FANCM MUTATIONS IN BREAST CANCER
RISK AND SURVIVAL

Johanna I. Kiiski

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Medicine of the University of Helsinki, for public examination in Lecture Hall 2, Biomedicum Helsinki 1, on 28 September 2018, at 13:00.

Helsinki, Finland 2018
Supervised by
Docent Heli Nevanlinna, PhD
Department of Obstetrics and Gynecology
Helsinki University Hospital
University of Helsinki
Helsinki, Finland

Reviewed by
Minna Kankuri-Tammilehto MD, PhD
Department of Clinical Genetics
Turku University Hospital
Institute of Biomedicine
University of Turku
Turku, Finland

Docent Synnöve Staff MD, PhD
Department of Obstetrics and Gynecology
Tampere University Hospital
Laboratory of Cancer Biology, BioMediTech
and Faculty of Medicine and Life Sciences
University of Tampere
Tampere, Finland

Official Opponent
Professor Minna Nyström, PhD
Faculty of Biological and Environmental Sciences
Research Programme in Molecular and Integrative Biosciences
University of Helsinki
Helsinki, Finland

Dissertationes Scholae Doctoralis Ad Sanitatem Investigandam Universitatis Helsinkiensis

ISSN 2342-3161 (Print)
ISSN 2342-317X (Online)
http://ethesis.helsinki.fi

Unigrafia
Helsinki 2018
4.1.2 Ovarian cancer patient samples ......................................................... 41
4.1.3 Population controls ....................................................................... 42
4.1.4 Ethical standards ........................................................................... 42
4.2 Methods ........................................................................................... 42
  4.2.1 Exome sequencing (I,III) .............................................................. 42
  4.2.2 DNA and RNA analyses and genotyping (I,III) ............................. 44
  4.2.3 Statistical analyses and bioinformatics .......................................... 45
5 Results ................................................................................................. 48
  5.1 Identifying FANCM as a novel breast cancer gene (I) ...................... 48
    5.1.1 FANCM c.5101C>T is associated with breast cancer risk .............. 48
    5.1.2 Familial segregation of the FANCM c.5101C>T mutation .............. 50
    5.1.3 Functional analysis of the FANCM c.5101C>T mutation .............. 52
  5.2 FANCM c.5101C>T survival (II) ....................................................... 52
    5.2.1 FANCM c.5101C>T mutation is associated with poor breast cancer prognosis................................................................. 52
    5.2.2 Interaction analyses ..................................................................... 54
    5.2.3 Histopathology of FANCM c.5101C>T mutation-positive tumors ................................................................. 55
    5.2.4 Immunohistochemical analyses .................................................... 55
  5.3 Truncating FANCM mutations (III) ................................................. 56
    5.3.1 Risk analyses ............................................................................... 56
    5.3.2 Combined risk analysis ................................................................ 57
6 Discussion ............................................................................................ 59
  6.1 FANCM and breast cancer risk (I,III) .............................................. 59
  6.2 FANCM and breast cancer survival (II) ........................................... 64
7 Summary and conclusions ..................................................................... 69
Acknowledgments .................................................................................. 71
References .............................................................................................. 73
LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications, which are referred to in the text by their Roman numerals:


ABBREVIATIONS

Gene names are written in italics.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKT</td>
<td>AKT!serine/threonine kinase 1</td>
</tr>
<tr>
<td>ATM</td>
<td>ATM!serine/threonine kinase</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>BARD1</td>
<td>BRCA1!Associated RING Domain 1</td>
</tr>
<tr>
<td>BER</td>
<td>base excision repair</td>
</tr>
<tr>
<td>BIC</td>
<td>Breast Cancer Information Core</td>
</tr>
<tr>
<td>BRCA1</td>
<td>BRCA1, DNA repair associated</td>
</tr>
<tr>
<td>BRCA2</td>
<td>BRCA2, DNA repair associated</td>
</tr>
<tr>
<td>BRIP1</td>
<td>BRCA1 interacting protein C-terminal helicase 1</td>
</tr>
<tr>
<td>BRRS</td>
<td>Bannayan–Riley–Ruvalcaba syndrome</td>
</tr>
<tr>
<td>CDH1</td>
<td>cadherin 1</td>
</tr>
<tr>
<td>CHEK2</td>
<td>checkpoint kinase 2</td>
</tr>
<tr>
<td>CI</td>
<td>confidence interval</td>
</tr>
<tr>
<td>CS</td>
<td>Cowden syndrome</td>
</tr>
<tr>
<td>DSB</td>
<td>double-strand break</td>
</tr>
<tr>
<td>DSBR</td>
<td>double-strand break repair</td>
</tr>
<tr>
<td>ER</td>
<td>estrogen receptor</td>
</tr>
<tr>
<td>FA</td>
<td>Fanconi anemia</td>
</tr>
<tr>
<td>FAAP24</td>
<td>Fanconi anemia core complex associated protein 24</td>
</tr>
<tr>
<td>FANCA/B/C/D/etc.</td>
<td>Fanconi anemia complementation group A/B/C/D</td>
</tr>
<tr>
<td>FIGO</td>
<td>Fédération Internationale de Gynécologie et d’Obstétrique</td>
</tr>
<tr>
<td>FIMM</td>
<td>Institute for Molecular Medicine Finland</td>
</tr>
<tr>
<td>GG-NER</td>
<td>global genome nucleotide excision repair</td>
</tr>
<tr>
<td>GWAS</td>
<td>genome-wide association analyses</td>
</tr>
<tr>
<td>HER2</td>
<td>human epidermal growth factor receptor 2</td>
</tr>
<tr>
<td>HNPCCC</td>
<td>hereditary non-polyposis colorectal cancer</td>
</tr>
<tr>
<td>HR</td>
<td>hazard ratio</td>
</tr>
<tr>
<td>HRR</td>
<td>homologous recombination repair</td>
</tr>
<tr>
<td>ICL</td>
<td>interstrand crosslink</td>
</tr>
<tr>
<td>LFS</td>
<td>Li-Fraumeni syndrome</td>
</tr>
<tr>
<td>LOF</td>
<td>loss-of-function</td>
</tr>
<tr>
<td>LS</td>
<td>Lynch syndrome</td>
</tr>
<tr>
<td>MLH1</td>
<td>MutL Homolog 1</td>
</tr>
<tr>
<td>MMR</td>
<td>mismatch repair</td>
</tr>
<tr>
<td>MRE11A</td>
<td>MRE11 homolog, double-strand break repair nuclease</td>
</tr>
<tr>
<td>MSH2/3/6</td>
<td>MutS Homolog 2/3/6</td>
</tr>
<tr>
<td><strong>MUTYH</strong></td>
<td>mutY DNA glycosylase</td>
</tr>
<tr>
<td><strong>NBN</strong></td>
<td>nibrin</td>
</tr>
<tr>
<td><strong>NER</strong></td>
<td>nucleotide excision repair</td>
</tr>
<tr>
<td><strong>NF1</strong></td>
<td>neurofibromin 1</td>
</tr>
<tr>
<td><strong>NHEJ</strong></td>
<td>non-homologous end joining</td>
</tr>
<tr>
<td><strong>NMD</strong></td>
<td>nonsense-mediated RNA decay</td>
</tr>
<tr>
<td><strong>OR</strong></td>
<td>odds ratio</td>
</tr>
<tr>
<td><strong>PALB2</strong></td>
<td>Partner and localizer of BRCA2</td>
</tr>
<tr>
<td><strong>PAR</strong></td>
<td>poly(ADP-ribose)</td>
</tr>
<tr>
<td><strong>PARP</strong></td>
<td>poly(ADP-ribose) polymerase</td>
</tr>
<tr>
<td><strong>PARPi</strong></td>
<td>poly(ADP-ribose) polymerase inhibitor</td>
</tr>
<tr>
<td><strong>PHTS</strong></td>
<td>PTEN hamartoma tumor syndrome</td>
</tr>
<tr>
<td><strong>PI3K</strong></td>
<td>phosphoinositide 3-kinase</td>
</tr>
<tr>
<td><strong>PJS</strong></td>
<td>Peutz–Jeghers syndrome</td>
</tr>
<tr>
<td><strong>PMS2</strong></td>
<td>PMS1 homolog 2, mismatch repair system component</td>
</tr>
<tr>
<td><strong>PR</strong></td>
<td>progesterone receptor</td>
</tr>
<tr>
<td><strong>PTEN</strong></td>
<td>phosphatase and tensin homolog</td>
</tr>
<tr>
<td><strong>RAD50</strong></td>
<td>RAD50 double-strand break repair protein</td>
</tr>
<tr>
<td><strong>RAD51</strong></td>
<td>RAD51 recombinase</td>
</tr>
<tr>
<td><strong>RAD51C/D</strong></td>
<td>RAD51 paralog C/D</td>
</tr>
<tr>
<td><strong>RECQL</strong></td>
<td>RecQ like helicase</td>
</tr>
<tr>
<td><strong>RNA Pol II</strong></td>
<td>RNA polymerase II</td>
</tr>
<tr>
<td><strong>RPA</strong></td>
<td>replication protein A</td>
</tr>
<tr>
<td><strong>SDSA</strong></td>
<td>synthesis-dependent strand annealing</td>
</tr>
<tr>
<td><strong>SNP</strong></td>
<td>single-nucleotide polymorphism</td>
</tr>
<tr>
<td><strong>ssDNA</strong></td>
<td>single-strand DNA</td>
</tr>
<tr>
<td><strong>STK11</strong></td>
<td>serine/threonine kinase 11</td>
</tr>
<tr>
<td><strong>TC-NER</strong></td>
<td>transcription coupled nucleotide excision repair</td>
</tr>
<tr>
<td><strong>TFIHH</strong></td>
<td>transcription initiation factor I1H</td>
</tr>
<tr>
<td><strong>TN</strong></td>
<td>triple-negative</td>
</tr>
<tr>
<td><strong>TNM</strong></td>
<td>tumor size, lymph node status, distant metastasis</td>
</tr>
<tr>
<td><strong>TP53</strong></td>
<td>tumor protein p53</td>
</tr>
<tr>
<td><strong>XP</strong></td>
<td>xeroderma pigmentosum</td>
</tr>
<tr>
<td><strong>XPC</strong></td>
<td>XPC complex subunit, DNA damage recognition and repair factor</td>
</tr>
<tr>
<td><strong>XRCC2</strong></td>
<td>X-ray repair cross complementing 2</td>
</tr>
</tbody>
</table>
ABSTRACT

Breast cancer is the most diagnosed malignancy and the leading cause of cancer mortality in women worldwide. In Finland, approximately 5,000 cases are diagnosed annually, constituting about 30% of all new female cancers. Breast cancer also occurs in men, albeit rarely; around 20 cases are met in Finland each year.

Although commonly referred to as a single disease, breast cancer is a clinically and morphologically heterogeneous disorder. Different histologic and molecular subtypes are associated with distinct risk factors, predisposing mutations, patient prognosis, and treatment outcome. The single most significant risk factor for breast cancer is familial predisposition. Inherited germline mutations can increase the lifetime risk of breast cancer by up to ~80%, often conferring increased risk also for ovarian cancer.

The aim of this study was to identify novel breast and/or ovarian cancer alleles in the Finnish population (I), to evaluate the cancer risk in large case-control datasets (I,III), and to examine the tumor characteristics, patient survival, and treatment outcome associated with the identified mutations (II).

Exome sequencing and further genotyping of large sample sets of breast and ovarian cancer patients as well as healthy population controls from Finland identified FANCM as a novel moderate-risk breast cancer gene (I). The frequency of the FANCM c.5101C>T nonsense mutation was higher in breast cancer patients (3.1%) than in controls (1.8%). The most significant association and a fourfold increased risk was seen for triple-negative breast cancer. This aggressive subtype does not respond to hormone therapy and has a poor prognosis.

In the follow-up study of a large dataset of breast cancer patients from different geographical areas of Finland (II), FANCM c.5101C>T mutation was associated with poor 10-year breast cancer-specific survival, especially among familial patients. The mutation also increases the risk for local recurrence of the disease in patients not receiving radiation treatment, but not in patients treated with radiation, indicating that mutation carriers may specifically benefit from radiotherapy. Based on immunohistochemistry analyses, mutation carriers exhibit reduced DNA repair-associated PARP-activity, suggesting that PARP inhibitor therapy could be utilized in the management of breast cancer in FANCM mutation-positive patients.

Another FANCM nonsense mutation c.5791C>T identified among familial cases in a multicenter study was investigated in a large case-control series of Finnish breast cancer patients (III). The mutation was particularly enriched in triple-negative cases, similar to the FANCM c.5101C>T mutation. Combined analysis of both mutations confirmed the association with triple-negative breast cancer. Two other deleterious FANCM variants identified in
Finnish cancer patients were too rare to allow statistical evaluation, but these mutations may suggest a wider mutation spectrum of the \textit{FANCM} gene.

Compared with other European populations, \textit{FANCM} c.5101C>T and c.5791C>T mutations are more common in Finland. The enrichment of the mutations in the Finnish population may be explained by several features of founder effects typical to restricted populations with recent bottlenecks.

The discovery of cancer predisposing variants is important for early diagnosis, individual breast cancer risk assessment, and precise treatment. This applies particularly to families with a history of breast cancer. Inherited mutation carriers may benefit from intense follow-up or preventive measures. Also studying the effects of breast cancer mutations on tumor phenotype, patient survival, and treatment outcome can improve the clinical management of cancer and survival of mutation carriers. Identifying new risk alleles will further improve knowledge of the genetic background of the disease, as well as its pathobiology. Additional studies are warranted to define the cancer risks associated with \textit{FANCM} mutations and investigate their prevalence in other populations, and also to confirm the mechanism associated with the observed aggressive phenotype of \textit{FANCM}-defective tumors.
FINNISH SUMMARY

Joka kolmas suomalainen sairastuu syöpään elinaikanaan, ja vuosittain syöpädiagnoosin saa noin 30,000 henkilöä Suomessa. Naisilla rintasyöpä on yleisin; joka vuosi noin 5,000 naista sairastuu. Rintasyövän esiintyvyys on kasvanut voimakkaasti muun muassa väestön ikääntyminen myötä, mutta taudin ennuste on jatkuu parantunut varhaisemman toteamisen sekä tehokkaampien hoitomenetelmien ansiosta. Miehillä rintasyöpä on harvinainen ja uusia tapauksia todetaan Suomessa vuosittain parikymmentä.

Rintasyöpä on monimuotoinen sairaus, ja sen eri alatyyppit vaikuttavat ennusteeeseen ja syöpähoitojen valintaan. Yleisiä riskitekijöitä ovat muun muassa runsas alkoholinkäyttö, ylipaino ja etenkin hormonaaliset tekijät. Merkittävin yksittäinen riskitekijä rintasyövälle on kuitenkin perinnöllinen alttius, joka voi lisätä sairastumisriskiiä moninkertaistamalla, ja kasvattaa usein myös risiköt sairastua munasarjasyöpään, jonka ennuste on rintasyöpää huonompi.

Tämän väitöskirjatyön tavoitteena oli löytää uusia rinta- ja/tai munasarjasyövälle altistavia geenimuutoksia suomalaisessa väestössä ja tarkemmin tutkia tunnistettuihin mutaatioihin liittyviä sairastumisriskejä sekä niiden vaikutusta hoitovasteeseen, kasvainten ominaisuuksiin ja potilaiden eloonjääntiin. Syövälle altistavien geneettisten muutosten tunnistaminen tukee henkilökohtaisen sairastumisriskin arvioimista, varhaista diagnoosia ja yksilöllistä hoitoa.


1 INTRODUCTION

Breast cancer is the most common type of invasive cancer in women, accounting for approximately 30% of all female cancers. It also has the highest mortality rate among cancers in women worldwide. Breast cancer is an important public health issue, and in Finland near 5,000 women are diagnosed annually.

Breast cancer is a heterogeneous disease, with several subtypes associated with different predisposing mutations, prognosis, and treatment outcome. The lifetime breast cancer risk for any woman in the Western world is approximately 10%, but family history of the disease can substantially increase the risk, and hereditary factors may often have an impact on sporadic cancer cases as well. Mutations in the two major susceptibility genes, *BRCA1* and *BRCA2*, explain ~20% of familial breast cancer cases worldwide. A large portion of the remaining cases can be explained with mutations in moderate-risk susceptibility genes and a polygenic model of predisposition, in which several genetic changes together with environmental factors have small and independent effects. Furthermore, inherited founder mutations may appear in certain populations with higher frequencies and increase the cancer risk in carrier families. Such mutations are usually encountered in restricted and inbred populations, e.g. in Finland and Iceland.

In the Western world, the prognosis of breast cancer has improved due to regular screening programs, earlier diagnosis, and advances in treatment. Prognosis is affected by several factors associated with the biological characteristics of the tumor, such as hormone receptor expression, tumor size, and presence of metastases. Many studies show that family history may influence also breast cancer survival, however further research is needed to determine the heritable component of the outcome of the disease.

Identification of new breast cancer susceptibility genes and prognostic biomarkers allow early diagnosis, more detailed prognosis, and accurate treatment. This study aimed to find novel breast and/or ovarian cancer susceptibility alleles in the Finnish population and further evaluate the association of the identified mutations with the disease, breast cancer-specific survival, and treatment outcome.
2 REVIEW OF THE LITERATURE

2.1 CANCER

Cancer is the second most common cause of death worldwide. It can be seen as a heterogeneous group of diseases with a vast number of subtypes, affected by diverse risk factors and varying epidemiology. Cancer can originate from almost all cell types and organs in the human body, and the disease is characterized by uncontrolled proliferation of cells that eventually obtain the ability to invade adjacent tissues and metastasize to distant organs 1, 2.

About 14.1 million new cancer cases and 8.2 million cancer deaths were registered worldwide in 2012. The overall number of cancers continues to rise as the population grows and life expectancy increases. Known risk factors, such as smoking, poor diet, and changes in reproductive patterns influence cancer burden in both developed and less developed countries 2, 3. Lung and breast cancer are the most common cancers and also the leading cause of cancer deaths 2. In Finland, one of three citizens will get cancer during their lifetime, and approximately 30,000 people are diagnosed with cancer each year. Fortunately, nearly two-thirds of affected persons will recover 4.

2.1.1 Biology of cancer

Cancer is a disease of our genes, the result of a collection of changes that occur in DNA sequences of the cancer cell genomes. These alterations affect the expression of the genes, allowing the cells to escape normal growth regulation systems.

Each cell in every tissue of a human body is a direct descendant of its progenitor through mitotic cell divisions. Human cells develop, grow, divide, and eventually die under the tight control of a cell cycle machinery: a set of regulatory signaling networks specific to each tissue. Genetic and epigenetic changes – inherited or somatic – are required in order for cells to acquire the ability for abnormal growth, following the principles of Darwinian natural selection 1–5.

The variability in cancer progression, histopathology, mutagenesis, and epidemiology is extensive and the pathway to cancer may differ substantially between different cancer types. However, the classical hallmarks of cancer initiation and progression are always similar. They include the ability of the cell to escape from mitogenic growth signaling control and to proliferate at increasing speed. This requires not only evading the antiproliferative signals, but also escaping from the programmed cell death, apoptosis, which normally terminates the life of a rebellious or damaged cell. To acquire an
unlimited proliferation ability, the cancer cell must also prevent shortening of telomeres, i.e. structures of a chromosome that normally function as a counting device for cell proliferation and lead to death of an aged cell after a sufficient number of cell divisions. When all of these acquired abilities eventually lead to a tumor formation, the growing mass needs to guarantee the supply of oxygen and nutrients by maintaining angiogenesis, i.e. the formation of blood vessels. Finally, to fully develop into a malignant cancer, the tumor cells must move out and invade adjacent tissues. Metastases are a typical sign of an advanced cancer, and the main reason for cancer deaths.

Furthermore, tumors must reprogram the glucose metabolism to maximize energy uptake as well as escape destruction by the body’s immune system.

Tumors may develop in any tissue of the human body. However, eukaryotic tissues usually consist of several cell types, and thus cancers are commonly classified by the type of cells from which the tumors originate from. Carcinomas derive from epithelial cells, being the most common group of cancers in adults; most breast, lung, and colon cancers are carcinomas. Sarcomas arise from connective tissue, lymphomas from lymphocytes, and leukemias from bone marrow. Blastosomas originate from immature or embryonic cells, consequently being more common in adolescents. Benign tumors, such as lipomas originating from fat cells, lack the ability to invade other tissues and metastasize. They can, however, cause severe health problems and may become malignant.

