Department für Diagnostische Labormedizin
der Universität Tübingen
Institut für Medizinische Genetik und angewandte Genomik

Role of Genomic Variants in DNA Repair Genes to Modify the Age at Onset of Hereditary Breast Cancer in BRCA1 Carriers

Thesis submitted as requirement to fulfill the degree "Doctor of Philosophy" in *Experimental Medicine* (PhD)

At the Faculty of Medicine
Eberhard Karls Universität
Tübingen

Presented by

Sepahi, Ilnaz

2024

Dean:	Professor Dr. B. Pichler
1. reviewer:	Professor Dr. O. Rieß
2. reviewer:	Professor Dr. K. Schulze-Osthoff

Date of oral defense: 08.12.2023

Contents

List of Figures

List of Tables

List of Abbreviation

BC	Breast Cancer
BCCR	Breast Cancer Cluster Region
BER	Base Excision Repair
BRCA1	Breast Cancer 1
BRCA2	Breast Cancer 2
CIMBA	Consortium of International Management Business Analysis
DDR	DNA-Damage Response
DNA	Deoxyribonucleic Acid
DSBR	Double Strand Break Repair
dsDNA	Double Strand DNA
ER	Estrogenereceptor
GC-HBOC	German Consortium for Hereditary Breast and
HBC	Hereditary Breast Cancer
HBOC	Hereditary Breast and Ovarian Cancer
HER2	Human Epidermal Growth Factor Receptor 2
HR	Homologous Recombination
NER	Nucleotide Excision Repair
NHEJ	Non-Homologous End Joining
OC	Ovarian Cancer
OCCR	Ovarian Cancer Cluster Region
PR	Progesterone Receptor
RRSO	Risk-Reducing Salpingo Oophorectomy
SSA	Single Strand Annealing

1 Introduction

1.1 Breast Cancer

1.1.1 Epidemiology

Breast cancer is the most commonly diagnosed malignancy among females, and it is the second common cancer (11.6 %) among both sexes. Based on the records in 2018, annually 2,088,849 cases are diagnosed with breast cancer in the world (Bray et al., 2018). In 154 of 185 countries, breast cancer is the most frequently diagnosed cancer and is also the major cause of cancer death in 100 countries (Bray et al., 2018). Most of exceptions are countries where the lung cancer is the main cause of cancer death such as Australia/New Zealand, Northern Europe, Northern America and many countries in Sub-Saharan Africa as a result of increased cervical cancer rates (Bray et al., 2018). Breast cancer incidence rate varies by regions, with almost four-fold change from being highest in Australia/New Zealand, Northern/ Western/ Southern Europe and Norther America (94-80 per 100,000 females) to the lowest in South Central Asia (26 per 100,000 females) (Bray et al., 2018). concerning mortality, less changeability has been shown in breast cancer rate, where Fiji shows the most mortality rates. Migration studies revealed that majority observed international and inter-ethnic of diversities in the incidence rate of breast cancer are due to non-genetic factors rather than genetic factors, despite the fact that genetic factors and personal and family history of breast and ovarian cancer and inheritance of cancer predisposition genes such as *BRCA1* and *BRCA2* accounts for 5-10% of breast cancer cases. Analysis of migration patterns from population with low incidence rate such as Asian population to the high-incidence rate population such as USA population uncovered that breast cancer incidence rates elevated in successive generation (Ziegler et al., 1993). In the past few decades, the incidence rates have risen as a result of changing exposures to the reproductive and nutrition factors over time. Developing countries experienced the most dramatic rises compare to industrialized countries as a result of “westernization of lifestyles” in regard to late childbearing, having fewer children, use of oral contraceptives,

smoking, calorie-dense foods, lack of physical activity, and obesity(Bray et al., 2004; Colditz et al., 2006; Jemal et al., 2010; Kirsi Määttä, 2016; Prentice, 2009). In Germany 70,000 women are being diagnosed with breast cancer annually with a Grude incidence rate of 169.1 per 100,000 female and the mortality rate of 43.2 per 100,000 cases in 2012, and a five-year relative survival rate of 88% (https://www.krebsdaten.de/Krebs/EN/Home/homepage_node.html; GERMAN CENTRE FOR CANCER REGISTRY DATA).

1.1.2 Breast Anatomy and Histology

Each breast consists of 15-20 major duct systems which empty at nipple, and they branch out until they form Terminal Ductal Lobular Unit (TDLU) which consists of a terminal duct and many small ductules (or acini). Cuboidal to columnar epithelial cells and an outer layer of myoepithelial cells line up the ducts and ductules.

1.1.3 Risk Factors

1.1.3.1 GENERAL

Age. The incidence of breast cancer is increasing with aging. Only about 10% of breast cancer cases occur in women younger than 45 years (National Cancer Institute, Surveillance, Epidemiology and End Results Program).

Gender. Being a woman is the biggest risk factor regarding breast cancer while men also develop breast cancer at an incidence rate of 1.1 per 100,000 cases women develop breast cancer at an incidence rate of 126 per 100,000 (National Cancer Institute, Surveillance, Epidemiology and End Results Program).

Ethnicity. The incidence rate is higher among white non-Hispanic women in comparison to Black, Asian/pacific Islanders and American Indian/ Alaska Natives and Hispanic women, while the mortality rate is higher in Black women (National Cancer Institute, Surveillance, Epidemiology and End Results Program). In addition, African American women are more likely to develop triple-negative breast cancer which is more aggressive than the other breast cancer types (https://www.breastcancer.org/risk/factors/race_ethnicity accessed 03.12.2018).

Lifestyle. Even light alcohol consumption increases the risk of breast cancer. Smoking and lack of physical activity is linked to the increased risk of breast cancer. Several studies demonstrated that physical activity leads to a reduced breast cancer risk in active women compared to inactive women; a meta-analysis of 31 prospective studies suggested 12% reduction in the risk of breast cancer among active women (Y. Wu et al., 2013). The association between obesity and incidence of many cancer types has been previously studied (Renehan et al., 2008). Epidemiological studies revealed a link between increased body mass index (BMI) and breast cancer risk in postmenopausal women (Key et al., 2003) partly due to elevated serum concentration of bioavailable estradiol, which in terms result in a growth in production of estrogen by aromatase in the adipose tissue (Key et al., 2003).

Previous Treatments. Radiation to the chest for therapeutic purposes for cancers other than breast cancer (such as Hodgkin or non-Hodgkin lymphoma) before the age 30 increases the risk of breast cancer (Clemons et al., 2000) (https://www.breastcancer.org/risk/factors/radiation accessed on 16.01.2019 and https://cdc.gov/cancer/breast/basics_info/risk_factors.htm accessed 16.01.2019). Results of Boice and colleagues' studies on women exposed to ionizing radiation either as a therapeutic procedure among tuberculosis patients or among Hiroshima and Nagasaki atomic bombs survivors demonstrated that that numerous fluoroscopies are significant risk factors for breast cancer (Boice et al., 1991; Land et al., 1980). Conclusion from the data collected from 2,573 women who were tested by X-ray fluoroscopy an average of 88 times in a course of therapy for tuberculosis and were followed up for an average of 30 years was that the younger women were at higher risk and the relative risk of 1Gy radiation was estimated as 1.61 when allowing for 10-year latency.

Oral contraceptive consumption in the last ten years may slightly increase the risk of breast cancer. The risk is dropping once the uptake is stopped (https://www.cancercenter.com/breast-cancer/risk-factors/ accessed on 16.01.2019) (McPherson et al., 2000). Diethylstilbestrol, a drug which was prescribed to pregnant women in 1940s to 1970s to prevent abortion marginally

elevated the risk of breast cancer (https://www.cancercenter.com/breast-cancer/risk-factors/ accessed on 16.01.2019) (Colton et al., 1993). Hormone replacement therapy (HRT) which is used to reduce the postmenopausal symptoms and prevent osteoporosis has been considered as a risk factor for breast cancer by International Agency for Research on Cancer (IARC, List of classifications, Volumes 1-123, accessed 17.01.2019). Results of a cohort studies demonstrated that breast cancer risk is 55-100% higher in current combination hormone therapy users compared with never users (Beral & Million Women Study Collaborators, 2003; Chlebowski et al., 2013). Estrogen-only HRT also increases the breast cancer risk but to a lower extent compared with combined HRT(Beral & Million Women Study Collaborators, 2003; Ross et al., 2000).

Reproductive Factors. Longer exposure to the estrogen either due to younger age at menarche (by 5% for each year younger at menarche) or older age at menopause (by around 3% for each year older at menopause) increases the risk of breast cancer (Collaborative Group on Hormonal Factors in Breast Cancer, 2012). Having children decreases the lifetime risk of breast cancer by 7% for each live birth (“Breast Cancer and Breastfeeding: Collaborative Reanalysis of Individual Data from 47 Epidemiological Studies in 30 Countries, Including 50 302 Women with Breast Cancer and 96 973 Women without the Disease,” 2002; Ma et al., 2006). It is also thought that having the first full-term pregnancy before the age 30 leads to reduced breast cancer risk (https://www.breastcancer.org/risk/factors/radiation accessed on 16.01.2019). Results of a meta-analysis showed that 25% reduction in the risk of estrogen and progesterone positive breast cancers in women who had children in comparison to those who did not have (Lambertini et al., 2016). Breast feeding also lowers the risk of breast cancer. A pooled analysis of 47 epidemiological studies showed that the relative risk for breast cancer is reduced by 4.3% in parous women for every 12 months of breast feeding (Collaborative Group on Hormonal Factors in Breast Cancer, 2002).

Surgery of Reproductive Organs. Results of a pooled analysis and a case-control study demonstrated that the risk of breast cancer among women who have undertaken hysterectomy and oophorectomy before menopause is 24-41% lower compared to women who have not had these surgeries (Nichols et al., 2012; Press et al., 2011).

Personal History. Women who have prior history of hyperplasia or a neoplastic disease in the breast are at higher risk of developing cancer in the opposite breast (contralateral) or in another part of the same breast (ipsilateral). Atypical ductal hyperplasia (ADH) is an identified risk factor for breast cancer. Menes and colleagues found that the cumulative risk of invasive breast cancer is 2.6 times higher in individuals with ADH history compared with women with no history of ADH (Menes et al., 2017). Women with a history of breast carcinoma *in situ* have a 2-3-fold increase in the risk of breast cancer when compared with general female population (Robinson et al., 2008; Soerjomataram et al., 2006). In the study by Soerjomataram and colleagues the second breast cancer was similarly common in the ipsilateral and contralateral breast (Soerjomataram et al., 2006). Ductal carcinoma *in situ* (DCIS) diagnosis is an indicator of an elevated risk of breast cancer as it is a preinvasive lesion. The epidemiological study by Innos and Horn-Ross indicated that the standard incidence ratio (SIR) of ipsilateral invasive breast carcinoma is 1.7 and the SIR for a contralateral invasive breast carcinoma is 1.4, in comparison with general female population (Innos & Horn-Ross, 2008). Lobular carcinoma *in situ* (LCIS) is not a preinvasive lesion but it confers an increased breast cancer risk (Singletary, 2003). The risk of breast cancer increases for all breast tissue and not only the original lesion. Most of the subsequent malignancies which occur about 15 years after the diagnosis of LCIS are ductal rather than lobular (Frykberg, 1999; Frykberg et al., 1987). Studies have shown that risk of second contralateral breast cancer differs according to the hormone receptor (HR) status of primary tumor. They demonstrated that having a primary HR negative invasive ductal or lobular breast tumor leads to a significantly higher risk of second contralateral breast cancer when compared with HR positive invasive ductal or lobular carcinoma (Coradini et al., 1998; Kurian et al., 2009; Swain et al., 2004).

Dense Breast. The mammographic density confers almost the greatest risk to breast cancer compared to most of other cancer risks, and the elevated risk has been shown to stay at least ten years after the date that mammogram classified the density (Byrne et al., 1995). The increased risk of breast cancer associated with qualitative classification of mammographic parenchymal pattern was first described by Wolf in 1976 (Wolfe, 1976a, 1976b). When evaluating mammographic density quantitatively, some studies demonstrated an increased risk of at least four-fold and in some studies the risk for the densest tissues was 1.8 to 6 times as high as the one for the least dense tissues (Boyd et al., 1982; Byrne et al., 1995; Kato et al., 1995; Saftlas et al., 1991). Only 20-30% of diversity in the age-adjusted in the mammographic density can be explained by menopausal status, weight and, number of live births (Boyd et al., 1998; Vachon et al., 2000), these finding resulted in conducting a classic twin study by Boyd and colleagues to determine that to which extent the genetic factors can explain the large unsolved proportion of diversity in mammographic density. They found that the heritability resolves 60% of variability in density among Australian twins, 67% among North American twins and 63% among all twins studied (Boyd et al., 2002). These results show that heritability of breast density varies among different populations.

1.1.3.2 GENETICS

Family History. The first familial breast cancer case was described in 1866 (Paul Broca, 1866; Singletary, 2003). Thereafter, numerous studies tried to define the breast cancer risk related to the positive family history. Pharoah and colleagues (Pharoah et al., 1997) systematically reviewed these studies and performed a meta-analysis using the data collected from the studies which have been done between 1966 to 1996. The data was collected from 52 case-control studies and 22 cohort studies which had calculated the breast cancer risk associated with family history. They discovered the breast cancer risk associated with family history is dependent on several factors such as the number of affected relatives, the type of affected relatives (first degree or second degree), and the age at cancer onset among affected relatives. Comparing the relative risk related to the family history with no family history they estimated relative risk associated with a

family history of having a first-degree relative with an age at onset of 50 years or older as high as 1.8. Similar results were obtained in a pooled analysis which was done by Collaborative Group on Hormonal Factors in Breast Cancer (Collaborative Group on Hormonal Factors in Breast Cancer, 2001).The relative risk increases to 3.3 if the affected first-degree relative developed breast cancer at an age younger than 50. The relative risk is as high as 3.6 for a person with two first degree relative affected with breast cancer in comparison with one without a family history. As compared to those without a family history, individuals with a second degree relative have a relative risk of 1.5 (Collaborative Group on Hormonal Factors in Breast Cancer, 2001).

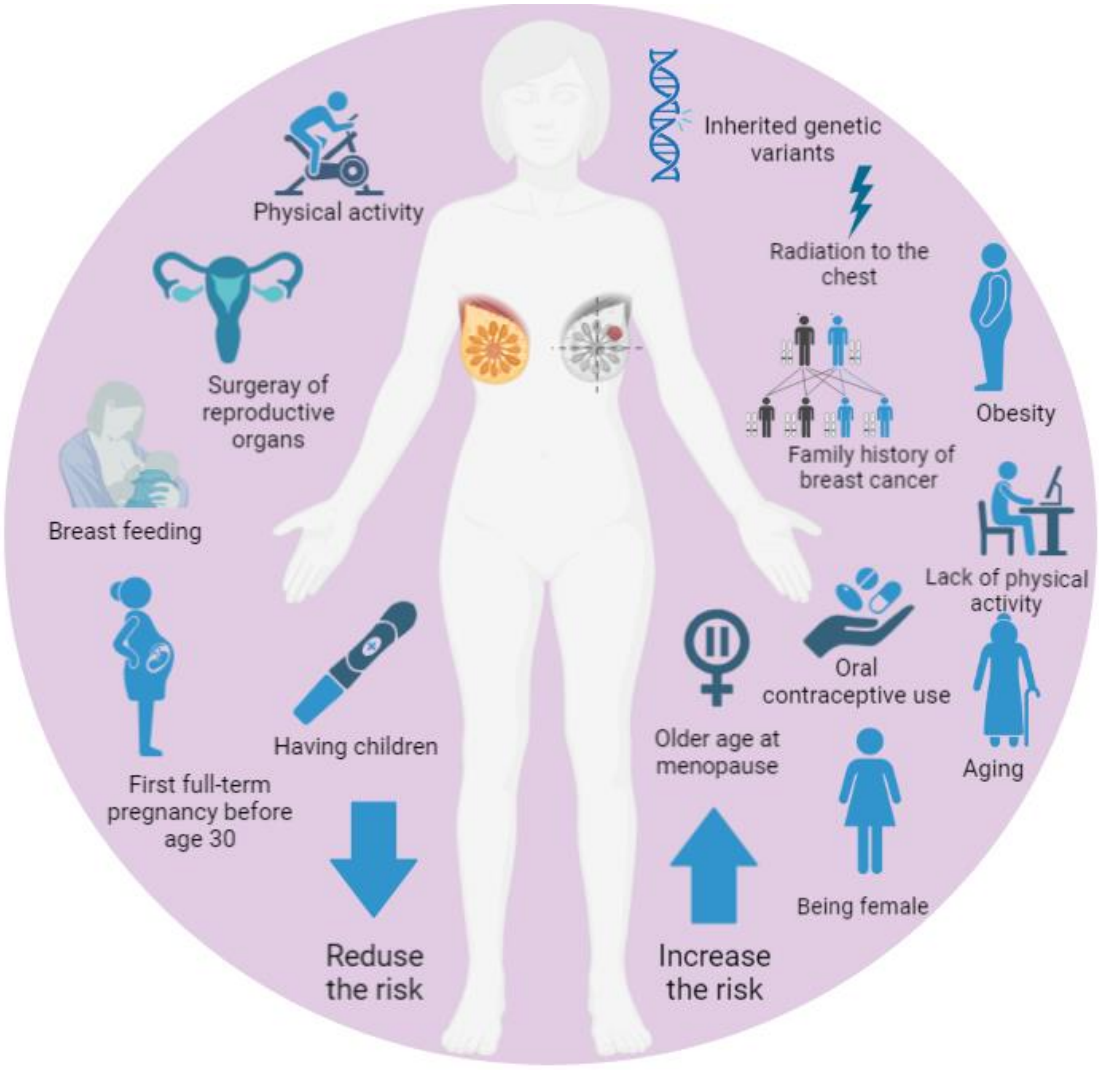

Figure 1. Effect of environmental factors on breast cancer risk (created with BioRender.com)
Factors which are associated with an increased risk of breast cancer are shown in the right side of the figure and factors associated with reduced breast cancer risk are depicted in the left side of the figure.

1.1.4 Sporadic Vs Familial Vs Hereditary Cancer Syndromes

In 2007 Berliner and Fay published a recommendation for risk assessment and genetic counseling for hereditary breast and ovarian cancer in a Journal of Genetic Counseling (Berliner & Fay, 2007). They have defined precisely the differences between sporadic, familial and hereditary cancer syndromes. According to their definition, hereditary cancer syndromes follow an autosomal dominant mode of inheritance. They are developed at an age younger than typical cancer age at onset while in the case of familial cancer syndrome there are more cancer cases of a specific cancer type in the family than what is statistically expected, but no specific pattern of inheritance is detectable. The age at cancer onset varies among familial cancer syndromes. Sporadic cancer cases in a family are more likely due to nonhereditary factors and even if there is more than one affected individual in the family, there is no specific pattern of inheritance. Moreover, cases usually develop cancer at a typical age. Individual affected with hereditary cancer syndromes are at risk of multiple primary cancers or bilateral or multifocal cancers and clustering of rare cancers are detectable while the familial cancer syndromes might be only due to chance clustering of sporadic cases, and they might be also due to a common genetic background and a shared environment and similar lifestyle. First-degree relatives of mutation carriers are at 50% risk of having the same mutation. One of the characteristics of causative genes in hereditary cancer is incomplete penetrance and variable expressivity which leads to different age at onset or no cancer incidence among obligate carriers of family mutation. Considering the size and structure of family, and the expected number of cases, Brewer and colleagues, defined a family history score (FHS) and concluded that breast cancer risk increased with greater FHS (Brewer et al., 2017).

1.1.5 Clinicopathological Features

According to American cancer society the most common symptoms of breast cancer are a lump or thickening of nipple and swelling of all or part of breast, skin changes including skin irritation or dimpling of skin, changes in shape of nipple such as nipple retraction, nipple discharge, and pain in the breast and nipple (https://www.cancer.org/cancer/breast-cancer/about/breast-cancer-signs-and-

symptoms.html, accessed on 24.01.2019). Breast cancer more commonly affects women at age between 55 to 64 years with the median age at diagnosis of 62 (National Cancer Institute, Surveillance, Epidemiology and End results program https://seer.cancer.gov/statfacts/html/breast.html, accessed on 24.01.2019). Breast cancer is considered a heterogenous disease both clinically and genetically (Stingl & Caldas, 2007), therefore a system was needed to organize this heterogeneity and systemize the language. This is why a system was developed over time to classify breast cancer sub-types based on histological and molecular features.

1.1.5.1 Histological Classification

Lack of markers which can define hyperplasia (typical and atypical), carcinoma *in situ*, and invasive cancer has made it impossible to define the progression of breast cancer (Malhotra et al., 2010). Nevertheless, histologically, breast cancer can be divided into two categories of breast carcinoma *in situ* and invasive (infiltrating) carcinoma. Based on cytological and growth pattern, the breast carcinoma *in situ* is subdivided into two groups of ductal and lobular carcinoma *in situ* (Malhotra et al., 2010). Ductal carcinoma *in situ* (DCIS) encompasses about 25% of all breast cancers diagnosed in the United States. It also covers a broad range of diseases ranging from low-grade lesions which are not lethal to high-grade lesions which may have foci of invasive breast carcinoma (Virnig et al., 2010). Based on architectural features of tumor, DCIS can be further classified into Comedo, Cribriform, Micropapillary, Papillary and Solid (Malhotra et al., 2010; Virnig et al., 2010). Regarding tumor differentiation, DCIS can be categorized into high, intermediate, and low-grade tumors (Virnig et al., 2010). The same as *in situ* carcinomas, invasive carcinomas are also histologically categorized into seven sub-groups including infiltrating ductal, invasive lobular, ductal/lobular, mucinous (colloid), tubular, medullary, and papillary carcinomas. Among them, infiltrating ductal carcinoma (IDC) is the most common sub-group which attributes to 70-80% of all invasive lesions (Malhotra et al., 2010). In comparison to IDC, invasive lobular carcinoma (ILC) explains 10-15% of all breast cancers (Ciriello et al., 2015; Desmedt et al., 2016; Du et al., 2018). Small, round, noncohesive tumor cells growing in stroma in a single-file pattern are

characteristics of ILC (Arpino et al., 2004; Rakha & Ellis, 2010), and in contrast to IDC they are hardly detectable using standard imaging techniques such as mammography (Arpino et al., 2004; Krecke & Gisvold, 1993; Lopez & Bassett, 2009). ILCs are more likely occurring at an older age compared to IDC and they are more likely to be estrogen and progesterone receptor positive (Arpino et al., 2004). According to the degree of differentiation and proliferative activity of tumor, IDC can be additionally classified into three sub-types of well-differentiated (grade 1), moderately differentiated (grade 2), and poorly differentiated (grade 3) (Malhotra et al., 2010; Weigelt & Reis-Filho, 2009). Grading is an indicator of tumor aggressiveness (Weigelt & Reis-Filho, 2009). The use of molecular markers such as estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor (HER2/neu) is well appreciated for the IDC while its utility is still debated for DCIS (Harris et al., 2007).

1.1.5.2 Molecular Classification

As the current classification of breast cancers which basically provides prognostic values, is not sufficient to predict responses to the newer targeted therapies and therapy planning, a new classification system to determine the molecular features of tumor was necessary. To address this issue the molecular features of tumor, metastatic propensity, and the signature associated to prognosis (van de Vijver et al., 2002; van ’t Veer et al., 2002; Wang et al., 2005), were broadly investigated via the high-throughput microarray-based gene expression analysis. Moreover, benefiting from complementary DNA microarrays and hierarchical clustering algorithm, Perou et al. and Sorlie et al. developed a new molecular classification system for breast cancer which consists of four molecular subtypes including basal-like, HER2-overexpressed, normal breast-like and luminal subtypes A and B (Perou et al., 2000; Sørlie et al., 2001; Sorlie et al., 2003). Luminal subtype A is the least aggressive subtype and is associated with high expression of ERα and related transcription factors (Santos et al., 2015) while the luminal B subtype is described by the over-expression of proliferation-related genes (Santos et al., 2015; Sørlie et al., 2001). Normal breast-like subtype exhibited the high expression of genes which are known to be expressed by adipose tissues. They were also characterized by over-expression of basal epithelia genes and low

expression of luminal epithelial genes (Sørlie et al., 2001). Whilst basal-like and HER2-overexpressed subtypes are similarly the most aggressive breast cancer subtypes and associated with a poor survival, at the molecular level they are different. HER2-overexpressed tumors are characterized by over-expression of several genes at the 17q22.24 locus including Erb-B2 Receptor Tyrosine Kinase 2 (*ERBB2*) and Growth Factor Receptor Bound Protein 7 (*GRB7*) (Santos et al., 2015; Sørlie et al., 2001), while the basal-like subtypes are portrayed by over-expression of keratins 5 and 17, laminin, and fatty acid binding protein 7 (Santos et al., 2015; Sørlie et al., 2001). Another breast cancer intrinsic subtype is claudin-low which was introduced by Herschkowitz and colleagues in 2007 (Herschkowitz et al., 2007). This subtype is associated with low expression of claudin genes which are critical for tight junctions and cell-cell adhesion (Herschkowitz et al., 2007). Regarding the mutation spectrum, data from Sorlie and colleagues (Sørlie et al., 2001) demonstrated the highest *TP53* mutation rate among basal-like and HER2-overexpressed subtype while luminal A subtype showed the lowest mutation rate in *TP53*. The TCGA consortium studied 825 breast cancer tumors and performed exome sequencing, DNA methylation, RNA expression arrays, genomic DNA copy number arrays, microRNA sequencing and reverse-phase protein arrays ("Comprehensive Molecular Portraits of Human Breast Tumours.," 2012). As for the somatic mutations' spectrum, they revealed the same results as Sorlie and colleagues. They reported the higher overall mutation rate in basal-like and HER2-overexpressed subtypes while the lowest overall mutation rate was observed among luminal A subtypes. The majority of basal like breast tumors are often triple-negative (TNB) and enriched for loss of p53 function. In regard to methylation, luminal B subtype showed enrichment of hypermethylated phenotype (Santos et al., 2015).

1.2 Hereditary Breast and/or Ovarian Cancer

The idea of familial predisposition to some cancers such as breast cancer was developed by Pierre Paul Broca, a French physician, whose wife developed breast cancer at an early age. in 1866 using the family history of his wife, having four generations of breast cancer in the pedigree, he described hereditary breast cancer in his outstanding two-volume treatise entitled *Traite des Tumeurs* (Paul

Broca, 1866). In early 1970s, when Knudson proposed the theory of genetic "two-hit" model for hereditary cancers, the role of tumor suppressor genes was established (Knudson, 1971). Meanwhile, Lynch and Krush reported their findings when studying 34 families with high incidence of breast cancer. In nine percent of the families, they observed interesting association with ovarian carcinoma (Lynch & Krush, 1972). These findings led to the identification of hereditary breast and/or ovarian cancer syndrome (HBOC). HBOC accounts for almost 90% of all hereditary cancers (Kenny & Bickerstaff, 2017). More recently, large twin-cohort studies provided an insight into heritability of different kind of cancers including breast and ovarian cancer (Lichtenstein et al., 2000; Möller et al., 2016; Mucci et al., 2016). By investigating concordance rate of monozygotic and dizygotic Nordic twins, the heritability of breast cancer has been calculated to be 27-31% (Lichtenstein et al., 2000; Möller et al., 2016; Mucci et al., 2016). The same study has been done on concordance rate of monozygotic and dizygotic Nordic twins for ovarian cancer and the heritability of ovarian cancer was estimated to be 39% (Mucci et al., 2016). Twenty percent of breast cancers are familial (family history of breast cancer) (Carroll et al., 2008). Monogenic predisposition to hereditary breast and/or ovarian cancer syndrome can explain 5-10% of all breast cancer cases and about 20% of all ovarian cancer cases (Nielsen et al., 2016). In early 1990s, the linkage analysis and positional cloning resulted in identification of high penetrance hereditary breast and ovarian cancer susceptibility genes *BRCA1* and *BRCA2* (Miki, Swensen, Shattuck-Eidens, Futreal, Harshman, Tavtigian, Liu, Cochran, Bennett, Ding, Bell, et al., 1994; Wooster et al., 1994, 1995), two DNA-repair genes. Pathogenic variants in *BRCA1* confer a breast cancer risk of 57-65% by age 70 years and a lifetime risk of 39-44% for ovarian cancer among female carriers. In women, *BRCA2* pathogenic mutations lead to 45-55% risk of developing breast cancer by age 70 years and a lifetime risk of 11-18% of developing ovarian cancer (Antoniou et al., 2003; Chen & Parmigiani, 2007; Mavaddat et al., 2013). *BRCA1* and *BRCA2* mutations in men also result in a 1.2% and 6.8% risk of breast cancer, respectively (Tai et al., 2007). Despite the fact that *BRCA1* and *BRCA2* are known as high penetrance HBOC associated genes, there is substantial

interindividual and intra-familial (Goldgar et al., 1994) variability in both the age at onset and the site of cancer incidence among mutation carriers. There are multiple lines of evidence proposing the influence of additional modifying factors in cancer penetrance among *BRCA1/2* pathogenic variant carriers. In 2002, Begg and colleagues reported that ignoring the relevant covariables in risk estimating, results in a bias in penetrance estimation and concluded that it is likely that modifiers exist which affect the penetrance of *BRCA1/2*-associated cancers (Begg, 2002).

1.3 High-Risk Genes

1.3.1 *BRCA1*

Breast cancer 1, Early Onset (*BRCA1*) is located on chromosome 17 spanning a region of 81 Kb of genomic DNA. It was isolated by positional cloning of chromosome 17q21 in 1994 (Miki, Swensen, Shattuck-Eidens, Futreal, Harshman, Tavtigian, Liu, Cochran, Bennett, Ding, & Et, 1994). The gene consists of 24 exons of which 22 are coding a transcript of 7.8 kb which results in a protein of 1863 amino acids (Miki, Swensen, Shattuck-Eidens, Futreal, Harshman, Tavtigian, Liu, Cochran, Bennett, Ding, & Et, 1994). *BRCA1* has a unique structure; most of exons are in the expected range of about 100- to 500bps except exon 11 which occupies almost 60% of the coding region of gene and consists of approximately 3500 pbs (Ouchi, 2006). The BRCA1 protein (Figure 2) contains four major domains: (i) RING domain which interacts with BARD1 and exhibits E3 ligase activity. The ligase activity of BRCA1 is intensified when a heterodimer is formed via the RING domain of its partner protein, BRCA1-associated RING domain protein 1 (BARD1) (L. C. Wu et al., 1996). This heterodimer produces poly ubiquitin chains at K6 linkages. These unconventional polyubiquitin chains at unconventional K6 sites may mediate downstream signaling events rather than signaling for protein degradation (L. C. Wu et al., 1996). (ii) the P300/CBP domain shows transcriptional regulatory activity. This domain interacts with two transcriptional cofactors: P300 (histone acetyl transferase p300) and CBP (CREB binding protein) (Trapp et al., 2011). (iii) the coiled-coil domain interacts with PALB2 which in turns binds to N-terminus of

BRCA2 to form a BRCA1-containing complex which functions in double strand-break repair via homologous recombination (Sy et al., 2009; Zhang et al., 2009). (iv) the BRCT domains (BRCA1 carboxy-terminal repeats), conserved in multiple proteins involved in DNA damage response (DDR), are responsible for interaction of BRCA1. BRCT domain is a phosphopeptide-binding domain and recognizes the proteins phosphorylated at serine in SXXF motifs by ATM. BRCA1 interacts with abraxas, BRIP1 and CtIP (also known as RBBP8) to form BRCA1 macro-protein complexes that have different and overlapping roles in DDR. These complexes are mutually exclusive as BRCT domains can bind one protein at a time (Pfeffer et al., 2017).

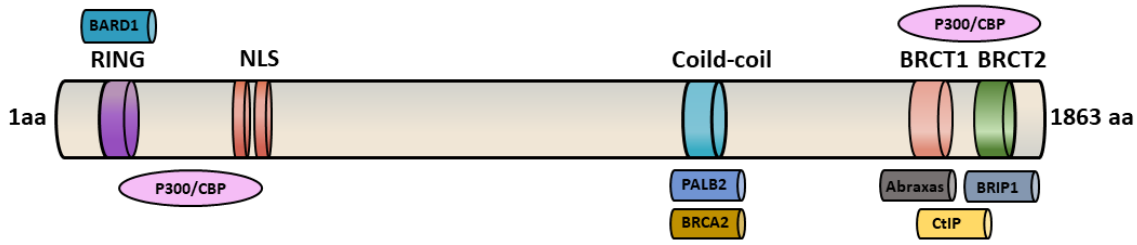

Figure 2. BRCA1 functional domains and interacting proteins.
Domain map of BRCA1 protein. RING domain which interacts with BARD1, NLS (nuclear localization sequence), coiled-coil domain which interacts with PALB2, and the BRCT domains (BRCA1 carboxy-terminal repeats) with the interacting proteins are depicted.

1.3.1.1 Normal Function

The role of *BRCA1* in maintaining genomic integrity is primarily attributed to its involvement in complexes required for the repair of DSBs and stalled replication forks (Roy et al., 2012). It is an extremely crucial component of several different protein complexes that are required for the performance of these functions. As a consequence of its BRCA1 C-terminus domains, highly conserved motifs found in several proteins involved in DSB repair, BRCA1 facilitates the assembly of complexes with proteins such as BTB domain-CNC homolog 1- (BACH1), retinoblastoma binding protein 8 (*RBBP8*) and BRCA1 A complex subunit (*ABRAXAS1*), each of which has distinct functions in the recognition and initiation of DSB repair (Savage & Harkin, 2015). The ubiquitination of histone H2A by the BRCA1-BARD1 complex has been suggested to be essential for repositioning

tumor-suppressor p53-binding protein 1 from the sites of DNA damage(Densham & Morris, 2017). It is believed that this favors homologous recombination (HR)-mediated repair of DSBs over the error-prone alternative non-homologous end joining (Densham & Morris, 2017; Savage & Harkin, 2015). The BRCA1-PALB2 (partner and localizer of BRCA2)-BRCA2 complex is required for the recruitment of RAD51 recombinase (*RAD51*) to the site of DSBs (Roy et al., 2012). By creating a protein complex, RAD51 facilitates the search for homologous DNA sequences, triggering strand invasion and initiating the repair process (Ohta et al., 2011; Roy et al., 2012; Savage & Harkin, 2015).

In DSB repair pathways, end resection promotes the use of homology while suppressing canonical NHEJ; therefore it is a critical step in DSB repair pathway choice (Kass & Jasin, 2010). BRCA1 colocalizes with MRE11-RAD50-NBS1 (MRN complex) after DNA damage and interacts directly with CtIP, suggesting that it is involved in resection (Sartori et al., 2007). The role of BRCA1 in resection has been suggested by the observation that mutant BRCA1 cells are defective in one of the homology-based DSB repair pathways, single-strand annealing (SSA), which, like HR, relies on a resection intermediate but diverges later in the process (Stark et al., 2004).

Along with DSB repair, BRCA1 is also involved in transcription regulation. The transcription regulating functions of BRCA1 are mediated either indirectly by the modulation of transcription regulating pathways or directly by interaction of BRCA1 with transcription factors. For example, by direct interaction, BRCA1 inhibits the transcriptional activity of the estrogen receptor. A good example of how BRCA1 modifies transcription factor activity is its interaction with p53. It was shown that BRCA1 (aa 224–500) interacts with p53's C-terminus and significantly alters p53's transcriptional activity (Mullan et al., 2006).

Additionally, several studies have shown that phosphorylation of BRCA1 activates specific DNA damage-induced checkpoints, such as the intra-S-phase checkpoint. In response to DNA damage, BRCA1 is phosphorylated on serine 1387 by ATM, and this phosphorylation event is required for activation of the intra-

S phase checkpoint (B. Xu et al., 2002). Moreover, phosphorylation of BRCA1 on serine 1423 is also critical for the arrest of the G2/M cell cycle (B. Xu et al., 2001).

1.3.1.2 Molecular Pathogenesis (BRCA1- Haploinsufficiency)

One mutant copy of the BReast CAncer gene 1 (BRCA1) is associated with a significant increase in the risk of developing early-onset breast and ovarian cancer (Narod & Foulkes, 2004). Tumors that develop in individuals with these conditions are different from sporadic cancers in that they tend to arise rapidly (in less than a year) and usually between mammographic screenings (Komenaka et al., 2004). Breast cancers associated with *BRCA1* are also unique in that they tend to be more aggressive than sporadic cancers and are more likely to be of the basal subtype (Campeau et al., 2008; Robson et al., 1998). The mechanism behind why pathogenic alterations in a single *BRCA1* allele create an increased and preferential risk for breast and ovarian cancer cannot be explained in mouse models (Cao et al., 2003; Drost & Jonkers, 2009). A large array of pathways essential to genomic maintenance, including homologous recombination, double-strand break repair, S-phase, G2/M, and spindle checkpoints, are regulated by BRCA1. Due to its essential role in coupling DNA damage sensing and repair to the cell-cycle machinery, biallelic inactivation of *BRCA1* leads to genomic instability and cancer development (Huen et al., 2009; Roy et al., 2012; Zhang & Powell, 2005). It is important to note, however, that BRCA1's proposed functions are not exclusive to breast epithelial cells. Hence, it remains unclear why *BRCA1* pathogenic variants are preferentially associated with cancer incidence in only a subset of tissues rather than a general increase in cancer incidence (as is detected with other tumor-suppressor proteins involved in DNA damage repair for example, p53, ATM) (Musolino et al., 2007; Tlsty, 2011). Moreover, for reasons that remain obscure, it is unclear why breast cancer onset is early and rapid in *BRCA1*-PV carriers (Komenaka et al., 2004; Whittemore et al., 1997) whereas loss of the intact wild-type allele of *BRCA1* occurs late during tumor development (Clarke et al., 2006; Martins et al., 2012). In *BRCA1*-associated breast cancers, inherited pathogenic variants lead to specific molecular and cellular changes in breast epithelial differentiation resulting in a tendency for basal-like tumors to form (Lim et al., 2009; Proia et al., 2011). *BRCA1*-PV carriers

have been shown to have gains and losses of gene copies in key tumor suppressors and oncogenes in carcinoma-free breast and ovarian tissues (Baldeyron et al., 2002; Konishi et al., 2011; Martins et al., 2012; Rennstam et al., 2010). Genetically engineered and primary *BRCA1*-haploinsufficient human mammary epithelial cells (HMEC) have both been shown to have defects in error-free DNA damage repair prior to BRCA1 loss (Baldeyron et al., 2002; Konishi et al., 2011; Pathania et al., 2014; Rennstam et al., 2010). According to a more recent study, premature senescence is not caused by LoH in *BRCA1*mut/+ HMECs. As a result, haploinsufficiency for *BRCA1* causes the onset of a premature senescence-like barrier (a process they call haploinsufficiency-induced senescence (HIS)). A novel form of haploinsufficiency-induced senescence (HIS) was discovered in epithelial cells, which is triggered by activation of the pRb pathway rather than p53 induction (Sedic et al., 2015).

Many of BRCA1's cellular interactions have been discovered since it was first cloned in 1994. Yet its highly specific role in tumorigenesis in breast tissue-*BRCA1*-PV carriers are predisposed to life-time cancer risks of up to 80%-relative to other tissues that remain unaffected, has not yet been fully understood. To address this question, Schemler and colleagues have applied a universal model of tissue-specificity of cancer genes to *BRCA1* and presented a systematic review of four categories of proposed concepts. First, tissue-specific differences in BRCA1 expression levels have been observed, followed by differences in the expression of proteins with redundant functions, and then cell-type dependent interactions of BRCA1 have been highlighted, followed by factors unique to the cell-type as well as the environment of the breast tissue being identified. According to their research, tissue-specificity does not affect BRCA1's role in DSB repair. Furthermore, neither the tissue-specific expression of BRCA1 nor the absence of redundant proteins could explain the observation. Rather, it results from the synergistic effects of BRCA1's interactions with breast epithelial cells and a promoting environment in the breast (Semmler et al., 2019).

Several cellular functions that are mediated by BRCA1 (e.g., DSB repair or cell-cycle checkpoint control) are observed at similar levels in *BRCA1*

haploinsufficient cells and non-mutated cells. *BRCA1* haploinsufficient cells, however, are more susceptible to replicational stress. Breast tissue may be subjected to increased replicational stress leading to genetic instability due to estrogen-induced high concentrations of genotoxic metabolites (Fridlich et al., 2015; Pathania et al., 2014; Savage et al., 2014).

In general, ER- α signaling promotes NHEJ, while HRR and BER pathways are down regulated. BRCA1 antagonizes the mitogenic effect of activated ER- α signaling by lowering NHEJ and activating high fidelity repairs like HRR in estrogen-responsive BRCA1 wild type cells. As BRCA1 inhibits the ligand dependent ER- α signaling, it prevents ER- α from sequestering p53 and blocking its subsequent anti-apoptotic functions, therefore maintaining genomic stability. Conversely, *BRCA1* haploinsufficiency or pathogenic alterations reverse the control over ER-α signaling, thereby stimulating NHEJ and suppressing HRR. As a result, ER- α, which is not under control of *BRCA1*, has the ability to suppress p53 activity and promote uncontrolled cell proliferation in Cyclin D1 dependent manner. The inhibition of p53 by ER- α might not exist in estrogen-nonresponsive *BRCA1* wild type cells; therefore, p53 might be the major mediator of DDR instead of ER-α. Ku80 is upregulated in these cells due to an ER- α deficiency, which stimulates uncontrolled NHEJ. It is possible, however, that the up-regulation of error-prone NHEJ does not affect cell survival, if these estrogen non-responsive cells are wild-type for both BRCA1 and p53. On the other hand, the loss of both BRCA1 and p53 raises concerns about genomic stability. This indicates that ER- α signaling can enhance BRCA1 function in estrogen-responsive cells; therefore, BRCA1 loss as well as deregulation of ER- α signaling together impact chromosomal integrity (Rajan et al., 2021).

