ORIGINAL ARTICLES

Analysis of *BRCA* gene missense mutations

Stella W.S. Lai, Rebecca M. Lopes, Elaine Doherty, Debra O. Prosser, Rongying Tang, Donald R. Love

Diagnostic Genetics, Lab PLUS, Auckland City Hospital, Auckland, New Zealand.

Correspondence: Donald R. Love. Address: Diagnostic Genetics, LabPLUS, Auckland City Hospital, PO Box 110031, Auckland 1148, New Zealand. Email: donaldl@adhb.govt.nz

Received: September 2, 2015 **Accepted:** October 7, 2015 **Online Published:** October 20, 2015
DOI: 10.5430/jbei.v2n1p91 **URL:** http://dx.doi.org/10.5430/jbei.v2n1p91

Abstract

With the significant progress in sequencing technologies over the last 10 years, a concomitant increase in the detection of variants of uncertain significance (VUSs) has been reported with an increasing amount of data. The interpretation of VUSs has been challenging due to the discordance of prediction results and their classification in different locus-specific databases (LSDBs). The evolving nature of variant classification systems poses the question as to the best strategies for variant interpretation. With the increased complexity of data analysis in a clinical setting, the pathogenicity of a variant should be determined through integrating and interpreting the data as a whole. Here we demonstrate the problems that are commonly encountered when interpreting VUSs and show that data integration helps in determining the pathogenicity of a variant.

Key words

VUS interpretation, Sequence variants, *BRCA1* gene, *BRCA2* gene, *In silico*

1 Introduction

Breast cancer is the most frequently registered cancer and the second leading cause of cancer death among women in New Zealand. Compared to the second half of last century, the incidence of breast cancer has been increasing in New Zealand [1]. Germline mutations in the *BRCA1/2* genes account for approximately 10% to 15% of all breast and ovarian cancers and are known as hereditary breast and ovarian cancers (HBOC) [2, 3].

The *BRCA1/2* genes, which are tumor-suppressor genes, were identified by positional cloning in the 1990s. These genes encode for proteins that are responsible for controlling cellular growth and differentiation [3-5]. Patients who have known pathogenic mutations identified in the *BRCA1/2* genes carry a genetic predisposition to developing breast, ovarian, prostate, and/or pancreatic cancer. According to Stratton and Rahman [6], patients carrying known pathogenic mutations have a 10 to 20-fold increased risk of breast/ovarian cancer compared to those in the general population. Mutation screening of the *BRCA1/2* genes using either Sanger-based or Massively Parallel Sequencing approaches provide improved prognosis and clinical management for HBOC patients. Patients who carry known pathogenic mutations are offered enhanced surveillance strategies, chemoprevention and risk-reducing surgery [7, 8].

The majority of germline pathogenic mutations in the *BRCA1/2* genes are either nonsense or frame-shift mutations, while approximately 5% to 6% of HBOC patients in the United States are reported as carrying an "unclassified variant" (UV) or a "variant of uncertain clinical significance" (VUS) in the *BRCA1/2* genes [9]. The remaining 80% of patients carry variants that are common polymorphisms. These polymorphisms are detected in greater than 1% of the population, which are not predicted to have any impact on protein function [10].

With the increasing demand of multi-gene panel sequencing and advanced sequencing technologies, such as whole-genome sequencing (WGS) and whole-exome sequencing (WES), there has been a concomitant increase in the detection of VUSs [7, 11, 12]. The detection frequency of VUSs ranges from 2% to 21% among laboratories [9, 12, 13]. VUSs are sometimes referred to as unclassified variants (UVs). The two terminologies are interchangeable but the interpretation differs between the two. VUSs refer to variants that may or may not be previously studied and their clinical significance is unknown, whereas UVs refers to unstudied variants. VUSs can be either i) missense substitutions or in-frame deletions and insertions (IFDIs), in which the effect on protein structure and function is unknown, ii) silent substitution or intronic variants, which may potentially affect mRNA splicing, or iii) variants located in regulatory regions [10].

The findings of a VUS always complicate genetic counselling and cancer risk estimation, as the clinical interpretation remains unclear in relation to the phenotype of the patient, thus bringing challenges to family counselling and decision-making regarding preventive surgery. A retrospective study [8] has compared the risk management strategies of patients with a deleterious mutation and patients with a VUS. Patients with a VUS were observed to have a twofold lower likelihood of having risk-reducing surgery and lower rates of surveillance in their first five years of being tested.

In order to interpret the pathogenicity of UVs and VUSs, and hence their roles in tumour development, different multifactorial likelihood models have been developed and applied in order to aid the interpretation [14-17]. The multifactorial likelihood model, also known as an integrated evaluation or posterior probability model, consists of three components: prior probability of causality, combined likelihood ratios of observational data, and posterior probability of causality [14-16].

The prior probability of causality primarily focuses on analysing a VUS at the protein level by evolutionary conservation and physiochemical properties of the amino acid [14]. If the substitution is located in a highly conserved position of the protein, such as the RING and BRCT domains of *BRCA1* or the DNA-binding domain of *BRCA2*, then *in silico* prediction tools (*e.g.* Align-GVGD) can be used to calculate the prior probability of being pathogenic [15]. With respect to calculating a combined likelihood ratio of observed data, four types of information can be included that comprise the following: i) co-segregation analysis, ii) co-occurrence (in *trans*) with known deleterious variants, iii) personal and family history, and iv) histopathology of the tumour [10, 15]. Co-segregation analysis relies on genotype data from the pedigree; if most family members who develop breast cancer carry the same VUS, it is highly suggestive that this VUS is disease-causing [10, 16]. The identification of co-occurrence (in *trans*) with known deleterious variant(s) is another powerful approach as it helps exclude the pathogenicity of a VUS. Individuals who are homozygotes for pathogenic mutations in the *BRCA1* or *BRCA2* genes are embryonically lethal or develop Fanconi anaemia, respectively [10, 16]. Information regarding particular features such as the age of onset, number of cancers and the types of cancers allows comparisons to be made between families with a deleterious mutation and families with a VUS, hence establishing the likelihood of a VUS with the disease phenotype [10, 16]. Histopathological features of the tumour from VUS carriers can be compared with tumours from patients who carry known pathogenic *BRCA* gene mutations. These features include estrogen receptor (ER) status, tumor grade and cytokine status. By deriving the likelihood ratios from these data, and combining the prior probability, the posterior probability of causality can be calculated for a VUS for classification purposes (see Figure 1) [10, 16].

A number of studies have used multifactorial likelihood modelling for variant classification. Lindor *et al*. [10] combined the odds or likelihood ratios of segregation analysis results, variant co-occurrence, personal and family history and pathology profiles to calculate the posterior probability of causality for each variant. This approach led to reclassifying VUSs into five classes according to the IARC (International Agency for Research on Cancer) Working Group on Unclassified Genetic Variants (classification classes will be discussed below). Kuo *et al*. [18] used a multifactorial model that involved

segregation studies (pedigree analysis), tumour histopathology assessment and bioinformatic analysis to predict the pathogenicity of unclassified variants, followed by validation using functional assays. Walker *et al*. [19] undertook a comprehensive analysis of splice site variants using multifactorial likelihood analysis, together with family studies and *in silico* bioinformatic predictions.

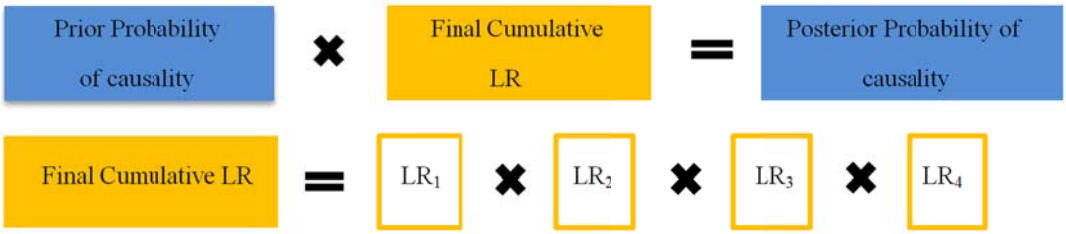

Figure 1. Calculating the posterior probability of causality. LR: likelihood ratio; Posterior Probability of causality is calculated by multiplying the prior probability of causality and the final cumulative likelihood ratios of a VUS. The Final Cumulative Likelihood Ratio is the product of LRs derived from the results of each study; each study should be an independent approach and is denoted as a subscript in the equation. Modified from Lindor *et al.* [10].

In diagnostic laboratories, the implementation of posterior probability modeling can be challenging: i) co-segregation analysis usually requires data from a large sample set in order to establish strong likelihoods to interpret a VUS to be disease-causing; ii) a VUS may be a hypomorphic variant that has subtle effect on protein function (embryonic lethality or Fanconi anemia will not be expressed); iii) the interpretation of personal and family histories varies between different pedigrees, hence different datasets are required from the families to interpret a VUS; iv) the histopathological features of tumour between VUS carriers and pathogenic mutation carriers are unclear, so further investigation of larger sample sets are required to support this correlation.

