Mikko Vuorela

ROLE OF THE RNF8, UBC13, MMS2 AND RAD51C DNA DAMAGE RESPONSE GENES AND RARE COPY NUMBER VARIANTS IN HEREDITARY PREDISPOSITION TO BREAST CANCER
MIKKO VUORELA

ROLE OF THE RNF8, UBC13, MMS2 AND RAD51C DNA DAMAGE RESPONSE GENES AND RARE COPY NUMBER VARIANTS IN HEREDITARY PREDISPOSITION TO BREAST CANCER

Academic dissertation to be presented with the assent of the Doctoral Training Committee of Health and Biosciences of the University of Oulu for public defence in Auditorium A101 of the Department of Anatomy and Cell Biology (Aapistie 7 A), on 13 December 2013, at 12 noon

UNIVERSITY OF OULU, OULU 2013
Vuorela, Mikko, Role of the RNF8, UBC13, MMS2 and RAD51C DNA damage response genes and rare copy number variants in hereditary predisposition to breast cancer.

University of Oulu Graduate School; University of Oulu, Faculty of Medicine, Institute of Diagnostics, Department of Clinical Chemistry; Biocenter Oulu; Oulu University Hospital

University of Oulu, P.O. Box 8000, FI-90014 University of Oulu, Finland

Abstract

Mutations in the currently known breast cancer susceptibility genes account for only 25–30% of all familial cases. Novel susceptibility genes can be identified by several methods, including candidate gene re-sequencing and genome-wide microarrays. We have applied microarrays for the detection of a new genomic variation class, copy number variants (CNVs), which potentially could disrupt genes in multiple pathways related to breast cancer susceptibility. The aim of the current study was to evaluate the role of the RNF8, UBC13, MMS2 and RAD51C DNA damage response genes in breast cancer susceptibility as well as to study if rare CNVs are associated with the predisposition to this disease.

The analysis of 123 familial breast cancer cases revealed altogether nine different changes in the RNF8 and UBC13 candidate genes. However, none of the observed alterations were considered pathogenic. No alterations were observed in MMS2. The obtained results suggest that breast cancer predisposing alterations in RNF8, UBC13 and MMS2 are rare, or even absent.

The RAD51C mutation screening of 147 familial breast cancer cases and 232 unselected ovarian cancer cases revealed two deleterious mutations: c.-13_14del27 was observed in a breast cancer case with familial history of ovarian cancer and c.774delT in an ovarian cancer case. Both mutations were absent in the control cohort. The results of the study support the hypothesis that rare variants of RAD51C predispose predominantly to ovarian cancer.

A genome-wide scan of CNVs was performed for 103 familial breast cancer cases and 128 controls. The biological networks of the genes disrupted by CNVs were different between the two groups. In familial breast cancer cases, the observed mutations disrupted genes, which were significantly overrepresented in cellular functions related to maintenance of genomic integrity (P=0.0211). Biological network analysis showed that the disrupted genes were closely related to estrogen signaling and TP53-centered tumor suppressor network, and this result was confirmed by the analysis of an independent young breast cancer cohort of 75 cases. These results suggest that rare CNVs represent an alternative source of genetic variation contributing to hereditary risk for breast cancer.

Keywords: breast neoplasms, DNA copy number variations, DNA mutational analysis, DNA-binding proteins, estrogens, Fanconi anemia, genetics, germ-line mutation, tumor suppressor protein p53, ubiquitin-conjugating enzymes
Vuorela, Mikko, RNF8-, UBC13-, MMS2- ja RAD51C-DNA-vauriovastegeenien sekä harvinaisten kopiolukuvarianttien rooli perinnöllisessä rintasyöpääalttiudessa.

Oulun yliopiston tutkijakoulu; Oulun yliopisto, Lääketieteellinen tiedekunta, Diagnostiikan laitos, Klininen kemia; Biocenter Oulu; Oulun yliopistollinen sairaala


Oulun yliopisto, PL 8000, 900414 Oulun yliopisto

Tiivistelmä


123 familiaalisen rintasyöpätapauksen analyysissä löytyi yhteensä yhdeksän muutosta RNF8- ja UBC13-geeneistä, joista yksikään ei osoittautunut patogeeniseksi. MMS2-geeneissä ei havaittu muutosia. Tulosten perusteella rintasyövälle altistavat muutokset RNF8- ja MMS2-geeneissä ovat joko erittäin harvinaisia tai niitä ei esiinny lainkaan.

RAD51C-geenin mutaatiokartoitus 147 familiaalisesta rintasyöpätapauksesta sekä 232 valikoimattomasta munasarjasyöpätapauksesta paljasti kaksi haitallista mutaatiota. c.-13_14del27 havaittiin rintasyöpäpotilaalla, jonka suvussa esiintyi munasarjasyöpää, ja c.774delT todettiin munasarjasyöpäpotilaita. Kumpaakaan mutaatiota ei havaittu verrokkiaineistossa. Tulokset vahvistavat hypoteesia RAD51C-geenin harvinaisten varianttien yhteydestä pääasiassa munasarjasyöpäriskiin.


Asiasanat: DNA-kopiomäärän vaihtelut, DNA-mutaatioanalyysi, DNA:n sitojaproteiinit, estrogeenit, Fanconin anemia, ituradan mutaatio, kasvainsalpaajaproteiini p53, perinnöllisyystiede, rinnan kasvaimet, ubikitiinikonjugaaatioensymit
If I have seen further it is by standing on the shoulders of giants.

- Isaac Newton
Acknowledgements

This study was carried out at the Laboratory of Cancer Genetics and Tumor Biology, Department of Clinical Chemistry, Department of Clinical Genetics, Biocenter Oulu, University of Oulu and Oulu University Hospital during August 2008 – December 2013. I want to express my sincere gratitude to all those who participated in this project.

I wish to express my deepest gratitude to my supervisor Professor Robert Winqvist for giving me the opportunity to explore the interesting world of cancer genetics as a member of his group. He encouraged me to independent scientific thinking and gave me time and freedom to mature as a scientist in my own way. He has truly been an encouraging and inspiring mentor. I am also more than grateful to my other supervisor, Docent Katri Pylkäs, who has been indispensable help throughout my time as a PhD student. Her guidance in scientific matters, from the beginning of this work to the comments on the thesis has been the best possible. I feel that during the years Robert, Katri and I spent together, we evolved to an excellent team. I honestly believe that we have done some good science together.

I also owe my sincere thanks to Professor Juha Risteli, MD, PhD, and current head of the Department of Clinical Chemistry, for the opportunity to prepare this thesis. Furthermore, I am grateful to Dr. Jukka Moilanen, MD, PhD, the head of the University of Oulu Department of Clinical Genetics, for his supportive attitude towards this PhD study.

Docent Auli Karhu and Docent Outi Monni are acknowledged for their expert review and valuable comments on the thesis. I also wish to thank Anna Vuolteenaho, for her careful revision of the language.

My co-authors and collaborators Professor Anne Kallioniemi, Professor Veli-Matti Kosma, Professor Karin Sundfeldt, Professor Annika Lindblom, Professor Annika Rosengren, Docent Arto Mannermaa, Docent Arja Jukkola-Vuorinen, Docent Ulla Puistola, Dr. Jaana Hartikainen, Dr. Anna von Wachenfeldt Wäppling and Dr. Maarit Anttila deserve my sincere thanks for their valuable contribution to the research projects included in this thesis.

I am ever so grateful for all my present and former colleagues for their supportive and the helpful atmosphere, their collaboration on various research projects and excellent company: Muthiah Bose, MSc, Tuomo Mantere, MSc, Anna Tervasmäki, MSc, Raman Devarajan, MSc, Andriana Poulimenos, BSc, Maria Haanpää, MD, Hellevi Peltokeeto, PhD, Sini Nurmenniemi, PhD, Jenni
Nikkilä, PhD, Hannele Erkko, PhD, Szilvia Solyom, PhD and Sanna-Maria Karppinen, PhD. Meeri Otsukka, Leena Keskitalo, Annika Väntänen and Helmi Konola are thanked for the excellent quality of their supportive laboratory work.

I also want thank other scientist and staff, from janitors to waitresses of the restaurant, who were working in the Kieppi building during the years. Researchers are generally humble people with big imagination. These people created warm, open minded and interesting environment, which I enjoyed a lot. I will warmly remember the lively and inspiring chats in the coffee room with Antti Viklund, Veli-Matti Ronkainen, Antti Railo, Mika Kaakinen, Kimmo Halt, Ilkka Pietiläinen, Renata Prunskaita-Hyyryläinen and many others. Thank you for that. Sometimes the little things make all the difference. Especially, I want thank Antti Viklund who was always helping me, whatever the case, inside the lab as well as outside.

I want express my deepest thanks to my family and friends. I am so grateful to my mother and father for their love, support and trust. I could always trust and rely on them. They gave me freedom to do whatever I wanted in my life, and I guess this freedom have been a fundamental reason why I became interested in biology.

Another job and vent hole during the PhD project has been my band Riutta. It has truly been an adventure to make music with you guys. Last, but not least, my dearest thanks goes my to fiancé Annukka who has together with our children Melli and Urho have had the greatest influence on me becoming the person who I am today. Life might not always be easy, but spending it with family you love and which loves you back, makes it much easier. Thank you for your endless support. Without you I would not be writing these words.

This study has been financially supported by the University of Oulu, Biocenter Oulu, Oulu University Hospital, the University of Oulu Graduate School, the Academy of Finland, the Maud Kuistila Memorial Foundation, the University of Oulu Support Foundation, the Cancer Foundation of Northern Finland, the Finnish Cancer Foundation, the Orion-Farmos Research Foundation, the Emil Aaltonen foundation and the Sigrid Juselius Foundation.
Abbreviations

ANKS1B  gene for ankyrin repeat and sterile alpha motif domain containing 1B
APC  gene for adenomatous polyposis coli protein
ATM  gene for ataxia telangiectasia mutated protein
BLM  gene for Bloom syndrome, RecQ helicase-like
BRCA1  gene for breast cancer protein 1
BRCA2/FANCD  gene for breast cancer protein 2/ Fanconi anemia subtype D1
BRCC36  gene for BRCA1/BRCA2-containing complex, subunit 3
BRCC45/BRE  gene for brain and reproductive organ-expressed (TNFRSF1A modulator)
BRIP1/FANCJ  gene for BRCA1 interacting protein C-terminal helicase 1/
  Fanconi anemia subtype J
CASP3  gene encoding caspase 3, apoptosis-related cysteine peptidase
CDH1  gene for cadherin 1/ E-cadherin
CHEK2  gene for checkpoint kinase 2
CI  confidence interval
CNV  copy number variant
CRC  colorectal cancer
CSGE  conformation-sensitive gel electrophoresis
CTIP/RBBP  gene for retinoblastoma binding protein 8
CTNNB1  gene for β-catenin
CYP2C19  gene for cytochrome P450, family 2, subfamily C, polypeptide 19
DAB2IP  gene for DAB2 interacting protein
DDR  DNA damage response
DCLRE1C  gene for DNA cross-link repair 1C/ ARTEMIS
DGV  database of genomic variants
DSB  DNA double-strand break
EIF2C2/AGO2  gene for argonaute RISC catalytic component 2
ESR2  gene for estrogen receptor 2
FA  Fanconi anemia
FAP  familial adenomatous polyposis
FRC  Finnish Red Cross
GWAS  genome-wide association study
H2A  histone H2A
H2AX  H2A histone family, member X
HECW2  gene for HECT, C2 and WW domain containing E3 ubiquitin protein ligase 2
HER2  gene for human epidermal growth factor receptor type 2
HRM  high-resolution melting
IR  ionizing radiation
HNPPCC  hereditary nonpolyposis colon cancer
HR  homologous recombination
ICL  interstrand crosslink
IPA  ingenuity pathways analysis software
kb  kilobase(s)
MAF  minor allele frequency
MCPH1  gene for microcephalin 1
LD  linkage disequilibrium
MDC1  gene for mediator of DNA-damage checkpoint 1
MET  MET proto-oncogene
MERIT40/BABAM1  gene for BRISC and BRCA1 A complex member 1
MMR  mismatch repair system
MMS2  gene for ubiquitin-conjugating enzyme E2 variant 2
MRE11  gene for meiotic recombination 11 homologue
MRN  MRE11-RAD50-NBS1-protein complex
NBS1  Gene for Nijmegen breakage syndrome protein
NER  nucleotide excision repair
NHEJ  nonhomologous end-joining
OR  odds ratio
P  statistical significance
PALB2/FANCN  gene for partner and localizer of BRCA2/Fanconi anemia subtype N
PCR  polymerase chain reaction
PTEN  gene for phosphatase and tensin homolog
qPCR  quantitative-PCR
RAD50  human homolog of S. cerevisiae Rad50 gene
RAD51  gene for RAD51 recombinase
RAD51B/C/D  genes for RAD51 paralogs B, C and D
RAP80  gene for receptor associated protein 80
RAS  rat sarcoma viral oncogene-family of small GTPases
RB  retinoblastoma

12
<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RB1</td>
<td>gene for retinoblastoma 1 protein</td>
</tr>
<tr>
<td>RECQL4</td>
<td>gene for RecQ protein-like 4</td>
</tr>
<tr>
<td>RET</td>
<td>gene for RET proto-oncogene</td>
</tr>
<tr>
<td>RNF8</td>
<td>gene for ring finger protein 8, E3 ubiquitin protein ligase</td>
</tr>
<tr>
<td>RNF168</td>
<td>gene for ring finger protein 168, E3 ubiquitin protein ligase</td>
</tr>
<tr>
<td>RR</td>
<td>relative risk</td>
</tr>
<tr>
<td>SCID</td>
<td>severe combined immunodeficiency</td>
</tr>
<tr>
<td>SNP</td>
<td>single nucleotide polymorphism</td>
</tr>
<tr>
<td>STK11</td>
<td>gene for serine/threonine kinase 11</td>
</tr>
<tr>
<td>STRN</td>
<td>gene for striatin, calmodulin binding protein</td>
</tr>
<tr>
<td>TLS</td>
<td>translesion synthesis</td>
</tr>
<tr>
<td>TP53</td>
<td>gene for tumor suppressor protein 53</td>
</tr>
<tr>
<td>UBC13</td>
<td>gene for ubiquitin-conjugating enzyme E2N</td>
</tr>
<tr>
<td>UTR</td>
<td>untranslated region</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet radiation</td>
</tr>
<tr>
<td>XRCC2/3</td>
<td>genes for X-ray repair complementing defective repair in Chinese hamster cells 2 and 3</td>
</tr>
</tbody>
</table>
List of original publications


*Shared first authorship.
Contents

Abstract
Tiivistelmä
Acknowledgements
Abbreviations
List of original publications
Contents
1 Introduction
2 Review of the literature
  2.1 Cancer is a complex genetic disease
  2.2 Inherited predisposition to cancer
  2.3 Characteristics of breast cancer
  2.4 Inherited breast cancer susceptibility
    2.4.1 High-penetrance genes
    2.4.2 Moderate-penetrance genes
    2.4.3 Low-penetrance genes
  2.5 Searching for the missing heritability of breast cancer susceptibility
  2.6 Cellular DNA damage response
    2.6.1 Repair of DNA double-strand breaks
  2.7 Genomic copy number variation and breast cancer susceptibility
3 Aims of the study
4 Materials and methods
  4.1 Cases and controls (I-III)
  4.1.1 Familial breast cancer cases (I-III)
  4.1.2 Unselected young breast cancer cases (III)
  4.1.3 Unselected breast and ovarian cancer cases (II)
  4.1.4 Healthy controls (I-III)
  4.2 DNA isolation (I-III)
  4.3 Mutation detection methods
    4.3.1 Conformation-sensitive gel electrophoresis (CSGE) (I)
    4.3.2 High-resolution melting (HRM) (I, II)
    4.3.3 Sequencing (I, II)
    4.3.4 MassARRAY analysis (II)
  4.4 CNV discovery with microarray analysis (III)
    4.4.1 CNV validation (III)
4.5 Statistical analysis (I, II, III) ................................................................. 54
4.6 Bioinformatical analysis ........................................................................ 54
  4.6.1 In Silico prediction for mutations pathogenicity (I, II) ................. 54
  4.6.2 Network analysis and functional profiling of rare CNVs (III) ....... 55
4.7 Ethical issues .......................................................................................... 55

5 Results
  5.1 Mutation screening of the RNF8, UBC13 and MMS2 genes (I) ....... 57
  5.2 Mutation screening of the RAD51C gene (II) ..................................... 57
  5.3 Screening of rare CNVs (III) ................................................................. 62

6 Discussion
  6.1 Pathogenic mutations in RNF8, UBC13 and MMS2 are rare .......... 73
  6.2 Mutation screening of RAD51C identifies two rare pathogenic
      deletions .............................................................................................. 74
  6.3 Estrogen signaling and TP53 tumor suppression network is
      disrupted by rare CNVs in breast cancer patients ......................... 77

7 Concluding remarks
8 Future prospects
References
Original publications
1 Introduction

About 12.7 million cancer cases and 7.6 million cancer deaths are estimated to have occurred worldwide in the year 2008. Most of the cases (56%) and of the deaths (64%) take place in the economically developing world. In females, breast cancer is world-wide the most commonly diagnosed cancer and leading cause of cancer death (Jemal et al. 2011). In Finland, breast cancer accounts for ~32% of all cancers and 16% of all cancer-related deaths. In 2011 alone, around 4,900 new breast cancer cases and 800 breast cancer-related deaths were reported (Finnish Cancer Registry 2013).

