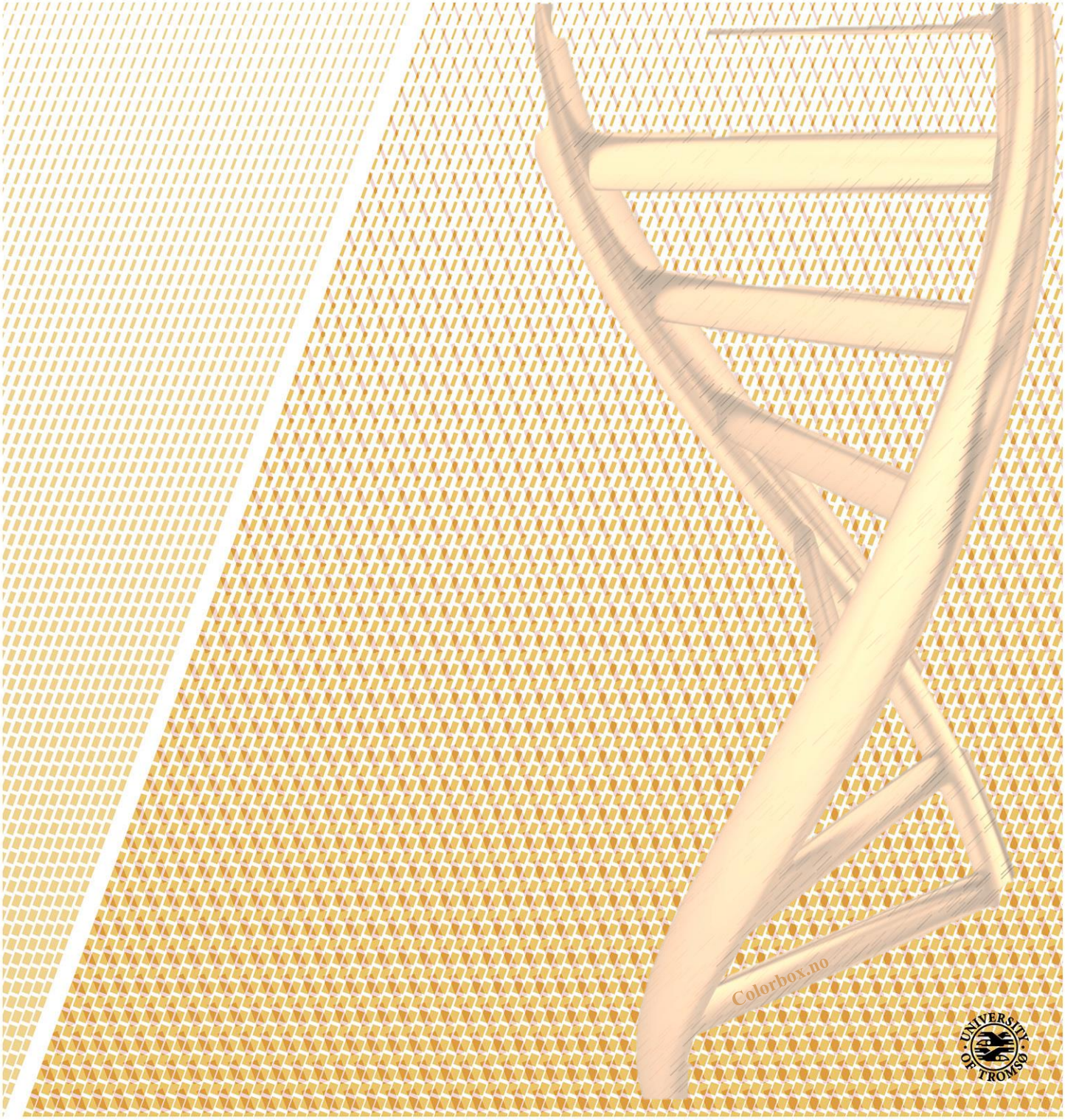
UiT
THE ARCTIC UNIVERSITY OF NORWAY
HELSE NORD
UNIVERSITETSSYKEHUSET NORD-NORGE
UNIVERSITY HOSPITAL OF NORTH NORWAY

Faculty of Health Sciences
Department of Clinical Medicine

What are the molecular consequences of germline mutations in breast and ovarian cancer susceptibility genes in a Norwegian breast and ovarian cancer population?

Elisabeth Jarhelle

A dissertation for the degree of Philosophiae Doctor – March 2018



What are the molecular consequences of germline mutations in breast and ovarian cancer susceptibility genes in a Norwegian breast and ovarian cancer population?

Elisabeth Jarhelle

A dissertation for the degree of Philosophiae Doctor

Department of Medical Genetics
University Hospital of North Norway

&

Department of Clinical Medicine, Faculty of Health Sciences
UiT – The Arctic University of Norway

Acknowledgements

This project was carried out at the Department of Medical Genetics, Division of Child and Adolescent Health at the University Hospital of North Norway (UNN). The project was financed by Helse Nord RHF, Odd Fellow and the Division of Child and Adolescent Health at UNN.

First, I would like to thank my supervisor Marijke Van Ghelue and co-supervisor Hilde Monica F. R. Stensland. I am forever grateful for sticking out with me all these years, as they also were my supervisors during my master thesis. They have helped and educated me enormously through all these years! Marijke Van Ghelue was the one who introduced me to the field of genetics and hereditary breast and ovarian cancer. I want to thank her for always being available to answer questions, fruitful discussions and giving me feedback. Second, I want to thank Hilde Monica F. R. Stensland for her extraordinary eye for details, which I truly appreciate!

I would also like to thank Nina Strømsvik and Geir Åsmund Myge Hansen for their massive amount of work regarding the Next Generation Sequencing project. Nina has a contagious enthusiasm for the research and people involved, and Geir has an excellent expertise and a never-ending supply of comic bits during the workday.

I would also like to thank the newly formed cancer research group, consisting of my supervisors MVG and HMFRS, Nina Strømsvik, Geir Å.M. Hansen, Christoffer Jonsrud, Randi Marlene Dahle, Siri Briskemyr and Beate Iren Vestli. It has been very rewarding to discuss the research with all of you.

I would like to give a special thanks to my former labmate, Vigdis Brox, who now is living the good life as a retiree, and my current labmates Elisabeth Kjeldsen Buvang and Veronika Franekova. You all have taught me a lot during my time at the department and I love the time we spend in the lab together!

To all my co-workers at the Department of Medical Genetics, I doubt the existence of a better workplace and it is all thanks to you.

I would also like to thank all patients and participants that have contributed to the study.

Last, but not least, I would like to thank family and friends for being there, supporting my work, entertaining and for keeping me safe and sound. I could not have done this without all of you! A special thanks to Nina Løvhaug for looking after me and my dog Alma, making sure we got the fresh air and activity needed to function during the last few months of writing.

Elisabeth Jarhelle

Tromsø, March 2018

VUS – a Very Unhelpful Statement

– Nazneen Rahman (2017)

Contents

Paper I	Characterization of *BRCA1* and *BRCA2* variants found in a Norwegian breast or ovarian cancer cohort.
Paper II	Comparing the quality of RNA preserved in PAXgene and Tempus Blood RNA tubes using *BRCA1* splicing events as a model system.
Paper III	Identifying sequence variants contributing to hereditary breast/ovarian cancer in *BRCA1/2* negative breast and ovarian cancer patients.

List of papers

This thesis is based on three papers, referred to in the text as Paper I, II and III.

Paper I **Jarhelle E, Stensland HMFR, Mæhle L, Van Ghelue M.**

Characterization of *BRCA1* and *BRCA2* variants found in a Norwegian breast or ovarian cancer cohort.

Familial Cancer, 2017. **16**(1): p. 1-16.

Paper II **Jarhelle E, Stensland HMFR, Van Ghelue M.**

Comparing the quality of RNA preserved in PAXgene and Tempus Blood RNA tubes using *BRCA1* splicing events as a model system.

[Manuscript]

Paper III **Jarhelle E, Stensland HMFR, Hansen GÅM, Skarsfjord S, Jonsrud C, Ingebrigtsen M, Strømsvik N and Van Ghelue M.**

Identifying sequence variants contributing to hereditary breast/ovarian cancer in *BRCA1/2* negative breast and ovarian cancer patients.

[Manuscript]

Abbreviations

Δ	Skipping of
3’ss	3’ Splice Site
5’ss	5’ Splice Site
A	Adenine
aa	Amino acid
AKT1	AKT serine/threonine kinase 1
AlignGVGD	Align Grantham Variation; Grantham Deviation
APC/C	Anaphase promoting complex/cyclosome
APRT	Adenine Phosphoribosyltransferase
ATM	ATM serine/threonine kinase
ATP	Adenine triphosphate
ATPase	Adenosine triphosphatase
BARD1	BRCA1 associated RING domain 1
BCDX2	RAD51B-RAD51C-RAD51D-XRCC2
BIC	Breast Cancer Information Core
BOC	Breast and ovarian cancer
bp	Base pairs
BPS	Branch point site
BRCA1	BRCA1, DNA repair associated
BRCA2	BRCA2, DNA repair associated
BRCT	BRCA1 C-Terminal
BRIP1	BRCA1 interacting protein C-terminal helicase 1
C	Cytosine
ca.	Circa
CDC20	Cell division cycle 20
CDC25A	Cell division cycle 25A
CDC25C	Cell division cycle 25C
CDH1	Cadherin 1
CDK”x”	Cyclin dependent kinase “x”
CDKN1A	Cyclin dependent kinase inhibitor 1A

cDNA	Complementary DNA
CE	Capillary electrophoresis
ChAM	Chromatin-association motif
CHEK2/CHK2	Checkpoint kinase 2
CRAL-TRIO	CRALBP combined with Trio (PF00650)
CTNND1	Catenin delta 1
CX3	RAD51C-XRCC3
DBD	DNA binding domain
dbSNP	Database of Single Nucleotide Polymorphisms
DEAH	Aspartate-Glutamate-Alanine-Histidine
delins	Deletion and insertion
DNA	Deoxynucleic Acid
DSB	Double Stranded Break
EOC	Epithelial ovarian cancer
ER	Estrogen receptor
ERCC5	ERCC excision repair 5, endonuclease
ESE	Exonic Splicing Enhancer
ESS	Exonic Splicing Silencer
ExAC	The Exome Aggregation Consortium
EXO1	Exonuclease 1
FA	Fanconi Anemia
FANC”x”	Fanconi anemia, complementation group “x”
FAT	Focal Adhesion Targeting
FATC	FAT C-terminal
FHA domain	Forkhead-associated domain
G	Guanine
G4	G-quadruplex
GAP	GTPase activating protein
GDP	Guanine diphosphate
GTP	Guanine triphosphate
gnomAD	The Genome Aggregation Database
H2AX	Histone 2A member X

HBOC	Hereditary Breast and Ovarian Cancer
HER2	Human epidermal growth factor receptor 2
HGMDp	Human Gene Mutation Database Professional
HNPCC	Hereditary Non-Polyposis Colorectal Cancer
HRR	Homology Recombination Repair
ICL	DNA interstrand crosslink
ISE	Intronic Splicing Enhancer
ISS	Intronic Splicing Silencer
LIG1	DNA ligase 1
LOH	Loss of heterozygosity
LPV	Likely pathogenic variant
MAD2L2 (REV7)	Mitotic arrest deficient 2 like 2
MDC1	Mediator of DNA-damage checkpoint 1
MES	MaxEntScan
MLH1	MutL homolog 1
MMR	Mismatch repair
MRE11	MRE11 homolog, double strand break repair nuclease
MRN	MRE11-RAD50-NBN
mRNA	Messenger RNA
MSH2	MutS homolog 2
MSH6	MutS homolog 6
mTOR	Mechanistic target of rapamycin kinase
MutLα	MLH1/PMS2 heterodimer
MutSα	MSH2/MSH6 heterodimer
MutSβ	MSH2/MSH3 heterodimer
NBN	Nibrin
NES	Nuclear Export Signal
NF1	Neurofibromin 1
NGS	Next Generation Sequencing
NHEJ	Non-Homologous End Joining
NLS	Nuclear Localization Signal
NMD	Nonsense mediated mRNA decay

NOP53	NOP53 ribosome biogenesis factor
OSC	Ovarian serous carcinoma
p	Start/5’end of exon/intron (in regard to splicing)
p53	Tumor protein 53
PALB2	Partner and localizer of BRCA2
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PDK1	Pyruvate dehydrogenase kinase 1
PI3K	Phosphatidylinositol 3-kinase
PIP	PCNA interaction protein
PIP_2	Phosphatidylinositol 4,5-bisphosphate
PIP_3	Phosphatidylinositol (3,4,5)-triphosphate
PMS2	PMS1 homolog 2, mismatch repair system component
Pol-δ (POLD1-4)	DNA polymerase delta
POLH	DNA polymerase eta
PR	Progesterone receptor
Pre-mRNA	Precursor messenger RNA
PRR	Proline-rich region
PSEN1	Presenilin 1
PTEN	Phosphatase and tensin homolog
PV	Pathogenic variant
PWWP	Proline-Tryptophan-Tryptophan-Proline
q	End/3’end of exon/intron (in regard to splicing)
RAD50	RAD50 double strand break repair protein
RAD51	RAD51 recombinase
RAD51C	RAD51 paralog C
RAD51D	RAD51 paralog D
RAF	Raf proto-oncogene, serine/threonine kinase
RAS	RAS type GTPase family
RBBP8 (CtIP)	RB binding protein 8, endonuclease
RIN	RNA integrity number
RING	Really interesting new gene

RNA	Ribonucleic acid
RNF8	Ring finger protein 8
RPA	Replication protein A
RTK	Receptor tyrosine kinase
SEM1	SEM1, 26S proteasome complex subunit
Ser	Serine
SF1	Splicing factor 1
snRNA	Small nuclear ribonucleic acid
snRNP	Small nuclear ribonucleoprotein particles
SQ/TQ cluster	Cluster of serine/threonine followed by glutamine
SR protein	Serine/Arginine-rich protein
ssDNA	Single stranded DNA
SSF	SpliceSiteFinder-like
STK11	Serine/threonine kinase 11
SUMO2	Small ubiquitin-like modifier 2
T	Thymine
TAD	Transcriptional Activation Domain
TET	Tetramerization domain
Thr	Threonine
TP53	Tumor protein 53
TP53BP1	Tumor protein p53 binding protein 1
U	Uracil
VUS	Variant of Unknown clinical Significance
WRN	Werner syndrome RecQ like helicase
XRCC"x"	X-ray repair cross complementing "x"

"x" = a number or letter. Several proteins have the same name, with a different number or letter.

1. Introduction

The human body is estimated to consist of trillions of cells [1]. There are different types of cells with different functions. Some cells are specialized in nutrient up-take, some function as a barrier against microbes, yet others are responsible for neural signaling, transporting oxygen through the body, eliciting the body's immune response etc. All these cells need to work together for the body to function properly. Cells are therefore under strict control and cell proliferation (cell growth and division), in particular, is strictly regulated by several proteins during the cell cycle. However, mutations (a change in the cell's DNA) may lead to dysregulation of the cell cycle and subsequently to abnormal cell growth and eventually cancer.

Cancer is an abnormal cell growth, where cells have acquired an unregulated cell proliferation and eventually the potential to invade the surrounding tissue (metastasize). For cells to reach this point, several events need to occur. In 2000, Hanahan and Weinberg proposed the hallmarks of cancer, which comprises six biological capabilities acquired during tumor formation (also called tumorigenesis) [2]. These hallmarks include sustaining proliferative signaling, evading growth suppressors, activating invasion/metastasis, enabling limitless replication abilities, inducing angiogenesis (development of new blood vessels) and evading cell death. In 2008, the hallmarks were revised, adding four new capabilities; including deregulating cellular energetics and avoiding immune destruction, and two enabling characteristics involving genome instability due to mutations and tumor-promoting inflammation [3].

Cancers can originate from several types of cells and are classified accordingly. The most common cancer types are carcinomas. **Carcinomas** arise from epithelial cells, which are cells lining the walls of the body cavities and channels, and cells lining the outside of the body (skin cells). Carcinomas can be further classified into squamous cell carcinoma (originating from cells responsible for protecting underlying cells) and adenocarcinoma (originating from cells secreting substances for their own protection, e.g. epithelial cells of the stomach). **Sarcomas** are another form of cancer, arising from connective tissue. This includes cancers arising from cells such as fibroblasts (cells secreting collagen), adipocytes (fat cells), osteoblasts (bone forming cells) and myocytes (muscle cells). Cancers of **hematopoietic origin** define another cancer type, including leukemia (from white blood cells), lymphomas (from B- and T-lymphocytes) and myelomas (from antibody-producing cells of the bone marrow). The last cancer group comprises the **neuroectodermal** tumors, arising from cells of the central and peripheral nervous system. This group of tumors includes glioma/glioblastoma (from non-neuronal glial cells), neuroblastoma/medulloblastoma (from primitive neuronal precursor cells) and schwannoma (from Schwann cells, cells forming sheets around the axons of neuros) [4].

All of these cancers are caused by a dysregulation of cell proliferation. Cell proliferation is under control of two different types of proteins encoded by tumor suppressor genes and proto-oncogenes. Proto-oncogenes encode proteins that stimulate cell division if conditions are right. However, upon an acquired gain-of-function mutation, these genes become oncogenes. Oncogenes encode proteins that will facilitate cell division, despite of unfavorable conditions and signals, and eventually might drive the cell to become a cancer cell. Oncogenes behave in a dominant manner; accordingly, a pathogenic variant in one of the alleles is enough for detrimental consequences. Tumor suppressor genes, on the other hand, code for proteins that negatively regulate the progression through the cell cycle. Accordingly, for tumor suppressor genes, both alleles need to lose their original function for the cell to become cancerous. Nevertheless, there are some exceptions, where pathogenic mutations in one allele of a tumor suppressor gene might be enough for development of cancerous cells, this is known as haploinsufficiency [5].

There are several types of mutations that can affect tumor suppressor genes and proto-oncogenes, ranging from large chromosomal mutations to point mutations affecting a single nucleotide (Figure 1). **Chromosomal mutations** may involve several genes and are divided into duplications (a region of a chromosome is repeated, resulting in an increased dosage of involved genes), inversions (a region of a chromosome is flipped, the gene dosage remains the same, but genes at the break points and/or regulatory elements may be affected), deletions (a region of a chromosome is lost, resulting in a decreased dosage of involved genes) and translocations (a region of one chromosome is misplaced to another chromosome, genes in the break points and/or regulatory elements may be affected). **Point mutations** are divided into substitutions, insertions/duplications and deletions of one or more nucleotides. Substitutions are further divided into synonymous mutations (encompasses nucleotide changes that do not change the encoded amino acid), nonsense mutations (nucleotide changes that change a codon into a stop codon), missense mutations (nucleotide changes that change the encoded amino acid) and splicing affecting mutations (nucleotide changes affecting the splicing pattern of messenger RNA (mRNA)). Insertions/duplications and deletions can be divided into in-frame mutations (the incorporation of or loss of additional amino acids) and frame-shifting mutations (resulting in changes in the following amino acids and eventually introducing a premature stop codon). **Copy number variations** are divided into gene amplification/deletion (or part of a gene) and expanding/decreasing trinucleotide repeats [6].

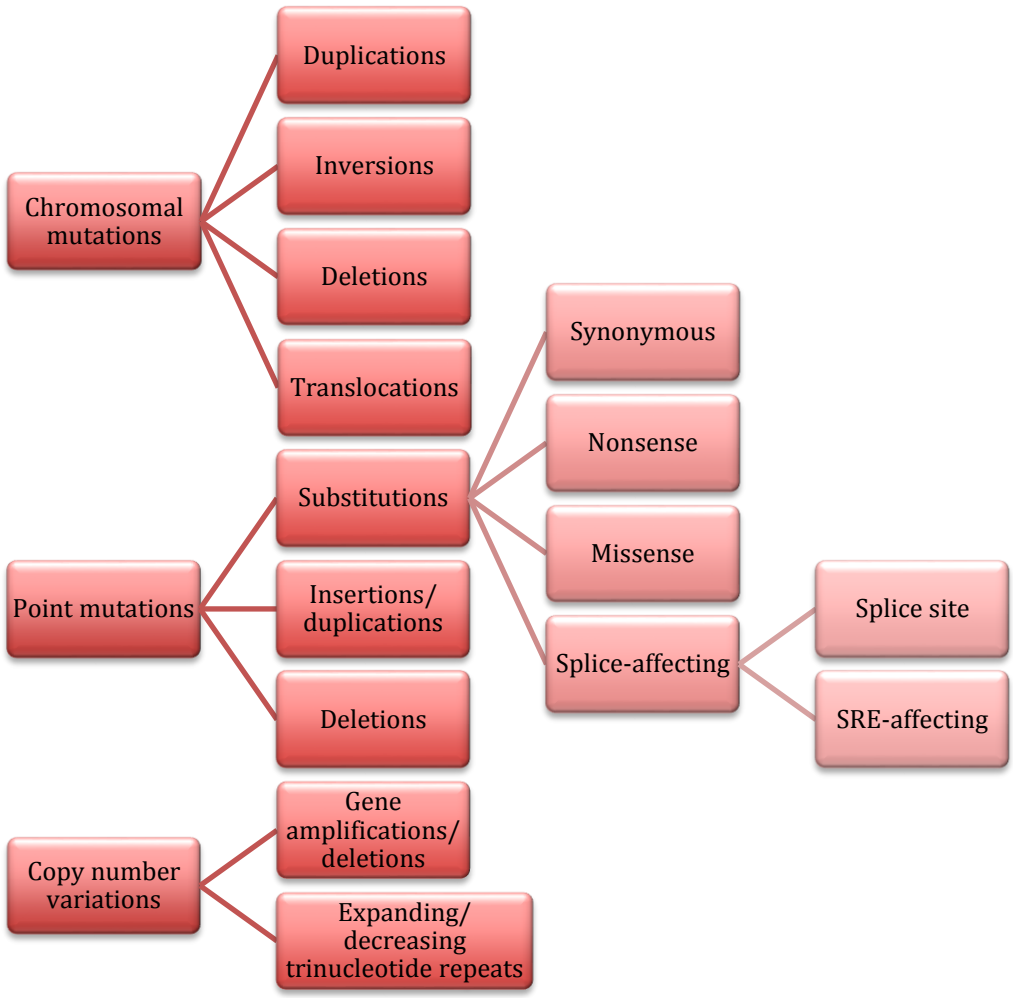


Figure 1 Types of sequence variants [6].

Mutations may affect gene expression on several levels, from transcription of the gene, splicing of the precursor mRNA (pre-mRNA), and further through changes in amino acid sequences. Transcription of the gene may be changed by mutations located in the promoter region or other regulator regions. Splicing of the pre-mRNA into mRNA can be affected by mutations changing splice sites (masking the original splice sites and/or activating cryptic splice sites) or changing splicing regulatory elements (splicing is described further in **section 1.4. Splicing**). Changes in the amino acid sequence may affect folding and/or function of the protein. For example, changing a hydrophobic amino acid (e.g. alanine, leucine or valine) with a polar or charged amino acid (e.g. arginine, threonine or serine) may produce an unstable protein which will be targeted for degradation [7]. Similarly, proteins may be rendered unstable by changing a rigid, large amino acid (e.g. proline) with a small and flexible amino acid (e.g. glycine), and vice versa. Mutations may also affect protein domains and regions, thereby altering the protein's subcellular localization, its binding properties, its enzymatic activity etc.

Mutations that affect tumor suppressor genes can be inherited through generations if they have occurred in germ cells. Such variants are inherited in an autosomal dominant fashion, even though the variants are recessive on a cellular level. Thus, for a cell to become cancerous, both copies of a tumor suppressor gene have to be mutated. This concept is known as the two-hit hypothesis, proposed by Knutson in 1971 [8]. The hypothesis explains how people with an inherited pathogenic mutation have an increased risk of cancer. Individuals without an inherited mutation need "two hits", meaning two pathogenic mutations need to occur in the same gene (on different alleles), in the same cell, before the tumor suppressor function of the gene is abolished. Individuals with an inherited pathogenic mutation already have one mutation in all cells and only need the "second hit" to lose the tumor suppression function in the cell, accordingly increasing the cancer risk.

1.1. Breast and ovarian cancer

Among women, breast cancer is the most frequently diagnosed cancer listed in the Norwegian cancer registry in 2016. A total of 3371 new female breast cancer cases and 488 new ovarian cancer cases were reported in Norway in 2016 [9]. Although breast cancer is a more common diagnosis, ovarian cancer is more lethal. In 2015 (2016 data not available), 504 new ovarian cancer cases were diagnosed, and 299 deaths were reported. Ovarian cancer is the 5th most common cause of cancer death with a five-year (2012-2016) relative survival of 48.7%, compared to 89.7% for breast cancer [9].

1.1.1. Symptoms of breast and ovarian cancer

There are many symptoms associated with breast cancer: a lump or thickening in the breast or armpit; a change in size, shape or feel of the breast; breast pain; skin changes, including puckering, dimpling, a rash or redness of the skin; change in the position of the nipple or the nipple might be inverted; fluid leaking from a nipple in a woman who is neither pregnant nor breast feeding [10].

Ovarian cancer has fewer and more diffuse symptoms, which might be the reason for the late diagnosis and the low 5-year survival percentage. Symptoms occurring with ovarian cancer includes visible increase and swelling of the abdomen, abdominal distension/increased girth, abdominal or pelvic pain, abdominal or pelvic bloating and loss of appetite [11]. Moderately informative symptoms are diarrhea, isolated abdominal pain, weight loss, change in bowl

habits, constipation, urinary frequency/urgency, vague discomfort in the upper abdomen/feeling of fullness (dyspepsia) and abnormal vaginal bleeding [11].

1.1.2. Anatomy of breasts and the female reproductive system and cancer subtypes

1.1.2.1. Breasts and breast cancer subtypes

The breasts' function is to produce and to secrete milk to feed infants. Milk is produced in glands of the breasts, also called lobules. Lobules consist of several alveoli. The produced milk is collected from the alveoli into the ductal system, which leads the milk to the nipples (Figure 2). Although breasts have an important function in feeding humans in their most vulnerable life stage, breasts are also the site of the most frequently diagnosed cancer type in women [9].

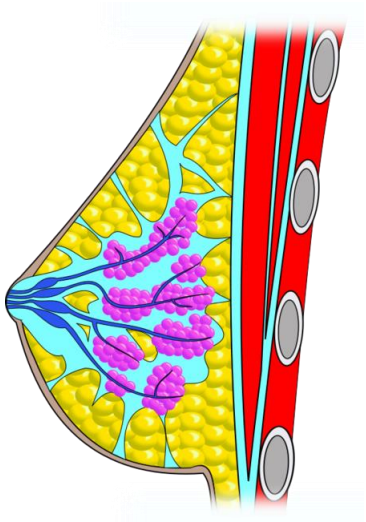


Figure 2 Breast anatomy. Glands (lobules) (purple) consists of several alveoli, which is the site of milk production. The milk travels out into the ductal system (blue), which leads the milk to the nipples. (from Wikimedia commons)

Breast cancer can broadly be categorized into in situ carcinoma (a group of abnormal cells, neoplasm) and invasive carcinoma (cancers that have spread to the surrounding breast tissue) (Figure 3) [12]. In situ carcinoma is further classified based on the location in the glands/lobules (lobular) or in the ductal system (ductal) (Figure 2), while invasive carcinoma is classified into seven subgroups: tubular, ductal lobular, invasive lobular, infiltrating ductal, mucinous, medullary and infiltrating ductal (the latter accounting for 70-80% of all invasive carcinomas) [12].

Breast cancer is also classified based on the receptors that are expressed on the tumor cells' surface (Figure 3), including estrogen receptor (ER), progesterone receptor (PR) and human epithelial growth factor receptor 2 (HER2) [13]. Based on the receptor status, there are five breast cancer subtypes: the Luminal A, the Luminal B, the Triple-negative/basal-like (TNBC), the HER2-enriched, and the Normal-like (which might be contamination of samples with normal mammary cells instead of a different subtype). Luminal A and B are both ER positive (ER+) and/or PR positive (PR+), while luminal A is HER2 negative (HER2-) and luminal B is HER2 positive (HER2+) [13]. The HER2-enriched subtype is ER negative (ER-), PR negative (PR-) and HER2+, while TNBC is negative for all three receptors, as the name suggests [13].

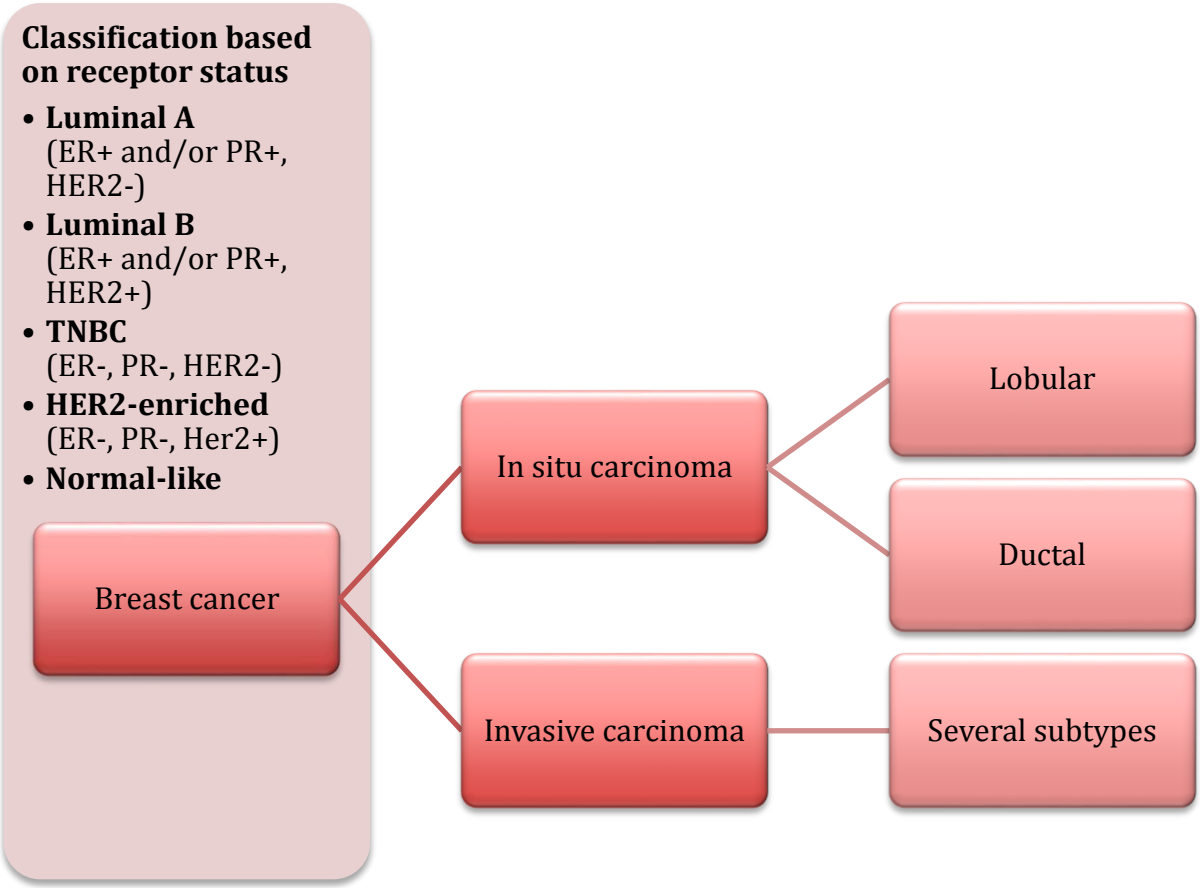


Figure 3 Different subtypes of breast cancer. Breast cancer is categorized based on histological features and molecular subtype (receptor status). ER= estrogen receptor; PR= progesterone receptor; HER2= Human epithelial growth factor receptor 2; "+"= positive; "-"= negative.

1.1.2.2. The female reproductive system and ovarian cancer subtypes

The female reproductive system (Figure 4) consists of two ovaries, two fallopian tubes, the uterus, the cervix and the vagina. Approximately once a month, an egg (also called the ova or oocyte) is released from one of the ovaries in a process is called ovulation. This released egg is caught by the nearby fallopian tube. If the egg is fertilized by a sperm cell while still localized in the fallopian tube, the egg will later attach to the uterus wall. However, if the egg is not fertilized, the egg will leave the body through menstruation.

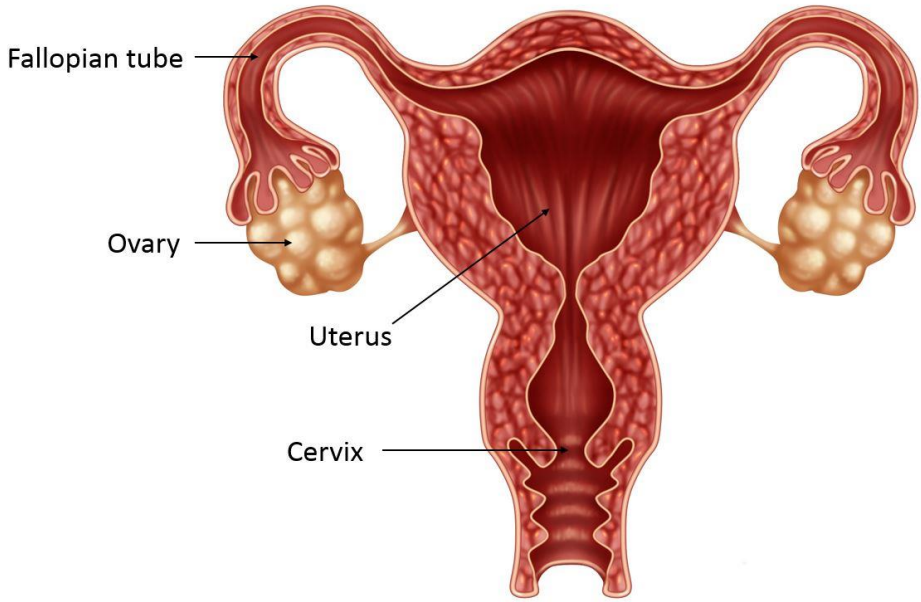


Figure 4 the female reproductive system. Two ovaries are connected to the uterus via the fallopian tubes. The ovaries are the site of egg maturation. Fallopian tubes are responsible for leading eggs to the uterus. If the egg is fertilized during this passage, it may attach to the uterus wall and develop into a fetus. Modified from: www.colourbox.com.