1.3.1.3 BRCA1 Pathogenic Variants

To date, there have been more than 1800 rare variants reported in *BRCA1* which include intronic alterations, missense variants and small insertions and deletions (Breast Cancer Information Core; https://research.nhgri.nih.gov/bic/). Majority of pathogenic missense variants in *BRCA1*, which confer a high risk of breast cancer, are located within the RING finger and BRCT domains (Futreal et al.,

1994; Miki et al., 1994). Large genomic rearrangements comprise about 14% of the pathogenic alterations in *BRCA1* (Judkins et al., 2012).

Allelic Variations. *BRCA1* genotype-phenotype correlation was reported first by Gayther and colleagues in late 1995. By studying 33 families they reported a lower ovarian: breast cancer ratio in families carrying mutations downstream of exon 12 (Gayther et al., 1995). Later, Thompson and colleagues found a significantly higher ratio of ovarian:breast cancer associated with variants in central portion of exon 11 as compared to other mutations. This association was not only attributed to increased ovarian cancer risk but also to decreased breast cancer risk (Thompson & Easton, 2002). Gayther and colleagues also reported an increased ovarian vs breast cancer risk associated with mutation in *BRCA2* exon 11. They called this region the Ovarian Cancer Cluster Region (OCCR) (Gayther et al., 1997). In 2015, Rebbeck and colleagues performed an observational study on more than 31,000 *BRCA1* and *BRCA2* mutation carriers and computed the hazard ratios for breast and ovarian cancer based on mutation type, function, and nucleotide position. They also calculated the ration of breast vs ovarian cancer hazard ratios. Based on their data, there are three Breast Cancer Cluster Regions (BCCR) in *BRCA1* located at c.179 to c.505 (BCCR1), c.4328 to c.4945 (BCCR2) and c.5261 to c.5563 (BCCR2'). They also identified an ovarian Cancer Cluster Region encompassing the region between c.1380 to c.4062 (approximately exon 11).

1.3.2 Non-Genetic Modifiers of Breast Cancer Risk Among *BRCA1*-PV Carriers

In 2014, Friebel and colleagues performed a meta-analysis of 44 articles which investigated the modifiers of cancer risk among *BRCA1/2*-PV carriers (Friebel et al., 2014). In their meta-analysis they have included several factors such as age at first live birth, menarche, parity, oral contraceptive use, tamoxifen, breast feeding, mammography, coffee and smoking. Several other factors such as abortion, infertility, menopause, miscarriage, hormone replacement therapy, weight, radiation, x-ray and tubal ligation were also investigated but not meta-

analyzed (due to lack of independent studies). Overall, many of the factors exhibited a null effect.

Age At First Live Birth. Meta-analysis of ***age at first live birth*** was performed using the data from two cohorts (Lecarpentier et al., 2012; Milne et al., 2010). Unlike the general population, the meta-analysis showed a reduced effect size (ES) for the parity at age 30 years or older compared to parity at a younger age (<20 or 20–24 years) (ES = 0.65, 95% CI = 0.42 to 0.99 the same results were obtained for parity at age 25 to 29 years with parity at a younger age: ES = 0.69, 95% CI = 0.48 to 0.99. Authors discussed that this inconsistency is either due to the fact that the effect of the age at first live birth is different among BRCA1 mutation carriers compare to general population or the application of risk reducing salpingo-oophorectomy or bias in ascertainment could have affected the results (Friebel et al., 2014).

Breast-Feeding. Friebel and colleagues reported the results of one study on ***breast-feeding*** (Kotsopoulos et al., 2012), since due to overlap in samples in 6 studies they were not able to perform a meta-analysis. The result of this case-control study revealed a risk reduction for breast feeding compared to never having breast feeding, among *BRCA1*-PV carriers (OR = 0.76; 95% CI 0.61-0.95) (Kotsopoulos et al., 2012).

Age at Menarche. Due to the fact that ***age at menarche*** was not coded consistently in different studies, Friebel and colleagues were not able to perform meta-analysis. The results of a match case-control study by Kotsopoulos and colleagues in 2005 revealed a reverse association between age at menarche and the risk of breast cancer among *BRCA1*-PV carriers (Kotsopoulos et al., 2005).

Parity. To study the impact of ***parity*** in the risk of breast cancer among *BRCA1*-PV carriers, two cohort studies were included in meta-analysis (Lecarpentier et al., 2012; Milne et al., 2010). The analysis of nulliparous against parous showed a null association (ES = 0.79; 95% CI 0.59-1.06) but the significant association was revealed when they further divided the *parity* into five subgroups of: nulliparous, one live birth, two live births, three live births and four or more live

births. According to the meta-analysis having three live births or more and four live birth or more reduces the breast cancer risk compared to nulliparity (ES = 0.57, 95% CI 0.39-0.85; ES = 0.56, 95% CI 0.36-0.86), respectively (Friebel et al., 2014). Each additional live birth leads to a 17% risk reduction (ES = 0.57, 95% CI 0.39-0.85; ES = 0.56, 95% CI 0.36-0.86 (Friebel et al., 2014).

Oral Contraceptive Use. The ***use of oral contraceptive*** (OC) was meta-analyzed using the data from seven publication (Friebel et al., 2014). The meta-analysis of case-control studies revealed a null effect for ever use of OC against the never use of OC (ES = 0.78; 95% CI 0.59-1.04) in *BRCA1* mutation carriers while the combined hazard ratio of cohort studies suggested an increased risk for ever users (ES = 1.59; 95% CI 1.32-1.92) (Friebel et al., 2014).

Prophylactic Oophorectomy. In 1999 Warner and colleagues conducted a study to investigate the prevalence and penetrance of *BRCA1* and *BRCA2* pathogenic variants among Ashkenazi Jewish women with breast cancer. They observed that incidence of breast cancer is maximal between age of 40 and 55 years and it declines afterwards (Warner et al., 1999). These observations are indicative for the effect of ovarian hormonal effect on promoting breast carcinogenesis. Meanwhile, Rebbeck and colleagues found an approximately 50% breast cancer risk reduction (OR = 0.53; 95% CI 0.33-0.84) among *BRCA1*-PV carriers who undergone a risk reducing salpingo oophorectomy (RRSO) (Rebbeck, Levin, et al., 1999). In 2017, Heemskerk-Gerritsen and colleagues followed by Terry and colleagues revealed that the strong association of RRSO and reduced breast cancer risk might have resulted from several types of bias. They found no association when they treated the RRSO as a time-dependent covariate (Heemskerk-Gerritsen et al., 2015; Terry et al., 2019).

Prophylactic Mastectomy. A study performed by Rebbeck and colleagues in 2004 reported that bilateral prophylactic mastectomy reduces the risk of breast cancer by approximately 90%. A prior or concurrent prophylactic oophorectomy could lead to 95 % breast cancer risk among *BRCA1/2*-PV carriers (Rebbeck et al., 2004).

According to a large cohort study of *BRCA1-2*-PV carriers, **diagnostic radiation exposure** before age 30 has been associated with increased risk of breast cancer among *BRCA1*-PV carriers (Pijpe et al., 2012). While results of a meta-analysis revealed a null association between **mammography exposure** and breast cancer risk (Friebel et al., 2014), studies reported a statistically significant increased breast cancer risk for those *BRCA1*-PV carriers who had experienced an **X-ray exposure** compared to those who never had an X-ray exposure (Andrieu et al., 2006; Lecarpentier et al., 2011).

1.3.3 Genetic Modifiers of Cancer Risk Among *BRCA1/2*-PV Carriers

Modifying Genes. The presence and absence of specific alleles of modifier genes can change the penetrance of major genes. The modifying gene can be linked to the major gene. In this case the modifying allele co-segregates with mutant allele. A more possible situation is that modifier gene and major genes are not linked. The modifier genes of hereditary breast and ovarian cancer are related to sex hormones metabolism and DNA-repair (Narod, 2002). Four genes have been proposed as potential modifiers of breast and ovarian cancer risk among *BRCA1/2*-PV carriers.

Androgen Receptor. Androgen receptor contains a polymorphic glutamine extension encoded by trinucleotide repeat (CAG)n, which ranges in length in general population from 17-26 glutamines. *In vitro* analysis showed that this polymorphism is reversely related to transactivation activity of androgen receptor (Kazemi-Esfarjani et al., 1995). Giguère and colleagues also reported an inverse association between CAG length and breast cancer risk (Giguère et al., 2001). Rebbeck and colleagues reported an earlier age of breast cancer onset in *BRCA1*-PV carriers with at least one long androgen receptor allele (>27 CAG repeats) (Rebbeck, Kantoff, et al., 1999). These results are indicative of protective effect of increased androgenic activity, although they were not confirmed by Dagan and colleagues among Jewish *BRCA1*-PV carriers (Dagan et al., 2002).

Nuclear Receptor Co-activator 3. *NCOA3* is also known as “amplified in breast cancer 1”, AIB1, and is needed for female reproductive function and mammary

gland development (J. Xu et al., n.d.). AIB1 contains a glutamine-rich region which is encoded by a (CAG)n trinucleotide expansion (Hayashi et al., 1999). Polymorphic CAG lengths which are ranging in size from 20 to 37 repeats are believed to have a functional effect on sex-steroid-hormone signaling (Rebbeck et al., 2001). Rebbeck and colleagues found that *BRCA1/2*-PV carriers who also carry a long AIB1 allele (≥ 29 CAG repeats) are at an approximately threefold greater risk of developing breast cancer compared to carriers of two short alleles (Rebbeck et al., 2001).

***RAD51*.** *RAD51* is involved in the double strand break repair pathway and interacts with both *BRCA1* and *BRCA2* (Baumann & West, 1998). While a positive association was found between presence of a single nucleotide polymorphism and breast cancer risk among *BRCA2*-PV carriers, no significant modifying effect for *BRCA1*-PV carriers was observed. (Levy-Lahad et al., 2001; W. W. Wang et al., 2001).

HRAS. The presence of a variable number of tandem repeat (VNTR) located one kilobase downstream of the *HRAS* proto-oncogene was reported to be associated with an increased risk of ovarian cancer for *BRCA1*-PV carriers. The association was not found for breast cancer (Phelan et al., 1996).

Catechol-*O*-methyltransferase (*COMT*). was first described by Axelrod and Tomchick in 1958 as an enzyme catalyzing the O-methylation of catecholamine neurotransmitters (Axelrod & Tomchick, 1958). COMT also catalyzes the methylation of catechol metabolites of estrone and estradiol (E1, E2)(M & Kaakkola, 1999). This single gene encodes both soluble COMT(S-COMT) and membrane-bound COMT (MB-COMT)(Lundström et al., 1991; Salminen et al., 1990). COMT is expressed in the range of tissues with the greatest expression in liver and kidney. It is also expressed in both normal breast epithelial cells and tumor tissue (Weisz et al., 2000). Using the methyl group derived from S-adenosylmethionine COMT is functioning as *O*-methylator of catechol estrogens such as 2-OHE1(E2) at the 2-OH and 3-OH positions and of 4-OHE1(E2) at the 4-OH position (Männistö & Kaakkola, 1999) and consequently inactivating the carcinogenic activity of these estrogen metabolites (J. J. Li & Li, 1987; Liehr et

al., 1986). Catechol estrogen are the main metabolites of E1 and E2 which can oxidize to catechol estrogen quinones such as E1(E2)-3,4-Q, which in turn can react with DNA and result in the depurinating adducts 4-OHE1(E2)-1-N3Ade and 4-OHE1(E2)-1-N7Gua (Cavalieri et al., 1997; K.-M. Li et al., 2003). Apurinic sites in the DNA are generated as a result of depurination of these adducts. These apurinic sites can be altered by error-prone base excision repair (BER) to changes which can lead to cancer initiation. *COMT* is a polymorphic gene. The mostly investigated variant in COMT is a common missense variant resulting from a G>A transition leading to a Val to Met alteration at codon 158(108)(Lachman et al., 1996) . It is a common variant in the European population with an allele frequency of approximately 53% (ExAc data accessed on 11.12.2018). Studies have shown that the $COMT^{Met}$ allele showed three-to-four-fold reduction in enzymatic activity compared to wild type allele $COMT^{Val}$ and it has also exhibited thermolability (Scanlon et al., 1979). Earlier, Syvanen and colleges also determined the same reduction of enzyme activity in homozygous individuals and intermediate enzymatic activity in heterozygous individuals (Syvänen et al., 1997). This reduction in methylation activity may lead to accumulation of 4-OHE2. These findings were enough for researchers to hypothesize the association between this *COMT* Val158Met and breast cancer risk; however, the data regarding the association between *COMT* polymorphism Val158Met is contradictory. Some studies have indicated increased breast cancer risk associated with one or two copies of low activity *COMT* allele (Val158Met) among postmenopausal women (Huang et al., 1999; Lavigne et al., 1997; Yim et al., 2001). In contrast, some studies found no association between one or more copies of low activity *COMT* allele and breast cancer risk (Bergman-Jungeström & Wingren, 2001; Millikan et al., 1998). The results of meta-analyses disagreed with the biological role of COMT in inactivating genotoxic catechol metabolites of E2/E1. Two meta-analyses showed a slight protective role for the low activity COMT SNP in recessive mode (Ding et al., 2010; He et al., 2012). More recently, in an attempt to find modifiers of *BRCA1/2* penetrance, Movassagh et al., identified a synonymous substitution p.Leu203Leu (rs165631) in the gene COMT to have a protective affect among *BRCA1/2*-PV carriers (Movassagh et al., 2017).

Common Genetic Modifiers. Unlike the previously used hypothesis-based association studies, genome wide association studies (GWAS) were successful in identifying genetic factors that modify breast cancer risk, both in general population and in *BRCA*-PV carriers. Genome wide association approach led to the identification of more than 100 common susceptibility variants for breast cancer risk in the general population. This number is 22 for ovarian cancer (Milne & Antoniou, 2016). These SNPs are generally common (minor allele frequency ≥ 0.13) and they have small effect size per one copy number of mutated allele (odds ratio, all < 1.30) (Milne & Antoniou, 2011). The Consortium of Investigators of Modifiers of *BRCA1/2* (CIMBA) was established to identify common genetic modifiers of breast and ovarian cancer among *BRCA1/2*-PV carriers. (Milne & Antoniou, 2016). The computed relative risks per copy of the minor allele is small and it ranges from 1.05 to 1.26 for breast cancer and 1.03 to 1.48 for ovarian cancer. These genetic modifiers can explain small proportion (< 10 %) of the modifying genetic variance for *BRCA1* and *BRCA2*-PV carriers (Milne & Antoniou, 2016).

Polygenic Risk Score (PRS). The combined effect of several common breast cancer susceptibility variants, identified in genome wide association studies (GWAS), is expressed as a polygenic risk score (PRS), which can be a substantial genetic risk for an individual person. There has been a consistent association between polygenic risk scores (PRS) and elevated breast cancer risk in cohort studies in both women with and without a family history of breast cancer (Mavaddat et al., 2019). As compared with the limited SNP profiles investigated to date, PRSs based on many SNPs are expected to result in even greater differences in absolute cancer risks for mutation carriers at the extremes of the combined SNP distributions (Milne & Antoniou, 2016). Using such genomic profiles, women can be stratified according to their risk of developing breast cancer. Thus, screening or other preventative strategies can be targeted at those women most likely to benefit from them, thereby improving breast cancer prevention and survival rates (Mavaddat et al., 2019). Nevertheless, PRS risk stratification is reduced in *BRCA1/2*-PV carriers compared to the general population (Sawyer et al., 2012). Coignard et al. have also recently conducted a

case-only study involving over 60,000 unselected BC cases and 13,000 cases containing *BRCA1/2* pathogenic variants. They demonstrated that several SNPs associated with breast cancer risk in the general population, which are therefore included in PRS calculations, are actually associated with *BRCA1/2*-PV status. Therefore, they have no impact on BC risk in carriers of *BRCA1/2*-PV (Coignard et al., 2021).

1.4 Major Signaling Pathways in Breast Cancer Development and Progression

Notable analogues exist between normal development and cancer progression (Huebner & Ewald, 2014; Macias & Hinck, 2012). Many of the signaling pathways which control the normal human development and allows the cells to communicate with each other and surrounding environment are either hijacked or dysregulated by cancer stem cells or cancer cells (Hunter, 2000, 2007; Sever & Brugge, 2015). In fact, cancer is caused by genetic and epigenetic alteration which affect the signaling pathways that command cell proliferation and division, cell death, cell differentiation and cell fate, and cell motility. These alterations allow cells to break out the mechanisms that normally control cell proliferation, survival and migration (Sever & Brugge, 2015). Therefore, mutations in pro-oncogenes can lead to hyperactivation of these pathways while inactivation of tumor suppressors removes critical negative regulator of signaling (Sever & Brugge, 2015). Some of the predominant signaling pathways that regulate normal mammary gland development and breast cancer stem cell functions are presented here.

1.4.1 ER Signaling and ER-Positive Breast Cancer

About 70% of all breast cancers express the estrogen receptor (ER), progesterone receptor (PgR), or both, and therefore are considered hormone receptor-positive (HR+) (Murphy & Dickler, 2016). There are two ERs (ERα and Erβ), which are members of the nuclear receptor superfamily of hormone-inducible transcription factors (Aranda & Pascual, 2001). Estrogen has high affinity and specificity for the estrogen receptor. It binds to ER and functions through two main types of pathways, the classical (or nuclear) pathway and the

alternative (nonnuclear) pathway (Nardone et al., 2015). Upon ligand binding, the estrogen-ER complex dimerizes and interacts with coregulator proteins and estrogen responsive elements within the DNA. This in turn results in activation of the transcription of a wide range of genes that participate in the regulation of the cell cycle, DNA replication, cellular differentiation, apoptosis, and angiogenesis. This is called nuclear (or classical) pathway; however, it starts in the cytoplasm to activate coregulator growth factor and G-protein coupled signaling. Coregulators in the nonnuclear pathways consist of receptors (e.g., insulin-like growth factor-1 receptor, fibroblast growth factor receptor [FGFR], HER2), and kinases (e.g., mitogen-activated protein kinases, receptor tyrosine kinase, PI3K, AKT, mTOR, Src, and CDK). ER can also be activated by ligand-independent mechanisms, which provides several opportunities for crosstalk between the ER, growth factors, and protein kinases, which can activate or modify ER activity (Glück, 2017; Nardone et al., 2015).

1.4.2 HER2 Signaling and HER2 Positive Breast Cancer

HER2 (also known as Neu, ErbB2) is a member of the epidermal growth factor receptor (EGFR; also known as ErbB) family of receptor tyrosine kinases (RTKs), which in humans consists of HER1 (EGFR, ERBB1), HER2, HER3 (ERBB3) and HER4 (ERBB4). Receptor tyrosine kinases are transmembrane single subunit glycoproteins. RTKs have an extracellular ligand-binding domain, a transmembrane domain, and an intracellular tyrosine kinase catalytic domain. Upon ligand activation, the receptors dimerize forming homodimers or heterodimers, which subsequently results in transphosphorylation. This activates a number of intracellular signaling pathways including Ras/mitogen-activated protein kinase pathway, the phosphatidylinositol 3 kinase (PI3K)/Akt pathway, the Janus kinase/signal transducer and activator of transcription pathway, and the phospholipase C pathway, which have roles in cell proliferation, survival, motility, and adhesion (Moasser, 2007). HER2 receptors have no ligand and do not form homodimers but they heterodimers with other HER family members. Overexpression of HER2 is found in approximately 20-30% of human breast cancers (Cho et al., 2003). Breast tumors can have up to 25 to 50 copies of the HER2 gene and up to a 40- to 100-fold increase in HER2 protein expression

(Gutierrez & Schiff, 2011; Kallioniemi et al., 1992). HER2 positivity is associated with more aggressive tumors and a poorer prognosis (Cho et al., 2003).

1.4.3 DNA Damage Response (DDR) Pathway

The stability of the genome is critical for the survival and reproduction of all cells(Valles et al., 2020). An abnormally altered genetic base sequence can impair cellular biological functions and cause carcinogenesis (Burgess et al., 2020). Several important DNA repair functions have evolved in mammalian cells to protect against different types of damage. Mismatch repair pathways, base excision repair pathways, and nucleotide excision pathways, for example, have been well characterized (Kottemann & Smogorzewska, 2013). However, cancer cells have frequently evolved in relation to abnormal DNA damage repair functions and processes (Tubbs & Nussenzweig, 2017). Environmental hazards and endogenous toxic agents, such as free radicals, can compromise DNA integrity, leading to a range of diseases, including cancer (Lin et al., 2020; Patel et al., 2021). Cells can generate various types of DNA damage under either endogenous or exogenous stress, including base pair changes, replication errors (Ragunathan et al., 2020), and distortion and breakage of DNA double strands (Marshall & Santangelo, 2020). Numerous studies have identified and documented the negative effects of common exogenous factors, including heavy metals and ionizing radiation (Huang & Zhou, 2021; Pariset et al., 2021). Several types of DNA damage have been reported previously, as follows: (i) single-strand breaks; (ii) double-strand breaks (DSBs) (Ceccaldi et al., 2016); (iii) base damage; (iv) sugar damage; (iv) DNA cross-linking and (v) clustered damaged sites (Frankenberg-Schwager, 1990). There is a growing body of evidence that DNA double-strand breaks are among the most damaging types of DNA damage and the most severe threat to cells (Bröckelmann et al., 2020). DSBs that are left without effective repair or repaired with error-prone repair can cause carcinogenesis or cell death (J. Li et al., 2019). Several mechanisms have evolved in order to ensure genomic stability within cells or even to use DNA damage as an opportunity for natural selection. These mechanisms have been identified as the DNA damage response (DDR) (R. Huang & Zhou, 2021). Pathogenic variants within the DDR genes such as *BRCA1* and *BRCA2* have

been described in various cancers, including breast cancer and prostate cancer (Kitagishi et al., 2013).

Repair pathways include direct reversal, base excision repair, nucleotide excision repair, mismatch repair, single-strand break repair, and double-strand break (DSB) repair (Lai et al., 2020; Sassa & Odagiri, 2020; Szewczuk et al., 2020). Repair of pyrimidine dimers caused by ultraviolet radiation or other factors, or repair of alkylated bases are considered direct repairs. DNA replication lesions or bulky adducts caused by distortions of the DNA structure are repaired by nucleotide excision repair (Kajitani et al., 2021). Repair of mismatched base pairs in double-stranded DNA, as well as some insertions or deletions of less than four nucleotides, is known as mismatch repair (Latham et al., 2021; Sena et al., 2021). Double-strand break repair refers to repair of DSB lesions (Gachechiladze et al., 2020; Latham et al., 2021; Yoshioka & Matsuno, 2021).

DSB repair has been classified into two broad categories, homologous recombination (HR) (Wright et al., 2018a) and non-homologous end joining (NHEJ) (Lieber, 2010a; Wright et al., 2018a). In contrast to the NHEJ pathway, HR is more conservative and error-free because it relies on the presence of sister chromatids. As a result, the HR pathway can repair DSBs only during the S/G2 phase of the cell cycle, whereas the NHEJ pathway can repair DSBs at any time (R. Huang & Zhou, 2021; Lieber, 2010a; Wright et al., 2018a)

1.4.3.1 Non-Homologous End Joining (NHEJ)

Nonhomologous end joining (NHEJ) is a mechanism for repairing double-strand breaks during the cell cycle (Davis & Chen, 2013). The repair process entails ligation of the ends of the broken DNA strands, which results in high rates of DNA loss and variation (Lord & Ashworth, 2012). There are two distinct NHEJ pathways: classical and alternative. Alternative NHEJ (also known as microhomology-mediated end joining (MMEJ)) is a less-well-described process that is more likely to cause translocations and large deletions (Gostissa et al., 2011). In situations where faithful repair via homologous recombination is no longer possible due to pathogenic changes or epigenetic alterations, NHEJ is called upon to repair double-strand breaks (Lieber, 2010).

1.4.3.2 Single Strand Annealing (SSA)

DSBs are repaired by non-homologous end joining (NHEJ) and homologous recombination repair (HRR) in humans and other higher organisms. Another DSB repair system, single-strand annealing (SSA), joins DNA ends using homologous repeats flanking the DSB. SSA generally occurs when at least two direct DNA repeats are present on the two ends of a DSB. SSA, unlike HRR, does not require a donor sequence. However, it is error-prone, as it removes DNA fragments between repeats along with one repeat. There is homology at breakpoint junctions of many DNA deletions in cancer cells, suggesting that SSA may play a role. SSA may result in chromosomal translocations when more than one DSB occurs on different chromosomes. These translocations are essential for many types of cancer to develop (Blasiak, 2021).

1.4.3.3 Homologous Recombination Repair

When both strands of the DNA double helix are compromised, homologous recombination (HR) is necessary in order to access the redundant genetic information contained in sister chromatids or homologous chromosomes. In somatic cells and during meiosis, HR is critical in supporting DNA replication and repairing DNA double-strand breaks (DSBs). HR consists of three processes: 1) double-strand break recognizing (DSBR); 2) synthesis-dependent strand annealing (SDSA); and 3) break-induced replication (BIR) (Wright et al., 2018). HR consists of three processes: 1) double-strand break recognizing (DSBR); 2) synthesis-dependent strand annealing (SDSA); and 3) break-induced replication (BIR) (R. Huang & Zhou, 2021). ATM and ATR recognize double-strand DNA breaks in the early stages of homologous recombination, phosphorylate downstream targets including CHEK2, P53, BRCA1, and H2AX, and initiate homologous recombination. With the help of BARD1 and BRIP1, BRCA1 organizes the remaining proteins to the site of repair. MRE11, RAD50, and NBS1 form the MRN complex which resects DNA to produce 3′overhangs that are bound by RPA or Replication Protein A (Krejci et al., 2012; Sung & Klein, 2006; Valerie & Povirk, 2003). RPA (consisted of a trimer of RPA1, RPA2 and RPA3) is a single-stranded DNA-binding protein (Dhingra et al., 2019). The recruitment of BRCA2 is mediated by PALB2 and the loading of RAD51 onto RPA-coated

DNA takes place with the help of RAD51B, RAD51C, and RAD51D. The RAD51 filament then invades the homologous DNA strand to allow the remainder of DNA repair to occur using the sister chromatid as a template for error-free repair (Krejci et al., 2012; Sung & Klein, 2006; Valerie & Povirk, 2003)

In the process of homologous recombination repair of double-stranded DNA breaks, *BRCA1* and *BRCA2* are critical proteins. About 5–10% of breast cancers are caused by germline *BRCA1* and *BRCA2* pathogenic variants (Alsop et al., 2012). Additionally, many other proteins involved in homologous recombination repair have now also been recognized to play a role in hereditary cancer risk, such as *ATM*, *CHEK2*, *BARD1*, *BRIP1*, *MRE11*, *RAD50*, *NBS1*, *RAD51C*, *RAD51D* and *PALB2* (Walsh, 2015).

1.5 Role of Next Generation Sequencing in Early Diagnosis of HBC

Many areas of medicine, such as oncology, have benefited from next-generation sequencing (NGS). By using very little nucleic acids, NGS technology allows the accurate characterization of the "status" of numerous genes, while being cost-effective and time efficient. A number of second and third generation platforms, based on a variety of chemical and physical principles, are currently commercially available. In many studies, NGS technology has been used to characterize the tumorigenic process and tumor heterogeneity because it is capable of detecting multiple variants simultaneously. Clinically useful information can be obtained from an analysis of the molecular landscape of tumors. As a result of NGS technology, we can identify and discover novel genes that cause cancer susceptibility, as well as counsel patients and their families about screening, surveillance, and risk-reduction measures. As NGS was developed in the last decade, targeted sequencing of several genes, including those that have already been known to contribute to breast and ovarian cancer susceptibility, as well as sequencing of entire genomes, transcriptomes, and exomes for identifying novel genes became possible.

In "targeted gene sequencing", selected genes or subsets of genes are analyzed whose involvement in specific diseases has already been detected or suggested. Since NGS gene panels offer a very high-throughput and cost-effective screening

method for sequencing specific targets of interest, they are widely used. The system allows massive parallel multigene analysis in a few days, reducing the cost and time requirements significantly.

Whole exome sequencing (WES), in which DNA coding regions are captured and sequenced at a deep level, can detect disease-causing variants and uncover new targets. WES provides a more complete analysis of the genomic landscape than targeted gene sequencing. Currently, WES is the most widely used NGS technique for identifying rare genetic variants. The WES analysis, however, only provides information about exons, focusing on coding regions. It is therefore critical to note that this approach omits variants in non-coding regulatory regions that can act as cancer drivers (Hu et al., 2021; Meldrum et al., 2011).

1.5.1 Next Generation Sequencing

Almost 25 years after the discovery of DNA's structure, the first method for sequencing DNA was published (Sanger et al., 1977; Watson & Crick, 1953). By incorporating chain-terminating and radioactively labeled (earlier approach) or fluorescently labeled (later approach) dideoxynucleotides, this method allowed the sequencing of a complementary DNA strand to the interrogated template strand. Following size separation, the fragments were analyzed by gel electrophoresis to determine their sequences. After the introduction of capillary electrophoresis, Sanger-sequencing developed into a “first-generation sequencing” technique that was widely used to sequence both small and large genomes in bacteria, phages, and humans. This method had a limited throughput since it could only analyze one sequencing reaction at a time. The advent of next-generation sequencing (NGS) technologies in 2004 and 2006 transformed biomedical inquiry, resulting in a dramatic increase in the amount of data produced from sequencing. This significant increase in data output was the result of nanotechnology principles and innovations that enabled the massively parallel sequencing of individual DNA molecules. Regardless of the sequencing platform used, high throughput and single-molecule DNA sequencing are hallmarks of NGS (Hu et al., 2021).

A “second generation” approach, such as those employed by Illumina or Ion Torrent, generally begins with DNA fragmentation, DNA end-repair, adapter ligation, surface attachment, and in-situ amplification. As a result of these new “short-read” sequencing technologies, millions of individual sequencing reactions can be performed simultaneously (Hu et al., 2021).

1.5.2 Library Preparation (DNA-seq)

There are four basic types of DNA-Seq: Whole Genome Sequencing (WGS), Whole Exome Sequencing (WES), Epigenome Sequencing (ES), and Targeted Sequencing (TS) (Lightbody et al., 2019). Polymerase chain reaction (PCR) and hybridization capture-based methods are two approaches to template preparation. The most common way to prepare TS templates is by PCR. Amplification sequencing was initially limited to a few genes or exons (short-range PCR) (Head et al., 2014). Due to long-range PCR (LR-PCR), shotgun sequencing (sequencing randomly broken fragments shorter than the amplicons) became dominant, enabling all regions of the gene, including intronic, untranslated, upstream, and downstream regions, to be sequenced. With LR-PCR, sequence ambiguities with short amplicon sequencing were resolved (Meldrum et al., 2011). A hybridization capture-based preparation of templates is applied for WES and TS using biotinylated probes that are hybridized with regions of interest, which are then isolated using streptavidin-coated magnetic beads. This approach provides cost-effective TS of larger genomic regions and more genes with less chance of allele dropout than PCR-based approaches. There are four core steps to DNA-Seq library construction: fragmentation, end-repair, adaptor ligation, and size selection (Podnar et al., 2014).

1.5.3 Sequencing

Clonal amplification and sequencing are two sequential elements of short-read sequencing. The purpose of clonal amplification is to produce strong, detectable signals during the sequencing of DNA fragments through solid-phase amplification. Flow cells (Illumina) and beads (Thermofisher's Ion Torrent) are two examples of solid-phase surfaces to which single DNA fragments can bind. A PCR emulsion (Ion Torrent) or PCR bridge (Illumina) is used to amplify the

anchored DNA fragments into spatially separated template fragments, depending on the sequencing platform. Both Ion Torrent and Illumina use the "sequencing by synthesis (SBS)" method, which relies on DNA-polymerase-dependent nucleotide incorporation into the extended DNA chain (Goodwin et al., 2016).

1.5.4 Data Analysis

The use of these technologies requires streamlined bioinformatics data management and data analysis given that massively parallel sequencing generates large volumes of data. A data analysis workflow typically involves three levels of analysis: primary, secondary, and tertiary (Dolled-Filhart et al., 2013). Within the primary analysis base-calling is performed on each clonally amplified DNA fragment by the instrument software following sequencing. During this phase, quality control procedures such as read filtering and trimming are also carried out. A FASTQ file is used to store the sequence information along with the quality scores (Phred values). Pair-end sequencing on the Illumina platform generates two FASTQ files linked by sequence identifiers. On Illumina instruments, three indicators are used to assess the quality of sequencing runs: cluster density, percent of clusters passing filters, and percentage of base calls with a quality score of at least Q30 (1 in 1000 chance of an incorrect base call). Read alignment and variant calling are included in secondary analysis. FASTQ files are used to store short reads, either paired-end or single-end. First, short reads are aligned against the human reference genome. In order to achieve fast and accurate alignment, Burrows-Wheeler Aligners (BWA) use a hash-table algorithm allowing gapped alignment (H. Li & Durbin, 2010). A binary alignment/map (BAM) format is used for storage of alignment results. User-friendly software, such as the Interactive Genome Viewer (IGV), can be used to view alignments. During this stage, coverage is another quality indicator to consider, which includes depth of coverage (number of times a base is sequenced) and breadth of coverage (percentage of a reference genome covered). A sufficient level of coverage ensures sufficient sensitivity and specificity for variant detection. Variant calling follows read alignment. Through comparison of reads to a reference genome, variants, SNPs, indels, or larger structural variants can be identified. GATK and Freebayes (Mckenna et al., n.d.;

Sandmann et al., 2017) are two open-source tools that can be used for this analysis. Variant Call Format (VCF) is used to store sequence variation data. Annotation and interpretation of variants are part of tertiary NGS analysis. The goal of this analysis is to determine the biological and pathological functions of found variants (e.g., SNPs, INDELs, and CNVs). In this analysis, ANNOVAR and VAT are commonly used tools.

2 Aims

In spite of the fact that pathogenic variants in *BRCA1* have the highest penetrance for hereditary breast and ovarian cancer, the cause of variation in penetrance between individuals and even within families remains a subject of research. As a result of this variation, genetic counseling and risk calculation are challenging.

Besides the environmental factors that have been extensively studied, genetic factors are another source of variation in the risk of breast cancer among *BRCA1* women. The Consortium of Investigators of Modifiers of *BRCA1/2* (CIMBA) carried out Genome Wide Association Studies (GWAS) and described several candidates: each adding a small part of risk variation in *BRCA1* mutation carriers. As opposed to GWAS studies that are based on common variants, this study attempted to predict *BRCA1* penetrance and AAO of breast cancer by analyzing rare variants in genes associated with DNA damage response and genome integrity maintenance pathways, as well as genes that interact with *BRCA1*. We hypothesized that the co-occurrence of a germline truncating variant in DNA-repair genes with *BRCA1* is associated with early onset of breast cancer.

Furthermore, HBOC molecular diagnostic laboratories face significant challenges when interpreting variants of uncertain significance, especially as the number of genes implicated in the syndrome grows. A well-established *in vivo* or *in vitro* functional study is considered to be a strong piece of evidence for variant interpretation in the ACMG/AMP framework. Investigating the effects of these variants on RNA splicing is a major contribution to understanding their functional significance. Here we introduce a targeted RNA-sequencing method to simultaneously evaluate the effect of several variants within several genes on mRNA and consequently on protein function.

We aim to identify:

1. Modifiers of risk of hereditary breast cancer among *BRCA1*-positive women
2. A method to evaluate functional effects of variants of uncertain significance in HBOC genes.

3 Materials and Methods

3.1 Study Population

3.1.1 *BRCA* PV-Carriers From GC-HBOC

BRCA1/2-PVcarriers were selected from more than 30,000 cases referred to and registered in 12 centers of the German Consortium for Hereditary Breast and/or Ovarian Cancer biobank. Patients were eligible for the study if (a) they had developed breast cancer before the age 35 years (early age at onset), (b) they had developed breast cancer after age 60 or they cancer free by the age 60 and above (control cohort), (c) patients are unrelated (d) patients who had done the prophylactic mastectomy or prophylactic oophorectomy prior to the age 45 were not included in the analyses (Heemskerk-Gerritsen et al., 2015) as they have reduced the risk of breast cancer. All the patients have been tested positive for *BRCA1/2* pathogenic variants in their centers prior to the recruitment and they have signed a written informed consent. The study was approved by the local ethics committee (ethic vote number 053/2017BO2)

3.1.2 Splicing Pattern in Cancer Predisposing Genes (Targeted RNA-Sequencing Vs. cDNA Sequencing)

3.1.2.1 Patients' Characteristics

The interpretation of variants of unknown significance (VUS) is one of the most challenging aspects of molecular diagnosis of hereditary breast and ovarian cancer (HBOC), especially with the increasing knowledge of genes that play a role in this syndrome. One of the keyways in which VUS can have a functional impact is through their effects on RNA splicing. In attempt to replace the conventional laborious method of RT-PCR and direct sequencing of isolated fragments from agarose gel with a powerful semi-quantitative high-throughput method, we designed a panel of baits to capture the exons of 34 targeted cancer related genes -without including the baits for known exon-exon junctions to avoid the enrichment of known transcripts-, along with a bioinformatics pipeline to analyze aberrant transcripts and splicing variants. We compared the results of RT-PCR and cDNA sequencing in parallel with targeted RNA-Seq in 5 patients

either with a personal or a family history of cancer referred to the Women Hospital in Tübingen for genetic testing. The patients carried variants in cancer susceptibility genes including *CHEK2*, *BRCA2*, *BRCA1*, *TP53*, and *STK11*.

3.1.2.2 RT-PCR Analysis

Whole blood samples which were collected using PAXgene Blood RNA Tube (IVD) (PreAnalytiX, A QIAGEN/ BD Company) were used as a source of total RNA. Total RNA extraction was performed using PAXgene Blood miRNA (PreAnalytiX, A QIAGEN/ BD Company) according to the manufacturer's recommended protocol. An Aliquot of 500ng was used for the first-strand cDNA synthesis by means of SuperScript III First-Strand Synthesis SuperMix (Invitrogen, Life Technologies, USA) according to the manufacturer's instructions.

3.1.2.3 PCR

Blood samples were used to analyze the splice variant for the genes which were found to carry a splice-site variant and the prediction tools predicted a possible exon skipping. Primers were designed using "Primer 3" or "primer design". Primers were encompassing the upstream and downstream exons of the exon, which was expected to be skipped. Table 1 shows the primers which were used for each specific variant. The 25µl PCR reaction contained about 20ng of cDNA template, 10x buffer + $MgCl_2$ (Roche, F. Hoffmann-La Roche AG), 10mM dNTPs (Roche, F. Hoffmann-La Roche AG), 5U/µl FastStart Taq polymerase (Roche, F. Hoffmann-La Roche AG), 5M of Betain, 5x Q-Solution, 10 pmol/µl of Forward and Reverse primers as well as DNAse-free water. The PCR reactions were performed in G-Storm thermocycler using a touchdown program starting with initial denaturation at 95°C for 5 minutes followed by 10 cycles of denaturation at 94°C for 30 seconds, annealing at variable temperatures dependent on the primer pairs, decreasing one degree at each cycle, for 30 seconds, and elongation at 72°C for either 30- or 60-seconds dependent on the size of PCR product. It was followed by 30 cycles of denaturation for 30 seconds at 94°C, annealing at the temperature which was reached in the final cycle of the first 10 cycles, for 30 seconds, and elongation at 72°C for 30- or 60- seconds following

by 7 minutes of elongation at 72°C. PDH was used as internal control. PCR products were analyzed on 1.5% agarose gels.

Table 1. PCR primers

Gene	Variant	Exon number	Forward Primer	Exon number	Reverse Primer
CHEK2	c.592 G>A	Ex2	GACCAAGAACCTGAGGACCA	Ex6	TGCCTCTCTTGCTGAACCAA
BRCA2	c.8755-2 A>G	Ex21	GCAAGATGGTGCAGAGCTTTA	Ex23	TTTTGTCGCTGCTAACTGTATG
BRCA1	c.4093 T>G		GTGAATTGGAAGACTTGACTG		TGTCACTCTGAGAGGATAGC
TP53	c.375+5 G>A	Ex1	GTGACACGCTTCCCTGGAT	Ex6	CCAAATACTCCACACGCAAA
STK11	c.597 G>A	Ex2	GGCCAACGTGAAGAAGGAAA	Ex7	TTCTTCCGGAACCAGCTGT

3.1.2.4 DNA Extraction and Sanger Sequencing

The visible fragments corresponding to each patient and matching control was extracted from agarose gel using a QIAquick Gel Extraction Kit (Qiagen, Inc.) according to the manufacturer's instructions. Purified PCR products were subjected to sequencing using a BigDye Terminator v.3.1 Cycle Sequencing Kit. PCR mixture (10µl) contains 2µl of purified DNA, BigDye buffer, 10 pmol/µl M13 primers, DMSO, BigDye 3.1, and DNase free water to reach the volume to 10 µl. PCR reaction was carried out in GStorm thermocycler, initiating with 2 minutes of denaturation at 96°C, followed by 30 cycles of denaturation at 96°C for 10 seconds, annealing for 5 seconds at 55°C, and elongation for 3 minutes at 60°C followed by a final extension of 1 minutes at 60°C. PCR products were cleaned using 10µl cleanDTR magnetic beads per sample which were mixed with 42µl of 85% ethanol, after 3 minutes incubation on a magnetic rack, the supernatant is discarded and beads are washed with 100µl of 85% ethanol. After 3 minutes incubation ethanol is removed, and DNA is eluted in 150µl HPLC water. 30µl of the DNA was transformed to the 4titude plate ready for sequencing.