Most solid tumors are assembled of several cell types that facilitate the cancerous growth and progression and collectively create the tumor microenvironment, which subsequently transforms when the tumor invades new tissues. Cancer stem cells are able to self-renew and give rise to cells that cannot promote tumor growth but may have some other functions and also constitute the heterogeneous tumor mass. The phenotypic plasticity enables the cancer stem cells to spawn functionally different subpopulations, and their complexity may affect cancer treatment via resistance to chemotherapeutic agents or radiation as well as the disease recurrence.

### 2.1.2 Cancer genes and mutations

Thousands of DNA lesions occur daily by internal physiological processes and external mutagens, leading to defects in DNA and subsequently to tumor formation if not properly corrected by the cell’s own repair systems. DNA repair mechanisms are discussed in detail in Section 2.1.4. Several different mutation types on a DNA or chromosome level may change the expression of the genes, including base substitutions, deletions, and insertions of a single nucleotide or longer DNA segments, and different rearrangements such as amplifications, inversions, translocations, and copy number alterations. In addition, the cancer genome commonly has epigenetic changes affecting gene
expression by modifying chromatin structure without disrupting the genetic content of the cells. This may lead to activation or silencing of the genes.

Not every uncontrolled genetic change in cancer cell genomes drives tumor progression; some may have an unspecified function or no role at all. The terms “driver” and “passenger” mutation describe the consequences of the mutations; a driver mutation actually grants growth advantage to the cancer cell and has been positively selected. Of ~20,000 protein coding genes in the human genome, approximately 200 have been shown to act as drivers through certain pathways controlling cell growth, replication, and death, and the mutated versions of the same driver genes are often discovered in different cancers. A passenger mutation is not selected and typically does not assist cancer development, subsequently being mostly harmless to the tumor. It usually, however, will get transferred to descendant cells in clonal evolution, until the final stages of cancer. The great amount of passenger mutations may complicate genetic studies aimed at identifying new mutations associated with cancer progression.

The number of mutations required for a normal cell to become malignant has been largely debated. With respect to solid tumors, three alterations in driver genes may be sufficient for cancer transformation. However, a tumor genome may contain thousands of passenger mutations and epigenetic changes not detected in the germline. These mutations typically are distinct in every tumor, in contrast to the cellular controlling pathways affected by driver mutations which are usually similar in all cancers.

Traditionally, the cancer genes have been classified into two categories. Oncogenes are mutated versions of normal proto-oncogenes that regulate cell growth, proliferation, apoptosis, and differentiation. They function dominantly, with only one defective copy of the gene being sufficient to promote the tumor formation. Oncogene mutations typically operate with gain-of-function model, switching the gene to a constantly active mode or altering its primary functions.

Conversely, defects in tumor suppressor genes are commonly recessive loss-of-function (LOF) mutations, altering tumor-preventing functions of the genes. Tumor suppressors in general maintain the genomic integrity by halting cells from dividing, activating DNA repair, and initiating apoptosis when necessary. The renowned “two-hit hypothesis” by Alfred G. Knudson demonstrated that loss of both alleles of a tumor suppressor gene is required to alter the phenotype. The first “hit” exists already in the germline, whereas the other is somatic, however there are several exceptions. Dominant-negative function of a protein product can prevent the function of the normal allele of tumor suppressor, which is often the case with missense mutations in TP53 gene. Furthermore, in the situation of haploinsufficiency, the gene product from the wild-type allele is not sufficient alone to completely maintain the protein function.

Tumor suppressor genes have been further classified into three categories: gatekeepers, caretakers, and landscapers. Gatekeepers inhibit the
growth of a cell or promote cell death. Thus, defective gatekeepers lose their
ability to prevent tumorigenesis. The inactivating caretaker gene does not
directly affect the tumor progression, instead leading to genomic instability
by increasing the speed of mutagenesis. Many tumor suppressors may have
abilities of both gatekeeping and caretaking, as is the case with BRCA1 and
BRCA2 genes, commonly mutated in breast cancer. The third type of
tumor suppressor genes are landscapers; genes that regulate the tumor
microenvironment, favoring abnormal growth when mutated.

2.1.3 Hereditary predisposition to cancer

The vast majority of all cancer cases arise from somatically acquired
mutations, but some individuals carry a specific inherited germline mutation
in each cell of their body, greatly increasing the cancer risk. However,
familial cancer incidence explains only a small proportion of all cancer cases
worldwide.

Familial susceptibility to cancer was suggested already in the 17th century.
Epidemiological studies in the 1940s and 1950s, based on data showing
increased cancer risk in relatives of cancer patients, have proved the
hypothesis. Later, in the 1980s and 1990s, several hereditary cancer
predisposition genes, such as Lynch syndrome genes MLH1 and MSH2 and
breast cancer susceptibility genes BRCA1 and BRCA2, were identified with
linkage analysis and positional cloning. Most of the susceptibility genes are
tumor suppressors (germline mutations in oncogenes are commonly lethal),
often associated with DNA repair pathways. The mutations in these genes are
highly penetrant, conferring significantly increased cancer risk. However,
such mutations are rare and their frequencies differ between populations.

Many familial cancers do not derive from high-risk mutations, instead
arising from lower risk variants with incomplete penetrance, elevating the
cancer risk from twofold to fivefold. Although these mutations are
overrepresented in cancer families, they do not segregate completely, so
detection with linkage analysis is not possible. Instead such variants have
been identified with candidate gene approaches. Moderate-penetrance
mutations act dominantly and independently and are often population
specific founder mutations (further discussed in Section 2.2.4). Biallelic,
homozygous, and compound heterozygous moderate-risk mutation carriers
commonly have distinct clinical phenotypes. Biallelic mutations in moderate-
penetrance cancer genes are often associated with distinct childhood-onset
syndromes that confer increased risk for different cancer types. A
number of targeted gene analyses are available for several cancer
predisposition disorders, and in families with high incidence of cancer cases,
counseling and genetic testing may facilitate preventive measures or early
detection of the cancer.
Much of the inherited cancer susceptibility is thought to result from the polygenic model of predisposition, in which several genetic variants have small and independent effects. They affect both familial and sporadic cancer cases and are common in the population, also among unaffected individuals. Large-scale genome-wide association studies (GWAS) in large datasets have identified hundreds of common low-penetrance loci. The risk from each variant is modest, but the combined effect may be considerable. Potentially thousands of these common loci exist, each contributing to the formation of tumors in concert with environmental and biological factors 24, 27, 28.

2.1.4 DNA repair

Thousands of DNA lesions occur daily in the \( \sim 10^{13} \) cells of the human body 29. The main goal of living organisms is to deliver their DNA correctly and intact to the next generation, and several mechanisms have thus evolved to detect and subsequently correct DNA damage and maintain genomic stability during the cell cycle (Figure 1). DNA repair mechanisms are highly conserved from bacteria to eukaryotes, and this universality of DNA repair processes among all life forms emphasizes the importance of genome stability 11.

**Figure 1.** Simplified illustration of the cell cycle clock of a typical mammalian cell. Several checkpoints are included in the replication cycle of a cell to ensure DNA damage repair and maintain genome stability.
2.1.4.1 Base excision repair

Base excision repair (BER) assesses the majority of endogenous DNA damage such as deaminations, depurinations, alkylations, and oxidative damage (Figure 2). It removes approximately 40,000 endogenous lesions per human cell each day, thus playing a clear role in cancer prevention 30. In the BER pathway, DNA glycosylases recognize damaged bases and excise them. At least 11 such enzymes are known, each devoted to a specific type of lesions. Endonucleases cleave the site to form a 3′-hydroxyl end and a 5′-deoxyribose phosphate end, after which the repairing DNA synthesis and DNA ligase-directed strand ligation will complete the process 30, 31.

BER knockout mouse models accumulate DNA damage and may develop, for example, gastric lesions and lymphomas, and lung, colon, and ovarian tumors. Therefore BER-associated mutations have been an attractive target for identifying new cancer-predisposing genes, but the results have been conflicting. However, monofunctional BER-glycosylase MUTYH has been identified as a colon cancer susceptibility gene 30, 32. Also, the breast cancer susceptibility gene BRCA1, without being a direct partner in the BER core complexes, is shown to stimulate several early steps in the BER pathway in human breast carcinoma cell lines 33. Furthermore, triple-negative (TN) and BRCA1-mutated breast cancer cell lines showed reduction in the ability to repair oxidative DNA damage with the BER pathway 34. Altogether, in the absence of functional BER machinery mutations will accumulate in cells, which become hypersensitive to DNA-damaging agents 30.

2.1.4.2 Nucleotide excision repair

Nucleotide excision repair (NER) is responsible for correcting a wide range of single-strand DNA lesions in mammals (Figure 2) by two different partially overlapping sub-pathways. Transcription-coupled NER (TC-NER) repairs lesions in the transcribed strand of active genes when encountering stalled RNA polymerase II (RNA Pol II). Global genome NER (GG-NER) can occur anywhere in the genome when helix distortions are detected 35, 36. In GG-NER, the main initiator of the repair is XPC, which binds next to the lesion. This allows the association of the TFIIF (transcription initiation factor IIH) complex, which unwinds the helix with helicase and excises the lesion with endonucleases, leaving a 22- to 30-nucleotide-long single-strand gap, which is then filled and new DNA is synthesized by specific proteins 36.

TC-NER is initiated by lesion-stalled RNA Pol II, which recruits the cascade of TC-NER machinery proteins. It is believed that RNA Pol II then backtracks, allowing the repair complex to operate 37. DNA is unwound to form a 20- to 30-nucleotide loop, the lesion in excised, and new nucleotides inserted with DNA polymerase complex. A set of ligases seals the DNA 38.
Defects in the GG-NER may allow the genome-wide accumulation of DNA lesions, leading to strong cancer predisposition. Patients with autosomal recessive GG-NER disorder xeroderma pigmentosum (XP) have high incidence of skin cancers due to extreme sensitivity to sunlight. In clinical studies, a nearly 10,000-fold increase of skin cancers have been found in XP patients aged under 20 years. In addition, carcinogens in cigarette smoke bind to DNA, causing damage that is usually repaired with the NER route. Thus XP patients have a higher smoking-induced cancer rate.

**Figure 2.** Overview of the main DNA repair mechanisms.

### 2.1.4.3 DNA mismatch repair

An incorrectly inserted nucleotide during DNA synthesis generates a non-complementary base pair within the DNA helix. The proofreading ability associated with some DNA polymerases can correct the misstep, however, the process is prone to errors. Mistakes escaping DNA polymerase proofreading machinery can be corrected with mismatch repair (MMR). This mechanism also corrects errors created as a natural outcome of genetic recombination or chemically modified bases in DNA, caused by agents such as O6-methylguanine, carcinogen adducts, and UV photo products (Figure 2).
In the MMR mechanism, a combination of Msh2-Msh6 complex (MutSα) and Msh2-Msh3 complex (MutSβ) together with MLH complexes and several associated proteins recognize the mispaired bases and promote an excision of the mispaired strand, resulting in a gap that is filled with the correct base by DNA polymerases. However, MMR is strand-specific, and in order to recognize and repair the mismatched base, the MMR machinery must distinguish the newly synthesized DNA strand from its parental strand. In eukaryotes, several discrimination mechanisms have been proposed; commonly a nick present in DNA would act as signal for the recognition of a new strand.

Defects in the MMR route can result in a threefold higher overall mutation rate, and germline mutations in MMR genes, most commonly in MLH1 and MSH2, are associated with hereditary non-polyposis colorectal cancer (HNPCC) and also development of sporadic tumors in different tissues. Loss of the functional MMR causes microsatellite instability, which is a hypermutable phenotype, leading to an increased number of microsatellite repeats (short repetitive DNA sequences) that are prone to frameshift mutations and substitutions during DNA replication.

2.1.4.4 Homologous recombination repair

Most DNA repair routes are targeted to single-strand DNA defects. Yet the most severe damage to the genome is caused by DNA double-strand breaks (DSBs). DSBs can occur as a result of exogenic sources, such as ionizing radiation and X-rays, or disturbances during DNA replication (Figure 2). They can be repaired by homologous recombination repair (HRR) and non-homologous end joining (NHEJ). NHEJ can work in the cells at any point of the cell cycle, but the HRR pathway is active only in the S/G2-phases (DNA synthesis/cell growth). The main difference between the mechanisms is indicated in their names: NHEJ can unite DNA ends without any requirements for homology, and thus, it is also prone to errors by rejoining the wrong ends of DNA, resulting in random translocations or small deletions and insertions. HRR is thought of as a “copy-and-paste” mechanism, which usually requires the replicated sister chromatid as a template for repair, and therefore, it only occurs during S/G2-phases. Using the sister chromatid as a template makes HRR a very accurate repair mechanism. In addition to DNA repair, homologous recombination has an important role in DNA replication, telomere maintenance, and meiotic chromosome segregation.

The two main HRR pathways are double-strand break repair (DSBR), which can produce crossover recombinants, and synthesis-dependent strand annealing (SDSA), which only produces identical DNA molecules and is the suggested model for mitotic double-strand repair. Both routes have similar initiation. When encountering a DSB, free DNA ends are detected by MRE11,
RAD50, and NBN-complex, promoting DNA damage checkpoint signaling by ATM. Both strands are resected to create 3′ overhangs. A replication protein A (RPA) is loaded to the single-stranded DNA (ssDNA) in the broken chromosome, and it facilitates the assembly of RAD51 filaments, which is controlled by interactions of BRCA1, PALB2, and BRCA2 proteins. Strand invasion of the sister chromatid by the 3′ ssDNA overhang allows copying of the genetic information and the strands are annealed and the gaps enzymatically sealed. In the DSBR model, the generation of DNA intermediates between the recipient and donor chromatids, called Holliday junctions, is vital for creating cross-over products, whereas in the SDSA model the invading strand is displaced after repairing the DSB, and the ssDNA tails are reannealed directly, without crossover events.

In addition to DSB repair, HRR can also resolve stalled replication forks and repair interstrand crosslinks. This requires the association of the Fanconi anemia (FA) protein complex, discussed in detail in Section 2.4.1.

### 2.2 BREAST CANCER

Breast cancer is the most frequently diagnosed cancer and also the leading cause of cancer death among women worldwide, accounting for approximately one-third of all cancer cases, with an estimated 1.7 million new diagnoses each year. In the Western world, the lifetime breast cancer risk for women is approximately 10%. Breast cancer rates are highest in North America, Australia/New Zealand, and Northern and Western Europe and lowest in most of Africa and Asia. The differences between breast cancer incidences can be explained by the availability of early detection and known risk factors such as obesity, physical activity, and hormonal factors including age of menarche and menopause and parity. The increasing breast cancer rates seen recently in South America, Africa, and Asia may be explained by lifestyle changes moving towards Westernization, however, all reasons for accumulation of breast cancer cases in these countries are not understood.

In Finland, 4,717 new breast cancers were diagnosed in 2015. This comprises approximately 30% of all new cancer cases. Time trends of breast cancer incidence per 100,000 persons in 1956-2015 are illustrated in Figure 3. Relative 5-year survival rate among women in Finland is 91%. Breast cancer also occurs in men, albeit very rarely. Around 20 cases are seen annually in Finland.
Figure 3. Time trends of female breast cancer incidence in Finland in 1956-2015, age-standardized (World) rates per 100,000 persons. Source: Finnish Cancer Registry, syoparekisteri.fi/tilastot/tautitilastot/. Information as of 05.02.2018.

2.2.1 Biology and classification of breast cancer

A human breast (or mammary gland) consists of fat, connective tissue, glands, and ducts, supported by a network of nerves and blood and lymphatic vessels. A female breast is divided into 15-25 lobules, which are connected by ducts that transport milk to the nipple. The female breast enlarges due to ovarian estrogen and progesterone production during puberty, leading to proliferation of epithelial and connective tissue components. The development is considered completed during pregnancy, and the ductal and glandular elements start to regress during menopause 56, 57. Considering all the developmental characteristics, female breast tissue experiences an extensive amount of remodeling and hormonal modifying during a person’s lifetime; in utero development, puberty, monthly pre-menopausal cycles, possible pregnancies and lactation, and menopause. Therefore cell proliferation, apoptosis, and differentiation in the breasts occur at higher rates than in most tissues in the human body, subsequently increasing the possibility of accumulation of DNA replication errors 58.

Most invasive breast cancers are carcinomas, deriving from epithelial cells. Around 70-80% can be classified as ductal and 10-15% as lobular
subtype. Medullary, tubular, papillary, or mucinous subtypes of breast cancer are rare. In situ carcinomas are preinvasive lesions that have not spread beyond the ducts into the surrounding breast tissue. The survival rate of in situ patients is close to 100%, however, having such a lesion can increase the risk of developing an invasive breast cancer later in life 59, 60.

Several molecular sub-categories of breast cancer are recognized. They differ in prognosis, and each subtype also markedly affects the treatment. The most important classifier of breast cancer is estrogen (ER) and progesterone (PR) hormone receptor expression of tumors, defined by immunohistochemical methods. Commonly, the expression of ER and PR receptors is strongly correlated. Most breast tumors (60-70%) are ER-positive and they have a better prognosis than ER-negative tumors. Another classifier and predictive marker is overexpression of the proto-oncogene HER2, an important driver in many tumors 60, 61, 62.

More detailed molecular classification of breast tumors is based on gene-expression profiling 63, 64. Most breast carcinomas are classified as luminal A, expressing ER and PR receptors, but are HER2-negative. Proliferation rate marker Ki67 is usually low. Luminal B-type cancers are also ER/PR positive, HER2 status may vary, but Ki67 expression is high 60, 64, 65. Triple-negative or basal-like carcinomas do not express ER or PR and are additionally negative for HER2. This subgroup of breast cancer, being usually of high grade, often has poor a prognosis at least partly due to the lack of effect of targeted therapies 65, 66. The fourth group is HER2-positive carcinomas, which are often ER-negative but show overexpression of HER2 60, 64, 65.

In the clinical approach, breast cancers are classified based on the TNM-system (tumor size, spreading of the disease to lymph nodes, and distant metastasis). The classification correlates with the survival of patients; smaller tumors predict better survival than larger ones. The absence of axillary nodal metastases or other localized metastases also usually indicates better prognosis. TNM classification is constantly updated and widely utilized when deciding on a breast cancer treatment 67, 68.

Histological grade, describing the differentiation status of the breast tumor, is another important prognostic marker. Grading is based on the degree of nuclear pleomorphisms, mitotic count, and tubule formation. Grade I tumors have better survival than poorly differentiated grade II and III tumors 69.

2.2.2 Risk factors for breast cancer

The etiology of breast cancer is multifactorial. Inherited predisposition is the most significant individual risk factor for breast cancer, discussed further in Section 2.2.3. Other risk factors are hormonal and environmental. The overall risk for individual is usually a combination of all of these elements.
Woman’s age is an evident risk factor for breast cancer; the incidence increases with age during the reproductive years until menopause, after which the incidence decreases. Breast cancers in young women are more likely to be of the triple-negative or HER2-overexpressing subtype than those in older women. Many other risk factors are related to reproductive and hormonal issues. Young age at first childbirth, breastfeeding, and having more children decrease the risk. Hormone replacement therapy, young age at menarche, and old age at menopause are associated with increased breast cancer risk, as these factors increase the number of menstrual cycles, thus increasing exposure to the endogenous female hormones stimulating cell growth in mammary tissue.

Lifestyle factors, such as obesity and alcohol use, may affect the risk of developing a breast cancer. Meta-analyses have also shown an association between reduced breast cancer risk and increased physical activity. However, environmental risk factors may not have the same effect on every person, due to the genetic components of each individual. For example, mammography density is known to considerably increase breast cancer risk, but it may be affected by both genetic and environmental issues.

2.2.3 Inherited susceptibility to breast cancer

A family history of breast and/or ovarian cancer can considerably increase the lifetime risk of developing breast cancer. A high prevalence of breast tumors in one family over four generations was first described by French physician Paul Broca in 1866. Identification of the most well-known breast cancer susceptibility genes in the 1990s, BRCA1 and BRCA2, established family history as a major risk factor for breast cancer and also allowed predictive genetic testing of susceptibility in families.

Genes harboring breast cancer-associated mutations typically encode proteins involved in intracellular DNA damage signaling and repair networks, especially those involved with DSBR with homologous recombination such as BRCA1, BRCA2, and PALB2 (Table 1). Rarely, breast cancer susceptibility genes may also have a role in genome maintenance or in other biological functions.