Ovarian cancer was originally thought to develop from cells of ovarian origin. However, the origin of ovarian cancer has later been debated and some of the "ovarian" cancers may have fallopian origin [14]. Ovarian cancer can be categorized based on its cell type origin: epithelial cells, germ cells or stromal cells [15]. In over 90% of cases, the cancers are of epithelial cell origin and are accordingly named epithelial ovarian cancer (EOC) [16]. Epithelial ovarian cancer comprises several histological subtypes and grades, including serous (30-70%), endometrioid (10-20%), mucinous (5-20%), clear cell (3-10%) and undifferentiated (1%) [16]. Ovarian serous carcinoma (OSC), the most common EOC, is further classified in low grade or high grade, where low-grade OSC shows fewer molecular abnormalities and has minimal nuclear atypia [16]. Mutational analysis of the two OSC groups indicates that low-grade OSC develops through a dysregulation of the RAS type GTPase family (RAS) - Raf proto-oncogene, serine/threonine kinase (RAF) signaling pathway. While high-grade OSC probably arises from mutations in *BRCA1, DNA repair associated* (*BRCA1*) or *BRCA2, DNA repair associated* (*BRCA2*), together with mutations in *Tumor protein 53* (*TP53*) [16]. High-grade OSC comprises 50-60% of all EOC [17].

Germ cell tumors originate from cells destined to form eggs. Stromal tumors originate from connective tissue cells and cells that produce the female hormones, estrogen and progesterone [15]. Together, these two subtypes comprise less than 10% of ovarian cancers and will therefore not be explained in more detail.

All ovarian cancers can also be classified based on their invasion of other organs of the body, as stage I-IV (Figure 5). Stage I are cancers confined to the ovary or ovaries (or fallopian tube(s)). Stage II are cancers with metastasis to the pelvis. Stage III are cancers that have metastasized beyond the pelvis. Stage IV are cancers that are widely spread throughout the body [15].

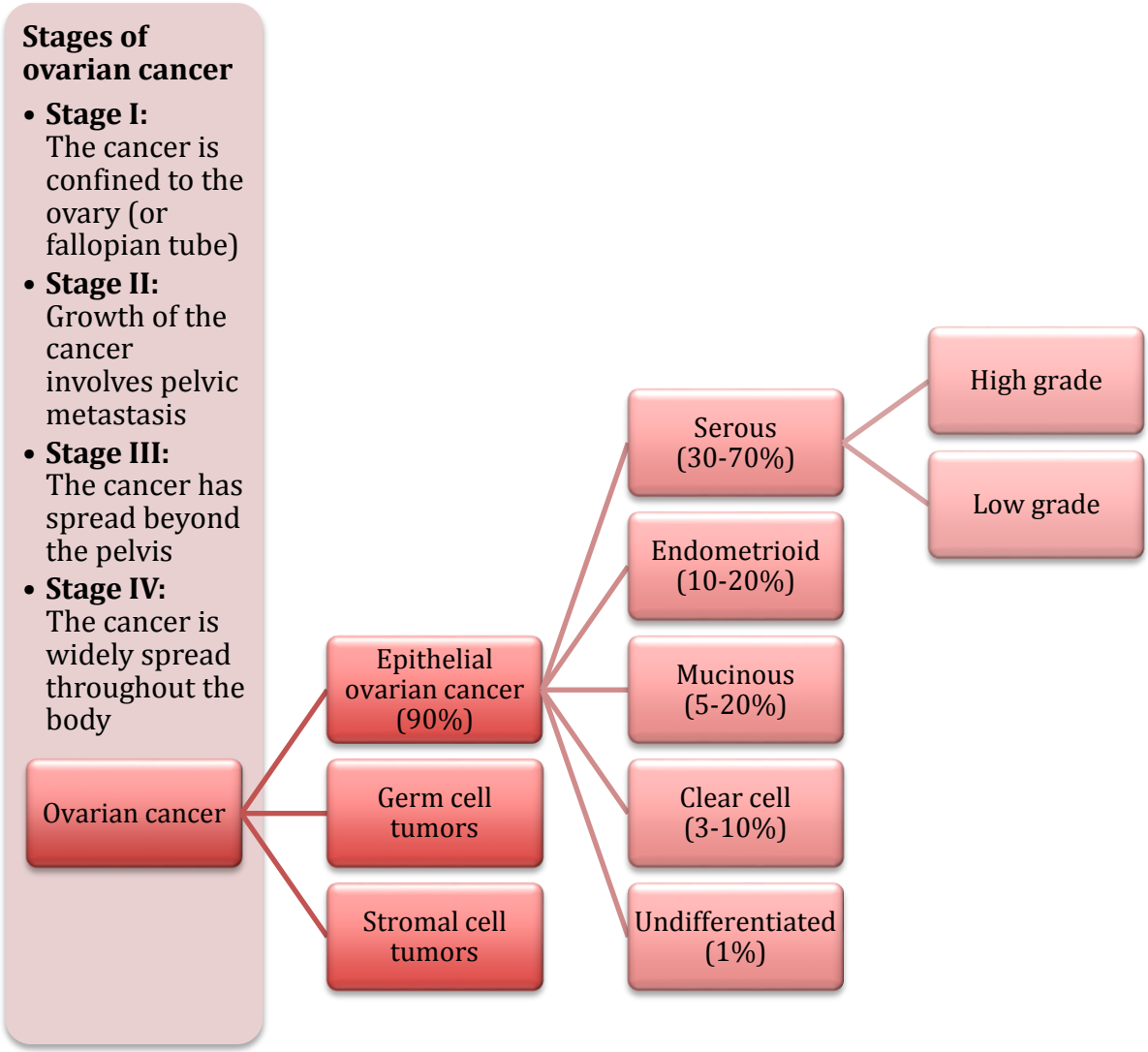


Figure 5 Different subtypes of ovarian cancer. Ovarian cancer is categorized by its cellular origin: epithelia cells, germ cells or stromal cells [15]. The largest group of ovarian cancer originates from epithelial cells and is further classified by the type of epithelial cells [16]. Ovarian cancers are also classified based on stage, related to the invasion of other organs [15].

1.1.3. Risk factors

There are several risk factors associated with both breast and ovarian cancer (BOC). The highest risk factor for BOC is **gender**. Although BOC are primarily female-associated cancers, there are some cases of breast cancer in men (31 in Norway in 2016 [9]).

Age is probably the best-known risk factor, but cancer is more common in older people and accordingly not specific for breast or ovarian cancer. In Norway, the majority of cancers are diagnosed in people above 50 years, 90% and 85% in men and women, respectively [9]. In addition, nearly half of all cancer cases are diagnosed after 70 years of age (48.4% of men, 45% of women [9]). **Other general risk factors** include overweight/lack of exercise, alcohol and smoking. The latter is known to drastically increase lung cancer risk.

More specific risk factors for BOC development are absence of **pregnancies, breast-feeding history and use of oral contraceptives**. Kent Athol published in 2012 a review entitled *Nuns and contraceptives* [18]. This article debates the higher BOC risk seen in nuns, probably due to the combination of lack of pregnancies and breast-feeding and lack of oral contraceptive use,

compared with the general female population. However, the exact mechanism is currently unknown.

Menstrual history may also play a role, where menarche before the age of 12 and menopause after the age of 55 were shown to increase the risk of breast cancer [19]. In addition, **previous personal history of breast cancer**, **family history of cancers** and **genetics** are additional cancer risk components (see below).

1.2. Hereditary breast and ovarian cancer

It is estimated that 5-10% of breast cancers and approximately 25% of ovarian cancers are due to inherited germline mutations [20, 21]. In the mid 90's, two breast and ovarian cancer associated genes were discovered, *BRCA1* and *BRCA2* [22, 23]. Together, pathogenic mutations in these genes are responsible for approximately 20-25% of hereditary breast and ovarian cancer (HBOC) cases [24, 25]. The lifetime risk of BOC for females carrying a pathogenic *BRCA1* mutation is calculated to be 57-65% for breast cancer and 39-59% for ovarian cancer, whereas a pathogenic *BRCA2* mutation confers a 45-55% lifetime risk of developing breast cancer and 11-18% lifetime risk of developing ovarian cancer [25, 26].

1.2.1. Genetic testing

In Norway, genetic testing for sequence variants in *BRCA1/BRCA2* that may increase BOC risk is offered to patients with a family history of BOC, to patients with ovarian cancer and to patients with incidental breast cancer that fulfill some extra criteria; The following criteria for genetic testing is practiced in Norway [27, 28]:

- Woman with breast cancer <50y
- Two close relatives* with breast cancer, mean diagnostic age <55y
- Three close relatives* with breast cancer, regardless of age
- Man with breast cancer
- Woman with bilateral breast cancer <60y
- Woman with breast cancer and a close relative* with ovarian cancer
- Woman with breast cancer and a close relative* with prostate cancer <55y
- Woman with ovarian cancer, regardless of age
- Woman with triple-negative breast cancer <60y

*Close relatives meaning first grade relatives, or second grade relatives if cancer risk is inherited via a man.

Testing for sequence variants in other genes requires individual evaluation after genetic counselling.

The variants identified during genetic testing are evaluated based on allele frequency in population databases such as the Exome Aggregation Consortium (ExAC)/the Genome Aggregation Database (gnomAD) [29] and the Database of Single Nucleotide Polymorphisms (dbSNP) [30]. In addition, the evaluation is based on information from the ClinVar database [31] (a public archive of classified variants, associated phenotypes, information about the submitter and other supporting data) and information from the Human Gene Mutation Database Professional (HGMDp) [32] (a repository of inherited mutation data). Furthermore, variants are evaluated based on missense/splice prediction programs, conservation of the nucleotide/amino acid, the variant's localization in the gene, experimental testing and reports

in the literature. After evaluation of the sequence variants, they are classified according to a 5-tier system: class 1/benign variants, class 2/likely benign variants, class 3/variants of unknown clinical significance (VUS), class 4/likely pathogenic variants and class 5/pathogenic variants. Variants with evidence for or against pathogenicity are classified as class 4/likely pathogenic or class 2/likely benign, respectively. If variants are well documented in databases and/or the literature, they can be further classified as class 5/pathogenic or class 1/benign, respectively.

1.2.2. Technology and additional genes

Previously, genetic testing for pathogenic sequence variants in *BRCA1/*2 was performed by Sanger sequencing, which was long regarded as the gold standard of genetic testing. However, Sanger sequencing required multiple sequencing reactions for each gene (each sequencing reaction typically covered a single exon). Fortunately, new technology has revolutionized the genetics field, by facilitating massive parallel sequencing. This new technology allows for sequencing of multiple genes from multiple patients at the same time, generating a massive amount of sequencing reads. This massive parallel sequencing, termed Next-Generation Sequencing (NGS), has become the new norm in genetic screening. Accordingly, numerous studies have investigated several genes in BOC cohorts [33-39]. Buys et al. (2017) [39] for example, investigated over 35,000 women with breast cancer with a 25-gene panel, where pathogenic variants were identified in 9.3% of the patients. While almost 50% of these pathogenic variants were identified in *BRCA1* and *BRCA2*, the rest were identified in other genes, indicating the need for panel testing of breast cancer patients. *Checkpoint kinase 2* (*CHEK2*), *ATM serine/threonine kinase* (*ATM*) and *Partner and localizer of BRCA2* (*PALB2*) were the most reoccurring genes with pathogenic variants [39]. These findings demonstrated, together with several other studies, that pathogenic variants in genes coding for proteins in the homologous recombination repair (HRR) (other than *BRCA1/2*) are commonly identified in HBOC patients [21, 33-36, 40, 41]. In addition, several other genes have been associated with HBOC (Table 1).

*Table 1 Hereditary breast and ovarian cancer risk genes. Exclusively genes included in the Illumina Cancer panel and with increased BC/OC risk are included in the table. BC= Breast cancer; OC= ovarian cancer; Gyn.= gynecological cancer. Modified from Katsuki and Takata (2016) [42]. * = lifetime risk calculated age varied between studies. **cumulative risk by age 60 years.*

Gene	Syndrome associated	Associated cancer type	Lifetime risk (70-80y*)
BRCA1/ FANCS	HBOC	BC, OC	BC: 57-65%, OC: 39-59% [25, 26]
BRCA2/ FANCD1	HBOC	BC, OC	BC: 45-55%, OC: 11-18% [25]
TP53	Li-Fraumeni syndrome	BC, OC	All: ~100%, BC: 54% [43], OC: no increased risk [44]
PALB2/ FANCN	Fanconi anemia	BC (OC)	BC: 33-58% [45, 46]
ATM	Ataxia telangiectasia	BC	BC: 30% [46]
CHEK2	HBCC	BC	BC: 18.3-31.8% [46]
RAD51C/ FANCO	Fanconi anemia-like	(BC), OC	OC: 6.12% [46]
RAD51D	HBOC	(BC), OC	OC: 10-13.6% [46, 47]
NBN	Nijmegen breakage	BC	BC: 30% [46]
BRIP1/ FANCJ	Fanconi anemia, HBOC	(BC), OC	OC: 4.06-12.7% [46]
MSH2	Lynch syndrome	BC, OC	BC: 11.5-22% [48, 49], OC: 10-24% [49-51]
MSH6	Lynch syndrome	(BC), OC	BC: 12.5%, OC: 1-10.5% [49, 51]
MLH1	Lynch syndrome	BC, OC	BC: 18.6-25% [49, 52], OC: 5-20% [49, 51]
PMS2	Lynch syndrome	OC	OC: <6% [50]
PTEN	Cowden syndrome, PTEN hamartoma	BC	BC: 85.2% [53]
CDH1	Hereditary diffuse gastric cancer	BC, OC	BC: 42% [54]
STK11	Peutz-Jeghers syndrome	BC, OC	ALL: 81%, BC:32%**, Gyn.: 13%** [55]
NF1	Neurofibromatosis type 1	BC	BC: 18% [56]
FANCM	(Fanconi anemia) HBOC	BC	Currently unknown

1.2.3. Clinical management

The Norwegian Directorate of Health has created the following recommendations for clinical management of BOC patients [27, 28]:

Patients diagnosed with breast cancer, without an identified pathogenic sequence variant and without a family history of cancer, follow the general population-screening program: biannual mammography from age 50-69 years. Patients diagnosed with breast cancer and with an estimated increased risk based on family history are offered annual mammography from 30-60 years (with a low threshold for including ultrasound). From 60 years of age, they are included in the normal biannual screening program [27]. However, these recommendations are under revision, with a national agreement that annual mammography should start from 40 years of age, unless family members of the patient have been diagnosed at an earlier age. The new recommendations are awaiting approval from the Norwegian Directorate of Health.

Patients diagnosed with ovarian cancer, without an identified pathogenic sequence variant and regardless of family history are offered ovarian cancer screening. This screening includes annual gynecological examination with vaginal ultrasound (10 years before youngest case of ovarian cancer in the family), measurement of serum biomarkers (CA125 and optionally HE4)

every 6 months and mammography/MR from 25-30 years of age (during first trimester of pregnancy and during breast-feeding, the mammography is replaced with ultrasound) [28].

If a high penetrant pathogenic variant is detected in a patient diagnosed with breast or ovarian cancer, the patient is offered annual magnetic resonance (MR) screening. The screening is offered from the age of 25 if the pathogenic variant is identified in the high-risk cancer associated *BRCA1* or *BRCA2* genes. If the pathogenic variant is detected in *TP53* or *Phosphatase and tensin homolog* (*PTEN*), however, the screening starts at age 20. This annual MR screening is offered until 75 years of age, if mastectomy is not performed. From 75-80 years of age either annual mammography or annual MR is offered [27]. Additionally, pathogenic *BRCA1* variant carriers are offered bilateral salpingo-oophorectomy and this should be considered from 40 years of age. The same applies for carriers of pathogenic *BRCA2* variants; however, bilateral salpingo-oophorectomy can be delayed to 45 years of age if mastectomy has been performed.

These recommendations are only implicated for patients carrying pathogenic variants in *BRCA1*, *BRCA2*, *TP53* and *PTEN*. Although *PALB2* is not currently included in the recommendations from the Norwegian Directorate of Health, patients with identified pathogenic sequence variants in *PALB2* are offered the same management as patients with pathogenic variants in *BRCA1/2* (except ovarian cancer screening and salpingo-oophorectomy). Clinical management programs for patients with breast and/or ovarian cancer and an identified pathogenic variant in another gene (Table 1) are currently unavailable. An exception exists for pathogenic variants identified in *MutS homolog 2* (*MSH2*), *MutS homolog 6* (*MSH6*), *MutL homolog 1* (*MLH1*) and *PMS1 homolog 2, mismatch repair system component* (*PMS2*), where there is a clinical management program from the Norwegian Directorate of Health. However, this clinical management program is management of colorectal cancer risk [57]. The lack of appropriate clinical management programs for carriers of variants not identified in *BRCA1*, *BRCA2*, *TP53* and *PTEN* are mostly due to the currently restricted knowledge of risk estimates for such variants.

1.3. Cell cycle

Cell proliferation, the cellular growth and division, is regulated through the cell cycle (Figure 6). The cell cycle consists of four major phases; the gap 1 phase (G1), the synthesis phase (S), the gap 2 phase (G2) and the mitotic and cytokinesis phase (M). The M phase itself consists of six distinct phases, the prophase, prometaphase, metaphase, anaphase (separation of sister chromatids), telophase and cytokinesis (division of cytoplasm) [7]. During G1 and G2, the cell grows and duplicates its cellular contents, during the S phase the DNA is replicated and during the M phase the division of the duplicated chromosomes and the cytoplasm occurs. Cells which are not actively dividing reside in the G0 phase, awaiting signals to re-enter the cell cycle [4].

The cell cycle is highly regulated through checkpoints (Figure 6). Progression through checkpoints are regulated by different active cyclins and cyclin dependent kinases (CDK) (cyclin-CDK). The first checkpoint, positively regulated by cyclin E and CDK2, is called the restriction point or R-point. After passage through the R-point, the cell is no longer dependent on external cues for continuing the cell division. The R-point is located before entrance into the S phase and halts the cell cycle progression if the genome is damaged. The second checkpoint is positively regulated by cyclin A and CDK2 and is located in the S-phase, where the replication of DNA is blocked if the genome is damaged. A third checkpoint is positively regulated by cyclin B and CDK1 and is located late in the G2 phase. The G2 checkpoint blocks entrance into the M phase if DNA replication is not completed or if the DNA is damaged.

Another checkpoint is located in the M phase. The M phase checkpoint blocks the transition from the metaphase to the anaphase if chromosomes are not properly attached to the mitotic spindle, by blocking the action of the anaphase promoting complex/cyclosome (APC/C) and cell division cycle 20 (CDC20) [4].

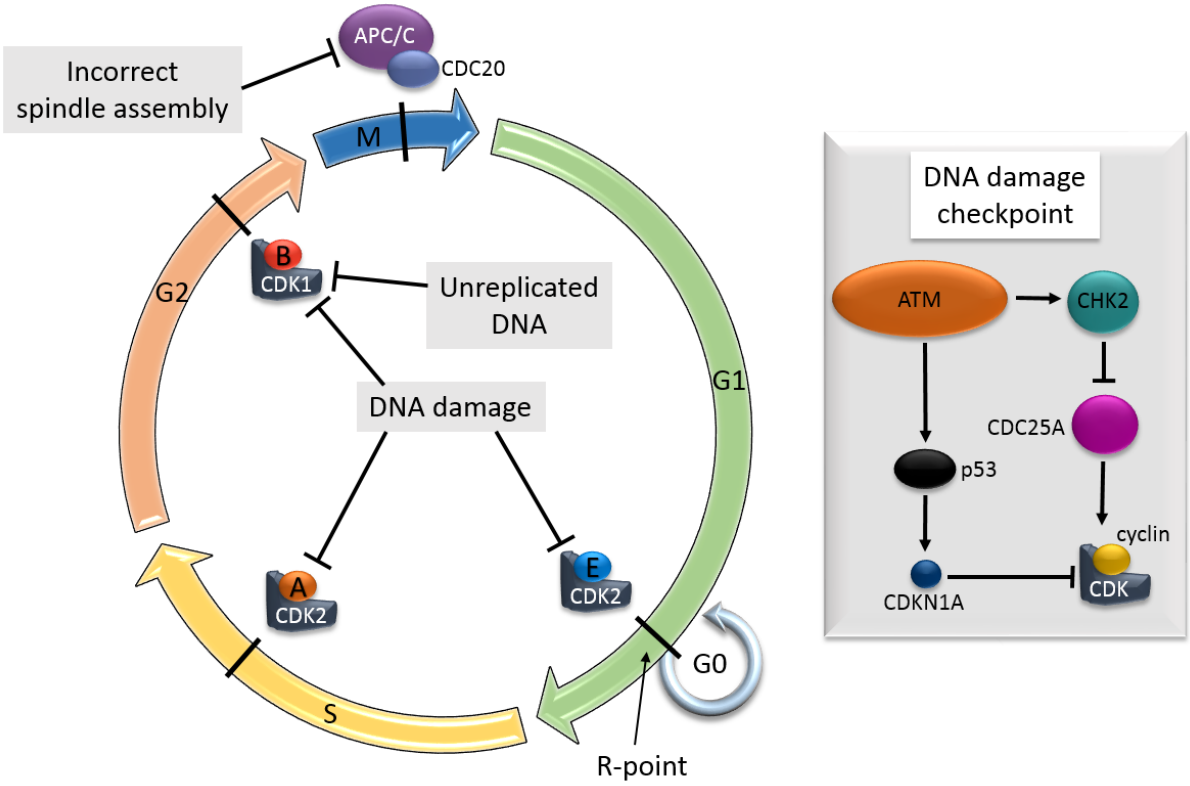


Figure 6. The cell cycle. The cell cycle consists of a gap 1 phase (G1), a synthesis phase (S), a gap 2 phase (G2) and a mitosis and cytokinesis phase (M). The M phase comprises an additional six phases. The cells resting phase is termed G0. The cells' checkpoints are marked with a black line. Four checkpoints are marked. Three of them (the R-point in late G1, the checkpoint in late S and in late G2) are DNA damage checkpoints. These checkpoints are controlled by the proteins displayed in the separate rectangle to the right. Double stranded breaks of the DNA results in activation of ATM, and subsequent activation of CHK2 and p53. CHK2 inhibits the CDC25A phosphatase action, which normally activates the corresponding cyclin-CDK complex. The p53 protein is a positive transcriptional regulator of CDKN1A, which in turn will inhibit the action of the cyclin-CDK complex. Active complexes of cyclin-CDKs promote the passage of the corresponding checkpoints and are negatively regulated during signaling of DNA damage. In addition, the checkpoint in late G2 is also a checkpoint for completed DNA replication. The checkpoint in the M phase is a checkpoint for correct spindle assembly, for transition from the metaphase to the anaphase. In anaphase, the duplicated sister chromatids are separated. Incorrect spindle assembly inhibits the action of the CDC20-APC/C complex.

Several of the genes investigated in the current study have protein products that are associated with the cell cycle regulation, either directly or indirectly through DNA damage repair. One of the DNA damages that can occur are double stranded DNA breaks (DSB), which is the most cellular lethal DNA damage if not properly repaired. The repair mechanism of DSB is discussed in section **1.3.2. DNA repair by Homologous recombination repair**. Double stranded breaks are recognized by the MRE11-RAD50-NBN (MRN) complex (Section **1.3.1.1**), which in turn recruits and activates the ATM protein (Figure 7) [25]. The ATM protein subsequently activates cell cycle arrest, either through Checkpoint kinase 2 (CHK2) or p53, which indirectly negatively regulate the cyclin-CDK complexes and consequently arrest the cell cycle (Figure 6 and 7) [58].

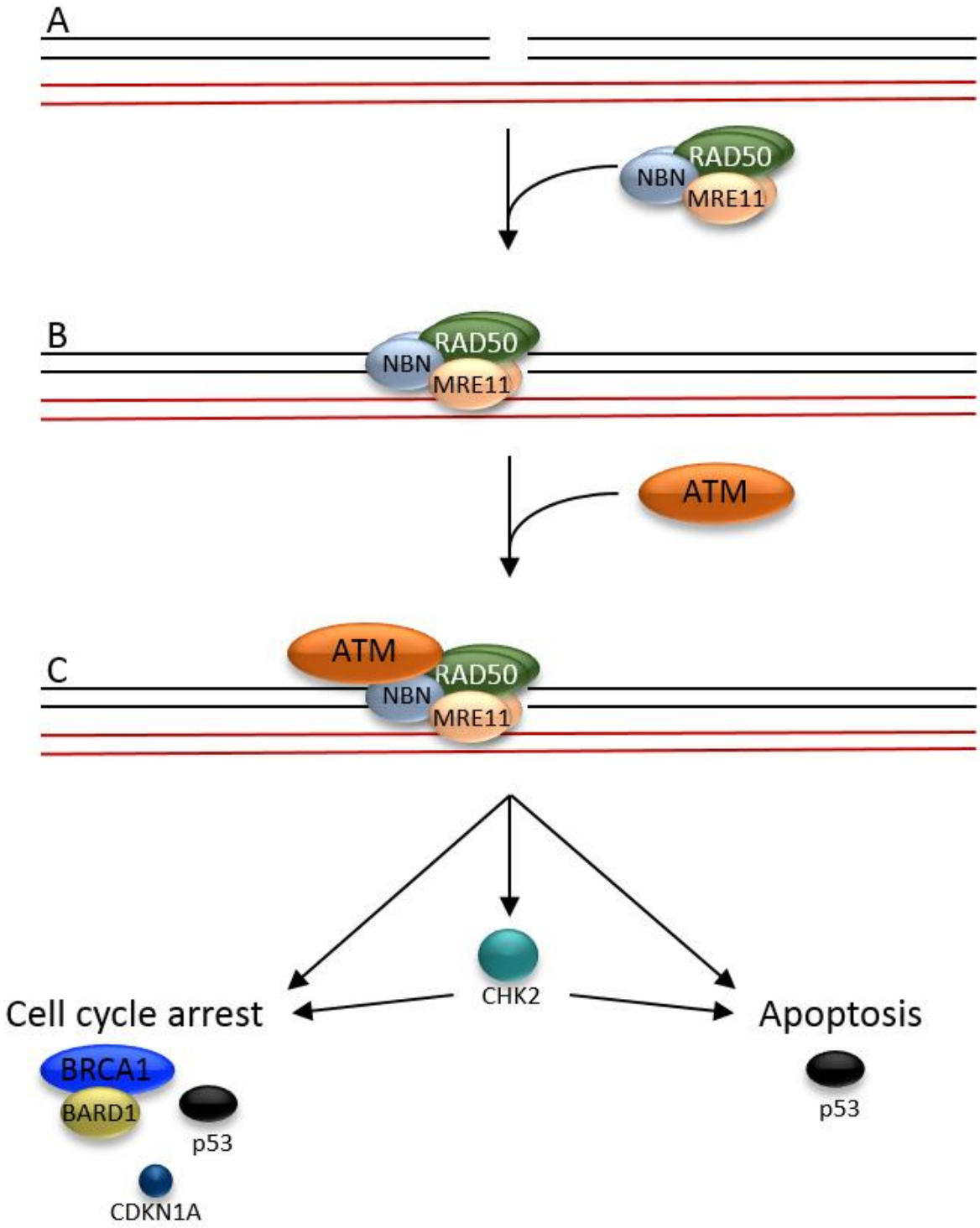


Figure 7 Response pathway to double stranded DNA breaks (DSB). The red and black strands represent the double stranded DNA of sister chromatids, which are present in the late S-phase and during the G2-phase. The heterohexameric MRN complex, consisting of MRE11, RAD50 and NBN, binds double stranded breaks and subsequently recruits ATM to the break site. ATM phosphorylates CHK2, and both ATM and CHK2 phosphorylates several targets, either activating apoptosis through p53 or activating cell cycle arrest through phosphorylation of BRCA1 and p53, amongst others. Cell cycle arrest through p53 phosphorylation leads to activation of CDKN1A, an inhibitor of cyclin dependent kinase 2 (CDK2) and CDK4.

1.3.1.1. The MRN complex – *NBN/NBS1*

The MRN complex is a heterohexamer, which consists of homodimer subunits of MRE11 homolog, double strand break repair nuclease (MRE11) and RAD50 double strand break repair protein (RAD50), together with two Nibrin (NBN) proteins [59]. This complex recognizes DSB (Figure 7) and recruits ATM to the breakpoint site. Additionally, the MRN complex recognizes stalled replication forks and is involved in telomere maintenance [59]. In the current study, applied NGS technology did not include probes to investigate the *MRE11* and *RAD50* genes. Accordingly, these genes and their gene products will not be further discussed.

The *NBN* (alias: *NBS1*) gene consists of 16 exons and is located on chromosome 8q21.3. The transcript (NM_002485.4) encodes an 85 kDa protein called Nibrin, consisting of 754 aa (Figure 8) [60]. At the N-terminal, Nibrin has a phosphoprotein-binding core, containing a Forkhead-associated (FHA) domain (aa 24-83), which is a phosphopeptide recognition motif recognizing phosphothreonine epitopes, a BRCA1 C-Terminal (BRCT)1 domain (aa 105-181) and a BRCT2 domain (aa 216-325) [61-63]. At the C-terminal end, it has an MRE11 and an ATM interacting motif (aa 683-746) [59, 62, 64]. Nibrin is responsible for the interaction and activation of ATM when DSBs occur. Nibrin is also responsible for the interaction and activation of ATR serine/threonine kinase (ATR) in response to replication fork stalling [59]. Nibrin is post-translationally modified by phosphorylation of nine serine and two threonine residues (Ser278, Thr337, Ser343, Ser347, Ser397, Thr402, Ser432, Ser509, Ser518, Ser615 and Ser673). Lastly, Nibrin also gets SUMOylated by cross-linking with Small ubiquitin-like modifier 2 (SUMO2) through Lys529, Lys571 and Lys582 [61].

Biallelic pathogenic variants in *NBN* lead to Nijmegen breakage syndrome, characterized by microcephaly, growth retardation in the uterus, short stature, recurrent nasal/sinus infections and increased risk of cancer [65]. Heterozygous pathogenic variants in *NBN* have previously been associated with breast cancer, with a moderate 3-fold increase/30% lifetime BC risk, where the estimated 30% risk is calculated from one single truncating pathogenic variant [25, 46]. The OC risk of pathogenic variants is, however, low or non-existent [25, 66]. In ClinVar, approximately 130 different germline variants are reported as likely pathogenic or pathogenic and the majority are reported to be associated with hereditary cancer-predisposing syndrome. Almost all of these variants are truncating variants [67]. Although ClinVar is a substantial sequence variant database, not all diagnostic laboratories have reported their findings there.

1.3.1.2. *ATM*

The ATM protein plays a crucial role in cell cycle regulation, through checkpoint regulation at G1, S and G2 phases (Figure 6). ATM is recruited to DSBs by the MRN complex (Figure 7). The protein subsequently phosphorylates downstream targets such as CHK2, p53, BRCA1 and PALB2 (Figure 6 and 7) [25].

The ATM protein (Figure 8) is encoded by the *ATM* gene, consisting of 63 exons and located on chromosome 11q22.3. The gene produces several transcripts. The most common transcript (NM_000051.3) encodes a 351 kDa protein, consisting of 3,056 aa [60, 68]. At the N-terminal end, ATM has a chromatin binding and substrate binding region (aa 5-224), followed by a nuclear localization signal (NLS) (aa 385-388) [69]. At the C-terminal end, ATM has a focal adhesion targeting (FAT) domain (aa 1960-2566), a Phosphatidylinositol 3-kinase (PI3K)-like serine/threonine kinase domain (aa 2712-2962), and a FAT C-terminal (FATC) domain (aa 3024-3056) [61, 70]. Additionally, ATM has several Nibrin-interaction regions (aa 248-522, aa 1436-1770, aa 1764-2138 and aa 2139-2427) [71]. ATM is inactive when dimerized and active in monomeric form, except when responding to oxidative stress, where ATM forms an active disulfide-linked dimer [70].

Biallelic pathogenic sequence variants in *ATM* are the cause of the recessive disorder Ataxia-Telangiectasia (A-T). Heterozygous pathogenic sequence variants in *ATM* have, since 1996, been linked to an increased risk of developing breast cancer [72]. The cumulative lifetime risk for breast cancer is estimated to 30%, whereas the risk of developing ovarian cancer is currently unknown (Table 1) [46]. The type of mutations (Figure 1) associated with increased breast cancer risk has been questioned. Goldgar et al. (2011) concluded that truncating ATM mutations are associated with a significant increased risk of breast cancer, with a penetrance similar to that of *BRCA2* [73]. However, most pathogenic *ATM* mutations confer a moderate, approximately 2-fold increased risk, according to Hollestelle et al. (2010) [74]. In addition, Tavtigian et al. (2009) found only marginal evidence for an increased breast cancer risk for *ATM* truncating sequence variants. According to Tavtigian et al., there is stronger evidence that a subset of rare missense substitutions confers an increased breast cancer risk [75]. In ClinVar, approximately 830 different germline variants are reported as likely pathogenic or pathogenic, most of them in association with A-T. Circa (ca.) 65% are also associated with hereditary cancer-predisposing syndrome, however, only around 2% are further specified to be associated with breast cancer, where most are truncating variants [67].

