Irene Catucci
Degree in Biotechnology

Identification of low-penetrance alleles, genetic modifiers and mutation analysis in familial breast cancer cases

Thesis presented to The Open University of London for the Degree of Doctor of Philosophy

Discipline: Life and Biomolecular Sciences

Date of Submission: 28 June 2013

Date of Award: 25 October 2013

Affiliated Research Centre: Fondazione IRCCS “Istituto Nazionale dei Tumori”
TABLE OF CONTENTS

ABSTRACT 4

CHAPTER 1. INTRODUCTION 6

1.1 Genetic risk for breast cancer 6

1.2 Breast cancer high-penetrance genes 7
  1.2.1 BRCA1 and BRCA2 genes 7
  1.2.2 Other high-penetrance genes 14

1.3 Breast cancer moderate-penetrance genes 15
  1.3.1 ATM, CHEK2 and BRIP1 genes 15
  1.3.2 Fanconi Anemia and breast cancer susceptibility 16
  1.3.3 The PALB2/FANCN gene 20
  1.3.4 The SLX4/FANCP gene 22

1.4 Low-penetrance alleles 24
  1.4.1 SNPs in microRNA as low-penetrance alleles 26

1.5 Genetic risk modifiers in BRCA genes mutation carriers 29
  1.5.1 The rs3834129 in the CASP8 promoter region as candidate genetic risk modifier in BRCA genes mutation carriers 35

CHAPTER 2. AIM OF THE STUDY 38

CHAPTER 3. MATERIALS AND METHODS 39

3.1 Study population 39
  3.1.1 Genetic counseling and eligibility to BRCA1 and BRCA2 genetic test 39
3.1.2 Recruiting and inclusion criteria in our study of individuals affected with breast cancer or at-risk for breast cancer 40
3.1.3 Recruiting of blood donors 41

3.2 Association studies 42

3.3 DNA samples preparation 43

3.4 Mutation screenings 43
3.4.1 PCRs conditions 43
3.4.2 Mutations detection by sequencing 44

3.5 Genotyping analyses 47

3.6 In silico analyses 48
3.6.1 Software for prediction of the missense mutations effect 48
3.6.2 Software for prediction of non-canonical splicing 49

3.7 Investigation of the PALB2 c.48G>A splicing mutation 49
3.7.1 B-lymphocytes immortalization 49
3.7.2 RNA extraction and cDNA synthesis 50
3.7.3 Amplification of cDNA and transcript analysis 51

3.8 Statistical analyses 52

CHAPTER 4. RESULTS 53

4.1 PALB2/FANCN mutation analysis 53
4.1.1 Mutation screening in a series of Italian BRCAX cases 53
4.1.2 Investigation of the c.72delG and the c.1027C>T recurrent mutations 61
4.1.3 PALB2 mutation analysis in breast and pancreatic cancer families 62

4.2 Sequencing analysis of SLX4/FANCP gene 69
4.3 Investigation of miR-27a rs895819 polymorphism as candidate low-penetrance allele 74

4.4 Analysis of the CASP8 rs3834129 as risk modifier in BRCA gene mutation carriers 76

CHAPTER 5. DISCUSSION 80

REFERENCES 92

LIST OF FIGURES 112

LIST OF TABLES 113

PUBLISHED MATERIALS FROM THIS THESIS 115
ABSTRACT

To date, germline mutations in known high-penetrance genes, mainly *BRCA1* and *BRCA2*, and in moderate- and low-penetrance genes are responsible for approximately 30-35% of breast cancer familial clustering, leaving the majority of them unexplained. In addition, the variability of the risk conferred by *BRCA1* and *BRCA2* mutations suggests the presence of genetic modifiers of this risk. Therefore, the identification and characterization of as many as possible of genetic factors is crucial for risk prediction in members of breast cancer families.

In this context, the aim of this thesis was firstly to investigate the role of the two Fanconi Anemia (FA) genes *PALB2* and *SLX4* as breast cancer predisposing loci. In the *PALB2* screening, I observed a frequency of deleterious mutation of 2.1% in familial cases recruited in cancer centers in Milan. Interestingly, I also identified the recurrent mutation c.1027C>T, detected with 10-fold increased frequency in cases from Bergamo with respect to those ascertained in Milan, suggesting a founder effect. On the contrary, the *SLX4* analysis failed to identify any clearly deleterious mutation, excluding a major role of this gene in breast cancer susceptibility in the Italian population. In addition, I genotyped the candidate low-risk rs895819 polymorphism, located in the gene coding for miR-27a, to evaluate its role in reducing breast cancer risk, previously reported in the German population. No such an association was observed in our sample set. Finally, I investigated the role of the *CASP8* rs3834129 ins/del polymorphism as a genetic modifier in Italian *BRCA1* and *BRCA2* mutation carries and I observed an association of this SNP with increased breast cancer risk only in individuals carrying *BRCA1* mutations.
In conclusion, our investigation contributed to assess the role of candidate predisposing loci and genetic modifiers of breast cancer risk, providing further knowledge on the susceptibility to this disease.
CHAPTER 1
INTRODUCTION

1.1 Genetic risk for breast cancer

Breast cancer is the second most common type of cancer and the fifth most common cause of cancer death in world population. In particular, in Italy about 45,000 novel breast cancer cases are diagnosed every year and breast cancer affects one in eight women during their lifetime. It represents about 29% of all female cancer and the first cause of cancer death in Italian women (AIRTUM Working Group, 2011).

Ten to 15% of all breast cancer cases can be considered familial since occurring in families where relatives of the index case are affected with the disease. It has been suggested that family history for breast cancer is one of the strongest risk factor. In fact, it has been observed that breast cancer risk among first-degree relatives of breast cancer patients is two to four-fold increased with respect to the general population (reviewed in Stratton and Rahman, 2008). This risk depends on the number of affected relatives, their age at diagnosis and their relationship proximity to the proband (Bradbury and Olopade, 2007; Mavaddat et al., 2010).

Recently, the understanding of breast cancer genetic predisposition was largely improved and novel breast cancer susceptibility loci have been identified. To date, breast cancer cases can be accounted by germline mutations that can be classified in three different groups according to the increase of relative risk (RR)\(^1\) that they confer. Mutations in high-penetrance genes are very rare, being detected with a cumulative frequency of no

---

\(^{1}\) Relative risk is the risk of an event (or of developing a disease) relative to exposure. In particular, it represents the ratio of the probability of the event occurring in the exposed group versus the non-exposed group.
more than 0.1-0.5%, and they increase the risk of more than 10 to 20-fold, accounting for no more the 3-5% of all breast cancer cases. Moderate-penetrance mutations are also rare, being detected with a cumulative frequency of 0.5-1%, but the increase of risk conferred is from two to three-fold. Finally, a large number of low-penetrance alleles have been identified. These are found with an individual frequency of more than 5% and each confers a RR ranging from approximately 1.05 to 1.3. Currently known high-penetrance genes are responsible for about 15-20% of breast cancer familial clustering whereas identified moderate-penetrance genes account for about 2% (reviewed in Mavaddat et al., 2010 and in Laloo and Evans, 2012); in addition, it has been estimated that low-penetrance alleles are responsible for an additional 28% of familial risk of which 14% can be explained by currently known loci (Michailidou et al., 2013) (Figure 1.1). Since only about 30-35% of familial breast cancer clustering can be explained by the above mentioned genes, the large majority of the high-risk breast cancer cases are molecularly unexplained.

1.2 Breast cancer high-penetrance genes

1.2.1 BRCA1 and BRCA2 genes

The two major genes involved in breast cancer susceptibility are the tumor suppressor genes BRCA1 and BRCA2, identified in the early 1990s.

In 1990, genetic linkage analysis, performed in families with multiple cases of early onset breast cancer, resulted in the localization of BRCA1 on chromosome 17q12-21 (Hall et al., 1990). Subsequently, an international collaborative study confirmed this data, and BRCA1 was cloned in 1994 (Miki et al., 1994). This gene is composed of 23 exons, of
which one is not coding, and extends for approximately 100 kb of the genomic DNA. The exon 11 of BRCA1 is unusually large, representing about 60% of the entire coding region.

BRCA1 encodes a protein of 1,863 amino acids with a very weak sequence conservation among species, except for highly conserved functional domains located at the terminal regions. In particular, the N-terminus region contains a RING domain, with E3 ubiquitin ligase activity, that binds the BRCA1-associated RING domain protein (BARD1) gene, and a nuclear localization sequence (NLS). The C-terminus region contains a coiled-coil domain that associates with the BRCA2 partner and localizer (PALB2) gene, and a BRCT
domain, responsible for the activation of several transcription factors (reviewed in Roy et al., 2012). BRCA1 is involved in several cellular mechanisms including DNA double-strand break repair, homologous recombination, cell cycle checkpoint control, transcriptional regulation and ubiquitination (Ahmed et al., 2009; Foulkes, 2008).

Since in early studies only 45% of all families with multiple cases of breast cancer were associated with \textit{BRCA1} mutations, the presence of an additional breast cancer susceptibility locus was proposed and confirmed by subsequent analyses (reviewed in Ahmed et al., 2009). The \textit{BRCA2} gene was mapped to chromosome 13q12-q13 in 1994 (Wooster et al., 1994), and cloned the following year (Wooster et al., 1995). This gene consists of 26 coding exons, it extends for approximately 70 kb of the genomic DNA and encodes a very large protein of 3,418 amino acids. \textit{BRCA2} contains a PALB2 binding domain at the N-terminus of the protein, eight BRC repeats from amino acid residues 1009 to 2083 that form the binding site of the DNA recombination repair protein RAD51, and a NLS domain in the C-terminus. \textit{BRCA2} is involved in DNA repair processes and maintenance of chromosome stability, being responsible for the double strand DNA breaks recognition, and participating in the homologous recombination repair (reviewed in Roy et al., 2012).

Germline deleterious mutations in \textit{BRCA1} and \textit{BRCA2} are associated with increased breast cancer risk and have high penetrance. It was estimated that these mutations increase the risk by approximately 10 to 20-fold. This means that carriers of \textit{BRCA1} and \textit{BRCA2} mutations have risk of developing breast cancer of 30-60% by the age of 60, compared with 3% observed in the general population (reviewed in Stratton and Rahman, 2008). It has been also reported that BRCA genes mutations confer a risk of developing other types of cancer. In particular, \textit{BRCA1} and \textit{BRCA2} mutation carriers have a cumulative ovarian cancer risk of 39% and 11%, respectively (Milne and Antoniou,
2011) whereas mutations in *BRCA2* are also responsible for increased risk of developing pancreatic and prostate cancer (Foulkes, 2008).

A large amount of *BRCA1* and *BRCA2* mutations have been identified. The majority of these mutations are very rare, with a cumulative frequency of approximately 0.2-0.4% in the general population and are mostly found in single families only (reviewed in Turnbull and Rahman, 2008 and in Laloo and Evans, 2012). However, this frequency varies depending on the tightness of the inclusion criteria in each study, the method of mutation analysis and also the screened population (reviewed in Roy et al., 2012).

While a large amount of different *BRCA1* and *BRCA2* mutations are detected in the majority of the screened populations, specific ethnic groups are characterized by few mutations with higher frequency, due to a 'founder' effect. The 'founder' concept was introduced to explain the loss of genetic variation that occurs when a new population was founded by a very small number of individuals, deriving from a larger group. This phenomenon can be caused by geographical isolation or by a dramatic decrease of the original population (reviewed in Fackenthal and Olopade, 2007).

One of the more interesting example of founder mutations is found in the Ashkenazi Jews that originates from Eastern and Central Europe (Germany, Poland, Lithuania, Ukraine and Russia). In this population, the *BRCA1* mutations 185delAG\(^2\) and 5382insC have been found with a frequency of 0.8-1% and 0.1-0.4%, and the *BRCA2* mutation 6174delT with a frequency of 1-1.5%. These three mutations are responsible for about 6.7-11.7% of all breast cancer cases and about 59% of high-risk cases (Table 1.1; reviewed in Fackenthal and Olopade, 2007 and in Ferla et al., 2007).

---

2 Throughout the text, *BRCA1* and *BRCA2* mutation are reported according to the Breast Cancer Information Core (BIC, http://research.nhgri.nih.gov/bic/) nomenclature.
Additional founder mutations have been detected in other European populations (Table 1.1). In Icelanders, the \textit{BRCA2} mutation 999del5 has been detected with a frequency of 8.5% and the \textit{BRCA1} mutation G5193A has been found in 1% of breast cancer cases. In Norway, four main founder mutations have been detected in \textit{BRCA1}, of which three, 1675delA, 816delGT, 3347delAG, originated from southwestern and the fourth, 1135insA, from the southeast. They represent about 68% of all \textit{BRCA1} mutations found in the country. In Finland, 11 recurrent mutations have been identified and four of them, IVS11+3A>G in \textit{BRCA1} and IVS23+1G>A, C7708T and T8555G in \textit{BRCA2} are exclusive to the Finnish population. In Sweden, the \textit{BRCA1} 3171ins5 accounts for 70% of the BRCA gene mutations. In The Netherlands, the \textit{BRCA1} 2804delAA accounts for about 24% of all mutations in \textit{BRCA1} and \textit{BRCA2} and probably originated more than 200 years ago. Two additional founder mutations, \textit{BRCA1} IVS12-1643del3835 and \textit{BRCA2} 5579insA, were found in the southwest of the country. In France, the \textit{BRCA1} founder mutations 3600del11 and G1710X represent 37% and 15% of all mutations found in high-risk breast cancer cases, respectively. In addition, several founder mutations, originating from France, were found in French-Canadians of Quebec. Among these, the most common are \textit{BRCA1} C4446T, \textit{BRCA2} 8765delAG and 3398delAAAAG. (Table 1.2; reviewed in Fackenthal and Olopade, 2007 and in Ferla et al., 2007).

In Italy, there are only few mutations that result recurrent in specific geographical areas of the country. In particular, the \textit{BRCA1} 5083del19 mutation, was found in four probands from families that originated from Calabria (Baudi et al., 2001); in Sardinia, the \textit{BRCA2} 8765delAG mutation was detected with a frequency of 1.7% (Pisano et al., 2000); \textit{BRCA1} V1688del was reported as recurrent in families from Northeast Italy (Malacrida et al., 2008); finally four distinct \textit{BRCA1} mutations, 1499insA, 3347delAG, 3404delA and 5181del3, have been shown to account for about 73% of familial breast and/or ovarian
cancers originating from Central-Eastern Tuscany (Table 1.1) (Caligo et al., 1996; Papi et al., 2009).

<table>
<thead>
<tr>
<th>Population</th>
<th>BRCA1 Mutations</th>
<th>BRCA2 Mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ashkenazi Jews</td>
<td>185delAG&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>6174delT&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>5382insC&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Icelandics</td>
<td></td>
<td>999del5&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Norwegians</td>
<td>1675delA&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>816delGT&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3347delAG&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1135insA&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Finnish</td>
<td>IVS11+3A&gt;G&lt;sup&gt;e&lt;/sup&gt;</td>
<td>IVS23+1G&gt;A&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C7708T&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T8555G&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Swedish</td>
<td>3171ins5&lt;sup&gt;f,g&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Dutch</td>
<td>2804delAA&lt;sup&gt;h&lt;/sup&gt;</td>
<td>5579insA&lt;sup&gt;h&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>IVS12-1643del3835&lt;sup&gt;h&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>French</td>
<td>3600del11&lt;sup&gt;i&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G1710X&lt;sup&gt;i&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Italians (Calabria)</td>
<td>5083del119&lt;sup&gt;j&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Italians (Sardinia)</td>
<td></td>
<td>8765delAG&lt;sup&gt;k&lt;/sup&gt;</td>
</tr>
<tr>
<td>Italians (Tuscany)</td>
<td>1499insA&lt;sup&gt;l,m&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3347delAG&lt;sup&gt;m&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3404delA&lt;sup&gt;n&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5181del13&lt;sup&gt;n&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Italians (Northeast Italy)</td>
<td>V1688del&lt;sup&gt;o&lt;/sup&gt;</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Struwein et al., 1995; <sup>b</sup>Roa et al., 1996; <sup>c</sup>Thorlacius et al., 1996; <sup>d</sup>Moller et al., 2007; <sup>e</sup>Sarantaus et al., 2000; <sup>f</sup>Bergman et al., 2001; <sup>g</sup>Bergman et al., 2005; <sup>h</sup>Zeegers et al., 2004; <sup>i</sup>Muller et al., 2004; <sup>j</sup>Baudi et al., 2001; <sup>k</sup>Pisano et al., 2000; <sup>l</sup>Caligo et al., 1996; <sup>m</sup>Papi et al., 2009; <sup>n</sup>Malacrida et al., 2008.

In non-European countries, founder mutations were identified in American-Hispanic breast cancer cases (BRCA1 2552delC and S995X), in Columbian-Hispanics (BRCA1 3450delCAAG and A1708E, BRCA2 3034delACAA), in Afro-Americans
(BRCA1 943ins10, 1832del5 and 5296del4), in South-Africans (BRCA1 E881X), in the Chinese population (BRCA1 1081delG) and in other Asian countries, such as Japan (BRCA1 Q934X and K63X, BRCA2 5802delAATT), Malaysia (BRCA1 2846insA), Philippines (BRCA1 5454delC, BRCA2 4265delCT and 4859delA) and Pakistan (BRCA1 S1503X and R1835X) (Table 1.2) (reviewed in Ferla et al., 2007).

Table 1.2. BRCA1 and BRCA2 founder mutations identified in non-European countries

<table>
<thead>
<tr>
<th>Population</th>
<th>BRCA1 Mutations</th>
<th>BRCA2 Mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>French-Canadians</td>
<td>C44461*</td>
<td>8765delAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3398delAAAAG</td>
</tr>
<tr>
<td>American-Hispanics</td>
<td>2552delC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S995X</td>
<td></td>
</tr>
<tr>
<td>Columbian-Hispanics</td>
<td>3450delCAAG</td>
<td>3034delACAA</td>
</tr>
<tr>
<td></td>
<td>A1708E</td>
<td></td>
</tr>
<tr>
<td>Afro-Americans</td>
<td>943ins10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1832del5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5296del4</td>
<td></td>
</tr>
<tr>
<td>South-Africans</td>
<td>E881X</td>
<td></td>
</tr>
<tr>
<td>Chineses</td>
<td>1081delG</td>
<td></td>
</tr>
<tr>
<td>Japanese</td>
<td>Q934X</td>
<td>5802delAATT</td>
</tr>
<tr>
<td></td>
<td>K63X</td>
<td></td>
</tr>
<tr>
<td>Malaysians</td>
<td>2846insA</td>
<td></td>
</tr>
<tr>
<td>Filipinos</td>
<td>5454delC</td>
<td>4265delCT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4859delA</td>
</tr>
<tr>
<td>Pakistanis</td>
<td>S1503X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R1835X</td>
<td></td>
</tr>
</tbody>
</table>

|                  |                  |                  |
| Tonin et al., 1999; | Tonin et al., 2001; | Oros et al., 2006; |
|                  |               | Weitzel et al., 2005; |
|                  |               | Torres et al., 2007; |
|                  |               | Olopade et al., 2003; |
|                  |               | Reeves et al., 2004; |
|                  |               | Khoo et al., 2002; |
|                  |               | Ikeda et al., 2001; |
|                  |               | Sekine et al., 2001; |
|                  |               | Lee et al., 2003; |
|                  |               | De Leon Matsuda et al., 2002; |
|                  |               | Rashid et al., 2006. |

To date, a very large number of BRCA1 and BRCA2 variants have been identified, including disease-causing, neutral and unclassified variants. Disease-causing mutations
include nonsense mutations, frameshifts caused by small insertions and/or deletions, and
splice-site mutations leading to the formation of a truncated protein. In addition to point
mutations, large genomic rearrangements, including large deletions (whole exon) and
insertions/duplications, have been also detected. In high-risk cases, these have been found
in approximately 2%-12% and 2%-5% of \textit{BRCA1} and \textit{BRCA2} mutation carriers,
respectively (reviewed in Fackenthal and Olopade, 2007). Also, a small number of unique
missense variants both in \textit{BRCA1} and \textit{BRCA2} have been classified as deleterious. These
are especially located in the functional domains of the two proteins. In addition to
deleterious mutations, common variants have been detected and classified as neutral or
with no clinical significance. However, there is a large number of missense mutations,
intronic variants and in-frame deletions or insertions with an uncertain clinical relevance.
These variants of unknown significance (VUS) represent a major problem in the
counseling and clinical management of at risk individuals, since it is unclear how these
variants can alter the protein function. Multi-factorial probability based models and
functional assays can be used to clarify the role of variants not yet classified, improving
breast cancer risk assessment in families that carry these mutations (Lindor et al., 2012;
Couch et al., 2008).

\subsection*{1.2.2 Other high-penetrance genes}

Another small fraction of familial breast cancer cases can be accounted by germline
mutations in other high-penetrance genes, responsible for hereditary syndromes that
include breast cancer as phenotype. These genes include \textit{TP53}, causing the Li-Fraumeni
syndrome; \textit{LKB1/STK11}, causing the Peutz-Jeghers syndrome and \textit{PTEN}, causing the
Cowden syndrome. All these genes are involved in multiple pathways that regulate cell
cycle, transcription and cell polarity. Mutations in these genes are very rare and account for
less than 1% of all familial breast cancer cases, increasing breast cancer risk by more than ten-fold (Turnbull and Rahman, 2008).

1.3 Breast cancer moderate-penetrance genes

Moderate-penetrance genes were considered of more interest in the last decade, when it became clear that mutations in high-risk genes could explain only about 20% of all familial breast cancer cases and linkage analyses failed to identify novel high-risk genes, although their existence cannot be completely excluded (Ahmed et al., 2009). Moderate penetrance genes have been firstly identified through the approach of candidate genes, based on the investigation of genes involved in BRCA1 and BRCA2 pathways. It has been reported that mutations in ATM, CHEK2, BRIP1 and PALB2 are associated with increased breast cancer risk. These mutations are less rare than BRCA1 and BRCA2 mutations, being detected with a cumulative frequency of approximately 1%, and increase breast cancer risk of no more than two to three folds. Carriers of moderate penetrance variants have 6-10% risk of developing breast cancer by age 60, compared with 3% in the general population (reviewed in Stratton and Rahman, 2008).

1.3.1 ATM, CHEK2 and BRIP1 genes

The gene ATM encodes a protein kinase with a crucial role in response to DNA double-strand breaks. In particular, ATM promotes the activation of a signaling cascade that causes the phosphorylation of multiple proteins, including BRCA1 and p53. Homozygous mutations in this gene are responsible for the Ataxia Telangiectasia, an autosomal recessive disease characterized by predisposition to cancer in childhood,
particularly lymphoid cancer, and an increased breast cancer risk. The role of *ATM* in breast cancer susceptibility was suggested in 2006, when mutations in *ATM* were found in 12/443 familial breast cancer cases, negative for mutations in *BRCA1* and *BRCA2*, and in 2/521 controls, with a relative risk of 2.37 (Renwick et al., 2006).