Approximately 10% of all breast cancers are today estimated to arise from inherited mutations in breast cancer susceptibility genes. The proportion may be threefold higher in patients diagnosed under 30 years of age, and breast cancer in a first-degree relative increases the risk threefold. A high number of both breast- and ovarian cancer cases in the same family is often clinically described as a hereditary breast and ovarian cancer syndrome, characterized by young age of disease onset, increased risk for both breast and ovarian cancer, male breast cancer, and higher rate of bilateral tumors.
2.2.3.1 High- and moderate-risk breast cancer genes

**BRCA1** and **BRCA2** are tumor suppressor genes, controlling genome integrity by arresting cells from dividing after DNA damage and participating in homologous recombination DNA repair. **BRCA1/2** mutations are rare in most populations, Ashkenazi Jewish being an exception, with one of 40 individuals carrying one of three main **BRCA**-founder mutations.

Germline mutations in **BRCA1/2** genes explain approximately 15-25% of hereditary breast cancer cases and predispose also to ovarian cancer. They confer significantly high risk for breast cancer (10-to 30-fold compared with the general population) in the families in which they segregate, however, the risk may be less excessive in families with moderate cancer history or in sporadic cases. The risk also differs between different **BRCA1/2** mutations and may be modified by other genetic or environmental risk factors aggregating in the families.

According to the recent **BRCA1** and **BRCA2** Cohort Consortium study, the cumulative risk for breast cancer by the age of 80 years is 72% for **BRCA1** mutation carriers and 69% for **BRCA2** mutation carriers. However, the participants in this study were mostly recruited from clinical genetic centers, and thus, the risk evaluations may be overestimated for patients without a family history. Antoniou et al. (2003) evaluated the cumulative breast cancer risk of unselected **BRCA1** carriers by age 70 years as being 65% and for **BRCA2** carriers 45%, and in a recent study, the cumulative risk of breast cancer at age 80 was found to be ~60% for **BRCA1/2** carriers with no affected first-degree relatives. Such individuals may, however, need similar screening and clinical management as those with a stronger family history.

Mutations in **BRCA2** are also found in men with breast cancer, whereas **BRCA1** mutations are rarer. In Finland, 8% of male breast cancer patients carry **BRCA2** mutations, and among individuals with a family history of the disease, **BRCA2** mutation can be found in ~40% of the male patients. In addition to increased breast and ovarian cancer risk, **BRCA1/2** mutation carriers are susceptible to developing some other cancers as well, such as pancreatic and prostate cancer. In **BRCA2** carriers, a risk for melanoma and colon cancer has also been observed. The risks for these cancers are, however, smaller than the risk of breast or ovarian cancer in **BRCA1/2**-mutation-positive patients; the relative risk for pancreatic cancer is approximately 3.5 – 5.9 and for prostate cancer 2.5-6.3. Biallelic mutations in **BRCA1/2** genes are known to be associated with the developmental disorder Fanconi anemia, discussed further in Section 2.4.1. Homozygosity of **BRCA1** on an embryonic level is typically lethal; one sufficient copy of the gene is essential for normal development. Studies show, however, that loss of the wild type allele later in life may lead to sensitivity to interstrand crosslinking agents and consequently tumor susceptibility, which are typical characteristics of Fanconi anemia.
The Breast Cancer Information Core (BIC) database includes more than 1,600 identified BRCA1 and 1,800 BRCA2 mutations and variants, which are evenly distributed across the coding sequences of the genes. Most inherited BRCA1/2 mutations are truncating frameshifts, nonsense mutations, or splice site alterations. Large-scale deletions and duplications have also been identified, although they are less frequent. It should be noted that while BRCA1 and BRCA2 fall into the high-risk breast cancer gene category, some identified mutations may confer rather moderate or lower increased risk.

Breast tumors in individuals with BRCA1 mutations are usually poorly differentiated (grade 3) ductal carcinomas with high mitotic count, and they are commonly triple-negative with young age at onset, whereas the tumors of BRCA2 mutation carriers are more heterogeneous and resemble those of non-carriers. Nevertheless, approximately 16-23% of BRCA2 positive breast tumors display triple-negative characteristics, and these commonly are grade two or three. Thus, altogether ~15% of unselected triple-negative breast cancers are associated with inherited BRCA1/2 mutations.

Li-Fraumeni syndrome (LFS) is a rare autosomal hereditary disorder that is highly penetrant and predisposes to a wide spectrum of solid and hematological cancers, most commonly soft tissue sarcomas, osteosarcoma, leukemia, and breast cancer. Most individuals with LFS carry mutations in the tumor suppressor gene TP53, and the risk of mutation carriers developing cancer is approximately 50% by the age of 30-31 years (females) or 46 years (males). Breast cancer is the most common cancer among Li-Fraumeni patients, and LFS accounts for approximately 1% of all female breast cancer cases.

PTEN is a tumor suppressor gene, producing a lipid- and protein phosphatase that regulates activation of the PI3K/AKT oncogenic pathway. Mutations in PTEN are associated with PTEN hamartoma tumor syndrome (PHTS). It is a spectrum of disorders caused by germline mutations in the PTEN gene. PHTS includes Cowden syndrome (CS), Bannayan-Riley-Ruvalcaba syndrome (BRRS), Proteus syndrome, and Proteus-like syndrome, all of which are characterized by a wide range of neurodevelopmental issues and benign tumors (hamartomas) affecting a variety of tissues. Most cases are inherited with an autosomal dominant pattern and high penetrance, but 10-40% may be due to de novo mutations. Among PHTS disorders, women with CS have a 30-50% lifetime risk of developing breast cancer, with an average age of diagnosis between 38 and 46 years, and a 67% lifetime risk for developing benign breast disease. Male breast cancer is also known to be associated with CS.

Another hamartomatous syndrome is Peutz-Jeghers syndrome (PJS), characterized by mucocutaneous pigmentation and gastrointestinal polyposis. The syndrome is caused by STK11 germline mutations, identified in 30-80% of PJS patients. Women with PJS have a 50% risk of developing breast cancer by the age of 60 years. STK11 is a cell metabolism, growth, and
survival-associated serine/threonine kinase, and it is also linked to DNA repair.\textsuperscript{52, 99, 100} 

\textit{CDH1} gene encodes E-cadherin, a cell-cell adhesion-regulating transmembrane glycoprotein. Mutations in \textit{CDH1} lead to compromised cell adhesion and increased cell motility and are known to predispose to hereditary diffuse gastric cancer, which is associated with increased risk for lobular breast cancer. The cumulative breast cancer risk for \textit{CDH1} mutation carriers is \textsim{40\%} by the age of 80 years.\textsuperscript{52}

\textit{PALB2} gene is functionally connected to \textit{BRCA2} in homologous recombination and DSB repair. Mutations in it may confer an approximately fivefold risk of female breast cancer compared with non-carriers. In Finland, \textit{PALB2} c.1592delT mutation has been identified in \textsim{1\%} of unselected breast cancer cases, conferring an approximately sixfold increased risk, which is similar to the risk of deleterious \textit{BRCA2} mutation carriers.\textsuperscript{86, 101, 102} Biallelic \textit{PALB2} mutations are found in a subset of Fanconi anemia cases resembling those caused by biallelic \textit{BRCA2} mutations, discussed in Section 2.4.1. In addition, \textit{PALB2} mutations are associated with male breast cancer.\textsuperscript{104} The risk of breast cancer for female \textit{PALB2} mutation carriers with two or more affected first-degree relatives is 58\%, whereas breast cancer for mutation carriers without family history is around 35\%. Altogether, \textit{PALB2} loss-of-function mutations account for approximately 2.4\% of the familial breast cancer cases, however, the estimates vary between populations.\textsuperscript{102}

\textit{CHEK2} is a tumor suppressor gene encoding a serine/threonine kinase that regulates cell division, apoptosis, and DNA repair. ATM activates \textit{CHEK2} kinase activity by phosphorylation, and it interacts with other cell cycle control proteins, including BRCA1, BRCA2, and TP53.\textsuperscript{105, 106} \textit{CHEK2} germline mutations have been connected to hereditary cancer predisposition since detecting the deleterious \textit{CHEK2} c.1100delC allele in Li-Fraumeni patients.\textsuperscript{107} Despite this finding, \textit{CHEK2} defects do not cause the syndrome, but carrying a deleterious \textit{CHEK2} mutation increases breast cancer risk by approximately 20\%, and the risk increases with number of affected first- and second-degree relatives. In addition to breast cancer patients, \textit{CHEK2} mutations have been observed in prostate, kidney, colon, and thyroid cancer patients.\textsuperscript{106, 108}

\textit{ATM} gene encodes a protein kinase involved in DSB repair. It is responsible for phosphorylating other DNA damage response and cell cycle proteins, including TP53, BRCA1, and CHEK2. Heterozygous mutations in \textit{ATM} confer an approximately two- to fivefold increased breast cancer risk, similar to \textit{CHEK2}. Biallelic \textit{ATM} mutations, either compound heterozygotes or homozygotes, cause Ataxia-telangiectasia (Louis-Bar syndrome), which is an autosomal recessive condition. The disorder is characterized by neuronal degeneration, immunodeficiency, sensitivity to ionizing radiation, and greatly increased cancer risk relative to the general population. In childhood, lymphoid cancers are common, and breast cancer is often found in adults.\textsuperscript{72, 109}
**NF1** gene product is a GTPase activating protein that regulates the RAS signaling pathway. Pathogenic **NF1** mutations are found in individuals with neurofibromatosis type 1, and mutation-positive women have an approximately sixfold increased risk of breast cancer 52.

Several other genes that may fall into the moderate-risk breast/ovarian cancer gene category have been identified in different populations. However, the evidence thus far is limited. These genes include **RAD51C, RAD51D, and BRIP1**, showing an association with ovarian cancer, as well as DNA repair genes **MRE11A, RAD50, NBN, XRCC2, RECQL, FANCC**, and also **FANCM** identified in this thesis, showing an association with breast cancer 86, 110.

**Table 1.** High- and moderate-risk breast cancer genes, their main functions in cells, associated hereditary syndromes, and cancer susceptibility.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Main function</th>
<th>Syndrome</th>
<th>Main cancer susceptibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRCA1</td>
<td>Homologous recombination</td>
<td>Fanconi anemia</td>
<td>Breast and ovarian</td>
</tr>
<tr>
<td>BRCA2</td>
<td>Homologous recombination</td>
<td>Fanconi anemia</td>
<td>Breast and ovarian</td>
</tr>
<tr>
<td>TP53</td>
<td>Several anticancer functions</td>
<td>Li-Fraumeni</td>
<td>Breast, sarcoma, leukemia</td>
</tr>
<tr>
<td>PTEN</td>
<td>AKT pathway control</td>
<td>Cowden</td>
<td>Breast, thyroid, endometrial</td>
</tr>
<tr>
<td>STK11</td>
<td>Cell polarity regulator</td>
<td>Peutz-Jeghers</td>
<td>Breast, colorectal, thyroid</td>
</tr>
<tr>
<td>CDH1</td>
<td>Cell-cell adhesion</td>
<td>Hereditary diffuse gastric cancer</td>
<td>Breast, gastric</td>
</tr>
<tr>
<td>PALB2</td>
<td>Homologous recombination</td>
<td>Fanconi anemia</td>
<td>Breast, pancreatic</td>
</tr>
<tr>
<td>CHEK2</td>
<td>Cell cycle checkpoint</td>
<td></td>
<td>Breast, prostate</td>
</tr>
<tr>
<td>ATM</td>
<td>DNA repair</td>
<td>Ataxia telangiectasia</td>
<td>Breast, lymphoma</td>
</tr>
<tr>
<td>NF1</td>
<td>Ras pathway control</td>
<td>Neurofibromatosis type 1</td>
<td>Breast, brain, leukemia</td>
</tr>
</tbody>
</table>

**2.2.3.2 Low-risk breast cancer alleles**

To date, around 200 common variants associated with breast cancer have been identified with large genome-wide studies. These variants are mainly single-nucleotide polymorphisms (SNPs) occurring often in non-coding sequences and enriched in distal regulatory elements of the genome 28. Such variants commonly have minor allele frequency >1% and usually confer ~1.5-fold increased breast cancer risk compared with the general population, however, their co-existence with high- or moderate-risk breast cancer alleles combined with lifestyle factors and family history affect the absolute individual cancer risk 86. The identified common susceptibility loci explain an estimated 18% of the familial relative risk. These SNPs can be incorporated into polygenic risk prediction models to identify women at increased breast cancer risk. Furthermore, common SNPs can modify the risks associated with moderate- and high-penetrance mutations 28, 111.
2.2.4 Finnish founder mutations

Some inherited mutations may appear in certain populations with higher frequencies. Founder mutations commonly arise from restricted and inbred populations, allowing accumulation of the mutation and simultaneously limiting its spreading to other populations. Founder mutations are often encountered, for example, among Ashkenazi Jews and in Iceland. The Finnish population is one of the most studied genetic isolates in the Western world, providing considerable advantages for human genetic research. Finland has most likely been inhabited since the last glacial period, with two main migratory waves 4,000 and 2,000 years ago. The Finnish gene pool has been shaped by several typical features of founder effects such as small number of the original inhabitants, isolated geographical location, regional sub-isolates, and genetic drift, creating a unique disease heritage. Perceptions of Finland’s genotype distribution have been supported by analyses of the paternal Y chromosome haplotypes and maternal mitochondrial sequences demonstrating small genetic diversity among the Finnish population compared with other European populations. In addition to well-recorded population history, comprehensive healthcare registries have facilitated extensive genetic research.

Approximately 40 rare inherited disorders (the so-called Finnish disease heritage) are known to be more prevalent in Finland than elsewhere. Most of these diseases have an autosomal recessive inheritance caused by one major mutation. Recurrent founder mutations have also been seen in Finnish cancer patients, including breast and ovarian cancer patients.

2.2.4.1 Finnish breast cancer founder mutations

A single founder mutation in the Finnish population has been identified in the PALB2 gene. The minor allele frequency of the c.1592delT mutation in the general population in Finland is 0.2%, whereas the frequency in other European populations is 0.0008%, according to the gnomAD database. The c.1592delT frameshift mutation in exon 4 creates a premature stop codon on Leu531, leading to a sixfold increased risk for breast cancer and explaining approximately 1% of all breast cancer cases in the Finnish population. This is comparable with the risk associated with BRCA2 mutations; however, the average age at the diagnosis of breast cancer for c.1592delT carriers is slightly higher than for BRCA2 mutation-positive patients in Finland, yet noticeably lower than for individuals with sporadic breast cancer.

A heterozygous CHEK2 c.1100delC frameshift mutation at cytosine residue at position 381 predisposes to familial breast cancer. The frequency of the mutation differs between populations, but seems to be highest in Finland (1.4%) and in the Netherlands (1.3-1.6%) in the respective
studies. Homozygous c.1100delC mutation increases the risk of breast cancer by twofold compared with heterozygous carriers\textsuperscript{125}. In addition, \textit{CHEK2}\textsuperscript{c.1100delC} mutation is associated with an increased male breast cancer risk, with an observed frequency of 5.9\% among male breast cancer patients in Finland. The frequency is very similar to that of female patients (5.5\%)\textsuperscript{126}. Iceland has another isolated founder population in which an unusual situation is seen; a single \textit{BRCA2} mutation, c.999del5, with an estimated 0.5\% frequency explains approximately 40\% of the familial breast cancer risk in Iceland, whereas \textit{BRCA1} mutations are very rare\textsuperscript{114}. In contrast, the mutation spectrum in \textit{BRCA1/2}-positive breast cancer patients is quite wide in the Finnish population. Approximately 30 Finnish founder mutations have been identified in \textit{BRCA1/2} genes. Some of these are present in other populations, while others are exclusively of Finnish origin and in addition, have regional clustering, such as the c.4261nt-2A>G mutation. This is one of the most frequent \textit{BRCA1} mutations in Finland, unique to the Finnish population, although it has started to spread recently, i.e. less than 10 generations ago\textsuperscript{127, 128}. The c.995del5 mutation common in Iceland is also the most frequent \textit{BRCA2} mutation among Finnish breast cancer patients, but unique Finnish founder mutations have been identified in this gene also \textsuperscript{129, 130}.

### 2.2.5 Treatment of breast cancer

Although breast cancer incidence is increasing worldwide, death rates have been declining in developed countries due to improvements in treatment and early detection\textsuperscript{2, 131}. In Finland, around 850 women die of breast cancer annually, yet the 5-year breast cancer survival rate is approximately 90\%, which is among the highest in Europe\textsuperscript{132}. Surgery is the primary treatment of breast cancer. The whole breast may be removed, or in the breast-conserving method the tumor tissue from the breast and possible metastases from regional lymph nodes are excised. Radiotherapy administered to the site of the tumor and lymph nodes after breast-conserving surgery halves the risk of recurrence and reduces the cancer death rate by ~17\%\textsuperscript{133, 134}. Several systemic adjuvant treatment options are available after surgery. These are administered based on the risk of disease recurrence and on biological, molecular, and histological characteristics of the tumor. Chemotherapy is most commonly applied as a combination treatment. Anthracycline- or anthracycline/taxane-based polychemotherapy are estimated to reduce mortality in early breast cancer patients by 36\% and 44\%, respectively. However, side effects of chemotherapy are common, and these may limit the treatment options for patients with comorbidity, high age, or poor overall condition\textsuperscript{135-136}. HER2-positive tumors can be effectively treated with trastuzumab (Herceptin), a humanized monoclonal antibody blocking the HER2 function
in cells \(^{136}\). A relatively new and promising group of targeted cancer therapies are PARP inhibitors, already in clinical use in the treatment of recurrent ovarian cancer, but the role of PARPi therapy in breast cancer is still undefined \(^{137}, 138\). PARP inhibitors block the enzyme poly(ADP-ribose) polymerase, which is required for the early steps of specific cellular DNA repair pathways. In the current view, the mechanism of PARPi is assumed to rely on trapping PARP1 enzyme on DNA to prevent PARylation, interfering with PARPi functions, and causing severe DNA damage such as blocked replication forks that cannot be repaired with defective DNA repair mechanisms. This cytotoxic effect leads to death of the tumor cell. The synthetic lethality of PARPi-treatment does not affect normal cells, which do not replicate as often as tumor cells and also do not lack the repair capacity of \(BRCA1/2\) or other DSB repair-related genes \(^{137}, 139, 140\).

Adjuvant hormone therapies can be included in the treatment regimen if hormone receptor-positive cells are found in the tumor. Around 60-70% of all breast cancers are ER-positive, and they are commonly treated with five-year tamoxifen or an aromatase inhibitor hormone therapy, depending on the menopausal status of the patient. For premenopausal patients, an additional five years of tamoxifen treatment is often beneficial. Post-menopausal patients are commonly treated with aromatase inhibitors. Hormone therapies have significantly reduced breast cancer mortality, however, they may have serious side effects; tamoxifen increases the risk of endometrial cancer and thromboembolism, especially with extended treatment. Aromatase inhibitors may cause osteoporosis \(^{62, 65, 141}\).

Among women with inherited \(BRCA1/2\) mutations, development of secondary ipsi- or contralateral breast cancer is more likely and bilateral mastectomy is often recommended. It reduces the lifetime risk of breast cancer by 90-95%. Oncological treatment for familial breast cancer patients is similar to that of sporadic cases, however, for individuals with high-risk mutations more intensive screening is usually recommended \(^{134}, 142\). Genetic testing for inherited mutations offers the opportunity for risk-reducing intervention for individuals at higher risk. In Finland, annual follow-up visits are recommended for high-risk mutation carriers from the age of 25 years. Follow-up lasts for the lifetime of the mutation carrier unless risk-reducing mastectomy is performed \(^{143}\).

Breast cancer screening has proven effective for detecting cancers and decreasing cancer mortality in developed countries. However, a balance between harms and benefits of the mammography screenings is still being sought, as overdiagnosis and false-positive findings may lead to overtreatment, emotional stress, and increased radiation dose predisposing to cancer \(^{143}, 144\). In Finland, the national screening program started in 1987, and currently invites all women aged 50-69 years to mammography screening every second year. Based on a recent study, the organized screening program in Finland has decreased breast cancer mortality by 33% in women attending the program \(^{132}, 145, 146\).
2.2.6 Breast cancer survival

Survival of breast cancer is dependent on tumor progression and metastasis. Most recurrences happen within five years, and the most common reason for cancer death is metastasized disease \(^6,68\). Prognostic factors for breast cancer are described in Section 2.2.1. However, it is difficult to determine the extent to which the disease outcome is influenced by differences in tumor characteristics, biological and genetic differences in patients, or different treatments received. In general, younger breast cancer patients (<40 years) have shown worse survival than older patients \(^{147}\). Breast cancer in young women is often diagnosed at an advanced stage, resulting in unfavorable prognosis. Recurrence of the disease is also seen especially among young patients. In addition, triple-negative disease associated with poor prognosis is more common among young or premenopausal women \(^{70,87}\).

