



Article

Detection of Germline Variants in 450 Breast/Ovarian Cancer Families with a Multi-Gene Panel Including Coding and Regulatory Regions

Chiara Guglielmi ¹, Rosa Scarpitta ², Gaetana Gambino ³, Eleonora Conti ¹, Francesca Bellè ⁴, Mariella Tancredi ¹, Tiziana Cervelli ⁴, Elisabetta Falaschi ¹, Cinzia Cosini ¹, Paolo Aretini ⁵, Caterina Congregati ⁶, Marco Marino ⁷, Margherita Patruno ⁸, Brunella Pilato ⁸, Francesca Spina ⁹, Luisa Balestrino ⁹, Elena Tenedini ⁷, Ileana Carnevali ¹⁰, Laura Cortesi ¹¹, Enrico Tagliafico ⁷, Maria Grazia Tibiletti ¹⁰, Stefania Tommasi ⁸, Matteo Ghilli ¹², Caterina Vivonet ⁹, Alvaro Galli ^{4,*} and Maria Adelaide Caligo ^{1,*}

- ¹ SOD Molecular Genetics, University Hospital of Pisa, 56126 Pisa, Italy; chiara.guglielmi.cg@gmail.com (C.G.); eleonoraconti4@gmail.com (E.C.); mariella.tancredi@gmail.com (M.T.); cunegonda_f@yahoo.it (E.F.); c.cosini@ao-pisa.toscana.it (C.C.)
- ² Division of Pathology, University of Pisa, 56126 Pisa, Italy; rose2312@hotmail.it
- ³ Department of Clinical and Experimental Medicine, University of Pisa, 56126 Pisa, Italy; gaetana.gambino@gmail.com
- ⁴ Functional Genetics and Genomics Lab, Institute of Clinical Physiology, IFC-CNR, 56127 Pisa, Italy; f.belle01@hotmail.com (F.B.); tizicerv@ifc.cnr.it (T.C.)
- ⁵ Section of Oncological Genomics, Fondazione Pisana per la Scienza, 56017 Pisa, Italy; p.aretini@fpscience.it
- ⁶ Division of Internal Medicine, University Hospital of Pisa, 56126 Pisa, Italy; c.congregati@ao-pisa.toscana.it
- ⁷ Department of Life Sciences, University of Modena and Reggio Emilia, 41125 Modena, Italy; marcomarino83@gmail.com (M.M.); elena.tenedini@unimore.it (E.T.); enrico.tagliafico@unimore.it (E.T.)
- ⁸ IRCCS Istituto Tumori "Giovanni Paolo II", 70124 Bari, Italy; m.patruno@oncologico.bari.it (M.P.); brunellapilato@gmail.com (B.P.); s.tommasi@oncologico.bari.it (S.T.)
- ⁹ SC Medical Genetics, ASSL Cagliari, 09126 Cagliari, Italy; francescaspina@hotmail.com (F.S.); luibale@hotmail.com (L.B.); cateviva@libero.it (C.V.)
- ¹⁰ Ospedale di Circolo ASST Settelaghi, 21100 Varese, Italy; ileana.carnevali@asst-settelaghi.it (I.C.); mariagrazia.tibiletti@asst-settelaghi.it (M.G.T.)
- ¹¹ Department of Oncology, Haematology and Respiratory Diseases, University Hospital of Modena, 41124 Modena, Italy; cortesilaura67@gmail.com
- ¹² Breast Cancer Center, University Hospital, 56126 Pisa, Italy; m.ghilli@ao-pisa.toscana.it
- * Correspondence: alvaro.galli@ifc.cnr.it (A.G.); adelaide.caligo@do.unipi.it (M.A.C.)

Citation: Guglielmi, C.; Scarpitta, R.; Gambino, G.; Conti, E.; Bellè, F.; Tancredi, M.; Cervelli, T.; Falaschi, E.; Cosini, C.; Aretini, P.; et al. Detection of Germline Variants in 450 Breast/Ovarian Cancer Families with a Multi-Gene Panel Including Coding and Regulatory Regions. *Int. J. Mol. Sci.* **2021**, *22*, 7693. <https://doi.org/10.3390/ijms22147693>

Academic Editor: Paola Ghiorzo

Received: 12 June 2021

Accepted: 14 July 2021

Published: 19 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Abstract: With the progress of sequencing technologies, an ever-increasing number of variants of unknown functional and clinical significance (VUS) have been identified in both coding and non-coding regions of the main Breast Cancer (BC) predisposition genes. The aim of this study is to identify a mutational profile of coding and intron-exon junction regions of 12 moderate penetrance genes (*ATM*, *BRIPI1*, *CDH1*, *CHEK2*, *NBN*, *PALB2*, *PTEN*, *RAD50*, *RAD51C*, *RAD51D*, *STK11*, *TP53*) in a cohort of 450 Italian patients with Hereditary Breast/Ovarian Cancer Syndrome, *wild type* for germline mutation in *BRCA1/2* genes. The analysis was extended to 5'UTR and 3'UTR of all the genes listed above and to the *BRCA1* and *BRCA2* known regulatory regions in a subset of 120 patients. The screening was performed through NGS target resequencing on the Illumina platform MiSeq. 8.7% of the patients analyzed is carriers of class 5/4 coding variants in the *ATM* (3.6%), *BRIPI1* (1.6%), *CHEK2* (1.8%), *PALB2* (0.7%), *RAD51C* (0.4%), *RAD51D* (0.4%), and *TP53* (0.2%) genes, while variants of uncertain pathological significance (VUSs)/class 3 were identified in 9.1% of the samples. In intron-exon junctions and in regulatory regions, variants were detected respectively in 5.1% and in 32.5% of the cases analyzed. The average age of disease onset of 44.4 in non-coding variant carriers is absolutely similar to the average age of disease onset in coding variant carriers for each proband's group with the same cancer type. Furthermore, there is not a statistically significant difference in the proportion of cases with a tumor onset under age of 40 between the two groups, but the presence of multiple non-coding variants in the same patient may affect the aggressiveness of the tumor and it is worth underlining that 25% of patients with an aggressive tumor are carriers of a

PTEN 3'UTR-variant. This data provides initial information on how important it might be to extend mutational screening to the regulatory regions in clinical practice.

Keywords: Hereditary breast and ovarian cancer; *BRCA1/2*; breast cancer predisposition genes; coding variants; non-coding variants; regulatory regions; gene panel; NGS

1. Introduction

Breast cancer (BC) is the most common malignant tumor in women representing 25% of the total number of cancer cases in women with an annual increase of 793,700 new cases and 197,600 deaths only in developed countries [1,2].

The presence of germline mutations in the *BRCA1* and *BRCA2* genes significantly increases the risk of developing breast and ovarian cancer [3–5], however only 25% of the risk of familial cancer can be attributed to pathogenetic mutations localized in the coding regions of these two genes. Most of the risk is probably due to a combination of high, moderate, and low risk variants which may also include non-coding variants located in regulatory regions [4].

In clinical practice, during the past years, the mutational analysis in patients with Hereditary Breast and Ovarian Cancer (HBOC) has been mostly limited to the coding regions and to the intron-exon junctions of *BRCA1* and *BRCA2*, precluding the identification of mutations in the non-coding and/or regulatory regions and in other genes that can confer a high or moderate risk to the disease. Under these conditions, mutation screening is negative in 80% of the cases analyzed [6].

Segregation analysis generated no evidence for further dominant genes with a risk penetrance profile comparable to *BRCA1* or *BRCA2* and this has been corroborated by linkage studies [7,8]. Over the years, efforts have been made to identify new intermediate-penetrance predisposition genes, although it is difficult to predict how many they might be and what proportion of the total familial risk is attributable to them. The known intermediate-penetrance genes are all involved in DNA repair or have a role in *BRCA1* or *BRCA2* pathway, so these genes represent plausible candidates and are the most intensively investigated [9]. In addition, in recent years, the interest in investigating the effect that regulatory variants could have on cancer risk has increased [10].

Generally, gene expression is controlled at various levels by key regulatory elements located in promoters, introns, in “long-range enhancers” and in Untranslated Regions (5' and 3' UTR). This is also true for *BRCA1* and *BRCA2*; in fact, it has been shown that the activity of the 5'UTR and of the *BRCA1/2* exon-flanking regions is controlled by a self-regulation mechanism and by an increasing number of transcription factors that act as activators or repressors. Interestingly, germline variants of 5'UTR causing a significant change in promoter activity have been identified in such regions [11–13].

Similarly, germline variants localized in consensus regions for “RNA binding protein” and miRNA located in 3'UTR of *BRCA1* [14,15] and *PTEN* [16,17] have been associated with breast cancer risk.

Therefore, germline regulatory variants can be associated with a deregulation of the transcriptional activity and modify the physiological expression both of *BRCA* genes and of other genes that can affect the risk of breast cancer [13].

With the progression of DNA sequencing technologies, it is possible to analyze in each patient multiple genomic regions by simultaneously screening multiple target genes.

An ever-increasing number of variants of unknown functional and clinical significance (VUS) have been identified in both coding and non-coding regions. Therefore, it is absolutely necessary to establish their risk level in carrier patients, for more effective genetic counselling.

In 2008, the International Agency for Research on Cancer (IARC) proposed to classify the missense variants on the basis of their probable pathogenicity, calculated by multifactorial models, in five risk classes: Class 1 (non-pathogenic or of no clinical significance), Class 2 (likely non-pathogenic or of little clinical significance), Class 3 (uncertain), Class 4 (likely pathogenic), and Class 5 (definitely pathogenic) [18].

Recently, the indications of the American College of Medical Genetics (ACMG) have been increasingly followed by several clinicians and research groups. They have recommended a five-tier classification system where variants can be classified as: 'Pathogenic', 'Likely pathogenic', 'Uncertain significance', 'Likely benign', and 'Benign' [19].

In this work we used this classification that, among those available, is the most complete and is based on an algorithm that takes into account data on penetrance, clinical information of carriers, and scientific evidence.

To classify VUSs, alternative methods have also been developed. They do not directly evaluate the role of the variant in the development of cancer, but the structural and functional modifications of the proteins of interest by providing indirect evidence of the pathogenicity of the variant [20,21]. Both in vitro functional assays and in silico predictors can be used [20].

Recently the international consortium ENIGMA (Evidence-based Network for the Interpretation of Germline Mutant Alleles) has become in charge of reviewing and comparing all the available and useful functional tests in order to define the functional meaning of a variant and to contribute to the interpretation of clinical data [22–25].