Breast cancer arises due to genetic and environmental factors. Five to ten percent of all breast cancer cases are clustered in families and expected to have a genetic background predisposing to the disease (Apostolou & Fostira 2013, Claus et al. 1996, Narod & Foulkes 2004). Most of the known breast cancer susceptibility genes have a function in DNA damage response pathways (Stratton & Rahman 2008).

The multifunctional and highly penetrant \textit{BRCA1} and \textit{BRCA2} are the two major breast cancer susceptibility genes, accounting for about 20% of breast cancer families. A small portion of the familial cases can be explained by mutations in moderate-risk breast cancer genes like \textit{ATM}, \textit{CHEK2}, \textit{BRIP1}, \textit{PALB2} and \textit{RAD50}, but for most of the remaining familial cases the underlying genetic defect remains unknown (Lalloo & Evans 2012, Walsh & King 2007). Most of the remaining cases might be explained by mutations in moderate and low penetrance genes together with environmental factors (Stratton & Rahman 2008).

Identification of new genetic risk factors is crucial for understanding the genetic etiology of breast cancer and ultimately for the design of more effective therapies. In addition, by knowing the predisposing germline defect, diagnostics and genetic counseling can be provided for patients at high risk.

Recent reports have suggested a role for genomic structural variants in susceptibility to various diseases (Girirajan et al. 2011). This relatively new class of genetic variation was discovered to exist in 2004 (Iafrate et al. 2004, Sebat et al. 2004). Copy number variants (CNVs) are the most prevalent type of structural variation in the human genome. Association of common germline CNVs with breast cancer risk has been ruled out by a large case-control study (Wellcome Trust Case Control Consortium et al. 2010). However, the role of the rare CNVs as a risk factor for breast cancer is not well studied.
The first aim of the current study was to identify further breast cancer susceptibility genes by searching for pathogenic germline mutations in the \textit{RNF8}, \textit{UBC13}, \textit{MMS2} and \textit{RAD51C} genes. These candidate genes were chosen based on their involvement in the DNA damage response. The coding regions and exon-intron boundaries of these genes were screened for mutations in familial breast cancer cases and the frequency of the observed variants was compared to that of healthy controls in order to evaluate their potential pathogenicity.

The second aim of the study was to investigate if rare CNVs in Finnish breast cancer patients might have an impact on breast cancer predisposition. CNV discovery was performed by using Illumina HumanOmni1-Quad BeadChips. The pathway, biological function and network analysis for the genes disrupted by rare CNVs was performed by using Ingenuity Pathway Analysis (IPA) software.
2 Review of the literature

2.1 Cancer is a complex genetic disease

Cancer includes more than 100 distinct diseases which originate from most of the somatic cell types and organs of the human body and which are defined by relatively uncontrollable proliferation of cells that can invade beyond normal tissue boundaries and metastasize to distant organs (Stratton et al. 2009). Different forms of cancer have one major feature in common: they are all diseases of the genome. Tumor development is associated with a breakdown of the mechanisms that control cell proliferation and the maintenance of genome integrity. The classic model of tumorigenesis describes multiple, successive clonal expansions driven by the accumulation of genomic alterations or ‘mutations’ that are selected by the tumor environment (Yates & Campbell 2012). Most cancers can be seen as ecosystems of evolving clones in which Darwinian selection operates (Aparicio & Caldas 2013, Greaves & Maley 2012).

The process from normal cell proliferation towards cancerous growth requires multiple alterations in the physiology of the cell. These alterations include: sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, and activating invasion and metastasis (Figure 1). Genome instability and tumor-promoting inflammation are considered as enabling characteristics which promote tumorigenesis. In addition, reprogramming of energy metabolism and evading immune destruction are now being considered as emerging hallmarks of cancer development (Hanahan & Weinberg 2011).

Instead of being static, the human genome is a very dynamic structure. Somatic mutations occur in the genomes of all cells through the rounds of cell division that take place during fetal development and during regeneration of tissues in postnatal life. It is estimated that each cell undergoes over 20,000 DNA-damaging events (Balmain et al. 2003, Bell et al. 1993, Branzei & Foiani 2008) and over 10,000 replication errors per day (Preston 2005). As a consequence, mutations gradually accumulate in the genome.
Fig. 1. The hallmark capabilities and enabling characteristics as well as emerging hallmarks that are essential in carcinogenesis. Sustaining proliferative signaling, evading growth suppressors, cell death resistance, replicative immortality, angiogenesis induction and invasion/metastasis activation are the six hallmark capabilities of cancer. Genome instability and tumor promoting inflammation are enabling characteristics which promote tumorigenesis. Reprogramming of energy metabolism and avoiding immune destruction are now being considered as emerging hallmarks. Modified from Hanahan & Weinberg 2011.

On average between 1,000 and 10,000 somatic substitutions are present in the genomes of most adult cancers, including breast, ovary, colorectal, pancreas, and glioma (Greenman et al. 2007, Wood et al. 2007). In contrast, there appear to be relatively few mutations in some cancers like medulloblastomas, testicular germ cell tumors, acute leukemias, and carcinoids (Greenman et al. 2007, Parsons et al. 2011). On the other hand, some cancer types, such as lung cancers and melanomas can sometimes have more than 100,000 mutations. Even within a particular cancer type, individual tumors often show wide variation in the occurrence of base substitutions and also other genomic alterations (Stratton 2011).
It is hypothesized that the mutation rate of non-malignant cells is inadequate to generate the large numbers of mutations that are present in human cancers. Thus, it is suggested that malignant cells express a mutator phenotype and as a result, these cells progressively accumulate large numbers of mutations during tumor progression (Loeb 2011). Especially mutations in genes encoding proteins that are responsible for maintaining genomic stability, such as those affecting DNA damage response, DNA repair and DNA replication, have been shown to be important in promoting mutation accumulation. Large-scale sequencing studies have identified high prevalence of mutations in DNA repair genes in cancer cell lines. The Cancer Genome Project reports that as much as 60% and 58% of cancer cell lines exhibit somatic mutations in DNA repair and replication genes, respectively (Wellcome Trust Sanger Institute 2013).

Mutations accumulating during tumorigenesis can be classified in two separate classes. Mutations in genes that contribute to cellular growth advantage are called “driver” mutations. The remaining and often representing the majority of the occurring mutations are referred to as “passengers”, which do not confer growth advantage to the neoplastic clone (Stratton 2011).

Genes that promote cancer development can be classified into two major classes: oncogenes and tumor suppressor genes. Normal cell proliferation is regulated by a finely tuned balance between these two groups of genes. Oncogenes and their normal counterparts, proto-oncogenes, encode proteins that support cell proliferation and differentiation. Proto-oncogenes are altered to oncogenes by dominant gain of function mutations. Oncogenes can be classified into five groups based on their function: growth factors (e.g. \textit{PDGFB}), growth factor receptors (\textit{EGFR, MET, RET}), signal transducers (\textit{RAS} family, \textit{SRC}), transcription factors (\textit{MYC, JUN, FOS}), and apoptosis regulators (\textit{BCL2}) (Kufe \textit{et al.} 2003). In contrast, tumor suppressors act in the normal cell as negative controllers of cell growth. Based on their roles, tumor suppressors are further classified into “gatekeepers” or “caretakers”. Gatekeeper genes directly regulate the growth of tumor cells by controlling cellular proliferation or promoting cell death. Caretakers are responsible for maintaining the genetic stability of a cell, i.e., they maintain the genome’s integrity (Hoeijmakers 2001).

The tumor suppressor genes are inactivated by loss-of-function alterations during cancer development. Some of the genes can have both gatekeeper and caretaker functions. Examples of tumor suppressor genes include \textit{RB1} (functions in cell division, DNA replication, cell death), \textit{TP53} (cell division, DNA repair, cell death), \textit{CDKN2A} (cell division, cell death), \textit{APC} (cell division, DNA damage,
cell migration, cell adhesion, cell death), MLH1, MSH2, MSH6 (DNA mismatch repair, cell cycle regulation), BRCA1, BRCA2 (repair of double-stranded DNA breaks, cell division, cell death), WT1, WT2 (cell division, transcriptional regulation), NF1, NF2 (RAS-mediated signal transduction, cell differentiation, cell division, developmental processes) and VHL (cell division, cell death, cell differentiation, response to cell stress) (Chial 2008).

Defective tumor suppressor genes usually act recessively and thus both copies need to be inactivated for a loss-of-function. This “two-hit” model was originally proposed by Alfred Knudson (Knudson 1971). Based on epidemiological studies of familial retinoblastoma (RB) he concluded that one mutated allele is inherited from one of the parents and the second one is somatically inactivated, leading to loss of heterozygosity (LOH) in the tumor. In the nonhereditary form, both mutations occur in somatic cells. Although hereditary RB is caused by recessive inactivating mutations in the RB1 gene, the disease predisposition is inherited dominantly. The two-hit model has provided a useful conceptual framework for four decades of research on tumor suppressor genes, but recently it has become evident that mutations in tumor suppressor genes are not always completely recessive. Tumor suppressor gene mutations can also be haploinsufficient, dominant-negative or associated with a gain of function. This means that under certain circumstances, one hit in a tumor suppressor gene may be sufficient to promote tumorigenesis. Haploinsufficiency occurs when one allele is insufficient to produce the quantity of protein needed for the full functionality produced from two wild-type alleles. A dominant-negative mutation changes the gene product so that it interferes with the function of the wild-type protein, while gain-of-function mutations alter the gene product in such a way that it gains a new and abnormal function (Payne & Kemp 2005).

The prevailing dogma of cancer development is “gradualism” in which acquisition of driver mutations occurs cumulatively over years to decades. Recent cancer genome analyses have revisited this dogma by proposing an alternative process that involves massive de novo structural rearrangement formation in a one-step catastrophic genomic event named chromothripsis (Korbel & Campbell 2013). The phenomenon of chromothripsis can be seen in at least 2–3% of all cancers, across many subtypes, and is for instance present in ~25% of all bone cancers (Stephens et al. 2011).
2.2 Inherited predisposition to cancer

Although cancer is always caused by defects in the genome of the cells, its incidence can be either sporadic or show familial clustering. Familial clustering can be due to both shared environmental or genetic factors that are clustered in families. Studies of twins make it possible to estimate the overall contribution of inherited gene defects to the development of malignant diseases (Neale & Cardon 1992).

A small proportion of the genetic tendency to develop cancer is explained by highly penetrant mutations that show Mendelian inheritance and cause strong genetic predisposition. Most of the genes associated with familial cancer syndromes are tumor suppressors. To date only three genes, \textit{RET} (in multiple endocrine neoplasia type 2), \textit{MET} (hereditary papillary renal cell carcinoma) and \textit{CDK4} (familial malignant melanoma) function as oncogenes (Marsh & Zori 2002). Inherited cases of retinoblastoma, Wilm’s tumor and xeroderma pigmentosum are examples of cancer syndromes caused by highly penetrant mutations in tumor suppressor genes. Natural selection keeps these mutations rare because they cause severe disease already early in life. Other examples of inherited cancer syndromes caused by highly penetrant mutations are the two colon cancer syndromes: familial adenomatous polyposis (FAP), caused by dominantly inherited mutations of the \textit{APC} gene, and hereditary nonpolyposis colon cancer (HNPCC), caused by mutations in the DNA mismatch repair (MMR) system (Frank 2004, Ollila et al. 2008). FAP and HNPCC both have a median age of onset of about 40–42 years (Frank 2004, Lynch et al. 1995). Because of later disease onset, the population frequency of these mutations can be higher than for mutations causing childhood cancers. Overall, more than 20 different hereditary cancer syndromes have been described and specific germline mutations in various genes have been identified as causative factor (Fearon 1997).

High-penetrance mutations that display Mendelian inheritance are easy to identify, and most of the loci that are associated with a high level of risk might already have been identified. However, occasionally cancers cluster in families but do not follow the rigid inheritance pattern characteristic of a mutation in a single gene. It is more difficult to identify alleles with lower penetrance or later disease onset and to assess their relative importance. Polygenic inheritance involving many allelic variants, each of which has a small effect, seems to dominate genetic predisposition for late-onset epithelial cancers (Frank 2004).
2.3 Characteristics of breast cancer

Breast cancer is the most common malignancy and the second leading cause of cancer mortality among women, with over 1,300,000 cases and 450,000 deaths each year worldwide (Jemal et al. 2011). One in eight women will develop breast cancer during her lifetime (Downs-Holmes & Silverman 2011). In Finland, breast cancer accounts for ~33% of all cancers and ~15% for all cancer-related deaths. In 2011, around 4,800 new breast cancer cases and 800 breast cancer-related deaths were reported (Finnish Cancer Registry 2013). Breast cancer incidences have been observed to increase in many Western countries; this is possibly a result of changes in reproductive factors (including the increased use of postmenopausal hormone therapy) as well as in systematic screening protocols. At the same time a reduction in mortality from breast cancer during the past two decades is observed, mostly due to diagnosis at earlier stage and associated with the widespread administration of adjuvant therapies (Benson & Jatoi 2012, Jemal et al. 2011).

The term breast cancer is used for all malignancies originating from breast tissue. Most commonly breast malignancies develop from the epithelial cells lining inside of the milk-producing ducts and lobules. Cancer originating from epithelium is called carcinoma. About 80% of breast cancers are ductal carcinomas, between 5–10% are infiltrating lobular carcinomas, and the remainder arise from diverse cell types, including medullar, mucinous, papillary and tubular carcinomas, as well as Paget’s disease (Berg & Hutter 1995). Breast cancer is clinically a very heterogeneous disease. More detailed classification of breast tumor subtypes is based on histopathology, gene-expression patterns or integrative classification based on genomic and transcriptomic data (Dvinge et al. 2013). Histopathologically breast tumors are classified according to estrogen receptor status (ER+/−), progesterone receptor status (PR+/−) and human epidermal growth factor receptor type 2 status (HER2+/−) (Onitilo et al. 2009). Global gene expression analysis of breast tumors has identified five “intrinsic” subtypes of breast cancer known as luminal A, luminal B, HER2-Enriched (HER2-E), basal-like and normal breast-like tumor (Hu et al. 2006, Sorlie et al. 2001). A clinically applicable gene expression-based predictor that robustly identifies these main intrinsic subtypes by quantitative measurement of 50 genes (PAM50) has been developed (Parker et al. 2009). Integrative classification by using genomic and transcriptomic data has also been explored. Curtis et al. (Curtis et al. 2012) combined copy number aberration and gene expression data in
Breast cancer has a complex etiology where predisposition is influenced by both environmental and genetic factors. Remarkable experimental and epidemiological evidence suggests that lifetime exposure to endogenous hormones, particularly estrogens and androgens, promotes breast carcinogenesis. In population-based studies, factors related to increased estrogen exposure throughout a woman’s lifetime, such as early menarche, late menopause, use of oral contraceptives, and hormone replacement therapy, have been associated with a ~2-fold increase in breast cancer risk among premenopausal women (Kaaks et al. 2005a, Kaaks et al. 2005b, Key et al. 2002). Other factors such as age, late age (>30 years) at first pregnancy or never being pregnant, and high breast density are also important in defining the risk for breast cancer (Rim & Chellman-Jeffers 2008). Variable and modifiable factors like nutrition, exercise, and alcohol or tobacco use also influence breast cancer risk (Bissonauth et al. 2008). All women have a similar number of ductal and lobular epithelial cells regardless of absolute breast size; breast size is thus not a significant risk-determining factor for breast cancer (Bunz 2008). Overall, the strongest risk factor is family history of breast cancer. The individual risk increases with an increasing number of relatives affected with breast cancer (McPherson et al. 2000).

### 2.4 Inherited breast cancer susceptibility

It is estimated that approximately 5–10% of breast cancer cases stem from a major hereditary predisposition to the disease (Apostolou & Fostira 2013, Claus et al. 1996, Narod & Foulkes 2004). Familial breast cancer is classified by a clustering of breast cancers in a family that is substantially higher than the observed incidence in the general population. Familial predisposition is also thought to manifest as early disease onset, occurrence of bilateral breast cancer or multiple primary tumors in the same individual or male breast cancer incidences in the family. In familial breast cancer the disease is usually seen in individuals of several consecutive generations, which indicates that inherited factors must contribute strongly to the malignancy predisposition (Thull & Vogel 2004).

Twin studies have confirmed genetic factors to be predominant component of familial aggregation in breast cancer. Statistically significant effects of hereditary factors were seen in 27% of the cases studied. In addition, a monozygotic twin is at higher risk of developing breast cancer by the age of 75 years than a dizygotic
twin of an affected individual. The estimated risk is 18% for a monozygotic twin and 9% for a dizygotic twin (Lichtenstein et al. 2000, Peto & Mack 2000).