3.1.2.5 Gene Selection and Panel Design, Library Preparation and Sequencing

Agilent eArray (SureDesign; Agilent was used to design 4690 SureSelect solution library baits each contained 120-nucleotides which target all known exons of 34 cancer-related genes (Appendix). Prior to the enrichment the RNA integrity number (RIN) was defined by Agilent 2100 BioAnalyzer (RNA 6000 Pico Kit, Agilent) and the samples with RIN>7 were selected for library construction. the

SureSelectXT RNA Target Enrichment for Illumina Multiplexed Sequencing (Agilent), poly A targeted sequencing libraries were generated according to the manufacturer's recommended protocol except that we used 13 cycles for the Pre-capture PCR. Libraries are sequenced Libraries either on HighSeq2500 or NextSeq500 (Illumina, San Diego, CA, USA) All libraries are sequenced on the Illumina HiSeq 2500 platform in paired-end mode with read length 125bp and at a depth of approx. 5 million clusters each. Library preparation and sequencing procedures are performed by the same individual and a design aimed at minimizing technical batch effects was chosen.

3.1.2.6 Bioinformatic Analysis

Quality of raw RNA-seq data in FASTQ files was assessed using ReadQC (ngs-bits version 0.1) to identify potential sequencing cycles with low average quality and base distribution bias. Reads were preprocessed with skewer (version 0.2.2) and aligned using STAR (version 2.5.4a) allowing spliced read alignment to the human reference genome (GRCh37). Alignment quality was analyzed using MappingQC (ngs-bits version 0.1) and visually inspected with Broad Integrative Genome Viewer (IGV, version 2.3.1). Based on the Ensembl genome annotation (GRCh37), junction counts were obtained with Sashimi plots generated in IGV.

3.1.3 DNA-Repair Genes Sequencing

3.1.3.1 Gene Selection and Panel Design for Next Generation Sequencing

Genes were selected based on a reported association with breast cancer in the literature as well as all DNA-repair related genes which were selected from KEGG GENES database (http://www.genome.jp/kegg/genes.html, last accessed: 26.11.2013). All genes are shown in table 1. Agilent eArray (SureDesign; Agilent) was used to design the baits to target the entire coding region as well as exon-intron boundaries ±25 bps. All the setting parameters were set as default except for "Masking" that "Most Stringent" was selected.

3.1.3.2 Library Preparation and Sequencing

DNA was isolated from blood and were shipped from each center to Tübingen center. Genomic DNA was quantified using Qubit 2.0 fluorometer and dsDNA Assay kit (Thermo Fischer Scientific, Waltham, MA, USA). 200 nanogram of

genomic DNA were sheared by a Covaris system (Covaris, Inc., Woburn, Massachusetts) in order to produce fragments of 120-150 bps size. Prior to capturing the fragmented DNA was tested for the quality on a TapeStation (Agilent, Santa Clara, CA). Regions of were captured using Agilent SureSelect custom RNA probes (Agilent, Santa Clara, CA) and according to the Agilent SureSelectXT protocol. Post-captured libraries were also tested for quality by a TapeStation and quantified by a Qubit High Sensitivity dsDNA Assay. Next Generation Sequencing was performed either on an Illumina Miseq, NextSeq500 or on a HiSeq2500 (Illumina, San Diego CA) platform using paired-end reads of 151 bps or 101 base pairs.

3.1.3.3 Bioinformatic Analysis

An open-access, free-to-use bioinformatics pipeline called megSAP was utilized for data analysis (version 0.1-379-gb459ce0, https://github.com/imgag/megSAP). The adapter trimming was performed by SeqPurge (Sturm et al., 2016) and the sequencing reads were mapped to the human reference genome version hg19 using BWA (H. Li & Durbin, 2010). PCR duplicates were removed by SAMBLASTER (Faust & Hall, 2014). Indel realignment was conducted by ABRA2 (v.2.05) (Mose et al., 2014). Variant calling was performed using freebayes (v.1.1.0) (Garrison & Marth, 2012). Variant annotation was conducted using snpEff/SnpSift (v.4.3i) (Cingolani et al., 2012). Splice site annotation was performed using Alamut batch (v. 1.5.1, Interactive Biosoftware). Quality control was performed in three levels of information including raw reads, mapped reads and variants.

3.1.3.4 Variant Interpretation

A modified version of American College of Medical Genetics and Genomics (ACMG) guidelines for variant classification (Richards et al., 2015) was used as a basis for writing an in-house algorithm to classify the obtained variants. To classify variants in DNA-repair genes the algorithm was written as follows: among the variants with the minor allele frequency (MAF) below 0.01 in the population databases such as 1000 Genomes Project (1KGP), Exome Aggregation Consortium (ExAc), and Exome Sequencing Projects 6500 (ESP6500), those

leading to a truncated protein due to (a) premature stop codon, (b) small insertion/deletion resulting in a frameshift, and (c) a start loss were classified as pathogenic/truncating variants. If there was additional evidence in favor of pathogenicity such as functional assessment in the literature or if the variant was clinically assessed and classified in ClinVar, it was a plus for the variant, but it was not essential. The Splice site variants in the canonical splice site as well as variants at +/-1 and +/-2, were classified as likely pathogenic if they were rare (MAF< 0.01) and they disrupt the protein function due to a truncation that they may cause. The indications for truncation or exon skipping were provided by Alamut visual incorporated tools such as MaxEntScan, Splice Site Finder Like, and Human Splicing Finder. For the variants located outside of consensus splice sites such as Cartegni splice sequences (Cartegni et al., 2002) differences between wild type splice site and the mutated splice site given by prediction tools such as MaxEntScan and Splice Site Finder-like was considered to classify them as variant of uncertain significant or benign in the case that nucleotide was not conserved (based on phyloP (Cooper et al., 2005; Siepel et al., 2006)score) and the variant had a MAF>0.05 in population and subpopulation databases. Synonymous variants were classified as benign if they had MAF>0.05 and they didn't cause a new cryptic splice donor or splice acceptor (based on the differences between wild type and mutant provided by splice prediction tools). Deep intronic variants in positions that nucleotide is not conserved, and the variant has a MAF between 0.01 and 0.05, and it doesn't cause a new cryptic splice acceptor or donor, were classified as likely benign.

3.1.3.5 Statistical Analysis

Descriptive statistics such as mean, median, IQR, as well as confidence intervals (95 % CI) have been calculated by GraphPad Prism (GraphPad Software, La Jolla California USA) to characterize the population studies and to investigate the differences of pathological characteristics of patients from different cohorts. To assess the mutational load, the association between mutational load and the age at breast cancer onset, as well as the association between tumor characteristics and age at breast cancer onset, we used Fisher's exact test with a two-sided p value using GraphPad Prism (GraphPad Software, La Jolla California USA). Rare

variant association study has been conducted by performing Burden and SKAT-O association tests implemented in the R package SKAT (https://www.hsph.harvard.edu/skat/download/) version 1.3.0. Maftools was operated to visualize *BRCA1/2* pathogenic variants (lollipop plots) with a modified database.

3.1.4 Analysis of Extreme Phenotype Sampling Data

3.1.4.1 Rare Variant Association Study (RVAS)

In order to acquire functional annotation (exonic, nonsynonymous, synonymous, splicing etc.), European population allele frequencies (from 1KGP, exome Variant Server (ESV) and ExAc database), in addition to functional impact score from CADD, variants obtained from freebayes in VCF format were annotated by use of eDiVA platform. As a consequence, those variants which were annotated as “exonic” or “splicing”, together with variants within segmental duplication (SegDup identity $\geq$ 0.9) were excluded from further analysis. Sample quality control has been done by searching for outliers in (a) number of variants per sample and (b) transition to transversion ratio per sample. Synonymous single nucleotide variants which were not in linkage disequilibrium and had an allele frequency of more than 0.005 in European Variant Server were used to determine the first 10 principal component analyses (PCA) of all samples. Eventually, the rare variant load per gene was compared between early and late age at onset cohorts. As a result, no outlier was detected in any QC test and late and early age at onset subjects were clustering in a single group in the PCA. Ensuing quality check, all the variants with the European allele frequency above 0.01 in any of population databases mentioned above, were removed from analysis. Furthermore, all the variants with a CADD score below 10 as well as synonymous variants were excluded. The remaining rare and likely deleterious variants were used to conduct the Burden and Sequence Kernel Association Test-O (SKAT-O) implemented in R package SKAT (https://www.hsph.harvard.edu/skat/download/) version 1.3.0. The SKAT_Null_Model function with output set to dichotomous outcome (out_type= “D”) and no sample adjustment (Adjustment= FALSE) was used to compute the Null model for both tests. SKATBinary function was utilized for the SKAT-O test.

Barring methods that were set to “optimal.adj” which is equivalent to SKAT-O method, all the parameters were set as default. Minor allele frequencies (MAF) which were transformed with Get_Logistic_Weights were used as weights. The same function and parameters, apart from method that was set to “Burden” was used for Burden test.

3.1.5 Amplicon Sequencing

3.1.5.1 Library Preparation and Sequencing

Amplicon sequencing was utilized to screen for the variants in *COMT* gene. Primers (Table 2) were designed to cover the entire coding region of *COMT* including 4 coding exons (out of 6 exons) as well as exon-intron boundaries. While designing the primers the sequencing adapters were added to the 5’ side of primers. Primers were designed by primer 3 software(Koressaar & Remm, 2007; Untergasser et al., 2012). DNA concentrations were examined either by Nanodrop or by Qubit Broad Range dsDNA Assay (Thermo Fischer Scientific, Waltham, MA, USA) ahead of PCR. To prepare the libraries for sequencing we used 4 ng of genomic DNA isolated from the blood, for the first PCR. PCR reactions (25 µl) contained 1X Q5 reaction buffer, 200µM dNTPs, 0.5 µM forward primer, 0.5 µM reverse primer, 0.02 U/µl Q5 High-Fidelity DNA Polymerase (NEB, New England BioLabs), 1X Q5 High GC Enhancer, and nuclease-free water. In the first PCR step, primers contained a Nextera adapter sequence as well as gene specific primer, amplified our region of interest. The PCR reaction was carried out in a GStorm Thermal Cycler using the program starting with an initial denaturation at 98°C for 30 seconds, following 30 cycles of denaturation at 98°C for 10 seconds, annealing at variable temperature dependent on primer pairs for 30 seconds, and elongation at 72°C for 30 seconds, and a final elongation of 2 minutes at 72°C. The PCR products were visualized on 1.5 % agarose gel to assure the PCR efficiency. Following the completion of amplification of all six fragments for each patient, 5µl of each amplicon corresponding to each patient were pooled and the pool was purified by Ampure XP beads (Beckman Coulter, Life Sciences) according to the manufacturer’s protocol except that we used 80 % freshly prepared ethanol for washing and DNA was eluted in 21 µl of nuclease-free water. The pooled, cleaned-up PCR products were used as an input for the

second PCR. In the second round of the PCR, adaptors were used as primers in order to add the Nextera XT indices (Nextera XT Index Kit, Illumina, Inc. San Diego, CA) to both 5' and 3' of each fragment to achieve dual indexing. In the second round of PCR, the 25 µl reaction consisted of 2.5 µl of each index (i5 and i7), 2 µl of pooled fragments, 1X Q5 reaction buffer, 200 µM dNTPs, Q5 High-Fidelity DNA polymerase (NEB, New England BioLabs), and nuclease-free water. The PCR reactions were performed in a GStorm Thermocycler using a program initiating with denaturation at 98°C for 30 seconds, following 10 cycles of denaturation at 98°C for 30 seconds, annealing at 55°C for 30 seconds and elongation at 72° for 30 seconds following a final elongation of 2minutes at 72°C. The PCR products were purified using the same protocol that was described earlier and then it was quantified using a Quant-iT Broad Range dsDNA Assay kit on a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). The equimolar of each patient's pool was pooled to construct the final library pool. Libraries were sequenced on Illumina MiSeq (Illumina, San Diego CA) platform using MiSeq Reagent Nano Kit v2 (300-cycles). To increase the diversity of the library 10% of genomic DNA from the phage PhiX was added to the library. Sequences were paired end and 150 bps.

Table 2. Primers used for amplicon sequencing.

Exon	Forward (5'->3')	Reverse (5'->3')
Exon 1-part 1	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGG CTGGCATTTCTGAACCTTG	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGG AACTCGTTCCAGCCGATAA
Exon 1-part 2	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGG CTGGAACGAGTTCATCCTG	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTC CTGTAAGGGCTTTGATGC
Exon 2	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGC ACCTGTGCTCACCTCTCCT	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCC CTTTTTCCAGGTCTGACA
Exon 3	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG CTGTTCCAGGTCACCCTTGT	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGG GCTCTACTGGAATGCCTGG
Exon 4-part 1	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGC TAGTGAGGAGCACCCATCC	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGC GCACGTGTGCTAGGAAGT
Exon 4-part 2	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGG CTACTGGCTGACAACGTGA	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGT ACCAGGCTGGGTGAGAGA

3.1.6 Bioinformatic Analysis

Short read sequencing data was generated on the Illumina platform (MiSeq, paired-end mode with 2x150bp reads). Quality of raw sequencing data in FASTQ format was assessed with ReadQC (ngs-bits version 2018_06). Reads were

preprocessed with SeqPurge (ngs-bits version 2018_06 https://www.ncbi.nlm.nih.gov/pubmed/27161244) using default parameters to remove adapter contamination and low-quality bases. Read alignment to the human reference genome (GRCh37) was performed with BWA-MEM (version 0.7.17 https://arxiv.org/abs/1303.3997). Alignment quality was analyzed using MappingQC (ngs-bits version 2018_06) and visually inspected with Broad Integrative Genome Viewer (version 2.2.4). FreeBayes (version 1.1.0 https://www.arxiv.org/ans/1207.3907) was used to call variants on the aligned data.

3.1.7 Statistical Analysis

Descriptive analyses such as medians, means and standard deviations for continuous data and proportion and 95% CI for categorical data were used to characterize the study population and sequencing results. Association of identified variants with breast cancer was tested by Fisher's exact test. P values are two-sided and a p value of 0.05 is considered to be statistically significant. All the statistical analyses were done using GraphPad Prism version 6.07 for Windows (GraphPad Software, La Jolla California USA).

4 Results

Partial results of the presented work have been published in (Sepahi et al., 2019).

4.1 Sequencing and Participants Characteristics

4.2 Sequencing Results

In this study, 311 genes (Appendix) were evaluated that either maintain genome integrity and/or have been associated with HBOC. The average depth of the sequencing was 456x ± 197.3 SD. Table 3 provides detailed results and quality parameters regarding the sequencing process. Of the 3703 variants identified, 43 (1.2%) were truncating variants (Appendix) found in 36 DNA-repair genes.

Table 3. The quality parameters of Next Generation Sequencing

Quality Control Parameter	Average for the Panel
Read counts	9.7± 5.6 Mio
Depth	456.0± 197.3
20x Coverage percentage	97.8± 0.3%
Insert size	178.4± 8.4
On target reads percentage	68.26± 4.4
Q30 base percentage	95.13± 1.7

4.2.1 Participants in *BRCA1* Positive Early Age at Onset and Control Cohorts

In total, 152 *BRCA1*-PV carriers were screened for 311 DNA-repair genes. For the purpose of extreme phenotyping patients were recruited based on their age of first breast cancer onset. Of these patients, eight from early age at onset cohort (≤35 years) and three from control cohort (≥ 60 years) were excluded due to not meeting the age at onset criteria. Two patients from control cohort were excluded from further analysis since they had developed ovarian cancer and two had undertaken prophylactic oophorectomy before the age 45. One patient from the control cohort was also removed from the analysis as she was carrying a missense *BRCA1* pathogenic variant (ENST00000357654: c.5365G>A:p.Ala1789Thr) which is known to be a variant of uncertain significance (ClinVar Accession: RCV000989865.3). Finally, three cases were removed since the reported *BRCA1* pathogenic could not be confirmed. Out of

these 133 patients, 73 women manifested breast cancer at an age younger than 35 years (median age at onset, 27 years; interquartile range (IQR) 25-27 years) and 60 women either were diagnosed with breast cancer at an age older than 60 years (n = 25; 41.7%, median age at onset, 64 years (IQR, 62-67)) or have not developed breast cancer by the time of sample taking (n = 35; 58.3%, median age, 70 years; IQR, 63-75 years). The demographic characteristics of the *BRCA1* positive patients are shown in Table 4.

Overall, 65 *BRCA1* pathogenic and likely pathogenic single nucleotide variants were detected in 117 patients from both cohorts (Appendix). Large exon deletions and duplications were found in 16 patients (including 13 large deletions and three large duplications). The most frequent single nucleotide variant in early AAO cohort is an insertion of C in exon 20 of the *BRCA1* gene (ENST00000357654:c.5266dupC:p.Gln1756fs) which was detected in 15.1% of the patients in this cohort. In the control cohort, the missense variant in exon 4 (ENST00000357654: c.181T>G: p.Cys61Gly) was the most frequently (10%) detected pathogenic variant.

Table 4. Demographic characteristics of the BRCA1 positive population study. Table has been reprinted from Sepahi et al., 2019

	Early age at onset cohort	**Control cohort**
Total Number	73	60
Breast cancer positive	100%	41.7%
Age at onset (range)	21-33	60-87.9
Median Age (IQR)	27 (25-27)	67.3 (62.7-73.2)
Median age of onset	27.8	67.3
BCCR1 (95 %-CI)	13.8 % (6.1-25.4 %)	11.5 % (4.4-23.4 %)
BCCR2 (95 %-CI)	8.6 % (2.9-19.0 %)	5.8 % (1.2-15.9%)
BCCR2' (95 %-CI)	22.4 % (12.5-35.3 %)	15.4 % (6.9-28.1%)
OCCR (95 %-CI)	25.9 % (15.3-39 %)	42.3 % (28.7-56.8 %)
***BRCA1* variant type % (95 %-CI)**		
Frame-Shift-Del	26.0 % (16.5-37.6 %)	35.0 % (23.1-48.4 %)
Frame-Shift-Ins	19.2 % (10.9-30.1 %)	16.7 % (8.3-28.5 %)
Missense variant	8.2 % (3.1-17.0 %)	13.3 % (5.9-24.6 %)
Nonsense variant	26.0 % (16.5-37.6 %)	21.7 % (12.1-34.20 %)
Splice-Site	5.5 % (1.5-13.4 %)	5.0 % (1.0-13.9 %)
CNV	15.1 % (7.8-25.4 %)	8.3 % (2.8-18.4 %)
Family History		
Data available for	73 (100 %)	60 (100 %)
First-degree relative with Breast and/or Ovarian cancer	41 (56.2 %)	59 (98.4 %)

BCCR: Breast Cancer Cluster region, BCCR1: c.179-505, BCCR2: c.4328-4945, BCCR2': c.5261-5563, Del: deletion, OCCR: c.1380-4062, Ins: insertion, CNV: Copy Number Variation. Table has been reprinted from Sepahi et al., 2019.

4.3 Allelic Variation

4.3.1 Pathogenic Variants in *BRCA1* Compared by Type and Location in Each Cohort

Allelic variation in type and location of pathogenic variants in *BRCA1* gene was assessed between both early and late age at onset cohorts by comparing the PV frequencies in Breast Cancer Cluster Regions and Ovarian Cancer Cluster Regions. Patients carrying large deletions/insertions as well as splice-site variant carriers were excluded from the analysis, since these pathogenic variants may encompass more than one region and their effect on protein function is not clear. Although there was no difference in the frequency of pathogenic variants in BCCRs between the cohorts, OCCR pathogenic variant frequency was higher in late age at onset; however, it was not statistically significant. (p value= 0.07). Twenty-two (45.3 %) patients in the control cohort carried a pathogenic variant within the OCCR compared to 15 (25.9 %) of patients in the early AAO cohort (Figure 3a).

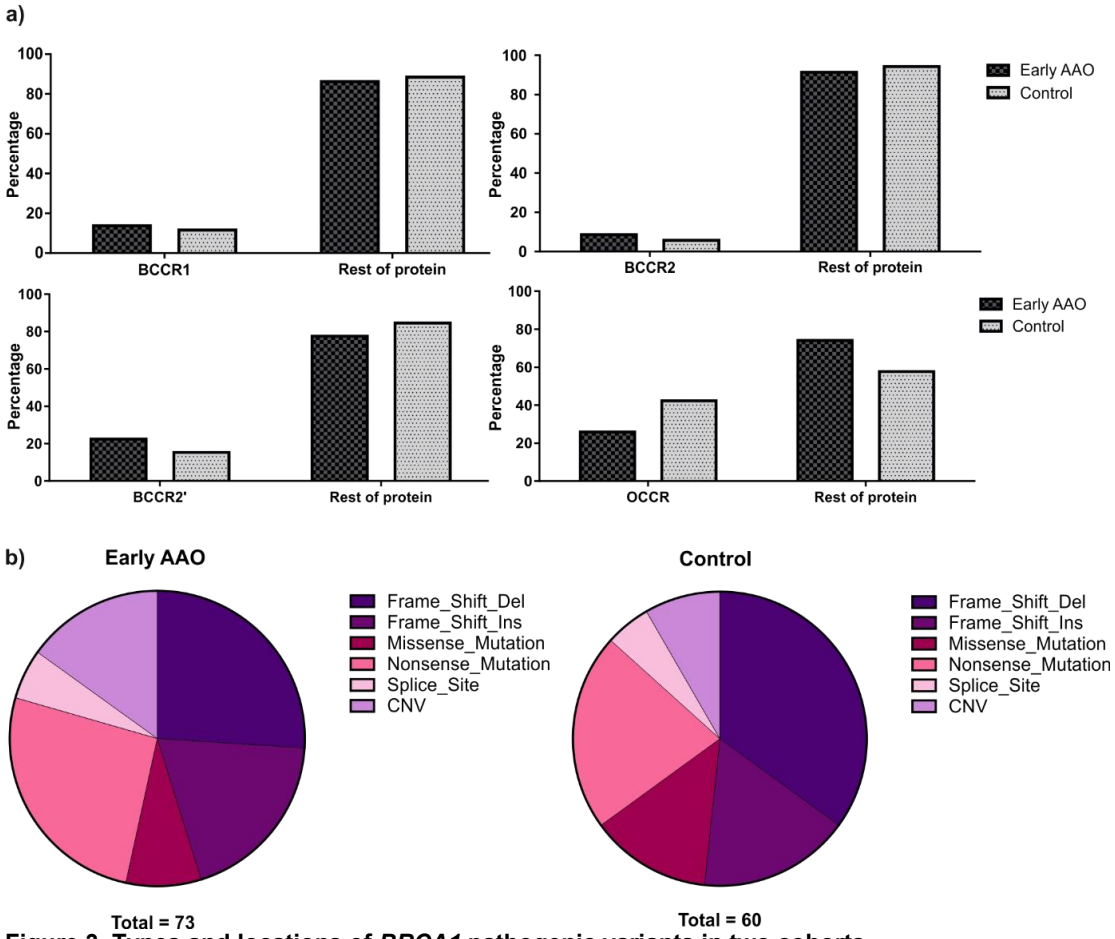

Figure 3. Types and locations of *BRCA1* pathogenic variants in two cohorts
a) The distribution of pathogenic variants in BCCR (Breast Cancer Cluster Region) and OCCR (Ovarian Cancer Cluster Region) are compared. b) Types of pathogenic variants in two cohorts; Del: deletion; Ins: insertion; CNV: Copy Number Variation. Figure has been reprinted from Sepahi et al., 2019.

Among the early AAO cohort, 76.7 % (95 %-CI 65.4 % to 85.3 %) of *BRCA1*-PV carriers carried a truncating variant whereas 8.2 % (95 %-CI 3.1 % to 13.3 %) carried a missense pathogenic variant (ENST00000357654: c.181T>G: p.Cys61Gly) and 15.1 % (95 %-CI 7.8 %-25.4 %) carried a structural variant. A truncating variant was found in 78.3 % (95 %-CI.65.8 % to 87.9) of the control cohort, while 13.3 % (95 %-CI 5.9 % to 24.6 %) carried a missense pathogenic variant (Figure 3b) including ENST00000357654: c.181T>G: p.Cys61Gly, and c.5096G>A: p.Arg1699Gln) and 8.3 % (95 % CI 2.8 % to 18.4 %) carried a structural variant (large deletion/duplication).

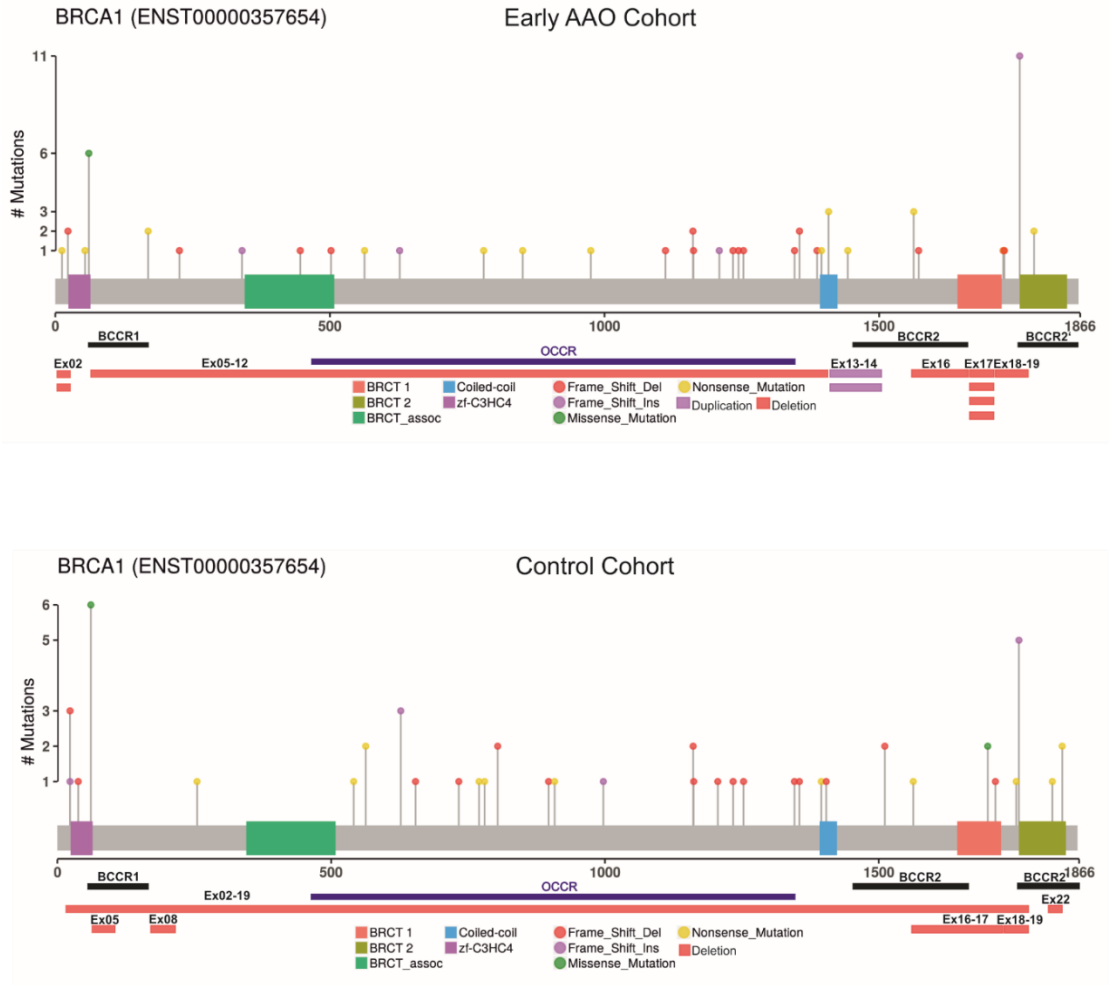

Figure 4. Pathogenic variants in *BRCA1*
The X-axis indicates the amino acid position and functional domains of the BRCA1 protein. The lollipops represent different types of pathogenic variants, which are depicted by different colors. Number of PV carriers is shown on the Y-axis. Copy number variation is shown by the horizontal bars. Different colors indicate deletion (red) and duplication (purple). Black bars represent Breast Cancer Cluster Regions (BCCRs), while the dark blue bar represents Ovarian Cancer Cluster Region (OCCR). The splice-site variants are not displayed. Figure has been reprinted from Sepahi et al., 2019.

4.4 Environmental Factors

4.4.1 Age at Menarche

There was a slightly significant difference in the mean age at menarche among the *BRCA1*-PV carriers in early AAO cohort compared with the control cohorts (p value: 0.047). Data regarding the age at menarche was available for 55 out of 73 patients in the early AAO cohort (73%) compared with 36 out 60 patients in the control cohort (60%). Patients in the early AAO cohort had an earlier age at menarche than the control cohort (12.6 vs 13.2 years; p-value) (Figure 5).

4.4.2 Oral Contraceptive Use

The data regarding the use of oral contraceptive (OC) was available for 66 patients (92%) in early AAO cohort and all the patients in control cohort. While 66% of the patients in early AAO whose data was available had ever used the OC, only 23% of the patients in control cohort had ever used OC (odds ratio: 6.3; 95%-CI 2.8 to 13.7; p-value for the Fisher exact test <0.0001) (Figure 5). We did not identify any difference when we compared the duration of OC use (months) between the patients in early AAO cohort and control cohort (p-value: 0.5; median length of OC use in early AAO: 107.5 vs median length of OC use in control cohort: 84.0).

4.4.3 Parity

The data regarding the parity was available for 59 (80%) of the patients in early AAO cohort and 44 (73%) of the patients in control cohort. Of these patients only 29% in early AAO cohort were parous compared to 93% in control cohort (odds ratio 0.03; 95%-CI 0.009 to 0.1; p-value: <0.0001). Parity has also been compared between the two groups by the number of pregnancies. Unpaired t test showed a significant difference between the number of pregnancies (p-value: < 0.0001; mean number of pregnancies in early AAO cohort 0.5 vs 2.5 pregnancies in control cohort). Age at first pregnancy was available for 14 out of 17 parous women in the early AAO cohort and for 35 out of 41 parous women in the control cohort. Women in the control cohort had a significantly younger at first pregnancy compared with women in the early AAO cohort (mean age at first pregnancy in control cohort: 23 years vs mean age at first pregnancy compared to 27 years; p-value 0.0003).

4.4.4 Birth Cohort

Women in the early AAO cohort were born after 1960 while women in control cohort were born between 1919 to 1954. Most women (50.7%) in the early AAO cohort were born between 1981 and 1993; while the majority of women in the control cohort (38.3%) were born between 1930 to 1940 (Table 5).

Table 5. Distribution of patients in each cohort based on the year of birth.

	Control cohort (n= 60) n (%)	Early AAO cohort (n= 73) n (%)
<1930	6 (10%)	
1930-1940	23 (38.3%)	
1941-1950	21 (35%)	
1951-1960	10 (16.7%)	
1961-1970	_	10 (13.7%)
1971-1980	_	26 (35.6%)
>1981	_	37 (50.7%)

4.4.5 Family History

The data from the total number of first-degree relatives was available for 67 patients out of 73 in early AAO cohort and for all the women in the control cohort. In general, patients in the early AAO cohort came from smaller families. The average number of first-degree relatives in early AAO was 3.7 compared to 6.8 persons in the control cohort. About 44% of the patients in the early AAO did not have a history of breast and/or ovarian cancer in their first-degree relatives, while this number in the control cohort was only 1.7%. In contrast, about 33% of the women in the control cohort had three or more affected first-degree relatives while only one patient in the early AAO cohort (approx. 1.4%) had three or more first-degree relatives with breast and/or ovarian cancer (Table 6).

Table 6. Family history of breast and/or ovarian cancer among BRCA1 positive patients

	Early AAO cohort (n=73) n (%)	Control cohort (n= 60) n (%)
Family history of breast and or ovarian cancer (no. of affected first-degree relatives)		
0	32 (43.8%)	1 (1.7%)
1	32 (43.8%)	15 (25%)
2	8 (11%)	24 (40%)
≥3	1 (1.4%)	20 (33.3%)

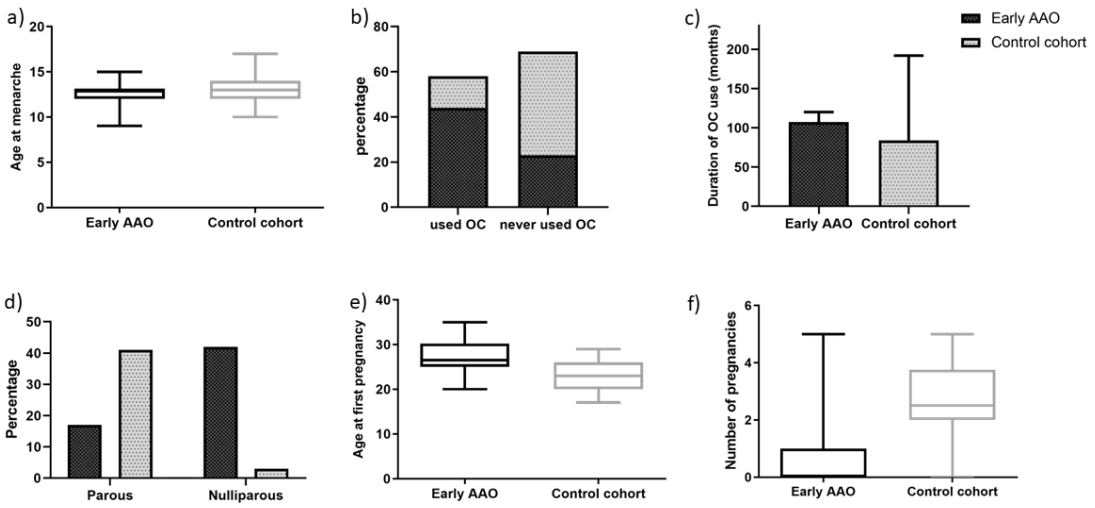

Figure 5. Environmental modifiers in two cohorts
Two cohorts were examined for environmental modifiers of breast cancer risk. a) The age at menarche was compared between the two cohorts. b) and c) the duration of oral contraceptive use was compared among those who reported using oral contraceptives. d), f) compare the parity status in each cohort and the age at first pregnancy and the number of pregnancies among the parous women.

4.5 Truncating Germline Variants in DNA-repair Genes in *BRCA1* PV-Carriers

Thirty-six DNA-repair genes were affected by truncating variants. Among a total of 3703 variants, 43 (1.2 %) of them were leading to a truncated protein. The truncating variants mainly affected genes in Single Strand Break Repair pathway (SSBR, 30.6 %), Double Strand Break Repair pathway (DSBR, 30.6%), and check-point factors (11.1 %). Other truncating variants were found in genes associated with other functions, such as *BRCA1* and *BRCA2* interactors, centrosome formation, and signal transduction. There were 42 women (in both cohorts) who had at least one additional truncating variant in the DNA repair pathway. Among the early AAO cohort, 26 individuals (35.6 %; 95 %-CI 24.7 % - 47.7 %) carried at least one additional truncating variant, including two cases who carried two additional truncating variants in DNA-repair genes (Figure 6a). Out of 60 participants in the control cohort,16 (26.7%; 95%-CI 16.1 to 39.7%) carried an additional DNA-repair germline truncating variant. Three participants in this cohort carried two germline DNA-repair truncating variants; at least one of these affected a DSBR pathway gene (Figure 6b).

The risk of developing breast cancer among *BRCA1*-PV carriers was examined in relation to the presence of additional DNA-repair truncating variants, adjusted for age at menarche, oral contraceptive use, parity, and family history. While it did not reach the conventionally accepted p-value of 0.05 for double heterozygote patients, the odds ratio indicates an increased risk of breast cancer (OR: 3.1; 95% CI 0.92 to 11.5, p-value = 0.07). We conducted a similar analysis on a subset of subjects matched for family history (early AAO cohort; n = 41 and control cohort; n = 59) adjusted for age at menarche, oral contraception use, and parity to confirm the validity of our model (OR: 3.3; 95% confidence interval 0.92 to 13.3; p-value = 0.07). In this subset of cohorts, similar results were obtained.

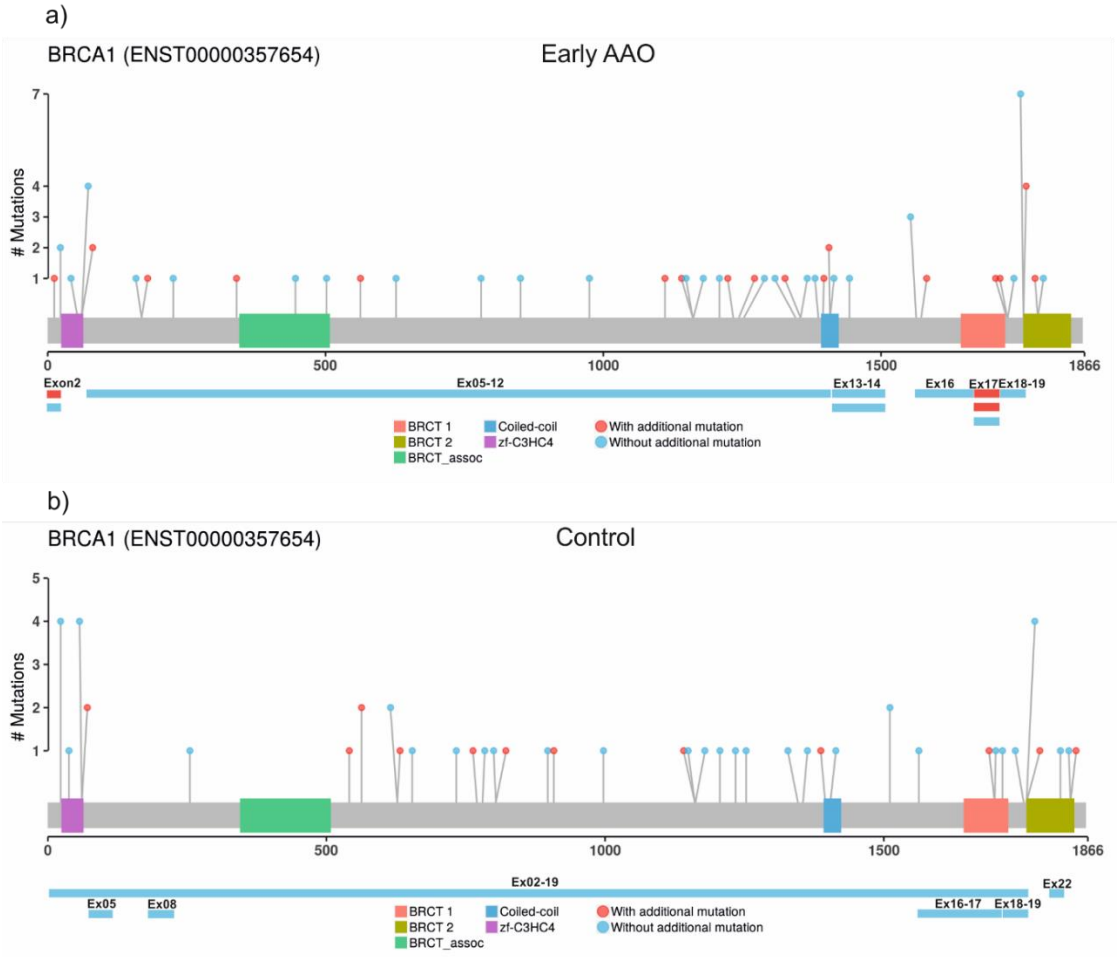

Figure 6. Carriers versus non-carriers of additional truncating variants.
This lollipop plot illustrates the position of BRCA1 pathogenic variants in two cohorts: (a) early AAO and (b) control cohort; with and without additional DNA-repair truncating variants. The X-axis shows the functional domain and amino acid positions of the BRCA1 protein, while the Y axis represents the number of carriers. Each lollipop represents the position of a pathogenic variant of BRCA1 in patients with (red) or without (blue) additional truncating variants. The horizontal bars show BRCA1 copy number variations in patients with (red)

and without (blue) additional truncating variants. Splice-site variants are not displayed. Figure has been reprinted from Sepahi et al., 2019.

4.6 Truncating Variants in Different DNA-repair Pathways

To assess the role of specific DNA-repair pathogenic variants in the age of breast cancer onset we broke down the mutational load into small fragments. The mutational load in DSBR and SSBR as two mainly affected DNA-repair pathways was compared between two cohorts. Among the early AAO cohort, 8/73 women (11.0 %; 95 %-CI 4.9 %-20.5 %) carried an additional truncating variant in DSBR compared to 5/60 women (8.3 %; 95 %-CI 2.8 %-18.4 %) in the control cohort. Regarding SSBR genes, we found 8/73 women (11.0% %; 95 %-CI 4.9 %-20.5 %) in the early AAO cohort carrying additional SSBR truncating variants compared to 5/60 women (8.3%; 95 %-CI 2. %-20.5 %) in the control cohort. The mutational load in DSBR and SSBR did not differ between both cohorts, respectively. In addition, we compared the age at onset of the DNA-repair PV carriers to non-carriers inside each cohort. We also tested the age at onset diversity of those carrying additional PV in DSBR with those carrying additional PV in SSBR in each cohort. Furthermore, we checked for intra-cohort diversity of AAO between those carrying additional truncating variants in DSBR and SSBR genes with the rest of the cohort (including non-carriers and other DNA-repair PV carriers). We also compared the mean age at onset of DSBR truncating variant carriers with those carrying additional truncating variants in other pathways. Moreover, the mean age at onset of DSBR truncating variant carriers was compared to those without an additional one. The same was done for SSBR truncating variant carriers (Figure 7). In none of the cases differences were statistically significant.