*CHEK2* was the first moderate risk breast cancer gene identified, in 2002. This gene encodes a cell cycle checkpoint protein kinase responsible for the phosphorylation of p53 and *BRCA1*, and it is involved in the response to DNA damage. Among all reported variants, the c.1100delC is the most common *CHEK2* mutation. It was found with a frequency of 0.2-1% in the European population and in about 4.2% of breast cancer families, with a relative risk of 2.34. However, this mutation frequency is highly variable among different populations (reviewed in Laloo and Evans, 2012) and extremely low in Italy (0.11%; Caligo et al., 2004).

In 2006, it was showed that *BRIP1*, which encodes a *BRCA1*-interacting helicase involved in DNA repair, was a moderate breast cancer risk gene. In a case-control study, *BRIP1* germline mutations were found in 9/1212 familial breast cancer cases and 1/2081 controls (Seal et al., 2006).

### 1.3.2 Fanconi Anemia and breast cancer susceptibility

In the last decade, a strong connection between breast cancer susceptibility and the Fanconi Anemia (FA) disease has been suggested. In 2002, it has been showed that the FA gene *FANCD1*, responsible for FA subtype D1, and *BRCA2* are the same genes (Howlett et al., 2002). Subsequently, it has been also reported that mutations in other FA genes, including *FANCN*/PALB2, *FANCJ*/*BRIP1* and *FANCO*/RAD51C, have a frequency of approximately 0.5-1% in familial breast and ovarian cancer cases (Tischkowitz and Xia, 2010; Hollestelle et al., 2010; Meindl et al., 2010). These data contributed to consider FA genes as interesting candidates in breast cancer susceptibility.
The Fanconi Anemia is an autosomal recessive hereditary disorder characterized by congenital defects, progressive bone marrow failure and cancer predisposition. FA is caused by biallelic mutations in one of the 15 FA or FA-like genes, that are *FANCA, FANCB, FANCC, FANCD1, FANCD2, FANCE, FANCF, FANCG, FANCI, FANCI, FANCL, FANCM, FANCN, FANCO* and *FANCP*. In particular, about 85% of FA cases are accounted by mutations in *FANCA, FANCC* and *FANCG*; about 10% are due to mutations in *FANCD1, FANCD2, FANCE, FANCF* and *FANCL*, and the remaining 5% by mutations in *FANCB*, responsible for the X-linked FA disease, *FANCI, FANCI, FANCM, FANCN, FANCO* and *FANCP*.

Recent studies suggested that all FA proteins are involved in a common pathway, the Fanconi Anemia pathway, required for DNA damage response and DNA repair mechanisms. Among all of the FA proteins, *FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL* and *FANCM* are assembled in a nuclear complex named FA core complex. The role of this complex is the mono-ubiquitination of two other FA proteins, *FANCI* and *FANCD2*, forming the ID complex. The activation of the ID complex leads to the recruitment of the downstream proteins of the FA pathway, that are *FANCD1, FANCI, FANCN, FANCO* and *FANCP*, and the promotion of the DNA repair.

Specifically, when a DNA damage is recognized, the activation of the FA pathway promotes the assembly of the FA core complex proteins and other associated proteins, such as *FAAP24, FAAP100* and the recently identified FANCM-associated proteins, *MHF1* and *MHF2*. Generally, *FANCA, FANCB, FANCC, FANCE, FANCF, FANCG* and *FANCL* are constitutively associated, forming a stable sub-complex, whereas the association of *FANCM* to the complex occurs only after the activation of the pathway. This binding is mediated by the associated proteins *FAAP24, FAAP100, MHF1* and *MHF2*. The formation of this sub-complex is required for recognizing of stalled replication forks, the recruitment
of the core complex and its binding with chromatin. It was also suggested that the mono-ubiquitination activity of the core complex is mediated by FANCL. This protein contains a ring finger domain for the E3 ubiquitin ligase activity required for the mono-ubiquitination of the ID complex, mediated by the UBE2T enzyme. This event results in the translocation of the complex to the chromatin, where it forms DNA repair foci. The correct modification of FANCD2 and FANCI in the ID complex is also mediated by ataxia telangiectasia and Rad3-related kinase ATR, responsible for the phosphorylation of FANCD2 and FANCI and the formation of the ID complex heterodimer. This process is reversible: the deubiquitination is mediated by the USP1 ubiquitin hydrolase that presumably deactivates the ID complex at the end of the repair process.

At the level of the DNA repair foci, the ID complex interacts with the downstream FA proteins FANCD1, FANCN, FANCJ, FANCO and FANCP, and other proteins involved in DNA repair, promoting homologous recombination repair. In particular, FANCD1 or FANCO are cofactor of the recombinase protein RAD51; FANCN interacts with FANCD1, promoting its nuclear localization and its stabilization and FANCJ is an helicase interacting with BRCA1, also involved in DNA damage response and homologous recombination (Figure 1.2; Kitao and Takata, 2011; Su and Huang, 2011; Cybulski and Howlett, 2011).

Very recently, in addition to mutations reported in FANCN/PALB2, FANCJ/BRIP1 and FANCO/RAD51C, mutations in other FA genes have been reported in breast cancer cases. As a result of exome sequencing analyses, the protein truncating mutation c.651_652del was identified in the X-ray repair cross completing gene-2 (XRCC2), a paralog gene of RAD51C, in an Australian breast cancer patient (Park et al., 2012). Subsequent analyses revealed the presence of other carriers of XRCC2 pathogenic mutations, both in cases and in controls. However, the role of XRCC2 mutations in breast
cancer susceptibility remains controversial. A recent screening of a large series of familial breast cancer cases and controls did not confirm an involvement of XRCC2 mutations in increasing the risk in familial breast cancer cases from The Netherlands, USA, Spain and Italy (Hilbers et al., 2012). In 2012, mutations in FANCA and FANCC were also reported in familial breast cancer cases. In particular, the p.Thr561Met, the p.Cys625Ser and the p.Ser1088Phe missense variants were found in FANCA and predicted to be deleterious by protein prediction programs (Litim et al., 2013), and three truncating mutations, c.535C>T (p.Arg179X), c.553C>T (p.Arg185X) and c.67delG, were detected in FANCC, suggesting a role in breast cancer susceptibility (Thompson et al., 2012).

Figure 1.2. A schematic model of the Fanconi Anemia (FA) pathway. The activation of the FA pathway promotes the assembly of FANCM with the associated proteins MHF1, MHF2, FAAP24 and FAAP100. Subsequently, the binding of this sub-complex with FANCA, FANCB, FANCC, FANCE, FANCF, FANCG and FANCL results in the formation of the FA core complex and its binding to the chromatin. After the phosphorylation mediated by ATR, the ID complex, composed by FANCD2 and FANCI, is mono-ubiquitinated by the E3 ligase activity, mediated by the core complex protein FANCL. This mechanism leads to the association of the ID complex with the chromatin and the recruiting of FA downstream proteins (FANCD1, FANCJ, FANCN, FANCO and FANCP) promoted by the associated protein FAN1. All of these events promote the DNA repair.

Modified from Kitao and Takata, Int J Hematol. 2011
1.3.3 The *PALB2/FANCN* gene

The gene *PALB2* (Partner and Localizer of BRCA2) was identified by searching novel components of endogenous BRCA2-containing complexes. This gene is located on chromosome 16p12.2 and consists of 13 exons. It encodes a protein of 1,186 amino acids, with no clearly functional domains, a coiled-coil motif at the N-terminus and a series of WD40 repeats at C-terminus of the protein. The major function of PALB2 is the stabilization of BRCA2 and the promotion of its nuclear localization. The PALB2 binding site on BRCA2 is localized on the extreme N-terminus of the protein and interacts with the C-terminus of PALB2, in the WD40 repeats region. This interaction is essential for BRCA2 role in homologous recombination. In particular, it has been proposed that PALB2 is able to recruit BRCA2 in the site of the DNA damage, promoting the assembly with RAD51 and the activation of the homologous recombination mechanisms (Oliver et al., 2009; Tischkowitz and Xia, 2010). Recent studies indicated that PALB2 can also interact with BRCA1. Specifically, a coiled-coil motif in the N-terminus of PALB2 directly binds the BRCT domain of BRCA1, physically linking BRCA1 and BRCA2. It has been also proposed that BRCA1 is responsible for the recruiting of PALB2 in the DNA damage site, causing the subsequent recruiting of BRCA2 (Zhang et al., 2009).

In 2007, it was reported that biallelic mutations in *PALB2/FANCN* are responsible for the FA disease (Reid et al., 2007), whereas monoallelic mutations are involved in breast cancer susceptibility. *PALB2* germline truncating mutations were identified in 10/923 English familial breast cancer cases, with a frequency of 1.1% and it has been estimated that these mutations increase breast cancer risk by approximately two-fold (Rahman et al., 2007). Subsequently, *PALB2* truncating mutations were also reported in 1/95 Spanish (Garcia et al., 2009), 8/976 German (Hellebrand et al., 2010; Bogdanova et al., 2011), 19/1512 Australian (Southey et al., 2010; Wong et al., 2011; Teo et al., 2013-a), 2/227 Italian (Papi et al., 2009; Balia et al., 2010) and 5/1,124 Danish familial breast cancer cases.
(Tischkowitz et al., 2012). *PALB2* mutations were also found in 34/995 cases with unselected ancestry (Tischkowitz et al., 2007; Casadei et al., 2011) and in 3/360 Chinese breast cancer cases with early onset of disease (Cao et al., 2008). Among the above mentioned studies, mutation frequency varied from 0.2% to 3.3%. In addition, founder mutations have been identified in different populations. In 2007, the c.1592delT mutation was identified in 3/113 Finnish familial breast cancer cases and 6/2,501 controls, with a frequency of 2.7% and 0.2%, respectively (Erkko et al., 2007). Subsequently, the c.2323C>T mutation (p.Gln775X) was reported in 4/564 (0.7%) French-Canadian breast cancer cases with early onset of disease (Ghadirian et al., 2009), and the c.509_510delGA mutation was detected in 4/648 (0.6%) Polish familial breast cancer cases and in 1/1,310 (0.08%) unrelated controls (Dansonka-Mieszkowska et al., 2010). Very recently, the c.3113G>A (p.Trp1038X) mutation was found in 5/1,403 (0.4%) Australian cases and in 0/764 controls from a population-based study (Southey et al., 2010), and in 8/871 (0.9%) high-risk familial cases (Teo et al., 2013-b).

In 2009, the *PALB2* pathogenic mutation c.172_175delTTGT was identified in a pancreatic cancer patient using exomic sequencing analysis (Jones et al., 2009). Following this observation the genomic region of the gene was sequenced in 96 familial pancreatic cancer cases. In this analysis, three different truncating mutations were found, for a frequency of 3.1%. Interestingly, among the four *PALB2* positive families, three presented cases of breast cancer in their pedigrees. Two subsequent studies reported *PALB2* truncating mutations in four other pancreatic cancer families, with a frequency of less than 1% and 3.7%, respectively, and in each of these families, cases of breast cancer were reported (Tischkowitz et al., 2009; Slater et al., 2010). Thus, of the eight pancreatic cancer families positive for *PALB2* mutations, reported in the above mentioned studies, seven had one or more cases of breast cancer in their pedigrees. More recently, *PALB2* was also screened in breast cancer cases with personal or family history of pancreatic cancer.
(Hofstatter et al., 2011; Stadler et al., 2011). In the first analysis, it was reported that 2/94 cases carried a \( \text{PALB2} \) pathogenic mutation, for a frequency of 2.1%, whereas no \( \text{PALB2} \) mutation was identified in the second study. All of these findings indicate a possibly co-occurrence of breast and pancreatic cancer in \( \text{PALB2} \) mutation positive families.

1.3.4 The \( \text{SLX4/FANCP} \) gene

\( \text{SLX4} \) encodes a 1,834 amino acids multidomain scaffold protein involved in DNA repair. In particular, it has been proposed that \( \text{SLX4} \) interacts with three different endonucleases, \( \text{SLX1}, \text{ERCC4/XPF-ERCC1} \) and \( \text{MUS81-EME1} \), promoting their enzymatic activity in the processing of DNA repair intermediates and in the repair mechanism (Fekairi et al., 2009; Munoz et al., 2009; Svendsen et al., 2009). In addition, it has been shown that \( \text{SLX4} \) directly interacts with the telomere-binding protein TRF2 and its partner TERF2IP/RAP1, and with the mismatch repair heterodimer MSH2-MSH3 (Cybulski and Howlett, 2011).

In early 2011, two different studies proposed an involvement of \( \text{SLX4} \) in the development of the Fanconi Anemia. Firstly, Stoepker and colleagues analyzed an individual with a Fanconi Anemia diagnosis but with no mutations in known Fanconi Anemia genes, proposing a possible involvement of the \( \text{SLX4} \) gene (Stoepker et al., 2011). Sequence analysis revealed the presence of a homozygous truncating mutation, c.268delA, in the first exon of the gene, resulting in the formation of a premature stop at codon 126. Additional evidences of the role of \( \text{SLX4} \) in the disease were obtained by sequencing a second Fanconi Anemia family, in which three siblings carried two \( \text{SLX4} \) mutations, the c.1093delC, inherited from their father, and the splicing site mutation c.1163+3dupT, inherited from their mother. Subsequently, Kim and colleagues identified two novel Fanconi Anemia individuals with unassigned Fanconi Anemia complementation group and carrying \( \text{SLX4} \) pathogenic mutations (Kim et al., 2011). In particular, they detected one
individual carrying two heterozygous deleterious mutations, c.514delC and c.2013+225_3147delinsCC, and a second individual carrying the homozygous mutation c.1163+2T>A. This evidence supports a role of the gene \textit{SLX4} in Fanconi Anemia development, and indicates the \textit{SLX4/FANCP} gene as responsible for the new FA subtype-P. Since germline mutations in some of the other FA genes, such as \textit{BRCA2}, \textit{PALB2} and \textit{BRIP1}, are associated with an increased breast cancer risk, \textit{SLX4} may be considered a good candidate as a breast cancer predisposing gene.

To date, four different studies focused on the involvement of \textit{SLX4} in breast cancer susceptibility have been performed and only two truncating mutations have been found. In 2011, 52 German and Byelorussian and 94 Spanish familial breast cancer cases were screened for the entire \textit{SLX4} coding region (Landwehr et al., 2011; Fernandez-Rodriguez et al., 2012). Although a large number of variants were found in both of these analyses, none of them resulted as a clearly pathogenic mutation. Subsequently, the screening of a large cohort of 729 familial breast cancer cases from The Netherland, Canada and Belgium identified a splicing site mutation, c.2013+2T>A, that causes the disruption of the splice donor site, loss of the reading frame and the introduction of a premature codon stop (Bakker et al., 2013). More recently, a Spanish study screened 486 familial breast cancer cases and identified the nonsense mutation p.Glu1517X and an additional missense mutation, the p.Arg372Trp, that was predicted to be pathogenic by \textit{in silico} analysis (de Garibay et al., 2012). In conclusion, two carriers of \textit{SLX4} clearly deleterious mutations were found in a total of 1,361 familial breast cancer cases, for a very low frequency of 0.15%.
1.4 Low-penetrance alleles

In the early 2000, a hypothesis began to emerge that a percentage of familial breast cancer cases was due to a large number of low-penetrance alleles, each conferring a very small increased or decreased breast cancer risk, and possibly acting simultaneously, under a polygenic model. In fact, germline mutations in high- and moderate-penetrance genes explain only approximately 20-25% of all familial breast cancer cases (reviewed in Varghese and Easton, 2010).

Association studies are the most important instrument for detection of low-penetrance alleles. In these studies, frequencies of genetic variants are measured and compared in cases versus controls, to evaluate their association with a specific phenotype. Early association studies were based on the analysis of a limited number of variants, mainly single nucleotide polymorphisms (SNPs), located in genes of interest for breast cancer. This method is known as the candidate gene approach and requires an a priori knowledge of the biological functions of the candidate gene (Ahmed et al., 2009; Mavaddat et al., 2010; Varghese and Easton, 2010). With this approach, the CASP8 variant D30211 (rs1045485) was investigated and it was reported as associated with breast cancer risk, although it is possible that other variants in linkage disequilibrium are those actually causative (Cox et al., 2007). In addition to the CASP8 rs1045485, other variants were reported as associated with increased risk but none of these associations was confirmed (Breast Cancer Association Consortium, 2006), probably because of the limited sample size of each single study that did not allow enough statistical power. To overcome single study limitations, consortia or multi-group collaborations have been established. These consortia collect a very large number of cases and controls and provide a higher statistical power to detect small increased or decreased breast cancer risk.
The identification of common genetic variations tagging genomic regions and the developing of high-throughput platforms has led to the advent of a novel strategy for detection of low-risk alleles, the genome wide association studies (GWASs), that allow the simultaneous genotyping of hundreds of thousands of SNPs throughout the genome. This strategy is based on the selection of subsets of SNPs to detect most common variations in the genome in a given population, taking advantage of the correlation among flanking genetic variants in linkage disequilibrium (reviewed in Mavaddat et al., 2010). This is an agnostic approach not based on prior knowledge of functions of a gene or its involvement in a particular pathway. The initial GWASs identified SNPs in 12 breast cancer susceptibility loci (Table 1.3). Some of these variants were located in known genes, such as FGFR2, TOX3, MAP3K1, that play a role in breast cancer development or in cells proliferation and apoptosis or in cancer progression and metastasis. Other variants were located in regions, such as the 8q24 and the 2q35, that do not contain any gene (reviewed in Varghese and Easton, 2010).

Very recently, a large-scale genotyping were performed on 10,052 breast cancer cases and 12,575 controls of European origins in the Breast Cancer Association Consortium (BCAC), as part of a collaborative project involving four different consortia (COGS). In this analysis, 41 novel SNPs associated with breast cancer risk were identified, both in genes and in desert gene regions, largely increasing the number of susceptibility loci known to date (Michailidou et al., 2013). It was estimated that these newly associated loci account for approximately 5% of the familial breast cancer risk, and individually conferring a small risk increase, with the higher odds ratio (OR) of 1.26.

The profiling of the current set of known susceptibility loci, under the assumption that their effects combine multiplicatively, may allow the identification of the 5% of female individuals with a 2.3-fold increased breast cancer risk and the 1% with a 3-fold increased risk with respect to the average population.
Chapter I - Introduction

Table 1.3. Low-penetrance breast cancer susceptibility loci identified in the initial GWASs

<table>
<thead>
<tr>
<th>Locus</th>
<th>Gene</th>
<th>SNP</th>
<th>Per allele OR</th>
<th>P-value</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1p11.2</td>
<td>NOTCH2/FCGR1B</td>
<td>rs11249433</td>
<td>1.16</td>
<td>7 x 10^{-10}</td>
<td>Thomas et al., 2009</td>
</tr>
<tr>
<td>2q35</td>
<td>none</td>
<td>rs13387042</td>
<td>1.20</td>
<td>1 x 10^{-13}</td>
<td>Stacey et al., 2007</td>
</tr>
<tr>
<td>3p24</td>
<td>NEK10/SLC4A7</td>
<td>rs4973768</td>
<td>1.11</td>
<td>4 x 10^{-25}</td>
<td>Ahmed et al., 2009</td>
</tr>
<tr>
<td>5q11</td>
<td>MAP3K1</td>
<td>rs889312</td>
<td>1.13</td>
<td>7 x 10^{-20}</td>
<td>Easton et al., 2007</td>
</tr>
<tr>
<td>5p12</td>
<td>MRPS30</td>
<td>rs10941579</td>
<td>1.19</td>
<td>3 x 10^{-11}</td>
<td>Stacey et al., 2008</td>
</tr>
<tr>
<td>6q25.1</td>
<td>ESR1</td>
<td>rs2046210</td>
<td>1.29</td>
<td>2 x 10^{-15}</td>
<td>Zheng et al., 2009</td>
</tr>
<tr>
<td>8q24</td>
<td>none</td>
<td>rs13281615</td>
<td>1.08</td>
<td>5 x 10^{-12}</td>
<td>Easton et al., 2007</td>
</tr>
<tr>
<td>10q26</td>
<td>FGFR2</td>
<td>rs2981582</td>
<td>1.26</td>
<td>2 x 10^{-76}</td>
<td>Easton et al., 2007;</td>
</tr>
<tr>
<td>11p15</td>
<td>LSP1</td>
<td>rs3817198</td>
<td>1.07</td>
<td>3 x 10^{-9}</td>
<td>Easton et al., 2007</td>
</tr>
<tr>
<td>14q24.1</td>
<td>RAD51L1</td>
<td>rs999737</td>
<td>1.06</td>
<td>2 x 10^{-7}</td>
<td>Thomas et al., 2009</td>
</tr>
<tr>
<td>16q12</td>
<td>TOX3</td>
<td>rs3803662</td>
<td>1.20</td>
<td>1 x 10^{-36}</td>
<td>Easton et al., 2007;</td>
</tr>
<tr>
<td>17q23.2</td>
<td>COX11</td>
<td>rs6504950</td>
<td>1.05</td>
<td>1 x 10^{-8}</td>
<td>Ahmed et al., 2009</td>
</tr>
</tbody>
</table>

OR, odds ratio

1.4.1 SNPs in microRNA as low-penetrance alleles

MicroRNAs (miRNAs) are a class of small non-coding RNAs of about 22 nucleotides, involved in the regulation of gene expression through a specific binding with the mRNA. Usually, miRNAs are transcribed by RNA polymerase II as long primary transcripts (pri-miR), processed and cleaved in the nucleus (pre-miR), and exported to the cytoplasm. Here, an additional enzymatic cleavage leads to the formation of the mature miRNA (Figure 1.3; reviewed in Ryan et al., 2010).

The activity of mature miRNA is due to the recognition of a 6-7 nucleotide target sequence located at the 3'-untranslated region (UTR) of the mRNA. The mRNA-miRNA binding results in the inhibition of translation and/or degradation of target mRNA (reviewed in Le Quesne and Caldas, 2010).
It has been suggested that deregulation of miRNAs is involved in etiology, progression and prognosis of different types of cancer. In addition, it has been observed that SNPs located in miRNA genes can affect miRNA function by modulating the transcription of the primary transcript, pri-miRNA and pre-miRNA processing and maturation, or miRNA-mRNA interaction. SNPs in miRNA have been investigated in case-control studies and an association of these SNPs with increased or decreased risk has been reported in different types of cancer (reviewed in Ryan et al., 2010).