The co-occurrence of sporadic and heritable breast cancer within families makes identification of genes associated with the breast cancer susceptibility difficult. Breast cancer affects one in eight women and thus approximately 11% of families will contain more than one female with breast cancer (Downs-Holmes & Silverman 2011, Ellsworth et al. 2010). Therefore, it may be difficult to distinguish disease attributable to an inherited cancer susceptibility gene from a coincidence clustering of sporadic breast cancer cases within a family.

Human disease genes can be located by linkage analysis to families in which the incidence of the disease is high. Linkage analysis can reveal the chromosomal location of the genes of interest by identifying polymorphic genetic markers of known location that are co-inherited with the disease in multiple-case families (Ott 1999). Using this approach, Hall et al. (1990) assembled 23 extended Caucasian families containing 146 individuals with breast cancer, including early onset, bilateral disease, and/or male breast cancer cases. Forty percent of these families indicated strong linkage to a marker on chromosome 17q21. This locus was formally labeled as early onset familial breast cancer 1 (BRCA1) (Solomon & Ledbetter 1991). In 1994, a novel gene with a zinc-finger domain and wide tissue expression, including breast and ovary, was identified as BRCA1 (Miki et al. 1994). Later on, pathogenic germline mutations in BRCA1 were found to account for about 7–10% of all familial breast cancers (Peto et al. 1999).

Regardless of the strong contribution of BRCA1 to hereditary breast and ovarian cancer, mutations in BRCA1 did not account for all cases of inherited breast cancer, implicating the existence of another major susceptibility gene or multiple different genes. Using techniques similar to those used in the discovery of BRCA1, Wooster et al. (1994) discovered a novel gene on chromosome 13q12-q13 that was linked mainly to hereditary breast cancer risk. The gene was identified as BRCA2 in 1995 (Wooster et al. 1995). BRCA1 and BRCA2 have shown notable difference between them. Mutations in BRCA1 explain few if any families with both male and female breast cancer, but are associated with increased ovarian cancer risk. In contrast, BRCA2 mutations were to some extent also connected to the occurrence of male breast cancer, but did not typically seem to associate with a high risk for ovarian cancer (Wooster et al. 1994).

After identifying the BRCA1 and BRCA2 genes it soon became evident that mutations in these two genes are not responsible for all familial clustering of breast and ovarian cancer (Serova et al. 1997). Many novel additional
predisposing genetic risk factors for particularly breast cancer have already been identified. These genetic risk factors can be divided into different classes according to the level of risk they confer and the prevalence of the disease causing variants in the population. Penetrance describes the likelihood that an individual who carries a mutation or a genetic variant will develop breast cancer as a consequence of its presence. Currently three different classes of disease risk factors have been described (Figure 2): rare high-penetrance mutations with a relative risk (RR) varying from 10- to 20-fold, rare moderate-penetrance mutations (2- to 4-fold RR) and common low-penetrance mutations (≤1.5-fold RR) (Apostolou & Fostira 2013, Stratton & Rahman 2008).

New breast cancer predisposing factors are actively searched for by several different approaches. Linkage analyses of \textit{BRCA1} and \textit{BRCA2} mutation-negative families have not revealed any new high-penetrance susceptibility genes and thus it is highly likely that genes exhibiting high-penetrance mutations similar to \textit{BRCA1} and \textit{BRCA2} will not be identified (Smith \textit{et al.} 2006). In the candidate gene re-sequencing approach a plausible breast cancer gene is selected and screened for mutations by sequencing. Whole exome or whole genome analysis by next-generation sequencing provides a good approach to study genes responsible for diseases with a high degree of genetic heterogeneity, a class to which also breast cancer belongs to (Gracia-Aznarez \textit{et al.} 2013). Genome-wide association studies (GWAS) have been used to detect common lower penetrance alleles. In addition, germline copy number variants can be detected by next-generation sequencing and microarray technologies.
Fig. 2. Breast cancer susceptibility loci and genes are expressed according to their disease risk and population frequency. Rare high risk alleles have been detected by performing linkage analysis in multiple-case breast cancer families. Moderate-risk alleles have been identified largely by using candidate gene re-sequencing in familial breast cancer cohorts. Common low-risk alleles have been identified mainly by genome-wide association studies in unselected cases. Common high-risk alleles are not believed to exist and rare low risk-alleles are difficult to detect. Modified from Foulkes 2008.

### 2.4.1 High-penetrance genes

Overall, high-risk susceptibility genes explain about 22% of all inherited breast cancers (Ellsworth et al. 2010). Mutations in the two major genes, \textit{BRCA1} and \textit{BRCA2}, exhibit a dominant disease inheritance pattern and account for 17% to 20% of all familial breast cancers (Lalloo & Evans 2012, Peto et al. 1999). Disease predisposing variants in \textit{BRCA1} and \textit{BRCA2} confer a high risk of breast cancer, approximately 10- to 20-fold relative risk. The relative risks are higher at younger ages. \textit{BRCA1/2} mutation carriers also have elevated risks of ovarian and
some other cancers (The Breast Cancer Linkage Consortium 1999, Thompson et al. 2002b). Additional high-penetrance germline mutations have been identified in genes causative for specific inherited cancer syndromes in which breast cancer development is a common feature. These include TP53 mutations found in Li-Fraumeni syndrome (Malkin 1994), PTEN mutations in Cowden syndrome (Nelen et al. 1996), STK11 mutations in Peutz-Jegher syndrome (Hemminki et al. 1998) and CDH1 (E-Cadherin) mutations in hereditary diffuse gastric cancer (Guilford et al. 1998). Nevertheless, the risk of mutations in these genes for causing breast cancer is often lower and less well defined than those of BRCA1 and BRCA2. Furthermore, germline mutations of TP53 (Rapakko et al. 2001), STK11 (Chen & Lindblom 2000, Guenard et al. 2010), PTEN (FitzGerald et al. 1998), and CDH1 (Masciari et al. 2007, Schrader et al. 2011) are very rarely observed in breast cancer patients without manifestations of the other typical features of these syndromes.

Of the major susceptibility genes, BRCA1 is very well characterized. It is a large gene with 24 exons encoding a protein of 1,863 amino acids. Exon 11 is unusually large. Mutations are found throughout the coding sequence of the gene but a majority of the pathogenic mutations are frameshift mutations resulting in truncated proteins. Approximately 2% of the pathogenic mutations of BRCA1 are missense mutations, but they may be difficult to distinguish from polymorphisms. Between 15% and 27% of the mutations may be due to large rearrangements, including large deletions (whole exon) and insertion/duplications (Lalloo & Evans 2012, Thompson et al. 2002b). Within certain populations particular mutations are more common (founder mutations), but hundreds of unique pathogenic BRCA1 mutations have also been described. Within the Ashkenazi Jewish population, two founder mutations, c.68_69delAG and c.5266dupC, exist in about 1.2% of the population (Lalloo & Evans 2012). The c.5266dupC mutation is also observed in other Eastern European populations, particularly in Poland (Hamel et al. 2011).

BRCA2 mutations account for ~10% of the families with breast and ovarian cancers. BRCA2 is larger than the BRCA1 gene with 27 exons encoding a 3,418 amino acid protein, exon 11 being the largest. Similar to BRCA1, mutations occur throughout the BRCA2 gene, the majority being frameshifts. Large numbers of missense mutations are found within BRCA2, but the pathogenicity of these may be difficult to establish. In addition, large gene rearrangements also occur in BRCA2, but they are less frequent than for BRCA1 (Lalloo & Evans 2012).
The main function of BRCA1 seems to be associated with DNA repair, including homologous recombination (HR) and nucleotide excision repair. However, it also has a function in the regulation of cell-cycle progression, particularly G2/M and S-phase checkpoint control (Roy et al. 2011). In contrast to the multifunctional activities of BRCA1, the primary function of BRCA2 is in homologous recombination. BRCA2 regulates RAD51 loading to DNA double-strand breaks (DSBs). RAD51 is an essential protein that covers appropriate sites of single-stranded DNA, thus preventing it from binding to double-stranded DNA, and therefore activates strand invasion in HR (Roy et al. 2011). Cells that lack BRCA1 or BRCA2 are hypersensitive to DNA-damaging agents that cause DSBs. Without proper ability to execute HR repair, DSBs are repaired by error-prone mechanisms such as non-homologous end joining, resulting in chromosomal rearrangements and instability (Lalloo & Evans 2012). As discussed earlier (chapter 2.1), chromosomal instability is a crucial feature in tumorigenesis.

Interestingly, biallelic mutations in BRCA2 (FANCD1) have been reported to cause Fanconi anemia (FA) complementation group D1 (Howlett et al. 2002). FA is a rare autosomal recessive disease characterized by developmental anomalies, such as abnormalities of the thumbs, bone marrow failure and predisposition to various childhood cancers. The classic diagnosis of FA utilizes the detection of increased chromosomal breakage in the presence of DNA interstrand cross-linking (ICL) reagents, such as diepoxybutane (DEB) and mitomycin C (MMC). FA is genetically heterogeneous disease and 15 genes have already been identified to be mutated in the known complementation groups: FANCA, FANCB, FANCC, FANCD1, FANCD2, FANCE, FANCF, FANCG, FANCI, FANCI, FANCL, FANCM, FANCN, FANCO and FANCP (Kee & D'Andrea 2012, Levy-Lahad 2010, Moldovan & D'Andrea 2009). The protein products of these genes have a function in ICL resolution which is a multistep DNA repair process including nucleotide excision repair (NER), translesion synthesis (TLS) and homologous recombination (HR).

Curiously, biallelic (compound heterozygous for deleterious c.2457delC and cancer associated missense mutation p.Val1736Ala) mutations in BRCA1 were recently reported in one early-onset (28 years) ovarian carcinoma case, whose clinical phenotype also manifested short stature, microcephaly, developmental delay, and significant toxicity from chemotherapy. This patient did not show any obvious abnormalities in her thumbs and had a normal complete blood count at the time of her cancer diagnosis, thus lacking some of the typical features of FA
Nevertheless, short stature and microcephaly are observed in some FA cases (Auerbach 2009).

Progress in the understanding of the functions of \( \text{BRCA1} \) and \( \text{BRCA2} \) in the DNA damage response pathway has been instrumental for the identification of a number of other breast cancer susceptibility genes, including \( \text{PALB2}, \text{BRIP1}, \text{ATM}, \) and \( \text{CHEK2} \), all with protein products interacting directly or indirectly with \( \text{BRCA1} \) and \( \text{BRCA2} \) (Foulkes & Shuen 2013, Walsh & King 2007).

### 2.4.2 Moderate-penetrance genes

The most important moderate risk breast cancer genes include \( \text{ATM}, \text{CHEK2}, \text{BRIP1}, \text{PALB2}, \) and \( \text{RAD50} \). Mutations in these genes have recently been identified to associate with a relative two- to fourfold risk of breast cancer. The pathogenic mutations in these genes are rare with a population frequency of \(<0.6\%\). However, as the phenotypes associated with most of these monoallelic mutations have not been adequately defined, the clinical utility of these genotypes has yet to be established (Lalloo & Evans 2012, Walsh & King 2007). Overall, moderate penetrance genes appear to explain less than 3\% of all familial breast cancers (Ellsworth et al. 2010). Most of the moderate penetrance genes have been found by candidate gene re-sequencing studies, on genes from the same pathways where \( \text{BRCA1} \) and \( \text{BRCA2} \) operate (Walsh & King 2007). Curiously, whereas monoallelic mutations in many of these moderate penetrance genes predispose to breast cancer, bilallelic mutations in the same gene result in early onset cancer predisposition syndromes, reminiscent of the high-penetrance \( \text{BRCA2} \) mutations (Howlett et al. 2002).

Biallelic mutations in \( \text{ATM} \) cause ataxia-telangiectasia, a rare congenital syndrome characterized by progressive cerebellar ataxia, immune deficiency and cancer predisposition that includes breast cancer (Easton 1994). Monoallelic female \( \text{ATM} \) mutation carriers do not display the ataxia-telangiectasia phenotype, but do have an increased susceptibility to breast cancer with a relative risk of 2.4 (Renwick et al. 2006), or as high as 12- or 15-fold for some mutations (Chenevix-Trench et al. 2002, Pylkäs et al. 2007).

Checkpoint kinase 2 (CHEK2) is a multifunctional enzyme whose functions are central in the induction of cell cycle arrest and apoptosis after DNA damage (Ahn et al. 2004). Somatic mutations in this tumor suppressor have been identified in a number of malignancies (Lalloo & Evans 2012). CHEK2 functions in the same pathway as \( \text{p53} \) and \( \text{BRCA1} \). Specific germline mutation 1100delC,
which generates truncating protein product, has been shown to have a relative risk for breast cancer of 2.3 (Meijers-Heijboer et al. 2002). CHEK2*1100delC variant is present in 0.2–1% of European populations and in 4.2% of breast cancer families, although the mutation frequency varies between populations. In addition, families with homozygous CHEK2*1100delC mutations have been identified (Adank et al. 2011). Although otherwise without any specific phenotypic features, individuals homozygous for the mutation have a much higher risk of breast cancer – estimated to be six-fold. Other CHEK2 mutations have also been reported in breast cancer families, but their clinical significance is still unclear (Lalloo & Evans 2012).

**BRIP1** (BRCA1-interacting protein) encodes for a protein that was identified as a binding partner of BRCA1 and in 2006, heterozygous truncating mutations were identified in breast cancer families (Seal et al. 2006). Segregation analysis found that the relative risk of breast cancer in heterozygous mutation carriers is 2.0. However, higher risk **BRIP1** alleles have also been reported (Rafnar et al. 2011). Biallelic mutations of **BRIP1** cause FA complementation group J. **BRIP1/FANCJ** was the second (after **BRCA2/FANCD1**) FA pathway-related gene that was associated with breast cancer susceptibility (Levitus et al. 2005, Levrán et al. 2005). The phenotype of the FANCJ complementation group is different from that of biallelic mutations in **BRCA2/FANCD1** and results in a much lower rate of childhood solid tumors. The FA pathway will be discussed in more detail in chapter 2.5.1. **BRIP1** mutations may also confer an increased risk to ovarian cancer (Rafnar et al. 2011).

Mutations in **PALB2** (partner and localizer of **BRCA2**) were identified in breast cancer families negative for mutations in **BRCA1/2** (Erkko et al. 2007, Rahman et al. 2007). PALB2 interacts with BRCA2 during homologous recombination mediated repair of DSBs. The relative risk of breast cancer associated with heterozygous **PALB2** germline mutations has been estimated to be ~2.3 (Rahman et al. 2007). However, some specific mutations can cause even higher risks. Indeed, the Finnish **PALB2** c.1592delT founder mutation is estimated to result in a 6-fold risk for breast cancer (Erkko et al. 2008), and the c.3113G>A mutation observed in the Australian population is estimated to cause a still much higher breast cancer risk (hazard ratio = 30.1 and cumulative risk estimates 49% to age 50 and 91% to age 70) (Southey et al. 2010). Biallelic mutations in **PALB2** cause FA complementation group N (FANCN), which is phenotypically similar to that of complementation group D1 caused by biallelic mutations in **BRCA2** (Xia...
et al. 2007). Reminiscent to BRCA2, heterozygous mutations in PALB2 have also been associated with susceptibility to pancreatic cancer (Jones et al. 2009).

RAD50 is an essential DNA double-strand break repair factor. Together with NBS1 (alias NBN = nibrin) and MRE11 it composes the Mre11 complex that has functions in sensing and early processing of DSBs and in cell cycle checkpoints, DNA recombination and maintenance of telomeres (Stracker et al. 2004, Wu et al. 2000, Zhu et al. 2000). Germline mutations in RAD50 have been associated with breast cancer by Heikkinen and co-workers who discovered a Finnish founder mutation, RAD50 687delT, to be associated with a 4.3-fold increase in breast cancer risk (Heikkinen et al. 2006). This mutation, however, has not been found in any other populations, making confirmation of the association difficult. One patient with a variant form of Nijmegen breakage syndrome, but without immunodeficiency, was shown to harbor biallelic mutations in RAD50 (Waltes et al. 2009).

Recently, by candidate gene re-sequencing mutations in two RAD51 paralogs, RAD51C and RAD51D, have been identified in breast/ovarian cancer kindreds, but not breast cancer only families (Loveday et al. 2011, Meindl et al. 2010). The role of RAD51C mutations in breast cancer susceptibility will be described in more detail later on in chapter 2.8.

2.4.3 Low-penetration genes

In addition to rare high- and moderate-penetrance mutations, the “common disease, common variant” model predicts that a large proportion of the variance in breast cancer risk in the population is due to the cumulative effect of the “normal” variation found in the genome (Antoniou & Easton 2006, Pharoah et al. 2002). A number of common variations have now been identified to be associated with a slightly increased or decreased risk of breast cancer. These may work together in a polygenic multiplicative way to account for the remainder of familial breast cancer (Lalloo & Evans 2012). Low-penetrance genes confer a relative risk less than 1.5 for breast cancer, but their heterozygote frequencies are high, ranging from 10% to 90% in the general population (Pharoah et al. 2008, Turnbull & Rahman 2008). Overall, low-penetrance genes have been predicted to explain ~5% of all familial breast cancers (Ellsworth et al. 2010).

