doi:10.25143/prom-rsu_2024-15_dts

Egija Berga-Švītiņa

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

Summary of the 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

Summary of the 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

The Doctoral Thesis was developed at Institute of Oncology and Molecular Genetics, Rīga Stradiņš University, Latvia

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

Official Reviewers:

Dr. med., Associate Professor **Madara Auzenbaha**, Department of Biology and Microbiology, Rīga Stradiņš University, Latvia

PhD **Raitis Pečulis**, Latvian Biomedical Research and Study Centre, Latvia

MD, PhD, Professor **Neeme Tõnisson**, Estonian Biobank; Institute of Genomics, University of Tartu, Estonia

Defence of the Doctoral Thesis will take place at the public session of the Promotion Council of the Basic Sciences of Medicine, Including Pharmacy on 20 September 2024 at 10.00 remotely via online platform Zoom.

The Doctoral Thesis is available in RSU Library and on RSU website: https://www.rsu.lv/en/dissertations

This research has been developed with financing from the European Social Fund and Latvian state budget within the project No 8.2.2.0/20/I/004 "Support for involving doctoral students in scientific research and studies" at Rīga Stradiņš University.

Secretary of the Promotion Council:

Dr. sc. soc., Associate Professor **Andrejs Ivanovs**

Table of Contents

Abbreviations used in the Thesis

ACMG	American College of Medical Genetics and Genomics
AIM	ancestry informative marker
ALT	alternative allele
ARL3	ADP ribosylation factor like GTPase 3
AUC	area under the curve
BC	breast cancer
BCAC	Breast Cancer Association Consortium
BRCA1	BRCA1 DNA repair associated
BRCA2	BRCA2 DNA repair associated
BRRM	bilateral risk-reducing mastectomy
BRRSO	bilateral risk-reducing salpingo-oophorectomy
CDPC	Centre for Disease Prevention and Control
CHEK2	checkpoint kinase 2
CI	confidence interval
CIMBA	Consortium of Investigators of Modifiers of *BRCA1/2*
DNA	deoxyribonucleic acid
DNAAF9	dynein axonemal assembly factor 9
DR2	dosage R-squared
eQTL	expression quantitative trait loci
FAM107B	family with sequence similarity 107 member B
FDR	false discovery rate
FUMA	functional mapping and annotation
GRCh37	genome reference consortium human build 37
GTEx	genotype-tissue expression project
GWAS	genome-wide association study
HBOC	hereditary breast and ovarian cancer

HR	hazard ratio
HWE	Hardy-Weinberg equilibrium
IBD	identity by descent
LD	linkage disequilibrium
LGDB	genome database of Latvian population
lncRNA	long non-coding RNA
LoF	loss-of-function
MAF	minor allele frequency
NES	normalised effect size
OC	ovarian cancer
OCAC	Ovarian Cancer Association Consortium
OR	odds ratio
PC	principal component
PCR	polymerase chain reaction
PRS	polygenic risk score
PV	pathogenic variant
QC	quality control
REF	reference allele
ROC	receiver-operating characteristic
SD	standard deviation
SE	standard error
SLC1A5	solute carrier family 1 member 5
SNV	single nucleotide variant
TNBC	triple-negative breast cancer
VEP	variant effect predictor
WES	whole exome sequencing
WGS	whole genome sequencing
ZNF514	zinc finger protein 514

Introduction

According to GLOBOCAN 2020, breast cancer (BC) is the most frequently diagnosed cancer in females, contributing to 15 % of cancer-related deaths worldwide, with approximately 522,000 reported deaths. Additionally, ovarian cancer (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 pathogenic variants (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 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 genome-wide association study (GWAS) approach. Additionally, we explored and compared the efficiency of two recently developed genome-wise polygenic risk score (PRS) models (BayesW vs. BayesRR-RC) to estimate the overall genetic risk in women carrying these two most frequently identified germline *BRCA1* PVs. 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.

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 Materials and methods

1.1 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). Study participants were clinical cohort recruited continuously between 2002 and 2022, who were ≥ 18 years old and underwent germline genetic testing for hereditary breast and ovarian cancer (HBOC) syndrome at the Breast Surgery Unit of the Pauls Stradiņš Clinical University Hospital. 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 bilateral risk-reducing mastectomy (BRRM) or bilateral risk-reducing salpingo-oophorectomy (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 loss-of-function (LoF) of the protein.

Experimental setup and data analysis workflow are presented in the Figure 1.1.

This research received ethical approval from the Central Medical Ethics Committee of Latvia under protocol No 2/18-09-19, with supplement No 01-29.1.2/282. Additionally, approval was granted by the Genome Research Council under protocol No A-1/18-10-19. The use of Estonian reference data

was authorised through approval No 1.1-12/624, along with amendment No 1.1-12/1478 by the Estonian Committee on Bioethics and Human Research (Estonian Ministry of Social Affairs).

