Comprehensive search for genetic predisposition to breast cancer based on multigene panel testing

Hata, Chihiro

Doctor of Philosophy

Department of Genetics
School of Life Science
The Graduate University for Advanced Studies,
SOKENDAI

Abstract

Breast cancer is the most common cancer in women. It has been a long-known disease about its pathological conditions. Approximately 10% of women develop breast cancer during their lifetime. There are several subtypes of breast cancer, and appropriate treatments are selected based on the subtypes. For example, PARP inhibitor has been demonstrated to be very effective in treating for patients with triple negative breast cancer (TNBC). The presence of germline mutations on specific genes can be a marker for selecting effective treatments. With advancements in DNA sequencing technologies, many genes including *BRCA1/2* have been shown to be associated with breast cancer. Recently, multigene panel testing has been conducted to detect pathogenic variants in breast cancer patients. However, there are several issues such as selection of appropriate genes and lack of information for accurate classification of pathogenic variants in populations other than European-descent populations. In this doctoral thesis, I attempted to search for pathogenic germline mutations in Chinese patients with breast cancer using multigene panel testing.

I investigated genotype-phenotype correlations in Chinese patients with breast cancer by using next-generation sequencing data for 54 breast cancer predisposition genes. First, I conducted multigene panel testing for 583 Chinese patients with breast cancer. I detected 78 pathogenic mutations comprising 43 single nucleotide variants (SNVs) and 35 short insertions and deletions (INDELs). In other words, 14.2% of the patients had pathogenic mutations (83/583). The utility of the multigene panel was demonstrated by the fact that the pathogenesis of a larger number of the patients was identified compared to genetic testing focusing only on *BRCA1/2*. Furthermore, I showed that pathogenic germline mutations were associated with several clinical features of breast cancer including TNBC and histological grades. On the other hand, I found that there were two issues emerged: 1) Even after the 54 genes were evaluated, the proportion of the patients whose pathogenesis were not clearly explained was still high, and 2) a large number of variant of uncertain significance (VUS) was detected as the bad side of using the multigene panel.

Second, I focused on the fact that a general application of pathogenic mutation search was limited to SNVs and short INDELs. I hypothesized that exploration of intermediate-sized INDELs (50 bp to 10,000 bp) in breast cancer patients was useful to resolve the first issue emerged in the previous chapter. By examining intermediate-sized

INDEL by using a state-of-the-art tool together with deep bioinformatics analyses, I identified a novel intermediate-sized deletion–insertion in *PTENα*, which can be a disease risk factor for breast cancer and illustrated the importance of examining intermediate-sized INDELs in multigene panel testing. To my best knowledge, this is the first report of pathogenic mutation on a translational variant of PTEN gene in breast cancer.

Third, I formulated a methodology to assess the association between the burden of rare and harmful VUSs and breast cancer risk. The biggest problem in multigene panel testing is that VUSs have been detected in about 30% of patients who underwent multigene panel testing, but VUS information has not been fully used to determine treatment strategy. I thought that a part of VUSs detected in the multigene panel testing were harmful mutations and could be involved in the etiology of breast cancer. I applied rare exome variant ensemble learner (REVEL) score that was an ensemble method for predicting the pathogenicity of missense variants to estimate functional damage for the protein by VUSs. The results showed that VUSs, which were predicted to have a detrimental effect on gene function at a very low frequency, were significantly more abundant in breast cancer patients who did not have other distinct pathogenic variants. The result in the first in the world to incorporate the concept of the burden test and to demonstrate the association between VUS and breast cancer.

In conclusion, the results of my study contribute to discoveries of many of novel pathogenic germline mutations in East Asian population. In addition, the result from the analysis for the association between VUS and breast cancer risk will provide clues for new therapeutic strategies targeted to POP with harmful VUS.

Table of Contents

Abbreviation table.

Abbreviation	Definition
1KG	1000 Genomes Project
95% CI	95% Confidence interval
ACMG	The American College of Medical Genetics and Genomics
AMP	The Association for Molecular Pathology
BWA	Burroughs–Wheeler aligner
CK5/6	Cytokeratin 5/6
EGFR	Epidermal growth factor
ENIGMA	The Evidence-based Network for the Interpretation of Germline Mutant Allele
ER	Estrogen receptor
ExAC	Exome Aggregation Consortium
FISH	Fluorescence in situ hybridization
gnomAD	The Genome Aggregation Database
HER2	Human epidermal growth factor receptor type 2
HR	Homologous recombination
IHC	Immunohistochemistry
INDEL	Insertion and deletion
MMR	Mismatch repair
NCCN	National Comprehensive Cancer Network
NGS	Next-generation sequencing
OR	Odds ratio
PARP	Poly ADP ribose polymerase
POP	Patients without pathogenic variants
PR	Progesterone receptor
PRS	Polygenic risk score
PWP	Patients with pathogenic variants
RefSeq	Reference Sequence Database
SNV	Single nucleotide variant
TNBC	Triple negative breast cancer
VCF	The variant call format
VUS	Variant of uncertain significant

Chapter 1: General Introduction

1.1. Incidence of breast cancer is worldwide

Breast cancer is the most common cancer in women. Its pathological conditions are known for a long time (Turnbull et al., 2008). The first description of breast cancer was in 1600 BC and was recorded in the Edwin Smith papyrus (Breasted., 1930). The origin of the term “cancer” can be traced back to ancient Greece. In 460 BC, the Greek physician Hippocrates described breast cancer. He named cancer karkinos, which is a Greek word for “crab”, probably because the appearance of invasive ductal breast cancer presented the finger-like spreading projections, which are reminiscent of crab legs. Subsequently, the Hippocratic physicians used the term karkinoma or carcinoma to describe malignant tumors (Lukong, 2017).

Presently, approximately 10% women develop breast cancer during their lifetime. European women have a higher risk of developing breast cancer than Asian or African women (Jemal et al., 2010). Breast cancer in males is rare and comprises 0.5%–1% of patients with breast cancer. The reason for the low incidence rate in men is that they have a relatively low amount of breast tissue and different hormonal environment as compared to women (Serdy et al., 2017). In 2018, 2,088,849 women were diagnosed with breast cancer; of these, 62,669 died worldwide (The GLOBOCAN 2018 database provided by The International Agency for Research on Cancer: http://gco.iarc.fr/today/home). It is estimated that 268,600 women in the United States will be diagnosed with invasive breast cancer and 62,930 women will be diagnosed with *in situ* breast cancer in 2019 (Cancer.Net provided by American Society of Clinical Oncology: https://www.cancer.net/cancer-types/breast-cancer/statistics).

Epidemiological studies have indicated that several environmental factors such as age, obesity, smoking, and alcohol consumption are associated with the risk of breast

cancer (Kispert et al., 2017). A contribution of genetic factors to the etiology of breast cancer has been suggested by the presence of familial aggregation of the disease (Hemminki et al., 1998). Breast cancer is considered to be a multifactorial disease caused by a complex interplay among genetic, epigenetic, and environmental factors.

1.2. Treatment for subtypes of breast cancer

Breast cancer is not a homogeneous disease, and it can be divided into multiple subtypes based on molecular characteristics (Perou et al., 2000). Effective therapeutic strategies are different based on the breast cancer subtype. Therefore, a better understanding of the molecular mechanisms underlying heterogeneity among breast cancer subtypes is necessary for the development of effective treatments and drug therapeutics.

Clinically, breast cancer is primarily categorized based on the expression of estrogen receptors (ERs), progesterone receptors (PRs), and human epidermal growth factor receptor type 2 (HER2; also known as ERBB2). Ki67 is a nuclear antigen encoded by *MKI67* and has been widely used a proliferation marker in breast cancer. Higher score of Ki67 indicates higher speed of cancer cell proliferation. The presence or absence of these receptors guide clinical decisions about treatment protocols (Goldhirsch et al., 2011). Standard subtypes are Luminal A, Luminal B (HER2 positive), Luminal B (HER2 negative), HER2 positive (non-luminal) and triple-negative breast cancer (TNBC), according to the St. Gallen International Expert Consensus 2011 (Table 1.1). Knowledge of these subtypes is important when deciding upon endocrine and molecularly targeted therapies.

Female hormones (i.e., estrogen and progesterone) are strongly associated with the development of breast cancer. Estrogen is secreted by the ovaries and plays a role in

promoting cell proliferation and cell differentiation, leading to the initiation and growth of breast cancer (Brisken et al., 2010). Tamoxifen was the first anti-estrogen drug discovered, and it has been used as standard endocrine therapy for decades (Meisel et al., 2018). This drug suppresses hormone-dependent growth of breast cancer by preventing estrogen from binding to ER. The catalytic conversion of androgen to estrogen is due to the enzyme aromatase. Aromatase inhibitors decrease circulating estrogen and are administered as endocrine therapy in breast cancer. These endocrine therapies offer advantages in ER and/or PR positive breast cancers that are categorized into Luminal A and Luminal B types, respectively (Table 1.1).

HER2 is a member of the epidermal growth factor family. It has been reported that HER2 was frequently amplified in breast cancer, and the amplification of this oncogene was associated with a poor prognosis (Slamon et al., 1987). Trastuzumab (brand name, Herceptin) was developed as a recombinant DNA-derived humanized monoclonal antibody that selectively targets the extracellular domain of HER2 (Meisel et al., 2018). Herceptin, which molecularly targets HER2, has been used as a therapeutic agent for HER2 positive patients (Table 1.1).

Patients with TNBC lack ER and PR expressions and the amplification of HER2. Therefore, the endocrine and molecularly targeted therapies cannot be applied to patients with TNBC. Although chemotherapy remains the standard therapy, patients with TNBC often have a poor prognosis (Bianchini et al., 2016). Recently, poly ADP ribose polymerase (PARP) inhibitors have been developed based on using a synthetic lethal approach for anticancer therapeutics, in which PARP inhibitors selectively kill cancer cells that have defects in the homologous recombination (HR) DNA repair pathway (Yap et al., 2011). About 20% of TNBC patients harbor either germline or somatic mutations of *BRCA1* or

BRCA2. *BRCA1/2* are involved in the HR DNA repair pathway, and these genes are known to cause hereditary breast and ovarian cancer syndrome as discussed in the following sections. TNBC patients with hereditary deficiency in the HR DNA repair pathway can be treated by administrating PARP inhibitors (Sunada et al., 2018). The discovery of PARP inhibitors has raised global attention toward genetic testing for cancer-associated genes, including *BRCA1/2*.

1.3. Relationship between genes and breast cancer

Approximately 10% breast cancers are considered heritable. Familial cases are more likely to develop early-onset and high-grade malignant tumors. If a woman has a family history of breast cancer in first-degree relatives (e.g., mother or sister), the risk of disease development nearly doubles (Hemminki et al., 1998). Thus, a family history of breast cancer is suggested as one of the greatest risk factors for the disease. Researchers at the University of California at Berkeley mapped a susceptibility locus for breast cancer on chromosome 17q21 from pedigree analysis based on 23 extended families with 146 cases of breast cancer (Hall et al., 1990). Mary-Claire King later named a gene on chromosome 17q21 that was said to be responsible for breast cancer as breast cancer gene 1 (*BRCA1*). Then, researchers at the University of Utah and Myriad Genetics succeeded in cloning *BRCA1* from the 17q21 region (Miki et al., 1994). Furthermore, in 1994, Stratton's group reported a second breast cancer susceptibility locus, *BRCA2*, on 13q12-13 (Wooster et al., 1994; Wooster et al., 1995). Presently, *BRCA1/2* are known to be tumor suppressor genes and are involved in the HR DNA repair pathway. There are remarkable genetic effects of *BRCA1/2* mutations on breast cancer risk. Cumulative risk for breast cancer development in women aged 70 years is estimated to be 57% for *BRCA1* mutation carriers and 49% for

BRCA2 mutation carriers (Chen et al., 2007). Furthermore, the Ashkenazi Jewish population has a high frequency of carriers with *BRCA1* and *BRCA2* mutations, accounting for the high incidence of breast cancer in this population (Rubinstein, 2004).

Genetic screening for *BRCA1/2* in many populations has indicated that 10%–50% of breast cancer patients with a family history of breast and/or ovarian cancer had a *BRCA1* or *BRCA2* mutation, which suggested that *BRCA1/2* mutations could not completely explain familial aggregations of breast cancer and/or ovarian cancer. Mutational screenings have been conducted by focusing on genes functionally related to *BRCA1/2* and genes associated with other hereditary cancer syndromes (Stratton et al, 2008).

The National Comprehensive Cancer Network (NCCN) annually updates the cancer treatment guidelines, which are used worldwide. The NCCN members determine the guidelines for genetic and familial high-risk assessment for breast and ovarian cancers. In addition to *BRCA1/2*, nine genes (*ATM*, *CDH1*, *CHEK2*, *NBN*, *NF1*, *PALB2*, *PTEN*, *STK11*, and *TP53*) have been reported to increase the risk of breast cancer in the NCCN guidelines (NCCN Guidelines, 2018). *TP53* is the most famous tumor suppressor gene and has been categorized as one of the high-risk genes based on the guidelines. Loss of function mutations in *TP53* produce defects in the regulation of apoptosis, cell-cycle checkpoints, and DNA damage repair, which results in aberrant cell growth. Germline mutations in *TP53* are known to cause a hereditary cancer predisposition syndrome (Li-Fraumeni syndrome). Because *ATM* and *CHEK2* work as the main upstream activators of *TP53*, their mutations increase the risk of breast cancer. *PALB2* is also well known as a tumor suppressor gene. *PALB2* is involved in DNA repair in conjunction with *BRCA1/2*. Mutations in *PALB2* cause Fanconi anemia type N and increases the risk of breast cancer (Antoniou, 2014). Germline mutations of *CHD1*, *NBN*, *NF1*, *PTEN*, and *STK11* cause

hereditary cancer syndromes that confer increased risks for many types of cancers including breast cancer.

1.4. Genetic testing for *BRCA1/2* and multigene panel testing

Genetic testing for *BRCA1* and *BRCA2* is used as standard clinical practice for women with personal or family history of breast or ovarian cancers. The identification of pathogenic mutations in *BRCA1/2* allows prophylactic surgical intervention that can reduce the incidence, morbidity, and mortality of the mutation carriers. A Hollywood actress, Angelina Jolie, revealed that she underwent preventive double mastectomy and salpingo-oophorectomy to reduce the risk of breast and ovarian cancer development after becoming aware that she carried a pathogenic mutation of *BRCA1*. In addition, the U.S. Supreme Court unanimously nullified Myriad Genetics' patents on *BRCA1* and *BRCA2*. These events increased the public awareness of genetic testing for *BRCA1/2* as a preventive, proactive strategy. Genetic screening for *BRCA1/2* mutations in the general population has been recently proposed, which may result in the early identification of carriers with pathogenic germline mutations (King et al., 2014; Ahmadloo et al., 2017).

Next-generation sequencing (NGS) technologies have dramatically decreased the cost of DNA sequencing. Compared with conventional Sanger sequencing, the use of NGS technologies allows simultaneous determination of the sequences of multiple cancer predisposition genes, including *BRCA1/2*, in a time- and cost-effective manner. The advantage of multigene panel testing is increased sensitivity for detecting mutations in a patient that underlie a disease. At the same time, however, several potential disadvantages have been debated (Michael et al., 2014). First, there has been a controversy and lack of conclusive evidence on the associations between some gene mutations and cancer risks.

Multigene panel testing offered by commercial companies and academic institutions evaluate and report on several genes that are not fully characterized in terms of their breast cancer risks or their management options. Second, sufficient data is lacking for populations other than European-descent populations to distinguish pathogenic variants of genes from benign ones. When the effect of a genetic variant on a function is not known or there are insufficient data to definitively confirm that a variant confers the risk of developing breast cancer or another disease, the variant is classified as a variant of uncertain significance (VUS) (Chang et al., *in press*). The number of VUSs has remarkably increased as the number of genes included in a panel has increased, as discussed in the following section. Finally, there is a possibility that multigene panel testing provides unexpected gene mutations that are connected to another disease than the one the multigene panel testing was done for. Such incidental findings are major concern for the medical community especially when the unexpectedly identified mutations lead to disorders for which preventive strategies or treatments are not available.

1.5. Difficulties of VUS

There are several lines of evidence showing the clinical utility of multigene panel testing (LaDuca et al., 2014). However, there has been increasing concern regarding the interpretations of VUS identified by multigene panel testing. A large part of the missense variants in breast cancer-associated genes detected by multigene panel testing are classified as VUSs. VUSs were identified in 30%–40% patients who underwent multigene panel testing (Tung et al., 2015). VUS is a possible cause of confusion in interpretation among patients and physicians. It has been reported that prophylactic medical treatments were conducted in unaffected individuals who carried VUSs. The Evidence-based Network for

the Interpretation of Germline Mutant Allele (ENIGMA) consortium stated that a framework to improve standardized reporting and better genetic literacy regarding VUSs for families, patients, and providers was necessary (Eccles et al., 2015). Under these circumstances, the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) made a declaration that clinicians should not make medical decisions based on VUSs (Richards et al., 2015). One of the concerns on increased frequency of VUSs by multigene panel testing is that the results of variant classification are not consistent among clinical laboratories. To overcome this disadvantage, enormous efforts on the re-classification of VUSs have been put forth, but VUSs remain a challenging and difficult issue (Maxwell et al., 2016). Several approaches to improve the reproducibility and the objectivity of variant classifications have been proposed (Nykamp et al., 2017; Li et al., 2017).

1.6. Guideline for pathogenicity classification

In 2015, standards and guidelines for interpretation of variants were formulated by the ACMG/AMP (Richards et al., 2015). The 2015 ACMG/AMP guidelines recommend that variants are to be classified into five categories (pathogenic, likely pathogenic, VUS, likely benign, and benign) based on 28 criteria by assessing features of the variants such as genomic annotations, frequency in general populations, clinical information, computational predictions, functional and experimental data and public databases. These pathogenic criteria are weighted as very strong (PVS1), strong (PS1-4), moderate (PM1-6), or supporting (PP1-5), and benign criteria are weighted as stand-alone (BA1), strong (BS1-4), or supporting (BP1-6) (Table 1.2). For each variant, the criteria are selected based on the observed evidence, and then combined according to the scoring rules in Table 1.3 to select

the classification for the variant from the five-tier system. For example, a nonsense mutation of *BRCA1* in a breast cancer patient, which is not observed in any databases for variant frequency, fulfills criteria "PVS1" and "PM2" (Table 1.2). After combining "PVS1" and "PM2", this variant is classified into "pathogenic" according to the scoring rules (Table 1.3). A missense variant of *BRCA2* in a breast cancer patient whose allele frequency in the general population is 30% satisfies a criterion "BA1" (Table 1.2), which leads to the classification of "benign". If a variant does not fulfill criteria in Table 1.2, or the evidence for benign and pathogenic is conflicting, the variant is classified into VUS.

1.7. Definition of germline mutations

Germline or hereditary mutations are mutations that are inherited from a parent and are present in virtually every cell in the body of a person. The term "germline" is used because these mutations are present in the parent's germ cells (egg or sperm cells) and are transmitted to the person. A part of germline mutations predisposes individuals to a high risk of developing certain types of cancer.

Somatic or acquired mutations occur during a person's life and are present only in certain cells. These mutations can be caused by environmental exposures such as ultraviolet radiation from sunlight or by errors during DNA replications. Cancers develop as a result of the accumulation of somatic mutations (Vogelstein et al., 2013). For example, breast cancers arise from mammary epithelial cells that acquire several somatic mutations. In cancer studies, DNA changes in tumor cells are called as somatic mutations. Genetic analyses for tumor cells obtained from cancer patients have been intensively conducted (Lawrence et al., 2013).

There is a concern that the term "mutation" is ambiguous and misleads readers.

Therefore it is required to clearly distinguish between inherited germline mutations and acquired somatic mutations in cancer studies.

In my doctoral thesis, I focus on germline variants in cancer predisposition genes by using DNA samples extracted from peripheral blood of patients with breast cancer. As I explained, germline mutations are defined as mutations that are inherited from either parent and are present in each cell of a patient. Therefore, germline genetic testing can be evaluated by using DNA from any cells in the body of a patient except for tumor cells. Although DNA extracted from peripheral blood, saliva, buccal mucosa or skin biopsy is commonly used, peripheral blood is the most common source of DNA for germline genetic testing in humans (Lynch et al., 2015). It is because that peripheral blood samples are collected in routine clinical practice and provide adequate amount of high-quality DNA with lower contamination by foreign DNA such as bacteria, fungi, and food remnants. Additionally, there are 4,000,000-10,000,000 (on average 7,000,000) white blood cells per milliliter of blood from an adult individual. Therefore, somatic mutations in a small fraction of white blood cells is unlikely to affect the result of germline variant detection. However, I cannot completely exclude a possibility of clonal hematopoiesis that is a phenomenon where a single hematopoietic stem cell lineage contributes predominantly to the population of mature blood cells (Zink et al., 2017).

In addition, there is a concern that if somatic mutations on cancer-associated genes occur in an early stage of development, such somatic mutations might be present in blood-derived DNA samples and falsely detected as germline mutations. Here, I consider this issue. It is reported that somatic mutation rate in human skin cells was 2.66×10^{-9} mutations per bp per mitosis (Milholland et al., 2017). Though somatic mutation rates may differ according to cell types, I rely on this estimation because reliable estimates of somatic

mutation rates for other cell types are not available. According to the estimated somatic mutation rate, about two mutations occur in the diploid human genome per mitosis. Protein coding exons account for only 1%-2% of the human genome. Moreover, I analyze only 54 genes out of 20,000-25,000 protein coding genes (0.2%-0.25%) (genes analyzed in this study is described in Chapter 2). The target region is very small (0.002%-0.005% of the human genome). Accordingly, I infer that 0.00004-0.0001 mutations occur per mitosis. It is estimated that more than 1,000 somatic mitoses necessary to generate one mutation in the targeted genes. Next, I consider that detection level for such somatic mutations by NGS. It is reported that the detection limit of the percentage of reads of mutant alleles over total number of reads covering the mutation sites by typical NGS for germline variant calling is about 10% (Zink et al., 2017). It is thought that the proportion of a cell population with a specific somatic mutation is reflected as the mutant allele frequency of the corresponding mutation (Nik-Zainal et al., 2012a, Nik-Zainal et al., 2012b). For example, if all the sequenced cells have a heterozygous mutation, its mutant allele frequency is expected to be 0.5. If a half of the sequenced cells have a heterozygous mutation, the expected mutant allele frequency is 0.25. Here, I consider a heterozygous mutation because it is thought that the probability of acquiring mutation at the same position is extremely rare. If one fourth of the cell population has a heterozygous mutation, one eighth of the reads (12.5%) covering the mutation site are expected to have the mutant allele, which is close to the detection limit of a heterozygous variant. If small proportion of cells (<1/4) has somatic mutations, germline variant call methods cannot detect such mutations. Based on these inferences, I think the probability that mutations occur at important positions of the target genes during one or two cell divisions is very low and therefore does not largely affect the result of my study.

1.8. The objective of this doctoral thesis

The objective of my doctoral thesis is to evaluate the usefulness of genetic testing using multigene panels by using DNA samples from Chinese breast cancer patients. First, I assess whether the genetic risk factors of patients whose etiologies are not clarified by conventional genetic tests containing only *BRCA1/2* are accounted for by multigene panel testing including recently identified breast-cancer associated genes. Second, I will evaluate population-specific mutation prevalence, which may be important because the mutational spectrum is still poorly understood in populations other than European-descent populations. Third, I investigate clinical features of patients with mutations on breast-cancer associated genes to find useful findings for selecting appropriate treatments for patients according to mutation status.

Another purpose of my doctoral thesis is to improve the effectiveness of multigene panel testing by solving technical issues in the analytical workflow. The first issue is that most applications of NGS-based multigene panel testing target only SNVs and small INDELs. Then, I explore intermediate-sized INDELs to detect patients whose etiologies are not explained by focusing only on SNVs and small INDELs. The second and biggest problem in multigene panel testing is that the information about VUSs is not utilized for clinical decision making though 30%-40% of breast cancer patients have VUSs by multigene panel testing. I propose a hypothesis in which a small fraction of VUSs confer harmful effects on functions of genes and contribute to breast cancer risk. In order to prove my hypothesis, I formulate an analytical method to prioritize rare and harmful VUSs and to evaluate whether such VUSs are associated with the risk for developing breast cancer.

Table 1.1. Systemic treatments for breast cancer subtypes by St. Gallen International Expert Consensus 2011.

Subtypes	Hormone receptor	HER2 receptor	Ki67	Treatment
Luminal A	Positive	Negative	Low	Endocrine therapy
Luminal B (HER2 negative)	Positive	Negative	High	Endocrine and cytotoxic therapy
Luminal B (HER2 positive)	Positive	Positive	Any	Anti-HER2 and endocrine therapy
HER2 positive (non-liminal)	Absent	Positive	Any	Anti-HER2, endocrine and cytotoxic therapy
Triple negative	Absent	Negative	Any	Cytotoxic therapy

Table. 1.2. Schematic table for pathogenic or benign assertion. FH, family history; LOF, loss of function; MAF, minor allele frequency. (From Richards S et al., 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. 2015.)

	Benign		**Pathogenic**			
	Strong	**Supporting**	**Supporting**	**Moderate**	**Strong**	**Very strong**
Population data	MAF is too high for disorder **BA1**/**BS1** OR observation in controls inconsistent with disease penetrance **BS2**			Absent in population databases **PM2**	Prevalence in affected statistically increased over controls **PS4**	
Computational data		Multiple lines of computational evidence suggest no impact on gene /gene product **BP4** Missense in gene where only truncating cause disease **BP1** Silent variant with non predicted splice impact **BP7** In-frame indels in repeat w/out known function **BP3**	Multiple lines of computational evidence support a deleterious effect on the gene /gene product **PP3**	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before **PM5** Protein length changing variant **PM4**	Same amino acid change as an established pathogenic variant **PS1**	Predicted null variant in a gene where LOF is a known mechanism of disease **PVS1**
Functional data	Well-established functional studies show no deleterious effect **BS3**		Missense in gene with low rate of benign missense variants and path, missenses common **PP2**	Mutational hot spot or well-studied functional domain without benign variation **PM1**	Well-established functional studies show a deleterious effect **PS3**	
Segregation data	Nonsegregation with disease **BS4**		Cosegregation with disease in multiple affected family members **PP1**	Increased segregation data →		
De novo data				De novo (without paternity & maternity confirmed) **PM6**	De novo (paternity & maternity confirmed) **PS2**	
Allelic data		Observed in *trans* with a dominant variant **BP2** Observed in *cis* with a pathogenic variant **BP2**		For recessive disorders, detected in trans with a pathogenic variant **PM3**		
Other database		Reputable source w/out shared data = benign **BP6**	Reputable source = pathogenic **PP5**			
Other data		Found in case with an alternate cause **BP5**	Patient's phenotype or FH highly specific for gene **PP4**			

Table 1.3. Decision rules for pathogenicity of variant. (From Richards S et al., 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. 2015.)

Pathogenic	(i) 1 Very strong (PVS1) AND (a) ≥1 Strong (PS1-PS4) OR (b) ≥2 Moderate (PM1-PM6) OR (c) 1 Moderate (PM1-PM6) and 1 supporting (PP1-PP5) OR (d) ≥2 Supporting (PP1-PP5) (ii) ≥2 Strong (PS1-PS4) OR (iii) 1 Strong (PS1-PS4) AND (a) ≥3 Moderate (PM1-PM6) OR (b) 2 Moderate (PM1-PM6) AND ≥2 Supporting (PP1-PP5) OR (c) 1 Moderate (PM1-PM6) AND ≥4 supporting (PP1-PP5)
Likely pathogenic	(i) 1 Very strong (PVS1) AND 1 moderate (PM1-PM6) OR (ii) 1 Strong (PS1-PS4) AND 1-2 moderate (PM1-PM6) OR (iii) 1 Strong (PS1-PS4) AND ≥2 supporting (PP1-PP5) OR (iv) ≥3 Moderate (PM1-PM6) OR (v) 2 Moderate (PM1-PM6) AND ≥2 supporting (PP1-PP5) OR (vi) 1 Moderate (PM1-PM6) AND ≥4 supporting (PP1-PP5)
Benign	(i) 1 Stand-alone (BA1) OR (ii) ≥2 Strong (BS1-BS4)
Likely benign	(i) 1 Strong (BS1-BS4) and 1 supporting (BP1-BP7) OR (ii) ≥2 Supporting (BP1-BP7)
Uncertain significance	(i) Other criteria shown above are not met OR (ii) The criteria for benign and pathogenic are contradictory

Chapter 2: Identification of germline mutations on 54 breast cancer susceptibility genes

The contents of this chapter are also described in the below paper:

Germline mutations of multiple breast cancer-related genes are differentially associated with triple-negative breast cancers and prognostic factors. Hata C et al. (2020). Journal of Human Genetics. https://doi.org/10.1038/s10038-020-0729-7

2.1. Introduction

Breast cancer is the most common noncutaneous malignancy among women (DeSantis et al., 2017). Approximately 10% of cases involve germline mutations of genes involved in DNA repair pathways (Stratton et al., 2008). The tumor suppressor genes *BRCA1/2* are essential for the repair of DNA double-strand breaks through the homologous recombination pathway (Prakash et al., 2015). Pathogenic germline mutations of *BRCA1/2* confer strong risks of breast cancer development (King, 2014). The lifetime risk of breast cancer can increase to 80% in mutation carriers of these genes (King et al., 2003). Genetic testing for *BRCA1/2* has now become a common practice worldwide (Wallace, 2016).

