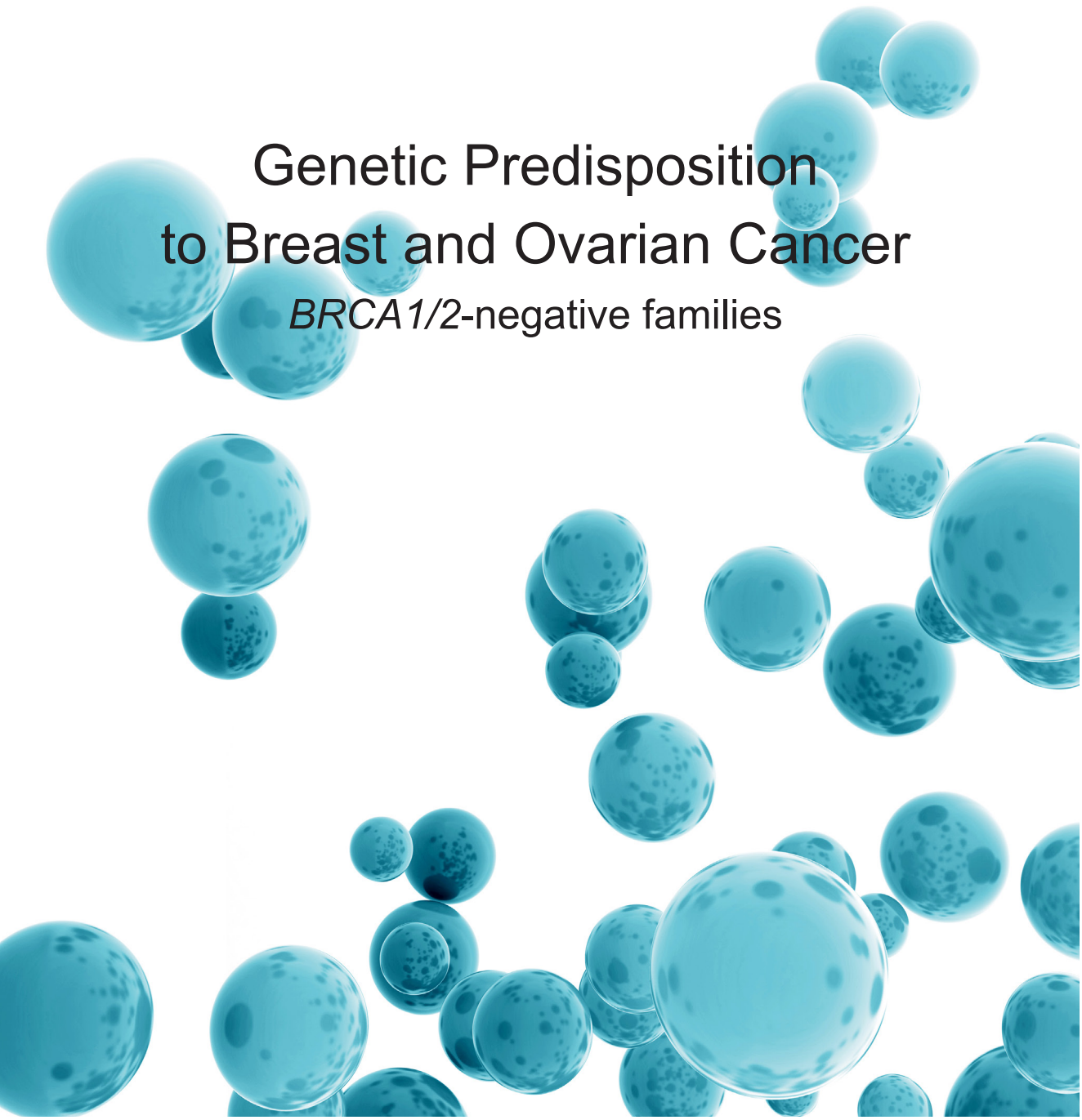KIRSI MÄÄTTÄ

Genetic Predisposition
to Breast and Ovarian Cancer
BRCA1/2-negative families

KIRSI MÄÄTTÄ

Genetic Predisposition
to Breast and Ovarian Cancer

BRCA1/2-negative families

■

UNIVERSITY OF TAMPERE

KIRSI MÄÄTTÄ

Genetic Predisposition
to Breast and Ovarian Cancer

BRCA1/2-negative families

UNIVERSITY
OF TAMPERE

The originality of this thesis has been checked using the Turnitin OriginalityCheck service in accordance with the quality management system of the University of Tampere.

Cover design by
Mikko Reinikka

CONTENTS

LIST OF ORIGINAL COMMUNICATIONS

This thesis is based on the following communications, which are referred by the corresponding Roman numerals.

I **Kuusisto KM**, Bebel A, Vihinen M, Schleutker J, Sallinen SL. Screening for *BRCA1, BRCA2, CHEK2, PALB2, BRIP1, RAD50*, and *CDH1* mutations in high-risk Finnish *BRCA1/2*-founder mutation-negative breast and/or ovarian cancer individuals (2011). Breast Cancer Res. 2011 Feb 28;13(1):R20.

II **Kuusisto KM**, Akinrinade O, Vihinen M, Kankuri-Tammilehto M, Laasanen SL, Schleutker J. Copy number variation analysis in familial *BRCA1/2*-negative Finnish breast and ovarian cancer. PLoS One. 2013 Aug 13;8(8):e71802.

III **Määttä KM***, Rantapero T*, Lindström A, Nykter M, Kankuri-Tammilehto M, Laasanen SL, Schleutker J. Whole exome sequencing of Finnish hereditary breast cancer families. Submitted.

* equal contribution

ABBREVIATIONS

AKT2	*V-Akt Murine Thymoma Viral Oncogene Homolog 2*
ALDH1	Aldehyde Dehydrogenase 1 Family, Member A1
A-T	ataxia-telangiectasia
ATM	*Ataxia Telangiectasia Mutated*
ATP	adenosine triphosphate
ATR	*Ataxia Telangiectasia And Rad3 Related*
ATRIP	ATR Interacting Protein
BABAM1	BRISC And BRCA1 Complex Member 1
BAP1	BRCA1-Associated Protein 1
BARD1	BRCA1 Associated RING Domain 1
BC	breast cancer
BCL2	B-Cell CLL/Lymphoma 2
BNIPL	*BCL2/Adenovirus E1B 19kD Interacting Protein Like*
BRAF	*B-Raf Proto-Oncogene, Serine/Threonine Kinase*
BRCA1	*Breast Cancer 1, Early Onset*
BRCA2	*Breast Cancer 2, Early Onset*
BRCC36	BRCA1/BRCA2-Containing Complex, Subunit 3
BRE	Brain And Reproductive Organ-Expressed (TNFRSF1A Modulator)
BRIP1	*BRCA1 Interacting Protein C-Terminal Helicase 1*
CASP8	*Caspase 8, Apoptosis-Related Cysteine Peptidase*
CCC	clear cell carcinoma
Cdc25A	Cell Division Cycle 25A
Cdc25C	Cell Division Cycle 25C
CDH1	*Cadherin 1, Type 1, E-Cadherin (Epithelial)*
CDKN2A	*Cyclin-Dependent Kinase Inhibitor 2A*
CDK2	Cyclin-Dependent Kinase 2
CHEK2/CHK2	*Checkpoint Kinase 2*
CHK1	*Checkpoint Kinase 1*
CI	confidence interval

CINP	*Cyclin-Dependent Kinase 2 Interacting Protein*
CNV	copy number variation
CSMD1	*CUB And Sushi Multiple Domains 1*
CtIP	Retinoblastoma Binding Protein 8
CTNNB1	*Catenin (Cadherin-Associated Protein), Beta 1, 88kDa*
DDR	DNA damage response
DENND2D	*DENN/MADD Domain Containing 2D*
DSB	double-strand break
DSS1	Deleted In Split-Hand/Foot 1
EC	endometrioid carcinoma
ECM	extracellular matrix
EDN3	*Endothelin 3*
EFCAB13	*EF-Hand Calcium Binding Domain 13*
EPHA3	*EPH Receptor A3*
EPSTI1	*Epithelial Stromal Interaction 1 (Breast)*
ER	estrogen receptor
ERBB2	*Erb-B2 Receptor Tyrosine Kinase 2*
ERBB4	*Erb-B2 Receptor Tyrosine Kinase 4*
ERVV-2	*Endogenous Retrovirus Group V, Member 2*
EXO1	*Exonuclease 1*
FA	Fanconi anemia
FAAP24	Fanconi Anemia-Associated Protein Of 24 KDa
FAM175A	*Family With Sequence Similarity 175, Member A*
FANCD1	*Fanconi Anemia, Complementation Group D1*
FANCD2	*Fanconi Anemia, Complementation Group D2*
FANCJ	*Fanconi Anemia, Complementation Group J*
FANCM	*Fanconi Anemia, Complementation Group M*
FANCN	*Fanconi Anemia, Complementation Group N*
FANCO	*Fanconi Anemia, Complementation Group O*
FFPE	formalin-fixed paraffin-embedded
FGFR2	*Fibroblast Growth Factor Receptor 2*
FOCAD	*Focadhesin*
GRB7	*Growth Factor Receptor-Bound Protein 7*
GWAS	genome-wide association study
HBOC	hereditary breast and/or ovarian cancer
HDGC	hereditary diffuse gastric cancer

HER2	Human Epidermal Growth Factor Receptor 2
HGSC	high-grade serous carcinoma
HNPCC	hereditary nonpolyposis colorectal cancer
HR	homologous recombination
HRM	high-resolution melt
HUS1	*HUS1 Checkpoint Homolog (S. Pompe)*
H2AX	H2A Histone Family, Member X
KEAP1	Kelch-Like ECH-Associated Protein 1
KRAS	*Kirsten Rat Sarcoma Viral Oncogene Homolog*
LAMA5	*Laminin, Alpha 5*
LFL	Li-Fraumeni Like-Syndrome
LFS	Li-Fraumeni Syndrome
LGSC	low-grade serous carcinoma
LKB1	*Liver Kinase B1*
LOF	loss-of-function
LSP1	*Lymphocyte-Specific Protein 1*
MAGEF1	*Melanoma Antigen Family F1*
MAP3K1	*Mitogen-Activated Protein Kinase Kinase Kinase 1, E3 Ubiquitin Protein Ligase*
MC	mucinous carcinoma
MDC1	Mediator Of DNA-Damage Checkpoint 1
MLH1	*MutL Homolog 1*
MLPA	multiplex ligation-dependent probe amplification
MRE11	Meiotic Recombination 11 Homolog A (S. Cerevisiae)
MRG15	Mortality Factor 4 Like 1
MSH2	*MutS Homolog 2*
MSH6	*MutS Homolog 6*
MYC	*V-Myc Avian Myelocytomatosis Viral Oncogene Homolog*
NBR1	*Neighbor Of BRCA1 Gene 1*
NBR2	*Neighbor Of BRCA1 Gene 2*
NBS1	Nijmegen Breakage Syndrome 1
NCOA3	*Nuclear Receptor Coactivator 3*
NGS	next-generation sequencing
NHEJ	nonhomologous end-joining
NLS	nuclear localization signal
OC	ovarian cancer

OR	odds ratio
PALB2	*Partner And Localizer Of BRCA2*
PARP	Poly(ADP-ribose) polymerase
PDZK1	*PDZ Domain Containing 1*
PIK3CA	*Phosphatidylinositol-4,5-Bisphosphate 3-Kinase, Catalytic Subunit Alpha*
PI3K	Phosphoinositide 3-Kinase
PJS	Peutz-Jeghers Syndrome
PLAU	*Plasminogen Activator, Urokinase*
PLD1	*Phospholipase D1, Phosphatidylcholine-Specific*
PMS2	*Postmeiotic Segregation Increased (S. Cerevisiae) 2*
PON-P	Pathogenic-or-Not-Pipeline
PR	progesterone receptor
PTEN	*Phosphatase And Tensin Homolog*
PTT	protein truncation test
RAD1	*RAD1 Checkpoint DNA Exonuclease*
RAD9	*RAD9 Homolog A (S. Pompe)*
RAD18	RAD18 E3 Ubiquitin Protein Ligase
RAD50	*RAD50 Homologue (S. Cerevisiae)*
RAD51	*RAD51 Recombinase*
RAD51B	*RAD51 Paralog B*
RAD51C	*RAD51 Paralog C*
RAD51D	*RAD51 Paralog D*
RAD52	*RAD52 Homologue (S. Cerevisiae)*
RAP80	Receptor Associated Protein 80
RBL2	*Retinoblastoma-Like 2*
RGMB	*Repulsive Guidance Molecule Family Member B*
RPA2	*Replication Protein A2*
RRM2B	*Ribonucleotide Reductase M2B (TP53 Inducible)*
SETBP1	*SET Binding Protein 1*
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
STK11	*Serine/Threonine Kinase 11*
S1PR5	*Sphingosine-1-Phosphate Receptor 5*
TICRR	*TOPBP1-Interacting Checkpoint And Replication Regulator*
TNBC	triple-negative breast cancer

TNRC9	*Trinucleotide Repeat Containing 9*
TopBP1	Topoisomerase (DNA) II Binding Protein 1
TOX3	*TOX High Mobility Group Box Family Member 3*
TP53	*Tumor Protein P53*
WNT3	*Wingless-Type MMTV Integration Site Family, Member 3*
WNT10A	*Wingless-Type MMTV Integration Site Family, Member 10A*
XRCC2	X-Ray Repair Complementing Defective Repair In Chinese Hamster Cells 2
XRCC3	X-Ray Repair Complementing Defective Repair In Chinese Hamster Cells 3
53BP1	Tumor Protein P53 Binding Protein 1

ABSTRACT

Breast cancer is the most frequent cancer and the most common cause of cancer deaths among females worldwide. Ovarian cancer is a highly lethal gynecologic malignancy that is the seventh most common cancer and the eight cause of death from cancer in women worldwide. In Finland, 4694 new breast cancer cases and 471 new ovarian cancer cases were diagnosed in 2012. Both breast and ovarian cancers are heterogeneous groups of diseases that can be divided into several subtypes; each subtype has distinct biological and clinical characteristics and responses to therapies. The major risk factors of breast and ovarian cancers include age and family history, and genetic predisposition accounts for as much as 10% of breast and 15% of ovarian cancers. The two major susceptibility genes for both diseases are *BRCA1* and *BRCA2*, and several other susceptibility genes have been identified. However, in the majority of high-risk breast and/or ovarian cancer (HBOC) families, the genetic predisposition factors remain unidentified, making the genetic counseling of these families challenging. The aim of this study was to obtain new information about the genetic factors that predispose individuals to breast and ovarian cancer in the high-risk Finnish *BRCA1/2*-negative HBOC families. The obtained information can be utilized in designing more efficient diagnostic, screening, prevention, and therapeutic strategies for breast and ovarian cancer, as well as new tools for genetic counseling.

Three different methodological approaches were utilized to identify genetic predisposition factors in high-risk Finnish *BRCA1/2* founder mutation-negative HBOC families: 1) mutational screening of candidate genes by Sanger sequencing, TaqMan genotyping assays, the HRM-method, and MLPA; 2) genome-wide copy number variation analysis using a SNP genotyping array; and 3) exome sequencing by target enrichment of the protein coding region of the genome and next-generation sequencing.

A candidate gene approach revealed that previously known pathogenic mutations in *BRCA1* and *CHEK2* contribute to 13.4% of cancer cases in HBOC families. The proportion of *CHEK2* mutations was remarkable and clinically relevant. Additionally, a novel and possibly pathogenic variant was detected in *BRCA2*. Copy number variation analysis identified several potential copy number variations that

likely increase the risk of HBOC susceptibility and explain the fraction of breast and ovarian cancer cases. Chromosomal aberrations at 3p11.1, 5q15, 8p23.2, and 19q13.41 were of special interest. Of these, deletions at 3p11.1 and 8p23.2 affected intronic regions of *EPHA3* and *CSMD1*, respectively, whereas duplication at 19q13.41 disrupted the coding region of the *ERVV-2* gene. Moreover, a deletion at 5q15 was located in a non-genic region but was determined to affect regulatory elements. Exome sequencing analysis focused on DNA damage repair (DDR) pathway genes. Five variants in DDR genes (*ATM, MYC, PLAU, RAD1*, and *RRM2B*) were enriched in a cohort of HBOC cases compared to controls, suggesting that these variants may be low-to-moderate risk alleles. A rare variant that may have clinical relevance was detected in *BRCA1*. Additionally, a rare variant in *RAD50* gene was suggested to predispose to male breast cancer. Moreover, defects in novel candidate genes targeting other pathways, such as DNA repair and replication, signaling, apoptosis, and the cell cycle, were identified in early-onset breast cancer patients. The interesting candidate genes included, for instance, *DENND2D, TICRR, BNIPL, EDN3,* and *FOCAD*.

In conclusion, potential germline sequence alterations and copy number variations were detected in known susceptibility genes, as well as in novel candidate genes, and the roles of the variations in HBOC predisposition were indicated. These findings warrant further confirmation and provide an excellent premise for further studies.

YHTEENVETO

Rintasyöpä on yleisin syöpä ja syöpäkuolemien syy naisilla maailmanlaajuisesti. Munasarjasyöpä on erittäin huonoennusteinen gynokologinen sairaus joka on vastaavasti naisten seitsemäksi yleisin syöpä ja kahdeksaksi yleisin syöpäkuolemien syy. Vuonna 2012 Suomessa diagnosoitiin 4694 uutta rintasyöpätapausta kun taas uusien munasarjasyöpätapausten määrä oli 471. Rinta- ja munasarjasyöpä ovat molemmat heterogeeninen joukko sairauksia, jotka voidaan jakaa useisiin alaluokkiin. Kullakin alaluokalla on tyypilliset biologiset ja kliiniset piirteet sekä terapiavasteet. Pääriskitekijöitä rinta- ja munasarjasyövälle ovat ikä ja perhehistoria. Perintötekijöiden vaikutus rintasyöpään on arviolta jopa 10 % ja munasarjasyöpään 15 %. Kaksi tärkeintä rinta- ja munasarjasyöpäalttiusgeeniä ovat *BRCA1* ja *BRCA2*. Myös lukuisia muita alttiusgeenejä tunnetaan. Kuitenkaan valtaosassa korkean rinta- ja/tai munasarjasyöpäriskin perheistä altistavia geenivirheitä ei tiedetä, jonka vuoksi perheiden perinnöllisyysneuvonta on haastavaa. Tutkimuksen tavoitteena oli saada uutta tietoa rinta- ja/tai munasarjasyövälle altistavista perintötekijöistä suomalaisissa korkean riskin rinta- ja/tai munasarjasyöpäriskin perheissä, joissa ei esiinny tunnettuja *BRCA1*- ja *BRCA2*-geenien muutoksia. Saatua tietoa voidaan hyödyntää suunniteltaessa uusia tehokkaampia strategioita rinta- ja munasarjasyövän diagnostiikkaan, seulontaan, ehkäisyyn ja hoitomuotoihin sekä perinnöllisyysneuvontaan.

Tutkimuksessa hyödynnyttiin kolmea eri menetelmällistä lähestymistapaa geneettisten alttiustekijöiden tunnistamiseksi suomalaisissa korkean rinta- ja/tai munasarjasyöpäriskin perheissä, joissa ei ole tunnettuja *BRCA1* ja *BRCA2* geenien perustajamuutoksia: 1) kandidaattigeenien mutaatioanalyysi hyödyntäen Sangerin sekvensointia, TaqMan kemiaa sekä HRM ja MLPA menetelmiä 2) genominlaajuinen kopiolukumuutosanalyysi käyttäen SNP genotyypitys sirua sekä 3) eksomisekvensointi hyödyntäen genomin proteiinia koodaavan alueen kohdennettua rikastusta sekä uuden sukupolven sekvensointia.

Kandidaattigeenien mutaatioanalyysissä löydettiin *BRCA1*- ja *CHEK2*-geenien tunnettujen haitallisten muutosten esiintyvän 13.4 %:lla korkean rinta- ja/tai munasarjasyöpäriskin perheistä. Näistä *CHEK2*-geenin muutosten osuus oli huomattava ja kliinisesti olennainen. Uusi ja haitalliseksi ennustettu muutos löytyi

BRCA2-geenistä. Kopiolukumuutosanalyysissä tunnistettiin potentiaalisia perinnölliseen rinta- ja/tai munasarjasyöpäalttiuteen vaikuttavia kopiolukumuutoksia. Kopiolukuumuutokset etenkin 3p11.1, 5q15, 8p23.2 ja 19q13.41 alueilla osoittautuivat mielenkiintoisiksi. Näistä muutokset 3p11.1 ja 8p23.2 alueilla sijaitsevat *EPHA3* ja *CSMD1*-geenien introneissa ja muutos 19q13.41 *ERVV-2* geenin koodaavalla alueella. Muutos 5q15 alueella sijaitsi geenien välisellä alueella mutta sen havaittiin mahdollisesti vaikuttavan säätelyelementteihin. Eksomisekvensoinnissa keskityttiin muutoksiin, jotka sijaitsivat DNA:n korjausreitin geeneissä. Tutkimuksessa tunnistettiin viisi muutosta *ATM, MYC, PLAU, RAD1*, ja *RRM2B* geeneissä, ja muutosten havaittiin rikastuneen rinta- ja/tai munasarjasyöpäperheissä kontrolleihin verrattuna viitaten siihen, että virheet liittyvät mahdollisesti matalasta keskisuuren syöpäriskiin. *BRCA1*-geenistä tunnistettiin lisäksi harvinainen virhe, joka voi olla kliinisesti tärkeä. Myös *RAD50*-geenistä tunnistettiin virhe, joka voi mahdollisesti altistaa miesrintasyövälle. Eksomianalyysissä keskityttiin myös potilasjoukkoon, johon kuului hyvin varhaisella iällä rintasyöpään sairastuneita naisia, ja näillä potilailla tunnistettiin virheitä uusissa potentiaalisissa kandidaattigeeneissä, jotka osallistuvat etenkin DNA:n korjaukseen ja replikaatioon, signalointiin, apoptoosiin ja solusykliin. Mielenkiintoisia kandidaattigeenejä ovat esimerkiksi *DENND2D, TICRR, BNIPL, EDN3* ja *FOCAD*.

Väitöskirjatyössä työssä tunnistettiin potentiaalisia perinnölliselle rinta- ja/tai munasarjasyövälle altistavia ituradan muutoksia jo tunnetuissa alttiusgeeneissä ja uusissa kandidaattigeeneissä. Uudet löydökset vaativat varmennusta ja tarjoavat erinomaisen lähtökohdan jatkotutkimuksille.

INTRODUCTION

Breast cancer is the most frequent cancer and ovarian cancer the seventh most frequent cancer among females worldwide, representing approximately 25% and 4% of all cancers, respectively (Ferlay et al, 2015). In Finland, 4694 new breast cancer cases and 471 new ovarian cancers were diagnosed in 2012 (Finnish Cancer Registry). Ovarian cancer is a particularly lethal gynecological malignancy and is commonly diagnosed when the disease is already at a late stage. Therefore, the survival rate for ovarian cancer is much lower than for breast cancer. Both breast and ovarian cancer are heterogeneous diseases composed of different tumor types with distinctive features and behaviors. The main risk factors for breast and ovarian cancer include age, family history, and genetics. The genetic components of both of the diseases have been well established, contributing to up to 10% of all breast cancer cases and 15% of all ovarian cancer cases (Claus et al, 1996, Lynch et al, 2009). Less than half of the genetic predisposition to breast cancer has been resolved. Predisposing factors can be classified into three different categories based on the risk associated with the disease: high-risk, moderate-risk, and low-risk genes. Two major high-risk genes are *Breast Cancer 1, Early Onset (BRCA1)* and *Breast Cancer 2, Early Onset (BRCA2)*, and rare defects in these genes explain a significant percentage (15-20%) of the genetic predisposition to breast cancer (Miki et al, 1994, Turnbull & Rahman, 2008, Wooster et al, 1994). *BRCA1* and *BRCA2* are tumor suppressor genes that have central roles in the DNA damage response (DDR) pathway. These genes were detected by linkage analysis and positional cloning in the mid-90s. Since the identification of *BRCA1* and *BRCA2*, the DDR pathway has been one of the most studied pathways in breast cancer pathogenesis. Therefore, other genes participating the DDR pathway have been considered good candidates for breast cancer susceptibility and have been studied through candidate gene approaches. In this way, rare moderate-risk defects in *Partner And Localizer Of BRCA2 (PALB2)*, *Checkpoint Kinase 2 (CHEK2)*, and *Ataxia Telangiectasia Mutated (ATM)* have been found to contribute a fraction of the breast cancer cases (Erkko et al, 2007, Renwick et al, 2006, Vahteristo et al, 2002). Additionally, rare mutations in high-to-moderate risk genes associated with cancer syndromes, such as *Tumor Protein P53 (TP53), Phosphatase And Tensin Homolog (PTEN),* and *Cadherin 1, Type 1, E-Cadherin (Epithelial) (CDH1)*, explain a fraction of the

hereditary breast cancer cases (Lynch et al, 1997, Masciari et al, 2007, McBride et al, 2014). Moreover, genome-wide association study (GWAS) approaches have identified common low-risk alleles in over 70 loci, and their contribution to breast cancer predisposition has been estimated to be approximately 14% (Michailidou et al, 2015). In ovarian cancer, genetic predisposition can be explained by defects in high-to-moderate-risk genes, such as *BRCA1, BRCA2, MutL Homolog 1 (MLH1), MutS Homolog 2 (MSH2), RAD51 Paralog C (RAD51C)*, and *BRCA1 Interacting Protein C-Terminal Helicase 1 (BRIP1)* (Lynch et al, 2009, Pelttari et al, 2011, Rafnar et al, 2011). Of these genes, *BRCA1* and *BRCA2* explain most (90%) of the genetic predisposition to ovarian cancer. Despite intensive efforts to identify additional breast and ovarian cancer susceptibility genes using different methodological approaches, the majority of predisposing factors remain unidentified, especially in high-risk breast and/or ovarian cancer families that do not carry mutations in the two major high-risk genes, *BRCA1* and *BRCA2*.

In recent years, next-generation sequencing (NGS) technologies have provided high-throughput applications for cancer genetic studies. Particularly, whole exome sequencing has proven to be a cost-effective method to identify novel susceptibility genes. The exome (i.e., the protein coding region of the genome) represents only 1-2% of the whole genome but harbors over 85% of disease-associated mutations, making it an attractive target in disease gene identification (Ng et al, 2009).

Several breast and ovarian cancer susceptibility genes have been identified, and both of these diseases are considered genetically heteregeneous. The unknown portion of the genetic contribution to breast and ovarian cancer is believed to consist a large number of family-specific, low-to-moderate risk factors that act in multiplicative fashion (i.e., a polygenic model). The aim of the current study was to utilize different methodological approaches to identify genetic factors that predispose individuals to hereditary breast and/or ovarian cancer (HBOC) in the high-risk Finnish *BRCA1/2* founder mutation-negative HBOC families. The overall aim was to obtain novel information on breast and ovarian cancer genetics, which could then be utilized in the design of more efficient clinical management strategies for hereditary breast and ovarian cancer.

REVIEW OF THE LITERATURE

1 Breast cancer

1.1 Breast overview

The breast consists of 15-20 lobes (consisting of smaller sections termed lobules), the nipple, ducts (thin tubes connecting the lobes and nipples), fatty- and fibrous tissue, as well as blood and lymphatic vessels (Figure 1). The main function of the breast is to produce milk.
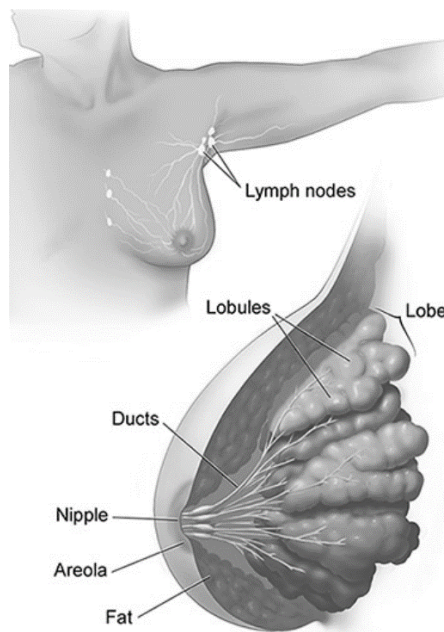
Figure 1. Breast and adjacent lymph nodes. SEER Cancer Statistics Factsheets: Female Breast Cancer (National Cancer Institute, Surveillance, Epidemiology, and End Results Program).

1.2 Epidemiology

Breast cancer (BC) is the most frequent cancer in the world among females, with an estimated 1.67 million new cancer cases diagnosed in 2012 (25.2% of all cancers) (Ferlay et al, 2015). BC is also the most common cause of cancer deaths among females worldwide, with an estimated 522,000 deaths in 2012 (14.7% of all cancer deaths) (Ferlay et al, 2015). The incidence rates of BC vary nearly fourfold across different regions, with rates ranging from 27 per 100,000 in Middle Africa and Eastern Asia to 96 per 100,000 in Western Europe (Ferlay et al, 2015). Differences in incidence rates between countries can be explained by several factors, including ethnicity, genetics, socio-economically correlated environmental factors related to lifestyle, nutrition, the use of exogenous hormones, reproduction, mammographic screening, and cancer treatment possibilities (Bray et al, 2004, Jemal et al, 2010). However, the incidence rates in developing countries have begun to increase in past few decades due to the increased adoption of lifestyles that are common in Western countries, including smoking, the consumption of saturated fat and calorie-dense food, physical inactivity, the use of oral contraceptives, late child bearing, and fewer pregnancies (Bray et al, 2004, Jemal et al, 2010). Mortality rates for BC also vary between countries ranging from 6 per 100,000 in Eastern Asia to 20 per 100,000 in Western Africa (Ferlay et al, 2015). In developed countries with high incidence rates, the mortality rates have been stable or are decreasing due to a reduction in the use of menopausal hormone therapy, early detection by mammographic screening, and improved treatment possibilities (Jemal et al, 2010). Moreover, five-year relative survival rates vary from approximately 40% to 90% in low- and high-income countries, respectively (Coleman et al, 2008). In Finland, 4694 new BC were diagnosed in 2012, with an incidence rate of 91.3 per 100,000, a mortality rate of 14 per 100,000, and a five-year relative survival rate of 89% (Finnish Cancer Registry).

1.3 Risk factors

Gender. Being a female is the major risk factor. BC also occurs among men, but it is much rarer, with an incidence of less than 1% that of female BC (Miao et al, 2011).

Age. BC is most common in middle-aged and older women. The median age at diagnosis is 61 years (National Cancer Institute, Surveillance, Epidemiology, and End Results Program).

Race/Ethnicity. White non-Hispanic females have the highest incidence rates of BC, whereas the highest mortality rates of BC are observed among African-American females (National Cancer Institute, Surveillance, Epidemiology, and End Results Program).

Personal history of BC. A woman with previous BC has an elevated risk of developing a second cancer of the contralateral breast (Molina-Montes et al, 2014). Moreover, *BRCA1/2*-mutation carriers are at higher risk of contralateral BC than non-carriers (Molina-Montes et al, 2014). Additionally, benign breast conditions and high breast density are strong risk factors for BC (Tice et al, 2013).

Family history of BC. Family history is a strong risk factor for BC. However, the extent of the risk varies according to the nature of the family history (i.e., the type of relative affected, the age at which the relative developed BC, and the number of relatives affected) (Pharoah et al, 1997). A woman's risk of BC is two or more times greater if she has first-degree relative (mother, sister, or daughter) who developed the disease before the age of 50. Moreover, the younger the relative is when she develops BC, the greater the risk (McPherson et al, 2000). The BC risk increases by between four and six times if two first-degree relatives develop the disease (McPherson et al, 2000). The risk is also increased, although to a lesser extent, if the BC is diagnosed in a second-degree relative or any relative at all (Pharoah et al, 1997). BC risk is age-specific, and the risk is higher in women under 50 years of age who have a relative with early-onset BC (Pharoah et al, 1997). Moreover, family history of ovarian cancer increases the risk of BC given that both cancers are a part of HBOC syndrome caused by defects in *BRCA1* and *BRCA2* (Lynch et al, 2009).

Genetics. The occurrence of several BC cases in the family with certain features can indicate a genetic predisposition to the disease. These features include 1) early age of onset; 2) a bilateral BC; or 3) the occurrence of other cancer including ovarian and male BC (McPherson et al, 2000). Approximately 5-10% of BCs in the general population are estimated to be caused by genetic factors, primarily related to the two major high-risk BC susceptibility genes, *BRCA1* and *BRCA2* (Claus et al, 1996, Miki et al, 1994, Wooster et al, 1994). Several other BC predisposition genes have been identified and can be classified into three categories based on the different levels of risk and prevalence in the population: rare high-penetrance risk genes, rare moderate-penetrance risk genes, and common low-penetrance risk genes (Stratton & Rahman, 2008). Known susceptibility genes explain less than half of the genetic predisposition to BC (Couch et al, 2014).

Hormonal and reproductive factors. A prolonged or increased exposure to estrogen, including early age of menarche, late age at menopause, nulliparity, and late

age at first birth are associated with and increased risk of BC. In contrast, reducing exposure to estrogen, such as through long-term breast-feedings is thought to be protective (Collaborative Group on Hormonal Factors in Breast Cancer, 2002, Martin & Weber, 2000, McPherson et al, 2000). Exposure to exogenous hormones, such as use of oral contraceptives or postmenopausal hormone replacement therapy (use for over 10 years), is thought to elevate the risk for BC (Collaborative Group on Hormonal Factors in Breast Cancer, 1996, McPherson et al, 2000). Furthermore, treatments with the synthetic estrogen diethylstilbestrol during pregnancy have been associated with increased BC risk both in pregnant woman and in the unborn daughter (Palmer et al, 2006). The differential risk of BC among *BRCA1* and *BRCA2* mutation carriers has been associated with reproductive factors (Pan et al, 2014). For example, a later age at first birth is associated with a lower risk of BC in *BRCA1* but not *BRCA2* mutation carriers, and breast feeding for at least 1-2 years conferred a 37% reduction in BC risk for *BRCA1* but not *BRCA2* mutation carriers (Pan et al, 2014).

Environmental factors. A growing number of studies have implicated dietary factors in BC development, but the results have been somewhat conflicting. High intake of red meat, animal fat and saturated fatty acids have been reported to increase the risk of BC, whereas high intake of vegetables, fruits, fiber, unsaturated fatty acids, and phyto-estrogens (obtained from soya products, sourdough rye bread, berries) are suggested to reduce the BC risk (Hanf & Gonder, 2005). Obesity is a known risk factor for several cancers, including BC, although the exact molecular mechanisms are poorly understood (Khandekar et al, 2011). Specifically, an increased body mass index has been associated with BC risk in postmenopausal women (Renehan et al, 2008). Moreover, physical activity has been associated with a decreased BC risk, and several epidemiologic studies have found a 25% average risk reduction amongst physically active women compared to the least active women (Lynch et al, 2011). Alcohol consumption during the time when breast tissue is particularly vulnerable to carcinogens (between menarche and first full-time pregnancy) has been associated with increased risk of BC (Liu et al, 2013). Smoking has been reported to increase the BC risk, although there is inconsistency between various epidemiologic studies (Terry & Rohan, 2002). Ionizing radiation (both diagnostic and therapeutic) is a well-known risk factor for the development of BC (Drooger et al, 2015). There is a clear positive dose-risk relation, which is modified by age, whereby young age at exposure is associated with an increased risk (Drooger et al, 2015). Moreover, it has been found that patients with *BRCA1* and *BRCA2* mutations might be more

sensitive to the deleterious effects of ionizing radiation due to an impaired capacity of repairing double strand DNA breaks (Drooger et al, 2015).

1.4 Clinical features and pathology

BC is most frequently diagnosed among women aged 55-64, and the median age at diagnosis is 61 years (National Cancer Institute, Surveillance, Epidemiology, and End Results Program). The first sign of BC is usually a lump in the breast. Other symptoms may also include changes in breast size and shape, changes in nipple shape and the surrounding area, skin changes, nipple discharge other than milk, and pain in the breast.

BC is not a single disease. Instead, it can be divided into various subtypes with distinctive histological and biological characteristics, clinical behaviors, and responses to therapy. Histologically, breast tumors can be broadly categorized into in situ carcinomas and invasive (infiltrating) carcinomas (Figure 2). In situ carcinoma is a pre-invasive BC in which malignant cells are confined within the site of origin and which can transform into an invasive cancer over a few years or even decades (Barnes et al, 2012). However, only a subset of in situ cancers become invasive, and recent studies do not support the notion that in situ carcinoma is an obligate precursor of invasive BC (To et al, 2014). In situ carcinomas account for approximately 20% of all diagnosed BCs (To et al, 2014). Moreover, in situ carcinomas are further divided into ductal and lobular based on the origin of the cancer cells (Figure 2). Ductal carcinoma in situ is a more common and heterogeneous group of tumors than lobular carcinoma in situ, which presents low histological variation (Figure 2). Similarly to in situ tumors, invasive carcinomas are classified into subtypes, with the major subtypes being infiltrating ductal, invasive lobular, ductal/lobular, mucinous, tubular, medullary, and papillary (Figure 2). Of these, infiltrating ductal carcinoma and lobular carcinoma are the most common subtypes, accounting for approximately 50-80% and 5-15% of all breast carcinomas, respectively (Weigelt & Reis-Filho, 2009). In addition to histological type, histological grade (grades 1-3) is used to classify breast carcinomas. Grade is an assessment of the degree of differentiation and proliferative activity of a tumor and reflects its aggressiveness (Weigelt et al, 2010). Moreover, several molecular markers are used to further classify invasive carcinomas to determine which patients are likely to respond to targeted therapies. The most commonly used markers include estrogen

receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) (Payne et al, 2008).
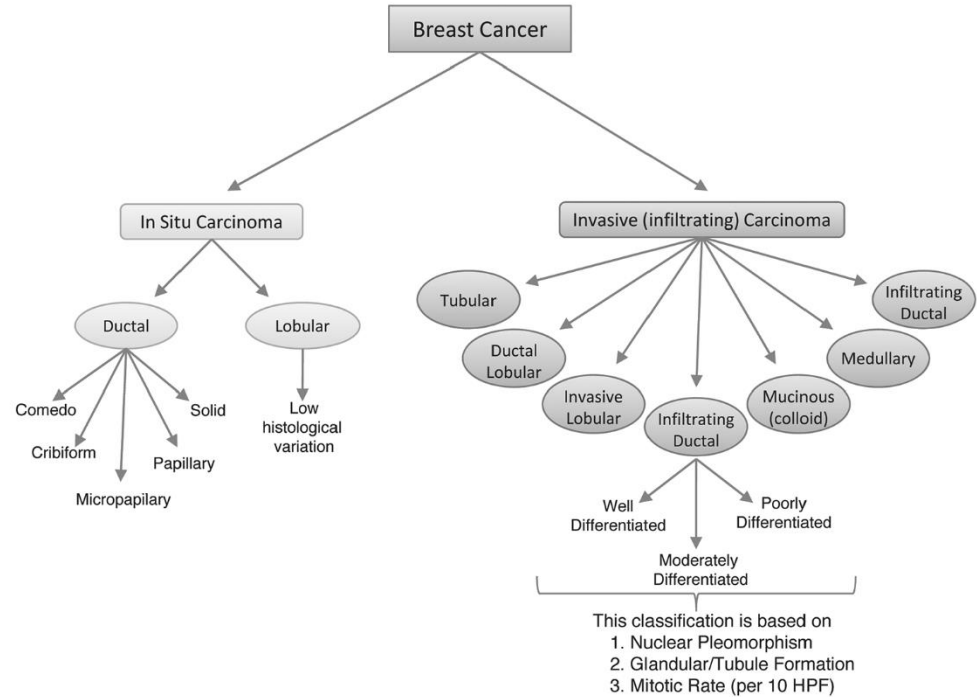
Figure 2. Histological classification of breast cancer subtypes (Malhotra et al, 2010). Reprinted by permission of Taylor and Francis LLC.

More recently, new technologies have been used to further differentiate the molecular subtypes of BC according to gene expression profiles. The five main molecular classes of BC have been determined to be luminal A (ER/PR+, HER2-), luminal B (ER/PR+, HER2+), HER2-overexpressing (ER/PR-, HER2+), basal-like (ER/PR-, HER2-), and normal breast-like tumors (unclassified) (Perou et al, 2000, Sorlie et al, 2001, Sorlie et al, 2003). Furthermore, a claudin-low subtype (ER/PR-, HER2-) has also been identified (Herschkowitz et al, 2007). The luminal A group of tumors show high expression of ERα and related transcription factors, whereas luminal B has high expression of a cluster of genes related to proliferation (Santos et al, 2015). The HER2-overexpressing subtype is characterized by a high expression of genes at 17q22.24, including *Erb-B2 Receptor Tyrosine Kinase 2 (ERBB2)* and *Growth Factor Receptor-Bound Protein 7 (GRB7),* whereas the basal-like subtype highly expresses

laminin, and keratins 5 and 17 (Santos et al, 2015). Normal breast-like tumors have high expression levels of adipose tissue genes. The claudin-low subtype is characterized by low expression of claudin genes, which are involved in tight junctions and cell-cell adhesion, and high expression of vimentin, N-cadherin, and immune system response genes (Santos et al, 2015). Basal-like, HER2-overexpressing, and claudin-low tumors are the most aggressive and are associated with a poor survival. In contrast, the luminal A type is the class of least aggressive tumors (Santos et al, 2015). Both basal-like and HER2-overexpressing subtypes are associated with a high frequency of *TP53* mutations, whereas the basal-like subtype is associated with only *BRCA1* mutations (Cancer Genome Atlas Network, 2012). Basal-like tumors are often referred to as triple-negative BCs (TNBCs) because most are negative for ER, PR and HER2 (Cancer Genome Atlas Network, 2012).

In addition to molecular classification of BC, a functional classification system based on the BC stem cells is emerging (Malhotra et al, 2010). In this system, BCs are classified based on the tumor-initiating cells, and there are currently two hypotheses in this regard. One suggests that BC heterogeneity arises from distinct mammary stem/progenitor cells at various levels within the mammary cell hierarchy, whereas the other hypothesis is that BC originates from a single mammary stem/progenitor cell that is transformed by various oncogenes to give rise to various types of cancer (Malhotra et al, 2010). Several markers for this classification system have been identified, including CD44/CD24 and Aldehyde Dehydrogenase 1 Family, Member A1 (ALDH1) (Malhotra et al, 2010).

2 Ovarian cancer

2.1 Ovaries overview

The ovaries are a pair of organs in the female reproductive system and are located in the pelvis, one on each side of the uterus (Figure 3). The main function of the ovaries is to produce egg cells and the female hormones, progesterone and estrogen.
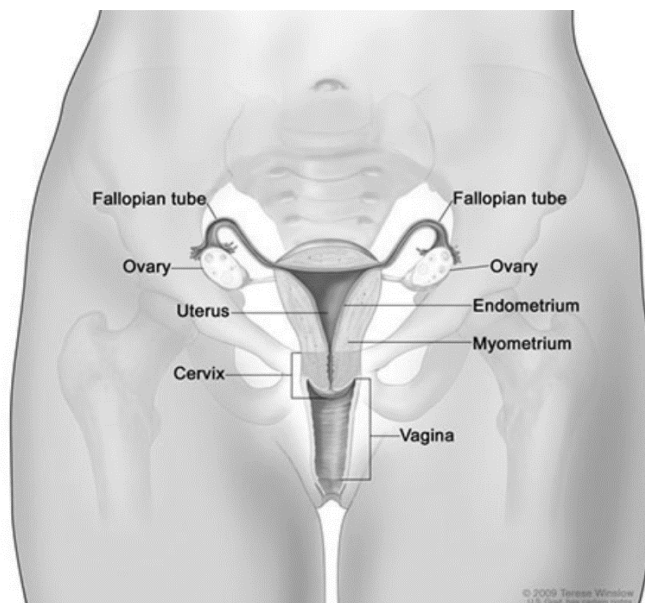
Figure 3. Female reproductive anatomy. SEER Cancer Statistics Factsheets: Ovary Cancer (National Cancer Institute, Surveillance, Epidemiology, and End Results Program).

2.2 Epidemiology

Ovarian cancer (OC) is a highly lethal gynecologic malignancy. It is the seventh most common cancer and the eighth highest cause of death from cancer in women worldwide, with an estimated 239,000 new cases (3.6% of all cancers) and 152,000

deaths (4.3% of all cancer deaths) in 2012 (Ferlay et al, 2015). The incidence rates of OC are highest in more developed regions, with rates in these areas exceeding 7.5 per 100,000, whereas the lowest incidence is in Africa, with rates below 5 per 100,000 (Ferlay et al, 2015). Differences in incidence rates around the world are due to ethnicity, genetic factors, socio-economically correlated environmental factors related to lifestyle, nutrition, the use of exogenous hormones, reproduction, and both diagnostic and medical treatment possibilities (La Vecchia, 2001, Lowe et al, 2013). The mortality rates of OC are higher in developed regions, such as North America and Europe (from 5 to 6 per 100,000) than in less developed regions, such as Eastern Asia (2 per 100,000) (Ferlay et al, 2015). More than 70% of women with OC are diagnosed with advanced disease (Rauh-Hain et al, 2011). The five-year survival rates for women with advanced disease vary from 20% to 30% (Rauh-Hain et al, 2011). In Finland, 471 new OC cases were diagnosed in 2012 (Finnish Cancer Registry), making OC the tenth most common cancer and the fifth highest cause of death from cancer among women. In Finland, the incidence of OC is 8.6 per 100,000, the mortality rate is 4.8 per 100,000, and the five-year survival rate is 49% (Finnish Cancer Registry).

2.3 Risk factors

Gender. OC is a sex-specific cancer, and being a female is a major risk factor.

Age. OC risk increases with age, and the majority of OCs are diagnosed in older women. The median age at diagnosis is 63 years (National Cancer Institute, Surveillance, Epidemiology, and End Results Program).

Race/Ethnicity. The highest incidence and mortality rates of OC are observed among white females (National Cancer Institute, Surveillance, Epidemiology, and End Results Program).

Family history of OC. A family history of OC is one of the strongest risk factors for the disease. A woman with an OC-affected first-degree relative (mother, sister, daughter) has a 5% risk, and a woman with two OC-affected first-degree relatives has a 7% risk for developing the disease, whereas the lifetime risk for developing OC in the general population is 1.6% (Prat et al, 2005). Additionally, a family history of BC increases the risk of developing OC given that these two cancers comprise the hereditary breast and ovarian cancer (HBOC) syndrome, which is caused by defects in *BRCA1* and *BRCA2* (Lynch et al, 2009).

Genetics. It is estimated that inherited susceptibility can explain 5-15% of all OCs (Lynch et al, 2009). Hereditary OC occurs in three different forms, site-specific OC and as a component of two cancer syndromes, HBOC and Lynch syndrome (also known as hereditary nonpolyposis colorectal cancer, or HNPCC syndrome) (Prat et al, 2005). The causative genes for site-specific hereditary OC and HBOC syndromes are two tumor suppressor genes, *BRCA1* and *BRCA2*, whereas Lynch syndrome is associated with defects in DNA mismatch repair genes, including *MLH1* and *MSH2* (Prat et al, 2005). The majority (65-85%) of all hereditary OCs are due to mutations in *BRCA1* and *BRCA2,* and another 10-15% of hereditary cases can be explained by Lynch syndrome gene mutations (Lynch et al, 2009). A carrier of a *BRCA1* or *BRCA2* mutation, unselected for family history, has an average cumulative lifetime risk of 39% and 11% for developing OC by the age of 70, respectively (Antoniou et al, 2003). Moreover, the risk estimates for mutation carriers are higher in HBOC families (Lynch et al, 2009).

Hormonal and reproductive factors. The ovarian epithelium responds strongly to the local hormonal environment. Long-term exposure to elevated estrogen levels, including early age of menarche, late age at natural menopause, and hormone replacement therapy increase the risk of OC (Hunn & Rodriguez, 2012). Pregnancies (especially after 35 years of age), having many children, breastfeeding, and the use of oral contraceptives have protective effects (Hunn & Rodriguez, 2012).

Environmental factors. Several environmental factors related to lifestyle have been reported to influence OC risk, but the study results are somewhat inconclusive or controversial. Overweight and obesity (body mass index greater than 30) in early adulthood have been reported to increase the risk of OC (Olsen et al, 2007). In contrast, moderate physical exercise has been suggested to lower OC risk (Cannioto & Moysich, 2015). Dietary factors, such as the intake of carbohydrates and dairy, have been suggested to increase the risk of OC, whereas the consumption of green leafy vegetables, vegetable oils and fish have been shown to have a protective effect (Hunn & Rodriguez, 2012). Additionally, a high dietary intake of vitamin D has been reported to be protective against specific histological subtypes of OC (Merritt et al, 2013). Smoking has been reported to increase the risk of mucinous subtype of OC, but its effect on overall OC risk is uncertain (Collaborative Group on Epidemiological Studies of Ovarian Cancer et al, 2012).

Inflammatory factors. Inflammatory factors have been reported to be involved in ovarian carcinogenesis. Endometriosis has been shown to be associated with an increased risk of ovarian carcinoma, especially endometrioid and clear cell types (Prowse et al, 2006). Moreover, pelvic inflammatory disease and perineal talc

exposure have been shown contribute to OC risk, although the results for the latter risk factor are somewhat inconclusive (Houghton et al, 2014, Huncharek et al, 2003, Lin et al, 2011).

2.4 Clinical features and pathology

OC develops at a later age and is most typically observed in women over 60 years of age. The hereditary form of OC is diagnosed about a decade earlier than non-hereditary forms (Jazaeri, 2009). The symptoms of OC are unclear and can be easily confused with those of common conditions, such as problems with digestion. Symptoms may include abdominal bloating, pelvic and abdominal pain, feeling full quickly, and difficulty eating (Goff, 2012). In the absence of clinically significant symptoms and a lack of efficient screening methods, OC is primarily detected when the disease is already at a late stage.

The origin and pathogenesis of OC are poorly understood making early detection and new therapeutic approaches difficult. OC is a heterogeneous disease composed of different types of tumors that have distinct features, behavior, and treatment responses. Over 90% of the ovarian tumors are considered to originate from ovarian surface epithelial cells, whereas other tumor types are considered to arise from germ cells and sex-cord-stromal cells (Lynch et al, 2009). Based on the histopathology and molecular genetic analyses, at least five main types of ovarian carcinomas are identified: high-grade serous carcinoma (HGSC; 70%), endometrioid carcinoma (EC; 10%), clear cell carcinoma (CCC; 10%), mucinous carcinoma (MC; 3%), and low-grade serous carcinoma (LGSC; <5%), accounting for a total of 98% of ovarian carcinomas (Prat, 2012). All these tumor types have different risk factors, precursor lesions, patterns of spread, molecular events during oncogenesis, response to chemotherapy, and prognosis (Table 1). HGSC is the most common ovarian carcinoma and is detected at an advanced stage in approximately 80% of patients (Prat, 2012). Thus, HGSC has the poorest prognosis of the different tumor types (Table 1). Moreover, HGSC type particularly has been associated with germline defects in *BRCA1/2* genes (Risch et al, 2006), whereas EC and CCC types are associated with endometriosis (Rosen et al, 2009). Relevant in this regard, EC is the major HNPCC type (Prat, 2012). Moreover, OCs can be further classified according to stages. The staging of OC (stages I-IV) is based on the FIGO nomenclature (Heintz et al, 2006). Stage I disease is limited to the ovaries, whereas stage IV indicates metastatic disease.

Based on the distinct morphologic and molecular genetic features, a dualistic model of ovarian tumorigenesis has been proposed in which OCs can be divided into two subgroups: Type I (low-grade pathway) and Type II (high-grade pathway) (Kurman & Shih, 2010). Type I tumors consist of low-grade serous, low-grade endometrioid, mucinous, clear cell and transitional (Brenner) carcinomas. Type II tumors consist of high-grade serous carcinoma, undifferentiated carcinoma and malignant mixed mesodermal tumors (carcinosarcomas). Type I tumors are generally indolent, present in stage I (tumor confined to the ovary), and develop from well-established precursors; these tumors are referred to as borderline tumors. Type I tumors are relatively genetically stable and present specific mutations in certain genes, such as *Kirsten Rat Sarcoma Viral Oncogene Homolog (KRAS), B-Raf Proto-Oncogene, Serine/Threonine Kinase (BRAF), ERBB2, Catenin (Cadherin-Associated Protein), Beta 1, 88 kDa (CTNNB1), PTEN,* and *Phosphatidylinositol-4,5-Bisphosphate 3-Kinase, Catalytic Subunit Alpha (PIK3CA)*; however, Type I tumors rarely exhibit *TP53* mutations. Type II tumors are aggressive, present in advanced stage, develop from intraepithelial cells, are genetically highly unstable, and have a high frequency of *TP53* mutations.

