UiT
THE ARCTIC UNIVERSITY OF NORWAY

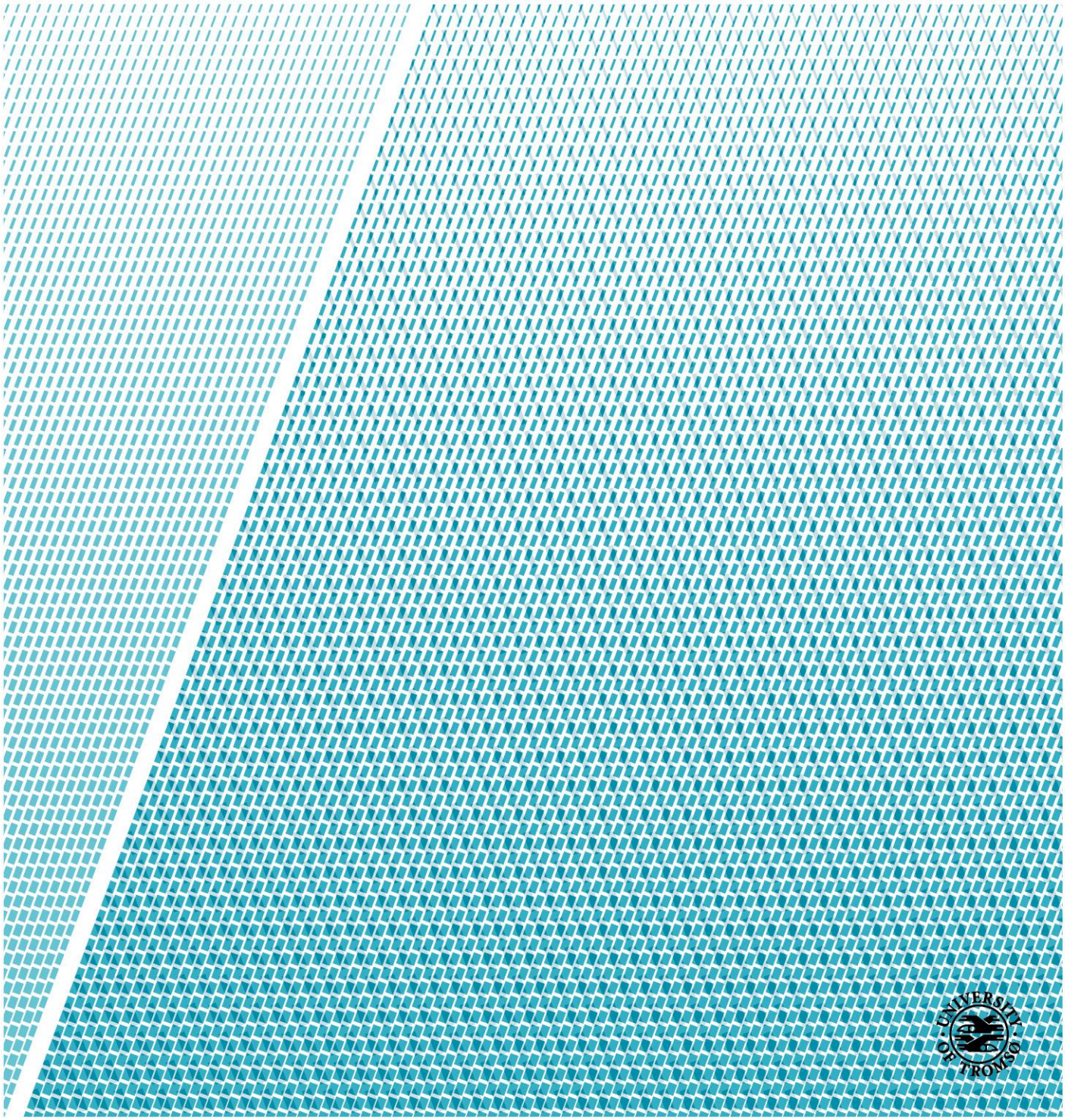
The Faculty of Health Sciences
The Department of Medical Biology
The Medical Genetic Department, University Hospital in Northern Norway

Hereditary breast and ovarian cancer

a. *Diversity of genetic causes of HBOC in a Norwegian breast and ovarian cancer patient cohort*
b. *BRCA2 c.8331+2C>T – a Norwegian founder mutation*

Siri Hermansen Skarsfjord

MBI-3911 Master's thesis in Biomedicine May 2017



Acknowledgements

The experimental work for this master's thesis was performed at the Medical Genetics department at the University Hospital in Northern Norway (UNN) in Tromsø from August 2016 to May 2017, and markers the end of two years of Master in Biomedicine at UiT – The Artic University of Norway.

First of all I would like to express my gratitude to my three supervisors Marijke Van Gehlue, Elisabeth Jarhelle and Hilde Monica Frostad Riise Stensland for their outstanding guidance, knowledge and feedback both during the experimental work and during the writing process of this thesis.

I would also like to thank my co-supervisor Ugo Lionel Moens from the department of medical biology, UiT, without whom it would not have been possible to work on my master's thesis at UNN.

I would like to thank all the members of the medical genetics department for all help during the experimental work, as well as a good working environment.

I would also like to mention my office-mates, Aud-Malin and Lotte, for lunch- and coffee breaks, often combined with knitting, as well as encouraging conversations during these two years.

Last, but not least, I would like to thank my family for their patience and support with encouraging phone calls and mailed packages.

Siri Hermansen Skarsfjord

Tromsø, May 2017

Index

Summary

Hereditary breast and ovarian cancer (HBOC) causes 5-10 % of breast cancer cases and 25 % of ovarian cancer cases. About 24 % of HBOC are caused by deleterious variants in *BRCA1* and *BRCA2*. Currently, more than 25 different genes have been associated with HBOC, including *BRCA1* and *BRCA2*, many of which encode proteins participating in homologous recombination repair (HRR) and mismatch repair (MMR).

In one part of this study, 16 genes associated with HBOC were scrutinised using next generation sequencing (NGS) on 48 patient samples where no deleterious variants or variants of unknown clinical significance (VUS's) had previously been found in *BRCA1* or *BRCA2*. Among five of the 48 patients included in this study, three different deleterious variants were identified including: *ATM* c.3245_3247delinsTGAT, *TP53* c.818G>A and *CHEK2* c.319+2T>A. In addition, eight different VUS's in 5 different genes were identified in *ATM*, *BRIP1*, *MLH1*, *NF1* and *PMS2*.

In the second part of this study, the *BRCA2* c.8331+2C>T variant, which has been identified in 29 families in Norway, was found to cause skipping of exon 18. Both the high frequency of the *BRCA2* c.8331+2C>T variant in Norwegian breast cancer families and the currently conducted microsatellite analysis with markers in close proximity to the *BRCA2* gene, indicated that this variant is a Norwegian founder mutation.The variant was estimated to be 97-215 generations old.

Abbreviations

aa	amino acid	dNTP	deoxynucleotide
Ala	Alanine	DSB	Double stranded break
APRT	Adenine phosphoribosyl transferase	dsDNA	double stranded DNA
Arg	Arginine	EDTA	ethylenediaminetetraacetic acid
Asn	Asparagine	EE1	enrichment elution buffer 1
Asp	Aspartate	EHB	Enrichment hybridization Buffer
A-T	Ataxia-telangiectasia	ESE	Exonic Splicing Enhancer
ATM	Ataxia-telangiectasia mutated	ESP	EVS/Exome sequencing project
ATPase	Adenosine thriphosphatase	EtOH	ethanol
BIC	Breast cancer information core	ET2	elute target buffer 2
bp	base pair	EWS	Enrichment wash solution
BRCA1	Breast cancer 1	ExAC	Exome aggregation consortium
BRCA2	Breast cancer 2	gDNA	genomic DNA
CCD	Charge-coupled device	G_1 phase	gap 1 phase
cDNA	complementary DNA	G_2 phase	gap 2 phase
CHEK2/CHK2	Checkpoint kinase 2	Gln	glutamine
CLL	chronic lymphocytic leukaemia	Glu	glutamic acid
COSMIC	catalogue of somatic mutations in cancer	Gly	glycine
CRC	colorectal cancer	gnomAD	The genome Aggregation Database
CSO	Costume selected oligoes	GVGD	Grantham variation/Grantham deviation
Cys	Cysteine	HBOC	Hereditary breast and ovarian cancer
dbSNP	Single Nucleotide Polymorphism Database	HDGC	Hereditary diffuse gastric cancer
ddNTP	dideoxynucleotide	HGMD	Human gene mutation database
DMLE	Disease mapping using linkage disequilibrium	HGNC	Hugo gene nomenclature committee
DNA	Deoxyribonucleic acid	HGVS	Human genome variation society
His	Histidine	NLS	Nuclear Localization Site

HNPCC	Hereditary nonpolyposis colorectal cancer	nM	nano Molar
HRR	Homologous recombination repair	NMD	nonsense-mediated mRNA decay
HS	High sensitivity	NNSPLICE	Neural Network Splice Site Prediction
HSF	Human Splice Finder	OD	Oligomerization domain
IGV	Integrative genomics viewer	p53	p53 protein
Ile	Isoleucine	PALB2	Partner and localizer of BRCA2
K	Kilo	PCR	Polymerase chain reaction
kb	kilo base pair	Phe	phenylalanine
Leu	Leucine	PI3-kinase	Phosphoinositide 3-kinase
LOVD	Leiden open variation database	PKC	Protein kinase C
LS	Lynch syndrome	PMS2	PMS1 homologue 2
M	Morgans	PMS2CL	PMS2 C-terminal-like pseudogene
MES	MaxEntScan	PoliP	Proline rich domain
Met	Methionine	PolyPhen-2	Polymorphism Phenotyping version 2
Min	Minutes	Pro	proline
MLH1	mutL homologue 1	PTEN	phosphate and tensin homolog
MMR	Mismatch repair	RAD51C	RAD51 paralog C
M phase	mitotic phase	RAD51D	RAD51 paralog D
MRE11	Meiotic recombination 11	REK	Regional Ethical Committee
MRN complex	MRE11, RA50 and NBN complex	RESCUE-ESE	Relative Enhancer and Silencer Classification by Unanimous Enrichment- ESE
mRNA	messenger RNA	RNA	Ribonucleic acid
MSH2	mutS homologue 2	RSB	Resuspension Buffer
MSH6	mutS homologue 6	RT	room temperature
NCBI	National Center for Biotechnology	SBS	Substrate binding site
NEG	Negative regulation domain	Sec	seconds
NGS	Next generation sequencing	Ser	serine
SIFT	Sorting intolerant from tolerant		

SMB	Streptavidin magnetic beads
SNP	single nucleotide polymorphism
SP	Sample preparation
SPB	Sample purification beads
S phase	DNA synthesis phase
ssDNA	single stranded DNA
SSF	Splice Site Finder
TAD1	Transactivation domain 1
TAD2	Transactivation domain 2
TAE-buffer	Tris-acetate-EDTA buffer
Thr	Threonine
TP53	Tumour protein p53
ssDNA	single stranded DNA
UV-light	Ultraviolet light
UVP	Ultraviolet Production Ltd.
Val	Valine
VUS	Variant of unknown clinical significance

1 Introduction

Cancer (from the Latin word for crab) is a group of diseases which is caused by abnormal cell growth. The abnormal cell growth arises as a result of deleterious variants in specific genes in the cell genome. These deleterious variants can give the cell novel and often abnormal phenotypes. Mutations of genes whose products are involved in cellular proliferation are the main reason for cancer development, since pathogenic mutations in these genes can cause the cell to proliferate into large populations of cells that no longer act as the cells around them [1].

Uncontrolled cell proliferation can arise from many specialized cell types throughout the body. Accordingly, there are currently more than 100 types of different cancers described. The cancers may be classified by their primary site of origin or by their histology or tissue types. The main groups of cancers, depending from which cell types they develop, are carcinomas (malignancies that originate from epithelial cells present in skin or tissue that cover internal organs), sarcomas (malignancies that originate from mesenchymal cells present in connective or supportive tissue such as bone, cartilage, fat, muscle or blood vessels), leukaemia (cancers that originates in blood forming tissue), lymphoma and myeloma (cancers that originates in cells of the immune system) and neuroectodermal tumours (central and peripheral nervous system originated cancers) [1]. Tumours can be either benign or malignant. A benign tumour grows in a confined specific site in the tissue, and does not invade adjacent tissue. Malignant tumours on the other hand, are locally invasive and can possibly spread to other parts of the organism (metastasis).

On a genetic level the genes which are most often the reason for cancer development can be divided in two groups; proto-oncogenes and tumour suppressor genes. Both groups of genes can be expressed in normal cells and are part of the cellular growth-control pathways [2]. Proto-oncogenes can cause cancer when the genes are mutated resulting in a gain-of-function of the encoded proteins. Such mutations convert the proto-oncogenes into oncogenes. Proto-oncogenes promote cellular growth and only one defect allele is needed for cancer development. Accordingly, the oncogenes act dominantly [2]. Tumour suppressor gene products usually inhibit cell proliferation. Only one functional allele is required in order for the inhibition to take place. Cancer development from defects in these genes therefore acts

recessively on the cellular level. Cancers caused by defects in tumour suppressor genes are the most common form of hereditary cancer [1].

Since tumour suppressor genes are recessive on a cellular level, both alleles have to be mutated in order for cancer to develop. Individuals with a familial mutation already have one defective allele, which means that only one mutational event leading to loss-of-function of the second allele is needed for development of cancer (Figure 1) [1, 2]. This "two hit" hypothesis was first formulated by Knudson in 1971 [3].

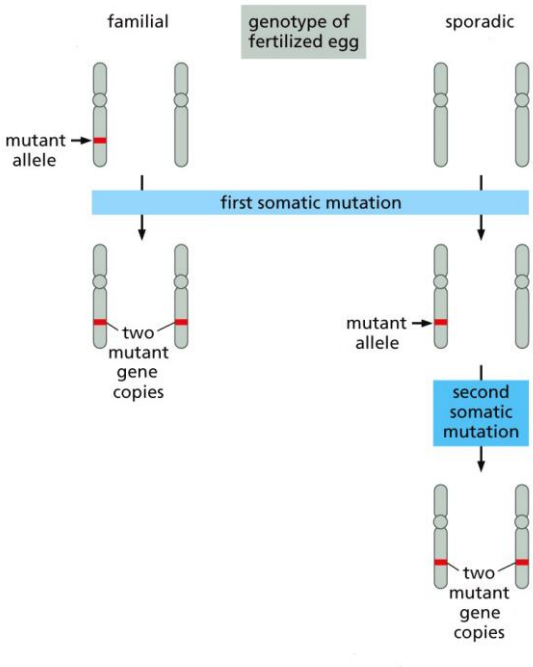


Figure 1. Individuals with one defect allele need only one somatic mutation in order to develop cancer, while individuals with two normal alleles require two somatic mutations in the same cell in order to develop cancer (adapted from [1]).

A familial mutation is present in all the cells of an individual, since the variant has been passed through the germ line from a parent to a fertilized egg. This means that individuals with a deleterious familial mutation can develop different types of cancer, depending on the type of mutated gene.

1.1 Hereditary breast and ovarian cancer

Breast cancer is the most commonly occurring cancer among women, with a 8.5 % chance of developing breast cancer before the age of 75 [4]. In 2015 3415 new cases of breast cancer (excluding 24 cases of male breast cancer) were reported in Norway [4]. However, in these patients, the five-year survival is quite high (89.0 %) [4]. This is probably due to a good population awareness together with the development of better diagnostic tools.

Ovarian cancer is less common than breast cancer, with a 1.3 % chance of developing ovarian cancer by the age of 75 [4]. In 2015, 504 new cases of ovarian cancer were reported in Norway. The survival rate is lower for ovarian cancer than it is for breast cancer, with a five-year survival of 48.1 % [4]. This is probably because the symptoms of ovarian cancer are more diffuse than the symptoms of breast cancer, and since there are no good programs for screening of ovarian cancer, thus the cancer is discovered at a later stage in the cancer development.

About 5-10 % of all breast cancer cases, and about 25 % of ovarian cancer occurrences are familial [5]. The genes often associated with hereditary breast or ovarian cancers (HBOC) are tumour suppressor genes. The first recognized genes associated with HBOC in the 1990s, were *BRCA1* and *BRCA2* [6, 7]. Mutations in these two genes comprise about 24 % of the HBOC cases [8]. Currently, more than 25 genes have been associated with HBOC [9]. Many of these genes encode tumour suppressors that participate in genome stability pathways, like homologous recombination repair and mismatch repair, two separate mechanisms essential to ensure the genetic integrity during cell division [9].

1.2 Cell cycle

The cell cycle is the reproductive cycle of a cell, and entails a series of events leading to duplication of DNA and cell division [10]. The cell cycle is divided into four phases: gap 1 phase (G_1 phase), DNA synthesis phase (S phase), gap 2 phase (G_2 phase) and mitotic phase (M phase) (Figure 2). The G_1, S, and G_2 phase constitute the interphase of the cell cycle. During the G_1 and G_2 phases the cell grows and duplicates its cytoplasm. In the S phase the nuclear DNA is replicated, which is essential for cell division. The M phase consists of mitosis, which is the process leading to the nucleus dividing, and the cell dividing into two identical daughter cells [10]. During the G phases the internal and external environment is

monitored by the cell to ensure that everything is ready for the cell to go into S phase and M phase. If the cell is not yet ready for the next phase, the cell can decide to pause to allow more preparation time [10].

To ensure that the DNA and organelles of cells are replicated and divided in an orderly manner, the cells have a cell-cycle control system [10]. This control system consists of a complex network of regulatory proteins, which ensures that the events of the cell cycle happen in the correct order, and that one process does not start before the last process has been completed. There are three checkpoints where the cell cycle can be stopped if a process is not completed. These checkpoints are in G_1 before entering S phase; in G_2 before mitosis; and during mitosis before the nucleus divides [10]. The checkpoint in G_1 can instruct the cell to delay progress, and even to enter a resting state called G_0, if extracellular conditions are unfavourable [10].

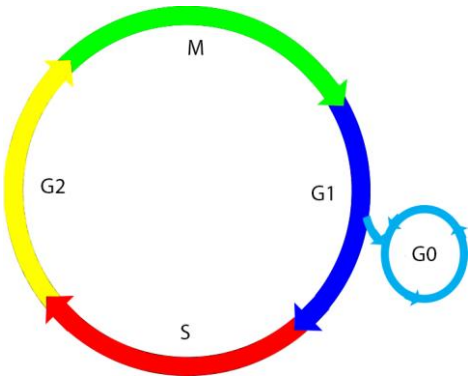


Figure 2. Cell cycle with the gap 1 phase (G_1 phase), DNA synthesis phase (S phase) and gap 2 phase (G_2 phase) and mitotic phase (M phase), as well as cell cycle arrest phase (G_0 phase).

The G_2 checkpoint ensures that the cell does not enter mitosis before damaged DNA has been repaired and DNA replication is completed [10]. Some of the DNA repair mechanisms are described in the upcoming paragraphs. If the DNA damage is too severe, the cell can be ordered to go into apoptosis at this checkpoint [9, 10]. Damaged DNA will activate ATM, which promotes DNA damage checkpoint signalling. ATM phosphorylates CHK2, p53 or PALB2, which leads to either cell cycle arrest or apoptosis (Figure 3) [9]. A brief overview of these proteins, their function and the genes encoding them will be described in the following paragraphs.

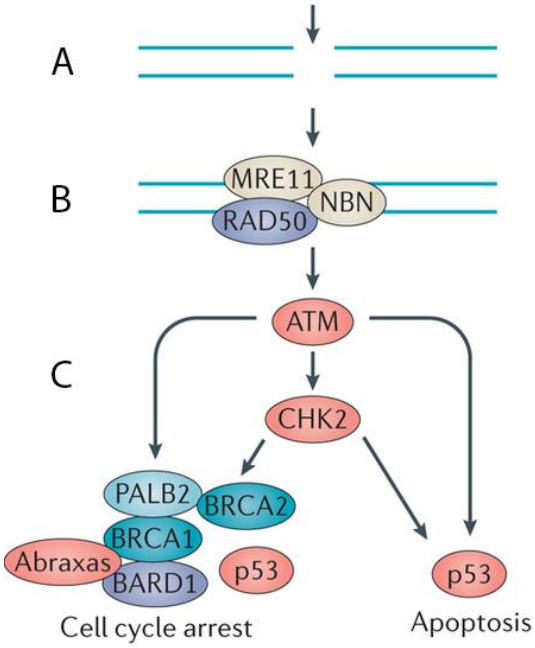


Figure 3. DNA damage signalling. The damaged DNA (A) is detected by a protein complex (B), and activates ATM. The activated ATM phosphorylates other proteins in different pathways, which can result in either cell cycle arrest or apoptosis (C) (adapted from [9]).

The mitotic checkpoint ensures that the replicated chromosomes are properly attached to the mitotic spindle before the chromatids are pulled apart and distributed equally between the two daughter cells [10]. The mitotic spindle is part of the cytoskeleton which forms from the centrosomes during mitosis.

1.2.1.1 ATM

The ataxia telangiectasia mutated gene (*ATM*) consists of 63 exons (NM_000051.3) and is located on chromosome 11q22.3. The gene encodes a 350 kD protein with 3 056 amino acids (aa) [11, 12]. The ATM protein is a serine/threonine kinase which is mainly involved in the response to DNA double stranded breaks (DSBs), and functions upstream of p53 and CHK2, among others (Figure 3) [1, 13, 14].

The recessive disorder ataxia-telangiectasia (A-T) is the result of biallelic *ATM* deleterious variants [13, 15]. A-T is characterized by progressive neurodegeneration, cell cycle checkpoint defects, augmented ionizing radiation and increased incidence of lymphoid malignancies. *ATM* mutation carriers are considered to have a moderate-increased risk of developing breast cancer [5, 16, 17], but their risk for ovarian cancer is currently not known [9].

A-T is most often caused by nonsense variants, or frame-shift variants including premature stop-codons, but also missense variants are present in these patients [18]. Missense mutations affecting the function of the protein and thus acting in a dominant-negative manner are believed to be the main reason for the increased breast cancer risk [15, 19, 20]. However, Pylkäs and colleagues performed a study where they found that breast cancer susceptibility is not completely restricted to dominant-negative variants [15]. There are no clear mutation hotspots in *ATM*, but deleterious variants which affect the PI3-kinase domain (Figure 4) seems to be over-represented [18, 21].

Figure 4. The domains of the ATM protein. Deleterious variants which affect the kinase domain seem to be over-represented. SBS: substrate binding site; FAT protein domain shared by representative FRAP, ATM and TRAPP; FATC homologous protein domain: like FAT, but located C-terminal (adapted from [22]).

1.2.1.2 CHEK2

The checkpoint kinase 2 gene (*CHEK2*) is located on chromosome 22q12.1 and consists of 16 exons (NM_001005735.1) [11, 12]. *CHEK2* encodes a 65 kD protein with 586 aa, which is also a serine/threonine kinase, like ATM. The protein, Chk2, is part of the DNA damage response pathway, and has its function downstream of ATM and upstream of p53 (Figure 3) [23, 24].

CHEK2 deleterious variants are considered to increase the risk of developing breast cancer, as well as sarcoma and brain tumours [5, 16, 17, 24, 25]. It is not quite clear whether *CHEK2* deleterious mutations can increase ovarian cancer risk, though *CHEK2* variants have been found in patients with both breast and ovarian cancer [9, 26].

1.2.1.3 TP53

The tumour protein p53 gene (*TP53*) consists of 11 exons (NM_000546.5) and is located on chromosome 17p13.1. The gene encodes p53, a 53 kD protein with 393 aa [11, 12]. The protein functions as part of the stress response for the cell and regulates a number of different genes, resulting in cell-cycle arrest, apoptosis, senescence, metabolic alterations and DNA damage repair [27].

Mutated *TP53* occurs in more than 50% of all human cancer tissues, and most often the gene has acquired a missense mutation [27]. These missense mutations frequently occur at one of six hotspots [27]. These hotspots are located in the DNA binding domain of the protein (Figure 5). The *TP53* gene typically acts as a tumour suppressor gene which can experience loss-of-function mutations, however, gain-of-function mutations which give the protein novel abilities, can also occur [27].

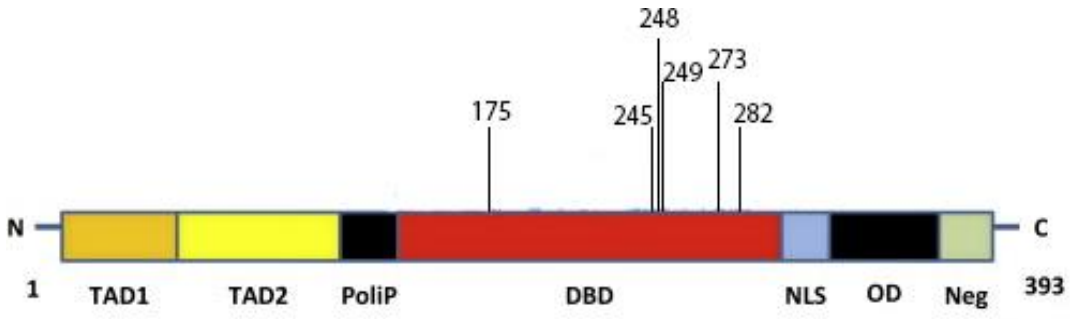


Figure 5. The p53 protein with domains and the amino acid position of the six mutational hotspots. The hotspots are all located in the DNA binding domain (DBD) of the protein. TAD1 and TAD2 = transactivation domain 1 and 2; PoliP = proline rich domain; NLS = nuclear localisation site; OD = oligomerization domain; NEG = negative regulation domain (adapted from [28]).

Deleterious variants in the *TP53* gene are associated with Li-Fraumeni syndrome, which is a autosomal dominant inherited condition that predisposes to breast cancer, sarcoma, brain tumours and adrenocortical carcinoma [9]. *TP53* deleterious variants have also been observed in cases of ovarian cancer, however, very rarely [9].

1.2.2 Homologous recombination repair

Homologous recombination repair is a mechanism which is activated when a double-strand break (DSB) of DNA has occurred during or after DNA replication. The HRR mechanism uses the intact sister chromatid as a template for the repair. Since the sister chromatid is only available after DNA replication and before mitosis, HRR is active during S and G_2 phase [9, 10].

After a DSB (Figure 6A), the free dsDNA ends are detected by the MRN protein complex (consisting of MRE11, RAD50 and NBN) which activates ATM (Figure 6B). The MRE11 nuclease processes the dsDNA ends to expose the 3’ ssDNA ends. This process is promoted both by the rest of the MRN complex and ATM [9, 10]. Subsequently, RAD51 proteins are

loaded onto one of the exposed ssDNA ends by BRCA2 in a complex with PALB2, forming the nucleoprotein filament (Figure 6C). The RAD51 filament then pairs with the complimentary DNA strand on the sister chromatid, which is used as a template for DNA synthesis [9, 10]. The repair is concluded by additional DNA synthesis of the other ssDNA strand on the original sister chromatid, followed DNA ligation (Figure 6D).

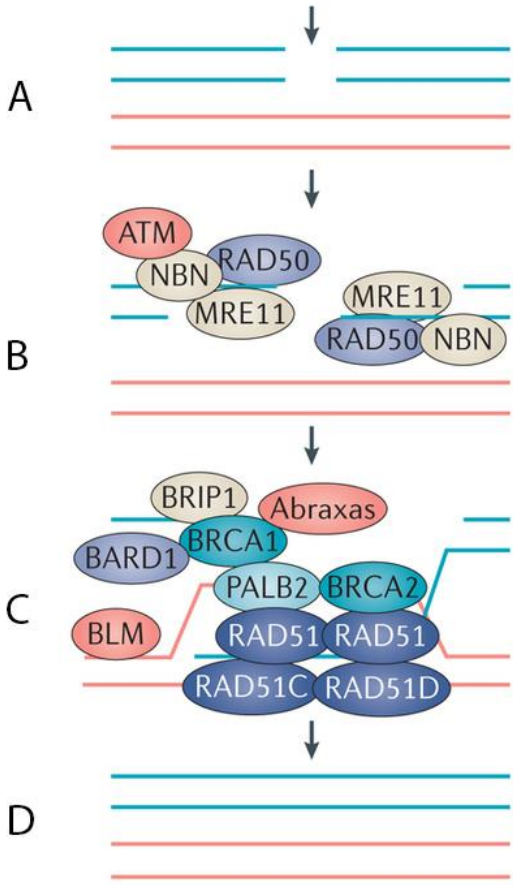


Figure 6. Homologous recombination repair. The DNA DSB (A) is detected by the MRN complex (MRE11, NBN, RAD50), which recruits ATM (B). The MRE11 protein processes the dsDNA ends to expose the ssDNA ends. RAD51, BRCA2 and PALB2 form a nucleoprotein filament which pairs one ssDNA from the chromatid which have undergone DSB with the complimentary DNA strand on the sister chromatid. The sister chromatid is used as a template for synthesis of the DSB ssDNA (C). The newly synthesised ssDNA returns to its original chromatid and the other strand is synthesized, before the ends are ligated, and the HRR is completed (D). The exact function of the remaining proteins (BRIP1, BARD1, BRCA1 and Abraxas) shown in this figure is not fully known, other than that they promote HRR (adapted from [9]).

Some genes encoding proteins involved in the HRR pathway are associated with an increased risk of developing breast and/or ovarian cancer. A brief overview of the genes, their products and function will be summarized next.

1.2.2.1 *BRCA1* and *BRCA2*

The breast cancer 1 gene (*BRCA1*) consists of 23 exons (NM_007294.3), is located on chromosome 17q21.3, and encodes a 207 kD protein with 1 863 aa [11, 12]. The protein is a part of the response to DSBs of DNA (Figure 6), and has a wide range of functions, however many are still poorly understood [1, 9, 29].

The breast cancer 2 gene (*BRCA2*) is located on chromosome 13q13.1, and consists of 26 exons (NM_000059.3). It encodes a 384 kD protein with 3 418 aa [11, 12]. BRCA2 is also part of the response to DSBs of DNA where it interacts with different proteins including BRCA1 (Figure 6) [9, 29].

