RĪGA STRADIŅŠ UNIVERSITY

Egija Berga-Švītiņa

Identifying Genetic Factors Associated with Breast or Ovarian Cancer Risk in *BRCA1* Pathogenic Variant Carriers

Doctoral Thesis for obtaining the scientific degree
"Doctor of Science (*PhD*)"

Sector Group – Medical and Health Sciences
Sector – Basic Medicine
Sub-Sector – Medicinal Genetics

Riga, 2024

Egija Berga-Švītiņa

ORCID 0000-0001-5150-0185

Identifying Genetic Factors Associated with Breast or Ovarian Cancer Risk in *BRCA1* Pathogenic Variant Carriers

Doctoral Thesis for obtaining the scientific degree
"Doctor of Science (*PhD)"*

Sector Group – Medical and Health Sciences

Sector – Basic Medicine

Sub-Sector – Medicinal Genetics

Supervisors of the Doctoral Thesis:

Dr. biol., Professor **Edvīns Miklaševičs,**
Institute of Oncology and Molecular Genetics,
Department of Biology and Microbiology,
Rīga Stradiņš University, Latvia

Dr. rer. nat., **Baiba Vilne,**
Bioinformatics Group at the Scientific centre,
Rīga Stradiņš University, Latvia

Scientific Advisor:

Dr. med., **Jeļena Maksimenko,**
Institute of Oncology and Molecular Genetics,
Rīga Stradiņš University;
Breast Surgery Unit, Department of Surgery,
Pauls Stradiņš Clinical University Hospital, Latvia

Riga, 2024

Abstract

Breast cancer (BC) is a most prevalent cancer among women globally and ovarian cancer (OC) is also a significant healthcare burden, ranking eighth in terms of incidence and mortality in females. The aetiology of these malignancies involves a complex interplay between modifiable and non-modifiable risk factors. Among these, genetic predisposition, particularly pathogenic variants (PVs) in *BRCA1* gene, significantly elevate a risk of BC or OC development. However, BC and OC risk for germline *BRCA1* PV carriers differ by individual and are affected by genetic factors. The aim of this study is to explore genetic factors that might modulate BC and OC risk and to assess the effect of polygenic risk score (PRS) to estimate the overall genetic risk of a women carrying region-specific germline *BRCA1* PVs to develop BC or OC due to additional genetic variations.

We performed a genome-wide association study (GWAS) in 406 female *BRCA1* PV (c.4035del or c.5266dup) carriers, affected with BC or OC vs. unaffected individuals, followed by functional annotations of the most significantly associated single nucleotide variants (SNVs). Next, we investigated recently developed novel genome-wise PRS association with BC and OC risk in *BRCA1* PV carriers. A binomial logistic regression model was applied to assess the association of PRS with BC or OC development risk.

In BC patients, the most significantly associated SNV was rs2609813 ($p = 2.33 \times 10^{-7}$, odds ratio (OR) = 0.28) in *FAM107B* gene (genomic position (GRCh37) 10:14800320). The variant is intronic in the protein coding gene and predicted to be a regulatory region variant. The second most significant BC-associated SNV was rs4688094 ($p = 7.76 \times 10^{-7}$, OR = 0.38) in long non-coding RNA (lncRNA) gene (genomic position (GRCh37) 3:118003477) and the most significant OC-associated SNV was rs79732499 ($p = 1.38 \times 10^{-7}$, OR = 0.00031) located in genomic position (GRCh37) 20:3404208 and is predicted to be a regulatory region variant located in enhancer. Both variants are in the non-coding genome. This suggests that they may influence gene expression or other regulatory processes rather than directly altering protein structure or function. Due to the small sample size, our results did not reach a genome-wide significance of $p = 5 \times 10^{-8}$. Regarding PRS calculations, best-fitting BayesW PRS model could effectively predict the individual's BC risk (OR = 1.37; 95 % confidence interval (CI) = 1.03–1.81, $p = 0.029$ with area under receiver-operator curve (AUC) = 0.76). At the same time, none of the applied PRS was a good predictor of OC development risk, suggesting the need for further investigation in larger OC cohort.

The results of this study can be used as preliminary data for a more comprehensive study and might contribute to customised PRS development for *BRCA1* PV carriers. Previously developed BayesW PRS model contributed to assessing the risk of developing BC for germline

BRCA1 PV (c.4035del or c.5266dup) carriers and may facilitate more precise and timelier patient stratification and decision-making to improve the current BC treatment or even prevention strategies.

Anotācija

Ģenētisko faktoru, kas saistīti ar krūts vai olnīcu vēža risku, identificēšana *BRCA1* patogēno variantu nesējās

Krūts vēzis (KV) ir visizplatītākais vēzis sieviešu vidū visā pasaulē, kā arī olnīcu vēzis (OV) ir nozīmīgs veselības aprūpes slogs, ieņemot astoto vietu incidences un mirstības rādītājos. Šo ļaundabīgo audzēju etioloģija iekļauj kompleksu mijiedarbību starp modificējamiem un nemodificējamiem riska faktoriem. Viens no šādiem riska faktoriem ir ģenētiskā predispozīcija, tai skaitā, *BRCA1* gēna patogēnie varianti (PV), kas ievērojami palielina KV vai OV attīstības risku. Tomēr risks *BRCA1* PV nesējās ir atšķirīgs, jo to ietekmē citi ģenētiskie faktori. Šīs disertācijas mērķis ir izpētīt ģenētiskos faktorus, kas ietekmē KV un OV attīstības risku sievietēs ar reģionam specifiskiem pārmantojamiem *BRCA1* PV, kā arī izvērtēt poligēnā riska modeļa (angl. *PRS*) ietekmi uz individualizētu kopējo ģenētisko risku.

Šajā disertācijā tika veikta genoma mēroga asociāciju analīze (angl. *GWAS*) 406 sievietēs ar pārmantojamu *BRCA1* PV (c.4035del un c.5266dup) un KV vai OV salīdzinājumā ar sievietēm ar pārmantojamu *BRCA1* PV bez audzēja diagnozes, kam sekoja statistiski nozīmīgi asociēto viena nukleotīda variantu (angl. *SNV*) funkcionālā anotācija. Tālāk tika pētīta nesen izveidoto genoma-mēroga PRS asociācija ar KV vai OV attīstības risku *BRCA1* PV nesējās, kas tika pārbaudīta ar binomiālās loģistiskās regresijas modeli.

KV pacientēs statistiski nozīmīgāk saistītais SNV bija rs2609813 ($p = 2{,}33 \times 10^{-7}$, izredžu attiecība (angl. *OR*) = 0,28), kas ir intronisks variants proteīnu kodējošā *FAM107B* gēnā (genomiskajā pozīcijā (GRCh37) 10:14800320) un tiek prognozēts kā regulējošā reģiona variants. Otrs statistiski nozīmīgākais ar KV saistītais SNV bija rs4688094 ($p = 7{,}76 \times 10^{-7}$, OR = 0,38), kas atrodas garās nekodējošās RNS (angl. *lncRNA*) gēnā (genomiskajā pozīcijā (GRCh37) 3:118003477) un nozīmīgākais ar OV saistītais SNV bija rs79732499 ($p = 1{,}38 \times 10^{-7}$, OR = 0,00031), kas atrodas genomiskajā pozīcijā (GRCh37) 20:3404208 un tiek prognozēts kā regulējošā reģiona variants enhanserī (angl. *enhancer*). Abi minētie varianti atrodas genoma nekodējošā daļā. Rezultāti liecina, ka atklātie varianti visticamāk ietekmē gēnu ekspresiju vai citus regulatorus procesus, nevis specifiski proteīna struktūru vai funkciju. Nelielās kohortas izmēra dēļ mūsu rezultāti nesasniedza genoma mēroga statistisko nozīmīgumu $p = 5 \times 10^{-8}$. Savukārt PRS aprēķinos atbilstošākais modelis bija BayesW PRS, ar kuru varēja efektīvi paredzēt indivīda KV risku (OR = 1.37; 95 % ticamības intervāls (angl. *CI*) = 1,03–1,81, p = 0,029 ar laukumu zem uztvērēja operatora līknes (angl. *AUC*) = 0,76). Vienlaicīgi neviens no izmantotajiem PRS nebija labs OV attīstības riska prognozētājs, kas liecina par nepieciešamību veikt padziļinātus pētījumus lielākā OV kohortā.

Šī pētījuma rezultātus ir iespējams izmantot kā preliminārus datus plašākiem pētījumiem, un tie varētu veicināt individualizētu PRS izstrādi un pielietošanu sievietēs ar pārmantojamu *BRCA1* PV. Iepriekš izstrādātais BayesW PRS ir efektīvs un palīdz novērtēt KV attīstības risku *BRCA1* PV (c.4035del vai c.5266dup) nesējās. Šis modelis var veicināt precīzāku un savlaicīgāku pacienšu riska stratifikāciju un palīdzēt lēmumu pieņemšanā par KV ārstēšanas vai profilakses stratēģiju.

Atslēgvārdi: poligēnā riska modelis, krūts vēzis, olnīcu vēzis, *BRCA1* patogēno variantu nesējas.

Table of Contents

Abbreviations used in the Thesis

ABL1	ABL proto-oncogene 1, non-receptor tyrosine kinase
ACMG	American College of Medical Genetics and Genomics
AD	autosomal dominant
AI	artificial intelligence
AIM	ancestry informative marker
AKT pathway	AKT serine/threonine kinase signalling pathway
ALT	alternative allele
ARID1A	AT-rich interaction domain 1A
ARL3	ADP ribosylation factor like GTPase 3
ATM	ATM serine/threonine kinase
AUC	area under the curve
BARD1	BRCA1 associated RING domain 1
BASC	BRCA1-associated genome surveillance complex
BC	breast cancer
BCAC	Breast Cancer Association Consortium
BIC	Breast Cancer Information Core
BL	basal-like
BMI	body mass index
BRAF	B-Raf proto-oncogene, serine/threonine kinase
BRCA1	BRCA1 DNA repair associated
BRCA2	BRCA2 DNA repair associated
BRIP1	BRCA1 interacting helicase 1
BRRM	bilateral risk-reducing mastectomy
BRRSO	bilateral risk-reducing salpingo-oophorectomy
CA-125	cancer antigen 125
CCND1	cyclin D1
CCOC	clear cell ovarian carcinoma
CDH1	cadherin 1
CDK	cyclin-dependent kinase
CDKN2A	cyclin dependent kinase inhibitor 2A
CDKN2B	cyclin dependent kinase inhibitor 2B
CDPC	Centre for Disease Prevention and Control
CHD	coronary heart disease

CHEK2	checkpoint kinase 2
CI	confidence interval
CIMBA	Consortium of Investigators of Modifiers of *BRCA1/2*
CN-LOH	copy neutral loss of heterozygosity
CTNNB1	catenin beta 1
DDR	DNA damage response
DNA	deoxyribonucleic acid
DNAAF9	dynein axonemal assembly factor 9
DR2	dosage R-squared
DSB	double-stranded break
EGFR	epidermal growth factor receptor
EMT	epithelial-mesenchymal transition
ENIGMA	evidence-based network investigating germline mutant alleles
EOC	epithelial ovarian cancer
EPCAM	epithelial cell adhesion molecule
eQTL	expression quantitative trait loci
ER	estrogen receptor
ERBB2	erb-b2 receptor tyrosine kinase 2
ESR1	estrogen receptor 1
EstBB	Estonian biobank
FAM107B	family with sequence similarity 107 member B
FDR	false discovery rate
FGFR1	fibroblast growth factor receptor 1
FGFR2	fibroblast growth factor receptor 2
FIGO	Federation of Gynaecology and Obstetrics
FUMA	functional mapping and annotation
GATA3	GATA binding protein 3
GMRM	grouped mixture of regressions model
GoF	gain-of-function
GRCh37	genome reference consortium human build 37
GTEx	genotype-tissue expression project
GWAS	genome-wide association study
HBOC	hereditary breast and ovarian cancer
HER2	human epidermal growth factor receptor 2
HNF-1β	HNF1 homeobox B protein

HR	hazard ratio
HRD	homologous recombination deficiency
HRR	homologous recombination DNA repair
HRT	hormone replacement therapy
HWE	Hardy-Weinberg equilibrium
IBD	identity by descent
IDC	invasive ductal carcinoma
IHC	immunohistochemistry
ILC	invasive lobular carcinoma
KRAS	KRAS proto-oncogene, GTPase
LD	linkage disequilibrium
LGDB	genome database of Latvian population
LLM	large language models
lncRNA	long non-coding RNA
LoF	loss-of-function
MAF	minor allele frequency
MLH1	MutL homolog 1
MRI	magnetic resonance imaging
mRNA	messenger RNA
MSH2	MutS homolog 2
MSH6	MutS homolog 6
mTOR pathway	mechanistic target of rapamycin signalling kinase pathway
MYC	MYC proto-oncogene, bHLH transcription factor
NCCN	National Comprehensive Cancer Network
NES	normalised effect size
NET	neuroendocrine tumour
NGS	next-generation sequencing
NMD	nonsense-mediated mRNA decay
OC	ovarian cancer
OCAC	Ovarian Cancer Association Consortium
OR	odds ratio
PALB2	partner and localiser of BRCA2
PARP	poly (ADP-ribose) polymerase
PC	principal component
PCR	polymerase chain reaction

PI3K pathway	phosphoinositide 3-kinase signalling pathway
PIK3CA	phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha
PMS2	PMS1 homolog 2, mismatch repair system component
PR	progesterone receptor
PRS	polygenic risk score
PTEN	phosphatase and tensin homolog
PV	pathogenic variant
QC	quality control
RAD51C	RAD51 paralog C
RAD51D	RAD51 paralog D
REF	reference allele
RNA	ribonucleic acid
ROC	receiver-operating characteristic
ROS	reactive oxygen species
SD	standard deviation
SE	standard error
SNV	single nucleotide variant
STK11	serine/threonine kinase 11
TIME	tumour immune microenvironment
TNBC	triple-negative breast cancer
TNM	tumour, node, metastasis
TP53	tumour protein p53
T2D	type 2 diabetes
VEP	variant effect predictor
VUS	variant of uncertain significance
WES	whole exome sequencing
WGS	whole genome sequencing
WHO	World Health Organization
ZNF514	zinc finger protein 514

Introduction

According to GLOBOCAN 2020, BC is the most frequently diagnosed cancer in females, contributing to 15 % of cancer-related deaths worldwide, with approximately 522,000 reported deaths. Additionally, OC ranks eighth in terms of incidence and mortality among females, contributing to 5 % of cancer-related deaths (Sung et al., 2021). In Latvia, BC and OC creates a significant healthcare burden, accounting for approximately 1200 new BC diagnoses and 300 OC diagnosis annually (CDPC, 2020).

Approximately 5–10 % of all BC cases and 10–15 % of all OC cases are estimated to be hereditary, being associated with germline PVs in a cancer predisposition gene, particularly *BRCA1* and *BRCA2* (Angeli et al., 2020; Leitsalu et al., 2021). Germline PVs in *BRCA1* gene are recognised as the most penetrant genetic predisposition for both BC and OC. The associated lifetime risks for cancer development have been estimated to range from 60 % to 75 % for BC and 34 % to 44 % for OC by the age of 80 in female carriers of germline *BRCA1* PVs (Barnes et al., 2020; Borde et al., 2022; Rebbeck et al., 2015). These data suggest an incomplete penetrance, where a subset of *BRCA1* PV carriers never develops BC or OC in their lifetime, presenting challenges in genetic counselling and risk assessment due to variability in penetrance among carriers. Penetrance refers to the likelihood of an individual carrying specific genetic PVs to develop particular trait or disease, in this case BC or OC. Subsequently, other genetic factors are suggested to contribute to this phenomenon (Chen et al., 2020; Downs et al., 2019; Narod, 2002).

Currently, the assessment of individuals' risk of developing BC or OC is based on personal history or the presence of first-degree relatives with specific cancer diagnosis, along with an age-related criteria, followed by screening to identify germline *BRCA1* PVs with founder effect (Jürgens et al., 2022). However, given the incomplete penetrance of *BRCA1* PVs, the assessment should also include other penetrance-modifying factors. As the preventive procedures are invasive and can have severe psychological and physiological effects, precise age-dependent estimations of cancer risk in *BRCA1* PV carriers are critical in genetic counselling. Enhanced risk prediction can help to identify high-risk women who may benefit from early clinical intervention and low-risk women who may decide to postpone prophylactic procedures or chemoprevention (Borde et al., 2022; Kuchenbaecker, McGuffog, et al., 2017).

Therefore, the aim of this Thesis was to contribute to research on potential genetic modifiers of BC or OC risk in *BRCA1* PV (c.4035del and c.5266dup) carriers using GWAS approach. Additionally, we explored and compared the efficiency of two PRS models (BayesW vs. BayesRR-RC) to estimate the overall genetic risk in women carrying these two most frequently identified germline *BRCA1* PVs that are region-specific for the Latvian population.

The goal of this Thesis was to evaluate the risk of developing BC or OC due to additional genetic variations.

Aim of the Thesis

To identify genetic factors that might influence the penetrance of two most prevalent *BRCA1* pathogenic variants (c.4035del and c.5266dup) within the study cohort.

Objectives of the Thesis

To achieve the overall aim of the Thesis, the following objectives have been set:

1. Assess the effect of three pathogenic/likely pathogenic variants of the *CHEK2* gene on the penetrance of *BRCA1* pathogenic variants (c.4035del and c.5266dup) in the study cohort;

2. Conduct a GWAS in breast cancer patients to identify additional genetic variants affecting the penetrance of *BRCA1* pathogenic variants (c.4035del and c.5266dup) in the study cohort;

3. Conduct a GWAS in ovarian cancer patients to identify additional genetic variants affecting the penetrance of *BRCA1* pathogenic variants (c.4035del and c.5266dup) in the study cohort;

4. Evaluate the association between novel genome-wise PRSs and the risk of breast and ovarian cancer in *BRCA1* pathogenic variant (c.4035del and c.5266dup) carriers within the study cohort.

Hypothesis of the Thesis

The penetrance of region-specific *BRCA1* pathogenic variants (c.4035del and c.5266dup) in the study cohort is affected by other genetic variants.

Novelty of the Thesis

Until now, there has been a lack of research into the genetic factors contributing to the incomplete penetrance of specific *BRCA1* PVs. This study represents the first comprehensive investigation of genetic modifiers of region-specific *BRCA1* PVs conducted on a cohort of the Latvian population. This cohort was specifically selected based on two founder variants in *BRCA1* gene (c.4035del and c.5266dup), resulting in genetically homogeneous cohort.

Author's contribution

The author has performed and participated in all stages of the study, including the study design, DNA processing, molecular analysis, including genotyping, as well as the bioinformatic and statistical data analysis. The author has prepared scientific publications and wrote this Thesis.

AI Disclosure / LLM statement

As the author is not a native English speaker Large Language Models (LLM models) and Artificial Intelligence (AI) assisted editing tools were used to correct grammatical and spelling errors and improve the writing style and readability throughout this PhD Thesis. These models/tools were not used to replace any key authoring tasks such as producing scientific and/or medical insights or drawing conclusions. The author has carefully reviewed and edited the content as needed and takes full responsibility for the validity and integrity of this work.

1 Literature review

1.1 Carcinogenesis

Carcinogenesis is the process by which normal cells transform into malignant cancer cells, which is an extremely complicated and multistep process influenced by various endogenous pathways, such as reactive oxygen species (ROS) or disturbance in the DNA repair pathway, as well as exogenous environmental factors (Peters & Gonzalez, 2018). A well-established hallmark of cancer involves dynamic changes in the cancer cell genome. Carcinogenesis has been characterised by six essential alterations in cell physiology, including self-sufficiency in growth signals, insensitivity to growth-inhibitory signals, avoidance of programmed cell death (also known as apoptosis), limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis (Hanahan & Weinberg, 2000).

To develop effective strategies for cancer prevention and novel targeted therapeutic treatment, it is essential to investigate the molecular mechanisms driving carcinogenesis. This involves exploring the roles and functions of proto-oncogenes, tumour suppressor genes, and cell cycle regulators to understand the mechanisms involved in cancer initiation and progression (Hanahan & Weinberg, 2000; Weinberg, 1994).

1.1.1 Proto-oncogenes and tumour suppressor genes

Proto-oncogenes and tumour suppressor genes are important in regulating cell growth and apoptosis, thereby preventing cancer development. Proto-oncogenes, that are frequently subjected to dominant gain-of-function (GoF) mutations, can undergo a transformation into oncogenes and promote uncontrolled cell growth, division, and survival. In contrast, tumour suppressor genes act as "guardians" of the genome by preventing uncontrolled cellular proliferation, promoting DNA repair, normal cell differentiation and activating cell cycle checkpoints. They are often inactivated by loss-of-function (LoF) mutations, most often requiring alteration in both alleles for tumour progression. Maintaining a delicate balance between proto-oncogenes and tumour suppressor genes is essential for cellular homeostasis (Couch, 1996; Lee & Muller, 2010; Pitot, 1993). The progression of BC and OC, that is the central theme of this Thesis, involves acquired genetic alterations, including the activation of oncogenes and impairment of specific tumour suppressor genes such as *BRCA1* or *TP53* (Lee & Muller, 2010; Polyak, 2007).

Proto-oncogenes

Oncogenic proteins promote cell proliferation, inhibit apoptosis, or prevent differentiation of cancer cells. When dominant GoF mutations are acquired, these proteins

undergo a functional shift, transitioning from their normal role in maintaining cellular homeostasis to promoting carcinogenic signalling (Couch, 1996; Hanahan & Weinberg, 2000).

Various classes of oncogenes have been identified, each with specific effect on different cellular processes. These include receptor tyrosine kinases (e.g. epidermal growth factor receptor (*EGFR*) gene, erb-b2 receptor tyrosine kinase 2 (*ERBB2*) gene) controlling growth and differentiation; cytoplasmic tyrosine kinases (e.g. ABL proto-oncogene 1, non-receptor tyrosine kinase (*ABL1*) gene) regulating cell proliferation, differentiation, migration, and survival; cytoplasmic serine/threonine kinases (e.g. B-Raf proto-oncogene, serine/threonine kinase (*BRAF*) gene) regulating cell cycle control, proliferation, differentiation, apoptosis, and survival; membrane-linked GTPases (e.g. KRAS proto-oncogene, GTPase (*KRAS*) gene) regulating cell proliferation; and transcription factors (e.g. MYC proto-oncogene, bHLH transcription factor (*MYC*) gene) indirectly modulating cell proliferation. Proto-oncogenes can undergo activation through various mechanisms, including chromosomal translocation, point mutation, and gene amplification. Over the past two to three decades, significant progress has been made in developing targeted therapies for specific molecular pathways activated by oncogenes, including successful use of tyrosine kinase inhibitors (TKI) (Hartmann et al., 2009; Kontomanolis et al., 2020; Patterson et al., 2018).

Tumour suppressor genes

Tumour suppressor genes play a critical role in regulating normal cell growth and differentiation, while actively inhibiting the development of cancer. Carcinogenesis requires the inactivation of both copies of a tumour suppressor gene, often involving a recessive LoF mutation in one allele, combined with the loss of the second allele through deletions or copy neutral loss of heterozygosity (CN-LOH) (Kontomanolis et al., 2020).

The important member of tumour suppressor genes is the *TP53* gene. In response to DNA damage, p53 induces either cell cycle arrest, allowing DNA repair, or apoptosis in the case of excessive damage. It is evident that the functionality of the p53 DNA damage signalling pathway is lost in the majority of human cancers, contributing to genome instability and variability, and leading to the generation of aggressive mutant cells with possible selective advantages (Hanahan & Weinberg, 2000; Harris, 1996).

Additionally, the *BRCA1* gene stands out among crucial tumour suppressor genes. *BRCA1* plays an essential role in repairing DNA damage, preserving proper cell cycle regulation, and maintaining genomic stability. Impairment of *BRCA1* function, often caused by deleterious germline PVs, initiates a cascade of events leading to increased susceptibility to

BC or OC (Fu et al., 2022; Lee & Muller, 2010). The role of *BRCA1* gene in carcinogenesis will be discussed in detail in following chapters.

1.1.2 The cell cycle regulators

Cell cycle checkpoints function as surveillance mechanisms that monitor the order, integrity, and fidelity of crucial events within the cell cycle. These events include ensuring cell growth, replication, and accurate segregation of chromosomes (Barnum & O'Connell, 2014).

The primary drivers regulating the progression of cell cycle are the cyclin-dependent kinases (CDKs). These serine/threonine protein kinases play an essential role in phosphorylating key substrates to promote DNA synthesis and facilitate mitotic progression (Barnum & O'Connell, 2014; Hanahan & Weinberg, 2000). Disruptions in the cell cycle can lead to uncontrolled cell division and contribute to cancer development.

In addition to CDKs, other critical players in maintaining the integrity of the genome include cell cycle checkpoint proteins like *CHEK2*. *CHEK2* has an important role in the DNA repair pathway in response to double-strand breaks. It acts by inhibiting the cell cycle and promoting the activation of DNA repair by phosphorylating crucial proteins, including p53 and BRCA1 (Aksoy et al., 2022; Cai et al., 2009). Dysfunction in cell cycle checkpoints can lead to diverse consequences, such as chromosomal rearrangements (e.g. deletions, amplifications, or translocations), potentially promoting cancer development (Hartwell & Kastan, 1994).

1.2 General overview of BC and OC

Next chapter provides a thorough exploration and general overview of BC and OC, two significant malignancies with substantial impact on women's health. The chapter will discuss the epidemiology, classification, risk factors, origins, and genetic component of these cancers, it aims to provide a comprehensive understanding of their complexities and impact.

1.2.1 Incidence, impact, and current status

In 2020, BC surpassed lung cancer as the most commonly diagnosed cancer, with an estimated 2.3 million new cases (11.70 %) (Yang et al., 2022). According to GLOBOCAN 2020, BC is the most prevalent neoplasm in females, contributing to 15 % of cancer-related deaths, with roughly 522,000 deaths worldwide. Furthermore, OC ranks eighth in terms of incidence and mortality among females (Sung et al., 2021). In Latvia, approximately 1200 women are diagnosed with BC, while OC affects around 300 women annually (CDPC, 2020).

The global cancer burden in women is rising, which is a trend observed across different countries regardless of income level. This increase is attributed to population growth, increasing family history, lifestyle factors and an aging demographic. However, in high-income countries,

BC is frequently diagnosed at an early stage, creating better prognosis and outcomes. The availability of advanced healthcare infrastructure and widespread screening practices contribute to the early detection and effective management of the disease (Harbeck et al., 2019).

However, despite significant progress in prevention, early detection, and personalised treatment, BC and OC continue to be a major cause of cancer-related deaths, primarily due to belated diagnosis, recurrence, distant metastasis, and resistance to chemotherapy (Liu et al., 2021).

1.2.2 Classification of BC

In the upcoming chapters, we will explore in more detail the classification of BC and OC. BC is primarily diagnosed based on histological testing that is the foundation of histological classification system. This classification not only confirms the diagnosis of malignancy but also characterises the tumour, providing implications for treatment options and other key prognostic features. The most common histological subtypes of BC include invasive ductal carcinoma (IDC), invasive lobular carcinoma (ILC), and many others, each exhibiting distinct characteristics and behaviours (Makki, 2015; Rakha et al., 2023).

The histological classification of BC is incorporated in various staging and grading systems, including the TNM (Tumour, Node, Metastasis) staging system and the Nottingham grading system. These systems integrate parameters such as histological grade, tumour size, lymph node involvement, distant metastasis, and other relevant factors to provide prognostic information and guide treatment decisions (Oluogun et al., 2019; Rakha et al., 2023).

In addition, the evaluation of hormone receptor status, including estrogen receptor (ER) and progesterone receptor (PR) status, as well as human epidermal growth factor receptor 2 (HER2) status by immunohistochemistry (IHC), is a standard procedure for characterizing BC cells. This information significantly influences treatment strategies, such as hormone therapy (Rakha et al., 2023). Furthermore, biomarkers are the key factors in classifying the BC into different molecular subtypes: luminal A and B, HER2-enriched, and triple-negative (TNBC) or basal-like (BL) BC, as mentioned in the Table 1.1 (Zubair et al., 2020). Moreover, the increasing availability of molecular profiling techniques, such as microarrays and next-generation sequencing (NGS), continue to facilitate a better understanding of BC molecular subtypes and the progress towards precision and personalised medicine (Rakha et al., 2023).

Table 1.1

Molecular classification of BC

Molecular subtype	Biomarker status	Description based on (Eliyatkın et al., 2015)
Luminal A	ER-positive and/or PR-positive, HER2-negative	Most prevalent molecular subtype of invasive BC, characterized by hormone receptor positivity and lower proliferative activity, and associated with better prognosis.
Luminal B	ER-positive and/or PR-positive, HER2-positive	Exhibits hormone receptor positivity, increased proliferation, and higher histologic grade than Luminal A; may have a more aggressive clinical course compared to Luminal A.
HER2-enriched	ER-negative, PR-negative, HER2-positive	Characterized by overexpression of HER2, high proliferation, *TP53* mutations, high histologic grade, and nodal positivity; tends to be more aggressive with unfavourable prognosis, but targeted therapies like Herceptin can be effective.
Triple negative (basal-like)	ER-negative, PR-negative, HER2-negative	Lacks hormone receptor and HER2 expression, high proliferation, *TP53* mutations, and *BRCA1* PVs (germline, sporadic); often more aggressive, but sensitive to platinum group chemotherapy and PARP inhibitors.

PARP – Poly (ADP-ribose) polymerase.

1.2.3 Classification of OC

Similar to BC, OC exhibits diverse histological subtypes, resulting in a complex classification with distinct characteristics and behaviours that provides valuable insights into diagnosis, prognosis, and treatment strategies (Hayashi & Konishi, 2023). The World Health Organization (WHO) has categorised OC into various types from which the most prevalent type is epithelial OC (EOC), accounting for approximately 90 % of cases (Zamwar & Anjankar, 2022). The classification of EOC mostly relies on both histological and molecular characteristics but it is a highly heterogeneous phenotype (Dareng et al., 2022).

Additionally, the clinical management and prognosis of OC depend on the stage and grade of the cancer. Cancer stage indicates the extent to which it has spread from the original site. OC staging follows the guidelines outlined by FIGO (International Federation of Gynaecology and Obstetrics) system, which includes four stages: Stage I (cancer is limited to one or both ovaries); Stage II (cancer has spread to other pelvic organs but is still within the pelvis); Stage III (cancer has spread beyond the pelvis to the abdominal lining, lymph nodes, or other nearby organs); Stage IV (cancer has spread to distant organs, such as the liver or lungs). The stage of OC not only guides treatment decisions but also provides important prognostic implications. Unfortunately, most cases of EOC are diagnosed in advanced stages

and typically require a combination of surgery and chemotherapy for treatment. Despite these aggressive interventions, survival rates for patients with advanced stage EOC remain low, highlighting the critical need for more effective diagnostic and therapeutic approaches (O'Shea, 2022).

1.2.4   Sporadic BC and OC

BC and OC are multifactorial and highly heterogeneous diseases, with most cases classified as sporadic. In sporadic cases, genetic alterations are somatic, meaning they are acquired during an individual's lifetime and are specific to cancer cells (Harbeck et al., 2019). The underlying cause of sporadic cancer involves a combination of internal factors, lifestyle factors such as smoking and obesity, and environmental factors, including exposure to sun, radiation, or certain chemicals (Bissonauth et al., 2008; Cohen et al., 2023; Maas et al., 2016).

The liability threshold model is a concept used to explain the interplay between genetic and environmental factors in the development of complex genetic disorders, including sporadic BC and OC. This model proposes the existence of a cumulative genetic and environmental threshold, which, when surpassed, results in the manifestation of the disease. The crossing of this threshold may arise from a combination of various genetic factors and environmental exposures (Dahlqwist et al., 2019).

Sporadic cancers emerge through a gradual accumulation of acquired and unrepaired somatic mutations, including activation of oncogenes, frequently accompanied by inactivation of tumour suppressor genes. These mutations are likely early events in sporadic tumours, followed by subsequent accumulation of independent mutations in several other genes (Kenemans et al., 2008).

Notably, individuals with sporadic BC typically present at a significantly higher mean age compared to BC patients with germline *BRCA1* PVs. For instance, the respective ages are 64 years for sporadic BC and 42 years for *BRCA1*-related BC. This age difference serves as a powerful discriminator between individuals with *BRCA1*-related and those with sporadic BC (Filippini & Vega, 2013; van der Groep et al., 2006).

1.2.5   Hereditary breast and ovarian cancer (HBOC)

Hereditary breast and ovarian cancer (HBOC) is an inherited condition characterised by a genetic predisposition to early onset BC or OC, particularly occurring at the younger age (especially before the age of 50 years) (Petrucelli et al., 2010). HBOC is estimated to account for approximately 5–10 % of all BC cases and 10–15 % of all OC cases occurring in the general population and is associated with germline PVs in certain cancer predisposition genes (Angeli et al., 2020; Leitsalu et al., 2021).

In most cases, HBOC is associated with germline PVs in the *BRCA1* and *BRCA2* genes, which are inherited in an autosomal dominant (AD) manner. This means that only one altered gene copy from either parent is sufficient to increase the lifetime risk of developing BC or OC. Moreover, individuals with HBOC have an elevated risk of developing other cancers, such as melanoma, pancreatic, and prostate cancers (Hampel et al., 2015; Sabiani et al., 2020; Yoshida, 2021).

Generally, *BRCA1/2*-associated HBOC is suspected in individuals with personal or family history, especially when the disease is diagnosed at a relatively young age (Leitsalu et al., 2021).

1.2.6 Risk factors

In the following chapters, we will discuss the numerous factors contributing to the risk of cancer development. These factors cover both modifiable and non-modifiable elements, including environmental, behavioural, and lifestyle factors. Additionally, the risk of cancer development is influenced by various genetic factors, as well as interactions between these genetic and environmental/behavioural/lifestyle factors.

Modifiable risk factors

Modifiable risk factors are important components contributing significantly to the complex landscape of cancer development – this represents the 'missing information' that is essential for assessing the individuals' risk.

For instance, smoking, both current and former, is identified as the primary modifiable risk factor for the development of BC or OC, followed by obesity (body mass index (BMI) > 30 kg/m^2). Additionally, physical inactivity and alcohol consumption are recognised as modifiable risk factor that increase the risk of BC development (Cohen et al., 2023; Maas et al., 2016).

Moreover, increasing parity and breastfeeding for one year and more are observed to have a protective effect against both BC and OC in *BRCA1* PV carriers. In contrast, the *BRCA2* PV carriers have an increased risk of BC with each full-term pregnancy before age of 50, along with an increased risk of OC and breastfeeding demonstrates no protective effect (Antoniou et al., 2006; Cullinane et al., 2005; Jernström et al., 2004; McLaughlin et al., 2007). Additionally, the use of oral contraceptives has been associated with a reduced risk of OC development but potentially a slightly increased risk of BC development in general population. However, while *BRCA1* PV carriers also experience the same reduced risk of OC, there is no clear evidence of an increased risk of BC (Iodice et al., 2010; Jürgens et al., 2022).

There is a conflicting information about the usage of hormone replacement therapy (HRT) and BC risk, where most studies indicate no association, a few suggest an increased risk in *BRCA1* PV carriers, and one study even indicates a decreased BC risk in the general population (Cohen et al., 2023). However, HRT has been listed as a modifiable risk factor for OC development. Other controversial risk factors include infertility, as well as fertility medications (Ali et al., 2023).

In summary, modifiable risk factors include behaviours and exposures that can be altered, such as tobacco usage, alcohol consumption, excess body weight, physical inactivity, and dietary habits, as well as access to routine cancer screening tests. Addressing these modifiable risk factors holds the potential to prevent a significant proportion of cancer cases and deaths (Stein & Colditz, 2004).

Non-modifiable risk factors

Extensive epidemiological research in BC and OC patients has identified numerous non-modifiable risk factors that significantly influence cancer development. These include advanced age, gender, race, ethnicity, family history of cancer, and genetic predisposition, which will be discussed in more detail in the next chapter. Other non-modifiable risk factors include time of menarche and age at menopause. Considering these factors is crucial when assessing an individual's cancer risk and implementing appropriate prevention and screening strategies (Ferris et al., 2023; Maas et al., 2016).