Moreover, there is compelling evidence that tumors that have been traditionally considered primary ovarian tumors (high-grade serous, clear cell, and endometrioid) actually originate in other pelvic organs and involve the ovary secondarily. Specifically, high-grade serous carcinomas are reported to arise from epithelial lesions in the distal fimbriated end of the fallopian tube, and clear cell and endometrioid carcinomas originate from ovarian endometriosis (Kurman & Shih, 2010, Prat, 2012)

Table 1.   Ovarian carcinoma: clinical and molecular features of the five most common types. (Prat, 2012) by permission of Oxford University Press.

	HGSC	LGSC	MC	EC	CCC
Risk factors	*BRCA1/2*	?	?	HNPCC	?
Precursor lesions	Tubal intraepithelial carcinoma	Serous borderline tumor	Cystadenoma/ borderline tumor?	Atypical endometriosis	Atypical endometriosis
Pattern of spread	Very early transcoelomic spread	Transcoelomic spread	Usually confined to ovary	Usually confined to pelvis	Usually confined to pelvis
Molecular abnormalities	*BRCA, p53*	*BRAF, KRAS*	*KRAS, HER2*	*PTEN, ARID1A*	*HNF1, ARID1A*
Chemosensitivity	High	Intermediate	Low	High	Low
Prognosis	Poor	Intermediate	Favorable	Favorable	Intermediate

Abbreviations: CCC, clear cell carcinoma; EC, endometrioid carcinoma; HGSC, high-grade serous carcinoma; HNPCC, hereditary polyposis colorectal carcinoma; LGSC, low-grade serous carcinoma; MC, mucinous carcinoma.

3   DNA Damage Response (DDR) pathway

The maintenance of genome integrity ensures the transmission of correct genetic material across generations. Genetic material is continuously threatened by spontaneous damages, such as occurs during DNA metabolism, and by damaging agents coming from outside (exogenic) or inside (endogenic) the cell. Exogenic threats include ultraviolet light, ionizing radiation, and chemicals, whereas endogenic threats include reactive oxygen species, which are side products of normal cellular metabolism. To protect genetic material from such damage, cells use a DNA damage response (DDR) system that detects DNA damage and promotes the appropriate cellular response, such as senescence, cell cycle checkpoint activation, DNA repair, apoptosis, or tolerance (Figure 4). If the DDR machinery does not work properly, the genome becomes unstable, which may result in uncontrolled behavior of the cell and, eventually, cancer development. These processes are reviewed in (Bartek et al, 2007, Ciccia & Elledge, 2010, Harper & Elledge, 2007, Zannini et al, 2014)
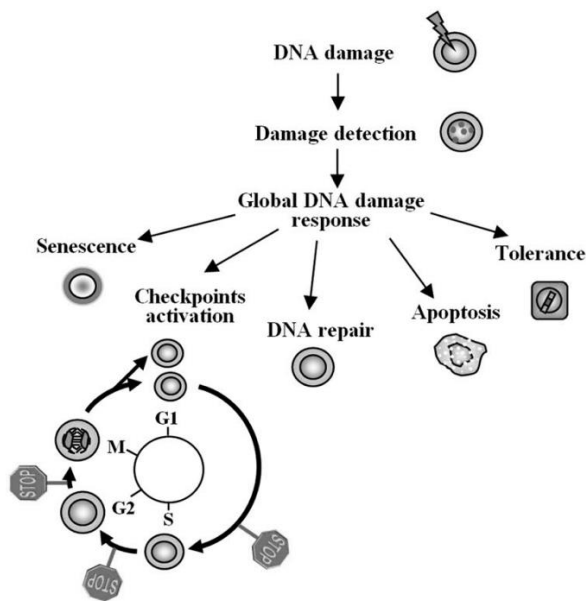
Figure 4.   The DNA damage response promotes the appropriate cellular response. (Zannini et al, 2014) by permission of Oxford University Press.

3.1 Different DDR repair mechanisms

DNA damage can be either single-stranded or double-stranded, and different repair mechanisms are activated depending on the type of damage. Mispaired DNA bases are changed to correct bases by mismatch repair, and small chemical alterations of DNA bases and single-strand breaks are corrected by base excision repair. More complex single-strand errors, such as pyrimidine dimers and intrastrand crosslinks, are repaired by nucleotide excision repair. For the interstrand crosslinks, interstrand crosslink repair is used. DNA double strand breaks (DSB) are the most deleterious form of DNA damage and are repaired by at least by four independent mechanisms: nonhomologous end joining (NHEJ), homologous recombination (HR), alternative NHEJ, and single strand annealing. Of these processes, NHEJ and HR are the two major mechanisms. Depending on the extent of DNA end processing, different mechanisms are used to repair DSBs. HR is considered the most error-free mechanism as it utilizes sister chromatids as a template for the synthesis of new DNA. These processes are reviewed in (Ciccia & Elledge, 2010, Lord & Ashworth, 2012).

3.2 Key proteins in DDR

The DDR machinery is a complex network comprising numerous pathways, proteins, and protein complexes that function in a well-coordinated manner. The DDR is involved in all steps of this process, from DNA damage detection to the activation of a cellular response to the damage. In basic terms, the DDR cascade consists of proteins termed sensors, apical kinases, mediators, downstream kinases, and effectors (Figure 5). The major regulators of DDR are the phosphoinositide 3-kinase (PI3K)-related proteins kinases, ataxia-telangiectasia mutated (ATM), and ATM and RAD3-related (ATR), which share many biochemical and functional similarities. ATM primarily functions in response to DSBs, whereas ATR is primarily activated in response to replication stress. However, both ATM and ATR target an overlapping set of substrates in the DDR cascade. DNA lesions are recognized by sensor proteins that vary with the different DDR regulators. For ATM, damage is recognized by the MRN complex, which consists of Meiotic Recombination 11 Homolog A (S. Cerevisiae) (MRE11), RAD50 Homologue (S. Cerevisiae) (RAD50), and Nijmegen Breakage Syndrome 1 (NBS1). For ATR, the damage-sensing proteins include replication protein A (RPA) and the 9-1-1 complex bound by ATR

interacting protein (ATRIP). The 9-1-1 complex consists of RAD9 Homolog A (S. Pompe) (RAD9), *RAD1 Checkpoint DNA Exonuclease (RAD1)*, and *HUS1 Checkpoint Homolog (S. Pomple) (HUS1)*. Following detection of the DNA damage, ATM and ATR initially phosphorylate mediator proteins, which can amplify the DDR by acting both as recruiters of ATM/ATR substrates and as scaffolds upon which to assemble complexes. At the site of DNA damage, phosphorylation of histone variant H2A Histone Family, Member X (H2AX) on Serine 139 by ATM and ATR kinases is required to recruit mediators, such as Mediator of DNA-Damage Checkpoint 1 (MDC1). Other mediator proteins include, for example, Tumor Protein P53 Binding Protein 1 (53BP1), BRCA1, Topoisomerase (DNA) II Binding Protein (TopBP1), and Claspin. Two kinases, CHK2 for ATM and Checkpoint Kinase 1 (CHK1) for ATR, are involved in spreading the DNA damage signal through a phosphorylation cascade. Along with ATM and ATR, CHK1 and CHK2 also phosphorylate effector proteins, such as p53 and Cell Division Cycle 25 (Cdc25), which execute DDR cellular responses. Additionally, a large number of other proteins are known to participate in the DDR cascade. The DDR has also been discovered to play a role in variety of different pathways, including RNA splicing, chromatin remodeling, transcription, ubiquitination, and circadian rhythms. These findings are reviewed in (Ciccia & Elledge, 2010, Cimprich & Cortez, 2008, Harper & Elledge, 2007, Sulli et al, 2012).
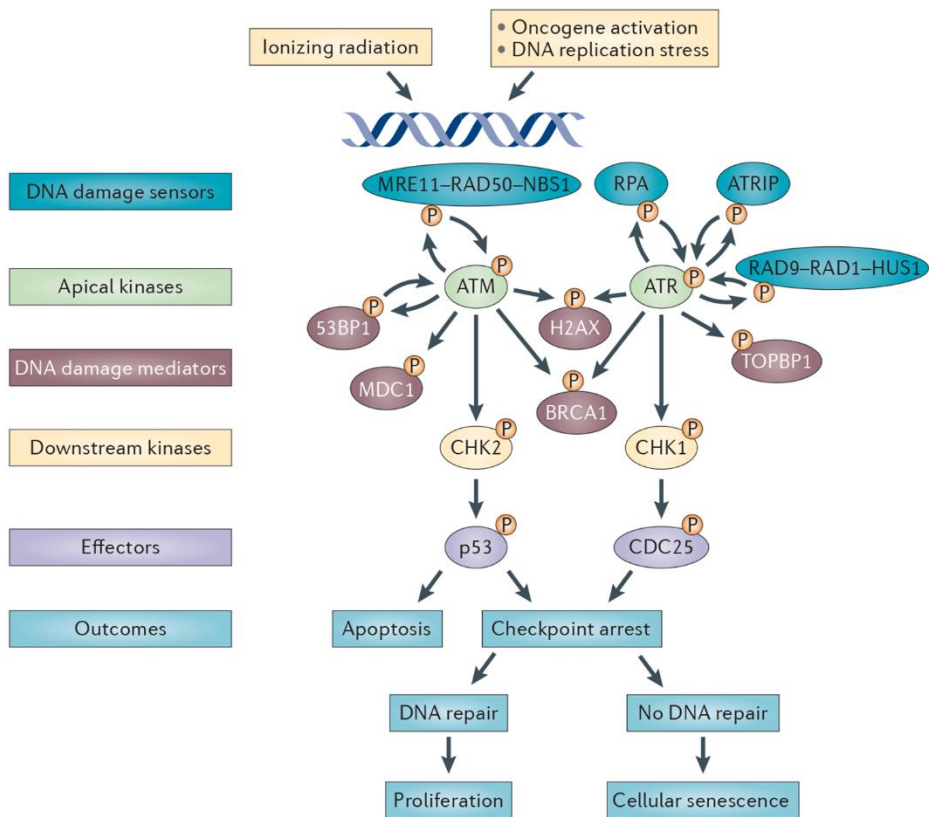
Figure 5. Key proteins in the DNA damage response. Reprinted by permission from Macmillan Publisher Ltd: [NATURE REVIEWS CANCER] (Sulli et al, 2012), copyright (2012).

3.3 DDR and cancer

The DDR plays a central role in human physiology. Hereditary defects in genes encoding key proteins in the DDR contribute to various human diseases, including neurological disorders, infertility, immune deficiency, premature aging, and cancer.

Cancer, in particular, is driven by genomic instability. Several DDR-related cancer syndromes have been recognized. The DDR syndromes, which can include breast and ovarian cancer, include HNPCC syndrome, familial breast cancer syndrome, Fanconi anemia (FA), Ataxia-telangiectasia (A-T), and Li-Fraumeni syndrome (LFS). Of these, HNPCC is related to defects in mismatch repair genes such as *MLH1, MSH2, MutS Homolog 6 (MSH6)*, and *Postmeiotic Segregation Increased (S. Cerevisiae) 2 (PMS2)*. Familial breast cancer syndrome is related to defects in homologous

recombination and damage signaling, and the causative genes include *ATM, BRCA1, BRCA2, BRIP1, CHK2, NBS1, PALB2, RAD50,* and *RAD51C.* Moreover, FA is related to defects in interstrand crosslink repair and homologous recombination, and several FA genes have been recognized, including *Fanconi Anemia, Complementation Group D1 (FANCD1* or *BRCA2)*, *Fanconi Anemia, Complementation Group J (FANCJ* or *BRIP1)*, and *Fanconi Anemia, Complementation Group N (FANCN* or *PALB2)*. Furthermore, A-T and Li-Fraumeni syndromes are associated with defects in DNA damage signaling and DSB repair; causative genes for these conditions include *ATM* and *TP53*, respectively. Reviewed in (Ciccia & Elledge, 2010, Jackson & Bartek, 2009)

4 Genetics of cancer

4.1 Hallmarks of cancer and mutation signature

Cancer can be considered a genetic disease. Cancer develops from a single somatic cell that acquires changes in its DNA sequence throughout its lifespan and thereby gains a growth advantage compared to other cells. Six alterations in cancer cell physiology are considered essential for malignant growth, including self-sufficiency in growth signals, insensitivity to growth-inhibitory signals, evasion of programmed cell death, limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis (Hanahan & Weinberg, 2000). Somatic mutations in the cancer cell genome may encompass several distinct classes of DNA sequence changes, including the substitution of one base by another, insertions or deletions of small or large segments of DNA, inter- or intrachromosomal rearrangements, and copy-number changes (Stratton et al, 2009). In solid tumors, such as those derived from the breast, colon, or pancreas, an average of 33 to 65 genes display somatic mutations, and approximately 95% of these mutations are single base substitutions (Vogelstein et al, 2013). Additionally, epigenetic changes, which alter chromatin structure and gene expression, the acquisition of new DNA from exogenous sources (e.g., from viruses) and defects in the mitochondrial genome contribute to cancer development (Stratton et al, 2009).

Somatic mutations in cancer cell can be classified into driver and passenger mutations according to their consequences for cancer development (Stratton et al, 2009). Driver mutations directly or indirectly confer a selective growth advantage to the cell in which they occur. The other mutations that accumulate in the cell but do not confer selective growth advantage are considered passengers. Typically, two to eight driver mutations are necessary for cancer development (Vogelstein et al, 2013). Two main types of genes participate cancer development; tumor suppressor genes and oncogenes. Specifically, cancer-inhibiting tumor suppressor genes are inactivated and cancer-promoting oncogenes are activated by mutations.

4.2 Tumor suppressor genes

Tumor suppressor genes encode proteins whose normal function is to inhibit cell transformation. The proteins participate in a variety of critical and highly conserved cell function, including regulation of the cell cycle and apoptosis, differentiation, surveillance of genomic integrity, repair of DNA errors, signal transduction, and cell adhesion (Oliveira et al, 2005). Tumor suppressor genes can be divided into three classes, caretakers, gatekeepers, and landscapers, based on different properties (Kinzler & Vogelstein, 1997, Kinzler & Vogelstein, 1998). Caretaker genes encode DNA repair proteins that act as caretakers of the genome. The inactivation of caretaker gene results in a greatly increased mutation rate and genomic instability. Gatekeepers prevent cancer through direct control of cell growth. The inactivation of gatekeeper gene contributes directly to the neoplastic growth of the tumor. Landscaper genes encode proteins that affect the microenvironment of the tumor. According to Knudson's two hit hypothesis, two mutational events are required to inactivate the both alleles in tumor suppressor genes (Knudson, 1971). In the dominantly inherited form, one mutation is inherited via germinal cells and the second occurs in somatic cells. In the nonhereditary form, both mutations occur in the somatic cell. One of the most commonly known tumor suppressor genes is *TP53*, which is mutated in more than half of all human cancers (Vousden & Lu, 2002). In BC, *BRCA1* and *BRCA2* are by far the two most commonly known tumor suppressor genes (Miki et al, 1994, Wooster et al, 1994).

4.3 Oncogenes

Oncogenes encode proteins that promote cell proliferation. These genes can be categorized into six groups: transcription factors, chromatin remodelers, growth factors, growth factor receptors, signal transducers, and apoptosis regulators (Croce, 2008). Oncogenes are derived from proto-oncogenes by point mutations, amplifications, or chromosomal rearrangements (Croce, 2008). Compared to tumor suppressor genes, an activating somatic mutation in one allele of an oncogene is generally sufficient to confer a selective growth advantage on the cell (Vogelstein & Kinzler, 2004). One of the most commonly known oncogenes is *V-Myc Avian Myelocytomatosis Viral Oncogene Homolog (MYC)* (Dang, 2010). Moreover, the amplification of the *ERBB2/HER2* oncogene in BC is a well-known biological

marker with therapeutic value. Amplification of this gene is seen in approximately 20% of BCs and is associated with an aggressive disease (Sauter et al, 2009).

5   The genetic predisposition to breast and ovarian cancer: susceptibility genes

The genetic predisposition to breast and ovarian cancer has been well established. Up to 10% of all BCs and up to 15% of all OCs are caused by inherited genetic defects (Claus et al, 1996, Lynch et al, 2009). Less than half of the genetic predisposition to BC has been resolved (Figure 6). BC predisposition factors can be classified into three categories according to the risk associated with the disease: high-risk genes (>10-fold elevated risk), moderate-risk genes (2-4-fold elevated risk) and low-risk genes (<1.5-fold elevated risk) (Turnbull & Rahman, 2008). Rare mutations in two major high-risk genes, *BRCA1* and *BRCA2*, explain the majority 15% of BC familial relative risk (i.e., the ratio of the risk of disease for a relative of an affected individual to that for the general population) (Figure 6). Additionally, rare mutations in high-to-moderate risk genes that are associated with cancer syndromes (e.g., *TP53, PTEN, Liver Kinase B (LKB1)*, and *CDH1*) and in moderate-risk genes (e.g., *CHEK2, ATM*, and *PALB2*) add another 3% and 4% to the BC familial relative risk, respectively (Figure 6). The remaining known genetic predisposition to BC is due to common single nucleotide polymorphisms (SNPs) in low-risk genes, such as *Fibroblast Growth Factor Receptor 2 (FGRF2)*, *Caspase 8, Apoptosis-Related Cysteine Peptidase (CASP8)*, and *TOX High Mobility Group Box Family Member 3 (TOX3)* (Figure 6). In epithelial ovarian cancer, approximately 90% of the genetic predisposition is explained by gene defects in the high-penetrance genes *BRCA1* and *BRCA2*, whereas the remaining 10% of the genetic predisposition is attributable to defects in HNPCC syndrome genes (Prat et al, 2005). Additionally, moderate-to-high risk alleles in genes such as *BRIP1* and *RAD51C* have been determined to contribute to a fraction of OC predisposition (Pelttari et al, 2011, Rafnar et al, 2011)

Locus	SNP	Odds ratio	Locus	SNP	Odds ratio	Locus	SNP	Odds ratio
6q14.1	rs17529111	1.97	8q24	rs13281615	1.09	CDCA7	rs1550623	0.94
*BRCA2	rs11571833	1.44	TOX3	rs3803662	1.09	10q26.12	rs11199914	0.94
CCND1	rs75915166	1.38	CDYL2	rs13329835	1.09	COX11	rs6504950	0.94
ESR1	rs3757318	1.33	1p13.2*	rs11552449	1.08	*SSBP4	rs4808801	0.94
FGFR2	rs2981579	1.33	2q35	rs16857609	1.08	2q35	rs13387042	0.93
CCND1	rs554219	1.33	RANBP1	rs204247	1.08	PDE4D	rs1353747	0.93
FGFR2	rs2981582	1.26	ZMIZ1	rs704010	1.08	11q13.1	rs3903072	0.92
*MERIT40	rs8170	1.26	12q24	rs1292011	1.08	11q24.3	rs11820646	0.92
TERT	rs10069690	1.24	3p26.2	rs6762644	1.07	PAX9	rs2236007	0.92
DNAJC1	rs11814448	1.22	8p21.1	rs9693444	1.07	RAD51L1	rs999737	0.92
CCDN1	rs614367	1.21	8q24.21	rs11780156	1.07	NRIP1	rs2823093	0.92
MKL1	rs6001930	1.21	LSP1	rs3817198	1.07	PEX14	rs616488	0.91
MDM4	rs4245739	1.19	CCDC88C	rs941764	1.07	FOXQ1	rs11242675	0.91
ESR1	rs2046210	1.16	TET2	rs9790517	1.06	8q21.11	rs6472903	0.91
HNF4G	rs2943559	1.14	CDKN2A/B	rs1011970	1.06	RAD51L1	rs2588809	0.91
5p12	rs10941679	1.13	TGFBR2	rs12493607	1.05	LGR6	rs6678914	0.9
12p13.1	rs12422552	1.13	9q31.2	rs10759243	1.05	NTN4	rs17356907	0.9
MAP3K1	rs889312	1.12	DNAJC1	rs7072776	1.05	2q14.2	rs4849887	0.89
TCF7L2	rs7904519	1.12	19q13.31	rs3760982	1.05	9q31	rs865686	0.89
22q12.2	rs132390	1.12	RAB3C	rs10472076	1.04	ZNF365	rs10995190	0.86
2p24.1	rs12710696	1.11	ANKRD16	rs2380205	0.98	ADAM29	rs6828523	0.84
2q31.1	rs2016394	1.1	*CASP8	rs1045485	0.97	MERIT40	rs2363956	0.82
SLC4A7	rs4973768	1.1	7q35	rs720475	0.96	TERT	rs2736108	0.77
EBF1	rs1432679	1.1	CHST9	rs1436904	0.96	PTHLH	rs10771399	0.72
FTO	rs11075995	1.1	FTO	rs17817449	0.95			
1p11.2	rs11249433	1.09	18q11.2	rs527616	0.95	* Denotes coding variant		

Figure 6. Genetic variants that predispose to breast cancer. From (Couch et al, 2014). Reprinted with the permission from AAAS.

5.1 High-risk genes: *BRCA1* and *BRCA2*

5.1.1 *BRCA1*

Gene. The *Breast Cancer 1, Early Onset (BRCA1)* gene is located on the long arm of chromosome 17 at 17q21. It was detected in the early-90s as a strong candidate gene for breast and ovarian cancer susceptibility by linkage and positional cloning methods (Miki et al, 1994). *BRCA1* is a large gene, spanning an 81 kb region of genomic DNA, consisting of 24 exons, and encoding a protein of 1,863 amino acids (Miki et al, 1994, Smith et al, 1996). Exon 1 is non-coding, and exon 11 is the largest exon, encoding over 60% of the BRCA1 protein (Miki et al, 1994, Thakur et al, 1997). The majority of clinically relevant mutations are in exon 11 (National Human Genome Research Institute, Breast Cancer Information Core Database).

Protein structure and function. The two most important regions of the BRCA1 protein are a RING domain in the amino terminus and two BRCT domains in the carboxyl terminus (Figure 7a). Additionally, in the middle of the protein are nuclear localization signals (NLS) and a coiled-coil domain (Figure 7a). BRCA1 has phosphorylation sites for ATM/ATR and CHK2 kinases and has a critical interaction with several proteins, including BRCA1 Associated RING Domain 1 (BARD1), PALB2, Abraxas, Retinoblastoma Binding Protein 8 (CtIP), and BRIP1 (Figure 7a).

BRCA1 has vital role in normal cellular development, and BRCA1 deficiency leads to early embryonic lethality in mice (Hakem et al, 1996, Ludwig et al, 1997). BRCA1's key role is to maintain genomic stability and function as a tumor suppressor (Wang, 2012). This protein acts as central mediator of the cellular response to DNA damage, regulating the activities of multiple repair and checkpoint pathways (Wang, 2012). BRCA1 is a substrate of the central DNA damage response kinases ATM/ATR, which control the DDR (Wang, 2012). The highly conserved zinc-binding RING domain (residues 20-68) has DNA-binding properties and is an interaction site for BRCA1-associated RING domain (BARD1) (Wu et al, 1996). BRCA1 and BARD1 form a heterodimeric RING finger complex that has ubiquitin E3 ligase activity. The enzymatic activity of the BRCA1-BARD1 complex has been suggested to be critical in DNA double strand break repair and the tumor suppression function of BRCA1, but the exact role of this enzymatic activity is not well understood (Hashizume et al, 2001, Morris & Solomon, 2004, Reid et al, 2008). Another protein that binds to the BRCA1 RING domain is BRCA1-associated Protein-1 (BAP1), which is a nuclear-localized deubiquitinating enzyme that has been suggested to have tumor suppressor properties (Eletr & Wilkinson, 2011, Jensen et al, 1998). Two tandem and highly conserved BRCA1 BRCT domains (residues 1699-1736 and 1818-1855) are essential for the tumor suppressor function of this protein (Glover, 2006). These BRCT domains are phospho-peptide-binding domains that recognize a phospho-SPxF motif (S, serine; P, proline; × varies; F, phenylalanine) (Wang, 2012). Similar BRCT domains are present similar in a wide variety of proteins and are involved in cell cycle regulation and DNA repair (Glover, 2006). BRCT-interacting proteins include Abraxas, BRIP1, and CtIP, which directly interact with BRCT domains through their phospho-SPxF motifs in a phosphorylation-dependent manner (Wang, 2012). Additionally, the BRCA1 coiled-coil domain is an interacting site for PALB2 and BRCA2 (Wang, 2012). Moreover, the tumor suppression function of BRCA1 has been shown to occur via heterochromatin-mediated silencing (Zhu et al, 2011).
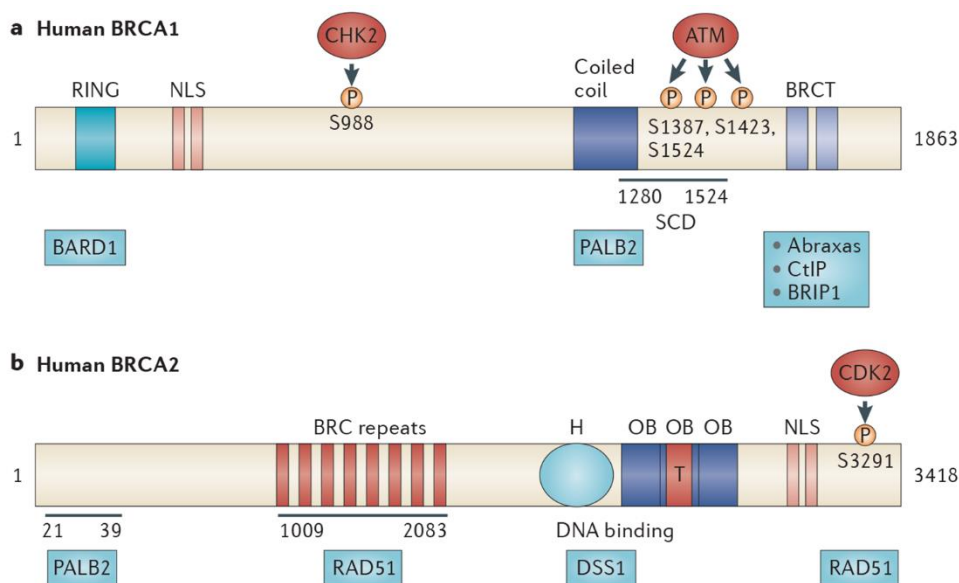
Figure 7. BRCA1 and BRCA2 functional domains and interacting proteins. Adapted and reprinted by permission from Macmillan Publishers Ltd: [NATURE REVIEWS CANCER] (Roy et al, 2011), copyright (2011).

5.1.2 *BRCA2*

Gene*. Breast Cancer 2, Early Onset (BRCA2)* is located on chromosome 13 at 13q12. This gene was detected as another major breast and ovarian cancer susceptibility gene by linkage and positional cloning methods in the 90s, immediately following *BRCA1* identification (Tavtigian et al, 1996, Wooster et al, 1994). *BRCA2* is a large gene, spanning approximately 70 kb of genomic DNA, consisting of 27 exons, and encoding a protein of 3,418 amino acids (Wooster et al, 1994). Similar to *BRCA1*, exon 1 is non-coding and exon 11 is the largest exon, encoding over half of the protein (Tavtigian et al, 1996). The majority of clinically relevant *BRCA2* mutations are located in exon 11 (National Human Genome Research Institute, Breast Cancer Information Core Database).

   Protein structure and function. The BRCA2 protein consists of a DNA binding domain, eight BRC repeats, and nuclear localization signals on the C-terminus (Figure 7b). A critical phosphorylation site for cyclin dependent kinase 2 (CDK) is located on the C-terminus (Figure 7b). BRCA2 has many interaction sites for different proteins, including RAD51 Recombinase (RAD51), PALB2, and

Deleted In Split-Hand/Foot 1 (DSS1); these sites are critical for BRCA2's proper function (Figure 7b).

BRCA2 maintains genome stability and has a key role in the HR pathway in DNA double-strand break repair. The HR pathway is dependent on the recombinase function of RAD51, which is an important interacting partner of BRCA2. BRCA2 binds directly to RAD51 through its C-terminus and conserved BRC repeats. BRCA2 is recruit RAD51 to the nucleus and to the site of DNA damage by utilizing its nuclear localization signals, which are lacking in RAD51. Moreover, BRCA2 regulates RAD51 function during recombination events. The N-terminal portion of BRCA2 interacts with PALB2, physically linking it to BRCA1. The BRCA1-PALB2-BRCA2 complex is needed for the recruitment of RAD51 to the site of DNA damage. DSS1 interacts with DNA-binding domain of BRCA2, and this interaction is critical for the proper functioning of BRCA2. On the C-terminus of BRCA2 is a CDK2 phosphorylation site (S3291), the phosphorylation status of which modulates BRCA2's C-terminal interaction with RAD51. These findings are reviewed in (Gudmundsdottir & Ashworth, 2006, Roy et al, 2011)

5.1.3 Contribution of germline *BRCA1/2* mutations to hereditary breast and ovarian cancer

Monoallelic germline *BRCA1* and *BRCA2* mutations predispose to hereditary breast and ovarian cancers in a highly penetrant manner. Pathogenic *BRCA1/2* mutations appear to account for approximately 30% of cases of these cancers in high-risk families (Couch et al, 2014). Most deleterious mutations are small insertions or deletions that result in the translation of a truncated protein (Mavaddat et al, 2013). The frequency of *BRCA1/2* mutations varies depending on the population in question. The frequencies are highest in founder populations, such as Ashkenazi Jews (Levy-Lahad et al, 1997). According to a study of breast/ovarian cancer patients who were unselected for family history, the average cumulative risk in *BRCA1-* and *BRCA2*-mutation carriers by age 70 years was 65% and 45% for breast cancer and 39% and 11% for ovarian cancer, respectively (Antoniou et al, 2003). The risk estimates for mutation carriers in multiple-case high-risk families are even higher (Antoniou et al, 2003). Moreover, risk estimates vary by age at diagnosis of the index case, the type and site of the cancer, and the site of the mutation (Mavaddat et al, 2013). In addition, several genetic modifiers of *BRCA1* and *BRCA2* contribute to cancer risk in mutation carriers (Antoniou et al, 2008). Biallelic mutations in both of

these genes are rare. Specifically, biallelic *BRCA1* mutations are extremely rare and have been reported only in one early-onset OC patient (Domchek et al, 2013). Moreover, a recent study reported a biallelic *BRCA1* mutation in one early-onset BC patient who presented with a Fanconi anemia-like disorder (Sawyer et al, 2015). Biallelic *BRCA2* mutations have been associated with Fanconi anemia subtypes (Howlett et al, 2002). Additionally, a recent study reported a presence of biallelic *BRCA2* mutations in single family presenting with early-onset colorectal cancer in the absence of Fanconi anemia features (Degrolard-Courcet et al, 2014).

BRCA1- and *BRCA2*-related breast tumors differ from each other and from sporadic cancers. Typically, compared to sporadic tumors, *BRCA1*-related tumors are poorly differentiated ductal carcinomas of higher grade and higher mitotic index. Such tumors also exhibit a greater degree of nuclear polymorphism, less tubule formation, and may exhibit features of medullary carcinoma (Mavaddat et al, 2013). *BRCA1*-related tumors are often negative for estrogen and progesterone receptors and HER2 and are more likely to be p53-positive (Lakhani et al, 2002). Moreover, *BRCA1*-related tumors resemble basal subtype of BC (Sorlie et al, 2003). The specific phenotype for *BRCA2*-related tumors has not been consistently described (Phillips, 2000).

5.1.4   Contribution of germline *BRCA1/2* mutations to other cancers

BRCA1 mutations are associated with an increased prostate cancer risk in male carriers (Leongamornlert et al, 2012). *BRCA2* mutations have been shown to contribute to prostate, pancreatic, gallbladder, bile duct, and stomach cancers, as well as melanoma (Breast Cancer Linkage Consortium, 1999). Both *BRCA1* and *BRCA2* mutations increase the male BC risk, and the risk is higher in *BRCA2* mutation carriers (Tai et al, 2007).

5.1.5   The germline *BRCA1/2* mutation spectrum in Finnish hereditary breast and/or ovarian cancer families

In Finland, germline *BRCA1/2*-mutations contribute approximately 20% to familial breast and ovarian cancer cases (Hartikainen et al, 2007, Vehmanen et al, 1997a). In the Finnish breast and ovarian cancer studies, several *BRCA1* and *BRCA2* deleterious mutations have been reported, and total of fourteen mutations, seven in each gene, are considered founder mutations. These mutations are designated as

such given that they are recurrent and account for the majority of all detected *BRCA1/2* mutation in Finland (Hartikainen et al, 2007, Huusko et al, 1998, Sarantaus et al, 2000, Vehmanen et al, 1997a, Vehmanen et al, 1997b). The most frequently observed Finnish founder mutations include 3604delA, 3744delT/3745delT, 4216-2ntA>G, 4446C>T, 5370C>T in *BRCA1* and 999del5, 7708C>T, 8555T>G, and 9346-2ntA>T in *BRCA2* (Sarantaus et al, 2000). There are geographical differences in the distribution of founder mutations across Finland. For example, in the Northern and Eastern part of Finland, the mutation spectrum is scarce and specific, whereas in Southern Finland, almost all founder mutations have been detected (Hartikainen et al, 2007, Huusko et al, 1998, Sarantaus et al, 2000). Moreover, one mutation, 4088insA in *BRCA2*, has been detected uniquely in the Eastern part of Finland (Hartikainen et al, 2007). Large genomic rearrangements in *BRCA1/2* are rare and only one study has reported a large *BRCA1* deletion in a Finnish family with a history of OC (Pylkas et al, 2008).

5.2 High-to-moderate-risk genes involved in cancer syndromes

5.2.1 *TP53* and Li-Fraumeni syndrome

Tumor protein p53 (TP53) gene is located on chromosome 17 at 17p13. This gene spans approximately 20 kb of genomic DNA and consists of 11 exons, encoding a 393-amino acid protein (i.e., p53) (McBride et al, 1986). Protein p53 is a tumor suppressor protein whose main function is to maintain cell integrity by controlling cell survival, proliferation, and death (Vousden & Prives, 2009). Protein p53 is activated by diverse cell stress signals, such as DNA damage, oxidative stress, hyperproliferative signals and hypoxia (Bieging et al, 2014). In its active form, p53 triggers the transcription of various target genes that further induce responses such as cell cycle arrest, DNA repair, senescence, and apoptosis. These effects ultimately either promote the repair and survival or the permanent removal of damaged cells (Vousden & Prives, 2009). In addition, p53 regulates processes such as cellular metabolism, stem cell function, invasion and metastasis, and cell-cell communication within the tumor microenvironment; these effects may also contribute to tumor suppression (Bieging et al, 2014). Protein p53 is inactive in more than half of all human cancers due either to mutations in the *TP53* gene or defects in signaling

pathways upstream or downstream of p53 (Vousden & Lu, 2002). Over 80% of *TP53* mutations in human tumors localize to the central region of the gene, which encodes a highly conserved, sequence-specific DNA-binding domain (exons 5-8) (Bieging et al, 2014, Vousden & Lu, 2002).

Germline mutations in *TP53* contribute to Li-Fraumeni Syndrome (LFS) and its variant, Li-Fraumeni Like-Syndrome (LFL). Both of these conditions are rare autosomal cancer predisposition syndromes characterized by familial clustering of a wide spectrum of tumors in children and young adults. These tumors are predominantly breast tumors, bone and soft-tissue sarcomas, brain tumors, leukemia, and adrenocortical tumors (Birch et al, 1994, Malkin et al, 1990). LFL resembles LSF but is defined by less stringent classification criteria (Birch et al, 1994, Eeles, 1995). Germline *TP53* mutations are observed in 71-77% and 22-40% of LFS and LFL families, respectively (Varley et al, 1997, Varley, 2003). In LFS families, nearly half of the individuals with germline *TP53* mutations develop cancer by the age of 30 years, and the lifetime cancer risk is 73% in males and almost 100% in females. This latter high-penetrance is primarily due to early-onset BC (McBride et al, 2014). BC is the most common cancer type among female *TP53* mutation carriers in LFS/LFL families, accounting for 27% of all cancers (McBride et al, 2014). Moreover, germline *TP53* mutations have been reported to contribute up to 8% of early-onset BC cases with family history of the disease (Lalloo et al, 2006, Lee et al, 2012, Mouchawar et al, 2010).

In Finland, germline *TP53* mutations are very rare in the general BC population but explain a small fraction of the hereditary BC cases (Rapakko et al, 2001). Pathogenic *TP53* mutations occurring in the conserved region of *TP53* have been reported in 3/108 (2.8%) Finnish *BRCA1/2*-negative hereditary BC families that also exhibit features of either LFS or LFL (Huusko et al, 1999, Rapakko et al, 2001).

5.2.2   *ATM* and Ataxia-telangiectasia

Ataxia Telangiectasia Mutated (ATM) gene is located at locus 11q22-23 and consists of 66 exons, encoding a large protein of 3,056 amino acids (Gatti et al, 1988, Savitsky et al, 1995, Uziel et al, 1996). The encoded protein belongs to a family of PI3K-related kinases that are involved in various cellular responses to genotoxic stress. These functions are carried out via phosphorylation of key proteins in cellular response pathways (Shiloh & Ziv, 2013). The protein kinase ATM has a central role as an activator of the DNA damage response cascade after DNA double-strand

breaks (Shiloh & Ziv, 2013). ATM exists as a homodimer that dissociates into active monomers upon activation, which occurs via autophosphorylation at Ser1981 after DNA damage (Shiloh & Ziv, 2013). In its active form, ATM is recruited to the DNA damage site, where it activates a signaling cascade that includes direct and indirect phosphorylation events of a large number of proteins. These phosphorylation events lead to the activation of cell-cycle checkpoints and the initiation of DNA repair (Shiloh & Ziv, 2013). Downstream targets of ATM include also proteins that are encoded by the known BC susceptibility genes *TP53*, *BRCA1*, and *CHEK2* (Ahmed & Rahman, 2006).

Homozygous germline defects or compound heterozygosity in *ATM* are a cause of Ataxia-telangiectasia (A-T), a rare autosomal recessive disorder characterized by cerebellar degeneration, immunodeficiency, chromosomal instability, cancer predisposition, radiation sensitivity, and cell cycle abnormalities (Savitsky et al, 1995, Swift et al, 1986). Moreover, heterozygous A-T-causing germline mutations have been shown to increase BC risk from 2-to-9-fold in female relatives of A-T patients and in hereditary BC patients without a family history of A-T in different populations (Broeks et al, 2000, Olsen et al, 2005, Renwick et al, 2006, Swift et al, 1987, Thompson et al, 2005, Thorstenson et al, 2003). However, despite the central role of ATM in DNA repair pathway, the contribution of *ATM* mutations to the risk of other cancers is not clearly understood (Thompson et al, 2005). Although *ATM* has been considered a moderate penetrance BC gene, various studies have been inconsistent regarding the degree of BC risk related to heterozygous *ATM* variants (Ahmed & Rahman, 2006, Prokopcova et al, 2007).

In Finland, the contribution of germline *ATM* variants to BC predisposition has been studied in different geographical BC cohorts. A common heterozygous *ATM* polymorphism, 5557G>A with IVS38-8T>C, was associated with bilateral BC (odds ratio (OR) = 10.2; 95% confidence interval (CI) = 3.1-33.8; p = 0.001) in a cohort of 185 breast or breast-ovarian cancer patients from 121 Northern Finnish families (Heikkinen et al, 2005). However, the findings were not supported by further case-control association analysis utilizing an extensive cohort of Southern Finnish familial and unselected BC patients (Tommiska et al, 2006a). Moreover, two pathogenic mutations, 7570G>C and 6903insA, that were originally identified in Finnish A-T families were observed in 3/162 (1.9%) BC families of Central and Northern Finnish origin (Allinen et al, 2002). The apparent yet overall limited role of heterozygous *ATM* variants in hereditary BC predisposition was confirmed by the identification of two pathogenic mutations, 7570G>C and 6903insA, and one A-T-causative mutation (8734A>G) in 6/541 (1.1%) familial (*P* = 0.006, OR 12.4, 95% CI 1.5-

103.3) and 7/1124 (0.6%) unselected BC cases of Finnish origin ($P = 0.07$, OR 6.9, 95% CI 0.9-56.4); in population controls, the prevalence of these mutations was 1/1107 (0.1%) (Pylkas et al, 2007).

5.2.3 *PTEN* and Cowden syndrome

Phosphatase And Tensin Homolog gene *(PTEN)* is located on the long arm of chromosome 10 at 10q23 and encodes a 403-amino acid protein (Li et al, 1997). PTEN is a dual specificity protein and lipid phosphatase that blocks the PI3K/Akt signaling pathway and thereby inhibits cell survival, growth, and proliferation (Hopkins et al, 2014). Moreover, PTEN is known to act in a PI3K-independent manner, inhibiting migration and affecting genomic stability, gene expression, and the cell cycle (Hopkins et al, 2014).

   PTEN is frequently altered in a variety of human cancers, including the brain, breast, and prostate. To date, more than 2700 *PTEN* mutations have been identified in different tumor types (Hopkins et al, 2014). *PTEN* has been detected as a causative gene for Cowden syndrome, a rare autosomal dominant familial cancer syndrome with a high risk of breast and thyroid cancer and the presence of benign hamartomas in several tissue, including the skin, breast, thyroid, oral mucosa, and intestinal epithelium (Hopkins et al, 2014, Nelen et al, 1996, Nelen et al, 1997). The lifetime risk for female BC among *PTEN* mutation carriers is as high as 85%, even higher than the BC risk estimates for *BRCA1/2*-mutation carriers (Antoniou et al, 2003, Tan et al, 2012). Germline *PTEN* mutations have been reported to be rare in high-risk non-*BRCA1/2* BC families (Blanco et al, 2010, Guenard et al, 2007). However, a Finnish hereditary BC study reported that germline *PTEN* promoter region variants affect tumor progression and the gene expression profile in BC (Heikkinen et al, 2011).

5.2.4 *STK11* and Peutz-Jeghers syndrome

Serine/Threonine Kinase 11 gene (*STK11*, also known as *Liver Kinase B1, LKB1*) is located on the telomeric end of chromosome 19 at 19p13 and encodes a serine/threonine kinase (STK11) of 433 amino acids (Hemminki et al, 1998). STK11 is tumor suppressor protein involved in several cellular responses, such as energy metabolism, cellular polarity and cellular growth. STK11 regulates these processes primarily via AMPK/mTOR signaling (Hemminki et al, 1998, Korsse et al, 2013).

Moreover, STK11 is involved in cell cycle regulation and apoptosis, and one of its interacting partners is the tumor suppressor protein p53 (Korsse et al, 2013).

STK11 is a causative gene for Peutz-Jeghers Syndrome (PJS), a rare autosomal dominant disorder characterized by hamartomatous polyps of the gastrointestinal tract, mucocutaneous melanin deposits, and an increased risk of benign and malignant neoplasms of the intestinal tract, pancreas, ovaries, testis, breast, and uterus (Giardiello et al, 1987, Giardiello & Trimbath, 2006, Hemminki et al, 1998, Jenne et al, 1998). Patients with PJS have a cumulative risk of 37-93% for all cancers and of 32-54% for BC by the age of 70 (van Lier et al, 2010). The BC risk among PJS patients is comparable to the BC risk associated with germline *BRCA1/2* mutations (Antoniou et al, 2003). Germline *STK11* mutations have been reported to be rare in high-risk non-*BRCA1/2* BC families (Guenard et al, 2010).

5.2.5 *CDH1* and hereditary diffuse gastric cancer syndrome

Cadherin 1, type 1, E-cadherin (epithelial) (CDH1) gene is located on chromosome 16 at 16q22 and encodes a calcium-dependent cell-cell adhesion molecule of approximately 728 amino acids (Berx et al, 1995). CDH1 is a transmembrane glycoprotein that plays an important role in the formation and maintenance of the normal architecture and function of epithelial cells (Berx et al, 1998). The adhesive function of CDH1 is dependent on the interaction of its cytoplasmic domain with catenin proteins, including β-catenin, which is a proto-oncogene that plays an important role in Wnt-signaling (Behrens, 1999, Berx et al, 1998). The CDH1 protein is a well-known growth and invasion suppressor in epithelial cells and acts through complex mechanism, including the promotion of tissue organization and inhibition of apoptosis (van Roy, 2014). The loss of CDH1 expression and/or its abnormal function play an important role in human cancer development (Paredes et al, 2012).

Germline mutations in *CDH1* have been identified as the genetic cause of hereditary diffuse gastric cancer (HDGC) syndrome, which is characterized by 1) the familial aggregation of diffuse gastric cancer cases among first or second degree relatives and 2) an increased risk of other cancers, including lobular breast, colon, prostate, and ovarian cancer (Fitzgerald et al, 2010, Pinheiro et al, 2014). Approximately 25-30% of families fulfilling the HDGC criteria harbor inactivating mutations in the *CDH1* gene region (Fitzgerald et al, 2010). Germline *CDH1* mutations are of high-penetrance, conferring high lifetime risks of developing gastric and lobular BCs (over 80% and 60% by 80 years of age, respectively) (Fitzgerald et

al, 2010). However, germline *CDH1* mutations are generally rare in lobular BC patients without a family history of HDGC (McVeigh et al, 2014, Schrader et al, 2011, Xie et al, 2011).

In Finland, germline *CDH1* alterations have been studied only in Finnish prostate cancer patients who also presented with gastric cancer in their families. One *CDH1* missense mutation has been identified to have a higher frequency in both familial and unselected prostate cancer cases than in controls in a large-scale population-based survey (Ikonen et al, 2001).

5.3 Moderate-risk genes

5.3.1 *CHEK2*

The *Checkpoint Kinase 2 (CHEK2)* gene is located on the long arm of chromosome 22 at 22q12 and encodes a serine/threonine-protein kinase that functions as a key mediator of the DNA damage response, cell cycle control and DNA repair (Bartek et al, 2001). The CHEK2 protein is activated via phosphorylation by the upstream kinase ATM in response to DNA double strand breaks (Bartek et al, 2001). After activation, CHEK2 kinase phosphorylates downstream effectors, including the tumor suppressor proteins BRCA1 and p53, the transcription factor E2F1, the PI3K, and the phosphatases Cell Division Cycle 25A (Cdc25A) and Cell Division Cycle 25C (Cdc25C). These effects thereby activate different cellular responses, such as damage-induced transcription, DNA repair, cell cycle arrest/delay, and apoptosis (Bartek et al, 2001, Bartek & Lukas, 2003). The *CHEK2* gene consists of 14 exons, and protein structure shows three characteristic domains: 1) an amino-terminal SQ/TQ regulatory domain with putative phosphorylation sites by ATM kinase, 2) a forkhead-associated domain for binding to other phosphorylated proteins, and 3) a carboxyl terminal serine/threonine kinase domain that has activation loop structure and shows structural homology with other serine/threonine kinases (Bartek et al, 2001, Nevanlinna & Bartek, 2006) Defects in the CHEK2 protein contribute to the development both hereditary and sporadic human cancers, and its role as a candidate tumor suppressor has been suggested (Bartek et al, 2001).

Germline *CHEK2* mutations were originally identified in Li-Fraumeni Syndrome, which is a highly penetrant familial cancer phenotype usually associated with inherited mutations in *TP53* (Bell et al, 1999). However, germline *CHEK2* variants

were subsequently identified in familial and unselected BC patients without Li-Fraumeni or Li-Fraumeni-Like Syndrome, as well as in healthy population controls. These results suggest that *CHEK2* is a low-to-intermediate penetrance BC susceptibility gene (Vahteristo et al, 2002). Two germline mutations are the most studied *CHEK2* variants with respect to hereditary BC and other cancers in different populations. These mutations include 1) a c.1100delC deletion in the kinase domain in exon 10, which results in a truncated protein; and 2) a missense mutation (c.470T>C (I157T)) in the forkhead-associated domain in exon 3, which disrupts critical interactions with its binding partners, such as BRCA1 (Nevanlinna & Bartek, 2006). The c.1100delC variant results in an approximately two-to-three-fold increase in the risk of BC in women without *BRCA1* and *BRCA2* mutations (Meijers-Heijboer et al, 2002). Similar observations for the c.1100delC variant were reported in Finnish *BRCA1/2*-negative hereditary BC patients, with observed frequencies 5.5% (28/507) in cases and 1.4% (26/1,885) in controls (Vahteristo et al, 2002). Additionally, a large international collaborative study by the CHEK2 Breast Cancer Consortium found the c.1100delC variant in 1.9% (201/10,860) of BC cases and of 0.7% (64/9,065) controls (estimated odds ratio 2.34; 95% CI 1.72-3.20). These data confirm that c.1100delC variant confers an increased risk (approximately two-fold) of BC, and this risk is apparent in women unselected for family history (CHEK2 Breast Cancer Case-Control Consortium, 2004). Moreover, trends have been observed towards an early age of onset and a bilateral form of BC among c.1100delC variant carriers (CHEK2 Breast Cancer Case-Control Consortium, 2004, Oldenburg et al, 2003, Vahteristo et al, 2002). Additionally, a higher frequency of c.1100delC variant carriers has been reported in cases with an affected first-degree relative (Vahteristo et al, 2002). As the c.1100delC is a low-to moderate risk variant, it has been suggested that it acts as a modifier in combination with other gene variants, thereby multiplying the BC risk in *BRCA1/2*-negative families (Antoniou et al, 2002, Oldenburg et al, 2003). Compared to c.1100delC, I157T is a lower risk variant and has been observed in different populations with varying frequencies. The highest frequencies have been observed in Finnish (7.4%), Polish (6.7%), German (2.2%) and Byelorussian (5.7%) populations (Bogdanova et al, 2005, Cybulski et al, 2004, Kilpivaara et al, 2004). A recent meta-analysis confirmed the I157T variant to be associated with a 1.5-fold increased BC risk both in familial and unselected BC cases and a 4-fold increased BC risk in lobular BC cases (Liu et al, 2012).

Moreover, *CHEK2* variants increase the risk of other cancers, including male breast, colon, prostate, thyroid, and kidney cancers (Cybulski et al, 2004, Meijers-Heijboer et al, 2002). The contribution of the *CHEK2* variants to hereditary prostate

and sporadic and familial colorectal cancers has also been observed in the Finnish population (Kilpivaara et al, 2003, Kilpivaara et al, 2006, Seppala et al, 2003).

5.3.2 *PALB2*

The *Partner and localizer of BRCA2 (PALB2)* gene is located on the short arm of chromosome 16 at 16p12 and consists of 13 exons that encode a 1186-residue polypeptide. PALB2 binds to the N-terminus of BRCA2 and ensures its proper function in DNA double-strand break repair (Xia et al, 2006). The interaction of PALB2 with BRCA2 occurs via its WD40 domain in the C-terminus (Xia et al, 2006). PALB2 colocalizes with BRCA2 in nuclear foci, promotes its localization and stability in nuclear structures. In addition, PALB2 enables BRCA2's recombinational repair and checkpoint functions, as well as its tumor suppression activity (Rahman et al, 2007, Xia et al, 2006). Moreover, PALB2 has an important role as a BRCA1-interacting partner, being part of the BRCA1-PALB2-BRCA2-complex, which plays a critical role in homologous recombination repair (Sy et al, 2009, Zhang et al, 2009). The interaction of PALB2 with BRCA1 occurs through the coiled-coil domain in the N-terminus of PALB2 (Sy et al, 2009). Moreover, PALB2 directly interacts through its WD40 domain with several other proteins that function in both cellular responses to DNA damage and homologous recombination DNA repair. These PALB2-interacting proteins include RAD51, RAD51C, X-Ray Repair Complementing Defective Repair In Chinese Hamster Cells 3 (XRCC3), and polη (Park et al, 2014c). The central portion of PALB2 interacts with Mortality Factor 4 Like 1 (MRG15), a component of histone modifying complexes; this interaction is important in regulating homologous recombination (Park et al, 2014c). Furthermore, PALB2 interacts with Kelch-Like ECH-Associated Protein 1 (KEAP1), a regulator of the response to oxidative stress, and plays a role in cellular redox homeostasis (Park et al, 2014c). PALB2 also functionally interacts with two tumor suppressors, Receptor Associated Protein 80 (RAP80) and Abraxas (Park et al, 2014c).