In this study involving five Italian hospital centers (Pisa, Cagliari, Varese, Modena, and Bari), a genomic screening was conducted on 450 Breast and/or Ovarian Cancer patients with cancer family history in at least two generations and wild type for germinal mutations in the coding regions and/or for large rearrangements in *BRCA1/2* genes.

The analysis was carried out by Next Generation Sequencing (NGS) using a panel that covers the coding regions and the intron-exon junctions of 10 genes mainly involved in DNA repair pathways: *ATM*, *BRIP1*, *CDH1*, *CHECK2*, *PALB2*, *PTEN*, *RAD51C*, *RAD51D*, *STK11*, *TP53*. For 120 of these patients, the screening was extended to *RAD50* and *NBN* genes and to 5'UTR and 3'UTR of all the genes listed above including the known regulatory regions of *BRCA1* and *BRCA2*. We hypothesized that variants located in the regulatory regions of the main BC predisposition genes can contribute to tumor development in a manner comparable to the variants located in coding regions. To support this hypothesis, we compared the average age of disease onset between the coding and the non-coding variant carriers with the same cancer type. Moreover, we analyzed the proportion of cases with a tumor onset under the age of 40 between the two groups and we verified any variant association in more aggressive tumours.

2. Results

2.1. Summary Overview of Germline Variants Found

Mutational screening was performed for 450 patients in the coding regions and 50 bp flanking intronic sequences of 10 genes: *ATM*, *BRIP1*, *CDH1*, *CHECK2*, *PALB2*, *PTEN*, *RAD51C*, *RAD51D*, *STK11*, *TP53*. For 120 of these patients, the analysis was extended to two additional genes, *RAD50* and *NBN*, and to the regulatory regions of *BRCA1* and *BRCA2*, 5'UTR, and 3'UTR for all 12 genes listed above.

We identified 88 coding variants in 91 patients distributed as follows:

- 16 class 4 (Likely-pathogenic) and 19 class 5 (Pathogenic) variants in 39 patients (8.7%)
- 37 class 3 (VUS) in 41 patients (9.1%)
- 16 Likely-benign variants in 19 patients (4.2%), but all of them potentially classifiable as VUS, according to ACMG classification, because of interpretation conflicts.

In the intron-exon junctions, 13 variants in 23 patients were identified (5.1%).

The extension of the screening in 120 patients to the known regulatory regions of *BRCA1/2* and to 5'UTR, 3'UTR for all the genes analyzed, allowed us to identify another 51 variants in 39 patients (32.5%).

2.2. Coding Variants

A total of 88 coding variants were identified: 35 Pathogenic or Likely-pathogenic variants, 37 VUS, and 16 Likely-benign variants.

Thirty out of 35 Pathogenic or Likely-pathogenic identified coding variants were encoded in dbSNP and 27 were already known in literature or reported in databases such as IARC, LOVD, MGeND, ClinVar Miner with more clinical information. Five variants have never been registered: c.8368delA in *ATM*, c.507delT and c.1430delC in *CHEK2*, c.1565dupC in *PALB2*, c.397_398delCA in *RAD51C*. The 35 class 4 and 5 variants are distributed in the following genes: 14 in *ATM*, 5 in *BRIP1*, 8 in *CHEK2*, 3 in *PALB2*, 2 in *RAD51C*, 2 in *RAD51D*, and 1 in *TP53*.

Thirty-four out of 37 VUS identified were encoded in dbSNP and 25 of these were already known and reported in databases previously cited. The following 3 VUSs were completely new: c.2317A >G in *ATM*, c.3029A >G in *BRIP1* and c.17_18insGCG in *PTEN*. Eighteen of the 37 VUSs identified are in *ATM*, 5 in *BRIP1*, 2 in *CDH1*, 1 in *CHEK2*, 3 in *NBN*, 3 in *PALB2*, 2 in *PTEN*, 2 in *RAD50*, and 1 in *STK11*.

Fifteen out of 16 Likely-benign variants found had a dbSNP entry and 12 of these were encoded in databases. One variant had never been registered (c.252T >C in *RAD51D*). The 16 Likely-benign variants are distributed as follows: 4 in *ATM*, 1 in *BRIP1*, 3 in *CDH1*, 1 in *CHEK2*, 2 in *RAD50*, 1 in *RAD51C*, 3 in *RAD51D*, and 1 in *STK11*.

Figure 1 shows the percentage of distributions of the variants stratified by gene, according to the current ACMG classification. As shown in Figure 1 *ATM* is the gene with the highest mutation rate.

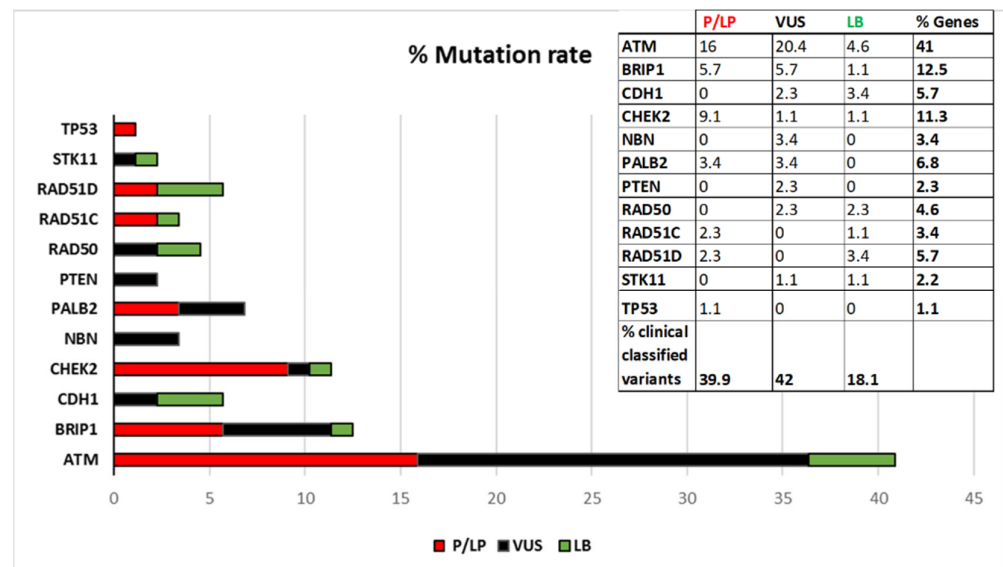


Figure 1. Percentage of identified coding variants distributed by gene and by clinical classification. P/LP = pathogenic/likely pathogenic, VUS = variant of uncertain significance, LB = likely benign.

Sixty-six out of 88 identified variants are missense variants (of which 18 Pathogenic and Likely-pathogenic, 35 VUS and 13 Likely-benign), 5 Nonsense variants (all Pathogenic), 4 Synonymous variants (1 VUS and 3 Likely-benign), 11 frameshift variants (all Pathogenic), and 2 in-frame variants (1VUS and 1 Pathogenic). The distribution by gene is shown in Figure 2.

The predictive analysis on the effect of amino acid change on the function of the protein was conducted with PROVEAN and Polyphen2 software, as described in Materials and Methods. Specifically, the information for each identified variant is shown in Table S1 in Supplementary Materials, where it is also reported whether the variant has already been identified and registered in database by other groups.

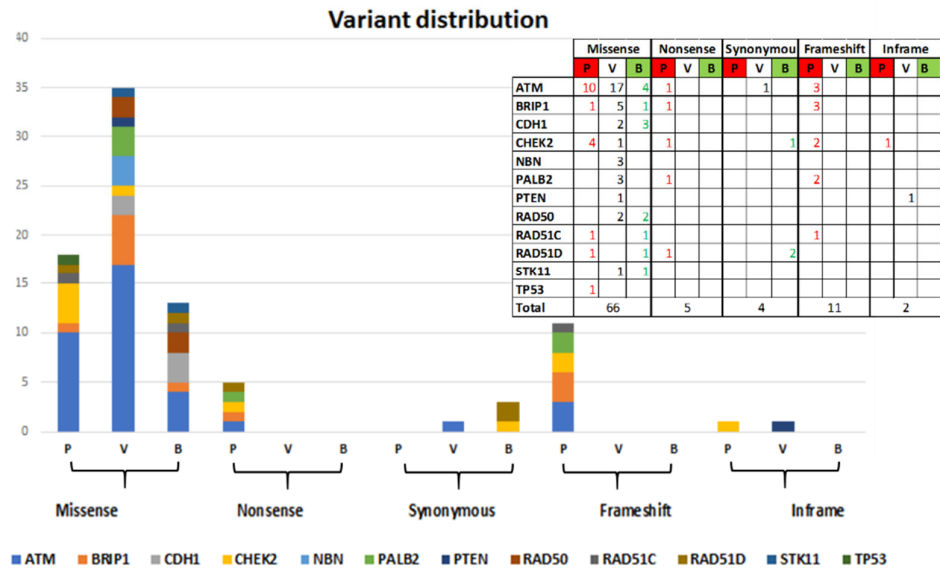


Figure 2. Distribution of 88 variants by effect and by clinical classification (P = Pathogenic/Likely-pathogenic, V = VUS, B = Likely-benign).

2.3. Demographics and Clinical Features of Patients with Coding Variants

The average age of tumor diagnosis for 91 patients with coding variants is 44. 8.8% (8) of the patients have a bilateral BC, 2.2% (2) have both breast and ovarian cancer, 4.4% (4) have ovarian cancer, 2.2% (2) have multiple tumors (in addition to breast cancers) and the remaining percentage, 82.4% (75) of total cases, has a BC. Among the latter patients, 16% (12) have Triple-Negative Breast Cancer (TNBC).

Their family histories are distributed as follows: 48.3% (44) have Hereditary Breast Cancer (HBC), 12.1% (11) have Hereditary Breast and Ovarian Cancer (HBOC), 18.7% (17) have Early Onset Breast Cancer (EOBC), and 20.9% (19) have other Hereditary Breast Cancer Syndromes (HBCS) (Table 1).

Table 1. Coding variant carrier characteristics: percentage distribution of probands according to the cancer they are affected by and to their cancer family history. The number of patients is shown in brackets. The average age at disease onset is reported for each proband kind of cancer (for patients with bilateral breast cancer and multiple tumors, the age of onset of initial tumor has been considered).

