr.html), 2010.

Chapter 7

Summary

The aim of the study was to elucidate the variation in phenotypic expression observed within *BRCA2* c.8162delG mutation positive families. The study attempted to identify possible genetic factors that contribute to the residual risk conferred by the *BRCA2* founder mutation. As BC is a polygenetic disorder, polymorphisms within various low penetrance genes may contribute to the expression of the disease. The selection of the SNPs were based on the results of the CIMBA consortium and have been proven to be associated with an increased BC risk in the general population (Easton *et al.*, 2007) and in *BRCA2* mutation carriers specifically (Antoniou *et al.*, 2008). Two SNPs (rs2234693 [*PvuII*] and rs9340799 [*XbaI*]) present within *ESR1* as well as SNPs present in *TNRC9* (rs3803662), *LSP1* (rs3817198), *MAP3K1* (rs889312) and *FGFR2* (rs2981582) identified by GWAS have been implicated in BC risk. These six polymorphisms have been selected to evaluate the risk within the Afrikaner *BRCA*2 8162delG (c.7934del, p.Arg2645AsnfsX3) mutation carriers specifically.

Genotyping of rs2234693 (*PvuII*) and rs9340799 (*XbaI*) was done by PCR-RFLP analysis whereas Taqman® assays were used for genotyping rs3803662 (*TNRC9*), rs3817198 (*LSP1*), rs889312 (*MAP3K1*) and rs2981582 (*FGFR2*). Automated allelic discrimination using the BioRad CFX Manager v1.1.308.1111 software were compared to manual discrimination methods to ensure robust genotyping. Cohen's kappa analysis suggested a combination of automated (Method 1) and manual (Method 3) genotyping was best suited for accurate allelic discrimination except for *LSP1*. Due to an putative SNP detected within *LSP1,* the validity of the *LSP1* results should be treated cautiously as no information on the frequency of the second putative SNP in white European individuals is available.

Of the six polymorphisms analyzed, only rs2234693 (*PvuII*), indicated a possible association with BC (*P*-value = 0.0896), which should be explored within a larger study group. For *FGFR2*, the HWE results indicated that the deviation observed in the *BRCA2* mutation carrier group could possibly be associated with BC. Haplotypes compiled for rs2234693 (*PvuII*) and rs9340799 (*XbaI*) as well as the remaining four SNPs were uninformative as it revealed no differences between the BC patients and the Cases. These results may have been due to the high allelic heterogeneity observed within the Afrikaner population, as well as the small test group used..

Although the results of this study did not deliver significant results, it did provide insight into allelic distributions of the SNPs in the Afrikaner *BRCA*2 8162delG (c.7934del, p.Arg2645AsnfsX3) mutation carriers specifically. Larger scale genotyping could lead to more significant findings to help elucidate the polygenetic nature of BC with the Afrikaner.

Keywords: familial breast cancer, genetic modifiers, SNPs, penetrance, haplotype, Taqman®, *ESR1*, *TNRC9*, *LSP1*, *MAP3K1*, *FGFR2*.

Chapter 8

Opsomming

Die doel van hierdie studie was om die variasie waargeneem in die fenotipiese uitdrukking onder *BRCA2* k.8162delG mutasie positiewe families toe te lig. Die studie het gepoog om verskeie genetiese faktore wat moontlik kan bydra tot die gesamentlike risiko wat toegeken word deur die *BRCA2* stigtersmutasie, te identifiseer. Aangesien borskanker 'n poligeniese siekte is, kan polimorfismes binne verskeie lae-penetrasie gene tot die uitdrukking van die siekte bydra. Die seleksie van die polimorfismes was gebaseer op die resultate van die CIMBA konsortium wat bewys het dat diè polimorfismes met verhoogde borskanker risiko in die algemene populasie (Easton *et al.*, 2007) sowel as in die *BRCA2* stigtersmutasie draers spesifiek geassosieerd is (Antoniou *et al.*, 2008). Twee polimorfismes (rs2234693 [*PvuII*] en rs9340799 [*XbaI*]) teenwoordig in *ESR1* asook die polimorfismes teenwoordig in *TNRC9* (rs3803662), *LSP1* (rs3817198), *MAP3K1* (rs889312) en *FGFR2* (rs2981582) wat deur GWAS geidentifiseer is, word met 'n verhoogde borskanker risiko geassosieer. Die ses polimorfismes is gekies om die addisionele risiko in die Afrikaner *BRCA2* k. 8162delG (k.7934del, p.Arg2645AsnfsX3) mutasie draers spesifiek te ondersoek.

Genotipering van rs2234693 (*PvuII*) en rs9340799 (*XbaI*) is uitgevoer met behulp van ensiem snydings (RFLP), terwyl Taqman® analises gebruik is om rs3803662 (*TNRC9*), rs3817198 (*LSP1*), rs889312 (*MAP3K1*) en rs2981582 (*FGFR2*) te genotipeer. Outomatiese alleliese diskriminasie gedoen deur die BioRad CFX Manager v1.1.308.1111 sagteware is vergelyk met semi- en nie-outomatiese diskriminasie metodes om sodoende robuuste genotipering te verseker. Cohen se kappa analises het die datastelle vergelyk en aangedui dat die metode van analise van Metode 1 en Metode 3 die meeste ooreenstem, met *LSP1* as die uitsondering. Die teenwoordigheid van 'n addisionele polimorfisme binne dieselfde gebied, impliseer dat die data ingewin vir *LSP1* met versigtigheid geïnterpreteer moet word. Geen inligting rakende die frekwensie van hierdie polimorfisme was vir die Europese individue beskikbaar nie.