Figure 1.3. MicroRNA maturation processes. RNA polymerase II (Pol II) produces a long transcript of about 500-3000 nucleotides, known as pri-miRNA. This molecule is subsequently cropped by the ribonuclease DROSHA, forming a 60-100 nucleotides molecule, named pre-miRNA. This double strand structure is exported from the nucleus to the cytoplasm by RAN GTPase and exportin 5 (XPO5). Within the cytoplasm, the pre-miRNA is cleaved by the enzyme DICER1, forming a mature miRNA of about 20 nucleotides. The mature miRNA is selected by an argonaute protein (AGO2) and incorporated, with other associated proteins, in the RNA-inducing silencing complex (RISC). The RISC complex, containing the mature miRNA, is then competent to target mRNA.
To date, different SNPs located in miRNAs have been investigated as associated with breast cancer risk. Shen et al. reported the association of rs2910164, located in miR-146a, with increased risk of developing breast cancer in Chinese breast cancer cases with early age at onset (Shen et al., 2008). However, this association was not confirmed by subsequent studies in Caucasian (Catucci et al., 2010) and in Chinese (Hu et al., 2009) populations, and in different meta-analyses (Gao et al., 2011; Tian et al., 2010; Xu et al., 2011; Lian et al., 2012; Wang et al., 2012).

In 2009, rs11614913, located in miR-196a2, was reported as associated with increased breast cancer risk in Chinese breast cancer cases by Hu et al. (Hu et al., 2009), with an OR = 1.23 (95% confidence interval [CI] 1.02-1.48, \( P = 0.032 \)). Subsequently, different meta-analyses confirmed this association in breast cancer and also in other types of cancer (Tian et al., 2010; Gao et al., 2011; Wang et al., 2012; Qiu et al., 2011), whereas these data were not confirmed in Italian breast cancer cases (Catucci et al., 2010).

A case-control study performed in Antwerp, Belgium, reported a strong association of rs12975333, located in miR-125a, with breast cancer risk (Li et al., 2009). In this study, the rare minor allele [T] was detected in 6/72 (8.3%) breast cancer cases and in none of 869 Caucasian controls, including 289 recruited in the Antwerp area and 587 collected in the USA. However, a large multicenter study, performed in 2011, failed to confirm this association in German, Italian, Spanish and Australian breast cancer cases (Peterlongo et al., 2011).

In 2009, the miR-27a SNP rs895819 was investigated in a series of 1,217 German familial breast cancer cases and 1,422 unrelated German controls (Yang et al., 2009). This analysis showed an association of the minor allele [G] with reduced breast cancer risk, with a OR = 0.88 (95% CI 0.78-0.99, \( P = 0.0287 \)). Additional analyses indicated also that the protective effect was limited to cases with age at diagnosis <50 years (OR = 0.83, 95%
CI 0.70-0.98, \( P = 0.0314 \)), whereas a stronger effect was observed in bilateral breast cancer cases (OR = 0.70, 95% CI 0.52-0.95, \( P = 0.238 \)).

1.5 Genetic risk modifiers in BRCA genes mutation carriers

To date, it has been estimated that the cumulative average risk to develop breast cancer by the age of 70 in BRCA gene mutation carriers is about 65% in \( BRCA1 \) and 45% in \( BRCA2 \) mutation carriers (reviewed in Milne and Antoniou, 2011). However, it has been also observed that the risk conferred by a specific \( BRCA1 \) or \( BRCA2 \) mutation may be variable among carriers from different families and also from the same family where the mutation segregates. Such evidence suggests the presence of additional factors, including genetic modifiers, that could modulate the risk conferred by \( BRCA1 \) and \( BRCA2 \) mutations in these families.

In this context, as reported for the detection of the low-risk alleles, two main different approaches, the candidate gene strategy and the GWASs, were used to identify genetic risk modifiers. With the gene candidate approach, several common SNPs were screened as plausible modifiers as they are located in genes possibly acting as risk factors for the disease and in genes involved in \( BRCA1 \) and \( BRCA2 \) pathway, or functionally interacting with them. Although these studies reported several positive associations between candidate SNPs and increased breast cancer risk, the large part of these associations failed to be confirmed by replication studies (reviewed in Chenevix-Trench et al., 2007). As an example of the investigation of modifiers in candidate genes, Rebbeck and colleagues analyzed two CAG repeat length polymorphisms located in the androgen receptor (\( AR \)) gene and in the nuclear receptor coactivator 3 (\( NCOA3 \), also referred to as \( AIB1 \)) gene,
encoding a hormone receptor and a receptor interacting protein, respectively. They suggested an association between the alleles containing more than 28 repeats and increased breast cancer risk (Rebbeck et al., 1999; Rebbeck et al., 2001), but this association failed to be confirmed in subsequent studies (reviewed in Chenevix-Trench et al., 2007). An interesting exception was represented by the RAD51C/135G>C polymorphism (rs1801320). Wang and colleagues reported the association of the minor allele [C] with increased breast cancer risk in BRCA2 mutation carriers (Wang et al., 2001). Subsequently, this association was confirmed in two additional studies (Levy-Lahad et al., 2001; Kadouri et al., 2004). The apparent difficulty of reproducing results obtained in the original studies was probably due to the limitations of the candidate gene approach and the small sample size of each single study, that often does not allow enough statistical power. As described for low-penetrance alleles, these issues were overcome by establishing larger studies and consortia.

The Consortium of Investigators of Modifiers of BRCA1 and BRCA2 (CIMBA) was established in 2005 to provide sufficient sample size, combining DNAs and data from several studies. As concerned the gene candidate approach, the RAD51C/135G>C SNP was re-genotyped in more than 8,500 female BRCA1 and BRCA2 mutation carriers from 19 different studies. This analysis indicated that carriers of two copies of the [C] allele had three-fold increased risk of developing breast cancer (hazard ratio [HR] = 3.18), confirming the previous data (Antoniou et al., 2007). In a following study, the CASP8/D302H polymorphism was investigated by CIMBA because of the reported association with reduced breast cancer risk in the general population (Frank et al., 2006). CIMBA analysis showed a significant association of the minor allele with reduced breast cancer risk, as reported in the general population, but only in BRCA1 mutation carriers, with a HR = 0.85 (Engel et al., 2010).
Recently, the investigation of novel candidate breast cancer risk modifiers has been focused on low-penetrance alleles that were found to be strongly associated with risk by GWASs. These common variants were screened as risk modifiers in a large series of BRCA1 and BRCA2 mutation carriers collected by the CIMBA.

In the first two studies, a total of six variants were genotyped. Of them, five, rs2981522 in FGFR2, rs3803662 in TOX3/TNRC9, rs889312 in MAP3K1, rs3817198 in LSP1 and rs13387042 in the 2q35 region, were associated with increased risk in BRCA2 mutation carriers whereas only two, rs3803662 in the TOX3/TNRC9 gene and rs13387042 in the 2q35 region, were associated with increased risk in BRCA1 mutation carriers. No association was found for rs13281615 in the 8q24 region. (Antoniou et al., 2008; Antoniou et al., 2009). The estimated relative risk conferred by these variants in BRCA1 and BRCA2 mutation carriers was comparable to that observed in the general population.

In additional studies, novel low-penetrance alleles were tested as genetic modifiers. This genotyping showed an association of two SNPs, rs2046210 and rs9397435 both located in the 6q25 region, with increased risk in BRCA1 mutation carriers and another four variants, rs4973768 in the NEK10/SLC4A7 gene, rs10941679 in the 5p12 region, rs9397435 in the 6q25.1 region and rs11249433 in the 1p11.2 region, associated with increased risk in BRCA2 mutation carriers (Antoniou et al., 2010; Antoniou et al., 2011). These findings strongly suggest that common variants that are associated with breast cancer risk in the general population may also act as genetic risk modifiers in BRCA mutation carriers. The differential association, observed between BRCA1 and BRCA2 mutation carriers, of the risk conferred by these genetic modifiers probably reflects the distinct biology features of BRCA1- and BRCA2-related tumors, related to estrogen receptor (ER) status (reviewed in Antoniou and Chenevix-Trench, 2010).

Very recently, two GWASs were performed in BRCA1 and BRCA2 mutation carriers, confirming the association with a few previously identified loci and identifying novel
genetic modifiers in both genes. In BRCA1 mutation carriers, two known SNPs located in the LSP1 and RAD51L1 genes were found to be associated with increased risk and one novel locus was also detected. This is located at 1q32 and contains the oncogene MDM4 (Couch et al., 2013). In BRCA2 mutation carriers, rs4733664, located in the 8q24 region, rs16917302 and rs17221319, located in the ZNF356 gene, rs311499, located in the 20q13 region and the rs27633, located in the 12p11 region, were found associated with increased breast cancer risk (Gaudet et al., 2013). All these identified breast cancer risk genetic modifiers are reported in Tables 1.4 and 1.5.

<table>
<thead>
<tr>
<th>Gene/Region</th>
<th>SNP</th>
<th>HR (95% CI)</th>
<th>P</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CASP8</td>
<td>rs1045485</td>
<td>0.85 (0.76-0.97)</td>
<td>0.028</td>
<td>Engel et al., 2010</td>
</tr>
<tr>
<td>TOX3/TNRC9</td>
<td>rs3803662</td>
<td>1.11 (1.03-1.19)</td>
<td>0.004</td>
<td>Antoniou et al., 2008</td>
</tr>
<tr>
<td>2q35</td>
<td>rs13387042</td>
<td>1.14 (1.04-1.25)*</td>
<td>0.005</td>
<td>Antoniou et al., 2009</td>
</tr>
<tr>
<td>6q25.1</td>
<td>rs2046210</td>
<td>1.17 (1.11-1.23)</td>
<td>4.5x10^-9</td>
<td>Antoniou et al., 2011</td>
</tr>
<tr>
<td>6q25.1</td>
<td>rs997435</td>
<td>1.28 (1.18-1.40)</td>
<td>1.3x10^-8</td>
<td>Antoniou et al., 2011</td>
</tr>
<tr>
<td>LSP1</td>
<td>rs3817198</td>
<td>1.09 (1.04-1.14)</td>
<td>9.4x10^-4</td>
<td>Couch et al., 2013</td>
</tr>
<tr>
<td>RAD51L1</td>
<td>rs999737</td>
<td>0.94 (0.89-0.99)</td>
<td>0.035</td>
<td>Couch et al., 2013</td>
</tr>
<tr>
<td>1q32</td>
<td>rs2290854</td>
<td>1.14 (1.09-1.20)</td>
<td>2.7x10^-8</td>
<td>Couch et al., 2013</td>
</tr>
</tbody>
</table>

HIR hazard ratio, *HIR under a dominant model

Likewise in the general population, it is estimated that each of the identified risk modifiers conferred a very small increase of breast cancer risk in BRCA1 and BRCA2 mutation carriers, with HIRs ≤ 1.32. The larger increase is conferred by rs2981522 in the FGFR2 gene in carriers of BRCA2 mutation. Nevertheless, different studies suggested that the combined effect of these alleles accounts for a larger risk increase, depending on the number of risk alleles carried, under a multiplicative model (reviewed in Milne and Antoniou, 2011). For example, Antoniou and colleagues estimated the breast cancer risk for BRCA2 mutation carriers combining the genotype distribution of seven different SNPs.
Chapter I

Introduction

(Antoniou et al., 2010). The risk of developing breast cancer by age 50 for the 5% of the mutation carriers at lower risk was predicted to be 10-13%, whereas this risk is 29-47% for the 5% of the mutation carriers at higher risk (Figure 1.4).

Table 1.5. SNPs associated with breast cancer risk modification in BRCA2 mutation carriers

<table>
<thead>
<tr>
<th>Gene/Region</th>
<th>SNP</th>
<th>HR (95% CI)</th>
<th>P</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAD51C</td>
<td>rs1801320</td>
<td>3.18 (1.39-7.27)</td>
<td>0.0007</td>
<td>Antoniou et al., 2007</td>
</tr>
<tr>
<td>FGFR2</td>
<td>rs2981522</td>
<td>1.32 (1.20-1.45)</td>
<td>2x10^-8</td>
<td>Antoniou et al., 2008</td>
</tr>
<tr>
<td>FGFR2</td>
<td>rs2420946</td>
<td>1.27 (1.19, 1.34)</td>
<td>2x10^-14</td>
<td>Gaudet et al., 2013</td>
</tr>
<tr>
<td>TOX3/TNRC9</td>
<td>rs3803662</td>
<td>1.15 (1.03-1.27)</td>
<td>0.009</td>
<td>Antoniou et al., 2008</td>
</tr>
<tr>
<td>MAP3K1</td>
<td>rs889312</td>
<td>1.12 (1.02-1.24)</td>
<td>0.02</td>
<td>Antoniou et al., 2008</td>
</tr>
<tr>
<td>LSP1</td>
<td>rs3817198</td>
<td>1.16 (1.07-1.25)</td>
<td>0.0003</td>
<td>Antoniou et al., 2009</td>
</tr>
<tr>
<td>2q35</td>
<td>rs13387042</td>
<td>1.18 (1.04-1.33)</td>
<td>0.008</td>
<td>Antoniou et al., 2009</td>
</tr>
<tr>
<td>NEK10/SLC4A7</td>
<td>rs4973768</td>
<td>1.10 (1.03-1.18)</td>
<td>0.0064</td>
<td>Antoniou et al., 2010</td>
</tr>
<tr>
<td>5p12</td>
<td>rs10941679</td>
<td>1.15 (1.04-1.27)</td>
<td>0.0083</td>
<td>Antoniou et al., 2010</td>
</tr>
<tr>
<td>6q25.1</td>
<td>rs9397435</td>
<td>1.14 (1.01-1.28)</td>
<td>0.031</td>
<td>Antoniou et al., 2011</td>
</tr>
<tr>
<td>1p11.2</td>
<td>rs11249433</td>
<td>1.09 (1.02-1.17)</td>
<td>0.015</td>
<td>Antoniou et al., 2011</td>
</tr>
<tr>
<td>MAP3K1</td>
<td>rs16886113</td>
<td>1.24 (1.11, 1.38)</td>
<td>1x10^-4</td>
<td>Gaudet et al., 2013</td>
</tr>
<tr>
<td>8q24</td>
<td>rs4733664</td>
<td>1.10 (1.04, 1.17)</td>
<td>1.7x10^-3</td>
<td>Gaudet et al., 2013</td>
</tr>
<tr>
<td>ZNF365</td>
<td>rs16917302</td>
<td>0.88 (0.80, 0.98)</td>
<td>0.01</td>
<td>Gaudet et al., 2013</td>
</tr>
<tr>
<td>ZNF365</td>
<td>rs17221319</td>
<td>1.09 (1.02, 1.15)</td>
<td>6x10^-3</td>
<td>Gaudet et al., 2013</td>
</tr>
<tr>
<td>20q13</td>
<td>rs13039229</td>
<td>0.90 (0.84, 0.97)</td>
<td>5x10^-3</td>
<td>Gaudet et al., 2013</td>
</tr>
<tr>
<td>12p11</td>
<td>rs27633</td>
<td>1.14 (1.07, 1.21)</td>
<td>4x10^-5</td>
<td>Gaudet et al., 2013</td>
</tr>
</tbody>
</table>

HR hazard ratio, *HR under a dominant model

While the impact of the SNP profiles in the assessment of breast cancer risk in the general population is at present very limited, because of the very small risk increase conferred, in BRCA mutation carriers this profiling may be of clinical relevance. In fact, the combined relative risk of several alleles results in much larger differences, compared to the general population, in the absolute lifetime risk of developing breast cancer, since BRCA mutation carriers are individuals with already high risk of the disease (reviewed in Milne and Antoniou, 2011; Figure 1.5).
Figure 1.4. Breast cancer age-specific cumulative risk for BRCA2 mutation carriers. This risk has been estimated by combining the genotype distribution of the following SNPs: rs2981582 in FGFR2, rs3803662 in TOX3/TNRC9, rs889312 in MAP3K1, rs3817198 in LSP1, rs13387042 in the 2q35 region, rs4973768 in NEK10/SLC4A7 and rs10941679 in the 5p12 region.

Figure 1.5. Comparison of predicted age-specific cumulative breast cancer risks in the general population (on the left) and in BRCA2 mutation carriers (on the right). This predicted risk is based on the combined genotypes of 18 different SNPs associated with breast cancer risk in the general population. The figure estimates that the absolute lifetime risk of developing breast cancer in the general population varies from 5.7% to 19%, whereas this risk in BRCA2 mutation carriers varies from 47% to 89%.
1.5.1 The rs3834129 in the CASP8 promoter region as candidate genetic risk modifier in BRCA genes mutation carriers

The CASP8 gene encodes a protein of the cysteine-aspartic acid protease (caspase) family. Caspases play a crucial role in apoptosis, a process of programmed cell death essential for controlling cell proliferation and cancer development. In particular, CASP8 is an apical caspase involved in the apoptosis activation mediated by death receptors and their ligands, cooperating with CASP10 and CFLAR (Hlo and Howkins, 2005). Because of the crucial role of these proteins, it has been suggested that mutations in their genes may be associated with breast cancer susceptibility.

The initial report observed the association of the CASP8/D302H (rs1045485) with a reduced breast cancer risk (OR = 0.83; 95% CI 0.74-0.94 for heterozygotes and OR = 0.58; 95% CI 0.39-0.88 for homozygotes) (MacPherson et al., 2004). This association was confirmed by two additional studies (Cox et al., 2007; Sergentanis et al., 2010).

Subsequently, an additional SNP in the CASP8 gene was studied as a susceptibility allele. Sun and colleagues investigated rs3834129, a common six-nucleotide (AGTAAG) insertion/deletion, located at position -652 in the promoter region of the gene (-652 6N ins/del) (Figure 1.6). Firstly, they observed that the deletion of these six nucleotides destroys the binding site for the transcriptional activator Sp1, decreasing the gene transcription. Furthermore, they performed a case-control analysis and found a significant association of the del allele with a reduced risk of developing breast cancer, with an OR of 0.65 (95% CI 0.54-0.78) for heterozygotes and 0.50 (95% CI 0.34-0.74) for homozygotes, in Chinese unselected cases (Sun et al., 2007). However, this association was not confirmed in three different studies performed in unselected Caucasian breast cancer cases (Frank et al., 2008; Cybulski et al., 2008; Haiman et al., 2008).
Italian familial breast cancer cases negative for mutation in *BRCA1* and *BRCA2* were also genotyped for this SNP and, although this analysis failed to confirm the association of rs3834129 with breast cancer risk, a case-only analysis suggested a significant association of the del/del genotype with increased age at diagnosis (De Vecchi et al., 2009). Recently, two different meta-analyses were also performed. The first one, based on the five previously described studies, showed a borderline association of the del allele with a reduced breast cancer risk, with an OR = 0.94 (95% CI 0.884-1.008), suggesting that, in each of the above mentioned studies, sample size was inadequate (Sergentanis and Economopoulos, 2009). In the second meta-analysis, four of the five above mentioned studies were included to test the association of rs3834129 and breast cancer risk. Here, it has been confirmed the protective effect of the del allele, with an OR of 0.95 (95% CI 0.83-1.08) for heterozygotes and 0.82 (95% CI 0.70-0.95) for homozygotes (Yin et al., 2010).

As mentioned above, recent studies have indicated that common low-risk alleles are also responsible for risk variability in *BRCA1* and *BRCA2* mutation carriers (Antoniou et al., 2008; Antoniou et al., 2009; Antoniou et al., 2010; Antoniou et al., 2011). Among them the *CASP8* D302H variant was reported, reducing breast cancer risk in *BRCA1* mutation
carriers (HR = 0.85) (Engel et al., 2010). These findings suggest that also other $CASP8$

gene variants, specifically the -652 6N ins/del, might be responsible for breast cancer risk

variation in $BRCA1$ and $BRCA2$ mutation carriers.
The major aim of this thesis was to investigate candidate moderate- and low-penetrance genes and alleles and genetic risk modifiers associated with breast cancer susceptibility in familial cases. This project can be considered as a part of a wider study aimed at the identification of as many as possible of these genetic factors to allow a more accurate prediction of the individual breast cancer risk both in the general population and in breast cancer families. In the first part of this thesis, I investigated the role of the two FA genes \textit{PALB2} and \textit{SLX4} as candidate moderate-penetrance loci, in affected individuals negative for mutations in \textit{BRCA1} and \textit{BRCA2}. A mutation screening of the entire coding region and splice sites of these genes were performed in a large series of familial breast cancer cases to verify the association of mutations of \textit{PALB2} and \textit{SLX4} with increased breast cancer risk.

In the second part of this thesis, a case-control study was performed in a large series of familial breast cancer cases to investigate the role of rs895819, located in the gene coding for miR-27a, as a low-penetrance allele. In particular, I wanted to verify the association of this SNP with reduced breast cancer risk, previously reported in the German population.

Finally, the rs3834129 SNP located in the promoter region of the \textit{CASP8} gene, was investigated as a genetic risk modifier in a large series of affected and unaffected individuals carrying a \textit{BRCA1} or \textit{BRCA2} mutation.
CHAPTER 3
MATERIALS AND METHODS

3.1 Study population

3.1.1 Genetic counseling and eligibility to BRCA1 and BRCA2 genetic test

Individuals affected with breast cancer or at-risk for breast cancer because of their family history underwent genetic counseling in the cancer genetic clinics of different centers, collaborating with our research group. The majority of individuals were recruited through the Medical Genetic Unit of the “Fondazione IRCCS Istituto Nazionale dei Tumori” (INT) and the Division of Cancer Prevention and Genetics of the “Istituto Europeo di Oncologia” (IEO), in Milan. Additional individuals were recruited through the Unit of Medical Oncology of Azienda Ospedaliera Ospedali Riuniti of Bergamo and the cancer genetic clinics of other six centers participating in the “Consorzio degli Studi Italiani sul Tumore Ereditario alla Mammella” (CONSID TEAM; Consortium of Italian Studies on Hereditary Breast Cancer) including: Università degli Studi in Turin, Centro Riferimento Oncologico in Aviano, Università “La Sapienza” and Istituto Nazionale Tumori “Regina Elena” in Rome, Università degli Studi in Florence and “Istituto Nazionale per la Ricerca sul Cancro” in Genoa; these individuals were interviewed following the specific protocol of each center, to collect a detailed family history of cancer or other disease, and to reconstruct detailed pedigrees. When possible, the diagnoses of reported cancers were verified by medical records.

Individuals recruited at INT and IEO were considered eligible for mutation screening in BRCA1 and BRCA2 when fulfilling the following criteria, based on tumor type, age at onset and family history of cancer:
• Female individuals affected with breast cancer <36 years, or breast cancer and ovarian cancer at any age, or male patients affected with breast cancer at any age, independently of family history;

• Members of families with three or more first degree relatives (or second degree, if in paternal lineage) affected with breast cancer or ovarian cancer at any age;

• Members of families with two first degree relatives (or second degree, if in paternal lineage) with the following features:
  -both affected with breast cancer <50 years,
  -one affected with breast cancer <50 years and the other with either bilateral breast cancer, or ovarian cancer, or male breast cancer at any age
  -both affected with ovarian cancer at any age.