Beyond inheriting the risk for breast cancer, it has also been been debated whether hereditary factors may be associated with survival of the disease. In a population-based cohort study in Sweden, the survival of women with familial breast cancer predicted the prognosis of a first-degree relative diagnosed with breast cancer. The concordance in prognosis was more evident in women with first-degree family members diagnosed with cancer at a young age \(^{148}\). Family members also appeared to share the good prognosis, even in cases where the mother was diagnosed prior to her daughters, indicating that the prognosis is not an effect of awareness of the familial predisposition and knowledge of the daughter’s future diagnosis \(^{149}\). However, estimating the extent to which an certain inherited (or somatic) mutation is associated with breast cancer outcome is challenging. Conflicting results have emerged in studies concerning, for example, BRCA1/2 mutation carriers. Many individual studies confirm the poor disease outcome of BRCA1/2-positive patients, yet in the meta-analysis the survival results are inconclusive, showing at best only moderate evidence for worse disease outcome in BRCA1 carriers \(^{150,151}\).

Several factors may influence outcome of the survival analyses; many studies are rather small and have inadequate power to show a true concordance. They often have insufficient data on tumor characteristics affecting prognosis, such as grade and stage, or on features that may affect a tumor’s metastatic abilities such as the patient’s access to treatments and dose response. Inherited differences in metabolism may explain the similarities in drug response, or family members can share a genetic predisposition to metastatic disease. Such confounding factors require large datasets for analysis and careful selection of familial patients \(^{150,152,153}\).
2.3 OVARIAN CANCER

2.3.1 Epidemiology of ovarian cancer

Ovarian cancer is the seventh most common cancer in women worldwide, with approximately 240,000 new diagnosed cases annually \(^2\). In Finland, more than 500 cases are diagnosed each year \(^54\). With a five-year survival rate below 45%, the overall prognosis of ovarian cancer is poor, and it is the eighth most common cause of cancer death among women \(^2, 154\). The high mortality rates are explained by the absence of specific symptoms followed by late diagnosis, unfavorable intra-abdominal location of the ovary, and molecular heterogeneity of ovarian tumors \(^155\).

The pathogenesis of ovarian cancer has been much debated. Generally, ovarian carcinoma was thought to originate from ovarian epithelial cells. Current evidence shows, however, that cancer in most cases seemingly arises from the fallopian tube epithelium \(^154, 156\). The fallopian tube-type cancer is noticeably elevated in patients with BRCA1/2 mutations, and the histology of most ovarian cancers is similar to that of the fallopian tube \(^155, 157\).

Epithelial ovarian tumors are classified according to their histology. High-grade serous (70%), endometrioid (10%), clear cell (10%), mucinous (3%), and low-grade serous (<5%) subtypes account for 98% of ovarian carcinomas. Rarer histological subtypes of ovarian cancer include transitional carcinoma, malignant Brenner tumor, and mixed epithelial carcinoma \(^158\). They are characteristically distinct diseases, differing in risk factors, precursor lesions, spreading patterns, molecular events during cancer progression, prognosis, and treatment response \(^158, 159\).

The staging of the ovarian cancer at diagnosis is one of the most important prognostic factors. Staging is based on the Fédération Internationale de Gynécologie et d’Obstetrique (FIGO) system (I-IV). A high proportion of patients are diagnosed at advanced stages III and IV, explaining the high mortality of the disease. Other important prognostic factors are histological type and grade, tumor size, and patient’s age \(^160\).

Ovarian cancer screening does not improve disease outcome \(^161\). Women at high risk may consider risk-reducing salpingo-oophorectomy, followed by hormone replacement therapy. Also, tubal ligation and use of oral contraceptives may reduce the risk \(^155, 161\).

The primary treatment of ovarian cancer is surgical reduction of the tumor, commonly followed by adjuvant platinum and taxane-based chemotherapy. Patient prognosis and survival are significantly associated with outcome of surgery and amount of residual disease. For advanced ovarian cancer (either high-grade primary disease or relapse), bevacizumab in combination with chemotherapy may improve survival. Unfortunately, after the primary therapy, the majority of patients with ovarian cancer will
experience a tumor recurrence and develop chemotherapy resistance, leading to death\textsuperscript{162, 163, 164}.

Novel targeted therapies combined with chemotherapy and alone have improved the treatment of ovarian cancer. A large group of patients with defects in the homologous recombination repair pathway (such as \textit{BRCA1/2} mutation carriers), may benefit from PARPi therapy; this is especially the case for those with platinum-sensitive disease\textsuperscript{164}.

\subsection*{2.3.2 Genetics of ovarian cancer}

The most common risk factors for ovarian cancer are nulliparity, endometriosis, menopausal hormone therapy, and overweight\textsuperscript{154}. However, the strongest risk factor for ovarian cancer is family history of ovarian and/or breast cancer. Inherited loss-of-function mutations in \textit{BRCA1} and \textit{BRCA2} confer a lifetime risk for ovarian carcinoma ranging between 20\% and 50\%. The risk is considerably increased among women with one affected first-degree relative, especially if that individual was diagnosed before the age of 50 years\textsuperscript{76, 77, 165, 166}. Hereditary ovarian cancer tumors are commonly high-grade serous type and develop primarily from the fimbriae of the fallopian tube. Conversely, low-grade serous ovarian cancers and serous borderline tumors are usually not associated with \textit{BRCA1/2} germline mutations and have a lower frequency of somatic \textit{TP53} and \textit{BRCA1/2} mutations as well\textsuperscript{161}. Lynch syndrome (LS), also known as hereditary nonpolyposis colorectal cancer (HNPCC), is an autosomal dominant cancer predisposition syndrome caused by mutations in the mismatch repair genes \textit{MLH1}, \textit{MSH2}, \textit{MSH6}, and \textit{PMS2}\textsuperscript{161, 167, 168}. LS is also associated with an increased risk for ovarian cancer, mainly among \textit{MLH1} and \textit{MSH2} mutation carriers (8-15\% risk for ovarian cancer) and to a lesser amount among \textit{MSH6} and \textit{PMS2} mutation carriers\textsuperscript{167}. Most cases have mixed histology. The overall ovarian cancer survival is better in women with LS than in \textit{BRCA1} or \textit{BRCA2} mutation carriers, based on earlier stage of diagnosis and nonserous subtype of the disease\textsuperscript{161, 169}.

In addition to \textit{BRCA1}, \textit{BRCA2}, and LS-associated genes, mutations in \textit{BRIP1}\textsuperscript{170}, \textit{RAD51C}\textsuperscript{171}, and \textit{RAD51D}\textsuperscript{172} are associated with an increased ovarian cancer risk. Several other genes, including \textit{FANCM}\textsuperscript{173}, \textit{MRE11A}, \textit{NBN}, \textit{RAD50}, \textit{CHEK2}, \textit{BARD1}, and \textit{PALB2}\textsuperscript{165}, have also been connected to ovarian cancer, however, conclusive evidence of the firm association has not yet been demonstrated. Genome-wide association studies have also identified several variants conferring a low relative risk of ovarian cancer\textsuperscript{161}. Identification of the mutation of the ovarian cancer patient is critical for diagnosis and treatment. For patients without a family history of the disease, identification of a specific mutation may support the more intensive treatment usually recommended for familial patients\textsuperscript{174}.
2.4 FANCM

FANCM (Fanconi anemia complementation group M), an ortholog of the archaeal DNA repair protein Hef, is an evolutionarily conserved ATP-dependent DNA binding translocase. FANCM, located in chromosome 14q21.2, is a member of the Fanconi anemia (FA) complementation group of genes that maintain genome stability throughout the cell cycle, principally by activating DNA repair in cells. Mutations in FA genes are associated with increased hypersensitivity to DNA crosslinking agents such as mitomycin C, and increased levels of chromosomal instability.

To date, 21 associated FA proteins (A, B, C, D1, D2, E, F, G, I, J, L, M, N, O, P, Q, R, S, T, U, V) have been identified as part of the FA pathway. Identification is commonly based on the molecular function of the proteins or manifestation of FA symptoms when the particular FA gene is mutated. The upstream part of the FA pathway consisting of nine subunits (FANCA/B/C/E/F/G/L/M/T) activates the FANCI–FANCD2 heterodimer through monoubiquitination during S-phase together with FAAP24 (homolog of ERRC1) and histone-fold protein complex MHF. Ubiquitination activates the downstream path of the remaining FA proteins regulating recombinational and nucleolytic reactions to complete DNA repair with homologous recombination (Figure 4). Homologous recombination repair is further discussed in Section 2.1.4.4.

Heterozygous mutations in several FA genes, including FANCD1/BRCA2, FANCN/PALB2, FANCI/BRIP1, FANCO/RAD51C, and FANCS/BRCA1, are associated with increased risk for breast and ovarian cancer. In the studies performed for this thesis, monoallelic mutations in the FA pathway gene FANCM were also identified in breast cancer patients.
Figure 4. Overview of the main steps of the FA repair pathway. FANCM, FAAP24, and MHF complex bind and unwind DNA at the double-strand break site. Accumulation of ss-DNA binding protein RPA activates ATR kinase-dependent checkpoint response. ATR phosphorylates MRN complex and FANCD2-FANCI heterodimer, which is associated with DSB resection coordinating CtIP. These actions are required for early stages of repair and damage signaling.

FANCM, FAAP24, and MHF bind the assembling FA core complex to DNA, resulting in FANCD2-FANCI monoubiquitination by FANCL and FANCT (UBE2T). FANCD2-I heterodimer is directed to the damage site, where it is associated with structure-specific endonucleases MUS81-EME1 and ERCC1-FANCQ/XPF interacting with FANCQ/SLX4. The core complex and FANCD2-I heterodimer assists the intermediate repair with TLS polymerases and homologous recombination repair proteins 179, 181.

2.4.1 Fanconi anemia

Biallelic mutations in certain FA pathway genes cause an autosomal (or seldom X-linked if associated with FANCB mutations) heterogeneous disorder characterized by bone marrow failure, congenital abnormalities, and cancer predisposition. The disease is very rare, with an incidence of 1-5 per one million births. Due to the wide range of symptoms, Fanconi anemia has been a successful model system for studying biological processes such as DNA repair and cancer progression 178, 189. On the cellular level, FA is a chromosomal instability disorder in which the cells of the patient rapidly accumulate DNA damage as a result to hypersensitivity to DNA interstrand crosslink (ICL) -creating agents. ICLs affect both strands of the DNA helix. They form covalent links that prevent DNA from unwinding, blocking DNA
replication and translation. These lesions are difficult to repair, and it is possible that the FA pathway has developed in high eukaryotes to assess precisely this type of DNA damage. In addition to clinical symptoms, the disease is commonly diagnosed if ICL hypersensitivity is observed at the cellular level \cite{178,189,190}.

Most FA patients harbor mutations in the upstream core complex, particularly in \textit{FANCA}. A major health threat for adult FA patients is cancer, most commonly acute myelogenous leukemia, but also squamous cell, liver, brain, skin, and renal tumors. Notably, most of the breast/ovarian susceptibility genes in the FA pathway identified so far are not included in the FA core complex, and FA patients with mutations in the core complex rarely develop breast or ovarian tumors. This may be a result of FA patients often being sterile, and female patients may have incomplete estrogen production \cite{189,191}. The symptoms of the disease vary greatly between the associated genes, and the subtyping of newly diagnosed patients is challenging. FA patients with \textit{FANCA} subtype usually have mild symptoms with later onset of bone marrow failure, whereas \textit{FANCC} and \textit{FANCG} subtypes are more severe. Rare \textit{BRCA1}/\textit{FANCS} mutation carriers with symptoms resembling those of Fanconi anemia usually display only so-called FA-like syndrome, lacking the bone marrow failure typical of most FA subtypes. Assessing \textit{BRCA1}/\textit{FANCS} as a FA disease gene has been difficult because rare biallelic germline mutation carriers typically do not live long enough to demonstrate the FA phenotype \cite{83,189,191,179}.

ICL agents have commonly been used as cancer chemotherapeutic drugs, but due to hypersensitivity to these agents the treatment options for cancer in FA patients are limited. Radio- and chemotherapy must be applied in lower doses to minimize toxic or even lethal side effects. Management of cancer should include careful monitoring of the disease occurrence and prevention. Some FA subtypes may alter patients’ susceptibility to certain types of cancers and identifying the genetic variation specific to cancer cells could enable more precise treatment \cite{189,191}.

2.4.2 Cellular functions of FANCM

Among the 21 FA proteins identified thus far, FANCM appears as a distinct component of the pathway with respect to its several independent functions in cells. FANCM was originally identified as FAAP250, a 250kDA polypeptide that coimmunoprecipitated with antibodies against multiple FA core complex components. \textit{FAAP250} was assigned as a Fanconi anemia complementation group gene, as a study patient (EUFA867) with Fanconi anemia carried biallelic mutations in \textit{FAAP250} and lacked the expression of the gene \cite{175}. To date, this is the only Fanconi anemia patient with \textit{FANCM} mutations described. However, the same patient also carries biallelic mutations in \textit{FANCA} gene \cite{175,192}. Furthermore, individuals without any symptoms of FA
are found to carry homozygous loss-of-function mutations in FANCM \(^{192, 193}\). It is therefore likely that FANCM is not associated with the FA disease per se, but in the absence of FANCM the nuclear localization and stability of several FA-associated proteins are altered, affecting DNA repair and genomic instability \(^{175, 194, 195}\).

FANCM is regulated by phosphorylation. During S-phase the phosphorylation level of the protein is restrained, but it increases during mitosis. This is correlated with the ubiquitination levels of the FANCD2. FANCM also appears hyperphosphorylated in response to DNA damage, e.g. when inducing ICL-causing agents. The N-terminal domain of the protein with translocase activity participates in localizing the FA core complex to the chromatin at the damage site, which is followed by the ubiquitination of the FANCD2 and repairing of the ICLs. Furthermore, the ATPase activity of FANCM appears to have a role in the repair reactions and cell cycle arrest \(^{175, 182, 194}\). The ERRC domain in the C-terminal end of the protein is responsible for targeting and recognizing branched DNA molecules. It is likely that the protein domains of FANCM act sequentially and co-operative in the activation of DNA repair \(^{194}\).

In addition to ICL resolution, FANCM has several other functions in cells. Studies in human and chicken DT40 cells have revealed that the ATPase activity of FANCM is needed to prevent replication forks from stalling during DNA replication in normal growth conditions. FANCM can further activate the ATR-mediated DNA damage checkpoint stabilizing replication forks. These functions of the FANCM protein appear to be independent of the other FA genes \(^{196-199}\). Furthermore, during homologous recombination repair the DNA translocase activity of FANCM and FAAP24 promote anti-crossover functions while resolving the Holliday junctions. Crossing over may lead to chromosome translocation and loss of heterozygosity, and therefore, the efforts of FANCM and FAAP24, again independently of other FA genes, may help to safeguard the integrity of the genome during HR repair. The independent role of FANCM is further supported by studies showing that lower eukaryotes lack many of the FA proteins, yet FANCM is conserved among them \(^{196, 199, 200}\).

The Fancm mouse models support the perception of FANCM being a tumor suppressor gene. Fancm-deficient mice showed gonadal abnormalities, increased chromosomal breakage, and crosslinker hypersensitivity. These features are typical also for other FA mouse models, but Fancm\(^{ΔαΔβ}\) mice additionally exhibit reduced life span and increased cancer incidence compared with littermate controls, as well as residual FANCD2 monoubiquitination and increased spontaneous sister chromatid exchanges, leading to the conclusion that with the absence of FANCM the stability of the genome is not properly maintained \(^{192}\).
3 AIMS OF THE STUDY

1. To identify novel susceptibility alleles for breast and ovarian cancer in the Finnish population.

2. To characterize such mutations by evaluating the risk of breast and/or ovarian cancer in large case-control datasets.

3. To investigate the role of FANCM c.5101C>T nonsense mutation (identified in Study I) in breast cancer prognosis by studying tumor phenotype, patient survival, and treatment outcome.
4 MATERIALS AND METHODS

4.1 SAMPLES

4.1.1 Breast cancer patient samples

Identifying new breast and/or ovarian cancer susceptibility alleles was investigated by genotyping germline DNA samples from unselected and familial breast cancer patients as well as from healthy population controls (Table 2) and statistically evaluating the risk (I,III). All breast cancer samples from each study are genomic DNA isolated from peripheral blood. Tumor phenotype, patient survival, and treatment outcome were examined based on histopathological and cancer treatment information and follow-up data of study participants (II). Additional samples from Finland (Kuopio, Oulu, and Tampere) and from Iceland were collected by the research groups from the respective facilities and used in collaboration in this thesis.

Table 2. Number of breast cancer and population control samples from Finnish studies used in this thesis, and respective collection periods. BC = breast cancer. The number of samples used in each analysis are denoted in respective tables in the Results Section.

<table>
<thead>
<tr>
<th>Study cohort</th>
<th>N (BC)</th>
<th>Collection period(s)</th>
<th>N (control samples)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Familial: ongoing</td>
<td></td>
</tr>
<tr>
<td>Oulu</td>
<td>1419</td>
<td>2000 – 2014</td>
<td>511</td>
</tr>
<tr>
<td>Kuopio</td>
<td>516</td>
<td>1990 – 1995</td>
<td>162</td>
</tr>
</tbody>
</table>

4.1.1.1 Helsinki series

The breast cancer patient series from the Helsinki area consisted of unselected and familial patients, recruited from the Helsinki University Central Hospital over several periods. In 1997 – 1998 and 2000, 884 unselected samples were collected at the Department of Oncology. In addition, 986 samples were collected at the Department of Surgery in 2001 – 2004. Of these, 397 cases had a familial background. Altogether 1,730 unselected patients were included in Study I, 1,712 in Study II, and 1,699 in Study III.

Additional cases with a family history of breast cancer were derived from the Departments of Oncology and Clinical Genetics. After combining additional familial cases and the cases from unselected series with familial
predisposition, 524 patients had a strong family background with at least three breast or ovarian cancers among first- or second-degree relatives (including the proband). These patients had tested negative for \textit{BRCA1/2} mutations. In addition, 568 patients had one first-degree relative affected with breast or ovarian cancer, and these patients had tested negative for Finnish \textit{BRCA1/2} founder mutations\cite{123, 129, 205}. Altogether 1,074 familial patients were included in Study I, 1,006 in Study II, and 1,078 in Study III.

All patient genealogies were confirmed with population registries or hospital records. Cancer diagnoses were retrieved from hospital records and the Finnish Cancer Registry. Pathology reports were used to collect ER and PR hormone receptor status (positive when >10% of cells were stained) and tumor histology information. HER2 status was ascertained with immunohistochemistry and gene amplification\cite{130, 206, 207}. Information on breast cancer death was obtained from the Finnish Cancer Registry.

\subsection*{4.1.1.2 Tampere series}

The unselected breast cancer patient series including 408 samples from the Tampere area was collected at Tampere University Hospital in 1997 – 1999\cite{201}. An additional 336 incident cases were derived in 1996 - 2004\cite{203}. ER and PR hormone receptor statuses (positive when >10% of cells were stained), HER2 status, and other clinicopathological information of the study participants were obtained from patient and pathology reports. Information on breast cancer death was retrieved from the Finnish Cancer Registry. A total of 674 samples were included in Study I, 650 in Study II, and 662 in Study III.

\subsection*{4.1.1.3 Oulu series}

The unselected breast cancer samples from Northern Finland were collected in 2000 – 2014 from patients operated on at Oulu University Hospital. Study participants were unselected for age at disease onset and family history of cancer. In total, 516 incident invasive cases were included in Study II. In Study III, 1,147 unselected, 153 familial, and 56 young breast cancer patients were included in the analyses. Familial patients were affected \textit{BRCA1/2}-negative individuals from breast and ovarian cancer families and the young breast cancer series consisted of breast cancer patients unselected for family history of cancer but with early onset of the disease (≤40 years), suggesting a possible hereditary predisposition\cite{208-210}.

ER and PR hormone receptor statuses (positive when >10% of cells were stained) and tumor histology information were collected from pathology reports\cite{211, 212}. HER2 status was obtained with immunohistochemistry.
analyses. Information on breast cancer death of study participants was received from Oulu University Hospital.