Biallelic germline defects in *PALB2* gene predispose to Fanconi anemia type N, a subtype of a rare, recessive chromosomal instability disorder characterized by growth retardation, congenital malformations, progressive bone marrow failure, cancer predisposition, and cellular DNA hypersensitivity to cross-linking agents (Reid et al, 2007). The Fanconi anemia type N subtype is similar to the D1 subtype, which is caused by biallelic mutations in *BRCA2*, and a common feature of both subtypes is a high risk of childhood cancer (Reid et al, 2007). Furthermore,

monoallelic germline loss-of-function (LOF) mutations in *PALB2* have been associated with an increased risk of BC, and *PALB2* is considered a strong BC susceptibility gene (Antoniou et al, 2014, Erkko et al, 2007, Rahman et al, 2007, Southey et al, 2010, Tischkowitz et al, 2012). In contrast, the contribution of other types of *PALB2* mutations, such as missense variants, to BC risk is uncertain (Tischkowitz et al, 2012). Germline *PALB2* LOF mutations are rare but have been identified in many populations, and the carrier frequencies vary from 0.1% to 3.0% in familial BC cases (Southey et al, 2013). Population-specific *PALB2* LOF founder mutations have been reported in Finland, Poland, Canada, UK, Australia, and USA, and these recurrent mutations show high enrichment in familial BC cases compared to healthy controls (Southey et al, 2013). For example, a Finnish founder mutation, c.1592delT, was reported in 6/2,501 (0.2%) healthy controls but in 3/313 (2.7%) Northern Finnish hereditary BC patients and in 18/1,918 (0.9%) BC patients unselected for family history ($P = 0.005$; OR 11.3; 95% CI 1.8–57.8 and $P = 0.003$, OR 3.94, 95% CI 1.5–12.1, respectively) (Erkko et al, 2007). Another Finnish study using a cohort of Southern Finnish BC patients identified the c.1592delT mutation in 2/1,079 (0.2%) healthy controls, 19/947 (2.0%) familial BC cases, and 8/1,274 (0.6%) sporadic BC cases ($P<0.0001$; OR 11.03; 95% CI 2.65-97.78 and $P=0.1207$; OR 3.40; 95% CI 0.68-32.95, respectively) (Heikkinen et al, 2009). The mean lifetime risk of BC among females with germline *PALB2* LOF mutation has been estimated to be 35% (95% CI, 26 to 46) by 70 years of age (Antoniou et al, 2014). The BC risk in *PALB2* mutation carrier is high for cases with a family history of BC, being 58% (95% CI, 50 to 66) compared to carriers with no affected relatives, for whom the rate is 33% (95% CI, 25 to 44) (Antoniou et al, 2014). For the Finnish founder mutation, c.1592delT, the BC risk has been estimated to be 40% (95% CI, 17-77) by 70 years of age (Erkko et al, 2008). As the BC risk associated with germline *PALB2* LOF mutations in the familial setting is comparable to germline *BRCA2* mutations (Antoniou et al, 2003), the clinical relevance and usefulness of *PALB2* mutation screening in hereditary BC families negative for *BRCA1/2* mutations has been suggested by a growing number of studies (Antoniou et al, 2014, Casadei et al, 2011, Haanpaa et al, 2013, Janatova et al, 2013, Southey et al, 2013, Teo et al, 2013).

Additionally, the contribution of *PALB2* variants to the risk for several other cancers has been studied, with the results indicating an increased risk of male BC and familial pancreatic cancer (Jones et al, 2009, Vietri et al, 2015); in contrast, there is little or no evidence showing an association with an increased risk of hereditary prostate (Erkko et al, 2007, Tischkowitz et al, 2008) or ovarian cancer (Dansonka-Mieszkowska et al, 2010, Prokofyeva et al, 2012).

5.3.3 *BRIP1*

The *BRCA1-interacting protein C-terminal helicase 1 (BRIP1,* also known as *BACH1)* gene is located on chromosome 17 at 17q22 and encodes a 1249-amino acid protein. BRIP1 is a member of the DEAH helicase family and is a binding partner of BRCA1 (Cantor et al, 2001). BRIP1 binds directly to the BRCT repeats of BRCA1 and contributes its DNA repair and tumor suppressor functions and checkpoint control (Cantor et al, 2001, Yu et al, 2003).

Biallelic germline mutations in *BRIP1* predispose to Fanconi anemia subtype J (Levitus et al, 2005), whereas rare monoallelic germline *BRIP1* mutations have been reported to confer susceptibility to BC (Cantor et al, 2004, Seal et al, 2006). A British study identified truncating germline *BRIP1* mutations among BC patients from *BRCA1/2*-negative families, and the estimated the relative BC risk is approximately twofold (Seal et al, 2006). Additionally, germline missense variants in *BRIP1* have been reported in *BRCA1/2*-negative breast or ovarian cancer patients from Jewish high cancer risk families (Catucci et al, 2012). In addition to rare variants, common polymorphisms in *BRIP1* have been reported to contribute to the BC risk, but the results are inconclusive (Pabalan et al, 2013, Ren et al, 2013, Sigurdson et al, 2004).

In Finland, *BRIP1* mutation screening analyses in the Finnish breast/ovarian cancer families have shown that germline mutations are very rare and their contribution to familial BC seems marginal (Karppinen et al, 2003, Solyom et al, 2010, Vahteristo et al, 2006). Moreover, germline *BRIP1* mutations have been observed to confer a high risk of OC in Iceland and Spain, and *BRIP1* has been identified as a tumor suppressor gene for OC (Rafnar et al, 2011). Additionally, *BRIP1* gene polymorphisms have been reported to contribute to an individual's predisposition to cervical cancer in the Chinese population (Ma et al, 2013).

5.3.4 *RAD50*

The *RAD50 Homolog (S. Cerevisiae)* gene is located on chromosome 5 at 5q31 and encodes a protein of 1,312 amino acids (Dolganov et al, 1996). RAD50 is part of RAD50-MRE11-NBS1 (MRN, also known as MRE11) protein complex, which plays an important role in maintaining genomic integrity. This complex acts by governing the activation of the central transducing kinase ATM, thereby enabling the detection of DNA double-strand breaks and controlling the DNA damage response (Stracker & Petrini, 2011). Moreover, the MRN complex has a role in DNA double-strand break metabolism, telomere homeostasis, meiosis, apoptosis, and

immune system development (Stracker & Petrini, 2011). The RAD50 protein shows both sequence and structural homology to structural maintenance of chromosomes family members, which control chromatin structure and dynamics (Lamarche et al, 2010). Similarly to other structural maintenance of chromosomes proteins, RAD50 contains N-terminal Walker A and C-terminal Walker B nucleotide-binding motifs that associate with one another. This association forms a bipartite ATP (adenosine triphosphate)-binding cassette (ABC)-type ATPase domain, which binds and unwinds double-stranded DNA termini (Lamarche et al, 2010). The amino acids between the N- and C-termini form an anti-parallel coiled-coil that terminates with a zinc-hook (CxxC) domain (Lamarche et al, 2010). Both the coiled-coil and hook domains of RAD50 are important for MRN complex function (Lamarche et al, 2010, Stracker & Petrini, 2011). In the MRN complex, RAD50 directly interacts with MRE11 through the coiled-coil region, whereas the interaction with NBS1 is mediated through MRE11 (Lamarche et al, 2010). Moreover, MRE11 possesses both endonuclease and exonuclease activity and is primarily responsible, together with RAD50, for DNA binding (Stracker & Petrini, 2011). NBS1 stimulates the DNA binding and nuclease activities of RAD50 and MRE11 and mediates protein-protein interactions (e.g., with ATM) at DNA breakage sites (Lamarche et al, 2010). In addition, the MRN complex functions as a part of multi-subunit genomic surveillance complex together with BRCA1 and CtIP (Wang, 2012).

Homozygous germline mutations in any of the MRN complex genes predispose to genomic instability disorders and severe cancer phenotypes. Biallelic mutations in *NBS1* are associated with a rare recessive autosomal Nijmegen breakage syndrome characterized by microcephaly, growth retardation, immunodeficiency, radiosensitivity, and cancer predisposition (Digweed & Sperling, 2004). Similarly, the phenotype associated with biallelic *RAD50* mutations resembles the Nijmegen breakage syndrome disorder (Waltes et al, 2009). Moreover, biallelic mutations in *MRE11* are associated with ataxia-telangiectasia-like disorder (Stewart et al, 1999). Additionally, mouse studies have shown that homozygous MRN gene knockouts are lethal (Stracker & Petrini, 2011). Heterozygous mutations in MRN genes have been reported to predispose to common types of cancer, including gastrointestinal, prostate, and breast cancer (Ebi et al, 2007, Heikkinen et al, 2006, Zuhlke et al, 2012). Specifically, truncating variants and missense substitutions that occur in key functional domains in the MRN genes have been associated with an intermediate (two- to four-fold) BC risk in certain populations (Damiola et al, 2014, Heikkinen et al, 2006, Zhang et al, 2012).

In Finland, two pathogenic *RAD50* mutations, c.687delT and IVS3-1G>A, have been observed among Northern Finnish BC patients unselected for family history (Heikkinen et al, 2006). The c.687delT mutation results in a stop codon at 234 and is a relatively common low-risk Finnish founder mutation. This mutation has been identified in 8/317 (2.5%) BC patients compared to 6/1000 (0.6%) healthy controls ($P = 0.008$, OR 4.3, 95% CI 1.5–12.5). The IVS3-1G>A mutation has been observed in 1/317 (0.3%) BC patients but in none of the 1000 healthy controls (Heikkinen et al, 2006). Additionally, the c.687delT has been identified in 3/590 (0.5%) familial BC patients of Southern Finnish origin compared to 1/560 (0.2%) healthy controls (Tommiska et al, 2006b).

5.3.5   *RAD51C*

The *RAD51 recombinase C (RAD51C)* gene, one of five *RAD51* paralogs, is located on chromosome 17 at 17q23 and encodes a 376-amino-acid protein that is involved in the recombinational repair of DNA damage (Dosanjh et al, 1998, Vaz et al, 2010). RAD51C interacts with other RAD51 paralogs to form two complexes. The first of these complexes consists of RAD51 Paralog B (RAD51B), RAD51C, RAD51 Paralog D (RAD51D), and X-Ray Repair Complementing Defective Repair In Chinese Hamster Cells 2 (XRCC2); the second consists of RAD51C and XRCC3. Both complexes work at different stages of BRCA2-RAD51-dependent homologous recombination (Chun et al, 2013). RAD51C plays a role in recruiting RAD51 to sites of DNA damage, regulates the resolution of recombination intermediates, and is required for *CHEK2* activation and checkpoint function (Somyajit et al, 2010). Moreover, RAD51C interacts with the RAD18 E3 Ubiquitin Protein Ligase (RAD18), which has a key role in transmitting the DNA damage signal to elicit homologous recombination repair (Huang et al, 2009). It has recently been shown that RAD51C interacts with BRCA2 and PALB2 to form a functionally important complex in homologous recombination (Park et al, 2014b).

Originally, biallelic germline *RAD51C* mutations were associated with a Fanconi anemia-like disorder (Vaz et al, 2010). Subsequently, rare pathogenic monoallelic germline *RAD51C* mutations were identified in German families with both breast and ovarian cancer but not BC alone. These results indicated the involvement of the *RAD51C* variants in the predisposition to breast and ovarian cancer in a high-penetrance manner (Meindl et al, 2010). Further studies have revealed that *RAD51C* is actually a moderate-to-high risk ovarian cancer susceptibility gene but that

mutations do not affect or only slightly increase BC risk (Loveday et al, 2012, Pelttari et al, 2011, Thompson et al, 2012a). However, recent studies in Spanish and Pakistani populations have supported *RAD51C*'s role as a breast and ovarian cancer susceptibility gene. These reports identified germline *RAD51C* mutations in *BRCA1/2*-negative families with breast and ovarian cancer and BC only (Blanco et al, 2014, Rashid et al, 2014).

In Finland, two recurrent deleterious *RAD51C* mutations, c.93delG and c.837+1G>A, were observed in 4/277 (1.4%) Southern Finnish breast or ovarian cancer families. Further screening of these two mutations in a series of breast or ovarian cancer patient cohorts from the Helsinki and Tampere regions revealed an association of these mutations with a 14-, 213- and 6-fold increased risk of familial breast and ovarian cancer, familial OC in the absence of BC, and with unselected OC, respectively (Pelttari et al, 2011). Another Finnish study confirmed the contribution of germline *RAD51C* mutations to hereditary breast and ovarian cancer susceptibility by identifying a rare novel pathogenic variant, c.-13_14del27, in 1/147 (0.7%) Northern Finnish breast and ovarian cancer families; in contrast, the variant was absent in 990 unselected BC patients and in 852 healthy controls (Vuorela et al, 2011). Germline *RAD51C* variants have not been implicated in an increased risk for other cancers, such as prostate or colorectal cancer in the Finnish population (Pelttari et al, 2012).

5.3.6 *FAM175A*

The *Family With Sequence Similarity 175, Member A (FAM175A)* gene (also known as *ABRA1, CCDC98* and *Abraxas*) is located at genomic position 4q21 and encodes a protein of 409 amino acids (Liu et al, 2007, Wang et al, 2007a). The Abraxas protein interacts with BRCA1 and is required for genomic stability and tumor suppression (Castillo et al, 2014). Abraxas binds directly to BRCA1 BRCT repeats through its phosphorylated SPxF motif at the C-terminus, forming part of the BRCA1-A complex (Kim et al, 2007, Liu et al, 2007, Wang et al, 2007a, Wang, 2012). The BRCA1-A complex also includes four other proteins, RAP80, BRISC And BRCA1 Complex Member 1 (BABAM1), Brain And Reproductive Organ-Expressed (TNFRSF1A Modulator) (BRE), and BRCA1/BRCA2-Containing Complex, Subunit 3 (BRCC36) (Wang, 2012). Abraxas interacts with RAP80, BABAM1, and BRE through its MPN domain at the N-terminus and with BRCC36 through its coiled-coil domain located in the middle of the protein (Castillo et al, 2014). Abraxas

is a central organizer adaptor protein in the BRCA1-A complex, mediating interactions between BRCA1 and other complex proteins and bridging interactions between each member of the complex (Wang, 2012). The main function of the BRCA1-A complex is to mediate DNA damage-induced ubiquitin signaling for recruitment of BRCA1 to the site of DNA double-strand breaks (Wang, 2012). The down-regulation any component of the BRCA1-A complex leads to increased cell sensitivity to ionizing radiation and an inability of cells to both arrest the cell cycle and mediate homologous recombination-dependent repair in response to DNA-damage (Wang, 2012). Defects in BRCA1-A complex proteins have been reported to contribute to breast tumor development (Wang, 2012).

Abraxas has been shown to be essential for DNA repair and tumor suppression in an *in vivo* mouse model (Castillo et al, 2014). Moreover, reduced expression, gene copy loss, and mutations in Abraxas are observed in multiple human tumors, including breast and ovarian cancer (Castillo et al, 2014). A germline *Abraxas* mutation that results in abrogated protein nuclear localization and DNA damage response functions has been reported in 2.4% (3/125) of Northern Finnish BC families and in 0.1% (1/991) BC patients unselected for family history; in contrast, this mutation was found in 0% (0/868) of healthy female controls (Solyom et al, 2012).

5.3.7   *FANCM*

The *Fanconi Anemia Complementation Group M (FANCM)* gene is located at 14q21 and encodes a 2,048-residue protein. FANCM belongs to a group of Fanconi anemia (FA) proteins, the main role of which is to repair DNA interstrand crosslinks (Deans & West, 2011). Currently, 17 genes encoding FA proteins have been identified. Many of these genes, including *FANCD1 (BRCA2), FANCJ (BRIP1), FANCN (PALB2),* and *Fanconi Anemia, Complementation Group O (FANCO* or *RAD51C)*, are known breast/ovarian cancer susceptibility genes (Schneider et al, 2015). FANCM plays a particularly central role in the FA network. FANCM recognizes interstrand crosslinks and recruits the FA core complex to the site of damage, ultimately leading to damage signaling, the recruitment of repair proteins, and checkpoint activation (Deans & West, 2011). In these crucial processes, FANCM has critical interactions with Fanconi Anemia-Associated Protein Of 24 KDa (FAAP24) and Apoptosis-inducing, TAF9-like Domain 1 and 2 proteins (also known as MHF1 and MHF2)

(Deans & West, 2011). The FANCM protein contains a helicase domain and DNA translocase activity, which are important for its proper function (Xue et al, 2008).

A heterozygous *FANCM* nonsense mutation, c.5101C>T, was originally associated with increased BC risk in the Finnish population (Kiiski et al, 2014). By screening a total of 3,079 Finnish BC patients, the mutation was found to be particularly enriched in patients with triple-negative BC (OR = 3.56; 95% CI = 1.81-6.98; *P* =0.0002). Moreover, a recent study identified a loss-of-function variant, c.5791C>T (p.Arg1931*), to be associated with familial BC risk in a multinational cohort of familial BC patients (N = 8,635; OR = 3.93; 95% CI = 1.28-12.11; *P* = 0.017). These results further verify the role of *FANCM* as BC susceptibility gene (Peterlongo et al, 2015).

5.4 Low-risk genes

Currently, common genetic variants in at least 79 loci have been identified to be associated with low BC risk, and these variants together explain approximately 14% of the inherited risk of the disease (Michailidou et al, 2015). These loci have been detected through GWAS, which use a massive number of BC patients from the general population and healthy controls. The detected common variants in these studies typically confer a less than 1.5-fold elevated BC risk, and the risk varies between estrogen receptor-positive and -negative disease (Garcia-Closas et al, 2013, Michailidou et al, 2013). One of the most strongly BC-associated common SNPs is located at the 10q26 locus in the *Fibroblast Growth Factor Receptor 2 (FGFR2)* gene region. The *FGFR2* risk locus predominantly predisposes individuals to estrogen receptor-positive BC (Meyer et al, 2013). Other well-known low-risk loci are, for example, 2q33 (*CASP8)*, 2q35, 5q11 (*Mitogen-Activated Protein Kinase Kinase Kinase 1, E3 Ubiquitin* or *MAPK3K1*), 8q24, 11p15 (*Lymphocyte-Specific Protein 1* or *LSP1*), and 16q12 (*Trinucleotide Repeat Containing 9* or *TNRC9)* (Cox et al, 2007, Easton et al, 2007, Stacey et al, 2007). A recent GWAS also highlighted two novel candidate loci, 1q21.1 (*PDZ Domain Containing 1* or *PDZK1*) and 18q12.3 (*SET Binding Protein 1* or *SETBP1*), in BC susceptibility (Michailidou et al, 2015).

6 Approaches for novel breast and ovarian cancer susceptibility gene identification

In the past decades, three principal experimental designs have been used in the molecular identification of genetic breast/ovarian cancer susceptibility factors: genome-wide linkage analysis, mutational screening of candidate genes, and association studies (Turnbull & Rahman, 2008). In recent years, NGS techniques have revolutionized cancer research. NGS technology provides high-throughput and cost-effective applications that are well suited for disease gene identification and thus provide a fruitful approach to further reveal genetic factors that contribute to breast and ovarian cancer susceptibility.

6.1 Linkage analysis

Linkage analysis is based on mapping the disease locus by studying the co-segregation of genomic markers with disease phenotype utilizing multiple members of large high-risk families (Turnbull & Rahman, 2008). The region of the linked marker is further analyzed by positional cloning to identify the likely causative gene. Linkage analyses are suitable for mapping only high-penetrance genes. The two major high-penetrance BC genes, *BRCA1* and *BRCA2,* were identified using this approach in the mid-90s (Miki et al, 1994, Wooster et al, 1994).

6.2 Mutational screening of candidate genes

Candidate gene studies have focused on mutational screening of genes that encode *BRCA1* and *BRCA2*-interacting partners or genes that function in the same DDR pathway as *BRCA1/2* (Turnbull & Rahman, 2008). Usually, the whole coding region of the gene is screened by methods such as direct (Sanger) sequencing or conformation-sensitive gel electrophoresis in a cohort of breast/ovarian cancer patients. The deleterious mutations are then further analyzed in a large number of cases and controls (Erkko et al, 2007). Since the identification of *BRCA1* and

BRCA2, mutational screening of candidate genes has revealed *CHEK2*, *PALB2*, and *BRIP1*, which are low-to-moderate breast/ovarian cancer risk genes (Erkko et al, 2007, Seal et al, 2006, Vahteristo et al, 2002).

6.3   Genome-wide association studies

GWASs are based on the analysis of variant frequency in a large number of cases and controls (Turnbull & Rahman, 2008). GWASs utilize linkage disequilibrium-based markers, and the association of the marker is primarily measured indirectly (i.e., the associated marker itself is not a causative variant) (Hirschhorn & Daly, 2005). GWASs are the most suitable approach for the detection of variants that are common in the population (allele frequency >5%). Since early 00s, GWASs have become large multinational collaborations that analyze millions of genetic markers in tens of thousands of cases and controls. For example, one such collaboration effort is the Collaborative Oncological Gene-environment Study, which uses the largest dataset ever seen in cancer research (239,832 individuals from 167 research groups all over the world) (Collaborative Oncological Gene-environment Study). This study aims to identify genetic factors contributing to breast, ovarian, and prostate cancer. Currently, GWASs have identified more than 70 and 10 loci that are associated with BC and serous epithelial OC, respectively (Kar et al, 2015, Michailidou et al, 2015).

6.4   Next-generation sequencing

Next-generation sequencing has become one of the primary tools for identifying defects that underlie genetic disorders. Whole exome sequencing (WES) (sequencing of the whole protein-coding region of the genome) has become a particularly widely used approach for the identification of novel disease genes. Such advances have primarily been made in Mendelian diseases and more recently, complex diseases have been examined (Cruchaga et al, 2014, Do et al, 2015, Ng et al, 2009, Ng et al, 2010). Novel BC susceptibility genes have been identified through WES of multiple-case BC families (Kiiski et al, 2014, Park et al, 2014a, Thompson et al, 2012b). The exome represents only 1-2% of the genome but contains approximately 85% of disease-causing mutations (Ng et al, 2009). Therefore, it is a more cost-effective method in disease gene identification than whole genome sequencing. When NGS techniques

develop further and costs decrease, the whole genome sequencing will likely be the ideal method for revealing the rest of the genetic predisposition to cancer, which cannot be explained by defects in the coding region. Interestingly, NGS applications have been widely adopted in the clinical setting as well. For instance, NGS has been demonstrated to be an efficient screening method in the molecular diagnosis of hereditary breast and ovarian cancer when using multigene panels of well-known candidate genes (Castera et al, 2014, Trujillano et al, 2015).

AIMS OF THE STUDY

The aim of this study was to provide new information about the genetic factors that predispose to hereditary breast and/or ovarian cancer (HBOC) in high-risk Finnish *BRCA1/2* founder mutation-negative families. The specific aims were the following:

1) To identify additional germline variants contributing to HBOC susceptibility in seven known hereditary breast cancer-associated genes (I).

2) To analyze the role of germline copy number variations in HBOC susceptibility (II).

3) To identify HBOC susceptibility genes and gene variants through exome sequencing (III).

MATERIALS AND METHODS

1 Study subjects

1.1 High-risk HBOC individuals from the Tampere region (I-III)

The study subjects were recruited from the Tampere University Hospital Genetics Outpatient Clinic (Tampere, Finland). The hospital records and pedigree information of breast and/or ovarian cancer individuals who obtained genetic counseling between 1997 and May 2008 (n=350) were reviewed. Those individuals who 1) had strong family history of breast and/or ovarian cancer, 2) fulfilled the high-risk hereditary BC criteria, and 3) previously tested negative for *BRCA1/2* mutations were selected for inclusion (n=120). Negative mutation individuals were identified by minisequencing of the previously known 28 Finnish *BRCA1/2* mutations and by protein truncation tests (PTTs) of exon 11 for *BRCA1* and exons 10 and 11 for *BRCA2*. The high-risk hereditary BC criteria for selecting study subjects were as follows: 1) the individual or her first-degree relative (only female family members were included when defining first-degree relatives) was diagnosed with breast or ovarian cancer before reaching 30 years of age; or 2) two first-degree relatives in the family were diagnosed with breast and/or ovarian cancer and at least one of the cancers had been diagnosed at younger than 40 years of age; or 3) three first-degree relatives in the family had breast and/or ovarian cancer and at least one of the cancers had been diagnosed at younger than 50 years of age; or 4) four or more first-degree relatives had breast and/or ovarian cancer at any age; or 5) the same individual had breast and ovarian cancer; or 6) male BC was observed in the family. Individuals with bilateral BC were considered to have two separate cancers. Selected individuals were informed of the study and were asked to give a written consent to participate, to use their existing DNA samples and medical records, and to provide additional blood samples. Initially, one individual from each family

(referred to as the index individuals) who tested negative for *BRCA1/2* mutations was recruited into the study. In total, index individuals from 87 of the 120 recruited high-risk families gave written consent and new blood samples. Subsequently (between May 2013 and March 2014), recruitment was extended to affected and healthy relatives of a subset of high-risk families. In total, 81 relatives from 14 high-risk families participated in this study. Cancer diagnoses of index individuals and their family members were confirmed from the hospital records and/or from the Finnish Cancer Registry. Additionally, Population Registry Centre data were used to confirm pedigree structures of the families. The numbers of analyzed high-risk HBOC individuals and sample types in studies I-III are presented in Table 2.

1.2 High-risk HBOC individuals from the Turku region (II-III)

The study subjects were recruited from the Turku University Hospital Department of Clinical Genetics (Turku, Finland) between May 2013 and February 2015. The subjects were selected according to the high-risk hereditary BC criteria as described in the previous section. In addition, these patients had been previously tested to be *BRCA1/2* mutation-negative according to the protocol designed by the Turku University Hospital Department of Clinical Genetics. A similar recruitment protocol was applied for the study subjects as described in the previous section. Originally, one individual per family was recruited into this study and, subsequently, recruitment was extended to affected and healthy relatives of a subset of high-risk families. Written consent and permission to use medical records and existing DNA samples and to collect new blood samples were obtained from all participants. The cancer diagnoses of the index individuals and family members were confirmed from the hospital records. The numbers of analyzed HBOC individuals and sample types in studies II and III are shown in Table 2.

1.3 Breast or breast and ovarian cancer patients (III)

Breast or breast and ovarian cancer patients were *BRCA1/2*-negative females of Finnish origin. Formalin-fixed paraffin-embedded (FFPE) breast tissue block samples of breast or breast and ovarian cancer patients were obtained from the Auria Biobank (Turku, Finland). FFPE tissue samples from 31 breast or breast and ovarian

cancer patients were utilized in study III (Table 2). Both tumor and normal tissue samples were analyzed in parallel.

1.4 Male breast cancer patients (III)

Forty-four male BC patients included in this study belonged to a previously described cohort of male BC patients (Syrjakoski et al, 2003, Syrjakoski et al, 2004). Five additional male BC patients were recruited from the Turku University Hospital Department of Clinical Genetics (Turku, Finland), as described in the previous section. In total, 49 male BC patients were utilized in study III (Table 2).

1.5 Population controls (I-III)

Population controls were female or male blood donors whose blood samples were obtained from the Finnish Red Cross. The donors' blood samples had been collected from the Tampere, Turku, and Kuopio regions during the years 1997-1998. An anonymous, voluntary blood donor is between 18 and 65 years of age and healthy at the time of blood draw. Up to 989 female population control samples were utilized in studies I-III, and 909 male population control samples were utilized in study III (Table 2).

Table 2. Study subjects and sample types used in studies I-III.

Subjects and sample types	Study I	Study II	Study III
Tampere HBOC individuals			
Germline DNA	82	81	81
FFPE tissue DNA	-	-	1
Turku HBOC individuals			
Germline DNA	-	20	66
Breast or breast and ovarian cancer patients			
FFPE tissue DNA	-	-	31
Male breast cancer patients			
Germline DNA	-	-	49
Female population controls			
Germline DNA	384	905	989
Male population controls			
Germline DNA	-	-	909

1.6 Ethical aspects (I-III)

Permission to collect data from high-risk HBOC families and to use data from the Finnish Cancer Registry and Population Register Centre was granted by the National Institute for Health and Welfare on January 12th, 2009 (license number 16/5.05.00/2009). Permission to collect and use blood samples and clinical data from high-risk HBOC families who visited the Tampere University Hospital Genetics Outpatient Clinic (Tampere, Finland) was received from the Ethical Committee of Tampere University Hospital on August 13th, 2007 (latest extension on April 9th, 2013) (license number R07121). Additionally, permission to carry out the research project at the Department of Pediatrics Tampere University Hospital was obtained on October 4th, 2007. Permission to use blood and clinical tissue samples of deceased individuals for medical research purposes was obtained from the National Authority for Medicolegal Affairs on January 28th, 2008 (license number 6107/04/046/07). For collaborative studies, the Ethical Committee of Turku University Hospital granted permission on June 20th, 2012 (license number T67/2012) to collect and use blood samples and clinical data from high-risk HBOC families who visited the Department of Clinical Genetics, Turku University Hospital (Turku, Finland). Additionally, the Auria Biobank (Turku, Finland) provided permission to use their samples in the study (license number AB14-9588). All individuals participating in this study were informed of the analyses and gave written consent to use their samples and medical records in the study.

2 Methods

2.1 DNA extraction (I-III)

Genomic DNA was extracted from peripheral blood leukocytes using Puregene™ kits (Gentra Systems, Inc., Minneapolis, MN, USA) and Wizard® Genomic DNA Purification Kit (Promega Corporation, Madison, WI, USA). DNA concentration and purity was measured using a NanoDrop® ND-1000 Spectrophotometer and NanoDrop 1000 V3.7.1 software (NanoDrop Technlogies Inc., Wilmington, DE, USA). DNA from the FFPE breast tissue block was extracted using the Arrow DNA kit and NorDiag Arrow Automated Magnetic Bead-Based Nucleic Acid Extraction System (DiaSorin, Saluggia (Vercelli), Italy) according to the manufacturer's instructions at Fimlab Laboratories (Tampere, Finland) (one Tampere HBOC individual sample). The DNA extraction of FFPE blocks obtained from the Auria Biobank was performed using a GeneRead DNA FFPE Kit (Qiagen Inc., Valencia, CA, USA) at the Department of Medical Biochemistry and Genetics, University of Turku (Turku, Finland). Prior to DNA extraction, the FFPE block was reviewed by a pathologist to confirm the presence of tumor cells.

2.2 Sanger sequencing (I, III)

Sanger sequencing was utilized to screen sequence variants from genomic DNA. In study I, the entire coding region and exon-intron boundaries were analyzed for *BRCA1* (excluding previously analyzed exon 11), *BRCA2* (excluding previously analyzed exons 10 and 11), *PALB2, BRIP1, RAD50,* and *CDH1*. In study III, 18 variants were detected by exome sequencing and confirmed by Sanger sequencing for the following genes: V-*Akt Murine Thymoma Viral Oncogene Homolog 2 (AKT2), ATM, BRCA1, Cyclin-Dependent Kinase Inhibitor 2A (CDKN2A), MYC, Nuclear Receptor Coactivator 3 (NCOA3), Plasminogen Activator, Urokinase (PLAU), RAD1 Checkpoint DNA Exonuclease (RAD1), RAD50, RAD52 Homolog (S. Cerevisiae) (RAD52), Retinoblastoma-Like 2 (RBL2), RPA2, Ribonucleotide Reductase M2B (TP53 Inducible) (RRM2B), Wingless-Type MMTV Integration Site Family, Member 3 (WNT3),* and

Wingless-Type MMTV Integration Site Family, Member 10A (WNT10A). Additionally, *AKT2* and *RRM2B* gene variants were screened in cohorts of cancer patients and healthy controls by Sanger sequencing. Primers for *BRCA1, BRCA2, CDH1, AKT2, ATM, CDKN2A, MYC, NCOA3, PLAU, RAD52, RBL2, RPA2, RRM2B, WNT3,* and *WNT10A* were designed with Primer3 software (Koressaar & Remm, 2007, Untergasser et al, 2012), whereas primer sequences for *PALB2, BRIP1,* and *RAD50* have been reported previously (Erkko et al, 2007, Tommiska et al, 2006b, Vahteristo et al, 2006). All primer sequences used for Sanger sequencing are presented in Table 3. The PCR products were purified using rAPid Alkaline Phosphatase (Roche Diagnostics GmbH, Mannheim, Germany) and Exonuclease I (New England Biolabs, Ipswich, MA, USA) according to the ExoSAP-protocol (Nucleics, Woollahra, Australia). Sequencing was carried out using the ABI PRISM® BigDye™ Terminator v3.1 Cycle Sequencing Kit and the Applied Biosystems 3130xl Genetic Analyzer according to the manufacturer's instructions (Life Technologies Corporation, Carlsbad, CA, USA). The sequences were analyzed with Sequencher software (different versions) (Gene Codes Corporation, Ann Arbor, MI, USA).

Table 3. Primer sequences used in Sanger sequencing.

Gene_exon	Forward (5'->3')	Reverse (5'->3')
Study I		
*BRCA1*_exon 3	atgaagttgtcattttataaacctttt	ggtcaattctgttcatttgcat
*BRCA1*_exon 4	cagttcctgacacagcagaca	tggagccacataacacattca
*BRCA1*_exon 5	ggctcttaagggcagttgtg	gtggttgcttccaacctagc
*BRCA1*_exon 6	cacttgctgagtgtgtttctca	tggacagcacttgagtgtca
*BRCA1*_exon 7	gcatacatagggtttctcttggtt	aaaattagcctggcatggtg
*BRCA1*_exon 8	gccaacaattgcttgactgtt	gctgcctaccacaaatacaaa
*BRCA1*_exon 9	cccatgcctttaaccacttc	tgcacatacatccctgaacc
*BRCA1*_exon 10	ttggtcatttgacagttctgc	tgggttgtaaaggtcccaaa
*BRCA1*_exon 12	aatccagtcctgccaatgag	aatgcaaaggacaccacaca
*BRCA1*_exon 13a	catgggcattaattgcatga	tttggccaacaatacacacc
*BRCA1*_exon 13b	cagcaggaaatggctgaact	gagcagggacaagaaccaag
*BRCA1*_exon 14	tgaattatcactatcagaacaaagca	gatgtcagataccacagcatcttt
*BRCA1*_exon 15a	attggcaggcaacatgaatc	tgtttgttccaatacagcagatg
*BRCA1*_exon 15b	cgatggtttctccttccat	gagctattttctaaagtgggctta
*BRCA1*_exon 16a	ttggatttccaccaacactg	ccctgctcacactttcttcc
*BRCA1*_exon 16b	aagacagagccccagagtca	cacagaactgtgattgttttctagatt
*BRCA1*_exon 17	tgtagaacgtgcaggattgc	tttatgcagcagatgcaagg
*BRCA1*_exon 18	tccagattgatcttgggagtg	taaagggaggaggggagaaa

Table 3.   Continued.

Gene_exon	Forward (5'->3')	Reverse (5'->3')
*BRCA1*_exon 19	agggaaggacctctcctctg	ggtgcattgatggaaggaag
*BRCA1*_exon 20	tgacgtgtctgctccacttc	cagagtggtggggtgagatt
*BRCA1*_exon 21	tctggacattggactgcttg	gctgttttgtttggagagtgg
*BRCA1*_exon 22	agagacagactctcccattgag	aggctacagtaggggcatcc
*BRCA1*_exon 23	ctgggtgacagagcaagacc	aaaccaaacccatgcaaaag
*BRCA1*_exon 24	ggcgacagagcaagactcc	gccaggacagtagaaggactg
*BRCA2*_exon 2	tggctagtggcaccggtttgg	tcatgtggttaacctgcaaacga
*BRCA2*_exon 3a	agattcgcaagagaatggattaatga	caggagattggtacagcggca
*BRCA2*_exon 3b	aagaactttcttcagaagctccaccc	cgtaattcagcaaatctcagttggca
*BRCA2*_exon 4	agaatgcaaatttataatccagagta	aatcagattcatctttatagaacaaa
*BRCA2*_exons 5&6	aacaatttatatgaatgagaatc	taaatctcagggcaaagg
*BRCA2*_exon 7	taagtgaaataaagagtgaa	aacagaagtattagagatgac
*BRCA2*_exon 8	aatagtagatgtgcttttga	acatataggaccaggtttagagac
*BRCA2*_exon 9	aagtgaaaccatggataagg	gcagaggttgcggtaaac
*BRCA2*_exon 12	aggtcactatttgttgtaag	agtggctcatgtctgtaat
*BRCA2*_exon 13	taaagcctataattgtctca	cttcttaacgttagtgtcatt
*BRCA2*_exon 14a	atgtagcaaatgagggtctg	gtctgcctgtagtaatcaagtgt
*BRCA2*_exon 14b	cttcaagcaatttagcagtttcag	caaaggggggaaaaccatcag
*BRCA2*_exon 15	ccaggggttgtgctttta	aaaattacactctgtcataaaagcc
*BRCA2*_exon 16	tttggtaaattcagttttggttt	aacacacaatcttttgcataga
*BRCA2*_exon 17	cagagaatagttgtagttgttgaa	agaaaccttaaccatactgc
*BRCA2*_exon 18	attcagtttttattctcagttattc	tttaactgaatcaatgactg
*BRCA2*_exon 19	aagtgaatattttttaaggcagtt	tatatggtaagtttcaagaatacatc
*BRCA2*_exon 20	cactgtgcctggcctgat	ggcttagacctgatatttctgtcc
*BRCA2*_exon 21	gggtgttttatgcttggttct	catttcaacatactccttcctg
*BRCA2*_exon 22	gatgagctctaattttgttgta	aatcattttgttagtaaggtcat
*BRCA2*_exon 23	cacttcttccattgcatctt	acaaaacaaaaattcaacatatgg
*BRCA2*_exon 24	caacaactaccggtacaaac	ggtagctccaactaatcataaga
*BRCA2*_exon 25	taaaattcatctaacacatctat	caactatccaatttgtataaaag
*BRCA2*_exon 26	aggaaatactttggaaacat	catgtttactaggtatacaacagaa
*BRCA2*_exon 27a	taggagttaggggagggagactg	caaatgggactaacaggtggag
*BRCA2*_exon 27b	gatgacttcaaagtcttgtaaa	gcagtcctagtggattcac
*BRCA2*_exon 27c	tcaatagctgacgaagaacttg	gtggtttgaaattatattccagtc
*PALB2*_exon 1	aggccgaatggtggattta	acacaaagccaggcctaaaa
*PALB2*_exons 2&3	ccacttgcccagtattgttg	cctgggaaatgaataataaagca
*PALB2*_exon 4a	ctgaatgaaatgtcactgattctt	tgatttcttcctgttcctttagtc
*PALB2*_exon 4b	agagactgtgtctttggcactg	aaggaggttatctgtagagacagtca
*PALB2*_exon 4c	tgtaagtttggaggcacaagg	tccacggctactttcctctg
*PALB2*_exon 4d	tcttgcacagtgcctgaag	gaagttggcaaaagtggttca

Table 3. Continued.

Gene_exon	Forward (5'->3')	Reverse (5'->3')
*PALB2*_exon 4e	actgaagataatgacttgtctaggaa	gtatgctggctttgcgagtt
*PALB2*_exon 5a	tgttgggttttgttactattttgtg	gttcctggcgggacagagt
*PALB2*_exon 5b	cgcatggatacagaaatgg	ggcaaatagtaattgttaactttcatc
*PALB2*_exon 5c	actaaacaattcgacagttcagg	acattcactaaggcatttcattc
*PALB2*_exon 6	acaatacgaagtagacattttgatga	caagaagctatatgactgaattctttt
*PALB2*_exon 7	tgctttgcataaaacagcact	tggtaagctgcccatctaca
*PALB2*_exon 8	attatccttgtacagtgagaatacaaa	tgcacttaaaaccagctgaca
*PALB2*_exon 9	aaaaggttactcctcacatcacc	cagcacagaaaaacgagatcc
*PALB2*_exon 10	cagttcaacaatgcggagaa	tcttcacaacaaccctgtaaaa
*PALB2*_exon 11	tttccctggtcacctcctaa	tgcttatgacttactgctctcactt
*PALB2*_exon 12	tcctgacatactcttgacagtc	tccattcttctaagtgacacaaaaa
*PALB2*_exon 13	tggatatgtaatctgaattatatcttc	cctaaaacccttttctcaaagt
*BRIP1*_exon 2	ctgtttccagatttctccc	gtgaacccagaaaatattctcc
*BRIP1*_exon 3	ccctggagtgcaatctcact	tagcgacagcatggctgaa
*BRIP1*_exon 4	cctgggtgaactgggctgtag	ttttaaaatcattccagagaaactaga
*BRIP1*_exon 5	ttgcctacctgtaagttatttatg	accatgttcagctgtaactaactg
*BRIP1*_exon 6	gagctgttttggcctttgaga	ctgagtgggttgctactgtcct
*BRIP1*_exon 7	gttctgattccatgtgaggtt	gtacatataaaacacatactgagt
*BRIP1*_exon 8	gatgttcctcaaattctgagataa	catctaaaagcttttacattcaac
*BRIP1*_exon 9	gcctatagtgtgaattttaaaatg	cctagttaaccaaagtttactaac
*BRIP1*_exon 10	gatcaacgcatgacaataatgatg	gggttactcactagatttaatctg
*BRIP1*_exon 11	gcatgttttgttgggtttcattgt	ggtatgtattaaacacatgctagc
*BRIP1*_exon 12	gtaccagctctttcaaatgag	ctatctttaaaagagtcaaccac
*BRIP1*_exon 13	acaggtgtgagccactttgc	ggcacttcaggtatcttctaacttg
*BRIP1*_exon 14	cttgttgcttgatcttttatgtac	ctaggaagcttactgtggtaa
*BRIP1*_exon 15	acagctctatgagatatattg	tcataggagaacaagtacaat
*BRIP1*_exon 16	ttttcaaaatgacaagaataagcaa	aagcaaagcgcaataaaatga
*BRIP1*_exon 17	ctgttagaagttaatatgatg	gaatacataccagttcctatg
*BRIP1*_exon 18	ccaattttctgtctgtcccac	gatagtagagctcatgttatgtg
*BRIP1*_exon 19	cttcactagaaaaagcaagtg	ccaccatatttaaggaattaatc
*BRIP1*_exon 20a	acctagcaattatgttagct	tctgtatcttcaggatcgta
*BRIP1*_exon 20b	attgatgccacccttacta	taacataagcatgatgacata
*RAD50*_exon 1	ttgcttcggcctcagttaag	ccgaaagatgtagcgacctg
*RAD50*_exon 2	tttatggtaaacttctgtggttctc	ttttccagtgccaagttttct
*RAD50*_exon 3	tgccttttctcagaaccaac	gaaaacaaccatcaacttacagacc
*RAD50*_exon 4	gccatttctaacgggatagg	gcatccaaattgcaaacaca
*RAD50*_exon 5	agtgacagcataatatcccactg	ttgatttagccagtccacga
*RAD50*_exon 6	tgcctggacctggagtatct	ggatggcaaaatggattcaa
*RAD50*_exon 7	tgaaggatattgaataaggtttggtt	ttcactcagcataagtccttgg

Table 3.   Continued.

Gene_exon	Forward (5'->3')	Reverse (5'->3')
*RAD50*_exon 8	cgtgaatctgcagctatctcaa	atgccaaaatggagtccaac
*RAD50*_exon 9	tcttcagtgtacatatattccttttgc	aaaatgctataatcttaggatcaaaaa
*RAD50*_exon 10	aattcttgcacaaaatatataacacc	tgagtgcaaggtaggctatttta
*RAD50*_exon 11	tttttgttcttgatataatgtggagat	aaacaattaaacaggtaagcatgaaa
*RAD50*_exon 12	tttgcctactcaaattttcaaac	atgaaaattcaaaggtgtcaaag
*RAD50*_exon 13	agaaaaagatacaaccgtattcaga	aggcatgagatgggtacctt
*RAD50*_exon 14	cctcagtctaactgtgagaaacatc	caagccaccatggaacaag
*RAD50*_exon 15	tttgctaaaattgtatctagaaatgg	ggaaggaacttcgttgtcca
*RAD50*_exon 16	ccgataaaaatgggaagaatga	gatcgtgctactgcactcca
*RAD50*_exon 17	tgacccctaaagtaagataatggaa	agtgtccgacgtggtgctat
*RAD50*_exon 18	ctgtgctctcctgttatgtgc	tctttgtgtttctcgcattca
*RAD50*_exon 19	ttcaagatgggaaagacgac	gcatttctattcaatggatcttct
*RAD50*_exon 20	tgcctgttacagatttcatgtt	ccaggctgaggctaaaattc
*RAD50*_exon 21	tgacttttccacttcaggttgtt	tgcaataagaaaatccccagtc
*RAD50*_exon 22	accattgaaaatattgaggaagtt	ggtcataaggggaagagctg
*RAD50*_exon 23	caaaaggctacagagcataggtt	tctctctttcagttacttgggtga
*RAD50*_exon 24	gcctgccatgagatgagaag	cctgtccccaaatgtgagat
*RAD50*_exon 25	caaatcaaagaaggggttatgc	cactttctgaggacctacatttct
*CDH1*_exon 1	gtgaaccctcagccaatcag	gacccgaactttcttggaag
*CDH1*_exon 2	ggtcttgaggggtgactc	ggtgtgggagtgcaatttct
*CDH1*_exon 3	gttcgctctttggagaagga	ggctgagaaacctggattaga
*CDH1*_exon 4	gtctggctaggttggactgt	cactgggtctttttccctttc
*CDH1*_exon 5	ggatttggcagagaagtacc	gttaagctcctcatgtgttcag
*CDH1*_exon 6	ctcagagcctaggaaggtgtg	gttacacaacctttgggctt
*CDH1*_exon 7	gtcccctcctttatccctca	gacaactggcctagcaggat
*CDH1*_exon 8	ctggttccatgtgttgggc	ccatgagcagtggtgacactt
*CDH1*_exon 9	cctttagcccccctgagact	gccaaagcgaatctacttct
*CDH1*_exon 10	gaaagtcatggcagaaacca	ggtcttgtacagacaaatgaca
*CDH1*_exon 11	gaagaagcgcttaagccgtt	gaagctcttccctccaaaagaa
*CDH1*_exon 12	ctgaagagccaggacaagat	gtatcaatggaaggggtgac
*CDH1*_exon 13	cttgcgggtgtctttagttc	ggagtctcttcccacatca
*CDH1*_exon 14	ctgtgatagctgctgcttct	gctgtttcaaatgcctacct
*CDH1*_exon 15	ggcatcatccaaccataatc	gctcaggcaagctgaaac
*CDH1*_exon 16	ccacaagtctgggtgcatt	gaaactcatctcaagggaagg
Study III		
*AKT2*_exon 3	gtggtaccccttgtgagtca	gctaacaaggagggagagca
*ATM*_exon 17	tgaccgtggagaagtagaatca	tgaggcctcttatactgcca
*ATM*_exon 22	aggcatctaacaaaggagagga	tgtaagacattctactgccatct

Table 3. Continued.

Gene_exon	Forward (5'->3')	Reverse (5'->3')
*ATM*_exon 29	aattcttcttgccatatgtgagc	tgggcggacagagtgagt
*ATM*_exon 37	tgtcaacggggcatgaaaat	gttgtggatcggctcgtttg
*BRCA1*_exon 9	acacccaggatcctttcttg	tgtaaaatgtgctccccaaaag
*CDKN2A*_exon 3	agcccatacgcaacgagatt	gccagcttgcgataaccaaa
*MYC*_exon 2	actcaagactgcctcccg	gaagggagaagggtgtgac
*NCOA3*_exon 18	gtagagaagtgatgcctggc	ctcatgatgttggctcgtgg
*PLAU*_exon 2	gagaggagagagacagctgg	cccgtggttctcaagcct
*RAD1*_exon 4	tggaattctctacacacaactgt	tcactcgtcatatccaattcaga
*RAD50*_exon 3	acctgatctcctaatgatgctga	acaaccatcaacttacagacct
*RAD52*_exon 6	gtctgccgttttcccacttt	ctgcgtgactgtgtaactacg
*RBL2*_exon 13	ggaagagcggctgtttcaaa	cacaaacctcttcacatgtagga
*RPA2*_exon 2	gaaggttagggagactcggg	acaggaggtggaaaatttggc
*RRM2B*_exon 1	ggatgaggtaaatgttgctgttg	aactccttgtcaccatccca
*WNT3A*_exon 2	tgttgggccacagtattcct	aggctcagtcctgcttcag
*WNT10A*_exon 2	ccaatgacattctggacctcc	aagaggaaggagagcagcag

2.3   Multiplex Ligation dependent Probe Amplification (MLPA) (I, II)

The MLPA method was used to detect large genomic rearrangements in *BRCA1* and *BRCA2*. The following SALSA® MLPA® kits were used according to the manufacturer's instructions: P002-B1 (lot 0508) for *BRCA1* (study I), P090-A2 (lot 0808) for *BRCA2* (study I), and POO2 probemix and EK1 reagent kit (lot C2-0811) for *BRCA1* (study II) (MRC-Holland, Amsterdam, the Netherlands). The analysis was performed with Applied Biosystems 3130xl Genetic Analyzer and with Applied Biosystems Genemapper® v.4.0 and Peak Scanner™ v1.0 software (Life Technologies), as indicated by the manufacturer. In study II, the National Genetics Reference Laboratory (Manchester, UK) Spreadsheet was utilized for data analysis according to the manufacturer's instructions.

2.4   High-Resolution Melt (HRM) analysis (I)

High-Resolution Melt (HRM) analysis was used to screen for genomic DNA sequence variants in *CHEK2*. The primers were designed with Beacon Designer™ software (PREMIER Biosoft, Palo Alto, CA, USA). The primer sequences are

presented in Table 4. The HRM analysis was performed on a Bio-Rad platform (Bio-Rad Laboratories Headquarters, Hercules, CA, USA) according to the manufacturer's instructions.

Table 4. *CHEK2* primer sequences used in HRM analysis.

Exon_amplicon	Forward (5' -> 3')	Reverse (5' -> 3')
1_1	acctttgttgttggacacttt	tggagtttggcatcgtg
1_2	cacgatgccaaactcca	taggctcctcaggttctt
1_3	aagaacctgaggagccta	caactccaatcagaacct
2_1	tgataattctgattgccttct	accagtagttgtcattcac
2_2	gtgtgaatgacaactact	ttgctgtatgttcggtat
2_3	ataccgaacatacagcaa	actttgaatagcagagattt
3_1	tttcccactataaatctctgcta	ttccattgccactgtgat
3_2	catacatagaagatcacagt	attccagtaaccataagataa
4_1	tgaaacccatttctactcttt	tgtattcatctcttaatgcctta
4_2	cctaaggcattaagagatgaatac	gagaaaccaccaatcacaaat
5_1	acattcagacaatcacta	caatagcaaacttccttt
5_2	tattggttcagcaagaga	ctcacaaattcatccatcta
6_1	actctgttattctgtttatcaaa	aggctttatactcttctcatat
7_1	gtttctcactactttcccttt	gcaagcctacattagattct
8_1	tttgtgtttcaggatgga	aatagagcttgcaggtag
8_2	ctacctgcaagctctatt	cgatttctgcctaattca
9_1	ttggagaatatggttgtg	cttcttgagatgacagtaa
9_2	gtattatacaccgtgactta	gaacaagaatctacaggaa
10_1	ggcaagttcaacattattcc	gttccacataaggttctcat
10_2	gaaccttatgtggaaccc	caacagaaacaagaacttcag
10_3	aagttcttgtttctgttg	tcatctaatcacctccta
11_1	ctgagaatgccacttgatt	acacttgagtcctatgct
11_2	agcataggactcaagtgt	cttctcatattcatacctttctctg
11_3	cagagaaaggtatgaatatgaaag	gaacagaattgacaggagaa
12_1	tctggcatactcttactgata	aaggcttcttctgtcgta
12_2	gaagttgttggtagtggat	cagggcttcccatgtatt
13_1	tattatccttcagacaca	atcatctttgcttatcag
13_2	gatgaagacatgaagaga	atcaggaatacgaatacc
14_1	tttcttttgaacatttctccatt	ttcacaacacagcagcaca
14_2	gtgctgctgtgttgtgaa	cccaggttccatcaggtt

2.5 SNP genotyping (I, III)

SNP genotyping was one of the methods used to determine the frequencies of the SNPs identified in population controls in study I and in patient cohorts and population controls in study III. The genotyping was performed with Applied Biosystems TaqMan® SNP Genotyping Assays, with ABI PRISM® 7900HT Sequence Detection System and SDS v2.2.2 software (Life Technologies) according to the manufacturer's instructions. Pre-designed and functionally tested assays were used for three *BRCA2* SNPs (rs28897749, rs11571833, rs1801426), and for one *PALB2* SNP (rs152451) in study I. For four SNPs, c.72A > T (*BRCA2*), c.814G > A (*PALB2*), c.1000T > G (*PALB2*), and rs45551636 (*PALB2*), no pre-designed assays were available and assays were designed using Applied Biosystems Custom TaqMan® Assay Design Tool (Life Technologies) according to the manufacturer's instructions (study I). In study III, the following assays were used: C_176281813_10 (*CDKN2A* c.C496T), C___1244825_20 (*RAD52* c.G538A), C___2283286_20 (*ATM* c.T2572C), C__45273750_10 (*ATM* c.C3161G), C_166902853_10, (*RPA2* c.C122T), C__30585831_10 (*ATM* c.A5558T), C__63879305_20 (*ATM* c.A4424G), C__32333045_10 (*RAD50* c.A280C), C_153129907_10 (*BRCA1* c.A3904C), C__32376001_10, (*MYC* n.A77G), C_102161615_10 (*RBL2* c.G1723C), C_167350917_10, (*NCOA3* c.A3353C), C__15956024_20, (*PLAU* c.G43T), C__15760210_10 (*RAD1* c.G341A), C_168146154_10 (*WNT10A* c.C337T), and C_190555357_10 (*WNT3A* c.G277A).