In summary, non-modifiable risk factors significantly contribute to cancer development and should be integrated into public health initiatives, clinical practice, and individual risk assessments.

Genetic risk factors

There are several well-known genetic factors associated with the risk of developing BC or OC. These genetic factors can contribute to disease predisposition through either monogenic risk variants, which disrupt important physiological pathways with substantial effect on disease progression, or polygenic risk, involving numerous variants with smaller effects across several pathways (Fahed et al., 2020).

Although best-know monogenic risk variants associated with BC and OC are *BRCA1* and *BRCA2* PVs, recent advances in molecular techniques, especially NGS, have revolutionised the field by identifying various new genes associated with genetic predisposition to BC and OC, each characterised with different penetrance estimates (Angeli et al., 2020). Accumulated evidence has highlighted recurrent alterations in multiple genes, including *TP53*, *ESR1*, *PIK3CA*, *PTEN*, *CDH1*, *GATA3*, *CCND1*, *FGFR1/2*, *ERBB2*, *CDKN2A/B*, and *MYC*, that can

lead to dysregulations of various signalling pathways (Fonseca-Montaño et al., 2023). For instance, studies have demonstrated that alterations in *TP53* gene can lead to dysregulation in DNA damage repair pathways, influencing the cell's ability to maintain genomic integrity (Hanahan & Weinberg, 2000). Similarly, mutations in *ESR1* and *GATA3* genes have been linked to disruptions in hormone receptor signalling pathways, contributing to abnormal cell survival and proliferation. These mutations may also contribute to a hormonal therapy resistance (Bianco et al., 2022; Miziak et al., 2023). Additionally, dysregulation in the PI3K/AKT/mTOR pathway, due to mutations in *PIK3CA* and *PTEN* genes, is frequently observed in various cancers, including BC and OC. This pathway is important in cell survival, proliferation, and differentiation, with its activation associated with therapy resistance. Consequently, this pathway has emerged as an attractive target for anticancer therapy (Sirico et al., 2023).

These molecular changes are associated with key cellular processes such as cell survival, proliferation, epithelial-mesenchymal transition (EMT), therapy resistance, immune evasion, and alterations in the tumour immune microenvironment (TIME) (Fonseca-Montaño et al., 2023). In summary, the insight into the functional consequences of genetic alterations provides a deeper understanding of the complex mechanisms driving cancer progression and informs potential therapeutic targets.

High- and moderate-penetrance genes affecting BC and OC risk

Female carriers of PVs in high and moderate penetrance susceptibility genes face significantly elevated risks of developing BC and OC compared to the general population (Dareng et al., 2022). Penetrance, in this context, refers to the probability that individuals carrying specific genetic PVs will develop a particular trait or condition, in this case, a higher risk of developing BC or OC, and it is usually defined in terms of specific age, for example, to age 70 (Narod, 2002). High-penetrance genes confer substantially increased risk of cancer development, while moderate-penetrance genes contribute to an elevated but comparatively lower risk (Shiovitz & Korde, 2015). This distinction and the incorporation of penetrance estimates in risk prediction models are crucial in risk assessment and genetic counselling, as carriers of high-penetrance genes may benefit from more intensive surveillance and preventive measures (Mavaddat et al., 2013; Narod, 2002).

High-penetrance genes

Approximately 5–10 % of all BC cases are believed to originate from high-impact germline PVs in BC susceptibility genes, making them hereditary. Within this percentage, up to 30 % can be linked to PVs in *BRCA1* and *BRCA2* genes, with a smaller proportion involving other susceptibility genes such as *TP53*, *STK11*, *PTEN*, and *CDH1*. While PVs in these genes

are rare, their high penetrance is associated with various genetic syndromes and a significantly elevated risk of developing various cancers, including BC and OC (Angeli et al., 2020; Jürgens et al., 2022).

However, rare variants in well-known high- and moderate-penetrance susceptibility genes such as *BRCA1*, *BRCA2*, *BRIP1*, *PALB2, RAD51C, RAD51D* and the mismatch repair genes contribute to approximately one third of the inherited component of EOC development risk. Additionally, various common low-penetrance susceptibility variants have been identified, explaining additional 6 % of EOC heritability (Dareng et al., 2022; Jones et al., 2017; Lyra et al., 2020).

Moderate- and low-penetrance genes

Moderate-penetrance genes such as *ATM*, *PALB2*, *CHEK2*, *NBN*, and *NF1* are associated with 2–5 times higher relative risk for BC and other cancers, depending on particular gene (Jürgens et al., 2022). For instance, *CHEK2* is a tumour suppressor gene, located on chromosome 22q12.1, and it encodes a serine/threonine-protein kinase critical for various cellular processes, including DNA repair, cell cycle arrest, and apoptosis. PVs in *CHEK2* can disrupt these vital functions, potentially contributing to an increased risk of BC and other malignancies (Apostolou & Papasotiriou, 2017; Mars, Widén, et al., 2020; Pavlovica et al., 2022). *CHEK2* is known to be involved in DNA damage repair pathway by interacting with various tumour suppressor genes, including *BRCA1* (Bartek & Lukas, 2003). It has been suggested that an additional gene defects in the same pathway, such as *CHEK2* variants, might have a multiplicative effect in *BRCA1* PV carriers, increasing the susceptibility to DNA damage and genomic instability. However, the results of previous studies have been inconsistent, and further research is needed to understand if allelic variants in *CHEK2* gene might have a potential role in influencing *BRCA1* PV penetrance (Cybulski et al., 2009; Sokolenko et al., 2014). Understanding the interplay between *CHEK2* and *BRCA1* could provide valuable insights into cancer susceptibility and improve personalised cancer risk assessment.

Other moderate-penetrance susceptibility genes to BC and OC are *BARD1*, *BRIP1*, *RAD51C, RAD51D,* and the mismatch repair genes, including *EPCAM*, *MLH1*, *MSH2*, *MSH6*, *PMS2* (Angeli et al., 2020).

1.3   *BRCA1* gene and its role in cancer development

The following chapter will focus on *BRCA1* gene PVs and the genetic factors influencing their penetrance, which is a central theme of this Thesis. This chapter will present a comprehensive exploration of the *BRCA1* gene, including its function, its crucial role in cancer development and HBOC, and its contribution to diagnosis, treatment, and prophylactic

strategies. Additionally, this chapter will explore *BRCA1* PV prevalence in various populations, with a specific emphasis on Latvia and the Baltic States. The discussion will extend to the varying penetrance of the *BRCA1* gene and underlying factors that might affect it.

1.3.1 Function of *BRCA1* gene

The *BRCA1* gene, located on chromosome 17q21.31, is a critical component in various cellular processes such as DNA repair, cell cycle checkpoint control, and the maintenance of genomic stability. Functioning as a tumour suppressor, the BRCA1 protein interacts with other tumour suppressors, DNA damage sensors, and signal transducers. Together, they form the *BRCA1*-associated genome surveillance complex (BASC), a large multi-subunit protein complex essential for the surveillance of genome integrity (Angeli et al., 2020). It is well-known that both BRCA1 and BRCA2 proteins suppress the formation of tumours by homologous recombination DNA repair (HRR) that ensures genomic stability (Krais & Johnson, 2020).

1.3.2 Homologous recombination deficiency (HRD) and cancer development

DNA damage is unavoidable, multifactorial, and affects genomic stability. HRR is an essential and evolutionarily conserved process that corrects double-stranded breaks (DSBs) by using a sister chromatid as a repair template, thereby ensuring the integrity of the genome (Creeden et al., 2021). HR is one of the major pathways for the repairing DNA DSBs in eukaryotic cells (Mekonnen et al., 2022).

The *BRCA1* gene plays an important role in the HRR pathway by interacting with other DNA repair proteins. Because the PVs in this gene cause homologous recombination deficiency (HRD), which accumulates DNA damage and genomic instability, they can contribute to various malignancies, increasing susceptibility to carcinogenesis. Numerous studies have underscored the important role of HRD in the development of different cancers, including BC and OC in individuals carrying germline *BRCA1* PVs (den Brok et al., 2017; Mekonnen et al., 2022; Prakash et al., 2015).

1.3.3 The prevalence of germline *BRCA1* PVs in different populations and in Latvia

Different ethnic groups and geographical regions have distinct *BRCA1* PV spectrum and prevalence. Research has shown that population genetics can vary significantly by region, highlighting the importance of considering local genetic structures in large-scale genetic studies. In the Baltic States, the genetic structure is closely correlated with the regional geography, resulting in reduced genetic heterogeneity and increased practical utility for genetic studies in this region (Janavičius, 2010; Janavičius et al., 2013; Pankratov et al., 2020). Consequently, extensive research has identified two region-specific germline PVs in

the *BRCA1* gene – c.4035del and c.5266dup – as founder variants. These variants account for approximately 80 % of all identified PVs in the *BRCA1* gene in BC and OC patients in Latvia and other Baltic countries (Gardovskis et al., 2005; Janavičius et al., 2014; Tamboom et al., 2010; Tikhomirova et al., 2005). Similar enrichment of these rare and low-frequency *BRCA1* PVs has been demonstrated in other populations (Kerr et al., 2023; Xue et al., 2017).

The prevalence of *BRCA1* PV c.5266dup is notably high in Central and Eastern Northern European countries, including Poland, Finland, Latvia, Estonia, and Lithuania, and Russia. Similarly, the *BRCA1* PV c.4035del is considered a founder variant in Central and Eastern Europe (Janavičius et al., 2014; Sokolenko et al., 2014). Founder PVs, such as these two, demonstrate a higher prevalence in certain racial/ethnic groups and account for the majority of observed PVs within these populations. The founder effect is most frequently observed in geographically, culturally or religiously isolated populations, which typically have less genetic diversity. Confirmation of their status as true founder PVs is based on the presence of common ancestral haplotypes (Janavičius et al., 2013; Rebbeck et al., 2018).

However, it is important to note that the prevalence of germline *BRCA1* PVs shows significant variation among different ethnic groups and geographical regions. A comprehensive overview of the 10 most common *BRCA1* PVs in each ethnic group is provided in Table 1.2. Among the *BRCA1* PVs identified in each racial/ethnic group, some are observed in multiple populations, including c.5266dup (Rebbeck et al., 2018). This observation suggests a common genetic ancestry or historical migration patterns leading to the presence of specific *BRCA1* PVs in different racial and ethnic groups.

Table 1.2

Ten most frequently observed *BRCA1* PVs by self-identified race/ethnicity
Based on (Rebbeck et al., 2018)

Caucasian	African American	Asian	Hispanic/Latino	Jewish	Other
c.5266dup (17 %)	c.815_824dup (16 %)	c.390C>A (4 %)	c.68_69del (12 %)	c.68_69del (72 %)	**c.5266dup (12 %)**
c.181T>G (6 %)	c.5324T>G (7 %)	c.5496_5506 delinsA (3 %)	c.3331_3334del (10 %)	**c.5266dup (24 %)**	c.68_69del (17 %)
c.68_69del (6 %)	c.5177_5180del (5 %)	c.470_471del (3 %)	c.5123C>A (9 %)	c.3756_3759del (0.3 %)	c.181T>G (5 %)
c.4035del (2 %)	c.4357+1G>A (5 %)	c.5503C>T (2 %)	c.548−?_4185+? del (7 %)	c.1757del (0.3 %)	c.5333−36_5406 +0del (3 %)
c.4065_4068del (2 %)	c.190T>G (3 %)	c.922_924 delinsT (2 %)	c.211A>G (5 %)	c.2934T>G (0.2 %)	c.3481_3491del (2 %)
c.3756_3759del (2 %)	c.68_69del (3 %)	c.68_69del (2 %)	c.815_824del (3 %)	c.5503C>T (0.1 %)	c.1687C>T (2 %)
c.1687C>T (2 %)	c.5467+1G>A (3 %)	c.3770_3771 del (2 %)	c.2433del (3 %)	c.4185+1G>T (0.1 %)	c.4065_4068del (2 %)
c.4327C>T (2 %)	c.182G>A (3 %)	c.2635G>T (2 %)	c.1960A>T (3 %)	c.4689C>G (0.1 %)	c.5277+1G>A (2 %)
c.2475del (2 %)	c.5251C>T (2 %)	c.2726dup (2 %)	c.3029_3030del (3 %)	c.3770_3771del (0.1 %)	c.2685_2686del (68 %)
c.4186−?_4357+ ?dup (1 %)	c.4484G>T (2 %)	c.3627dup (2 %)	c.4327C>T (2 %)	c.4936del (0.1 %)	c.4327C>T (1 %)

The most prevalent *BRCA1* PV c.5266dup in the Baltic region*,* is believed to have originated in Scandinavia or northern Russia approximately 1800 years ago and subsequently entered the Ashkenazi Jewish population in Poland around 400–500 years ago (Hamel et al., 2011; Rebbeck et al., 2018). Furthermore, Lithuania colleagues conducted a haplotype analysis in 78 unrelated *BRCA1* PV c.4035del carriers from Lithuania, Latvia, Poland and Russia to estimate the age of this variant. By applying the maximum likelihood method, they speculated that c.4035del arose 1550 years ago, most likely in the territory of Lithuania from ancient Balts, and subsequently gradually entered the genetic pool of neighbouring countries (Janavičius et al., 2013).

It is important to note that the frequencies reported in many studies are not predominantly population-based, especially in settings where founder PVs are selectively screened. Consequently, they may be influenced by testing biases. While most studies of *BRCA1* PVs are based on clinical cohorts, which may not fully reflect the populations they aim to represent, they do provide some estimates of the population frequency of PVs (Kerr et al., 2023; Rebbeck et al., 2018).

The overall prevalence of *BRCA1* mutations is estimated at 1 in 300, but this number may vary due to enrichment on founder PVs in specific geographic locations, leading to an increased prevalence of PVs in those populations (Hampel et al., 2015). Unfortunately, many studies lack comprehensive population data on the general prevalence of *BRCA1* PVs. Reported prevalence might significantly vary, as estimates are strongly affected by the specific characteristics of the study cohort (Jürgens et al., 2022; Leitsalu et al., 2021).

The increasing use of NGS technologies has led to a substantial increase in reported PVs in the *BRCA1* gene. In 2010, the Breast Cancer Information Core (BIC) website (http://research.nhgri.nih.gov/projects/bic/) and Cancer Genetics Web for *BRCA1* described over 1,500 PVs in the *BRCA1* gene alone (Tamboom et al., 2010). In 2018, the Consortium of Investigators of Modifiers of *BRCA1/2* (CIMBA) presented an overview of the existing *BRCA1* disease-associated PVs. They reported 1,650 unique PVs in *BRCA1* gene (Rebbeck et al., 2018). This highlights the increasing and diverse mutational spectrum associated with *BRCA1* PVs.

1.3.4 The role of germline *BRCA1* PVs in HBOC

As discussed in the previous chapters, the inherited PVs in the *BRCA1* gene represent the most prevalent cause of HBOC. The role of *BRCA1* gene in OC and early-onset BC susceptibility was identified in the early 1990s (Futreal et al., 1994; Miki et al., 1994). Associated lifetime risk for BC development has been evaluated by various studies, indicating the recent estimates to be from 60 % to 75 % by the age of 80 for female carriers of germline *BRCA1* PVs (Borde et al., 2022; Kuchenbaecker, Hopper, et al., 2017; Mavaddat et al., 2013). The corresponding OC risk has been estimated to be between 34 % to 44 % by the age of 80 for female carriers of germline *BRCA1* PVs (Barnes et al., 2020; Rebbeck et al., 2015). However, risk estimates from different studies have wide confidence intervals that could be explained by different sampling strategies (e.g. population based vs. clinical samples), population and PV characteristics, analytic methods, and other (Kuchenbaecker, Hopper, et al., 2017).

BC in *BRCA1* PV carriers is typically characterised by TN and more aggressive high-grade carcinomas. Individuals with *BRCA1* and *BRCA2* PVs are diagnosed at a younger age compared to non-carriers, and the distinctive tumour characteristics associated with *BRCA1* and *BRCA2* PVs are more prevalent in the younger age groups (Muranen et al., 2023).

1.3.5 The role of germline *BRCA1* in BC and OC diagnosis

Germline genetic testing for *BRCA1* PVs has been an integral part of clinical practice, with widespread availability in most developed countries. Referral considerations include individuals with a personal history of or first-degree relatives with specific cancer types and criteria related to age, ancestry, and family history patterns. However, the introduction of panel

tests by various clinical diagnostic services has expanded the use of these tests to a much broader group of individuals (Hampel et al., 2015; Jürgens et al., 2022; H. Li et al., 2022).

Considering the prevalence of specific PVs, a targeted approach, such as testing for the *BRCA1* PV c.5266dup in Central Eastern European populations, is a valuable approach before considering full gene sequencing. In the Baltic States, screening for *BRCA1* PV c.4035del, the second most common PV in the region, complements this basic testing (Rebbeck et al., 2018). This strategy is already applied in Latvian *BRCA1* PV screening, and once a specific *BRCA1* PV is identified in a family, other members, both affected and unaffected, can undergo 'cascade testing' to determine their carrier status.

Several biobanks globally, including those in Australia, Northern Europe, and the United States have applied a genotype-first approach. This approach involves testing and recontacting individuals carrying clinically significant PVs in actionable genes, irrespective of family history or medical indication, presenting an alternative to common clinical practice. In a future perspective, genotype-first screening is an appealing approach to enhance long-term outcomes for high-risk individuals in the population, who may be unaware of their genetic risk (Leitsalu et al., 2021; Manickam et al., 2018; Rowley et al., 2019).

1.3.6   The role of germline *BRCA1* PVs in BC and OC treatment

BRCA1 PVs significantly influence cancer treatment decisions, particularly regarding the use of platinum agents or poly (ADP-ribose) polymerase (PARP) inhibitors as the germline carriers of *BRCA1* PVs exhibit a favourable response to therapies targeting HRR pathways (den Brok et al., 2017; Rebbeck et al., 2018; Yoshida, 2021). Information on the germline *BRCA1* PV status has a positive impact on treatment options, leading to improved prognosis. Moreover, PARP inhibitor therapies, initially successful in *BRCA1*-related BC and OC, are now expanding to treat pancreatic, prostate, and potentially stomach cancers in near future (S. Li et al., 2022).

The rationale behind using PARP inhibitors in germline *BRCA1* PV carriers is based on the concept of synthetic lethality. It suggests that the cells treated with PARP inhibitors with only one mutated allele remain compatible with life, but a somatic mutation in the other allele triggers cellular death. PARP inhibitors exploit *BRCA1* PVs and DNA damage response (DDR) deficiencies, causing cell death in HRD cancers (Konecny & Kristeleit, 2016; Lord & Ashworth, 2017).

Additionally, platinum salts like cisplatin and carboplatin are effective treatments for BC and OC. These agents act as DNA cross-linking agents, forming intra-strand crosslinks with the purine bases on DNA, which induces DNA damage and subsequently triggers apoptosis in

affected cells. They exhibit good efficacy, particularly in cells lacking functional HR mechanism (Dasari & Tchounwou, 2014; Lord & Ashworth, 2017).

The characteristic mutational signature of cancers with *BRCA1* PVs, marked by HRD, makes them responsive to platinum-based therapy or PARP-inhibitors. However, despite this responsiveness, many *BRCA1* PV carriers still undergo treatment based on standard indications (Muranen et al., 2023).

1.3.7   The role of germline *BRCA1* PVs in BC or OC prevention strategies

Currently, the clinical management of women carrying *BRCA1* PVs focuses on a comprehensive strategy involving early diagnosis and cancer risk reduction by intensified medical surveillance, risk-reducing surgeries, and chemoprevention (Kuchenbaecker, McGuffog, et al., 2017). In accordance with the National Comprehensive Cancer Network (NCCN) guidelines, women with germline *BRCA1* PVs are recommended to undergo an extensive surveillance protocol, including clinical breast examination every 6–12 months and annual breast magnetic resonance imaging (MRI), starting at the age of 25, annual mammography with tomosynthesis, also known as 3D mammography, starting at the age of 30, and annual transvaginal ultrasound and the measurements of serum cancer antigen 125 (CA-125) concentration, starting at the age of 30–35, although the benefits of latter are uncertain. Additionally, women with *BRCA1* PVs should consider the option of a bilateral risk-reducing mastectomy (BRRM) that involves surgery to remove healthy breast tissues and a bilateral risk-reducing salpingo-oophorectomy (BRRSO) that involves surgical removal of both fallopian tubes and ovaries. These preventive procedures are typically considered between 35 and 40 years and upon completion of childbearing (Angeli et al., 2020; Ludwig et al., 2016).

Previous studies have demonstrated effectiveness of risk-reducing procedures in lowering the overall cancer risk and mortality for germline *BRCA1/2* PV carriers. In most recent studies, BRRSO has shown a significant reduction of OC risk by 72–88 %, and BRRM has been associated with a 90–95 % reduction in BC risk for women carrying *BRCA1/2* PVs (Domchek et al., 2010; Ludwig et al., 2016; Wang et al., 2022). However, it's important to note that these prophylactic procedures are invasive and can have severe psychological and physiological effects. Accurate age-dependent estimations of cancer penetrance in *BRCA1* PV carriers are critical in genetic counselling, allowing to make informed decisions about preventive measures that correspond to personalised BC and OC risk. Enhanced risk prediction can help identify high-risk women who may benefit from early clinical intervention and low-risk women who may decide to postpone prophylactic procedures or chemoprevention (Borde et al., 2022; Kuchenbaecker, McGuffog, et al., 2017).

1.3.8 The penetrance of germline *BRCA1* PVs

Germline PVs in *BRCA1* gene is recognised as the most penetrant genetic predisposition for both BC and OC. However, incomplete penetrance is observed, where a subset of *BRCA1* PV carriers never develops BC or OC in their lifetime. This AD inheritance pattern of *BRCA1*-associated HBOC presents challenges in genetic counselling and risk assessment due to the variability in penetrance among individuals carrying *BRCA1* PVs. Other genetic factors are suggested to contribute to this phenomenon (Chen et al., 2020; Downs et al., 2019).

Despite the high estimated cancer development risk associated with germline *BRCA1* PVs, with the highest estimates in majority of studies not exceeding 75 % by the age of 80 years (Borde et al., 2022), a significant portion of carriers are unlikely to develop BC or OC. Investigating the causes of incomplete penetrance is crucial, as it can explain the genetic factors affecting cancer development in *BRCA1* PV carriers. This understanding is essential for predicting the likelihood of disease development and may suggest potential strategy for disease prevention (Downs et al., 2019).

As reliable penetrance estimates of *BRCA1* PVs are critical for informed decision-making, and various retrospective and prospective studies have been performed using either clinical cohorts or population-based studies of cancer patients. For instance, Antoniou et al. reported an overall penetrance to age of 70 years for all tested *BRCA1* PVs to be 59 % in BC patients and 34 % in OC patients. A more comprehensive analysis by Mavaddat et. al. reported penetrance estimates ranging from 40 % to 87 % for BC and from 22 % to 65 % for OC (Antoniou et al., 2008; Mavaddat et al., 2013). Furthermore, a more recent study of PV-specific penetrance of BC and OC among *BRCA1* PV carriers demonstrates variability, ranging from 56–83 % and 10–43 %, respectively (Rebbeck et al., 2015).

In conclusion, current research on PV-specific penetrance is limited, highlighting the need for further investigation into the factors influencing the variable penetrance of germline *BRCA1* PVs.

Factors that might affect the penetrance of *BRCA1* PVs

As established in previous chapters, PVs in the *BRCA1* gene are associated with an elevated risk of both BC and OC, but not all PVs within the gene have equal penetrance. The overall penetrance of *BRCA1* PVs might be affected by the type or localization of the variant, as well as family history, influence of age and gender, environmental and various genetic factors (Cooper et al., 2013).

In this Thesis, we will concentrate more on exploring the modifying genetic factors, which are essential for accurate risk assessment and personalised medical management for individuals with germline *BRCA1* PVs. These findings enable more targeted interventions, such

as personalised surveillance and preventive strategies, based on the specific characteristics of the identified PVs and the individual estimates of cancer development risk.

The type and localization of germline BRCA1 PVs

Early studies by Gayther et al. have already indicated a connection between the location of a specific *BRCA1* PV and the risk of both BC and OC development (Gayther et al., 1995). However, the knowledge is still limited about how the type of *BRCA1* PV affects the risk of BC or OC development (Rebbeck et al., 2015). Accumulated evidence and *in vitro* studies suggest that highly penetrant PVs in *BRCA1* gene predominantly disrupts BRCA1 protein activity (Kerr et al., 2023).

Typically, most analyses estimating penetrance have grouped all PVs together, regardless of variant type, and assumed similar associated risks. The majority of *BRCA1* PVs included in these studies were variants predicted to result in a transcript encoding a protein termination codon, leading to the nonsense-mediated decay (NMD) or encoding a truncated, inactive protein (H. Li et al., 2022). The penetrance of several specific germline *BRCA1* PVs has been shown to be associated to the characteristics of these variants themselves (Kerr et al., 2023). Different variants within the *BRCA1* gene may have varying impacts on protein function and, consequently, on the associated cancer risk. Certain PVs may have a more significant impact on the function of the BRCA1 protein, leading to a higher penetrance of cancer risk. Results of other studies suggest that cancer development risk varies by PV location, suggesting the potential benefit of extended genetic testing and individualised counselling (Kuchenbaecker, Hopper, et al., 2017).

Protein truncating PVs in the BRCA1 gene

The most common disease associated *BRCA1* PVs are those causing premature protein truncation, predicted to result in the loss of normal protein function. The predominant type of *BRCA1* PVs include nonsense variants, out-of-frame insertions/deletions, and variants affecting splicing. The majority of *BRCA1* PVs lead to premature translation termination, triggering NMD of an mRNA due to the presence of a stop codon within the first ~90 % of the coding region (Leitsalu et al., 2021; Rebbeck et al., 2018). However, accumulating evidence from genotype/phenotype studies suggests that even protein truncating variants may not all be associated with the same risks, depending on their position within the gene (H. Li et al., 2022).

Missense PVs in the BRCA1 gene

Analysis of missense variants in the *BRCA1* gene presents a challenge, as most such variants are initially considered of little or uncertain clinical significance (Aljarf et al., 2022).

However, through efforts of Evidence-Based Network Investigating Germline Mutant Alleles (ENIGMA) Consortium and following studies, numerous missense variants have been classified or reclassified as pathogenic using multifactorial methods. These consider factors like co-segregation, family history, and tumour histopathology to classify variants of uncertain significance (VUS) (Caputo et al., 2021; H. Li et al., 2022; Parsons et al., 2019).

Research has shown that missense *BRCA1* PVs, especially those located in functionally important domains (RING and BRCT), are associated with a reduced risk of BC compared to protein truncating variants. Many missense *BRCA1* PVs exhibit relative BC risks closer to those estimated for PVs in moderate-penetrance genes such as *ATM* and *CHEK2*, suggesting that surveillance might be an optimal approach for carriers of these PVs. However, limited data indicate that the risk of OC development is comparable to that found in literature for protein truncating variants (H. Li et al., 2022).

Recently, the functional evaluation of missense variants in *BRCA1* gene has been restricted by the limited number of variants subjected to experimental assessment. Moreover, nowadays the interpretation of variant pathogenicity relies more on *in silico* tools designed to predict functional effects, complemented by family-based data analysis. Growing use of *in silico* tools is expected to contribute to an increased identification of missense PVs (Aljarf et al., 2022). However, further genetic data collection and comprehensive analysis are essential to clarify the status of missense *BRCA1* PVs and their associated penetrance.

Variable expressivity and pleiotropy

Two important genetic concepts that should be mentioned in the context of incomplete penetrance is variable expressivity and pleiotropy. While distinct, they often interact and are challenging to distinguish in practice.

Variable expressivity refers to the phenomenon where affected individuals with shared genotype exhibit diverse severity of the phenotype, even among relatives. In the context of *BRCA1* PVs, variable expressivity can manifest as differences in the age of onset, tumour aggressiveness, and other characteristics of cancer presentation in carriers of the PV. This variability can lead to differences in disease penetrance, where some individuals may develop cancer at a relatively young age, while others may never develop cancer despite carrying the PV (Al-Mulla et al., 2009; Cooper et al., 2013; Kingdom & Wright, 2022).

However, pleiotropy refers to a concept where PVs in the same gene, such as *BRCA1*, can cause multiple traits or phenotypes (Ittisoponpisan et al., 2017). Previous research has demonstrated that pleiotropy is observed in significant fraction of genes and SNVs associated with different cancers, including approximately 34.8 % cancer related genes and 4.8 % cancer

related SNVs (Sivakumaran et al., 2011). In the case of *BRCA1* PVs, pleiotropy can frequently manifest as an increased risk not only for BC, but also for OC, prostate cancer, and other cancers (Yoshida, 2021). This phenomenon can contribute to the variability in disease presentation and penetrance of different *BRCA1* PVs among carriers.

Together, pleiotropy and variable expressivity contribute to the complexity of *BRCA1*-associated HBOC and can influence the incomplete penetrance observed in carriers of *BRCA1* PVs.

Other genetic factors

The penetrance of germline *BRCA1* PVs is significantly influenced by the presence of other genetic factors, such as epigenetic changes as well as modifier genes, which can either increase or decrease the risk of cancer development in these individuals. Multiple studies have demonstrated that the disease risk for *BRCA1* PV carriers follows a polygenic pattern and mode of inheritance, with increased cancer development risk observed in individuals with a higher number of affected first- and second-degree relatives. This observation suggests the contribution of other genetic factors in modulating cancer development risk for *BRCA1* PV carriers (Barnes et al., 2020; Doraczynska-Kowalik et al., 2022; Kingdom & Wright, 2022; Lavoro et al., 2022; Lee et al., 2020).

Moreover, PVs in actionable genes like *BRCA1* are often considered to have higher penetrance within the clinical context of a family history of the relevant condition compared to population-based cohorts. This higher penetrance may be attributed to the co-inheritance of multiple low-penetrance genetic modifiers (Forrest et al., 2022).

Consistent with this observation, an increasing number of common BC and OC susceptibility SNVs have been identified through population based GWAS. These studies persistently demonstrate the impact of common SNVs on the development risk of BC and OC in individuals carrying *BRCA1* PVs (Barnes et al., 2020; Kuchenbaecker, Hopper, et al., 2017; Mars, Widén, et al., 2020). By recognizing the significance of identifying these common genetic modifiers, the GWAS approach has been increasingly applied in recent years.

1.4   GWAS in BC and OC patients

GWAS have emerged as a powerful tool to discover genetic factors, mostly common SNVs (with minor allele frequency (MAF) > 0.01), that are associated with a particular disease status or phenotypic trait. Large-scale GWAS have successfully identified thousands of genetic loci associated with the risk of complex diseases, including BC and OC. These studies primarily compare disease cases with controls, providing valuable insights into molecular mechanisms and genetic factors influencing disease susceptibility. The extensive data generated through

GWAS present a powerful tool for enhancing clinical risk assessment in various diseases (Marees et al., 2018; Mars, Koskela, et al., 2020).

While a small percentage (5–10 %) of BC cases exhibit a significant hereditary component in the form of rare genetic variants (with MAF < 0.01), the majority include a substantial polygenic component. Large GWAS findings support this notion, as they have identified associations of over 100 genomic risk loci with BC in the European population (Läll et al., 2019). The broader understanding of the genetic architecture provided by GWAS will contribute to understanding the complex biological mechanisms underlying disease risk.

Given that high- and moderate-penetrance gene PVs contribute only a small proportion of inherited cancer risk, common low-penetrance variants identified through GWAS may explain a missing component in understanding cancer susceptibility (Kerr et al., 2023). This highlights the complementary role of GWAS in uncovering a spectrum of genetic factors that contribute to disease risk beyond traditional high- and moderate-penetrance genes.

Identifying inherited prognostic and predictive genetic biomarkers in patients, rather than in the tumour itself, is a promising approach. It provides insights into the mechanisms of tumour progression and facilitates more effective treatment stratification, potentially leading to increased therapeutic benefits (Escala-Garcia et al., 2019).

1.4.1 Previous GWAS in BC and OC patients to identify cancer susceptibility variants in the general population

GWAS have been useful in identifying genetic risk factors for BC and OC within the general population. These studies involve analysing genotyping data from large case-control studies to identify common small-effect SNVs that are statistically associated with an increased or decreased risk of specific trait or disease. The results of large-scale GWAS have successfully identified more than a hundred loci associated with the risk of BC or OC development. These variants represent common genetic SNVs that have an effect on a diverse range of molecular pathways, including various signalling pathways, in contrast to rare, high-risk PVs discovered in genes associated with high cancer susceptibility, often disrupting critical pathways involved in maintaining the integrity of DNA repair processes (Jurj et al., 2020; Mars, Widén, et al., 2020).

A considerable number of identified risks SNVs are located in non-coding regions of the genome, such as distal regulatory elements, including enhancers, promoters, and silencers. These elements may influence cancer development risk by controlling expression of target susceptibility genes. However, identifying their target genes is a major challenge (Amos et al., 2017; Edwards et al., 2013). Interestingly, by using *in silico* data to predict the target genes of

identified risk SNVs in BC patients, Michailidou et al. demonstrated a strong overlap between candidate target genes and somatic BC driver genes (Michailidou et al., 2017).

Among all cancer types, BC research has led to the identification of the greatest number of risk loci, accounting for more than 200 common SNVs associated with BC development risk (Jia et al., 2022; Yang et al., 2022). In contrast, large-scale GWAS in OC have identified only approximately 35 susceptibility loci, with common SNVs explaining approximately 3.9 % of the inherited component of EOC development risk. This suggests the likely existence of additional susceptibility loci yet to be discovered (Kuchenbaecker et al., 2015; Lu et al., 2018; Phelan et al., 2017).

Understanding the driving mechanisms responsible for malignant transformation holds promise in addressing challenges related to cancer recurrence and treatment resistance (Yang et al., 2022). However, further functional studies, including genome editing, oncogenic assays, and/or animal models, are crucial to evaluate whether the identified risk SNVs and their candidate genes are causal for BC or OC susceptibility (Michailidou et al., 2017).

1.4.2 Previous GWAS in BC and OC patients to identify cancer susceptibility variants in *BRCA1* PV carriers

Association studies have been conducted to evaluate common BC and OC susceptibility variants identified in the general population as potential modifiers or additional risk factors for individuals carrying *BRCA1* PVs. The primary objective of these studies was to understand whether these common genetic variants modify the risk of developing BC or OC in *BRCA1* PV carriers (Coignard et al., 2021).

Most of these studies have primarily focused on replicating associations already observed in the general population and evaluating their combined effects on the risk prediction for BC and OC development among *BRCA1* PV carriers. While 70–80 % of BC cases in *BRCA1* PV carriers are ER-negative, most common BC susceptibility SNVs have been identified through GWAS predominantly involving sporadic BC cases, the majority of which have ER-positive disease. Subsequently, specific GWAS have been conducted in germline *BRCA1* PV carriers to identify common genetic modifiers specific to this population (Couch et al., 2013; Milne & Antoniou, 2016; Milne et al., 2017).

Coignard et al. have demonstrated that over 50 common SNVs modify the risk of developing BC in *BRCA1* PV carriers. Among these, three SNVs were identified as carrier-specific susceptibility variants, suggesting the existence of additional variants specific to *BRCA1* PV carriers. Despite the inclusion of a large number of *BRCA1* PV carriers, the power to detect these genetic modifiers remains limited compared to those identified in the general population. These findings underscore the necessity of identifying additional genetic modifiers

specific to *BRCA1* PV carriers, which could contribute to improved personalised risk assessment for this high-risk population (Coignard et al., 2021).

However, studies of genetic modifiers of OC are underpowered and more challenging because of the smaller number of *BRCA1* PV carriers diagnosed with OC compared to BC. In 2016, a total of 11 common SNVs were identified as associated with OC risk in *BRCA1* PV carriers. Future studies are likely to yield additional common OC risk modifiers (Milne & Antoniou, 2011, 2016). Further functional studies of the identified SNVs should lead to a better understanding of the biological mechanisms underlying cancer susceptibility in *BRCA1* PV carriers (Milne et al., 2017).