4.1.4 Kuopio series

Breast cancer samples from the Kuopio area were collected in 1990 – 1995 from women entering Kuopio University Hospital due to breast symptoms. Altogether 1,919 women were invited to take part in the Kuopio Breast Cancer Project and 516 were eventually diagnosed with breast cancer. Of these, 430 invasive breast cancer patients were included in Studies II and III. Clinicopathological features of the breast tumors, surgical and oncological treatments, and follow-up times were obtained from hospital registries. ER and PR hormone receptor statuses were classified as positive when >10% of cells were stained and HER2 statuses was determined with immunohistochemistry 213, 214.

4.1.5 Icelandic breast cancer samples

Altogether 965 unselected breast cancer patients from Iceland, diagnosed at the Landspitali University Hospital, Reykjavik, during 1987 – 2004, were included in Study I. Of these, 92 patients had a family history of breast cancer (first- and/or second-degree relative diagnosed with breast cancer). Patients tested negative for Icelandic BRCA1/2 founder mutations as previously described 215. Patient genealogies were provided by the Genetic Committee of the University of Iceland, and cancer diagnoses by the Department of Pathology of Landspitali University Hospital and the Icelandic Cancer Registry.

4.2 Ovarian cancer patient samples

The unselected ovarian cancer series consists of distinct DNA and tumor samples collected at the Department of Obstetrics and Gynecology of Helsinki University Central Hospital. Invasive epithelial ovarian carcinoma patients were treated at the Helsinki University Central Hospital during 1989 – 1998, and 233 blood samples were collected during routine follow-up visits to the clinic in 1998 130. Additional blood and tumor DNA samples were collected between 1998 and 2006 from ovarian cancer patients treated at the Department of Obstetrics and Gynecology of Helsinki University Central Hospital. Of 548 samples included in Study I, 345 were genomic and 204 tumor DNA. In Study III, of the 526 samples 408 were genomic and 118 tumor DNA.
4.1.3 Population controls

Population controls consisted of cancer-free blood donors from designated areas of Finland (Helsinki, Tampere, Kuopio, and Oulu). A total of 1,271 peripheral blood samples from donors around Helsinki and 809 samples from the Tampere area were used for analyses in Study I. In Study III, 1,258 control samples from Helsinki, 808 from Tampere, 158 from Kuopio, and 510 from Oulu were genotyped for case-control analyses.

4.1.4 Ethical standards

All procedures studies involving human participants were performed in accordance with the ethical standards of the institutional research committees and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. All participating patients and their family members signed an informed consent.

4.2 METHODS

4.2.1 Exome sequencing (I,III)

In Study I, 24 BRCA1/2-negative female breast cancer patients from eleven breast/ovarian cancer families were selected for exome sequencing. Two cases from each of nine families and three cases from two families were chosen. All patients had at least three breast or ovarian cancer cases in first- or second-degree relatives. In Study III, exome sequencing was performed for 28 patients from 24 breast and uterine cancer families and 16 patients with a family history of female and male breast cancer.

Exomic regions were captured from 3 µg of genomic DNA with the Agilent SureSelect Human All Exon 50-Mb kit (I) or with Roche Nimblegen SeqCap EZ Exome v3 (III). The sequencing was performed on Illumina HiSeq2000 sequencer with 100-bp paired-end reads. All exome sequencing and sequence raw data quality checking, alignment, and variant calling were carried out at Genome Quebec Innovation Centre, Montreal, Canada.

4.2.1.1 Variant filtering (I)

Plausible breast or ovarian cancer variants identified in exome sequencing were prioritized to be further validated and genotyped in the large case-
control datasets. Only frameshift deletions and insertions, missense and nonsense variants predicted to be pathogenic, and splicing alterations located in the DNA repair genes were included in the further analysis. DNA repair gene definition is based on the Gene Ontology project data from AmiGO browser. Annovar was used to annotate the selected variants. Variants with mean read coverage of less than 15 and population allele frequency > 1% were excluded. The frequencies were based on the data of 1000 Genomes and Exome Variant Server as well as population-matched exome sequenced in-house control patients. Altogether 22 variants from 21 DNA repair genes (Table 3) were selected and manually examined to visualize and verify their structure and position with Interactive Genomics Viewer and in-house variant visualization software BasePlayer.

**Table 3.** Variants selected from exome sequencing for further analysis. Genomic positions are indicated according to genome build hg19. ExAC frequency includes non-Finnish European individuals. SISu frequency refers to the Sequencing Initiative Suomi search engine (SISu), Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Finland (http://sisuproject.fi), SISU v4.1. Information as of Aug2017.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Variant</th>
<th>Position</th>
<th>Rs-number</th>
<th>Protein level</th>
<th>SISu freq</th>
<th>ExAC freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATM</td>
<td>c.7547T&gt;G</td>
<td>11:108202202</td>
<td>rs774312539</td>
<td>p.Phe2516Cys</td>
<td>0.00038</td>
<td>0.00003</td>
</tr>
<tr>
<td>CDK9</td>
<td>c.586G&gt;A</td>
<td>9:130549857</td>
<td>-</td>
<td>p.Val196Ile</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DCLRE1</td>
<td>c.412C&gt;T</td>
<td>10:115612530</td>
<td>rs41292634</td>
<td>p.Arg138*</td>
<td>0.00226</td>
<td>0.00375</td>
</tr>
<tr>
<td>DDX10</td>
<td>c.973G&gt;T</td>
<td>11:108559787</td>
<td>-</td>
<td>p.Glu325*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>FANCA</td>
<td>c.1682C&gt;T</td>
<td>16:89846310</td>
<td>rs148154682</td>
<td>p.Thr561Met</td>
<td>0.00057</td>
<td>0.00001</td>
</tr>
<tr>
<td>MCM7</td>
<td>c.748G&gt;C</td>
<td>11:108559787</td>
<td>-</td>
<td>p.Glu250Gln</td>
<td>0.00012</td>
<td>-</td>
</tr>
<tr>
<td>NEIL3</td>
<td>c.943C&gt;T</td>
<td>4:178272607</td>
<td>rs141868607</td>
<td>p.Arg315Trp</td>
<td>0.00048</td>
<td>0.00042</td>
</tr>
<tr>
<td>PARP3</td>
<td>c.1180A&gt;G</td>
<td>3:51980263</td>
<td>rs751503675</td>
<td>p.Ile394Val</td>
<td>0.00005</td>
<td>0.00002</td>
</tr>
<tr>
<td>POLG2</td>
<td>c.590T&gt;C</td>
<td>17:62489111</td>
<td>rs139282177</td>
<td>p.Leu197Pro</td>
<td>0.00019</td>
<td>0.00117</td>
</tr>
<tr>
<td>RAD54B</td>
<td>c.904G&gt;C</td>
<td>8:95416345</td>
<td>-</td>
<td>p.Glu302Gln</td>
<td>0.00010</td>
<td>-</td>
</tr>
<tr>
<td>RBBP8</td>
<td>c.2162A&gt;G</td>
<td>18:20581567</td>
<td>rs764171110</td>
<td>p.Asn721Ser</td>
<td>0.00100</td>
<td>0</td>
</tr>
<tr>
<td>SMC1B</td>
<td>c.3056C&gt;A</td>
<td>22:45750901</td>
<td>-</td>
<td>p.Pro1019Gln</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TREX1</td>
<td>c.139G&gt;A</td>
<td>3:48508028</td>
<td>-</td>
<td>p.Gly47Ser</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SHPRH</td>
<td>c.3577_3580 delCTTA</td>
<td>6:14624394</td>
<td>-</td>
<td>p.Leu119Glnfs*7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PAXIP1</td>
<td>c.803C&gt;T</td>
<td>7:154767677</td>
<td>rs61752015</td>
<td>p.Pro268Leu</td>
<td>0.00425</td>
<td>0.00086</td>
</tr>
<tr>
<td>BARD1</td>
<td>c.2282G&gt;A</td>
<td>2:215593452</td>
<td>rs142155101</td>
<td>p.Ser761Asn</td>
<td>0.00896</td>
<td>0.00209</td>
</tr>
<tr>
<td>RUVBL1</td>
<td>c.950G&gt;A</td>
<td>3:127816209</td>
<td>-</td>
<td>p.Arg316His</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SLX4</td>
<td>c.2484G&gt;C</td>
<td>16:3641155</td>
<td>rs199656607</td>
<td>p.Glu828Asp</td>
<td>0.00115</td>
<td>0.00007</td>
</tr>
<tr>
<td>FANCA</td>
<td>c.4228T&gt;G</td>
<td>16:89805322</td>
<td>-</td>
<td>p.Cys1410Gly</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>FANCM</td>
<td>c.5101C&gt;T</td>
<td>14:45658326</td>
<td>rs147021911</td>
<td>p.Gln1701*</td>
<td>0.00768</td>
<td>0.00142</td>
</tr>
<tr>
<td>MPG</td>
<td>c.401G&gt;T</td>
<td>16:129423</td>
<td>rs146600185</td>
<td>p.?</td>
<td>0.00643</td>
<td>0.00111</td>
</tr>
<tr>
<td>NEIL1</td>
<td>c.314dupC</td>
<td>15:75641560</td>
<td>-</td>
<td>p.Pro106Alafs*50</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
4.2.2 DNA and RNA analyses and genotyping (I,III)

In Study I, 22 chosen variants were genotyped in the Helsinki and Tampere case-control datasets with three consecutive phases, allowing concurrent exclusion of variants for which no additional carriers were found or those prevalent in the population controls (I, Fig. S1).

The chosen 22 DNA repair variants (Table 3) were first genotyped in 524 familial BRCA1/2-negative cases from Helsinki to determine whether they were recurrent in the Finnish families. Altogether 14 variants were excluded from further analysis as no additional carriers were detected, and FANCA c.1682C>T variant was excluded because of an unreliable result. In phase two, the remaining eight variants were genotyped in 552 population-matched healthy female controls to further examine the general population frequency. Variants with control sample frequencies similar to phase one familial patients were excluded, as were variants too infrequent to allow statistical evaluation. In phase three, the six remaining variants (RUVBL1 c.950G>A, SLX4 c.2484G>C, FANCA c.4228T>G, FANCM c.5101C>T, MPG c.40-1G>T, and NEIL1 c.314dupC) were genotyped in additional 233 familial BRCA1/2-negative and 1,730 unselected breast cancer patients from the Helsinki area, as well as in 679 unselected breast cancer cases from Tampere. Genotyping was performed also for healthy population controls from Helsinki and Tampere. In addition, these six variants were genotyped in 569 unselected ovarian cancer patients from Helsinki.

Genotyping was performed with Sequenom MassArray system with iPLEX Gold assays at Institute for Molecular Medicine Finland (FIMM) (University of Helsinki), Sanger sequencing by using BioTools DNA polymerase, or with TaqMan real-time PCR by using TaqMan SNP Genotyping assays and TaqMan Genotyping Master Mix (Applied Biosystems) (I, Table S1). For TaqMan, the genotype calling was performed with 7500 Fast RealTime PCR System and ABI Prism 7500 SDS software 1.4 (Applied Biosystems). For Sanger sequencing, PCR products were visualized on 2% (wt/vol) agarose gel and purified with Exo-SAP-IT (Affymetrix). ABI BigDye Terminator 3.1 Cycle sequencing kit (Applied Biosystems) was used for purified PCR products, and the final sequencing was performed at FIMM with 3730xl DNA Analyzer (Applied Biosystems). Variant Reporter (Applied Biosystems) and FinchTV (Geospiza) were used for visualizing the sequence chromatograms.

In addition, FANCM c.5101C>T mutation was genotyped in 965 unselected breast cancer patients (including 92 familial cases) from Iceland. TaqMan SNP Genotyping assays and TaqMan Genotyping Master Mix were used, and PCR was performed with StepOne RealTime PCR system (Life Technologies). Data were collected and analyzed with StepOne software 2.0 (Life Technologies). All Iceland sample genotyping was performed at Landspitali University, Iceland.
**FANCM** c.5791C>T mutation was previously identified as a familial breast cancer allele as well as enriched loss-of-function mutation in the Finnish population. In Study III, this mutation was genotyped in Helsinki and Tampere breast cancer samples and healthy controls with Sequenom MassARRAY system using iPLEX Gold assays (Sequenom) at FIMM (University of Helsinki). For Oulu and Kuopio sample genotyping, performed at the University of Oulu, High Resolution Melt analysis (CFX96, Bio-Rad) with Type-it HRM reagents (Qiagen) was used.

TaqMan real-time PCR was utilized for **FANCM** indel mutations c.4025_4026delCT and c.5293dupA by using TaqMan SNP Genotyping Custom assays and TaqMan Genotyping Master Mix (Applied Biosystems). 7500 RealTime PCR System and 7500 software (version 2.06, Applied Biosystems) were used for genotype calling (III). The c.4025_4026delCT variant was originally identified with panel sequencing at Lund University in one Finnish patient and the c.5293dupA variant with exome sequencing in Study III.

Positive controls were used in all analyses (I,III) and all mutations were confirmed with Sanger sequencing and/or Interactive Genomics Viewer.

Quantitative allele-specific RT-PCR was performed for wild-type and mutant **FANCM** c.5101C>T alleles to determine whether the mutation leads to nonsense-mediated mRNA decay (I). RNA was extracted from lymphoblastoid cells from a heterozygous carrier of the c.5101C>T mutation, a non-carrier from the Finnish population, and five Caucasians (non-Finnish) non-carriers. The analysis was run with modified forward primers on a 7900HT system (Applied Biosystems) and repeated using RNA from cells treated with cycloheximide (100 µg/ml) for 4 hours to inhibit nonsense-mediated RNA decay. **FANCM** allele levels were normalized relative to GAPDH levels. All nonsense-mediated mRNA decay analyses were performed at Mayo Clinic, Rochester, Minnesota, USA.

### 4.2.3 Statistical analyses and bioinformatics

#### 4.2.3.1 Risk analyses (I,III)

Breast/ovarian cancer risk associated with the studied mutations was statistically evaluated by comparing the frequencies of the genotyped cases and controls. All study participants were also divided into subgroups according to family history of breast or ovarian cancer, ER status, and TN subtype to assess the cancer risk associated with breast cancer phenotypes. Two-sided P-values and odds ratios to evaluate the risk in all breast cancer cases as well as in subgroups were calculated with Pearson chi-squared test or Fisher exact test (if n ≤ 5) (I,III). As Study I initially included altogether 22
Materials and methods

variants, multiple testing correction was conducted with Bonferroni method, with P-value of <0.0023 considered as significant.

To increase the statistical power, a fixed-effect meta-analysis combining Helsinki and Tampere datasets for FANCM c.5101C>T mutation was performed with R.3.01-environment (http://www.r-project.org/) by using inverse variance-weighted method (I). In Study III, all datasets (Helsinki, Tampere, Kuopio, Oulu) were pooled for combined analysis and the odds ratios and P-values were estimated with logistic regression model stratified by study. Similar analyses were performed for subgroups defined by histopathology and family history of breast cancer (I,III). In addition, genotyping results for both FANCM c.5791C>T and c.5101C>T mutations were pooled and analyzed with logistic regression in all studied patients (III).

The heterogeneity between ER-positive and TN subgroups was evaluated with two-sided z-test (I) or between studies and mutations (III) by combining betas and standard errors of the different studies in “rmeta” package in R.3.02-environment (http://www.r-project.org/). In Study I, the mean age of FANCM c.5101C>T carriers and non-carriers at the time of diagnosis was compared with Student paired t-test.

4.2.3.2 Survival analyses (II)

Study participant data, including 3,933 invasive breast cancer patients from Helsinki, Tampere, Kuopio, and Oulu, were pooled for survival analyses. The date of the patient diagnosis was used to calculate the time-to-event and all follow-up times were left-truncated at the time of the study entry to account for the latency between diagnosis and study recruitment of the patients. Hazard ratios (HRs) were estimated with univariate Cox’s proportional hazard model among datasets stratified by study. HRs and 95% confidence intervals (CIs) were further visualized with Kaplan-Meier curves. In multivariate Cox’s proportional hazard models, clinically relevant factors (ER, grade, tumor size, nodal status) and/or cancer treatments (radiotherapy, hormone therapy, and chemotherapy) were fitted in as categorical co-variates to test the independence of the variables. All analyses were further visualized with forest plots.

Breast cancer death with 10-year follow-up time was the primary end point of the survival analyses in all datasets. Local recurrence (in five years) was used as an endpoint in the radiotherapy-based subgroup analysis in the Helsinki dataset. The interaction between FANCM c. 5101C>T mutation and radiation treatment was studied with Cox’s proportional hazard models, one including the treatment and c.5101C>T genotype as individual covariates and the other including an interaction term between these two. Models were compared with two-way ANOVA test.

Histopathological features of the tumors were compared with Pearson’s chi-squared test or Fisher exact test (if n ≤ 5), or with logistic regression for
variables with two or more categories. The correlation between \textit{FANCM} mutation status and immunohistochemical expression of DNA damage and repair response markers was tested with Kruskal-Wallis test (continuously scored markers) or chi-squared test (categorical scoring).

All statistical analyses were performed using the R.3.0.2 environment (http://www.r-project.org/).
5 RESULTS

5.1 IDENTIFYING FANCM AS A NOVEL BREAST CANCER GENE (I)

Exome sequencing of 24 BRCA1/2-negative patients from 11 breast cancer families produced a list of 80,918 variants in 80,867 unique positions. Of these, 22 variants from 21 genes participating in DNA repair were selected for further genotyping in the case-control datasets of Finnish breast and ovarian cancer patients and healthy population controls. In the FANCM gene, 96 carriers of the c.5101C>T mutation (14:45658326C>T, rs147021911, p.Gln1701*) were identified in 3,079 breast cancer cases (3.1%), 12 in 548 ovarian cancer cases (2.2%), and 38 in 2,080 population controls (1.8%).

5.1.1 FANCM c.5101C>T is associated with breast cancer risk

After statistical analyses of the genotyping results, FANCM c.5101C>T nonsense mutation was found to be associated with breast cancer. The frequency of the mutation was significantly higher among all breast cancer cases (2.9%) than in controls (1.4%) in the Helsinki sample set, demonstrating an approximately twofold increased risk for breast cancer (OR=2.06, 95% CI=1.22 – 3.47, P=0.006) (Table 4).

All genotyped patients (n=3,166) were divided into subgroups based on family history of breast cancer, ER receptor status, and triple-negative subtype to identify the patient subgroups with the highest risk of the disease (Table 4). A significant association was seen in the Helsinki dataset among familial patients (OR=2.21, 95% CI=1.24 – 3.94, P=0.006) as well as among unselected patients (OR=2.01, 95% CI=1.16 – 3.47, P=0.011). Among patients with a strong family history of breast cancer, c.5101C>T carrier frequency was 3.2% and among patients with one affected first-degree relative the frequency was 2.9%. Frequencies did not notably differ between families with both breast and ovarian cancer (3.5%) and families with breast cancer only (3.2%).

In the breast cancer phenotype subgroups, a significant association and a higher risk were seen among ER-negative patients (OR=2.38, 95% CI=1.17 – 4.83, P=0.013) relative to the ER-positive subgroup (OR=1.83, 95% CI=1.05 – 3.17, P=0.029). The most significant association and the highest risk were seen in the triple-negative subgroup of patients (OR=4.13, 95% CI=1.76 – 9.67, P=0.0004) (Table 4). In the Helsinki dataset, 143 patients were diagnosed as triple-negative. Of these, 24 patients had a family history of the disease. Altogether eight triple-negative patients carried the mutation, including two familial patients.
In the independent sample set from Tampere, results from risk analyses were consistent (Table 4). The highest risk was again seen among triple-negative cases (OR=2.77, 95% CI=0.92 – 8.37, P=0.086).

The frequency of the carriers was slightly lower in unselected ovarian cancer patients (2.2%), and a consistent but not significant association with ovarian cancer was seen in 548 patients included in the analysis (OR=1.56, 95% CI=0.75 – 3.26, P=0.235). Among 965 unselected breast cancer cases from Iceland, no FANCM c.5101C>T carriers were identified. However, this country is genetically isolated, and it is possible that other population-specific FANCM mutations may be present.

Table 4. Frequencies of the FANCM c.5101C>T mutation in all breast cancer patients, breast cancer subgroups, and ovarian cancer patients. MUT = mutation carrier, WT = wild type.