In the other collaborating centers, individuals were recruited following very similar criteria. Family history information and pedigrees were subsequently communicated to the diagnostic and research laboratories. For all eligible individuals, the \textit{BRCA1} and \textit{BRCA2} genetic tests were performed. Individuals from INT, IEO and Azienda Ospedaliera Ospedali Riuniti of Bergamo were screened in the diagnostic laboratory located at the Fondazione Istituto FIRC di Oncologia Molecolare, in Milan (IFOM).

\textbf{3.1.2 Recruitment and inclusion criteria to research studies of individuals affected with breast cancer or at-risk for breast cancer}

For the recruitment of individuals who underwent a BRCA gene test into the studies here described an informed consent for the use of their biological samples for research purpose was required. Blood samples of these individuals, including index cases, and their relatives, were sent to our laboratory. In the present project, the following two groups of individuals were included:
- Female individuals with disease causing mutations in *BRCA1* or *BRCA2*, with or without a diagnosis of cancer (BRCA mutation carriers);
- Female individuals negative for disease causing mutations in *BRCA1* and *BRCA2*, with a personal history of breast cancer as the first diagnosed (BRCAX cases).

The entire recruitment process is described in Figure 3.1.

![Figure 3.1.](image-url)  
*Based on individuals recruited at the Medical Genetic Unit of INT from 2009 to 2012. **Based on individuals recruited from all collaborating centers in 2012. *BRCA1* and *BRCA2* mutation positive individuals include all ascertained carriers of examined families. As for BRCAX cases, only the family probands were considered.

### 3.1.3 Recruiting of blood donors

Normal controls were female blood donors consecutively recruited through the Immunohematology and Transfusion Medicine Unit of INT and the Associazione
Volontari Italiani Sangue (AVIS) agencies, in Milan and Bergamo. All individuals signed an inform consent to the use of their biological samples for research purpose.

3.2 Association studies

Association studies (or case-controls studies) are analyses that allow evaluation of the association of a specific variant with a particular phenotype. These analyses are performed measuring the frequencies of candidate genetic variants in two different groups represented by individuals affected with a disease (cases) and individuals unaffected (controls) (Mann et al., 2003). Statistically significant differences in the frequencies measured in cases versus controls indicate the association with the disease. In addition, the odds ratio\(^3\) (OR) calculation allows to define the role of the associated variant in increasing or decreasing the risk of developing the disease. In particular, an OR > 1 indicates that the variant is a risk factor, whereas an OR < 1 indicates that the variant act as a protective factor for the disease (Turnbull and Rahman, 2008). However, one of the major limitations of these studies is the sample size. In fact, underpowered studies did not allow enough statistical power to demonstrate the association of a genetic variant with the disease. This limitation is even more evident for variants with very low frequencies (Turnbull and Rahman, 2008). The establishment of national and international consortia became indispensable for collecting a large number of cases and controls, hundreds of thousands of samples, to ensure higher level of statistical significance.

\(^3\) The odds ratio represents the measure of the association between the exposure to a risk factor and the disease in association studies.
3.3 DNA samples preparation

Blood samples were obtained from the diagnostic laboratory, after mutation screening in \textit{BRCA1} and \textit{BRCA2} genes. DNA was extracted from peripheral blood, using the Gentra Puregene Blood Kit (QIAGEN), according to manufacturer's protocol. After DNA extraction, samples were quantified by Thermo Scientific NanoDrop\textsuperscript{TM} 1000 (NanoDrop), diluted at a final concentration of 25 ng/\textmu l in the DNA Hydration Solution, included in the commercial kit, and distributed in 96-well plates.

3.4 Mutation screenings

3.4.1 PCRs conditions

DNA amplification was performed using the polymerase chain reaction (PCR), for the entire coding region and intron/exon junctions of the \textit{PALB2} and \textit{SLX4} genes. For the \textit{CASP8} gene, only the promoter region including rs3834129 was amplified. Each PCR was performed in a total volume of 15 \textmu l, containing 30 ng of genomic DNA, 1X PCR Buffer, 2.5 mM MgCl\textsubscript{2}, 0.2 mM of each dNTP, 0.5 \mu M of each primer, 0.75 units of \textregistered Taq (EuroClone). Amplification was carried out as follow: 32 PCR cycles consisting of a denaturation step at 94°C for 30 seconds, an annealing step at different temperature for each fragment for 30 seconds and an extension step at 72°C for 30 seconds, followed by a final extension step at 72°C for 5 minutes. The \textit{CASP8} region containing rs3834129 was amplified using primers specifically designed. The amplification of \textit{PALB2} fragments was performed using primers described by Reid et al. (Reid et al., 2007). \textit{SLX4} fragments were amplified using primers described by Stoepker et al. (Stoepker et al., 2011), with the
exception of primers for exons 5, 8, 12B-F, 13 and 15 that were redesigned. The amplification of SLX4 exon 8 was optimized using 5% DMSO. Primer sequences and PCR conditions are described in Tables 3.1, 3.2 and 3.3. All PCR products were checked by electrophoresis on 1.5% agarose gels.

### 3.4.2 Mutation detection by sequencing

Mutation screening of the PALB2 and SLX4 genes and the genotyping of rs3834129 were performed at the DNA Sequencing Unit of the Technological Service at IFOM, using single strand sequence analysis.

PCR fragments were sequenced using the chain termination sequencing (Sanger sequencing) method on ABI 3730xl or ABI 3500dx sequence analyzers (Life Technologies). Sequencing results were analyzed using the DNA Sequence Analysis Software Sequencher 5.0 (GeneCode Corporation). Identified truncating mutations were confirmed by double strand DNA sequencing.

<p>| Table 3.1. CASP8 primer sequences and PCR conditions |
|------------------------------|---------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer Sequence</th>
<th>PCR Fragment (bp)</th>
<th>Annealing temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CASP8F</td>
<td>TCCCCGCTGTTAAACATTTTG</td>
<td>228</td>
<td>60</td>
</tr>
<tr>
<td>CASP8R</td>
<td>CTGCATCCAGAGCTAAGT</td>
<td>228</td>
<td>60</td>
</tr>
<tr>
<td>Primer Name</td>
<td>Primer Sequence</td>
<td>PCR Fragment (bp)</td>
<td>Annealing temperature (°C)</td>
</tr>
<tr>
<td>------------</td>
<td>-------------------------------------</td>
<td>------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>1F</td>
<td>GGATTTAATTGCGGCCGGAGTT</td>
<td>309</td>
<td>59</td>
</tr>
<tr>
<td>1R</td>
<td>GACACAAAGCCAGGCTGAAAA</td>
<td>400</td>
<td>59</td>
</tr>
<tr>
<td>2-3F</td>
<td>ACCTTTCCACTTGCCAGTA</td>
<td>495</td>
<td>59</td>
</tr>
<tr>
<td>2-3R</td>
<td>GGGAAAAAGAACAATAGGCCAAA</td>
<td>387</td>
<td>59</td>
</tr>
<tr>
<td>4AF</td>
<td>GCCTGAATGAAATGTCAGTATT</td>
<td>357</td>
<td>59</td>
</tr>
<tr>
<td>4AR</td>
<td>GCAAAAATCTCTGCTAGTACACC</td>
<td>388</td>
<td>58</td>
</tr>
<tr>
<td>4BF</td>
<td>CCCCTAGGGTGAGCACAAGG</td>
<td>589</td>
<td>65</td>
</tr>
<tr>
<td>4BR</td>
<td>TCAAGGTGCTGACTACTACGG</td>
<td>398</td>
<td>58</td>
</tr>
<tr>
<td>4CF</td>
<td>ACCAAGTGCACACACACAG</td>
<td>395</td>
<td>59</td>
</tr>
<tr>
<td>4CR</td>
<td>GGTTTTTCTATTGCTGGTAGG</td>
<td>395</td>
<td>59</td>
</tr>
<tr>
<td>4DF</td>
<td>AAAGAGGGAAGCTGAGCCAATTGGA</td>
<td>398</td>
<td>58</td>
</tr>
<tr>
<td>4DR</td>
<td>TTTTTCCCTTGGATCAGGAAAATGACTCGC</td>
<td>389</td>
<td>59</td>
</tr>
<tr>
<td>4EF</td>
<td>GCAGAAAAACATTCTGTGCCACA</td>
<td>589</td>
<td>65</td>
</tr>
<tr>
<td>4ER</td>
<td>AAGGAAGTGCCAAGCCAAATGAA</td>
<td>213</td>
<td>59</td>
</tr>
<tr>
<td>5AF</td>
<td>GATTGTCTGTTGGGGTTTTG</td>
<td>293</td>
<td>62</td>
</tr>
<tr>
<td>5AR</td>
<td>GTGCTGCTTTCAAGCTTCCATT</td>
<td>221</td>
<td>58</td>
</tr>
<tr>
<td>5BF</td>
<td>AAAGAGGGAAGCTGAGCCAATTGGA</td>
<td>287</td>
<td>64</td>
</tr>
<tr>
<td>5BR</td>
<td>CACCTGTCTTTCCATCTTTA</td>
<td>250</td>
<td>58</td>
</tr>
<tr>
<td>5CR</td>
<td>GGCATTTCCATTCCTTCCAGAGA</td>
<td>244</td>
<td>58</td>
</tr>
<tr>
<td>6F</td>
<td>AGTGGGTATGCAGGCGACA</td>
<td>240</td>
<td>59</td>
</tr>
<tr>
<td>6R</td>
<td>TGAATGATTCTTCTGCATCTATT</td>
<td>213</td>
<td>59</td>
</tr>
<tr>
<td>7F</td>
<td>TGCTTTGCAAAAAACACGACT</td>
<td>293</td>
<td>62</td>
</tr>
<tr>
<td>7R</td>
<td>TGTTAAAGCTGCCCATCTACA</td>
<td>221</td>
<td>58</td>
</tr>
<tr>
<td>8F</td>
<td>TGGAAAAATCTGGATTAACGAAAAA</td>
<td>250</td>
<td>58</td>
</tr>
<tr>
<td>8R</td>
<td>TGACCTAAAACCGAGCTGACA</td>
<td>287</td>
<td>64</td>
</tr>
<tr>
<td>9F</td>
<td>ATTTAAAAAGTTTACTCTCCATCATCAC</td>
<td>244</td>
<td>58</td>
</tr>
<tr>
<td>9R</td>
<td>CCCAATTTTCTCGAAACCTCAGT</td>
<td>241</td>
<td>58</td>
</tr>
<tr>
<td>10F</td>
<td>CAGAGACTGTTTATTAGTGCAGAA</td>
<td>250</td>
<td>58</td>
</tr>
<tr>
<td>10R</td>
<td>TTCACAAACACCTGTGAAAATTAG</td>
<td>281</td>
<td>59</td>
</tr>
<tr>
<td>11F</td>
<td>TTTTTGCAAACTCGTTGTTTGGA</td>
<td>400</td>
<td>58</td>
</tr>
<tr>
<td>11R</td>
<td>CGGGAAGTTGTTGTTTCATTA</td>
<td>281</td>
<td>59</td>
</tr>
<tr>
<td>12F</td>
<td>TGCCAGATCTTTTATTTTTCTCTGA</td>
<td>400</td>
<td>58</td>
</tr>
<tr>
<td>12R</td>
<td>TGGTTATCGACAGTGGCCTTT</td>
<td>281</td>
<td>59</td>
</tr>
<tr>
<td>13F</td>
<td>TGGTTTTGGGAAAAACGTTTT</td>
<td>400</td>
<td>58</td>
</tr>
<tr>
<td>13R</td>
<td>TTAAGTGTCACTCATATTTCTTCCTTT</td>
<td>281</td>
<td>59</td>
</tr>
</tbody>
</table>
Table 3.3. *SLX4* primer sequences and PCR conditions

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer Sequence</th>
<th>PCR Fragment (bp)</th>
<th>Annealing Temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S2F</td>
<td>TGTTTAACCAACAGGCCCAAT</td>
<td>707</td>
<td>60</td>
</tr>
<tr>
<td>S2R</td>
<td>GCCCTTTCCAGGAAGTTTTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S3F</td>
<td>ACCAACAGCAGACACATC</td>
<td>524</td>
<td>62</td>
</tr>
<tr>
<td>S3R</td>
<td>ATCCAGTGAGTTGGCAAAAGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S4F</td>
<td>TTCCCCGAGTGCTGATTAGT</td>
<td>500</td>
<td>62</td>
</tr>
<tr>
<td>S4R</td>
<td>ACAACAAAGCAGGTTGGCTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S5F</td>
<td>GACCCACATTTGCTCCAAATC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S5R</td>
<td>GTTGTTAAGCTACTAGTCCCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S6F</td>
<td>AACTTCTGAGCTGGAATGGA</td>
<td>518</td>
<td>59</td>
</tr>
<tr>
<td>S6R</td>
<td>ATACCCGGGTTTCTCCTTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S7F</td>
<td>CCAGAACAGGTTGTGGTGA</td>
<td>534</td>
<td>59</td>
</tr>
<tr>
<td>S7R</td>
<td>CTTTCTGAGCTTTTCCATCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S8F</td>
<td>GTGTTTTTAGGTCCAGCGTGATCA</td>
<td>494</td>
<td>64</td>
</tr>
<tr>
<td>S8R</td>
<td>AAAATGAAAGGCCCAGGAGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S9F</td>
<td>TCTCTTACTCTCCCTGGTTGA</td>
<td>440</td>
<td>62</td>
</tr>
<tr>
<td>S9R</td>
<td>CTCAACGGATGATCGATGTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S10F</td>
<td>GGTCCACTCGAGGCTGAGG</td>
<td>444</td>
<td>62</td>
</tr>
<tr>
<td>S10R</td>
<td>GCAGGAAAGTGAGGGAGAGTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S11F</td>
<td>AGGCTGCAAGATTACGATGAT</td>
<td>492</td>
<td>65.5</td>
</tr>
<tr>
<td>S11R</td>
<td>CTGACCTGAGCTTGAGGATT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S12AF</td>
<td>TGTTTCTGCGAAGAGTGTG</td>
<td>548</td>
<td>62</td>
</tr>
<tr>
<td>S12AR</td>
<td>CTCCACCTTGTCCTCCACTTT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S12BF</td>
<td>TACCTAGCGAAAGCTCTCTCA</td>
<td>587</td>
<td>62</td>
</tr>
<tr>
<td>S12BR</td>
<td>ACGACCCACTTTGTGATGAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S12CF</td>
<td>GAACAAAGTGGGCGCTGC</td>
<td>556</td>
<td>62</td>
</tr>
<tr>
<td>S12CR</td>
<td>GTCACAGGACCTAGGGCTAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S12DF</td>
<td>TCTTACTGCGACTGATGAGGA</td>
<td>547</td>
<td>62</td>
</tr>
<tr>
<td>S12DR</td>
<td>CGTCGAAAGTTTCCTGGAGAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S12EF</td>
<td>AAACAGGGGAAAGGAAAGGT</td>
<td>540</td>
<td>62</td>
</tr>
<tr>
<td>S12ER</td>
<td>GGGTGGGTGTCGAGGAGT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S12FF</td>
<td>AATCCAAATTGAGCGACTGCT</td>
<td>505</td>
<td>62</td>
</tr>
<tr>
<td>S12FR</td>
<td>AAGTGTCAGCCTACAGTCAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S13F</td>
<td>ACCACTGTTGCTTTTCACTGG</td>
<td>353</td>
<td>62</td>
</tr>
<tr>
<td>S13R</td>
<td>ACCAGACCAAGGACACACAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S14F</td>
<td>ATAGGGAAAGCGTGAGTGTG</td>
<td>588</td>
<td>62</td>
</tr>
<tr>
<td>S14R</td>
<td>GACGCGGTTTTTGAAGATT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S15F</td>
<td>CATGGACCCCCGAGACACC</td>
<td>522</td>
<td>62</td>
</tr>
<tr>
<td>S15R</td>
<td>CAGGCTCCCTTTGCAATG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.5 Genotyping analyses

Genotyping analyses were performed at the Real-Time PCR Unit of the Technological Service at IFOM.

The rs895819 SNP, located in the miR-27a gene region, was genotyped using a pre-designed TaqMan SNP Genotyping Assay (Life Technologies). The genotyping of the two recurrent PALB2 mutations, c.72delG and c.1027C>T, was performed using two custom TaqMan SNP Genotyping Assays (Table 3.4 and Table 3.5). Primers design was provided by the manufacturer.

Table 3.4. TaqMan assay information of the PALB2 c.72delG mutation

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>72delG_F</td>
<td>TGGTGGTTTTTCTTCTTCCAGTTAAAGGA</td>
</tr>
<tr>
<td>72delG_R</td>
<td>GCGGGCTAGTGTCTTGCT</td>
</tr>
<tr>
<td>Reporter Name</td>
<td>Reporter Sequence</td>
</tr>
<tr>
<td>72delG_V</td>
<td>TTCCCTTTTCAAGAAATG</td>
</tr>
<tr>
<td>72delG_M</td>
<td>ATTCCTTTT-AAGAATG</td>
</tr>
</tbody>
</table>

Table 3.5. TaqMan assay information of the PALB2 c.1027C>T mutation

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1027CT_F</td>
<td>AACTCACCTACAATAAACTTACCAGCAA</td>
</tr>
<tr>
<td>1027CT_R</td>
<td>CAAGAGTGTCACTGGAGATTTAAAGA</td>
</tr>
<tr>
<td>Reporter Name</td>
<td>Reporter Sequence</td>
</tr>
<tr>
<td>1027CT_V</td>
<td>TTCTTTAAGTTTGGTTTC</td>
</tr>
<tr>
<td>1027CT_M</td>
<td>TTCTTTAAGTTTGTGTTTC</td>
</tr>
</tbody>
</table>
For all the genotyping tests, the reactions were performed in a total volume of 8µl, containing 20 ng of genomic DNA, using the ABI 7500 Fast Real Time PCR System (Life Technologies). Amplification was carried out as follow: an initial hold step at 95°C for 20 seconds, followed by 40 PCR cycles consisting of a denaturation step at 95°C for 3 seconds and an annealing/extension step at 60°C for 30 seconds and by a final post-PCR read step at 60°C for 1 minute. In each of the 96-well plates, three duplicate samples, one non-DNA blank control and one positive control sample were included. Genotypes for all duplicates and positive controls were completely concordant.

3.6 In silico analyses

3.6.1 Software for prediction of the missense mutations effect

For the protein prediction analysis, the following software were used:

- PolyPhen-2 (Polymorphism Phenotyping v2; http://genetics.bwh.harvard.edu/pph2/) predicts the results of an amino acid substitution based of the comparison between the wild-type and the mutant allele, using physical and evolutionary comparative characteristics (Adzhubei et al., 2010);
- SIFT (Sorts Intolerant From Tolerant; http://sift.jcvi.org/) predicts the consequences of an amino acid substitution based on the amino acid conservation, assuming that functionally important positions should be conserved in an alignment of the protein family, whereas unimportant positions should appear diverse in an alignment (Kumar et al., 2009);
- SNPs&GO (http://snps.biofold.org/snps-and-go/snps-and-go.html) predicts the results of an amino acid substitution collecting data from protein sequence, protein sequence profile, and protein function (Calabrese et al., 2009).
3.6.2 Software for prediction of non-canonical splicing

Analyses of candidate splicing aberrations were performed using the following prediction software:

- Berkeley Drosophila Genome Project (BDGP; http://www.fruitfly.org/seq_tools/splice.html; Reese et al., 1997);
- NetGene2 (http://www.cbs.dtu.dk/services/NetGene2/; Brunak et al., 1991; Hebsgaard et al., 1996);
- SpliceView (http://zeus2.itb.cnr.it/~webgene/wwwspliceview_ex.html; Rogozin et al., 1997);
- MaxEntScan (http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html; Yeo et al., 2004);
- SplicePredictor (http://deepe2.psi.iastate.edu/cgi-bin/sp.cgi; Brendel et al., 2004);
- GeneSplicer (http://www.cbcb.umd.edu/software/GeneSplicer/gene_spl.shtml; Pertea et al., 2001);
- Automated Splice Site and Exon Definition Analyses (ASSEDA) (http://splice.uwo.ca/; Mucaki et al., 2013).

3.7 Investigation of the PALB2 c.48G>A splicing mutation

3.7.1 B-lymphocytes immortalization

The lymphoblastoid cell line carrying the PALB2 c.48G>A mutation was obtained by transformation of peripheral B-lymphocytes obtained from a carrier of the variant using Epstein-Barr virus (EBV) infection, according to the following protocols. This was kindly made available by the Dr. Mara Colombo of INT.
Chapter 3 - Materials and Methods

I. Collection of B-lymphocytes from peripheral blood sample. Lymphocytes were isolated from whole blood sample by gradient centrifugation with Ficoll-Paque (Amersham Pharmacia Biotech AB). Briefly, blood sample was diluted 1:2 in RPMI 1640 medium (Biowhittaker Europe), stratified on equal volume of Ficoll-Paque and centrifuged at 1800 rpm for 25 minutes. After the centrifugation, B-lymphocytes were recovered, washed twice with RPMI 1640 medium and centrifuged at 1800 rpm for 10 minutes.

II. Preparation of the viral solution. The viral stock was obtained from the B95-8 cell line. These cells were expanded in RPMI 1640 medium supplemented with 10% Fetal Bovine Serum (FBS, EuroClone) and 1% penicillin/streptomycin, at 37°C, 5% CO₂. After cells became confluent, the recovered supernatant was centrifuged at 1500 rpm for 5 minutes and filtered with 0.22μm filter (Millipore).

III. Immortalization with EBV. B-lymphocytes were resuspended with 1.5ml of RPMI 1640 medium, 15% FBS, 4% Glutamine and 1% penicillin/streptomycin, supplemented with 1.5ml of the viral solution and 2μg of CSA (Ciclosporin A, Sandimmun Sandoz-Wander PHARMA S.A.) and cultured at 37°C, 5% CO₂. For the complete transformation, one month was required.

IV. Cell culture conditions. Lymphoblastoid cell line was cultured using RPMI 1640, 15% FBS, 1% penicillin/streptomycin, 25mM Hepes, until their use for RNA extraction.