2.6 Copy number variation (CNV) analysis (II)

SNP array protocol. A genome-wide SNP genotyping to identify CNVs from genomic DNA was performed at the Technology Centre, Institute for Molecular Medicine Finland, University of Helsinki (Helsinki, Finland). For this analysis, an Illumina HumanCytoSNP-12 v2.1 Beadchip, an iScan system, and standard reagents and protocols provided by Illumina Inc. (San Diego, CA, USA) were used. Illumina's HumanCytoSNP-12 v2.1 Beadchip contains approximately 300,000 markers distributed throughout the genome and has been optimized to target regions of cytogenetic importance.

    Data-analysis. The genotypes were visualized and analyzed with GenomeStudio™ v2010.2 software, as indicated by the manufacturer (Illumina). The call rates were greater than 99.5%. It was ensured that the samples met high quality

criteria (Wang et al, 2007b). CNV calling was performed with the PennCNV (2009Aug27) program (Wang et al, 2007b). In addition, the QuantiSNP v2.3 (Colella et al, 2007) and cnvPartition v3.1.6 (Illumina) programs were utilized to confirm the PennCNV results when selecting CNVs for validation. All three programs were used with default parameters. CNVs spanning less than three SNPs were filtered out. Identified CNVs were queried against the Database of Genomic Variants (Database of Genomic Variants) using the NCBI Genome Build 36 (hg 18). CNVs were annotated using the NCBI RefSeq gene database (NCBI Reference Sequence Database) to identify genes and exons overlapping the observed CNV loci. For intergenic CNVs, the loci were expanded upstream and downstream of the CNV to identify neighboring genes. CNV-affected genes were further studied by enrichment analyses, including Gene Ontology terms, KEGG pathways, Pathway Commons, and Wikipathways to reveal common functions of the gene products using the Web-based Gene Set Analysis Toolkit V2 (WebGestalt2) (Zhang et al, 2005). Additionally, CNV-affected genes were queried against the NCBI Online Mendelian Inheritance in Man database (NCBI OMIM Database) and the Genetic Association Database (Genetic Association Database) to determine whether the genes have previously been associated with the disease. The Encyclopedia of DNA Elements database (Encyclopedia of DNA elements at UCSC) was utilized to search possible functional elements in the CNV region

CNV validation and genotyping. Selected CNVs were validated, and their frequencies were determined in an additional cohort of HBOC patients and population controls using TaqMan® Copy Number Assays, an ABI PRISM® 7900HT Real-Time PCR system, and SDS v2.2.2 software, according to the manufacturer's instructions (Life Technologies). The following pre-designed TaqMan® Copy Number Assays were used: Hs04703682_cn (2q34), Hs03458738_cn (3p11.1), Hs03253932_cn (5q15), Hs06178677_cn (8p23.2), Hs02640223_cn (17q21.31), and Hs04482315_cn (19q13.41). As an internal standard, a TaqMan® RNaseP Reference Assay was run with the TaqMan® Copy Number Assays in a duplex, real-time PCR reaction following the manufacturer's guidelines (Life Technologies). The analysis was performed with Copy Caller v1.0 software, and copy numbers were calculated using the comparative CT method, as indicated by the manufacturer (Life Technologies).

2.7 Exome sequencing (III)

Exome capture and next-generation sequencing protocol. Exome capture and next-generation sequencing was performed at the BGI Tech Solutions (Hong Kong) Co. Ltd using a SureSelect Human All Exon 51M kit (Agilent Technologies, Inc., Santa Clara, CA) and HiSeq 2000 technology (Illumina). Protocols provided by Agilent, Illumina, and the BGI were followed. The genome sequencing coverage depth was 50x per sample.

   Data-analysis. The reads were aligned with Bowtie2 against the human reference genome (hg19) using default parameters (Langmead & Salzberg, 2012). The quality control was performed by FastQC (Babraham Bioinformatics FastQC). As a preprocessing step for variant calling, PCR duplicates and reads with mapping quality less than 10 were filtered out using SAMtools (Li et al, 2009). Variant calling was performed using a bioinformatics toolkit (Pypette) that was developed in-house (GitHub annalam/pypette). Variant annotation was conducted with Annovar (Wang et al, 2010) using the Refseq genes as the reference gene set. Several filtering steps were used to reduce the number of candidate variants for downstream analyses. Based on the variant annotation, functional variants (non-synonymous single nucleotide variants (SNVs), splicing site SNVs/indels, stop gain/loss variants, and indels inducing frameshift) were prioritized. Furthermore, rare variants (minor allele frequency ≤ 0.05 were selected by screening the variants against the data from the 1000 Genomes Project (1000 Genomes Project Consortium et al, 2012) (August 2014 version) and the Exome Sequencing Project 6500 (included in Annovar) (NHLBI Exome Sequencing Project). Additionally, the variants were screened against the Sequencing Initiative Suomi database to obtain information about allele frequencies in the Finnish population (Sequencing Initiative Suomi). From the remaining set of variants, only those variants for which the host genes participate in the DNA damage response pathway were selected. The pathway data were retrieved from the IntPath-database, which includes data integrated from the following several sources: KEGG, Wikipathways and BioCyc (Zhou et al, 2012). To further reduce the number of candidate variants, the deleteriousness of the variants was evaluated by utilizing a precompiled set of pathogenicity predictions included in the Annovar's ljb26_all dataset. The dataset covers predictions of the deleteriousness of all possible SNVs for 10 pathogenicity predictor methods and 3 methods evaluating the evolutionary conservation of the variant locus. Only variants that were predicted to be deleterious by at least one of the predictors or were frameshift indels, were selected for further assessment. Moreover, variants present only in healthy relatives

were excluded. Furthermore, variants that had a reliable genotype call rate and were shared between affected family members or between families were prioritized. Based on the above-mentioned filtering criteria, candidate variants were selected for further validation experiments (described in sections 2.2 and 2.5). Moreover, variants present in only early-onset patients were analyzed in detail. Similar filtering criteria based on functionality, rarity, and predicted pathogenicity were used, as described above. However, variants targeting any pathway genes were considered.

2.8 Statistical analyses (I-III)

Association of the identified variants with breast/ovarian cancer was tested using Fisher's exact test, and the chi-square test (R v2.15.2 (The R Project for Statistical Computing) (R Core Team, R Foundation for Statistical Computing, Vienna, Austria)) and PLINK v1.07 (Purcell et al, 2007). CNV distribution and median lengths were compared between HBOC individuals and controls using the Wilcoxon test (R v2.15.2) (study II). When Fisher's exact test resulted a non-numerical value for the enrichment analysis, A VCD package was implemented in R to estimate the numerical values of the odds ratios (Meyer et al, 2012) (study II). P-values were two-sided. $P<0.05$ was considered statistically significant.

2.9 Bioinformatics tools (I, III)

A Basic Local Alignment Search Tool (NCBI BLAST) tool was utilized to determine whether the identified novel variants were located in genomic regions that are conserved across different species (study I). The pathogenicity of missense variants was predicted using the Pathogenic-or-Not-Pipeline (PON-P) (Thusberg & Vihinen, 2009) (study I). PON-P integrates amino-acid tolerance predictors (PolyPhen version 2, Sift, PhD-SNP, SNAP) and a protein stability predictor (I-Mutant version 3) to predict the probability that missense variants affect protein function. ESEfinder was used to predict the effects of variants on exonic splicing enhancer elements (Cartegni et al, 2003) (study III).

2.10 MicroRNA database search (I)

A microRNA target site search of the microRNA database was performed for the genomic positions of the novel variants (miRBase).

RESULTS

1 Germline sequence variants in *BRCA1*, *BRCA2*, *CHEK2*, *PALB2*, *BRIP1*, *RAD50*, and *CDH1*, and their contribution to HBOC susceptibility in high-risk families (I)

In study I, coding region and exon-intron boundaries of seven known BC susceptibility genes were analyzed to identify additional variants that predispose to inherited forms of breast and ovarian cancer. Index individuals from 82 of the high-risk hereditary breast and/or ovarian cancer families analyzed in this study were previously screened to be negative for 28 Finnish *BRCA1* and *BRCA2* founder mutations by minisequencing. These individuals also tested negative for protein-truncating variants in the largest exons (exon 11 for *BRCA1* and exons 10 and 11 for *BRCA2*), suggesting that variants other than these already tested defects contribute to breast and/or ovarian cancer susceptibility in these families. By thoroughly screening 1) *BRCA1* and *BRCA2* using MLPA and Sanger sequencing (with the exception of the previously analyzed largest exons); 2) *PALB2, BRIP1, RAD50*, and *CDH1* by Sanger sequencing; and 3) *CHEK2* using the HRM method, 54 different variants in 82 HBOC individuals were identified. The variant spectrum was as follows: 19 non-synonymous coding variants, 11 synonymous coding variants, 21 intronic variants, and 3 5'-untranslated region variants. All the detected variants were single nucleotide changes or indels. In total, 14/54 (25.9%) of the identified variants were novel.

Based on the variant type, novelty, observed frequency in HBOC individuals, and available information in the databases, the most promising variants were selected (n=18), and their frequency was determined in 384 controls. The primary focus was on non-synonymous coding variants as these are most likely to be disease-causing.

Variant frequencies between HBOC individuals and controls were compared using Fisher's exact test, but no statistically significant difference was observed for any of the variants. This result was probably due to the rareness of the variants and the limited sample size used in this study. All the identified non-synonymous variants and their association with breast and/or ovarian cancer risk are presented in Table 5. Three previously known intermediate breast/ovarian cancer risk variants, *BRCA1* Arg1699Trp, *CHEK2* Ile157Thr, and *CHEK2* 1100delC (bold in Table 5) were observed in 11/82 (13.4%) HBOC individuals. One individual carried both of the *CHEK2* variants. The contribution of the two *CHEK2* variants to BC risk was notably high in the analyzed HBOC families, 10/82 (12.2%) (Table 5). The effects of novel missense variants (n=5) on protein function was predicted using the PON-P program: of the analyzed variants, only one possible pathogenic variant (*BRCA2* Leu24Phe) was predicted. The substitution of the nonpolar leucine by the nonpolar phenylalanine with an aromatic ring structure in the side chain at position 24 was predicted significantly alter the properties of the side chain; this substitution was predicted to be untolerated. The *BRCA2* Leu24Phe variant was observed in one BC patient (1/82, 1.2%) but in none of the 380 healthy controls (Table 5). The BC patient had been diagnosed with ductal, grade III disease with ER+, PR- and HER2+ status at the age of 53 years. Additionally, the patient had two BC-affected first-degree relatives.

The clinical characteristics of the HBOC individuals carrying the identified non-synonymous variants were reviewed, but no clear trend was observed among individuals carrying the same variants. However, one interesting case was an early-onset (diagnosed at the age of 26 years) BC patient carrying both of the BC-associated *CHEK2* variants, Ile157Thr and 1100delC. Thus, the combination of the two *CHEK2* variants was reported to contribute to early disease onset. Novel variant genomic positions were queried against the known microRNA target sites from the miRBase, but no hits were found. Moreover, a Basic Local Alignment Search Tool was utilized to determine whether the identified novel variants were located in conserved genomic positions, which may indicate a pathogenic role for the variant. Sequence similarities between different species were observed for several novel variant genomic positions, including *BRCA2* Leu24Phe.

Detailed analysis of the *BRCA1* and *BRCA2* genes (Sanger sequencing and MLPA, excluding the previously analyzed largest exons) revealed additional variants that were not detected by the genetic testing protocol that targeted only the Finnish founder-mutations. One such variant was Arg1699Trp in *BRCA1,* which is a known intermediate breast/ovarian cancer risk variant.

Table 5. Identified non-synonymous variants in *BRCA1, BRCA2, CHEK2, PALB2, BRIP1*, and *RAD50*, and their association with hereditary breast and/or ovarian cancer risk.

Gene, variant	Carrier frequency (%)		OR	95%CI	*P*
	HBOC individuals	**Controls**			
BRCA1, Ser1613Gly	52/82 (63.4%)	-	-	-	-
BRCA1, Met1628Thr	4/82 (4.9%)	6/367 (1.6%)	3.09	0.85-11.19	0.090
***BRCA1*, Arg1699Trp**	**1/82 (1.2%)**	**-**	**-**	**-**	**-**
BRCA2, Leu24Phe*	1/82 (1.2%)	0/380 (0%)	-	-	0.177
BRCA2, Val2728Ile	1/82 (1.2%)	1/378 (0.3 %)	4.65	0.29-75.19	0.325
BRCA2, Lys3326Stop	1/82 (1.2%)	11/378 (2.9%)	0.41	0.05-3.24	0.702
BRCA2, Ile3412Val	1/82 (1.2%)	8/379 (2.1%)	0.57	0.07-4.64	1.000
***CHEK2*, Ile157Thr**	**8/81 (9.8%)**	**21/381 (5.5%)**	**1.88**	**0.80-4.41**	**0.203**
***CHEK2*, 1100delC**	**3/82 (3.7%)**	**6/380 (1.6%)**	**2.37**	**0.58-9.67**	**0.203**
CHEK2, Val455Ile*	79/81 (97.5%)	373/382 (97.6%)	0.95	0.20-4.50	1.000
PALB2, Glu272Lys*	1/82 (1.2%)	0/372 (0%)	-	-	0.181
PALB2, Tyr334Asp*	1/82 (1.2%)	4/380 (1.1%)	1.16	0.13-10.52	1.000
PALB2, Leu337Ser	6/82 (7.3%)	-	-	-	-
PALB2, Gln559Arg	10/82 (12.2%)	64/371 (17.3%)	0.67	0.33-1.36	0.323
PALB2, Val932Met	3/82 (3.7%)	-	-	-	-
PALB2, Gly998Glu	1/82 (1.2%)	14/372 (3.8%)	0.32	0.04-2.44	0.491
BRIP1, Leu195Pro	2/82 (2.4%)	-	-	-	-
BRIP1, Pro919Ser	32/82 (39.0%)	-	-	-	-
RAD50, Asp515Gly*	1/82 (1.2%)	4/384 (1.0%)	1.17	0.13-10.63	1.000

Abbreviations: CI, confidence interval; HBOC, hereditary breast and/or ovarian cancer; OR, odds ratio.
*Novel variant.

Bolded variants are known breast/ovarian cancer risk variants.

2 Germline copy number variations and their contribution to HBOC susceptibility (II)

In study II, a genome-wide SNP genotyping was performed to identify germline CNVs that confer susceptibility to HBOC. SNP genotyping was performed for index individuals from 84 HBOC families and 36 healthy controls. After applying strict sample quality criteria, 81 HBOC individuals and 35 controls were included in the data analysis. In total, 545 autosomal CNVs, 300 deletions and 245 duplications, were identified at 273 separate genomic regions in the analyzed HBOC individuals and controls. The primary focus was on the CNVs that affected potential or known genes in breast/ovarian cancer predisposition. Additionally, clinical features of the CNV carriers were inspected. After several filtering steps, 6 CNVs were selected for further validation by quantitative PCR. Five of these CNVs were genotyped in an additional cohort of 20 HBOC patients and in up to 869 additional controls by quantitative PCR (Table 6). The carrier frequencies of the validated CNVs were compared between 101 HBOC individuals and up to 899 controls (Table 6). Validated CNVs showed enrichment in HBOC cases compared to controls (Table 6). No statistically significant difference was observed, an effect that was probably due to the limited sample size used in this study. Therefore, the results require verification in a larger sample set.

Three of the validated CNVs were intronic deletions affecting *Erb-B2 Receptor Tyrosine Kinase 4 (ERBB4), EPH Receptor A3 (EPHA3),* and *CUB and Sushi multiple domains 1 (CSMD1). ERBB4* and *EPHA3* encode proteins that are important regulators in signaling pathways, whereas *CSMD1* is a tumor suppressor gene. A novel *ERBB4*-affecting deletion was slightly more frequent in HBOC individuals (5/101, 5.0%) than in controls (12/358, 3.4%) (OR=1.49 95%CI= 0.52-4.28) (Table 6). The *EPHA3*-affecting deletion was nearly twice as common in HBOC individuals (12/101, 11.9%) as in controls (27/432, 6.3%) (OR=1.96 95%CI=0.97-3.94) (Table 6). A novel *CSMD1*-affecting deletion was observed in one BC patient (1/101, 1.0%), but the variant was very rare in controls (1/436, 0.2%) (OR=4.33 95%CI=0.27-69.57) (Table 6). One of the validated CNVs was an intergenic deletion at a 5q15 region where regulatory elements were reported to be present according to the The Encyclopedia of DNA Elements database. The 5q15 deletion was observed

in a homozygous form in one BC patient (1/101, 1.0%) who exhibited an interesting clinical feature. Specifically, the patient had been diagnosed and died of BC at the age of 29. The 5q15 homozygous variant was very rare in controls (1/899, 0.1%) (Table 6). Two of the selected CNVs were exonic; one deletion affected *BRCA1, Neighbor Of BRCA1 Gene 1 (NBR1)* and *Neighbor Of BRCA1 Gene 2 (NBR2)* and one duplication affected *Endogenous Retrovirus Group V, Member 2 (ERVV-2).* Large deletions affecting *BRCA1* are known to predispose to hereditary breast and ovarian cancer, and the deletion was identified in one BC patient (1/101, 1.0%) (Table 6) with a family history of OC. *ERVV-2* belongs to the endogenous retrovirus group of proteins, and it has been implicated that in human cancer. The *ERVV-2*-affecting duplication was observed to be slightly more frequent in HBOC individuals (11/101, 10.9%) than in controls (34/334, 10.2%) (OR=1.37 95%CI 0.73-2.55) (Table 6). However, homozygous *ERVV-2*-affecting duplications were over four times more common in HBOC individuals (4/101, 4.0%) than in controls (3/334, 0.9%) (Table 6).

Some findings were made with respect to the clinical features of the HBOC individuals with the identified CNVs. For instance, a ductal tumor type was a common feature for all 3p11.1 deletion (*EPHA3* locus) carriers with BC (n=11). Of these tumors, 9/11 were ER and PR positive. Similarly, the ductal tumor type predominated in BC-affected females with a 19q13.41 duplication at the *ERVV-2* locus (10/11 carriers). Interestingly, 3/4 homozygous 19q13.41 duplication carriers had ductal high-grade tumors (grade III), implying a more aggressive disease. Two of five 2q34 deletion (*ERBB4* locus) carriers had a bilateral form of BC diagnosed at ≤43 years of age. Cosegregation of the novel 2q34 deletion with breast and/or ovarian cancer was studied in one family in which an intermediate breast and ovarian cancer risk variant, *BRCA1* Arg1699Trp, was identified (study I). The 2q34 deletion was identified in one OC-affected and two BC-affected relatives, some of whom were carriers of the *BRCA1* variant and some of whome were not.

Table 6. Association of the validated CNVs with hereditary breast and/or ovarian cancer risk.

CNV position, affected gene, type and size	Carrier frequency		OR; 95%CI	P
	HBOC individuals	Controls		
2q34 *ERBB4* intronic deletion[a] 28.7-59.0 kb	5/101 (5.0%)	12/358 (3.4%)	1.49; 0.52-4.28	0.457
3p11.1 *EPHA3* intronic deletion 14.6 kb	12/101 (11.9%)	27/432 (6.3%)	1.96; 0.97-3.94	0.055
5q15 intergenic deletion 49.8 kb	5/101 (5.0%)[b]	57/899 (6.3%)[c]	0.92; 0.39-2.16	0.845
8p23.2 *CSMD1* intronic deletion[a] 10.8 kb	1/101 (1.0%)	1/436 (0.2%)	4.33; 0.27-69.57	0.259
17q21.31 *BRCA1, NBR1, NBR2* exonic deletion[d], 99.0 kb	1/101 (1.0%)	0/35) (0%)	-	0.555
19q13.41 *ERVV-2* exonic duplication 15.8-26.9 kb	11/101 (10.9%)[e]	34/334 (10.2%)[f]	1.37; 0.73-2.55	0.322

Abbreviations: CI, confidence interval; CNV, copy number variation; HBOC, hereditary breast and/or ovarian cancer; OR, odds ratio.

[a] Novel variant.

[b] Homozygous in 1/101 (1.0%) of the HBOC individuals.

[c] Homozogyous in 1/899 (0.1%) of the controls.

[d] Large *BRCA1*-affecting deletions are known to be associated with HBOC susceptibility and there was no need to genotype additional controls.

[e] Homozygous in 4/101 (4.0%) of the HBOC individuals.

[f] Homozygous in 3/334 (0.9%) of the controls.

3 Identification of HBOC susceptibility genes and gene variants by exome sequencing (III)

In study III, with the aim of identifying gene variants that contribute to HBOC susceptibility, whole exome sequencing was performed for 37 individuals from 13 high-risk *BRCA1/2*-negative HBOC families. Of the studied individuals, 23 were female breast or breast and ovarian cancer patients, one was a male BC patient, and 13 were healthy relatives. Of note, six of the BC patients were diagnosed at a very early age ($\leq$ 29 years). In total, 736,963 sequence variants were detected in 37 individuals. Several filtering steps were used to reduce the number of candidate variants for downstream analyses. The primary focus was on DDR pathway gene variants that were shared between affected family members or families. Eighteen candidate variants were selected for further analyses, genotyping of these variants was performed in a cohort of 129 female HBOC cases and up to 989 healthy female controls (Table 7). Additionally, two of the variants, which were detected in a male BC patient, were also screened in a cohort of 49 male BC patients and 909 healthy male controls (Table 7). Carrier frequencies of the variants were compared between the cancer cases and controls (Table 7).

Five variants, *ATM* D1853V, *MYC* N26S, *PLAU* V15L, *RAD1* G114D, and *RRM2B* R71fs, were enriched in female HBOC cases compared to controls (OR 1.16-2-16) implying that the variants may be low-to-moderate risk alleles (Table 7). A *BRCA1* T1302P variant, detected in a two affected females in a single BC family by exome sequencing, was absent in both female HBOC patients and female controls (Table 7). This result implies that *BRCA1* T1302P is an extremely rare BC susceptibility variant. The *RAD50* I94L variant, which was detected in a male BC patient, was absent in a cohort of male BC cases and extremely rare in male controls (1/909, 0.1%) (Table 7), suggesting it may contribute to male BC risk. The other variant detected in a male BC patient, *ATM* Y1475C, was absent both in male BC cases and male controls but was detected in female controls (Table 7). This results suggests that *ATM* Y1475C is not specific for male BC. No statistically significant association was reached for any of the variants, a result that can likely be explained by the rareness of the variants and the limited sample size. Therefore, the results need further verification. Eighteen variants were also screened from breast tumor

tissue samples from an additional 31 BC patients, but no wild type allele loss was observed.

Table 7. Validated variants and their association with breast/ovarian cancer risk.

| | Carrier frequency | | | | | |
| | Females | | Males | | | |
Gene, variant	HBOC cases	Controls	BC cases	Controls	*P*	OR; 95%CI
AKT2, P50T	2/127	5/280	-	-	1	0.88; 0.17-4.57
ATM, F858L	1/129	10/975	-	-	1	0.75; 0.10-5.92
ATM, P1054R	1/129	14/981	-	-	1	0.54; 0.07-4.13
ATM, Y1475C	0/129	1/278	0/49	0/909	1/1[e]	na
ATM, D1853V	1/129	5/989	-	-	0.52	1.54; 0.18-13.19
BRCA1, T1302P	0/128	0/986	-	-	1	na
CDKN2A, H166Y	3/129	7/280	-	-	1	0.93; 0.24-3.62
MYC, N26S	5/129[b]	23/987	-	-	0.14	2.02; 0.81-5.01
NCOA3, Q1118P	0/129	7/279	-	-	0.10	na
PLAU, V15L	2/129	11/984	-	-	0.60	2.16; 0.30-15.45
RAD1, G114D	5/129	15/464	-	-	0.79	1.16; 0.42-3.22
RAD50, I94L	0/129	0/187	0/49	1/909	1/1[e]	na
RAD52, G180R	4/129	15/269	-	-	0.33	0.55; 0.18-1.67
RBL2, E575Q	8/129	22/261	-	-	0.55	0.73; 0.32-1.66
RPA2, S41F	0/129	5/467	-	-	0.59	na
RRM2B, R71fs[a]	16/128[c]	22/247[d]	-	-	0.31	1.39; 0.73-2.64
WNT3A, D93N	1/129	4/468	-	-	1	0.91; 0.10-8.15
WNT10A, R113C	1/129	10/988	-	-	1	0.77; 0.10-6.00

Abbreviations: BC, breast cancer; CI, confidence interval; HBOC, hereditary breast and/or ovarian cancer; OR, odds ratio.
[a] Novel variant
[b] Homozygous in 1/129 of the female HBOC cases
[c] Homozygous in 1/128 of the female HBOC cases
[d] Homozygous in 2/247 of the female controls
[e] Females/Males

An interesting observation related to the *MYC* N26S variant was made in a single family by exome sequencing. Genotyping detected the *MYC* variant in five female HBOC patients (5/129, 3.9%), of whom one carried a homozygous form of the variant (Table 7). Interestingly, a homozygous variant carrier was diagnosed with

triple-negative BC at the age of 61 years and a family history of multiple cancers (three BCs, colon cancer, and possibly throat and stomach cancers). Further analysis identified the homozygous *MYC* variant in the index's healthy sister (current age 81 years) and in heterozygous form in the index's healthy young relatives. Unfortunately, no DNA samples of BC-affected relatives were available at that time. Of note, the homozygous form of the *MYC* variant was not detected among 987 healthy controls. A novel frameshift-inducing duplication in *RRM2B* was found to be a very common variant in both female HBOC patients and healthy controls (Table 7), although it was originally considered rare based on the allele frequencies in the databases. A homozygous form of the *RRM2B* variant was also detected, but the frequencies were similar both in female HBOC cases and controls (Table 7). Another interesting observation related to the *ATM* D1853V variant. The *ATM* D1853V was reported to affect splicing. The clinical features of the variant carriers were inspected. One interesting finding was that one of the *PLAU V15L* variant carriers, detected by genotyping, exhibited a mucinous subtype of BC (grade 2, ER+/PR+/HER2-) that was diagnosed at the age of 55. The mucinous subtype of BC is rare and is associated with a good prognosis.

Specific attention was also given to six early-onset BC patients (diagnosed ≤ 29 years) in the exome sequencing cohort. The aim of this analysis was to identify variants that were present only in these early-onset patients and could explain the drastic clinical outcome. After several filtering steps, variants of host genes that participate in certain pathways were considered good candidates for BC susceptibility. These pathways and functions included the cell cycle, proliferation, apoptosis, adhesion, different signaling pathways, and the DNA damage response. The candidate variants are presented in Table 8. Of these, 50 were non-synonymous SNVs, 4 were stop-gains, and 3 were frameshift indels. The majority of the variants were detected in a single patient only. Nonsynonymous variants in genes such as *Cyclin-Dependent Kinase 2 Interacting Protein (CINP), Focadhesin (FOCAD), Laminin, Alpha 5 (LAMA5), Phospholipase D1, Phosphatidylcholine-Specific (PLD1), RBL2* and *Sphingosine-1-Phosphate Receptor 5 (S1PR5)* were identified in two of six patients (Table 8). Additionally, enrichment was observed in certain pathway gene variants in the same individual. For instance, enrichment was observed for extracellular matrix (ECM)-receptor interaction and focal adhesion pathway genes. Additionally, of special interest were variants that 1) induced a premature stop codon in 4 genes, including *DENN/MADD Domain Containing 2D (DENND2D), EF-Hand Calcium Binding Domain 13 (EFCAB13), Epithelial Stromal Interaction 1 (Breast) (EPSTI1)*, and *TOPBP1-Interacting Checkpoint And Replication Regulator (TICRR)*; and 2) caused

frameshifts in 3 genes, including *BCL2/Adenovirus E1B 19 kD Interacting Protein Like (BNIPL), Endothelin 3 (EDN3)* and *Melanoma Antigen Family F1 (MAGEF1)* (Table 8).

Table 8. Candidate variants in early-onset breast cancer patients.

Gene	Variant	*N* in patients	Pathway or function
EPSTI1	E395_L396delinsAX	1	apoptosis
MBD4	I358T	1	base excision repair
NEIL3	Q172H	1	base excision repair
EFCAB13	K337X	1	calcium ion binding
MAD1L1	R59C	1	cell cycle, progesterone-mediated oocyte maturation
CDKN2B	A19D	1	cell cycle, TGF beta signaling pathway, pathways in cancer
FANCD2	G901V	1	DNA damage response
MAP3K4	H906P	1	DNA damage response, MAPK signaling pathway
CHEK2	I157T	1	DNA damage response, p53 signaling pathway, cell cycle
RBL2	D33A	2	DNA damage response, TGF beta signaling pathway, cell cycle
RBL2	A34P	2	DNA damage response, TGF beta signaling pathway, cell cycle
RBL2	E60G	1	DNA damage response,TGF beta signaling pathway, cell cycle
TICRR	R665X	1	DNA replication
CDC45	E109G	1	DNA replication, cell cycle
CINP	N53K	2	DNA replication, checkpoint signaling
LIG1	V281M	1	DNA replication, mismatch repair, base and nucleotide excision repair
TNC	V548M	1	ECM-receptor interaction, focal adhesion
TNC	V993L	1	ECM-receptor interaction, focal adhesion
COL11A2	L11H	1	ECM-receptor interaction, focal adhesion
COL6A2	D227N	1	ECM-receptor interaction, focal adhesion
COL4A6	I1161V	1	ECM-receptor interaction, pathways in cancer, focal adhesion
LAMA1	R729H	1	ECM-receptor interaction, pathways in cancer, focal adhesion
LAMA5	R1679W	2	ECM-receptor interaction, pathways in cancer, focal adhesion
LAMA5	A1021V	1	ECM-receptor interaction, pathways in cancer, focal adhesion
LAMA5	R2456H	1	ECM-receptor interaction, pathways in cancer, focal adhesion
LAMA5	S2138I	1	ECM-receptor interaction, pathways in cancer, focal adhesion
LAMB1	D957N	1	ECM-receptor interaction, pathways in cancer, focal adhesion
LAMB2	G436S	1	ECM-receptor interaction, pathways in cancer, focal adhesion
LAMC3	R563W	1	ECM-receptor interaction, pathways in cancer, focal adhesion
MAGEF1	E18fs	1	enhancer of ubiquitin ligase activity
AKAP13	G191R	1	G protein signaling pathways

Table 8.   Continued.

Gene	Variant	*N* in patients	Pathway or function
AKAP8	N505D	1	G protein signaling pathways
PLD1	R398C	2	glycerophospholipid metabolism, pathways in cancer
DHH	P9A	1	hedgehog signaling pathway
LRP2	T2284A	1	hedgehog signaling pathway
LRP2	P1703S	1	hedgehog signaling pathway
RASGRP3	G282S	1	integrated cancer pathway, MAPK signaling pathway
BNIPL	T11fs	1	interacts with BCL2, promotes cell death
EXO1	N279S	1	mismatch repair
NLRP4	G638R	1	NOD pathway
PRDM1	P580L	1	NOD pathway
TAB3	T248M	1	NOD-like receptor signaling pathway
DTX4	R415C	1	notch signaling pathway
NUMBL	F449L	1	notch signaling pathway
RET	R959Q	1	pathways in cancer
DENND2D	R16X	1	promotes the exchange of GDP to GTP
S1PR5	L318Q	2	signal transduction of S1P receptor
RICTOR	D1074G	1	TOR signaling, mTOR signaling pathway
APEX1	I64V	1	TSH signaling pathway, base excision repair
FOCAD	A1683T	2	tumor suppressor in glioma and colorectal cancer
FBXW8	T470M	1	ubiquitin mediated proteolysis
UBE2Q1	N243H	1	ubiquitin mediated proteolysis
UBE3A	A178T	1	ubiquitin mediated proteolysis
BRCA1	P1052L	1	ubiquitin mediated proteolysis, DNA damage response
BIRC6	E892G	1	ubiquitin mediated proteolysis, apoptosis modulation and signaling
EDN3	E187fs	1	variety of cellular roles including proliferation, migration, differentiation
SOX17	G28V	1	wnt signaling pathway

DISCUSSION

1 Contribution of variants in well-known breast cancer susceptibility genes to high-risk Finnish HBOC families (I, II, III)

Mutations in well-known BC susceptibility genes, such as *BRCA1, BRCA2, ATM, PALB2,* and *CHEK2*, explain approximately 20-30% of the genetic predisposition to BC (Couch et al, 2014, Mavaddat et al, 2010, Turnbull & Rahman, 2008). *BRCA1* and *BRCA2* make a major contribution to this predisposition and in Finland contribute to BC in approximately 20% of HBOC families (Hartikainen et al, 2007, Vehmanen et al, 1997a). Additionally, Finnish founder mutations have been identified in *CHEK2*, *PALB2*, or *RAD50* and contribute to a fraction of BC cases in Finnish HBOC families (Erkko et al, 2007, Heikkinen et al, 2006, Vahteristo et al, 2002). However, the genetic predisposition factors remain unknown in majority of the high-risk families. Well-established clinical management strategies exist for *BRCA1* and *BRCA2* mutation carriers (Couch et al, 2014). However, the clinical management of high-risk *BRCA1/2*-negative HBOC patients is problematic. Therefore, new information of breast and ovarian cancer predisposing factors is urgently needed to improve the clinical assessment of high-risk families and to widen the genetic testing and guidance protocol for other genes and mutations as well.

Additional variants were identified in *BRCA1* and *BRCA2* genes in HBOC individuals who had been tested to be founder mutation-negative for both of these genes. The detected variants in these two genes were reported to contribute to BC in a fraction of the high-risk Finnish HBOC families. The *BRCA1* Arg1699Trp variant, which was detected in study I in three BC-affected females in a single HBOC family, has been classified as clinically significant variant in the Breast Cancer Information Core database. This variant was further confirmed to be an intermediate

breast and ovarian cancer risk variant by functional analyses (Spurdle et al, 2012). Additionally, two other nonsynonymous variants, *BRCA1* Met1628Thr and *BRCA2* Val2728Ile, which were detected in study I, showed enrichment in HBOC families compared to controls. This result implies that these variants may be low-to-moderate risk alleles. Although the clinical significance of both variants is unknown (Deffenbaugh et al, 2002, Ostrow et al, 2004, Phelan et al, 2005), further study is clearly warranted. Additionally, a rare and possibly pathogenic variant in *BRCA1*, T1302P, was detected in study III in two BC-affected relatives in a single family. The role of this variant in BC predisposition is suspected given that it was not observed among healthy controls. Additionally, one novel possibly pathogenic *BRCA2* variant (Leu24Phe) was detected in a single HBOC family in study I. The novel variant was particularly interesting as it is located in the N-terminal portion of BRCA2, which is an interaction site for PALB2. PALB2 is essential for key BRCA2 functions, and sequence variants disrupting this interaction have been reported to play a role in cancer predisposition (Xia et al, 2006). Thus, additional analyses of the Leu24Phe variant are warranted. Moreover, a large deletion affecting *BRCA1* was detected in study II in a BC patient with family history of ovarian cancer. The same deletion, which removes most of the gene, including the promoter, and prevents transcription of *BRCA1*, has been reported in the previous Finnish study of ovarian cancer family (Pylkas et al, 2008). Large genomic rearrangements in both *BRCA1* and *BRCA2* are rare, but the findings confirm the contribution of *BRCA1* rearrangements to a fraction of the HBOC families in the presence of multiple cases of ovarian cancer.

An important finding in study I was that the two known BC-associated *CHEK2* variants, Ile157Thr, and *CHEK2* 1100delC, contributed to BC predisposition in the Finnish *BRCA1/2*-negative HBOC families to a remarkable degree (10/82, 12.2%); these results suggest the clinical relevance of these variants. Among heterozygous 1100delC carriers with a family history of BC, the cumulative risk of BC by age 70 is estimated to be 37%, which is comparable to the BC risk among *BRCA1* and *BRCA2* mutation carriers (Weischer et al, 2008). Moreover, 1100delC carriers seem to have poorer disease-free and overall survival than non-carriers and an increased risk of developing a second, primarily contralateral, BC (Ripperger et al, 2009). Therefore, genotyping of the 1100delC variant for clinical assessment of BC risk has been suggested particularly in Northern and Eastern European populations in which the *CHEK2* variants are observed with high frequencies (Weischer et al, 2008). The 1100delC increases the BC risk approximately two-fold, whereas the Ile157Thr is the lower risk variant (CHEK2 Breast Cancer Case-Control Consortium, 2004, Kilpivaara et al, 2004). Thus, the *CHEK2* variants are suggested to act in

combination with other susceptibility alleles to multiply the BC risk (i.e., a polygenic risk model) (CHEK2 Breast Cancer Case-Control Consortium, 2004). Here, a particularly interesting finding was that both of the *CHEK2* variants, 1100delC and Ile157Thr, were observed in the same family, and a female patient carrying both of the variants exhibited a dramatic clinical outcome (BC at the age of 26 years). This set of results suggest a multiplicative effect of these two variants. Moreover, bilateral BC was diagnosed in one of the three patients with 1100delC variant, which is in line with previous observations. However, due to unclear clinical consequences related to the incomplete segregation of the *CHEK2* variants with BC and the fact that these variants are very rare in most populations, genetic testing of *CHEK2* variants has not been justified in the clinical practice (Ripperger et al, 2009). Here, the results favor a more profound segregation analyses of the two *CHEK2* variants, 1100delC and Ile157Thr, in the high-risk Finnish *BRCA1/2*-negative hereditary BC families.

ATM has been considered a moderate-risk BC gene, and heterozygous mutations have been reported to confer an approximately two-fold elevated BC risk (Ahmed & Rahman, 2006). An *ATM* D1853V (c.A5558T) variant was reported to contribute a low-to-moderate BC predisposition in study III. The variant was identified altogether in four HBOC families. Interestingly, a common *ATM* polymorphism (c.G5557A) was observed among two of the c.A5558T variant carriers at the adjacent nucleotide position. The c.G5557A variant is a common polymorphism. The association of this variant with BC has been widely studied and has been associated with bilateral BC in the Finnish study (Heikkinen et al, 2005). In line with these results, one of the BC patients carrying both of the *ATM* variants, c.G5557A and c.A5558T, had been diagnosed with bilateral disease. Furthermore, the c.G5557A variant has been reported to have an effect on splicing (Thorstenson et al, 2003); a similar observation was seen for the c.A5558T variant, which it was predicted to create a new splicing site. Splicing mutations have been reported to be particularly common in *ATM* (Thorstenson et al, 2003), which makes the c.A5558T variant an intriguing candidate. Therefore, further studies are necessary to fully analyze the role of the c.A5558T variant in splicing, either alone or in combination with the c.G5557A variant.

BC-associated Finnish founder mutations have been reported in *PALB2* and *RAD50* genes (Erkko et al, 2007, Heikkinen et al, 2006), but these founders were not detected in the present studies. This difference, is likely explained by the rarity of the mutations and the limited number of the studied families. Instead, a *RAD50* I94L variant was reported to contribute to male BC in study III, but this finding warrants further confirmation. There was no evidence that the identified coding

variants in the well-known BC susceptibility genes *PALB2, BRIP1, RAD50*, and *CDH1* contributed to female breast or ovarian cancer risk in the analyzed HBOC families.

2 Contribution of germline copy number variations to HBOC susceptibility and the identification of candidate genes (II)

Copy number variation is defined as situation in which a segment of DNA (1 kb or larger) shows an altered copy number compared to the reference genome (Feuk et al, 2006). CNVs are known to play a role in cancer predisposition, but their role in HBOC is largely unexplored (Krepischi et al, 2012a, Krepischi et al, 2012b). By studying CNV disrupted genes, defective biological processes and novel candidate genes underlying breast/ovarian cancer predisposition can be detected. A previous Finnish study provided evidence of rare CNVs that contribute to BC susceptibility by identifying disrupted genes in estrogen signaling and the TP53 tumor suppressor network (Pylkas et al, 2012).

The current study identified CNV-affected genes as good candidates for HBOC susceptibility. Of particular interest were *ERBB4* at 2q34, *EPHA3* at 3p11.1, *CSMD1* at 8p23.2, and *ERVV-2* at 19q13.41.

ERBB4 encodes an epidermal growth factor receptor tyrosine kinase subfamily member that functions in several cellular processes, including proliferation, survival, migration, and differentiation (Sundvall et al, 2008). ERBB4 has a role in mammary carcinogenesis and has been suggested to have both tumor-promoting and tumor-suppressing functions (Sundvall et al, 2008). An *ERBB4*-affecting deletion was observed in 5% (5/101) of the HBOC families. Segregation analysis of the novel *ERBB4*-affecting intronic deletion in one HBOC family suggested that the deletion may be a low-risk variant and may modify cancer risk when occurring in combination with the intermediate breast and ovarian cancer risk variant, *BRCA1* Arg1699Trp, which was identified in study I.

EPHA3 encodes a protein that functions in signaling pathways. EPHA3 belongs to the ephrin receptor subfamily of the receptor tyrosine kinase family, which has central roles in normal cell physiology and disease pathogenesis (Pasquale, 2008). Ephrin receptor signaling and ephrin ligands regulate tumor growth in variety of cancers, including BC (Pasquale, 2010). Additionally, altered expression levels of EPHA3 have been associated with gastric and colorectal cancers, and CNVs in the

EPHA3 region have been reported in hematological malignancies (Guan et al, 2011, Xi & Zhao, 2011, Xi et al, 2012). Moreover, CNVs at the *EPHA3* region have been implicated in predisposition to hereditary prostate cancer (Laitinen et al, 2015), making this CNV a plausible candidate also for HBOC susceptibility. Here, the *EPHA3*-disrupting deletion occurred in 11.9% (12/101) of the HBOC families.

CSMD1 is a known tumor suppressor gene. The loss of CSMD1 has been primarily associated with head and neck squamous cell carcinoma but has also been observed in several epithelial cancers, including BC (Ma et al, 2009). Furthermore, the decreased expression of CSMD1 has been associated with a high-tumor grade and poor survival in invasive ductal breast carcinoma; this gene has also been used as a prognostic marker (Kamal et al, 2010). Therefore, *CSMD1* is an excellent candidate gene for HBOC susceptibility. Here, the *CSMD1*-affecting deletion was rare and was observed in a single (1/101, 1.0%) BC family

ERVV-2 encodes a protein in the human endogenous retrovirus (HERV) group. HERVs and related genetic elements occupy approximately 8% of the human genome, but their role in human cells is poorly understood (Suntsova et al, 2015). HERVs encode active retroviral proteins that are likely involved in important physiological functions and may be involved in the progression of cancer and several other human diseases (Suntsova et al, 2015). Moreover, HERVs regulate the expression of neighboring host genes and modify the genomic regulatory landscape (e.g., transcription factor binding sites) (Suntsova et al, 2015). Therefore, *ERVV-2* is a very interesting candidate gene in HBOC susceptibility, providing an intriguing link between breast and ovarian cancer predisposition and endogenous retroviruses. Here, the *ERVV-2*-affecting duplication was observed in 10.9% (11/101) of the HBOC families. A particularly interesting observation was that the homozygous form of the duplication (observed in 4/101, 4.0% of the HBOC families) was over four times more common in HBOC cases compared to controls and seemed to correlate with high-grade tumors.

One interesting CNV was detected at the non-genic 5q15 region. The CNV was considered interesting due to clinical outcome of the single HBOC patient (diagnosed and died of BC at the age of 29) carrying the homozygous form of the CNV. The homozygous deletion at 5q15 was extremely rare in controls (1/899, 0.1%), implying that it may be disease-related and may have clinical significance for families with early-onset BC cases. This intergenic deletion may disrupt the transcriptional control of target gene expression, and this hypothesis was supported by the finding that regulatory element activity is present at the 5q15 deletion region. Gene expression regulation is a complex process, and regulatory elements can be far

from the target genes (Kleinjan & van Heyningen, 2005). An interesting observation was that *Repulsive Guidance Molecule Family Member B (RGMB),* the aberrant expression of which has been found in BC (Li et al, 2011), was observed to be the nearest neighboring gene in the 5q15 deletion region (1.0 Mb). Further studies are necessary to identify the target gene of the possible regulatory elements that are disrupted by the deletion at 5q15.

3 Identification of novel candidate genes and gene variants in high-risk HBOC families (III)

3.1 DNA damage response pathway

DDR pathway is one of the central pathways in breast and ovarian cancer predisposition. Several susceptibility genes from this pathway have been recognized including *BRCA1, BRCA2, ATM, PALB2, BRIP1, CHEK2,* and *RAD50* (Ciccia & Elledge, 2010, Jackson & Bartek, 2009). Since the DDR pathway is a complex network consisting of numerous proteins and complexes, there are likely unknown candidates in this pathway that contribute to breast and ovarian cancer susceptibility.

In study III, the exome sequencing approach was utilized to analyze 13 high-risk HBOC families, and the primary focus was in DDR pathway gene variants that were shared between affected family members or between families. Family-specific enrichments of multiple rare DDR pathway gene variants were observed, confirming the central role of the DDR pathway defects in HBOC families. Moreover, different combinations of possible low-to-moderate risk variants in critical pathway genes were implicated in cancer predisposition and phenotypic variation (e.g., disease onset) in the analyzed families. One such interesting combination was observed in a family in which three BC-affected females were exome sequenced. Two females who had been diagnosed with BC at a later age (>50 years) carried variants in both *CDKN2A* (H166Y) and *AKT2* (P50T), whereas the third female who had early-onset disease (diagnosed at age 28) carried a variant (G901V) in *Fanconi Anemia, Complementation Group D2 (FANCD2)*. AKT2 is a serine threonine kinase that functions in a key signaling pathway in mammary gland development and cancer (Wickenden & Watson, 2010). *CDKN2A* is a tumor suppressor gene in melanoma. Germline defects in *CDKN2A* have been reported in female BC patients with melanoma (Nagore et al, 2009). Of note, melanoma was not observed among BC patients in the family in which the *CDKN2A* variant was observed but one unconfirmed melanoma case was reported in their male relative. Moreover, *FANCD2* belongs to the Fanconi Anemia complementation gene group, in which biallelic mutations predispose to Fanconi Anemia (Schneider et al, 2015). Heterozygous mutations in several Fanconi Anemia genes, such as *FANCD1*

(BRCA2), *FANCN (PALB2),* and *FANCM,* have been associated with BC predisposition (Erkko et al, 2007, Kiiski et al, 2014, Wooster et al, 1994). These results make *FANCD2* an interesting candidate gene to further study in terms of BC susceptibility, particularly in early-onset cases.

The most promising candidate variants (n=18) in DDR genes were observed in additional cohorts of female HBOC cases, male BC cases, breast tumor samples, and controls. Due to the rarity of these variants and the limited number of analyzed cases, no statistically significant association with the disease was reached. In addition, no wild type allele loss was observed for any of the variants in breast tumor samples. However, five variants, *ATM* D1853V*, MYC* N26S*, PLAU* V15L*, RAD1* G114D, and *RRM2B* F71fs, were enriched in HBOC cases compared to controls (OR 1.16-2-16). These results suggest that the variants may confer low-to-moderate risk for breast and ovarian cancer and may contribute to a fraction of the cases in HBOC families. However, further confirmative studies for the variants are warranted. All five genes in which the variants occur play important roles in the DDR pathway, making them good candidates genes for disease susceptibility. In addition to *ATM*, which is a well-known BC susceptibility gene, *MYC* is an established oncogene that plays a role in a variety of human cancers (Hoffman & Liebermann, 2008). *PLAU* encodes a protein that plays a role in cancer metastasis (Moquet-Torcy et al, 2014). *RAD1* encodes a component of the cell cycle checkpoint complex that participates in cell cycle checkpoint activation and DNA repair (Xu et al, 2009), and *RRM2B* encodes a protein involved in the p53 checkpoint for the repair of damaged DNA (Tanaka et al, 2000).

3.2   Other pathways

Pathways related to cell cycle, proliferation, apoptosis, signaling, and adhesion are interesting candidate pathways for breast and ovarian cancer susceptibility, although these pathways are primarily unexplored. By focusing on six exomes from early-onset BC patients (diagnosed ≤ 29 years), rare pathogenic variants in these potential pathway genes were explored. For instance, deleterious frameshift variants were detected in *BNIPL, EDN3*, and *MAGEF1*, and variants that induce a premature stop codon were detected in *DENND2D, EFCAB13, EPST11,* and *TICRR.* Of these, *BNIPL* encodes an apoptosis-associated protein that interacts with B-Cell CLL/Lymphoma 2 (BCL2) and promotes the invasion and metastasis of human hepatocellular carcinoma cells (Qin et al, 2003, Xie et al, 2007), whereas EDN3 is an

important signaling molecule that participates in several key cellular processes, such as proliferation, migration, and differentiation (Levin, 1995). Furthermore, *DENND2D* is involved in cell signaling, and altered expression levels of this gene have been reported in cancers (Hibino et al, 2014). *TICRR* encodes a protein that is a crucial regulator of DNA replication and cell cycle checkpoints (Sansam et al, 2010). Furthermore, non-synonymous variants were observed in *CINP, FOCAD*, and *Exonuclease 1 (EXO1),* which are also of great interest. *CINP* is involved in DNA replication and checkpoint signaling (Lovejoy et al, 2009). *FOCAD* is a tumor suppressor gene that has been associated with polyposis and colorectal cancer development (Weren et al, 2015). Furthermore, *EXO1* plays a role in mismatch repair, and its role in susceptibility to BC has been reported (Michailidou et al, 2015). The novel findings encourage the study of other pathways in breast/ovarian cancer predisposition and provide information on potential candidate genes. These data can be used an excellent foundation for future studies.

4 Limitations of the study

Breast and ovarian cancers are genetically heterogeneous diseases in which numerous genetic factors play multiplicative roles. Genetic predisposition factors are often rare and family specific, and there is phenotypic variation within families, making the identification of predisposition factors challenging. Sample selection is one of the critical steps when identifying disease-causing genetic factors. In the current study, breast and/or ovarian cancer families were selected according to strict high-risk hereditary BC criteria. There criteria were chosen to select families with similar characteristics and that are most likely to share common hereditary defects. Although well-selected high-risk HBOC families were used in this study, the number of analyzed patients was somewhat limited and the statistical analyses lacked significant association results. For these reasons, these findings require further confirmation in a larger sample sets. Moreover, the availability of samples from both affected and healthy relatives was a limiting step in the segregation analyses of certain variants. Thus, there is a need to continue recruiting more high-risk HBOC families and more relatives from the studied families. Additionally, tumor samples from these patients would be essential to further study the deleteriousness of the variants identified in the germline. When using large sample cohorts, clinically different subtypes (e.g., tumors with hormone receptor status) can be utilized. For instance, a previous Finnish study identified a novel BC susceptibility gene to be particularly associated with triple-negative BC (Kiiski et al, 2014). Here, for example, the results implicated that copy number variation at 3p11.1 would be more common in BC patients with ductal and estrogen- and progesterone-receptor positive tumor. Obviously, this result should be confirmed in the larger cohort of BC patients to see the trend more clearly. Overall, the Finnish population, which is a homogenous group, provides an excellent basis for genetic studies. Although, it should be noted that there can be great variation in the mutation spectrum between different parts of Finland. For instance, the prevalence of *BRCA1* and *BRCA2* founder mutations varies across the country (Hartikainen et al, 2007, Huusko et al, 1998, Sarantaus et al, 2000).

Since the identification of the two major breast and ovarian cancer susceptibility genes, *BRCA1* and *BRCA2*, a variety of different methodological approaches have been utilized in susceptibility gene studies. Each of the approaches have their

strengths and limitations. In the current study, three different approaches were utilized, including a candidate gene approach, genome-wide copy number variation analysis and exome sequencing. The mutational screening of candidate genes is accurate but often costly and time-consuming when using, for example, direct sequencing. Genome-wide approaches provide a cost-effective and rapid approach for the detection of a massive number of genetic variants in a single run. When using genome-wide approaches, the challenge is to use appropriate filtering strategies to trim down the number of candidate variants. If using filtering criteria that are too stringent, disease-causing variants can be excluded already in the data-analysis phase. Most commonly, variants are filtered based on allele frequency, predicted pathogenicity, and function. Therefore, it is possible that causative variants were missed in this study during these filtering steps. Moreover, exome sequencing detects only variants that are located in the protein-coding region. Therefore, causative variants that are located outside the coding region and affect, for example, transcription, may have been missed. Additionally, focusing on certain pathway genes (study III) might be a useful step in reducing the number of candidate variants. However, variants in other pathways remain unexplored. Therefore, a subset of exome-sequenced patients (study III) that included early-onset BC cases were re-analyzed. In this analysis, variants in all pathways were considered potential candidates when identifying novel candidate genes and pathways that likely play a role in BC pathogenesis.