Coding Variant Carrier Characteristics		
Patient number	91	
Average age of cancer onset	44	
Proband cancer kind	Average age at onset	
Breast cancer	82.4% (75)	42.8
Bilateral Breast Cancer	8.8% (8)	45.7
Ovarian cancer	4.4% (4)	47.7
Breast and ovarian cancer	2.2% (2)	49.5
Breast/Ovarian Cancer and other tumors	2.2% (2)	54.5

Proband cancer history	
HBC	48.3% (44)
HBOC	12.1% (11)
EOBC	18.7% (17)
Other HBCS	20.9% (19)

Thirty-nine patients are carriers of variants classified as “Pathogenic” or “Likely-pathogenic”, 41 patients are carriers of VUS, while “Likely-benign” variants have been identified in 19 patients. If we consider the total number of patients screened in the coding regions (450) the percentage of class 4 and 5 variant carriers is 8.7% and it is distributed in this way: 3.6% have variants in *ATM*, 1.6% in *BRIP1*, 1.8% in *CHEK2*, 0.7% in *PALB2*, 0.4% in *RAD51C*, 0.4% in *RAD51D*, and 0.2% in *TP53*. In Figure 3 the frequencies of patients with Pathogenic and Likely-pathogenic, VUS and Likely-benign variants are reported.

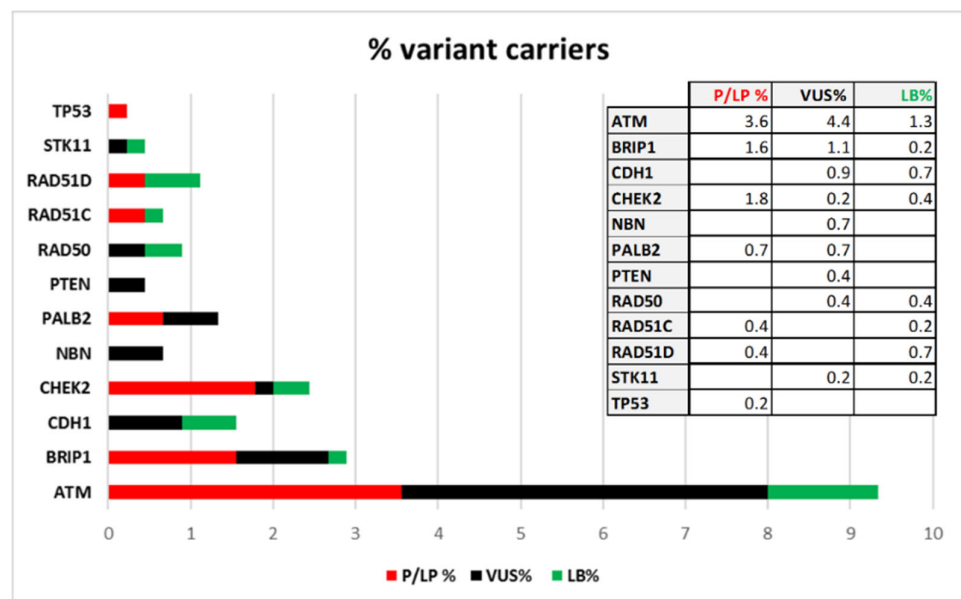


Figure 3. Percentages of coding-variant carriers distributed by gene and by clinical classification out of the total of 450 cases screened.

Among the 75 patients with BC, we identified 82 variants of which 30 Pathogenic and Likely-pathogenic, broken down as follows: 13 in *ATM* gene, 5 in *BRIP1*, 2 in *PALB2*, 6 in *CHEK2*, 2 in *RAD51C*, 1 in *RAD51D* and 1 in *TP53*. In these same patients we identified 38 VUS: 15 in *ATM*, 6 in *BRIP1*, 4 in *CDH1*, 3 in *NBN*, 3 in *PALB2*, 2 in *CHEK2*, 2 in *PTEN*, 1 in *RAD50*, 1 in *RAD51D* and 1 in *STK11*. The 14 remaining variants found are classified as Likely-benign and distributed as follows: 6 in *ATM*, 3 in *RAD51D*, 2 in *CDH1* and 3 respectively in *BRIP1*, *RAD50*, and *RAD51C* genes.

Among the four cases of patients with ovarian cancer, we found a Pathogenic variant in *RAD51C*, 2 VUSs, respectively in *ATM* and *RAD50* genes, and a Likely-benign variant in *STK11*.

Among the 12 cases of multiple tumors in coding variant carriers (9 cases of bilateral breast cancer, 2 of breast and ovarian cancer, and 1 case of breast tumor associated with pancreatic cancer), 9 have Pathogenic or Likely-pathogenic variants, 1 VUS and 2 Likely-benign variants. The statistical analysis has found a significant association of these cases with Pathogenic variant carriers (χ^2 -test, $p < 0.01$) in comparison with the cases of single tumors.

Three out of nine patients with Pathogenic variants have variants in *ATM*, 4 in *CHEK2*, 1 in *BRIP1* and 1 in *PALB2* (Table 2). Three of these nine variants are new variants

never reported in dbSNP. One case has a VUS in *ATM* gene and two patients with multiple tumors have Likely-benign variants in *CDH1* and *CHEK2* genes. Frequency of coding variants in *CHEK2* gene is significant in these patients (Fisher test, $p < 0.05$).

Table 2. List of variants found among the 12 cases of patients with multiple tumors in coding-variant carriers. The table reports the protein change of variant, the clinical classification (P = Pathogenic, LP = Likely-pathogenic, VUS = Variant of Unknown Clinical Significance, and LB = Likely-benign), the variant frequency among these 12 cases, the patient cancer type in which they were detected and the eventual database registration. The variants are reported in HGVS nomenclature.

VARIANTS IN 12 CASES OF MULTIPLE TUMORS							
Variant	Protein Change	Gene	Classification	n.patients	Tumor	dbSNP	Database
c.3894dupT	p.Ala1299CysfsTer3	ATM	P	1	Bil. Breast	rs587781823	ClinVar Miner
c.8368delA	p.Arg2790AspfsTer16	ATM	P	1	Bil. Breast		
c.3284G > A	p.Arg1095Lys	ATM	LP	1	Br/Ov	rs587781815	LOVD
c.2237_2240delTCAA	p.Ile746AsnfsTer12	BRIP1	P	1	Bil.Breast	rs587782726	MGeND
c.1312G > T	p.Val438Phe	CHEK2	LP	1	Bil.Breast	rs58778017	LOVD
c.483_485delAGA	p.Glu204del	CHEK2	LP	1	Bil.Breast/Panocr.	rs587782008	MGeND
c.507delT	p.Asp169GlufsTer5	CHEK2	LP	1	Br/Ov		
c.1430delC	p.Thr477LysfsTer7	CHEK2	LP	1	Bil.Breast		
c.1451T > A	p.Leu484Ter	PALB2	P	1	Bil.Breast	rs786203714	ClinVar Miner
c.1516G > T	p.Gly506Cys	ATM	VUS	1	Bil.Breast	rs587779816	LOVD
c.131G > T	p.Arg44Leu	CDH1	LB	1	Bil.Breast	rs1375178645	
c.132C > T	p.Ser44=	CHEK2	LB	1	Bil.Breast	rs199715101	ClinVar Miner

By analyzing cases of more aggressive tumors such as the 12 Triple-negative breast cancer patients (TNBC) (Table 3), all with G2 or G3 infiltrating ductal carcinoma, we identified 6 Pathogenic and Likely-pathogenic variants: 1 in *ATM*, 1 in *BRIP1*, 2 in *CHEK2*, 1 in *RAD51D* and 1 in *TP53*. In 6 patients, 4 VUSs (3 in *ATM* and 1 in *PTEN* never recorded before) and 2 Likely-benign variants were identified. The latter all have a conserved evolutionary profile (GERP > 3.8) and a conflict of interpretation according to Varsome and ClinVar.

Table 3. List of variants identified in 12 cases of patients with Triple-Negative Breast cancer (TNBC). The table also reports the protein change for each variant, their clinical classification (P = Pathogenic, LP = Likely-pathogenic, VUS = Variant of Unknown clinical Significance, LB = Likely-benign), their frequency identified in these patients and whether they have been previously recorded in the database LOVD. The variants are reported in HGVS nomenclature.

VARIANTS IN 12 CASES OF TNBC						
Variant	Protein Change	Gene	Classification	n.patients	dbSNP	Database
c.6067G > A	p.Gly2023Arg	ATM	P	1	rs11212587	LOVD
c.2392C > T	p.Arg798Ter	BRIP1	P	1	rs137852986	LOVD
c.409C > T	p.Arg137Ter	CHEK2	P	1	rs730881701	LOVD
c.1118A > G	p.Gln373Arg	CHEK2	LP	1	rs894075046	LOVD
c.694C > T	p.Arg232Ter	RAD51 D	P	1	rs587780104	LOVD
c.817C > G	p.Arg273Gly	TP53	P	1	rs121913343	LOVD
c.1272T > C	p.Pro424=	ATM	VUS	1	rs35578748	LOVD
c.4703A > G	p.His1568Arg	ATM	VUS	1	rs368830730	LOVD
c.4060C > A	p.Pro1354Thr	ATM	VUS	1	rs145119475	LOVD
c.17_18insGC G	p.Lys6_Glu7insAr g	PTEN	VUS	1		
c.2119T > C	p.Ser707Pro	ATM	LB	1	rs4986761	LOVD
c.1810C > T	p.Pro604Ser	ATM	LB	1	rs2227922	LOVD

2.4. Non-Coding and/or Regulatory Variants

The screening of non-coding regions has allowed us to identify 64 variants. Forty-six of these have a dbSNP code and only 24 are known variants already registered in databases by other groups.

The 64 variants are distributed as follows: 18 are located at 5'UTR, 21 at 3'UTR, 6 in splice site regions, and 19 in intron regions.

Figure 4 shows the distribution of the variants by gene divided according to their location.