3.7.2 RNA extraction and cDNA synthesis

RNA extraction was performed using the NucleoSpin RNA II Kit (Macherey-Nagel), according to the manufacturer’s protocol. After the RNA extraction, a cDNA synthesis with random primers was performed, using the ImProm-II Reverse Transcription System.
Chapter 3 – Materials and Methods

(Promega). About 140 ng of RNA and 1 µl of random primers were combined in a initial pre-mix with a final volume of 5 µl, heated at 70°C for 5 minutes and chilled on ice for 5 minutes. Subsequently, the reverse transcription mix was prepared in a total reaction volume of 15 µl, containing 4 µl of ImProm-II Buffer 1X, 3 mM MgCl₂, 0.5 mM of each dNTP, 0.5 units of Recombinant RNasin Ribonuclease Inhibitor and 1 µl of ImProm-II Reverse Transcriptase. Reverse transcription was carried out as follow: 25°C for 5 minutes, 40°C for 1 hour and 70°C for 15 minutes.

3.7.3 Amplification of cDNA and transcript analysis

The amplification of the cDNA obtained from the above described reverse transcription was performed in a total reaction volume of 15 µl, containing 1X PCR Buffer, 2.5 mM MgCl₂, 0.2 mM of each dNTP, 0.5 µM of each primer, 0.75 units of @Taq (EuroClone) and 2 µl of non-diluted cDNA. Amplification was carried out as follow: 32 PCR cycles, including a denaturation step at 94°C for 30 seconds, an annealing step at 58°C for 30 seconds and an extension step at 72°C for 45 seconds, followed by a final extension step at 72°C for 5 minutes. Primers for amplification are reported in Table 3.6.

The amplification was checked by electrophoresis on 1.5% agarose gels. The transcripts found with this amplification were analyzed by direct sequencing, as described in paragraph 3.4.2.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer Sequence</th>
<th>PCR Fragment (bp)</th>
<th>Annealing temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FW</td>
<td>GCTGCTTTTTCTTTCTGTC</td>
<td>223</td>
<td>58</td>
</tr>
<tr>
<td>RV</td>
<td>GGTGAGAGATCCTGCTGAGAC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.8 Statistical analyses

Statistical analyses were performed by the Unit of Medical Statistics and Biometry of INT by Dr. Paolo Verderio and his collaborators. Comparison of frequencies in cases and controls was performed by resorting to a logistic regression model both in univariate and multivariate fashion (Hosmer and Lemeshow, 1989). In this model, fitted by method of maximum likelihood, each regression coefficients is the logarithmic of the OR. Under the null hypothesis (absence of association between cancer and alleles), OR is expected to be 1.00. A final parsimonious model was obtained using appropriate selection procedures.
4.1 *PALB2/FANCN* mutation analysis

4.1.1 Mutation screening in a series of Italian BRCAX cases

Mutation screening of all coding exons and flanking intronic sequences of the *PALB2* gene was performed in a series of 575 Italian BRCAX cases collected in cancer centers in Milan. In addition, gene fragments where a deleterious mutation was found were also tested in 784 controls recruited in Milan. In this analysis, a total of 34 variants were found. Eight of these were previously reported as common (allelic frequency > 1%) with comparable distribution in cases and controls and, therefore, considered as neutral polymorphisms (Table 4.1) (Rahman et al., 2007; Erkko et al., 2007; Garcia et al., 2008).

Table 4.1. Frequencies of common *PALB2* polymorphisms in 575 BRCAX cases and 784 controls

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Protein change</th>
<th>Annotation status</th>
<th>Cases (%)</th>
<th>Controls (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>c.-47G&gt;A</td>
<td>na</td>
<td>rs8053188</td>
<td>554 (96.3)</td>
<td>744 (94.9)</td>
</tr>
<tr>
<td>c.212-58A&gt;C</td>
<td>na</td>
<td>none*</td>
<td>519 (90.3)</td>
<td>776 (99.0)</td>
</tr>
<tr>
<td>c.1010T&gt;C</td>
<td>p.Leu337Scr</td>
<td>rs45494092</td>
<td>568 (98.8)</td>
<td>776 (99.0)</td>
</tr>
<tr>
<td>c.1676A&gt;G</td>
<td>p.Gln559Arg</td>
<td>rs45494092</td>
<td>449 (78.1)</td>
<td>613 (78.2)</td>
</tr>
<tr>
<td>c.2014G&gt;C</td>
<td>p.Glu672Gln</td>
<td>rs45532440</td>
<td>528 (91.8)</td>
<td>699 (89.2)</td>
</tr>
<tr>
<td>c.2794G&gt;A</td>
<td>p.Val932Met</td>
<td>rs45624036</td>
<td>563 (98.0)</td>
<td>773 (98.6)</td>
</tr>
<tr>
<td>c.2993G&gt;A</td>
<td>p.Gly998Glu</td>
<td>rs45551636</td>
<td>527 (91.7)</td>
<td>773 (98.6)</td>
</tr>
<tr>
<td>c.3300T&gt;G</td>
<td>p.Thr1100Thr</td>
<td>rs45516100</td>
<td>522 (90.8)</td>
<td>773 (98.6)</td>
</tr>
</tbody>
</table>

*na not applicable, nor normal, het heterozygote, hom homozygote, nd not done*

*reported in Erkko et al., 2007; Garcia et al., 2008; Tischkowitz et al., 2009; Sluiter et al., 2009; Dansonka-Mieszkowska et al., 2010; Silvestri et al., 2010; Catucci et al., 2012.*
In addition, we detected 26 rare or unique variants. Of these, eight, none of which previously reported, introduced a premature protein termination codon (Table 4.2). Two of these mutations, c.72delG (p.Arg26fs) and c.1027C>T (p.Gln343X), were recurrent being detected in two and three cases, respectively. None of the truncating mutations were found in controls.

<table>
<thead>
<tr>
<th>Exon</th>
<th>Mutation</th>
<th>Protein change</th>
<th>Class</th>
<th>Families with mutation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>c.72delG</td>
<td>p.Arg26fs</td>
<td>frameshift</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>c.1027C&gt;T</td>
<td>p.Gln343X</td>
<td>nonsense</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>c.1037_1041delAAGAA</td>
<td>p.Leu346fs</td>
<td>frameshift</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>p.Gln370X</td>
<td>p.Gln370X</td>
<td>nonsense</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>c.2074C&gt;T</td>
<td>p.Gln692X</td>
<td>nonsense</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>c.2167_2168delAT</td>
<td>p.Met723fs</td>
<td>frameshift</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>c.3497delG</td>
<td>p.Gly1166fs</td>
<td>frameshift</td>
<td>1</td>
</tr>
</tbody>
</table>

Of the remaining 18 rare or unique variants (11 previously reported and seven novel) seven were synonymous, seven were missense and four were intronic (Table 4.3). *In silico* analyses were performed to identify potential splicing mutations, using the bioinformatics tools described in paragraph 3.6.2. The c.48G>A (p.Lys16Lys) synonymous mutation, located at the last base of exon one, was predicted to affect the canonical mRNA splicing by causing the loss of the physiological donor splice site by all five bioinformatics tools that correctly detected the natural site (Figure 4.1). To verify this prediction, we performed a reverse transcriptase (RT)-PCR analysis in which we amplified a cDNA fragment spanning exons 1/2 junction from a lymphoblastoid cell line (LCL) of the c.48G>A mutation carrier. An aberrant transcript not present in control cDNA was observed (Figure
Sequencing of this transcript showed the loss of 17 nucleotides at the 3’ end of the exon 1 (Figure 4.2B). These results indicate that the c.48G>A mutation abolishes the canonical donor splice site and activates an alternative site within exon 1, causing the loss of the reading frame and the formation of a premature termination codon at the amino acid residue 36. In addition, we observed that this transcript was absent in LCLs from seven individuals affected with breast cancer, but negative for the investigated mutation (data not shown). Furthermore, two additional synonymous mutations, c.2379C>T (p.Gly793Gly) and c.2418G>T (p.Pro806Pro), were predicted to affect mRNA splicing. In particular, these mutations were predicted to cause the activation of a cryptic donor and a cryptic acceptor splice site, respectively by five of the seven bioinformatics tools considered. No LCLs of carriers of either mutations were available to verify these outputs. All the remaining variants were predicted not to affect normal mRNA splicing by the majority of bioinformatics analyses.

*In silico* analyses were also performed to investigate all of the missense mutations found, using three different software packages (Polyphen2, SIFT and SNPs&GO) predicting the effect of mutations on the protein structure and functioning. In this analysis, two mutations, c.2792T>G (p.Leu931Arg) and c.2816G>T (p.Leu939Trp), were predicted to be damaging by all these tools.
Table 4.3. Frequencies of rare or unique non-truncating \textit{PALB2} variants in cases and controls

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Protein change</th>
<th>Class</th>
<th>Annotation</th>
<th>Cases (n=575)</th>
<th>Controls (n=784)</th>
<th>Found previous studies</th>
<th>Polyphen2/SIFT/ SNPs&amp;GO predictions*</th>
</tr>
</thead>
<tbody>
<tr>
<td>c.13C&gt;T</td>
<td>p.Pro5Ser</td>
<td>missense</td>
<td>none</td>
<td>2</td>
<td>0</td>
<td>Yes\textsuperscript{a}</td>
<td>B/T/N</td>
</tr>
<tr>
<td>c.48G&gt;A</td>
<td>p.Lys16Lys</td>
<td>synonymous</td>
<td>none</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>na</td>
</tr>
<tr>
<td>c.243G&gt;A</td>
<td>p.Lys81Lys</td>
<td>synonymous</td>
<td>none</td>
<td>2</td>
<td>nd</td>
<td>No</td>
<td>na</td>
</tr>
<tr>
<td>c.292A&gt;G</td>
<td>p.Leu98Val</td>
<td>missense</td>
<td>none</td>
<td>1</td>
<td>nd</td>
<td>No</td>
<td>B/T/N</td>
</tr>
<tr>
<td>c.1001A&gt;G</td>
<td>p.Tyr334Cys</td>
<td>missense</td>
<td>rs200620434</td>
<td>1</td>
<td>1</td>
<td>Yes\textsuperscript{b-e}</td>
<td>B/T/N</td>
</tr>
<tr>
<td>c.1194G&gt;A</td>
<td>p.Val398Val</td>
<td>synonymous</td>
<td>rs61755173</td>
<td>1</td>
<td>0</td>
<td>Yes\textsuperscript{b,e,f,i}</td>
<td>na</td>
</tr>
<tr>
<td>c.1572A&gt;G</td>
<td>p.Ser524Ser</td>
<td>synonymous</td>
<td>rs45472400</td>
<td>3</td>
<td>3</td>
<td>Yes\textsuperscript{b,e,k}</td>
<td>na</td>
</tr>
<tr>
<td>c.1684+42_1684+43insTGA</td>
<td>na</td>
<td>intronic</td>
<td>none</td>
<td>4</td>
<td>7</td>
<td>Yes\textsuperscript{e,h}</td>
<td>na</td>
</tr>
<tr>
<td>c.2091C&gt;G</td>
<td>p.Gly697Gly</td>
<td>synonymous</td>
<td>none</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>na</td>
</tr>
<tr>
<td>c.2379C&gt;T</td>
<td>p.Gly793Gly</td>
<td>synonymous</td>
<td>none</td>
<td>1</td>
<td>nd</td>
<td>No</td>
<td>na</td>
</tr>
<tr>
<td>c.2418G&gt;T</td>
<td>p.Pro806Pro</td>
<td>synonymous</td>
<td>none</td>
<td>1</td>
<td>nd</td>
<td>Yes\textsuperscript{f}</td>
<td>na</td>
</tr>
<tr>
<td>c.2587-38G&gt;C</td>
<td>na</td>
<td>intronic</td>
<td>rs180177119</td>
<td>5</td>
<td>nd</td>
<td>No</td>
<td>na</td>
</tr>
<tr>
<td>c.2587-25A&gt;G</td>
<td>na</td>
<td>intronic</td>
<td>none</td>
<td>1</td>
<td>nd</td>
<td>No</td>
<td>na</td>
</tr>
<tr>
<td>c.2590C&gt;T</td>
<td>p.Pro864Ser</td>
<td>missense</td>
<td>rs45568339</td>
<td>5</td>
<td>nd</td>
<td>Yes\textsuperscript{d,f,i,k-o}</td>
<td>PrD/T/N</td>
</tr>
<tr>
<td>c.2792T&gt;G</td>
<td>p.Leu931Arg</td>
<td>missense</td>
<td>none</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>PoD/APF/D</td>
</tr>
<tr>
<td>c.2816T&gt;G</td>
<td>p.Leu939Trp</td>
<td>missense</td>
<td>rs45478192</td>
<td>2</td>
<td>2</td>
<td>Yes\textsuperscript{d,f,h,i,m}</td>
<td>PoD/APF/D</td>
</tr>
<tr>
<td>c.2996+17T&gt;C</td>
<td>na</td>
<td>intronic</td>
<td>rs180177128</td>
<td>2</td>
<td>nd</td>
<td>Yes\textsuperscript{f}</td>
<td>na</td>
</tr>
<tr>
<td>c.3428T&gt;A</td>
<td>p.Leu1143Ile</td>
<td>missense</td>
<td>rs62625284</td>
<td>2</td>
<td>1</td>
<td>Yes\textsuperscript{b,f,e,a}</td>
<td>PoD/APF/N</td>
</tr>
</tbody>
</table>

*\textsuperscript{a}benign, \textsuperscript{PrD} probably damaging, \textsuperscript{PoD} possibly damaging, \textsuperscript{U} unclassified (PolyPhen-2); \textsuperscript{T} tolerated, \textsuperscript{APF} affecting protein function (SIFT); \textsuperscript{N} neutral, \textsuperscript{D} disease, \textsuperscript{U} unclassified (SNP&GO); \textsuperscript{na} not available; \textsuperscript{nd} not done. \textsuperscript{Casadei et al., 2011; Hellebrand et al., 2011; Balia et al., 2010; Tischkowitz et al., 2012; Catucci et al., 2012; Rahman et al., 2007; Garcia et al., 2008; Tischkowitz et al., 2009; Blanco et al., 2011; Hofstatter et al., 2011; Bogdanova et al., 2010; Papi et al., 2009; Sauty de Chalon et al., 2009; Zheng et al., 2012; Teo et al., 2013. The variants predicted by \textit{in silico} analyses to be deleterious, including the c.48G>A verified \textit{in vitro}, are shown in bold.

In summary, we found nine \textit{PALB2} mutations classifiable as pathogenic, including eight truncating and one splicing variant (Figure 4.3) and 12 carriers of these mutations, for a frequency of 2.1%. None of these mutations were detected in 784 tested controls. Families in which the index case carried a \textit{PALB2} truncating mutation are showed in Table 4.4 and in Figure 4.4.
Figure 4.1. Bioinformatics analyses of the \textit{PALB2} c.48G>A splicing mutation. \textit{In silico} analyses were performed using the indicated tools. In the A, B, D and E panels, the canonical splicing site (indicated by the red arrow) is detected only in the wild-type sequence (wt) and not in the mutated sequence (mut). In the C panel, there is a substantial decrease of the site recognition score in the mutated compared to the wild type sequence.
Chapter 4 - Results

Figure 4.2. Characterization of the mRNA transcripts caused by the PALB2 c.48G>A mutation. (A) RT-PCR results: M, molecular marker (ΦX-174 HaeIII); lane 1, no template; lane 2, genomic DNA used as negative control of the RT-PCR; lane 3, cDNA from a PALB2 wild-type LCL used as a normal control of the RT-PCR lane 4, cDNA from the LCL carrying the heterozygous PALB2 c.48G>A mutation. The sizes of the full-length (FL) and aberrant (Ex1_17bp del) transcripts are indicated. The additional band due to the improper annealing of the two transcripts is indicated by the asterisk. (B) Sequencing results: the mutated cDNA (upper panel) shows two overlapping sequences, one corresponding to the full-length transcript (the unique fragment presents in the normal control cDNA, lower panel) and one to the aberrant transcript due to the skipping of 17nt (GTGAGGAGAAGGAAAAG) at the 3'-end of exon 1.

 Relatives of the index case were available for mutation testing in four of the families with a PALB2 pathogenic mutation (family D, F, H and K; Figure 4.4). In particular, we tested a paternal aunt affected with breast cancer at age 45, in family D; a sister affected with bilateral breast cancer at age 39 and 44, in family F; a sister and a maternal cousin affected with breast cancer at age 41 and 30, respectively, in family H; all of these individuals carried the same mutation found in the index case. In addition, three nieces of the proband were available for testing in family K, of whom only two were affected with breast cancer. None of them carried the PALB2 mutation, but they were carriers of the BRCA1 2335_2336delAA mutation, inherited from the other family branch.
Table 4.4. Characteristics of families with PALB2 truncating mutations

<table>
<thead>
<tr>
<th>Family</th>
<th>Mutation</th>
<th>Index case status</th>
<th>Relatives affected with breast cancer*</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>c.1108C&gt;T</td>
<td>Br, 36</td>
<td>none</td>
</tr>
<tr>
<td>B</td>
<td>c.1037_1041delAAGAA</td>
<td>Br, 47</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>c.72delG</td>
<td>Br, 56</td>
<td>3</td>
</tr>
<tr>
<td>D</td>
<td>c.2167_2168delAT</td>
<td>Br, 26</td>
<td>2</td>
</tr>
<tr>
<td>E</td>
<td>c.1027C&gt;T</td>
<td>Br, 37</td>
<td>2</td>
</tr>
<tr>
<td>F</td>
<td>c.3497delG</td>
<td>Br, 39</td>
<td>4</td>
</tr>
<tr>
<td>G</td>
<td>c.1027C&gt;T</td>
<td>Br bil, 33</td>
<td>4</td>
</tr>
<tr>
<td>H</td>
<td>c.1027C&gt;T</td>
<td>Br bil 42</td>
<td>5</td>
</tr>
<tr>
<td>I</td>
<td>c.72delG</td>
<td>Br, 46</td>
<td>2</td>
</tr>
<tr>
<td>J</td>
<td>c.2074C&gt;T</td>
<td>Br, 31</td>
<td>none</td>
</tr>
<tr>
<td>K</td>
<td>c.2787_2788delTA</td>
<td>Br bil, 41</td>
<td>10</td>
</tr>
<tr>
<td>L</td>
<td>c.48G&gt;A</td>
<td>Br, 41</td>
<td>6</td>
</tr>
</tbody>
</table>

*index case excluded

Br breast cancer, Br bil bilateral breast cancer
Figure 4.3. Electropherograms of the sequences where a *PALB2* pathogenic mutation was found. For each mutation, the mutated (*mut*) and the wild-type (*wt*) sequence are showed.
4.1.2 Investigation of the c.72delG and the c.1027C>T recurrent mutations

In the *PALB2* screening, the c.72delG and the c.1027C>T recurrent mutations were identified. To better investigate the frequency of these two variants, we further genotyped an additional 332 BRCAX cases and 176 controls recruited in Milan, using a TaqMan SNP Genotyping Assay. No other carriers of these two mutations were found. Interestingly, we also observed that, among the three carriers of the c.1027C>T mutation, two individuals were from families that self-reported as being originally from the province of Bergamo, in Northern Italy. To explore the possibility that the c.1027C>T is more frequent in this area, we genotyped 112 BRCAX cases recruited at the Ospedali Riuniti of Bergamo and 477 controls recruited through the AVIS of Bergamo. Interestingly, we found that 5/112 cases and 2/477 controls carried the c.1027C>T mutation, for a frequency of 4.5% and 0.4%, respectively (Table 4.5). Pedigrees of BRCAX families collected in Bergamo in which the index case carried the c.1027C>T mutation are showed in Figure 4.5. Only in family O, three relatives of the proband were available for mutation testing: the father, affected with breast cancer at 86, was a non-carrier, while two unaffected sisters carried the c.1027C>T mutation.

<table>
<thead>
<tr>
<th>Group</th>
<th>Carriers/total samples</th>
<th>Carriers %</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRCAX from Milan</td>
<td>3/907</td>
<td>0.3</td>
</tr>
<tr>
<td>Controls from Milan</td>
<td>0/960</td>
<td>0.0</td>
</tr>
<tr>
<td>BRCAX from Bergamo</td>
<td>5/112</td>
<td>4.5</td>
</tr>
<tr>
<td>Controls from Bergamo</td>
<td>2/477</td>
<td>0.4</td>
</tr>
</tbody>
</table>
To verify the statistical power of the study, the null hypothesis of an equal mutation rate in cases and controls was assessed by resorting to Fisher's exact test with a Type I error probability of 0.05. This made possible to reject the null hypothesis with a probability (power) of 0.824.

### 4.1.3 *PALB2* mutation analysis in breast and pancreatic cancer families

Although the families included in the screening for *PALB2* mutations were selected based on family history of breast cancer, several other cancers types were diagnosed in the relatives of the proband, including pancreatic cancers. Following the analysis of the pedigrees of the 575 screened families, we identified 39 families with both breast and pancreatic cancer cases, including three in which the index case carried a *PALB2* truncating mutation (family C, F and G; Figure 4.4), for a frequency of 7.7% (3/39), higher than that observed in the overall group. This observation suggested that breast cancer families with cases of pancreatic cancer could be enriched in *PALB2* mutations. To verify this hypothesis, we screened for *PALB2* mutations an additional 23 BRCAX index cases, selected from families in which at least one case of pancreatic cancer was reported in first- or second-degree relatives, independently of the breast cancer family branch. No other truncating mutation was found, for an overall frequency in the two combined groups of breast and pancreatic cancer families of 4.8% (3/62; Figure 4.6).
Chapter 4 - Results

A

B

C

Breast cancer
Bilateral breast cancer
Pancreatic cancer
Others cancers
Figure 4.4. Pedigrees of the 12 families in which the index case carried a PALB2 truncating mutation. Index cases are indicated by arrow and PALB2 mutations are described. Cancer type and age at diagnosis are reported, when known. Br breast cancer, Br bil bilateral breast cancer, Ch choledoch cancer, CR colorectal cancer, Ep epidermoid cancer, GC granulosa cell cancer, Ki kidney cancer, La larynx cancer, Le leukemia, LH Hodgkin's lymphoma, Li liver cancer, Lu lung cancer, Ov ovarian cancer, Os osteosarcoma, Pa pancreatic cancer, Pr prostate cancer, St stomach cancer, Ut uterine cancer, VC vocal cords cancer.
Chapter 4 - Results

Breast cancer
Bilateral breast cancer
Pancreatic cancer
Others cancers

M

N

O
Figure 4.5. Pedigrees of the five families recruited at the Ospedali Riuniti of Bergamo, in which the index case carried the c.1027C>T PALB2 mutation. Index cases are indicated by arrow. Cancer type and age at diagnosis are reported when known. Br breast cancer, Br bil bilateral breast cancer, La larynx cancer, LH Hodgkin's lymphoma, Li liver cancer, Lu lung cancer, St stomach cancer, In intestinal cancer, Pa pancreatic cancer, Pr prostate cancer.