5   Future prospects

There is still a very large gap in knowledge concerning genetic predisposition factors in high-risk HBOC families that do not carry defects in *BRCA1* or *BRCA2*. There is an immense need to obtain novel information of additional breast and ovarian cancer susceptibility genes and gene defects to identify at-risk individuals as early as possible. In this way, these individuals and their family members can be provided with more efficient prevention, screening strategies, and therapeutic options. For instance, patients with defective genes in the DNA-damage repair pathway, including *BRCA1* and *BRCA2* mutation carriers, benefit from treatment with Poly (ADP-ribose) polymerases (PARP) inhibitors (Livraghi & Garber, 2015).

Novel next-generation sequencing technologies provide fast and cost-effective applications for the detection of genetic predisposition factors. Whole exome sequencing, which focuses on only the protein-coding region of the genome, is currently an attractive method for identifying novel susceptibility genes by (Ng et al, 2009). When additional technologies are developed and costs are reduced, whole genome sequencing will be a major method for revealing the remaining genetic defects that underlie these diseases. NGS technologies have been applied in clinical settings as well. For instance, NGS has been shown to be an efficient tool in HBOC diagnostics (Castera et al, 2014, Trujillano et al, 2015). However, the major challenge related to NGS lies in transferring sequencing data into medical diagnoses. There is a need to develop appropriate and validated guidelines related to data-analysis and the interpretation of variants that are of unknown significance (Rehm et al, 2013). Additionally, the amount of data will increase notably for whole genome sequencing, requiring appropriate storage and handling strategies in the clinical setting. Furthermore, when the number of candidate variants discovered through NGS increases, there will be a need to functionally validate the findings and design novel clinical assessment guidelines.

Moreover, multinational collaborations, such as the Collaborative Oncological Gene-environment Study (Collaborative Oncological Gene-environment Study), are likely to play an essential role in identifying predisposition factors for breast and ovarian cancer. These types of consortia are able to study a massive number of patients from a variety of populations. Additionally, new technologies have made it

possible to collect variety of different sample types from patients. The collected samples and specific clinical information of the patients are increasingly being stored in biobanks, providing an excellent source for research purposes. Moreover, sequencing efforts have led to the creation of several databases, such as the 1000 Genomes (1000 Genomes Project Consortium et al, 2012) and The Cancer Genome Atlas (The Cancer Genome Atlas). These provide large datasets that can be freely used in cancer genetic studies. It is likely that the available information in these and other databases will increase massively in the coming years and will provide unprecedented possibilities in terms of integrating data from different sources.

Constantly increasing genetic information will enable healthcare to move towards well-tailored medical care. In the future, it will likely be possible to design personalized screening, prevention, and therapeutic strategies and thereby optimize the clinical management of hereditary breast and ovarian cancer.

CONCLUSIONS

The current study was conducted to provide novel information about the genetic factors predisposing to HBOC in high-risk *BRCA1/2* founder mutation-negative Finnish families.

The major findings of this study were following:

1. Previously known breast cancer-associated mutations in *BRCA1* and *CHEK2* contributed to 13.4% of the HBOC families. Proportion of the *CHEK2* mutations was remarkable and clinically relevant. A novel possibly pathogenic variant was identified in *BRCA2*.

2. Copy number variations at 3p11.1, 5q15, 8p23.2, and 19q13.41 were of special interest and their role in HBOC predisposition was indicated. Deletions at 3p11.1 and 8p23.2 affected intronic regions of *EPHA3* and *CSMD1* genes, whereas duplication at 19q13.41 disrupted the coding region of *ERVV-2* gene. Additionally, deletion at 5q15 occurred at intergenic region and was reported to affect regulatory elements.

3. Five variants in DNA damage response pathway genes *ATM, MYC, PLAU, RAD1*, and *RRM2B* may act as low-to-moderate risk alleles. A rare variant that may have clinical relevance was detected in *BRCA1*. Additionally, a rare variant in *RAD50* was suggested to predispose to male breast cancer. Variants in novel candidate genes targeting DNA repair and replication, signaling, apoptosis, and cell cycle were observed in early-onset breast cancer patients. Novel candidate genes included, for example *DENND2D, TICRR, BNIPL, EDN3,* and *FOCAD*.

ACKNOWLEDGEMENTS

Tampere, January 2016

Kirsi Määttä

REFERENCES

1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, & McVean GA (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**: 56-65

Ahmed M & Rahman N (2006) ATM and breast cancer susceptibility. *Oncogene* **25**: 5906-5911

Allinen M, Launonen V, Laake K, Jansen L, Huusko P, Kaariainen H, Borresen-Dale AL, & Winqvist R (2002) ATM mutations in Finnish breast cancer patients. *J Med Genet* **39**: 192-196

Antoniou A, Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL, Loman N, Olsson H, Johannsson O, Borg A, Pasini B, Radice P, Manoukian S, Eccles DM, Tang N, Olah E, Anton-Culver H, Warner E, Lubinski J, Gronwald J et al (2003) Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. *Am J Hum Genet* **72**: 1117-1130

Antoniou AC, Casadei S, Heikkinen T, Barrowdale D, Pylkas K, Roberts J, Lee A, Subramanian D, De Leeneer K, Fostira F, Tomiak E, Neuhausen SL, Teo ZL, Khan S, Aittomaki K, Moilanen JS, Turnbull C, Seal S, Mannermaa A, Kallioniemi A et al (2014) Breast-cancer risk in families with mutations in PALB2. *N Engl J Med* **371**: 497-506

Antoniou AC, Pharoah PD, McMullan G, Day NE, Stratton MR, Peto J, Ponder BJ, & Easton DF (2002) A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. *Br J Cancer* **86**: 76-83

Antoniou AC, Spurdle AB, Sinilnikova OM, Healey S, Pooley KA, Schmutzler RK, Versmold B, Engel C, Meindl A, Arnold N, Hofmann W, Sutter C, Niederacher D, Deissler H, Caldes T, Kampjarvi K, Nevanlinna H, Simard J, Beesley J, Chen X et al (2008) Common breast cancer-predisposition alleles are associated with breast cancer risk in BRCA1 and BRCA2 mutation carriers. *Am J Hum Genet* **82**: 937-948

Babraham Bioinformatics FastQC http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/, accessed on January 2015.

Barnes NL, Ooi JL, Yarnold JR, & Bundred NJ (2012) Ductal carcinoma in situ of the breast. *BMJ* **344**: e797

Bartek J, Bartkova J, & Lukas J (2007) DNA damage signalling guards against activated oncogenes and tumour progression. *Oncogene* **26**: 7773-7779

Bartek J, Falck J, & Lukas J (2001) CHK2 kinase--a busy messenger. *Nat Rev Mol Cell Biol* **2**: 877-886

Bartek J & Lukas J (2003) Chk1 and Chk2 kinases in checkpoint control and cancer. *Cancer Cell* **3**: 421-429

Behrens J (1999) Cadherins and catenins: role in signal transduction and tumor progression. *Cancer Metastasis Rev* **18**: 15-30

Bell DW, Varley JM, Szydlo TE, Kang DH, Wahrer DC, Shannon KE, Lubratovich M, Verselis SJ, Isselbacher KJ, Fraumeni JF, Birch JM, Li FP, Garber JE, & Haber DA (1999) Heterozygous germ line hCHK2 mutations in Li-Fraumeni syndrome. *Science* **286**: 2528-2531

Berx G, Becker KF, Hofler H, & van Roy F (1998) Mutations of the human E-cadherin (CDH1) gene. *Hum Mutat* **12**: 226-237

Berx G, Staes K, van Hengel J, Molemans F, Bussemakers MJ, van Bokhoven A, & van Roy F (1995) Cloning and characterization of the human invasion suppressor gene E-cadherin (CDH1). *Genomics* **26**: 281-289

Bieging KT, Mello SS, & Attardi LD (2014) Unravelling mechanisms of p53-mediated tumour suppression. *Nat Rev Cancer* **14**: 359-370

Birch JM, Hartley AL, Tricker KJ, Prosser J, Condie A, Kelsey AM, Harris M, Jones PH, Binchy A, & Crowther D (1994) Prevalence and diversity of constitutional mutations in the p53 gene among 21 Li-Fraumeni families. *Cancer Res* **54**: 1298-1304

Blanco A, Grana B, Fachal L, Santamarina M, Cameselle-Teijeiro J, Ruiz-Ponte C, Carracedo A, & Vega A (2010) Beyond BRCA1 and BRCA2 wild-type breast and/or ovarian cancer families: germline mutations in TP53 and PTEN. *Clin Genet* **77**: 193-196

Blanco A, Gutierrez-Enriquez S, Santamarina M, Montalban G, Bonache S, Balmana J, Carracedo A, Diez O, & Vega A (2014) RAD51C germline mutations found in Spanish site-specific breast cancer and breast-ovarian cancer families. *Breast Cancer Res Treat* **147**: 133-143

Bogdanova N, Enssen-Dubrowinskaja N, Feshchenko S, Lazjuk GI, Rogov YI, Dammann O, Bremer M, Karstens JH, Sohn C, & Dork T (2005) Association of two mutations in the CHEK2 gene with breast cancer. *Int J Cancer* **116**: 263-266

Bray F, McCarron P, & Parkin DM (2004) The changing global patterns of female breast cancer incidence and mortality. *Breast Cancer Res* **6**: 229-239

Breast Cancer Linkage Consortium (1999) Cancer risks in BRCA2 mutation carriers. *J Natl Cancer Inst* **91**: 1310-1316

Broeks A, Urbanus JH, Floore AN, Dahler EC, Klijn JG, Rutgers EJ, Devilee P, Russell NS, van Leeuwen FE, & van 't Veer LJ (2000) ATM-heterozygous germline mutations contribute to breast cancer-susceptibility. *Am J Hum Genet* **66**: 494-500

Cancer Genome Atlas Network (2012) Comprehensive molecular portraits of human breast tumours. *Nature* **490**: 61-70

Cannioto RA & Moysich KB (2015) Epithelial ovarian cancer and recreational physical activity: A review of the epidemiological literature and implications for exercise prescription. *Gynecol Oncol* **137**: 559-573

Cantor S, Drapkin R, Zhang F, Lin Y, Han J, Pamidi S, & Livingston DM (2004) The BRCA1-associated protein BACH1 is a DNA helicase targeted by clinically relevant inactivating mutations. *Proc Natl Acad Sci U S A* **101**: 2357-2362

Cantor SB, Bell DW, Ganesan S, Kass EM, Drapkin R, Grossman S, Wahrer DC, Sgroi DC, Lane WS, Haber DA, & Livingston DM (2001) BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. *Cell* **105**: 149-160

Cartegni L, Wang J, Zhu Z, Zhang MQ, & Krainer AR (2003) ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res* **31**: 3568-3571

Casadei S, Norquist BM, Walsh T, Stray S, Mandell JB, Lee MK, Stamatoyannopoulos JA, & King MC (2011) Contribution of inherited mutations in the BRCA2-interacting protein PALB2 to familial breast cancer. *Cancer Res* **71**: 2222-2229

Castera L, Krieger S, Rousselin A, Legros A, Baumann JJ, Bruet O, Brault B, Fouillet R, Goardon N, Letac O, Baert-Desurmont S, Tinat J, Bera O, Dugast C, Berthet P, Polycarpe F, Layet V, Hardouin A, Frebourg T, & Vaur D (2014) Next-generation sequencing for the diagnosis of hereditary breast and ovarian cancer using genomic capture targeting multiple candidate genes. *Eur J Hum Genet* **22**: 1305-1313

Castillo A, Paul A, Sun B, Huang TH, Wang Y, Yazinski SA, Tyler J, Li L, You MJ, Zou L, Yao J, & Wang B (2014) The BRCA1-interacting protein Abraxas is required for genomic stability and tumor suppression. *Cell Rep* **8**: 807-817

Catucci I, Milgrom R, Kushnir A, Laitman Y, Paluch-Shimon S, Volorio S, Ficarazzi F, Bernard L, Radice P, Friedman E, & Peterlongo P (2012) Germline mutations in BRIP1 and PALB2 in Jewish high cancer risk families. *Fam Cancer* **11**: 483-491

CHEK2 Breast Cancer Case-Control Consortium (2004) CHEK2*1100delC and susceptibility to breast cancer: a collaborative analysis involving 10,860 breast cancer cases and 9,065 controls from 10 studies. *Am J Hum Genet* **74**: 1175-1182

Chun J, Buechelmaier ES, & Powell SN (2013) Rad51 paralog complexes BCDX2 and CX3 act at different stages in the BRCA1-BRCA2-dependent homologous recombination pathway. *Mol Cell Biol* **33**: 387-395

Ciccia A & Elledge SJ (2010) The DNA damage response: making it safe to play with knives. *Mol Cell* **40**: 179-204

Cimprich KA & Cortez D (2008) ATR: an essential regulator of genome integrity. *Nat Rev Mol Cell Biol* **9**: 616-627

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

Colella S, Yau C, Taylor JM, Mirza G, Butler H, Clouston P, Bassett AS, Seller A, Holmes CC, & Ragoussis J (2007) QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res* **35**: 2013-2025

Coleman MP, Quaresma M, Berrino F, Lutz JM, De Angelis R, Capocaccia R, Baili P, Rachet B, Gatta G, Hakulinen T, Micheli A, Sant M, Weir HK, Elwood JM, Tsukuma H, Koifman S, E Silva GA, Francisci S, Santaquilani M, Verdecchia A et al (2008) Cancer survival in five continents: a worldwide population-based study (CONCORD). *Lancet Oncol* **9**: 730-756

Collaborative Group on Epidemiological Studies of Ovarian Cancer, Beral V, Gaitskell K, Hermon C, Moser K, Reeves G, & Peto R (2012) Ovarian cancer and smoking: individual participant meta-analysis including 28,114 women with ovarian cancer from 51 epidemiological studies. *Lancet Oncol* **13**: 946-956

Collaborative Group on Hormonal Factors in Breast Cancer (2002) Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50302 women with breast cancer and 96973 women without the disease. *Lancet* **360**: 187-195

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

Collaborative Oncological Gene-environment Study http://www.cogseu.org/, accessed on June 2015.

Couch FJ, Nathanson KL, & Offit K (2014) Two decades after BRCA: setting paradigms in personalized cancer care and prevention. *Science* **343**: 1466-1470

Cox A, Dunning AM, Garcia-Closas M, Balasubramanian S, Reed MW, Pooley KA, Scollen S, Baynes C, Ponder BA, Chanock S, Lissowska J, Brinton L, Peplonska B, Southey MC, Hopper JL, McCredie MR, Giles GG, Fletcher O, Johnson N, dos Santos Silva I et al (2007) A common coding variant in CASP8 is associated with breast cancer risk. *Nat Genet* **39**: 352-358

Croce CM (2008) Oncogenes and cancer. *N Engl J Med* **358**: 502-511

Cruchaga C, Karch CM, Jin SC, Benitez BA, Cai Y, Guerreiro R, Harari O, Norton J, Budde J, Bertelsen S, Jeng AT, Cooper B, Skorupa T, Carrell D, Levitch D, Hsu S, Choi J, Ryten M, UK Brain Expression Consortium, Hardy J et al (2014) Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. *Nature* **505**: 550-554

Cybulski C, Gorski B, Huzarski T, Masojc B, Mierzejewski M, Debniak T, Teodorczyk U, Byrski T, Gronwald J, Matyjasik J, Zlowocka E, Lenner M, Grabowska E, Nej K, Castaneda J, Medrek K, Szymanska A, Szymanska J, Kurzawski G, Suchy J et al (2004) CHEK2 is a multiorgan cancer susceptibility gene. *Am J Hum Genet* **75**: 1131-1135

Damiola F, Pertesi M, Oliver J, Le Calvez-Kelm F, Voegele C, Young EL, Robinot N, Forey N, Durand G, Vallee MP, Tao K, Roane TC, Williams GJ, Hopper JL, Southey MC, Andrulis IL, John EM, Goldgar DE, Lesueur F, & Tavtigian SV (2014) Rare key functional domain missense substitutions in MRE11A, RAD50, and NBN contribute to breast cancer susceptibility: results from a Breast Cancer Family Registry case-control mutation-screening study. *Breast Cancer Res* **16**: R58

Dang CV (2010) Enigmatic MYC Conducts an Unfolding Systems Biology Symphony. *Genes Cancer* **1**: 526-531

Dansonka-Mieszkowska A, Kluska A, Moes J, Dabrowska M, Nowakowska D, Niwinska A, Derlatka P, Cendrowski K, & Kupryjanczyk J (2010) A novel germline PALB2 deletion in Polish breast and ovarian cancer patients. *BMC Med Genet* **11**: 20-2350-11-20

Database of Genomic Variants http://dgv.tcag.ca/, accessed on September 2011.

Deans AJ & West SC (2011) DNA interstrand crosslink repair and cancer. *Nat Rev Cancer* **11**: 467-480

Deffenbaugh AM, Frank TS, Hoffman M, Cannon-Albright L, & Neuhausen SL (2002) Characterization of common BRCA1 and BRCA2 variants. *Genet Test* **6**: 119-121

Degrolard-Courcet E, Sokolowska J, Padeano MM, Guiu S, Bronner M, Chery C, Coron F, Lepage C, Chapusot C, Loustalot C, Jouve JL, Hatem C, Ferrant E, Martin L, Coutant C, Baurand A, Couillault G, Delignette A, El Chehadeh S, Lizard S et al (2014) Development of primary early-onset colorectal cancers due to biallelic mutations of the FANCD1/BRCA2 gene. *Eur J Hum Genet* **22**: 979-987

Digweed M & Sperling K (2004) Nijmegen breakage syndrome: clinical manifestation of defective response to DNA double-strand breaks. *DNA Repair (Amst)* **3**: 1207-1217

Do R, Stitziel NO, Won HH, Jorgensen AB, Duga S, Angelica Merlini P, Kiezun A, Farrall M, Goel A, Zuk O, Guella I, Asselta R, Lange LA, Peloso GM, Auer PL, NHLBI Exome Sequencing Project, Girelli D, Martinelli N, Farlow DN, DePristo MA et al (2015) Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature* **518**: 102-106

Dolganov GM, Maser RS, Novikov A, Tosto L, Chong S, Bressan DA, & Petrini JH (1996) Human Rad50 is physically associated with human Mre11: identification of a conserved multiprotein complex implicated in recombinational DNA repair. *Mol Cell Biol* **16**: 4832-4841

Domchek SM, Tang J, Stopfer J, Lilli DR, Hamel N, Tischkowitz M, Monteiro AN, Messick TE, Powers J, Yonker A, Couch FJ, Goldgar DE, Davidson HR, Nathanson KL, Foulkes WD, & Greenberg RA (2013) Biallelic deleterious BRCA1 mutations in a woman with early-onset ovarian cancer. *Cancer Discov* **3**: 399-405

Dosanjh MK, Collins DW, Fan W, Lennon GG, Albala JS, Shen Z, & Schild D (1998) Isolation and characterization of RAD51C, a new human member of the RAD51 family of related genes. *Nucleic Acids Res* **26**: 1179-1184

Drooger JC, Hooning MJ, Seynaeve CM, Baaijens MH, Obdeijn IM, Sleijfer S, & Jager A (2015) Diagnostic and therapeutic ionizing radiation and the risk of a first and second primary breast cancer, with special attention for BRCA1 and BRCA2 mutation carriers: a critical review of the literature. *Cancer Treat Rev* **41**: 187-196

Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R, Wareham N, Ahmed S, Healey CS, Bowman R, SEARCH collaborators, Meyer KB, Haiman CA, Kolonel LK, Henderson BE, Le Marchand L et al (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**: 1087-1093

Ebi H, Matsuo K, Sugito N, Suzuki M, Osada H, Tajima K, Ueda R, & Takahashi T (2007) Novel NBS1 heterozygous germ line mutation causing MRE11-binding domain loss predisposes to common types of cancer. *Cancer Res* **67**: 11158-11165

Eeles RA (1995) Germline mutations in the TP53 gene. *Cancer Surv* **25**: 101-124

Eletr ZM & Wilkinson KD (2011) An emerging model for BAP1's role in regulating cell cycle progression. *Cell Biochem Biophys* **60**: 3-11

Encyclopedia of DNA elements at UCSC https://genome.ucsc.edu/ENCODE/, accessed on September 2011.

Erkko H, Dowty JG, Nikkila J, Syrjakoski K, Mannermaa A, Pylkas K, Southey MC, Holli K, Kallioniemi A, Jukkola-Vuorinen A, Kataja V, Kosma VM, Xia B, Livingston DM, Winqvist R, & Hopper JL (2008) Penetrance analysis of the PALB2 c.1592delT founder mutation. *Clin Cancer Res* **14**: 4667-4671

Erkko H, Xia B, Nikkila J, Schleutker J, Syrjakoski K, Mannermaa A, Kallioniemi A, Pylkas K, Karppinen S, Rapakko K, Miron A, Sheng Q, Li G, Mattila H, Bell DW, Haber DA, Grip M, Reiman M, Jukkola-Vuorinen A, Mustonen A et al (2007) A recurrent mutation in PALB2 in Finnish cancer families. *Nature* **446**: 316-319

Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, & Bray F (2015) Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* **136**: E359-86

Feuk L, Carson AR, & Scherer SW (2006) Structural variation in the human genome. *Nat Rev Genet* **7**: 85-97

Finnish Cancer Registry www.cancerregistry.fi, accessed on June 2015.

Fitzgerald RC, Hardwick R, Huntsman D, Carneiro F, Guilford P, Blair V, Chung DC, Norton J, Ragunath K, Van Krieken JH, Dwerryhouse S, Caldas C, & International Gastric Cancer Linkage Consortium (2010) Hereditary diffuse gastric cancer: updated consensus guidelines for clinical management and directions for future research. *J Med Genet* **47**: 436-444

Garcia-Closas M, Couch FJ, Lindstrom S, Michailidou K, Schmidt MK, Brook MN, Orr N, Rhie SK, Riboli E, Feigelson HS, Le Marchand L, Buring JE, Eccles D, Miron P, Fasching PA, Brauch H, Chang-Claude J, Carpenter J, Godwin AK, Nevanlinna H et al (2013) Genome-wide association studies identify four ER negative-specific breast cancer risk loci. *Nat Genet* **45**: 392-8, 398e1-2

Gatti RA, Berkel I, Boder E, Braedt G, Charmley P, Concannon P, Ersoy F, Foroud T, Jaspers NG, & Lange K (1988) Localization of an ataxia-telangiectasia gene to chromosome 11q22-23. *Nature* **336**: 577-580

Genetic Association Database http://geneticassociationdb.nih.gov/, accessed on September 2011.

Giardiello FM & Trimbath JD (2006) Peutz-Jeghers syndrome and management recommendations. *Clin Gastroenterol Hepatol* **4**: 408-415

Giardiello FM, Welsh SB, Hamilton SR, Offerhaus GJ, Gittelsohn AM, Booker SV, Krush AJ, Yardley JH, & Luk GD (1987) Increased risk of cancer in the Peutz-Jeghers syndrome. *N Engl J Med* **316**: 1511-1514

GitHub annalam/pypette https://github.com/annalam/pypette, accessed on January 2015.

Glover JN (2006) Insights into the molecular basis of human hereditary breast cancer from studies of the BRCA1 BRCT domain. *Fam Cancer* **5**: 89-93

Goff B (2012) Symptoms associated with ovarian cancer. *Clin Obstet Gynecol* **55**: 36-42

Guan M, Liu L, Zhao X, Wu Q, Yu B, Shao Y, Yang H, Fu X, Wan J, & Zhang W (2011) Copy number variations of EphA3 are associated with multiple types of hematologic malignancies. *Clin Lymphoma Myeloma Leuk* **11**: 50-53

Gudmundsdottir K & Ashworth A (2006) The roles of BRCA1 and BRCA2 and associated proteins in the maintenance of genomic stability. *Oncogene* **25**: 5864-5874

Guenard F, Labrie Y, Ouellette G, Beauparlant CJ, Bessette P, Chiquette J, Laframboise R, Lepine J, Lesperance B, Pichette R, Plante M, Durocher F, & INHERIT BRCAs (2007) Germline mutations in the breast cancer susceptibility gene PTEN are rare in high-risk non-BRCA1/2 French Canadian breast cancer families. *Fam Cancer* **6**: 483-490

Guenard F, Pedneault CS, Ouellette G, Labrie Y, Simard J, INHERIT, & Durocher F (2010) Evaluation of the contribution of the three breast cancer susceptibility genes CHEK2, STK11, and PALB2 in non-BRCA1/2 French Canadian families with high risk of breast cancer. *Genet Test Mol Biomarkers* **14**: 515-526

Haanpaa M, Pylkas K, Moilanen JS, & Winqvist R (2013) Evaluation of the need for routine clinical testing of PALB2 c.1592delT mutation in BRCA negative Northern Finnish breast cancer families. *BMC Med Genet* **14**: 82-2350-14-82

Hakem R, de la Pompa JL, Sirard C, Mo R, Woo M, Hakem A, Wakeham A, Potter J, Reitmair A, Billia F, Firpo E, Hui CC, Roberts J, Rossant J, & Mak TW (1996) The tumor suppressor gene Brca1 is required for embryonic cellular proliferation in the mouse. *Cell* **85**: 1009-1023

Hanahan D & Weinberg RA (2000) The hallmarks of cancer. *Cell* **100**: 57-70

Hanf V & Gonder U (2005) Nutrition and primary prevention of breast cancer: foods, nutrients and breast cancer risk. *Eur J Obstet Gynecol Reprod Biol* **123**: 139-149

Harper JW & Elledge SJ (2007) The DNA damage response: ten years after. *Mol Cell* **28**: 739-745

Hartikainen JM, Kataja V, Pirskanen M, Arffman A, Ristonmaa U, Vahteristo P, Ryynanen M, Heinonen S, Kosma VM, & Mannermaa A (2007) Screening for BRCA1 and BRCA2 mutations in Eastern Finnish breast/ovarian cancer families. *Clin Genet* **72**: 311-320

Hashizume R, Fukuda M, Maeda I, Nishikawa H, Oyake D, Yabuki Y, Ogata H, & Ohta T (2001) The RING heterodimer BRCA1-BARD1 is a ubiquitin ligase inactivated by a breast cancer-derived mutation. *J Biol Chem* **276**: 14537-14540

Heikkinen K, Rapakko K, Karppinen SM, Erkko H, Knuutila S, Lundan T, Mannermaa A, Borresen-Dale AL, Borg A, Barkardottir RB, Petrini J, & Winqvist R (2006) RAD50 and NBS1 are breast cancer susceptibility genes associated with genomic instability. *Carcinogenesis* **27**: 1593-1599

Heikkinen K, Rapakko K, Karppinen SM, Erkko H, Nieminen P, & Winqvist R (2005) Association of common ATM polymorphism with bilateral breast cancer. *Int J Cancer* **116**: 69-72

Heikkinen T, Greco D, Pelttari LM, Tommiska J, Vahteristo P, Heikkila P, Blomqvist C, Aittomaki K, & Nevanlinna H (2011) Variants on the promoter region of PTEN affect breast cancer progression and patient survival. *Breast Cancer Res* **13**: R130

Heikkinen T, Karkkainen H, Aaltonen K, Milne RL, Heikkila P, Aittomaki K, Blomqvist C, & Nevanlinna H (2009) The breast cancer susceptibility mutation PALB2 1592delT is associated with an aggressive tumor phenotype. *Clin Cancer Res* **15**: 3214-3222

Heintz AP, Odicino F, Maisonneuve P, Quinn MA, Benedet JL, Creasman WT, Ngan HY, Pecorelli S, & Beller U (2006) Carcinoma of the ovary. FIGO 26th Annual Report on the Results of Treatment in Gynecological Cancer. *Int J Gynaecol Obstet* **95 Suppl 1**: S161-92

Hemminki A, Markie D, Tomlinson I, Avizienyte E, Roth S, Loukola A, Bignell G, Warren W, Aminoff M, Hoglund P, Jarvinen H, Kristo P, Pelin K, Ridanpaa M, Salovaara R, Toro T, Bodmer W, Olschwang S, Olsen AS, Stratton MR et al (1998) A serine/threonine kinase gene defective in Peutz-Jeghers syndrome. *Nature* **391**: 184-187

Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, Rasmussen KE, Jones LP, Assefnia S, Chandrasekharan S, Backlund MG, Yin Y, Khramtsov AI, Bastein R, Quackenbush J, Glazer RI, Brown PH, Green JE, Kopelovich L, Furth PA et al (2007) Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol* **8**: R76

Hibino S, Kanda M, Oya H, Takami H, Shimizu D, Nomoto S, Hishida M, Niwa Y, Koike M, Yamada S, Nishikawa Y, Asai M, Nakayama G, Fujii T, Sugimoto H, Fujiwara M, & Kodera Y (2014) Reduced expression of DENND2D through promoter hypermethylation is an adverse prognostic factor in squamous cell carcinoma of the esophagus. *Oncol Rep* **31**: 693-700

Hirschhorn JN & Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* **6**: 95-108

Hoffman B & Liebermann DA (2008) Apoptotic signaling by c-MYC. *Oncogene* **27**: 6462-6472

Hopkins BD, Hodakoski C, Barrows D, Mense SM, & Parsons RE (2014) PTEN function: the long and the short of it. *Trends Biochem Sci* **39**: 183-190

Houghton SC, Reeves KW, Hankinson SE, Crawford L, Lane D, Wactawski-Wende J, Thomson CA, Ockene JK, & Sturgeon SR (2014) Perineal powder use and risk of ovarian cancer. *J Natl Cancer Inst* **106**: 10.1093/jnci/dju208. Print 2014 Sep

Howlett NG, Taniguchi T, Olson S, Cox B, Waisfisz Q, De Die-Smulders C, Persky N, Grompe M, Joenje H, Pals G, Ikeda H, Fox EA, & D'Andrea AD (2002) Biallelic inactivation of BRCA2 in Fanconi anemia. *Science* **297**: 606-609

Huang J, Huen MS, Kim H, Leung CC, Glover JN, Yu X, & Chen J (2009) RAD18 transmits DNA damage signalling to elicit homologous recombination repair. *Nat Cell Biol* **11**: 592-603

Huncharek M, Geschwind JF, & Kupelnick B (2003) Perineal application of cosmetic talc and risk of invasive epithelial ovarian cancer: a meta-analysis of 11,933 subjects from sixteen observational studies. *Anticancer Res* **23**: 1955-1960

Hunn J & Rodriguez GC (2012) Ovarian cancer: etiology, risk factors, and epidemiology. *Clin Obstet Gynecol* **55**: 3-23

Huusko P, Castren K, Launonen V, Soini Y, Paakkonen K, Leisti J, Vahakangas K, & Winqvist R (1999) Germ-line TP53 mutations in Finnish cancer families exhibiting features of the Li-Fraumeni syndrome and negative for BRCA1 and BRCA2. *Cancer Genet Cytogenet* **112**: 9-14

Huusko P, Paakkonen K, Launonen V, Poyhonen M, Blanco G, Kauppila A, Puistola U, Kiviniemi H, Kujala M, Leisti J, & Winqvist R (1998) Evidence of founder mutations in Finnish BRCA1 and BRCA2 families. *Am J Hum Genet* **62**: 1544-1548

Ikonen T, Matikainen M, Mononen N, Hyytinen ER, Helin HJ, Tommola S, Tammela TL, Pukkala E, Schleutker J, Kallioniemi OP, & Koivisto PA (2001) Association of E-cadherin germ-line alterations with prostate cancer. *Clin Cancer Res* **7**: 3465-3471

Jackson SP & Bartek J (2009) The DNA-damage response in human biology and disease. *Nature* **461**: 1071-1078

Janatova M, Kleibl Z, Stribrna J, Panczak A, Vesela K, Zimovjanova M, Kleiblova P, Dundr P, Soukupova J, & Pohlreich P (2013) The PALB2 gene is a strong candidate for clinical testing in BRCA1- and BRCA2-negative hereditary breast cancer. *Cancer Epidemiol Biomarkers Prev* **22**: 2323-2332

Jazaeri AA (2009) Molecular profiles of hereditary epithelial ovarian cancers and their implications for the biology of this disease. *Mol Oncol* **3**: 151-156

Jemal A, Center MM, DeSantis C, & Ward EM (2010) Global patterns of cancer incidence and mortality rates and trends. *Cancer Epidemiol Biomarkers Prev* **19**: 1893-1907

Jenne DE, Reimann H, Nezu J, Friedel W, Loff S, Jeschke R, Muller O, Back W, & Zimmer M (1998) Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase. *Nat Genet* **18**: 38-43

Jensen DE, Proctor M, Marquis ST, Gardner HP, Ha SI, Chodosh LA, Ishov AM, Tommerup N, Vissing H, Sekido Y, Minna J, Borodovsky A, Schultz DC, Wilkinson KD, Maul GG, Barlev N, Berger SL, Prendergast GC, & Rauscher FJ,3rd (1998) BAP1: a novel ubiquitin hydrolase which binds to the BRCA1 RING finger and enhances BRCA1-mediated cell growth suppression. *Oncogene* **16**: 1097-1112

Jones S, Hruban RH, Kamiyama M, Borges M, Zhang X, Parsons DW, Lin JC, Palmisano E, Brune K, Jaffee EM, Iacobuzio-Donahue CA, Maitra A, Parmigiani G, Kern SE, Velculescu VE, Kinzler KW, Vogelstein B, Eshleman JR, Goggins M, & Klein AP (2009) Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. *Science* **324**: 217

Kamal M, Shaaban AM, Zhang L, Walker C, Gray S, Thakker N, Toomes C, Speirs V, & Bell SM (2010) Loss of CSMD1 expression is associated with high tumour grade and poor survival in invasive ductal breast carcinoma. *Breast Cancer Res Treat* **121**: 555-563

Kar SP, Tyrer JP, Li Q, Lawrenson K, Aben KK, Anton-Culver H, Antonenkova N, Chenevix-Trench G, Australian Cancer Study, Australian Ovarian Cancer Study Group, Baker H, Bandera EV, Bean YT, Beckmann MW, Berchuck A, Bisogna M, Bjorge L, Bogdanova N, Brinton L, Brooks-Wilson A et al (2015) Network-Based Integration of GWAS and Gene Expression Identifies a HOX-Centric Network Associated with Serous Ovarian Cancer Risk. *Cancer Epidemiol Biomarkers Prev* **24**: 1574-1584

Karppinen SM, Vuosku J, Heikkinen K, Allinen M, & Winqvist R (2003) No evidence of involvement of germline BACH1 mutations in Finnish breast and ovarian cancer families. *Eur J Cancer* **39**: 366-371

Khandekar MJ, Cohen P, & Spiegelman BM (2011) Molecular mechanisms of cancer development in obesity. *Nat Rev Cancer* **11**: 886-895

Kiiski JI, Pelttari LM, Khan S, Freysteinsdottir ES, Reynisdottir I, Hart SN, Shimelis H, Vilske S, Kallioniemi A, Schleutker J, Leminen A, Butzow R, Blomqvist C, Barkardottir RB, Couch FJ, Aittomaki K, & Nevanlinna H (2014) Exome sequencing identifies FANCM as a susceptibility gene for triple-negative breast cancer. *Proc Natl Acad Sci U S A* **111**: 15172-15177

Kilpivaara O, Alhopuro P, Vahteristo P, Aaltonen LA, & Nevanlinna H (2006) CHEK2 I157T associates with familial and sporadic colorectal cancer. *J Med Genet* **43**: e34

Kilpivaara O, Laiho P, Aaltonen LA, & Nevanlinna H (2003) CHEK2 1100delC and colorectal cancer. *J Med Genet* **40**: e110

Kilpivaara O, Vahteristo P, Falck J, Syrjakoski K, Eerola H, Easton D, Bartkova J, Lukas J, Heikkila P, Aittomaki K, Holli K, Blomqvist C, Kallioniemi OP, Bartek J, & Nevanlinna H (2004) CHEK2 variant I157T may be associated with increased breast cancer risk. *Int J Cancer* **111**: 543-547

Kim H, Huang J, & Chen J (2007) CCDC98 is a BRCA1-BRCT domain-binding protein involved in the DNA damage response. *Nat Struct Mol Biol* **14**: 710-715

Kinzler KW & Vogelstein B (1998) Landscaping the cancer terrain. *Science* **280**: 1036-1037

Kinzler KW & Vogelstein B (1997) Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature* **386**: 761, 763

Kleinjan DA & van Heyningen V (2005) Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am J Hum Genet* **76**: 8-32

Knudson AG,Jr (1971) Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* **68**: 820-823

Koressaar T & Remm M (2007) Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23**: 1289-1291

Korsse SE, Peppelenbosch MP, & van Veelen W (2013) Targeting LKB1 signaling in cancer. *Biochim Biophys Acta* **1835**: 194-210

Krepischi AC, Achatz MI, Santos EM, Costa SS, Lisboa BC, Brentani H, Santos TM, Goncalves A, Nobrega AF, Pearson PL, Vianna-Morgante AM, Carraro DM, Brentani RR, & Rosenberg C (2012a) Germline DNA copy number variation in familial and early-onset breast cancer. *Breast Cancer Res* **14**: R24

Krepischi AC, Pearson PL, & Rosenberg C (2012b) Germline copy number variations and cancer predisposition. *Future Oncol* **8**: 441-450

Kurman RJ & Shih I (2010) The origin and pathogenesis of epithelial ovarian cancer: a proposed unifying theory. *Am J Surg Pathol* **34**: 433-443

La Vecchia C (2001) Epidemiology of ovarian cancer: a summary review. *Eur J Cancer Prev* **10**: 125-129

Laitinen VH, Akinrinade O, Rantapero T, Tammela TL, Wahlfors T, & Schleutker J (2015) Germline copy number variation analysis in Finnish families with hereditary prostate cancer. *Prostate*

Lakhani SR, Van De Vijver MJ, Jacquemier J, Anderson TJ, Osin PP, McGuffog L, & Easton DF (2002) The pathology of familial breast cancer: predictive value of immunohistochemical markers estrogen receptor, progesterone receptor, HER-2, and p53 in patients with mutations in BRCA1 and BRCA2. *J Clin Oncol* **20**: 2310-2318

Lalloo F, Varley J, Moran A, Ellis D, O'dair L, Pharoah P, Antoniou A, Hartley R, Shenton A, Seal S, Bulman B, Howell A, & Evans DG (2006) BRCA1, BRCA2 and TP53 mutations in very early-onset breast cancer with associated risks to relatives. *Eur J Cancer* **42**: 1143-1150

Lamarche BJ, Orazio NI, & Weitzman MD (2010) The MRN complex in double-strand break repair and telomere maintenance. *FEBS Lett* **584**: 3682-3695

Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357-359

Lee DS, Yoon SY, Looi LM, Kang P, Kang IN, Sivanandan K, Ariffin H, Thong MK, Chin KF, Mohd Taib NA, Yip CH, & Teo SH (2012) Comparable frequency of BRCA1, BRCA2 and TP53 germline mutations in a multi-ethnic Asian cohort suggests TP53 screening should be offered together with BRCA1/2 screening to early-onset breast cancer patients. *Breast Cancer Res* **14**: R66

Leongamornlert D, Mahmud N, Tymrakiewicz M, Saunders E, Dadaev T, Castro E, Goh C, Govindasami K, Guy M, O'Brien L, Sawyer E, Hall A, Wilkinson R, Easton D, UKGPCS Collaborators, Goldgar D, Eeles R, & Kote-Jarai Z (2012) Germline BRCA1 mutations increase prostate cancer risk. *Br J Cancer* **106**: 1697-1701

Levin ER (1995) Endothelins. *N Engl J Med* **333**: 356-363

Levitus M, Waisfisz Q, Godthelp BC, de Vries Y, Hussain S, Wiegant WW, Elghalbzouri-Maghrani E, Steltenpool J, Rooimans MA, Pals G, Arwert F, Mathew CG, Zdzienicka MZ, Hiom K, De Winter JP, & Joenje H (2005) The DNA helicase BRIP1 is defective in Fanconi anemia complementation group J. *Nat Genet* **37**: 934-935

Levy-Lahad E, Catane R, Eisenberg S, Kaufman B, Hornreich G, Lishinsky E, Shohat M, Weber BL, Beller U, Lahad A, & Halle D (1997) Founder BRCA1 and BRCA2 mutations in Ashkenazi Jews in Israel: frequency and differential penetrance in ovarian cancer and in breast-ovarian cancer families. *Am J Hum Genet* **60**: 1059-1067

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, & 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079

Li J, Ye L, Mansel RE, & Jiang WG (2011) Potential prognostic value of repulsive guidance molecules in breast cancer. *Anticancer Res* **31**: 1703-1711

Li J, Yen C, Liaw D, Podsypanina K, Bose S, Wang SI, Puc J, Miliaresis C, Rodgers L, McCombie R, Bigner SH, Giovanella BC, Ittmann M, Tycko B, Hibshoosh H, Wigler MH, & Parsons R (1997) PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science* **275**: 1943-1947

Lin HW, Tu YY, Lin SY, Su WJ, Lin WL, Lin WZ, Wu SC, & Lai YL (2011) Risk of ovarian cancer in women with pelvic inflammatory disease: a population-based study. *Lancet Oncol* **12**: 900-904

Liu C, Wang Y, Wang QS, & Wang YJ (2012) The CHEK2 I157T variant and breast cancer susceptibility: a systematic review and meta-analysis. *Asian Pac J Cancer Prev* **13**: 1355-1360

Liu Y, Colditz GA, Rosner B, Berkey CS, Collins LC, Schnitt SJ, Connolly JL, Chen WY, Willett WC, & Tamimi RM (2013) Alcohol intake between menarche and first pregnancy: a prospective study of breast cancer risk. *J Natl Cancer Inst* **105**: 1571-1578

Liu Z, Wu J, & Yu X (2007) CCDC98 targets BRCA1 to DNA damage sites. *Nat Struct Mol Biol* **14**: 716-720

Livraghi L & Garber JE (2015) PARP inhibitors in the management of breast cancer: current data and future prospects. *BMC Med* **13**: 188-015-0425-1

Lord CJ & Ashworth A (2012) The DNA damage response and cancer therapy. *Nature* **481**: 287-294

Loveday C, Turnbull C, Ruark E, Xicola RM, Ramsay E, Hughes D, Warren-Perry M, Snape K, Breast Cancer Susceptibility Collaboration (UK), Eccles D, Evans DG, Gore M, Renwick A, Seal S, Antoniou AC, & Rahman N (2012) Germline RAD51C mutations confer susceptibility to ovarian cancer. *Nat Genet* **44**: 475-6; author reply 476

Lovejoy CA, Xu X, Bansbach CE, Glick GG, Zhao R, Ye F, Sirbu BM, Titus LC, Shyr Y, & Cortez D (2009) Functional genomic screens identify CINP as a genome maintenance protein. *Proc Natl Acad Sci U S A* **106**: 19304-19309

Lowe KA, Chia VM, Taylor A, O'Malley C, Kelsh M, Mohamed M, Mowat FS, & Goff B (2013) An international assessment of ovarian cancer incidence and mortality. *Gynecol Oncol* **130**: 107-114

Ludwig T, Chapman DL, Papaioannou VE, & Efstratiadis A (1997) Targeted mutations of breast cancer susceptibility gene homologs in mice: lethal phenotypes of Brca1, Brca2, Brca1/Brca2, Brca1/p53, and Brca2/p53 nullizygous embryos. *Genes Dev* **11**: 1226-1241

Lynch BM, Neilson HK, & Friedenreich CM (2011) Physical activity and breast cancer prevention. *Recent Results Cancer Res* **186**: 13-42

Lynch ED, Ostermeyer EA, Lee MK, Arena JF, Ji H, Dann J, Swisshelm K, Suchard D, MacLeod PM, Kvinnsland S, Gjertsen BT, Heimdal K, Lubs H, Moller P, & King MC (1997) Inherited mutations in PTEN that are associated with breast cancer, cowden disease, and juvenile polyposis. *Am J Hum Genet* **61**: 1254-1260

Lynch HT, Casey MJ, Snyder CL, Bewtra C, Lynch JF, Butts M, & Godwin AK (2009) Hereditary ovarian carcinoma: heterogeneity, molecular genetics, pathology, and management. *Mol Oncol* **3**: 97-137

Ma C, Quesnelle KM, Sparano A, Rao S, Park MS, Cohen MA, Wang Y, Samanta M, Kumar MS, Aziz MU, Naylor TL, Weber BL, Fakharzadeh SS, Weinstein GS, Vachani A, Feldman MD, & Brose MS (2009) Characterization CSMD1 in a large set of primary lung, head and neck, breast and skin cancer tissues. *Cancer Biol Ther* **8**: 907-916

Ma XD, Cai GQ, Zou W, Huang YH, Zhang JR, Wang DT, & Chen BL (2013) BRIP1 variations analysis reveals their relative importance as genetic susceptibility factor for cervical cancer. *Biochem Biophys Res Commun* **433**: 232-236

Malhotra GK, Zhao X, Band H, & Band V (2010) Histological, molecular and functional subtypes of breast cancers. *Cancer Biol Ther* **10**: 955-960

Malkin D, Li FP, Strong LC, Fraumeni JF,Jr, Nelson CE, Kim DH, Kassel J, Gryka MA, Bischoff FZ, & Tainsky MA (1990) Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science* **250**: 1233-1238

Martin AM & Weber BL (2000) Genetic and hormonal risk factors in breast cancer. *J Natl Cancer Inst* **92**: 1126-1135

Masciari S, Larsson N, Senz J, Boyd N, Kaurah P, Kandel MJ, Harris LN, Pinheiro HC, Troussard A, Miron P, Tung N, Oliveira C, Collins L, Schnitt S, Garber JE, & Huntsman D (2007) Germline E-cadherin mutations in familial lobular breast cancer. *J Med Genet* **44**: 726-731

Mavaddat N, Antoniou AC, Easton DF, & Garcia-Closas M (2010) Genetic susceptibility to breast cancer. *Mol Oncol* **4**: 174-191

Mavaddat N, Peock S, Frost D, Ellis S, Platte R, Fineberg E, Evans DG, Izatt L, Eeles RA, Adlard J, Davidson R, Eccles D, Cole T, Cook J, Brewer C, Tischkowitz M, Douglas F, Hodgson S, Walker L, Porteous ME et al (2013) Cancer risks for BRCA1 and

BRCA2 mutation carriers: results from prospective analysis of EMBRACE. *J Natl Cancer Inst* **105**: 812-822

McBride KA, Ballinger ML, Killick E, Kirk J, Tattersall MH, Eeles RA, Thomas DM, & Mitchell G (2014) Li-Fraumeni syndrome: cancer risk assessment and clinical management. *Nat Rev Clin Oncol* **11**: 260-271

McBride OW, Merry D, & Givol D (1986) The gene for human p53 cellular tumor antigen is located on chromosome 17 short arm (17p13). *Proc Natl Acad Sci U S A* **83**: 130-134

McPherson K, Steel CM, & Dixon JM (2000) ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. *BMJ* **321**: 624-628

McVeigh TP, Choi JK, Miller NM, Green AJ, & Kerin MJ (2014) Lobular breast cancer in a CDH1 splice site mutation carrier: case report and review of the literature. *Clin Breast Cancer* **14**: e47-51

Meijers-Heijboer H, van den Ouweland A, Klijn J, Wasielewski M, de Snoo A, Oldenburg R, Hollestelle A, Houben M, Crepin E, van Veghel-Plandsoen M, Elstrodt F, van Duijn C, Bartels C, Meijers C, Schutte M, McGuffog L, Thompson D, Easton D, Sodha N, Seal S et al (2002) Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nat Genet* **31**: 55-59

Meindl A, Hellebrand H, Wiek C, Erven V, Wappenschmidt B, Niederacher D, Freund M, Lichtner P, Hartmann L, Schaal H, Ramser J, Honisch E, Kubisch C, Wichmann HE, Kast K, Deissler H, Engel C, Muller-Myhsok B, Neveling K, Kiechle M et al (2010) Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. *Nat Genet* **42**: 410-414

Merritt MA, Cramer DW, Vitonis AF, Titus LJ, & Terry KL (2013) Dairy foods and nutrients in relation to risk of ovarian cancer and major histological subtypes. *Int J Cancer* **132**: 1114-1124

Meyer D, Zeileis A, & Hornik K (2012) vcd: Visualizing Categorical Data. **R package version 1.2-13**

Meyer KB, O'Reilly M, Michailidou K, Carlebur S, Edwards SL, French JD, Prathalingham R, Dennis J, Bolla MK, Wang Q, de Santiago I, Hopper JL, Tsimiklis H, Apicella C, Southey MC, Schmidt MK, Broeks A, Van 't Veer LJ, Hogervorst FB, Muir K et al (2013) Fine-scale mapping of the FGFR2 breast cancer risk locus: putative functional variants differentially bind FOXA1 and E2F1. *Am J Hum Genet* **93**: 1046-1060

Miao H, Verkooijen HM, Chia KS, Bouchardy C, Pukkala E, Laronningen S, Mellemkjaer L, Czene K, & Hartman M (2011) Incidence and outcome of male breast cancer: an international population-based study. *J Clin Oncol* **29**: 4381-4386

Michailidou K, Beesley J, Lindstrom S, Canisius S, Dennis J, Lush MJ, Maranian MJ, Bolla MK, Wang Q, Shah M, Perkins BJ, Czene K, Eriksson M, Darabi H, Brand JS, Bojesen SE, Nordestgaard BG, Flyger H, Nielsen SF, Rahman N et al (2015) Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat Genet* **47**: 373-380

Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, Schmidt MK, Chang-Claude J, Bojesen SE, Bolla MK, Wang Q, Dicks E, Lee A, Turnbull C, Rahman N, Breast and Ovarian Cancer Susceptibility Collaboration, Fletcher O, Peto J, Gibson L, Dos Santos Silva I et al (2013) Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* **45**: 353-61, 361e1-2

Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, & Ding W (1994) A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**: 66-71

miRBase http://www.mirbase.org/, accessed on August 2010.

Molina-Montes E, Perez-Nevot B, Pollan M, Sanchez-Cantalejo E, Espin J, & Sanchez M (2014) Cumulative risk of second primary contralateral breast cancer in BRCA1/BRCA2 mutation carriers with a first breast cancer: A systematic review and meta-analysis. *Breast* **23**: 721-742

Moquet-Torcy G, Tolza C, Piechaczyk M, & Jariel-Encontre I (2014) Transcriptional complexity and roles of Fra-1/AP-1 at the uPA/Plau locus in aggressive breast cancer. *Nucleic Acids Res* **42**: 11011-11024

Morris JR & Solomon E (2004) BRCA1 : BARD1 induces the formation of conjugated ubiquitin structures, dependent on K6 of ubiquitin, in cells during DNA replication and repair. *Hum Mol Genet* **13**: 807-817

Mouchawar J, Korch C, Byers T, Pitts TM, Li E, McCredie MR, Giles GG, Hopper JL, & Southey MC (2010) Population-based estimate of the contribution of TP53 mutations to subgroups of early-onset breast cancer: Australian Breast Cancer Family Study. *Cancer Res* **70**: 4795-4800

Nagore E, Montoro A, Garcia-Casado Z, Botella-Estrada R, Insa A, Lluch A, Lopez-Guerrero JA, & Guillen C (2009) Germline mutations in CDKN2A are infrequent in female patients with melanoma and breast cancer. *Melanoma Res* **19**: 211-214

National Cancer Institute, Surveillance, Epidemiology, and End Results Program www.seer.cancer.gov, accessed on June 2015.

National Human Genome Research Institute, Breast Cancer Information Core Database http://research.nhgri.nih.gov/bic/, accessed on June 2015.

NCBI BLAST http://blast.ncbi.nlm.nih.gov/Blast.cgi, accessed on August 2010.

NCBI OMIM Database http://www.ncbi.nlm.nih.gov/omim, accessed on September 2011.

NCBI Reference Sequence Database http://www.ncbi.nlm.nih.gov/refseq/, accessed on September 2011.

Nelen MR, Padberg GW, Peeters EA, Lin AY, van den Helm B, Frants RR, Coulon V, Goldstein AM, van Reen MM, Easton DF, Eeles RA, Hodgsen S, Mulvihill JJ, Murday VA, Tucker MA, Mariman EC, Starink TM, Ponder BA, Ropers HH, Kremer H et al (1996) Localization of the gene for Cowden disease to chromosome 10q22-23. *Nat Genet* **13**: 114-116

Nelen MR, van Staveren WC, Peeters EA, Hassel MB, Gorlin RJ, Hamm H, Lindboe CF, Fryns JP, Sijmons RH, Woods DG, Mariman EC, Padberg GW, & Kremer H (1997) Germline mutations in the PTEN/MMAC1 gene in patients with Cowden disease. *Hum Mol Genet* **6**: 1383-1387

Nevanlinna H & Bartek J (2006) The CHEK2 gene and inherited breast cancer susceptibility. *Oncogene* **25**: 5912-5919

Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura K, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ et al (2010) Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* **42**: 790-793

Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, & Shendure J (2009)

Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**: 272-276

NHLBI Exome Sequencing Project https://esp.gs.washington.edu/drupal/, accessed on January 2015.