The research of low-penetrance allelic variants has been conducted mainly through genome-wide association studies (GWAS) that use a large number of common SNPs, spreading throughout the human genome, to identify associations
with the disease that rely upon patterns of linkage disequilibrium (LD). The association of genetic variants at different loci is evaluated in large series of cases and controls, analyzing panels that can include as much as hundred thousand SNPs, to identify new alleles of susceptibility to breast cancer (Fanale et al. 2012). Examples of common low risk alleles include genes TOX3, LSP1, MAP3K as well as chromosomal regions with no annotated function 2q35 and 6q22 (Figure 2)(Foulkes 2008, Stacey et al. 2007).


**2.5 Searching for the missing heritability of breast cancer susceptibility**

Pathogenic mutations in currently known breast cancer predisposing genes account for 25–30% of familial breast cancers (Ellsworth et al. 2010, Turnbull & Rahman 2008). Thus, approximately 70% of the genetic predisposition to breast cancer is still unknown, highlighting the need for a search for new susceptibility genes and other disease predisposing factors. A majority of the remaining familial breast cancer cases is predicted to involve mutations in moderate- and low-penetrance susceptibility genes together with influence from environmental factors, even though the data from large multiple-case families strongly suggest that there might still be some novel high-penetrance genes to be identified (Michailidou et al. 2013, Stratton & Rahman 2008). There is strong evidence that rare genetic variation is important in breast and ovarian cancer predisposition (Gayther & Pharoah 2010, Turnbull & Rahman 2008).

Virtually all of the at least moderate breast cancer risk alleles identified to date are related to fundamental processes in cellular proliferation and DNA integrity control. Natural selection keeps the frequency of these alleles low as biallelic mutations in most of these genes have severe consequences, these
individuals being either non-viable or having reduced fertility. It is likely that additional rare breast cancer predisposing alleles will be found and the genes involved in the FA- and BRCA-pathway are evidently the most plausible candidates for harboring such mutations. Recently, two distinct strategies have been used to harvest new breast cancer susceptibility alleles: direct re-sequencing of genes believed to be strong candidates based on their biological function, and genome-wide tag SNP association studies. The former has led to the identification of rare moderate-penetrance alleles and the latter to identification of common low-penetrance alleles (Stratton & Rahman 2008). In addition, the next-generation sequencing technologies enable sequencing of whole genomes and exomes. So far four exome sequencing studies investigating breast cancer predisposition have been published (Gracia-Aznarez et al. 2013, Ruark et al. 2013, Snape et al. 2012, Thompson et al. 2012b).

2.6 Cellular DNA damage response

Damage to genetic material poses a serious threat to cells. External sources of DNA damage are exposure to ionizing radiation (IR), ultraviolet radiation (UV), or environmental toxins. Endogenous sources such as reactive oxygen species or errors during DNA replication can also lead to DNA damage. A wide range of DNA lesions can be generated by these events, including modified bases or sugar residues, the formation of DNA adducts, crosslinking of the DNA strands, and production of single- and double-strand breaks (Price & D'Andrea 2013). Cells have therefore developed the DNA damage response (DDR) as a defense system against these threats. DDR is an elaborate signaling network activated by DNA damage, which swiftly modulates many physiological processes, such as cell cycle, DNA replication, DNA-repair and apoptosis, to prevent lesion transmission to daughter cells (Ciccia & Elledge 2010, Hoeijmakers 2001). At least six different DNA-repair pathways to deal with distinct types of DNA damage are known: homologous recombination (HR), non-homologous end joining (NHEJ), base excision repair (BER), mismatch repair (MMR), nucleotide excision repair (NER) and translesion DNA synthesis (TLS). The FA/BRCA pathway that is also involved in repair of DNA crosslinks is important in coordinating the functions of other pathways (Kennedy & D'Andrea 2006). In these DNA repair pathways, complex signal transduction cascades are operating that have the ability to sense DNA damage, transduce this information and finally repair the damage. The most relevant pathways so far associated with hereditary predisposition to breast cancer
involve both HR and FA pathways. However, a direct link to the entire FA pathway is questionable as it appears that it is only the downstream part of the pathway (equaling to HR) that is more important for breast cancer predisposition (Figure 4) (Li & Greenberg 2012).

2.6.1 Repair of DNA double-strand breaks

NHEJ and HR are the two major pathways to repair DSBs. In homologous recombination, DSBs are repaired through the alignment of homologous sequences of DNA and it occurs primarily during the late S to M phase of the cell cycle. HR will be discussed in more detail in subsequent chapters. In comparison, NHEJ is active throughout the entire cell cycle but it is thought to be the preferred choice as a repair pathway during G0, G1 and early S phase. NHEJ is highly efficient pathway but it is more error-prone than HR. Thus, NHEJ can also introduce mutations during repair (Lieber & Wilson 2010). The characteristics of NHEJ are the direct ligation of broken DNA ends and minimal DNA end processing. Broken DNA ends are first recognized, captured and brought together by the Ku70/Ku80 heterodimer which then recruits and activates the DNA-dependent protein kinase (DNA-PK) holoenzyme (Gottlieb & Jackson 1993). The catalytic subunits of DNA-PK juxtapose broken DNA ends. The XLF-XRCC4-LigaseIV complex then ligates these juxtaposed DNA. If the NHEJ is hindered, alternative end-joining pathways can take over; they usually rely on terminal microhomologies for the joining and involve certain factors that function in HR and single-strand break (SSB) repair, such as MRE11-RAD50-NBS1 (MRN) complex (Polo & Jackson 2011).

RNF8, UBC13 and MMS2 function in DDR signal transduction pathway

The most important feature of HR is that it precisely restores the genomic sequence of the broken DNA ends by utilizing sister chromatids as a template for the repair (Ciccia & Elledge 2010). DSBs can be recognized and stabilized by the MRN sensor complex, which promotes the recruitment and activation of the ATM protein kinase at the site of DSB (Williams et al. 2007). The MRN complex is also involved in later steps of the pathway when it prepares DNA ends to suitable form for HR. Once activated ATM can phosphorylate many proteins (Figure 3) (Matsuoka et al. 2007), including the histone H2A variant H2AX and MDC1. Phosphorylated H2AX (γ-H2AX) marks chromatin at the DSB site which then
serves as a docking site for MDC1, a mediator of the repair process (Stucki & Jackson 2006).

Phosphorylated MDC1 is recognized by E3 ubiquitin ligase RNF8 by its FHA domain (Figure 3). RNF8 (RING finger 8), together with its E2, UBC13, then ubiquitinates histones H2A and H2AX. Ubiquitinated H2A histones are recognized by another E3 ubiquitin ligase, RNF168, which then interacts with UBC13 to amplify RNF8-dependent histone ubiquitination (Doil et al. 2009). The ubiquitin E2 variant (UEV) MMS2 is important for the functions of UBC13. MMS2 forms a complex with UBC13 and this heterodimer formation has been shown to be essential for the DNA damage repair function of UBC13 (Andersen et al. 2005, Hofmann & Pickart 1999). The RNF8/RNF168-mediated ubiquitination cascade results in an ubiquitin-mediated recruitment of effector complexes, including the ABRAXAS–BRCA1–BRCC36-BRCC45-MERIT40-RAP80 complex through the ubiquitin-interaction motif (UIM) of RAP80 (Chen et al. 2006, Dong et al. 2003, Kim et al. 2007, Liu et al. 2007, Shao et al. 2009, Sobhian et al. 2007, Wang et al. 2007).

Breast cancer risk increasing mutations have recently been identified in genes which encode proteins operating closely with RNF8, UBC13 and MMS2. For instance, germline mutations in RAP80 and ABRAXAS have been shown to increase breast cancer risk (Nikkilä et al. 2009, Solyom et al. 2012). In addition, the rare allele of a single-nucleotide polymorphism (SNP) rs8170 in MERIT40 has been associated with breast cancer and serous epithelial ovarian cancer risk (Antoniou et al. 2010, Bolton et al. 2010). Because the protein products of the RNF8, UBC13 and MMS2 genes have an integral role in the DDR they are potential tumor suppressor genes (caretakers) and thus good candidates for breast cancer susceptibility genes.

Interestingly, mutations in RNF168, which operates closely with RNF8, UBC13 and MMS2, are causally linked to the human RIDDLE syndrome. The RNF168−/− mice, which model the RIDDLE syndrome, display immunodeficiency and exhibit increased radiosensitivity and are cancer-prone. This highlights the importance of proper DNA break signaling in maintaining genomic integrity in mammals (Bohgaki et al. 2011).
Fig. 3. Schematic representation of the DNA damage response pathway. In the first steps of DNA damage response ATM phosphorylates histone H2AX. Phosphorylated H2AX (γH2AX) is recognized by MDC1 and is subsequently phosphorylated by ATM. The ubiquitin E3 ligase RNF8 binds to the phosphorylated MDC1 and together with the E2 ubiquitin-conjugating enzyme UBC13, they synthesize K63-linked ubiquitin on histone H2A and its variants. E3 ligase RNF168 recognizes the K63-linked ubiquitin chains synthesized by RNF8/UBC13, leading to its recruitment to sites of damage. RNF168/UBC13 further synthesizes K63-linked ubiquitin chains which RAP80 specifically binds. Via RAP80 a protein complex including ABRAXAS, MERIT40, BRCC36, BRCC45, and BRCA1 is recruited to sites of damage. Modified from Tang & Greenberg 2010. P, phosphate; Ub, ubiquitin.

RAD51C functions in HR repair and FA pathway

After the signal transduction the actual repair of the DSB is executed. In addition to stabilizing DNA ends, MRE11 has endonuclease and exonuclease activities important for the initial steps of DNA end resection that is crucial for HR (Williams et al. 2007). In resection DNA is cleaved in 5’-3’ direction to produce recombinogenic 3’ single-stranded tails. Complex formation between BRCA1,
CTIP and MRN is important for facilitating DSB end resection (Chen et al. 2008). RECQ family helicases and the nucleases ExoI and Dna2 are also essential in DNA end resection (Longhese et al. 2010). After resection, the ssDNA overhangs formed are rapidly coated by the ssDNA-binding complex RPA, which is then substituted with RAD51 nucleofilaments that mediate a homology search in the sister chromatid. In this process 3’ single-stranded tails from the damaged chromatid invade the complementary sequence of the sister chromatid (Davies et al. 2001, Moynahan et al. 2001). RAD51 paralogs and other proteins, which comprise the downstream part of the classical FA pathway, such as FANCD1/BRCA2 and FANCN/PALB2, are needed in this step (Figure 4) (Moldovan & D'Andrea 2009).

RAD51 is the central recombinase in mammalian cells that mediates homologous pairing of DNA sequences and strand exchange at the sites of DNA damage. In addition to RAD51, RAD51 paralogs, RAD51B, RAD51C, RAD51D, XRCC2 and XRCC3, have an essential role in the DNA repair reactions through HR. These paralogs act in transducing the DNA damage signal to effector kinases and in promoting break repair (Suwaki et al. 2011). RAD51 paralogs form two distinct complexes: the four-protein complex RAD51B/RAD51C/RAD51D/XRCC2 (BCDX2) and the two-protein complex RAD51C/XRCC3 (CX3) (Masson et al. 2001, Thacker 2005). RAD51C is thought to have more prominent cellular functions, as it is the only paralog found in both BCDX2 and CX3 complexes. In addition to HR, RAD51C has a role in checkpoint regulation by activating CHEK2 and in regulation of resolution of recombination intermediates (Somyajit et al. 2010).
Fig. 4. Simplified schematic representation of the key players of the FA protein complexes and FA-BRCA DNA damage response pathway. FANCM and FAAP24 recruit the FA core complex to DNA lesions. The core complex monoubiquitinates FANCD2 and FANCI, which then localize to DNA repair foci together with BRCA2, BRIP1 and PALB2. HR, the final stage in interstrand cross-link DNA repair, requires both RAD51 and RAD51C, and their interaction with BRCA1 and BRCA2. RAD51C contributes to the cell cycle arrest also by activating the checkpoint kinase 2 (CHEK2). Many of the genes that operate in the downstream part of this pathway (HR), such as BRCA1/2, BRIP1 and PALB2, are well-known breast/ovarian cancer predisposing genes. Not all protein interactions are shown, and proximity does not imply direct interaction. Modified from Levy-Lahad 2010. FA proteins are indicated with the complementation group letter (for example, A represents FANC-A); Ub, ubiquitin. 

Some of the HR repair pathway proteins are defected in FA. The main function of FA proteins is in the repair of extraordinarily deleterious lesions, interstrand crosslinks, and in maintaining genomic stability during DNA replication (Kottemann & Smogorzewska 2013). FANCM and FAAP24 have a role in recruiting the multisubunit FA core complex to DNA lesions in the early stages of the FA pathway. The core complex consists of eight proteins (FANCA/B/C/D/E/F/G/L/M) and has E3 ubiquitin ligase activity. The core complex monoubiquitinates FANCD2 and FANCI which then localize to DNA repair foci together with FANCD1 (BRCA2), FANCJ (BRIP1), and FANCN (PALB2). HR, the final stage in interstrand cross-link DNA repair, requires
RAD51 and RAD51C as well as interaction with BRCA1, BRCA2 and PALB2 (Levy-Lahad 2010, Moldovan & D'Andrea 2009).

In 2010, Vaz et al. found that biallelic mutations in the RAD51C gene cause a disorder reminiscent of FA (2010). Because the affected individuals in the family with the disease did not have hematological symptoms, which are classically observed in FA, they referred to the clinical phenotype as a Fanconi anemia-like disorder, with the provisional assignment FA-O. Nevertheless, the affected family members showed some of the severe congenital abnormalities, such as imperforate anus and cystic kidneys with renal failure, also described in BRCA2- and PALB2-defective FA patients. Simultaneously, heterozygous mutations in RAD51C were reported to associate with high risk of breast and ovarian cancer (Meindl et al. 2010). Consequently, RAD51C is the fourth gene (in addition to BRCA2, BRIP and PALB2) connecting the pathways involved in both FA and breast cancer susceptibility.

Six monoallelic dominant RAD51C mutations were identified in 480 German families displaying both breast and ovarian tumors. Mutations were not observed in families with breast cancer only (Meindl et al. 2010). As regards ovarian cancer occurrence the RAD51C families show major similarities with the families carrying BRCA1 and BRCA2 mutations, whereas families affected by mutations in the PALB2 and BRIP1 genes lack this feature (Erkko et al. 2007, Seal et al. 2006). Of the subsequent replication studies (Akbari et al. 2010, Pang et al. 2011, Romero et al. 2011, Silvestri et al. 2011, Wong et al. 2011, Zheng et al. 2010), only one (Akbari et al. 2010, Romero et al. 2011) was able to detect a single pathogenic mutation in one Spanish breast and ovarian cancer family. Accordingly, the mutations in the RAD51C might be much rarer than suggested by the initial report.

2.7 Genomic copy number variation and breast cancer susceptibility

In 2004, the study of genomic variability by array hybridization methods demonstrated the existence of copy number variants (CNVs) (Iafrate et al. 2004, Sebat et al. 2004), which are the most prevalent type of structural variation in the human genome. Outstanding evolution of microarray technologies has enabled high-resolution screening of the entire human genome for CNVs. These technologies include oligonucleotide array comparative genomic hybridization (CGH) and SNP genotyping arrays (Stankiewicz & Lupski 2010). CNVs have
been studied intensively in the last ten years and they have been catalogued in public databases such as the Database of Genomic Variants (DGV) (formerly known as Toronto Database of Genomic Variants) (Iafrate et al. 2004).

CNV breakpoint sequencing has revealed that microhomology of sequence at the breakpoints, non-allelic homologous recombination, and L1 retrotransposition are the most common mechanisms leading to the formation of CNVs (Conrad et al. 2010, Kidd et al. 2010). In addition, fork-stalling and template switching, microhomology-mediated break-induced replication (MMBIR) mechanisms and non-homologous end-joining (NHEJ) can all lead to CNVs (Weischenfeldt et al. 2013).

CNVs can mediate biological outcome by several molecular mechanisms (Figure 5). One of the most commonly recognized mechanisms is via altering the copy number of a gene (or genes) sensitive to a dosage effect (Lupski et al. 1992). CNVs can also disrupt the regulatory elements as far as ~1Mb away either upstream or downstream from the affected gene (Stankiewicz & Lupski 2010). Another molecular mechanism by which CNVs may alter gene function is the unmasking of recessive mutations or functional polymorphisms of the remaining allele when a deletion occurs (Kurotaki et al. 2005). In addition, CNVs can affect transvection (communication between alleles on homologous chromosomes) by deleting regulatory elements required for this process, or a break-point can simply disrupt a coding sequence (Lupski & Stankiewicz 2005).
Fig. 5. CNVs and their potential consequences to gene expression. A) Simplified representation of duplication and deletion. Individual 1 (normal) has two copies of the genetic element. As a result of duplication individual 2 has three copies of the element and as a result of a deletion individual 3 has only one copy of the element. B) Duplication or deletion of the whole gene can potentially lead to overexpression or haploinsufficiency of the gene product, respectively. If only a portion of a gene is affected by CNV it can lead to disruption of the coding region of the gene by the break point, which can lead to an aberrant protein product, altered splicing, or in the case of non-coding RNAs (ncRNA), imbalanced production. CNVs affecting extragenic regulatory elements can lead to altered gene expression regulation. Modified from Cabianca & Gabellini 2010.