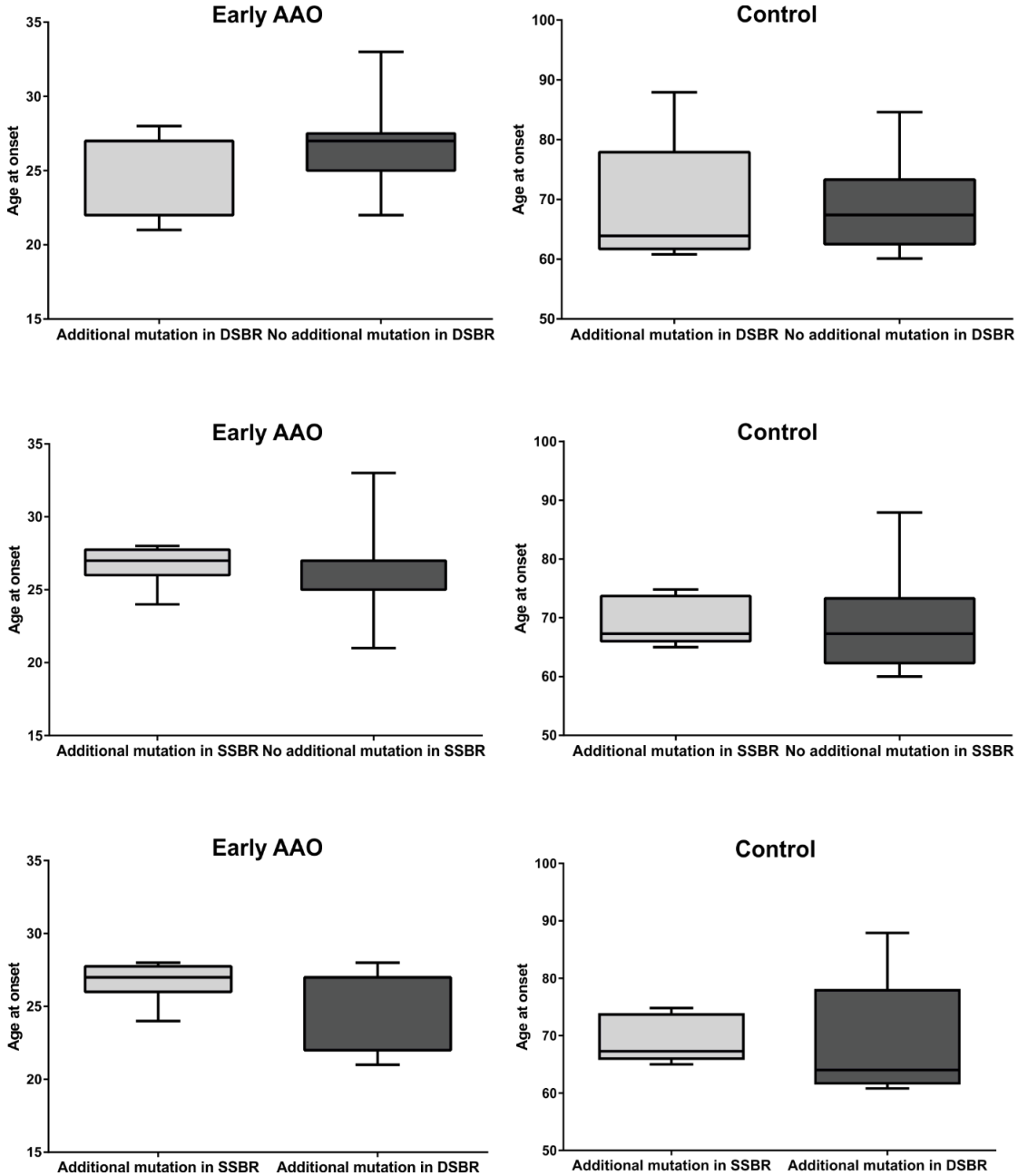

Figure 7. Comparison of AAO between DSBR/SSBR gene PV carriers and non-carriers

DSBR: double strand break repair, SSBR: single strand break repair. Figure has been reprinted from Sepahi et al., 2019.

4.7 Pathological Characteristics

There were 25 (41.7%) patients in the control group who developed breast cancer at a median age of 64. This group of patients was compared with the early AAO patients in terms of tumor characteristics. The immunohistochemical staining of estrogen and progesterone receptors did not differ significantly with regard to the AAO, however, there was a much higher frequency of ER and PR negativity among early AAO patients than among affected control patients (p-value = 0.28 and 0.76, respectively, Table 5). In contrast to the tumors of the affected control patients, the early AAO group had higher histological grades (Table 7), but the difference was not statistically significant (p-value = 0.24).

Table 7. Histopathological characteristics of tumors

	Early AAO cohort Number (%)	**Control cohort Number (%)**	**P value**
Histological Type			
Data available	*62 out of 73*	*22 out of 25*	
Ductal	53 (85 .5%)	22 (100 %)	0.10
Medullary	6 (9.7 %)	0	
Lobular	2 (3.2 %)	0	
Others	1 (1.6 %)	0	
Histological grade			
Data available	*66 out of 73*	*22 out 25*	
Grade III	53 (80.3 %)	14 (63.7 %)	0.24
Grade II	13 (19.7 %)	7 (31.8 %)	
Grade I	0	1 (4.5 %)	
Steroid receptors			
Data available	*64 out of 73*	*22 out 25*	
ER negative	47(73.4 %)	13 (59.1 %)	0.28
PR negative	52 (81.3 %)	17 (77.3 %)	0.76
Human Epidermal Receptor ***Data available***	*52 out of 73*	*19 out of 25*	
HER2/neu negative	49 (94.2 %)	17 (89.5 %)	0.60

Data were available for 67 out of 73 patients from the early age at onset cohort and from 28 cases that developed breast cancer in the control cohort. ER: Estrogen receptor; PR: Progesterone receptor; HER2: Human Epidermal growth factor receptor 2. Table has been reprinted from (Sepahi et al., 2019)

There were no significant differences between patients with truncating variants in DNA-repair genes and patients without additional truncating variants in DNA-

repair genes regarding expression of estrogen and progesterone receptors (Table 8).

Table 8. Comparison of histopathological characteristics of DNA-repair PV carriers with non-carriers

	Early age at onset (<35) n =.73		Controls (>60) with cancer manifestation n = 25	
	With additional truncating variant n= 26	**Without additional truncating variant n= 47**	**With additional truncating variant and breast cancer n = 6**	**Without additional truncating variant and breast cancer n = 19**
Data Available	21	43	5	17
ER negativity	15 (71.4 %) 95 %-CI (47.8-88.7 %)	32 (74.4 %) 95 %-CI (58.8-86.5 %)	4 (80.0 %) 95 %-CI (28.4-99.5 %)	9 (52.9 %) 95 %-CI (27.8-77.0 %)
Data Available	21	43	5	17
PR negativity	17 (81.0 %) 95 %-CI (58.1-94.6 %)	35 (81.4 %) 95 %-CI (66.6-91.6 %)	5 (100 %)	12 (70.6 %) 95 %-CI (44.0-89.7 %)
Data Available	19	33	4	15
HER2 negativity	17 (89.5 %) 95 %-CI (66.9-98.7 %)	32 (97.0 %) 95 %-CI (84.2-99.9 %)	4 (100 %)	13 (86.7 %) 95 %-CI (59.5-98.3 %)
Data available	22	44	5	16
Grade 3	17 (77.3 %) 95 %-CI (54.6-92.2 %)	36 (81.8 %) 95 %-CI (67.3-91.8 %)	3 (60.0 %) 95 %-CI (14.7-94.7 %)	11 (68.8 %) 95 %-CI (41.3-89.0%)
Data Available	22	42	5	19
Ductal carcinoma	18 (81.8 %) 95 %-CI (59.7-94.8 %)	35 (83.3 %) 95 %-CI (68.6-93.0%)	5 (100 %)	17 (89.5 %) 95 %-CI (66.9-98.7 %)

There was no significant difference in tumors of patients carrying additional truncating variant in DNA-repair genes compared to non-carriers in each cohort. ER: Estrogen receptor; PR: Progesterone receptor; HER2: Human Epidermal growth factor receptor 2. Table has been reprinted from Sepahi et al., 2019.

4.8 Rare Variant Association Study (RVAS)

For determination of the influence of rare missense (VUS + pathogenic variants) variants in DNA-repair genes on the AAO of breast cancer in *BRCA1*-positive patients, we used the Burden test and SKAT-O (SNP-set (sequence) Kernel Association Test). As a result, a comprehensive quality control was conducted on the early AAO cohort and controls (see Methods). In terms of (a) variants per sample, (b) rare variant load per gene, (c) transition-transversion ratio, and (d) top 10 PCA components, there were no significant differences observed between the early AAO cohort and controls (Figure 8).

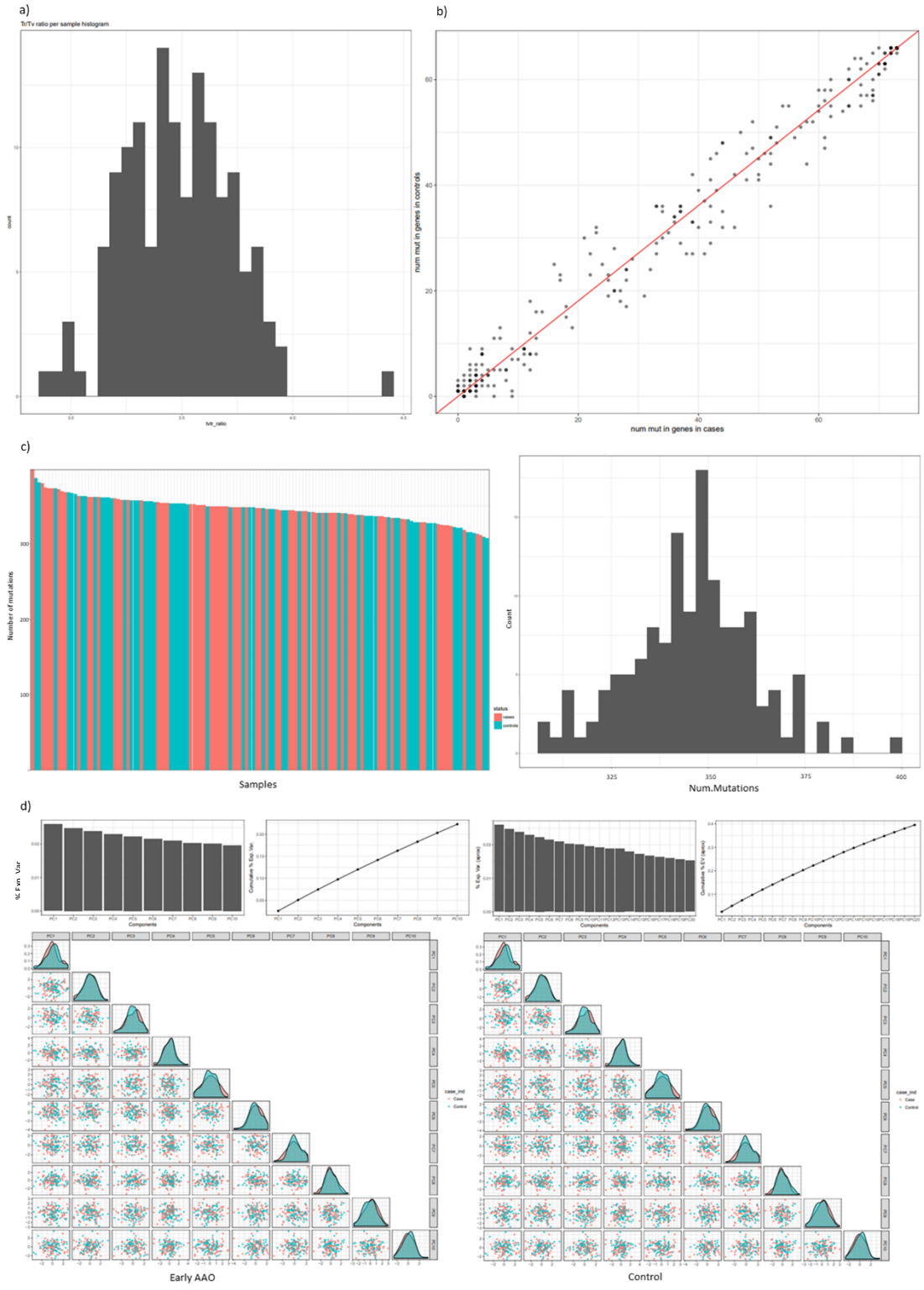

Figure 8. Quality control for the Burden test.
a) transversion(tv)-transition (tr) ratio (x-axis shows the tvtr ratio and y-axis shows the count), b) variant counts in cases (early AAO) x-axis and variant counts in controls (late AAO) y-axis, c) number of variants (y-axis) per sample (x-axis); cases are shown in red and controls in blue, and the histogram of the number of variants (x-axis) per sample, and d) top 10 PCA components (upper panel x-axis: components and y-axis % expected variants) cases depicted in red and controls in blue.

The next step was to remove all common variants (MAF > 1% in EVS, 1KGP, or ExAc) as well as all synonymous variants from both early AAO and control cohorts (Table 6). We used patients from the early AAO cohort as cases and patients from the control cohort as controls in order to search for genes associated with increased risk. Even though there were no significant genes detected after FDR correction, several genes showed significant uncorrected p-values in at least one of the two RVAS tests, suggesting the need for further investigation in independent, larger cohorts. Among these candidate genes (Table 8) are *MYBBP1A* (early AAO: 13, controls: 3), *MRE11* (7:0), TDG (5:0), *WRN* (7:1), *TP53BP1* (10:3), and *REV1* (8:2). A potential risk reducing factor is *PTCH1* (early AAO: 1, controls: 8).

Table 9. The top eight genes identified by the Burden test. q value after FDR correction

Gene	no. affected individual in early AAO cohort	no. affected individual in control cohort	Total no. Early AAO	Total no. Control	p value	q value	Potential effect
MRE11	7	0	73	60	0.0093	0.9	Risk
PTCH1	1	8	73	60	0.0129	0.9	Protective
MYBBP1A	13	3	73	60	0.0169	0.9	Risk
WRN	7	1	73	60	0.0342	0.9	Risk
TDG	5	0	73	60	0.0409	0.9	Risk
TP53BP1	10	3	73	60	0.0415	0.9	Risk
REV1	8	2	73	60	0.0488	0.9	Risk

4.9 Double Heterozygote Patients

In either cohort, there were two cases carrying pathogenic variants in both *BRCA* genes. Patient 1 was diagnosed with breast cancer at the age of 26. She had two first-degree relatives who had breast cancer. The patient had no family history of ovarian cancer, nor did any second-degree relative have any type of cancer. She carried a *BRCA1* pathogenic variant (ENST00000357654: c.1016dupA) as well as a *BRCA2* pathogenic variant (ENST00000544455.1: c.3585_3686delAAAT). The tumor characteristics of this patient were not available. Patient 2 was diagnosed with breast cancer at the age of 63.9 years. There were three first-degree relatives with ovarian cancer in her family, as well as a first-degree relative with breast cancer. Additionally, there was a second-degree relative diagnosed

with breast cancer. A nonsense variant of *BRCA1* (ENST00000357654: c.1687C > T) as well as a nonsense variant of *BRCA2* (ENST00000544455.1: c.8875G > T) were detected in her. Additionally, a truncating variant was found in *EME2*, a DSBR gene (ENST00000568449: c.541_544delGCTG). The immunohistochemical staining revealed a triple negative tumor.

4.10 Splicing Pattern in Genes Involved in Cancer

Here we present a targeted RNA-Seq assay to analyze abnormal splicing junctions either predicted by *in silico* tools or were expected to have a deleterious effect of splicing based on variant characteristics. These variants were initially detected in the patients during the course of diagnosis.

Five variants detected in cancer susceptibility genes detected in patients either with a personal history or a family history of cancer (predictive test or co-segregation analysis) were selected for RNA analysis. These variants have been classified as variants of uncertain significance based on the scientific and familial data. To further assess the clinical significance of these variants, cDNA- and targeted RNA-sequencing have been performed for these variants. cDNA analysis was performed prior to RNA-sequencing.

The RNA splicing junctions of 34 (Appendix) HBOC genes were targeted with an exon capture enrichment approach. The average number of sequence-reads generated per sample was about 29 million, with a minimum of 15 million reads per sample.

Table 10 shows the summary of results of cDNA- and targeted RNA-sequencing.

Table 10. Summary of the targeted RNA-seq and cDNA analysis on 5 variants

Gene	Variant	(Predicted) RNA effect	cDNA sequencing results	Targeted RNA-seq results
TP53	c.375+5 G>A	not predictable	r.176_375del,p.Gly59Val*fs**23	7.4% of the detected transcripts are aberrant
STK11	c.597 G>A	synonymous p.Glu199=	r.465_597del;p.Tyr156His*fs**87	7.7% of the detected transcripts are aberrant
CHEK2	c.592 G>A	missense p.Val198Ile	r.445_592del;p.Glu149Phe*fs**7	Not detected
BRCA2	c.8755-2 A>G	skip of coding exon 21	r.8755_8953del;p.Gly2919Leu*fs**2	Not detected
BRCA1	c.4093 T>G	missense p.Leu365Val	r.4093_4096del;p.Leu1365Val*fs**27	Not detected

Variant 1. This intronic variant (*TP53*: c.375+5G>A) results from a substitution of a G to A five nucleotides after the exon 4 (3rd coding exon, transcript ID: NM_000546.6) of *TP53* gene. In addition to the normal transcript, cDNA analysis revealed the presence of an aberrant transcript. Sequencing of this aberrant transcript showed that this variant results in a 200bps deletion within the exon 4

leading to a shift in the reading frame. Consistent with the cDNA-sequencing results, 464 RNA-seq reads (7.4%) supported a new splice junction within exon 4 which results in deletion of 200 bps (Figure 7c). This transcript was not observed in the controls. Moreover, Sashimi plots of this variant shows coverage within the intronic region, which may imply presence of another set of aberrant transcripts with intronic region inclusion (Figure 7a and b).

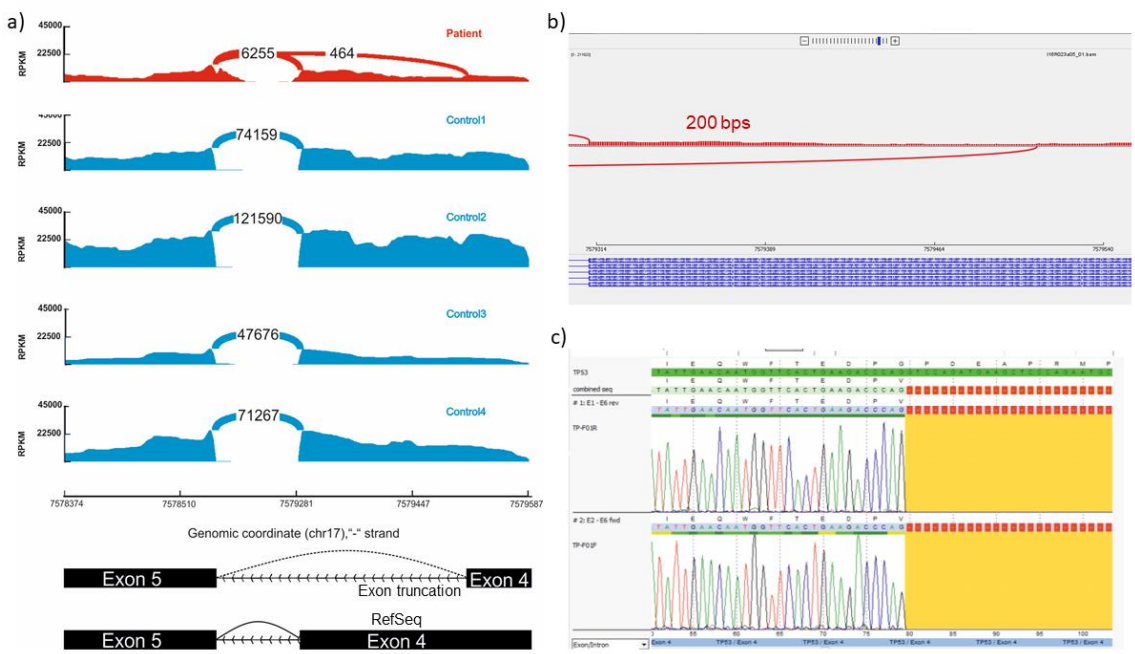

Figure 9. Aberrant splicing detection *TP53*
a) TP53. Sashimi plot of exon truncation due to activation of a cryptic splice junction (patient in red and controls in blue). The RNA coverage is given as the RPKM-value and the number of reads spanning the given intron is indicated on the exon-connecting lines. At the bottom the gene model of the RefSeq annotation is depicted and the aberrant splice junction is depicted by dotted line. b) a close-up of the Sashimi plot at the cryptic splice site. C) Sanger sequencing of the isolated cDNA is consistent with RNA-seq results

Variant 2. This variant (*STK11*: c.597G>A) is a synonymous variant which results from a substitution of a highly conserved G to A at the last nucleotide of exon 4. In silico splice site prediction tools such as SpliceSite Finder Like, MaxEntScan, NNSPLICE, GeneSplicer, HSF predicted a reduction in the capacity of the canonical splice site. cDNA analysis revealed the presence of an aberrant transcript. Sequencing of this aberrant transcript showed that this variant results in results in the deletion of exon 4. Consistent with the cDNA-sequencing results, 1122 RNA-seq reads (7.7%) supported skipping of exon 4. Deletion of exon 4

disrupts the reading frame of the protein (Figure 8c). This transcript was observed in the controls as very weak splice transcripts (Figure 8a).

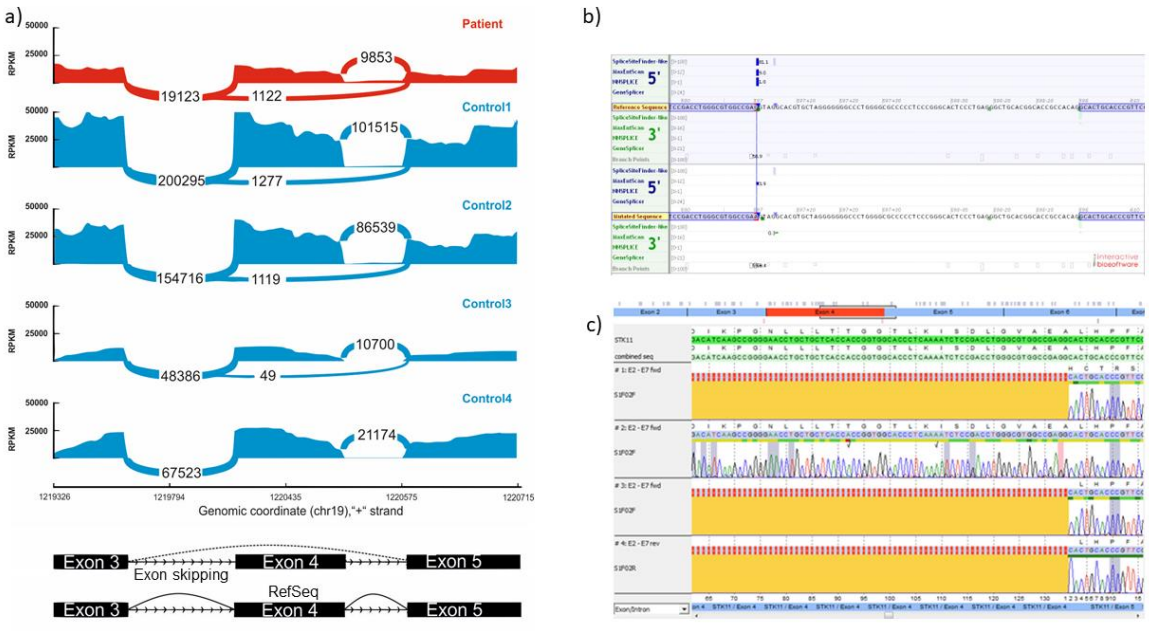

Figure 10. Aberrant splicing detection *STK11*
a) *STK11.* Sashimi plot of exon skipping (patient in red and controls in blue). The RNA coverage is given as the RPKM-value and the number of reads spanning the given intron is indicated on the exon-connecting lines. At the bottom the gene model of the RefSeq annotation is depicted and the aberrant splice junction is depicted by dotted line. b) In silico splice prediction tools predict reduction in the capacity of the canonical splice site. C) Sanger sequencing of the isolated cDNA is consistent with RNA-seq results

Variant 3. This variant (*CHEK2*: c.592G>A; p.Val198Ile) affects the first nucleotide of the fourth coding exon. It is a missense variant resulting in the substitution of a highly conserved aspartic acid with an asparagine. Splice prediction tools did not predict an impact on splicing. However, RT-PCR revealed an aberrant fragment whose sequencing demonstrated that this variant results in the deletion of coding exon 3. This aberrant transcript was not detected in the controls. Targeted RNA-seq could not detect the aberrant transcript.

Variant 4. This variant (BRCA2:c.8755-2A>G) located within the canonical splice acceptor of coding exon 21. Splice prediction tools predict a deleterious effect on the splicing. Sequencing of the aberrant fragment detected by RT-PCR and was absent from the controls, revealed skipping of coding exon 21. Skipping of this out of frame exon results in a shift in the reading frame of the protein and thus a

premature termination (Table 10). This result however could not be confirmed via RNA-sequencing.

Variant 5. This variant (*BRCA1*: c.4093 T>G; p.Leu365Val) is a missense variant that affects a highly conserved nucleotide and resulting in the substitution of a moderately conserved leucine with valine. This nucleotide substitution results in the activation of a new splice junction. Activation of this splice donor leads to an out of frame deletion of 4 nucleotide and premature translation termination (Table 10). However, these results could not be confirmed via targeted RNA-sequencing.

4.11 Association Between Catechol-O-Methyltransferase (*COMT*) Polymorphisms and Breast Cancer in *BRCA1/2* Positive Patients

4.11.1 Participants Characteristics

Using amplicon sequencing, where the material was available, the entire coding region of *COMT* was sequenced in patients from early AAO and the control cohort. In addition, 32 *BRCA2*-PV carrier women who were eligible for the age criteria in early AAO and the control cohort were also screened for *COMT* gene. Out of these 32 *BRCA2*-PV carriers, 17 fell in the control cohort and 15 were eligible for early AAO cohort. Patients with prophylactic oophorectomy and mastectomy, and ovarian cancer were excluded from the analysis. In total we have sequenced 157 patients. Seventy-seven patients in the control cohort either were diagnosed with breast cancer at an age older than 60 years (n = 32; 41.6%, median age at onset, 70.6 years (IQR, 62.8-74.1)) or have not developed breast cancer by the time of sample taking (n = 45; 58.4%, median age, 68 years; IQR, 63.3-73.2 years). Eighty patients manifested breast cancer at an age younger than 35 years (median age at onset, 27 years; interquartile range (IQR) 25-27 years).

4.11.2 Association of *COMT* Val158Met Polymorphism and Breast Cancer

Allele frequency was calculated for the mutated low-activity allele ($COMT^{Met}$) in each cohort. It was present at an allele frequency of 42.5% in early AAO cohort compared to 50% in control cohort.

Table 11. Calculated odds ratio (ORs), COMT genotype- and allele frequencies for BRCA1/2-PV carriers in early AAO and control cohort.

Genotype	Control (%) (n = 77)	Early AAO (%) (n=80)	OR	95% CI	P
Codominant model					
$COMT^{Val/Val}$	15 (19.5%)	28 (35%)	1		
$COMT^{Val/Met}$ vs. $COMT^{Val/Val}$	47 (61%) vs 15 (19.5%)	36 (45%) vs 28 (35%)	0.4	0.2-0.9	0.02*
$COMT^{Met/Met}$ vs. $COMT^{Val/Val}$	15 (19.5%) vs 15 (19.5%)	16 (20%) vs 28 (35%)	0.6	0.2-1.4	0.34
$COMT^{Val/Met}$ vs. $COMT^{Met/Met}$	47 (61%) vs 15 (19.5%)	36 (45%) vs 16 (20%)	0.7	0.3-1.6	0.53
Dominant model					
$COMT^{Val/Met}$ + $COMT^{Met/Met}$ vs. $COMT^{Val/Val}$	62 (80.5%) vs. 15 (19.5%)	52 (65%) vs. 28 (35%)	0.4	0.2-0.9	0.03*
Recessive model					

$COMT^{Met/Met}$ vs. $COMT^{Val/Val}$+ $COMT^{Val/Met}$	62 (80.5%) vs. 15 (19.5%)	64 (80%) vs. 16 (29%)	1.3	0.5-2.2	>0.9
Allele frequencies					
COMT Val	77 (50%)	92 (57.5%)			
COMT Met	77 (50%)	68 (42.5%)			

The odds ratio was calculated to assess the risk of breast cancer among patients in early breast cancer onset and control cohort. The frequencies of the *COMT* genotype among each cohort are shown in Table 11. The odds ratio is calculated for the presence of one or two mutated low-activity allele under codominant, dominant and homozygous models (Table 11). Slightly significant results were obtained when we compared the presence of one low-activity allele with no low-activity allele, meaning that 61% of the individuals in control cohort carried one low activity allele compared to 19.5% of the women in this group who had no mutated allele. These results were significantly lower in the early AAO cohort. 45% of the patients in this cohort carried one low activity allele while 35% had no low activity allele (odds ratio: 0.4; 95%-CI 0.2-0.9; p-value for the Fisher exact test 0.02). A similar result was obtained under the dominant model where 80.5% of the individuals in control cohort carried either one or two mutated low-activity allele while 65% of the patients in early AAO cohort carried either one or two copies of the low-activity allele (odds ratio: 0.4; 95%-CI 0.2-0.9; p-value for the Fisher exact test 0.03). When we combined all the affected patients from the early AAO and control cohort (n= 112) and compared them with the small subset of the patients in the control cohort who were above 60 years of age and remained unaffected (n= 45) by the time of sample taking, we did not find any significant results regarding the presence of one or two copies of mutated low-activity COMT allele (Table 12).

Table 12. Calculated odds ratio (ORs), COMT genotype- and allele frequencies for affected and unaffected BRCA1/2-PV carriers.

Genotype	Unaffected (%) (n = 45)	Affected (%) (n=112)	OR	95% CI	P
Codominant model					
$COMT^{Val/Val}$	8 (17.8%)	35 (31.25%)	1		
$COMT^{Val/Met}$ vs. $COMT^{Val/Val}$	27 (60%)	56 (50%)	0.5	0.2-1.1	0.14
$COMT^{Met/Met}$ vs. $COMT^{Val/Val}$	10 (22.2%)	21 (18.75%)	0.5	0.2-1.4	0.27
$COMT^{Val/Met}$ vs. $COMT^{Met/Met}$	27 (60%)	56 (50%)	0.9	0.4-2.5	>0.9

Dominant model					
$COMT^{Val/Met}$ + $COMT^{Met/Met}$ vs. $COMT^{Val/Val}$	37 (82.2%) vs. 8 (17.8)	77 (68.75%) vs 35 (31.25%)	0.5	0.2-1.1	0.11
Recessive model					
$COMT^{Met/Met}$ vs $COMT^{Val/Val}$+ $COMT^{Val/Met}$	35 (77.8%) vs 10 (22.2%)	91 (81.25) vs. 21 (18.75)	0.8	0.3-1.9	0.66
Allele frequencies					
COMT Val	43 (47.8%)	126 (56.25%)			
COMT Met	47 (52.2%)	98 (34.75%)			

4.11.3 Association of *COMT* c.609C>T; p.Leu203Leu and Breast Bancer

The allele frequency for the synonymous variant c.609C>T in COMT was determined for the *BRCA1/2*-PV carriers in the two cohorts of early AAO and controls. This variant was present among the *BRCA1/2*-PV carriers with early AAO at a minor allele frequency of 2.6%. This variant was detected at a minor allele frequency of 1.25% in control cohort (Table 13). When the patients were divided in two groups of affected and unaffected, the minor allele frequency was slightly higher in unaffected patients (3.3% vs 1.3%); however, this difference was not statistically significant (odds ratio: 0.4; 95%-CI 0.08-1.4; p-value for the Fisher exact test 0.3) (Table 13).

Table 13. Calculated odds ratio (ORs), COMT genotype- and allele frequencies for Early and late AAO as well as affected and unaffected BRCA1/2-PV carriers.

Genotype	**Late AAO (%) (n = 77)**	**Early AAO (%) (n=80)**	**OR**	**95% CI**	**P**
Codominant model					
COMT CC	73 (94.8)	78 (97.5%)			
COMT CT	4 (5.2%)	2 (2.5%)	0.5	0.08-2.06	0.4
Allele frequencies					
COMT T	2.6 %	1.25%			
Genotype	**Unaffected (%) (n = 45)**	**Affected (%) (n=112)**	**OR**	**95% CI**	**P**
Codominant model					
COMT CC	42 (93.3%)	109 (97.3%)			
COMT CT	3 (6.7%)	3 (2.7%)	0.4	0.08-1.7	0.3
Allele frequencies					
COMT T	3 (3.3%)	3 (1.3%)			

5 Discussion

Rare deleterious variants of high-penetrance genes such as *BRCA1* and *BRCA2* account for 5% to 10% of breast cancer cases. *BRCA1* and *BRCA2* are critical components of the DNA repair pathway. The risk of developing breast cancer by age 80 is 72% for female *BRCA1* PV carriers (Kuchenbaecker, Hopper, et al., 2017). However, although pathogenic variants of *BRCA1* are associated with the highest penetrance of HBOC, the reasons for inter-individual and intra-familial variation in penetrance remain unknown. Genetic and environmental factors have been described as contributing to the small risk difference observed among *BRCA1*-PV carriers. Determining the penetrance of pathogenic variants within *BRCA1* is critical clinically. This is critical to avoid overtreating carriers who will not develop breast cancer in their lifetime or who may develop breast cancer in the future. By examining 311 DNA-repair genes which contribute to genome stability via DNA-repair pathways along with *BRCA1* and *BRCA2*, we hoped to address this issue by investigating the differences in AAO of breast cancer among *BRCA1*-PV carriers.

5.1 Extreme Phenotype Sampling and Cohorts' Characteristics

The prognosis of cancer patients is influenced by a number of intrinsic and extrinsic factors due to the complexity and multifactorial nature of this disease. Due to their limited penetrance, it is very difficult to identify genetic alterations associated with significantly different prognosis. In order to identify genetic features associated with characteristic outcomes, concentrating studies on a small number of individuals with extremely differentiated phenotypes may be a more efficient strategy. Instead of performing studies on the entire population, studies should focus on individuals with extremely differentiated phenotypes. These individuals are more likely to carry the characteristic genotypes responsible for the differences in prognosis seen in clinics (Perez-Gracia et al., 2002).

An extreme phenotype sampling design is used when the covariate of interest is either too expensive or unattainable for all individuals in a genetic association study (Bjørnland et al., 2018).

This study utilized a preference-selection approach to enrich rare variants by selecting carriers who are the most informative cases. As a result, we chose two highly selected cohorts at the extreme ends of age at onset of hereditary breast cancer in order to identify differences in mutational load. The hypothesis was that inherited truncating variants in DNA-repair genes, which are partners in maintaining genome integrity with BRCA1, might reduce the onset age of hereditary breast cancer by interacting with BRCA1.

In this study 152 patients who carried a PV in *BRCA1* were recruited. Patients were stratified based on the age at breast cancer onset either diagnosed with primary breast cancer at an age younger than 35 years (early AAO cohort) or remained cancer-free until the age of 60 years. Therefore, patients who had developed breast cancer between age 35 and 60 were excluded from the analysis. Since the inclusion criteria was breast cancer onset, patients with ovarian cancer onset were disqualified. None of the patients has undergone prophylactic mastectomy prior to the diagnosis of breast cancer. We have excluded patients who underwent risk reducing salpingo-oophorectomy before the age of 45. The risks of OC have been proven to be reduced by risk-reducing mastectomy (RRM) and risk-reducing salpingo-oophorectomy (RRSO), although the association between the latter and BC risk is unclear. A consistent reduction in BC risk of up to 50% has been reported in early studies following RRSO (Rebbeck, Levin, et al., 1999). A later analysis that considered RRSO as a time-varying covariate refuted these results, i.e., assuming that the association between RRSO and breast cancer only begins following surgery rather than counting all years without cancer. RRSO has been proven not to be associated with BC risk, when considered as a time-varying covariate, among women with BRCA1 pathogenic variants and women with other high-risk factors. The cumulative breast cancer risk curves, however, indicate that RRSO has a slightly protective effect when performed prior to menopause (Heemskerk-Gerritsen et al., 2015).

In the first stage sampling the inclusion criteria was presence of a pathogenic variant in the patients. One patient who carried a missense variant

c.5365G>A:p.Ala1789Thr within the BRCT2 domain (UniProt: P38398) of BRCA1 protein. This variant results in the substitution of a neutral and non-polar amino acid with a neutral and polar amino acid. At the time of sample collection this variant has been given an uncertain classification. The data regarding the pathogenicity of this variant is conflicting (ClinVar variation ID: 55552). The variant has been initially reported in two patients from a family with breast and ovarian cancer. In vitro functional analyses revealed that this variant alters the DNA DSBR (Guidugli et al., 2011).

Since this variant was detected in one of the individuals with no breast cancer onset, considering the uncertain clinical significance of this variant, we have excluded this individual from further analyses.

5.2 Allelic Variation Among *BRCA1* Positive Patients

In 2001, Thompson and Easton reported the existence of allelic variation in *BRCA1* pathogenic variants, which was the subject of more recent research by Rebbeck and colleagues (2015), in which allelic variation in *BRCA1* pathogenic variants was found to be an important contributing factor in the variation in breast cancer risk among HBOC patients with respect to ovarian cancer. According to Rebbeck and colleagues, several regions have been identified as being associated with a higher risk for breast cancer compared to ovarian cancer (breast cancer cluster regions = BCCRs) and one region as having a higher risk for ovarian cancer compared to breast cancer (OCCR) (Rebbeck et al., 2015; Thompson & Easton, 2002).

Here we have evaluated the distribution of the identified variant along BRCA1 gene between the two cohorts. About 22% of the identified BRCA1 PVs in early AAO cohort were located within the BCCR2' (c.5261-c.5563) while this region encompasses about 15% in control cohort. The majority of BRCA1 PVs in control cohort (~42%) were located within the OCCR, this number is much less in early AAO (~26%). Among the two cohorts, there was no difference in mutational position between BCCR and OCCR; however, there was a non-significant difference in variant load between the two cohorts (p-value = 0.07). While the difference was not statistically significant, it is worthwhile to note that pathogenic

variants in OCCR are not only linked to increased ovarian cancer risk, but they are also linked to a reduction in breast cancer risk as well. Therefore, these results are consistent with the age at breast cancer onset in each cohort.

Regarding the type of pathogenic variants, previous studies have shown that *BRCA1* missense PVs were associated with lower breast cancer risks compared to *BRCA1* truncating variants, particularly for missense PVs located within the *BRCA1* C-terminal domain (H. Li et al., 2022; Scott et al., 2003). There were no differences among our cohorts in terms of the type of *BRCA1* PVs. In both cohorts, the majority of the PVs were truncating variants (76.7% in the early AAO vs 78.3% in the control cohort) mainly located within exon 11 corresponding to the OCCR (Figure 2). It was the most common pathogenic missense variant in both cohorts c.181T>G: p.Cys61Gly, whereas the missense variant c.5090G > A: p.Arg1699Gln was found exclusively in two of the patients in the control cohort.

The most frequently detected PV (c.5266dupC:p.Gln1756fs a.k.a c.5266_5267insC, and c.5382insC or c.5385insC using alternate nomenclature) in early AAO cohort is a common founder variant in the Ashkenazi Jewish population (Roa, Benjamin, Boyd Alfred A., Volcik Kelly, 1996), while the most frequently identified PV (c.181T>G: p.Cys61Gly) in control cohort is a European founder variant (Bogdanova et al., 2010; Domchek et al., 2010) located within the RING domain of the BRCA1 protein. This variant has been shown to confer a significantly lower hazard ratio in women aged≥50 compared to protein truncating variants in BRCA1 (H. Li et al., 2022). It was the most common pathogenic missense variant in both early AAO and control cohorts c.181T>G: p.Cys61Gly, whereas the missense variant c.5090G > A: p.Arg1699Gln was found exclusively in two of the patients in the control cohort. This variant has been shown to be a hypomorphic *BRCA1* variant which confers a reduced cumulative risk of breast cancer compared with the average *BRCA1* truncating variant (Moghadasi et al., 2017). The presence of this variant only in the control cohort is in line with the expected decreased breast cancer risk for this variant.

5.3 Environmental Factors

Several studies have been published in recent years that examine the factors that modify cancer risk among carriers of the *BRCA1/2* pathogenic variant (Friebel et al., 2014). Researchers included factors such as age at first live birth, menarche, parity, oral contraceptive use, tamoxifen, breastfeeding, mammography, and coffee in their meta-analysis. Several other factors were also investigated, but not meta-analyzed, including abortion, infertility, menopause, miscarriage, hormone replacement therapy, weight, radiation and x-rays. Many factors showed no effect.

Previously, a reverse association between age at menarche and the risk of breast cancer among *BRCA1*-PV carriers has been reported (Kotsopoulos et al., 2005). They have found that a later age at menarche is associated with a reduced breast cancer risk among *BRCA1*-PV carriers. Consistent with the previous study by Kotsopoulos and colleagues, in our study patients in the control cohort (who were cancer free by age 60) showed a significantly older age at menarche in contrast to patients in early AAO. This difference in the age at menarche among the two cohorts may explain a small proportion of the differences in the age at breast cancer onset. This implies that this factor can act as a confounder in our analysis of the effect of DNA-repair PV on the age at breast cancer onset.