The classification system for VUSs vary slightly between countries, depending on the guidelines that the laboratory adopts [12]. The majority of variant classification systems follow a five-category system: 1) clearly not pathogenic, 2) likely not to be pathogenic, 3) uncertain significance, 4) likely to be pathogenic and 5) clearly pathogenic. The American College of Medical Genetics (ACMG) suggests a classification system with an additional category for those variants that are not expected to cause the disorder but are reported to be associated with a clinical presentation [12]. Regardless of the classification system that is used, the interpretation of VUSs and the clinical management of patients carrying a VUS remain a challenge.

Due to the limited resources of many diagnostic laboratories, the classification of VUSs usually involves three components: i) locus-specific database (LSDB) searches; ii) population database searches; and iii) performing *in silico* bioinformatic prediction analysis.

Searching LSDBs is essential in diagnostic laboratories in order to determine the clinical relevance of variants, hence providing appropriate medical surveillance. However, discrepancies exist in the classification of variants so caution is required [20, 21]. Population databases searches are recommended as the presence of a variant in the majority of the healthy population can suggest non-pathogenicity. Any variant that is present at a frequency of at least one percent in the general population is usually considered a polymorphism. The Single Nucleotide Polymorphism database (dbSNP) is one of the common population databases; however, this database includes variants that are pathogenic as well as variants with multiple classifications (*e.g.* the variant c.2612C>T in the *BRCA1* gene is listed as "benign, uncertain significance and other"). Therefore, multiple population databases should be considered.

In silico bioinformatic prediction tools are designed to predict the impact of changes in either protein function or splicing and they use different algorithms for the predictions. The algorithms of *in silico* protein bioinformatic prediction tools can

be categorised into three major groups: i) evolutionary conservation and sequence homology-based, ii) protein structure-based and iii) supervised learning [22].

Against the background described above, 29 unique missense variants (see Table 1) detected by the authors in the *BRCA1/2* genes were analysed by interrogating multiple LSDs and *in silico* prediction programmes The aims here were two-fold: first, to achieve a classification status for the 29 variants; and secondly, to determine an optimum strategy for future variant analysis.

Table 1. Summary of missense variants

***BRCA1* gene**		***BRCA2* gene**	
Nucleotide change	**Predicted Protein change**	**Nucleotide change**	**Predicted Protein change**
c.140G>A	p.(Cys47Tyr)	c.865A>C	p.(Asn289His)
c.1067A>G	p.(Gln356Arg)	c.1114A>C	p.(Asn372His)
c.1487G>A	p.(Arg496His)	c.2680G>A	p.(Val894Ile)
c.2077G>A	p.(Asp693Asn)	c.2971A>G	p.(Asn991Asp)
c.2315T>C	p.(Val772Ala)	c.4258G>T	p.(Asp1420Tyr)
c.2612C>T	p.(Pro871Leu)	c.5744C>T	p.(Thr1915Met)
c.3113A>G	p.(Glu1038Gly)	c.6100C>T	p.(Arg2034Cys)
c.3119G>A	p.(Ser1040Asn)	c.6101G>A	p.(Arg2034His)
c.3548A>G	p.(Lys1183Arg)	c.6323G>A	p.(Arg2108His)
c.4039A>G	p.(Arg1347Gly)	c.8149G>T	p.(Ala2717Ser)
c.4535G>T	p.(Ser1512Ile)	c.8215G>A	p.(Val2739Ile)
c.4837A>G	p.(Ser1613Gly)	c.8351G>A	p.(Arg2784Gln)
c.4956G>A	p.(Met1652Ile)	c.8359C>T	p.(Arg2787Cys)
c.5525T>C	p.(Val1842Ala)	c.8851G>A	p.(Ala2951Thr)
		c.9038C>T	p.(Thr3013Ile)

2 Methods

Patients were referred to Genetic Health Services New Zealand (Northern Hub) for *BRCA1/2* gene mutation screening. DNA was extracted from peripheral ethylenediaminetetraacetic acid (EDTA) blood samples using the Gentra® Puregene® Blood Kit (3 ml) (Qiagen, Venlo, Limburg, Netherlands), according to manufacturer's instructions. Informed consent underpinned the diagnostic referrals. The National Multi-Region Ethics Committee has ruled that cases of patient management do not require formal ethics committee approval. The quality and quantity of extracted gDNA were measured using a NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA).

Genomic DNA from 120 patients were subjected to *BRCA1/2* gene sequencing using Massively Parallel Sequencing (MPS) technology. Any identified variants were subsequently confirmed by bi-directional Sanger-based sequencing. Sequence data was aligned against the reference sequences NC_000017.10 (*BRCA1*; LRG_292t1; NM_007294.3) and NC_000013.10 (*BRCA2*; LRG_293t1; NM_000059.3) from the Human Genome assembly (HG19 build). HGVS v2.0 nomenclature was used to describe all variants with nucleotide numbering starting from the first nucleotide of the translated sequence.

2.1 MPS sequence data

Amplicons encompassing *BRCA1/2* gene exons with flanking intronic regions of 3-20bp upstream and downstream were analysed using SeqPilot (SeqNext module, Version 3.4.2 Build 504; JSI medical systems GmbH). Customised settings, as described in other studies, were used to achieve a Phred score equivalent of 33 [29, 30].

2.2 Sanger-based sequencing data

Amplicons encompassing *BRCA1/2* gene exons and, if necessary, 20 bp of flanking intronic DNA were analysed using commercially available software (Variant Reporter; Applied Biosystems, USA).

2.3 Pathogenicity prediction

The interrogation of databases and online bioinformatic programmes were carried out using Reference Sequences indicated above, together with RefSeq protein and Uniprot accession numbers: *BRCA1* (NP_009225.1; P35398) and *BRCA2* (NP_000050.2; P51587).

2.4 Classification based on data from Locus-specific Databases (LSDs)

Five locus-specific databases were assessed for variant classification: Breast Cancer Information Core (BIC) Database [23, 31], Human Gene Mutation Database (HGMD®) Professional [26], BRCA Share™ (formally known as Universal Mutation Database [UMD]) [27], Leiden Open Variation Database (LOVD), and ex-VUS LOVDatabase (known as LOVD-IARC) [28].

The BIC database [23] has been the leading locus-specific database for breast cancer susceptibility genes and, to date, more than 1500 variants are listed in the database as of unknown clinical significance [14, 19, 24]. This database has evolved to be one of the variant classification platforms for scientists and clinicians [25]. Prior to 2006, the pathogenicity of a variant was solely based on the submitter's data, which could be potentially biased due to insufficient data and incorrect use of the BIC Classification system; interestingly, the BIC database uses a unique nomenclature to describe each variant. The HGMD® Professional [26] is a paid subscription database that is maintained by the Institute of Medical Genetics in Cardiff, containing comprehensive mutation data with published literature and *in silico* prediction results. The LOVD is maintained by the Leiden University Medical Center, The Netherlands, in which variants are listed with dual HGVS and BIC nomenclature, together with information from the literature. BRCA Share™ (formally known as UMD) [27], is maintained by the French BRCA GGC Consortium and contains data collected from 16 French laboratories. Finally, the ex-VUS LOVDatabase (known as LOVD-IARC) [28] contains missense variants that are listed in LOVD but have been reclassified using a quantitative "posterior probability model".

For simplicity, the classification of these missense variants in five locus-specific databases were categorised as "benign", "pathogenic", "uncertain", and "not listed" (see Table 2).

Table 2. Definition of variant classifications between five locus-specific databases

Classification	HGMD® Professional	BIC	BRCA Share	LOVD	ex-VUS LOVDatabase
Benign	DP-1 DF-1 DFP-1	Not Path	1-neutral 2-likely neutral polymorphism	-/?	Class 1 Class 2
Uncertain	DM? DP-2 DFP	Unknown	3-UV	Combination of +/? , -/? and/or ?/?	Class 3
Pathogenic	DM DP FTV	Path	4-likely causal 5-causal	+/?	Class 4 Class 5
Not listed	Not listed	Not listed	Not listed	Not listed	Not listed

Note. HGMD® = Human Gene Mutation Database Professional 2015.1. Variant Classes: DM = disease causing mutation, DM? = disease causing mutation?, DP= disease-associated polymorphism; DFP = disease-associated polymorphism with additional supporting functional evidence, FTV = frameshift or truncating variant, 1 = associated with a decreased risk, 2 = functional polymorphism; BIC = breast cancer information core database. Variant Classes: Not Path = Not pathogenic, Unknown = unknown pathogenic significance, Path = pathogenic; LOVD = leiden open variant database. Variant Classification: +/? = predicted to be deleterious, -/? = predicted to be neutral,?/? = inconclusive or no comment on pathogenicity; ex-VUS LOVDatabase Variant Classes: Class 1 = no known pathogenic, Class 2 = probably no pathogenicity, Class 3 = effect unknown, Class 4 = probably pathogenic, Class 5 = pathogenic

2.5 Classification based on data from population databases

Three population databases were accessed for allele frequency data: Database of Single Nucleotide Polymorphisms (dbSNP) [32, 33], Exome Aggregation Consortium [34] and Exome Variant Server [35]. Variants listed with a minor allele frequency (MAF) of 1% or greater were assigned a benign classification. Data from the 1,000 Genomes project was available on the dbSNP database, therefore a separate entry was not carried out.