<table>
<thead>
<tr>
<th>Patient series</th>
<th>MUT (%)</th>
<th>WT (%)</th>
<th>P</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helsinki</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>18 (1.4)</td>
<td>1253 (98.6)</td>
<td>0.0059</td>
<td>2.06 (1.22-3.47)</td>
</tr>
<tr>
<td>All breast cancer cases</td>
<td>69 (2.9)</td>
<td>2336 (97.1)</td>
<td>0.0061</td>
<td>2.21 (1.24-3.94)</td>
</tr>
<tr>
<td>Familial breast cancer</td>
<td>33 (3.1)</td>
<td>1041 (96.9)</td>
<td>0.0109</td>
<td>2.23 (1.19-4.57)</td>
</tr>
<tr>
<td>≥3 cases</td>
<td>17 (3.2)</td>
<td>507 (96.8)</td>
<td>0.0307</td>
<td>2.09 (1.06-4.12)</td>
</tr>
<tr>
<td>≥2 cases</td>
<td>16 (2.9)</td>
<td>534 (97.1)</td>
<td>0.0109</td>
<td>2.01 (1.16-3.47)</td>
</tr>
<tr>
<td>Unselected breast cancer</td>
<td>48 (2.8)</td>
<td>1665 (97.2)</td>
<td>0.0293</td>
<td>1.83 (1.05-3.17)</td>
</tr>
<tr>
<td>ER-positive</td>
<td>46 (2.6)</td>
<td>1753 (97.4)</td>
<td>0.0132</td>
<td>2.38 (1.17-4.83)</td>
</tr>
<tr>
<td>ER-negative</td>
<td>14 (3.3)</td>
<td>409 (96.7)</td>
<td>0.0004</td>
<td>4.13 (1.76-9.67)</td>
</tr>
<tr>
<td>Triple-negative</td>
<td>8 (5.6)</td>
<td>135 (94.4)</td>
<td>0.0806</td>
<td>2.77 (0.92-8.37)</td>
</tr>
<tr>
<td>Unselected ovarian cancer</td>
<td>12 (2.2)</td>
<td>536 (97.8)</td>
<td>0.0932</td>
<td>1.65 (0.91-2.96)</td>
</tr>
<tr>
<td>Tampere</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>20 (2.5)</td>
<td>789 (97.5)</td>
<td>0.0651</td>
<td>1.96 (0.95-4.08)</td>
</tr>
<tr>
<td>All breast cancer cases</td>
<td>27 (4.0)</td>
<td>647 (96.0)</td>
<td>0.0497</td>
<td>2.36 (0.98-5.70)</td>
</tr>
<tr>
<td>Familial breast cancer</td>
<td>12 (4.7)</td>
<td>241 (95.3)</td>
<td>0.2610</td>
<td>1.45 (0.76-2.76)</td>
</tr>
<tr>
<td>ER-positive</td>
<td>18 (3.5)</td>
<td>491 (96.5)</td>
<td>0.0497</td>
<td>2.36 (0.98-5.70)</td>
</tr>
<tr>
<td>ER-negative</td>
<td>7 (5.6)</td>
<td>117 (94.4)</td>
<td>0.0806</td>
<td>2.77 (0.92-8.37)</td>
</tr>
<tr>
<td>Triple-negative</td>
<td>4 (6.6)</td>
<td>57 (93.4)</td>
<td>0.0806</td>
<td>2.77 (0.92-8.37)</td>
</tr>
</tbody>
</table>

Estimates from the Helsinki and Tampere datasets were combined for meta-analysis (Table 5), confirming the association of the c.5101C>T mutation with breast cancer (OR=1.86, 95% CI=1.26 – 2.75, P=0.0018). The risk was similarly higher in familial cases (OR=2.11, 95% CI=1.34 – 3.32, P=0.0012) and especially among triple-negative patients (OR=3.56, 95% CI=1.81 – 6.98, P=0.0002). P-values remained significant after multiple testing correction. Borderline heterogeneity was seen in the risks between ER-positive and triple-negative breast cancer (P_{het}=0.059), emphasizing an increased risk especially for the triple-negative subtype of breast cancer. The mean age of
Results

FANCM c.5101C>T mutation carriers was 55.2 years at the age of diagnosis, and the mean age of non-carriers was 56.4 years (P=0.416).

Table 5. Meta-analysis of Helsinki and Tampere breast cancer patients.

<table>
<thead>
<tr>
<th>Patient series</th>
<th>N</th>
<th>P</th>
<th>OR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>All breast cancer cases</td>
<td>3079</td>
<td>0.0018</td>
<td>1.86</td>
<td>1.26-2.75</td>
</tr>
<tr>
<td>Familial breast cancer</td>
<td>1327</td>
<td>0.0012</td>
<td>2.11</td>
<td>1.34-3.32</td>
</tr>
<tr>
<td>ER-positive</td>
<td>2308</td>
<td>0.0182</td>
<td>1.66</td>
<td>1.09-2.52</td>
</tr>
<tr>
<td>ER-negative</td>
<td>547</td>
<td>0.0021</td>
<td>2.37</td>
<td>1.37-4.12</td>
</tr>
<tr>
<td>Tripe-negative</td>
<td>204</td>
<td>0.0002</td>
<td>3.56</td>
<td>1.81-6.98</td>
</tr>
</tbody>
</table>

The other 21 genotyped variants did not reach statistical significance (I, Table S2). However, FANCA c.4228T>G was detected in two breast cancer patients and RUVBL1 c.950G>A in three breast cancer patients, but not in any population controls. Also, three variants showed some co-segregation in the families; SHPRH deletion c.3577_3580delCTTA, MPG splicing mutation c.40-1G>T, and RUVBL1 missense variant c.950T>G were all discovered in both exome-sequenced breast cancer patients in the respective families. The association of these variants with breast cancer remains to be studied in larger datasets. However, the FANCA c.4228T>G variant most likely is not a breast cancer allele since it is Fanconi anemia core complex gene, which are rarely associated with breast or ovarian cancer predisposition.

5.1.2 Familial segregation of the FANCM c.5101C>T mutation

Familial segregation of the FANCM c.5101C>T mutation was evaluated in 45 individuals from eight mutation carrier families. Of 11 female relatives with breast cancer, five first-degree relatives of the index patients carried the mutation. Of 16 healthy female relatives aged between 33 and 80 years, seven were mutation carriers. Most of the studied families showed incomplete segregation of the mutation; in one family, however, all three sisters affected with breast cancer were mutation carriers (Figure 5). In addition, one male diagnosed with prostate cancer and one female with undefined connective tissue cancer carried the c.5101C>T mutation (Figure 6). Several other cancer types were also present in the families, including pancreatic, lung, colorectal, prostate, bone marrow, liver, and kidney cancer, but no samples were available to determine whether these relatives were c.5101C>T carriers. The potential role of FANCM in these cancer types remains to be elucidated.
**Figure 5.** Pedigree of a family with three sisters affected with breast cancer (BC) carrying the *FANCM* c.5101C>T mutation. Also, one non-affected sister is a carrier of *FANCM* c.5101C>T. The age at diagnosis of the index patient was 49 years, the other two sisters were diagnosed at 60 and 56 years. Also ovarian (OC), lung, prostate (PR), and bone marrow (Bm) cancers are present in the family.

**Figure 6.** A family with a prostate cancer patient with the *FANCM* c.5101C>T mutation and a female breast cancer patient with also skin cancer carrying the c.5101C>T mutation. BC = breast cancer, CRC = colorectal cancer, PR = prostate cancer.
5.1.3 Functional analysis of the FANCM c.5101C>T mutation

Allelic-specific quantitative real-time PCR was performed for wild-type and mutation alleles to determine whether the FANCM c.5101C>T mutation leads to nonsense-mediated mRNA decay (NMD) in cells. Both alleles were detected in equal proportions in RNA extracted from a heterozygous c.5101C>T carrier. Only the wild-type allele was detected in non-carrier control samples, indicating that the c.5101C>T allele does not cause nonsense-mediated mRNA decay and the expression of the mutant allele may affect the breast cancer risk.

5.2 FANCM C.5101C>T SURVIVAL (II)

5.2.1 FANCM c.5101C>T mutation is associated with poor breast cancer prognosis

FANCM c.5101C>T mutation and breast cancer outcome were investigated in 3,933 invasive breast cancer patients from Finland (areas of Helsinki, Tampere, Kuopio, and Oulu). Descriptions of the datasets and the tumor characteristics of all patients are presented in Publication II, in Tables 1 and 2, respectively.

Ten-year breast cancer-specific survival was first examined with Cox’s univariate proportional hazard analysis in the pooled dataset stratified by study (Figure 7). The mutation was associated with poor prognosis in the univariate model (HR=1.66, 95% CI=1.09 – 2.52, P=0.018), but not in the multivariate model that included the common clinical features (ER status, grade, tumor size, nodal status) and the anticancer treatments (radiotherapy, chemotherapy, hormone therapy) in 3,268 patients (HR=1.44, 95% CI=0.91 – 2.26, P=0.133). This suggests that the mutation may not be independently prognostic, although the hazard ratio and confidence intervals remain similar to those in the main univariate analysis. However, the statistical power of the multivariate analysis is lower. It is also possible that some other biological factors may affect the disease outcome in concert with the mutation.

In Study I, the FANCM c.5101C>T mutation was associated especially with the triple-negative subtype of breast cancer. As this tumor phenotype is known to be associated with poor breast cancer prognosis as such, survival among ER-positive breast cancer patients was analyzed to determine whether the mutation is associated with reduced survival independently of the triple-negative phenotype. Adverse disease outcome was also observed in ER-positive patients (HR=1.8, 95% CI=1.09 – 2.98, P=0.021) and therefore, the poor survival in the main analysis is most likely not only a result of the higher number of triple-negative tumors among FANCM c.5101C>T carriers.
Also, in Study I, the frequency of the \textit{FANCM} c.5101C>T mutation was slightly increased among familial breast cancer patients. A univariate Cox’s proportional hazard analysis was utilized to examine the breast cancer survival in 1,006 invasive familial cases from Helsinki. Even more reduced survival was seen in this group of patients (HR=2.93, 95% CI= 1.5 – 5.76, P=0.0018) (Figure 7).

![Cumulative breast cancer survival in 10 years in all breast cancer patients in the pooled dataset (n=3,933) and in familial patients from the Helsinki dataset (n=1,006), presented as Kaplan-Meier plots. WT=wild type, MUT=mutation carrier.](image)

**Figure 7.** Cumulative breast cancer survival in 10 years in all breast cancer patients in the pooled dataset (n=3,933) and in familial patients from the Helsinki dataset (n=1,006), presented as Kaplan-Meier plots. WT=wild type, MUT=mutation carrier.

To further define the patient subgroups associated with survival effect, univariate Cox’s proportional hazard analyses with 10-year breast cancer-specific survival as an endpoint were conducted in different groups of tumor histology clinically associated with the disease prognosis (ER, PR, TN, nodal status, tumor size, grade) in the pooled dataset. A similar analysis was performed with the anticancer treatments (radiotherapy, chemotherapy, and hormone therapy) to further examine the association of the \textit{FANCM} c.5101C>T mutation with the patient response to breast cancer treatment (Table 6). Identical analyses for different subgroups and treatments were conducted among ER-positive patients \textit{(II, Supplementary Fig.1)}.

Poor survival was seen especially in patients who had not received radiotherapy (HR=3.43, 95% CI=1.6 – 7.34, P=0.0014), but no such effect was seen in patients treated with radiation (HR=1.35, 95% CI=0.82 – 2.23, P=0.237) (Table 6). In the other treatment subgroups (hormone and chemotherapy), no significant difference was seen between patients receiving treatment and those who did not.
### Table 6. Univariate Cox’s proportional hazard analyses for 10-year specific breast cancer survival in all study patients and in histology and cancer treatment subgroups.

<table>
<thead>
<tr>
<th>Tumor subgroup</th>
<th>P</th>
<th>HR</th>
<th>lower95</th>
<th>upper95</th>
<th>N</th>
<th>Deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>All cases</td>
<td>0.023</td>
<td>1.62</td>
<td>1.069</td>
<td>2.461</td>
<td>3868</td>
<td>580</td>
</tr>
<tr>
<td>ER-positive</td>
<td>0.022</td>
<td>1.79</td>
<td>1.087</td>
<td>2.960</td>
<td>2977</td>
<td>370</td>
</tr>
<tr>
<td>ER-negative</td>
<td>0.981</td>
<td>1.01</td>
<td>0.448</td>
<td>2.277</td>
<td>738</td>
<td>190</td>
</tr>
<tr>
<td>PR-positive</td>
<td>0.255</td>
<td>1.44</td>
<td>0.768</td>
<td>2.713</td>
<td>2418</td>
<td>280</td>
</tr>
<tr>
<td>PR-negative</td>
<td>0.145</td>
<td>1.54</td>
<td>0.862</td>
<td>2.742</td>
<td>1293</td>
<td>279</td>
</tr>
<tr>
<td>Triple-negative</td>
<td>0.211</td>
<td>1.79</td>
<td>0.720</td>
<td>4.443</td>
<td>308</td>
<td>69</td>
</tr>
<tr>
<td>Nodes positive</td>
<td>0.118</td>
<td>1.47</td>
<td>0.906</td>
<td>2.394</td>
<td>1659</td>
<td>409</td>
</tr>
<tr>
<td>Nodes negative</td>
<td>0.455</td>
<td>1.40</td>
<td>0.576</td>
<td>3.422</td>
<td>2133</td>
<td>158</td>
</tr>
<tr>
<td>Tumorsize T1</td>
<td>0.055</td>
<td>2.00</td>
<td>0.986</td>
<td>4.066</td>
<td>2239</td>
<td>183</td>
</tr>
<tr>
<td>Tumorsize T2+</td>
<td>0.486</td>
<td>1.20</td>
<td>0.709</td>
<td>2.061</td>
<td>1554</td>
<td>385</td>
</tr>
<tr>
<td>Grade 1_2</td>
<td>0.318</td>
<td>1.40</td>
<td>0.722</td>
<td>2.725</td>
<td>2537</td>
<td>277</td>
</tr>
<tr>
<td>Grade 3</td>
<td>0.243</td>
<td>1.41</td>
<td>0.792</td>
<td>2.520</td>
<td>1064</td>
<td>267</td>
</tr>
<tr>
<td><strong>Treatment subgroup</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemotherapy</td>
<td>0.301</td>
<td>1.38</td>
<td>0.736</td>
<td>2.606</td>
<td>1284</td>
<td>252</td>
</tr>
<tr>
<td>No chemotherapy</td>
<td>0.054</td>
<td>1.74</td>
<td>0.999</td>
<td>3.030</td>
<td>2516</td>
<td>326</td>
</tr>
<tr>
<td>Hormone therapy</td>
<td>0.031</td>
<td>1.95</td>
<td>1.065</td>
<td>3.576</td>
<td>1595</td>
<td>227</td>
</tr>
<tr>
<td>No hormone therapy</td>
<td>0.285</td>
<td>1.37</td>
<td>0.770</td>
<td>2.435</td>
<td>2201</td>
<td>350</td>
</tr>
<tr>
<td>Radiotherapy</td>
<td>0.237</td>
<td>1.35</td>
<td>0.820</td>
<td>2.227</td>
<td>2963</td>
<td>422</td>
</tr>
<tr>
<td>No radiotherapy</td>
<td>0.001</td>
<td>3.43</td>
<td>1.607</td>
<td>7.336</td>
<td>841</td>
<td>155</td>
</tr>
</tbody>
</table>

Radiation treatment is aimed to prevent local recurrence of the disease, and to this end, a survival analysis was performed with local recurrence (within 5 years) as an endpoint in 2,337 patients from Helsinki. *FANCM* mutation carriers who had not received radiotherapy showed an increased risk for local recurrence (HR=6.19, 95% CI=1.46 – 26.2, P=0.013). Radiotherapy-treated patients did not show any recurrence risk (HR=0.98, 95% CI=0.24 – 4.00, P=0.979).

#### 5.2.2 Interaction analyses

The interaction between the *FANCM* c.5101C>T mutation and radiation treatment was tested with Cox’s proportional hazard model stratified by study in the pooled dataset in 2,996 radiotherapy-treated patients and 864 patients who did not receive the treatment. The interaction between the mutation and radiation treatment was statistically significant, with a protective hazard ratio (HR=0.37, 95% CI=0.15 – 0.92, P=0.032). Two-way ANOVA likelihood ratio test between the interaction term and independent
co-variates in the analysis showed a significant interaction ($P_{\text{interaction}}=0.040$). A similar analysis was performed in the Helsinki dataset ($n=2,069$) for the c.5101C>T mutation and radiation treatment using local recurrence (within 5 years) as an endpoint. A stronger protective hazard ratio was seen for the interaction between the mutation and radiation treatment (HR=0.16), than for the mutation alone (HR=5.96). The sample set was, however, smaller, providing less statistical power, and a significant result for the interaction effect was not observed (likelihood ratio P-value test=0.090).

5.2.3 Histopathology of FANCM c.5101C>T mutation-positive tumors

To study the FANCM c.5101C>T association with specific histopathological features that could possibly be associated with disease outcome, all mutation carriers were compared with wild-type patients in relation to grade, tumor size, nodal status, distant metastases, hormone receptor status, and morphology. The mutation carrier tumors did not clearly associate with any of the clinical features, but they were more frequently of the triple-negative subtype compared with non-carrier tumors ($P=0.06$) (II, Table 2).

5.2.4 Immunohistochemical analyses

FANCM is associated with cellular DNA repair in a Fanconi anemia pathway. To this end, the nuclear immunohistochemical staining of 13 DNA damage markers and the association with the FANCM c.5101C>T mutation were investigated in mutation carrier and non-carrier tumors in 1,240 invasive breast cancer cases $^{223, 224}$. Of the studied markers, FANCM c.5101C>T mutation was associated with nuclear poly-ADP-ribose (PAR), a measure of PARP activity. This group of proteins acts as a rapid response to DNA damage in cells. PAR staining was reduced in mutation carrier tumors for a proportion of positively stained tumor nuclei ($P=0.016$) and staining intensity ($P=0.011$). Other studied immunohistochemical markers were not associated with FANCM c.5101C>T. The list of studied markers and the immunohistochemistry methods are included in Publication II, Supplementary Appendix I and Supplementary Table 3.
Results

5.3 TRUNCATING FANCM MUTATIONS (III)

5.3.1 Risk analyses

FANCM c.5791C>T (rs144567652, p.Arg1931*) mutation was studied in 4,806 Finnish breast cancer patients and 2,734 population controls from Helsinki, Tampere, Oulu, and Kuopio. Altogether eight mutation carriers were identified in controls (0.3%) and 28 in all breast cancer patients (0.6%). Of these, eight carriers were BRCA1/2-negative familial patients from Helsinki and Oulu. The highest population frequency was seen in Northern and Eastern Finland (Oulu 0.6% and Kuopio 0.6%) and the lowest in Southern and South Western Finland (Helsinki 0.2% and Tampere 0.1%). In addition, two mutation carriers were identified in 526 unselected ovarian cancer patients from Helsinki.

Breast cancer risk was evaluated among all patients from each study as well as in breast cancer subgroups including family history, ER status, and triple-negative subtype (Table 7). A significant association was seen between triple-negative breast cancer and the FANCM c.5791C>T mutation in the Helsinki dataset (OR=9.09, 95% CI=1.82 – 45.49, P=0.02). Two additional triple-negative cases carrying the mutation were identified in the Tampere and Oulu datasets. These results suggest that the c.5791C>T mutation is associated with the risk of triple-negative breast cancer in Finland, similar to the FANCM c.5101C>T mutation identified in Study I. The P-values of other subgroups did not reach statistical significance, however, the odds ratios were comparable with the previous FANCM c.5791C>T data of Peterlongo et al. 222 and Neidhardt et al. 225, suggesting increased risk for breast cancer. The ORs were consistently increased in all breast cancer patients from Helsinki (OR=2.11, 95% CI=0.59 – 7.49, P=0.24) and Tampere (OR=6.14, 95% CI=0.72 – 52.70, P=0.10) as well as in ER-negative, ER-positive, and familial breast cancer patients from Helsinki (Table 7). In the Kuopio dataset, only two mutations carriers were identified (OR=0.73, 95% CI=0.07 – 8.15, P=1) (Table 7). No heterogeneity was observed between the studies (P=0.9).

The analysis of ovarian cancer cases from Helsinki did not yield a significant result, as only two c.5791C>T mutation carriers were identified (OR=1.60, 95% CI=0.27 – 9.58, P=0.64). Larger datasets are needed to evaluate the risks associated with the mutation and ovarian cancer in Finland.
Table 7. FANCM c.5791C>T mutation frequencies in the studied sample sets.