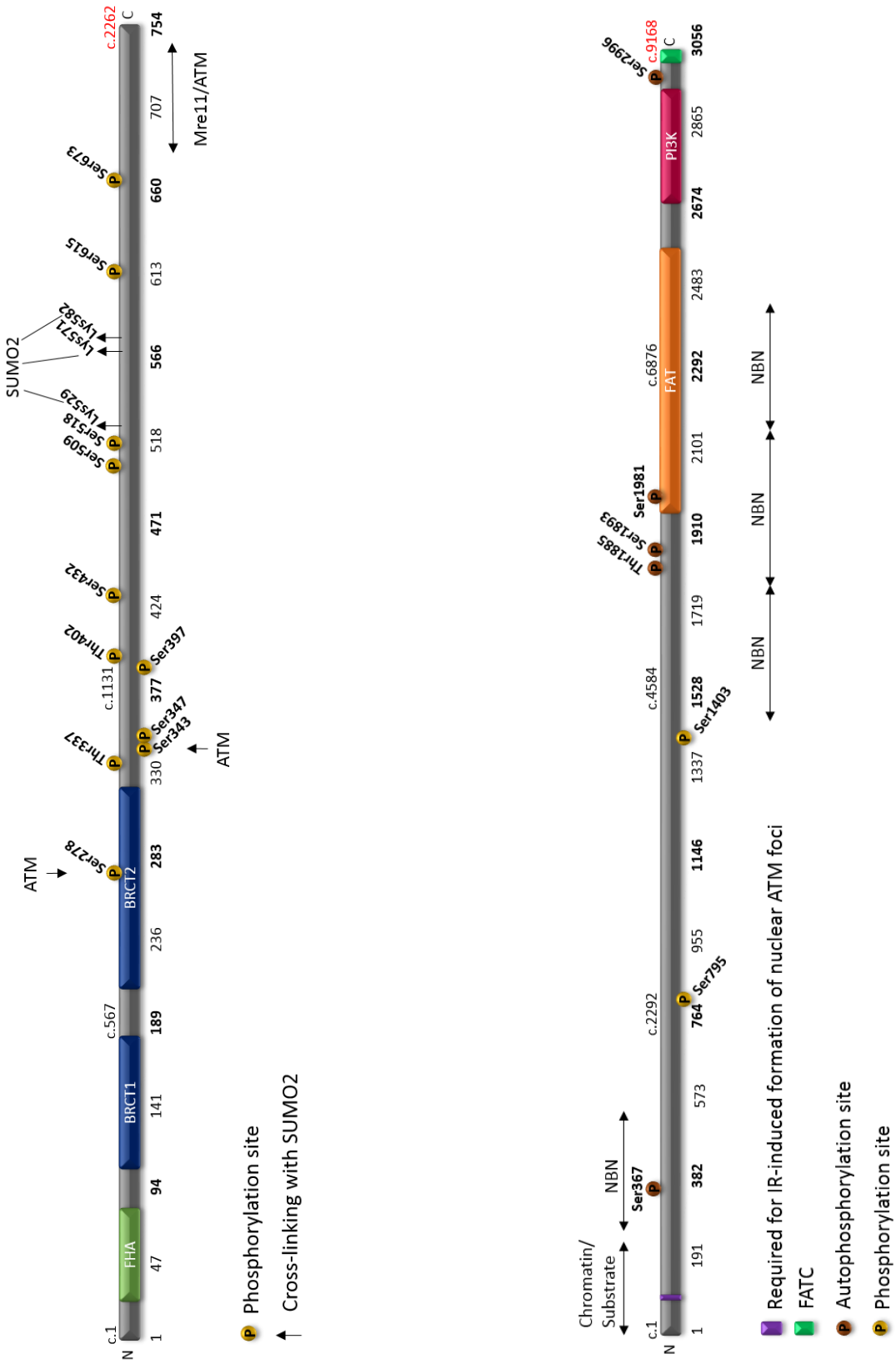


Figure 8 Domain structure of NBN and ATM. The amino acid numbers are listed underneath, while the corresponding position in the gene is listed above. ***To the left:*** *Domain structure of NBN. The protein contains three main domains; FHA domain, BRCT1 domain and BRCT2 domain. At the C-terminal end, the protein has an Mre11/ATM interaction region. The protein has three SUMOylation sites and several phosphorylation sites.* ***To the right:*** *Domain structure of ATM. The protein contains three main domains; the FAT domain, the PI3K domain and a FATC. N-terminally, the protein interacts with chromatin or substrates of the protein. The protein has several NBN-interaction regions and a region required for irradiation (IR)-induced formation of nuclear ATM foci. Additionally, ATM has several phosphorylation sites.*

1.3.1.3. *CHEK2*

The CHK2 protein is a serine/threonine-protein kinase that functions as a signal transducer for ATM. The CHK2 protein is responsible for cell cycle control (in G1, S and G2 phase) through activation of cell cycle arrest, DNA repair and apoptosis (controlled cell death), in response to DSB (Figure 6, 7 and 9).

The CHK2 protein is encoded by the *CHEK2* gene, which consists of 15 exons and is located on chromosome 22q12.1. The transcript (NM_007194) encodes a 61 kDa protein, consisting of 543 aa [60]. The CHK2 protein (Figure 10) contains an N-terminal serine-glutamine/threonine-glutamine (SQ/TQ) cluster domain (aa20-75), which contains seven serine or threonine residues followed by glutamine [76]. CHK2 is activated by phosphorylation of these serine- and threonine residues by ATM. This phosphorylation, especially Thr68, of CHK2 promotes homodimerization and intermolecular phosphorylation of Thr383, Thr387 and Ser516 [76, 77]. The CHK2 protein also contains an FHA domain (aa 113-175) and a protein kinase domain (aa 220-486) with a T-loop/activation segment (aa 368-394), important for signal transduction [61].

CHK2 phosphorylates several target proteins. Some of the target proteins are involved in DNA repair, such as BRCA1 and BRCA2. Other target proteins are required in the cell cycle regulation, such as CDC25A and CDC25C. Additionally, CHK2 phosphorylates proteins involved in p53 signaling and in the apoptosis pathway. CHK2 also phosphorylates proteins, such as the Bloom syndrome RecQ like helicase and CDK11, for which the function still remains to be elucidated [78].

Pathogenic sequence variants in *CHEK2* are linked to moderate breast and prostate cancer risk (Table 1). The lifetime breast cancer risk is estimated to 18.3-31.8%, depending on the type of sequence variants. For example, the missense variant p.Ile157Thr has a low estimated breast cancer risk, whereas the truncating p.1100delC confers a higher estimated breast cancer risk [46]. However, 31.8% may be an overestimation since it is calculated based on a constant relative risk, but the relative risk may decline with age leading to a lifetime risk of 23.4% [46]. In ClinVar, approximately 230 different germline variants are reported as likely pathogenic or pathogenic. Of these, ca. 55% are associated with breast cancer. The breast cancer associated variants include ca. 14% missense variants: ca. 1% are located in the SQ/TQ cluster domain, ca. 3% in the FHA domain, ca. 7% in the protein kinase domain and ca. 3% outside known domains [67].

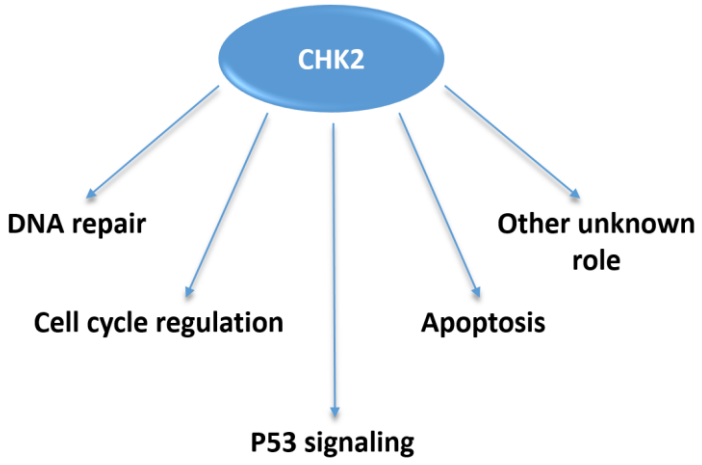


Figure 9 The CHK2 signal transducer.

1.3.1.4. *TP53*

The tumor protein p53 (p53) has long been referred to as the "guardian of the genome" [79]. p53 has a central role in cell cycle control (Figure 6) and it is one of the most commonly mutated genes in cancer cells [80]. The p53 protein is a transcription factor and is active as a homotetramer [81, 82]. Amongst its target genes is the *Cyclin Dependent Kinase inhibitor 1A* (*CDKN1A*) gene. The *CDKN1A* encodes a protein responsible for direct inhibition of cyclin-CDKs, leading to cell cycle arrest (Figure 6) [58].

The p53 protein is encoded by the *TP53* gene, consisting of 11 exons and located on chromosome 17p13.1. The gene transcript (NM_000546.5) encodes a 44 kDa protein, consisting of 393 aa [60]. The p53 protein (Figure 10) has two transactivation domains (TAD), important for its transcription factor activity, each TAD contains 9 aa; TADI at aa 17-25 and TADII at aa 48-56 [61]. The two TADs are followed by a proline-rich region (PRR) (aa64-92), a DNA binding domain (DBD) (aa 94-292) and a tetramerization domain (TET) (aa 325-355) [81-83]. In addition, a bipartite NLS is located at aa 305-321 and a nuclear export signal (NES) is located at aa340-351 [61, 81]. The amino acid Trp91 folds back onto the Arg174 of the DBD [82].

Pathogenic variants in *TP53* are linked to Li-Fraumeni syndrome, which is a cancer predisposition syndrome. Common Li-Fraumeni tumors include soft tissue sarcoma, osteosarcoma, pre-menopausal breast cancer, brain tumors, adrenocortical carcinoma and leukemia [84]. Pathogenic variants in *TP53* confer a lifetime cancer risk of 100% and a breast cancer risk of 54% (Table 1) [43]. According to the National Comprehensive Cancer Network, there is no increased risk of ovarian cancer associated with pathogenic *TP53* variants [44].

Although pathogenic mutations are identified in the entire length of *TP53*, approximately 80% (including somatic mutations) are clustered in exons 5-8, the region encoding the core segment of the p53 protein, the DBD (aa 94-292), and 28% are found in one of the six mutational hotspots in the DBD in *TP53* (Figure 10) [80, 85]. Furthermore, approximately 80% are missense mutations [85]. In ClinVar, approximately 360 different germline variants are reported as likely pathogenic or pathogenic; ca. 25% are reported in association with neoplasm in breast tissue. Of the latter variants, 98% are missense variants located in the DBD [67].

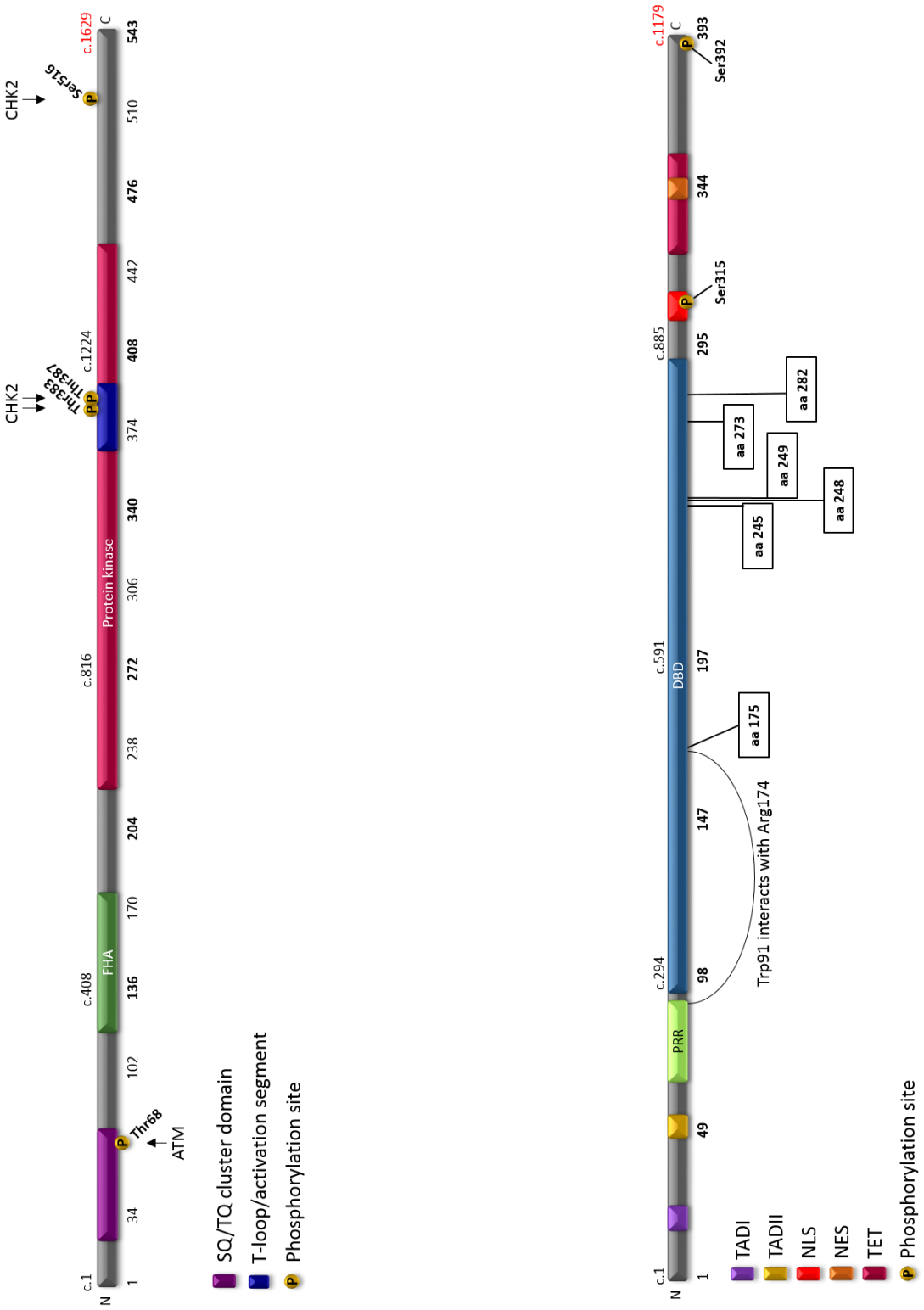


Figure 10 Domain structure of CHK2 and p53. The amino acid numbers are listed underneath, while the corresponding position in the gene is listed above. ***To the left:*** *domain structure of CHK2. The protein contains three main domains; the SQ/TQ cluster domain, the FHA domain and a protein kinase domain, with a T-loop/activation segment. ATM phosphorylates several amino acids of CHK2, especially Thr68. After ATM activation, the CHK2 proteins dimerize and trans-phosphorylate Thr383, Thr387 and Ser516.* ***To the right:*** *domain structure of p53. The protein contains several domains and motifs, starting with the TADI, the TADII, and a PRR. This is followed by the large DBD, which contains the mutational hot-spots marked underneath. Additionally, p53 has a bipartite NLS and a NES. The latter is located within the proteins TET domain.*

1.3.2. DNA repair by Homologous recombination repair

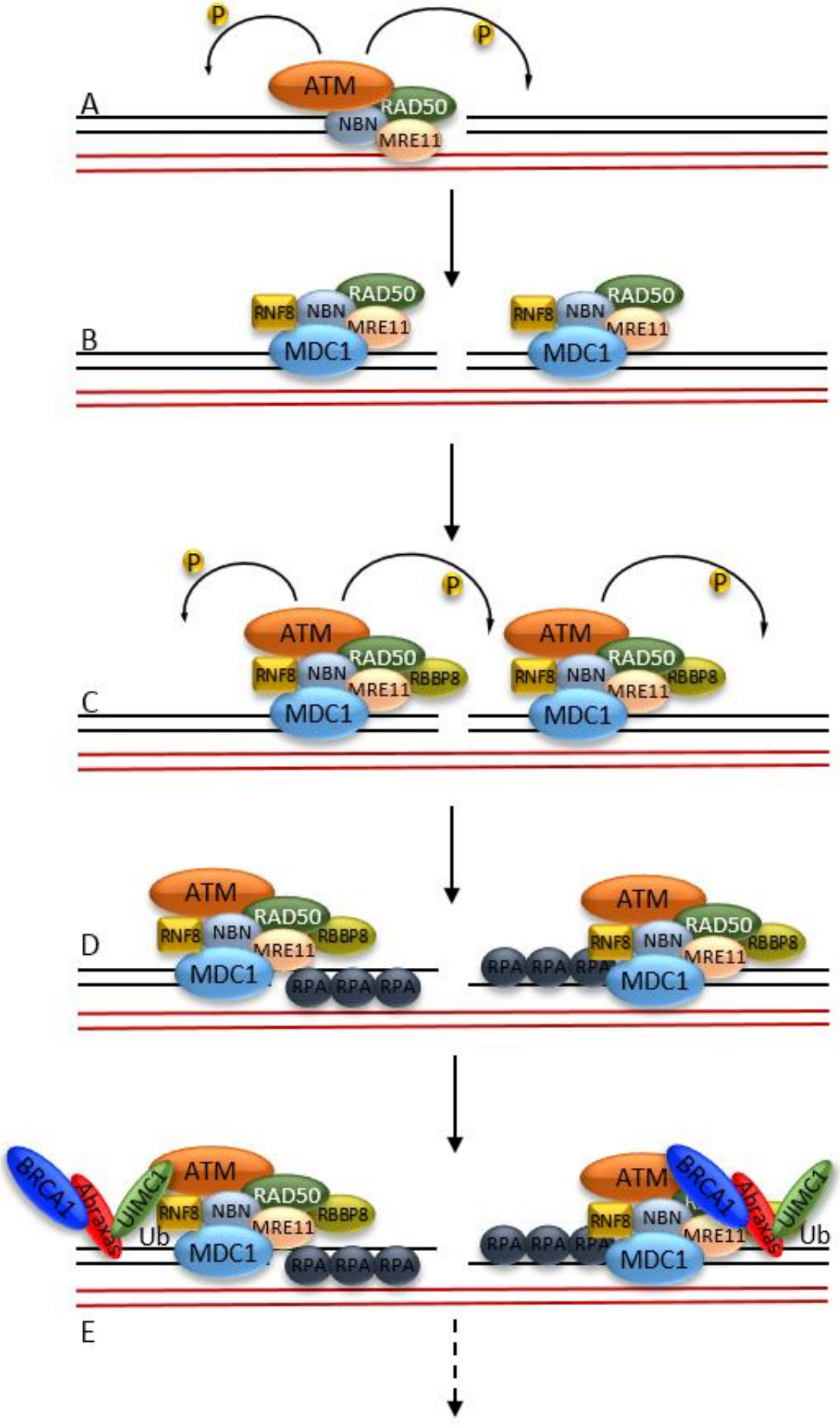
Double stranded DNA breaks (Figure 7) are the most cellular lethal DNA damages and need rapid repair. The DSBs can be resolved by two mechanisms, the non-homologous end joining (NHEJ) pathway and the homologous recombination repair (HRR) pathway. The NHEJ is active in all the phases of the cell cycle. However, it is the most error-prone of the two and may result in the loss of nucleotides when DNA ends are ligated without the use of a homologous template [7]. The HRR, on the other hand, is a conservative repair pathway, which utilizes the sister chromatid as a template, ensuring exact copies of nucleotides at the DNA break site. Consequently, HRR is active during the late S-phase and the G2-phase (Figure 6), when the sister chromatid is present [25].

Homologous recombination repair (Figure 11) engages several proteins, which are recruited in several "waves". The first "wave" consists of recognition of the DSB by the MRN complex (see section **1.3.1.1. The MRN complex – *NBN/NBS1***). The MRN complex rapidly recognizes free DNA ends at the DSB and in turn recruits ATM to the DSB site (Figure 7). The ATM protein marks the DSB by phosphorylating the Histone 2A member X (H2AX) chromatin protein, resulting in the γH2AX (Figure 11A). The γH2AX histones are recognized by the Mediator of DNA-damage checkpoint 1 (MDC1) complex, consisting of the MDC1 protein, the MRN complex and the Ring finger protein 8 (RNF8) (Figure 11B). The MRN recruits the RB binding protein 8, endonuclease (RBBP8; alias CtIP) to the DSB and can additionally recruit more ATM to the DSB and lead to phosphorylation of even more H2AX – constituting the "second wave" of protein recruitment (Figure 11C) [86]. After recruitment of these proteins, single stranded DNA (ssDNA) formation can start. The MRE11 proteins of the MRN complexes are responsible for the ssDNA formation in cooperation with the RBBP8, which in turn promotes loading of the replication protein A (RPA) to the newly formed ssDNA (Figure 11D) [87].

The RNF8 protein in the MDC1 is an E3 ubiquitin ligase, responsible for poly-ubiquitination of H2A histones. This poly-ubiquitination attracts the Ubiquitin interaction motif containing 1 (UIMC1; alias RAP80), in complex with BRCA1 and Abraxas, as a "third wave" of protein recruitment (Figure 11E) [86]. ATM phosphorylates RBBP8 at Ser327, and RBBP8 can then bind BRCA1 (Figure 11F) [88]. A prerequisite for phosphorylation by ATM is phosphorylation of RBBP8 by CDK [89]. The BRCA1-RBBP8 complex subsequently binds one of the many NBNs in MRN complexes [90]. The PALB2 protein binds BRCA1 and recruits BRCA2 with bound RAD51 recombinase (RAD51) proteins (Figure 11G). Subsequently, BRCA2 loads RAD51 onto the ssDNA, displacing RPA (Figure 11H). The RAD51-ssDNA filament is required for strand invasion of the sister chromatid (Figure 11I) and this invasion allows the sister chromatid to function as the polymerase template when starting DNA synthesis to recover the lost nucleotides of the DSB and restore the missing sequence (Figure 11J) [91].

The tumor protein p53 binding protein 1 (TP53BP1) is also recruited to DSB, most likely by binding phosphorylated H2A [86]. The TP53BP1 is involved in the DBS repair pathway choice, promoting the error-prone NHEJ pathway. One of the suggested roles of BRCA1 in HRR is to prevent the binding of TP53BP1 to the DSB [25].

Although HRR is a very accurate repair mechanism important during the cell cycle, there is a disadvantage to the mechanism. If a tumor suppressor gene has a heterozygous pathogenic variant, HRR may lead to loss of heterozygosity (LOH), resulting in the knock-out of both alleles, and subsequently two non-functional proteins or non-existent proteins. The LOH in tumor suppressor genes is an important factor in cancer development [92].


A
P
P
ATM
RAD50
NBN
MRE11
B
RNF8
NBN
RAD50
MRE11
MDC1
C
RBBP8
D
RPA
E
BRCA1
Abraxas
UIMC1
Ub

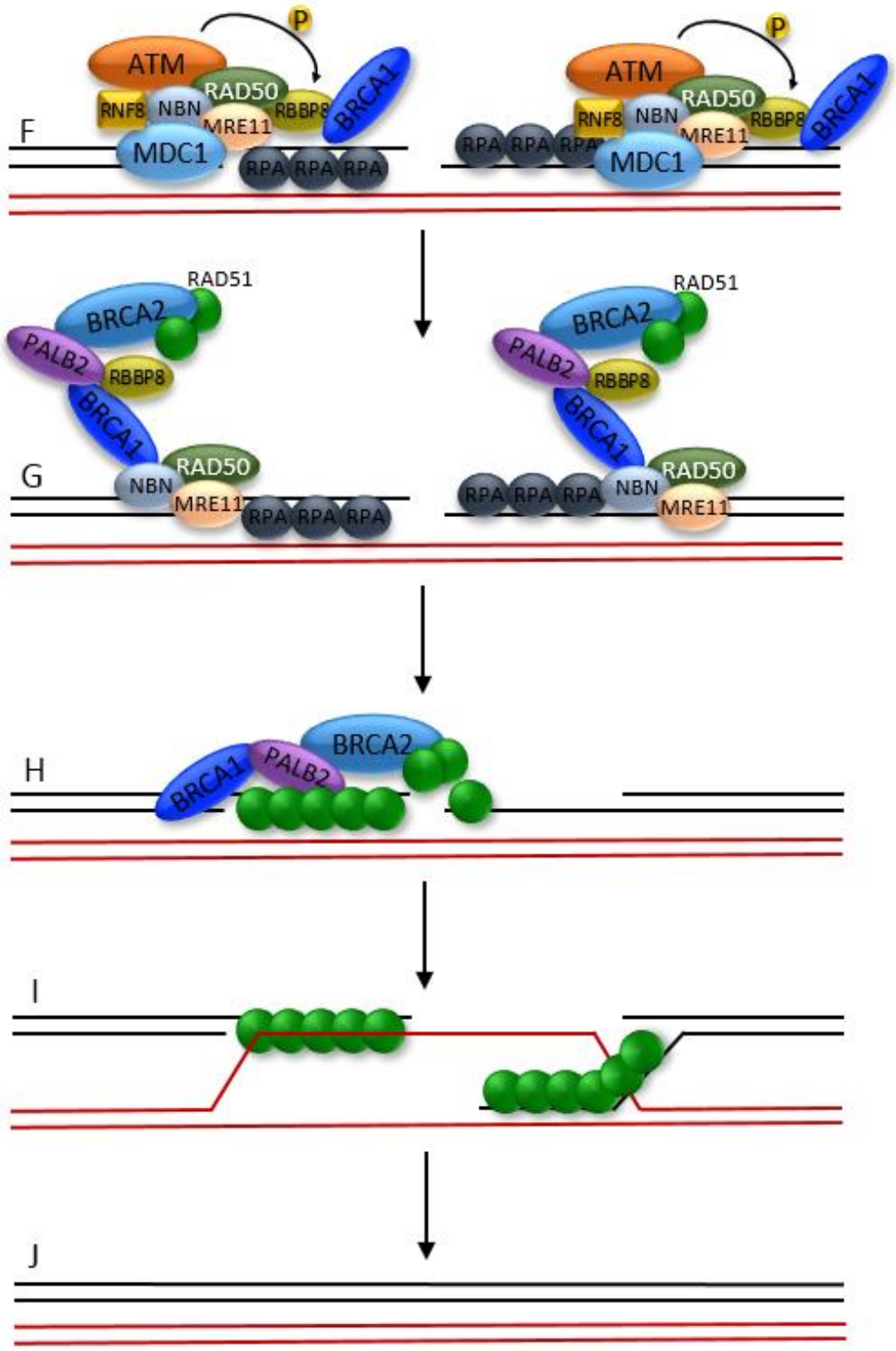


Figure 11 Homologous recombination repair after a DBS. The red strands represent the sister chromatid. A) A DSB is recognized by the MRN complex, ATM is recruited by the MRN complex and subsequently phosphorylates H2AX histones (γH2AX). B) γH2AX recruits the MDC1 complex (MDC1, MRN, and RNF8). C) The MRN complex recruits ATM that phosphorylates even more H2AX. D) MRE11 and RBBP8 produce ssDNA and RPA is loaded to the newly formed ssDNA. E) RNF8 poly-ubiquitinates H2A histones, which recruits UIMC1 in complex with BRCA1 and Abraxas. F) ATM phosphorylates RBBP8, which can then bind BRCA1. G) BRCA1-RBBP8 binds NBN of the MRN complex. BRCA2-RAD51 is recruited through PALB2 binding BRCA1. H) BRCA2 loads RAD51 to the produced ssDNA, displacing RPA. I) The RAD51-ssDNA filaments can perform strand invasion of the sister chromatid. J) The polymerase can use the sister chromatid as template for synthesis and the DSB is restored.

1.3.2.1. *BRCA1/FANCS*

The BRCA1 protein plays a central role in HRR (Figure 11), during the late S-phase and G2-phase (Figure 6). However, BRCA1 has several additional functions. The BRCA1 protein has a really interesting new gene (RING) domain with an E3 ubiquitin ligase activity, which is potentiated by BRCA1-Associated RING Domain-1 (BARD1) binding and not required for HRR. In addition, the C-terminal end encompasses two tandem BRCT domains, responsible for interacting with a number of cell cycle checkpoint proteins and repair proteins [93, 94].

The BRCA1 protein (Figure 12) is encoded by the *BRCA1* (Alias: *FANCS*) gene, consisting of 23 exons and located on chromosome 17q21.31. The main transcript (NM_007294.3) encodes a 208 kDa protein, consisting of 1,863 aa [60]. N-terminally, the BRCA1 protein has a RING domain (aa 8-96) with a zinc finger motif (aa 24-65) [61, 95]. The BRCA1 protein has three NLS signals; two canonical NLS (aa 503-508 and 606-615) and a non-canonical NLS (aa 252-257) [96-98]. The NLS are responsible for locating the protein where it can affect genome maintenance. C-terminally, BRCA1 has a coiled-coil domain (aa 1364-1437), which can interact with PALB2 (aa 1397-1424), and two tandem BRCT domains (aa 1646-1736 and aa 1760-1855) [61, 99]. The two BRCT domains display phospho-specific binding activity, important for mutually exclusive interactions with phosphorylated BRCA1 interacting protein C-terminal helicase 1 (BRIP1) (Ser990), RBBP8 or Abraxas [93]. The interaction between BRCA1 and BRIP1 is essential for the G2- to M-phase checkpoint control [100]. BRCA1 also has a Werner syndrome RecQ like helicase (WRN) binding region (aa 452-1079) [101]. In addition, BRCA1 endures several post-translational modifications, such as acetylation of Met1 and SUMOylation through cross-linking with SUMO2 at several lysine residues (Lys109, Lys301, Lys339, Lys443, Lys459, Lys519, Lys583, Lys654, Lys734, Lys739, Lys918, Lys987 and Lys1079). BRCA1 is also phosphorylated by several kinases, such as CHK2 (Ser988), ATM (Ser1387, Ser1423, Ser1524) and ATR (Ser1143, Ser1280, Ser1387, Thr1394, Ser1423, Ser1457). Additionally, BRCA1 has several other phosphorylation sites: Ser114, Ser308, Ser395, Ser398, Ser423, Ser434, Ser551, Ser694, Ser708, Ser725, Ser753, Ser840, Ser1009, Ser1189, Ser1191, Ser1211, Ser1217, Ser1218, Ser1328, Ser1336, Ser1342 and Ser1542 [61].

Patients with a heterozygous pathogenic variant in *BRCA1* have a highly increased risk of developing breast and ovarian cancer. The lifetime risk for breast cancer is estimated to 57-65%, while the risk of ovarian cancer is 39-59% (Table 1) [25, 26]. However, different variants may confer different cancer risks. For example, the *BRCA1* sequence variant c.5095C>T p.(Arg1699Trp) has been shown to be an intermediate penetrant variant [102]. It has long been assumed that carrying biallelic pathogenic variants in *BRCA1* was lethal [103]. However, in 2012, Domchek et al. reported the first patient with validated biallelic pathogenic variants in *BRCA1* and in 2015, Sawyer et al. identified the first patient with Fanconi anemia subtype S (Fanconi anemia is discussed in section **1.3.3. The Fanconi anemia pathway**) and biallelic pathogenic variants in *BRCA1* [103, 104]. In ClinVar, approximately 2,600 different germline variants are reported as likely pathogenic or pathogenic, ca. 97% are associated with breast and/or ovarian cancer. Most reported variants are truncating variants. Only approximately 4% of these are missense variants [67]. Of these 4%, ca. 30% are located in the RING domain (presumably disturbing the BARD1 interaction) and ca. 50% in the BRCT1 and BRCT2 domains (presumably disturbing the transactivation capability of BRCA1). Another database with *BRCA1* variants is the Breast Cancer Information Core (BIC) [105]. The BIC database is a repository for all sequence variants identified in *BRCA1* and *BRCA2*. In the BIC database, approximately 8,600 (ca. 900 unique) variants are reported classified as variants of clinical importance. Approximately 5% of these variants are missense variants and ca. 90% of these are located in the RING domain and the two BRCT domains.

1.3.2.2. *BRCA2/FANCD1*

Like BRCA1, BRCA2 has a central role in HRR. In HRR, BRCA2 is recruited to the DSB by PALB2 and is responsible for RAD51 loading to the ssDNA (Figure 11G-I).

The BRCA2 protein (Figure 12) is encoded by the *BRCA2* (Alias: *FANCD1*) gene, consisting of 27 exons and located on chromosome 13q13.1. The transcript (NM_000059.3) encodes a 384 kDa protein, consisting of 3,418 aa [60]. N-terminally, the BRCA2 protein has a binding region for the PALB2 protein (aa 10-40) [106], followed by eight BRC repeats; I (aa 1002-1036), II (aa 1212-1246), III (aa 1421-1455), IV (aa 1517-1551), V (aa 1664-1698), VI (aa 1837-1871), VII (aa 1971-2005) and VIII (aa 2051-2085) [61]. Each of the BRC repeats binds one RAD51 molecule by mimicking the oligomerization motifs of RAD51, enabling BRCA2 to form RAD51 filaments at the ssDNA at DSB (Figure 11G) [107]. C-terminally, BRCA2 has a DBD (aa 2460-3170) and a NES (aa 2682-2698) masked by binding to SEM1, 26S proteasome complex subunit (SEM1). BRCA2 also possesses several phosphorylation sites; Ser70, Ser445, Ser492, Ser755, Ser1970, Thr2035, Ser2095, Ser3291 (by CDK1/CDK2, preventing HRR), Ser3319 and Thr3387 (the latter by Checkpoint kinase1/CHK2) [61, 108]. In addition, the protein has two NLS at the C-terminal end, NLS1 (aa 3263-3269) and NLS2 (aa 3381-3385) [109]. However, the exact location of NLS2 is still disputed [110].

Patients with heterozygous pathogenic variants in *BRCA2* have a highly increased risk of breast and ovarian cancer, while patients with biallelic pathogenic *BRCA2* variants suffer from Fanconi anemia, subtype D1. The lifetime risk of heterozygote carriers for developing breast cancer is 45-55% and for developing ovarian cancer is 11-18% (Table 1) [25]. In ClinVar, approximately 3,000 different germline variants are reported as likely pathogenic or pathogenic, where ca. 97% are reported to be associated with breast and/or ovarian cancer. Most reported variants are truncating variants. Only approximately 2.5% of these are missense variants [67]. Of these 2.5%, ca. 60% are located in the DBD. In the BIC database, approximately 4,500 (ca. 850 unique) variants are reported classified as variants of clinical importance. Approximately 1% of these variants are missense variants. However, none of them were reported in the DBD [105].

For both *BRCA1* and *BRCA2*, several naturally occurring alternatively spliced transcripts have been reported (see **Section 1.4.2.1. Naturally occurring alternative splicing in *BRCA1* and *BRCA2***). Additionally, several splice-affecting sequence variants have been reported for both genes [111-115].