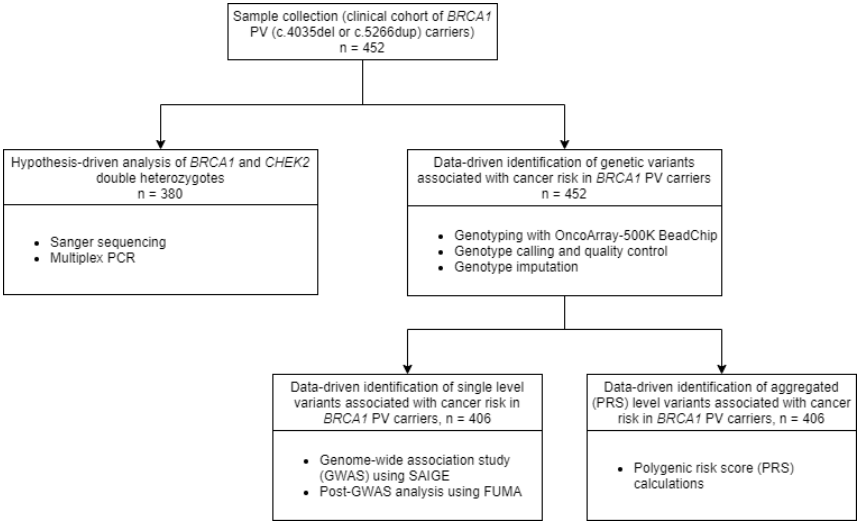

Figure 1.1 **Diagram of analysis workflow presented in this Doctoral Thesis**

Each participant who enrolled in this study gave a written informed consent for the use of their clinical and genomic information for research purposes.

1.2 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 1.1). 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 used 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).

1.3 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 single nucleotide variants (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), Consortium of Investigators of Modifiers of *BRCA1/2* (CIMBA), and Ovarian Cancer Association Consortium (OCAC).

1.4 Genotype calling and quality control (QC)

A modified genotype quality control (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).

First, sample genotype calling was done using GenomeStudio software, which performed automatic clustering. Individuals with a call-rate below 98 % or mismatched sex data were manually excluded, and variant calls with a GenTrain score below 0.7 were inspected and re-clustered before final export to PLINK format.

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.

To ensure data quality, several QC steps were conducted: gender mismatch was checked using inbreeding estimates for the X chromosome using PLINK command *--check-sex*, retaining females with an inbreeding estimate < 0.2. Race mismatch was assessed through principal component (PC) analysis using EIGENSOFT software (Price et al., 2006) and 687 ancestry informative markers (AIMs), with a threshold of > mean ± 6 standard deviation (SD) for race mismatch. Relatedness and duplicates were identified by pair-wise identity by descent (IBD) calculations after linkage disequilibrium (LD) pruning to enhance

marker independence, excluding probable duplicates with PI_HAT value close to 1. Hardy-Weinberg equilibrium (HWE) outliers were identified using a *p* value threshold of $< 1 \times 10^{-7}$ for unaffected individuals and $< 1 \times 10^{-12}$ for cases, excluding 503 SNVs. Samples with extreme heterozygosity (± 4.89 SD from the mean) and inbreeding coefficient > 0.1 were also excluded, leaving 406 samples for analysis.

1.5 Genotype imputation

For the imputation, additional SNVs with minor allele frequency (MAF) < 0.01 were excluded. Missing genotypes were imputed using the Estonian population based high coverage whole genome sequencing (WGS) dataset (n = 2,244) 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.

1.6 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 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 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 $p < 1 \times 10^{-6}$.

1.7 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 used (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.

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.

Additionally, 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.

1.8 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.

1.9 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 odds ratios (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 % confidence interval (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).

1.10 Data availability

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

2 Results

2.1 Study cohort characteristics

2.1.1 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 2.1. 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.

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.

Table 2.1

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.

2.1.2 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 2.1.

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 2.2. 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 2.2

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.

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

Next, 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.

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

Additional statistical tests consistently confirmed the significance of this association. The likelihood ratio, Wald, and 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* PV c.5266dup is a statistically significant predictor of earlier cancer onset.

In the subsequent analysis, we explored the impact of the *BRCA1*:c.5266dup PV on BC and OC groups individually (see Figure 2.1).