1.5   The concept of polygenic risk scores (PRS)

Individually, GWAS have identified multiple common cancer susceptibility SNVs that modify disease risks only slightly, with OR typically close to 1. However, their combined effect, when summarised as a PRS, can be substantial (Borde et al., 2022; Mavaddat et al., 2019). PRS analysis does not aim to identify individual SNVs but rather to aggregate genetic risk across the genome into a single individual polygenic score for a specific trait (Mars, Koskela, et al., 2020).

A PRS is typically constructed as a weighted sum of a collection of genetic variants, calculated by multiplying the effect size of each variant from GWAS results by the individual's genotype score. This provides a single score that represents the individual's cumulative genetic risk for the disease. The resulting score is approximately normally distributed in the general population, where higher scores indicates a higher risk (Collister et al., 2022). These cumulative scores, reflecting the overall genetic burden, have demonstrated significant associations between high polygenic scores and disease status in various common complex diseases, including coronary heart disease (CHD), type 2 diabetes (T2D) and BC (Mars, Koskela, et al., 2020; Mars, Widén, et al., 2020).

Based on the GWAS results, several efficient PRSs have been developed for common complex diseases, offering potential improvements to existing risk prediction algorithms and the possibility of integration into future clinical risk assessments (Läll et al., 2019). PRSs have been explored as a tool to enhance individual risk stratification in complex disease, with potential application to both the general population and individuals carrying high-penetrance PVs in cancer susceptibility genes such as *BRCA1* (Dareng et al., 2022; Mars, Widén, et al., 2020; Pujol-Gualdo et al., 2022).

1.5.1 Different approaches of PRS calculations

The introduction of global biobank projects, as well as growing number of GWAS, has created the opportunity for more accurate assessment of the effects of genetic markers, offering new tools for developing personalised risk estimates (Cline et al., 2018; Lavoro et al., 2022; Szabo et al., 2000; Wang et al., 2023). PRSs hold the potential to be integrated into clinical risk models alongside independent biomarkers, such as *BRCA1* PV carrier status. Additionally, several techniques have been developed to calculate PRSs, each with distinct strengths and weaknesses (Wang et al., 2023).

The calculation of PRS involves determining the weighted sum of risk alleles for specific variants. However, the best approach for identifying the variant set and their weights to maximise the predictive power of a PRS remains unknown (Dareng et al., 2022). In PRS development, a common strategy is to use GWAS results, combining the effect sizes of numerous genetic markers that have reached a genome-wide significance and are statistically associated with a specific trait or disease. This approach typically incorporates tens to several hundred genetic markers (Läll et al., 2019; Mars, Koskela, et al., 2020; Uffelmann et al., 2021). Recently, more complex PRSs with potentially enhanced prediction accuracy have been developed using random-effects models. The authors used a Bayesian grouped mixture of regressions model (GMRM) to create joint PRS models containing a genome-wise set of 2,174,072 SNVs (Orliac et al., 2022). For joint PRS model calculations, two Bayesian approaches were implemented: the age-at-onset BayesW and case–control BayesRR-RC model. In simple terms, both models use a Bayesian approach to make probabilistic statements or estimates about the likelihood of outcomes, such as the status and age-of-onset of BC and/or OC, based on input data, such as genetic factors.

In this study, the BayesW model was used to predict the age at which a woman could develop BC and/or OC based on her genetic information. The BayesW model employs the Weibull distribution to simulate the time until the event – the onset of BC and/or OC. Furthermore, the BayesW model uses a unique representation of the Weibull distribution to model the age-at-onset of the disease. This involves taking the natural logarithm of the time until the disease occurs (time-to-event) and combining it with a measure of the distribution's shape (its moment), to define the parameters of the distribution. This approach enables a more precise modelling of the age-at-onset of BC and/or OC, as well as the genetic factors contributing to it (Ojavee et al., 2021).

In contrast, the grouped Dirac spike-and-slab model, referred to as BayesRR-RC and developed by Patxot et al., provides probabilistic insights into the genetic architecture. It implements an extended version of the BayesR model, offering estimates for group-specific

variance. This feature allows flexible prioritization of certain genomic regions, including intronic, exonic, and distal regulatory regions, resulting in robust model performance (Patxot et al., 2021). To simplify, we used the BayesRR-RC model to better understand the genetic factors contributing to the onset of BC and/or OC. The model adopts a statistical approach known as a grouped Dirac spike-and-slab model to analyse the data. This approach acknowledges that genetic markers might have different effects on different traits.

The model is called a "spike-and-slab" because it utilises two distinct probability distributions to represent the effects of genetic markers on a trait or disease. The "spike" distribution represents markers that have no effect, while the "slab" distribution represents markers influencing the onset of BC and/or OC. Additionally, it is called a "grouped" model because it allows the grouping of genetic markers into different categories based on characteristics, like their location in the genome (e.g. intronic, exonic, and distal regulatory regions) or function. This method identifies which genetic markers are more likely to influence a trait or disease and to estimate the size of that effect (Patxot et al., 2021).

In this study, we explored and compared the efficiency of these two PRS models (BayesW vs. BayesRR-RC) to estimate the overall genetic risk of women carrying the two most prevalent germline *BRCA1* PVs (c.4035del or c.5266dup) in the Latvian population. The goal was to evaluate the risk of developing BC or OC due to additional genetic variations.

1.5.2 Previous studies

Numerous studies have investigated the correlation between various PRSs and the risk of BC and OC development in individuals carrying *BRCA1* PVs. Previous research has demonstrated that PRSs contribute to improved BC risk prediction, not only in the general population, but also among women with germline PVs in high-risk genes and those with affected close relatives (Mars, Widén, et al., 2020). These studies often implement SNVs with established genome-wide significance to calculate PRSs for BC or OC risk prediction. While varying numbers of SNVs have been used, most PRSs consistently demonstrate a strong ability to predict future BC cases. Despite the availability of several proposed PRSs for BC risk prediction, there are no comprehensive comparison of the scores in existing literature (Läll et al., 2019). Additionally, research by Mars et al. indicates that individuals with an elevated PRS face a higher risk of developing bilateral BC following an initial diagnosis, with the PRS significantly improving risk assessment, especially among female first-degree relatives (Mars, Widén, et al., 2020).

Genetic risk profiling using PRSs has provided actionable insights for various cancers, including BC and prostate cancer. While PRSs for invasive EOC risk are still under

development, some studies have explored their potential both in the general population and among *BRCA1* PV carriers. For instance, Barnes et al. recently developed a PRS using 22 SNVs significantly associated with high-grade serous EOC development risk, demonstrating its efficacy and potential in predicting EOC risk, specifically in individuals with *BRCA1* PVs (Barnes et al., 2020; Dareng et al., 2022).

1.5.3 PRS potential in clinical practice

Personalised approaches based on individual risk levels deserve further assessment. Ideally, those should integrate available information from clinical risk factors and genetic information. Individualised genomic profiles, that implement PRSs, can be used to stratify women according to their risk of developing BC or OC. The genetic information could include both moderate- and high-penetrance germline PV testing, as well as PRSs (Läll et al., 2019; Mars, Widén, et al., 2020; Mavaddat et al., 2019). Increased PRS may recommend intensified medical surveillance and consideration of preventive procedures (such as risk-reducing surgery), as well as improved risk assessment of first-degree relatives (Mars, Koskela, et al., 2020). This in turn holds the promise of improved BC prevention and survival, by targeting screening or other preventative strategies for those women most likely to benefit (Mavaddat et al., 2019).

While PRSs have not yet been integrated into routine clinical practice and randomised clinical trials are needed, they represent a promising tool for improving preventative and personalised risk assessment strategies for various cancer types, particularly BC (Daly et al., 2021; Padrik et al., 2023). PRSs have demonstrated significant predictive accuracy, especially among women of European ancestry, suggesting their potential inclusion in risk prediction and prevention approaches for BC and OC in the future. However, further studies are required to improve and optimise existing PRSs, accounting for ancestral diversity. Additionally, validating the performance of PRSs, in combination with the inclusion of other genetic and lifestyle risk factors, is essential for ensuring reliable risk assessment before their implementation in clinical practice (Dareng et al., 2022; Leitsalu et al., 2021; Pujol-Gualdo et al., 2022).

To summarise, in this study, we address the significant healthcare burden of BC and OC, both globally and in Latvia, with a particular focus on the hereditary component associated to germline PVs in the *BRCA1* gene. While *BRCA1* PVs are known as the most penetrant genetic predisposition for both BC and OC, the variability in penetrance among PV carriers presents challenges in genetic counselling and risk assessment. Current risk assessment primarily relies on age, personal, and family history, but the incomplete penetrance of *BRCA1*

PVs highlights the need for exploration of additional penetrance modifying factors. Therefore, this Thesis aimed to explore potential genetic modifiers of BC or OC development risk in *BRCA1* PV carriers, particularly focusing on the region-specific *BRCA1* PVs in the Latvian population (c.4035del and c.5266dup). This was achieved through hypothesis-driven targeted candidate gene approach focusing on *BRCA1* and *CHEK2* double heterozygotes, followed by a data-driven GWAS approach. Additionally, we explored and compared the efficiency of two recently developed genome-wise PRSs, BayesW vs. BayesRR-RC, to estimate the overall genetic risk in women carrying the two most frequently identified germline *BRCA1* PVs. The goal was to contribute to enhanced risk prediction and stratification in this high-risk population.

2 Materials and methods

2.1 Study cohort

The study cohort consisted of 452 women who were selected based on two germline *BRCA1* PVs - NM_007294.4:c.4035del (rs80357711, previously referred to as c.4154delA) and NM_007294.4:c.5266dup (rs80357906, previously referred to as c.5382insC), present in a heterozygous state. Study participants were clinical cohort recruited continuously between 2002 and 2022, who were ≥ 18 years old and underwent germline genetic testing for HBOC syndrome at the Breast Surgery Unit of the Pauls Stradiņš Clinical University Hospital. Click here to enter text.Participants were diagnosed as affected with primary BC (n = 196), primary OC (n = 129) vs. unaffected (n = 127). The age of participants was censored at recruitment, and the follow-up data was not available. At the time of recruitment, none of the participants had undergone BRRM or BRRSO. DNA was isolated from peripheral blood by the FlexiGene DNA Kit (Qiagen, Germany) in accordance with the manufacturer's protocol.

Both tested variants are frameshift variants that result in a premature stop codon, leading to truncated (c.5266dup) or reduced (c.4035del) BRCA1 protein. Both variants are classified as pathogenic based on the American College of Medical Genetics and Genomics (ACMG) criteria (Richards et al., 2015), and their biological effect is LoF of the protein.

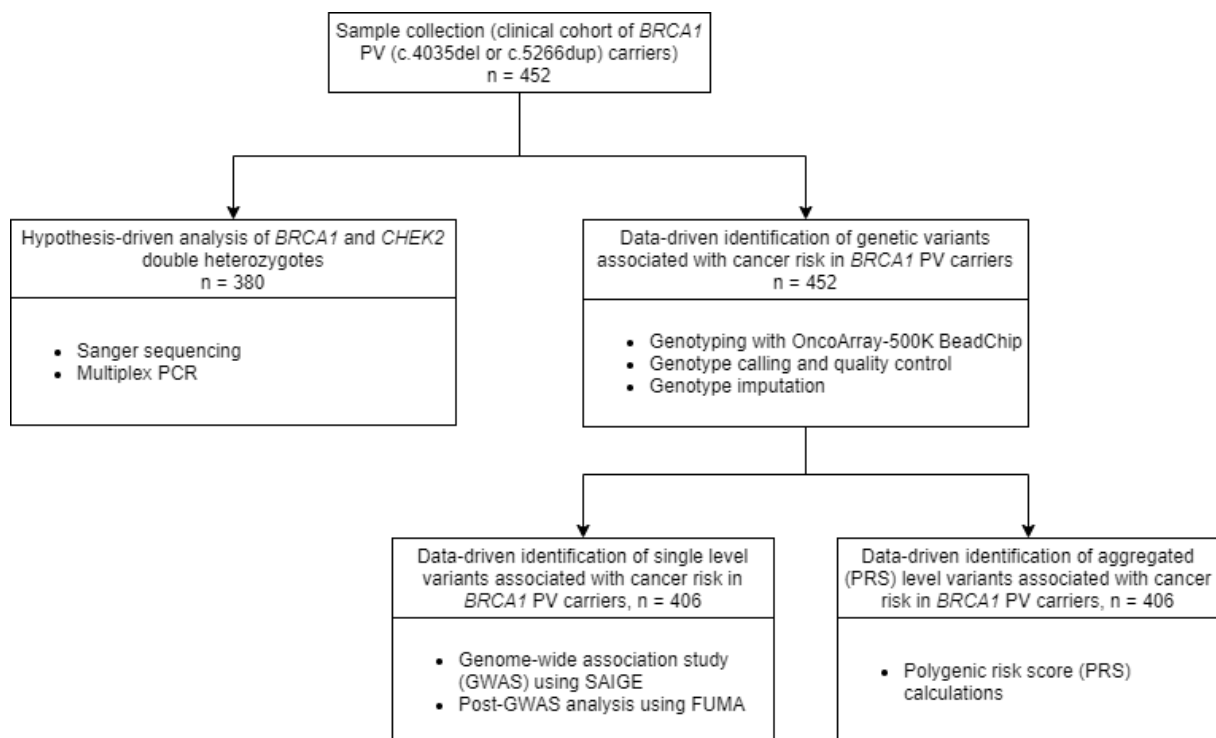Experimental setup and data analysis workflow are presented in the Figure 2.1.

Figure 2.1 **Diagram of analysis workflow presented in this Doctoral Thesis**

2.1.1 Ethics statement

This research received ethical approval from the Central Medical Ethics Committee of Latvia under protocol No 2/18-09-19, with Annex No 01-29.1.2/282 (please see Annexes 1–2). Additionally, approval was granted by the Genome Research Council under protocol No A-1/18-10-19 (please see Annex 3). The use of Estonian reference data was authorised through approval No 1.1-12/624, along with amendment 1.1-12/1478 by the Estonian Committee on Bioethics and Human Research (Estonian Ministry of Social Affairs).

2.1.2 Informed consent statement

Every participant who enrolled in this study provided written informed consent for the utilization of their clinical and genomic information for research purposes. The template of informed consent and accompanying description is provided in the Annexes 4–5.

2.2 Analysis of *BRCA1* and *CHEK2* double heterozygotes

At the study initiation in 2019, a hypothesis-driven analysis of *BRCA1* and *CHEK2* double heterozygotes was performed in 380 participants who were enrolled up to study onset (see Figure 2.1). The pathogenic/likely pathogenic and risk variants (Pavlovica et al., 2022) of *CHEK2* gene (splice site variant NM_007194.4:c.444+1G>A, p.(?), rs121908698 and missense variant NM_007194.4:c.470T>C, p.(Ile157Thr), rs17879961) were identified by Sanger's sequencing using BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA) along with primers as previously described (Cybulski et al., 2004). The sequencing results were analysed using the 3500 Genetic Analyzer (Applied Biosystems, USA). Data processing and editing was carried out using Sequencing Analysis Software and SeqScape™ Software (Applied Biosystems, USA) with reference to the Genome Reference Consortium Human Build 37 (GRCh37)/hg19 (released in 2009). Despite the availability of the more recent GRCh38 reference genome, we utilised GRCh37/hg19 not only for compatibility with earlier studies but also to ensure analysis tool compatibility and minimise potential errors associated with transitioning between genome builds. To detect the PV NM_007194.4:c.(908+1_909-1)_(1095+1_1096-1)del in *CHEK2* gene, which results in the deletion of exon 9-10 (also referred to as del5395), we used a multiplex polymerase chain reaction (PCR) approach (Veriti, Applied Biosystems, USA) as described elsewhere (Cybulski et al., 2007; Plonis et al., 2015). The products of the PCR reaction were separated using 2 % agarose gel. Confirmation of the deletion in multiplex PCR-positive samples was achieved through subsequent Sanger's sequencing. Detailed information about this methodology has been previously described (Cybulski et al., 2006), and the primer information is listed in Table 2.1.

Table 2.1

Primers used for analysis of *BRCA1* and *CHEK2* double heterozygotes

CHEK2 variant	Forward primer	Reverse primer
c.444+1G>A	ATTTATGAGCAATTTTTAAACG	TCCAGTAACCATAAGATAATAATATTAC
c.470T>C	ACCCATGTATCTAGGAGAGCTG	CCACTGTGATCTTCTATGTATGCA
del5395 primer pair 1	CTCTGTTGTGTACAAGTGAC	GTCTCAAACTTGGCTGCG
del5395 primer pair 2	TGTAATGAGCTGAGATTGTGC	CAGAAATGAGACAGGAAGTT

2.3 Genotyping with OncoArray-500K BeadChip

At the Institute of Oncology and Molecular Genetics, Rīga Stradiņš University, all 452 study samples were genotyped using the Infinium OncoArray-500K BeadChip (Illumina, San Diego, CA, USA) between 2019 and 2022. With a genome-wide backbone of 250,000 tag SNVs of common variants, the array has approximately 500,000 SNVs. The remaining markers are genetic variants linked to BC, OC, and other cancers that have been discovered through previous GWAS and other methods (Guo et al., 2015; Michailidou et al., 2015; Michailidou et al., 2013). The array has been developed in collaboration with leading experts from OncoArray consortium, including Breast Cancer Association Consortium (BCAC), CIMBA, and Ovarian Cancer Association Consortium (OCAC).

2.4 Genotype calling and quality control (QC)

A modified genotype QC process was followed for our dataset which have been described in detail elsewhere (Guo et al., 2014). In essence, this involved a sample based and variant based QC steps primarily using GenomeStudio software (Illumina, Genotyping module v2.0.5) and command-line program PLINK v1.07 and v1.9 (Purcell et al., 2007). Sample based and variant based QC workflow is presented in the Figure 2.2.

2.4.1 Primary genotype calling and QC in GenomeStudio

Sample genotype calling and PLINK format files were created using GenomeStudio software. Automatic clustering was performed by GenomeStudio during the manual data loading. Individuals were excluded from the analysis if their call-rate was < 98 % or if their sex defined by heterozygosity of X chromosomes did not match their sex in the phenotype data. Variant calls were filtered by GenTrain score and poor-quality variants (GenTrain score < 0.7) that appeared adjustable were selected for manual inspection and re-clustering. Final data was exported from GenomeStudio in PLINK format.
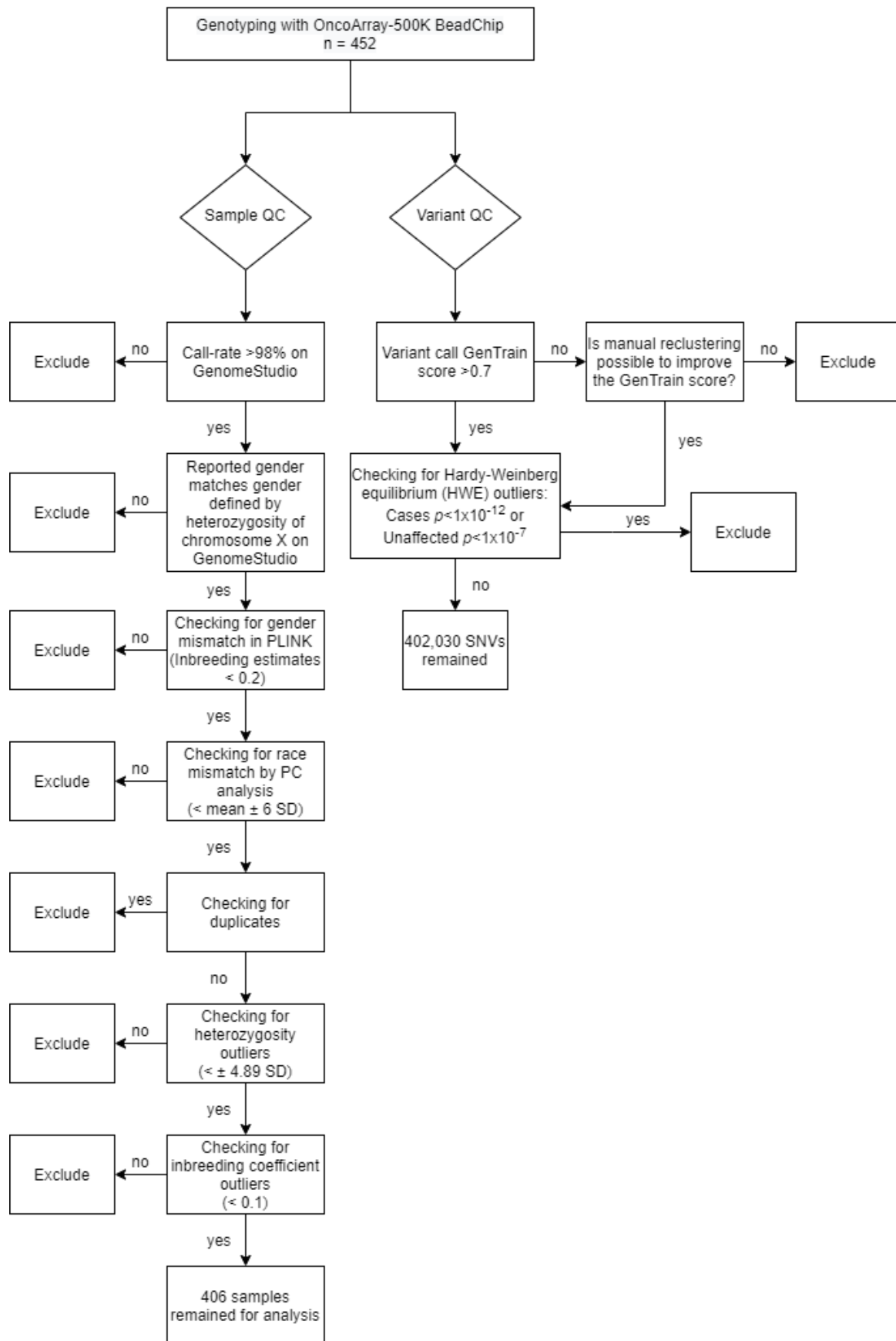
Figure 2.2 **Diagram of sample based, and variant based QC workflow applied in this Doctoral Thesis**

2.4.2 Converting all SNVs to the forward strand

Next, variant positions were updated to the human reference genome assembly GRCh37/hg19 and all variants were changed from the TOP strand to hg19 plus strand using GSAMD-24v1-0_20011747_A1-b37.strand.RefAlt.zip files that can be found at the https://www.well.ox.ac.uk/~wrayner/strand/ webpage.

2.4.3 Checking for gender mismatch

To check between reported and genotype-based gender mismatch, the inbreeding estimates for X chromosome were calculated using PLINK command --*check-sex*. Individuals were called as females and remained in the study dataset if an inbreeding estimate was < 0.2.

2.4.4 Checking for race mismatch by principal component (PC) analysis

Ancestry was calculated using a PC analysis by EIGENSOFT software (Price et al., 2006). To calculate PCs, the 687 ancestry informative markers (AIMs) from Illumina were applied. For the main analyses, first ten PCs were used and the threshold for race mismatch was set to (> mean ±6 standard deviation (SD)).

2.4.5 Checking for relatedness and probable duplicates

Relatedness and probable duplicates were identified through pair-wise identity by descent (IBD) calculation. Since IBD calculations are not aware of LD, we performed a LD pruning step. LD pruning involves filtering out variants in strong LD to enhance the independence of markers. In this study we utilised specific parameters for LD pruning, including a window size of 200 variants, a step size of 5 variants to shift the window at the end of each step, and a pairwise r^2 threshold of 0.2. In this process, pairs of variants in the current window with a squared correlation greater than the threshold were identified at each step. Variants were then pruned from the window until no such highly correlated pairs remain. The final dataset retained relatives, but probable duplicates with PI_HAT value close to 1 were excluded from the consecutive analysis.

2.4.6 Checking for Hardy-Weinberg equilibrium (HWE) outliers

The HWE test was performed to identify SNVs that deviate from HWE ($p < 1 \times 10^{-7}$ for unaffected individuals and $p < 1 \times 10^{-12}$ for cases) using the PLINK function --*hardy*. 503 SNVs were excluded from the dataset and 402,030 SNVs passed the initial QC.

2.4.7 Checking for heterozygosity and inbreeding outliers

The samples with extreme heterozygosity (deviation ±4.89 SD from the samples' heterozygosity rate mean) and inbreeding coefficient (> 0.1) were excluded from the dataset and 406 samples remained for the analysis.

2.5 Genotype imputation

For the imputation, additional SNVs with MAF<0.01 were excluded. Missing genotypes were imputed using the Estonian population based high coverage whole genome sequencing (WGS) dataset (n = 2244) as the reference panel, as described previously (Mitt et al., 2017). A two-stage imputation approach was implemented: phasing with EAGLE (Loh et al., 2016) and imputation with BEAGLE (Browning et al., 2018). Estimated genotypes were generated for approximately 38 million SNVs. Post-imputation QC was done, excluding SNVs with $MAF < 0.01$ and dosage R-squared (DR2) < 0.8. Filtered dataset contained 7,911,505 good quality SNVs for subsequent analysis.

2.6 Genome-wide association study (GWAS) using SAIGE

A total of 406 individuals were available for association analysis after dataset cleaning and imputation. Association analysis was carried out using software program R v4.0.2 (R. C. Team, 2020) package SAIGE v0.38 (Chen et al., 2016) to implement a mixed logistic regression model. The model was adjusted for relatedness, the first 4 PCs, age at recruitment/disease onset, and type of *BRCA1* PV. In this study, relatedness is adjusted for to minimise the risk of false positive associations and ensure that the genetic variants tested are genuinely associated with the outcomes (e.g. BC and OC) rather than being confounded by familial relationships. The implementation of a mixed logistic regression model, along with adjustment for relatedness and other covariates, helps to control for potential sources of bias by providing more reliable results. For association analysis a stringent significance threshold of $p < 5 \times 10^{-8}$ was used that in following Post-GWAS analysis was reduced to genome-wide suggestive significance threshold of $p < 1 \times 10^{-6}$.

2.7 Post-GWAS analysis using free access platform FUMA and VEP

The Functional Mapping and Annotation (FUMA) platform was used to annotate, prioritise, visualise, and interpret GWAS results. To identify independent significant SNVs, the SNVs with p values less than or equal to 1×10^{-6} and an $r^2 < 0.6$ were selected from GWAS results. Furthermore, to define lead SNVs from independent significant SNVs, the pairwise SNV threshold of $r^2 < 0.1$ was used. Next, the genomic risk loci in which SNVs were in LD with an r^2 coefficient exceeding 0.6 with the independent significant SNVs were detected.

The maximum distance of 250 kb between LD blocks to consolidate them into a single genomic locus was used. To conduct the LD analysis, the genetic data from 1000 Genome Project phase 3 was applied as a reference data.

Additionally, the Ensembl Variant Effect Prediction (VEP) tool was utilised (https://www.ensembl.org/info/docs/tools/vep/index.html) to assess the effect of GWAS-identified top variants on genes, transcripts, and regulatory regions.

2.7.1 Gene prioritization and positional mapping

SNP2GENE function was used to compute LD structure, characterise the risk loci, annotate functions to SNVs, and prioritise candidate genes. For positional mapping, genes within each genomic risk locus were determined based on SNVs that were physically located within a 10 kb distance from the gene.

2.7.2 Expression quantitative trait loci (eQTL) mapping

Expression quantitative trait loci (eQTL) mapping was conducted to explore the associations between GWAS-identified SNVs and changes in gene expression levels. This analysis helps to understand the functional consequences of identified genetic variants and to provide insights into the possible biological mechanisms underlying the observed genetic associations. eQTL data from 2 tissue types, including Genotype-Tissue Expression (GTEx) project v8 Breast, and GTEx v8 Ovary data sources were used for eQTL mapping. Only eQTL values with a false discovery rate (FDR) less than 0.05 were considered significant and used to map SNVs to genes.

2.8 Polygenic risk score (PRS) calculations

The PRS estimates employed in this study incorporated information from 2,174,072 SNVs that are present in both the UK Biobank (https://www.ukbiobank.ac.uk/ (Bycroft et al., 2018)) and Estonian Biobank individuals (https://genomics.ut.ee/en/content/estonian-biobank (Mitt et al., 2017)). These PRSs were developed using data from 428,747 UK Biobank individuals and 105,000 Estonian Genome Centre participants (Orliac et al., 2022). For the calculations conducted in this study, 2,041,044 SNVs were used due to the missingness of the remaining 133,028 variants in our dataset. The PLINK v2.00 function --*score* was used for all PRS calculations.

2.9 Statistical analysis

For the statistical analysis, R v4.0.2 (R Core Team, Vienna, Austria) (R. C. Team, 2020) and RStudio v1.3.1093 (RStudio Team, Boston, MA, USA) (R. Team, 2020) software programs

were used. All statistical tests conducted were two-sided, and *p* values below 0.05 were considered statistically significant.

A variety of statistical techniques and R packages were employed to address specific study objectives. For instance, the Kruskal-Wallis test (base R 'stats' package) was used to assess differences in age distribution among the study groups, followed by post-hoc pairwise comparisons with the Wilcoxon rank-sum test with Bonferroni correction. The prevalence of *BRCA1* and *CHEK2* double heterozygous variants and their association with BC and/or OC were evaluated using a two-tailed Fisher's exact test to determine ORs and their statistical significance. Additionally, the Bioconductor package 'Survival', version 3.2-3 (Therneau, 2020), was used to investigate the impact of the PVs on the cumulative risk of BC and/or OC using Kaplan-Meier estimates, with curve differences assessed using the Log-rank test. For the prediction of cumulative hazard (time-to-event probability), Cox-regression analysis was performed.

The association between PRS and the presence of BC and/or OC in *BRCA1* PV carriers was evaluated by using a binomial logistic regression model. The outcome variable had three categories: 0 (no cancer), 1 (BC), and/or 2 (OC). The model was adjusted for age, age squared, *BRCA1* PV (c.4035del or c.5266dup), and the first two PCs. OR and their 95 % CI were calculated using the R package Epi (Carstensen, 2022). Receiver operating characteristic (ROC) curve analysis was performed to select the most optimal binomial logistic regression analysis model using the R package pROC (Robin et al., 2011).

2.10 Data availability

Summary statistics will be available from https://dataverse.rsu.lv/ repository.

3 Results

3.1 Study group characteristics

3.1.1 Patient characteristics

Our study cohort comprised 452 women who were carriers of one of *BRCA1* PVs – c.4035del or c.5266dup. These women had been diagnosed with either BC, OC, or had no cancer diagnosis at the time of recruitment. Among the study cohort, 196 women (43.4 %) were diagnosed with BC, 129 women (28.5 %) were diagnosed with OC, and 127 women (28.1 %) had no cancer diagnosis, serving as unaffected group for comparison. The mean ages at onset of BC or OC were 46.52 years (range 25–92, SD = 11.71) and 50.62 years (range 27–79, SD = 8.80), respectively. The mean age of the unaffected group was 38.36 years (range 18–73, SD = 11.05). Pairwise comparisons of patient age between different groups, conducted using the Wilcoxon rank sum test with continuity correction, revealed statistically significant $p < 0.01$ for all 3 groups, indicating substantial differences in age between each group. These age differences between BC and OC and unaffected groups were adjusted and standardised when performing subsequent analyses. Key characteristics of the study cohort are summarised in Table 3.1. The specific patient characteristics outlined in the study are crucial for understanding the diversity within the cohort and for drawing meaningful conclusions related to the impact of *BRCA1* PVs on cancer development.

Table 3.1

Study cohort characteristics

	Total	***BRCA1*:c.4035del**	***BRCA1*:c.5266dup**
Study sample	452	173 (38.28 %)	279 (61.72 %)
Breast cancer	196 (43.36 %)	53 (11.73 %)	143 (31.64 %)
Ovarian cancer	129 (28.54 %)	69 (15.27 %)	60 (13.27 %)
Unaffected	127 (28.10 %)	51 (11.28 %)	76 (16.81 %)
Mean age	45.40 ± 11.72	47.67 ± 12.02	43.99 ± 11.35
Breast cancer*	46.52 ± 11.71	49.68 ± 12.56	45.34 ± 11.19
Ovarian cancer*	50.62 ± 8.80	52.00 ± 9.57	49.03 ± 7.58
Unaffected*	38.36 ± 11.05	39.73 ± 10.64	37.45 ± 11.30

* Represents statistically significant (of $p < 0.01$) age difference between all 3 study groups.

Following multi-step QC and comprehensive data cleaning, our dataset was reduced to 406 samples. The final study cohort used for subsequent GWAS and PRS analysis consisted of 171 women (42.1 %) with a BC diagnosis, 121 women (29.8 %) with an OC diagnosis, and 114 women (28.1 %) with no cancer diagnosis. The mean ages at disease onset were 46.67 years (range 25–92) for BC and 50.55 years (range 27–79) OC. The primary characteristics of the final study cohort are presented in Table 3.2.

Table 3.2

Patient characteristics after data QC and the filtering steps

	Total	*BRCA1*:c.4035del	*BRCA1*:c.5266dup
Study sample	406	161 (39.65 %)	245 (60.35 %)
Breast cancer	171 (42.12 %)	49 (12.07 %)	122 (30.05 %)
Ovarian cancer	121 (29.80 %)	64 (15.76 %)	57 (14.04 %)
Unaffected	114 (28.08 %)	48 (11.82 %)	66 (16.26 %)
Mean age	45.38	47.55	43.52
Breast cancer	46.67	49.53	45.52
Ovarian cancer	50.55	51.53	49.44
Unaffected	37.98	40.61	36.28

3.1.2 The penetrance of *BRCA1* PVs c.4035del and c.5266dup in the study cohort

The study cohort was divided into two subgroups based on the specific founder alleles of the *BRCA1* gene: c.4035del and c.5266dup. The overall study population consisted of 173 women carrying the c.4035del PV (53 in the BC group, 69 in the OC group, and 51 in the unaffected group) and 279 women carrying the c.5266dup PV (143 in the BC group, 60 in the OC group and, 76 in the unaffected group), as shown in Table 3.1.

Penetrance, defined as the proportion of individuals carrying specific disease-associated PV who develop the corresponding disease phenotype (Cooper et al., 2013) of either BC or OC, was calculated in this study cohort. The results are presented in Table 3.3. Among the carriers of *BRCA1* c.4035del and c.5266dup PVs, the estimated penetrance in the study cohort was 31 % for BC and 40 % for OC, and 51 % for BC and 22 % for OC, respectively.

Table 3.3

Penetrance of *BRCA1* PVs c.4035del and c.5266dup in the study cohort

BRCA1 PV	Breast cancer (%)	Ovarian cancer (%)
c.4035del	30.64	39.88
c.5266dup	51.25	21.51

PV – pathogenic variant.

3.1.3 Age related cumulative incidence of BC or OC among *BRCA1* c.4035del and c.5266dup PV carriers

We conducted a Cox proportional hazards regression analysis to investigate the relationship between the *BRCA1* PV – c.4035del and c.5266dup – and the time to an event (cancer diagnosis) in our study cohort of 452 individuals. Of these, 325 individuals had cancer diagnosis (196 BC cases and 129 OC cases).

The analysis revealed a significant association between the *BRCA1* PV c.5266dup and the age of cancer onset, with a regression coefficient of 0.3626 ($p = 0.00169$**). The hazard ratio (HR) for the *BRCA1*:c.5266dup variant was estimated to be 1.437 (95 % CI: 1.15–1.80),

indicating that individuals with this variant had a 43.70 % higher risk of cancer development at younger age compared to individuals with other c.4035del variant (Figure 3.1).

The concordance index, a measure of the model's predictive accuracy, was 0.562 (standard error = 0.015), indicating moderate predictive ability.

Additional statistical tests consistently confirmed the significance of this association. The likelihood ratio, Wald, and score (Log Rank) tests all showed significant associations between c.5266dup variant and the cancer occurrence ($p = 0.001$, $p = 0.002$, and $p = 0.002$, respectively). These results suggest that the *BRCA1* variant c.5266dup is a statistically significant predictor of earlier cancer onset.