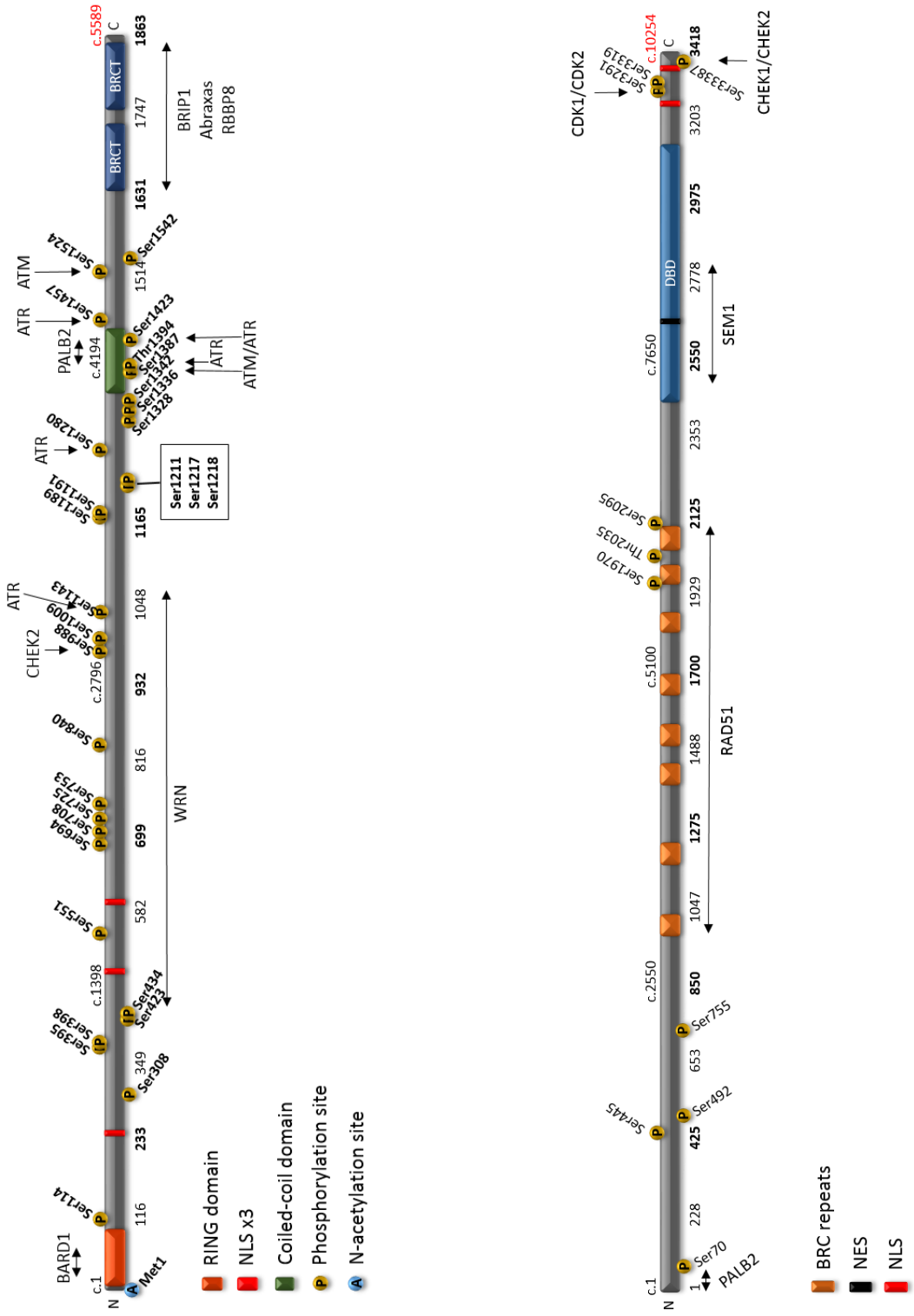


Figure 12 Domain structure of BRCA1 and BRCA2. The amino acid numbers are listed underneath, while the corresponding position in the gene is listed above. ***To the left:*** *Domain structure of BRCA1. The protein contains several domains and motifs: the RING domain, three NLS, a coiled-coil domain and two BRCT domains. N-terminally, the protein interacts with BARD1. BRCA1 also has other interaction partners, such as WRN, CHEK2, ATM, ATR, BRIP1 and PALB2, the latter linking BRCA1 to BRCA2 during HRR. The protein has a SUMOylation site at 13 Lysine residues (not displayed).* ***To the right****: Domain structure of BRCA2. The protein has eight BRC repeats, a DBD, a NES and two NLS. The protein has a SEM1 binding region, which overlaps with the NES, trapping the protein in the nucleus. N-terminally, the protein has a PALB2 interacting region. PALB2 is responsible for linking BRCA1 and BRCA2 during HRR.*

1.3.2.3. *PALB2/FANCN*

The PALB2 protein is essential in linking BRCA1 and BRCA2 together during HRR and stimulating DNA synthesis during strand invasion (Figure 11G-I).

The PALB2 protein (Figure 13) is encoded by the *PALB2* (Alias: *FANCN*) gene, consisting of 13 exons and located on chromosome 16p12.2. The gene transcript (NM_024675.3) encodes a 131 kDa protein, consisting of 1186 aa [60]. PALB2 has a coiled-coil domain (aa 9-41) and seven WD repeats (1: aa 854-915; 2: aa 917-961; 3: aa 962-1009; 4: aa 1010-1052; 5: aa 1058-1109; 6: aa 1115-1153; 7: aa 1155-1186) responsible for binding BRCA2 and RAD51 and recruiting them to the DSBs [61, 116]. PALB2 has a DNA-binding region (aa 1-579) where aa 1-160 is required for its oligomerization and focal concentration at DNA damage sites. The DNA-binding region overlaps with a BRCA1 interaction region (aa 1-319) and a RAD51 interaction region (aa 1-200) [61, 117]. PALB2 also has a chromatin-association motif (ChAM) that mediates nucleosome association (aa 395-446) [118]. In addition, PALB2 has been shown to interact with DNA polymerase eta (POLH) (aa 775-1186) and to stimulate POLH DNA synthesis during HRR [61, 119].

Biallelic pathogenic variants in *PALB2* have been associated with Fanconi anemia, subtype N. Fanconi anemia, subtype N is associated with an unusually severe predisposition to pediatric malignancies [120]. Monoallelic pathogenic variants confer an increased risk of breast cancer, with a lifetime risk of 33-58% [45, 46]. There is conflicting evidence for increased risk of ovarian cancer associated with monoallelic pathogenic variants in *PALB2* [46]. In ClinVar, approximately 400 different germline variants are reported as likely pathogenic or pathogenic. Approximately 90% are reported to be associated with hereditary cancer and ca. 50% of the germline variants are further characterized to be associated with breast cancer. Most of the mutations are truncating variants [67].

1.3.2.4. *RAD51C/FANCO* and *RAD51D*

Five proteins share approximately 25% aa sequence identity with RAD51: RAD51 paralog B (RAD51B), RAD51 paralog C (RAD51C), RAD51 paralog D (RAD51D), X-ray repair cross complementing 2 (XRCC2) and X-ray repair cross complementing 3 (XRCC3). These proteins are involved in two different complexes, the RAD51B-RAD51C-RAD51D-XRCCC2 (BCDX2) and RAD51C-XRCC3 (CX3) [121]. It has been suggested that the BCDX2 complex acts upstream of RAD51 recruitment in HRR, while CX3 acts downstream of the RAD51 recruitment to DSB [122].

The **RAD51C** protein (Figure 13) is encoded by the *RAD51C* (alias: *FANCO*) gene, consisting of 9 exons and located on chromosome 17q22. Its main transcript (NM_058216.1) encodes a 42 kDa protein, consisting of 376 aa [60]. N-terminally, the protein has a region required for Holliday junction resolution (aa1-126) [61]. Holliday junction is branched double stranded DNA consisting of four arms that occurs at the end of the HRR. In addition, RAD51C has a phosphorylation site at Ser20, a nucleotide binding region (aa 125-132) and a region for interaction with RAD51B, RAD51D and XRCC3 (aa 79-136), required for the formation of BCDX2 and CX3 [61, 121]. C-terminally, the protein has a NLS (aa 366-370) [61].

Biallelic pathogenic sequence variants in *RAD51C* are associated with Fanconi anemia complementation group O [123]. The same year as RAD51C was identified as a Fanconi anemia gene, the first unambiguous evidence that monoallelic pathogenic sequence variants in *RAD51C* were associated with increased risk for breast and ovarian cancer was presented by Meindl and colleagues [124]. Meindl and colleagues (2010) reported six pedigrees with six

different pathogenic sequence variants in *RAD51C* (two frame-shift variants, two splice site variants and two missense variants). In all the families, the variant segregated with the disease. In addition, all families presented with both breast and ovarian cancer. The mean age at first diagnosis for breast cancer was 53 years (ranging from 33-78), whereas the mean age for diagnosis of ovarian cancer was 60 years (ranging from 50-81) [124]. The lifetime risk of ovarian cancer was later estimated to 6.12%, while the associated breast cancer risk is still unknown or non-existent [46, 125]. In ClinVar, approximately 100 different germline variants are reported as likely pathogenic or pathogenic, where ca. 70% are reported to be associated with hereditary cancer and ca. 18% of the germline variants are reported to be associated with BOC. Most of the listed BOC associated variants are truncating variants [67].

The **RAD51D** protein (Figure 14) is encoded by the *RAD51D* gene, consisting of 10 exons and located on chromosome 17q12. The gene transcript (NM_002878.3) encodes a 37 kDa protein, consisting of 348 aa [60]. N-terminally, RAD51D has a region that preferentially binds ssDNA (aa 1-83) and a second nucleotide binding region (aa 107-114) [61]. Additionally, it is found that murine RAD51D aa 4-77 has the ability to bind human XRCC2, and aa 77-328 has the ability to bind aa 79-376 of human RAD51C, indicating that this might be the corresponding XRCC2- and RAD51C-binding regions for human RAD51D [121].

Pathogenic mutations in RAD51D are associated with increased risk of ovarian cancer, while the breast cancer risk is currently unknown [125]. The lifetime risk of ovarian cancer is estimated to 10-13.6% [47, 126]. In ClinVar, approximately 70 different germline variants are reported as likely pathogenic or pathogenic, ca. 90% of these are associated with hereditary cancer. Approximately 55% of the variants are reported to be associated with BOC. Most of the reported BOC associated variants are truncating variants [67].

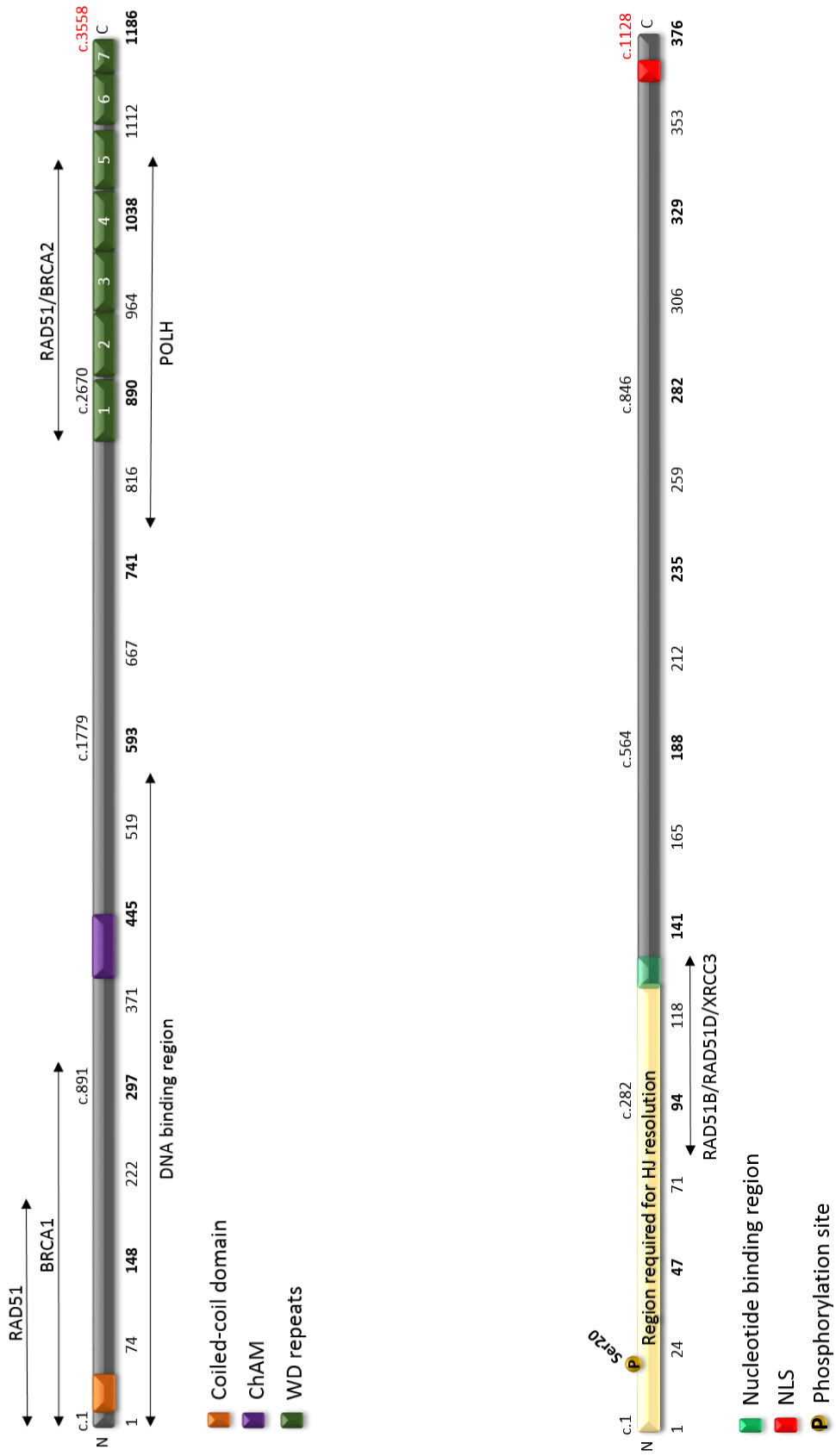


Figure 13 Domain structure of PALB2 and RAD51C. The amino acid numbers are listed underneath, while the corresponding position in the gene is listed above. ***To the left:*** *Domain structure of PALB2. The protein contains a Coiled-coil domain, a ChAM, and seven WD repeats. PALB2 has several interaction regions: RAD51, BRCA1, DNA, BRCA2 and POLH.* ***To the right:*** *Domain structure of RAD51C. The protein contains a region required for Holliday junction resolution, a region for nucleotide binding and a region for interaction with RAD51B/RAD51D/XRCC3. C-terminally, the protein has a NLS.*

1.3.2.5. *BRIP1/BACH1/FANCJ*

The BRIP1 protein was identified by Cantor and colleagues (2001) when they screened for proteins that interact with the BRCT repeats of BRCA1 [127]. BRIP1 is a helicase, which is an enzyme that unwinds DNA strands. BRIP1 plays a role in the Fanconi anemia pathway of interstrand cross-link (covalent linkage of the two DNA strands) repair and in the HRR pathway (Figure 11). BRIP1 also resolves G-quadruplex (G4) DNA structures, a secondary structure consisting of four ssDNA strands bound together. The G-quadruplex negatively affects replication and transcription and may play a role in preservation of telomeres [128].

The BRIP1 protein (Figure 14) is encoded by the *BRIP1* (Alias: *BACH1*, *FANCJ*) gene, consisting of 20 exons and located on chromosome 17q23.2. The gene transcript (NM_032043.2) encodes a 141 kDa protein consisting of 1,249 aa [60]. N-terminally, BRIP1 has a MLH1 binding site (aa 141-142) that interacts with aa 478-744 of MLH1, indicating a role in mismatch repair (MMR), and a NLS (aa 158-175) [61, 128, 129]. As BRIP1 functions as a helicase, it also has a conserved Adenosine Triphosphatase (ATPase) helicase core domain (aa 11-836) comprised of eight motifs; 0 (Q) (including an invariant glutamine, Q25), I (aa39-56), Ia (aa 245-258), II (aa 385-398) with an Aspartate-Glutamate-Alanine-Histidine (DEAH) box (aa 393-396), III (aa 610-624), IV (aa 689-699),V (aa 748-775) and VI (aa 819-836) [61, 127]. It's helicase activity is markedly stimulated by RPA [128]. Between the IA and II motif of the helicase core domain, a Fe-S ([4Fe-4S] iron sulfur) cluster resides at Cys283, Cys298, Cys310 and Cys350, which allows BRIP1 to resolve G4s. C-terminally, BRIP1 has a BRCA1 interaction region (aa 888-1063). Post-translational modifications include phosphorylation of Ser505, Ser927, Ser930, Ser956, Ser990, Ser1004, Ser1032, Thr1133 and Ser1237, together with an acetylation of Lys1249 [61, 128].

Biallelic pathogenic sequence variants in *BRIP1* are associated with Fanconi anemia, subtype J [130, 131]. Monoallelic pathogenic variants in *BRIP1* are associated with HBOC. Although monoallelic pathogenic variants in *BRIP1* are infrequent in HBOC cohorts, in a study where they tested over 35 000 women with breast cancer, several pathogenic variants were identified [39]. Moreover, deleterious variants in *BRIP1* are the third most common cause of hereditary ovarian cancer (after *BRCA1/BRCA2*) [66, 132]. The associated ovarian cancer risk is 4.06-12.7% [46]. In ClinVar, approximately 200 different germline variants are registered as likely pathogenic or pathogenic, where ca. 90% are associated with hereditary cancer, including 55%, which are associated with BOC. Only ca. 3% of the BOC associated variants are missense variants, which are too few to comment on distribution in the gene [67].

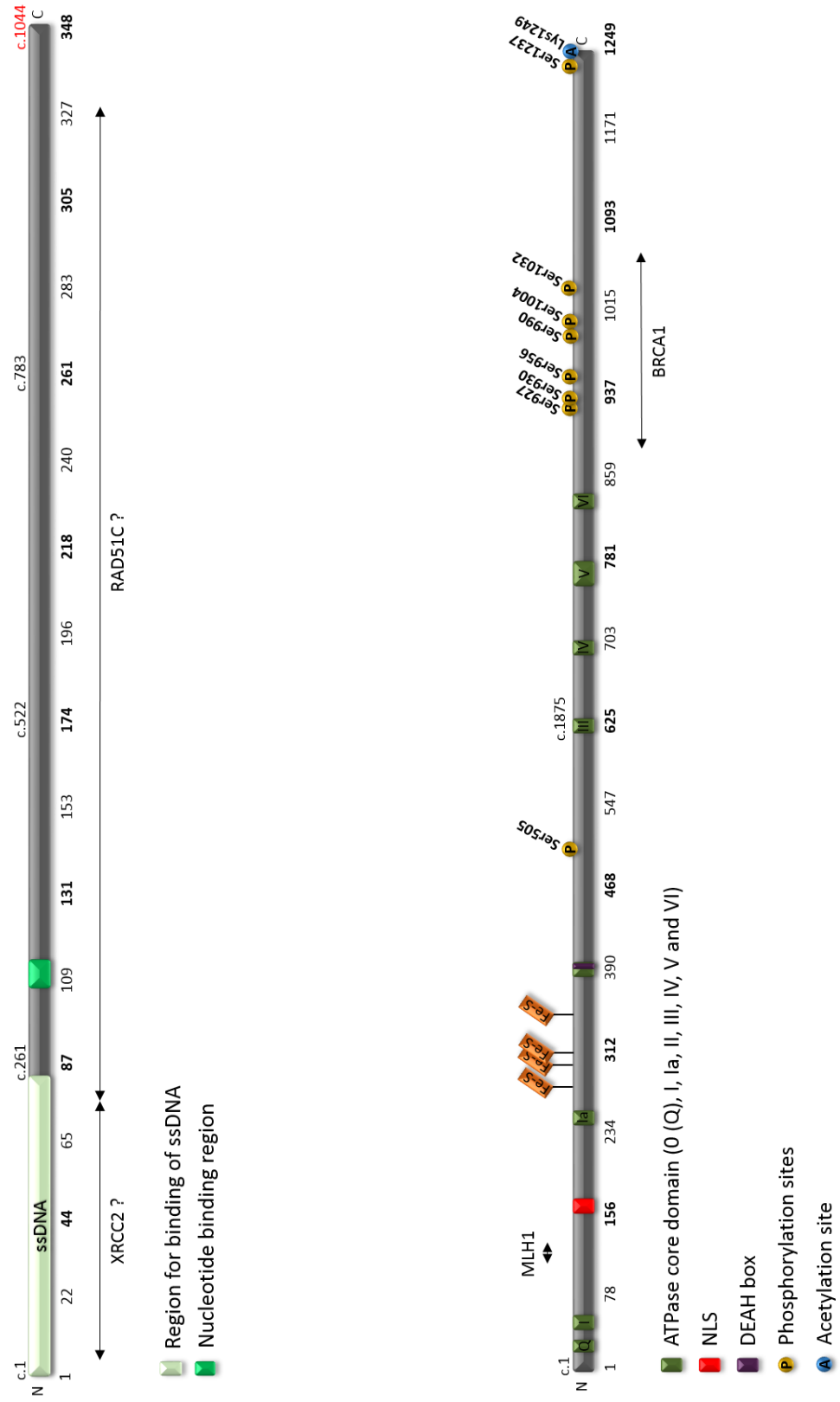


Figure 14 Known features of RAD51D and domain structure of BRIP1. The amino acid numbers are listed underneath, while the corresponding position in the gene is listed above. **To the left:** *Known features of RAD51D. The protein has a region for binding ssDNA and a nucleotide-binding region. Binding regions for XRCC2 and RAD51C have been demonstrated in murine RAD51D.* ***To the right:*** *Domain structure of BRIP1. The protein has an ATPase core domain consisting of eight motifs, with a DEAH box imbedded in motif II. The protein also contains a NLS and an Fe-S cluster consisting of four cysteine each binding 4Fe-4S. BRIP1 has an MLH1 interaction site and a BRCA1 interaction site. In addition, BRIP1 has several phosphorylation sites and one acetylation site.*

1.3.2.6. *RBBP8/CtIP*

The RBBP8 (Alias CtIP) protein is an endonuclease, involved in ssDNA formation at DBS (Figure 11) [133] . Although this protein is involved in HRR, it is not included in the Illumina Cancer panel and was accordingly not further investigated in this study.

1.3.3. The Fanconi anemia pathway

During DNA replication, the replication forks meet obstacles and stall. Disassembly of stalled replication forks may result in DNA damage and to avoid such damages, the Fanconi anemia (FA) pathway is needed to resolve stalled replication forks [134]. Fanconi anemia is a recessive disease characterized by congenital abnormalities, chromosome instability, progressive bone marrow failure and a strong predisposition to cancer. Fanconi Anemia disease will not be discussed further. However, as the FA genes encodes proteins overlapping with the HRR pathway (Table 2), it is possible to assume that some of the FA genes, in addition to genes coding for proteins involved in HRR, are involved in increased breast and ovarian cancer risk.

Table 2 Fanconi Anemia (FA) genes and FA-like genes, and their synonyms. Modified from Katsuki and Takata [42] by including FANCV [135]. There are categorized 21 FA/FA-like genes so far, and some overlap with HR genes. However, biallelic mutations in FANCM has not been proven to cause FA. ICL = DNA interstrand crosslink

FA-Gene name	Synonym	Functions	Symptoms	Heterozygous germline mutation
FANCA		FA core complex	FA pathologies	
FANCB		FA core complex	FA pathologies	
FANCC		FA core complex	FA pathologies	
FANCD1	*BRCA2*	HR repair and protection of stalled replication fork	FA pathologies, not all present with bone marrow failure	HBOC
FANCD2		Protection of stalled replication fork	FA pathologies	
FANCE		FA core complex	FA pathologies	
FANCF		FA core complex	FA pathologies	
FANCG	*XRCC9*	FA core complex	FA pathologies	
FANCI		Required for FA core complex activation	FA pathologies	
FANCJ	*BRIP1, BACH1*	HR repair, ICL repair, 3'-5' helicase activity.	FA pathologies	HBOC
FANCL	*PHF9*	FA core complex	FA pathologies, but no cancers	
FANCM	*Hef*	DNA translocase	HBOC	HBOC
FANCN	*PALB2*	HR repair	FA pathologies	HBOC
FANCO	*RAD51C*	HR repair (RAD51 paralog)	FA pathologies, but no bone marrow failure	HBOC?
FANCP	*SLX4*		FA pathologies	
FANCQ	*XPF, ERCC4*	ICL repair/unhooking, Endonuclease	FA pathologies	HBOC
FANCR	*RAD51*	HR repair, protection of stalled forks	FA-like syndrome, no bone marrow failure.	
FANCS	*BRCA1*	HR repair	FA-like syndrome, no bone marrow failure.	HBOC
FANCT	*UBE2T*	E2 ubiquitin-conjugating enzyme for the FANCD2 complex	FA pathologies	
FANCU	*XRCC2*	HR repair (RAD51 paralog)	FA-like syndrome, no bone marrow failure.	
FANCV	*MAD2L2, REV7*		FA pathologies	

1.3.4. DNA repair by mismatch repair

The MMR pathway maintains genome stability through removal of erroneous base pairing after DNA replication [136]. Consequently, MMR is active during the S-phase and G2-phase of the cell cycle (Figure 6).

The MMR (Figure 15) consists of MutSα or MutSβ, MutLα, Exonuclease 1 (EXO1), Replication factor C (RFC), Proliferating cell nuclear antigen (PCNA), RPA, polymerase-δ (pol- δ) and DNA ligase 1 (LIG1). The MutSα heterodimer consists of MSH2 and MSH6, and predominantly recognizes small 1-2 nucleotide mismatches. The MutSβ heterodimer consists of the MSH2 and the MutS homolog 3 (MSH3) and recognizes larger mismatches than MutSα. The MutLα heterodimer consist of MLH1 and PMS2 [136].

During DNA replication, the DNA-polymerase makes mistakes in spite of its proofreading activity, such as mismatches of bases. The MutSα heterodimer is an ATPase, which recognizes such mismatches and initiates MMR (Figure 15A). The MutSα forms a clamp around the discovered mismatch while exchanging the bound ADP with ATP (Figure 15B). The MutLα is subsequently recruited to the site (Figure 15C). Together, the MutSα and MutLα scan the DNA for nicks in the DNA strand in both directions (5'→ 3' and 3'→ 5') until it encounters PCNA (located at nicks). Depending on whether the identified nick is located 5' or 3' of the mismatch, EXO1 loading is performed differently. If the nick is 5' of the mismatch, EXO1 is loaded directly at the nick (Figure 15D). The EXO1 degrades the DNA strand in a 5' → 3' direction, approximately 150 nucleotides past the mismatch. However, if the identified nick is situated 3' of the mismatch, the ternary complex (MutSα, MutLα and PCNA) travels along the DNA and introduce additional breaks 5' of the located nick and eventually 5' of the mismatch (not shown in figure). EXO1 is loaded at the newly introduced breaks and degrades the strand in a 5'→ 3' direction (Figure 15E). Both these recruitment options of EXO1 leaves a large gap covering the former mismatch, which can be filled by the PCNA/pol-δ complex and the ends are ligated by LIG1 (Figure 15F) [137].

One of the functions of MMR is to repair mismatches at microsatellite positions. Microsatellites are mononucleotide, dinucleotide or trinucleotide repeats. These positions may cause the polymerase to stutter during replication and lead to lower or higher copy-number of the repeated sequence. Accordingly, deficiency in the MMR caused by pathogenic variants in the *MSH2, MSH6, MLH1* and *PMS2* genes may lead to microsatellite instability (MSI).

The MMR machinery is also involved in response to methylated guanine at the oxygen-6 position (O6MG). If left unrepaired, the polymerase incorporates a thymine at the opposite strand of O6MG during DNA replication, which subsequently leads to G:C to A:T transition. Processing of the O6MG is done by either the O^6-alkylguanine DNA alkyltransferase (AGT) or the MMR machinery, where the latter is potentiated by BRIP1 and results in apoptosis (see section **1.3.2.5. *BRIP1/BACH1/FANCJ***) [129, 138].

Dysregulation of the MMR due to pathogenic sequence variants in the *MSH2, MSH6, MLH1* and *PMS2* genes is associated with Lynch syndrome, also called hereditary nonpolyposis colorectal cancer (HNPCC) [139]. Lynch syndrome increases the risk of several types of cancers, including colorectal, stomach, small intestine, liver, gallbladder ducts, upper urinary tract, brain and skin cancers [140]. Furthermore, several reports of pathogenic variants in MMR genes in HBOC patients have now been published [34-36, 38, 39, 132, 141]. These four genes were subsequently included in our study.

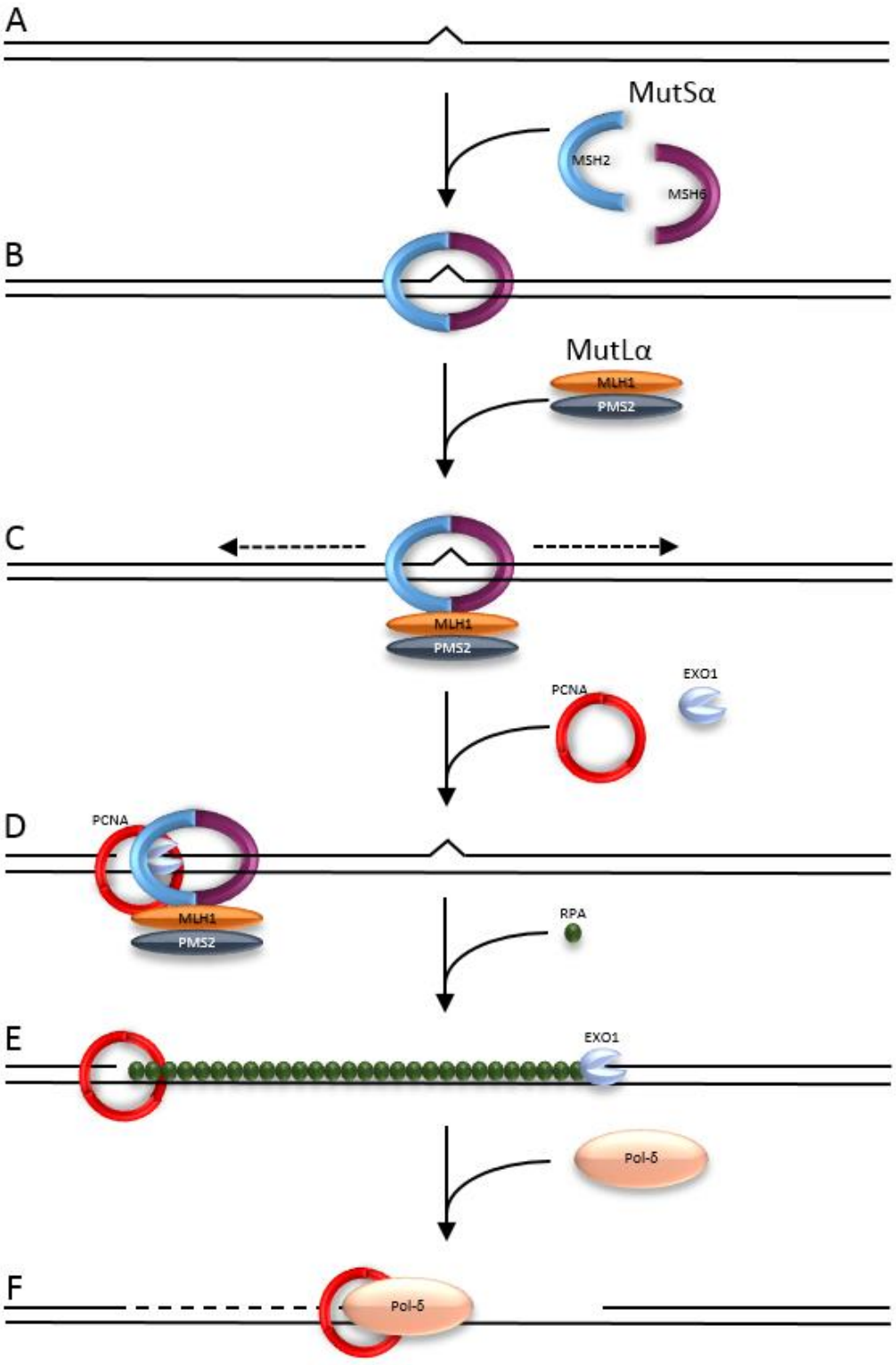


Figure 15 Schematic presentation of the Mismatch repair mechanism. A) A mismatch occurring during replication leads to a bulge in the DNA. B) This mismatch is recognized by MutSα, which consists of MSH2 and MSH6. MutSα forms a clamp around the mismatch. C) MutLα is recruited to the complex and the entire complex scans the DNA for single stranded breaks (this can occur in both a 5' and 3' direction). D) When the MutSα/MutLα complex finds a single stranded break/nick with PCNA already loaded, EXO1 is recruited to the nick. This figure displays the identified nick located 5' of the mismatch. If the nick is encountered 3' of the mismatch, then the MutSα/MutLα/PCNA-complex inserts several breaks 5' of the identified nick. EXO1 is recruited to these breaks. E) EXO1 degrades the nicked strand in a 5'→3' direction. At the same time, RPA is loaded onto the remaining strand. F) Polymerase-δ is recruited and synthesizes a new DNA strand with the help of PCNA. The nick is afterwards ligated by DNA ligase 1 (not shown).

1.3.4.1. *MSH2*

The MSH2 protein plays a central part in the MMR pathway, where it forms a heterodimer with MSH6, called MutSα (Figure 15), or MSH3, called MutSβ.

The MSH2 protein is encoded by the *MSH2* gene, consisting of 16 exons and located on chromosome 2p21. The gene transcript (NM_000251.2) encodes a 105 kDa protein, consisting of 934 aa [60]. The MSH2 protein (Figure 16) has five domains, similar to the MSH6 C-terminal. N-terminally the protein has a DNA mismatch binding domain (aa 17-132), followed by a connector domain (aa 145-290) containing three surface loops that may mediate protein-protein interaction (aa 150–160, 207–217, and 243–262 [142]) and a Lever (aa 305-609) with an integrated Clamp domain (aa 473-569). C-terminally, MSH2 has an ATPase domain (aa 665-852) [61, 62, 143]. In addition, the MSH2 protein has an EXO1 interaction region (aa 601-671). Post-translational modifications include removal of Met1, acetylation of Ala2, Lys555 and Lys567, cross-link formation of Lys430 with SUMO2 and phosphorylation of Ser921 [61]

Pathogenic variants in *MSH2* are mainly associated with colorectal cancer. However, these variants are also associated with an increased risk of breast and ovarian cancer, with an estimated lifetime risk of 11.5-22% and 10-24%, respectively (Table 1) [48-51]. In ClinVar, approximately 900 different germline variants are reported as likely pathogenic or pathogenic. None of the variants are reported specifically associated with breast or ovarian cancer [67]. For the MMR genes, an expert panel called the International Society for Gastrointestinal Hereditary Tumours (InSiGHT) has reviewed several of the variants, which are reported associated with Lynch syndrome. In total, they have reviewed approximately 540 variants and classified as likely pathogenic or pathogenic, ca. 9% of which are missense variants. Most (ca. 85%) of the missense variants are reported in the Connector domain, the Lever domain and the ATPase domain.

1.3.4.2. *MSH6*

The MSH6 protein plays a central role in the MMR pathway, where it forms a heterodimer with the MSH2, called MutSα (Figure 15).