Van die ses polimorfismes geanaliseer, het slegs rs2234693 (*PvuII*) 'n moontlike assosiasie met borskanker getoon (*P*-waarde = 0.0896) wat in 'n toekomstige groter studie verder ondersoek moet word. Die HWE resultate het aangedui dat die afwyking waargeneem vir *FGFR2* in die *BRCA2* mutasie draer groep, moontlike assosiasie met borskanker kan beteken. Saamgestelde haplotipes vir rs2234693 (*PvuII*) en rs9340799 (*XbaI*) asook die oorblywende vier polimorfismes was oninsiggewend omdat daar geen verkil tussen die borskanker pasiënte en gevalle opgemerk is nie. Die bevindinge kan moontlik die gevolg wees van die hoë alleliese variasie waargeneem in die Afrikaner populasie, sowel as die klein toetsgroep wat gebruik is.

Alhoewel die bevindinge van die studie nie statisties betekenisvolle resultate opgelewer het nie, dui dit die alleliese verpreiding van die polimorfimses in die Afrikaner *BRCA2* k.8162delG (k.7934del, p.Arg2645AsnfsX3) mutasie draers aan. Grootskaalse genotipering kan lei tot meer insiggewende bevindige om te help om die poligeniese aard van borskanker in die Afrikaner te ontrafel.

Sleutelwoorde: oorerflike borskanker, genetiese veranderlikes, polimorfismes, penetrasie, haplotipe, Taqman®, *ESR1*, *TNRC9*, *LSP1*, *MAP3K1*, *FGFR2*.

Appendix A

FREE STATE PROVINCE

Ref. no.: H4/3/2

20 March 2006

Dr. NC van der Merwe
Principal Investigator
Division of Human Genetics
Faculty of Health Sciences
University of the Free State
BLOEMFONTEIN
9300

Dear Dr. van der Merwe

PERMISSION FOR USING LABORATORY SPACE AND BLOOD SAMPLES FROM CLINIC PATIENTS

Your letter dated 10 March 2006 in the abovementioned refers.

Herewith permission for the mentioned project to be done at Universitas Academic Hospital on condition that approval is obtained from the Ethics Committee.

No findings can be published without permission of the Chief Executive Officer.

Yours sincerely

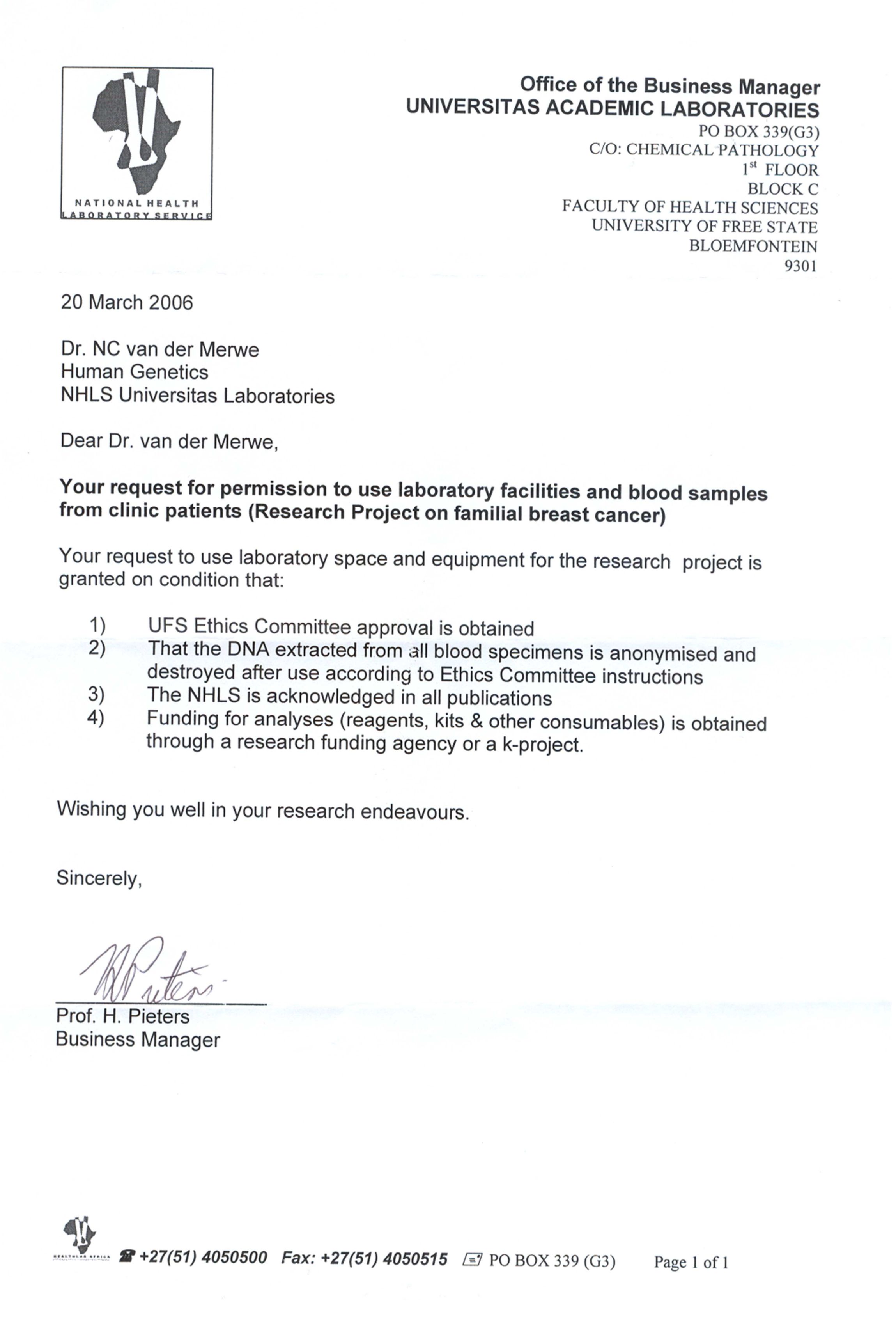
DR NIC R J VAN ZYL
HEAD: CLINICAL SERVICES