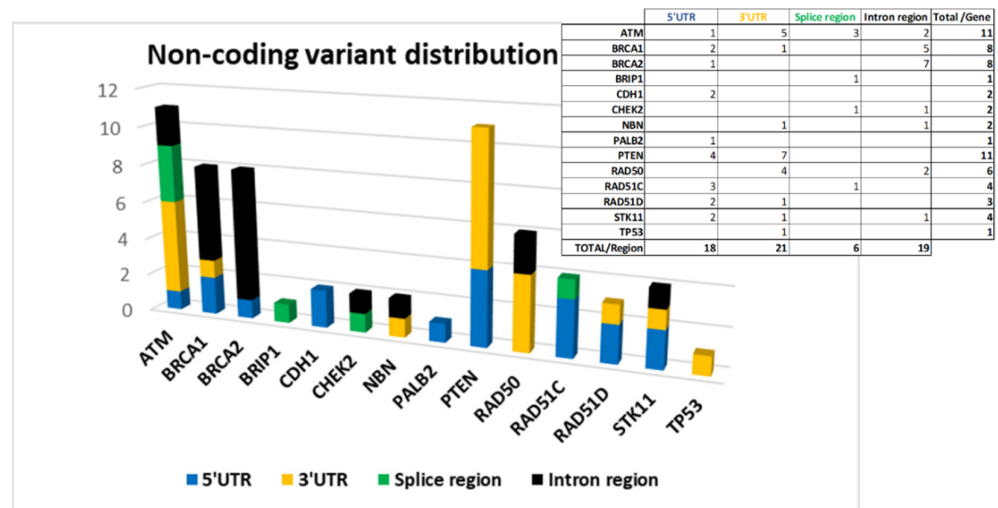


Figure 4. Non-coding variant distribution by gene and by gene region (5'UTR, 3'UTR, Splice regions, and Intron regions).

ATM, *BRCA1*, *BRCA2*, and *PTEN* are among the genes in which we have found more non-coding variants (Figure 4).

The variant putative effect on gene expression, was carried out by using different software (described in Materials and Methods) depending on variant position. The results are reported in Table S2a–d in Supplementary Materials.

An unexpectedly high number of variants in 5' UTR of *PTEN* (4) and *RAD51C* (3) was found and, according to our analysis, all of them are localized in TF binding regions (Table S2a/Supplementary Materials).

The in-silico analysis performed by 'TFbind' and 'Promo' software, has identified TF consensus sequences in the regions where the 5'UTR of *BRCA1* (c.-676T > A) and *BRCA2* (c.-1193C > T) variants are localized (Table S2a/Supplementary Materials).

In 3'UTR, the greatest contribution is due to variants located in *ATM*, *PTEN*, and *RAD50* genes (Figure 4) on which putative miRNA-binding regions have been identified by 'TargetScanHuman' and 'miRDB' software (Table S2b/Supplementary Materials).

In splicing regions, we identified six variants and only one, c.904+5G > T, was predicted to inactivate the wild-type donor site in the intron 6 of the *RAD51C* gene, and therefore is classified by the ACMG criteria as 'Pathogenic'.

Three variants were in *ATM* but only the c.7788+ 8G > T, localized in intron 52, was predicted by 'Human Splicing Finder' as a possible modifier of an exonic ESE site with a potential alteration of splicing. The remaining 2 in *CHEK2* and *BRIP1* genes were predicted neutral on splicing (Table S2c/Supplementary materials).

In the remaining intronic regions, variants in *ATM*, *BRCA1*, *BRCA2*, *CHEK2*, *NBN*, *RAD50*, *STK11*, and *TP53* genes were identified. The possible presence of enhancer regions, compatible with bonds to transcription factors, was analyzed by the software mentioned above. The four variants identified in intron 2 of *BRCA1* (c.81-3510_81-3505delCTTTT; c.81-3790G > C; c.81-4100A > C; c.80 + 247C > T) deserve attention. In fact, all four variants identified in this study map in two conserved enhancer regions, that have already been demonstrated to be involved in gene expression regulation (Table S2d/Supplementary Materials) [26].

In conclusion, we found only one Pathogenic variant in the *RAD51C* gene, whereas the large majority were classified as VUSs (50), 2 as Likely-benign and 11 as Benign variants.

The high number of VUSs is due to the fact that very few studies have analyzed these non-coding regions, and for this reason the classification is based only on in silico prediction.

2.5. Deep Intronic Regions

The panel of 9071 probes used was designed to optimize coverage of exonic regions, 50 bp flanking each intron-exon junction, 5'UTR, 3'UTR of all the genes of interest and the known regulatory regions of *BRCA1* and *BRCA2*. However, the position of the probes managed to cover additional regions which led us to identify another nine variants in intronic regions (all with a coverage $\geq 50X$) that proved to be interesting for the predictions obtained by 'TFbind' and 'Promo' software and for the information obtained from rVarBase and ENCODE (Table S2e/Supplementary Materials). Two variants are in the intron 61 of *ATM*, one in the intron 16 of *BRIP1*, two in intron 2 respectively of *RAD51C* and *RAD51D* genes and four variants are in the intron 1 of *TP53*. Eight out of nine variants are classified as VUS and four are not registered in dbSNP.

2.6. Demographics and Clinical Features of Patients with Non-Coding Variants

Non-coding variants were found in 64 patients (excluding two cases of co-occurrence of Pathogenic coding variants). According to this data, in these patients there is a significant increase in both ovarian cancer (χ^2 -test, $p < 0.05$) and HBOC families (χ^2 -test, $p < 0.01$), compared to patients with variants in coding regions (Table 1/Table 4).

The mean age of tumor onset for these patients is 44.4. If we compare the coding and non-coding variant carriers in each proband group with the same kind of cancer, the average age of disease onset is completely comparable (Table 1/Table 4). In particular, if we compare the 64 non-coding variant carriers to the 39 Pathogenic coding variant carriers, there is not a statistically significant difference in the proportion of cases with a tumor onset under age 40 (Table 5, χ^2 -test, $p > 0.05$).

Table 4. Non-coding variant carrier characteristics: percentage distribution of probands according to cancer type and cancer family history. The number of patients is shown in brackets. The average age at disease onset is reported for each proband cancer type (for patients with bilateral breast cancer and multiple tumors, the age of onset of initial tumor has been considered).

Non-coding variant carrier characteristics		
Patient number	64	
Average age of cancer onset	44.4	
Proband cancer kind	Average age at onset	
Breast cancer	67.2% (43)	43.9
Bilateral Breast Cancer	10.9% (7)	43.1
Ovarian cancer	18.7% (12)	44
Breast and ovarian cancer	1.6% (1)	48
Breast/Ovarian Cancer and other tumors	1.6% (1)	49
Proband cancer history		
HBC	53.1% (34)	
HBOC	29.7% (19)	
EOBC	4.7% (3)	
Other HBCS	12.5% (8)	

Table 5. Age of tumor onset: comparison between Pathogenic coding variant carriers and non-coding variant carriers.

	Pathogenic Coding Variant Carriers	Non-Coding Variant Carriers
≤40	15	22
>40	24	42
Total	39	64

Figure 5 shows the distribution by gene and by gene location of the variants identified respectively in 43 patients with BC, in 12 patients with ovarian cancer, and in 9 cases of patients with multiple tumors (seven with bilateral cancer, one with breast and ovarian cancer, and one with breast and pancreatic cancer). From this data emerges the high frequency of variants identified in the *ATM* gene, as well as in *BRCA1* and *BRCA2*, in patients with BC and the high contribution of variants localized in the 3'UTR of *PTEN* in patients with multiple tumors. In particular, three out of nine variants found in these patients are localized in *PTEN* 3'UTR and the bioinformatics predictions found, for all three, consensus sequences compatible with miRNA binding sites (Figure 5, Table S2b/Supplementary Materials). In patients with ovarian cancer, instead, variants were identified mainly in the *RAD50* and *RAD51C* genes.

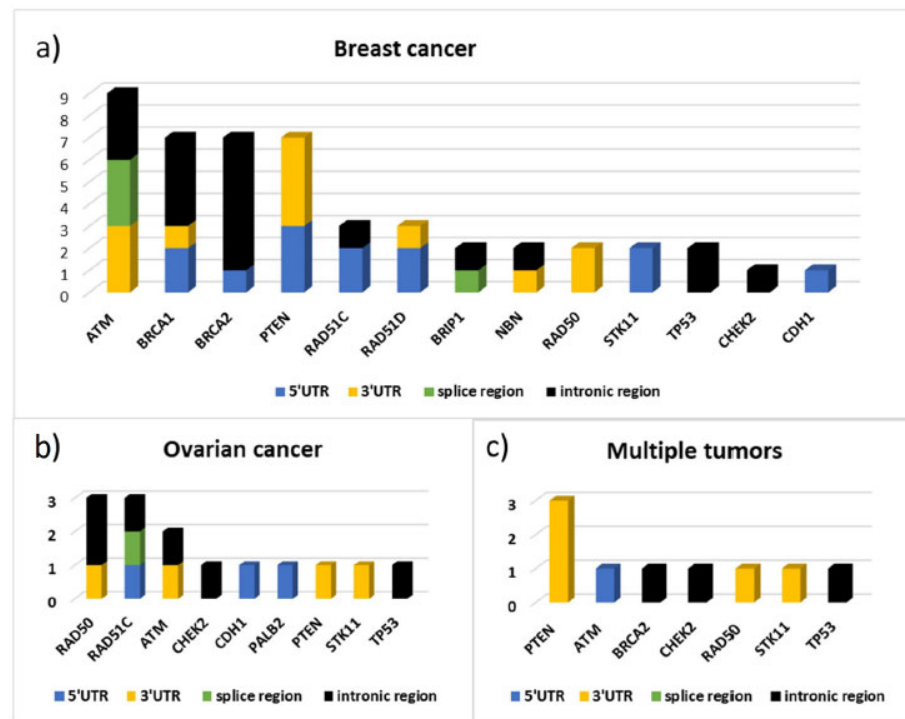


Figure 5. Variant distribution by gene and by gene region (5'UTR, 3'UTR, Splice regions and Intron regions) respectively in patients with breast cancer (a), ovarian cancer (b), and with multiple tumors (c).

Among the 12 cases of TNBC, all with G2 or G3 infiltrating ductal carcinoma, 4 variants out of 19 identified are on the *PTEN* gene: one located at 5'UTR and the others at 3'UTR. Other interesting variants, for the type of feedback we obtained through prediction analysis, are the variants in the 5'UTR of *RAD51D* and *STK11* genes. The complete list of variants found in these patients is shown in Table 6. Four variants are in intronic regions of *BRCA2*. Furthermore, among all the non-coding variants filtered in this work, the only variant in the *RAD51C* gene classified as "Pathogenic" by ACMG criteria was found in one of these patients.