Unlike the previous meta-analyses which have shown a null association between the parous and nulliparous with the risk of breast cancer among *BRCA1*-PV carriers (Lecarpentier et al., 2012; Milne & Antoniou, 2011), our cohorts showed a significant difference when we compared the number of parous against the nulliparous women in both cohorts (p-value: <0.0001). Patients in control cohort tend to be more parous (93% vs 29%).

Friebel and colleagues showed that having three live births or more decreases the breast cancer risk in *BRCA1*-PV carriers when compared to nulliparous *BRCA1*-PV carriers (Friebel et al., 2014). Due to the small sample size, we were not able to categorize cohorts into the 5 subgroups suggested by Friebel and colleagues; however, our results showed that patients in the control cohort have a greater number of pregnancies (p-value: < 0.0001). We cannot exclude the

possibility that these results are biased due to the fact that patients in the control cohort were at an older age at the time of the analysis.

When comparing parity at age 30 or older with parity at a younger age, a meta-analysis of two cohort analyses revealed a statistically significant reduction in effect size (Friebel et al., 2014). Because of the small sample size and the extreme phenotypic sampling based on breast cancer onset age, we could not evaluate the same age groups as previously reported in the meta-analyses. However, when we compared the mean age at first pregnancy among parous women in our cohorts, we found different results. In our cohorts, patients in control cohort showed a significantly younger age at first pregnancy in contrast with the patients in control cohort. These results are in contrast with the previously published meta-analysis results which suggested a protective effect against the breast cancer for *BRCA1*-PV carriers aged 30 years or older vs women younger than 39 years (Friebel et al., 2014).

Depending on the type of the analysis (prospective or retrospective; case-control or cohort analysis) the data regarding the effect of oral contraceptive use and the risk of breast cancer among *BRCA1*-positive patients can vary (Schrijver et al., 2018). Among *BRCA1*-PV carriers, a meta-analysis of case-control studies revealed no difference between those who ever used OC and those who never used OC, but a combined hazard ratio of cohort studies suggested an increased risk (Friebel et al., 2014).

We have identified a significant difference in the use of oral contraceptive among *BRCA1*-positive women in the early AAO and control cohort. Our results showed that a vast majority of women in the early AAO cohort (66%) had ever an experience of OCP use, while only less than a third of women in the control cohort (23%) had ever used OCP. High prevalence of oral contraceptive use among patients in the early AAO cohort is in line with the observed fewer number of pregnancies in this cohort.

These differences in the OCP use and the number of pregnancies among the two cohorts can be a result of significant difference in the birth cohorts. Women in the

control cohort were born between 1919 and 1954 and patients in early AAO cohort were born between 1960 and 1993. These results are consistent with the previous studies which have shown that female *BRCA1*-PV carriers born in 1958 or later are subject to a higher risk of breast cancer (A. Antoniou et al., 2003; Kroiss et al., 2005). A more recent study has demonstrated a genetic anticipation of breast cancer among *BRCA1/2*-PV carriers. In this study it was found that both *BRCA*-PV carriers and non-carriers were diagnosed with breast cancer at an earlier age in successive generations. Observing a downshift in age at diagnosis in non-carrier pairs suggests that other factors (environmental, lifestyle, or social) may also be involved (Kedmi et al., 2022).

Moreover, oral contraceptive pills were first introduced in 1950 and it seems that they were initially used as a mean of menstrual regulation and only by late 1960s they were prescribed for the indication of contraception (Liao & Dollin, 2012). Considering the year of birth distribution in our cohorts, this can explain the differences between the two cohorts regarding the oral contraceptive use.

The presence of a positive family history of breast cancer is an additional risk factor for breast cancer. Multiple studies using a variety of study designs have demonstrated an increased risk of breast cancer in women with a family history of the disease. There is, however, a wide variation in the extent of this risk according to the nature of the family history (type of family member affected, age at which the relative developed breast cancer, and number of relatives affected) and may also vary based on the individual's age. The family history of having a first-degree relative with an age at onset of 50 years or older increases the relative risk of breast cancer as high as 1.8 (Pharoah et al., 1997).

In our study, patients from the early AAO cohort came from smaller families and had less strong family history when compared to the control cohort. In the control cohort, most patients had two or more first-degree relatives with breast and/or ovarian cancer. While the majority of patients in the early AAO cohort had either one or no first-degree relatives with breast and/or ovarian cancer. Having a strong family history of breast cancer in the first degree is not an indicator of early breast cancer onset in *BRCA1* positive patients, according to these results. In 2010

Panchal and colleagues evaluated the effect of a positive family history on the age at breast cancer onset among *BRCA1/2* positive patients. Although they identified a modest trend for *BRCA1* carriers, it was not statistically significant (Panchal et al., 2010). Moreover, Brewer and colleagues showed that having several affected family members may not represent a very high risk if a woman comes from a very large family (Brewer et al., 2017).

5.4 Rare Truncating Variants in DNA-Repair Genes

Genome-wide association studies on the modifiers of breast cancer risk identified several loci that all together can explain approx. 2.2% of the variation in the risk of breast cancer among *BRCA1* positive patients (Milne & Antoniou, 2011). To assess the effect of rare variants on the differences in breast cancer risk among *BRCA1* carriers, we focused on the subset of genes, which contribute to genome stability through different DNA-repair pathways and inherited and acquired alterations in these genes are critical mechanisms in the genesis of malignant tumors. It is hypothesized that women with an early age at breast cancer onset are enriched for heritable defects in DNA repair genes in addition to the main causal *BRCA1* variant.

As it is already known for *BRCA1* and *BRCA1*, haploinsufficiency of genes involved in DNA damage response is expected to have a deleterious effect on their protein function. Therefore, for this analysis, we limited the study of the sum-effect of rare variants only to truncating variants in DNA-repair genes. We observed that these truncating variants are not highly recurrent in a single gene or in specific DNA-repair pathways. However, variants affecting DNA double-strand repair and single-strand repair were highly enriched in both cohorts. Although truncating variants in DSBR and SSBR were found more frequently in the early AAO cohort, these results were not statistically significant.

The sum-effect of rare variants in DNA-repair genes has previously been studied in different cancer types. Pritchard and colleagues showed that the incidence of germline pathogenic variants in genes mediating DNA-repair processes was significantly higher among men with metastatic prostate cancer than the incidence among men with localized prostate cancer (Pritchard et al., 2016).

Brohl and colleagues also noted that their cohort of Ewing sarcoma patients was highly enriched for pathogenic/likely pathogenic variants involved in DNA-repair responses. In pathway analysis, significant enrichment was observed for hereditary breast cancer signaling, DNA repair pathways, and, in particular, DNA double-strand break repair pathways (Brohl et al., 2017).

To the best of our knowledge, it is the first study to investigate the effect of rare variants on the penetrance and the age at onset of breast cancer among *BRCA1*-PV carriers. We observed that patients in the early AAO cohort tend to be more enriched for truncating variants in DNA-repair pathways (35.6% in early AAO vs 26.7% in the control cohort). As the patients in each cohort were not selected and matched for the modifying environmental factors, and the cohorts were highly heterogeneous for these factors we conducted a multivariable logistic regression model to adjust for these confounding factors. The result of this analysis was replicated for a subset of the cohorts who were matched for the positive family history. We observed an enrichment of truncating variants in DNA-repair genes among women positive for *BRCA1*-PV and earlier breast cancer onset, although it was not statistically significant. Larger cohorts may prove that carrying an additional truncating variant in DNA-repair genes may decrease the age at breast cancer onset among *BRCA1*-PV carriers.

5.5 Rare Variants Vs. Common Variants (Polygenic Risk Score)

Genome wide association studies (GWAS) have successfully identified genetic factors that modify breast cancer risk, both in the general population and in *BRCA-PV* carriers. More than 100 common susceptibility variants for breast cancer each has a small effect size (Milne & Antoniou, 2016). Likewise, over 50 common breast cancer susceptibility variants have been linked with breast cancer risk in *BRCA1*- and *BRCA2*-PV carriers. The combined effect of these factors summed up as polygenic risk scores (PRS) which results in significant differences in the absolute risk of developing breast cancer for *BRCA1/2*-PV carriers at the extremes of the PRS distribution (Kuchenbaecker, McGuffog, et al., 2017). Coignard and colleagues used a novel case-only analysis strategy to uncover novel genetic modifiers of breast cancer risk for *BRCA1*- and *BRCA2*-

PV carriers and to clarify the effects of known breast cancer susceptibility SNPs on breast cancer risk for carriers. They found that most SNPs associated with risk in the general population are also associated with risk for PV carriers, however, their average effect sizes are smaller. As a result, a large proportion of risk variation among BRCA-PV carriers remains unknown. By analyzing rare variants in genes involved in DNA damage response and genome integrity maintenance pathways as well as genes that interact with *BRCA1* in contrast to GWAS studies that rely on common variants, our study aimed to predict breast cancer risk and breast cancer age at onset in *BRCA1*-PV carriers. To enrich for rare variants, we preferentially selected those *BRCA1*-PV carriers who showed the extreme phenotype. For this reason, the extreme ends of age at onset of hereditary breast cancer were chosen. We aimed to identify differences in the mutational load in these two highly selected cohorts. We hypothesized that inherited truncating variants in DNA-repair genes, which are partner components of *BRCA1* in the maintenance of genome integrity, are likely to interact with *BRCA1* by reducing the age at onset of hereditary breast carcinoma.

5.6 Rare Variant Association Study

Rare genetic variants, defined here as alleles with a frequency below 1% to 5%, have the potential to play a significant role in the development of complex diseases and traits. For rare genetic variants, standard methods used to test for association with common variants are underpowered unless sample sizes or effect sizes are very large. As an alternative, burden tests that assess the cumulative effects of multiple variants within a genomic region can be utilized. A burden test is a method that collapses or sums the rare variants within a region into a single value and then tests its association with the trait in question (Lee et al., 2014).

Despite the relatively small sample size a rare variant association study (RVAS) was conducted using the SKAT-O and Burden tests to shed light on the role of rare variants in the risk of early onset breast cancer among *BRCA1* positive patients. The SKAT-O as a linear combination of the burden and SKAT tests achieves robust power regardless of whether a given gene has a significant

proportion of causal variants exerting effects in the same direction or many noncausal variants or noncausal variants exerting effects in the opposite direction (Lee et al., 2014).

Following multiple testing corrections, neither the SKAT-O nor Burden tests were statistically significant. However, *MRE11* was the top-ranked gene in the Burden test. The MRE11 protein is part of the MRN complex (MRE11, RAD50, and NBS1). The complex is responsible for sensing DNA double-strand breaks and initiating the repair process (Trujillo et al., 1998). Genomic instability syndromes associated with germline variants in *MRE11*, *NBS1* or *RAD50* are characterized by immunodeficiency, hypersensitivity to radiation, and cancer predisposition. Pathogenic variants in the *MRE11* gene are associated with ataxia-telangiectasia-like disorder (Stewart et al., 1999). Pathogenic variations in *NBS1* are responsible for Nijmegen breakage syndrome (NBS) (Varon et al., 1998). RAD50 deficiency has also been described in a case of NBS-like disorder. In addition to breast cancer predisposition, variants in the *MRE11*, *RAD50*, and *NBS1* genes may also increase other cancers risk. Recently, Gupta and colleagues reported an association between *MRE11* variants and triple negative breast cancer (Gupta et al., 2017). An extensive investigation of MRN in sporadic breast cancer has revealed that low MRN is associated with high tumor grade, high mitotic index, as well as ER negative breast cancer (Alblihy et al., n.d.). According to a univariate analysis, low nuclear MRE11 expression and low nuclear RAD50 expression are associated with an adverse prognosis.

5.7 Pathological Characteristics

This study also included an assessment of the tumor histology and immunohistochemistry characteristics of the tumors. This was done in order to determine whether these tumor characteristics in the two age groups of *BRCA1*-PV carriers were distinct from one another. *BRCA1*-associated breast cancers have been extensively studied for their clinicopathological characteristics. Previous studies have found that *BRCA1*-positive tumors exhibit higher tumor grade, lower estrogen receptor (ER) expression, and lower progesterone receptor (PR) expression (Foulkes et al., 2000; Loman et al., 1998; Noguchi et

al., 1999). However, there is less research on ER and PR expression in young and older *BRCA1*-associated breast cancer patients. There was a significant difference in the prevalence of ER and PR positivity between *BRCA1*-positive patients whose age at onset was younger than 50 years of age and *BRCA1*-positive patients whose age at onset was over 50 years of age, as reported by Vaziri and colleagues (Vaziri et al., 2001). Eerola and colleagues reported similar results from their study in 2005, in which they compared the results of *BRCA1/2* positive families to those of *BRCA1/2*-negative families. *BRCA1*-positive, premenopausal patients (age at diagnosis under 50 years of age) showed a significant difference in the level of ER negativity. In addition, these patients were also found to have higher-grade tumors when compared to patients who were postmenopausal (Eerola et al., 2005). In a study by Székely and colleagues, they stratified breast cancer patients into two categories: early AAO (<35) and late AAO (>70). In terms of menstrual and reproductive factors as well as the histological characteristics and immunophenotype of the tumors, there were statistically significant differences between the two groups. According to their results, triple negative breast tumors were more common among patients with early AAO than among those with late AAO. The tumors in the early AAO cohort were also of a higher grade than those in the late AAO cohort (Borbála Székely et al., 2010).

Here, in this study we have stratified the *BRCA1*-positive BC patients based on the AAO. Patients in the control cohort were either cancer free or had developed breast cancer as late as 60 years of age. Tumor characteristics in this group were compared with tumor characteristics in the early AAO cohort. Unlike the previous studies, we did not identify a significant difference in the immunohistochemical features of the tumors in the two age groups. While tumors in control cohort are solely of ductal type, tumors in the early AAO cohort are mainly of ductal type (85.5%). A small subset of the tumors in the early AAO cohort was of medullary 9.7%) and lobular type (3.2%). The majority of the tumors in both cohorts were of III grade (80.3% in early AAO cohort vs 63.7% in control cohort). Estrogen, progesterone and human epidermal receptor expression were not significantly different.

We are also able to demonstrate that carrying a truncating PV in a DNA-repair gene in addition to carrying a PV in BRCA1 does not change the characteristics of the tumor. In patients with additional truncating PV in DNA-repair genes in contrast to those with no additional truncating PV, there were no differences in the histology or histochemical features of the tumors.

5.8 Splicing Pattern in Genes Involved in Cancer

HBOC molecular diagnostic laboratories face significant challenges when interpreting VUS, especially given the increasing number of genes implicated in the syndrome. According to ACMG/AMP framework, well established in vivo or in vitro functional studies are considered a strong piece of evidence for variant interpretation (Richards et al., 2015). A major contribution to understanding the functional significance of these variants is an investigation of their effects on RNA splicing. This is done to determine if they could cause aberrant RNAs, which could lead to protein dysfunction or loss of function. In theory, any detected variant could affect RNA splicing (Soukarieh et al., 2016). In this study, we developed a targeted RNA-Seq approach that detects and quantifies splicing junctions in many genes at once.

Here we have studied a few variants in genes which are known to confer moderate to high breast and ovarian cancer risk. These variants have been detected in patients with various types of cancer during the course of diagnosis and have been evaluated as “variants of unknown significance” based on the available data. The identified variant in *TP53* is an intronic variant within the splicing region of coding exon 4. This variant is absent from the general population (gnomAD). To date, no functional assessment of this variant has been reported. Previous studies suggest that variants located at position +5 are prone to deleterious splicing, more likely as a result of sequential interactions with U1 and U6 snRNAs (Buratti et al., 2007). Our RT-PCR and cDNA sequencing results which could be confirmed via targeted RNA-seq was consistent with the expected deleterious effect on splicing based on the variant position and conservation of this nucleotide position through different species (UCSC). These results imply an activation of a cryptic splice site within the coding exon 4 resulting in deletion of

200bps from the coding exon 4 leading to a shift in the reading frame and a premature translation termination. This aberrant transcript was supported by 7.4% of the reads covering coding exon 4 an exon-intron junction. Recently the same results were reported for a synonymous variant affecting the last nucleotide of coding exon 4 (*TP53*: c.375G>A; p.Thr125Thr). RNA analysis on the patient derived and cultured lymphocyte treated with nonsense-mediated decay inhibitor agents showed partial skipping of exon 4 (r.176_375del; p. Gly59Val*fs*∗23) (Rofes et al., 2020). These results together with our results demonstrate the existence of a cryptic splice junction within exon 4 at the position c.176. This junction is activated in the absence or deactivation of the canonical splice donor.

The second variant which was successfully confirmed via RNA-sequencing to result in aberrant splicing is a synonymous substitution resulting from a G to A transition affecting the last nucleotide of coding exon 4. The result obtained by RT-PCR showed an aberrant transcript and cDNA sequencing showed that this variant results in coding exon 4 skipping. RNA-sequencing could also detect similar aberrant transcript supported by 7.7% of the sequencing reads. Skipping of the exon 4 disrupts the reading frame and leads to a premature stop codon. This in turn will likely result in either a truncated protein or nonsense-mediated decay of the mRNA transcript. This variant has recently been reported in two affected siblings with Peutz–Jeghers syndrome and similar results were obtained in the RNA analysis reported by the authors (Cerasuolo et al., 2020).

Three other variants in *BRCA1*, *BRCA2,* and *CHEK2* have been tested. While we were able to detect the aberrant transcripts using RT-PCR, targeted RNA-sequencing was not able to confirm these results. RT-PCR analysis and cDNA sequencing revealed that the missense variant c.592G>A; p.Val198Ile results in the out of frame coding exon 3 skipping, which leads to a premature stop codon (r.465_597del;p.Tyr156Hisfs*87), and thus likely leads either to a truncated protein or nonsense-mediated decay of the mRNA transcript. The *BRCA2* variant which affects the canonical splice acceptor of exon 21 was shown by RT-PCR to result in the out of frame skipping of coding exon 21, which consequently leads to a premature stop codon. We could not detect the aberrant transcript using

targeted RNA-sequencing. Similarly, the *BRCA1* missense variant c.4093T>G; p.Leu365Val which was shown by RT-PCR and cDNA sequencing to result in an activation of a cryptic splice donor, and an aberrant transcript with partial skipping of exon 9 (r.4093_4096del) could not be detected in the targeted RNA-sequencing data.

For these analyses we utilized the extracted from the whole blood using the PAXgene system, which provides a snapshot of the transcripts at the time of sampling. However, a study by Davy and colleagues found that, the PAXgene system appears not to be well adapted to RNA-Seq, especially for HBOC-associated genes because these genes are very lowly expressed, especially in whole blood (Davy et al., 2017).

In comparison to TP53 and STK11 (with 7.698 TPM and 23.66 TPM, respectively) BRCA1, BRCA2 and CHEK2 have a very low expression level within the whole blood cells (0.8826 TPM, 0.2086 TPM and 1.259 TPM) (GTEx portal accessed on 13.12.2022). This may partly explain why we were able to detect aberrant transcripts for *TP53* and *STK11*, but not *BRCA1*, *BRCA2* and *CHEK2*. A technical adjustment would be required to address this issue. This would be with a decrease in the number of sequenced samples per run and an increase in sequencing capacity. The cost of sequencing, however, will be significantly affected by these changes. Alternatively, one can use patients lymphoblastoid cell lines (LCL). It is possible, however, that Epstein-Barr virus transformation, which modifies global gene expression, might have an effect on the splicing pattern (Davy et al., 2017).

According to recent large-scale studies, gene expression and mRNA isoforms vary widely between tissues, thus suggesting that for many diseases, sequencing disease-relevant tissues will be beneficial for interpreting genetic differences (Melé et al., 2015). Here in this study, we have performed analyses on easy-to-obtain blood samples. The data from the study by Davy and colleagues suggested that the predominant alternative splicing in 11 HBOC-associated genes (including *BRCA1*, *BRCA2*, *CHEK2*, and *STK11*) is similar in blood and breast tissue (Davy et al., 2017).

According to our findings, RNA-seq can be used to interpret both coding and noncoding variants and can significantly enhance the diagnosis rate for patients for whom exome or whole-genome analysis has been inconclusive.

5.9 The Effect of Rare and Common *COMT* Variant on Breast Cancer Risk Among *BRCA1/2* Positive Women

The effect of a polymorphism (p.Val158Met) in the *COMT* gene has long been studied on the risk of breast cancer. Based on the biological function of COMT, it has been hypothesized that women carrying one or two copies of mutated low activity COMT Met are at increased risk of breast cancer due to accumulation of potentially carcinogenic estrogen metabolites. Many epidemiological studies have studied the role of *COMT* Val158Met in breast cancer risk in both postmenopausal and premenopausal women unselected for their *BRCA* status. However, the results remain controversial. Some original studies reported that the *COMT* Val158Met polymorphism was associated with increased breast cancer risk, but others reported different results (Lavigne et al., 1997; Millikan et al., 1998; Yim et al., 2001). All these studies were hypothesized based on a presumed causal role for *COMT* SNP but not as a genetic modifier of breast cancer risk. In none of these studies the subjects were tested for the high-, moderate- and low-risk factor genes involved in hereditary breast cancer. Here, for the first time, we investigated the effect of *COMT* polymorphism Val158Met and the recently reported rare synonymous variant Leu203Leu on breast cancer risk and the age at onset among *BRCA1/2* positive women. We have identified a slightly significant association between the presence of one or two copies of the mutated low-activity allele (Val158Met) and later breast cancer onset among *BRCA1/2* positive women. Similar results were reported by Ding and colleagues, who conducted a meta-analysis of the data obtained from 26 studies, consisting of 16,693 breast cancer cases and 18,261 controls. Their findings showed a slight protective role for the low activity *COMT* SNP in recessive mode among European populations (Ding et al., 2010). Although their results and their interpretation of results were criticized by Xi and colleagues (Xi et al., 2011) as they have pooled studies from populations with significantly different allele frequency for Val158Met. In addition, they also questioned the finding of a

protective role for the low activity allele of *COMT* to be in contrast with the biological role of *COMT* in inactivating genotoxic catechol metabolites of E2/E1 (Xi et al., 2011). Moreover, He and colleagues conducted a meta-analysis of 30,199 breast cancer cases and 38,922 controls and reported a similar protective effect for the mutated allele under the recessive model (He et al., 2012). However, Jungeström and Wingren studied the association between *COMT* polymorphism and early-onset breast cancer. They conducted a case-control analysis of 126 young breast cancer patients and 117 healthy controls. Looking for (low and high) allele frequency differences in cases and controls they found no significant difference in the frequency of low- and high- activity alleles (Bergman-Jungeström & Wingren, 2001).

When the patients were divided into two groups of affected and unaffected based on their breast cancer status, we did not detect any association with the presence of one or two copies of a mutated low-activity allele. Similarly, an updated meta-analysis by Qin and colleagues showed no association between low activity *COMT* polymorphism and risk of breast cancer (Qin et al., 2012).

Similar to the previous studies on the role of *COMT* polymorphism on breast cancer risk our results were conflicting. The confounding factors, which were also not addressed in our analysis, may account for the inconsistency of these results: some genetic and non-genetic determinants of the carcinogenic estrogen metabolites such as the activity of cytochrome P450 and its involvement in catechol and quinone synthesis (e.g. *CYP1B1* and BMI), as well as other *COMT* polymorphisms, which may alter *COMT* mRNA conformation and translation, and possibly exposure to exogenous catechols such as drugs and environmental chemicals which could have influence on *COMT* activity and SNP penetrance (Yager, 2012).

In an attempt to search for protective genetic signatures among *BRCA1/2*-PV carriers by exploring genome variation in women who carry deleterious variants in *BRCA1/2* and in whom early-onset breast cancer has not developed Movassagh and colleagues identified a rare synonymous variant (Leu203Leu) as a protective factor (Movassagh et al., 2017). Although our sample size is limited

for the purpose of analyzing rare genetic modifiers of the risk of breast and ovarian cancer among carriers of *BRCA1/2*, relying on our extreme phenotype sampling method that seeks to increase power of analysis, we have tested our early and late AAO breast cancer cohorts that are positive for *BRCA1/2* pathogenic variants for the presence of the previously described synonymous variant in *COMT*. We did not identify similar results as reported by Movassagh and colleagues (Movassagh et al., 2017). However, our results are in line with the results reported by the Consortium of Investigators of Modifiers of *BRCA1/2* (CIMBA) which were presented in response to Movassagh and colleagues (Georgia Chenevix-Trench et al., 2017).

All together, these conflicting results indicate that investigated genetic variants in *COMT* may not alone be significant breast cancer risk factors or protective factors. Larger studies controlled for other confounding factors are required to investigate the effect of genetic variants in *COMT* as a modifier of breast cancer risk.

5.10 Implications and Future Research

Over the past few years, several attempts have been made to understand the variable penetrance of *BRCA1* pathogenic variants. Based on GWA analyses, several loci were identified that have the potential to modify the penetrance of *BRCA1/2* pathogenic variants and the age at onset of hereditary breast and ovarian cancer to some extent. This is the first study to assess the effect of germline truncating variants in DNA-repair pathways on the onset of breast cancer among *BRCA1* carriers. Based on the odds ratio observed in this study, there may be an association between co-occurring DNA-repair truncating variants and pathogenic variants in *BRCA1* and an earlier onset of breast cancer. This study has several limitations, including a small sample size due to the low number of asymptomatic *BRCA1* mutation carriers. Larger cohorts consisting of women of different ethnicities would be beneficial. Another limiting factor was a large number of missense variants in DNA-repair genes that are of uncertain significance and some of which may in time be proven to be clinically significant. In addition, we did not evaluate copy-numbers as part of our analysis, because

we believe that copy-number detection from targeted sequencing is more likely to generate false positives than small-variant detection. Furthermore, we did not take into account variants located within intronic regions, enhancers, and promoters, which can have an effect on gene expression and regulation; however, based on the current scientific knowledge the evaluation of these variant types is challenging.

Future studies may take the same approach by extreme phenotype sampling to evaluate the complete genomic signature of *BRCA1/2* positive women with early and late breast cancer onset to identify genetic modifiers other than DNA-repair pathways.

6 Summary

Breast cancer is the most common cancer among women, accounting for 30% of all new cancer diagnoses. Hereditary genetic factors are estimated to explain 5–10% of breast cancer cases. Currently genetic testing for HBOC includes 13-17 genes, the majority of which are involved in DNA-repair pathways. Pathogenic variants in *BRCA1* and *BRCA2*, two genes which are involved in homologous recombination repair (HRR), explain about 24% of all HBOC cases. Yet, the penetrance of pathogenic variants in *BRCA1* is not full and the overall risk of developing breast cancer by the age of 80 is 72% in *BRCA1* carriers. Several factors including genetic and environmental factors have been described to modify the risk of breast cancer among *BRCA1* carriers. The environmental factors include birth cohort, age at menarche, number of pregnancies, therapeutic abortion, oral contraceptives, and prophylactic oophorectomy. The genetic factors are either the variation of type and location of variants within the *BRCA1* gene, or other modifier genes. Here we studied different genetic modifiers of *BRCA1* penetrance. In the first part of the study, we investigated double heterozygosity for *BRCA1* pathogenic variants and rare truncating variants in 313 DNA-repair genes. The selected genes have been linked to DNA-repair machinery and genome integrity maintenance which are related to HBOC pathogenesis. Applying extreme phenotype sampling, 113 women registered in the German Consortium for Hereditary Breast and/or Ovarian Cancer (GC-HBOC) with *BRCA1*-PV either with early or late breast cancer onset were recruited for this study. The patients were sequenced for 313 DNA-repair genes. Of a total of 3703 detected variants, 43 (1.2%) were rare truncating variants in the early onset cohort. As compared to 16 controls, 26 women in the early AAO group had truncating variants (35.6%; 95%-CI 24.7 - 47.7%). By controlling for the environmental factors, we compared the DNA-repair mutational load in these cohorts and observed a trend towards carrying additional truncating variants in DNA-repair genes among patients with early onset (OR: 3.1; 95%-CI 0.92 to 11.5; p-value = 0.07). However, due to the small sample size, this did not reach the conventionally acceptable significance level of 0.05. In the second part of this study, we investigated the role of a known polymorphism (p.Val158Met) and a

recently described rare synonymous variant (p.Leu203Leu) in the *COMT* gene, on the breast cancer onset among *BRCA1*- and *BRCA2*-PV carriers. A significant association between the presence of one or two copies of mutated low-activity allele (p.Val158Met) and later breast cancer onset was detected among *BRCA1/2* positive women (p-value 0.02 and 0.03 respectively). In consistent with the results presented by the Consortium of Investigators of Modifiers of *BRCA1/2* (CIMBA), we did not identify any protective role for the rare synonymous variant in (p.Leu203Leu) in the *COMT* gene against the breast cancer onset among *BRCA1/2* positive women. Our study suggests that additional truncating variants in DNA-repair pathways could affect the penetrance of *BRCA1* pathogenic variants. Larger studies are needed to confirm this effect. And finally, the tested variants in COMT, another candidate modifier of *BRCA1*-penetrance, were not related to breast cancer risk in this study. An increasingly precise penetrance estimate will help to avoid over- and undertreatment of *BRCA1* carriers for breast cancer.

7 Zusammenfassung

Brustkrebs ist die häufigste Krebsart bei Frauen und macht 30 % aller neuen Krebsdiagnosen aus. Es wird geschätzt, dass 5-10 % der Brustkrebsfälle auf erbliche genetische Faktoren zurückzuführen sind. Etwa 25 von insgesamt mehr als 300 Genen, die mit der DNA-Reparaturmaschinerie und der Aufrechterhaltung der Genom-Integrität in Verbindung gebracht werden, werden mit HBOC in Verbindung gebracht. Pathogene Varianten in *BRCA1* und *BRCA2*, zwei Gene, die an der homologen Rekombination Reparatur (HRR) beteiligt sind, erklären etwa 24 % aller HBOC-Fälle. Die Penetranz der pathogenen Varianten in *BRCA1* ist jedoch nicht vollständig und das Gesamtrisiko, bis zum Alter von 80 Jahren an Brustkrebs zu erkranken, beträgt bei *BRCA1*-PV-Trägerinnen 72 %. Es wurde beschrieben, dass mehrere Faktoren, darunter genetische Einflüsse und Umweltfaktoren, das Brustkrebsrisiko von *BRCA1*-PV-Trägerinnen beeinflussen. Zu den Umweltfaktoren gehören die Geburtskohorte, das Alter bei der Menarche, die Anzahl der Schwangerschaften, Schwangerschaftsabbrüche, orale Kontrazeptiva und prophylaktische Oophorektomie. Bei den genetischen Faktoren handelt es sich entweder um die Variation der Art und des Ortes der Varianten innerhalb des *BRCA1*-Gens oder um andere Modifikatorgene. In dieser Studie untersuchten wir die doppelte Heterozygotie für pathogene *BRCA1*-Varianten und seltene trunkierende Varianten in 313 DNA-Reparaturgenen. Hierfür wurden im Rahmen einer extremen Phänotyp-Selektion 113 Frauen mit *BRCA1*-PV mit frühem oder spätem Auftreten von Brustkrebs für diese Studie ausgewählt. Die Patientinnen wurden mit einem Gen-Panel mittes NGS sequenziert. Von den insgesamt 3703 detektierten Varianten waren 43 (1,2 %) seltene trunkierende Varianten. 26 Frauen in der Gruppe mit frühem AAO wiesen Im Vergleich zu 16 Kontrollpersonen trunkierende Varianten auf (35,6%; 95%-CI 24,7 - 47,7%). Unter Berücksichtigung der Umweltfaktoren haben wir die Variantenlast in Genen der DNA-Reparatur in diesen Kohorten verglichen. Dabei wurde ein Trend zu einer früheren Erkrankung an Brustkrebs bei zusätzlich vorhandenen trunkierenden Varianten in DNA-Reparatur-Genen nachgewiesen (OR: 3,1; 95%-CI 0,92 bis 11,5; p-Wert = 0,07).

Im zweiten Teil dieser Studie untersuchten wir die Rolle eines bekannten Polymorphismus (p.Val158Met) und einer kürzlich beschriebenen seltenen synonymen Variante (p.Leu203Leu) im *COMT*-Gen für das Auftreten von Brustkrebs bei *BRCA1*- und *BRCA2*-PV-Trägern. Ein leicht signifikanter Zusammenhang zwischen dem Vorhandensein von einer oder zwei Kopien des mutierten Allels mit geringer Aktivität (p.Val158Met) und dem späteren Auftreten von Brustkrebs wurde bei *BRCA1/2*-positiven Frauen festgestellt. In Übereinstimmung mit den vom Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA) vorgelegten Ergebnissen konnten wir keine schützende Rolle für die seltene synonyme Variante (p.Leu203Leu) im *COMT*-Gen vor dem Auftreten von Brustkrebs bei BRCA1/2-positiven Frauen feststellen.

Unsere Studie zeigt, dass zusätzliche trunkierende Varianten in DNA-Reparatursignalwegen die Penetranz von pathogenen *BRCA1*-Varianten beeinflussen können. Größere Studien sind erforderlich, um diesen Effekt zu bestätigen. Die getesteten Varianten in *COMT*, einem weiteren Kandidaten für die Modifizierung der *BRCA1*-Penetranz, standen in dieser Studie nicht mit dem Brustkrebsrisiko in Verbindung. Eine immer präzisere Abschätzung des Erkrankungsrisikos ermöglicht ein individualisiertes Vorgehen bei *BRCA1*-Anlageträgerinnen und trägt dazu bei, eine Über- oder Unterbehandlung zu vermeiden.

.

8 Bibliography

Alblihy, A., Shoqafi, A., Toss, M. S., Algethami, M., Harris, A. E., Jeyapalan, J. N., Abdel-Fatah, T., Servante, J., Chan, S. Y. T., Green, A., Mongan, N. P., Rakha, E. A., & Madhusudan, S. (n.d.). *ARTICLE Untangling the clinicopathological significance of MRE11-RAD50-NBS1 complex in sporadic breast cancers*. https://doi.org/10.1038/s41523-021-00350-5

Alsop, K., Fereday, S., Meldrum, C., DeFazio, A., Emmanuel, C., George, J., Dobrovic, A., Birrer, M. J., Webb, P. M., Stewart, C., Friedlander, M., Fox, S., Bowtell, D., & Mitchell, G. (2012). BRCA mutation frequency and patterns of treatment response in BRCA mutation-positive women with ovarian cancer: a report from the Australian Ovarian Cancer Study Group. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, *30*(21), 2654–2663. https://doi.org/10.1200/JCO.2011.39.8545

Andrieu, N., Easton, D. F., Chang-Claude, J., Rookus, M. A., Brohet, R., Cardis, E., Antoniou, A. C., Wagner, T., Simard, J., Evans, G., Peock, S., Fricker, J.-P., Nogues, C., Veer, L. V., van Leeuwen, F. E., & Goldgar, D. E. (2006). Effect of Chest X-Rays on the Risk of Breast Cancer Among *BRCA1/2* Mutation Carriers in the International *BRCA1/2* Carrier Cohort Study: A Report from the EMBRACE, GENEPSO, GEO-HEBON, and IBCCS Collaborators' Group. *Journal of Clinical Oncology*, *24*(21), 3361–3366. https://doi.org/10.1200/JCO.2005.03.3126

Antoniou, A., Pharoah, P. D. P., Narod, S., Risch, H. A., Eyfjord, J. E., Hopper, J. L., Loman, N., Olsson, H., Johannsson, O., Borg, Å., Pasini, B., Radice, P., Manoukian, S., Eccles, D. M., Tang, N., Olah, E., Anton-Culver, H., Warner, E., Lubinski, J., … Easton, D. F. (2003). Average Risks of Breast and Ovarian Cancer Associated with BRCA1 or BRCA2 Mutations Detected in Case Series Unselected for Family History: A Combined Analysis of 22 Studies. *The American Journal of Human Genetics*, *72*(5), 1117–1130.

Aranda, A., & Pascual, A. (2001). Nuclear hormone receptors and gene expression. *Physiological Reviews*, *81*(3), 1269–1304. https://doi.org/10.1152/PHYSREV.2001.81.3.1269/ASSET/IMAGES/LARGE/9J0310150010.JPEG

Arpino, G., Bardou, V. J., Clark, G. M., & Elledge, R. M. (2004). Infiltrating lobular carcinoma of the breast: tumor characteristics and clinical outcome. *Breast Cancer Research*, *6*(3), R149. https://doi.org/10.1186/bcr767

Axelrod, J., & Tomchick, R. (1958). *Enzymatic 0-Methylation of Epinephrine and Other Catechols**. http://www.jbc.org/

Baumann, P., & West, S. C. (1998). Role of the human RAD51 protein in homologous recombination and double-stranded-break repair. *Trends in Biochemical Sciences*, *23*(7), 247–251. http://www.ncbi.nlm.nih.gov/pubmed/9697414

Begg, C. B. (2002). On the Use of Familial Aggregation in Population-Based Case Probands for Calculating Penetrance. *CancerSpectrum Knowledge Environment*, *94*(16), 1221–1226. https://doi.org/10.1093/jnci/94.16.1221

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. http://www.ncbi.nlm.nih.gov/pubmed/12927427

Bergman-Jungeström, M., & Wingren, S. (2001). Catechol-O-Methyltransferase (COMT) gene polymorphism and breast cancer risk in young women. *British Journal of Cancer*. https://doi.org/10.1054/bjoc.2001.2009

Berliner, J. L., & Fay, A. M. (2007). Risk Assessment and Genetic Counseling for Hereditary Breast and Ovarian Cancer: Recommendations of the National Society of Genetic Counselors. *Journal of Genetic Counseling*, *16*(3), 241–260. https://doi.org/10.1007/s10897-007-9090-7

Bjørnland, T., Bye, A., Ryeng, E., Wisløff, U., & Langaas, M. (2018). Powerful extreme phenotype sampling designs and score tests for genetic association studies. *Statistics in Medicine*, *37*(28), 4234–4251. https://doi.org/10.1002/SIM.7914

Blasiak, J. (2021). Single-Strand Annealing in Cancer. *International Journal of Molecular Sciences*, *22*(4), 1–24. https://doi.org/10.3390/IJMS22042167

Boice, J. D., Preston, D., Davis, F. G., & Monson, R. R. (1991). Frequent chest X-ray fluoroscopy and breast cancer incidence among tuberculosis patients in Massachusetts. *Radiation Research*, *125*(2), 214–222. http://www.ncbi.nlm.nih.gov/pubmed/1996380

Borbála Székely, Lilla Madaras, G. S. A. M. S. Z. B. L. S. L. T. E. Z. E. B. I. K. A. K. Z. L., & Ferenc Bánhidy, J. K. M. D. (2010). *Comparison of breast cancer in young and old women based on clinicopathological features*.

Boyd, N. F., Dite, G. S., Stone, J., Gunasekara, A., English, D. R., McCredie, M. R. E., Giles, G. G., Tritchler, D., Chiarelli, A., Yaffe, M. J., & Hopper, J. L. (2002). Heritability of Mammographic Density, a Risk Factor for Breast Cancer. *New England Journal of Medicine*, *347*(12), 886–894. https://doi.org/10.1056/NEJMoa013390

Boyd, N. F., Lockwood, G. A., Byng, J. W., Tritchler, D. L., & Yaffe, M. J. (1998). Mammographic densities and breast cancer risk. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *7*(12), 1133–1144. http://www.ncbi.nlm.nih.gov/pubmed/9865433

Boyd, N. F., O'Sullivan, B., Campbell, J. E., Fishell, E., Simor, I., Cooke, G., & Germanson, T. (1982). Mammographic signs as risk factors for breast cancer. *British Journal of Cancer*, *45*(2), 185–193. https://doi.org/10.1038/bjc.1982.32

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *68*(6), 394–424. https://doi.org/10.3322/caac.21492

Bray, F., McCarron, P., & Parkin, D. M. (2004). The changing global patterns of female breast cancer incidence and mortality. *Breast Cancer Research*, *6*(6), 229. https://doi.org/10.1186/bcr932

Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50 302 women with breast cancer and 96 973 women without the disease. (2002). *The Lancet*, *360*(9328), 187–195. https://doi.org/10.1016/S0140-6736(02)09454-0

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

Brohl, A. S., Patidar, R., Turner, C. E., Wen, X., Song, Y. K., Wei, J. S., Calzone, K. A., & Khan, J. (2017). Frequent inactivating germline mutations in DNA repair genes in patients with Ewing sarcoma Germline mutations in Ewing sarcoma. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, *19*(8), 955–958.

Buratti, E., Chivers, M., Královičová, J., Romano, M., Baralle, M., Krainer, A. R., & Vořechovský, I. (2007). Aberrant 5' splice sites in human disease genes: mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. *Nucleic Acids Research*, *35*(13), 4250–4263. https://doi.org/10.1093/NAR/GKM402

Burgess, J. T., Rose, M., Boucher, D., Plowman, J., Molloy, C., Fisher, M., O'Leary, C., Richard, D. J., O'Byrne, K. J., & Bolderson, E. (2020). The Therapeutic Potential of DNA Damage Repair Pathways and Genomic Stability in Lung Cancer. *Frontiers in Oncology*, *10*. https://doi.org/10.3389/FONC.2020.01256

Byrne, C., Schairer, C., Wolfe, J., Parekh, N., Salane, M., Brinton, L. A., Hoover, R., & Haile, R. (1995). Mammographic Features and Breast Cancer Risk: Effects With Time, Age, and Menopause Status. *JNCI Journal of the National Cancer Institute*, *87*(21), 1622–1629. https://doi.org/10.1093/jnci/87.21.1622

Carroll, J. C., Cremin, C., Allanson, J., Blaine, S. M., Dorman, H., Gibbons, C. A., Grimshaw, J., Honeywell, C., Meschino, W. S., Permaul, J., & Wilson, B. J. (2008). Hereditary breast and ovarian cancers. *Canadian Family Physician*, *54*(12), 1691. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2602616/

Cartegni, L., Chew, S. L., & Krainer, A. R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet*, *3*(4), 285–298.

