Uncovering the clinical relevance of unclassified variants in DNA repair genes: a focus on *BRCA* negative Tunisian cancer families

Maroua Boujemaa[1]*, Fatma Nouira[2], Nouha Jandoubi[1], Nesrine Mejri[1,3], Hanen Bouaziz[1,4], Cherine Charfeddine[1,5], Sonia Ben Nasr[1,6], Soumaya Labidi[1,3], Houda El Benna[1,3], Yosra Berrazega[3], Haifa Rachdi[3], Nouha Daoud[3], Farouk Benna[7], Abderrazek Haddaoui[6], Sonia Abdelhak[1], Mohamed Samir Boubaker[1,8], Hamouda Boussen[1,3] and Yosr Hamdi[1,8]*

[1]Laboratory of Biomedical Genomics and Oncogenetics, LR20IPT05, Institut Pasteur de Tunis, University of Tunis El Manar, Tunis, Tunisia, [2]Laboratory of Bioactive Substances, Center of Biotechnology of Borj Cedria, University of Tunis El Manar, Hamam, Tunisia, [3]Medical Oncology Department, Abderrahman Mami Hospital, Faculty of Medicine Tunis, University Tunis El Manar, Tunis, Tunisia, [4]Surgical Oncology Department, Salah Azaiez Institute of Cancer, Tunis, Tunisia, [5]High Institute of Biotechnology of Sidi Thabet, Biotechpole of Sidi Thabet, University of Manouba, Ariana, Tunisia, [6]Department of Medical Oncology, Military Hospital of Tunis, Tunis, Tunisia, [7]Radiation Oncology Department, Salah Azaiez Institute, Tunis, Tunisia, [8]Laboratory of Human and Experimental Pathology, Institut Pasteur de Tunis, Tunis, Tunisia

Introduction: Recent advances in sequencing technologies have significantly increased our capability to acquire large amounts of genetic data. However, the clinical relevance of the generated data continues to be challenging particularly with the identification of Variants of Uncertain Significance (VUSs) whose pathogenicity remains unclear. In the current report, we aim to evaluate the clinical relevance and the pathogenicity of VUSs in DNA repair genes among Tunisian breast cancer families.

Methods: A total of 67 unsolved breast cancer cases have been investigated. The pathogenicity of VUSs identified within 26 DNA repair genes was assessed using different *in silico* prediction tools including SIFT, PolyPhen2, Align-GVGD and VarSEAK. Effects on the 3D structure were evaluated using the stability predictor DynaMut and molecular dynamics simulation with NAMD. Family segregation analysis was also performed.

Results: Among a total of 37 VUSs identified, 11 variants are likely deleterious affecting *ATM, BLM, CHEK2, ERCC3, FANCC, FANCG, MSH2, PMS2 and RAD50* genes. The *BLM* variant, c.3254dupT, is novel and seems to be associated with increased risk of breast, endometrial and colon cancer. Moreover, c.6115G>A in *ATM* and c.592+3A>T in *CHEK2* were of keen interest identified in families with multiple breast cancer cases and their familial cosegregation with disease has been also confirmed. In addition, functional *in silico* analyses revealed that the *ATM* variant may lead to protein immobilization and rigidification thus decreasing its activity. We have also shown that *FANCC* and *FANCG* variants may lead to protein destabilization and alteration of the structure compactness which may affect FANCC and FANCG protein activity.

Conclusion: Our findings revealed that VUSs in DNA repair genes might be associated with increased cancer risk and highlight the need for variant

reclassification for better disease management. This will help to improve the genetic diagnosis and therapeutic strategies of cancer patients not only in Tunisia but also in neighboring countries.

1 Introduction

Genetic testing of hereditary breast cancer has become broadly available especially with the emergence of Next Generation Sequencing (NGS) technologies that have revolutionized the field of genomics (Yadav and Couch, 2019). To identify women at increased risk of developing breast cancer, the National Comprehensive Cancer Network and the American Society of Clinical Oncology both recommend complete mutation screening of *ATM, CDH1, CHEK2, PALB2, PTEN, STK11* and *TP53* in addition to *BRCA1/2* genes. There is also accumulative evidence that mutations in *BARD1, BRIP1, MSH2, MLH1, MSH6, NBN, NF1, PMS2, RAD51C,* and *RAD51D* predispose to hereditary breast cancer (Urbina-Jara et al., 2019; Yadav and Couch, 2019). Carriers of risk variants may benefit from enhanced screening, chemoprevention and/or preventive surgery. Genetic testing is also becoming increasingly relevant in cancer therapy since patients with defects in DNA repair genes particularly those with *BRCA* mutations and/or with Homologous Recombination Deficiency (HRD) are more sensitive to platinum-based therapies and may also benefit from PARP inhibitors (Monteiro et al., 2020; Gonzalez and Stenzinger, 2021). Certainly, that progress made in NGS technologies has increased our capability to understand the genetic basis of cancer (Gerasimavicius et al., 2020). Nevertheless, the main challenge in clinical genetic testing remains the interpretation of the identified variants particularly those with uncertain significance (Ernst et al., 2018; Gerasimavicius et al., 2020; Monteiro et al., 2020). Disease causing variants are usually nonsense mutations that result in a premature stop codon, small indels causing a frameshift and a premature termination of translation, splicing site mutations that occur inside of the canonical splice sites, large deletions or known deleterious missense variants. Whereas variants with insufficient or conflicting evidence of pathogenicity that support their association with disease, are classified as Variants of Uncertain Significance (VUSs). These variants are usually missense substitutions and small in-frame indels (Girolimetti et al., 2014; Joynt et al., 2022). In public databases more than 90% of missense variants identified by clinical genetic testing are described as variants of uncertain significance (Girolimetti et al., 2014; Landrum et al., 2020), of these a significant fraction is identified within *BRCA* genes and other DNA repair genes involved in cancer predisposition (Hart et al., 2019; Monteiro et al., 2020). The lack of evidence on the pathogenicity of VUSs in these cancer associated genes represent a critical clinical challenge and efforts to review the classification of these variants are urgently needed. With this in mind, several initiatives have been established such as "ENIGMA consortium" aiming to determine the clinical significance of sequence variants in *BRCA1, BRCA2* and other known or candidate breast cancer susceptibility genes (Spurdle et al., 2012; Richards et al., 2015; Hart et al., 2019; Heczkova et al., 2019). Moreover, the genetic and cancer group (GGC) – Unicancer have launched the COVAR project in order to classify the maximum of VUSs, detected in *BRCA1/2* and in genes included in hereditary cancer panels, in terms of their probability to be pathogenic based on co-segregation analysis (Cohen-Haguenauer, 2019).

In the current report, we aim to assess the clinical relevance and the pathogenicity of VUSs in DNA repair genes in Tunisian cancer families.

2 Materials and methods

2.1 Patients

A total of 67 unsolved hereditary breast cancer cases were included in this study. Written informed consents were obtained from all participants. The study was conducted according to the Declaration of Helsinki Principles and ethical approval was obtained from the biomedical ethics committee of Institut Pasteur de Tunis (2017/16/E/Hopital A-M & 2019/1/I/LR16IPT05).

2.2 DNA isolation

Total genomic DNA was isolated from peripheral blood using QIAamp DNA Mini Kit (Qiagen) according to the manufacturer's instructions. DNA purity and concentration were measured using a NanoDrop™ spectrophotometer.

2.3 Next Generation Sequencing and data analysis

Targeted sequencing using a panel of 24 genes was performed on 53 patients. This panel included the following genes: *BRCA1, BRCA2, PALB2, CDH1, PTEN, TP53, RAD51C, RAD51D, MLH1, MSH2, MSH6, PMS2, EPCAM, ATM, BRIP1, CHEK2, STK11, MRE11A, NBN, RAD50, BARD1, BLM, XRCC2* and *MUTYH*. Target enrichment of coding regions and intron/exon junctions (±50bp) was performed by Twist technology followed by a pair-end sequencing reaction (2 × 150bp) on a Nextseq2000 platform (Illumina).

Whole Exome Sequencing was carried out on 14 patients. Samples were prepared according to Agilent SureSelectXT Human All Exon V6 Protocol and enrichment was done according to Agilent SureSelect protocols. Paired-end (2 × 150) sequencing was performed on enriched samples on the Illumina NovaSeq6000 system.

Bioinformatic analysis was performed using an In-House pipeline. Data quality control and preprocessing (including adapter trimming, quality trimming and removal of very short reads) were performed using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and BBDuk (https://jgi.doe.gov/data-and-tools/software-tools/bbtools/) tools respectively. DNA sequences were mapped to their location in the build of the human genome (hg19/b37) using the Burrows–Wheeler Aligner (BWA) package (Li and Durbin, 2009). The subsequent SAM files were converted to BAM files using Samtools (Li et al., 2009). Duplicate reads were removed using Picard version 2.6 (http://broadinstitute.github.io/picard/). Post-Alignment quality control and variant calling were performed using GATK version 4.1.2 (McKenna et al., 2010). VarAFT version 2.16 was subsequently used for variant annotation (Desvignes et al., 2018). First, we have looked for known pathogenic variants in 37 genes frequently analyzed in high-risk breast and ovarian cancer families (Suszynska et al., 2019). Then, we have focused our analysis on a set of DNA repair genes to unravel the genetic etiology of the investigated cases.