Figure 3.1 **Cumulative incidence of either BC or OC in *BRCA1* PV carriers**

Red line indicates *BRCA1*:c.4035del variant carriers; the blue line indicates *BRCA1*:c.5266dup variant carriers.

In the subsequent analysis, we explored the impact of the *BRCA1*:c.5266dup variant on BC and OC groups individually, see Figure 3.2. The Cox proportional hazards regression models were applied to each group separately, yielding the following results. For the BC group consisting of 323 individuals, with 196 cancer events, the Cox regression analysis revealed a significant association between the *BRCA1*:c.5266dup variant and the age of cancer onset. consisting of 323 individuals, with 196 cancer events, the Cox regression analysis revealed a significant association between the *BRCA1*:c.5266dup variant and the age of cancer onset.
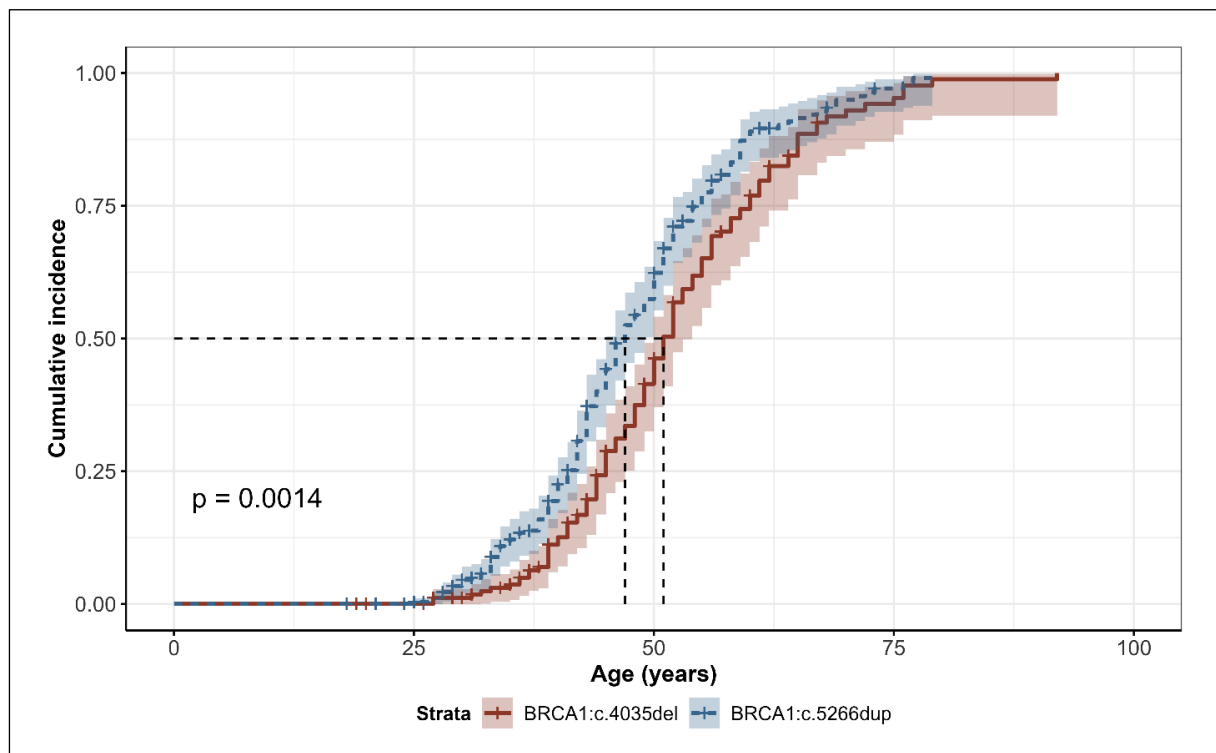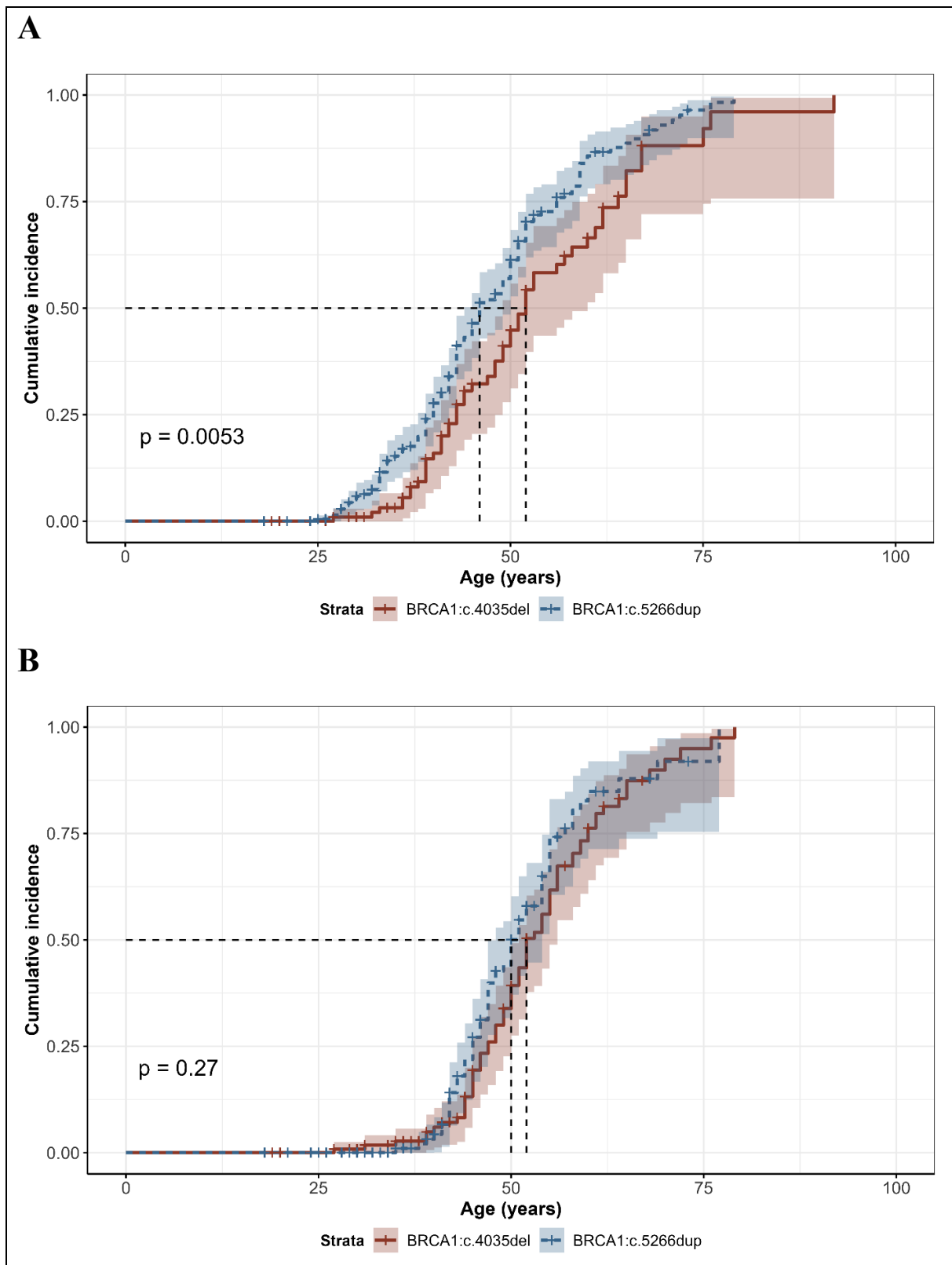
Figure 3.2 **Cumulative incidence of BC or OC in *BRCA1* PV carriers**

Red line indicates *BRCA1*:c.4035del variant carriers; the blue line indicates *BRCA1*:c.5266dup variant carriers.
A) Plot visualizing the cumulative incidence of BC development in *BRCA1* PV carriers;
B) Plot visualizing the cumulative incidence of OC development in *BRCA1* PV carriers.

The HR for the *BRCA1*:c.5266dup variant was 1.564 (95 % CI: 1.14–2.15), indicating a 56.40 % higher hazard of developing BC compared to individuals with another *BRCA1* variant (c.4035del).The statistical tests further confirmed the significance of this association. The likelihood ratio test yielded a *p* value of 0.005, the Wald and Log Rank tests resulted in a *p* value of 0.006, and the concordance index was 0.56.

In the OC group, which included 256 individuals with 129 cancer events, the Cox regression analysis did not show a similar trend. The association between the *BRCA1*:c.5266dup variant and the age of cancer onset in this group was not statistically significant. The HR for the *BRCA1*:c.5266dup variant was 1.2198 (95 % CI: 0.86–1.73), with a *p* = 0.265.

The likelihood ratio test, Wald test, and Log Rank test all produced consistent *p* values around 0.3, indicating that the evidence is not strong enough to conclude that the presence of the c.5266dup variant has a significant impact on the age of OC onset in this particular study. These distinct results suggest that the *BRCA1*:c.5266dup variant plays a significant role in the age of cancer onset within the BC group, but its impact is less evident in the OC group.

3.2 Study design: Hypothesis-driven vs. data-driven analysis

This section provides an overview of the main design framework for the study and categorises it as either hypothesis-driven or data-driven. These two approaches differ significantly in how they define the study objectives and conduct the investigation. A hypothesis-driven study is characterised by the formulation of specific research hypotheses prior to data collection and analysis. Conversely, a data-driven study is distinguished by its exploration of data without a presumptive hypothesis.

Our study employed a hybrid methodology that combined hypothesis-driven and data-driven techniques. This combination allowed us to test specific hypotheses while also exploring unexpected patterns and associations within the dataset.

3.2.1 Hypothesis-driven analysis of *BRCA1* and *CHEK2* double heterozygotes

Here, our primary focus was on a hypothesis-driven analysis, specifically investigating individuals with double heterozygosity for *BRCA1* and *CHEK2* genes, as both genes are involved in the same DNA repair pathway. This analysis was conducted at the outset of the study, using data from 380 individuals.

The studied *CHEK2* variants were discovered in 13 double heterozygous cases (including c.444 + 1G>A, n = 1, c.470T>C, n = 11, del5395, n = 1), as listed in Table 3.4. None of the samples contained more than one simultaneous *CHEK2* variant. To estimate the penetrance of *CHEK2* allelic variants in relation to BC or OC development risk among

BRCA1 PV carriers, we compared the prevalence of these *CHEK2* variants in the BC and OC groups to an unaffected group within the cohort. While the prevalence of *CHEK2* variants was relatively high in the OC group (5.41 %), the increase in OC risk did not reach statistical significance (OR = 1.56; 95 % CI: 0.32–9.94; *p* = 0.73). Additionally, the prevalence of the studied *CHEK2* variants in BC patients did not significantly differ from that in the unaffected group (OR = 0.88; 95 % CI: 0.15–6.15; *p* = 1).

Table 3.4

Frequencies of *CHEK2* variants in the study cohort

Variant and case	No of carriers/total	Frequency (%)
c.444+1G>A		
Unaffected	1/87	1.15
BC cases	0/132	0.00
OC cases	0/111	0.00
c.470T>C		
Unaffected	2/87	2.30
BC cases	3/132	2.27
OC cases	6/111	5.41
del5395		
Unaffected	0/87	0.00
BC cases	1/129	0.78
OC cases	0/109	0.00

BC – breast cancer; OC – ovarian cancer.

The impact of a specific *CHEK2* variant on the age at cancer onset was not consistent. Among carriers of the *BRCA1*:c.4035del variant, the presence of any studied *CHEK2* variant did not significantly alter the median age at the onset of any cancer (*p* > 0.3 by Log-rank test).

In contrast, for carriers of the *BRCA1*:c.5266dup variant with any of the studied *CHEK2* variants, the median age at onset of OC was notably lower, with an 8.5 year difference compared to *BRCA1*:c.5266dup carriers without the *CHEK2* variant. The HR for this effect was 3.93 (95 % CI: 0.93–16.65). Although a Log-rank test indicated a statistically significant difference (*p* = 0.043) and a trend suggested an association between identified *CHEK2* variants and a younger age at OC onset, alternative Cox regression modelling did not yield statistical significance (regression coefficient: 1.37, *p* = 0.064).

3.2.2 Data-driven identification of single level variants associated with cancer risk in *BRCA1* PV carriers

This section transitions to a data-driven analysis, with a primary focus on the identification of genetic variants associated with the risk of BC and OC development in individuals carrying *BRCA1* PVs without predefined hypotheses. The study employed a genome-wide association study (GWAS) approach to identify such variants.

A total of 7,911,505 SNVs were tested for associations with BC or OC development risk in 406 *BRCA1* PV carriers. Our analytical approach included the incorporation of covariates such as age at recruitment/disease onset, relatedness among participants, and the specific type of *BRCA1* PV in the models.

Results of the most significant top SNVs associated with BC or OC development risk are presented in the Manhattan plots in Figure 3.3 and later detailed in following chapter, within Table 3.7 of this manuscript. For genome-wide significance, we employed a stringent significance level of $p < 5 \times 10^{-8}$, while p values ranging from 5×10^{-8} to $\leq 1 \times 10^{-6}$ were considered suggestive of association. The most significant SNVs for a suggestive association with BC development risk was located on chromosomes 3 and 10, with the most significant association for an SNV located on chromosome 10 (see Figure 3.3A). However, in the OC study group, chromosome 20 exhibited the most significant suggestive association, as shown in Figure 3.3B.
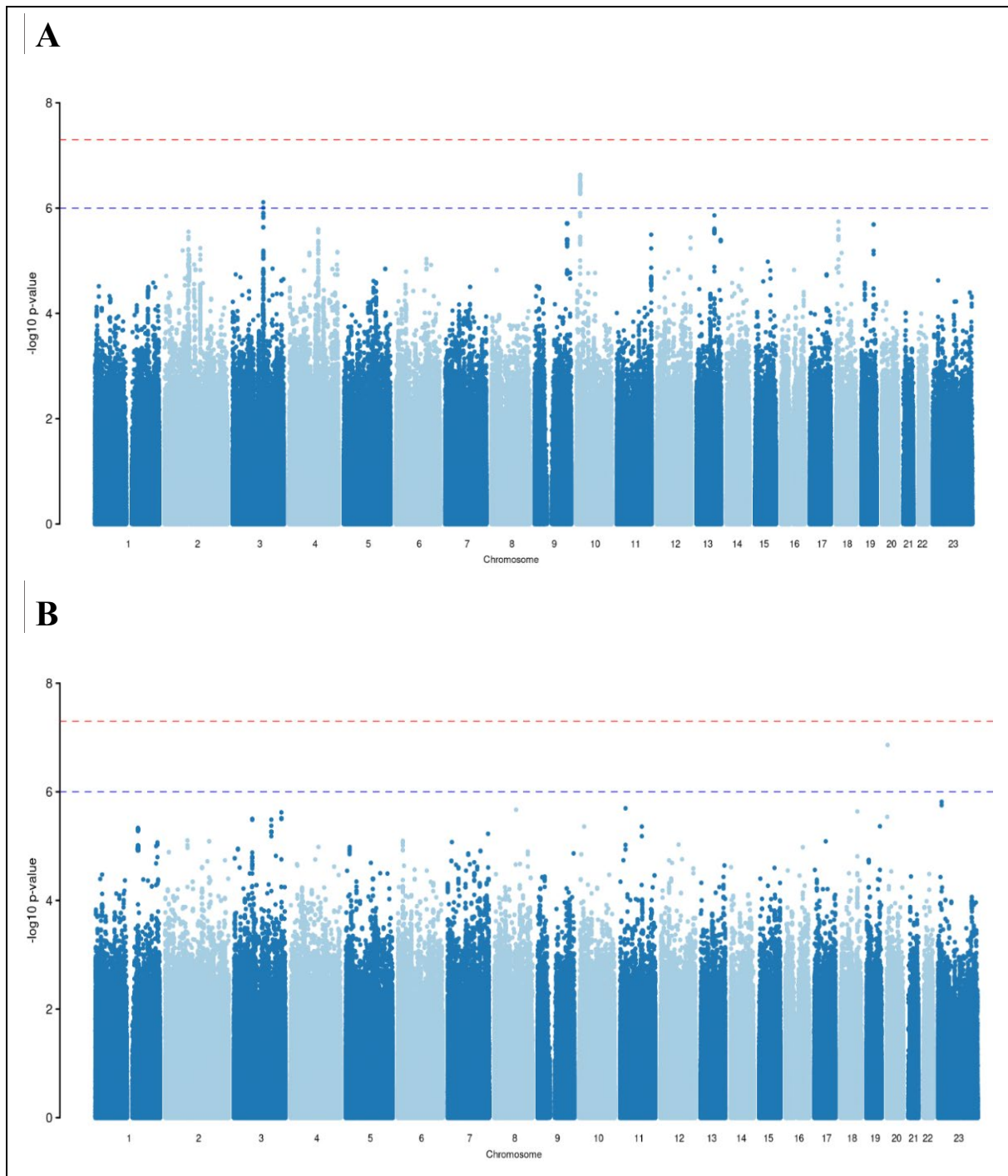
Figure 3.3 **SNV association with BC or OC development risk**

A) Manhattan plot visualizing $-\log_{10} p$ values for SNV associations with BC development risk.
B) Manhattan plot visualizing $-\log_{10} p$ values for SNV associations with OC development risk. The red line denotes genome-wide significance ($p = 5 \times 10^{-8}$); the blue line denotes genome-wide suggestive significance ($p = 1 \times 10^{-6}$); chromosome 23 represents chromosome X.

To estimate potential biases in our dataset-specific analysis, we generated quantile-quantile (Q-Q) plots and estimated genomic factors for both BC and OC groups. In Figure 3.4A and B, we present Q-Q plots, which display the observed p values against the expected p values for associations with BC and OC development risk.
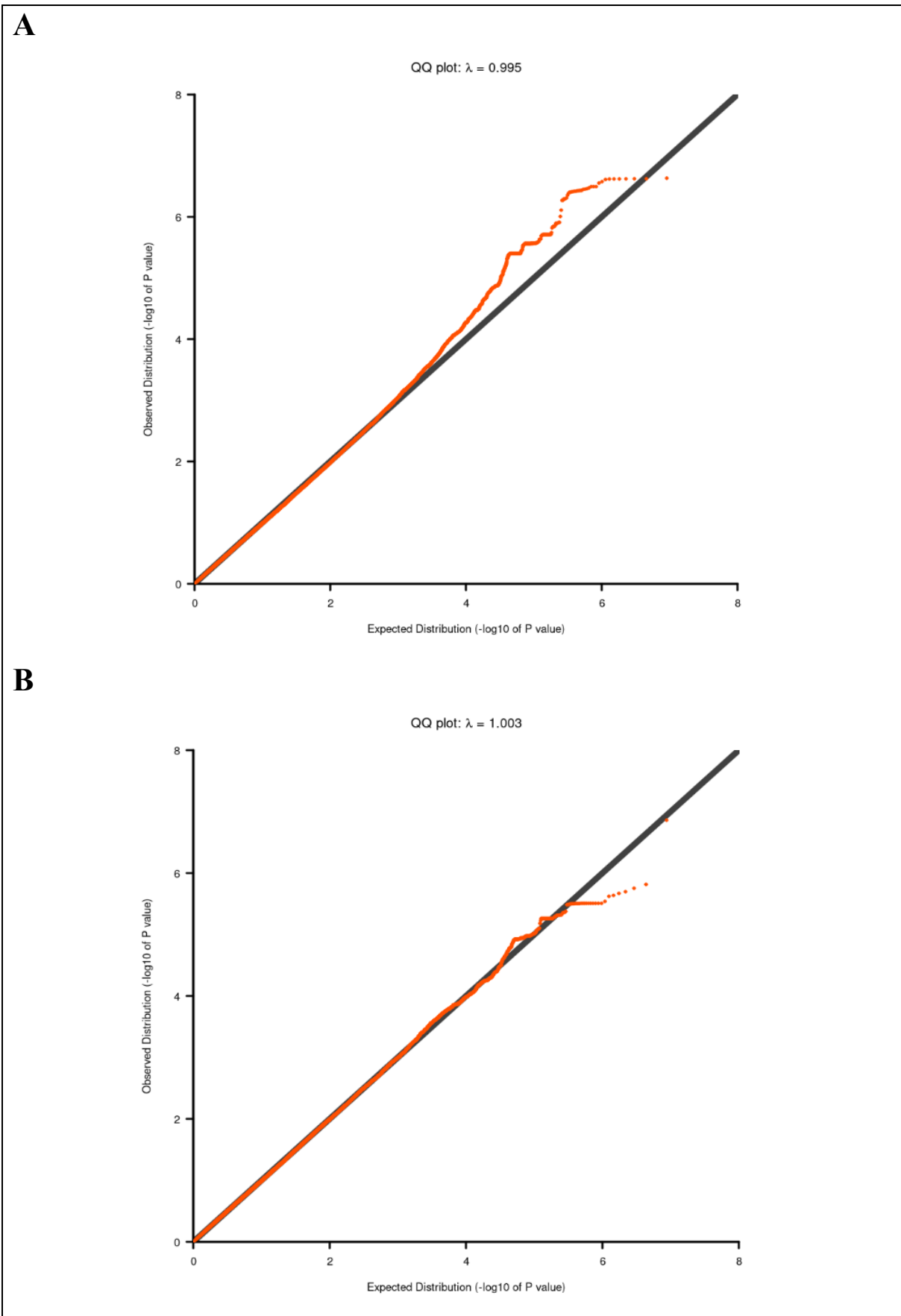
Figure 3.4 **Quantile-quantile (Q-Q) plots for GWAS of the BC or OC development risk**

For A) BC patients and B) OC patients. The black line represents the expected distribution under the null hypothesis of no association; the red line represents observed distribution. λ = lambda (genomic inflation factor).

The calculated inflation factors (λ) for BC and OC were 0.995 and 1.003, respectively. These values indicate that there was no substantial genomic inflation in our analysis.

Identification of single level variants that are associated with BC and OC risk

Table 3.5 presents the main findings of our association analysis in the study cohort. Using the FUMA platform in our post-GWAS analysis, we identified 18 genomic risk loci associated with BC and 21 genomic risk loci associated with OC development risk. These loci include independent SNVs physically close or overlapping within a locus, each represented by the top lead SNV with the lowest p value. We identified 27 independents significant SNVs in the BC group and 25 independents significant SNVs in OC group that reached our predefined genome-wide suggestive significance threshold of $p < 1 \times 10^{-6}$ and were independent from each other at $r^2 < 0.6$. Additionally, from these independent significant SNVs, 19 in BC and 22 in OC group were identified to be lead SNVs that are independent from each other at $r^2 \leq 0.1$. More detailed information about all identified lead SNVs can be found in Supplementary Tables 1 and 2.

Table 3.5

Results of association analysis ($p = 1 \times 10^{-6}$)

No of	BC vs. Unaffected	OC vs. Unaffected
genomic risk loci	18	21
lead SNVs	19	22
independent significant SNVs	27	25
candidate SNVs	1152	633
candidate GWAS tagged SNVs	934	543

SNV – single nucleotide variant; GWAS – genome-wide association analysis.

The number of the candidate SNVs that exhibit LD ($r^2 > 0.6$) with one of the previously mentioned independent significant SNVs is 1152 in the BC group and 633 in the OC group (see Table 3.5). These candidate SNVs include 934 and 543 candidate GWAS tagged SNVs in BC and OC groups as well as non-GWAS tagged variants obtained from the 1000 genomes reference panel. Table 3.6 provides a summary of the distribution of candidate SNVs based on their functional consequences and genomic localization, highlighting that most of these variants are positioned within non-coding regions of the genome.

Table 3.6

Functional consequences of SNVs on genes

	BC vs. Unaffected	**OC vs. Unaffected**
intergenic	690 (60.21 %)	519 (84.39 %)
intronic	120 (10.47 %)	71 (11.54 %)
ncRNA_intronic	274 (23.91 %)	13 (2.11 %)
exonic	8(0.70 %)	4 (0.65 %)
downstream	12 (1.05 %)	3 (0.49 %)
3' UTR	1 (0.09)	2 (0.33 %)
ncRNA_exonic	21 (1.83)	2 (0.33 %)
upstream	20 (1.75)	1 (0.16 %)

ncRNA – non-coding RNA; 3' UTR – a three prime untranslated region.

Top associated variants with BC or OC development risk

Table 3.7 highlights three most significant ($p < 1 \times 10^{-7}$) genetic variants that were associated with the risk of developing BC or OC. All the significant lead SNVs from our study that were suggestive for association with BC or OC development risk ($p < 1 \times 10^{-6}$) are presented in Supplementary Tables 1 and 2.

Table 3.7

Top associated variants with BC or OC development risk

Group	**rsID**	**Chr**	**Position**	**REF**	**ALT**	**MAF**	***p* value**	**Beta**	**SE**	**Nearest gene**
BC	rs2609813	10	14800320	A	G	0.07952	2.33×10^{-7}	−1.26	0.24	*FAM107B*
BC	rs4688094	3	118003477	G	C	0.4523	7.76×10^{-7}	−0.96	0.19	*RP11-384F7.1*
OC	rs79732499	20	3404208	G	T	0.01789	1.38×10^{-7}	−8.09	1.54	*C20orf194*

BC – breast cancer; OC – ovarian cancer; rsID – reference SNV ID number; Chr – chromosome; REF – reference allele; ALT – alternative allele; MAF – minor allele frequency; Beta – multivariate linear regression coefficient; SE – standard error.

Annotation of candidate SNVs to the nearest gene in GWAS is a common practice. The decision to report the nearest gene is often practical, relying on the assumption that the proximity correlates with a higher likelihood of affecting gene's function. However, it's crucial to recognise that the nearest gene may not always be the functional gene influencing the observed association (Watanabe et al., 2017).

These lead variants present valuable candidates for future functional studies, providing a foundation for understanding the complex molecular mechanisms that contribute to the effect on *BRCA1* PV penetrance.

Lead variant rs2609813 and FAM107B gene

The strongest association in BC group was observed for rs2609813 variant, which exhibited the most significant association with BC development risk (beta = −1.26; $p = 2.33 \times 10^{-7}$; risk allele G frequency = 0.08). Detailed information is available in Table 3.7. Imputation efficacy for this variant was average, with DR2 value of 0.95.

As illustrated in regional plot in Figure 3.5A, the lead variant rs2609813 is located on chromosome 10 and it is an intronic variant of the *FAM107B* (*Family with Sequence Similarity 107 Member B*) protein coding gene (ENSG00000065809). Notably, an additional 56 SNVs, exhibiting high LD with the lead variant, were mapped to this intronic region. Based on VEP tool, the variant is predicted to be an intronic variant, as well as the regulatory region variant in enhancer.

Lead variant rs4688094 and lncRNA RP11-384F7.1

The second strongest association with BC development risk was identified for the rs4688094 variant (beta = −0.96; $p = 7.76 \times 10^{-7}$; risk allele C frequency = 0.45) as presented in Table 3.7. The imputation efficacy for this variant was the same as the previous variant, with a DR2 value of 0.95.

As illustrated in the regional plot in Figure 3.5B, the rs4688094 variant is situated on chromosome 3 and is particularly located within the novel lncRNA *RP11-384F7.1* (ENSG00000243276), which exhibits high LD with 295 other SNVs.

Lead variant rs79732499 and C20orf194

The only variant that reached genome-wide suggestive significance of $p < 1.38 \times 10^{-7}$ in the OC group was the lead variant rs79732499. This variant exhibited the lowest p value observed in this study (beta = −8.09; $p = 1.39 \times 10^{-7}$) with a risk allele T frequency of 0.018 (see Table 3.7). The imputation efficacy for this variant was average, with a DR2 value of 0.88.

The lead variant rs79732499 is located on chromosome 20 within an intergenic region. The nearest mapped gene *DNAAF9 (Dynein Axonemal Assembly Factor 9,* previously known as *C20orf194*) is a protein coding gene (ENSG00000088854). Figure 3.5C illustrates that the lead variant rs79732499 is in LD with four SNVs mapped within this gene. Based on VEP tool, the variant is predicted to be an intergenic variant that is located between genes within a regulatory region (enhancer).
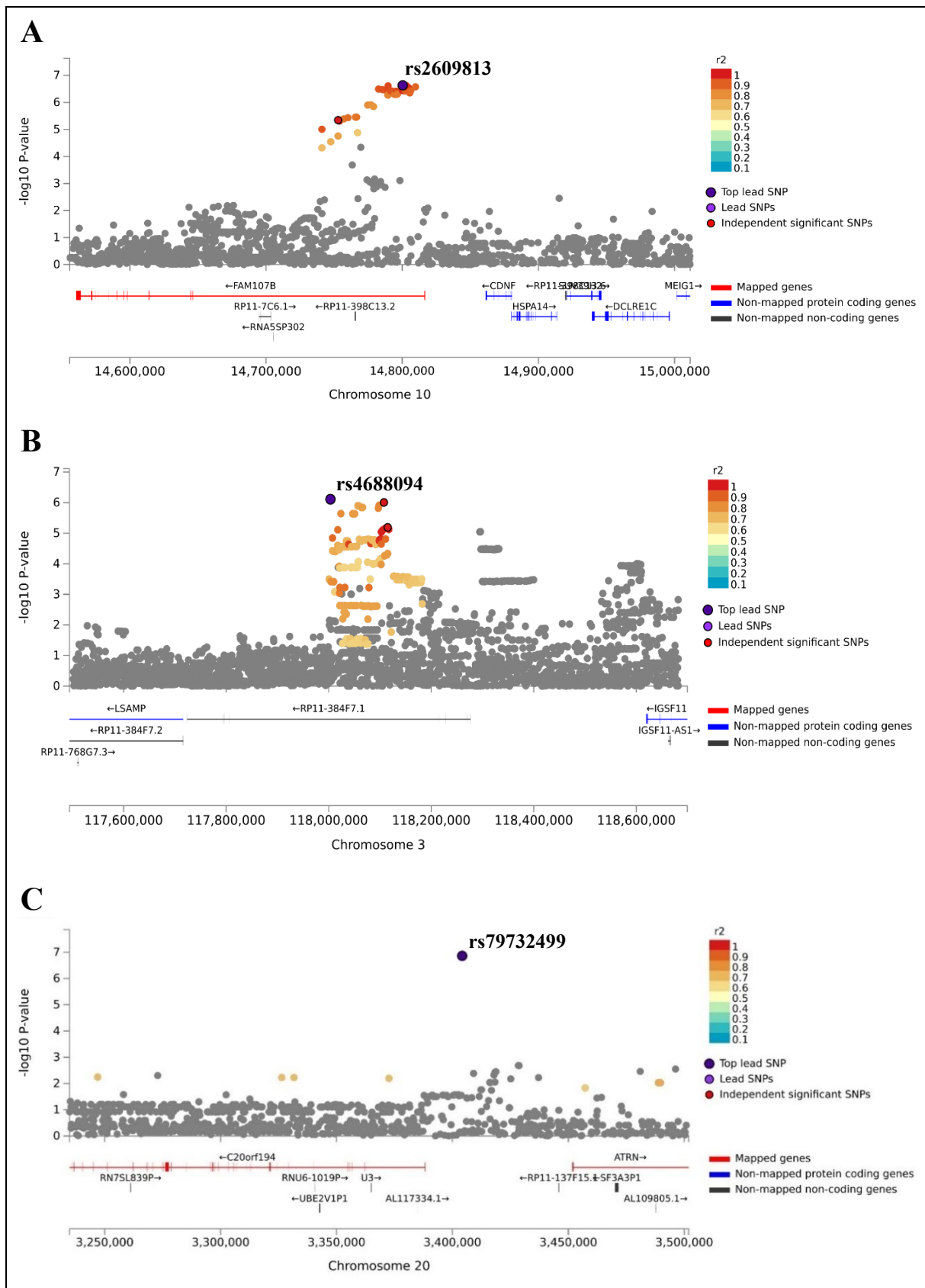
Figure 3.5 **The regional plots of the −log10 *p* values for SNVs at top associated genomic risk loci**

For A) and B) BC patients and C) OC patients. The top lead SNVs with the highest −log10 *p* value is coloured dark blue and identified by its rsID. Colours of other SNVs reflect the level of correlation with the top lead SNV.

The eQTL results in breast tissue

Next, we performed expression quantitative trait loci (eQTL) mapping, focusing on the influence of genetic variants on gene expression using publicly available GTEx breast and ovary tissue data. The GTEx dataset comprised 563 genotyped samples, of which tissue samples from normal breast (n = 396), and ovary (n = 167) were used. The mapping was done in order to highlight potentially functional variants in our dataset, predict target genes and prioritise future experimental validations. Among all candidate SNVs, no significant SNV-gene pairs of cis-eQTL values were found in ovarian tissue by applying a false discovery rate (FDR) threshold of less than 0.05. However, we observed two significant eQTL values in the BC group (see Table 3.8).

Table 3.8

The eQTL results in breast tissue

rsID	Chr	Position	REF	ALT	MAF	*p* value	FDR	NES	Gene
rs10178186	2	95467255	C	T	0.10	3.83×10^{-7}	1.55×10^{-16}	−0.36	*ZNF514*
rs434451	19	47328835	T	C	0.035	2.90×10^{-6}	0.011	−0.42	*SLC1A5*

rsID – reference SNV ID number; Chr – chromosome; REF – reference allele; ALT – alternative allele; MAF – minor allele frequency; FDR – false discovery rate; NES – normalized effect size, is defined as the slope of the linear regression, and is computed as the effect of the alternative allele (ALT) relative to the reference allele (REF) in the human genome reference (i.e. the eQTL effect allele is the ALT allele).

It is crucial to recognise that breast and ovary tissues contain various cell types, and they are not entirely homogeneous. The breast, for instance, consists of several structural components, including epithelial cells, stromal cells, adipocytes, and various connective tissues (Boyd et al., 2010). Similarly, ovaries consist of different cell subpopulation, including oocytes, granulosa cells, stromal cells, endothelial cells, vascular smooth muscle cells, and various immune cell types (Gong et al., 2022). When working with GTEx datasets containing breast or ovary tissue samples, it is essential to consider the heterogeneity of the tissue and potential variations in cell types and structures. Understanding this heterogeneity is crucial for interpreting genetic studies and eQTL mapping results.

Variant rs10178186 and ZNF514 gene expression

The most significant association was determined for the top lead SNV rs10178186 with a raw *p* value of 3.83×10^{-7} and a risk allele T frequency of 0.10 (Table 3.7). The imputation efficacy for this variant was high, with a DR2 value of 0.99.