2.6 Classification based on data from *in silico* splice site bioinformatic analysis

Four *in silico* splice prediction programmes were used to check for possible splicing effects of the missense variants: the Splice Site Prediction by Neural Network online tool of the Berkeley Drosophila Genome Project [36, 37]; the Alternative Splice Site Predictor [38, 39] tool; and Human Splicing Finder [40, 41] using the prediction algorithms of HSF and MaxEnt [42]. Prediction outcomes from these programmes were compared for each variant.

2.7 Classification based on data from *in-silico* protein bioinformatic analysis

Thirteen online *in silico* protein analysis programmes were used to predict the pathogenicity of each missense variant and the prediction algorithm for each of these programmes is shown in Table 3.

Table 3. Summary of *in silico* protein prediction program algorithms used in our analysis

Online *in silico* programmes	Programmes input	Type of prediction algorithms		
		Evolutionary conservation & sequence homology	protein sequence & protein structure	Supervised learning
PolyPhen2	UniProt annotation		✓	
Mutation Assessor	RefSeq Protein ID	✓		
I-Mutant 2.0	FASTA Protein sequence		✓	
PhD SNP	UniProt annotation	✓		✓
MutPred	FASTA Protein sequence		✓	✓
SNP&GO	UniProt annotation		✓	✓
PANTHER	FASTA Protein sequence	✓		
Align-GVGD	FASTA Protein sequence	✓[1]		
SNAP	FASTA Protein sequence		✓	✓
SIFTBLink	RefSeq Protein ID	✓		
PROVEAN	RefSeq Protein ID	✓		
Mutation Taster	NCBI Gene ID	✓[2]		✓

Note. √[1] together with biophysical characteristics of the amino acid; √[2] together with data from different mutation databases. PolyPhen2 = polymorphism phenotyping v2; Mut Ass = mutation assessor; PhD-SNP = predictor of human deleterious single nucleotide polumorphisms; MutPred = application tool for classifying an amino acid substitution as disease-associated or neutral; SNPs&GO = server for predicting human disease-related mutations in proteins with functional annotations; PANTHER = protein analysis through evolutionary relationships; Align-GVGD = Align-Grantham variation grantham deviation; SNAP = predicts effect of non-synonymous polymorphisms on protein function; SIFTBLink = sorting intolerant from tolerant analysis on single protein using precomputed BLAST from NCBI Blink; PROVEAN = protein variation effect analyzer.

Polymorphism Phenotyping Version 2 (PolyPhen 2) [43] predicts the effect of a missense variant on protein structure and function, based on sequence conservation using a Naïve Bayes Classifier, with prediction outcomes as either "Probably damaging", "Possibly damaging", or "Benign" for the variant of interest. Two pairs of trained PolyPhen-2 models were available: HumDiv- and HumVar-trained models. HumDiv model predicts pathogenicity by comparing Mendelian disease variants to the divergence of close mammalian homologs of the human protein, whereas the HumVar model compares all disease-associated variants to reported benign polymorphisms. Predictions made using the HumVar model are considered more suitable for diagnostic purposes.

Align-GVGD [44, 45] predicts the pathogenicity of missense variants based on multiple protein sequence alignments and the biophysical characteristics of the amino acids. The classification ranges from Class C65 (most likely disease-causing) to Class C0 (less likely disease-causing).

Mutation assessor [46] predicts the functional effect of a missense variant on a protein based on evolutionary conservation patterns derived from multiple sequence alignments.

I-Mutant 2.0 [47, 48] assesses the stability of a missense variant based on the changes in protein sequence and structure and classifies the variant as either "neutral" or "disease" with a reliability index that ranges from 0 (less reliable) to 9 (most reliable).

MutPred [49] predicts the pathogenicity of a missense variant based on the protein sequence and structure, and classifies the change as either disease-associated (denoted as D) or neutral (denoted as N), with a probability score.

SNPs&GO [50] predicts the pathogenicity of a missense variant based on information derived from the sequence and function of a protein from the Gene Ontology (GO database). The prediction result is presented as either a neutral polymorphism or a disease-related polymorphism with a reliability index ranging from 0 (unreliable) to 10 (reliable).

Protein analysis through evolutionary relationships (PANTHER) [51, 52] predicts the functional impact of a missense variant on the protein based on the alignment of evolutionarily-related proteins, and calculates a subSPEC (substitution position-specific evolutionary conservation) score ranging from -10 (most likely to be deleterious) to 0 (neutral), while -3 is the cut-off value for functional significance.

Screening for non-acceptable polymorphism (SNAP) [53] predicts the effect of a missense variant on protein function and structural annotation, which classifies the variant as non-neutral or neutral with reliability index and accuracy calculated.

Predictor of human Deleterious Single Nucleotide Polymorphisms (PhD-SNP) [54] works in a similar fashion iMutant 2.0 as it assesses sequence homology to classify a missense variant as disease-related (Disease) or a neutral polymorphism (Neutral), with a reliability index.

Protein variation effect analyser (PROVEAN) [55] predicts the functional impact of a missense variant. The PROVEAN Human Protein Batch tool compares homologous sequences between human and mouse and generates a PROVEAN score with a predefined threshold of -2.5. A deleterious prediction corresponds to a PROVEAN score of less than or equal to -2.5, otherwise it is considered neutral. This programme also provides a prediction based on the SIFT algorithm.

Sorting intolerant from tolerant (SIFT) blink [56] predicts the pathogenicity of a missense variant based on the sequence homology from multiple sequence alignments, and a conservation value and scaled probability are calculated. The variants are classified as either "tolerated" or "affect protein function" with a SeqRep score. This score refers to the fraction of sequences containing one of the basic amino acids. Poorer predictions are made from unaligned sequences or are severely gapped for the position of interest; the poorer the predictions, the lower the SeqRep scores.

MutationTaster2 [57] classifies missense variants as either neutral or disease-causing, with an associated *P* value. The *P* value refers to the probability of the prediction, which is not the probability of error as used in *t*-test statistics. The closer to the value of 1, the higher the security of the prediction, but not the reliability of the prediction; incorrect predictions are not usually reflected by low probabilities. This programme assesses evolutionary conservation and integrates data from different databases: dbSNP, 1,000 Genome, ClinVar and HGMD® Pro, in order to provide a comprehensive analysis of the variant.

3 Results

3.1 Searching Locus-Specific Databases (LSDs)

Twenty-nine missense variants were checked for pathogenicity in five locus-specific databases. Detailed data are given in Table 4. The results for each variant are shown graphically in Figure 2 to provide a visual summary regarding the classification of each variant.

Table 4. Classification of 29 *BRCA1/2* gene missense variants in five locus-specific databases