<table>
<thead>
<tr>
<th>Study</th>
<th>MUT (%)</th>
<th>WT (%)</th>
<th>P</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Helsinki</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>3 (0.2)</td>
<td>1255 (99.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All breast cancer cases</td>
<td>12 (0.5)</td>
<td>2379 (99.5)</td>
<td>0.24</td>
<td>2.11 (0.59 – 7.49)</td>
</tr>
<tr>
<td>Unselected breast cancer</td>
<td>9 (0.5)</td>
<td>1690 (99.5)</td>
<td>0.22</td>
<td>2.23 (0.60 – 8.25)</td>
</tr>
<tr>
<td>Familial breast cancer</td>
<td>5 (0.5)</td>
<td>1073 (99.5)</td>
<td>0.48</td>
<td>1.95 (0.46 – 8.18)</td>
</tr>
<tr>
<td>ER-positive</td>
<td>9 (0.5)</td>
<td>1786 (99.5)</td>
<td>0.25</td>
<td>2.11 (0.57 – 7.80)</td>
</tr>
<tr>
<td>ER-negative</td>
<td>3 (0.7)</td>
<td>409 (99.3)</td>
<td>0.16</td>
<td>3.07 (0.62 – 15.26)</td>
</tr>
<tr>
<td>Triple-negative</td>
<td>3 (2.1)</td>
<td>138 (97.9)</td>
<td>0.02</td>
<td>9.09 (1.82 – 45.49)</td>
</tr>
<tr>
<td>Unselected ovarian cancer</td>
<td>2 (0.4)</td>
<td>524 (99.6)</td>
<td>0.64</td>
<td>1.60 (0.27 – 9.58)</td>
</tr>
<tr>
<td><strong>Tampere</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>1 (0.1)</td>
<td>807 (99.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All breast cancer cases</td>
<td>5 (0.8)</td>
<td>657 (99.2)</td>
<td>0.10</td>
<td>6.14 (0.72 - 52.70)</td>
</tr>
<tr>
<td>ER-positive</td>
<td>4 (0.8)</td>
<td>490 (99.2)</td>
<td>0.07</td>
<td>6.59 (0.73 - 59.11)</td>
</tr>
<tr>
<td>ER-negative</td>
<td>1 (0.8)</td>
<td>121 (99.2)</td>
<td>0.25</td>
<td>6.67 (0.41 – 107.34)</td>
</tr>
<tr>
<td>Triple-negative</td>
<td>1 (1.5)</td>
<td>67 (98.5)</td>
<td>0.15</td>
<td>12.04 (0.74 – 194.74)</td>
</tr>
<tr>
<td><strong>Oulu</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>3 (0.6)</td>
<td>507 (99.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All breast cancer cases</td>
<td>9 (0.7)</td>
<td>1314 (99.3)</td>
<td>1</td>
<td>1.16 (0.31 – 4.30)</td>
</tr>
<tr>
<td>Unselected</td>
<td>6 (0.5)</td>
<td>1141 (99.5)</td>
<td>0.87</td>
<td>0.89 (0.22 – 3.57)</td>
</tr>
<tr>
<td>Familial breast cancer</td>
<td>3 (2)</td>
<td>150 (98.0)</td>
<td>0.14</td>
<td>3.38 (0.68 – 16.92)</td>
</tr>
<tr>
<td>Young breast cancer</td>
<td>0 (0)</td>
<td>56 (100)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ER-positive</td>
<td>2 (0.5)</td>
<td>432 (99.5)</td>
<td>1</td>
<td>0.78 (0.13 – 4.70)</td>
</tr>
<tr>
<td>ER-negative</td>
<td>1 (0.9)</td>
<td>108 (99.1)</td>
<td>0.54</td>
<td>1.56 (0.16 – 15.19)</td>
</tr>
<tr>
<td>Triple-negative</td>
<td>1 (1.4)</td>
<td>68 (98.6)</td>
<td>0.40</td>
<td>2.49 (0.25 – 24.23)</td>
</tr>
<tr>
<td><strong>Kuopio</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>1 (0.6)</td>
<td>157 (99.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All breast cancer cases</td>
<td>2 (0.5)</td>
<td>428 (99.5)</td>
<td>1</td>
<td>0.73 (0.07 – 8.15)</td>
</tr>
<tr>
<td>ER-positive</td>
<td>1 (0.3)</td>
<td>317 (99.7)</td>
<td>1</td>
<td>0.50 (0.03 – 7.97)</td>
</tr>
<tr>
<td>ER-negative</td>
<td>0 (0)</td>
<td>95 (100)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Triple-negative</td>
<td>0 (0)</td>
<td>47 (100)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

No additional carriers were found for the FANCM c.5293dupA variant among 862 familial patients from Helsinki, and thus, it may represent a unique mutation in the family in which it was identified. It is also noteworthy that the variant was originally identified in exome sequencing of a breast cancer family member with uterine cancer. For the c.4025_4026delCT variant, one additional carrier was found in the same dataset of familial patients. Both identified carriers had a family history of breast cancer, but no additional samples were available for genotyping the relatives and this variant was not studied further.

5.3.2 Combined risk analysis

In the combined analysis of all four datasets, a significantly increased risk for the triple-negative subgroup was observed (OR=5.14, 95% CI=1.65 – 16.0,
P=0.005) (Table 8). P-values of the other subgroups did not reach statistical significance, however, the odds ratios were consistently increased, especially in the familial and ER-negative groups of patients.

Table 8. Combined analysis of all sample sets in all breast cancer cases and subgroups, with number of mutation carriers and wild-type patients of each study. BC = breast cancer.

<table>
<thead>
<tr>
<th>Study</th>
<th>Helsinki wt / mut</th>
<th>Tampere wt / mut</th>
<th>Oulu wt / mut</th>
<th>Kuopio wt / mut</th>
<th>OR (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All BC</td>
<td>2379 / 12</td>
<td>657 / 5</td>
<td>1314 / 9</td>
<td>428 / 2</td>
<td>1.94 (0.87 – 4.32)</td>
<td>0.11</td>
</tr>
<tr>
<td>Familial BC</td>
<td>1073 / 5</td>
<td>-</td>
<td>150 / 3</td>
<td>-</td>
<td>2.50 (0.83 – 7.51)</td>
<td>0.10</td>
</tr>
<tr>
<td>Unselected BC</td>
<td>1690 / 9</td>
<td>657 / 5</td>
<td>1141 / 6</td>
<td>428 / 2</td>
<td>1.87 (0.82 - 4.26)</td>
<td>0.14</td>
</tr>
<tr>
<td>ER-positive</td>
<td>1786 / 9</td>
<td>490 / 4</td>
<td>432 / 2</td>
<td>317 / 1</td>
<td>1.86 (0.78 - 4.41)</td>
<td>0.16</td>
</tr>
<tr>
<td>ER-negative</td>
<td>409 / 3</td>
<td>121 / 1</td>
<td>108 / 1</td>
<td>95 / 0</td>
<td>2.34 (0.75 - 7.35)</td>
<td>0.14</td>
</tr>
<tr>
<td>Triple-negative</td>
<td>138 / 3</td>
<td>67 / 1</td>
<td>68 / 1</td>
<td>47 / 0</td>
<td>5.14 (1.65 - 16.0)</td>
<td>0.005</td>
</tr>
</tbody>
</table>

As the number of the c.5791C>T mutation carriers is low, the current results and the genotyping results of the FANCM c.5101C>T mutation from Study I (Helsinki and Tampere) and c.5101C>T genotypes from Oulu and Kuopio were combined to increase the power and further investigate the association with truncating FANCM mutations and breast cancer (Table 9). Breast cancer risk was significantly increased among study patients carrying either of the mutations (OR=1.86, 95% CI=1.32 – 2.49, P=0.0002). The risk was also consistently and significantly increased in breast cancer subgroups, especially among triple-negative patients (OR=3.08, 95% CI=1.77 – 5.35, P=0.00007). No heterogeneity was observed between mutations (P=0.7). It is to be noted that one breast cancer patient carries both c.5101C>T and c.5791C>T mutations. In the analyses this patient was counted as one individual carrier.

Table 9. Combined analysis of FANCM c.5101C>T and c.5791C>T mutations in all sample sets, among different subgroups with number of mutation carriers and wild-type patients.

<table>
<thead>
<tr>
<th>BC subgroup</th>
<th>P-value</th>
<th>OR (95% CI)</th>
<th>Helsinki wt / mut</th>
<th>Tampere wt / mut</th>
<th>Oulu wt / mut</th>
<th>Kuopio wt / mut</th>
</tr>
</thead>
<tbody>
<tr>
<td>All BC</td>
<td>0.0002</td>
<td>1.86 (1.32 - 2.49)</td>
<td>2285 / 81</td>
<td>656 / 35</td>
<td>1293 / 30</td>
<td>419 / 11</td>
</tr>
<tr>
<td>Familial BC</td>
<td>0.004</td>
<td>1.99 (1.24 - 3.19)</td>
<td>1020 / 38</td>
<td>-</td>
<td>148 / 5</td>
<td>-</td>
</tr>
<tr>
<td>Unselected BC</td>
<td>0.0006</td>
<td>1.77 (1.28 - 2.45)</td>
<td>1652 / 58</td>
<td>656 / 35</td>
<td>1124 / 23</td>
<td>419 / 11</td>
</tr>
<tr>
<td>ER-positive</td>
<td>0.005</td>
<td>1.64 (1.17 - 2.32)</td>
<td>1721 / 55</td>
<td>480 / 22</td>
<td>426 / 8</td>
<td>292 / 8</td>
</tr>
<tr>
<td>ER-negative</td>
<td>0.004</td>
<td>2.02 (1.25 - 3.25)</td>
<td>405 / 16</td>
<td>114 / 8</td>
<td>108 / 1</td>
<td>99 / 2</td>
</tr>
<tr>
<td>Triple-negative</td>
<td>0.00007</td>
<td>3.08 (1.77 - 5.35)</td>
<td>130 / 10</td>
<td>60 / 5</td>
<td>68 / 1</td>
<td>58 / 2</td>
</tr>
</tbody>
</table>

Control samples  | Helsinki wt / mut | Tampere wt / mut | Oulu wt / mut | Kuopio wt / mut |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1234 / 21</td>
<td>782 / 21</td>
<td>498 / 12</td>
<td>157 / 1</td>
<td></td>
</tr>
</tbody>
</table>
6 DISCUSSION

6.1 FANCM AND BREAST CANCER RISK (I,III)

Study I in this thesis was the first to demonstrate an association of a deleterious mutation in the FANCM gene with an increased breast cancer risk. The identified mutation, c.5101C>T (rs147021911, p.Gln1701*), encoding a premature stop codon, was found with exome sequencing of 24 Finnish BRCA1/2-negative patients from 11 families with a history of breast cancer. Further genotyping of a large dataset of breast and ovarian cancer cases and population controls revealed an approximately twofold increased mutation frequency among unselected and familial breast cancer cases. The association was especially strong in the triple-negative subgroup, with an approximately fourfold increased risk. Altogether 96 FANCM c.5101C>T carriers were identified in 3,079 genotyped breast cancer patients (3.1%) and 38 in 2,080 healthy population controls (1.8%).

After the discovery of the c.5101C>T allele, another but more infrequent FANCM mutation, c.5791C>T (rs144567652, p.Arg1931*), was also found to be associated with familial and triple-negative breast cancer in several populations. Genotyping of 4,806 invasive breast cancer cases and 2,734 healthy population controls from Finland in Study III showed increased frequency of the FANCM c.5791C>T mutation in breast cancer patients, and again, the increased breast cancer risk was seen especially in the triple-negative subgroup (OR=5.14). Altogether 28 mutation carriers were identified in patients (0.6%) and eight in controls (0.3%).

These results identify FANCM as a moderate-risk breast cancer gene in the Finnish population and are convergent with murine studies suggesting FANCM is a tumor suppressor and plausible cancer susceptibility gene. Furthermore, monoallelic mutations in several other Fanconi anemia pathway genes are known to predispose to breast and ovarian cancer. Recently, a study of a large exome sequencing data set with rare truncating germline variants identified FANCM as the third most mutated gene (after BRCA1 and BRCA2) in human breast cancer.

FANCM c.5101C>T mutation appears more prevalent in Finland than in other countries. The minor allele frequency of the c.5101C>T mutation in Finland is 0.8%, whereas the frequency in other European populations is 0.3% in the gnomAD database. No c.5101C>T carriers have been identified in female breast cancer patients from Italy, Netherlands, Australia, Spain, Poland, Ukraine, or Pakistan, however, the mutation was detected in rare German and Czech cases. Also, in Study I, c.5101C>T carriers were not identified among 965 breast cancer patients from Iceland.
In light of these findings, it appears that \textit{FANCM} c.5101C>T is enriched in the Finnish population, and this may be explained by founder effects. Isolated populations with recent bottlenecks may accumulate a higher proportion of deleterious alleles, and allele/locus heterogeneity may be reduced in complex disease genetics \cite{115, 116}. The frequencies of the \textit{FANCM} c.5101C>T in Finland resemble those of \textit{CHEK2} c.1100delC frameshift mutation; both have similar control sample frequency in Finnish studies, and they are enriched in the Finnish population. In addition, the fourfold increased risk of breast cancer among familial \textit{CHEK2} c.1100delC carriers is comparable with that of \textit{FANCM} c.5101C>T triple-negative carriers \cite{123}.

The \textit{FANCM} c.5791C>T allele is more infrequent in Finland than the c.5101C>T mutation, and c.5791C>T is found also in other countries such as Italy, France, and Germany \cite{222, 225}. However, in the gnomAD database, the frequency of the mutation is again higher in the Finnish population; the minor allele frequency in Finland is 0.4%, versus 0.1% in other European populations \cite{120, 121}. Due to the rarity of the c.5791C>T allele, most studies show only borderline significance when estimating breast cancer risk; however, the risk is consistently increased in all studies. Discovery of \textit{FANCM} mutations as a breast cancer-predisposing factor is a relatively new finding, and identifying more carriers from different populations and larger datasets as well as possibly finding new deleterious \textit{FANCM} mutations will improve the risk estimates in the future. Age-specific risk estimates require even larger datasets and comprehensive segregation analyses in \textit{FANCM}-positive families. It is also to be noted that common genetic variants as well as lifestyle and family history factors may affect the absolute cancer risk, and exact risk estimation is challenging especially for individuals with infrequent and recently identified variants due to limited knowledge of population-specific risks. Furthermore, the segregation of the c.5101C>T mutation was incomplete in the studied Finnish families, which is usually the case with moderate-risk breast cancer alleles. Other unknown susceptibility alleles may also segregate in these families. Gene panel testing may, however, help in identifying such variation and allow early diagnosis and genetic counseling for breast cancer families.

The frequencies of the \textit{FANCM} c.5101C>T and c.5791C>T mutations have been particularly high among triple-negative patients in the studies in this thesis and elsewhere \cite{225, 231}. Approximately 10-20% of all breast cancers are triple-negative, but the genetic susceptibility to this subtype of breast cancer is poorly known. Triple-negative disease is usually aggressive and does not respond to hormonal therapy, as the growth of the cancer is not supported by estrogen and progesterone hormones. It is also more likely to recur than other subtypes, however, chemotherapy may improve short-term survival \cite{66, 87, 232}. Triple-negative tumors commonly have DNA repair deficiencies, especially somatic or germline \textit{BRCA1} mutations, but also defects in other homologous recombination repair genes. Triple-negative tumors with FA
mutations have shown lower mRNA expression of DNA repair genes than luminal A-type tumors. The decreased ability to repair DNA defects increases genomic instability, and this may also partly explain both the chemosensitivity and poor prognosis. Despite the good response to chemotherapy (mainly short-term), genomic instability accelerates mutagenesis, and tumors may become more aggressive. Interestingly, in a recent study a male breast cancer patient carrying a deleterious FANCM mutation had a triple-negative tumor, which is a particularly rare subtype in male breast cancer. The results in this thesis have identified FANCM as a triple-negative breast cancer susceptibility gene, however, additional studies are warranted to elucidate the association between FANCM mutations and triple-negative disease. Treatment of triple-negative breast cancer is difficult, and identification of the variants associated especially with this subtype is important for better understanding of the development of the disease as well as improved treatment.

The frequencies of the FANCM c.5101C>T and c.5791C>T mutations were also slightly increased among familial breast cancer patients in Finland, with somewhat higher odds ratios relative to unselected cases. Noticeably increased breast cancer risk for familial patients with FANCM mutations has been shown in other studies. The samples in these other two studies have, however, been selected based on positive family history, and thus, the included cases may be enriched with a number of genetic factors compared with unselected cases. In the Finnish studies presented in this thesis, the breast cancer risk has been examined in both unselected and familial cases, allowing more accurate evaluation of the associations. Furthermore, all breast cancer patients were divided into subgroups based on the subtype of the disease (ER-positive/negative, triple-negative). Breast cancer is a heterogeneous disorder, and while the subgroup analyses may reduce the statistical power, they allow identification of the cancer phenotype and the individuals at highest risk.

In addition to the c.5101C>T and c.5791C>T mutations, several deleterious variants in the FANCM gene have been identified such as c.5293dupA and c.4025_4026delCT presented in Study III. However, due to their rarity an association with breast or ovarian cancer cannot be shown. A mutational analysis of the entire coding region of the FANCM gene in German female breast cancer patients and healthy controls revealed other rare truncating mutations with increased risk of breast cancer, with possible higher association in early-onset cases. Notably, two FANCM-positive male breast cancer patients both carrying mutations in the 5’ region of the gene developed an early-onset disease, although an association between the FANCM mutations and male breast cancer was not confirmed. It is possible that DNA defects affecting the 5’ area of the gene may be associated with a greater impact on breast cancer risk and earlier age at onset than
Discussion

3´ regional mutations. This is consistent with the results of the distal 3´ FANCM c.5101C>T allele in Study I, where the mean age of mutation carriers at time of breast cancer diagnosis (55.2 years) was close to that of non-carriers (56.4 years), and only six mutation carriers (out of 96) were diagnosed under the age of 40 years.

A significant association between FANCM mutations and ovarian cancer risk was not observed in the studies of this thesis, however, a slightly increased OR for ovarian cancer was seen for both c.5101C>T and c.5791C>T mutations. The ovarian cancer sample size was relatively small in both Studies I and III, and therefore, all ovarian cancer subtypes were included in the analyses instead of examining the subtypes individually. Hence, the risk for ovarian cancer cannot be excluded until FANCM mutations have been evaluated in a larger dataset, also among subgroups. Interestingly, Dicks et al. 173 observed higher FANCM mutation frequency in a high-grade serous ovarian cancer patient dataset (0.96%) than in controls (0.38%). They estimated that relative risk of high-grade serous ovarian cancer associated with deleterious FANCM mutations would be 2.5 (95% CI=1.3 – 5.0, P=0.006). Thus, mutations in FANCM may confer risk for ovarian cancer as well, and if the association can be confirmed, FANCM mutation analysis could be implemented as clinical genetic testing 173.

Similarly for breast cancer, in populations where FANCM mutations are found, diagnostic gene panel testing including FANCM may be useful for families with breast cancer or patients with young age at onset of the disease. In Finland, FANCM is available in clinical gene panel testing 236. Previously, breast and ovarian cancer prevention has been offered mainly to high-risk individuals such as BRCA1/2 mutation carriers with a familial background. Clinical familial-based genetic testing may, however, miss some of the mutation carriers, especially those without a family history of cancer. A recent study compared the lifetime costs and effects of current clinical/familial-based BRCA1/2 testing and multi-gene panel testing including several high-penetrance breast and ovarian cancer susceptibility genes (BRCA1, BRCA2, RAD51C, RAD51D, BRIP1, PALB2) in non-Jewish women age 30 years or older. Multi-gene testing was more cost-effective than BRCA1/2-testing alone, and could prevent ~2% of breast cancer and ~4% of ovarian cancer cases on a population level 237. However, population-based genetic testing raises several ethical and clinical questions. Prevalence and penetrance vary by gene, mutation, and population, and anxiety, unnecessary procedures, and extra costs are likely 238.

For now, the exact breast cancer risks for FANCM mutation carriers cannot be defined and as such a positive mutation result is not clinically actionable, and predictive testing of family members is not warranted. Nevertheless, due to the relatively high frequency of the FANCM variants in the Finnish population, homozygous or compound heterozygous carriers may be identified, and in these cases the cancer risks may be higher. Specifically,
the reported chemotoxicity associated with \textit{FANCM} mutations and 5´mutations associated with possible early-onset disease may also have clinical relevance in the management of such cases 225, 239, 240.