Cavalieri, E. L., Stack, D. E., Devanesan, P. D., Todorovic, R., Dwivedy, I., Higginbotham, S., Johansson, S. L., Patil, K. D., Gross, M. L., Gooden, J. K., Ramanathan, R., Cerny, R. L., & Rogan, E. G. (1997). Molecular origin of cancer: catechol estrogen-3,4-quinones as endogenous tumor initiators. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(20), 10937–10942. https://doi.org/10.1073/PNAS.94.20.10937

Ceccaldi, R., Rondinelli, B., & D'Andrea, A. D. (2016). Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends in Cell Biology*, *26*(1), 52–64. https://doi.org/10.1016/J.TCB.2015.07.009

Cerasuolo, A., Cammarota, F., Duraturo, F., Staiano, A., Martinelli, M., Miele, E., Izzo, P., & de Rosa, M. (2020). Implications of splicing alterations in the onset and phenotypic variability of a family with subclinical manifestation of Peutz–

Jeghers syndrome: Bioinformatic and molecular evidence. *International Journal of Molecular Sciences*, *21*(21), 1–12. https://doi.org/10.3390/ijms21218201

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

Chlebowski, R. T., Manson, J. E., Anderson, G. L., Cauley, J. A., Aragaki, A. K., Stefanick, M. L., Lane, D. S., Johnson, K. C., Wactawski-Wende, J., Chen, C., Qi, L., Yasmeen, S., Newcomb, P. A., & Prentice, R. L. (2013). Estrogen Plus Progestin and Breast Cancer Incidence and Mortality in the Women's Health Initiative Observational Study. *JNCI: Journal of the National Cancer Institute*, *105*(8), 526–535. https://doi.org/10.1093/jnci/djt043

Cho, H. S., Mason, K., Ramyar, K. X., Stanley, A. M., Gabelli, S. B., Denney, D. W., & Leahy, D. J. (2003). Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. *Nature*, *421*(6924), 756–760. https://doi.org/10.1038/NATURE01392

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*, *6*(2), 80–92.

Ciriello, G., Gatza, M. L., Beck, A. H., Wilkerson, M. D., Rhie, S. K., Pastore, A., Zhang, H., McLellan, M., Yau, C., Kandoth, C., Bowlby, R., Shen, H., Hayat, S., Fieldhouse, R., Lester, S. C., Tse, G. M. K., Factor, R. E., Collins, L. C., Allison, K. H., … Zmuda, E. (2015). Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell*, *163*(2), 506–519. https://doi.org/10.1016/j.cell.2015.09.033

Clemons, M., Loijens, L., & Goss, P. (2000). Breast cancer risk following irradiation for Hodgkin's disease. *Cancer Treatment Reviews*, *26*(4), 291–302. https://doi.org/10.1053/ctrv.2000.0174

Coignard, J., Lush, M., Beesley, J., O'Mara, T. A., Dennis, J., Tyrer, J. P., Barnes, D. R., McGuffog, L., Leslie, G., Bolla, M. K., Adank, M. A., Agata, S., Ahearn, T., Aittomäki, K., Andrulis, I. L., Anton-Culver, H., Arndt, V., Arnold, N., Aronson, K. J., … Antoniou, A. C. (2021). A case-only study to identify genetic modifiers of breast cancer risk for BRCA1/BRCA2 mutation carriers. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-020-20496-3

Colditz, G. A., Sellers, T. A., & Trapido, E. (2006). Epidemiology — identifying the causes and preventability of cancer? *Nature Reviews Cancer*, *6*(1), 75–83. https://doi.org/10.1038/nrc1784

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. *The Lancet*, *358*(9291), 1389–1399. https://doi.org/10.1016/S0140-6736(01)06524-2

Collaborative Group on Hormonal Factors in Breast Cancer. (2002). Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50 302 women with breast cancer and 96 973 women without the disease. *The Lancet*, *360*(9328), 187–195. https://doi.org/10.1016/S0140-6736(02)09454-0

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

Colton, T., Greenberg, E. R., Noller, K., Resseguie, L., van Bennekom, C., Heeren, T., & Zhang, Y. (1993). Breast cancer in mothers prescribed diethylstilbestrol in pregnancy. Further follow-up. *JAMA*, *269*(16), 2096–2100. http://www.ncbi.nlm.nih.gov/pubmed/8468763

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

Cooper, G. M., Stone, E. A., Asimenos, G., NISC Comparative Sequencing Program, Green, E. D., Batzoglou, S., & Sidow, A. (2005). Distribution and intensity of constraint in mammalian genomic sequence. *Genome Research*, *15*(7), 901–913. https://doi.org/10.1101/gr.3577405

Coradini, D., Oriana, S., Mariani, L., Miceli, R., Bresciani, G., Marubini, E., & di Fronzo, G. (1998). Is steroid receptor profile in contralateral breast cancer a marker of independence of the corresponding primary tumour? *European Journal of Cancer (Oxford, England : 1990)*, *34*(6), 825–830. http://www.ncbi.nlm.nih.gov/pubmed/9797693

Dagan, E., Friedman, E., Paperna, T., Carmi, N., & Gershoni-Baruch, R. (2002). Androgen receptor CAG repeat length in Jewish Israeli women who are BRCA1/2 mutation carriers: association with breast/ovarian cancer phenotype. *European Journal of Human Genetics*, *10*(11), 724–728. https://doi.org/10.1038/sj.ejhg.5200880

Davis, A. J., & Chen, D. J. (2013). DNA double strand break repair via non-homologous end-joining. *Translational Cancer Research*, *2*(3), 130–143. https://doi.org/10.3978/J.ISSN.2218-676X.2013.04.02

Davy, G., Rousselin, A., Goardon, N., Castéra, L., Harter, V., Legros, A., Muller, E., Fouillet, R., Brault, B., Smirnova, A. S., Lemoine, F., De La Grange, P., Guillaud-Bataille, M., Caux-Moncoutier, V., Houdayer, C., Bonnet, F., Blanc-Fournier, C., Gaildrat, P., Frebourg, T., … Krieger, S. (2017). Detecting splicing patterns in genes involved in hereditary breast and ovarian cancer. *European Journal of Human Genetics*, *25*(10), 1147–1154. https://doi.org/10.1038/ejhg.2017.116

Densham, R. M., & Morris, J. R. (2017). The BRCA1 Ubiquitin ligase function sets a new trend for remodelling in DNA repair. *Nucleus (Austin, Tex.)*, *8*(2), 116–125. https://doi.org/10.1080/19491034.2016.1267092

Desmedt, C., Zoppoli, G., Gundem, G., Pruneri, G., Larsimont, D., Fornili, M., Fumagalli, D., Brown, D., Rothé, F., Vincent, D., Kheddoumi, N., Rouas, G.,

Majjaj, S., Brohée, S., van Loo, P., Maisonneuve, P., Salgado, R., van Brussel, T., Lambrechts, D., … Sotiriou, C. (2016). Genomic Characterization of Primary Invasive Lobular Breast Cancer. *Journal of Clinical Oncology*, *34*(16), 1872–1881. https://doi.org/10.1200/JCO.2015.64.0334

Dhingra, N., Wei, L., & Zhao, X. (2019). Replication protein A (RPA) sumoylation positively influences the DNA damage checkpoint response in yeast. *The Journal of Biological Chemistry*, *294*(8), 2690–2699. https://doi.org/10.1074/JBC.RA118.006006

Ding, H., Fu, Y., Chen, W., & Wang, Z. (2010). COMT Val158Met polymorphism and breast cancer risk: Evidence from 26 case-control studies. *Breast Cancer Research and Treatment*, *123*(1), 265–270. https://doi.org/10.1007/s10549-010-0759-5

Dolled-Filhart, M. P., Lee, M., Ou-Yang, C. W., Haraksingh, R. R., & Lin, J. C. H. (2013). Computational and bioinformatics frameworks for next-generation whole exome and genome sequencing. *TheScientificWorldJournal*, *2013*. https://doi.org/10.1155/2013/730210

Du, T., Zhu, L., Levine, K. M., Tasdemir, N., Lee, A. v., Vignali, D. A. A., Houten, B. van, Tseng, G. C., & Oesterreich, S. (2018). Invasive lobular and ductal breast carcinoma differ in immune response, protein translation efficiency and metabolism. *Scientific Reports*, *8*(1), 7205. https://doi.org/10.1038/s41598-018-25357-0

Eerola, H., Heikkilä, P., Tamminen, A., Aittomäki, K., Blomqvist, C., & Nevanlinna, H. (2005). Relationship of patients' age to histopathological features of breast tumours in BRCA1 and BRCA2and mutation-negative breast cancer families. *Breast Cancer Research*, *7*(4), R465.

Faust, G. G., & Hall, I. M. (2014). SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics (Oxford, England)*, *30*(17), 2503–2505. https://doi.org/10.1093/bioinformatics/btu314

Foulkes, W. D., Chappuis, P. O., Wong, N., Brunet, J.-S., Vesprini, D., Rozen, F., Yuan, Z. Q., Pollak, M. N., Kuperstein, G., Narod, S. A., & Begin, L. R. (2000). Primary node negative breast cancer in BRCA1 mutation carriers has a poor outcome. *Annals of Oncology*, *11*, 307–314. https://doi.org/10.1023/A:1008340723974

Frankenberg-Schwager, M. (1990). Induction, repair and biological relevance of radiation-induced DNA lesions in eukaryotic cells. *Radiation and Environmental Biophysics*, *29*(4), 273–292. https://doi.org/10.1007/BF01210408

Friebel, T. M., Domchek, S. M., & Rebbeck, T. R. (2014). Modifiers of cancer risk in BRCA1 and BRCA2 mutation carriers: systematic review and meta-analysis. *Journal of the National Cancer Institute*, *106*(6), dju091. https://doi.org/10.1093/jnci/dju091

Frykberg, E. R. (1999). Lobular Carcinoma In Situ of the Breast. *The Breast Journal*, *5*(5), 296–303. http://www.ncbi.nlm.nih.gov/pubmed/11348305

Frykberg, E. R., Santiago, F., Betsill, W. L., & O'Brien, P. H. (1987). Lobular carcinoma in situ of the breast. *Surgery, Gynecology & Obstetrics*, *164*(3), 285–301. http://www.ncbi.nlm.nih.gov/pubmed/3547723

Futreal, P. A., Liu, Q., Shattuck-Eidens, D., Cochran, C., Harshman, K., Tavtigian, S., Bennett, L. M., Haugen-Strano, A., Swensen, J., Miki, Y., Eddington, K., McClure, M., Frye, C., Weaver-Feldhaus, J., Ding, W., Gholami, Z., Soderkvist, P., Terry, L., Jhanwar, S., … Wiseman, R. (1994). BRCA1 Mutations in Primary Breast and Ovarian Carcinomas. *Science*, *266*(5182), 120–122. https://doi.org/10.1126/SCIENCE.7939630

Gachechiladze, M., Skarda, J., Bouchalova, K., Soltermann, A., & Joerger, M. (2020). Predictive and Prognostic Value of DNA Damage Response Associated Kinases in Solid Tumors. *Frontiers in Oncology*, *10*. https://doi.org/10.3389/FONC.2020.581217

Garrison, E., & Marth, G. (2012). *Haplotype-based variant detection from short-read sequencing*.

Gayther, S. A., Mangion, J., Russell, P., Seal, S., Barfoot, R., Ponder, B. A. J., Stratton, M. R., & Easton, D. (1997). Variation of risks of breast and ovarian cancer associated with different germline mutations of the BRCA2 gene. *Nature Genetics*, *15*(1), 103–105. https://doi.org/10.1038/ng0197-103

Gayther, S. A., Warren, W., Mazoyer, S., Russell, P. A., Harrington, P. A., Chiano, M., Seal, S., Hamoudi, R., van Rensburg, E. J., Dunning, A. M., Love, R., Evans, G., Easton, D., Clayton, D., Stratton, M. R., & Ponder, B. A. J. (1995). Germline mutations of the BRCA1 gene in breast and ovarian cancer families provide evidence for a genotype–phenotype correlation. *Nature Genetics*, *11*(4), 428–433. https://doi.org/10.1038/ng1295-428

Georgia Chenevix-Trench, Daniel R Barnes, & Antonis C Antoniou. (2017). More on Co-Occurrence of COMT and BRCA1/2 Variants in a Population. *New England Journal of Medicine*, *377*(8), 793–796. https://doi.org/10.1056/NEJMc1708425

GERMAN CENTRE FOR CANCER REGISTRY DATA. (n.d.). https://doi.org/10.17886/rkipubl-2016-015

Giguère, Y., Dewailly, E., Brisson, J., Ayotte, P., Laflamme, N., Demers, a, Forest, V. I., Dodin, S., Robert, J., & Rousseau, F. (2001). Short polyglutamine tracts in the androgen receptor are protective against breast cancer in the general population. *Cancer Research*, *61*(15), 5869–5874. http://www.ncbi.nlm.nih.gov/pubmed/11479228

Glück, S. (2017). Consequences of the Convergence of Multiple Alternate Pathways on the Estrogen Receptor in the Treatment of Metastatic Breast Cancer. *Clinical Breast Cancer*, *17*(2), 79–90. https://doi.org/10.1016/J.CLBC.2016.08.004

Goldgar, D. E., Fields, P., Lewis, C. M., Tran, T. D., Cannon-Albright, L. A., Ward, J. H., Swensen, J., & Skolnick, M. H. (1994). A Large Kindred With 17q-Linked Breast and Ovarian Cancer: Genetic, Phenotypic, and Genealogical Analysis. *JNCI Journal of the National Cancer Institute*, *86*(3), 200–209. https://doi.org/10.1093/jnci/86.3.200

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics 2016 17:6*, *17*(6), 333–351. https://doi.org/10.1038/nrg.2016.49

Gostissa, M., Alt, F. W., & Chiarle, R. (2011). Mechanisms that promote and suppress chromosomal translocations in lymphocytes. *Annual Review of Immunology*, *29*, 319–350. https://doi.org/10.1146/ANNUREV-IMMUNOL-031210-101329

Guidugli, L., Rugani, C., Lombardi, G., Aretini, P., Galli, A., & Caligo, M. A. (2011). A recombination-based method to characterize human BRCA1 missense variants. *Breast Cancer Research and Treatment*, *125*(1), 265–272. https://doi.org/10.1007/S10549-010-1112-8/TABLES/3

Gupta, G. P., Ho, A. Y., Feng, W., Fan, C., Akram, M., Brogi, E., Powell, S. N., Perou, C. M., Wen, Y. H., & Petrini, J. H. J. (2017). Mre11 Dysfunction Is Associated With Triple-Negative Breast Cancer and Confers Sensitivity to DNA Damaging Therapy. *International Journal of Radiation Oncology • Biology • Physics*, *96*(2), S43.

Gutierrez, C., & Schiff, R. (2011). HER2: biology, detection, and clinical implications. *Archives of Pathology & Laboratory Medicine*, *135*(1), 55–62. https://doi.org/10.5858/2010-0454-RAR.1

Harris, L., Fritsche, H., Mennel, R., Norton, L., Ravdin, P., Taube, S., Somerfield, M. R., Hayes, D. F., Bast, R. C., & American Society of Clinical Oncology. (2007). American Society of Clinical Oncology 2007 Update of Recommendations for the Use of Tumor Markers in Breast Cancer. *Journal of Clinical Oncology*, *25*(33), 5287–5312. https://doi.org/10.1200/JCO.2007.14.2364

Hayashi, Y., Yamamoto, M., Ohmori, S., Kikumori, T., Imai, T., Funahashi, H., & Seo, H. (1999). Polymorphism of homopolymeric glutamines in coactivators for nuclear hormone receptors. *Endocrine Journal*, *46*(2), 279–284. http://www.ncbi.nlm.nih.gov/pubmed/10460012

He, X.-F., Wei, W., Li, S.-X., Su, J., Zhang, Y., Ye, X.-H., Liu, Y., & Wang, W. (2012). Association between the COMT Val158Met polymorphism and breast cancer risk: a meta-analysis of 30,199 cases and 38,922 controls. *Molecular Biology Reports*, *39*(6), 6811–6823. https://doi.org/10.1007/s11033-012-1506-2

Head, S. R., Kiyomi Komori, H., LaMere, S. A., Whisenant, T., van Nieuwerburgh, F., Salomon, D. R., & Ordoukhanian, P. (2014). Library construction for next-generation sequencing: Overviews and challenges. *BioTechniques*, *56*(2), 61–77. https://doi.org/10.2144/000114133/ASSET/IMAGES/LARGE/FIGURE7.JPEG

Heemskerk-Gerritsen, B. A., Seynaeve, C., van Asperen, C. J., Ausems, M. G., Collee, J. M., van Doorn, H. C., Gomez Garcia, E. B., Kets, C. M., van Leeuwen, F. E., Meijers-Heijboer, H. E., Mourits, M. J., van Os, T. A., Vasen, H. F., Verhoef, S., Rookus, M. A., & Hooning, M. J. (2015). Breast cancer risk after salpingo-oophorectomy in healthy BRCA1/2 mutation carriers: revisiting the evidence for risk reduction. *J Natl Cancer Inst*, *107*(5).

Herschkowitz, J. I., Simin, K., Weigman, V. J., Mikaelian, I., Usary, J., Hu, Z., Rasmussen, K. E., Jones, L. P., Assefnia, S., Chandrasekharan, S., Backlund, M. G., Yin, Y., Khramtsov, A. I., Bastein, R., Quackenbush, J., Glazer, R. I., Brown, P. H., Green, J. E., Kopelovich, L., … Perou, C. M. (2007). Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biology*, *8*(5), R76. https://doi.org/10.1186/gb-2007-8-5-r76

Hu, T., Chitnis, N., Monos, D., & Dinh, A. (2021). Next-generation sequencing technologies: An overview. *Human Immunology*, *82*(11), 801–811. https://doi.org/10.1016/j.humimm.2021.02.012

Huang, C. S., Chern, H. D., Chang, K. J., Cheng, C. W., Hsu, S. M., & Shen, C. Y. (1999). Breast cancer risk associated with genotype polymorphism of the estrogen-metabolizing genes CYP17, CYP1A1, and COMT: a multigenic study on cancer susceptibility. *Cancer Research*, *59*(19), 4870–4875. http://www.ncbi.nlm.nih.gov/pubmed/10519398

Huang, R., & Zhou, P. K. (2021). DNA damage repair: historical perspectives, mechanistic pathways and clinical translation for targeted cancer therapy. *Signal Transduction and Targeted Therapy 2021 6:1*, *6*(1), 1–35. https://doi.org/10.1038/s41392-021-00648-7

Huebner, R. J., & Ewald, A. J. (2014). Cellular foundations of mammary tubulogenesis. *Seminars in Cell & Developmental Biology*, *31*, 124–131. https://doi.org/10.1016/J.SEMCDB.2014.04.019

Hunter, T. (2000). Signaling—2000 and Beyond. *Cell*, *100*(1), 113–127. https://doi.org/10.1016/S0092-8674(00)81688-8

Hunter, T. (2007). The Age of Crosstalk: Phosphorylation, Ubiquitination, and Beyond. *Molecular Cell*, *28*(5), 730–738. https://doi.org/10.1016/J.MOLCEL.2007.11.019

Innos, K., & Horn-Ross, P. L. (2008). Risk of second primary breast cancers among women with ductal carcinoma in situ of the breast. *Breast Cancer Research and Treatment*, *111*(3), 531–540. https://doi.org/10.1007/s10549-007-9807-1

Jemal, A., Center, M. M., Desantis, C., & Ward, E. M. (2010). Global Patterns of Cancer Incidence and Mortality Rates and Trends. *Cancer Epidemiol Biomarkers Prev*, *19*(8), 1893–1907. https://doi.org/10.1158/1055-9965.EPI-10-0437

Judkins, T., Rosenthal, E., Arnell, C., Burbidge, L. A., Geary, W., Barrus, T., Schoenberger, J., Trost, J., Wenstrup, R. J., & Roa, B. B. (2012). Clinical significance of large rearrangements in BRCA1 and BRCA2. *Cancer*, *118*(21), 5210–5216. https://doi.org/10.1002/CNCR.27556

Kajitani, G. S., Nascimento, L. L. de S., Neves, M. R. de C., Leandro, G. da S., Garcia, C. C. M., & Menck, C. F. M. (2021). Transcription blockage by DNA damage in nucleotide excision repair-related neurological dysfunctions. *Seminars in Cell & Developmental Biology*, *114*, 20–35. https://doi.org/10.1016/J.SEMCDB.2020.10.009

Kallioniemi, O. P., Kallioniemi, A., Kurisu, W., Thor, A., Chen, L. C., Smith, H. S., Waldman, F. M., Pinkel, D., & Gray, J. W. (1992). ERBB2 amplification in breast cancer analyzed by fluorescence in situ hybridization. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(12), 5321–5325. https://doi.org/10.1073/PNAS.89.12.5321

Kass, E. M., & Jasin, M. (2010). Collaboration and competition between DNA double-strand break repair pathways. *FEBS Letters*, *584*(17), 3703–3708. https://doi.org/10.1016/J.FEBSLET.2010.07.057

Kato, I., Beinart, C., Bleich, A., Su, S., Kim, M., & Toniolo, P. G. (1995). A nested case-control study of mammographic patterns, breast volume, and breast cancer (New York City, NY, United States). *Cancer Causes and Control*, *6*(5), 431–438. https://doi.org/10.1007/BF00052183

Kazemi-Esfarjani, P., Trifiro, M. A., & Pinsky, L. (1995). Evidence for a repressive function of the long polyglutamine tract in the human androgen receptor: possible pathogenetic relevance for the (CAG) $_n$ -expanded neuronopathies. *Human Molecular Genetics*, *4*(4), 523–527. https://doi.org/10.1093/hmg/4.4.523

Kedmi, A., Kadouri, L., Sagy, I., Hamburger, T., Levin, G., Zimhony-Nissim, N., & Peretz, T. (2022). Genetic anticipation of breast cancer among BRCA1/BRCA2 mutation carriers: A retrospective study. *International Journal of Gynecology and Obstetrics*. https://doi.org/10.1002/IJGO.14179

Kenny, L., & Bickerstaff, H. (2017). *Gynaecology by Ten Teachers*. CRC Press/Taylor & Francis Group. https://books.google.de/books?id=Vx4SMQAACAAJ

Key, T. J., Appleby, P. N., Reeves, G. K., Roddam, A., Dorgan, J. F., Longcope, C., Stanczyk, F. Z., Stephenson, H. E., Falk, R. T., Miller, R., Schatzkin, A., Allen, D. S., Fentiman, I. S., Key, T. J., Wang, D. Y., Dowsett, M., Thomas, H. v, Hankinson, S. E., Toniolo, P., … Endogenous Hormones Breast Cancer Collaborative Group. (2003). Body mass index, serum sex hormones, and breast cancer risk in postmenopausal women. *Journal of the National Cancer Institute*, *95*(16), 1218–1226. http://www.ncbi.nlm.nih.gov/pubmed/12928347

Kirsi Määttä. (2016). *Genetic Predisposition to Breast and Ovarian Cancer BRCA1/2-negative families. Academic Dissertation, University of Tampere*.

Kitagishi, Y., Kobayashi, M., & Matsuda, S. (2013). Defective DNA repair systems and the development of breast and prostate cancer (review). *International Journal of Oncology*, *42*(1), 29–34. https://doi.org/10.3892/IJO.2012.1696

Knudson, A. G. (1971). *Mutation and Cancer: Statistical Study of Retinoblastoma* (Vol. 68, Issue 4). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC389051/pdf/pnas00079-0129.pdf

Koressaar, T., & Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics*, *23*(10), 1289–1291. https://doi.org/10.1093/bioinformatics/btm091

Kotsopoulos, J., Lubinski, J., Lynch, H. T., Neuhausen, S. L., Ghadirian, P., Isaacs, C., Weber, B., Kim-Sing, C., Foulkes, W. D., Gershoni-Baruch, R., Ainsworth, P., Friedman, E., Daly, M., Garber, J. E., Karlan, B., Olopade, O. I., Tung, N., Saal, H. M., Eisen, A., … Narod, S. A. (2005). Age at menarche and the risk of breast cancer in BRCA1 and BRCA2 mutation carriers. *Cancer Causes {&} Control*, *16*(6), 667–674.

Kotsopoulos, J., Lubinski, J., Salmena, L., Lynch, H. T., Kim-Sing, C., Foulkes, W. D., Ghadirian, P., Neuhausen, S. L., Demsky, R., Tung, N., Ainsworth, P., Senter, L., Eisen, A., Eng, C., Singer, C., Ginsburg, O., Blum, J., Huzarski, T., Poll, A., … Hereditary Breast Cancer Clinical Study Group. (2012). Breastfeeding and the risk of breast cancer in BRCA1 and BRCA2 mutation carriers. *Breast Cancer Research : BCR*, *14*(2), R42. https://doi.org/10.1186/bcr3138

Kottemann, M. C., & Smogorzewska, A. (2013). Fanconi anaemia and the repair of Watson and Crick DNA crosslinks. *Nature*, *493*(7432), 356–363. https://doi.org/10.1038/NATURE11863

Krecke, K. N., & Gisvold, J. J. (1993). Invasive lobular carcinoma of the breast: mammographic findings and extent of disease at diagnosis in 184 patients. *American Journal of Roentgenology*, *161*(5), 957–960. https://doi.org/10.2214/ajr.161.5.8273634

Krejci, L., Altmannova, V., Spirek, M., & Zhao, X. (2012). Homologous recombination and its regulation. *Nucleic Acids Research*, *40*(13), 5795–5818. https://doi.org/10.1093/NAR/GKS270

Kroiss, R., Winkler, V., Bikas, D., Fleischmann, E., Mainau, C., Frommlet, F., Muhr, D., Fuerhauser, C., Tea, M., Bittner, B., Kubista, E., Oefner, P. J., Bauer, P., Wagner, T. M. U., Bachmann, A., Matthae, K., Bachner, M., Baldinger, C., Baldinger, M., … Berning, S. (2005). Younger birth cohort correlates with higher breast and ovarian cancer risk in European BRCA1 mutation carriers. *Human Mutation*, *26*(6), 583–589. https://doi.org/10.1002/HUMU.20261

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., McGuffog, L., Evans, D. G., Barrowdale, D., Frost, D., … Olsson, H. (2017). Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. *Jama*, *317*(23), 2402.

Kuchenbaecker, K. B., McGuffog, L., Barrowdale, D., Lee, A., Soucy, P., Dennis, J., Domchek, S. M., Robson, M., Spurdle, A. B., Ramus, S. J., Mavaddat, N., Terry, M. B., Neuhausen, S. L., Schmutzler, R. K., Simard, J., Pharoah, P. D. P., Offit, K., Couch, F. J., Chenevix-Trench, G., … Antoniou, A. C. (2017). Evaluation of Polygenic Risk Scores for Breast and Ovarian Cancer Risk Prediction in BRCA1 and BRCA2 Mutation Carriers. *Journal of the National Cancer Institute*, *109*(7). https://doi.org/10.1093/jnci/djw302

Kurian, A. W., McClure, L. A., John, E. M., Horn-Ross, P. L., Ford, J. M., & Clarke, C. A. (2009). Second Primary Breast Cancer Occurrence According to Hormone Receptor Status. *JNCI Journal of the National Cancer Institute*, *101*(15), 1058–1065. https://doi.org/10.1093/jnci/djp181

Lachman, H. M., Papolos, D. F., Saito, T., Yu, Y. M., Szumlanski, C. L., & Weinshilboum, R. M. (1996). Human catechol-O-methyltransferase pharmacogenetics: description of a functional polymorphism and its potential application to neuropsychiatric disorders. *Pharmacogenetics*, *6*(3), 243–250.

Lai, Y., Beaver, J. M., Laverde, E., & Liu, Y. (2020). Trinucleotide repeat instability via DNA base excision repair. *DNA Repair*, *93*, 102912. https://doi.org/10.1016/J.DNAREP.2020.102912

Lambertini, M., Santoro, L., del Mastro, L., Nguyen, B., Livraghi, L., Ugolini, D., Peccatori, F. A., & Azim, H. A. (2016). Reproductive behaviors and risk of developing breast cancer according to tumor subtype: A systematic review and meta-analysis of epidemiological studies. *Cancer Treatment Reviews*, *49*, 65–76. https://doi.org/10.1016/j.ctrv.2016.07.006

Land, C. E., Boice, J. D., Shore, R. E., Norman, J. E., & Tokunaga, M. (1980). Breast cancer risk from low-dose exposures to ionizing radiation: results of parallel analysis of three exposed populations of women. *Journal of the National Cancer Institute*, *65*(2), 353–376. http://www.ncbi.nlm.nih.gov/pubmed/6931253

Latham, A., Shia, J., Patel, Z., Reidy-Lagunes, D. L., Segal, N. H., Yaeger, R., Ganesh, K., Connell, L., Kemeny, N. E., Kelsen, D. P., Hechtman, J. F., Nash, G. M., Paty, P. B., Zehir, A., Tkachuk, K. A., Sheikh, R., Markowitz, A. J., Mandelker, D., Offit, K., … Stadler, Z. K. (2021). Characterization and Clinical Outcomes of DNA Mismatch Repair-deficient Small Bowel Adenocarcinoma. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research*, *27*(5), 1429–1437. https://doi.org/10.1158/1078-0432.CCR-20-2892

Lavigne, J. A., Helzlsouer, K. J., Huang, H. Y., Strickland, P. T., Bell, D. A., Selmin, O., Watson, M. A., Hoffman, S., Comstock, G. W., & Yager, J. D. (1997). An association between the allele coding for a low activity variant of catechol-O-methyltransferase and the risk for breast cancer. *Cancer Research*, *57*(24), 5493–5497. http://www.ncbi.nlm.nih.gov/pubmed/9407957

Lecarpentier, J., Noguès, C., Mouret-Fourme, E., Gauthier-Villars, M., Lasset, C., Fricker, J.-P., Caron, O., Stoppa-Lyonnet, D., Berthet, P., Faivre, L., Bonadona, V., Buecher, B., Coupier, I., Gladieff, L., Gesta, P., Eisinger, F., Frénay, M., Luporsi, E., Lortholary, A., … Andrieu, N. (2012). Variation in breast cancer risk associated with factors related to pregnancies according to truncating mutation location, in the French National BRCA1 and BRCA2 mutations carrier cohort (GENEPSO). *Breast Cancer Research*, *14*(4), R99. https://doi.org/10.1186/bcr3218

Lecarpentier, J., Noguès, C., Mouret-Fourme, E., Stoppa-Lyonnet, D., Lasset, C., Caron, O., Fricker, J.-P., Gladieff, L., Faivre, L., Sobol, H., Gesta, P., Frenay, M., Luporsi, E., Coupier, I., Lidereau, R., Andrieu, N., & Andrieu, N. (2011). Variation in breast cancer risk with mutation position, smoking, alcohol, and chest X-ray history, in the French National BRCA1/2 carrier cohort (GENEPSO). *Breast Cancer Research and Treatment*, *130*(3), 927–938. https://doi.org/10.1007/s10549-011-1655-3

Lee, S., Abecasis, G. R., Boehnke, M., & Lin, X. (2014). Rare-Variant Association Analysis: Study Designs and Statistical Tests. *The American Journal of Human Genetics*, *95*, 5–23. https://doi.org/10.1016/j.ajhg.2014.06.009

Levy-Lahad, E., Lahad, A., Eisenberg, S., Dagan, E., Paperna, T., Kasinetz, L., Catane, R., Kaufman, B., Beller, U., Renbaum, P., & Gershoni-Baruch, R. (2001). A single nucleotide polymorphism in the RAD51 gene modifies cancer risk in BRCA2 but not BRCA1 carriers. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(6), 3232–3236. https://doi.org/10.1073/pnas.051624098

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, *26*(5), 589–595. http://dx.doi.org/10.1093/bioinformatics/btp698

Li, H., Engel, C., de la Hoya, M., Peterlongo, P., Yannoukakos, D., Livraghi, L., Radice, P., Thomassen, M., Hansen, T. V., Gerdes, A.-M., Nielsen, H. R., Caputo, S. M., Zambelli, A., Borg, A., Solano, A., Thomas, A., Parsons, M. T., Antoniou, A. C., Leslie, G., … Consortium, C. (2022). *ARTICLE Risks of breast and ovarian cancer for women harboring pathogenic missense variants in BRCA1 and BRCA2 compared with those harboring protein truncating variants*. https://doi.org/10.1016/j.gim.2021.08.016

Li, J. J., & Li, S. A. (1987). Estrogen carcinogenesis in Syrian hamster tissues: role of metabolism. *Federation Proceedings*, *46*(5), 1858–1863.

Li, J., Sun, H., Huang, Y., Wang, Y., Liu, Y., & Chen, X. (2019). Pathways and assays for DNA double-strand break repair by homologous recombination. *Acta Biochimica et Biophysica Sinica*, *51*(9), 879–889. https://doi.org/10.1093/ABBS/GMZ076

Li, K.-M., Todorovic, R., Devanesan, P., Higginbotham, S., Köfeler, H., Ramanathan, R., Gross, M. L., Rogan, E. G., & Cavalieri, E. L. (2003). Metabolism and DNA binding studies of 4-hydroxyestradiol and estradiol-3,4-quinone in vitro and in female ACI rat mammary gland in vivo. *Carcinogenesis*, *25*(2), 289–297. https://doi.org/10.1093/carcin/bgg191

Liao, P. V., & Dollin, J. (2012). *Half a century of the oral contraceptive pill Historical review and view to the future*. *58*.

Lichtenstein, P., Holm, N. v., Verkasalo, P. K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A., & Hemminki, K. (2000). Environmental and Heritable Factors in the Causation of Cancer — Analyses of Cohorts of Twins from Sweden, Denmark, and Finland. *New England Journal of Medicine*, *343*(2), 78–85. https://doi.org/10.1056/NEJM200007133430201

Lieber, M. R. (2010a). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual Review of Biochemistry*, *79*, 181–211. https://doi.org/10.1146/ANNUREV.BIOCHEM.052308.093131

Lieber, M. R. (2010b). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual Review of Biochemistry*, *79*, 181–211. https://doi.org/10.1146/ANNUREV.BIOCHEM.052308.093131

Liehr, J. G., Fang, W. F., Sirbasku, D. A., & Ari-Ulubelen, A. (1986). Carcinogenicity of catechol estrogens in Syrian hamsters. *Journal of Steroid Biochemistry*, *24*(1), 353–356.

Lightbody, G., Haberland, V., Browne, F., Taggart, L., Zheng, H., Parkes, E., & Blayney, J. K. (2019). Review of applications of high-throughput sequencing in personalized medicine: Barriers and facilitators of future progress in research and clinical application. In *Briefings in Bioinformatics* (Vol. 20, Issue 5, pp. 1795–1811). Oxford University Press. https://doi.org/10.1093/bib/bby051

Lin, L., Cheng, X., & Yin, D. (2020). Aberrant DNA Methylation in Esophageal Squamous Cell Carcinoma: Biological and Clinical Implications. *Frontiers in Oncology*, *10*. https://doi.org/10.3389/FONC.2020.549850

Loman, N., Johannsson, O., r-Ola Bendahl, P., Borg, Å., Fernö, M., & Olsson, H. (1998). *Steroid Receptors in Hereditary Breast Carcinomas Associated with BRCA1 or BRCA2 Mutations or Unknown Susceptibility Genes*. https://doi.org/10.1002/(SICI)1097-0142(19980715)83:2

Lopez, J. K., & Bassett, L. W. (2009). Invasive Lobular Carcinoma of the Breast: Spectrum of Mammographic, US, and MR Imaging Findings. *RadioGraphics*, *29*(1), 165–176. https://doi.org/10.1148/rg.291085100

Lord, C. J., & Ashworth, A. (2012). The DNA damage response and cancer therapy. *Nature*, *481*(7381), 287–294. https://doi.org/10.1038/NATURE10760

Lundström, K. , Salminen, M. , Jalanko, A. , Savolainen, R. , & Ulmanen, I. (1991). Cloning and Characterization of Human Placental Catechol--Methyltransferase cDNA. *DNA and Cell Biology*, *10*(3), 181–189. https://doi.org/10.1089/dna.1991.10.181

Lynch, H. T., & Krush, A. J. (1972). Carcinoma of the breast and ovary in three families. *Obstetrical and Gynecological Survey*, *27*(3), 174–176. https://doi.org/10.1097/00006254-197203000-00014

M, P. T., & Kaakkola, S. (1999). *Catechol-O-methyltransferase (COMT): Biochemistry, Molecular Biology, Pharmacology, and Clinical Efficacy of the New Selective COMT Inhibitors*.

Ma, H., Bernstein, L., Pike, M. C., & Ursin, G. (2006). Reproductive factors and breast cancer risk according to joint estrogen and progesterone receptor status: a meta-analysis of epidemiological studies. *Breast Cancer Research : BCR*, *8*(4), R43. https://doi.org/10.1186/bcr1525

Macias, H., & Hinck, L. (2012). Mammary gland development. *Wiley Interdisciplinary Reviews: Developmental Biology*, *1*(4), 533–557. https://doi.org/10.1002/wdev.35

Malhotra, G. K., Zhao, X., Band, H., & Band, V. (2010). Histological, molecular and functional subtypes of breast cancers. *Cancer Biology & Therapy*, *10*(10), 955–960. https://doi.org/10.4161/CBT.10.10.13879

Männistö, P. T., & Kaakkola, S. (1999). Catechol-O-methyltransferase (COMT): Biochemistry, Molecular Biology, Pharmacology,

and Clinical Efficacy of the New Selective COMT Inhibitors. *Pharmacological Reviews*, *51*(4), 593 LP – 628.

Marshall, C. J., & Santangelo, T. J. (2020). Archaeal DNA Repair Mechanisms. *Biomolecules*, *10*(11), 1–23. https://doi.org/10.3390/BIOM10111472

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., Peock, S., Frost, D., Ellis, S., Platte, R., Fineberg, E., Evans, D. G., Izatt, L., Eeles, R. A., Adlard, J., Davidson, R., Eccles, D., Cole, T., Cook, J., Brewer, C., Tischkowitz, M., Douglas, F., Hodgson, S., Walker, L., … Easton, D. F. (2013). Cancer risks for BRCA1 and BRCA2 mutation carriers: Results from prospective analysis of EMBRACE. *Journal of the National Cancer Institute*, *105*(11), 812–822.

Mckenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & Depristo, M. A. (n.d.). *The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data*. https://doi.org/10.1101/gr.107524.110

McPherson, K., Steel, C. M., & Dixon, J. M. (2000). ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. *BMJ (Clinical Research Ed.)*, *321*(7261), 624–628. http://www.ncbi.nlm.nih.gov/pubmed/10977847

Meldrum, C., Doyle, M. A., & Tothill, R. W. (2011). Next-Generation Sequencing for Cancer Diagnostics: a Practical Perspective. *The Clinical Biochemist Reviews*, *32*(4), 177. /pmc/articles/PMC3219767/

Melé, M., Ferreira, P. G., Reverter, F., DeLuca, D. S., Monlong, J., Sammeth, M., Young, T. R., Goldmann, J. M., Pervouchine, D. D., Sullivan, T. J., Johnson, R., Segrè, A. v., Djebali, S., Niarchou, A., Wright, F. A., Lappalainen, T., Calvo, M., Getz, G., Dermitzakis, E. T., … Guigó, R. (2015). The human transcriptome across tissues and individuals. *Science (New York, N.Y.)*, *348*(6235), 660. https://doi.org/10.1126/SCIENCE.AAA0355

Menes, T. S., Kerlikowske, K., Lange, J., Jaffer, S., Rosenberg, R., & Miglioretti, D. L. (2017). Subsequent Breast Cancer Risk Following Diagnosis of Atypical Ductal Hyperplasia on Needle Biopsy. *JAMA Oncology*, *3*(1), 36–41. https://doi.org/10.1001/jamaoncol.2016.3022

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L. M., Ding, W., Bell, R., Rosenthal, J., Hussey, C., Tran, T., Mcclure, M., Frye, C., Hattier, T., Phelps, R., Haugen-Strano, A., … Bogden, R. (1994). *A Strong Candidate for the Breast and Ovarian Cancer Susceptibility Gene BRCA1*. *266*(5182), 66–71.

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L., Ding, W., & Et, Al. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*, *266*(5182), 66–71.