***BRCA1* gene**							
Nucleotide	**Predicted Protein**	**HGMD® Professional 2015.1**	**BIC**	**BRCA Share**	**LOVD Database**	**ex-VUS LOVDatabase**	**Nucleotide**
c.140G>A	p.(Cys47Tyr)	DM	Not listed	5 - Causal	Not listed	Not listed	
c.1067A>G	p.(Gln356Arg)	DP	Unknown	1 - Neutral	Mixed	-/? & ?/? &+/?	1
c.1487G>A	p.(Arg496His)	DM?	Unknown	1 - Neutral	Mixed	-/? & ?/?	1
c.2077G>A	p.(Asp693Asn)	DP	Not Path	1 - Neutral	Mixed	-/? & ?/? &+/?	1
c.2315T>C	p.(Val772Ala)	DM	Unknown	1 - Neutral	Mixed	-/? & ?/? &+/?	1
c.2612C>T	p.(Pro871Leu)	DFP-1	Not Path	1 - Neutral	Mixed	-/? & ?/? &+/?	Not listed
c.3113A>G	p.(Glu1038Gly)	DP	Not Path	1 - Neutral	Mixed	-/? & ?/? &+/?	1
c.3119G>A	p.(Ser1040Asn)	DM?	Unknown	1 - Neutral	Mixed	-/? & ?/?	1
c.3548A>G	p.(Lys1183Arg)	DP-1	Not Path	1 - Neutral	Mixed	-/? & ?/? &+/?	1
c.4039A>G	p.(Arg1347Gly)	DM?	Unknown	1 - Neutral	Mixed	?/?, -/?, +/?	1
c.4535G>T	p.(Ser1512Ile)	DM?	Not Path	1 - Neutral	Mixed	-/? & ?/?	1
c.4837A>G	p.(Ser1613Gly)	DM?	Not Path	1 - Neutral	Mixed	?/?, -/?, +/?	1
c.4956G>A	p.(Met1652Ile)	DM?	Unknown	1 - Neutral	Mixed	?/?, -/?, +/?	1
c.5525T>C	p.(Val1842Ala)	Not listed	Not listed	Not listed	Not listed		Not listed
***BRCA2* gene**							
Nucleotide	**Predicted Protein**	**HGMD® Professional 2015.1**	**BIC**	**BRCA Share**	**LOVD Database**	**ex-VUS LOVDatabase**	**Nucleotide**
c.865A>C	p.(Asn289His)	DP-1	Not Path	1 - Neutral	Mixed	-/? & ?/? &+/?	Not listed
c.1114A>C*	p.(Asn372His)	DFP	Not listed	1 - Neutral	Mixed	-/? & ?/?	Not listed
c.2680G>A	p.(Val894Ile)	Not listed	Unknown	2 - Likely Neutral	Neutral	-/?	1
c.2971A>G	p.(Asn991Asp)	DM?	Not Path	Polymorphism	Mixed	-/? & ?/? &+/?	Not listed
c.4258G>T	p.(Asp1420Tyr)	DM?	Not Path	1 - Neutral	Mixed	-/? & ?/? &+/?	1
c.5744C>T	p.(Thr1915Met)	Unknown	Unknown	1 - Neutral	Mixed	-/? & ?/?	Not listed
c.6100C>T	p.(Arg2034Cys)	DM?	Unknown	1 - Neutral	Mixed	-/? & ?/?	1
c.6101G>A	p.(Arg2034His)	DM?	Unknown	Not listed	Not listed		Not listed
c.6323G>A	p.(Arg2108His)	DM?	Unknown	1 - Neutral	Mixed	-/? & ?/?	1
c.8149G>T	p.(Ala2717Ser)	DM?	Not Path	1 - Neutral	Mixed	-/? & ?/?	1
c.8215G>A	p.(Val2739Ile)	Not listed	Not listed	3 - UV	Mixed	-/? & ?/?	Not listed
c.8351G>A	p.(Arg2784Gln)	DM	Unknown	3 - UV	Pathogeni	+/?	Not listed
c.8359C>T	p.(Arg2787Cys)	Not listed	Unknown	3 - UV	Pathogeni	+/?	Not listed
c.8851G>A	p.(Ala2951Thr)	DM?	Not Path	1 - Neutral	Mixed	-/? & ?/?	Not listed
c.9038C>T	p.(Thr3013Ile)	DM?	Not Path	1 - Neutral	Mixed	-/? & ?/?	Not listed

Note. HGMD® = human gene mutation database professional 2015. Variant Classes: DM = disease causing mutation, DM? = disease causing mutation?, DP= disease-associated polymorphism; DFP = disease-associated polymorphism with additional supporting functional evidence, 1 = associated with a decreased risk; BIC = breast cancer information core database. Variant Classes: Not Path = not pathogenic, Unknown = unknown pathogenic significance, Path = pathogenic; LOVD = leiden open variant database. Variant Classification: +/? = predicted to be deleterious, -/? = predicted to be neutral, ?/? = inconclusive or no comment on pathogenicity.

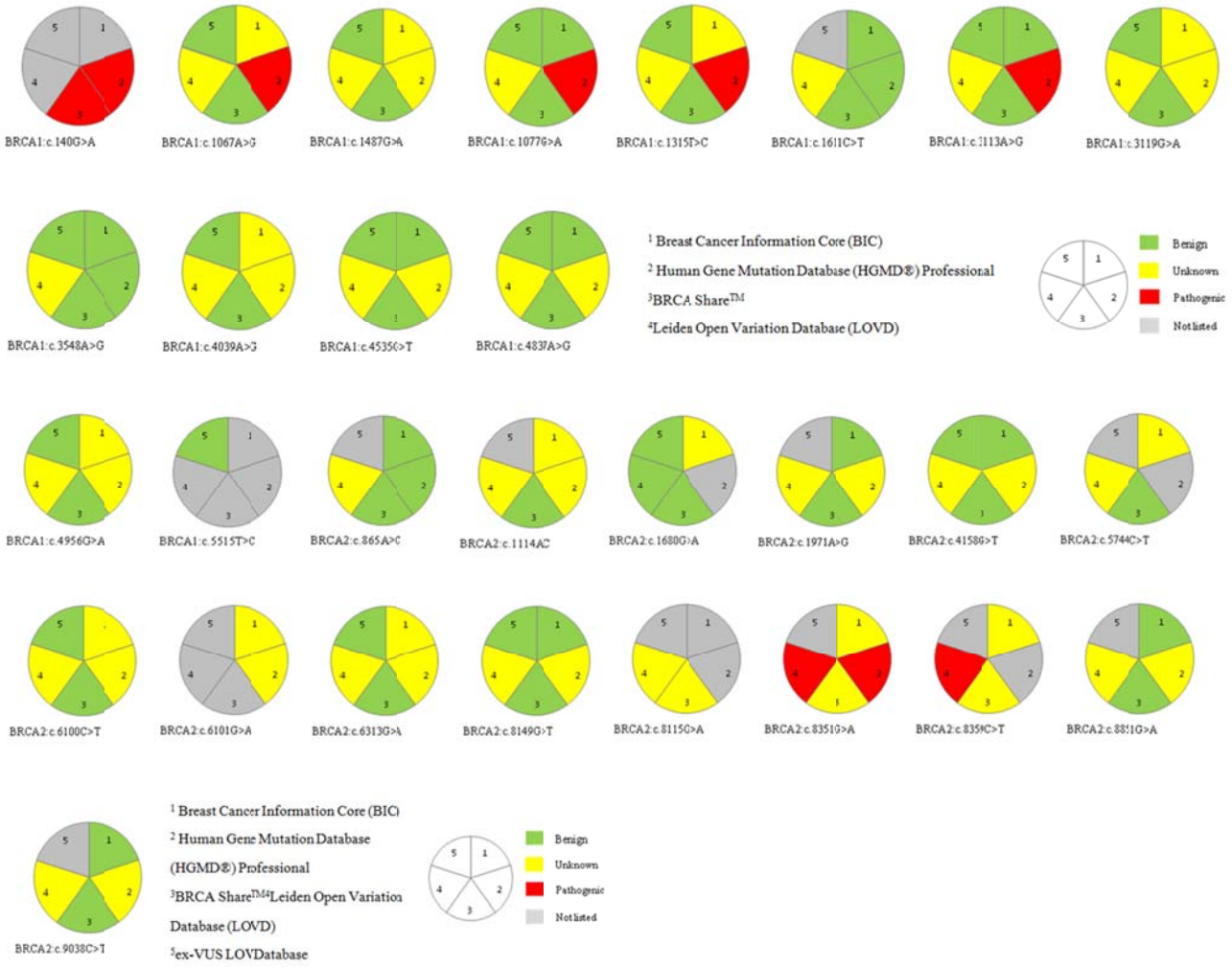

Figure 2. Diagrammatic representation of missense variant classifications in five locus-specific databases

3.2 Searching population databases

Three population databases were assessed for the minor allele frequency of each missense variant. Of the 29 variants, 12 variants could be classified as "likely to be benign" based on their allele frequency. The results for these variants in each database are summarised in Table 5. Detailed data are provided in Table 6.

Table 5. Summary of data from three population databases

	Number of variants		
	dbSNP	**ExAC**	**EVS**
Not listed or N/A	10	3	4
MAF >1%	10	12	13
MAF <1%	9	14	12

Note. dbSNP: Database of Single Nucleotide Polymorphisms [32, 33]; ExAC: Exome Aggregation Consortium [34]; EVS: Exome Variant Server [35].

3.3 *In silico* splice site bioinformatic analysis

As mentioned earlier, four *in silico* splice site prediction programmes were used to assess possible splicing effects of the missense variants. None of the variants was predicted to result in a splicing effect (data not shown).

3.4 *In silico* protein bioinformatic analysis

Thirteen *in silico* protein prediction programmes were used to predict the pathogenicity of the missense variants, the results are shown graphically in Figure 3. Interestingly, certain *in silico* protein prediction programmes (*e.g.* SNPs&GO) appear to overestimate the pathogenicity of a variant (see Figure 4).