DR NRJ VAN ZYL
2006 -03- 2 0
HEAD: CLINICAL SERVICES
UNIVERSITAS ACADEMIC HOSPITAL

Department of Health · Departement van Gesondheid · Lefapha La Bophelo Bo Botle

Head: Clinical Services, Dr. NRJ van Zyl, Universitas Tertiary Hospital • P/Bag X20660, Bloemfontein 9300 •
Tel: 051-405 2866 • Fax: 051-444 0792, e-mail: vanzylnr@fshealth.gov.za • Room 1129, 1st Floor,
Universitas Tertiary Hospital, Bloemfontein

Appendix B



NATIONAL HEALTH
LABORATORY SERVICE

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 STATE
BLOEMFONTEIN
9301

20 March 2006

Dr. NC van der Merwe
Human Genetics
NHLS Universitas Laboratories

Dear Dr. van der Merwe,

Your request for permission to use laboratory facilities and blood samples from clinic patients (Research Project on familial breast cancer)

Your request to use laboratory space and equipment for the research project is granted on condition that:

1) UFS Ethics Committee approval is obtained
2) That the DNA extracted from all blood specimens is anonymised and destroyed after use according to Ethics Committee instructions
3) The NHLS is acknowledged in all publications
4) Funding for analyses (reagents, kits & other consumables) is obtained through a research funding agency or a k-project.

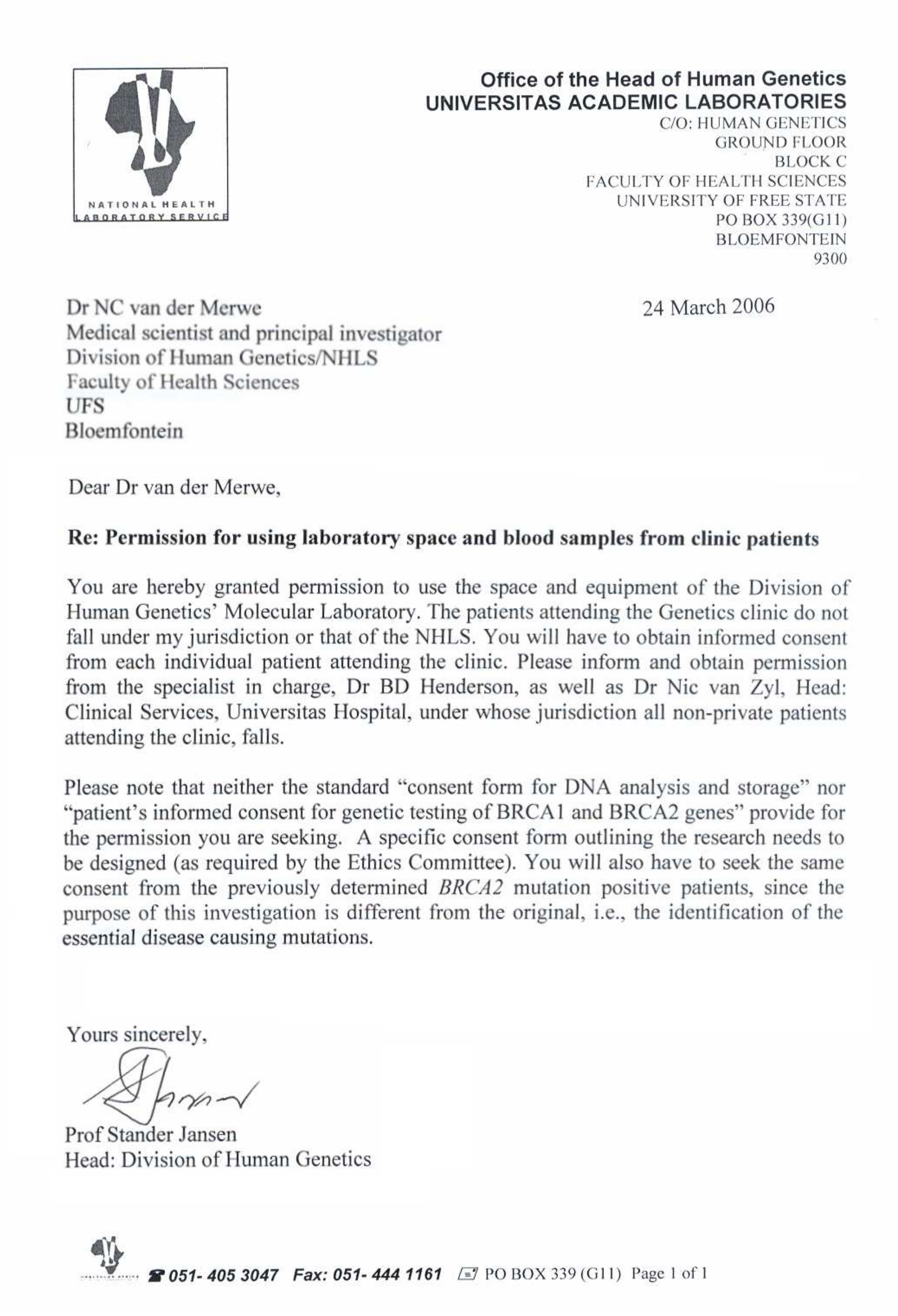
Wishing you well in your research endeavours.

Sincerely,

Prof. H. Pieters
Business Manager

☎ +27(51) 4050500 Fax: +27(51) 4050515 PO BOX 339 (G3) Page 1 of 1

Appendix C



NATIONAL HEALTH LABORATORY SERVICE

Office of the Head of Human Genetics
UNIVERSITAS ACADEMIC LABORATORIES
C/O: HUMAN GENETICS
GROUND FLOOR
BLOCK C
FACULTY OF HEALTH SCIENCES
UNIVERSITY OF FREE STATE
PO BOX 339(G11)
BLOEMFONTEIN
9300

24 March 2006

Dr NC van der Merwe
Medical scientist and principal investigator
Division of Human Genetics/NHLS
Faculty of Health Sciences
UFS
Bloemfontein

Dear Dr van der Merwe,

Re: Permission for using laboratory space and blood samples from clinic patients

You are hereby granted permission to use the space and equipment of the Division of Human Genetics' Molecular Laboratory. The patients attending the Genetics clinic do not fall under my jurisdiction or that of the NHLS. You will have to obtain informed consent from each individual patient attending the clinic. Please inform and obtain permission from the specialist in charge, Dr BD Henderson, as well as Dr Nic van Zyl, Head: Clinical Services, Universitas Hospital, under whose jurisdiction all non-private patients attending the clinic, falls.

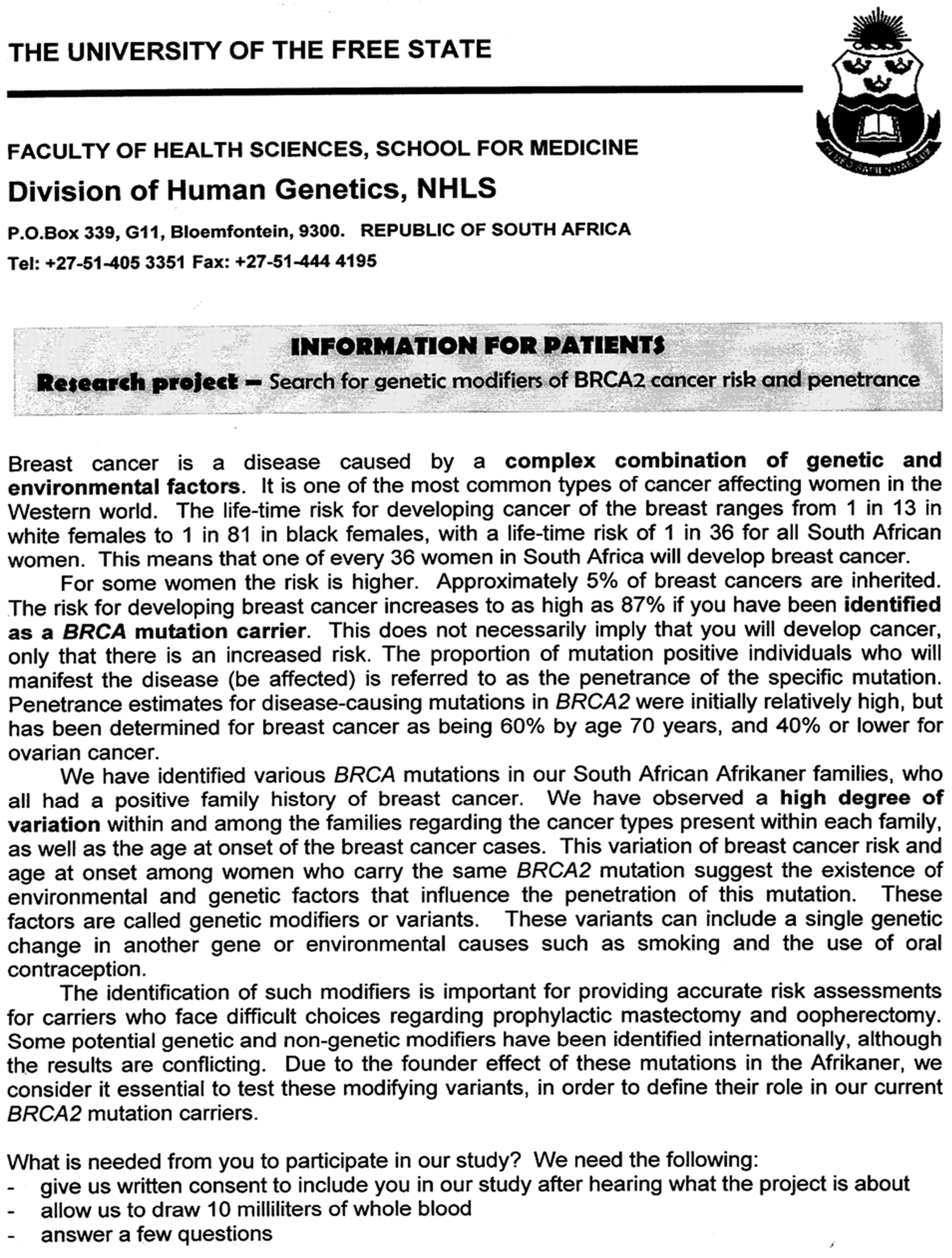
Please note that neither the standard "consent form for DNA analysis and storage" nor "patient's informed consent for genetic testing of BRCA1 and BRCA2 genes" provide for the permission you are seeking. A specific consent form outlining the research needs to be designed (as required by the Ethics Committee). You will also have to seek the same consent from the previously determined *BRCA2* mutation positive patients, since the purpose of this investigation is different from the original, i.e., the identification of the essential disease causing mutations.

Yours sincerely,

Prof Stander Jansen
Head: Division of Human Genetics

☎ ***051- 405 3047 Fax: 051- 444 1161*** 🖂 PO BOX 339 (G11) Page 1 of 1

Appendix D



THE UNIVERSITY OF THE FREE STATE

FACULTY OF HEALTH SCIENCES, SCHOOL FOR MEDICINE

Division of Human Genetics, NHLS

P.O.Box 339, G11, Bloemfontein, 9300. REPUBLIC OF SOUTH AFRICA

Tel: +27-51-405 3351 Fax: +27-51-444 4195

INFORMATION FOR PATIENTS

Research project – Search for genetic modifiers of BRCA2 cancer risk and penetrance

Breast cancer is a disease caused by a **complex combination of genetic and environmental factors**. It is one of the most common types of cancer affecting women in the Western world. The life-time risk for developing cancer of the breast ranges from 1 in 13 in white females to 1 in 81 in black females, with a life-time risk of 1 in 36 for all South African women. This means that one of every 36 women in South Africa will develop breast cancer.

For some women the risk is higher. Approximately 5% of breast cancers are inherited. The risk for developing breast cancer increases to as high as 87% if you have been **identified as a *BRCA* mutation carrier**. This does not necessarily imply that you will develop cancer, only that there is an increased risk. The proportion of mutation positive individuals who will manifest the disease (be affected) is referred to as the penetrance of the specific mutation. Penetrance estimates for disease-causing mutations in *BRCA2* were initially relatively high, but has been determined for breast cancer as being 60% by age 70 years, and 40% or lower for ovarian cancer.

We have identified various *BRCA* mutations in our South African Afrikaner families, who all had a positive family history of breast cancer. We have observed a **high degree of variation** within and among the families regarding the cancer types present within each family, as well as the age at onset of the breast cancer cases. This variation of breast cancer risk and age at onset among women who carry the same *BRCA2* mutation suggest the existence of environmental and genetic factors that influence the penetration of this mutation. These factors are called genetic modifiers or variants. These variants can include a single genetic change in another gene or environmental causes such as smoking and the use of oral contraception.

The identification of such modifiers is important for providing accurate risk assessments for carriers who face difficult choices regarding prophylactic mastectomy and oopherectomy. Some potential genetic and non-genetic modifiers have been identified internationally, although the results are conflicting. Due to the founder effect of these mutations in the Afrikaner, we consider it essential to test these modifying variants, in order to define their role in our current *BRCA2* mutation carriers.

What is needed from you to participate in our study? We need the following:

- give us written consent to include you in our study after hearing what the project is about
- allow us to draw 10 milliliters of whole blood
- answer a few questions

UNIVERSITEIT VAN DIE VRYSTAAT

FAKULTEIT GESONDHEIDSWETENSKAPPE, MEDIESE SKOOL
Afdeling Mensgenetika, NHLS
Posbus 339, G11, Bloemfontein, 9300. SUID-AFRIKA
Tel: +27-51-405 3351 Faks: +27-51-444 4195

Geagte pasiënt **01/06/06**

Bekendstelling van 'n verdere navorsingsprojek op oorerflike borskanker

Ons weet dat borskanker veroorsaak word deur 'n komplekse kombinasie van genetiese- en omgewingsfaktore en dat dit een van die algemeenste kanker tipes in vroue van die Westerse wêreld is. Die borskanker risiko varieer van 1 in 13 vir blanke vrouens, tot 1 in 81 vir swart vrouens, met 'n gemiddelde van 1 in 36 vir die Suid-Afrikaanse vrou.

Vir sekere vroue soos u is die risiko heelwat hoër (tot so hoog as 87% in *BRCA* mutasie draers). U het reeds op diagnostiese vlak kliniese raadgewing ontvang en is deur ons laboratorium getoets vir die teenwoordigheid van siekte-veroorsakende mutasies in die oorerflike borskanker gene, *BRCA1* en *BRCA2.* Die genetiese toetse het daarop gedui dat u wel 'n *BRCA2* mutasie draer is.

Ons benodig u hulp. Ons vind dat positiewe mutasie draers binne 'n spesifieke familie nie noodwendig op dieselfde ouderdom siek word nie, of dat party nooit aantasting toon nie, selfs al is hulle reeds op 'n gevorderde ouderdom. Met ander woorde die penetrasie van hierdie spesifieke mutasie varieer tussen individue binne dieselfde familie en selfs tussen verskillende families. Wanneer daar verwys word na die penetrasie van 'n mutasie, word die aantal mutasie draers wat wel die siekte ontwikkel gereflekteer in die persentasie. Internasionaal word aanvaar dat die penetrasie van mutasies in *BRCA2* redelik hoog is, naamlik 60% vir borskanker teen die ouderdom 70.

Die variasie tussen en binne een familie is 'n aanduiding dat ander faktore (geneties sowel as omgewingsfaktore) ook moontlik 'n rol kan speel in die uiteindelike ontwikkeling van borskanker in 'n mutasie draer, m.a.w. op die penetrasie van die mutasie. Hierdie onbekende of modifiserende faktore kan enige iets insluit, van 'n genetiese polimorfisme in 'n ander geen tot 'n omgewingsfaktor soos rookgewoontes of die gebruik van orale voorbehoed.

Die identifisering van hierdie modifiserende faktore is baie belangrik aangesien dit die risiko-bepaling van mutasie draers meer akkuraat kan maak. Dit sal u, die pasiënt in staat stel tot beter besluitneming rakende voorkomende behandeling soos profilaktiese mastektomie en ovarektomie. Internasionale navorsers het reeds 'n paar moontlike modifiserende faktore uitgewys, alhoewel die resultate baie kontrasterend is vir die verskillende bevolkingsgroepe en nasionaliteite. Ons voel dat die navorsing van waarde kan wees vir die Afrikaner bevolking, omdat dit so ‘n unieke populasie is wat stigterseffekte toon.

Wat word van u benodig om deel te wees van hierdie studie? Ons benodig die volgende:

- ingeligte toestemming om u deel te maak van die studie (sal geskied nadat u breedvoerig oor die studie ingelig is)
- ‘n nuwe/vars bloedmonster (10 milliliter bloed vir DNA ekstraksie)
- antwoorde op ’n algemene vraelys rakende hormoongebruik, obstetries en reproduktiewe geskiedenis (sal later gedoen word)

Daar sal geen kostes verbonde wees aan u deelname nie, aangesien dit ‘n navorsingsprojek is. Indien ons wel betekenisvolle antwoorde tydens die uitvoer van die studie kry, sal dit aan u bekend gemaak word via u huisdokter of ons eie genetiese raadgewer. Indien daar ander vroulike familielede binne u onmiddelike familie is wat onlangs gediagnoseer is met borskanker waarvan ons nie weet nie, sal ons dit hoog op prys stel as u hulle kontak besonderhede aan ons kan deurgee. Ons is bereid om hulle kosteloos te toets vir die siekte-veroorsakende mutasie. Ons stel ook belang in onaangetaste positiewe vroulike mutasie-draers. Hulle vorm ook deel van ons projek.

Ek hoop dat u ons versoek gunstig sal oorweeg, aangesien dit ook vir u eie kinders en die nageslag voordeel kan inhou. Die mediese veld as geheel kan by u deelname baat vind. Ons uiteindelike doel is om die risiko bepaling en diagnostiese toetsing van hierdie spesifieke mutasie te verfyn. Dit sal ons instaat stel om u vrae beter te beantwoord en meer sinvolle siftingsmetodes of voorkomende prosedures voor te stel.

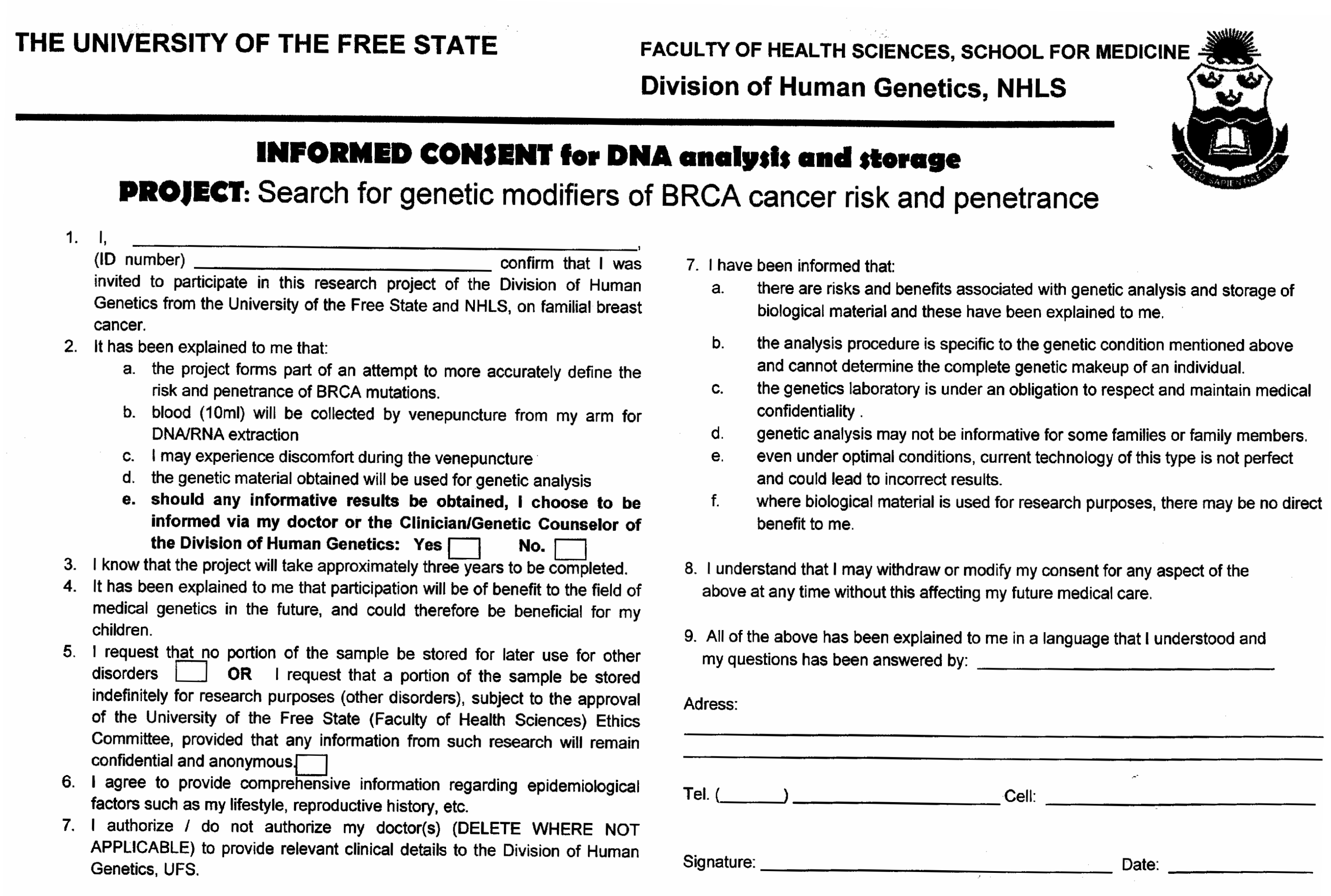
Ek gaan oor twee weke telefonies met u kontak maak om u antwoord te verneem. Indien u wel wil deelneem, sal ons dan reëlings tref vir die inligtings-gesprek, die teken van die ingeligte toestemmingsbrief en die trek van die bloed.

Baie dankie by voorbaat.

Die uwe

Dr Nerina van der Merwe
Hoof Navorser
e-mail: gnmgncv.MD@mail.uovs.ac.za
Tel. (051) 405 3351

Appendix E



THE UNIVERSITY OF THE FREE STATE

FACULTY OF HEALTH SCIENCES, SCHOOL FOR MEDICINE
Division of Human Genetics, NHLS

INFORMED CONSENT for DNA analysis and storage

PROJECT: Search for genetic modifiers of BRCA cancer risk and penetrance

1. I, ____,
(ID number) ______________________________ confirm that I was invited to participate in this research project of the Division of Human Genetics from the University of the Free State and NHLS, on familial breast cancer.
2. It has been explained to me that:
    a. the project forms part of an attempt to more accurately define the risk and penetrance of BRCA mutations.
    b. blood (10ml) will be collected by venepuncture from my arm for DNA/RNA extraction
    c. I may experience discomfort during the venepuncture
    d. the genetic material obtained will be used for genetic analysis
    e. **should any informative results be obtained, I choose to be informed via my doctor or the Clinician/Genetic Counselor of the Division of Human Genetics: Yes ☐ No. ☐**
3. I know that the project will take approximately three years to be completed.
4. It has been explained to me that participation will be of benefit to the field of medical genetics in the future, and could therefore be beneficial for my children.
5. I request that no portion of the sample be stored for later use for other disorders ☐ **OR** I request that a portion of the sample be stored indefinitely for research purposes (other disorders), subject to the approval of the University of the Free State (Faculty of Health Sciences) Ethics Committee, provided that any information from such research will remain confidential and anonymous.☐
6. I agree to provide comprehensive information regarding epidemiological factors such as my lifestyle, reproductive history, etc.
7. I authorize / do not authorize my doctor(s) (DELETE WHERE NOT APPLICABLE) to provide relevant clinical details to the Division of Human Genetics, UFS.