Deleterious mutations in *BRCA1* and *BRAC2* are associated with an increased lifetime risk of developing breast and ovarian cancer, and are associated with 24 % of HBOC cases [5, 8, 16]. The risk of developing breast cancer by age 70 is calculated to be 57-65 % for *BRCA1* mutation carriers and 45-55 % for *BRCA2* mutation carriers, while the lifetime risk of developing ovarian cancer is 39-44 % for *BRCA1* and 11-18 % for *BRCA2* [9, 16]. *BRCA2* mutation carriers have a tendency to develop breast cancer later in life than *BRCA1* carriers [30].

1.2.2.2 PALB2

The partner and localizer of BRCA2 gene (*PALB2*) consists of 13 exons (NM_024675.3), is located on chromosome 16p12.2, and encodes a 131 kD protein with 1186 aa [11, 12]. The protein interacts both with BRCA1 and BRCA2 (Figure 6) [9, 31].

Deleterious mutations in *PALB2* are associated with a moderate-risk of breast cancer, though the susceptibility to ovarian cancer is suggested to be low [9, 16, 17].

1.2.2.3 *RAD51C* and *RAD51D*

The RAD51 paralog C gene (*RAD51C*) is located on chromosome 17q22 and consists of 9 exons (NM_058216.1). *RAD51C* encodes a 42 kD protein with 376 aa [11, 12]. The RAD51 paralog D gene (*RAD51D*) consists of 10 exons (NM_002878.3), is located on chromosome 17q12, and encodes a 35 kD protein with 328 aa [11, 12]. RAD51C and RAD51D are members of the RAD51 protein family, which are crucial for RAD51 nucleoprotein filament formation (Figure 6) [9].

Deleterious mutations in *RAD51C* and *RAD51D* are associated with an increased risk of ovarian cancer [16]. The role of mutated *RAD51C* and *RAD51D* play a role in breast cancer is still unclear, as deleterious variants in these genes mainly have been found in families with either exclusively ovarian cancer or combined breast and ovarian cancer [9, 32].

1.2.2.4 BRIP1

The BRCA1-interaction protein 1 gene (*BRIP1*) consists of 20 exons (NM_032043.2), is located on chromosome 17q23.2, and encodes 140 kD protein with 1249 aa [11, 12]. The BRIP1 protein is also known as FANCJ, and is suggested to be involved in HRR by recruiting BRCA1 to DSBs (Figure 6) [9].

Deleterious mutations in *BRIP1* are associated with increased risk of ovarian cancer, but do not seem to have an effect on breast cancer risk [9, 16]. These variants often have an effect on the ATPase helicase core domain which is comprised of eight motifs: 0 (Q), I, Ia, II, III, IV, V and VI (Figure 7) [33].

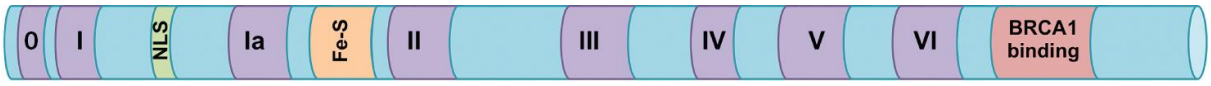


Figure 7. The BRIP1 protein with domains, including: the ATPase helicase core domain comprised by eight motifs (0, I, Ia, II, III, IV, V, VI); the nuclear localisation site (NLS); the Iron-Sulphur (Fe-S) cluster; and the BRCA1 binding domain (adapted from [34]).

1.2.2.5 NBN

The nibrin gene (*NBN*) consists of 16 exons (NM_002485.4), is located on chromosome 8q21.3, and encodes a 85 kD protein with 754 aa [11, 12]. Nibrin is part of the MRN complex which is included in the DNA DSBs repair (Figure 6), meiotic recombination, cell cycle checkpoints (Figure 3), and the maintenance of telomeres [9, 35].

Deleterious variants in *NBN* are considered to results in a moderate-risk for development of breast cancer [16, 35] and are also associated with ovarian cancer [9]. However, one study performed by Ramus *et al.* [34], suggests that *NBN* does not increase the risk of ovarian cancer as they found no difference in mutation frequency between patients and controls.

The remaining proteins involved in HRR in Figure 6, have not been reported to increase the risk of cancer development or where not included in this study and are therefore not described further.

1.2.3 DNA mismatch repair

DNA mismatch repair (MMR) is a repair mechanism dedicated to correcting mismatches which occur during DNA replication (Figure 8) [9, 10]. These mismatches can be insertions, deletions and disincorporation of nucleotides. During replication, one mismatch is made per 10^7 nucleotides synthesized [10]. The MMR mechanism repairs about 99 % of these errors, lowering the occurrence of mismatches to $1/10^9$ nucleotides [10]. A complex of MMR proteins recognizes these DNA mismatches, excises the newly synthesised strand which contains the mismatch, and resynthesizes the missing strand (Figure 8B-D) [10].

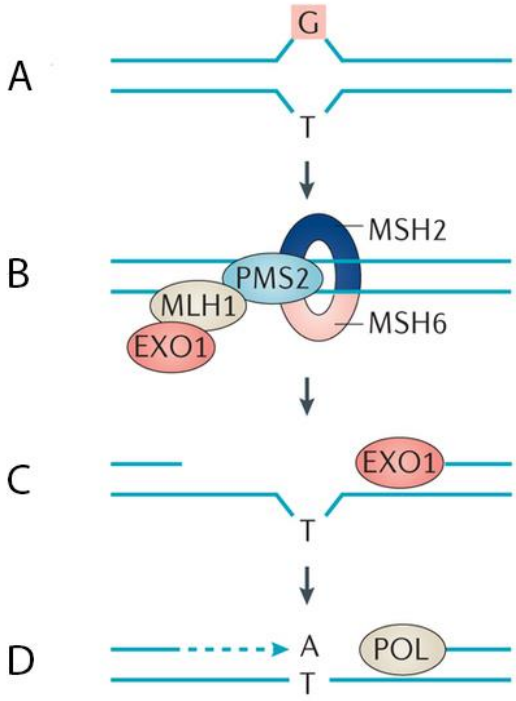


Figure 8. Mismatch repair mechanism. The mismatch (A) is detected by the heterodimer complex formed by MSH2 and MSH6 (B). The newly synthesised strand with the mismatch is then excised by the heterodimer complex formed by MLH1 and PMS2 (C), and a new strand is synthesised (D). The remaining factors are depicted as they are necessary for the repair reaction (addapted from [9]).

Deleterious variants in the MMR genes cause Lynch syndrome (LS), also called hereditary nonpolyposis colorectal cancer (HNPCC) [9, 36, 37]. LS is an inherited syndrome which predisposes to development of early-on-set cancer. There are many different sites of cancer development associated with LS, including colon, rectum, endometrium, ovarian, stomach,

small bowel and urinary tract [36, 37]. It is also believed to give a small increase in the risk of breast cancer, but this is still debated [9, 38-40].

1.2.3.1 MLH1

The mutL homologue 1 gene (*MLH1*) is located on chromosome 3p22.2 and consists of 19 exons (NM_000249.3). *MLH1* encodes a 84 kD protein with 756 aa [11, 12, 36]. The protein forms a heterodimer complex with PMS2 which removes the mismatch during MMR (Figure 8) [41, 42].

Deleterious variants in *MLH1* cause LS and increase the lifetime risk of developing ovarian cancer, as well as breast cancer [38, 39]. Individuals with deleterious *MLH1*variants have a similar risk of developing ovarian cancer as *BRCA1* mutation carriers [43].

1.2.3.2 MSH2

The mutS homologue 2 gene (*MSH2*) consists of 16 exons (NM_000251.2), is located on chromosome 2p21-p16, and encodes a 104 kD protein with 934 aa [11, 12, 36]. The MSH2 protein is part of a heterodimer complex which detects mismatches after DNA replication, together with MSH6 (Figure 8) [41, 42].

Deleterious *MSH2* variants cause LS and give an increased lifetime risk of both breast cancer and ovarian cancer [38, 39, 43]. The risk of cancer is quite similar to individuals with deleterious variants in *MLH1*.

1.2.3.3 MSH6

The mutS homologue 6 gene (*MSH6*) consists of 10 exons (NM_000179.2), is located on chromosome 2p16.3, and encodes a 152 kD protein with 1360 aa [11, 12]. The protein participates in a heterodimer complex together with MSH2, and detects mismatches after DNA replication (Figure 8) [41, 42].

The risk of developing breast cancer is believed to be higher in *MSH6* mutation carriers than in the general population [38]. It is uncertain if the risk of developing ovarian cancer is increased with deleterious *MSH6* variants. Some studies report that there is a higher lifetime

risk of ovarian cancer [39], while other studies report that the risk of developing ovarian cancer is close to the lifetime risk estimated for the general population [43].

1.2.3.4 PMS2

The PMS1 homolog 2 gene (*PMS2*) is located on chromosome 7p22.1, and consists of 15 exons (NM_000535.6). *PMS2* encodes a 95 kD protein with 862 aa [11, 12]. The PMS2 protein takes part in MMR in a complex with MLH1 and removes the newly synthesised strand containing mismatches (Figure 8) [41, 42].

Deleterious variants in *PMS2* are associated with an increased risk of developing breast and ovarian cancer [40]. These variants are often difficult to detect due to strong homology between the gene and several pseudogenes. There are a total of 15 different pseudogenes detected in the human genome, which share homology with exon 1-5, 9 and 11-15 of *PMS2* [37]. The main problem for mutation detection is the pseudogene *PMS2CL*, which has a homology level of about 98% with the *PMS2* gene. The *PMS2CL* corresponds to exon 9 and 11-15 in the *PMS2* gene [44].

1.3 Other HBOC related genes

1.3.1 NF1

The neurofibromatosis type 1 gene (*NF1*) consists of 58 exons (NM_001042492.2), is located on chromosome 17q11, and encodes a 319 kD protein with 2839 aa [11, 12]. The encoded protein, neurofibromin, is a GTPase activating protein which regulates the Ras signalling pathways (Figure 9) [9, 45]. The Ras protein control signalling pathways that are key regulators of several aspects of normal cell growth and malignant transformation.

Deleterious mutations in *NF1*cause neurofibromatosis type 1, an autosomal dominant disorder characterised by cutaneous neurofibromas and pigmentary changes [9, 45]. Carriers of deleterious *NF1*variants also have an increased risk of different cancers, including breast and ovarian cancer [9].

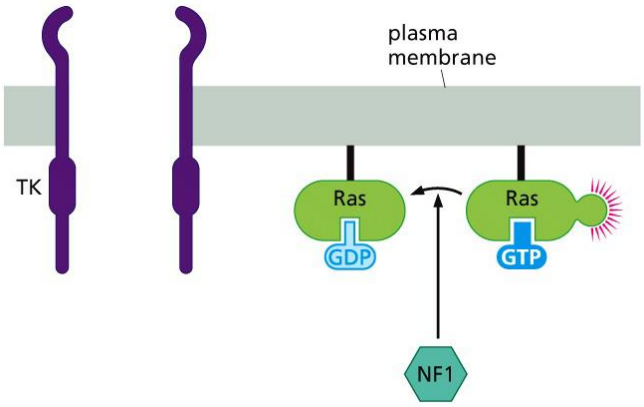


Figure 9. The RAS signalling pathway. NF1 ensures that the level of GTP-loaded Ras is held low in cells which are not experiencing mitogenic signalling by inducing GTP hydrolysis by Ras proteins (adapted from [1]).

1.3.2 CDH1

The cadherin 1 gene (*CDH1*) is located on chromosome16q22.1, consists of 16 exons (NM_004360.3), and encodes a 97 kD protein, called E-cadherin, with 882 aa [11, 12]. E-cadherin is a transmembrane glycoprotein which is part of calcium-dependent cell-cell adhesion (Figure 10) [9, 46].

Deleterious mutations in *CDH1* are associated with hereditary diffuse gastric cancer (HDGC). Families with HDGC often have cases of lobular breast cancer, as well as diffuse gastric cancer [9, 46, 47]. The majority of reported deleterious variants found in HDGC families are nonsense variants [47].

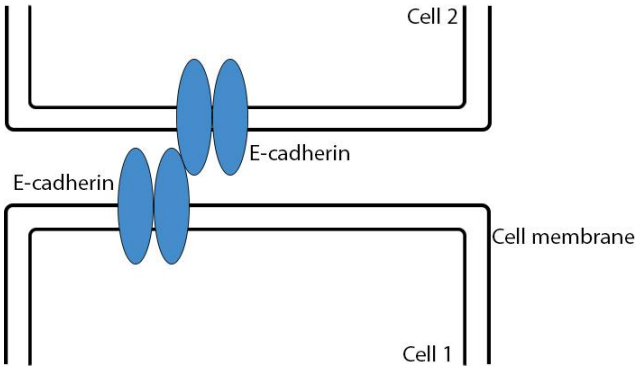


Figure 10. Cell-cell adhesion. E-cadherin is a transmembrane glycoprotein which is involved in calcium-dependent cell-cell adhesion.

1.3.3 PTEN

The phosphate and tensin homolog gene (*PTEN*) consists of 9 exons (NM_000314.4), and is located on chromosome 10q23.3. *PTEN* encodes a 47 kD protein with 403 aa [11, 12]. The protein is the primary negative regulator of the PI3K-AKT-mTOR pathway which promote cell proliferation (Figure 11) [9, 48].

Deleterious mutations in *PTEN* are associated with Cowden syndrome. Cowden syndrome is characterised by macrocephaly, skin hamartomas, gastrointestinal polyps, and increased risk of several different cancers, including kidney, lung, breast and prostate cancer [9, 48].

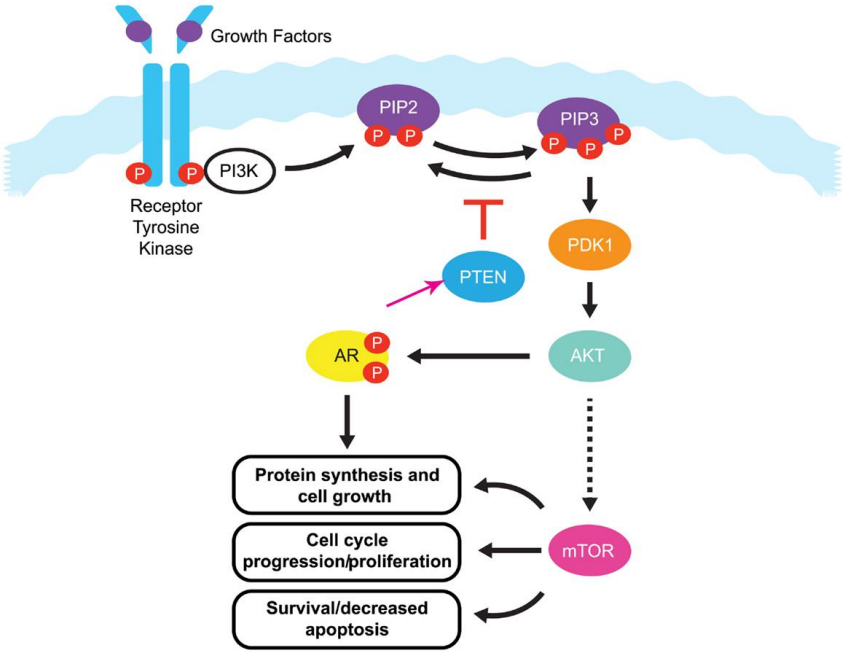


Figure 11. PI3K-AKT-mTOR pathway. The pathway promotes cell proliferation, and is negatively regulated by PTEN (marked with a red arrow) (adapted from [48]).

1.3.4 STK11

The serine/threonine kinase 11 gene (*STK11*) consists of 10 exons (NM_000455.4), is located on chromosome 19p13.3, and encodes a 48 kD protein with 433 aa [11, 12]. The protein regulates the adenosine monophosphate-activated protein kinase pathway, which regulates a broad spectrum of functions in the cell, including growth, metabolism, autophagy and polarity (Figure 12) [49].

Deleterious mutations in *STK11* are associated with Peutz-Jeghers syndrome, which is characterised by polyps in the gastrointestinal tract, mucocytaneous pigmentation and

increased risk of cancer. Breast and ovarian cancer are among the cancer associated with Peutz-Jeghers syndrome [9, 49, 50].

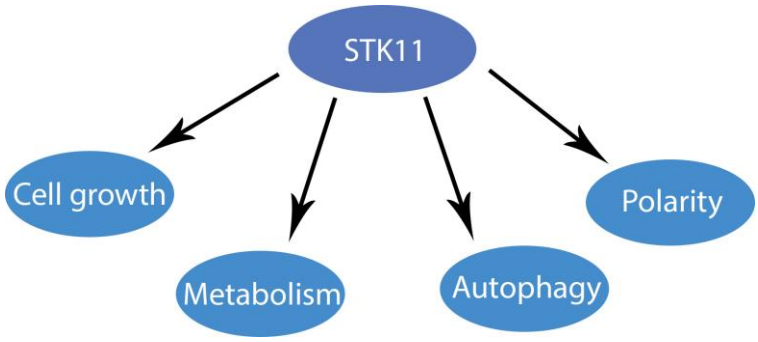


Figure 12. STK11 has a broad spectrum of functions in cells, including growth, metabolism, autophagy and polarity.

1.4 Founder mutations

In one part of this study we investigated a specific variant in *BRCA2* which has been identified in several families in Norway.

There are some populations that have a few recurring mutations. These mutations can be due to several occurrences at a mutational unstable hotspot, or can be because of a founder effect [30, 51]. Founder effects are the result of a geographically or culturally isolated population that was founded by a small group of individuals. This leads to a less diverse genetic population. If one or more of the ancestors of the population were carriers of a specific mutation, this mutation would have a relatively high frequency in the population, and is called a founder mutation [52-54].

There are several examples of founder mutations in cancer-associated genes in different populations around the world. The Ashkenazi Jews have three characterised founder mutations in *BRCA1* and *BRCA2* (two in *BRCA1* and one in *BRCA2*) [30, 51, 54]. One study found proof of eight founder mutations in *BRCA1* and *BRCA2* (five in *BRCA1* and three in *BRCA2*) in a population of breast and/or ovarian cancer patients in the North-East of Italy [30]. Another example of founder mutations are four *BRCA1* mutations found in the south of Norway [51].

In order to decide if a recurrent mutation is a founder mutation or a hotspot mutation, it is important to determine if mutation carriers have a common ancestor. This can be done by

studying genetic markers which are associated with the mutation through several generations due to the close proximity to the mutation [54]. These markers can be either single nucleotide polymorphisms (SNPs) and/or microsatellite markers. A series of markers close to the mutation which segregate together through generations, constitute the haplotype following the mutation [54]. If the mutation occurs on only one haplotype, it is considered a founder mutation. A recurrent mutation which occurs on different haplotypes has occurred several times independently and is not considered a founder mutation.

1.4.1 Microsatellites

Microsatellite markers (also called short tandem repeats) are sequences with 2-7 repeated nucleotides [52, 55]. A microsatellite marker can have a high variety of lengths in a population. By analysing the length of different microsatellite markers the genotype of an individual can be determined. This is often used both in family and forensic studies [55, 56].

The length of microsatellite markers can change over time. There are two types of events which can cause variations in the size of the microsatellite marker: meiotic recombination between two homologous chromosomes with different alleles, or polymerase stutter during DNA replication [56].

The haplotype of the microsatellite markers and the variant located in close proximity at the same chromosome will segregate together. Due to the changes which happen to the markers over time, the haplotype of the chromosome with the variant will also change. This change happens first to the microsatellite located furthest away from the variant. The time which has passed since the variant arose can be calculated by looking at the haplotype for the microsatellite in close proximity to the variant among carriers. The smaller the common haplotype among the carriers, the older the variant.

2 Aim

The aims of this study were:

a. To investigate patients with breast and/or ovarian cancer for contributing genetic factors using next generation sequencing (NGS). These patients had all previously been tested for deleterious variants in *BRCA1* and *BRCA2*, however no such variants were found in these two genes. A total of 16 genes were scrutinised in order to determine if other genes could be the cause of HBOC in these patients.

b. *BRCA2* c.8331+2C>T, which is present in several breast cancer patients in Norway, was more thoroughly investigated to determine if:
   i. the variant had an influence on RNA splicing
   ii. the high prevalence of the variant in the Norwegian HBOC population was due to a reoccurring mutation event or due to a founder effect.

3 Materials and methods

3.1 Patient samples

Forty eight DNA samples from patients with a family history of breast and/or ovarian cancer were investigated by NGS. None of these patients had previously identified pathogenic variants or variants of unknown (VUS) significance in *BRCA1/2*. The samples originated from two different groups for which no NGS analysis had been performed previously. Group 1 (n=32) consisted of samples collected from deceased breast/ovarian cancer patients. Due to exemption for consent, received from the Regional Ethical Committee (REK), these samples could be investigated without informed consent from relatives. Group 2 (n=16) consisted of breast/ovarian cancer patient samples. Patients belonging to this group were mailed information about the study and got the opportunity to passively consent. The NGS study was approved by the Regional Ethical Committee (REK no 2016/980/REK).

For the second part of the study, DNA samples from families with the *BRCA2* c.8331+2C>T were used for microsatellite analysis. A total of 94 DNA samples from family members with and without the variant were included in this study. One RNA sample from a patient with the mutation and three normal control RNA samples were used to study the effect this sequence variant had on splicing events.

3.2 DNA-extraction

In order to extract DNA from patient samples an extraction instrument called QIAsymphony Sample preparation (SP) (by QIAGEN) was used. Blood samples collected in EDTA containing blood sample tubes were used.

DNA isolation was performed using the QIAsymphony® DSP DNA Midi kit (by QIAGEN) according to the manufacturer's protocol. In brief: QIAsymphony SP uses magnetic particles and a magnetic rod to extract the DNA (Figure 13). Inside the instrument the blood samples are first transferred to well 1, to which a lysis buffer and magnetic particles are added. When the DNA containing cells are lysed, the negatively charged DNA molecules attach to the magnetic particles. A magnetic rod removes and transfers the DNA/magnetic particles to well 2, where the DNA/magnetic particles are released from the rod. Subsequently, several washing steps purify the DNA from residual of other cell components. In the last step the

DNA/magnetic particles are again attached to the magnetic rod, before the DNA is eluted from the magnetic particles with an elution buffer into new tubes.

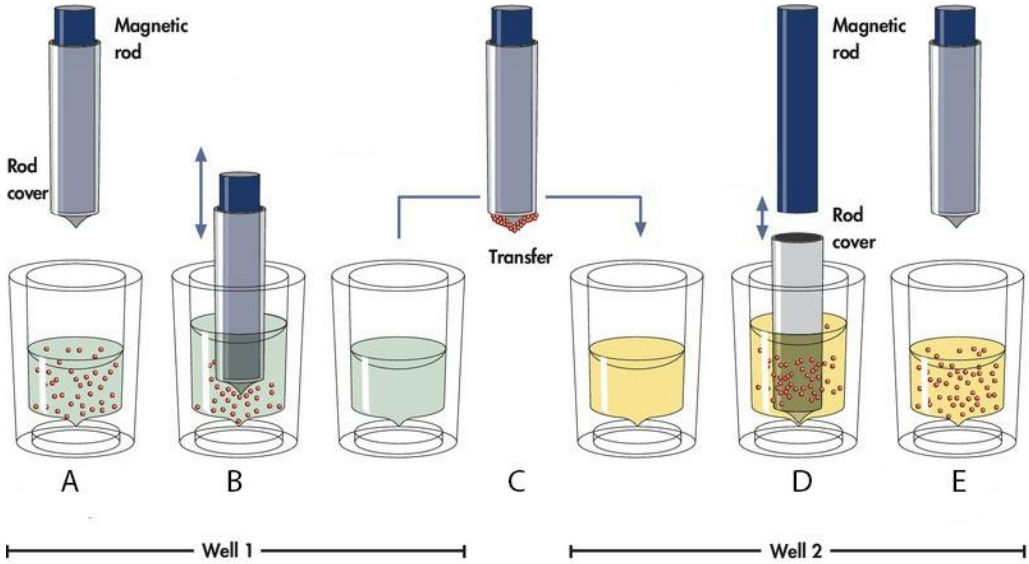


Figure 13. DNA-extraction on a QIAsymphony. (A) The cells are lysed and the released DNA attaches to the magnetic particles in well 1. (B) The magnetic rod then attaches the magnetic particles and (C) transfers these to a new well, called well 2. (D) In the new well the magnetic particles with DNA molecules are released from the magnetic rod, and are subsequently (E) washed several times (Adapted from [57]).

3.3 Quantification

3.3.1 Qubit Fluorometric method

The Qubit 3.0 Fluorometer (Invitrogen by Thermo Fisher Scientific) was used for the quantitation of the DNA samples during the NGS Library preparation. The samples were quantified before tagmentation and before the libraries were pooled, as well as during the library validation.

The fluorometric method quantifies the amount of DNA using dye which is only fluorescent when bound to DNA. The Qubit dsDNA High Sensitivity (HS) Assay Kit (Invitrogen by Thermo Fisher Scientific) was used according to the manufacturer's protocol. In brief: the reagents were equilibrated to room temperature. A Qubit Working Solution was made by diluting Qubit dsDNA Reagent 1:200 in Qubit dsDNA buffer. The method uses two standards to calculate the concentration of the samples. The amount of working solution and standard/sample needed for the analysis are listed in Table 1.

Table 1. The volumes working solution and sample needed for the Qubit quantification performed during the different steps of NGS analysis

	Working solution (µL)	*Sample (µL)*
Standard solution	190	10
Normalization 1st dilution	198	2
Normalization 2nd and 3rd dilution*	195	5
Before pooled libraries	199	1
Library validation	190	10

*-if necessary

After the standards and samples have been mixed with the Qubit Working Solution, they were vortexed and incubated at room temperature (RT) for 5 minutes before quantitation. The two standards were measured first, followed by the samples. Each sample was measured twice. There had to be at least 30 seconds between each time the same standard or sample was re-measured.

3.3.2 Bioanalyzer

The Bioanalyzer (Agilent Technologies) was used to determine the size of the DNA fragments after the tagmentation step, and during the NGS library validation (Figure 15C).

The Bioanalyzer uses the same principle as a classic gel electrophoresis, the main difference being that it comes in a chip format. The method uses a polymer gel mixed with a fluorescent dye. The chip has 16 wells, consisting of sample wells, gel wells and a well for an external standard (ladder). There are micro-channels connecting the wells, which are filled with the gel-dye mix during chip preparation [58]. A 16-stick electrode (inside the Bioanalyzer instrument), one for each well, applies a voltage through the chip in order for the DNA molecules to be separated. The molecules are then detected by laser induced fluorescence detection. Due to the applied ladder, the size of the DNA fragments in each sample can be calculated.

The HS DNA kit and the 2100 Bioanalyzer from Agilent Technologies were used according to the manufacturer's protocol. A gel-dye mix was made mixing the gel matrix and the dye which came with the kit. Before being used, the reagents of the kit were stabilized for 30 minutes at room temperature. Subsequently, 9 µL of the Gel-Dye mix was added to one well

of the HS DNA chip, which was placed in a priming station (Figure 14A). The chip-priming station, with a syringe, was then used to add pressure to the gel in order to transfer it into the micro-channels. Thereafter, 9 µL Gel-Dye was added to three other wells on the same column (Figure 14B). Before 1 µL of the HS ladder and 1 µL of the samples were added to their respective wells (Figure 14D and 14E), 5 µL of a marker was added to the same wells (Figure 14C). The chip was vortexed before it was analysed using the 2100 Bioanalyzer.

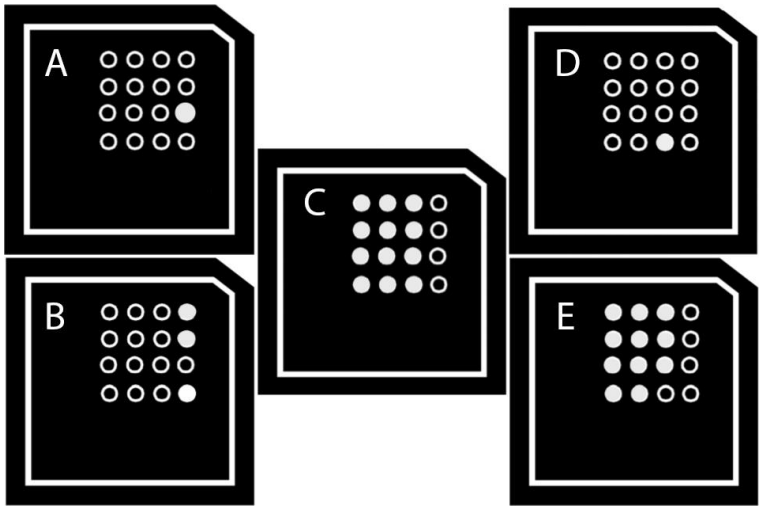


Figure 14. Loading to the Bioanalyzer chip. (A) 9 µL gel is added and pressurized before (B) the remaining gel wells are each added 9 µL gel. (C) 5 µL marker is added to the remaining wells before (D) 1 µL of the ladder and (E) 1 µL of the samples are added to the same wells (adapted from [59]).

3.3.3 Nanodrop

The Nanodrop 2000 (Thermo Fisher Scientific) was used to determine the DNA concentration of the samples used in the microsatellite analysis.

DNA absorbs light at 260 nm [60]. Nanodrop 2000 uses this DNA property in order to measure its concentration in a sample. The emitted light is detected by a spectrophotometer using a linear CCD (Charge-coupled device) assay [60]. The concentration of the sample is decided by how much light has been absorbed by the sample.

The Nanodrop 2000 was used according to manufacturer's protocol. Before the samples were measured 1µL of water was used to calibrate the instrument. Subsequently, 1 µL of each patient DNA sample was measured.

3.4 Next Generation Sequencing

Next generation sequencing was used to identify variants in cancer associated genes in a breast and ovarian cancer patient cohort, negative for *BRCA1/2* pathogenic variants and VUSs. The preparation of NGS libraries of the patient samples takes five days, and the different steps of the preparation are illustrated in Figure 15. DNA libraries were produced using the TruSight Rapid Capture Sample Preparation kit together with the TruSight Cancer sequencing panel from Illumina. This panel targets 94 different genes and 284 SNPs which have been associated with cancer development [61].