Recent reports have suggested a role for genomic structural variants in susceptibility to various diseases, particularly neurodevelopmental disorders (Girirajan et al. 2011, Ionita-Laza et al. 2009, Lupski 2007a, Lupski 2007b, Sharp 2009, Stankiewicz & Lupski 2010, Wain et al. 2009, Weischenfeldt et al. 2013). In addition, germline CNVs of large DNA segments have recently been reported as factors that might have a role in predisposing individuals to different cancers,
including colorectal cancer, aggressive prostate cancer, neuroblastoma, pancreatic cancer and ovarian cancer.

The association of CNVs and colorectal cancer (CRC) have been investigated in two studies. A germline copy number loss in region at 3q26 that potentially harbors an element that regulates the expression of an upstream candidate tumor suppressor, \textit{PPM1L}, has been found to be associated with CRC (Thean \textit{et al.} 2010). In addition, Venkatachalam \textit{et al.} (2011) identified six protein and two microRNA coding loci by genome-wide copy number profiling to be potential novel candidate CRC susceptibility genes. A germline deletion at 2p24.3, which is close to the \textit{AY928977} and \textit{FAM84A} genes, but with unknown biological function, has been associated with aggressive prostate cancer (Liu \textit{et al.} 2009). A common germline CNV in 1q21.1 was identified to associate with the childhood cancer neuroblastoma. This CNV exists in different copy number states and leads to altered expression of \textit{NBPF23} (neuroblastoma breakpoint family, member 23) (Diskin \textit{et al.} 2009). In addition, several potential new candidates for susceptibility to pancreatic cancer have been identified by genome-wide CNV analysis (Lucito \textit{et al.} 2007). Also genomic profiles of CNV in \textit{BRCA1} carriers are shown to be qualitatively distinct from those in sporadic ovarian cancer patients (Yoshihara \textit{et al.} 2011).

Association of common germline CNVs with breast cancer risk has been ruled out by large case-control studies (Wellcome Trust Case Control Consortium \textit{et al.} 2010). However, Krepischi \textit{et al.} (2012b) have provided evidence of rare CNVs’ contribution to familial and early-onset breast cancer. Furthermore, large genomic rearrangements disrupting the \textit{BRCA1} and \textit{BRCA2} genes have also been observed in some families (Hogervorst \textit{et al.} 2003). Thus it is possible that rare CNVs might disrupt the function of other important genes that are associated with breast cancer predisposition.
3 Aims of the study

In search for predisposing genes that could explain an additional portion of familial breast cancer clustering in Finland, we set out to evaluate the presence of germline mutations in \textit{RNF8, MMS2, UBC13} and \textit{RAD51C} DNA damage response genes. Furthermore, the role of a relatively new class of genomic variation, copy number variation, in breast cancer susceptibility was assessed. The specific aims of this study were:

1. To perform the first mutation screening analysis of novel candidate genes \textit{RNF8, UBC13} and \textit{MMS2}, and to evaluate their potential involvement in breast cancer susceptibility.

2. To screen the suggested high-risk gene, \textit{RAD51C}, for conventional mutations, and to evaluate its potential involvement in breast and ovarian cancer susceptibility in the Finnish population.

3. To study whether rare copy number variants in Finnish breast cancer patients might have an impact on breast cancer predisposition.
4 Materials and methods

4.1 Cases and controls (I-III)

4.1.1 Familial breast cancer cases (I-III)

Patients from 123–147 breast cancer families from northern Finland were screened for germline mutations in the coding regions and exon-intron boundaries of the RNF8, MMS2, UBC13 and RAD51C candidate genes. Genome-wide scans of rare CNVs were performed for 103 of these families. In study I, fourteen of the studied index cases had previously been tested positive for known breast cancer-associated germline mutations in BRCA1 or BRCA2 (eleven) and PALB2 (three) whereas in studies II and III these cases were excluded from the analysis. The studied families were categorized as high- and moderate-risk families and the inclusion criteria and the number of families are summarized in Table 1.

Table 1. Inclusion criteria and a summary of the breast cancer families included in studies I, II and III.

<table>
<thead>
<tr>
<th>Risk</th>
<th>Study (no. of families)</th>
<th>Inclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>I (77)</td>
<td>1. Three or more cases of breast cancer and/or ovarian cancer in first- or second-degree relatives.</td>
</tr>
<tr>
<td></td>
<td>II (99)</td>
<td>2. Two cases of breast cancer in first- or second-degree relatives, of which at least one with early disease onset ($\leq$35 years), bilateral disease or multiple primary tumors.</td>
</tr>
<tr>
<td></td>
<td>III (73)</td>
<td></td>
</tr>
<tr>
<td>Moderate</td>
<td>I (46)</td>
<td>Two cases of breast cancer in first- or second-degree relatives.</td>
</tr>
<tr>
<td></td>
<td>II (48)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>III (30)</td>
<td></td>
</tr>
<tr>
<td>Total number of families</td>
<td>I (123)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>II (147)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>III (103)</td>
<td></td>
</tr>
</tbody>
</table>

4.1.2 Unselected young breast cancer cases (III)

The cohort of young breast cancer cases consisted of 75 Northern Finnish patients who were diagnosed with breast cancer at or under the age of 40. These breast cancer patients were not selected for or against family history of the disease, and tested negative for Finnish BRCA1, BRCA2 and PALB2 founder mutations. This
independent breast cancer cohort was collected as a validation group for the studied familial cases in study III based on the assumption that when a woman under the age of 40 years develops breast cancer, a hereditary predisposition may be suspected regardless of whether there is a family history or not (Brunet 2010).

4.1.3 Unselected breast and ovarian cancer cases (II)

For the \textit{RAD51C} c.-13\_14del27 and T287A genotyping as well as for the five tagSNPs (study II) samples from an unselected cohort of breast cancer patients (N = 993) were analyzed. This cohort consisted of 542 Northern Finnish breast cancer patients diagnosed at the Oulu University Hospital between the years 2000 and 2007, and of 451 prospective breast cancer patients from the province of Northern Savo in Eastern Finland, diagnosed at the Kuopio University Hospital between April 1990 and December 1995 (Hartikainen \textit{et al.} 2006).

The presence of the \textit{RAD51C} c.-13\_14del27 allele was also tested in an unselected ovarian cancer cohort of 332 patients diagnosed at the Oulu University Hospital. Furthermore, for the confirmation of \textit{RAD51C} involvement in ovarian cancer susceptibility, an unselected cohort of 232 ovarian cancer patients was comprehensively screened for mutations. The study set included 204 Swedish cases diagnosed at the Sahlgrenska University Hospital in Gothenburg, and 28 Finnish cases diagnosed at the Kuopio University Hospital.

4.1.4 Healthy controls (I-III)

The control samples were derived from anonymous female cancer-free Finnish Red Cross (FRC) blood donors (age ≥45 years) originating from Northern Finland. In the study I, 104–299 controls were used depending on the tested mutation. In study II, 871 control samples were used; of these 507 were FRC blood donors and 364 were control subjects from the province of Northern Savo in Eastern Finland, who were age- and area-of-residence-matched, selected from the National Population Register during the same time period as the Kuopio unselected breast cancer patient material. In study III, 128 FRC control samples were used.
4.2 DNA isolation (I-III)

Genomic DNA from peripheral blood lymphocytes was extracted using the standard phenol-chloroform method or the Puregene D-50K purification kit (Gentra).

4.3 Mutation detection methods

4.3.1 Conformation-sensitive gel electrophoresis (CSGE) (I)

Mutation screening of candidate genes in study I was partially performed by using CSGE on PCR products covering the exons and exon-intron boundaries. The assay is based on the assumption that mildly denaturing solvents can accentuate the conformational changes produced by mismatches in double-stranded DNA and hence increase the differential migration of heteroduplexes and homoduplexes (Ganguly et al. 1993, Körkkö et al. 1998).

4.3.2 High-resolution melting (HRM) (I, II)

In HRM, PCR product melting analysis is combined with real-time PCR. Melting of the double-stranded DNA can be detected in fluorescence changes by increasing the melting temperature. Wild-type and heterozygous samples can be differentiated in the melting plots (Er & Chang 2012, Wittwer 2009). CFX96 real-time thermal cycler was used in this assay for PCR amplification and mutation detection (Bio-Rad). Depending on the amplicon, the Ssofast EvaGreen Supermix (Bio-Rad) or Type-it HRM-kit (Qiagen) were used in real-time PCR reactions.

4.3.3 Sequencing (I, II)

Samples exhibiting deviating CSGE patterns or melting curves in HRM were sequenced to detect the mutation status. Additionally, some of the studied sequences were impossible to analyze with CSGE or HRM due to GC-richness or large size. These difficult sequences were studied by direct sequencing. Sequencing was performed with ABI3130xl (ABI Perkin Elmer) using the BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems).
4.3.4 MassARRAY analysis (II)

Investigation of the prevalence of two RAD51C sequence alterations, c.-13_14del27 and T287A (rs28363317), observed in genomic sequencing, as well as the genotyping of five tag-SNPs (rs12946522, rs304270, rs304283, rs17222691, and rs28363312) for the association analysis was performed in an unselected breast cancer patient cohort and in healthy female controls by using MassARRAY® mass spectrometer (Sequenom Inc.) and iPLEX® Gold (Sequenom, Inc.) on a 384-well plate format. TagSNPs were selected with the HapMap Genome Browser release 2. TagSNPs for region chr17:54118611–54173040 were picked out for the CEU population using the Tagger multimarker algorithm. MassARRAY was used for spectra acquisitions from the SperloCHIP. Data analysis and genotype calling were performed by using TyperAnalyzer Software version 4.0.3.18 (Sequenom, Inc.). Each 384-well plate contained at least eight non-template controls. For the c.-13_14del27 and for T287A, DNA samples from heterozygous mutation carriers were used as positive controls on each plate. For quality control, duplicate analysis was performed for 6.5% of the Oulu samples and for 6.7% of the Kuopio samples.

4.4 CNV discovery with microarray analysis (III)

CNV discovery for both the familial and young breast cancer cohort as well as for the healthy controls was carried out by using Illumina HumanOmni1-Quad BeadChips (Illumina Inc). These chips provide high-resolution coverage of the genome with over one million genetic markers, including those derived from the 1,000 Genomes Project and all three HapMap phases, and enables exact definition of the breakpoints. All samples included in the array had to go through the standard quality control measures, which included agarose gel runs to confirm the integrity of the DNA sample, and precise concentration determination with three-step dilution measurements. To control the confounding effects resulting from the sample handling and subsequent CNV analysis, all cases and controls were given new IDs and were blindly analyzed without knowing their disease status. All samples were analyzed following the Illumina-provided protocol in the same laboratory (Laboratory of Cancer Genetics, University of Oulu) with same arrays at the same period of time, with random places on the chip.

Analysis of the samples was performed with GenomeStudio Genotyping module (Illumina) and Nexus Copy Number Discovery Edition 5.1 software.
(BioDiscovery Inc.). Projects were generated in GenomeStudio, and samples having Call Rates over 98% were transported to Nexus where samples with quality score under 0.15 were passed on for further analysis. The analysis was restricted to CNVs called by two independent algorithms, GenomeStudio CNVpartition algorithm and The SNP-FASST2 segmentation algorithm of Nexus, to make sure that a high-quality CNV dataset was obtained. The sensitivity of detection in Nexus was examined by analyzing 11 samples containing known deletions/amplifications confirmed by independent methods, and all changes were detected under the parameters used. All identified CNVs had to be confirmed by Illumina cnvPartition 2.4.4 software, using a confidence level of over 50 in order to be included in the analysis: values of 50 or higher tend to display a region with high confidence. The breakpoints of the observed aberrations were defined using the information acquired from both Nexus and GenomeStudio, and CNVs that appeared to be artificially split by the algorithm were joined.

For the analysis, rare events were defined as those which were called by two independent algorithms and did not overlap more than 60% with the common CNVs in the CNV track defined in Nexus, based on the DGV. Nevertheless, as the DGV presents several known cancer susceptibility genes as containing polymorphic CNVs, each CNV not fulfilling the rare variant criteria was individually inspected before removal. This led to inclusion of “common” CNVs fulfilling the following criteria: 1) the CNV disrupts the involved gene partially, or deletes it entirely, 2) the affected gene is a known breast cancer predisposing gene, or based on its biological function, it is a highly likely breast cancer predisposing gene, and 3) biallelic defects in the involved gene lead to a rare genomic disorder, indicating that the defective allele is highly unlikely to be polymorphism. All “rare” events which were present at polymorphic frequencies in the pooled population of 250 cases and controls, apart from those that were specific or showed a clear enrichment in cancer cases, were excluded from further analyses.

4.4.1 CNV validation (III)

All rare CNV variants were validated by another separate method, either by Affymetrix Genome-Wide Human SNP Array 6.0 platform (Affymetrix) or quantitative real time PCR (qPCR). Affymetrix chip analysis was carried out following all the QC measures recommended by the protocol, and Affymetrix
CEL files were transported to Nexus for analysis with the SNP-FASST2 segmentation algorithm.

Confirmation with qPCR was done with CFX96 real-time thermal cycler (Biorad) using SsoFast EvaGreen Supermix (BioRad). The analyses of the samples with rare CNVs and at least 3 wild-type controls were performed in triplicates, and quantitation was done with CFX manager software (version 1.5). RAD50 and CTIP were used as reference genes.

4.5 Statistical analysis (I, II, III)

Statistical significance ($P$), odds ratio (OR), and confidence intervals (CI) of the observed differences in mutation and rare CNV frequencies between cancer patients and control individuals were evaluated by Pearson’s Chi-Square or Fisher’s exact test (I, II, III). The frequency of common CNVs and the size of duplications and deletions was monitored both in cases and controls and tested for differences with Mann-Whitney U-test (III). All the statistical tests were two-sided and considered to be statistically significant with a $P$-value ≤ 0.05. All analyses were carried out with PASW Statistics software (PASW Statistics 18.0 for Windows, SPSS Inc.) and this software was also used for the generation of odds ratios (OR) and confidence intervals (CI). For tagSNP data (II), the overall association as well as the Hardy-Weinberg equilibrium, allele specific $P$, OR, and CI were computed using Cochran-Armitage trend test (http://ihg.gsf.de/cgi-bin/hw/hwa1.pl).

4.6 Bioinformatical analysis

4.6.1 In Silico prediction for mutations pathogenicity (I, II)

All alterations were checked with NNSplice software for potential splicing effects (http://www.fruitfly.org/seq_tools/splice.html). All missense mutations were evaluated by using PolyPhen (http://genetics.bwh.harvard.edu/pph) and SIFT (http://sift.jcvi.org/). FPROM human promoter prediction software (Softberry, Inc.) and hidden Markov model (HMM)-based gene structure prediction software FGENESH (Softberry, Inc) were used to predict the effect of the RAD51C c.-13_14del27.
4.6.2 Network analysis and functional profiling of rare CNVs (III)

For pathway and biological function analysis of the genes disrupted by rare CNVs, Ingenuity Pathway Analysis (IPA) was used. The disrupted genes were defined as genes, including their promoter region, disturbed by the breakpoints or deleted entirely, and not shared between cases and controls. The list of disrupted genes was uploaded to IPA, which is an online exploratory tool with a curated database for over 20,000 mammalian genes and 1.9 million published literature references. IPA integrates transcriptomics data with mining techniques to predict and build up networks, pathways and biological function clusters. IPA is working together with several databases, including Entrez Gene, Gene Ontology and GWAS database, and the software maps the biological relationships of the uploaded genes according to published literature included in the Ingenuity database. The output results are presented as scores and $P$-values computed based on the numbers of uploaded genes in the cluster or network and the size of network or cluster in the Ingenuity knowledge database. Benjamini-Hochberg multiple testing correction $P$-values (to monitor the false discovery rate) were applied to determine the probability that each biological function or overrepresentation in diseases is due to change alone. Scores for IPA networks are the negative logarithm of the $P$-value, and they indicate the probability of the genes analyzed in a network for being found together due to random chance. Scores 2 or higher have at least a 99% probability of not being generated by chance alone.

4.7 Ethical issues

All of the biological specimens and clinical data of the familial breast cancer cases investigated were collected at the Oulu University Hospital, with the informed consent of the patients. Family members were only contacted with permission of the index case. Each DNA sample was labeled with a research code and handled anonymously. Results from the analyses were only used for research purposes and no information on the outcome of the mutation analyses was given to patients or family members.