As depicted in Figure 3.6A, the eQTL is mapped to the protein coding gene *ZNF514* (*Zinc Finger Protein 514*) (ENSG00000144026) on chromosome 2, along with 99 other variants exhibiting high LD with this lead variant.
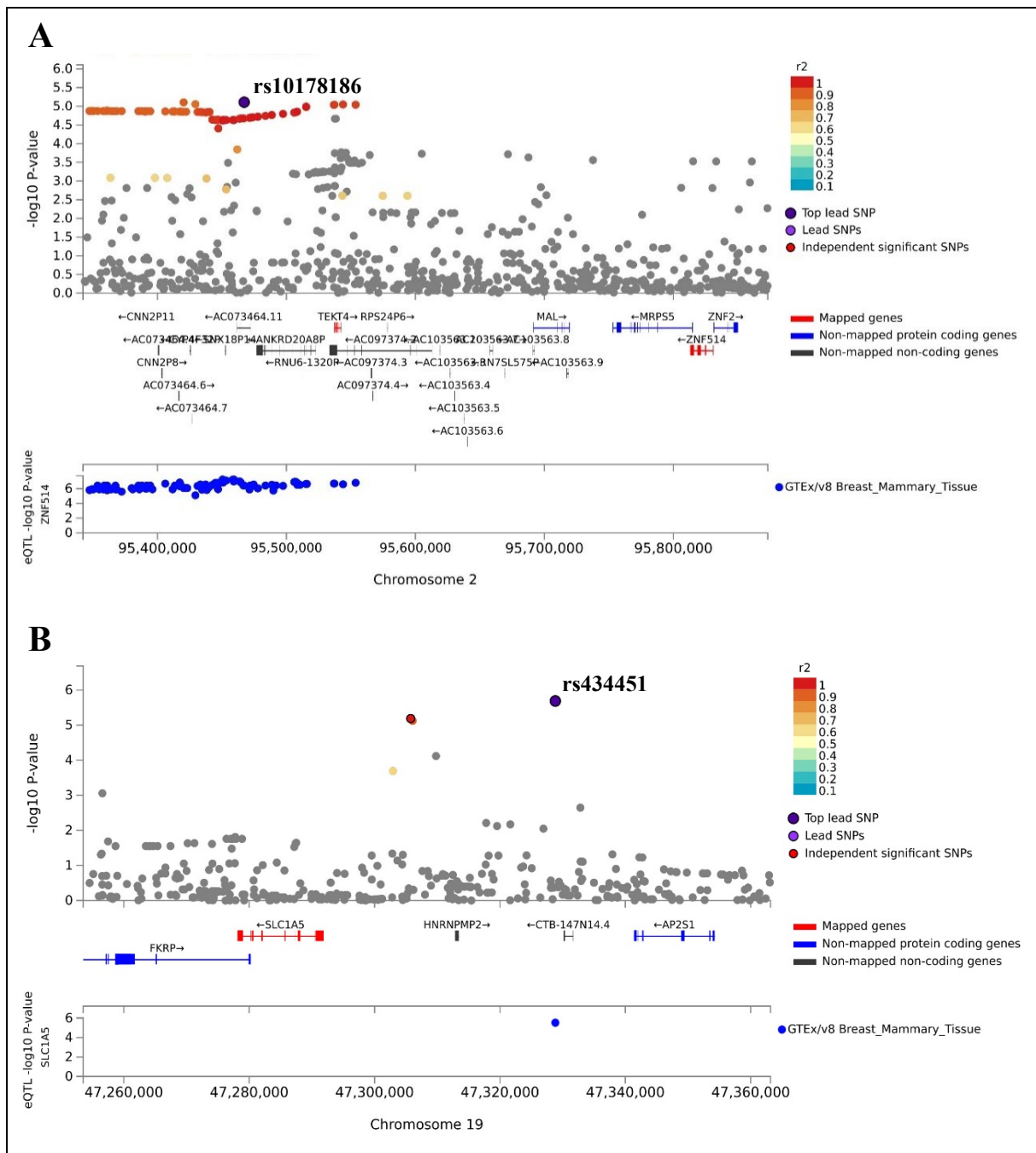
Figure 3.6 **The regional plots of association for the eQTL results in normal breast tissue from GTEx database**

A) Results for top lead SNV rs10178186 in breast cancer patients and B) results for top lead SNV rs434451 in breast cancer patients. The top lead SNVs with the highest −log10 *p* value is coloured dark blue and identified by its rsID. Colours of other SNVs reflect the level of correlation with the top lead SNV.

The normalised effect size of −0.36 indicates a negative association between the rs10178186 variant and *ZNF514* gene expression.

Variant rs434451 and SLC1A5 gene expression

The second significant eQTL association was identified for the top lead SNV rs434451 with a raw *p* value of 2.90×10^{-6} and a risk allele C frequency of 0.96 (refer to Table 3.7), while the imputation efficacy for this variant was average, with a DR2 value of 0.79.

As shown in Figure 3.6A, the lead SNV, intriguingly, was the sole variant mapped to the protein coding gene *SLC1A5* (*Solute Carrier Family 1 Member 5*) (ENSG00000105281) on chromosome 19.

The normalised effect size of −0.42 underscores a negative association between the rs434451 variant and the expression of *SLC1A5* gene.

3.2.3 Data-driven identification of aggregated (PRS) level variants associated with cancer risk in *BRCA1* PV carriers

Exploring diverse joint models for score calculations and key findings

In this study, we used four different PRS joint models, denoted as score1 to score4, to estimate the genetic risk of developing BC or OC in carriers of *BRCA1* PVs. Notably, these PRS models represent a significant advancement as they are the first genome-wide models that encompass over 2,000,000 SNVs, providing comprehensive coverage of the genetic landscape. Further details of each score are provided in Table 3.9.

Table 3.9

Joint model characteristics employed for the risk calculations

Score	Description
score1	The weighted effect size calculated in BC patients with BayesW model
score2	The weighted effect size calculated in BC patients with BayesRR-RC model
score3	The weighted effect size calculated in OC patients with BayesW model
score4	The weighted effect size calculated in OC patients with BayesRR-RC model

BC – breast cancer; OC – ovarian cancer.

We assessed the association of four PRSs (score1–4) with the risk of developing BC or OC using binomial logistic regression analysis. Our goal was to determine the effectiveness of the recently developed PRS models (BayesW vs. BayesRR-RC) in predicting BC and OC risk in *BRCA1* PV carriers in the Latvian population. This was achieved by comparing the PRS weighted effect size in PV carriers with cancer (BC and/or OC) vs. in PV carriers without cancer (unaffected).

As a result, we observed that overall, the average PRSs (score1 and score2) calculated for BC patients were significantly higher in the BC group compared to the average PRS in the unaffected group (see Figure 3.7). This difference between the BC and unaffected groups reached statistical significance, with *p* values of 0.029 for score1 and 0.042 for score2.

However, in the OC group, no statistically significant difference was observed (refer to Figure 3.8, $p > 0.05$).
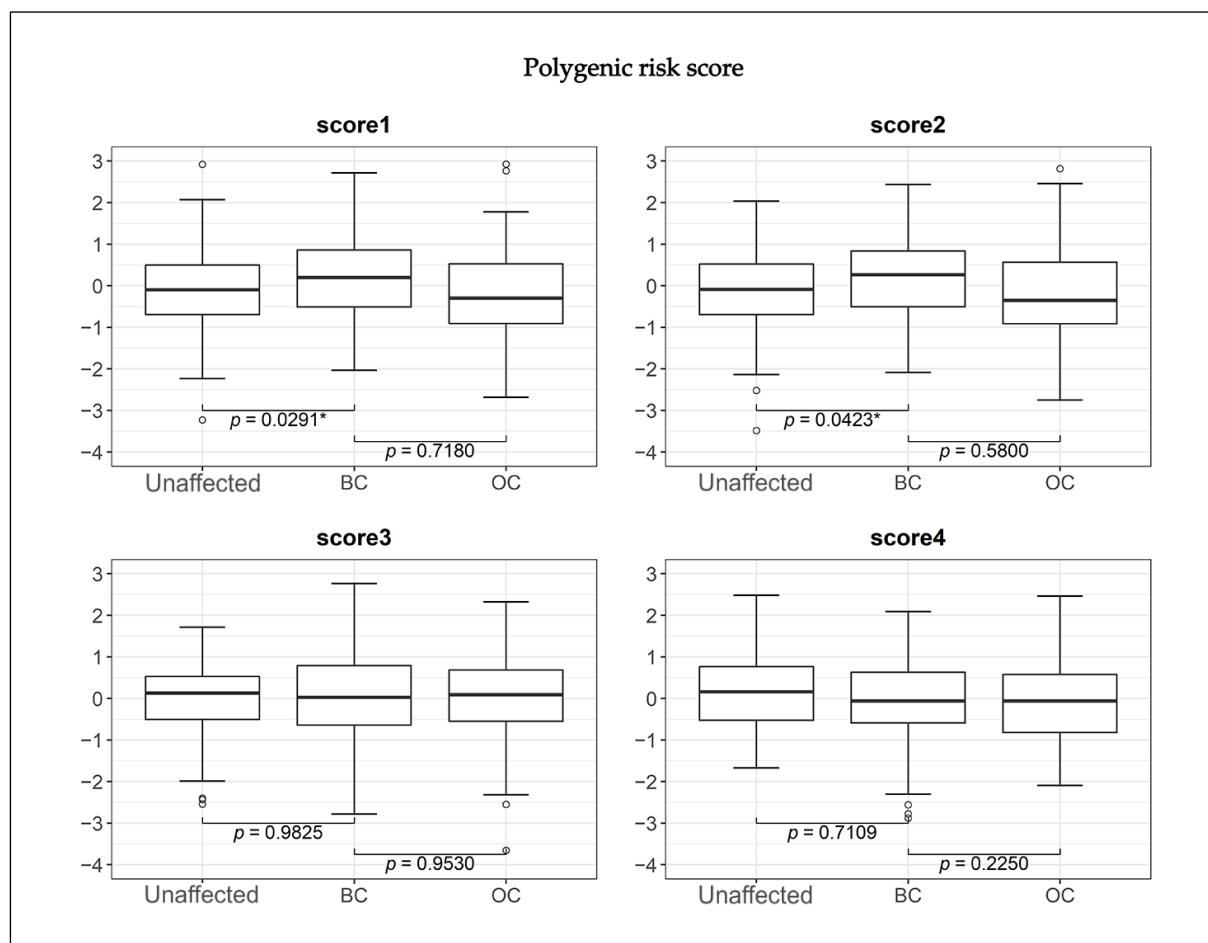
Figure 3.7 **Boxplots and binomial logistic regression analysis *p* values of polygenic risk scores in 406 *BRCA1* PV carriers**

Unaffected – no cancer diagnosis; BC – breast cancer; OC – ovarian cancer. * *p* value below 0.05.

Among the four tested PRSs, it was evident that score1 exhibited the strongest association with the susceptibility to BC. The OR for score1 was 1.37 (95 % CI = 1.03–1.81, $p = 0.0291$) as detailed in Table 3.10. Regardless of the specific PRS employed, none of the models exhibited a statistically significant association with the risk of OC, as presented in Table 3.10.

Table 3.10

Binomial logistic regression analysis results in three different study groups

|	OR	95 % CI	*p* value
BC + OC vs. Unaffected			
score1	1.14	0.89–1.46	0.3119
score2	1.11	0.86–1.42	0.4205
score3	1.00	0.78–1.28	0.9781
score4	0.89	0.69–1.14	0.3514
BRCA1:c.5266dup	1.73	1.03–2.91	0.0375*
BC vs. Unaffected			
score1	1.37	1.03–1.81	0.0291*
score2	1.33	1.01–1.76	0.0423*
score3	1.00	0.76–1.31	0.9825
score4	0.95	0.72–1.25	0.7109
BRCA1:c.5266dup	2.55	1.44–4.53	0.0013**
OC vs. Unaffected			
score1	0.94	0.68–1.31	0.7180
score2	0.91	0.65–1.27	0.5800
score3	0.99	0.71–1.38	0.9530
score4	0.81	0.57–1.14	0.2250
BRCA1:c.5266dup	0.93	0.48–1.79	0.8170

BC – breast cancer; OC – ovarian cancer; BC + OC – both cancers combined; OR – odds ratios; 95 % CI–95 % confidence interval for the associations of PRS with BC and OC risk in *BRCA1* PV carriers. Four different PRS joint models were employed for the risk calculations (see Table 3.9). * *p* value below 0.05; ** *p* value below 0.01.

Next, we conducted an analysis of the area under the receiver operating characteristic curve (AUC) to evaluate the predictive accuracy of three distinct models incorporating various covariates, including the PRS (Figure 3.8). Notably, the model that encompassed age at onset, age squared, *BRCA1* PV status, and the most effective PRS (score1) demonstrated the highest AUC value of 0.7587.

In our comparative analysis of the three models using a bootstrap method, we identified a statistically significant difference ($p = 0.0368$), particularly in the AUC values between the model that included age and age squared as covariates and the model that included age at onset, age squared, *BRCA1* PV status, and the highest performing PRS (score1).
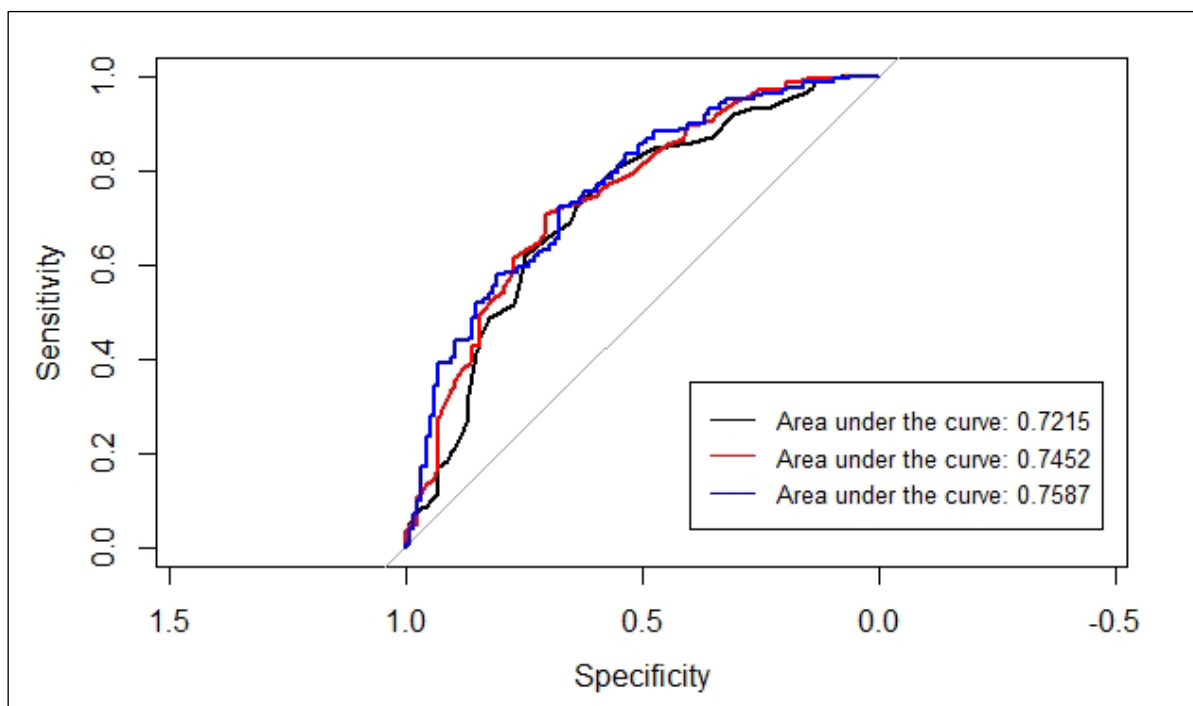
Figure 3.8 **A Comparison of the AUC (area under the receiver operating characteristic curve) to select the most optimal binomial logistic regression analysis model**

In black – the model with only age and age squared as covariates; in red – the model with the *BRCA1* PV added; in blue – the model with the *BRCA1* PV and the best performing PRS added (i.e. score1).

4 Discussion

4.1 Main findings in the study cohort

Our study presents an essential exploration of the association between specific *BRCA1* PVs (c.4035del and c.5266dup) and the development of BC or OC. Table 3.1 displays the distribution of these PVs among the participants in our study, and it is consistent with patterns identified in previous research, confirming their relevance and founder effect within Latvian population (Gardovskis et al., 2005; Tikhomirova et al., 2005). Observed differences in the frequency of these PVs among distinct cancer diagnoses (BC or OC) and an unaffected group (see Table 3.3) suggest unique potential implications of these *BRCA1* PVs in affecting the development of BC or OC. These observations will be discussed in more detail in subsequent chapters.

Our dataset, after stringent QC procedure, created a comprehensive cohort of 406 samples (refer to Table 3.2), allowing an exploration of genetic associations related to *BRCA1* PVs. This analysis investigated their penetrance and the potential impact on the age of onset of BC or OC. Additionally, this dataset provided the foundation for following GWAS and PRS analyses to identify genetic factors affecting the penetrance of region-specific *BRCA1* PVs. This detailed analysis may improve our knowledge of the relationship between specific *BRCA1* PVs (c.4035del and c.5266dup) and the development risk of BC or OC. Such insights might direct further research, personalised risk assessment, and development of focused preventative strategies, contributing to individualised approaches for the management and prevention of cancer.

4.1.1 The penetrance of *BRCA1* PVs c.4035del and c.5266dup in the study cohort

The objective of this study was to investigate the penetrance of distinct *BRCA1* PVs for BC and OC within the study cohort. However, it is important to acknowledge that the penetrance estimates derived from this study may not fully represent the objective lifetime risk associated with these PVs due to the relatively young age range of the unaffected group. Penetrance in known to be age-dependent, with clinical signs appearing more frequently with increasing age, which could lead to potentially higher estimates in older cohorts (Cooper et al., 2013). The estimates calculated in this study might be skewed because the unaffected patients in our cohort were significantly younger than those with BC or OC diagnosis. Nevertheless, despite this limitation, this study offers valuable insights into the penetrance of region-specific *BRCA1* PVs.

As discussed in the literature review, penetrance can vary depending on the specific *BRCA1* PV. Our findings support this concept, demonstrating different penetrance for the *BRCA1* PVs (c.4035del and c.5266dup) in the BC and OC groups. The data presented in

Table 3.3 indicates that the *BRCA1*:c.5266dup PV exhibits higher penetrance in the BC group compared to the OC group, while the *BRCA1*:c.4035del PV demonstrates similar penetrance in both cancer types. This study confirms the previous observations indicating that the penetrance of BC increases with more distal PV locations in the *BRCA1* gene (Gayther et al., 1995; Risch et al., 2001). However, the observed heterogeneity in PV penetrance highlights the importance of further research to identify the source of this variance (Chen & Parmigiani, 2007).

This observation could potentially be explained by the impact of these mutations on the BRCA1 protein. In particular, the *BRCA1* PV c.5266dup, located in exon 19, causes a frameshift and introduces a premature stop codon at position 74 of the new reading frame, which is found within the terminal exon. The resulting mutant transcript is predicted to escape NMD, by likely producing a stable truncated protein lacking the C-terminal BRCT domain (Perrin-Vidoz et al., 2002; Rodriguez et al., 2004). Similarly, a premature stop codon is introduced at position 20 of the new reading frame due to frameshift caused by the *BRCA1* PV c.4035del, located in exon 10. But since truncating mutations within exon 10 are known to undergo NMD, this frameshift variant will result in reduced production of the protein (Perrin-Vidoz et al., 2002).

Consequently, these genetic variants exhibit a genotype–phenotype correlation and differing clinical presentation, potentially arising from their position and subsequent effects on the structural and functional aspects of the mutated BRCA1 protein. Previous research has indicated that PVs positioned towards the 3′ end of the *BRCA1* gene (e.g. c.5266dup) are linked to a higher risk developing BC, while PVs in exon 10 (e.g. c.4035del) present almost equal incidences of BC and OC among PV carriers (Milne & Antoniou, 2016; Plakhins et al., 2011).

In our dataset, the *BRCA1*:c.4035del PV did not show statistically significant evidence of an increased risk for BC development compared to OC, supporting the observation that this specific *BRCA1* PV is associated with relatively balanced risks for both cancer types. This highlights the potential significance of the position of the *BRCA1* PV in the risk assessment (Kuchenbaecker, Hopper, et al., 2017).

4.1.2 Age related cumulative incidence of BC or OC among *BRCA1* c.4035del and c.5266dup PV carriers

Next, we performed a Cox proportional hazards regression analysis of 452 individuals that revealed a statistically significant influence of the *BRCA1*:c.5266dup variant on earlier cancer onset (combining BC and OC groups), compared to the *BRCA1*:c.4035del variant. Notably, *BRCA1*:c.5266dup carriers had a median age of cancer onset at 46.52 years, while *BRCA1*:c.4035del carriers presented at 50.62 years (Table 3.1). The HR of 1.437 indicated

a 43.70 % increased risk of earlier cancer onset among *BRCA1*:c.5266dup carriers, supported by various statistical tests.

When examining the variants individually within the BC and OC groups, the *BRCA1*:c.5266dup variant demonstrated a substantial 56.40 % higher hazard for BC development. This finding is consistent with previous observations in Latvian BC patients in 2011 by Plakhins et al., where study participants presented similar age at onset: 46.51 years for *BRCA1*:c.5266dup carriers and 51.76 years for *BRCA1*:c.4035del carriers (Plakhins et al., 2011). The confirmation of these previous observations in the Latvian population indicates the necessity of personalised approach in genetic counselling about available risk-reducing strategies based on the *BRCA1* PV. Incorporation of *BRCA1* PV into risk management could involve intensified surveillance or potentially risk-reducing bilateral mastectomy.

These results underscore once again the variant-specific effects of *BRCA1* PVs, particularly in driving the development of distinct cancer patterns. While the *BRCA1*:c.5266dup variant significantly influenced earlier cancer onset of BC, this association was not evident in the OC group, highlighting the genotype-phenotype correlation discussed in the previous chapter (Milne & Antoniou, 2016; Plakhins et al., 2011). Several potential reasons may underlie the absence of a significant association in the OC group, including genetic modifying factors and age-related variation. For instance, interactions with other genetic or environmental factors could potentially modify the impact of the *BRCA1*:c.5266dup variant on OC development risk. Additionally, OC is usually diagnosed at an advanced stage that most likely influence observed age of onset, possibly decreasing the effect of *BRCA1*:c.5266dup. In contrast, the impact of this variant on age of onset might be more apparent in BC, where early detection is more common (Thulesius et al., 2004).

4.2 Hypothesis-driven analysis of *BRCA1* and *CHEK2* double heterozygotes

Unfortunately, although risk-reducing strategies such as bilateral mastectomy and salpingo-oophorectomy are proven to be effective and sufficient in preventing BC or OC development, they could reduce the quality of life for the patients. These procedures are related to physiological, sexual, and psychosocial distress, influencing patients' decision-making and post-surgery adaptation (Alves-Nogueira et al., 2023). This aspect highlights the need for further research to evaluate the factors influencing individual *BRCA1* PV penetrance.

Therefore, our study focused on examining the impact of *CHEK2* gene variants on *BRCA1* PV penetrance as *CHEK2* is a cell cycle regulator and it is involved in the same DNA repair pathway as *BRCA1* gene. *CHEK2* variants are frequently observed among BC and OC patients, therefore extensively studied and documented in several European countries like

Poland, the Czech Republic, Belarus, Germany, the Slovak Republic, Finland, Netherlands, and Denmark (Myszka et al., 2011; Narod & Lynch, 2007). However, within the Latvian population, there have been only few studies investigating *CHEK2* variants and their association with cancer risk. In 2006, Irmejs et al. did a pilot study to investigate the potential predisposing effect of the *CHEK2* variant c.470T>C on BC and colorectal cancer development risk (Irmejs et al., 2006). Additionally, a study in 2011 investigated the presence of a large deletion of exons 9 and 10 (del5395) of *CHEK2* gene among different cancer patient groups, including BC and OC (Plonis et al., 2015).

A growing number of studies indicate that additional SNVs in modifier genes frequently impact the penetrance of different gene PVs on an individual's susceptibility to disease. However, there has been limited focus on patients harbouring double heterozygous PVs in both the *CHEK2* and *BRCA1* genes across different populations. Such double heterozygotes are predicted to be extremely rare, as evidenced by previous studies examining thousands of patients, primarily focusing on BC. Based on varying frequencies of *CHEK2* variants in different populations, the identification of *CHEK2* and *BRCA1* double heterozygotes was considerably rare, ranging from 1 to 15 cases per study (Cybulski et al., 2009; Meijers-Heijboer et al., 2002; Sokolenko et al., 2014; Turnbull et al., 2012).

In contrast to previous studies, who have predominantly compared the frequency of *CHEK2* and *BRCA1* double heterozygotes among BC patients with that of healthy controls from the general population, the primary objective of this study was to evaluate the hypothesis that *CHEK2* variants might influence the penetrance of *BRCA1* PVs. This study was designed to assess the presence of *CHEK2* and *BRCA1* double heterozygotes in BC and OC patients compared to an unaffected group without a cancer diagnosis at the time of the recruitment, all consisting of women carrying *BRCA1* PVs.

In total, we identified 13 cases of *CHEK2* and *BRCA1* double heterozygotes, which is consistent with the frequencies observed in other studies, as discussed in the following paragraph. While our findings imply a tendency toward an association between the double heterozygous state of *CHEK2* and *BRCA1* variants and a younger age of onset for OC compared to the heterozygous state for the *BRCA1* PV alone, the data did not produce statistically significant evidence supporting the influence of *CHEK2* variants on the penetrance of *BRCA1* PVs. These observations are consistent with the results of previous studies and indicate that the studied variants of *CHEK2* do not appear to decrease the age of cancer onset in BC or OC patients who are carriers of *BRCA1* PVs (Cybulski et al., 2009; Sokolenko et al., 2014, Sukumar et al., 2021).

According to a previous study by Plonis et al., the frequency of the *CHEK2* del5395 variant (0.3 %) detected in women carrying *BRCA1* PV was lower than that of BC patients (0.68 %) without *BRCA1* PV and an unaffected group (0.76 %) (Plonis et al., 2015). Moreover, in contrast to earlier study by Irmejs et al., where percentages were considerably higher in BC patients (7.60 %) and healthy controls (6.40 %) without identified *BRCA1* PVs, the missense variant *CHEK2*:c.470T>C was less frequently observed in *BRCA1* PV carriers with diagnosed BC (2.27 %) and unaffected group (2.30 %) (Irmejs et al., 2006). Similar trends were observed in the previously cited research from other populations, which consistently demonstrated that BC patients who carried *BRCA1* PV had a significantly lower frequency of *CHEK2* variants, whereas BC patients without *BRCA1* PV had an increased frequency of *CHEK2* variants, suggesting a negative interaction between these variants (Meijers-Heijboer et al., 2002; Cybulski et al., 2009; Turnbull et al., 2012; Sokolenko et al., 2014). This observation suggests a potential interplay between *BRCA1* and *CHEK2* in cancer development and could be explained by a model proposing reduced viability of *CHEK2* and *BRCA1* double heterozygous cancer cells compared to cells with only one variant, as both gene products are involved in the same DNA repair pathway. It has been suggested that inhibiting the CHEK2 protein may cause cell death in malignancies lacking genes involved in the DNA repair pathway, such as *TP53* or *BRCA1*. This is achieved by increasing genomic instability and DNA damage accumulation, ultimately leading to cellular death (Bartek & Lukas, 2003; Collins & Garrett, 2005; Lee et al., 2000). Additionally, the overexpression of *CHEK2* in tumours with germline *BRCA1* PVs have been reported previously and it supports the hypothesis that optimal wild-type *CHEK2* expression is essential for preserving the viability of cancer cells in individuals carrying *BRCA1* PVs (Cybulski et al., 2009; Honrado et al., 2006). Further research is needed to investigate the precise mechanisms underlying this interaction and its implications for BC development in *BRCA1* and *CHEK2* double heterozygotes.

According to our data, no statistically significant evidence has emerged regarding the impact of pathogenic/likely pathogenic *CHEK2* variants on the risk of BC or OC development in carriers of *BRCA1*:c.4035del or *BRCA1*:c.5266dup. The relatively modest sample size within BC and OC subgroups may limit the study's statistical power, therefore, increasing the sample size could enhance the credibility of the findings.

4.3 Data-driven identification of single level variants associated with cancer risk in *BRCA1* PV carriers

To perform a data-driven identification of single level variants associated with BC or OC development risk in *BRCA1* PV carriers, we conducted a GWAS analysis. The objective of this study was to evaluate common genetic variants associated with BC or OC susceptibility as

potential modifiers of cancer development risk in *BRCA1* PV carriers. Due to a relatively small size of the study cohort, the GWAS power was sufficient only for the identification of common genetic variants.

4.3.1 Results of association analysis (Identification of single level variants that are associated with BC or OC risk)

Our study explored the genetic landscape of region-specific *BRCA1* PVs (c.4035del and c.5266dup) and their association with the risk of BC or OC development within a clinical cohort from the Latvian population. By employing the GWAS approach, we identified 18 genomic risk loci associated with BC development risk and 21 risk loci associated with OC development risk. Despite numerous large-scale GWAS conducted both in the general population and among *BRCA1* PV carriers, which have successfully identified over a hundred loci associated with BC and OC development risk, none of the risk loci identified in our study have been previously reported. Furthermore, our cohort did not replicate previous GWAS results (Couch et al., 2013; Kuchenbaecker et al., 2015; Milne & Antoniou, 2016; Milne et al., 2017; Yang et al., 2022).

The absence of previously reported risk loci in our study can likely be explained by our unique study design and possible differences in methodology. Firstly, most previously identified susceptibility SNVs were discovered within the general population (Amos et al., 2017; Jurj et al., 2020; Michailidou et al., 2017; Phelan et al., 2017). However, it has been demonstrated that SNVs commonly identified in the general population may not consistently elevate BC or OC risk in *BRCA1* PV carriers (Coignard et al., 2021). Additionally, most association studies in *BRCA1* PV carriers have used a case-control design, where controls consist of healthy women from the general population without diagnosed *BRCA1* PVs (Milne & Antoniou, 2016). In contrast, our study design specifically focused on *BRCA1* PV carriers, allowing to identify carrier-specific susceptibility SNVs (Coignard et al., 2021). Consequently, our study might not be directly comparable with the results of most studies. Furthermore, while other studies may have focused on broad consortium sample pools with diverse *BRCA1* PVs (Rebbeck et al., 2018), our analysis focused on the region-specific *BRCA1* PVs characteristic of the Latvian population and Baltic region (Gardovskis et al., 2005; Janavičius et al., 2014; Tamboom et al., 2010; Tikhomirova et al., 2005).

After exceeding the genome wide suggestive significance threshold of $p < 1 \times 10^{-6}$, our analysis identified 27 independent significant SNVs in the BC group and 25 in the OC group, suggesting a potential role for these SNVs in cancer susceptibility. Additionally, 19 lead SNVs in BC and 22 in OC were identified, highlighting their impact on the risk of developing BC or OC. Furthermore, the dataset contained a substantial number of candidate SNVs in LD ($r^2 > 0.6$) with the identified independent significant SNVs, resulting in 1152 candidates in the

BC group and 633 candidates in the OC group. Most of these candidates were located in non-coding regions of the genome, suggesting the importance of regulatory regions outside of coding areas in influencing the risk of cancer development and highlighting the need for further in-depth functional exploration. Moreover, a comprehensive examination of global GWAS data has revealed that most common variants associated with cancer susceptibility are found within non-coding regions of the genome and are believed to affect cancer risk through the regulation of certain gene expression (Amos et al., 2017; Edwards et al., 2013; Yang et al., 2022).

In the following chapter, we will explore the most significant GWAS results in detail, offering valuable resources for future research and novel insights into the complex interplay between genetic modifiers of cancer risk and region-specific *BRCA1* PVs in the Latvian population.

Top associated variants with BC or OC development risk

A comprehensive GWAS analysis of 7,911,505 SNVs identified numerous top associated variants with genome-wide suggestive significance ($p < 5 \times 10^{-6}$) that were linked to the risk of BC or OC development. Table 3.7 presents three of the most significant genetic variants associated with BC or OC development risk that will be discussed in the upcoming chapters. Interestingly, all three variants exhibited a negative beta, suggesting a potential protective effect on cancer development. The prevalence of these variants within our study cohort indicates their probable influence in the development of BC or OC and highlights their potential as genetic risk markers.

Lead variant rs2609813 and FAM107B gene

This chapter explores the specific intronic variant rs2609813 of the *FAM107B* gene and its effect on BC development risk, suggesting the potential role in carcinogenesis. A protein coding gene *FAM107B*, a member of the Family with Sequence Similarity 107 (*FAM107*) family of proteins, remains understudied with limited available biological data. Despite this, the N-terminal domain (DUF1151) structure of these gene family members is highly conserved between species and suggests their role in regulating gene transcription. The FAM107B protein appears to affect the rearrangement of the cytoskeleton and plays a role in cell migration and proliferation. However, the molecular mechanisms underlying the biological functions of *FAM107B* remain unclear. In particular, further exploration into the functions and molecular interactions of conserved DUF1151 domain is required to understand its role in signal transduction and gene transcription modulation (Nakajima & Koizumi, 2014).

Previous studies have suggested the *FAM107* gene family as potential candidate tumour suppressor genes. For instance, the downregulation of the *FAM107A* gene, previously known

as *DRR1*, correlates with tumour development and proliferation in various malignancies, including non-small cell lung cancer, renal cell cancer, prostate cancers, and astrocytoma (Liu et al., 2009; van den Boom et al., 2006; Wang et al., 2000). While accumulating information supports *FAM107A* as a candidate tumour suppressor gene, limited biological information is available for *FAM107B* gene.

Nakajima et al. observed decreased expression of *FAM107B* in various tumour tissues, including breast, thyroid, gastric, and colon cancer cells, suggesting its involvement in tumour development and proliferation. Additionally, forced expression of *FAM107B* has demonstrated inhibitory effects on cancer cell proliferation *in vitro* and *in vivo* (Nakajima et al., 2010; Nakajima et al., 2012). Moreover, Guo et al. provided experimental evidence that inhibition of *FAM107B* significantly increases proliferation and migratory ability of gastric cancer cells, supporting the hypothesis that *FAM107B* acts as a tumour suppressor gene (Guo et al., 2017).

Furthermore, *FAM107B* is characterised by a unique promoter region with heat shock transcription factor 1 (*HSF1*)-binding sites, resulting in transcriptional induction following heat-shock or hyperthermia treatment. This distinctive feature has led to its designation as Heat Shock-Inducible Tumour Small protein (HITS) (Nakajima et al., 2010). Decreased expression of HITS has been observed in two prevalent histological types of BC, invasive ductal and lobular carcinomas, compared to normal breast tissue. Correlation analyses with TNM staging revealed an inverse relationship between HITS expression scores and primary tumour size (T-value), suggesting its potential as a marker for tumour progression. Although no correlation with histological grade or tumour differentiation has been observed, further analysis with other pathological parameters of BC indicated elevated HITS expression in aggressive BC phenotypes, characterised by HER2 positive, Ki-67 positive, PR negative, and desmoplastic reaction-positive BC, indicating an increased risk of disease recurrence and shortened survival. However, authors hypothesised that HITS expression influences primary tumour growth during tumour development but does not impact invasion or metastasis (Nakajima et al., 2012). These observations suggest that *FAM107B* may play a role in modulating the aggressiveness of various BC subtypes.

Additionally, massively parallel DNA sequencing of basal-like BC has revealed a recurrent point mutation in the C-terminal region of the *FAM107B* gene, suggesting its potential role in carcinogenesis through the transcriptional regulation of oncogenes or tumour suppressor genes (Ding et al., 2010; Nakajima & Koizumi, 2014).

Given that the identified SNV has a negative effect size with beta coefficient of -1.26 and it has been predicted to be a regulatory region variant, it could be speculated that the variant has potential protective effect by affecting other gene expression in *BRCA1* PV carriers.

However, a comprehensive functional analysis is required to precisely assess the impact of this intronic variant located in the regulatory region.

In conclusion, *FAM107B* emerges as a promising candidate tumour suppressor gene in BC, displaying evidence of its involvement in regulating gene transcription and suppressing cancer cell proliferation (Nakajima et al., 2010; Nakajima et al., 2012). Despite an incomplete understanding of its exact molecular mechanisms and functions in BC, further research should focus on studying these mechanisms and conducting additional functional analyses of this regulatory region variant rs2609813. Such understanding may facilitate the development of targeted therapies and prognostic markers in addressing this heterogeneous disease.

Lead variant rs4688094 and lncRNA RP11-384F7.1

The second most significant SNV suggestively associated with BC development risk was rs4688094 ($p = 7.76 \times 10^{-7}$, OR = 0.38) as presented in Table 3.7. It is located within the novel lncRNA *RP11-384F7.1* and its biological function is unknown. Therefore, it is difficult to predict the functional consequence of this variant.

LncRNAs have emerged as important regulators in cancer development and progression, participating in variety of biological processes, including proliferation, apoptosis, metastasis, and drug resistance (Arun et al., 2018; Liu et al., 2021). Previous studies have associated the dysregulation of certain lncRNAs with different subtypes and clinical outcomes in BC (Su et al., 2014). These findings suggest a potential role for lncRNAs as diagnostic and prognostic biomarkers in BC (Zhao et al., 2021).

The novel lncRNA reported in this study has not been previously associated with BC. The observed negative beta coefficient of −0.96 suggests a potential protective effect associated with the risk allele C, emphasizing the need for further investigation into the functional implications of the rs4688094 variant and its impact on *RP11-384F7.1* expression.