2.4 DNA repair genes investigation

In order to evaluate whether variants in DNA repair genes might be associated with hereditary predisposition in the Tunisian population, a set of 169 genes (Supplementary Table S1) have been investigated. This list included known genes associated with hereditary predisposition to cancer in addition to other genes belonging to 7 major DNA repair pathways retrieved from KEGG GENES Database (https://www.genome.jp/kegg/genes.html) and mdanderson data (https://www.mdanderson.org/documents/Labs/Wood-Laboratory/human-dna-repair-genes.html) and it was limited to 24 genes for patients investigated by gene panel. The 7 major DNA repair pathways are as follows: 1) Base-excision repair (BER) 2) Mismatch repair (MMR) 3) Nucleotide-excision repair (NER) 4) Homologous recombination (HR) 5) Nonhomologous end joining (NHEJ) 6) Translesional synthesis (TLS) and 7) Fanconi Anemia pathways. Among these genes, unclassified genetic variants described as variants of uncertain significance, having conflicting interpretations of pathogenicity in the ClinVar database or not described in public databases were selected for further analyses.

Protein-protein interaction network and functional enrichment analyses were performed using String (Jensen et al., 2009) and EnrichR databases (Kuleshov et al., 2016).

2.5 Assessment of the functional impact of variants

2.5.1 Pathogenicity predictions (sequence based prediction)

The pathogenicity of variants was assessed using several *in silico* predictions tools including: SIFT (Vaser et al., 2016), PolyPhen2 (Adzhubei et al., 2013), LRT (Chun and Fay, 2009), MutationAssessor (Reva et al., 2011) PROVEAN (Choi et al., 2012), MutationTaster (Schwarz et al., 2010), FATHMM (Rogers

et al., 2018), CADD (Kircher et al., 2014), MetaLR/MetaSVM (Dong et al., 2015), UMD-predictor (Salgado et al., 2016) and Align-GVGD (Tavtigian et al., 2006).

In addition, we have used VarSEAK (https://varseak.bio/), SpliceAI Lookup (https://spliceailookup.broadinstitute.org/) and Human Splicing Finder (HSF) (Desmet et al., 2009) to predict the effect of variants on splicing events. TraP v3 (Gelfman et al., 2017) and FATHMM-XF (Rogers et al., 2018) were used to predict the pathogenicity of synonymous as well as non-coding variants.

2.5.2 *In-silico* functional analyses (structure based prediction)

In silico functional analyses were performed to evaluate the impact of the identified variants on the protein structure. Experimentally derived three-dimensional protein structures were extracted from the Protein Data Bank (PDB) archive which represents the single repository of information about the 3D structures of proteins, nucleic acids, and complex assemblies (Berman et al., 2000). HHpred server (Soding et al., 2005) in combination with MODELLER software (Webb and Sali, 2016) were employed for structure prediction and homology modeling by using the MPI Bioinformatics Toolkit (Gabler et al., 2020) (https://toolkit.tuebingen.mpg.de). The impact of mutations on protein stability and flexibility was assessed using the DynaMut web server which also provides inter-residue interactions for both wild-type and mutant structures (Rodrigues et al., 2018). Stability and flexibility are evaluated respectively through the estimation of the difference in $\Delta\Delta G$ and $\Delta\Delta SVib$ between the normal and the mutant protein. The $\Delta\Delta G$ represents the folding free energy of the protein and is expressed in kcal/mol. DynaMut defines mutations with $\Delta\Delta G \geq 0$ as Stabilizing and those with $\Delta\Delta G < 0$ as deStabilizing. The effect on protein stability is considered as significant if $\Delta\Delta G \geq 0.5$ (Stabilisation) or $\Delta\Delta G < -0.5$ (destabilization). The $\Delta\Delta SVib$ or vibrational entropy energy contributes significantly to protein binding free energies. DynaMut implements ENCoM, to calculate $\Delta\Delta SVib$ as the difference between the vibrational energy of the wild-type and mutant structures. A $\Delta\Delta SVib < 0$ represents a rigidification of the protein structure while a $\Delta\Delta SVib \geq 0$ represents a gain in flexibility. To enhance the reliability of our results, we have also used the FoldX plugin to predict the effects of mutations on protein stability (Van Durme et al., 2011). Furthermore, a molecular dynamics simulation was performed using the two software packages VMD (Humphrey et al., 1996) and NAMD (Phillips et al., 2020) to better understand the effects of the identified variants on the 3D structure of the protein. Different parameters have been studied all over the simulation trajectory including root mean square deviation (RMSD) and the radius of gyration (Rg).

3 Results

The genetic investigation in 67 breast cancer cases using TGS and WES showed the absence of know pathogenic variants in genes associated with hereditary predisposition to cancer. To assess whether novel mutations or misclassified variants are responsible for disease susceptibility a set of 169 DNA repair genes have been investigated. This revealed a total of 37 unclassified variants
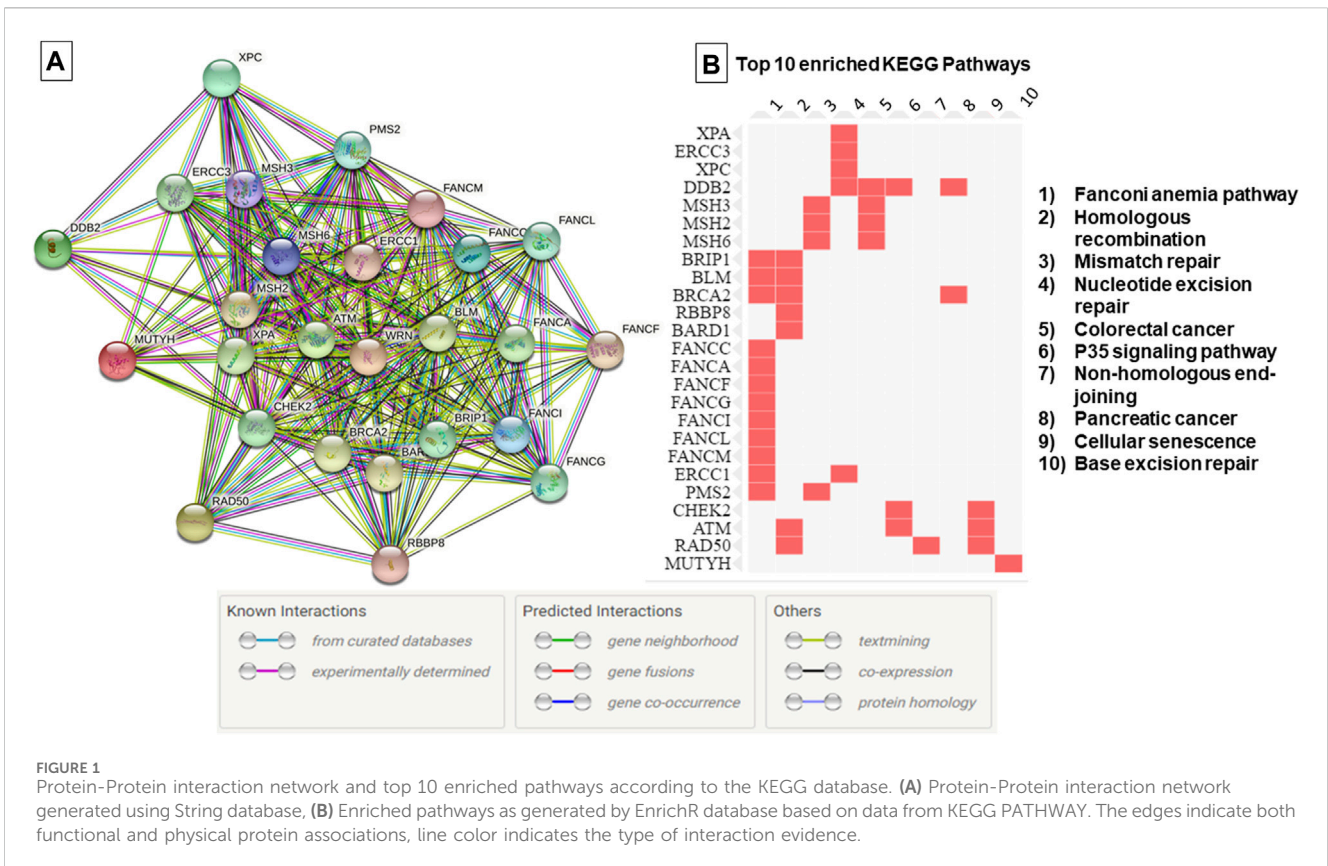
FIGURE 1
Protein-Protein interaction network and top 10 enriched pathways according to the KEGG database. **(A)** Protein-Protein interaction network generated using String database, **(B)** Enriched pathways as generated by EnrichR database based on data from KEGG PATHWAY. The edges indicate both functional and physical protein associations, line color indicates the type of interaction evidence.

localized within 26 DNA repair genes (Supplementary Table S2, Figure 1). It is noteworthy that the number of unclassified variants detected appears to be relatively few when compared to the total number of genes investigated. This observation is expected given the fact that most patients were subjected to TGS. In the current study, most of the identified variants were exonic (62%), 11% were splicing [±10bp from the exon-intron junction) and the remaining variations intronic and/or located within 5′ and 3′ untranslated regions (Supplementary Figure S1)]. According to the KEGG database, most of the identified variants are localized within genes involved in Fanconi anemia pathway followed by Homologous recombination and Mismatch repair pathway (Figure 1). Identified variants were prioritized based on frequency in public databases and predicted pathogenicity. Only those with a MAF <0.01 in gnomAD database and predicted as deleterious by at least 6 *in silico* tools and/or predicted to alter the protein structure and function were kept for further analyses.