Recently, evidence of associations between breast cancer and genes other than *BRCA1/2* has been accumulated (Easton et al., 2015). Germline mutations in genes causing familial cancers, such as *TP53* and *PTEN*, confer high risks of breast cancer (Malkin et al., 1990; Liaw et al., 1997). *PALB2* mutation carriers had a >5-fold increased risk of breast cancer (Rahman et al., 2007; Antoniou et al., 2014). Furthermore, mutations of other DNA repair genes such as *CHEK2* and *ATM* have been reported to be associated with a 2- to 4-fold increased risk of breast cancer (Meijers-Heijboer et al., 2002; Weischer et al., 2008; Renwick et al., 2006). These findings suggest the clinical significance of performing multigene panel testing of *BRCA1/2* and the other breast cancer-related genes (Easton et al., 2015). The development of NGS technologies has rendered multigene panel testing feasible for the assessment of hereditary cancer risk (Goodwin et al., 2016).

Breast cancer is a heterogeneous disease at the genomic, molecular, and cellular levels, with varying histology, treatment response, and patient survival outcomes. Clinically, breast cancer is primarily categorized based on the expressions of ER, PR and HER2, which guide clinical decisions such as treatment protocol (Goldhirsch et al., 2011).

Breast cancer patients with *BRCA* mutations often occur in younger women. The tumors in *BRCA1/2* mutation carriers are more frequently high grade and likely to be ER-, PR-, and HER2-negative, i.e., TNBC. In particular, *BRCA1*-mutated breast cancers show stronger associations with a basal-like phenotype and TNBC (Bianchini et al., 2016; Lord et al., 2016). Although the characteristics of *BRCA1/2*-mutated breast cancers have been well investigated, the clinical features of tumors with mutations in other cancer predisposition genes remain unclear. Multigene panel testing can facilitate the identification of associations between the mutation status, clinicopathologic features, and outcomes of this heterogeneous disease.

The objective of this chapter is to demonstrate the usefulness of multigene panel testing for breast cancer. For the purpose, I evaluated the possibility of identifying genetic risk factors for patients whose etiologies could not be determined through conventional genetic test including only *BRCA1/2* by adding recently identified breast cancer associated genes. Additionally, I examined mutation prevalence in Chinese breast cancer patients in which the mutation landscape was not still fully understood. Finally, I assessed whether the mutation status was associated with clinicopathologic features and outcomes to provide useful information to help guide treatment of breast cancer.

2.2. Materials and Methods

2.2.1. Study patients

In total, I analyzed 583 Han Chinese patients with breast cancer between December 2016 and September 2017 at the First Affiliated Hospital of Chongqing Medical University and Affiliated Cancer Hospital & Institute of Guangzhou Medical University; all patients provided informed consent for participation in the study. All procedures were performed in

accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The Ethics Committees of the First Affiliated Hospital of Chongqing Medical University, the Affiliated Cancer Hospital & Institute of Guangzhou Medical University, and the National Institute of Genetics, approved the study protocols.

Patients' mean age at diagnosis was 49.1 years (standard deviation, 9.2 years). The baseline characteristics of the 583 patients are shown in Table 2.1.

2.2.2. Clinicopathologic characteristics

All clinicopathologic data were retrospectively collected from medical records. Pathological factors, including tumor grade; tumor histology; positive axillary lymph nodes; and ER, PR, HER2, p53, cytokeratin 5/6 (CK5/6), E-cadherin, and Ki67 status, were obtained. All pathological findings were determined by the expert panel of the Pathology Department of First Affiliated Hospital of Chongqing Medical University and Affiliated Cancer Hospital & Institute in Guangzhou Medical University, who adopted uniform diagnostic criteria as described in collaborator's previous study (Zheng et al., 2018). Briefly, the cutoff values for ER and PR positivity were set to >1% of positive tumor cells with nuclear staining. HER2 status was evaluated by immunohistochemistry (IHC) and fluorescence *in situ* hybridization (FISH). Tumors with no (0) or weak (1+) staining were considered to be HER2-negative, whereas those with strong (3+) staining were considered to be HER2-positive. For tumors with IHC staining of 2+, HER2 positivity was determined by FISH. Histological grade was assessed by the Nottingham grading system (Grades I, II, and III).

A family history of cancer was obtained from telephone or in-person interviews.

The proportions of the patients with a family history of any type of cancer, breast cancer, and breast and/or ovarian cancers were 22.0%, 7.2%, and 6.5%, respectively. Summary statistics for the clinical variables are shown in Table 2.1.

2.2.3. DNA samples

DNA samples were extracted from each patient's peripheral blood using a FavorPrepTM Tissue Genomic DNA Extraction Mini Kit (Favorgen, Ping-Tung, Taiwan) according to the manufacturer's instructions. The extracted DNA was quantified using a Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA) on a FilterMax F5 Multi-Mode Microplate Reader (Molecular Devices, Sunnyvale, CA, USA).

2.2.4. Target gene sequencing of 54 breast cancer predisposition genes

In total, 54 predisposition genes were selected based on previous studies of multigene panel testing for hereditary breast and/or ovarian cancer (Walsh et al., 2010; Castéra et al., 2014; Chong et al., 2014; LaDuca et al., 2014; Couch et al., 2015; Cybulski et al., 2015; Desmond et al., 2015; Easton et al., 2015; Judkins et al., 2015; Lincoln et al., 2015; Maxwell et al., 2015; Minion et al., 2015; Tung et al., 2015; Norquist et al., 2016; Tung et al., 2016) and are listed in Table 2.2.

Target sequencing of the 54 genes was performed using the pre-capture pooling method described in previous studies (Ahmadloo et al., 2017; Suda et al., 2018). In brief, 20-ng DNA was simultaneously fragmented and adapter-ligated with the SureSelect QXT Library Prep Kit (Agilent Technologies). The fragmented libraries with distinct indexed adapters were pooled in equimolar amounts, followed by target enrichment using the

SeqCap EZ Choice System (Roche Diagnostics). The target regions were designed to cover the exons and 50 bp of flanking intronic regions of all 54 genes based on annotations from NCBI's Reference Sequence Database (RefSeq) in the UCSC Genome Browser (http://genome.ucsc.edu/index.html). A DNA probe set complementary to the target regions was designed by NimbleDesign (http://design.nimblegen.com). The libraries were sequenced on an Illumina HiSeq 2500 platform performed in the rapid run mode with a 2 × 100-bp paired-end protocol.

2.2.5. NGS data processing and variant calling

Illumina adapter-containing sequences reads were trimmed using a Trimmomatic version 0.36 (Bolger et al., 2014). After the quality control step for excluding or trimming low quality sequences, the sequence reads were aligned to the human reference genome (hg19) using the Burroughs–Wheeler aligner (BWA) version 0.7.15 (Li et al., 2010). Aligned read information was converted into compressed binary form (BAM format) and sorted based on genomic coordinates using SAMtools version 1.4.1 (Li et al., 2009). The aligned reads were processed for PCR-duplicated and erroneous read removal using Picard tools version 1.8.0_131 (https://broadinstitute.github.io/picard/). Local realignment and base quality recalibration were implemented using GATK version 3.4-46 (McKenna et al., 2010; DePristo et al., 2011). Coverage and average depth over the target regions were calculated with the CallableLoci and DepthOfCoverage tools in the GATK package, respectively. SNVs and INDELs were detected using the HaplotypeCaller tool in the GATK package.

2.2.6. Functional annotation of identified variants

Functional annotation was implemented using ANNOVAR (version released 29 September 2017) (Wang et al., 2010). Estimates of variant frequencies in general populations were based on publicly available databases provided by two whole-genome and exome sequencing projects: the 1000 Genomes Project (1KG) and the Exome Aggregation Consortium (ExAC) (1000 Genomes Project Consortium, 2015; Lek et al., 2016). The potential effects of the variants were estimated by four prediction and conservation tools (SIFT, PolyPhen2, GERP++, and REVEL) using ANNOVAR.

2.2.7. Variant classification

The identified variants were classified according to pathogenicity using the following procedure. First, a variant was classified as "benign" if its frequency was >1.0% in any 1KG or ExAC population. Second, a variant was classified as "pathogenic" if it resulted in loss of function (i.e., nonsense SNVs, frameshift INDELs, and SNVs and INDELs at canonical ± 1 or 2 splice sites or the initiation codon). Third, missense variants with previously established pathogenic or benign effects were explored based on the ClinVar database version 20180521 (Landrum et al., 2016). Fourth, variants were classified as 'Conflicting interpretations of pathogenicity' if multiple submitters reported discordant interpretations for the same variants in the ClinVar. Finally, variants with 'uncertain significance' were classified as 'VUS'. Variants without reliable information in the ClinVar were classified as 'NA'.

I filtered to exclude variants whose frequencies in the general populations were greater than 1.0%. I explain the reasoning of this threshold. Therefore, approximately 8%-10% women develop breast cancer during their lifetime (Jemal et al., 2010).

Given that allele "a" is a pathogenic variant whose allele frequency is 1% in a

population and allele “A” is a reference allele, the frequency of each genotype is estimated according to Hardy-Weinberg equilibrium, are as follows:

AA=0.9801、Aa=0.0198、aa=0.0001

Assuming an autosomal dominant inheritance model for the pathogenic variant like *BRCA1/2*, it is estimated that approximately 2% of the population are at high risk for hereditary breast cancer due to allele “a” (genotypes: Aa + aa). Assuming that there is a pathogenic mutation with an allele frequency of 1%, about 2% of individuals from the population would develop breast cancer due to the single pathogenic variant. Considering that the prevalence of hereditary breast cancer is 0.8%-2.0% as described above, a single genetic mutation would explain almost all the hereditary breast cancer patients. It is well known, however, that a large number of pathogenic mutations on cancer-associated genes such as *BRCA1/2* are involved in breast cancer, the prevalence of hereditary breast cancer becomes unrealistically high. From these considerations, it is reasonable to consider that the frequency of pathogenic variation of hereditary breast cancer is very low; consistent with this idea, the actually reported pathogenic mutations are very low in their frequencies. Additionally, it is thought that up to 20% of patients with breast cancer may have strong genetic risks (Cobain et al., 2016). These findings imply that up to 2% of people are at high risk of developing breast cancer due to strong genetic factors. Therefore, allele frequency of 1% seems to be reasonable as the upper limit when searching for germline pathogenic mutations in breast cancer.

2.2.8. Statistical analyses

To evaluate associations between mutation status and clinical variables, I classified the patients with pathogenic germline variants into six groups: i) all mutation carriers; ii)

BRCA1 mutation carriers; iii) *BRCA2* mutation carriers; iv) non-*BRCA1/2* HR (homologous recombination) gene mutation carriers; v) MMR (mismatch repair) gene mutation carriers; and vi) other mutation carriers. The assignment of the 54 breast cancer predisposition genes to these groups is shown in Table 2.2. I compared the clinical variables in these six groups between mutation carriers and noncarriers.

I investigated associations between genotypes and results of IHC and FISH assays of seven biomarkers: ER, PR, HER2, p53, CK5/6, E-cadherin, and Ki67 (Table 2.1). TNBC was determined based on ER, PR, and HER2 profiles. Furthermore, eight prognostic factors (age at diagnosis; family history of any type of cancer, breast cancer, and breast and/or ovarian cancers; histologic grade; preoperative axillary lymph nodal status; bone metastasis; and vascular invasion) were analyzed (Table 2.1).

For dichotomous variables, I used Fisher's exact test. I estimated the odds ratio (OR) and 95% confidence interval (95% CI). The Wilcoxon–Mann–Whitney test was used to evaluate differences in quantitative variables. When subgroup analysis was conducted, results from each of the strata were combined by a meta-analysis based on Fisher's combined probability method. I conducted all standard statistical tests with the R program (http://www.r-project.org). The threshold of statistical significance was set at $P < 0.05$.

2.3. Results

2.3.1. Identification of pathogenic variants

The average depth for the target regions was 117.6, and the mean proportion of the targeted regions covered by at least 20 reads was 99.3%, supporting confident variant detection. I detected 3,742 variants comprising 3,250 SNVs and 492 INDELs. According to the annotations, I filtered and prioritized genetic variants (Figure 2.1). First, I retained

variants whose frequencies were <1% in any 1KG and ExAC population (2,442 SNVs and 369 INDELs). Second, I aimed to identify protein-truncating variants, namely nonsense mutations, splice site SNVs within 2 bp from the exon–intron boundary, and frameshift INDELs. Finally, I retrieved missense and intronic SNVs that were recorded as "pathogenic" in ClinVar. Ultimately I detected 78 pathogenic mutations comprising 43 SNVs (32 nonsense SNVs, five splice site SNVs, five missense SNVs, and one intronic SNV) and 35 INDELs (33 frameshift and two splice site INDELs) in *BRCA2*, *BRCA1*, *MSH6*, *PTEN*, *PALB2*, *GEN1*, *RAD51D*, *MUTYH*, *CHEK2*, *BLM*, *PMS2*, *PMS1*, *MSH2*, *NF1*, *AXIN1*, *FANCM*, *MET*, *MRE11A*, *BRIP1*, *RECQL*, and *TP53* (Table 2.3). Furthermore, I detected a intermediated-sized deletion-insertion in a translational variant of *PTEN* described in detail in Chapter 3 and added to subsequent analyses. Overall, 84 patients harbored these 79 pathogenic mutations.

2.3.2. Annotation of the identified variants by ClinVar

Of the 37 nonsense and splice site SNVs, 22 were not registered in ClinVar and were considered to be novel pathogenic variants (Figure 2.1A, Table 2.4). Similarly, 83.3% of the frameshift INDELs were novel pathogenic variants (Figures 2.1B and 2.1D, and Table 2.5). These results suggest that a substantial proportion of the variants of pathogenicity for breast cancer remain undiscovered, particularly in individuals of non-European descent, which requires further exploration of pathogenic variants.

Classification of rare missense SNVs has been challenging. I retrieved known pathogenic variants based on ClinVar (Figure 2.1C). Only 0.8% (5/617) of the rare missense SNVs had already been reported as "pathogenic" in ClinVar (Table 2.6). The other missense SNVs were classified as "Benign" (5.2%), "conflict" (12.0%), and

"uncertain significance" (28.5%); however, 53.5% were not registered ("NA") in ClinVar. Overall, I was unable to classify 94.0% of the rare missense SNVs (Figure 2.1C).

I detected 1,788 rare SNVs in untranslated, intronic, and intergenic regions. Only one SNV located on intron 4 of *BRCA1* (c.213-15A>G) was "pathogenic" according to the ClinVar classification (Table 2.6). According to *in silico* analysis, this SNV was predicted to use a cryptic site at 59 nucleotides upstream of the natural splice site, resulting in a severe splicing defect (Houdayer et al., 2012).

I examined the clinical condition field for the pathogenic variants retrieved from ClinVar searches. Two pathogenic mutations on *BRCA1* (c.213-15A>G and c.212G>A) and one pathogenic mutation on *BRCA2* (c.8243G>A) were registered from breast-ovarian cancer studies. Two pathogenic mutations in *TP53* (c.743G>A and c.542G>A) and one pathogenic mutation in *MSH6* (c.3226C>T) are registered from hereditary cancer-predisposing syndrome studies. This result suggests that the pathogenic variants retrieved from ClinVar are linked with high risk of breast cancer.

2.3.3. Proportion of patients with pathogenic variants

The identified pathogenic mutations accounted for 14.4% of the breast cancer cases (Table 2.3); of these 7.7% of the patients harbored *BRCA1/2* mutations (45/583), and the remaining 6.5% of the patients carried pathogenic mutations in 19 genes (38/583), suggesting the usefulness of my multigene panel testing.

I detected a mutation in *BRCA2* (c.1399A>T; p.K467X) that was previously found to be common in Chinese patients with breast cancer (Kwong et al., 2016). Four pathogenic SNVs were shared in two patients: *BRCA1* p.R1443X, *BRCA2* p.Q1037X, *BRCA2* p.Q1129X, and *PALB2* p.Q251X. Furthermore, four frameshift INDELs were shared in two

patients: *BRCA1* p.N1355fs, *BRCA2* p.A1689fs, *GEN1* p.I643fs, and *PTEN* p.R52fs. Two patients were identified with double pathogenic mutations in different genes: i) *AXIN1* p.R841X and *BRCA2* p.G2748D and ii) *MSH2* p.R929X and *FANCM* p.E1903X. One patient exhibited two mutations in *GEN1* (p.I643fs and p.K833fs).

2.3.4. Association between pathogenic variants and clinical variables

2.3.4.1. *Biomarkers*

The results from association analyses of all mutation carriers and their specific groups with biomarkers are shown in Table 2.7 and Tables 2.8 and 2.9, respectively.

Germline mutations carriers in any of the 54 genes were negatively associated with HER2-overexpressing breast cancers ($P = 8.2 \times 10^{-3}$; OR, 0.46; 95% CI, 0.24–0.84). This association was particularly strong in *BRCA1* mutation carriers ($P = 0.013$; OR, 0.19; 95% CI, 0.021–0.81). *BRCA1* mutation carriers were less likely to develop ER-positive breast cancers ($P = 1.1 \times 10^{-4}$; OR, 0.14; 95% CI, 0.033–0.44) and had a significantly lower frequency of PR-positive breast cancers ($P = 8.2 \times 10^{-3}$; OR, 0.23; 95% CI, 0.054–0.73). The association of non-*BRCA1/2* HR mutation carriers with PR expression levels was significant ($P = 0.040$; OR, 0.36; 95% CI, 0.11–1.03). Consequently, a strong association was observed between *BRCA1* mutation carriers and TNBC ($P = 6.0 \times 10^{-9}$; OR, 21.58; 95% CI, 6.49–93.18); however, the association between *BRCA2* carriers and TNBC was not statistically significant ($P = 0.33$; OR, 1.66; 95% CI, 0.69–6.78). Furthermore, non-*BRCA1/2* HR mutation carriers showed a significant association with TNBC ($P = 0.034$; OR, 3.11; 95% CI, 0.92–9.38). Although the number of MMR mutation carriers was small, they all had ER-positive breast cancers ($P = 0.021$).

A significant association was observed between *BRCA1* mutation carriers and

CK5/6-positive breast cancers (P = 0.015; OR, 3.84; 95% CI, 1.15–11.46), confirming previous findings (Foulkes et al., 2003; Laakso et al., 2005).

The breast cancers in all germline mutation carriers were less likely to present E-cadherin-positive cells ($P = 7.8 \times 10^{-3}$; OR, 0.16; 95% CI, 0.036–0.72). However, the association was stronger in *BRCA2* mutation carriers ($P = 4.6 \times 10^{-3}$; OR, 0.075; 95% CI, 0.014–0.51).

BRCA1 mutation carriers showed a tendency of a higher proportion of Ki67-positive cells than noncarriers ($P = 1.6 \times 10^{-6}$; Figure 2.2A). In order to evaluate whether higher Ki67 values in *BRCA1* mutation carriers were a consequence of overrepresentation of TNBC in the mutation carriers, I conducted a subgroup analysis stratified by ER status. Ki67 values were higher for *BRCA1* mutation carriers than those for non-carriers in both ER positive (P = 0.089) and negative ($P = 4.9 \times 10^{-8}$) subgroups (Figure 2.3A). Meta-analysis combining the results from these two subgroups by Fisher's method reinforced significant increase of Ki67 expressions in *BRCA1* mutation carriers ($P = 8.9 \times 10^{-9}$).

Similarly, I detected weak correlations in the number of PR-positive cells between non-*BRCA1/2* HR mutation carriers and noncarriers (P = 0.072; Figure 2.2C) and ER-positive cells between MMR mutation carriers and noncarriers (P = 0.021; Figure 2.2D). Furthermore, *BRCA1* mutation carriers showed remarkably lower ER and PR expression levels than noncarriers ($P = 1.1 \times 10^{-3}$ and P = 0.011, respectively).

2.3.4.2. *Prognostic factors*

Statistically significant difference in the age at diagnosis was observed between *BRCA1* mutation carriers and noncarriers ($P = 7.5 \times 10^{-4}$). On average, patients with *BRCA1* mutations developed breast cancer 6.4 years earlier than non-carrier patients

(Figure 2.2E). To assess whether the association between *BRCA1* mutations and age at diagnosis was independent from molecular subtypes based on receptor expression profiles, I conducted a subgroup analysis stratified by ER status. Notably, the ages at diagnosis for *BRCA1* mutation carriers were younger than those for non-carriers in both ER positive (P = 0.058) and negative (P = 0.02) subgroups (meta-analysis, $P = 9.0 \times 10^{-3}$) (Figure 2.3B).

The results of association analyses of all mutation carriers and specific group of mutation carriers with prognostic factors are shown in Table 2.10 and Tables 2.11 and 2.12, respectively.

BRCA1 mutations were strongly associated with a family history of any type of cancer ($P = 3.6 \times 10^{-3}$; OR, 4.55; 95% CI, 1.46–15.62), breast cancer ($P = 6.3 \times 10^{-5}$; OR, 11.28; 95% CI, 3.29–37.23), and breast and/or ovarian cancers ($P = 1.2 \times 10^{-5}$; OR, 12.89; 95% CI, 3.91–42.62). Additionally, *BRCA2* mutation carriers were overrepresented by a family history of breast cancer (P = 0.023; OR, 3.82; 95% CI, 1.03–11.85) and breast and/or ovarian cancers (P = 0.015; OR, 4.30; 95% CI, 1.15–13.45).

Tumors in patients with germline pathogenic mutations had a high (grade III vs. I/II) histological grade (P = 0.070; OR, 1.74; 95% CI, 0.90–3.26), particularly in those with *BRCA1* mutations ($P = 6.5 \times 10^{-3}$; OR, 4.37; 95% CI, 1.34–14.64).

Axillary lymph node metastasis was more frequent in all pathogenic mutation carriers ($P = 5.3\times 10^{-3}$; OR, 2.14; 95% CI, 1.23–3.68), *BRCA2* mutation carriers (P = 0.074; OR, 2.25; 95% CI, 0.84–5.67), and non-*BRCA1/2* HR mutation carriers (P = 0.082; OR, 2.38; 95% CI, 0.76–6.93). Notably, bone metastases were highly enriched in *BRCA2* carriers (P = 0.020; OR, 14.25; 95% CI, 1.14–130.91).

2.4. Discussion

2.4.1. Proportion of breast cancer patients with pathogenic variants

I detected 79 pathogenic mutations in 21 cancer-related genes by using NGS of a multigene panel including 54 cancer-related genes in 583 unselected Chinese patients with breast cancer. The identified pathogenic mutations accounted for 14.4% of the breast cancer cases; of these 7.7% of the patients harbored *BRCA1/2* mutations (45/583), and the remaining 6.5% of the patients carried pathogenic mutations in 19 genes (38/583), suggesting the usefulness of my multigene panel testing. When I considered genes with more concrete evidence of associations with breast cancer that were listed in NCCN guidelines as having "increased risk of breast cancer" (NCCN Guidelines, 2018), the frequency of mutation carriers on *BRCA1*, *BRCA2*, *PALB2*, *CHEK2*, *TP53*, *PTEN*, and *NF1* was 10.6% in total.

The result demonstrated that genetic risk factors of the patients whose etiologies were not clarified by conventional *BRCA1/2* testing were successfully explained by multigene panel testing containing recently identified breast cancer predisposing genes, supporting the utility of the multigene panel.

2.4.2. Population-specific mutation prevalence and landscape

The findings in my study indicate that the frequency distribution of pathogenic mutations differs substantially between populations in recently identified breast cancer predisposition genes. In Chinese breast cancer patients, *PALB2* was the most commonly (1.2%) mutated gene other than *BRCA1/2*, and the frequencies of pathogenic mutations in *CHEK2* (0.51%) and *ATM* (0%) were low, which were consistent with two recent large-scale studies of East Asian populations (Sun et al., 2017; Momozawa et al., 2018). On the other hand, the most frequently mutated genes other than *BRCA1/2* were *CHEK2* (2.82%),

ATM (1.06%), and *PALB2* (0.87%) in European-descent populations (Couch et al., 2017). These results suggest that the contribution of mutations of *CHEK2* and *ATM* to breast cancers is lower in East Asian populations than that in European-descent populations. This may be because several founder mutations of *CHEK2* and *ATM* are prevalent in European-descent populations (Bogdanova et al., 2009; Cybulski et al. 2011).

It has been well described that the frequencies of *BRCA1/2* mutations also differ by population, where *BRCA1/2* mutations were most prevalent (11.7%) in reports regarding Ashkenazi Jewish patients (Warner et al., 1999; Frey et al., 2018) compared to other populations (up to 9.8% of unselected Asian patients including my study [Kwong et al., 2016; Sun et al., 2017]) and up to 6.0% of unselected European patients [Antoniou et al., 2002]).

Furthermore, I showed that the majority of identified pathogenic mutations were novel. The pathogenic mutations identified in my study were not largely overlapped with those in a large-scale Chinese breast cancer study (Sun et al., 2017). There is a possibility that the difference in the observed mutations between the two studies reflect regional heterogeneity in Han Chinese population. While Sun et al. collected samples in Beijing, the samples in my study were collected at Chongqing and Guangzhou. Recently, it is shown that Han Chinese population is significantly diversified across regions in an isolation-by-distance manner (Chiang et al., 2018). Taken together, the results in my study indicate that regional subpopulations represent distinct mutational landscapes of breast cancer predisposition genes even in the same country.

The findings in my study suggest that mutation screening by using multigene panel testing is important particularly for non-European populations, in which the mutation prevalence and landscape have not been fully understood.

2.4.3. Association between pathogenic variants and clinical variables

Notably, I identified several novel associations between pathogenic variants and clinical variables. I detected an association between *BRCA2* mutation carriers and E-cadherin negative breast cancers. Additional analyses including subtypes of breast cancer revealed that patients with invasive lobular carcinoma were likely to harbor *BRCA2* mutations (P = 0.049; OR, 7.25; 95% CI, 0.68–43.7) and were represented by a lower E-cadherin expression ($P = 7.3 \times 10^{-8}$; OR, 0.008; 95% CI, 0.0012–0.048). Invasive lobular carcinoma was previously reported to show a lower expression level of E-cadherin (Qureshi et al., 2006). Overall, these results suggest that *BRCA2* mutation carriers are likely to develop invasive lobular carcinomas characterized by a lower of E-cadherin expression. Furthermore, patients who carry *BRCA2* mutations have an association with bone metastasis and lymph node involvement, suggesting that *BRCA2* mutations affect metastatic phenotype.

I showed that non-*BRCA1/2* HR mutation carriers were associated with TNBC. In particular, four of seven patients with *PALB2* mutations had TNBC (P = 0.01; OR, 8.27; 95% CI, 1.4–58.0). Unfortunately, while I was preparing this doctoral thesis, the association between *PALB2* mutations and TNBC was reported in European-descent population (Shimelis et al., 2018). The finding from independent two studies corroborates the strength of evidence that can be translated to clinical application where breast cancer patients with *PALB2* mutations are targeted to the administration of PARP inhibitors. Additionally, TNBC was highly enriched in *MUTYH* mutation carriers (2/4), although the association did not achieve the statistical significance threshold because of the limited number of patients with breast cancer with *MUTYH* mutations in my study (P = 0.097; OR, 6.21; 95% CI, 0.44–87.5). To the best of my knowledge, this is the first study to suggest an

association between *MUTYH* mutation and TNBC. A large-scale study is required to verify this association.

Additionally, I confirmed previously reported associations that were mostly related to *BRCA1/2* mutations. The results of the association analyses between *BRCA1/2* mutation status and clinical parameters showed distinct characteristics of breast cancers with *BRCA1* and *BRCA2* mutations when both *BRCA1 and BRCA2* mutations were strongly associated with a family history of breast and/or ovarian cancers. On average, *BRCA1* mutation carriers developed breast cancer 6.4 years earlier than non-carrier patients. *BRCA1* mutation carriers were strongly predisposed to TNBC; however, *BRCA2* mutation carriers were not (Stevens et al., 2013). Tumors in the patients with *BRCA1* mutations had a high histological grade and showed a higher proportion of Ki67-positive cells. The results from the subgroup analyses stratified by ER status suggest that *BRCA1* mutation carriers are likely to develop breast cancers at younger ages and with higher Ki67 levels after considering the differences in molecular subtypes. *BRCA1* mutation carriers were associated with positivity for CK5/6 (markers for myoepithelial and luminal epithelial cells) suggesting that *BRCA1*-mutated cases are likely to develop basal-like ductal invasive carcinomas (Foulkes et al., 2003; Lakhani et al., 2005).

Overall, these results demonstrated differential associations between mutations of *BRCA1/2* and other breast cancer predisposition genes with biomarkers and prognostic factors, further corroborating the clinical usefulness of multigene panel testing. On the other hand, two issues were highlighted: 1) large proportion of patients with unclear etiology after assessing the 54 genes and 2) extremely large number of VUS detected as a disadvantage of using multigene panel. I will tackle these two issues in the subsequent two chapters.

Table 2.1. Clinical information of 583 patients.

Variables			n	%
Subject characteristics				
	Sex, n (%)	Female	583	100.0
	Mean (SD) age, y	49.1	(±9.2)	
	Ethnicity, n (%)	Han Chinese	583	100.0
	Family history, n (%)	Yes	128	26.6
		No	353	73.4
	Family history of HBOC, n (%)	Yes	42	8.7
		No	439	91.3
	Family history of BC, n (%)	Yes	38	7.9
		No	443	92.1
Biomarkers				
	HER2, n (%)	Positive	186	37.7
		Negative	307	62.3
	ER, n (%)	Positive	370	64.7
		Negative	202	35.3
	PR, n (%)	Positive	303	53.0
		Negative	269	47.0
	TNBC, n (%)	Positive	85	17.3
		Negative	407	82.7
	p53, n (%)	Positive	412	73.0
		Negative	152	27.0
	CK5/6, n (%)	Positive	58	10.8
		Negative	478	89.2
	E-cadherin, n (%)	Positive	550	98.2
		Negative	10	1.8
Clinical variables				
	Grade, n (%)	I	19	4.3
		II	323	73.6
		III	97	22.1
	Vascular invatsion, n (%)	Positive	31	5.7
		Negative	517	94.3
	Axially lymph node, n (%)	Positive	129	23.2
		Negative	428	76.8
	Bone metastasis, n (%)	Positive	5	0.9
		Negative	553	99.1

Table 2.2. The 54 breast cancer predisposition genes analyzed in this study.