Table 6. Data (Minor allele frequency) of 29 missense variants from three population databases

***BRCA1* gene**				***BRCA2* gene**			
Nucleotide	**dbSNP**	**ExAC**	**EVS**	**Nucleotide**	**dbSNP**	**ExAC**	**EVS**
c.140G>A	N/A	Not listed	Not listed	c.865A>C	7.37%	5.18%	3%
c.1067A>G	2.18%	4.41%	4.59%	c.1114A>C	24.94%	27.79%	23%
c.1487G>A	N/A	0.05%	0.06%	c.2680G>A	N/A	0.00%	0.038%
c.2077G>A	3.35%	5.68%	5.43%	c.2971A>G	8.01%	5.34%	4%
c.2315T>C	N/A	0.01%	0.02%	c.4258G>T	0.40%	0.68%	5%
c.2612C>T	45.61%	41.00%	49.32%	c.5744C>T	0.86%	1.79%	2%
c.3113A>G	45.61%	34.29%	27.90%	c.6100C>T	0.14%	0.32%	0.40%
c.3119G>A	0.98%	1.32%	1.65%	c.6101G>A	N/A	0.00%	0.40%
c.3548A>G	33.57%	34.90%	29.52%	c.6323G>A	0.38%	0.13%	0.031%
c.4039A>G	0.06%	0.40%	0.48%	c.8149G>T	0.06%	0.12%	0.15%
c.4535G>T	0.06%	0.22%	0.28%	c.8215G>A	N/A	0.00%	N/A
c.4837A>G	35.58%	34.96%	29.82%	c.8351G>A	N/A	0.00%	0.02%
c.4956G>A	1.12%	1.76%	1.08%	c.8359C>T	N/A	N/A	N/A
c.5525T>C	Not listed	Not listed	Not listed	c.8851G>A	0.01%	0.79%	0.44%
				c.9038C>T	N/A	0.02%	0.046%

Note. N/A = minor allele frequency not available; dbSNP = Database of Single Nucleotide Polymorphisms (dbSNP, 2015; Sherry *et al.*, 2001), ExAC = Exome Aggregation Consortium (ExAC, 2015); EVS = Exome Variant Server (EVS, 2015). Minor allele frequency (MAF) of each variant is presented as a percentage for direct comparison. MAF value of greater than 1% is highlighted in grey.

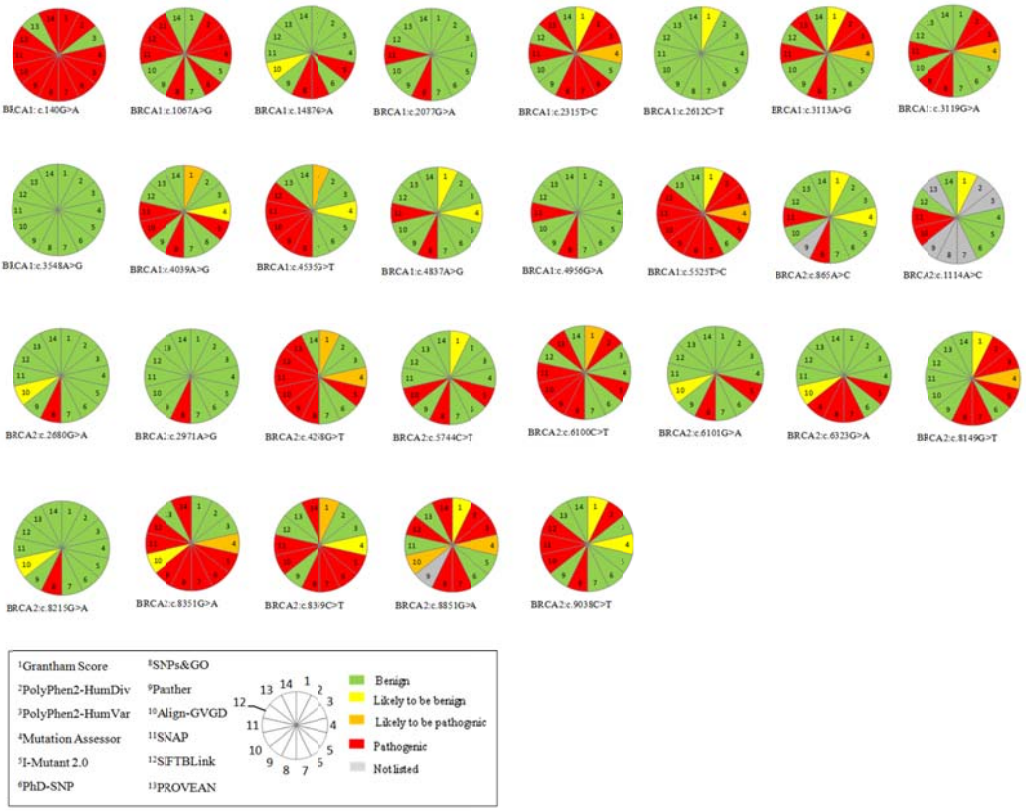

Figure 3. Prediction outcomes using 13 *in silico* protein bioinformatic programmes, together with Grantham Score

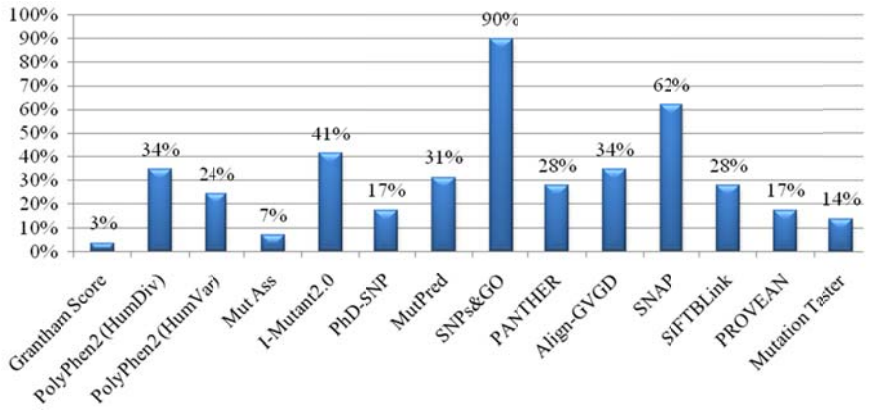

Figure 4. Percentage of 29 missense variants predicted to be pathogenic using *in silico* protein prediction programmes. PolyPhen2 = polymorphism phenotyping v2; Mut Ass = mutation assessor; PhD-SNP = predictor of human deleterious single nucleotide polumorphisms; MutPred = application tool for classifying an amino acid substitution as disease-associated or neutral; SNPs&GO = server for predicting human disease-related mutations in proteins with functional annotations; PANTHER = protein analysis through evolutionary relationships; Align-GVGD = Align-Grantham variation grantham deviation; SNAP = predicts effect of non-synonymous polymorphisms on protein function; SIFTBLink = sorting intolerant from tolerant analysis on single protein using precomputed BLAST from NCBI Blink; PROVEAN = protein variation effect analyzer.

3.5 Integrated data analysis

A simplistic approach was taken to represent the diversity of "calls" that were made for the 29 missense variants. This approach gave equal weight to each "call" that was made in LSDs, population databases and *in silico* predictions. The "calls" were assigned to one of three classifications, as shown in Table 7. The results of this analysis are shown in Table 8. In the case of *BRCA1*: c.140G>A, 23% of all calls were assigned a classification of "benign" while 54% were classified as "pathogenic".

Table 7. Definition of classifications for integrating data from different categories

Classification	LSD	Population Data	*In silico* splicing prediction results	*In silico* protein prediction results
Benign	Benign	MAF >1%	No effect on splicing	Benign
Uncertain	Uncertain	Not listed or MAF <1%	-	Likely to be benign and likely to be pathogenic
Pathogenic	Pathogenic	-	May affect splicing	Pathogenic

Note. LSD= locus-specific databases; MAF = minor allele frequency

The integrated data analysis results highlight the discordant prediction outcomes in classifying variants, which led us to the following questions:

1) Should VUS classification rely on information obtained from one database? If not, then how many and which locus-specific database should be accessed for classification?

2) What is the ideal number, and type, of *in silico* prediction program to use for classifying *BRCA1/2* gene variants?

In order to answer these questions, further data analysis was carried out for the 29 missense variants. First, the data entries from five locus-specific databases were compared to determine the "gold standard" database for variant classification (see Figure 5). The BIC database [23, 31] was considered at the outset of the research presented here to be the principal

database; however, during the course of this study, data in the BIC database was found to be "out-of-date". The HGMD® Professional was found to over-score the variants, with common SNPs listed as mutations [58].