Approval for the study was obtained from the Ethical Board of the Northern Ostrobothnia Health Care District and the Finnish Ministry of Social Affairs and Health. This study also included Swedish patients diagnosed at the Sahlgrenska University Hospital in Gothenburg and Finnish patients diagnosed at the Kuopio
University Hospital, and approval to perform the study was obtained from the Ethical Boards of the University Hospital Health Care Districts involved
5 Results

5.1 Mutation screening of the RNF8, UBC13 and MMS2 genes (I)

The RNF8, UBC13 and MMS2 genes were screened from index cases of 123 breast cancer families. Two exonic, two intronic and three 5’UTR variants were revealed by the mutation analysis of RNF8. Only one of these variants, the 5’ UTR variant c.1-36C>T, was unreported (not listed in the NCBI SNP database, http://www.ncbi.nlm.nih.gov/SNP/), and both of the identified exonic variants of RNF8, c.1344G>A and c.1377G>A were synonymous. In UBC13, one novel (c.421-18G>T) and one known (c.418+17C>T) intronic variant were observed. Interestingly, no sequence alterations were observed in MMS2. All observed variants in RNF8 and UBC13 were assessed in silico for possible effects on consensus splice sites, but none of them seemed to have effect on splicing. In order to evaluate possible pathogenicity of the observed changes, their frequencies were compared between familial cases and healthy female controls. None of the observed sequence variants appeared to associate with breast cancer susceptibility, however (Table 3 of study I).

5.2 Mutation screening of the RAD51C gene (II)

RAD51C mutations were screened from index cases of 112 breast and 35 breast–ovarian cancer families (Ntot = 147). Two sequence alterations were identified: one leading to a single amino acid change (T287A) and one to a 27-bp (c.-13_14del27) deletion in exon 1 (Table 2, Figure 6). These two alterations were also genotyped from an additional breast cancer cohort of 993 patients unselected for family history of cancer and from a healthy Finnish control cohort of 871 samples. The missense variant, T287A (rs28363317), was observed in two familial cancer index cases diagnosed with breast cancer, and was considered a benign alteration based on its frequency in healthy controls (2/860, \( P = 0.104 \)) and absence from the unselected breast cancer cohort (0/984).
Table 2. Observed sequence changes in the RAD51C gene in the Finnish familial breast cancer patients.

<table>
<thead>
<tr>
<th>Location</th>
<th>Nucleotide change</th>
<th>Effect on protein</th>
<th>Novelty</th>
<th>Mutation carrier frequency</th>
<th>P (OR; 95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Familial Br</td>
<td>Unselected Br</td>
</tr>
<tr>
<td>Ex1</td>
<td>c.-13_14del27</td>
<td>No transcript</td>
<td>Novel</td>
<td>1/147</td>
<td>0.147 (NA)</td>
</tr>
<tr>
<td>Ex6</td>
<td>c.859A&gt;G</td>
<td>T287A</td>
<td>rs28363317</td>
<td>2/147</td>
<td>0.104 (5.9; 0.8–42.0)</td>
</tr>
</tbody>
</table>

Br, breast cancer; NA, not available; OR, odds ratio; CI, confidence interval. 5Based on FPROM human promoter prediction software and HMM-based gene structure prediction software, FGENESH. 6Fisher’s Exact Test, familial Br cases vs. controls.
Fig. 6. RAD51C protein structure and predicted functional domains. RAD51C contains Walker ATP binding motifs, which have the ability to bind and hydrolyze ATP. BRC repeat interfaces (multimer BRC interface) enable possible oligomerization between individual RAD51C monomers. The observed sequence changes affecting the protein structure are indicated. The cohort from which each alteration was found is marked. BRCA, familial breast cancer cohort; OVCA, unselected ovarian cancer cohort; Putative NLS (nuclear localization signal).

The c.-13_14del27, which deletes 27 bp around the translation initiation site, was observed in the index case of one family (Figure 7). The human promoter prediction software FPROM predicts that this deletion transfers the transcription start site 95 nucleotides downstream compared to that of the wild-type allele. In addition, all sequences needed for reliable predictions for this exon made by the Hidden Markov model-based gene structure prediction software, FGENESH, are abolished. Thus the c.-13_14del27 leads to a null allele, and is pathogenic.
Fig. 7. Pedigree of breast cancer patient carrying the RAD51C c.-13_14del27 allele. Arrow points to the index patient. Black circles represent patients with breast or ovarian cancer; other cancer types are marked with gray. Genotype of the family members is shown if it has been possible to analyze. +, carrier; –, non-carrier of the RAD51C c.-13_14del27 allele; br, breast cancer; bone, bone cancer; es, esophagus cancer; leu, leukemia; ov, ovarian cancer; neck, neck tumor; pro, prostate cancer; sp, spinocellular carcinoma. If known, age at diagnosis or the time of monitoring is shown.

The index case with the RAD51C c.-13_14del27 allele was diagnosed with breast cancer at the age of 53 years and leukemia at the age of 54 years. Mother of the index case had breast cancer at the age of 69, and reminiscently, was also diagnosed with leukemia at the age of 76 years. This family showed very strong genetic predisposition to cancer, breast and ovarian in particular, but the segregation of the mutation with the disease phenotype could not be confirmed because of the lack of appropriate specimens from deceased individuals. No carriers of RAD51C c.-13_14del27 was observed in unselected breast cancer cohort and in control cohort. In addition, RAD51C c.-13_14del27 was screened in 332 ovarian cancer cases, but no additional carriers were identified.
Besides screening the exons and exon-intron boundaries of the entire gene for rare deleterious mutations, the association of common SNPs in $RAD51C$ with breast cancer susceptibility was also investigated. The allele distribution of five $RAD51C$ tagSNPs locating within the gene, (rs12946522, rs304270, rs304283, rs17222691, and rs28363312) was compared between 993 unselected breast cancer cases and 871 geographically matched unaffected controls. There was no evidence of a statistically significant association for any of these tagSNPs with breast cancer susceptibility. However, rs12946522 and rs17222691 were in strong linkage disequilibrium and suggested a borderline significant association with cancer risk (Table 3).

### Table 3. Association of five $RAD51C$ tagSNPs with breast cancer.

<table>
<thead>
<tr>
<th>TagSNP</th>
<th>Minor allele</th>
<th>MAF cases</th>
<th>MAF controls</th>
<th>$P$ value$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs12946522</td>
<td>G</td>
<td>0.26</td>
<td>0.23</td>
<td>0.071</td>
</tr>
<tr>
<td>rs304270</td>
<td>G</td>
<td>0.33</td>
<td>0.34</td>
<td>0.369</td>
</tr>
<tr>
<td>rs304283</td>
<td>T</td>
<td>0.27</td>
<td>0.26</td>
<td>0.944</td>
</tr>
<tr>
<td>rs17222691</td>
<td>T</td>
<td>0.28</td>
<td>0.25</td>
<td>0.078</td>
</tr>
<tr>
<td>rs28363312</td>
<td>A</td>
<td>0.09</td>
<td>0.09</td>
<td>0.969</td>
</tr>
</tbody>
</table>

$^{a}$Cochran-Armitage trend test, MAF, minor allele frequency.

An additional ovarian cancer cohort (N = 232) unselected for a family history of the disease collected from Finland (N = 28) and from Sweden (N = 204) was used to investigate the association of $RAD51C$ germline mutations with ovarian cancer. The ovarian cancer samples were comprehensively screened for $RAD51C$ mutations by using the HRM pre-screening method, which revealed seven alterations (Figure 6 and Table 4): five amino acid changes, one intronic variant, and one frameshift mutation. All sequence alterations were observed in Swedish ovarian cancer patients. Three of the observed changes were previously unreported and two of them, M10R and E191K, led to amino acid substitutions and both affected residues showed evolutional conservation among mammals and zebra fish. The M10R change was predicted to have a deleterious effect on the protein by SIFT and PolyPhen algorithms. E191K was observed in two index patients both showing a clear indication for hereditary predisposition to cancer. The other patient had been diagnosed with both breast and ovarian cancer at the age of 43 and 50 years, respectively, and the other was diagnosed with ovarian cancer at the age of 36 years. E191K was considered possibly damaging by PolyPhen. It is clear that in order, to reliably demonstrate the pathological nature
of the M10R and E191K alterations, functional analyses would be required. Until then both remain missense alterations of uncertain clinical significance. From the four changes previously observed by other studies, the c.774delT has a clearly deleterious effect on the protein as it leads to a frameshift and thus protein truncation. The carrier of c.774delT had been diagnosed with ovarian cancer at the age of 61 years. In addition, both of her parents and also one sibling had been diagnosed with cancer, but no additional details of the type of cancers or of the age at malignancy diagnosis were available from this family.

Table 4. Observed sequence changes in the RAD51C gene in unselected ovarian cancer patients.

<table>
<thead>
<tr>
<th>Location</th>
<th>Nucleotide change</th>
<th>Effect on protein</th>
<th>SIFT a</th>
<th>PolyPhen b</th>
<th>Novelty</th>
<th>N c</th>
<th>MAF d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ex1</td>
<td>c.29T&gt;G</td>
<td>M10R</td>
<td>Yes</td>
<td>Pb</td>
<td>Novel</td>
<td>1/232</td>
<td>-</td>
</tr>
<tr>
<td>Ex2</td>
<td>c.376G&gt;A</td>
<td>A126T</td>
<td>No</td>
<td>No</td>
<td>rs61758784</td>
<td>2/232</td>
<td>0.004</td>
</tr>
<tr>
<td>Ex3</td>
<td>c.571&gt;A</td>
<td>E191K</td>
<td>Yes</td>
<td>Ps</td>
<td>Novel</td>
<td>2/232</td>
<td>0.004</td>
</tr>
<tr>
<td>IVS3</td>
<td>c.571-17G&gt;T</td>
<td>Intronic d</td>
<td>-</td>
<td>-</td>
<td>rs193023469 d</td>
<td>2/232</td>
<td>0.004</td>
</tr>
<tr>
<td>Ex5</td>
<td>c.774delT</td>
<td>R258fs</td>
<td>-</td>
<td>-</td>
<td>1/232</td>
<td>0.004</td>
<td>-</td>
</tr>
<tr>
<td>Ex5</td>
<td>c.790G&gt;A</td>
<td>G264S</td>
<td>No</td>
<td>No</td>
<td>rs147241704 d</td>
<td>2/232</td>
<td>0.004</td>
</tr>
<tr>
<td>Ex6</td>
<td>c.859A&gt;G</td>
<td>T287A</td>
<td>No</td>
<td>Pb</td>
<td>rs28363317</td>
<td>4/232</td>
<td>0.009</td>
</tr>
</tbody>
</table>

Predictive algorithms used: SIFT and PolyPhen. aAll changes observed originating from Sweden (n = 204). bMinor allele frequency not displayed for singletons. cNo frequency data. dNo effect on splicing consensus sequences (NNSplice software), always observed together with c.376G>A. eSIFT: Yes, affects function; No, tolerated. fPolyPhen: Ps, possibly damaging; Pb, probably damaging; No, benign.

unreported at the time of publication.

5.3 Screening of rare CNVs (III)

Rare CNV discovery

103 familial breast cancer cases and 128 controls were scanned for structural variants by using high-resolution Illumina HumanOmni1-Quad BeadChips. Strict quality control criteria were applied to assure that ascertainment of CNVs was uniform between cases and controls. The frequencies of common CNVs did not significantly differ between the two groups (mean 9.7 CNVs for cases and 9.13 CNVs for controls). CNV that did not overlap more than 60% with the common CNVs in DGV were defined as rare CNVs. All the CNVs fulfilling the rare
variant criteria were confirmed by independent method. A total of 65 microdeletions and microduplications, ranging in size from 25 kb to 612 kb, were observed in the studied 231 subjects. There were 15 deletions (mean length 123 kb, median 61 kb) and 20 duplications (mean 216 kb, median 173 kb) in the cases. In controls 14 deletions (mean 146 kb, median 133 kb) and 16 duplications (mean 242 kb, median 186 kb) were observed.

The total number of rare CNVs was slightly higher in familial breast cancer cases than in controls. In addition, the number of familial cases was also higher when only considering those rare CNVs involving genes, and those directly disrupting genes. A similar trend was observed when analyzing the independent young breast cancer cohort of 75 patients (Table 5). The most profound difference was seen when considering only CNVs disrupting genes and restricting the analysis to variants not shared between cases and controls. The number of rare CNVs was almost twice in familial cases and 1.5-fold in young breast cancer cases compared to controls, but none of the differences were statistically significant. The functions and pathways of the disrupted genes were identified (Tables S1, S2 and S3 of study III) and were assessed by using the Ingenuity Pathway Analysis (IPA) classification system.
<table>
<thead>
<tr>
<th>Subjects</th>
<th>N</th>
<th>Observed rare CNVs</th>
<th>Involving genes</th>
<th>All disrupting genes</th>
<th>Disrupting genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fam BC cases</td>
<td>103</td>
<td>0.34 (35/103)</td>
<td>0.29 (30/103)</td>
<td>0.24 (25/103)</td>
<td>0.25 (26/103)</td>
</tr>
<tr>
<td>Young BC cases</td>
<td>75</td>
<td>0.32 (24/75)</td>
<td>0.24 (18/75)</td>
<td>0.23 (17/75)</td>
<td>0.23 (17/75)</td>
</tr>
<tr>
<td>Controls</td>
<td>128</td>
<td>0.23 (30/128)</td>
<td>0.21 (27/128)</td>
<td>0.16 (20/128)</td>
<td>0.16 (21/128)</td>
</tr>
</tbody>
</table>

*Observed only in cancer cases, or only in controls. The genomic loci has annotated genes. Gene disruptions include rare CNVs having breakpoints with the genes.*

BC, breast cancer; Fam, familial.
Disrupted genes in familial cases are involved in genomic integrity maintenance functions and diabetes

Mutations disrupting only part of the gene are likely to have biological consequences, and entirely deleted genes in the case of tumor suppressors follow the rationale of Knudson’s two-hit model or haploinsufficiency (Fodde & Smits 2002). Therefore, only those genes which were either disrupted by the breakpoints or deleted entirely were selected for analysis. Only a small number of the disrupted genes were part of known canonical pathways, and neither cases nor controls showed significant increase in any of them. Nevertheless, the genes disrupted in familial cases showed a significant overrepresentation in functions involving the maintenance of genomic integrity (Table 6). No particular functions were overrepresented among controls. Three of the genes disrupted in cases, BLM, RECQL4 and DCLRE1C, are directly involved in double-strand break (DSB) repair signaling. BLM is involved in BRCA1-mediated DNA damage response (Wang et al. 2000), RECQL4 participates in DNA replication and DSB repair (Abe et al. 2011), and DCLRE1C has a role in DSB repair by NHEJ (Ma et al. 2002).

Table 6. Molecular and cellular functions, and diseases and disorders overrepresented among the genes disrupted in familial breast cancer cases.

<table>
<thead>
<tr>
<th>Molecular and cellular functions</th>
<th>P-values*</th>
<th>Genes involved</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organization of chromosomes</td>
<td>0.0133</td>
<td>BLM, DCLRE1C</td>
</tr>
<tr>
<td>Maintenance of telomeres</td>
<td>0.0133</td>
<td>BLM, DCLRE1C</td>
</tr>
<tr>
<td>Repair of DNA</td>
<td>0.0178</td>
<td>RECQL4, BLM, DCLRE1C</td>
</tr>
<tr>
<td>Double-stranded DNA break repair</td>
<td>0.0211</td>
<td>BLM, DCLRE1C</td>
</tr>
<tr>
<td>Quantity of corpus luteum</td>
<td>0.00367</td>
<td>CASP3, ESR2</td>
</tr>
<tr>
<td>Diseases and disorders</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>0.000268</td>
<td>ACSL1,ANKS1B,ARHGAP39,BLM,CASP3,ESR2,KCNIP4, KHL1,MARCH6,MLF1IP,RBFOX1,STRN,SYNE2</td>
</tr>
</tbody>
</table>

*Statistically significant false discovery rate (FDR) adjusted P-values; correction for multiple testing was done using the Benjamini-Hochberg method.

Interestingly, both BLM and RECQL4 are RecQ family DNA helicases with a crucial role in the maintenance of genomic stability. Defects in BLM result in Bloom syndrome (Ellis et al. 1995), while defects in RECQL4 result in
Rothmund-Thompson syndrome (Kitao et al. 1999), both of which are recessive cancer predisposition syndromes. DCLRE1C encodes ARTEMIS, which is an essential factor for V(D)J recombination. Biallelic mutations of DCLRE1C result in severe combined immunodeficiency (SCID), in which lymphoma has been described (Moshous et al. 2003). Interestingly, the observed DCLRE1C allele is one of the most frequent mutations reported among SCID patients. This null allele comprises a gross deletion of exons 1–4 and the upstream MEIG1 gene. This deletion is known to arise from homologous recombination of DCLRE1C with the pseudo-DCLRE1C gene, located 61.2 kb upstream (Pannicke et al. 2010). BLM, RECQL4 and DCLRE1C are all attractive susceptibility genes according to their biological functions. However, to date clearly deleterious, breast cancer-related mutations have not been reported in any of them. It should be pointed out that another DNA repair gene, MCPH1, was found to be disrupted in one of the studied controls. MCPH1 has function in early stages of DNA damage response, and dysfunction in it leads to recessive primary microcephaly without any reported malignancies (Gavvovidis et al. 2010). The observed deletion removes exon 13 and is predicted to lead to out-of-frame translation of the last exon, number 14, thereby disrupting one of the three BRCT domains of MCPH1. The carrier of the MCPH1 deletion was still healthy at the age of 59 years. Thus, this mutation might not predispose to the disease.