The MSH6 protein is encoded by the *MSH6* gene, consisting of 10 exons and located on chromosome 2p16.3. The gene transcript (NM_000179.2) encodes a 153 kDa protein, consisting of 1,360 aa [60]. The MSH6 protein (Figure 16) has several domains in common with MSH2, only with a different and longer N-terminal region. The N-terminal region contains a PCNA interaction protein (PIP) motif (aa 4-11) that recognizes and binds PCNA and a Proline-Tryptophan-Tryptophan-Proline (PWWP) sequence (aa 104-107) that is characteristic for proteins that associate with chromatin. The common domains shared with MSH2 is the DNA mismatch binding domain (aa 407-526), a Connector domain (aa 537-703) containing three surface loops that may mediate protein-protein interaction (aa 545-555, 602-612 and 650-675 [142]), a Lever domain (aa 738-1064) with an integrated Clamp domain (aa 932-1024), and an ATPase domain (aa 1130-1324) [62, 143]. Post-translational modifications include phosphorylation (of Ser14, Ser41, Ser43, Ser79, Ser91, Ser137, Ser200, Ser219, Ser227, Ser252, Ser254, Ser256, Ser261, Thr269, Ser274, Ser275, Ser279, Ser280, Ser309, Thr488, Ser830, Ser935 and Thr1010) and acetylation (of Lys70 and Lys504) [61].

Pathogenic variants in *MSH6* are frequently associated with colorectal cancer, but have also been associated with increased ovarian cancer risk (Table 1). The lifetime risk of breast cancer is estimated to 12.5% by age 80 years [49]. The lifetime risk of ovarian cancer is estimated from 1% (lower than the general population) to 10.5%, equivalent to a high risk [49, 51]. In

ClinVar, approximately 600 different germline variants are listed as likely pathogenic or pathogenic. None of the variants registered are specifically associated with breast or ovarian cancer [67]. InSiGHT has reviewed approximately 170 likely pathogenic and pathogenic *MSH6* variants, ca. 4% of which are missense variants (too few to comment on distribution).

1.3.4.3. *MLH1*

The MLH1 protein is a central component of the MMR, where it forms a heterodimer with PMS2, called MutLα (Figure 15).

The MLH1 is encoded by the *MLH1* gene, consisting of 19 exons and located on chromosome 3p22.2. The gene transcript (NM_000249.3) encodes an 85 kDa protein, consisting of 756 aa [60]. N-terminally, the MLH1 protein (Figure 17) contains a histidine kinase domain (aa 25-138), including two ATP binding sites (aa 38 and 63) and two ATP binding regions (aa 82-84 and 100-104) [61, 62]. In addition, MLH1 has a DNA mismatch repair domain (aa 216-335) and an MLH1 C-terminal domain (aa 502-756), the latter forming a part of the endonuclease active site of MutLα [62]. The protein interacts with EXO1 (aa 410-650) and PMS2 (aa 506-675) [144, 145]. Post-translational modifications include removal of Met1, acetylation of Ser2 and phosphorylation of Ser477 [61].

Pathogenic variants in *MLH1* are frequently associated with colorectal cancer [139]. However, they have also been reported to be associated with breast and ovarian cancer (Table 1), with an estimated lifetime risks of 18.6-25% and 5-20%, respectively [49, 51, 52]. In ClinVar, approximately 800 different germline variants are listed as likely pathogenic or pathogenic. None have been entered as associated with breast or ovarian cancer [67]. InSiGHT has reviewed approximately 580 likely pathogenic and pathogenic *MLH1* variants, ca. 14% of which are missense variants. Approximately 90% of the missense variants are reported in the Histidine kinase domain (ca. 40%), DNA mismatch domain (ca. 20%) and the MLH1 C-terminal domain (ca. 30%).

1.3.4.4. *PMS2*

The PMS2 protein is a central component of the MMR, where it forms a heterodimer with MLH1, called MutLα (Figure 15).

The PMS2 protein is encoded by the *PMS2* gene, consisting of 15 exons and located on chromosome 7p22.1. The gene transcript (NM_000535.6) encodes a 96 kDa protein, consisting of 862 aa [60]. The PMS2 protein (Figure 17) contains the same domains as MLH1; a histidine kinase domain (aa 33-161), a DNA mismatch repair domain (aa 248-363) and a MutL C terminal dimerization domain (aa 676-822), responsible for interaction with MLH1 [62, 144]. Post-translational modifications include phosphorylation of Thr573 and Thr597 [61].

The *PMS2* gene has several pseudogenes, where 14 pseudogenes comprise pseudocopies of some or all of exons 1-5. Another pseudogene is however of greater importance, due to the sequence alignment problems it may cause, the *PMS2CL*. The *PMS2CL* shares 98% sequence identity with exon 9 and 11-15 of *PMS2* and is located approximately 700 kb centromeric to *PMS2* on chromosome 7, as an inverted duplication [146, 147].

Pathogenic variants in *PMS2* are frequently associated with colorectal cancer [139]. However, they have also been reported in association with ovarian cancer, with a lifetime ovarian cancer risk of <6% (Table 1) [50]. In ClinVar, approximately 260 different germline variants are listed as likely pathogenic or pathogenic. However, none of them have been entered as associated

with breast or ovarian cancer [67]. InSiGHT has reviewed approximately 70 likely pathogenic and pathogenic *PMS2* variants, ca. 3% are missense variants, which is too few to comment on distribution in the gene.

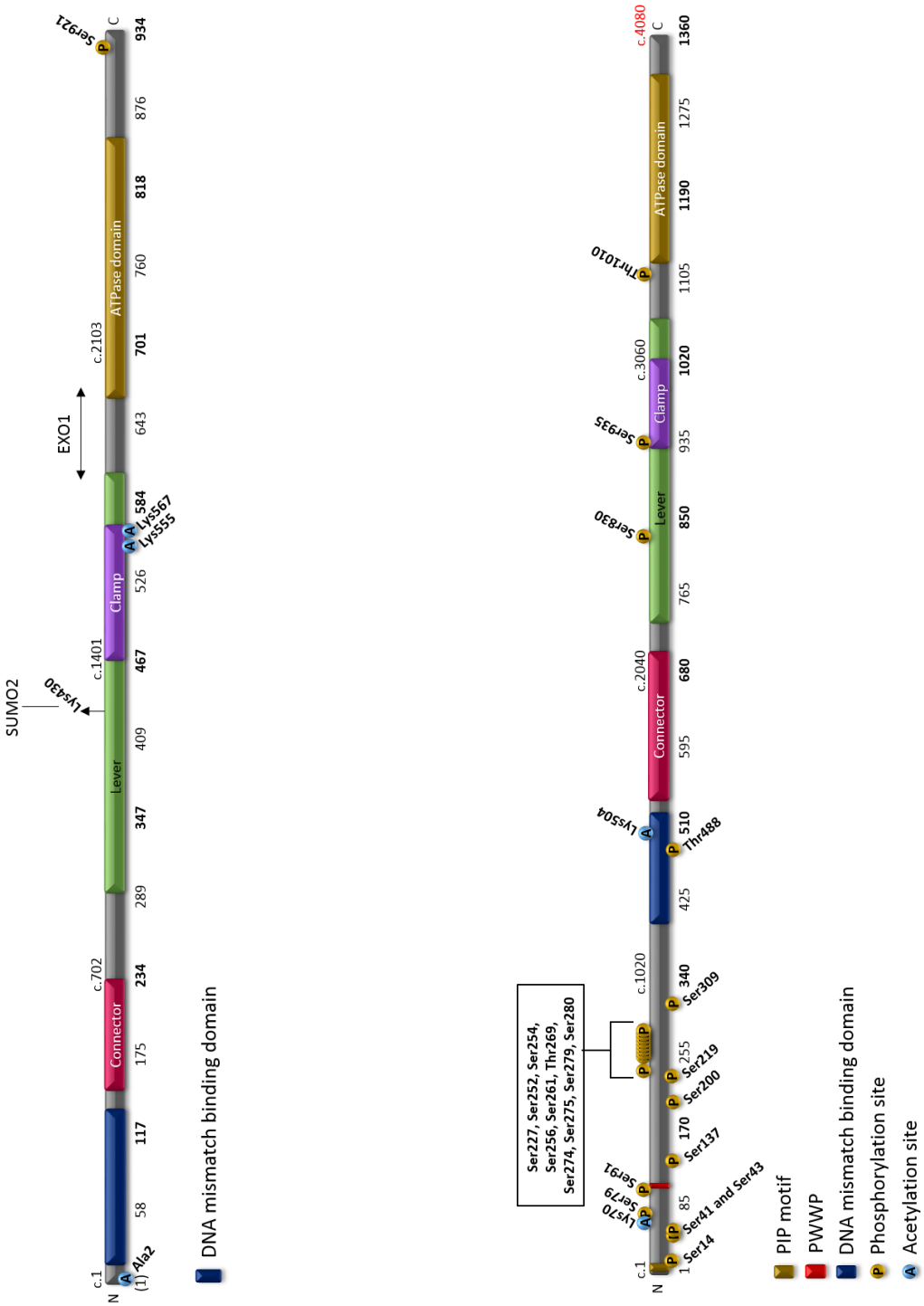


Figure 16 Domain structure of MSH2 and MSH6, which form the MutSα. The amino acid numbers are listed underneath, while the corresponding position in the gene is listed above. ***To the left:*** *Domain structure of MSH2. The protein contains five domains, including the DNA mismatch binding domain, the connector domain, the lever domain with the incorporated Clamp domain and C-terminally the ATPase domain. MSH2 has an EXO1 interaction region and has several post-translational modifications, including removal of Met1, acetylation at three sites, a phosphorylation site and a SUMOylation site.* ***To the right:*** *Domain structure of MSH6. The protein contains the same domain structure as MSH2, but with an additional N-terminal region, including a PIP motif, a PWWP sequence and several phosphorylation sites. Additional domains include the DNA mismatch binding domain, the connector domain, the lever domain with the incorporated Clamp domain and C-terminally the ATPase domain.*

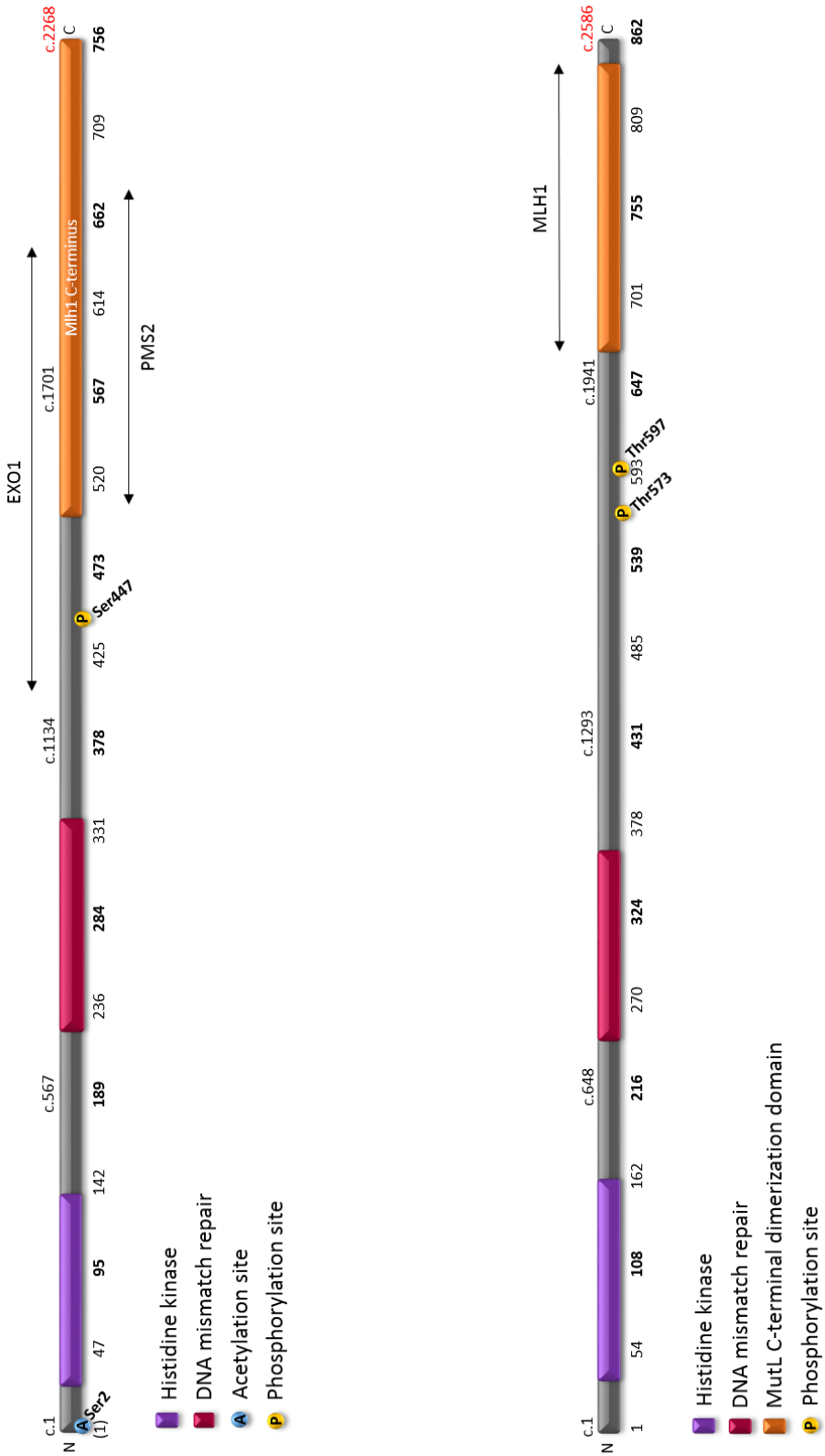


Figure 17 Domain structure of MLH1 and PMS2, which form the MutLα. The amino acid numbers are listed underneath, while the corresponding position in the gene is listed above. ***To the left:*** *Domain structure of MLH1. N-terminally, MLH1 has a histidine kinase domain, with four ATP binding sites/regions (not shown), and a DNA mismatch repair domain. C-terminally the protein has an MLH1 C-terminus domain. In addition, MLH1 has an EXO1 and a PMS2 interaction region C-terminally. Post-translational modifications include removal of Met1, acetylation of Ser2 and phosphorylation of Ser447. ,* ***To the right:*** *Domain structure of PMS2. PMS2 has similar domains as MLH1, with a histidine kinase domain, followed by a DNA mismatch repair domain and a MutL C-terminal dimerization domain responsible for interaction with MLH1. PMS2 has two phosphorylation sites.*

1.3.5. Other mechanisms involved in HBOC

1.3.5.1. *CDH1*

The Epithelial-cadherin (E-cadherin) is a transmembrane protein involved in calcium-dependent cell-cell adhesion (Figure 18). This is an important mechanism to ensure the immobilization of cells. Dysregulation of this adhesion leads to increased cell mobility [25]. Increased mobility is an important capability for cancer metastasis.

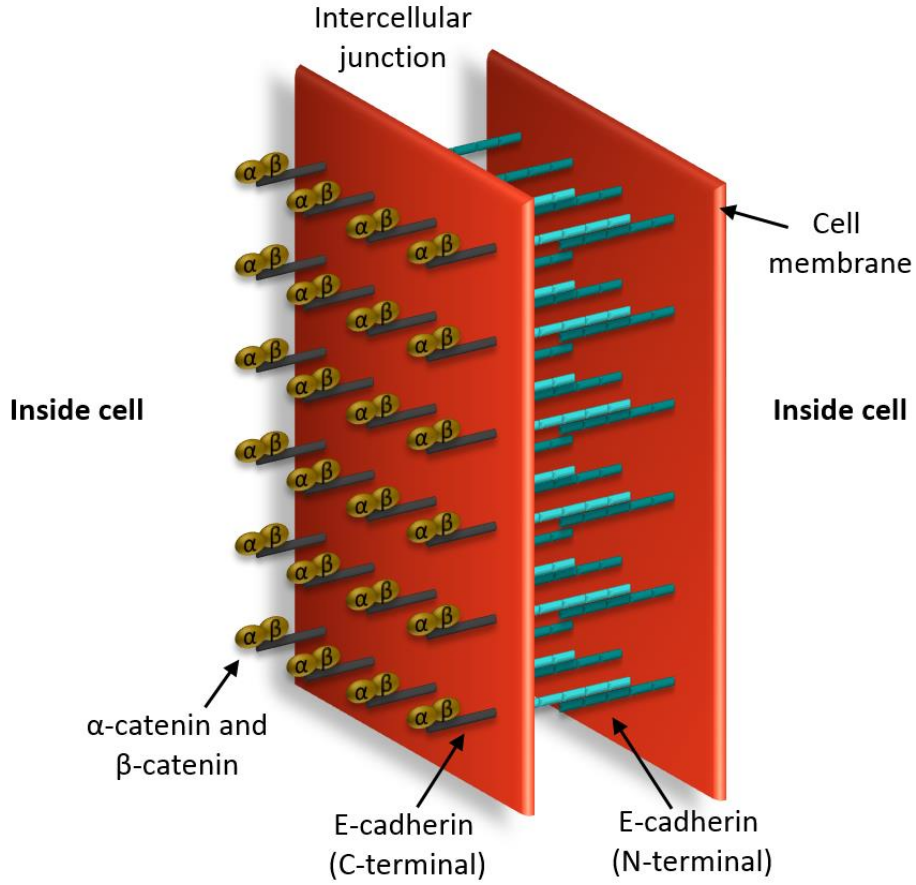


Figure 18 E-cadherin in calcium-dependent cell-cell adhesion. The cytoplasmic part of the protein binds to α-catenin and β-catenin. These catenins link E-cadherin to the cytoskeleton.

The E-cadherin is encoded by the *Cadherin 1* (*CDH1*) gene, consisting of 16 exons and located on chromosome 16q22.1. The gene transcript (NM_004360.3) encodes a 97 kDa protein, consisting of 882 aa [60]. E-cadherin (Figure 20) is a membrane-bound protein, with a signal peptide from aa 1-22 and a propeptide from aa 23-154. After cleavage of the signal/propeptide, E-cadherin consists of an extracellular region (aa 155-709), a helical transmembrane region (aa 710-730) and an intracellular region (aa 731-882). The extracellular region comprises five Cadherin domains: Cadherin 1 (aa 155-262), Cadherin 2 (aa 263-375), Cadherin 3 (aa 376-486), Cadherin 4 (aa 487-593) and Cadherin 5 (aa 594-697). The E-cadherin protein has a Catenin delta 1 (CTNND1) and Presenilin 1 (PSEN1) (aa 758-769) interaction region, and a region for α-, β- and γ-catenin interaction (aa 811-882). Additionally, an inter-chain disulfide bond is formed from Cys163 for homodimerization of the extracellular domains (Figure 18). Post-translational modifications of the protein include several glycosylation events (of Asn558,

Asn570, Asn622 and Asn637) and phosphorylation events (of Tyr753, Tyr754, Tyr755, Ser770, Ser793, Ser838, Ser840 and Ser846). [61].

If apoptosis is induced, the E-cadherins are cleaved at three sites by different proteins; a site recognized by a metalloproteinase (aa 700-701), another site by gamma-secretase/PS1 (aa 731-732) and a third site by caspase-3 (aa 750-751) [61].

Pathogenic sequence variants in *CDH1* lead to hereditary diffuse gastric cancer syndrome, with an increased risk of diffuse gastric cancer and of lobular breast cancer [25]. Hereditary diffuse gastric cancer syndrome leads to a lifetime risk for breast cancer of 42% (Table 1) [54]. In ClinVar, approximately 160 different germline variants are catalogued as likely pathogenic or pathogenic, only ca. 1% are reported as breast or ovarian cancer associated. However, several variants are association with diffuse gastric cancer syndrome, which may include cases of lobular breast cancers [67].

1.3.5.2. *PTEN*

The PTEN protein is responsible for regulating the PI3K - AKT serine/threonine kinase 1 (AKT1) - Mechanistic target of rapamycin kinase (mTOR) signaling pathway (Figure 19) [25]. The PI3K-AKT1-mTOR pathway is responsible for inhibiting apoptosis in response to extracellular survival signals. Extracellular survival signals bind the monomeric receptor tyrosine kinases (RTK), leading to homodimerization of receptors with bound signal molecules. Upon homodimerization, the RTKs intracellular tyrosine kinase domain is cross-phosphorylated and recruits the PI3K protein. The PI3K is responsible for phosphorylating phosphatidylinositol 4,5-biphosphate (PIP_2), leading to phosphatidylinositol (3,4,5)-triphosphate (PIP_3). The PIP_3 functions as a docking site for AKT1 and pyruvate dehydrogenase kinase 1 (PDK1), and the signal is subsequently forwarded. The PTEN protein is responsible for dephosphorylating the third carbon atom of the inositol ring of PIP_3 and dysregulation of its activity prolongs the survival signal, promoting uncontrolled cell growth [7].

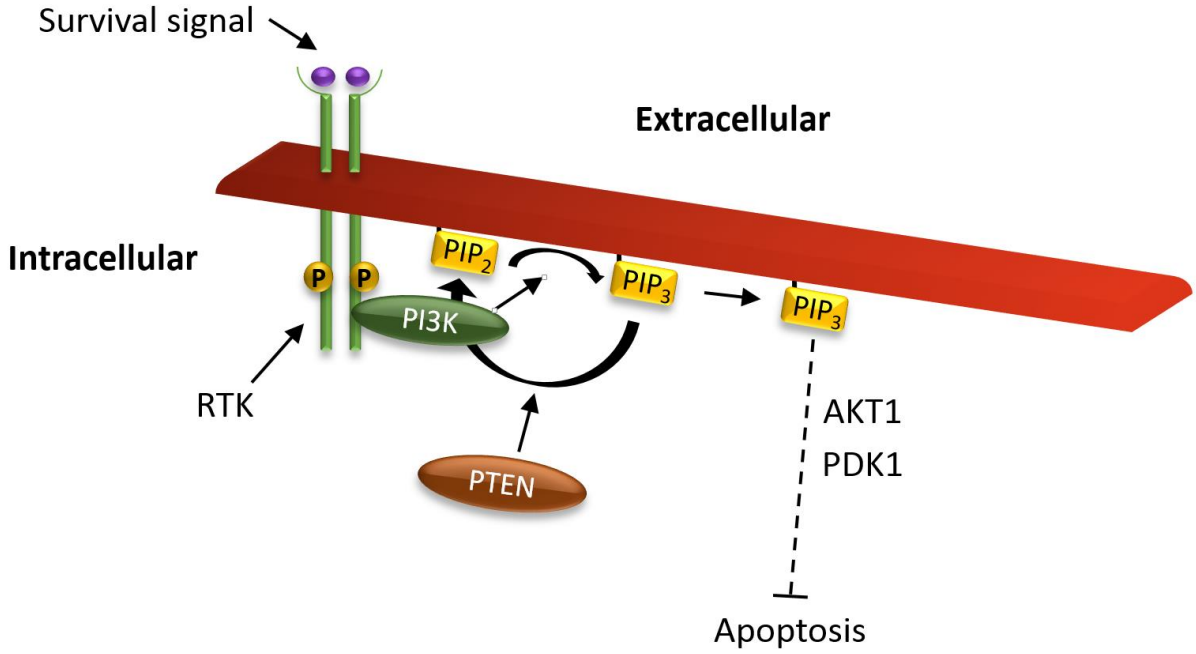


Figure 19. The PI3K-AKT1-mTOR pathway leads to inhibition of apoptosis. The role of PTEN is regulating the signaling of survival signals. Survival signal molecules bind RTK, leading to phosphorylation of the intracellular domains. The PI3K is recruited and phosphorylates PIP_2 to PIP_3, forwarding the survival signal. PTEN dephosphorylates PIP_3.

The PTEN protein (Figure 20) is encoded by the *PTEN* gene, consisting of 9 exons and located on chromosome 10q23.31. The gene transcript (NM_00314.4) encodes a 47 kDa protein, consisting of 403 aa [60]. The PTEN protein contains a phosphatase tensin-type domain (aa 14-185), a C2 tensin-type domain (aa 190-350), a region required for interaction with NOP53 ribosome biogenesis factor (NOP53) (aa 338-348) and a region for PDZ domain-binding (aa 401-403) [61]. Post-translational modifications include removal of Met1, acetylation of Thr2, ubiquitination through cross-link formation of Lys13 and Lys289 (both interchain, linked to Gly C-terminally in ubiquitin) and phosphorylation of Ser294, Tyr336, Thr366, Ser370, Ser380, Thr382, Thr383, Ser385 and Thr401 [61].

Pathogenic sequence variants in *PTEN* are associated with Cowden syndrome, characterized by macrocephaly, skin hamartomas, gastrointestinal polyps and increased risk of thyroid, endometrial (inner membrane of the uterus, Figure 4), kidney and breast cancer [25]. The lifetime risk of breast cancer is estimated to be 85.2% [53]. Tan and colleagues (2012) [53], found that approximately 30% of pathogenic variants in female breast cancer cases were nonsense mutations and approximately 30% were missense. In addition, significant correlations between breast cancer and promoter mutations were identified (13% of the cases had promoter mutations). Approximately 380 different germline variants are reported in ClinVar as likely pathogenic or pathogenic. Merely 1.6% are listed to be associated with neoplasms of the breast or ovaries. However, several variants are registered to be associated with Cowden syndrome, which may include breast cancers [67].

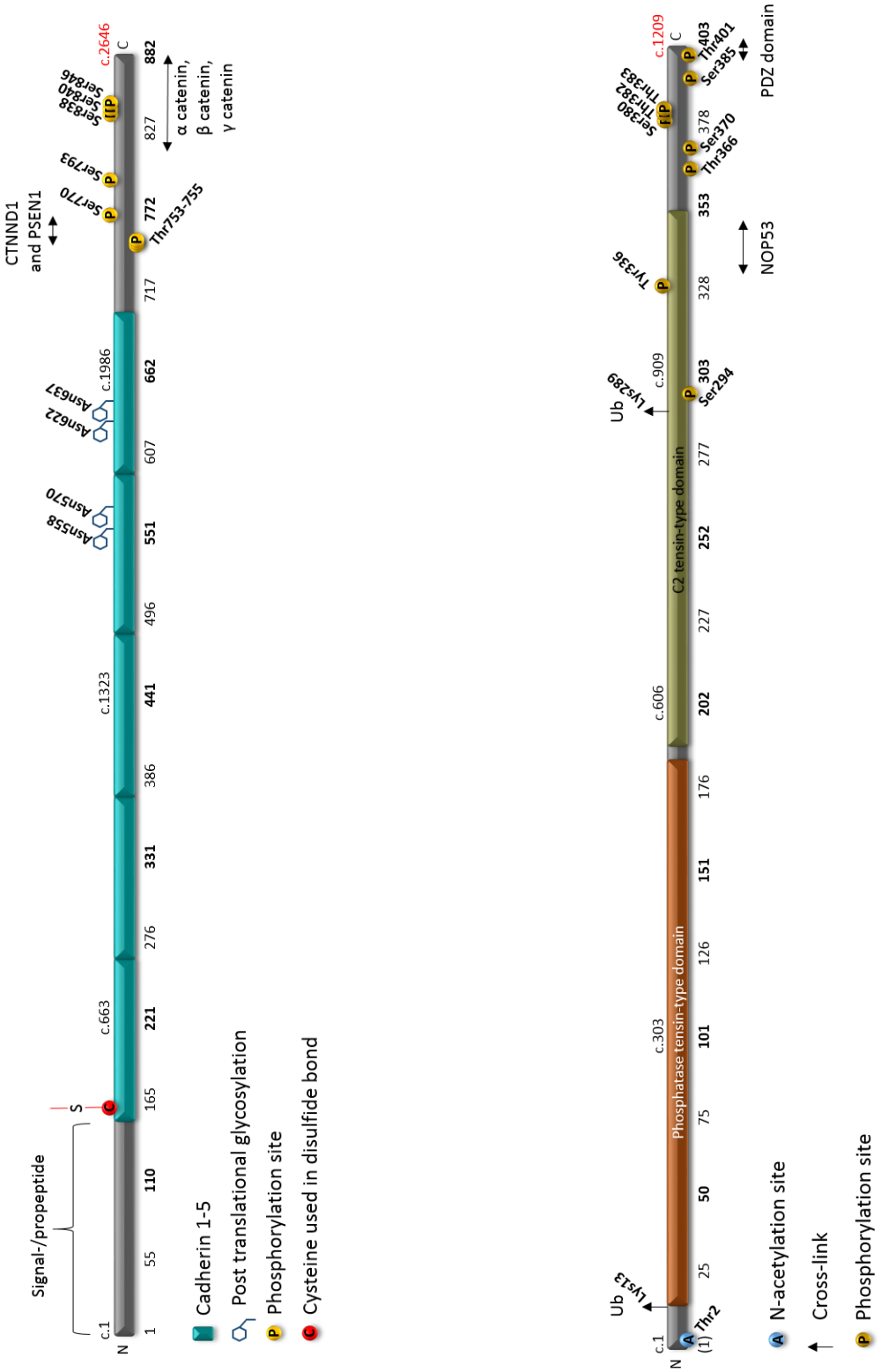


Figure 20 Domain structure of CDH1 and PTEN. The amino acid numbers are listed underneath, while the corresponding position in the gene is listed above. ***To the left:*** *Domain structure of CDH1. N-terminally the protein has a signal-/propeptide sequence to guide it to the membrane. The protein has five cadherin domains, a cysteine (Cys163) for disulfide linkage for homodimerization, several glycosylation and phosphorylation sites and binding sites for CTNND1, PSEN1, α catenin, β catenin and γ catenin.* ***To the right:*** *Domain structure of PTEN. The protein has two domains, the phosphatase tensin-type domain and C2 tensin-type domain. Met1 is removed as a post-translational modification, Thr2 is acetylated, Lys13 and Lys289 are linked to ubiquitin, and C-terminally, the protein has several phosphorylation sites.*

1.3.5.3. *STK11*

The Serine/threonine kinase 11 (STK11) has been directly linked to HRR. This became clear when cells defective in STK11 were shown to be radiation sensitive, accumulate DSBs, and have an increased mutation rate, in addition, the cells did not show any homology-directed repair [148]. The STK11 protein positively regulates the AMP-activated protein kinase (AMPK), which further indirectly regulates the mTOR protein (overlapping with the PI3K-AKT1-mTOR pathway that PTEN regulates (Figure 19)). This regulation subsequently functions as an antagonistic regulator of survival signals [149].

The STK11 protein (Figure 22) is encoded by the *STK11* gene, consisting of 10 exons and located on chromosome 19p13.3. The gene transcript (NM_000455.4) encodes a 49 kDa protein, consisting of 433 aa [60]. STK 11 is translated as a propeptide, where three aa (aa 431-433) are removed from the C-terminal during protein maturation. The protein has a protein kinase domain spanning most of its length (aa 49-309). The kinase domain has a binding site for ATP (aa 7), an active site (aa 176) that serves as a proton acceptor and a nucleotide binding region (aa 55-63). In addition, post-translational modifications include phosphorylation (of Ser31, Thr189, Ser325, Thr336, Thr363, Ser401 and Ser428), acetylation (of Lys44, Lys48, Lys96, Lysin97, Lys296, Lys311, Lys416 and Lys423), lipidation (of Cys418 and Cys430) and addition of methyl ester at Cys430 [61].

Pathogenic variants in STK11 are linked to Peutz-Jeghers syndrome (PJS). Peutz-Jeghers syndrome is an autosomal dominant disorder characterized by gastrointestinal polyposis, mucocutaneous (region where mucosa transitions to skin, e.g. lips) pigmentation (which may fade in puberty and adulthood) and an increased cancer risk [150]. The increased cancer risk is associated with colorectal, gastric, pancreatic, breast and ovarian cancers [150]. The cumulative risk of female breast cancer by age 60 years is 32% and gynecological cancer (including ovarian cancer) is 13% (Table 1) [55]. In ClinVar, approximately 130 different germline variants are reported as likely pathogenic or pathogenic. None of the variants are documented in association with BOC. However, ca. 60% of the variants are associated with hereditary cancer-predisposition syndrome [67].

1.3.5.4. *NF1*

The Neurofibromin 1 (NF1) is a GTPase activating protein (GAP) that regulates the RAS type GTPase family (RAS) signaling pathway. The RAS protein is responsible for relaying signals from RTKs at the cellular surface to the nucleus, frequently in response to signals promoting cell division (Figure 6). RAS is inactive when guanine diphosphate (GDP) is bound and active when guanine triphosphate (GTP) is bound (Figure 21). The switch between GDP-RAS and GTP-RAS is regulated by Guanine nucleotide exchange factors (GEFs) and GAPs. The GEF is responsible for removing GDP, allowing GTP to take its place, and the GAP removes one phosphate group from the GTP, rendering RAS inactive [7].

The NF1 protein (Figure 22) is encoded by the *NF1* gene, consisting of 58 exons and located on chromosome 17q11.2. The gene transcript (NM_001042492.2) encodes a 319 kDa protein, consisting of 2,839 aa [60]. The NF1 protein has a Ras-GAP domain (aa 1235-1451) and a CRAL-TRIO domain (aa 1580-1738) that is a structural domain that binds small lipophilic molecules, a lipid binding region (aa 1580-1837) and a bipartite NLS (aa 2555-2571) [61, 62]. Post-translational modifications include removal of Met1, acetylation of Ala2, and phosphorylation of several aa (Ser864, Ser876, Ser2188, Ser2467, Thr2514, Ser2515, Ser2521, Ser2523, Ser2543, Thr2565, Ser2597, Ser2802 and Ser2817) [61].

Pathogenic variants in *NF1* are linked to Neurofibromatosis 1. Neurofibromatosis 1 is characterized by multiple café au lait spots (pigmented birthmarks), axillary (armpit) and inguinal (groin) freckling, multiple cutaneous neurofibromas (benign nerve sheath tumor), iris Lisch nodules (iris hamartoma), and choroidal (vascular layer of the eye) freckling. Over 50% of individuals with Neurofibromatosis 1 have learning disabilities [151]. The cumulative risk of female breast cancer is estimated to be 18% [56]. There is no increased ovarian cancer risk linked to pathogenic variants in *NF1* [44]. In ClinVar, approximately 700 different germline *NF1* variants are described as likely pathogenic or pathogenic. None of the variants are listed in association with BOC. However, ca. 30% of the variants are associated with hereditary cancer-predisposition syndrome [67].