Millikan, R. C., Pittman, G. S., Tse, C. K., Duell, E., Newman, B., Savitz, D., Moorman, P. G., Boissy, R. J., & Bell, D. A. (1998). Catechol-O-methyltransferase and breast cancer risk. *Carcinogenesis*, *19*(11), 1943–1947. http://www.ncbi.nlm.nih.gov/pubmed/9855007

Milne, R. L., & Antoniou, A. C. (2011). Genetic modifiers of cancer risk for BRCA1 and BRCA2 mutation carriers. *Annals of Oncology*, *22*(suppl 1), i11–i17. https://doi.org/10.1093/annonc/mdq660

Milne, R. L., & Antoniou, A. C. (2016). Modifiers of breast and ovarian cancer risks for BRCA1 and BRCA2 mutation carriers. *Endocrine-Related Cancer*, *23*(10), T69–T84. https://doi.org/10.1530/erc-16-0277

Milne, R. L., Osorio, A., Ramón y Cajal, T., Baiget, M., Lasa, A., Diaz-Rubio, E., de la Hoya, M., Caldés, T., Teulé, A., Lázaro, C., Blanco, I., Balmaña, J., Sánchez-Ollé, G., Vega, A., Blanco, A., Chirivella, I., Esteban Cardeñosa, E., Durán, M., Velasco, E., … Benítez, J. (2010). Parity and the risk of breast and ovarian cancer in BRCA1 and BRCA2 mutation carriers. *Breast Cancer Research and Treatment*, *119*(1), 221–232. https://doi.org/10.1007/s10549-009-0394-1

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

Moghadasi, S., Meeks, H. D., Vreeswijk, M. P., Janssen, L. A., Borg, Å., Ehrencrona, H., Paulsson-Karlsson, Y., Wappenschmidt, B., Engel, C., Gehrig, A., Arnold, N., van Overeem Hansen, T., Thomassen, M., Jensen, U. B., Kruse, T. A., Ejlertsen, B., Gerdes, A. M., Pedersen, I. S., Caputo, S. M., … García, E. G. (2017, January). The BRCA1 c. 5096G>A p.Arg1699Gln (R1699Q) intermediate risk variant: Breast and ovarian cancer risk estimation and recommendations for clinical management from the ENIGMA consortium. *Journal of Medical Genetics*, *55*(1), 15–20. https://doi.org/10.1136/jmedgenet-2017-104560

Möller, S., Mucci, L. A., Harris, J. R., Scheike, T., Holst, K., Halekoh, U., Adami, H.-O., Czene, K., Christensen, K., Holm, N. v, Pukkala, E., Skytthe, A., Kaprio, J., & Hjelmborg, J. B. (2016). *The Heritability of Breast Cancer among Women in the Nordic Twin Study of Cancer*. https://doi.org/10.1158/1055-9965.EPI-15-0913

Mose, L. E., Wilkerson, M. D., Neil Hayes, D., Perou, C. M., & Parker, J. S. (2014). ABRA: Improved coding indel detection via assembly-based realignment. *Bioinformatics*, *30*(19), 2813–2815. https://doi.org/10.1093/bioinformatics/btu376

Movassagh, M., Mudvari, P., & Horvath, A. (2017). Co-Occurrence of COMT and BRCA1/2 Variants in a Population. *New England Journal of Medicine*, *376*(21), 2090–2091. https://doi.org/10.1056/NEJMc1701592

Mucci, L. A., Hjelmborg, J. B., Harris, J. R., Czene, K., Havelick, D. J., Scheike, T., Graff, R. E., Holst, K., Möller, S., Unger, R. H., McIntosh, C., Nuttall, E., Brandt, I., Penney, K. L., Hartman, M., Kraft, P., Parmigiani, G., Christensen, K., Koskenvuo, M., … Kaprio, J. (2016). Familial Risk and Heritability of Cancer

Among Twins in Nordic Countries. *JAMA*, *315*(1), 68. https://doi.org/10.1001/jama.2015.17703

Mullan, P. B., Quinn, J. E., & Harkin, D. P. (2006). The role of BRCA1 in transcriptional regulation and cell cycle control. *Oncogene*, *25*(43), 5854–5863. https://doi.org/10.1038/SJ.ONC.1209872

Murphy, C. G., & Dickler, M. N. (2016). Endocrine resistance in hormone-responsive breast cancer: mechanisms and therapeutic strategies. *Endocrine-Related Cancer*, *23*(8), R337–R352. https://doi.org/10.1530/ERC-16-0121

Nardone, A., de Angelis, C., Trivedi, M. v., Osborne, C. K., & Schiff, R. (2015). The Changing Role of ER in Endocrine Resistance. *Breast (Edinburgh, Scotland)*, *24*(0 2), S60. https://doi.org/10.1016/J.BREAST.2015.07.015

Narod, S. A. (2002). MODIFIERS OF RISK OF HEREDITARY BREAST AND OVARIAN CANCER. *Nature Reviews Cancer*, *2*(2), 113–123.

Nichols, H. B., Trentham-Dietz, A., Newcomb, P. A., Titus, L. J., Egan, K. M., Hampton, J. M., & Visvanathan, K. (2012). Postoophorectomy Estrogen Use and Breast Cancer Risk. *Obstetrics & Gynecology*, *120*(1), 27–36. https://doi.org/10.1097/AOG.0b013e31825a717b

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.

Noguchi, S., Kasugai, T., Miki, Y., Fukutomi, T., Emi, M., Nomizu, T., & Koyama, H. (1999). *Clinicopathologic Analysis of BRCA1-or BRCA2-Associated Hereditary Breast Carcinoma in Japanese Women*. https://doi.org/10.1002/(SICI)1097-0142(19990515)85:10

Ohta, T., Sato, K., & Wu, W. (2011). The BRCA1 ubiquitin ligase and homologous recombination repair. *FEBS Letters*, *585*(18), 2836–2844. https://doi.org/10.1016/J.FEBSLET.2011.05.005

Ouchi, T. (2006). BRCA1 phosphorylation: Biological consequences. *Cancer Biology and Therapy*, *5*(5), 470–475. https://doi.org/10.4161/cbt.5.5.2845

Panchal, S., Bordeleau, L., Poll, A., Llacuachaqui, M., Shachar, O., Ainsworth, P., Armel, S., Eisen, A., Sun, P., & Narod, S. A. (2010). Does family history predict the age at onset of new breast cancers in BRCA1 and BRCA2 mutation-positive families? *Clinical Genetics*, *77*(3), 273–279. https://doi.org/10.1111/j.1399-0004.2009.01328.x

Pariset, E., Malkani, S., Cekanaviciute, E., & Costes, S. v. (2021). Ionizing radiation-induced risks to the central nervous system and countermeasures in cellular and rodent models. *International Journal of Radiation Biology*, *97*(sup1), S132–S150. https://doi.org/10.1080/09553002.2020.1820598

Patel, S. M., Dash, R. C., & Hadden, M. K. (2021). Translesion synthesis inhibitors as a new class of cancer chemotherapeutics. *Expert Opinion on Investigational Drugs*, *30*(1), 13–24. https://doi.org/10.1080/13543784.2021.1850692

Paul Broca. (1866). *Traité des tumeurs*. Paris : P. Asselin, 1866-1869. https://archive.org/details/traitdestumeurs02brocgoog/page/n12

Perez-Gracia, J. L., Ruiz-Ilundain, M. G., Garcia-Ribas, I., & Carrasco, E. M. (2002). The role of extreme phenotype selection studies in the identification of clinically relevant genotypes in cancer research. *Cancer*, *95*(7), 1605–1610. https://doi.org/10.1002/CNCR.10877

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), 747–752. https://doi.org/10.1038/35021093

Pfeffer, C. M., Ho, B. N., & Singh, A. T. K. (2017). The evolution, functions and applications of the breast cancer genes BRCA1 and BRCA2. In *Cancer Genomics and Proteomics* (Vol. 14, Issue 5, pp. 293–298). International Institute of Anticancer Research. https://doi.org/10.21873/cgp.20040

Pharoah, P. D., Day, N. E., Duffy, S., Easton, D. F., & Ponder, B. A. (1997). Family history and the risk of breast cancer: a systematic review and meta-analysis. *International Journal of Cancer*, *71*(5), 800–809. http://www.ncbi.nlm.nih.gov/pubmed/9180149

Phelan, C. M., Rebbeck, T. R., Weber, B. L., Devilee, P., Ruttledge, M. H., Lynch, H. T., Lenoir, G. M., Stratton, M. R., Easton, D. F., Ponder, B. A. J., Cannon-Albright, L., Larsson, C., Goldgar, D. E., & Narod, S. A. (1996). Ovarian cancer risk in BRCA1 carriers is modified by the HRAS1 variable number of tandem repeat (VNTR) locus. *Nature Genetics*, *12*(3), 309–311. https://doi.org/10.1038/ng0396-309

Pijpe, A., Andrieu, N., Easton, D. F., Kesminiene, A., Cardis, E., Noguès, C., Gauthier-Villars, M., Lasset, C., Fricker, J.-P., Peock, S., Frost, D., Evans, D. G., Eeles, R. A., Paterson, J., Manders, P., Asperen, C. J. van, Ausems, M. G. E. M., Meijers-Heijboer, H., Thierry-Chef, I., … Leeuwen, F. E. van. (2012). Exposure to diagnostic radiation and risk of breast cancer among carriers of BRCA1/2 mutations: retrospective cohort study (GENE-RAD-RISK). *The BMJ*, *345*. https://doi.org/10.1136/BMJ.E5660

Podnar, J., Deiderick, H., & Hunicke-Smith, S. (2014). Next-Generation Sequencing Fragment Library Construction. *Current Protocols in Molecular Biology*, *107*(1), 7.17.1-7.17.16. https://doi.org/10.1002/0471142727.MB0717S107

Prentice, A. M. (2009). Obesity in emerging nations: evolutionary origins and the impact of a rapid nutrition transition. *Nestle Nutrition Workshop Series. Paediatric Programme*, *63*, 47–54; discussion 54-7, 259–268. https://doi.org/10.1159/000209972

Press, D. J., Sullivan-Halley, J., Ursin, G., Deapen, D., McDonald, J. A., Strom, B. L., Norman, S. A., Simon, M. S., Marchbanks, P. A., Folger, S. G., Liff, J. M., Burkman, R. T., Malone, K. E., Weiss, L. K., Spirtas, R., & Bernstein, L. (2011). Breast Cancer Risk and Ovariectomy, Hysterectomy, and Tubal Sterilization in the Women's Contraceptive and Reproductive Experiences Study. *American Journal of Epidemiology*, *173*(1), 38–47. https://doi.org/10.1093/aje/kwq339

Pritchard, C. C., Mateo, J., Walsh, M. F., de Sarkar, N., Abida, W., Beltran, H., Garofalo, A., Gulati, R., Carreira, S., Eeles, R., Elemento, O., Rubin, M. A., Robinson, D., Lonigro, R., Hussain, M., Chinnaiyan, A., Vinson, J., Filipenko, J., Garraway, L., … Nelson, P. S. (2016). Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer. *New England Journal of Medicine*, *375*(5), 443–453.

Qin, X., Peng, Q., Qin, A., Chen, Z., Lin, L., Deng, Y., Xie, L., Xu, J., Li, H., Li, T., Li, S., & Zhao, J. (2012). Association of COMT Val158Met polymorphism and breast cancer risk: an updated meta-analysis. *Diagnostic Pathology*, *7*, 136. https://doi.org/10.1186/1746-1596-7-136

Ragunathan, K., Upfold, N. L. E., & Oksenych, V. (2020). Interaction between Fibroblasts and Immune Cells Following DNA Damage Induced by Ionizing Radiation. *International Journal of Molecular Sciences*, *21*(22), 1–15. https://doi.org/10.3390/IJMS21228635

Rakha, E. A., & Ellis, I. O. (2010). Lobular breast carcinoma and its variants. *Seminars in Diagnostic Pathology*, *27*(1), 49–61. http://www.ncbi.nlm.nih.gov/pubmed/20306830

Rebbeck, T. R., Friebel, T., Lynch, H. T., Neuhausen, S. L., van ’t Veer, L., Garber, J. E., Evans, G. R., Narod, S. A., Isaacs, C., Matloff, E., Daly, M. B., Olopade, O. I., & Weber, B. L. (2004). Bilateral prophylactic mastectomy reduces breast cancer risk in BRCA1 and BRCA2 mutation carriers: the PROSE Study Group. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, *22*(6), 1055–1062. https://doi.org/10.1200/JCO.2004.04.188

Rebbeck, T. R., Kantoff, P. W., Krithivas, K., Neuhausen, S., Blackwood, M. A., Godwin, A. K., Daly, M. B., Narod, S. A., Garber, J. E., Lynch, H. T., Weber, B. L., & Brown, M. (1999). Modification of BRCA1-Associated Breast Cancer Risk by the Polymorphic Androgen-Receptor CAG Repeat. *The American Journal of Human Genetics*, *64*(5), 1371–1377. https://doi.org/10.1086/302366

Rebbeck, T. R., Levin, A. M., Eisen, A., Snyder, C., Watson, P., Cannon-Albright, L., Isaacs, C., Olopade, O., Garber, J. E., Godwin, A. K., Daly, M. B., Narod, S. A., Neuhausen, S. L., Lynch, H. T., & Weber, B. L. (1999). Breast cancer risk after bilateral prophylactic oophorectomy in BRCA1 mutation carriers. *Journal of the National Cancer Institute*, *91*(17), 1475–1479. http://www.ncbi.nlm.nih.gov/pubmed/10469748

Rebbeck, T. R., Mitra, N., Wan, F., Sinilnikova, O. M., Healey, S., McGuffog, L., Mazoyer, S., Chenevix-Trench, G., Easton, D. F., Antoniou, A. C., Nathanson, K. L., & Author, C. (2015). Association of Type and Location of BRCA1 and BRCA2 Mutations With Risk of Breast and Ovarian Cancer Original Investigation. *JAMA*, *313*(13), 1347–1361. https://doi.org/10.1001/jama.2014.5985

Rebbeck, T. R., Wang, Y., Kantoff, P. W., Krithivas, K., Neuhausen, S. L., Godwin, A. K., Daly, M. B., Narod, S. A., Brunet, J. S., Vesprini, D., Garber, J. E., Lynch, H. T., Weber, B. L., & Brown, M. (2001). Modification of BRCA1- and BRCA2-associated breast cancer risk by AIB1 genotype and reproductive history.

Cancer Research, *61*(14), 5420–5424. http://www.ncbi.nlm.nih.gov/pubmed/11454686

Renehan, A. G., Tyson, M., Egger, M., Heller, R. F., & Zwahlen, M. (2008). Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *The Lancet*, *371*(9612), 569–578. https://doi.org/10.1016/S0140-6736(08)60269-X

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., & Rehm, H. L. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*, *17*(5), 405–423.

Robinson, D., Holmberg, L., & Møller, H. (2008). The occurrence of invasive cancers following a diagnosis of breast carcinoma in situ. *British Journal of Cancer*, *99*(4), 611–615. https://doi.org/10.1038/sj.bjc.6604524

Rofes, P., Menéndez, M., González, S., Tornero, E., Gómez, C., Vargas-Parra, G., Montes, E., Salinas, M., Solanes, A., Brunet, J., Teulé, A., Capellá, G., Feliubadaló, L., del Valle, J., Pineda, M., & Lázaro, C. (2020). Improving Genetic Testing in Hereditary Cancer by RNA Analysis: Tools to Prioritize Splicing Studies and Challenges in Applying American College of Medical Genetics and Genomics Guidelines. *Journal of Molecular Diagnostics*, *22*(12), 1453–1468. https://doi.org/10.1016/j.jmoldx.2020.09.007

Ross, R. K., Paganini-Hill, A., Wan, P. C., & Pike, M. C. (2000). Effect of hormone replacement therapy on breast cancer risk: estrogen versus estrogen plus progestin. *Journal of the National Cancer Institute*, *92*(4), 328–332. http://www.ncbi.nlm.nih.gov/pubmed/10675382

Roy, R., Chun, J., & Powell, S. N. (2012). BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nature Reviews. Cancer*, *12*(1), 68–78. https://doi.org/10.1038/nrc3181

Saftlas, A. F., Hoover, R. N., Brinton, L. A., Szklo, M., Olson, D. R., Salane, M., & Wolfe, J. N. (1991). Mammographic densities and risk of breast cancer. *Cancer*, *67*(11), 2833–2838. https://doi.org/10.1002/1097-0142(19910601)67:11<2833::AID-CNCR2820671121>3.0.CO;2-U

Salminen, M., Lundström, K., Tilgmann, C., Savolainen, R., Kalkkinen, N., & Ulmanen, I. (1990). Molecular cloning and characterization of rat liver catechol-O-methyltransferase. *Gene*, *93*(2), 241–247.

Sandmann, S., de Graaf, A. O., Karimi, M., van der Reijden, B. A., Hellström-Lindberg, E., Jansen, J. H., & Dugas, M. (2017). Evaluating Variant Calling Tools for Non-Matched Next-Generation Sequencing Data. *Scientific Reports*, *7*. https://doi.org/10.1038/SREP43169

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, *74*(12), 5463–5467. https://doi.org/10.1073/PNAS.74.12.5463

Santos, C., Sanz-Pamplona, R., Nadal, E., Grasselli, J., Pernas, S., Dienstmann, R., Moreno, V., Tabernero, J., & Salazar, R. (2015). Intrinsic cancer subtypes-next steps into personalized medicine. *Cellular Oncology*, *38*(1), 3–16. https://doi.org/10.1007/s13402-014-0203-7

Sartori, A. A., Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., Baer, R., Lukas, J., & Jackson, S. P. (2007). Human CtIP promotes DNA end resection. *Nature*, *450*(7169), 509–514. https://doi.org/10.1038/NATURE06337

Sassa, A., & Odagiri, M. (2020). Understanding the sequence and structural context effects in oxidative DNA damage repair. *DNA Repair*, *93*, 102906. https://doi.org/10.1016/J.DNAREP.2020.102906

Savage, K. I., & Harkin, D. P. (2015). BRCA1, a "complex" protein involved in the maintenance of genomic stability. *The FEBS Journal*, *282*(4), 630–646. https://doi.org/10.1111/FEBS.13150

Sawyer, S., Mitchell, G., McKinley, J., Chenevix-Trench, G., Beesley, J., Chen, X. Q., Bowtell, D., Trainer, A. H., Harris, M., Lindeman, G. J., & James, P. A. (2012). A role for common genomic variants in the assessment of familial breast cancer. *Journal of Clinical Oncology*, *30*(35), 4330–4336. https://doi.org/10.1200/JCO.2012.41.7469

Scanlon, P. D., Raymond, F. A., & Weinshilboum, R. M. (1979). Catechol-O-methyltransferase: thermolabile enzyme in erythrocytes of subjects homozygous for allele for low activity. *Science (New York, N.Y.)*, *203*(4375), 63–65.

Schrijver, L. H., Olsson, H., Phillips, K. A., Terry, M. B., Goldgar, D. E., Kast, K., Engel, C., Mooij, T. M., Adlard, J., Barrowdale, D., Davidson, R., Eeles, R., Ellis, S., Evans, D. G., Frost, D., Izatt, L., Porteous, M. E., Side, L. E., Walker, L., … Zattara-Cannoni, H. (2018). Oral contraceptive use and breast cancer risk: Retrospective and prospective analyses from a BRCA1 and BRCA2 mutation carrier cohort study. *JNCI Cancer Spectrum*, *2*(2). https://doi.org/10.1093/jncics/pky023

Scott, C. L., Jenkins, M. A., Southey, M. C., Davis, T. A., Leary, J. A., Easton, D. F., Phillips, K. A., & Hopper, J. L. (2003). Average age-specific cumulative risk of breast cancer according to type and site of germline mutations in BRCA1 and BRCA2 estimated from multiple-case breast cancer families attending Australian family cancer clinics. *Human Genetics*, *112*(5–6), 542–551. https://doi.org/10.1007/s00439-003-0908-6

Sena, L. A., Fountain, J., Isaacsson Velho, P., Lim, S. J., Wang, H., Nizialek, E., Rathi, N., Nussenzveig, R., Maughan, B. L., Velez, M. G., Ashkar, R., Larson, A. C., Pritchard, C. C., Adra, N., Bryce, A. H., Agarwal, N., Pardoll, D. M., Eshleman, J. R., Lotan, T. L., & Antonarakis, E. S. (2021). Tumor Frameshift Mutation Proportion Predicts Response to Immunotherapy in Mismatch Repair-Deficient Prostate Cancer. *The Oncologist*, *26*(2), e270–e278. https://doi.org/10.1002/ONCO.13601

Sepahi, I., Faust, U., Sturm, M., Bosse, K., Kehrer, M., Heinrich, T., Grundman-Hauser, K., Bauer, P., Ossowski, S., Susak, H., Varon, R., Schröck, E., Niederacher, D., Auber, B., Sutter, C., Arnold, N., Hahnen, E., Dworniczak, B.,

Wang-Gorke, S., … Schroeder, C. (2019). Investigating the effects of additional truncating variants in DNA-repair genes on breast cancer risk in BRCA1-positive women. *BMC Cancer*, *19*(1). https://doi.org/10.1186/s12885-019-5946-0

Sever, R., & Brugge, J. S. (2015). Signal transduction in cancer. *Cold Spring Harbor Perspectives in Medicine*, *5*(4). https://doi.org/10.1101/cshperspect.a006098

Siepel, A., Pollard, K. S., & Haussler, D. (2006). *New Methods for Detecting Lineage-Specific Selection* (pp. 190–205). https://doi.org/10.1007/11732990_17

Singletary, S. E. (2003). Rating the risk factors for breast cancer. *Annals of Surgery*, *237*(4), 474–482. https://doi.org/10.1097/01.SLA.0000059969.64262.87

Soerjomataram, I., Louwman, W. J., van der Sangen, M. J. C., Roumen, R. M. H., & Coebergh, J. W. W. (2006). Increased risk of second malignancies after in situ breast carcinoma in a population-based registry. *British Journal of Cancer*, *95*(3), 393–397. https://doi.org/10.1038/sj.bjc.6603231

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 of the United States of America*, *98*(19), 10869–10874. https://doi.org/10.1073/pnas.191367098

Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J. S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., & Geisler, S. (2003). Repeated observation of breast tumour subtypes in independent gene expression data sets. *Proc Natl Acad Sci*, *100*. https://doi.org/10.1073/pnas.0932692100

Soukarieh, O., Gaildrat, P., Hamieh, M., Drouet, A., Baert-Desurmont, S., Frébourg, T., Tosi, M., & Martins, A. (2016). Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools. *PLoS Genetics*, *12*(1). https://doi.org/10.1371/JOURNAL.PGEN.1005756

Stark, J. M., Pierce, A. J., Oh, J., Pastink, A., & Jasin, M. (2004). Genetic steps of mammalian homologous repair with distinct mutagenic consequences. *Molecular and Cellular Biology*, *24*(21), 9305–9316. https://doi.org/10.1128/MCB.24.21.9305-9316.2004

Stewart, G. S., Maser, R. S., Stankovic, T., Bressan, D. A., Kaplan, M. I., Jaspers, N. G. J., Raams, A., Byrd, P. J., Petrini, J. H. J., & Taylor, A. M. R. (1999). The DNA double-strand break repair gene hMRE11 is mutated in individuals with an ataxia-telangiectasia-like disorder. *Cell*, *99*(6), 577–587.

Stingl, J., & Caldas, C. (2007). Molecular heterogeneity of breast carcinomas and the cancer stem cell hypothesis. *Nature Reviews Cancer*, *7*(10), 791–799. https://doi.org/10.1038/nrc2212

Sturm, M., Schroeder, C., & Bauer, P. (2016). SeqPurge: highly-sensitive adapter trimming for paired-end NGS data. *BMC Bioinformatics*, *17*, 208. https://doi.org/10.1186/s12859-016-1069-7

Sung, P., & Klein, H. (2006). Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nature Reviews. Molecular Cell Biology*, *7*(10), 739–750. https://doi.org/10.1038/NRM2008

Swain, S. M., Wilson, J. W., Mamounas, E. P., Bryant, J., Wickerham, D. L., Fisher, B., Paik, S., & Wolmark, N. (2004). Estrogen receptor status of primary breast cancer is predictive of estrogen receptor status of contralateral breast cancer. *Journal of the National Cancer Institute*, *96*(7), 516–523. http://www.ncbi.nlm.nih.gov/pubmed/15069113

Sy, S. M. H., Huen, M. S. Y., & Chen, J. (2009). PALB2 is an integral component of the BRCA complex required for homologous recombination repair. *Proceedings of the National Academy of Sciences*, *106*(17), 7155–7160. https://doi.org/10.1073/pnas.0811159106

Syvänen, A. C., Tilgmann, C., Rinne, J., & Ulmanen, I. (1997). Genetic polymorphism of catechol-O-methyltransferase (COMT): Correlation of genotype with individual variation of S-COMT activity and comparison of the allele frequencies in the normal population and Parkinsonian patients in Finland. In *Pharmacogenetics* (Vol. 7, Issue 1, pp. 65–71). https://doi.org/10.1097/00008571-199702000-00009

Szewczuk, M., Boguszewska, K., Kaźmierczak-Barańska, J., & Karwowski, B. T. (2020). The role of AMPK in metabolism and its influence on DNA damage repair. *Molecular Biology Reports*, *47*(11), 9075–9086. https://doi.org/10.1007/S11033-020-05900-X

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

Terry, M. B., Daly, M. B., Phillips, K. A., Ma, X., Zeinomar, N., Leoce, N., Dite, G. S., MacInnis, R. J., Chung, W. K., Knight, J. A., Southey, M. C., Milne, R. L., Goldgar, D., Giles, G. G., Weideman, P. C., Glendon, G., Buchsbaum, R., Andrulis, I. L., John, E. M., … Hopper, J. L. (2019). Risk-Reducing Oophorectomy and Breast Cancer Risk Across the Spectrum of Familial Risk. *JNCI: Journal of the National Cancer Institute*, *111*(3), 331–334. https://doi.org/10.1093/jnci/djy182

Thompson, D., & Easton, D. (2002). Variation in BRCA1 cancer risks by mutation position. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *11*(4), 329–336.

Trapp, O., Seeliger, K., & Puchta, H. (2011). Homologs of Breast Cancer Genes in Plants. *Frontiers in Plant Science*, *2*, 19. https://doi.org/10.3389/fpls.2011.00019

Trujillo, K. M., Yuan, S. S. F., Lee, E. Y. H. P., & Sung, P. (1998). Nuclease activities in a complex of human recombination and DNA repair factors Rad50, Mre11, and p95. *Journal of Biological Chemistry*, *273*(34), 21447–21450.

Tubbs, A., & Nussenzweig, A. (2017). Endogenous DNA Damage as a Source of Genomic Instability in Cancer. *Cell*, *168*(4), 644–656. https://doi.org/10.1016/J.CELL.2017.01.002

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Research*, *40*(15), e115. https://doi.org/10.1093/nar/gks596

Vachon, C. M., Kuni, C. C., Anderson, K., Anderson, V. E., & Sellers, T. A. (2000). Association of mammographically defined percent breast density with epidemiologic risk factors for breast cancer (United States). *Cancer Causes and Control*, *11*(7), 653–662. https://doi.org/10.1023/A:1008926607428

Valerie, K., & Povirk, L. F. (2003). Regulation and mechanisms of mammalian double-strand break repair. *Oncogene*, *22*(37), 5792–5812. https://doi.org/10.1038/SJ.ONC.1206679

Valles, G. J., Bezsonova, I., Woodgate, R., & Ashton, N. W. (2020). USP7 Is a Master Regulator of Genome Stability. *Frontiers in Cell and Developmental Biology*, *8*. https://doi.org/10.3389/FCELL.2020.00717

van de Vijver, M. J., He, Y. D., van ’t Veer, L. J., Dai, H., Hart, A. A. M., Voskuil, D. W., Schreiber, G. J., Peterse, J. L., Roberts, C., Marton, M. J., Parrish, M., Atsma, D., Witteveen, A., Glas, A., Delahaye, L., van der Velde, T., Bartelink, H., Rodenhuis, S., Rutgers, E. T., … Bernards, R. (2002). A Gene-Expression Signature as a Predictor of Survival in Breast Cancer. *New England Journal of Medicine*, *347*(25), 1999–2009. https://doi.org/10.1056/NEJMoa021967

van ’t Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A. M., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R., & Friend, S. H. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, *415*(6871), 530–536. https://doi.org/10.1038/415530a

Varon, R., Vissinga, C., Platzer, M., Cerosaletti, K. M., Chrzanowska, K. H., Saar, K., Beckmann, G., Seemanová, E., Cooper, P. R., Nowak, N. J., Stumm, M., Weemaes, C. M. R., Gatti, R. A., Wilson, R. K., Digweed, M., Rosenthal, A., Sperling, K., Concannon, P., & Reis, A. (1998). Nibrin, a novel DNA double-strand break repair protein, is mutated in Nijmegen breakage syndrome. *Cell*, *93*(3), 467–476.

Vaziri, S. A., Krumroy, L. M., Elson, P., Budd, G. T., Darlington, G., Myles, J., Tubbs, R. R., & Casey, G. (2001). Breast tumor immunophenotype of BRCA1-mutation carriers is influenced by age at diagnosis. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research*, *7*(7), 1937–1945. http://www.ncbi.nlm.nih.gov/pubmed/11448907

Virnig, B. A., Tuttle, T. M., Shamliyan, T., & Kane, R. L. (2010). Ductal Carcinoma In Situ of the Breast: A Systematic Review of Incidence, Treatment, and Outcomes. *JNCI Journal of the National Cancer Institute*, *102*(3), 170–178. https://doi.org/10.1093/jnci/djp482

Walsh, C. S. (2015). Two decades beyond BRCA1/2: Homologous recombination, hereditary cancer risk and a target for ovarian cancer therapy. *Gynecologic Oncology*, *137*(2), 343–350. https://doi.org/10.1016/J.YGYNO.2015.02.017

Wang, W. W., Spurdle, A. B., Kolachana, P., Bove, B., Modan, B., Ebbers, S. M., Suthers, G., Tucker, M. A., Kaufman, D. J., Doody, M. M., Tarone, R. E., Daly, M., Levavi, H., Pierce, H., Chetrit, A., Yechezkel, G. H., Chenevix-Trench, G., Offit, K., Godwin, A. K., & Struewing, J. P. (2001). A single nucleotide polymorphism in the 5' untranslated region of RAD51 and risk of cancer among BRCA1/2 mutation carriers. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *10*(9), 955–960. http://www.ncbi.nlm.nih.gov/pubmed/11535547

Wang, Y., Klijn, J. G., Zhang, Y., Sieuwerts, A. M., Look, M. P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M. E., Yu, J., Jatkoe, T., Berns, E. M., Atkins, D., & Foekens, J. A. (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *The Lancet*, *365*(9460), 671–679. https://doi.org/10.1016/S0140-6736(05)17947-1

Warner, E., Foulkes, W., Goodwin, P., Meschino, W., Blondal, J., Paterson, C., Ozcelik, H., Goss, P., Allingham-Hawkins, D., Hamel, N., di Prospero, L., Contiga, V., Serruya, C., Klein, M., Moslehi, R., Honeyford, J., Liede, A., Glendon, G., Brunet, J.-S., & Narod, S. (1999). Prevalence and Penetrance of BRCA1 and BRCA2 Gene Mutations in Unselected Ashkenazi Jewish Women With Breast Cancer. *JNCI Journal of the National Cancer Institute*, *91*(14), 1241–1247. https://doi.org/10.1093/jnci/91.14.1241

Watson, J. D., & Crick, F. H. C. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature 1953 171:4356*, *171*(4356), 737–738. https://doi.org/10.1038/171737a0

Weigelt, B., & Reis-Filho, J. S. (2009). Histological and molecular types of breast cancer: is there a unifying taxonomy? *Nature Reviews Clinical Oncology*, *6*(12), 718–730. https://doi.org/10.1038/nrclinonc.2009.166

Weisz, J., Fritz-Wolz, G., Gestl, S., Clawson, G. A., Creveling, C. R., Liehr, J. G., & Dabbs, D. (2000). Nuclear Localization of Catechol-O-Methyltransferase in Neoplastic and Nonneoplastic Mammary Epithelial Cells. *The American Journal of Pathology*, *156*(6), 1841–1848. https://doi.org/10.1016/S0002-9440(10)65057-2

Wolfe, J. N. (1976a). Breast patterns as an index of risk for developing breast cancer. *AJR. American Journal of Roentgenology*, *126*(6), 1130–1137. https://doi.org/10.2214/ajr.126.6.1130

Wolfe, J. N. (1976b). Risk for breast cancer development determined by mammographic parenchymal pattern. *Cancer*, *37*(5), 2486–2492. https://doi.org/10.1002/1097-0142(197605)37:5<2486::AID-CNCR2820370542>3.0.CO;2-8

Wooster, R., Bignell, G., Lancaster, J., Swift, S., Seal, S., Mangion, J., Collins, N., Gregory, S., Gumbs, C., Micklem, G., Barfoot, R., Hamoudi, R., Patel, S., Rices, C., Biggs, P., Hashim, Y., Smith, A., Connor, F., Arason, A., … Stratton, M. R. (1995). Identification of the breast cancer susceptibility gene BRCA2. *Nature*, *378*(6559), 789–792.

Wooster, R., Neuhausen, S., Mangion, J., Quirk, Y., Ford, D., Collins, N., Nguyen, K., Seal, S., Tran, T., Averill, D., & Et, Al. (1994). Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science*, *265*(5181), 2088–2090.

Wright, W. D., Shah, S. S., & Heyer, W. D. (2018a). Homologous recombination and the repair of DNA double-strand breaks. *The Journal of Biological Chemistry*, *293*(27), 10524–10535. https://doi.org/10.1074/JBC.TM118.000372

Wright, W. D., Shah, S. S., & Heyer, W. D. (2018b). Homologous recombination and the repair of DNA double-strand breaks. *The Journal of Biological Chemistry*, *293*(27), 10524. https://doi.org/10.1074/JBC.TM118.000372

Wu, L. C., Wang, Z. W., Tsan, J. T., Spillman, M. A., Phung, A., Xu, X. L., Yang, M.-C. W., 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

Wu, Y., Zhang, D., & Kang, S. (2013). Physical activity and risk of breast cancer: a meta-analysis of prospective studies. *Breast Cancer Research and Treatment*, *137*(3), 869–882. https://doi.org/10.1007/s10549-012-2396-7

Xi, B., Zeng, T., & Liu, W. (2011). Catechol-O-methyltransferase Val158Met polymorphism in breast cancer risk. *Breast Cancer Research and Treatment*, *126*(3), 839–841. https://doi.org/10.1007/s10549-010-1337-6

Xu, B., Kim, S., & Kastan, M. B. (2001). Involvement of Brca1 in S-Phase and G 2 -Phase Checkpoints after Ionizing Irradiation . *Molecular and Cellular Biology*, *21*(10), 3445–3450. https://doi.org/10.1128/MCB.21.10.3445-3450.2001

Xu, B., O'donnell, A. H., Kim, S.-T., & Kastan, M. B. (2002). Phosphorylation of Serine 1387 in Brca1 Is Specifically Required for the Atm-mediated S-Phase Checkpoint after Ionizing Irradiation 1. *CANCER RESEARCH*, *62*, 4588–4591. http://aacrjournals.org/cancerres/article-pdf/62/16/4588/2497170/ch1602004588.pdf

Xu, J., Liao, L., Ning, G., Yoshida-Komiya, H., Deng, C., & O'malley, B. W. (n.d.). *The steroid receptor coactivator SRC-3 (pCIPRAC3AIB1ACTRTRAM-1) is required for normal growth, puberty, female reproductive function, and mammary gland development*. Retrieved June 25, 2019, from www.pnas.orgcgidoi10.1073pnas.120166297

Yager, J. D. (2012). Catechol-O-methyltransferase: characteristics, polymorphisms and role in breast cancer. *Drug Discovery Today. Disease Mechanisms*, *9*(1–2), e41–e46. https://doi.org/10.1016/j.ddmec.2012.10.002

Yim, D. S., Parkb, S. K., Yoo, K. Y., Yoon, K. S., Chung, H. H., Kang, H. L., Ahn, S. H., Noh, D. Y., Choe, K. J., Jang, I. J., Shin, S. G., Strickland, P. T., Hirvonen, A., & Kang, D. (2001). Relationship between the Val158Met polymorphism of catechol O-methyl transferase and breast cancer. *Pharmacogenetics*, *11*(4), 279–286. http://www.ncbi.nlm.nih.gov/pubmed/11434504

Yoshioka, K. ichi, & Matsuno, Y. (2021). Genomic destabilization and its associated mutagenesis increase with senescence-associated phenotype expression. *Cancer Science*, *112*(2), 515–522. https://doi.org/10.1111/CAS.14746

Zhang, F., Ma, J., Wu, J., Ye, L., Cai, H., Xia, B., & Yu, X. (2009). PALB2 Links BRCA1 and BRCA2 in the DNA-Damage Response. *Current Biology*. https://doi.org/10.1016/j.cub.2009.02.018

Ziegler, R. G., Hoover, R. N., Pike, M. C., Hildesheim, A., Nomura, A. M. Y., West, D. W., Wu-Williams, A. H., Kolonel, L. N., Horn-Ross, P. L., Rosenthal, J. F., & Hyer, M. B. (1993). Migration Patterns and Breast Cancer Risk in Asian-American Women. *JNCI Journal of the National Cancer Institute*, *85*(22), 1819–1827. https://doi.org/10.1093/jnci/85.22.1819

9 Publication List

Parts of this thesis have been published in form of the following scientific article:

Sepahi, I., Faust, U., Sturm, M., Bosse, K., Kehrer, M., Heinrich, T., Grundman-Hauser, K., Bauer, P., Ossowski, S., Susak, H., Varon, R., Schröck, E., Niederacher, D., Auber, B., Sutter, C., Arnold, N., Hahnen, E., Dworniczak, B., Wang-Gorke, S., Schroeder, C. (2019). Investigating the effects of additional truncating variants in DNA-repair genes on breast cancer risk in BRCA1-positive women. BMC Cancer, 19(1). https://doi.org/10.1186/s12885-019-5946-0

10 Declaration of Contributions to the Dissertation

The dissertation work was carried out at the Institute of Medical Genetics and Applied Genomics under the supervision of Prof. Olaf Riess.

The study was designed by Dr. Christopher Schroeder.

I carried out the amplicon sequencing library preparation with the assistance of Bernadette Dahl and contributed to the establishment of the protocol for the targeted RNA-sequencing under the supervision of the staff of NCCT core facility. cDNA analysis was supported by Silja Gauß.

I have designed and ordered the panel for the sequencing. Panel sequencing was performed by NCCT, and the sequencing data analyses were supported by Dr. Marc Strum and Jakob Admard. After training by Dr. Ulrike Faust, I carried out the interpretation of the sequencing data.

Statistical analyses were carried out by me after a consultation with the institute for Biometry. RVAS analysis was supported by Hana Susak.

I confirm that I wrote the manuscript myself under the supervision of Prof. Olaf Riess and that any additional sources of information have been duly cited.

11 Acknowledgment

I would like to express my deepest appreciation to everyone who helped me and inspired me during the years that I have been working on this thesis.

This endeavor would not have been possible without the support of Prof. Olaf Riess and Dr. Christopher Schroeder. Thanks to them for their patience, support, encouragement, and professional expertise.

There is no way I could express my appreciation for Dr. Ulrike Faust's continuous support and guidance.

Many thanks to the German Cancer Consortium for Hereditary Breast and/or Ovarian Cancer for providing us with the patients' samples and data. I am also grateful to the patients who participated in this study.

Thanks should also be given to the NGS Competence Center Tübingen (NCCT) for their support with the sequencing facilities.

I would like to express my deepest gratitude to my parents for all the support and love they have shown me throughout this journey.

It is my sincere wish to thank my husband for always being there to support me in all my challenges.

12 Appendix

Table 1. List of 311 DNA repair and cancer predisposition syndrome genes as well as the pathways. DSBR: Double Strand Break Repair, SSBR: Single Strand Break Repair, HR: Homologous Recombination, NER: Nucleotide Excision Repair, BER, Base Excision Repair, FA: Fancony Anemia, NHEJ: Non-Homologous End Joining

Gene	Details	source
AKT1	Eukaryotic Type->Protein kinases->Serine/threonine protein kinases	Literature
ALKBH2	Eukaryotic Type->SSBR->Direct repair	kegg
ALKBH3	Eukaryotic Type->SSBR->Direct repair	kegg
ANKRD28	Eukaryotic Type->DSBR ->Other DSBR factors->Protein phosphatase 6	kegg
ANKRD44	Eukaryotic Type->DSBR ->Other DSBR factors->Protein phosphatase 6	kegg
ANKRD52	Eukaryotic Type->DSBR->Other DSBR factors->Protein phosphatase 6	kegg
APC	Eukaryotic Type->Centrosome formation and ciliogenesis proteins	Diagnostic
APEX1	Eukaryotic Type->SSBR->BER->AP endonucleases	kegg
APEX2	Eukaryotic Type->SSBR->BER->AP endonucleases	kegg
APITD1	Eukaryotic Type->DSBR->FA pathway->FA core complex binding factors	kegg
APLF	Eukaryotic Type->SSBR->BER->Other BER factors: Eukaryotic Type->DSBR->NHEJ->Other NHEJ factors	kegg
APTX	Eukaryotic Type->SSBR->BER->Other BER factors: Eukaryotic Type->DSBR->NHEJ->Other NHEJ factors	kegg
ATF1	Others	
ATM	Eukaryotic Type->Check point factors->Other check point factors	kegg
ATR	Eukaryotic Type->Check point factors->Other check point factors	kegg
ATRIP	Eukaryotic Type->Check point factors->Other check point factors	kegg
BABAM1	Eukaryotic Type->BRISC and BRCA1 A complex member 1	Diagnostic
BACH1	Others	
BAP1	Eukaryotic Type->Ubiquitin system-> Deubiquitinating enzyme (DUB)-> Ubiquitin-specific proteases (UBPs)	Diagnostic
BARD1	Eukaryotic Type->BRCA1/2 interactor	Literature; diagnostic
BCAS2	Eukaryotic Type->Other factors with a suspected DNA repair function->PSO4 complex	kegg
BIVM-ERCC5	Eukaryotic Type->SSBR->NER->Other NER factors	kegg
BLM	Eukaryotic Type->DSBR->HR: Eukaryotic Type->DSBR->HR->Bloom's syndrome complex (BTR): Eukaryotic Type->DSBR->FA pathway	kegg
BMPR1A	Eukaryotic Type->Protein kinases->Serine/threonine protein kinases: TKL group->TGFBR1 family	Diagnostic
BRCA1	Eukaryotic Type->DSBR->HR->Other HR factors: Eukaryotic Type->DSBR->FA pathway->Downstream FA components	kegg

BRCA2	Eukaryotic Type->DSBR->HR->Other HR factors: Eukaryotic Type->DSBR->FA pathway->Downstream FA components	kegg
BRCC3	Eukaryotic Type->Deubiquitinating enzyme->Ubiquitin-specific proteases (UBPs)	Diagnostic
BRIP1	Eukaryotic Type->DSBR->FA pathway->Downstream FA components	kegg
C17orf70	Eukaryotic Type->DSBR->FA pathway->FA core complex	kegg
C19orf40	Eukaryotic Type->DSBR->FA pathway->FA core complex	kegg
CCNH	Eukaryotic Type->SSBR->NER->TFIIH complex	kegg
CCNO	Eukaryotic Type->SSBR->BER->DNA glycosylases	kegg
CDC25A	Eukaryotic Type->manually selected	
CDC25C	Eukaryotic Type->manually selected	
CDC5L	Eukaryotic Type->Other factors with a suspected DNA repair function->PSO4 complex	kegg
CDH1	Eukaryotic Type->Cellular antigens->Proteins	Diagnostic
CDK4	Eukaryotic Type->Protein kinases->Serine/threonine protein kinases: CMGC group-> CDK family	Diagnostic
CDK7	Eukaryotic Type->SSBR->NER->TFIIH complex	kegg
CDKN1A	Eukaryotic Type->BRCA1/2 interactor	Literature
CDKN2A	Eukaryotic Type->Cellular Processes->Cell growth and death->Cell cycle->p53 signaling pathway	Diagnostic
CETN2	Eukaryotic Type->SSBR->NER->GGR (global genome repair) factors->XPC-HR23B-CETN2 complex	kegg
CHEK1	Eukaryotic Type->Check point factors->Other check point factors	kegg
CHEK2	Eukaryotic Type->Check point factors->Other check point factors	kegg
CNTLN	Eukaryotic Type->Centrosome formation and ciliogenesis proteins-> Centriole proteins	Diagnostic
CRY1	Eukaryotic Type->SSBR->Direct repair	kegg
CRY2	Eukaryotic Type->SSBR->Direct repair	kegg
CTNNA1	Eukaryotic cytoskeleton proteins->Actin filaments / Microfilaments	Diagnostic
CUL3	Eukaryotic Type->SSBR->NER->Other NER factors	kegg
CUL4A	Eukaryotic Type->SSBR->NER->GGR (global genome repair) factors: Eukaryotic Type->SSBR->NER->TCR (transcription coupled repair) factors	kegg
CUL4B	Eukaryotic Type->SSBR->NER->GGR (global genome repair) factors: Eukaryotic Type->SSBR->NER->TCR (transcription coupled repair) factors	kegg
CUL5	Eukaryotic Type->SSBR->NER->Other NER factors	kegg
DCLRE1A	Eukaryotic Type->Other factors with a suspected DNA repair function->Nucleases	kegg
DCLRE1B	Eukaryotic Type->Other factors with a suspected DNA repair function->Nucleases	kegg
DCLRE1C	Eukaryotic Type->DSBR->NHEJ->Other NHEJ factors	kegg
DDB1	Eukaryotic Type->SSBR->NER->GGR (global genome repair) factors: Eukaryotic Type->SSBR->NER->TCR (transcription coupled repair) factors	kegg

DDB2	Eukaryotic Type->SSBR->NER->GGR (global genome repair) factors->Cul4-DDB2 complex	kegg
DMC1	Eukaryotic Type->DSBR->HR->RecA family proteins	kegg
DNTT	Eukaryotic Type->DSBR->NHEJ->X-family DNA polymerases	kegg
DUT	Eukaryotic Type->Other factors with a suspected DNA repair function	kegg
E2F1	Others	
EME1	Eukaryotic Type->DSBR->HR->Other HR factors	kegg
EME2	Eukaryotic Type->DSBR->HR->Other HR factors	kegg
EPCAM	Eukaryotic Type->Cellular antigens->Proteins	Diagnostic
ERCC1	Eukaryotic Type->SSBR->NER->Other NER factors	kegg
ERCC2	Eukaryotic Type->SSBR->NER->TFIIH complex	kegg
ERCC3	Eukaryotic Type->SSBR->NER->TFIIH complex	kegg
ERCC4	Eukaryotic Type->SSBR->NER->Other NER factors	kegg
ERCC5	Eukaryotic Type->SSBR->NER->Other NER factors	kegg
ERCC6	Eukaryotic Type->SSBR->NER->TCR (transcription coupled repair) factors->Other TCR factor	kegg
ERCC8	Eukaryotic Type->SSBR->NER->TCR (transcription coupled repair) factors->Cul4-CSA complex	kegg
ESR1	Others	
EXO1	Eukaryotic Type->SSBR->MMR->Other MMR factors	kegg
EXO5	Eukaryotic Type->Other factors with a suspected DNA repair function->Nucleases	kegg
FAM175A	Eukaryotic Type->BRCA1/2 interactor	Literature; diagnostic
FAM175B	Others	
FAN1	Eukaryotic Type->DSBR->FA pathway->Other FA pathway factors	kegg
FANCA	Eukaryotic Type->DSBR->FA pathway->FA core complex	kegg
FANCB	Eukaryotic Type->DSBR->FA pathway->FA core complex	kegg
FANCC	Eukaryotic Type->DSBR->FA pathway->FA core complex	kegg
FANCD2	Eukaryotic Type->DSBR->FA pathway->FANCD2-I complex	kegg
FANCE	Eukaryotic Type->DSBR->FA pathway->FA core complex	kegg
FANCF	Eukaryotic Type->DSBR->FA pathway->FA core complex	kegg
FANCG	Eukaryotic Type->DSBR->FA pathway->FA core complex	kegg
FANCI	Eukaryotic Type->DSBR->FA pathway->FANCD2-I complex	kegg
FANCL	Eukaryotic Type->DSBR->FA pathway->FA core complex	kegg
FANCM	Eukaryotic Type->DSBR->FA pathway->FA core complex	kegg
FEN1	Eukaryotic Type->SSBR->BER->Long Patch-BER factors	kegg
FOXM1	Eukaryotic Type->manually selected	
GADD45A	Others	
GEN1	Eukaryotic Type->DSBR->HR->Other HR factors	kegg
GTF2H1	Eukaryotic Type->SSBR->NER->TFIIH complex	kegg
GTF2H2	Eukaryotic Type->SSBR->NER->TFIIH complex	kegg
GTF2H2C	Eukaryotic Type->SSBR->NER->TFIIH complex	kegg
GTF2H2D	Eukaryotic Type->SSBR->NER->TFIIH complex	kegg

GTF2H3	Eukaryotic Type->SSBR->NER->TFIIH complex	kegg
GTF2H4	Eukaryotic Type->SSBR->NER->TFIIH complex	kegg
GTF2H5	Eukaryotic Type->SSBR->NER->TFIIH complex	kegg
H2AFX	Eukaryotic Type->BRCA1/2 interactor	Literature
HDAC9	Others	
HES1	Eukaryotic Type->DSBR->FA pathway->FA core complex binding factors	kegg
HFM1	Eukaryotic Type->DSBR->HR->Other HR factors	kegg
HLTF	Eukaryotic Type->TLS (translesion DNA synthesis) factors->Rad6 epistasis group	kegg
HMGB1	Eukaryotic Type->SSBR->BER->Other BER factors: Eukaryotic Type->SSBR->MMR->Other MMR factors	kegg
HOXB13		Diagnostic
HUS1	Eukaryotic Type->Check point factors->Rad9-Hus1-Rad1 complex	kegg
HUS1B	Eukaryotic Type->Check point factors->Rad9-Hus1-Rad1 complex	kegg
KANK4		Diagnostic
LIG1	Eukaryotic Type->SSBR->BER->Long Patch-BER factors: Eukaryotic Type->SSBR->NER->Other NER factors: Eukaryotic Type->SSBR->MMR->Other MMR factors	kegg
LIG3	Eukaryotic Type->SSBR->BER->Short Patch-BER factors	kegg
LIG4	Eukaryotic Type->DSBR->NHEJ->DNA Ligase 4 complex	kegg
MBD4	Eukaryotic Type->SSBR->BER->DNA glycosylases	kegg
MDC1	Eukaryotic Type->BRCA1/2 interactor	Literature
MDM4	Eukaryotic Type->manually selected	
MEN1	Eukaryotic Type-> Histone modification proteins->HMT complexes->MLL-HCF complex	Diagnostic
MGMT	Eukaryotic Type->SSBR->Direct repair: Prokaryotic Type->SSBR->Direct repair	kegg
MLH1	Eukaryotic Type->SSBR->MMR->MutL homologs	kegg
MLH3	Eukaryotic Type->SSBR->MMR->MutL homologs	kegg
MMS19	Eukaryotic Type->SSBR->NER->Other NER factors	kegg
MNAT1	Eukaryotic Type->SSBR->NER->TFIIH complex	kegg
MPG	Eukaryotic Type->SSBR->BER->DNA glycosylases	kegg
MRE11	Eukaryotic Type->DSBR->HR->MRN(MRX) complex: Eukaryotic Type->DSBR->NHEJ->MRX complex	kegg
MSH2	Eukaryotic Type->SSBR->MMR->Mismatch and loop recognition factors	kegg
MSH3	Eukaryotic Type->SSBR->MMR->Mismatch and loop recognition factors	kegg
MSH4	Eukaryotic Type->SSBR->MMR->MutS homologs specialized for meiosis	kegg
MSH5	Eukaryotic Type->SSBR->MMR->MutS homologs specialized for meiosis	kegg
MSH6	Eukaryotic Type->SSBR->MMR->Mismatch and loop recognition factors	kegg

MUS81	Eukaryotic Type->DSBR->HR->Other HR factors	kegg
MUTYH	Eukaryotic Type->SSBR->BER->DNA glycosylases: Prokaryotic Type->SSBR->BER->DNA glycosylases	kegg
MYBBP1A	Eukaryotic Type->Other factors with a suspected DNA repair function->DNA polymerases	kegg
MYC	Others	
MYCT1		Diagnostic
NBN	Eukaryotic Type->DSBR->HR->MRN(MRX) complex	kegg
NEIL1	Eukaryotic Type->SSBR->BER->DNA glycosylases	kegg
NEIL2	Eukaryotic Type->SSBR->BER->DNA glycosylases	kegg
NEIL3	Eukaryotic Type->SSBR->BER->DNA glycosylases	kegg
NF1	Eukaryotic Type->Environmental Information Processing->Signal transduction	Diagnostic
NHEJ1	Eukaryotic Type->DSBR->NHEJ->DNA Ligase 4 complex	kegg
NOTCH2	Eukaryotic Type->Environmental Information Processing->Signal transduction->Notch signaling pathway	Diagnostic
NTHL1	Eukaryotic Type->SSBR->BER->DNA glycosylases: Prokaryotic Type->SSBR->BER->DNA glycosylases	kegg
NUDT1	Eukaryotic Type->Other factors with a suspected DNA repair function	kegg
NUDT15	Eukaryotic Type->Other factors with a suspected DNA repair function	kegg
NUDT18	Eukaryotic Type->Other factors with a suspected DNA repair function	kegg
OBSL1	Eukaryotic Type-> Centrosome formation and ciliogenesis proteins	Diagnostic
OGG1	Eukaryotic Type->SSBR->BER->DNA glycosylases	kegg
PALB2	Eukaryotic Type->DSBR->FA pathway->Downstream FA components	kegg
PAPD7	Eukaryotic Type->Other factors with a suspected DNA repair function->DNA polymerases	kegg
PARP1	Eukaryotic Type->SSBR->BER->Other BER factors	kegg
PARP2	Eukaryotic Type->SSBR->BER->Other BER factors	kegg
PARP3	Eukaryotic Type->SSBR->BER->Other BER factors	kegg
PARP4	Eukaryotic Type->SSBR->BER->Other BER factors	kegg
PCNA	Eukaryotic Type->SSBR->BER->Long Patch-BER factors: Eukaryotic Type->SSBR->MMR->Other MMR factors	kegg
PCSK7	Eukaryotic Type->Chaperones and folding catalysts-> Intramolecular chaperones->Subtilisin family	Diagnostic
PER1	Eukaryotic Type->Check point factors->Other check point factors	kegg
PER2	Eukaryotic Type->Check point factors->Other check point factors	kegg
PER3	Eukaryotic Type->Check point factors->Other check point factors	kegg
PIK3CA	Eukaryotic Type->Transferases->Transferring phosphorus-containing groups	Diagnostic
PLK1	Others	
PLRG1	Eukaryotic Type->Other factors with a suspected DNA repair function->PSO4 complex	kegg

PLS3	Eukaryotic Type->Check point factors->Other check point factors	kegg
PML	Eukaryotic Type->manually selected	
PMS1	Eukaryotic Type->SSBR->MMR->MutL homologs	kegg
PMS2	Eukaryotic Type->SSBR->MMR->MutL homologs	kegg
PNKP	Eukaryotic Type->SSBR->BER->Other BER factors	kegg
POLB	Eukaryotic Type->SSBR->BER->Short Patch-BER factors: Eukaryotic Type->DSBR->NHEJ->X-family DNA polymerases	kegg
POLD1	Eukaryotic Type->SSBR->BER: Eukaryotic Type->SSBR->MMR	kegg
POLD2	Eukaryotic Type->SSBR->BER: Eukaryotic Type->SSBR->MMR	kegg
POLD3	Eukaryotic Type->SSBR->BER: Eukaryotic Type->SSBR->MMR	kegg
POLD4	Eukaryotic Type->SSBR->BER: Eukaryotic Type->SSBR->MMR	kegg
POLE	Eukaryotic Type->SSBR->BER->Long Patch-BER factors->DNA polymerase epsilon complex	kegg
POLE2	Eukaryotic Type->SSBR->BER->Long Patch-BER factors->DNA polymerase epsilon complex	kegg
POLE3	Eukaryotic Type->SSBR->BER->Long Patch-BER factors->DNA polymerase epsilon complex	kegg
POLE4	Eukaryotic Type->SSBR->BER->Long Patch-BER factors->DNA polymerase epsilon complex	kegg
POLH	Eukaryotic Type->TLS (translesion DNA synthesis) factors->Y-family DNA polymerases	kegg
POLI	Eukaryotic Type->TLS (translesion DNA synthesis) factors->Y-family DNA polymerases	kegg
POLK	Eukaryotic Type->TLS (translesion DNA synthesis) factors->Y-family DNA polymerases	kegg
POLL	Eukaryotic Type->SSBR->BER->Other BER factors: Eukaryotic Type->DSBR->NHEJ->X-family DNA polymerases	kegg
POLM	Eukaryotic Type->DSBR->NHEJ->X-family DNA polymerases	kegg
POLN	Eukaryotic Type->TLS (translesion DNA synthesis) factors->A-family DNA polymerase	kegg
POLQ	Eukaryotic Type->Other factors with a suspected DNA repair function->DNA polymerases	kegg
POLR2A	Eukaryotic Type->SSBR->NER->TCR (transcription coupled repair) factors->DNA-directed RNA polymerase II complex	kegg
POLR2B	Eukaryotic Type->SSBR->NER->TCR (transcription coupled repair) factors->DNA-directed RNA polymerase II complex	kegg
POLR2C	Eukaryotic Type->SSBR->NER->TCR (transcription coupled repair) factors->DNA-directed RNA polymerase II complex	kegg
POLR2D	Eukaryotic Type->SSBR->NER->TCR (transcription coupled repair) factors->DNA-directed RNA polymerase II complex	kegg
POLR2E	Eukaryotic Type->SSBR->NER->TCR (transcription coupled repair) factors->DNA-directed RNA polymerase II complex	kegg
POLR2F	Eukaryotic Type->SSBR->NER->TCR (transcription coupled repair) factors->DNA-directed RNA polymerase II complex	kegg
POLR2G	Eukaryotic Type->SSBR->NER->TCR (transcription coupled repair) factors->DNA-directed RNA polymerase II complex	kegg
POLR2H	Eukaryotic Type->SSBR->NER->TCR (transcription coupled repair) factors->DNA-directed RNA polymerase II complex	kegg

POLR2I	Eukaryotic Type->SSBR->NER->TCR (transcription coupled repair) factors->DNA-directed RNA polymerase II complex	kegg
POLR2J	Eukaryotic Type->SSBR->NER->TCR (transcription coupled repair) factors->DNA-directed RNA polymerase II complex	kegg
POLR2J2	Eukaryotic Type->SSBR->NER->TCR (transcription coupled repair) factors->DNA-directed RNA polymerase II complex	kegg
POLR2J3	Eukaryotic Type->SSBR->NER->TCR (transcription coupled repair) factors->DNA-directed RNA polymerase II complex	kegg
POLR2K	Eukaryotic Type->SSBR->NER->TCR (transcription coupled repair) factors->DNA-directed RNA polymerase II complex	kegg
POLR2L	Eukaryotic Type->SSBR->NER->TCR (transcription coupled repair) factors->DNA-directed RNA polymerase II complex	kegg
POU2F1	Others	
PPM1D	Eukaryotic Type->Protein phosphatases and associated proteins->Protein Ser/ Thr phosphatases	Diagnostic
PPP4C	Eukaryotic Type->DSBR->HR->Protein phosphatase 4	kegg
PPP4R1	Eukaryotic Type->DSBR->HR->Protein phosphatase 4	kegg
PPP4R2	Eukaryotic Type->DSBR->HR->Protein phosphatase 4	kegg
PPP4R4	Eukaryotic Type->DSBR->HR->Protein phosphatase 4	kegg
PPP6C	Eukaryotic Type->DSBR->Other DSBR factors->Protein phosphatase 6	kegg
PPP6R1	Eukaryotic Type->DSBR->Other DSBR factors->Protein phosphatase 6	kegg
PPP6R2	Eukaryotic Type->DSBR->Other DSBR factors->Protein phosphatase 6	kegg
PPP6R3	Eukaryotic Type->DSBR->Other DSBR factors->Protein phosphatase 6	kegg
PRKDC	Eukaryotic Type->DSBR->NHEJ->DNA-PK complex	kegg
PRPF19	Eukaryotic Type->Other factors with a suspected DNA repair function->PSO4 complex	kegg
PRSS1	Eukaryotic Type->Environmental Information Processing->Signaling molecules and interaction	Diagnostic
PSMC3IP	Eukaryotic Type->Eukaryotic proteasome->Proteasome interacting proteins (PIPs)-> Other PIPs	Diagnostic
PTCH1	Eukaryotic Type->Environmental Information Processing->Signal transduction	Diagnostic
PTEN	Eukaryotic Type->Protein phosphatases and associated proteins	Diagnostic
RAD1	Eukaryotic Type->Check point factors	kegg
RAD17	Eukaryotic Type->Check point factors	kegg
RAD18	Eukaryotic Type->TLS (translesion DNA synthesis) factors	kegg
RAD23A	Eukaryotic Type->SSBR->NER->GGR (global genome repair) factors->XPC-HR23B-CETN2 complex	kegg
RAD23B	Eukaryotic Type->SSBR->NER->GGR (global genome repair) factors->XPC-HR23B-CETN2 complex	kegg
RAD50	Eukaryotic Type->DSBR->HR->MRN(MRX) complex: Eukaryotic Type->DSBR->NHEJ->MRX complex	kegg
RAD51	Eukaryotic Type->DSBR->HR->RecA family proteins: Eukaryotic Type->DSBR->FA pathway->Other FA pathway factors	kegg

RAD51B	Eukaryotic Type->DSBR->HR->RecA family proteins	kegg
RAD51C	Eukaryotic Type->DSBR->HR->RecA family proteins: Eukaryotic Type->DSBR->FA pathway->Other FA pathway factors	kegg
RAD51D	Eukaryotic Type->DSBR->HR->RecA family proteins	kegg
RAD52	Eukaryotic Type->DSBR->HR->Rad52 family proteins	kegg
RAD54B	Eukaryotic Type->DSBR->HR->Rad54 family proteins	kegg
RAD54L	Eukaryotic Type->DSBR->HR->Rad54 family proteins	kegg
RAD54L2	Eukaryotic Type->DSBR->HR->Rad54 family proteins	kegg
RAD9A	Eukaryotic Type->Check point factors->Rad9-Hus1-Rad1 complex	kegg
RAD9B	Eukaryotic Type->Check point factors->Rad9-Hus1-Rad1 complex	kegg
RB1	Others	
RBBP8	Eukaryotic Type->BRCA1/2 interactor	Literature; diagnostic
RBX1	Eukaryotic Type->SSBR->NER->GGR (global genome repair) factors: Eukaryotic Type->SSBR->NER->TCR (transcription coupled repair) factors	kegg
RDM1	Eukaryotic Type->DSBR->HR->Rad52 family proteins	kegg
RECQL	Eukaryotic Type->DSBR->HR->RecQ family DNA helicases	kegg
RECQL4	Eukaryotic Type->DSBR->HR->RecQ family DNA helicases	kegg
RECQL5	Eukaryotic Type->DSBR->HR->RecQ family DNA helicases	kegg
RET	Eukaryotic Type->Cytokine receptors->Receptor tyrosine kinase->RTK class XIV (RET receptor family)	Diagnostic
REV1	Eukaryotic Type->TLS (translesion DNA synthesis) factors->Y-family DNA polymerases	kegg
REV3L	Eukaryotic Type->TLS (translesion DNA synthesis) factors->B-family DNA polymerases	kegg
RFC1	Eukaryotic Type->SSBR->MMR->RFC (replication factor C)	kegg
RFC2	Eukaryotic Type->SSBR->MMR->RFC (replication factor C):Eukaryotic Type->Check point factors	kegg
RFC3	Eukaryotic Type->SSBR->MMR->RFC (replication factor C):Eukaryotic Type->Check point factors	kegg
RFC4	Eukaryotic Type->SSBR->MMR->RFC (replication factor C):Eukaryotic Type->Check point factors	kegg
RFC5	Eukaryotic Type->SSBR->MMR->RFC (replication factor C):Eukaryotic Type->Check point factors	kegg
RINT1	Eukaryotic Type->RAD50 interactor 1	Diagnostic
RMI1	Eukaryotic Type->DSBR->HR->Bloom's syndrome complex (BTR):Eukaryotic Type->DSBR->FA pathway	kegg
RMI2	Eukaryotic Type->DSBR->HR->Bloom's syndrome complex (BTR):Eukaryotic Type->DSBR->FA pathway	kegg
RNF8	Eukaryotic Type->BRCA1/2 interactor	Literature
RPA1	Eukaryotic Type->SSBR->NER->RPA (replication factor A):Eukaryotic Type->SSBR->MMR->RPA: Eukaryotic Type->DSBR->HR->RPA	kegg

RPA2	Eukaryotic Type->SSBR->NER->RPA (replication factor A): Eukaryotic Type->SSBR->MMR->RPA: Eukaryotic Type->DSBR->HR->RPA	kegg
RPA3	Eukaryotic Type->SSBR->NER->RPA (replication factor A): Eukaryotic Type->SSBR->MMR->RPA: Eukaryotic Type->DSBR->HR->RPA	kegg
RPA4	Eukaryotic Type->SSBR->NER->RPA (replication factor A): Eukaryotic Type->SSBR->MMR->RPA: Eukaryotic Type->DSBR->HR->RPA	kegg
RPS15	Eukaryotic Type->BRCA1/2 interactor	Literature
RRM1	Eukaryotic Type->Other factors with a suspected DNA repair function->Modulation of nucleotide pools	kegg
RRM2	Eukaryotic Type->Other factors with a suspected DNA repair function->Modulation of nucleotide pools	kegg
RRM2B	Eukaryotic Type->Other factors with a suspected DNA repair function->Modulation of nucleotide pools	kegg
SBDS	Eukaryotic Type-> Pre-60S particles->Export and cytoplasmic maturation factors	Diagnostic
SHFM1	Eukaryotic Type->DSBR->HR->Other HR factors	kegg
SLX1A	Eukaryotic Type->DSBR->HR->Other HR factors	kegg
SLX1B	Eukaryotic Type->DSBR->HR->Other HR factors	kegg
SLX4	Eukaryotic Type->DSBR->HR->Other HR factors	kegg
SMAD4	Eukaryotic Type->Environmental Information Processing-> Signal transduction->Wnt signaling pathway->TGF-beta signaling pathway	Diagnostic
SMARCA1	Others	
SMEK1	Eukaryotic Type->DSBR->HR->Protein phosphatase 4	kegg
SMEK2	Eukaryotic Type->DSBR->HR->Protein phosphatase 4	kegg
SMO	Eukaryotic Type->smoothelin-like	Diagnostic
SMUG1	Eukaryotic Type->SSBR->BER->DNA glycosylases	kegg
SPO11	Eukaryotic Type->DSBR->HR->Other HR factors	kegg
SSBP1	Prokaryotic Type->SSBR->MMR->Other MMR factors: Prokaryotic Type->TLS (translesion DNA synthesis) factors->Other SOS response factors	kegg
STAT1	Others	
STK11	Eukaryotic Type->Environmental Information Processing->Signal transduction->PI3K-Akt signaling pathway->AMPK signaling pathway->AMPK signaling pathway	Diagnostic
STRA13	Eukaryotic Type->DSBR->FA pathway->FA core complex binding factors	kegg
TCEB1	Eukaryotic Type->SSBR->NER->Other NER factors	kegg
TCEB2	Eukaryotic Type->SSBR->NER->Other NER factors	kegg
TCEB3	Eukaryotic Type->SSBR->NER->Other NER factors	kegg
TCEB3B	Eukaryotic Type->SSBR->NER->Other NER factors	kegg
TCEB3C	Eukaryotic Type->SSBR->NER->Other NER factors	kegg
TCEB3CL	Eukaryotic Type->SSBR->NER->Other NER factors	kegg

TDG	Eukaryotic Type->SSBR->BER->DNA glycosylases: Prokaryotic Type->SSBR->BER->DNA glycosylases	kegg
TDP1	Eukaryotic Type->SSBR->BER->Other BER factors	kegg
TELO2	Eukaryotic Type->DSBR->FA pathway->Other FA pathway factors	kegg
TIMELESS	Eukaryotic Type->Check point factors->FPC (fork protection complex)	kegg
TIPARP	Eukaryotic Type->Enzymes ->Transferases->Glycosyltransferases->Pentosyltransferases->NAD+ ADP-ribosyltransferase	Diagnostic
TIPIN	Eukaryotic Type->Check point factors->FPC (fork protection complex)	kegg
TMEM189	Eukaryotic Type->TLS (translesion DNA synthesis) factors->Rad6 epistasis group	kegg
TMEM189-UBE2V1	Eukaryotic Type->TLS (translesion DNA synthesis) factors->Rad6 epistasis group	kegg
TOP3A	Eukaryotic Type->DSBR->HR->Bloom's syndrome complex (BTR): Eukaryotic Type->DSBR->FA pathway	kegg
TOP3B	Eukaryotic Type->DSBR->HR->Bloom's syndrome complex (BTR): Eukaryotic Type->DSBR->FA pathway	kegg
TOPBP1	Eukaryotic Type->BRCA1/2 interactor	Literature
TP53	Eukaryotic Type->Check point factors->Other check point factors	kegg
TP53BP1	Eukaryotic Type->BRCA1/2 interactor	Literature; diagnostic
TREX1	Eukaryotic Type->Other factors with a suspected DNA repair function->Nucleases	kegg
TREX2	Eukaryotic Type->Other factors with a suspected DNA repair function->Nucleases	kegg
UBE2A	Eukaryotic Type->TLS (translesion DNA synthesis) factors->Rad6 epistasis group	kegg
UBE2B	Eukaryotic Type->TLS (translesion DNA synthesis) factors->Rad6 epistasis group	kegg
UBE2N	Eukaryotic Type->TLS (translesion DNA synthesis) factors->Rad6 epistasis group	kegg
UBE2NL	Eukaryotic Type->TLS (translesion DNA synthesis) factors->Rad6 epistasis group	kegg
UBE2T	Eukaryotic Type->DSBR->FA pathway->FA core complex binding factors	kegg
UBE2V1	Eukaryotic Type->TLS (translesion DNA synthesis) factors->Rad6 epistasis group	kegg
UBE2V2	Eukaryotic Type->TLS (translesion DNA synthesis) factors->Rad6 epistasis group	kegg
UIMC1	Eukaryotic Type->BRCA1/2 interactor	Literature; diagnostic
UNG	Eukaryotic Type->SSBR->BER->DNA glycosylases: Prokaryotic Type->SSBR->BER->DNA glycosylases	kegg
USP1	Eukaryotic Type->DSBR->FA pathway->Other FA pathway factors	kegg
VHL	Eukaryotic Type->Genetic Information Processing->Folding, sorting and degradation->Ubiquitin mediated proteolysis	Diagnostic
WDR48	Eukaryotic Type->DSBR->FA pathway->Other FA pathway factors	kegg

WRN	Eukaryotic Type->DSBR->HR->RecQ family DNA helicases	kegg
XPA	Eukaryotic Type->SSBR->NER->Other NER factors	kegg
XPC	Eukaryotic Type->SSBR->NER->GGR (global genome repair) factors->XPC-HR23B-CETN2 complex	kegg
XRCC1	Eukaryotic Type->SSBR->BER->Short Patch-BER factors	kegg
XRCC2	Eukaryotic Type->DSBR->HR->RecA family proteins	kegg
XRCC3	Eukaryotic Type->DSBR->HR->RecA family proteins	kegg
XRCC4	Eukaryotic Type->DSBR->NHEJ->DNA Ligase 4 complex	kegg
XRCC5	Eukaryotic Type->DSBR->NHEJ->DNA-PK complex	kegg
XRCC6	Eukaryotic Type->DSBR->NHEJ->DNA-PK complex	kegg
ZNF350	Others	

Table 3 List of putative truncating variants in DNA repair genes. 43 truncating variants were detected in 36 DNA repair genes. This table has been reprinted from Sepahi et al., 2019

Gene	Coding position	Variant Type	Pathway
MSH6	MSH6: ENST00000234420: exon4/10: c.2764C>T: p.Arg922*	stop-gain	SSBR
FANCL	FANCL: ENST00000402135: exon14/14: c.1111_1114dupATTA: p.Thr372fs	frameshift	DSBR
OBSL1	OBSL1: ENST00000404537: exon20/21: c.5583delC: p.Thr1862fs	frameshift	Centrosome formation
OBSL1	OBSL1: ENST00000404537: exon12/21: c.3906G>A: p.Trp1302*	stop-gain	Centrosome formation
OBSL1	OBSL1: ENST00000404537: exon5/20: c.2135-3_2135-2delCA	Splice-acceptor	Centrosome formation
XPC	XPC: ENST00000285021: exon3/16: c.342_343delAG: p.Ala116fs	frameshift	SSBR
ATRIP	ATRIP: ENST00000320211: exon8/13: c.1152_1155delTGGA: p.Gly385fs	frameshift	Check point factors
POLQ	POLQ: ENST00000264233: exon16/30: c.4262_4268delTACTATT: p.Ile1421fs	frameshift	DNA polymerases
RAD1	RAD1:ENST00000341754: exon4/6:c .325C>T: p.Arg109*	stop-gain	Check point factors
RAD1	RAD1: ENST00000341754: exon2/6: c.168_172delAAAGT: p.Lys57fs	frameshift	Check point factors
POLK	POLK: ENST00000241436: exon10/15: c.1243delA: p.Ser415fs	frameshift	Y-family DNA polymerases
UIMC1	UIMC1: ENST00000377219: exon15/15: c.1996C>T: p.Gln667*	stop-gain	BRCA1/2 interactor
MDC1	MDC1: ENST00000376406: exon10/15: c.3861dupA: p.Ala1288fs	frameshift	BRCA1/2 interactor
MSH5	MSH5: ENST00000375742: exon7/25: c.592C>T: p.Arg198*	stop-gain	SSBR
MYCT1	MYCT1: ENST00000367245: exon2/2: c.623G>A: p.Trp208*	stop-gain	no kegg information
SBDS	SBDS: ENST00000246868: exon2/4: c.258+2T>C:	splice-donor	Pre-60S particles
RINT1	RINT1: ENST00000257700.4: exon9/15: c.1332dupA: p.Phe445fs	frameshift	RAD50 interactor 1

WRN	WRN: ENST00000298139: exon6/35: c.522_523dupCT: p.Trp175fs	frameshift	DSBR
NBN	NBN: ENST00000265433: exon6/16: c.612delT: p.Ile205fs	frameshift	DSBR
CNTLN	CNTLN: ENST00000380647: exon7/26: c.1087C>T: p.Gln363*	stop-gain	Centrosome formation
DCLRE1C	DCLRE1C: ENST00000378278 : exon10/14: c.796C>T: p.Gln266*	stop-gain	DSBR
ALKBH3	ALKBH3:ENST00000302708: exon2/10: c.19C>T: p.Arg7*	stop-gain	SSBR
RAD52	RAD52: ENST00000545564: exon6/7: c.508_511delCTCT: p.Leu170fs	frameshift	DSBR
GTF2H3	GTF2H3: ENST00000543341: exon1/13: c.3G>A: p.Met1?	start-loss	SSBR
BRCA2	BRCA2: ENST00000380152: exon11/27: c.3865_3868delAAAT: p.Lys1289fs	frameshift	DSBR
BRCA2	BRCA2: ENST00000380152: exon22/27: c.8875G>T: p.Glu2959*	stop-gain	DSBR
FAN1	FAN1: ENST00000362065: exon2/15: c.929C>G: p.Ser310*	stop-gain	DSBR
TIPIN	TIPIN: ENST00000261881: exon2/8: c.20delA: p.Asn7fs	frameshift	Check point factors
NEIL1	NEIL1: ENST00000355059: exon2/10: c.572dupC: p.Pro192fs	frameshift	SSBR
NEIL1	NEIL1: ENST00000355059M_001256552.1: exon4/9: c.876+1G>T	splice-donor	SSBR
MPG	MPG: ENST00000219431M_002434.3: exon3/5: c.253C>T: p.Arg85*	stop-gain	SSBR
EME2	EME2: ENST00000307394: exon4/8: c.541_544delGCTG: p.Ala181fs	frameshift	DSBR
EME2	EME2: ENST00000307394: exon7/8: c.929delC: p.Ala310fs	frameshift	DSBR
NTHL1	NTHL1: ENST00000219066: exon2/6: c.268C>T: p.Gln90*	stop-gain	SSBR
SLX1B	SLX1B: ENST00000330181: exon4/5: c.711-1G>C	splice-acceptor	DSBR
FANCA	FANCA: ENST00000389301: exon27/43: c.2571C>A: p.Cys857*	stop-gain	DSBR
PER1	PER1: ENST00000317276: exon12/23: c.1397delT: p.Leu466fs	frameshift	Check point factors
NF1	NF1: ENST00000358273: exon35/58: c.4600C>T: p.Arg1534*	stop-gain	Signal transduction
BRIP1	BRIP1: ENST00000259008: exon9/20: c.1236delA: p.Val413fs	frameshift	DSBR
TCEB3B	TCEB3B: ENST00000332567: exon1/1: c.319delG: p.Glu107fs	frameshift	SSBR
ERCC2	ERCC2: ENST00000391945: exon18/23: c.1703_1704delTT: p.Phe568fs	frameshift	SSBR
PNKP	PNKP: ENST00000322344: exon17/17: c.1510delC: p.Arg504fs	frameshift	SSBR
PNKP	PNKP: ENST00000322344: exon11/16: c.1029+2T>C	Splice-donor	SSBR

ATM, BARD1, BRCA1, BRCA2, BRIP1, CDH1, CHEK2, ERCC2, FAM175A, FANCA, FANCL, FANCM, GPRC5A, MAP3K1, MLH1, MRE11, MSH2, MSH6, MUTYH, MYCT1, NBN, PALB2, PIK3CA, PMS2, PPM1D, PTEN, RAD50, RAD51C, RAD51D, RINT1, SMARCA4, STK11, TP53 **and XRCC2 (Targeted-RNA-seq genes)**

Table 4-BRACa1 pathogenic variants

Genomic position	Start	End	Variant Type	Coding region (ENST00000357654)	Protein Change	Early AAO co	Control cohort
chr17	41197784	41197784	Stop -gain	c.5503C>T	Arg1835*	no	yes
chr17	41199678	41199678	Stop-gain	c.5449G>T	Glu1817*	no	yes
chr17	41201136	41201136	Splice-donor	c.5406+2delT		no	yes
chr17	41201198	41201198	Stop-gain	c.5346G>A	Trp1782*	yes	no
chr17	41201199	41201199	Stop-gain	c.5345G>A	Trp1782*	yes	no
chr17	41209079	41209079	Frameshift	c.5266dupC	Gln1756fs	yes	yes
chr17	41209095	41209095	Stop-gain	c.5251C>T	Arg1751*	no	yes
chr17	41215361	41215361	Frameshift	c.5182delA	Met1728fs	yes	no
chr17	41215363	41215366	Frameshift	c.5177_5180delGAAA	Arg1726fs	yes	no
chr17	41215364	41215364	Stop-gain	c.5179A>T	Lys1727*	yes	no
chr17	41215392	41215392	Splice-acceptor	c.5153-2delA		yes	no
chr17	41215890	41215890	Splice-donor	c.5152+1G>C		yes	no
chr17	41215906	41215906	Frameshift	c.5137delG	Val1713fs	no	yes
chr17	41215947	41215947	Missense	c.5096G>A	Arg1699Gln	no	yes
chr17	41222939	41222939	Splice region	c.4986+6T>C		yes	
chr17	41223187	41223217	Frameshift	c.4714_4744delTCTGATGACCCTGA ATCTGATCCTTCTGAAG	Ser1572fs	yes	no
chr17	41223242	41223242	Stop-gain	c.4689C>G	Tyr1563*	yes	yes
chr17	41226489	41226490	Frameshift	c.4533_4534delCA	His1511fs	no	yes
chr17	41234451	41234451	Stop-gain	c.4327C>T	Arg1443*	yes	no
chr17	41234556	41234556	Stop-gain	c.4222C>T	Gln1408*	yes	no
chr17	41234568	41234568	Frameshift	c.4210delC	Leu1404fs	no	yes
chr17	41242963	41242963	Stop-gain	c.4183C>	Gln1395*	yes	yes
chr17	41242984	41242988	Frameshift	c.4158_4162delCTCTC	Ser1387fs	yes	no
chr17	41243480	41243483	Frameshift	c.4065_4068delTCAA	Asn1355fs	yes	yes
chr17	41243513	41243513	Frameshift	c.4035delA	Glu1346fs	no	yes
chr17	41243789	41243792	Frameshift	c.3756_3759delGTCT	Ser1253fs	no	yes
chr17	41243805	41243817	Frameshift	c.3731_3743delATAGCACCGTTGC	His1244fs	yes	no
chr17	41243844	41243848	Frameshift	c.3700_3704delGTAAA	Val1234fs	yes	no
chr17	41243923	41243923	Frameshift	c.3624dupA	Leu1209fs	yes	no
chr17	41243936	41243936	Frameshift	c.3612delA	Ala1206fs	no	yes
chr17	41244057	41244067	Frameshift	c.3481_3491delGAAGATACTAG	Glu1161fs	no	yes
chr17	41244063	41244063	Frameshift	c.3485delA	Asp1162fs	yes	yes
chr17	41244214	41244217	Frameshift	c.3331_3334delCAAG	Gln1111fs	yes	no
chr17	41244557	41244557	Frameshift	c.2989_2990dupAA	Asn997fs	no	yes
chr17	41244625	41244625	Stop-gain	c.2923C>T	Gln975*	yes	no
chr17	41244826	41244826	Stop-gain	c.2722G>T	Glu908	no	yes
chr17	41244862	41244863	Frameshift	c.2685_2686delAA	Pro897fs	no	yes
chr17	41244997	41244997	Stop-gain	c.2551G>T	Glu851*	yes	no
chr17	41245136	41245137	Frameshift	c.2411_2412delAG	Gln804fs	no	yes
chr17	41245210	41245210	Stop-gain	c.2338C>T	Gln780*	yes	yes
chr17	41245239	41245239	Stop-gain	c.2309C>A	Ser770*	no	yes
chr17	41245347	41245351	Frameshift	c.2197_2201delGAGAA	Glu733fs	no	yes
chr17	41245587	41245587	Frameshift	c.1961delA	Lys654fs	no	yes

chr17	41245670	41245670	Frameshift	c.1874_1877dupTAGT	Val627fs	yes	yes
chr17	41245861	41245861	Stop-gain	c.1687C>T	Gln563*	yes	yes
chr17	41245927	41245927	Stop-gain	c.1621C>T	Gln541*	no	yes
chr17	41246040	41246044	Frameshift	c.1504_1508delTTAAA	Leu502fs	yes	no
chr17	41246212	41246212	Frameshift	c.1336delA	Arg446fs	yes	no
chr17	41246531	41246531	Frameshift	c.1016dupA	Val340fs	yes	no
chr17	41246533	41246596	Frameshift	c.952_1015delCATAACAGATGGGCTGGAAGTAAGGAAACATGTAATGATAGGCGGACTCCCAGCACAGAAAAAA	His318fs	no	yes
chr17	41246785	41246785	Stop-gain	c.763G>T	Glu255*	no	yes
chr17	41246872	41246872	Frameshift	c.676delT	Cys226fs	yes	no
chr17	41251834	41251834	Stop-gain	c.505C>T	Gln169*	yes	no
chr17	41256985	41256985	Splice region	c.213-12A>G		yes	no
chr17	41258504	41258504	Missense	c.181T>G	Cys61Gly	yes	yes
chr17	41258525	41258525	Stop-gain	c.160C>T	Gln54*	yes	no
chr17	41267764	41267765	Frameshift	c.112_113delAA	Lys38fs	no	yes
chr17	41267797	41267797	Splice-acceptor	c.81-1G>A		no	yes
chr17	41276032	41276032	Splice-donor	c.80+2T>A		no	yes
chr17	41276045	41276046	Frameshift	c.68_69delAG	Glu23fs	yes	yes
chr17	41276047	41276047	Frameshift	c.66dupA	Glu23fs	no	yes
chr17	41276080	41276080	Stop-gained	c.34C>T	Gln12*	yes	no