Figure 4.6. Venn diagram representing the distribution of PALB2 truncating mutations in breast and breast/pancreatic cancer families recruited in this study.
4.2 Sequencing analysis of SLX4/FANCP gene

A series of 526 BRCAX cases were screened for mutations in SLX4/FANCP gene by sequencing of coding exons and flanking intronic sequences. In this analysis, a total of 81 different variants were detected, but no one could be considered as a clear pathogenic mutation (Table 4.6). Of these variants, 35 were previously annotated in public databases including dbSNP (Sherry et al., 2001), 1000 Genomes (Altshuler et al., 2010) and Exome Variant Server (http://evs.gs.washington.edu/EVS/) with a carrier frequency ≥ 1% in Caucasians and thus considered as likely neutral polymorphisms. We used the above mentioned frequency as threshold value because it was estimated that the overall allelic frequency of variants of each high/moderate-penetrance gene does not exceed 0.5% (reviewed in Mavaddat et al., 2010). Among the 46 remaining variants, 29 were missense, 14 were silent, two were intronic and one was a 3-nucleotide in-frame deletion causing the loss of a single conserved amino acid (p.Ile1195del). All 46 variants were analyzed in silico to investigate their potential impact on mRNA splicing, using the following four programs: Berkeley Drosophila Genome Project (BDGP), NetGene2, SplicePredictor and GeneSplicer (Table 4.7). In this analysis, three programs predicted the c.833G>A to create a new donor splice site, whereas two of them predicted the c.5155T>A to abolish the physiological acceptor site of exon 15. Both predictions were confirmed using an additional software (MaxEntScan). The 29 missense mutations were also analyzed using three different protein prediction programs: PolyPhen-2, SIFT and SNP&GO (Table 4.7). Of all the tested variants, the c.5155T>A (p.Ser1719Tyr) was classified as a variant with pathogenic effect by all three programs.
To evaluate the statistical power of this study, we estimated that in our sample size of 526 BRCAX cases, we had a probability of detecting \( SLX4 \) truncating mutations of 0.6%, that corresponds to the upper limit of the 95% confidence interval of the event probability in which no events (no carriers of truncating \( SLX4 \) mutations) have been observed.
Table 4.6. Frequencies of SLX4 variants in BRCAX cases

<table>
<thead>
<tr>
<th>Nucleotide change</th>
<th>Amino acid change</th>
<th>Mutation type/considered as a neutral polymorphism</th>
<th>Annotation status</th>
<th>Number of Nor/Het/Hom genotypes (allelic frequency)</th>
</tr>
</thead>
<tbody>
<tr>
<td>c.60G&gt;A</td>
<td>p.Leu20Leu</td>
<td>silent/no</td>
<td>not annotated</td>
<td>525/1/0 (0.001)</td>
</tr>
<tr>
<td>c.90C&gt;T</td>
<td>p.Ser30Ser</td>
<td>silent/yes</td>
<td>rs118089506</td>
<td>511/15/0 (0.014)</td>
</tr>
<tr>
<td>c.244A&gt;G</td>
<td>p.Asn82Asp</td>
<td>missense/no</td>
<td>not annotated</td>
<td>525/1/0 (0.001)</td>
</tr>
<tr>
<td>c.247G&gt;A</td>
<td>p.Gly83Ser</td>
<td>missense/no</td>
<td>not annotated</td>
<td>525/1/0 (0.001)</td>
</tr>
<tr>
<td>c.299C&gt;A</td>
<td>p.Thr100Asn</td>
<td>missense/no</td>
<td>not annotated</td>
<td>525/1/0 (0.001)</td>
</tr>
<tr>
<td>c.421G&gt;T</td>
<td>p.Gly141Trp</td>
<td>missense/no</td>
<td>not annotated</td>
<td>524/2/0 (0.002)</td>
</tr>
<tr>
<td>c.452C&gt;T</td>
<td>p.Pro151Leu</td>
<td>missense/no</td>
<td>not annotated</td>
<td>525/1/0 (0.001)</td>
</tr>
<tr>
<td>c.553G&gt;A</td>
<td>p.Asp185Asn</td>
<td>missense/no</td>
<td>not annotated</td>
<td>525/1/0 (0.001)</td>
</tr>
<tr>
<td>c.555C&gt;T</td>
<td>p.Asp185Asp</td>
<td>silent/yes</td>
<td>rs74640850</td>
<td>484/42/0 (0.040)</td>
</tr>
<tr>
<td>c.590T&gt;C</td>
<td>p.Val197Ala</td>
<td>missense/no</td>
<td>not annotated</td>
<td>523/3/0 (0.003)</td>
</tr>
<tr>
<td>c.610C&gt;T</td>
<td>p.Arg204Cys</td>
<td>missense/yes</td>
<td>rs79842542</td>
<td>484/42/0 (0.040)</td>
</tr>
<tr>
<td>c.678C&gt;T</td>
<td>p.His226His</td>
<td>silent/yes</td>
<td>rs28516641</td>
<td>498/27/1 (0.028)</td>
</tr>
<tr>
<td>c.707C&gt;T</td>
<td>p.Ala236Val</td>
<td>missense/no</td>
<td>not annotated</td>
<td>525/1/0 (0.001)</td>
</tr>
<tr>
<td>c.710G&gt;A</td>
<td>p.Arg237Gln</td>
<td>missense/no</td>
<td>not annotated</td>
<td>521/5/0 (0.005)</td>
</tr>
<tr>
<td>c.734C&gt;T</td>
<td>p.Pro245Leu</td>
<td>missense/no</td>
<td>not annotated</td>
<td>524/2/0 (0.002)</td>
</tr>
<tr>
<td>c.742G&gt;A</td>
<td>p.Glu248Lys</td>
<td>missense/no</td>
<td>not annotated</td>
<td>525/1/0 (0.001)</td>
</tr>
<tr>
<td>c.753G&gt;A</td>
<td>p.Ala251Ala</td>
<td>silent/yes</td>
<td>rs8061528</td>
<td>329/172/25 (0.211)</td>
</tr>
<tr>
<td>c.761-32T&gt;G</td>
<td>none</td>
<td>intronic/no</td>
<td>rs118098382</td>
<td>514/12/0 (0.011)</td>
</tr>
<tr>
<td>c.833G&gt;A</td>
<td>p.Arg278Gln</td>
<td>missense/no</td>
<td>not annotated</td>
<td>525/1/0 (0.001)</td>
</tr>
<tr>
<td>c.999C&gt;T</td>
<td>p.Ile333Ile</td>
<td>silent/yes</td>
<td>rs7198338</td>
<td>525/1/0 (0.001)</td>
</tr>
<tr>
<td>c.1065G&gt;A</td>
<td>p.Gln355Gln</td>
<td>silent/no</td>
<td>not annotated</td>
<td>525/1/0 (0.001)</td>
</tr>
<tr>
<td>c.1152A&gt;G</td>
<td>p.Pro384Pro</td>
<td>silent/yes</td>
<td>rs112511042</td>
<td>484/41/1 (0.041)</td>
</tr>
<tr>
<td>c.1153C&gt;A</td>
<td>p.Pro385Thr</td>
<td>missense/yes</td>
<td>rs115694169</td>
<td>520/6/0 (0.006)</td>
</tr>
<tr>
<td>c.1156A&gt;G</td>
<td>p.Met386Val</td>
<td>missense/yes</td>
<td>rs113490934</td>
<td>484/41/1 (0.041)</td>
</tr>
<tr>
<td>c.1163+10C&gt;T</td>
<td>none</td>
<td>intronic/yes</td>
<td>rs80116508</td>
<td>484/41/1 (0.041)</td>
</tr>
<tr>
<td>c.1164-75C&gt;G</td>
<td>none</td>
<td>intronic/yes</td>
<td>rs59622164</td>
<td>484/42/0 (0.040)</td>
</tr>
<tr>
<td>c.1164-66T&gt;A</td>
<td>none</td>
<td>intronic/no</td>
<td>not annotated</td>
<td>524/2/0 (0.002)</td>
</tr>
<tr>
<td>c.1366+11T&gt;C</td>
<td>none</td>
<td>intronic/yes</td>
<td>rs76350200</td>
<td>477/48/1 (0.048)</td>
</tr>
<tr>
<td>c.1371T&gt;G</td>
<td>p.Asn457Lys</td>
<td>missense/yes</td>
<td>rs74319927</td>
<td>488/38/0 (0.036)</td>
</tr>
<tr>
<td>c.1641G&gt;A</td>
<td>p.Thr547Thr</td>
<td>silent/no</td>
<td>not annotated</td>
<td>525/1/0 (0.001)</td>
</tr>
<tr>
<td>c.1755C&gt;T</td>
<td>p.Pro585Pro</td>
<td>silent/yes</td>
<td>rs114016359</td>
<td>520/6/0 (0.006)</td>
</tr>
<tr>
<td>c.1755C&gt;A</td>
<td>p.Pro585Pro</td>
<td>silent/no</td>
<td>not annotated</td>
<td>525/1/0 (0.001)</td>
</tr>
<tr>
<td>c.1803G&gt;A</td>
<td>p.Ser601Ser</td>
<td>silent/no</td>
<td>not annotated</td>
<td>520/6/0 (0.006)</td>
</tr>
<tr>
<td>c.1832C&gt;A</td>
<td>p.Ala611Asp</td>
<td>missense/no</td>
<td>not annotated</td>
<td>525/1/0 (0.001)</td>
</tr>
<tr>
<td>c.1846G&gt;A</td>
<td>p.Val616Met</td>
<td>missense/no</td>
<td>not annotated</td>
<td>525/1/0 (0.001)</td>
</tr>
<tr>
<td>c.1896G&gt;C</td>
<td>p.Gly632Gly</td>
<td>silent/no</td>
<td>not annotated</td>
<td>525/1/0 (0.001)</td>
</tr>
<tr>
<td>c.1898G&gt;A</td>
<td>p.Gly633Asp</td>
<td>missense/yes</td>
<td>rs1056085</td>
<td>525/1/0 (0.001)</td>
</tr>
<tr>
<td>c.1911G&gt;A</td>
<td>p.Ser637Ser</td>
<td>silent/no</td>
<td>not annotated</td>
<td>525/1/0 (0.001)</td>
</tr>
<tr>
<td>c.2006G&gt;A</td>
<td>p.Arg669Asp</td>
<td>missense/no</td>
<td>not annotated</td>
<td>525/1/0 (0.001)</td>
</tr>
<tr>
<td>c.2012T&gt;C</td>
<td>p.Leu671Ser</td>
<td>missense/yes</td>
<td>rs77985244</td>
<td>482/42/0 (0.042)</td>
</tr>
<tr>
<td>c.2013+23G&gt;A</td>
<td>none</td>
<td>intronic/yes</td>
<td>rs112226642</td>
<td>483/43/0 (0.041)</td>
</tr>
<tr>
<td>SNP</td>
<td>Coding/Splicing</td>
<td>Genotype</td>
<td>Minor allele</td>
<td>Minor allele frequency</td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------------</td>
<td>----------</td>
<td>--------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>c.2160+50C&gt;T</td>
<td>none/intronic</td>
<td>rs75762935</td>
<td>484/42/0</td>
<td>0.040</td>
</tr>
<tr>
<td>c.2235C&gt;T</td>
<td>p.Thr745Thr</td>
<td>silent/no</td>
<td>rs75184268</td>
<td>524/2/0</td>
</tr>
<tr>
<td>c.2359G&gt;A</td>
<td>p.Glu787Lys</td>
<td>missense/no</td>
<td>not annotated</td>
<td>518/8/0</td>
</tr>
<tr>
<td>c.2597A&gt;C</td>
<td>p.Gln866Pro</td>
<td>missense/no</td>
<td>not annotated</td>
<td>525/1/0</td>
</tr>
<tr>
<td>c.2854G&gt;A</td>
<td>p.Ala952Thr</td>
<td>missense/yes</td>
<td>rs59939128</td>
<td>482/44/0</td>
</tr>
<tr>
<td>c.2855C&gt;T</td>
<td>p.Ala952Val</td>
<td>missense/yes</td>
<td>rs78637028</td>
<td>484/42/0</td>
</tr>
<tr>
<td>c.2924C&gt;T</td>
<td>p.Pro975Leu</td>
<td>missense/no</td>
<td>rs114472821</td>
<td>519/7/0</td>
</tr>
<tr>
<td>c.2975G&gt;A</td>
<td>p.Gly992Glu</td>
<td>missense/yes</td>
<td>rs58735123</td>
<td>525/1/0</td>
</tr>
<tr>
<td>c.3062G&gt;A</td>
<td>p.Arg1021His</td>
<td>missense/no</td>
<td>not annotated</td>
<td>525/1/0</td>
</tr>
<tr>
<td>c.3109G&gt;C</td>
<td>p.Leu1037Leu</td>
<td>silent/no</td>
<td>rs714181</td>
<td>462/63/1</td>
</tr>
<tr>
<td>c.3162G&gt;A</td>
<td>p.Ser1054Ser</td>
<td>silent/yes</td>
<td>rs76488917</td>
<td>500/25/1</td>
</tr>
<tr>
<td>c.3189C&gt;T</td>
<td>p.Gly1063Gly</td>
<td>silent/no</td>
<td>rs77699867</td>
<td>520/6/0</td>
</tr>
<tr>
<td>c.3308G&gt;A</td>
<td>p.Arg1103His</td>
<td>missense/no</td>
<td>not annotated</td>
<td>525/1/0</td>
</tr>
<tr>
<td>c.3316G&gt;A</td>
<td>p.Val1106Met</td>
<td>missense/no</td>
<td>not annotated</td>
<td>525/1/0</td>
</tr>
<tr>
<td>c.3365C&gt;T</td>
<td>p.Pro1122Leu</td>
<td>missense/yes</td>
<td>rs714181</td>
<td>462/63/1</td>
</tr>
<tr>
<td>c.3420A&gt;G</td>
<td>p.Lys1140Lys</td>
<td>silent/no</td>
<td>rs525/1/0</td>
<td>0.001</td>
</tr>
<tr>
<td>c.3583_3585delATT</td>
<td>p.Ile1195del</td>
<td>in frame del/no</td>
<td>not annotated</td>
<td>525/1/0</td>
</tr>
<tr>
<td>c.3662C&gt;T</td>
<td>p.Ala1221Val</td>
<td>missense/yes</td>
<td>rs3827530</td>
<td>491/35/0</td>
</tr>
<tr>
<td>c.3783G&gt;A</td>
<td>p.Pro1261Pro</td>
<td>silent/yes</td>
<td>rs77699867</td>
<td>520/6/0</td>
</tr>
<tr>
<td>c.3812C&gt;T</td>
<td>p.Ser1271Phe</td>
<td>missense/yes</td>
<td>rs3810813</td>
<td>474/52/0</td>
</tr>
<tr>
<td>c.3849C&gt;G</td>
<td>p.Ala1283Ala</td>
<td>silent/no</td>
<td>rs525/1/0</td>
<td>0.001</td>
</tr>
<tr>
<td>c.3963G&gt;A</td>
<td>p.Pro1321Pro</td>
<td>silent/yes</td>
<td>rs116781836</td>
<td>516/10/0</td>
</tr>
<tr>
<td>c.4068G&gt;A</td>
<td>p.Pro1356Pro</td>
<td>silent/no</td>
<td>rs115491049</td>
<td>525/1/0</td>
</tr>
<tr>
<td>c.4338C&gt;T</td>
<td>p.Thr1446Ser</td>
<td>missense/no</td>
<td>rs77718962</td>
<td>525/1/0</td>
</tr>
<tr>
<td>c.4500T&gt;C</td>
<td>p.Asn1500Asn</td>
<td>silent/yes</td>
<td>rs3810812</td>
<td>135/246/145</td>
</tr>
<tr>
<td>c.4563T&gt;C</td>
<td>p.Pro1521Pro</td>
<td>silent/yes</td>
<td>rs525/1/0</td>
<td>0.001</td>
</tr>
<tr>
<td>c.4580C&gt;T</td>
<td>p.Pro1527Leu</td>
<td>missense/no</td>
<td>not annotated</td>
<td>525/1/0</td>
</tr>
<tr>
<td>c.4581G&gt;A</td>
<td>p.Pro1527Pro</td>
<td>silent/yes</td>
<td>rs78635099</td>
<td>517/9/0</td>
</tr>
<tr>
<td>c.4597G&gt;T</td>
<td>p.Ala1533Ser</td>
<td>missense/no</td>
<td>not annotated</td>
<td>525/1/0</td>
</tr>
<tr>
<td>c.4600G&gt;A</td>
<td>p.Gly1534Ser</td>
<td>missense/no</td>
<td>rs78770603</td>
<td>525/1/0</td>
</tr>
<tr>
<td>c.4648C&gt;T</td>
<td>p.Arg1550Trp</td>
<td>missense/yes</td>
<td>rs77021998</td>
<td>525/1/0</td>
</tr>
<tr>
<td>c.4739+24G&gt;T</td>
<td>none/intronic</td>
<td>rs12933120</td>
<td>371/143/12</td>
<td>0.159</td>
</tr>
<tr>
<td>c.4865A&gt;G</td>
<td>p.Gln1622Arg</td>
<td>missense/no</td>
<td>not annotated</td>
<td>525/1/0</td>
</tr>
<tr>
<td>c.5040G&gt;C</td>
<td>p.Arg1680Ser</td>
<td>missense/no</td>
<td>not annotated</td>
<td>525/1/0</td>
</tr>
<tr>
<td>c.5146T&gt;A</td>
<td>p.Ser1716Thr</td>
<td>missense/yes</td>
<td>rs75182789</td>
<td>525/1/0</td>
</tr>
<tr>
<td>c.5154-28C&gt;T</td>
<td>none/intronic</td>
<td>not annotated</td>
<td>525/1/0</td>
<td>0.001</td>
</tr>
<tr>
<td>c.5155T&gt;A</td>
<td>p.Ser1719Tyr</td>
<td>missense/no</td>
<td>not annotated</td>
<td>525/1/0</td>
</tr>
<tr>
<td>c.5183T&gt;G</td>
<td>p.Phe1728Cys</td>
<td>missense/no</td>
<td>not annotated</td>
<td>525/1/0</td>
</tr>
<tr>
<td>c.5501A&gt;G</td>
<td>p.Asn1834Ser</td>
<td>missense/yes</td>
<td>rs111738042</td>
<td>522/4/0</td>
</tr>
<tr>
<td>c.5505+8A&gt;G</td>
<td>none/intronic</td>
<td>rs3751839</td>
<td>476/50/0</td>
<td>0.048</td>
</tr>
</tbody>
</table>

*annotated in dbSNP, 1000 Genomes and Exome Variant Server with carrier frequency ≥ 1% in Caucasians. Nor common homozygotes, Het heterozygotes, Hom rare homozygotes.
Table 4.7. *In silico* analyses of SLX4 variants in BRCAX cases

<table>
<thead>
<tr>
<th>Nucleotide change</th>
<th>Amino acid change</th>
<th>Predicted effect on mRNA transcript*</th>
<th>Protein prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>c.60G&gt;A</td>
<td>p.Leu20Leu</td>
<td>none</td>
<td>Polyphen 2 SIFT SNP&amp; GO</td>
</tr>
<tr>
<td>c.244A&gt;G</td>
<td>p.Asn82Asp</td>
<td>none</td>
<td>B T N</td>
</tr>
<tr>
<td>c.247G&gt;A</td>
<td>p.Gly83Ser</td>
<td>none</td>
<td>B T N</td>
</tr>
<tr>
<td>c.299C&gt;A</td>
<td>p.Thr100Asn</td>
<td>none</td>
<td>B T N</td>
</tr>
<tr>
<td>c.421G&gt;T</td>
<td>p.Gly141Trp</td>
<td>none</td>
<td>PrD APF N</td>
</tr>
<tr>
<td>c.452C&gt;T</td>
<td>p.Pro151Leu</td>
<td>none</td>
<td>PoD APF N</td>
</tr>
<tr>
<td>c.553G&gt;A</td>
<td>p.Asp185Trp</td>
<td>none</td>
<td>B T N</td>
</tr>
<tr>
<td>c.590T&gt;C</td>
<td>p.Val197Leu</td>
<td>none</td>
<td>B T N</td>
</tr>
<tr>
<td>c.707C&gt;T</td>
<td>p.Ala236Val</td>
<td>none</td>
<td>B T N</td>
</tr>
<tr>
<td>c.710G&gt;A</td>
<td>p.Arg237Gln</td>
<td>none</td>
<td>B T N</td>
</tr>
<tr>
<td>c.734C&gt;T</td>
<td>p.Pro245Leu</td>
<td>none</td>
<td>PrD T N</td>
</tr>
<tr>
<td>c.742G&gt;A</td>
<td>p.Glu248Lys</td>
<td>none</td>
<td>B T N</td>
</tr>
<tr>
<td>c.833G&gt;A</td>
<td>p.Arg278Gln</td>
<td>creation of a new donor splice site</td>
<td>not done</td>
</tr>
<tr>
<td>c.1065G&gt;A</td>
<td>p.Gln355Gln</td>
<td>none</td>
<td>not done</td>
</tr>
<tr>
<td>c.1164-66T&gt;A</td>
<td>none</td>
<td>none</td>
<td>not done</td>
</tr>
<tr>
<td>c.1641G&gt;A</td>
<td>p.Thr547Thr</td>
<td>none</td>
<td>not done</td>
</tr>
<tr>
<td>c.1755C&gt;A</td>
<td>p.Pro585Pro</td>
<td>none</td>
<td>not done</td>
</tr>
<tr>
<td>c.1803G&gt;A</td>
<td>p.Ser601Ser</td>
<td>none</td>
<td>not done</td>
</tr>
<tr>
<td>c.1832C&gt;A</td>
<td>p.Ala611Asp</td>
<td>none</td>
<td>PrD T N</td>
</tr>
<tr>
<td>c.1846C&gt;A</td>
<td>p.Val161Met</td>
<td>none</td>
<td>PoD APF N</td>
</tr>
<tr>
<td>c.1896G&gt;C</td>
<td>p.Gly632Gly</td>
<td>none</td>
<td>not done</td>
</tr>
<tr>
<td>c.1911G&gt;A</td>
<td>p.Ser637Ser</td>
<td>none</td>
<td>not done</td>
</tr>
<tr>
<td>c.2006G&gt;A</td>
<td>p.Arg669Asp</td>
<td>none</td>
<td>B T N</td>
</tr>
<tr>
<td>c.2235C&gt;T</td>
<td>p.Thr745Thr</td>
<td>none</td>
<td>not done</td>
</tr>
<tr>
<td>c.2559G&gt;A</td>
<td>p.Glu787Lys</td>
<td>none</td>
<td>not done</td>
</tr>
<tr>
<td>c.2597A&gt;C</td>
<td>p.Gln866Pro</td>
<td>none</td>
<td>B T N</td>
</tr>
<tr>
<td>c.2975G&gt;A</td>
<td>p.Gly992Glu</td>
<td>none</td>
<td>not done</td>
</tr>
<tr>
<td>c.3062G&gt;A</td>
<td>p.Arg1021His</td>
<td>none</td>
<td>B T N</td>
</tr>
<tr>
<td>c.3109T&gt;C</td>
<td>p.Leu1037Leu</td>
<td>none</td>
<td>not done</td>
</tr>
<tr>
<td>c.3189C&gt;T</td>
<td>p.Gly1063Gly</td>
<td>none</td>
<td>not done</td>
</tr>
<tr>
<td>c.3308G&gt;A</td>
<td>p.Arg1103His</td>
<td>none</td>
<td>not done</td>
</tr>
<tr>
<td>c.3316G&gt;A</td>
<td>p.Val1106Met</td>
<td>none</td>
<td>PoD T D</td>
</tr>
<tr>
<td>c.3420G&gt;A</td>
<td>p.Lys1140Lys</td>
<td>none</td>
<td>not done</td>
</tr>
<tr>
<td>c.3583_3585delATT</td>
<td>p.Ile1195del</td>
<td>none</td>
<td>not done</td>
</tr>
<tr>
<td>c.3849C&gt;G</td>
<td>p.Ala1283Ala</td>
<td>none</td>
<td>not done</td>
</tr>
<tr>
<td>c.4068G&gt;A</td>
<td>p.Pro1356Pro</td>
<td>none</td>
<td>not done</td>
</tr>
<tr>
<td>c.4338C&gt;T</td>
<td>p.Thr1446Ser</td>
<td>none</td>
<td>PrD T N</td>
</tr>
<tr>
<td>c.4563T&gt;C</td>
<td>p.Pro1521Pro</td>
<td>none</td>
<td>not done</td>
</tr>
<tr>
<td>c.4580C&gt;T</td>
<td>p.Pro1527Leu</td>
<td>none</td>
<td>B T N</td>
</tr>
<tr>
<td>c.4597G&gt;T</td>
<td>p.Ala1533Ser</td>
<td>none</td>
<td>B APF N</td>
</tr>
<tr>
<td>c.4600G&gt;A</td>
<td>p.Gly1534Ser</td>
<td>none</td>
<td>B T N</td>
</tr>
<tr>
<td>c.4865A&gt;G</td>
<td>p.Gln1622Arg</td>
<td>none</td>
<td>B T N</td>
</tr>
<tr>
<td>c.5040G&gt;C</td>
<td>p.Arg1680Ser</td>
<td>none</td>
<td>B T D</td>
</tr>
<tr>
<td>c.5154-28C&gt;T</td>
<td>none</td>
<td>none</td>
<td>not done</td>
</tr>
<tr>
<td>c.5155T&gt;A</td>
<td>p.Ser1719Tyr</td>
<td>abolition of the natural acceptor splice site</td>
<td>PrD APF D</td>
</tr>
<tr>
<td>c.5183T&gt;G</td>
<td>p.Phe1728Cys</td>
<td>none</td>
<td>B T D</td>
</tr>
</tbody>
</table>

*Based on the outputs of the majority of utilized bioinformatics programs.