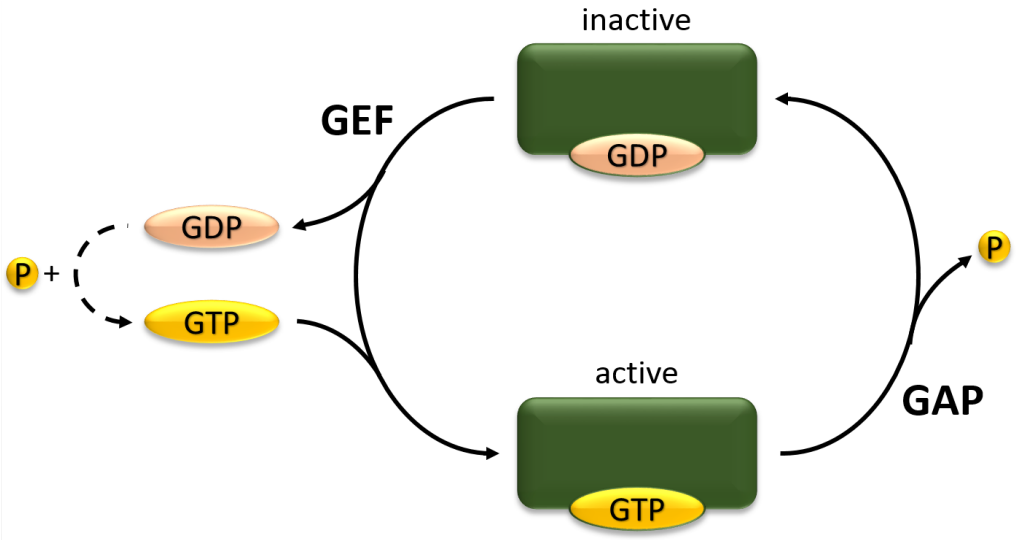


Figure 21 One of the major intracellular signaling mechanisms – signaling by GTP-binding protein. The GTP-binding protein is inactive when GDP is bound and active when GTP is bound. This change between active and inactive is regulated by guanine nucleotide exchange factors (GEFs) and GTPase-activating proteins (GAPs), respectively. Upon removal of the GDP by GEF, GTP immediately fills the nucleotide binding site [7].

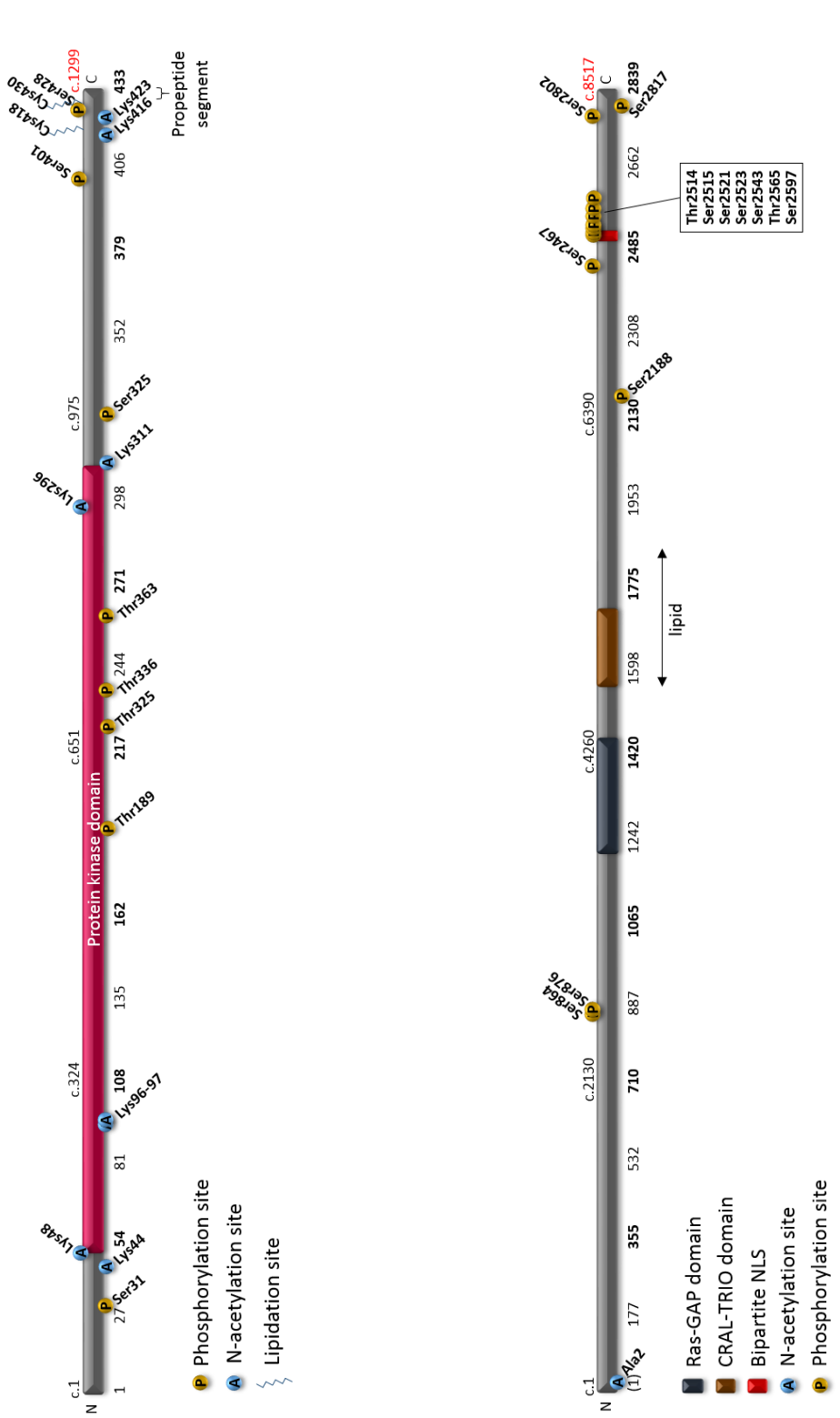


Figure 22 Domain structure of STK11 and NF1. The amino acid numbers are listed underneath, while the corresponding position in the gene is listed above. ***To the left:*** *Domain structure of STK11. The protein has a protein kinase domain spanning most of its length, a binding site for ATP at aa 78, an active site at aa 176 that serves as a proton acceptor and a nucleotide binding region. Three amino acids (aa 431-433) are removed from the C-terminal during protein maturation. In addition, the protein has an added methyl ester, two lipids, several phosphate groups and several acetyl groups.* ***To the right:*** *Domain structure of NF1. Met1 is removed as a post-translational modification, Ala2 is acetylated and several other amino acids are phosphorylated. The NF1 protein has a Ras-GAP domain and a CRAL-TRIO domain that is a structural domain that binds small lipophilic molecules, a lipid binding region and a bipartite NLS.*

1.4. Splicing

Genes consist on average of 8.8 exons with a mean size of 145 base pairs (bp), interspaced by non-coding introns with a mean size of 3,365 bp [152]. The genes are transcribed into pre-mRNA transcripts, and during processing of the pre-mRNA, it is essential that the introns are correctly removed. This processing is performed by the splicing machinery, also known as the spliceosome. The spliceosome consists of several small nuclear ribonucleoprotein particles (snRNPs); U1, U2, U4, U5 and U6. These snRNPs assemble onto the pre-mRNA together with other proteins, to remove introns. The result is a mature mRNA, which is on average 1,340 nucleotides [152]. The intron removal process takes place mainly during transcription, however, post-transcriptional removal is also possible [153]. Based on the emerging evidence for co-transcriptional splicing, Herzel et al. (2017) have reviewed how transcription elongation dynamics and RNA folding might impact identification of introns and exons by the spliceosome [154].

1.4.1. Splice site recognition and the spliceosome

The spliceosome recognizes exon-intron boundaries through splice site (ss) sequences at the start and end of introns, and through splicing regulatory elements (SRE), which is described below.

1.4.1.1. Splice sites

Intron-exon boundaries are often recognized by the spliceosome due to the highly conserved +1, +2, -1 and -2 positions in introns. The general structure of introns follows the GT-AG rule, where the intron starts with guanine (G) and thymine (T) (uracil in RNA) at positions +1 and +2, respectively, and ends with adenine (A) and guanine (G) in positions -2 and -1, respectively (Figure 23). According to Lander et al. (2001) [152], approximately 98% of introns use the canonical GT-AG pattern, while approximately 1% utilize the closely resembling GC-AG pattern and even less utilize a AT-AC pattern. The AT-AC pattern is recognized by the minor spliceosome, instead of the major spliceosome that recognizes the GT-AG dinucleotides. In addition to the splice site sequences, the intron also contains a polypyrimidine tract near the 3' end of the intron, as well as a highly conserved adenosine (A) called the branch point site (BPS), located 18-40 nucleotides upstream of the 3'ss [154]. Sequence variants in these positions may potentially interfere with the spliceosome recognition and binding, and consequently interfere with the splicing of the corresponding exon.

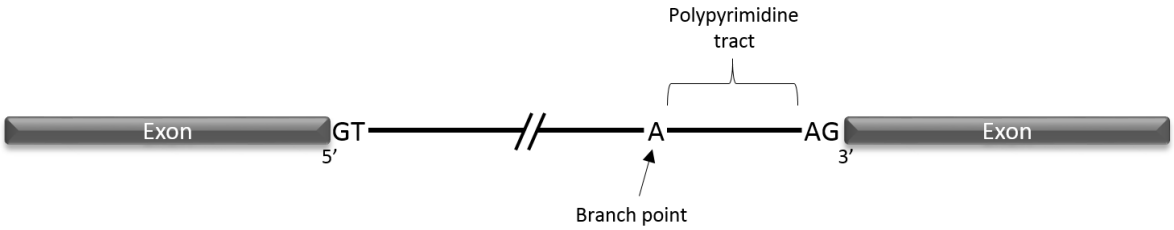


Figure 23 Schematic presentation of the splice sites in the canonical pattern, GT-AG.

1.4.1.2. Splicing regulatory sequences

In addition to splice site sequences, exons and introns are also recognized by other factors – splicing regulatory elements (SRE). These SRE are short degenerate sequences in the exons or introns, which are recognized by Serine/Arginine-rich protein (SR proteins) or heterogeneous nuclear ribonucleoproteins. Splicing regulatory sequences are divided into four categories based on location and function; SREs located in introns are called intronic splicing enhancers (ISE) or intronic splicing silencers (ISS), while SREs located in exons are called exonic splicing enhancers (ESE) or exonic splicing silencers (ESS), depending on their action. It is however difficult to establish if a SRE-binding protein might interfere or enhance inclusion of an exon, since this largely depend on placement in regards to splice sites [154].

The function of proteins that recognize SREs, is to help the spliceosome recognize the correct splice sites. The spliceosome can either be recruited by signaling over exons or over introns, which is called exon definition and intron definition, respectively. Exon definition relies on SR proteins recognizing ESE elements and subsequently recruiting the spliceosome components U1 and U2AF to the 5'ss and 3'ss at either side of the exon, respectively (Figure 24). Exon definition is common in mammalian genes, whereas intron definition is more common for the short introns found in lower eukaryotes [154].

1.4.1.3. The spliceosome

Humans have both a major and a minor spliceosome. The major spliceosome is an U2 snRNA-dependent spliceosome, which splices approximately 98% of introns (introns with the GT-AG pattern). The minor spliceosome is an U12 snRNA-dependent spliceosome, which splice a rare class of introns (introns with the AT-AC pattern). The components of the major spliceosome consists of the following snRNPs: U1, U2, U4, U5 and U6, while the minor spliceosome consists of U11, U12, $U4_{atac}$, U5 (common for both major and minor) and $U6_{atac}$ [152]. The following paragraphs describe the assembly and action of the major spliceosome.

The spliceosome assembly starts with the formation of the commitment complex, also called the **E-complex** (E for early). The E-complex consists of the U1 snRNP, the U2AF and the splicing factor 1 (SF1). The U1 snRNP consists of three SNRPs and the snRNA RNU1-1, also called U1. The U1 snRNP binds the 5' splice site (ss). The U2AF consists of two proteins and binds to the poly pyrimidine tract at the 3' end of the intron. Finally, the SF1, a single branch-point binding protein, binds to the BPS [154, 155].

After the E-complex is formed, it is converted to the **A-complex**. The A-complex prepares the BPS adenosine for the nucleophilic attack on the 5'ss. The conversion from the E-complex to the A-complex consists of the dissociation of SF1 while the snRNA of the U2 snRNP base pairs with the BPS (Figure 24A). This causes the adenosine to bulge out, ready to carry out the nucleophilic attack on the 5'ss [154]. The U2 snRNP consists of two SNRPs, three splicing factors, together with the snRNA RNU2-1, also called U2 [155].

Conversion to the **B-complex** requires the recruitment of the tri-snRNP, which contains the U4/U6 and U5 snRNPs (Figure 24B) [154]. The U4/U6 snRNP consists of the snRNA RNU4-1 (U4) and RNU6-1 (U6), together with three pre-mRNA processing factors and two other proteins. The U5 snRNP consists of the snRNA RNU5-1 (U5), together with seven proteins [155]. The maturation of the B-complex into the **B^{act}-complex** is preformed through several rearrangements within the spliceosome and the release of the U1 snRNP, together with almost all of the components of the U4/U6 snRNP. Although the catalytical center (U2-U6) is present, the spliceosome is not catalytically active since splice factors conceal the BPS. By destabilizing

these splice factors, the BPS is exposed and the **B*-complex** is formed, which is the first catalytically active complex. During step I catalysis, the 2'OH of the adenosine at the BPS carries out a nucleophilic attack on the 5'ss guanosine , yielding a 5' exon with a free 3'OH [154].

Following the first catalytical step is the **C-complex.** Here, the RNU5-1 base pairs with the 5' exon, the RNU6-1interacts with the 5'ss of the intron lariat and the RNU2-1 immobilizes the intron lariat. The C-complex stimulates conformational rearrangement of the catalytic site, leading to the **C*-complex**. The C*-complex initiates the second catalytical step, where the 3'OH of the 5' exon attacks the first nucleotide of the 3' exon [154].

This leads to the **P-complex** (P for post), where the two exons are assembled together (Figure 24C) and the lariat intron can be released.

In addition to the already described components of the spliceosome, numerous other factors are involved in this sophisticated process. These components can be found in the spliceosome database and at the HUGO Gene Nomenclature Committee website [155, 156].

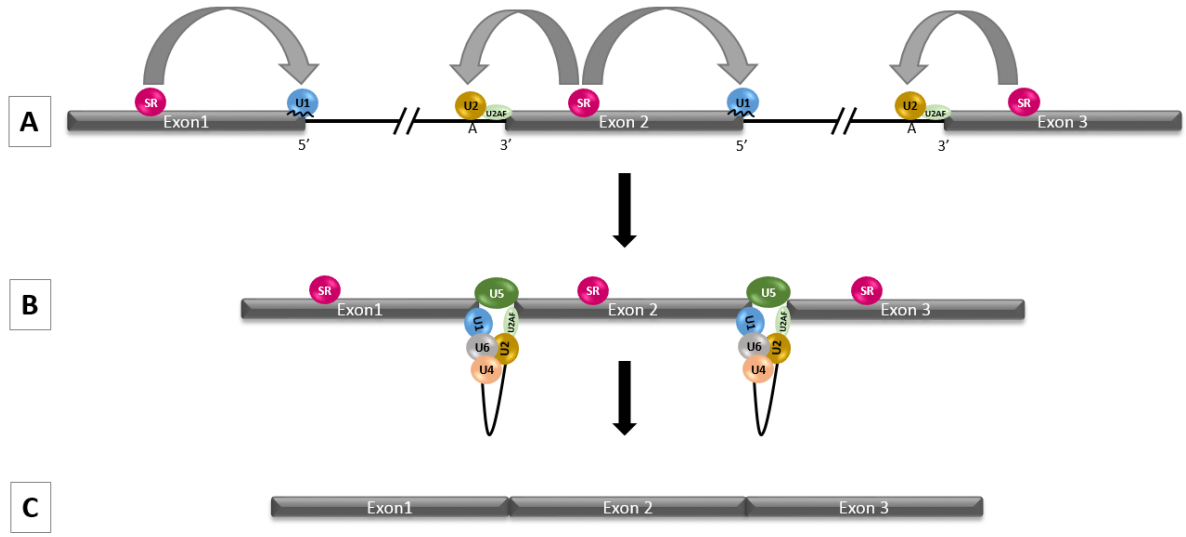


Figure 24 Assembly of the spliceosome on the pre-mRNA to the completed mRNA, through exon definition. Serine- and arginine-rich (SR) proteins recognize and binds ESE elements. A) The SR proteins facilitate recruitment and stabilization of the U1 snRNP and U2AF to the 5' and 3' splice sites, respectively. Together with the splice factor 1, this constitute the commitment complex (E-complex). B) The other snRNPs are subsequently recruited stepwise, to perform the removal of introns and C) joining of exons. Inspired from [157].

1.4.2. Alternative splicing

1.4.2.1. Naturally occurring alternative splicing in *BRCA1* and *BRCA2*

Naturally occurring splicing events are well documented for several genes. In the later years, several splicing events have been reported for both *BRCA1* and *BRCA2*. The Evidence-based Network of Germline Mutant Alleles (ENIGMA) has published two comprehensive articles on normal alternative splicing events of transcripts from *BRCA1* and *BRCA2* [158, 159]. For *BRCA1*, 63 splicing events were reported, where 10 were predominant events (Δ1q (GTAAAG), Δ5, Δ5q, Δ8p (CAG), Δ9, Δ9,10, Δ9-11, Δ11q and Δ13p (CAG)) [158]. For *BRCA2*, 24 different splicing events were reported, where four were predominant events (Δ3,

Δ6q,7, Δ12 and Δ17,18) [159]. Additionally, recently reported nanopore sequencing of full-length *BRCA1* mRNA transcripts revealed co-occurrence of some of these splicing events [160].

The documentation of naturally occurring splicing events is essential concerning evaluation of variants that might lead to aberrant splicing. Additionally, the information about naturally occurring splicing events is important for evaluation of nonsense sequence variants, as they may be rescued by normal alternative in-frame skipping, removing the nonsense variant. de la Hoya and colleagues (2016) reported that the normal alternative in-frame *BRCA1* Δ9,10 skipping may rescue the effect from nonsense sequence variants in exon 9 and 10 [161].

1.4.2.2. Alternative splicing and disease

While alternative splicing is a normal event occurring in several genes, dysregulation of splicing might lead to disease, which is well documented for *BRCA1* and *BRCA2* [111-115]. Aberrant splicing of pre-mRNA transcripts may produce frame-shifts, introducing a premature stop codon, and consequently potentially lead to nonsense-mediated mRNA decay (NMD). Nonsense-mediated mRNA decay is a degradation pathway that targets transcripts with premature termination codons [162], either produced by nonsense mutations or frame-shift mutations (including splicing affecting mutations leading to frame-shift). Another possibility for a pathogenic variant to affect the finished protein product by aberrant splicing is the removal of important coding regions, without leading to frame-shifts. This may e.g. influence folding of the protein, subcellular localization, binding properties or its enzymatic activity.

Variants in the +/- 1 and 2 positions (the GT-AG dinucleotides) of introns are recognized as variants affecting splicing. However, variants in both exons and other parts of the introns might lead to aberrant splicing, by affecting SREs or creating cryptic splice sites, e.g. *BRCA2* c.5434C>G that leads to exon skipping by affecting an ESE [163].

Evaluation of the effect of sequence variants on splicing can be performed in numerous ways. This prompted the need for common recommendations. In 2014, Whiley and colleagues compared different protocols across multiple laboratories to deduce common recommendations [164]. These recommendations are based on the evaluation of nine variants known to affect splicing, evaluated at 23 different participating laboratories. The deduced recommendations for best practice for mRNA splicing assay protocols include:

- NMD inhibitor: The need depends of sensitivity of detection method.
- RNA extraction: RNA extraction protocols were indistinguishable.
- DNase treatment: Recommended.
- cDNA synthesis primers: Gene-specific OR oligo(dT) + random hexamers.
- cDNA synthesis: Superscript reverse transcriptase is better for longer transcripts.
- PCR primers: Forward and reverse primers must be at least 1 whole exon 5' or 3' of variant, respectively.
- PCR conditions: Extension time long enough to copy amplicon, at least 30 cycles.
- Detection: Capillary electrophoresis (CE) was the most sensitive, followed by Qiaxcel (capillary electrophoresis system), and then sequencing and agarose gel electrophoresis.
- Sequencing: Cloning and sequencing is more sensitive than direct sequencing, but need to sequence at least 40 clones.

2. Aims

The overall goal of this project was to identify and investigate the possible consequences of pathogenic germline sequence variants leading to HBOC in a Norwegian cancer cohort.

The specific aims of this project were:

1. To advance the understanding of variants of unknown clinical significance (VUS) identified in the two common HBOC genes, *BRCA1* and *BRCA2*. By evaluation of their effect on splicing and their possible influence on the normal function of their gene products (Paper I).
2. To investigate how alternative transcripts are best preserved for further downstream cDNA studies by comparing two RNA collection systems. In addition, to develop a fragment analysis assay capable of investigating aberrant splicing of *BRCA1* as a possible screening tool for the identification of variants affecting normal splicing (Paper II).
3. To investigate the possible cancer cause in patients with no pathogenic sequence variants in *BRCA1/2*, an NGS study of genes included in the TruSight cancer panel from Illumina (Paper III).

3. Summary of papers

Paper I

Characterization of BRCA1 and BRCA2 variants found in a Norwegian breast or ovarian cancer cohort

Germline mutations in *BRCA1* and *BRCA2* cause hereditary breast and ovarian cancer (HBOC). Molecular screening of these two genes in patients with a family history of breast or ovarian cancer has revealed pathogenic variants as well as genetic variants of unknown significance (VUS). These VUS may cause a challenge in the genetic counseling process regarding clinical management of the patient and the family. In this study, we investigated 32 variants previously detected in 33 samples from patients with a family history of breast or ovarian cancer. cDNA was analyzed for alternative transcripts and selected missense variants located in the BRCT domains of *BRCA1* were assessed for their trans-activation ability.

An extensive cDNA analysis revealed that three of the 32 variants appeared to affect the splice-process (*BRCA1* c.213-5T>A, *BRCA1* c.5434C>G and *BRCA2* c.68-7T>A). In addition, two variants located in the BRCT domains of BRCA1 (c.5075A>C p.Asp1692Ala and c.5513T>G p.Val1838Gly) were shown to abolish the BRCT domain trans-activation ability, whereas *BRCA1* c.5125G>A (p.Gly1709Arg) exhibited equal trans-activation capability as the wild type domain. These functional studies may offer further insights into the pathogenicity of certain identified variants; however, this assay is only applicable for a subset of missense variants.

Paper II (Manuscript)

Comparing the quality of RNA preserved in PAXgene and Tempus Blood RNA tubes using *BRCA1* splicing events as a model system

Samples from patients with inherited breast and ovarian cancer (HBOC) are routinely analyzed for variants in *BRCA1* and *BRCA2*. In both these genes, variants affecting splicing have been associated with an increased risk of development of breast and ovarian cancer. To identify variants affecting pre-mRNA splicing it is important to ensure the preservation of RNA from the acquired samples. Comparison of different RNA preservation methods has been performed in several studies, but these studies mainly focused on the level of the main transcripts and not on the presence of different alternatively spliced transcripts. In this study, we wanted to compare RNA preservation and the detection of alternative spliced transcripts from PAXgene Blood RNA tube and Tempus Blood RNA tube collected blood samples. Blood samples were collected in parallel in the two different collection tubes from 48 anonymous blood donors at the University Hospital of North Norway. RNA concentrations, RNA Integrity Number (RIN) values and mRNA transcripts were evaluated. For the latter we used fragment analysis of *BRCA1* cDNA. *BRCA1* has earlier been demonstrated to produce many alternative spliced transcripts. We found that blood samples collected in Tempus tubes had a significantly higher average RNA yield (1.8-fold; $p=1.14\text{x}10^{-14}$) and RIN value (8.2 vs 7.7; $p=5.05\text{x}10^{-7}$) compared with blood samples collected in PAXgene tubes. However, less alternatively spliced fragments were identified in samples collected in Tempus tubes.

Paper III (Manuscript)

Identifying sequence variants contributing to hereditary breast/ovarian cancer in BRCA1/2 negative breast and ovarian cancer patients

Families with breast and ovarian cancer are often tested for disease associated sequence variants in *BRCA1* and *BRCA2*. Pathogenic sequence variants (PV) in these two genes are known to increase breast and ovarian cancer risks in females. However, in most families no pathogenic variants are detected in these two genes. Currently, several studies have identified other genes involved in hereditary breast and ovarian cancer (HBOC). To identify genetic risk factors for breast and ovarian cancer in a Norwegian HBOC cohort, 101 breast and/or ovarian cancer patients negative for pathogenic variants or variants of unknown clinical significance (VUS) in *BRCA1*/*BRCA2* were screened for PVs in 94 genes using next-generation sequencing (NGS). Sixteen genes were closely scrutinized. Ten different deleterious germline PV/likely pathogenic variants (LPVs) were identified in seven genes in 13 patients: three in *ATM*, two in *CHEK2* and one in *ERCC5, FANCM, RAD51C, TP53* and *WRN*. Six of the 13 patients also carried a VUS in another gene. In total, 30 different VUSs were identified and these require further characterization. For carriers of PV/LPV in many of these genes, there are no clinical management programs in Norway. The diversity of genetic risk factors possibly involved in cancer development show the necessity for more knowledge to improve the clinical follow-up of this genetically diverse patient group.

4. Patient/participant cohorts

4.1. Paper I

Blood samples, collected from an HBOC cohort at the Oslo University Hospital (OUS), were investigated for possible aberrant splicing. Additionally, novel variants located in the BRCT domains of BRCA1 were further investigated for their influence on the function of the BRCA1 protein.

Table 3 Sequence variants in BRCA1 and BRCA2 analyzed for their effect on splicing.

Sequence variant	Location	Protein	Patient
BRCA1 c.-20+521_-20+525delAAAAA	Intron 1	-	1
BRCA1 c.140G>T	Exon 5	p.Cys47Phe	2
BRCA1 c.213-5T>A	Intron 5	-	3
BRCA1 c.486G>T	Exon 8	p.= (p.Val162Val)	4
BRCA1 c.548-17G>T	Intron 8	-	5
BRCA1 c.734A>T	Exon 11	p.Asp245Val	6
BRCA1 c.1419C>T	Exon 11	p.= (p.Asn473Asn)	7
BRCA1 c.1487G>A	Exon 11	p.Arg496His	8
BRCA1 c.2521C>T	Exon 11	p.Arg841Trp	9
BRCA1 c.3418A>G	Exon 11	p.Ser1140Gly	10
BRCA1 c.3708T>G	Exon 11	p.Asn1236Lys	11
BRCA1 c.5075A>C	Exon 18	p.Asp1692Ala	12
BRCA1 c.5096G>A	Exon 18	p.Arg1699Gln	13
BRCA1 c.5117G>C	Exon 18	p.Gly1706Ala	9
BRCA1 c.5123C>T	Exon 18	p.Ala1708Val	14
BRCA1 c.5125G>A	Exon 18	p.Gly1709Arg	15
BRCA1 c.5434C>G	Exon 23	p.Pro1812Ala	16
BRCA1 c.5513T>G	Exon 24	p.Val1838Gly	17
BRCA2 c.40A>G	Exon 2	p.Ile14Val	18
BRCA2 c.68-7T>A	Intron 2	-	19
BRCA2 c.750G>A	Exon 9	p.= (p.Val250Val)	20, 33
BRCA2 c.2680G>A	Exon 11	p.Val894Ile	21
BRCA2 c.3568C>T	Exon 11	p.Arg1190Trp	22
BRCA2 c.4068G>A	Exon 11	p.= (p.Leu1356Leu)	23, 10
BRCA2 c.4828G>A	Exon 11	p.Val1610Met	24
BRCA2 c.5272_5274delAAT	Exon 11	p.Asn1758del	25
BRCA2 c.6100C>T	Exon 11	p.Arg2034Cys	26
BRCA2 c.6821G>T	Exon 11	p.Gly2274Val	27
BRCA2 c.7301A>C	Exon 14	p.Lys2434Thr	28
BRCA2 c.8177A>G	Exon 18	p.Tyr2726Cys	29
BRCA2 c.8323A>G	Exon 18	p.Met2775Val	30
BRCA2 c.9116C>T	Exon 23	p.Pro3039Leu	31, 32

4.2. Paper II

Participants were recruited from blood-donors at the blood bank at the University Hospital of North Norway (UNN). Healthy males and females donated blood in two parallel tubes: PAXgene Blood RNA tubes and Tempus Blood RNA tubes.

The study included blood from 24 female and 24 male participants, in total 96 samples.

4.3. Paper III

Patients for paper III were divided into three subgroups, based on how/when they were included in the study. All groups included patients with neither pathogenic sequence variants nor VUSs identified in *BRCA1* and *BRCA2*. All participants were diagnosed with breast and/or ovarian cancer. A total of 101 patients were included, from 93 unrelated families.

Group 1: Patients previously investigated using Sanger sequencing. Patients in this group were deceased. Accordingly, samples could be included without obtaining active or passive consent. This was in accordance with the decision of the regional ethics committee.

Group 2: Patients previously investigated using Sanger sequencing. This patient group had not been in contact with the department for some years. These patients were mailed an information leaflet about the study and the option to contact the department if they wanted to be excluded. None of the contacted patients opted not to be included in the current study.

Group 3: Patients previously investigated using the TruSight Cancer panel (Illumina) analyzed in a MiSeq instrument. Exclusively *BRCA1/2* had previously been analyzed, but the sequencing data for the whole TruSight cancer panel was readily available. Patients were mailed an information leaflet, a consent form and a pre-paid envelope for the return of the consent form. Exclusively patients who returned a signed consent form were included in the study.

Table 4 Patients included in the NGS study (Paper III). Group 1: P-1 – P32; Group 2: P-33 – P-78; Group 3: P-79 – P-101. Diagn. = diagnosis: includes first breast cancer (BC) or ovarian cancer (OC) diagnosis. Other diagn. = Other cancer diagnoses: includes BC and OC diagnosed after initial BC/OC diagnosis, together with other types of cancer diagnoses. Age at diagn.= age at diagnosis: includes age at first BC/OC diagnosis, age for "other diagnoses" are included in parentheses. CC = cervical cancer.

Patient	Diagn.	Other diagn.	Age at diagn.	Patient	Diagn.	Other diagn.	Age at diagn.
P-1	BC		36	P-51	BC		61
P-2	BC		57	P-52	OC		50
P-3	BC		63	P-53	BC		38
P-4	OC		60	P-54	OC		37
P-5	OC		55	P-55	OC		35
P-6	BC		65	P-56	BC		61
P-7	BC		50	P-57	BC		57
P-8	BC		56	P-58	Bilateral BC		57
P-9	BC		44	P-59	BC		58
P-10	BC	Ventricle cancer	55 (64)	P-60	BC		61
P-11	OC		53	P-61	OC		52
P-12	OC	Sarcoma (unspecified)	27 (61)	P-62	OC	Skin cancer (unspecified) + CC	38 (48 + 65)

Patient	Diagn.	Other diagn.	Age at diagn.	Patient	Diagn.	Other diagn.	Age at diagn.
P-13	BC		36	P-63	BC		25
P-14	BC		35	P-64	BC		35
P-15	Bilateral BC		74/80	P-65	Bilateral BC		51/51
P-16	OC		70	P-66	BC		53
P-17	BC		49	P-67	BC		51
P-18	OC	BC	67 (75)	P-68	OC	Bilateral BC	55 (58/58)
P-19	Bilateral BC		48/52	P-69	OC		52
P-20	BC		56	P-70	BC		53
P-21	OC		52	P-71	OC		48
P-22	BC		86	P-72	BC	Kidney cancer + Uterus cancer	48 (53 + 62)
P-23	BC		48	P-73	BC		50
P-24	OC		57	P-74	BC		54
P-25	BC		51	P-75	BC	Malign melanoma	(42) 49
P-26	Bilateral BC		37/47	P-76	BC		55
P-27	BC		56	P-77	BC		41
P-28	OC		54	P-78	BC		49
P-29	BC	Lung cancer	39 (58)	P-79	BC		55
P-30	OC		51	P-80	BC		58
P-31	OC		47	P-81	BC		64
P-32	OC		52	P-82	BC		37
P-33	BC	Colon cancer	(47) 57	P-83	BC		59
P-34	BC		42	P-84	BC		69
P-35	Bilateral BC	BC	41/42 (50)	P-85	BC		68
P-36	BC		57	P-86	BC		39
P-37	BC		61	P-87	BC		43
P-38	BC		48	P-88	BC		48
P-39	BC		51	P-89	BC		47
P-40	BC		59	P-90	BC		57
P-41	BC		69	P-91	BC		54
P-42	BC		63	P-92	BC		51
P-43	BC		48	P-93	BC		32
P-44	BC		49	P-94	BC		55
P-45	BC		63	P-95	Bilateral BC		57/62
P-46	BC		50	P-96	BC		59
P-47	BC		50	P-97	BC		50
P-48	BC		39	P-98	BC		53
P-49	BC		41	P-99	BC		52
P-50	BC		51	P-100	BC		52
				P-101	BC		53

5. Discussion

Breast cancer is the most commonly diagnosed cancer in women. Ovarian cancer is not as frequently diagnosed, but the outcome is more severe. The five-year relative survival is 48.7% for ovarian cancer, compared to 89.7% for breast cancer [9]. It is estimated that 5-10% of breast cancer cases and approximately 20-25% of ovarian cancer cases are due to inherited germline pathogenic sequence variants, which increase the lifetime risk of BOC (Table 1) [24, 25]. For example, pathogenic *BRCA1* variants increase the lifetime risk of breast cancer from 8.6%, found in women in the general population, up to 65% and ovarian cancer risk from 1.3% in the general population up to 59% for carriers of pathogenic sequence variants [9, 25, 26]. However, only in approximately 14% (median; ranging from 9.3 – 41.5%) of HBOC cases, pathogenic sequence variants are identified Furthermore, on average, approximately 45% of the identified pathogenic sequence variants are identified in other genes than the commonly investigated *BRCA1* and *BRCA2* [33, 35, 36, 39, 41, 165]. Identification of carriers of sequence variants conferring increased risk of breast and/or ovarian cancer is important, as these carriers may be offered a more intensive cancer surveillance program or prophylactic mastectomy and salpingo-oophorectomy in order to reduce the cancer risks drastically [27]. Therefore, the overall goal of this project was to identify and investigate the possible consequences of germline sequence variants leading to hereditary breast and ovarian cancer in a Norwegian cancer cohort.