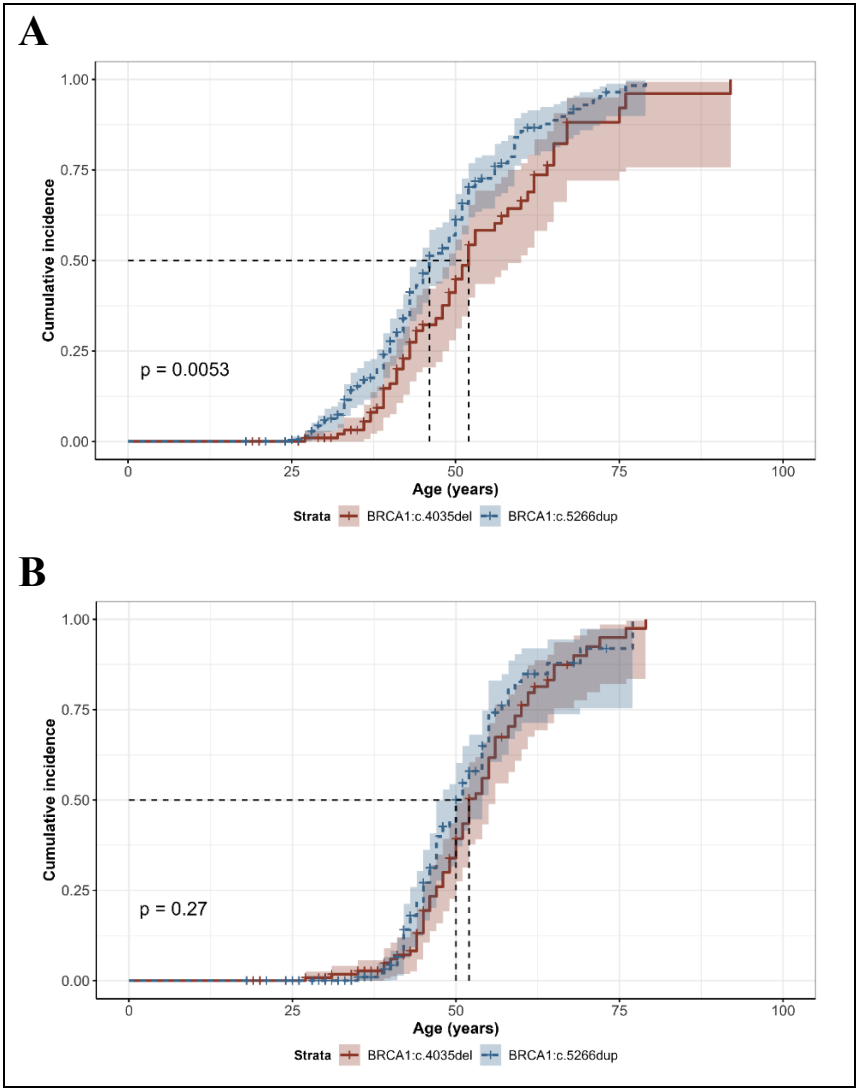

Figure 2.1 **Cumulative incidence of BC or OC in *BRCA1* PV carriers**

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

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 PV and the age of cancer onset.

The HR for the *BRCA1*:c.5266dup PV was 1.564 (95 % CI: 1.14–2.15), indicating a 56.40 % higher hazard of developing BC compared to individuals with another *BRCA1* PV (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 PV and the age of cancer onset in this group was not statistically significant. The HR for the *BRCA1*:c.5266dup PV 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 PV has a significant impact on the age of OC onset in this particular study. These distinct results suggest that the *BRCA1*:c.5266dup PV plays a significant role in the age of cancer onset within the BC group, but its impact is less evident in the OC group.

2.2 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 hypothesis while also exploring unexpected patterns and associations within the dataset.

2.2.1 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 2.3.

Table 2.3

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.

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; $p = 1$).

The impact of a specific *CHEK2* variant on the age at cancer onset was not consistent. Among carriers of the *BRCA1*:c.4035del PV, 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 PV 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 PV 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, $p = 0.064$).

2.2.2 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 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 2.2 and later detailed in following chapter, within Table 2.4 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 2.2A). However, in the OC group, chromosome 20 exhibited the most significant suggestive association, as shown in Figure 2.2B.

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 (not shown in this summary). 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.

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 contained 27 independent significant SNVs in the BC group and 25 independent 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, we identified 1152 candidate SNVs in BC and 633 in OC that exhibited LD with previously mentioned independent significant SNVs.