Twelve patients with multiple variants in non-coding regions were filtered to assess whether the presence of multiple mutations could affect tumor pathogenicity. Ten out of 12 filtered patients have BC and 6 of these have TNBC. According to our data, there is a statistically significant association between TNBC and multiple variants in non-coding regions (χ^2 -test, $p < 0.05$). In fact, 50% of BC patients with multiple variants in non-coding regions have TNBC, while, among the 52 BC non-coding variants carriers, only 12 (23.1%) have TNBC (χ^2 -test, $p < 0.05$; Table 7).

Table 6. Variants identified in the 12 patients with TNBC among the non-coding-variant carriers. The table shows the variant position in the gene (ss = splice site), the clinical classification (P = Pathogenic, VUS = Variant of Unknown clinical Significance, B = Benign), the frequency of each variant and any database registration. The variants are reported in HGVS nomenclature and those highlighted with the same color belong to the same patient.

VARIANTS IN 12 CASES OF TNBC						
Variant	Gene	Position	Classification	Patient	dbSNP	Database
c.-323C > T	PTEN	5'UTR	VUS	Varese 36		
c.-116C > T	RAD51 D	5'UTR	VUS	Varese 36		
c.*1122_*1123insT	PTEN	3'UTR	VUS	Varese 36		
c.9257-3610G > A	BRCA2	Intron 24	VUS	Varese 36	rs170775191	BRCAexchange
c.-239C > T	RAD51 D	5'UTR	LB	Varese 55	rs550346401	
c.681+697C > T	BRCA2	Intron 8	VUS	Varese 55	rs11571632	BRCAexchange
c.-636C > T	STK11	5'UTR	VUS	Pisa 1489	rs981125953	
c.*1674 > G	RAD50	3'UTR	VUS	Pisa 1489		
c.*26A > G	ATM	3'UTR	VUS	Pisa 2348	rs1247328981	
c.*724dupT	NBN	3'UTR	VUS	Varese 10	rs564878448	
c.*4606_*4609delGCT T	PTEN	3'UTR	B	Bari 1		
c.*1152C > T	PTEN	3'UTR	B	Varese 17	rs552354954	
c.7788+8G > T	ATM	Intron 52ss	VUS	Bari 4	rs112775908	LOVD
c.317-1021A > G	BRCA2	Intron 3	VUS	Bari 4	rs11571604	BRCAexchange
c.8850+389G > A	ATM	Intron 61	LB	Varese 14		
c.317-512A > T	BRCA2	Intron 3	VUS	Varese 14		
c.320-5T > C	CHEK2	Intron 24ss	VUS	Pisa 2280		LOVD
c.904+5G > T	RAD51 C	Intron 6ss	LP	Pisa 1564	rs587782702	LOVD
c.-29+475C > A	TP53	Intron 1	VUS	Varese 20		

Table 7. Patients with multiple non-coding variants, age at disease diagnosis and tumor characteristics. The table also reports the cases of co-occurrence of VUS and Likely-benign coding variants with multiple non-coding variants in the same patient. The variants are reported in HGVS nomenclature.

Patients	Non-Coding Variants	Coding Variants	Cancer	Age at Diagnosis	Histological Characteristics	TNBC
Pisa 710	c.-1193C > T BRCA2; c.-408T > A BRCA1		Breast	57	infiltrating ductal carcinoma	
Pisa 1489	c.*1674A > G RAD50; c.-636C > T STK11		Breast	50	infiltrating ductal carcinoma	V
Bari 4	c.7788+8G > T ATM, c.317-1021A > G BRCA1	c.2074G > A CDH1, c.17_18insGCG PTEN	Breast	37	infiltrating ductal carcinoma	V
Varese 7	c.-138G > C PALB2, c.-29+622G > C TP53	c.1094G > A RAD50	Ovary	48	serous carcinoma	
Varese 8	c.81-4100A > C BRCA1, c.*114A > G RAD50	c.4060C > A ATM	Breast	45	infiltrating ductal carcinoma	
Varese 14	c.8850+389G > A ATM, c.317-512A > T BRCA2		Breast	41	infiltrating ductal carcinoma	V
Varese 36	c.9257-3610G>A BRCA2, c.-323C > T PTEN, c.*1122_*1123insT PTEN, c.-116C > T RAD51D		Breast	44	infiltrating ductal carcinoma	V
Varese 43	c.560+19_560+22delTACT ATM, c.904+5G > T RAD51C		Ovary	49	serous carcinoma	
Varese 45	c.*2842A > G ATM, c.*750A > G BRCA1, c.*106G > A RAD51D	c.1810C > T ATM, c.6226A > G ATM	Breast	46	infiltrating ductal carcinoma	
Varese 51	c.8850+689A > C ATM, c.*537A > C ATM		Breast	34	infiltrating ductal carcinoma	
Varese 55	c.681+697C > T BRCA2, c.-239C > T RAD51D	c.1810C > T ATM	Breast	45	infiltrating ductal carcinoma	V
Pisa 2582	c.-24A>C ATM, c.*261C>T STK11		Bil.Breast	46	infiltrating ductal carcinoma	

3. Discussion

This study regards the analysis of coding regions and 50 bp flanking intron–exon junction of 10 BC predisposition genes in a cohort of 450 HBOC patients. For a portion of these patients (120) the screening was extended to additional two genes (*NBN* and *RAD50*), to the 5'UTR and 3'UTR of each gene analyzed and to the known regulatory regions of *BRCA1* and *BRCA2*. Eighty eight coding variants were identified in 91/450 patients (20.2%): 35 of these were classified as pathogenic (class 5) or likely-pathogenic (class 4) and found in 39 patients (8.7%), 37 were classified as VUSs in 41 patients (9.1%) and 16 as “Likely-benign” in 19 patients (4.2%).

ATM was the more mutated gene (16 carriers), followed by *CHEK2* (8 carriers), *BRIP1* (7 carriers), *PALB2* (3 carriers). *RAD51C* and *RAD51D* were found mutated in 2 patients/each and *TP53* in one patient only.

Thirteen variants in the intron-exon junctions were identified in 23/450 patients (5.1%) and 51 non-coding variants in 39 out of 120 patients analyzed for this purpose (32.5%).

Our results support what was recently published by the Breast Cancer Association Consortium in a larger and more geographically heterogeneous population consisting of 60,466 women. With a panel of 34 putative susceptibility genes, a significant association between the increased risk of breast cancer and truncating mutations in *ATM*, *CHEK2*, *PALB2*, *BARD1*, *RAD51C*, *RAD51D*, and *TP53* was identified [27]. Similar results were found through a screening of 65,057 white woman with breast cancer where a high or moderately increased risk of disease was found in relation with pathogenic mutations in *ATM*, *CHEK2*, *PALB2*, and *RAD51D* [28].

ATM, *CHEK2*, *PALB2*, and *TP53* were confirmed as breast cancer predisposition genes also by the German Consortium for Hereditary Breast and Ovarian Cancer, through a screening of eight cancer predisposition genes conducted on 5589 breast cancer patients negative for pathogenic *BRCA1/2* mutations [29]. An American retrospective study performed on 35,000 women who underwent genetic testing through a 25-gene hereditary cancer panel found the same correlation [30].

The analysis of non-coding regions reports a consistent number of variants never registered previously (18/64 variants).

In details, in the 5'UTR 2 variants were found, one in *BRCA1* (c.-676t > a) and one in *BRCA2* (c.-1193c > t). Both were registered in BRCA-exchange and LOVD database as VUSs. The in-silico analysis, conducted by ‘TFbind’ and ‘Promo Prediction’, has identified putative consensus sequences for Elk-1, NF-AT1, STAT4 transcription factors, for the first variant, and NRF2, GATA1/2, USF for the second.

Whereas the 5'UTR and 3'UTR of the *BRCA1/2* genes have been two of the most intensely investigated non-coding regions over the last few years [12,13,31–34], this is the first fully comprehensive study of 5' and 3'UTRs of so called “minor genes for breast/ovarian cancer susceptibility” and a number of variants were also identified in the 5'UTRs: 4 in the *PTEN* gene and 3 in the *RAD51C* gene and the in-silico predictions obtained regarding their involvement in TF binding are interesting (Table S2a in Supplementary Materials).

In the 3'UTRs of genes other than *BRCA1/2*, the greatest contribution was due to variants located in the *ATM* (5), *PTEN* (7) and *RAD50* (4) genes. The use of software such as ‘TargetScanHuman’ and ‘miRDB’ has made it possible to identify in those regions binding sites for miRNAs (Table S2b in Supplementary Materials).

The variants localized in splice regions mainly affect the *ATM* (3) gene and, in particular the c.7788+8G > T variant, localized in intron 52 of *ATM*, could modify an exonic ESE site (Table S2c in Supplementary Materials). Another variant of interest is the c.904+5G > T of *RAD51C*, classified by ACMG criteria as ‘Pathogenic’.

According to our data, variants in the intronic regions have been identified mainly in the *BRCA1* (5), *BRCA2* (7), *ATM* (2), and *RAD50* (2) genes.

Four out of 5 *BRCA1* variants are localized in intron 2 (c.81-3510_81-3505delCTTTTT; c.81-3790G > C; c.81-4100A > C; c.80+247C > T) and the in-silico predictive analysis led us to identify TF binding sites in those regions (Table S2d/Supplementary Materials). Those variants deserve to be studied in depth through a functional assay because in *BRCA1* intron 2, two conserved enhancer regions (CNS1 and CNS2) are localized and involved in the regulation of gene expression [26].

Eleven out of 12 intronic variants found in *BRCA1/2* genes could be classified as 'deep intronic'. The deep intronic regions of the BRCA genes are intensely been studied in the recent years. These investigations have led to the identification of variants capable to modify the expression level or the functionality of the BRCA1 protein [35].

Finally, we incidentally found two new variants in deep intronic regions of the *ATM* and *TP53* genes in two patients with TNBC (due to the fact that additional regions that were not of our interest were included in the panel design). The variants are respectively in intron 61 of *ATM* (c.8850+389G > A) and intron 1 of *TP53* (c.-29+475C > A) and are both predicted as TF binding regions. This result underlines how important is to broaden screening to deeper intronic regions (not only in BRCA genes).

By analyzing the characteristics of patients for the type of variants identified, distinguishing between coding variant carriers versus non-coding variant carriers, we can make some interesting considerations.

The average age of breast cancer onset in coding variants carriers is 44 and most of them have breast cancer only family history as shown in Table 1.