The biological complexity of lncRNAs is a challenge when assessing the exact impact of mutations on their expression (Zhao et al., 2021). Regulatory networks involving lncRNAs in cancer, including BC, are still not fully understood. In particular, the dysregulation of lncRNAs has been linked to a variety of cancer related characteristics, acting as both oncogenes and tumour suppressors (Fonseca-Montaño et al., 2023). LncRNAs can regulate other gene expression at various levels, including chromatin modification, as well as transcription and post-transcriptional processing of RNA. The diversity of lncRNAs in modulating cancer signalling pathways underscores their potential as therapeutic targets (Gutschner & Diederichs, 2012).

In conclusion, the identification of the rs4688094 variant within the lncRNA *RP11-384F7.1* locus highlights the potential protective effect of lncRNAs in the development

of BC. More research is needed to fully understand the functional implications of this variant, its impact on *RP11-384F7.1* expression, and the underlying molecular mechanisms that link lncRNAs to BC development risk.

Lead variant rs79732499 and C20orf194

The lead variant surpassing the genome-wide suggestive significance threshold ($p < 1 \times 10^{-6}$) within the OC group was rs79732499*,* located in an intergenic regulatory region. Notably, this variant demonstrates high LD with several other variants within the *C20orf194* gene, also known as *DNAAF9*, suggesting a potential impact on *DNAAF9*.

DNAAF9, an uncharacterised protein coding gene localised on chromosome 20p13, has limited available information about its functions, but current knowledge suggests its interaction with microtubules and its function in tubulin assembly and cytokinesis (Casalou et al., 2020). According to UniProt database, *DNAAF9* may function as an effector for *ARL3* (*ADP Ribosylation Factor Like GTPase 3*). While the functional role of *ARL3* in cancer remains unknown, observations in glioma indicate that *ARL3* plays a role in angiogenesis and immune cell infiltration in the tumour microenvironment (Casalou et al., 2020; Wang et al., 2019).

Furthermore, the identified SNV has a negative effect size with beta coefficient of −8.09, suggesting a potential protective effect in *BRCA1* PV carriers. However, the precise function of *DNAAF9* in cancer is not well understood, and limited information is available regarding its potential role in cancer development or progression. Subsequent investigation is necessary to gain a better understanding of the function of *DNAAF9* and its potential downstream implications in OC.

To increase the reliability of our findings, it is essential to validate them in an independent cohort, such as EstBB. The inclusion of an independent cohort would not only help assess the robustness and reproducibility of the identified potential associations but also provide additional evidence supporting the plausibility of our findings.

The eQTL results in breast tissue

Next, we investigated the effect of genetic variants on gene expression in breast and ovary tissue using eQTL mapping with the GTEx breast and ovary tissue dataset. While no significant SNV-gene associations were observed in ovarian tissue analysis, the study identified two significant genetic variants affecting gene expression in breast tissue. Table 3.8 summarises the results of eQTL mapping, highlighting the significance of the rs10178186 variant on chromosome 2 and the rs434451 variant on chromosome 19. These two variants demonstrated significant eQTL values associated with the *ZNF514* and *SLC1A5* genes, respectively. Their

statistical significance suggests a potential connection to alterations in gene expression in breast tissue, underlining their importance for further exploration in BC research.

Variant rs10178186 and ZNF514 gene expression

The first identified eQTL variant (rs10178186), located on chromosome 2, is associated with reduced *ZNF514* expression. According to information available in UniProt database (https://www.uniprot.org/uniprotkb/Q96K75/entry#function), Zinc finger protein 514 is predicted to be active in the nucleus and involved in regulation of transcription by RNA polymerase II (Gene Ontology annotation GO:0006357).

Zinc finger proteins (*ZNF*s) constitute the largest family of transcription factors in the human genome, playing a mechanistic role in the development of many cancers. Despite their large number, most of the *ZNF*s are not well studied (Luo et al., 2018). Numerous studies suggest that *ZNF*s play a crucial role in carcinogenesis, cancer progression, and metastasis across various cancers. Some *ZNF*s are known to recruit transcriptional co-repressors or act as transcriptional activators, influencing the regulation of multiple downstream genes (Ye et al., 2021).

While prior studies have implicated the role of certain *ZFN*s in cancer development, additional functional studies are needed to fully understand their potential role and impact on BC development. For instance, *ZNF165* has been identified as modulator of gene transcription, associated with the promotion of TNBC cell development. Elevated *ZNF165* mRNA expression has been correlated with decreased BC patient survival, indicating a potential impact on more aggressive carcinogenesis (Gibbs et al., 2020). In other study, hypermethylated *ZNF154* promoter was observed in numerous tumour cell lines, and the silenced gene was intriguingly associated with increased survival rates in resectable pancreatic cancer (Wiesmueller et al., 2019).

Our GWAS results indicate a negative association between the effect size of the identified variant and the risk of BC development, implying a protective effect of reduced *ZNF514* expression in BC carcinogenesis. However, a comprehensive understanding of the biological mechanisms underlying this observation requires further in-depth research. Additionally, the prognostic value of *ZNF*s in BC has yet to be systematically approached. Given the limited research on the roles of *ZNF*s in BC onset and development, further investigation on their biological functions is essential for the interpretation of our findings.

Variant rs434451 and SLC1A5 gene expression

The second eQTL variant (rs434451), located on chromosome 19, is associated with reduction in *SLC1A5* expression. *SLC1A5* encodes a cell surface solute-carrying transporter crucial for maintaining the uptake of neutral amino acids, including glutamine. While glutamine

is a non-essential amino acid in normal cells, its demand rapidly increases during malignant transformation to support increased metabolic demands of tumour cells (Alfarsi et al., 2021; van Geldermalsen et al., 2016).

The observed negative association between the rs434451 variant and *SLC1A5* expression in *BRCA1* PV carriers suggests a potential role in modulating glutamine metabolism and, consequently, tumour growth in BC. Geldermalsen et al. investigated pharmacological inhibitors of *SLC1A5*-mediated transport and observed significant reduction in glutamine uptake in human BC cell lines. This reduction led to decreased mTORC1 signalling, cellular proliferation, and induced cell death. Importantly, these effects were subtype-specific, underscoring the crucial role of *SLC1A5* transport in TNBC compared to luminal BC cells (van Geldermalsen et al., 2016). The subtype-dependent effects observed highlight the potential therapeutic implications of targeting *SLC1A5* in specific BC subtypes.

Other studies have indicated that *SLC1A5* expression is associated with endocrine therapy sensitivity in luminal BC, suggesting its potential utility as a predictive marker for treatment response (Alfarsi et al., 2021). Given the crucial roles of *SLC1A5*, *SLC3A2*, and *SLC7A5* in cancer cell metabolism, growth, and proliferation, pharmacological targeting of these transporters has been explored to block cancer cell growth and survival. While there are currently no clinical trials testing *SLC1A5* inhibitors, a few inhibitors have shown promising results in preclinical studies (Nachef et al., 2021).

However, further studies are needed to fully understand the potential importance of *SLC1A5* expression in BC development risk for *BRCA1* PV carriers. More in-depth investigations into the precise mechanisms underlying its role in tumour growth, progression, and response to therapy are needed.

4.4 Data-driven identification of aggregated (PRS) level variants associated with cancer risk in *BRCA1* PV carriers

In this study, we investigated the association between two recently reported novel genome-wise PRSs (Orliac et al., 2022), containing 2,174,072 SNVs, with the risk of BC and OC in *BRCA1* PV carriers. While the best approach to select the SNV set and to determine their weights to generate the most effective PRS remains uncertain, our hypothesis focused on the joint estimation of the effects of genome-wise SNVs in the PRS models. Our goal was to increase prediction accuracy compared to commonly used approaches for PRS development (Dareng et al., 2022).

Since the majority of PRSs, including those under evaluation in this research, are derived from cohorts within the general population, it is important to carefully review and validate their performance, particularly in individuals carrying *BRCA1* PVs (Jones et al., 2017;

Mavaddat et al., 2019; Michailidou et al., 2017). The variable penetrance of germline PVs in the *BRCA1* gene poses a significant challenge in estimating the likelihood, age, and site of cancer onset for each individual. As a result, it is important to explore effective strategies for initiating prophylactic screening and clinical management in high-risk women (Chen et al., 2020; Downs et al., 2019). PRS has the potential in stratifying individuals based on their disease risk (Mars, Widén, et al., 2020). However, to achieve this goal and integrate PRSs into clinical practice, it is essential to identify the most optimal set of SNVs that contribute to the best performing PRS.

4.4.1 Exploring diverse joint models for PRS calculations and key findings

The results of this study demonstrate the effectiveness of the best fitting BayesW PRS model in accurately predicting an individual's susceptibility to developing BC. This improves the understanding about polygenic contribution on the manifestation of BC phenotype in individuals carrying germline *BRCA1* PVs (Barnes et al., 2020; Kuchenbaecker, McGuffog, et al., 2017). While the BayesRR-RC PRS model performed well in predicting the risk of developing BC, the BayesW PRS model remained superior (see Table 3.10).

Previous studies have shown a limited focus on evaluating PRS in individuals carrying *BRCA1* PVs. Notably, Kuchenbaecker et al. conducted a study where they developed three PRSs for overall BC, ER-positive and ER-negative BC, as well as one for OC patients. Their research involved data from 15,252 female *BRCA1* PV carriers (7797 females with BC diagnosis, and 2462 females with OC diagnosis), revealing strong associations between the PRS and the risk of both BC and OC. Particularly, the PRS for ER-negative BC exhibited the strongest association with the BC risk (HR = 1.27, 95 % CI = 1.23–1.31, $p = 8.2 \times 10^{-53}$) (Kuchenbaecker, McGuffog, et al., 2017).

Similar findings were replicated by Barnes et al. in a study that included 9473 female *BRCA1* PV carriers with diagnosed BC (Barnes et al., 2020), highlighting that the ER-negative PRS demonstrated the strongest association with BC risk in *BRCA1* PV carriers (HR = 1.29, 95 % CI = 1.25–1.33, $p = 3 \times 10^{-72}$). Considering that ER-negative BC is the predominant tumour subtype in *BRCA1* PV carriers (Foulkes et al., 2004), these studies highlight the strong association of BC subtype-specific PRS with the risk of BC development. This underscores that the most accurate prediction of BC development risk involved integrating comprehensive clinical data into the analysis (Barnes et al., 2020; Kuchenbaecker, McGuffog, et al., 2017). Unfortunately, because of insufficient clinical data, our study was unable to incorporate the information regarding ER status. The available information on ER status was only accessible for a small fraction (< 80) of BC patients.

Other study by Mavaddat et al. demonstrated a strong association between PRS and the overall risk of developing BC in the general population (OR = 1.61, 95 % CI = 1.57–1.65, with AUC = 0.630, 95 % CI = 0.628–0.651) (Mavaddat et al., 2019). Our results are consistent with previous research, indicating that the calculated OR for BC in individuals with *BRCA1* PVs are lower than previously published estimates in the general population. This suggests the existence of a potential subset of SNVs within the PRS that might not combine multiplicatively with the status of *BRCA1* PVs. However, it is essential to acknowledge that potential limitations to direct comparisons may arise from variations in study designs and sample sizes (Kuchenbaecker, McGuffog, et al., 2017).

Our study did not identify any statistically significant association with OC, in contrast to previous studies that have consistently indicated a substantial association between PRS and the risk of OC. One such study was conducted by Barnes et al., which demonstrated a strong association between their high-grade serous PRS and OC development risk (HR = 1.32, 95 % CI = 1.25–1.40, $p = 3 \times 10^{-22}$) (Barnes et al., 2020; Kuchenbaecker, McGuffog, et al., 2017). We observed that the genome-wise PRS was more effective in predicting the risk of developing BC than OC in *BRCA1* PV carriers (OR = 1.37, 95 % CI = 1.03–1.81, $p = 0.029$ for BC vs. OR = 0.99, 95 % CI = 0.71–1.38, $p = 0.95$ for OC). The observed results might be influenced by the limited sample size of 121 *BRCA1* PV carriers diagnosed with OC in our study cohort.

4.5   Strengths and weaknesses of the study

4.5.1   Strengths of the study

As of November 2023, the NCBI ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/) contained 3,264 germline *BRCA1* PV records (including deletions, duplications, indels, insertions, and SNVs, all < 50 bp). However, our study focused on a genetically homogenous cohort consisting of women carrying one of two region-specific *BRCA1* PVs (c.4035del or c.5266dup). Moreover, most existing penetrance estimates are derived from large-scale studies combining data from multiple populations. These studies often include a wide range of *BRCA1* PVs, potentially overlooking the penetrance specific to founder PVs within the distinct populations. Furthermore, population-specific genetic structures can influence the study outcomes, as certain SNVs of modifier genes may be more prevalent in one population while rare in another (Narod, 2002). This statement is supported by Pankratov et al., who have studied differences in local population history and suggested that it's highly region-specific and highlighted the importance of considering local genetic structure in association analysis (Pankratov et al., 2020). Therefore, studies in founder populations are beneficial for evaluating region-specific penetrance of *BRCA1* PVs and identifying modifying genetic factors.

Additionally, while the genome-wise PRSs used in our study were initially developed within a population-based framework using data from the UK Biobank and Estonian Genome Centre participants (Orliac et al., 2022), our results represent an independent evaluation of these PRSs specifically within the subset of region-specific *BRCA1* PV carriers from the Latvian population. We believe that these genome-wise PRSs have the potential to provide equivalent, if not superior, predictive capabilities compared to previously developed PRSs.

4.5.2 Limitations of the study

In our study, we encountered several limitations, including the small number of patients harbouring both the *CHEK2* and *BRCA1* double heterozygous genotype. It can be mostly attributed to the limited cohort size, which is composed of women carrying the two most prevalent region-specific *BRCA1* PVs in Latvia. Also, individuals with *CHEK2* and *BRCA1* double heterozygosity are exceptionally rare (Cybulski et al., 2009). To verify the hypothesis that these double heterozygotes might exhibit higher risk of developing BC or OC, a larger study cohort is essential. This could potentially be achieved through analysing samples from consortiums or larger collective studies. More samples of these double heterozygotes may be identified through growing application of whole exome sequencing (WES) or WGS. Moreover, conducting a meta-analysis could help to overcome the limitations encountered in this study.

This research has several limitations that should be considered, as they could have influenced the results obtained. In particular, the number of women with BC or OC who also carry germline *BRCA1* PV was relatively small in this study. Furthermore, the study cohort may not accurately represent the general population of *BRCA1* PV carriers since the samples were obtained during diagnostic germline variant testing in a clinical setting, potentially introducing selection biases.

Although previous studies in our region have indicated that the tested *BRCA1* PVs, specifically c.4035del and c.5266dup, account for approximately 80 % of identified *BRCA1* PVs (Gardovskis et al., 2005; Janavičius et al., 2014; Jürgens et al., 2022; Tamboom et al., 2010; Tikhomirova et al., 2005), it is important to acknowledge that this study exclusively focused on these two region-specific variants. We did not investigate individuals with additional *BRCA1* PVs that could be relevant to the development of BC or OC. Consequently, this approach may result in an incomplete understanding of the genetic landscape and might not fully capture the entire population of *BRCA1* PV carriers.

Another aspect to mention is that our analysis included 2,041,044 SNVs out of the 2,174,072 SNVs that were integrated into the PRS joint model. The absence of the remaining 133,028 SNVs in our dataset was due to their missingness. Factors such as DNA

sample quality or the specific microarray used might have impacted the availability of these SNVs. Moreover, the missing SNVs could be attributed to imputation quality as well. Although we utilised a genetically similar reference panel derived from WGS data of 2244 Estonian biobank participants (Mitt et al., 2017), it is important to acknowledge that possible genetic differences exist within the Latvian population. These variances could potentially influence the performance of the PRS. As the potential future improvement, enhancing imputation accuracy could involve incorporating a more population-specific reference panel from the Genome Database of Latvian Population (LGDB) once the relevant WGS data are obtained and becomes available for research purposes (Rovite et al., 2018).

Another limitation was the significant age difference among the three analysed groups in our study cohort. While these age differences between BC, OC and unaffected groups were adjusted and standardised during subsequent analyses, they may have still influenced the penetrance estimates, leading to a potentially inaccurate representation of the true lifetime risk associated with the studied *BRCA1* PVs. This limitation could be addressed in future studies by selecting the unaffected group from LGDB (Rovite et al., 2018), which could provide a more age-matched study cohort.

A significant limitation in our study was the absence of detailed clinical information on specific tumour phenotypes in a considerable portion of our patient data. Consequently, our results are average estimations across all BC and OC phenotypes.

Currently, many SNVs exhibit associations with either other SNVs or specific genetic regions. This interplay creates a challenge in precisely characterizing the isolated influence of an identified SNV on a particular phenotype. Furthermore, related SNVs can also potentially impact the trait through mechanisms such as LD and other associated factors. GWAS-identified tag SNVs may not always represent actual risk variants, highlighting the critical need to evaluate these results with great scrutiny and extensive post-GWAS analysis (Yang et al., 2022). As a result, the interpretation of the specific impact of an individual SNV on a particular phenotype becomes more complicated.

Additionally, given that most of the identified associated variants are in the non-coding genome, it has been challenging to link GWAS results to plausible candidate genes and in-depth functional studies are needed (Milne & Antoniou, 2016).

4.6   Future perspectives

Understanding how the identified genetic variants impact the penetrance of specific *BRCA1* PVs (c.4035del and c.5266dup) is critical for more precise risk assessment and

the development of potential prophylactic and therapeutic strategies for individuals carrying these *BRCA1* PVs (Mars, Widén, et al., 2020).

One future direction could involve longitudinal studies, where participants are recruited at a younger age and observed over an extended period of time. Additionally, incorporating more comprehensive data of modifiable risk factors, including smoking status, alcohol consumption, physical activity, and dietary habits, would enhance risk assessment (Milne & Antoniou, 2016). Another perspective could be the deployment of different technology, such as WGS, which has the potential to reveal additional variants that have not been covered by microarray technology. The LGDB is a promising initiative for studying *BRCA1* PVs in the Latvian population (Rovite et al., 2018), as well as exploring other genetic factors influencing their penetrance, including PRSs. This initiative, using WGS data, increases the likelihood of discovering more clinically significant variants.

However, further validation using a larger study group consisting of region-specific *BRCA1* PV carriers is necessary, and our study can serve as preliminary data for a more extensive comparison of all available PRSs. It is important to highlight that the risks of subsequent secondary malignancies were not considered in our analysis, but instead, it focused solely on the first occurrence of BC or OC. In future perspective, exploring whether the tested PRSs also contribute to the prediction of subsequent secondary cancers among *BRCA1* PV carriers would be beneficial.

Conclusions

1. Based on this study, none of the tested *CHEK2* variants demonstrate a significant influence on the penetrance of *BRCA1* pathogenic variants (c.4035del and c.5266dup).

2. Among breast cancer patients, the intronic variant rs2609813 in the *FAM107B* gene exhibits the most significant association with *BRCA1* pathogenic variant penetrance in this study ($p = 2.33 \times 10^{-7}$, OR = 0.28).

3. Among ovarian cancer patients, the variant rs79732499, located in the non-coding regulatory region of the genome, exhibits the most significant association with *BRCA1* pathogenic variant penetrance in this study ($p = 1.38 \times 10^{-7}$, OR = 0.00031).

4. Among the genome-wise PRSs tested in this study, the BayesW PRS model contributes to assessing the risk of breast cancer development for germline *BRCA1* pathogenic variant (c.4035del or c.5266dup) carriers, and it may improve patient stratification and decision-making regarding breast cancer treatment and prevention strategies for female carriers of *BRCA1* pathogenic variant.

Proposals

As the LGDB initiative is evolving and the number of Latvian donors increases, we propose the potential implementation of a genotype-first approach for systematic screening of region-specific *BRCA1* PVs in our population. This could be a progressive step toward a more personalised and effective healthcare system, that has been inspired by several successful implementations in other global biobank projects, such as in Estonia or Australia (Leitsalu et al., 2021; Rowley et al., 2019). The genotype-first approach offers an innovative way to identify individuals carrying clinically significant PVs in high-penetrance *BRCA1* gene, regardless of their family history or medical indication. Additionally, it allows for cancer risk stratification based on their PRSs.

The primary objective of this proposal is to enhance risk stratification and long-term outcomes for region-specific *BRCA1* PV carriers in the Latvian population who may be unaware of their genetic predisposition to BC or OC. This strategy of enhanced risk stratification will empower individuals and healthcare providers to adopt more targeted and effective preventive measures, potentially reducing the incidence and impact of BC and OC.

Recontacting the individuals that have been identified as a clinically significant PV or high-risk PRS carriers will ensure that they receive a comprehensive genetic counselling about their cancer risk.

This strategy can improve the long-term outcomes of high-risk individuals and their relatives by prioritizing the genetic screening and recontacting individuals carrying clinically significant PVs, thereby contributing to the overall health of the Latvian population.

Publications and reports on topics of doctoral Thesis

Publications

1. **Berga-Švītiņa, E.,** Pirsko, V., Nakazawa-Miklaševiča, M., Maksimenko, J., Gardovskis, J., and Miklaševičs, E. 2023. Penetrance of *CHEK2* and *BRCA1* Double Heterozygotes in Breast and/or Ovarian Cancer Patients. Proceedings of the Latvian Academy of Sciences. Section B. Natural, Exact, and Applied Sciences; vol.77, no.2, 137–140. doi: 10.2478/prolas-2023-0020.

2. **Berga-Švītiņa, E.,** Maksimenko, J., Miklaševičs, E., Fischer, K., Vilne, B., Mägi, R. 2023. Polygenic Risk Score Predicts Modified Risk in *BRCA1* Pathogenic Variant c.4035del and c.5266dup Carriers in Breast Cancer Patients. Cancers (Basel); 15(11):2957. doi: 10.3390/cancers15112957.

Reports and theses at international congresses and conferences

1. **Berga-Švītiņa, E.,** Pirsko, V., Nakazawa-Miklaševiča, M., Maksimenko, J., Gardovskis, J., and Miklaševičs, E. 2019. *CHEK2* Pathogenic variants do not Change Penetrance of *BRCA1* variants c.4034delA and c.5266dupC. Poster presentation at Rīga Stradiņš University International Research Conference on Medical and Health Care Sciences "Knowledge for Use in Practice": Abstracts, 1.–3.04.2019, 70.

2. **Berga-Švītiņa, E.,** Pirsko, V., Nakazawa-Miklaševiča, M., Maksimenko, J., Gardovskis, J., and Miklaševičs, E. 2023. Identifying Genetic Factors Associated with Breast or Ovarian Cancer Risk in *BRCA1* Pathogenic Variant Carriers. Oral presentation at Rīga Stradiņš University International Research Conference on Medical and Health Care Sciences "Knowledge for Use in Practice": Abstracts, 29.–31.03.2023.

3. **Berga-Švītiņa, E.,** Loža, P., Maksimenko, J., Fischer, K., Miklaševičs, E., Mägi, R., and Vilne, B. 2023. Exploring genetic factors and polygenic risk scores to predict breast and ovarian cancer risk in *BRCA1* pathogenic variant carriers. Oral presentations at Precision Medicine Networking Forum, 12.–13.10.2023.

References

1. Aksoy, F., Tezcan Unlu, H., Cecener, G., Guney Eskiler, G., Egeli, U., Tunca, B., . . . Gokgoz, M. S. (2022). Identification of CHEK2 germline mutations in BRCA1/2 and PALB2 negative breast and ovarian cancer patients. *Hum Hered*. https://doi.org/10.1159/000521369

2. Al-Mulla, F., Bland, J. M., Serratt, D., Miller, J., Chu, C., & Taylor, G. T. (2009). Age-dependent penetrance of different germline mutations in the BRCA1 gene. *J Clin Pathol*, *62*(4), 350–356. https://doi.org/10.1136/jcp.2008.062646

3. Alfarsi, L. H., El Ansari, R., Craze, M. L., Mohammed, O. J., Masisi, B. K., Ellis, I. O., . . . Green, A. R. (2021). SLC1A5 co-expression with TALDO1 associates with endocrine therapy failure in estrogen receptor-positive breast cancer. *Breast Cancer Res Treat*, *189*(2), 317–331. https://doi.org/10.1007/s10549-021-06298-1

4. Ali, A. T., Al-Ani, O., & Al-Ani, F. (2023). Epidemiology and risk factors for ovarian cancer. *Prz Menopauzalny*, *22*(2), 93–104. https://doi.org/10.5114/pm.2023.128661

5. Aljarf, R., Shen, M., Pires, D. E. V., & Ascher, D. B. (2022). Understanding and predicting the functional consequences of missense mutations in BRCA1 and BRCA2. *Sci Rep*, *12*(1), 10458. https://doi.org/10.1038/s41598-022-13508-3

6. Alves-Nogueira, A. C., Melo, D., Carona, C., & Figueiredo-Dias, M. (2023). The Psychosocial Impact of the Decision to Undergo Risk-Reducing Salpingo-Oophorectomy Surgery in. *Curr Oncol*, *30*(2), 2429–2440. https://doi.org/10.3390/curroncol30020185

7. Amos, C. I., Dennis, J., Wang, Z., Byun, J., Schumacher, F. R., Gayther, S. A., . . . Easton, D. F. (2017). The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers. *Cancer Epidemiol Biomarkers Prev*, *26*(1), 126–135. https://doi.org/10.1158/1055-9965.EPI-16-0106

8. Angeli, D., Salvi, S., & Tedaldi, G. (2020). Genetic Predisposition to Breast and Ovarian Cancers: How Many and Which Genes to Test? *Int J Mol Sci*, *21*(3). https://doi.org/10.3390/ijms21031128

9. Antoniou, A. C., Cunningham, A. P., Peto, J., Evans, D. G., Lalloo, F., Narod, S. A., . . . Easton, D. F. (2008). The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. *Br J Cancer*, *98*(8), 1457–1466. https://doi.org/10.1038/sj.bjc.6604305

10. Antoniou, A. C., Shenton, A., Maher, E. R., Watson, E., Woodward, E., Lalloo, F., . . . Evans, D. G. (2006). Parity and breast cancer risk among BRCA1 and BRCA2 mutation carriers. *Breast Cancer Res*, *8*(6), R72. https://doi.org/10.1186/bcr1630

11. Apostolou, P., & Papasotiriou, I. (2017). Current perspectives on CHEK2 mutations in breast cancer. *Breast Cancer (Dove Med Press)*, *9*, 331–335. https://doi.org/10.2147/BCTT.S111394

12. Arun, G., Diermeier, S. D., & Spector, D. L. (2018). Therapeutic Targeting of Long Non-Coding RNAs in Cancer. *Trends Mol Med*, *24*(3), 257–277. https://doi.org/10.1016/j.molmed.2018.01.001

13. Barnes, D. R., Rookus, M. A., McGuffog, L., Leslie, G., Mooij, T. M., Dennis, J., . . . CIMBA (2020). Polygenic risk scores and breast and epithelial ovarian cancer risks for carriers of BRCA1 and BRCA2 pathogenic variants. *Genet Med*, *22*(10), 1653–1666. https://doi.org/10.1038/s41436-020-0862-x

14. Barnum, K. J., & O'Connell, M. J. (2014). Cell cycle regulation by checkpoints. *Methods Mol Biol*, *1170*, 29–40. https://doi.org/10.1007/978-1-4939-0888-2_2

15. Bartek, J., & Lukas, J. (2003). Chk1 and Chk2 kinases in checkpoint control and cancer. *Cancer Cell*, *3*(5), 421–429. https://doi.org/10.1016/s1535-6108(03)00110-7

16. Bianco, G., Coto-Llerena, M., Gallon, J., Kancherla, V., Taha-Mehlitz, S., Marinucci, M., . . . Piscuoglio, S. (2022). GATA3 and MDM2 are synthetic lethal in estrogen receptor-positive breast cancers. *Commun Biol*, *5*(1), 373. https://doi.org/10.1038/s42003-022-03296-x

17. Bissonauth, V., Shatenstein, B., & Ghadirian, P. (2008). Nutrition and breast cancer among sporadic cases and gene mutation carriers: an overview. *Cancer Detect Prev*, *32*(1), 52–64. https://doi.org/10.1016/j.cdp.2008.01.005

18. Borde, J., Laitman, Y., Blümcke, B., Niederacher, D., Weber-Lassalle, K., Sutter, C., . . . Ernst, C. (2022). Polygenic risk scores indicate extreme ages at onset of breast cancer in female BRCA1/2 pathogenic variant carriers. *BMC Cancer*, *22*(1), 706. https://doi.org/10.1186/s12885-022-09780-1

19. Boyd, N. F., Martin, L. J., Bronskill, M., Yaffe, M. J., Duric, N., & Minkin, S. (2010). Breast tissue composition and susceptibility to breast cancer. *J Natl Cancer Inst*, *102*(16), 1224–1237. https://doi.org/10.1093/jnci/djq239

20. Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am J Hum Genet*, *103*(3), 338–348. https://doi.org/10.1016/j.ajhg.2018.07.015

21. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., . . . Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, *562*(7726), 203–209. https://doi.org/10.1038/s41586-018-0579-z

22. Cai, Z., Chehab, N. H., & Pavletich, N. P. (2009). Structure and activation mechanism of the CHK2 DNA damage checkpoint kinase. *Mol Cell*, *35*(6), 818–829. https://doi.org/10.1016/j.molcel.2009.09.007

23. Caputo, S. M., Golmard, L., Léone, M., Damiola, F., Guillaud-Bataille, M., Revillion, F., . . . Stoppa-Lyonnet, D. (2021). Classification of 101 BRCA1 and BRCA2 variants of uncertain significance by cosegregation study: A powerful approach. *Am J Hum Genet*, *108*(10), 1907–1923. https://doi.org/10.1016/j.ajhg.2021.09.003

24. Carstensen, B., Plummer, M., Laara, E., Hills, M. R Package, Version 2.46; Epi: A Package for Statistical Analysis in Epidemiology; Oxford University Press: Oxford, UK, 2022; Available online: https://CRAN.R-project.org/package=Epi (accessed on 22 May 2023).

25. Casalou, C., Ferreira, A., & Barral, D. C. (2020). The Role of ARF Family Proteins and Their Regulators and Effectors in Cancer Progression: A Therapeutic Perspective. *Front Cell Dev Biol*, *8*, 217. https://doi.org/10.3389/fcell.2020.00217

26. CDPC. (2020). Statistical Data. Centre for Disease Prevention and Control. Available online: https://www.spkc.gov.lv/lv/statistikas-dati (accessed on 3 March 2023).

27. Chen, H., Wang, C., Conomos, M. P., Stilp, A. M., Li, Z., Sofer, T., . . . Lin, X. (2016). Control for Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models. *Am J Hum Genet*, *98*(4), 653–666. https://doi.org/10.1016/j.ajhg.2016.02.012

28. Chen, J., Bae, E., Zhang, L., Hughes, K., Parmigiani, G., Braun, D., & Rebbeck, T. R. (2020). Penetrance of Breast and Ovarian Cancer in Women Who Carry a. *JNCI Cancer Spectr*, *4*(4), pkaa029. https://doi.org/10.1093/jncics/pkaa029

29. Chen, S., & Parmigiani, G. (2007). Meta-analysis of BRCA1 and BRCA2 penetrance. *J Clin Oncol*, *25*(11), 1329–1333. https://doi.org/10.1200/JCO.2006.09.1066

30. Cline, M. S., Liao, R. G., Parsons, M. T., Paten, B., Alquaddoomi, F., Antoniou, A., . . . Spurdle A. B. (2018). BRCA Challenge: BRCA Exchange as a global resource for variants in BRCA1 and BRCA2. *PLoS Genet*, *14*(12), e1007752. https://doi.org/10.1371/journal.pgen.1007752

31. Cohen, S. Y., Stoll, C. R., Anandarajah, A., Doering, M., & Colditz, G. A. (2023). Modifiable risk factors in women at high risk of breast cancer: a systematic review. *Breast Cancer Res*, *25*(1), 45. https://doi.org/10.1186/s13058-023-01636-1

32. Coignard, J., Lush, M., Beesley, J., O'Mara, T. A., Dennis, J., Tyrer, J. P., . . . Antoniou, A. C. (2021). A case-only study to identify genetic modifiers of breast cancer risk for BRCA1/BRCA2 mutation carriers. *Nat Commun*, *12*(1), 1078. https://doi.org/10.1038/s41467-020-20496-3

33. Collins, I., & Garrett, M. D. (2005). Targeting the cell division cycle in cancer: CDK and cell cycle checkpoint kinase inhibitors. *Curr Opin Pharmacol*, *5*(4), 366–373. https://doi.org/10.1016/j.coph.2005.04.009

34. Collister, J. A., Liu, X., & Clifton, L. (2022). Calculating Polygenic Risk Scores (PRS) in UK Biobank: A Practical Guide for Epidemiologists. *Front Genet*, *13*, 818574. https://doi.org/10.3389/fgene.2022.818574

35. Cooper, D. N., Krawczak, M., Polychronakos, C., Tyler-Smith, C., & Kehrer-Sawatzki, H. (2013). Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum Genet*, *132*(10), 1077–1130. https://doi.org/10.1007/s00439-013-1331-2

36. Couch, D. B. (1996). Carcinogenesis: basic principles. *Drug Chem Toxicol*, *19*(3), 133–148. https://doi.org/10.3109/01480549608998231

37. Couch, F. J., Wang, X., McGuffog, L., Lee, A., Olswold, C., Kuchenbaecker, K. B., . . . CIMBA. (2013). Genome-wide association study in BRCA1 mutation carriers identifies novel loci associated with breast and ovarian cancer risk. *PLoS Genet*, *9*(3), e1003212. https://doi.org/10.1371/journal.pgen.1003212

38. Creeden, J. F., Nanavaty, N. S., Einloth, K. R., Gillman, C. E., Stanbery, L., Hamouda, D. M., . . . Nemunaitis, J. (2021). Homologous recombination proficiency in ovarian and breast cancer patients. *BMC Cancer*, *21*(1), 1154. https://doi.org/10.1186/s12885-021-08863-9

39. Cullinane, C. A., Lubinski, J., Neuhausen, S. L., Ghadirian, P., Lynch, H. T., Isaacs, C., . . . Narod, S. A. (2005). Effect of pregnancy as a risk factor for breast cancer in BRCA1/BRCA2 mutation carriers. *Int J Cancer*, *117*(6), 988–991. https://doi.org/10.1002/ijc.21273

40. Cybulski, C., Górski, B., Huzarski, T., Byrski, T., Gronwald, J., Debniak, T., Lubinski, J. (2009). Effect of CHEK2 missense variant I157T on the risk of breast cancer in carriers of other CHEK2 or BRCA1 mutations. *J Med Genet*, *46*(2), 132–135. https://doi.org/10.1136/jmg.2008.061697

41. Cybulski, C., Górski, B., Huzarski, T., Masojć, B., Mierzejewski, M., Debniak, T., . . . Lubiński, J. (2004). CHEK2 is a multiorgan cancer susceptibility gene. *Am J Hum Genet*, *75*(6), 1131–1135. https://doi.org/10.1086/426403

42. Cybulski, C., Wokołorczyk, D., Huzarski, T., Byrski, T., Gronwald, J., Górski, B., Lubiński, J. (2006). A large germline deletion in the Chek2 kinase gene is associated with an increased risk of prostate cancer. *J Med Genet*, *43*(11), 863–866. https://doi.org/10.1136/jmg.2006.044974

43. Cybulski, C., Wokołorczyk, D., Kładny, J., Kurzawski, G., Suchy, J., Grabowska, E., . . . Lubiński, J. (2007). Germline CHEK2 mutations and colorectal cancer risk: different effects of a missense and truncating mutations? *Eur J Hum Genet*, *15*(2), 237–241. https://doi.org/10.1038/sj.ejhg.5201734

44. Dahlqwist, E., Magnusson, P. K. E., Pawitan, Y., & Sjölander, A. (2019). On the relationship between the heritability and the attributable fraction. *Hum Genet*, *138*(4), 425–435. https://doi.org/10.1007/s00439-019-02006-8

45. Daly, M. B., Pal, T., Berry, M. P., Buys, S. S., Dickson, P., Domchek, S. M., . . . Dwyer, M. A. (2021). Genetic/Familial High-Risk Assessment: Breast, Ovarian, and Pancreatic, Version 2.2021, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw*, *19*(1), 77–102. https://doi.org/10.6004/jnccn.2021.0001