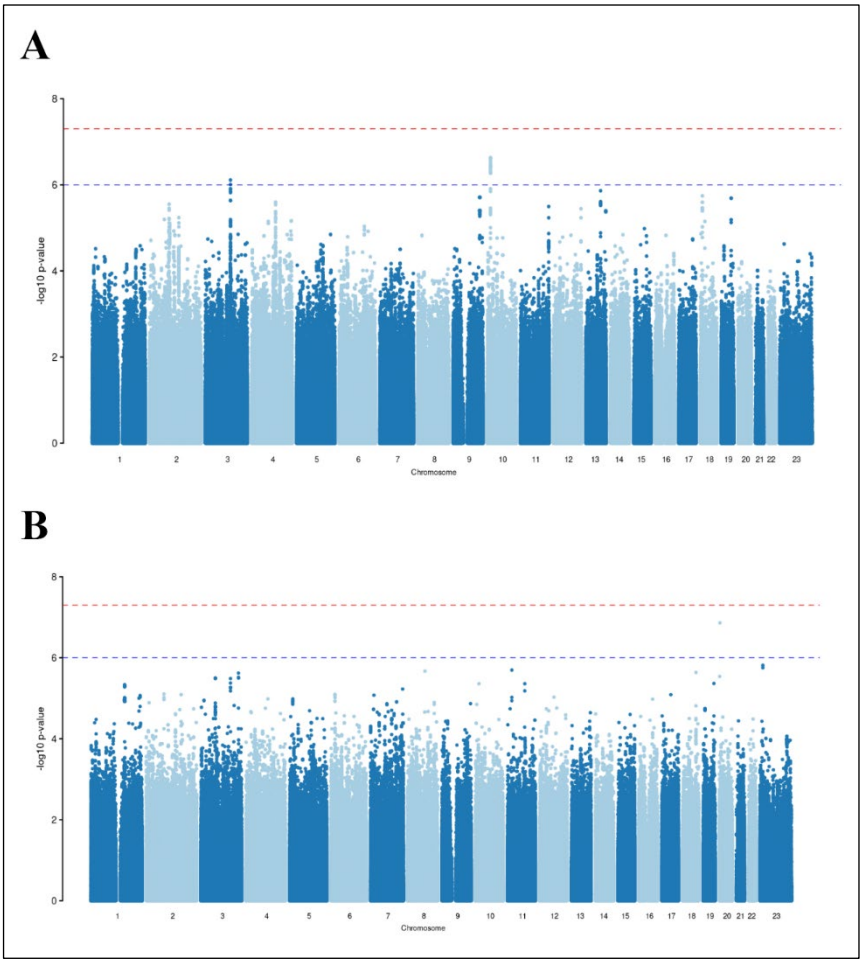

Figure 2.2 **SNV association with BC or OC development risk**

A) Manhattan plot visualising -$\log_{10}p$ values for SNV associations with BC development risk. B) Manhattan plot visualising -$\log_{10}p$ 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.

Table 2.4 highlights three most significant ($p < 1 \times 10^{-7}$) genetic variants that were associated with the risk of developing BC or OC.

Table 2.4

Top associated variants with BC or OC development risk

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

BC – breast cancer; OC – ovarian cancer; rsID – reference SNV ID number; 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).

Our identified 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.

The strongest association with BC development risk was observed for rs2609813 variant (beta = -1.26; $p = 2.33 \times 10^{-7}$; risk allele G frequency = 0.08). Detailed information is available in Table 2.4. This 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.

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 2.4. The rs4688094 variant is situated on chromosome 3 and is located within the novel long non-coding RNA (lncRNA) *RP11-384F7.1* (ENSG00000243276), which exhibits high LD with 295 other SNVs.

The only variant that reached genome-wide suggestive significance of $p < 1 \times 10^{-7}$ 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 2.4). It 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). 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).

Next, we performed 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 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 FDR threshold of less than 0.05. However, we observed two significant eQTL values in the BC group (see Table 2.5).

The most significant association was identified 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 2.5). This variant was 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. The normalised effect size (NES) of −0.36 indicates a negative association between the rs10178186 variant and *ZNF514* gene expression.

Table 2.5

The eQTL results in breast tissue

rsID	REF	ALT	MAF	*p* value	FDR	NES	Nearest gene
rs10178186	C	T	0.10	3.83×10^{-7}	1.55×10^{-16}	−0.36	*ZNF514*
rs434451	T	C	0.035	2.90×10^{-6}	0.011	−0.42	*SLC1A5*

rsID – reference SNV ID number; REF – reference allele; ALT – alternative allele; MAF – minor allele frequency; FDR – false discovery rate; NES – normalised 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).

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 2.5). Intriguingly, this variant was the sole variant mapped to the protein coding gene *SLC1A5* (Solute Carrier Family 1 Member 5) (ENSG00000105281) on chromosome 19. The NES of −0.42 underscores a negative association between the rs434451 variant and the expression of *SLC1A5* gene.

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

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 2.6.

Table 2.6

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 region-specific *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).

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 2.7. Regardless of the specific PRS employed, none of the models exhibited a statistically significant association with the risk of OC development ($p > 0.05$), as presented in Table 2.7.

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 2.3). 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.

Table 2.7

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 2.6). * *p* value below 0.05; ** *p* value below 0.01.

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).

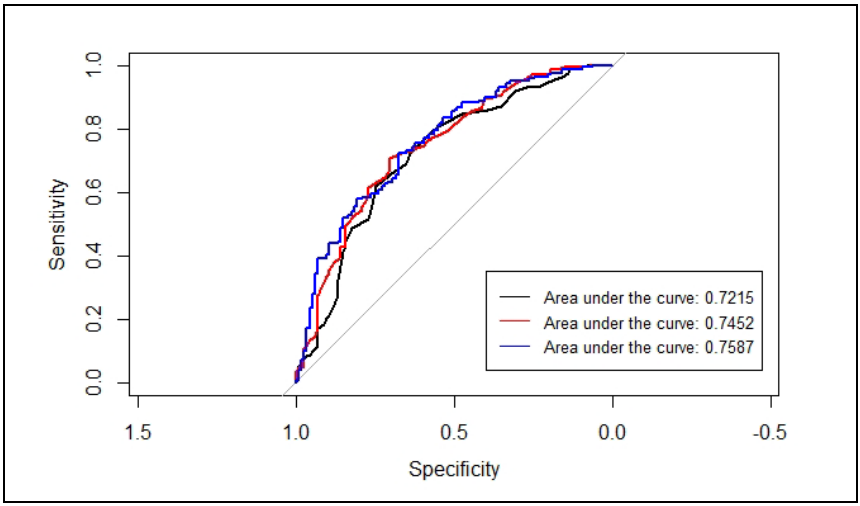

Figure 2.3 **A Comparison of the AUC (area under the receiver operating characteristic curve) to select the most optimal binomial logistic regression analysis model**

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).

3 Discussion

3.1 Main findings in the study cohort

Our study investigated the association between specific *BRCA1* PVs (c.4035del and c.5266dup) and the development of BC or OC. The distribution of these PVs among study cohort (see Table 2.1) is consistent with previous research, confirming their relevance within the Latvian population (Gardovskis et al., 2005; Tikhomirova et al., 2005). We explored the penetrance and impact of these *BRCA1* PVs on the age of onset for BC or OC. Additionally, this dataset served as the basis for GWAS and PRS analyses to identify genetic factors affecting the penetrance of region-specific *BRCA1* PVs. This detailed analysis may improve our understanding of the relationship between these specific *BRCA1* PVs and BC and OC risk, guiding further research, personalised risk assessment, and preventative strategies.

The objective of this study was to investigate the penetrance of specific *BRCA1* PVs for BC and OC within the cohort. While the younger age of unaffected individuals might shew the penetrance estimates, this study offers valuable insights into the penetrance of region-specific *BRCA1* PVs.

Our findings support the concept that penetrance can vary depending on the localisation of specific *BRCA1* PV. For instance, *BRCA1*:c.5266dup PV exhibits higher penetrance in the BC group compared to the OC group, while *BRCA1*:c.4035del PV demonstrates similar penetrance in both cancer types (see Table 2.2). These PVs exhibit a genotype–phenotype correlation and differing clinical presentations, potentially due to their position and effects on the 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 of developing BC, while PVs in exon 10 (e.g. c.4035del) present nearly 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 risk assessment (Kuchenbaecker, Hopper, et al., 2017).

Next, we performed a Cox proportional hazards regression analysis of the study cohort (452 individuals), which revealed a statistically significant influence of the *BRCA1*:c.5266dup PV on earlier cancer onset (combining BC and OC groups) compared to the *BRCA1*:c.4035del PV. *BRCA1*:c.5266dup PV carriers had a median age of cancer onset at 46.52 years, while *BRCA1*:c.4035del PV carriers presented at 50.62 years (see Table 2.1). The hazard ratio of 1.437 indicated a 43.70 % increased risk of earlier cancer onset among *BRCA1*:c.5266dup PV carriers.

Within the BC group, the *BRCA1*:c.5266dup PV demonstrated a 56.40 % higher hazard for BC development, consistent with previous findings in Latvian BC patients (Plakhins et al., 2011). These results highlight the necessity of a personalised approach in genetic counselling, incorporating two region-specific *BRCA1* PVs into risk management strategies such as intensified surveillance or risk-reducing bilateral mastectomy.

The *BRCA1*:c.5266dup PV significantly influenced earlier cancer onset in BC but not in OC, underscoring the variant-specific effects of *BRCA1* PVs on genotype-phenotype correlation (Milne & Antoniou, 2016; Plakhins et al., 2011). Interactions with other genetic or environmental factors could potentially modify the impact of the *BRCA1*:c.5266dup PV on OC development risk. The advanced stage at which OC is typically diagnosed might also influence the observed age of onset, possibly decreasing the effect of *BRCA1*:c.5266dup PV compared to BC, where early detection is more common (Thulesius et al., 2004).

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

Next, our study examined the impact of *CHEK2* gene variants on *BRCA1* PV penetrance, as *CHEK2* is involved in the same DNA repair pathway as *BRCA1* gene. *CHEK2* variants are frequently observed in BC and OC patients and have been extensively studied in several European countries (Myszka et al., 2011; Narod & Lynch, 2007), but research in the Latvian population is limited. Only two previous studies by Irmejs et al. and Plonis et al. have investigated the association of specific *CHEK2* variants with BC, OC, and colorectal cancer development risk (Irmejs et al., 2006; Plonis et al., 2015).

While many studies show that SNVs in modifier genes impact PV penetrance, research on double heterozygous *BRCA1* and *CHEK2* PV carriers is rare. Previous studies have identified only 1 to 15 cases of such double heterozygotes among thousands of patients per study, mostly focusing on BC (Cybulski et al., 2009; Meijers-Heijboer et al., 2002; Sokolenko et al., 2014; Turnbull et al., 2012).

In contrast to previous studies predominantly comparing the frequency of *BRCA1* and *CHEK2* double heterozygotes among BC patients to 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 *BRCA1* and *CHEK2* 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.

We identified 13 cases of *BRCA1* and *CHEK2* double heterozygotes (see Table 2.3), which is consistent with previous research. Our findings suggest a tendency towards earlier OC onset in double heterozygotes compared to *BRCA1* PV carriers alone, but this was not statistically significant. The results

align with previous studies showing no significant impact of *CHEK2* variants on cancer onset in *BRCA1* PV carriers (Cybulski et al., 2009; Sokolenko et al., 2014, Sukumar et al., 2021).

Previous research in other populations, as well as in Latvia, has demonstrated lower frequencies of *CHEK2* variants in *BRCA1* PV carriers compared to BC patients without *BRCA1* PVs, suggesting a negative interaction between these variants (Cybulski et al., 2009; Irmejs et al., 2006; Meijers-Heijboer et al., 2002; Plonis et al., 2015; Sokolenko et al., 2014; Turnbull et al., 2012). This could be due to reduced viability of cells with both variants, as both gene products are involved in the same DNA repair pathway (Bartek & Lukas, 2003; Collins & Garrett, 2005; Lee et al., 2000).

In conclusion, no statistically significant evidence has emerged regarding the impact of pathogenic/likely pathogenic *CHEK2* variants on the risk of BC or OC development in *BRCA1* PV carriers. The modest sample size may limit the statistical power, and larger studies are needed to enhance the credibility of these findings.

3.3 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.

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. 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 paragraphs, 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. Table 2.4 presents three of the most significant genetic variants associated with BC or OC development risk. Interestingly, all three variants exhibited a negative beta, suggesting a potential protective effect on cancer development.

The most significantly associated lead variant with BC development risk was intronic variant rs2609813 of the *FAM107B* gene. 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 across 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 (Nakajima & Koizumi, 2014).

Previous studies have suggested the *FAM107* gene family as candidate tumour suppressors, with *FAM107A*, previously known as *DRR1*, showing a correlation between downregulation and increased tumourigenesis across 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). Although *FAM107B* gene shares structural similarities, its specific role as tumour suppressor remains unclear. Studies have observed reduced expression of *FAM107B* in different tumour tissues, including breast, thyroid, gastric, and colon cancer cells, suggesting its potential involvement in tumour development. Experimental evidence also indicates that increased *FAM107B* expression inhibits cancer cell proliferation (Nakajima et al., 2010; Nakajima et al., 2012), and the 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, decreased expression of *FAM107B* has been observed in BC tissues, particularly in aggressive phenotypes associated with increased risk of disease recurrence and shortened survival (Nakajima et al., 2012).

The identified negative effect size of the variant rs2609813, predicted to be a regulatory region variant, suggests its potential protective effect by affecting other gene expression in *BRCA1* PV carriers. Overall, *FAM107B* emerges as a promising candidate tumour suppressor gene in BC, and further research should focus on studying the mechanisms of this regulatory region variant.

The second most significant SNV suggestively associated with BC development risk was rs4688094 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, influencing various biological processes such as proliferation, apoptosis, metastasis, and drug resistance (Arun et al., 2018; Liu et al., 2021). 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). They can regulate other gene expression at various levels, including chromatin modification, transcription, and post-transcriptional processing of RNA, indicating their potential as therapeutic targets (Gutschner & Diederichs, 2012). Moreover, dysregulation of specific lncRNAs has been associated with different BC subtypes and clinical outcomes, supporting their potential as diagnostic and prognostic biomarkers (Su et al., 2014; Zhao et al., 2021).

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

The only lead variant surpassing the genome-wide suggestive significance threshold within the OC group was rs79732499*,* located in an intergenic regulatory region. This variant demonstrates high LD with several variants within the *C20orf194* gene, also known as *DNAAF9*, suggesting a potential impact on *DNAAF9*.

The information about *DNAAF9* functions is limited, but current knowledge suggests it is involved in microtubule interaction and tubulin assembly (Casalou et al., 2020). According to UniProt database, it 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).

Despite the negative effect size of the identified rs79732499 variant, suggesting a potential protective effect in *BRCA1* PV carriers, the precise function of *DNAAF9* in cancer remains unknown. Further investigation is necessary to understand its function and potential implications in OC.

Next, we investigated the effect of SNVs on gene expression in breast and ovary tissue using eQTL mapping with the GTEx dataset. While no significant SNV-gene associations were observed in ovarian tissue, two SNVs, rs10178186 and rs434451, were identified to affect gene expression in breast tissue (see Table 2.5).

The first identified eQTL variant rs10178186 is associated with reduced *ZNF514* expression. Zinc finger proteins (*ZNF*s), including *ZNF514*, are transcription factors suggested to play a role in carcinogenesis, cancer progression, and metastasis across various cancers. Despite their large number, most of the *ZNF*s are not well studied (Luo et al., 2018; Ye et al., 2021).

While previous studies have implicated the role of certain *ZNF*s in carcinogenesis, further functional studies are needed to fully understand their potential role and impact on BC development. For instance, *ZNF165* has been associated with promoting triple-negative BC (TNBC) development, potentially promoting more aggressive carcinogenesis (Gibbs et al., 2020). Conversely, the hypermethylated *ZNF154* promoter has been associated with increased survival rates in resectable pancreatic cancer (Wiesmueller et al., 2019).

The second eQTL variant, rs434451, is associated to reduced expression of *SLC1A5*, a gene encoding a cell surface solute-carrying transporter important for maintaining the uptake of neutral amino acids, particularly glutamine, crucial for cancer cell metabolism (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. Pharmacological studies have demonstrated that inhibiting *SLC1A5*-mediated transport reduces glutamine uptake, leading to decreased cancer cell proliferation and increased cell death, especially in TNBC cells (van Geldermalsen et al., 2016). Additionally, *SLC1A5* expression is associated with sensitivity to endocrine therapy in luminal BC, suggesting its potential utility as a predictive biomarker of treatment response (Alfarsi et al., 2021).

Previous research highlights the significance of *SLC1A5* and its associated transporters in cancer cell metabolism, growth, and proliferation. While preclinical studies have shown promising results with SLC1A5 inhibitors, there are currently no clinical trials testing them (Nachef et al., 2021). Further research is required to fully understand the role of *SLC1A5* and its potential as a therapeutic target in BC.

In conclusion, additional studies are necessary to understand the potential importance of *ZNF514* and *SLC1A5* expression in BC development risk for *BRCA1* PV carriers. More in-depth investigations into the precise mechanisms underlying their role in tumour growth, progression, and response to therapy are needed.

3.4 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 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.

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. While the BayesRR-RC PRS model performed well in predicting the risk of developing BC, the BayesW PRS model remained superior (see Table 2.7).

In previous study Kuchenbaecker et al. 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, 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 of 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 (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.

3.5 Strengths and weaknesses 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). Most existing penetrance estimates are derived from large-scale studies combining data from multiple populations, often overlooking the penetrance specific to founder PVs within the distinct populations. Population-specific genetic structures can influence study outcomes, as certain SNVs of modifier genes may be more prevalent in one population while rare in another (Narod, 2002). This is supported by Pankratov et al., who suggested that local population history and genetic structure can significantly influence 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.

However, we encountered several limitations, primarily due to the small number of patients with the double heterozygous *BRCA1* and *CHEK2* genotype. This was mainly due to the limited cohort size and the rarity of this genotype (Cybulski et al., 2009). A larger cohort is necessary to verify the hypothesis that double heterozygotes may have a higher risk of BC or OC, potentially achievable through consortiums or larger collaborative studies. Additional double heterozygotes may be identified with the increasing use of whole exome sequencing (WES) or WGS.

Other limitation includes a small number of BC and OC cases with germline *BRCA1* PVs and potential selection biases, since the samples were obtained during diagnostic germline variant testing in a clinical setting. Our focus on two specific *BRCA1* PVs might not fully represent the entire population of *BRCA1* PV carriers. Additionally, our PRS analysis was limited by missing SNVs, which could be improved by incorporating a more population-specific reference panel for the imputation step from the Genome Database of Latvian Population (LGDB) once the relevant WGS data becomes available (Rovite et al., 2018).

The significant age difference among the groups might have influenced the penetrance estimates, which could be addressed in future studies by using an age-matched study cohort. Moreover, the absence of detailed clinical information on tumour phenotypes made our results an average estimation across all BC and OC phenotypes.

Finally, the complex interplay of SNVs and the predominance of identified non-coding variants highlight the critical need for extensive post-GWAS analysis and in-depth functional studies to interpret the specific impact of identified SNVs on a particular phenotypes (Milne & Antoniou, 2016; Yang et al., 2022).

3.6 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 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 enable individuals and healthcare providers to take 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 prioritising the genetic screening and recontacting individuals carrying clinically significant PVs, thereby contributing to the overall health of the Latvian population.

List of publications, reports and patents on the topic of the 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. 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

2. 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

3. 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

4. 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

5. 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

6. 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

7. 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

8. 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

9. 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

10. Carstensen. (2022). *Epi: A Package for Statistical Analysis in Epidemiology*. Oxford University Press: Oxford, UK. Available online: https://CRAN.R-project.org/package=Epi (accessed on 22 May 2023).

11. 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

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

13. 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

14. 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

15. 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

16. 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

17. 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

18. 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

19. 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

20. 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

21. 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

22. 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

23. 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

24. 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

25. 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

26. 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

27. 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

28. 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

29. 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

30. 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

31. 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

32. 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

33. 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

34. 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

35. 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

36. 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

37. 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

38. 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

39. 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

40. 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

41. Kuchenbaecker, K. B., Ramus, S. J., Tyrer, J., Lee, A., Shen, H. C., Beesley, J., . . . BRCA2, 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

42. 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

43. 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

44. 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.

45. 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

46. 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

47. 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

48. 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

49. 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

50. 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

51. 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

52. 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

53. 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

54. 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

55. 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

56. 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

57. 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

58. 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

59. 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

60. 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

61. 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

62. 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

63. 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

64. 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

65. 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

66. 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

67. 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

68. 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

69. 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

70. 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

71. 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

72. 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

73. 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

74. 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

75. 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

76. 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

77. 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

78. 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

79. 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

80. 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

81. 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

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

83. Team, R. C. (2020). 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).

84. 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).

85. 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

86. 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

87. 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

88. 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

89. 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

90. 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

91. 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

92. 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

93. 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

94. 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

95. 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

96. 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

Acknowledgements

First of all, I would like to express my sincere gratitude to my scientific supervisors, **Edvīns Miklaševičs** and **Baiba Vilne**, for their support and guidance throughout my PhD journey. I extend my thanks to **E. Miklaševičs** for giving me the opportunity to pursue a PhD, for providing the initial research idea, and for teaching me the independence in research. I am deeply grateful to **B. Vilne** for coming to RSU and taking me, a desperate PhD student, under her "wing" and patiently guiding me through the overwhelming world of bioinformatics. I can confidently say, that without your mentorship, the completion of the analysis and this Thesis would not have been possible.

I also extend my appreciation to my scientific supervisor, **Jeļena Maksimenko**, and my colleagues at the Breast Surgery Unit of Pauls Stradiņš Clinical University Hospital for their invaluable help with sample collection and selection.

My sincere gratitude goes to my reviewer, Assoc. Prof. **Madara Auzenbaha**, for dedicating time to review my Thesis and providing valuable comments and advice. Her insights have undoubtedly improved the quality of this work. I also express gratitude to all members of the Doctoral Thesis Discussion Committee for their assessment and valuable recommendations.

I would like to thank my colleagues from Rīga Stradiņš University, Institute of Oncology and Molecular Genetics for their continuous support and guidance.

A special thank you goes to my colleagues at the Children's University Hospital for the enthusiasm we share for genetics, the endless support, and the motivation that has been a source of strength during challenging times.

I am also deeply grateful to our colleagues at the University of Tartu, Institute of Genomics, for their collaborative spirit, invaluable guidance, training,

and expertise in GWAS analysis. Without their support, I would not be able to complete this work. In particular, I would like to thank **Tiit Nikopensius**, **Reedik Mägi**, **Krista Fischer**, **Triin Laisk**, and **Urmo Võsa**.

I am deeply grateful for the unending support and belief of my friends and family. I owe a special thank you to my husband for his remarkable patience, tireless understanding, and continuous encouragement throughout this journey.

Finally, I would like to thank everyone who has supported and cheered me as I have progressed towards my degree. Each of you has played a significant role in shaping my academic and personal growth, and for that, I am truly grateful.