3.1 Variant distribution and prioritization

3.1.1 Exonic variants

We have identified 23 exonic variants among which 9 are of particular interest, predicted as deleterious by at least 6 *in silico* prediction tools and/or predicted to affect the protein function (Table 1). They include 6 missense variants localized within *ATM, ERCC3, FANCC, FANCG, MSH2* and *PMS2* genes, 1 novel frameshift deletion in the *BLM* gene, 1 novel in-frame deletion in *CHEK2* gene and 1 frameshift deletion in *PMS2*.

3.1.2 Splicing variants

In the current report 4 splicing variants, located ±10bp from the exon-intron junction, have been identified (Supplementary Table S2) among which 2 variants, c.592+3A>T in *CHEK2* and c.3036+5G>A in *RAD50* gene, were predicted to alter the wild-type donor site and to probably affect splicing.

3.1.3 Non-coding variants

Ten non-coding variants have been identified. Among these variants, 4 were rare (MAF <0.01) and *in silico* predictions have not revealed any significant pathogenic effect (Supplementary Table S2).

In summary, among all the unclassified variants identified in this study, 11 are likely the most relevant (9 coding and 2 splicing).

3.2 Genotype phenotype correlation and segregation analysis

A total of 11 variants likely associated with hereditary predisposition have been identified and are localized within *ATM, BLM, CHEK2, ERCC3, FANCC, FANCG, MSH2, PMS2* and *RAD50* genes.

The *ATM* variant, c.6115G>A, results in the change of a Glutamic Acid to a Lysine. It was identified in 2 unrelated patients belonging to the same geographical region. A strong family history of breast cancer was observed among these two cases. The first patient is 41 years old with 6 cases of breast cancer in the family and the second is 38 years old having 8 cases of breast cancer in the family. A targeted mutation

TABLE 1 Relevant variants likely associated with cancer predisposition.

Gene	cDNA change	Proteine change	ClinVar	Frequency (GnomAD Exome)	Align GVGD[a]	CADD[b]	Mutaton taster	Mutation assessor	Polyphen2	PROVEAN	SIFT
ATM	c.6115G>A	Glu2039Lys	Conflicting interpretations of pathogenicity	0.000004	C55	26.2[b]	D[c]	M[d]	D[e]	D[f]	D[g]
BLM	c.3254dupT	Arg1086Lysfs[a]7	Not reported	unknown	—	—	D	—	—	—	D
CHEK2	c.493_498delGAATAT	Glu165_Tyr166del	Not reported	unknown	—	—	—	—	—	—	—
ERCC3	c.2111C>T	Ser704Leu	Conflicting interpretations of pathogenicity	0.002196	C65	24	D	M	P[h]	D	D
FANCC	c.1156T>C	Ser386Pro	Conflicting interpretations of pathogenicity	0.000553	C65	25.4	D	M	D	D	D
FANCG	c.366G>C	Trp122Cys	Conflicting interpretations of pathogenicity	0.00080	C65	31	D	M	D	D	D
MSH2	c.728G>A	Arg243Gln	Uncertain significance	0.00002	C35	34	D	M	D	D	D
PMS2	c.1004A>T	Asn335Ile	Uncertain significance	0.00025	C65	33	D	H[i]	D	D	D
PMS2	c.2186_2187del	Leu729Glnfs[a]6	Uncertain significance	0.0018	—	—	D	—	—	—	D

[a]C55-C65 Most likely affecting function.
[b]CADD cutoff on deleteriousness 15.
[c]Disease_causing.
[d]Medium.
[e]Probably damaging.
[h]Possibly damaging.
[f]Deleterious.
[g]Deleterious.
[i]High.
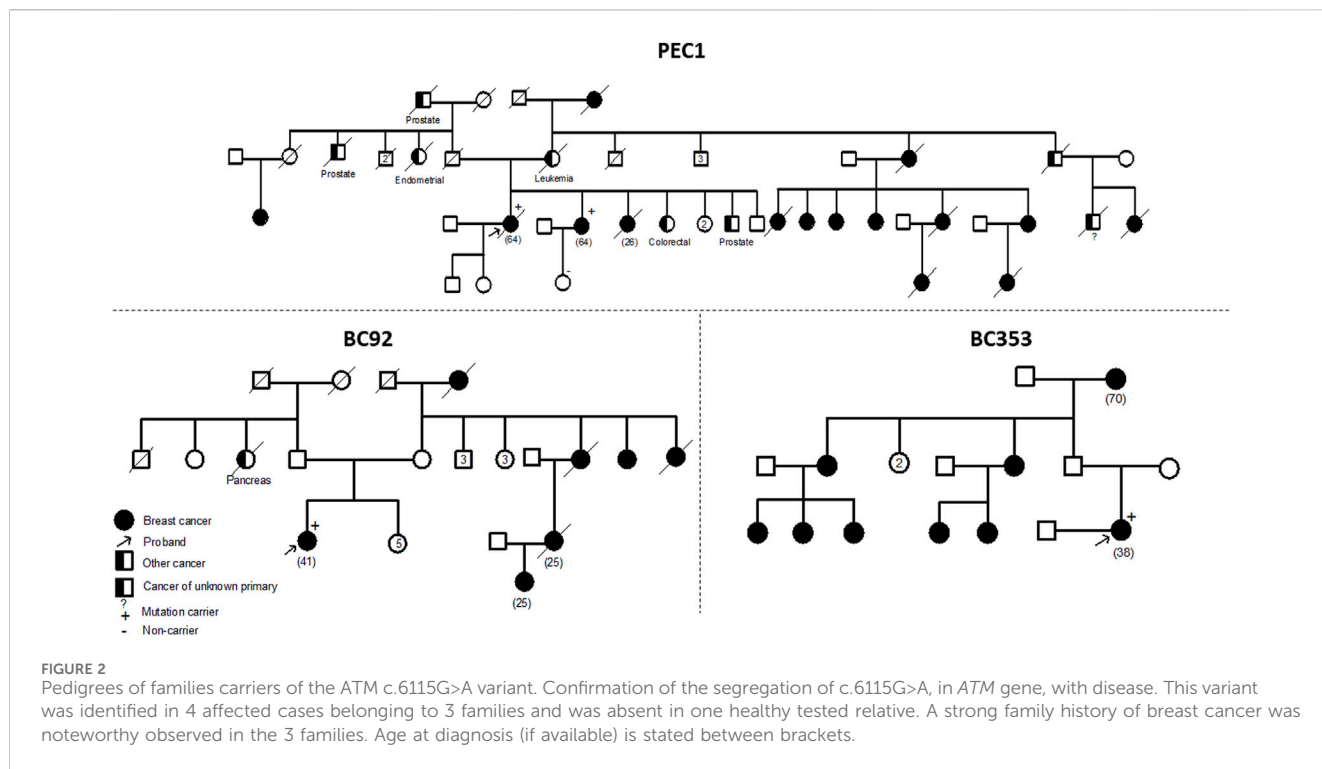BC, Breast Cancer; CC, Colorectal Cancer; ED, Endometrial Cancer; PC, Prostate Cancer.

FIGURE 2
Pedigrees of families carriers of the ATM c.6115G>A variant. Confirmation of the segregation of c.6115G>A, in *ATM* gene, with disease. This variant was identified in 4 affected cases belonging to 3 families and was absent in one healthy tested relative. A strong family history of breast cancer was noteworthy observed in the 3 families. Age at diagnosis (if available) is stated between brackets.

TABLE 2 Splice sites variations likely associated with disease predisposition.