46. Dareng, E. O., Tyrer, J. P., Barnes, D. R., Jones, M. R., Yang, X., Aben, K. K. H., . . . Pharaoah, P. D. P. (2022). Polygenic risk modeling for prediction of epithelial ovarian cancer risk. *Eur J Hum Genet*, *30*(3), 349–362. https://doi.org/10.1038/s41431-021-00987-7

47. Dasari, S., & Tchounwou, P. B. (2014). Cisplatin in cancer therapy: molecular mechanisms of action. *Eur J Pharmacol*, *740*, 364–378. https://doi.org/10.1016/j.ejphar.2014.07.025

48. Den Brok, W. D., Schrader, K. A., Sun, S., Tinker, A. V., Zhao, E. Y., Aparicio, S., & Gelmon, K. A. (2017). Homologous Recombination Deficiency in Breast Cancer: A Clinical Review. *JCO Precis Oncol*, *1*, 1–13. https://doi.org/10.1200/PO.16.00031

49. Ding, L., Ellis, M. J., Li, S., Larson, D. E., Chen, K., Wallis, J. W., . . . Mardis, E. R. (2010). Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature*, *464*(7291), 999–1005. https://doi.org/10.1038/nature08989

50. Domchek, S. M., Friebel, T. M., Singer, C. F., Evans, D. G., Lynch, H. T., Isaacs, C., . . . Rebbeck, T. R. (2010). Association of risk-reducing surgery in BRCA1 or BRCA2 mutation carriers with cancer risk and mortality. *JAMA*, *304*(9), 967–975. https://doi.org/10.1001/jama.2010.1237

51. Doraczynska-Kowalik, A., Michalowska, D., Matkowski, R., Czykalko, E., Blomka, D., Semeniuk, M., . . . Laczmanska, I. (2022). Detection of BRCA1/2 pathogenic variants in patients with breast and/or ovarian cancer and their families. Analysis of 3,458 cases from Lower Silesia (Poland) according to the diagnostic algorithm of the National Cancer Control Programme. *Front Genet*, *13*, 941375. https://doi.org/10.3389/fgene.2022.941375

52. Downs, B., Sherman, S., Cui, J., Kim, Y. C., Snyder, C., Christensen, M., . . . Wang, S. M. (2019). Common genetic variants contribute to incomplete penetrance: evidence from cancer-free BRCA1 mutation carriers. *Eur J Cancer*, *107*, 68–78. https://doi.org/10.1016/j.ejca.2018.10.022

53. Edwards, S. L., Beesley, J., French, J. D., & Dunning, A. M. (2013). Beyond GWASs: illuminating the dark road from association to function. *Am J Hum Genet*, *93*(5), 779–797. https://doi.org/10.1016/j.ajhg.2013.10.012

54. Eliyatkın, N., Yalçın, E., Zengel, B., Aktaş, S., & Vardar, E. (2015). Molecular Classification of Breast Carcinoma: From Traditional, Old-Fashioned Way to A New Age, and A New Way. *J Breast Health*, *11*(2), 59–66. https://doi.org/10.5152/tjbh.2015.1669

55. Escala-Garcia, M., Guo, Q., Dörk, T., Canisius, S., Keeman, R., Dennis, J., . . . Schmidt, M. K. (2019). Genome-wide association study of germline variants and breast cancer-specific mortality. *Br J Cancer*, *120*(6), 647–657. https://doi.org/10.1038/s41416-019-0393-x

56. Fahed, A. C., Wang, M., Homburger, J. R., Patel, A. P., Bick, A. G., Neben, C. L., . . . Khera, A. V. (2020). Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. *Nat Commun*, *11*(1), 3635. https://doi.org/10.1038/s41467-020-17374-3

57. Ferris, J. S., Morgan, D. A., Tseng, A. S., Terry, M. B., Ottman, R., Hur, C., . . . Genkinger, J. M. (2023). Risk factors for developing both primary breast and primary ovarian cancer: A systematic review. *Crit Rev Oncol Hematol*, *190*, 104081. https://doi.org/10.1016/j.critrevonc.2023.104081

58. Filippini, S. E., & Vega, A. (2013). Breast cancer genes: beyond BRCA1 and BRCA2. *Front Biosci (Landmark Ed)*, *18*(4), 1358–1372. https://doi.org/10.2741/4185

59. Fonseca-Montaño, M. A., Vázquez-Santillán, K. I., & Hidalgo-Miranda, A. (2023). The current advances of lncRNAs in breast cancer immunobiology research. *Front Immunol*, *14*, 1194300. https://doi.org/10.3389/fimmu.2023.1194300

60. Forrest, I. S., Chaudhary, K., Vy, H. M. T., Petrazzini, B. O., Bafna, S., Jordan, D. M., . . . Do, R. (2022). Population-Based Penetrance of Deleterious Clinical Variants. *JAMA*, *327*(4), 350–359. https://doi.org/10.1001/jama.2021.23686

61. Foulkes, W. D., Metcalfe, K., Sun, P., Hanna, W. M., Lynch, H. T., Ghadirian, P., . . . Narod, S. A. (2004). Estrogen receptor status in BRCA1- and BRCA2-related breast cancer: the influence of age, grade, and histological type. *Clin Cancer Res*, *10*(6), 2029–2034. https://doi.org/10.1158/1078-0432.ccr-03-1061

62. Fu, X., Tan, W., Song, Q., Pei, H., & Li, J. (2022). BRCA1 and Breast Cancer: Molecular Mechanisms and Therapeutic Strategies. *Front Cell Dev Biol*, *10*, 813457. https://doi.org/10.3389/fcell.2022.813457

63. Futreal, P. A., Liu, Q., Shattuck-Eidens, D., Cochran, C., Harshman, K., Tavtigian, S., . . . Miki, Y. (1994). BRCA1 mutations in primary breast and ovarian carcinomas. *Science*, *266*(5182), 120–122. https://doi.org/10.1126/science.7939630

64. Gardovskis, A., Irmejs, A., Miklasevics, E., Borosenko, V., Bitina, M., Melbarde-Gorkusa, I., . . . Gardovskis, J. (2005). Clinical, molecular and geographical features of hereditary breast/ovarian cancer in latvia. *Hered Cancer Clin Pract*, *3*(2), 71–76. https://doi.org/10.1186/1897-4287-3-2-71

65. Gayther, S. A., Warren, W., Mazoyer, S., Russell, P. A., Harrington, P. A., Chiano, M., . . . Ponder, B. A. (1995). Germline mutations of the BRCA1 gene in breast and ovarian cancer families provide evidence for a genotype-phenotype correlation. *Nat Genet*, *11*(4), 428–433. https://doi.org/10.1038/ng1295-428

66. Gibbs, Z. A., Reza, L. C., Cheng, C. C., Westcott, J. M., McGlynn, K., & Whitehurst, A. W. (2020). The testis protein ZNF165 is a SMAD3 cofactor that coordinates oncogenic TGFβ signaling in triple-negative breast cancer. *Elife*, *9*. https://doi.org/10.7554/eLife.57679

67. Gong, X., Zhang, Y., Ai, J., & Li, K. (2022). Application of Single-Cell RNA Sequencing in Ovarian Development. *Biomolecules*, *13*(1). https://doi.org/10.3390/biom13010047

68. Guo, J., Bian, Y., Wang, Y., Chen, L., Yu, A., & Sun, X. (2017). FAM107B is regulated by S100A4 and mediates the effect of S100A4 on the proliferation and migration of MGC803 gastric cancer cells. *Cell Biol Int*, *41*(10), 1103–1109. https://doi.org/10.1002/cbin.10816

69. Guo, Q., Schmidt, M. K., Kraft, P., Canisius, S., Chen, C., Khan, S., . . . Pharoah, P. D. P. (2015). Identification of novel genetic markers of breast cancer survival. *J Natl Cancer Inst*, *107*(5). https://doi.org/10.1093/jnci/djv081

70. Guo, Y., He, J., Zhao, S., Wu, H., Zhong, X., Sheng, Q., . . . Long, J. (2014). Illumina human exome genotyping array clustering and quality control. *Nat Protoc*, *9*(11), 2643–2662. https://doi.org/10.1038/nprot.2014.174

71. Gutschner, T., & Diederichs, S. (2012). The hallmarks of cancer: a long non-coding RNA point of view. *RNA Biol*, *9*(6), 703–719. https://doi.org/10.4161/rna.20481

72. Hamel, N., Feng, B. J., Foretova, L., Stoppa-Lyonnet, D., Narod, S. A., Imyanitov, E., . . . Foulkes, W. D. (2011). On the origin and diffusion of BRCA1 c.5266dupC (5382insC) in European populations. *Eur J Hum Genet*, *19*(3), 300–306. https://doi.org/10.1038/ejhg.2010.203

73. Hampel, H., Bennett, R. L., Buchanan, A., Pearlman, R., Wiesner, G. L., & Guideline Development Group, American College of Medical Genetics and Genomics Professional Practice and Guidelines Committee and National Society of Genetic Counselors Practice Guidelines Committee (2015). A practice guideline from the American College of Medical Genetics and Genomics and the National Society of Genetic Counselors: referral indications for cancer predisposition assessment. *Genet Med*, *17*(1), 70–87. https://doi.org/10.1038/gim.2014.147

74. Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *Cell*, *100*(1), 57–70. https://doi.org/10.1016/s0092-8674(00)81683-9

75. Harbeck, N., Penault-Llorca, F., Cortes, J., Gnant, M., Houssami, N., Poortmans, P., . . . Cardoso, F. (2019). Breast cancer. *Nat Rev Dis Primers*, *5*(1), 66. https://doi.org/10.1038/s41572-019-0111-2

76. Harris, C. C. (1996). p53 tumor suppressor gene: from the basic research laboratory to the clinic – an abridged historical perspective. *Carcinogenesis*, *17*(6), 1187–1198. https://doi.org/10.1093/carcin/17.6.1187

77. Hartmann, J. T., Haap, M., Kopp, H. G., & Lipp, H. P. (2009). Tyrosine kinase inhibitors – a review on pharmacology, metabolism and side effects. *Curr Drug Metab*, *10*(5), 470–481. https://doi.org/10.2174/138920009788897975

78. Hartwell, L. H., & Kastan, M. B. (1994). Cell cycle control and cancer. *Science*, *266*(5192), 1821–1828. https://doi.org/10.1126/science.7997877

79. Hayashi, T., & Konishi, I. (2023). Molecular Histopathology for Establishing Diagnostic Method and Clinical Therapy for Ovarian Carcinoma. *J Clin Med Res*, *15*(2), 68–75. https://doi.org/10.14740/jocmr4853

80. Honrado, E., Osorio, A., Palacios, J., & Benitez, J. (2006). Pathology and gene expression of hereditary breast tumors associated with BRCA1, BRCA2 and CHEK2 gene mutations. *Oncogene*, *25*(43), 5837–5845. https://doi.org/10.1038/sj.onc.1209875

81. Iodice, S., Barile, M., Rotmensz, N., Feroce, I., Bonanni, B., Radice, P., . . . Gandini, S. (2010). Oral contraceptive use and breast or ovarian cancer risk in BRCA1/2 carriers: a meta-analysis. *Eur J Cancer*, *46*(12), 2275–2284. https://doi.org/10.1016/j.ejca.2010.04.018

82. Irmejs, A., Miklasevics, E., Boroschenko, V., Gardovskis, A., Vanags, A., Melbarde-Gorkusa, I., . . . Gardovskis, J. (2006). Pilot study on low penetrance breast and colorectal cancer predisposition markers in latvia. *Hered Cancer Clin Pract*, *4*(1), 48–51. https://doi.org/10.1186/1897-4287-4-1-48

83. Ittisoponpisan, S., Alhuzimi, E., Sternberg, M. J., & David, A. (2017). Landscape of Pleiotropic Proteins Causing Human Disease: Structural and System Biology Insights. *Hum Mutat*, *38*(3), 289–296. https://doi.org/10.1002/humu.23155

84. Janavičius, R. (2010). Founder BRCA1/2 mutations in the Europe: implications for hereditary breast-ovarian cancer prevention and control. *EPMA J*, *1*(3), 397–412. https://doi.org/10.1007/s13167-010-0037-y

85. Janavičius, R., Rudaitis, V., Feng, B. J., Ozolina, S., Griškevičius, L., Goldgar, D., & Tihomirova, L. (2013). Haplotype analysis and ancient origin of the BRCA1 c.4035delA Baltic founder mutation. *Eur J Med Genet*, *56*(3), 125–130. https://doi.org/10.1016/j.ejmg.2012.12.007

86. Janavičius, R., Rudaitis, V., Mickys, U., Elsakov, P., & Griškevičius, L. (2014). Comprehensive BRCA1 and BRCA2 mutational profile in Lithuania. *Cancer Genet*, *207*(5), 195–205. https://doi.org/10.1016/j.cancergen.2014.05.002

87. Jernström, H., Lubinski, J., Lynch, H. T., Ghadirian, P., Neuhausen, S., Isaacs, C., . . . Narod, S. A. (2004). Breast-feeding and the risk of breast cancer in BRCA1 and BRCA2 mutation carriers. *J Natl Cancer Inst*, *96*(14), 1094–1098. https://doi.org/10.1093/jnci/djh211

88. Jia, G., Ping, J., Shu, X., Yang, Y., Cai, Q., Kweon, S. S., . . . Zheng, W. (2022). Genome- and transcriptome-wide association studies of 386,000 Asian and European-ancestry women provide new insights into breast cancer genetics. *Am J Hum Genet*, *109*(12), 2185–2195. https://doi.org/10.1016/j.ajhg.2022.10.011

89. Jones, M. R., Kamara, D., Karlan, B. Y., Pharoah, P. D. P., & Gayther, S. A. (2017). Genetic epidemiology of ovarian cancer and prospects for polygenic risk prediction. *Gynecol Oncol*, *147*(3), 705–713. https://doi.org/10.1016/j.ygyno.2017.10.001

90. Jurj, M. A., Buse, M., Zimta, A. A., Paradiso, A., Korban, S. S., Pop, L. A., & Berindan-Neagoe, I. (2020). Critical Analysis of Genome-Wide Association Studies: Triple Negative Breast Cancer. *Int J Mol Sci*, *21*(16). https://doi.org/10.3390/ijms21165835

91. Jürgens, H., Roht, L., Leitsalu, L., Nõukas, M., Palover, M., Nikopensius, T., . . . Tõnisson, N. (2022). Precise, Genotype-First Breast Cancer Prevention: Experience With Transferring Monogenic Findings From a Population Biobank to the Clinical Setting. *Front Genet*, *13*, 881100. https://doi.org/10.3389/fgene.2022.881100

92. Kenemans, P., Verstraeten, R. A., & Verheijen, R. H. (2008). Oncogenic pathways in hereditary and sporadic breast cancer. *Maturitas*, *61*(1–2), 141–150. https://doi.org/10.1016/j.maturitas.2008.11.010

93. Kerr, S. M., Cowan, E., Klaric, L., Bell, C., O'Sullivan, D., Buchanan, D., . . . Miedzybrodzka, Z. (2023). Clinical case study meets population cohort: identification of a BRCA1 pathogenic founder variant in Orcadians. *Eur J Hum Genet*, *31*(5), 588–595. https://doi.org/10.1038/s41431-023-01297-w

94. Kingdom, R., & Wright, C. F. (2022). Incomplete Penetrance and Variable Expressivity: From Clinical Studies to Population Cohorts. *Front Genet*, *13*, 920390. https://doi.org/10.3389/fgene.2022.920390

95. Konecny, G. E., & Kristeleit, R. S. (2016). PARP inhibitors for BRCA1/2-mutated and sporadic ovarian cancer: current practice and future directions. *Br J Cancer*, *115*(10), 1157–1173. https://doi.org/10.1038/bjc.2016.311

96. Kontomanolis, E. N., Koutras, A., Syllaios, A., Schizas, D., Mastoraki, A., Garmpis, N., . . . Fasoulakis, Z. (2020). Role of Oncogenes and Tumor-suppressor Genes in Carcinogenesis: A Review. *Anticancer Res*, *40*(11), 6009–6015. https://doi.org/10.21873/anticanres.14622

97. Krais, J. J., & Johnson, N. (2020). BRCA1 Mutations in Cancer: Coordinating Deficiencies in Homologous Recombination with Tumorigenesis. *Cancer Res*, *80*(21), 4601–4609. https://doi.org/10.1158/0008-5472.CAN-20-1830

98. Kuchenbaecker, K. B., Hopper, J. L., Barnes, D. R., Phillips, K. A., Mooij, T. M., Roos-Blom, M. J., . . . Olsson, H. (2017). Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. *JAMA*, *317*(23), 2402–2416. https://doi.org/10.1001/jama.2017.7112

99. Kuchenbaecker, K. B., McGuffog, L., Barrowdale, D., Lee, A., Soucy, P., Dennis, J., . . . Antoniou, A. C. (2017). Evaluation of Polygenic Risk Scores for Breast and Ovarian Cancer Risk Prediction in BRCA1 and BRCA2 Mutation Carriers. *J Natl Cancer Inst*, *109*(7). https://doi.org/10.1093/jnci/djw302

100. Kuchenbaecker, K. B., Ramus, S. J., Tyrer, J., Lee, A., Shen, H. C., Beesley, J., . . . CIMBA (2015). Identification of six new susceptibility loci for invasive epithelial ovarian cancer. *Nat Genet*, *47*(2), 164–171. https://doi.org/10.1038/ng.3185

101. Lavoro, A., Scalisi, A., Candido, S., Zanghì, G. N., Rizzo, R., Gattuso, G., . . . Falzone, L. (2022). Identification of the most common BRCA alterations through analysis of germline mutation databases: Is droplet digital PCR an additional strategy for the assessment of such alterations in breast and ovarian cancer families? *Int J Oncol*, *60*(5). https://doi.org/10.3892/ijo.2022.5349

102. Lee, A., Moon, B. I., & Kim, T. H. (2020). /. *Ann Lab Med*, *40*(2), 114–121. https://doi.org/10.3343/alm.2020.40.2.114

103. Lee, E. Y., & Muller, W. J. (2010). Oncogenes and tumor suppressor genes. *Cold Spring Harb Perspect Biol*, *2*(10), a003236. https://doi.org/10.1101/cshperspect.a003236

104. Lee, J. S., Collins, K. M., Brown, A. L., Lee, C. H., & Chung, J. H. (2000). hCds1-mediated phosphorylation of BRCA1 regulates the DNA damage response. *Nature*, *404*(6774), 201–204. https://doi.org/10.1038/35004614

105. Leitsalu, L., Palover, M., Sikka, T. T., Reigo, A., Kals, M., Pärn, K., . . . Tõnisson, N. (2021). Genotype-first approach to the detection of hereditary breast and ovarian cancer risk, and effects of risk disclosure to biobank participants. *Eur J Hum Genet*, *29*(3), 471–481. https://doi.org/10.1038/s41431-020-00760-2

106. Li, H., Engel, C., de la Hoya, M., Peterlongo, P., Yannoukakos, D., Livraghi, L., . . . CIMBA (2022). Risks of breast and ovarian cancer for women harboring pathogenic missense variants in BRCA1 and BRCA2 compared with those harboring protein truncating variants. *Genet Med*, *24*(1), 119–129. https://doi.org/10.1016/j.gim.2021.08.016

107. Li, S., Silvestri, V., Leslie, G., Rebbeck, T. R., Neuhausen, S. L., Hopper, J. L., . . . Antoniou, A. C. (2022). Cancer Risks Associated With. *J Clin Oncol*, *40*(14), 1529–1541. https://doi.org/10.1200/JCO.21.02112

108. Liu, Q., Zhao, X. Y., Bai, R. Z., Liang, S. F., Nie, C. L., Yuan, Z., . . . Wei, Y. Q. (2009). Induction of tumor inhibition and apoptosis by a candidate tumor suppressor gene DRR1 on 3p21.1. *Oncol Rep*, *22*(5), 1069–1075.

109. Liu, S., Sun, Y., Hou, Y., Yang, L., Wan, X., Qin, Y., . . . Liu, M. (2021). A novel lncRNA ROPM-mediated lipid metabolism governs breast cancer stem cell properties. *J Hematol Oncol*, *14*(1), 178. https://doi.org/10.1186/s13045-021-01194-z

110. Loh, P. R., Palamara, P. F., & Price, A. L. (2016). Fast and accurate long-range phasing in a UK Biobank cohort. *Nat Genet*, *48*(7), 811–816. https://doi.org/10.1038/ng.3571

111. Lord, C. J., & Ashworth, A. (2017). PARP inhibitors: Synthetic lethality in the clinic. *Science*, *355*(6330), 1152–1158. https://doi.org/10.1126/science.aam7344

112. Lu, Y., Beeghly-Fadiel, A., Wu, L., Guo, X., Li, B., Schildkraut, J. M., . . . Long, J. (2018). A Transcriptome-Wide Association Study Among 97,898 Women to Identify Candidate Susceptibility Genes for Epithelial Ovarian Cancer Risk. *Cancer Res*, *78*(18), 5419–5430. https://doi.org/10.1158/0008-5472.CAN-18-0951

113. Ludwig, K. K., Neuner, J., Butler, A., Geurts, J. L., & Kong, A. L. (2016). Risk reduction and survival benefit of prophylactic surgery in BRCA mutation carriers, a systematic review. *Am J Surg*, *212*(4), 660–669. https://doi.org/10.1016/j.amjsurg.2016.06.010

114. Luo, A., Zhang, K., Zhao, Y., Zhu, Z., Fu, L., & Dong, J. T. (2018). ZNF121 interacts with ZBRK1 and BRCA1 to regulate their target genes in mammary epithelial cells. *FEBS Open Bio*, *8*(12), 1943–1952. https://doi.org/10.1002/2211-5463.12530

115. Lyra, P., LB, R., & ANA, M. (2020). Functional landscape of common variants associated with susceptibility to epithelial ovarian cancer. In (Vol. 7, 49–57): Curr Epidemiol Rep.

116. Läll, K., Lepamets, M., Palover, M., Esko, T., Metspalu, A., Tõnisson, N., . . . Fischer, K. (2019). Polygenic prediction of breast cancer: comparison of genetic predictors and implications for risk stratification. *BMC Cancer*, *19*(1), 557. https://doi.org/10.1186/s12885-019-5783-1

117. Maas, P., Barrdahl, M., Joshi, A. D., Auer, P. L., Gaudet, M. M., Milne, R. L., . . . Chatterjee, N. (2016). Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States. *JAMA Oncol*, *2*(10), 1295–1302. https://doi.org/10.1001/jamaoncol.2016.1025

118. Makki, J. (2015). Diversity of Breast Carcinoma: Histological Subtypes and Clinical Relevance. *Clin Med Insights Pathol*, *8*, 23–31. https://doi.org/10.4137/CPath.S31563

119. Manickam, K., Buchanan, A. H., Schwartz, M. L. B., Hallquist, M. L. G., Williams, J. L., Rahm, A. K., . . . Murray, M. F. (2018). Exome Sequencing-Based Screening for BRCA1/2 Expected Pathogenic Variants Among Adult Biobank Participants. *JAMA Netw Open*, *1*(5), e182140. https://doi.org/10.1001/jamanetworkopen.2018.2140

120. Marees, A. T., de Kluiver, H., Stringer, S., Vorspan, F., Curis, E., Marie-Claire, C., & Derks, E. M. (2018). A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *Int J Methods Psychiatr Res*, *27*(2), e1608. https://doi.org/10.1002/mpr.1608

121. Mars, N., Koskela, J. T., Ripatti, P., Kiiskinen, T. T. J., Havulinna, A. S., Lindbohm, J. V., . . . Ripatti, S. (2020). Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. *Nat Med*, *26*(4), 549–557. https://doi.org/10.1038/s41591-020-0800-0

122. Mars, N., Widén, E., Kerminen, S., Meretoja, T., Pirinen, M., Della Briotta Parolo, P., . . . Ripatti, S. (2020). The role of polygenic risk and susceptibility genes in breast cancer over the course of life. *Nat Commun*, *11*(1), 6383. https://doi.org/10.1038/s41467-020-19966-5

123. Mavaddat, N., Michailidou, K., Dennis, J., Lush, M., Fachal, L., Lee, A., . . . Easton, D. F. (2019). Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am J Hum Genet*, *104*(1), 21–34. https://doi.org/10.1016/j.ajhg.2018.11.002

124. Mavaddat, N., Peock, S., Frost, D., Ellis, S., Platte, R., Fineberg, E., . . . EMBRACE. (2013). Cancer risks for BRCA1 and BRCA2 mutation carriers: results from prospective analysis of EMBRACE. *J Natl Cancer Inst*, *105*(11), 812–822. https://doi.org/10.1093/jnci/djt095

125. McLaughlin, J. R., Risch, H. A., Lubinski, J., Moller, P., Ghadirian, P., Lynch, H., . . . Hereditary Ovarian Cancer Clinical Study Group (2007). Reproductive risk factors for ovarian cancer in carriers of BRCA1 or BRCA2 mutations: a case-control study. *Lancet Oncol*, *8*(1), 26–34. https://doi.org/10.1016/S1470-2045(06)70983-4

126. Meijers-Heijboer, H., van den Ouweland, A., Klijn, J., Wasielewski, M., de Snoo, A., Oldenburg, R., . . . CHEK2-Breast Cancer Consortium (2002). Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nat Genet*, *31*(1), 55–59. https://doi.org/10.1038/ng879

127. Mekonnen, N., Yang, H., & Shin, Y. K. (2022). Homologous Recombination Deficiency in Ovarian, Breast, Colorectal, Pancreatic, Non-Small Cell Lung and Prostate Cancers, and the Mechanisms of Resistance to PARP Inhibitors. *Front Oncol*, *12*, 880643. https://doi.org/10.3389/fonc.2022.880643

128. Michailidou, K., Beesley, J., Lindstrom, S., Canisius, S., Dennis, J., Lush, M. J., . . . Easton, D. F. (2015). Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat Genet*, *47*(4), 373–380. https://doi.org/10.1038/ng.3242

129. Michailidou, K., Hall, P., Gonzalez-Neira, A., Ghoussaini, M., Dennis, J., Milne, R. L., . . . Easton, D. F. (2013). Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet*, *45*(4), 353–361, 361e351–352. https://doi.org/10.1038/ng.2563

130. Michailidou, K., Lindström, S., Dennis, J., Beesley, J., Hui, S., Kar, S., . . . Easton, D. F. (2017). Association analysis identifies 65 new breast cancer risk loci. *Nature*, *551*(7678), 92–94. https://doi.org/10.1038/nature24284

131. Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., . . . Ding, W. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*, *266*(5182), 66–71. https://doi.org/10.1126/science.7545954

132. Milne, R. L., & Antoniou, A. C. (2011). Genetic modifiers of cancer risk for BRCA1 and BRCA2 mutation carriers. *Ann Oncol*, *22 Suppl 1*, i11–17. https://doi.org/10.1093/annonc/mdq660

133. Milne, R. L., & Antoniou, A. C. (2016). Modifiers of breast and ovarian cancer risks for BRCA1 and BRCA2 mutation carriers. *Endocr Relat Cancer*, *23*(10), T69–84. https://doi.org/10.1530/ERC-16-0277

134. Milne, R. L., Kuchenbaecker, K. B., Michailidou, K., Beesley, J., Kar, S., Lindström, S., . . . Simard, J. (2017). Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. *Nat Genet*, *49*(12), 1767–1778. https://doi.org/10.1038/ng.3785

135. Mitt, M., Kals, M., Pärn, K., Gabriel, S. B., Lander, E. S., Palotie, A., . . . Palta, P. (2017). Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. *Eur J Hum Genet*, *25*(7), 869–876. https://doi.org/10.1038/ejhg.2017.51

136. Miziak, P., Baran, M., Błaszczak, E., Przybyszewska-Podstawka, A., Kałafut, J., Smok-Kalwat, J., . . . Stepulak, A. (2023). Estrogen Receptor Signaling in Breast Cancer. *Cancers (Basel)*, *15*(19). https://doi.org/10.3390/cancers15194689

137. Muranen, T. A., Morra, A., Khan, S., Barnes, D. R., Bolla, M. K., Dennis, J., . . . Nevanlinna, H. (2023). PREDICT validity for prognosis of breast cancer patients with pathogenic BRCA1/2 variants. *NPJ Breast Cancer*, *9*(1), 37. https://doi.org/10.1038/s41523-023-00546-x

138. Myszka, A., Karpinski, P., Slezak, R., Czemarmazowicz, H., Stembalska, A., Gil, J., . . . Sasiadek, M. M. (2011). Irrelevance of CHEK2 variants to diagnosis of breast/ovarian cancer predisposition in Polish cohort. *J Appl Genet*, *52*(2), 185–191. https://doi.org/10.1007/s13353-010-0013-1

139. Nachef, M., Ali, A. K., Almutairi, S. M., & Lee, S. H. (2021). Targeting SLC1A5 and SLC3A2/SLC7A5 as a Potential Strategy to Strengthen Anti-Tumor Immunity in the Tumor Microenvironment. *Front Immunol*, *12*, 624324. https://doi.org/10.3389/fimmu.2021.624324

140. Nakajima, H., Ishigaki, Y., Xia, Q. S., Ikeda, T., Yoshitake, Y., Yonekura, H., . . . Motoo, Y. (2010). Induction of HITS, a newly identified family with sequence similarity 107 protein (FAM107B), in cancer cells by heat shock stimulation. *Int J Oncol*, *37*(3), 583–593. https://doi.org/10.3892/ijo_00000707

141. Nakajima, H., & Koizumi, K. (2014). Family with sequence similarity 107: A family of stress responsive small proteins with diverse functions in cancer and the nervous system (Review). *Biomed Rep*, *2*(3), 321–325. https://doi.org/10.3892/br.2014.243

142. Nakajima, H., Koizumi, K., Tanaka, T., Ishigaki, Y., Yoshitake, Y., Yonekura, H., . . . Motoo, Y. (2012). Loss of HITS (FAM107B) expression in cancers of multiple organs: tissue microarray analysis. *Int J Oncol*, *41*(4), 1347–1357. https://doi.org/10.3892/ijo.2012.1550

143. Narod, S. A. (2002). Modifiers of risk of hereditary breast and ovarian cancer. *Nat Rev Cancer*, *2*(2), 113–123. https://doi.org/10.1038/nrc726

144. Narod, S. A., & Lynch, H. T. (2007). CHEK2 mutation and hereditary breast cancer. *J Clin Oncol*, *25*(1), 6–7. https://doi.org/10.1200/JCO.2006.08.8229

145. O'Shea, A. S. (2022). Clinical Staging of Ovarian Cancer. *Methods Mol Biol, 2424*, 3–10. https://doi.org/10.1007/978-1-0716-1956-8_1

146. Ojavee, S. E., Kousathanas, A., Trejo Banos, D., Orliac, E. J., Patxot, M., Läll, K., . . . Robinson, M. R. (2021). Genomic architecture and prediction of censored time-to-event phenotypes with a Bayesian genome-wide analysis. *Nat Commun, 12*(1), 2337. https://doi.org/10.1038/s41467-021-22538-w

147. Oluogun, W. A., Adedokun, K. A., Oyenike, M. A., & Adeyeba, O. A. (2019). Histological classification, grading, staging, and prognostic indexing of female breast cancer in an African population: A 10-year retrospective study. *Int J Health Sci (Qassim), 13*(4), 3–9.