Many of the sequence variants identified in BOC patients are VUS. To establish if patients with VUS in *BRCA1* or *BRCA2* (Table 3) had an increased cancer risk, we attempted to re-categorize the variants by combining in silico evaluation, mRNA studies and a functional assay (**Paper I**). For the evaluation of transcripts from the *BRCA1* and *BRCA2* genes, samples had been collected in PAXgene tubes. PAXgene tubes are part of the PAXgene Blood RNA System, which is intended for collection, transport and storage of blood and the stabilization of RNA. We speculated if the sample collection system could influence the preservation of alternative transcripts and consequently we compared PAXgene with a corresponding system, the Tempus Blood RNA system. For this evaluation, 96 samples were tested, using *BRCA1* as a model system (**Paper II**). Furthermore, since not only variants in *BRCA1* or *BRCA2* can cause increased risk of breast and ovarian cancer, we sequenced additional genes in 101 patients negative for pathogenic variants and VUSs in *BRCA1/2* (**Paper III**).

5.1. Paper I: Characterization of *BRCA1* and *BRCA2* variants

Even though *BRCA1* and *BRCA2* have been known since the mid 90's, several VUSs are still identified in these two genes. In an effort to re-categorize some of these VUSs, we evaluated their effect on splicing and their influence on the normal function of their associated gene products.

5.1.1. cDNA analysis

Sequence variants may influence the intended splicing of pre-mRNAs, by either masking original splice site, creating cryptic splice sites or changing SREs. In Paper I [115], 18 *BRCA1* and 14 *BRCA2* variants were investigated for their impact on splicing of pre-mRNA, using patient blood collected in PAXgene tubes. The study resulted in identification of three variants affecting splicing: the novel *BRCA1* c.213-5T>A, *BRCA1 c.5434C>G* and *BRCA2* c.68-7T>A.

The novel variant ***BRCA1* c.213-5T>A** was located in intron 5. In silico analysis tools predicted a reduction in the strength of the 3'ss and increased predictions for a cryptic 3'ss 59 nucleotides

upstream of the original site (Table 4 in paper I). The cDNA analysis revealed an inclusion of 59 nucleotides (Figure 1A in paper I), causing a frame-shift resulting in the introduction of a premature stop-codon after 75 codons (r.212_213ins213-59_213-1, p.Arg71Serfs*11). This variant could either activate the NMD pathway, and accordingly no protein would be generated from this allele, or alternatively, the variant could lead to the usage of a downstream initiation site at c.298-300 (original aa 100) [166]. The latter would result in a protein without the RING finger domain (aa 8-96) (Figure 12), responsible for interaction with BARD1, amongst others [95]. The effect of this variant on the protein was, however, not further investigated.

***BRCA1* c.5434C>G** in exon 23 resulted in skipping of exon 23 (r.5407_5467del p.Gly1803Glnfs*11) (Figure 1B in paper I), even though prediction programs suggested strengthening of a cryptic 3'ss at the c.5434 position (Table 4 in paper I). This variant was previously reported by Gaildrat et al. (2010), who demonstrated that the variant caused exon skipping, most likely by affecting an SRE [163]. The c.5434C>G variant thereby nicely demonstrates the importance of experimentally assessing the effect of exonic variants on splicing, which *ideally* should be done for all exonic variants with a low frequency in the general population.

***BRCA2* c.68-7T>A** in intron 2 resulted in increased skipping of exon 3 (Figure 1C in paper I) [115, 167, 168]. Exon 3 skipping is also common in samples from healthy controls and is one of the four predominant naturally occurring splicing events in *BRCA2* [159]. Exon 3 encodes most of the PALB2-binding region of BRCA2 (Figure 12), a region crucial for the recruitment of BRCA2 to DSB. However, this variant is considered to be a likely benign variant due to several reasons. Firstly, the exon skipping is incomplete and result in an in-frame deletion. Secondly, healthy controls also exhibit skipping of this exon, albeit at lower levels. Thirdly, the variant is present at a relatively high frequency in the Finnish population (gnomAD allele frequency: 0.66%). Fourthly, the *BRCA2* c.68-7T>A co-occurs in *trans* with other deleterious *BRCA2* variants [169, 170]. The variant is reported several times in ClinVar, where 17 of 21 submitters classify this variant as benign or likely benign (ClinVar ID: 52187). Unfortunately, 21 Norwegian women have endured mastectomy and salpingo-oophorectomy based on an incorrect interpretation of this variant [171]. However, Møller and Hovig (2017) disagree with our interpretation of the variant as likely benign and concluded that this variant is associated with breast cancer (although with a reduced penetrance) [172]. However, they did not consider all of the abovementioned arguments against pathogenicity. The authors also acknowledge that the segregation analysis in their cohort was inconclusive [172].

To further complicate the interpretation of variants affecting splicing of pre-mRNA, incomplete or 'leaky' skipping (normally spliced transcripts present) can still show tumor suppressor haplosufficiency [161]. De la Hoya et al. (2016) demonstrated how some apparently pathogenic high risk variants, exemplified by the in *cis BRCA1* c.[594-2A>C; 641A>G] that cause complete lack of full length transcripts from the mutant allele, may not be considered high risk variants after all. The above allele resulted in 20-30% transcripts with the in-frame Δ9,10 and tumor suppressor function. The authors argued that truncating variants in these two exons might not markedly increase the risk of BOC. They additionally suggested that truncating variants in other exons with similar results (residual 20-30% tumor suppressor functioning transcripts ('leaky' splicing)) might have the same effect [161].

In conclusion, great caution has to be taken when evaluating the effect variants may have on splicing. Several controls should be included to determine if there are naturally occurring alternatively spliced transcripts produced, albeit at lower levels. An example of this is the aforementioned *BRCA2* c.68-7T>A [115, 173].

5.1.2. Trans-activation assay

In paper I, we also performed a transcription activation assay on novel variants located in the two BRCT domains of *BRCA1* (Figure 12). Three variants were analyzed: p.Asp1692Ala (c.5075A>C), p.Gly1709Arg (c.5125G>A) and p.Val1838Gly (c.5513T>G).

By studying the expression and trans-activation of mutated BRCT domains in cell cultures, we demonstrated that both p.Asp1692Ala and p.Val1838Gly exhibited a clear loss of trans-activation activity in the *in vitro* assay (Figure 2B in paper I), which was not due to lack of protein expression (Figure 2C in paper I). Even though some of the prediction programs indicated pathogenicity for p.Gly1709Arg, it did not affect trans-activation activity in this assay.

Lee et al. (2010) [174] have previously validated this functional assay, and Woods et al. (2016) [175] have proposed that variants that abolish function in validated functional studies should at least be classified as likely pathogenic variants. However, by our criteria, this is not enough to re-classify the variant from VUS to likely pathogenic. Therefore, the two variants, p.Asp1692Ala and p.Val1838Gly, remained classified as VUS.

5.1.3. Additional observations

During this study, all exonic heterozygous positions (reported variants) were used as biallelic markers in the cDNA analysis to monitor if transcripts from both alleles were present. All samples tested were shown to express both alleles. However, with the applied technology we were unable to assess the relative quantity of individual transcripts.

Houdayer et al. (2012) proposed specific criteria for the selection of variants that should be investigated by mRNA analysis, using splicing prediction programs. For *BRCA1* and *BRCA2*, the criteria for inclusion is a predicted reduction of the original splice site value of more than 15% by MaxEntScan (MES) and more than 5% by SpliceSiteFinder-like (SSF). However, it is important that the originally predicted splice site has a predicted value of over three by MES and >60 by SSF [176]. Both *BRCA1* c.213-5T>A and *BRCA2* c.68-7T>A would have been included based on these criteria (Table 4 in paper I). However, the *BRCA1* c.5434C>G would have been omitted from cDNA analysis, since this variant most likely affects an SRE. Consequently, most variants affecting SRE would probably not be included following these criteria.

5.2. Paper II: Comparison of the PAXgene and Tempus Blood RNA systems

Experimental investigation of variants is important in the evaluation of their potential effect on splicing. In 2014, Whiley and colleagues compared different protocols for analysis of aberrant splicing at multiple laboratories to develop recommendations for best practice [164]. However, they did not assess if the use of the RNA collection and preservation systems PAXgene Blood RNA and Tempus Blood RNA may influence results, as these systems were not used by the participating laboratories. We speculated if the sample collection system could influence the preservation of alternative transcripts and accordingly compared the PAXgene and Tempus Blood RNA systems. Additionally, we wanted to develop a fragment analysis assay capable of investigating aberrant splicing of *BRCA1* as a possible screening tool for the identification of variants affecting normal splicing.

To investigate if collection systems influence preservation of alternative transcripts, we compared RNA/cDNA from blood from 48 participants. From each participant, blood was collected in parallels in PAXgene tubes and Tempus tubes (n=96). For evaluation of these collection methods, we measured RNA yield and RIN values (a measure of the integrity and level of degradation). Additionally, fragment analysis of *BRCA1* transcripts was performed. Both PAXgene and Tempus collection methods produced RNA of sufficient quality and yield for fragment analysis. Nevertheless, Tempus samples had a significantly higher average RNA yield (1.8-fold; $p=1.14x10^{-14}$) and RIN value (8.2 vs 7.7; $p=5.05x10^{-7}$). The higher RNA yield from blood collected in Tempus tubes were in concordance with previous findings (2.3-5.4 fold) [177-184], however, previous studies did not report significant differences in RIN values [181-184].

None of the studies that compared PAXgene and Tempus tubes had investigated if these collection methods affected the amount of alternatively spliced transcripts. We designed a fragment analysis assay, which covered the entire *BRCA1* transcript (Figure 1 and table 1 in paper II). More alternative transcripts were identified from RNA from blood collected in PAXgene tubes than from Tempus tubes (Table 2 in paper II). This might be a result of the significantly lower quality of RNA isolated from blood collected in PAXgene tubes, since the lower RIN values may indicate more fragmentation and therefore less templates for the full-length PCR products, which were used as internal controls.

Currently, no biological function has been attributed to the alternative transcripts of *BRCA1* (except BRCA1-IRIS [185], not investigated in this study and accordingly not discussed further). Accordingly, the difference identified in the comparison of PAXgene and Tempus collection systems are not of significance. The smaller amount of alternative transcripts seen in Tempus samples is, however, beneficial during the investigation of aberrant splicing, especially if utilizing Sanger sequencing to determine the exact consequences of splicing affecting variants. Samples with less background from naturally occurring alternatively spliced transcripts are easier to assess, as two different sequences are possible to read simultaneously, but three or more are becoming comprehensively harder, if not impossible. However, if some of the alternative transcripts have a biological effect, it is crucial that they are properly preserved. Consequently, the PAXgene system might be a better choice, if the reason for the increased number of alternative transcripts is due to better preservation of these less presented transcripts. Conversely, should the increased number of alternative transcripts be due to higher degradation of full-length transcripts, the Tempus system might be preferred.

The currently developed fragment analysis assay provides a screening tool for patients with a family history noticeably indicating high-risk variants in *BRCA1*, but for whom no pathogenic variants have been identified in *HBOC* associated genes. They may carry variants affecting splicing, but not detected in the routine screening of exons. Gambino et al. (2015) [186] have already investigated this in 13 women negative for pathogenic variants in *BRCA1* and *BRCA2*. Although their primer sets did not have the preferred overlap (at least one whole exon), they did detect aberrant splicing in two patients. Nonetheless, it appears that the aberrantly spliced transcripts represented only a limited fraction of the full-length transcripts.

5.3. Paper III: An extensive search for pathogenic variants in HBOC patients

During the last two years, several NGS studies of HBOC patients have been published [21, 33-36, 40, 41, 165]. This has resulted in the identification of causative variants for HBOC patients in several different genes. The identification of causative variants in several different genes has shown how diverse the landscape of genetic cancer risk factors is. From the identification of *BRCA1* and *BRCA2* in the mid 90's [22, 187], variants in genes coding for proteins involved in the Fanconi anemia pathway, the MMR pathway, cell signaling and others are now included (Table 1). However, the BOC risk associated with variants in most of the genes associated with these pathways remains elusive.

We investigated 101 patients (Table 4) diagnosed with breast and/or ovarian cancer and who had tested negative for pathogenic or uncertain variants in *BRCA1*/*BRCA2*. This was performed using a 94-gene panel, where sixteen genes were more closely scrutinized (Table 1, paper III), by evaluation of all types of sequence variants that passed filter settings (see Material and methods in paper III). Additionally, *BRCA1/BRCA1* were closely scrutinized, to confirm that no pathogenic or uncertain variants were present. The additional 76 genes included in the TruSight Cancer panel were investigated exclusively for the presence of frame-shift variants and nonsense variants. For the investigated 94 genes, on average 203.4 variants (range 170-247) were detected for each of the 101 patients. After filtration, on average 1.1 variant remained per patient, which resulted in 79 unique variants, of which 10 were classified as LPV/PV, 30 as VUS and 39 as likely benign/benign. The 10 LPV/PV were identified in seven genes in 13 patients: three in *ATM*, two in *CHEK2* and one in *ERCC5*, *FANCM*, *RAD51C*, TP53 and *WRN* (Table 3/Figure 2 in paper III).

In total, we identified LPV/PV in 12.9% of the investigated patients, which is in accordance with previous findings. Other studies with findings in *BRCA1/2*-negative patients reported PV/LPV in 7.7% of patients, on average [33-36, 40, 41, 165]. The results ranged from 4.6% - 15.1% of patients. Bias of ascertainment may be the reason for the large range in findings. Some of the blood samples included in our study were from deceased patients with a strong family history of cancer (group 1). Six of the LPV/PV were from these patients (n=32), resulting in findings in 18.8%. Other blood samples were from patients referred to the department several years ago and previously analyzed using Sanger Sequenced (group 2). Five of the LPV/PV were from these patients (n=46), resulting in findings in 10.9%. Samples from patients in the last group (group 3; n=23) are more likely be from breast or ovarian cancer cases lacking a strong family history. Two of the LPV/PV were from such patients, resulting in findings in only 4.4%. Nevertheless, the small sample size prevented a robust estimate of true disease associated allele frequencies in this population.

5.3.1. Variants identified in genes coding for cell cycle regulators

After *BRCA1* and *BRCA2*, the most commonly identified genes with pathogenic variants in HBOC cohorts are *ATM* and *CHEK2* [36, 39, 40]. Protein products from both *ATM* and *CHEK2* are involved in cell cycle regulation (Figure 6) by mediating DNA damage signaling (Figure 7), accordingly they play important roles as tumor suppressors. In paper III we identified four patients with pathogenic *ATM* variants: c.3245_3247delinsTGAT identified in two unrelated patients (P-2, P-31), c.5932G>T (P-91) and c.8432delA (P-62). Additionally, three patients were identified with likely pathogenic *CHEK2* variants: c.319+2T>A identified in two unrelated patients (P-12, P-16) and c.599T>C (P-59).

The ***ATM*** c.3245_3247delinsTGAT p.His1082Leufs*14 has previously been identified as a pathogenic variant (ClinVar: RCV000003172.7/RCV000159638.5) and is a Norwegian founder mutation [188, 189]. This variant was identified in two patients, one diagnosed with BC (P-2, 57y) and one with OC (P-31, 47y). Interestingly, the latter patient had a sister who was also included in the study (P-32) and equally diagnosed with OC (52y). However, the *ATM* variant was exclusively diagnosed in one of the sisters (P-31). Moreover, we found a novel VUS in *BRIP1* (c.2087C>T p.(Pro696Leu)) present in both sisters (discussed later).

The *ATM* c.5932G>T is predicted to produce a premature stop-codon (p.(Glu1978*)), however, the variant has previously been demonstrated to affect splicing of *ATM* (p.Ser1974Ilefs*4) [190]. The variant was identified in our cohort in a woman diagnosed with BC (54y) and has previously been associated with HBOC [191, 192].

The last variant, *ATM* c.8432delA was identified in a woman diagnosed with OC (38y) and has previously been identified as a pathogenic variant in Norwegian A-T patients [188, 193].

In general, heterozygous pathogenic variants in *ATM* are associated with 30% increased BC risk (Table 1), while there is currently no association with increased ovarian cancer risk [46]. It is therefore interesting that we identified pathogenic *ATM* variants in four patients, where two of them were diagnosed with OC (P-31, P-62). This could indicate that either the *ATM* variants in these patients are not the reason for their cancer development or that pathogenic *ATM* variants may increase ovarian cancer risk. Furthermore, it has been debated which types of pathogenic variants (Figure 1) in *ATM* that confer increased breast cancer risk. Some indicate that only a few, evolutionary rare, missense variants confer increased risk due to a dominant negative effect (protein without kinase activity) [75]. While others suggests that, also truncating variants influence cancer risk [73, 74, 194]. One way to investigate this is to sequence tumor tissue from patients identified with pathogenic germline variants, especially truncating variants.

The likely pathogenic germline ***CHEK2*** c.319+2T>A p.(?) was identified in two unrelated patients both diagnosed with OC (P-12, 27y; P-16, 70y). The variant has previously been identified in the Norwegian population, in a patient diagnosed with thyroid cancer (31y), BC (43y, 48y) and a family history of both BC and endometrial cancer [195].

The second *CHEK2* variant was the c.470T>C, p.(Ile157Thr), which is a well-documented likely pathogenic variant, associated with a marginal increased lifetime risk of breast cancer (18.2%) [46]. Since the variant has an allele frequency of 2.51% in the Finnish population, it is classified as a LPV (gnomAD; Table 3 in paper III). However, the high frequency might be in concordance with the low increase in breast cancer risk. In our study, the variant was identified in a patient diagnosed with BC (P-59, 58y).

On the opposite side of the scale of the low-/moderate-risk *ATM* and *CHEK2* variants, are pathogenic variants in ***TP53***. Pathogenic variants in this gene are linked to Li-Fraumeni syndrome [84]. Li-Fraumeni syndrome is a cancer predisposition syndrome, which predisposes to several cancer types, such as brain tumors, leukemias, pre-menopausal breast cancer and several others (see **section 1.3.1.4 *TP53***), with a 100% lifetime risk of cancer (Table 1) and a lifetime risk of breast cancer of 54% [43].

We identified *TP53* c.818G>A p.Arg23His in our patient cohort. This is one of the well-known mutational hotspot for *de novo* mutations in *TP53* (Figure 10) [196]. The variant was identified in 33% of the reads in our patient's blood (P-13). This is above the set threshold for reads for detection of variants, but skewed from the expected 50-50% ratio seen for germline variants. Either this is due to a technical artefact or it may be a somatically acquired sequence variant.

Somatic pathogenic sequence variants in *TP53* have been shown to increase in blood of women who endured chemotherapy treatment [197].

Our patient did not meet the classic clinical Li-Fraumeni syndrome (LFS) criteria, when diagnosed with BC (36y). However, pre-menopausal BC is categorized as a LFS tumor based on the Chompret criteria for *TP53* testing, however, these criteria also requires additional LFS tumor family history. The patients family history does not meet either the classical Li-Fraumeni criteria nor the revised Chompret criteria for LFS [84]. However, the patient's knowledge of the family history was sparse. With extended family examination, we would be able to determine if the mutation in *TP53* is an acquired or a germline variant.

5.3.2. HRR genes

In several NGS studies, pathogenic variants in ***PALB2*** are often detected and assigned third place after pathogenic findings in *ATM* and *CHEK2*. However, in our study, neither pathogenic, likely pathogenic nor uncertain variants were identified in *PALB2*. Overall, pathogenic variants in *PALB2* are identified in ≤1% of HBOC patients in several studies [35, 36, 39, 41]. Accordingly, the size of the currently investigated cohort might be too small to identify any LPV/PV in *PALB2* or it might be that pathogenic *PALB2* variants are not as prevalent in the Norwegian population. To investigate this, the sample size needs to be increased.

We did not identify any pathogenic, likely pathogenic or uncertain variants in ***RAD51D*** either. However, we did identify a likely pathogenic splice-affecting variant in ***RAD51C***, the c.1026+5_1026+7delGTA p.Arg322Serfs*22. This variant was identified in a woman diagnosed with OC (P-69, 52y) and had previously been reported by Janatova et al. (2015) [198]. No likely pathogenic or pathogenic variants were identified in ***BRIP1***. However, a VUS was identified in two sisters (P-31, P-32), both diagnosed with OC (discussed later).

5.3.3. Fanconi anemia genes

Fanconi anemia is a recessive disease characterized by congenital abnormalities, chromosome instability, progressive bone marrow failure and a strong predisposition to cancer. The FA genes code for proteins involved in the interstrand cross-link repair pathway. This repair results in DBS in one of the involved sister chromatids, which is subsequently repaired by the HRR (Figure 11). Accordingly, FA genes code for proteins overlapping with the HRR pathway (Table 2), it is therefore possible to assume that some FA genes, in addition to those coding for proteins involved in HRR, are involved in increased breast and ovarian cancer risk.

In our study, we identified the previously described ***FANCM*** c.5101C>T p.(Gln1701*) in two unrelated patients [199]. Although it is called a FA-gene and has a central role in the FA/DNA interstrand crosslink (ICL) repair pathway, biallelic variants in this gene have not been proven to cause FA syndrome. The only FA patient described with biallelic pathogenic variants in *FANCM* also harbored biallelic pathogenic variants in *FANCA* [200]. Several patients with homozygous pathogenic *FANCM* variants have been reported. However, these patients were not diagnosed with FA, but with early onset cancers, including breast cancer [201, 202].

We also identified a nonsense variant in *FANCF*, the c.1087C>T, p.(Gln363*), however, this variant is classified as a VUS and is further discussed under the section **5.3.6. VUS – a never-ending story.**

5.3.4. Genes coding for proteins involved in other cellular mechanisms

In the extended panel (the additional 76 genes of the Illumina cancer panel), we identified a PV and a LPV in *WRN* (c.1105C>T p.(Arg369*)) and *ERCC5 excision repair 5, endonuclease* (*ERCC5*) (c.67G>T p.(Glu23*)), respectively.

Biallelic pathogenic variants in the ***WRN*** gene are linked to Werner syndrome. Werner syndrome is a disease associated with premature aging symptoms, including cancers [203]. It is assumed that monoallelic pathogenic variants therefore confer an increased cancer risk [204]. Supporting this assumption, an NGS study of breast cancer patients has identified a deleterious *WRN* mutation (c.4245dupT, p.(Asp1416*)) [141]. In addition, Ding and colleagues (2007) have also reported an association between WRN and breast cancer [205].

BRCA1 has a WRN binding region (aa 452-1079) (Figure 12), and binding will stimulate both the WRN helicase and exonuclease activities. Furthermore, the amount of interacting BRCA1 and WRN increases in cells exposed to DNA cross-linking agents, inducing ICLs [101]. The ICLs are resolved by the FA pathway and subsequently the HRR. WRN also plays a role in Holliday junction resolution, which is an important last step of HRR [206].

In this study, we identified the *WRN* c.1105C>T p.(Arg396*) in a woman diagnosed with BC (P-90, 57y). The variant introduces and early stop codon and has previously been reported both in ClinVar and the Human Gene Mutation Database Professional (HGMDp) as a pathogenic variant and a disease mutation, respectively, causing Werner syndrome.

The identified ***ERCC5*** variant, c.67G>T p.(Glu23*), is a novel nonsense variant located in exon 1. Transcripts containing this variant most likely induce NMD. However, in the odd chance that NMD is not induced, several other outcomes have been evaluated; 1) the transcript is translated into a protein, using the newly formed stop-codon. This would, however, render the protein severely truncated, as it will lack most of the protein sequence. 2) *In silico* analysis shows that with the variant present, a new, cryptic 5'ss one nucleotide up-stream of the variant is predicted. The usage of this 5'ss would lead to a frame-shift and subsequently a new stop-codon (p.(Glu23Tyrfs*2)). 3) Some transcripts have an alternative initiation site downstream of the original initiation site, such as described for *BRCA1*. The first following methionine codon is Met169 in exon 5, however, the usage of this as a possible initiation site has not previously been reported.

Biallelic pathogenic variants in *ERCC5* are linked to Xeroderma pigmentosum complementation group G (XP-G), which is characterized by sun sensitivity of the skin and eyes, including increased risk of skin neoplasms [207]. The *ERCC5* c.67G>T variant was identified in a woman diagnosed with BC (P-44, 49y), if this variant is the cause of the patients BC is however unclear.

5.3.5. Negative results

Several NGS studies done on HBOC cohorts identified pathogenic variants in mismatch repair genes [34-36, 39]. This prompted us to investigate the four Lynch syndrome associated genes: *MSH2, MSH6, MLH1* and *PMS2*. These four genes encode proteins involved in the MMR pathway and pathogenic variants in them are known to increase the risk of developing colorectal cancer [139]. They have additionally been associated with increased risk of ovarian cancer, and an occasional breast cancer (Table 1). However, in our study, only VUSs were identified in these four genes (Table 4 in paper III).

Additionally, no sequence variants passed filter settings for ***PTEN*** and ***STK11***. Furthermore, no PV/LPV were identified in the ***CDH1*** gene. However, a *CDH1* VUS was identified in a patient diagnosed with OC (P-12, 27y), this variant was identified together with the previously discussed pathogenic *ATM* c.319+2T>A.

In total, 40 unique variants passed filter settings were classified as likely benign or benign. These are listed in the supplementary table S2 in paper III.

In 88% of the currently investigated patient cohort, no PV/LPV were identified (Figure 2 in paper III). As not all of the patients had a family history of BOC, it may be that many of these are patients with no increased BOC risk. However, for some of the patients it is possible (although not very likely) that negative results might stem from wrongly classified intermediate *BRCA1/2* variants, which we have not seen since they are removed by filter settings. Shimelis et al. (2017) have provided evidence that reduced BRCA2 function can be associated with an intermediate risk of breast cancer [108]. One of the variants they identified conferring an intermediate breast cancer risk was the *BRCA2* p.Tyr3035Ser. However, in ClinVar (Variant ID: 38211), none of the eight submitters have reported this variant as more than a VUS (per Feb 17, 2018), and even two well-known submitters, Ambry Genetics and Invitae, have reported the variant as likely benign [67]. This indicates that we still have a long way to go for the mapping of the mutational landscape in *BRCA1* and *BRCA2* is complete.

5.3.6. VUS – a never-ending story

Variants of unknown clinical significance pose an even larger problem as gene panels are becoming more widely used in a clinical setting. In our pilot study, we identified 30 variants classified as VUS. The largest number of VUS was identified in *ATM* (10) and *NF1* (7). Even though most of the VUSs probably are benign variants, some will need further evaluation.

In two sisters diagnosed with OC (P-31, 47y; P-32, 52y) a VUS in ***BRIP1*** was identified (c.2087C>T, Table 4 in paper III). P-31 was also a carrier of the pathogenic *ATM* variant c.3245_3247delinsTGAT (p.His1082Leufs*14, Table 3 in paper III). Pathogenic variants in *BRIP1* have been associated with increased risk of ovarian cancer, as opposed to pathogenic variants in *ATM*, a gene mainly associated with an increased risk of breast cancer (Table 1) [46, 66, 132, 208]. If the identified variant in *BRIP1* is the cause for the ovarian cancers in both sisters, remains to be investigated.

Another VUS that should be further investigated is the previously discussed ***FANCF*** c.1087C>T (p.(Gln363*)), a nonsense mutation located at the end of the only exon in that gene. If translated, the protein would lack only the last 12 amino acids. No pathogenic *FANCF* variants have been reported further 3'. Furthermore, this variant is probably not targeted by NMD since *FANCF* is a single-exon gene and NMD's mode of action requires exonic junction complexes [209]. The variant was classified as a VUS due to its location and the fact that no pathogenic variants had been reported in this region, as far as we know.

5.3.7. Limitations of the study

This study has revealed several pathogenic and likely pathogenic variants; however, there are some limitations to the study. The main limitation of the study is sample size, which currently prevents the detection and robust estimate of true disease associated allele frequencies in this population. However, future prospective studies will include additional patient samples.

The study was performed using the TruSight cancer panel from Illumina. Although several important HBOC associated genes are represented with exonic enrichment probes, there is still a limitation in the genes we could investigate. The *BARD1*, *RAD51* and possible *RBBP8* would have been interesting to investigate for germline pathogenic variants, but were not included in the TruSight cancer panel. Additionally, some regions were not covered by the included probes, for example most of the untranslated regions and some additional regions listed in table 1 in paper III.

Furthermore, the patients were anonymized after enrollment in the study, and we were only allowed by the ethical committee to access family history upon positive findings. However, it would have given the study more weight if the cancer history of all patients could have been included, also for the patients with negative results. The anonymization additionally prohibited us from investigating other tissues, which should have been done for the patient with the possibly somatic *TP53* variant c.818G>A p.Arg23His.

6. Concluding remarks

During this study, we have identified three variants that affect splicing: *BRCA1* c.213-5T>A, *BRCA1* c.5434C>G and *BRCA2* c.68-7T>A. *BRCA2* c.68-7T>A clearly demonstrated why caution should be taken when interpreting splice-affecting variants. The transactivation assay utilized in the study was a rapid method for the detection of the effect of variants in the BRCT domains of *BRCA1* and helped us demonstrate that two of the variants (*BRCA1* p.Asp1692Ala (c.5075A>C) and p.Val1838Gly (c.5513T>G)) abolished the transactivation activity (**Paper I**).

Additionally, we have demonstrated that RNA conserved in Tempus tubes resulted in a significantly higher yield and higher RIN values than RNA conserved in PAXgene tubes. However, both collection systems gave sufficient RNA yield and acceptable RIN values. Less alternatively spliced transcripts were present in cDNA synthesized from RNA conserved in Tempus tubes during fragment analysis. However, if this is beneficial or not remains to be elucidated (**Paper II**). We are currently planning to expand this study, including cultured blood, which will allow the use of NMD inhibitors. Additionally, the developed fragment analysis design should be tested on patients with a clear family history of BOC, but without an identified pathogenic variant in a HBOC gene.

In the last part of this study, we investigated a BOC cohort for PV/LPV in other genes than *BRCA1* and *BRCA2*. This study resulted in the identification of 13 patients with PV/LPV. The *ATM* and *CHEK2* genes seemed to be the most commonly mutated genes, however, their contribution to the cancer risk in the respective families is currently unclear. Furthermore, two of the pathogenic ATM variants were identified in patients diagnosed with OC, even though pathogenic variants in *ATM* variants are mainly associated with increased BC risk. In addition, the *FANCM* c.5101C>T was identified in two unrelated patients. The *FANCM* gene is not routinely investigated in the Norwegian cancer cohorts, accordingly there might be other patients carrying this variant. It would be interesting to screen numerous patients for this variant, to investigate if this is a widespread mutation. Furthermore, 30 VUS were identified, which will need further characterization (**Paper III**). We are currently planning to expand the cohort to a total of over 300 patients and investigate several of the VUS using functional assays.

7. References

1. Bianconi, E., et al., *An estimation of the number of cells in the human body.* Ann Hum Biol, 2013. **40**(6): p. 463-71.
2. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.
3. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.
4. Weinberg, R.A., *The biology of cancer*. 2007, New York: Garland Science.
5. Santarosa, M. and A. Ashworth, *Haploinsufficiency for tumour suppressor genes: when you don't need to go all the way.* Biochim Biophys Acta, 2004. **1654**(2): p. 105-22.
6. Clancy, S., *Genetic mutation*. 2008, Nature Education: [Internet].
7. Alberts, B., et al., *Molecular Biology of the Cell*. 5th ed. 2008, New York, USA: Garland Science
8. Knudson, A.G., Jr., *Mutation and cancer: statistical study of retinoblastoma.* Proc Natl Acad Sci U S A, 1971. **68**(4): p. 820-3.
9. Larsen, I., et al., *Cancer in Norway 2016 - Cancer incidence, mortality, survival and prevalence in Norway*, I.K. Larsen, Editor. 2017.
10. *Cancer research UK*. 2017 [cited 2017 1st of August]; Breast cancer symptoms]. Available from: https://www.cancerresearchuk.org/.
11. Ebell, M.H., M.B. Culp, and T.J. Radke, *A Systematic Review of Symptoms for the Diagnosis of Ovarian Cancer.* Am J Prev Med, 2016. **50**(3): p. 384-394.
12. Malhotra, G.K., et al., *Histological, molecular and functional subtypes of breast cancers.* Cancer Biol Ther, 2010. **10**(10): p. 955-60.
13. Hon, J.D., et al., *Breast cancer molecular subtypes: from TNBC to QNBC.* Am J Cancer Res, 2016. **6**(9): p. 1864-1872.
14. Klotz, D.M. and P. Wimberger, *Cells of origin of ovarian cancer: ovarian surface epithelium or fallopian tube?* Arch Gynecol Obstet, 2017. **296**(6): p. 1055-1062.
15. *Types & Stages of Ovarian Cancer*. [cited 2018 Jan 15; Available from: http://ovarian.org/about-ovarian-cancer/what-is-ovarian-cancer/types-a-stages.
16. Rosen, D.G., et al., *Ovarian cancer: pathology, biology, and disease models.* Front Biosci (Landmark Ed), 2009. **14**: p. 2089-102.
17. Krzystyniak, J., et al., *Epithelial ovarian cancer: the molecular genetics of epithelial ovarian cancer.* Ann Oncol, 2016. **27 Suppl 1**: p. i4-i10.
18. Kent, A., *Nuns and contraceptives.* Rev Obstet Gynecol, 2012. **5**(3-4): p. e166-7.
19. Hunt, K.K. and G.L. Robb, *Breast Cancer*. MD Anderson Cancer Care Series, ed. A.U. Buzdar and R.S. Freedman. 2001, New York: Springer Sciences+Business Media.
20. Walsh, T., et al., *Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing.* Proc Natl Acad Sci U S A, 2011. **108**(44): p. 18032-7.
21. Caminsky, N.G., et al., *Prioritizing Variants in Complete Hereditary Breast and Ovarian Cancer Genes in Patients Lacking Known BRCA Mutations.* Hum Mutat, 2016. **37**(7): p. 640-52.
22. Miki, Y., et al., *A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1.* Science, 1994. **266**(5182): p. 66-71.
23. Wooster, R., et al., *Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13.* Science, 1994. **265**(5181): p. 2088-90.