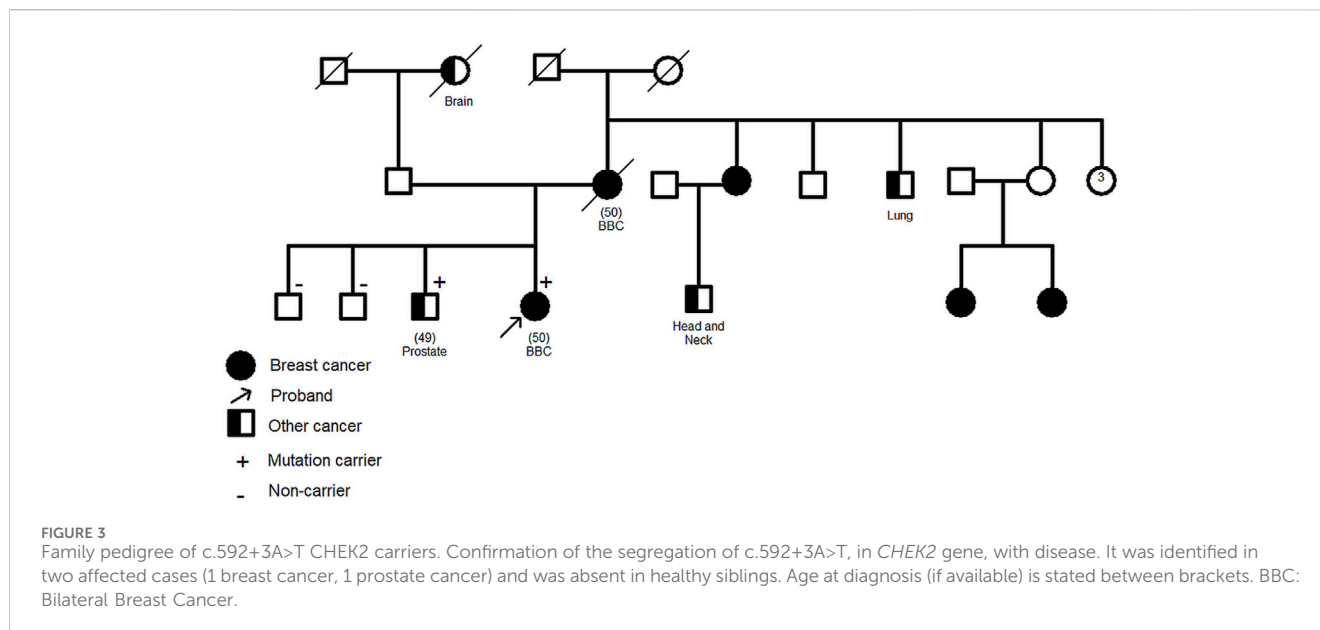| Variation	Carriers	Disease	VarSEAK	Human splicing finder	SpliceAI lookup
NM_007194.4(*CHEK2*): c.592+3A>T	BC419-1	BC	Class 5: Splicing effect	Broken WT Donor Site: Alteration of the WT Donor site, most probably affecting splicing	High probability of Donor site loss
	BC419-2	PC	Loss of function for authentic Splice Site: Exon Skipping		
			VarSEAK Classification: Likely Pathogenic		
NM_005732.4(*RAD50*): c.3036+5G>A	BC40	BC	Class 4: Likely splicing effect	Broken WT Donor Site: Alteration of the WT Donor site, most probably affecting splicing	No Splicing Effect
			Likely loss of function for authentic Splice Site. Exon Skipping		

BC, Breast Cancer; PC, Prostate Cancer.

screening was conducted in this geographical region and allowed the identification of a third carrier who had developed breast cancer at the age of 64. This variation was subsequently confirmed in her sister diagnosed also with breast cancer and was absent in the healthy niece. In this family, a strong family history of breast cancer was also observed where 14 breast cancer cases were detected. These findings support the pathogenicity of this variation and suggest a possible founder effect as well as a plausible association with a breast cancer only phenotype (Figure 2).

The *CHEK2* variant, c.592+3A>T, consists of an A>T nucleotide substitution at the +3 position of intron 4 of the *CHEK2* gene. *In silico* predictions revealed that this variant may alter the wild type of the donor site leading to exon skipping (Table 2). It was identified in a 50 year old woman diagnosed with bilateral breast cancer. This same mutation was present in her brother diagnosed with prostate cancer at 49 years old and was absent in healthy siblings (Figure 3) which support the pathogenicity of this variant.

For the other identified variants, segregation analysis was not accessible due to the fact that most of the affected members of the investigated families were deceased at the time of the analysis. Among these high-risk variants, one novel homozygous frameshift deletion was identified in the *BLM* gene, c.3254dupT, that leads to premature termination of translation at codon position 1092. Interestingly, this variant was detected in one breast cancer patient with a personal history of endometrial and colorectal cancer. No other affected cases were identified within the family. *BLM* gene mutations are known to be associated with increased risk of developing multiple cancers including breast (Kluzniak et al., 2019), colorectal cancers (de Voer et al., 2015) and endometrial cancer (Long et al., 2019) which is consistent with our findings and support the pathogenicity of the identified variant.

Another novel deletion c.493_498delGAATAT was detected in the *CHEK2* gene. This in-frame deletion is located in the exon 4 of *CHEK2* gene and predicted to lead to the loss of two amino acids,

FIGURE 3
Family pedigree of c.592+3A>T CHEK2 carriers. Confirmation of the segregation of c.592+3A>T, in *CHEK2* gene, with disease. It was identified in two affected cases (1 breast cancer, 1 prostate cancer) and was absent in healthy siblings. Age at diagnosis (if available) is stated between brackets. BBC: Bilateral Breast Cancer.

Glu and Tyr in positions 165 and 166 of the protein. This deletion lies within the forkhead-associated (FHA) domain which is critically involved in dimerization of CHK2 molecules in phosphorylation-dependent manner. The patient harboring this variant was diagnosed with RH-/HER2+ breast cancer at 27 years old. Family history of early onset breast cancer was also observed among relatives in addition to other types of cancer.

Among the selected variants, c.728G>A in *MSH2* gene was identified in one breast cancer case with a family history of ovarian, gastric and lung cancers. Interestingly, this same variant was recently described in a Tunisian patient with gastric cancer that has a family history of breast, ovarian, and colon cancers (Kabbage et al., 2022). In this same study, structural bioinformatics analyses revealed that this variant is involved in the MSH2-MLH1 complex stability and may impact on the binding of MSH2 protein with MLH1 by disrupting the electrostatic potential which is suggestive of a pathogenic effect.

Moreover, one breast cancer patient (BC310) with family history of stomach, colon and breast cancer, harbored a frameshift deletion in the *PMS2* gene (c.2186_2187del). This latter is classified as a variant of uncertain significance in the ClinVar database although it is predicted to result in the premature termination of the protein at amino acid position 734 leading to protein function alteration. Genetic data and family history of BC310 are consistent with Lynch syndrome which is known to be associated with an increased risk of developing colorectal cancer as well as other tumor types including breast and gastric cancers (Fornasarig et al., 2018; Sheehan et al., 2020). In the same gene, a missense variation, Asn335Ile, was identified in a patient diagnosed with early onset breast cancer and with family history of breast, skin and pancreatic cancers. Considering *ERCC3* gene, a missense variation Ser704Leu was identified in a breast cancer patient with multiple affected relatives. In the same context, *ERCC3* germline mutations were previously identified in families with multiple breast cancer cases (Vijai et al., 2016). Two missense variations in *FANCC* and *FANCG* genes, predicted to be deleterious, were further identified in breast cancer cases with family history of different tumors including leukemia, breast, prostate, and colon cancers. Finally, a splicing mutation

predicted to lead to exon skipping was identified in *RAD50* gene (Tables 2, 3) in a breast cancer patient aged 37 years.

3.3 *In-silico* functional analyses

In-silico functional analyses including stability and flexibility analysis as well as molecular dynamics simulations were performed to predict the impact of coding missense variants on the 3D structure of the corresponding proteins mainly ATM, ERCC3, FANCC, FANCG, and PMS2. Among these variants, c.6115G>A (E2039K) in the *ATM* gene is of particular interest since it was identified in families with multiple cases of breast cancer (more than 6 cases). This variant sits within the regulatory FAT domain of the ATM protein. Predictions of protein stability and flexibility changes upon mutation revealed that E2039K may increase protein stability and lead to the rigidification of the protein structure (Table 4; Figure 4). Predicting stability changes using the FoldX plugin was also in line with the DynaMut results revealing an increase in stability upon mutation with a $\Delta\Delta G = -29,37$ (FoldX provides negative $\Delta\Delta G$ values for stabilizing mutations and positive values for destabilizing mutants). Our findings revealed also that E2039K may result in a change of interatomic molecular interactions with surrounding amino acids consisting of interatomic interactions loss with E2037, R2748 and R2929 and gain of interactions with M2041, as shown in Figure 4. Analysis of molecular dynamics results has also supported the previous findings. Indeed, the RMSD profile showed lower RMSD values of the mutant structure compared to the native protein consistent with a gain in stability. In addition, the radius of gyration (Rg) analysis, which gives information on protein compactness, revealed that the mutant protein is more compact than the native form thus correlating with the stability analysis results (Supplementary Figure S2). Among the other investigated variants, and based on DynaMut predictions, we have found that the amino acid changes S386P in FANCC and W122C in FANCG may lead to protein destabilization associated with a significant increase in flexibility particularly for W122C (FANCG). Furthermore, molecular dynamics simulation results have shown

TABLE 3 Clinicopathological features of patients carriers of relevant variants.