A possible association of *ATM-BRIP1-PALB2* and a significant association of *CHEK2* pathogenic/likely-pathogenic coding variants with multiple tumors, including bilateral breast cancer, was observed. These data are in agreement with the findings of German Consortium for Hereditary Breast and Ovarian Cancer, that demonstrated a significant association of *CHEK2* deleterious variants with bilateral breast cancer [29].

Among the 12 cases of TNBC patients, 6 were carriers of pathogenic coding variants: 2 in *CHEK2* and the remaining 4 in the *ATM*, *BRIP1*, *RAD51D* and *TP53* genes respectively (Table 3).

Studies conducted on a larger number of TNBC patients have come to concordant results regarding the association of these cases with pathogenic variants in the *PALB2*, *BRIP1*, *RAD51C*, and *RAD51D* genes [30,36–38]. Our results partially agree with these data, having identified pathogenic variants in TNBC patients with *ATM*, *RAD51D*, *BRIP1*, and *TP53* variants. It is likely that the analysis of a higher number of TNBC cases could have led us to a greater degree of concordance.

However, it is certainly very interesting that 50% of the TNBC cases analyzed had pathogenic coding variants: it is a higher proportion compared to that obtained in other cohorts of TNBC patients filtered in the population solely on the basis of their tumor phenotype [36,37]. It is probable that the higher percentage registered in our cases depends on the double filtering, based not only on patient tumor phenotype but also on the patient breast cancer family history.

It is known that some TNBCs BRCAWT show the so-called 'BRCAness' phenotype, exhibiting clinical and pathological properties similar to BRCA-mutated tumors such as the homologous recombination deficiency and the sensitivity to DNA damaging agents [39,40]. Therefore, our data could be of interest at a therapeutic level. In fact, the data supports what we have already reported in Spugnese et al. [38] where germline mutations in DNA repair genes, in *BRCA1/2* WT tumors, were associated with a group of TNBC patients who responded to anthracyclines/taxanes neoadjuvant therapy. It is becoming increasingly evident that genes involved in DNA repair could become response biomarkers for DNA-damaging therapy in these patients [41].

Although we cannot make any consideration about the variants identified in ovarian cancer, having only analyzed four patients with ovarian cancer and two with breast and ovarian cancer, the detection of pathogenic variants in *RAD51C* and *ATM* is in agreement with what is reported in a case-control study performed on 2051 women with ovarian

cancer. The authors of this paper identified an increased risk for ovarian cancer associated with *MSH6*, *RAD51C*, *TP53*, and *ATM* [42].

Analyzing the characteristics of non-coding variants carriers, we recorded an average age of disease onset of 44.4 absolutely similar to the average age of disease onset in coding variant carriers for each proband's group with the same cancer type. Furthermore, there is not a statistically significant difference in the proportion of cases with a tumor onset under the age of 40 between the two groups (χ^2 -test, $p > 0.05$; Table 5).

A statistically significant increase in cases of ovarian cancer and of HBOC families in non-coding variant carriers compared to coding variant carriers was found (Table 1/Table 4). This is an interesting observation but very preliminary considering the limited number of samples analyzed, and it needs to be further investigated.

One of the results that emerges clearly when analyzing the characteristics of non-coding variant carriers, is the relationship between variants in 3'UTR of *PTEN* and multiple tumors or TNBC tumor phenotype. This result is in agreement with what has already been described in previous studies in which the deregulation of *PTEN* expression, via binding sites for RNA-binding-protein and miRNA located in 3'UTR, was associated with aggressive phenotype and poor outcome for BC patients [16,17]. Moreover, four out of seven *BRCA2* deep intronic variants identified (c.9257-3610G > A, c.681+697C > T, c.317-1021A > G, c.317-512A > T) have been found in TNBC patients.

Finally, by analyzing the 12 patients carrying multiple non-coding variants we observed a statistically significant association with a TNBC phenotype (χ^2 -test, $p < 0.05$).

This work has the merit of having highlighted certain variants never recorded in the coding and regulatory regions which are worthy of further study. It also underlines the importance, increasingly evident in recent years, of also screening the regulatory regions in patients with clearly familial tumors with no relevant variation in the coding sequences of the main breast/ovarian cancer predisposition genes. Certainly, in the future it will be necessary to perform a segregation analysis in order to understand the real involvement of certain genomic regions in breast and ovarian cancer predisposition. This work surely could help to pave the way for other in-depth projects on the functional meaning of the variants identified, by providing, thanks to the many notes obtained, a lot of useful information for a better understanding of their clinical significance.

4. Materials and Methods

4.1. Selection of Patients Included in the Study

We selected a total of 450 patients eligible for this study who underwent genetic counseling between 2001 and 2019 at five different Italian centers distributed as follows:

- 163 patients from University Hospital of Pisa
- 210 patients from Azienda Tutela Salute Sardegna, ASST of Cagliari
- 55 patients from Ospedale di Circolo, ASST-Sette Laghi of Varese
- 13 patients from University Hospital of Modena
- 9 patients from IRCCS Istituto Tumori "Giovanni Paolo II" of Bari

All patients signed an informed consent. The eligibility criteria included:

- family history of BC in at least two generations: (a) two or more BC cases in first degree relatives, one of which diagnosed before age 40 (b) one ovarian cancer case associated with a BC in a first degree relative at any age;
- a priori risk of being carriers of *BRCA1/2* mutation >10% calculated with BRCAPro (<http://www4.utsouthwestern.edu/breast-health/cagene>, accessed on March 2017-December 2019) and BOADICEA (<https://pluto.srl.cam.ac.uk/cgi-bin/BD2/v2/bd.cgi>, accessed on March 2017-December 2019) software;
- to be wild type for point mutations and large rearrangements in the *BRCA1/2* coding regions.

4.2. Library Preparation and NGS Sequencing

DNA was extracted from blood samples. Genomic DNA screening was carried out with a NGS target resequencing approach on the Illumina MiSeq machine. 450 patients were analyzed with a custom panel optimized to sequence the coding regions and 50 bp flanking each intron-exon junction of 10 genes involved in breast and ovarian cancer predisposition: *ATM* (NM_000051.3), *BRIP1* (NM_032043.2), *CDH1* (NM_004360.3), *CHECK2* (NM_007194.3), *PALB2* (NM_024675.3), *PTEN* (NM_000314.6), *RAD51C* (NM_058216.2), *RAD51D* (NM_002878.3), *STK11* (NM_000455.4), *TP53* (NM_000546.5). For a subset of 120 patients (43 from Pisa, 55 from Varese, 13 from Modena and 9 from Bari), the screening was also extended to *NBN* (NM_002485.5) and *RAD50* (NM_005732) genes, including 5'UTR and 3'UTR of each gene and the known regulatory regions of *BRCA1* (NM_007294.4) and *BRCA2* (NM_000059) [12–15,26]. To amplify the target sequences, a panel of 9071 probes was designed using SureSelect DNA Advanced Design Wizard in order to ensure the total coverage of 178.155 kb corresponding to 99% of the entire region of interest (Supplementary Materials Table S2a,b). Library preparation was performed according to SureSelectQXT Target Enrichment for Illumina Multiplexed Sequencing protocol (SureSelectQXT Reagent Kit, Agilent Technologies, Santa Clara, CA, USA). To assess library DNA quantity and quality after fragmentation and adaptor-tag addition phase and after amplification of the indexed library, we used the Agilent 2100 Bioanalyzer. Following purification and quantification of pooled libraries, we sequenced them by 2 × 100 bp paired-end reads on a Miseq (Illumina, San Diego, CA, USA) instrument. Sequence reads allocated to each individual were aligned to the human reference sequence (hg19) using Agilent Sure Call (ver.4.1.1). All the regions considered for the analysis had a coverage ≥50X and a Quality Score (QS) ≥30. Sanger sequencing was performed to confirm the identified variants in lymphocyte DNA.

4.3. Bioinformatic Annotation and Prioritization of the Identified Variants

Each identified variant was evaluated in terms of Minor Allele Frequency (MAF): reference allelic frequency is that of total population reported in 1000 Genomes (<http://www.internationalgenome.org>, accessed on January 2020) and in gnomAD (<https://gnomad.broadinstitute.org>, accessed on January 2020). All variants with a MAF >1% were excluded from further analysis as they were considered common polymorphisms. To speed up the reading of the variants and their filtering on the basis of the known annotations, we used BaseSpace Variant Interpreter platform (<https://variantinterpreter.informatics.illumina.com>, accessed on January 2020). All variants were also evaluated in term of evolutionary conservation profile (GERP score) [43]. The clinical significance of each variant was annotated considering information from ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar>, accessed on August 2020) [44] and according to ACMG (American College of Medical Genetics) criteria [19,45] also used by Varsome (<https://varsome.com>, accessed on August 2020) [46]. The latter helped us mostly to summarize all the known information and the pathogenicity predictions of the identified variants.

For the coding variant, we annotated the possible impact of an amino acid substitution on the protein structure and function obtained by PolyPhen2 (<http://genetics.bwh.harvard.edu/pph2>, accessed on May 2020) and PROVEAN (<http://provean.jcvi.org/index.php>, accessed on May 2020) software [47,48]. For the non-coding variants, in order to obtain information about chromatin states of the surrounding regions, methylation and acetylation status and hypersensitivity to DNaseI, we used rVarBase (<http://rv.psych.ac.cn>, accessed on February 2020) and ENCODE (<https://www.encodeproject.org>, accessed on February 2020) database [49–51].

To predict possible TF binding site in 5'UTR and in intronic enhancer regions, we employed three different software: Promo (http://alggen.lsi.upc.es/cgi-bin/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3, accessed on February 2020) [52],

Tfbind (<http://tfbind.hgc.jp/>) and Lasagna-Search2.0 (https://biogrid-lasagna.engr.uconn.edu/lasagna_search/index.php, accessed on February 2020) [53]. The presence of miRNA binding site in 3'UTR variants, was evaluated by TargetScanHuman (http://www.targetscan.org/vert_72, accessed on February 2020) and miRDB (<http://www.mirdb.org>, accessed on February 2020) [54,55]. Finally, in order to identify consensus regions for potential splice sites, we made use of Human Splicing Finder (<http://www.umd.be/HSF>, accessed on February 2020) [56]. All the programs used to obtain predictions about the variants found are free software and they are summarized in Table 8.

All this information allowed us to have more complete annotations of the variants, making it possible to suggest future functional tests, especially those never reported in literature or in databases such as ENIGMA (<https://enigmaconsortium.org>, accessed on December 2020), BRCA Exchange (<https://brcaexchange.org>, accessed on December 2020), LOVD (<https://www.lovd.nl>, accessed on December 2020), MgeND (<https://mgend.med.kyoto-u.ac.jp>, accessed on December 2020), and ClinVar Miner (<https://clinvarminer.genetics.utah.edu>, accessed on December 2020). For each variant, the dbSNP (<https://www.ncbi.nlm.nih.gov/snp>, accessed on January 2021) encoding was also checked.

Table 8. Tools used for variant analysis, annotations and prioritization.

Coding Variant Annotations	Non-Coding Variant Annotations
Variant information from human genomic data (Variant Interpreter, Varsome)	Variant information from human genomic data (Variant Interpreter, Varsome)
Relationships among human variants and phenotypes, clinical significance (ClinVar, ACMG criteria-Varsome)	Regulatory annotations, chromatin states of the surrounding regions, methylation and acetylation status, hypersensitivity to the DNaseI (rVarBase, Encode)
Possible impact of an amino acid substitution on the structure and function of a human protein (PholyPhen2, PROVEAN)	TF binding sites (PromoPrediction, TF Bind, Lasagna)
	miRNA binding sites (TargetScan Human, miRDB)
	Consensus regions for potential splice sites (Human Splicing Finder)

4.4. Statistical Analysis

Tumor type percentages, family distribution, and age at diagnosis in coding and non-coding variant carriers were compared by χ^2 -test. $p < 0.05\%$ was considered statistically significant. The same test was used to estimate the effect of Pathogenic, VUS, and Likely Benign variants on clinical characteristics of coding variant carriers and to evaluate the presence of multiple variants in more aggressive tumors in non-coding variant carriers.

The gene association with tumor aggressiveness was performed through the Fisher's test. The results were considered significant with a threshold of $p < 0.05\%$. All analyses were performed with the free software Social Science Statistics (<https://www.socscistatistics.com>, accessed on January 2021).

Supplementary Materials: The following are available online at www.mdpi.com/article/10.3390/ijms22147693/s1: Table S1: List of 89 coding variants identified in 450 of patients; Table S2: List of 75 non-coding variants identified in 5'UTRs (a), 3'UTRs (b), splice regions (c), intronic regions (d), and the incidental findings (e); Table S3: Genomic coordinates of the regulatory regions covered with the probe panel.

Author Contributions: Conceptualization, M.A.C. and A.G.; methodology, C.G., R.S., G.G., E.C., M.T., F.B., T.C., E.F., and C.C. (Cinzia Cosini); Software, C.G., P.A., and E.C.; Validation, C.G. and E.C.; Investigation C.V., F.S., L.B., S.T., M.G.T., E.T. (Enrico Tagliafico), L.C., I.C., C.C. (Caterina Congregati), M.M., M.P., B.P., and M.G.; Data curation, C.G., E.C., M.T., and P.A.; Writing—original draft preparation, C.G. and M.A.C.; Writing—review and editing S.T., M.G.T., E.T. (Enrico Tagliafico), E.T. (Elena Tenedini), L.C., I.C., and T.C.; Supervision, A.G. and M.A.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by FONDAZIONE PISA, grant number no. 127/16 (Italy).

Institutional Review Board Statement: The study was approved by Ethics Committee of Area Vasta Nord Ovest (CEAVNO-protocol code: 13616-date of approval: 10/10/2019).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Acknowledgments: The authors would like to acknowledge all laboratory technicians of the Molecular Genetics Section AOUP.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Torre, L.A.; Bray, F.; Siegel, R.L.; Ferlay, J.; Lortet-Tieulent, J.; Jemal, A. Global cancer statistics, 2012. *CA Cancer J. Clin.* **2015**, *65*, 87–108.
2. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2018**, *68*, 394–424, doi:10.3322/caac.21492.
3. Mavaddat, N.; Peock, S.; Frost, D.; Ellis, S.; Platte, R.; Fineberg, E.; Evans, D.G.; Izatt, L.; Eeles, R.A.; Adlard, J.; et al. Cancer risks for BRCA1 and BRCA2 mutation carriers: Results from prospective analysis of EMBRACE. *J. Nat. Cancer Inst.* **2013**, *105*, 812–822.
4. Nielsen, F.C.; van Overeem Hansen, T.; Sorensen, C.S. Hereditary breast and ovarian cancer: New genes in confined pathways. *Nat. Rev. Cancer* **2016**, *16*, 599–612.
5. Lord, C.J.; Ashworth, A. BRCAness revisited. *Nat. Rev. Cancer* **2016**, *16*, 110–120.
6. Campeau, P.M.; Foulkes, W.D.; Tischkowitz, M.D. Hereditary breast cancer: New genetic developments, new therapeutic avenues. *Hum. Genet.* **2008**, *124*, 31–42.
7. Antoniou, A.C.; Pharoah, P.D.; McMullan, G.; Day, N.E.; Ponder, B.A.; Easton, D. Evidence for further breast cancer susceptibility genes in addition to BRCA1 and BRCA2 in a population-based study. *Genet. Epidemiol.* **2001**, *21*, 1–18.
8. Bonadona, V.; Lasset, C. Inherited predisposition to breast cancer: After the BRCA1 and BRCA2 genes, what next? *Bull. Cancer* **2003**, *90*, 587–594.
9. Turnbull, C.; Rahman, N. Genetic Predisposition to Breast Cancer: Past, Present, and Future. *Hum. Genet.* **2008**, *9*, 321–345.
10. Diederichs, S.; Bartsch, L.; Berkmann, J.C.; Fröse, K.; Heitmann, J.; Hoppe, C.; Iggena, D.; Jazmati, D.; Karschnia, P.; Linsenmeier, M.; et al. The dark matter of the cancer genome: Aberrations in regulatory elements, untranslated regions, splice sites, non-coding RNA and synonymous mutations. *EMBO Mol. Med.* **2016**, *8*, 442–457.
11. De Siervi, A.; De Luca, P.; Byun, J.S.; Di, L.J.; Fufa, T.; Haggerty, C.M.; Vazquez, E.; Moiola, C.; Longo, D.L.; Gardner, K. Transcriptional autoregulation by BRCA1. *Cancer Res.* **2010**, *70*, 532–542.
12. Wang, J.; Bian, C.; Li, J.; Couch, F.J.; Wu, K.; Zhao, R.C. Poly(ADP-ribose) polymerase-1 down-regulates BRCA2 expression through the BRCA2 promoter. *J. Biol. Chem.* **2008**, *283*, 36249–36256.
13. Burke, L.J.; Sevcik, J.; Gambino, G.; Tudini, E.; Mucaki, E.J.; Shirley, B.C.; Whiley, P.; Parsons, M.T.; Leeneer, K.D.; Gutiérrez-Enríquez, S.; et al. BRCA1 and BRCA2 5' noncoding region variants identified in breast cancer patients alter promoter activity and protein binding. *Hum. Mutat.* **2018**, *39*, 2025–2039.
14. Lheureux, S.; Lambert, B.; Krieger, S.; Legros, A.; Vaur, D.; Denoyelle, C.; Berthet, P.; Poulain, L.; Hardouin, A. Two novel variants in the 3'UTR of the BRCA1 gene in familial breast and/or ovarian cancer. *Breast Cancer Res. Treat.* **2011**, *125*, 885–891.
15. Brewster, B.L.; Rossiello, F.; French, J.D.; Edwards, S.L.; Wong, M.; Wronski, A.; Whiley, P.; Waddell, N.; Chen, X.; Bove, B.; et al. Identification of fifteen novel germline variants in the BRCA1 3'UTR reveals a variant in a breast cancer case that introduces a functional miR-103 target site. *Hum. Mutat.* **2012**, *33*, 1665–1675.
16. Zhou, X.J.; Wu, J.; Shi, L.; Li, X.-X.; Zhu, L.; Sun, X.; Qian, J.-Y.; Wang, Y.; Wei, J.-F.; Ding, Q. PTEN expression is upregulated by a RNA-binding protein RBM38 via enhancing its mRNA stability in breast cancer. *J. Exp. Clin. Cancer Res.* **2017**, *36*, 149.