24. Kast, K., et al., *Prevalence of BRCA1/2 germline mutations in 21 401 families with breast and ovarian cancer.* J Med Genet, 2016. **53**(7): p. 465-71.
25. Nielsen, F.C., T. van Overeem Hansen, and C.S. Sorensen, *Hereditary breast and ovarian cancer: new genes in confined pathways.* Nat Rev Cancer, 2016. **16**(9): p. 599-612.
26. Mavaddat, N., et al., *Cancer risks for BRCA1 and BRCA2 mutation carriers: results from prospective analysis of EMBRACE.* J Natl Cancer Inst, 2013. **105**(11): p. 812-22.
27. NBCG, *Nasjonalt handlingsprogram med retningslinjer for diagnostikk, behandling og oppfølging av pasienter med brystkreft*. 2017 June, Helsedirektoratet: http://www.helsebiblioteket.no/retningslinjer/brystkreft/.
28. Kærn, J., et al., *Nasjonalt handlingsprogram med retningslinjer for gynekologisk kreft*. 2016 September, Helsedirektoratet: https://helsedirektoratet.no/retningslinjer.
29. Lek, M., et al., *Analysis of protein-coding genetic variation in 60,706 humans.* Nature, 2016. **536**(7616): p. 285-91.
30. Sherry, S.T., et al., *dbSNP: the NCBI database of genetic variation.* Nucleic Acids Res, 2001. **29**(1): p. 308-11.
31. Landrum, M.J., et al., *ClinVar: improving access to variant interpretations and supporting evidence.* Nucleic Acids Res, 2018. **46**(D1): p. D1062-D1067.
32. Stenson, P.D., et al., *The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies.* Hum Genet, 2017. **136**(6): p. 665-677.
33. Pinto, P., et al., *Implementation of next-generation sequencing for molecular diagnosis of hereditary breast and ovarian cancer highlights its genetic heterogeneity.* Breast Cancer Res Treat, 2016. **159**(2): p. 245-56.
34. Tedaldi, G., et al., *Multiple-gene panel analysis in a case series of 255 women with hereditary breast and ovarian cancer.* Oncotarget, 2017.
35. Kraus, C., et al., *Gene panel sequencing in familial breast/ovarian cancer patients identifies multiple novel mutations also in genes others than BRCA1/2.* Int J Cancer, 2017. **140**(1): p. 95-102.
36. Tung, N., et al., *Frequency of Germline Mutations in 25 Cancer Susceptibility Genes in a Sequential Series of Patients With Breast Cancer.* J Clin Oncol, 2016. **34**(13): p. 1460-8.
37. Aloraifi, F., et al., *Detection of novel germline mutations for breast cancer in non-BRCA1/2 families.* FEBS Journal, 2015. **282**(17): p. 3424-3437.
38. Susswein, L.R., et al., *Pathogenic and likely pathogenic variant prevalence among the first 10,000 patients referred for next-generation cancer panel testing.* Genet Med, 2015.
39. 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.
40. Yadav, S., et al., *Outcomes of retesting BRCA negative patients using multigene panels.* Fam Cancer, 2017. **16**(3): p. 319-328.
41. O'Leary, E., et al., *Expanded Gene Panel Use for Women With Breast Cancer: Identification and Intervention Beyond Breast Cancer Risk.* Ann Surg Oncol, 2017. **24**(10): p. 3060-3066.
42. Katsuki, Y. and M. Takata, *Defects in homologous recombination repair behind the human diseases: FA and HBOC.* Endocr Relat Cancer, 2016. **23**(10): p. T19-37.

43. Mai, P.L., et al., *Risks of first and subsequent cancers among TP53 mutation carriers in the National Cancer Institute Li-Fraumeni syndrome cohort.* Cancer, 2016. **122**(23): p. 3673-3681.
44. NCCN.org. *Genetic/Familial High-Risk Assessment: Breast and Ovarian*. NCCN Clinical Practice Guidelnes in Oncology (NCCN Guidelines) 2017 [cited 2018 30th January]; Available from: www.nccn.org/professionals/physician_gls/pdf/genetics_screening.pdf.
45. Antoniou, A.C., et al., *Breast-cancer risk in families with mutations in PALB2.* N Engl J Med, 2014. **371**(6): p. 497-506.
46. Tung, N., et al., *Counselling framework for moderate-penetrance cancer-susceptibility mutations.* Nat Rev Clin Oncol, 2016. **13**(9): p. 581-8.
47. Loveday, C., et al., *Germline mutations in RAD51D confer susceptibility to ovarian cancer.* Nat Genet, 2011. **43**(9): p. 879-882.
48. Goldberg, M., et al., *Association between the Lynch syndrome gene MSH2 and breast cancer susceptibility in a Canadian familial cancer registry.* J Med Genet, 2017. **54**(11): p. 742-746.
49. Engel, C., et al., *Risks of less common cancers in proven mutation carriers with lynch syndrome.* J Clin Oncol, 2012. **30**(35): p. 4409-15.
50. Provenzale, D., et al. *NCCN Guidelines. Genetic/Familial High-Risk Assessment: Colorectal, Version 2.2015*. 2017; Available from: www.nccn.org.
51. Bonadona, V., et al., *Cancer risks associated with germline mutations in MLH1, MSH2, and MSH6 genes in Lynch syndrome.* JAMA, 2011. **305**(22): p. 2304-10.
52. Harkness, E.F., et al., *Lynch syndrome caused by MLH1 mutations is associated with an increased risk of breast cancer: a cohort study.* J Med Genet, 2015. **52**(8): p. 553-6.
53. Tan, M.H., et al., *Lifetime cancer risks in individuals with germline PTEN mutations.* Clin Cancer Res, 2012. **18**(2): p. 400-7.
54. Hansford, S., et al., *Hereditary Diffuse Gastric Cancer Syndrome: CDH1 Mutations and Beyond.* JAMA Oncol, 2015. **1**(1): p. 23-32.
55. Lim, W., et al., *Relative frequency and morphology of cancers in STK11 mutation carriers.* Gastroenterology, 2004. **126**(7): p. 1788-94.
56. Uusitalo, E., et al., *Breast cancer in neurofibromatosis type 1: overrepresentation of unfavourable prognostic factors.* Br J Cancer, 2017. **116**(2): p. 211-217.
57. NGICG-CR, *Nasjonalt handlingsprogram med retningslinjer for diagnostikk, behandling og oppfølging av kreft i tykktarm og endetarm*. 2017 June, Helsedirektoratet: http://www.helsedirektoratet.no/publikasjoner/.
58. Asashima, M., et al., *The Cell Cycle*, in *A comprehensive approach to Life Science - web textbook*, S. Ishiura, et al., Editors. 2011, The University of Tokyo: [online].
59. Williams, G.J., S.P. Lees-Miller, and J.A. Tainer, *Mre11-Rad50-Nbs1 conformations and the control of sensing, signaling, and effector responses at DNA double-strand breaks.* DNA Repair (Amst), 2010. **9**(12): p. 1299-306.
60. Yates, A., et al., *Ensembl 2016. (release 89).* Nucleic Acids Res, 2016. **44**(D1): p. D710-6.
61. The UniProt, C., *UniProt: the universal protein knowledgebase.* Nucleic Acids Res, 2017. **45**(D1): p. D158-D169.
62. *Pfam protein domain database*. [cited 2017; Available from: http://pfam.xfam.org/.
63. Durocher, D. and S.P. Jackson, *The FHA domain.* FEBS Lett, 2002. **513**(1): p. 58-66.

64. You, Z., et al., *ATM activation and its recruitment to damaged DNA require binding to the C terminus of Nbs1.* Mol Cell Biol, 2005. **25**(13): p. 5363-79.
65. Varon, R., I. Demuth, and K.H. Chrzanowska. *Nijmegen Breakage Syndrome*. [Internet] 1999 May 17 [updated 2017 Feb 2]; In: Adam MP, Ardinger HH, Pagon RA, et al., editors. GeneReviews®. Seattle (WA): University of Washington, Seattle; 1993-2017 [Available from: https://www.ncbi.nlm.nih.gov/books/NBK1176/.
66. Ramus, S.J., et al., *Germline Mutations in the BRIP1, BARD1, PALB2, and NBN Genes in Women With Ovarian Cancer.* J Natl Cancer Inst, 2015. **107**(11).
67. Landrum, M.J., et al., *ClinVar: public archive of interpretations of clinically relevant variants.* Nucleic Acids Res, 2016. **44**(D1): p. D862-8.
68. Safran, M., et al., *GeneCards Version 3: the human gene integrator.* Database (Oxford), 2010. **2010**: p. baq020.
69. Young, D.B., et al., *Identification of domains of ataxia-telangiectasia mutated required for nuclear localization and chromatin association.* J Biol Chem, 2005. **280**(30): p. 27587-94.
70. Cremona, C.A. and A. Behrens, *ATM signalling and cancer.* Oncogene, 2014. **33**(26): p. 3351-60.
71. Falck, J., J. Coates, and S.P. Jackson, *Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage.* Nature, 2005. **434**(7033): p. 605-11.
72. Athma, P., R. Rappaport, and M. Swift, *Molecular genotyping shows that ataxia-telangiectasia heterozygotes are predisposed to breast cancer.* Cancer Genet Cytogenet, 1996. **92**(2): p. 130-4.
73. Goldgar, D.E., et al., *Rare variants in the ATM gene and risk of breast cancer.* Breast Cancer Res, 2011. **13**(4): p. R73.
74. Hollestelle, A., et al., *Discovering moderate-risk breast cancer susceptibility genes.* Curr Opin Genet Dev, 2010. **20**(3): p. 268-76.
75. Tavtigian, S.V., et al., *Rare, evolutionarily unlikely missense substitutions in ATM confer increased risk of breast cancer.* Am J Hum Genet, 2009. **85**(4): p. 427-46.
76. Nevanlinna, H. and J. Bartek, *The CHEK2 gene and inherited breast cancer susceptibility.* Oncogene, 2006. **25**(43): p. 5912-9.
77. Lee, C.H. and J.H. Chung, *The hCds1 (Chk2)-FHA domain is essential for a chain of phosphorylation events on hCds1 that is induced by ionizing radiation.* J Biol Chem, 2001. **276**(32): p. 30537-41.
78. Zannini, L., D. Delia, and G. Buscemi, *CHK2 kinase in the DNA damage response and beyond.* J Mol Cell Biol, 2014. **6**(6): p. 442-57.
79. Lane, D.P., *p53, guardian of the genome.* Nature, 1992. **358**(6381): p. 15-6.
80. Vousden, K.H. and X. Lu, *Live or let die: the cell's response to p53.* Nat Rev Cancer, 2002. **2**(8): p. 594-604.
81. Joerger, A.C. and A.R. Fersht, *The tumor suppressor p53: from structures to drug discovery.* Cold Spring Harb Perspect Biol, 2010. **2**(6): p. a000919.
82. Natan, E., et al., *Interaction of the p53 DNA-binding domain with its n-terminal extension modulates the stability of the p53 tetramer.* J Mol Biol, 2011. **409**(3): p. 358-68.
83. Wells, M., et al., *Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain.* Proc Natl Acad Sci U S A, 2008. **105**(15): p. 5762-7.
84. Schneider, K., et al. *Li-Fraumeni Syndrome*. [Internet] 1999 Jan 19 [updated 2013 Apr 11]; In: Adam MP, Ardinger HH, Pagon RA, et al., editors. GeneReviews®.

Seattle (WA): University of Washington, Seattle; 1993-2017 [Available from: https://www.ncbi.nlm.nih.gov/books/NBK1311/].
85. Zhang, Y., et al., *TP53 mutations in epithelial ovarian cancer.* Translational Cancer Research, 2016. **5**(6): p. 650-663.
86. Bekker-Jensen, S. and N. Mailand, *Assembly and function of DNA double-strand break repair foci in mammalian cells.* DNA Repair (Amst), 2010. **9**(12): p. 1219-28.
87. Sartori, A.A., et al., *Human CtIP promotes DNA end resection.* Nature, 2007. **450**(7169): p. 509-14.
88. Yu, X. and J. Chen, *DNA damage-induced cell cycle checkpoint control requires CtIP, a phosphorylation-dependent binding partner of BRCA1 C-terminal domains.* Mol Cell Biol, 2004. **24**(21): p. 9478-86.
89. Wang, H., et al., *The interaction of CtIP and Nbs1 connects CDK and ATM to regulate HR-mediated double-strand break repair.* PLoS Genet, 2013. **9**(2): p. e1003277.
90. Caestecker, K.W. and G.R. Van de Walle, *The role of BRCA1 in DNA double-strand repair: past and present.* Exp Cell Res, 2013. **319**(5): p. 575-87.
91. Buisson, R. and J.Y. Masson, *PALB2 self-interaction controls homologous recombination.* Nucleic Acids Res, 2012. **40**(20): p. 10312-23.
92. Bishop, A.J. and R.H. Schiestl, *Homologous Recombination and Its Role in Carcinogenesis.* J Biomed Biotechnol, 2002. **2**(2): p. 75-85.
93. Huen, M.S., S.M. Sy, and J. Chen, *BRCA1 and its toolbox for the maintenance of genome integrity.* Nat Rev Mol Cell Biol, 2010. **11**(2): p. 138-48.
94. Prakash, R., et al., *Homologous recombination and human health: the roles of BRCA1, BRCA2, and associated proteins.* Cold Spring Harb Perspect Biol, 2015. **7**(4): p. a016600.
95. Narod, S.A. and W.D. Foulkes, *BRCA1 and BRCA2: 1994 and beyond.* Nat Rev Cancer, 2004. **4**(9): p. 665-76.
96. Korlimarla, A., et al., *Identification of a non-canonical nuclear localization signal (NLS) in BRCA1 that could mediate nuclear localization of splice variants lacking the classical NLS.* Cell Mol Biol Lett, 2013. **18**(2): p. 284-96.
97. Henderson, B.R., *Regulation of BRCA1, BRCA2 and BARD1 intracellular trafficking.* Bioessays, 2005. **27**(9): p. 884-93.
98. Chen, C.F., et al., *The nuclear localization sequences of the BRCA1 protein interact with the importin-alpha subunit of the nuclear transport signal receptor.* J Biol Chem, 1996. **271**(51): p. 32863-8.
99. Millot, G.A., et al., *A guide for functional analysis of BRCA1 variants of uncertain significance.* Hum Mutat, 2012. **33**(11): p. 1526-37.
100. Yu, X., et al., *The BRCT domain is a phospho-protein binding domain.* Science, 2003. **302**(5645): p. 639-42.
101. Cheng, W.H., et al., *Collaboration of Werner syndrome protein and BRCA1 in cellular responses to DNA interstrand cross-links.* Nucleic Acids Res, 2006. **34**(9): p. 2751-60.
102. Spurdle, A.B., et al., *BRCA1 R1699Q variant displaying ambiguous functional abrogation confers intermediate breast and ovarian cancer risk.* J Med Genet, 2012. **49**(8): p. 525-32.
103. Domchek, S.M., et al., *Biallelic deleterious BRCA1 mutations in a woman with early-onset ovarian cancer.* Cancer Discov, 2013. **3**(4): p. 399-405.
104. Sawyer, S.L., et al., *Biallelic mutations in BRCA1 cause a new Fanconi anemia subtype.* Cancer Discov, 2015. **5**(2): p. 135-42.
105. *The Breast Cancer Information Core database*. [[internet]] [cited 2018 Mar 03.

106. Xia, B., et al., *Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2.* Mol Cell, 2006. **22**(6): p. 719-29.
107. Pellegrini, L., et al., *Insights into DNA recombination from the structure of a RAD51-BRCA2 complex.* Nature, 2002. **420**(6913): p. 287-93.
108. Shimelis, H., et al., *BRCA2 Hypomorphic Missense Variants Confer Moderate Risks of Breast Cancer.* Cancer Res, 2017. **77**(11): p. 2789-2799.
109. Spain, B.H., et al., *Truncated BRCA2 is cytoplasmic: implications for cancer-linked mutations.* Proc Natl Acad Sci U S A, 1999. **96**(24): p. 13920-5.
110. Yano, K., et al., *Nuclear localization signals of the BRCA2 protein.* Biochem Biophys Res Commun, 2000. **270**(1): p. 171-5.
111. Thomassen, M., et al., *Characterization of BRCA1 and BRCA2 splicing variants: a collaborative report by ENIGMA consortium members.* Breast Cancer Res Treat, 2012. **132**(3): p. 1009-23.
112. Steffensen, A.Y., et al., *Functional characterization of BRCA1 gene variants by mini-gene splicing assay.* Eur J Hum Genet, 2014. **22**(12): p. 1362-8.
113. Quiles, F., et al., *Investigating the effect of 28 BRCA1 and BRCA2 mutations on their related transcribed mRNA.* Breast Cancer Res Treat, 2016. **155**(2): p. 253-60.
114. 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 Genet, 2017. **13**(3): p. e1006691.
115. Jarhelle, E., et al., *Characterization of BRCA1 and BRCA2 variants found in a Norwegian breast or ovarian cancer cohort.* Fam Cancer, 2017. **16**(1): p. 1-16.
116. Oliver, A.W., et al., *Structural basis for recruitment of BRCA2 by PALB2.* EMBO Rep, 2009. **10**(9): p. 990-6.
117. Sy, S.M., M.S. Huen, and J. Chen, *PALB2 is an integral component of the BRCA complex required for homologous recombination repair.* Proc Natl Acad Sci U S A, 2009. **106**(17): p. 7155-60.
118. Bleuyard, J.Y., et al., *ChAM, a novel motif that mediates PALB2 intrinsic chromatin binding and facilitates DNA repair.* EMBO Rep, 2012. **13**(2): p. 135-41.
119. Buisson, R., et al., *Breast cancer proteins PALB2 and BRCA2 stimulate polymerase eta in recombination-associated DNA synthesis at blocked replication forks.* Cell Rep, 2014. **6**(3): p. 553-64.
120. Tischkowitz, M. and B. Xia, *PALB2/FANCN: recombining cancer and Fanconi anemia.* Cancer Res, 2010. **70**(19): p. 7353-9.
121. Miller, K.A., et al., *Domain mapping of the Rad51 paralog protein complexes.* Nucleic Acids Res, 2004. **32**(1): p. 169-78.
122. Chun, J., E.S. Buechelmaier, and S.N. Powell, *Rad51 paralog complexes BCDX2 and CX3 act at different stages in the BRCA1-BRCA2-dependent homologous recombination pathway.* Mol Cell Biol, 2013. **33**(2): p. 387-95.
123. Vaz, F., et al., *Mutation of the RAD51C gene in a Fanconi anemia-like disorder.* Nat Genet, 2010. **42**(5): p. 406-9.
124. Meindl, A., et al., *Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene.* Nat Genet, 2010. **42**(5): p. 410-4.
125. Easton, D.F., et al., *Gene-panel sequencing and the prediction of breast-cancer risk.* N Engl J Med, 2015. **372**(23): p. 2243-57.
126. Loveday, C., et al., *Germline mutations in RAD51D confer susceptibility to ovarian cancer.* Nat Genet, 2011. **43**(9): p. 879-82.

127. Cantor, S.B., et al., *BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function.* Cell, 2001. **105**(1): p. 149-60.
128. Brosh, R.M., Jr. and S.B. Cantor, *Molecular and cellular functions of the FANCJ DNA helicase defective in cancer and in Fanconi anemia.* Front Genet, 2014. **5**: p. 372.
129. Xie, J., et al., *An MLH1 mutation links BACH1/FANCJ to colon cancer, signaling, and insight toward directed therapy.* Cancer Prev Res (Phila), 2010. **3**(11): p. 1409-16.
130. Levitus, M., et al., *Heterogeneity in Fanconi anemia: evidence for 2 new genetic subtypes.* Blood, 2004. **103**(7): p. 2498-503.
131. Levitus, M., et al., *The DNA helicase BRIP1 is defective in Fanconi anemia complementation group J.* Nat Genet, 2005. **37**(9): p. 934-5.
132. Norquist, B.M., et al., *Inherited Mutations in Women With Ovarian Carcinoma.* JAMA Oncol, 2016. **2**(4): p. 482-90.
133. Makharashvili, N. and T.T. Paull, *CtIP: A DNA damage response protein at the intersection of DNA metabolism.* DNA Repair (Amst), 2015. **32**: p. 75-81.
134. Garcia-Rubio, M.L., et al., *The Fanconi Anemia Pathway Protects Genome Integrity from R-loops.* PLoS Genet, 2015. **11**(11): p. e1005674.
135. Bluteau, D., et al., *Biallelic inactivation of REV7 is associated with Fanconi anemia.* J Clin Invest, 2016. **126**(9): p. 3580-4.
136. Li, G.M., *Mechanisms and functions of DNA mismatch repair.* Cell Res, 2008. **18**(1): p. 85-98.
137. Pena-Diaz, J. and J. Jiricny, *Mammalian mismatch repair: error-free or error-prone?* Trends Biochem Sci, 2012. **37**(5): p. 206-14.
138. Jena, N.R. and M. Bansal, *Mutagenicity associated with O6-methylguanine-DNA damage and mechanism of nucleotide flipping by AGT during repair.* Phys Biol, 2011. **8**(4): p. 046007.
139. Kohlmann, W. and S.B. Gruber. *Lynch Syndrome*. [Internet] 2004 Feb 5 [updated 2018 Feb 1]; In: Adam MP, Ardinger HH, Pagon RA, et al., editors. GeneReviews®. Seattle (WA): University of Washington, Seattle; 1993-2017 [Available from: https://www.ncbi.nlm.nih.gov/books/NBK1211/.
140. *National Library of Medicine (US). Genetics Home Reference [Internet]*. 2017 Sep 5; Available from: https://ghr.nlm.nih.gov/condition/lynch-syndrome.
141. Yang, X., et al., *Identification of a comprehensive spectrum of genetic factors for hereditary breast cancer in a Chinese population by next-generation sequencing.* PLoS One, 2015. **10**(4): p. e0125571.
142. Warren, J.J., et al., *Structure of the human MutSalpha DNA lesion recognition complex.* Mol Cell, 2007. **26**(4): p. 579-92.
143. Edelbrock, M.A., S. Kaliyaperumal, and K.J. Williams, *Structural, molecular and cellular functions of MSH2 and MSH6 during DNA mismatch repair, damage signaling and other noncanonical activities.* Mutat Res, 2013. **743-744**: p. 53-66.
144. Guerrette, S., S. Acharya, and R. Fishel, *The interaction of the human MutL homologues in hereditary nonpolyposis colon cancer.* J Biol Chem, 1999. **274**(10): p. 6336-41.
145. Apweiler, R., et al., *UniProt: the Universal Protein knowledgebase.* Nucleic Acids Res, 2004. **32**(Database issue): p. D115-9.
146. Hayward, B.E., et al., *Extensive gene conversion at the PMS2 DNA mismatch repair locus.* Hum Mutat, 2007. **28**(5): p. 424-30.
147. Nakagawa, H., et al., *Mismatch repair gene PMS2: disease-causing germline mutations are frequent in patients whose tumors stain negative for PMS2 protein,*

but paralogous genes obscure mutation detection and interpretation. Cancer Res, 2004. **64**(14): p. 4721-7.
148. Gupta, R., et al., *LKB1 preserves genome integrity by stimulating BRCA1 expression.* Nucleic Acids Res, 2015. **43**(1): p. 259-71.
149. Shackelford, D.B. and R.J. Shaw, *The LKB1-AMPK pathway: metabolism and growth control in tumour suppression.* Nat Rev Cancer, 2009. **9**(8): p. 563-75.
150. McGarrity, T.J., C.I. Amos, and M.J. Baker. *Peutz-Jeghers Syndrome*. [Internet] 2001 Feb 23 [updated 2016 Jul 14]; In: Adam MP, Ardinger HH, Pagon RA, et al., editors. GeneReviews®. Seattle (WA): University of Washington, Seattle; 1993-2017 [Available from: https://www.ncbi.nlm.nih.gov/books/NBK1266/.
151. Friedman, J. *Neurofibromatosis 1*. [Internet] 1998 Oct 2 [updated 2018 Jan 11]; In: Adam MP, Ardinger HH, Pagon RA, et al., editors. GeneReviews®. Seattle (WA): University of Washington, Seattle; 1993-2017 [Available from: https://www.ncbi.nlm.nih.gov/books/NBK1109/.
152. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.
153. Brugiolo, M., L. Herzel, and K.M. Neugebauer, *Counting on co-transcriptional splicing.* F1000Prime Rep, 2013. **5**: p. 9.
154. Herzel, L., et al., *Splicing and transcription touch base: co-transcriptional spliceosome assembly and function.* Nat Rev Mol Cell Biol, 2017. **18**(10): p. 637-650.
155. *Gene Family: Major spliceosome*. [[internet]] [cited 2018 Feb 14]; Available from: https://www.genenames.org/cgi-bin/genefamilies/set/1518.
156. Cvitkovic, I. and M.S. Jurica, *Spliceosome database: a tool for tracking components of the spliceosome.* Nucleic Acids Res, 2013. **41**(Database issue): p. D132-41.
157. Ast, G., *How did alternative splicing evolve?* Nat Rev Genet, 2004. **5**(10): p. 773-82.
158. Colombo, M., et al., *Comprehensive annotation of splice junctions supports pervasive alternative splicing at the BRCA1 locus: a report from the ENIGMA consortium.* Hum Mol Genet, 2014. **23**(14): p. 3666-80.
159. Fackenthal, J.D., et al., *Naturally occurring BRCA2 alternative mRNA splicing events in clinically relevant samples.* J Med Genet, 2016.
160. de Jong, L.C., et al., *Nanopore sequencing of full-length BRCA1 mRNA transcripts reveals co-occurrence of known exon skipping events.* Breast Cancer Res, 2017. **19**(1): p. 127.
161. de la Hoya, M., et al., *Combined genetic and splicing analysis of BRCA1 c.[594-2A>C; 641A>G] highlights the relevance of naturally occurring in-frame transcripts for developing disease gene variant classification algorithms.* Hum Mol Genet, 2016.
162. Hug, N., D. Longman, and J.F. Caceres, *Mechanism and regulation of the nonsense-mediated decay pathway.* Nucleic Acids Res, 2016. **44**(4): p. 1483-95.
163. Gaildrat, P., et al., *The BRCA1 c.5434C->G (p.Pro1812Ala) variant induces a deleterious exon 23 skipping by affecting exonic splicing regulatory elements.* J Med Genet, 2010. **47**(6): p. 398-403.
164. Whiley, P.J., et al., *Comparison of mRNA splicing assay protocols across multiple laboratories: recommendations for best practice in standardized clinical testing.* Clin Chem, 2014. **60**(2): p. 341-52.
165. Singh, J., et al., *Screening of over 1000 Indian patients with breast and/or ovarian cancer with a multi-gene panel: prevalence of BRCA1/2 and non-BRCA mutations.* Breast Cancer Res Treat, 2018.

166. Buisson, M., et al., *The 185delAG mutation (c.68_69delAG) in the BRCA1 gene triggers translation reinitiation at a downstream AUG codon.* Hum Mutat, 2006. **27**(10): p. 1024-9.
167. Vreeswijk, M.P., et al., *Intronic variants in BRCA1 and BRCA2 that affect RNA splicing can be reliably selected by splice-site prediction programs.* Hum Mutat, 2009. **30**(1): p. 107-14.
168. Sanz, D.J., et al., *A high proportion of DNA variants of BRCA1 and BRCA2 is associated with aberrant splicing in breast/ovarian cancer patients.* Clin Cancer Res, 2010. **16**.
169. Muller, D., et al., *An entire exon 3 germ-line rearrangement in the BRCA2 gene: pathogenic relevance of exon 3 deletion in breast cancer predisposition.* BMC Med Genet, 2011. **12**: p. 121.
170. Colombo, M., et al., *The BRCA2 c.68-7T > A variant is not pathogenic: A model for clinical calibration of spliceogenicity.* Hum Mutat, 2018.
171. Storvik, A.G., *21 kvinner fikk fjernet bryst og eggstokker - skulle ikke vært operert*, in *Dagens Medisin*. 2017: [internet].
172. Møller, P. and E. Hovig, *The BRCA2 variant c.68-7 T>A is associated with breast cancer.* Hereditary Cancer in Clinical Practice, 2017. **15**(1): p. 20.
173. Moller, P. and E. Hovig, *The BRCA2 variant c.68-7 T>A is associated with breast cancer.* Hered Cancer Clin Pract, 2017. **15**: p. 20.
174. Lee, M.S., et al., *Comprehensive analysis of missense variations in the BRCT domain of BRCA1 by structural and functional assays.* Cancer Res, 2010. **70**(12): p. 4880-90.
175. Woods, N.T., et al., *Functional assays provide a robust tool for the clinical annotation of genetic variants of uncertain significance.* npj Genomic Medicine, 2016. **1**: p. 16001.
176. Houdayer, C., et al., *Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants.* Hum Mutat, 2012. **33**(8): p. 1228-38.
177. Asare, A.L., et al., *Differential gene expression profiles are dependent upon method of peripheral blood collection and RNA isolation.* BMC Genomics, 2008. **9**: p. 474.
178. Meyer, A., et al., *Evaluation of Existing Methods for Human Blood mRNA Isolation and Analysis for Large Studies.* PLoS One, 2016. **11**(8): p. e0161778.
179. Skogholt, A.H., et al., *Gene expression differences between PAXgene and Tempus blood RNA tubes are highly reproducible between independent samples and biobanks.* BMC Res Notes, 2017. **10**(1): p. 136.
180. Nikula, T., et al., *Genome-wide comparison of two RNA-stabilizing reagents for transcriptional profiling of peripheral blood.* Transl Res, 2013. **161**(3): p. 181-8.
181. Menke, A., et al., *Peripheral blood gene expression: it all boils down to the RNA collection tubes.* BMC Res Notes, 2012. **5**: p. 1.
182. Duale, N., et al., *Human blood RNA stabilization in samples collected and transported for a large biobank.* BMC Res Notes, 2012. **5**: p. 510.
183. Hantzsch, M., et al., *Comparison of whole blood RNA preservation tubes and novel generation RNA extraction kits for analysis of mRNA and MiRNA profiles.* PLoS One, 2014. **9**(12): p. e113298.
184. Bayatti, N., et al., *Comparison of blood RNA extraction methods used for gene expression profiling in amyotrophic lateral sclerosis.* PLoS One, 2014. **9**(1): p. e87508.
185. ElShamy, W.M. and D.M. Livingston, *Identification of BRCA1-IRIS, a BRCA1 locus product.* Nat Cell Biol, 2004. **6**(10): p. 954-67.

186. Gambino, G., et al., *Characterization of three alternative transcripts of the BRCA1 gene in patients with breast cancer and a family history of breast and/or ovarian cancer who tested negative for pathogenic mutations.* Int J Mol Med, 2015. **35**(4): p. 950-6.
187. Wooster, R., et al., *Identification of the breast cancer susceptibility gene BRCA2.* Nature, 1995. **378**(6559): p. 789-92.
188. Stray-Pedersen, A., et al., *The impact of an early truncating founder ATM mutation on immunoglobulins, specific antibodies and lymphocyte populations in ataxia-telangiectasia patients and their parents.* Clin Exp Immunol, 2004. **137**(1): p. 179-86.
189. Laake, K., et al., *Characterization of ATM mutations in 41 Nordic families with ataxia telangiectasia.* Hum Mutat, 2000. **16**(3): p. 232-46.
190. Telatar, M., et al., *Ataxia-telangiectasia: identification and detection of founder-effect mutations in the ATM gene in ethnic populations.* Am J Hum Genet, 1998. **62**(1): p. 86-97.
191. Bogdanova, N., et al., *A nonsense mutation (E1978X) in the ATM gene is associated with breast cancer.* Breast Cancer Res Treat, 2009. **118**(1): p. 207-11.
192. Susswein, L.R., et al., *Pathogenic and likely pathogenic variant prevalence among the first 10,000 patients referred for next-generation cancer panel testing.* Genet Med, 2016. **18**(8): p. 823-32.
193. Stray-Pedersen, A., *Ataxia-telangiectasia and other primary immunodeficiency diseases*. 2008, Oslo University Hospital: Oslo. p. 218.
194. Pylkas, K., et al., *Evaluation of the role of Finnish ataxia-telangiectasia mutations in hereditary predisposition to breast cancer.* Carcinogenesis, 2007. **28**(5): p. 1040-5.
195. Dominguez-Valentin, M., et al., *Potentially pathogenic germline CHEK2 c.319+2T>A among multiple early-onset cancer families.* Fam Cancer, 2018. **17**(1): p. 141-153.
196. Baugh, E.H., et al., *Why are there hotspot mutations in the TP53 gene in human cancers?* Cell Death Differ, 2018. **25**(1): p. 154-160.
197. Swisher, E.M., et al., *Somatic Mosaic Mutations in PPM1D and TP53 in the Blood of Women With Ovarian Carcinoma.* JAMA Oncol, 2016. **2**(3): p. 370-2.
198. Janatova, M., et al., *Mutation Analysis of the RAD51C and RAD51D Genes in High-Risk Ovarian Cancer Patients and Families from the Czech Republic.* PLoS One, 2015. **10**(6): p. e0127711.
199. Kiiski, J.I., et al., *Exome sequencing identifies FANCM as a susceptibility gene for triple-negative breast cancer.* Proc Natl Acad Sci U S A, 2014. **111**(42): p. 15172-7.
200. Singh, T.R., et al., *Impaired FANCD2 monoubiquitination and hypersensitivity to camptothecin uniquely characterize Fanconi anemia complementation group M.* Blood, 2009. **114**(1): p. 174-80.
201. Bogliolo, M., et al., *Biallelic truncating FANCM mutations cause early-onset cancer but not Fanconi anemia.* Genet Med, 2017.
202. Catucci, I., et al., *Individuals with FANCM biallelic mutations do not develop Fanconi anemia, but show risk for breast cancer, chemotherapy toxicity and may display chromosome fragility.* Genet Med, 2017.
203. Oshima, J., G.M. Martin, and F.M. Hisama. *Werner Syndrome*. [Internet] 2002 Dec 2 [updated 2016 Sep 29]; In: Adam MP, Ardinger HH, Pagon RA, et al., editors. GeneReviews®. Seattle (WA): University of Washington, Seattle; 1993-2017 [Available from: https://www.ncbi.nlm.nih.gov/books/NBK1514/.

204. Wang, Z., et al., *A polymorphism in Werner syndrome gene is associated with breast cancer susceptibility in Chinese women.* Breast Cancer Res Treat, 2009. **118**(1): p. 169-75.
205. Ding, S.L., et al., *Genetic variation in the premature aging gene WRN: a case-control study on breast cancer susceptibility.* Cancer Epidemiol Biomarkers Prev, 2007. **16**(2): p. 263-9.
206. Saintigny, Y., et al., *Homologous recombination resolution defect in werner syndrome.* Mol Cell Biol, 2002. **22**(20): p. 6971-8.
207. Kraemer, K.H. and J.J. DiGiovanna. *Xeroderma Pigmentosum*. [Internet] 2003 Jun 20 [updated 2016 Sep 29]; In: Adam MP, Ardinger HH, Pagon RA, et al., editors. GeneReviews®. Seattle (WA): University of Washington, Seattle; 1993-2017 [Available from: https://www.ncbi.nlm.nih.gov/books/NBK1397/.
208. Rafnar, T., et al., *Mutations in BRIP1 confer high risk of ovarian cancer.* Nat Genet, 2011. **43**(11): p. 1104-7.
209. Wilke, C.O., *Transcriptional robustness complements nonsense-mediated decay in humans.* PLoS Genet, 2011. **7**(10): p. e1002296.