Variant	Patient ID	Screening method	Pathology	Age at diagnosis	Family history of cancer	Hormone receptor status	HER2 status	Ki67
ATM c.6115G>A	BC92	WES	Breast Cancer	41	6 Breast 1 Pancreas	NA	NA	NA
	BC353	TGS	Breast Cancer	38	8 Breast	ER+/PR+	HER2-	15%
	PEC1-1	Sanger sequencing	Breast Cancer	64	14 Breast 2 Prostate 1 Endometrium 1 Colorectal 1 Lung	ER+/PR-	HER2-	NA
	PEC1-2	Sanger sequencing	Breast Cancer	64		NA	NA	NA
BLM 3254dupT	BC418	TGS	Endometrial Cancer Breast Cancer Colorectal Cancer	47 48 51	-	NA	NA	NA
CHEK2 c.493_498delGAATAT	BC401	TGS	Breast Cancer	27	1 Breast 1 Pancreas	ER-/PR-	HER2+	60%
CHEK2 c.592+3A>T	BC419-1	TGS	Breast Cancer	50	4 Breast 1 Prostate 1 Head and neck 1 Brain 1 Lung	NA	NA	NA
	BC419-2	Sanger sequencing	Prostate Cancer	49		-	-	-
ERCC3 c.2111C>T	BC19	WES	Breast Cancer	49	3 Breast	ER-/PR-	NA	NA
FANCC c.1156T>C	BC252-1	WES	Breast Cancer	40	2 Prostate 3 Breast 1 Lung 1 Leukemia	ER+/ER+	HER2-	30%
FANCG c.366G>C	BC39	WES	Breast Cancer	60	1 Breast 1 colon 1 Endometrium 1 Testes	ER+/PR+	HER2-	50%
MSH2 c.728G>A	BC47	WES	Breast Cancer	48	1 Ovarian 1 Stomach 1 Lung	ER-/PR-	HER2+	NA
PMS2 c.1004A>T	BC22	WES	Breast Cancer	29	1 Breast 1 Pancreas 1 Skin	ER+/PR+	NA	NA
PMS2 c.2186_2187del	BC310	WES	Breast Cancer	55	5 Stomach 1 Colon 1 Breast	NA	NA	NA
RAD50 c.3036+5G>A	BC40	WES	Breast Cancer	37	1 Breast 1 Stomach	ER+/PR+	HER2+	22%

higher RMSD values of the mutant structures due to S386P (FANCC) and W122C (FANCG) compared to native proteins. This decrease in stability correlated with decrease in FANCC compactness for S386P while a high increase in compactness was observed for W122C compared to the native FANCG protein as observed in the Rg plots. The assessment of changes in stability and flexibility upon the amino acid changes S704L in ERCC3 and N335I in PMS2 revealed no significant effects on the protein structure (Table 3). Indeed, Molecular dynamic simulation results showed a slight increase in protein stability and decrease in protein compactness for S704L and N335I (Supplementary Figure S2).

For the missense variant identified within the *MSH2* gene, no further investigations were performed, as findings of the study of Kabbage et al. (2022) are suggestive of a likely pathogenic effect.

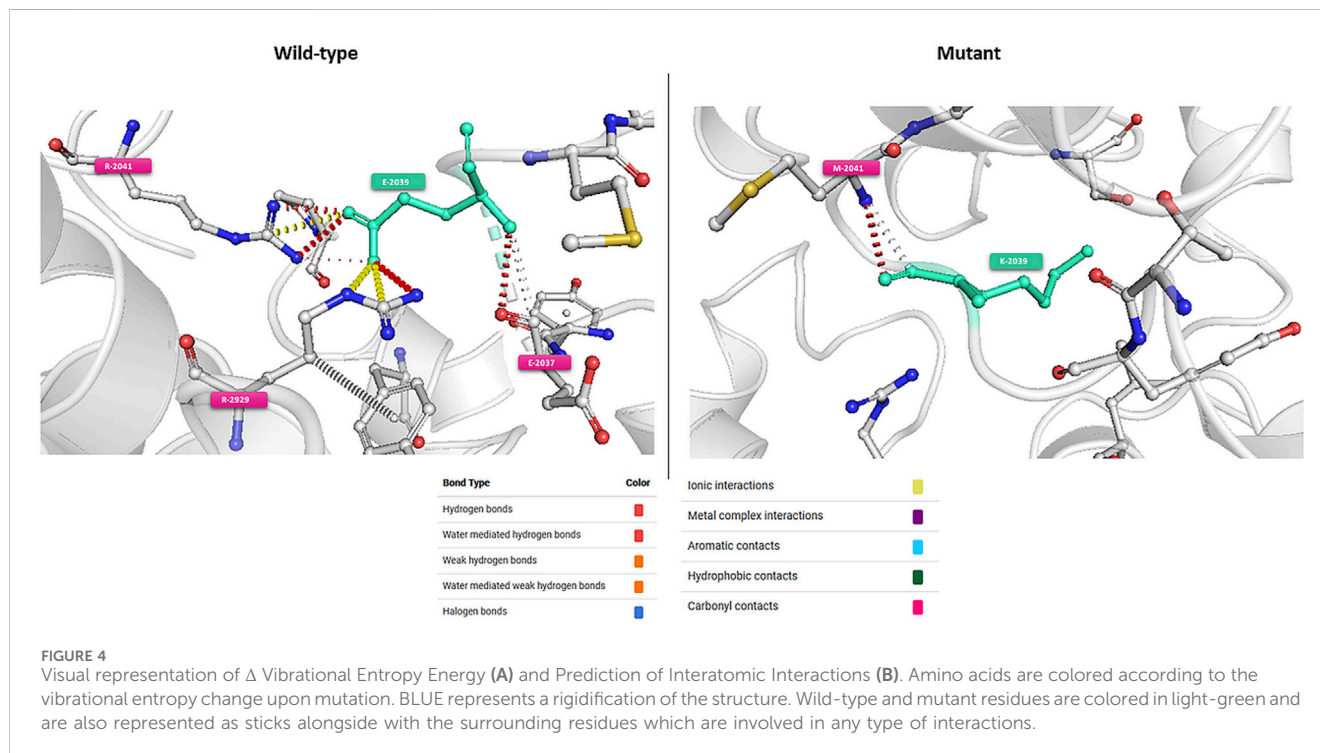## 3.4 Variant reclassification

Taking into account *in silico* analysis findings, genotype phenotype correlation and segregation analysis results a novel classification based on the American College of Medical Genetics and Genomics (ACMG) guidelines (Richards et al., 2015) was

TABLE 4 Prediction of the impact of missense variants on protein stability and flexibility.

Variant	Protein	ΔΔG DynaMut (kcal/mol)	ΔΔG ENCoM (kcal/mol)	ΔΔS ENCoM (kcal.mol-1.K-1)
E2039K	ATM	**1.903 (Stabilizing)**	**4.519 (Stabilizing)**	**−4.148 (Decrease of molecule flexibility)**
S704L	ERCC3	0.297 (Stabilizing)	−0.056 (Destabilizing)	0.069 (Increase of molecule flexibility)
S386P	FANCC	**−0.783 (Destabilizing)**	−0.320 (Destabilizing)	0.399 (Increase of molecule flexibility)
W122C	FANCG	−0.284 (Destabilizing)	**−1.032 (Destabilizing)**	**1.290 (Increase of molecule flexibility)**
N335I	PMS2	−0.245 (Destabilizing)	−0.044 (Destabilizing)	0.056 (Increase of molecule flexibility)

Bold values refer to significant effects on protein stability.

FIGURE 4
Visual representation of Δ Vibrational Entropy Energy (A) and Prediction of Interatomic Interactions (B). Amino acids are colored according to the vibrational entropy change upon mutation. BLUE represents a rigidification of the structure. Wild-type and mutant residues are colored in light-green and are also represented as sticks alongside with the surrounding residues which are involved in any type of interactions.

suggested. This novel classification was submitted in the ClinVar database under the submission ID SUB13085050. It highlights the clinical relevance of *ATM*, *BLM* and *CHEK2* variants for which strong evidence of pathogenicity were obtained. The suggested classification and criteria used to reclassify these variants are detailed in Supplementary Table S3.