Nor common homozygotes, Het heterozygotes, Hom rare homozygotes; B benign, PrD probably damaging, PoD possibly damaging, U unclassified (PolyPhen-2); T tolerated, APF affecting protein function (SIFT); N neutral, D disease, U unclassified (SNP&GO).
4.3 Investigation of miR-27a rs895819 polymorphism as candidate low-penetrance allele

A series of 1,027 BRCAx cases and 1,593 controls were screened for the miR-27a rs895819 polymorphism, using a TaqMan SNP Genotyping Assay. Two samples were excluded because the genotyping failed. Thus, a total of 1,025 BRCAx cases and 1,593 controls were tested. Genotypes and allelic frequencies were analyzed applying a logistic regression model (Hosmer and Lemeshow, 1989). As performed by Yang and colleagues (Yang et al., 2009), we investigated the association between rs895819 and breast cancer risk in the entire case group, in cases with age at diagnosis >50 years and in cases with bilateral breast cancer. None of these analyses was statistically significant (Table 4.8). An additional analysis to investigate the association with bilaterality was performed in cases only, testing 144 bilateral and 881 unilateral breast cancer cases. Here, we found that the [G] allele was marginally significant associated with an increased risk of bilateral breast cancer versus unilateral breast cancer, with an OR of 1.33 (95 %CI 1.01-1.74, P = 0.041) (Table 4.9).

Based on the size of the examined sample set and the observed frequency of the minor allele [G] in cases and controls, we were able to reject the null hypothesis that this odds ratio equals 1 with probability (power) of 0.183 and a Type I error probability of 0.05.
Table 4.8. Genotype frequencies of miR-27a in 1,025 breast cancer cases and 1,593 controls

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Cases (%)a</th>
<th>Controls (%)b</th>
<th>OR</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All cases</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>547 (53.37)</td>
<td>803 (50.41)</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AG</td>
<td>388 (37.85)</td>
<td>633 (39.74)</td>
<td>0.90</td>
<td>0.76-1.06</td>
<td>0.214</td>
</tr>
<tr>
<td>GG</td>
<td>90 (8.78)</td>
<td>157 (9.86)</td>
<td>0.84</td>
<td>0.63-1.11</td>
<td>0.223</td>
</tr>
<tr>
<td>[G] vs [A]</td>
<td></td>
<td></td>
<td>0.91</td>
<td>0.80-1.02</td>
<td>0.114</td>
</tr>
<tr>
<td>Age at diagnosis ≥ 50 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>126 (55.51)</td>
<td>268 (52.34)</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AG</td>
<td>88 (38.77)</td>
<td>197 (38.48)</td>
<td>0.96</td>
<td>0.68-1.33</td>
<td>0.778</td>
</tr>
<tr>
<td>GG</td>
<td>13 (5.73)</td>
<td>47 (9.18)</td>
<td>0.58</td>
<td>0.30-1.12</td>
<td>0.104</td>
</tr>
<tr>
<td>[G] vs [A]</td>
<td></td>
<td></td>
<td>0.84</td>
<td>0.65-1.09</td>
<td>0.188</td>
</tr>
<tr>
<td>Age at diagnosis &lt; 50 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>421 (52.76)</td>
<td>535 (49.49)</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AG</td>
<td>300 (37.59)</td>
<td>436 (40.43)</td>
<td>0.86</td>
<td>0.71-1.05</td>
<td>0.144</td>
</tr>
<tr>
<td>GG</td>
<td>77 (9.65)</td>
<td>110 (10.18)</td>
<td>0.88</td>
<td>0.64-1.22</td>
<td>0.0447</td>
</tr>
<tr>
<td>[G] vs [A]</td>
<td></td>
<td></td>
<td>0.91</td>
<td>0.78-1.05</td>
<td>0.178</td>
</tr>
<tr>
<td>Bilateral breast cancer cases</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>68 (47.22)</td>
<td>843 (51.62)</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AG</td>
<td>59 (40.97)</td>
<td>633 (38.76)</td>
<td>1.11</td>
<td>0.77-1.59</td>
<td>0.582</td>
</tr>
<tr>
<td>GG</td>
<td>17 (11.81)</td>
<td>157 (9.61)</td>
<td>1.29</td>
<td>0.74-2.26</td>
<td>0.367</td>
</tr>
<tr>
<td>[G] vs [A]</td>
<td></td>
<td></td>
<td>1.13</td>
<td>0.88-1.47</td>
<td>0.338</td>
</tr>
</tbody>
</table>

P trend =0.123

P trend =0.196

P trend =0.191

P trend =0.350

OR, odds ratio adjusted for age; CI, confidence interval

* Median age 41, range: 18-80 years

b Median age 43, range: 18-71 years

Table 4.9. Analyses of genotype frequencies of miR-27a in 144 bilateral and 881 unilateral breast cancer cases

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Bilateral (%)a,b</th>
<th>Unilateral (%)c</th>
<th>OR</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>68 (47.22)</td>
<td>479 (54.37)</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AG</td>
<td>59 (40.97)</td>
<td>329 (37.34)</td>
<td>1.25</td>
<td>0.85-1.83</td>
<td>0.250</td>
</tr>
<tr>
<td>GG</td>
<td>17 (11.81)</td>
<td>73 (8.29)</td>
<td>1.81</td>
<td>1.01-3.28</td>
<td>0.049</td>
</tr>
<tr>
<td>[G] vs [A]</td>
<td></td>
<td></td>
<td>1.33</td>
<td>1.01-1.74</td>
<td>0.041</td>
</tr>
</tbody>
</table>

P trend =0.045

OR, odds ratio adjusted for age; CI, confidence interval

* Median age 45, range: 26-74 years

b Median time interval between first and contralateral cancer 6, range: 0-32 years

c Median age 41, range: 18-80 years

75
4.4 Analysis of the *CASP8* rs3834129 as risk modifier in BRCA genes mutation carriers

A series of 1,241 Italian *BRCA1* and *BRCA2* mutation carriers were recruited for this analysis by different collaborating centers (CONSIT TEAM, Consortium of Italian Studies on Hereditary Breast Cancer; Table 4.10) and screened by direct sequencing. We excluded 18 samples in which the genotyping analysis failed, 13 women who self-reported as "non-Caucasian" and three women who carried both *BRCA1* and *BRCA2* mutations. Overall, 1,207 carriers, of which 508 unaffected and 699 affected with breast cancer, remained included. Of these cases, 740 carried a *BRCA1* mutation and 467 carried a *BRCA2* mutation.

<table>
<thead>
<tr>
<th>Centre</th>
<th><em>BRCA1</em> mutation carriers</th>
<th><em>BRCA2</em> mutation carriers</th>
<th>Total of carriers (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fondazione IRCCS Istituto Nazionale dei Tumori, Milano</td>
<td>349</td>
<td>152</td>
<td>501 (40.4)</td>
</tr>
<tr>
<td>Università degli Studi, Torino</td>
<td>100</td>
<td>69</td>
<td>169 (13.6)</td>
</tr>
<tr>
<td>Istituto Europeo di Oncologia, Milano</td>
<td>95</td>
<td>73</td>
<td>168 (13.5)</td>
</tr>
<tr>
<td>Centro Riferimento Oncologico, Aviano</td>
<td>52</td>
<td>66</td>
<td>118 (9.5)</td>
</tr>
<tr>
<td>Università degli Studi “La Sapienza”, Roma</td>
<td>61</td>
<td>44</td>
<td>105 (8.5)</td>
</tr>
<tr>
<td>Università degli Studi, Firenze</td>
<td>59</td>
<td>30</td>
<td>89 (7.2)</td>
</tr>
<tr>
<td>Istituto Nazionale per la Ricerca sul Cancro, Genova</td>
<td>37</td>
<td>24</td>
<td>61 (4.9)</td>
</tr>
<tr>
<td>Istituto Nazionale Tumori “Regina Elena”, Roma</td>
<td>14</td>
<td>16</td>
<td>30 (2.4)</td>
</tr>
<tr>
<td>All</td>
<td>767</td>
<td>474</td>
<td>1,241 (100)</td>
</tr>
</tbody>
</table>

In this analysis, there was evidence for association of the del allele of rs3834128 with increased breast cancer risk, under the dominant model, with a HR of 1.35 (95 % CI 1.04-
1.76, \( P = 0.023 \) for \( BRCA1 \) and \( BRCA2 \) mutation carriers combined and 1.52 (95\% CI 1.14-2.02, \( P = 0.004 \)) for \( BRCA1 \) mutation carriers only (Table 4.11). Oppositely, the analysis failed to suggest any association in \( BRCA2 \) mutation carriers.

Table 4.11. Genotype frequencies of rs3834129 by \( BRCA \) mutation and disease status, and corresponding hazard ratios estimated in the overall group of \( BRCA \) mutation carriers

<table>
<thead>
<tr>
<th>( BRCA ) Group</th>
<th>Genotype</th>
<th>Unaffected (%</th>
<th>Affected (%</th>
<th>HR</th>
<th>95% CI</th>
<th>( P )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( BRCA1 ) and ( BRCA2 ) (N=1,207)</td>
<td>nor/nor</td>
<td>177 (34.8)</td>
<td>208 (29.8)</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>nor/del</td>
<td>224 (44.1)</td>
<td>346 (49.5)</td>
<td>1.40</td>
<td>1.06-1.85</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>del/del</td>
<td>107 (21.1)</td>
<td>145 (20.7)</td>
<td>1.26</td>
<td>0.90-1.77</td>
<td>0.176</td>
</tr>
<tr>
<td></td>
<td>del vs nor (per allele)</td>
<td></td>
<td></td>
<td>1.15</td>
<td>0.96-1.38</td>
<td>0.121</td>
</tr>
<tr>
<td></td>
<td>Dominant (del)</td>
<td></td>
<td></td>
<td>1.35</td>
<td>1.04-1.76</td>
<td>0.023</td>
</tr>
<tr>
<td>( BRCA1 ) (N = 740)</td>
<td>nor/nor</td>
<td>125 (35.5)</td>
<td>105 (27.1)</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>nor/del</td>
<td>152 (43.2)</td>
<td>193 (49.7)</td>
<td>1.56</td>
<td>1.16-2.13</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>del/del</td>
<td>75 (21.3)</td>
<td>90 (23.2)</td>
<td>1.42</td>
<td>1.00-2.03</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>del vs nor (per allele)</td>
<td></td>
<td></td>
<td>1.23</td>
<td>1.02-1.49</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td>Dominant (del)</td>
<td></td>
<td></td>
<td>1.52</td>
<td>1.14-2.02</td>
<td>0.004</td>
</tr>
<tr>
<td>( BRCA2 ) (N = 467)</td>
<td>nor/nor</td>
<td>52 (33.3)</td>
<td>103 (33.1)</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>nor/del</td>
<td>72 (46.2)</td>
<td>153 (49.2)</td>
<td>1.09</td>
<td>0.66-1.78</td>
<td>0.746</td>
</tr>
<tr>
<td></td>
<td>del/del</td>
<td>32 (20.5)</td>
<td>55 (17.7)</td>
<td>0.80</td>
<td>0.42-1.55</td>
<td>0.516</td>
</tr>
<tr>
<td></td>
<td>del vs nor (per allele)</td>
<td></td>
<td></td>
<td>0.92</td>
<td>0.66-1.29</td>
<td>0.619</td>
</tr>
<tr>
<td></td>
<td>Dominant (del)</td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.62-1.60</td>
<td>0.988</td>
</tr>
</tbody>
</table>

**HR** hazard ratio, **CI** confidence interval, nor/nor common homozygotes, nor/del heterozygotes, del/del rare homozygotes

We performed an additional analysis and classified \( BRCA1 \) and \( BRCA2 \) mutations in two different groups, based on the expected functional effect of each mutation, according to criteria established by CIMBA (Antoniou et al., 2009). Class 1 mutations include those potentially causing the complete loss-of-function of the protein, while class 2 mutations include those, potentially responsible for the formation of stable mutant protein with a possible dominant negative effect. The group of mutation carriers of class 1 was analyzed
separately, whereas the group of mutation carries of class 2 was too small to justify a separate analysis. Here, we found an association of the del allele, under the dominant model, with an increased breast cancer risk with a HR of 1.46 (95 % CI 1.08-1.99, \( P = 0.015 \)) for \( BRCA1 \) and \( BRCA2 \) mutation carriers combined and 1.74 (95 % CI 1.24-2.46, \( P = 0.002 \)) for only \( BRCA1 \) mutation carriers (Table 4.12). Consistently with previous analyses, no association was found in \( BRCA2 \) mutation carriers.

Table 4.12. Genotype frequencies of rs3834129 by BRCA mutation and disease status, and corresponding hazard ratios estimated in carriers of \( BRCA \) loss-of-function (class 1) mutations

<table>
<thead>
<tr>
<th>( BRCA ) group</th>
<th>Genotype</th>
<th>Unaffected (%)</th>
<th>Affected (%)</th>
<th>HR</th>
<th>95% CI</th>
<th>( P )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( BRCA1 ) and ( BRCA2 ) (N=920)</td>
<td>nor/nor</td>
<td>143 (36.8)</td>
<td>156 (29.4)</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>nor/del</td>
<td>165 (42.4)</td>
<td>265 (49.9)</td>
<td>1.56</td>
<td>1.12-2.15</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>del/del</td>
<td>81 (20.8)</td>
<td>110 (20.7)</td>
<td>1.29</td>
<td>0.87-1.93</td>
<td>0.208</td>
</tr>
<tr>
<td></td>
<td>del vs nor (per allele) Dominant (del)</td>
<td></td>
<td></td>
<td>1.18</td>
<td>0.95-1.46</td>
<td>0.138</td>
</tr>
<tr>
<td></td>
<td>nor/nor</td>
<td>94 (38.4)</td>
<td>69 (26.6)</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( BRCA1 ) (N = 504)</td>
<td>nor/del</td>
<td>100 (40.8)</td>
<td>128 (49.4)</td>
<td>1.83</td>
<td>1.27-2.64</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>del/del</td>
<td>51 (20.8)</td>
<td>62 (23.9)</td>
<td>1.60</td>
<td>1.03-2.48</td>
<td>0.035</td>
</tr>
<tr>
<td></td>
<td>del vs nor (per allele) Dominant (del)</td>
<td></td>
<td></td>
<td>1.33</td>
<td>1.05-1.69</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td>nor/nor</td>
<td>49 (34.0)</td>
<td>87 (32.0)</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( BRCA2 ) (N = 416)</td>
<td>nor/del</td>
<td>65 (45.1)</td>
<td>137 (50.4)</td>
<td>1.24</td>
<td>0.73-2.10</td>
<td>0.419</td>
</tr>
<tr>
<td></td>
<td>del/del</td>
<td>30 (20.8)</td>
<td>48 (17.6)</td>
<td>0.79</td>
<td>0.40-1.57</td>
<td>0.506</td>
</tr>
<tr>
<td></td>
<td>del vs nor (per allele) Dominant (del)</td>
<td></td>
<td></td>
<td>0.92</td>
<td>0.65-1.32</td>
<td>0.662</td>
</tr>
<tr>
<td></td>
<td>nor/nor</td>
<td>1.09</td>
<td>0.66-1.80</td>
<td>0.738</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( HR \) hazard ratio, \( CI \) confidence interval, nor/nor common homozygotes, nor/del heterozygotes, del/del rare homozygotes

It was possible to reject the null hypothesis that the odds ratio observed in affected carriers of the del allele compared to non-carriers equals 1 with probability (power) of 0.727 with a Type I error probability of 0.05. However, the analysis of \( BRCA1 \) and \( BRCA2 \) modifiers is potentially more complex of the classical case-control association study, in
which genotype frequencies are compared between cases and controls, because a high proportion of carriers become affected. More powerful analyses can be conducted by treating breast cancer as a survival (age at onset), rather than a simple binary, endpoint. To increase the power of the tested hypothesis, a weighted Cox regression model can be used, as reported in our study performed in collaboration with the CIMBA Consortium (Osorio et al., 2009; Antoniou et al., 2005).
CHAPTER 5
DISCUSSION

In the present study, the role of different candidate breast cancer susceptibility loci has been investigated. In particular, the two Fanconi Anemia genes *PALB2* and *SLX4* have been screened to assess their impact in breast cancer predisposition genes in the Italian population. In addition, I sought to elucidate the role of rs895819, located in the miR-27a gene, and rs3834129, located in the promoter region of the *CASP8* gene, as low-risk allele and genetic risk modifier, respectively.

The entire coding region and corresponding splice sites of the *PALB2* gene was screened for mutations in 575 familial BRCAx cases and a total of 34 different variants were detected. While eight were previously reported as common and considered as neutral polymorphisms, 26 were rare or unique mutations. Of these, eight were novel truncating mutations, including two, c.72delG (p.Arg26fs) and c.1027C>T (p.Gln343X), that were recurrent, and one, c.48G>A, was ascertained to alters the canonical mRNA splicing and to introduce a premature termination codon. None of above mutations were found in 784 controls, recruited in Milan.

Overall, we found 12 individuals carrying a *PALB2* truly pathogenic mutation, for a frequency of 2.1%. Even if this frequency appears to be higher with respect to that observed in the other two Italian studies, where mutation frequencies of 0.75% and 1.1% were found (Papi et al., 2009; Balia et al., 2010), this result is comparable to that observed
in other populations, where \textit{PALB2} mutations were detected with frequency from 0.2 to 3.3%, and confirms the role of this gene in breast cancer susceptibility.

Interestingly, the actual proportion of carriers of pathogenic \textit{PALB2} mutations in our group could be even higher. In fact, \textit{in silico} analyses indicated that the c.2379C>T (p.Gly793Gly) and the c.2418G>T (p.Pro806Pro) affect the canonic mRNA splicing, causing the activation of a cryptic donor and acceptor splice site, respectively. In addition, the missense mutations c.2792T>G (p.Leu931Arg) and c.2816G>T (p.Leu939Trp) were predicted to be deleterious by all of the three software used in this analysis. Interestingly, the c.2792T>G is located in a highly conserved residue (in all species from \textit{H. sapiens} to \textit{D. rerio}) in the WD40-repeat domain of the protein, that is responsible for the BRCA2 binding.

Additional analyses performed on the c.1027C>T (p.Gln343X) showed that this is a recurrent mutation, with an higher frequency in cases and controls recruited in the Bergamo area, with respect to that observed in cases and controls recruited in Milan. In particular, we identified 5/112 (4.5%) and 2/477 (0.4%) carriers of the c.1027C>T in cases and controls from Bergamo versus 3/907 (0.3%) and 0/960 (0.0%) in those recruited in cancer centers in Milan, which recruit patients from all of the country, with an increased frequency in cases of more than 10-fold.