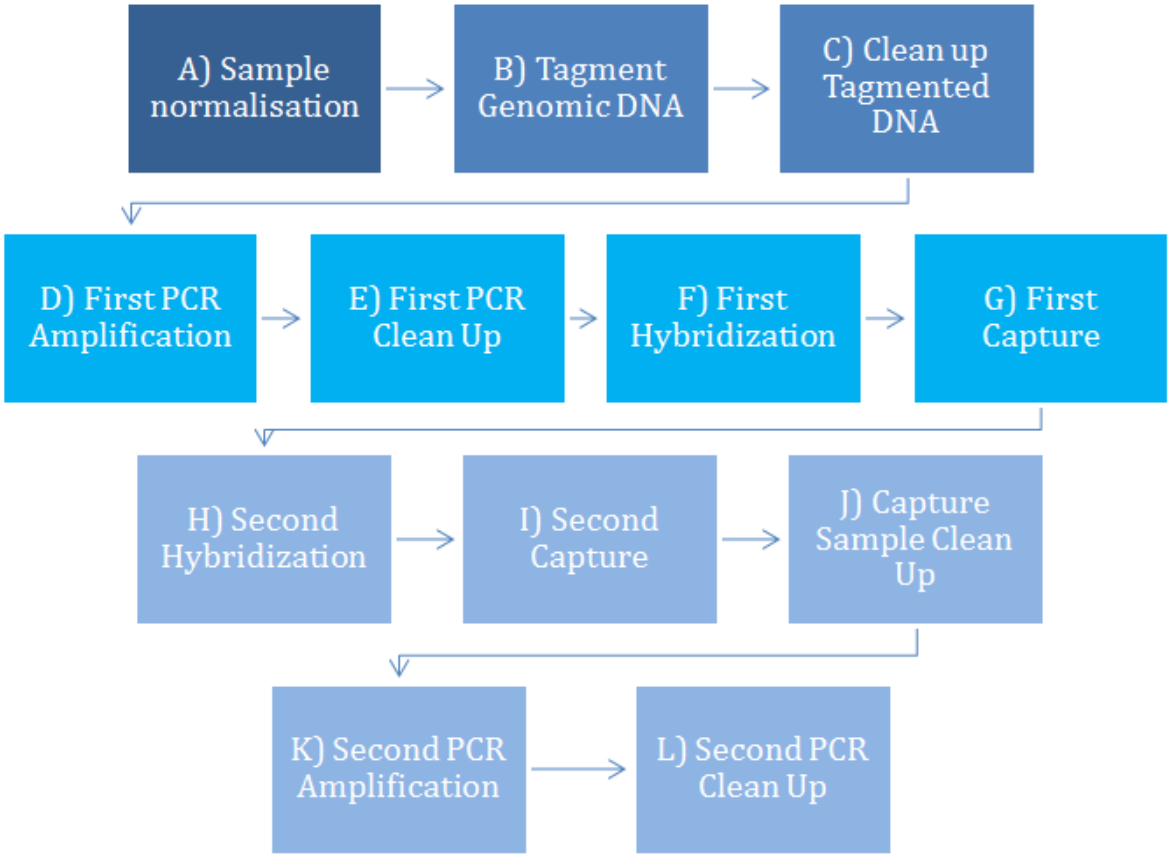


Figure 15. The different steps included in the library preparation applying TruSight Rapid Capture Sample preparation.

3.4.1 Library preparation

3.4.1.1 Sample normalization

Patients DNA samples were normalized to obtain a concentration of 5-7 ng/µL (Figure 15A). Thirty-five µL of 10 mM Tris-HCL was added to 15µL of each DNA sample, and was stored overnight at 4°C. Subsequently, each sample was quantified using the Qubit fluorometric

method (3.3.1 Qubit Fluorometric method). If the samples had a concentration higher then 5-7ng/µL, they were diluted and re-quantified until they had the required concentration.

3.4.1.2 Tagmentation

On the third day of the sample preparation, the tagmentation of genomic DNA was started (Figure 15B). Tagmentation was performed according to manufacturer's protocol. The first step of the library preparation is tagmentation of the genomic DNA (gDNA). This was done by adding transposomes which cut the gDNA into fragments. In addition, adapter sequences were added to the ends of the created fragments (Figure 16). Twenty-five µL Tagment DNA Buffer and 15 µL Tagment DNA Enzyme 1 were added to 10 µL of each sample, and the mixture was incubated at 58°C for 9 minutes and 40 seconds before 15 µL of a Stop Tagment Buffer was added.

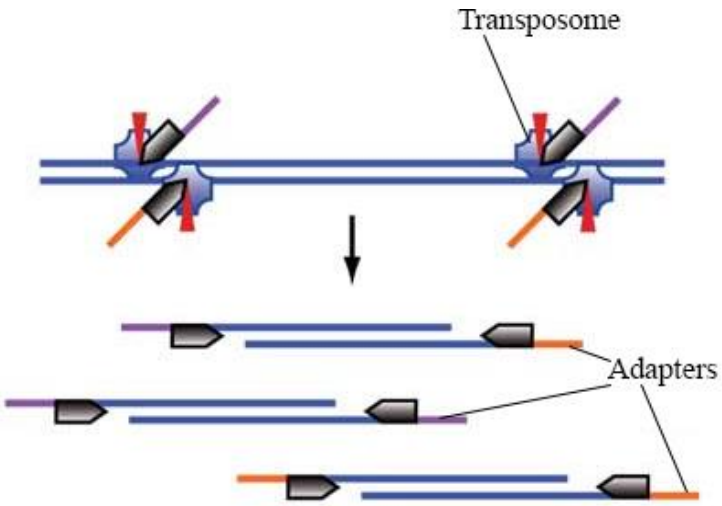


Figure 16.Tagmentation. Transposomes cut the gDNA randomly and ad adapter sequences at the ends of the fragments (adapted from [62]).

After the tagmentation process had been stopped, the DNA fragments were purified during the "Clean Up Tagmentation DNA"-step (Figure 15C). This step was performed in order to eradicate the transposomes as they can bind tightly to the DNA ends. If the enzymes are not removed they might interfere later in the library preparation process. The samples were subsequently cleaned using 65 µL Sample Purification Beads (SPBs), which binds the tagmented DNA, and washed twice with 200 µL 80 % ethanol (EtOH). The SPBs with DNA were then dried for 10 minutes, and resuspended by using the Resuspension Buffer (RSB).

The RSB removed the DNA fragments from the SPB, the beads were collected using a magnet, and the supernatant was subsequently transferred to a new well.

When the "Clean Up Tagmentation DNA" had been performed, the size of the tagmented DNA was investigated using the Bioanalyzer HS DNA kit (3.3.2 Bioanalyzer). The anticipated length of the DNA fragments should be approximately 400 bp (with a range of 150 bp – 1 kb).

3.4.1.3 First PCR Amplification

Subsequently, to each sample two different index primers were added as illustrated in Table 2. To each sample, 5 µL of each primer was added. The combination of these two index primers made it possible to identify which DNA fragments belonged to which patient after the libraries had been pooled together and sequenced. Nextera Library Amplification Mix, containing the common adapters (P5 and P7) which were needed for cluster generation, was also added in this step. The samples were amplified using a cycle PCR program (see Appendix B – PCR-programs, Table 16).

Table 2. Index primers 1 (i7, N701-N708) and 2 (i5, N501-N502) are added to the different samples in a combination of one i7 and one i5 for each sample. This was used to be able to tell the libraries apart after sequencing.

	N701	*N702*	*N703*	*N704*	*N705*	*N706*	*N707*	*N708*
N501	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	Sample 7	Sample 8
N502	Sample 9	Sample 10	Sample 11	Sample 12	Sample 13	Sample 14	Sample 15	Sample 16

After PCR amplification (Figure 15D), unincorporated primers were removed by using the same clean-up process as described above: a two-step washing using SPBs and 80 % EtOH; and resuspension using RSB (Figure 15E). The samples were subsequently quantified using Qubit Fluorometric method (see 3.3.1 Qubit Fluorometic method). The expected range of the sample concentration was 50-100 ng/µL.

3.4.1.4 Hybridization

The results of the quantification in the previous step were used in order to pool the libraries (Figure 15F). The concentrations were used to calculate the volume required in order to get 500 ng of each sample in the pooled library. If the total volume of the pool was higher than 40

µL, the pool was concentrated through an Amicon Ultra-0.5 centrifugal filter unit (Sigma Aldrich). The centrifugal filter unit was used according to the manufacturer's protocol. After concentration, the volume was brought up to 40 µL with Resuspension Buffer (RSB).

After the tagmented DNA samples had been pooled and concentrated to 40 µL, the pool was transferred to a strip tube. Fifty µL Enrichment Hybridization Buffer (EHB) and 10 µL TruSight Content Set CSO (Custom Selected Oligos) were then added to the strip tube. The tube was placed in a pre-programmed Agilent Technologies SureCycler 8800 on a cycle PCR program (see Appendix B – PCR programs, Table 17). This program hybridized the DNA library with capture probes to targeted regions of interest (Figure 15F).

3.4.1.5 First and second capture

At day four of the preparation process, the strip tube was removed from the PCR instrument after incubating at 58°C over night. The library then went through the first capture process (Figure 15G). In this step, 250 µL Streptavidin Magnetic Beads (SMB) were used to capture probes hybridized to the targeted regions of interest. The SMB with the pooled library was incubated at room temperature for 25 minutes before undergoing two heated wash procedures, using 200 µL Enrichment Wash Solution (EWS) for each wash. These wash procedures were performed in order to remove non-specific binding between the beads and the probes. For each wash the sample was incubated at 50°C for 30 minutes. After the heated wash procedures, the sample was eluted as followed: 28.5 µL Enrichment Elution Buffer 1 (EE1) and 1.5 µL 2N NaOH was mixed producing an elution pre-mix. Twenty-three µL of the mix was added to the sample in order to elute the targeted regions from the SMB. The clear supernatant was then transferred to a strip tube, before 4 µL Elute Target Buffer 2 (ET2) was added in order to neutralize the elution.

The eluted DNA library was subsequently combined with additional capture probes to the regions of interest (Figure 15H). In this step EHB and CSO was added to the library and the strip tube was placed in the PCR instrument on the PCR-program listed in Appendix B –PCR-programs (Table 17). The library was then incubated at 58°C for 14.5 -24 hours. This second hybridization was done to ensure high specificity of the captured regions.

The fifth day the library was removed from the PCR instrument and went through the second capture process, which is similar to the first capture using SMB, a two-step heated wash and elution step (Figure 15I).

3.4.1.6 Second PCR

The captured library was purified using the two-step 80 % EtOH wash described earlier (see 3.3.1.3 Tagmentation) (Figure 15J), before the library was amplified using a 12-cycle PCR program (see Appendix B – PCR-programs, Table 18) (Figure 15K). Subsequently, a new clean-up step using SPB and 80 % EtOH was done (Figure 15L). This step purified the library by washing away unwanted products.

3.4.1.7 Validate Library

In order to achieve the highest data quality it is important to achieve optimal cluster densities in the flow cell, which is where sequencing takes place. This requires accurate quantitation of DNA library templates. The sample containing the pooled libraries was diluted 1:13, before it was quantified using the Qubit fluorometric method (3.3.1 Qubit fluorometric method), and the concentration was expected to be between 0.5-1.0 ng/µL. The Bioanalyzer (3.3.2 Bioanalyzer) was used in order to determine the mean fragments size of the DNA libraries. The distribution of DNA fragments was expected to have a size range from about 200 bp-1kb (with the optimal size at 400 bp). The results from the Qubit analysis and the Bioanalyzer were used to convert the concentration from ng/µL to nM. The libraries were diluted to 1.2 nM.

3.4.2 Preparation for Sequencing on MiSeq

The libraries were subsequently prepared as illustrated in Table 3. The reagents were added to the library in the order they are presented in the table. The mixture was put on ice until it could be loaded into the reagent cartridge.

Table 3. The different solutions which were mixed before the libraries were sequenced. The reagents were added to the pooled libraries in the order they are presented.

	Volume	
1.2 nM library	10 µL	Incubated at room temperature for 5 min
0.1 N NaOH	10µL	
Hybridization Buffer (HT1)	980µL	
*20pM PhiX**	10µL	
Total	1010µL	Put on ice until loaded into reagent cartridge

* PhiX is an icosahedral, nontailed bacteriophage which contains a single stranded DNA (ssDNA). It has a small, well-defined genome (5386 nucleotides) and is commonly used as a control for NGS (Illumina) [63].

The flow cell used for the amplification and sequencing of the combined libraries, is stored in a buffer, and must be cleaned before use. Milli-Q water, EtOH and lens paper were used to clean the flow cell. It is important that there are no marks on the flow cell before use, as these can interfere with the analysis. The flow cell, the PR2 buffer (incorporation buffer) and waste bottle was placed inside the MiSeq, as instructed by the instrument. A sample sheet created by using Illumina Experiment Manager v.1.12.0 (Illumina), which contained the information of the composition of the index primers for each sample (Table 2), was opened in the software of the MiSeq. The reagent cartridge was placed inside the cooling chamber after the prepared library mix had been added to the correct well of the cartridge. After the MiSeq had made a pre-run to check different elements and connections, the NGS run was started.

3.4.3 MiSeq analysis

The principle of this method is based on sequencing by synthesis and uses reversible terminator nucleotides, where each of the four nucleotides are labelled with a different fluorescent dye as well as a reversible terminator. A DNA polymerase enzyme, which is capable of incorporating these nucleotides, is also used [64].

The instrument adds the validated library to a flow cell inside the MiSeq. The flow cell is covered in with covalently attached oligoes complementary to the adapters on the ends of the DNA fragments (Figure 17C). The DNA fragments from the libraries hybridizes to the oligoes on the flow cell by an active heating and cooling step [65]. Following the attachment of the DNA fragments, the pooled libraries are amplified using “bridge-amplification” (Figure 17). The bridge-amplification creates clusters, which are crucial for there to be a strong enough signal during the sequencing.

During the amplification the specific DNA fragments, which are attached to the flow cell at one end, bend over and the other fragment end is hybridized to another adapter on the flow cell, creating a "bridge-like" structure [66] (Figure 17D). These "bridge-like" structures are the templates for the amplification, and the complementary adapters act as primers [64]. The DNA fragments are then amplified by an isothermal polymerase (Figure 17E). Following the amplification the double stranded DNA (dsDNA) fragments are denatured and there will be two ssDNA fragments for each specific DNA fragment present at the flow cell (Figure 17F). The amplification is repeated over and over until clusters of about 1000 copies each of ssDNA fragments are created on the surface of the flow cell (Figure 17G). One cluster represents one type of ssDNA fragment [64, 65].

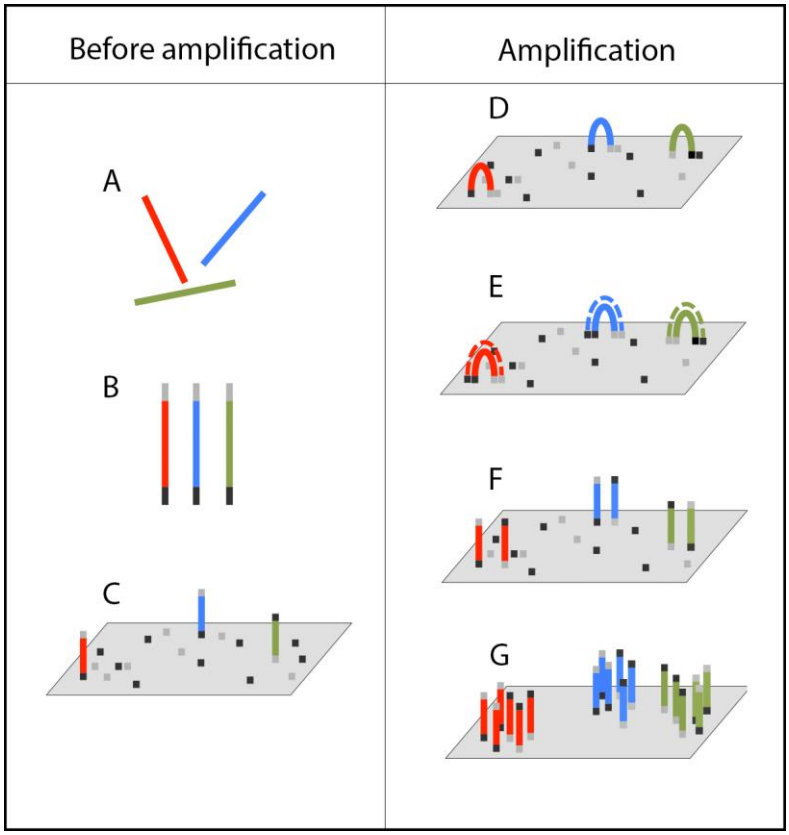


Figure 17. Simplified view of library preparation and bridge-amplification. Before amplification the DNA is (A) fragmented and (B) ligated to adapters at both ends. The fragments with adapters are then (C) hybridized to a flow cell coated with complementary adapters. The fragments will then (D) bend over and the other end of the fragment will bind to another adapter on the flow cell creating a "bridge-like" structure. (E) A DNA polymerase will then amplify the fragments using the adapters as primers. (F) After completed amplification the dsDNA is denaturized into two ssDNA fragments. (G) The amplification is repeated until there are about 1000 copies of each fragment forming clusters on the flow cell (Adapted from [67]).

The sequencing uses reversible terminator nucleotides, each with a different fluorescent dye. The 3'-OH of the nucleotides are chemically inactivated to ensure that only one nucleotide is

incorporated at the time. For each cycle of the sequencing, a reaction mixture containing the four nucleotides, primers, and the DNA polymerase is added to the flow cell [64] (Figure 18A). One nucleotide is then incorporated to each DNA fragment on the flow cell, followed by a washing step (Figure 18B) before the dye of the nucleotide incorporated to each cluster is detected by a CCD camera (Figure 18C). After the imaging step the terminator and the fluorescent dye on the nucleotide are removed, and the cycle can be repeated until the fragments have been fully sequenced (Figure 18D) [64-66].

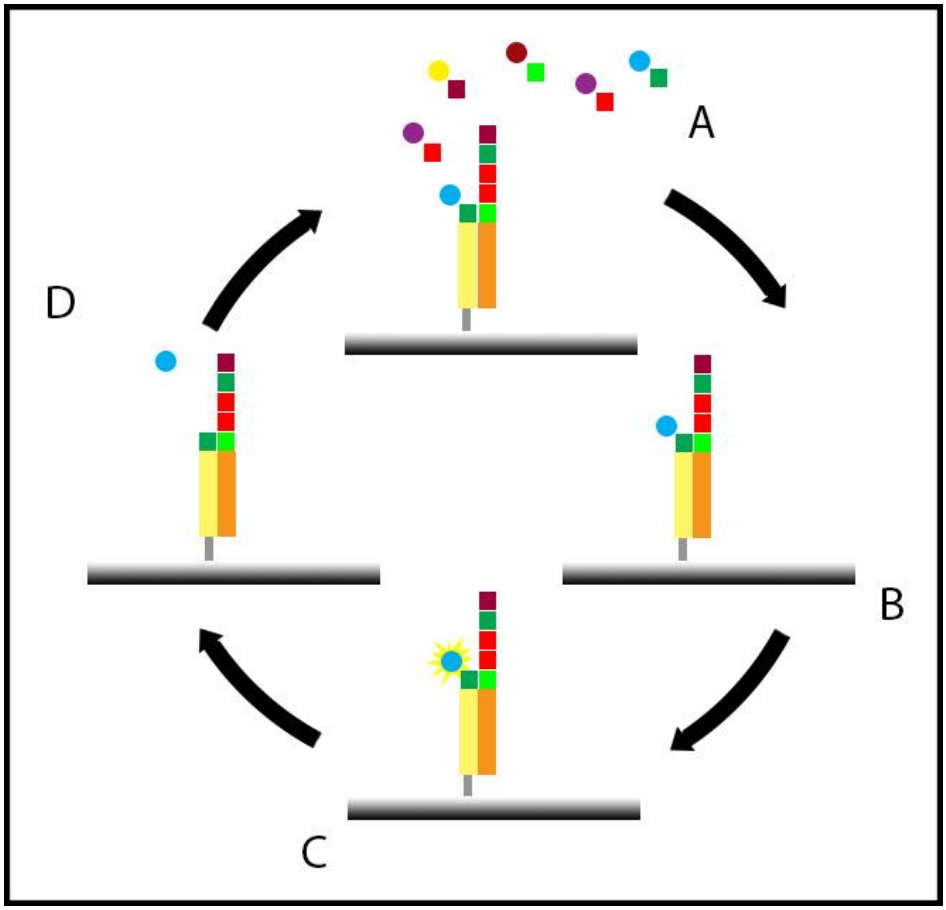


Figure 18. Sequencing by synthesis. Primers, nucleotides and DNA polymerase are added to the flow cell, and a single nucleotide, which is attached to a dye and terminator group, is incorporated to the template (A). This is followed by a wash step (B) before the dye attached to the nucleotide is detected by a CCD camera (C). The terminator group and the dye are then removed from the nucleotide (D) before the cycle can start over and a new nucleotide can be incorporated.

3.4.4 Quality of a run

The quality of a run was determined based on several parameters. One of these parameters was the depth of coverage. The mean of the depth of the *BRCA1* and *BRCA2* genes of all the samples in one run was compared to each sample, and used as a parameter to determine the coverage of each sample.

The gaps of the genes scrutinised were also checked manually using Integrative Genomics Viewer (IGV). Gaps are exons and introns (+/-20) which have less than 30 reads. There were some gaps which were present in all of the samples analysed. *MSH6* was not completely covered by the probes in the kit used in this analysis, and a large part of exon 1 was left out. *CDH1*, *RAD51D* and *STK11* had gaps which were present in most samples. *CDH1* and *RAD51D* had two gaps each. *CDH1* had one gap in the 5'of intron 11 and one in the exon/intron border of exon 2. *RAD51D* had one gap at the 3' of intron 4 and one gap at the start of exon 1. *STK11* could have up to three gaps in a run, all of which were at the exon/intron border of exon 4 (at the 3'end), exon 7 and exon 9 (at the 5' end).

Other parameters included were cluster density (theoretical at 1200 K/mm^2) and alignment (theoretical at 1 %). In addition, the total number of variants found among the 94 genes in the TrueSight Cancer kit, and the number of variants found in *BRCA1* and *BRCA2* was checked in order to determine if there was contamination between the samples in one run.

3.4.5 Cartagenia

After the MiSeq had finished sequencing the DNA fragments, the MiSeq Control Software (by Illumina) aligned the sequenced fragments to a human reference sequence. It also identified variations from the reference sequence (Table 4). In order to find variants which could be of clinical significance, the variations were uploaded into the Cartagenia software v.5.0 (Agilent Technologies). The Cartagenia software provides the possibility to filtrate the detected variants according to different criteria such as bad sequence quality, or high frequency in the normal population. The filtration tree was made according to which genes were scrutinized in the present project, and is illustrated in Appendix C – Filtration tree (Figure 26).

In the present project, variants in 18 genes, including *BRCA1* and *BRCA2* (Table 4), were further investigated. *BRCA1* and *BRCA2* were included in the gene filter in order to make sure that no pathogenic variants or VUSs were present in these genes.

Table 4. Genes scrutinized in this study, including BRCA1 and BRCA2.

Gene	*Reference sequence*
BRCA1	NM_00724.3
BRCA2	NM_000059.3
TP53	NM_000546.5
PTEN	NM_00314.4
CDH1	NM_004360.3
STK11	NM_000455.4
NF1	NM_001042492.2
PALB2	NM_024675.3
ATM	NM_000051.3
CHEK2	NM_001005735.1
NBN	NM_002485.4
BRIP1	NM_032043.2
RAD51C	NM_058216.1
RAD51D	NM_002878.3
MLH1	NM_000249.3
MSH2	NM_000251.2
MSH6	NM_000179.2
PMS2	NM_000535.6

Variants in the 18 genes were filtered using several quality filters (Appendix C – Filtration tree, Figure 26). The call quality >30 filter removes the variants that were identified for less than 30 % of the reads at the specific location. The quality read depth >18 filtered out variants with a read depth below 18. The genotype quality filter removed false genotypes. The R8 quality filter removed variants located after eight or more repeated mononucleotides and the standard bias filter, and variant allele frequency filters removed reads where the reads containing the allele were lower than the cut-off to prevent artefacts. The variants which did not get through the quality filters were not included in this study.

The variants which passed the quality filters were subsequently removed if they occurred with a frequency > 1 % in population databases (ESP, ExAC, 1000 Genomes, dbSNP, which will be described in 3.4.6.1 Alamut Visual). The remaining variants were sorted according to the effect they were predicted to have on their encoded proteins.

3.4.5.1 Criteria for variant inclusion

All the exonic variants which passed the quality filters and had a population frequency beneath 1 % were included in the Cartagenia report. In addition, the intronic variants which were within +2 and -10 of the exon/intron border were included.

3.4.6 Classification of variants

The variants were evaluated and classified in five different classes [68] mainly based on their occurrence in population and patient databases, and their occurrence in literature, as well as their effect on the protein.

Class 1 – Benign variant

Variants in this class are not of clinical importance, and are sufficiently documented in databases or in literature. These variants are often been reported at a high frequency in population databases.

Class 2 – Likely benign variant

These variants are unlikely to be of clinical importance, but limited documentation is available. Variants in this class may occur together with pathogenic variants in the same gene, or they can be synonymous variants or intronic variants with no effect on splicing after performed cDNA analysis. The variants in class 2 are often reported in population databases and often have a high frequency.

Class 3 – Variant of unknown clinical significance (VUS)

This class includes variants which cannot be placed in any of the other classes, and are not frequent in the population or documented in the literature. These variants are often missense variants and possible splice variants (located outside +/-1, 2). The variants can have conflicting predictions according to the different prediction programs applied.

Class 4 – Likely pathogenic variant

Variants in this class are likely of clinical importance, but the documentation is lacking or insufficient for class 5. This class includes nonsense variants, frameshift variants, splice variants (located +1, +2, -1, -2), and larger duplications and deletions. The variants can still be reported in population databases but at very low frequencies.

Class 5 – Pathogen variant

This class includes variants which are of clinical importance, and are well documented in the literature or in patient databases. The variant is classified as pathogenic if it has occurred in several unrelated patients with the same phenotype, and if the variant has been found to

segregate with the disease in several families. The variant can still be reported in population databases, but at a very low frequency.

3.4.6.1 Alamut Visual

Variants remaining after filtration in Cartagenia were interpreted using Alamut® Visual v.2.7-rev.1 (©2015 Interactive Biosoftware). Alamut® Visual is a support software that integrates information from different sources to help with the interpretation of sequence variants. The software described the gene variants according to the HGNC- (Hugo Gene Nomenclature Committee) and HGVS-(Human Genome Variation Society) nomenclatures [69].

The databases scrutinized by Alamut Visual for the specified nucleotide changes are: population databases, such as NCBI dbSNP (Single nucleotide polymorphism database), ExAC (Exome Aggregation Consortium) and ESP (EVS – NHLBI Exome Sequencing Project); and other databases like HGMD (Human Gene Mutation Database) professional, ClinVar (a public collection of reported human sequence variants and phenotypes, with supporting evidence) and SwissProt/UniProtKB (a curated protein sequence database).

There are also different prediction tools available in Alamut Visual. The following prediction programs are used for missense variants. Align GVGD (Grantham Variation, Grantham Deviation) compares the biochemical variations between the original amino acid and the new amino acid, and estimates whether the change will affect the function of the protein. SIFT (Sorting Intolerant from Tolerant) predictions are based on the degree of conservation of the amino acid residues in sequence alignment derived from closely related sequences. MutationTaster evaluates conservation, changes in splice sites, changes in protein properties, and changes in translation in order to predict if the changes in amino acids can cause disease [70]. PolyPhen-2 (Polymorphism Phenotyping v.2) evaluates the change in amino acid caused by the variant based on the predicted changes in structure and function of the evaluated protein.

Alamut Visual also provides the following tools which predict if the variant can have an effect on the splice sites: MaxEntScan (MES), Neural Network Splice Site Prediction (NNSPLICE), Splice Site Finder-like (SSF-like), Human Splicing Finder (HSF), Gene

Splicer, Exonic Splicing Enhancers finder (ESE finder) and Relative Enhancer and Silencer Classification by Unanimous Enrichment-ESE (RESCUE-ESE). MaxEntScan, NNSPLICE and SSF-like predict acceptor- and donor sites by using knowledge on the base composition of the splice site sequence. This knowledge is used to run an algorithm which generates a scoring matrix for acceptor- and donor sites. SSF-like also generates a score matrix for branch points [71]. Human Splicing Finder predicts and assesses the strength of acceptor- and donor sites based on position weight matrices, but it seems to predict splice sites at a higher frequency than the other programs and is therefore not put much weight on during evaluation of the splice-signals [72]. Gene Splicer considers a small window around the splice junctions by combining several splice site detection techniques [73]. Each of these programs generate a score, and the higher the score the more certain the prediction is [71]. The two remaining tools, ESE finder (which looks for binding sites for ESE) and RESCUE-ESE (which identifies specific hexamer sequences as candidate ESEs), as well as the branch point predictions, are still too unspecific to be used as a the main basis for further investigation of a variant [71].

An overview of the functional protein domains for each gene is also included in Alamut, which makes it possible to determine the variant's position in regard to functional domains.

The genome Aggregation Database (gnomAD) is a population database, which is not connected to Alamut Visual, but was used along with ExAC and ESP for the classification of variants.

3.5 Verification of variants

In order to verify variants identified using NGS and classified as likely pathogenic or pathogenic, PCR products encompassing the detected variant were sequenced by Sanger sequencing.

3.5.1 Polymerase chain reaction (PCR)

The patient samples which were to be Sanger sequenced, were first amplified using polymerase chain reaction (PCR) (Figure 19). The method is based on DNA replication and requires a template (DNA in the sample), DNA polymerase, all four deoxynucleotides and two oligonucleotide primers [74].

DNA polymerase is an enzyme which incorporates the correct deoxynucleotide complementary to the DNA template. The two primers are complementary to the 3' ends of the two DNA strands of the DNA template. The primers help the DNA polymerase to amplify the specific sequence [10]. The PCR can be divided into three steps which are repeated as many times as is required, giving millions of copies of the same sequence (Figure 19) [74]. In the first step of a PCR the sample is heated to 90-95°C to denature the dsDNA, forming ssDNA strings (Figure 19A). In the next step, the temperature is lowered to about 55°C, allowing the primers to anneal to the ssDNA (Figure 19B). The exact annealing temperature is dependent on the primers and their optimal temperature for the primers to hybridize to the DNA. In the third and final step the sample is heated again, this time to about 72°C, which is the optimal temperature for the DNA polymerase. In this step the DNA polymerase synthesizes the new DNA strands by extending the primers (Figure 19C).