Table 8. Summary of integrated data for 29 missense variants

Nucleotide change	Classification			
	Benign	Uncertain	Pathogenic	Not listed
BRCA1:c.140G>A	23%	0%	54%	23%
BRCA1:c.1067A>G	58%	8%	35%	0%
BRCA1:c.1487G>A	62%	23%	12%	4%
BRCA1:c.2077G>A	85%	4%	12%	0%
BRCA1:c.2315T>C	42%	23%	31%	4%
BRCA1:c.2612C>T	88%	8%	0%	4%
BRCA1:c.3113A>G	65%	12%	23%	0%
BRCA1:c.3119G>A	62%	19%	19%	0%
BRCA1:c.3548A>G	96%	4%	0%	0%
BRCA1:c.4039A>G	54%	31%	15%	0%
BRCA1:c.4535G>T	54%	27%	19%	0%
BRCA1:c.4837A>G	77%	15%	8%	0%
BRCA1:c.4956G>A	81%	12%	8%	0%
BRCA1:c.5525T>C	31%	8%	35%	27%
BRCA2:c.865A>C	73%	12%	8%	8%
BRCA2:c.1114A>C	46%	15%	8%	31%
BRCA2:c.2680G>A	73%	15%	4%	8%
BRCA2:c.2971A>G	85%	8%	4%	4%
BRCA2:c.4258G>T	50%	23%	27%	0%
BRCA2:c.5744C>T	65%	15%	12%	8%
BRCA2:c.6100C>T	46%	27%	27%	0%
BRCA2:c.6101G>A	58%	19%	8%	15%
BRCA2:c.6323G>A	58%	27%	15%	0%
BRCA2:c.8149G>T	54%	27%	19%	0%
BRCA2:c.8215G>A	62%	15%	4%	19%
BRCA2:c.8351G>A	31%	23%	38%	8%
BRCA2:c.8359C>T	35%	15%	31%	19%
BRCA2:c.8851G>A	38%	31%	23%	8%
BRCA2:c.9038C>T	50%	23%	19%	8%

Six of the 29 missense variants were listed as disease-causing mutations in the HGMD® Professional database. BRCA Share™ classified the majority of the variants, with only 5 of the 29 missense variants recorded as "classification unknown" (comprising those of "uncertain significance" and "not listed"). The data available from BRCA Share™, based on the French population, is not a true reflection of the general population. Only three of the 29 missense variants were not listed in the LOVD database; however, more than three-quarters of the variants were listed as "unknown clinical significance", so this database was of limited use. In contrast, the ex-VUS LOVDatabase, which uses a rigorous posterior-probability approach, classified 17 of the 29 missense variants as benign; the remaining 12 missense variants were unreported in this database. Due to the perceived clarity of the classifications made in the ex-VUS LOVDatabase, it was considered to be the "gold standard" database for variant classification.

 ISSN 2377-9381 E-ISSN 2377-939X

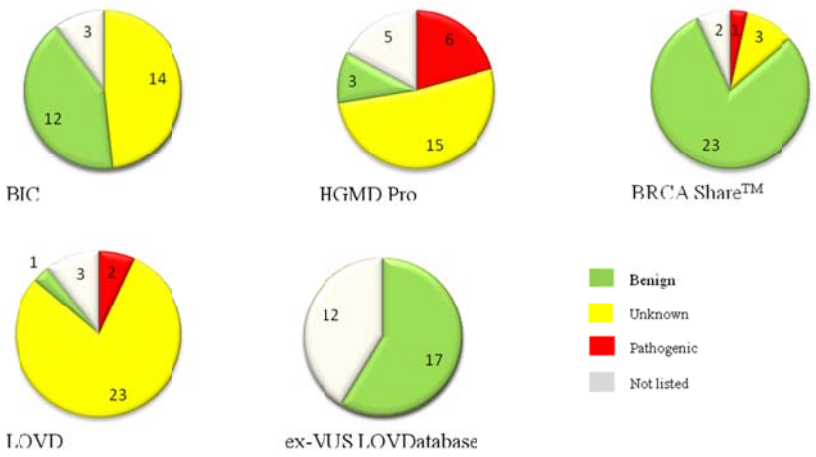

Figure 5. Classification categories for 29 missense *BRCA1/2* gene variants recorded in five locus-specific databases

3.6 A comparison of prediction results against entries in the ex-VUS LOVDatabase

In order to resolve the classification of the 12 missense variants that were not present in the ex-VUS LOVDatabase, it was decided to determine which *in silico* protein prediction programmes yielded variant classifications that were consistent with those reported in the ex-VUS LOVDatabase. In this way it was thought that unreported variants in the ex-VUS LOVDatabase could be confidently assigned a classification category. The splicing effect and population data were excluded from this strategy as all splicing predictions were uninformative for all variants and the weight given to using minor allele frequency for variant classification was unclear at the time of this study.

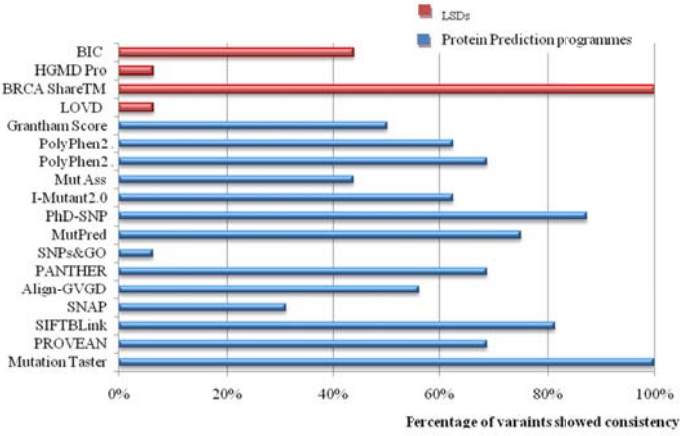

Figure 6. Percentage of variants that showed classifications that were consistent with those reported in the ex-VUS LOVDatabase. BIC = breast cancer information core; HGMD® Pro = human genetic mutation database professional 2015; LOVD = leiden open variant database; PolyPhen2 = polymorphism phenotyping v2; Mut Ass = mutation assessor; PhD-SNP = predictor of human deleterious single nucleotide polumorphisms; MutPred = application tool for classifying an amino acid substitution as disease-associated or neutral; SNPs&GO= server for predicting human disease-related mutations in proteins with functional annotations; PANTHER= protein analysis through evolutionary relationships; Align-GVGD = Align-Grantham variation grantham deviation; SNAP = predicts effect of non-synonymous polymorphisms on protein function; SIFTBLink = sorting intolerant from tolerant analysis on single protein using precomputed BLAST from NCBI Blink; PROVEAN = protein variation effect analyzer.

Pursuing the above strategy indicated that the BRCA Share™ database showed 100% consistency with the classification of reported variants in the ex-VUS LOVDatabase. In addition, the predictions from four protein bioinformatic programmes showed a high degree of consistency with the classification of reported variants in the ex-VUS LOVDatabase: PhD-SNP (88%), MutPred (75%), SIFTBLink (81%) and Mutation Taster (100%). These four programmes covered the three types of protein prediction tools: sequence/evolutionary conservation-based, protein-structure-based and supervised learning. Figure 6 graphically shows the percentage of variants with classifications consistent with those reported in the ex-VUS LOVDatabase.

3.7 Integrating the data from selected complementary database and *in silico* protein prediction programmes

For those variants not listed in the ex-VUS LOVDatabase, data from BRCA ShareTM and the classification provided by the *in silico* protein prediction programmes (PhD-SNP, MutPred, SIFTBLink and Mutation Taster) were combined to provide an the interpretation of these variants. The results are shown graphically in Figure 7. This approach classified one variant to be pathogenic, four variants to be benign and two variants to be "likely benign", leaving four variants of uncertain significance (*BRCA1*:c.5525T>C, *BRCA2*:c.8359C>T, *BRCA2*:c.8851G>A, and *BRCA2*:c.9038C>T).

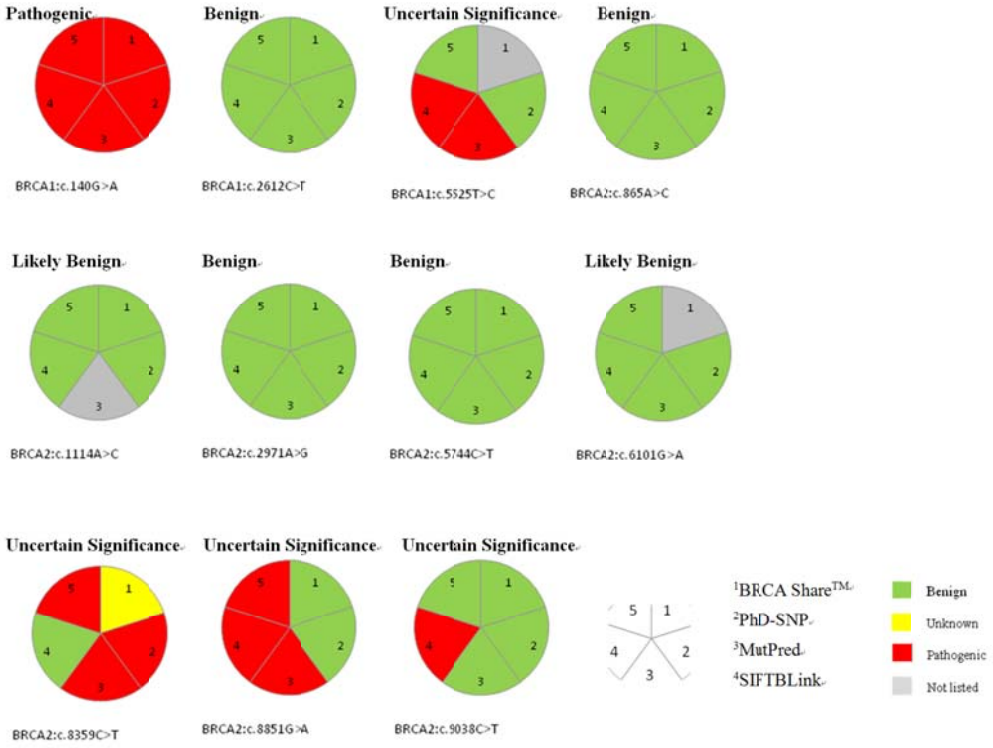

Figure 7. Reclassification of variants using BRCA Share™ database and four *in silico* protein prediction programs. SIFTBLink = sorting intolerant from tolerant analysis on single protein using precomputed BLAST from NCBI Blink; PhD-SNP = predictor of human deleterious single nucleotide polumorphisms; MutPred = application tool for classifying an amino acid substitution as disease-associated or neutral.