4 Discussion

DNA damage repair genes play a key role in cancer biology and have critical implications in cancer diagnosis and therapy. Cancer cells are usually deficient in normal DNA repair function which causes their genomic instability. DNA repair deficiency also explains the increased sensitivity of cancer cells to genotoxic agents including ionizing radiation and chemotherapy. It is also well established that cells deficient in homologous recombination DNA repair are hypersensitive to polyADP ribose polymerase (PARP) inhibitors. Therefore, resolving the classification of variants of uncertain significance in DNA repair genes is crucial and will lead to better

clinical management of cancer (D'Andrea, 2015). Indeed, in some cases therapeutic management may change dramatically, when a variant is upgraded from VUS to pathogenic or likely pathogenic variant. Variant reclassification might allow clinicians to better counsel patients and make recommendations about appropriate medical care (Chiang et al., 2021). In the current report, 11 variants of uncertain clinical significance in *ATM*, *BLM*, *CHEK2, ERCC3, FANCC, FANCG, MSH2, PMS2* and *RAD50* genes were the most relevant based on *in silico* functional predictions and segregation analysis. Previous studies have shown that heterozygous carriers of *ATM* gene mutations have a 2-5-fold increased risk of developing breast cancer (Moslemi et al., 2021). In the present study, c.6115G>A *ATM* variant, was identified in 4 breast cancer cases belonging to 3 unrelated families all originating from the South of Tunisia suggesting a possible founder effect. A strong family history of breast cancer was clearly observed among these families and segregation analysis has confirmed the co-segregation of this variant with the disease. The presence of multiple breast cancer cases in proband's relatives in all the 3 families may also suggest a "breast cancer only phenotype"

associated with this variation. This same variation seems to be rare in other populations. Indeed, in a large meta-analysis, this variant was observed in 1/2531 breast cancer cases and was absent in 2245 controls (Tavtigian et al., 2009). *In silico* functional analysis performed in this study revealed substantial structural changes due to this amino acid substitution in the mutant protein which made it more stable, less flexible, and more compact. This may in turn alter the protein function. In fact, proteins are highly dynamic molecules, whose function is basically related to their molecular movements (Rodrigues et al., 2018). Several studies have shown that missense variants can lead to protein dysfunction by affecting their stabilities and interactions with other biomolecules. These variants are thought to be deleterious due to reducing or increasing the stability of the corresponding protein (Chen et al., 2020; Birolo et al., 2021). The negative effect of stabilizing mutations is manifested by the rigidification of cooperative movements of subunits, the deregulation of protein-protein interactions and the activity-stability trade-off which implies that an increase in activity is accompanied by a concomitant decrease in the stability of proteins (Siddiqui, 2017; Gerasimavicius et al., 2020). Accordingly, the over-stability of ATM as well as its rigidification may affect the kinase activity of the protein as well as its interactions with other biological molecules. This may disrupt its key role in repairing DNA double-strand breaks and thus promote the risk of developing cancer. All this evidence, along with familial segregation results and *in silico* functional analyses support the pathogenicity of the *ATM* c.6115G>A variant. Considering *CHEK2* gene, two relevant variants were identified, c.493_498delGAATAT and c.592+3A>T. The c.493_498delGAATAT is a novel variant described for the first time in this study, it was neither reported in public databases nor in published literature. This in-frame deletion leads to the loss of two amino acids. It is localized within the forkhead-associated (FHA) domain that is involved in binding to other downstream phosphorylated proteins. Indeed, it mediates ATM-dependent CHK2 phosphorylation and directing of CHK2 to binding partners such as BRCA1, playing hence a critical role in DNA damage response (Li et al., 2002).

The second variant in *CHEK2*, c.592+3A>T, was identified in one family with both breast and prostate cancer cases and segregation analysis has confirmed the cosegregation of the variant with the disease. This variant was predicted to lead to exon skipping and a recent RNA analysis in the study of Apostolou et al. (2021) has confirmed that it leads to exon 3 skipping. In this same study, it was shown that c.592+3A>T is recurrent and founder mutation in the Greek population rising approximately 35 generations ago (Apostolou et al., 2021). These findings along with our results support the pathogenic effect of this variant and highlight the need for variant reclassification for better risk assessment and patients' management. Indeed, women with *CHEK2* mutations have a 28%–37% lifetime risk of developing breast cancer. This risk is higher in women with a strong family history of the disease (Moslemi et al., 2021). In male carriers, the risk of prostate cancer is higher given that *CHEK2* upregulation reduces cell growth whereas its downregulation alters androgen receptor activity (Apostolou and Papasotiriou, 2017).

Disease risk associated with the identified variants must be appropriately assessed especially that prophylactic mastectomy may also be considered for *ATM* and *CHEK2* mutations carriers

depending on family history (Daly et al., 2017). Interestingly, in the current study we have identified a new homozygous likely pathogenic mutation in the *BLM* gene in a patient with multiple primary tumors, breast endometrial and colon cancers. Homozygous *BLM* mutations cause a rare autosomal recessive inherited disorder "Bloom Syndrome" that is characterized by chromosomal instability, immunodeficiency, and a predisposition to different types of malignancies, including breast and colon cancers (de Voer et al., 2015; Kluzniak et al., 2019). It was also shown that *BLM* may be a potential endometrial cancer predisposing gene which is consistent with our findings (Long et al., 2019). Other hereditary cancer syndromes predispose to cancer such as Lynch Syndrome. This latter is caused by pathogenic germline variants in Mismatch Repair Genes, *MLH1, MSH2, MSH6 or PMS2*, and predispose mainly to colorectal cancer and to other malignancies including endometrial, gastrointestinal, ovarian, pancreatic and skin cancers (Pande et al., 2012). In the current report, two variants of uncertain significance were identified in the *PMS2* gene, c.1004A>T and c.2186_2187del. This frameshift deletion is predicted to lead to the premature termination of the protein which in turn will alter the protein function. This same variant was previously reported in a homozygous state in a prostate cancer patient (Leongamornlert et al., 2014) and in a compound heterozygous state in cases with constitutional mismatch repair deficiency syndrome (CMMRD) (Bakry et al., 2014). In this CMMRD syndrome, homozygous carriers of *PMS2* gene mutation develop brain cancers in the first decade of life and 40% of patients may develop a second primary tumor (Ramchander et al., 2017). It was also previously described in a patient with colon cancer; yet, immunohistochemistry showed retained nuclear staining thus conflicting its pathogenicity (Haraldsdottir et al., 2017). In the current report, pathogenicity predictions for both *PMS2* variants were suggestive of a deleterious effect, however molecular dynamics simulation showed inconclusive results for c.1004A>T (N335I). Nevertheless, family history of both probands were consistent with lynch syndrome tumor spectrum which in turn may support the pathogenicity of the identified variants. Among the other MMR genes, we have identified a missense variant, c.728G>A in *MSH2* gene that seems to be likely pathogenic based on the findings of Kabbage et al. (2022). Indeed, this variant, which was described in a gastric cancer patient, appears to affect the MSH2-MLH1 complex as well as DNA-complex stability. Interestingly, both cases carrying this variant (current report & Kabbage et al., 2022) showed family history of cancer suggestive of lynch syndrome involving breast, gastric and ovarian malignancies which in turn may support the pathogenicity of the c.728G>A variant. In addition to MMR genes, we have identified a likely pathogenic splicing variant in *RAD50* gene that was predicted to lead to exon skipping (c.3036+5G>A). *RAD50* gene was associated with intermediate risk of developing breast cancer and updating the classification of this variant is important for better disease risk management (Damiola et al., 2014). Regarding *FANCC* and *FANCG* genes, identified variants were predicted to destabilize the protein structure and to modify its compactness. These alterations in protein structural integrity may have a significant effect on FANCC and FANCG activities. They could lead to loss of interactions among the Fanconi Anemia Core Complex alerting in consequence its important role in maintaining the genome activity. Mutations in

the *FANCC* gene have been associated with an increased risk of developing breast cancer yet they seem to be rare compared to mutations in other DNA repair genes (Fang et al., 2020). Regarding *FANCG* gene, although it is usually included in the genetic diagnostic panels for hereditary breast cancer, its contribution to the genetic susceptibility of the disease is not well defined (Del Valle et al., 2020). Finally, considering *ERCC3* gene (c.2111C>T) *in silico* predictions revealed a potential deleterious effect yet stability and molecular dynamics simulation were not suggestive of a pathogenic effect. To better elucidate the functional impact of the studied variants *in vivo* or *in vitro* analysis, such as CRISPR/Cas9 genome editing tools, are required in order to validate the pathogenicity of these variants. This is crucial especially when variants are identified in unique families or when segregation analysis could not be performed. In the presence of strong evidence of pathogenicity, it is important to reclassify variants in order to avoid misinterpretation and to ensure appropriate patient care and adequate cancer risk assessment. Indeed, reviewing variant classification is of particular medical interest since carriers of pathogenic variants could undertake prophylactic surgery and may benefit from targeted therapy. Finally, and since variant reclassification has a significant impact on clinical management, physicians need to stay up to date with variant updates. Moreover, and as recommended by the National Comprehensive Cancer Network (Chiang et al., 2021), VUS carriers are invited to re-contact their genetics service providers after a few years asking for new updates regarding the pathogenicity of the identified variants.

5 Conclusion

In the current report, we have shown that variants with uncertain interpretations of pathogenicity may explain a part of the missing heritability of breast cancer in Tunisia. Our findings were supported by clinical and family history data along with segregation analysis, as well as *in silico* predictions and structural analysis results. This was particularly the case for *ATM* and *CHEK2* variants where we have found strong evidence arguing their pathogenicity. Therefore, for a better variant interpretation and classification, it is crucial to consider genomics data of African populations. It is also important to establish expert panel groups in understudied populations for variant curation and interpretation.