Group	Gene name	Gene description	Associated diseases
BRCA1/2	*BRCA1*	BREAST CANCER 1 GENE	Fanconi anemia; Breast-ovarian cancer; Pancreatic cancer
BRCA1/2	*BRCA2*	BRCA2 GENE	Fanconi anemia; Wilms tumor; Breast cancer; Breast-ovarian cancer; Glioblastoma; Medulloblastoma; Pancreatic cancer; Prostate cancer
MMR	*EPCAM*	EPITHELIAL CELLULAR ADHESION MOLECULE	Colorectal cancer; Diarrhea
MMR	*MLH1*	MutL, E. COLI, HOMOLOG OF, 1	Colorectal cancer; Mismatch repair cancer syndrome; Muir-Torre syndrome
MMR	*MSH2*	MutS, E. COLI, HOMOLOG OF, 2	Colorectal cancer; Mismatch repair cancer syndrome; Muir-Torre syndrome
MMR	*MSH6*	MutS, E. COLI, HOMOLOG OF, 6	Colorectal cancer; Endometrial cancer; Mismatch repair cancer syndrome
MMR	*PMS1*	POSTMEIOTIC SEGREGATION INCREASED, S. CEREVISIAE, 1	Breast cancer
MMR	*PMS2*	POSTMEIOTIC SEGREGATION INCREASED, S. CEREVISIAE, 2	Colorectal cancer; Mismatch repair cancer syndrome
non-BRCA1/2 HR	*AKT1*	V-AKT MURINE THYMOMA VIRAL ONCOGENE HOMOLOG 1	Cowden syndrome
non-BRCA1/2 HR	*ATM*	ATAXIA-TELANGIECTASIA MUTATED GENE	Ataxia-telangiectasia; Breast cancer
non-BRCA1/2 HR	*BARD1*	BRCA1-ASSOCIATED RING DOMAIN 1	Breast cancer
non-BRCA1/2 HR	*BLM (RECQL3)*	BLM GENE (RECQ PROTEIN-LIKE 3)	Bloom syndrome
non-BRCA1/2 HR	*BRIP1*	BRCA1-INTERACTING PROTEIN 1	Breast cancer; Fanconi anemia
non-BRCA1/2 HR	*CHEK2*	CHECKPOINT KINASE 2, S. POMBE, HOMOLOG OF	Li-Fraumeni syndrome; Breast and colorectal cancer, susceptibility to; Breast cancer, susceptibility to; Prostate cancer, familial, susceptibility to
non-BRCA1/2 HR	*FANCC*	FANCC GENE	Fanconi anemia
non-BRCA1/2 HR	*FANCM*	FANCM GENE	Premature ovarian failure; Spermatogenic failure
non-BRCA1/2 HR	*GEN1*	GEN1, DROSOPHILA, HOMOLOG OF	Breast cancer
non-BRCA1/2 HR	*MEN1*	MULTIPLE ENDOCRINE NEOPLASIA, TYPE I	Multiple endocrine neoplasia
non-BRCA1/2 HR	*MRE11A*	MEIOTIC RECOMBINATION 11, S. CEREVISIAE, HOMOLOG OF, A	Ataxia-telangiectasia-like disorder 1
non-BRCA1/2 HR	*NBN*	NIBRIN	Aplastic anemia; Leukemia; Nijmegen breakage syndrome
non-BRCA1/2 HR	*PALB2*	PARTNER AND LOCALIZER OF BRCA2	Fanconi anemia; Breast cancer; Pancreatic cancer
non-BRCA1/2 HR	*RAD50*	RAD50, S. CEREVISIAE, HOMOLOG OF	Nijmegen breakage syndrome-like disorder
non-BRCA1/2 HR	*RAD51*	RAD51, S. CEREVISIAE, HOMOLOG OF	Fanconi anemia; Mirror movements; Breast cancer
non-BRCA1/2 HR	*RAD51B*	RAD51, S. CEREVISIAE, HOMOLOG OF, B	Breast cancer
non-BRCA1/2 HR	*RAD51C*	RAD51, S. CEREVISIAE, HOMOLOG OF, C	Fanconi anemia; Breast-ovarian cancer
non-BRCA1/2 HR	*RAD51D*	RAD51, S. CEREVISIAE, HOMOLOG OF, D	Breast-ovarian cancer
non-BRCA1/2 HR	*RECQL*	RECQ PROTEIN-LIKE	Breast cancer
non-BRCA1/2 HR	*SLX4*	SLX4, S. CEREVISIAE, HOMOLOG OF	Fanconi anemia
non-BRCA1/2 HR	*XRCC2*	X-RAY REPAIR, COMPLEMENTING DEFECTIVE, IN	Fanconi anemia

non-BRCA1/2 HR	*XRCC3*	CHINESE HAMSTER, 2 X-RAY REPAIR, COMPLEMENTING DEFECTIVE, IN CHINESE HAMSTER, 3	Breast cancer; Melanoma
Other	*APC*	APC GENE	Adenomatous polyposis coli; Brain tumor-polyposis syndrome; Desmoid disease; Gardner syndrome
Other	*ATR*	ATR GENE	Cutaneous telangiectasia and cancer syndrome; Seckel syndrome
Other	*AXIN1*	AXIS INHIBITOR 1	Caudal duplication anomaly
Other	*AXIN2*	AXIS INHIBITOR 2	Oligodontia-colorectal cancer syndrome
Other	*BAP1*	BRCA1-ASSOCIATED PROTEIN 1	Tumor predisposition syndrome
Other	*BMPR1A*	BONE MORPHOGENETIC PROTEIN RECEPTOR, TYPE IA	Juvenile polyposis syndrome; Polyposis syndrome; Polyposis
Other	*CDH1*	CADHERIN 1	Blepharocheilodontic syndrome; Gastric cancer; Breast cancer; Prostate cancer
Other	*CDK4*	CYCLIN-DEPENDENT KINASE 4	Melanoma
Other	*CDKN2A*	CYCLIN-DEPENDENT KINASE INHIBITOR 2A	Melanoma and neural system tumor syndrome; Orolaryngeal cancer; Pancreatic cancer/melanoma syndrome; Melanoma
Other	*CTNNB1*	CATENIN, BETA-1	Exudative vitreoretinopathy; Mental retardation
Other	*FAM175A*	FAMILY WITH SEQUENCE SIMILARITY 175, MEMBER A	Breast cancer
Other	*HOXB13*	HOMEOBOX B13	Breast cancer
Other	*MET*	MET PROTOONCOGENE	Deafness; Osteofibrous dysplasia
Other	*MUTYH*	MutY, E. COLI, HOMOLOG OF	Adenomas; Colorectal adenomatous polyposis
Other	*NF1*	NEUROFIBROMATOSIS, TYPE I	Neurofibromatosis
Other	*PALLD*	PALLADIN, MOUSE, HOMOLOG OF	Pancreatic cancer
Other	*PIK3CA*	PHOSPHATIDYLINOSITOL 3-KINASE, CATALYTIC, ALPHA	Cowden syndrome
Other	*PPM1D*	PROTEIN PHOSPHATASE, MAGNESIUM/MANGANESE-DEPENDENT, 1D	Intellectual developmental disorder with gastrointestinal difficulties and high pain threshold
Other	*PTCH1*	PATCHED, DROSOPHILA, HOMOLOG OF, 1	Basal cell nevus syndrome; Holoprosencephaly
Other	*PTEN*	PHOSPHATASE AND TENSIN HOMOLOG	Bannayan-Riley-Ruvalcaba syndrome; Cowden syndrome; Lhermitte-Duclos syndrome; Macrocephaly/autism syndrome; PTEN hamartoma tumor syndrome; VATER association with macrocephaly and ventriculomegaly; Glioma; Meningioma
Other	*RET*	REARRANGED DURING TRANSFECTION PROTOONCOGENE	Central hypoventilation syndrome; Medullary thyroid carcinoma; Multiple endocrine neoplasia; Pheochromocytoma; Hirschsprung disease
Other	*SMAD4*	MOTHERS AGAINST DECAPENTAPLEGIC, DROSOPHILA, HOMOLOG OF, 4	Juvenile polyposis/hereditary hemorrhagic telangiectasia syndrome; Myhre syndrome; Polyposis
Other	*STK11*	SERINE/THREONINE PROTEIN KINASE 11	Pancreatic cancer; Peutz-Jeghers syndrome
Other	*TP53*	TUMOR PROTEIN p53	Adrenal cortical carcinoma; Breast cancer; Choroid plexus papilloma; Colorectal cancer; Hepatocellular carcinoma; Li-Fraumeni syndrome; Nasopharyngeal carcinoma; Osteosarcoma; Pancreatic cancer; Basal cell carcinoma; Glioma

Table 2.3. Number of the pathogenic SNVs and INDELs and number of mutation carriers on 21 breast cancer predisposition genes.[†]

Gene	No. of SNVs	No. of INDELs	No. of carriers	Proportion (%)
BRCA1/2				
BRCA2	9	12	24	4.1
BRCA1	8	11	21	3.6
HR pathway genes				
PALB2	4	2	7	1.2
CHEK2	2	1	3	0.5
GEN1	0	2	3	0.3
BLM	2	0	2	0.3
RAD51D	2	0	2	0.3
RECQL	1	0	1	0.2
MRE11A	1	0	1	0.2
FANCM	1	0	1	0.2
BRIP1	1	0	1	0.2
MMR pathway genes				
MSH6	1	3	4	0.7
PMS2	1	1	2	0.3
MSH2	1	0	1	0.2
PMS1	0	1	1	0.2
Other genes				
MUTYH	4	0	4	0.7
TP53	3	0	3	0.5
PTEN	0	2	3	0.5
MET	1	0	1	0.2
NF1	0	1	1	0.2
AXIN1	1	0	1	0.2
Sum	43	36	84	14.4

[†] Genes are sorted in descending order number of carriers stratified by gene groups.

SNV, single nucleotide variant; INDEL, short insertion and deletion; HR, homologous recombination; and MMR, mismatch repair.

Table 2.4. List of 41 pathogenic SNVs identified in this study.

Patient	Gene	Refseq ID	Type of mutation	Nucleotide change	Amino acid change	SIFT	Polyohen2 HVAR	Polyohen2 HDIV	GERP++	Freq. in 1000G EAS	Freq. in ExAC EAS	Clinvar
584	*BRCA1*	NM_007294	Nonsense	c.C4327T	p.R1443X	0.87	NA	NA	2.27	0.000	0.0000	Pathogenic
675	*BRCA1*	NM_007294	Nonsense	c.C4327T	p.R1443X	0.87	NA	NA	2.27	0.000	0.0000	Pathogenic
663	*BRCA1*	NM_007294	Splice site	c.4185+1G>A		NA	NA	NA	4.66	0.000	0.0000	Pathogenic
322	*BRCA1*	NM_007294	Nonsense	c.G2158T	p.E720X	0.13	NA	NA	3.50	0.000	0.0000	Pathogenic
711	*BRCA1*	NM_007294	Nonsense	c.A1519T	p.R507X	1.00	NA	NA	4.62	0.000	0.0000	not provided
142	*BRCA1*	NM_007294	Nonsense	c.G1154A	p.W385X	1.00	NA	NA	4.71	0.000	0.0000	NA
210	*BRCA1*	NM_007294	Splice site	c.135-2A>C		NA	NA	NA	4.55	0.000	0.0000	NA
350	*BRCA2*	NM_000059	Nonsense	c.A1399T	p.K467X	1.00	NA	NA	2.89	0.000	0.0001	Pathogenic
180	*BRCA2*	NM_000059	Nonsense	c.C3109T	p.Q1037X	0.15	NA	NA	3.97	0.000	0.0000	Pathogenic
256	*BRCA2*	NM_000059	Nonsense	c.C3109T	p.Q1037X	0.15	NA	NA	3.97	0.000	0.0000	Pathogenic
20	*BRCA2*	NM_000059	Nonsense	c.C3385T	p.Q1129X	0.00	NA	NA	5.75	0.000	0.0000	NA
332	*BRCA2*	NM_000059	Nonsense	c.C3385T	p.Q1129X	0.00	NA	NA	5.75	0.000	0.0000	NA
265	*BRCA2*	NM_000059	Nonsense	c.C5682G	p.Y1894X	0.97	NA	NA	NA	0.000	0.0000	Pathogenic
250	*BRCA2*	NM_000059	Nonsense	c.A7252T	p.R2418X	0.98	NA	NA	NA	0.000	0.0000	NA
403	*BRCA2*	NM_000059	Nonsense	c.A7486T	p.K2496X	1.00	NA	NA	3.03	0.000	0.0000	NA
254	*BRCA2*	NM_000059	Splice site	c.8331+1G>T		NA	NA	NA	5.68	0.000	0.0000	Pathogenic
413	*BRCA2*	NM_000059	Nonsense	c.C9382T	p.R3128X	1.00	NA	NA	4.97	0.000	0.0000	Pathogenic
593	*MUTYH*	NM_001128425	Splice site	c.1186+2T>C		NA	NA	NA	5.82	0.000	0.0000	NA
469	*MUTYH*	NM_001128425	Nonsense	c.G522A	p.W174X	0.00	0.967	1.000	5.01	0.000	0.0000	NA
155	*MUTYH*	NM_001128425	Nonsense	c.C289T	p.R97X	1.00	NA	NA	4.41	0.000	0.0000	Pathogenic
611	*MUTYH*	NM_001128425	Nonsense	c.C55T	p.R19X	1.00	NA	NA	4.30	0.000	0.0000	Pathogenic
506	*MSH2*	NM_000251	Nonsense	c.C2785T	p.R929X	0.34	NA	NA	2.62	0.001	0.0011	Uncertain significance
266	*PMS2*	NM_000535	Nonsense	c.C2404T	p.R802X	1.00	NA	NA	2.90	0.000	0.0000	Pathogenic
369	*MET*	NM_001127500	Nonsense	c.C4198T	p.R1400X	0.42	NA	NA	4.04	0.001	0.0001	Uncertain significance

220	*MRE11A*	NM_005590	Nonsense	c.C1090T	p.R364X	1.00	NA	NA	5.07	0.000	0.0001	Pathogenic
554	*RECQL*	NM_002907	Nonsense	c.C796T	p.Q266X	0.58	NA	NA	4.57	0.001	0.0006	NA
506	*FANCM*	NM_020937	Nonsense	c.G5707T	p.E1903X	0.29	NA	NA	4.38	0.000	0.0000	NA
133	*BLM*	NM_000057	Splice site	c.1221-2A>C		NA	NA	NA	4.75	0.000	0.0000	NA
143	*BLM*	NM_000057	Nonsense	c.C3678A	p.C1226X	0.30	NA	NA	5.00	0.000	0.0000	NA
405	*AXIN1*	NM_003502	Nonsense	c.C2521T	p.R841X	0.91	NA	NA	3.45	0.000	0.0000	NA
688	*PALB2*	NM_024675	Nonsense	c.G2968T	p.E990X	0.01	NA	NA	4.87	0.000	0.0000	Pathogenic
124	*PALB2*	NM_024675	Nonsense	c.C1516T	p.Q506X	0.82	NA	NA	2.09	0.000	0.0000	NA
495	*PALB2*	NM_024675	Nonsense	c.C778T	p.Q260X	0.33	NA	NA	NA	0.000	0.0000	NA
348	*PALB2*	NM_024675	Nonsense	c.C751T	p.Q251X	0.29	NA	NA	NA	0.000	0.0000	Pathogenic
547	*PALB2*	NM_024675	Nonsense	c.C751T	p.Q251X	0.29	NA	NA	NA	0.000	0.0000	Pathogenic
401	*TP53*	NM_000546	Nonsense	c.G592T	p.E198X	0.00	NA	NA	5.28	0.000	0.0000	NA
31	*RAD51D*	NM_002878, NM_001142571	Nonsense	c.G793T, c.G853T	p.G265X, p.G285X	0.00	NA	NA	4.86	0.000	0.0000	NA
185	*RAD51D*	NM_001142571	Nonsense	c.C184T	p.Q62X	0.52	NA	NA	NA	0.000	0.0001	NA
395	*BRIP1*	NM_032043	Nonsense	c.C2392T	p.R798X	NA	NA	NA	3.51	0.000	0.0000	Conflicting interpretations of pathogenicity
75	*CHEK2*	NM_007194	Nonsense	c.C417A	p.Y139X	0.49	NA	NA	NA	0.000	0.0000	NA
564	*CHEK2*	NM_007194	Nonsense	c.C283T	p.R95X	1.00	NA	NA	5.42	0.000	0.0000	Pathogenic

Table 2.5. List of 39 identified INDELs.

Patient	Gene	Refseq ID	Type of mutation	Nucleotide change	Amino acid change	SIFT	Polyohen2 HVAR	Polyohen2 HDIV	GERP++	Freq. in 1000G EAS	Freq. in ExAC EAS	Clinvar
174	*BRCA1*	NM_007294	Frameshift deletion	c.5470_5477del	p.I1824fs	NA	NA	NA	NA	0.0000	0.0000	Pathogenic
352	*BRCA1*	NM_007294	Frameshift insertion	c.4997dupA	p.Y1666_K1667delinsX	NA	NA	NA	NA	0.0000	0.0000	NA
411	*BRCA1*	NM_007294	Frameshift deletion	c.4873_4880del	p.Y1625fs	NA	NA	NA	NA	0.0000	0.0000	NA
528	*BRCA1*	NM_007294	Frameshift deletion	c.4065_4068del	p.N1355fs	NA	NA	NA	NA	0.0000	0.0000	Pathogenic
644	*BRCA1*	NM_007294	Frameshift deletion	c.4065_4068del	p.N1355fs	NA	NA	NA	NA	0.0000	0.0000	Pathogenic
378	*BRCA1*	NM_007294	Frameshift deletion	c.3980delA	p.Q1327fs	NA	NA	NA	NA	0.0000	0.0000	NA
587	*BRCA1*	NM_007294	Frameshift deletion	c.3359_3363del	p.V1120fs	NA	NA	NA	NA	0.0000	0.0000	not provided
178	*BRCA1*	NM_007294	Frameshift deletion	c.2945delC	p.P982fs	NA	NA	NA	NA	0.0000	0.0000	Pathogenic
368	*BRCA1*	NM_007294	Frameshift deletion	c.1934delC	p.S645fs	NA	NA	NA	NA	0.0000	0.0000	NA
122	*BRCA1*	NM_007294	Frameshift deletion	c.1338_1339del	p.R446fs	NA	NA	NA	NA	0.0000	0.0000	NA
634	*BRCA1*	NM_007294	Frameshift insertion	c.1299dupC	p.S434fs	NA	NA	NA	NA	0.0000	0.0000	NA
344	*BRCA1*	NM_007294	Frameshift insertion	c.1293dupA	p.L432fs	NA	NA	NA	NA	0.0000	0.0000	NA
341	*BRCA2*	NM_000059	Frameshift deletion	c.903delT	p.D301fs	NA	NA	NA	NA	0.0000	0.0000	NA
488	*BRCA2*	NM_000059	Frameshift deletion	c.1567_1579del	p.H523fs	NA	NA	NA	NA	0.0000	0.0000	NA
648	*BRCA2*	NM_000059	Frameshift deletion	c.2806_2809del	p.K936fs	NA	NA	NA	NA	0.0000	0.0000	NA
407	*BRCA2*	NM_000059	Frameshift deletion	c.2866delA	p.K956fs	NA	NA	NA	NA	0.0000	0.0000	NA
109	*BRCA2*	NM_000059	Frameshift deletion	c.5067delA	p.A1689fs	NA	NA	NA	NA	0.0000	0.0000	NA
507	*BRCA2*	NM_000059	Frameshift insertion	c.5067dupA	p.A1689fs	NA	NA	NA	NA	0.0000	0.0000	NA
737	*BRCA2*	NM_000059	Frameshift deletion	c.5825_5826del	p.V1942fs	NA	NA	NA	NA	0.0000	0.0000	NA

508	*BRCA2*	NM_000059	Splice site	c.7806-2->G		NA	NA	NA	NA	0.0000	0.0000	NA
42	*BRCA2*	NM_000059	Frameshift deletion	c.8472_8485del	p.R2824fs	NA	NA	NA	NA	0.0000	0.0000	NA
251	*BRCA2*	NM_000059	Frameshift deletion	c.9009delA	p.G3003fs	NA	NA	NA	NA	0.0000	0.0000	NA
414	*BRCA2*	NM_000059	Frameshift insertion	c.9090dupA	p.T3030fs	NA	NA	NA	NA	0.0000	0.0000	NA
483	*BRCA2*	NM_000059	Frameshift insertion	c.9090dupA	p.T3030fs	NA	NA	NA	NA	0.0000	0.0000	NA
421	*BRCA2*	NM_000059	Frameshift deletion	c.9400delG	p.G3134fs	NA	NA	NA	NA	0.0000	0.0000	NA
1	*GEN1*	NM_182625	Frameshift deletion	c.1929_1932del	p.I643fs	NA	NA	NA	NA	0.0000	0.0001	NA
1	*GEN1*	NM_182625	Frameshift deletion	c.2497_2500del	p.K833fs	NA	NA	NA	NA	0.0000	0.0000	NA
164	*GEN1*	NM_182625	Frameshift deletion	c.2497_2500del	p.K833fs	NA	NA	NA	NA	0.0000	0.0000	NA
650	*MSH6*	NM_000179	Frameshift deletion	c.2672_2673del	p.I891fs	NA	NA	NA	NA	0.0000	0.0000	NA
535	*MSH6*	NM_000179	Frameshift insertion	c.3254dupC	p.T1085fs	NA	NA	NA	NA	0.0000	0.0004	NA
457	*MSH6*	NM_000179	Frameshift insertion	c.4082_4083insGACT	p.X1361delinsX	NA	NA	NA	NA	0.0000	0.0002	NA
450	*PMS1*	NM_000534	Frameshift deletion	c.2749delC	p.H917fs	NA	NA	NA	NA	0.0000	0.0000	NA
154	*PMS2*	NM_000535	Frameshift deletion	c.1864_1865del	p.M622fs	NA	NA	NA	NA	0.0000	0.0000	Pathogenic
29	*PTEN*	NM_001304717	Frameshift insertion	c.154_155insCGG	p.R52fs	NA	NA	NA	NA	0.0000	0.0000	NA
462	*PTEN*	NM_001304717	Frameshift insertion	c.154_155insCGG	p.R52fs	NA	NA	NA	NA	0.0000	0.0000	NA
375	*PALB2*	NM_024675	Frameshift deletion	c.3076_3077del	p.L1026fs	NA	NA	NA	NA	0.0000	0.0000	NA
157	*PALB2*	NM_024675	Frameshift insertion	c.2760dupA	p.Q921fs	NA	NA	NA	NA	0.0000	0.0000	NA
129	*NF1*	NM_000267	Frameshift deletion	c.7850delT	p.I2617fs	NA	NA	NA	NA	0.0000	0.0000	NA
679	*CHEK2*	NM_007194	Splice site	c.444+1G>-		NA	NA	NA	NA	0.0000	0.0000	Pathogenic

Table 2.6. List of 6 identified pathogenic SNVs.

Patient	Gene	Refseq ID	Type of mutation	Nucleotide change	Amino acid change	SIFT	Polyohen2 HVAR	Polyohen2 HDIV	GERP++	Freq. in 1000G EAS	Freq. in ExAC EAS	Clinvar
156	*BRCA1*	NM_007294	intronic	c.213-15A>G	NA	NA	NA	NA	2.62	0.0000	0.0000	Pathogenic/Likely pathogenic
30	*BRCA1*	NM_007294	Missense	c.G212A	p.R71K	NA	0.987	0.999	4.64	0.0000	0.0000	Pathogenic
405	*BRCA2*	NM_000059	Missense	c.G8243A	p.G2748D	0.00	1.000	1.000	5.42	0.0000	0.0000	Pathogenic
433	*MSH6*	NM_000179	Missense	c.C3226T	p.R1076C	0.10	0.818	0.991	5.37	0.0000	0.0003	Likely pathogenic
438	*TP53*	NM_000546	Missense	c.G743A	p.R248Q	0.01	0.999	1.000	3.65	0.0000	0.0002	Pathogenic/Likely pathogenic
712	*TP53*	NM_000546	Missense	c.G542A	p.R181H	0.00	0.966	1.000	4.28	0.0000	0.0000	Pathogenic/Likely pathogenic

Table 2.7. Associations of all germline pathogenic mutation status with expressions of biomarkers.

	No. and proportion (%) of patients					
Variables	Carriers		Non-carriers		OR (95% CI)	*P*-value
ER						
Negative	36	(43.9)	166	(33.9)	1.00	
Positive	46	(56.1)	324	(66.1)	0.66 (0.40-1.09)	0.08
PR						
Negative	44	(53.7)	225	(45.9)	1.00	
Positive	38	(46.3)	265	(54.1)	0.23 (0.45-1.20)	0.23
HER2						
Negative	55	(76.4)	252	(59.9)	1.00	
Positive	17	(23.6)	169	(40.1)	0.46 (0.24-0.84)	0.0082
TNBC						
No	45	(62.5)	362	(86.2)	1.00	
Yes	27	(37.5)	58	(13.8)	3.73 (2.06-6.70)	6.2×10^{-6}
p53						
Negative	25	(30.5)	127	(26.3)	1.00	
Positive	57	(69.5)	355	(73.7)	0.82 (0.48-1.42)	0.42
CK5/6						
Negative	68	(88.3)	410	(89.3)	1.00	
Positive	9	(11.7)	49	(10.7)	1.11 (0.46-2.42)	0.84
E-cadherin						
Negative	5	(6.1)	5	(1.0)	1.00	
Positive	76	(93.9)	474	(99.0)	0.16 (0.036-0.72)	7.8×10^{-3}

OR, odds ratio; and CI, confidence interval.

Table 2.8. Association of groups of homologous recombination (HR) mutation carriers for expressions of biomarkers compared to non-carriers.

Variables	*BRCA1* mutation				*BRCA2* mutation				Non-*BRCA1/2* HR mutation			
	Carriers		OR (95% CI)	*P*-value	Carriers		OR (95% CI)	*P*-value	Carriers		OR (95% CI)	*P*-value
ER												
Negative	15	(78.9)	1.00		5	(20.8)	1.00		10	(50.0)	1.00	
Positive	4	(21.1)	0.14 (0.033-0.44)	1.1×10^{-4}	19	(79.2)	1.94 (0.69-6.78)	0.27	10	(50.0)	0.51 (0.19-1.40)	0.15
PR												
Negative	15	(78.9)	1.00		8	(33.3)	1.00		14	(66.7)	1.00	
Positive	4	(21.1)	0.23 (0.054-0.73)	8.2×10^{-3}	16	(66.7)	1.69 (0.67-4.65)	0.29	6	(33.3)	0.36 (0.11-1.03)	0.040
HER2												
Negative	16	(88.9)	1.00		15	(78.9)	1.00		13	(72.2)	1.00	
Positive	2	(11.1)	0.19 (0.021-0.81)	0.013	4	(21.1)	0.40 (0.095-1.28)	0.15	5	(27.8)	0.57 (0.16-1.76)	0.34
TNBC												
No	4	(22.2)	1.00		15	(78.9)	1.00		12	(66.7)	1.00	
Yes	14	(77.8)	21.58 (6.49-93.18)	6.0×10^{-9}	4	(21.1)	1.66 (0.39-5.46)	0.33	6	(33.3)	3.11 (0.92-9.38)	0.034
p53												
Negative	8	(42.1)	1.00		6	(25.0)	1.00		6	(30.0)	1.00	
Positive	11	(57.9)	0.49 (0.18-1.44)	0.18	18	(75.0)	1.07 (0.40-3.38)	1.0	14	(70.0)	0.84 (0.29-2.71)	0.80
CK5/6												
Negative	13	(68.4)	1.00		19	(86.4)	1.00		19	(95.0)	1.00	
Positive	6	(31.6)	3.84 (1.15-11.46)	0.015	3	(13.6)	1.30 (0.24-4.73)	0.72	1	(5.0)	0.44 (0.010-2.90)	0.71
E-cadherin												
Negative	1	(5.3)	1.00		3	(12.5)	1.00		0	(0.0)	1.00	
Positive	18	(94.7)	0.19 (0.020-9.47)	0.21	21	(87.5)	0.075 (0.014-0.51)	4.6×10^{-3}	20	(100.0)	∞ (0.036-∞)	1.0

OR, odds ratio; and CI, confidence interval.

Table 2.9. Association of groups of mismatch repair (MMR) gene and other mutation carriers for expressions of biomarkers compared to non-carriers.

	MMR genes mutation				Others mutation			
Variables	Carriers		OR (95% CI)	*P*-value	Carriers		OR (95% CI)	*P*-value
ER								
Negative	0	(0.0)	1.00		6	(46.2)	1.00	
Positive	7	(100.0)	∞ (0.73-∞)	0.10	7	(53.8)	0.60 (0.17-2.19)	0.38
PR								
Negative	2	(28.6)	1.00		6	(46.2)	1.00	
Positive	5	(71.4)	2.12 (0.34-22.46)	0.46	7	(53.8)	0.99 (0.28-3.62)	1.0
HER2								
Negative	6	(85.7)	1.00		6	(54.5)	1.00	
Positive	1	(14.3)	0.25 (0.0054-2.08)	0.25	5	(45.5)	1.24 (0.29-4.97)	0.76
TNBC								
No	7	(100.0)	1.00		8	(72.7)	1.00	
Yes	0	(0.0)	0.00 (0.00-4.45)	0.60	3	(27.3)	2.33 (0.39-10.09)	0.19
p53								
Negative	2	(28.6)	1.00		3	(23.1)	1.00	
Positive	5	(71.4)	0.89 (0.14-9.51)	1.00	10	(76.9)	1.19 (0.30-6.85)	1.0
CK5/6								
Negative	6	(100.0)	1.00		12	(100.0)	1.00	
Positive	0	(0.0)	0.00 (0.00-7.32)	1.0	0	(0.0)	0.00 (0.00-3.12)	0.62
E-cadherin								
Negative	0	(0.0)	1.00		1	(8.3)	1.00	
Positive	7	(100.0)	∞ (0.011-∞)	1.0	11	(91.7)	0.12 (0.012-5.98)	0.14

OR, odds ratio; and CI, confidence interval.