4 Discussions

Here we demonstrate the problems commonly encountered when interpreting VUSs: the disparity between databases and the discordance in prediction results. Data integration was shown to help in determining the pathogenicity of a variant.

However, the analysis was carried out using a relatively small sample size and was limited to missense variants. Further analysis should be undertaken using a larger sample set with different types of VUSs. Furthermore, the control group was not well established in the analysis. Control cohorts using clearly pathogenic and clearly benign classifications could be considered to determine those *in silico* protein prediction programs that should be used for classifying variants in the *BRCA1/2* genes. As has been described elsewhere [22], the most appropriate repertoire of *in silico* programmes to use must be determined for each gene.

In the diagnostic environment, data from locus-specific databases and *in silico* prediction programmes are given weight in establishing the classification of a VUS and hence aiding clinicians in supporting a diagnosis and for subsequent predictive testing in family members of the proband. Population data has gradually been implemented as part of the analysis approach in diagnostic laboratories; however, allele frequency data are ethnic-specific. Furthermore, while data can be aggregated from a disease cohort, caution should be taken during interpretation.

The current classification approaches using population allele frequencies, entries in disease databases and computational analysis cannot always clearly classify missense variants. Segregation data, as well as functional data, would be beneficial to assist in the interpretation of the clinical significance of variants.

The newly introduced online visualisation tool, *BRCA1*Circos [59], might change the face of the current analysis approach in diagnostic laboratories. This tool compiles and displays all the functional data for all documented variants in the *BRCA1* gene, which allows direct comparisons between functional data and strengthens the classification system of VUSs. Furthermore, an international collaboration by the ENIGMA (Evidence-Based Network for the Interpretation of Germline Mutant Alleles) Consortium has been established to facilitate studies of the clinical significance of VUSs [24]. The consortium comprises six working groups focusing on either: VUS interpretation for cancer risk, VUS classification in relation to clinical details, ENIGMA database maintenance, functional assays for VUS, histopathological studies of VUS, and large-scale splicing studies. Recently published guidelines by the American College of Medical Genetics (ACMG) has introduced a comprehensive evaluation system for variant interpretation [60]. The system involves assessing the strength of all the available evidence and integrating it to classify a sequence variant by following pre-defined criteria. Furthermore, two different systems have been suggested to classify variants as “pathogenic or likely pathogenic” and “benign or likely benign”. This published system reflects the increasing complexity of data analysis in a clinical setting, and suggests that the pathogenicity of VUSs should be determined through integrating and interpreting the data as a whole. With increasingly accessible functional data, the multidisciplinary approach by the ENGIMA consortium, and a more comprehensive classification system, determining the pathogenicity of VUS should improve in the near-future.

References

[1] Cunningham R, Shaw C, Blakely T, *et al*. Ethnic and socioeconomic trends in breast cancer incidence in New Zealand. BMC Cancer. 2010; 10: 674. PMid:21138590. http://dx.doi.org/10.1186/1471-2407-10-674

[2] Michils G, Hollants S, Dehaspe L, *et al*. Molecular analysis of the breast cancer genes *BRCA1* and *BRCA2* using amplicon-based massive parallel pyrosequencing. J Mol Diagn. 2012; 14(6): 623-30. PMid:23034506. http://dx.doi.org/10.1016/j.jmoldx.2012.05.006

[3] Taylor MR. Genetic testing for inherited breast and ovarian cancer syndromes: important concepts for the primary care physician. Postgrad Med J. 2001; 77(903): 11-5. PMid:11123386. http://dx.doi.org/10.1136/pmj.77.903.11

[4] Miki Y, Swensen J, Shattuck-Eidens D, *et al*. A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*. Science. 1994; 266: 66-71. PMid:7545954. http://dx.doi.org/10.1126/science.7545954

[5] Tavtigian SV, Simard J, Rommens J, *et al*. The complete *BRCA2* gene and mutations in chromosome 13q-linked kindreds. Nature genetics. 1996; 12(3): 333-7. PMid:8589730. http://dx.doi.org/10.1038/ng0396-333

[6] Stratton MR, Rahman N. The emerging landscape of breast cancer susceptibility. Nature genetics. 2008; 40(1): 17-22. PMid:18163131. http://dx.doi.org/10.1038/ng.2007.53

[7] Culver JO, Brinkerhoff CD, Clague J, *et al*. Variants of uncertain significance in BRCA testing: evaluation of surgical decisions, risk perception, and cancer distress. Clinical genetics. 2013; 84(5): 464-72. PMid:23323793. http://dx.doi.org/10.1111/cge.12097

[8] Garcia C, Lyon L, Littell RD, *et al*. Comparison of risk management strategies between women testing positive for a BRCA variant of unknown significance and women with known BRCA deleterious mutations. Genet Med. 2014; 16(12): 896-902. PMid:24854227. http://dx.doi.org/10.1038/gim.2014.48

[9] Lindor NM, Goldgar DE, Tavtigian SV, *et al*. *BRCA1/2* Sequence Variants of Uncertain Significance: A Primer for Providers to Assist in Discussions and in Medical Management. The Oncologist. 2013; 18(5): 518-24. PMid:23615697. http://dx.doi.org/10.1634/theoncologist.2012-0452

[10] Lindor N, Guidugli L, Wang X, *et al*. A review of a multifactorial probability-based model for classification of BRCA1 and BRCA2 variants of uncertain significance (VUS). Hum Mutat. 2012; 33(1): 8-21. PMid:21990134 . http://dx.doi.org/10.1002/humu.21627

[11] Frebourg T. The Challenge for the Next Generation of Medical Geneticists. Human Mutation. 2014; 35(8): 909-11. PMid:24838402. http://dx.doi.org/10.1002/humu.22592

[12] Cheon J, Mozersky J, Cook-Deegan R. Variants of uncertain significance in BRCA: a harbinger of ethical and policy issues to come? Genome Medicine. 2014; 6(12): 121. PMid:25593598. http://dx.doi.org/10.1186/s13073-014-0121-3

[13] Eggington J, Bowles K, Moyes K, *et al*. A comprehensive laboratorybased program for classification of variants of uncertain significance in hereditary cancer genes. Clinical genetics. 2014; 86: 229-37. PMid:24304220. http://dx.doi.org/10.1111/cge.12315

[14] Easton DF, Deffenbaugh AM, Pruss D, *et al*. A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes. American journal of human genetics. 2007; 81(5): 873-83. PMid:17924331. http://dx.doi.org/10.1086/521032

[15] Goldgar DE, Easton DF, Deffenbaugh AM, *et al*. Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. American journal of human genetics. 2004; 75(4): 535-44. PMid:15290653. http://dx.doi.org/10.1086/424388

[16] Goldgar DE, Easton DF, Byrnes GB, *et al*. Genetic evidence and integration of various data sources for classifying uncertain variants into a single model. Hum Mutat. 2008; 29(11): 1265-72. PMid:18951437. http://dx.doi.org/10.1002/humu.20897

[17] Spurdle AB. Clinical relevance of rare germline sequence variants in cancer genes: evolution and application of classification models. Curr Opin Genet Dev. 2010; 20(3): 315-23. PMid:20456937. http://dx.doi.org/10.1016/j.gde.2010.03.009

[18] Kuo WH, Lin PH, Huang AC, *et al*. Multimodel assessment of BRCA1 mutations in Taiwanese (ethnic Chinese) women with early-onset, bilateral or familial breast cancer. Journal of human genetics. 2012; 57(2): 130-8. PMid:22277901. http://dx.doi.org/10.1038/jhg.2011.142

[19] Walker LC, Whiley PJ, Couch FJ, *et al*. Detection of splicing aberrations caused by *BRCA1* and *BRCA2* sequence variants encoding missense substitutions: implications for prediction of pathogenicity. Hum Mutat. 2010; 31(6): E1484-505. PMid:20513136. http://dx.doi.org/10.1002/humu.21267

[20] Vail PJ, Morris B, van Kan A, *et al*. Comparison of locus-specific databases for *BRCA1* and *BRCA2* variants reveals disparity in variant classification within and among databases. Journal of community genetics. 2015. PMid:25782689. http://dx.doi.org/10.1007/s12687-015-0220-x

[21] Brookes C, Lai S, Doherty E, *et al*. Predicting the Pathogenic Potential of *BRCA1* and *BRCA2* Gene Variants Identified in Clinical Genetic Testing. Sultan Qaboos University medical journal. 2015; 15(2): e218-25. PMid:26052455.