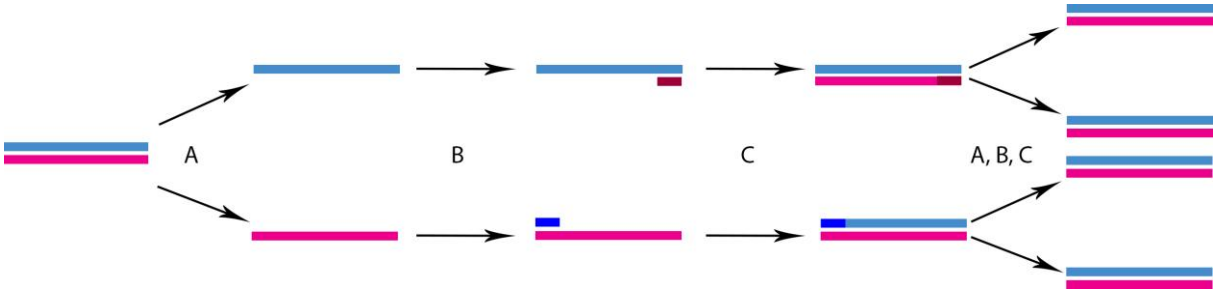


Figure 19. PCR amplification. The dsDNA is denatured forming two ssDNA (A) followed by the annealing of primers to the ssDNA (B), and the DNA polymerase synthesises a new DNA strand (C). Each PCR cycle contains these three steps, and for each cycle the number of DNA strands doubles.

The reaction mix used is illustrated in Table 5. This table shows the amount of the different ingredients required for one sample.

Table 5. PCR reaction mix for each sample

	µL for each sample
Primer 1	0.5
Primer 2	0.5
Jumpstart Redtaq Ready-mix	5.5
Sterilised H_2O	7.5
Volume reaction mix	14.0
DNA sample	1.0
Total volume	15.0

The Jumpstart Redtaq Ready-mix (Sigma Aldrich) contains the DNA polymerase (from *Thermus aquaricus*) [74] and all four of the deoxynucleotides, as well as a red staining dye which is used during capillary gel electrophoresis. The two primers used the forward and reverse primers, are specific for the DNA sequence to be amplified.

For the verification of NGS findings by Sanger sequencing we used primers, and PCR programs, which are shown in the Appendix A – Primers (Table 13) and Appendix B – PCR-programs (Table 19).

3.5.2 Agarose gel electrophoresis

PCR-products were analysed using agarose gel electrophoresis. Both PCR products for verification of variants which were identified by NGS, and PCR products for the microsatellite analysis, were assessed by agarose gel electrophoresis.

Gel electrophoresis is used to separate macromolecules according to size, where the smallest molecules migrate fastest [75]. The gel used contains a microscopic network of pores, where the DNA fragments migrate through when a voltage is applied across the gel. Since DNA is negatively charged, the fragments will migrate towards the positive electrode. Depending on the size of the fragments, the fragments will spread across the gel accordingly [10]. A ladder with DNA fragments with known sizes is used to decide the fragment sizes for the sample analysed.

An agarose gel was made using UltraPure™ Agarose (Invitrogen by life technologies) and 1×TAE (Tris-acetate-EDTA) buffer (from 10×TAE buffer (Invitrogen by life technologies)). For the PCR products intended for Sanger sequencing a 0.8 % agarose gel was used. The agarose gel solution was boiled until the agarose powder had dissolved in the 1×TAE buffer. Subsequently, 100 mL slightly cooled agarose was poured into a casting mould.

The PCR products and a 1kb plus DNA ladder (Invitrogen by life technologies) were added to wells, created by a well comb, in the polymerized gel. After the samples and ladder had been added to the gel, a voltage (90 V) was applied across the gel for an appropriate time (35 minutes – 75 minutes).

In order to visualize the bands of DNA fragments, the gel was stained using GelRed™ Nucleic Acid Gel Stain (Biotium) mixed with 0.1 M NaCl (800mL 0.1 M NaCl + 200 µL GelRed). DNA bands were then visualized using ultraviolet-light (UV-light) in the BioDoc-It™ – Imaging system UVP (Ultra-Violet Production Ltd.).

3.5.3 Sanger sequencing

Before Sanger sequencing was performed, the PCR-product was purified using an A'SAP PCR purification kit (ArticZymes). 1 µL A'SAP solution was added to 11 µL of PCR-product which was vortexed and spun down, before the A'SAP PCR-program was used (Appendix B –PCR-programs, Table 20).

Sanger sequencing uses dideoxynucleotides (ddNTPs) to determine the sequence of a DNA fragment. The sequencing reaction requires one primer (either forward or reverse), a template, DNA-polymerase, deoxynucleotides (dNTPs) and ddNTPs, where each of the four different ddNTPs has different dyes attached. ddNTPs are missing a 3' OH-group which is vital for further incorporation of other nucleotides. When a ddNTP is incorporated the DNA-synthesis will be terminated. Since the incorporation of ddNTP happens by chance, the products from the sequence reaction will differ in length. Subsequently, the sequencing products are separated using capillary gel electrophoresis before the different dyes are detected, giving the sequence of the DNA fragment. The method can only sequence ssDNA in the same direction in one sample. This means that the forward fragments and the reverse fragments must be separated [75].

After the PCR-products had been purified, 2 µL of the sample was used in the sequencing mix (Table 6) containing either the forward or the revers sequencing primer. The primers used for the amplification (Appendix B – Primers, Table 13) have all attached a M13 tail. The same primers can therefore be used for sequencing all the fragments. The M13 tail is different for the forward and the reverse primers.

Table 6. The reagents needed for sequencing products with M13-tails.

Reagents	*Volume per sample*
M13-primer F/R (Life by Invitrogen)	0.5
BigDye v3.1	0.5
5xsequencing buffer for BigDye v3.1	3.0
H_2O	14.0
Total volume	18.0

Subsequently, the samples were run through a sequencing PCR-program, which is illustrated in the Appendix B – PCR-programs (Table 21).

3.5.4 Capillary gel electrophoresis

Capillary gel electrophoresis was used to separate DNA fragments of different size for Sanger sequencing. It was also used to determine the size of the fragments analysed during the RNA analysis (3.6.1 RNA analysis) and the microsatellite analysis (3.6.2 Microsatellite analysis). The principle is similar as for agarose gel electrophoresis. The main difference is the accuracy of the fragment size as capillary electrophoresis can detect single nucleotide differences. The fragments are detected by a CCD camera near the end of the capillaries due to a fluorescent dye attached to the fragments [75]. For Sanger sequencing the different terminal nucleotides have attached one of four fluorescent colours, which are detected and decides the sequence of the fragment.

3.6 *BRCA2* c.8331+2C>T

3.6.1 RNA analysis

In order to determine if the *BRCA2* c.8331+2C>T variant affects the RNA splicing process of the *BRCA2* mRNA in blood from PAXgene Blood RNA Tubes (PreAnalytiX by QIAGEN/BD company).

3.6.1.1 RNA-extraction

RNA was extracted from the patient sample using QIAsymphony SP (see 3.2 DNA-extraction) and the QIAsynphony® PAXgene® Blood RNA kit (96) (PreAnalytiX by QIAGEN/BD company).

3.6.1.2 cDNA synthesis

The mRNA extracted from the sample was used as a template to make complementary DNA (cDNA), which is more stable than RNA. In order to make cDNA from RNA, the SuperScript® VILO™ cDNA Synthesis kit (Invitrogen by Thermo Fisher Scientific) was used according to the manufacturer's protocol. The SuperScript® VILO™ cDNA Synthesis kit utilize a reverse transcriptase, along with random non-specific hexamer primers and dNTPs.

In brief, a reaction mix was made (Table 7) and the PCR-program is illustrated in Appendix B – PCR-programs (Table 22). The SuperScript® VILO™ buffer contains primers and dNTPs, while the SuperScript® enzyme mix contains reverse transcriptase.

Table 7. The reaction mix for one sample used for the cDNA synthesis.

	Volume (µL)
SuperScript® VILO™ buffer	4
10× SuperScript® enzyme mix	2
H_2O (free of RNase)	9
RNA (up to 2.5µg)	5
Total	20

As a control for the cDNA synthesis and residual of gDNA contamination, a PCR was set up for *APRT* (Adenine phosphoribosyl transferase) (see 3.4.1 Polymerase chain reaction). *APRT* is ubiquitously expressed in every cell, and can therefore be used as a control. The PCR reaction mix is described in Table 5. The PCR-program used is listed in Appendix B – PCR-programs (Table 23) and the primers used can be seen in Appendix A – Primers (Table 14). The PCR-product was analysed by gel electrophoresis (see 3.5.2 Agarose gel electrophoresis), using a 1 % gel. If any gDNA was present in the sample, the gel would show two bands, a cDNA band (218 bp) and a genomic band including intron (721 bp).

The primers used for the analysis of *BRCA2* c.8331+2C>T, which aligns to exon 17 (forward) and exon 19 (reverse), are shown in Appendix A – Primers (Table 15). The PCR was performed as described in "3.5.1 Polymerase chain reaction". The PCR-program used is illustrated in the Appendix B – PCR programs (Table 24).

Subsequently, the PCR-products were analysed by agarose gel electrophoresis (3.5.2 Agarose gel electrophoresis) before being analysed using fragment analysis as described in "3.6.2 Microsatellite analysis" and "3.5.4 Capillary gel electrophoresis". The expected length of the cDNA without skipping of exon 18 was 540 bp. The expected length of cDNA with exon skipping was 185 bp, since exon 18 consists of 355 bp [11].

3.6.2 Microsatellite analysis

After the DNA samples for microsatellite analysis had been quantified using Nanodrop (3.3.3 Nanodrop), they were diluted to a concentration of 50 ng/µL before being amplified using PCR as described in "3.4.1 Polymerase chain reaction", and the reaction mix in Table 5. The primers used for the amplification are listed in Table 8, while the PCR-program is listed in Appendix B – PCR-programs (Table 25). The primers were used to amplify microsatellites located on chromosome 13q12.3 –q13.2.

Table 8. Primer setts used in the microsatellite analysis. The primers are listed in 5' to 3' direction.

Primers	*Forward*	*Reverse*	*Product size (bp)*
BRCA2.1m	CATGGGAAAAGGGTGTTTTT	GGCTCACATTGAAGCACAAA	209
BRCA2.2m	TGGTTTTAAAAAGAAGACCGTTA	TCAAAGGATGATCCCATTATAACA	200
BRCA2.3m	GCCTGGAAGCTACGAAGATG	CCCCAAATATCATTACCGTGA	238
BRCA2.4m	CTGGAGTGCAGTGGTGTGAT	GGAGGGTAGGAGGGTGAAGA	451
BRCA2.5m	GGGGACTTGTTTGGAGGAAT	CCTCCCACCACAGACTCAAT	236
BRCA2.6m	AGGCTCAGAGGAGAAATCAGA	TAAGGAGAAGCCAGGCCAAG	179
BRCA2.7m	CCCTTTCATGAAGGAAAACTACA	GGGCTATTCCTCAAGGTGTTT	290
BRCA2.8M	CCTCAATGGATGCTGAAACC	AACCCTTCTTGCCTTCCCTA	340
D13S290	CCTTAGGCCCCATAATCT	CAAATTCCTCAATTGCAAAAT	176-194
D13S1246	GAGCATGTGTGACTTTCATATTCAG	AGTGGCTATTCATTGCTACAGG	177-213
D13S260	AGATATTGTCTCCGTTCCATGA	CCCAGATATAAGGACCTGGCTA	158-173
D13S267	GGCCTGAAAGGTATCCTC	TCCCACCATAAGCACAAG	148-162
D13S220	CCAACATCGGGAACTG	TGCATTCTTTAAGTCCATGTC	191-203

Amplified samples were analysed using gel electrophoresis in order to see if the PCR had been successful (3.5.2 Agarose gel electrophoresis). For PCR products used for microsatellite analysis a 1 % agarose gel was used. The fragment size was then decided using capillary gel electrophoresis (see 3.5.4 Capillary gel electrophoresis). In order to be able to decide the size of the fragments an internal standard (ROX GeneScan™ 500 ROX™ dye Size Standard by Thermo Fisher Scientific) was used for the fragment analysis.

The PCR products amplified for microsatellite analysis were diluted 1:5, if necessary. Thereafter 0.7 µL of each diluted sample were added to 9.4 µL ROX-pool. The ROX-pool is

made from 850 µL HiDi Formamide (Life Technologies by Thermo Fisher Scientific) and 50 µL 500 ROX™.

After the PCR products and the ROX-pool were mixed, the plate was vortexed and centrifuged for 1 minute. Subsequently, the samples were denatured for 3 minutes at 86°C and put on ice for 2 minutes. The fragments were then analysed by capillary gel electrophoresis (3.5.4 Capillary gel electrophoresis).

3.6.2.1 GeneMapper

In order to be able to interpret the results from the fragment analysis the software GeneMapper® Software 5 (Thermo Fisher Scientific) was used according to the manufacturer's protocol. The analysis method was changed to Microsatellite default, and Size Standard was set to GS500ROX. The fragment files were then analysed and fragment sizes were registered. A pedigree was made for each family illustrating the length of the different fragments/repeated sequences for each family member (Appendix E – *BRCA2* c.8331+2C>T results, Figures 35-40).

3.6.3 DMLE – Disease mapping using linkage disequilibrium

In order to estimate the age of the *BRCA2* c.8331+2C>T variant in the investigated Norwegian cancer population, the DMLE +2.2 software [76] was used. This method compares the differences in linkage disequilibrium between the mutation and flanking markers in the DNA samples from mutation carriers and controls by using a Bayesian approach [30, 77]. The software compares the samples based on the following parameters [30]: observed haplotypes or genotypes in normal and affected groups; distance between markers and mutation site in Morgans (M); estimated population growth rate and an estimated proportion of mutation bearing chromosomes. The software is available online and was used according to the manufacturer's protocol.

The size of the microsatellites and genotypes for family members with the mutation was used for this purpose. The haplotypes of the family members without the mutation were set as a normal population. The program only uses the normal population to check against normal allele counts. Consequently, the genotypes are exactly equivalent to haplotypes. Since the

software only allows haplotype data for the normal population, the genotype data was split into any of the possible haplotypes. This means that one genotype gives two haplotypes.

The different sizes of the fragments were given integer numbers, as required by the software. Each marker, in total five, had the fragments numbered from 1 to n, where n was the total number of different alleles (fragment sizes). There could be no gaps in the numbering system, which included both the normal population and the patient population.

The map distance between markers and the mutation site, was calculated from the chromosome positions of the microsatellites and the *BRCA2* gene, with the first marker (D13S290) being placed at 0.0 (Table 9). The distance was given in Morgans (M), given that 1cM=1000kb.

Table 9. The map distance between markers and mutation site, given in bp and M, as required by the software.

Marker	*Distance (bp)*	*Distance (M)*
D13S290	0	0.0
BRCA2.6m	1232944	0.01233
BRCA2.1m	1482874	0.01483
BRCA2	1784180	0.01784
BRCA2.5m	2219549	0.02220
D13S267	2584545	0.02585

In order to estimate the age of the mutation the growth rate (r^{gen}) had to be calculated. The equation below was used for this purpose [78].

$$r^{gen} = \frac{log \frac{P_t}{P_0}}{g}$$

4 Results

4.1 NGS

Forty-eight DNA samples from patients with breast and/or ovarian cancer were sequenced using NGS. In thirty-one (64.6 %) of the patient samples, sequence variants were identified among the 16 genes scrutinised in this study. In these 31 patients a total of 36 different variants were detected in 12 genes (Table 11). In addition, a more extensive assessment of these identified sequence variants, including their localisation in the corresponding gene, their protein associated change, the number of times the variants were detected in the current study, their occurrence in a population database (gnomAD), their classification in our study and in ClinVar, and previous references where these variants have been described is summarized in Table 10. The predictions from the prediction programs are listed in Appendix D – NGS results, Table 26. Eleven of the currently investigated patients had more than one variant among the 16 genes scrutinised. None of the 48 patients had sequence variants in *PTEN*, *STK11*, *NBN* and *RAD51D*.

Three different deleterious variants were found in five patients (10.4 %), two were pathogenic variants (class 5, *TP53* c.818G>A and *ATM* c.3245_3247delinsTGAT) and one was a likely pathogenic variant (class 4, *CHEK2* c.319+2T>A). In addition, ten variants classified as VUS (class 3) were recognized in ten different patients (20.8 %). These ten VUSs were found in *ATM*, *BRIP1*, *CDH1*, *MLH1*, *NF1* and *PMS2*. *BRIP1* c.2087C>T (VUS) and *PMS2* c.1437C>G (VUS) were identified twice in four different patients. The genomic localization of the sequence variants classified as class 3, 4 and 5 are illustrated in Figure 23. The remaining 32 variants (68.1 %) of which four (*ATM* c.5071A>C, *PALB2* c.2794G>A, *PMS2* c.52A>G and *PMS2* c.1569C>G) were each recognized in 2 patients and one (*MSH6* c.3246G>T) was seen in three patients, were classified as benign variants. In total, three class 1 variants and 29 class 2 variants with no increased cancer risk were identified.

4.1.1 Deleterious variants

In one patient a single sequence variant in *TP53* was discovered. The *TP53* c.818G>A p.(Arg273His) variant is a missense variant located in exon 8 (Figure 23H) at one of the mutation hotspots of *TP53* (Figure 5) [27, 79-81] and introduce a change in aa 273 from Arg to His in p53. The variant has been reported in several studies. The prediction programs classified the variant as a deleterious variant (Appendix D- NGS results, Table 26). The *TP53*

c.818G>A variant was reported in the population databases, but with a very low frequency (0.0016 % in gnomAD, Table 10). The variant, which was found in heterozygous form (32.7 %) in patient sample TSC13, was classified as a class 5 variant (Table 11). The variant was verified using Sanger sequencing by using primers which annealed at intron 7 (forward) and intron 8 (reverse). The results from the sequencing, illustrated in Figure 20, shows that the variant identified by NGS was not an artefact.

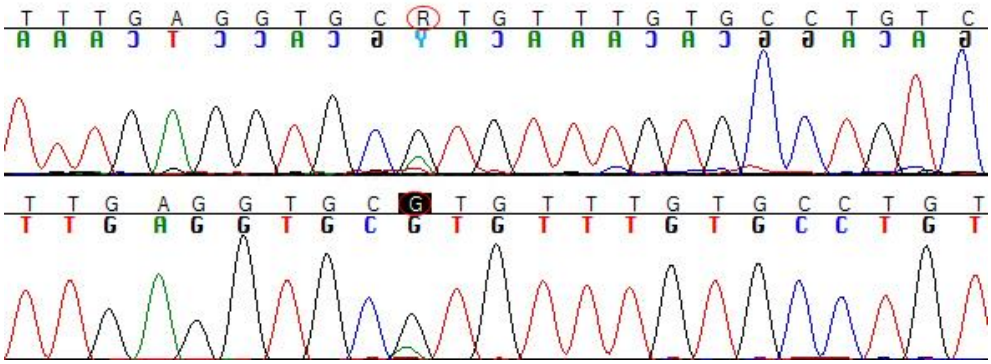


Figure 20. Sanger sequencing of part of exon 8 of the TP53 gene from patient sample TSC13, where the TP53 c.818G>A variant was identified. Both the reverse and the forward sequence are illustrated. The site of the variant is marked in red.

The *ATM* c.3245_3247delinsTGAT (p.His1082Leu*14) variant in exon 22 (Figure 23A) changes the reading frame and results in a stop codon after 13 aa. The mutated protein, if produced, would not have the kinase domain shown in Figure 4. The variant had been reported in literature [14, 18, 82], but had not been reported in gnomAD (Table 10). Two patient samples (TSC2 and TSC31) in our studied cohort were heterozygous for the *ATM* c.3245_3247delinsTGTA variant. The variant was verified using Sanger sequencing on patient sample TSC2, where the primers were chosen to anneal to the transition between intron 20 and exon 21 (forward) and to intron 22 (reverse). The result from the sequencing, illustrated in Figure 21, showed that the variant identified by NGS was not an artefact.

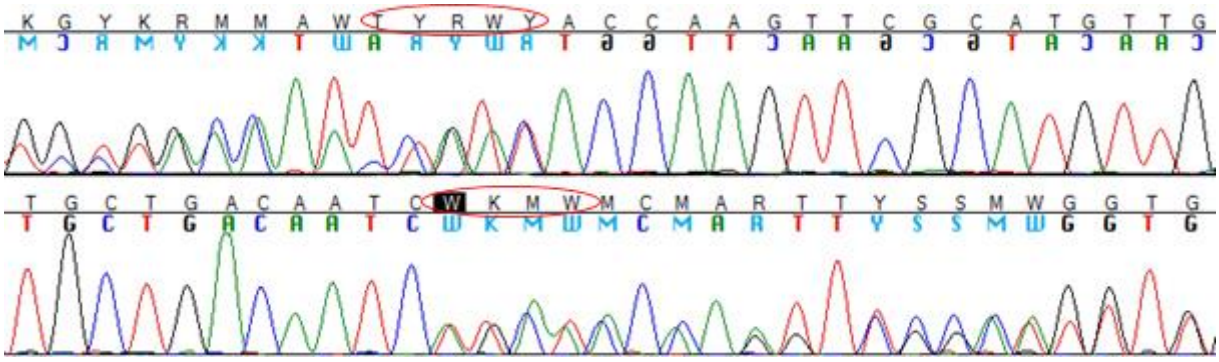


Figure 21. Sanger sequencing of part of exon 22 of the ATM gene in patient sample TSC2, where the ATM c.3245_3247delinsTGAT was found. Both the reverse and the forward sequence are illustrated. The site of the variant is marked in red.

CHEK2 c.319+2T>A in intron 2 (Figure 23D) was predicted to have an effect on the splice site of exon 2 (Appendix D – NGS results, Table 26). The variant had not been reported in literature or in any of the population databases. The variant was found in heterozygous form in two patient samples (TSC12 and TSC16). The variant, which was classified as a class 4 variant, was verified using Sanger sequencing on patient sample TSC12, where the primers were chosen to anneal to exon 2 (forward) and intron 2 (revers). The result for the sequencing, illustrated in Figure 22, showed that the variant identified by NGS was not an artefact.

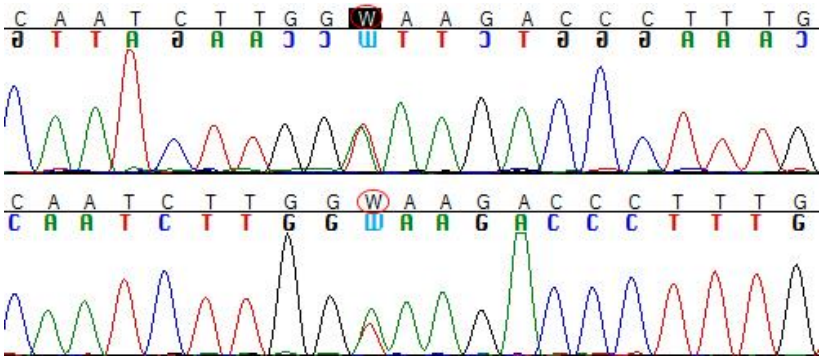


Figure 22. Sanger sequencing of part of exon 2 and intron 2 of the CHEK2 gene in patient sample TSC12, where the CHEK2 c.319+2T>A variant was identified. Both the reverse and the forward sequence are illustrated. The site of the variant is marked in red.

4.1.2 Variants of unknown clinical significance

Patient TSC14 was shown to be heterozygous for *ATM* c.8734A>G p.(Arg2912Gly). This missense variant located in *ATM* exon 60 (Figure 23A) changes aa 2912 from Arg to Gly. The variant had previously been reported in the literature as a class 3 variant [15, 20, 83]. One study estimated this variant to give a 7-fold increased risk of breast cancer [84]. Although

various prediction programs classified *ATM* c.8734A>G as a deleterious variant (Appendix D – NGS results, Table 26), it was found in the population databases with a frequency of 0.02 % (gnomAD, Table 10). Moreover, one individual was shown to be homozygous for this variant. The variant was classified as a VUS in this study.

The *BRIP1* c.2087C>T p.(Pro696Leu) variant is located in exon 14 (Figure 23B), between two of the eight motifs of the ATPase helicase core domain (Figure 7). Two patients (TSC31 an TSC32) were heterozygous for *BRIP1*c.2087C>T. This missense variant, which changes aa 696 from Pro to Leu, had not been reported in the literature, but had been reported in gnomAD with a frequency of 0.0047 %. According to the prediction programs this was a deleterious variant (Appendix D – NGS results, Table 26). The variant was classified as a VUS.

The *CDH1* c.136C>G p.(Leu46Val) missense variant n exon 2 (Figure 23C) changed aa 46 from Leu to Val. The variant has not been reported in the literature or in the population databases. According to the prediction programs this was a tolerated variant (Appendix D – NGS results, Table 26). This *CDH1* c.136C>G variant, which we classified as a VUS, was found in heterozygous form in patient sample TSC12, along with *CHEK2* c.319+2T>A (Table 11).

MLH1 c.-7C>T at the 5'UTR (Figure 23E) has been described in the literature as a possible pathogenic variant [85]. Additionally, the variant has been reported in gnomAD with a frequency of 0.13 % (Table 10), and the splice prediction programs did not show any significant changes in splice sites (Appendix D – NGS results, Table 26). The variant, which was classified as a VUS, was found together with *CHEK2* c.319+2T>A in patient sample TSC16 (Table 11).

Three VUSs were found in *NF1*, neither of which was reported in the literature. Variant c.5225A>G p.(Ans1742Ser) is a missense variant which is located in exon 37 (Figure 23F), and changes aa 1742 from Asn to Ser. The variant, which was found together with the *ATM* c.3245_3247delinsTGTA variant in patient sample TSC2 (Table 11), has been reported in gnomAD with a frequency of 0.0029 % (Table 10). The prediction programs indicated that *NF1* c.5225A>G was a tolerant variant (Appendix D – NGS results, Table 26).

NF1 c.587-10_587-9delTT is located in intron 5 (Figure 23F), and is a possible splice variant as the splice site prediction programs predicting a weakened splice site at the start of exon 6 (Appendix D – NGS results, Table 26). The variant has not been reported in population databases. *NF1* c.587-10_587-9delTT was found in patient sample TSC15 (Table 11).

The *NF1* c.7354C>T p.(Arg2452Cys) variant, found in patient sample TSC1, is located in exon 50 (Figure 23F). This is a missense variant which change aa 2452 from Arg to Cys. The variant have been reported in gnomAD with a frequency of 0.00081 % (Table 10). The variant was predicted to be deleterious by the prediction programs (Appendix D – NGS results, Table 26).

PMS2 c.1437C>G p.(His479Gln) is a missense variant found in two patient samples (TSC4 and TSC28). The variant, which is located in exon 11 (Figure 23G), changes aa 479 from His to Gln. The variant has been reported with a frequency of 0.46 % in gnomAD (Table 10), and has also been reported in the literature [37, 86]. According to the prediction programs this was a tolerated variant (Appendix D – NGS results, Table 26). Currently, we classified this variant as a class 3 variant.

Table 10. Sequence variants found during NGS summarized by gene, nucleotide change, localisation, amino acid change, classification (class), number of patients with the variant among 48 patients investigated in this study, frequency in gnomAD, classification in ClinVar, and literature references for each variant

Gene	*cDNA*	*Localisation*	*Associated Protein Change*	*Class*	*No. of patients with variant*	*Frequency in gnomAD*	*Classification in ClinVar (cl.)*	*Reference*
ATM	c.162T>C	Exon 3	p.(Tyr54Tyr)	2	1	0.18%	1	[87, 88]
ATM	c.2572T>C[a]	Exon 17	p.(Phe858Leu)	1	2	0.86%	1	[89-91]
ATM	c.3245_3247delinsTGAT	Exon 22	p.(His1082Leu*14)	5	2	-	5	[14, 18, 82]
ATM	c.5071A>C	Exon 34	p.(Ser1691Arg)	2	2	0.18%	½	[21, 89, 90]
ATM	c.5229A>G	Exon 35	p.(Thr1743Thr)	2	1	-	2	-
ATM	c.5793T>C	Exon 39	p.(Ala193Ala)	2	1	0.51%	½	[92, 93]
ATM	c.8734A>G	Exon 60	p.(Arg2912Gly)	3	1	0,02 %	3	[15, 20, 83, 84]
BRIP1	c.517C>T	Exon 6	p.(Arg173Cys)	2	1	0.27%	½	[89, 90, 94]
BRIP1	c.577G>A	Exon 6	p.(Val193Ile)	2	1	0.36%	½	[31]
BRIP1	c.2087C>T	Exon 14	p.(Pro696Leu)	3	2	0.0047%	3	-
CDH1	c.136C>G	Exon 2	p.(Leu46Val)	3	1	-	-	-
CDH1	c.345G>A	Exon 3	p.(Thr115Thr)	2	1	0.36%	½	[46]
CDH1	c.699C>T	Exon 6	p.(His233His)	2	1	0.0061%	½	-
CHEK2	c.319+2T>A	Intron 2	-	4	2	0.011%	4/5	-
MLH1	c.-7C>T	5'UTR	-	3	1	0.13%	3	[85]
MLH1	c.803A>G	Exon 10	p.(Glu268Gly)	2	1	0.023%	1[b]/2	[95, 96]
MLH1	c.1379A>C	Exon 12	p.(Glu460Ala)	2	1	0.014%	2/3	[97, 98]
MSH2	c.339G>A	Exon 2	p.(Lys113Lys)	2	1	0.28%	1/3	[36, 99]
MSH2	c.1680T>C	Exon 11	p.(Ans560Ans)	2	1	0.029%	1/2[b]	-
MSH6	c.3246G>T	Exon 5	p.(Pro1082Pro)	2	3	0.33%	1/2[b]	-
NF1	c.378A>G	Exon 4	p.(Glu126Glu)	3	1	-	-	-
NF1	c.528T>A	Exon 5	p.(Asp176Glu)	2	1	0.37%	½	[89, 90, 100, 101]
NF1	c.587-10_587-9delTT	Intron 5	-	3	1	-	-	-
NF1	c.1810T>C	Exon 16	p.(Leu604Leu)	2	1	0.13%	1/2/3	[100]
NF1	c.5225A>G	Exon 37	p.(Ans1742Ser)	3	1	0.0029%	3	-

Table 10 continues

Gene	*cDNA*	*Localisation*	*Associated Protein change*	*Class*	*No. of patients with variant*	*Frequency in gnomAD*	*Classification in ClinVar (cl.)*	*Reference*
NF1	c.5793T>C	Exon 39	p.(Ile193Ile)	2	1	0.0029%	2	-
NF1	c.7354C>T	Exon 50	p.(Arg2452Cys)	3	1	0.00081%	3	-
PALB2	c.2794G>A	Exon 8	p.(Val932Met)	2	2	0.52%	½	[89, 90, 102]
PMS2	c.52A>G	Exon2	p.(Ile18Val)	2	2	0.91%	1/3[b]	[90, 103]
PMS2	c.1437C>G	Exon 11	p.(His479Gln)	3	2	0.46%	1/3[b]	[37, 86]
PMS2	c.1569C>G	Exon 11	p.(Ser523Ser)	2	2	0.62 %	1/2[b]	
PMS2	c.1688G>T	Exon 11	p.(Arg563Leu)	2	1	0.64%	1/2[b]	[37, 89, 90, 101]
PMS2	c.2324A>G	Exon 14	p.(Asn775Ser)	1	1	0.032 %	1[b]/3	
RAD51C	c.376G>A	Exon 2	p.(Ala126Thr)	2	1	0.35%	½	[90]
RAD51C	c.790G>A	Exon 5	p.(Gly264Ser)	2	1	0.17%	1/3	[90, 104-107]
TP53	c.818G>A	Exon 8	p.(Arg273His)	5	1	0.0016%	5	[27, 79-81, 108]

[a]-found both homozygous and heterozygous

[b]-classified by expert panel

4.1.3 Benign variants

A total of 32 variants were classified as likely benign (class 2) or benign (class 1). Of the 32 benign variants, 25 different variants were identified. Two of these variants had already been classified by the Medical Genetics department, University Hospital in Northern Norway. Both variants were found in *PMS2* (c.1569C>G and c.2324A>G) and were classified as class 2 and class 1, respectively. The details of the variants without clinical significance are listed in Table 10, Table 11 and Appendix D – NGS results (Table 26). In *ATM* there were five different benign variants (four class 2 variants, and one class 1 variants) among the investigated patient samples. There were two likely benign variants each identified in *BRIP1*, *CDH1 MLH1*, *MSH2*, *PMS2* and *RAD51C*. One synonymous and one missense class 2 variant was identified in *MSH6* and *PALB2*, respectively. In *NF1* three synonymous and one missense variant were identified, and classified as likely benign.

Table 11. Sequence variants found by NGS in the genes scrutinised in 48 patients included in this study. Variants were classified as class 1, class 2, class 3, class 4 and class 5. The 17 patient samples where no variants were found in the 16 genes investigated together with the four genes without any identified variants were not included in this table.

Patient sample	*TP53*	*CDH1*	*NF1*	*PALB2*	*ATM*	*CHEK2*	*BRIP1*	*RAD51C*	*PMS2*	*MLH1*	*MSH2*	*MSH6*
TSC1		c.699C>T	c7354C>T									
TSC2			c.5225A>G		c.3245_3247delinsTGAT							c.3246G>T
TSC4									c.1437C>G			c.3246G>T
TSC5							c.577G>A					
TSC6									c.1569C>G			
TSC11									c.52A>G			
TSC12		c.136C>G				c.319+2T>A						
TSC13	c.818G>A											
TSC14					c.8734A>G							
TSC15			c.587-10_587-9delTT		c.5071A>C				c.1688G>T		c.339G>A	
TSC16						c.319+2T>A				c.-7C>T		
TSC19				c.2794G>A								
TSC22			c.5793T>C						c.52A>G			
TSC23			c.528T>A		c.5071A>C							
TSC25			c.1810T>C									
TSC26											c.1680T>C	
TSC27					c.162T>C							
TSC28									c.1437C>G			
TSC30								c.376G>A				
TSC31					c.3245_3247delinsTGAT		c.2087C>T		c.2324A>G			
TSC32							c.2087C>T					
TSC33					c.5793T>C					c.1379A>C		
TSC34												c.3246G>T
TSC38				c.2794G>A								
TSC39							c.517C>T					
TSC40								c.790G>A				
TSC42					c.5229A>G							
TSC44		c.345G>A	c.378A>G									
TSC46										c.803A>G		
TSC47					c.2572T>C				c.1569C>G			
TSC48					c.2572T>C							

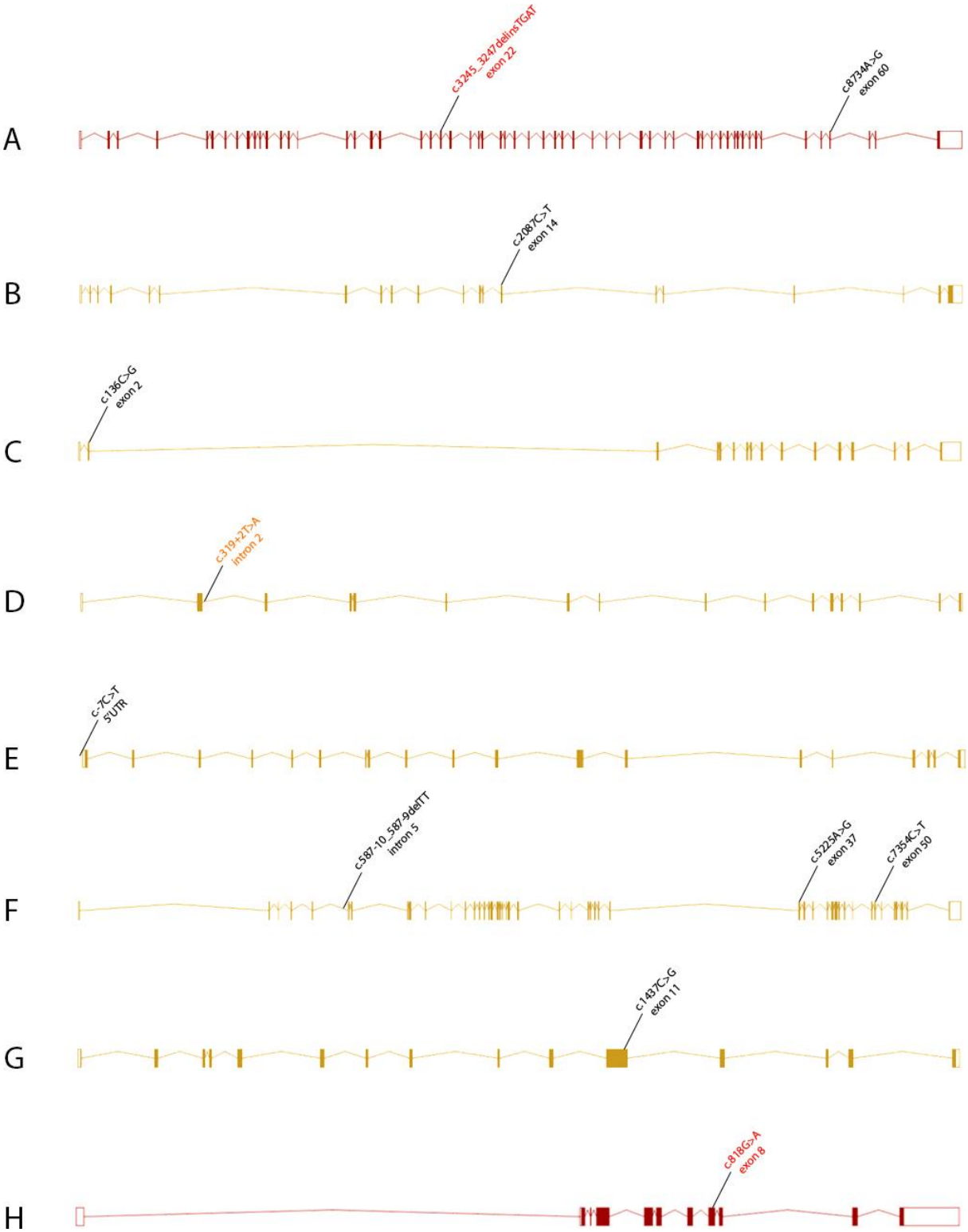


Figure 23. Class 3-5 variants and their position in their respective genes identified in the current NGS study of 48 cancer patients. Class 3 variants are illustrated with black writing, class 4 variants are illustrated with orange writing, and class 5 variants are illustrated with red writing. A) In the ATM gene one variant was found located in exon 22 (c.3245_3247delinsGAT) and one variant was found in exon 60 (c.8734A>G). B) In the BRIP1 gene a class 3 variant was found in exon 14 (c.2087C>T). C) CDH1 had a variant in exon 2 (c.136C>G). D) CHEK2 had a class 4 variant in intron 2 (c.319+2T>A). E) In MLH1 a variant was found in the 5'UTR (c.-7C>T). F) In NF1 class 3 variants were found in exon 4 (c.378A>G), intron 5 (c.587-10_587-9delTT), exon 37 (c.5225A>G), and exon 50 (c.7354C>T). G) PMS2 had one class 3 variant in exon 11 (c.1437C>G). In TP53 a class 5 variant was identified in exon 8 (c.818G>A) (adapted from [11]).

4.2 *BRCA2* c.8331+2C>T

4.2.1 cDNA analysis

The *BRCA2* c.8331+2C>T variant have been found in several families with HBOC in Norway. In order to investigate the effect of the *BRCA2* c.8331+2C>T variant on RNA splicing, a cDNA analysis was performed. Primers were chosen to anneal to exon 17 (forward) and exon 19 (reverse). The hm*APRT* control showed that RNA extraction and cDNA synthesis was successfully performed (Appendix E – *BRCA2* c.8331+2C>T results, Figure 28).

The results from the analysis of *BRCA2* are shown in Figure 24. The figure illustrates two signal tops (at 185 bp and 545 bp) with about the same signal strength for the patient sample with the *BRCA2* c.8331+2C>T variant, while one signal top (at 545 bp) was present for each of the control samples without the variant. The signal at 545 bp represents the full length cDNA fragment, while the signal at 185 bp represents the fragment where exon 18 is skipped.

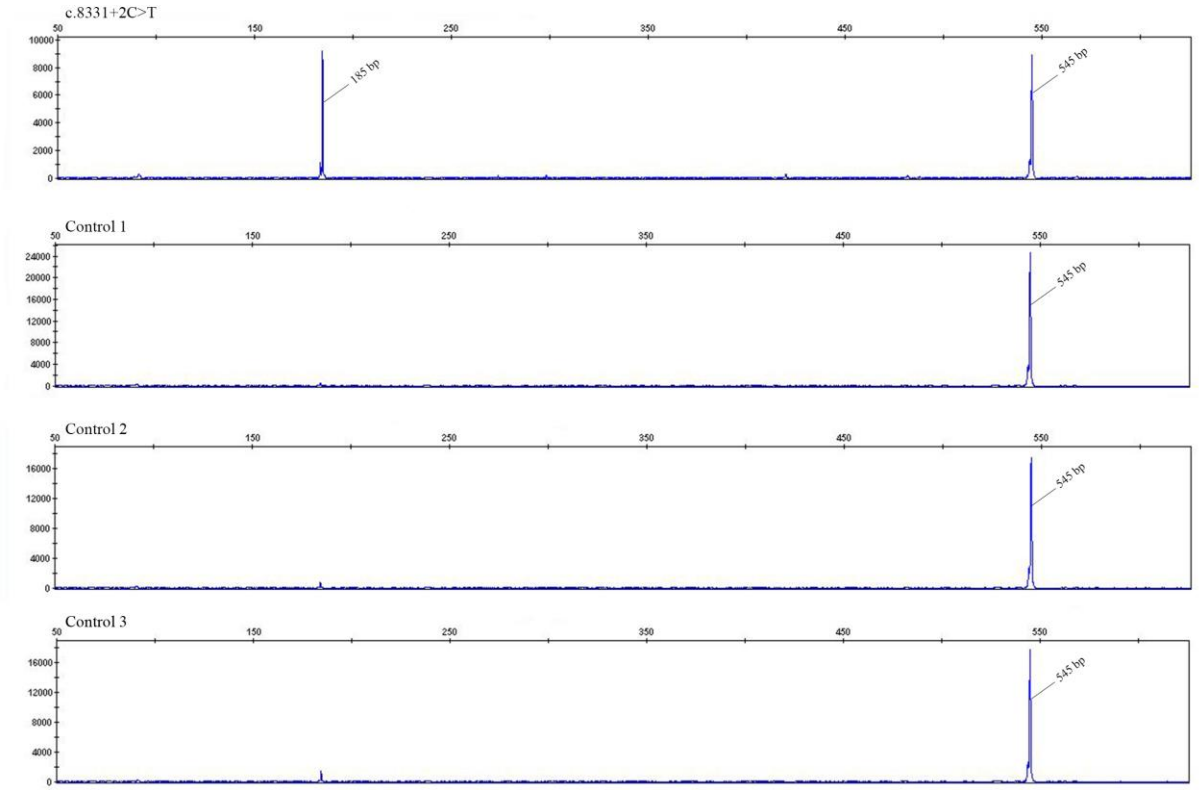


Figure 24. Results from the cDNA analysis. One sample with the BRCA2 c.8331+2C>T variant, and three control samples without the variant were analysed.

4.2.2 Microsatellite analysis

From a total of 29 families with 134 confirmed carriers of the *BRCA2* c.8331+2C>T variant we investigated a selection of 94 samples from 26 different families (Appendix E – *BRCA2* c.8331+2C>T results, Table 27). DNA samples were analysed using fragment analysis to find the genotypes of each individual for several microsatellite markers. Sixty of the samples analysed were from individuals with the *BRCA2* c8331+2C>T variant, while the rest (34 individuals) did not carry the variant. The 13 markers described in the method (3.6.2 Microsatellite analysis) were tested on several patient samples, and we found that primer sets BRCA2.2m, D13S260, BRCA2.7m and D13S220 did not give interpretable results after fragment analysis, and primer sets BRCA2.3m, BRCA2.4m, BRCA2.8 and D13S1246 were not informative. Accordingly, the results of these markers were not used further after testing. The remaining primer sets included: BRCA2.1m, BRCA2.5m, BRCA2.6m, D13S267 and D13S290.

An representative of the agarose gels, as well as fragments electropherogram depicted in the GeneMapper software, are illustrated in Appendix E –*BRCA2* c.8331+2C>T results (Figures 30-34).

The identified microsatellite fragment sizes (Appendix E –*BRCA2* c.8331+2C>T results, Table 27) were included in the pedigrees to compare haplotypes of the family members with and without the *BRCA2* c.8331+2C>T variant. The pedigrees of the families are illustrated in Appendix E – *BRCA2* c.8331+2C>T results (Figures 35-40). No pedigrees for Family 6 and Family 22 were included because from Family 6 exclusively family members without the variant were available for analysis, and only one family member with *BRCA2* c.8331+2C>T was analysed from Family 22.

Nine families (families 1, 4, 16, 17, 19, 20, 22, 23 and 24) were shown to share a common haplotype encompassing *BRCA2* c.8331+2C>T (Table 12). The largest common haplotype in this study included all 5 microsatellites investigated and had allele sizes for D13S290=189 bp, *BRCA2*.6m=177 bp, *BRCA2*.1m=193 bp, *BRCA2.5m*=245 bp and D13S267=157. However, in eight other families (families 2, 5, 8, 11, 14, 21, 25 and 26) crossing over events had occurred between *BRCA2*.6m and *BRCA2*.1m at the centromeric side of *BRCA2* (Table 12). Ten families (families 7, 8, 9, 11, 12, 14, 15, 16, 18 and 19) demonstrated crossing over events between *BRCA2*.5m and D13S167 on the telomeric side of *BRCA2*. Moreover, in a few

families (families 3 and 8) a crossing over event was demonstrated between *BRCA2* and *BRCA2*.5m. Accordingly, the maximum linked area on chromosome 13 with *BRCA2* c.8331+2C>T was 986.6 kb and the minimum chromosome area linked to *BRCA2* c.8331+2C>T was 188 kb.

Table 12. The alleles for the markers of the individuals carrying the BRCA2 c.8331+2C>T variant (mB). The alleles marked in orange are common among the individuals in the different families. The alleles marked in white show the crossing over events. The families which numbers are listed more than one time, have affected individuals with different alleles.

Family		1	2	3	4	5	7	8	**8**	9	10	11	12	13	14	15	16	**16**	17	**17**	18	19	**19**	20	20	21	22	23	24	25	26
	Chromosome 13 location																														
D13S290	31862621-31862968	189	189	175	189	175	189	189	189	175	189	189	189	185	175	189	189	189	189	175	189	189	189	191	189	175	189	189	189	175	175
BRCA2.6m	32338381-32338426	177	183	177	177	181	177	177	183	177	177	179	177	177	181	177	177	177	177	177	177	177	177	177	177	181	177	177	177	181	183
BRCA2.1m	32588311-32588358	193	193	193	193	193	181	193	193	193	193	193	193	193	193	193	193	193	193	193	193	193	193	193	193	193	193	193	193	193	193
BRCA2	32315474-32400266	mB	mB	mB	mB	mB	mB	mB	mB	mB	mB	mB	mB	mB	mB	mB	mB	mB	mB	mB	mB	mB	mB	mB	mB	mB	mB	mB	mB	mB	mB
BRCA2.5m	33324986-33325026	245	245	239	245	245	245	245	239	245	245	245	245	245	245	245	245	245	245	245	245	245	245	245	245	245	245	245	245	245	245
D13S267	33689982-33690250	157	?	151	157	157	159	155	149	155	?	155	155	157	143	151	157	147	157	157	159	157	147	157	157	157	157	157	157	157	157

4.2.2.1 Mutation age

The age of the *BRCA2* c.8331+2C>T was estimated using the DMLE +2.2 software. In order to calculate the growth rate the equation in "2.5.4 DMLE – Disease mapping using linkage disequilibrium" was used. The growth rate was calculated from year 1000 (P_0) to today (P_t). It is presumed that in year 1000 there were about 150 000 people living in Norway. According to Statistics Norway, by 1st of January 2017, the population of Norway was at 5 258 317 [109]. The number of generations between these two points in time (g) was calculated assuming one generation spans over 25 years.

$$g = \frac{2017 - 1000}{25} = 40.68$$

This gave a growth rate of 0.038.

$$r^{gen} = \frac{log\frac{P_t}{P_0}}{g} = \frac{log\frac{5\ 258\ 317}{150\ 000}}{40.68} = 0.038$$

The estimated prevalence of *BRCA2* carriers in the general population is 0.12 % [110]. Assuming this prevalence is correct the number of carriers in Norway was calculated to be 6309.98.

$$Number\ of\ mutated\ BRCA2\ carriers\ in\ Norway = \frac{5\ 258\ 317 \times 0.12\ \%}{100} = 6309.98$$

Subsequently, the proportion of sampled mutation-carrying chromosomes in this study was calculated to be 0.00951, based on the 60 mutation-carrying samples.

$$Proportion\ of\ sampled\ chromosomes = \frac{60\ samples}{6309.98} = 0.00951$$

The acquired information was submitted into the DMLE +2.2 software. According to this analysis the age of *BRCA2* c.8331+2C>T was estimated to be between 97 and 215 generations old (Figure 25). Because we defined that one generation was 25 years, the overall age of this *BRCA2* variant is estimated to be between 2425 and 5275 years old.

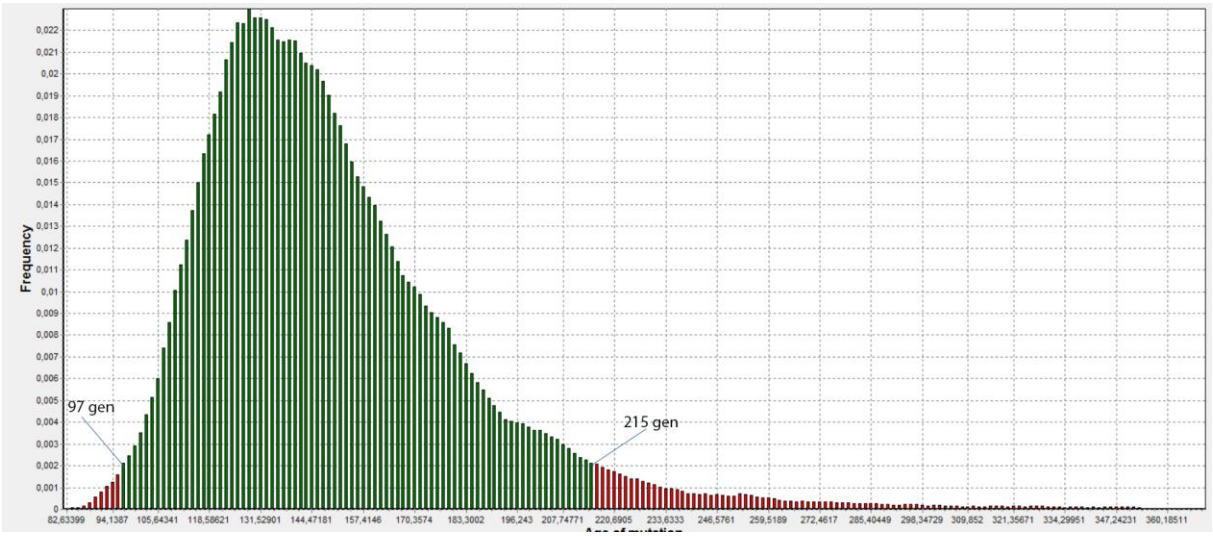


Figure 25. The age of the BCRA2 c.8331+2C>T variant estimated by the DMLE +2.2 software with a 95 % confidence interval. The age is given in generations (gen).

5 Discussion

5.1 NGS

We investigated DNA samples from 48 patients with breast and/or ovarian cancer using NGS. Eighteen genes where scrutinised (Table 4), including *BRCA1* and *BRCA2*, which were included to make sure that none of the patients had deleterious variants in these two genes. Thirty-one of 48 patients (64.6 %) had a total of 36 different variants in the remaining 16 genes. Among the 36 variants, three variants (identified in five patients) were characterized as clinically significant. Accordingly, we discovered deleterious variants in 10.4 % of the investigated patient material. However, the findings in 10.4 % of the investigated patients is an overestimation compared to the literature, where 4.9-10 % of the HBOC patients were identified with deleterious variants in gene other than *BRCA1* and *BRCA2* [16, 25, 86, 111], due to a bias of ascertainment since we selected patients from families with several affected family members. In addition 8 different variants identified in 10 different patients were classified as VUSs. Two of the deleterious variants were found in patients were also a VUS was identified (TSC2, TSC12, TSC16 and TSC31, Table 11). Over half of these variants (68.1 %) were classified as benign or likely benign, and will not be discussed further.

Currently, several reports of NGS analysis of patients with breast and/or ovarian cancer have been published. Kraus and colleagues, for example, found that 7.8 % (32/409) of their patients had deleterious variants in genes other than *BRCA1/2* [86]. Pinto and colleagues, who performed a study with 94 patients, found deleterious variants in 10 % of their families among 15 scrutinised genes [16]. Buys and colleagues, who used a 25-gene hereditary cancer panel, found that 4.9 % of their over 35 000 patients had deleterious variants in genes other than *BRCA1* and *BRCA2* [25]. Tung and colleagues, who also used a 25-gene hereditary cancer panel, identified 5.7 % of their patient cohort without *BRCA1* or *BRCA2* deleterious variants [111]. Three of these four studies have a lower percentage of deleterious variants among their patient cohorts then the 10.4 % found in our study. The higher portion of patients with deleterious variants in this study is probably due to bias of ascertainment of limited amount of patients currently investigated. This might also be the case for the study performed by Pinto and colleagues, since their patient cohort is lower than the other three studies.

5.1.1 Deleterious variants

Three class 4 and 5 variants were found among five patients. One of these variants was identified in a high penetrance cancer susceptibility gene (*TP53*) and confers a 4.3-9.3 increased relative cancer risk [108]. Two deleterious variants were identified in moderate penetrance breast cancer genes, *ATM* (c.3245_3247delinsTGAT) and *CHEK2* (c.319+2T>A), where each identified in two different patients. *ATM* and *CHEK2* are associated with a 1.5 – 3.8 and 2.7 – 4.8 fold increased cancer risk, respectively [111].

5.1.1.1 *TP53* c.818G>A

Mutation of the tumour suppressor *TP53* gene occurs in more than half of all human cancers. The variant *TP53* c.818G>A p.(Arg273His) is the second most frequent *TP53* variant reported in the catalogue of somatic mutations in cancer (COSMOS) [80], and is located at one of the mutation hotspots of *TP53* earlier described (Figure 5) [27]. The currently identified variant c.818G>A p.(Arg273His), has been shown to be gain-of function mutation [27]. In addition, transcriptional analysis of the TP53-Arg273 mutants indicated that the apoptosis pathway was less active in p.Arg273His compared to WT TP53 [27]. The same study performed molecular dynamics simulations, suggesting that WT TP53 binds more tightly to DNA than TP53-Arg273His. Accordingly, the deleterious nature of the TP53 variant currently found in one of the patients studied has been well documented [27, 79-81].

Our patient (TSC13) was only 36 years old when diagnosed with breast cancer. In addition, several cases of different cancers were identified in the family, including a 6 years old sister with Wilms tumour. Deleterious variants in *TP53* are associated with several different cancers, including Wilms tumour [112, 113]. This suggests that the *TP53* c.818G>A variant may be the cause of the cancers in this family. The results from NGS showed that patient TSC13 had 30 % reads with the *TP53* mutation, whereas, 70 % showed WT *TP53*. This, on the other hand, might indicate that the patient might be carrying a somatic mosaic mutation in *TP53*. Further investigations on a different tissue from our patient and family mutation analysis for the present *TP53* variant should be performed in order to resolve this discrepancy. No additional variants were found in the 16 genes scrutinised in this patient. It is of course possible that there is a undermined pathogenic variant in another gene in this patient, since some of the scrutinised genes were not completely covered by the kit used for the sequencing, or in a gene not included in this study.

5.1.1.2 *ATM* c.3254_3247delinsTGAT

In patients TSC2 and TSC31 the *ATM* c.3245_3247delinsTGAT p.(His1082Leu*14) variant was found. This variant is a Norwegian founder mutation from south-east Norway, and has been reported in several A-T families [14, 18]. The *ATM* c.3245_3247delinsTGAT variant theoretically results in a frame-shift and stop of translation after 14 aa. This stop occurs before the PI3 kinase domain of the protein (Figure 4), which will give a dysfunctional product. Although A-T is an autosomal recessive disease, there is evidence of moderate increased risk of cancer development for heterozygote carriers, particularly for breast cancer development [18].

ATM variants which lead to breast cancer are believed to be missense variants, due to deleterious missense variants giving a dominant negative effect on the protein, even though other variants have also been found in *ATM* among breast cancer families [15, 19, 20]. The reason for which most deleterious *ATM* variants causing breast cancer are believed to be missense variants is probably because the nonsense or frameshift transcripts are most likely to be removed by nonsense-mediated mRNA decay (NMD). Nonsense-mediated mRNA decay is a mechanism which prevents the production of truncated proteins by degrading the mRNA of a dysfunctional transcript before it is translated into a protein [114].

Patient TSC2, who had been diagnosed with breast cancer at age 57 years, had in addition to the *ATM* c.3245_3247delinsTGAT variant also a VUS in *NF1* c.5225A>G p.(Ans1742Ser). The variant changes the aa 1742 from Asn to Ser. Asn172 is located within the lipid binding domain of neurofibromin. Both aa are polar, and the change has been predicted to be tolerated (Appendix D – NGS results, Table 26). However, TSC2 had a family history of breast, ventricle and pancreas cancer. Although the *ATM* variant is most likely responsible for the increased cancer risk in this family, the variant in *NF1* c.5225A>G has not been described in the literature before and may possibly add to the already *ATM* associated cancer risk in this family. Mutation analysis of several affected family members have to be performed in order to establish which variant segregates with increased cancer risk. In addition, we cannot exclude that the increased cancer risk in this family may be associated with another cancer gene which was not investigated in the current study.

Patients TSC31 and TSC32 are sisters both diagnosed with ovarian cancer at age 45 years and 52 years, respectively. We have established that TSC31 was heterozygous for *ATM*

c.3245_3247delinsTGAT. However, this *ATM* variant was not identified in her sister (TSC32). Nevertheless, both sisters are heterozygous for a missense variation in *BRIP1*. *BRIP1* c.2087C>T p.(Pro696Leu) changes the aa from Pro to Leu, which have some structural difference from each other. The change in *BRIP1* is predicted to be deleterious (Appendix D – NGS results, Table 26), and Pro696 is located in between two of the eight motifs (III and IV) which form the ATPase helicase core domain of the BRIP1 protein (Figure 7). Accordingly, one might speculate that the variant can cause a different folding of the protein. However, the variant is also reported with a low frequency in gnomAD (0.0047 %), indicating that this can be a deleterious variant. Due to the lack of literature the variant was classified as a VUS. The absence of the *ATM* c.3245_3247delinsTGAT variant in one of the two affected sisters indicates that this *ATM* variant might not be the cause of cancer in this family. In addition, deleterious *ATM* variants may increase the risk for breast cancer, but its association with ovarian cancer is not established [9]. The family history of ovarian cancer might indicate that *BRIP1* c.2087C>T is a better candidate, since deleterious variants in *BRIP1* give an increased risk of developing ovarian cancer [5, 9, 16, 17]. On the other hand, other genetic reasons for development of ovarian cancer cannot completely be excluded. Functional studies for the mutant BRIP1 protein will be necessary to establish its importance in cancer development.

5.1.1.3 *CHEK2* c.319+2T>A

The class 4 variant in *CHEK2* c.319+2T>A was found in two patients with ovarian cancer (TSC12 and TSC16). The variant was classified as a class 4 variant because of its predicted effect on splicing (Appendix D –NGS results, Table 26). *CHEK2* c.319+2T>A has not been reported in literature. However, variants in *CHEK2* are considered to increase risk of breast cancer and sarcoma, while its role in ovarian cancer is not quite clear [5, 9, 16, 17, 24-26]. Besides that both patients were heterozygous for *CHEK2* c.319+2T>A variant, each patient was also heterozygous for a VUS. In patient sample TSC12 a *CDH1* c.136C>G p.(Leu46Val) variant was also identified. This variant has not been reported in either literature or in the population databases, but is seen as a tolerated change by the prediction programs (Appendix D –NGS results, Table 26). Patient TSC12 has been diagnosed with sarcoma (61 years) as well as ovarian cancer (27 years), and has a family history of breast cancer on her father's side. Her mother's family has a cervix, kidney, prostate, breast and colorectal cancer (CRC) history. The roles of *CDH1* and *CHEK2* in ovarian cancer are currently unknown. However,

CHEK2 variants have been associated with sarcoma which the patient developed later in life. Further family investigation are needed to establish the possible genetic cause associated with the increased cancer risk in this family.

Patient TSC16 was besides heterozygous for *CHEK2* c.319+2T>A also heterozygous for a VUS in *MLH1* c.-7C>T. *MLH1* is one of the MMR genes which, when mutated, can cause LS. Deleterious variants in *MLH1*have also been associated with an increased risk for ovarian cancer [9, 36, 37]. The *MLH1* c.-7C>T variant has been reported by Hesson and colleagues to possibly be a deleterious variant due to partial effect on the expression of MLH1 [85]. It might be that *MLH1* c.-7C>T is the cause of ovarian cancer in this patient, and not *CHEK2*, since the role of deleterious variants in *CHEK2* in ovarian cancer is not quite known [9]. Further family investigations are necessary to establish the possible role of both variants in this family.

5.1.2 Variants of unknown clinical significance

Several other VUSs were identified among the patient samples. However, the family history of these patients was not available. Accordingly, exclusively the possible effect of the sequence variants will be discussed in this section.

Patient TSC14 was heterozygous for *ATM* c.8734A>G p.(Arg2912Gly). This variant changes aa 2912 from Arg to Gly, which encompasses a change from a charged aa to a hydrophobic aa. Arg2912 is located in the PI3-kinase catalytic domain of the protein (Figure 4) and might accordingly have an effect on the kinase function of *ATM* [15, 20]. The *ATM* c.8734A>G variant has been reported in the literature among families with breast cancer, but not all the affected members in one of the families had the variant [15, 20, 83]. The variant has also been found in an unaffected individual, whose mother had bilateral breast cancer, together with a *BRCA1* frameshift mutation (c.185delAG*39) [15, 20]. This can indicate that *ATM* c.8734A>G may not be associated with an increased breast cancer risk, or that it can give a moderate increased risk. The studies had no other affected family members in the family available for sequencing for these two variants, so the occurrence of the variants in the affected individuals are not known [15, 20]. In a study performed by Hollestelle and colleagues, they estimated that the variant may give a 7-fold increased risk of developing breast cancer, however they also state that this is a rough estimate due to limited data [84].

Patient TSC14, who was heterozygous for *ATM* c.8734A>G, did not have other candidate variants present in the 16 scrutinised genes.

Patient TSC15 was heterozygous for VUS in *NF1* (c.587-10_587-9delTT). The variant is located in intron 5, and it has not been reported in literature. This intronic variant might have an effect on splicing since four splicing prediction programs: SSF-like, GeneSplicer and HSF, as well as NNSPLICE, predict a reduction in the strength of the splice sites (Appendix D – NGS results, Table 26). Without cDNA studies, it is difficult to determine whether this is a splice variant or not.

The *NF1* c.7354C>T p.(Arg2452Cys) variant was found in patient TSC1 and has not been reported in literature before. This variant changes aa 2452 from Arg, a charged aa, to Cys, a polar aa. There is a big physiochemical difference between these two aa. However, the variant is not located in an identified functional domain of the protein. Further functional investigations are necessary to establish the effect of this missense mutation and its role in cancer development.

We identified *PMS2* c.1437C>G p.(His479Gln) in two patients, TSC4 and TSC28. This missense variant results in change of aa 479 from His to Gln in mismatch repair endonuclease PMS2. The two aa are both polar and there is only a small difference in structure between them. The *PMS2*c.1437C>G has been reported in literature both as a class 1 variant and a class 3 variant [37, 86]. Further examinations are necessary to establish its association with cancer development.

Possibly, there are more sequence variants present in the genes scrutinised in this study, but these variants may not have been detected due to the limitations of the NGS kit used. For example *PMS2* has a strong homology with several pseudogenes [37, 44]. Exon 9 and 11-15 have an especially high homology with one of the pseudogenes, *PMS2CL*. Accordingly, variants within the *PMS2* gene may be miss-aligned and subsequently been filtrated out by the MiSeq software. A possible way to investigate this is to sequence the region by using long-range PCR to amplify the real gene [44]. Clendenning and colleagues performed long-range PCR by using four overlapping sets of primers which were designed to amplify the complete coding region of *PMS2* while avoiding amplification of pseudogenes [44]. The pseudogenes were avoided by designing primers which would anneal to the regions without

homology to the pseudogenes. However, in our study the *PMS2* gene in the patient samples was not investigated further after the filtration of the variants.

During each NGS run there were some regions which were not covered well enough (3.4.4 Quality of a run). It is possible that there were variants in these regions which were not recognised. One gap, which was part of *MSH6* exon 1, was present in every sample analysed. Conversely, variants localized in this region could not be detected with the currently applied TruSight cancer technology used.

5.2 BRCA2 c.8331+2C>T

5.2.1 cDNA analysis

The *BRCA2* c.8331+2C>T variant has been found in several families in Norway and was expected to have an effect on the splice site. Accordingly, cDNA analysis was performed. The cDNA analysis established that the *BRCA2* c.8331+2C>T variant leads to skipping of exon 18. This is consistent with the study performed by Fraile-Bethencourt and colleagues [115], where they found 30 spliceogenic variants in *BRCA2* using a minigene assay. In their study they also found that the *BRCA2* c.8331+2C>T variant results in skipping of exon 18, as well as skipping of the 3' part of exon 17 together with exon 18 [115]. In our study the partial skipping of exon 17 was not seen due to the positions of the forward primer. Optimally, the primers should be placed with one exon between primer location and variant location, which in this case would have been exon 16 and 20. Since we performed fragment analysis, this was not possible due to the correspondingly large PCR-product.

The loss of exon 18 will give a shift in the reading frame and an early truncation after 43 aa. This premature stop happens within the DNA-binding domain, which probably affects the BRCA2 proteins function during HRR.

5.2.2 Microsatellite analysis

In Norway, 29 families including 134 individuals have been found to be heterozygous for *BRCA2* c.8331+2C>T. This can be an indication that the *BRCA2* c.8331+2C>T variant may be a founder mutation. Microsatellite markers were used to determine if the mutation had a common ancestor.

We initiated a haplotype study with 13 different microsatellite markers in the proximity of the *BRCA2* gene, however due to diverse causes, including the lack of informatively of some of these markers, we determined the haplotypes according to the results of 5 markers. Although a common haplotype was found in eight families, several crossing over events could be identified. Only a single marker together with the *BRCA2* c.8331+2C>T defined the area on chromosome 13 which was linked to this mutation event. Accordingly, we could establish that the maximum linked area on chromosome 13 with *BRCA2* c.8331+2C>T is 986.6 kb and the minimum chromosome area linked to *BRCA2* c.8331+2C>T is 188 kb. The common haplotype and the small associated chromosome area indicated a common ancestral origin for *BRCA2* c.8331+2C>T. However, additional investigations establishing the allele frequency for the different markers investigated should be performed in order to make sure that not several mutational events have occurred among common haplotypes at the associated chromosomes 13.

5.2.2.1 Mutation age

Currently, several studies have applied DMLE +2.2 software in order to estimate the age of certain founder mutations [30, 78, 116]. Due to the high prevalence of the *BRCA2* c.8331+2C>T variant occurring in the Norwegian breast cancer population we assumed we had a founder mutation. The result form the analysis of the five microsatellite markers shows that the chromosome 13 associated area linked with the *BRCA2* c.8331+2C>T variant is less than 1Mb, indicating that the origin of variant is quite old. To determine the age of the variant, the alleles of the different microsatellite markers for the affected individuals were put into the DMLE +2.2 software, along with the unaffected family members as frequency controls. The age of the variant was calculated to range from 97 to 215 generations.

Due to the limitations of the statistical analysis, it is possible that the age of the *BRCA2* c.8331+2C>T variant is over or underestimated. Especially, because the control samples used for the frequency of the different alleles of the five markers were relatives of the affected patients. Accordingly, the frequency of the investigated alleles might therefore not represent the frequency of the Norwegian population, and could give a higher age of the variant than the actual age. The number of markers can also have an effect as one of the families (Family 3) did not have a common allele with either of the other families at the telomeric side of the *BRCA2* gene. This could possibly make the variant older.

6 Concluding remarks

In the present study, three different deleterious variants were identified in five of 48 patients. The variants were *ATM* c.3245_3247delinsTGAT, *TP53* c.818G>A and *CHEK2* c.319+2T>A. In addition, we identified eight different VUSs in *ATM*, *BRIP1*, *MLH1*, *NF1* and *PMS2*. Some of these VUSs, especially *ATM* c.8734A>G and *MLH1* c.-7C>T, are strong candidates to be classified as deleterious. In order to determine if the deleterious variants or the VUSs increase the risk of HBOC, further investigation of affected family members are needed. In addition, functional analyses are necessary to determine the consequences of the identified variants on the function of the protein. Whether, besides mutations in to *BRCA1* and *BRCA2*, sequencing variants in *ATM*, *TP53* and *CHEK2* are the most additional representative causes of HBOC needs to be substantiated by investigating a larger HBOC patient cohort.

The cDNA analysis of the *BRCA2* c.8331+2C>T variant showed that the variant causes skipping of exon 18. This variant has been identified in several families in Norway. Microsatellite analysis showed that the families had a common associated chromosome area smaller than 1 Mb linked to the *BRCA2* c.8331+2C>T variant. This indicates that the variant is most likely a common Norwegian founder. With the current generated data we were able to estimate that the *BRCA2* c.8331+2C>T variant was 97-215 generations old. Extended haplotype investigations are needed both of heterozygous carriers and an independent control population to be able to give a more exact estimation of age of this mutational event.

References

1. Weinberg, R., *The biology of cancer*, ed. 2nd. 2014: Garland science.
2. Lodish, H., A. Berk, and S.L. Zipursky, *Section 24.2, Proto-Oncogenes and Tumor-Suppressor Genes*, in *Molecular Cell Biology*, 4th, Editor. 2000, W. H. Freeman: New York.
3. Knudson, A.G., *Mutation and cancer: statistical study of retinoblastoma.* Proceedings of the National Academy of Sciences, 1971. **68**(4): p. 820-823.
4. Larsen, I.K., et al., *Cancer Registry of Norway. Cancer in Norway 2015 - Cancer incidence, mortality, survival and prevalence in Norway*, I.K. Larsen, Editor. 2016.
5. Caminsky, N.G., et al., *Prioritizing variants in complete hereditary breast and ovarian cancer genes in patients lacking known BRCA mutations.* Human mutation, 2016. **37**(7): p. 640-652.
6. Gene, S., *A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1.* Science, 1994. **266**: p. 7.
7. Wooster, R., et al., *Identification of the breast cancer susceptibility gene BRCA2.* Nature, 1995. **378**(6559): p. 789.
8. Kast, K., et al., *Prevalence of BRCA1/2 germline mutations in 21 401 families with breast and ovarian cancer.* Journal of medical genetics, 2016. **53**(7): p. 465-471.
9. Nielsen, F.C., T. van Overeem Hansen, and C.S. Sørensen, *Hereditary breast and ovarian cancer: new genes in confined pathways.* Nature Reviews Cancer, 2016.
10. Alberts, B., et al., *Essential cell biology*. 3rd ed. 2010: Garland Science.
11. Ensembl. *Ensemble, release 88. Ref: https://doi.org/10.1093/nar/gkv1157*. [cited 2017; Available from: www.ensemble.org.
12. UniProt. *Ref: https://doi.org/10.1093/nar/gkw1099*. [cited 2017; Available from: www.uniprot.org.
13. Prodosmo, A., et al., *Detection of ATM germline variants by the p53 mitotic centrosomal localization test in BRCA1/2-negative patients with early-onset breast cancer.* Journal of Experimental & Clinical Cancer Research, 2016. **35**(1): p. 135.
14. Laake, K., et al., *Characterization of ATM mutations in 41 Nordic families with ataxia telangiectasia.* Human mutation, 2000. **16**(3): p. 232.
15. Pylkäs, K., et al., *Evaluation of the role of Finnish ataxia-telangiectasia mutations in hereditary predisposition to breast cancer.* Carcinogenesis, 2007. **28**(5): p. 1040-1045.
16. Pinto, P., et al., *Implementation of next-generation sequencing for molecular diagnosis of hereditary breast and ovarian cancer highlights its genetic heterogeneity.* Breast cancer research and treatment, 2016. **159**(2): p. 245-256.
17. Southey, M.C., et al., *PALB2, CHEK2 and ATM rare variants and cancer risk: data from COGS.* Journal of medical genetics, 2016. **53**(12): p. 800-811.
18. Laake, K., et al., *Identical mutation in 55% of the ATM alleles in 11 Norwegian AT families: evidence for a founder effect.* European Journal of Human Genetics, 1998. **6**(3): p. 235-244.
19. Scott, S.P., et al., *Missense mutations but not allelic variants alter the function of ATM by dominant interference in patients with breast cancer.* Proceedings of the National Academy of Sciences, 2002. **99**(2): p. 925-930.
20. Thorstenson, Y.R., et al., *Contributions of ATM mutations to familial breast and ovarian cancer.* Cancer Research, 2003. **63**(12): p. 3325-3333.
21. Barone, G., et al., *Modeling ATM mutant proteins from missense changes confirms retained kinase activity.* Human mutation, 2009. **30**(8): p. 1222-1230.

22. Lavin, M.F., *Ataxia-telangiectasia: from a rare disorder to a paradigm for cell signalling and cancer.* Nature reviews Molecular cell biology, 2008. **9**(10): p. 759-769.
23. Roeb, W., J. Higgins, and M.-C. King, *Response to DNA damage of CHEK2 missense mutations in familial breast cancer.* Human molecular genetics, 2012. **21**(12): p. 2738-2744.
24. Al-Rakan, M.A., S.-F. Hendrayani, and A. Aboussekhra, *CHEK2 represses breast stromal fibroblasts and their paracrine tumor-promoting effects through suppressing SDF-1 and IL-6.* BMC cancer, 2016. **16**(1): p. 575.
25. Buys, S.S., et al., *A study of over 35,000 women with breast cancer tested with a 25-gene panel of hereditary cancer genes.* Cancer, 2017.
26. Vahteristo, P., et al., *A CHEK2 genetic variant contributing to a substantial fraction of familial breast cancer.* The American Journal of Human Genetics, 2002. **71**(2): p. 432-438.
27. Li, J., et al., *Mutants TP53 p. R273H and p. R273C but not p. R273G enhance cancer cell malignancy.* Human mutation, 2014. **35**(5): p. 575-584.
28. Silva, J.L., et al., *Prion-like aggregation of mutant p53 in cancer.* Trends in biochemical sciences, 2014. **39**(6): p. 260-267.
29. Rebbeck, T.R., et al., *Modification of ovarian cancer risk by BRCA1/2-interacting genes in a multicenter cohort of BRCA1/2 mutation carriers.* Cancer research, 2009. **69**(14): p. 5801-5810.
30. Cini, G., et al., *Tracking of the origin of recurrent mutations of the BRCA1 and BRCA2 genes in the North-East of Italy and improved mutation analysis strategy.* BMC medical genetics, 2016. **17**(1): p. 11.
31. Wong, M.W., et al., *BRIP1, PALB2, and RAD51C mutation analysis reveals their relative importance as genetic susceptibility factors for breast cancer.* Breast cancer research and treatment, 2011. **127**(3): p. 853-859.
32. Thompson, E.R., et al., *Analysis of RAD51C germline mutations in high-risk breast and ovarian cancer families and ovarian cancer patients.* Human mutation, 2012. **33**(1): p. 95-99.
33. Brosh Jr, R.M. and S.B. Cantor, *Molecular and cellular functions of the FANCJ DNA helicase defective in cancer and in Fanconi anemia.* 2014.
34. Ramus, S.J., et al., *Germline mutations in the BRIP1, BARD1, PALB2, and NBN genes in women with ovarian cancer.* Journal of the National Cancer Institute, 2015. **107**(11): p. djv214.
35. Damiola, F., et al., *Rare key functional domain missense substitutions in MRE11A, RAD50, and NBN contribute to breast cancer susceptibility: results from a Breast Cancer Family Registry case-control mutation-screening study.* Breast Cancer Research, 2014. **16**(3): p. R58.
36. Valentin, M.D., et al., *Characterization of germline mutations of MLH1 and MSH2 in unrelated south American suspected Lynch syndrome individuals.* Familial cancer, 2011. **10**(4): p. 641-647.
37. Hansen, M.F., et al., *A massive parallel sequencing workflow for diagnostic genetic testing of mismatch repair genes.* Molecular genetics & genomic medicine, 2014. **2**(2): p. 186-200.
38. Engel, C., et al., *Risks of less common cancers in proven mutation carriers with lynch syndrome.* Journal of Clinical Oncology, 2012. **30**(35): p. 4409-4415.
39. Cohen, S.A. and A. Leininger, *The genetic basis of Lynch syndrome and its implications for clinical practice and risk management.* The application of clinical genetics, 2014. **7**: p. 147.

40. Ten Broeke, S.W., et al., *Lynch syndrome caused by germline PMS2 mutations: delineating the cancer risk.* Journal of Clinical Oncology, 2014. **33**(4): p. 319-325.
41. Merg, A., et al., *Hereditary colon cancer—part II.* Current problems in surgery, 2005. **42**(5): p. 267-334.
42. Lynch, H.T., T. Smyrk, and J. Lynch, *An update of HNPCC (Lynch syndrome).* Cancer genetics and cytogenetics, 1997. **93**(1): p. 84-99.
43. Bonadona, V., et al., *Cancer risks associated with germline mutations in MLH1, MSH2, and MSH6 genes in Lynch syndrome.* Jama, 2011. **305**(22): p. 2304-2310.
44. Clendenning, M., et al., *Long-range PCR facilitates the identification of PMS2-specific mutations.* Human mutation, 2006. **27**(5): p. 490-495.
45. Bianchessi, D., et al., *126 novel mutations in Italian patients with neurofibromatosis type 1.* Molecular genetics & genomic medicine, 2015. **3**(6): p. 513-525.
46. Valente, A.L., et al., *Sequence-based detection of mutations in cadherin 1 to determine the prevalence of germline mutations in patients with invasive lobular carcinoma of the breast.* Hereditary cancer in clinical practice, 2014. **12**(1): p. 17.
47. Figueiredo, J., et al., *The importance of E-cadherin binding partners to evaluate the pathogenicity of E-cadherin missense mutations associated to HDGC.* European Journal of Human Genetics, 2013. **21**(3): p. 301-309.
48. Phin, S., M. Moore, and P.D. Cotter, *Genomic rearrangements of PTEN in prostate cancer.* Frontiers in oncology, 2013. **3**: p. 240.
49. Momcilovic, M. and D. Shackelford, *Targeting LKB1 in cancer–exposing and exploiting vulnerabilities.* British journal of cancer, 2015. **113**(4): p. 574-584.
50. Beggs, A., et al., *Peutz–Jeghers syndrome: a systematic review and recommendations for management.* Gut, 2010. **59**(7): p. 975-986.
51. Janavičius, R., *Founder BRCA1/2 mutations in the Europe: implications for hereditary breast-ovarian cancer prevention and control.* EPMA journal, 2010. **1**(3): p. 397.
52. Nussbaum, R.L., R.R. McInnes, and H.F. Willard, *Thompson & Thompson genetics in medicine*, ed. 7th. 2007: Elsevier Health Sciences.
53. A.J.F., G., et al., *A synthesis of forces: variation and divergence of populations*, in *An introduction to genetic analysis*. 2000, W. H. Freeman: New York.
54. Fackenthal, J.D. and O.I. Olopade, *Breast cancer risk associated with BRCA1 and BRCA2 in diverse populations.* Nature Reviews Cancer, 2007. **7**(12): p. 937-948.
55. Butler, J.M., *Fundamentals of forensic DNA typing*. 2009: Academic Press.
56. Strachan, T. and A. Read, *Human molecular genetics*. 4th ed. 2011, New York: Garland Science.
57. PreAnalytiX. *QIAsymphony PAXgene Blood ccfDNA kit*. 2016 11.03 [cited 2017 02.09.]; Available from: http://www.preanalytix.com/products/blood/ccfDNA/qiasymphony-paxgene-blood-ccfdna-kit.
58. AgilentTechnologies, *2100 Expert User's Guide*. Agilent 2100 Bioanalyzer. 2005.
59. AgilentTechnologies, *Agilent High Sensitivity DNA Kit Guide*. 2013, Germany: Agilent Technologies inc.
60. Scientific, T.F., *NanoDrop 2000/2000c Spectrophotometer V1. 0 User Manual.* Wilmington, Deleware, 2009.
61. Illumina. *TrueSight Cancer*. 2017 [cited 2017 03. 27.]; Available from: https://www.illumina.com/products/by-type/clinical-research-products/trusight-cancer.html.
62. Syed, F., H. Grunenwald, and N. Caruccio, *Optimized library preparation method for next-generation sequencing.* Nature Methods, 2009. **6**(10).

63. Mukherjee, S., et al., *Large-scale contamination of microbial isolate genomes by Illumina PhiX control.* Standards in genomic sciences, 2015. **10**(1): p. 18.
64. Ansorge, W.J., *Next-generation DNA sequencing techniques.* New biotechnology, 2009. **25**(4): p. 195-203.
65. Mardis, E.R., *The impact of next-generation sequencing technology on genetics.* Trends in genetics, 2008. **24**(3): p. 133-141.
66. Ku, C.-S., et al., *The Evolution of High-Throughput Sequencing Technologies: From Sanger to Single-Molecule Sequencing*, in *Next Generation Sequencing in Cancer Research*. 2013, Springer. p. 1-30.
67. Lu, Y., et al., *Next Generation Sequencing in Aquatic Models.* 2015.
68. Richards, S., et al., *Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology.* Genetics in Medicine, 2015. **17**(5): p. 405-423.
69. Interactive-Biosoftware. *Alamut Visual*. [cited 2017 04.26.]; Available from: http://www.interactive-biosoftware.com/alamut-visual/.
70. Schwarz, J.M., et al., *MutationTaster evaluates disease-causing potential of sequence alterations.* Nature methods, 2010. **7**(8): p. 575-576.
71. Houdayer, C., *In silico prediction of splice-affecting nucleotide variants.* In Silico Tools for Gene Discovery, 2011: p. 269-281.
72. Desmet, F.-O., et al., *Human Splicing Finder: an online bioinformatics tool to predict splicing signals.* Nucleic acids research, 2009. **37**(9): p. e67-e67.
73. Pertea, M., X. Lin, and S.L. Salzberg, *GeneSplicer: a new computational method for splice site prediction.* Nucleic acids research, 2001. **29**(5): p. 1185-1190.
74. Pratt, C.W. and K. Cornely, *Essential biochemistry*. 2nd ed. 2011, Hoboken, N.J: Wiley.
75. Harris, D.C., *Quantitative chemical analysis*. 8th ed. ed. 2010, New York: Freeman.
76. Reeve, J. and B. Rannala. *DMLE + Linkage disequilibrium mapping software*. 2005 10.18. [cited 2017 04.01.]; Available from: http://www.dmle.org/.
77. Reeve, J.P. and B. Rannala, *DMLE+: Bayesian linkage disequilibrium gene mapping.* Bioinformatics, 2002. **18**(6): p. 894-895.
78. Pin, E., et al., *MUTYH c. 933+ 3A> C, associated with a severely impaired gene expression, is the first Italian founder mutation in MUTYH-Associated Polyposis.* International journal of cancer, 2013. **132**(5): p. 1060-1069.
79. Melhem-Bertrandt, A., et al., *Early onset HER2-positive breast cancer is associated with germline TP53 mutations.* Cancer, 2012. **118**(4): p. 908-913.
80. Olfson, E., et al., *Identification of medically actionable secondary findings in the 1000 genomes.* PloS one, 2015. **10**(9): p. e0135193.
81. Park, K.-J., et al., *Germline TP53 mutation and clinical characteristics of Korean patients with Li-Fraumeni syndrome.* Annals of Laboratory Medicine, 2016. **36**(5): p. 463-468.
82. Chen, J., et al., *The role of ataxia-telangiectasia heterozygotes in familial breast cancer.* Cancer research, 1998. **58**(7): p. 1376-1379.
83. Tavtigian, S.V., et al., *Rare, evolutionarily unlikely missense substitutions in ATM confer increased risk of breast cancer.* The American Journal of Human Genetics, 2009. **85**(4): p. 427-446.
84. Hollestelle, A., et al., *Discovering moderate-risk breast cancer susceptibility genes.* Current opinion in genetics & development, 2010. **20**(3): p. 268-276.
85. Hesson, L.B., et al., *Lynch syndrome associated with two MLH1 promoter variants and allelic imbalance of MLH1 expression.* Human mutation, 2015. **36**(6): p. 622-630.

86. Kraus, C., et al., *Gene panel sequencing in familial breast/ovarian cancer patients identifies multiple novel mutations also in genes others than BRCA1/2.* International Journal of Cancer, 2017. **140**(1): p. 95-102.
87. Tommiska, J., et al., *ATM variants and cancer risk in breast cancer patients from Southern Finland.* BMC cancer, 2006. **6**(1): p. 209.
88. Liberzon, E., et al., *Molecular variants of the ATM gene in Hodgkin's disease in children.* British journal of cancer, 2004. **90**(2): p. 522-525.
89. Bodian, D.L., et al., *Germline variation in cancer-susceptibility genes in a healthy, ancestrally diverse cohort: implications for individual genome sequencing.* PloS one, 2014. **9**(4): p. e94554.
90. Maxwell, K.N., et al., *Evaluation of ACMG-guideline-based variant classification of cancer susceptibility and non-cancer-associated genes in families affected by breast cancer.* The American Journal of Human Genetics, 2016. **98**(5): p. 801-817.
91. Navrkalova, V., et al., *ATM mutations uniformly lead to ATM dysfunction in chronic lymphocytic leukemia: application of functional test using doxorubicin.* Haematologica, 2013: p. haematol. 2012.081620.
92. Mitui, M., et al., *Independent mutational events are rare in the ATM gene: haplotype prescreening enhances mutation detection rate.* Human mutation, 2003. **22**(1): p. 43-50.
93. Thorstenson, Y.R., et al., *Global analysis of ATM polymorphism reveals significant functional constraint.* The American Journal of Human Genetics, 2001. **69**(2): p. 396-412.
94. Lei, H. and I. Vorechovsky, *BACH1 517C→ T transition impairs protein translocation to nucleus: A role in breast cancer susceptibility?* International journal of cancer, 2003. **104**(3): p. 389-391.
95. Amendola, L.M., et al., *Actionable exomic incidental findings in 6503 participants: challenges of variant classification.* Genome research, 2015. **25**(3): p. 305-315.
96. Takahashi, M., et al., *Functional analysis of human MLH1 variants using yeast and in vitro mismatch repair assays.* Cancer Research, 2007. **67**(10): p. 4595-4604.
97. Kansikas, M., R. Kariola, and M. Nyström, *Verification of the three-step model in assessing the pathogenicity of mismatch repair gene variants.* Human mutation, 2011. **32**(1): p. 107-115.
98. Christensen, L.L., et al., *Functional characterization of rare missense mutations in MLH1 and MSH2 identified in Danish colorectal cancer patients.* Familial cancer, 2009. **8**(4): p. 489.
99. Pastrello, C., et al., *Integrated analysis of unclassified variants in mismatch repair genes.* Genetics in Medicine, 2011. **13**(2): p. 115-124.
100. Mattocks, C., et al., *Automated comparative sequence analysis identifies mutations in 89% of NF1 patients and confirms a mutation cluster in exons 11–17 distinct from the GAP related domain.* Journal of medical genetics, 2004. **41**(4): p. e48-e48.
101. Jelsig, A.M., et al., *Germline variants in Hamartomatous Polyposis Syndrome-associated genes from patients with one or few hamartomatous polyps.* Scandinavian journal of gastroenterology, 2016. **51**(9): p. 1118-1125.
102. Bogdanova, N., et al., *PALB2 mutations in German and Russian patients with bilateral breast cancer.* Breast cancer research and treatment, 2011. **126**(2): p. 545-550.
103. Niessen, R.C., et al., *PMS2 involvement in patients suspected of Lynch syndrome.* Genes, Chromosomes and Cancer, 2009. **48**(4): p. 322-329.

104. Jønson, L., et al., *Identification of six pathogenic RAD51C mutations via mutational screening of 1228 Danish individuals with increased risk of hereditary breast and/or ovarian cancer.* Breast cancer research and treatment, 2016. **155**(2): p. 215-222.
105. Cunningham, J.M., et al., *Clinical characteristics of ovarian cancer classified by BRCA1, BRCA2, and RAD51C status.* Scientific reports, 2014. **4**: p. 4026.
106. Loveday, C., et al., *Germline RAD51C mutations confer susceptibility to ovarian cancer/Meindl et al. reply.* Nature genetics, 2012. **44**(5): p. 475.
107. Meindl, A., et al., *Germline RAD51C mutations confer susceptibility to ovarian cancer.* Nature Genetics, 2012. **44**(5): p. 476-476.
108. Gonzalez, K.D., et al., *Beyond Li Fraumeni Syndrome: clinical characteristics of families with p53 germline mutations.* Journal of Clinical Oncology, 2009. **27**(8): p. 1250-1256.
109. Dybendal, K. *Key figures for the population*. 2017 03.13. [cited 2017 04.01.]; Available from: https://www.ssb.no/en/befolkning/nokkeltall/population.
110. Peto, J., et al., *Prevalence of BRCA1 and BRCA2 gene mutations in patients with early-onset breast cancer.* Journal of the National Cancer Institute, 1999. **91**(11): p. 943-949.
111. Tung, N., et al., *Frequency of mutations in individuals with breast cancer referred for BRCA1 and BRCA2 testing using next-generation sequencing with a 25-gene panel.* Cancer, 2015. **121**(1): p. 25-33.
112. Ooms, A.H., et al., *Significance of TP53 Mutation in Wilms Tumors with Diffuse Anaplasia: A Report from the Children's Oncology Group.* Clinical Cancer Research, 2016. **22**(22): p. 5582-5591.
113. Birch, J.M., et al., *Relative frequency and morphology of cancers in carriers of germline TP53 mutations.* Oncogene, 2001. **20**(34): p. 4621.
114. Miller, J.N. and D.A. Pearce, *Nonsense-mediated decay in genetic disease: friend or foe?* Mutation Research/Reviews in Mutation Research, 2014. **762**: p. 52-64.
115. Fraile-Bethencourt, E., et al., *Functional classification of DNA variants by hybrid minigenes: Identification of 30 spliceogenic variants of BRCA2 exons 17 and 18.* PLoS Genetics, 2017. **13**(3): p. e1006691.
116. Aretz, S., et al., *MUTYH-associated polyposis (MAP): evidence for the origin of the common European mutations p. Tyr179Cys and p. Gly396Asp by founder events.* European Journal of Human Genetics, 2014. **22**(7): p. 923-929.

Appendix A - Primers

Table 13. Primers used for PCR amplification for Sanger sequencing verification of variants identified by NGS and classified as class 4 and 5. The primers are listed in 5' to 3' direction.

Primers	*Sequence*	*Product size (bp)*
M13-ATM in20ex21.F	TGTAAAACGACGGCCAGTCAGGCATCTAACAAAGGAGAGG	450
M13-ATM in22.R	CAGGAAACAGCTATGACCGGCACACCGTATATACTCAACA	
M13-TP53 in7.F	TGTAAAACGACGGCCAGTTTGGGAGTAGATGGAGCCTG	300
M13-TP53 in8.R	CAGGAAACAGCTATGACCGGTGATAAAAGTGAATCTGAGGC	
M13-CHEK2 ex2.F	TGTAAAACGACGGCCAGTAAACTCCAGCCAGTCCTCTC	389
M13-CHEK2 in2.R	CAGGAAACAGCTATGACCGCTCCCATCCTGTGACATGT	

Table 14. Primers used for the APRT amplification used for control of the cDNA analysis. The primers are listed in 5' to 3' direction. hmAPRT1 is the forward primer, and hmAPRT2 is the reverse primer.

Primer	*Sequence*	*Product size cDNA*	*Product size gDNA*
hmAPRT1	GGGGAAGCTGCCAGGCCCCACT	218 bp	721 bp
hmAPRT2	GCGAGGTCAGCTCCACCAGGCT		

Table 15. Primers used for the BRCA2 c.8331+2C>T cDNA analysis. The primers are listed in 5' to 3' direction.

Primer	*Sequence*	*Product size (bp)*
BRCA2 ex17.F	ATGGAAACTGGCAGCTATGG	540
FAM-BRCA2 ex19.R	AGCGATGATAAGGGCAGAGG	

Appendix B - PCR-programs

Table 16. PCR-program for library preparation (First amplification).

	Temperature	*Time*
Denature	72˚C	3 min
	98˚C	30sec
Cycle x10	98˚C	10 sec
	60˚C	30 sec
	72˚C	30 sec
Elongation	72˚C	5min
Storage	10˚C	∞

Table 17. PCR-program or library preparation (Hybridization and first capture).

	Temperature	*Time*
Denature	95˚C	10 min
Cycle x18	94˚C	1 min
Storage/Pause	12˚C	∞

Table 18. PCR-program for library preparation (Second PCR).

	Temperature	*Time*
Denature	98˚C	30 sec
Cycle x12	98˚C	10 sec
	60˚C	30 sec
	72˚C	30 sec
Elongation	72˚C	5 min
Storage/Pause	10˚C	∞

Table 19. PCR-program used for amplifying specific gene sequences of verification of variants by Sanger sequencing.

	Temperature	*Time*
Hot Start	96˚C	8 min
Cycle x2	95˚C	20 sec
	63˚C	20 sec
	72˚C	30 sec
Cycle x2	95°C	20 sec
	61°C	20 sec
	72°C	30 sec
Cycle x2	95°C	20 sec
	59°C	20 sec
	72°C	30 sec
Cycle x28	95°C	20 sec
	58°C	20 sec
	72°C	30 sec
Storage/Pause	12˚C	∞

Table 20. The A'SAP PCR-program

	Temperature	*Time*
	37°C	15 min
	80°C	15 min
Storage/pause	12°C	∞

Table 21. Sanger sequencing PCR-program with M13-primers

	Temperature	*Time*
Hot Start	96˚C	3 min
Cycle x30	96˚C	10 sec
	50˚C	5 sec
	60˚C	4 min
Storage/Pause	12˚C	∞

Table 22. PCR-program used for cDNA synthesis

	Temperature	*Time*
Annealing	25°C	10 min
enzyme activity	42°C	1 h
Denaturation	85°C	5 min

Table 23. PCR-program for APRT cDNA amplification

	Temperature	*Time*
Cycle ×30	94°C	30 sec
	63°C	30 sec
	72°C	1 min
End cycle	72°C	7 min
Storage/Pause	12°C	∞

Table 24. PCR-program used for splicing analysis of BRCA2 c.8331+2C>T

	Temperature	*Time*
Hot Start	95˚C	? min
Cycle x36	95˚C	20 sec
	58˚C	20 sec
	72˚C	30 sec
End cycle	72˚C	30 min
Storage/Pause	12˚C	∞

Table 25. PCR-program used for the microsatellite analysis

	Temperature	*Time*
Hot Start	94˚C	3 min
Cycle x30	94˚C	30 sec
	55˚C	30 sec
	72˚C	30 sec
End cycle	72˚C	30 min
Storage/Pause	4˚C	∞

Appendix C – Filtration tree

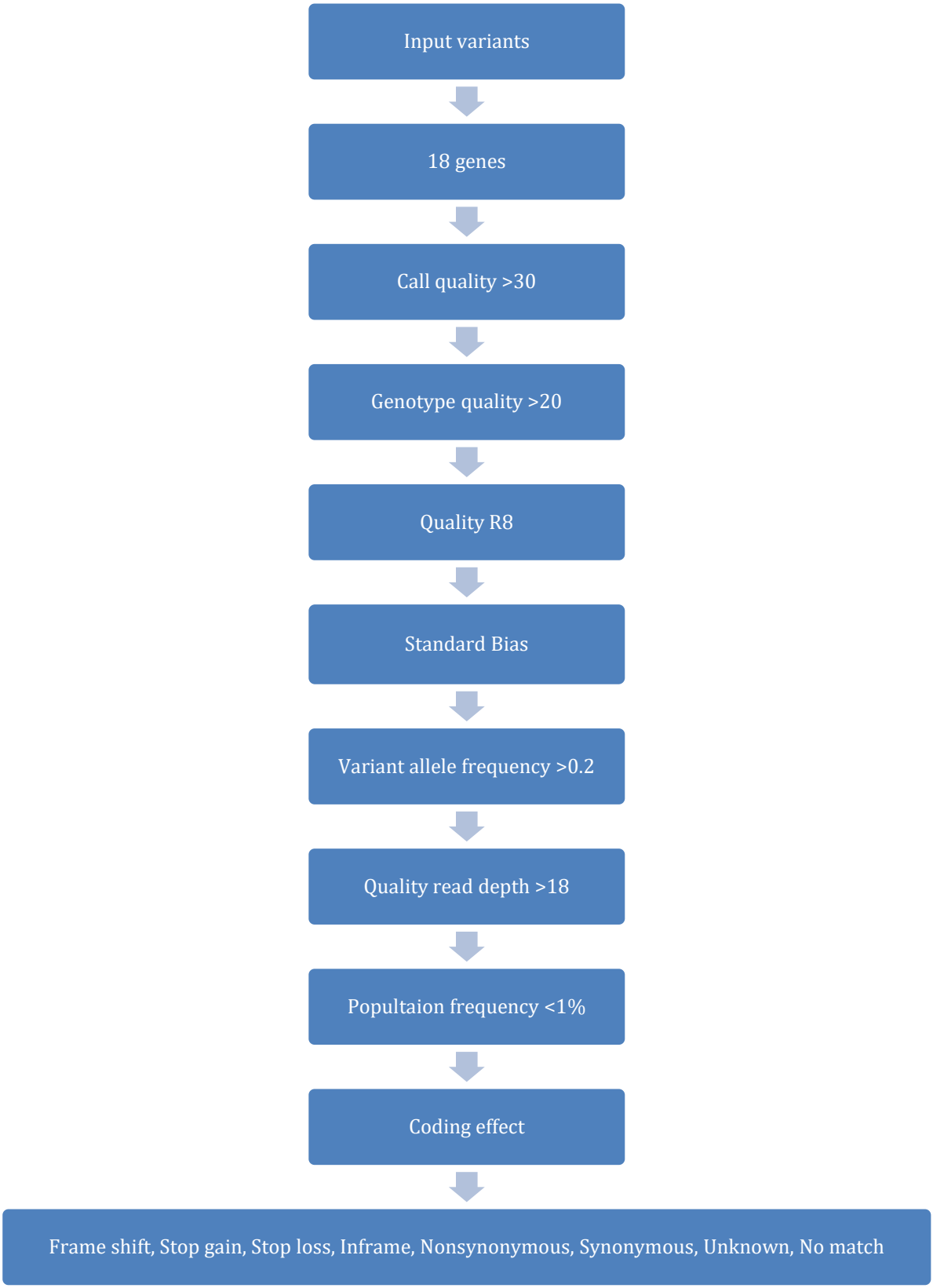


Figure 26. Filtration tree used in Cartagenia. The variants which did not meet the filters criteria were excluded. Call quality >30 (variants that are identified for less than 30 % of the reads). Genotype quality filter (filters out false genotypes). R8 (variants found after eight or more repeated mononucleotides). Standard bias and variant allele frequency filters (removes artefacts). Quality read depth >18 (read depth below 18 reads). Population frequency (removes variants with a frequency over 1 % in population databases).

Appendix D – NGS results

Agarose gel

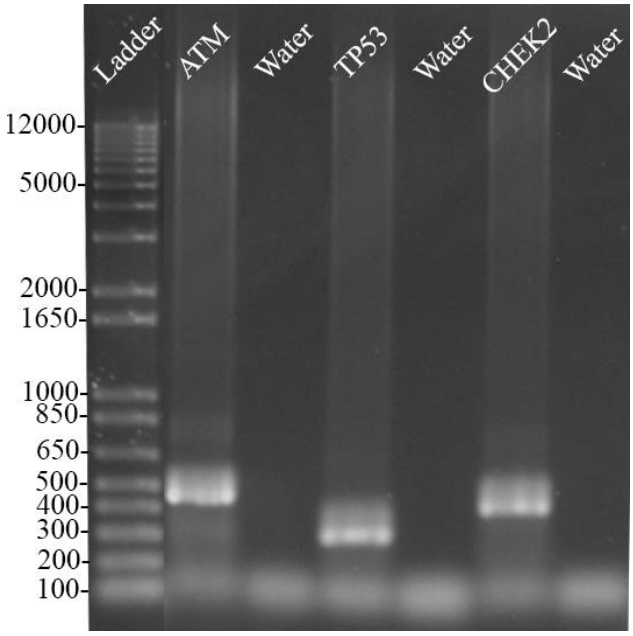


Figure 27. Agarose gel electrophoresis of the PCR-products before A'SAP purification and Sanger sequencing. The figure shows the samples with controls consisting of the master mix with water and no DNA. The PCR products include the ATM sample from TSC2 (lane 2), the TP53 sample from TSC13 (lane 4) and the CHEK2 sample from TSC12 (lane 6). The ladder used is 1Kb plus DNA molecular marker (Invitrogen by Thermo Fisher Scientific).

Table 26. Overview of the predictions of the variants identified in the 48 patient investigated by NGS. The predictions were used as part of the classification of the variants.

Variant	*Align GVGD*	*SIFT*	*MutationTaster*	*PolyPhen2*		*Comment on splice prediction*
				Hum Div	Hum Var	
ATM c.162T>C	-	-	-	-	-	-
ATM c.2572T>C	C0	Tolerated (0.73)	Polymorphism (1.0)	Possibly damaging (0.732)	Benign (0.325)	-
ATM c.3245_3247delinsTGAT	-	-	-	-	-	-
ATM c.5071A>C	C0	Tolerated (0.58)	Polymorphism (1.0)	Benign (0.001)	Benign (0.002)	-
ATM c.5229A>G	-	-	-	-	-	-
ATM c.5793T>C	-	-	-	-	-	-
ATM c.8734A>G	C65	Deleterious (0.00)	Disease causing (1.0)	Probably damaging (1.000)	Probably damaging (1.000)	-
BRIP1 c.517C>T	C15	Deleterious (0.01)	Disease causing (1.0)	Probably damaging (1.000)	Probably damaging (0.997)	-
BRIP1 c.577G>A	C0	Tolerated (0.17)	Polymorphism (1.0)	Benign (0.000)	Benign (0.001)	-
BRIP1 c.2087C>T	C65	Deleterious (0.00)	Disease causing (1.0)	Probably damaging (1.000)	Probably damaging (1.000)	-
CDH1 c.136C>G	C0	Tolerated (0.17)	Polymorphism (0.997)	Benign (0.285)	Benign (0.099)	-
CDH1 c.345G>A	-	-	-	-	-	-
CDH1 c.699C>T	-	-	-	-	-	-
CHEK2 c.319+2T>A	-	-	-	-	-	Indication of loss of splice site
MLH1 c.-7C>T	-	-	-	-	-	-
MLH1 c.803A>G	C15	Deleterious (0.02)	Disease causing (1.0)	Probably damaging (0.993)	Probably damaging (0.955)	Indication of new 5' splice site (SSF, NNSPLICE)
MLH1 c.1379A>C	C0	Tolerated (0.46)	Polymorphism (1.0)	Benign (0.000)	Benign (0.000)	-
MSH2 c.339G>A	-	-	-	-	-	-
MSH2 c.1680T>C	-	-	-	-	-	-
MSH6 c.3246G>T	-	-	-	-	-	-
NF1 c.378A>G	-	-	-	-	-	-
NF1 c.528T>A	C0	Tolerated (0.53)	Disease causing (0.991)	Probably damaging (0.979)	Possibly damaging (0.625)	-
NF1 c.587-10_587-9delTT	-	-	-	-	-	Weaker 3' splice site downstream

Table 26 continues.

Variant	*Align GVGD*	*SIFT*	*MutationTaster*	*PolyPhen2*		*Comment on splice prediction*
				Hum Div	Hum Var	
NF1 c.1810T>C	-	-	-	-	-	-
NF1 c.5225A>G	C0	Tolerated (0.07)	Disease causing (1.0)	Benign (0.000)	Benign (0.000)	Indication of new 3' splice site (MES and HSF)
NF1 c.5793T>C	-	-	-	-	-	-
NF1 c.7354C>T	C0	Deleterious (0.00)	Disease causing (1.0)	Probably damaging (1.000)	Probably damaging (0.991)	-
PALB2 c.2794G>A	C0	Tolerated (0.23)	Disease causing (0.717)	Probably damaging (1.000)	Probably damaging (0.993)	Indication of new 5' splice site (SSF)
PMS2 c.52A>G	C25	Deleterious (0.00)	Disease causing (1.0)	Probably damaging (0.998)	Probably damaging (0.994)	-
PMS2 c.1437C>G	C0	Tolerated (0.59)	Polymorphism (1.0)	Benign (0.000)	Benign (0.000)	-
PMS2 c.1688G>T	C0	Tolerated (0.21)	Polymorphism (1.0)	Benign (0.029)	Benign (0.016)	Indication of new 3' splice site (MES×2)
RAD51C c.376G>A	C0	Tolerated (0.54)	-	Benign (0.066)	Benign (0.217)	-
RAD51c c.790G>A	C0	Tolerated (0.33)	-	Benign (0.057)	Benign (0.100)	-
TP53 c.818G>A	C25	Deleterious (0.00)	Disease causing (1.0)	Possibly damaging (0.831)	Possibly damaging (0.625)	-

Appendix E – *BRCA2* c.8331+2C>T results

Agarose gel

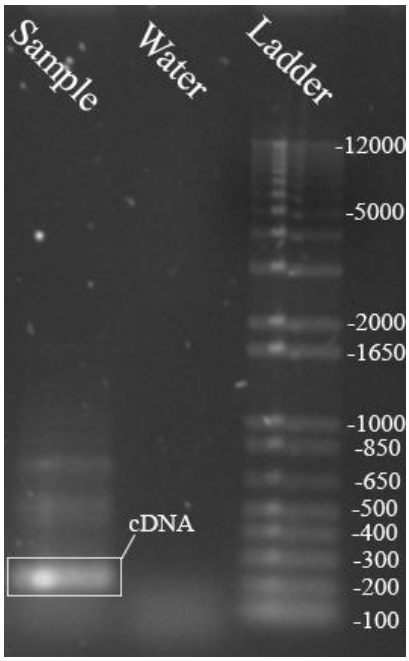


Figure 28. 1 % agarose gel for the APRT test. The figure shows the sample with the BRCA2 c.8331+2C>T variant and a water control for the master mix. The band which contains the cDNA is marked with a white square. The ladder is 1Kb plus DNA molecular marker (Invitrogen by Thermo Fisher Scientific).

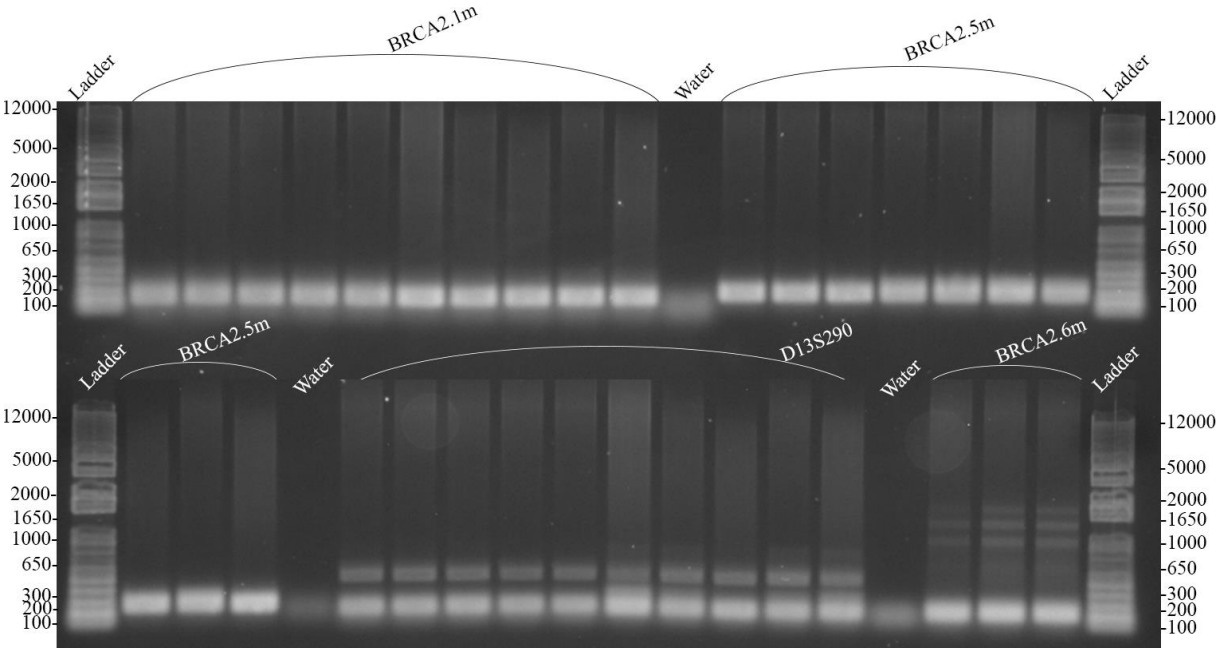


Figure 29. 1 % agarose gel from some of the samples and markers analysed using microsatellite analysis. The samples were analysed with gel electrophoresis after PCR (Appendix B – PCR-programs, Table 25), to see if there were any products present in the samples. This figure show samples run with primer sets BRCA2.1m, BRCA2.5m, D13S290 and BRCA2.6m, as well as one water control per primer set. The ladder used is 1Kb plus DNA molecular marker (Invitrogen by Thermo Fisher Scientific).

Fragments

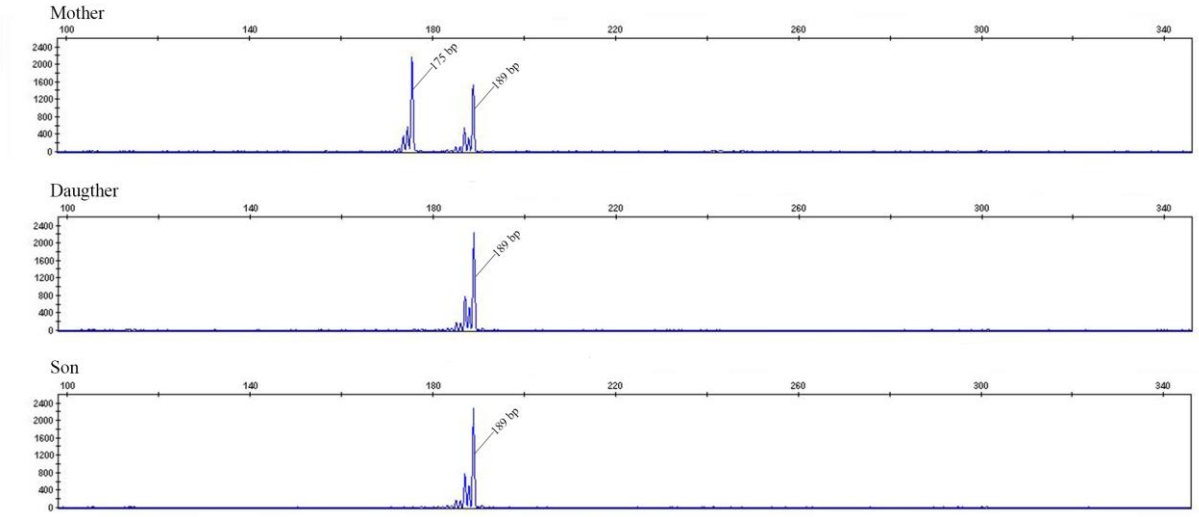


Figure 30. Family 11, marker D13S290.

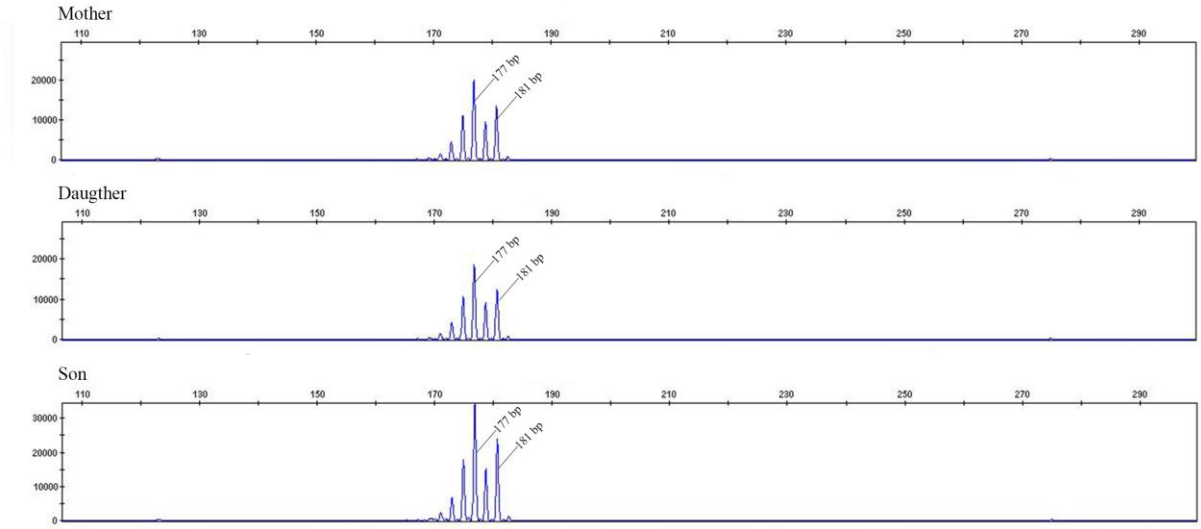


Figure 31. Family 13, marker BRCA2.6m

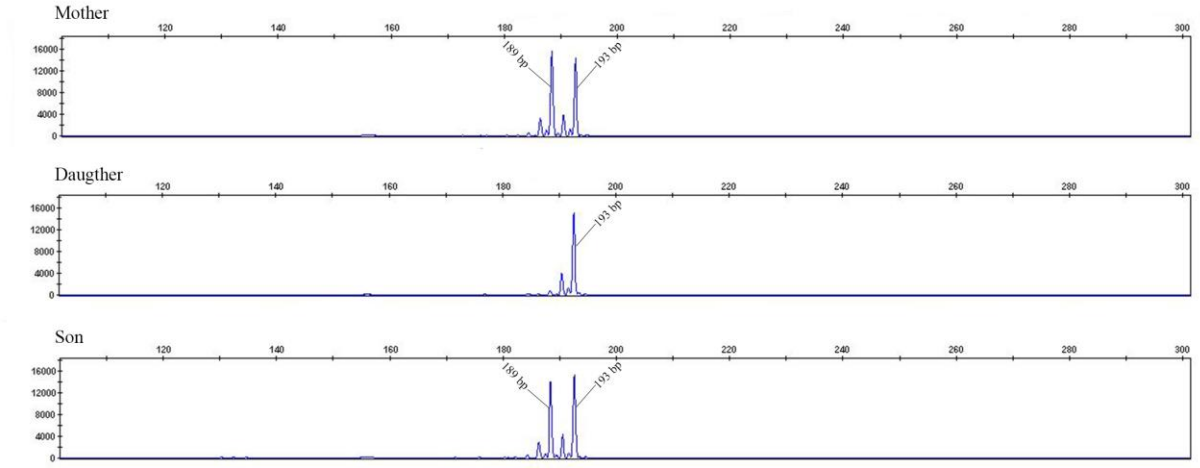


Figure 32. Family 11, marker BRCA2.1m

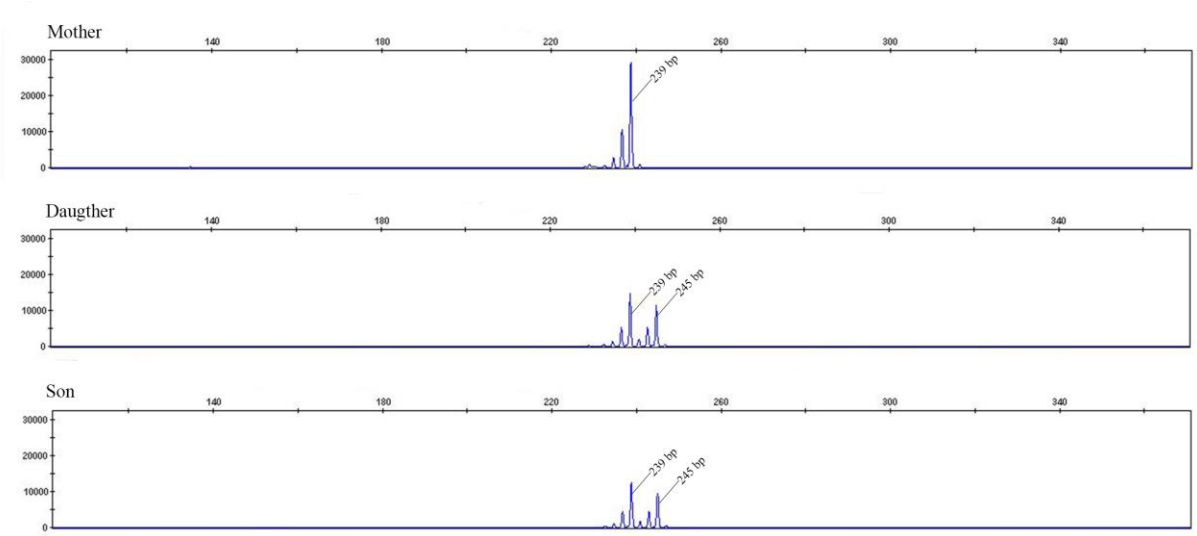


Figure 33. Family 11, marker BRCA2.5m

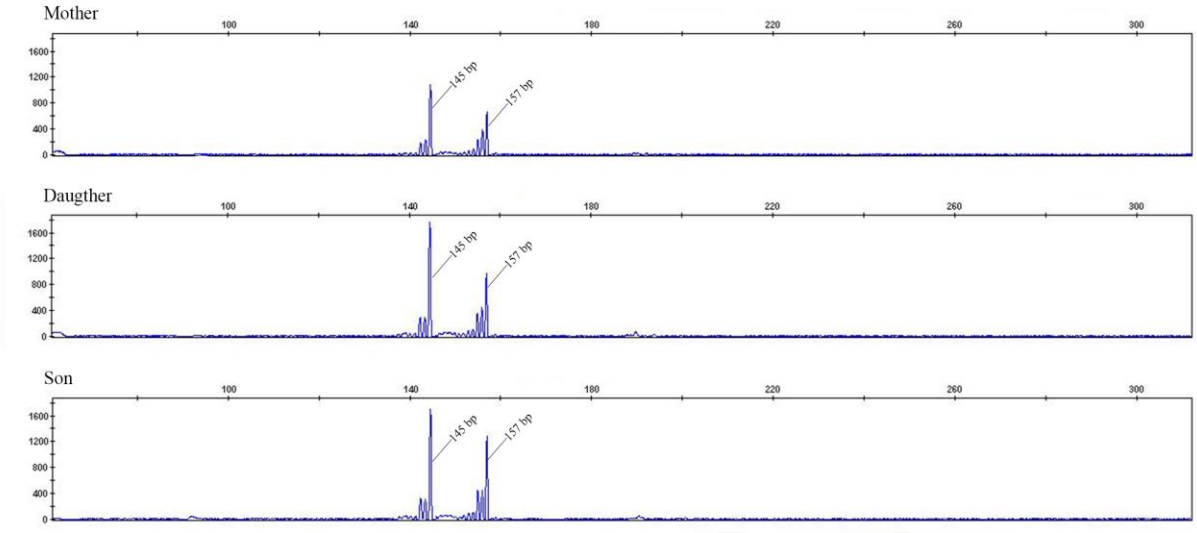


Figure 34. Family 5, marker D13S267

Table 27. The results from fragment analysis for the five primer sets used in this this part of the study. The results presented according to which family the individual belongs.

Family	*Family member*	*D13S290*	*BRCA2.6m*	*BRCA2.1m*	*BRCA2.5m*	*D13S267*
1	II:1	175 / 187	167 / 183	191/ 193	237/ 239	145 / 151
	II:2	175 / 187	167 / 177	191/ 193	239/ 241	145 / 151
	I:1*	187 / 189	167 / 177	191/ 193	239/ 245	151 / 157
2	II:2	175 / 189	183	191/193	237/ 235	151
	II:1*	175 / 189	183	191/ 193	239/ 245	?
3	I:1	175	177 / 181	189/ 193	237/ 239	145 / 151
	I:2*	175	177 / 183	189/ 193	237/ 239	145 / 151
	II:1	175	?	189	237/ 239	145 / 151
4	II:2*	175 / 189	177 / 187	189/ 193	239/ 245	145 / 157
	II:1*	175 / 189	177 / 187	189/ 193	239/ 245	145 / 157
5	II:1	175	177 / 183	191/ 193	239/ 243	145 / 157
	II:2	175	167 / 183	191/ 193	239/ 243	145 / 157
	I:2*	175	181 /183	193	239/ 245	145 / 157
6	II:1	189	179 / 181	189 / 193	?	145 / 149
	II:2	189	179 / 181	189 / 193	239	145 / 149
	II:3	189	179 / 181	189 / 193	239	144 / 149
7	II:2	189	177 / 181	189 / 193	239 / 245	159
	II:1*	189	177 / 181	189 / 193	239 / 245	159
8	I:1*	175 / 189	177 / 183	193	239 / 245	149 / 155
	II:1*	175 / 189	175 / 177	191 / 193	239 / 245	149 / 155
	II:3	175	177 / 183	193	239 / 241	149
	II:2*	175 / 189	177 / 183	191 / 193	241 / 245	149/ 155
	II:4*	175 / 189	181 / 183	193	239	145 / 149
9	III:2*	175 / 189	177 / 181	193	239 / 245	145 / 155
	IV:3	175 / 189	167 / 177	193	239 / 241	145
	III:5*	175	177 / 181	193	239 / 245	145 / 155
	III:7	175	177	193	239	145 / 149
10	I:1*	175 / 189	177	189 / 193	245 /247	
	II:1	175	177	189 / 193	239 / 247	145 / 149
	II:2*	175 / 189	177	193	239 / 245	
11	I:1	175 / 189	179	189 / 193	239	145
	II:2*	189	167 / 179	193	239 / 245	145 / 155
	II:1*	189	167 / 179	189 / 193	239 / 245	145 / 155
12	IV:2*	189	177	193 / 207	239 / 245	149 / 155
	VI:1	175 / 189	177	193	239	145
	VI:2*	189	177	193	239 / 245	151 / 155
	IV:5*	175 / 189	177 /179	189 / 193	239 / 245	155
	IV:4*	189	177	193 / 207/0	239 / 245	151 / 155
	V:5*	175 / 189	167 / 177	193	239 / 245	144 / 155
	V:6	175 / 177	179	189 / 193	239	145 / 155
	III:5*	175 / 189	177 / 183	193	239 / 245	145 / 155
13	I:1*	175 / 185	177 / 181	193 / 209	239 / 245	155 / 157
	II:2*	185 / 189	177 / 181	193	245	157
	II:1*	175 / 189	177 / 181	193	245	157

*-have the *BRCA2* c.8331+2C>T variant

Table 27 continues.

Family	*Family member*	*D13S290*	*BRCA2.6m*	*BRCA2.1m*	*BRCA2.5m*	*D13S267*
14	I:1*	175	167 / 181	193	241 / 245	143 / 151
	II:4*	175	165 / 181	193	239 / 245	143
	II:1	175	165 / 167	193	239 / 241	145 / 151
	II:2	175 / 189	167 / 179	189 / 193	239 / 241	143 / 151
	II:3*	175	165 / 181	193	239 / 245	143
15	I:1*	189	177 / 185	193	239 / 245	143 / 151
	II:3*	187 / 189	177 / 183	193	239 / 245	151 / 157
	III:1	177 / 187	177 / 183	193	239	157
	II:2	187 / 189	183 / 185	193	239	143 / 157
16	II:4*	189	177 / 187	193	239 / 245	147 / 157
	III:3*	177 / 189	177	193	239 / 245	157
	III:2*	189	177 / 183	193	239 / 245	147
	III:1*	177 / 189	177	193	239 / 245	157
	II:1	175 / 189	177 / 181	189 / 205	239	
	II:2	175 / 189	177 / 181	189 / 205	239	143 / 151
17	II:3*	175 / 189	167 / 177	193	241 / 245	157
	III:1*	189	177	189 / 193	239 / 245	151 / 157
	II:2	175 / 189	177 / 181	189 / 193	237	143 / 157
	II:1*	175 / 177	167 / 177	193	241 / 245	157
18	II:2	175	181	193	241 / 245	159
	II:1*	175 / 189	177 / 181	193	241 / 245	159
19	III:3*	189	177	193	239 / 245	144 /157
	IV:1	189	177	189 / 193	239	143 / 151
	IV:2*	175 / 189	175 /177	189 / 193	239 / 245	143 / 147
	III:6*	177 / 189	177 / 181	193	239 / 245	157
	IV:3*	175 / 189	177 / 181	189 / 193	239 / 245	143 / 157
	III:7*	175 / 189	177	193 /205	239 / 245	157
	III:8*	175 / 189	177	193 / 205	239	157
20	II:1*	175 / 191	177 / 183	193	245	145 / 157
	II:2*	175 / 191	177	193	241 / 245	159
	III:1*	189 / 191	177 / 183	193	239 / 245	151 / 157
	III:2	175	177 / 185	193 / 209	239 / 241	151 / 159
	III:3*	175 / 189	177 / 183	193	239 / 245	151 / 157
21	I:1*	175	181 / 183	193	239 / 245	157
	II:2	175 / 177	177 / 183	189 / 193	239	
22	Male*	175 / 189	177 / 183	193 / 195	239 / 245	143 / 157
23	I:1*	175 / 189	177 / 183	193	239 / 245	157
	II:1*	175 / 189	177	193	241 / 245	151 / 157
	II:2	175	177 / 183	193	239 / 241	151 / 157
24	II:1*	189	177	189 / 193	239 / 245	157
	II:2*	175 / 189	177	193 / 209	241 / 245	143 / 157
	II:3*	189	177	189 / 193	239 / 245	157

*-have the *BRCA2* c.8331+2C>T variant

Table 27 continues

Family	*Family member*	*D13S290*	*BRCA2.6m*	*BRCA2.1m*	*BRCA2.5m*	*D13S267*
25	II:2*	175	181 / 183	193	241 / 245	157
	III:1	175	177 / 183	193	239 /241	157
	III:3*	175 / 189	179	193	239 / 245	151 / 157
	II:4*	175	181 / 183	193	241 / 245	151 / 157
26	II:1*	175	183	193	239 / 245	151 / 157
	II:2	175 / 189	177 / 183	193	239	151
	II:3*	175	177 / 183	193	239 / 245	143 / 157
	III:1	175 / 189	167 / 177	193	239 / 247	143

*-have the *BRCA2* c.8331+2C>T variant

Pedigrees

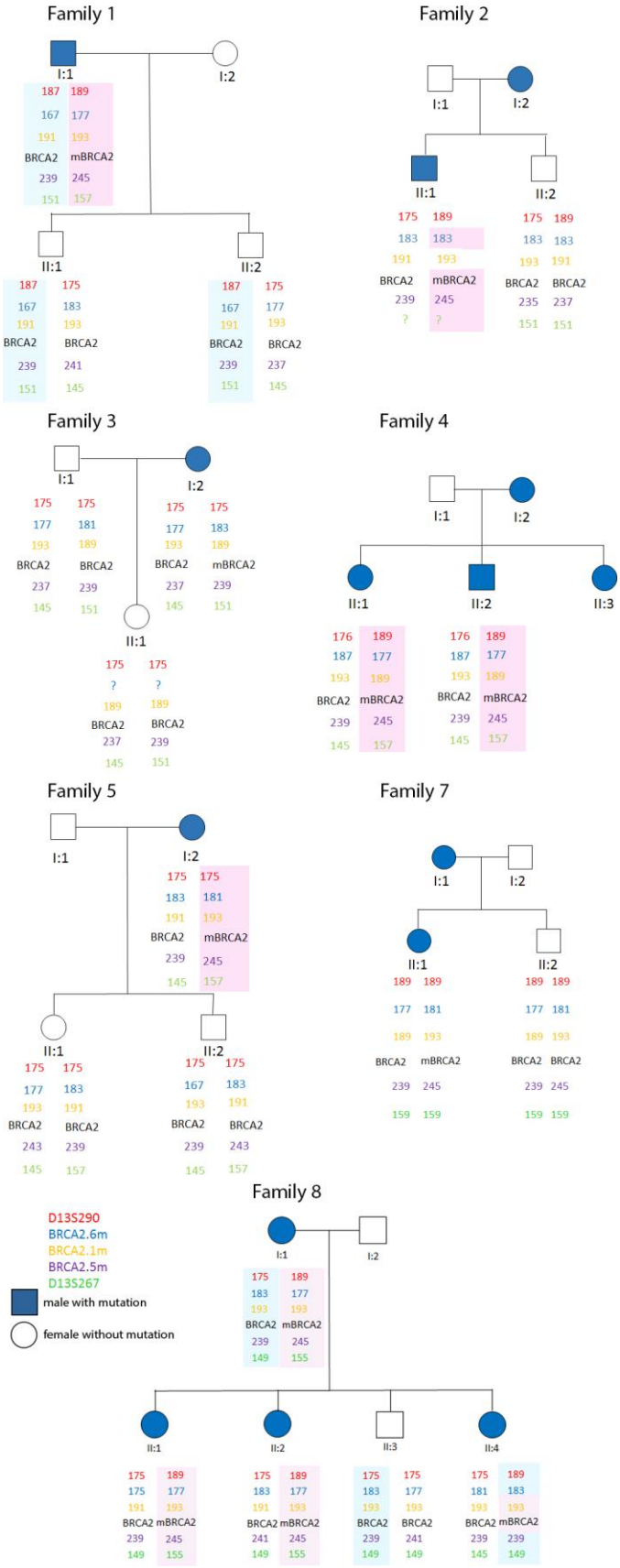


Figure 35. Families 1, 2, 3, 4, 5, 7 and 8 with the fragment size from the microsatellite analysis. The alleles which segregate with the BRCA2 c.8331+2C>T variant are marked in red. The alleles marked in blue segregated without the variant. mBRCA2: BRCA2 c.8331+2C>T.

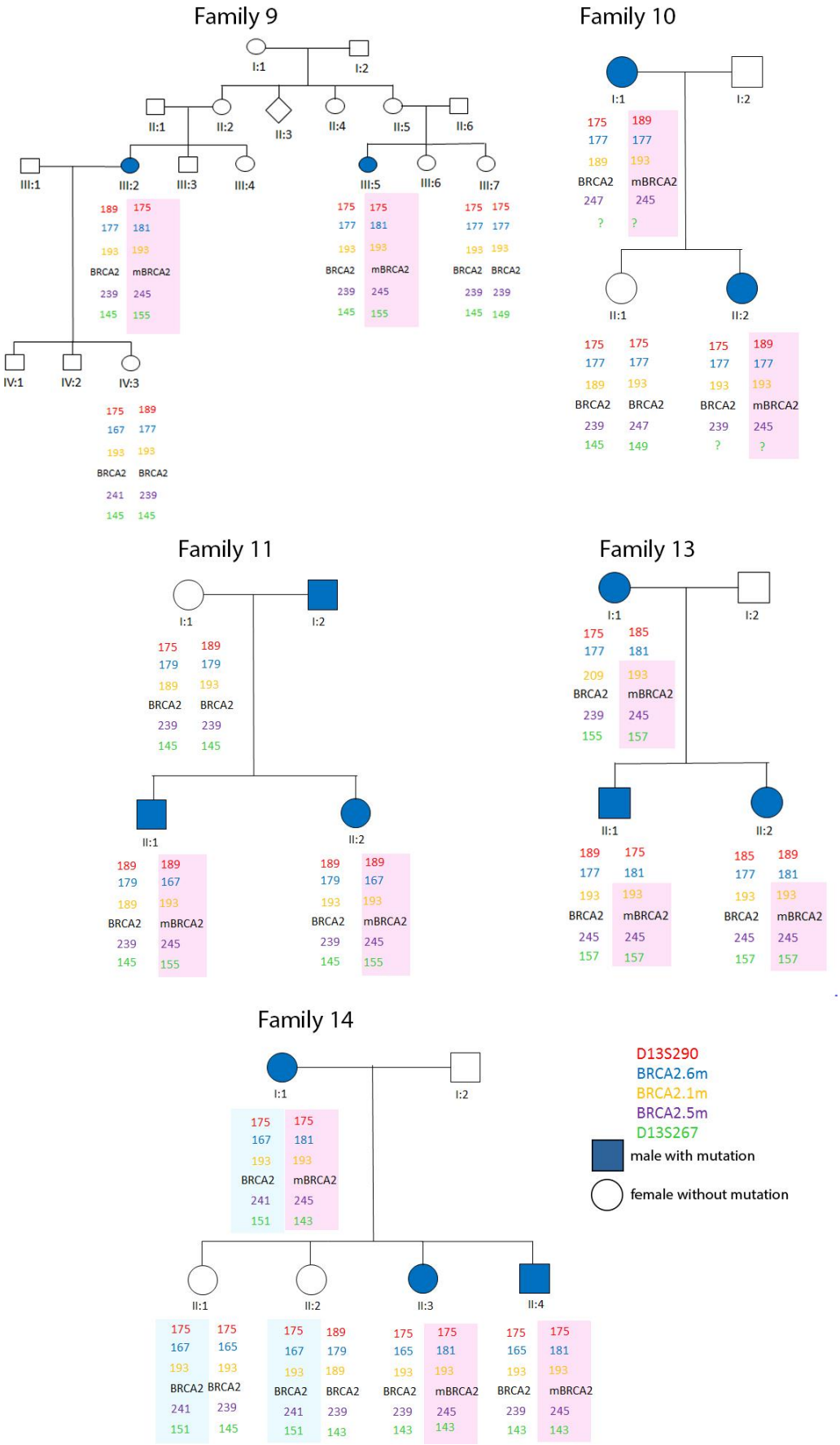


Figure 36. Families 9, 10, 11, 13 and 14 with the fragment size of the markers from the microsatellite analysis. The alleles which segregate with the BRCA2 c.8331+2C>T variant are marked in red. The alleles marked in blue segregated without the variant. mBRCA2: BRCA2 c.8331+2C>T.

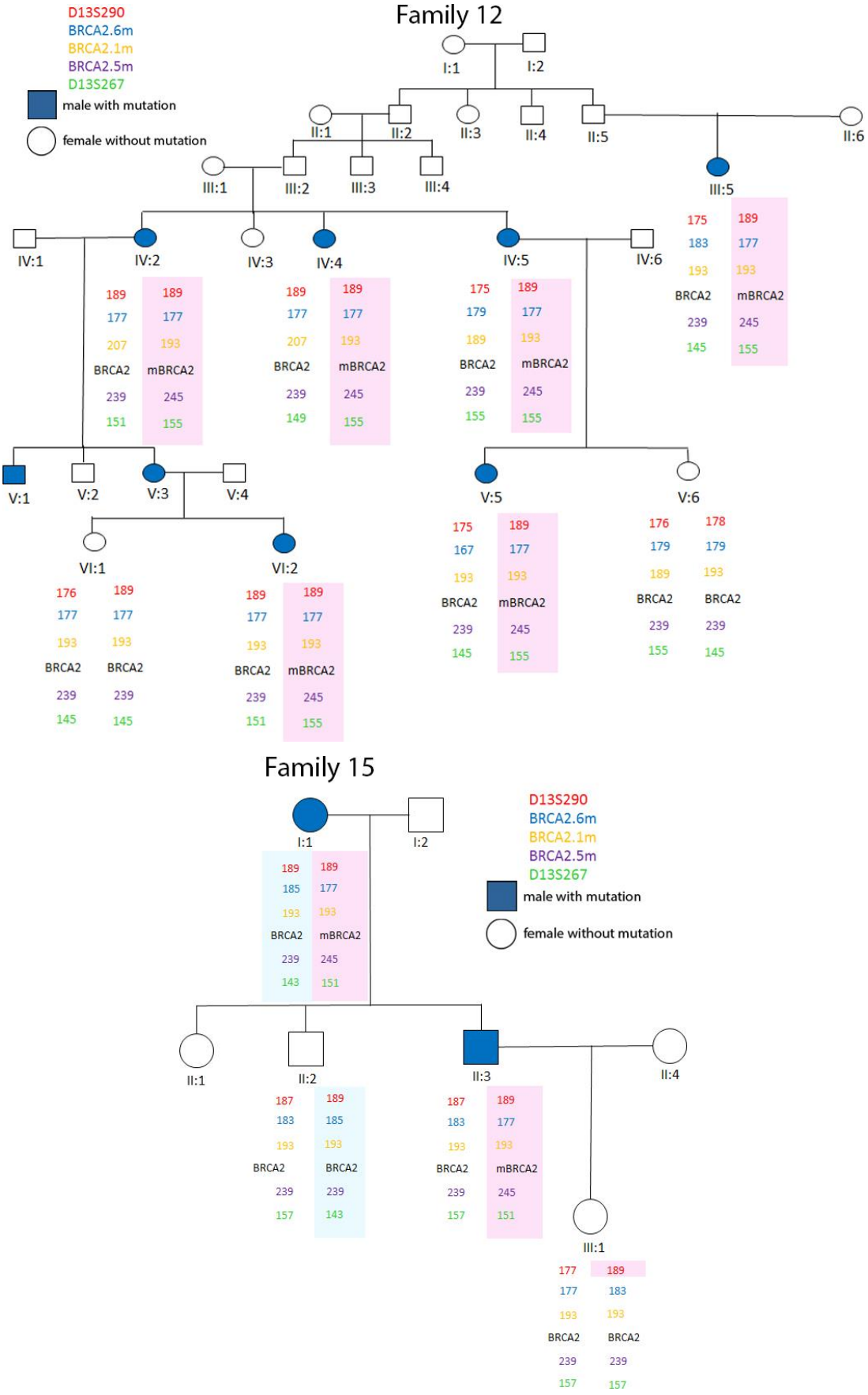


Figure 37. The pedigree of Families 12 and 15 with the different sizes of the microsatellite markers. The alleles, which were segregated with the BRCA2 c.8331+2C>T variant for this family, are marked in red. The alleles marked in blue segregated without the variant. mBRCA2: BRCA2 c.8331+2C>T.

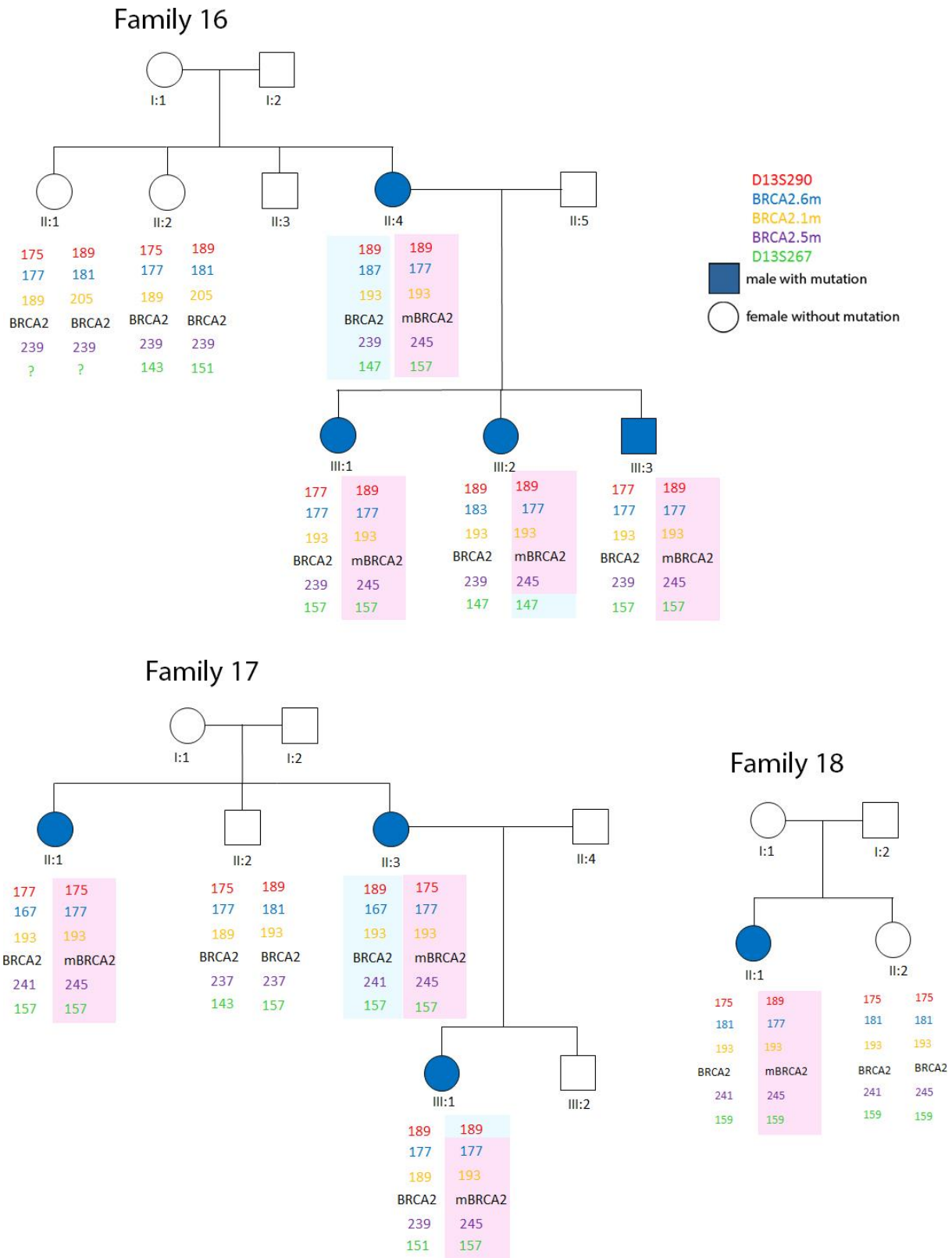


Figure 38. Families 16, 17 and 18 with the fragment size of the markers used for the microsatellite analysis. The alleles which segregate with the BRCA2 c.8331+2C>T variant are marked in red. The alleles marked in blue segregated without the variant. mBRCA2: BRCA2 c.8331+2C>T.

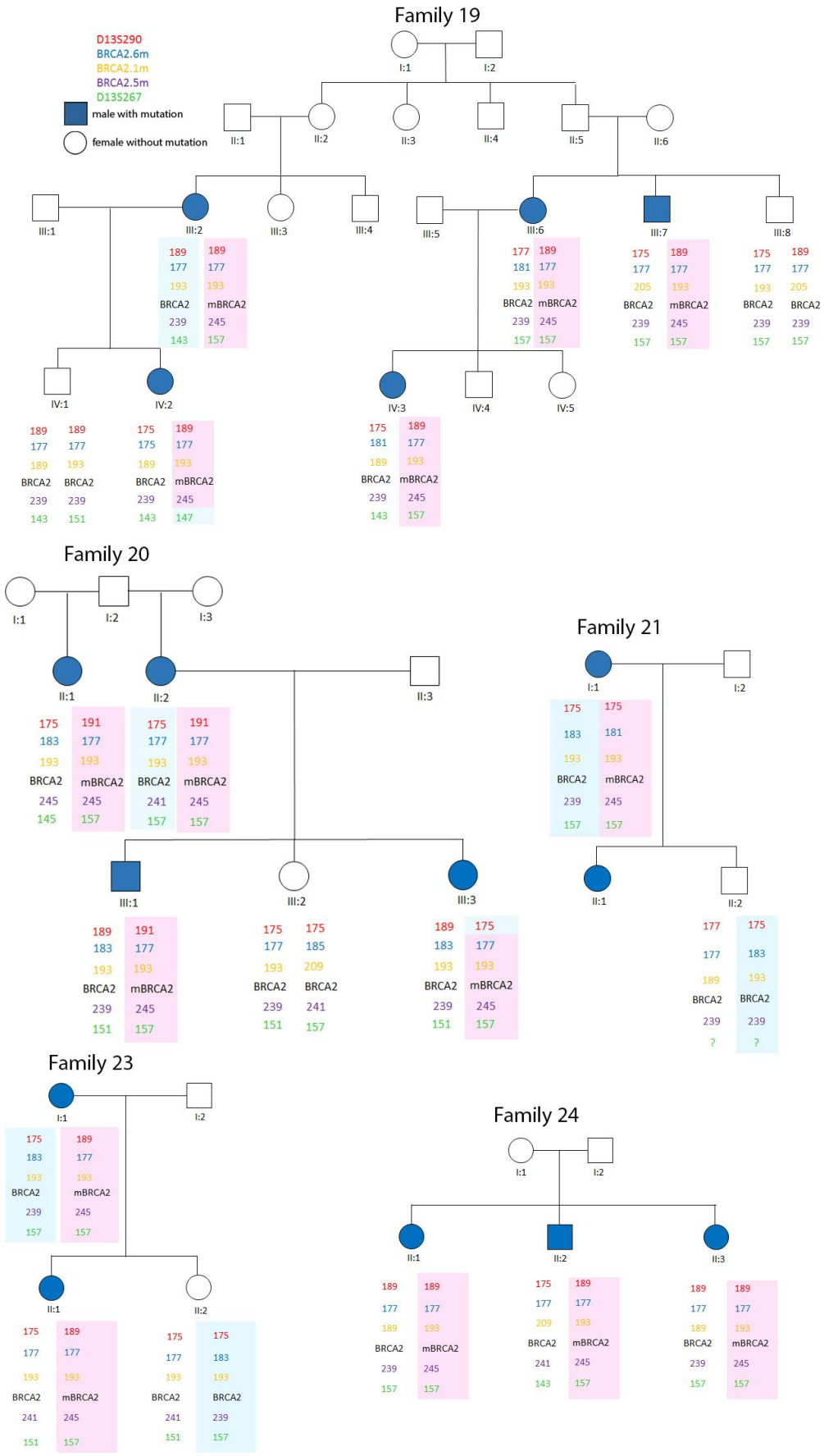


Figure 39. Families 19, 20, 21, 23 and 24 with the fragment size of the markers used for the microsatellite analysis. The alleles which segregate with the BRCA2 c.8331+2C>T variant are marked in red. The alleles marked in blue segregated without the variant. mBRCA2: BRCA2 c.8331+2C>T.

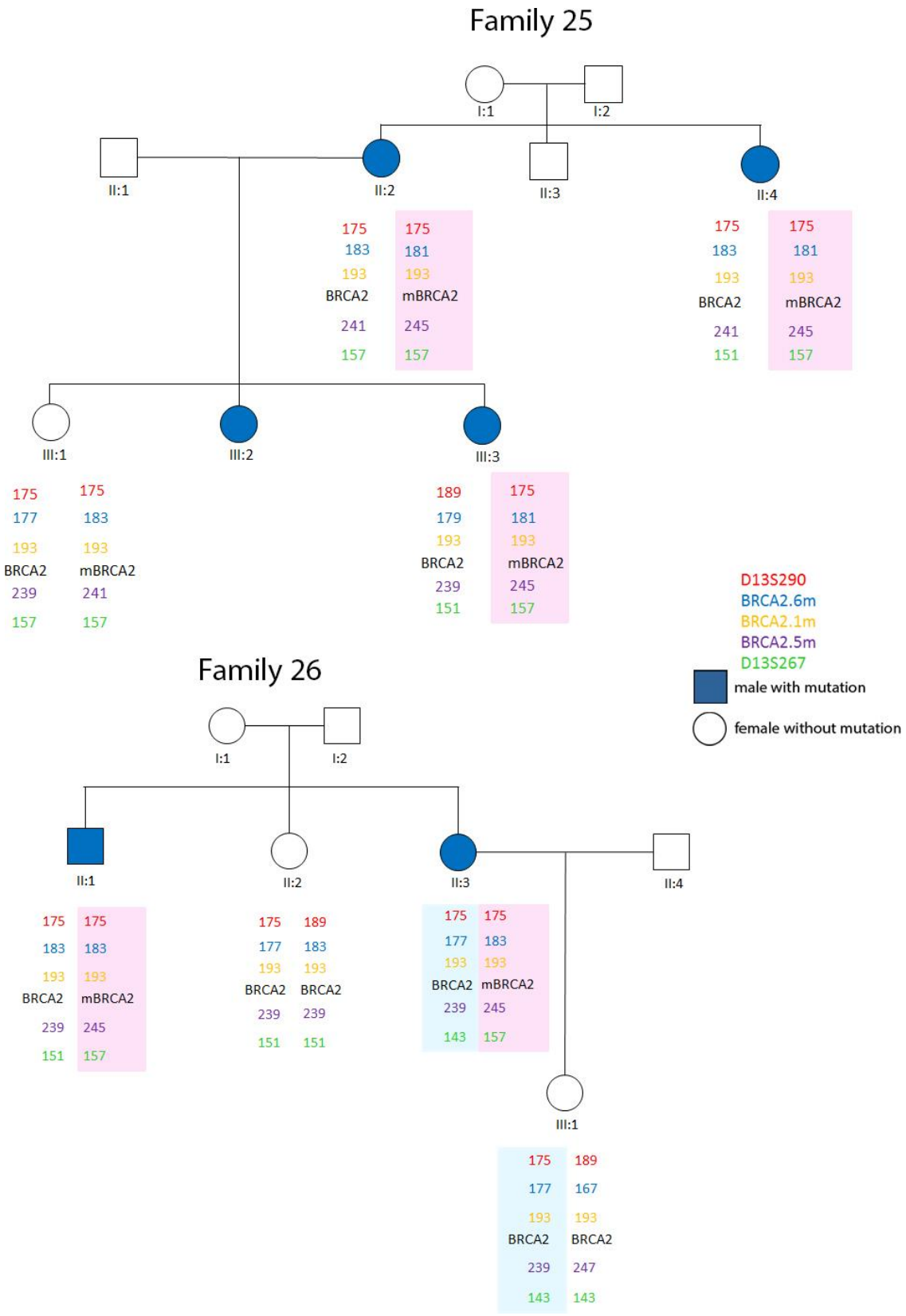


Figure 40. Families 25 and 26 with the fragment size of the markers used for the microsatellite analysis. The alleles which segregate with the BRCA2 c.8331+2C>T variant are marked in red. The alleles marked in blue segregated without the variant. mBRCA2: BRCA2 c.8331+2C>T.