Interestingly, a similar frequency difference was also observed for a few \textit{BRCA1} and \textit{BRCA2} mutations carriers. In particular, two of these mutations, identified through routine BRCA gene testing, occur with a 10-fold higher frequency in the Bergamo area with respect to that observed in cases recruited in Milan. These are the \textit{BRCA1} C64R, reported in 15/2,065 (0.7%) cases recruited in Milan and in 15/158 (9.5%) of those recruited in Bergamo, and the \textit{BRCA2} V1969fs, detected in 5/2,065 (0.2%) cases recruited in Milan and in 5/158 (3.2%) of those recruited in Bergamo (unpublished data; Table 5.1).
Table 5.1. Frequency of BRCA1 C64R, BRCA2 V1969fs and PALB2 c.1027C>T recurrent mutations in probands recruited in Milan and Bergamo

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Carriers/Probands from Milan</th>
<th>Carriers/Probands from Bergamo</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRCA1 C64R</td>
<td>15/2,065 (0.7%)</td>
<td>15/158 (9.5%)</td>
</tr>
<tr>
<td>BRCA2 V1969fs</td>
<td>5/2,065 (0.2%)</td>
<td>5/158 (3.2%)</td>
</tr>
<tr>
<td>PALB2 c.1027C&gt;T</td>
<td>3/575 (0.5%)</td>
<td>5/112 (4.5%)</td>
</tr>
</tbody>
</table>

Previous studies showed the presence of PALB2 founder mutations in Finnish (2.7%; Erkko et al., 2007), French-Canadian (0.7%; Ghadirian et al., 2009), Polish (0.6%; Dansonka-Mieszkowska et al., 2010) and Australian (0.4%; Southey et al., 2010) breast cancer cases and also in Finnish (0.2%; Erkko et al., 2007) and Polish (0.08%; Dansonka-Mieszkowska et al., 2010) unrelated controls. All these populations are geographically isolated and/or characterized by rapid expansion and low grade of immigration. However, compared to the previously mentioned populations, Italians are much more heterogeneous, characterized by ethnic mixture, a higher grade of internal immigration and lower probability of finding founder mutations. Nevertheless, some BRCA genes founder mutations were observed in different regional areas. The first Italian founder mutation, the BRCA1 5083del19, was identified by Baudi and colleagues, with the high frequency of 16.7% in individuals from Calabria (Baudi et al., 2001). Other BRCA1 and BRCA2 founder mutations were also described in breast cancer cases from other Italian regions, including Sardinia, Sicily, Tuscany, Calabria and the Northeast of Italy (Pisano et al. 2000; Russo et al. 2007; Papi et al. 2009; Russo et al. 2009; Malacrida et al., 2008).

In conclusion, we identified the novel PALB2 c.1027C>T mutation and observed that this is recurrent in the area of Bergamo, being detecting with a 10-fold increased frequency with respect to that observed in breast cancer cases recruited in Milan. This result, combined with the identification of two other recurrent mutations in BRCA1 and BRCA2 in
the same area, suggests that this region could be characterized by a lower grade of genetic heterogeneity. This hypothesis is reinforced by the fact that this area includes several deep and isolated valleys that could prevent genetic admixture and promote the circulation of a small number of mutations with higher frequencies. However, further analyses are necessary to confirm these results.

Through the analysis of families screened for PALB2 mutations, we observed an apparent excess of pancreatic cancer cases in families with pathogenic PALB2 pathogenic mutations. In particular, we identified 39/575 families with both breast and pancreatic cancer cases and in three of them the index case carried a truncating mutation, with a frequency of 7.7% (3/39). However, no other truncating mutation was found in additional 23 index cases from breast and pancreatic cancer families, for a final frequency of 4.8% (3/62).

In early 2011, breast cancer cases with personal/family history of pancreatic cancer were screened for PALB2 mutations in two different studies, showing a cumulative mutation frequency of 1.2% (Hofstatter et al., 2011; Stadler et al., 2011). Although our reported frequency appears to be increased with respect to that observed in these studies, this result is not significant, possibly due to the limited sample size. If a preferential co-occurrence of breast and pancreatic cancer in PALB2 positive families exists, it could be detected only in larger analyses.

To evaluate its contribution in breast cancer susceptibility, the entire coding region and intron/exon junctions of the SLX4 gene were screened in 526 Italian familial breast cancer cases. Even if a large amount of different variants were identified (n=81), none of them could be classified as truly pathogenic. In silico analyses predicted as deleterious two
of the identified variants. These are c.833G>A (p.Arg278Gln), with a possible effect on canonical splicing, and c.5155T>A (p.Ser1719Tyr), with a predicted negative effect both on the splicing and on the protein structure. Since it appears unlikely that the c.5155T>A mutation may have a pathological effect both at splicing and protein level, the latter observation underlines the actual limitations of *in silico* tools of predicting the pathogenic role of genetic variants.

These data indicate that *SLX4* is unlikely to act as a breast cancer susceptibility gene in the Italian population. In addition, we could estimate that truncating *SLX4* mutations are very rare in our population. In fact, if such mutations were present in our sample group of 526 individuals with a frequency of no less than 0.6% (corresponding to the 95% CI upper limit of the event probability in a sample of the examined size negative for truncating mutations), we would have a 95% probability to detect at least one such mutation.

To date, only two *SLX4* truncating mutations were reported in breast cancer cases from four different studies (Landwehr et al., 2011; Fernandez-Rodriguez et al., 2012; de Garibay et al., 2012; Bakker et al., 2013) even if a very large amount of non-truncating variants were detected in each of them (Table 5.2). Therefore, considering our results, combined with those reported in these studies, a cumulative mutation frequency of 0.1% was found in a total of 1,887 tested cases. These results are consistent with the hypothesis that *SLX4* pathogenic mutations are very rare and this gene plays a marginal role in breast cancer susceptibility.

In the present study, the analysis of the rs895819 variant, located in the miR-27a, failed to support any association of this SNP with breast cancer risk in Italian familial breast cancer cases, as reported in the German population (Yang et al., 2009). A subsequent analysis, performed considering cases only, showed a marginal association of
the minor allele with increased risk of bilateral breast cancer (OR = 1.33, 95 % CI 1.01-1.74, P = 0.041).

Table 5.2. Frequency of SLX4 truncating and non-truncating mutations

<table>
<thead>
<tr>
<th>Study</th>
<th>Number of tested cases</th>
<th>Truncating mutations (frequency)</th>
<th>Non-truncating mutations (frequency)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present study</td>
<td>526</td>
<td>0 (0%)</td>
<td>81 (15.4%)</td>
</tr>
<tr>
<td>Landwehr et al., 2011</td>
<td>52</td>
<td>0 (0%)</td>
<td>27 (51.9%)</td>
</tr>
<tr>
<td>Fernandez-Rodriguez et al., 2012</td>
<td>94</td>
<td>0 (0%)</td>
<td>49 (52.1%)</td>
</tr>
<tr>
<td>de Garibay et al., 2012</td>
<td>486</td>
<td>1 (0.2%)</td>
<td>56 (11.5%)</td>
</tr>
<tr>
<td>Bakker et al., 2013</td>
<td>729</td>
<td>1 (0.1%)</td>
<td>102 (14%)</td>
</tr>
</tbody>
</table>

To date, studies that have explored the association of the rs895819 with breast cancer risk, reported controversial results. As mentioned above, an association of the minor allele with a reduced risk has been reported in familial breast cancer cases from Germany (Yang et al., 2009). However, a study involving a series of 252 Chinese breast cancer cases and 248 unrelated controls failed to confirm this association (Zhang et al., 2012).

Similar controversial data were also observed in other types of cancer. In 2010, 304 Chinese gastric cancer cases and 304 controls were screened and an association of the SNP with an increased risk was detected (OR = 1.48, 95 % CI 1.06-2.05, P = 0.019; Sun et al., 2010). Nevertheless, a subsequent analysis involving 311 Chinese gastric cancer cases and 427 controls indicated an opposite result, showing an association of the minor allele of rs895819 with a reduced risk (OR = 0.771, 95 % CI 0.604-0.985, P = 0.037; Zhou et al., 2012). In 2012, Zhong and colleagues performed a meta-analysis considering six different studies, of which three involving Caucasian and three Asian individuals, for breast and gastric cancer. The overall analysis did not show any association of rs895819 with cancer
susceptibility whereas stratification by cancer type indicated only a marginal association with reduced risk of developing breast cancer (OR = 0.92, 95% CI 0.74-1.14, \( P = 0.04 \); Zhong et al., 2013). Very recently, in addition to breast and gastric cancer, an association of rs895819 with reduced risk has been reported in renal cell cancer (OR = 0.71, 95% CI 0.56-0.90, \( P = 0.004 \); Shi et al., 2012), whereas no associations were found with colorectal and lung cancer (Hezova et al., 2012; Yoon et al., 2012). Finally, an additional meta-analysis was carried out in 2013, including results from seven studies involving Caucasian and Asian cases affected with breast, gastric, colorectal and renal cell cancer. Here, no association was found both in the overall analysis and after the stratification of samples by tumor type (Wang et al., 2013).

To date, several SNPs in microRNA genes were reported to be associated with breast cancer risk, even if these associations failed to be confirm in subsequent analyses. Interestingly, a very recent GWAS reported 41 novel SNPs associated with breast cancer, including two SNPs located in miRNA genes. This are the rs11780156, in miR-1208, and the rs17817449, in the miR-1972-2 (Michailidou et al., 2013).

In conclusion, these studies did not provide conclusive results on the impact of rs895819 both in breast and in other types of cancer. The apparent conflicting results could be due to an underpowered sample size of each single study and could be resolved only by the establishment of larger collaboration studies, involving a larger number of individuals.

The analysis of the rs3834129 ins/del polymorphism, located in the promoter region of \( \text{CASP}8 \), suggested an association of the del allele with an increased breast cancer risk in \( \text{BRCA1} \) mutation carriers.

Our study indicated that the SNP rs3834129 del allele can be also considered a genetic risk modifier of breast cancer in BRCA genes mutation carriers, in a similar way as
reported for other low-penetrance alleles (Antoniou et al., 2008; Antoniou et al., 2009, Antoniou et al., 2010; Antoniou et al., 2011). However, the association of this SNP with breast cancer risk in BRCA1 mutation carriers was in the opposite direction with respect to that described in the general population, where it has been reported that the rs895819 del allele reduces the risk of developing breast cancer (Sun et al., 2007; Sergentanis and Economopoulos, 2009; Yin et al., 2010). These contradictory results may be due to the different clinical and pathological characteristics of BRCA1-related tumors compared with breast cancer in the general population. Typically, the majority of BRCA1-related tumors are of the “triple-negative” type, being characterized by the absence immunohistochemical positivity for estrogen receptor (ER), progesterone receptor (PR) and a negativity for HER2 expressing cells. In addition, they present high proliferation grade, higher frequency of TP53 somatic mutations and are generally associated with a poor prognosis. In contrast to BRCA1-related tumors, breast cancer in the general population, and also in BRCA2 mutation carriers, are more frequently ER-positive and PR-positive, with lower frequency of TP53 mutations and low proliferation grade (reviewed in Mavaddat et al., 2010 and in Lalloo and Evans, 2012). Consistently, it has been suggested that low-penetrance alleles in unselected breast cancer are more frequently genetic modifiers for BRCA2 than for BRCA1 mutation carriers. In fact, it has been reported that low-risk alleles associated with ER-positive tumors in the general population tend to be associated with increased risk in BRCA2 mutation carriers, whereas those associated with ER-negative tumors appear preferentially associated with BRCA1 mutation carriers (reviewed in Milne and Antoniou, 2011). As an example, it has been reported that a polymorphism located in the CASP8 gene (p.D302I; rs1045485) and associated with decreased breast cancer risk only in BRCA1 mutation carriers, appears to have a stronger protective effect in ER-negative (OR = 0.90, 95 % CI 0.84-0.96, P = 0.001) than in ER-positive breast cancer cases of the general population (Brocks et al., 2011).
To our knowledge, this is the first report of an the association of the rs3834129 del allele, which would appear to be a protective factor in the general population, with increased breast cancer risk in BRCA1 mutation carriers.

CONCLUSIONS AND FUTURE DIRECTIONS

The evidence of familial predisposition to breast cancer encouraged an extensive research of genes underlying this susceptibility. In this context, a large number of susceptibility loci have emerged, classified on the basis of the risk that they conferred. The present investigation provides a contribution to the general knowledge on breast cancer predisposition. Firstly, it confirms the role of PALB2 mutations in increasing breast cancer risk and led to the identification of a possible founder mutation in a specific area in the Northern Italy, the province of Bergamo. In addition, this analysis contributes to assess the marginal role of the Fanconi Anemia gene SLX4 in breast cancer susceptibility. Finally, it provides support to the role of rs3834129 as risk modifier in BRCA1 mutations carriers increasing the possibility to improve risk assessment of breast cancer in these individuals.

One of the major limitation of this study is represented by the relatively reduced size of the sample tested. This issue firstly emerges in the investigation of the association of PALB2 mutations with pancreatic cancer, due to the very small series of recruited families with both breast and pancreatic cancer cases. A similar limitation regards also the analysis of the role of SLX4 gene in breast cancer susceptibility. While we excluded a major role of mutations in this gene in breast cancer predisposition, truncating mutations have been subsequently found in breast cancer cases, even if rare, resulting in a difficult assessment of the SLX4 role. Considering the low frequency of these truncating mutations and the costs inherent to their detection by conventional re-sequencing, the role of this gene could
be established using next generation sequencing approaches. Finally, the limited sample size represents the main issue also for association studies. In fact, all of the candidate low-penetrance alleles and genetic modifiers found to be associated with breast cancer risk in single studies, only the CASP8/p.D302H (rs1045485) and the RAD51C/c.135G>C (rs1801320) polymorphisms, respectively, were confirmed to be associated with breast cancer in larger analyses (Mavaddat et al., 2010).

In the last two decades, great advances have been made in understanding breast cancer genetic predisposition and several novel genes have been identified. Nevertheless, mutations in these genes are not sufficient to explain all familial breast cancer cases. To date, known high-, moderate- and low-penetrance alleles account for no more than 30-35% of breast cancer familial clustering, leaving most of them unexplained (reviewed in Mavaddat et al., 2010 and in Laloo and Evans, 2012; Michaillidou et al., 2013).

An interesting field that remains to be explored is to identify novel breast cancer predisposition alleles in the FA pathway. It is known that mutations in several FA genes, including FANCD1/BRCA2, FANCJ/BRIPl, FANCN/PALB2 and FANCO/RAD51C, increase the risk of developing breast cancer (reviewed in Hollestelle et al., 2010). These are downstream genes of the FA pathway, directly involved in the DNA repair mechanism by mediating homologous recombination. However, this pathway consists of a large number of genes with different functions, all potentially candidates as breast cancer predisposing factors. As an example, potential breast cancer susceptibility alleles have been recently reported in two genes forming the FA core complex, FANCA and FANCC (Litim et al., 2013; Thompson et al., 2012), supporting the notion that a fraction of the unexplained hereditary breast cancer cases could be due to mutations in other genes of the FA pathway.
Together with this candidate gene approach, technological improvements have provided agnostic strategies, such as whole exome and/or whole genome sequencing, that allow efficient mutation identification without any prior biological or molecular knowledge. It has been suggested that these analyses may represent a good approach to the detection of loci responsible for heterogenic diseases, such as hereditary breast cancer (Gracia-Aznarez at al., 2013).

Of all of identified mutations in cancer predisposing genes, only a fraction has a clearly pathogenic role. Thus, the characterization of variants of unclassified significance (VUSs) represents another essential issue for the understanding breast cancer susceptibility. To date, a large amount of VUSs have been identified in breast cancer genes, both in BRCA genes and in the other high- and moderate-penetrance genes. Most of them are extremely rare and their effect on breast cancer risk cannot be measured by association studies. The assessment of the role of these variants may be improved using a combined approach based on genetic and epidemiological data, the use of in silico tools for prediction of the effects of these mutations on canonical splicing and/or protein structure and functioning and, finally, the development of novel functional assays to test the biological effects of the variants (Radice et al., 2011). In particular, multifactorial likelihood prediction models have been developed, in which the probability of a genetic variant to be pathogenic is estimated combining the prior probability based on in silico analyses with likelihood ratios derived from the analysis of different features including: a) co-segregation of the variant with the disease; b) co-occurrence of the variant with pathogenic mutations; c) personal family history in variant carriers; d) pathological characteristics of tumor in variant carriers (Lindor et al., 2012). The usefulness of multifactorial models depend on the amount of data that are available in order to reach the odds ratios that are required for a realible classification of VUSs. This has prompted the
established of large international consortia such as the Evident-based Network for the Interpretation of Germline Mutant Alleles (ENIGMA; Spurdle et al., 2012).

The advent of large international consortia and the GWASs has allowed identification of over 70 low-risk alleles for breast cancer. It has been speculated that, although these loci contribute only a small risk increase, they might act with a combined effect (reviewed in Varghese and Easton, 2010). Further large scale genotyping efforts may disclose a larger number of novel low-penetrance loci, substantially increasing individual risk prediction (Michailidou et al., 2013). An identical strategy has to be pursued for the identification of additional genetic risk modifiers in \textit{BRCA1} and \textit{BRCA2} mutation carriers, particularly because the contribution of these loci in high-risk families results in larger risk differences with respect to that observed in the general population (reviewed in Milne and Antoniou, 2011).

In conclusion, clearer understanding of breast cancer genetic susceptibility will be obtained using a combined strategy based on the identification of candidate loci in small selected populations by large scale sequencing, followed by the validation of these loci in large association studies, including several thousands of samples, and on the assessment of the clinical relevance of identified variants by integrated genetic and functional analyses.

With the discovery of novel predisposition genes and genetic risk modifiers and their characterization, a more complete polygenic risk profile may be obtained for each affected or at-risk individual and in different populations. In addition, a personal risk assessment by profiling of known predisposition loci might lead to the development of more accurate surveillance programs and more adequate treatments, specifically designed for each patient.


**References**
References


References


References


References


References


### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Breast cancer genetic susceptibility loci</td>
<td>8</td>
</tr>
<tr>
<td>1.2</td>
<td>A schematic model of the Fanconi Anemia (FA) pathway</td>
<td>19</td>
</tr>
<tr>
<td>1.3</td>
<td>MicroRNA maturation processes</td>
<td>27</td>
</tr>
<tr>
<td>1.4</td>
<td>Breast cancer age-specific cumulative risk for BRCA2 mutation carriers</td>
<td>34</td>
</tr>
<tr>
<td>1.5</td>
<td>Comparison of predicted age-specific cumulative breast cancer risks in the general population and in BRCA2 mutation carriers</td>
<td>34</td>
</tr>
<tr>
<td>1.6</td>
<td>Structure of the CASP8 promoter region</td>
<td>36</td>
</tr>
<tr>
<td>3.1</td>
<td>The entire process of recruitment and selection of individuals affected with breast cancer or at-risk of breast cancer included in the present studies</td>
<td>41</td>
</tr>
<tr>
<td>4.1</td>
<td>Bioinformatics analysis of the PALB2 c.48G&gt;A splicing mutation</td>
<td>57</td>
</tr>
<tr>
<td>4.2</td>
<td>Characterization of the mRNA transcripts caused by the PALB2 c.48G&gt;A mutation</td>
<td>58</td>
</tr>
<tr>
<td>4.3</td>
<td>Electropherograms of the sequences where a PALB2 pathogenic mutation was found.</td>
<td>60</td>
</tr>
<tr>
<td>4.4</td>
<td>Pedigrees of the 12 families in which the index case carried a PALB2 truncating mutation</td>
<td>63</td>
</tr>
<tr>
<td>4.5</td>
<td>Pedigrees of the five families recruited at the Ospedali Riuniti of Bergamo, in which the index case carried the c.1027C&gt;T PALB2 mutation</td>
<td>68</td>
</tr>
<tr>
<td>4.6</td>
<td>Venn diagram represents the distribution of PALB2 truncating mutations in breast and breast/pancreatic cancer families recruited in this study</td>
<td>68</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1.1  *BRCA1* and *BRCA2* founder mutations identified in European countries  12

Table 1.2  *BRCA1* and *BRCA2* founder mutations identified in non-European countries  13

Table 1.3  Low-penetrance breast cancer susceptibility loci identified in the initial GWASs  26

Table 1.4  SNPs associated with breast cancer risk modification in *BRCA1* mutation carriers  32

Table 1.5  SNPs associated with breast cancer risk modification in *BRCA2* mutation carriers  33

Table 3.1  *CASP8* primer sequences and PCR conditions  44

Table 3.2  *PALB2* primer sequences and PCR conditions  45

Table 3.3  *SLX4* primer sequences and PCR conditions  46

Table 3.4  TaqMan assay information of the *PALB2* c.72delG mutation  47

Table 3.5  TaqMan assay information of the *PALB2* c.1027C>T mutation  47

Table 3.6  *PALB2* PCR primer and conditions for c.48G>A variant characterization  51

Table 4.1  Frequencies of common *PALB2* polymorphisms in 575 BRCAX cases and 784 controls  53

Table 4.2  *PALB2* truncating mutations found in 575 BRCAX cases  54

Table 4.3  Frequencies of rare or unique non-truncating *PALB2* variants in cases and controls  56

Table 4.4  Characteristics of families with *PALB2* truncating mutations  59

Table 4.5  Number of carriers of 1027C>T mutation in BRCAX and  61
controls recruited in Milan and Bergamo

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 4.6</td>
<td>Frequencies of SLX4 variants in BRCAX cases</td>
<td>71</td>
</tr>
<tr>
<td>Table 4.7</td>
<td><em>In silico</em> analyses of SLX4 variants in BRCAX cases</td>
<td>73</td>
</tr>
<tr>
<td>Table 4.8</td>
<td>Genotype frequencies of miR-27a in 1,025 breast cancer cases and 1,593 controls</td>
<td>75</td>
</tr>
<tr>
<td>Table 4.9</td>
<td>Analyses of genotype frequencies of miR-27a in 144 bilateral and 881 unilateral breast cancer cases</td>
<td>75</td>
</tr>
<tr>
<td>Table 4.10</td>
<td>List of centers participating to CONSIT TEAM with numbers of BRCA1 and BRCA2 mutation carriers contributed</td>
<td>76</td>
</tr>
<tr>
<td>Table 4.11</td>
<td>Genotypes frequencies of rs3834129 by BRCA mutation and disease status, and corresponding hazard ratio estimated in the overall group of BRCA mutation carriers</td>
<td>77</td>
</tr>
<tr>
<td>Table 4.12</td>
<td>Genotype frequencies of rs3834129 by BRCA mutation and disease status, and corresponding hazard ratio estimated in carriers of BRCA loss-of-function (class 1) mutations</td>
<td>78</td>
</tr>
<tr>
<td>Table 5.1</td>
<td>Frequency of BRCA1 C64R, BRCA2 V1969fs and PALB2 c.1017C&gt;T recurrent mutations in probands recruited in Milan and Bergamo</td>
<td>82</td>
</tr>
<tr>
<td>Table 5.2</td>
<td>Frequency of SLX4 truncating and non-truncating mutations</td>
<td>85</td>
</tr>
</tbody>
</table>
MATERIALS PUBLISHED FROM THIS THESIS

Published from CHAPTER 4 – RESULTS, paragraph 4.1.3 “PALB2 mutation analysis in breast and pancreatic cancer families”:

Published from CHAPTER 4 – RESULTS, paragraph 4.2 “Sequencing analysis of SLX4/FANCP gene”:

Published from CHAPTER 4 – RESULTS, paragraph 4.3 “Investigation of miR-27a rs895819 polymorphism as candidate low-penetrance allele”:
Published from CHAPTER 4 – RESULTS, paragraph 4.4 “Analysis of the CASP8 rs3834129 ins/del polymorphism as genetic modifier”: