Exploring the Role of Novel Candidate Genes in Susceptibility to Breast Cancer

by

Neda Zamani

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Institute of Medical Science
University of Toronto

Exploring the Role of Novel Candidate Genes in Susceptibility to Breast Cancer

Neda Zamani

Doctor of Philosophy

Institute of Medical Science
University of Toronto

2025

Abstract

While several breast cancer susceptibility genes have been identified, more are likely to exist. Our research team recently identified ATRIP as a novel breast cancer susceptibility gene candidate through whole-exome sequencing (WES) of familial breast cancer patients and healthy controls from the Polish founder population. A recurrent rare mutation, ATRIP c.1152_1155del, was observed in 42 of 16,085 Polish breast cancer patients and in 11 of 9,285 controls (OR = 2.14, 95% CI = 1.13-4.28, p = 0.02). In another study among the Bahamian population, our research team identified three novel candidate genes, DCLRE1C, DDX19B, and XRCC3, with potential roles in susceptibility to breast cancer.

To follow up with these findings, we conducted targeted full-gene sequencing on germline DNA from 1,021 familial breast cancer patients in Ontario, 1,390 French-Canadian patients in Quebec, and 421 Iranian patients. Controls included 930 healthy women from Ontario, 800 individuals from the Iranome database, and 450 from the

Gen3G study. Additionally, we analyzed UK Biobank WES data from 15,643 Caucasian breast cancer cases and 157,943 matched controls.

Gene-based association analysis revealed that of the four candidate genes under study, only ATRIP loss of function (LoF) variants were associated with a significantly increased risk of breast cancer among the British population. We identified ATRIP LoF carriers among 13 of 15,643 Caucasian breast cancer patients and 40 of 157,943 ethnicity-matched healthy individuals within the UK Biobank (UKB) cohort (age-adjusted OR = 3.28, 95% CI = 1.76-6.14, $p < 0.001$). As an exploratory aim, we also examined ATRIP in 1,488 unselected ovarian cancer patients from Ontario. We identified four LoF carriers, with none among 930 local controls (OR = 5.6, $p = 0.14$).

Tumor DNA analysis of ten ATRIP-mutated breast tumors showed loss of heterozygosity (LOH) in at least four and homologous recombination deficiency (HRD) in seven tumors. Mutational signature analysis indicated defects in both homologous recombination repair (HRR) and mismatch repair pathways.

ATRIP is essential for activating the ATR-mediated DNA damage checkpoint at stalled replication forks. Altogether, our findings support ATRIP as a breast cancer susceptibility gene linking DNA replication stress to hereditary breast cancer risk.

Acknowledgments

First and foremost, I am profoundly appreciative of my supervisor, Dr. Mohammad R. Akbari, for his unwavering support, expertise, and guidance throughout this study. I wish to express my heartfelt gratitude for your warm reception at the University of Toronto, the thrilling opportunity to engage in this study, and the privilege of conducting research in your laboratory. Your extensive knowledge and considerable experience have inspired me in my academic pursuits and personal endeavors.

I would like to extend my heartfelt gratitude to my program advisory committee members, Dr. Jennifer Brooks and Dr. Lisa Strug, for their invaluable feedback and counsel, which have been instrumental in refining my project and ensuring its success. I appreciate your support and encouragement, and I am thankful for the knowledge and thoughtful advice you have imparted to me during my study.

My sincere appreciation goes to Agata Szymiczek and Dr. Shiyu Zhang for their support and guidance throughout my wet lab training and for their assistance with the challenges I encountered during my laboratory work. Your guidance and expertise have been invaluable, and I am grateful for all that I have learned from you.

Finally, I wish to express my profound gratitude to my family. I consider myself fortunate to have been born to my parents, Ahmad and Farzaneh, who gave me life and cultivated my intellect throughout my journey. My deepest gratitude goes to my husband and closest companion, Saeid, for his unwavering support during the challenging phases of my studies. My accomplishments would not have been achievable without your affection, motivation, and steadfast support. I am also immensely grateful to my sister, Hoda, for her consistent spiritual support not only throughout my studies but also in every aspect of my life. Lastly, I express profound gratitude to my daughter, Rosha, whose smiles perpetually brighten my days and provide an eternal source of motivation.

I extend my profound gratitude to all of you for your consistent guidance, support, and encouragement throughout my PhD journey. Your invaluable contributions have been pivotal to the completion of this work. Thank you all once again for everything.

Contributions

I, Neda Zamani, am the author of this thesis. I prepared this thesis in all aspects and contributed either wholly or in part to the conduction of laboratory experiments, data analysis, results, and write-up.

Dr. Mohammad R. Akbari (Supervisor) – mentorship, laboratory resources, data analysis, guidance in planning and execution of the study, assistance in thesis preparation

Dr. Jennifer Brooks (Program Advisory Committee member) – assistance in thesis planning, preparation and execution

Dr. Lisa Strug (Program Advisory Committee member) – assistance in thesis planning, preparation and execution

Declaration of Publications

The discovery of ATRIP as a novel breast cancer gene candidate among the Polish population (i.e., the preliminary work by our research team discussed in Chapter 2), validation of this finding among the UK population (Chapter 3), and the subsequent analysis of tumor DNA, including loss of heterozygosity analysis and homologous recombination deficiency status of ATRIP-mutated breast tumors (Chapter 6) have been published as an original article in the American Journal of Human Genetics. The author of this thesis is the second author of this multi-author publication.

Additionally, two manuscripts have been prepared: one addressing ATRIP's role in susceptibility to ovarian cancer (Chapter 3), and another presenting a case report of an Iranian family with clustered breast cancer and carriers of a truncating ATRIP variant (Chapter 4). Both manuscripts are currently undergoing review for submission, with the author of this thesis as the primary author. A third manuscript on tumor mutational signatures observed in ATRIP-mutated breast tumors (Chapter 6) is under preparation. The author of this thesis is the primary author of this manuscript.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACMG	American College of Medical Genetics and Genomics
ASCO	American Society of Clinical Oncology
BCAC	Breast Cancer Association Consortium
CGC	Cancer Gene Census
CI	Confidence Interval
CRG	Cancer Relevant Genes
DSBs	Double-Strand Breaks
EGFR	Epidermal Growth Factor Receptor
ER	Estrogen Receptor
FFPE	Formalin-Fixed, Paraffin-Embedded
GWAS	Genome Wide Association Study
HBC	Hereditary Breast Cancer
HBOC	Hereditary Breast and Ovarian Cancer
HER2	Human Epidermal Growth Factor Receptor 2
HR	Hormone Receptor
HR	Hazard Ratio
HRD	Homologous Recombination Deficiency
HRR	Homologous Recombination Repair
HRT	Hormone Replacement Therapy
IHC	Immunohistochemistry
ISH	In Situ Hybridization
LFS	Li-Fraumeni Syndrome
LoF	Loss of Function
LOH	Loss of Heterozygosity
MAF	Minor Allele Frequency
MMR	Mismatch Repair
MRI	Magnetic Resonance Imaging
MSI	Microsatellite Instability
NGS	Next Generation Sequencing

NHEJ	Non-Homologous End-Joining
NMF	Non-negative Matrix Factorization
OC	Oral Contraceptive
OR	Odds Ratio
PARP	Poly ADP Ribose Polymerase
PR	Progesterone Receptor
RR	Relative Risk
SBS	Single-Base Substitution
SNP	Single Nucleotide Polymorphism
SNVs	Single Nucleotide Variants
TCGA	The Cancer Genome Atlas
TNBC	Triple-Negative Breast Cancer
UKB	UK Biobank
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
WHI	Women's Health Initiative

Chapter 1 – Literature Review

1.1 Breast Cancer

1.1.1 Breast Cancer Statistics

Breast cancer is the most frequently diagnosed malignancy among women in both developed and developing countries (Bray et al., 2024). According to the most recent GLOBOCAN report, there were approximately 2.3 million new cases of breast cancer in 2022 worldwide, which accounted for about 1 in 4 cancer cases among women. Breast cancer is also ranked as the leading cause of cancer-related mortalities among women, contributing to roughly 1 out of 6 cancer deaths in women globally (Bray et al., 2024).

Overall, breast cancer incidence rates are higher in developed countries than in developing countries (55.9 and 29.7 per 100,000 females, respectively), whereas mortality rates are reported to be higher in developing countries compared to the developed ones (15.0 and 12.8 per 100,000 females, respectively) (Sung et al., 2021). Elevated rates of breast cancer incidence in developed countries can be due to a higher frequency of lifestyle, hormonal, and reproductive risk factors, as well as increased detection through organized or opportunistic mammographic screening. Moreover, the availability of screening programs for early detection and subsequent access to personalized therapy upon diagnosis may contribute to the lower mortality rates of the disease in such countries (Bray et al., 2024; Sung et al., 2021; Wilkinson & Gathani, 2022).

In Canada, breast cancer is the most frequent cancer and the second major cause of cancer-related mortalities among women. Based on the most recent report of the Canadian Cancer Society, in 2024, approximately 30,800 Canadian women are expected to be diagnosed with breast cancer, which accounts for ~25% of all new cancer cases among women. Also, about 5,500 women are expected to die from the disease, which represents ~13% of all cancer-related deaths among Canadian women in 2024. It is

expected that 1 out of 8 Canadian women will probably develop breast cancer during their life, and 1 out of 34 will die from the disease (Brenner et al., 2024).

Breast cancer incidence rates in Canada rose in the mid-1980s, and since then, there has been a fluctuation in the incidence rates with an overall slight reduction. Increased incidence rates in the mid-1980s were mostly due to more frequent Mammography, which led to the detection of more breast cancer cases. The slight decrease in the incidence rate in 2002 coincided with the reduction in using hormone replacement therapies among post-menopausal women as their role in susceptibility to breast cancer became publicized (Canadian Cancer Statistics Advisory Committee, 2018).

Breast cancer death rates in Canada have declined by approximately 45% since their peak in the mid-1980s. Based on the latest report, the 5-year survival rate of breast cancer was estimated to be 89% in 2022. Moreover, more than 80% of affected women were diagnosed at earlier stages of breast cancer (Stages I and II), and less than 5% presented with the late stage of the disease (stage IV) (Canadian Cancer Statistics Advisory Committee, 2018). These trends reflect enhancements in screening technology and continuous research in the field, leading to the improvement of breast cancer prevention, early diagnosis, and personalized treatment.

1.1.2 Breast Cancer Conventional (Clinical) Subtypes

In the late 19th century, a British surgeon proposed that female gonadal hormones may have a significant impact on breast cancer and hypothesized that the ovaries' secretions can influence other cells, leading to breast tumorigenesis (Beatson, 1896). The discovery of hormone receptor status as a predictive factor for the risk of recurrence and response to treatment in breast cancer patients did not occur until the late 1960s (Osborne et al., 1980). The groundbreaking research by Elwood Jensen, which involved estrogen receptor (ER) isolation and investigating its regulatory function in gene expression, laid the foundation for classifying breast cancer according to hormone receptors (Greene, 2013). The identification of antibodies against ER facilitated the

advancement of quantitative and immunohistochemistry (IHC)-based assays for assessing ER expression (Greene, 2013). ER is a nuclear hormone receptor that functions as a transcriptional regulator of numerous genes, including the gene that encodes another nuclear hormone receptor known as the progesterone receptor (PR) (Mueller et al., 2018). It has been observed that breast tumors classified as ER-negative are also likely to be PR-negative. Subsequently, the correlation between ER and PR expression was elucidated, as the induction of the PR gene relies on the estrogen pathway (Hammond et al., 2010).

The identification of epidermal growth factor receptors unveiled the correlation between human epidermal growth factor receptor 2 (HER2)/Neu (also referred to as ERBB2) and a decreased survival rate in breast cancer patients (Slamon et al., 1987). ERBB2 is a proto-oncogene located on chromosome 17 and encodes the tyrosine kinase receptor HER2, which plays a role in activating pathways that promote cell proliferation and influences cell differentiation, invasiveness, and survival (Fragomeni et al., 2018).

These discoveries led to the advancement of targeted therapies against hormone receptors and HER2 in breast cancer patients. An expert panel from the American Society of Clinical Oncology (ASCO) and the College of American Pathologists (CAP) released the latest clinical practice guidelines defining the application of endocrine therapy in patients with ER/PR-positive tumors and HER2-targeted therapy in patients with HER2-positive tumors (Allison et al., 2020; Wolff et al., 2018). ER, PR, and HER2 status of the breast tumor cells are assessed by established techniques like IHC and in situ hybridization (ISH) (Mueller et al., 2018). The clinical categorization of breast cancer was established based on the expression levels of ER, PR, and HER2, which are the only biomarkers known to have a clinically validated connection with prognosis and treatment choices. This conventional classification differentiates the following subtypes of breast cancer:

Hormone receptor-positive (HR-positive) breast cancer is characterized by being ER-positive, PR-positive, and HER2-negative (HR+, HER2-). Approximately 70% of all breast cancer cases are of this clinical subtype (Cho, 2016). These tumors have a

favorable prognosis and generally show a positive response to hormone therapy. However, the effectiveness of the therapy may vary depending on the level of ER expression (Mueller et al., 2018).

HER2-positive breast cancer is characterized by being ER-negative, PR-negative, and HER2-positive (HR-, HER2+). About 20%–25% of breast tumors manifest HER2 overexpression (Garrett & Arteaga, 2011). The abundance of HER2 in the cell membrane is associated with a higher tumor grade, an increased number of cell divisions, and the presence of cancerous cells in the lymph nodes, all of which are indicators of a negative prognosis (Nassar et al., 2014). HER2-positive tumors exhibit limited responsiveness to chemotherapy and endocrine therapy; however, they are suitable candidates for anti-HER2 treatment (Mueller et al., 2018).

Triple-Negative Breast Cancer (TNBC) is characterized by the lack of ER, PR, and HER2 expression. This subtype has the most unfavorable outlook and is treated with systemic chemotherapy. Because there are no well-established molecular targets for therapy, TNBC is a subtype of particular interest for clinical trials looking for innovative treatment methods (Yin et al., 2020).

Lastly, there are ER-positive, PR-positive, and HER2-positive (HR+, HER2+) breast cancers, which are commonly referred to as Triple-Positive tumors. These tumors generally have a moderate prognosis and can be managed using a combination of hormone therapy, chemotherapy, and/or anti-HER2 treatment (Dieci & Guarneri, 2020).

1.1.3 Breast Cancer Molecular (Intrinsic) Subtypes

There are notable differences in the rates of treatment response among the subgroups categorized based on the conventional classification, which necessitated further research on breast cancer subtyping. It was hypothesized that monitoring relevant signaling pathway activities through gene expression profiling could provide accurate distinctions between breast cancer subtypes (Nicolini et al., 2017). In a groundbreaking research, Perou et al. performed a gene expression profiling study where they utilized hierarchical clustering analysis of microarray gene expression data to identify four distinct

subtypes of breast cancer: ER+/Luminal-Like, HER2-Enriched, Basal-Like, and Normal Breast-Like (Perou et al., 2000). Sorlie et al. expanded upon this research by utilizing cDNA microarrays of gene expression patterns. Their findings revealed that the ER+/Luminal-Like subtype could be subdivided into two distinct luminal subtypes, namely Luminal A and Luminal B (Sørlie et al., 2001). The Cancer Genome Atlas (TCGA) team conducted a study where they analyzed genomic DNA, mRNA, and protein arrays. This analysis revealed four intrinsic subtypes, including Luminal A, Luminal B, HER2-Enriched, and Basal-Like, that explain the phenotypic diversity observed in breast tumors. Because the Normal-Like subgroup had a low level of somatic mutations, it was considered enriched for benign breast tissue and therefore excluded (The Cancer Genome Atlas Network, 2012).

Luminal A tumors are the most common molecular subtype, representing 40%-50% of all invasive breast cancers (Voduc et al., 2010). These tumors are characterized by being low-grade and having high expression of luminal epithelial genes, as well as ER and PR, while showing low expression of HER2. Somatic mutations in the PIK3CA, GATA3, and MAP3K1 genes are commonly observed in these tumors (Norum et al., 2014). Another feature of Luminal A tumor cells is that they have low levels of Ki-67, a protein that is involved in cell proliferation (Inic et al., 2014). Having these features, Luminal A tumors typically exhibit a favorable prognosis and high responsiveness to hormone and endocrine therapy (Fragomeni et al., 2018).

Luminal B tumors are characterized by their higher grade and lower expression of ER and PR compared to Luminal A tumors. These tumors show frequent mutations in the PIK3CA and TP53 genes (Norum et al., 2014). A subset of Luminal B tumors exhibits HER2 overexpression. These HER2-positive Luminal B tumors have a less favorable prognosis compared to HER2-negative Luminal B cancers (Guiu et al., 2012). In general, Luminal B tumors tend to be more aggressive and have a worse prognosis than Luminal A tumors, but compared to all other subtypes, their prognosis is considered to be intermediate. As this subtype is more proliferative, Luminal B breast cancer patients are likely to benefit from a combination of chemotherapy and hormone treatment. If the tumor exhibits HER2 expression, administering an anti-HER2 treatment will also serve as an effective intervention (Norum et al., 2014).

HER2-Enriched tumors are not solely determined by the status of hormone receptors or the expression of the HER2 gene. These tumors are driven by the epidermal growth factor receptor (EGFR)/HER2 signaling pathway, which consists of four primary proteins: EGFR, HER2, HER3, and HER4 (Godoy-Ortiz et al., 2019; Moasser, 2007). Enhanced dimerization of HER2 and EGFR leads to increased proliferative activity. The expression of HER2-regulated genes can be used as a reliable predictive marker to assess the potential effectiveness of treatments that target the HER2 pathway (Llombart-Cussac et al., 2017). HER2-Enriched tumors typically display elevated levels of HER2 expression and do not show expression of luminal epithelial genes. These tumors exhibit rapid cell division and are likely to be high-grade, more aggressive, and have a poorer prognosis compared to luminal tumors. Treatment options for HER2-Enriched patients may include HER2 blockade therapies, such as the administration of anti-HER2 antibodies or small molecule inhibitors (Llombart-Cussac et al., 2017). Notably, HER2-Enriched cancers frequently harbor mutations in TP53, which exacerbates the unfavorable prognosis of the disease (Nik-Zainal et al., 2016; Silwal-Pandit et al., 2014).

Basal-Like tumors are characterized by an increased expression of genes commonly associated with basal cells (i.e., KRT5, KRT17, and LAMC1). These tumors exhibit distinct features such as high grade, high proliferation, the presence of necrotic regions, and complex genomic rearrangements. There is also increased activation of the WNT pathway and high expression of EGFR (Norum et al., 2014). Basal-Like tumors lack the expression of ER, PR, and HER2 and often exhibit mutations in the TP53 and BRCA1 genes (Dai et al., 2015). These tumors exhibit a high level of aggressiveness and are strongly correlated with an unfavorable prognosis and a high chance of relapse. Chemotherapy is the recommended therapeutic approach for tumors categorized as Basal-Like (Norum et al., 2014).

1.1.4 Risk Factors for Breast Cancer

Epidemiological studies have identified various modifiable and non-modifiable factors that may increase a woman’s chance of developing breast cancer. Some modifiable risk factors include alcohol consumption, smoking, excess body fat, low

physical activity, use of oral contraceptives (OCs), and postmenopausal hormone replacement therapy (HRT) (Rojas & Stuckey, 2016; Sun et al., 2017).

A pooled analysis of 53 worldwide epidemiological studies demonstrated that compared with women who reported no alcohol drinking, the relative risk (RR) of breast cancer was 1.32 (95% confidence interval (CI) = 1.19-1.45) in women with alcohol intake of 35-44 grams per day, and 1.46 (95% CI = 1.33-1.61) in women who drank at least 45 grams of alcohol per day. It was estimated that the relative risk of breast cancer elevated by 7.1% (95% CI = 5.5-8.7) with each additional 10 grams of alcohol intake per day (Hamajima et al., 2002).

Based on 22 cohort and 27 case-control studies, the 2014 Surgeon General's report concluded that ever smoking was associated with a significant increase of ~10% in breast cancer relative risk. This effect seemed to be stronger for current smokers (12%) than for former smokers (9%). The risk also elevated by 16% for smoking duration of 20 or more years and by 13% for smoking 20 or more cigarettes per day (National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health, 2014). The WECARE Study, a large population-based case-control study with 51,521 female breast cancer cases and 52,212 controls, also showed an elevated risk of breast cancer in women who both smoked and drank alcohol (RR = 1.54; 95% CI = 1.18-1.99) (Knight et al., 2017).

Obesity is another strong risk factor for breast cancer, especially in postmenopausal women. The analysis of data from 162,691 postmenopausal women within the UK Biobank cohort study demonstrated a continuous increase in the risk of breast cancer with increasing adiposity. In this study, it was observed that the relative risk of developing breast cancer was 1.70 (95% CI = 1.52-1.90) among women in the top quartile of body fat mass (median 38.5 kg) compared to those in the lowest quartile of body fat mass (median 17.3 kg) (W. Guo et al., 2018). On the other hand, it has been shown that regular engagement in strenuous physical activity decreased the risk of breast cancer by 14% (RR = 0.86; 95% CI = 0.78-0.95). This preventive effect seemed to be greater with more hours of exercise (McTiernan et al., 2003).

The findings regarding the correlation between various types of menopausal HRT and breast cancer risk are inconsistent. The association between the use of conjugated estrogens and an elevated risk of breast cancer was first reported in 1976 (Hoover et al., 1976), with subsequent studies replicating this finding (Bergkvist et al., 1989; Colditz et al., 1995). The Women's Health Initiative (WHI) study, launched between 1993 and 1998, sought to evaluate the health risks associated with HRT among more than 16,000 postmenopausal women aged 50 to 79 years old (Rossouw et al., 2002). Following a 5.2-year follow-up, the WHI report in 2002 showed that women who underwent combined HRT exhibited a 26% elevated risk of developing breast cancer compared to the placebo group. Estrogen-alone therapy elevated the risk of breast cancer but required a longer duration of exposure compared with combined HRT for detrimental effects. Moreover, HRT-associated breast cancers were diagnosed at an earlier stage, were mainly HR-positive, and exhibited a favorable prognosis (Rossouw et al., 2002). Subsequently, in the Million Women Study, a cohort of about one million women aged 50-64 years and recruited from 66 breast cancer screening centers across the UK, Beral et al. demonstrated that both estrogen-alone and combined HRT were associated with significantly elevated breast cancer risks, with the associated risk significantly higher for combined HRT compared to the estrogen-alone therapy (Beral & Million Women Study Collaborators, 2003). The Million Women Study report was updated in 2019 after a roughly 20-year follow-up on 907,167 postmenopausal women with no breast cancer at the time of recruitment. It was demonstrated that both estrogen-alone and combined HRT were associated with excess breast cancer mortality rates (Beral et al., 2019). In 2020, the WHI study released the findings of two randomized clinical trials involving 27,347 postmenopausal cancer-free women with a cumulative 20-year follow-up period (Chlebowski et al., 2020). The 2020 WHI report showed that the use of combined HRT was associated with an elevated risk of developing breast cancer, and the risk was not restricted to HR-positive cancers. Combined HRT interfered with mammographic detection of breast tumors, leading to diagnoses at more advanced stages of the disease. Moreover, combined HRT was associated with an increased breast cancer mortality rate during a 7-year follow-up. However, its effect on mortality rate subsequently attenuated and ceased to be statistically significant. Surprisingly, the findings regarding the use of

estrogen-alone therapy were inconsistent with previous reports. During a 7.2-year follow-up, it was shown that the use of estrogen-alone therapy reduced the risk of breast cancer, with the greatest effect on ER-positive and PR-negative breast tumors, and decreased breast cancer mortality over a 20-year follow-up period. Moreover, it was demonstrated that estrogen-alone therapy did not significantly interfere with breast cancer detection via mammography (Chlebowski et al., 2020).

The Nurses' Health Study II, encompassing 116,608 U.S. female nurses, reported a 33% increase in the risk of developing breast cancer in women who were using any type of OCs (RR = 1.33; 95% CI = 1.03-1.73) (Hunter et al., 2010). A case-control study of 897 breast cancer patients and 1,569 controls demonstrated that current users of OCs were at a 3.1-fold increased risk of developing TNBC (odds ratio (OR) = 3.1; 95% CI = 1.2-7.6) (Dolle et al., 2009). A meta-analysis of 79 case-control studies, including 72,030 breast cancer cases and 123,650 population- or hospital-based controls reported no significant difference in breast cancer risk between OC ever-users and never-users (crude OR = 1.01, 95% CI = 0.95-1.07, p = 0.69). No significant increase in breast cancer risk was observed among premenopausal women who had ever used OCs compared to controls who had never used them (crude OR = 1.06, 95% CI = 0.92-1.22, p = 0.44). Moreover, the risk of developing breast cancer in postmenopausal women or women over 50 years old who had ever used OCs was not statistically different from that of controls who had never used OCs (crude OR = 0.99, 95% CI = 0.89–1.10, p = 0.90). However, the use of OCs prior to the first full-term pregnancy was associated with a significantly elevated risk of breast cancer compared to non-users (crude OR = 1.14, 1.01–1.28, p = 0.04). Moreover, the use of OCs for more than 5 years resulted in a modest yet significant elevation in breast cancer risk (crude OR = 1.09, 95% CI = 1.01–1.18, p = 0.02) (Kanadys et al., 2021).

Reproductive factors such as late age at menopause, early age at menarche, late age at first pregnancy, and low parity may increase a woman's chance of developing breast cancer. A pooled analysis of 117 epidemiological studies, including 118,964 women with breast cancer, reported an increase of 1.029 in the relative risk of breast cancer (95% CI = 1.025-1.032) for every one year delay in menopause. The same pooled analysis also concluded that every one year delay in menarche can decrease the risk of

breast cancer by 5% (RR = 1.050; 95% CI = 1.044-1.057) (Collaborative Group on Hormonal Factors in Breast Cancer, 2012). The age at which a woman has her first child can also play a role in susceptibility to breast cancer, with the risk increasing as the age of the first pregnancy increases (Subramani & Lakshmanaswamy, 2017). A Norwegian cohort study showed that women ≥ 35 years old at the time of their first delivery had a 54% higher chance of developing breast cancer compared with women < 20 years old at their first delivery (HR = 1.54; 95% CI = 1.13-2.11) (Horn et al., 2013). However, the overall protective effect of childbirth is thought to be preceded by an immediate increased risk of the disease during the first 5 years after delivery. The proposed etiology for this transient risk increase includes hormonal stimulation and proinflammatory stimuli during wound healing (Lyons et al., 2011). Using pooled data of ~900,00 women from 15 prospective studies, Nichols et al. observed an elevated risk of breast cancer in parous women compared with nulliparous women for more than 20 years after childbirth. The risk peaked 5 years after giving birth, with an HR of 1.80 (95% CI = 1.63-1.99), but crossed over from positive to negative about 24 years after childbirth and reached an HR of 0.77 (95% CI = 0.67-0.88) after 34 years from childbirth (Nichols et al., 2019).

Breastfeeding is also inversely associated with the overall risk of breast cancer. A meta-analysis of 36,881 breast cancer cases from 27 studies, including 8 cohort and 19 case-control studies, showed that ever breastfeeding had an inverse association with triple-negative breast tumors (OR = 0.78; 95% CI = 0.66-0.91). No association was observed between ever breastfeeding and hormone receptor-positive breast cancers in this meta-analysis (Islami et al., 2015). As such, breastfeeding can have a protective effect against hormone receptor-negative breast tumors, which are more prevalent among younger females and generally have a poorer prognosis than other tumor subtypes.

Lastly, having a family history of breast cancer plays a significant role in susceptibility to the disease. A woman’s risk of developing breast cancer is specifically high if at least one first-degree (i.e., mother, daughter, and full-sister) and/or second-degree (i.e., grandmother, granddaughter, aunt, niece, and half-sister) relatives have been diagnosed with breast or ovarian cancer (Fakhri et al., 2022). A cohort study of

~113,000 women in UK demonstrated that women with one affected first-degree relative had a 1.77-fold higher risk of developing breast cancer compared with women with no affected relatives (HR = 1.77; 95% CI = 1.58-1.97), and the risk became 2.52-fold in women with at least two affected first-degree relatives (HR = 2.52; 95% CI = 1.83-3.47). Interestingly, participants who had a relative diagnosed with breast cancer before 45 years old showed a relative risk of 2.47 (95% CI = 2.01–3.04), which was significantly higher than those with an affected relative diagnosed over the age of 45 (HR = 1.63; 95% CI = 1.45–1.84) (Brewer et al., 2017). A meta-analysis of 58,209 women with breast cancer and 101,986 controls reported that risk ratios for breast cancer increased as the number of affected first-degree relatives increased: for one, two, and three or more affected first-degree relatives, the ratios were 1.80 (95% CI = 1.69-1.91), 2.93 (95% CI = 2.36-3.64), and 3.90 (95% CI = 2.03-7.49), respectively, compared with women with no affected family members (Collaborative Group on Hormonal Factors in Breast Cancer, 2001).

1.2 Hereditary Breast Cancer

Hereditary breast cancer (HBC) accounts for 5%-10% of all breast cancer cases. An individual who inherits a deleterious mutation in a breast cancer predisposition gene from either parent has an increased risk of developing breast cancer (Corso & Magnoni, 2021). Specific disease characteristics can suggest a hereditary component to breast cancer, such as an early age at diagnosis and having bilateral breast cancer (Lynch et al., 2008). This inherited predisposition to breast cancer is often connected with other hereditary cancer syndromes (Lynch et al., 2008). In particular, there is a well-established genetic association between breast and ovarian cancers known as the hereditary breast and ovarian cancer (HBOC) syndrome, which is characterized by the presence of breast and/or ovarian cancers in multiple generations on the same side of a family. Additionally, common features of HBOC syndrome include early age at disease onset, one or more women affected with multiple primary breast cancer, bilateral breast cancer, or both breast and ovarian cancers, and a family history of specific cancers, such as male breast cancer, prostate cancer, and pancreatic cancer (Lu et al., 2019). Deleterious mutations in BReast CAncer gene 1 (BRCA1) and BReast CAncer gene 2 (BRCA2) account for

20%-25% of HBOC patients (Torabi Dalivandan et al., 2021). Over the past decades, several other breast cancer predisposition genes have been discovered, some of which are also associated with the risk of ovarian cancer (explained in detail in Section 1.3).

In search for genetic risk factors associated with breast cancer susceptibility, early linkage analysis and positional cloning in families with multiple affected individuals discovered BRCA1 and BRCA2 as the two major highly penetrant breast cancer susceptibility genes (Miki et al., 1994; Narod et al., 1991; Wooster et al., 1994). Nevertheless, many breast cancer patients with a family history of the disease or an early age at onset were tested negative for mutations in BRCA1 or BRCA2, which suggested the potential presence of other breast cancer predisposition genes. Given that genetic linkage analyses have failed to identify additional breast cancer susceptibility genes with rare and highly penetrant variants, it was speculated that no further highly penetrant genes of comparable significance to BRCA1 and BRCA2 exist (Smith et al., 2006).

It is believed that predisposition to breast cancer can be attributed to three main categories of genetic factors that confer varying degrees of RR: variants with high penetrance, which are associated with an RR greater than 4; variants with moderate penetrance, which are associated with RRs ranging from 2 to 4; and variants with low penetrance, which correspond to an RR less than 1.5 (Apostolou & Fostira, 2013; Easton et al., 2015; Shiovitz & Korde, 2015). Variants of moderate to high penetrance are typically rare, having a minor allele frequency (MAF) of less than 1%. The majority of variants in these two categories have been discovered within the coding or splice junction sequences of breast cancer susceptibility genes (Easton et al., 2015). Various genes have been proposed to harbor such breast cancer risk alleles. Nevertheless, to date, a core panel of nine genes is regarded as well-established moderately to highly penetrant breast cancer susceptibility genes that possess variants with a confirmed association that satisfies rigorous statistical significance and burden testing criteria (Breast Cancer Association Consortium et al., 2021; Hu et al., 2021). These genes will be discussed in detail in Section 1.3. In total, rare high- and moderate-risk variants in established breast cancer susceptibility genes are found in less than half the hereditary breast cancer cases (Shiovitz & Korde, 2015; Torabi Dalivandan et al., 2021).

Over the past few years, new tools such as high-density single nucleotide polymorphism (SNP) arrays have become available for performing genome-wide association studies (GWASs) among large cohorts of patients and controls to identify low-risk alleles associated with an elevated breast cancer risk (Filippini & Vega, 2013). The first three breast cancer GWASs were published simultaneously in 2007. In a study by Stacey et al., 4,554 breast cancer patients and 17,577 healthy individuals of European ancestry were analyzed to discover two SNPs, rs13387042 and rs3803662, associated with a modest increase in breast cancer risk with ORs of 1.2 and 1.28, respectively (Stacey et al., 2007). In the second study, Easton et al. discovered five additional susceptibility loci using a cohort of 4,398 breast cancer patients and 4,316 controls of European ancestry at the discovery phase, and roughly 20,000 cases and 20,000 controls at the validation. These loci included SNPs within or adjacent to MAP3K1, LSP1, FGFR2, TNRC9, and H19 (Easton et al., 2007). In the third GWAS, Hunter et al. performed a GWAS on 2,921 European breast cancer patients and 3,214 healthy individuals and detected four intronic FGFR2 SNPs, thereby validating Easton et al.'s finding (Hunter et al., 2007). To date, more than 100 breast cancer GWASs have been conducted, resulting in the discovery of over 150 low-penetrant SNPs linked to breast cancer risk at a genome-wide level of significance (Yanes et al., 2020). Each breast cancer-associated SNP exerts a little impact on the risk of breast cancer. However, their cumulative effect, represented as the polygenic risk score (PRS), has demonstrated the ability to differentiate risk levels, enabling the stratification of individuals into different categories of disease risk (Dite et al., 2016; Mavaddat et al., 2015). Studies regarding the impact of PRS on breast cancer susceptibility have consistently indicated higher PRSs among breast cancer-affected women relative to population controls (Allman et al., 2015; Läll et al., 2019; van Veen et al., 2018). Research on European populations has shown a minimum two-fold difference in the risk of developing breast cancer between the lowest and highest quartiles of PRS distribution. Similar results have also been observed among African American and Asian women (Chan et al., 2018; Muranen et al., 2016; S. Wang et al., 2018). Collectively, low-risk variants account for approximately 18% of breast cancer heritability (Mavaddat et al., 2019).

The hypothesis that rare variants may significantly contribute to disease risk is compelling, as variants that strongly predispose individuals to disease are expected to be maintained at low allele frequencies due to purifying selection (Skol et al., 2016). Hence, the search for additional moderate- to high-risk breast cancer predisposition genes is still ongoing, given that mutations in the already known susceptibility genes can not explain a large proportion of familial cases (Torabi Dalivandan et al., 2021).

1.2.1 Significance of Hereditary Breast Cancer Research

Research efforts to identify breast cancer susceptibility genes are of great importance. Firstly, unaffected women with an underlying genetic predisposition may benefit from primary preventative measures or early diagnosis through intensive screening. For instance, women with deleterious mutations in BRCA1 or BRCA2 are estimated to have a 60%-80% lifetime risk of developing breast cancer (Daly et al., 2021). Once diagnosed, such patients confer a high risk of developing cancer in the contralateral breast (Kuchenbaecker et al., 2017). If an individual is tested positive for BRCA pathogenic mutations, they may be suited to undergo preventive measures, which may involve chemoprevention, prophylactic bilateral mastectomy, and oophorectomy. Bilateral mastectomy, as the most effective risk-reducing measure, has been shown to reduce the risk of breast cancer by roughly 90% among BRCA mutation carriers (Gronwald et al., 2006). Studies have also suggested that prophylactic salpingo-oophorectomy may decrease the risk of breast cancer by 56% and 46% among BRCA1 and BRCA2 mutation carriers, respectively (Eisen et al., 2005; Kotsopoulos et al., 2012; Rebbeck et al., 2009). Although there are no clear recommendations on risk-reducing surgery in individuals with pathogenic variants in moderate-risk non-BRCA genes, these individuals can benefit from regular screening, including annual breast magnetic resonance imaging (MRI) and mammogram (Daly et al., 2021).

Secondly, the identification of new breast cancer susceptibility genes can enhance our understanding of novel carcinogenesis pathways, which may consequently lead to the development of new therapeutics and individualized treatments for patients with the underlying genetic predisposition. For example, after the discovery of BRCA1 and BRCA2 as breast cancer predisposition genes, clinical trials revealed that breast tumors harboring

germline mutations in these genes can be effectively targeted using DNA damaging agents such as poly ADP ribose polymerase (PARP) inhibitors through a mechanism known as synthetic lethality (Farmer et al., 2005; Tan et al., 2008). BRCA1 and BRCA2 play essential roles in maintaining genomic stability by participating in the homologous recombination repair (HRR) of DNA double-strand breaks (DSBs). HRR pathway is crucial for the error-free repair of DSBs and requires functional BRCA1 and BRCA2 proteins (Jensen et al., 2010). Cells with a BRCA1 or BRCA2 germline mutation have a compromised HRR pathway (Radhakrishnan et al., 2014; Rose et al., 2020). If PARP enzymes are inhibited in cells, single strand DNA breaks are not repaired and transformed to DSBs. In cells lacking functional BRCA1 or BRCA2 proteins, DSBs are repaired by an alternative error-prone mechanism known as the non-homologous end-joining pathway (NHEJ). However, the error-prone nature of this repair pathway leads to the accumulation of DNA damage and ultimate cell death (Fong et al., 2009; Rose et al., 2020). As a result, PARP inhibitors are being considered as a targeted therapeutic approach in the treatment of breast cancer patients who harbor BRCA1 or BRCA2 germline mutations. Similarly, the identification of other breast cancer susceptibility genes and their underlying carcinogenesis pathway may eventually help to develop or enhance personalized medication strategies.

Lastly, gaining knowledge of the hereditary components of breast cancer can also benefit the broader community of breast cancer patients. Again, the use of PARP inhibitors in the treatment of breast cancer patients is a good example. PARP inhibitors were initially proven to be very effective for the treatment of breast cancer patients with deleterious germline mutations in BRCA1 or BRCA2 genes. The application of PARP inhibitors is now being investigated in settings other than germline BRCA-mutated breast tumors. Several clinical studies have reported positive findings for the use of PARP inhibitors in the treatment of patients with somatic BRCA1/2 mutations (Cortesi et al., 2021).

Due to the reasons mentioned above, continuous research for identifying genetic factors that contribute to breast cancer susceptibility has been, and continue to be, of

great importance for reducing incidence as well as mortality rates through early detection and improved personalized treatment strategies.

1.2.2 Research Methods in Hereditary Breast Cancer

Two main strategies have been used in searching for genes that may be associated with an increased risk of developing breast cancer, which include genetic linkage analysis and association studies (Turnbull & Rahman, 2008).

Linkage analysis involves genotyping a set of known genome-wide markers among multiple breast cancer-affected family members. Markers that flank the unknown risk allele tend to highly segregate with disease status. Subsequently, the genomic region surrounding the linked markers is studied further to identify the likely causative gene (Pulst, 1999). Using large breast cancer families, genome-wide linkage analysis could discover BRCA1 and BRCA2 as the two high-penetrance breast cancer susceptibility genes (Hall et al., 1990; Wooster et al., 1994). It is important to note that linkage analysis is suitable for mapping only high-penetrance breast cancer predisposition genes, in which deleterious mutations confer a high risk of developing breast cancer. If a susceptibility gene is of lower penetrance, the correlation between breast cancer and risk alleles within these genes may not be strong enough to generate a significant linkage signal. For this reason, linkage analysis has had no success in identifying moderate- or low-penetrance breast cancer susceptibility genes (Gonzalez-Neira et al., 2007; Hilbers et al., 2013).

The other main approach for identifying breast cancer susceptibility genes has been through case–control association studies. These studies compare the frequency of putative susceptibility variant(s) among breast cancer cases versus unaffected controls. If the frequency is greater among cases compared to controls, it provides evidence that the variant(s) under study may be associated with a higher chance of developing breast cancer. Association studies comprise two main strategies, which include the candidate gene approach and GWAS (Turnbull & Rahman, 2008).

In candidate gene studies, a set of genes are selected based on their location in a region of linkage or their function and potential role in carcinogenesis, such as genes

involved in the estrogen pathway, DNA repair, apoptosis, and cell-cycle control. The selected genes are then screened to search for a variant or set of variants that is enriched or depleted in breast cancer cases compared to unaffected controls (Apostolou & Fostira, 2013). Overall, this approach is subject to several limitations. First, the selection of genes is based on a prior knowledge of their function in cancer-related pathways, and therefore, it may preclude the discovery of genes in currently unknown pathogenesis pathways. Second, the causative genes might be in the upstream or downstream of signaling pathways of the selected genes, which can not be identified through this approach. Finally, due to the small sample sizes, most candidate gene studies are likely to be underpowered, which makes their validation and replication in future studies a challenge (Alghamdi & Padmanabhan, 2014).

On the other hand, GWAS offers a hypothesis-free approach in which hundreds of thousands of already known single nucleotide variants (SNVs) are genotyped across the entire genome of the case and control subjects with no prior knowledge of the location of the causal variants. These studies are best suited for testing known common variants (i.e., with MAF $\geqslant$ 5%) that confer small effect sizes (Pe'er et al., 2006). Despite the success of GWASs in identifying many genotype-phenotype associations, this study design has not been without controversy. One important limitation of GWASs is the high cost required to genotype hundreds of thousands of SNVs among thousands of individuals, which may cause pressure to confine the sample size with a consequent reduction in the study power to detect association for common variants that confer modest effect sizes (Tam et al., 2019). Another challenge is that GWASs are not able to identify high-risk rare variants, as these variants are not tagged well by the common tagging SNVs that are currently used in GWASs. Moreover, the lack of thorough knowledge about the entire human genes, their biological functions, and relativity to human disease complicates the interpretation of many loci found to be disease-associated in GWAS (Alghamdi & Padmanabhan, 2014).

1.3 Well-Established Breast Cancer Susceptibility Genes

Various breast cancer susceptibility genes have been identified using the above-mentioned approaches. To this date, a core panel of nine genes is regarded as confirmed moderately to highly penetrant breast cancer susceptibility genes (Breast Cancer Association Consortium et al., 2021; Hu et al., 2021). A brief description of each gene is summarized below.

1.3.1 BReast CAncer Gene 1 (BRCA1)

BRCA1 is a tumor suppressor gene and encodes a nuclear phosphoprotein, which is involved in maintaining genomic stability through HRR of DNA double-strand breaks (Jensen et al., 2010). The BRCA1 protein acts with other tumor suppressors and signal transducers to form a large protein complex known as the BRCA1-associated genome surveillance complex (BASC) (Y. Wang et al., 2000). In BRCA1-deficient cells, DNA double-strand breaks are repaired by an alternative mechanism known as non-homologous end joining. However, the error-prone nature of this mechanism leads to the accumulation of DNA damage across the genome, which is believed to be a crucial feature of carcinogenesis (Fong et al., 2009). BRCA1 protein is also involved in other cellular functions, such as cell-cycle control, centrosome duplication, protein ubiquitylation, and chromatin remodeling at sites of double-strand DNA breaks (Venkitaraman, 2002, 2004).

In 1988, segregation analysis of breast cancer in high-risk families provided evidence that the familial clustering of breast cancer in 4% of the families could be due to a highly penetrant autosomal dominant susceptibility gene (Newman et al., 1988). Subsequently, linkage analysis on 23 families with a strong family history of breast cancer provided the first report of linkage of early-onset familial breast cancer to chromosome 17q21 (Hall et al., 1990). Shortly thereafter, linkage analysis in several other high-risk families confirmed that chromosome 17q21 was the locale of a gene for inherited susceptibility to breast cancer and reported for the first time that the patients with the linkage signal conferred a significantly higher risk of developing ovarian cancer as well

(Narod et al., 1991). Finally, in 1994, positional cloning identified BRCA1 as the breast and ovarian cancer susceptibility gene candidate (Miki et al., 1994).

BRCA1 is a highly penetrant breast cancer susceptibility gene. Early studies of large cancer families have suggested that the risk of developing breast cancer by the age of 70 years can be as high as 87% among women who carry BRCA1 pathogenic/likely pathogenic mutations (Easton et al., 1995; Ford et al., 1998). In 2021, a seminal study by the Breast Cancer Association Consortium (BCAC), comprising 48,826 breast cancer patients unselected for family history and 50,703 population controls, identified 515 carriers of BRCA1 protein-truncating variants among breast cancer cases compared to 58 carriers among controls (OR = 10.57, 95% CI = 8.02-13.93, $p = 1.1\times10^{-62}$). The study also estimated a lifetime risk of approximately 55% for developing breast cancer by the age of 80 among BRCA1 protein-truncating mutation carriers (Breast Cancer Association Consortium et al., 2021). Breast cancer risk in women harboring BRCA1 mutations is considerably higher below the age of 40 years and becomes less dramatic with increasing age. Patients with BRCA1 mutations are at an increased risk of developing contralateral breast cancer (Kuchenbaecker et al., 2017). BRCA1 is also a highly penetrant ovarian cancer susceptibility gene, and deleterious mutations in this gene confer a lifetime risk of about 40% for developing ovarian cancer (Gabai-Kapara et al., 2014; Pietragalla et al., 2020). Moreover, BRCA1 mutations are associated with an elevated risk of male breast cancer, which is estimated to be 1.2% by the age of 70 years (Tai et al., 2007). BRCA1-associated tumors share distinct pathological features and are typically high-grade invasive ductal carcinomas with triple-negative phenotypes (i.e., lack of ER, PR, and HER2 expression) (Fostira et al., 2012). These tumors are among the most aggressive breast cancer subtypes and are associated with a poor prognosis (X. Li et al., 2017). As explained in Section 1.2.1, the use of DNA-damaging agents such as PARP inhibitors is a targeted therapeutic approach for the treatment of breast cancer patients with BRCA1 mutations (Rose et al., 2020).

1.3.2 BReast CAncer Gene 2 (BRCA2)

BRCA2 is a tumor suppressor gene and encodes a nuclear protein involved in the repair of damaged DNA through HRR (Apostolou & Fostira, 2013). Other roles of the

BRCA2 protein in maintaining genomic integrity include mediating cellular resistance to various DNA damaging agents, regulating DNA damage checkpoint arrest, stabilizing stalled replication forks, and protecting them from degradation of nascent DNA (Andreassen et al., 2021).

Shortly after the identification of BRCA1 as a breast cancer predisposition gene, in 1994, an international team of researchers performed a linkage analysis on 15 families with early onset breast cancer who were negative for BRCA1 mutations. This effort provided evidence of the linkage between breast cancer and a locus on chromosome 13q12, which led to the discovery of BRCA2 as the second breast cancer susceptibility gene candidate (Wooster et al., 1994).

BRCA2 is a highly penetrant breast cancer susceptibility gene. Early studies of large cancer families have suggested that the risk of developing breast cancer by the age of 70 years can be as high as 84% among women who carry BRCA2 mutations (Ford et al., 1998). The BCAC seminal study in 2021 reported BRCA2 protein-truncating mutations among 754/48,826 unselected breast cancer patients and 135/50,703 population controls (OR = 5.85, 95% CI = 4.85-7.06, $p = 2.2\times10^{-75}$), confirming BRCA2 as a high-risk breast cancer gene. The study also estimated a lifetime risk of ~45% for developing breast cancer by the age of 80 among BRCA2 mutation carriers (Breast Cancer Association Consortium et al., 2021). Unlike BRCA1 mutations, the risk of breast cancer varies substantially for different BRCA2 mutations. For example, breast cancer risk associated with a deleterious founder BRCA2 mutation (BRCA2 c.6714delT) in Ashkenazy Jewish women has been reported to be half of the risk in women with other BRCA2 mutations (Narod & Salmena, 2011). BRCA2-associated tumors do not appear to exhibit specific pathological characteristics, with the exception of an increased frequency of high-grade tumors, but this association is weaker than that of BRCA1-associated tumors (Kim et al., 2020). Carriers of BRCA2 mutations are also at an increased risk of ovarian cancer and male breast cancer. Pathogenic and likely pathogenic mutations in BRCA2 are associated with a lifetime risk of 11%-18% for developing ovarian cancer (Chen & Parmigiani, 2007; Pietragalla et al., 2020). For male breast cancer, the lifetime risk is 6.8%, which is higher than that of BRCA1 mutations (Tai et al., 2007). Regarding the therapeutic approach, BRCA2-associated tumors can benefit from treatment with DNA-damaging agents such

as PARP inhibitors that act through the mechanism of synthetic lethality (Rose et al., 2020).

Together, BRCA1 and BRCA2 mutations account for 5%-10% of all female breast cancers, 15%-20% of all familial breast cancers, and 20%-25% of all hereditary breast cancer cases (Nielsen et al., 2016; N. Tung et al., 2016; Turnbull & Rahman, 2008).

1.3.3 BRCA1-Associated RING Domain (BARD1)

BARD1 is a tumor suppressor gene located on chromosome 2q35 and encodes a protein that interacts with the N-terminal of BRCA1, creating a heterodimer that functions during the repair of DNA double-strand breaks and apoptosis initiation (Wu et al., 1996). The first report of BARD1 germline mutations in hereditary breast and/or ovarian cancer was released in 2002 after analyzing 40 breast cancer patients from 40 high-risk families who tested negative for BRCA1/2 mutations. BARD1 screening revealed four different germline mutations in four of the studied patients with no evidence of loss of heterozygosity (Ghimenti et al., 2002). Further studies identified several BARD1 pathogenic mutations among BRCA1/2 negative familial breast cancer patients from different populations, suggesting that deleterious mutations in BARD1 may be linked to an increased risk of breast cancer (Brakeleer et al., 2010; Ratajska et al., 2012). Although BARD1 is regarded as a breast cancer predisposition gene, the rarity of its mutations and lack of sufficient study sample sizes have hampered a precise estimation of the associated breast cancer risk. In 2021, the BCAC study identified BARD1 protein-truncating mutations among 62/48,826 breast cancer patients unselected for family history and 32/50,703 population controls (OR = 2.09, 95% CI = 1.35-3.23, p = 0.00098), confirming BARD1 as a moderate-risk breast cancer susceptibility gene. The study also estimated a lifetime risk of ~25% for developing breast cancer by the age of 80 among BARD1 mutation carriers (Breast Cancer Association Consortium et al., 2021). Deleterious mutations in BARD1 have also been reported among ovarian cancer patients (Ratajska et al., 2012; Walsh et al., 2011), but no confirmed association between BARD1 mutations and susceptibility to ovarian cancer has been concluded (Pietragalla et al., 2020). Since BARD1 is involved in the HRR pathway and is associated with BRCA1

protein, clinical trials are evaluating the efficacy of PARP inhibitors in the treatment of breast cancer patients with BARD1 germline mutations. However, the lack of sufficient numbers of patients with BARD1 mutations has hampered the ability to assess therapeutic response (N. M. Tung et al., 2020).

1.3.4 Ataxia-Telangiectasia Mutated (ATM)

ATM is a tumor suppressor gene located on chromosome 11q22. The encoded protein is a cell-cycle checkpoint kinase that plays a critical role in sensing and signaling the presence of DNA double-strand breaks through downstream interaction with TP53, BRCA1, and CHEK2 proteins involved in DNA repair and cell cycle control (Zaki-Dizaji et al., 2017). ATM mutations can cause “ataxia-telangiectasia”, an autosomal recessive disorder characterized by cerebellar degeneration (ataxia), dilated blood vessels in the eyes and skin, immunodeficiency, increased sensitivity to ionizing radiation, and predisposition to leukemia and lymphoma (H. H. Chun & Gatti, 2004). The proposal of ATM as a breast cancer susceptibility gene first came in 1976 from an epidemiological study that reported an increased risk of breast cancer among female relatives of patients with ataxia telangiectasia who harbored ATM mutations (Easton, 1994; Swift et al., 1990). It took almost two decades after this preliminary observation to compile more evidence from large epidemiological studies and classify ATM as a breast cancer susceptibility gene (Hollestelle et al., 2010). In 2006, a group of researchers from the United Kingdom screened ATM on the germline DNA of 443 highly familial breast cancer patients negative for BRCA1/2 mutations and 521 controls and detected 12 carriers of ATM variants known to cause ataxia-telangiectasia among cases and two carriers among controls (OR = 2.37, 95% CI = 1.51–3.78, p = 0.0003) (Renwick et al., 2006). The BCAC seminal study in 2021 reported ATM protein-truncating mutations among 294/48,826 unselected breast cancer patients and 150/50,703 population controls (OR = 2.10, 95% CI = 1.71-2.57, p = 9.2×10^{-13}), concluding that ATM is a moderately penetrant breast cancer susceptibility gene. The study also estimated a lifetime risk of ~25% for developing breast cancer by the age of 80 among ATM mutation carriers (Breast Cancer Association Consortium et al., 2021). Moreover, large studies have identified ATM as a moderate-risk ovarian cancer gene, conferring an approximately two- to three-fold increase in the risk of ovarian cancer

among ATM mutation carriers when compared to population controls (Lilyquist et al., 2017; Lu et al., 2019).

1.3.5 Checkpoint Kinase 2 (CHEK2)

CHEK2 is a tumor suppressor gene located on chromosome 22q12 and encodes a G2 checkpoint kinase with a crucial role in the repair of damaged DNA. The CHEK2 protein is activated by ATM in response to DNA double-strand breaks or replicative stress, which in turn triggers the phosphorylation of downstream substrates, including tumor suppressors TP53 and BRCA1 (Nevanlinna & Bartek, 2006). Originally, heterozygous germline mutations in CHEK2 were found by Bell et al. in families with Li-Fraumeni Syndrome (LFS), an autosomal dominant hereditary disorder that predisposes the affected individuals to various early-onset malignancies, including sarcoma, breast cancer, and brain tumors (Bell et al., 1999). The role of CHEK2 in breast cancer predisposition was first proposed by the same group of researchers when they identified a truncating mutation in this gene, the CHEK2 c.1100delC mutation, in three individuals in one of the studied families, two of whom were affected by breast cancer (Bell et al., 1999). Shortly thereafter, another group of researchers screened for the CHEK2 c.1100delC mutation in a cohort of 1,035 unselected breast cancer cases and 1,885 controls from Finland. Although the CHEK2 c.1100delC mutation frequency was not statistically different between the entire cohort of cases compared to controls, a significantly higher frequency was observed among 358 patients with a positive family history of breast cancer compared to the population controls (11/358 cases versus 26/1,885 controls; OR = 2.27, 95% CI = 1.11–4.63, p = 0.021). Further analysis of the CHEK2 c.1100delC mutation in an independent series of 507 familial breast cancer patients tested negative for BRCA1/2 mutations also demonstrated a significantly elevated frequency of CHEK2 c.1100delC mutation compared to the controls (28/507 cases versus 26/1,885 controls; OR = 4.2, 95% CI = 2.4–7.2, p = 0.0002) (Vahteristo et al., 2002). Since then, the CHEK2 c.1100delC mutation has been reproducibly associated with the risk of breast cancer among different populations. A meta-analysis of 26 studies on CHEK2 c.1100delC mutation and breast cancer risk, comprising 118,735 breast cancer cases and 195,807 controls, concluded that this variant is a moderately penetrant

breast cancer risk allele among unselected breast cancer cases (OR = 2.88, 95% CI = 2.63-3.16) and contributes to a higher risk among familial breast cancer cases (OR = 3.21, 95% CI = 2.41-4.29). The CHEK2 c.1100delC mutation is also associated with a three-fold increased risk for male breast cancer (OR = 3.13, 95% CI = 1.94-5.07) (Liang et al., 2018). The BCAC study in 2021 reported CHEK2 protein-truncating mutations among 704/48,826 unselected breast cancer patients and 315/50,703 population controls (OR = 2.54, 95% CI = 2.21-2.91, $p = 3.1\times10^{-39}$), confirming CHEK2 as a moderate-risk breast cancer susceptibility gene. The CHEK2 c.1100delC mutation accounted for ~80% of the total protein-truncating variants. The effect size for this mutation (OR =2.66, 95% CI = 2.27-3.11, $p = 1.1\times10^{-33}$) did not differ significantly from the effect size for all other CHEK2 protein-truncating variants (OR = 2.13; 95% CI = 1.60-2.84; $p = 3.0\times10^{-7}$). The study also estimated a lifetime risk of ~27% for developing breast cancer by the age of 80 among CHEK2 mutation carriers (Breast Cancer Association Consortium et al., 2021). CHEK2 mutation carriers are more likely to develop breast cancer at a younger age and are at an increased risk of contralateral breast cancer and a poorer prognosis compared to non-carriers (Kriege et al., 2014; Yadav et al., 2023). Studies on the role of CHEK2 pathogenic mutations in susceptibility to ovarian cancer has been inconsistent, with some studies reporting a moderately elevated risk and others reporting no statistically significant associations (Lilyquist et al., 2017; Szymanska-Pasternak et al., 2006).

1.3.6 Partner and Localizer of BRCA2 (PALB2)

PALB2 is a tumor suppressor gene located on chromosome 16p12. The precipitation of BRCA2-containing complexes revealed PALB2 as a novel protein with a key role in the localization and stability of BRCA2, thus promoting BRCA2-mediated DNA repair through HRR pathway (Xia et al., 2006). Knockdown of the PALB2 gene has shown to induce chromosomal damage, which can predispose to Fanconi Anemia, a rare recessive hereditary disorder characterized by chromosomal instability, growth retardation, congenital malformation, bone marrow complications, and susceptibility to early-onset malignancies including breast cancer (Reid et al., 2007; Xia et al., 2006). Biallelic PALB2 mutations predispose to a subset of Fanconi Anemia with a phenotype similar to that caused by biallelic BRCA2 mutations (Reid et al., 2007). Prompted by this

observation, in 2007, Rahman et al. investigated the role of monoallelic PALB2 mutations in susceptibility to breast cancer. Sequencing of the gene revealed PALB2 truncating mutations in 10/923 British familial breast cancer patients negative for BRCA1/2 mutations compared with 0/1,084 controls (OR = 2.3, 95% CI = 1.4-3.9, p = 0.0025), suggesting that PALB2 can be considered as a moderately penetrant breast cancer susceptibility gene (Rahman et al., 2007).

Since then, the role of PALB2 deleterious mutations in susceptibility to breast cancer has been replicated in various studies largely consisting of familial breast cancer cases, with some studies reporting PALB2 as a moderately penetrant breast cancer gene and others suggesting PALB2 as a highly penetrant breast cancer gene (Antoniou et al., 2014; Casadei et al., 2011; Wilcox et al., 2023; X. Yang, Leslie, et al., 2020). The BCAC seminal study in 2021 reported PALB2 protein-truncating mutations among 274/48,826 unselected breast cancer patients and 55/50,703 population controls (OR = 5.02, 95% CI = 3.73-6.76, $p = 1.6\times10^{-26}$), therefore confirming PALB2 as a highly penetrant breast cancer susceptibility gene. The study also estimated a lifetime risk of ~40% for developing breast cancer by the age of 80 among PALB2 mutation carriers (Breast Cancer Association Consortium et al., 2021).

Patients harboring PALB2 mutations are more likely to develop high-grade, triple-negative cancers at a younger age and generally have worse survival rates compared to noncarriers. Given the involvement of PALB2 in the HRR pathway, some clinical trials are evaluating the efficacy of PARP inhibitors in patients with PALB2 germline mutations, and promising responses have been reported (N. M. Tung et al., 2020). Studies regarding the role of PALB2 mutations in predisposition to ovarian cancer have been inconsistent with their findings. Some studies have reported that PALB2 mutations occur at similar frequencies among ovarian cancer patients and population controls (Ramus et al., 2015). However, a recent study with the largest cohort of individuals with PALB2 pathogenic variants to date, comprising 976 individuals from 524 families, demonstrated a relative risk of 2.91 (95% CI = 1.40–6.04, $p = 4.1\times10^{-3}$) for developing ovarian cancer among PALB2 mutation carriers compared to population controls (X. Yang, Leslie, et al., 2020).

1.3.7 DNA Repair Protein RAD51 Paralogs C and D (RAD51C & RAD51D)

RAD51C and RAD51D genes are located on chromosomes 17q22 and 17q12, respectively, and encode two proteins that are involved in fundamental roles such as HRR through creating complexes with other RAD51 paralogs (J. Chun et al., 2013). In human cell lines, mutant RAD51C and RAD51D genes, along with other mutant RAD51 paralogs, have been linked to compromised HRR, increased double-strand breaks, genomic instability, and growth defects (Garcin et al., 2019).

RAD51C/D were first discovered as ovarian cancer susceptibility genes, and it took almost a decade to clarify their association with breast cancer risk. Initial studies on the role of RAD51C/D pathogenic mutations in susceptibility to breast cancer were inconsistent, with some studies reporting an elevated risk and others reporting no statistically significant associations (Couch et al., 2015; N. Li et al., 2019; Shimelis et al., 2018). Subsequently, in 2020, Yang et al. screened RAD51C/D among ~6,500 families from 12 countries with multiple breast cancer affected members and identified 73 female breast cancer patients with RAD51C pathogenic mutations (from 125 families) and 30 female breast cancer patients with RAD51D pathogenic mutations (from 60 families). Considering 42,325 cancer-free individuals from the UK Biobank database, the relative risk of developing breast cancer was estimated to be 1.99 for RAD51C mutation carriers (95% CI = 1.39-2.85, $p = 1.55 \times 10^{-4}$) and 1.83 for RAD51D mutation carriers (95% CI = 1.24-2.72, p = 0.002). For both genes, breast cancer risk increased by the age of 60, resulting in the cumulative breast cancer risk of 21% for RAD51C and 20% for RAD51D mutation carriers by the age of 80 years (X. Yang, Song, et al., 2020). The BCAC study in 2021 identified RAD51C protein-truncating mutations among 54/48,826 unselected breast cancer patients and 26/50,703 population controls (OR = 1.93, 95% CI = 1.20-3.11, p = 0.0070), confirming RAD51C as a moderate-risk breast cancer susceptibility gene. For RAD51D, protein-truncating variants were observed among 51/48,826 breast cancer patients and 25/50,703 controls (OR = 1.80, 95% CI = 1.11-2.93, p = 0.018), concluding that RAD51D is also a moderate-risk breast cancer gene. The study estimated lifetime risks of ~22% and ~20% for developing breast cancer by the age of 80 among

RAD51C and RAD51D mutation carriers, respectively (Breast Cancer Association Consortium et al., 2021).

RAD51C/D pathogenic mutations are strongly associated with the risk of ovarian cancer (Loveday et al., 2011; Meindl et al., 2010). Studies with large cohorts have confirmed RAD51C/D pathogenic variants as high-risk ovarian cancer susceptibility alleles that confer about seven-fold increased risk of ovarian cancer. Carriers of RAD51C and RAD51D deleterious variants have approximately 11% and 13% lifetime risks of developing ovarian cancer by the age of 80, respectively (X. Yang, Song, et al., 2020). Moreover, studies have shown that RAD51C/D deficient cells tend to be sensitive to PARP inhibitors, suggesting a potential therapeutic approach for RAD51C/D-associated tumors (Loveday et al., 2011; Somyajit et al., 2015).

1.3.8 Tumor Suppressor 53 (TP53)

The tumor suppressor gene TP53 is located on chromosome 17p13 and encodes a protein activated in response to diverse cellular stress signals such as DNA damage, hyperproliferation, hypoxia, and oncogene activation. Due to its crucial role in maintaining genomic stability, TP53 protein has been referred to as “the guardian of the genome”, and loss of its function is believed to suppress protection mechanisms against the accumulation of genetic alterations (Tonnessen-Murray et al., 2017). Inherited TP53 mutations are associated with Li-Fraumeni syndrome (LFS), a rare autosomal dominant disorder characterized by a high predisposition to several early-onset malignancies, such as sarcomas, osteosarcomas, breast cancer, brain tumors, leukemia, and adrenocortical carcinomas (Malkin et al., 1990).

Breast cancer is the most frequent malignancy among LFS families, suggesting that TP53 must have an important role in susceptibility to breast cancer (Chompret et al., 2000). In 1990, researchers analyzed TP53 expression in 11 human breast cancer cell lines by immunohistochemistry and observed abnormalities in all the studied cell lines, speculating that TP53 must be a potent breast cancer susceptibility gene (Bartek et al., 1990). Subsequently, in 1991, another group of researchers studied the relationship between TP53 expression, gene mutation, and loss of the wild-type allele in 49 primary

human breast tumors. TP53 screening in seven of these tumors revealed the presence of a TP53 mutation in each of the sequenced tumors that altered the coding sequence in a highly conserved region of the gene, and loss of the TP53 wild-type allele was evident in four of the seven tumors (Davidoff et al., 1991).

TP53 pathogenic mutations are responsible for ~1% of hereditary breast cancer and 5%-8% of early-onset breast cancer cases with no family history of the disease (Martin et al., 2003; McCuaig et al., 2012). The BCAC seminal study in 2021 reported TP53 protein-truncating mutations among 7/48,826 unselected breast cancer patients and 2/50,703 population controls (OR = 3.06, 95% CI = 0.63-14.91, p = 0.17). The low estimate may reflect the rarity of germline TP53 pathogenic mutations and the significant early mortality among carriers of these mutations. The absolute risk of developing breast cancer was not calculated for TP53 in this study because of the wide 95% CI for the OR and the known association of TP53 mutations with a considerable risk of malignancy in childhood (Breast Cancer Association Consortium et al., 2021). However, previous studies have reported a lifetime breast cancer risk of up to 85% by the age of 60 years in women with deleterious mutations in TP53 compared to non-carriers (Aloraifi et al., 2015; Schon & Tischkowitz, 2018). Breast cancer in these patients manifests at a very early age, with a median age at diagnosis of 34 years, and tends to be ductal in histology, hormone receptor positive, and/or HER2 positive (Schon & Tischkowitz, 2018). TP53-associated tumors are more aggressive and have increased resistance to chemotherapy than TP53 wild-type tumors (Shahbandi et al., 2020). Due to the rarity of TP53 mutations, many studies have failed to find a statistically significant association between TP53 mutations and the risk of ovarian cancer (Castéra et al., 2018; Norquist et al., 2016). However, in a study of 2,051 ovarian cancer patients and 3,988 unaffected controls, pathogenic mutations in TP53 were associated with an approximately 18-fold increased risk of developing ovarian cancer (OR = 18.50, 95% CI 2.56-808.1) (Lu et al., 2019).

1.4 Advances of New DNA Sequencing Technologies

The advent of Next-Generation Sequencing (NGS) technology has revolutionized efforts in the discovery of disease-associated genes over the past decade (Kamps et al.,

2017). In contrast to traditional sequencing approaches, such as Sanger sequencing, that were capable of sequencing only one DNA fragment at a time, NGS technology permits massively parallel sequencing of millions of DNA fragments, with considerable improvements in efficiency, affordability, yield, and sensitivity. These key features have made NGS technology the pioneering approach for the identification of novel genes responsible for cancer susceptibility (Kerkhof et al., 2017). Two main NGS strategies are commonly used in scientific research, which include whole-genome sequencing (WGS) and whole-exome sequencing (WES).

WGS is the most comprehensive NGS approach as it screens the entire genome, including both coding and non-coding regions. Compared to WES, sequencing the entire genome is substantially more expensive and resource-intensive, (Cirulli & Goldstein, 2010). Moreover, interpreting the functional significance of many intronic variants, variants detected in genes with unknown functions or the ones outside of the gene coding regions can be very challenging due to the lack of adequate knowledge about the entirety of the human genome. In other words, if WGS identifies disease-associated variants in genomic regions with unknown function or biological relevance, such findings would be deemed unusable in finding disease susceptibility loci (Petersen et al., 2017).

Currently, WES is the most frequently used NGS approach to look for rare disease-associated variants. Compared to WGS, WES is considerably more cost-effective as it only captures coding genomic regions, which represent ~1%-2% of the entire human genome, and the detected variants are much easier to interpret (Kiezun et al., 2012). Although the majority of deleterious mutations that confer disease risk appear to be in the coding regions of the genome, overlooking the impact of non-coding variants on disease susceptibility is considered the major limitation of the WES approach (Koch, 2014).

In search of novel breast cancer susceptibility genes, various WES studies have been conducted so far with the hypothesis that a large portion of unexplained familial breast cancer cases can be attributed to moderately or highly penetrant rare risk alleles in novel susceptibility genes that are not discovered yet (Sokolenko et al., 2015). In this regard, two different WES study designs have been utilized, which include the family-based approach and WES a cohort of unrelated familial breast cancer patients (Chandler

et al., 2016). In the family-based approach, multiple breast cancer-affected and healthy family members will be exome-sequenced to achieve a list of shared variants among the sequenced breast cancer-affected individuals, which presumably include the inherited breast cancer risk allele. Subsequently, variants of highest interest (i.e., variants that segregated well with the disease) will be investigated further through independent case-control association studies (Chandler et al., 2016). As more distantly related breast cancer-affected family members are exome sequenced, the list of shared variants will substantially become narrower, making it more likely to identify the familial breast cancer risk allele. However, WES more distantly related individuals may also increase the chance of sequencing phenocopies (i.e., individuals with a sporadic breast cancer) (Sham & Purcell, 2014). Overall, this strategy is most powerful in the identification of highly penetrant risk alleles. Owing to breast cancer mortality rates and the genetically heterogeneous nature of the disease with reduced penetrance and the existence of phenocopies, WES a small number of family members seems not to be promising in the identification of novel breast cancer risk genes with intermediate or low penetrance (Chandler et al., 2016).

The second study design involves WES a cohort of unrelated familial breast cancer patients and healthy individuals. This strategy does not assume high penetrance and therefore can identify moderate risk alleles as well. Because of locus and allelic heterogeneity, these studies are more powerful when more genetically homogenous populations are studied and/or when the cohort size increases. Ultimately, variant(s) of highest interest are investigated further to determine their significance in susceptibility to breast cancer; this can involve genotyping the variant(s) or performing full-gene sequencing in an independent and usually larger cohort of cases and controls (Chandler et al., 2016).

1.4.1 Lessons Learned from Hereditary Breast Cancer WES Studies

While the aforementioned DNA sequencing advances appeared promising and raised hope for the discovery of additional breast cancer susceptibility genes, most studies have been unsuccessful in identifying novel moderately or highly penetrant breast

cancer predisposition genes through the application of WES (Gracia-Aznarez et al., 2013; Snape et al., 2012; Thompson et al., 2012).

Although most of these studies recruited familial breast cancer patients, many factors can be blamed for their disappointing results. Firstly, an optimal sample size is a crucial requirement when looking for moderately or highly penetrant disease susceptibility genes. Since pathogenic risk alleles seem to be very rare, it is critical to have a large enough sample size to statistically identify disease-associated variants with very low MAFs (Petersen et al., 2017). However, the majority of the studies have not met this requirement and thus could not report the discovery of any novel breast cancer risk genes.

Furthermore, an ideal study design should recruit cases and controls with closely matched characteristics such as age, gender, and population background that would otherwise act as confounding factors. Population-level genetic studies often consider ethnicity as a confounding factor due to observed variations in allele frequencies across populations. However, it is essential to recognize that human genetic diversity exists on a continuum, and sociocultural categories like ethnicity may not always align with genetic variation. As such, the reliance on ethnic categories in genetic-association studies can obscure the continuous and overlapping nature of genetic variations. These complexities highlight the importance of considering more precise metrics, such as genetic ancestry, rather than sociocultural labels, to more accurately capture the nuances of genetic variation and reduce the risk of spurious findings (Keen-Kim et al., 2006; National Academies of Sciences, 2024). To address these challenges, some studies have proposed using relatives, such as the cases' siblings or parents, as controls. However, it is frequently difficult to recruit a sufficiently large number of family members to achieve an optimal sample size. Another disadvantage of using family members is that, due to incomplete penetrance, unaffected family members harboring a disease-associated mutation are more likely to be recruited as controls compared to studies with population-based controls (Hirschhorn, 2005).

Most importantly, the majority of unsuccessful studies recruited their study subjects from populations with a high genetic variation background (Gracia-Aznarez et al., 2013; Snape et al., 2012; Thompson et al., 2012). It is extremely difficult to detect recurrent risk

alleles existing at very low frequencies in such genetically heterogeneous populations, which may explain, for the most part, why these studies failed to find any novel breast cancer susceptibility gene. This issue can be solved by recruiting study subjects from more genetically homogeneous populations, such as founder populations. Such a strategy can provide a more homogenous cohort with great potential in the discovery of breast cancer susceptibility genes (Chandler et al., 2016). A founder population forms when a population is established by a small group of people separated from a larger population or when a bottleneck occurs, and inbreeding becomes inevitable due to geographic or religious isolation. In the process of population expansion (inbreeding over generations), specific rare mutations can amplify and become more prevalent in the population (Neuhausen, 2000a). This means that there is less locus and allelic heterogeneity in a founder population compared to large mixed populations, and therefore, a larger proportion of diseased individuals will carry a recurrent genetic alteration within a specific susceptibility gene (Neuhausen, 2000a). Once a founder mutation or a novel predisposition gene is identified in a founder population, researchers can further investigate and validate their findings among other populations or the same founder population at a lower cost (Neuhausen, 2000b). Examples of founder populations are the Finnish population, Icelandic population, French-Canadians, Ashkenazi Jews, Polish population, and Bahamian population.

Chapter 2 – Preliminary Work, Rationale, Hypothesis and Research Aims

2.1 Preliminary Work

2.1.1 Polish Breast Cancer Gene Discovery Study

Poland is a relatively homogeneous country from a genetic perspective. The genetic variation background in the Polish founder population is much less than what is seen in western European countries or North America, and the spectrum of mutant alleles may be limited (Szwiec et al., 2015). To date, researchers at Women's College Hospital, in collaboration with scientists at Pomeranian Medical University in Poland, have identified several founder mutations which predispose to breast cancer (3 in BRCA1, 4 in CHEK2, 2 in PALB2, 1 in NBN, and 1 in RECQL) (Cybulski, Carrot-Zhang, et al., 2015; Cybulski et al., 2011; Cybulski, Kluźniak, et al., 2015; Górski et al., 2003). Nonetheless, a large proportion of Polish familial breast cancer patients (i.e., patients with high familial clustering of breast cancer) are negative for mutations in known susceptibility genes, indicating that not all candidate breast cancer susceptibility genes in this population have been identified. Therefore, it is hypothesized that a proportion of hereditary breast cancers in the Polish population (in which no known mutations are detected) may be due to rare founder mutations in a number of moderately to highly penetrant genes that are not identified yet.

In search for novel breast cancer susceptibility genes, our team at Women's College Research Institute in Toronto and Pomeranian Medical University in Poland performed a gene-discovery project among the Polish founder population (Cybulski et al., 2023). At the discovery phase of the study, WES was conducted on the germline DNA of 510 probands with a strong family history of breast cancer and 308 unaffected women. Cases (age range 28–76, mean age 49.1 years) were diagnosed with invasive breast cancer from 2000 to 2017 at 16 different centers in Poland, and all were negative for

genetic variants in known genes associated with breast cancer risk (BRCA1, BRCA2, BARD1, ATM, CHEK2, TP53, PALB2, PTEN, NBN, RECQL, RAD50, and CDH1). Control subjects were selected randomly from a registry of individuals who participated in a population-based study in Poland (Wokołorczyk et al., 2020). These individuals were selected randomly based on the criteria that they were cancer-free females ≥ 40 years of age (mean age of 59.5 years) and had no family history of cancer in their first-degree relatives. The genetic profile of the study participants was then analyzed. The focus was on loss of function (LoF) variants (frameshift indels, stop codon gain, stop codon loss, essential splicing site, and start codon loss mutations) with a MAF of ≤ 1% among different ethnic groups in gnomAD and 1,000 Genomes Project databases. Genes with confirmed LoF variants among study subjects were selected for gene-based association analyses (N = 7635), and the ones that showed a positive association with breast cancer risk (i.e., age-adjusted OR > 4, with any p-value) were chosen for the validation phases of the project (N = 75). Among the candidate genes, the focus was on those with known functions related to cancer pathogenesis and with a recurrent LoF variant. Based on this criteria, a short list of nine genes, with nine recurrent variants, was chosen for further evaluation. On top of the list, there was the ATRIP gene (NM_130384.3) with three individuals harboring LoF variants among the 510 affected women and no carriers among the 308 control subjects at the discovery phase (OR = 4.23, p = 0.29). One affected woman had ATRIP c.69_75dup (p.Thr26Alafs*23) frameshift insertion and two others had ATRIP c.1152_1155del (p.Gly385Ter) frameshift deletion in this gene. Of note, the frequency of ATRIP LoF variants among 22,928 cancer-free non-Finnish European females of the ExAC database is 1 in every 2,293 women (Karczewski et al., 2017).

Considering this observation, the recurrent ATRIP c.1152_1155del variant was genotyped in a much larger (non-overlapping) cohort of 16,085 Polish women with unselected breast cancer and 9,285 unaffected Polish women at the validation phase. The ATRIP c.1152_1155del variant (Figure 1) was detected in 42 of 16,085 affected women (1 in every 383 women with breast cancer, 0.3%) compared to 11 of 9,285 control subjects (1 in every 844 individuals, 0.1%) (age-adjusted OR = 2.14, 95% CI = 1.13–4.28, p = 0.02). All 42 patients who carried this germline founder variant were tested negative for mutations in the already known breast cancer susceptibility genes (Cybulski et al.,

2023). Detailed information regarding ATRIP and its cell function is provided in Section 3.5.

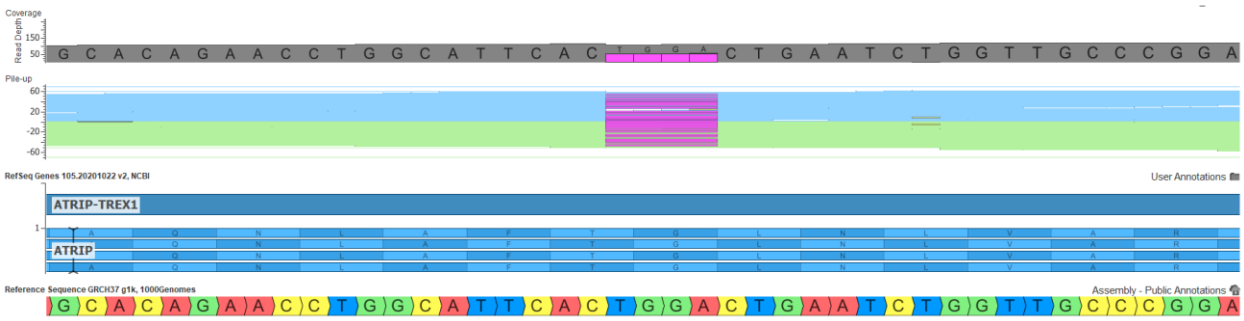


Figure 1: Next generation sequencing reads pileup of ATRIP c.1152_1155del mutation identified among the Polish founder population along with RefSeq V105 and GRCH37 human genome reference regions were shown.

2.1.2 Bahamian Breast Cancer Gene Discovery Study

The Bahamian population is another genetically homogenous founder population inhabiting the Bahamas, a group of islands in the Caribbean with roughly 300,000 inhabitants. There are approximately 100 new cases of breast cancer diagnosed among the Bahamian population every year (Akbari et al., 2014). More importantly, Bahamian women have shown to develop breast cancer at a younger age, with an average age of 42 years at the time of diagnosis as compared with 62 years in the Canadian population (Donenberg et al., 2011; Seung et al., 2020; Smigal et al., 2006). This young age at onset suggests the involvement of genetic components in the development of breast cancer in the Bahamian population. About 27% of Bahamian women with breast cancer carry a deleterious mutation in BRCA1 or BRCA2 genes, which exceeds that of women in any other country (Akbari et al., 2014). It is possible that mutations in other novel genes may predispose to hereditary breast cancer in this population.

In search for novel moderately to highly penetrant breast cancer susceptibility genes among the Bahamian population, our team at Women's College Hospital

performed WES on the germline DNA of a cohort of 96 Bahamian women with a strong family history of breast cancer. All patients were previously tested negative for BRCA1/2 mutations. A strict set of inclusion criteria was applied to increase the chance that the patients' breast cancer is of the hereditary kind; each patient must have had at least two first or second-degree relatives affected with breast/ovarian cancer. If a patient had no family history of breast/ovarian cancer, they had to have a diagnosis prior to the age of 40 years old. The main goal was to search for LoF variants in novel genes that may be associated with breast cancer risk. To approach this aim, two sets of genes were used to select candidate genes. The first set of genes was obtained from the Cancer Gene Census (CGC) database (Sondka et al., 2018) and the second set, called Cancer Relevant Genes (CRG), was a panel constructed by our research team. This panel included genes which are involved in DNA repair and cell cycle control pathways (http://www.genome.jp/kegg/). Both CRG and CGC include genes that are related, in some way, to cancer pathogenesis. The selection of the candidate genes from these two sets was based on the number of LoF mutation carriers per gene and biological involvement in mechanisms relevant to carcinogenesis. Ten candidate genes were ultimately selected from the prioritized list of candidate genes for further investigation in the second phase of the study.

At the validation phase of the project, targeted full-gene sequencing of the ten candidate genes was performed on the germline DNA of 387 unselected breast cancer patients and 653 unaffected controls. All patients were negative for BRCA1/2 mutations. Subsequently, gene-based association analysis was performed using Fisher's exact test. For three of the ten candidate genes, a marginally significant higher frequency was observed in the number of carriers in cases versus control subjects. The first gene was the DNA Cross-Link Repair 1C (DCLRE1C) gene. Three patients carried DCLRE1C LoF mutations, but no carriers were found among the control subjects (OR = 11.9; p-value = 0.05). The same finding was observed for the DEAD-Box Helicase 19B (DDX19B) gene: three patients harbored a recurrent frameshift deletion mutation, whereas no mutation was observed among the controls (OR = 11.9; p-value = 0.05). The third gene was the X-Ray Repair Cross Complementing 3 (XRCC3) gene, with four LoF mutations detected in four patients and one mutation carrier among the control subjects (OR = 6.8; p-value =

0.067). Detailed information regarding these three genes and their cellular functions is provided in Section 3.5. Despite marginal p-values, the higher frequency of deleterious mutations in these three genes observed among the studied patients compared to healthy controls suggests their potential role in susceptibility to breast cancer, which necessitates further investigation using larger cohorts from other populations.

2.2 Rationale

To follow up with these preliminary findings, we intend to investigate the role of the four novel breast cancer gene candidates in susceptibility to breast cancer among other founder and non-founder populations. We have DNA repositories of familial breast cancer patients in the genetically homogenous population of French-Canadians living in Quebec and the non-founder populations of Ontario and Iran. This will enable us to look for deleterious mutations in our candidate genes among these populations. Therefore, this study aims to apply targeted full-gene sequencing of the four novel breast cancer susceptibility gene candidates identified among the Polish and Bahamian populations on cohorts of familial breast cancer patients and healthy controls from Canada and Iran. Through gene-based association analyses, we will evaluate the association of these candidate genes with breast cancer risk among the populations under study. Additionally, we have access to genetic and phenotypic data from the large UK Biobank study to include in our association analysis. The UK Biobank is a prospective population-based cohort study involving approximately 500,000 individuals aged 40-69 years old at the time of recruitment. Participants were recruited between 2006 and 2010 from 22 assessment centers across England, Scotland, and Wales (Bycroft et al., 2018). Breast cancer is the most prevalent malignancy and the most common incident cancer diagnosed in women within the UK Biobank study. We will access individual-level exome sequence and phenotypic data deposited within the UK Biobank to investigate the role of the four candidate genes under study in susceptibility to breast cancer among the UK Biobank participants.

We also intend to perform loss of heterozygosity (LOH) analysis, investigate homologous recombination deficiency (HRD) status, and analyze tumor mutational signatures to investigate further the role of deleterious mutations in the four novel gene

candidates identified among the Polish and Bahamian populations in susceptibility to breast cancer.

Somatic loss of the wild-type allele is a common mechanism of biallelic inactivation of tumor suppressor genes involved in the development of hereditary cancers such as breast tumors. For example, LOH is observed in roughly 90% and 55% of BRCA1 and BRCA2 mutated breast tumors, respectively, and its occurrence in CHEK2-associated breast tumors varies between 27%–81% (Iyevleva et al., 2022; Maxwell et al., 2017). Conversely, although PALB2 is a tumor suppressor gene and a well-established breast cancer susceptibility gene, there is controversy as to whether PALB2-associated breast tumors harbor biallelic PALB2 inactivation through LOH, with some studies reporting that most germline mutations in PALB2 are not accompanied by LOH, while others have observed biallelic inactivation of PALB2 in more than 60% of PALB2-associated breast tumors (Lee et al., 2018; A. Li et al., 2019). Therefore, the presence of LOH in tumor DNA of germline mutation carriers will give additional proof of the role of the candidate gene's mutations in susceptibility to breast cancer. However, the absence of LOH can not rule out such an association. If a tumor harboring a germline mutation in any of the candidate genes undergoes LOH, it can cause the tumor cells to lack both functional copies of the gene, thereby rendering them deficient in HRR.

Tumor mutational signatures are the imprint of mutational processes that have occurred in the course of cancer evolution (please refer to Chapter 5 for detailed information). The capability to accurately discern these signatures and measure their magnitude in a specific cancer genome is of great importance, as they provide valuable insight into the causes or origins of the tumor and can reveal underlying biological abnormalities that can be targeted for therapy. As an example, BRCA1/2 mutated tumors mostly show Signature 3 and, to a lesser extent, Signature 8 (Nik-Zainal et al., 2016; Van Hoeck et al., 2019). These signatures correlate with defects in HRR pathways and tumors harboring them show favorable responses to PARP inhibitors and Platinum-based chemotherapy (Polak et al., 2017). Identification of tumor mutational signatures correlated to the known function of any of the four candidate genes, repeated among germline mutation carriers of the gene, can provide additional evidence supporting the causal relationship between the candidate gene mutations and breast cancer risk.

As an exploratory aim, we want to study if inherited deleterious mutations in the four breast cancer susceptibility gene candidates may increase the risk of developing ovarian cancer. As discussed earlier in Section 1.2, HBOC syndrome is an inherited genetic condition in which there is a high chance of developing certain types of malignancies, most notably breast and ovarian cancers (Lu et al., 2019). Deleterious mutations in BRCA1/2 account for the majority of HBOC syndrome (Pujol et al., 2021). Over the past decades, several other breast cancer predisposition genes have been linked with the risk of ovarian cancer as well. Therefore, it is important to evaluate the impact of mutations in the four breast cancer gene candidates on the risk of ovarian cancer. Identification of additional HBOC susceptibility genes may ultimately empower us to expand preventive strategies to a broader group of women at high risk of HBOC syndrome. We have DNA samples of unselected cases of invasive epithelial ovarian cancer diagnosed in Ontario, which will enable us to look for deleterious mutations in our candidate genes among these patients and study their impact on the risk of ovarian cancer.

2.3 Hypothesis and Objectives

2.3.1 Hypothesis

Almost half of familial breast cancer cases are negative for pathogenic mutations in the already known breast cancer susceptibility genes, and therefore, we hypothesize that a proportion of these cases is due to rare mutations in moderately to highly penetrant predisposition genes that are not identified yet. In this regard, our research team has identified four novel breast cancer susceptibility gene candidates by performing association studies among the founder populations of Poland and Bahamas. We speculate that deleterious mutations in these candidate genes may predispose to breast cancer and intend to gather more supporting evidence regarding these preliminary findings.

2.3.2 Objectives

The current project is comprised of three research objectives:

1- To evaluate the association of the four novel breast cancer susceptibility gene candidates identified among the Polish and Bahamian populations with breast cancer risk among other populations.
2- To provide more evidence beyond association tests, supporting the causal relationship between deleterious mutations identified in the candidate genes and breast cancer risk.
3- To explore the role of the four novel breast cancer susceptibility gene candidates identified among the Polish and Bahamian populations in susceptibility to ovarian cancer.

2.4 Research Aims

The following research aims will be pursued to achieve our three research objectives:

Aim 1 - Evaluating the role of the four novel candidate genes in susceptibility to breast cancer among the Canadian, Iranian, and British populations (Chapter 3): To validate the findings of association tests performed among the Polish and Bahamian populations, we intend to investigate the role of the four novel candidate genes in susceptibility to breast cancer among other founder and non-founder populations. To achieve this aim, we will perform targeted sequencing of the entire coding regions of the four candidate genes on the germline DNA of 1,021 familial breast cancer patients in Ontario, 1,390 familial French-Canadian breast cancer patients from Quebec, and 421 familial breast cancer patients from Iran. These include all the available DNA repositories of the French-Canadian, Ontario, and Iranian familial breast cancer patients stored at our biobank in Dr. Akbari's Molecular Genetics Laboratory at Women's College Hospital. We will also perform targeted full-gene sequencing of the four candidate genes on the germline DNA of 930 healthy women from Ontario, which will be served as the control group for association analysis with the Ontario breast cancer patients. We will use the sequence data of 800 healthy Iranian individuals reported in the Iranome database as controls for comparison with the Iranian breast cancer patients (Fattahi et al., 2019). Also, we will use the sequence data of 450 healthy individuals from the Genetics of Glucose Regulation in

Gestation and Growth (Gen3G) study as the control group for association analysis with the French-Canadian breast cancer patients (Guillemette et al., 2016). Within the UK Biobank study, we will include all Caucasian women with invasive breast cancer (both incident and prevalent cases) and compare each candidate gene's mutation frequency among cases with that observed among all ethnicity-matched healthy women within the UK Biobank study. This aim will address our first research objective.

Aim 2- Detection of tumor mutational signatures in germline mutation carriers (Chapter 6): While performing association analysis is critical to assess cancer susceptibility, we want to provide more evidence supporting the causal role of the candidate genes mutations in predisposition to breast cancer. To address this objective, we will look for patterns of somatic mutations in breast tumors of germline mutation carriers in the candidate genes under study. This will give us a deeper perception of the biological mechanisms that might have gone awry in the course of cancer development in these patients. Identification of signatures correlated to the known functions of the candidate genes, repeated among germline mutation carriers of the gene, can provide additional evidence in support of the causal relationship between the candidate gene mutations and breast cancer risk. This aim will address our second research objective.

Aim 3- LOH analysis of the germline mutation carriers and HRD score calculation (Chapter 6): Another approach to provide more evidence supporting the causal role of the identified deleterious mutations in the four breast cancer susceptibility gene candidates, is the LOH analysis to investigate if breast tumors harboring germline mutations in any of the candidate genes receive a second hit in their remaining functional copy of the gene. We will also investigate HRD status in tumor cells of the germline mutation carriers. The presence of HRD, especially in tumors with LOH and in the absence of mutations in other HRR-related genes, can further support the causal role of the candidate gene's mutations in susceptibility to breast cancer. This aim will address our second research objective.

Aim 4 - Study the association of the four novel candidate genes in susceptibility to ovarian cancer among Ontario population (Chapter 3): To study whether the four candidate genes under study contribute to the risk of ovarian cancer, we will sequence them on the germline DNA of 1,488 unselected ovarian cancer patients from Ontario. We will use the same control group as the Ontario familial breast cancer patients described in aim 1 for

the association analysis with ovarian cancer patients. This aim will address our third research objective.

Chapter 3 – Association Studies

3.1 Study Subjects

3.1.1 Ontario Study Subjects

The Ontario breast cancer patients included 1,021 women diagnosed with invasive breast cancer between 2013 and 2018 at four academic centers in Toronto, Canada (Women's College Hospital, Sunnybrook Health Sciences Centre, St. Michael's Hospital, and University Health Network). All cases had been previously tested negative for mutations in BRCA1/2 and PALB2 and had at least one of the following criteria: Jewish ethnicity, triple-negative breast cancer, age of 50 years or younger at the time of diagnosis, synchronous bilateral breast cancer, or a positive family history defined as having at least one first or second degree relative with breast cancer diagnosed at the age of 50 years old or younger, ovarian cancer, or male breast cancer.

The Ontario ovarian cancer patients comprised 1,488 women with epithelial ovarian tumors who were unselected for family history of the disease and were identified through the Ontario Cancer Registry (OCR) in Ontario province, Canada. These patients were diagnosed between 1995 and 2004 and were 20-79 years old at the time of diagnosis. All had been previously tested negative for mutations in 12 ovarian cancer predisposition genes, including ATM, BRCA1, BRCA2, BRIP1, CHEK2, MSH2, MSH6, MLH1, PALB2, RAD51C, RAD51D, and TP53.

The control group for comparison with the Ontario breast/ovarian cancer patients included 930 cancer-free women who had been referred to Women's College Hospital (Toronto, Canada) for routine checkup between 1999 and 2001 and donated their DNA for cancer research. These individuals had no personal history of cancer and were 40 years or older at the time of recruitment.

3.1.2 French-Canadian Study Subjects

The French-Canadian familial breast cancer patients included 1,390 women with invasive breast cancer treated and monitored at a breast cancer clinic affiliated with the Research Center of the University of Montreal (Montreal, Quebec). Women aged more than 50 years were included if they had at least two first or second-degree relatives with breast or ovarian cancer. All women aged 50 years or younger were eligible, regardless of their family history. All patients were negative for seven French-Canadian founder mutations in BRCA1, BRCA2, and PALB2 (three in BRCA1, three in BRCA2, and one in PALB2).

As the control group for comparison with the French-Canadian patients, we used the sequence data of 450 cancer-free women from Gen3G, which is a prospective cohort study of 1,034 pregnant women aged ≥ 18 years old and 854 newborns of these women living in Sherbrooke (Québec, Canada). Women enrolled in Gen3G represent the general population of women in reproductive age from the Eastern Townships region in Québec. These individuals were recruited between 2010 and 2013 to study the effects of maternal dysglycaemia during pregnancy on offspring adiposity and metabolic profile (Guillemette et al., 2016).

3.1.3 Iranian Study Subjects

Iranian familial breast cancer patients comprised 421 women referred to a national cancer institute between 2012 and 2016 due to their breast cancer diagnosis. All patients were negative for BRCA1/2 and PALB2 mutations and had at least one of the hereditary breast cancer risks criteria: age of 40 years or younger at the time of diagnosis, positive family history of breast or ovarian cancers in their first or second-degree relatives, triple-negative breast cancer diagnosed below 60 years of age, bilateral breast cancer diagnosed at ≤ 50 years of age, or multiple primary tumors in the same individual.

We used the sequence data of 800 healthy individuals reported in the Iranome database as the control group for comparison with the Iranian breast cancer patients. This database is created after performing WES on the germline DNA of 800 healthy individuals

from eight major ethnic groups in Iran (100 individuals from each ethnic group). These ethnic groups represent over 80 million Iranians and include Arabs, Azeris, Baluchs, Kurds, Lurs, Persians, Persian Gulf Islanders, and Turkmen (Fattahi et al., 2019).

3.1.4 UK Biobank Study Subjects

We used the UK Biobank database to explore the association of the four candidate genes under study with breast cancer risk among the UK population. Detailed information regarding the UK Biobank cohort study has been provided in Section 2.2. Most participants within this cohort have self-reported their ethnic background as 'White British' (88.26%). Principal component analysis (PCA) conducted by the UK Biobank researchers has shown population structure within this most common ethnic background category. Therefore, the researchers used a combination of self-reported ethnicity and PCA results to identify a subset of individuals, referred to as 'Caucasians', who have self-reported as British and shared similar ancestral backgrounds based on the results of PCA (84%). We accessed the UK Biobank database on May 30, 2022. To reduce the risk of confounding due to differences in ancestral background, we confined our association studies within the UK Biobank database to Caucasian individuals. We defined our case group as all Caucasian women with invasive breast cancer (both incident and prevalent cases) and no pathogenic or likely pathogenic mutations in any of the highly penetrant breast cancer susceptibility genes BRCA1, BRCA2, and PALB2. Breast cancer cases were defined based on the International Classification of Diseases (ICD10 and ICD9) and self-reported status of cancer. In addition, all ethnicity-matched women with no personal history of cancer and no pathogenic or likely pathogenic mutations in BRCA1/2 and PALB2 were considered as the control group for association analysis. Our association study within the UKB database was not designed as a nested case-control or case-cohort study. Instead, given the large number of prevalent breast cancer patients in the UKB database, we recruited both prevalent and incident cases from the dataset. Unlike a nested case-control study, which selects controls at the time each case occurs, our study used all eligible cancer-free Caucasian women as the control group. This approach to recruit study subjects for a case-control study has been commonly employed by other researchers (Hassanin et al., 2022; Wilcox et al., 2023). After randomly excluding one individual from

each pair of related individuals, there were 15,643 breast cancer cases and 157,943 cancer-free controls for association analyses. Detailed information on how we dealt with kinship among the study subjects will be provided in Section 3.3.

3.2 Methods

3.2.1 Sequencing Laboratory Analysis

We applied NGS technology for efficient screening of the four novel breast cancer susceptibility gene candidates among our case and control subjects. We targeted the coding exons of the four candidate genes plus 20 bp of the introns at each end of the exons. The Agilent SureSelect XT HS2 Library Preparation Kit was used for library preparation and to capture target sequence regions. Beginning with 150 ng of gDNA for each sample, the genomic regions were enzymatically fragmented, end-repaired, A-tailed at their 3' ends, and subsequently ligated at their 5' ends with platform-specific adaptors. Unique molecular barcode sequences were incorporated into both ends of each DNA fragment at the ligation step. To ensure complete removal of free adapters, the samples were purified using AMPure XP beads. The constructed libraries were PCR-amplified for 12 cycles using unique dual index primer pairs and purified with AMPure XP beads one more time. Next, every 16 indexed samples were pooled together, and biotinylated RNA strand probes were used to hybridize to the target sequence regions. Using streptavidin-coated magnetic beads, the hybridized DNAs with biotinylated probes were captured and separated from the rest of the genomic DNA. Post-capture PCR amplification was performed for 15 cycles and the amplified captured libraries were subsequently purified using AMPure XP beads. Lastly, DNA quality check was conducted using Agilent's D1000 ScreenTapes on the TapeStation 2200 system (Agilent Technologies, Santa Clara, CA), and DNA was quantified using Qubit dsDNA High Sensitivity Assay Kits with the Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA). Twenty-four pooled libraries (each containing 16 samples) were combined together and sequenced on a mid-throughput cartridge of an Illumina NextSeq sequencer for paired-end sequencing of 300 cycles (2×150).

3.2.2 Variant Calling

Sequence reads were aligned to the human genome's reference sequence (hg19) using Burrows–Wheeler Aligner (BWA) incorporated into the Sentieon package (Sentieon Inc. San Jose, CA, United States). The same package was used to convert the sequence alignment map (SAM) files to BAM format and to sort/index the BAM files. Next, all the unmapped reads, reads aligned to more than one region in the human genome, and all duplicate reads were filtered out from the BAM file. Exact duplicate reads were collapsed to avoid inflated coverage. The HaplotypeCaller module of The Genome Analysis Toolkit (GATK) incorporated into the Sentieon package was used for calling both SNVs and indels. The low-quality called variants were removed using a range of metrics reported by the Sentieon package during variant calling. Regions with at least 20-fold depth of coverage were used for calling variants, and a different nucleotide from the reference sequence observed in at least 25% of the reads aligned to a given position was considered as a variant. For the UK Biobank patients, we used variants reported in VCF files available for study participants. To exclude potential sequencing artifacts within homopolymer regions, stretches of identical nucleotides repeated consecutively and prone to sequencing errors, we applied additional filtering strategies beyond allele fraction. These included examining VCF files for evidence of strand bias or co-occurring mutations within the homopolymer region. Variants exhibiting strand bias or accompanied by other mutations in the homopolymer region were excluded from further analysis. Finally, for individuals with germline DNA samples available in our biobank, rare germline mutations were confirmed using Sanger sequencing. The SNP & Variation Suite (SVS) (GoldenHelix Inc., Bozeman, MT, United States) was applied for annotating called variants based on RefSeq Genes 105v2 and NCBI data source. Annotation was used for determining the related genes and the effect of each variant on the coding proteins.

3.3 Data Analysis

The focus of our study was on LoF variants. These variants include frame-shift insertions and deletions (indels), stop codon gain, stop codon loss, start codon loss, and essential splicing site mutations, all of which could potentially change the encoded protein and disrupt its function. We excluded all variants with MAF greater than 1% reported in publicly available databases, assuming these to be benign or of low penetrance, and kept

our focus only on rare variants (MAF ≤ 1%). For this purpose, we used the dbSNP 154 database (http://www.ncbi.nlm.nih.gov/projects/SNP/), 1000 Genomes Project (http://www.1000genomes.org/), and Genome Aggregation Database (gnomAD) (https://gnomad.broadinstitute.org/). Finally, all the identified LoF variants were assessed using the American College of Medical Genetics and Genomics (ACMG) guidelines for variant classification, and only those classified as pathogenic or likely pathogenic were used for the association analyses.

A close relationship among the study subjects could potentially bias the association analysis. Therefore, it is important to determine individuals who have a kinship with each other and keep only one of them in the study. However, since the amount of genetic data produced from targeted sequencing of the four candidate genes under study is limited and does not provide sufficient power, we were not able to examine cryptic relatedness among the Canadian and Iranian study subjects (please refer to the "Limitations" of the study in Section 7.2). Within the UK Biobank study, related individuals were identified by calculating kinship coefficients for all pairs of study subjects using the KING software (Manichaikul et al., 2010). Kinship coefficient is the probability that two alleles selected at random and one from each individual are identical by descent. To control for cryptic relatedness among the UK Biobank study subjects, we randomly excluded one from each pair of related individuals if the genetic relationship was closer than third-degree relatives, defined as kinship coefficient > 0.044 calculated by the UK Biobank study.

We estimated ORs for developing breast/ovarian cancer to investigate the association of the four novel candidate genes with breast/ovarian cancer risk. To do so, we performed rare-variant association analyses (burden test) to compare the cumulative effect of total LoF mutations in each candidate gene among breast/ovarian cancer patients (cases) to that observed among healthy individuals (controls). We performed logistic regression analysis, and ORs were adjusted for age and self-reported ethnicity among the Canadian and Iranian study subjects. For the UK Biobank participants, ORs were adjusted only for age, considering that cases and controls had been already matched for genetically inferred ethnicity. We addressed multiple hypothesis testing by

using the Bonferroni method. For this, we divided the acceptable type I error rate (0.05) by the number of genes being tested. Therefore, a p-value less than 0.0125 was considered statistically significant. All analyses were done by SVS (Golden Helix Inc., Bozeman, MT, USA).

3.4 Results

3.4.1 Association Studies within the Ontario Population

Screening of the entire coding exons of the four candidate genes under study did not reveal any mutations of interest among the 1,021 familial breast cancer patients and 930 controls from Ontario. We did, however, find four carriers of ATRIP LoF variants among the 1,488 unselected ovarian cancer patients from Ontario, with one patient harboring the ATRIP c.1152_1155delTGGA (rs753154444) variant previously identified by our research team among the Polish population (please refer to Section 2.1.1 for detailed information). As there were no ATRIP mutation carriers among the Ontario control subjects, we added 0.5 to all cells of the 2×2 contingency table and calculated the odds ratio for developing ovarian cancer over these adjusted cell counts (OR = 5.6, p = 0.14). This strategy is known as Haldane-Anscombe correction (Haldane, 1956). Of note, ATRIP LoF variants have a frequency of about 0.061% among the 22,928 cancer-free non-Finnish European females of the ExAC database (1 in every 1,638 women) (Karczewski et al., 2017). Detailed information regarding the ATRIP LoF variants identified among the Ontario ovarian cancer patients is shown in Table 1.

Table 1: ATRIP LoF variants identified among 1,488 women with ovarian cancer from Ontario

DNA Variant [a]	Protein Effect	Exon	Number of Cases with ATRIP Mutation	Sequence Ontology (Clinically Relevant)
c.1152_1155delTGGA	p.Gly385Terfs	8	1	Frameshift variant
c.1956G>A	p.Trp652Ter	10	2	Stop-codon variant
c.2041C>T	p.Gln681Ter	11	1	Stop-codon variant

[a] Nucleotide positions are based on the GenBank: NM_130384.3 transcript of ATRIP.

3.4.2 Association Studies within the French-Canadian Population

Of the four candidate genes under study, we identified an ATRIP frameshift insertion (rs1454682560) in only one individual out of the 1,390 French-Canadian familial breast cancer cases. This mutation is a base pair duplication at position 290 in the coding sequence of ATRIP (NM_130384.3:c.290dupA) (Figure 2). This variant is not reported in ClinVar and is classified as a pathogenic variant based on the ACMG guidelines.

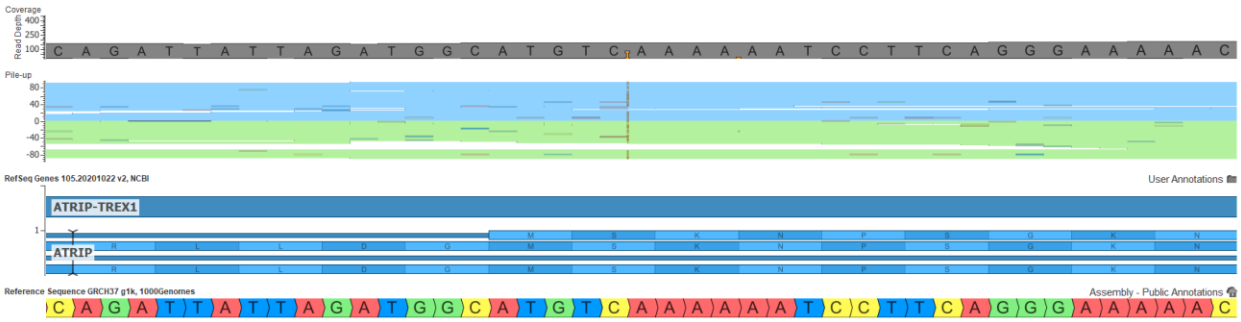


Figure 2: Next generation sequencing reads pileup of ATRIP c.290dupA mutation identified in a French-Canadian familial breast cancer patient along with RefSeq V105 and GRCH37 human genome reference regions were shown.

More information regarding this ATRIP mutation is shown in Table 2. We did not detect any LoF mutations in the four candidate genes among the 450 cancer-free women from Gen3G study that served as control group for the French-Canadian familial breast cancer cases in our study.

Table 2: ATRIP LoF variant identified among 1,390 French-Canadian women with familial breast cancer

DNA Variant [a]	Protein Effect	Exon	Number of Cases with ATRIP Mutation	Sequence Ontology (Clinically Relevant)
c.290dupA	p.Asn97LysfsTer13	2	1	Frameshift variant

[a] Nucleotide positions are based on the GenBank: NM_130384.3 transcript of ATRIP.

3.4.3 Association Studies within the Iranian Population

Of the four candidate genes under study, we identified a DCLRE1C frameshift insertion (rs760288938) among 3/421 Iranian familial breast cancer cases and 2/800 controls from the Iranome database. This mutation is on exon 14, the last coding exon of DCLRE1C, and is classified as a variant of unknown significance (VUS). Association analysis did not show a statistically significant association between carrying this mutation and developing breast cancer among the Iranian study subjects (OR = 2.86, 95% CI = 0.47-17.21, p-value = 0.34). More information regarding this mutation is depicted in Table 3. We did not detect any LoF mutations in the other three candidate genes among the Iranian study subjects.

Table 3: DCLRE1C LoF variant identified among 421 Iranian familial breast cancer cases and 800 controls from the Iranome database

DNA Variant [a]	Protein Effect	Exon	Number of Cases with DCLRE1C Mutation	Number of Controls with DCLRE1C Mutation	Sequence Ontology (Clinically Relevant)
c.1903dupA	p.Ser635Lysfs*6	14	3	2	Frameshift variant

[a] Nucleotide positions are based on the GenBank: NM_001033855.3 transcript of DCLRE1C.

3.4.4 Association Studies within the UK Population

Accounting for the ancestral background is essential in genetic association studies. To reduce the risk of confounding due to differences in ancestral background, we confined our association studies within the UK Biobank database to only a set of individuals with a relatively homogeneous ancestry. Most participants in the UK Biobank self-reported their ethnic background as 'White British' (88.26%). Principal component analysis (PCA) performed by the UK Biobank researchers has revealed population structure within this most common ethnic background category. Therefore, the researchers used a combination of self-reported ethnicity and PCA results to identify a subset of individuals, referred to as 'Caucasians', who self-reported as British and shared similar ancestral

backgrounds based on the results of PCA (84%). In Figure 3, we plotted all the participants within the UK Biobank cohort (n = 502,396) according to their principal component (PC) scores (eigenvalues) in each of the top three genetic principal components. The UK Biobank case and control subjects recruited in our study are shown in green and blue dots, respectively, and the remaining UK Biobank participants are in orange dots. As depicted, our case and control subjects form a common overlapping cluster in each plot. This observation confirms that our study subjects share similar ancestral background and there is no bias or confounding due to population structure in our analysis.

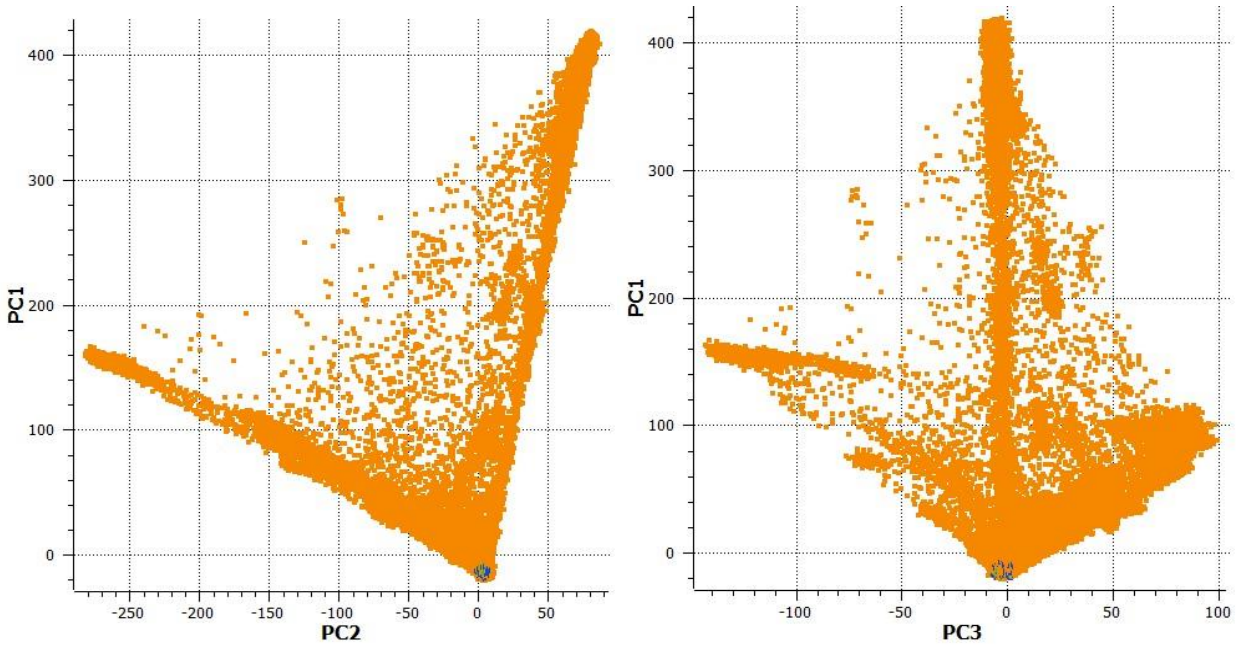


Figure 3: Genetic principal components within the entire UK Biobank database. The plots display the first principal component (PC1) against the second (PC2) and the third (PC3) principal components, respectively. In each plot, blue points represent the Caucasian healthy females who served as controls (n = 157,943). Green points indicate the Caucasian females affected with breast cancer (n = 15,643). Orange points represent the remaining individuals within the UK Biobank database (n = 328,810).

To have a better visualization of the study subjects, we created the same plots shown in Figure 3 for merely the UK Biobank case and control subjects. As depicted in Figure 4, our cases (green dots) and controls (blue dots) were uniformly distributed within a shared cluster in each plot. This observation reconfirms that there is no confounding due to population structure in our association analysis, as our UK Biobank study subjects have similar ancestral background.

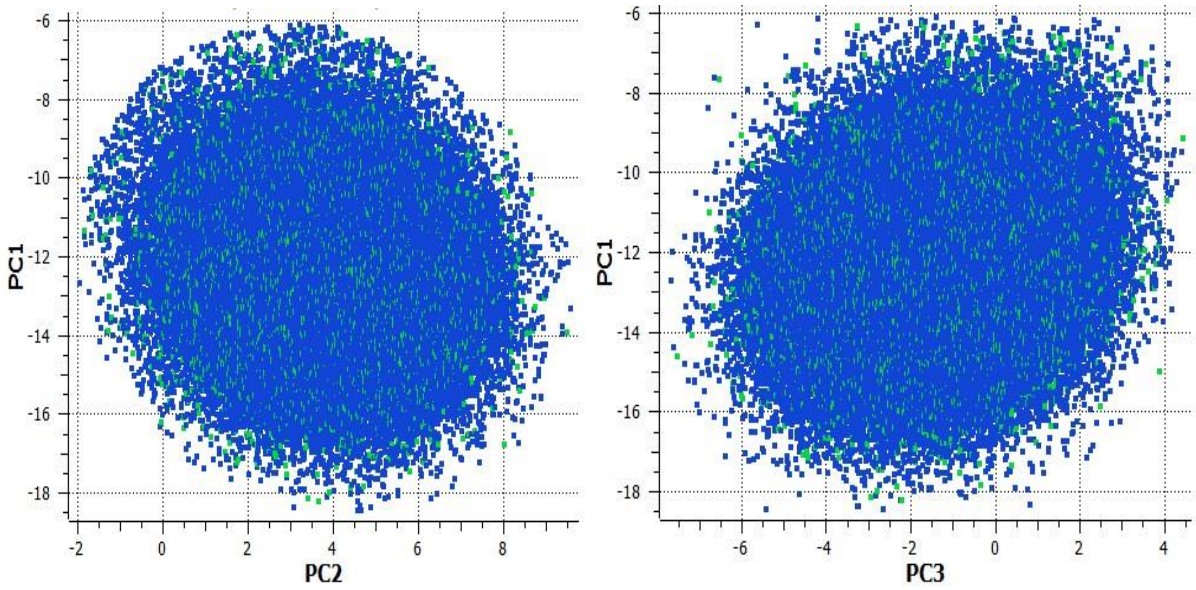


Figure 4: Genetic principal components within the UK Biobank case and control subjects recruited in our study. The plots display the first principal component (PC1) against the second (PC2) and the third (PC3) principal components, respectively. In each plot, blue points represent the Caucasian healthy females who served as controls (n = 157,943). Green points indicate the Caucasian females affected with breast cancer (n = 15,643).

Within the UK Biobank database, 18 individuals out of 15,643 breast cancer-affected Caucasian women harbored germline LoF variants in DCLRE1C (1 in every 869 Caucasian women with breast cancer). We compared that to the frequency of DCLRE1C's LoF variants among the 157,943 Caucasian females with no personal history of cancer and no pathogenic or likely pathogenic mutations in BRCA1/2 and PALB2.

There were 120 women with germline DCLRE1C LoF variants among the 157,943 cancer-free individuals that served as the control group (1 in every 1,316 healthy individuals). We concluded that LoF variants in DCLRE1C were not associated with a higher risk of breast cancer among the UK Biobank study subjects (OR = 1.54, 95% CI = 0.93–2.53, p-value = 0.1). Detailed information about the DCLRE1C LoF variants identified among the UK Biobank study subjects is presented in Table 4.

Table 4: DCLRE1C LoF variants identified among 15,643 women with breast cancer and 157,943 controls from the UK Biobank database

DNA Variant [a]	Protein Effect	Exon	Number of Cases with DCLRE1C Mutation	Number of Controls with DCLRE1C Mutation	Sequence Ontology (Clinically Relevant)
c.-4_3delCGCTATG	p.Met1fs	1	1	1	Start-codon variant
c.22dupA	p.Met8fs	1	0	1	Frameshift variant
c.109+1G>T	—	1	0	1	Splice-donor variant
c.110-1G>C	—	2	0	1	Splice-acceptor variant
c.184_191delTCACCTGT	p.Ser62fs	3	0	1	Frameshift variant
c.241C>T	p.Arg81Ter	3	3	9	Stop-codon variant
c.240delA	p.Lys80fs	3	0	1	Frameshift variant
c.247-2A>G	—	4	3	23	Splice-acceptor variant
c.362+2T>A	—	5	0	6	Splice-donor variant
c.465-2A>C	—	7	0	7	Splice-acceptor variant
c.571C>T	p.Arg191Ter	8	1	2	Stop-codon variant
c.671delG	p.Gly224fs	8	0	1	Frameshift variant
c.678+1G>C	—	9	0	1	Splice-donor variant
c.679-1G>C	—	9	0	1	Splice-acceptor variant
c.787G>T	p.Glu263Ter	10	0	1	Stop-codon variant
c.940_941delGC	p.Ala314fs	11	0	1	Frameshift variant
c.1062-1G>T	—	13	0	1	Splice-acceptor variant
c.1147C>T	p.Arg383Ter	13	0	3	Stop-codon variant
c.1156+1G>C	—	14	0	1	Splice-donor variant
c.1378G>T	p.Glu460Ter	14	0	1	Stop-codon variant
c.1424_1425dupCT	p.Val476fs	14	0	1	Frameshift variant
c.1506_1513delGAGTTTGG	p.Ser503fs	14	1	0	Frameshift variant
c.1539delA	p.Gly515fs	14	0	2	Frameshift variant
c.1582G>T	p.Glu528Ter	14	1	0	Stop-codon variant

c.1640delG	p.Gly547fs	14	0	1	Frameshift variant
c.1652G>A	p.Trp551Ter	14	0	1	Stop-codon variant
c.1653G>A	p.Trp551Ter	14	0	1	Stop-codon variant
c.1749dupA	p.Glu584fs	14	1	3	Frameshift variant
c.1774delG	p.Glu592fs	14	0	1	Frameshift variant
c.1777C>T	p.Gln593Ter	14	0	2	Stop-codon variant
c.1812_1830del	p.Asp604fs	14	0	1	Frameshift variant
c.1903dupA	p.Ser635fs	14	1	15	Frameshift variant
c.1990C>T	p.Arg664Ter	14	5	17	Stop-codon variant
c.2012_2013delAT	p.Tyr671fs	14	1	0	Frameshift variant
c.2035_2036delAG	p.Ser679fs	14	0	5	Frameshift variant
c.2049dupA	p.Arg685fs	14	0	5	Frameshift variant
c.2054delG	p.Arg685fs	14	0	1	Frameshift variant

[a] Nucleotide positions are based on the GenBank: NM_001033855.3 transcript of DCLRE1C.

Next, we studied the association of the XRCC3 LoF variants with breast cancer risk. There were 10 individuals with germline LoF variants in this candidate gene among the 15,643 Caucasian breast cancer patients within the UK Biobank database (1 in every 1,564 Caucasian women with breast cancer). We observed germline XRCC3 LoF variants in 70 out of the 157,943 Caucasian cancer-free females that served as the control group (1 in every 2,256 healthy individuals), concluding that the frequencies of LoF variants in XRCC3 were not statistically different between the Caucasian breast cancer patients (cases) and cancer-free individuals (controls) within the UK Biobank database (OR = 1.55, 95% CI = 0.79–3.01, p-value = 0.2). Further information regarding the XRCC3 LoF variants identified among the UK Biobank study subjects is provided in Table 5.

Table 5: XRCC3 LoF variants identified among 15,643 women with breast cancer and 157,943 controls from the UK Biobank database

DNA Variant [a]	Protein Effect	Exon	Number of Cases with XRCC3 Mutation	Number of Controls with XRCC3 Mutation	Sequence Ontology (Clinically Relevant)
c.3G>T	p.Met1Ile	4	1	9	Start-codon variant
c.56-2A>G	—	5	0	1	Splice-acceptor variant
c.68delC	p.Ser23fs	5	0	3	Frameshift variant
c.83delT	p.Leu28fs	5	2	6	Frameshift variant

c.106_109delAAGA	p.Lys36fs	5	0	1	Frameshift variant
c.113_114insA	p.Thr39fs	5	0	1	Frameshift variant
c.141G>A	p.Trp47Ter	5	0	1	Stop-codon variant
c.155delC	p.Thr52fs	5	2	6	Frameshift variant
c.160_161insGAGG	p.Ser54fs	5	2	6	Frameshift variant
c.228delC	p.Thr77fs	6	0	1	Frameshift variant
c.302dupG	p.Ile102fs	6	0	1	Frameshift variant
c.457C>T	p.Gln153Ter	7	0	1	Stop-codon variant
c.479_491del	p.Arg160fs	7	0	1	Frameshift variant
c.527delG	p.Gly176fs	7	0	2	Frameshift variant
c.558delT	p.Asp186fs	7	0	1	Frameshift variant
c.561+1_561+3delGTG	—	8	1	4	Splice-donor variant
c.676delC	p.Gln226fs	8	0	2	Frameshift variant
c.728delG	p.Arg243fs	8	1	0	Frameshift variant
c.772C>T	p.Gln258Ter	8	0	4	Stop-codon variant
c.775-2A>G	—	9	0	2	Splice-acceptor variant
c.802_806delGCAGC	p.Ala268fs	9	0	1	Frameshift variant
c.821+1dupG	—	10	0	1	Splice-donor variant
c.864G>A	p.Trp288Ter	10	0	2	Stop-codon variant
c.871delC	p.Gln291fs	10	0	1	Frameshift variant
c.919_920dupGC	p.Leu308fs	10	0	1	Frameshift variant
c.923_932delTCGGCTGCCC	p.Leu308fs	10	0	2	Frameshift variant
c.964dupC	p.His322fs	10	0	2	Frameshift variant
c.990C>A	p.Tyr330Ter	10	0	1	Stop-codon variant
c.990_991dupCA	p.Ile332fs	10	0	1	Frameshift variant
c.1010dupT	p.Arg338fs	10	1	5	Frameshift variant

[a] Nucleotide positions are based on the GenBank: NM_001100119.2 transcript of XRCC3.

For our next candidate gene, DDX19B, no individuals carried germline LoF variants among the 15,643 breast cancer-affected Caucasian women within the UK Biobank database, and we only identified germline DDX19B LoF variants in 16 out of the 157,943 Caucasian cancer-free females (1 in every 9,291 healthy individuals) (OR = 0.3, p = 0.39). The DDX19B variants identified among the UK Biobank control subjects are presented in Table 6.

Table 6: DDX19B LoF variants identified among the 157,943 controls from the UK Biobank database

DNA Variant [a]	Protein Effect	Exon	Number of Controls with DDX19B Mutation	Sequence Ontology (Clinically Relevant)
c.25delC	p.Gln9fs	1	1	Frameshift variant
c.260delC	p.Pro87fs	4	3	Frameshift variant
c.389+1G>T	—	6	2	Splice-donor variant
c.490dupC	p.Gln164fs	6	2	Frameshift variant
c.592dupG	p.Val198fs	7	1	Frameshift variant
c.1089dupG	p.Ser364fs	10	4	Frameshift variant
c.1138C>G	p.Arg380Ter	10	1	Stop-codon variant
c.1160dupT	p.Leu387fs	10	2	Frameshift variant

[a] Nucleotide positions are based on the GenBank: NM_007242.7 transcript of DDX19B.

Lastly, for ATRIP, there were 13 individuals harboring germline LoF variants among the 15,643 breast cancer-affected Caucasian women within the UK Biobank database (1 in every 1,203 women with breast cancer). We compared that with the ATRIP's LoF variant frequency among the 157,943 Caucasian cancer-free females with no pathogenic or likely pathogenic mutations in the highly penetrant BRCA1/2 and PALB2 genes. There were 40 women with germline ATRIP LoF variants among the 157,943 individuals that served as the control group (1 in every 3,948 healthy individuals). We concluded that LoF variants in ATRIP were associated with a significantly increased risk of developing breast cancer among the UK Biobank participants (OR = 3.28, 95% CI = 1.76–6.14, p-value < 0.001). Detailed information regarding the ATRIP LoF variants identified among the UK Biobank study subjects is provided in Table 7. Of note, carrying ATRIP mutation was not associated with the age of breast cancer onset as the mean age at the time of breast cancer diagnosis among the ATRIP mutation carriers did not differ significantly with that of non-carriers (57.89 vs 58.6, p-value = 0.49).

Table 7: ATRIP LoF variants identified among 15,643 women with breast cancer and 157,943 controls from the UK Biobank database

DNA Variant [a]	Protein Effect	Exon	Number of Cases with ATRIP Mutation	Number of Controls with ATRIP Mutation	Sequence Ontology (Clinically Relevant)

c.16_20delinsATGCAAG	p.Ala6Metfs*67	1	2	2	Frameshift variant
c.51_63del	p.Pro18Argfs*50	1	0	2	Frameshift variant
c.217C>T	p.Gln73Ter	1	2	4	Stop-codon variant
c.243_244delGT	p.Ser82Glnfs*2	1	1	3	Frameshift variant
c.247+2T>C	—	1	1	0	Splice-donor variant
c.253delC	p.His85Ilefs*7	2	0	1	Frameshift variant
c.269T>G	p.Leu90Ter	2	0	1	Stop-codon variant
c.290dupA	p.Asn97Lysfs*13	2	3	4	Frameshift variant
c.337G>T	p.Glu113Ter	2	1	3	Stop-codon variant
c.376_379delGAAA	p.Glu126Argfs*2	2	0	2	Frameshift variant
c.436C>T	p.Arg146Ter	3	0	2	Stop-codon variant
c.514dupC	p.Gln172Profs*5	3	0	2	Frameshift variant
c.672-1G>C	—	5	1	0	Splice-acceptor variant
c.721C>T	p.Gln241Ter	5	0	2	Stop-codon variant
c.829+1G>C	—	5	0	1	Splice-donor variant
c.830-2A>G	—	6	1	0	Splice-acceptor variant
c.1103_1112del	p.Gly368Alafs*5	8	0	1	Frameshift variant
c.1152_1155delTGGA	p.Gly385Ter	8	0	1	Frameshift variant
c.1270C>T	p.Gln424Ter	8	0	1	Stop-codon variant
c.1703delT	p.Leu568Trpfs*2	8	0	1	Frameshift variant
c.1710delA	p.Lys570Asnfs*2	8	0	1	Frameshift variant
c.2055+1G>A	—	11	0	2	Splice-donor variant
c.2073_2074delGG	p.Val692Aspfs*57	12	0	1	Frameshift variant
c.2123dupC	p.Arg709Lysfs*41	12	0	1	Frameshift variant
c.2192C>A	p.Ser731Ter	12	0	1	Stop-codon variant
c.2263delG	p.Val755Serfs*11	12	0	1	Frameshift variant
c.2278C>T	p.Arg760Ter	12	1	0	Stop-codon variant

[a] Nucleotide positions are based on the GenBank: NM_130384.3 transcript of ATRIP.

3.5 Discussion

Our team has recently introduced ATRIP as a novel breast cancer susceptibility gene candidate. Briefly, at the discovery phase, WES was performed on the germline DNA of 510 highly familial breast cancer patients and 300 controls from the founder population of Poland to find high-risk candidate genes. On top of the list, there was the ATRIP gene with three breast cancer patients harboring LoF variants in this gene and no

carriers among the control subjects (OR = 4.23, p = 0.29). One affected woman had the ATRIP c.69_75dup (p.Thr26Alafs*23) frameshift insertion and two others had the ATRIP c.1152_1155del (p.Gly385Ter) frameshift deletion. At the validation, the ATRIP c.1152_1155del mutation was observed in 42 of 16,085 unselected Polish breast cancer patients and in 11 of 9,285 controls (OR = 2.14, 95% CI = 1.13– 4.28, = 0.02) (please refer to Section 2.1.1 for more information) (Cybulski et al., 2023). In this study, we extended our investigation to evaluate the role of ATRIP in susceptibility to breast cancer in other populations. Among different cohorts under study, we identified ATRIP LoF variants in 13/15,643 breast cancer-affected Caucasian women and in 40/157,943 cancer-free Caucasian females within the UK Biobank cohort (Table 7), concluding that ATRIP LoF variants were associated with a higher risk of breast cancer among the British population (OR = 3.28, 95% CI = 1.76–6.14, p < 0.001) (Cybulski et al., 2023). However, due to the rarity of LoF variants in ATRIP (1 in every 1,000–3,000 of the general population) and the small sample sizes, we did not detect ATRIP mutations at a significant level to warrant association analyses among the Ontario, French-Canadian, and Iranian studied cohorts. We also found four carriers of ATRIP LoF variants among the 1,488 unselected ovarian cancer patients from Ontario, with one patient harboring the ATRIP c.1152_1155delTGGA (rs753154444) variant that our research team formerly identified among the Polish population (please refer to Section 2.1.1 for detailed information) (OR = 5.6, p = 0.14).

ATRIP is on chromosome 3p21 and encodes an essential component of the DNA damage checkpoint. In instances of DNA damage or stalled replication forks, single-stranded DNA (ssDNA) is coated with high-affinity replication protein A (RPA), suppressing further resections via negative feedback (Krasikova et al., 2016; T. Liu & Huang, 2016). The RPA-coated ssDNA recruits ATRIP (ATR-interacting protein) to the site of ssDNA damage, followed by localization of ATR (ataxia telangiectasia and Rad3-related) to the RPA-coated ssDNA via forming a complex with ATRIP. Proper positioning of this complex on damaged ssDNA triggers the ATR-mediated CHEK1 activation, initiating the DNA damage checkpoint, which arrests the cell cycle and coordinates a response to the replicative stress (Ball et al., 2005; Cortez et al., 2001). Pathogenic variants in ATRIP and ATR predispose to Seckel syndrome, a rare genetic condition

characterized by growth defects, neurodevelopmental malformations, and short life expectancy (O'Driscoll et al., 2003; Ogi et al., 2012). Several cancers have also been reported among carriers of ATR pathogenic variants, including oropharyngeal, nonmelanoma skin, cervical, and breast cancers (Tanaka et al., 2012).

Depleting ATRIP in human cell lines using small interfering RNA (siRNA) results in ATR destabilization and loss of ATR-mediated checkpoint responses to DNA damage. This observation suggests that the interaction between ATRIP and RPA-ssDNA has a critical role in recruiting ATR to sites of DNA damage and initiating DNA damage signals (Cortez et al., 2001; X. H. Yang & Zou, 2006). Multiple sites on ATRIP have been shown to interact with RPA-ssDNA; ATRIP mutant cell lines lacking the N-terminal 107 amino acids fail to bind to RPA-ssDNA (Ball et al., 2005; Krasikova et al., 2016). In vitro translation of different ATRIP fragments and testing their binding capability to ssDNA have shown the presence of two RPA-ssDNA interacting domains in ATRIP outside of Its N terminus, ATRIP 108–390 and ATRIP 390–791 (Namiki & Zou, 2006). Also, studying Random fragment libraries of ATRIP has shown that ATR interacting domains exist at the C terminus of ATRIP (Ball et al., 2005; Falck et al., 2005). The removal of ATRIP exon 11 in mammalian cells that codes for amino acids 658–684 severely compromised ATRIP binding to ATR (Ball et al., 2005; Cortez et al., 2001). Thus, amino acids 658–684 of ATRIP truncated in the ATRIP c.1152_1155del variant seem essential for efficient binding to ATR.

Interestingly, after publishing our results and introducing ATIRP as a novel breast cancer susceptibility gene candidate (Cybulski et al., 2023), a meta-analysis across two other large WES datasets replicated the positive association of ATRIP deleterious mutations with the risk of developing breast cancer (Wilcox et al., 2023). More specifically, the datasets in this meta-analysis comprised the Breast Cancer Risk after Diagnostic Gene Sequencing (BRIDGES) dataset, including samples from eight studies in the Breast Cancer Association Consortium (BCAC) and the Personalized Risk Assessment for Prevention and Early Detection of Breast Cancer: Integration and Implementation (PERSPECTIVE) dataset, including three BCAC studies. Protein truncating variants in ATRIP were observed among 13 of 8,410 breast cancer cases and 3 of 8,238 controls in these two datasets (OR = 3.69, 95% CI = 1.43–9.52, p = 0.007) (Wilcox et al., 2023). In

total, although ATRIP LoF variants seem to be relatively rare, the significantly increased breast cancer risk associated with these variants suggests that ATRIP can be regarded as a moderately penetrant gene in the context of hereditary breast cancer.

Our next candidate gene was DCLRE1C. This gene is on chromosome 10p13 and encodes for the nuclear protein Artemis, which has 5′-3′ exonuclease activity and participates in V(D)J recombination and NHEJ repair of DSBs caused by radiation and oxidative stress (Riballo et al., 2004; Woodbine et al., 2011). Germline alterations in DCLRE1C have been reported in patients with an autosomal recessive condition known as the Athabascan-type severe combined immunodeficiency (SCIDA) and the Omenn syndrome (Ege et al., 2005; L. Li et al., 2002). In addition to causing severe combined immunodeficiency, individuals with Artemis deficiency have a higher vulnerability to lymphomas (Powell et al., 2017). Studies have demonstrated that murine embryonic stem cells lacking Artemis display chromosomal instability, such as telomere fusion, increased sensitivity to radiomimetic drugs, and susceptibility to ionizing radiation (Zha et al., 2007). Due to the pivotal function of Artemis in DNA repair and its biological significance, there is an increasing belief that this protein might also play a role in the formation of non-lymphoid cancers (de Miranda et al., 2011).

So far, no study has specifically investigated the role of Artemis in the development of breast cancer. However, several research groups have reported the detection of DCLRE1C deleterious mutations among their studied breast cancer cohorts. In 2012, Pylkäs et al. conducted a study to assess the impact of copy number variations (CNVs) on susceptibility to breast cancer. A genome-wide scan was performed on the germline DNA of 103 familial and early onset breast cancer patients who tested negative for BRCA1, BRCA2, and PALB2 mutations, along with 128 controls who were geographically and ethnicity-matched to the patients. The researchers noted a higher occurrence of disrupting rare CNVs in DCLRE1C among the cases compared to the control group, concluding that such structural variations in DCLRE1C may play a role in predisposition to breast cancer (Pylkäs et al., 2012). Later, in 2014, the same group of researchers investigated the frequency of rare DCLRE1C CNVs in 842 Finnish breast cancer cases, unselected for family history of cancer and age at disease onset, along with 497 geographically matched healthy female controls, and observed no significant difference

in the carrier frequencies between the case and control subjects (Tervasmäki et al., 2014). More recently, a study conducted in Latin America recruited 403 breast cancer patients with HBOC criteria from 11 health institutions in Argentina, Colombia, Guatemala, Mexico, and Peru. A pilot cohort of 222 patients was then screened for a panel of 143 genes, including DCLRE1C, that were selected based on their putative role in susceptibility to different hereditary cancers. With regard to DCLRE1C, the study only found a singleton likely pathogenic variant in one HBOC patient (Oliver et al., 2019). Similarly, other studies have detected deleterious DCLRE1C mutations among their breast cancer cohorts. However, these mutations were not present at a significant level to warrant further analysis (Barrett et al., 2018; Lhota et al., 2016; Shahi et al., 2019).

In the preliminary study by our research team among the Bahamian population (please refer to Section 2.1.2 for more information), DCLRE1C LoF variants were detected in three of the 387 unselected breast cancer patients versus none of the 653 control subjects (OR = 11.9; p-value = 0.05). In the present study, we identified a DCLRE1C frameshift insertion (rs760288938) among 3/421 Iranian familial breast cancer cases and 2/800 controls from the Iranome database (Table 3). However, we did not observe a statistically significant difference in carrier frequencies between the Iranian case and control subjects (OR = 2.86, 95% CI = 0.47-17.21, p = 0.34). Within the UK Biobank database, 18 out of 15,643 breast cancer-affected Caucasian women (0.12%) and 120 out of 157,943 cancer-free Caucasian females (0.07%) harbored germline LoF variants in DCLRE1C (Table 4). Similar to the Iranian cohort, LoF variants in DCLRE1C were not associated with a significantly higher risk of breast cancer among the UK Biobank study subjects (OR = 1.54, 95% CI = 0.93–2.53, p = 0.1).

The third candidate gene that we studied was DDX19B. This gene is on chromosome 16q22, a genomic region associated with frequent LOH in various tumor types, including those of the breast, prostate, and liver (Cheung et al., 2014; Fawzy et al., 2015; Kluth et al., 2015). Since the 1980s, several studies have reported LOH at chromosome 16q in breast tumors, with a prevalence ranging from 28% to as high as 90% (Bürger et al., 2013; Hungermann et al., 2011; Voutsadakis, 2022). These observations suggest the presence of tumor-suppressor genes in the long arm of chromosome 16, in which deleterious alterations may be linked to breast tumorigenesis.

As such, extensive research has focused on this genomic region in the hope of identifying new breast cancer susceptibility gene candidates. DDX19B is proposed as one such potential candidate due to its specific location and biological role in crucial cellular functions (Rakha et al., 2006). Nevertheless, the rarity of recurring LoF mutations in this gene has hindered researchers from examining its contribution to breast cancer susceptibility.

DDX19B is an RNA helicase and plays a crucial role in multiple biological pathways. It primarily functions as an ATP-dependent RNA unwinding enzyme and is also responsible for transporting mRNA from the nucleus to the cytoplasm. In addition to its nuclear import/export functions, studies have shown that DDX19B plays a fundamental role in terminating the translation process through the recognition of the stop codon. It then brings in the release factors eRF1 and eRF3 to the translation complexes and interacts with them to ensure the efficient termination of the translation process. Furthermore, it has been proposed that DDX19B might participate in activating translation through the interaction with ribosomes and tRNA (Mikhailova et al., 2017).

mRNA transportation involves the interaction between various mRNA export factors and nucleoporins, the building blocks of nuclear pore complexes. Consequently, any disruption in this mechanism can have harmful effects and may contribute to the development of cancer (Conneely & Stevens, 2021; Mendes & Fahrenkrog, 2019). For instance, chromosomal translocation of the NUP214 gene, which encodes a nucleoporin that directly interacts with DDX19B during mRNA export, is shown to result in the formation of a fusion protein with potentially oncogenic properties. Chromosomal translocation of NUP214 is frequently reported in acute myeloid and nonlymphocytic leukemia. Earlier research has also demonstrated that NUP214 knockout mice are not capable of surviving (Culjkovic-Kraljacic & Borden, 2013; Mendes & Fahrenkrog, 2019). Although DDX19B's role in breast cancer remains unclear, disruptions in mRNA export involving DDX19B and related proteins may lead to carcinogenesis, highlighting the potential impact of dysfunctions in these pathways.

In the preliminary study by our research team among the Bahamian population (please refer to Section 2.1.2 for more information), a recurrent DDX19B frameshift

variant (NM_007242.5:c.988delG) was detected in three of the 387 unselected breast cancer patients and no carriers were found among the 653 control subjects (OR = 11.9; p-value = 0.05). This variant is located in the highly conserved C-terminus domain of DDX19B. The role of the C-terminus domain of DDX19B has yet to be determined. Nevertheless, due to its high level of conservation, mutations occurring within this region are likely to be deleterious. In the present study, however, due to the rarity of DDX19B deleterious mutations, we were not able to detect any DDX19B LoF variants among the Ontario, French-Canadian, and Iranian study subjects and only observed germline DDX19B LoF variants in 16 out of the 157,943 Caucasian cancer-free females within the UK Biobank database (Table 6).

Our fourth and last candidate gene was XRCC3. This gene is on chromosome 14q32 and belongs to the RAD51 gene family members that play a crucial role in preserving genomic stability through HRR. The paralogs within this family engage in mutual interaction and assemble distinct complexes, each of which independently fulfills a crucial function in the HRR pathway. For instance, XRCC3 and RAD51C create a complex known as CX3, which is one of the complexes recruited to the site of DNA damage at the time of BRCA1 and BRCA2-dependent HRR (J. Chun et al., 2013). Consequently, any abnormalities in these genes may compromise cells' ability to repair DNA double-strand breaks, which can ultimately contribute to carcinogenesis (Suwaki et al., 2011; Thacker, 2005). Studies have reported that XRCC3 mutant hamster cells demonstrate chromosomal instability, defective HRR pathway, and increased sensitivity to radiation and DNA cross-linking agents (Stackpole et al., 2007). Such studies provide further evidence that XRCC3 plays a critical role in maintaining genomic stability, and as such, deleterious mutations in this gene may be detrimental and lead to tumorigenesis.

For years, scientists have been investigating the role of XRCC3 in the development of various cancer types, including malignancies of the breast tissue. The polymorphism p.Thr241Met (c.722C>T; rs861539) in this gene has undergone extensive research and has been shown to increase vulnerability to different malignancies such as breast cancer among different populations (Bashir et al., 2015; Dashti et al., 2019; Ozoran et al., 2022; L. Yang et al., 2011). As a result, the literature contains very little information about other XRCC3 variants and their association with susceptibility to breast cancer.

In the preliminary study by our research team among the Bahamian population (please refer to Section 2.1.2 for more information), four XRCC3 LoF mutations were detected among four of the 387 unselected Bahamian breast cancer cases and one LoF mutation was observed in one individual out of the 653 control subjects (OR = 6.8; p-value = 0.067). In the present study, due to the rarity of XRCC3 deleterious mutations, we were not able to detect any XRCC3 LoF variants among the Ontario, French-Canadian, and Iranian study subjects. Within the UK Biobank database, we identified 10 individuals harboring XRCC3 LoF variants among the 15,643 Caucasian breast cancer patients (1 in every 1,564 Caucasian women with breast cancer). We observed germline XRCC3 LoF variants in 70 out of the 157,943 Caucasian cancer-free females that served as the control group (1 in every 2,256 healthy individuals), concluding that the frequencies of LoF variants in XRCC3 were not statistically different between the Caucasian breast cancer patients (cases) and cancer-free individuals (controls) within the UK Biobank database (OR = 1.55, 95% CI = 0.79–3.01, p-value = 0.2).

To sum up, our research team identified ATRIP as a novel breast cancer susceptibility gene candidate among the founder population of Poland. In this study, we were able to replicate this observation among the Caucasian population. By analyzing the UK Biobank sequence data, we showed that germline LoF mutations in ATRIP also predispose to an increased risk of breast cancer among the Caucasian participants within the UK Biobank cohort study. Figure 5 illustrates the approximate locations of ATRIP functional domains, along with all germline pathogenic and likely pathogenic LoF variants identified across the Canadian, Iranian, and British cohorts in our study.