[22] Leong IU, Stuckey A, Lai D, *et al*. Assessment of the predictive accuracy of five *in silico* prediction tools, alone or in combination, and two metaservers to classify long QT syndrome gene mutations. BMC Med Genet. 2015; 16(1): 34. PMid:25967940. http://dx.doi.org/10.1186/s12881-015-0176-z

[23] Szabo C, Masiello A, Ryan JF, *et al*. The Breast Cancer Information Core: Database design, structure, and scope. Human Mutation. 2000; 16(2): 123-31. http://dx.doi.org/10.1002/1098-1004(200008)16:2<123::AID-HUMU4>3.0.CO;2-Y

[24] Spurdle A, Healey S, Devereau A, *et al*. 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. 2012; 33: 2-7. PMid:21990146. http://dx.doi.org/10.1002/humu.21628

[25] Greenblatt MS, Brody LC, Foulkes WD, *et al*. Locus-Specific Databases (LSDBs) and Recommendations to Strengthen Their Contribution to the Classification of Variants in Cancer Susceptibility Genes. Human mutation. 2008; 29(11): 1273-81. PMid:18951438. http://dx.doi.org/10.1002/humu.20889

[26] Stenson PD, Mort M, Ball EV, *et al*. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Hum Genet. 2014; 133(1): 1-9. PMid:24077912. http://dx.doi.org/10.1007/s00439-013-1358-4

[27] Caputo S, Benboudjema L, Sinilnikova O, *et al*. Description and analysis of genetic variants in French hereditary breast and ovarian cancer families recorded in the UMD-*BRCA1/BRCA2* databases. Nucleic acids research. 2012; 40(Database issue): D992-1002. PMid:22144684. http://dx.doi.org/10.1093/nar/gkr1160

[28] Vallée MP, Francy TC, Judkins MK, *et al*. Classification of Missense Substitutions in the BRCA Genes: a Database Dedicated to Ex-UVs. Human mutation. 2012; 33(1): 22-8. PMid:21990165. http://dx.doi.org/10.1002/humu.21629

[29] De Leeneer K, De Schrijver J, Clement L, *et al*. Practical tools to implement massive parallel pyrosequencing of PCR products in next generation molecular diagnostics. PloS one. 2011; 6(9): e25531. PMid:21980484. http://dx.doi.org/10.1371/journal.pone.0025531

[30] Mardis ER. The impact of next-generation sequencing technology on genetics. Trends Genet. 2008; 24(3): 133-41. PMid:18262675. http://dx.doi.org/10.1016/j.tig.2007.12.007

[31] NHGRI: Breast Cancer Information Core 2015 (updated 13 March, 2015). Available from: http://research.nhgri.nih.gov/bic/

[32] Sherry ST, Ward MH, Kholodov M, *et al*. dbSNP: the NCBI database of genetic variation. Nucleic acids research. 2001; 29(1): 308-11. PMid:11125122. http://dx.doi.org/10.1093/nar/29.1.308

[33] dbSNP Home Page 2015 (cited 6 June, 2015). Available from: http://www.ncbi.nlm.nih.gov/SNP/

[34] ExAC Browser 2015 (cited 6 June, 2015). Available from: http://exac.broadinstitute.org/

[35] Exome Variant Server 2015 (6 June, 2015). Available from: http://evs.gs.washington.edu/EVS/

[36] Reese MG, Eeckman FH, Kulp D, *et al*. Improved splice site detection in Genie. J Comput Biol. 1997; 4(3): 311-23. PMid:9278062. http://dx.doi.org/10.1089/cmb.1997.4.311

[37] BDGP: Splice Site Prediction by Neural Network 2015 (cited 20 May, 2015). Available from: http://www.fruitfly.org/seq_tools/splice.html

[38] Wang M, Marin A. Characterization and prediction of alternative splice sites. Gene. 2006; 366(2): 219-27. PMid:16226402. http://dx.doi.org/10.1016/j.gene.2005.07.015

[39] Alternative Splice Site Predictor (ASSP) - Prediction 2015 (cited 22 May, 2015). Available from: http://wangcomputing.com/assp/

[40] Desmet FO, Hamroun D, Lalande M, *et al*. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic acids research. 2009; 37(9): e67. PMid:19339519. http://dx.doi.org/10.1093/nar/gkp215

[41] Human Splicing Finder - Version 3.0 2015(cited 7 June, 2015). Available from: http://www.umd.be/HSF3/HSF.html

[42] Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol. 2004; 11(2-3): 377-94. PMid:15285897. http://dx.doi.org/10.1089/1066527041410418

[43] Adzhubei IA, Schmidt S, Peshkin L, *et al*. A method and server for predicting damaging missense mutations. Nat Methods. 2010; 7(4): 248-9. PMid:20354512. http://dx.doi.org/10.1038/nmeth0410-248

[44] Mathe E, Olivier M, Kato S, *et al*. Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods. Nucleic acids research. 2006; 34(5): 1317-25. PMid:16522644. http://dx.doi.org/10.1093/nar/gkj518

[45] Tavtigian SV, Deffenbaugh AM, Yin L, *et al*. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. Journal of Medical Genetics. 2006; 43(4): 295-305. PMid:16014699. http://dx.doi.org/10.1136/jmg.2005.033878

[46] Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic acids research. 2011. PMid:21727090. http://dx.doi.org/10.1093/nar/gkr407

[47] Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic acids research. 2005; 33(suppl 2): W306-10. PMid:15980478. http://dx.doi.org/10.1093/nar/gki375

[48] Capriotti E, Fariselli P, Casadio R. A neural-network-based method for predicting protein stability changes upon single point mutations. Bioinformatics. 2004; 20 Suppl 1: 63-8. PMid:15262782. http://dx.doi.org/10.1093/bioinformatics/bth928

[49] Li B, Krishnan VG, Mort ME, *et al*. Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinformatics. 2009; 25(21): 2744-50. PMid:19734154. http://dx.doi.org/10.1093/bioinformatics/btp528

[50] Calabrese R, Capriotti E, Fariselli P, *et al*. Functional annotations improve the predictive score of human disease-related mutations in proteins. Hum Mutat. 2009; 30(8): 1237-44. PMid:19514061. http://dx.doi.org/10.1002/humu.21047

[51] Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic acids research. 2013; 41(Database issue): 377-86. PMid:23193289. http://dx.doi.org/10.1093/nar/gks1118

[52] Thomas PD, Campbell MJ, Kejariwal A, *et al*. PANTHER: a library of protein families and subfamilies indexed by function. Genome research. 2003; 13(9): 2129-41. PMid:12952881. http://dx.doi.org/10.1101/gr.772403

[53] Bromberg Y, Rost B. SNAP: predict effect of non-synonymous polymorphisms on function. Nucleic acids research. 2007; 35(11): 3823-35. PMid:17526529. http://dx.doi.org/10.1093/nar/gkm238

[54] PhD-SNP: Predictor of human Deleterious Single Nucleotide Polymorphisms 2015 (cited 2015 5 June 2015). Available from: http://snps.biofold.org/phd-snp/phd-snp.html

[55] Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics. 2015. http://dx.doi.org/10.1093/bioinformatics/btv195

[56] Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009; 4(7): 1073-81. PMid:19561590. http://dx.doi.org/10.1038/nprot.2009.86

[57] Schwarz JM, Cooper DN, Schuelke M, *et al*. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Meth. 2014; 11(4): 361-2. PMid:24681721. http://dx.doi.org/10.1038/nmeth.2890

[58] BIOBASE. Getting Started Guide: HGMD® Professional - Frequently Asked Questions 2014. 6 June, 2015. Available from: http://www.biobase-international.com/wp-content/uploads/2013/02/HGMD_FAQ_2013.4.pdf

[59] Jhuraney A, Velkova A, Johnson RC, *et al*. BRCA1 Circos: a visualisation resource for functional analysis of missense variants. J Med Genet. 2015; 52(4): 224-30. Epub 2015/02/04. PMid:25643705. http://dx.doi.org/10.1136/jmedgenet-2014-102766

[60] Richards S, Aziz N, Bale S, *et al*. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015; 17(5): 405-23. PMid:25741868. http://dx.doi.org/10.1038/gim.2015.30