7. I have been informed that:
    a. there are risks and benefits associated with genetic analysis and storage of biological material and these 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. even under optimal conditions, current technology of this type is not perfect and could lead to incorrect results.
    f. where biological material is used for research purposes, there may be no direct benefit to me.

8. 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.

9. All of the above has been explained to me in a language that I understood and my questions has been answered by: ______________________________

Adress:
____
____

Tel. (______) ____________________ Cell: __________________________

Signature: __________________________________ Date: ______________

UNIVERSITEIT VAN DIE VRYSTAAT FAKULTEIT GESONDHEIDSWETENSKAPPE, SKOOL VIR GENEESKUNDE
Afdeling Mensgenetika, NHLS

INGELIGTE TOESTEMMING vir ondersoeke op en die bewaar van DNA

NAVORSINGSPROJEK: – Ondersoek van genetiese polimorfismes in ander gene wat die penetrasie en risiko van 'n BRCA2 siekte-veroorsakende mutasie kan beinvloed

1. Ek, ________________________________,
   (ID nommer) ____________________ bevestig dat ek uitgenooi is om deel te neem aan bogenemelde navorsingsprojek van die Afdeling Mensgenetika (Universiteit van die Vrystaat/NHLS) op oorerflike borskanker.
2. Daar is aan my verduidelik dat:
   a. die projek deel uitmaak van 'n studie waarin daar gepoog sal word om die risiko en penetrasie van die Suid-Afrikaanse BRCA mutasies meer akkuraat te bepaal
   b. bloed (10ml) uit die voorarm versamel sal word vir DNA/RNA ekstraksie
   c. die proses van bloedtrek effense ongemak mag meebring
   d. die genetiese materiaal gebruik sal word vir genetiese analises
   e. indien enige betekenisvolle resultate verkry word, ek daarvan in kennis gestel sal word via my huisdokter of die Genetiese raadgewer van die Afdeling Mensgenetika: Ja ☐ Nee ☐
3. Ek weet dat die projek ongeveer drie jaar sal neem om te voltooi.
4. Daar is verder aan my verduidelik dat deelname aan die projek kan bydra tot die uitbouing van mediese kennis, wat moontlik voordele vir my kinders kan inhou.
5. Ek versoek dat geen gedeelte van die monster vir latere gebruik geberg mag word nie ☐ **OF** ek versoek dat 'n gedeelte van die monster onbepaald geberg mag word vir navorsingsdoeleindes onderhewig aan die goedkeuring van die Etiekkomitee van die Fakulteit Gesondheidswetenskappe, Universiteit van die Vrystaat, op voorwaarde dat enige inligting voortspruitend uit hierdie navorsing vertroulik en anoniem sal bly. ☐
6. Ek stem in om so volledig moontlik inligting rakende epidemiologiese faktore soos lewenswyse, reproduktiewe geskiedenis, ens. te verskaf.
7. Ek magtig / magtig nie my geneesheer(e) (SKRAP WAT NIE VAN TOEPASSING IS NIE) om toepaslike kliniese besonderhede aan die Afdeling Mensgenetika, UV beskikbaar te stel.
8. Ek is ingelig dat:
   a. daar risiko's en voordele verbonde is aan die genetiese ondersoeke en berging van biologiese material en dat dit aan my verduidelik is
   b. die ondersoekprosedure spesifiek is tot die oorerflike toestand hierbo gemeld en nie die volledige genetiese samestelling van 'n individue kan bepaal nie
   c. die genetika laboratorium verplig is om mediese vertroulikheid te respekteer en te handhaaf.
   d. genetiese ondersoeke nie vir alle families of familielede antwoorde kan verskaf nie
   e. selfs onder optimale toestande die bestaande tegnologie feilbaar is en tot foutiewe resultate kan lei
   f. waar biologiese materiaal vir navorsingsdoeleindes aangewend word, dit geen direkte voordele vir my mag inhou nie.
9. Ek verstaan dat ek te enige tyd my toestemming mag terugtrek of wysig sonder inbreuk op my toekomstige mediese sorg.
10. Al die bogenoemde aspekte is aan my verduidelik in 'n taal wat ek verstaan en my vrae is beantwoord deur: ____**Dr NC van der Merwe**________

Adres: ____

____

Tel. (______) ____________________ Sel: ____________________

Handtekening: ______________________________ Datum: __________