Oldenburg RA, Kroeze-Jansema K, Kraan J, Morreau H, Klijn JG, Hoogerbrugge N, Ligtenberg MJ, van Asperen CJ, Vasen HF, Meijers C, Meijers-Heijboer H, de Bock TH, Cornelisse CJ, & Devilee P (2003) The CHEK2*1100delC variant acts as a breast cancer risk modifier in non-BRCA1/BRCA2 multiple-case families. *Cancer Res* **63**: 8153-8157

Oliveira AM, Ross JS, & Fletcher JA (2005) Tumor suppressor genes in breast cancer: the gatekeepers and the caretakers. *Am J Clin Pathol* **124 Suppl**: S16-28

Olsen CM, Green AC, Whiteman DC, Sadeghi S, Kolahdooz F, & Webb PM (2007) Obesity and the risk of epithelial ovarian cancer: a systematic review and meta-analysis. *Eur J Cancer* **43**: 690-709

Olsen JH, Hahnemann JM, Borresen-Dale AL, Tretli S, Kleinerman R, Sankila R, Hammarstrom L, Robsahm TE, Kaariainen H, Bregard A, Brondum-Nielsen K, Yuen J, & Tucker M (2005) Breast and other cancers in 1445 blood relatives of 75 Nordic patients with ataxia telangiectasia. *Br J Cancer* **93**: 260-265

Ostrow KL, McGuire V, Whittemore AS, & DiCioccio RA (2004) The effects of BRCA1 missense variants V1804D and M1628T on transcriptional activity. *Cancer Genet Cytogenet* **153**: 177-180

Pabalan N, Jarjanazi H, & Ozcelik H (2013) Association between BRIP1 (BACH1) polymorphisms and breast cancer risk: a meta-analysis. *Breast Cancer Res Treat* **137**: 553-558

Palmer JR, Wise LA, Hatch EE, Troisi R, Titus-Ernstoff L, Strohsnitter W, Kaufman R, Herbst AL, Noller KL, Hyer M, & Hoover RN (2006) Prenatal diethylstilbestrol exposure and risk of breast cancer. *Cancer Epidemiol Biomarkers Prev* **15**: 1509-1514

Pan H, He Z, Ling L, Ding Q, Chen L, Zha X, Zhou W, Liu X, & Wang S (2014) Reproductive factors and breast cancer risk among BRCA1 or BRCA2 mutation carriers: results from ten studies. *Cancer Epidemiol* **38**: 1-8

Paredes J, Figueiredo J, Albergaria A, Oliveira P, Carvalho J, Ribeiro AS, Caldeira J, Costa AM, Simoes-Correia J, Oliveira MJ, Pinheiro H, Pinho SS, Mateus R, Reis CA, Leite M, Fernandes MS, Schmitt F, Carneiro F, Figueiredo C, Oliveira C et al (2012) Epithelial E- and P-cadherins: role and clinical significance in cancer. *Biochim Biophys Acta* **1826**: 297-311

Park DJ, Tao K, Le Calvez-Kelm F, Nguyen-Dumont T, Robinot N, Hammet F, Odefrey F, Tsimiklis H, Teo ZL, Thingholm LB, Young EL, Voegele C, Lonie A, Pope BJ, Roane TC, Bell R, Hu H, Shankaracharya, Huff CD, Ellis J et al (2014a) Rare mutations in RINT1 predispose carriers to breast and Lynch syndrome-spectrum cancers. *Cancer Discov* **4**: 804-815

Park JY, Singh TR, Nassar N, Zhang F, Freund M, Hanenberg H, Meetei AR, & Andreassen PR (2014b) Breast cancer-associated missense mutants of the PALB2 WD40 domain, which directly binds RAD51C, RAD51 and BRCA2, disrupt DNA repair. *Oncogene* **33**: 4803-4812

Park JY, Zhang F, & Andreassen PR (2014c) PALB2: the hub of a network of tumor suppressors involved in DNA damage responses. *Biochim Biophys Acta* **1846**: 263-275

Pasquale EB (2010) Eph receptors and ephrins in cancer: bidirectional signalling and beyond. *Nat Rev* Cancer **10**: 165-180

Pasquale EB (2008) Eph-ephrin bidirectional signaling in physiology and disease. *Cell* **133**: 38-52

Payne SJ, Bowen RL, Jones JL, & Wells CA (2008) Predictive markers in breast cancer--the present. *Histopathology* **52**: 82-90

Pelttari LM, Heikkinen T, Thompson D, Kallioniemi A, Schleutker J, Holli K, Blomqvist C, Aittomaki K, Butzow R, & Nevanlinna H (2011) RAD51C is a susceptibility gene for ovarian cancer. *Hum Mol Genet* **20**: 3278-3288

Pelttari LM, Nurminen R, Gylfe A, Aaltonen LA, Schleutker J, & Nevanlinna H (2012) Screening of Finnish RAD51C founder mutations in prostate and colorectal cancer patients. *BMC Cancer* **12**: 552-2407-12-552

Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, & Botstein D (2000) Molecular portraits of human breast tumours. *Nature* **406**: 747-752

Peterlongo P, Catucci I, Colombo M, Caleca L, Mucaki E, Bogliolo M, Marin M, Damiola F, Bernard L, Pensotti V, Volorio S, Dall'Olio V, Meindl A, Bartram C, Sutter C, Surowy H, Sornin V, Dondon MG, Eon-Marchais S, Stoppa-Lyonnet D et al (2015) FANCM c.5791C>T nonsense mutation (rs144567652) induces exon skipping, affects DNA repair activity and is a familial breast cancer risk factor. *Hum Mol Genet* **24**: 5345-5355

Pharoah PD, Day NE, Duffy S, Easton DF, & Ponder BA (1997) Family history and the risk of breast cancer: a systematic review and meta-analysis. *Int J Cancer* **71**: 800-809

Phelan CM, Dapic V, Tice B, Favis R, Kwan E, Barany F, Manoukian S, Radice P, van der Luijt RB, van Nesselrooij BP, Chenevix-Trench G, kConFab, Caldes T, de la Hoya M, Lindquist S, Tavtigian SV, Goldgar D, Borg A, Narod SA, & Monteiro AN (2005) Classification of BRCA1 missense variants of unknown clinical significance. *J Med Genet* **42**: 138-146

Phillips KA (2000) Immunophenotypic and pathologic differences between BRCA1 and BRCA2 hereditary breast cancers. *J Clin Oncol* **18**: 107S-12S

Pinheiro H, Oliveira C, Seruca R, & Carneiro F (2014) Hereditary diffuse gastric cancer - Pathophysiology and clinical management. *Best Pract Res Clin Gastroenterol* **28**: 1055-1068

Prat J (2012) New insights into ovarian cancer pathology. *Ann Oncol* **23 Suppl 10**: x111-7

Prat J, Ribe A, & Gallardo A (2005) Hereditary ovarian cancer. *Hum Pathol* **36**: 861-870

Prokofyeva D, Bogdanova N, Bermisheva M, Zinnatullina G, Hillemanns P, Khusnutdinova E, & Dork T (2012) Rare occurrence of PALB2 mutations in ovarian cancer patients from the Volga-Ural region. *Clin Genet* **82**: 100-101

Prokopcova J, Kleibl Z, Banwell CM, & Pohlreich P (2007) The role of ATM in breast cancer development. *Breast Cancer Res Treat* **104**: 121-128

Prowse AH, Manek S, Varma R, Liu J, Godwin AK, Maher ER, Tomlinson IP, & Kennedy SH (2006) Molecular genetic evidence that endometriosis is a precursor of ovarian cancer. *Int J Cancer* **119**: 556-562

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, & Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**: 559-575

Pylkas K, Erkko H, Nikkila J, Solyom S, & Winqvist R (2008) Analysis of large deletions in BRCA1, BRCA2 and PALB2 genes in Finnish breast and ovarian cancer families. *BMC Cancer* **8**: 146-2407-8-146

Pylkas K, Tommiska J, Syrjakoski K, Kere J, Gatei M, Waddell N, Allinen M, Karppinen SM, Rapakko K, Kaariainen H, Aittomaki K, Blomqvist C, Mustonen A, Holli K, Khanna KK, Kallioniemi OP, Nevanlinna H, & Winqvist R (2007) Evaluation of the role of Finnish ataxia-telangiectasia mutations in hereditary predisposition to breast cancer. *Carcinogenesis* **28**: 1040-1045

Pylkas K, Vuorela M, Otsukka M, Kallioniemi A, Jukkola-Vuorinen A, & Winqvist R (2012) Rare Copy Number Variants Observed in Hereditary Breast Cancer Cases Disrupt Genes in Estrogen Signaling and TP53 Tumor Suppression Network. *PLoS Genet* **8**: e1002734

Qin W, Hu J, Guo M, Xu J, Li J, Yao G, Zhou X, Jiang H, Zhang P, Shen L, Wan D, & Gu J (2003) BNIPL-2, a novel homologue of BNIP-2, interacts with Bcl-2 and Cdc42GAP in apoptosis. *Biochem Biophys Res Commun* **308**: 379-385

Rafnar T, Gudbjartsson DF, Sulem P, Jonasdottir A, Sigurdsson A, Jonasdottir A, Besenbacher S, Lundin P, Stacey SN, Gudmundsson J, Magnusson OT, le Roux L, Orlygsdottir G, Helgadottir HT, Johannsdottir H, Gylfason A, Tryggvadottir L, Jonasson JG, de Juan A, Ortega E et al (2011) Mutations in BRIP1 confer high risk of ovarian cancer. *Nat Genet* **43**: 1104-1107

Rahman N, Seal S, Thompson D, Kelly P, Renwick A, Elliott A, Reid S, Spanova K, Barfoot R, Chagtai T, Jayatilake H, McGuffog L, Hanks S, Evans DG, Eccles D, Breast Cancer Susceptibility Collaboration (UK), Easton DF, & Stratton MR (2007) PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat Genet* **39**: 165-167

Rapakko K, Allinen M, Syrjakoski K, Vahteristo P, Huusko P, Vahakangas K, Eerola H, Kainu T, Kallioniemi OP, Nevanlinna H, & Winqvist R (2001) Germline TP53 alterations in Finnish breast cancer families are rare and occur at conserved mutation-prone sites. *Br J Cancer* **84**: 116-119

Rashid MU, Muhammad N, Faisal S, Amin A, & Hamann U (2014) Deleterious RAD51C germline mutations rarely predispose to breast and ovarian cancer in Pakistan. *Breast Cancer Res Treat* **145**: 775-784

Rauh-Hain JA, Krivak TC, Del Carmen MG, & Olawaiye AB (2011) Ovarian cancer screening and early detection in the general population. *Rev Obstet Gynecol* **4**: 15-21

Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, Friez MJ, Funke BH, Hegde MR, Lyon E, & Working Group of the American College of Medical Genetics and Genomics Laboratory Quality Assurance Commitee (2013) ACMG clinical laboratory standards for next-generation sequencing. *Genet Med* **15**: 733-747

Reid LJ, Shakya R, Modi AP, Lokshin M, Cheng JT, Jasin M, Baer R, & Ludwig T (2008) E3 ligase activity of BRCA1 is not essential for mammalian cell viability or homology-directed repair of double-strand DNA breaks. *Proc Natl Acad Sci U S A* **105**: 20876-20881

Reid S, Schindler D, Hanenberg H, Barker K, Hanks S, Kalb R, Neveling K, Kelly P, Seal S, Freund M, Wurm M, Batish SD, Lach FP, Yetgin S, Neitzel H, Ariffin H, Tischkowitz M, Mathew CG, Auerbach AD, & Rahman N (2007) Biallelic mutations in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer. *Nat Genet* **39**: 162-164

Ren LP, Xian YS, Diao DM, Chen Y, Guo Q, & Dang CX (2013) Further evidence for the contribution of the BRCA1-interacting protein-terminal helicase 1 (BRIP1) gene in breast cancer susceptibility. *Genet Mol Res* **12**: 5793-5801

Renehan AG, Tyson M, Egger M, Heller RF, & Zwahlen M (2008) Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet* **371**: 569-578

Renwick A, Thompson D, Seal S, Kelly P, Chagtai T, Ahmed M, North B, Jayatilake H, Barfoot R, Spanova K, McGuffog L, Evans DG, Eccles D, Breast Cancer Susceptibility Collaboration (UK), Easton DF, Stratton MR, & Rahman N (2006) ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nat Genet* **38**: 873-875

Ripperger T, Gadzicki D, Meindl A, & Schlegelberger B (2009) Breast cancer susceptibility: current knowledge and implications for genetic counselling. *Eur J Hum Genet* **17**: 722-731

Risch HA, McLaughlin JR, Cole DE, Rosen B, Bradley L, Fan I, Tang J, Li S, Zhang S, Shaw PA, & Narod SA (2006) Population BRCA1 and BRCA2 mutation frequencies and cancer penetrances: a kin-cohort study in Ontario, Canada. *J Natl Cancer Inst* **98**: 1694-1706

Rosen DG, Yang G, Liu G, Mercado-Uribe I, Chang B, Xiao XS, Zheng J, Xue FX, & Liu J (2009) Ovarian cancer: pathology, biology, and disease models. *Front Biosci (Landmark Ed)* **14**: 2089-2102

Roy R, Chun J, & Powell SN (2011) BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nat Rev Cancer* **12**: 68-78

Sansam CL, Cruz NM, Danielian PS, Amsterdam A, Lau ML, Hopkins N, & Lees JA (2010) A vertebrate gene, ticrr, is an essential checkpoint and replication regulator. *Genes Dev* **24**: 183-194

Santos C, Sanz-Pamplona R, Nadal E, Grasselli J, Pernas S, Dienstmann R, Moreno V, Tabernero J, & Salazar R (2015) Intrinsic cancer subtypes--next steps into personalized medicine. *Cell Oncol (Dordr)* **38**: 3-16

Sarantaus L, Huusko P, Eerola H, Launonen V, Vehmanen P, Rapakko K, Gillanders E, Syrjakoski K, Kainu T, Vahteristo P, Krahe R, Paakkonen K, Hartikainen J, Blomqvist C, Lopponen T, Holli K, Ryynanen M, Butzow R, Borg A, Wasteson Arver B et al (2000) Multiple founder effects and geographical clustering of BRCA1 and BRCA2 families in Finland. *Eur J Hum Genet* **8**: 757-763

Sauter G, Lee J, Bartlett JM, Slamon DJ, & Press MF (2009) Guidelines for human epidermal growth factor receptor 2 testing: biologic and methodologic considerations. *J Clin Oncol* **27**: 1323-1333

Savitsky K, Bar-Shira A, Gilad S, Rotman G, Ziv Y, Vanagaite L, Tagle DA, Smith S, Uziel T, Sfez S, Ashkenazi M, Pecker I, Frydman M, Harnik R, Patanjali SR, Simmons A, Clines GA, Sartiel A, Gatti RA, Chessa L et al (1995) A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science* **268**: 1749-1753

Sawyer SL, Tian L, Kahkonen M, Schwartzentruber J, Kircher M, University of Washington Centre for Mendelian Genomics, FORGE Canada Consortium, Majewski J, Dyment DA, Innes AM, Boycott KM, Moreau LA, Moilanen JS, & Greenberg RA (2015) Biallelic mutations in BRCA1 cause a new Fanconi anemia subtype. *Cancer Discov* **5**: 135-142

Schneider M, Chandler K, Tischkowitz M, & Meyer S (2015) Fanconi anaemia: genetics, molecular biology, and cancer - implications for clinical management in children and adults. *Clin Genet* **88**: 13-24

Schrader KA, Masciari S, Boyd N, Salamanca C, Senz J, Saunders DN, Yorida E, Maines-Bandiera S, Kaurah P, Tung N, Robson ME, Ryan PD, Olopade OI, Domchek SM,

Ford J, Isaacs C, Brown P, Balmana J, Razzak AR, Miron P et al (2011) Germline mutations in CDH1 are infrequent in women with early-onset or familial lobular breast cancers. *J Med Genet* **48**: 64-68

Seal S, Thompson D, Renwick A, Elliott A, Kelly P, Barfoot R, Chagtai T, Jayatilake H, Ahmed M, Spanova K, North B, McGuffog L, Evans DG, Eccles D, Breast Cancer Susceptibility Collaboration (UK), Easton DF, Stratton MR, & Rahman N (2006) Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nat Genet* **38**: 1239-1241

Seppala EH, Ikonen T, Mononen N, Autio V, Rokman A, Matikainen MP, Tammela TL, & Schleutker J (2003) CHEK2 variants associate with hereditary prostate cancer. *Br J Cancer* **89**: 1966-1970

Sequencing Initiative Suomi http://www.sisuproject.fi/, accessed on January 2015.

Shiloh Y & Ziv Y (2013) The ATM protein kinase: regulating the cellular response to genotoxic stress, and more. *Nat Rev Mol Cell Biol* **14**: 197-210

Sigurdson AJ, Hauptmann M, Chatterjee N, Alexander BH, Doody MM, Rutter JL, & Struewing JP (2004) Kin-cohort estimates for familial breast cancer risk in relation to variants in DNA base excision repair, BRCA1 interacting and growth factor genes. *BMC Cancer* **4**: 9

Smith TM, Lee MK, Szabo CI, Jerome N, McEuen M, Taylor M, Hood L, & King MC (1996) Complete genomic sequence and analysis of 117 kb of human DNA containing the gene BRCA1. *Genome Res* **6**: 1029-1049

Solyom S, Aressy B, Pylkas K, Patterson-Fortin J, Hartikainen JM, Kallioniemi A, Kauppila S, Nikkila J, Kosma VM, Mannermaa A, Greenberg RA, & Winqvist R (2012) Breast cancer-associated Abraxas mutation disrupts nuclear localization and DNA damage response functions. *Sci Transl Med* **4**: 122ra23

Solyom S, Pylkas K, & Winqvist R (2010) Screening for large genomic rearrangements of the BRIP1 and CHK1 genes in Finnish breast cancer families. *Fam Cancer* **9**: 537-540

Somyajit K, Subramanya S, & Nagaraju G (2010) RAD51C: a novel cancer susceptibility gene is linked to Fanconi anemia and breast cancer. *Carcinogenesis* **31**: 2031-2038

Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Lonning PE, & Borresen-Dale AL (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **98**: 10869-10874

Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lonning PE, Brown PO, Borresen-Dale AL, & Botstein D (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* **100**: 8418-8423

Southey MC, Teo ZL, Dowty JG, Odefrey FA, Park DJ, Tischkowitz M, Sabbaghian N, Apicella C, Byrnes GB, Winship I, Baglietto L, Giles GG, Goldgar DE, Foulkes WD, Hopper JL, & kConFab for the Beast Cancer Family Registry (2010) A PALB2 mutation associated with high risk of breast cancer. *Breast Cancer Res* **12**: R109

Southey MC, Teo ZL, & Winship I (2013) PALB2 and breast cancer: ready for clinical translation! *Appl Clin Genet* **6**: 43-52

Spurdle AB, Whiley PJ, Thompson B, Feng B, Healey S, Brown MA, Pettigrew C, kConFab, Van Asperen CJ, Ausems MG, Kattentidt-Mouravieva AA, van den Ouweland AM, Dutch Belgium UV Consortium, Lindblom A, Pigg MH, Schmutzler RK, Engel C, Meindl A, German Consortium of Hereditary Breast and Ovarian Cancer, Caputo S

et al (2012) BRCA1 R1699Q variant displaying ambiguous functional abrogation confers intermediate breast and ovarian cancer risk. *J Med Genet* **49**: 525-532

Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, Masson G, Jakobsdottir M, Thorlacius S, Helgason A, Aben KK, Strobbe LJ, Albers-Akkers MT, Swinkels DW, Henderson BE, Kolonel LN, Le Marchand L, Millastre E, Andres R, Godino J et al (2007) Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* **39**: 865-869

Stewart GS, Maser RS, Stankovic T, Bressan DA, Kaplan MI, Jaspers NG, Raams A, Byrd PJ, Petrini JH, & Taylor AM (1999) The DNA double-strand break repair gene hMRE11 is mutated in individuals with an ataxia-telangiectasia-like disorder. *Cell* **99**: 577-587

Stracker TH & Petrini JH (2011) The MRE11 complex: starting from the ends. *Nat Rev Mol Cell Biol* **12**: 90-103

Stratton MR, Campbell PJ, & Futreal PA (2009) The cancer genome. *Nature* **458**: 719-724

Stratton MR & Rahman N (2008) The emerging landscape of breast cancer susceptibility. *Nat Genet* **40**: 17-22

Sulli G, Di Micco R, & d'Adda di Fagagna F (2012) Crosstalk between chromatin state and DNA damage response in cellular senescence and cancer. *Nat Rev Cancer* **12**: 709-720

Sundvall M, Iljin K, Kilpinen S, Sara H, Kallioniemi OP, & Elenius K (2008) Role of ErbB4 in breast cancer. *J Mammary Gland Biol Neoplasia* **13**: 259-268

Suntsova M, Garazha A, Ivanova A, Kaminsky D, Zhavoronkov A, & Buzdin A (2015) Molecular functions of human endogenous retroviruses in health and disease. *Cell Mol Life Sci* **72**: 3653-3675

Swift M, Morrell D, Cromartie E, Chamberlin AR, Skolnick MH, & Bishop DT (1986) The incidence and gene frequency of ataxia-telangiectasia in the United States. *Am J Hum Genet* **39**: 573-583

Swift M, Reitnauer PJ, Morrell D, & Chase CL (1987) Breast and other cancers in families with ataxia-telangiectasia. *N Engl J Med* **316**: 1289-1294

Sy SM, Huen MS, & Chen J (2009) PALB2 is an integral component of the BRCA complex required for homologous recombination repair. *Proc Natl Acad Sci U S A* **106**: 7155-7160

Syrjakoski K, Hyytinen ER, Kuukasjarvi T, Auvinen A, Kallioniemi OP, Kainu T, & Koivisto PA (2003) Androgen receptor gene alterations in Finnish male breast cancer. *Breast Cancer Res Treat* **77**: 167-170

Syrjakoski K, Kuukasjarvi T, Waltering K, Haraldsson K, Auvinen A, Borg A, Kainu T, Kallioniemi OP, & Koivisto PA (2004) BRCA2 mutations in 154 finnish male breast cancer patients. *Neoplasia* **6**: 541-545

Tai YC, Domchek S, Parmigiani G, & Chen S (2007) Breast cancer risk among male BRCA1 and BRCA2 mutation carriers. *J Natl Cancer Inst* **99**: 1811-1814

Tan MH, Mester JL, Ngeow J, Rybicki LA, Orloff MS, & Eng C (2012) Lifetime cancer risks in individuals with germline PTEN mutations. *Clin Cancer Res* **18**: 400-407

Tanaka H, Arakawa H, Yamaguchi T, Shiraishi K, Fukuda S, Matsui K, Takei Y, & Nakamura Y (2000) A ribonucleotide reductase gene involved in a p53-dependent cell-cycle checkpoint for DNA damage. *Nature* **404**: 42-49

Tavtigian SV, Simard J, Rommens J, Couch F, Shattuck-Eidens D, Neuhausen S, Merajver S, Thorlacius S, Offit K, Stoppa-Lyonnet D, Belanger C, Bell R, Berry S, Bogden R, Chen Q, Davis T, Dumont M, Frye C, Hattier T, Jammulapati S et al (1996) The

complete BRCA2 gene and mutations in chromosome 13q-linked kindreds. *Nat Genet* **12**: 333-337

Teo ZL, Park DJ, Provenzano E, Chatfield CA, Odefrey FA, Nguyen-Dumont T, kConFab, Dowty JG, Hopper JL, Winship I, Goldgar DE, & Southey MC (2013) Prevalence of PALB2 mutations in Australasian multiple-case breast cancer families. *Breast Cancer Res* **15**: R17

Terry PD & Rohan TE (2002) Cigarette smoking and the risk of breast cancer in women: a review of the literature. *Cancer Epidemiol Biomarkers Prev* **11**: 953-971

Thakur S, Zhang HB, Peng Y, Le H, Carroll B, Ward T, Yao J, Farid LM, Couch FJ, Wilson RB, & Weber BL (1997) Localization of BRCA1 and a splice variant identifies the nuclear localization signal. *Mol Cell Biol* **17**: 444-452

The Cancer Genome Atlas http://cancergenome.nih.gov/, accessed on June 2015.

The R Project for Statistical Computing https://www.r-project.org/, accessed on September 2011.

Thompson D, Duedal S, Kirner J, McGuffog L, Last J, Reiman A, Byrd P, Taylor M, & Easton DF (2005) Cancer risks and mortality in heterozygous ATM mutation carriers. *J Natl Cancer Inst* **97**: 813-822

Thompson ER, Boyle SE, Johnson J, Ryland GL, Sawyer S, Choong DY, kConFab, Chenevix-Trench G, Trainer AH, Lindeman GJ, Mitchell G, James PA, & Campbell IG (2012a) Analysis of RAD51C germline mutations in high-risk breast and ovarian cancer families and ovarian cancer patients. *Hum Mutat* **33**: 95-99

Thompson ER, Doyle MA, Ryland GL, Rowley SM, Choong DY, Tothill RW, Thorne H, kConFab, Barnes DR, Li J, Ellul J, Philip GK, Antill YC, James PA, Trainer AH, Mitchell G, & Campbell IG (2012b) Exome sequencing identifies rare deleterious mutations in DNA repair genes FANCC and BLM as potential breast cancer susceptibility alleles. *PLoS Genet* **8**: e1002894

Thorstenson YR, Roxas A, Kroiss R, Jenkins MA, Yu KM, Bachrich T, Muhr D, Wayne TL, Chu G, Davis RW, Wagner TM, & Oefner PJ (2003) Contributions of ATM mutations to familial breast and ovarian cancer. *Cancer Res* **63**: 3325-3333

Thusberg J & Vihinen M (2009) Pathogenic or not? And if so, then how? Studying the effects of missense mutations using bioinformatics methods. *Hum Mutat* **30**: 703-714

Tice JA, O'Meara ES, Weaver DL, Vachon C, Ballard-Barbash R, & Kerlikowske K (2013) Benign breast disease, mammographic breast density, and the risk of breast cancer. *J Natl Cancer Inst* **105**: 1043-1049

Tischkowitz M, Capanu M, Sabbaghian N, Li L, Liang X, Vallee MP, Tavtigian SV, Concannon P, Foulkes WD, Bernstein L, WECARE Study Collaborative Group, Bernstein JL, & Begg CB (2012) Rare germline mutations in PALB2 and breast cancer risk: a population-based study. *Hum Mutat* **33**: 674-680

Tischkowitz M, Sabbaghian N, Ray AM, Lange EM, Foulkes WD, & Cooney KA (2008) Analysis of the gene coding for the BRCA2-interacting protein PALB2 in hereditary prostate cancer. *Prostate* **68**: 675-678

To T, Wall C, Baines CJ, & Miller AB (2014) Is carcinoma in situ a precursor lesion of invasive breast cancer? *Int J Cancer* **135**: 1646-1652

Tommiska J, Jansen L, Kilpivaara O, Edvardsen H, Kristensen V, Tamminen A, Aittomaki K, Blomqvist C, Borresen-Dale AL, & Nevanlinna H (2006a) ATM variants and cancer risk in breast cancer patients from Southern Finland. *BMC Cancer* **6**: 209

Tommiska J, Seal S, Renwick A, Barfoot R, Baskcomb L, Jayatilake H, Bartkova J, Tallila J, Kaare M, Tamminen A, Heikkila P, Evans DG, Eccles D, Aittomaki K, Blomqvist C,

Bartek J, Stratton MR, Nevanlinna H, & Rahman N (2006b) Evaluation of RAD50 in familial breast cancer predisposition. *Int J Cancer* **118**: 2911-2916

Trujillano D, Weiss ME, Schneider J, Koster J, Papachristos EB, Saviouk V, Zakharkina T, Nahavandi N, Kovacevic L, & Rolfs A (2015) Next-generation sequencing of the BRCA1 and BRCA2 genes for the genetic diagnostics of hereditary breast and/or ovarian cancer. *J Mol Diagn* **17**: 162-170

Turnbull C & Rahman N (2008) Genetic predisposition to breast cancer: past, present, and future. *Annu Rev Genomics Hum Genet* **9**: 321-345

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, & Rozen SG (2012) Primer3--new capabilities and interfaces. *Nucleic Acids Res* **40**: e115

Uziel T, Savitsky K, Platzer M, Ziv Y, Helbitz T, Nehls M, Boehm T, Rosenthal A, Shiloh Y, & Rotman G (1996) Genomic Organization of the ATM gene. *Genomics* **33**: 317-320

Vahteristo P, Bartkova J, Eerola H, Syrjakoski K, Ojala S, Kilpivaara O, Tamminen A, Kononen J, Aittomaki K, Heikkila P, Holli K, Blomqvist C, Bartek J, Kallioniemi OP, & Nevanlinna H (2002) A CHEK2 genetic variant contributing to a substantial fraction of familial breast cancer. *Am J Hum Genet* **71**: 432-438

Vahteristo P, Yliannala K, Tamminen A, Eerola H, Blomqvist C, & Nevanlinna H (2006) BACH1 Ser919Pro variant and breast cancer risk. *BMC Cancer* **6**: 19

van Lier MG, Wagner A, Mathus-Vliegen EM, Kuipers EJ, Steyerberg EW, & van Leerdam ME (2010) High cancer risk in Peutz-Jeghers syndrome: a systematic review and surveillance recommendations. *Am J Gastroenterol* **105**: 1258-64; author reply 1265

van Roy F (2014) Beyond E-cadherin: roles of other cadherin superfamily members in cancer. *Nat Rev Cancer* **14**: 121-134

Varley JM (2003) Germline TP53 mutations and Li-Fraumeni syndrome. *Hum Mutat* **21**: 313-320

Varley JM, McGown G, Thorncroft M, Santibanez-Koref MF, Kelsey AM, Tricker KJ, Evans DG, & Birch JM (1997) Germ-line mutations of TP53 in Li-Fraumeni families: an extended study of 39 families. *Cancer Res* **57**: 3245-3252

Vaz F, Hanenberg H, Schuster B, Barker K, Wiek C, Erven V, Neveling K, Endt D, Kesterton I, Autore F, Fraternali F, Freund M, Hartmann L, Grimwade D, Roberts RG, Schaal H, Mohammed S, Rahman N, Schindler D, & Mathew CG (2010) Mutation of the RAD51C gene in a Fanconi anemia-like disorder. *Nat Genet* **42**: 406-409

Vehmanen P, Friedman LS, Eerola H, McClure M, Ward B, Sarantaus L, Kainu T, Syrjakoski K, Pyrhonen S, Kallioniemi OP, Muhonen T, Luce M, Frank TS, & Nevanlinna H (1997a) Low proportion of BRCA1 and BRCA2 mutations in Finnish breast cancer families: evidence for additional susceptibility genes. *Hum Mol Genet* **6**: 2309-2315

Vehmanen P, Friedman LS, Eerola H, Sarantaus L, Pyrhonen S, Ponder BA, Muhonen T, & Nevanlinna H (1997b) A low proportion of BRCA2 mutations in Finnish breast cancer families. *Am J Hum Genet* **60**: 1050-1058

Vietri MT, Caliendo G, Casamassimi A, Cioffi M, De Paola ML, Napoli C, & Molinari AM (2015) A novel PALB2 truncating mutation in an Italian family with male breast cancer. *Oncol Rep* **33**: 1243-1247

Vogelstein B & Kinzler KW (2004) Cancer genes and the pathways they control. *Nat Med* **10**: 789-799

Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA,Jr, & Kinzler KW (2013) Cancer genome landscapes. *Science* **339**: 1546-1558

Vousden KH & Lu X (2002) Live or let die: the cell's response to p53. *Nat Rev Cancer* **2**: 594-604

Vousden KH & Prives C (2009) Blinded by the Light: The Growing Complexity of p53. *Cell* **137**: 413-431

Vuorela M, Pylkas K, Hartikainen JM, Sundfeldt K, Lindblom A, von Wachenfeldt Wappling A, Haanpaa M, Puistola U, Rosengren A, Anttila M, Kosma VM, Mannermaa A, & Winqvist R (2011) Further evidence for the contribution of the RAD51C gene in hereditary breast and ovarian cancer susceptibility. *Breast Cancer Res Treat* **130**: 1003-1010

Waltes R, Kalb R, Gatei M, Kijas AW, Stumm M, Sobeck A, Wieland B, Varon R, Lerenthal Y, Lavin MF, Schindler D, & Dork T (2009) Human RAD50 deficiency in a Nijmegen breakage syndrome-like disorder. *Am J Hum Genet* **84**: 605-616

Wang B (2012) BRCA1 tumor suppressor network: focusing on its tail. *Cell Biosci* **2**: 6-3701-2-6

Wang B, Matsuoka S, Ballif BA, Zhang D, Smogorzewska A, Gygi SP, & Elledge SJ (2007a) Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response. *Science* **316**: 1194-1198

Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, Hakonarson H, & Bucan M (2007b) PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* **17**: 1665-1674

Wang K, Li M, & Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**: e164

Weigelt B, Geyer FC, & Reis-Filho JS (2010) Histological types of breast cancer: how special are they? *Mol Oncol* **4**: 192-208

Weigelt B & Reis-Filho JS (2009) Histological and molecular types of breast cancer: is there a unifying taxonomy? *Nat Rev Clin Oncol* **6**: 718-730

Weischer M, Bojesen SE, Ellervik C, Tybjaerg-Hansen A, & Nordestgaard BG (2008) CHEK2*1100delC genotyping for clinical assessment of breast cancer risk: meta-analyses of 26,000 patient cases and 27,000 controls. *J Clin Oncol* **26**: 542-548

Weren RD, Venkatachalam R, Cazier JB, Farin HF, Kets CM, de Voer RM, Vreede L, Verwiel ET, van Asseldonk M, Kamping EJ, Kiemeney LA, Neveling K, Aben KK, Carvajal-Carmona L, Nagtegaal ID, Schackert HK, Clevers H, van de Wetering M, Tomlinson IP, Ligtenberg MJ et al (2015) Germline deletions in the tumour suppressor gene FOCAD are associated with polyposis and colorectal cancer development. *J Pathol* **236**: 155-164

Wickenden JA & Watson CJ (2010) Key signalling nodes in mammary gland development and cancer. Signalling downstream of PI3 kinase in mammary epithelium: a play in 3 Akts. *Breast Cancer Res* **12**: 202

Wooster R, Neuhausen SL, Mangion J, Quirk Y, Ford D, Collins N, Nguyen K, Seal S, Tran T, & Averill D (1994) Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science* **265**: 2088-2090

Wu LC, Wang ZW, Tsan JT, Spillman MA, Phung A, Xu XL, Yang MC, Hwang LY, Bowcock AM, & Baer R (1996) Identification of a RING protein that can interact in vivo with the BRCA1 gene product. *Nat Genet* **14**: 430-440

Xi HQ, Wu XS, Wei B, & Chen L (2012) Aberrant expression of EphA3 in gastric carcinoma: correlation with tumor angiogenesis and survival. *J Gastroenterol* **47**: 785-794

Xi HQ & Zhao P (2011) Clinicopathological significance and prognostic value of EphA3 and CD133 expression in colorectal carcinoma. *J Clin Pathol* **64**: 498-503

Xia B, Sheng Q, Nakanishi K, Ohashi A, Wu J, Christ N, Liu X, Jasin M, Couch FJ, & Livingston DM (2006) Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. *Mol Cell* **22**: 719-729

Xie L, Qin W, Li J, He X, Zhang H, Yao G, Shu H, Yao M, Wan D, & Gu J (2007) BNIPL-2 promotes the invasion and metastasis of human hepatocellular carcinoma cells. *Oncol Rep* **17**: 605-610

Xie ZM, Li LS, Laquet C, Penault-Llorca F, Uhrhammer N, Xie XM, & Bignon YJ (2011) Germline mutations of the E-cadherin gene in families with inherited invasive lobular breast carcinoma but no diffuse gastric cancer. *Cancer* **117**: 3112-3117

Xu M, Bai L, Gong Y, Xie W, Hang H, & Jiang T (2009) Structure and functional implications of the human rad9-hus1-rad1 cell cycle checkpoint complex. *J Biol Chem* **284**: 20457-20461

Xue Y, Li Y, Guo R, Ling C, & Wang W (2008) FANCM of the Fanconi anemia core complex is required for both monoubiquitination and DNA repair. *Hum Mol Genet* **17**: 1641-1652

Yu X, Chini CC, He M, Mer G, & Chen J (2003) The BRCT domain is a phospho-protein binding domain. *Science* **302**: 639-642

Zannini L, Delia D, & Buscemi G (2014) CHK2 kinase in the DNA damage response and beyond. *J Mol Cell Biol* **6**: 442-457

Zhang B, Kirov S, & Snoddy J (2005) WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res* **33**: W741-8

Zhang F, Fan Q, Ren K, & Andreassen PR (2009) PALB2 functionally connects the breast cancer susceptibility proteins BRCA1 and BRCA2. *Mol Cancer Res* **7**: 1110-1118

Zhang ZH, Yang LS, Huang F, Hao JH, Su PY, & Sun YH (2012) Current evidence on the relationship between two polymorphisms in the NBS1 gene and breast cancer risk: a meta-analysis. *Asian Pac J Cancer Prev* **13**: 5375-5379

Zhou H, Jin J, Zhang H, Yi B, Wozniak M, & Wong L (2012) IntPath--an integrated pathway gene relationship database for model organisms and important pathogens. *BMC Syst Biol* **6 Suppl 2**: S2-0509-6-S2-S2. Epub 2012 Dec 12

Zhu Q, Pao GM, Huynh AM, Suh H, Tonnu N, Nederlof PM, Gage FH, & Verma IM (2011) BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. *Nature* **477**: 179-184

Zuhlke KA, Johnson AM, Okoth LA, Stoffel EM, Robbins CM, Tembe WA, Salinas CA, Zheng SL, Xu J, Carpten JD, Lange EM, Isaacs WB, & Cooney KA (2012) Identification of a novel NBN truncating mutation in a family with hereditary prostate cancer. *Fam Cancer* **11**: 595-600

Breast Cancer
R E S E A R C H

Screening for *BRCA1, BRCA2, CHEK2, PALB2, BRIP1, RAD50*, and *CDH1* mutations in high-risk Finnish *BRCA1/2*-founder mutation-negative breast and/or ovarian cancer individuals

Kirsi M Kuusisto[1,2], Aleksandra Bebel[1], Mauno Vihinen[1], Johanna Schleutker[1,2], Satu-Leena Sallinen[3*]

Abstract

Introduction: Two major high-penetrance breast cancer genes, *BRCA1* and *BRCA2*, are responsible for approximately 20% of hereditary breast cancer (HBC) cases in Finland. Additionally, rare mutations in several other genes that interact with *BRCA1* and *BRCA2* increase the risk of HBC. Still, a majority of HBC cases remain unexplained which is challenging for genetic counseling. We aimed to analyze additional mutations in HBC-associated genes and to define the sensitivity of our current *BRCA1/2* mutation analysis protocol used in genetic counseling.

Methods: Eighty-two well-characterized, high-risk hereditary breast and/or ovarian cancer (HBOC) *BRCA1/2*-founder mutation-negative Finnish individuals, were screened for germline alterations in seven breast cancer susceptibility genes, *BRCA1, BRCA2, CHEK2, PALB2, BRIP1, RAD50*, and *CDH1*. *BRCA1/2* were analyzed by multiplex ligation-dependent probe amplification (MLPA) and direct sequencing. *CHEK2* was analyzed by the high resolution melt (HRM) method and *PALB2, RAD50, BRIP1* and *CDH1* were analyzed by direct sequencing. Carrier frequencies between 82 (HBOC) *BRCA1/2*-founder mutation-negative Finnish individuals and 384 healthy Finnish population controls were compared by using Fisher's exact test. *In silico* prediction for novel missense variants effects was carried out by using Pathogenic-Or-Not -Pipeline (PON-P).

Results: Three previously reported breast cancer-associated variants, *BRCA1* c.5095C > T, *CHEK2* c.470T > C, and *CHEK2* c.1100delC, were observed in eleven (13.4%) individuals. Ten of these individuals (12.2%) had *CHEK2* variants, c.470T > C and/or c.1100delC. Fourteen novel sequence alterations and nine individuals with more than one non-synonymous variant were identified. One of the novel variants, *BRCA2* c.72A > T (Leu24Phe) was predicted to be likely pathogenic *in silico*. No large genomic rearrangements were detected in *BRCA1/2* by multiplex ligation-dependent probe amplification (MLPA).

Conclusions: In this study, mutations in previously known breast cancer susceptibility genes can explain 13.4% of the analyzed high-risk *BRCA1/2*-negative HBOC individuals. *CHEK2* mutations, c.470T > C and c.1100delC, make a considerable contribution (12.2%) to these high-risk individuals but further segregation analysis is needed to evaluate the clinical significance of these mutations before applying them in clinical use. Additionally, we identified novel variants that warrant additional studies. Our current genetic testing protocol for 28 Finnish *BRCA1/2*-founder mutations and protein truncation test (PTT) of the largest exons is sensitive enough for clinical use as a primary screening tool.

* Correspondence: Satu-Leena.Sallinen@pshp.fi
[3]Department of Pediatrics, Genetics Outpatient Clinic, Tampere University Hospital, Biokatu 8, Tampere, 33520, Finland
Full list of author information is available at the end of the article

Introduction

Breast cancer (BrCa) is the most common cancer among women in Finland, with about 4,000 cases diagnosed yearly (Finnish Cancer Registry). It has been estimated that a monogenic trait accounts for 5 to 10% of all BrCa cases [1]. The two major high-penetrance BrCa genes, *BRCA1* (*breast cancer 1*) and *BRCA2* (*breast cancer 2*), are responsible for 30% of hereditary breast cancer (HBC) cases worldwide, but only for about 20% in Finland [2-4]. *BRCA2* mutations have been found to be more common in the Finnish population than *BRCA1* [5]. In addition to *BRCA1* and *BRCA2* mutations, there are certain hereditary cancer syndromes, such as Li-Fraumeni, Cowden, Peutz-Jeghers and diffuse gastric cancer syndromes, associated with a high risk of BrCa [6-9]. However, these syndromes very seldom explain HBC.

BRCA1 and *BRCA2* have many DNA damage response functions in the cell [10]. Therefore, it has been hypothesized that genes coding for proteins that interact with *BRCA1* or *BRCA2* or act in the same DNA repair pathway would be likely candidate genes for HBC susceptibility. As expected, *CHEK2* (*checkpoint kinase 2*), *PALB2* (*partner and localizer of BRCA2*), *BRIP1* (*BRCA1-interacting protein 1*), and *RAD50* (*human homolog of Saccharomyces cerevisiae RAD50*) have been shown to have rare, moderate-risk BrCa-associated variants, which have also been studied in the Finnish population [11-14]. In addition, BrCa-associated variants have been reported in the *CDH1* (*cadherin-1*) [15].

Although mutations in many genes have been found to predispose an individual to BrCa, approximately 75 to 80% of HBC cases remain unexplained [16]. It is likely that additional BrCa susceptibility gene mutations remain unidentified, especially in the category of moderate- to low-penetrance gene variants that individually confer only minimal risk but that, through multiplicative and/or cumulative effects, can cause relatively high risk for the carriers [17]. Genome-wide association studies (GWAs) have revealed multiple low penetrance, single nucleotide polymorphisms (SNPs) in many genes and chromosomal loci with increased risk of BrCa. For example, SNPs in the *fibroblast growth factor receptor 2* (*FGFR2*) gene have shown significant association with increased risk among BrCa cases with strong family history [18].

To address the problem of heterogeneous HBC in genetic counseling, we wanted to investigate possible additional mutations in HBC-associated genes. The aim of this study was to screen seven known BrCa susceptibility genes for additional mutations in 82 well-characterized, Finnish, high-risk hereditary breast and/or ovarian cancer (HBOC) individuals tested to be *BRCA1/2*-founder mutation negative. In addition, the sensitivity of our current *BRCA1/2* mutation analysis protocol was defined for genetic counseling purposes.

Materials and methods
Patients and controls

Index individuals of 82 high-risk Finnish HBOC families were screened for germline alterations in BrCa-associated genes. All individuals had been detected to be founder mutation-negative by minisequencing of the previously known 28 Finnish *BRCA1/2* mutations and by protein truncation test (PTT) of exon 11 for *BRCA1* and exons 10 and 11 for *BRCA2*. Study material had been collected from the individuals, who visited the Tampere University Hospital Genetics Outpatient Clinic between January 1997 and May 2008. The hospital district, in the area of Pirkanmaa, consists of over 20% (1.23 million) of the Finnish population. Individuals were chosen to be included in this study according to the following criteria of high-risk HBC: (a) the individual or her first-degree relative (only female family members were included when defining first-degree relatives) had BrCa or ovarian cancer (OvCa) at younger than 30 years of age; or (b) two first-degree relatives in the family had BrCa and/or OvCa and at least one of the cancers had been diagnosed at younger than 40 years of age; or (c) three first-degree relatives in the family had BrCa and/or OvCa and at least one of the cancers had been diagnosed at younger than 50 years of age; or (d) four or more first-degree relatives had BrCa and/or OvCa at any age; or (e) the same individual had BrCa and OvCa. Patient with bilateral BrCa was considered to have two separate cancers. According to these criteria, our study material also included 11 non-affected females in addition to 71 BrCa and/or OvCa patients. We were also able to get blood samples from two affected relatives in 2 out of 11 separate families with healthy index. These relatives with BrCa were screened for the same variant as that identified in the index. The clinical data of the studied individuals are presented in Table 1. As controls, 384 blood samples from anonymous healthy females, collected from the Finnish Red Cross, were used. All individuals have been informed of the analyses, and they have given written consent to use their already existing DNA samples. Permission for the research project has been received from the Ethical Committee of Tampere University Hospital and the National Authority for Medicolegal Affairs.

Mutation detection

DNA samples of the individuals were kindly received from the Tampere University Hospital Genetics Outpatient Clinic. Mutation screening for *BRCA1*, *BRCA2*, *PALB2*, *BRIP1*, *RAD50*, and *CDH1* was performed by direct sequencing. Whole-coding regions and exon-

Table 1 Characteristics of the studied individuals

Number of index individuals (n = 82)	BrCa	Bil. BrCa	OvCa	BrCa and OvCa	Non-affected
	57	8	1	5	11
Age at diagnosis					
(BrCa/OvCa)	n = 57	n = 16[a]	n = 1	n = 10[b]	-
<30	13	0	0	1	-
<40 years	11	2	0	2	-
<50 years	15	7	0	2	-
≥50 years	18	7	1	5	-
Type of BrCa	n = 57	n = 16[a]	-	n = 5	-
Ductal	41	8	-	5	-
Intraductal	2	2	-	0	-
Lobular	9	3	-	0	-
Papillary	1	0	-	0	-
Medullary	1	0	-	0	-
Unknown	3	3	-	0	-
Ductal carcinomas grade known	n = 38	n = 6		n = 5	-
Grade 1	7	3		0	-
Grade 2	14	3		3	-
Grade 3	17	0		2	-
ER status known	n = 51	n = 11	-	n = 4	-
ER +	35	10	-	2	-
ER-	16	1	-	2	-
PR status known	n = 50	n = 11	-	n = 4	-
PR+	30	10	-	2	-
PR-	20	1	-	2	-
HER2, status known	n = 46	n = 11	-	n = 4	-
HER2+	14	1	-	1	-
HER2-	32	10	-	3	-
Type of OvCa	-	-	n = 1	n = 5	-
Serous	-	-	0	0	-
Endometrioid	-	-	0	0	-
Mucinous	-	-	0	2	-
Clear cell	-	-	0	1	-
Other	-	-	0	2	-
Unknown	-	-	1	0	-
Number of affected (BrCa/OvCa)					
first-degree relatives	n = 57	n = 8	n = 1	n = 5	n = 11
≥2	25	3	1	0	5
≥1	24	4	0	0	6
0	8	1	0	5	0
Number of affected (BrCa/OvCa)					
second-degree relatives	n = 57	n = 8	n = 1	n = 5	n = 11
≥2	5	0	0	0	4
≥1	11	1	0	0	0
0	41	7	1	5	7

Bil. BrCa, bilateral breast cancer; BrCa, breast cancer; ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; OvCa, ovarian cancer; PR, progesterone receptor. [a]8 Bilateral breast cancer cases, together 16 cancers, [b]5 Breast and ovarian cancer cases, together 10 cancers.

intron boundaries were analyzed. Primer sequences for *PALB2*, *BRIP1*, and *RAD50* have been reported previously [12,13,19]. Primers for *BRCA1* and *BRCA2* (excluding previously analyzed exon 11 for *BRCA1* and exons 10 and 11 for *BRCA2*) and *CDH1* were designed by using Primer3 software (Rozen and Skaletsky, Whitehead Institute for Biomedical Research, Cambridge, MA, USA) [20]. *CHEK2* was screened by using high-resolution melt (HRM) analysis on a Bio-Rad platform (Bio-Rad Laboratories Headquarters, Hercules, CA, USA). Sequencing was carried out

using the Big Dye Terminator v.3.1 Cycle Sequencing Kit and ABIPRISM 3130 × l Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Sequences were analyzed with Sequencher v.4.7 software (Gene Codes Corporation, Ann Arbor, MI, USA). Primer sequences, detailed HRM and PCR reaction conditions are available upon request.

Control frequencies were determined for 18 variants by HRM (*CHEK2* variants), direct sequencing (*BRCA1* c.4883T > C and *RAD50* c.1544A > G) and TaqMan® SNP genotyping assays (Applied Biosystems, Foster City, CA, USA) and with an ABI7900 instrument (Applied Biosystems, Foster City, CA, USA). Assays were already designed and functionally tested for the following SNPs: c.8182G > A (rs28897749), c.9976A > T (rs11571833), c.10234A > G (rs1801426), and c.1676A > G (rs152451). As for the c.72A > T, c.814G > A, c.1000T > G, and c.2993G > A (rs45551636) variants, assays were designed by Custom TaqMan® Assay Design Tool (Applied Biosystems, Foster City, CA, USA) according to manufacturer's instructions.

The multiplex ligation-dependent probe amplification (MLPA) analysis was performed for *BRCA1* and *BRCA2* (SALSA MLPA kit P002-B1 for *BRCA1* (lot 0508) and kit P090-A2 for *BRCA2* (lot 0808), MRC-Holland, Amsterdam, the Netherlands) according to manufacturer's instructions and analyzed with ABIPRISM 3130xl Genetic Analyzer and Genemapper® v.4.0 software (Applied Biosystems, Foster City, CA, USA).

Statistical analyses
Carrier frequencies between 82 studied individuals and 384 population controls were compared by using Fisher's exact test [21]. All *P*-values were two sided. Odds ratios (OR) were generated by two-by-two table.

In silico prediction of novel missense variants effects
The effects of five novel-coding missense variants, *BRCA2* c.72A > T (Leu24Phe), *CHEK2* c.1363G > A (Val455Ile), *PALB2* c.814G > A (Glu272Lys), *PALB2* c.1000T > G (Tyr334Asp), and *RAD50* c.1544A > G (Asp515Gly), were predicted with a number of tools by using Pathogenic-Or-Not-Pipeline (PON-P) [22]. The predictions included those for amino acid tolerance (programs PolyPhen version 2, Sift, PhD-SNP, SNAP) and protein stability (I-Mutant version 3). PON-P allows simultaneous submission of a number of variations and proteins to selected predictors. PON-P utilizes machine learning to combine results from several individual predictions.

MicroRNA database and BLAST search for novel variants
MicroRNA (miRNA) target site search was performed for the novel variant genomic positions from the microRNA database (miRBase) [23]. Also BLAST search [24] was performed for the novel human variant genomic positions to see if these sites are conserved among different organisms including mouse, rat, cow, and chicken.

Results
Index individuals of 82 high-risk HBOC families were screened for germline alterations in *BRCA1*, *BRCA2*, *CHEK2*, *PALB2*, *BRIP1*, *RAD50*, and *CDH1* genes. Detailed clinical information of analyzed individuals is shown in Table 1. All of the identified 54 sequence variants with their observed genotype frequencies and rs-numbers are presented in Supplementary Table S1 in Additional file 1. All of the identified non-synonymous and novel sequence alterations are summarized in Table 2. Table 2 variants are presented in Table 3 with index individual and family cancer history. In addition, as our study material also included healthy index individuals from 11 families, we made an effort to get blood samples from two affected relatives in 2 out of 11 separate families. These relatives with BrCa were screened for the same variant as that identified in the healthy index. Analysis was performed for the new cases in family 112 (*CHEK2* c.470T > C and *PALB2* c.1676A > G variants) and family 231 (*BRCA1* c.4883T > C variant; Table 3). In family 112, the case proved to have the same *PALB2* c.1676A > G variant as the index individual but in family 231, the affected relative did not carry the *BRCA1* c.4883T > C variant (data not shown). To further evaluate the impact of these 11 healthy index cases, we recalculated the frequencies without these 11 individuals for those variants accepted to be meaningful for BrCa risk. Supplementary Table S2 in Additional file 2 shows these re-calculated frequencies for *BRCA1* c.5095C > T, *CHEK2* c.470T > C, and *CHEK2* c.1100delC variants. No statistically significant effect was seen for exclusion of the 11 cases.