17. Fang, H.; Xie, J.; Zhang, M.; Zhao, Z.; Wan, Y.; Yao, Y. miRNA-21 promotes proliferation invasion of triple-negative breast cancer cells through targeting, P.T.E.N. *Am. J. Transl. Res.* **2017**, *9*, 953–961.
18. Plon, S.E.; Eccles, D.M.; Easton, D.; Foulkes, W.D.; Genuardi, M.; Greenblatt, M.S.; Hogervorst, F.B.L.; Hoogerbrugge, N.; Spurdle, A.B.; Tavtigian, S.V.; et al. Sequence variant classification and reporting: Recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum. Mutat.* **2008**, *29*, 1282–1291.
19. Tavtigian, S.V.; Greenblatt, M.S.; Harrison, S.M.; Nussbaum, R.L.; Prabhu, S.A.; Boucher, K.M.; Biesecker, L.G. ClinGen Sequence Variant Interpretation Working Group. Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genet. Med.* **2018**, *20*, 1054–1060.
20. Iversen, E.S., Jr.; Couch, F.J.; Goldgar, D.E.; Tavtigian, S.V.; Monteiro, A.N. A computational method to classify variants of uncertain significance using functional assay data with application to BRCA1. *Cancer Epidemiol. Biomark. Prev.* **2001**, *20*, 1078–1088.
21. Parsons, M.T.; Tudini, E.; Li, H.; Hahnen, E.; Wappenschmidt, B.; Feliubadaló, L.; Aalfs, C.M.; Agata, S.; Aittomäki, K.; Alducci, E.; et al. Large scale multifactorial likelihood quantitative analysis of BRCA1 and BRCA2 variants: An ENIGMA resource to support clinical variant classification. *Hum. Mutat.* **2019**, *40*, 1557–1578.
22. Millot, G.A.; Carvalho, M.A.; Caputo, S.M.; Vreeswijk, M.P.; Brown, M.A.; Webb, M.; Rouleau, E.; Neuhausen, S.L.; Hansen, T.V.O.; Galli, A.; et al. A guide for functional analysis of BRCA1 variants of uncertain significance. *Hum. Mutat.* **2012**, *33*, 1526–1537.
23. Guidugli, L.; Carreira, A.; Caputo, S.M.; Ehlen, A.; Galli, A.; Monteiro, A.N.A.; Neuhausen, S.L.; Hansen, T.V.O.; Couch, F.J.; Vreeswijk, M.P.G.; et al. Functional assays for analysis of variants of uncertain significance in BRCA2. *Hum. Mutat.* **2014**, *35*, 151–164.
24. Jhuraney, A.; Velkova, A.; Johnson, R.C.; Kessing, B.; Carvalho, R.S.; Whiley, P.; Spurdle, A.B.; Vreeswijk, M.P.G.; Caputo, S.M.; Millot, G.A.; et al. BRCA1 Circos: A visualisation resource for functional analysis of missense variants. *J. Med. Genet.* **2015**, *52*, 224–230.
25. Lindor, N.M.; Guidugli, L.; Wang, X.; Vallée, M.P.; Monteiro, A.N.A.; Tavtigian, S.; Goldgar, D.E.; Couch, F.J. A review of a multifactorial probability-based model for classification of BRCA1 and BRCA2 variants of uncertain significance (VUS). *Hum. Mutat.* **2012**, *33*, 8–21.
26. Wardrop, S.L.; kConFab Investigators; Brown, M.A. Identification of two evolutionarily conserved and functional regulatory elements in intron 2 of the human BRCA1 gene. *Genomics* **2005**, *86*, 316–328.
27. Breast Cancer Association Consortium; Dorling, L.; Carvalho, S.; Allen, J.; González-Neira, A.; Luccarini, C.; Wahlström, C.; Pooley, K.A.; Parsons, M.T.; Fortuno, C.; et al. Breast Cancer Risk Genes-Association Analysis in More than 113,000 Women. *N. Engl. J. Med.* **2021**, *384*, 428–439.
28. Couch, F.J.; Shimelis, H.; Hu, C.; Hart, S.N.; Polley, E.C.; Na, J.; Hallberg, E.; Moore, R.; Thomas, A.; Lilyquist, J.; et al. Associations Between Cancer Predisposition Testing Panel Genes and Breast Cancer. *JAMA Oncol.* **2017**, *3*, 1190–1196.
29. Hauke, J.; Horvath, J.; Groß, E.; Gehrig, A.; Honisch, E.; Hackmann, K.; Schmidt, G.; Arnold, N.; Faust, U.; Sutter, C.; et al. Gene panel testing of 5589 BRCA1/2-negative index patients with breast cancer in a routine diagnostic setting: Results of the German Consortium for Hereditary Breast and Ovarian Cancer. *Cancer Med.* **2018**, *7*, 1349–1358.
30. Buys, S.S.; Sandbach, J.F.; Gammon, A.; Patel, G.; Kidd, G.; Brown, K.L.; Sharma, L.; Saam, J.; Lancaster, J.; Daly, M.B. A Study of Over 35,000 Women With Breast Cancer Tested With a 25-Gene Panel of Hereditary Cancer Genes. *Cancer* **2017**, *123*, 1721–1730.
31. Santana Dos Santos, E.; Caputo, S.M.; Castera, L.; Gendrot, M.; Briaux, A.; Breault, M.; Krieger, S.; Rogan, P.K.; Mucaki, E.J.; Burke, L.J.; et al. Assessment of the functional impact of germline BRCA1/2 variants located in non-coding regions in families with breast and/or ovarian cancer predisposition. *Breast Cancer Res. Treat.* **2018**, *168*, 311–325.
32. Evans, D.G.R.; van Veen, E.M.; Byers, H.J.; Wallace, A.J.; Ellingford, J.; Beaman, G.M.; Santoyo-Lopez, J.; Aitman, T.J.; Eccles, D.; Lalloo, F.; et al. A Dominantly Inherited 5' UTR Variant Causing Methylation-Associated Silencing of BRCA1 as a Cause of Breast and Ovarian Cancer. *Am. J. Hum. Genet.* **2018**, *103*, 213–220.
33. Laner, A.; Benet-Pages, A.; Neitzel, B.; Holinski-Feder, E. Analysis of 3297 individuals suggests that the pathogenic germline 5'-UTR variant BRCA1 c.-107A > T is not common in south-east Germany. *Fam. Cancer* **2020**, *19*, 211–213.
34. Yoshikawa, Y.; Kozuma, H.; Morimatsu, M.; Sugawara, K.; Orino, K. Reduced translation efficiency due to novel splicing variants in 5' untranslated region and identification of novel cis-regulatory elements in canine and human BRCA2. *BMC Mol. Cell Biol.* **2021**, *22*, 2.
35. Montalban, G.; Bonache, S.; Moles-Fernández, A.; Gisbert-Beamud, A.; Tenés, A.; Bach, V.; Carrasco, E.; López-Fernández, A.; Stjepanovic, N.; Balmaña, J.; et al. Screening of BRCA1/2 deep intronic regions by targeted gene sequencing identifies the first germline BRCA1 variant causing pseudoexon activation in a patient with breast/ovarian cancer. *J. Med. Genet.* **2019**, *56*, 63–74.
36. Shimelis, H.; LaDuca, H.; Hu, C.; Hart, S.N.; Na, J.; Thomas, A.; Akinhanmi, M.; Moore, R.; Brauch, H.; Cox, A.; et al. Triple-Negative Breast Cancer Risk Genes Identified by Multigene Hereditary Cancer Panel Testing. *J. Natl. Cancer Inst.* **2018**, *110*, 855–862.
37. Zanti, M.; Loizidou, M.A.; Michailidou, K.; Pirpa, P.; Machattou, C.; Marcou, Y.; Kyriakou, F.; Kakouri, E.; Tanteles, G.A.; Spanou, E.; et al. NGS Panel Testing of Triple-Negative Breast Cancer Patients in Cyprus: A Study of BRCA-Negative Cases. *Cancers* **2020**, *12*, 3140.

38. Spugnese, L.; Gabriele, M.; Scarpitta, R.; Tancredi, M.; Maresca, L.; Gambino, G.; Collavoli, A.; Aretini, P.; Bertolini, I.; Salvadori, B.; et al. Germline mutations in DNA repair genes may predict neoadjuvant therapy response in triple negative breast patients. *Genes Chromosomes Cancer* **2016**, *55*, 915–924.
39. Silver, D.P.; Richardson, A.L.; Eklund, A.C.; Wang, Z.C.; Szallasi, Z.; Li, Q.; Juul, N.; Leong, C.-O.; Calogrias, D.; Buraimoh, A.; et al. Efficacy of neoadjuvant cisplatin in triple-negative breast cancer. *J. Clin. Oncol.* **2010**, *28*, 1145–1153.
40. Byrski, T.; Dent, R.; Blecharz, P.; Foszczynska-Kloda, M.; Gronwald, J.; Huzarski, T.; Cybulski, C.; Marczyk, E.; Chrzan, R.; Eisen, A.; et al. Results of a phase II open-label, non-randomized trial of cisplatin chemotherapy in patients with BRCA1-positive metastatic breast cancer. *Breast Cancer Res.* **2012**, *14*, R110.
41. Priyanka, S. Biology and Management of Patients With Triple-Negative Breast Cancer. *Oncologist* **2016**, *21*, 1050–1062.
42. Lu, H.-M.; Li, S.; Black, M.H.; Lee, S.; Hoiness, R.; Wu, S.; Mu, W.; Huether, R.; Chen, J.; Sridhar, S.; et al. Association of Breast and Ovarian Cancers With Predisposition Genes Identified by Large-Scale Sequencing. *JAMA Oncol.* **2019**, *5*, 122.
43. Cooper, G.M.; Stone, E.A.; Asimenos, G.; NISC Comparative Sequencing Program; Green, E.D.; Batzoglou, S. Arend Sidow Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* **2005**, *15*, 901–913.
44. Landrum, M.J.; Lee, J.M.; Benson, M.; Brown, G.; Chao, C.; Chitipiralla, S.; Gu, B.; Hart, J.; Hoffman, D.; Hoover, J.; et al. ClinVar: Public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **2016**, *44*, D862–D868.
45. Maxwell, K.N.; Hart, S.N.; Vijai, J.; Schrader, K.A.; Slavin, T.P.; Thomas, T.; Wubbenhorst, B.; Ravichandran, V.; Moore, R.M.; Hu, C.; et al. Evaluation of ACMG-Guideline-Based Variant Classification of Cancer Susceptibility and Non-Cancer-Associated Genes in Families Affected by Breast Cancer. *Am. J. Hum. Genet.* **2016**, *98*, 801–817.
46. Kopanos, C.; Tsiolkas, V.; Kouris, A.; Chapple, C.E.; Aguilera, M.A.; Meyer, R.; Massouras, A. VarSome: The human genomic variant search engine. *Bioinformatics* **2019**, *35*, 1978–1980.
47. Adzhubei, I.; Jordan, D.M.; Sunyaev, S.R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **2013**, doi:10.1002/0471142905.hg0720s76.
48. Choi, Y.; Chan, A.P. PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **2015**, *31*, 2745–2747.
49. Guo, L.; Du, Y.; Qu, S.; Wang, J. rVarBase: An updated database for regulatory features of human variants. *Nucleic Acids Res.* **2016**, *44*, D888–D893.
50. Ward, L.D.; Kellis, M. Interpreting noncoding genetic variation in complex traits and human disease. *Nat. Biotechnol.* **2012**, *30*, 1095–1106.
51. Rojano, E.; Seoane, P.; Ranea JA, G.; Perkins, J.R. Regulatory variants: From detection to predicting impact. *Brief Bioinform.* **2018**, *20*, 1639–1654.
52. Farré, D.; Roset, R.; Huerta, M.; Adsuaara, J.E.; Roselló, L.; Albà, M.M.; Messeguer, X. Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. *Nucleic Acids Res.* **2003**, *31*, 3651–3653.
53. Lee, C.; Huang, C.H. LASAGNA-Search 2.0: Integrated transcription factor binding site search and visualization in a browser. *Bioinformatics* **2014**, *30*, 1923–1925.
54. Agarwal, V.; Bell, G.W.; Nam, J.; Bartel, D.P. Predicting effective microRNA target sites in mammalian mRNAs. *eLife* **2015**, *4*, e05005.
55. Yuhao, C.; Xiaowei, W. miRDB: An online database for prediction of functional microRNA targets. *Nucleic Acids Res.* **2020**, *48*, D127–D131.
56. Ohno, K.; Takeda, J.I.; Masuda, A. Rules and tools to predict the splicing effects of exonic and intronic mutations. *Wiley Interdiscip. Rev. RNA* **2018**, *9*, doi:10.1002/wrna.1451.