Table 2.10. Associations of all germline pathogenic mutation status with prognostic factors.

Variables	No. and proportion (%) of patients				OR (95% CI)	*P*-value
	Carriers		Non-carriers			
Histological grade						
I + II	42	(68.9)	300	(79.4)	1.00	
III	19	(31.1)	78	(20.6)	1.74 (0.90-3.26)	0.070
Axillary lymph node status						
Negative	49	(63.6)	379	(79.0)	1.00	
Positive	28	(36.4)	101	(21.0)	2.14 (1.23-3.68)	5.3×10^{-3}
Bone metastasis						
Negative	75	(97.4)	478	(99.4)	1.00	
Positive	2	(2.6)	3	(0.6)	4.23 (0.35-37.54)	0.14
Vascular invasion						
Negative	75	(96.2)	442	(94.0)	1.00	
Positive	3	(3.8)	28	(6.0)	0.63 (0.12-2.13)	0.60
Family history of all cancer						
Negative	54	(74.0)	299	(73.3)	1.00	
Positive	19	(26.0)	109	(26.7)	0.97 (0.52-1.74)	1.00
Family history of BC						
Negative	61	(83.6)	382	(93.6)	1.00	
Positive	12	(16.4)	26	(6.4)	2.88 (1.26-6.30)	7.6×10^{-3}
Family history of HBOC						
Negative	60	(82.2)	379	(92.9)	1.00	
Positive	13	(17.8)	29	(7.1)	2.82 (1.27-5.99)	5.9×10^{-3}

OR, odds ratio; and CI, confidence interval. BC, breast cancer. HBOC, hereditary breast and ovarian cancer.

Table 2.11. Association of groups of homologous recombination (HR) gene mutation carriers for prognostic factors compared to non-carriers.

	BRCA1 mutation				*BRCA2* mutation				Non-*BRCA1/2* HR mutation			
Variables	Carriers		OR (95% CI)	*P*-value	Carriers		OR (95% CI)	*P*-value	Carriers		OR (95% CI)	*P*-value
Histological grade												
I + II	7	(46.7)			17	(85.0)	1.00		10	(83.3)	1.00	
III	8	(53.3)	4.37 (1.34-14.64)	6.5×10^{-3}	3	(15.0)	0.68 (0.12-2.44)	0.78	2	(16.7)	0.77 (0.080-3.72)	1.0
Axillary lymph node status												
Negative	12	(70.6)	1.00		15	(62.5)	1.00		11	(61.1)	1.00	
Positive	5	(29.4)	1.56 (0.42-4.90)	0.38	9	(37.5)	2.25 (0.84-5.67)	0.074	7	(38.9)	2.38 (0.76-6.93)	0.082
Bone metastasis												
Negative	17	(100.0)	1.00		22	(91.7)	1.00		19	(100.0)	1.00	
Positive	0	(0.0)	0.00 (0.00-71.29)	1.0	2	(8.3)	14.25 (1.14-130.91)	0.020	0	(0.0)	0.00 (0.00-63.41)	1.0
Vascular invasion												
Negative	19	(100.0)	1.00		23	(95.8)	1.00		17	(94.4)	1.00	
Positive	0	(0.0)	0.00 (0.00-3.60)	0.62	1	(4.2)	0.69 (0.016-4.56)	1.0	1	(5.6)	0.94 (0.023-6.82)	1.0
Family history of all cancer												
Negative	6	(37.5)	1.00		17	(77.3)	1.00		15	(88.2)	1.00	
Positive	10	(62.5)	4.55 (1.46-15.62)	3.6×10^{-3}	5	(22.7)	0.81 (0.23-2.35)	0.81	2	(11.8)	0.37 (0.040-1.62)	0.26
Family history of BC												
Negative	9	(56.3)	1.00		17	(77.3)	1.00		17	(100.0)	1.00	
Positive	7	(43.8)	11.28 (3.29-37.23)	6.3×10^{-5}	5	(22.7)	4.30 (1.15-13.45)	0.015	0	(0.0)	0.00 (0.00-3.81)	0.61
Family history of HBOC												
Negative	8	(50.0)	1.00		17	(77.3)	1.00		17	(100.0)	1.00	
Positive	8	(50.0)	12.89 (3.91-42.62)	1.2×10^{-5}	5	(22.7)	3.82 (1.03-11.85)	0.023	0	(0.0)	0.00 (0.00-3.37)	0.62

OR, odds ratio; and CI, confidence interval. BC, breast cancer. HBOC, hereditary breast and ovarian cancer.

Table 2.12. Association of groups of mismatch repair (MMR) gene and other mutation carriers for prognostic factors compared to non-carriers.

	MMR genes mutation				Other mutation			
Variables	Carriers		OR (95% CI)	*P*-value	Carriers		OR (95% CI)	*P*-value
Histological grade								
I + II	1	(33.3)	1.00		7	(63.6)	1.00	
III	2	(66.7)	7.64 (0.39-453.88)	0.11	4	(36.4)	2.19 (0.46-8.88)	0.25
Axillary lymph node status								
Negative	4	(66.7)	1.00		8	(61.5)	1.00	
Positive	2	(33.3)	1.87 (0.17-13.28)	0.61	5	(38.5)	2.34 (0.59-8.32)	0.17
Bone metastasis								
Negative	6	(100.0)	1.00		13	(100.0)	1.00	
Positive	0	(0.0)	0.00 (0.00-220.07)	1.0	0	(0.0)	0.00 (0.00-94.62)	1.0
Vascular invasion								
Negative	6	(100.0)	1.00		12	(92.3)	1.00	
Positive	0	(0.0)	0.00 (0.00-14.05)	1.0	1	(7.7)	1.31 (0.030-9.50)	0.56
Family history of all cancer								
Negative	7	(100.0)	1.00		10	(83.3)	1.00	
Positive	0	(0.0)	0.00 (0.00-1.94)	0.20	2	(16.7)	0.55 (0.058-2.64)	0.74
Family history of BC								
Negative	7	(100.0)	1.00		12	(100.0)	1.00	
Positive	0	(0.0)	0.00 (0.00-10.76)	1.0	0	(0.0)	0.00 (0.00-5.63)	1.0
Family history of HBOC								
Negative	7	(100.0)	1.00		12	(100.0)	1.00	
Positive	0	(0.0)	0.00 (0.00-9.52)	1.0	0	(0.0)	0.00 (0.00-4.98)	1.0

OR, odds ratio; and CI, confidence interval. BC, breast cancer. HBOC, hereditary breast and ovarian cancer.

Figure 2.1

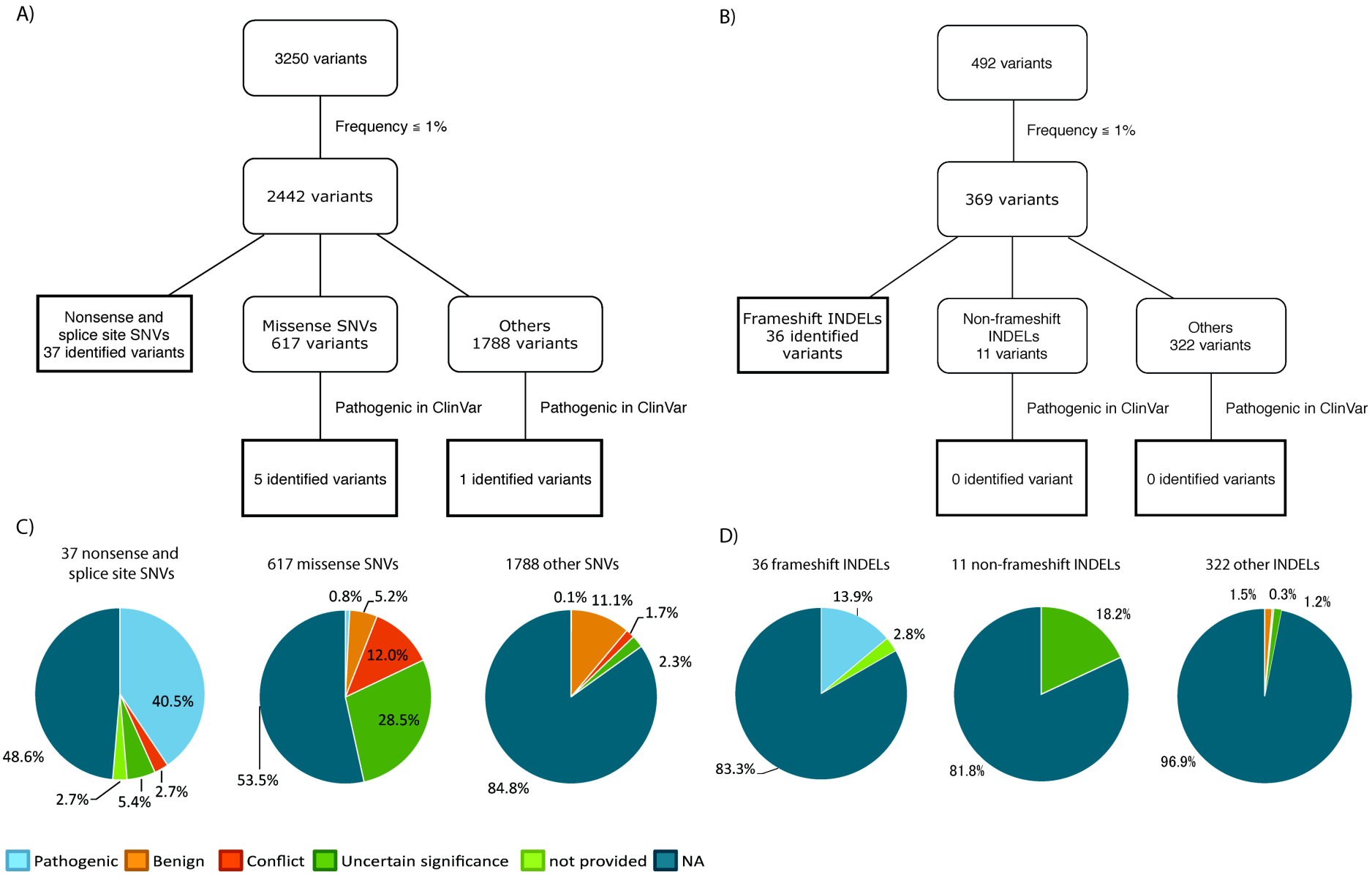

A)
3250 variants
Frequency ≦ 1%
2442 variants
Nonsense and splice site SNVs 37 identified variants
Missense SNVs 617 variants
Others 1788 variants
Pathogenic in ClinVar
Pathogenic in ClinVar
5 identified variants
1 identified variants
B)
492 variants
Frequency ≦ 1%
369 variants
Frameshift INDELs 36 identified variants
Non-frameshift INDELs 11 variants
Others 322 variants
Pathogenic in ClinVar
Pathogenic in ClinVar
0 identified variant
0 identified variants
C)
37 nonsense and splice site SNVs
40.5%
48.6%
2.7%
5.4%
2.7%
617 missense SNVs
0.8%
5.2%
12.0%
28.5%
53.5%
1788 other SNVs
0.1%
11.1%
1.7%
2.3%
84.8%
D)
36 frameshift INDELs
13.9%
2.8%
83.3%
11 non-frameshift INDELs
18.2%
81.8%
322 other INDELs
1.5%
0.3%
1.2%
96.9%
Pathogenic
Benign
Conflict
Uncertain significance
not provided
NA

Figure 2.1. Flowchart illustrating the filtering and variant selection processes used to identify germline pathogenic variants. A) Flowchart for single nucleotide variants (SNVs). B) Flowchart for insertions and deletions (INDELs). C) Pie charts representing ClinVar classification of 37 nonsense and splice site SNVs (left), 617 missense SNVs (middle), and 1,788 other SNVs (synonymous, untransrated region, intronic and intergenic SNVs) (right). D) Pie charts representing ClinVar classification of 36 frameshift and splice site INDELs (left), 11 non-frameshift INDELs (middle), and 322 untransrated region, intronic and intergenic INDELs (right).

Figure 2.2

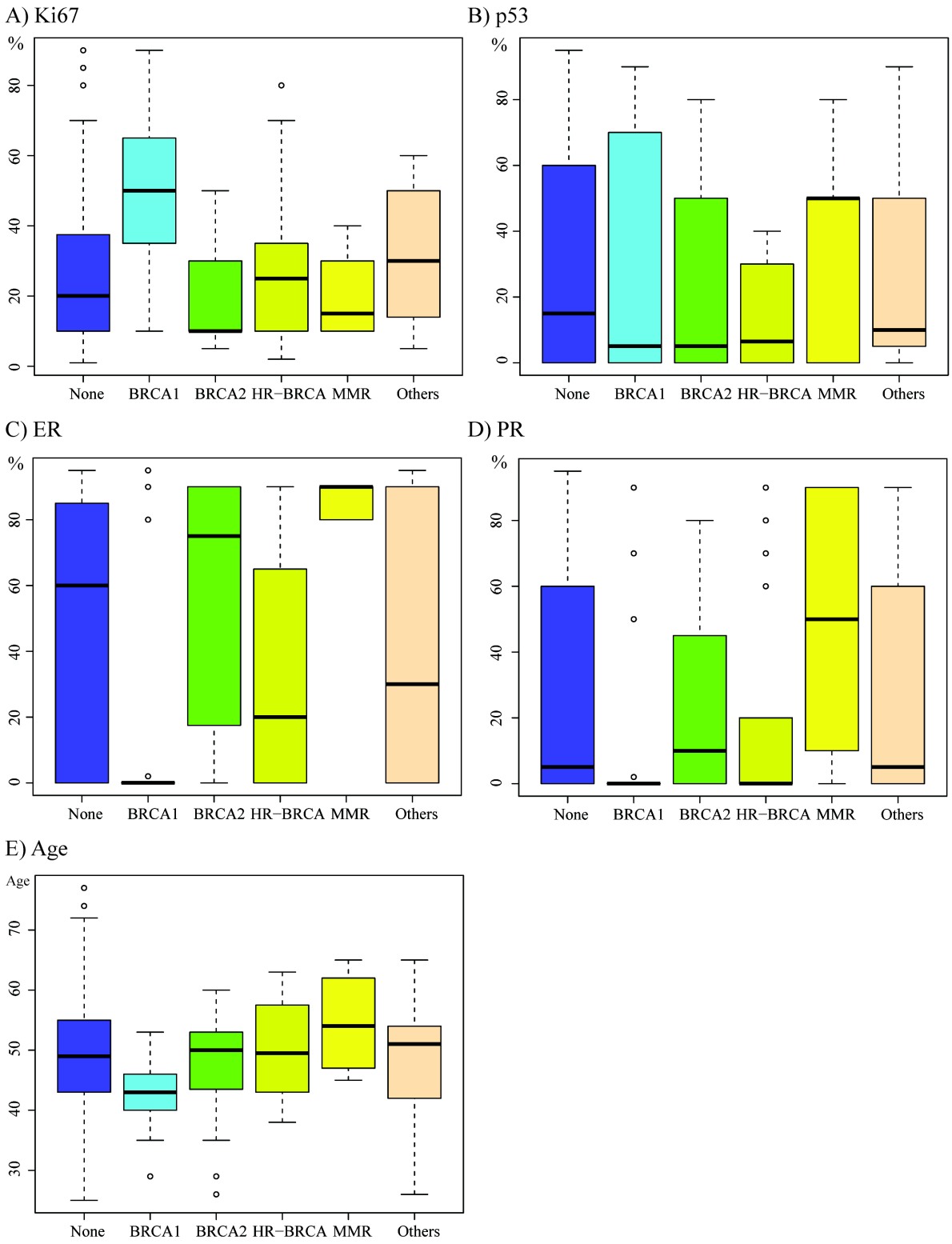
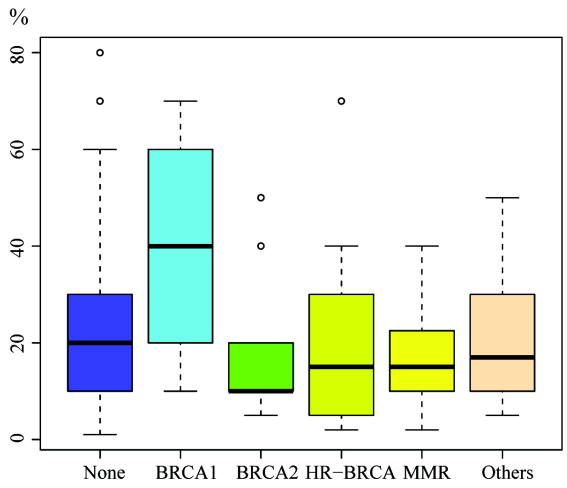

A) Ki67
%
0
20
40
60
80
None
BRCA1
BRCA2
HR−BRCA
MMR
Others
B) p53
%
0
20
40
60
80
None
BRCA1
BRCA2
HR−BRCA
MMR
Others
C) ER
%
0
20
40
60
80
None
BRCA1
BRCA2
HR−BRCA
MMR
Others
D) PR
%
0
20
40
60
80
None
BRCA1
BRCA2
HR−BRCA
MMR
Others
E) Age
Age
30
40
50
60
70
None
BRCA1
BRCA2
HR−BRCA
MMR
Others

Figure 2.2. Distributions of the expression levels of biomarkers and prognostic factors according to groups stratified by mutation status: (A) Ki67; B) p53; C) estrogen receptor (ER); D) progesterone receptor (PR); and E) age at diagnosis. Box plots represent the five-number summary: the minimum, first quartile, median, third quartile, and maximum. Points which falls more than 1.5 times the interquartile range above the third quartile or below the first quartile are defined as outliers.

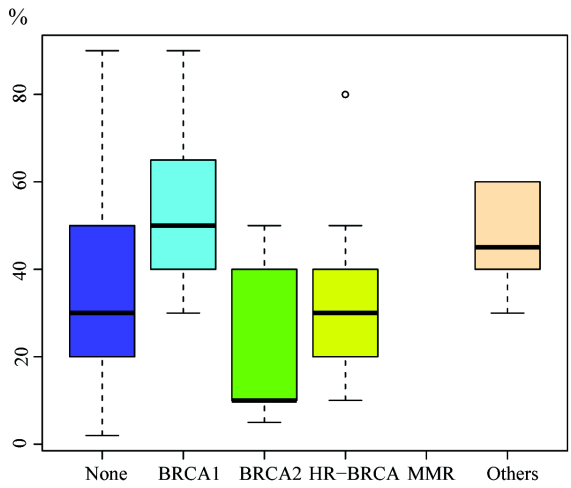
Figure 2.3

A) Ki67

I. ER positive

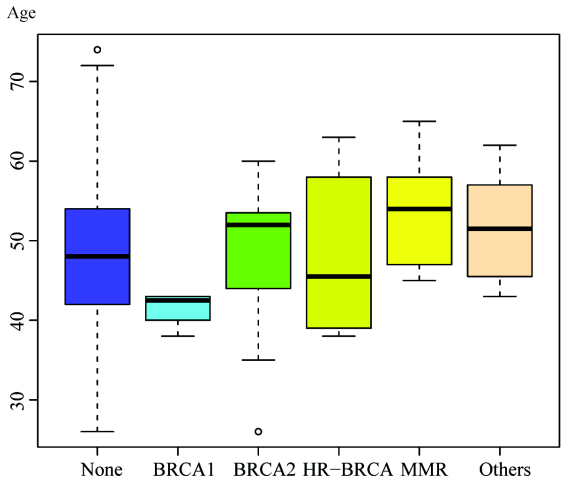


II. ER negative



B) Age

I. ER positive



II. ER negative

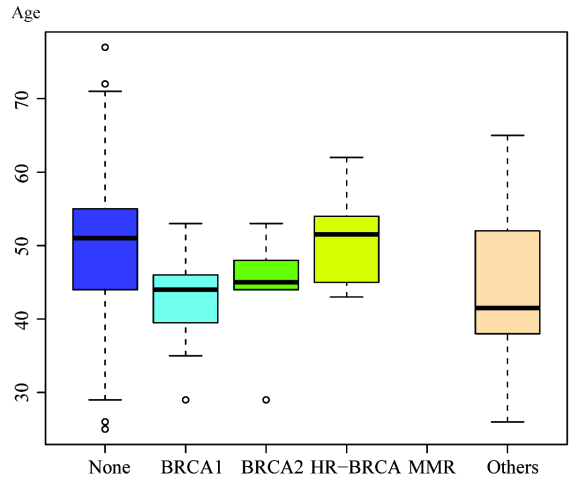


Figure 2.3. Distributions of the expression levels of biomarker and prognostic factor according to groups stratified by mutation status:(A) Ki67; B) age at diagnosis; I) ER positive; II) ER negative. Box plots represent the five-number summary: the minimum, first quartile, median, third quartile, and maximum. Points which falls more than 1.5 times the interquartile range above the third quartile or below the first quartile are defined as outliers.

Chapter 3: Exploration of Intermediate-Sized INDELs

The contents of this chapter are also described in the below paper:

Exploration of Intermediate-Sized INDELs by Next-Generation Multigene Panel Testing in Han Chinese Patients with Breast Cancer. Hata C et al. (2019). Human genome variation. 6:51. doi: 10.1038/s41439-019-0080-8. eCollection 2019.

3.1. Introduction

Breast cancer is the most common cancer among women (DeSantis et al., 2017), and approximately 10-15% of cases are associated with hereditary mutations in DNA repair genes including *BRCA1/2* (Ellisen et al. 1998). Genetic testing of *BRCA1/2* has been conducted all over the world with the advent of NGS technologies. For example, it was shown that germline mutations in genes involved in homologous recombination pathways such as *BARD1*, *BRCA1*, *BRCA2*, *PALB2*, and *RAD51D* were strongly associated with triple negative breast cancer (Shimelis et al., 2018).

Most applications of NGS-based multigene panel testing focus only on small-sized variants containing SNVs and short INDELs. Additionally, it has been reported that high-risk patients with hereditary breast and ovarian cancers harbor germline large rearrangement in *BRCA1/2* (Judkins et al., 2012). However, effects of intermediate-sized INDELs (50 bp to 10,000 bp) on the pathogenicity of breast cancer were still not fully investigated due to technical difficulties in detection from NGS data (Claudia et al., 2016 and Shigemizu et al., 2018). There is a possibility that intermediate-sized INDELs are involved in the pathology of breast cancer patients. In order to clarify clinical significance of intermediate-sized INDELs on breast cancer, I attempted to identify intermediate-sized INDELs in 54 cancer predisposition genes among 583 Han Chinese patients with breast cancer. I identified a novel deletion-insertion in *PTENα* in one patient.

In Chapter 2, I detected pathogenic mutations in 83 patients with breast cancer, and majority of the identified mutations were novel. Furthermore, I showed that mutations in genes that were involved in specific biological pathways were strongly associated with clinical variables. These results indicate the usefulness of multigene panel testing. On the other hands, pathogenic mutations were not identified in 85.8% of patients.

The objective of Chapter 3 is to detect intermediated-size INDELs in breast cancer patients whose pathogenic mutations were not identified by focusing only on SNVs and small INDELS.

3.2. Materials and Methods

3.2.1. Study patients

Information regarding the study subjects and target-gene sequencing has been described in Chapter 2. In brief, 583 Han Chinese patients with breast cancer were recruited between December 2016 and September 2017 at the First Affiliated Hospital of Chongqing Medical University and Affiliated Cancer Hospital and Institute of Guangzhou Medical University. All patients provided informed consent for participation in this study. The Ethics Committees of the First Affiliated Hospital of Chongqing Medical University, the Affiliated Cancer Hospital & Institute of Guangzhou Medical University, and the National Institute of Genetics approved the study protocols. The patients' mean age at diagnosis was 49.1 (standard deviation: 9.2) years.

3.2.2. DNA samples

Fifty-four cancer predisposition genes were selected based on previous studies of multigene panel testing for hereditary breast and/or ovarian cancer (Table 2.2). Target sequencing of these genes was performed using the pre-capture pooling method described in previous studies by using DNA samples isolated from peripheral blood (Ahmadloo et al., 2017 and Suda et al., 2018). The libraries were sequenced on an Illumina HiSeq 2500 platform operating in the rapid-run mode using a 2 × 100-bp paired-end protocol.

3.2.3. Variants validation

Sanger sequencing was performed using BigDye Terminator Cycle Sequencing or Ready Reaction kit (Life Technologies, Carlsbad, CA, USA) on ABI 3130x Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). The oligonucleotide sequences of the PCR primers are shown in Table 3.1.

3.2.4. Sequencing data analysis

NGS data processing and variant calling were performed using BWA version 0.7.15 (Li and Drubin, 2010) and GATK version 3.4-46 (McKenna et al., 2010 and DePristo et al., 2011). Functional annotation was implemented using ANNOVAR (version released 29 September 2017) (Wang et al., 2010). The estimation of variant frequencies in general populations was based on publicly available databases provided by ExAC (Lek et al., 2016). Nonsense and splice-site SNVs and frameshifting INDELs were considered to be pathogenic. The variants with previously established pathogenic or benign effects were explored based on ClinVar version 20180521 (Landrum et al., 2016). I attempted to detect intermediate- to large-sized INDELs from mapped paired-end sequencing reads via bioinformatics analysis using Manta (Chen et al., 2016).

3.3. Results

3.3.1. Identification of intermediate-sized INDELs

The average depth for the target regions was 117.6, and the mean proportion of the targeted regions covered by at least 20 reads was 99.3%, supporting confident variant detection. I identified 78 pathogenic mutations (43 SNVs and 35 short INDELs) in 21 genes containing *BRCA1/2* in Chapter 2. However, pathogenic SNVs and short INDELs were not detected in 85.8% (500/583) of the patients.

By using Manta, I identified two intermediate-sized INDELs from the patients without pathogenic SNVs or short INDELs.

One was an 89-bp heterozygous deletion present in intron 14 of *APC* (NM_000038.5:c.1743+15_1743+103del89). The patient with this mutation was 53 years old and diagnosed with TNBC because ER, PR, and HER2 were negative in IHC and FISH. Adenomatosis polyposis coli (APC) is a tumor suppressor protein that acts as an antagonist of the Wnt signaling pathway. A lack of this gene causes familial adenomatous polyposis (OMIM175100). The identified intermediate-sized deletion in the intronic region of APC might change the splicing behavior of the gene. However, this deletion is registered as "likely benign" in ClinVar. The allele frequency of this deletion is 0.9% in the general East Asian population from the ExAC project. Based on these findings, the significance of this deletion for the pathogenicity does not seem to be high.

The other INDEL was a combination of a 47-bp deletion and a 68-bp insertion in PTENα (also known as PTEN-Long) (Fig. 3.1A) [NM_001304717: c.8_54delins AGTAATGTTAGCGGTTAGGCGTACGGCCAGGGCTATTGGTTGAATGAGTAGGC TGATGGTTTCGATAG (p.R3_P18delinsQX); (Fig. 3.1A-D)]. The patient was heterozygote for this deletion–insertion. I attempted to determine the junctions of the intermediate-sized INDELs via in-house bioinformatics analysis that leverages split-reads

of paired-end sequencing. First, I extracted soft-clipped reads containing a part of the unmatched sequence with the reference genome. Second, I divided the soft-clipped reads into unmatched and matched sequences by using an in-house Perl script. Third, I aligned these two types of the read sequences with the reference genome by using BWA. Finally, I searched the genomic positions where these reads were mapped and successfully determined the deletion junctions at *PTENα* from the realignment of the matched sequences of the soft-clipped reads (Fig. 3.1B, C). The presence of the unmatched sequences of the soft-clipped reads at the deletion junctions indicated that a DNA fragment derived from another region was inserted into the *PTENα* deletion site (Fig. 3.1D).

3.3.2. Origin of the inserted DNA fragment in *PTENα*

Next, I investigated the origin of the inserted DNA fragment (Fig. 3.2) by assembling the unmatched sequences of the soft-clipped reads at the deletion site of *PTENα* to determine a plausible sequence of the inserted DNA fragment. I then searched for the sequence against the human genome by using BLAT (Kent et al., 2002). As a result, the inserted sequence matched with two candidate regions: (i) a reverse complement of a region [chr1:569503-569570 (hg19)] within the nuclear mitochondrial sequence (chr1:564465-570304) (Simone D et al., 2011) and (ii) a reverse complement of a part of the mitochondrial genome (chrM: 8955-9022). Considering the combinations of either of the two candidate-inserted segments with the *PTENα* sequence based on the human reference genome could not accurately account for the observed deletion–insertion, it resulted in one unresolved mismatch (G allele). Based on these results, I developed two hypotheses about the structure of the deletion–insertion. Hypothesis 1 assumed a 46-bp deletion and 67-bp insertion, in which the G allele has originated from an alteration at the

breakpoint of *PTENα*, whereas hypothesis 2 assumed a 47-bp deletion and 68-bp insertion, in which the G allele has originated from an alteration in either of the two candidate insertions. Therefore, I scrutinized common genetic polymorphisms deposited in dbSNP. As a consequence, hypothesis 1 was rejected because there was no genetic polymorphism rendering the G allele at the breakpoint of *PTENα* (Fig. 3.2B). When considering hypothesis 2, the two candidate insertions were identical; however, an SNP located in the nuclear mitochondrial sequence on chromosome 1 (rs1198320487: NC_000001.10:g.569503T>C) differentiated the sequences (Fig. 3.2C). Finally, I determined that the G allele had originated from the alternative allele of rs1198320487 in the nuclear mitochondrial sequence on chromosome 1.

3.4. Discussion

The patient with the deletion–insertion in *PTENα* was diagnosed at 42 years of age, which is earlier than the average age of diagnosis in this study. IHC of ER and PR were negative, and I could not retrieve the results of HER2 from the patient's clinical charts. The frequency of the deletion–insertion of *PTENα* was not observed in any of the ExAC and the other publicly available populations. The deletion–insertion was also not registered in either dbSNP or ClinVar, indicating that this mutation was a novel germline mutation. The identified deletion–insertion on exon 1 of *PTENα* was predicted to create a stop codon at the fourth amino acid of the PTENα protein.

PTEN is known to be a tumor suppressor gene. *PTEN* mutations are commonly found in patients with inherited cancer syndromes, such as Cowden syndrome (OMIM158350). PTENα is a translational variant of *PTEN* and has an additional 173 amino acids at the N-terminus, labeled as alternatively translated region (ATR) (Fig. 3.1A).

PTENα prevents cancer growth by antagonizing phosphoinositide-3 kinase signaling as well as canonical *PTEN*. More importantly, ATR contains a protein-binding domain and a cleavage site. These regions allow *PTENα* to bind to the cell membrane and to be released into the extracellular space. Because ATR contains sequences that have homology with known cell-permeable peptides, *PTENα* enters into and acts in neighboring cells. Furthermore, PTENα without the cleavage site could not suppress tumor cell growth as compared with normal PTENα in vivo (Hopkins et al., 2013). From these results, I assumed that this novel protein-truncating mutation in *PTENα* could lead to the development of breast cancer due to the lack of a tumor-suppressive function attributed to the PTENα protein. However, the functional significance of this deletion–insertion on canonical *PTEN* was unknown because the deletion–insertion was located on the 5' untranslated region of canonical *PTEN* (Fig. 3.1A). Two pathogenic germline mutations on the N-terminal residues of *PTENα* were identified in a Chinese cohort of patients with autism spectrum disorder, although a definitive association between *PTENα* and neurodevelopment remains unknown (Zhou et al., 2019).

Based on telephonic interviews and clinical charts, the patient with the deletion–insertion in *PTENα* did not report any family history of breast or ovarian cancer. Although I could not obtain DNA samples from her family members, there is a possibility that the deletion–insertion was inherited through her father's lineage. Other plausible explanations are that the mutation occurred de novo or arose at a very early stage of her development. Further examination of the genotypes of the mutation among her family members together with a review of her status is needed to assess the clinical significance of this novel deletion–insertion.

In conclusion of this chapter, I identified a novel intermediate-sized deletion–

insertion in *PTENα*, which can be a disease risk factor for breast cancer. This deletion–insertion may not be detected by general pipelines targeting SNVs and short INDELs in multigene panel testing. The breakpoint of the deletion and the possible source of the inserted fragment were determined by in-depth analyses. Therefore, my results suggest that patient-specific risk factors can be detected via detailed bioinformatics analyses. I hypothesized that a part of VUSs were harmful mutations that could deleterious to gene functions, and that the burden of rare and harmful VUSs might be associated with breast cancer risk. I demonstrated whether VUS was related to breast cancer susceptibility in Chapter 4. Based on the results of my study in Chapter 1 and 2, I identified pathogenic mutations in a total of 84 breast cancer patients. A large part of the remaining 499 patients harbored VUSs.

Table 3.1. Primers for Sanger sequencing of the detected genes

Target	Forward sequence (5' to 3')	Reverse sequence (5' to 3')
PTENα	ATGTGGCGGGACTCTTTATG	GGAATGGGGAGAAGACGAAT
APC	TTTGCAGGTTATTGCGAGTG	TGCGGTACTCTAAAACTATGGACT

Figure 3.1

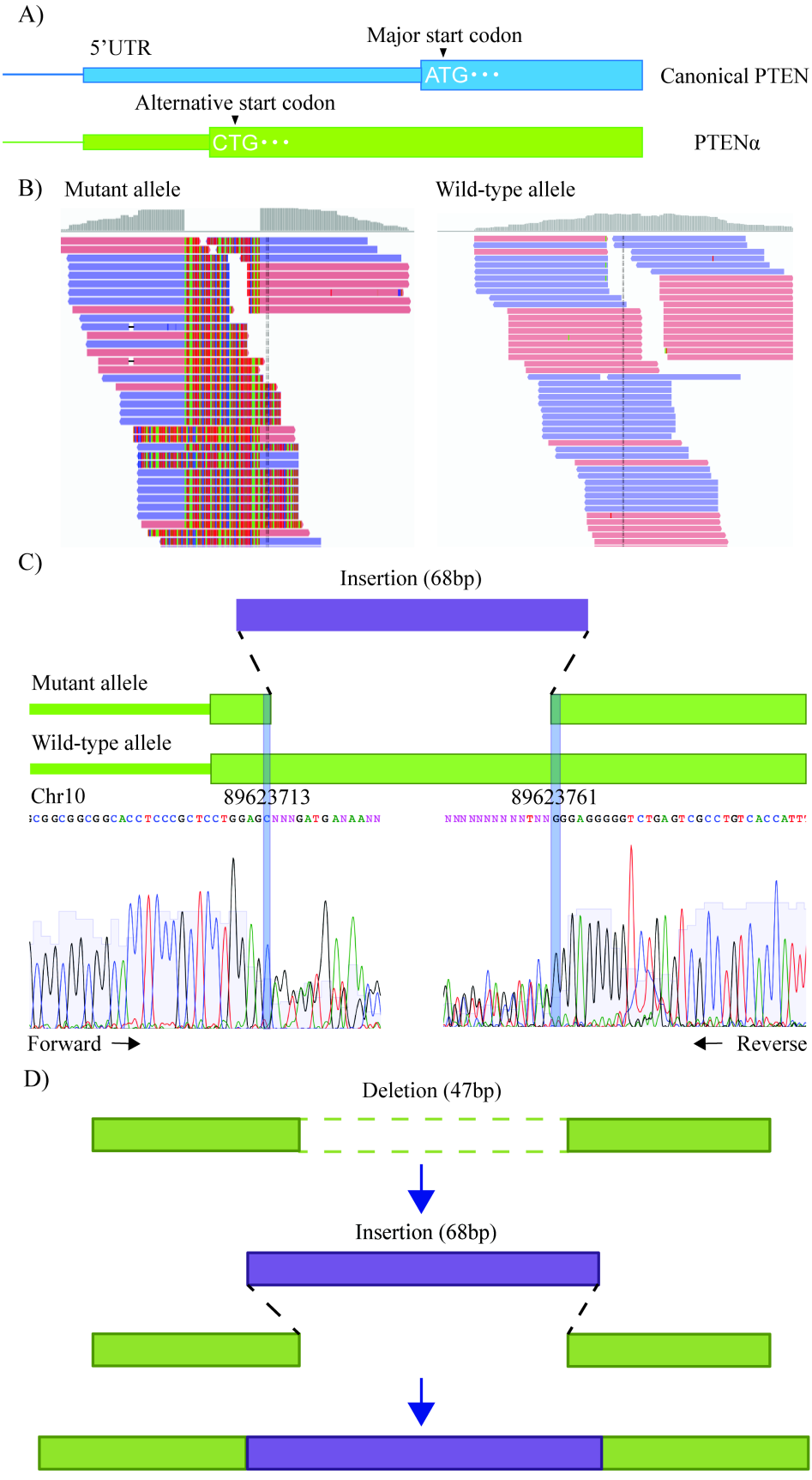

A)
5'UTR
Major start codon
ATG•••
Canonical PTEN
Alternative start codon
CTG•••
PTENα
B)
Mutant allele
Wild-type allele
C)
Insertion (68bp)
Mutant allele
Wild-type allele
Chr10
89623713
89623761
Forward
Reverse
D)
Deletion (47bp)
Insertion (68bp)

Figure 3.1. Overview of the novel intermediate-sized INDEL in PTENα.

A) Differences in the structure between canonical PTEN (top) and PTENα (bottom). PTENα has an alternative start codon (CTG).

B) Detection of deletion junctions in PTENα. The alignment result of soft-clipped reads derived from the mutant allele (i.e., deletion–insertion) is shown. For comparison, the alignment result of reads that were not soft-clipped from the wild-type allele is also shown. The sequences color-coded by light red and blue indicate the sequences matched with that of the human reference genome. Highlighted bases indicate the sequences that are mismatched with the reference genome (i.e., inserted sequences).

C) A schematic representation with plausible junctions of the deletion–insertion in PTENα. The purple bar indicates an inserted sequence (68 bp), whereas the light green bars indicate PTENα sequences. Sanger sequencing confirms the breakpoints of the deletion–insertion. Sanger sequencing using forward (left) and reverse (right) primers reveal aberrant electropherograms after the breakpoints of the deletion–insertion because the fluorescent signals from mutant and wild-type alleles are mixed. The breakpoints of the deletion–insertion are depicted as blue vertical lines.

D) A plausible mechanism of deletion and insertion at the same position. Light green and purple sequences indicate the reference and inserted sequences, respectively. Double-strand DNA breaks may result in a 47-bp deletion accompanied by a 68-bp insertion.

Figure 3.2

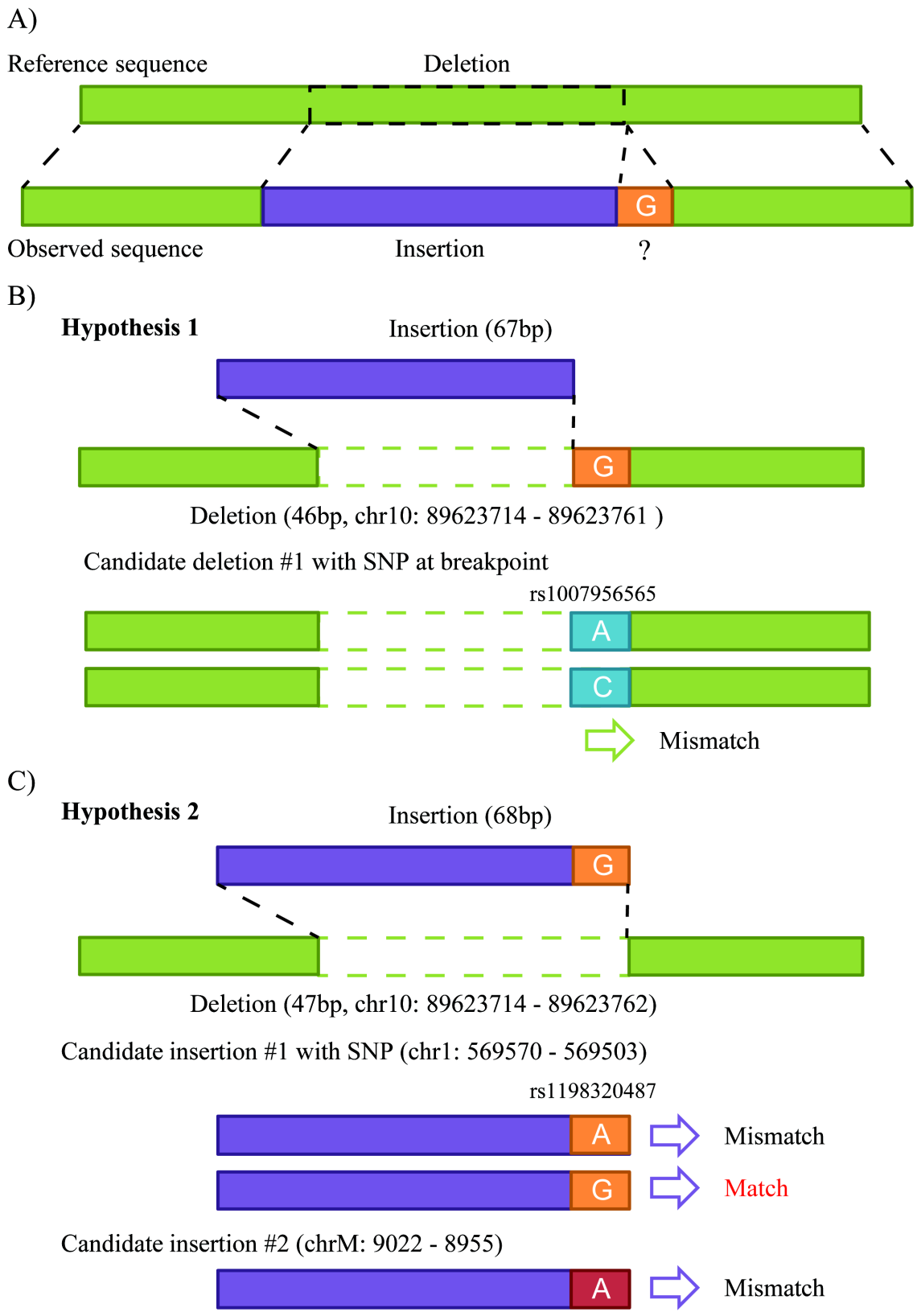

A)
Reference sequence
Deletion
G
Observed sequence
Insertion
?
B)
Hypothesis 1
Insertion (67bp)
G
Deletion (46bp, chr10: 89623714 - 89623761)
Candidate deletion #1 with SNP at breakpoint
rs1007956565
A
C
Mismatch
C)
Hypothesis 2
Insertion (68bp)
G
Deletion (47bp, chr10: 89623714 - 89623762)
Candidate insertion #1 with SNP (chr1: 569570 - 569503)
rs1198320487
A
Mismatch
G
Match
Candidate insertion #2 (chrM: 9022 - 8955)
A
Mismatch

Figure 3.2. Inference on the origin of the inserted sequence by the two hypotheses.

A) Differences in the structure between the reference (top) and observed (bottom) sequences. Light green and purple sequences indicate the reference and inserted sequences, respectively. The unresolved mismatch (G allele) around the breakpoint of PTENα is highlighted by an orange box.

B) Hypothesis 1 for the origin of the unresolved G allele. Hypothesis 1 assumes a 46-bp deletion and 67-bp insertion, in which the G allele arises from an alteration in the PTENα sequence by an SNP (rs1007956565, A/C). There are two possible sequences by the SNP (rs1007956565) at the breakpoint of PTENα; however, these two sequences cannot account for the unresolved mismatch (G allele).

C) Hypothesis 2 for the origin of the unresolved G allele. Hypothesis 2 assumes a 47-bp deletion and 68-bp insertion, in which the G allele arises from an alteration within the inserted sequence (purple). The two candidate sources of the inserted sequence are as follows: (i) a reverse complement of a region (chr1:569503-569570) within a nuclear mitochondrial sequence (chr1:564465-570304) and (ii) a reverse complement of a part of the mitochondrial genome (chrM: 8955-9022). These two candidate regions have identical sequences; however, there is an SNP (rs1198320487, A/G on the reverse strand) in the nuclear mitochondrial sequence on chromosome 1. One of the two possible sequences by the SNP (rs1198320487) can account for the unresolved mismatch (G allele). As a result, the source of the inserted sequence is likely to be the reverse complement of a region (chr1:569503-569570) within the nuclear mitochondrial sequence (chr1:564465-570304) with the alternative G allele at the SNP rs1198320487 site.

Chapter 4. Reclassification of VUS highlights increased burden of harmful variants in breast cancer patients

4.1. Introduction

NGS technologies together with target enrichment systems dramatically reduce the cost to determine the sequences of multiple genes simultaneously. Thus, multigene panel testing allows us to detect pathogenic mutations that are not discovered by single gene testing. On the other hand, several drawbacks of multigene panel testing have been considered. There are significant differences among commercially available multigene panel tests in terms of the number of genes to be analyzed, and the variant classification protocol. Most importantly, the use of multigene panel produces a large number of VUSs that have unknown functional effects and uncertain associations with cancer risk (Chang et al., *in press*). The classification of VUS is one of the most challenging issues in cancer genomics (Maxwell et al., 2016).

When considering the association with cancer risk, the identified variants are classified into five pathogenicity tiers: benign, likely benign, VUS, likely pathogenic, and pathogenic (Richards et al., 2015). In general, common variants that are frequently observed in general populations are considered to be benign. Additionally, synonymous and intronic variants that do not reliably affect protein sequence and abundance are considered to be benign. Pathogenic variants take the form of nonsense mutations, short frameshifting INDELs, larger gene rearrangements, and splicing alterations that all truncate or remove important domains of breast cancer associated proteins. In addition, missense substitutions, which are confidently predicted to disrupt the function of breast cancer associated proteins, may lead to increased risk of breast cancer and classified into pathogenic variants (Lindor et al., 2012). Therefore, VUSs are mainly rare and missense variants that result in single amino acid changes with unknown functional impacts.

The effects of missense substitutions on disease pathogenicity cannot be easily

predicted compared to nonsense or splicing mutations. However, a small fraction of missense variants at important domains may affect protein functions and lead to cancer predisposition. Current *in silico* prediction tools for the classification of missense mutations mainly rely on the information about difference in biochemical properties between reference and substituted amino acids and evolutionary conservation across species, represented by Align-GVGD (Mathe et al., 2006), Polyphen2 (Adzhubei et al., 2010), SIFT (Kumar et al., 2009), and MutationTaster2 (Schwarz et al., 2014). A multifactorial likelihood classification model was developed for *BRCA1/2* variant classification, where the prior probability of pathogenicity of missense variants predicted by means of biochemical properties and evolutionary conservation via the abovementioned tools are combined with likelihood ratios based on clinical information, including variant co-occurrence with a known pathogenic mutation in the same gene, segregation analysis, tumor features and functional assays (Lindor et al., 2012).

Furthermore, it is necessary to consider the fact that breast cancer is not a single-gene Mendelian disorder but a multifactorial disease caused by a complex interplay among genetic, epigenetic and environmental factors. While truncating mutations in DNA repair pathway genes are highly penetrant for breast cancer, it is possible that a part of missense variants are moderately penetrant alleles. Combined effects of such missense variants with other genetic and/or environmental risk factors may lead to the development of breast cancer.

It is thought that mutations that affect the functions of breast cancer-related genes involved in DNA repair pathways are not evolutionarily advantageous and are considered to be very rare in the general population. This idea is reflected in current knowledge on the genetic architecture of breast cancer risk, in which high-penetrance protein-truncating

variants are very rare but common variants identified by genome-wide association studies have weak effects on breast cancer risks (Hindorff et al., 2011). It may be probable that rare missense mutations that affect the functions of cancer-related genes are as-yet-unidentified genetic risk factors for breast cancer. Such variants are to be over-represented in breast cancer patients, compared to those in the general population.

The biggest problem in multi-gene panel inspection is the handling of VUS. Many of the low-frequency missense mutations are considered VUS because their clinical significance is unknown, and are not used to determine treatment strategies. In multigene panel testing, there are reports that as many as 30% to 40% of patients have VUS. According to the result of my analysis in Chapter 2, 77.0% of the patients with breast cancer had VUSs in 54 cancer-related genes. There is a possibility that a small fraction of VUSs detected in my multigene panel are harmful (hereinafter called as "harmful VUS"). I hypothesized that the burden of such harmful VUSs would be associated with breast cancer risk.

Multigene panel used in this study contains several genes that have been suggested to be associated with breast cancer but have not been fully proven. Although such multigene panel may be useful to evaluate prevalence of mutation carriers in Chinese breast cancer patients, it does not seem to be appropriate to evaluate effects of missense variants in the genes whose evidence of associations with breast cancer risk are suggestive. The NCCN consortium provides a list of genes that are highly evidenced for their associations with breast cancer and have actionable medical management options (NCCN Guidelines, 2018). Therefore, I focused on the 11 genes in the NCCN guidelines to reclassify VUSs.

The objective of this chapter is to elucidate the association of the burden of rare and harmful VUSs with breast cancer risk. For the purpose, I aim to formulate an analytical

flamework from prioritization of rare and harmful missense variants to burden testing of the selected variants in breast cancer patients whose etiologies are not clarified. Furthermore, the ultimate goal is to utilize the variants that are not fully explored in current applications of multigene panel testing for comprehensive search for genetic predisposition to breast cancer.

4.2. Materials and Methods

4.2.1. Study population

The information on study subjects and target-gene sequencing is described in Chapter 2.2. Briefly, 583 Han Chinese patients with breast cancer were recruited between December 2016 and September 2017 at the First Affiliated Hospital of Chongqing Medical University and Affiliated Cancer Hospital & Institute of Guangzhou Medical University. All patients provided informed consent for participation in the study. The Ethics Committees of the First Affiliated Hospital of Chongqing Medical University, the Affiliated Cancer Hospital & Institute of Guangzhou Medical University, and the National Institute of Genetics approved the study protocols.

As ethnicity-matched control samples, I used whole-genome sequencing data for 105 South and 103 North Han-Chinese individuals from 1KG (1000 Genomes Project Consortium, 2015). VCF file for these 208 individuals were downloaded by using Data Slicer tool (http://grch37.ensembl.org/Homo_sapiens/Tools/DataSlicer).

4.2.2. Analysis for sequencing data

NGS data processing and variant calling were performed using BWA version 0.7.15

(Li et al., 2010) and GATK version 3.4-46 (McKenna et al., 2010; DePristo et al., 2011) as described in the previous section. Functional annotation was implemented using ANNOVAR (version released 29 September 2017) (Wang et al., 2010). Estimates of variant frequencies in the general populations were based on publicly available databases provided by the Genome Aggregation Database (gnomAD) (Karczewski et al., 2019). Variants with previously established pathogenic or benign effects were explored based on the ClinVar database version 20180521 (Landrum et al., 2016).

4.2.3. Variant classification

As conducted in the previous section, variants whose frequencies were greater than 1% in any of gnomAD populations were considered to be benign. Nonsense and splice site SNVs and frameshifting INDELs were considered to be pathogenic. I sought for information about the pathogenicity of missense variants in the ClinVar database. The remining missense variants whose frequencies were less than 1% and whose impacts on pathogenicity were unknown were classified as VUS.

I examined reclassification of the detected VUSs in order to select harmful VUSs using the following procedure. First, I selected VUSs in *BRCA1/2* or 11 genes that were reported to be strongly associated with breast cancer by the NCCN guideline (NCCN Guidelines, 2018). In other words, I excluded any VUSs in genes whose evidence of associations with breast cancers were not fully established. Second, I scored the VUSs based on their frequencies: i) raw VUSs were less than 1% and ii) ultrarare VUSs were less than 0.1% in any of gnomAD populations. Third, ultrarare VUSs were classified by rare exome variant ensemble learner (REVEL) score (Ioannidis et al. 2016). REVEL is an ensemble method for predicting the pathogenicity of missense variants by integrating

results from the following 13 prediction tools (MutPred, FATHMM, VEST, PolyPhen, SIFT, PROVEAN, MutationAssessor, MutationTaster, LRT, GERP, SiPhy, phyloP, and phastCons). I applied the above mentioned VUS reclassification method to 583 breast cancer patients and 208 controls.

4.2.4. Localization of mutations

For the 11 NCCN genes, amino acid sequences based on the canonical transcripts below were obtained from Uniprot database (https://www.uniprot.org/): *ATM* (NM000051), *BRCA1* (NM007297), *BRCA2* (NM000059), *CDH1* (NM004360), *CHEK2* (NM001257387), *NBN* (NM002485), *NF1* (NM000267), *PALB2* (NM024675), *PTEN* (NM001304717), *STK11* (NM000455) and *TP53* (NM001276761). Domain information was predicted by Pfam database (https://pfam.xfam.org/). For each amino acid position, conservation score was calculated by the ConSurf Server (http://consurf.tau.ac.il/).

4.2.5. Statistical analyses

Association studies for common genetic variants generally assess whether allele frequency of a single SNV or INDEL is different between cases and controls by using Fisher's exact test or χ^2 test (blue squares in Figure 4.1). However, the statistical power to detect a significant association with a single rare variant is poor due to the small number of carrier of the variant (Raychaudhuri, 2011). To overcome this issue, analytical approaches collectively known as "burden tests" have been developed, where the combined effect of multiple rare variants is examined (Lee et al., 2014). As illustrated in Figure 4.1, all rare variants are singletons and therefore association test for any variant lacks of the statistical power. If the combined effect of multiple harmful variants was tested, the number of

individuals with harmful variants is larger in breast cancer patients than that in normal healthy controls (Figure 4.1). In other words, it can be statistically evaluated whether the burden of harmful variants is larger in breast cancer patients than that in controls.

In the context of the burden test, the critical issue is that researchers should predefine what variants are combined into the burden test (Price et al., 2010). I examined three inclusion criteria to evaluate the combined effect of reclassified VUS on breast cancer risk: i) raw VUS; ii) ultrarare VUS; and iii) ultrarare VUS satisfying several REVEL score cutoff values.

The burden of each individual was tested by means of the cohort allelic sum test (CAST), where the presence of any rare variants passing each of the thresholds is assumed to increase disease risk and frequencies of individuals harboring at least one variant passing the threshold are compared between breast cancer patients and normal healthy controls (Morgenthaler & Thilly, 2007). Generally, the burden test evaluates the combined effect of multiple variants in a gene or region. However, the statistical power is still limited because the number of individuals carrying ultrarare variants are very small. Then, I propose an aggregation test where the cumulative effect of multiple variants in a set of genes: i) *BRCA1/2*; and 11 NCCN genes.

For dichotomous variables, I used Fisher's exact test. I estimated the OR and 95% CI. The Wilcoxon–Mann–Whitney test was used to evaluate differences in quantitative variables. I conducted all standard statistical tests with the R program (http://www.r-project.org). The threshold of statistical significance was set at $P < 0.05$.

4.3. Results

4.3.1. The distribution of VUS

The distribution of VUS among the 54 genes is shown in Figure 4.2. VUS was detected in all the 54 genes except for *CDK4*. The VUS in *BRCA1/2* accounted for 11.7% of the overall VUS. Figure 4.3 shows that the number of patients with at least one VUS increases as the number of target genes increased. In particular, the proportion of patients with VUS in 54 genes (56.2%) was much larger than that in the NCCN genes (29.3%) (Figure 4.3 B & C).

4.3.2. Evaluation of ultrarare VUS in *BRCA1/2*

In order to increase the statistical power, I tested the association between VUS and breast cancer by combining multiple rare variants in a set of genes based on the CAST. The frequency of raw VUS (<1%) in *BRCA1/2* was not different between breast cancer patients and normal healthy controls (P = 0.72; OR, 0.91; 95% CI, 0.56-1.50) (Table 4.1). When focusing on ultrarare VUS (<0.1%) in *BRCA1/2*, the frequency was slightly higher in breast cancer patients particularly in patients without pathogenic variants (POP) than that in normal healthy controls (P = 0.36; OR, 1.41; 95% CI, 0.73-2.93) (Table 4.2). Furthermore, the burden of ultrarare VUSs with predicted harmful effects (REVEL score > 0.5) was three-fold higher in POP than that in controls, although the difference was not statistically significant (P = 0.17; OR, 2.97; 95% CI, 0.67-27.16) (Table 4.3).

Motivated by the result showing that patients with pathogenic mutations in *BRCA1/2* were likely to develop TNBC, I examined an association between ultrarare VUS with REVEL score above 0.5 and TNBC. As a result, I found a suggestive evidence that the burden of ultrarare harmful VUS was higher in patients with TNBC than that in controls (P = 0.071; OR, 5.57; 95% CI, 0.62-68.31). Additionally, POP with ultrarare harmful VUS in *BRCA1/2* developed breast cancer 2.9 years earlier than POP without harmful VUS (P =

0.21) (Figure 4.4).

4.3.3. Evaluation of VUS in 11 NCCN genes

When focusing only on *BRCA1/2*, the number of samples with ultrarare variants was very small. Therefore, I broadened the number of targeted genes. As high-risk genes, I selected 11 genes (*ATM, BRCA1, BRCA2, CDH1, CHEK2, NBN, NF1, PALB2, PTEN, STK11, TP53*) based on the NCCN guideline (NCCN Guidelines, 2018).

I evaluated difference in the frequency of raw VUS between patients and controls by means of the CAST. Comparison of the frequency of raw VUS between breast cancer patients and controls did not show statistically significant difference (P = 0.43; OR, 1.16; 95% CI, 0.82-1.68) for number of VUS (Table 4.4). When I stratified the patients into POP and PWP, there was no difference in the frequency of VUS for comparison between PWP and controls (P = 0.31; OR, 0.71; 95% CI, 0.37-1.31) and for comparison between POP and controls (P = 0.22; OR, 1.26; 95% CI, 0.87-1.82).

4.3.4. Evaluation of ultrarare VUS in 11 NCCN genes

I considered whether very rare VUS was associated with breast cancer risk based on the view that harmful allelic variants are subject to purifying selection pressures (Pritchard, 2001; Kryukov et al., 2007). I sought ultrarare VUS with allele frequency of less than 0.1% in the 11 NCCN genes. I detected 144 ultrarare VUS in 127 patients and 34 ultrarare VUS in 30 controls. The burden of ultrarare VUS was significantly higher in the patients (P = 0.026; OR, 1.65; 95% CI, 1.06-2.64) (Table 4.5). Particularly, POP showed significantly higher burden of ultrarare VUS compared to controls (P = 0.014; OR, 1.74; 95% CI, 1.10-2.80). Although the difference did not reach statistically significant level, the burden was

slightly higher in POP than that in PWP (P = 0.25; OR, 1.46; 95% CI, 0.78-2.92). As a consequence, the difference in the burden between PWP and control was not statistically different (P = 0.72; OR, 1.19; 95% CI, 0.55-2.47).

4.3.5. Evaluation of ultrarare harmful VUS in 11 NCCN genes

I introduced REVEL scores to further narrow down a list of harmful VUS (Figure 4.5). REVEL score takes value from zero to one, in which variants with higher scores are more likely to be deleterious (Ioannidis et al., 2016). As I changed the cutoff value for REVEL score from zero to one in increments of 0.1, OR remarkably increased from 0.5 to 0.8 (Figure 4.5), suggesting that ultrarare variants with harmful effects on these genes were over-represented in breast cancer patients. Notably, ultrarare VUS with REVEL >= 0.6 was observed in breast cancer patients but was absent in healthy controls. The burden of ultrarare harmful VUS was overrepresented especially when the cutoff value of REVEL was set to be 0.6 (P = 0.0094 for comparison between breast cancer patients and controls; and P = 0.0079 for comparison between POP and controls).

The developers of REVEL did not explicitly determine the cutoff value to distinguish deleterious variants from benign ones. However, they used a score above 0.5 to evaluate the predictive ability of REVEL (Ioannidis et al., 2016). Therefore, I considered variants with REVEL score above 0.5 as harmful in this study. The number of individuals carrying ultrarare variants with harmful effect was three times greater in breast cancer patients than that in controls (P = 0.033; OR, 2.96; 95% CI, 1.03-11.66) under the condition of REVEL score above 0.5 (Table 4.6). The burden of ultrarare harmful VUS was significantly higher in POP than that in controls (P = 0.030; OR, 3.03; 95% CI, 1.04-12.03). This result showed the OR of combined effect of ultrarare variants for breast cancer risk

clearly increased by incorporating REVEL score (3.03 versus 1.74).

Additionally, I evaluated whether the burden of ultrarare VUS with predicted harmful effect is strongly associated with specific subtypes of breast cancer or prognostic factors (Table 4.7). I showed a suggestive evidence that POP with high-REVEL VUS were likely to develop TNBC (P = 0.071; OR, 3.75; 95% CI, 0.68-20.84). For prognostic factors, POP with high-REVEL VUS showed stronger associations with blood vascular invasion (P = 0.038; OR, 6.04; 95% CI, 0.84-37.98) and high histological grade (P = 0.028; OR, 4.22; 95% CI, 0.97-20.95).

I examined the distribution of VUS along the protein domains of each gene. I found that ultrarare VUS with high REVEL score were located on the highly conserved region of each gene (Figure 4.6). These results suggest that missense mutations, which are extremely rare in the general population and predicted to affect gene functions by REVEL, can be a possible strong risk factor for breast cancer.

4.4. Discussion

Most challenging issue in cancer multigene panel testing is the handling of a large number of VUS. VUSs are mostly rare missense variants in cancer predisposition genes. Reclassification of VUS to distinguish from harmful variants from benign ones is required. Mutations that affect the functions of breast cancer-related genes involved in the DNA repair pathway would be not evolutionarily favorable and therefore considered to be very rare in the general population due to purifying selection pressures. In other words, very rare missense mutations that affect the functions of cancer-related genes can be associated with breast cancer risk and highly concentrated in breast cancer patients than in the general population. Based on these, I hypothesized that VUS could be appropriately reclassified by

selecting alleles that were extremely rare in the general population and predicted to be damaging to protein functions with a state-of-the-art bioinformatics tool, REVEL.

Additionally, there was a concern that the statistical power was limited when the burden of variants in a single gene was evaluated. Therefore, I evaluated the combined effect of genes listed in the NCCN guidelines to solve this problem. The NCCN working group has scrutinized current evidence of genetic associations and has proposed a gene list that is highly evidenced for its association with breast cancer and that is clinically worth considering. Moreover, the genes in the NCCN guidelines are all involved in DNA damage repair pathways; therefore, I thought that the burden of variants of genes in the NCCN guidelines could be a good surrogate for the combined effect on DNA damage repair pathways.

By incorporating these two ideas, I formulated the analytical framework from prioritization of rare and harmful missense variants to burden testing in breast cancer patients. As described below, the result of this chapter is novel and therefore this is the first study, to my knowledge, showing significant associations between the burden of VUS and breast cancer.

Firstly, I focused on VUSs in *BRCA1/2*. The combined effect of ultrarare VUSs that were less than <0.1% in the general populations in *BRCA1/2* conferred a modest-to-weak breast cancer risk (OR = 1.4). Notably, the effect size sharply increased (OR = 3.0) by focusing on ultrarare VUS with high REVEL score, demonstrating the usefulness of functional prediction score for the reclassification of VUS. I found that the burden of ultrarare VUS with high REVEL score was particularly higher in patients with TNBC than that in controls (OR = 5.6), which was consistent with the result in Chapter 2 that pathogenic mutations in *BRCA1* were strongly associated with TNBC. Similarly, POP with

VUS in *BRCA1/2* developed breast cancer 2.9 years earlier than POP without VUS. In Chapter 2, I showed that patients with pathogenic mutations developed breast cancer 6.4 years earlier. It is interesting that ultrarare harmful missense variants exerted similar directional effects on the development of breast cancer and its subtypes as pathogenic loss-of-function mutations. These results suggest that functions of *BRCA1/2* and the risk for breast cancer are affected by ultrarare harmful missense variants to weaker extent than by pathogenic loss-of-function mutations. However, I could not obtain statistically significant difference. This is probably due to the small number of samples, which was recognized as a limitation of my study.

To solve this issue, I increased the number of target genes by considering genes that were listed in the NCCN guidelines as having "increased risk of breast cancer" (NCCN Guidelines, 2018). Out of the 11 genes, I identified pathogenic mutations in *BRCA1*, *BRCA2*, *CHEK2*, *NF1*, *PALB2*, *PTEN* and *TP53* in Chapter 2. In addition to these seven genes, I detected harmful VUS in *ATM*, *CDH1*, *NBN*, and *STK11* in which pathogenic variants were not detected in Chapter 2.

There was no difference in VUS load among controls and patients when raw VUS was considered (or reclassification of VUS was not conducted). When I reclassified harmful VUS by selecting ultrarare variants in the NCCN genes according to the prediction score by REVEL, I demonstrated that ultrarare harmful VUS was highly enriched in POP than in controls. In addition, although statistically significant differences were weak, ultrarare harmful VUS tended to be more abundant in POP than that in PWP. It may be partially because that PWP harbored pathogenic loss-of-function mutations of the cancer predisposition genes, which strongly led to develop breast cancer and therefore additional risk factors were depleted. However, the burden of harmful VUS was higher in PWP than

that in controls (OR = 2.5). As it is evident from the presence of cancer-free *BRCA1/2* loss-of-function mutation carriers (Lang et al., 2017), there is a possibility that some of the pathogenic mutations identified in this study are incompletely penetrant alleles and additional harmful VUS increase the likelihood of developing breast cancer. This finding may prompt a rethink of a classical idea of VUS reclassification, where if a VUS is found to co-occur with a pathogenic mutation in a patient, the VUS is thought to be unlikely to be pathogenic (Easton et al., 2007).

The results in this chapter suggest that ultrarare missense mutations that are predicted to affect gene functions by REVEL in the NCCN genes can be a risk factor for breast cancer and strongly associated with TNBC and high histological grade. These findings provide a new possibility that a part of VUSs can be useful for clinical applications, where appropriate treatments can be selected for patients who do not have pathogenic variants but carry harmful VUS. For example, administration of PARP inhibitors may benefit POP with harmful VUS in *BRCA1/2*. The results in this chapter highlight the significance of reclassification of VUS for genetic risk prediction in breast cancer.

I should clarify the limitations of this study. I adopted the burden test, in which the combined effect of multiple variants in a gene or a set of genes was examined. Therefore, I could not directly explain the effect of each of the harmful VUS. In future, I need to experimentally validate functional impact of each harmful variant or collect detailed clinical information for the patients and their family members. Recent studies showed that functional experiments were useful to evaluate the effects of missense mutations on the protein functions of *BRCA1* (Woods et al., 2016; Findlay et al., 2018; Starita et al., 2018) and *PTEN* (Matreyek et al., 2018). An approach that combines experimental methods and

bioinformatics prediction tools can facilitate the appropriate reclassification of VUS in breast cancer research. Such an approach will be useful for the development of more effective treatment for patients with breast cancer stratified by mutation profiling.

Table 4.1. Associations between rare VUS in *BRCA1/2* and affected status. †

	Rare VUS		OR (95% CI)	P-value
	Carrier	Non-carrier		
1KG Chinese	29	179	Reference	-
All patients	75	508	0.91 (0.56-1.50)	0.72
PWP	7	77	0.56 (0.20-1.39)	0.24
POP	68	431	0.97 (0.60-1.62)	0.91

† Rare VUS was defined as missense variants whose frequencies were less than 1% in any gnomAD populations and whose clinical impacts were unknown based on ClinVar. OR and 95% CI were calculated by comparing frequency of carriers between each of three groups of breast cancer patients (i.e., all patients, PWP, and POP) and controls (1KG Chinese). 1KG Chinese was set as referent group.

Table 4.2. Associations between ultrarare VUS in *BRCA1/2* and affected status.[†]

	Ultrarare VUS		OR (95% CI)	P-value
	Carrier	Non-carrier		
1KG Chinese	13	195	Reference	-
All patients	49	534	1.38 (0.72-2.83)	0.37
PWP	6	78	1.15 (0.35-3.39)	0.80
POP	43	456	1.41 (0.73-2.93)	0.36

[†] Ultrarare VUS was defined as missense variants whose frequencies were less than 0.1% in any gnomAD populations and whose clinical impacts were unknown based on ClinVar.

OR and 95% CI were calculated by comparing frequency of carriers between each of three groups of breast cancer patients (i.e., all patients, PWP, and POP) and controls (1KG Chinese). 1KG Chinese was set as referent group.

Table 4.3. Associations between harmful VUS in *BRCA1/2* and affected status.[†]

	Harmful VUS		OR (95% CI)	P-value
	Carrier	Non-carrier		
1KG Chinese	2	206	Reference	-
All patients	16	567	2.90 (0.67-26.25)	0.18
PWP	2	82	2.50 (0.18-35.09)	0.33
POP	14	485	2.97 (0.67-27.16)	0.17

[†] Harmful VUS was defined as ultrarare VUS with high REVEL score (REVEL > 0.5). OR and 95% CI were calculated by comparing frequency of carriers between each of three groups of breast cancer patients (i.e., all patients, PWP, and POP) and controls (1KG Chinese). 1KG Chinese was set as referent group.

Table 4.4. Associations between rare VUS in 11 NCCN genes and affected status. †

	Rare VUS		OR (95% CI)	P-value
	Carrier	Non-carrier		
1KG Chinese	61	147	Reference	-
All patients	190	393	1.16 (0.82-1.68)	0.43
PWP	19	65	0.71 (0.37-1.31)	0.31
POP	171	328	1.26 (0.87-1.82)	0.22

† Rare VUS was defined as missense variants whose frequencies were less than 1% in any gnomAD populations and whose clinical impacts were unknown based on ClinVar. OR and 95% CI were calculated by comparing frequency of carriers between each of three groups of breast cancer patients (i.e., all patients, PWP, and POP) and controls (1KG Chinese). 1KG Chinese was set as referent group.

Table 4.5. Associations between ultrarare VUS in 11 NCCN genes and affected status.[†]

	Ultrarare VUS		OR (95% CI)	P-value
	Carrier	Non-carrier		
1KG Chinese	30	178	Reference	-
All patients	127	456	1.65 (1.06-2.64)	0.026
PWP	14	70	1.19 (0.55-2.47)	0.72
POP	113	386	1.74 (1.10-2.80)	0.014

[†] Ultrarare VUS was defined as missense variants whose frequencies were less than 0.1% in any gnomAD populations and whose clinical impacts were unknown based on ClinVar.

OR and 95% CI were calculated by comparing frequency of carriers between each of three groups of breast cancer patients (i.e., all patients, PWP, and POP) and controls (1KG Chinese). 1KG Chinese was set as referent group.

Table 4.6. Associations between harmful VUS in 11 NCCN genes and affected status.

	Harmful VUS		OR (95% CI)	P-value
	Carrier	Non-carrier		
1KG Chinese	4	204	Reference	-
All patients	32	551	2.96 (1.03-11.66)	0.033
PWP	4	80	2.54 (0.46-13.99)	0.23
POP	28	471	3.03 (1.04-12.03)	0.030

† Harmful VUS was defined as ultrarare VUS with high REVEL score (REVEL > 0.5). OR and 95% CI were calculated by comparing frequency of carriers between each of three groups of breast cancer patients (i.e., all patients, PWP, and POP) and controls (1KG Chinese). 1KG Chinese was set as referent group.

Table 4.7. Associations between harmful VUS and POP stratified by clinical variables.

	Positive POP		Negative POP	
Variables	Odds (95% CI)	P-value	Odds (95% CI)	P-value
ER	3.17 (1.03-13.00)	0.03	2.92 (0.80-13.19)	0.088
PR	2.84 (0.87-12.03)	0.088	3.38 (1.04-14.31)	0.03
HER2	2.54 (0.67-11.75)	0.15	3.21 (1.00-13.48)	0.035
TNBC	3.75 (0.68-20.84)	0.071	2.81 (0.92-11.53)	0.075
p53	2.72 (0.88-11.21)	0.073	3.87 (1.05-17.59)	0.037
E-cadherin	3.08 (1.05-12.25)	0.028	0 (0.00-76.81)	1
Calponin	4.04 (0.35-29.88)	0.14	2.91 (0.98-11.69)	0.04
CK5/6	3.31 (0.47-20.27)	0.13	3.03 (1.01-12.20)	0.037
Axillary lymph node status	1.54 (0.22-9.31)	0.69	3.3 (1.11-13.30)	0.023
Bone metastasis	0 (0.00-153.40)	1	2.93 (1.00-11.70)	0.042
Vascular invasion	6.04 (0.84-37.98)	0.038	2.67 (0.89-10.79)	0.084
Histological grade	4.22 (0.97-20.95)	0.028	3.25 (1.05-13.39)	0.027

Figure 4.1

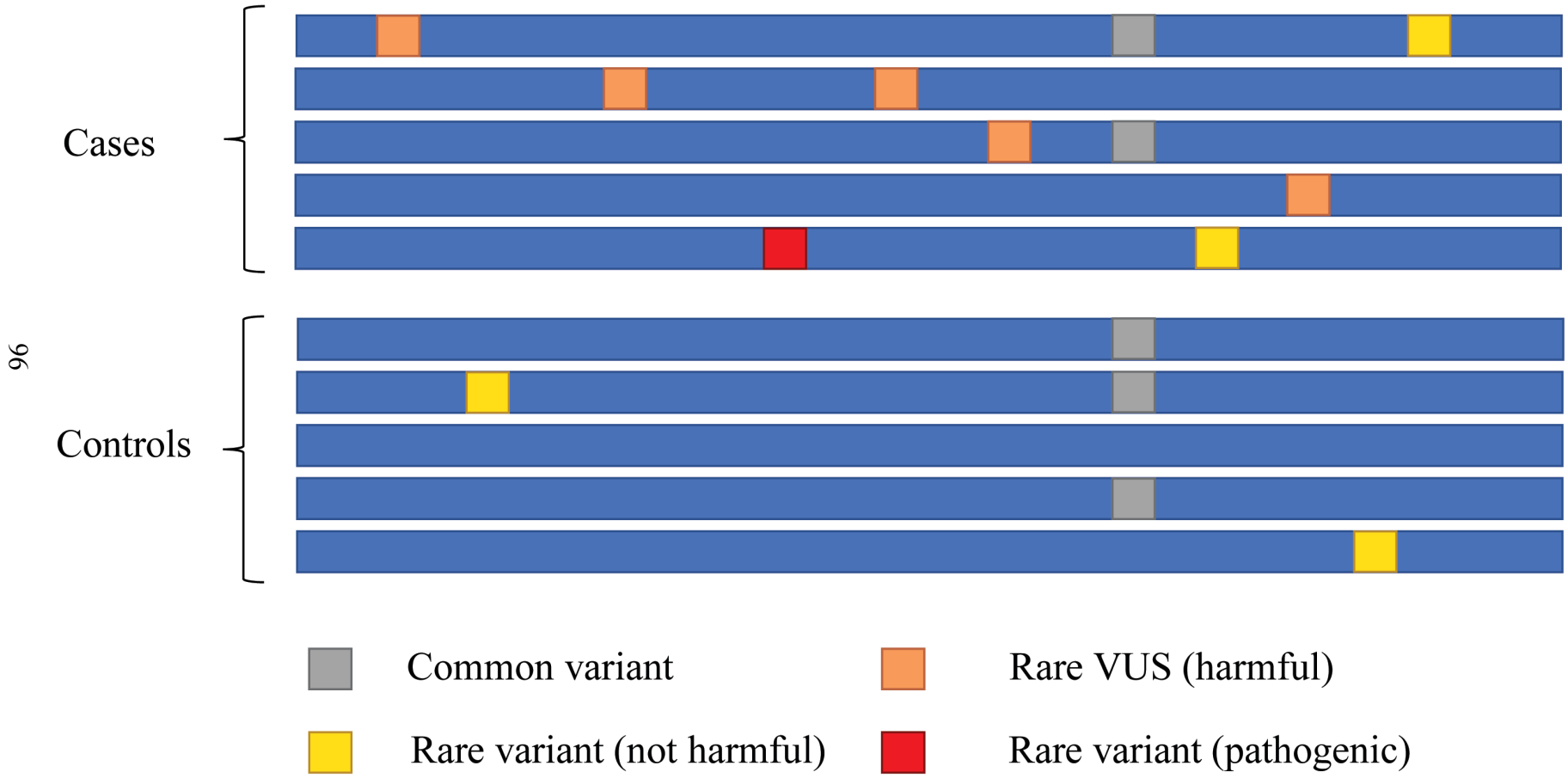

Cases
Controls
Common variant
Rare VUS (harmful)
Rare variant (not harmful)
Rare variant (pathogenic)

Figure 4.1. Schematic of association analysis in common and rare variants.

Upper and bottom panels indicate variant localization of cases and controls, respectively. Blue bar indicates DNA sequences (chromosomes). For a common variant (gray boxes), there are several chromosomes carrying the variant allele at the same position. Pathogenic rare variant (red box) is detected specifically in cases because the variant is highly penetrant to the disease. The number of chromosomes carrying benign rare variants (yellow boxes) at any site is similar between both cases and controls because these variants are not associated with the disease. Harmful rare VUSs (orange boxes) are over-represented in cases because these variants are strongly associated with the disease.

For common variants, association studies can be conducted by comparing allele frequency at single locus between cases and controls. On the other hand, the number of chromosomes carrying rare variants at the same position are quite small and therefore it is not appropriate to conduct association studies at single position separately. Therefore, a burden test is a powerful approach for association analysis of rare variants in which the combined effect of rare variants at multiple sites is examined. If benign VUS are not successfully excluded, the effect of harmful VUS is diminished. The burden test evaluating the combined effect of VUS that are more likely to be deleterious to the protein functions by reclassifying VUS may increase statistical power to detect gene-disease associations.

Figure 4.2

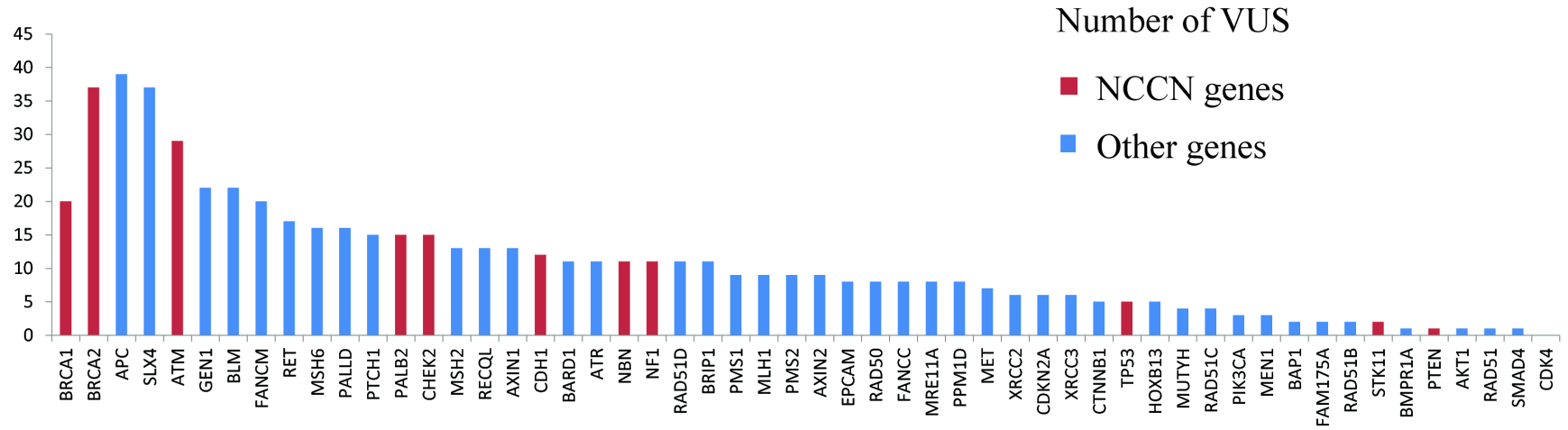


Figure 4.2. Numbers of detected VUSs for 54 genes.
Bars are color-coded as follows: Red for NCCN genes, and blue for the other genes.

Figure 4.3

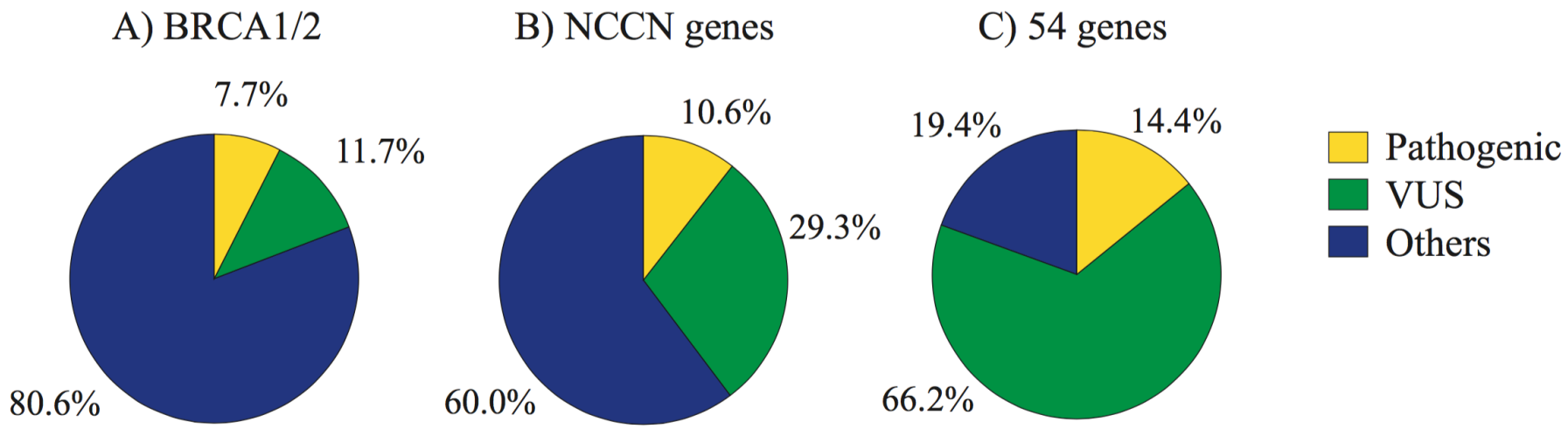


Figure 4.3. Proportions of breast cancer patients with pathogenic variants and VUS.

Pie charts represents proportions of carriers of pathogenic variants, carriers of VUS and non-carriers in A) *BRCA1/2*, B) 11 NCCN genes, and C) 54 genes.

Figure 4.4

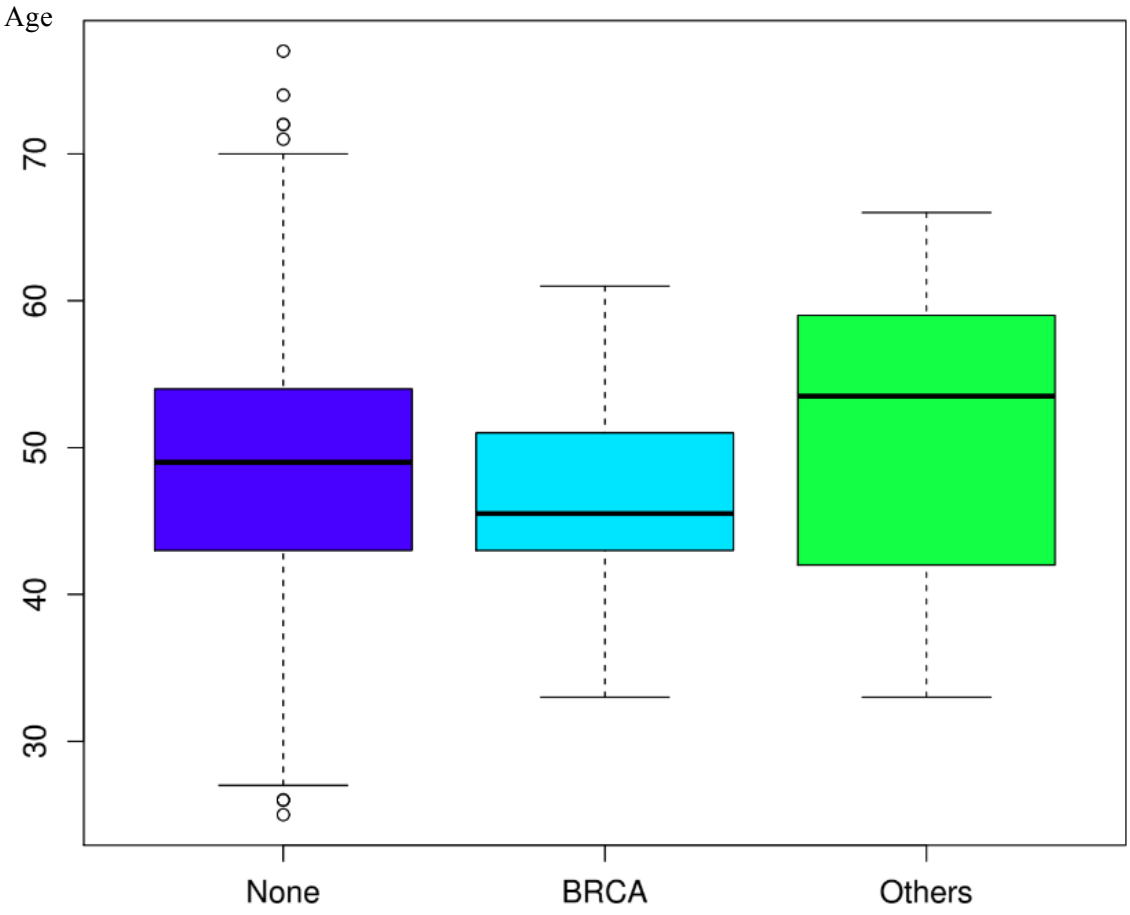
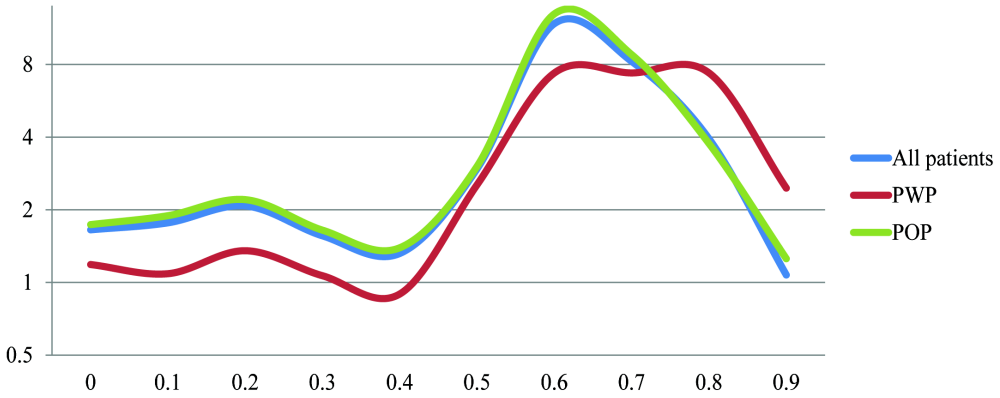


Figure 4.4. Distributions of the age at diagnosis of breast cancer in POP according to mutation status.

Box plot showed the distribution of the age at diagnosis of breast cancer in POP with harmful VUS in each gene groups. Box plot represents the five-number summary: the minimum, first quartile, median, third quartile, and maximum. Points which falls more than 1.5 times the interquartile range above the third quartile or below the first quartile are defined as outlier. Average of ages are 49.2 years old for non-carriers (None), 46.3 years old for *BRCA1/2* mutation carriers (BRCA), 50.1 years old for NCCN genes excluding *BRCA1/2* mutation carriers (Others).

Figure 4.5

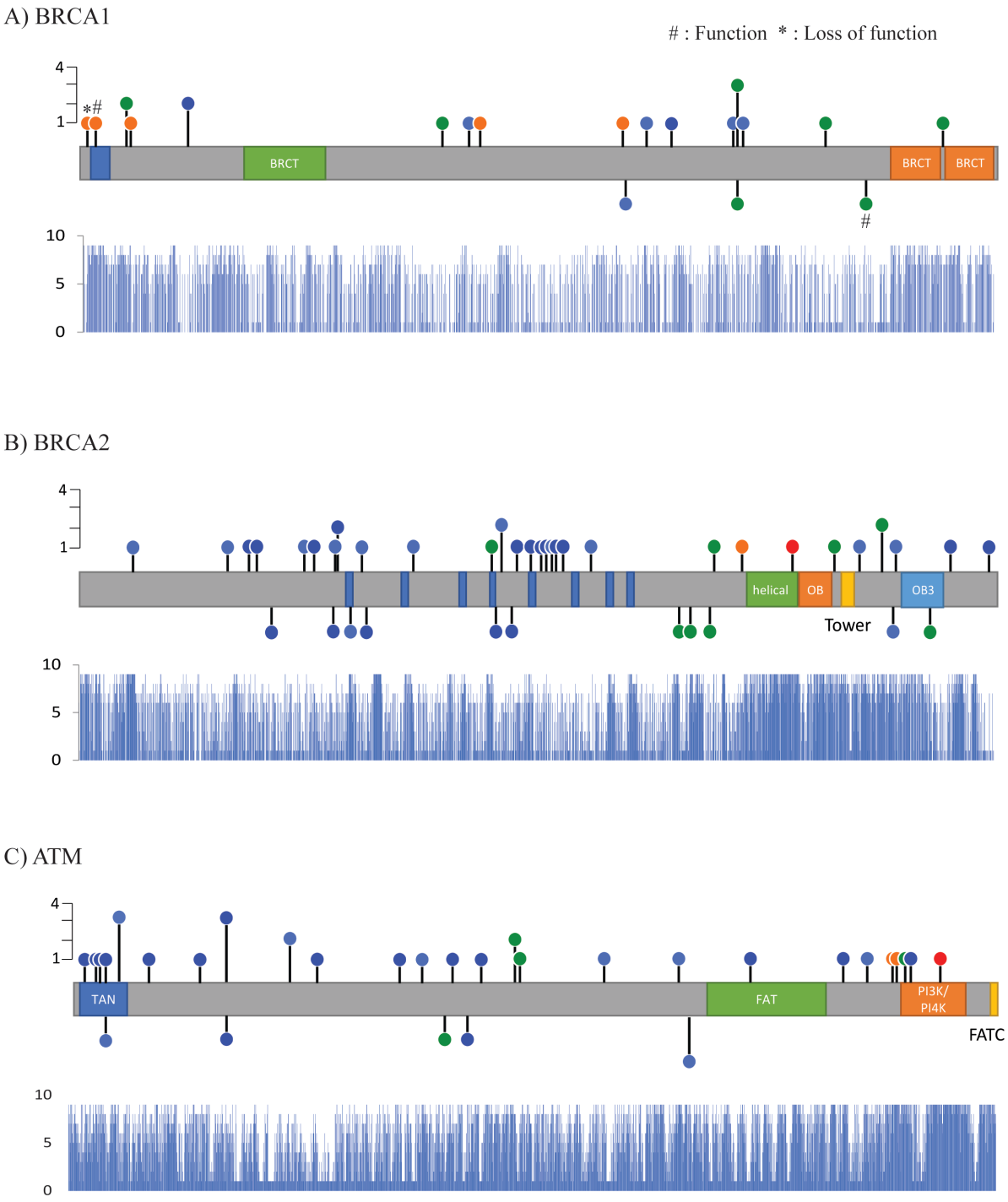


1KG Chinese (208)	30	23	15	14	11	4	0	0	0	0
All patients (583)	127	105	81	59	40	32	16	11	5	1
PWP (84)	14	10	8	6	4	4	1	1	1	0
POP (499)	113	95	73	53	36	28	15	10	4	1

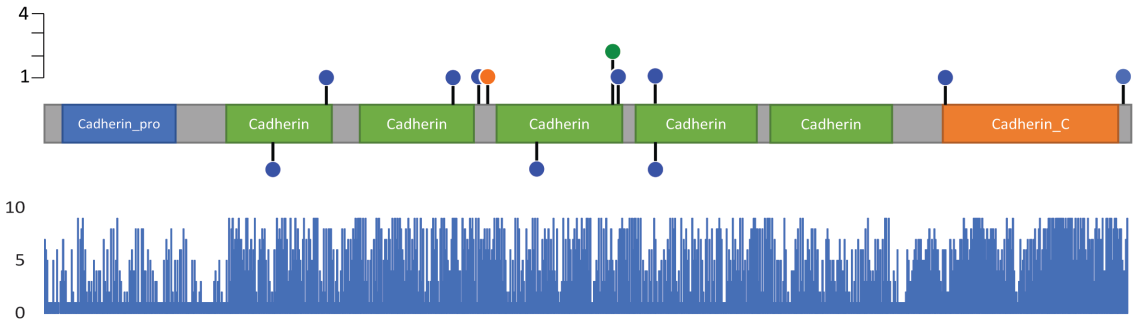
Figure 4.5. Change in OR of ultrarare harmful VUS for breast cancer risk as a function of REVEL score.

Horizontal and vertical axes correspond to REVEL score and OR, respectively. OR were calculated for four types of comparisons: i) All breast cancer patients vs. Controls [blue line], ii) PWP vs. Controls [red line], and iii) POP vs. Controls [green line]. The numbers of individuals with ultrarare VUS satisfying each REVEL score in each of the four groups (1KG Chinese, all breast cancer patients, PWP and POP) are shown in the bottom panel.

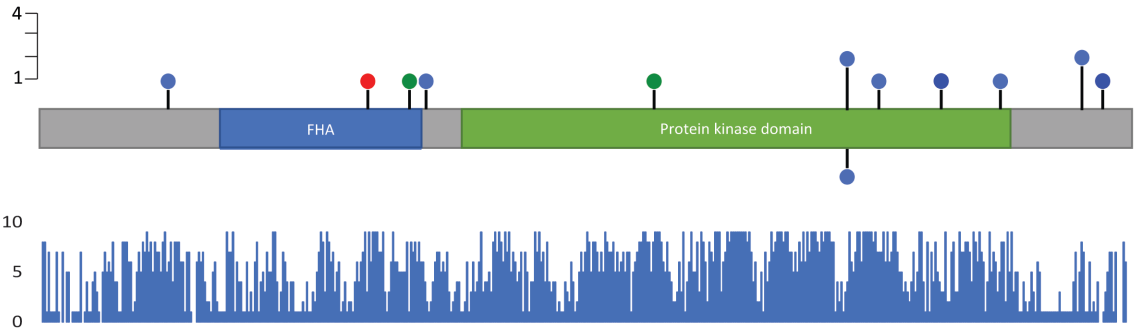
Figure 4.6


A) BRCA1
: Function * : Loss of function
4
1
*#
BRCT
BRCT
BRCT
#
10
5
0
B) BRCA2
4
1
helical
OB
OB3
Tower
10
5
0
C) ATM
4
1
TAN
FAT
PI3K/
PI4K
FATC
10
5
0

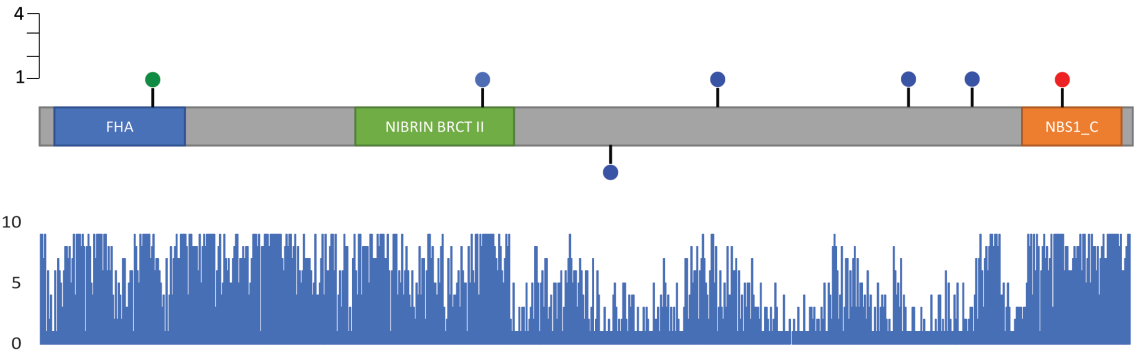
D) CDH1



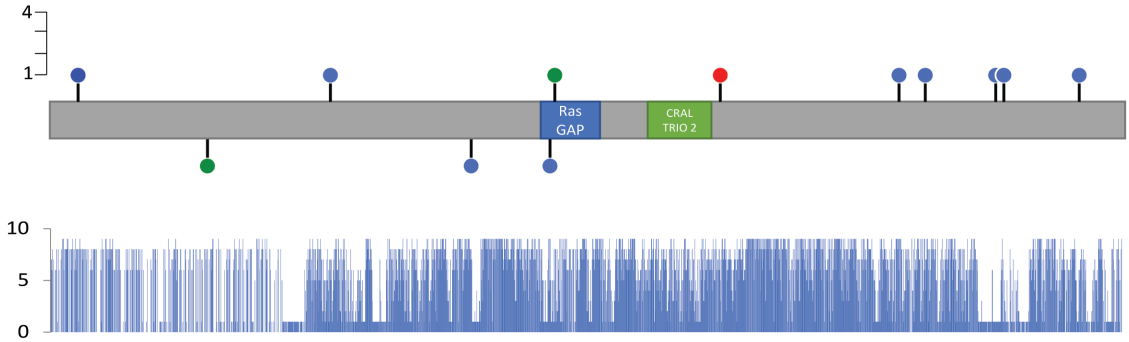
E) CHEK2



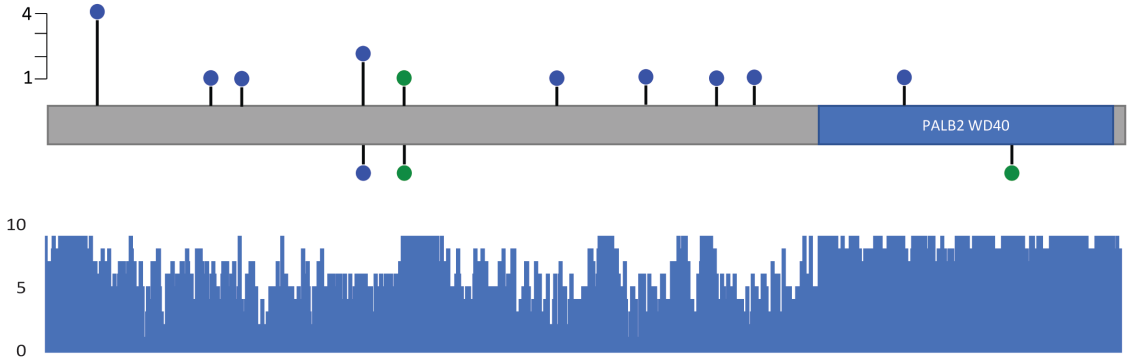
F) NBN



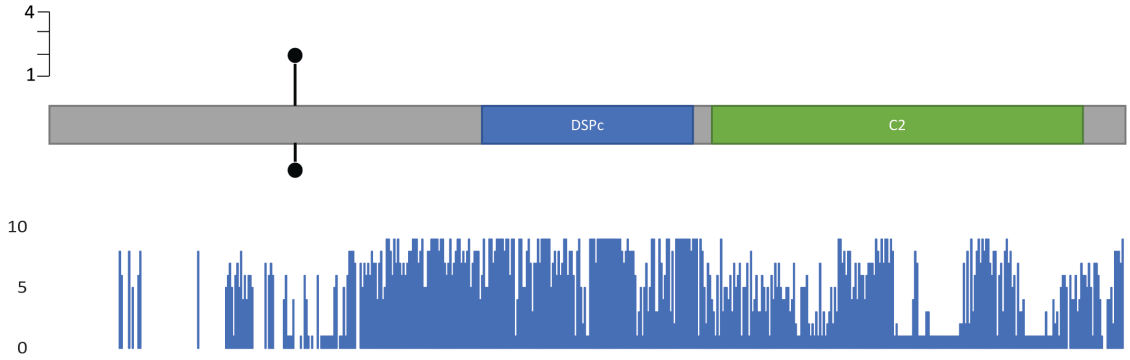
G) NF1



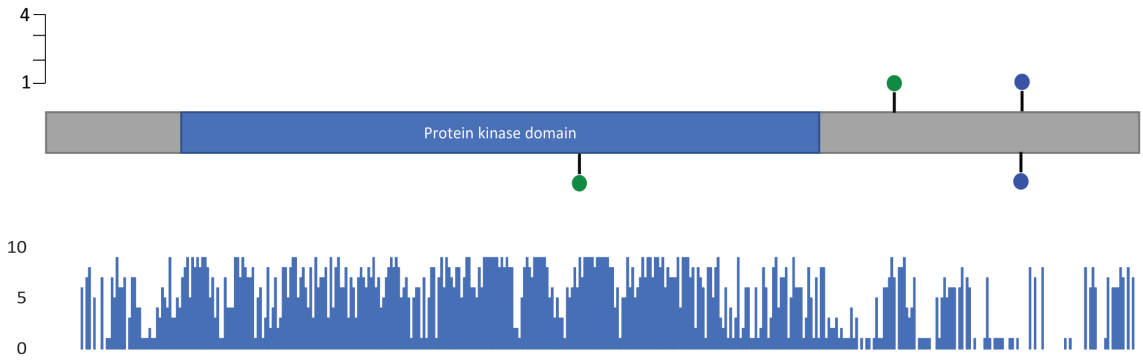
H) PALB2



I) PTEN



J) STK11



K) TP53

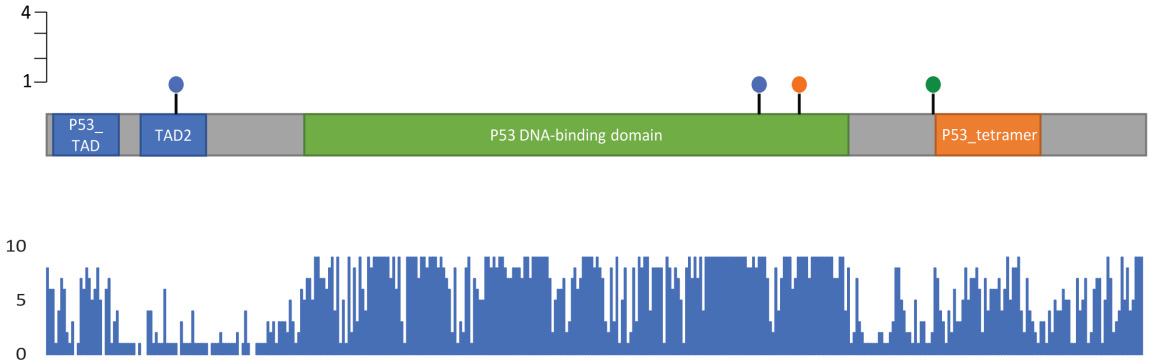


Figure 4.6. Lollipop plots for ultrarare VUSs along the protein domains of 11 NCCN genes.

A lollipop plot shows identified ultrarare VUSs relative to a schematic representation of the protein encoded by the gene. Position with a VUS is depicted by a circle, the length of the vertical line depends on the number of patients with the VUS. The color inside the circle indicates the REVEL score. The REVEL score was displayed in order to red (≧0.8), orange (≧0.6), green (≧0.4), light blue (≧0.2), and blue (≧0) on lollipops. Black indicates a REVEL score was not calculated. In the bottom, histogram showed conservation score at that codon based on the ConSurf Server (http://consurf.tau.ac.il/). Ultrarare VUSs were identified in A) BRCA1, B) BRCA2, C) ATM, D) CDH1, D) CHEK2, E) NBN, F) NF1, G) PALB2, H) PTEN, I) STK11, and J) TP53.

Chapter 5: General discussion

In this doctoral thesis, I investigated genotype-phenotype correlations in Chinese patients with breast cancer using NGS data for 54 breast cancer predisposition genes. I focused on Chinese breast cancer patients as a representative of East Asian population to assess the prevalence of population-specific mutations. The objectives of Chapter 2 are to evaluate the usefulness of genetic testing based on a multigene panel, understand the clinical characteristics of tumors in the patients with germline pathogenic mutations on breast cancer predisposition genes, and to obtain useful findings for clinical applications such as targeted therapy according to mutation status.

Of note, I obtained several novel findings from three viewpoints.

(i) The inclusion of recently identified breast cancer-related genes in a multigene panel testing was useful to identify genetic risk factors in a larger number of patients, demonstrating the utility of the multigene panel testing. It is meaningful to report the analysis results in Han Chinese population, one of East Asian populations, with few reports on multigene panel testing for breast cancer.

(ii) *PALB2* was the most common breast cancer related gene except for *BRCA1/2*. While the frequencies of pathogenic mutations on *CHEK2* and *ATM* were reported to be high in breast cancer patients of European population, patients with mutations in these genes were very low in this study. This may be because there are several founder mutations of *CHEK2* and *ATM* that are specific to patients of European population. Most of the mutations identified in this study were novel mutations that were not registered in the sequence database of ClinVar or the general populations from the 1KG and ExAC projects. Although this study targeted breast cancer patients collected in Chongqing and Guangzhou, China, there was little overlap with the mutations identified by a large-scale study conducted in Beijing (Sun et al., 2017). This result may reflect the fact that the population

structure of Han Chinese people is differentiated by region (Wang et al., 2018). The results of my study shed light on the importance of accumulating mutation information for each population especially from non-European descent populations where multigene panel testing has not been fully reported.

(iii) I discovered several novel clinical findings. First, I demonstrated that *BRCA2* mutation was a risk factor for developing invasive lobular carcinoma characterized by low expression of E-cadherin, and more likely to form high-grade tumors with axillary lymph node metastasis and bone metastasis. Second, the pathogenic mutations on *PALB2* were strongly associated with TNBC. The association between *PALB2* and TNBC has been recently reported in a study based on European population (Shimelis et al., 2018), suggesting that this correlation may be general across multiple populations. Since *PALB2* is a gene involved in DNA repair through the same pathway as *BRCA1/2*, PARP inhibitors can be applied to patients with *PALB2* mutation-positive TNBC as well as *BRCA1/2* mutation-positive TNBC. Third, I obtained a suggestive evidence that *MUTYH* mutations was associated with TNBC. Although verification in a large case-control study is needed, this result is a novel association that has not been reported. A large number of clinical characteristics associated with *BRCA1/2* mutations demonstrated by previous studied were reproduced in this study as described in the main text of Chapter 2.

Through the analyses in Chapter 2, I recognized that two issues emerged: 1) Even after the 54 genes were evaluated, the proportion of the patients whose pathogenesis was not clearly explained was still high, and 2) a large number of VUSs were detected when using the multigene panel testing. I tackled these two issues in the subsequent two chapters.

I focused on the fact that general applications of pathogenic mutation search were limited to SNVs and short INDELs. I hypothesized that exploration of intermediate-sized

INDELs (50 bp to 10,000 bp) in patients with breast cancer was useful to resolve the first issue that emerged from the discussion in Chapter 2. As a result of Chapter 3, I succeeded in identifying a deletion-insertion mutation in *PTENα*, an isoform of *PTEN* that was a tumor suppressor gene. To my best knowledge, this is the first report of a pathogenic mutation on a translational variant of *PTEN* in breast cancer. By examining intermediate-sized INDELs with a state-of-the-art tool together with deep bioinformatics analyses, I illustrated the importance of examining intermediate-sized INDELs in multigene panel testing. In addition to the identification of a novel intermediated-sized INDEL, the novelty of Chapter 3 is that I performed a unique bioinformatics analysis to identify the breakpoint of the deletion and the origin of the inserted sequence.

The biggest problem in multigene panel testing is that VUSs have been detected in about 30% of patients who underwent multigene panel testing, but VUS information has not been fully used to determine treatment strategy. I hypothesized that a small fraction of VUSs detected in the multigene panel testing were harmful mutations and could be involved in the etiology of breast cancer. Thus, I proposed a hypothesis in which the burden of rare and harmful VUSs would be higher in patients with breast cancer than that in unaffected individuals. I formulated a methodology to assess the association between the burden of rare and harmful VUSs and breast cancer risk. I demonstrated that the burden of rare and harmful VUSs significantly contributed to breast cancer susceptibility in Chapter 4. The results showed that VUSs, which were predicted to have a detrimental effect on gene function at a very low frequency, were significantly more abundant in breast cancer patients who did not have other distinct pathogenic variants. This result is the first to incorporate the concept of the burden test and to demonstrate the association between VUS and breast cancer. This illuminates the importance of risk factors that have not been fully explored and

overlooked by conventional approaches, which achieved one of the objectives of my doctoral thesis.

Multigene panel testing in cancers attracts increased attention. For example, a genetic test of *BRCA1/2* was approved to select advanced breast cancer patients for PARP inhibitor therapy in Japan. However, there is a lack of information on breast cancer-related gene mutations in non-European population. It is necessary to evaluate population-specific gene mutation prevalence. Large-scale analyses were performed by using 7,000 and 8,000 breast cancer patients in Japan and China, respectively (Momozawa et al., 2018, Sun et al., 2017). The large-scale analysis in China collected breast cancer patients in Beijing. I analyzed Han Chinese breast cancer patients in Chongqing and Guangzhou, China. The fact that there is little overlap with the identified pathogenic mutations in Beijing showed a genetic difference even within China, suggesting that the investigation in East Asian populations is still insufficient to catalogue a comprehensive list of germline pathogenic mutations.

I performed additional analyses of intermediate-sized INDEL and VUS to explore additional genetic risk factors in breast cancer patients without pathogenic SNVs or short INDELs that led to resulting truncated proteins. There are different promising strategies to study the genotype-phenotype correlations in breast cancer. First strategy is that target regions for searching pathogenic SNVs or INDELs can be extended by using whole-exome and whole-genome sequencing analyses. Such strategy have been intensely conducted since there is a possibility that there are as-yet-unknown breast cancer predisposition genes. Second strategy is that the combinatorial effects of variants with weak effects can be evaluated. There are two possibilities that variants with weak effects can lead to breast cancer. The first is that multiple mutations with weak effects may accumulate on a single

breast cancer predisposition gene, which would result in loss of function of the gene. The second is that mutations with weak effects on multiple genes that are involved in the same biological pathway such as DNA damage repair pathway, which would impair the function of the pathway and lead to the development of breast cancer.

Recently, polygenic risk scores (PRSs) have become the standard for predicting disease risks (Torkamani et al., 2018). PRSs are generally calculated as weighted sum of risk alleles using effect sizes based on genome-wide association studies (GWASs). The PRSs utilizes thousands of risk-conferring variants, most of which are common variants with very small effects. While the risk prediction by PRSs for breast cancer has been well established for European population, there is a problem in the accuracy of PRSs for non-European populations because majority of the GWAS data used for the construction of PRSs were derived from European descent populations (Martin et al., 2019, Mavaddat et al., 2019). It is necessary to establish population-specific PRSs including not only common but also rare variants to improve the validity of PRSs, where germline protein-truncating and harmful missense variants are incorporated.

In this study, I demonstrated links between germline pathogenic variants and molecular subtypes. Further integration of germline and somatic mutations with RNA expression profiles will be more effective in clarifying the relationships between breast cancer molecular phenotypes and mutation status.

References

1. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. (2010) A method and server for predicting damaging missense mutations. Nat Methods. 7(4):248-9. doi: 10.1038/nmeth0410-248.

2. Ahmadloo S, Nakaoka H, Hayano T, Hosomichi K, You H, Utsuno E, Sangai T, Nishimura M, Matsushita K, Hata A, Nomura F, Inoue I. (2017). Rapid and cost-effective high-throughput sequencing for identification of germline mutations of BRCA1 and BRCA2. Journal of Human Genetics. 62(5), 561-567. doi: 10.1038/jhg.2017.5.

3. Antoniou AC, Casadei S, Heikkinen T, Barrowdale D, Pylkäs K, Roberts J, Lee A, Subramanian D, De Leeneer K, Fostira F, Tomiak E, Neuhausen SL, Teo ZL, Khan S, Aittomäki K, Moilanen JS, Turnbull C, Seal S, Mannermaa A, Kallioniemi A, Lindeman GJ, Buys SS, Andrulis IL, Radice P, Tondini C, Manoukian S, Toland AE, Miron P, Weitzel JN, Domchek SM, Poppe B, Claes KB, Yannoukakos D, Concannon P, Bernstein JL, James PA, Easton DF, Goldgar DE, Hopper JL, Rahman N, Peterlongo P, Nevanlinna H, King MC, Couch FJ, Southey MC, Winqvist R, Foulkes WD, Tischkowitz M. (2014). Breast-cancer risk in families with mutations in PALB2. New England Journal of Medicine. 371(17), 1651-1652. doi: 10.1056/NEJMc1410673.

4. Antoniou A, Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL, Loman N, Olsson H, Johannsson O, Borg A, Pasini B, Radice P, Manoukian S, Eccles DM, Tang N, Olah E, Anton-Culver H, Warner E, Lubinski J, Gronwald J, Gorski B, Tulinius H, Thorlacius S, Eerola H, Nevanlinna H, Syrjäkoski K, Kallioniemi OP, Thompson D, Evans C, Peto J, Lalloo F, Evans DG, Easton DF. 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.

5. Bianchini G, Balko JM, Mayer IA, Sanders ME, Gianni L. (2016). Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. Nature Reviews Clinical Oncology. 13(11), 674-690. doi: 10.1038/nrclinonc.2016.66.
6. Bogdanova N, Cybulski C, Bermisheva M, Datsyuk I, Yamini P, Hillemanns P, Antonenkova NN, Khusnutdinova E, Lubinski J, Dörk T. (2009). A nonsense mutation (E1978X) in the ATM gene is associated with breast cancer. Breast Cancer Res Treat. 2009 Nov;118(1):207-11. doi: 10.1007/s10549-008-0189-9.
7. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. (2014). Bioinformatics. 30(15), 2114-2120. doi: 10.1093/bioinformatics/btu170.
8. Carter CL, Allen C, Henson DE. (1989). Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. Cancer. 63(1), 181-187.
9. Castéra L, Krieger S, Rousselin A, Legros A, Baumann JJ, Bruet O, Brault B, Fouillet R, Goardon N, Letac O, Baert-Desurmont S, Tinat J, Bera O, Dugast C, Berthet P, Polycarpe F, Layet V, Hardouin A, Frébourg T, Vaur D. (2014). Next-generation sequencing for the diagnosis of hereditary breast and ovarian cancer using genomic capture targeting multiple candidate genes. European Journal of Human Genetics. 22(11), 1305-1313. doi: 10.1038/ejhg.2014.16.
10. Chang J, Seng S, Yoo J, Equivel P, Lum SS. (2019) Clinical Management of Patients at Risk for Hereditary Breast Cancer with Variants of Uncertain Significance in the Era of Multigene Panel Testing. Ann Surg Oncol. doi:

10.1245/s10434-019-07595-2.

11. Chen S, Parmigiani G. (2007) Meta-analysis of BRCA1 and BRCA2 penetrance. J Clin Oncol. 25(11):1329-33.
12. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M et al. (2016). Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics. 32(8):1220-2.
13. Chiang CWK, Mangul S, Robles C, Sankararaman S. (2018). A Comprehensive Map of Genetic Variation in the World's Largest Ethnic Group-Han Chinese. Mol Biol Evol. 35(11):2736-2750. doi: 10.1093/molbev/msy170.
14. Chong HK, Wang T, Lu HM, Seidler S, Lu H, Keiles S, Chao EC, Stuenkel AJ, Li X, Elliott AM. (2014). The validation and clinical implementation of BRCAplus: a comprehensive high-risk breast cancer diagnostic assay. PLoS One. 9(5), e97408. doi: 10.1371/journal.pone.0097408.
15. Carvalho CM, Lupski JR. (2016). Mechanisms underlying structural variant formation in genomic disorders. Nat Rev Genet. 17(4): 224–238.
16. Cobain EF, Milliron KJ, Merajver SD. (2016). Updates on breast cancer genetics: Clinical implications of detecting syndromes of inherited increased susceptibility to breast cancer. Semin Oncol. 43(5):528-535. doi: 10.1053/j.seminoncol.2016.10.001.
17. Couch FJ, Hart SN, Sharma P, Toland AE, Wang X, Miron P, Olson JE, Godwin AK, Pankratz VS, Olswold C, Slettedahl S, Hallberg E, Guidugli L, Davila JI, Beckmann MW, Janni W, Rack B, Ekici AB, Slamon DJ, Konstantopoulou I, Fostira F, Vratimos A, Fountzilas G, Pelttari LM, Tapper WJ, Durcan L, Cross SS, Pilarski R, Shapiro CL, Klemp J, Yao S, Garber J, Cox A, Brauch H, Ambrosone

C, Nevanlinna H, Yannoukakos D, Slager SL, Vachon CM, Eccles DM, Fasching PA. (2015). Inherited mutations in 17 breast cancer susceptibility genes among a large triple-negative breast cancer cohort unselected for family history of breast cancer. Journal of Clinical Oncology. 33(4), 304-311. doi: 10.1200/JCO.2014.57.1414.

18. Couch FJ, Shimelis H, Hu C, Hart SN, Polley EC, Na J, Hallberg E, Moore R, Thomas A, Lilyquist J, Feng B, McFarland R, Pesaran T, Huether R, LaDuca H, Chao EC, Goldgar DE, Dolinsky JS. (2017). Associations Between Cancer Predisposition Testing Panel Genes and Breast Cancer. JAMA Oncol. 3(9), 1190-1196. doi: 10.1001/jamaoncol.2017.0424.
19. Cybulski C, Carrot-Zhang J, Kluźniak W, Rivera B, Kashyap A, Wokołorczyk D, Giroux S, Nadaf J, Hamel N, Zhang S, Huzarski T, Gronwald J, Byrski T, Szwiec M, Jakubowska A, Rudnicka H, Lener M, Masojć B, Tonin PN, Rousseau F, Górski B, Dębniak T, Majewski J, Lubiński J, Foulkes WD, Narod SA, Akbari MR. (2015). Germline RECQL mutations are associated with breast cancer susceptibility. Nature Genetics. 47(6), 643-646. doi: 10.1038/ng.3284.
20. Cybulski C, Wokołorczyk D, Jakubowska A, Huzarski T, Byrski T, Gronwald J, Masojć B, Deebniak T, Górski B, Blecharz P, Narod SA, Lubiński J. (2011). Risk of breast cancer in women with a CHEK2 mutation with and without a family history of breast cancer. J Clin Oncol. 29(28):3747-52. doi: 10.1200/JCO.2010.34.0778.
21. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. (2011). A

framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics. 43(5), 491-498. doi: 10.1038/ng.806.

22. DeSantis CE, Ma J, Goding Sauer A, Newman LA, Jemal A. (2017). Breast cancer statistics, 2017, racial disparity in mortality by state. CA: A Cancer Journal for Clinicians. 67(6), 439-448. doi: 10.3322/caac.21412.
23. Desmond A, Kurian AW, Gabree M, Mills MA, Anderson MJ, Kobayashi Y, Horick N, Yang S, Shannon KM, Tung N, Ford JM, Lincoln SE, Ellisen LW. (2015). Clinical Actionability of Multigene Panel Testing for Hereditary Breast and Ovarian Cancer Risk Assessment. JAMA Oncology. 1(7), 943-951. doi: 10.1001/jamaoncol.2015.2690.
24. Easton DF, Deffenbaugh AM, Pruss D, Frye C, Wenstrup RJ, Allen-Brady K, Tavtigian SV, Monteiro AN, Iversen ES, Couch FJ, Goldgar DE. Easton DF1, Deffenbaugh AM, Pruss D, Frye C, Wenstrup RJ, Allen-Brady K, Tavtigian SV, Monteiro AN, Iversen ES, Couch FJ, Goldgar DE. (2007) A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes. Am J Hum Genet. 81(5):873-83.
25. Easton DF, Pharoah PD, Antoniou AC, Tischkowitz M, Tavtigian SV, Nathanson KL, Devilee P, Meindl A, Couch FJ, Southey M, Goldgar DE, Evans DG, Chenevix-Trench G, Rahman N, Robson M, Domchek SM, Foulkes WD. (2015). Gene-panel sequencing and the prediction of breast-cancer risk. New England Journal of Medicine. 372(23), 2243-2257. doi: 10.1056/NEJMsr1501341.
26. Eccles DM, Mitchell G, Monteiro AN, Schmutzler R, Couch FJ, Spurdle AB, Gómez-García EB; ENIGMA Clinical Working Group. (2015) BRCA1 and

BRCA2 genetic testing-pitfalls and recommendations for managing variants of uncertain clinical significance. Ann Oncol. 26(10):2057-65.

27. Ellisen LW, Haber DA. Hereditary breast cancer. (1998) Annu Rev Med. 49:425-36.

28. Findlay GM, Daza RM, Martin B, Zhang MD, Leith AP, Gasperini M, Janizek JD, Huang X, Starita LM, Shendure J. (2018). Accurate classification of BRCA1 variants with saturation genome editing. Nature. 562(7726), 217-222. doi: 10.1038/s41586-018-0461-z.

29. Foulkes WD, Stefansson IM, Chappuis PO, Bégin LR, Goffin JR, Wong N, Trudel M, Akslen LA. (2003). Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. Journal of the National Cancer Institute. 95(19), 1482-1485.

30. Frey MK, Kopparam RV, Ni Zhou Z, Fields JC, Buskwofie A, Carlson AD, Caputo T, Holcomb K, Chapman-Davis E (2019). Prevalence of nonfounder BRCA1/2 mutations in Ashkenazi Jewish patients presenting for genetic testing at a hereditary breast and ovarian cancer center. Cancer. doi: 10.1002/cncr.31856.

31. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thürlimann B, Senn HJ; Panel members. (2011). Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. Annals of Oncology. 22(8), 1736-1747. doi: 10.1093/annonc/mdr304.

32. Goodwin S, McPherson JD, McCombie WR. (2016). Coming of age: ten years of next-generation sequencing technologies. Nature Reviews Genetics. 17(6), 333-351. doi: 10.1038/nrg.2016.49.

33. Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, King MC. (1990) Linkage of early-onset familial breast cancer to chromosome 17q21. Science. 250(4988):1684-9.

34. Hemminki K, Vaittinen P, Kyyrönen P. (1998) Age-specific familial risks in common cancers of the offspring. Int J Cancer. 78(2):172-5.

35. Hindorff LA, Gillanders EM, Manolio TA. (2011) Genetic architecture of cancer and other complex diseases: lessons learned and future directions. Carcinogenesis. 32(7):945-54. doi: 10.1093/carcin/bgr056.

36. Hopkins BD, Fine B, Steinbach N, Dendy M, Rapp Z, Shaw J, Pappas K, Yu JS, Hodakoski C, Mense S, Klein J, Pegno S, Sulis ML, Goldstein H, Amendolara B, Lei L, Maurer M, Bruce J, Canoll P, Hibshoosh H, Parsons R. (2013). A secreted PTEN phosphatase that enters cells to alter signaling and survival. Science. 341(6144), 399-402. doi: 10.1126/science.1234907.

37. Houdayer C, Caux-Moncoutier V, Krieger S, Barrois M, Bonnet F, Bourdon V, Bronner M, Buisson M, Coulet F, Gaildrat P, Lefol C, Léone M, Mazoyer S, Muller D, Remenieras A, Révillion F, Rouleau E, Sokolowska J, Vert JP, Lidereau R, Soubrier F, Sobol H, Sevenet N, Bressac-de Paillerets B, Hardouin A, Tosi M, Sinilnikova OM, Stoppa-Lyonnet D. (2012). Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. Humam Mutation. 33(8), 1228-1238. doi: 10.1002/humu.22101.

38. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, Musolf A, Li Q, Holzinger E, Karyadi D, Cannon-Albright LA, Teerlink CC, Stanford JL, Isaacs WB, Xu J, Cooney KA, Lange EM, Schleutker J, Carpten JD,

Powell IJ, Cussenot O, Cancel-Tassin G, Giles GG, MacInnis RJ, Maier C, Hsieh CL, Wiklund F, Catalona WJ, Foulkes WD, Mandal D, Eeles RA, Kote-Jarai Z, Bustamante CD, Schaid DJ, Hastie T, Ostrander EA, Bailey-Wilson JE, Radivojac P, Thibodeau SN, Whittemore AS, Sieh W. (2016) REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am J Hum Genet. 6;99(4):877-885. doi: 10.1016/j.ajhg.2016.08.016.

39. Jemal A, Center MM, DeSantis C, Ward EM. (2010) Global Patterns of Cancer Incidence and Mortality Rates and Trends. Cancer Epidemiol Biomarkers Prev. 19(8):1893-907. doi: 10.1158/1055-9965.EPI-10-0437.

40. Judkins T, Rosenthal E, Arnell C, Burbidge LA, Geary W, Barrus T, Schoenberger J, Trost J, Wenstrup RJ, Roa BB. (2012). Clinical significance of large rearrangements in BRCA1 and BRCA2. Cancer. 118(21): 5210–5216.

41. Judkins T, Leclair B, Bowles K, Gutin N, Trost J, McCulloch J, Bhatnagar S, Murray A, Craft J, Wardell B, Bastian M, Mitchell J, Chen J, Tran T, Williams D, Potter J, Jammulapati S, Perry M, Morris B, Roa B, Timms K. (2015). Development and analytical validation of a 25-gene next generation sequencing panel that includes the BRCA1 and BRCA2 genes to assess hereditary cancer risk. BMC Cancer. 15, 215. doi: 10.1186/s12885-015-1224-y.

42. Kent WJ. (2002) BLAT--the BLAST-like alignment tool. Genome Res. 12(4):656-64.

43. King, MC. (2014). "The race" to clone BRCA1. Science. 343(6178), 1462-1465. doi: 10.1126/science.1251900.

44. King MC, Levy-Lahad E, Lahad A. (2014) Population-based screening for BRCA1 and BRCA2: 2014 Lasker Award. JAMA. 312(11):1091-2.

45. King MC1, Marks JH, Mandell JB; New York Breast Cancer Study Group. (2003). Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. Science. 302(5645), 643-646.

46. Kispert S, McHowat J. (2017) Recent insights into cigarette smoking as a lifestyle risk factor for breast cancer. Breast Cancer (Dove Med Press). 9:127-132. doi: 10.2147/BCTT.S129746.

47. Konrad JK, Laurent CF, Grace T, Beryl BC, Jessica A, Qingbo W, Ryan LC, Kristen ML, Andrea G, Daniel PB, Laura DG, Harrison B, Matthew S, Nicholas AW, Daniel R, Moriel SB, Eleanor GS, Jack AK, Raymond KW, Katherine T, Yossi F, Eric B, Timothy P, Arcturus W, Cotton S, Nicola W, Jessica XC, Kaitlin ES, Emma PH, Zachary Z, O'Donnell-Luria AH, Eric VMl, Ben W, Monkol L, James SW, Christopher V, Irina MA, Louis B, Kristian C, Kristen MC, Miguel C, Stacey D, Steven F, Stacey G, Jeff G, Namrata G, Thibault J, Diane K, Christopher L, Ruchi M, Sam N, Nikelle P, David R, Valentin RR, Andrea S, Molly S, Jose S, Kathleen T, Charlotte T, Gordon W, Michael ET, The Genome Aggregation Database Consortium, Benjamin MN, Mark JD, Daniel GM. (2019) Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv. doi: https://doi.org/10.1101/531210.

48. Kryukov GV, Pennacchio LA, Sunyaev SR. (2007) Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. Am J Hum Genet. 80(4):727-39.

49. Kumar P, Henikoff S, Ng PC. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc.

4(7):1073-81. doi: 10.1038/nprot.2009.86.

50. Kwong A, Shin VY, Ho JC, Kang E, Nakamura S, Teo SH, Lee AS, Sng JH, Ginsburg OM, Kurian AW, Weitzel JN, Siu MT, Law FB, Chan TL, Narod SA, Ford JM, Ma ES, Kim SW. (2016). Comprehensive spectrum of BRCA1 and BRCA2 deleterious mutations in breast cancer in Asian countries. Journal of Medical Genetics. 53(1), 15-23. doi: 10.1136/jmedgenet-2015-103132.
51. Laakso M, Loman N, Borg A, Isola J. (2005). Cytokeratin 5/14-positive breast cancer: true basal phenotype confined to BRCA1 tumors. Modern Pathology. 18(10), 1321-1328.
52. LaDuca H, Stuenkel AJ, Dolinsky JS, Keiles S, Tandy S, Pesaran T, Chen E, Gau CL, Palmaer E, Shoaepour K, Shah D, Speare V, Gandomi S, Chao E. (2014). Utilization of multigene panels in hereditary cancer predisposition testing: analysis of more than 2,000 patients. Genetics in Medicine. 16(11), 830-837. doi: 10.1038/gim.2014.40.
53. Lakhani SR, Reis-Filho JS, Fulford L, Penault-Llorca F, van der Vijver M, Parry S, Bishop T, Benitez J, Rivas C, Bignon YJ, Chang-Claude J, Hamann U, Cornelisse CJ, Devilee P, Beckmann MW, Nestle-Krämling C, Daly PA, Haites N, Varley J, Lalloo F, Evans G, Maugard C, Meijers-Heijboer H, Klijn JG, Olah E, Gusterson BA, Pilotti S, Radice P, Scherneck S, Sobol H, Jacquemier J, Wagner T, Peto J, Stratton MR, McGuffog L, Easton DF; Breast Cancer Linkage Consortium. (2005). Prediction of BRCA1 status in patients with breast cancer using estrogen receptor and basal phenotype. Clinical Cancer Research. 11(14), 5175-5180.
54. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J, Jang W, Katz K, Ovetsky M, Riley G, Sethi A, Tully R,

Villamarin-Salomon R, Rubinstein W, Maglott DR. (2016). ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Research. 44(D1), D862-868. doi: 10.1093/nar/gkv1222.

55. Lang GT, Shi JX, Hu X, Zhang CH, Shan L, Song CG, Zhuang ZG, Cao AY, Ling H, Yu KD, Li S, Sun MH, Zhou XY, Huang W, Shao ZM. (2017) The spectrum of BRCA mutations and characteristics of BRCA-associated breast cancers in China: Screening of 2,991 patients and 1,043 controls by next-generation sequencing. Int J Cancer. 141(1):129-142. doi: 10.1002/ijc.30692.

56. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG; Exome Aggregation Consortium. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature. 536(7616), 285-291. doi: 10.1038/nature19057.

57. Li H, Durbin R. (2010). Fast and accurate long-read alignment with Burrows-

Wheeler transform. Bioinformatics. 26(5), 589-595. doi: 10.1093/bioinformatics/btp698.

58. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25(16), 2078-2079. doi: 10.1093/bioinformatics/btp352.
59. Li Q, Wang K. (2017) InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. Am J Hum Genet. 100(2):267-280.
60. Liang H, He S, Yang J, Jia X, Wang P, Chen X, Zhang Z, Zou X, McNutt MA, Shen WH, Yin Y. (2014). PTENα, a PTEN isoform translated through alternative initiation, regulates mitochondrial function and energy metabolism. Cell Metabolism. 19(5), 836-848. doi: 10.1016/j.cmet.2014.03.023.
61. Liaw D, Marsh DJ, Li J, Dahia PL, Wang SI, Zheng Z, Bose S, Call KM, Tsou HC, Peacocke M, Eng C, Parsons R. (1997). Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome. Nature Genetics. 16(1), 64-67.
62. Lincoln SE, Kobayashi Y, Anderson MJ, Yang S, Desmond AJ, Mills MA, Nilsen GB, Jacobs KB, Monzon FA, Kurian AW, Ford JM, Ellisen LW. (2015). A systematic comparison of traditional and multigene panel testing for hereditary breast and ovarian cancer genes in more than 1000 patients. The Journal of Molecular Diagnostics. 17(5), 533-544. doi: 10.1016/j.jmoldx.2015.04.009.
63. Lindor NM, Guidugli L, Wang X, Vallée MP, Monteiro AN, Tavtigian S, Goldgar DE, Couch FJ. (2012) A review of a multifactorial probability - based model for classification of BRCA1 and BRCA2 variants of uncertain significance (VUS).

Hum Mutat. 33(1):8-21. doi: 10.1002/humu.21627.

64. Lord CJ, Ashworth A. (2016). BRCAness revisited. Nature Reviews Cancer. 16(2), 110-120. doi: 10.1038/nrc.2015.21.

65. Lukong KE. (2017) Understanding breast cancer - The long and winding road. BBA Clin. 7:64-77.

66. Malkin D, Li FP, Strong LC, Fraumeni JF. Jr, Nelson CE, Kim DH., Kim JK, Magdalena AG, Farideh ZB, Michael AT, Friend SH. (1990). Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. Science. 250(4985), 1233-1238.

67. Mathe E, Olivier M, Kato S, Ishioka C, Hainaut P, Tavtigian SV. (2006) Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods. Nucleic Acids Res. 34(5):1317-25.

68. Matreyek KA, Starita LM, Stephany JJ, Martin B, Chiasson MA, Gray VE, Kircher M, Khechaduri A, Dines JN, Hause RJ, Bhatia S, Evans WE, Relling MV, Yang W, Shendure J, Fowler DM. (2018). Multiplex assessment of protein variant abundance by massively parallel sequencing. Nature Genetics. 50(6), 874-882. doi: 10.1038/s41588-018-0122-z.

69. Maxwell KN, Wubbenhorst B, D'Andrea K, Garman B, Long JM, Powers J, Rathbun K, Stopfer JE, Zhu J, Bradbury AR, Simon MS, DeMichele A, Domchek SM, Nathanson KL. (2015). Prevalence of mutations in a panel of breast cancer susceptibility genes in BRCA1/2-negative patients with early-onset breast cancer. Genetics in Medicine. 17(8), 630-638. doi: 10.1038/gim.2014.176.

70. Maxwell KN, Hart SN, Vijai J, Schrader KA, Slavin TP, Thomas T, Wubbenhorst

B, Ravichandran V, Moore RM, Hu C, Guidugli L, Wenz B, Domchek SM, Robson ME, Szabo C, Neuhausen SL, Weitzel JN, Offit K, Couch FJ, Nathanson KL. (2016) Evaluation of ACMG-Guideline-Based Variant Classification of Cancer Susceptibility and Non-Cancer Associated Genes in Families Affected by Breast Cancer. Am J Hum Genet. 98(5):801-817. doi: 10.1016/j.ajhg.2016.02.024.

71. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research. 20(9), 1297-1303. doi: 10.1101/gr.107524.110.
72. Meijers-Heijboer H, van den Ouweland A, Klijn J, Wasielewski M, de Snoo A, Oldenburg R, Hollestelle A, Houben M, Crepin E, van Veghel-Plandsoen M, Elstrodt F, van Duijn C, Bartels C, Meijers C, Schutte M, McGuffog L, Thompson D, Easton D, Sodha N, Seal S, Barfoot R, Mangion J, Chang-Claude J, Eccles D, Eeles R, Evans DG, Houlston R, Murday V, Narod S, Peretz T, Peto J, Phelan C, Zhang HX, Szabo C, Devilee P, Goldgar D, Futreal PA, Nathanson KL, Weber B, Rahman N, Stratton MR; CHEK2-Breast Cancer Consortium. (2002). Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. Nature Genetics. 31(1), 55-59.
73. Meisel JL, Venur VA, Gnant M, Carey L. (2018) Evolution of Targeted Therapy in Breast Cancer: Where Precision Medicine Began. Am Soc Clin Oncol Educ Book. 38:78-86.
74. Milholland B, Dong X, Zhang L, Hao X, Suh Y, Vijg J. (2017) Differences between germline and somatic mutation rates in humans and mice. Nat Commun. 8:15183. doi: 10.1038/ncomms15183.

75. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, Ding W, et al. (1994) A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science. 266(5182):66-71.

76. Minion LE, Dolinsky JS, Chase DM, Dunlop CL, Chao EC, Monk BJ. (2015). Hereditary predisposition to ovarian cancer, looking beyond BRCA1/BRCA2. Gynecologic Oncology. 137(1), 86-92. doi: 10.1016/j.ygyno.2015.01.537.

77. Momozawa Y, Iwasaki Y, Parsons MT, Kamatani Y, Takahashi A, Tamura C, Katagiri T, Yoshida T, Nakamura S, Sugano K, Miki Y, Hirata M, Matsuda K, Spurdle AB, Kubo M. (2018). Germline pathogenic variants of 11 breast cancer genes in 7,051 Japanese patients and 11,241 controls. Nat Commun. 9(1), 4083. doi: 10.1038/s41467-018-06581-8.

78. Morgenthaler S, Thilly WG. (2007) A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). Mutat Res. 615(1-2):28-56.

79. NCCN Guidelines. NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®) Genetic/Familial High-Risk Assessment: Breast and Ovarian. (2017, October 3). Retrieved from https://www.genomeweb.com/sites/default/files/nccn_2017.pdf

80. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, Jones D, Hinton J, Marshall J, Stebbings LA, Menzies A, Martin S, Leung K, Chen L, Leroy C, Ramakrishna M, Rance R, Lau KW, Mudie LJ, Varela I, McBride DJ, Bignell GR, Cooke SL, Shlien A, Gamble J, Whitmore I, Maddison M, Tarpey PS, Davies HR, Papaemmanuil E, Stephens PJ, McLaren S, Butler AP, Teague JW, Jönsson G, Garber JE, Silver D, Miron P, Fatima A, Boyault S, Langerød A, Tutt

A, Martens JW, Aparicio SA, Borg Å, Salomon AV, Thomas G, Børresen-Dale AL, Richardson AL, Neuberger MS, Futreal PA, Campbell PJ, Stratton MR; Breast Cancer Working Group of the International Cancer Genome Consortium. (2012) Mutational processes molding the genomes of 21 breast cancers. Cell. 149(5):979-93. doi: 10.1016/j.cell.2012.04.024.

81. Nik-Zainal S1, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, Raine K, Jones D, Marshall J, Ramakrishna M, Shlien A, Cooke SL, Hinton J, Menzies A, Stebbings LA, Leroy C, Jia M, Rance R, Mudie LJ, Gamble SJ, Stephens PJ, McLaren S, Tarpey PS, Papaemmanuil E, Davies HR, Varela I, McBride DJ, Bignell GR, Leung K, Butler AP, Teague JW, Martin S, Jönsson G, Mariani O, Boyault S, Miron P, Fatima A, Langerød A, Aparicio SA, Tutt A, Sieuwerts AM, Borg Å, Thomas G, Salomon AV, Richardson AL, Børresen-Dale AL, Futreal PA, Stratton MR, Campbell PJ; Breast Cancer Working Group of the International Cancer Genome Consortium. (2012) The life history of 21 breast cancers. Cell. 149(5):994-1007. doi: 10.1016/j.cell.2012.04.023.

82. Norquist BM, Harrell MI, Brady MF, Walsh T, Lee MK, Gulsuner S, Bernards SS, Casadei S, Yi Q, Burger RA, Chan JK, Davidson SA, Mannel RS, DiSilvestro PA, Lankes HA, Ramirez NC, King MC, Swisher EM, Birrer MJ. (2016). Inherited mutations in women with ovarian carcinoma. JAMA Oncology. 2(4), 482-490. doi: 10.1001/jamaoncol.2015.5495.

83. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. (2015). A global reference for human genetic variation. Nature. 526(7571), 68-74. doi: 10.1038/nature15393.

84. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lønning PE, Børresen-Dale AL, Brown PO, Botstein D. (2000) Molecular portraits of human breast tumours. Nature. 406(6797):747-52.

85. Prakash R, Zhang Y, Feng W, Jasin M. (2015). Homologous recombination and human health: the roles of BRCA1, BRCA2, and associated proteins. Cold Spring Harbor Perspectives in Biology. 7(4), a016600. doi: 10.1101/cshperspect.a016600.

86. Price AL, Kryukov GV, de Bakker PI, Purcell SM, Staples J, Wei LJ, Sunyaev SR. (2010) Pooled association tests for rare variants in exon-resequencing studies. Am J Hum Genet. 86(6):832-8. doi: 10.1016/j.ajhg.2010.04.005.

87. Pritchard JK. (2001) Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet. 69(1):124-37.

88. Qureshi HS, Linden MD, Divine G, Raju UB. (2006). E-cadherin status in breast cancer correlates with histologic type but does not correlate with established prognostic parameters. American Journal of Clinical Pathology. 125(3), 377-385.

89. Rahman N, Seal S, Thompson D, Kelly P, Renwick A, Elliott A, Reid S, Spanova K, Barfoot R, Chagtai T, Jayatilake H, McGuffog L, Hanks S, Evans DG, Eccles D; Breast Cancer Susceptibility Collaboration (UK), Easton DF, Stratton MR. (2007). PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. Nature Genetics. 39(2), 165-167.

90. Raychaudhuri S. (2011) Mapping rare and common causal alleles for complex human diseases. Cell. 147(1):57-69. doi: 10.1016/j.cell.2011.09.011.

91. Renwick A, Thompson D, Seal S, Kelly P, Chagtai T, Ahmed M, North B, Jayatilake H, Barfoot R, Spanova K, McGuffog L, Evans DG, Eccles D; Breast

Cancer Susceptibility Collaboration (UK), Easton DF, Stratton MR, Rahman N. (2006). ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. Nature Genetics. 38(8), 873-875.

92. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL; ACMG Laboratory Quality Assurance Committee. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 17(5):405-24. doi: 10.1038/gim.2015.30.
93. Rubinstein WS. Hereditary breast cancer in Jews. (2004) Fam Cancer. 3(3-4):249-57.
94. Schwarz JM, Cooper DN, Schuelke M, Seelow D. (2014) MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods. 11(4):361-2. doi: 10.1038/nmeth.2890.
95. Serdy KM, Leone JP, Dabbs DJ, Bhargava R. (2017) Male Breast Cancer. Am J Clin Pathol. 147(1):110-119. doi: 10.1093/ajcp/aqw207.
96. Shigemizu D, Miya F, Akiyama S, Okuda S, Boroevich KA, Fujimoto A et al. (2018). IMSindel: An accurate intermediate-size indel detection tool incorporating de novo assembly and gapped global-local alignment with split read analysis. Sci Rep. 8(1):5608.
97. Shimelis H, LaDuca H, Hu C, Hart SN, Na J, Thomas A, Akinhanmi M, Moore RM, Brauch H, Cox A, Eccles DM, Ewart-Toland A, Fasching PA, Fostira F, Garber J, Godwin AK, Konstantopoulou I, Nevanlinna H, Sharma P, Yannoukakos D, Yao S, Feng BJ, Tippin Davis B, Lilyquist J, Pesaran T, Goldgar DE, Polley EC,

Dolinsky JS, Couch FJ. (2018). Triple-Negative Breast Cancer Risk Genes Identified by Multigene Hereditary Cancer Panel Testing. Journal of the National Cancer Institute. doi: 10.1093/jnci/djy106.

98. Simone D, Calabrese FM, Lang M, Gasparre G, Attimonelli M. (2011). The reference human nuclear mitochondrial sequences compilation validated and implemented on the UCSC genome browser. BMC Genomics. 517. doi: 10.1186/1471-2164-12-517.
99. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. (1987) Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science. 235(4785):177-82.
100. Stratton MR, Rahman N. (2008). The emerging landscape of breast cancer susceptibility. Nat Genet. 40(1), 17-22. doi: 10.1038/ng.2007.53
101. Starita LM, Islam MM, Banerjee T, Adamovich AI, Gullingsrud J, Fields S, Shendure J, Parvin JD. (2018). A multiplex homology-directed DNA repair assay reveals the impact of more than 1,000 BRCA1 missense substitution variants on protein function. The American Journal of Human Genetics. 103(4), 498-508. doi: 10.1016/j.ajhg.2018.07.016.
102. Stevens KN, Vachon CM, Couch FJ. (2013). Genetic susceptibility to triple-negative breast cancer. Cancer Research. 73(7), 2025-2030. doi: 10.1158/0008-5472.CAN-12-1699.
103. Suda K, Nakaoka H, Yoshihara K, Ishiguro T, Tamura R, Mori Y, Yamawaki K, Adachi S, Takahashi T, Kase H, Tanaka K, Yamamoto T, Motoyama T, Inoue I, Enomoto T. (2018). Clonal expansion and diversification of cancer-associated mutations in endometriosis and normal endometrium. Cell Reports. 24(7), 1777-

1789. doi: 10.1016/j.celrep.2018.07.037.

104. Sun J, Meng H, Yao L, Lv M, Bai J, Zhang J, Wang L, Ouyang T, Li J, Wang T, Fan Z, Fan T, Lin B, Xie Y. (2017). Germline mutations in cancer susceptibility genes in a large series of unselected breast cancer patients. Clinical Cancer Research. 23(20), 6113-6119. doi: 10.1158/1078-0432.CCR-16-3227.

105. Sunada S, Nakanishi A, Miki Y. (2018) Crosstalk of DNA double-strand break repair pathways in poly(ADP-ribose) polymerase inhibitor treatment of breast cancer susceptibility gene 1/2-mutated cancer. Cancer Sci. 109(4):893-899.

106. Torkamani A, Wineinger NE, Topol EJ. (2018) The personal and clinical utility of polygenic risk scores. Nat Rev Genet. 19(9): 581-590.

107. Tung N, Battelli C, Allen B, Kaldate R, Bhatnagar S, Bowles K, Timms K, Garber JE, Herold C, Ellisen L, Krejdovsky J, DeLeonardis K, Sedgwick K, Soltis K, Roa B, Wenstrup RJ, Hartman AR. (2015). Frequency of mutations in individuals with breast cancer referred for BRCA1 and BRCA2 testing using next-generation sequencing with a 25-gene panel. Cancer. 121(1), 25-33. doi: 10.1002/cncr.29010.

108. Tung N, Lin NU, Kidd J, Allen BA, Singh N, Wenstrup RJ, Hartman AR, Winer EP, Garber JE. (2016). Frequency of germline mutations in 25 cancer susceptibility genes in a sequential series of patients with breast cancer. Journal of Clinical Oncology. 34(13), 1460-1468. doi: 10.1200/JCO.2015.65.0747.

109. Turnbull C, Rahman N. Genetic predisposition to breast cancer: past, present, and future. (2008) Annu Rev Genomics Hum Genet. 9:321-45. doi: 10.1146/annurev.genom.9.081307.164339.

110. Wallace AJ. (2016). New challenges for BRCA testing: a view from the diagnostic laboratory. European Journal of Human Genetics. 24 Suppl 1, S10-8. doi:

10.1038/ejhg.2016.94.

111. Walsh T, Lee MK, Casadei S, Thornton AM, Stray SM, Pennil C, Nord AS, Mandell JB, Swisher EM, King MC. (2010). Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. Proceedings of the National Academy of Sciences of the United States of America. 107(28), 12629-12633. doi: 10.1073/pnas.1007983107.

112. Wang K, Li M, Hakonarson H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Research. 38(16), e164. doi: 10.1093/nar/gkq603.

113. 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 JS, Narod S. (1999). Prevalence and penetrance of BRCA1 and BRCA2 gene mutations in unselected Ashkenazi Jewish women with breast cancer. Journal of the National Cancer Institute. 91(14), 1241-1247.

114. Weischer M, Bojesen SE, Ellervik C, Tybjaerg-Hansen A, Nordestgaard BG. (2008). CHEK2*1100delC genotyping for clinical assessment of breast cancer risk: meta-analyses of 26,000 patient cases and 27,000 controls. Journal of Clinical Oncology. 26(4), 542-548. doi: 10.1200/JCO.2007.12.5922.

115. Woods NT, Baskin R, Golubeva V, Jhuraney A, De-Gregoriis G, Vaclova T, Goldgar DE, Couch FJ, Carvalho MA, Iversen ES, Monteiro AN. (2016). Functional assays provide a robust tool for the clinical annotation of genetic variants of uncertain significance. NPJ Genomic Medicine. 1. pii: 16001. doi: 10.1038/npjgenmed.2016.1.

116. Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, Collins N, Gregory S, Gumbs C, Micklem G. (1995) Identification of the breast cancer susceptibility gene BRCA2. Nature. 378(6559):789-92.

117. Wooster R, Neuhausen SL, 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):2088-90.

118. Yap TA, Sandhu SK, Carden CP, de Bono JS. (2011) Poly (ADP-ribose) polymerase (PARP) inhibitors: Exploiting a synthetic lethal strategy in the clinic. CA Cancer J Clin. 61(1):31-49.

119. Yin Y, Zeng K, Wu M, Ding Y, Zhao M, Chen Q. (2014). The levels of Ki-67 positive are positively associated with lymph node metastasis in invasive ductal breast cancer. Cell Biochemistry and Biophysics. 70(2), 1145-1151. doi: 10.1007/s12013-014-0034-1.

120. Zheng K, Tan JX, Li F, Li HY, Zeng XH, Ma BL, Ou JH, Li H, Yang SS, Jiang AM, Ni Q, Liu JL, Liu JP, Zheng H, Yue-Yang, Ling R, He JJ, Li ZG, Zeng J, Zou TN, Jiang J, Song ZJ, Liu QL, Ren GS. (2018). Clinicopathologic factors related to the histological tumor grade of breast cancer in western China: an epidemiological multicenter study of 8619 female patients. Translational Oncology. 11(4), 1023-1033. doi: 10.1016/j.tranon.2018.06.005.

121. Zhou WZ, Zhang J, Li Z, Lin X, Li J, Wang S. (2019). Targeted resequencing of 358 candidate genes for autism spectrum disorder in a Chinese cohort reveals diagnostic potential and genotype-phenotype correlations. Hum Mutat. 40(6):801-815.