BRCA1 and *BRCA2* mutation analysis
Analysis of *BRCA1* and *BRCA2* revealed altogether 16 different sequence variants, seven in *BRCA1* and nine in *BRCA2* [see Supplementary Table S1 in Additional file 1]. All but two of the identified variants in *BRCA1*, c.4883T > C and c.5095C > T, have been reported to be neutral in the databases. Heterozygous c.4883T > C variant was observed in 4 of 82 (4.9%) women of which three had BrCa and one had a family history of breast, cervix and skin cancers (Tables 2 and 3). In population controls, the frequency of the c.4883T > C variant was 6 of 367 (1.6%). The c.5095C > T variant has been classified as a deleterious mutation in the Breast Cancer Information Core (BIC) database. The heterozygous c.5095C > T mutation was observed in 1 of 82 (1.2%) women. The mutation carrying woman had BrCa diagnosed at the age of 42 years and a strong family history of cancer (Tables 2 and 3,

Table 2 Identified non-synonymous and novel sequence alterations

Gene/Nucleotide change[a]	Effect on protein	rs Number[b]	Carrier frequency		P-values	OR; 95%CI	Status
			Individuals	Controls			
BRCA1							
4837A > G	Ser1613Gly	rs1799966	0.634 (52/82)[g]	na	-	-	Reported[c, d]
4883T > C	Met1628Thr	rs4986854	0.049 (4/82)	0.016 (6/367)	0.090	3.09; 0.85-11.19	Reported[c, d]
5095C > T	Arg1699Trp	rs55770810	0.012 (1/82)	na	-	-	Reported[c, d]
BRCA2							
68-80insT[f]	-	-	0.012 (1/82)	na	-	-	Novel
72A > T	Leu24Phe	-	0.012 (1/82)	0 (0/380)	0.177	na	Novel
793 + 34T > G	-	-	0.012 (1/82)	na	-	-	Novel
8182G > A	Val2728Ile	rs28897749	0.012 (1/82)	0.003 (1/378)	0.325	4.65; 0.29-75.19	Reported[c, d]
9976A > T	Lys3326Stop	rs11571833	0.012 (1/82)	0.029 (11/378)	0.702	0.41; 0.05-3.24	Reported[c, d]
10234A > G	Ile3412Val	rs1801426	0.012 (1/82)	0.021 (8/379)	1.000	0.57; 0.07-4.64	Reported[c, d]
CHEK2							
444 + 85T > A	-	-	0.012 (1/82)	0.005 (2/364)	0.457	2.23; 0.20-24.94	Novel
470T > C	Ile157Thr	-	0.098 (8/81)	0.055 (21/381)	0.203	1.88; 0.80-4.41	Reported[e]
792 + 39C > T	-	-	0.012 (1/82)	0.021 (8/375)	1.000	0.57; 0.07-4.60	Novel
1100delC[f]	Fs, stop at codon 381	-	0.037 (3/82)	0.016 (6/380)	0.203	2.37; 0.58-9.67	Reported[e]
1290T > C	His430His	-	0.951 (77/81)[g]	0.974 (372/382)	0.281	0.52; 0.16-1.69	Novel
1314T > C	Asp438Asp	-	0.951 (77/81)[g]	0.974 (372/382)	0.281	0.52; 0.16-1.69	Novel
1363G > A	Val455Ile	-	0.975 (79/81)[g]	0.976 (373/382)	1.000	0.95; 0.20-4.50	Novel
PALB2							
814G > A	Glu272Lys	-	0.012 (1/82)	0 (0/372)	0.181	na	Novel
1000T > G	Tyr334Asp	-	0.012 (1/82)	0.011 (4/380)	1.000	1.16; 0.13-10.52	Novel
1010T > C	Leu337Ser	rs45494092	0.073 (6/82)	na	-	-	Reported[c]
1676A > G	Gln559Arg	rs152451	0.122 (10/82)	0.173 (64/371)	0.323	0.67; 0.33-1.36	Reported[c]
2205A > G	Pro735Pro	-	0.012 (1/82)	na	-	-	Novel
2794G > A	Val932Met	rs45624036	0.037 (3/82)	na	-	-	Reported[c]
2993G > A	Gly998Glu	rs45551636	0.012 (1/82)	0.038 (14/372)	0.491	0.32; 0.04-2.44	Reported[c]
BRIP1							
584T > C	Leu195Pro	rs4988347	0.024 (2/82)	na	-	-	Reported[c]
2755C > T	Pro919Ser	rs4986764	0.390 (32/82)[g]	na	-	-	Reported[c]
RAD50							
1544A > G	Asp515Gly	-	0.012 (1/82)	0.010 (4/384)	1.000	1.17; 0.13-10.63	Novel
2398-32A > G	-	-	0.012 (1/82)	na	-	-	Novel
3475 + 33C > G	-	-	0.012 (1/82)	na	-	-	Novel

CI, confidence interval; Fs, frameshift; na, not analyzed; OR, odds ratio. [a]The reference nucleotide sequences were obtained from the UCSC Genome Browser [44] and the accession numbers were following: BRCA1: [UCSC Genome Browser:NM_007295.2], BRCA2: [UCSC Genome Browser:NM_000059.3], CHEK2: [UCSC Genome Browser:NM_007194.3], PALB2: [UCSC Genome Browser:NM_024675.3], BRIP1: [UCSC Genome Browser:NM_032043.1], RAD50: [UCSC Genome Browser: NM_005732.3], and CDH1: [UCSC Genome Browser:NM_004360.3]. The accession numbers for the protein sequences obtained from the Swiss-Prot Protein knowledgebase [45] were following: BRCA1: [Swiss-Prot:P38398], BRCA2: [Swiss-Prot:P51587], CHEK2: [Swiss-Prot:O96017], PALB2: [Swiss-Prot:Q86YC2], BRIP1: [Swiss-Prot:Q9BX63], RAD50: [Swiss-Prot:Q92878], and CDH1: [Swiss-Prot:P12830]. [b]The RefSNP number, obtained from the NCBI Single Nucleotide Polymorphism database (dbSNP) [46]. [c]The NCBI dbSNP [46]. [d]The Breast Cancer Information Core database [47]. [e]Reported in the Finnish population by Vahteristo et al. [11]. [f]Heterozygous deletion or insertion.[g]Due to the high frequency of the variant observed in analyzed individuals, variant is not presented in Table 3.

Figure 1, Family 249). Additional mutation analysis also revealed two other affected women carrying the c.5095C > T mutation in the same family. In *BRCA2*, three of the nine identified variants were novel, c.68-80insT, c.72A > T, and c.793 + 34T > G (Tables 2 and 3.). The heterozygous missense variant c.72A > T (Leu24Phe), was observed in 1 of 82 (1.2%) women but not in population controls. The c.72A > T variant carrying woman had BrCa diagnosed at the age of 53 years. She had also two affected first-degree

relatives (mother and sister). Protein predictions by PON-P suggested that substitution of leucine by phenylalanine in position 24 changes significantly the properties of the side chain and the substitution would not be tolerated. All the other identified variants in *BRCA2* have been reported previously and they are either neutral or the clinical significance of the variants is yet uncertain especially with the three missense variants, c.8182G > A, c.9976A > T and c.10234A > G (Tables 2 and 3). No deletions or duplication

Table 3 Identified variants in the studied individuals

Family id	Gene and variant	Type of cancer	BrCa/OvCa Histology,Grade	Receptor status	Other cancer cases in the family (Age at diagnosis if available)
202	*BRCA1*, 4883T > C	Br (26)	Ductal, 3	ER-, PR-, HER2-	Skin (54)
206	*BRCA1*, 4883T > C	Br (53)	Ductal, 1	ER-, PR-, HER2-	Bil. Ov (64), <u>Br (49)</u>
231	*BRCA1*, 4883T > C	-			<u>Br x2 (33, 46)</u>, Cer (60), Skin (73)
232	*BRCA1*, 4883T > C	Br (34)	Ductal, na	ER +, PR +, HER2 na	<u>Br (39)</u>
249 (Figure 1)	*BRCA1*, 5095C > T	Br (42)	Medullary, na	na	Br x5 (*35*, **44**, 57, 67, 71,), Co (78), Kid (67), Mel (63), *Ov (45)*, Skin, <u>To (51)</u>, Ute (39)
115	*BRCA2*, 68-80insT	Br (59)	Lobular, 2	ER+, PR+, HER2-	Br x3 (<50), <u>Br x2 (60, 60)</u>
240	*BRCA2*, 72A > T	Br (53)	Ductal, 3	ER+, PR-, HER2+	Br x2 (<u>42</u>, 62)
207	*BRCA2*, 793 + 34T > G	Br (38)	na	na	Bil. Br (64)
5 (Figure 8)	*BRCA2*, 8182G > A *BRCA2*, 10234A > G	-			Bil. Br x2 (43, 48 and 54, 76), Br (43), *Brain (75)*, Lip (45), *Lung (81)*, *Skin (75)*, Sto (56)
4	*BRCA2*, 9976A > T *CHEK2*, 470T > C	-			Bil. Br x2 (53, 69 and <70), Br (<70)
212	*CHEK2*, 444 + 85T > A *PALB2*, 2794G > A	Bil. Br (43)	Ductal, 2 and na	ER+, PR+, HER2- and na	<u>Br (52)</u>
110 (Figure 3)	*CHEK2*, 792 + 39C > T *CHEK2*, 470T > C *CHEK2*, 1100delC *RAD50*, 2398-32A > G	Br (26)	Ductal, 2	ER+, PR+, HER2+	*Br (48)*, Ca (84), *Pr (64)*
112	*CHEK2*, 470T > C *PALB2*, 1676A > G	-			Br x3 (<u>35</u>, <u>43</u>, 83), Skin (76), *Lung (71)*
120	*CHEK2*, 470T > C	-			Br (64), Ov (72)
122	*CHEK2*, 470T > C	Br (25)	Ductal, 2	ER-, PR-, HER2+	*Brain (66, Ca (83)*,Cer (31), Pr (93), Re (73), Skin (87)
126	*CHEK2*, 470T > C	Br (48)	Ductal, na	ER, PR, HER2 na	*Bil. Br*, Br x2 (<u>51</u>, <53)
129 (Figure 2)	*CHEK2*, 470T > C *PALB2*, 1676A > G	Skin (70), Bil. Br (78), Sto (82)	Lobular, 2 and Ductal, 1	ER-, PR-, HER2- and ER+, PR+, HER2-	<u>Bil. Br (59)</u>, **Co (58)**, **Skin (48)**
262 (Figure 6)	*CHEK2*, 470T > C *PALB2*, 1000T > G	Bil. Br (45, 58)	Intraductal, na and Ductal, na	ER+, PR+, HER2+ and ER+, PR+, HER2-	Br (57), Panc (83), *Si (79)*
264 (Figure 4)	*CHEK2*, 1100delC	Bil. Br (44)	Lobular, 2	ER+, PR+, HER2-	Br x2 (44, *52*)
265 (Figure 5)	*CHEK2*, 1100delC *PALB2*, 1676A > G	Br (45)	Ductal, 3	ER+, PR+, HER2+	Br (38)
237	*PALB2*, 814G > A	Br (28)	Ductal, 2	ER+, PR+, HER2+	-
133	*PALB2*, 1010T > C	Br (48)	Ductal, 2	ER+, PR+, HER2-	*Int, Br x2 (73, 79)*, Skin (60)
235	*PALB2*, 1010T > C	Br (52)	na	na	<u>Bil. Br (28)</u>, Br (56)
239	*PALB2*, 1010T > C	Br (37)	Ductal, 2	ER+, PR+, HER2-	*Br (> 90)*
250	*PALB2*, 1010T > C	Br (24)	Ductal, 3	ER+, PR+, HER2+	<u>Cer (30)</u>, Ov (83)
260	*PALB2*, 1010T > C	Br (29)	Ductal, 3	ER-, PR-, HER2-	Br (58), *Lung (60)*
267	*PALB2*, 1010T > C	Br (48)	Ductal, 1	ER+, PR+, HER2 na	<u>Br x2 (51, 58)</u>
113	*PALB2*, 1676A > G[a]	Br (51), Skin (55)	Ductal, 3	ER-, PR-, HER2+	Br (35)
131 (Figure 7)	*PALB2*, 1676A > G *BRIP1*, 584T > C	Bil. Br (54)	Intraductal, na and Ductal, 2	na	<u>Bil. Br (46), Br (48)</u>
229	*PALB2*, 1676A > G	Bil. Br (68)	na	na	<u>Bil. Br (50, 70), Br x2 (45, 50)</u>
236	*PALB2*, 1676A > G	Br (29)	Intraductal, na	na	Br (52)
246	*PALB2*, 1676A > G	Thy (30), Cer (33), Br (39)	Ductal, 3	ER-, PR-, HER2+	Br x2 (<u>49</u>, 54), *Re (61)*

Table 3 Identified variants in the studied individuals (*Continued*)

268	*PALB2*, 1676A > G	Br (62)	Papillary, na	ER+, PR+, HER2-	Br x2 (36, <u>38</u>)
	PALB2, 2205A > G				
271	*PALB2*, 1676A > G	Thy (62),	Lobular, 2	ER+, PR+, HER2-	**Br x2 (43, 44)**
		Br (65)			
102	*PALB2*, 2794G > A	Br (29)	Lobular, na	ER-, PR-, HER2+	*Br (72)*
	BRIP1, 584T > C				
244	*PALB2*, 2794G > A	Br (45)	Ductal, 2	ER+, PR+, HER2-	Bil. Br (<45), Br x2 (<35, <u>46</u>),
					Brain (67)
270	*PALB2*, 2993G > A	Br (66)	Ductal, 3	ER+, PR+, HER2-	Br x2 (<u>48</u>, <66)
257	*RAD50*, 1544A > G	Br (39)	Lobular, 2	ER+, PR+, HER2-	Br (69)
225	*RAD50*, 3475+33C > G	Br (43)	Ductal, 1	ER+, PR+, HER2-	Br x2 (<u>52</u>, 77), *Kid (64)*

ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; na, not available; Bil, Bilateral; Br, breast; Ca, cancer with unknown primary site; Cer, cervix; Co, colon; Int, intestines; Kid, kidney; Mel, melanoma; Ov, ovary; Panc, pancreas; Pr, prostate; Re, rectum; Si, Sigma; Sto, stomach; Thy, thyroid; To, tongue; Ute, uterus. ^aHomozygous variant. *Cancers diagnosed in the paternal side of the family are presented in italics.* <u>Cancers diagnosed in siblings or their children of the index patients are underlined.</u> Cancers diagnosed in the children of the index patients are presented in bold.

were identified either in *BRCA1* or *BRCA2* by multiplex ligation-dependent probe amplification (MLPA).

CHEK2 mutation analysis

In *CHEK2*, two previously reported BrCa-associated variants in the Finnish population, c.470T > C and c.1100delC, were identified in 10 of 82 (12.1%) individuals (Tables 2 and 3). The heterozygous c.470T > C variant was observed in eight women of which three were healthy. Two of the c.470T > C variant carriers had bilateral BrCa and they carried also *PALB2* missense variants (an example of the family pedigree of the index individual carrying the both variants is presented in Figure 2, Family 129).

The heterozygous c.1100delC variant was detected in three women (Tables 2 and 3). One woman carrying c.1100delC with an early-onset disease of 26 years of age also carried the c.470T > C and the novel c.792 + 39C > T *CHEK2* variants as well as the *RAD50*, c.2398-32A > G variant (Figure 3, Family 110). A second patient with the c.1100delC variant had bilateral BrCa at the age of 44 years and two other affected individuals in her family (mother and father's sister; Figure 4, Family 264). A third patient with the c.1100delC variant had BrCa diagnosed at the age of 45 years and one affected individual (mother) in her family. This woman carried also the *PALB2* c.1676A > G variant (Figure 5, Family 265). In addition to c.470T > C
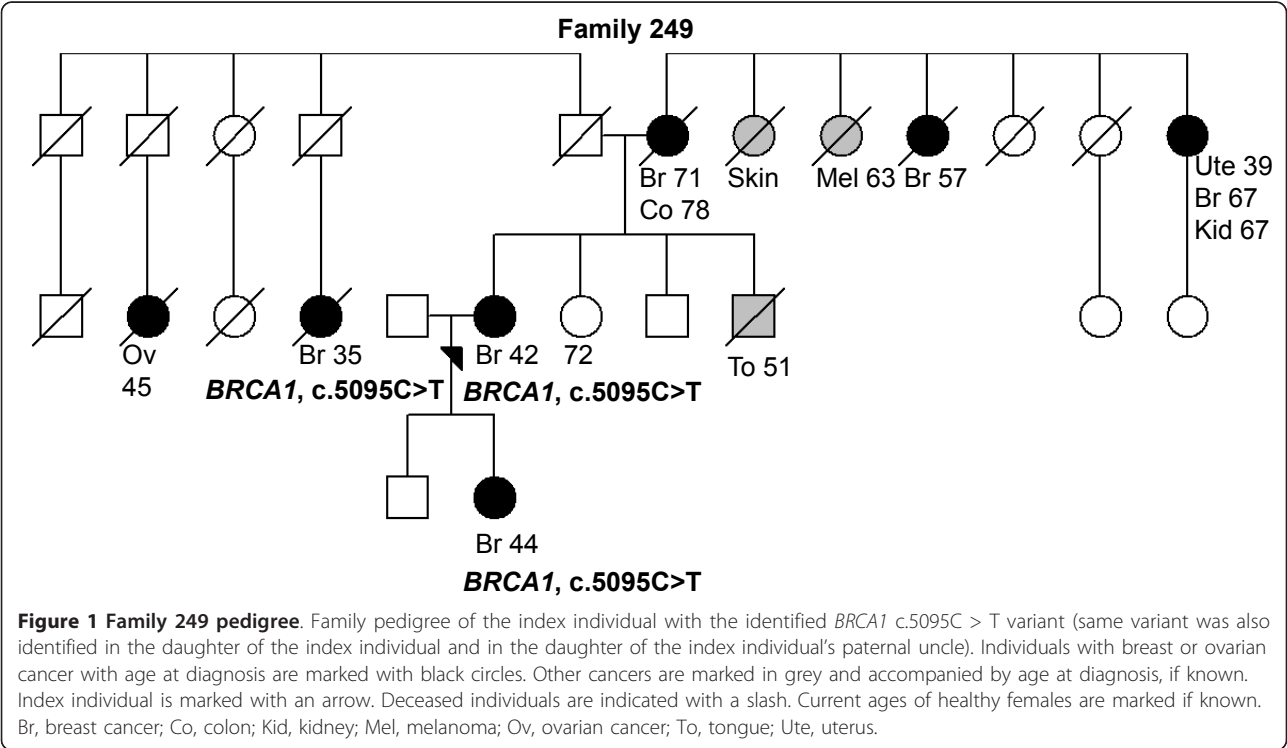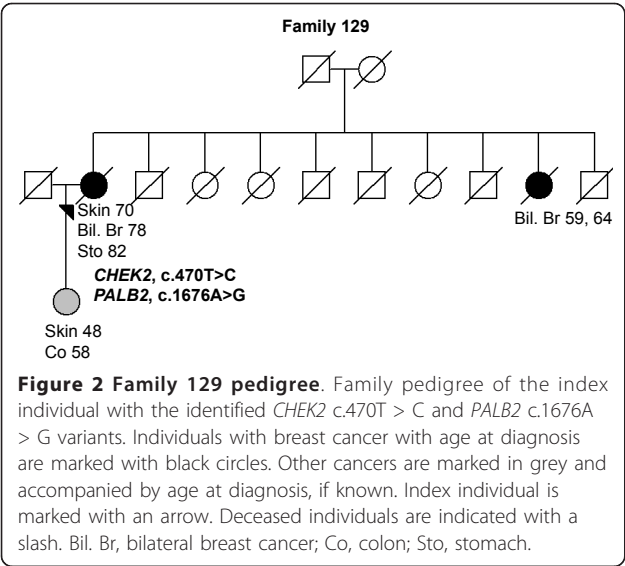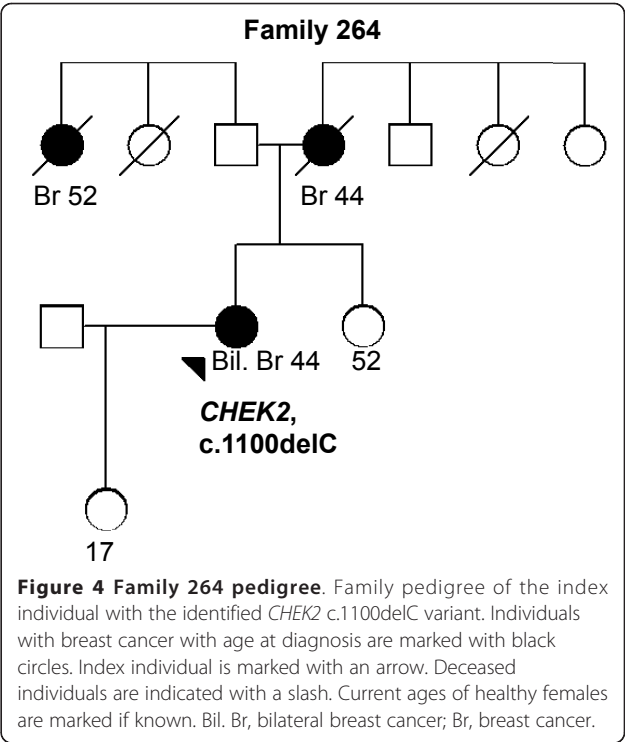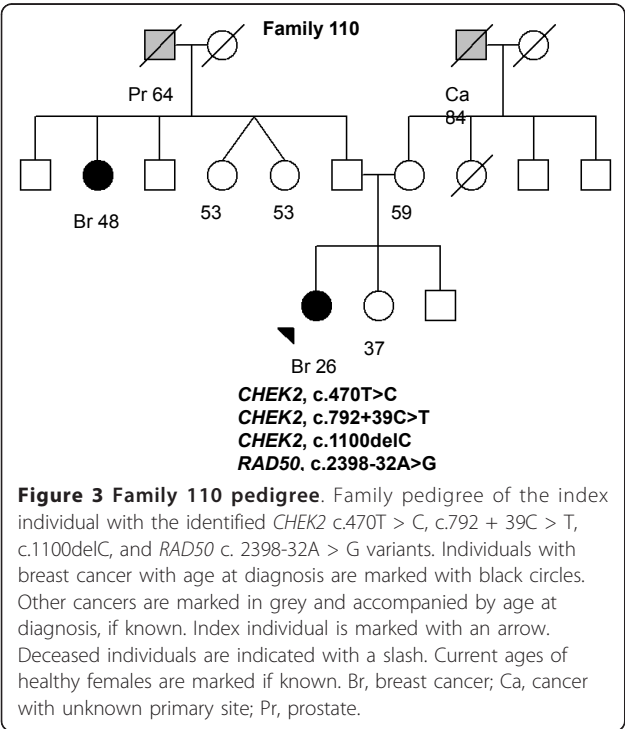

Figure 1 Family 249 pedigree. Family pedigree of the index individual with the identified *BRCA1* c.5095C > T variant (same variant was also identified in the daughter of the index individual and in the daughter of the index individual's paternal uncle). Individuals with breast or ovarian cancer with age at diagnosis are marked with black circles. Other cancers are marked in grey and accompanied by age at diagnosis, if known. Index individual is marked with an arrow. Deceased individuals are indicated with a slash. Current ages of healthy females are marked if known. Br, breast cancer; Co, colon; Kid, kidney; Mel, melanoma; Ov, ovarian cancer; To, tongue; Ute, uterus.

Figure 2 Family 129 pedigree. Family pedigree of the index individual with the identified *CHEK2* c.470T > C and *PALB2* c.1676A > G variants. Individuals with breast cancer with age at diagnosis are marked with black circles. Other cancers are marked in grey and accompanied by age at diagnosis, if known. Index individual is marked with an arrow. Deceased individuals are indicated with a slash. Bil. Br, bilateral breast cancer; Co, colon; Sto, stomach.
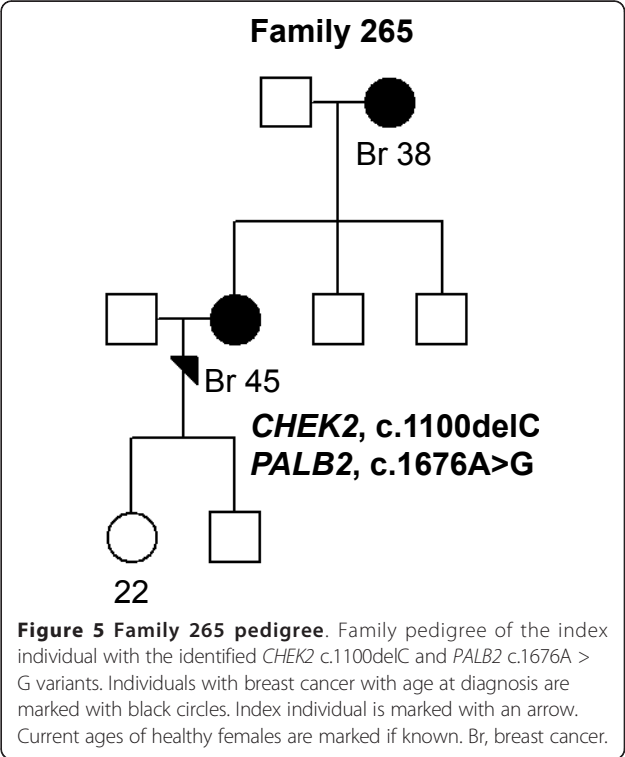
and c.1100delC, five novel variants (Table 2) and one common polymorphism [see Supplementary Table S1 in Additional file 1] were identified in *CHEK2*. The novel nonsynonymous variant, c.1363G > A (Val455Ile), is based on the computational predictions, and is likely benign.

PALB2 mutation analysis

In *PALB2*, altogether nine different variants, including three novel ones, were identified [see Supplementary Table S1 in Additional file 1]. Only one of the identified variants reported previously, c.2586 + 58C > T, has been

Figure 3 Family 110 pedigree. Family pedigree of the index individual with the identified *CHEK2* c.470T > C, c.792 + 39C > T, c.1100delC, and *RAD50* c. 2398-32A > G variants. Individuals with breast cancer with age at diagnosis are marked with black circles. Other cancers are marked in grey and accompanied by age at diagnosis, if known. Index individual is marked with an arrow. Deceased individuals are indicated with a slash. Current ages of healthy females are marked if known. Br, breast cancer; Ca, cancer with unknown primary site; Pr, prostate.

Figure 4 Family 264 pedigree. Family pedigree of the index individual with the identified *CHEK2* c.1100delC variant. Individuals with breast cancer with age at diagnosis are marked with black circles. Index individual is marked with an arrow. Deceased individuals are indicated with a slash. Current ages of healthy females are marked if known. Bil. Br, bilateral breast cancer; Br, breast cancer.

associated with a 36% increase of BrCa risk (odds ratio (OR): 1.36; 95% confidence intervals (CIs), 1.13-1.64; *P* = 0.001) in a Chinese population [25]. We identified the c.2586 + 58C > T variant in 36 of 82 (43.9%) women. A novel heterozygous c.814G > A variant was identified in

Figure 5 Family 265 pedigree. Family pedigree of the index individual with the identified *CHEK2* c.1100delC and *PALB2* c.1676A > G variants. Individuals with breast cancer with age at diagnosis are marked with black circles. Index individual is marked with an arrow. Current ages of healthy females are marked if known. Br, breast cancer.

1 of 82 (1.2%) women but not in population controls (Tables 2 and 3). The c.814G > A variant carrying woman had BrCa diagnosed at the age of 28 years, but no other affected individuals were seen in her family. The c.814G > A variant results in amino acid substitution of glutamic acid to lysine at position 272, which causes a significant change to side chain properties including size and change of the charge to opposite. However, protein predictions by PON-P suggest that variation is neutral. The second novel heterozygous variant, c.1000T > G (Tyr334Asp), was observed in 1 of 82 (1.2%) women and in 4 of 380 (1.1%) population controls. The c.1000T > G variant carrying woman had bilateral BrCa diagnosed at the ages of 45 and 58 years and a family history of three other cancers (Tables 2 and 3, Figure 6, Family 262). She carried also the *CHEK2* c.470T > C variant. However, the protein predictions for the c.1000T > G (Tyr334Asp) variant suggest it to be neutral. A third novel heterozygous variant, c.2205A > G (Pro735Pro), is silent and likely to be neutral. It was observed in 1 of 82 (1.2%) women (Tables 2 and 3). Previously reported *PALB2* missense variants, c.1010T > C, c.1676A > G, c.2794G > A, and c.2993G > A were identified here with frequencies from 1.2% to 12.2% in analyzed individuals (Tables 2 and 3) but the variants have not been associated with BrCa risk (an example of the family pedigree of the index individual
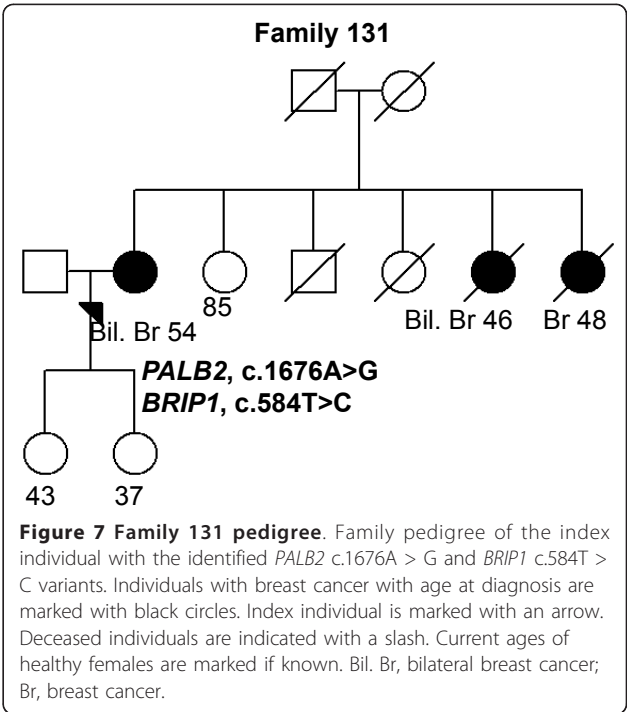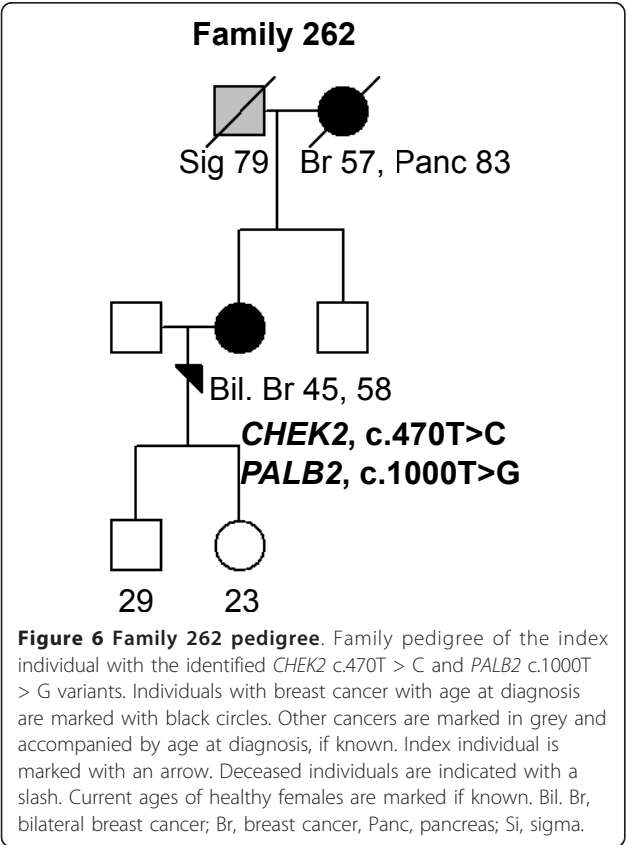
carrying the c.1676A > G variant in addition to the *BRIP1* c.584T > C variant is presented in Figure 7, Family 131).

BRIP1, *RAD50*, and *CDH1* mutation analysis

In *BRIP1*, two silent [see Supplementary Table S1 in Additional file 1] and two missense variants (Tables 2 and 3) were identified. All of the identified variants have been reported previously and they are likely to be neutral. In *RAD50*, altogether seven sequence alterations were observed [see Supplementary Table S1 in Additional file 1] and three of these were novel (Table 2). The novel missense variant, c.1544A > G (Asp515Gly), was observed in 1 of 82 (1.2%) women and in 4 of 384 (1.1%) population controls. The c.1544A > G variant carrying woman had BrCa diagnosed at the age of 39 years and one affected first-degree relative (Table 3). According to protein predictions, c.1544A > G variant is likely to be neutral. Two other novel variants, c.2398-32A > G and c.3475 + 33C > G, were both observed with the frequency of 1 of 82 (1.2%) in analyzed individuals (Tables 2 and 3). In *CDH1*, 10 different sequence alterations were identified [see Supplementary Table S1 in Additional file 1]. All of the variants have been reported previously and they are likely neutral.

MicroRNA database and BLAST search for novel variants

No known miRNA target sites were found in the identified novel variant genomic positions. In BLAST search, *BRCA2* c.72A > T variant position was found to have sequence similarities between rat and cow. *RAD50*

Figure 6 Family 262 pedigree. Family pedigree of the index individual with the identified *CHEK2* c.470T > C and *PALB2* c.1000T > G variants. Individuals with breast cancer with age at diagnosis are marked with black circles. Other cancers are marked in grey and accompanied by age at diagnosis, if known. Index individual is marked with an arrow. Deceased individuals are indicated with a slash. Current ages of healthy females are marked if known. Bil. Br, bilateral breast cancer; Br, breast cancer, Panc, pancreas; Si, sigma.

Figure 7 Family 131 pedigree. Family pedigree of the index individual with the identified *PALB2* c.1676A > G and *BRIP1* c.584T > C variants. Individuals with breast cancer with age at diagnosis are marked with black circles. Index individual is marked with an arrow. Deceased individuals are indicated with a slash. Current ages of healthy females are marked if known. Bil. Br, bilateral breast cancer; Br, breast cancer.

c.1544A > G variant position shared similarities with mouse, rat, cow and chicken. Three novel variant positions in *CHEK2* exon 11 and the *RAD50* c.3475 + 33C > G variant shared sequence similarity between mouse, rat and cow. Variants that occur in the genomic regions that are conserved across species may indicate a pathogenic role.
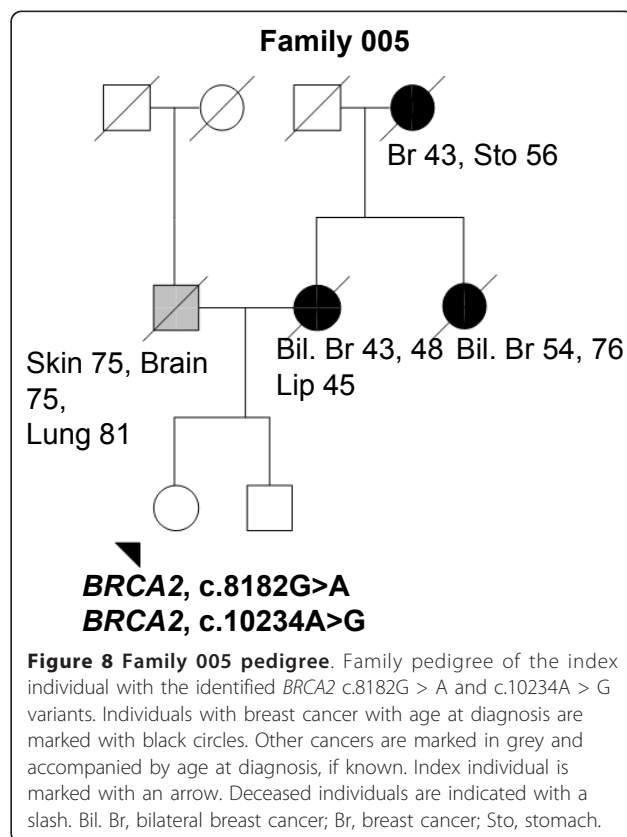
Discussion

In the present study, we screened BrCa susceptibility genes in 82 Finnish high-risk HBOC individuals with no known Finnish *BRCA1/2*-founder mutations. As genetic counseling and surveillance is greatly needed for these individuals and their families, we decided to study *BRCA1/2* in more detail and also to analyze five additional genes that had previously been associated with BrCa risk.

The majority of known *BRCA1/2* alterations are small insertions and deletions or point mutations (BIC database). Also, large genomic rearrangements have been reported in both genes with varying frequencies in different populations [26]. In Finland, so far only Pylkäs *et al.* have reported a large deletion in *BRCA1* identified in a Finnish OvCa family [27]. In our study, no deletions or duplications were found in either *BRCA1* or *BRCA2* by MLPA, which suggests the existence of more restricted alterations. A total of 16 different sequence variants were identified from these two genes [see Supplementary Table S1 in Additional file 1] and only one of the identified variants, c.5095C > T in *BRCA1*, has been classified as a clinically significant mutation in the BIC database. In line with this classification, our BrCa patient carrying this variant had a strong family history of cancer (Tables 2 and 3, Figure 1, Family 249) and two other variant carriers with BrCa were also observed in the same family. The c.5095C > T mutation thus can explain a fraction of the BrCa cases also in the Finnish population. The clinical significance of the *BRCA1* c.4883T > C variant in BrCa predisposition is uncertain [28,29]. Our data support the idea that it is a low-penetrant risk allele, because the variant was observed to be three times more common in analyzed high-risk individuals than healthy population controls (Tables 2 and 3). Novel variant findings in *BRCA2* (Tables 2 and 3) warrant additional studies, especially the novel missense variant, c.72A > T (Leu24Phe), which was shown not to be tolerated by protein prediction. Prediction indicated that the substitution decreases the stability of the produced protein and this might be the mechanism behind the disease for this variant. The amino acid position 24 is located near the N-terminal part of BRCA2. Amino acids 1 to 40 interact with PALB2, and sequence variants in this region have been shown to have effects on the PALB2 and BRCA2 interaction and thus are

suspected to have a role in cancer predisposition [30]. The role of the three *BRCA2* missense variants, c.8182G > A, c.9976A > T, and c.10234A > G, in HBOC risk, is uncertain [31-33]. All three heterozygous variants were observed in two healthy women with a history of BrCa, one carrying the c.9976A > T variant and the other both the c.8182G > A and c.10234A > G variants (Tables 2 and 3, Figure 8, Family 005). At this stage, because we only have samples from the index individuals, no segregation analyses of the variants have been performed, but these families clearly warrant additional studies. In recent risk models, it has been suggested that multiple low-risk variants within the same individual may actually cause a significantly elevated risk for the carrier [17]. The overall low frequency of new variants identified in *BRCA1/2* genes suggests that the present protocol for testing 28 Finnish *BRCA1/2*-founder mutations and PTT of the largest exons is adequate for clinical use to detect the majority of harmful mutations in these two genes in the Finnish population.

Two of the *CHEK2* variants, c.470T > C and c.1100delC, have been widely studied in BrCa predisposition in Finland and elsewhere. Previous studies have shown that the c.1100delC allele confers about a twofold elevated BrCa risk in women, whereas c.470T > C is a lower risk variant [34,35]. Both variants also

Figure 8 Family 005 pedigree. Family pedigree of the index individual with the identified *BRCA2* c.8182G > A and c.10234A > G variants. Individuals with breast cancer with age at diagnosis are marked with black circles. Other cancers are marked in grey and accompanied by age at diagnosis, if known. Index individual is marked with an arrow. Deceased individuals are indicated with a slash. Bil. Br, bilateral breast cancer; Br, breast cancer; Sto, stomach.

associate with other cancers in the Finnish population [36-38]. In our study, two of the *CHEK2* variants, c.470T > C and c.1100delC, were identified in 10 out of 82 analyzed individuals (12.2%) suggesting that the contribution of the two *CHEK2* variants to BrCa risk is remarkable in the high-risk Finnish *BRCA1/2*-founder mutation-negative individuals. However, clinical screening of the *CHEK2* variants has not yet been justified due to unclear clinical consequences related to incomplete segregation of the variants with BrCa in the high-risk BrCa families [39,40]. Based on the findings of this study, we agree that interpretation of the *CHEK2* mutation analysis results is very difficult, because many other gene variants were also identified in individuals with either c.470T > C or c.1100delC variants and some of the variant carriers had not (yet) been diagnosed with BrCa. Thus profound segregation analysis of the c.470T > C and c.1100delC variants for *BRCA1/2*-founder mutation-negative families would be needed to further study clinical significance of these variants. Also the novel variants identified in *CHEK2* should be further analyzed.

PALB2 has been associated with BrCa predisposition in Finland by Erkko *et al.* [12] and the c.1592delT variant was classified as a Finnish founder mutation. In this study the founder deletion was not found, which is probably explained by the limited number of analyzed high-risk HBOC individuals. We identified two novel *PALB2* missense variants, c.814G > A (Glu272Lys) and c.1000T > G (Tyr334Asp), in affected individuals (Tables 2 and 3). Protein predictions suggested a non-pathogenic role of these substitutions but further studies are needed to confirm these findings. None of the four previously reported *PALB2* missense variants, c.1010T > C, c.1676A > G, c.2794G > A, and c.2993G > A, have been associated with BrCa risk [12,41]. Interestingly, these variants were identified also together with other variants in analyzed individuals (Tables 2 and 3). One of the identified intronic variants, c.2586 + 58C > T, has been associated with an increase of BrCa risk in a Chinese population [25] but there is no evidence of that in the Finnish population.

BRIP1 and *RAD50* genes have been shown to have rare BrCa associated variants in familial BrCa patients [14,42]. Here, *BRIP1* mutation analysis revealed only previously reported likely neutral variants. Whereas analysis of *RAD50* identified three novel sequence alterations, including one missense variant, c.1544A > G (Asp515Gly) (Tables 2 and 3). To further study the role of these novel variants, additional analyses are needed. Germline mutations in *CDH1* have been previously found to associate with hereditary diffuse gastric cancer syndrome, but mutations have been also identified in familial invasive lobular BrCa patients without hereditary diffuse gastric

cancer [15,43]. Here, only neutral variants were identified, and all of them have been reported earlier [see Supplementary Table S1 in Additional file 1]. No clear results were found that any of the identified genetic variants in *BRIP1*, *RAD50*, or *CDH1* would increase the BrCa/OvCa risk in the analyzed high-risk Finnish HBOC individuals.

Conclusions

In this study, 13.4% of the analyzed, high-risk *BRCA1/2*-founder mutation-negative HBOC cases can be explained by previously reported mutations in BrCa susceptibility genes. *CHEK2* mutations, c.470T > C and c.1100delC, make a considerable contribution (12.2%) to these high-risk individuals but further segregation analysis are needed to evaluate the clinical significance of these mutations before applying them in clinical use. Novel variant findings warrant additional studies with special interest in the novel missense variant, *BRCA2* c.72A > T (Leu24Phe), which was predicted to bear untolerated mutations and to destabilize the protein. The complex nature of HBOC addresses the need for genome-wide approaches to further study these individuals and to create new tools for genetic counseling. This study also confirmed that our current genetic testing protocol for the 28 Finnish *BRCA1/2*-founder mutations and PTT of the largest exons is sensitive enough for clinical use in the majority of Finnish HBC/HBOC individuals.

Additional material

> **Additional file 1: Supplementary Table S1. All of the identified 54 sequence alterations**. Supplementary Table S1 include detailed information about all of the identified sequence alterations.
>
> **Additional file 2: Supplementary Table S2. Identified breast cancer associated variants in affected 71 individuals**. Supplementary Table S2 includes re-calculated frequencies for *BRCA1* c.5095C > T, *CHEK2* c.470T > C, and *CHEK2* c.1100delC variants in affected 71 index individuals (11 unaffected index individuals excluded).

Author details

[1]Institute of Biomedical Technology, University of Tampere, Biokatu 8, Tampere, 33520, Finland. [2]Centre for Laboratory Medicine, Tampere University Hospital, Biokatu 4, Tampere, 33520, Finland. [3]Department of Pediatrics, Genetics Outpatient Clinic, Tampere University Hospital, Biokatu 8, Tampere, 33520, Finland.

Authors' contributions

KMK participated patient collection, carried out the sequencing, MLPA and statistical analysis and drafted the manuscript. AB carried out and interpreted the HRM analysis of the *CHEK2* gene and helped to draft the methods section of the manuscript. MV performed and interpreted protein prediction analysis *in silico* and helped to draft the manuscript. JS and S-L S participated in the study design and coordination, and helped to draft the manuscript. S-L S also participated in patient collection and was responsible for genetic counseling of patients. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

References

1. Claus EB, Schildkraut JM, Thompson WD, Risch NJ: **The genetic attributable risk of breast and ovarian cancer.** *Cancer* 1996, **77(11)**:2318-2324.
2. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, Ding W: **A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1.** *Science* 1994, **266(5182)**:66-71.
3. Tavtigian SV, Simard J, Rommens J, Couch F, Shattuck-Eidens D, Neuhausen S, Merajver S, Thorlacius S, Offit K, Stoppa-Lyonnet D, Belanger C, Bell R, Berry S, Bogden R, Chen Q, Davis T, Dumont M, Frye C, Hattier T, Jammulapati S, Janecki T, Jiang P, Kehrer R, Leblanc JF, Mitchell JT, McArthur-Morrison J, Nguyen K, Peng Y, Samson C, Schroeder M, Snyder SC, *et al*: **The complete BRCA2 gene and mutations in chromosome 13q-linked kindreds.** *Nat Genet* 1996, **12**:333-337.
4. Vehmanen P, Friedman LS, Eerola H, McClure M, Ward B, Sarantaus L, Kainu T, Syrjakoski K, Pyrhonen S, Kallioniemi OP, Muhonen T, Luce M, Frank TS, Nevanlinna H: **Low proportion of BRCA1 and BRCA2 mutations in Finnish breast cancer families: evidence for additional susceptibility genes.** *Hum Mol Genet* 1997, **6**:2309-2315.
5. Syrjakoski K, Vahteristo P, Eerola H, Tamminen A, Kivinummi K, Sarantaus L, Holli K, Blomqvist C, Kallioniemi OP, Kainu T, Nevanlinna H: **Population-based study of BRCA1 and BRCA2 mutations in 1035 unselected Finnish breast cancer patients.** *J Natl Cancer Inst* 2000, **92**:1529-1531.
6. Li FP, Fraumeni JF Jr: **Rhabdomyosarcoma in children: epidemiologic study and identification of a familial cancer syndrome.** *J Natl Cancer Inst* 1969, **43**:1365-1373.
7. Starink TM, van der Veen JP, Arwert F, de Waal LP, de Lange GG, Gille JJ, Eriksson AW: **The Cowden syndrome: a clinical and genetic study in 21 patients.** *Clin Genet* 1986, **29**:222-233.
8. Giardiello FM, Brensinger JD, Tersmette AC, Goodman SN, Petersen GM, Booker SV, Cruz-Correa M, Offerhaus JA: **Very high risk of cancer in familial Peutz-Jeghers syndrome.** *Gastroenterology* 2000, **119**:1447-1453.
9. Keller G, Vogelsang H, Becker I, Hutter J, Ott K, Candidus S, Grundei T, Becker KF, Mueller J, Siewert JR, Hofler H: **Diffuse type gastric and lobular breast carcinoma in a familial gastric cancer patient with an E-cadherin germline mutation.** *Am J Pathol* 1999, **155**:337-342.
10. Welcsh PL, Owens KN, King MC: **Insights into the functions of BRCA1 and BRCA2.** *Trends Genet* 2000, **16**:69-74.
11. Vahteristo P, Bartkova J, Eerola H, Syrjakoski K, Ojala S, Kilpivaara O, Tamminen A, Kononen J, Aittomaki K, Heikkila P, Holli K, Blomqvist C, Bartek J, Kallioniemi OP, Nevanlinna H: **A CHEK2 genetic variant**

12. contributing to a substantial fraction of familial breast cancer. *Am J Hum Genet* 2002, **71**:432-438.
12. Erkko H, Xia B, Nikkila J, Schleutker J, Syrjakoski K, Mannermaa A, Kallioniemi A, Pylkas K, Karppinen S, Rapakko K, Miron A, Sheng Q, Li G, Mattila H, Bell DW, Haber DA, Grip M, Reiman M, Jukkola-Vuorinen A, Mustonen A, Kere J, Aaltonen LA, Kosma V, Kataja V, Soini Y, Drapkin Rl, Livingston DM, Winqvist R: **A recurrent mutation in PALB2 in Finnish cancer families.** *Nature* 2007, **446**:316-319.
13. Vahteristo P, Yliannala K, Tamminen A, Eerola H, Blomqvist C, Nevanlinna H: **BACH1 Ser919Pro variant and breast cancer risk.** *BMC Cancer* 2006, **6**:19.
14. Heikkinen K, Rapakko K, Karppinen SM, Erkko H, Knuutila S, Lundan T, Mannermaa A, Borresen-Dale AL, Borg A, Barkardottir RB, Petrini J, Winqvist R: **RAD50 and NBS1 are breast cancer susceptibility genes associated with genomic instability.** *Carcinogenesis* 2006, **27**:1593-1599.
15. Guilford P, Hopkins J, Harraway J, McLeod M, McLeod N, Harawira P, Taite H, Scoular R, Miller A, Reeve AE: **E-cadherin germline mutations in familial gastric cancer.** *Nature* 1998, **392**:402-405.
16. Easton DF: **How many more breast cancer predisposition genes are there?** *Breast Cancer Res* 1999, **1**:14-17.
17. Antoniou AC, Easton DF: **Models of genetic susceptibility to breast cancer.** *Oncogene* 2006, **25**:5898-5905.
18. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R, Wareham N, Ahmed S, Healey CS, Bowman R, SEARCH collaborators, Meyer KB, Haiman CA, Kolonel LK, Henderson BE, Le Marchand L, Brennan P, Sangrajrang S, Gaborieau V, Odefrey F, Shen CY, Wu PE, Wang HC, Eccles D, Evans DG, Peto J, *et al*: **Genome-wide association study identifies novel breast cancer susceptibility loci.** *Nature* 2007, **447**:1087-1093.
19. Tommiska J, Seal S, Renwick A, Barfoot R, Baskcomb L, Jayatilake H, Bartkova J, Tallila J, Kaare M, Tamminen A, Heikkila P, Evans DG, Eccles D, Aittomaki K, Blomqvist C, Bartek J, Stratton MR, Nevanlinna H, Rahman N: **Evaluation of RAD50 in familial breast cancer predisposition.** *Int J Cancer* 2006, **118**:2911-2916.
20. Rozen S, Skaletsky H: **Primer3 on the www for general users and for biologist programmers.** *Methods Mol Biol* 2000, **132**:365-386.
21. **Two by two table analysis online. Including chi-square, odds ratio, risk ratio.** [http://www.quantitativeskills.com/sisa/statistics/twoby2.htm].
22. Thusberg J, Vihinen M: **Pathogenic or not? And if so, then how? Studying the effects of missense mutations using bioinformatics methods.** *Hum Mutat* 2009, **30**:703-714.
23. **miRBase.** [http://www.mirbase.org/].
24. **BLAST: Basic Local Alignment Search Tool.** [http://blast.ncbi.nlm.nih.gov/Blast.cgi].
25. Chen P, Liang J, Wang Z, Zhou X, Chen L, Li M, Xie D, Hu Z, Shen H, Wang H: **Association of common PALB2 polymorphisms with breast cancer risk: a case-control study.** *Clin Cancer Res* 2008, **14**:5931-5937.
26. Mazoyer S: **Genomic rearrangements in the BRCA1 and BRCA2 genes.** *Hum Mutat* 2005, **25**:415-422.
27. Pylkas K, Erkko H, Nikkila J, Solyom S, Winqvist R: **Analysis of large deletions in BRCA1, BRCA2 and PALB2 genes in Finnish breast and ovarian cancer families.** *BMC Cancer* 2008, **8**:146.
28. Ostrow KL, McGuire V, Whittemore AS, DiCioccio RA: **The effects of BRCA1 missense variants V1804D and M1628T on transcriptional activity.** *Cancer Genet Cytogenet* 2004, **153**:177-180.
29. Phelan CM, Dapic V, Tice B, Favis R, Kwan E, Barany F, Manoukian S, Radice P, van der Luijt RB, van Nesselrooij BP, Chenevix-Trench G, kConFab , Caldes T, de la Hoya M, Lindquist S, Tavtigian SV, Goldgar D, Borg A, Narod SA, Monteiro AN: **Classification of BRCA1 missense variants of unknown clinical significance.** *J Med Genet* 2005, **42**:138-146.
30. Xia B, Sheng Q, Nakanishi K, Ohashi A, Wu J, Christ N, Liu X, Jasin M, Couch FJ, Livingston DM: **Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2.** *Mol Cell* 2006, **22**:719-729.
31. Deffenbaugh AM, Frank TS, Hoffman M, Cannon-Albright L, Neuhausen SL: **Characterization of common BRCA1 and BRCA2 variants.** *Genet Test* 2002, **6**:119-121.
32. Wagner TM, Hirtenlehner K, Shen P, Moeslinger R, Muhr D, Fleischmann E, Concin H, Doeller W, Haid A, Lang AH, Mayer P, Petru E, Ropp E, Langbauer G, Kubista E, Scheiner O, Underhill P, Mountain J, Stierer M, Zielinski C, Oefner P: **Global sequence diversity of BRCA2: analysis of 71 breast cancer families and 95 control individuals of worldwide populations.** *Hum Mol Genet* 1999, **8**:413-423.