Both of the studied main \textit{FANCM} mutations are located in the C-terminus of the protein and they create premature stop codons. In Study I, the c.5101C>T mutation was not found to be subject to nonsense-mediated RNA decay, a surveillance mechanism in eukaryotic cells recognizing and degrading faulty mRNAs that encode deficient polypeptides 241. The negative NMD test result suggests that expression of the mutant allele is associated with the cancer risk related to this mutation.

Instead of functioning as a traditional premature stop codon, the c.5791C>T mutation creates binding site for splicing factor hnRPA1, causing the skipping of exon 22. This leads to a truncated FANCM protein lacking the C-terminal domains responsible for recognizing and targeting branched DNA molecules 194, 222, 242. However, an aberrant FANCM protein most likely does not completely destroy the DNA binding ability of the protein, nor does it entirely prevent the monoubiquitination of FANCD2 because FANCM has a stimulatory rather than an essential role in these functions 192, 194. The wide range of DNA repair actions and the overall genome integrity protection associated with the operational FANCM protein robustly support its role as a tumor suppressor and as a candidate cancer susceptibility gene, although the exact cellular mechanisms by which it affects cancer predisposition remains to be further studied. To date, it seems that in the absence or dysfunction of the FANCM protein the association of the Fanconi anemia repair complex with the damage site is not complete and checkpoint damage signaling is disturbed, promoting genomic instability, and thus, cancer predisposition 175, 192, 194, 198.

It has been debated whether \textit{FANCM} should be removed from the Fanconi anemia disease gene listing because no \textit{FANCM} mutation carriers with clinical manifestation of FA have been identified. The only FA patient with a \textit{FANCM} mutation also carries biallelic \textit{FANCA} mutations, which most likely cause the symptoms of the disease in this individual 175, 192, 195. Furthermore, seven Finnish homozygous \textit{FANCM} mutation carriers did not exhibit any symptoms of blood diseases, increased cancer incidence, or evidence of other chronic diseases 193, and in two recent related studies 239, 240, homozygous \textit{FANCM} mutation carriers did not present the bone marrow failure or congenital abnormalities characteristic of FA. However, some homozygous individuals had head and neck squamous cell carcinomas and other solid tumors, chemotherapy-related side effects, early menopause, and chromosome fragility sensitivity related to DNA ICLs. All of these symptoms are included in the FA spectrum. Therefore, biallelic \textit{FANCM} mutations do not seem to cause classical FA, but they may be associated with FA-like cancer predisposition and early-onset breast cancer 239, 240.
In these studies, exome sequencing was utilized to identify new possible breast and ovarian cancer alleles. It is a widely used targeted sequencing method, which has proven effective for identification of genetic variants, especially among diseases in which a genetic model is already known and the mutation spectrum is distinctive from unaffected individuals. Over the past years, it has also been successfully used in studies searching for rare genetic variants with moderate or low penetrance in common diseases. However, exome sequencing necessitates a large dataset, or families with several affected individuals, as is the case with the studies in this thesis. Currently, several international or population-specific exome sequencing databases are also available for research purposes. Exome sequencing has its limitations for detecting translocations, copy number alterations, or other large-size genetic alterations. It is also targeted only for exons and intron-exon boundaries, which cover approximately 1.5% of human genome. Therefore, possible disease-causing variants affecting regulatory elements or other distal genetic components outside the coding region must be identified with other methods. Predicting the pathogenicity of such variants is also challenging and requires approaches beyond coding region alleles.

Here, several possible disease-causing variants were identified with exome sequencing. Further genotyping of the variants in large case-control datasets revealed a novel moderate-penetrance breast cancer mutation, demonstrating the benefits of such methods. However, these investigations utilized a strong founder population, amplifying the likelihood of making such findings. It is possible that variants in the non-coding areas of the genome further modify the risks and molecular mechanisms associated with common diseases, but identification of such modifying factors requires the analysis of the whole genomic region of susceptibility genes such as FANCM.

6.2 FANCM AND BREAST CANCER SURVIVAL (II)

Distinct tumor characteristics and genetic factors may influence the prognosis, survival, and management of breast cancer patients. In Study II, tumor phenotype, patient survival, and treatment outcome were studied in 3,933 invasive breast cancer patients from Finland to determine the association of the FANCM c.5101C>T mutation with breast cancer prognosis. The association of the mutation with immunohistochemical staining of DNA repair markers was also examined in 1,240 breast tumors.

FANCM c.5101C>T carriers have poor breast cancer survival compared with non-carriers, especially among patients with family history of the disease. In the multivariate analysis, however, FANCM c.5101C>T was not independently prognostic, indicating that other factors may be associated with the outcome of the disease in FANCM-positive cancer patients as well.
However, the statistical power is lower in the multivariate model, most likely affecting the results.

The tumors of the c.5101C>T mutation carriers were slightly more often of the triple-negative subtype than the tumors of non-carriers. This is consistent with FANCM risk studies showing an association especially with this subtype of breast cancer. Triple-negativity is commonly characterized by aggressive phenotype and poor prognosis, but here, the adverse survival outcome of FANCM c.5101C>T carriers was not only a result of the higher incidence of such tumors, with poor disease outcome also seen among estrogen receptor-positive patients. Whether poor disease outcome is a consequence of increased genomic instability and DNA repair deficiencies warrants further investigations. These results are, however, in line with murine studies, showing decreased overall survival in FancmΔαΔα mice with increased cancer incidence, reduced lifespan, and hypersensitivity to cross-linking agents.

The overall impact of genetic variation on breast cancer survival is still poorly understood. Breast cancer prognosis is commonly affected by tumor grade, stage, and hormone receptor status, which influence the treatment choices as well. Hence, the genetic variation affecting these factors could directly or indirectly modify the prognosis of the disease or the response to cancer treatments. Information on poor survival of mutation carriers may be helpful in the clinic when predicting tumor progression in the newly diagnosed cancer patient. The conclusions of FANCM survival should be confirmed in larger datasets and also in different populations. This, however, has proven difficult due to the rarity of the mutations in populations outside of Finland.

When examining the association between the FANCM c.5101C>T mutation with breast cancer treatment outcome, impaired survival was seen especially in patients who had not received radiotherapy, but not among patients treated with radiation. In the other treatment subgroups (chemotherapy and hormone therapy), no significant difference was observed between patients who had received treatment and those who had not. Interaction analyses between mutation and radiation treatment further verified the protective effect of radiotherapy, which was even more pronounced when investigating the interaction with local recurrence of the disease as an endpoint. Therefore, the risk of local recurrence and consequently also breast cancer death in FANCM mutation carriers may be reduced with postoperative radiotherapy.

Surgery is the primary treatment of breast cancer, aiming to reduce disease recurrence and death, and it is commonly followed by radiation treatment, further improving survival and decreasing the risk of local recurrence as well as breast cancer death in all patient groups. On average, about one breast cancer death can be avoided by year 15 for every four recurrences avoided by year 10. Radiotherapy reduces the 10-year risk of any
first recurrence from 35% to 19% and the 15-year risk of breast cancer death by about 25%. Overall, radiotherapy reduces the breast cancer death rate by about one-sixth. The radiotherapy findings in Study II warrant further research, and the benefits of radiation treatment may vary according to the disease characteristics of the breast cancer patients. So far, it can be only assumed that genomic instability and possibly aggressive subtype of the disease caused by FANCM mutations affect the risk of local recurrence and breast cancer death. Furthermore, when carrying an inherited mutation, the risk for recurrence in the other breast is always a possibility, and it may be difficult to determine whether one has a new primary cancer or a recurrence. In the case of hereditary mutations, preventive mastectomy can be an option.

Interestingly, genetic markers associated with outcome of radiotherapy in cancer patients have not been previously identified, but the treatment response results of FANCM c.5101C>T indicate that these tumors may have features that respond well to radiotherapy. This does not necessarily apply to homozygous FANCM mutation carriers, however, as these individuals suffer from chemotherapy-related hematological side effects and chromosome instability, and radiation may be acutely toxic to the bone marrow and other tissues of these individuals.

The Fanconi anemia pathway facilitates homologous recombination repair in cells. Hence, the association of mutated FANCM and immunohistochemical staining profiles of DNA repair markers of FANCM c.5101C>T carriers was investigated, revealing lowered expression of poly(ADP-ribose) marker (PAR). PAR measures the activity of several PARP enzymes contributing widely to DNA repair and cell survival. For example, PARP1 enzyme binds and responds to many DNA damage types, such as stalled replication forks and DNA crosslinks, and PARP3 responds selectively to double-strand breaks. The immunohistochemical method used in these studies measures only overall poly(ADP-ribosyl)ation (PAR-associated posttranslational multiprotein modification that provides an instant response to DNA damage) levels in tumor nuclei, and therefore, no assumptions of the impact of the mutation on specific PAR enzymes can be made.

The observed reduced PAR expression may indicate defects in PARP-signaling in FANCM-mutated tumors. This is consistent with the study showing hypersensitivity to PARP inhibition in FANCM-defective lymphoblastoid cell lines, suggesting that FANCM has a role in cellular defense against PARP inhibition. If this is the case, clinical trials may show whether FANCM mutation carriers could benefit from PARP inhibition therapy.

PARPi therapies have been mostly utilized in the treatment of BRCA1/2-positive cancers, especially ovarian cancer, but recent studies have suggested that PARP inhibition could function as a targeted anti-cancer procedure for tumors with other homologous recombination repair deficiencies as well, such as the Fanconi anemia pathway mutations.
Predictive biomarkers for tumors responding to PARPi therapy are required, and the utility of PARPi treatment in breast cancer needs further clinical testing\textsuperscript{139, 247-249}.

Previously, PARP inhibition was proposed to cause inability to repair single-stranded DNA breaks, which are converted to double-strand breaks during S-phase. These cannot be properly repaired in cells with defects in the homologous recombination repair pathway, consequently leading to death of a tumor cell. This model has been currently challenged with the data presenting that some PARP inhibitors in fact trap PARP1 on DNA. This disturbs the catalytic functions of PARP1, preventing poly(ADP-ribosyl)ation, and the trapped PARP1 is thought to cause the cytotoxic effects of the PARPi drugs; commonly, PARP1-trapping causes stalling of the replication forks, which can only be repaired with homologous recombination. This view is further supported by findings indicating that PARP1-defective cells are resistant to PARP inhibition\textsuperscript{139, 248, 250}. This is not necessarily the case with FANCM-mutated tumors, based on the observations of PARPi hypersensitivity in FANCM-defective cells\textsuperscript{248}. Furthermore, poly(ADP-ribosyl)ation is associated with transcription, telomere integrity, chromatin remodeling, and mitosis\textsuperscript{251}, and many of these functions overlap with those of FANCM. Therefore, the exact biological association(s) of FANCM mutations and changes in PAR activity can only be speculated. However, both FANCM and PARP are involved in resolving replication stress\textsuperscript{252, 253}, and in lymphoblastoid cell line studies the double-strand break formation following PARP inhibition increased greatly in BRCA2-deficient cells, consistent with the assumptions of the PARPi mechanism, but this was not seen in FANCM-mutated cells. It has been suggested that different mechanisms for PARPi sensitivity exist for different mutations, and in FANCM-deficient cells the sensitivity may be related to DNA replication and checkpoint defects\textsuperscript{247, 248}.

Data from mouse strain, lymphoblastoid cell line, and human cancer patient studies demonstrate associations between increased breast cancer risk, poor prognosis, and aggressive phenotype of FANCM mutations. However, based on current information, exact risk estimates and prognosis of carrying a FANCM mutation are difficult to assign, and the studies in this thesis of FANCM risk and survival also have limitations. Retrospective research may be compromised by confounding factors or information bias, and large sample sets are commonly needed for rare diseases or variants. The rarity of mutations complicates risk estimates and may affect the power of subgroup and multivariate analyses. Although the number of breast cancer samples in these studies is extensive, the lack of patient data may affect the analyses. For example, several study patients in this thesis lacked information on HER2 status, and thus, survival analysis in this subgroup could not be performed.
Nevertheless, the suggested poor prognosis and adverse treatment response of *FANCM* mutation carriers may have clinical implications. Especially homozygous *FANCM* mutation carriers seem to have more severe clinical consequences than heterozygous mutation carriers, including early-onset breast cancer predisposition. Chemotherapy and radiation may be risky, or possibly lethal, to *FANCM* homozygotes because of expected acute toxicity in bone marrow and other tissues\(^\text{239, 240}\). Furthermore, if such an individual has the triple-negative subtype of the disease, using hormone therapy is not an option due to the lack of hormone receptor expression. Therefore, treating cancer in these patients can be difficult. Further investigations will show whether *FANCM* mutation carriers could benefit from PARPi therapies. In the future, the clinical results of the *FANCM* studies may also be useful for developing preventive measures for deleterious *FANCM* mutation carriers, especially those with familial, triple-negative, and homozygous subtype.
7 SUMMARY AND CONCLUSIONS

Objectives of this study were to identify novel breast and/or ovarian cancer alleles in the Finnish population, to evaluate the cancer risk in large case-control datasets, and to determine tumor characteristics, patient survival, and treatment response associated with the identified mutations.

1. A nonsense mutation c.5101C>T in the FANCM gene was found to be associated with breast cancer in the Finnish population, particularly increasing the risk for the triple-negative subtype of the disease. The mutation is more prevalent in Finland than in other European populations.

2. The FANCM c.5101C>T mutation was associated with poor breast cancer-specific survival, especially among familial patients. Carrying the mutation also increased the risk of disease recurrence in patients not treated with radiation, however the situation may be improved with postoperative radiotherapy. Immunohistochemical analyses demonstrated lowered PAR activity in FANCM-mutated tumors, suggesting that further clinical testing may show if these tumors could be a target for PARPi therapy.

3. The FANCM c.5791C>T mutation identified among familial cases in a multicenter study was associated with triple-negative breast cancer in the Finnish population. Two other studied pathogenic FANCM mutations were too rare to allow statistical evaluation, nevertheless, they suggest a wider mutation spectrum in the FANCM gene.

Results from the studies of this thesis identify FANCM as a moderate-penetrance risk gene for breast cancer in the Finnish population. Truncating FANCM mutations confer an approximately twofold increased risk for breast cancer relative to the general population, and the risk is further increased for the triple-negative subtype of breast cancer.

Current knowledge of the population- or age-specific breast cancer risk associated with mutations in the FANCM gene is limited and further investigations are warranted. Future research may help health professionals and patients to make an informed decision concerning preventive measures or cancer treatment as well as follow-up of affected individuals. This is relevant especially in Finland with a strong enrichment of the FANCM mutations in the population, presumably explained by several features of founder effects present in the country’s colonization history.
Summary and conclusions

Most FANCM studies to date have examined mainly the two truncating germline variants presented in this thesis, however, characterizing the entire mutation spectrum of the FANCM gene could improve risk estimates and knowledge of genotype-phenotype effects, especially in the triple-negative subtype. Moreover, the ongoing international consortium studies may confirm the proposed association with ovarian cancer in the near future. Functional analyses can further elucidate the biological role of the pleiotropic FANCM protein and its precise role in the progression of breast cancer.

FANCM c.5101C>T is associated with poor breast cancer prognosis, particularly in familial patients. The results from several studies confirm the increased breast cancer risk and aggressive phenotype of FANCM-mutated tumors. The risk for disease recurrence and having a triple-negative disease directly suggest a poor prognosis, and treating the cancer of FANCM mutation carriers can be especially difficult. Based on the survival analysis, increased risk for local recurrence was seen among breast cancer patients who had not received radiotherapy, but not among patients treated with radiation, indicating that postoperative radiotherapy could improve survival of breast cancer patients with FANCM c.5101C>T mutation. This does not necessarily apply to homozygous carriers, in whom the expected acute toxicity of chemo- and radiotherapy can exacerbate the situation. However, the decrease in PAR activity in FANCM-defective tumors suggests that further clinical testing for PARP inhibition treatment in FANCM mutation carriers could provide more information about the utility of such a treatment for these individuals. Further studying the long-term survival in FANCM mutation carriers is also warranted.

In conclusion, pathogenic mutations in the FANCM gene are associated with breast cancer risk and survival. Identification and characterization of such mutations as FANCM c.5101C>T and c.5791C>T can improve knowledge of breast cancer and facilitate the disease prevention, early diagnosis, more accurate prognosis, and individualized treatment of breast cancer patients.
ACKNOWLEDGMENTS

This study was carried out at the Department of Obstetrics and Gynecology, Helsinki University Hospital and University of Helsinki, during 2013 - 2017. I sincerely thank all of the people contributing to this project:

Professors Jorma Paavonen and Juha Tapanainen from the Department of Obstetrics and Gynecology and Klaus Olkkola, Head of Clinicum at the Faculty of Medicine, for providing excellent research facilities.

My supervisor, Heli Nevanlinna, for the opportunity to work and study as a member of the hereditary breast cancer research group and for trusting me with many important and interesting projects. Her enthusiasm is truly appreciated.

The senior clinicians of the group, Docents Carl Blomqvist and Päivi Heikkilä, and Professor Kristiina Aittomäki for their clinical expertise. I am also grateful to Professor Aittomäki for accepting the important role of Custos.

Research nurses Irja Erkkilä, Virpi Palola, and Outi Malkavaara for their valuable help with patient information. I also thank all patients and their families for participating in the studies.

All former and present members of the HN group: Liisa Pelttari for her never-ending patience and valuable help with my research. Sofia Khan for generous help with bioinformatics and delightful company. Maral Jamshidi, my evil twin, for good advices and good laughs. I am also thankful for their company at the conference trips to Scotland and Japan. Rainer “The Brain” Fagerholm for solving any scientific problem. Taru Muranen for wise advices. Sara “Seppo” Vilske for her negotiation skills and important conversations both in and outside the lab. Tuomas Heikkinen, Xiaofeng Dai, Anna Nurmi, Maija Suvanto, Erja Nynäs, and Salla Ranta for their support and sharing the time in the lab and at the office.

My thesis committee, Docents Minna Pöyhönén and Minna Tanner for their insightful and helpful comments and support during my studies.

Doctor Minna Kankuri-Tammilehto and Docent Synnöve Staff for carefully reviewing this thesis.

Professor Minna Nyström for accepting the post of Opponent at my thesis defense.

All colleagues and coworkers at the Departments of Obstetrics and Gynecology and Clinical Chemistry, especially Krisse Nokelainen for being an awesome boss lady in the lab, and the Tapanainen group, especially Kirsi Sepponen for lending a helping hand.
Acknowledgements

All fellow PhD students with whom I collaborated at several courses.

All coauthors and collaborators of the studies: Professors Fergus Couch, Rosa Barkardóttir, Anne Kallioniemi, Johanna Schleutker, Arto Leinenen, Ralf Bützow, Jiri Bartek, Arto Mannermaa, and Robert Winqvist, and their research assistants, postdocs, and students for the successful collaborations.

My friends and family; my parents for encouragement and support. My sister Katariina and her family for the help with logistics. My bae Kimi Seeck for being there when the last light warms the rocks and the rattlesnakes unfold. Dr. Anna Tikka for sharing the whole journey to Mordor and back with me. Saima The Doggo for the rare genuinely happy face. Santeri for the epic laughs. Tormilainen family for support at a thirsty moment, and respectively, Grym family for the Damage Control. Tuire Ahjolinna for forest therapy. I would also like to take a moment here to note that Bruce Lee could perform one hand push-ups using only his thumb and index finger, perform 50 reps of one-arm chin-ups, and throw a single grain of rice in the air and catch it with chopsticks. The efforts required for finishing this thesis do not pale in comparison with that.

The following organizations are gratefully acknowledged for financially supporting my work: the Finnish Cultural Foundation, the Orion Research Foundation, Maud Kuistila Memorial Foundation, Oskar Öflund Foundation, Ida Montin Foundation, The Finnish Cancer Foundation, Biomedicum Helsinki Foundation, and the Doctoral Programme of Biomedicine.

Helsinki, August 2018

Johanna Kiiski
REFERENCES


44. Fang WH, Modrich P. Human strand-specific mismatch repair occurs by a bidirectional mechanism similar to that of the bacterial reaction. J Biol Chem 1993;268:11838-44.


conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-
analysis of individual patient data for 10,801 women in 17 randomised trials. *Lancet*
2011;378:1707-16.


89


251. Hassa PO, Hottiger MO. The diverse biological roles of mammalian PARPS, a small but powerful family of poly-ADP-ribose polymerases. *Front Biosci* 2008;13:3046-82.