The genes disrupted in familial cases were also highly overrepresented among genes connected to diabetes mellitus ($P = 0.000268$) (Table 6), which was mediated mainly through SNP associations observed in GWAS (Wellcome Trust Case Control Consortium 2007). The same phenomenon was also seen in the young breast cancer cohort ($P = 0.0246$), but not in controls. Interestingly, 6 genes of the 16 diabetes associated were under β-estradiol regulation.

**TP53 and β-estradiol centered networks are identified by network analysis in breast cancer cases**

The function of many genes is currently unknown and cannot be assigned to any predetermined pathways. Thus, an approach that is strictly focused on pathways is limited to identify biological relationships of gene clusters (Menashe et al. 2010). Consequently, we next used network analysis tools of IPA, which can map the biological relationships of the uploaded genes. Interestingly, analysis with familial cases revealed a network centered on TP53 and β-estradiol (score 29)
(Figure 8). To make things even more interesting, the same $TP53$ and $\beta$-estradiol centered network was identified when analyzing genes disrupted in the young breast cancer cohort (score 28) (Figure 9). When analyzing both case cohorts together, the network with the highest scores centered on $TP53$, $\beta$-estradiol and $CTNNB1$ (score 35) and the other around $\beta$-estradiol (score 31) (Figures 10 and 11, Table 7). Curiously, $CTNNB1$ encodes $\beta$-catenin, the oncogenic nuclear accumulation of which occurs in several malignancies, including breast cancer (Lin et al. 2000).

**Fig. 8.** $TP53$ and $\beta$-estradiol centered network in familial breast cancer cases. The centers of the network are colored black. The disrupted genes are colored gray.
Fig. 9. TP53 and β-estradiol centered network in young breast cancer cases. The centers of the network are colored black. The disrupted genes are colored gray.

Table 7. Genes disrupted or deleted entirely in breast cancer cases.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Aberration type</th>
<th>Involved exons*</th>
<th>Predicted consequence to transcriptb</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLM</td>
<td>disruption</td>
<td>promoter dup</td>
<td>unknown</td>
</tr>
<tr>
<td>EIF2C2</td>
<td>disruption</td>
<td>ex2-ex16, 3' UTR dup</td>
<td>unknown</td>
</tr>
<tr>
<td>HECW2</td>
<td>disruption</td>
<td>promoter, ex1 dup</td>
<td>unknown</td>
</tr>
<tr>
<td>RECQL4</td>
<td>deletion</td>
<td>entire gene</td>
<td>null allele</td>
</tr>
<tr>
<td>DAB2IP</td>
<td>disruption</td>
<td>promoter, ex1 del</td>
<td>null allele</td>
</tr>
<tr>
<td>LRRRC14</td>
<td>deletion</td>
<td>entire gene</td>
<td>null allele</td>
</tr>
<tr>
<td>ITGA9</td>
<td>disruption</td>
<td>ex19-ex23 del</td>
<td>in frame deletion</td>
</tr>
<tr>
<td>ACSL1</td>
<td>deletion</td>
<td>entire gene</td>
<td>null allele</td>
</tr>
<tr>
<td>KLHL1</td>
<td>disruption</td>
<td>promoter, ex1 del</td>
<td>null allele</td>
</tr>
<tr>
<td>ESR2</td>
<td>disruption</td>
<td>ex2-ex9, 3' UTR dup</td>
<td>unknown</td>
</tr>
<tr>
<td>ARHGAP39</td>
<td>disruption</td>
<td>ex12-ex13, 3' UTR del</td>
<td>premature termination</td>
</tr>
<tr>
<td>LRRFIP1</td>
<td>disruption</td>
<td>promoter, ex1 dup</td>
<td>unknown</td>
</tr>
<tr>
<td>CASP3</td>
<td>deletion</td>
<td>entire gene</td>
<td>null allele</td>
</tr>
<tr>
<td>TRPM3</td>
<td>disruption</td>
<td>promoter, ex1 dup</td>
<td>unknown</td>
</tr>
</tbody>
</table>
Neither the TP53 nor \(\beta\)-estradiol centered network was observed in controls, strongly in favor of the possibility that dysregulation of these networks is disease-related. Because p53 itself is a key regulator in preventing cells from malignancy, the TP53-centered network appears to have more obvious tumor-suppressive function. TP53 is frequently mutated during carcinogenesis in human malignancies, and germline lesions associate with the cancer-prone Li-Fraumeni syndrome (Malkin et al. 1990).

In the studied breast cancer cases, six genes (CASP3, BLM, EIF2C2, HECW2, RECQL4 and DAB2IP) disrupted by the observed rare CNVs were directly linked to TP53 (Figure 10). All of these genes encode proteins functioning in pathways with a potential role in malignancy prevention. RECQL4 and BLM were DNA damage response proteins and their network interactions were based on the repression of RECQL4 transcription by p53 (Sengupta et al. 2005), and the requirement of BLM for p53 localization to stalled replication forks (Sengupta et al. 2003). The rest of the interactions were based on direct binding of p53 with HECW2 (Miyazaki et al. 2003), DAB2IP and EIF2C2 (Lunardi et al. 2010). For CASP3, p53 has been shown to increase its activation (Cummings & Schnellmann 2002). The HECW2 disrupting allele was observed

<table>
<thead>
<tr>
<th>Gene</th>
<th>Aberration type</th>
<th>Involved exons</th>
<th>Predicted consequence to transcript</th>
</tr>
</thead>
<tbody>
<tr>
<td>KCNIP4</td>
<td>disruption</td>
<td>ex6-ex9, 3' UTR dup</td>
<td>unknown</td>
</tr>
<tr>
<td>DCLRE1C</td>
<td>disruption</td>
<td>promoter, ex1-ex4 del</td>
<td>null allele</td>
</tr>
<tr>
<td>TRAPPC9</td>
<td>disruption</td>
<td>promoter, ex1-ex3 dup</td>
<td>unknown</td>
</tr>
<tr>
<td>STRN</td>
<td>disruption</td>
<td>ex14-ex18, 3' UTR dup</td>
<td>unknown</td>
</tr>
<tr>
<td>PPP1R16A</td>
<td>disruption</td>
<td>ex6-ex10, 3' UTR del</td>
<td>premature termination</td>
</tr>
<tr>
<td>SYNE2</td>
<td>disruption</td>
<td>ex50-ex114, 3' UTR del</td>
<td>unknown</td>
</tr>
<tr>
<td>MARCH6</td>
<td>disruption</td>
<td>ex4-ex26, 3' UTR dup</td>
<td>unknown</td>
</tr>
<tr>
<td>GPT</td>
<td>deletion</td>
<td>entire gene</td>
<td>null allele</td>
</tr>
<tr>
<td>MLF11P</td>
<td>deletion</td>
<td>entire gene</td>
<td>null allele</td>
</tr>
<tr>
<td>SEMA4B</td>
<td>disruption</td>
<td>ex3-ex15, 3' UTR dup</td>
<td>unknown</td>
</tr>
<tr>
<td>RBF1OX1</td>
<td>disruption</td>
<td>ex11-ex13, 3' UTR dup</td>
<td>unknown</td>
</tr>
<tr>
<td>ANKS1B</td>
<td>disruption</td>
<td>promoter, ex1 dup</td>
<td>unknown</td>
</tr>
<tr>
<td>NXPH1</td>
<td>disruption</td>
<td>promoter, ex1-ex2 del</td>
<td>null allele</td>
</tr>
<tr>
<td>MEP1B</td>
<td>deletion</td>
<td>entire gene</td>
<td>null allele</td>
</tr>
</tbody>
</table>

Disruption, the gene is disrupted by the CNV breakpoints; deletion, the entire gene is deleted; del, partial gene deletion; dup, partial gene duplication. aBased on human genome assembly 19 (February 2009). Although detailed effects of partial gene duplication to gene transcription are not clear, duplication has potential to disrupt transcription by several mechanisms.
in two familial cases, whereas the other five p53-linked alleles were all singletons (Table S1 of study III).

The other network observed in both of the studied breast cancer case cohorts centered on β-estradiol (Figure 11), which is the primary biologically active form of estrogen. Exposure to both exogenous and endogenous estrogens is a well-established predisposing factor for breast cancer, and malfunctions in estrogen signaling and metabolism have a potential to affect this risk.

Estrogens mediate their physiological effects by binding to estrogen receptors, which then function as transcription factors and alter the expression of their target genes. Most of the estrogen actions are mediated by intracellular estrogen receptors ESR1 and ESR2 (Barros & Gustafsson 2011). The proliferation and differentiation of healthy breast epithelium is mediated by estrogens, but they also contribute to the progression of breast cancer by promoting the growth of transformed cells (Anderson 2002).

The network that centered on the β-estradiol consisted of several β-estradiol responsive genes: ANKS1B (Sauve et al. 2009), NXPH1, MEP1B (Thompson et al. 2002a), CASP3 (Childs et al. 2010) and ACSL1 (Thompson et al. 2002a). Interestingly, when separately tested in IPA none of the genes disrupted in controls were found to be under β-estradiol regulation. ESR2, STRN and ANKS1B of the β-estradiol network exhibited recurrent disrupting alleles among cancer cases (Table S1 of study III), emphasizing their potential role in breast cancer predisposition.
Fig. 10. TP53, β-estradiol and β–catenin centered network. IPA was used to analyze the connection between the genes disrupted in breast cancer cases (familial and young breast cancer cohort). The centers of the network are colored black. The disrupted genes are colored gray.
Fig. 11. β-estradiol centred network. The center of the network is colored black. The disrupted genes are colored gray.
Discussion

6.1 Pathogenic mutations in RNF8, UBC13 and MMS2 are rare

The maintenance of genomic integrity and cell-cycle checkpoint control is dependent on functions of RNF8, UBC13 and MMS2 (Al-Hakim et al. 2010, Andersen et al. 2005, Huen et al. 2007, Kolas et al. 2007), and previous studies have suggested that the RNF8 gene could be a novel tumor suppressor (Li et al. 2010). Curiously, another E3 ligase, RNF168, which acts together with UBC13 to amplify the RNF8-dependent histone ubiquitination, has been shown to be defected in RIDDLE syndrome, which is an immunodeficiency and radiosensitivity disorder. However, it is still unclear whether RIDDLE syndrome is associated with genome instability or increased tumor incidence, which are common features for most of the cancer syndromes caused by mutations in the genes that are also associated with breast cancer risk like Li-Fraumeni (TP53), Cowden (PTEN) and Peuz-Jegher syndromes (STK11) (Stewart et al. 2009).

Based on their role in DDR and importance in BRCA1-mediated DNA damage recognition it was considered possible that mutations in the RNF8, UBC13 and MMS2 genes might contribute to hereditary predisposition to breast cancer. The whole coding region of these genes was systematically screened for mutations in 123 breast cancer families in the current study. No clearly deleterious sequence alterations were observed in any of the studied genes. Thus, the protein products of these genes seem to be essential for the DNA damage response and other functions maintaining genomic integrity. Interestingly, mutation screenings have been performed for the other players in the BRCA1-dependent ubiquitin-mediated DNA damage recognition pathway, such as ABRAXAS, MERIT40 and RAP80, and some of these small studies have failed to identify any clear pathogenic mutations (Akbari et al. 2009, Novak et al. 2009, Osorio et al. 2009, Solyom et al. 2010), while others have found rare predisposing alleles (Nikkilä et al. 2009, Solyom et al. 2012). Therefore, it might still be possible that some rare mutations exist in RNF8, MMS2 and UBC13 which are not seen in the current study, and larger patient cohort would be needed for their identification. However, based on our findings rare mutations in these genes are unlikely to make any sizeable contribution to cancer predisposition. This is the first study reporting the mutation screening of the RNF8, UBC13 and
MM2 genes in familial breast cancer cases and according to literature, no additional mutation screening studies of these genes have been published.

6.2 Mutation screening of RAD51C identifies two rare pathogenic deletions

RAD51C was recently identified as a FA-like disorder and breast cancer predisposing factor (Meindl et al. 2010, Vaz et al. 2010). In the current study, we identified RAD51C c.-13_14del27 from one Finnish breast and ovarian cancer family and RAD51C c.774delT from a Swedish ovarian cancer case. These deletions are clearly deleterious and thus provide additional evidence that defects in RAD51C are implicated in cancer predisposition.

RAD51C c.-13_14del27 was identified in one of the 147 familial breast cancer index cases. The family of the index case had a history of both breast and ovarian cancer. Interestingly, in the initial report by Meindl et al. (2010) all the deleterious mutations were also identified in breast-ovarian cancer families, which markedly resembles the clinical presentations of people carrying BRCA1 and BRCA2 mutations (King et al. 2003, Turnbull & Rahman 2008). The RAD51C c.-13_14del27 family also had exceptionally high prevalence of leukemia cases, but co-segregation of the mutation with these could not be confirmed. The index had leukemia one year after breast cancer diagnosis, suggesting a possible treatment association. Association with leukemia and treatment-associated leukemia has previously been suggested also for BRCA1 and BRCA2 mutations (Shih et al. 2000). The frequency of RAD51C c.-13_14del27 mutation was also tested in 990 geographically matched breast and 332 ovarian cancer patients unselected for a family history of the disease, but additional carriers were not observed. Thus, RAD51C c.-13_14del27 appears to be very rare, or even a private mutation, rather than a common Finnish founder alteration that would explain a more significant fraction of breast or ovarian cancer susceptibility in our population. From the 147 families studied, a subset of 35 families displayed both breast and ovarian cancer, and the frequency of RAD51C mutations in them (1/35, 2.9%) was comparable to that reported previously (6/480, 1.25%) (Meindl et al. 2010). In the families with breast cancer only, no RAD51C mutations were observed, which highlights the connection between RAD51C defects and predisposition to ovarian cancer.
In addition, the identification of another deleterious alteration, RAD51C c.774delT in the unselected ovarian cancer cohort (N = 232) gives strength for this connection. The RAD51C c.774delT mutation carrier originates from Sweden. Romero et al. (Romero et al. 2011) observed independently the same RAD51C c.774delT deletion in a Spanish breast and ovarian cancer family. Interestingly, the Spanish mutation positive breast cancer patient inherited the c.774delT allele from her Swedish mother, who was still healthy at the age of 70 years. Because of the lack of suitable samples, segregation of the RAD51C c.774delT with the cancer phenotype could not be analyzed in the Spanish family. Nevertheless, this data strongly suggests that c.774delT is a recurrent Swedish mutation, and should be studied further in this population.

The exciting initial finding by Meindl et al. (2010) has motivated many research groups to perform replication studies in different populations for the confirmation of the RAD51C germline mutations’ association with breast-and ovarian cancer susceptibility. Nevertheless, many of these studies have failed to detect clearly deleterious mutations in RAD51C (Akbari et al. 2010, Clague et al. 2011, De Leeneer et al. 2012, Kushnir et al. 2012, Lu et al. 2012, Pang et al. 2011, Silvestri et al. 2011, Wong et al. 2011, Zheng et al. 2010), whereas others have reported highly penetrant mutations, although at lower frequency than in the initial study (Coulet et al. 2013, Loveday et al. 2012, Osorio et al. 2012, Pelttari et al. 2011, Romero et al. 2011, Thompson et al. 2012a). Interestingly, two recurrent deleterious RAD51C mutations, c.93delG and c.83711G>A, were identified to be novel moderate-to-high penetrance ovarian cancer susceptibility alleles in Finnish population (Pelttari et al. 2011). The frequency of these two Finnish founder mutations was very rare in the general population (2/1279, 0.2%). They were observed predominantly in ovarian cancer cases (6/417, 1.4%) and only in two breast cancer families with family history of ovarian cancer (2/1553, 0.1%).

A majority of the variants described as pathogenic in the replication studies affect splicing, causing protein truncations by frameshifts or by creation premature stop codons. However, most of the observed variations in these studies are missense mutations leading to amino acid substitutions and only few of them have been investigated in depth by functional assays. Lack of functional assays on missense mutations can lead to underestimation of the relevance of this mutation class. Only three studies so far have studied the functional consequence of RAD51C missense mutations (Clague et al. 2011, Meindl et al. 2010, Osorio et al. 2012). Thus, the role of the RAD51C missense mutations in breast and ovarian cancer susceptibility is yet to be fully understood.
cancer risk is still unclear. Although the frequency of individual harmful missense mutation can be rare, their collective frequency can be high and thus they might have a significant contribution to breast and ovarian cancer risk.

In addition to the two clearly harmful deletions identified in the current study, a total of five amino acid alterations from the Swedish unselected ovarian cancer cohort were observed: M10R, A126T, E191N, G264S and T287A, of which T287A was also observed in the Finnish familial cohort. The A126T missense mutation has previously been studied by functional assays and has proved not to be pathogenic. Functional analysis of G264S and T287A mutations has demonstrated that mutant carrying cells were more susceptible to the cytotoxic effects of mitomycin C when compared with the wild-type cells (Meindl et al. 2010). In addition, yeast two-hybrid analysis has shown that T287A decreases slightly but statistically significantly RAD51C interaction with XRCC3 (Clague et al. 2011). G264S has been implicated to have increased frequency also in other ovarian cancer cohorts and breast/ovarian cancer families (Meindl et al. 2010, Pelttari et al. 2011, Thompson et al. 2012a). Interestingly, predictive algorithms could not predict the reduced survival of G264S in complementation test with ΔRad51c DT40 cells (Meindl et al. 2010), indicating that many other observed missense mutations should be tested in vitro, even if in silico prediction implicates no harmful effect. M10R and E191N are novel alterations and both of these missense mutations may have potential clinical relevance, as they target evolutionally conserved residues and are predicted to have damaging effects on the protein. So far, M10R and E191K mutations have not been observed in any other replication studies. It is also important to notice that PCR-based screening methods used in our and other studies cannot detect large genomic rearrangements and the papers published so far cannot rule out their existence in the RAD51C gene as predisposing factors for breast and ovarian cancer.