Data availability statement

The data presented in the study are deposited in the ClinVar repository, accession number SUB13085050.

Ethics statement

The studies involving humans were approved by The Biomedical Ethics committee of Institut Pasteur de Tunis. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

MB: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing–original draft. FN: Writing–review and editing, Formal Analysis, Software, Methodology. NJ: Data curation, Writing–review and editing. NM: Resources, Writing–review and editing. HnB: Resources, Writing–review and editing. CC: Writing–review and editing. SB: Resources, Writing–review and editing. SL: Resources, Writing–review and editing. HE: Resources, Writing–review and editing. YB: Funding acquisition, Writing–review and editing. HR: Resources, Writing–review and editing. ND: Resources, Writing–review and editing. FB: Resources, Writing–review and editing. AH: Resources, Writing–review and editing. SA: Funding acquisition, Supervision, Writing–review and editing. MS: Funding acquisition, Supervision, Writing–review and editing. HmB: Funding acquisition, Supervision, Writing–review and editing. YH: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Writing–review and editing.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2024.1327894/full#supplementary-material

References

Adzhubei, I., Jordan, D. M., and Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* Chapter 7, Unit7.20. doi:10.1002/0471142905.hg0720s76

Apostolou, P., Dellatola, V., Papadimitriou, C., Kalfakakou, D., Fountzilas, E., Faliakou, E., et al. (2021). CHEK2 pathogenic variants in Greek breast cancer patients: evidence for strong associations with estrogen receptor positivity, overuse of risk-reducing procedures and population founder effects. *Cancers (Basel).* 13 (9), 2106. doi:10.3390/cancers13092106

Apostolou, P., and Papasotiriou, I. (2017). Current perspectives on CHEK2 mutations in breast cancer. *Breast Cancer* 9, 331–335. doi:10.2147/BCTT.S111394

Bakry, D., Aronson, M., Durno, C., Rimawi, H., Farah, R., Alharbi, Q. K., et al. (2014). Genetic and clinical determinants of constitutional mismatch repair deficiency syndrome: report from the constitutional mismatch repair deficiency consortium. *Eur. J. Cancer* 50 (5), 987–996. doi:10.1016/j.ejca.2013.12.005

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein Data Bank. *Nucleic Acids Res.* 28 (1), 235–242. doi:10.1093/nar/28.1.235

Birolo, G., Benevenuta, S., Fariselli, P., Capriotti, E., Giorgio, E., and Sanavia, T. (2021). Protein stability perturbation contributes to the loss of function in haploinsufficient genes. *Front. Mol. Biosci.* 8, 620793. doi:10.3389/fmolb.2021.620793

Chen, Y., Lu, H., Zhang, N., Zhu, Z., Wang, S., and Li, M. (2020). PremPS: predicting the impact of missense mutations on protein stability. *PLoS Comput. Biol.* 16 (12), e1008543. doi:10.1371/journal.pcbi.1008543

Chiang, J., Chia, T. H., Yuen, J., Shaw, T., Li, S. T., Binte Ishak, N. D., et al. (2021). Impact of variant reclassification in cancer predisposition genes on clinical care. *JCO Precis. Oncol.* 5, 577–584. doi:10.1200/PO.20.00399

Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., and Chan, A. P. (2012). Predicting the functional effect of amino acid substitutions and indels. *PloS one* 7 (10), e46688. doi:10.1371/journal.pone.0046688

Chun, S., and Fay, J. C. (2009). Identification of deleterious mutations within three human genomes. *Genome Res.* 19 (9), 1553–1561. doi:10.1101/gr.092619.109

Cohen-Haguenauer, O. (2019). Hereditary predisposition to breast cancer (1): genetics. *Med. Sci. Paris.* 35 (2), 138–151. doi:10.1051/medsci/2019003

Daly, M. B., Pilarski, R., Berry, M., Buys, S. S., Farmer, M., Friedman, S., et al. (2017). NCCN guidelines insights: genetic/familial high-risk assessment: breast and ovarian, version 2.2017. *J. Natl. Compr. Cancer Netw.* 15 (1), 9–20. doi:10.6004/jnccn.2017.0003

Damiola, F., Pertesi, M., Oliver, J., Le Calvez-Kelm, F., Voegele, C., Young, E. L., et al. (2014). Rare key functional domain missense substitutions in MRE11A, RAD50, and NBN contribute to breast cancer susceptibility: results from a Breast Cancer Family Registry case-control mutation-screening study. *Breast Cancer Res.* 16 (3), R58. doi:10.1186/bcr3669

D'Andrea, A. D. (2015). *DNA repair pathways and human cancer*, 47–66.e2.

Del Valle, J., Rofes, P., Moreno-Cabrera, J. M., Lopez-Doriga, A., Belhadj, S., Vargas-Parra, G., et al. (2020). Exploring the role of mutations in Fanconi anemia genes in hereditary cancer patients. *Cancers (Basel)* 12 (4), 829. doi:10.3390/cancers12040829

Desmet, F. O., Hamroun, D., Lalande, M., Collod-Beroud, G., Claustres, M., and Beroud, C. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* 37 (9), e67. doi:10.1093/nar/gkp215

Desvignes, J. P., Bartoli, M., Delague, V., Krahn, M., Miltgen, M., Beroud, C., et al. (2018). VarAFT: a variant annotation and filtration system for human next generation sequencing data. *Nucleic Acids Res.* 46 (W1), W545–W53. doi:10.1093/nar/gky471

de Voer, R. M., Hahn, M. M., Mensenkamp, A. R., Hoischen, A., Gilissen, C., Henkes, A., et al. (2015). Deleterious germline BLM mutations and the risk for early-onset colorectal cancer. *Sci. Rep.* 5, 14060. doi:10.1038/srep14060

Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K., et al. (2015). Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.* 24 (8), 2125–2137. doi:10.1093/hmg/ddu733

Ernst, C., Hahnen, E., Engel, C., Nothnagel, M., Weber, J., Schmutzler, R. K., et al. (2018). Performance of *in silico* prediction tools for the classification of rare BRCA1/2 missense variants in clinical diagnostics. *BMC Med. Genomics* 11 (1), 35. doi:10.1186/s12920-018-0353-y

Fang, C. B., Wu, H. T., Zhang, M. L., Liu, J., and Zhang, G. J. (2020). Fanconi anemia pathway: mechanisms of breast cancer predisposition development and potential therapeutic targets. *Front. Cell Dev. Biol.* 8, 160. doi:10.3389/fcell.2020.00160

Fornasarig, M., Magris, R., De Re, V., Bidoli, E., Canzonieri, V., Maiero, S., et al. (2018). Molecular and pathological features of gastric cancer in lynch syndrome and familial adenomatous polyposis. *Int. J. Mol. Sci.* 19 (6), 1682. doi:10.3390/ijms19061682

Gabler, F., Nam, S. Z., Till, S., Mirdita, M., Steinegger, M., Soding, J., et al. (2020). Protein sequence analysis using the MPI bioinformatics Toolkit. *Curr. Protoc. Bioinforma.* 72 (1), e108. doi:10.1002/cpbi.108

Gelfman, S., Wang, Q., McSweeney, K. M., Ren, Z., La Carpia, F., Halvorsen, M., et al. (2017). Annotating pathogenic non-coding variants in genic regions. *Nat. Commun.* 8 (1), 236. doi:10.1038/s41467-017-00141-2

Gerasimavicius, L., Liu, X., and Marsh, J. A. (2020). Identification of pathogenic missense mutations using protein stability predictors. *Sci. Rep.* 10 (1), 15387. doi:10.1038/s41598-020-72404-w

Girolimetti, G., Perrone, A. M., Santini, D., Barbieri, E., Guerra, F., Ferrari, S., et al. (2014). BRCA-associated ovarian cancer: from molecular genetics to risk management. *BioMed Res. Int.* 2014, 787143. doi:10.1155/2014/787143

Gonzalez, D., and Stenzinger, A. (2021). Homologous recombination repair deficiency (HRD): from biology to clinical exploitation. *Genes, Chromosomes Cancer* 60 (5), 299–302. doi:10.1002/gcc.22939

Haraldsdottir, S., Rafnar, T., Frankel, W. L., Einarsdottir, S., Sigurdsson, A., Hampel, H., et al. (2017). Comprehensive population-wide analysis of Lynch syndrome in Iceland reveals founder mutations in MSH6 and PMS2. *Nat. Commun.* 8, 14755. doi:10.1038/ncomms14755

Hart, S. N., Hoskin, T., Shimelis, H., Moore, R. M., Feng, B., Thomas, A., et al. (2019). Comprehensive annotation of BRCA1 and BRCA2 missense variants by functionally validated sequence-based computational prediction models. *Genet. Med.* 21 (1), 71–80. doi:10.1038/s41436-018-0018-4

Heczkova, M., Machackova, E., Macinga, P., Gallmeier, E., Cahova, M., Spicak, J., et al. (2019). Functional evaluation of variants of unknown significance in the BRCA2 gene identified in genetic testing. *Cancer Biol. Ther.* 20 (5), 633–641. doi:10.1080/15384047.2018.1550566

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14 (1), 33–38. doi:10.1016/0263-7855(96)00018-5

Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., et al. (2009). STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* 37, D412–D416. doi:10.1093/nar/gkn760

Joynt, A. C. M., Axford, M. M., Chad, L., and Costain, G. (2022). Understanding genetic variants of uncertain significance. *Paediatr. Child Health* 27 (1), 10–11. doi:10.1093/pch/pxab070

Kabbage, M., Ben Aissa-Haj, J., Othman, H., Jaballah-Gabteni, A., Laarayedh, S., Elouej, S., et al. (2022). A rare MSH2 variant as a candidate marker for lynch syndrome II screening in Tunisia: a case of diffuse gastric carcinoma. *Genes* 13 (8), 1355. doi:10.3390/genes13081355

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* 46 (3), 310–315. doi:10.1038/ng.2892

Kluzniak, W., Wokolorczyk, D., Rusak, B., Huzarski, T., Kashyap, A., Stempa, K., et al. (2019). Inherited variants in BLM and the risk and clinical characteristics of breast cancer. *Cancers (Basel)* 11 (10), 1548. doi:10.3390/cancers11101548

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 44 (W1), W90–W97. doi:10.1093/nar/gkw377

Landrum, M. J., Chitipiralla, S., Brown, G. R., Chen, C., Gu, B., Hart, J., et al. (2020). ClinVar: improvements to accessing data. *Nucleic Acids Res.* 48 (D1), D835–D44. doi:10.1093/nar/gkz972

Leongamornlert, D., Saunders, E., Dadaev, T., Tymrakiewicz, M., Goh, C., Jugurnauth-Little, S., et al. (2014). Frequent germline deleterious mutations in DNA repair genes in familial prostate cancer cases are associated with advanced disease. *Br. J. Cancer* 110 (6), 1663–1672. doi:10.1038/bjc.2014.30

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25 (14), 1754–1760. doi:10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25 (16), 2078–2079. doi:10.1093/bioinformatics/btp352

Li, J., Williams, B. L., Haire, L. F., Goldberg, M., Wilker, E., Durocher, D., et al. (2002). Structural and functional versatility of the FHA domain in DNA-damage signaling by the tumor suppressor kinase Chk2. *Mol. Cell* 9 (5), 1045–1054. doi:10.1016/s1097-2765(02)00527-0

Long, B., Lilyquist, J., Weaver, A., Hu, C., Gnanaolivu, R., Lee, K. Y., et al. (2019). Cancer susceptibility gene mutations in type I and II endometrial cancer. *Gynecol. Oncol.* 152 (1), 20–25. doi:10.1016/j.ygyno.2018.10.019

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20 (9), 1297–1303. doi:10.1101/gr.107524.110

Monteiro, A. N., Bouwman, P., Kousholt, A. N., Eccles, D. M., Millot, G. A., Masson, J. Y., et al. (2020). Variants of uncertain clinical significance in hereditary breast and ovarian cancer genes: best practices in functional analysis for clinical annotation. *J. Med. Genet.* 57 (8), 509–518. doi:10.1136/jmedgenet-2019-106368

Moslemi, M., Moradi, Y., Dehghanbanadaki, H., Afkhami, H., Khaledi, M., Sedighimehr, N., et al. (2021). The association between ATM variants and risk of breast cancer: a systematic review and meta-analysis. *BMC Cancer* 21 (1), 27–12. doi:10.1186/s12885-020-07749-6

Pande, M., Wei, C., Chen, J., Amos, C. I., Lynch, P. M., Lu, K. H., et al. (2012). Cancer spectrum in DNA mismatch repair gene mutation carriers: results from a hospital based Lynch syndrome registry. *Fam. Cancer* 11 (3), 441–447. doi:10.1007/s10689-012-9534-6

Phillips, J. C., Hardy, D. J., Maia, J. D. C., Stone, J. E., Ribeiro, J. V., Bernardi, R. C., et al. (2020). Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* 153 (4), 044130. doi:10.1063/5.0014475

Ramchander, N. C., Ryan, N. A., Crosbie, E. J., and Evans, D. G. (2017). Homozygous germ-line mutation of the PMS2 mismatch repair gene: a unique case report of constitutional mismatch repair deficiency (CMMRD). *BMC Med. Genet.* 18 (1), 40. doi:10.1186/s12881-017-0391-x

Reva, B., Antipin, Y., and Sander, C. (2011). Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* 39 (17), e118. doi:10.1093/nar/gkr407

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of medical genetics and genomics and the association for molecular pathology. *Genet. Med.* 17 (5), 405–424. doi:10.1038/gim.2015.30

Rodrigues, C. H., Pires, D. E., and Ascher, D. B. (2018). DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Res.* 46 (W1), W350–W5. doi:10.1093/nar/gky300

Rogers, M. F., Shihab, H. A., Mort, M., Cooper, D. N., Gaunt, T. R., and Campbell, C. (2018). FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. *Bioinformatics* 34 (3), 511–513. doi:10.1093/bioinformatics/btx536

Salgado, D., Desvignes, J. P., Rai, G., Blanchard, A., Miltgen, M., Pinard, A., et al. (2016). UMD-predictor: a high-throughput sequencing compliant system for pathogenicity prediction of any human cDNA substitution. *Hum. Mutat.* 37 (5), 439–446. doi:10.1002/humu.22965

Schwarz, J. M., Rodelsperger, C., Schuelke, M., and Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* 7 (8), 575–576. doi:10.1038/nmeth0810-575

Sheehan, M., Heald, B., Yanda, C., Kelly, E. D., Grobmyer, S., Eng, C., et al. (2020). Investigating the link between lynch syndrome and breast cancer. *Eur. J. Breast Health* 16 (2), 106–109. doi:10.5152/ejbh.2020.5198

Siddiqui, K. S. (2017). Defying the activity-stability trade-off in enzymes: taking advantage of entropy to enhance activity and thermostability. *Crit. Rev. Biotechnol.* 37 (3), 309–322. doi:10.3109/07388551.2016.1144045

Soding, J., Biegert, A., and Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 33, W244–W248. doi:10.1093/nar/gki408

Spurdle, A. B., Healey, S., Devereau, A., Hogervorst, F. B., Monteiro, A. N., Nathanson, K. L., et al. (2012). ENIGMA--evidence-based network for the interpretation of germline mutant alleles: an international initiative to evaluate risk and clinical significance associated with sequence variation in BRCA1 and BRCA2 genes. *Hum. Mutat.* 33 (1), 2–7. doi:10.1002/humu.21628

Suszynska, M., Klonowska, K., Jasinska, A. J., and Kozlowski, P. (2019). Large-scale meta-analysis of mutations identified in panels of breast/ovarian cancer-related genes - providing evidence of cancer predisposition genes. *Gynecol. Oncol.* 153 (2), 452–462. doi:10.1016/j.ygyno.2019.01.027

Tavtigian, S. V., Deffenbaugh, A. M., Yin, L., Judkins, T., Scholl, T., Samollow, P. B., et al. (2006). Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *J. Med. Genet.* 43 (4), 295–305. doi:10.1136/jmg.2005.033878

Tavtigian, S. V., Oefner, P. J., Babikyan, D., Hartmann, A., Healey, S., Le Calvez-Kelm, F., et al. (2009). Rare, evolutionarily unlikely missense substitutions in ATM confer increased risk of breast cancer. *Am. J. Hum. Genet.* 85 (4), 427–446. doi:10.1016/j.ajhg.2009.08.018

Urbina-Jara, L. K., Rojas-Martinez, A., Martinez-Ledesma, E., Aguilar, D., Villarreal-Garza, C., and Ortiz-Lopez, R. (2019). Landscape of germline mutations in DNA repair genes for breast cancer in Latin America: opportunities for PARP-like inhibitors and immunotherapy. *Genes* 10 (10), 786. doi:10.3390/genes10100786

Van Durme, J., Delgado, J., Stricher, F., Serrano, L., Schymkowitz, J., and Rousseau, F. (2011). A graphical interface for the FoldX forcefield. *Bioinformatics* 27 (12), 1711–1712. doi:10.1093/bioinformatics/btr254

Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M., and Ng, P. C. (2016). SIFT missense predictions for genomes. *Nat. Protoc.* 11 (1), 1–9. doi:10.1038/nprot.2015.123

Vijai, J., Topka, S., Villano, D., Ravichandran, V., Maxwell, K. N., Maria, A., et al. (2016). A recurrent ERCC3 truncating mutation confers moderate risk for breast cancer. *Cancer Discov.* 6 (11), 1267–1275. doi:10.1158/2159-8290.CD-16-0487

Webb, B., and Sali, A. (2016). Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci.* 86, 1–2. doi:10.1002/cpps.20

Yadav, S., and Couch, F. J. (2019). Germline genetic testing for breast cancer risk: the past, present, and future. *Am. Soc. Clin. Oncol. Educ. Book* 39, 61–74. doi:10.1200/EDBK_238987