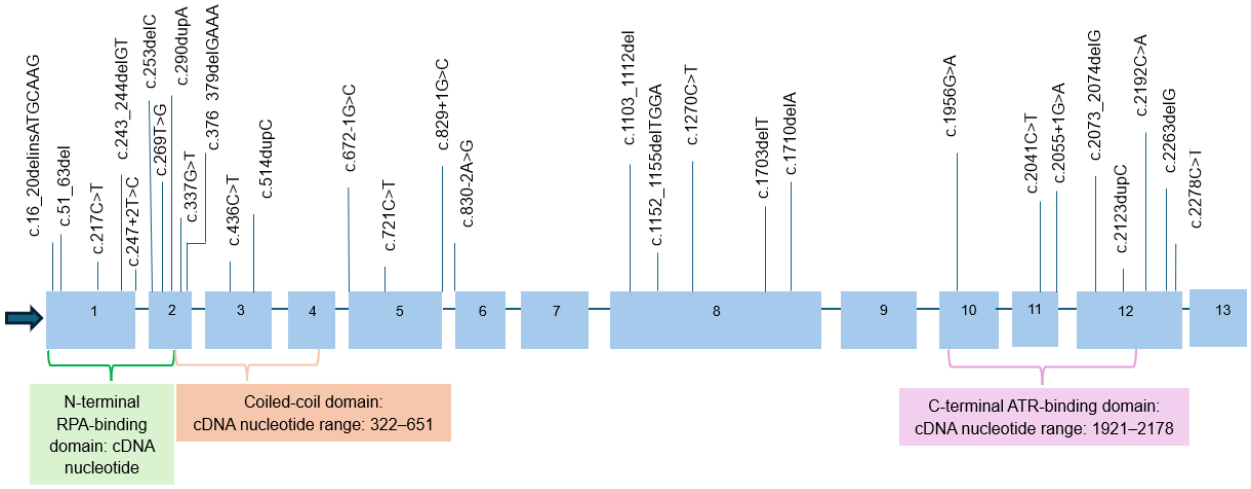


Figure 5: Location of germline pathogenic/likely pathogenic LoF variants in ATRIP (GenBank: NM_130384.3) identified in the Canadian, Iranian, and British cohorts. Positions of functional domains and variants are an approximation (retrieved from https://www.uniprot.org/uniprot/ and https://www.ensembl.org/).

Chapter 4 – A Novel ATRIP Mutation Detected by WES in an Iranian Family with Familial Clustering of Breast Cancer: A Case Report

Here, we report a novel ATRIP mutation (NM_130384.3:c.1033delC) among several members of an Iranian family with the clustering of breast cancer but negative for mutations in the already known breast cancer risk genes. To the best of our knowledge, this is the first report of the detection of a deleterious variant in ATRIP among several members of an Iranian family with a family history of breast cancer and negative for mutations in the already known breast cancer susceptibility genes.

4.1 Case Report

The proband (Figure 6, II-6) was a 52-year-old cancer-free Iranian woman who referred to Noor genetic laboratory (Tehran, Iran) due to concern regarding multiple first- and second-degree family members affected with breast cancer. She wanted to be screened for mutations in breast cancer susceptibility genes and be assessed for her risk of developing breast cancer. The proband's mother (Figure 6, I-4) was diagnosed with grade III invasive ductal carcinoma of the left breast at the age of 65 years old, with the largest tumor dimension of 5 cm and extensive tumor necrosis. IHC staining revealed expression of the estrogen receptor (ER+) and a lack of expression of the progesterone receptor (PR-). The patient died of metastatic complications. One of the proband's sisters (Figure 6, II-5) was diagnosed with grade I invasive ductal carcinoma of the right breast at the age of 46 years old and died of cancer-related complications. Two of the three daughters of this patient (nieces of the proband, Figure 6, III-4 and III-6), whose details are unknown, developed invasive ductal carcinoma when they were 36 and 40 years old, respectively, and both were alive at the time of this study. The father of these two breast cancer-affected sisters (husband of the proband's sister, Figure 6, II-4) also developed prostate cancer at the age of 64 years old, and his mother (Figure 6, I-2), who was not alive at the time of this study, was affected with breast cancer when she was 56 years

old. Moreover, the paternal cousin of the proband's affected nieces (Figure 6, III-3) was diagnosed with breast cancer at the age of 32 years old.

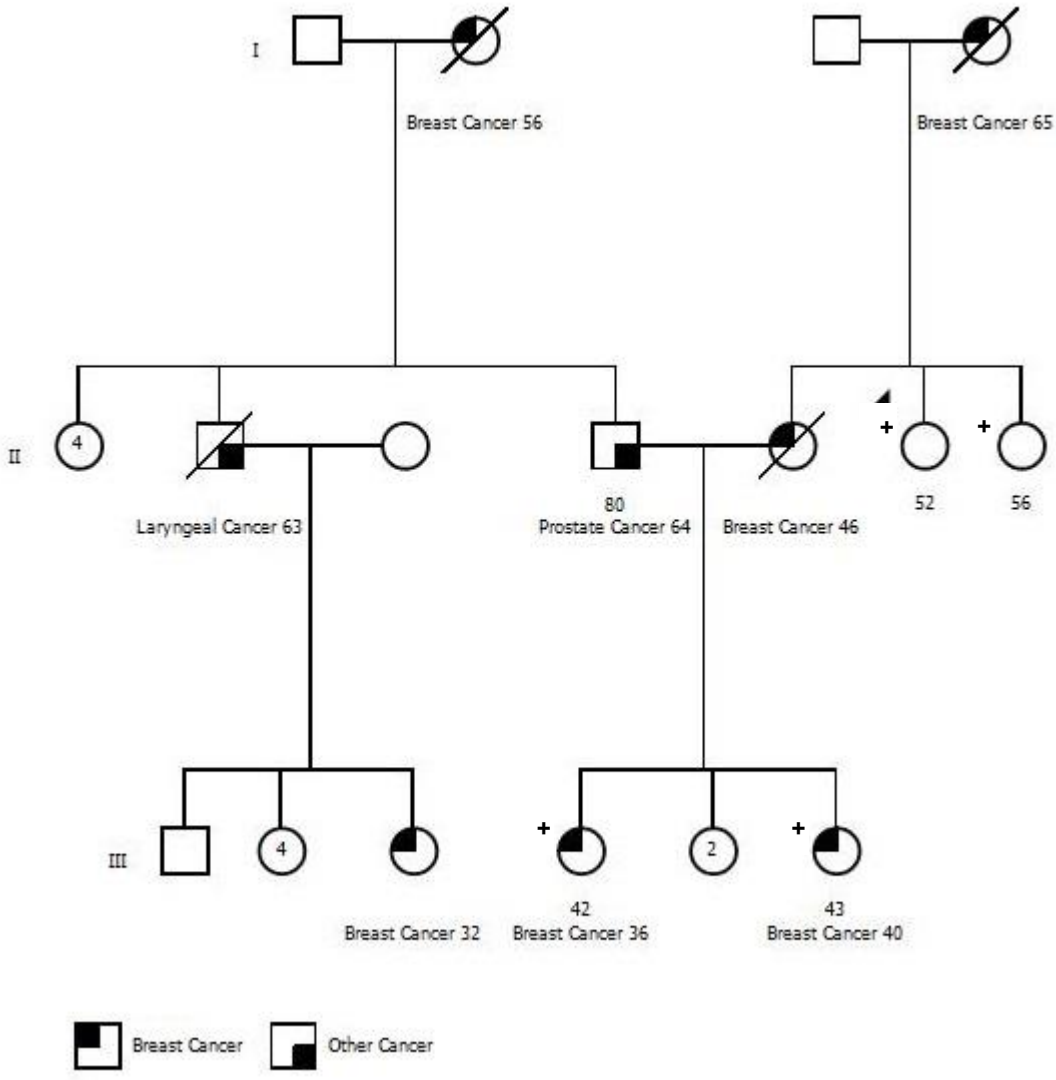


Figure 6: Pedigree for an Iranian family with clustering of breast cancer and carriers of the germline ATRIP c.1033delC mutation. The proband is shown with an arrow. The type of cancer and age at the time of diagnosis are indicated for individuals with cancer diagnosis. Square represents male and circle represents female. Deceased individuals are indicated by a diagonal line through the symbol. Carriers of the ATRIP c.1033delC mutation are indicated by a + sign next to the symbol.

To unveil the underlying genetic predisposition, the proband was screened for mutations in a panel of 45 cancer-related genes, including the already known breast cancer susceptibility genes, in a diagnostic laboratory. The gene panel sequencing did not reveal any putatively deleterious mutations in the 45 cancer-related genes that were tested. To further investigate the breast cancer hereditary component in this family, blood or saliva samples from six family members were sent to our Molecular Genetics Research laboratory at Women's College Hospital, Toronto, Canada. Germline DNA was extracted from the samples, and WES was performed using NGS technology (Illumina NextSeq 500). Considering the very low mutation frequency of ATRIP, our newly discovered breast cancer susceptibility gene candidate, and finding no ATRIP mutation carrier among over 400 Iranian familial breast cancer cases that we tested previously, we did not expect to find an ATRIP mutation in this family and performed WES in search of other new candidate genes. But, surprisingly, we identified a novel deleterious mutation in this gene, a base pair deletion at position 1033 in the ATRIP coding sequence (NM_130384.3:c.1033delC), among four of the six tested family members (Figure 7). This mutation is classified as likely pathogenic according to the ACMG guidelines. Interestingly, all the six exome-sequenced family members were negative for mutations in the already known breast or any other cancer susceptibility genes.

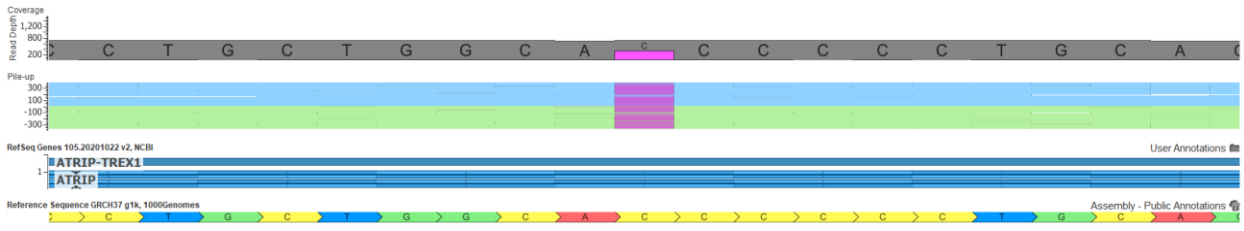


Figure 7: Next generation sequencing reads pileup of ATRIP c.1033delC mutation identified among four members of an Iranian family with clustering of breast cancer along with RefSeq V105 and GRCH37 human genome reference regions were shown.

The four ATRIP mutation carriers included the cancer-free proband (Figure 6, II-6), the proband's unaffected sister (Figure 6, II-7), and the two breast cancer-affected nieces of the proband (Figure 6, III-4 & III-6). As the father of these two affected sisters (Figure 6, II-4) was not an ATRIP mutation carrier, we concluded that their mother (proband's sister, Figure 6, II-5), who had died of breast cancer before this study, was the carrier of the ATRIP c.1033delC mutation. The ATRIP mutation identified by NGS among the four family members was confirmed by Sanger sequencing. Tumor samples of the proband's two breast cancer-affected nieces that harbored the ATRIP c.1033delC mutation were not available to perform additional studies, including LOH and mutational signature analyses.

Chapter 5 – Landscape of Somatic Point Mutations in Breast Cancer

5.1 Definition of Mutational Signatures

Until recently, the primary focus of cancer research was on identifying “driver” mutations. These are the mutations that play a causal role in cancer development and confer selective advantages in the course of cancer evolution. Additionally, a cancer genome harbors numerous “passenger” mutations, which may not be directly responsible for the initiation and development of cancer but provide valuable insight into the origins and causes of cancer. These mutations are the results of, and therefore, exhibit the effects of biological disruptions (i.e., the mutational processes) that have taken place during the progression of a cancer. It is hypothesized that every mutational process imprints a distinct pattern of somatic mutations, known as a mutational signature, on the cancer genome. This signature is determined by the specific type of DNA damage caused by external or internal DNA damaging agents, as well as by deficiencies in DNA repair or replication mechanisms. Additionally, the final catalogue of mutations in a cancer genome is influenced by the intensity and duration of exposure to each mutational process. In other words, the number of mutations that contribute to each signature (i.e., the weight of each signature in a cancer genome) serves as an indicator of the level of exposure to each mutational process. Lastly, because multiple internal and/or external mutational processes can act on the genome at the same time or in succession, the final mutational catalogue of a cancer genome bears a combination of multiple signatures with different weights generated by various mutational processes. Hence, the capability to accurately discern mutational signatures and measure their magnitude in a specific cancer genome is of great importance as these signatures provide valuable insight into the causes or origins of the tumor and can reveal underlying biological abnormalities that can be targeted for therapy.

While mutational signatures are relatively a new concept in the field of cancer biology, the notion of linking mutational processes to specific mutational patterns is not

novel. For example, exposure to aristolochic acid (AA), a carboxylic acid found in certain herbal remedies and contaminated flour, has been associated with carcinomas of the upper urinary tract (UUT). It has been shown that TP53 mutations in UUT patients exposed to AA have a distinctive pattern of mutations characterized by A>T transversions on the non-transcribed strand, clustering at hotspots rarely mutated in other cancers. This unique pattern, originally discovered among UUT patients in Taiwan and Balkans, has been replicated experimentally by exposing mouse cells with human TP53 sequences to AA (Hollstein et al., 2013). Similarly, other studies have demonstrated initial evidence that specific mutational processes can leave distinctive patterns on the genome, which can be observed and analyzed in tumor samples (Giglia-Mari & Sarasin, 2003; Pfeifer et al., 2002).

5.2 Mutational Signatures of Single-Base Substitutions

Of all different types of genetic alterations, a specific emphasis has been placed on the signatures of single-base substitutions. These substitutions have been categorized into six types according to the pyrimidine base that is mutated (C>A, C>G, C>T, T>A, T>C, and T>G). Nevertheless, the sequence context, that is the flanking bases immediately adjacent to the mutated base, has been found to influence mutation rates in the genome and should be accounted for when defining a mutational signature. Given that there are 6 classes of base substitutions and 16 potential sequence contexts for each mutated base (A, C, G, or T at the 5′ base and A, C, G, or T at the 3′ base), a total of 96 distinct single-base substitutions can occur in a trinucleotide sequence context. Subsequently, computational models such as matrix decomposition algorithms can be applied to detect and analyze these patterns (detailed information is provided in Section 5.3). Such a mathematical framework was originally utilized on ~1,000 whole genomes and ~10,000 whole exomes from 40 distinct cancer types obtained from large databases, including the Cancer Genome Atlas (TCGA), the International Cancer Genome Consortium (ICGC), and a large set of publicly available somatic mutations published in peer-reviewed journals, resulting in the discovery of 30 distinct single-base substitution signatures (Alexandrov, Nik-Zainal, Wedge, Aparicio, et al., 2013; Alexandrov, Nik-Zainal, Wedge, Campbell, et al., 2013; Nik-Zainal, Alexandrov, et al., 2012). These validated

mutational signatures are documented in a repository on the COSMIC (Catalogue of Somatic Mutations in Cancer) website and have been extensively used as reference signatures. A more recent study by Alexandrov et al. has introduced an updated set of signatures derived from a much larger dataset of ~19,000 whole exomes and ~4,600 whole genomes, comprising the sequences from the ICGC/TCGA Pan Cancer Analysis of Whole Genomes (PCAWG) project. The updated repertoire includes 49 additional mutational signatures of single-base substitutions (Alexandrov et al., 2020). Approximately half of the already known signatures have been linked to internal mutagenic processes such as the actions of the APOBEC family of deaminases, as well as deficiencies in DNA repair mechanisms like mismatch repair, homologous recombination, or faulty DNA polymerases, or external DNA damaging agents such as exposure to tobacco carcinogens, UV radiation, aflatoxin B1, alkylating chemotherapy drugs, and aristolochic acid. However, the etiology of the remaining signatures has not been determined yet. Uncovering the mechanisms responsible for these signatures will provide fresh insights into the origins of cancer.

5.3 Mathematical Definition of Mutational Catalogues and Signatures

Here, we briefly review the mathematical concept for discerning signatures of base substitutions based on the original framework developed by Alexandrov et al. in 2013 (Alexandrov, Nik-Zainal, Wedge, Campbell, et al., 2013). The input data to extract mutational signatures is referred to as the "mutational catalogue" of a genome or exome, with the counts for each of the 96 base substitution types as previously described in Section 5.2.

The mutational catalogue of the genome or exome "g" is defined as the vector $\left(m_g^1, \cdots, m_g^K\right)^T$, where each m_g^k is the number of mutations of type k observed in the sample "g", and k equals to 96, which is the number of possible base substitution types. The superscript T denotes the transpose of the matrix so that the vector is considered as a column vector.

The mutational catalogue of a sample is formed by the combination of various mutational processes operative throughout the lifetime. Consequently, this catalogue can be regarded as the weighted combination of multiple mutational signatures, each exclusively corresponding to a specific mutational process. The weight of a signature is greater if the underlying mutational process is more intense or lasts for a longer period of time compared to the other mutational processes.

The signature of a mutational process "n" is a vector $P_n = (P_n^1, \dots, P_n^K)^T$, where each P_n^k represents the probability that the mutational process "n" will induce a mutation of type k. In other words, P_n^k is the expected relative frequency of type k mutation in a genome exposed to "n". Of note, $\sum_{k=1}^{K} P_n^k = 1$ and $0 \leq P_n^k \leq 1$ for all k.

The intensity of exposure to a mutational process "n" in a sample "g" is measured by the number of mutations in "g" that is due to "n", depicted as e_g^n. For this reason, e_g^n is referred to as the exposure of "g" to "n". The expected number of mutations of type k due to the mutational process "n" in sample "g" is therefore $p_n^k e_g^n$. If sample "g" has been exposed to N mutational processes, then the total number of mutations of type k is

$$m_g^k = \sum_{n=1}^{N} p_n^k e_g^n + \epsilon_g^k \quad (1)$$

where ϵ_g^k is an error term reflecting sampling variability and non-systematic errors in sequencing or subsequent analysis.

5.3.1 De Novo Signature Extraction vs. Refitting

De novo signature extraction methods are often applied when researchers aim to discover novel mutational signatures using large datasets without prior knowledge of the potential sources of mutations. More specifically, de novo methods seek to estimate the values of p and e in equation (1) based on the given value of m (Omichessan et al., 2019). Non-negative matrix factorization (NMF) has been the most widely used mathematical approach to solve this equation. NMF identifies the two matrices p and e that minimize the distance between the original matrix, m, and the product of p and e. For this purpose, NMF necessitates the predefinition of the number of signatures N, which is

an unknown parameter. A method for determining this parameter involves obtaining the factorization of m for multiple values and subsequently selecting the optimal N based on a performance measure, such as the reconstruction error or the overall reproducibility.

Refitting approaches entail the extraction of mutational signatures using predefined templates, that is mutational signatures that have been already identified and validated in previous studies and are currently used as reference signatures. More specifically, refitting methods involve fitting an established and validated set of mutational signatures to an observed mutational catalogue with the aim of determining the contribution of each pre-defined signature to the dataset. In other words, refitting approaches assume that the signatures p are already known and aim to estimate e based on the given values of m and p. Unlike de novo methods, refitting can be done for individual mutational catalogues (i.e., individual samples) (Omichessan et al., 2019).

A current practice consists in first performing a de novo extraction of signatures and then comparing the newly identified signatures with the reference signatures (such as COSMIC signatures) using a similarity score. This score is usually measured using cosine similarity, which ranges from 0 (indicating complete dissimilarity) to 1 (indicating identical signatures) (Alexandrov, Nik-Zainal, Wedge, Aparicio, et al., 2013; P.-J. Huang et al., 2018). Ultimately, a "novel" signature is deemed to represent a particular reference signature if the cosine similarity score exceeds a predetermined threshold. If similarity is observed with more than one reference signature, the one with the highest similarity score is chosen. The selection of the cut-off value for this assignment is a critical factor, and there is currently inconsistency in the literature regarding its choice. Some studies have used a value of 0.75, while others have used 0.80 (Han et al., 2018; Letouzé et al., 2017).

5.4 Signatures of Base Substitutions in Breast Cancer

In 2012, a proof-of-principle exercise utilized 183,016 single-base substitutions found in 21 breast cancer whole genomes to establish the presence of mutational signatures of base substitutions (Nik-Zainal, Alexandrov, et al., 2012; Nik-Zainal, Van Loo, et al., 2012). The researchers investigated different mathematical techniques

and ultimately employed NMF to identify five substitution signatures found in these tumors, designated as Signatures A-E, later referred to as Signatures 1A/B, 2, 3, 8, and 13 (Nik-Zainal, Alexandrov, et al., 2012; Nik-Zainal, Van Loo, et al., 2012).

Following this preliminary finding, the same mathematical approach was utilized on 507 whole genomes and 6,535 whole exomes from 30 different types of cancer, encompassing 100 whole-genome and 800 whole-exome sequenced breast tumors. Encouragingly, the previously identified five substitution signatures were consistently detected in this expanded dataset, further strengthening confidence in the notion of mutational signatures and in the techniques employed to extract them (Alexandrov, Nik-Zainal, Wedge, Aparicio, et al., 2013).

Subsequently, in a seminal study examining 560 whole-genome sequenced breast tumors, a total of twelve single-base substitution signatures were found from a pool of 3,479,652 base substitutions, comprising the formerly identified five signatures and an additional seven signatures, of which five had previously been detected in other cancer types (Signatures 5, 6, 17, 18 and 20) and two novel signatures (Signatures 26 and 30) (Nik-Zainal et al., 2016). At first glance, this may seem a significant increase in the discovery of unique signatures in breast tumors. However, upon careful examination, several of the newly identified signatures were relatively uncommon and were found only in a small number of samples. Therefore, it is probable that researchers have reached maximum capacity in identifying prevalent signatures of base substitutions in breast cancer, and obtaining additional sequencing data from primary breast tumors is unlikely to reveal any new major signature. The study of larger datasets may, however, increase the power that potentially enables the clarification of closely correlated signatures. As a good example, the initially identified Signatures 1A/B are recently split into two separate Signatures 1 and 5 after the analysis of larger cohorts of breast cancers (Nik-Zainal et al., 2016; Nik-Zainal & Morganella, 2017). Of note, the identified breast cancer signatures do not show any specificity towards a particular type of breast cancer as defined by hormone receptor status.

Among the twelve identified signatures in breast cancer (Figure 8), Signature 1 (characterized by C>T transitions at N<u>C</u>G) and Signature 5 (primarily characterized by C>T and T>C mutations) are known as “clock-like” signatures in that the number of

mutations correlates with the age of the individual at the time of diagnosis. The suggested etiology for Signature 1 is an internal mutational process triggered by the spontaneous or enzymatic conversion of 5-methylcytosine to thymine, resulting in the formation of G:T mismatches in double-stranded DNA. If these mismatches are not detected and corrected before DNA replication, the C>T substitution becomes permanently established. The proposed etiology of Signature 5 is unknown. These two signatures are not specific to breast tumors and have been observed in all the 40 distinct cancer types studied by Alexandrov et al., discussed earlier in Section 5.2 (Alexandrov et al., 2020; Nik-Zainal et al., 2016).

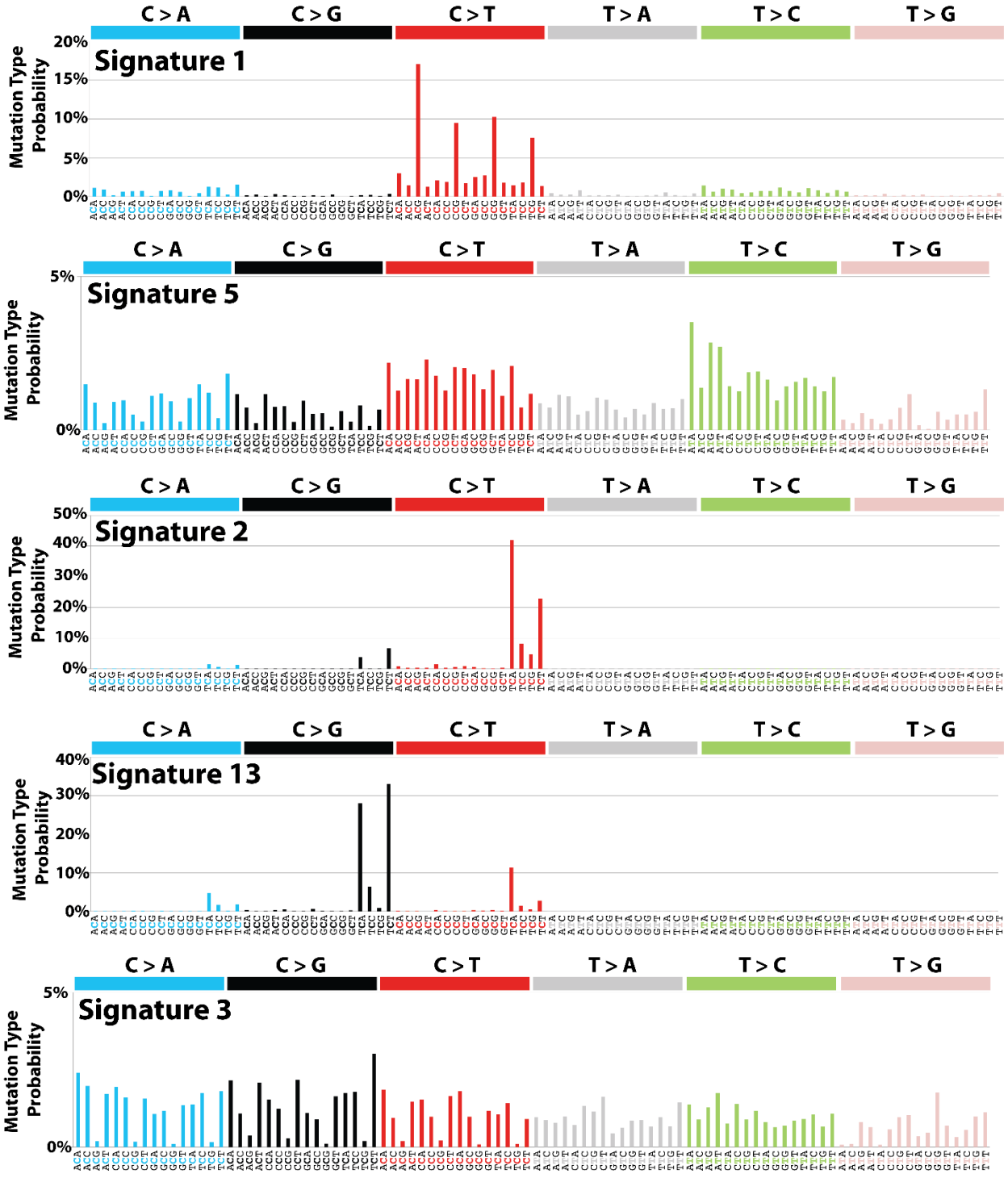
Signature 2 (predominantly composed of C>T substitutions at TCN) and Signature 13 (primarily characterized by C>G mutations at TCN) are attributed to the activity of the members of the APOBEC family of cytidine deaminases that deaminate cytosine to uracil. Mutations that contribute to these two signatures are likely produced by DNA replication across uracil or by error prone polymerases replicating across abasic sites generated by base excision repair removal of uracil. Interestingly, Signature 2 is usually observed in the same samples harboring Signature 13. In addition to breast tumors, Signatures 2 and 13 have been observed in several other cancer types, most importantly, bladder and lung carcinomas (Alexandrov et al., 2020; Nik-Zainal et al., 2016).

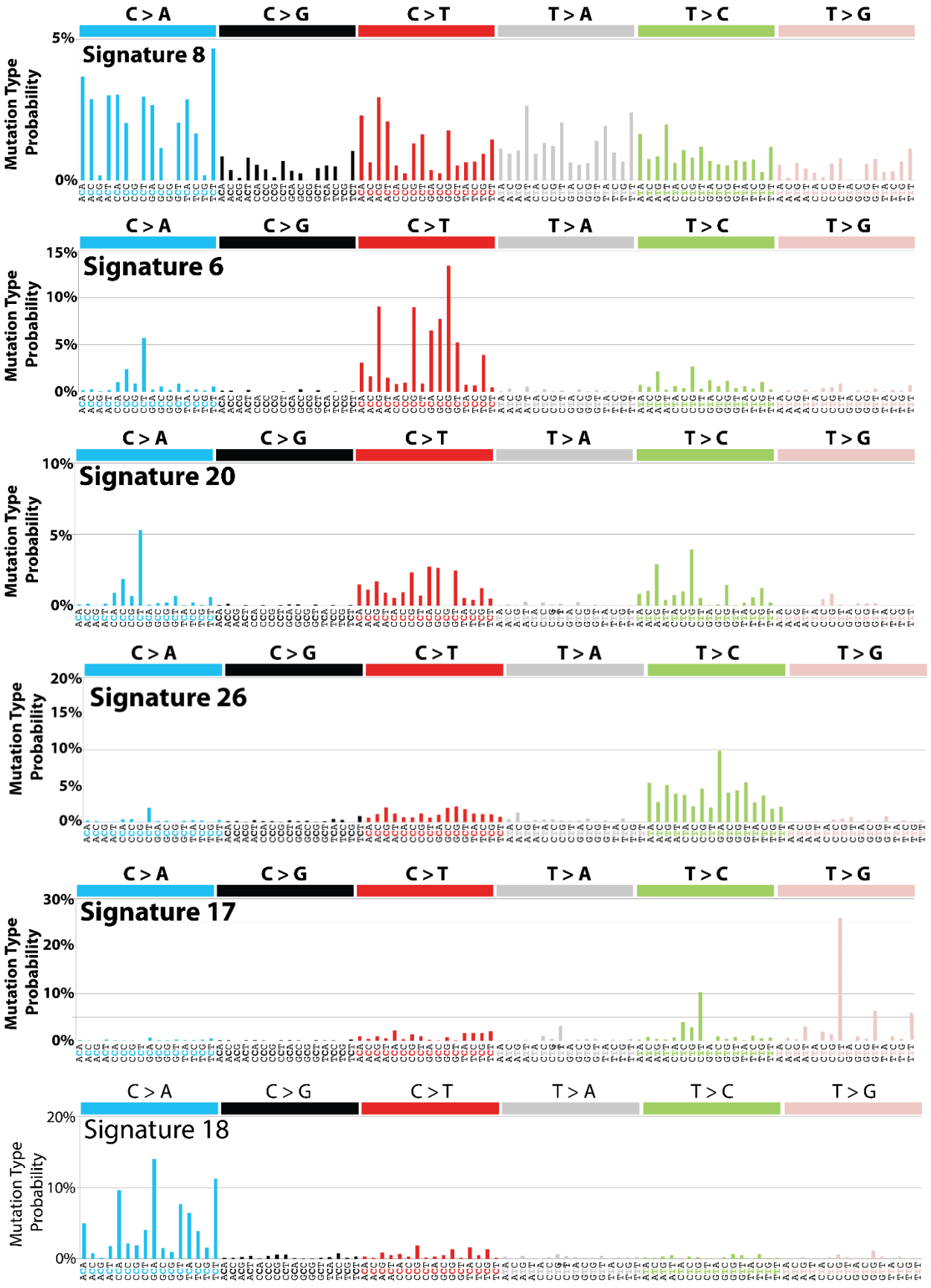
Signature 3 is strongly associated with defective homologous recombination repair pathway and is characterized by a fairly uniform distribution of mutations across all 96 possible single-base substitution types. This signature is observed in breast cancers (as well as in pancreatic and ovarian cancers) with complete BRCA1/2 loss of function due to germline or somatic point mutations in BRCA1/2 combined with loss of the wild-type allele (Nik-Zainal et al., 2016). Additionally, Signature 3 has been identified in breast tumors that showed epigenetic silencing of BRCA1 or RAD51C or harbored germline or somatic mutations in PALB2 or RAD51C (Nones et al., 2019; Polak et al., 2017). Interestingly, Signature 3 has been observed in BRCA-null isogenic cell lines, providing direct evidence of its association with HRR deficiency (Zámborszky et al., 2017). To a lesser extent, Signature 8 (enriched in C>A and T>A substitutions, but smaller contributions of other substitutions compared to Signature 3) is also associated with the absence of BRCA1 or BRCA2 function. The proposed etiology of Signature 8 is defects

in the HRR or nucleotide excision repair (NER) pathways lacking experimental validation. This signature is not specific to breast cancer and is also found in ovarian, pancreatic, prostate, and lung cancers (Alexandrov et al., 2020; Nik-Zainal et al., 2016).

Signature 6 (characterized by C>T mutations at NCG), Signature 20 (characterized by C>A mutations at CCN, as well as C>T and T>C mutations), and Signature 26 (primarily characterized by T>C mutations) are associated with defects in the DNA mismatch repair (MMR) pathway. In addition to breast tumors, these MMR-associated signatures are reported in carcinomas of colon, stomach, uterine, kidney, liver, prostate, esophagus, and pancreas (Alexandrov et al., 2020; Nik-Zainal et al., 2016). The association between MMR deficiency and these signatures has been validated in CRISPR-modified human stem cell colon organoids deficient in the MMR gene MLH1 (Drost et al., 2017). Using PMS2 and MLH1 knockout C. elegans, an experimental study revealed that defects in these two MMR genes give rise to the mutations that contribute to Signatures 6,20, and 26, further strengthening evidence that these signatures are linked to defective MMR pathway (Meier et al., 2018).

The etiology of Signature 17 (characterized by T>G mutations at NTT) is currently unknown. The proposed etiology of Signature 18 (exhibiting a high proportion of C>A mutations) is possibly the damage induced by reactive oxygen species that has been validated experimentally in human cell lines (Kucab et al., 2019). Finally, Signature 30 (primarily characterized by C>T transitions) is due to deficiencies in the base excision repair (BER) pathway. Using CRISPR-mediated knockout NTHL1, a major BER component, it has been demonstrated that NTHL1 deficiency in colon organoids leaves a specific pattern of mutations that contribute to Signature 30 (Drost et al., 2017). This signature was detected in only a single breast cancer patient in the seminal analysis of 560 whole-genome sequenced breast tumors by Nik-Zainal et al., and upon examining the germline DNA of this patient, a premature stop codon in NTHL1 was detected with clear evidence of loss of heterozygosity in the tumor (Nik-Zainal et al., 2016).


C > A
C > G
C > T
T > A
T > C
T > G
Signature 1
Mutation Type Probability
20%
15%
10%
5%
0%
Signature 5
5%
0%
Signature 2
50%
40%
30%
20%
10%
0%
Signature 13
40%
30%
20%
10%
0%
Signature 3
5%
0%


C > A
C > G
C > T
T > A
T > C
T > G
Signature 8
Signature 6
Signature 20
Signature 26
Signature 17
Signature 18
Mutation Type Probability
5%
0%
15%
10%
20%
30%

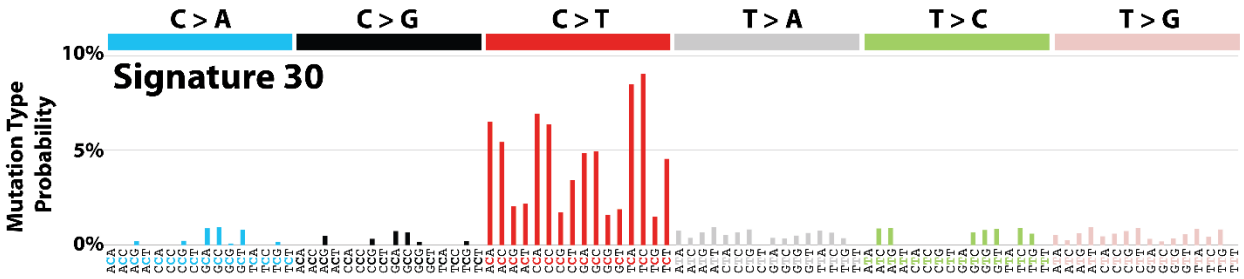


Figure 8: Twelve single-base substitution signatures identified in human breast tumors. These signatures are listed as reference and validated signatures in COSMIC database. In each signature, the 96 single-base substitution types are shown on the x-axis and their proportions are shown on the y-axis.

Chapter 6 – Tumor DNA Analysis

While performing association analysis is critical for assessing cancer susceptibility, we aimed to provide more lines of evidence supporting the causal role of deleterious mutations in our newly identified breast cancer susceptibility gene candidate, ATRIP. To address this objective, we performed LOH and mutational signature analyses on breast tumors of the germline ATRIP c.1152_1155del mutation carriers. If the germline ATRIP c.1152_1155del mutation coincides with LOH at the site of ATRIP, it suggests that our candidate gene may indeed play a role in breast tumorigenesis. Therefore, this observation strengthens the evidence for the causality of ATRIP's deleterious mutations in susceptibility to breast cancer. However, the absence of LOH can not rule out such an association. Moreover, analyzing the mutational signatures present in tumor samples of the germline ATRIP c.1152_1155del mutation carriers would give us a deeper perception of the underlying biological mechanisms driving tumorigenesis. Identification of signatures correlated to the known functions of ATRIP, shared among the germline ATRIP c.1152_1155del mutation carriers, provides further evidence supporting the involvement of ATRIP deleterious mutations in breast tumorigenesis.

As the preliminary findings regarding the three Bahamian breast cancer gene candidates, DCLRE1C, DDX19B, and XRCC3, were not replicated in our larger-scale association studies, we excluded these three genes from further analysis.

6.1 LOH Analysis at the Site of ATRIP Mutation and HRD Score Calculation

6.1.1 LOH Laboratory Analysis

Ten available tumor samples from Polish women with breast cancer who had the germline ATRIP c.1152_1155del variant were subjected to WGS (mean depth of coverage of 40×) for LOH analysis. Unstained formalin-fixed, paraffin-embedded (FFPE) sections were macro-dissected for tumor DNA extraction. Expert pathologists previously examined the hematoxylin and eosin (H&E) tumor slides to detect tumor/normal areas.

Tumor DNA was extracted from tumor cells using the QIAamp DNA Micro Kit (Qiagen) and subsequently quantified with the NanoDrop ND-1000 Spectrophotometer (Thermo Scientific Inc.). We conducted LOH analysis using NGS technology and prepared libraries using the xGen Prism DNA Library Prep Kit (Integrated DNA Technologies) according to the manufacturer's protocol. More specifically, the library construction comprised four major steps following the enzymatic fragmentation of DNA samples. These steps included end-repair/A-tailing, two consecutive ligation steps, and PCR amplification with the xGen Unique Dual Index (UDI) Primer Pairs. DNA purification was performed following the end repair, second ligation, and PCR steps using AMPure XP beads. Next, every five indexed samples were pooled together and used for paired-end sequencing for 300 cycles (2 × 150) on one lane of S4 flow cell of the Illumina NovaSeq 6000 sequencer for whole-genome sequencing.

For five tumor DNA samples with borderline evidence of LOH using their WGS data (explained in Section 6.3), we performed WES (mean depth of coverage of 200×) to further evaluate LOH in these samples. For this purpose, we prepared captured libraries using the xGen Hybridization and Wash v2 kit (Integrated DNA Technologies) according to the manufacturer's protocol. Post-capture PCR was performed with the xGen Library Amplification Primer Mix and xGen 2x HiFi PCR Mix (Integrated DNA Technologies) according to the xGen protocol. After PCR clean up with AMPure XP beads, DNA quality check was performed using Agilent's D1000 ScreenTapes on the TapeStation 2200 system (Agilent Technologies, Santa Clara, CA), and DNA was quantified using Qubit dsDNA High Sensitivity Assay Kits with the Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA). We pooled five captured libraries together and used for paired-end sequencing for 300 cycles (2×150) on the Illumina NovaSeq 6000 sequencer for whole-exome sequencing.

6.1.2 LOH Data Analysis

We looked for another pathogenic mutation in ATRIP in addition to the ATRIP c.1152_1155del germline mutation or the deletion of the entire wild-type allele. We considered LOH if the variant allele frequency (VAF) was greater than 60%. The VAF cut-off point of 60% is based on the formula VAF = (%tumor cells) / (%tumor cells + 2 × (100

- %tumor cells)), which takes into consideration that the deletion of the wild-type allele in hereditary syndromes is an early phenomenon, and therefore, it is present in most tumor cells. This represents 75% purity of the tumor cells.

6.1.3 HRD Score Calculation

For each tumor, the HRD score was calculated as the unweighted sum of three independent DNA-based measures of genomic instability that includes genomic LOH, telomeric allelic imbalance (TAI), and large-scale state transition (LST) scores (Telli et al., 2016). The HRD-LOH score was defined as the number of LOH regions across the genome that are longer than 15 Mb but shorter than the entire chromosome (Abkevich et al., 2012). The HRD-TAI score was defined as the number of regions with an allelic imbalance extending to one of the sub-telomeres but not crossing the centromere (Birkbak et al., 2012; Timms et al., 2014). The HRD-LST score was calculated as the number of breakpoints between regions longer than 10 Mb after filtering out regions shorter than 3 Mb (Popova et al., 2012). NxClinical 6.2 (BioDiscovery LLC) was used for HRD score analysis. All tumor DNA samples were screened for mutations in 15 HRR genes (ATM, BARD1, BRCA1, BRCA2, BRIP1, CHEK1, CKEK2, CDK12, FANCL, PALB2, RAD51B, RAD51C, RAD51D, RAD54L, and PPP2R2A) that could result in high HRD score if they are mutated (Hirsch et al., 2021; Qing et al., 2021).

6.2 Tumor Mutational Signature Analysis in ATRIP Germline Mutation Carriers

6.2.1 Tumor Mutational Signature Laboratory Analysis

We used the same WGS data of the ten available FFPE tumor samples from Polish breast cancer cases with the germline ATRIP c.1152_1155del variant (explained in Section 6.1.1) for mutational signature analysis. We only had access to the germline DNA sample, extracted from white blood cells, for two of these ten tumor DNA samples to perform matched tumor-normal somatic variant calling. For these two germline DNA samples, we prepared libraries using the xGen Prism DNA Library Prep Kit (Integrated

DNA Technologies) according to the manufacturer's protocol, as discussed in Section 6.1.1., and performed paired-end sequencing for 300 cycles (2x150) on the Illumina NovaSeq 6000 sequencer.

6.2.2 Identification of Tumor Mutational Signatures

Sentieon's TNseq pipeline was used to call somatic variants for both tumor-only and paired tumor-normal samples. Briefly, sequence reads were aligned to the human genome's reference sequence (hg19 version) using BWA incorporated into the Sentieon Package (Sentieon Inc., San Jose, CA, United States), followed by the removal of duplicates, realignment around indels, and recalibration of the base quality score. The quality of the aligned sequence data (mean base quality for each flowcell cycle, base quality score distribution, GC bias metrics, alignment metrics, and insert size metrics) was calculated for each sample with Sentieon's quality control algorithms based on Picard's various alignment metrics. Somatic variants, including SNVs and indels (up to about 50 bp in size), were called by the TNhaplotyper2 module of Sentieon's TNseq pipeline. For tumor-only analysis, the somatic variant calling process was performed with a user-built Panel of Normal (PoN) VCF. VCFs generated from TNseq were filtered to retain only variants that passed all internal filters and quality standards.

Tumor mutational signature analysis was performed based on the non-negative matrix factorization (NMF) framework, as explained in detail in Section 5.3.1. More specifically, we used SigProfilerMatrixGenerator (Khandekar et al., 2023) to create a mutational profile comprising all 96 possible single-base substitution (SBS) types for each tumor DNA sample, as described previously in Section 5.2. SBS signatures were inferred from both synonymous (silent) and nonsynonymous somatic single-nucleotide substitutions using SigProfilerAssignment (Díaz-Gay et al., 2023) and the online tool Signal (https://signal.mutationalsignatures.com/). Detected signatures were compared with the already known and validated signatures listed as references in the COSMIC database (https://cancer.sanger.ac.uk/signatures/sbs/) using cosine similarity. Each signature was assigned to the most similar COSMIC signature, provided that its cosine score was above the threshold of 0.75. We looked for the "best" combination of reference

signatures that could explain the observed mutational catalog and determined the signature profile and contribution in each cancer genome.

6.3 Results

We performed WGS (mean depth of coverage of 40×) on the tumor DNA of ten Polish women with breast cancer and available tumor samples who had the ATRIP c.1152_1155del variant in their germline DNA. Sequence data showed evidence of LOH at the site of ATRIP in six samples (Table 8). However, for five samples, either the VAF was borderline (around 60%) or the depth of coverage at the mutation site was too low to confirm the presence of LOH. To further evaluate LOH in these five samples, we performed WES on the tumor DNA (mean depth of coverage of 200×), and LOH was confirmed in three of the five samples (Table 8). All four samples with confirmed LOH through whole-genome and whole-exome sequencing had an HRD score greater than 42 indicating the presence of genomic instability (Telli et al., 2016). Three of the other six samples with no LOH also showed high HRD scores, two of the three were the samples with borderline evidence of LOH in WGS, which was not confirmed in their WES data (Table 8). In addition, none of the ten samples have any mutations in any of the 15 HRR genes, including ATM, BARD1, BRCA1, BRCA2, BRIP1, CHEK1, CKEK2, CDK12, FANCL, PALB2, RAD51B, RAD51C, RAD51D, RAD54L, and PPP2R2A (Hirsch et al., 2021; Qing et al., 2021).

Table 8: LOH analysis and HRD scores for ten tumor samples of Polish women with breast cancer harboring the germline ATRIP c.1152_1155del mutation

Tumor Sample	Tumor Histology [a]	LOH-WGS [b] AD (VAF)	LOH-WES [b] AD (VAF)	LOH Score	TAI Score	LST Score	HRD [c] Score
Sample 1	ER+/HER2-	10/5 (33.3%)	NA	3	1	20	24
Sample 2	ER+/HER2-	6/6 (50.0%)	NA	2	5	13	20
Sample 3	ER+/HER2-	5/8 (61.5%)	4,18 (81.8%)	15	12	35	62
Sample 4	ER+/HER2-	7/8 (53.3%)	NA	1	2	3	6

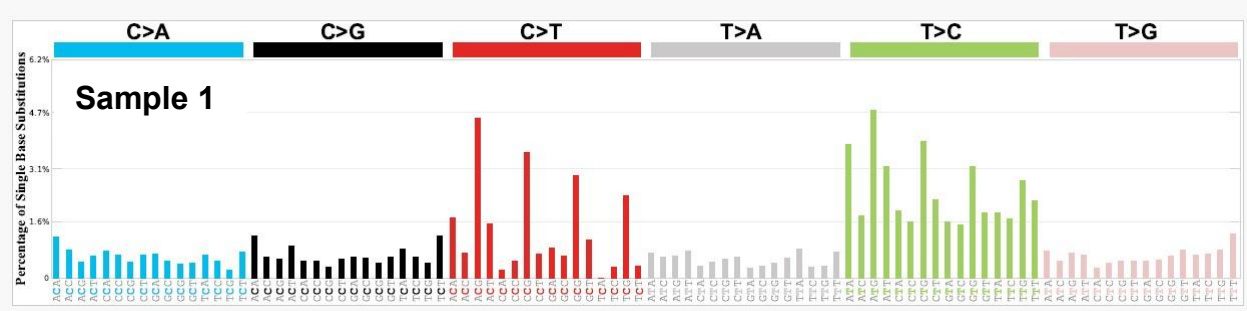
Sample 5	ER+/HER2-	0/6 (100%)	17,44 (72.1%)	18	19	49	86
Sample 6	ER+/HER2-	2/6 (75.0%)	57,70 (55.1%)	11	13	28	52
Sample 7	TN	8/14 (63.6%)	17,59 (77.6%)	10	17	53	80
Sample 8	ER+/HER2-	9/21 (70.0%)	NA	11	8	45	64
Sample 9	TN	11/8 (42.1%)	NA	4	13	60	77
Sample 10	ER+/HER2+	4/6 (60.0%)	91,98 (51.9%)	6	9	47	62

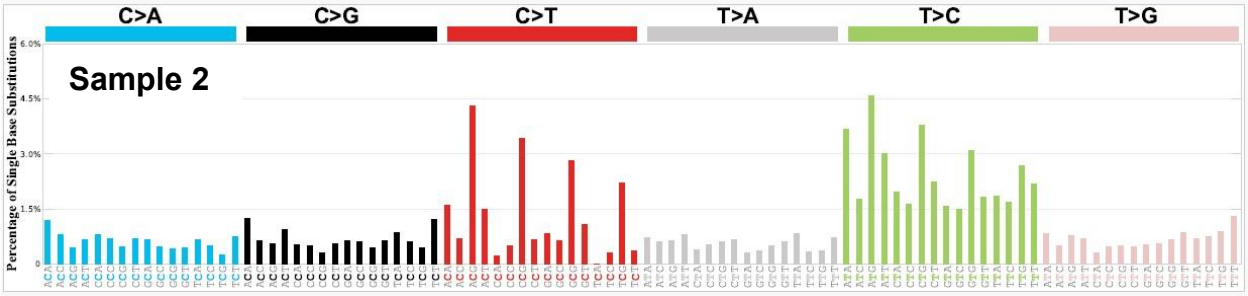
[a]Histology of the tested tumors based on presence of estrogen receptor (ER) and HER2; TN: Triple Negative

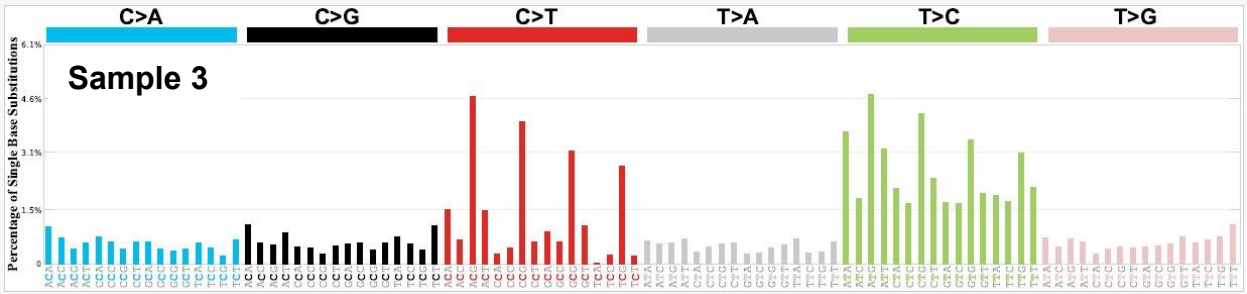
[b] Loss of heterozygosity (LOH) at the site of ATRIP mutation was detected using whole-genome sequencing (WGS). For five samples for which either the variant allele frequency (VAF) was borderline (around 60%), or the allelic depth of coverage (AD: Reference Allele/Alternative Allele) at the mutation site was too low, whole-exome sequencing (WES) was carried out to confirm LOH status at the site of ATRIP mutation.

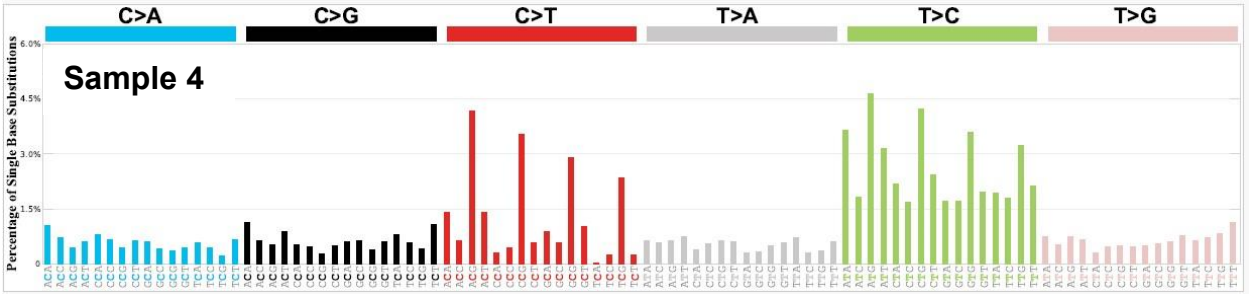
[c] Homologous recombination deficiency (HRD) score was calculated as the unweighted sum of the number of genomic LOH larger than 15 Mbp, telomeric allelic imbalance (TAI), and large-scale (at least 10 Mbp) state transition (LST) breakpoints across the tumor whole genome. None of the tumor DNA samples had any mutations in any of the 15 clinically important HRR genes (ATM, BARD1, BRCA1, BRCA2, BRIP1, CHEK1, CKEK2, CDK12, FANCL, PALB2, RAD51B, RAD51C, RAD51D, RAD54L, and PPP2R2A). NA: not applicable.

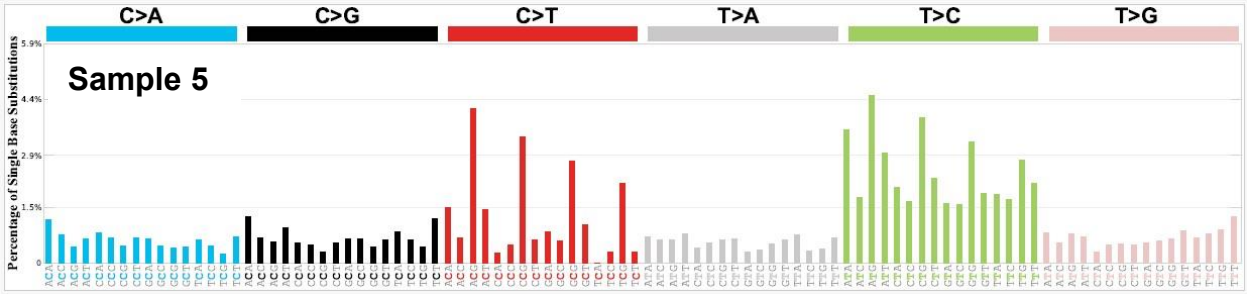
Next, for tumor mutational signature analysis, we called somatic point mutations using tumor-only WGS data for each of the ten FFPE tumor samples that harbored the germline ATRIP c.1152_1155del mutation. Then, for every tumor, we created a mutational profile with the counts of mutations for each of the 96 SBS types, which was used as the input for SBS signature analysis (Figure 9). We applied a tissue-specific approach and extracted SBS signatures using the set of twelve base substitution signatures known to occur in breast cancers, including Signatures 1, 2, 3, 5, 6, 8, 13, 17, 18, 20, 26, and 30 (Nik-Zainal et al., 2016). Detailed information regarding these signatures is provided in Section 5.4.

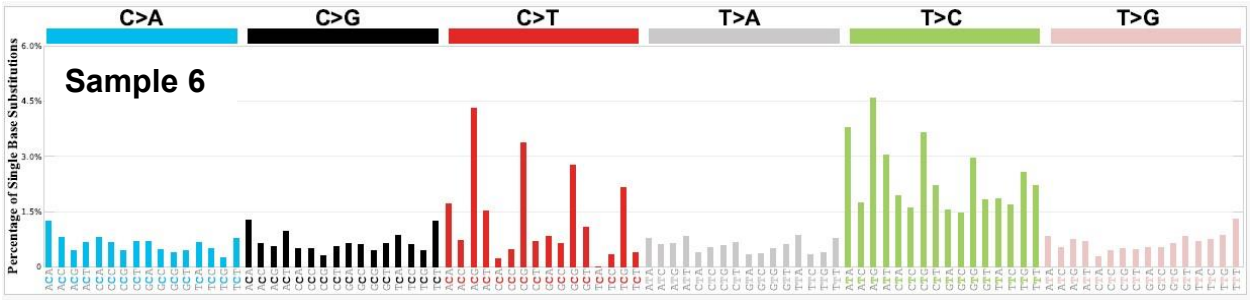

C>A
C>G
C>T
T>A
T>C
T>G
Sample 1
Percentage of Single Base Substitutions
6.2%
4.7%
3.1%
1.6%
0

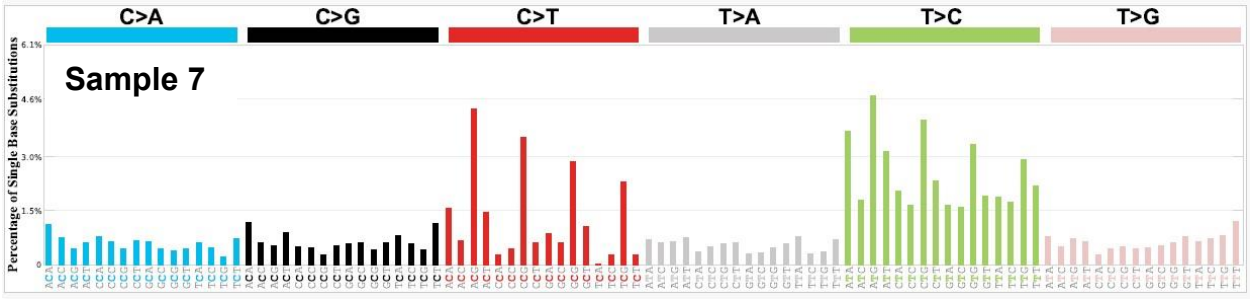

C>A
C>G
C>T
T>A
T>C
T>G
Sample 2
Percentage of Single Base Substitutions
6.0%
4.5%
3.0%
1.5%
0

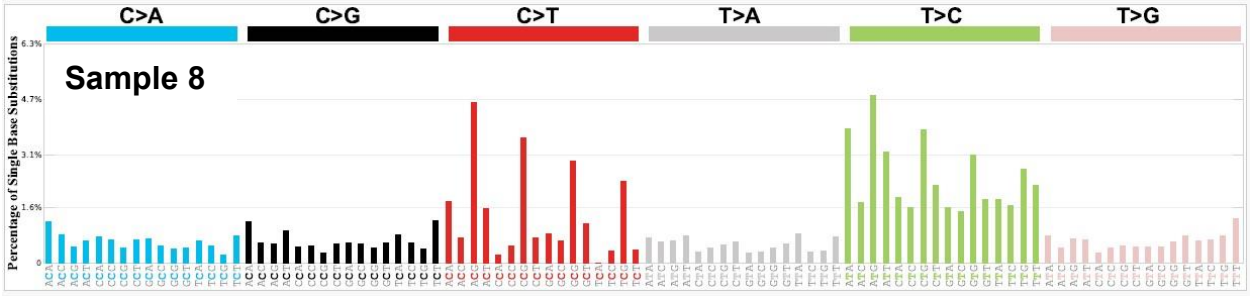

C>A
C>G
C>T
T>A
T>C
T>G
Sample 3
Percentage of Single Base Substitutions
6.1%
4.6%
3.1%
1.5%
0

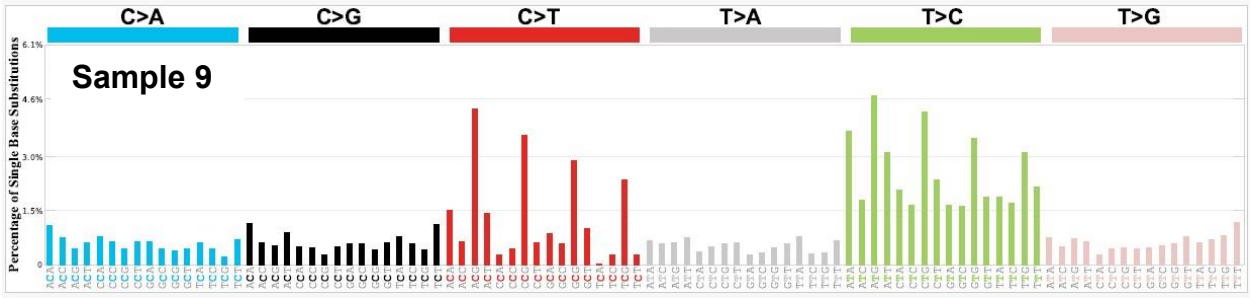

C>A
C>G
C>T
T>A
T>C
T>G
Sample 4
Percentage of Single Base Substitutions
6.0%
4.5%
3.0%
1.5%
0


C>A
C>G
C>T
T>A
T>C
T>G
Sample 5
Percentage of Single Base Substitutions
5.9%
4.4%
2.9%
1.5%
0


C>A
C>G
C>T
T>A
T>C
T>G
Sample 6
Percentage of Single Base Substitutions
6.0%
4.5%
3.0%
1.5%
0


C>A
C>G
C>T
T>A
T>C
T>G
Sample 7
Percentage of Single Base Substitutions
6.1%
4.6%
3.0%
1.5%
0


C>A
C>G
C>T
T>A
T>C
T>G
Sample 8
Percentage of Single Base Substitutions
6.3%
4.7%
3.1%
1.6%
0


C>A
C>G
C>T
T>A
T>C
T>G
Sample 9
Percentage of Single Base Substitutions
6.1%
4.6%
3.0%
1.5%
0

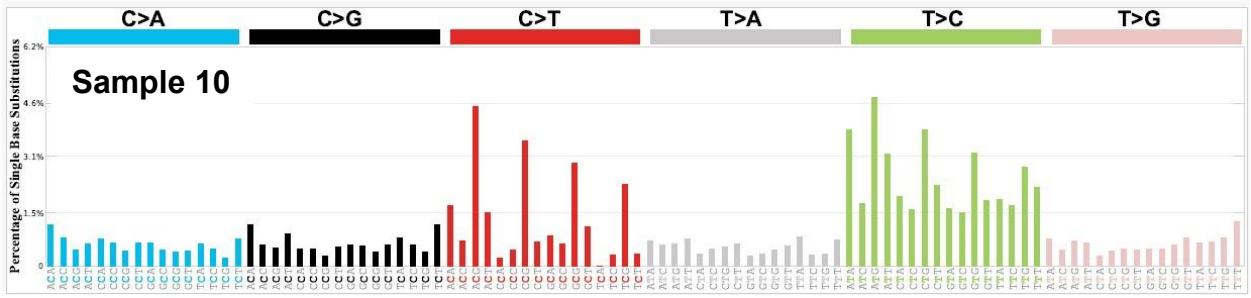

C>A
C>G
C>T
T>A
T>C
T>G
Sample 10
Percentage of Single Base Substitutions
6.2%
4.6%
3.1%
1.5%
0

Figure 9: Mutational profile of the ten breast tumor samples harboring the germline ATRIP c.1152_1155del variant. Each profile is created by counting the number of somatic mutations for each of the 96 single-base substitution types using tumor-only data. In each breast tumor mutational profile, the 96 single-base substitutions are shown on the x-axis and their percentages are shown on the y-axis.

Among all the ten breast cancer genomes, Signature 26 (SBS26) was the dominant signature, with a contribution ranging from 30.6% to 37.3%. This signature is associated with defects in the mismatch repair pathway and microsatellite instability (MSI), the presence of alternate-sized repetitive DNA sequences when mismatch repair is not working properly. Signatures 5 (SBS5) and 3 (SBS3) were the next observed signatures, with SBS5 having a slightly higher contribution. This signature is termed a "clock-like" signature as it positively correlates with aging. However, its underlying etiology is still unknown. SBS3 is associated with defects in the HRR pathway, an observation that is in accordance with the high HRD scores we reported for seven of these ten samples (Alexandrov, Nik-Zainal, Wedge, Aparicio, et al., 2013). Lastly, we observed Signature 1 (SBS1), which is due to the spontaneous deamination of 5-methylcytosine. Similar to SBS5, SBS1 is also known as a "clock-like" signature that follows an age-related dynamic (Ng & Chan, 2023). Table 9 depicts the inferred SBS signatures and their contributions in the ten breast cancer genomes using the set of twelve substitution signatures known to occur in breast cancers as reference signatures.

Table 9: Single-base substitution signatures in ten tumor DNA samples of Polish women with breast cancer harboring the germline ATRIP c.1152_1155del variant

Tumor Sample	SBS1 [a]	SBS3 [b]	SBS5 [c]	SBS26 [d]
Sample 1	16.5%	23.5%	27.3%	32.7%
Sample 2	15.7%	25.8%	26.9%	31.6%
Sample 3	18.2%	21.7%	23.2%	36.9%
Sample 4	15.5%	23.4%	23.8%	37.3%

Sample 5	14.7%	26.6%	25.3%	33.4%
Sample 6	15.3%	25.3%	28.8%	30.6%
Sample 7	15.6%	23.4%	26.4%	34.6%
Sample 8	16.8%	22.4%	28.7%	32.1%
Sample 9	15.7%	23%	25.4%	35.9%
Sample 10	16.2%	22.5%	28.1%	33.2%

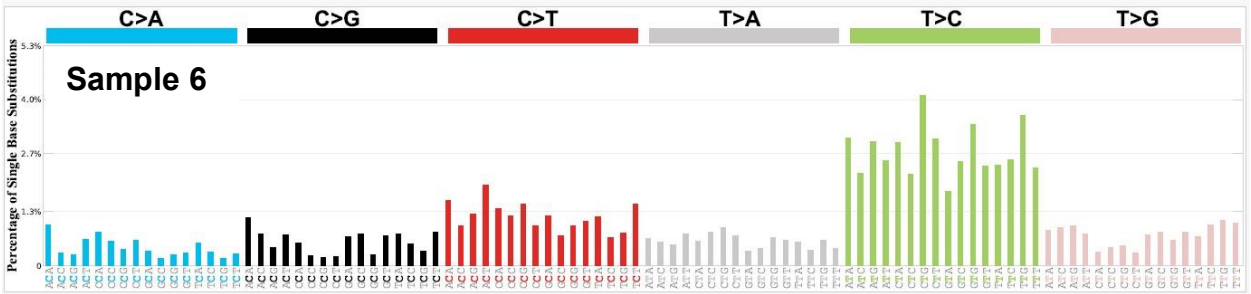
[a] Proposed etiology: Deamination of 5-methylcytosine

[b] Proposed etiology: Defective HRR pathway

[c] Proposed etiology: Unknown

[d] Proposed etiology: Defective DNA mismatch repair pathway

For a more precise mutational signature analysis, we asked our collaborators in Poland to provide us with matched germline DNA samples for the ten breast tumors that were tested. However, germline DNA samples were available for only two of the breast tumors (Sample 6 and Sample 10). We performed somatic variant calling using matched tumor-normal DNA for these two samples and subsequently conducted mutational signature analysis. Figure 10 depicts the mutational profile for these two samples using their matched tumor-normal data.



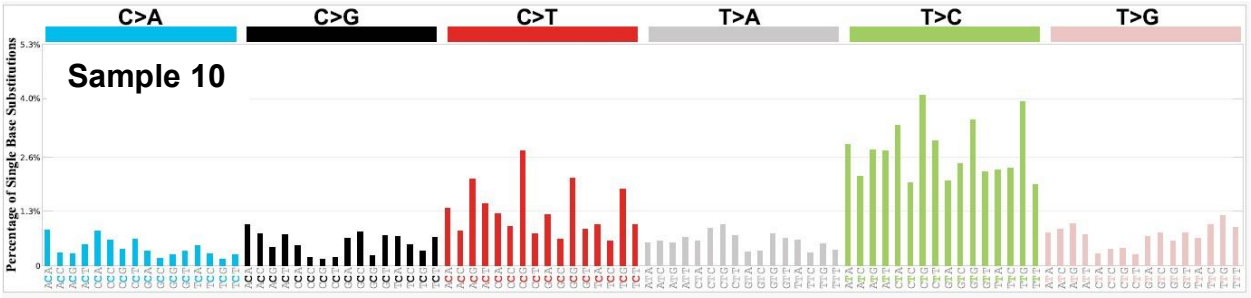


Figure 10: Mutational profile for two of the ten breast tumor samples harboring the germline ATRIP c.1152_1155del variant and with matched germline DNA. Each profile is created by counting the number of somatic mutations for each of the 96 single-base substitution types using matched tumor-normal data. In each breast tumor mutational profile, the 96 single-base substitutions are shown on the x-axis and their percentages are shown on the y-axis.

Considering the twelve substitution signatures that are known to happen in breast cancers as reference signatures, we observed the same contributing signatures as in our analysis of the ten tumor samples without matched germline DNA, except for a slight difference in the order of the first two contributing signatures. Table 10 depicts the inferred SBS signatures and their contributions in the two breast cancer genomes that were analyzed using their tumor-normal WGS data. SBS5 (unknown etiology) was the dominant signature among the two breast cancer genomes. SBS26 (defective mismatch repair pathway and MSI) and SBS3 (defective HRR pathway) were the second and third most contributing signatures, respectively, followed by SBS1 (deamination of 5-methylcytosine), which was the least contributing signature in both samples.

Table 10: Inferred single-base substitution signatures in the two tumor DNA samples of Polish women with breast cancer harboring the germline ATRIP c.1152_1155del variant and with matched germline DNA for somatic variant calling

Signature	Etiology	Sample 6	Sample 10

SBS1	Deamination of 5-methylcytosine	4.4%	7.4%
SBS3	Defective HRR pathway	23.1%	23.3%
SBS5	Unknown	40.2%	35.5%
SBS26	Defective DNA mismatch repair pathway	32.3%	33.8%

After observing a mismatch repair pathway-associated signature (SBS26), we performed MSI analysis using the WES data of the ten tumor samples to evaluate their status. We limited the analysis to MS sites with a minimum depth of coverage of 60×. Table 11 shows the number of MS sites observed for each sample, considering that level of coverage, as well as the number of unstable MS sites that showed different sizes in the tumor DNA compared to the Panel of Normal (PoN). Considering 20% as the cutoff for the presence of MSI, nine of the ten samples were classified as MSI-low, and only one sample was classified as MSI-high. We acknowledge that the use of WES data, coupled with our focus on regions with at least 60× depth of coverage, may result in conservative MSI estimates. By using WGS data, which covers a broader range of microsatellite loci across the genome, it is likely that MSI estimates would be higher due to the inclusion of unstable MS sites in noncoding regions and other areas more prone to instability. Nevertheless, the 20% cutoff is widely utilized in clinical practice and has proven effective in identifying MSI-high cases. Using this threshold, our laboratory have successfully detected MSI in all tested tumors harboring mutations in MMR genes, supporting the robustness of our approach despite the potential for underestimation.

Table 11: MSI status in the ten breast tumors harboring the ATRIP c.1152_1155del variant

Tumor Sample	Observed MS Sites	Unstable MS Sites	Percentage	MSI Status
Sample 1	79	14	17.72	MSI-Low

Sample 2	109	18	16.51	MSI-Low
Sample 3	118	25	21.19	MSI-High
Sample 4	110	19	17.27	MSI-Low
Sample 5	63	12	19.05	MSI-Low
Sample 6	73	11	15.07	MSI-Low
Sample 7	140	21	15	MSI-Low
Sample 8	148	19	12.84	MSI-Low
Sample 9	144	20	13.89	MSI-Low
Sample 10	157	22	14.01	MSI-Low

6.4 Discussion

DNA double-strand breaks (DSBs) are one of the most detrimental forms of DNA damage, resulting from the simultaneous breakage of both strands of the DNA molecule's phosphate backbones. DSBs can arise from internal physiological and metabolic

stressors within an organism or from external factors such as ionizing radiation and damage caused by chemicals (R.-X. Huang & Zhou, 2020). Failure to address this damage can result in genomic instability and, ultimately, tumorigenesis. Several pathways are involved in the repair of DSBs. The homologous recombination repair (HRR) pathway is the predominant mechanism employed to restore DNA to its original structure (Elbakry & Löbrich, 2021).

The HRR process involves complex sequential steps (Figure 11) that begin with the identification of a DSB by a protein complex known as MRN (MRE11-RAD50-NBN), which localizes BLM helicase and EXO1 onto the breaks to start 5'–3' double-stranded DNA resection. The single-stranded DNA (ssDNA) overhangs will then be coated by the replication protein A (RPA1-3) complex, which halts further DNA resection. The ataxia telangiectasia-mutated and Rad3-related (ATR) protein kinase loads onto the RPA-coated ssDNA overhangs through RPA-ATRIP interaction, which subsequently activates topoisomerase DNA II binding protein 1 (TopBP1). This activation switches on the G2/M-phase DNA damage checkpoint to arrest the cell cycle and stop further progression of cell division. Following cell cycle arrest, the radiation-sensitive protein 51 (RAD51) will displace RPA on the DNA overhangs. This is a major rate-limiting step in the HRR pathway mediated by several tumor suppressor proteins, including the BRCA1/BARD1 complex, BRCA2/PALB2 complex, and RAD51 paralogs. The steps mentioned so far are referred to as the presynaptic filament formation phase. The presynaptic strands undergo a homology search, and once the complementary homologous DNA strands are identified, the homologous strand invades the RAD51-coated DNA overhang of the presynaptic strand to form a synapse. This leads to DNA replication in the 5′→3′ direction, followed by DNA ligation-forming Holliday junctions. Finally, the Holliday junctions are resolved to complete the HRR pathway (Ali et al., 2024).

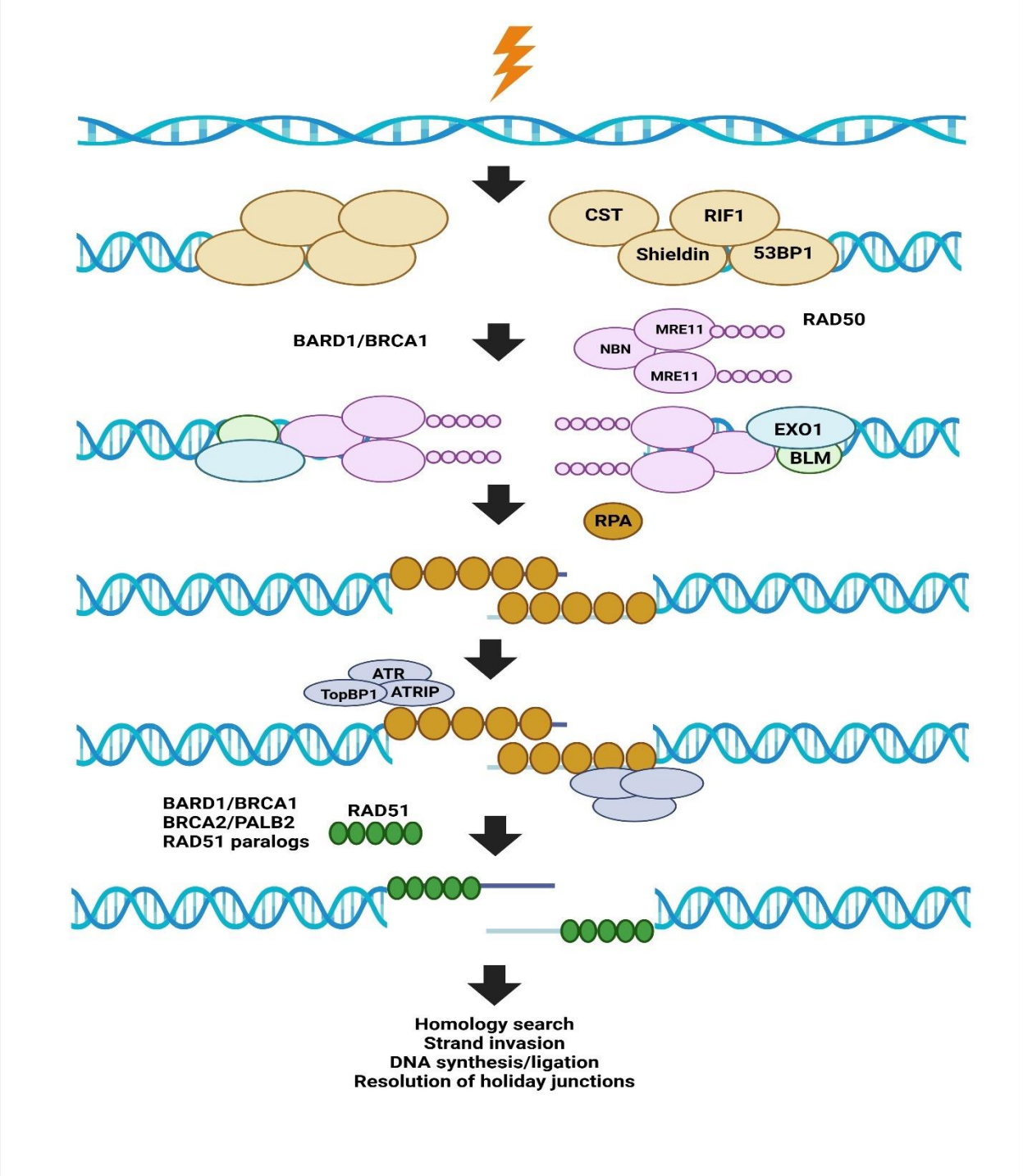


Figure 11: Sequential steps illustrating the homologous recombination repair (HRR) pathway. The process starts with the detection of double-stranded breaks by the MRN complex (MRE11-RAD50-NBN), followed by the recruitment of BLM helicase and EXO1 to initiate 5'-3' DNA resection. The resulting single-stranded DNA will be coated

by RPA, which stops any further resection. ATR is subsequently recruited to the site of DNA damage, leading to the activation of TopBP1 and initiating a cell cycle checkpoint to halt cell division temporarily. RAD51, with the assistance of tumor suppressor proteins like BRCA1, BRCA2, and PALB2, displaces RPA, resulting in the formation of a presynaptic filament. RAD51 executes a homology search, followed by strand invasion, DNA replication, and the creation and resolution of Holliday junctions, completing the HRR process.

Homologous recombination deficiency (HRD) refers to the cellular-level dysfunction of the HRR pathway that leads to a high degree of genomic instability and the accumulation of genetic alterations across the genome, thus playing an important role in the initiation and progression of cancer. As a hallmark of cancer, HRD is observed in several cancer types, most importantly ovarian, breast, pancreatic, and prostate cancers, and renders tumors more sensitive to treatment with platinum-based chemotherapies and PARP inhibitors through synthetic lethality (Rempel et al., 2022). When the HRR pathway is defective, DSBs rely on low-fidelity and error-prone alternative DNA damage repair mechanisms, such as non-homologous end joining (NHEJ), microhomology-mediated end joining (MMEJ), and single-strand annealing (SSA). These error-prone DNA repair pathways are likely to result in the insertion or deletion of nucleic acid sequences, abnormal duplication or deletion of genetic material, and the formation of cross-links between chromosomes, ultimately causing genomic instability and characteristic patterns of genomic scars (Ali et al., 2024; Radhakrishnan et al., 2014).

According to Knudson’s second-hit theory, tumor suppressor genes involved in DNA repair pathways such as HRR typically require not only a germline or somatic LoF mutation in a single allele but rather a complete inactivation of both mutated and wild-type alleles (biallelic inactivation) for complete loss of their function and subsequent carcinogenesis (Knudson, 2001). LOH of the wild-type allele (locus-specific LOH) is the most frequent second-hit mechanism, which involves the deletion of the functional allele of a mutated tumor suppressor gene. The size of this deletion could be variable, ranging from a single exon to the entire chromosome (Glodzik et al., 2023; Maxwell et al., 2017;

Nowak et al., 2004). To a lesser extent, biallelic inactivation can result from a compound heterozygous mutation (i.e., an inactivating somatic mutation in the remaining functional allele of a mutated gene), rearrangements, and epigenetic alterations such as DNA methylation or histone modifications that lead to the silencing of gene expression without changing the DNA sequence (Glodzik et al., 2023; Jia & Zhao, 2019). Also, a gene conversion may occur, in which the first inactivating mutation is transferred to the sequence of the intact allele. This usually happens through HRR and creates a homozygous state for the LoF mutation, rendering both copies of the gene non-functional (Takahashi & Innan, 2022).

BRCA1 and BRCA2 are the two major tumor suppressor genes involved in the HRR pathway. Biallelic inactivation through LOH at the BRCA1/2 locus has been reported in approximately 90% and 55% of BRCA1 and BRCA2 mutated breast and/or ovarian tumors, respectively (Maxwell et al., 2017). Evidence for loss of the wild-type allele via LOH and compromised HRR mechanisms has also been observed in breast/ovarian tumors harboring germline or somatic pathogenic mutations in tumor suppressor genes other than BRCA1/2, such as ATM, BRIP1, FANCM, PALB2, RAD51C, CDK12, and FANCD2 (Takamatsu et al., 2022).

We showed that four of the ten breast tumor samples harboring the germline ATRIP c.1152_1155del mutation received a second hit through LOH at the site of ATRIP. The observed 40% frequency of LOH in the ATRIP-mutated tumors in our study is similar to the LOH frequency reported in the breast cancer susceptibility genes PALB2 and BRCA2 (A. Li et al., 2019; Maxwell et al., 2017). We also showed that seven of the ten ATRIP-mutated tumors, including all four samples with confirmed LOH, had features of genomic instability and compromised HRR mechanisms, which will be discussed later in this section. Of note, two of the HRR-deficient samples (Table 8: Sample 6 and Sample 10) showed borderline evidence of LOH using their WGS data that was not confirmed in their WES data. As WGS is the gold standard for LOH detection (Rusch et al., 2018), we speculate that these two samples also have LOH, but because our WGS was low-depth, or maybe due to low tumor purity, the evidence for LOH in these two samples was borderline. Moreover, for one sample (Table 8: Sample 9) that showed features of

genomic instability and HRD in the absence of LOH, it is plausible that other second-hit mechanisms that we didn't examine may explain the observed HRD in this tumor.

Advancements in SNP microarray and NGS technologies have enabled researchers to quantify HRD status in a tumor genome (Sztupinszki et al., 2018). As explained in Section 6.1.3, HRD score is a numerical measure for assessing the extent of HRD in a tumor, calculated by the unweighted sum of three independent measures of genomic instability, including genomic LOH (loss of heterozygosity), TAI (telomeric allelic imbalance), and LST (large-scale state transitions) scores (Abkevich et al., 2012; Birkbak et al., 2012; Popova et al., 2012). To determine a threshold for HRD score, Telli et al. used a training set comprising 497 breast and 561 ovarian cancer patients (78 of the breast and 190 of the ovarian tumors were BRCA-deficient with a BRCA1/2 mutation and/or BRCA1 methylation). To achieve a minimum sensitivity of 95%, the threshold was set at 42, which was the 5th percentile of HRD scores in combined breast and ovarian BRCA-deficient tumors. Consequently, HRR deficiency was defined as having an HRD score ≥ 42, and HRR proficiency was defined as having an HRD score < 42 (Telli et al., 2016). Although the HRD score of 42 has been regarded as the canonical threshold for HRD status, recent studies have reported that any cutoff between 36 and 42 could lead to ~95% sensitivity for the identification of HRR-deficient tumors (Wen et al., 2022). Moreover, some clinical trials have proposed that lowering the HRD score threshold from 42 to 33 (1st percentile of HRD scores observed in BRCA1/2-mutant tumors) may even improve sensitivity for HRD characterization and detect a broader range of tumors that can benefit from treatment with PARP inhibitors (Stronach et al., 2018; Telli et al., 2018). In our study, we considered the canonical threshold of 42 to categorize tumor samples harboring the germline ATRIP c.1152_1155del mutation as HRR-deficient or HRR-proficient. Based on this threshold, seven of the ten ATRIP-mutated tumors were classified as HRR-deficient. Of note, none of the remaining three samples had HRD scores that could be regarded as HRR-deficient considering any of the exploratory thresholds discussed above (Table 8).

The direct association between loss of ATRIP function and resultant HRR deficiency in ATRIP-mutated tumor cells is indeed an area warranting further investigation. However, there are studies that provide indirect support for this association.

For example, conditional ATRIP knockout cell lines lacking the ATR binding domains of ATRIP at the C terminus show a remarkably reduced monoubiquitination of FANCD2 and FANCI and undetectable FANCI phosphorylation, suggesting that the ATR-ATRIP complex may trigger activation of the Fanconi Anemia DNA repair pathway (Beumer et al., 2015; Tomida et al., 2013). It has been shown that defects in this pathway may lead to compromised HRR and increased sensitivity of cancer cells to DNA-damaging agents such as PARP inhibitors (McCabe et al., 2006; Murai & Pommier, 2023). Another possible mechanism that can lead to HRD in ATRIP-mutated tumors might be the disruption of CHEK1 function, the protein activated by the ATRIP-ATR complex that has a key role in genomic maintenance via the HRR mechanism. Abrogation of CHEK1 function has been shown to inhibit HRR, leading to the accumulation of unrepaired DSBs and genomic instability (Gralewska et al., 2020; Sørensen et al., 2005). It is plausible that compromised localization of the ATR-ATRIP complex on damaged ssDNA may suppress ATR-mediated CHEK1 activation and subsequently lead to HRD. Lastly, ATR inhibitors have been shown to induce replication stress and genomic instability by causing DSBs at the site of stalled replication forks and inhibiting the repair of the broken DNA. ATR inhibitors render tumor cells sensitive to PARP inhibitors and sensitize BRCA-deficient cancer cells with acquired resistance to PARP inhibitors (Sørensen et al., 2005). Overall, although direct evidence may still be lacking, the existing literature suggests that there could be a connection between loss of ATRIP function and compromised HRR in tumor cells.

We further studied cancer etiology in the ten breast tumors with the germline ATRIP c.1152_1155del mutation by analyzing their mutational signature profiles. As previously stated in Section 5.1, each mutational signature is the impact of a particular mutational process on the cancer genome. Therefore, the mutational catalog of a cancer genome represents the main mutational mechanisms that have directed disease onset and progression. Deciphering these signatures provides valuable insight into the causes or origins of the tumor and can reveal underlying biological abnormalities that can be medically actionable (Alexandrov, Nik-Zainal, Wedge, Aparicio, et al., 2013).

One important consideration in performing tumor signature analysis is the existence of inter-tissue variations in mutational signatures (Alexandrov et al., 2020). Recently, Degasperi et al. performed signature analysis on 3,107 whole genome

sequenced samples of 21 organs that were obtained from the International Cancer Genome Consortium (ICGC) and reported inter-organ variability of signatures (Degasperi et al., 2020). To validate this observation, the researchers analyzed an independent WGS dataset obtained from the Hartwig consortium that comprised 3,096 metastatic cancers originating from 15 different tissue types. Their goal was to investigate whether the observed variations in mutational signatures between tissues are consistent and can be replicated. Organ-wise signature extractions were conducted based on the primary tissue of origin. For example, if a lung metastasis was caused by breast cancer, it was classified under the "breast" category. Comparing the signatures extracted from each organ in the metastatic cohort to the signatures extracted from the ICGC cohort, it was found that signatures from the same organ were the most similar to each other. The organ-specific signatures showed the highest similarities between the two cohorts for 14 out of the 15 organs that were compared. Pancreatic cancers seemed to deviate from the norm, potentially because of the large number of signatures present and the smaller number of samples in the metastatic cohort (75 compared to 313 in the ICGC cohort) (Degasperi et al., 2020). These findings supported the idea that there are tissue-specificities in the mutational signatures of diverse organs. In another seminal study, the same group of researchers conducted signature analysis on 12,222 WGS tumor-normal pairs from patients recruited via the UK National Health Service and compared the results with the two previously studied datasets (i.e., the ICGC and Hartwig consortium). Similar to their previous studies, they demonstrated that signatures from the same organ in different cohorts were more similar to each other than to those in a different tissue type, reassuring that mutational signatures in a specific organ are highly reproducible, have tissue specificities, and are discernible regardless of the sequencing platform or mutation calling algorithms (Degasperi et al., 2022). Therefore, the practical suggestion when analyzing a tumor genome is to utilize organ-specific signatures that correspond to the tissue of origin of the sample (Degasperi et al., 2020, 2022).

Considering the set of twelve single-base substitution signatures known to occur in breast cancers as the reference, tumor mutational signature analysis revealed four SBS signatures, including SBS5, SBS26, SBS3, and SBS1, in the ten breast tumors that we

tested using their tumor-only data. These signatures were further replicated in two of the ten samples, for which we repeated the analysis using their matched tumor-normal data.

Signature 5 (SBS5) appears in various cancer types, and its underlying etiology remains elusive. Using mutations in more than 10,000 cancer genomes obtained from 36 different cancer types, Alexandrov et al. reported that only two mutational signatures (i.e., Signatures 1 and 5) showed clock-like features correlating with the age of disease onset, albeit at different mutation rates (Alexandrov et al., 2020). It has been shown that there is a positive correlation between SBS5 and smoking history in lung cancer patients. This observation suggests that in lung cancer, Signature 5, which is characterized predominantly by C>T and T>C substitutions, may be due to carcinogens found in tobacco. However, SBS5 is not restricted to lung cancer and is also reported in other cancer types, most of which are not strongly associated with tobacco consumption (Alexandrov et al., 2015).

Recently, Perry et al. analyzed the WGS data of tumor/normal samples obtained from 37 breast cancer patients. Signature 5 was present in the majority (69%) of HR-positive patients and only in two (8%) patients with the triple-negative phenotype. In the latter group, Signature 5 was found at low rates, with less than 10% of the total mutational signature contributions (Perry et al., 2022). To validate this observation, the researchers analyzed a different cohort of 560 breast cancer patients and demonstrated a significant association between estrogen-receptor positive (ER+) status, mostly of luminal A subtype, and the dominance of Signature 5 (Perry et al., 2022). To achieve a deeper insight into the contribution of Signature 5 in HR-positive breast cancer, Perry et al. studied potential correlations between Signature 5 proportion and gene expression. The cluster that showed a positive correlation with Signature 5 contained the ESR1 gene and other genes associated with estrogen receptors. The ESR1 gene encodes an estrogen receptor that plays an important role in breast and endometrial cancers, as well as osteoporosis. Furthermore, STRING analysis revealed several protein-protein interaction clusters that showed a positive and high correlation with Signature 5. The largest cluster corresponded to an enriched estrogen signaling pathway. Limiting the analysis to ER+ patients revealed that higher Signature 5 ratios were significantly correlated with higher expression of ESR1. These observations suggest that Signature 5 could be a

characteristic feature of HR-positive breast cancers, exhibiting association with ER-related gene clusters and neoplastic processes regulated by estrogen signaling (Perry et al., 2022). Interestingly, eight of the ten breast tumors that we analyzed were ER+ tumors (Table 8).

There is a growing body of literature that suggests estrogen signaling can suppress the DNA damage response by preventing the activation of the ATRIP-ATR-TopBP1 complex. This endogenous inhibitory mechanism may predispose normal cells to tumorigenesis (Caldon, 2014; Perry et al., 2022). As discussed earlier, ATR is a key initiator of the DNA damage response that loads onto the RPA-coated ssDNA overhangs through interaction with ATRIP, subsequently activating TopBP1. This activation switches on the G2/M-phase DNA damage checkpoint to arrest the cell cycle and stop further progression of cell division to coordinate a replicative stress response (Ali et al., 2024). Studies have shown that Estradiol (E2) induces the phosphorylation of TopBP1 at the site of interaction with ATR. Phosphorylated TopBP1 is unable to bind to the ATR-ATRIP complex at the site of DNA damage. As a result, the activation of the cell cycle checkpoint is hindered, causing cells to continue through the cell cycle after DNA damage, ultimately leading to tumor formation (Jiménez-Salazar et al., 2021; Pedram et al., 2009).

Based on these observations, E2 can be regarded as an endogenous inhibitory factor that prevents the activation of ATR and downstream effectors, leading to a compromised DNA damage response in mammary epithelial cells. As discussed in Section 3.5, deleterious mutations in ATRIP may impair the stabilization of ATR on the RPA-coated ssDNA overhangs. It is possible that in the ATRIP-mutated breast tumors with the ER+ subtype (8/10 tumors that we tested), ATR-mediated checkpoint responses will be compromised further via the estrogen signaling pathway, giving grounds for observing Signature 5 in these tumors. The presence of Signature 5 in the two ATRIP-mutated tumors exhibiting the triple-negative phenotype might be due to other underlying mechanisms independent of the estrogen signaling pathway.

Signature 3 (SBS3) features a fairly uniform distribution of mutations across all 96 base substitution types and is strongly associated with germline and somatic BRCA1/2 mutations, as well as BRCA1 promoter hypermethylation in breast, pancreatic, ovarian,

and stomach cancers. Signature 3 was initially introduced by Nik-Zainal et al. after analyzing WGS data obtained from 21 primary breast tumors that included nine cases with germline mutations in the BRCA1/2 genes, all accompanied by loss of the wild-type allele as expected by Knudson second hit theory (Nik-Zainal, Alexandrov, et al., 2012). Experimentally, Signature 3 has been observed in $BRCA^{-/-}$ isogenic chicken DT40 cell lines, providing direct evidence of its association with HRR deficiency in BRCA-null tumors (Zámborszky et al., 2017). In accordance with these observations, a landmark study of 560 whole genomes and subsequent studies further confirmed the association of Signature 3 with deficiencies in the HRR machinery and reported the contribution of this signature in tumors with complete BRCA1/2 inactivation through germline or somatic point mutations in these genes coupled with loss of heterozygosity or hypermethylation of the BRCA1 promoter (Alexandrov et al., 2020; Degasperi et al., 2020, 2022; Nik-Zainal et al., 2016). Furthermore, Signature 3 has been reported in breast tumors with complete loss of function of other HRR components, including PALB2, RAD51D, RAD51C, and BARD1 (A. Li et al., 2019; Nones et al., 2019; Polak et al., 2017).

The existing literature suggests that there could be an indirect connection between loss of ATRIP function and compromised HRR. ATRIP knockout cell lines show a remarkably reduced monoubiquitination of FANCD2 and FANCI and undetectable FANCI phosphorylation. Research has shown that defects in the Fanconi Anemia DNA repair pathway may lead to compromised HRR and increased sensitivity of cancer cells to PARP inhibitors (Beumer et al., 2015; McCabe et al., 2006; Murai & Pommier, 2023). Another plausible mechanism leading to HRD in ATRIP-mutated tumors might be the disruption of CHEK1 function, as compromised localization of the ATR-ATRIP complex on damaged ssDNA may suppress ATR-mediated CHEK1 activation. Subsequently, abrogation of CHEK1 function has been shown to inhibit HRR, leading to unrepaired DSBs and genomic instability (Gralewska et al., 2020; Sørensen et al., 2005). Lastly, ATR inhibitors have been shown to induce replication stress and genomic instability by causing DSBs at the site of stalled replication forks and inhibiting the repair of the broken DNA. ATR inhibitors render tumor cells sensitive to PARP inhibitors and sensitize BRCA-deficient cancer cells with acquired resistance to PARP inhibitors (Sørensen et al., 2005).

Overall, we observed Signature 3 in all ten ATRIP-mutated breast tumors that we tested. Seven of the ten samples also showed HRD cellular phenotypes, defined as having an HRD score ≥ 42 (70% concordance). There were, however, three tumor samples that harbored Signature 3 but had HRD scores below 42 (Sample 1, Sample 2, and Sample 4). This discordance may reflect differences in the accuracy of the methods used in identifying HRD status (Batalini et al., 2022). While signature 3 has been suggested as a biomarker for HRD, it is unlikely to serve as a strong clinical biomarker on its own. Firstly, it probably lacks specificity as the vast majority of breast and ovarian tumors have some level of contribution from Signature 3 (Miller et al., 2020). Furthermore, determining an appropriate threshold will be challenging, as the relatively indistinct nature of signature 3 makes it highly sensitive to a decrease in the number of mutations that occurs in low tumor cellularity or when overshadowed by other mutational signatures (Davies et al., 2017). Since HRD leads to various types of genomic alterations, it is expected that an assay that utilizes a wide range of genomic information will provide higher levels of specificity for assigning HRD status.

Signature 26 (SBS26, primarily characterized by T>C mutations) is linked to impaired DNA mismatch repair (MMR) (Alexandrov, Nik-Zainal, Wedge, Aparicio, et al., 2013; Nik-Zainal, Alexandrov, et al., 2012). The post-replicative MMR pathway identifies and corrects mis-incorporated bases in DNA replication during the S phase, as well as insertions/deletions that occur during DNA replication and DNA recombination repair. The core of the MMR system involves two distinct protein complexes, MutS (MSH2/MSH6) and MutL (MLH1/PMS2), that identify nucleotide mismatches and insertion/deletion loops happening during DNA replication and coordinate their repair process (Iyer et al., 2015). Defective MMR mechanisms caused by mutations in MMR-related genes, most importantly MLH1, MSH2, MSH6, and PMS2, or through epigenetic silencing of these genes via promoter hypermethylation, increase the spontaneous mutation rate across the genome, leading to genomic instability (Ellis et al., 2023). Microsatellites (short tandem repeating sequences of 1–6 nucleotides) are particularly vulnerable to polymerase slippage and thus accumulate a significant number of mutations caused by compromised MMR. In MMR-deficient tumors, the length of these nucleotide motifs can not be

maintained during DNA replication, a characteristic known as microsatellite instability (MSI) (Giner-Calabuig et al., 2022).

Signature 26 has been reported in various cancer types, including colon, endometrial, breast, stomach, uterine, kidney, liver, prostate, esophagus, and pancreas malignancies (Alexandrov et al., 2020; Nik-Zainal et al., 2016). The association between this signature and defective MMR has been found and replicated in seminal studies involving thousands of whole genome sequenced tumor-normal matched pairs across different cancer types (Degasperi et al., 2020, 2022). The correlation between MMR deficiency and Signature 26 has also been validated in experimental studies of CRISPR-modified human stem cell colon organoids deficient in the MMR gene MLH1 (Drost et al., 2017). Moreover, using PMS2- and MLH1-knockout C. elegans, Meier et al. demonstrated that defects in these two MMR genes give rise to mutations that contribute to Signature 26, further strengthening evidence that this signature is linked to defective MMR pathways (Meier et al., 2018).

The role of MMR in DNA damage response pathways has also been studied extensively, revealing its interaction with the ATR-dependent checkpoint signaling pathway. Early findings by Wang and Qin suggested that MSH2 functions as a damage sensor for ATR-dependent signaling in response to DNA-methylating agents, facilitating the phosphorylation of CHEK1, the downstream target for ATR (Y. Wang & Qin, 2003). Subsequent research built upon this foundation. Fang et al. generated human cell lines in which a single ATR allele was inactivated on an MMR-deficient background. Monoallelic ATR gene targeting in MLH1-deficient HCT-116 colon carcinoma cells resulted in hypersensitivity to genotoxic stress, followed by dramatic increases in fragile site instability, chromosomal amplifications, and rearrangements. The $ATR^{+/-}$ HCT-116 cells also showed impaired activation of CHEK1. In a complementary study using a mouse model system, the same group of researchers demonstrated that mice harboring the $ATR^{+/-}/MLH1^{-/-}$ genotype were highly susceptible to embryonic lethality and early tumorigenesis. The researchers concluded that MMR proteins and ATR have a functional interaction during cellular response to genotoxic stress (Fang et al., 2004). Further research corroborates this interaction, emphasizing the interconnected roles of the MMR

pathway and ATR-mediated DNA damage responses (A. Liu et al., 2008; Yan et al., 2014).

Despite extensive research into MMR and ATR interactions, no study has yet investigated the potential functional interaction between MMR machinery and ATRIP. In our analysis of mutational signatures, all ten breast tumors with the germline ATRIP c.1152_1155del variant showed imprints of MMR deficiency on their genomes, and all of them were microsatellite instable (Table 11). These findings highlight a potential novel link between ATRIP mutations and MSI, mediated through defective MMR pathways.

Moreover, it is plausible that the HRD observed in ATRIP-mutated breast tumors may contribute to MSI and subsequently to the development of Signature 26. First, in the absence of a functional HRR pathway, cells increasingly rely on error-prone DNA repair mechanisms, such as NHEJ, which introduce additional errors and drive DNA damage accumulation across the genome. Microsatellite regions are particularly vulnerable to such errors and become hotspots for mutations. Second, HRD induces replication stress, which causes the stalling of the replication machinery at sites of DNA damage. Due to their repetitive nature, microsatellites are highly susceptible to replication slippage under such stress. In the absence of efficient HRR to resolve stalled replication forks, these replication errors remain uncorrected, resulting in MSI. Finally, HRD may indirectly exacerbate MSI by increasing the overall mutation burden across the genome, including mutations occurring within the MMR pathway genes themselves (Scully et al., 2019).

Signature 1 (SBS1) arises from the spontaneous deamination of 5-methylcytosine, leading to C:G>T:A transitions that occur predominantly within the NpCpG trinucleotide context (Alexandrov, Nik-Zainal, Wedge, Aparicio, et al., 2013). The frequent occurrence of this substitution throughout evolution is believed to be the main reason for the reduction in the number of methylated CpG sites observed in the human genome (Lutsenko & Bhagwat, 1999). Although the total number of these sites has decreased, they still remain one of the most mutagenic sequence motifs in the genome. Signature 1 is one of the most frequent mutational signatures that has been observed in at least 25 different cancer types, such as acute myeloid leukemia, breast cancer, glioma, head and neck cancers, malignant melanoma, kidney, and ovarian cancers. Interestingly, there is a correlation between the number of mutations linked to this signature and the age of patients,

regardless of their sex, when diagnosed with cancer (Alexandrov, Nik-Zainal, Wedge, Aparicio, et al., 2013).

Moreover, it has been shown that FFPE-induced artifacts share a similar pattern with Signature 1 (Q. Guo et al., 2022). FFPE preserves tumor tissue morphology and allows for immunohistochemical analysis, which is useful for clinical diagnosis. However, in comparison to fresh frozen material, genetic analysis of DNA extracted from FFPE blocks is challenging due to the negative effects of formalin fixation on DNA quality and quantity (Mathieson & Thomas, 2019). During the fixation process, a solution of buffered formalin containing 4% formaldehyde is used to penetrate into the biospecimen. This causes the formation of chemical bonds, known as cross-links, between large molecules within the cells, such as DNA-DNA, DNA-RNA, and DNA-protein. These cross-links impede the progress of DNA polymerases during the amplification of a library (Do & Dobrovic, 2015; Kennedy-Darling & Smith, 2014). Consequently, the diversity and the number of templates that can be amplified by PCR from FFPE-extracted DNA are substantially reduced (Bhagwate et al., 2019). Moreover, formalin induces the process of hydrolytic deamination of 5-methylcytosine (exclusively present in CpG dinucleotides), converting them into uracil. This leads to the formation of U:G mismatches, where during PCR amplification, DNA polymerase inserts adenine opposite to uracil, producing artifactual C:G>T:A substitutions in sequencing data (Do & Dobrovic, 2015). In order to reduce deamination artifacts, certain FFPE sequencing library preparations, including our protocol, incorporate a repair treatment that involves the addition of uracil DNA glycosylase (UDG) to eliminate uracil bases before the amplification process (Prentice et al., 2018). Nevertheless, formalin-induced deamination of 5-methylcytosine can also be converted directly to thymine rather than uracil. This direct conversion can not be corrected by UDG treatment, leading to an FFPE-induced artifactual pattern that highly resembles Signature 1 (Bhagwate et al., 2019; Q. Guo et al., 2022).

Chapter 7 – Concluding Summary, Limitations, and Future Directions

7.1 Concluding Summary

Breast cancer is one of the most prevalent malignancies among women worldwide. As genetic susceptibility has an important role in the etiology of this disease, increasing efforts have been placed into the identification of hereditary components that underlie breast cancer susceptibility. Deleterious mutations in known breast cancer susceptibility genes account for roughly half of the familial breast cancer cases, and the genetic etiology for the remaining women with high familial clustering of breast cancer remains a major knowledge gap.

The aim of this study was to contribute to this growing body of literature as we sought to investigate further the role of four novel gene candidates in susceptibility to breast cancer. These genes included DCLRE1C, DDX19B, and XRCC3 that were previously identified by our research team by performing WES on a cohort of highly familial breast cancer Bahamian patients, as well as ATRIP, which was identified through conducting WES on a cohort of highly familial breast cancer patients and healthy individuals from the founder population of Poland, followed by validation.

During our efforts, we utilized a case-control approach to investigate each candidate gene's association with breast cancer susceptibility. We conducted targeted full-gene sequencing of the four candidate genes on the germline DNA of 1,021 familial breast cancer patients in Ontario, 1,390 familial French-Canadian breast cancer patients from Quebec, and 421 familial breast cancer patients from Iran. We also sequenced the entire coding regions of the four candidate genes on the germline DNA of 930 healthy women from Ontario to serve as the control group for association analysis with the Ontario breast cancer patients. We used the sequence data of 800 healthy Iranian individuals reported in the Iranome database and 450 healthy individuals from the Gen3G study as controls for comparison with the Iranian and French-Canadian breast cancer patients,

respectively. Furthermore, we analyzed the UK Biobank whole exome sequence data of 15,643 genetically inferred Caucasian breast cancer cases and 157,943 ethnicity-matched cancer-free controls. Through performing gene-based association analyses, we showed that of the four candidate genes under study, only ATRIP LoF variants were associated with a significantly higher risk of breast cancer among the British population (OR = 3.28, 95% CI = 1.76–6.14, $p < 0.001$). Although LoF variants in ATRIP seem relatively rare (1 in every 1,000-3,000 of the general population), these mutations are associated with a significant risk of breast cancer, suggesting ATRIP as a moderately penetrant gene for hereditary breast cancer.

As an exploratory aim, we wanted to study if inherited deleterious mutations in the four gene candidates may increase the risk of developing ovarian cancer. For this purpose, we looked for deleterious mutations in our candidate genes among 1,488 unselected cases of invasive epithelial ovarian cancer diagnosed in Ontario. Screening of the entire coding exons of the four candidate genes under study revealed four carriers of ATRIP LoF variants among these patients, with one patient harboring the ATRIP c.1152_1155delTGGA (rs753154444) variant previously identified by our research team among the Polish population. No carriers of LoF variants in the other three genes were identified among the ovarian cancer patients.

We performed additional analyses to further confirm the role of ATRIP in breast cancer development. Specifically, we did LOH analysis by performing WGS on the tumor DNA of ten available FFPE samples of the ATRIP c.1152_1155del mutation carriers, followed by WES with a higher depth of coverage for samples with borderline evidence for LOH. Sequence data showed LOH in at least 40% of the tumor samples. We also showed HRD, as defined by having an HRD score ≥ 42, in seven of the ten tumors, including all four samples with confirmed LOH. Two of the samples with high HRD scores showed borderline evidence of LOH using their WGS data, which was not confirmed after performing WES.

Lastly, we looked for patterns of somatic point mutations in the ten available tumors of the ATRIP c.1152_1155del mutation carriers to achieve a deeper perception of the

biological mechanisms that have gone awry in the course of cancer development in these tumors. To achieve this goal, we called somatic point mutations using tumor-only WGS data for each of the ten FFPE tumor samples that harbored the germline ATRIP c.1152_1155del mutation. We then created a mutational profile comprising all 96 possible single-base substitution types for each tumor DNA sample. Signatures of single-base substitutions were inferred from each cancer genome and compared with the already known and validated signatures listed as references in the COSMIC database. For two of the tumor samples with available germline DNA, we repeated mutational signature analysis using their tumor-normal WGS data. In both scenarios, the contributing signatures in the studied breast cancer genomes included Signature 5 (unknown etiology), Signature 3 (defective HRR pathway), Signature 26 (defective mismatch repair pathway), and Signature 1 (deamination of 5-methylcytosine). This observation gave us the insight that these genomes were confronted with defects in the HRR and mismatch repair pathways at some point during their cells' lifetime.

7.2 Limitations of the Study

There are several limitations to this study. First, targeted sequencing of the four candidate genes among the Canadian and Iranian study cohorts did not yield a sufficient amount of genetic data to perform PCA to adjust for confounding due to unrecognized ancestry differences (i.e., population stratification). Therefore, we could only use information on self-reported ethnicity to calculate adjusted ORs and report our findings observed among these populations.

The second limitation concerning the Canadian and Iranian study cohorts is the potential existence of unrecognized kinship (i.e., cryptic relatedness) among study subjects, which we were not able to examine due to the inadequate amount of sequenced data. In the presence of cryptic relatedness, genotypes are not drawn independently from the population frequencies, and tests of association may have inflated type I or II error rates. This may lead to spurious associations, which is a major concern mostly in GWAS of common variants and in studies involving founder populations dealing with a founder mutation. However, in gene-based association studies of rare variants in mixed

populations such as the Ontario and Iranian populations, it is very unlikely that any two case or control subjects carry an identical rare risk allele in the same susceptibility gene. If cryptic relatedness is present among non-carrier cases or controls in such cohorts, this may lead to inflated false negative or false positive findings, respectively.

Another limitation of our genetic association study within the UK population is that we did not adjust for the top principal components at the data analysis step, even though we matched our cases and controls based on genetically inferred ethnicity using their PCA results. This could result in residual confounding caused by a subtle population structure that may not have been adequately controlled for through matching alone. By adjusting for the top PCs, the risk of residual confounding could have been minimized by controlling for any remaining population stratification, ensuring more robust and reliable findings.

We understand that a study's statistical power increases with sample size. The necessary sample size depends on the frequency of the risk allele (or the combined allele frequency of a class of alleles in a gene) as well as the allele's relative risk of disease. In order to achieve definitive evidence for the association of a novel moderately to highly penetrant susceptibility gene with breast cancer risk, thousands of cases and controls need to be recruited. Although we used all the available samples in our biobank, we realized that the small cohort size in our association analyses among the Canadian and Iranian populations weakened the study's power to detect associations among these populations.

Moreover, we realize that a proportion of the Gen3G study participants used as the control group for comparison with the French-Canadian breast cancer patients have hyperglycemia, which may raise concern as a potential confounder in the association analysis. However, the four candidate genes under study have known biological functions, and to the best of our knowledge, none of these genes may be associated with impaired glucose tolerance or dysglycaemia. However, we can not completely rule out the existence of such an association.

Lastly, we acknowledge that FFPE tumor samples present certain limitations compared to fresh frozen material, potentially affecting the results of our tumor DNA

analyses. DNA degradation that happens during the fixation process leads to the fragmentation and cross-linking of DNA molecules, which can compromise the quality and integrity of genetic material and impact the accuracy of downstream analyses, leading to a high rate of false positive somatic variant calls. This limitation is even more serious in the absence of a matched germline DNA sample, as was the case for eight of our tumor samples.

7.3 Future Directions

The findings of different lines of studies that we have performed propose that ATRIP could potentially serve as a breast cancer gene candidate. Given that ATRIP's role in susceptibility to breast cancer is further replicated among other populations, it will contribute to reducing the unexplained component of the hereditary breast cancer landscape and improving our knowledge of hereditary breast cancer genetics. Identifying novel breast cancer genetic determinants will ultimately empower us to better detect individuals with a higher risk of developing breast cancer through genetic screening measures, allowing for tailored preventive strategies for carriers. Moreover, identification of a new breast cancer gene candidate may lead to the discovery of novel underlying biological mechanisms related to the gene's function. Understanding these carcinogenesis pathways may consequently enable the development of novel drugs and individualized treatments for germline mutation carriers and even non-carriers that might harbor somatic mutations in the susceptibility gene.

In parallel to our studies, our collaborators at Laval University Cancer Research Center (Québec, Canada) performed in vitro functional assays that showed the truncated ATRIP protein resulted from ATRIP c.1152_1155del failed to perform its normal function to prevent replicative stress (Cybulski et al., 2023). Additionally, immunohistochemistry analysis of the ATRIP protein in breast tumor cells of the mutation carriers showed that ATRIP was weakly expressed compared to the tumor cells of the non-carriers. These observations provide further evidence supporting the pathogenicity of the ATRIP LoF variants. Other possible future directions for this study could include:

Validation studies: Perform additional association studies on larger and more heterogeneous populations to confirm the correlation between ATRIP LoF variants and susceptibility to breast cancer among other populations. This may involve collaborating with other research groups or utilizing larger population-based datasets.

Ovarian cancer studies: Conduct ovarian cancer association studies to investigate the relationship between ATRIP mutations and the risk of developing ovarian cancer in larger cohorts. An effective strategy may involve performing targeted sequencing on ovarian cancer patients from various populations to ascertain the frequency of ATRIP mutations and their association with susceptibility to ovarian cancer.

Clinical significance and therapeutic targets: Analyze the clinical importance of ATRIP mutations in terms of prognosis and response to treatment. Also, investigate the possibility of utilizing targeted therapies or personalized treatment strategies for breast cancer patients who have deleterious ATRIP mutations. This may entail evaluating the sensitivity of ATRIP-mutated tumors to currently available therapies that specifically target DNA repair pathways, such as PARP inhibitors or platinum-based chemotherapies that have been used efficiently in the treatment of HRR-deficient tumors.

In summary, the above-mentioned future directions can help provide a clearer understanding of the involvement of ATRIP in breast cancer susceptibility, enhance the evaluation of risk and management of patients, and potentially discover novel targets for personalized treatment strategies.

References

Abkevich, V., Timms, K. M., Hennessy, B. T., Potter, J., Carey, M. S., Meyer, L. A., Smith-McCune, K., Broaddus, R., Lu, K. H., Chen, J., Tran, T. V., Williams, D., Iliev, D., Jammulapati, S., FitzGerald, L. M., Krivak, T., DeLoia, J. A., Gutin, A., Mills, G. B., & Lanchbury, J. S. (2012). Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. *British Journal of Cancer*, *107*(10), 1776–1782. https://doi.org/10.1038/bjc.2012.451

Akbari, M., Donenberg, T., Lunn, J., Curling, D., Turnquest, T., Krill-Jackson, E., Zhang, S., Narod, S., & Hurley, J. (2014). The spectrum of BRCA1 and BRCA2 mutations in breast cancer patients in the Bahamas. *Clinical Genetics*, *85*(1), 64–67. https://doi.org/10.1111/cge.12132

Alexandrov, L. B., Jones, P. H., Wedge, D. C., Sale, J. E., Campbell, P. J., Nik-Zainal, S., & Stratton, M. R. (2015). Clock-like mutational processes in human somatic cells. *Nature Genetics*, *47*(12), 1402–1407. https://doi.org/10.1038/ng.3441

Alexandrov, L. B., Kim, J., Haradhvala, N. J., Huang, M. N., Tian Ng, A. W., Wu, Y., Boot, A., Covington, K. R., Gordenin, D. A., Bergstrom, E. N., Islam, S. M. A., Lopez-Bigas, N., Klimczak, L. J., McPherson, J. R., Morganella, S., Sabarinathan, R., Wheeler, D. A., Mustonen, V., Getz, G., … Stratton, M. R. (2020). The repertoire of mutational signatures in human cancer. *Nature*, *578*(7793), Article 7793. https://doi.org/10.1038/s41586-020-1943-3

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A. J. R., Behjati, S., Biankin, A. V., Bignell, G. R., Bolli, N., Borg, A., Børresen-Dale, A.-L., Boyault, S., Burkhardt, B., Butler, A. P., Caldas, C., Davies, H. R., Desmedt, C., Eils, R., Eyfjörd, J. E., Foekens, J.

A., … Stratton, M. R. (2013). Signatures of mutational processes in human cancer. *Nature*, *500*(7463), 415–421. https://doi.org/10.1038/nature12477

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J., & Stratton, M. R. (2013). Deciphering signatures of mutational processes operative in human cancer. *Cell Reports*, *3*(1), 246–259. https://doi.org/10.1016/j.celrep.2012.12.008

Alghamdi, J., & Padmanabhan, S. (2014). Fundamentals of Complex Trait Genetics and Association Studies. In *Handbook of Pharmacogenomics and Stratified Medicine* (pp. 235–257). Elsevier. https://doi.org/10.1016/B978-0-12-386882-4.00012-8

Ali, U., Vungarala, S., & Tiriveedhi, V. (2024). Genomic Features of Homologous Recombination Deficiency in Breast Cancer: Impact on Testing and Immunotherapy. *Genes*, *15*(2), Article 2. https://doi.org/10.3390/genes15020162

Allison, K. H., Hammond, M. E. H., Dowsett, M., McKernin, S. E., Carey, L. A., Fitzgibbons, P. L., Hayes, D. F., Lakhani, S. R., Chavez-MacGregor, M., Perlmutter, J., Perou, C. M., Regan, M. M., Rimm, D. L., Symmans, W. F., Torlakovic, E. E., Varella, L., Viale, G., Weisberg, T. F., McShane, L. M., & Wolff, A. C. (2020). Estrogen and Progesterone Receptor Testing in Breast Cancer: ASCO/CAP Guideline Update. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, *38*(12), 1346–1366. https://doi.org/10.1200/JCO.19.02309

Allman, R., Dite, G. S., Hopper, J. L., Gordon, O., Starlard-Davenport, A., Chlebowski, R., & Kooperberg, C. (2015). SNPs and breast cancer risk prediction for African American and Hispanic women. *Breast Cancer Research and Treatment*, *154*(3), 583–589. https://doi.org/10.1007/s10549-015-3641-7

Aloraifi, F., Boland, M. R., Green, A. J., & Geraghty, J. G. (2015). Gene analysis techniques and susceptibility gene discovery in non-BRCA1/BRCA2 familial breast cancer. *Surgical Oncology*, *24*(2), 100–109. https://doi.org/10.1016/j.suronc.2015.04.003

Andreassen, P. R., Seo, J., Wiek, C., & Hanenberg, H. (2021). Understanding BRCA2 Function as a Tumor Suppressor Based on Domain-Specific Activities in DNA Damage Responses. *Genes*, *12*(7), 1034. https://doi.org/10.3390/genes12071034

Antoniou, A. C., Casadei, S., Heikkinen, T., Barrowdale, D., Pylkäs, K., Roberts, J., Lee, A., Subramanian, D., De Leeneer, K., Fostira, F., Tomiak, E., Neuhausen, S. L., Teo, Z. L., Khan, S., Aittomäki, K., Moilanen, J. S., Turnbull, C., Seal, S., Mannermaa, A., … Tischkowitz, M. (2014). Breast-Cancer Risk in Families with Mutations in *PALB2*. *New England Journal of Medicine*, *371*(6), 497–506. https://doi.org/10.1056/NEJMoa1400382

Apostolou, P., & Fostira, F. (2013). Hereditary Breast Cancer: The Era of New Susceptibility Genes. *BioMed Research International*, *2013*, 1–11. https://doi.org/10.1155/2013/747318

Ball, H. L., Myers, J. S., & Cortez, D. (2005). ATRIP Binding to Replication Protein A-Single-stranded DNA Promotes ATR–ATRIP Localization but Is Dispensable for Chk1 Phosphorylation. *Molecular Biology of the Cell*, *16*(5), 2372–2381. https://doi.org/10.1091/mbc.e04-11-1006

Barrett, M. T., Lenkiewicz, E., Malasi, S., Basu, A., Yearley, J. H., Annamalai, L., McCullough, A. E., Kosiorek, H. E., Narang, P., Wilson Sayres, M. A., Chen, M., Anderson, K. S., & Pockaj, B. A. (2018). The association of genomic lesions and PD-1/PD-L1 expression in resected triple-negative breast cancers. *Breast Cancer Research: BCR*, *20*(1), 71. https://doi.org/10.1186/s13058-018-1004-0

Bartek, J., Iggo, R., Gannon, J., & Lane, D. P. (1990). Genetic and immunochemical analysis of mutant p53 in human breast cancer cell lines. *Oncogene*, *5*(6), 893–899.

Bashir, H., Majid, S., Hamid, R., Farooq, R., Wani, H. A., Shoib, S., & Bhat, A. A. (2015). Polymorphism of the XRCC3 gene and risk of gastric cancer in a Kashmiri population: A case-control study. *European Journal of Cancer Prevention: The Official Journal of the European Cancer Prevention Organisation (ECP)*, *24*(3), 167–175. https://doi.org/10.1097/CEJ.0000000000000115

Batalini, F., Gulhan, D. C., Mao, V., Tran, A., Polak, M., Xiong, N., Tayob, N., Tung, N. M., Winer, E. P., Mayer, E. L., Knappskog, S., Lønning, P. E., Matulonis, U. A., Konstantinopoulos, P. A., Solit, D. B., Won, H., Eikesdal, H. P., Park, P. J., & Wulf, G. M. (2022). Mutational Signature 3 Detected from Clinical Panel Sequencing is Associated with Responses to Olaparib in Breast and Ovarian Cancers. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, *28*(21), 4714–4723. https://doi.org/10.1158/1078-0432.CCR-22-0749

Beatson, G. T. (1896). On the Treatment of Inoperable Cases of Carcinoma of the Mamma: Suggestions for a New Method of Treatment, with Illustrative Cases. *Transactions. Medico-Chirurgical Society of Edinburgh*, *15*, 153–179.

Bell, D. W., Varley, J. M., Szydlo, T. E., Kang, D. H., Wahrer, D. C., Shannon, K. E., Lubratovich, M., Verselis, S. J., Isselbacher, K. J., Fraumeni, J. F., Birch, J. M., Li, F. P., Garber, J. E., & Haber, D. A. (1999). Heterozygous germ line hCHK2 mutations in Li-Fraumeni syndrome. *Science (New York, N.Y.)*, *286*(5449), 2528–2531. https://doi.org/10.1126/science.286.5449.2528

Beral, V. & Million Women Study Collaborators. (2003). Breast cancer and hormone-replacement therapy in the Million Women Study. *Lancet (London, England)*, *362*(9382), 419–427. https://doi.org/10.1016/s0140-6736(03)14065-2

Beral, V., Peto, R., Pirie, K., & Reeves, G. (2019). Menopausal hormone therapy and 20-year breast cancer mortality. *Lancet (London, England)*, *394*(10204), 1139. https://doi.org/10.1016/S0140-6736(19)32033-1

Bergkvist, L., Adami, H. O., Persson, I., Hoover, R., & Schairer, C. (1989). The risk of breast cancer after estrogen and estrogen-progestin replacement. *The New England Journal of Medicine*, *321*(5), 293–297. https://doi.org/10.1056/NEJM198908033210505

Beumer, J. H., Fu, K. Y., Anyang, B. N., Siegfried, J. M., & Bakkenist, C. J. (2015). Functional analyses of ATM, ATR and Fanconi anemia proteins in lung carcinoma: ATM, ATR and FA in lung carcinoma. *BMC Cancer*, *15*(1), 649. https://doi.org/10.1186/s12885-015-1649-3

Bhagwate, A. V., Liu, Y., Winham, S. J., McDonough, S. J., Stallings-Mann, M. L., Heinzen, E. P., Davila, J. I., Vierkant, R. A., Hoskin, T. L., Frost, M., Carter, J. M., Radisky, D. C., Cunningham, J. M., Degnim, A. C., & Wang, C. (2019). Bioinformatics and DNA-extraction strategies to reliably detect genetic variants from FFPE breast tissue samples. *BMC Genomics*, *20*(1), 689. https://doi.org/10.1186/s12864-019-6056-8

Birkbak, N. J., Wang, Z. C., Kim, J.-Y., Eklund, A. C., Li, Q., Tian, R., Bowman-Colin, C., Li, Y., Greene-Colozzi, A., Iglehart, J. D., Tung, N., Ryan, P. D., Garber, J. E., Silver, D. P., Szallasi, Z., & Richardson, A. L. (2012). Telomeric allelic imbalance indicates defective DNA repair and sensitivity to DNA-damaging agents. *Cancer Discovery*, *2*(4), 366–375. https://doi.org/10.1158/2159-8290.CD-11-0206

Brakeleer, S. D., Grève, J. D., Loris, R., Janin, N., Lissens, W., Sermijn, E., & Teugels, E. (2010). Cancer predisposing missense and protein truncating BARD1 mutations in non-BRCA1 or BRCA2 breast cancer families. *Human Mutation*, *31*(3), E1175–E1185. https://doi.org/10.1002/humu.21200

Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *74*(3), 229–263. https://doi.org/10.3322/caac.21834

Breast Cancer Association Consortium, Dorling, L., Carvalho, S., Allen, J., González-Neira, A., Luccarini, C., Wahlström, C., Pooley, K. A., Parsons, M. T., Fortuno, C., Wang, Q., Bolla, M. K., Dennis, J., Keeman, R., Alonso, M. R., Álvarez, N., Herraez, B., Fernandez, V., Núñez-Torres, R., … Easton, D. F. (2021). Breast Cancer Risk Genes—Association Analysis in More than 113,000 Women. *New England Journal of Medicine*, *384*(5), 428–439. https://doi.org/10.1056/NEJMoa1913948

Brenner, D. R., Gillis, J., Demers, A. A., Ellison, L. F., Billette, J.-M., Zhang, S. X., Liu, J. L., Woods, R. R., Finley, C., Ma, N. F., Saint-Jacques, N., Shack, L., & Turner, D. (2024). Projected estimates of cancer in Canada in 2024. *Central Nervous System*, *196*(18).

Brewer, H. R., Jones, M. E., Schoemaker, M. J., Ashworth, A., & Swerdlow, A. J. (2017). Family history and risk of breast cancer: An analysis accounting for family structure. *Breast Cancer Research and Treatment*, *165*(1), 193–200. https://doi.org/10.1007/s10549-017-4325-2

Bürger, H., de Boer, M., van Diest, P. J., & Korsching, E. (2013). Chromosome 16q loss—A genetic key to the understanding of breast carcinogenesis. *Histology and Histopathology*, *28*(3), 311–320. https://doi.org/10.14670/HH-28.311

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., & 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

Caldon, C. E. (2014). Estrogen Signaling and the DNA Damage Response in Hormone Dependent Breast Cancers. *Frontiers in Oncology*, *4*, 106. https://doi.org/10.3389/fonc.2014.00106

Canadian Cancer Statistics Advisory Committee. (2018). Canadian Cancer Statistics 2018. cancer.ca/Canadian-Cancer-Statistics-2018-EN

Casadei, S., Norquist, B. M., Walsh, T., Stray, S., Mandell, J. B., Lee, M. K., Stamatoyannopoulos, J. A., & King, M.-C. (2011). Contribution of Inherited Mutations in the BRCA2-Interacting Protein PALB2 to Familial Breast Cancer. *Cancer Research*, *71*(6), 2222–2229. https://doi.org/10.1158/0008-5472.CAN-10-3958

Castéra, L., Harter, V., Muller, E., Krieger, S., Goardon, N., Ricou, A., Rousselin, A., Paimparay, G., Legros, A., Bruet, O., Quesnelle, C., Domin, F., San, C., Brault, B., Fouillet, R., Abadie, C., Béra, O., Berthet, P., French Exome Project Consortium, … Vaur, D. (2018). Landscape of pathogenic variations in a panel of 34 genes and cancer risk estimation from 5131 HBOC families. *Genetics in Medicine: Official Journal of the*

American College of Medical Genetics, *20*(12), 1677–1686. https://doi.org/10.1038/s41436-018-0005-9

Chan, C. H. T., Munusamy, P., Loke, S. Y., Koh, G. L., Yang, A. Z. Y., Law, H. Y., Yoon, C. S., Wong, C. Y., Yong, W. S., Wong, N. S., Ng, R. C. H., Ong, K. W., Madhukumar, P., Oey, C. L., Ho, G. H., Tan, P. H., Tan, M. H., Ang, P., Yap, Y. S., & Lee, A. S. G. (2018). Evaluation of three polygenic risk score models for the prediction of breast cancer risk in Singapore Chinese. *Oncotarget*, *9*(16), 12796–12804. https://doi.org/10.18632/oncotarget.24374

Chandler, M. R., Bilgili, E. P., & Merner, N. D. (2016). A Review of Whole-Exome Sequencing Efforts Toward Hereditary Breast Cancer Susceptibility Gene Discovery. *Human Mutation*, *37*(9), 835–846. https://doi.org/10.1002/humu.23017

Chen, S., & Parmigiani, G. (2007). Meta-analysis of BRCA1 and BRCA2 penetrance. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, *25*(11), 1329–1333. https://doi.org/10.1200/JCO.2006.09.1066

Cheung, P. F.-Y., Yip, C. W., Ng, L. W.-C., Lo, K. W., Wong, N., Choy, K. W., Chow, C., Chan, K. F., Cheung, T. T., Poon, R. T.-P., Fan, S. T., & Cheung, S. T. (2014). Establishment and characterization of a novel primary hepatocellular carcinoma cell line with metastatic ability in vivo. *Cancer Cell International*, *14*(1), 103. https://doi.org/10.1186/s12935-014-0103-y

Chlebowski, R. T., Anderson, G. L., Aragaki, A. K., Manson, J. E., Stefanick, M. L., Pan, K., Barrington, W., Kuller, L. H., Simon, M. S., Lane, D., Johnson, K. C., Rohan, T. E., Gass, M. L. S., Cauley, J. A., Paskett, E. D., Sattari, M., & Prentice, R. L. (2020). Association of Menopausal Hormone Therapy With Breast Cancer Incidence and

Mortality During Long-term Follow-up of the Women's Health Initiative Randomized Clinical Trials. *JAMA*, *324*(4), 369–380. https://doi.org/10.1001/jama.2020.9482

Cho, N. (2016). Molecular subtypes and imaging phenotypes of breast cancer. *Ultrasonography*, *35*(4), 281–288. https://doi.org/10.14366/usg.16030

Chompret, A., Brugières, L., Ronsin, M., Gardes, M., Dessarps-Freichey, F., Abel, A., Hua, D., Ligot, L., Dondon, M. G., Bressac-de Paillerets, B., Frébourg, T., Lemerle, J., Bonaïti-Pellié, C., & Feunteun, J. (2000). P53 germline mutations in childhood cancers and cancer risk for carrier individuals. *British Journal of Cancer*, *82*(12), 1932–1937. https://doi.org/10.1054/bjoc.2000.1167

Chun, H. H., & Gatti, R. A. (2004). Ataxia-telangiectasia, an evolving phenotype. *DNA Repair*, *3*(8–9), 1187–1196. https://doi.org/10.1016/j.dnarep.2004.04.010

Chun, J., Buechelmaier, E. S., & Powell, S. N. (2013). Rad51 Paralog Complexes BCDX2 and CX3 Act at Different Stages in the BRCA1-BRCA2-Dependent Homologous Recombination Pathway. *Molecular and Cellular Biology*, *33*(2), 387–395. https://doi.org/10.1128/MCB.00465-12

Cirulli, E. T., & Goldstein, D. B. (2010). Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature Reviews. Genetics*, *11*(6), 415–425. https://doi.org/10.1038/nrg2779

Colditz, G. A., Hankinson, S. E., Hunter, D. J., Willett, W. C., Manson, J. E., Stampfer, M. J., Hennekens, C., Rosner, B., & Speizer, F. E. (1995). The use of estrogens and progestins and the risk of breast cancer in postmenopausal women. *The New England Journal of Medicine*, *332*(24), 1589–1593. https://doi.org/10.1056/NEJM199506153322401

Collaborative Group on Hormonal Factors in Breast Cancer. (2001). Familial breast cancer: Collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. *Lancet (London, England)*, *358*(9291), 1389–1399. https://doi.org/10.1016/S0140-6736(01)06524-2

Collaborative Group on Hormonal Factors in Breast Cancer. (2012). Menarche, menopause, and breast cancer risk: Individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. *The Lancet. Oncology*, *13*(11), 1141–1151. https://doi.org/10.1016/S1470-2045(12)70425-4

Conneely, S. E., & Stevens, A. M. (2021). Acute Myeloid Leukemia in Children: Emerging Paradigms in Genetics and New Approaches to Therapy. *Current Oncology Reports*, *23*(2), 16. https://doi.org/10.1007/s11912-020-01009-3

Corso, G., & Magnoni, F. (2021). Hereditary breast cancer: Translation into clinical practice of recent American Society of Clinical Oncology, American Society of Radiation Oncology, and Society of Surgical Oncology recommendations. *European Journal of Cancer Prevention*, *30*(4), 311. https://doi.org/10.1097/CEJ.0000000000000624

Cortesi, L., Rugo, H. S., & Jackisch, C. (2021). An Overview of PARP Inhibitors for the Treatment of Breast Cancer. *Targeted Oncology*, *16*(3), 255–282. https://doi.org/10.1007/s11523-021-00796-4

Cortez, D., Guntuku, S., Qin, J., & Elledge, S. J. (2001). ATR and ATRIP: Partners in checkpoint signaling. *Science (New York, N.Y.)*, *294*(5547), 1713–1716. https://doi.org/10.1126/science.1065521

Couch, F. J., Hart, S. N., Sharma, P., Toland, A. E., Wang, X., Miron, P., Olson, J. E., Godwin, A. K., Pankratz, V. S., Olswold, C., Slettedahl, S., Hallberg, E., Guidugli, L., Davila, J. I., Beckmann, M. W., Janni, W., Rack, B., Ekici, A. B., Slamon, D. J., … Fasching, P. A. (2015). Inherited mutations in 17 breast cancer susceptibility genes among a large triple-negative breast cancer cohort unselected for family history of breast cancer. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, *33*(4), 304–311. https://doi.org/10.1200/JCO.2014.57.1414

Culjkovic-Kraljacic, B., & Borden, K. L. B. (2013). Aiding and abetting cancer: mRNA export and the nuclear pore. *Trends in Cell Biology*, *23*(7), 328–335. https://doi.org/10.1016/j.tcb.2013.03.004

Cybulski, C., Carrot-Zhang, J., Kluźniak, W., Rivera, B., Kashyap, A., Wokołorczyk, D., Giroux, S., Nadaf, J., Hamel, N., Zhang, S., Huzarski, T., Gronwald, J., Byrski, T., Szwiec, M., Jakubowska, A., Rudnicka, H., Lener, M., Masojć, B., Tonin, P. N., … Akbari, M. R. (2015). Germline RECQL mutations are associated with breast cancer susceptibility. *Nature Genetics*, *47*(6), 643–646. https://doi.org/10.1038/ng.3284

Cybulski, C., Kluźniak, W., Huzarski, T., Wokołorczyk, D., Kashyap, A., Jakubowska, A., Szwiec, M., Byrski, T., Dębniak, T., Górski, B., Sopik, V., Akbari, M. R., Sun, P., Gronwald, J., Narod, S. A., & Lubiński, J. (2015). Clinical outcomes in women with breast cancer and a PALB2 mutation: A prospective cohort analysis. *The Lancet Oncology*, *16*(6), 638–644. https://doi.org/10.1016/S1470-2045(15)70142-7

Cybulski, C., Wokołorczyk, D., Jakubowska, A., Huzarski, T., Byrski, T., Gronwald, J., Masojć, B., Dębniak, T., Górski, B., Blecharz, P., Narod, S. A., & Lubiński, J. (2011). Risk of Breast Cancer in Women With a CHEK2 Mutation With and Without a Family History of

Breast Cancer. *Journal of Clinical Oncology*, *29*(28), 3747–3752. https://doi.org/10.1200/JCO.2010.34.0778

Cybulski, C., Zamani, N., Kluźniak, W., Milano, L., Wokołorczyk, D., Stempa, K., Rudnicka, H., Zhang, S., Zadeh, M., Huzarski, T., Jakubowska, A., Dębniak, T., Lener, M., Szwiec, M., Domagała, P., Samani, A. A., Narod, S., Gronwald, J., Masson, J.-Y., … Akbari, M. R. (2023). Variants in ATRIP are associated with breast cancer susceptibility in the Polish population and UK Biobank. *American Journal of Human Genetics*, *110*(4), 648–662. https://doi.org/10.1016/j.ajhg.2023.03.002

Dai, X., Li, T., Bai, Z., Yang, Y., Liu, X., Zhan, J., & Shi, B. (2015). Breast cancer intrinsic subtype classification, clinical use and future trends. *American Journal of Cancer Research*, *5*(10), 2929–2943.

Daly, M. B., Pal, T., Berry, M. P., Buys, S. S., Dickson, P., Domchek, S. M., Elkhanany, A., Friedman, S., Goggins, M., Hutton, M. L., CGC, Karlan, B. Y., Khan, S., Klein, C., Kohlmann, W., CGC, Kurian, A. W., Laronga, C., Litton, J. K., … Dwyer, M. A. (2021). Genetic/Familial High-Risk Assessment: Breast, Ovarian, and Pancreatic, Version 2.2021, NCCN Clinical Practice Guidelines in Oncology. *Journal of the National Comprehensive Cancer Network: JNCCN*, *19*(1), 77–102. https://doi.org/10.6004/jnccn.2021.0001

Dashti, S., Taherian-Esfahani, Z., Keshtkar, A., & Ghafouri-Fard, S. (2019). Associations between XRCC3 Thr241Met polymorphisms and breast cancer risk: Systematic-review and meta-analysis of 55 case-control studies. *BMC Medical Genetics*, *20*(1), 79. https://doi.org/10.1186/s12881-019-0809-8

Davidoff, A. M., Humphrey, P. A., Iglehart, J. D., & Marks, J. R. (1991). Genetic basis for p53 overexpression in human breast cancer. *Proceedings of the National Academy of Sciences*, *88*(11), 5006–5010. https://doi.org/10.1073/pnas.88.11.5006

Davies, H., Glodzik, D., Morganella, S., Yates, L. R., Staaf, J., Zou, X., Ramakrishna, M., Martin, S., Boyault, S., Sieuwerts, A. M., Simpson, P. T., King, T. A., Raine, K., Eyfjord, J. E., Kong, G., Borg, Å., Birney, E., Stunnenberg, H. G., van de Vijver, M. J., … Nik-Zainal, S. (2017). HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nature Medicine*, *23*(4), 517–525. https://doi.org/10.1038/nm.4292

de Miranda, N. F., Björkman, A., & Pan-Hammarström, Q. (2011). DNA repair: The link between primary immunodeficiency and cancer. *Annals of the New York Academy of Sciences*, *1246*, 50–63. https://doi.org/10.1111/j.1749-6632.2011.06322.x

Degasperi, A., Amarante, T. D., Czarnecki, J., Shooter, S., Zou, X., Glodzik, D., Morganella, S., Nanda, A. S., Badja, C., Koh, G., Momen, S. E., Georgakopoulos-Soares, I., Dias, J. M. L., Young, J., Memari, Y., Davies, H., & Nik-Zainal, S. (2020). A practical framework and online tool for mutational signature analyses show intertissue variation and driver dependencies. *Nature Cancer*, *1*(2), 249–263. https://doi.org/10.1038/s43018-020-0027-5

Degasperi, A., Zou, X., Amarante, T. D., Martinez-Martinez, A., Koh, G. C. C., Dias, J. M. L., Heskin, L., Chmelova, L., Rinaldi, G., Wang, V. Y. W., Nanda, A. S., Bernstein, A., Momen, S. E., Young, J., Perez-Gil, D., Memari, Y., Badja, C., Shooter, S., Czarnecki, J., … Nik-Zainal, S. (2022). Substitution mutational signatures in whole-genome-sequenced cancers in the UK population. *Science (New York, N.Y.)*, *376*(6591), science.abl9283. https://doi.org/10.1126/science.abl9283

Díaz-Gay, M., Vangara, R., Barnes, M., Wang, X., Islam, S. M. A., Vermes, I., Narasimman, N. B., Yang, T., Jiang, Z., Moody, S., Senkin, S., Brennan, P., Stratton, M. R., & Alexandrov, L. B. (2023). *Assigning mutational signatures to individual samples and individual somatic mutations with SigProfilerAssignment* (p. 2023.07.10.548264). bioRxiv. https://doi.org/10.1101/2023.07.10.548264

Dieci, M. V., & Guarneri, V. (2020). Should triple-positive breast cancer be recognized as a distinct subtype? *Expert Review of Anticancer Therapy*, *20*(12), 1011–1014. https://doi.org/10.1080/14737140.2020.1829484

Dite, G. S., MacInnis, R. J., Bickerstaffe, A., Dowty, J. G., Allman, R., Apicella, C., Milne, R. L., Tsimiklis, H., Phillips, K.-A., Giles, G. G., Terry, M. B., Southey, M. C., & Hopper, J. L. (2016). Breast Cancer Risk Prediction Using Clinical Models and 77 Independent Risk-Associated SNPs for Women Aged Under 50 Years: Australian Breast Cancer Family Registry. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *25*(2), 359–365. https://doi.org/10.1158/1055-9965.EPI-15-0838

Do, H., & Dobrovic, A. (2015). Sequence Artifacts in DNA from Formalin-Fixed Tissues: Causes and Strategies for Minimization. *Clinical Chemistry*, *61*(1), 64–71. https://doi.org/10.1373/clinchem.2014.223040

Dolle, J. M., Daling, J. R., White, E., Brinton, L. A., Doody, D. R., Porter, P. L., & Malone, K. E. (2009). Risk factors for triple-negative breast cancer in women under the age of 45 years. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *18*(4), 1157–1166. https://doi.org/10.1158/1055-9965.EPI-08-1005

Donenberg, T., Lunn, J., Curling, D., Turnquest, T., Krill-Jackson, E., Royer, R., Narod, S. A., & Hurley, J. (2011). A high prevalence of BRCA1 mutations among breast cancer patients from the Bahamas. *Breast Cancer Research and Treatment*, *125*(2), 591–596. https://doi.org/10.1007/s10549-010-1156-9

Drost, J., van Boxtel, R., Blokzijl, F., Mizutani, T., Sasaki, N., Sasselli, V., de Ligt, J., Behjati, S., Grolleman, J. E., van Wezel, T., Nik-Zainal, S., Kuiper, R. P., Cuppen, E., & Clevers, H. (2017). Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. *Science (New York, N.Y.)*, *358*(6360), 234–238. https://doi.org/10.1126/science.aao3130

Easton, D. F. (1994). Cancer risks in A-T heterozygotes. *International Journal of Radiation Biology*, *66*(6 Suppl), S177-182. https://doi.org/10.1080/09553009414552011

Easton, D. F., Ford, D., & Bishop, D. T. (1995). Breast and ovarian cancer incidence in BRCA1-mutation carriers. Breast Cancer Linkage Consortium. *American Journal of Human Genetics*, *56*(1), 265–271.

Easton, D. F., Pharoah, P. D. P., Antoniou, A. C., Tischkowitz, M., Tavtigian, S. V., Nathanson, K. L., Devilee, P., Meindl, A., Couch, F. J., Southey, M., Goldgar, D. E., Evans, D. G. R., Chenevix-Trench, G., Rahman, N., Robson, M., Domchek, S. M., & Foulkes, W. D. (2015). Gene-Panel Sequencing and the Prediction of Breast-Cancer Risk. *New England Journal of Medicine*, *372*(23), 2243–2257. https://doi.org/10.1056/NEJMsr1501341

Easton, D. F., Pooley, K. A., Dunning, A. M., Pharoah, P. D. P., Thompson, D., Ballinger, D. G., Struewing, J. P., Morrison, J., Field, H., Luben, R., Wareham, N., Ahmed, S., Healey, C. S., Bowman, R., Meyer, K. B., Haiman, C. A., Kolonel, L. K., Henderson, B. E., Marchand, L. L., … Webb, P. (2007). Genome-wide association study identifies novel

breast cancer susceptibility loci. *Nature*, *447*(7148), 1087–1093. https://doi.org/10.1038/nature05887

Ege, M., Ma, Y., Manfras, B., Kalwak, K., Lu, H., Lieber, M. R., Schwarz, K., & Pannicke, U. (2005). Omenn syndrome due to ARTEMIS mutations. *Blood*, *105*(11), 4179–4186. https://doi.org/10.1182/blood-2004-12-4861

Eisen, A., Lubinski, J., Klijn, J., Moller, P., Lynch, H. T., Offit, K., Weber, B., Rebbeck, T., Neuhausen, S. L., Ghadirian, P., Foulkes, W. D., Gershoni-Baruch, R., Friedman, E., Rennert, G., Wagner, T., Isaacs, C., Kim-Sing, C., Ainsworth, P., Sun, P., & Narod, S. A. (2005). Breast cancer risk following bilateral oophorectomy in BRCA1 and BRCA2 mutation carriers: An international case-control study. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, *23*(30), 7491–7496. https://doi.org/10.1200/JCO.2004.00.7138

Elbakry, A., & Löbrich, M. (2021). Homologous Recombination Subpathways: A Tangle to Resolve. *Frontiers in Genetics*, *12*. https://doi.org/10.3389/fgene.2021.723847

Ellis, H., Klempner, S. J., Quintanilha, J. C. F. C. F., Myer, P., Lin, D. I., Panarelli, N., Fisher, V., Li, G., Huang, R. S. P., Ross, J. S., Oxnard, G. R., & Graf, R. (2023). Microsatellite instability (MSI), mismatch repair (MMR), and tumor mutational burden (TMB) as predictive biomarkers for immune checkpoint inhibitor (ICI) effectiveness in real-world patients with metastatic colorectal cancer (mCRC). *Journal of Clinical Oncology*. https://doi.org/10.1200/JCO.2023.41.4_suppl.46

Fakhri, N., Chad, M. A., Lahkim, M., Houari, A., Dehbi, H., Belmouden, A., & El Kadmiri, N. (2022). Risk factors for breast cancer in women: An update review. *Medical Oncology*

(Northwood, London, England), *39*(12), 197. https://doi.org/10.1007/s12032-022-01804-x

Falck, J., Coates, J., & Jackson, S. P. (2005). Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage. *Nature*, *434*(7033), 605–611. https://doi.org/10.1038/nature03442

Fang, Y., Tsao, C., Goodman, B. K., Furumai, R., Tirado, C. A., Abraham, R. T., & Wang, X. (2004). ATR functions as a gene dosage-dependent tumor suppressor on a mismatch repair-deficient background. *The EMBO Journal*, *23*(15), 3164–3174. https://doi.org/10.1038/sj.emboj.7600315

Farmer, H., McCabe, N., Lord, C. J., Tutt, A. N. J., Johnson, D. A., Richardson, T. B., Santarosa, M., Dillon, K. J., Hickson, I., Knights, C., Martin, N. M. B., Jackson, S. P., Smith, G. C. M., & Ashworth, A. (2005). Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature*, *434*(7035), 917–921. https://doi.org/10.1038/nature03445

Fattahi, Z., Beheshtian, M., Mohseni, M., Poustchi, H., Sellars, E., Nezhadi, S. H., Amini, A., Arzhangi, S., Jalalvand, K., Jamali, P., Mohammadi, Z., Davarnia, B., Nikuei, P., Oladnabi, M., Mohammadzadeh, A., Zohrehvand, E., Nejatizadeh, A., Shekari, M., Bagherzadeh, M., … Najmabadi, H. (2019). Iranome: A catalog of genomic variations in the Iranian population. *Human Mutation*, *40*(11), 1968–1984. https://doi.org/10.1002/humu.23880

Fawzy, M., Bahanassy, A., Samir, A., & Hafez, H. (2015). Profiling Loss of Heterozygosity Patterns in a Cohort of Favorable Histology Nephroblastoma Egyptian Patients: What is Consistent With the Rest of the World. *Pediatric Hematology and Oncology*, *32*(8), 548–556. https://doi.org/10.3109/08880018.2015.1071902

Filippini, S. E., & Vega, A. (2013). Breast cancer genes: Beyond BRCA1 and BRCA2. *Frontiers in Bioscience (Landmark Edition)*, *18*(4), 1358–1372. https://doi.org/10.2741/4185

Fong, P. C., Boss, D. S., Yap, T. A., Tutt, A., Wu, P., Mergui-Roelvink, M., Mortimer, P., Swaisland, H., Lau, A., O'Connor, M. J., Ashworth, A., Carmichael, J., Kaye, S. B., Schellens, J. H. M., & De Bono, J. S. (2009). Inhibition of Poly(ADP-Ribose) Polymerase in Tumors from *BRCA* Mutation Carriers. *New England Journal of Medicine*, *361*(2), 123–134. https://doi.org/10.1056/NEJMoa0900212

Ford, D., Easton, D. F., Stratton, M., Narod, S., Goldgar, D., Devilee, P., Bishop, D. T., Weber, B., Lenoir, G., Chang-Claude, J., Sobol, H., Teare, M. D., Struewing, J., Arason, A., Scherneck, S., Peto, J., Rebbeck, T. R., Tonin, P., Neuhausen, S., … Zelada-Hedman, M. (1998). Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. *American Journal of Human Genetics*, *62*(3), 676–689. https://doi.org/10.1086/301749

Fostira, F., Tsitlaidou, M., Papadimitriou, C., Pertesi, M., Timotheadou, E., Stavropoulou, A. V., Glentis, S., Bournakis, E., Bobos, M., Pectasides, D., Papakostas, P., Pentheroudakis, G., Gogas, H., Skarlos, P., Samantas, E., Bafaloukos, D., Kosmidis, P. A., Koutras, A., Yannoukakos, D., … Fountzilas, G. (2012). Prevalence of BRCA1 mutations among 403 women with triple-negative breast cancer: Implications for genetic screening selection criteria: a Hellenic Cooperative Oncology Group Study. *Breast Cancer Research and Treatment*, *134*(1), 353–362. https://doi.org/10.1007/s10549-012-2021-9

Fragomeni, S. M., Sciallis, A., & Jeruss, J. S. (2018). Molecular Subtypes and Local-Regional Control of Breast Cancer. *Surgical Oncology Clinics of North America*, *27*(1), 95–120. https://doi.org/10.1016/j.soc.2017.08.005

Gabai-Kapara, E., Lahad, A., Kaufman, B., Friedman, E., Segev, S., Renbaum, P., Beeri, R., Gal, M., Grinshpun-Cohen, J., Djemal, K., Mandell, J. B., Lee, M. K., Beller, U., Catane, R., King, M.-C., & Levy-Lahad, E. (2014). Population-based screening for breast and ovarian cancer risk due to BRCA1 and BRCA2. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(39), 14205–14210. https://doi.org/10.1073/pnas.1415979111

Garcin, E. B., Gon, S., Sullivan, M. R., Brunette, G. J., Cian, A. D., Concordet, J.-P., Giovannangeli, C., Dirks, W. G., Eberth, S., Bernstein, K. A., Prakash, R., Jasin, M., & Modesti, M. (2019). Differential Requirements for the RAD51 Paralogs in Genome Repair and Maintenance in Human Cells. *PLoS Genetics*, *15*(10), e1008355. https://doi.org/10.1371/journal.pgen.1008355

Garrett, J. T., & Arteaga, C. L. (2011). Resistance to HER2-directed antibodies and tyrosine kinase inhibitors: Mechanisms and clinical implications. *Cancer Biology & Therapy*, *11*(9), 793–800. https://doi.org/10.4161/cbt.11.9.15045

Ghimenti, C., Sensi, E., Presciuttini, S., Brunetti, I. M., Conte, P., Bevilacqua, G., & Caligo, M. A. (2002). Germline mutations of the BRCA1-associated ring domain (BARD1) gene in breast and breast/ovarian families negative for BRCA1 and BRCA2 alterations. *Genes, Chromosomes & Cancer*, *33*(3), 235–242. https://doi.org/10.1002/gcc.1223

Giglia-Mari, G., & Sarasin, A. (2003). TP53 mutations in human skin cancers. *Human Mutation*, *21*(3), 217–228. https://doi.org/10.1002/humu.10179

Giner-Calabuig, M., De Leon, S., Wang, J., Fehlmann, T. D., Ukaegbu, C., Gibson, J., Alustiza-Fernandez, M., Pico, M.-D., Alenda, C., Herraiz, M., Carrillo-Palau, M., Salces, I., Reyes, J., Ortega, S. P., Obrador-Hevia, A., Cecchini, M., Syngal, S., Stoffel, E., Ellis, N.

A., … Xicola, R. M. (2022). Mutational signature profiling classifies subtypes of clinically different mismatch-repair-deficient tumours with a differential immunogenic response potential. *British Journal of Cancer*, *126*(11), 1595–1603. https://doi.org/10.1038/s41416-022-01754-1

Glodzik, D., Selenica, P., Rogge, R. A., Silverman, I. M., Mandelker, D., Harris, S., Zhao, J., Zinda, M., Veloso, A., Malani, N., Riaz, N., Koehler, M., Daber, R. D., Johnson, V., Rimkunas, V., & Reis-Filho, J. S. (2023). Detection of Biallelic Loss of DNA Repair Genes in Formalin-Fixed, Paraffin-Embedded Tumor Samples Using a Novel Tumor-Only Sequencing Panel. *The Journal of Molecular Diagnostics*, *25*(5), 295–310. https://doi.org/10.1016/j.jmoldx.2023.02.004

Godoy-Ortiz, A., Sanchez-Muñoz, A., Chica Parrado, M. R., Álvarez, M., Ribelles, N., Rueda Dominguez, A., & Alba, E. (2019). Deciphering HER2 Breast Cancer Disease: Biological and Clinical Implications. *Frontiers in Oncology*, *9*. https://doi.org/10.3389/fonc.2019.01124

Gonzalez-Neira, A., Rosa-Rosa, J. M., Osorio, A., Gonzalez, E., Southey, M., Sinilnikova, O., Lynch, H., Oldenburg, R. A., van Asperen, C. J., Hoogerbrugge, N., Pita, G., Devilee, P., Goldgar, D., & Benitez, J. (2007). Genomewide high-density SNP linkage analysis of non-BRCA1/2 breast cancer families identifies various candidate regions and has greater power than microsatellite studies. *BMC Genomics*, *8*, 299. https://doi.org/10.1186/1471-2164-8-299

Górski, B., Dębniak, T., Masojć, B., Mierzejewski, M., Mędrek, K., Cybulski, C., Jakubowska, A., Kurzawski, G., Chosia, M., Scott, R., & Lubiński, J. (2003). Germline 657del5

mutation in the NBS1 gene in breast cancer patients. *International Journal of Cancer*, *106*(3), 379–381. https://doi.org/10.1002/ijc.11231

Gracia-Aznarez, F. J., Fernandez, V., Pita, G., Peterlongo, P., Dominguez, O., de la Hoya, M., Duran, M., Osorio, A., Moreno, L., Gonzalez-Neira, A., Rosa-Rosa, J. M., Sinilnikova, O., Mazoyer, S., Hopper, J., Lazaro, C., Southey, M., Odefrey, F., Manoukian, S., Catucci, I., … Benitez, J. (2013). Whole exome sequencing suggests much of non-BRCA1/BRCA2 familial breast cancer is due to moderate and low penetrance susceptibility alleles. *PloS One*, *8*(2), e55681. https://doi.org/10.1371/journal.pone.0055681

Gralewska, P., Gajek, A., Marczak, A., & Rogalska, A. (2020). Participation of the ATR/CHK1 pathway in replicative stress targeted therapy of high-grade ovarian cancer. *Journal of Hematology & Oncology*, *13*(1), 39. https://doi.org/10.1186/s13045-020-00874-6

Greene, G. L. (2013). In memoriam: Elwood Jensen (1920-2012). *Endocrine Reviews*, *34*(6), 761–763. https://doi.org/10.1210/er.2013-1031

Gronwald, J., Tung, N., Foulkes, W. D., Offit, K., Gershoni, R., Daly, M., Kim-Sing, C., Olsson, H., Ainsworth, P., Eisen, A., Saal, H., Friedman, E., Olopade, O., Osborne, M., Weitzel, J., Lynch, H., Ghadirian, P., Lubinski, J., Sun, P., … Hereditary Breast Cancer Clinical Study Group. (2006). Tamoxifen and contralateral breast cancer in BRCA1 and BRCA2 carriers: An update. *International Journal of Cancer*, *118*(9), 2281–2284. https://doi.org/10.1002/ijc.21536

Guillemette, L., Allard, C., Lacroix, M., Patenaude, J., Battista, M.-C., Doyon, M., Moreau, J., Ménard, J., Bouchard, L., Ardilouze, J.-L., Perron, P., & Hivert, M.-F. (2016). Genetics of Glucose regulation in Gestation and Growth (Gen3G): A prospective prebirth cohort of

mother–child pairs in Sherbrooke, Canada. *BMJ Open*, *6*(2), e010031. https://doi.org/10.1136/bmjopen-2015-010031

Guiu, S., Michiels, S., André, F., Cortes, J., Denkert, C., Di Leo, A., Hennessy, B. T., Sorlie, T., Sotiriou, C., Turner, N., Van de Vijver, M., Viale, G., Loi, S., & Reis-Filho, J. S. (2012). Molecular subclasses of breast cancer: How do we define them? The IMPAKT 2012 Working Group Statement†. *Annals of Oncology*, *23*(12), 2997–3006. https://doi.org/10.1093/annonc/mds586

Guo, Q., Lakatos, E., Bakir, I. A., Curtius, K., Graham, T. A., & Mustonen, V. (2022). The mutational signatures of formalin fixation on the human genome. *Nature Communications*, *13*(1), 4487. https://doi.org/10.1038/s41467-022-32041-5

Guo, W., Key, T. J., & Reeves, G. K. (2018). Adiposity and breast cancer risk in postmenopausal women: Results from the UK Biobank prospective cohort. *International Journal of Cancer*, *143*(5), 1037–1046. https://doi.org/10.1002/ijc.31394

Haldane, J. B. (1956). The estimation and significance of the logarithm of a ratio of frequencies. *Annals of Human Genetics*, *20*(4), 309–311. https://doi.org/10.1111/j.1469-1809.1955.tb01285.x

Hall, J. M., Lee, M. K., Newman, B., Morrow, J. E., Anderson, L. A., Huey, B., & King, M. C. (1990). Linkage of early-onset familial breast cancer to chromosome 17q21. *Science (New York, N.Y.)*, *250*(4988), 1684–1689. https://doi.org/10.1126/science.2270482

Hamajima, N., Hirose, K., Tajima, K., Rohan, T., Calle, E. E., Heath, C. W., Coates, R. J., Liff, J. M., Talamini, R., Chantarakul, N., Koetsawang, S., Rachawat, D., Morabia, A., Schuman, L., Stewart, W., Szklo, M., Bain, C., Schofield, F., Siskind, V., … Collaborative Group on Hormonal Factors in Breast Cancer. (2002). Alcohol, tobacco

and breast cancer—Collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease. *British Journal of Cancer*, *87*(11), 1234–1245. https://doi.org/10.1038/sj.bjc.6600596

Hammond, M. E. H., Hayes, D. F., Dowsett, M., Allred, D. C., Hagerty, K. L., Badve, S., Fitzgibbons, P. L., Francis, G., Goldstein, N. S., Hayes, M., Hicks, D. G., Lester, S., Love, R., Mangu, P. B., McShane, L., Miller, K., Osborne, C. K., Paik, S., Perlmutter, J., … Wolff, A. C. (2010). American Society of Clinical Oncology/College of American Pathologists Guideline Recommendations for Immunohistochemical Testing of Estrogen and Progesterone Receptors in Breast Cancer. *Journal of Clinical Oncology*, *28*(16), 2784–2795. https://doi.org/10.1200/JCO.2009.25.6529

Han, M.-R., Shin, S., Park, H.-C., Kim, M. S., Lee, S. H., Jung, S. H., Song, S. Y., Lee, S. H., & Chung, Y.-J. (2018). Mutational signatures and chromosome alteration profiles of squamous cell carcinomas of the vulva. *Experimental & Molecular Medicine*, *50*(2), Article 2. https://doi.org/10.1038/emm.2017.265

Hassanin, E., May, P., Aldisi, R., Spier, I., Forstner, A. J., Nöthen, M. M., Aretz, S., Krawitz, P., Bobbili, D. R., & Maj, C. (2022). Breast and prostate cancer risk: The interplay of polygenic risk, rare pathogenic germline variants, and family history. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, *24*(3), 576–585. https://doi.org/10.1016/j.gim.2021.11.009

Hilbers, F. S., Meijers, C. M., Laros, J. F. J., van Galen, M., Hoogerbrugge, N., Vasen, H. F. A., Nederlof, P. M., Wijnen, J. T., van Asperen, C. J., & Devilee, P. (2013). Exome sequencing of germline DNA from non-BRCA1/2 familial breast cancer cases selected

on the basis of aCGH tumor profiling. *PloS One*, *8*(1), e55734. https://doi.org/10.1371/journal.pone.0055734

Hirsch, S., Gieldon, L., Sutter, C., Dikow, N., & Schaaf, C. P. (2021). Germline testing for homologous recombination repair genes—Opportunities and challenges. *Genes, Chromosomes and Cancer*, *60*(5), 332–343. https://doi.org/10.1002/gcc.22900

Hirschhorn, J. N. (2005). Genetic approaches to studying common diseases and complex traits. *Pediatric Research*, *57*(5 Pt 2), 74R-77R. https://doi.org/10.1203/01.PDR.0000159574.98964.87

Hollestelle, A., Wasielewski, M., Martens, J. W. M., & Schutte, M. (2010). Discovering moderate-risk breast cancer susceptibility genes. *Current Opinion in Genetics & Development*, *20*(3), 268–276. https://doi.org/10.1016/j.gde.2010.02.009

Hollstein, M., Moriya, M., Grollman, A. P., & Olivier, M. (2013). Analysis of TP53 mutation spectra reveals the fingerprint of the potent environmental carcinogen, aristolochic acid. *Mutation Research*, *753*(1), 41–49. https://doi.org/10.1016/j.mrrev.2013.02.003

Hoover, R., Gray, L. A., Cole, P., & MacMahon, B. (1976). Menopausal estrogens and breast cancer. *The New England Journal of Medicine*, *295*(8), 401–405. https://doi.org/10.1056/NEJM197608192950801

Horn, J., Åsvold, B. O., Opdahl, S., Tretli, S., & Vatten, L. J. (2013). Reproductive factors and the risk of breast cancer in old age: A Norwegian cohort study. *Breast Cancer Research and Treatment*, *139*(1), 237–243. https://doi.org/10.1007/s10549-013-2531-0

Hu, C., Hart, S. N., Gnanaolivu, R., Huang, H., Lee, K. Y., Na, J., Gao, C., Lilyquist, J., Yadav, S., Boddicker, N. J., Samara, R., Klebba, J., Ambrosone, C. B., Anton-Culver, H., Auer, P., Bandera, E. V., Bernstein, L., Bertrand, K. A., Burnside, E. S., … Couch, F. J. (2021).

A Population-Based Study of Genes Previously Implicated in Breast Cancer. *New England Journal of Medicine*, *384*(5), 440–451. https://doi.org/10.1056/NEJMoa2005936

Huang, P.-J., Chiu, L.-Y., Lee, C.-C., Yeh, Y.-M., Huang, K.-Y., Chiu, C.-H., & Tang, P. (2018). mSignatureDB: A database for deciphering mutational signatures in human cancers. *Nucleic Acids Research*, *46*(Database issue), D964–D970. https://doi.org/10.1093/nar/gkx1133

Huang, R.-X., & Zhou, P.-K. (2020). DNA damage response signaling pathways and targets for radiotherapy sensitization in cancer. *Signal Transduction and Targeted Therapy*, *5*(1), 60. https://doi.org/10.1038/s41392-020-0150-x

Hungermann, D., Schmidt, H., Natrajan, R., Tidow, N., Poos, K., Reis-Filho, J. S., Brandt, B., Buerger, H., & Korsching, E. (2011). Influence of whole arm loss of chromosome 16q on gene expression patterns in oestrogen receptor-positive, invasive breast cancer. *The Journal of Pathology*, *224*(4), 517–528. https://doi.org/10.1002/path.2938

Hunter, D. J., Colditz, G. A., Hankinson, S. E., Malspeis, S., Spiegelman, D., Chen, W., Stampfer, M. J., & Willett, W. C. (2010). Oral contraceptive use and breast cancer: A prospective study of young women. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *19*(10), 2496–2502. https://doi.org/10.1158/1055-9965.EPI-10-0747

Hunter, D. J., Kraft, P., Jacobs, K. B., Cox, D. G., Yeager, M., Hankinson, S. E., Wacholder, S., Wang, Z., Welch, R., Hutchinson, A., Wang, J., Yu, K., Chatterjee, N., Orr, N., Willett, W. C., Colditz, G. A., Ziegler, R. G., Berg, C. D., Buys, S. S., … Chanock, S. J. (2007). A genome-wide association study identifies alleles in FGFR2 associated with risk of

sporadic postmenopausal breast cancer. *Nature Genetics*, *39*(7), 870–874. https://doi.org/10.1038/ng2075

Inic, Z., Zegarac, M., Inic, M., Markovic, I., Kozomara, Z., Djurisic, I., Inic, I., Pupic, G., & Jancic, S. (2014). Difference between Luminal A and Luminal B Subtypes According to Ki-67, Tumor Size, and Progesterone Receptor Negativity Providing Prognostic Information. *Clinical Medicine Insights: Oncology*, *8*, CMO.S18006. https://doi.org/10.4137/CMO.S18006

Islami, F., Liu, Y., Jemal, A., Zhou, J., Weiderpass, E., Colditz, G., Boffetta, P., & Weiss, M. (2015). Breastfeeding and breast cancer risk by receptor status—A systematic review and meta-analysis. *Annals of Oncology: Official Journal of the European Society for Medical Oncology*, *26*(12), 2398–2407. https://doi.org/10.1093/annonc/mdv379

Iyer, R. R., Pluciennik, A., Napierala, M., & Wells, R. D. (2015). DNA Triplet Repeat Expansion and Mismatch Repair. *Annual Review of Biochemistry*, *84*, 199–226. https://doi.org/10.1146/annurev-biochem-060614-034010

Iyevleva, A. G., Aleksakhina, S. N., Sokolenko, A. P., Baskina, S. V., Venina, A. R., Anisimova, E. I., Bizin, I. V., Ivantsov, A. O., Belysheva, Y. V., Chernyakova, A. P., Togo, A. V., & Imyanitov, E. N. (2022). Somatic loss of the remaining allele occurs approximately in half of CHEK2-driven breast cancers and is accompanied by a border-line increase of chromosomal instability. *Breast Cancer Research and Treatment*. https://doi.org/10.1007/s10549-022-06517-3

Jensen, R. B., Carreira, A., & Kowalczykowski, S. C. (2010). Purified human BRCA2 stimulates RAD51-mediated recombination. *Nature*, *467*(7316), 678–683. https://doi.org/10.1038/nature09399

Jia, P., & Zhao, Z. (2019). Characterization of Tumor-Suppressor Gene Inactivation Events in 33 Cancer Types. *Cell Reports*, *26*(2), 496-506.e3. https://doi.org/10.1016/j.celrep.2018.12.066

Jiménez-Salazar, J. E., Damian-Ferrara, R., Arteaga, M., & Damián-Matsumura, P. (2021). Non-Genomic Actions of Estrogens on the DNA Repair Pathways Are Associated With Chemotherapy Resistance in Breast Cancer. *Frontiers in Oncology*, *11*. https://doi.org/10.3389/fonc.2021.631007

Kamps, R., Brandão, R. D., Bosch, B. J. van den, Paulussen, A. D. C., Xanthoulea, S., Blok, M. J., & Romano, A. (2017). Next-Generation Sequencing in Oncology: Genetic Diagnosis, Risk Prediction and Cancer Classification. *International Journal of Molecular Sciences*, *18*(2), 308. https://doi.org/10.3390/ijms18020308

Kanadys, W., Barańska, A., Malm, M., Błaszczuk, A., Polz-Dacewicz, M., Janiszewska, M., & Jędrych, M. (2021). Use of Oral Contraceptives as a Potential Risk Factor for Breast Cancer: A Systematic Review and Meta-Analysis of Case-Control Studies Up to 2010. *International Journal of Environmental Research and Public Health*, *18*(9), 4638. https://doi.org/10.3390/ijerph18094638

Karczewski, K. J., Weisburd, B., Thomas, B., Solomonson, M., Ruderfer, D. M., Kavanagh, D., Hamamsy, T., Lek, M., Samocha, K. E., Cummings, B. B., Birnbaum, D., The Exome Aggregation Consortium, Daly, M. J., & MacArthur, D. G. (2017). The ExAC browser: Displaying reference data information from over 60 000 exomes. *Nucleic Acids Research*, *45*(D1), D840–D845. https://doi.org/10.1093/nar/gkw971

Keen-Kim, D., Mathews, C. A., Reus, V. I., Lowe, T. L., Herrera, L. D., Budman, C. L., Gross-Tsur, V., Pulver, A. E., Bruun, R. D., Erenberg, G., Naarden, A., Sabatti, C., & Freimer,

N. B. (2006). Overrepresentation of rare variants in a specific ethnic group may confuse interpretation of association analyses. *Human Molecular Genetics*, *15*(22), 3324–3328. https://doi.org/10.1093/hmg/ddl408

Kennedy-Darling, J., & Smith, L. M. (2014). Measuring the Formaldehyde Protein–DNA Cross-Link Reversal Rate. *Analytical Chemistry*, *86*(12), 5678–5681. https://doi.org/10.1021/ac501354y

Kerkhof, J., Schenkel, L. C., Reilly, J., McRobbie, S., Aref-Eshghi, E., Stuart, A., Rupar, C. A., Adams, P., Hegele, R. A., Lin, H., Rodenhiser, D., Knoll, J., Ainsworth, P. J., & Sadikovic, B. (2017). Clinical Validation of Copy Number Variant Detection from Targeted Next-Generation Sequencing Panels. *The Journal of Molecular Diagnostics: JMD*, *19*(6), 905–920. https://doi.org/10.1016/j.jmoldx.2017.07.004

Khandekar, A., Vangara, R., Barnes, M., Díaz-Gay, M., Abbasi, A., Bergstrom, E. N., Steele, C. D., Pillay, N., & Alexandrov, L. B. (2023). Visualizing and exploring patterns of large mutational events with SigProfilerMatrixGenerator. *BMC Genomics*, *24*(1), 469. https://doi.org/10.1186/s12864-023-09584-y

Kiezun, A., Garimella, K., Do, R., Stitziel, N. O., Neale, B. M., McLaren, P. J., Gupta, N., Sklar, P., Sullivan, P. F., Moran, J. L., Hultman, C. M., Lichtenstein, P., Magnusson, P., Lehner, T., Shugart, Y. Y., Price, A. L., de Bakker, P. I. W., Purcell, S. M., & Sunyaev, S. R. (2012). Exome sequencing and the genetic basis of complex traits. *Nature Genetics*, *44*(6), 623–630. https://doi.org/10.1038/ng.2303

Kim, E.-K., Park, S. Y., & Kim, S.-W. (2020). Clinicopathological characteristics of BRCA-associated breast cancer in Asian patients. *Journal of Pathology and Translational Medicine*, *54*(4), 265–275. https://doi.org/10.4132/jptm.2020.04.07

Kluth, M., Runte, F., Barow, P., Omari, J., Abdelaziz, Z. M., Paustian, L., Steurer, S., Christina Tsourlakis, M., Fisch, M., Graefen, M., Tennstedt, P., Huland, H., Michl, U., Minner, S., Sauter, G., Simon, R., Adam, M., & Schlomm, T. (2015). Concurrent deletion of 16q23 and PTEN is an independent prognostic feature in prostate cancer. *International Journal of Cancer*, *137*(10), 2354–2363. https://doi.org/10.1002/ijc.29613

Knight, J. A., Fan, J., Malone, K. E., John, E. M., Lynch, C. F., Langballe, R., Bernstein, L., Shore, R. E., Brooks, J. D., Reiner, A. S., Woods, M., Liang, X., Bernstein, J. L., & WECARE Study Collaborative Group. (2017). Alcohol consumption and cigarette smoking in combination: A predictor of contralateral breast cancer risk in the WECARE study. *International Journal of Cancer*, *141*(5), 916–924. https://doi.org/10.1002/ijc.30791

Knudson, A. G. (2001). Two genetic hits (more or less) to cancer. *Nature Reviews Cancer*, *1*(2), 157–162. https://doi.org/10.1038/35101031

Koch, L. (2014). Cancer genomics: Non-coding mutations in the driver seat. *Nature Reviews. Genetics*, *15*(9), 574–575. https://doi.org/10.1038/nrg3801

Kotsopoulos, J., Lubinski, J., Lynch, H. T., Kim-Sing, C., Neuhausen, S., Demsky, R., Foulkes, W. D., Ghadirian, P., Tung, N., Ainsworth, P., Senter, L., Karlan, B., Eisen, A., Eng, C., Weitzel, J., Gilchrist, D. M., Blum, J. L., Zakalik, D., Singer, C., … Narod, S. A. (2012). Oophorectomy after Menopause and the Risk of Breast Cancer in *BRCA1* and *BRCA2* Mutation Carriers. *Cancer Epidemiology, Biomarkers & Prevention*, *21*(7), 1089–1096. https://doi.org/10.1158/1055-9965.EPI-12-0201

Krasikova, Y. S., Rechkunova, N. I., & Lavrik, O. I. (2016). [Replication protein A as a major eukaryotic single-stranded DNA-binding protein and its role in DNA repair]. *Molekuliarnaia Biologiia*, *50*(5), 735–750. https://doi.org/10.7868/S0026898416030083

Kriege, M., Hollestelle, A., Jager, A., Huijts, P. E. A., Berns, E. M., Sieuwerts, A. M., Meijer-van Gelder, M. E., Collée, J. M., Devilee, P., Hooning, M. J., Martens, J. W. M., & Seynaeve, C. (2014). Survival and contralateral breast cancer in CHEK2 1100delC breast cancer patients: Impact of adjuvant chemotherapy. *British Journal of Cancer*, *111*(5), 1004–1013. https://doi.org/10.1038/bjc.2014.306

Kucab, J. E., Zou, X., Morganella, S., Joel, M., Nanda, A. S., Nagy, E., Gomez, C., Degasperi, A., Harris, R., Jackson, S. P., Arlt, V. M., Phillips, D. H., & Nik-Zainal, S. (2019). A Compendium of Mutational Signatures of Environmental Agents. *Cell*, *177*(4), 821-836.e16. https://doi.org/10.1016/j.cell.2019.03.001

Kuchenbaecker, K. B., Hopper, J. L., Barnes, D. R., Phillips, K.-A., Mooij, T. M., Roos-Blom, M.-J., Jervis, S., van Leeuwen, F. E., Milne, R. L., Andrieu, N., Goldgar, D. E., Terry, M. B., Rookus, M. A., Easton, D. F., Antoniou, A. C., BRCA1 and BRCA2 Cohort Consortium, McGuffog, L., Evans, D. G., Barrowdale, D., … 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

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

Lee, J. E. A., Li, N., Rowley, S. M., Cheasley, D., Zethoven, M., McInerny, S., Gorringe, K. L., James, P. A., & Campbell, I. G. (2018). Molecular analysis of PALB2-associated breast cancers. *The Journal of Pathology*, *245*(1), 53–60. https://doi.org/10.1002/path.5055

Letouzé, E., Shinde, J., Renault, V., Couchy, G., Blanc, J.-F., Tubacher, E., Bayard, Q., Bacq, D., Meyer, V., Semhoun, J., Bioulac-Sage, P., Prévôt, S., Azoulay, D., Paradis, V., Imbeaud, S., Deleuze, J.-F., & Zucman-Rossi, J. (2017). Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nature Communications*, *8*(1), Article 1. https://doi.org/10.1038/s41467-017-01358-x

Lhota, F., Zemankova, P., Kleiblova, P., Soukupova, J., Vocka, M., Stranecky, V., Janatova, M., Hartmannova, H., Hodanova, K., Kmoch, S., & Kleibl, Z. (2016). Hereditary truncating mutations of DNA repair and other genes in BRCA1/BRCA2/PALB2-negatively tested breast cancer patients. *Clinical Genetics*, *90*(4), 324–333. https://doi.org/10.1111/cge.12748

Li, A., Geyer, F. C., Blecua, P., Lee, J. Y., Selenica, P., Brown, D. N., Pareja, F., Lee, S. S. K., Kumar, R., Rivera, B., Bi, R., Piscuoglio, S., Wen, H. Y., Lozada, J. R., Gularte-Mérida, R., Cavallone, L., Rezoug, Z., Nguyen-Dumont, T., Peterlongo, P., … Reis-Filho, J. S. (2019). Homologous recombination DNA repair defects in PALB2-associated breast cancers. *Npj Breast Cancer*, *5*(1), Article 1. https://doi.org/10.1038/s41523-019-0115-9

Li, L., Moshous, D., Zhou, Y., Wang, J., Xie, G., Salido, E., Hu, D., de Villartay, J.-P., & Cowan, M. J. (2002). A founder mutation in Artemis, an SNM1-like protein, causes SCID in Athabascan-speaking Native Americans. *Journal of Immunology (Baltimore, Md.: 1950)*, *168*(12), 6323–6329. https://doi.org/10.4049/jimmunol.168.12.6323

Li, N., McInerny, S., Zethoven, M., Cheasley, D., Lim, B. W. X., Rowley, S. M., Devereux, L., Grewal, N., Ahmadloo, S., Byrne, D., Lee, J. E. A., Li, J., Fox, S. B., John, T., Antill, Y., Gorringe, K. L., James, P. A., & Campbell, I. G. (2019). Combined Tumor Sequencing and Case-Control Analyses of RAD51C in Breast Cancer. *Journal of the National Cancer Institute*, *111*(12), 1332–1338. https://doi.org/10.1093/jnci/djz045

Li, X., Yang, J., Peng, L., Sahin, A. A., Huo, L., Ward, K. C., O'Regan, R., Torres, M. A., & Meisel, J. L. (2017). Triple-negative breast cancer has worse overall survival and cause-specific survival than non-triple-negative breast cancer. *Breast Cancer Research and Treatment*, *161*(2), 279–287. https://doi.org/10.1007/s10549-016-4059-6

Liang, M., Zhang, Y., Sun, C., Rizeq, F. K., Min, M., Shi, T., & Sun, Y. (2018). Association Between CHEK2*1100delC and Breast Cancer: A Systematic Review and Meta-Analysis. *Molecular Diagnosis & Therapy*, *22*(4), 397–407. https://doi.org/10.1007/s40291-018-0344-x

Lilyquist, J., LaDuca, H., Polley, E., Davis, B. T., Shimelis, H., Hu, C., Hart, S. N., Dolinsky, J. S., Couch, F. J., & Goldgar, D. E. (2017). Frequency of mutations in a large series of clinically ascertained ovarian cancer cases tested on multi-gene panels compared to reference controls. *Gynecologic Oncology*, *147*(2), 375–380. https://doi.org/10.1016/j.ygyno.2017.08.030

Liu, A., Yoshioka, K., Salerno, V., & Hsieh, P. (2008). The Mismatch Repair-Mediated Cell Cycle Checkpoint Response to Fluorodeoxyuridine. *Journal of Cellular Biochemistry*, *105*(1), 245–254. https://doi.org/10.1002/jcb.21824

Liu, T., & Huang, J. (2016). Replication protein A and more: Single-stranded DNA-binding proteins in eukaryotic cells. *Acta Biochimica et Biophysica Sinica*, *48*(7), 665–670. https://doi.org/10.1093/abbs/gmw041

Llombart-Cussac, A., Cortés, J., Paré, L., Galván, P., Bermejo, B., Martínez, N., Vidal, M., Pernas, S., López, R., Muñoz, M., Nuciforo, P., Morales, S., Oliveira, M., de la Peña, L., Peláez, A., & Prat, A. (2017). HER2-enriched subtype as a predictor of pathological complete response following trastuzumab and lapatinib without chemotherapy in early-stage HER2-positive breast cancer (PAMELA): An open-label, single-group, multicentre, phase 2 trial. *The Lancet Oncology*, *18*(4), 545–554. https://doi.org/10.1016/S1470-2045(17)30021-9

Loveday, C., Turnbull, C., Ramsay, E., Hughes, D., Ruark, E., Frankum, J. R., Bowden, G., Kalmyrzaev, B., Warren-Perry, M., Snape, K., Adlard, J. W., Barwell, J., Berg, J., Brady, A. F., Brewer, C., Brice, G., Chapman, C., Cook, J., Davidson, R., … Rahman, N. (2011). Germline mutations in RAD51D confer susceptibility to ovarian cancer. *Nature Genetics*, *43*(9), 879–882. https://doi.org/10.1038/ng.893

Lu, H.-M., Li, S., Black, M. H., Lee, S., Hoiness, R., Wu, S., Mu, W., Huether, R., Chen, J., Sridhar, S., Tian, Y., McFarland, R., Dolinsky, J., Tippin Davis, B., Mexal, S., Dunlop, C., & Elliott, A. (2019). Association of Breast and Ovarian Cancers With Predisposition Genes Identified by Large-Scale Sequencing. *JAMA Oncology*, *5*(1), 51–57. https://doi.org/10.1001/jamaoncol.2018.2956

Lutsenko, E., & Bhagwat, A. S. (1999). Principal causes of hot spots for cytosine to thymine mutations at sites of cytosine methylation in growing cells: A model, its experimental

support and implications. *Mutation Research/Reviews in Mutation Research*, *437*(1), 11–20. https://doi.org/10.1016/S1383-5742(99)00065-4

Lynch, H. T., Silva, E., Snyder, C., & Lynch, J. F. (2008). Hereditary Breast Cancer: Part I. Diagnosing Hereditary Breast Cancer Syndromes. *The Breast Journal*, *14*(1), 3–13. https://doi.org/10.1111/j.1524-4741.2007.00515.x

Lyons, T. R., O'Brien, J., Borges, V. F., Conklin, M. W., Keely, P. J., Eliceiri, K. W., Marusyk, A., Tan, A.-C., & Schedin, P. (2011). Postpartum mammary gland involution drives progression of ductal carcinoma in situ through collagen and COX-2. *Nature Medicine*, *17*(9), 1109–1115. https://doi.org/10.1038/nm.2416

Malkin, D., Li, F. P., Strong, L. C., Fraumeni, J. F., Nelson, C. E., Kim, D. H., Kassel, J., Gryka, M. A., Bischoff, F. Z., & Tainsky, M. A. (1990). Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science (New York, N.Y.)*, *250*(4985), 1233–1238. https://doi.org/10.1126/science.1978757

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W.-M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics*, *26*(22), 2867–2873. https://doi.org/10.1093/bioinformatics/btq559

Martin, A.-M., Kanetsky, P. A., Amirimani, B., Colligon, T. A., Athanasiadis, G., Shih, H. A., Gerrero, M. R., Calzone, K., Rebbeck, T. R., & Weber, B. L. (2003). Germline TP53 mutations in breast cancer families with multiple primary cancers: Is TP53 a modifier of BRCA1? *Journal of Medical Genetics*, *40*(4), e34. https://doi.org/10.1136/jmg.40.4.e34

Mathieson, W., & Thomas, G. (2019). Using FFPE Tissue in Genomic Analyses: Advantages, Disadvantages and the Role of Biospecimen Science. *Current Pathobiology Reports*, *7*(3), 35–40. https://doi.org/10.1007/s40139-019-00194-6

Mavaddat, N., Michailidou, K., Dennis, J., Lush, M., Fachal, L., Lee, A., Tyrer, J. P., Chen, T.-H., Wang, Q., Bolla, M. K., Yang, X., Adank, M. A., Ahearn, T., Aittomäki, K., Allen, J., Andrulis, I. L., Anton-Culver, H., Antonenkova, N. N., Arndt, V., … Easton, D. F. (2019). Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *American Journal of Human Genetics*, *104*(1), 21–34. https://doi.org/10.1016/j.ajhg.2018.11.002

Mavaddat, N., Pharoah, P. D. P., Michailidou, K., Tyrer, J., Brook, M. N., Bolla, M. K., Wang, Q., Dennis, J., Dunning, A. M., Shah, M., Luben, R., Brown, J., Bojesen, S. E., Nordestgaard, B. G., Nielsen, S. F., Flyger, H., Czene, K., Darabi, H., Eriksson, M., … Garcia-Closas, M. (2015). Prediction of breast cancer risk based on profiling with common genetic variants. *Journal of the National Cancer Institute*, *107*(5), djv036. https://doi.org/10.1093/jnci/djv036

Maxwell, K. N., Wubbenhorst, B., Wenz, B. M., De Sloover, D., Pluta, J., Emery, L., Barrett, A., Kraya, A. A., Anastopoulos, I. N., Yu, S., Jiang, Y., Chen, H., Zhang, N. R., Hackman, N., D'Andrea, K., Daber, R., Morrissette, J. J. D., Mitra, N., Feldman, M., … Nathanson, K. L. (2017). BRCA locus-specific loss of heterozygosity in germline BRCA1 and BRCA2 carriers. *Nature Communications*, *8*(1), 319. https://doi.org/10.1038/s41467-017-00388-9

McCabe, N., Turner, N. C., Lord, C. J., Kluzek, K., Białkowska, A., Swift, S., Giavara, S., O'Connor, M. J., Tutt, A. N., Zdzienicka, M. Z., Smith, G. C. M., & Ashworth, A. (2006). Deficiency in the Repair of DNA Damage by Homologous Recombination and Sensitivity to Poly(ADP-Ribose) Polymerase Inhibition. *Cancer Research*, *66*(16), 8109–8115. https://doi.org/10.1158/0008-5472.CAN-06-0140

McCuaig, J. M., Armel, S. R., Novokmet, A., Ginsburg, O. M., Demsky, R., Narod, S. A., & Malkin, D. (2012). Routine TP53 testing for breast cancer under age 30: Ready for prime time? *Familial Cancer*, *11*(4), 607–613. https://doi.org/10.1007/s10689-012-9557-z

McTiernan, A., Kooperberg, C., White, E., Wilcox, S., Coates, R., Adams-Campbell, L. L., Woods, N., Ockene, J., & Women's Health Initiative Cohort Study. (2003). Recreational physical activity and the risk of breast cancer in postmenopausal women: The Women's Health Initiative Cohort Study. *JAMA*, *290*(10), 1331–1336. https://doi.org/10.1001/jama.290.10.1331

Meier, B., Volkova, N. V., Hong, Y., Schofield, P., Campbell, P. J., Gerstung, M., & Gartner, A. (2018). Mutational signatures of DNA mismatch repair deficiency in *C. elegans* and human cancers. *Genome Research*, *28*(5), 666–675. https://doi.org/10.1101/gr.226845.117

Meindl, A., Hellebrand, H., Wiek, C., Erven, V., Wappenschmidt, B., Niederacher, D., Freund, M., Lichtner, P., Hartmann, L., Schaal, H., Ramser, J., Honisch, E., Kubisch, C., Wichmann, H. E., Kast, K., Deissler, H., Engel, C., Müller-Myhsok, B., Neveling, K., … Hanenberg, H. (2010). Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. *Nature Genetics*, *42*(5), 410–414. https://doi.org/10.1038/ng.569

Mendes, A., & Fahrenkrog, B. (2019). NUP214 in Leukemia: It's More than Transport. *Cells*, *8*(1), 76. https://doi.org/10.3390/cells8010076

Mikhailova, T., Shuvalova, E., Ivanov, A., Susorov, D., Shuvalov, A., Kolosov, P. M., & Alkalaeva, E. (2017). RNA helicase DDX19 stabilizes ribosomal elongation and

termination complexes. *Nucleic Acids Research*, *45*(3), 1307–1318. https://doi.org/10.1093/nar/gkw1239

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

Miller, R. E., Leary, A., Scott, C. L., Serra, V., Lord, C. J., Bowtell, D., Chang, D. K., Garsed, D. W., Jonkers, J., Ledermann, J. A., Nik-Zainal, S., Ray-Coquard, I., Shah, S. P., Matias-Guiu, X., Swisher, E. M., & Yates, L. R. (2020). ESMO recommendations on predictive biomarker testing for homologous recombination deficiency and PARP inhibitor benefit in ovarian cancer. *Annals of Oncology*, *31*(12), 1606–1622. https://doi.org/10.1016/j.annonc.2020.08.2102

Moasser, M. M. (2007). The oncogene HER2: Its signaling and transforming functions and its role in human cancer pathogenesis. *Oncogene*, *26*(45), 6469–6487. https://doi.org/10.1038/sj.onc.1210477

Mueller, C., Haymond, A., Davis, J. B., Williams, A., & Espina, V. (2018). Protein biomarkers for subtyping breast cancer and implications for future research. *Expert Review of Proteomics*, *15*(2), 131–152. https://doi.org/10.1080/14789450.2018.1421071

Murai, J., & Pommier, Y. (2023). BRCAness, Homologous Recombination Deficiencies, and Synthetic Lethality. *Cancer Research*, *83*(8), 1173–1174. https://doi.org/10.1158/0008-5472.CAN-23-0628

Muranen, T. A., Mavaddat, N., Khan, S., Fagerholm, R., Pelttari, L., Lee, A., Aittomäki, K., Blomqvist, C., Easton, D. F., & Nevanlinna, H. (2016). Polygenic risk score is associated

with increased disease risk in 52 Finnish breast cancer families. *Breast Cancer Research and Treatment*, *158*(3), 463–469. https://doi.org/10.1007/s10549-016-3897-6

Namiki, Y., & Zou, L. (2006). ATRIP associates with replication protein A-coated ssDNA through multiple interactions. *Proceedings of the National Academy of Sciences*, *103*(3), 580–585. https://doi.org/10.1073/pnas.0510223103

Narod, S. A., Feunteun, J., Lynch, H. T., Watson, P., Conway, T., Lynch, J., & Lenoir, G. M. (1991). Familial breast-ovarian cancer locus on chromosome 17q12-q23. *Lancet (London, England)*, *338*(8759), 82–83. https://doi.org/10.1016/0140-6736(91)90076-2

Narod, S. A., & Salmena, L. (2011). BRCA1 and BRCA2 mutations and breast cancer. *Discovery Medicine*, *12*(66), 445–453.

Nassar, A., Khoor, A., Radhakrishnan, R., Radhakrishnan, A., & Cohen, C. (2014). Correlation of HER2 overexpression with gene amplification and its relation to chromosome 17 aneuploidy: A 5-year experience with invasive ductal and lobular carcinomas. *International Journal of Clinical and Experimental Pathology*, *7*(9), 6254–6261.

National Academies of Sciences, E., and Medicine. (2024). *Rethinking Race and Ethnicity in Biomedical Research*. The National Academies Press. https://doi.org/10.17226/27913

National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health. (2014). *The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General*. Centers for Disease Control and Prevention (US). http://www.ncbi.nlm.nih.gov/books/NBK179276/

Neuhausen, S. L. (2000a). Founder populations and their uses for breast cancer genetics. *Breast Cancer Research*, *2*(2), 77. https://doi.org/10.1186/bcr36

Neuhausen, S. L. (2000b). Founder populations and their uses for breast cancer genetics. *Breast Cancer Research*, *2*(2), 77. https://doi.org/10.1186/bcr36

Nevanlinna, H., & Bartek, J. (2006). The CHEK2 gene and inherited breast cancer susceptibility. *Oncogene*, *25*(43), 5912–5919. https://doi.org/10.1038/sj.onc.1209877

Newman, B., Austin, M. A., Lee, M., & King, M. C. (1988). Inheritance of human breast cancer: Evidence for autosomal dominant transmission in high-risk families. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(9), 3044–3048. https://doi.org/10.1073/pnas.85.9.3044

Ng, A. S., & Chan, D. K. H. (2023). Commonalities and differences in the mutational signature and somatic driver mutation landscape across solid and hollow viscus organs. *Oncogene*, *42*(37), 2713–2724. https://doi.org/10.1038/s41388-023-02802-7

Nichols, H. B., Schoemaker, M. J., Cai, J., Xu, J., Wright, L. B., Brook, M. N., Jones, M. E., Adami, H.-O., Baglietto, L., Bertrand, K. A., Blot, W. J., Boutron-Ruault, M.-C., Dorronsoro, M., Dossus, L., Eliassen, A. H., Giles, G. G., Gram, I. T., Hankinson, S. E., Hoffman-Bolton, J., … Sandler, D. P. (2019). Breast Cancer Risk After Recent Childbirth: A Pooled Analysis of 15 Prospective Studies. *Annals of Internal Medicine*, *170*(1), 22–30. https://doi.org/10.7326/M18-1323

Nicolini, A., Ferrari, P., Diodati, L., & Carpi, A. (2017). Recent Advances in Comprehending the Signaling Pathways Involved in the Progression of Breast Cancer. *International Journal of Molecular Sciences*, *18*(11), Article 11. https://doi.org/10.3390/ijms18112321

Nielsen, F. C., van Overeem Hansen, T., & Sørensen, C. S. (2016). Hereditary breast and ovarian cancer: New genes in confined pathways. *Nature Reviews. Cancer*, *16*(9), 599–612. https://doi.org/10.1038/nrc.2016.72

Nik-Zainal, S., Alexandrov, L. B., Wedge, D. C., Van Loo, P., Greenman, C. D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L. A., Menzies, A., Martin, S., Leung, K., Chen, L., Leroy, C., Ramakrishna, M., Rance, R., Lau, K. W., Mudie, L. J., … Breast Cancer Working Group of the International Cancer Genome Consortium. (2012). Mutational processes molding the genomes of 21 breast cancers. *Cell*, *149*(5), 979–993. https://doi.org/10.1016/j.cell.2012.04.024

Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., Martincorena, I., Alexandrov, L. B., Martin, S., Wedge, D. C., Van Loo, P., Ju, Y. S., Smid, M., Brinkman, A. B., Morganella, S., Aure, M. R., Lingjærde, O. C., Langerød, A., Ringnér, M., … Stratton, M. R. (2016). Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*, *534*(7605), 47–54. https://doi.org/10.1038/nature17676

Nik-Zainal, S., & Morganella, S. (2017). Mutational Signatures in Breast Cancer: The Problem at the DNA Level. *Clinical Cancer Research*, *23*(11), 2617–2629. https://doi.org/10.1158/1078-0432.CCR-16-2810

Nik-Zainal, S., Van Loo, P., Wedge, D. C., Alexandrov, L. B., Greenman, C. D., Lau, K. W., Raine, K., Jones, D., Marshall, J., Ramakrishna, M., Shlien, A., Cooke, S. L., Hinton, J., Menzies, A., Stebbings, L. A., Leroy, C., Jia, M., Rance, R., Mudie, L. J., … Campbell, P. J. (2012). The Life History of 21 Breast Cancers. *Cell*, *149*(5), 994–1007. https://doi.org/10.1016/j.cell.2012.04.023

Nones, K., Johnson, J., Newell, F., Patch, A. M., Thorne, H., Kazakoff, S. H., de Luca, X. M., Parsons, M. T., Ferguson, K., Reid, L. E., McCart Reed, A. E., Srihari, S., Lakis, V., Davidson, A. L., Mukhopadhyay, P., Holmes, O., Xu, Q., Wood, S., Leonard, C., … Simpson, P. T. (2019). Whole-genome sequencing reveals clinically relevant insights into

the aetiology of familial breast cancers. *Annals of Oncology: Official Journal of the European Society for Medical Oncology*, *30*(7), 1071–1079. https://doi.org/10.1093/annonc/mdz132

Norquist, B. M., Harrell, M. I., Brady, M. F., Walsh, T., Lee, M. K., Gulsuner, S., Bernards, S. S., Casadei, S., Yi, Q., Burger, R. A., Chan, J. K., Davidson, S. A., Mannel, R. S., DiSilvestro, P. A., Lankes, H. A., Ramirez, N. C., King, M. C., Swisher, E. M., & Birrer, M. J. (2016). Inherited Mutations in Women With Ovarian Carcinoma. *JAMA Oncology*, *2*(4), 482. https://doi.org/10.1001/jamaoncol.2015.5495

Norum, J. H., Andersen, K., & Sørlie, T. (2014). Lessons learned from the intrinsic subtypes of breast cancer in the quest for precision therapy. *BJS (British Journal of Surgery)*, *101*(8), 925–938. https://doi.org/10.1002/bjs.9562

Nowak, M. A., Michor, F., Komarova, N. L., & Iwasa, Y. (2004). Evolutionary dynamics of tumor suppressor gene inactivation. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(29), 10635–10638. https://doi.org/10.1073/pnas.0400747101

O'Driscoll, M., Ruiz-Perez, V. L., Woods, C. G., Jeggo, P. A., & Goodship, J. A. (2003). A splicing mutation affecting expression of ataxia–telangiectasia and Rad3–related protein (ATR) results in Seckel syndrome. *Nature Genetics*, *33*(4), 497–501. https://doi.org/10.1038/ng1129

Ogi, T., Walker, S., Stiff, T., Hobson, E., Limsirichaikul, S., Carpenter, G., Prescott, K., Suri, M., Byrd, P. J., Matsuse, M., Mitsutake, N., Nakazawa, Y., Vasudevan, P., Barrow, M., Stewart, G. S., Taylor, A. M. R., O'Driscoll, M., & Jeggo, P. A. (2012). Identification of the First ATRIP–Deficient Patient and Novel Mutations in ATR Define a Clinical

Spectrum for ATR–ATRIP Seckel Syndrome. *PLoS Genetics*, *8*(11), e1002945. https://doi.org/10.1371/journal.pgen.1002945

Oliver, J., Quezada Urban, R., Franco Cortés, C. A., Díaz Velásquez, C. E., Montealegre Paez, A. L., Pacheco-Orozco, R. A., Castro Rojas, C., García-Robles, R., López Rivera, J. J., Gaitán Chaparro, S., Gómez, A. M., Suarez Obando, F., Giraldo, G., Maya, M. I., Hurtado-Villa, P., Sanchez, A. I., Serrano, N., Orduz Galvis, A. I., Aruachan, S., … Vaca Paniagua, F. (2019). Latin American Study of Hereditary Breast and Ovarian Cancer LACAM: A Genomic Epidemiology Approach. *Frontiers in Oncology*, *9*, 1429. https://doi.org/10.3389/fonc.2019.01429

Omichessan, H., Severi, G., & Perduca, V. (2019). Computational tools to detect signatures of mutational processes in DNA from tumours: A review and empirical comparison of performance. *PLOS ONE*, *14*(9), e0221235. https://doi.org/10.1371/journal.pone.0221235

Osborne, C. K., Yochmowitz, M. G., Knight, W. A., & McGuire, W. L. (1980). The value of estrogen and progesterone receptors in the treatment of breast cancer. *Cancer*, *46*(12 Suppl), 2884–2888. https://doi.org/10.1002/1097-0142(19801215)46:12+<2884::aid-cncr2820461429>3.0.co;2-u

Ozoran, E., Trabulus, F. D. C., Erhan, D., Batar, B., & Guven, M. (2022). Association of XRCC3, XRCC4, BAX, and BCL-2 Polymorphisms with the Risk of Breast Cancer. *International Journal of Breast Cancer*, *2022*, 5817841. https://doi.org/10.1155/2022/5817841

Pedram, A., Razandi, M., Evinger, A. J., Lee, E., & Levin, E. R. (2009). Estrogen Inhibits ATR Signaling to Cell Cycle Checkpoints and DNA Repair. *Molecular Biology of the Cell*, *20*(14), 3374–3389. https://doi.org/10.1091/mbc.E09-01-0085

Pe'er, I., de Bakker, P. I. W., Maller, J., Yelensky, R., Altshuler, D., & Daly, M. J. (2006). Evaluating and improving power in whole-genome association studies using fixed marker sets. *Nature Genetics*, *38*(6), 663–667. https://doi.org/10.1038/ng1816

Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., Fluge, Ø., Pergamenschikov, A., Williams, C., Zhu, S. X., Lønning, P. E., Børresen-Dale, A.-L., Brown, P. O., & Botstein, D. (2000). Molecular portraits of human breast tumours. *Nature*, *406*(6797), Article 6797. https://doi.org/10.1038/35021093

Perry, G., Dadiani, M., Kahana-Edwin, S., Pavlovski, A., Markus, B., Hornung, G., Balint-Lahat, N., Yosepovich, A., Hout-Siloni, G., Jacob-Hirsch, J., Sklair-Levy, M., Friedman, E., Barshack, I., Kaufman, B., Gal-Yam, E. N., & Paluch-Shimon, S. (2022). Divergence of mutational signatures in association with breast cancer subtype. *Molecular Carcinogenesis*, *61*(11), 1056–1070. https://doi.org/10.1002/mc.23461

Petersen, B. S., Fredrich, B., Hoeppner, M. P., Ellinghaus, D., & Franke, A. (2017). Opportunities and challenges of whole-genome and -exome sequencing. *BMC Genetics*, *18*(1), 14. https://doi.org/10.1186/s12863-017-0479-5

Pfeifer, G. P., Denissenko, M. F., Olivier, M., Tretyakova, N., Hecht, S. S., & Hainaut, P. (2002). Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene*, *21*(48), 7435–7451. https://doi.org/10.1038/sj.onc.1205803

Pietragalla, A., Arcieri, M., Marchetti, C., Scambia, G., & Fagotti, A. (2020). Ovarian cancer predisposition beyond BRCA1 and BRCA2 genes. *International Journal of Gynecologic Cancer*, *30*(11), 1803–1810. https://doi.org/10.1136/ijgc-2020-001556

Polak, P., Kim, J., Braunstein, L. Z., Tiao, G., Karlic, R., Rosebrock, D., Livitz, D., Kübler, K., Mouw, K. W., Haradhvala, N. J., Kamburov, A., Maruvka, Y. E., Leshchiner, I., Lander, E. S., Golub, T., Zick, A., Orthwein, A., Lawrence, M. S., Batra, R. N., … Getz, G. (2017). A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. *Nature Genetics*, *49*(10), 1476–1486. https://doi.org/10.1038/ng.3934

Popova, T., Manié, E., Rieunier, G., Caux-Moncoutier, V., Tirapo, C., Dubois, T., Delattre, O., Sigal-Zafrani, B., Bollet, M., Longy, M., Houdayer, C., Sastre-Garau, X., Vincent-Salomon, A., Stoppa-Lyonnet, D., & Stern, M.-H. (2012). Ploidy and large-scale genomic instability consistently identify basal-like breast carcinomas with BRCA1/2 inactivation. *Cancer Research*, *72*(21), 5454–5462. https://doi.org/10.1158/0008-5472.CAN-12-1470

Powell, E. J., Graham, J., Ellinwood, N. M., Hostetter, J., Yaeger, M., Ho, C.-S., Gault, L., Norlin, V., Snella, E. N., Jens, J., Waide, E. H., Boettcher, A. N., Kerrigan, M., Rowland, R. R. R., Ross, J. W., Dekkers, J. C. M., & Tuggle, C. K. (2017). T Cell Lymphoma and Leukemia in Severe Combined Immunodeficiency Pigs following Bone Marrow Transplantation: A Case Report. *Frontiers in Immunology*, *8*. https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2017.00813

Prentice, L. M., Miller, R. R., Knaggs, J., Mazloomian, A., Aguirre Hernandez, R., Franchini, P., Parsa, K., Tessier-Cloutier, B., Lapuk, A., Huntsman, D., Schaeffer, D. F., & Sheffield,

B. S. (2018). Formalin fixation increases deamination mutation signature but should not lead to false positive mutations in clinical practice. *PloS One*, *13*(4), e0196434. https://doi.org/10.1371/journal.pone.0196434

Pujol, P., Barberis, M., Beer, P., Friedman, E., Piulats, J. M., Capoluongo, E. D., Garcia Foncillas, J., Ray-Coquard, I., Penault-Llorca, F., Foulkes, W. D., Turnbull, C., Hanson, H., Narod, S., Arun, B. K., Aapro, M. S., Mandel, J.-L., Normanno, N., Lambrechts, D., Vergote, I., … Alés Martínez, J. E. (2021). Clinical practice guidelines for BRCA1 and BRCA2 genetic testing. *European Journal of Cancer*, *146*, 30–47. https://doi.org/10.1016/j.ejca.2020.12.023

Pulst, S. M. (1999). Genetic Linkage Analysis. *Archives of Neurology*, *56*(6), 667–672. https://doi.org/10.1001/archneur.56.6.667

Pylkäs, K., Vuorela, M., Otsukka, M., Kallioniemi, A., Jukkola-Vuorinen, A., & Winqvist, R. (2012). Rare Copy Number Variants Observed in Hereditary Breast Cancer Cases Disrupt Genes in Estrogen Signaling and TP53 Tumor Suppression Network. *PLoS Genetics*, *8*(6), e1002734. https://doi.org/10.1371/journal.pgen.1002734

Qing, T., Wang, X., Jun, T., Ding, L., Pusztai, L., & Huang, K.-L. (2021). Genomic Determinants of Homologous Recombination Deficiency across Human Cancers. *Cancers*, *13*(18), 4572. https://doi.org/10.3390/cancers13184572

Radhakrishnan, S. K., Jette, N., & Lees-Miller, S. P. (2014). Non-homologous end joining: Emerging themes and unanswered questions. *DNA Repair*, *17*, 2–8. https://doi.org/10.1016/j.dnarep.2014.01.009

Rahman, N., Seal, S., Thompson, D., Kelly, P., Renwick, A., Elliott, A., Reid, S., Spanova, K., Barfoot, R., Chagtai, T., Jayatilake, H., McGuffog, L., Hanks, S., Evans, D. G., Eccles,

D., Easton, D. F., & Stratton, M. R. (2007). PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nature Genetics*, *39*(2), 165–167. https://doi.org/10.1038/ng1959

Rakha, E. A., Green, A. R., Powe, D. G., Roylance, R., & Ellis, I. O. (2006). Chromosome 16 tumor-suppressor genes in breast cancer. *Genes, Chromosomes & Cancer*, *45*(6), 527–535. https://doi.org/10.1002/gcc.20318

Ramus, S. J., Song, H., Dicks, E., Tyrer, J. P., Rosenthal, A. N., Intermaggio, M. P., Fraser, L., Gentry-Maharaj, A., Hayward, J., Philpott, S., Anderson, C., Edlund, C. K., Conti, D., Harrington, P., Barrowdale, D., Bowtell, D. D., Alsop, K., Mitchell, G., Cicek, M. S., … Gayther, S. A. (2015). Germline Mutations in the BRIP1, BARD1, PALB2, and NBN Genes in Women With Ovarian Cancer. *JNCI: Journal of the National Cancer Institute*, *107*(11). https://doi.org/10.1093/jnci/djv214

Ratajska, M., Antoszewska, E., Piskorz, A., Brozek, I., Borg, Å., Kusmierek, H., Biernat, W., & Limon, J. (2012). Cancer predisposing BARD1 mutations in breast-ovarian cancer families. *Breast Cancer Research and Treatment*, *131*(1), 89–97. https://doi.org/10.1007/s10549-011-1403-8

Rebbeck, T. R., Kauff, N. D., & Domchek, S. M. (2009). Meta-analysis of Risk Reduction Estimates Associated With Risk-Reducing Salpingo-oophorectomy in BRCA1 or BRCA2 Mutation Carriers. *JNCI Journal of the National Cancer Institute*, *101*(2), 80–87. https://doi.org/10.1093/jnci/djn442

Reid, S., Schindler, D., Hanenberg, H., Barker, K., Hanks, S., Kalb, R., Neveling, K., Kelly, P., Seal, S., Freund, M., Wurm, M., Batish, S. D., Lach, F. P., Yetgin, S., Neitzel, H., Ariffin, H., Tischkowitz, M., Mathew, C. G., Auerbach, A. D., & Rahman, N. (2007).

Biallelic mutations in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer. *Nature Genetics*, *39*(2), 162–164. https://doi.org/10.1038/ng1947

Rempel, E., Kluck, K., Beck, S., Ourailidis, I., Kazdal, D., Neumann, O., Volckmar, A. L., Kirchner, M., Goldschmid, H., Pfarr, N., Weichert, W., Hübschmann, D., Fröhling, S., Sutter, C., Schaaf, C. P., Schirmacher, P., Endris, V., Stenzinger, A., & Budczies, J. (2022). Pan-cancer analysis of genomic scar patterns caused by homologous repair deficiency (HRD). *Npj Precision Oncology*, *6*(1), 1–13. https://doi.org/10.1038/s41698-022-00276-6

Renwick, A., Thompson, D., Seal, S., Kelly, P., Chagtai, T., Ahmed, M., North, B., Jayatilake, H., Barfoot, R., Spanova, K., McGuffog, L., Evans, D. G., Eccles, D., Easton, D. F., Stratton, M. R., & Rahman, N. (2006). ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nature Genetics*, *38*(8), 873–875. https://doi.org/10.1038/ng1837

Riballo, E., Kühne, M., Rief, N., Doherty, A., Smith, G. C. M., Recio, M.-J., Reis, C., Dahm, K., Fricke, A., Krempler, A., Parker, A. R., Jackson, S. P., Gennery, A., Jeggo, P. A., & Löbrich, M. (2004). A pathway of double-strand break rejoining dependent upon ATM, Artemis, and proteins locating to gamma-H2AX foci. *Molecular Cell*, *16*(5), 715–724. https://doi.org/10.1016/j.molcel.2004.10.029

Rojas, K., & Stuckey, A. (2016). Breast Cancer Epidemiology and Risk Factors. *Clinical Obstetrics and Gynecology*, *59*(4), 651. https://doi.org/10.1097/GRF.0000000000000239

Rose, M., Burgess, J. T., O'Byrne, K., Richard, D. J., & Bolderson, E. (2020). PARP Inhibitors: Clinical Relevance, Mechanisms of Action and Tumor Resistance. *Frontiers in Cell and*

Developmental Biology, *8*. https://www.frontiersin.org/articles/10.3389/fcell.2020.564601

Rossouw, J. E., Anderson, G. L., Prentice, R. L., LaCroix, A. Z., Kooperberg, C., Stefanick, M. L., Jackson, R. D., Beresford, S. A. A., Howard, B. V., Johnson, K. C., Kotchen, J. M., Ockene, J., & Writing Group for the Women's Health Initiative Investigators. (2002). Risks and benefits of estrogen plus progestin in healthy postmenopausal women: Principal results From the Women's Health Initiative randomized controlled trial. *JAMA*, *288*(3), 321–333. https://doi.org/10.1001/jama.288.3.321

Rusch, M., Nakitandwe, J., Shurtleff, S., Newman, S., Zhang, Z., Edmonson, M. N., Parker, M., Jiao, Y., Ma, X., Liu, Y., Gu, J., Walsh, M. F., Becksfort, J., Thrasher, A., Li, Y., McMurry, J., Hedlund, E., Patel, A., Easton, J., … Zhang, J. (2018). Clinical cancer genomic profiling by three-platform sequencing of whole genome, whole exome and transcriptome. *Nature Communications*, *9*(1), 3962. https://doi.org/10.1038/s41467-018-06485-7

Schon, K., & Tischkowitz, M. (2018). Clinical implications of germline mutations in breast cancer: TP53. *Breast Cancer Research and Treatment*, *167*(2), 417–423. https://doi.org/10.1007/s10549-017-4531-y

Scully, R., Panday, A., Elango, R., & Willis, N. A. (2019). DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nature Reviews. Molecular Cell Biology*, *20*(11), 698–714. https://doi.org/10.1038/s41580-019-0152-0

Seung, S. J., Traore, A. N., Pourmirza, B., Fathers, K. E., Coombes, M., & Jerzak, K. J. (2020). A Population-Based Analysis of Breast Cancer Incidence and Survival by Subtype in Ontario Women. *Current Oncology*, *27*(2), 191–198. https://doi.org/10.3747/co.27.5769

Shahbandi, A., Nguyen, H. D., & Jackson, J. G. (2020). TP53 Mutations and Outcomes in Breast Cancer: Reading beyond the Headlines. *Trends in Cancer*, *6*(2), 98–110. https://doi.org/10.1016/j.trecan.2020.01.007

Shahi, R. B., De Brakeleer, S., Caljon, B., Pauwels, I., Bonduelle, M., Joris, S., Fontaine, C., Vanhoeij, M., Van Dooren, S., Teugels, E., & De Grève, J. (2019). Identification of candidate cancer predisposing variants by performing whole-exome sequencing on index patients from BRCA1 and BRCA2-negative breast cancer families. *BMC Cancer*, *19*(1), 313. https://doi.org/10.1186/s12885-019-5494-7

Sham, P. C., & Purcell, S. M. (2014). Statistical power and significance testing in large-scale genetic studies. *Nature Reviews. Genetics*, *15*(5), 335–346. https://doi.org/10.1038/nrg3706

Shimelis, H., LaDuca, H., Hu, C., Hart, S. N., Na, J., Thomas, A., Akinhanmi, M., Moore, R. M., Brauch, H., Cox, A., Eccles, D. M., Ewart-Toland, A., Fasching, P. A., Fostira, F., Garber, J., Godwin, A. K., Konstantopoulou, I., Nevanlinna, H., Sharma, P., … Couch, F. J. (2018). Triple-Negative Breast Cancer Risk Genes Identified by Multigene Hereditary Cancer Panel Testing. *Journal of the National Cancer Institute*, *110*(8), 855–862. https://doi.org/10.1093/jnci/djy106

Shiovitz, S., & Korde, L. A. (2015). Genetics of breast cancer: A topic in evolution. *Annals of Oncology: Official Journal of the European Society for Medical Oncology*, *26*(7), 1291–1299. https://doi.org/10.1093/annonc/mdv022

Silwal-Pandit, L., Vollan, H. K. M., Chin, S.-F., Rueda, O. M., McKinney, S., Osako, T., Quigley, D. A., Kristensen, V. N., Aparicio, S., Børresen-Dale, A.-L., Caldas, C., & Langerød, A. (2014). TP53 mutation spectrum in breast cancer is subtype specific and

has distinct prognostic relevance. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, *20*(13), 3569–3580. https://doi.org/10.1158/1078-0432.CCR-13-2943

Skol, A. D., Sasaki, M. M., & Onel, K. (2016). The genetics of breast cancer risk in the post-genome era: Thoughts on study design to move past BRCA and towards clinical relevance. *Breast Cancer Research*, *18*(1), 99. https://doi.org/10.1186/s13058-016-0759-4

Slamon, D. J., Clark, G. M., Wong, S. G., Levin, W. J., Ullrich, A., & McGuire, W. L. (1987). Human breast cancer: Correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science (New York, N.Y.)*, *235*(4785), 177–182. https://doi.org/10.1126/science.3798106

Smigal, C., Jemal, A., Ward, E., Cokkinides, V., Smith, R., Howe, H. L., & Thun, M. (2006). Trends in Breast Cancer by Race and Ethnicity: Update 2006. *CA Cancer J Clin*, *55*, 168–183.

Smith, P., McGuffog, L., Easton, D. F., Mann, G. J., Pupo, G. M., Newman, B., Chenevix-Trench, G., kConFab Investigators, Szabo, C., Southey, M., Renard, H., Odefrey, F., Lynch, H., Stoppa-Lyonnet, D., Couch, F., Hopper, J. L., Giles, G. G., McCredie, M. R. E., Buys, S., … Stratton, M. R. (2006). A genome wide linkage search for breast cancer susceptibility genes. *Genes, Chromosomes & Cancer*, *45*(7), 646–655. https://doi.org/10.1002/gcc.20330

Snape, K., Ruark, E., Tarpey, P., Renwick, A., Turnbull, C., Seal, S., Murray, A., Hanks, S., Douglas, J., Stratton, M. R., & Rahman, N. (2012). Predisposition gene identification in common cancers by exome sequencing: Insights from familial breast cancer. *Breast*

Cancer Research and Treatment, *134*(1), 429–433. https://doi.org/10.1007/s10549-012-2057-x

Sokolenko, A. P., Suspitsin, E. N., Kuligina, E. S., Bizin, I. V., Frishman, D., & Imyanitov, E. N. (2015). Identification of novel hereditary cancer genes by whole exome sequencing. *Cancer Letters*, *369*(2), 274–288. https://doi.org/10.1016/j.canlet.2015.09.014

Somyajit, K., Mishra, A., Jameei, A., & Nagaraju, G. (2015). Enhanced non-homologous end joining contributes toward synthetic lethality of pathological RAD51C mutants with poly (ADP-ribose) polymerase. *Carcinogenesis*, *36*(1), 13–24. https://doi.org/10.1093/carcin/bgu211

Sondka, Z., Bamford, S., Cole, C. G., Ward, S. A., Dunham, I., & Forbes, S. A. (2018). The COSMIC Cancer Gene Census: Describing genetic dysfunction across all human cancers. *Nature Reviews Cancer*, *18*(11), 696–705. https://doi.org/10.1038/s41568-018-0060-1

Sørensen, C. S., Hansen, L. T., Dziegielewski, J., Syljuåsen, R. G., Lundin, C., Bartek, J., & Helleday, T. (2005). The cell-cycle checkpoint kinase Chk1 is required for mammalian homologous recombination repair. *Nature Cell Biology*, *7*(2), 195–201. https://doi.org/10.1038/ncb1212

Sørlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Thorsen, T., Quist, H., Matese, J. C., Brown, P. O., Botstein, D., Lønning, P. E., & Børresen-Dale, A.-L. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences*, *98*(19), 10869–10874. https://doi.org/10.1073/pnas.191367098

Stacey, S. N., Manolescu, A., Sulem, P., Rafnar, T., Gudmundsson, J., Gudjonsson, S. A., Masson, G., Jakobsdottir, M., Thorlacius, S., Helgason, A., Aben, K. K., Strobbe, L. J., Albers-Akkers, M. T., Swinkels, D. W., Henderson, B. E., Kolonel, L. N., Le Marchand, L., Millastre, E., Andres, R., … Stefansson, K. (2007). Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nature Genetics*, *39*(7), 865–869. https://doi.org/10.1038/ng2064

Stackpole, M. M., Wise, S. S., Goodale, B. C., Duzevik, E. G., Munroe, R. C., Thompson, W. D., Thacker, J., Thompson, L. H., Hinz, J. M., & Wise, J. P. (2007). Homologous recombination repair protects against particulate chromate-induced chromosome instability in Chinese hamster cells. *Mutation Research*, *625*(1–2), 145–154. https://doi.org/10.1016/j.mrfmmm.2007.06.003

Stronach, E. A., Paul, J., Timms, K. M., Hughes, E., Brown, K., Neff, C., Perry, M., Gutin, A., El-Bahrawy, M., Steel, J. H., Liu, X., Lewsley, L.-A., Siddiqui, N., Gabra, H., Lanchbury, J. S., & Brown, R. (2018). Biomarker Assessment of HR Deficiency, Tumor BRCA1/2 Mutations, and CCNE1 Copy Number in Ovarian Cancer: Associations with Clinical Outcome Following Platinum Monotherapy. *Molecular Cancer Research*, *16*(7), 1103–1111. https://doi.org/10.1158/1541-7786.MCR-18-0034

Subramani, R., & Lakshmanaswamy, R. (2017). Pregnancy and Breast Cancer. *Progress in Molecular Biology and Translational Science*, *151*, 81–111. https://doi.org/10.1016/bs.pmbts.2017.07.006

Sun, Y.-S., Zhao, Z., Yang, Z.-N., Xu, F., Lu, H.-J., Zhu, Z.-Y., Shi, W., Jiang, J., Yao, P.-P., & Zhu, H.-P. (2017). Risk Factors and Preventions of Breast Cancer. *International Journal of Biological Sciences*, *13*(11), 1387–1397. https://doi.org/10.7150/ijbs.21635

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: A Cancer Journal for Clinicians*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660

Suwaki, N., Klare, K., & Tarsounas, M. (2011). RAD51 paralogs: Roles in DNA damage signalling, recombinational repair and tumorigenesis. *Seminars in Cell & Developmental Biology*, *22*(8), 898–905. https://doi.org/10.1016/j.semcdb.2011.07.019

Swift, M., Chase, C. L., & Morrell, D. (1990). Cancer predisposition of ataxia-telangiectasia heterozygotes. *Cancer Genetics and Cytogenetics*, *46*(1), 21–27. https://doi.org/10.1016/0165-4608(90)90004-t

Sztupinszki, Z., Diossy, M., Krzystanek, M., Reiniger, L., Csabai, I., Favero, F., Birkbak, N. J., Eklund, A. C., Syed, A., & Szallasi, Z. (2018). Migrating the SNP array-based homologous recombination deficiency measures to next generation sequencing data of breast cancer. *Npj Breast Cancer*, *4*(1), 1–4. https://doi.org/10.1038/s41523-018-0066-6

Szwiec, M., Jakubowska, A., Górski, B., Huzarski, T., Tomiczek-Szwiec, J., Gronwald, J., Dębniak, T., Byrski, T., Kluźniak, W., Wokołorczyk, D., Birkenfeld, B., Akbari, M. r., Narod, S. a., Lubiński, J., & Cybulski, C. (2015). Recurrent mutations of BRCA1 and BRCA2 in Poland: An update. *Clinical Genetics*, *87*(3), 288–292. https://doi.org/10.1111/cge.12360

Szymanska-Pasternak, J., Szymanska, A., Medrek, K., Imyanitov, E. N., Cybulski, C., Gorski, B., Magnowski, P., Dziuba, I., Gugala, K., Debniak, B., Gozdz, S., Sokolenko, A. P., Krylova, N. Y., Lobeiko, O. S., Narod, S. A., & Lubinski, J. (2006). CHEK2 variants

predispose to benign, borderline and low-grade invasive ovarian tumors. *Gynecologic Oncology*, *102*(3), 429–431. https://doi.org/10.1016/j.ygyno.2006.05.040

Tai, Y. C., Domchek, S., Parmigiani, G., & Chen, S. (2007). Breast cancer risk among male BRCA1 and BRCA2 mutation carriers. *Journal of the National Cancer Institute*, *99*(23), 1811–1814. https://doi.org/10.1093/jnci/djm203

Takahashi, K. K., & Innan, H. (2022). Frequent somatic gene conversion as a mechanism for loss of heterozygosity in tumor suppressor genes. *Genome Research*, *32*(6), 1017–1025. https://doi.org/10.1101/gr.276617.122

Takamatsu, S., Brown, J. B., Yamaguchi, K., Hamanishi, J., Yamanoi, K., Takaya, H., Kaneyasu, T., Mori, S., Mandai, M., & Matsumura, N. (2022). Utility of Homologous Recombination Deficiency Biomarkers Across Cancer Types. *JCO Precision Oncology*, *6*, e2200085. https://doi.org/10.1200/PO.22.00085

Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., & Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics*, *20*(8), Article 8. https://doi.org/10.1038/s41576-019-0127-1

Tan, D. S. P., Rothermundt, C., Thomas, K., Bancroft, E., Eeles, R., Shanley, S., Ardern-Jones, A., Norman, A., Kaye, S. B., & Gore, M. E. (2008). "BRCAness" syndrome in ovarian cancer: A case-control study describing the clinical features and outcome of patients with epithelial ovarian cancer associated with BRCA1 and BRCA2 mutations. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, *26*(34), 5530–5536. https://doi.org/10.1200/JCO.2008.16.1703

Tanaka, A., Weinel, S., Nagy, N., O'Driscoll, M., Lai-Cheong, J. E., Kulp-Shorten, C. L., Knable, A., Carpenter, G., Fisher, S. A., Hiragun, M., Yanase, Y., Hide, M., Callen, J., &

McGrath, J. A. (2012). Germline Mutation in ATR in Autosomal- Dominant Oropharyngeal Cancer Syndrome. *The American Journal of Human Genetics*, *90*(3), 511–517. https://doi.org/10.1016/j.ajhg.2012.01.007

Telli, M. L., Metzger, O., Timms, K., Evans, B., Vogel, D., Wei, H., Jones, J. T., Wenstrup, R. J., McKee, M. D., Sullivan, D. M., Fallstrom, M., Maag, D., Ansell, P. J., Sohn, J., Chen, S.-T., Martinez, N., Geyer, C. E., Loibl, S., & Golshan, M. (2018). Evaluation of homologous recombination deficiency (HRD) status with pathological response to carboplatin +/- veliparib in BrighTNess, a randomized phase 3 study in early stage TNBC. *Journal of Clinical Oncology*, *36*(15_suppl), 519–519. https://doi.org/10.1200/JCO.2018.36.15_suppl.519

Telli, M. L., Timms, K. M., Reid, J., Hennessy, B., Mills, G. B., Jensen, K. C., Szallasi, Z., Barry, W. T., Winer, E. P., Tung, N. M., Isakoff, S. J., Ryan, P. D., Greene-Colozzi, A., Gutin, A., Sangale, Z., Iliev, D., Neff, C., Abkevich, V., Jones, J. T., … Richardson, A. L. (2016). Homologous Recombination Deficiency (HRD) Score Predicts Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with Triple-Negative Breast Cancer. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, *22*(15), 3764–3773. https://doi.org/10.1158/1078-0432.CCR-15-2477

Tervasmäki, A., Winqvist, R., Jukkola-Vuorinen, A., & Pylkäs, K. (2014). Recurrent CYP2C19 deletion allele is associated with triple-negative breast cancer. *BMC Cancer*, *14*(1), 902. https://doi.org/10.1186/1471-2407-14-902

Thacker, J. (2005). The RAD51 gene family, genetic instability and cancer. *Cancer Letters*, *219*(2), 125–135. https://doi.org/10.1016/j.canlet.2004.08.018

The Cancer Genome Atlas Network. (2012). Comprehensive molecular portraits of human breast tumours. *Nature*, *490*(7418), 61–70. https://doi.org/10.1038/nature11412

Thompson, E. R., Doyle, M. A., Ryland, G. L., Rowley, S. M., Choong, D. Y. H., Tothill, R. W., Thorne, H., kConFab, Barnes, D. R., Li, J., Ellul, J., Philip, G. K., Antill, Y. C., James, P. A., Trainer, A. H., Mitchell, G., & Campbell, I. G. (2012). Exome Sequencing Identifies Rare Deleterious Mutations in DNA Repair Genes FANCC and BLM as Potential Breast Cancer Susceptibility Alleles. *PLoS Genetics*, *8*(9), e1002894. https://doi.org/10.1371/journal.pgen.1002894

Timms, K. M., Abkevich, V., Hughes, E., Neff, C., Reid, J., Morris, B., Kalva, S., Potter, J., Tran, T. V., Chen, J., Iliev, D., Sangale, Z., Tikishvili, E., Perry, M., Zharkikh, A., Gutin, A., & Lanchbury, J. S. (2014). Association of BRCA1/2defects with genomic scores predictive of DNA damage repair deficiency among breast cancer subtypes. *Breast Cancer Research*, *16*(6), 475. https://doi.org/10.1186/s13058-014-0475-x

Tomida, J., Itaya, A., Shigechi, T., Unno, J., Uchida, E., Ikura, M., Masuda, Y., Matsuda, S., Adachi, J., Kobayashi, M., Meetei, A. R., Maehara, Y., Yamamoto, K. -i., Kamiya, K., Matsuura, A., Matsuda, T., Ikura, T., Ishiai, M., & Takata, M. (2013). A novel interplay between the Fanconi anemia core complex and ATR-ATRIP kinase during DNA cross-link repair. *Nucleic Acids Research*, *41*(14), 6930–6941. https://doi.org/10.1093/nar/gkt467

Tonnessen-Murray, C. A., Lozano, G., & Jackson, J. G. (2017). The Regulation of Cellular Functions by the p53 Protein: Cellular Senescence. *Cold Spring Harbor Perspectives in Medicine*, *7*(2), a026112. https://doi.org/10.1101/cshperspect.a026112

Torabi Dalivandan, S., Plummer, J., & Gayther, S. A. (2021). Risks and Function of Breast Cancer Susceptibility Alleles. *Cancers*, *13*(16), Article 16. https://doi.org/10.3390/cancers13163953

Tung, N., Lin, N. U., Kidd, J., Allen, B. A., Singh, N., Wenstrup, R. J., Hartman, A.-R., Winer, E. P., & Garber, J. E. (2016). Frequency of Germline Mutations in 25 Cancer Susceptibility Genes in a Sequential Series of Patients With Breast Cancer. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, *34*(13), 1460–1468. https://doi.org/10.1200/JCO.2015.65.0747

Tung, N. M., Robson, M. E., Ventz, S., Santa-Maria, C. A., Nanda, R., Marcom, P. K., Shah, P. D., Ballinger, T. J., Yang, E. S., Vinayak, S., Melisko, M., Brufsky, A., DeMeo, M., Jenkins, C., Domchek, S., D'Andrea, A., Lin, N. U., Hughes, M. E., Carey, L. A., … Garber, J. E. (2020). TBCRC 048: Phase II Study of Olaparib for Metastatic Breast Cancer and Mutations in Homologous Recombination-Related Genes. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, *38*(36), 4274–4282. https://doi.org/10.1200/JCO.20.02151

Turnbull, C., & Rahman, N. (2008). Genetic Predisposition to Breast Cancer: Past, Present, and Future. *Annual Review of Genomics and Human Genetics*, *9*(1), 321–345. https://doi.org/10.1146/annurev.genom.9.081307.164339

Vahteristo, P., Bartkova, J., Eerola, H., Syrjäkoski, K., Ojala, S., Kilpivaara, O., Tamminen, A., Kononen, J., Aittomäki, K., Heikkilä, P., Holli, K., Blomqvist, C., Bartek, J., Kallioniemi, O.-P., & Nevanlinna, H. (2002). A CHEK2 Genetic Variant Contributing to a Substantial Fraction of Familial Breast Cancer. *The American Journal of Human Genetics*, *71*(2), 432–438. https://doi.org/10.1086/341943

Van Hoeck, A., Tjoonk, N. H., van Boxtel, R., & Cuppen, E. (2019). Portrait of a cancer: Mutational signature analyses for cancer diagnostics. *BMC Cancer*, *19*(1), 457. https://doi.org/10.1186/s12885-019-5677-2

van Veen, E. M., Brentnall, A. R., Byers, H., Harkness, E. F., Astley, S. M., Sampson, S., Howell, A., Newman, W. G., Cuzick, J., & Evans, D. G. R. (2018). Use of Single-Nucleotide Polymorphisms and Mammographic Density Plus Classic Risk Factors for Breast Cancer Risk Prediction. *JAMA Oncology*, *4*(4), 476–482. https://doi.org/10.1001/jamaoncol.2017.4881

Venkitaraman, A. R. (2002). Cancer susceptibility and the functions of BRCA1 and BRCA2. *Cell*, *108*(2), 171–182. https://doi.org/10.1016/s0092-8674(02)00615-3

Venkitaraman, A. R. (2004). Tracing the network connecting BRCA and Fanconi anaemia proteins. *Nature Reviews. Cancer*, *4*(4), 266–276. https://doi.org/10.1038/nrc1321

Voduc, K. D., Cheang, M. C. U., Tyldesley, S., Gelmon, K., Nielsen, T. O., & Kennecke, H. (2010). Breast Cancer Subtypes and the Risk of Local and Regional Relapse. *Journal of Clinical Oncology*, *28*(10), 1684–1691. https://doi.org/10.1200/JCO.2009.24.9284

Voutsadakis, I. A. (2022). Loss of genes in chromosome arms 5q and 16q in breast cancer. *Breast Disease*, *41*(1), 331–341. https://doi.org/10.3233/BD-210047

Walsh, T., Casadei, S., Lee, M. K., Pennil, C. C., Nord, A. S., Thornton, A. M., Roeb, W., Agnew, K. J., Stray, S. M., Wickramanayake, A., Norquist, B., Pennington, K. P., Garcia, R. L., King, M.-C., & Swisher, E. M. (2011). Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(44), 18032–18037. https://doi.org/10.1073/pnas.1115052108

Wang, S., Qian, F., Zheng, Y., Ogundiran, T., Ojengbede, O., Zheng, W., Blot, W., Nathanson, K. L., Hennis, A., Nemesure, B., Ambs, S., Olopade, O. I., & Huo, D. (2018). Genetic variants demonstrating flip-flop phenomenon and breast cancer risk prediction among women of African ancestry. *Breast Cancer Research and Treatment*, *168*(3), 703–712. https://doi.org/10.1007/s10549-017-4638-1

Wang, Y., Cortez, D., Yazdi, P., Neff, N., Elledge, S. J., & Qin, J. (2000). BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. *Genes & Development*, *14*(8), 927–939.

Wang, Y., & Qin, J. (2003). MSH2 and ATR form a signaling module and regulate two branches of the damage response to DNA methylation. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(26), 15387–15392. https://doi.org/10.1073/pnas.2536810100

Wen, H., Feng, Z., Ma, Y., Liu, R., Ou, Q., Guo, Q., Shen, Y., Wu, X., Shao, Y., Bao, H., & Wu, X. (2022). Homologous recombination deficiency in diverse cancer types and its correlation with platinum chemotherapy efficiency in ovarian cancer. *BMC Cancer*, *22*(1), 550. https://doi.org/10.1186/s12885-022-09602-4

Wilcox, N., Dumont, M., González-Neira, A., Carvalho, S., Joly Beauparlant, C., Crotti, M., Luccarini, C., Soucy, P., Dubois, S., Nuñez-Torres, R., Pita, G., Gardner, E. J., Dennis, J., Alonso, M. R., Álvarez, N., Baynes, C., Collin-Deschesnes, A. C., Desjardins, S., Becher, H., … Simard, J. (2023). Exome sequencing identifies breast cancer susceptibility genes and defines the contribution of coding variants to breast cancer risk. *Nature Genetics*, *55*(9), Article 9. https://doi.org/10.1038/s41588-023-01466-z

Wilkinson, L., & Gathani, T. (2022). Understanding breast cancer as a global health concern. *The British Journal of Radiology*, *95*(1130), 20211033. https://doi.org/10.1259/bjr.20211033

Wokołorczyk, D., Kluźniak, W., Huzarski, T., Gronwald, J., Szymiczek, A., Rusak, B., Stempa, K., Gliniewicz, K., Kashyap, A., Morawska, S., Dębniak, T., Jakubowska, A., Szwiec, M., Domagała, P., Lubiński, J., Narod, S. A., Akbari, M. R., & Cybulski, C. (2020). Mutations in ATM, NBN and BRCA2 predispose to aggressive prostate cancer in Poland. *International Journal of Cancer*, *147*(10), 2793–2800. https://doi.org/10.1002/ijc.33272

Wolff, A. C., Hammond, M. E. H., Allison, K. H., Harvey, B. E., Mangu, P. B., Bartlett, J. M. S., Bilous, M., Ellis, I. O., Fitzgibbons, P., Hanna, W., Jenkins, R. B., Press, M. F., Spears, P. A., Vance, G. H., Viale, G., McShane, L. M., & Dowsett, M. (2018). Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Focused Update. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, *36*(20), 2105–2122. https://doi.org/10.1200/JCO.2018.77.8738

Woodbine, L., Brunton, H., Goodarzi, A. A., Shibata, A., & Jeggo, P. A. (2011). Endogenously induced DNA double strand breaks arise in heterochromatic DNA regions and require ataxia telangiectasia mutated and Artemis for their repair. *Nucleic Acids Research*, *39*(16), 6986–6997. https://doi.org/10.1093/nar/gkr331

Wooster, R., Neuhausen, S. L., Mangion, J., Quirk, Y., Ford, D., Collins, N., Nguyen, K., Seal, S., Tran, T., & Averill, D. (1994). Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science (New York, N.Y.)*, *265*(5181), 2088–2090. https://doi.org/10.1126/science.8091231

Wu, L. C., Wang, Z. W., Tsan, J. T., Spillman, M. A., Phung, A., Xu, X. L., Yang, M. C., Hwang, L. Y., Bowcock, A. M., & Baer, R. (1996). Identification of a RING protein that can interact in vivo with the BRCA1 gene product. *Nature Genetics*, *14*(4), 430–440. https://doi.org/10.1038/ng1296-430

Xia, B., Sheng, Q., Nakanishi, K., Ohashi, A., Wu, J., Christ, N., Liu, X., Jasin, M., Couch, F. J., & Livingston, D. M. (2006). Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. *Molecular Cell*, *22*(6), 719–729. https://doi.org/10.1016/j.molcel.2006.05.022

Yadav, S., Boddicker, N. J., Na, J., Polley, E. C., Hu, C., Hart, S. N., Gnanaolivu, R. D., Larson, N., Holtegaard, S., Huang, H., Dunn, C. A., Teras, L. R., Patel, A. V., Lacey, J. V., Neuhausen, S. L., Martinez, E., Haiman, C., Chen, F., Ruddy, K. J., … Couch, F. J. (2023). Contralateral Breast Cancer Risk Among Carriers of Germline Pathogenic Variants in ATM, BRCA1, BRCA2, CHEK2, and PALB2. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, *41*(9), 1703–1713. https://doi.org/10.1200/JCO.22.01239

Yan, S., Sorrell, M., & Berman, Z. (2014). Functional interplay between ATM/ATR-mediated DNA damage response and DNA repair pathways in oxidative stress. *Cellular and Molecular Life Sciences : CMLS*, *71*(20), 3951–3967. https://doi.org/10.1007/s00018-014-1666-4

Yanes, T., Young, M.-A., Meiser, B., & James, P. A. (2020). Clinical applications of polygenic breast cancer risk: A critical review and perspectives of an emerging field. *Breast Cancer Research*, *22*(1), 21. https://doi.org/10.1186/s13058-020-01260-3

Yang, L., Liu, L., Mi, Y., Li, J., Ma, X., Ai, X., Qin, T., Xu, Z., Wang, J., & Xiao, Z. (2011). [Relationship between RAD51-G135C/XRCC3-C241T polymorphisms and development of acute myeloid leukemia with recurrent chromosome translocation]. *Zhonghua Xue Ye Xue Za Zhi = Zhonghua Xueyexue Zazhi*, *32*(5), 299–303.

Yang, X. H., & Zou, L. (2006). Recruitment of ATR-ATRIP, Rad17, and 9-1-1 Complexes to DNA Damage. In *Methods in Enzymology* (Vol. 409, pp. 118–131). Elsevier. https://doi.org/10.1016/S0076-6879(05)09007-5

Yang, X., Leslie, G., Doroszuk, A., Schneider, S., Allen, J., Decker, B., Dunning, A. M., Redman, J., Scarth, J., Plaskocinska, I., Luccarini, C., Shah, M., Pooley, K., Dorling, L., Lee, A., Adank, M. A., Adlard, J., Aittomäki, K., Andrulis, I. L., … Tischkowitz, M. (2020). Cancer Risks Associated With Germline PALB2 Pathogenic Variants: An International Study of 524 Families. *Journal of Clinical Oncology*, *38*(7), 674–685. https://doi.org/10.1200/JCO.19.01907

Yang, X., Song, H., Leslie, G., Engel, C., Hahnen, E., Auber, B., Horváth, J., Kast, K., Niederacher, D., Turnbull, C., Houlston, R., Hanson, H., Loveday, C., Dolinsky, J. S., LaDuca, H., Ramus, S. J., Menon, U., Rosenthal, A. N., Jacobs, I., … Antoniou, A. C. (2020). Ovarian and Breast Cancer Risks Associated With Pathogenic Variants in RAD51C and RAD51D. *Journal of the National Cancer Institute*, *112*(12), 1242–1250. https://doi.org/10.1093/jnci/djaa030

Yin, L., Duan, J.-J., Bian, X.-W., & Yu, S.-C. (2020). Triple-negative breast cancer molecular subtyping and treatment progress. *Breast Cancer Research: BCR*, *22*(1), 61. https://doi.org/10.1186/s13058-020-01296-5

Zaki-Dizaji, M., Akrami, S. M., Abolhassani, H., Rezaei, N., & Aghamohammadi, A. (2017). Ataxia telangiectasia syndrome: Moonlighting ATM. *Expert Review of Clinical Immunology*, *13*(12), 1155–1172. https://doi.org/10.1080/1744666X.2017.1392856

Zámborszky, J., Szikriszt, B., Gervai, J. Z., Pipek, O., Póti, Á., Krzystanek, M., Ribli, D., Szalai-Gindl, J. M., Csabai, I., Szallasi, Z., Swanton, C., Richardson, A. L., & Szüts, D. (2017). Loss of BRCA1 or BRCA2 markedly increases the rate of base substitution mutagenesis and has distinct effects on genomic deletions. *Oncogene*, *36*(6), 746–755. https://doi.org/10.1038/onc.2016.243

Zha, S., Alt, F. W., Cheng, H.-L., Brush, J. W., & Li, G. (2007). Defective DNA repair and increased genomic instability in Cernunnos-XLF-deficient murine ES cells. *Proceedings of the National Academy of Sciences*, *104*(11), 4518–4523. https://doi.org/10.1073/pnas.0611734104