Interestingly, Hansmann et al. (2012) identified RAD51C to be hypermethylated in 3 of >600 (0.5%) familial cases breast cancer and/or ovarian cancer patients. They speculated that hypermethylation of RAD51C together with hypermethylation of BRCA1 provides a causal explanation for 1-2% of the cases in their cohort. Furthermore, they suggest that screening of epimutations in addition to sequence alterations should be included in mutation screening of RAD51C and other candidate genes. In addition, RAD51C was also identified as a new candidate for susceptibility loci for testicular germ cell tumor in large GWAS (Chung et al. 2013). Thus, evidence from many directions suggests that RAD51C
is an important predisposing factor for cancer development. Indeed, identification of *RAD51C* as breast and ovarian cancer susceptibility gene supports the notion that virtually all breast cancer susceptibility genes identified to date are related to fundamental processes in cellular proliferation and DNA integrity control, and particularly involved in HR pathway. Excitingly, studying of other *RAD51* paralogs has also highlighted the importance of these DNA repair factors in cancer predisposition. *RAD51D* mutations have been associated with ovarian cancer (Loveday *et al.* 2011) and the locus where *RAD51B* is located has been associated with breast cancer susceptibility in large GWAS analysis (Michailidou *et al.* 2013).

Overall, our results support the hypothesis that *RAD51C* is predominantly an ovarian cancer susceptibility gene, while the risk of breast cancer is not clearly elevated. Our *RAD51C* mutation analysis together with numerous previous studies support the idea that rare gene variants, with either moderate or potentially even high disease penetrance, are an important component of familial cancer.

**6.3 Estrogen signaling and TP53 tumor suppression network is disrupted by rare CNVs in breast cancer patients**

The analysis of the data from our high-resolution genome-wide scans for structural variants provide evidence that rare CNVs contribute to breast cancer susceptibility. The studied breast cancer cases showed a slight but consistent increase in the frequency of rare CNVs compared to controls. The difference was not as fundamental as seen in psychiatric disorder studies where the observed changes can have very clear and severe effects on patients’ phenotype, typically involving large genomic regions and numerous genes, and often arise as *de novo* mutations (International Schizophrenia Consortium 2008, Walsh *et al.* 2008). Nevertheless, in the current study the biological networks affected by the disrupted genes differed clearly between breast cancer cases and controls, supporting their role in cancer predisposition.

In theory, some of the observed CNVs in the young breast cancer cohort could have been generated by *de novo* events; in order to totally exclude this possibility, segregation analysis of the CNVs would be needed. However, by studying case-parent trios Stadler *et al.* (2012) found that *de novo* CNVs are not present in early-onset sporadic breast cancer cases.

Of the annotated biological functions, the maintenance of genomic integrity was significantly overrepresented among the genes disrupted in familial cases.
This included DSB repair, which is in line with the prevailing paradigm that defects in DSB pathway contribute to breast cancer predisposition (Stratton & Rahman 2008). The three DSB pathway genes, BLM, RECQL4 and DCLRE1C, disrupted in the familial cases all represent attractive breast cancer susceptibility genes. In addition, IPA analysis indicated that the genes disrupted by rare CNVs in the studied breast cancer cases formed a network around TP53 and β-estradiol. The existence of these networks was confirmed in both independent breast cancer cohorts. Both networks are coherent and provide a biologically meaningful explanation for the breast cancer risk. Furthermore, their identification through the genome-wide approach used provides strong evidence for a role in breast cancer predisposition and players in these networks should also be considered as relevant candidates in the future studies.

TP53 network genes encode proteins that have functions in pathways with potential role in malignancy prevention. These include HECW2 that is involved in DNA damage response and apoptosis (Miyazaki et al. 2003). In addition, EIF2C2, also known as AGO2, encodes an Argonaute family protein that has a role in RNA interference (Liu et al. 2004). DAB2IP and CASP3 were particularly interesting genes in the TP53 network. DAB2IP is a member of the Ras GTPase-activating gene family and has been indicated to act as a tumor suppressor. Promoter methylation mediated inactivation of DAB2IP occurs in several malignancies, including prostate and breast cancer (Dote et al. 2004). In addition, DAB2IP has been shown to modulate epithelial-to-mesenchymal transition and prostate-cancer metastasis (Xie et al. 2010). CASP3 encodes caspase 3, which is a member of a highly conserved caspase protease family. Caspases are key intermediaries of the apoptotic process. A failure in this process can lead to cancer (Hengartner 2000). Various molecular epidemiological studies have suggested that germline SNPs in caspases may contribute to cancer risk, and a common missense variant in caspase 8 has been associated with breast cancer susceptibility (Cox et al. 2007, Hengartner 2000). Interestingly, apoptosis is also one of the many genomic integrity maintenance functions of BRCA1. Caspase 3 has been indicated to mediate the cleavage of BRCA1 during UV-induced apoptosis, and the cleaved C-terminal fragment triggers the apoptotic response through activation of downstream effectors (Zhan et al. 2002). The deletion disrupting the promoter and exon 1 of DAB2IP is predicted to result in null allele. Furthermore, the deletion abolishing the whole CASP3 gene clearly leads to haploid genotype (Table 7). The Knudson two-hit model or haploinsufficiency could explain the
predisposition to the disease in carriers of DAB2IP and CASP3 deletions. All
TP53 network members represent attractive susceptibility genes, which could also
harbor other cancer predisposing mutations, therefore being excellent candidates
for re-sequencing studies.

Multiple lines of evidence support the profound role of estrogen in breast
cancer development, and disruptions in estrogen signaling and metabolism have
long been considered to affect breast cancer risk. Most of the genes in the
estrogen network were under β-estradiol regulation, but two of the disrupted
genes, ESR2 and STRN, had a more straightforward role in estrogen signaling.
ESR2 encodes the estrogen receptor β, which is one of the principal mediators of
estrogen signaling within the cell (Barros & Gustafsson 2011). It binds estrogens
with a similar affinity as estrogen receptor α, and promotes expression of estrogen
response element containing genes (Ogawa et al. 1998). Curiously, ESR2 has
previously been suggested to harbor common breast cancer predisposing variants
(Gold et al. 2004, Maguire et al. 2005). In addition, ESR2 variation has been
suggested to influence the development of breast cancer also by in vitro studies
(Speirs & Kerin 2000). In contrast, STRN encodes striatin, which acts as
molecular scaffold in non-genomic estrogen-mediated signaling (Lu et al. 2004).
Striatin physically interacts with calmodulin 1 which is a transducer of Ca^{2+}
signals (Castets et al. 1996). Furthermore, striatin interacts with estrogen receptor
α and forms a complex with protein phosphatase 2A, which also regulates the
function of estrogen receptor α (Lu et al. 2003, Lu et al. 2004).

The role of estrogen in breast cancer predisposition was further emphasized
by the identification of a recurrent deletion allele in CYP2C19, encoding an
enzyme (cytochrome P450, family 2, subfamily C, polypeptide 19) involved in
estrogen metabolism (Cheng et al. 2001, Cribb et al. 2006). This deletion in
CYP2C19 was observed with an increased frequency in familial cases (Table S2
of study III). Interestingly, the CYP2C19*17 allele, which defines an ultra-rapid
metabolizer phenotype, has previously been associated with a decreased risk for
breast cancer. The increased catabolism of estrogens by CYP2C19 may lead to
decreased estrogen levels and therefore to reduced breast cancer risk (Justenhoven
et al. 2009). Similarly, decreased activity of CYP2C19 through haploinsufficiency
might potentially increase the risk of breast cancer. Interestingly, based on their
function both ESR2 (Gold et al. 2004, Maguire et al. 2005) and CYP2C19
(Justenhoven et al. 2009) have long been considered good candidate genes for
breast cancer susceptibility. Nevertheless, structural variation has not previously
been reported in either of them, but it is possible that CNVs might represent a

79
new class of cancer predisposing variation in both of these genes. Functionally relevant structural variants might also be present in other cytochrome P450 genes that locate in gene clusters, such as CYP2C19 (Gray et al. 1995). It is known that clustering of similar genes increases the potential for unequal crossing-over between sister chromatids and thus for creation of CNV alleles.

As already mentioned, estrogens are essential for the proliferation and differentiation of healthy breast epithelium. Estrogens may also promote the growth of malignant cells. It is known that estradiol exposure and the lifetime risk for breast cancer development are closely related. Menarche at young age and an older age at menopause cause an increased window of lifetime estradiol exposure and are associated with an increased risk of developing breast cancer. In contrast, menarche at older age and a younger age at menopause decrease a woman's lifetime estradiol exposure and are thus associated with a decreased risk of developing breast cancer. Furthermore, chemicals known as endocrine disruptors, like bisphenol-A which is used in production of plastic and mimics physiological estrogen, are speculated to promote breast carcinogenesis by increasing the estrogen signaling (Jenkins et al. 2012). Thus, the genes that operate under estrogen regulation or are related in the estrogen signaling pathways are plausible candidates for breast cancer predisposition genes.

The genes damaged in both studied breast cancer cohorts were also significantly overrepresented among genes connected to diabetes mellitus. This unexpected observation could be likely explained by shared risk factors predisposing to both breast cancer and diabetes. Curiously, several non-genetic risk factors, including obesity and a sedentary lifestyle, have already been reported to be shared between these two diseases. Several hormonal systems are altered in diabetes, including insulin, insulin-like growth factors, and other growth factors as well as estrogen. These altered systems may also affect the development of breast cancer (Larsson et al. 2007, Xue & Michels 2007). The results of the current study suggest estrogen being the key link in the association between diabetes and breast cancer, as over one third of the diabetes-associated genes in the two studied breast cancer cohorts were part of the β-estradiol network.

In conclusion, the current study suggests that rare CNVs should be recognized as an alternative source of genetic variation influencing breast cancer risk. This observation is further supported by a recent study which also provided evidence for rare CNVs’ contribution to familial and early-onset breast cancer
(Krepischi et al. 2012a). Furthermore, a recent study identified germline CNVs as a potential prognostic factor for disease recurrence in the early-stage non-metastatic breast cancer (Sapkota et al. 2013). Other studies have also found that germline CNVs are risk factors for developing cancer. So far neuroblastoma (Diskin et al. 2009), colorectal cancer (Thean et al. 2010, Venkatachalam et al. 2011), hepatocellular carcinoma (Clifford et al. 2010), aggressive prostate cancer (Liu et al. 2009), nasopharyngeal carcinoma (Tse et al. 2011) and BRCA1-associated ovarian cancer (White et al. 2007) have been associated with germline CNVs.

The network analysis of the current study with two independent breast cancer cohorts provides strong evidence for the role of estrogen-mediated signaling in breast cancer predisposition and reinforces the concept of TP53 centered tumor suppression in the prevention of malignancy. The variety of disrupted genes belonging to these networks highlights that diverse mechanisms and genetic heterogeneity are relevant to breast cancer pathogenesis.
7 Concluding remarks

In the study of the involvement of putative candidate genes *RNF8*, *UBC13*, *MMS2* and *RAD51C*, as well as rare CNVs, in predisposition to breast cancer, the major observations and conclusions were as follows:

1. Germline mutations in *RNF8*, *UBC13* and *MMS2* genes do not make any major contribution to familial breast cancer, at least not in the studied population.

2. Identification of two rare deleterious *RAD51C* alleles, c.-13_14del27 and c.774delT, in the studied cohorts support the idea that *RAD51C* is predominantly an ovarian cancer susceptibility gene.

3. Genes disrupted by rare CNVs in familial cases showed enrichment in genomic integrity maintenance functions (*P* = 0.0211). Three of the disrupted genes in familial cases, *BLM*, *RECQL4* and *DCLRE1C* are directly involved in DSB repair signaling, which underlines the essential role of this pathway in breast cancer predisposition.

4. The genes disrupted by rare CNVs in familial cases were also highly overrepresented among genes linked to diabetes mellitus (*P* = 0.000268). The same phenomenon was seen in the young breast cancer cohort (*P* = 0.0246), but not in controls. This result probably represents shared genetic risk factors predisposing to both breast cancer and diabetes.

5. Network analysis of genes disrupted by rare CNVs revealed *TP53* and β-estradiol centered networks in breast cancer cases. Neither of these networks were observed in controls, strongly arguing in favor of the possibility that dysregulation of these networks is disease related. All the disrupted genes included in these networks represent plausible candidate genes for breast cancer susceptibility, possibly harboring additional predisposing germline mutations.

6. The results from the genome-wide CNV screening suggest that rare CNVs represent an alternative source of genetic variation influencing hereditary risk for breast cancer.
8 Future prospects

Overall, the results of the present study have to some extent elucidated the background of familial breast cancer and created novel research paths to be pursued. However, much more research is needed before all the genetic factors predisposing to breast cancer are defined. This goal might even be unreachable because de novo mutations arise constantly in the germline. Nevertheless, it should be possible to describe at least all the founder mutations of different populations and ethnic groups.

This study found that RAD51C is predominantly an ovarian cancer susceptibility gene. Other studies must be conducted in larger series to evaluate whether RAD51C confers predisposition to ovarian cancer alone or to both breast and ovarian cancer, which remains controversial. Currently, however, it seems evident that RAD51C mutations are only rarely observed in families having only breast cancer. In addition, our finding suggests that other genes involved in the HR-repair pathway, as well as other RAD51 paralogs, might represent excellent candidate genes for re-sequencing studies.

The identified rare CNVs should be studied further in larger data sets to evaluate their frequency in different populations. In addition, the segregation of them should be analyzed in the families of the index cases. All the CNV-disrupted genes that associated with breast cancer in this study are good candidates for re-sequencing studies. These genes could for instance be investigated in the future by exome or targeted-sequencing.

The population of Northern Finland has for a long time been relatively isolated and therefore constitutes a good study material for genetic analyses. Modern whole-genome and exome sequencing techniques have recently revolutionized the studies of disease genetics and biology. At the same time the prices for utilizing these comprehensive and very accurate high-throughput technologies have significantly dropped. Consequently, these advanced research methods should now be utilized also for studies of Northern Finnish familial breast cancer cohort in further detail. This next-generation sequencing can also be targeted to specific genomic regions or biological pathways of interest. For instance, in targeted-sequencing, a panel of certain pre-selected genes is used (Ruark et al. 2013). There are at least two distinct strategies for how these methods could be applied. First, analysis of multi-case breast cancer families can reveal the disease predisposing genetic factors in those particular families (Gracia-Aznarez et al. 2013, Thompson et al. 2012b). Second, investigating large
cohorts of breast cancer patients by utilizing a case-control approach can reveal novel predisposing factors associated with breast cancer in that particular cohort (Ruark et al. 2013). Besides all the enthusiasm related to these new technologies, it should be mentioned that there are considerable challenges with respect to study design, data analysis and study replication with the next-generation sequencing approaches, and particularly concerning genetic founder populations (Cirulli & Goldstein 2010).
References


95


Wellcome Trust Sanger Institute The Cancer Genome Project. URI: http://www.sanger.ac.uk/genetics/CGP/.


Original publications


Reprinted with the permission from Springer Science + Business Media (II).

Original publications are not included in the electronic version of the dissertation.
1206. Penttilä, Matti (2013) Duration of untreated psychosis: association with clinical and social outcomes and brain morphology in schizophrenia


1209. Hietikko, Elina (2013) Genetic and clinical features of familial Meniere’s disease in Northern Ostrobothnia and Kainuu


1212. Tikkanen, Jani (2013) Early repolarization in the inferolateral leads of the electrocardiogram: prevalence, prognosis and characteristics


1214. Kaakinen, Pirjo (2013) Pitkäaikaissairaiden aikuisten ohjauksen laatu sairaalassa

1215. Pasanen, Anna Kaisa (2013) A translational study on the roles of redox molecules, cell cycle regulators and chemokine receptors as prognostic factors in diffuse large B-cell lymphoma

1216. Malo, Elina (2013) The role of low birth weight and resistin in metabolic syndrome


1219. Koskela, Sanna (2013) Granulosa cell anti-Müllerian hormone secretion in ovarian development and disease

1220. Soini, Heidi (2013) Mitochondrial DNA sequence variation in Finnish patients with maternally inherited type 2 diabetes, epilepsy and mitochondrial disease: risk and novel mutations

Book orders:
Granum: Virtual book store
http://granum.uta.fi/granum/
Mikko Vuorela

ROLE OF THE RNF8, UBC13, MMS2 AND RAD51C DNA DAMAGE RESPONSE GENES AND RARE COPY NUMBER VARIANTS IN HEREDITARY PREDISPOSITION TO BREAST CANCER