33. Mazoyer S, Dunning AM, Serova O, Dearden J, Puget N, Healey CS, Gayther SA, Mangion J, Stratton MR, Lynch HT, Goldgar DE, Ponder BA, Lenoir GM: **A polymorphic stop codon in BRCA2.** *Nat Genet* 1996, **14**:253-254.

34. CHEK2 Breast Cancer Case-Control Consortium: **CHEK2*1100delC and susceptibility to breast cancer: a collaborative analysis involving 10,860 breast cancer cases and 9,065 controls from 10 studies.** *Am J Hum Genet* 2004, **74**:1175-1182.

35. Kilpivaara O, Vahteristo P, Falck J, Syrjakoski K, Eerola H, Easton D, Bartkova J, Lukas J, Heikkila P, Aittomaki K, Holli K, Blomqvist C, Kallioniemi OP, Bartek J, Nevanlinna H: **CHEK2 variant I157T may be associated with increased breast cancer risk.** *Int J Cancer* 2004, **111**:543-547.

36. Seppala EH, Ikonen T, Mononen N, Autio V, Rokman A, Matikainen MP, Tammela TL, Schleutker J: **CHEK2 variants associate with hereditary prostate cancer.** *Br J Cancer* 2003, **89**:1966-1970.

37. Kilpivaara O, Laiho P, Aaltonen LA, Nevanlinna H: **CHEK2 1100delC and colorectal cancer.** *J Med Genet* 2003, **40**:e110.

38. Kilpivaara O, Alhopuro P, Vahteristo P, Aaltonen LA, Nevanlinna H: **CHEK2 I157T associates with familial and sporadic colorectal cancer.** *J Med Genet* 2006, **43**:e34.

39. Nevanlinna H, Bartek J: **The CHEK2 gene and inherited breast cancer susceptibility.** *Oncogene* 2006, **25**:5912-5919.

40. Ripperger T, Gadzicki D, Meindl A, Schlegelberger B: **Breast cancer susceptibility: current knowledge and implications for genetic counselling.** *Eur J Hum Genet* 2009, **17**:722-731.

41. Rahman N, Seal S, Thompson D, Kelly P, Renwick A, Elliott A, Reid S, Spanova K, Barfoot R, Chagtai T, Jayatilake H, McGuffog L, Hanks S, Evans DG, Eccles D, Breast Cancer Susceptibility Collaboration (UK), Easton DF, Stratton MR: **PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene.** *Nat Genet* 2007, **39**:165-167.

42. Cantor SB, Bell DW, Ganesan S, Kass EM, Drapkin R, Grossman S, Wahrer DC, Sgroi DC, Lane WS, Haber DA, Livingston DM: **BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function.** *Cell* 2001, **105**:149-160.

43. Masciari S, Larsson N, Senz J, Boyd N, Kaurah P, Kandel MJ, Harris LN, Pinheiro HC, Troussard A, Miron P, Tung N, Oliveira C, Collins L, Schnitt S, Garber JE, Huntsman D: **Germline E-cadherin mutations in familial lobular breast cancer.** *J Med Genet* 2007, **44**:726-731.

44. UCSC Genome Browser Home. [http://genome.ucsc.edu/].

45. ExPASy - UniProt Knowledgebase: Swiss-Prot and TrEMBL. [http://au.expasy.org/sprot/].

46. SNP Home. [http://www.ncbi.nlm.nih.gov/snp].

47. Breast Cancer Information Core. [http://research.nhgri.nih.gov/bic/].

PLOS ONE

Copy Number Variation Analysis in Familial *BRCA1/2*-Negative Finnish Breast and Ovarian Cancer

Kirsi M. Kuusisto[1], Oyediran Akinrinade[1], Mauno Vihinen[2], Minna Kankuri-Tammilehto[3], Satu-Leena Laasanen[4], Johanna Schleutker[1,5]*

1 Institute of Biomedical Technology/BioMediTech, University of Tampere and Fimlab Laboratories, Tampere, Finland, 2 Department of Experimental Medical Science, Lund University, Lund, Sweden, 3 Department of Clinical Genetics, Turku University Hospital, Turku, Finland, 4 Department of Pediatrics, Genetics Outpatient Clinic, and Department of Dermatology, Tampere University Hospital, Tampere, Finland, 5 Department of Medical Biochemistry and Genetics, Institute of Biomedicine, University of Turku, Turku, Finland

Abstract

Background: Inherited factors predisposing individuals to breast and ovarian cancer are largely unidentified in a majority of families with hereditary breast and ovarian cancer (HBOC). We aimed to identify germline copy number variations (CNVs) contributing to HBOC susceptibility in the Finnish population.

Methods: A cohort of 84 HBOC individuals (negative for *BRCA1/2*-founder mutations and pre-screened for the most common breast cancer genes) and 36 healthy controls were analysed with a genome-wide SNP array. CNV-affecting genes were further studied by Gene Ontology term enrichment, pathway analyses, and database searches to reveal genes with potential for breast and ovarian cancer predisposition. CNVs that were considered to be important were validated and genotyped in 20 additional HBOC individuals (6 CNVs) and in additional healthy controls (5 CNVs) by qPCR.

Results: An intronic deletion in the *EPHA3* receptor tyrosine kinase was enriched in HBOC individuals (12 of 101, 11.9%) compared with controls (27 of 432, 6.3%) (OR = 1.96; P = 0.055). *EPHA3* was identified in several enriched molecular functions including receptor activity. Both a novel intronic deletion in the *CSMD1* tumor suppressor gene and a homozygous intergenic deletion at 5q15 were identified in 1 of 101 (1.0%) HBOC individuals but were very rare (1 of 436, 0.2% and 1 of 899, 0.1%, respectively) in healthy controls suggesting that these variants confer disease susceptibility.

Conclusion: This study reveals new information regarding the germline CNVs that likely contribute to HBOC susceptibility in Finland. This information may be used to facilitate the genetic counselling of HBOC individuals but the preliminary results warrant additional studies of a larger study group.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Johanna.Schleutker@utu.fi

Introduction

Breast cancer (BC) is the most common cancer among women in western countries, including Finland. Inherited BC risk is known to be associated with rare, highly penetrant variants, mainly single nucleotide polymorphisms (SNPs) and small insertions and deletions (indels) in *BRCA1* and *BRCA2*, which account for nearly 20% of hereditary breast and/or ovarian cancer (HBOC) cases in Finland [1–3]. Additionally, variants in other *BRCA1/2* interacting genes, including *CHEK2*, *PALB2*, *RAD51C*, and *Abraxas*, are known to account for a low proportion of HBOC susceptibility in the Finnish population [4–7].

In addition to SNPs and small indels, copy number variations (CNVs) contribute to susceptibility to complex diseases and disorders [8]. A CNV is a segment of DNA (1 kb or larger) that presents an altered copy number compared with the reference genome [9]. Depending on the location, CNVs may affect target gene expression through a dosage effect or by disrupting gene regulatory elements [10]. CNVs were initially associated with neurological disorders, but studies have demonstrated the role of CNVs also in other diseases, including several cancers [11–15].

Despite the fact that several heritable risk factors for breast and ovarian cancer have been recognised, in the majority (up to 80%) of HBOC families, inherited risk is likely explained by yet unknown factors, which makes genetic counselling and clinical surveillance challenging. The contribution of rare germline CNVs to breast and ovarian cancer susceptibility has also been established in the Finnish population, but their role is mostly unexplored [16,17]. Therefore, new information regarding germline CNVs and their role in HBOC predisposition is needed to identify CNVs that may be used clinically to facilitate the genetic counselling of HBOC families.

To determine additional genetic factors contributing to HBOC susceptibility in the Finnish population and gain new information

for genetic counselling, we analysed germline CNVs in a cohort of 84 well-characterised HBOC *BRCA1/2*-founder mutation-negative Finnish individuals who have been pre-screened for the most common high- and moderate-penetrant genes [18].

Materials and Methods

Study Material

Index individuals from 84 HBOC families were collected from the Tampere University Hospital Genetics Outpatient Clinic between January 1997 and May 2008. Individuals were selected according to previously reported high-risk hereditary BC criteria [18]. All individuals had been determined to be founder mutation-negative by minisequencing the 28 previously known Finnish *BRCA1/2* mutations and a protein truncation test (PTT) for *BRCA1* exon 11 and *BRCA2* exons 10 and 11. Eighty-one of the individuals included in this study have previously been characterised and screened for germline alterations in seven known BC-associated genes [18]. In addition, the index individuals from three additional HBOC families were included (described in File S1). For CNV validation analysis, index individuals from 20 additional HBOC families, collected from Turku University Hospital Genetics Outpatient Clinic between 2007 and 2011 were utilised. Clinical characteristics of the 20 additional HBOC individuals (negative for *BRCA1/2*-mutations) are described in File S2. As controls, 905 DNA samples from anonymous healthy females, collected from the Finnish Red Cross, were used. All of the HBOC individuals studied have been informed of the analyses, and they have given written consent to use their existing DNA samples. Permission for the research project has been received from the Ethical Committees of Tampere and Turku University Hospitals and the National Authority for Medicolegal Affairs.

Copy Number Variation Analysis

The DNA samples from 84 HBOC individuals and 36 controls were genotyped by using the genome-wide SNP array HumanCytoSNP-12 v.2.1 Beadchip (Illumina, Inc, San Diego, CA, USA), which targets regions of known cytogenetic importance. Sample preparation was performed according to the Infinium II assay protocol (Illumina, Inc, San Diego, CA, USA) at the Institute for Molecular Medicine, Finland. Log R Ratios (LRRs), B Allele frequencies (BAF), and X and Y channel intensities for each sample were exported from normalised Illumina data using GenomeStudio software (GSGTv1.7.4) to perform CNV calling. All of the samples had call rates greater than 99.5%. High sample quality was ensured by applying previously reported quality criteria [19]. Thus, 81 HBOC individuals and 35 controls were suitable for analysis. CNV calling was performed with the PennCNV (2009Aug27) program [19]. Additionally, two other programs, QuantiSNP v2.3 [20] and cnvPartition v3.1.6 (Illumina Inc, San Diego, CA, USA) were used to confirm the PennCNV results when selecting CNVs for validation. Programs were used with default parameters. CNVs spanning less than three SNPs were filtered out.

Statistical Analyses

CNV distribution and median lengths were compared between HBOC individuals and controls using the Wilcoxon test (R v2.15.2, R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria). CNV carrier frequencies between HBOC individuals and controls were compared with the Fisher's exact or χ^2 tests (R v2.15.2 and PLINK v.1.07 [21]). All *P*-values were two-sided. A *P*-value<0.05 was considered statistically significant. Furthermore, a VCD package was implemented in R

to estimate the numerical values of the odds ratios for enrichment analysis in case a non-numerical value was returned from the Fisher's exact test [22].

CNV Validation and Genotyping by Quantitative Real-time PCR (qPCR)

Selected CNVs were validated (6 CNVs) and genotyped in 20 additional HBOC individuals (6 CNVs) and in 299–869 additional healthy female controls (5 CNVs) by TaqMan® Copy Number Assays and TaqMan® real-time PCR, respectively, on an ABI PRISM 7900 sequence detection system (Applied Biosystems, Foster City, CA, USA). The following pre-designed TaqMan® Copy Number Assays were used: Hs04703682_cn (2q34), Hs03458738_cn (3p11.1), Hs03253932_cn (5q15), Hs06178677_cn (8p23.2), Hs02640223_cn (17q21.31), and Hs04482315_cn (19q13.41). As an internal standard, a TaqMan® RNaseP Reference Assay (Applied Biosystems, Part Number 4403326) was run with the pre-designed TaqMan® Copy Number Assays in a duplex, real-time PCR reaction (see File S3 for more details).

Data Analysis

Identified CNVs were queried for overlap with the Database of Genomic Variants (DGV), Toronto (http://projects.tcag.ca/variation/) using NCBI Genome Build 36 (hg 18). A CNV locus was considered novel if it did not overlap with any of the established CNV loci in the DGV. CNVs were annotated using NCBI RefSeq genes (http://www.ncbi.nlm.nih.gov/RefSeq/) to identify genes/exons overlapping the observed CNV loci. For intergenic CNVs, the loci were expanded upstream and downstream of the CNV to identify neighbouring genes.

Enrichment analyses, including Gene Ontology (GO) terms, KEGG pathways, Pathway Commons, and Wikipathways, were performed for CNV-affecting genes to reveal common functions of the gene products using the Web-based Gene Set Analysis Toolkit V2 (WebGestalt2) [23]. Furthermore, CNV-affected genes were queried for overlap against genes listed in the NCBI Online Mendelian Inheritance in Man (OMIM) database (http://www.ncbi.nlm.nih.gov/omim) to identify genomic loci associated with genetic disorders. In addition, a Genetic Association Database (GAD) (http://geneticassociationdb.nih.gov/) search was performed to identify genes analysed in previous association studies for complex diseases and disorders.

Results

We performed genome-wide CNV analysis with a SNP array targeting regions of known cytogenetic importance for individuals from 84 Finnish HBOC families and 36 healthy controls. After applying the quality control criteria, 81 HBOC individuals and 35 controls (n = 116) were included in the data analysis. The aim of this study was to identify germline CNVs contributing to HBOC susceptibility in Finnish families.

The PennCNV program was used to detect 545 autosomal CNVs at 273 different genomic regions in HBOC individuals and controls (n = 116). All of the identified CNVs are presented in detail in Table S1. A summary of the CNVs identified by PennCNV are shown in Table 1. The most important observations are that the average number of CNVs was slightly higher in HBOC individuals compared with controls, and deletions were more frequent in HBOC individuals. There was no statistically significant difference in the median size the CNVs between the HBOC individuals and controls (52.3 kb vs. 50.5 kb; *P* = 0.90). However, the median deletion size in HBOC individuals was

Table 1. Summary of the identified copy number variations (CNVs) by PennCNV in 81 hereditary breast and/or ovarian cancer (HBOC) individuals and 35 controls.

	Average no per sample	Median size (kb)	Gene-affecting (%)	Novel CNVs (%)
All CNVs (n = 545)				
HBOC individuals	392/81 (4.8)	52.3	228/392 (0.58)	37/392 (0.09)
Controls	153/35 (4.4)	50.5	85/153 (0.56)	11/153 (0.07)
HBOC individuals only	215/81 (2.7)	52.5	141/215 (0.66)	36/215 (0.17)
Deletions (n = 300)				
HBOC individuals	222/81 (2.7)	39.2	109/222 (0.56)	30/222 (0.14)
Controls	78/35 (2.2)	56.8	34/78 (0.44)	4/78 (0.05)
HBOC individuals only	116/81 (1.4)	34.6	72/116 (0.62)	29/116 (0.25)
Duplications (n = 245)				
HBOC individuals	170/81 (2.1)	68.7	119/170 (0.70)	7/170 (0.04)
Controls	75/245 (2.1)	47.5	51/75 (0.68)	7/75 (0.09)
HBOC individuals only	99/245 (1.2)	60.8	69/99 (0.70)	7/99 (0.07)

Abbreviations: no = number.
doi:10.1371/journal.pone.0071802.t001

smaller compared with the controls (39.2 kb vs. 56.8 kb; $P = 0.07$). In contrast, the median duplication size was significantly larger ($P = 0.01$) in HBOC individuals compared with controls (68.7 kb vs. 47.5 kb).

Annotation of all of the 545 CNVs against the genes in the NCBI RefSeq database revealed 313 (57.4%) gene-affecting CNVs (Table 1). Most importantly, gene-affecting deletions were more common in HBOC individuals compared with controls (Table 1). The identified CNVs were compared with healthy control sample data collected in the Database of Genomic Variants. The main observation was that the proportion of novel deletions to all deletions in HBOC individuals was nearly three times larger compared with controls (Table 1). In contrast, novel duplications in HBOC individuals were observed less frequently compared with controls (Table 1).

In this study, we focused on CNVs with the following characteristics: they were enriched in HBOC individuals compared with controls and 1) affected known or potential genes contributing to HBOC predisposition (3 CNVs); or 2) they were homozygous, and carriers presented with interesting clinical outcomes (1 CNV); or 3) they were not reported in the Database of Genomic Variants and affected genes related to BC (2 CNVs). CNVs of interest were confirmed by another program (QuantiSNP or cnvPartition). In total, six CNVs were selected for further validation by qPCR, and they were genotyped in additional cohort of index individuals from 20 HBOC families and five of the CNVs were genotyped in 299–869 additional healthy female controls (Table 2). The CNVs were correlated with clinical data from the HBOC individuals (Table 3).

All six validated CNVs listed in Table 2 are located in genomic regions related to BC. CNVs in the intronic regions of *ERBB4* and *EPHA3* were enriched in HBOC individuals compared with controls (Table 2). *EPHA3* and *ERBB4* encode proteins that are involved in important signalling pathways. A homozygous deletion in the 5q15 locus was identified in one BC patient (1 out of the 101, 1.0%) with drastic clinical characteristics (Tables 2 and 3). This homozygous deletion was observed only in 1 out of the 899 (0.1%) healthy controls (Table 2). Deletions affecting the intronic region of the *CSMD1* tumor suppressor gene and exonic regions of the highly penetrant *BRCA1* were observed only in 1 out of the 101

(1.0%) HBOC individuals and *CSMD1* deletion was identified in 1 out of the 436 (0.2%) controls (Table 2). Because large deletions in *BRCA1* are known to predispose to HBOC, there was no need to screen for the deletion in additional controls. A duplication affecting the coding region of the *ERVV-2* gene, which belongs to endogenous retroviruses, was more commonly homozygous in HBOC individuals compared with controls (Table 2).

The clinical characteristics and family cancer history for individuals with HBOC with the six validated CNVs are presented in Table 3 (only CNVs identified in our original cohort of 81 HBOC individuals first analysed in the SNP array are presented). Most importantly, 2 out of the 5 individuals with HBOC with a novel deletion at the 2q34 *ERBB4* locus had bilateral BC that was diagnosed at ≤43 years of age (Table 3; families 221 and 212). We were able to analyse the segregation of the 2q34 deletion in family 249 (Table 3) in which a deleterious *BRCA1* variant was previously identified in three individuals (Figure 1) [18]. The 2q34 deletion was identified in the index's mother (homozygous) and two paternal cousins (heterozygous) (Figure 1). However, the index's daughter did not carry the deletion (Figure 1). A common feature for all of the 3p11.1 deletion (at *EPHA3* locus) carriers was ductal BC diagnosed at ≤50 years and positive hormone receptor status (6 out of the 8 carriers) in the cohort of 81 HBOC individuals (Table 3). In the second cohort of 20 additional HBOC individuals, 3p11.1 deletion was identified in two BC patients, one ovarian cancer patient and a patient who had both breast and ovarian cancer (File S2). Interestingly, all three patients with BC presented ductal form of the cancer and estrogen and progesterone receptor positive status (File S2). Intergenic deletion in the 5q15 region was of great interest because it was found as a homozygous deletion in a BC patient who was diagnosed at age 29 years and died of BC at the same age (Table 3, family 123). Additionally, one heterozygous 5q15 deletion carrier had BC diagnosed at an early age (24 years) and the other had thyroid and cervical cancers in addition to BC diagnosed before age 40 years (Table 3; families 250 and 246). A novel deletion of high interest at 8p23.2, which affects the *CSMD1* intronic region, was identified in a patient with ductal grade 2, hormone receptor positive BC diagnosed at a relatively early age (36 years) with a paternal family history of BC (Table 3, family 128 and Figure 2). A deletion affecting *BRCA1*,

Table 2. Validated copy number variations.

Cytoband[a]	Gene(s)	Type	Size (kb)[b]	Carrier frequency HBOC ind[c]	Controls[d]	P-values	OR; 95%CI	Status[e]
2q34	ERBB4	intronic deletion	28.7–59.0	0.050 (5/101)	0.034 (12/358)	0.457	1.49; 0.52–4.28	Novel
3p11.1	EPHA3	intronic deletion	14.6	0.119 (12/101)	0.063 (27/432)	0.055	1.96; 0.97–3.94	Reported
5q15	–	intergenic deletion	49.8	0.050 (5/101)[f]	0.063 (57/899)	0.845	0.92; 0.39–2.16	Reported
8p23.2	CSMD1	intronic deletion	10.8	0.010 (1/101)	0.002 (1/436)	0.259	4.33; 0.27–69.57	Novel
17q21.31	BRCA1, NBR1, NBR2	exonic deletion	99.0	0.010 (1/101)	0 (0/35)	0.555	na	Reported
19q13.41	ERVV-2	exonic duplication	15.8–26.9	0.109 (11/101)[g]	0.102 (34/334)	0.322	1.37; 0.73–2.55	Reported

Abbreviations: CI = confidence interval; na = not available; OR = odds ratio.
[a]According to the NCBI Genome Build 36.1 (hg 18). Exact start and end positions of the CNVs are provided in Table S1.
[b]Size reported in HBOC individuals analysed in the SNP array (may vary between individuals).
[c]Combined frequencies of original cohort of 81 HBOC individuals (analysed in the SNP array) and cohort of 20 additional HBOC individuals (genotyped by TaqMan® Copy Number Assays). CNVs in the 2q34, 5q15, 8p23.2, and 17q21.31 regions were not observed in additional cohort of 20 HBOC individuals. Heterozygous deletion (copy number 1) in the 3p11.1 region was also identified in 4 out of the 20 additional HBOC individuals (File S2). Heterozygous duplication (copy number 3) in the 19q13.41 region was also identified in 3 out of the 20 additional HBOC individuals (File S2). Homozygous duplication (copy number 4) in the 19q13.41 region was identified in 1 out of the 20 additional HBOC individuals (File S2).
[d]Thirty-five controls were first analyzed in the SNP array. CNVs were also screened in additional controls by TaqMan® Copy Number Assays (excluding BRCA1 affecting CNV since large deletions in BRCA1 coding regions are known to associate with breast and ovarian cancer susceptibility).
[e]Search against the Database of Genomic Variants (DGV).
[f]Deletion in the 5q15 region was homozygous (copy number 0) in 1 out of the 101 (0.010) HBOC individuals and in 1 out of the 899 (0.001) controls and heterozygous (copy number 1) in 4 out of the 101 (0.040) HBOC individuals and in 56 out of the 899 (0.062) controls.
[g]Duplication in the 19q13.41 region was homozygous (copy number 4) in 4 out of the 101 (0.040) HBOC individuals and in 3 out of the 334 (0.009) controls and heterozygous (copy number 3) in 7 out of the 101 (0.069) HBOC individuals and in 31 out of the 334 (0.093) controls.
doi:10.1371/journal.pone.0071802.t002

NBR1, and NBR2 at 17q21.31 was identified in a patient with hormone receptor-negative BC with a family history of ovarian cancer (Table 3, family 252).

Enrichment analysis was performed for the CNV-affecting genes to identify common functions of the gene products. EPHA3, ERBB4 and BRCA1 were identified in several GO term categories and pathways that were significantly overrepresented (P<0.05) (presented in detail in Table S2). Both EPHA3 and ERBB4 were identified to have molecular functions related to receptor activity, transmembrane receptor activity, molecular transducer activity, and signal transducer activity. In contrast, BRCA1 was identified in several pathways related to DNA double-strand breaks and repair. In addition, Online Mendelian Inheritance in Man (OMIM) and The Genetic Association database searches revealed the role of CSMD1 in BC.

Discussion

In this study, we aimed to identify CNVs contributing to HBOC susceptibility in Finland and obtain new information for the genetic counselling of HBOC families. We utilised a cohort of well-characterised BRCA1/2-founder mutation-negative individuals from 84 Finnish hereditary breast and/or ovarian cancer families who had been previously screened for variations in seven known BC genes [18].

Here, we identified more gene-disrupting deletions in HBOC individuals compared with controls suggesting that altered function of their protein products, particularly in critical pathways, could explain pathogenic events in HBOC individuals. Additionally, a proportion of novel gene-affecting deletions, which were not reported in healthy controls in the database, was higher in HBOC individuals compared with controls, suggesting that these novel CNVs are more likely to be disease -related.

We focused on CNVs that were enriched in HBOC individuals compared with controls and affected genes that likely play a role in HBOC predisposition. In addition, one intergenic deletion was

also included for further validation based on the homozygous form of the aberration and notably poor clinical characteristics of the carrier. Thus, six CNVs were considered to be the most relevant for further validation. Because our sample number in the SNP array was limited, we also genotyped the six CNVs in a cohort of 20 additional HBOC individuals. Furthermore, five of the CNVs were genotyped in 299–869 additional healthy controls. Because clinical characteristics of the additional cohort of 20 HBOC individuals were comparable to our original cohort of 81 HBOC individuals, we combined the observed frequencies of the CNVs in both cohorts in Table 2. Additionally, we performed segregation analysis of one family to determine how the CNV co-segregated with the disease and another BC-associated variant. The CNVs were compared with the clinical data of the HBOC individuals.

In this study, the most frequently observed aberration in HBOC individuals was a deletion disrupting the EPHA3 intronic region (Table 2). EPHA3 belongs to the ephrin receptor subfamily of the receptor tyrosine kinase (RTK) family, which plays an important role in normal cell physiology and disease pathogenesis [24]. Ephrin receptor signalling together with ephrin-ligands is known to regulate both tumour growth and suppression in several different cancers including BC [25]. According to recent studies, altered EPHA3 expression is associated with gastric and colorectal cancers, and CNVs in the EPHA3 region have been found to be associated with haematologic malignancies [26–28]. However, haematologic malignancies were not observed in EPHA3 deletion carriers in this study. Our data suggest that an intronic deletion may disrupt the EPHA3 regulatory elements, thus leading to altered protein function and pathogenic BC events. Thus, considering the important role of EPHA3 in signalling pathways, the segregation of the intronic deletion should be studied in the families and the deletion should be further screened in a larger sample set.

The intergenic 5q15 deletion, particularly as a homozygous deletion, is highly interesting from a clinical perspective. This deletion was identified in a patient who had been diagnosed with

Table 3. The clinical characteristics and family cancer history for HBOC individuals analysed in the SNP array with the six validated copy number variations.

Family	Variation	Cancer (age at dg)	Br/Ov Ca histology/grade	Receptor Status	Ca cases in the family (age at dg if known)
221	2q34 del	Bil. Br (39, 42)	duct, gr 1 and	ER+, PR+, HER2− and	Br (51), *Panc (54)*
			duct, gr 2	ER+, PR+, HER2−	
212	2q34 del	Bil. Br (43)	duct, gr na and na	ER+, PR+, HER2− and na	Br (52)
263	2q34 del	Ov (69), Br (72)	duct, gr 3	ER−, PR−, HER2−	–
249	2q34 del	Br (42)	medullary, na	na	Br (35, **44,** 57, 67, 71), Ute (39), Kid (67), Mel (63)
					Ov (45), Skin, To (51), Co (78)
132	2q34 del	Br (47)	duct, gr 1	ER+, PR+, HER2 na	Br (38)
232	3p11.1 del	Br (34)	duct, na	ER+, PR+, HER2 na	Br (39)
244	3p11.1 del	Br (45)	duct, gr 2	ER+, PR+, HER2−	Bil. Br (<45), Br (<35, 46), Brain (67)
121	3p11.1 del	Br (50)	duct, gr 3	ER−, PR−, HER2+	4xBr (36, 39, 40, 48)
207	3p11.1 del	Br (38)	duct, gr 3	na	Bil.Br (64)
230	3p11.1 del	Br (33), Kid (37)	duct, gr 1	ER+, PR+, HER2−	*Br (70)*
118	3p11.1 del	Ov (32), Br (40), Mel (41)	Mucinous and	ER+, PR+, HER2−	–
	19q13.41 dup		duct, gr 2		
269	3p11.1 del	Br (36)	duct, gr 1	ER+, PR+, HER2−	*3xBr (52, 70, 72), Skin (66)*
225	3p11.1 del	Br (43)	duct, gr 1	ER+, PR+, HER2−	*2xBr (52, 77), Kid (64)*
123	5q15 del*	Br (29)	duct, gr 2	ER+, PR−, HER2−	Br (65), Eso (73)
250	5q15 del	Br (24)	duct, gr 3	ER+, PR+, HER2+	Cer (30), *Ov (83)*
264	5q15 del	Bil. Br (44)	lob, gr 2	ER+, PR+, HER2−	Br (44, *52*)
	19q13.41 dup				
129	5q15 del	BCC (70), Bil. Br (78),	left: lob, gr 2,	left: ER−, PR−, HER2−,	Bil. Br (59), **BCC (48), Co (58)**
		Sto (82)	right: duct, gr 1	right: ER+, PR+, HER2−	
246	5q15 del	Thy (30), Cer (33), Br (39)	duct, gr 3	ER−, PR−, HER2+	2xBr (**49**, 54), *Rectum (61)*
128	8p23.2 del	Br (36)	duct, gr 2	ER+, PR+, HER2−	*2x Br (45, 58), GI (57)*, Mel (69)
252	17q21.31 del	Br (46)	duct, gr 3	ER−, PR−, HER2−	Bil. Ov (46), Ov (44)
240	19q13.41 dup*	Br (53)	duct, gr 3	ER+, PR−, HER2+	*2xBr (42, 62)*
206	19q13.41 dup	Br (53)	duct, gr 1	ER−, PR−, HER2−	Bil. Br (64), **Br (49)**
133	19q13.41 dup	Br (48)	duct, gr 2	ER+, PR+, HER2−	*2xBr (73, 79), Int*, BCC (60)
113	19q13.41 dup*	Br (51), BCC (55)	duct, gr 3	ER−, PR−, HER2+	Br (35)
239	19q13.41 dup*	Br (37)	duct, gr 2	ER+, PR+, HER2−	*Br (>90), Co*

*Homozygous CNV.
Abbreviations: BCC = Basal-cell carcinoma; Bil. Br = bilateral breast; Br = breast; Ca = cancer; Cer = cervix in situ carcinoma/cervix carcinoma; Co = colon; Dg = diagnosis; Del = deletion; Duct = ductal; Dup = duplication; Eso = esophagus; GI = gastrointestinal; gr = grade; Int = intestine; Kid = kidney; Lob = lobular; Mel = melanoma; na = not available; Ov = ovary; Panc = pancreatic; Sto = stomach; Thy = thyroid; To = tongue; Ute = ute. Cancers diagnosed in the paternal side of the family are presented in italics. Cancers diagnosed in siblings or their children of the index patients are underlined. Cancers diagnosed in the children of the index patients are presented in bold.
doi:10.1371/journal.pone.0071802.t003

BC at age 29 and died of the disease at the same age. Homozygous deletion of the 5q15 locus was extremely rare in healthy controls (1 out of the 899, 0.1%) (Table 2), which emphasises the importance of the variation. Moreover, it is possible that a fraction of the anonymous controls may develop breast or ovarian cancer later in life although they were healthy at the time of the blood draw. The 5q15 deletion may affect the transcriptional control of target gene expression. Regulatory elements of the target gene can extend to long distances outside of the transcription unit [29], which makes gene expression regulation a complex process. Interestingly, aberrant expression of the nearest neighbouring gene (1.0 Mb distance), *RGMB*, has been implicated in BC [30]. Additional analysis is needed to determine whether *RGMB* regulatory elements exist in the 5q15 deletion locus. Moreover, a previous copy number study of breast tumours-associated aberrations in the

5q15–5q21 locus with p53 status and patient survival suggests that the 5q15 region may be important in BC predisposition [31]. Furthermore, to reveal possible functional elements located in the deletion region, the Encyclopedia of DNA Elements (ENCODE) (http://genome.ucsc.edu/ENCODE/) was utilised. Preliminary analysis revealed enhancer and promotor-associated histone mark (H3K4Me1) activity and DNase hypersensitivity, which indicate that regulatory elements are active in this genomic region. Thus, the 5q15 homozygous deletion requires special attention because it may have clinical significance for screening families with BC with early disease onset. Interestingly, two heterozygous 5q15 loss carriers with lobular BC (Table 3, families 264 and 129) were previously found to carry BC-associated *CHEK2* variants [18].

The novel 8p23.2 deletion affects an intronic region in the *CSMD1* tumour suppressor gene. *CSMD1* has mainly been
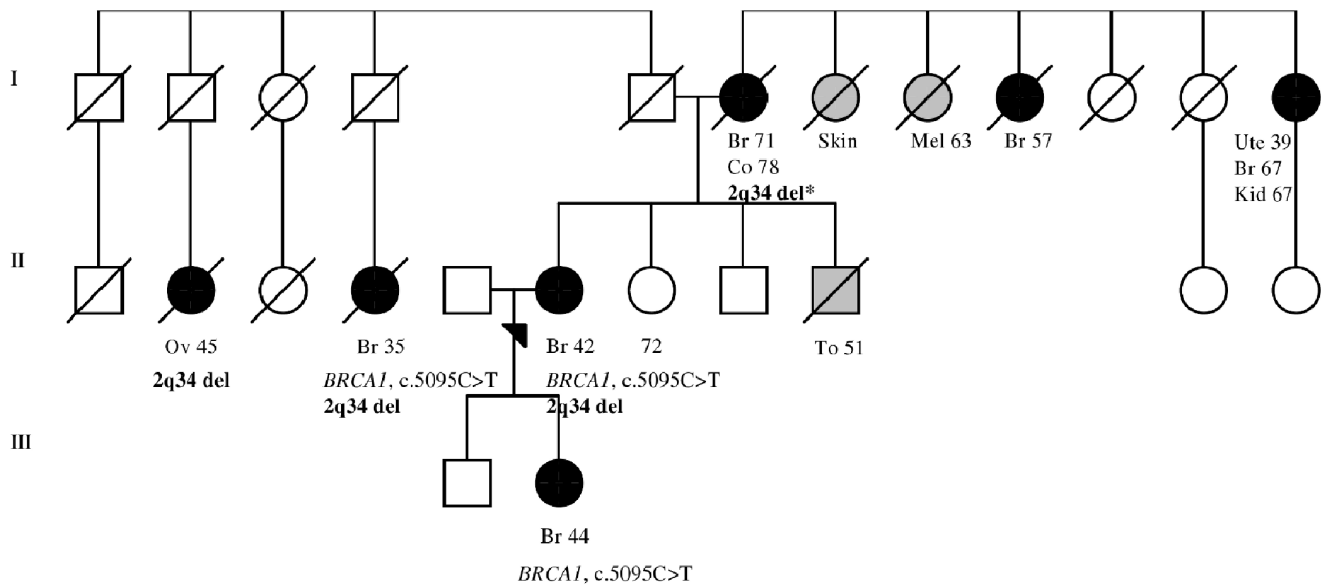
Figure 1. Family 249 pedigree. Index individual carries a novel 59.0 kb deletion in the 2q34 locus. The deletion affects intronic region of the *ERBB4* gene, which encodes a receptor tyrosine kinase family member that plays an important role in several cellular signalling pathways. The deletion was also identified in index's mother and two paternal cousins. Mother carried homozygous deletion (indicated with an asterisk). Index's daughter was tested to be negative for the deletion. Additionally, deleterious *BRCA1* c.5095C>T variant has been previously identified in three individuals in the family. Females are marked with circles and males are marked with squares. Index individual is marked with an arrow. Breast and ovarian cancers are marked with black circles with the age at diagnosis. Other cancers are marked with grey and specified with the age at diagnosis (Br: breast, Co: colon, Kid: kidney, Mel: melanoma, Ov: ovarian, To: tongue, Ute: uterus). Deceased individuals are marked with a slash. Current age of index's healthy sister is indicated. Generations are marked with the Roman numerals on the left. The pedigree figure has been modified from Kuusisto et al, 2011 [18].
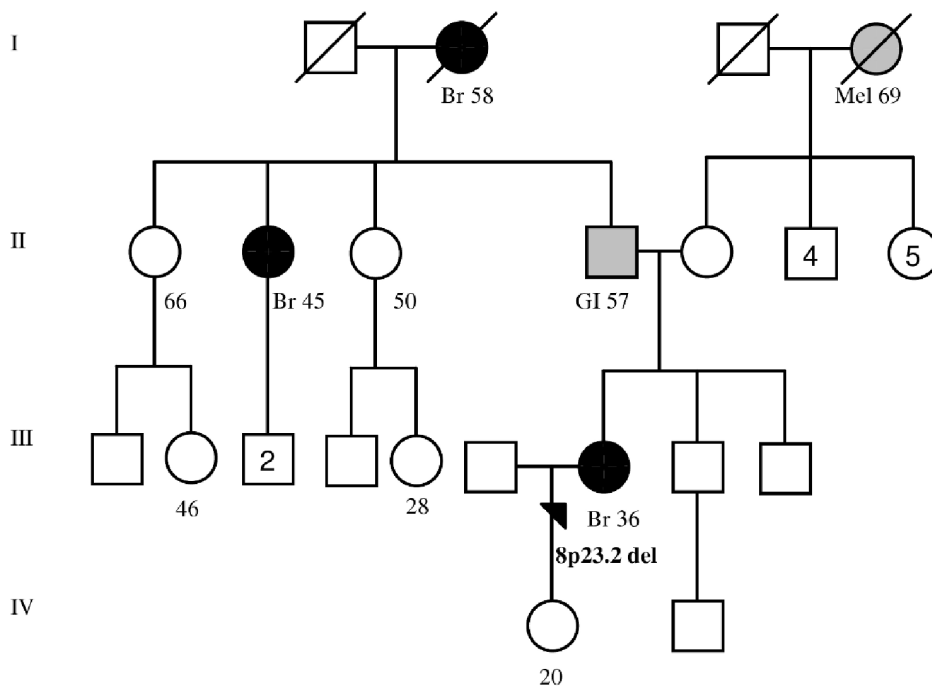doi:10.1371/journal.pone.0071802.g001

Figure 2. Family 128 pedigree. Index individual carries a novel 10.8 kb deletion in the 8p23.2. The deletion affects intronic region of the *CSMD1* tumor suppressor gene. Females are marked with circles and males are marked with squares. Number in circle or squares indicates descendants. Index individual is marked with an arrow. Breast cancers are marked with black circles with the age at diagnosis. Other cancers are marked with grey and specified with the age at diagnosis (Br: breast, GI: gastrointestinal, Mel: melanoma). Deceased individuals are marked with a slash. Current ages of healthy females are presented in the paternal side of the family. In addition, the current age of index's healthy daughter is indicated. Generations are marked with the Roman numerals on the left.
doi:10.1371/journal.pone.0071802.g002

associated with head and neck squamous cell carcinoma, but CSMD1 losses is also reported to contribute to the tumourigenesis of several other epithelial cancers, including BC [32]. In addition, *CSMD1* deletions and aberrant splicing have been shown to contribute to altered CSMD1 function in vivo [32]. Moreover, decreased CSMD1 expression has been associated with high tumour grade and the poor survival of invasive ductal breast carcinoma, and the role of CSMD1 expression as a potential BC prognostic marker has been suggested [33]. In this study, the *CSMD1*-affecting intronic deletion was identified in the index individual for one BC family (1 out of the 101, 1.0%) (family 128, Figure 2 and Table 3). In this family, the index patient and her paternal aunt and grandmother had been diagnosed with BC at ages 36, 45, and 58 years, respectively (Figure 2). In addition, gastrointestinal cancer was diagnosed on the paternal side of the family (father) (Figure 2). Interestingly, the *CSMD1*-affecting deletion was observed only in 1 out of the 436 (0.2%) healthy controls, suggesting that this rare variant likely predisposes individuals to BC. We are currently seeking DNA samples from the other family members (family 128) to determine whether the variation co-segregates with BC in the family. In addition, although the deletion should be screened for in larger sample set, the *CSMD1* gene is a potential candidate for the further study of HBOC susceptibility in Finnish families.

A novel deletion at 2q34 affects the intronic region of the *ERBB4* gene, which is known to play a role in BC [34]. *ERBB4* encodes an epidermal growth factor RTK subfamily member that regulates several cellular processes and plays an important role in cancer [35]. We found that the aberration in *ERBB4* is 1.5 times more common in HBOC individuals compared with controls suggesting that it may be a disease-related low-risk variant (Table 2). In addition, the clinical features of the *ERBB4* deletion carriers were interesting because two of the HBOC individuals had bilateral BC diagnosed at a relatively early age (Table 3). To further analyse the deletion, we were able to perform a segregation analysis in one family in which a deleterious *BRCA1* c.5095C>T variant was previously recognised (Figure 1) [18]. Thus, three BC cases in the family (index, index's daughter and paternal cousin) are explained by the paternally inherited high-penetrant *BRCA1* variant. The *ERBB4* deletion was observed on the maternal and paternal sides of the family (Figure 1). However, in the mother, who had BC diagnosed at an older age, the *ERBB4* deletion was homozygous, suggesting that the deletion could contribute to BC development at an older age, particularly in its homozygous form. Thus, it would be interesting to screen for the deletion in other BC cases diagnosed at an older age on the mother's side of the family as well. Additionally, an ovarian cancer patient who was negative for the highly -penetrant *BRCA1* variant was found to carry a heterozygous form of the 2q34 deletion, suggesting that the deletion may also contribute to ovarian cancer risk to some extent (Figure 1).

BRCA1 deletions are known to predispose to breast/ovarian cancer [36]. In this study, a large deletion overlapping exons 1A-13 of *BRCA1* was observed in one individual with BC diagnosed at age 46 years and with ovarian cancers diagnosed in her mother and half-sister (Table 3, family 252). In our previous analysis, the sample was excluded from the MLPA analysis due to a low sample quality value [18]. The *BRCA1* deletion encompassing exons 1A-13 has been reported in a Finnish breast/ovarian cancer family [37]. Here, the deletion was found to affect also the neighbouring genes *NBR1* (entire gene) and *NBR2* (exons 1–10) according to the PennCNV, QuantiSNP and cnvPartition programs. Similar findings have been reported worldwide in a few studies [38,39]. Because the *BRCA1* deletion is known to be clinically relevant,

MLPA analysis was performed to validate the *BRCA1* deletion (Figure S1). Genetic counselling was offered for the deletion carrier patient.

The duplication identified at 19q13.41 affects exon 1 of the *ERVV-2* gene. *ERVV-2* belongs to the human endogenous retrovirus (ERV) family and the involvement of ERVs in the pathogenesis of human cancer has been suggested but their roles in biological disease processes are poorly understood [40]. Because 19q13 genomic region has been previously associated with BC [41], this prompted us to further examine the duplication affecting the *ERVV-2* coding region. Screening for the duplication in additional controls revealed that it was as common in controls compared with HBOC individuals (Table 2). However, the homozygous form of the variation was 4.4 times more common in HBOC individuals compared with controls (Table 2), suggesting that the aberration may contribute to breast and ovarian cancer risk to some extent, but further studies are needed to confirm the findings. Of interest, one of the homozygous duplication carriers (Table 3, family 240) had been reported to carry a novel *BRCA2* variant predicted to be pathogenic [18].

In conclusion, this study is a continuation of our previous work with the aim of elucidating genetic factors contributing to HBOC susceptibility in Finland. We have identified several potential CNVs that likely increase the risk of HBOC susceptibility that may thus explain a fraction of breast and ovarian cancer cases. The aberrations at 3p11.1, 5q15, and 8p23.2 regions require special attention because they may be utilised for the genetic counselling of HBOC families, but more studies are needed to confirm the preliminary findings.

Supporting Information

Figure S1 *BRCA1* deletion (exons 1A-13) confirmation by MLPA.
(PDF)

Table S1 All of the identified 545 copy number variations (CNVs) at 273 different genomic regions (listed according to *P*-values).
(PDF)

Table S2 Enriched GO term categories and pathways (*P*-value less than 0.05) involving *EPHA3*, *ERBB4* and *BRCA1*.
(PDF)

File S1 Clinical characteristics of three additional individuals.
(PDF)

File S2 Clinical characteristics of 20 additional HBOC individuals utilised for CNV validation analysis.
(PDF)

File S3 Copy number variation validation protocol by quantitative RT-PCR.
(PDF)

Acknowledgments

Author Contributions

Conceived and designed the experiments: KMK MV SLL JS. Performed the experiments: KMK OA. Analyzed the data: KMK OA. Contributed reagents/materials/analysis tools: MKT MV SLL JS. Wrote the paper: KMK OA SLL JS. Revised the manuscript: MKT MV.

References

1. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, et al. (1994) A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science 266: 66–71.
2. Tavtigian SV, Simard J, Rommens J, Couch F, Shattuck-Eidens D, et al. (1996) The complete BRCA2 gene and mutations in chromosome 13q-linked kindreds. Nat Genet 12: 333–337.
3. Vehmanen P, Friedman LS, Eerola H, McClure M, Ward B, et al. (1997) Low proportion of BRCA1 and BRCA2 mutations in finnish breast cancer families: Evidence for additional susceptibility genes. Hum Mol Genet 6: 2309–2315.
4. Vahteristo P, Bartkova J, Eerola H, Syrjakoski K, Ojala S, et al. (2002) A CHEK2 genetic variant contributing to a substantial fraction of familial breast cancer. Am J Hum Genet 71: 432–438.
5. Erkko H, Xia B, Nikkila J, Schleutker J, Syrjakoski K, et al. (2007) A recurrent mutation in PALB2 in finnish cancer families. Nature 446: 316–319.
6. Pelttari LM, Heikkinen T, Thompson D, Kallioniemi A, Schleutker J, et al. (2011) RAD51C is a susceptibility gene for ovarian cancer. Hum Mol Genet 20: 3278–3288.
7. Solyom S, Aressy B, Pylkas K, Patterson-Fortin J, Hartikainen JM, et al. (2012) Breast cancer-associated abraxas mutation disrupts nuclear localization and DNA damage response functions. Sci Transl Med 4: 122ra23.
8. Stankiewicz P, Lupski JR (2010) Structural variation in the human genome and its role in disease. Annu Rev Med 61: 437–455.
9. Feuk L, Carson AR, Scherer SW (2006) Structural variation in the human genome. Nat Rev Genet 7: 85–97.
10. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, et al. (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 315: 848–853.
11. Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, et al. (2008) Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science 320: 539–543.
12. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, et al. (2010) Functional impact of global rare copy number variation in autism spectrum disorders. Nature 466: 368–372.
13. Diskin SJ, Hou C, Glessner JT, Attiyeh EF, Laudenslager M, et al. (2009) Copy number variation at 1q21.1 associated with neuroblastoma. Nature 459: 987–991.
14. Liu W, Sun J, Li G, Zhu Y, Zhang S, et al. (2009) Association of a germ-line copy number variation at 2p24.3 and risk for aggressive prostate cancer. Cancer Res 69: 2176–2179.
15. Venkatachalam R, Verwiel ET, Kamping EJ, Hoenselaar E, Gorgens H, et al. (2011) Identification of candidate predisposing copy number variants in familial and early-onset colorectal cancer patients. Int J Cancer 129: 1635–1642.
16. Krepischi AC, Achatz MI, Santos EM, Costa SS, Lisboa BC, et al. (2012) Germline DNA copy number variation in familial and early-onset breast cancer. Breast Cancer Res 14: R24.
17. Pylkas K, Vuorela M, Otsukka M, Kallioniemi A, Jukkola-Vuorinen A, et al. (2012) Rare copy number variants observed in hereditary breast cancer cases disrupt genes in estrogen signaling and TP53 tumor suppression network. PLoS Genet 8: e1002734.
18. Kuusisto KM, Bebel A, Vihinen M, Schleutker J, Sallinen SL (2011) Screening for BRCA1, BRCA2, CHEK2, PALB2, BRIP1, RAD50, and CDH1 mutations in high-risk finnish BRCA1/2-founder mutation-negative breast and/or ovarian cancer individuals. Breast Cancer Res 13: R20.
19. Wang K, Li M, Hadley D, Liu R, Glessner J, et al. (2007) PennCNV: An integrated hidden markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res 17: 1665–1674.
20. Colella S, Yau C, Taylor JM, Mirza G, Butler H, et al. (2007) QuantiSNP: An objective bayes hidden-markov model to detect and accurately map copy number variation using SNP genotyping data. Nucleic Acids Res 35: 2013–2025.
21. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559–575.
22. Meyer D, Zeileis A, Hornik K (2012) Vcd: Visualizing categorical data. R package version 1.2–13.
23. Zhang B, Kirov S, Snoddy J (2005) WebGestalt: An integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res 33: W741–8.
24. Pasquale EB (2008) Eph-ephrin bidirectional signaling in physiology and disease. Cell 133: 38–52.
25. Pasquale EB (2010) Eph receptors and ephrins in cancer: Bidirectional signalling and beyond. Nat Rev Cancer 10: 165–180.
26. Xi HQ, Wu XS, Wei B, Chen L (2012) Aberrant expression of EphA3 in gastric carcinoma: Correlation with tumor angiogenesis and survival. J Gastroenterol.
27. Xi HQ, Zhao P (2011) Clinicopathological significance and prognostic value of EphA3 and CD133 expression in colorectal carcinoma. J Clin Pathol 64: 498–503.
28. Guan M, Liu L, Zhao X, Wu Q, Yu B, et al. (2011) Copy number variations of EphA3 are associated with multiple types of hematologic malignancies. Clin Lymphoma Myeloma Leuk 11: 50–53.
29. Kleinjan DA, van Heyningen V (2005) Long-range control of gene expression: Emerging mechanisms and disruption in disease. Am J Hum Genet 76: 8–32.
30. Li J, Ye L, Mansel RE, Jiang WG (2011) Potential prognostic value of repulsive guidance molecules in breast cancer. Anticancer Res 31: 1703–1711.
31. Jain AN, Chin K, Borresen-Dale AL, Erikstein BK, Eynstein Lonning P, et al. (2001) Quantitative analysis of chromosomal CGH in human breast tumors associates copy number abnormalities with p53 status and patient survival. Proc Natl Acad Sci U S A 98: 7952–7957.
32. Ma C, Quesnelle KM, Sparano A, Rao S, Park MS, et al. (2009) Characterization CSMD1 in a large set of primary lung, head and neck, breast and skin cancer tissues. Cancer Biol Ther 8: 907–916.
33. Kamal M, Shaaban AM, Zhang L, Walker C, Gray S, et al. (2010) Loss of CSMD1 expression is associated with high tumour grade and poor survival in invasive ductal breast carcinoma. Breast Cancer Res Treat 121: 555–563.
34. Sundvall M, Iljin K, Kilpinen S, Sara H, Kallioniemi OP, et al. (2008) Role of ErbB4 in breast cancer. J Mammary Gland Biol Neoplasia 13: 259–268.
35. Hynes NE, Lane HA (2005) ERBB receptors and cancer: The complexity of targeted inhibitors. Nat Rev Cancer 5: 341–354.
36. Woodward AM, Davis TA, Silva AG, Kirk JA, Leary JA, et al. (2005) Large genomic rearrangements of both BRCA2 and BRCA1 are a feature of the inherited breast/ovarian cancer phenotype in selected families. J Med Genet 42: e31.
37. Pylkas K, Erkko H, Nikkila J, Solyom S, Winqvist R (2008) Analysis of large deletions in BRCA1, BRCA2 and PALB2 genes in finnish breast and ovarian cancer families. BMC Cancer 8: 146.
38. Garcia-Casado Z, Romero I, Fernandez-Serra A, Rubio L, Llopis F, et al. (2011) A de novo complete BRCA1 gene deletion identified in a spanish woman with early bilateral breast cancer. BMC Med Genet 12: 134.
39. Gad S, Bieche I, Barrois M, Casilli F, Pages-Berhouet S, et al. (2003) Characterisation of a 161 kb deletion extending from the NBR1 to the BRCA1 genes in a french breast-ovarian cancer family. Hum Mutat 21: 654.
40. Ruprecht K, Mayer J, Sauter M, Roemer K, Mueller-Lantzsch N (2008) Endogenous retroviruses and cancer. Cell Mol Life Sci 65: 3366–3382.
41. Nexo BA, Vogel U, Olsen A, Nyegaard M, Bukowy Z, et al. (2008) Linkage disequilibrium mapping of a breast cancer susceptibility locus near RAI/PPP1R13L/iASPP. BMC Med Genet 9: 56-2350-9-56.