148. Orliac, E. J., Trejo Banos, D., Ojavee, S. E., Läll, K., Mägi, R., Visscher, P. M., & Robinson, M. R. (2022). Improving GWAS discovery and genomic prediction accuracy in biobank data. *Proc Natl Acad Sci U S A, 119*(31), e2121279119. https://doi.org/10.1073/pnas.2121279119

149. Padrik, P., Puustusmaa, M., Tõnisson, N., Kolk, B., Saar, R., Padrik, A., & Tasa, T. (2023). Implementation of Risk-Stratified Breast Cancer Prevention With a Polygenic Risk Score Test in Clinical Practice. *Breast Cancer (Auckl), 17*, 11782234231205700. https://doi.org/10.1177/11782234231205700

150. Pankratov, V., Montinaro, F., Kushniarevich, A., Hudjashov, G., Jay, F., Saag, L., . . . Metspalu, M. (2020). Differences in local population history at the finest level: the case of the Estonian population. *Eur J Hum Genet, 28*(11), 1580–1591. https://doi.org/10.1038/s41431-020-0699-4

151. Parsons, M. T., Tudini, E., Li, H., Hahnen, E., Wappenschmidt, B., Feliubadaló, L., . . . Spurdle, A. B. (2019). Large scale multifactorial likelihood quantitative analysis of BRCA1 and BRCA2 variants: An ENIGMA resource to support clinical variant classification. *Hum Mutat, 40*(9), 1557–1578. https://doi.org/10.1002/humu.23818

152. Patterson, A. D., Gonzalez, F. J., Perdew, G. H., & Peters, J. M. (2018). Molecular Regulation of Carcinogenesis: Friend and Foe. *Toxicol Sci, 165*(2), 277–283. https://doi.org/10.1093/toxsci/kfy185

153. Patxot, M., Banos, D. T., Kousathanas, A., Orliac, E. J., Ojavee, S. E., Moser, G., . . . Robinson, M. R. (2021). Probabilistic inference of the genetic architecture underlying functional enrichment of complex traits. *Nat Commun, 12*(1), 6972. https://doi.org/10.1038/s41467-021-27258-9

154. Pavlovica, K., Irmejs, A., Noukas, M., Palover, M., Kals, M., Tonisson, N., . . . Gardovskis, J. (2022). Spectrum and frequency of CHEK2 variants in breast cancer affected and general population in the Baltic States region, initial results and literature review. *Eur J Med Genet, 65*(5), 104477. https://doi.org/10.1016/j.ejmg.2022.104477

155. Perrin-Vidoz, L., Sinilnikova, O. M., Stoppa-Lyonnet, D., Lenoir, G. M., & Mazoyer, S. (2002). The nonsense-mediated mRNA decay pathway triggers degradation of most BRCA1 mRNAs bearing premature termination codons. *Hum Mol Genet, 11*(23), 2805–2814. https://doi.org/10.1093/hmg/11.23.2805

156. Peters, J. M., & Gonzalez, F. J. (2018). The Evolution of Carcinogenesis. *Toxicol Sci, 165*(2), 272–276. https://doi.org/10.1093/toxsci/kfy184

157. Petrucelli, N., Daly, M. B., & Feldman, G. L. (2010). Hereditary breast and ovarian cancer due to mutations in BRCA1 and BRCA2. *Genet Med, 12*(5), 245–259. https://doi.org/10.1097/GIM.0b013e3181d38f2f

158. Phelan, C. M., Kuchenbaecker, K. B., Tyrer, J. P., Kar, S. P., Lawrenson, K., Winham, S. J., . . . Pharoah, P. D. P. (2017). Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. *Nat Genet, 49*(5), 680–691. https://doi.org/10.1038/ng.3826

159. Pitot, H. C. (1993). The molecular biology of carcinogenesis. *Cancer, 72*(3 Suppl), 962-970. https://doi.org/10.1002/1097-0142(19930801)72:3+<962::aid-cncr2820721303>3.0.co;2-h

160. Plakhins, G., Irmejs, A., Gardovskis, A., Subatniece, S., Rozite, S., Bitina, M., . . . Gardovskis, J. (2011). Genotype-phenotype correlations among BRCA1 4153delA and 5382insC mutation carriers from Latvia. *BMC Med Genet, 12*, 147. https://doi.org/10.1186/1471-2350-12-147

161. Plonis, J., Kalniete, D., Nakazawa-Miklasevica, M., Irmejs, A., Vjaters, E., Gardovskis, J., & Miklasevics, E. (2015). The. *Balkan J Med Genet*, *18*(2), 33–36. https://doi.org/10.1515/bjmg-2015-0083

162. Polyak, K. (2007). Breast cancer: origins and evolution. *J Clin Invest*, *117*(11), 3155–3163. https://doi.org/10.1172/JCI33295

163. Prakash, R., Zhang, Y., Feng, W., & Jasin, M. (2015). Homologous recombination and human health: the roles of BRCA1, BRCA2, and associated proteins. *Cold Spring Harb Perspect Biol*, *7*(4), a016600. https://doi.org/10.1101/cshperspect.a016600

164. Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*, *38*(8), 904–909. https://doi.org/10.1038/ng1847

165. Pujol-Gualdo, N., Läll, K., Lepamets, M., Rossi, H. R., Arffman, R. K., Piltonen, T. T., . . . Laisk, T. (2022). Advancing our understanding of genetic risk factors and potential personalized strategies for pelvic organ prolapse. *Nat Commun*, *13*(1), 3584. https://doi.org/10.1038/s41467-022-31188-5

166. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, *81*(3), 559–575. https://doi.org/10.1086/519795

167. Rakha, E. A., Tse, G. M., & Quinn, C. M. (2023). An update on the pathological classification of breast cancer. *Histopathology*, *82*(1), 5–16. https://doi.org/10.1111/his.14786

168. Rebbeck, T. R., Friebel, T. M., Friedman, E., Hamann, U., Huo, D., Kwong, A., . . . Nathanson, K. L. (2018). Mutational spectrum in a worldwide study of 29,700 families with BRCA1 or BRCA2 mutations. *Hum Mutat*, *39*(5), 593–620. https://doi.org/10.1002/humu.23406

169. Rebbeck, T. R., Mitra, N., Wan, F., Sinilnikova, O. M., Healey, S., McGuffog, L., . . . Andrulis, I. (2015). Association of type and location of BRCA1 and BRCA2 mutations with risk of breast and ovarian cancer. *JAMA*, *313*(13), 1347–1361. https://doi.org/10.1001/jama.2014.5985

170. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., . . . ACMG Laboratory Quality Assurance Committee (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*, *17*(5), 405–424. https://doi.org/10.1038/gim.2015.30

171. Risch, H. A., McLaughlin, J. R., Cole, D. E., Rosen, B., Bradley, L., Kwan, E., . . . Narod, S. A. (2001). Prevalence and penetrance of germline BRCA1 and BRCA2 mutations in a population series of 649 women with ovarian cancer. *Am J Hum Genet*, *68*(3), 700–710. https://doi.org/10.1086/318787

172. Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*, 77. https://doi.org/10.1186/1471-2105-12-77

173. Rodriguez, J. A., Au, W. W., & Henderson, B. R. (2004). Cytoplasmic mislocalization of BRCA1 caused by cancer-associated mutations in the BRCT domain. *Exp Cell Res*, *293*(1), 14–21. https://doi.org/10.1016/j.yexcr.2003.09.027

174. Rovite, V., Wolff-Sagi, Y., Zaharenko, L., Nikitina-Zake, L., Grens, E., & Klovins, J. (2018). Genome Database of the Latvian Population (LGDB): Design, Goals, and Primary Results. *J Epidemiol*, *28*(8), 353–360. https://doi.org/10.2188/jea.JE20170079

175. Rowley, S. M., Mascarenhas, L., Devereux, L., Li, N., Amarasinghe, K. C., Zethoven, M., . . . Campbell, I. G. (2019). Population-based genetic testing of asymptomatic women for breast and ovarian cancer susceptibility. *Genet Med*, *21*(4), 913–922. https://doi.org/10.1038/s41436-018-0277-0

176. Sabiani, L., Barrou, J., Mathis, J., Eisinger, F., Bannier, M., Lambaudie, E., & Houvenaeghel, G. (2020). How to manage BRCA mutation carriers? *Horm Mol Biol Clin Investig*, *41*(3). https://doi.org/10.1515/hmbci-2019-0065

177. Shiovitz, S., & Korde, L. A. (2015). Genetics of breast cancer: a topic in evolution. *Ann Oncol*, *26*(7), 1291–1299. https://doi.org/10.1093/annonc/mdv022

178. Sirico, M., D'Angelo, A., Gianni, C., Casadei, C., Merloni, F., & De Giorgi, U. (2023). Current State and Future Challenges for PI3K Inhibitors in Cancer Therapy. *Cancers (Basel)*, *15*(3). https://doi.org/10.3390/cancers15030703

179. Sivakumaran, S., Agakov, F., Theodoratou, E., Prendergast, J. G., Zgaga, L., Manolio, T., . . . Campbell, H. (2011). Abundant pleiotropy in human complex diseases and traits. *Am J Hum Genet*, *89*(5), 607–618. https://doi.org/10.1016/j.ajhg.2011.10.004

180. Sokolenko, A. P., Bogdanova, N., Kluzniak, W., Preobrazhenskaya, E. V., Kuligina, E. S., Iyevleva, A. G., . . . Imyanitov, E. N. (2014). Double heterozygotes among breast cancer patients analyzed for BRCA1, CHEK2, ATM, NBN/NBS1, and BLM germ-line mutations. *Breast Cancer Res Treat*, *145*(2), 553–562. https://doi.org/10.1007/s10549-014-2971-1

181. Stein, C. J., & Colditz, G. A. (2004). Modifiable risk factors for cancer. *Br J Cancer*, *90*(2), 299–303. https://doi.org/10.1038/sj.bjc.6601509

182. Su, X., Malouf, G. G., Chen, Y., Zhang, J., Yao, H., Valero, V., . . . Esteva, F. J. (2014). Comprehensive analysis of long non-coding RNAs in human breast cancer clinical subtypes. *Oncotarget*, *5*(20), 9864–9876. https://doi.org/10.18632/oncotarget.2454

183. Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660

184. Szabo, C., Masiello, A., Ryan, J. F., & Brody, L. C. (2000). The breast cancer information core: database design, structure, and scope. *Hum Mutat*, *16*(2), 123–131. https://doi.org/10.1002/1098-1004(200008)16:2<123::AID-HUMU4>3.0.CO;2-Y

185. Tamboom, K., Kaasik, K., Aršavskaja, J., Tekkel, M., Lilleorg, A., Padrik, P., . . . Veidebaum, T. (2010). BRCA1 mutations in women with familial or early-onset breast cancer and BRCA2 mutations in familial cancer in Estonia. *Hered Cancer Clin Pract*, *8*(1), 4. https://doi.org/10.1186/1897-4287-8-4

186. Team, RStudio Team. RStudio: Integrated Development Environment for R; RStudio, PBC: Boston, MA, USA, 2020; Available online: http://www.rstudio.com/ (accessed on 27 May 2023).

187. R Core Team. R: A Language and Environment for Statistical Computing; R Foundation for Statistical Computing: Vienna, Austria, 2020; Available online: https://www.R-project.org/ (accessed on 27 May 2023).

188. Therneau. (2020). A Package for Survival Analysis in R. R package version 3.2-3. In. Available online: https://CRAN.R-project.org/package=survival (accessed on 17 October 2023).

189. Thulesius, H. O., Lindgren, A. C., Olsson, H. L., & Håkansson, A. (2004). Diagnosis and prognosis of breast and ovarian cancer--a population-based study of 234 women. *Acta Oncol*, *43*(2), 175–181. https://doi.org/10.1080/02841860310022481

190. Tikhomirova, L., Sinicka, O., Smite, D., Eglitis, J., Hodgson, S. V., & Stengrevics, A. (2005). High prevalence of two BRCA1 mutations, 4154delA and 5382insC, in Latvia. *Fam Cancer*, *4*(2), 77–84. https://doi.org/10.1007/s10689-004-2758-3

191. Turnbull, C., Seal, S., Renwick, A., Warren-Perry, M., Hughes, D., Elliott, A., . . . Rahman, N. (2012). Gene-gene interactions in breast cancer susceptibility. *Hum Mol Genet*, *21*(4), 958–962. https://doi.org/10.1093/hmg/ddr525

192. Uffelmann, E., Huang, Q. Q., Munung, N. S., Vries, J. d., Okada, Y., Martin, A. R., . . . Posthuma, D. (2021). Genome-wide association studies. In (Vol. 1): Nature Reviews Methods Primers.

193. Van den Boom, J., Wolter, M., Blaschke, B., Knobbe, C. B., & Reifenberger, G. (2006). Identification of novel genes associated with astrocytoma progression using suppression subtractive hybridization and real-time reverse transcription-polymerase chain reaction. *Int J Cancer*, *119*(10), 2330–2338. https://doi.org/10.1002/ijc.22108

194. Van der Groep, P., Bouter, A., van der Zanden, R., Siccama, I., Menko, F. H., Gille, J. J., . . . van Diest, P. J. (2006). Distinction between hereditary and sporadic breast cancer on the basis of clinicopathological data. *J Clin Pathol*, *59*(6), 611–617. https://doi.org/10.1136/jcp.2005.032151

195. Van Geldermalsen, M., Wang, Q., Nagarajah, R., Marshall, A. D., Thoeng, A., Gao, D., . . . Holst, J. (2016). ASCT2/SLC1A5 controls glutamine uptake and tumour growth in triple-negative basal-like breast cancer. *Oncogene*, *35*(24), 3201–3208. https://doi.org/10.1038/onc.2015.381

196. Wang, L., Darling, J., Zhang, J. S., Liu, W., Qian, J., Bostwick, D., . . . Smith, D. I. (2000). Loss of expression of the DRR 1 gene at chromosomal segment 3p21.1 in renal cell carcinoma. *Genes Chromosomes Cancer*, *27*(1), 1–10. https://doi.org/10.1002/(sici)1098-2264(200001)27:1<1::aid-gcc1>3.0.co;2-6

197. Wang, Y., Namba, S., Lopera, E., Kerminen, S., Tsuo, K., Läll, K., . . . Wolford, B. N. (2023). Global Biobank analyses provide lessons for developing polygenic risk scores across diverse cohorts. In (Vol. 3, 100241): Cell Genomics.

198. Wang, Y., Song, Z., Zhang, S., Wang, X., & Li, P. (2022). Risk-reducing salpingo-oophorectomy and breast cancer risk in BRCA1 or BRCA2 mutation carriers: A systematic review and meta-analysis. *Eur J Surg Oncol*, *48*(6), 1209–1216. https://doi.org/10.1016/j.ejso.2022.02.019

199. Wang, Y., Zhao, W., Liu, X., Guan, G., & Zhuang, M. (2019). ARL3 is downregulated and acts as a prognostic biomarker in glioma. *J Transl Med*, *17*(1), 210. https://doi.org/10.1186/s12967-019-1914-3

200. Watanabe, K., Taskesen, E., van Bochoven, A., & Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*, *8*(1), 1826. https://doi.org/10.1038/s41467-017-01261-5

201. Weinberg, R. A. (1994). Oncogenes and tumor suppressor genes. *CA Cancer J Clin*, *44*(3), 160–170. https://doi.org/10.3322/canjclin.44.3.160

202. Wiesmueller, F., Kopke, J., Aust, D., Roy, J., Dahl, A., Pilarsky, C., & Grützmann, R. (2019). Silenced ZNF154 Is Associated with Longer Survival in Resectable Pancreatic Cancer. *Int J Mol Sci*, *20*(21). https://doi.org/10.3390/ijms20215437

203. Xue, Y., Mezzavilla, M., Haber, M., McCarthy, S., Chen, Y., Narasimhan, V., . . . Zeggini, E. (2017). Enrichment of low-frequency functional variants revealed by whole-genome sequencing of multiple isolated European populations. *Nat Commun*, *8*, 15927. https://doi.org/10.1038/ncomms15927

204. Yang, W., Zhang, T., Song, X., Dong, G., Xu, L., & Jiang, F. (2022). SNP-Target Genes Interaction Perturbing the Cancer Risk in the Post-GWAS. *Cancers (Basel)*, *14*(22). https://doi.org/10.3390/cancers14225636

205. Ye, M., Li, L., Liu, D., Wang, Q., Zhang, Y., & Zhang, J. (2021). Identification and validation of a novel zinc finger protein-related gene-based prognostic model for breast cancer. *PeerJ*, *9*, e12276. https://doi.org/10.7717/peerj.12276

206. Yoshida, R. (2021). Hereditary breast and ovarian cancer (HBOC): review of its molecular characteristics, screening, treatment, and prognosis. *Breast Cancer*, *28*(6), 1167–1180. https://doi.org/10.1007/s12282-020-01148-2

207. Zamwar, U. M., & Anjankar, A. P. (2022). Aetiology, Epidemiology, Histopathology, Classification, Detailed Evaluation, and Treatment of Ovarian Cancer. *Cureus*, *14*(10), e30561. https://doi.org/10.7759/cureus.30561

208. Zhao, Z., Guo, Y., Liu, Y., Sun, L., Chen, B., Wang, C., . . . Li, X. (2021). Individualized lncRNA differential expression profile reveals heterogeneity of breast cancer. *Oncogene*, *40*(27), 4604–4614. https://doi.org/10.1038/s41388-021-01883-6

209. Zubair, M., Wang, S., & Ali, N. (2020). Advanced Approaches to Breast Cancer Classification and Diagnosis. *Front Pharmacol*, *11*, 632079. https://doi.org/10.3389/fphar.2020.632079

Acknowledgments

Annexes

Approval of the Central Medical Ethics committee of Latvia

Centrālā medicīnas ētikas komiteja

Brīvības iela 72, Rīga, LV-1011 • Tālr. 67876182 • Fakss 67876071 • E-pasts: vm@vm.gov.lv

Rīgā

19.09.2018. Nr.2/18-09-19

Rīgas Stradiņa universitātes
Onkoloģijas institūtam

Atzinums par pētījumu
„BRCA1 gēna patogēno alēlisko variantu
penetranci ietekmējošie faktori"

Centrālā medicīnas ētikas komiteja 2018.gada 5.septembrī ir izskatījusi Rīgas Stradiņa Universitātes Onkoloģijas institūta iesniegto pētījumu *„BRCA1 gēna patogēno alēlisko variantu penetranci ietekmējošie faktori"*.

Pamatojoties uz Centrālās medicīnas ētikas komitejas 2018.gada 5.septembra sēdes protokola Nr.2018-4 punktu Nr.3 un iesniegtajiem labojumiem, tiek izsniegts atzinums, ka Rīgas Stradiņa Universitātes Onkoloģijas institūta iesniegtais pētījums *„BRCA1 gēna patogēno alēlisko variantu penetranci ietekmējošie faktori"* nav pretrunā ar bioētikas normām.

Centrālās medicīnas ētikas
komitejas priekšsēdētājs

V.Sīlis

Strautiņš, 67876190
Edgars.Strautins@vm.gov.lv

Annex of approval of the Central Medical Ethics committee of Latvia

Centrālā medicīnas ētikas komiteja

Brīvības iela 72, Rīga, LV-1011 • Tālr. 67876182 • Fakss 67876071 • E-pasts: vm@vm.gov.lv

Rīgā

Datums skatāms laika zīmogā   Nr. 01-29.1.2/282

Uz ___04.01.2022.___ Nr. _____

Rīgas Stradiņa universitāte
Onkoloģijas institūts
Egijai Bergai-Švītiņai
Egija.Berga-Svitina@rsu.lv

Par atļaujas saņemšanu pētījumam
"BRCA1 gēna patogēno alēlisko
variantu penetranci ietekmējošie faktori"

Centrālā medicīnas ētikas komiteja izskatīja Rīgas Stradiņa universitātes Onkoloģijas institūta papildinājumu pētījumam *"BRCA1 gēna patogēno alēlisko variantu penetranci ietekmējošie faktori"* (reģistrēts Veselības ministrijā 2021.gada 15.novembrī Nr.16588) ar precizējumiem (reģistrēts Veselības ministrijā 2022.gada 4.janvārī Nr.101).

Atbilstoši Centrālā medicīnas ētikas komitejas 2022.gada 18.janvāra sēdes protokola Nr.2022-15 II.daļas *"PRECIZĒTIE IESNIEGUMI"* 4.punktam tiek sniegts atzinums, ka pētījums nav pretrunā bioētikas normām.

Vienlaikus aicinām informētās piekrišanas veidlapu papildināt ar informāciju par pētījuma ietvaros paredzēto sadarbību ar Igaunijas Genoma centru, kā arī, ievērojot datu drošības aspektus, norādīt datu pārziņa kontaktinformāciju, ietverot pētāmās personas piekrišanu datu apstrādei un informāciju par iegūto paraugu uzglabāšanas ilgumu, kā arī norādām, ka informēto piekrišanu paraksta pētījuma veicējs. Papildus aicinām protokola 9.4.punktā veikt korektu atsauci uz normatīvo aktu, saskaņā ar kuru pētījuma ietvaros tiks veikta bioloģisko audu paraugu utilizācija.

Centrālās medicīnas ētikas
komitejas priekšsēdētājs   (paraksts*)   V. Sīlis

Kristīne Kalniņa   67876116
kristine.kalnina@vm.gov.lv

*Dokuments ir parakstīts ar drošu elektronisko parakstu un satur laika zīmogu

Approval of the Genome Research Council of Latvia

Genoma izpētes padome

Rātsupītes iela 1 k-1, Rīga, LV-1067 • Tālr. 67473083 • E-pasts: genoma.padome@biomed.lu.lv

Rīgā

19.10.2018. Nr.A-1/18-10-19

RSU Onkoloģijas Institūtam

Atzinums par pētījumu: *"BRCA1 gēna alēlisko patogēno variantu penetranci ietekmējošie faktori"*

Genoma izpētes padome izskatīja RSU Onkoloģijas Institūta iesniegumu par pētījumu *"BRCA1 gēna alēlisko patogēno variantu penetranci ietekmējošie faktori"*.

Pamatojoties uz Genoma izpētes padomes locekļu balsojumu tiek izsniegts atzinums, ka Genoma izpētes padome atbalsta RSU Onkoloģijas Institūta pētījuma *"BRCA1 gēna alēlisko patogēno variantu penetranci ietekmējošie faktori"* īstenošanu.

Genoma izpētes padomes
Priekšsēdētājs   J.Kloviņš

Rovite, 67473083
vita.rovite@biomed.lu.lv

ŠIS DOKUMENTS IR ELEKTRONISKI PARAKSTĪTS AR DROŠU ELEKTRONISKO PARAKSTU UN SATUR LAIKA ZĪMOGU

Approval of the Rīga Stradiņš University

RĪGAS STRADIŅA
UNIVERSITĀTE

RĪGAS STRADIŅA UNIVERSITĀTE
Reģistrācijas Nr. 90000013771
Dzirciema 16, Rīga, LV-1007, Latvija
Tālr. 67409230, fakss 67471815
E-pasts: rsu@rsu.lv, www.rsu.lv

Latvijas Republikas
Centrālajai medicīnas ētikas komitejai

07 -09- 2018 Nr. 51-7 /200/2018

Rīgas Stradiņa universitātes
Prorektores prof. Tatjanas Koķes

atzinums.

Atbalstu RSU Onkoloģijas institūta Molekulārās ģenētikas laboratorijas iesaistīšanos pētījumā „*BRCA1 gēna patogēno alēlisko variantu penetranci ietekmējošie faktori*".

Rīgas Stradiņa universitāte
Rektora p.i.
profesore
TATJANA KOĶE

Prorektore   T.Koķe

Sagatavoja: E.Berga-Švītiņa
29865861

Approval of the Pauls Stradiņš Clinical University Hospital

VSIA „Paula Stradiņa klīniskā universitātes slimnīca"
Zinātniskā institūta direktoram
Addressee: "Pauls Stradins Clinical University Hospital"
Director of Research Institute

IESNIEGUMS AKADĒMISKĀ PĒTĪJUMA ATĻAUJAS SAŅEMŠANAI /
APPLICATION FOR AUTHORIZATION OF ACADEMIC RESEARCH (THESIS) IN HOSPITAL.

03 / 08 / 2018
DD MM YYYY

IESNIEGUMA IESNIEDZĒJS / PĒTĪJUMA AUTORS VAI AUTORU KOLEKTĪVA VADĪTĀJS
APPLICANT/AUTHOR OR HEAD OF AUTHOR'S GROUP

Vārds, uzvārds / *Applicant name surname*	Egija Berga - Švītiņa		
Tālr. Nr. / *phone No.*	29865864	*e-pasts/e-mail*	Egija.Berga-Svitina@rsu.lv
Izglītības iestādes nosaukums / *University RSU or LU* *	RSU		
Studiju programma / *Study programme* *	Medicīna		

PĒTĪJUMA ZINĀTNISKAIS VADĪTĀJS* / *Research/Thesis supervisor* Profesors -

	Edvīns Miklaševičs
Vārds, uzvārds/ *name surname*	
Izglītības iestādes nosaukums, amats / *University: RSU or LU*	RSU Onkoloģijas institūta direktors
Zinātniskā vadītāja paraksts, datums / *supervisor signature and date*	

PĒTĪJUMA NOSAUKUMS / *THE TITLE OF THESIS*

BRCA1 gēna patogēno alēlisko variantu penetranci ietekmējošie faktori.

PĒTĪJUMA NORISES VIETA / *THE UNIT OF HOSPITAL WHERE RESEARCH WILL TAKE PLACE*

Klīnika/centrs/nodaļa *CLINIC/CENTRE/UNIT*	KRŪTS SLIMĪBU CENTRS
Klīnikas/centra/nodaļas vadītāja saskaņojums par pētījuma veikšanu *HEAD OF CLINIC/CENTRE/UNIT*	
Slimnīcas darbinieka- galvotājs** paraksts, datums / *guarantor** – employee of Hospital*	

NORISES LAIKS/ *PERIOD OF TIME OF RESEARCH*

Pētījuma norises laiks, norādot sākuma un beigu datumu *PERIOD OF TIME OF RESEARCH / start and end date*	09.2018. - 06.2021.

IZMANTOJAMĀS PACIENTU SLIMĪBAS VĒSTURES / *MEDICAL RECORDS TO BE USED*

Pētījumā paredzēts izmantot/ *NUMBER OF MEDICAL RECORDS PLANNED TO BE REVIEWED* ≤ 100 □ ≥ 100*** ☒

Slimības vēsturu diagnožu kodi (SSK-10) vai Slimības vēsturu numuri ja zināmi / ICD-10 codes and/or No. of medical records if known

ICD-10	C50

* Šo sadaļu aizpilda studenti, maģistranti un doktoranti / *to be filled by students, graduate students, doctoral students*

** ja iesnieguma iesniedzējs pats vai zinātniskā darba vadītājs nav Slimnīcas darbinieks, nepieciešams galvotājs (Slimnīcas darbinieks, firstniecības persona, kura ir atbildīga par pētniecības norisi un juridiskajiem sekām)/ *if applicant and/or supervisor is not employee of Hospital application has to be signed by guarantor – employee of Hospital*

*** ja paredzēts izmantot vairāk kā 100 Slimības vēstures, iesniegumam nepieciešams pievienot rakstisku pamatojumu brīvā formā/ *if there is planned to use more than 100 medical records please submit written rationale.*

03.08.2018.

Datums/Date   /Egija berga-Švītiņa/ Švītiņa
Authorization of Hospital   Iesniedzēja vārds, uzvārds, paraksts/ Applicant name Surname and signature

07.08. 2018

_____
Datums/ Date   Zinātniskā institūta kompetentā darbinieka paraksts un atšifrējums/ On behalf of Hospital's
          Research Institute, position, name surname, signature

Zin-1-006 versija 02

Template of informed consent for study participant

Pacienta piekrišana dalībai pētījumā

1. Esmu saņēmis un iepazinies ar rakstisku informāciju par daudzpusīgu vēža izpētes projekta mērķi, saturu un iespējamiem riskiem. Uz visiem maniem jautājumiem esmu saņēmis izsmeļošas atbildes. Man bija pietiekami daudz laika, lai pārdomātu savu lēmumu piekrist nodot bioloģisko materiālu vēža izpētei.

2. Es piekrītu bez atlīdzības nodot asinis un/vai audu paraugus (bioloģisko materiālu) šūnu iegūšanai ar mērķi tās identificēt, kā arī nukleīnskābju un/vai proteīnu izdalīšanai molekulāri – ģenētiskiem pētījumiem. Es saprotu, ka asins vai audu paraugu nodošana nerada risku manai veselībai, kas man ir izskaidrots.

3. Asins un/vai audu paraugi, ko nododu izpētei, ir nepieciešami, lai noskaidrotu labākās šūnu izdalīšanas metodes, identificētu vēža šūnas, kā arī izdalītu nukleīnskābes (DNS; RNS) un/vai proteīnus, ar ko veikt molekulāri ģenētiskās analīzes.

4. Es saprotu, ka jebkura mani identificējoša informācija būs konfidenciāla, un ka visi mani paraugi būs kodēti. Es apzinos, ka es jebkurā brīdī bez paskaidrojumiem varu pārtraukt piedalīšanos pētījumā, zinot, ka tas neietekmēs manu turpmāko ārstēšanos. Zinu, ka šādā gadījumā mani nodotie asins un/vai audu paraugi, izdalītās nukleīnskābes un/vai proteīni, veselības stāvokļa apraksts un jebkura mani identificējoša informācija tiks iznīcināta.

5. Veselības stāvokļa apraksti:

   ☐ Es ATĻAUJU sava veselības stāvokļa apraksta papildināšanu, atjaunošanu vai pārbaudi:

   ☐ Es AIZLIEDZU sava veselības stāvokļa apraksta papildināšanu, atjaunošanu vai pārbaudi:

6. Gadījumā, ja mana bioloģiskā materiāla izpētē tiks atklāta informācija par kādu man līdz šim nezināmu apdraudējumu manai un / vai manu radinieku veselībai (vajadzīgo atzīmēt):

   ☐ Es PIEKRĪTU, ka man tiek paziņota šī informācija;

   ☐ Es PIEKRĪTU, ka man tiek paziņota šāda informācija tikai tādā gadījumā, ja risks veselībai ir novēršams;

   ☐ Es NEVĒLOS saņemt nekādu papildus informāciju.

7. Bioloģiskā materiāla uzglabāšana:

   ☐ Es PIEKRĪTU, ka mans bioloģiskais materiāls turpmāk glabāsies RSU Onkoloģijas institūta Molekulārās ģenētikas laboratorijā un tiks izmantots pētījumos, kas saistīti ar vēža ģenētisko izpēti bez ierobežojuma;

1 no 2

111

☐ Es PIEKRĪTU mana bioloģiskā materiāla un veselības stāvokļa apraksta nosūtīšanai izpētei ārpus Latvijas;

☐ Es NEPIEKRĪTU mana bioloģiskā materiāla un veselības stāvokļa apraksta nosūtīšanai izpētei ārpus Latvijas;

Bioloģiskā materiāla donors (nepieciešamības gadījumā viņa aizbildnis vai aizgādnis):

Vārds un uzvārds (drukātiem burtiem):

_____

Personas kods: ☐☐☐☐☐☐ — ☐☐☐☐☐

Paraksts: _____

Adrese: _____

Datums: __ __ / __ __ / 20__   (DD/MM/GGG)

Ārsts/klīnicists:

Vārds un uzvārds, amats (drukātiem burtiem vai spiedogs):

_____

Datums: __ __ / __ __ / 20__   (DD/MM/GGG)

Apstiprinu, ka esmu informējis pacientu par šo pētījumu (paraksts):

_____

Pacientam piešķirtais kods   ☐☐☐☐☐☐☐

Ja Jums rodas kādas komplikācijas bioloģisko paraugu noņemšanas dēļ, lūdzu nekavējoties sazināties ar ārstu, kurš šos paraugus noņēma vai Urzulu Lakuču (tālr. 67704028, e-pasts: urzula.lakuca@rsu.lv).

2 no 2

Explanation of the study for participants

PĒTĪJUMA SKAIDROJUMS

Pētījuma īstenotājs: Rīgas Stradiņa universitāte, Onkoloģijas institūts.

Lūdzu rūpīgi izlasiet doto informāciju un, ja vēlaties, apspriediet to ar draugiem un radiniekiem. Lūdziet ārstam paskaidrot nesaprotamos vārdus vai sniegt papildus informāciju.

Pētījuma mērķis ir daudzpusīgi pētīt vēzi pēc to klīniskajām īpašībām un izmaiņām gēnos, kas dos iespēju nodrošināt precīzu un personalizētu riska izvērtējumu katram pacientam. Tā kā galvenais riska faktors pārmantotā krūts un olnīcu vēža gadījumā ir patogēna mutācija *BRCA1* gēnā, tad pētījuma ietvaros tiks apkopota un analizēta informācija par genoma izmaiņām, lai noskaidrotu pārmantošanas ceļus un tos ietekmējošos faktorus.

Pētījumā tiks izmantots no venozajām asinīm izdalīta DNS. Asinis tiks iegūtas diagnostiskos nolūkos veikto manipulāciju (asins paņemšanas) rezultātā, kas ietilpst vēža pacientu diagnostikas plānā. No iegūtā bioloģiskā materiāla izdalīs dezoksiribonukleīnskābes (DNS), kuras tālāk izmantos molekulāri ģenētiskās analīzēs.

Vienīgais risks, kas iespējams pacientam piedaloties pētījumā, ir saistīts ar asins parauga paņemšanu no vēnas. Asins ņemšanas vietā iespējams pārejošs diskomforts vai sāpes, var izveidoties zemādas asinsizplūdums, reti – infekcija. Procedūra tiks veikta ārstniecības iestādē, procedūru veiks sertificēts medicīnas personāls.

Pētījumā iegūtā informācija palīdzēs identificēt un raksturot tās genoma izmaiņas, kas ietekmē ar *BRCA1* patogēno mutāciju saistītā pārmantotā krūts vai olnīcu vēža risku. Balsoties uz šādām zināšanām, būs iespējams objektīvi konsultēt pacientu par piemērotāko profilakses pasākumu iespējām.

Kas ar mani notiks, ja piekritīšu piedalīties pētījumā?

Piekrītot piedalīties pētījumā:
- Tiks veikta slimību diagnosticējoša manipulācija, kuras laikā paredzēta vienreizēja 2-3 ml asins parauga paņemšana no intravenoza katetra vai ar sterilu adatu.

Ja vēlaties saņemt vairāk informācijas, lūdzu sazināties ar Urzulu Lakuču (tālr. 67704028, e-pasts: urzula.lakuca@rsu.lv).

Supplementary Table 1

Top associated variants with BC development risk

rsID	Chr	Position	REF	ALT	MAF_ALT	*p* value	Beta	SE	Nearest gene	Position
rs2609813	10	14800320	A	G	0.07952	2.33203991876e-07	-1.26	0.24	*FAM107B*	intronic
rs4688094	3	118003477	G	C	0.4523	7.76098782977e-07	-0.96	0.19	*RP11-384F7.1*	ncRNA_intronic
rs16951137	13	87833137	G	A	0.04274	1.37428743363e-06	-3.21	0.66	*UBBP5*	intergenic
rs62078923	18	9712033	G	A	0.0169	1.80836291877e-06	-3.86	0.81	*RAB31*	intronic
rs7035512	9	125433467	G	A	0.05666	1.93840164663e-06	-2.94	0.62	*OR1L3*	intergenic
rs434451	19	47328835	T	C	0.03479	2.05000432988e-06	2.59	0.54	*CTB-147N14.4*	intergenic
rs4326074	4	112496443	C	T	0.329	2.53520922997e-06	-0.93	0.20	*RP11-255I10.1*	intergenic
rs148017297	2	92278832	CAG	C	0.09742	2.81781609197e-06	-1.38	0.30	*IGKV1OR2-2*	intergenic
rs75850861	11	133448830	G	A	0.04374	3.19089930542e-06	-1.86	0.40	*OPCML*	intergenic
rs10773765	12	130767334	T	C	0.2664	3.60009934859e-06	1.08	0.23	*RP11-143E21.6*	intergenic
rs1151446	13	112128860	A	G	0.01292	4.0385563336e-06	3.09	0.67	*RP11-65D24.2*	intergenic
rs11690984	2	138125564	A	G	0.492	5.75794208576e-06	-0.93	0.21	*THSD7B*	intronic
rs4625963	2	69567461	G	A	0.1332	6.36694772388e-06	1.36	0.30	*GFPT1*	intronic
rs4241799	4	186401205	C	T	0.05964	6.82529320624e-06	2.40	0.53	*RP11-279O9.4*	intergenic
rs115746414	4	78115381	G	A	0.02286	7.1085584512e-06	-3.51	0.78	*CCNG2*	intergenic
rs140590760	18	21636355	G	A	0.01491	7.10879642189e-06	-3.16	0.70	*TTC39C*	intronic
rs10178186	2	95467255	C	T	0.1044	7.79155565451e-06	-1.31	0.29	*AC073464.11*	ncRNA_intronic
rs76379709	3	118296217	T	C	0.07753	8.95955880288e-06	-1.72	0.39	*RP11-384F7.1*	intergenic
rs139811148	6	118781263	G	A	0.1412	9.24630471659e-06	-1.22	0.27	*CEP85L*	downstream

BC – breast cancer; OC – ovarian cancer; rsID – reference SNV ID number; Chr – chromosome; REF – reference allele; ALT – alternative allele; MAF – minor allele frequency; Beta – multivariate linear regression coefficient; SE – standard error

Supplementary Table 2

Top associated variants with OC development risk

rsID	Chr	Position	REF	ALT	MAF_ALT	*p* value	Beta	SE	Nearest gene	Position
rs79732499	20	3404208	G	T	0.01789	1.37593865968e-07	-8.09	1.53	*C20orf194*	intergenic
rs139549679	23	16322198	T	C	0.02187	1.53087007856e-06	-9.27	1.93	*AC078993.1*	intergenic
rs148380651	11	20985559	C	A	0.008946	2.0087464698e-06	-4.16	0.88	*NELL1*	intronic
rs115514245	8	83785903	TG	T	0.3082	2.13896909542e-06	-1.33	0.28	*RP11-731N10.1*	ncRNA_intronic
rs141885477	18	64012510	C	T	0.1551	2.30370882881e-06	-1.60	0.34	*CDH19*	intergenic
rs9874498	3	181939180	T	C	0.1342	2.38852906941e-06	-1.75	0.37	*RP11-416O18.2*	intergenic
rs117285553	20	1864465	T	G	0.04175	2.88606187243e-06	-6.77	1.45	*SIRPA*	intergenic
rs73107837	3	70526995	A	G	0.1113	3.15301948301e-06	1.86	0.40	*RP11-231I13.2*	intergenic
rs192860697	3	143765548	C	G	0.001988	3.25692955046e-06	-3.80	0.82	*C3orf58*	intergenic
rs118143913	19	53074608	T	C	0.04374	4.31023600453e-06	-4.07	0.89	*ZNF701*	intronic
rs28544080	10	17542126	G	A	0.03181	4.36706963041e-06	-3.47	0.76	*ST8SIA6*	intergenic
rs188926167	11	83570558	G	C	0.005964	4.37685115834e-06	-4.28	0.93	*DLG2*	intronic
rs6686539	1	164131343	A	T	0.08847	4.62336118364e-06	-2.12	0.46	*NMNAT1P2*	intergenic
rs75438760	7	154915632	C	T	0.07952	5.91141189815e-06	-2.20	0.49	*AC092628.3*	intergenic
rs149406877	2	84563836	G	C	0.005964	7.86358139471e-06	-3.87	0.87	*AC106874.1*	ncRNA_intronic
rs12212198	6	18891857	A	G	0.4771	7.98851333955e-06	1.06	0.24	*RP11-254A17.1*	intergenic
rs72884678	2	168102774	G	T	0.03976	8.11894142429e-06	-2.65	0.59	*XIRP2*	exonic
17:43728452:C:T	17	43728452	C	T	0.000994	8.14659041686e-06	-7.41	1.66	*CRHR1:RP11-105N13.4*	ncRNA_intronic
rs75135717	1	238700592	T	G	0.02485	8.57460947684e-06	-4.18	0.94	*RP11-177F15.1*	ncRNA_exonic
rs11177790	12	70077983	G	A	0.1044	9.39211224177e-06	1.68	0.38	*BEST3*	UTR3
rs139132019	11	20864338	A	T	0.002982	9.5157833902e-06	-4.75	1.07	*NELL1*	intronic
rs75030318	1	233701975	A	G	0.02584	9.82981159967e-06	-2.77	0.63	*KCNK1*	intergenic

BC – breast cancer; OC – ovarian cancer; rsID – reference SNV ID number; Chr – chromosome; REF – reference allele; ALT – alternative allele; MAF – minor allele frequency; Beta – multivariate linear regression coefficient; SE – standard error