Description and analysis of genetic variants in French hereditary breast and ovarian cancer families recorded in the UMD-BRCA1/BRCAD databases

Sandrine Caputo¹,*, Louisa Benboudjema¹, Olga Sinilnikova²,³, Etienne Rouleau¹, Christophe Béroud⁴,⁵,⁶, Rosette Lidereau¹ and the French BRCA GGC Consortium†

¹Institut Curie—Hôpital René Huguenin, Service d’Oncogénétique, U735 INSERM—Saint-Cloud, ²Unité Mixte de Génétique Constitutionnelle des Cancers Fréquents, Hospices Civils de Lyon—Centre Léon Bérard, Lyon, ³INSERM U1052, CNRS UMR5286, Université Lyon 1, Centre Recherche en Cancérologie de Lyon, Lyon, ⁴INSERM U827—Montpellier, ⁵CHU Montpellier, Hôpital Arnaud de Villeneuve, Laboratoire de Génétique Moléculaire—Montpellier and ⁶Université Montpellier1, UFR Médecine—Montpellier, F-34000, France

ABSTRACT

BRCA1 and BRCA2 are the two main genes responsible for predisposition to breast and ovarian cancers, as a result of protein-inactivating monoallelic mutations. It remains to be established whether many of the variants identified in these two genes, so-called unclassified/unknown variants (UVs), contribute to the disease phenotype or are simply neutral variants (or polymorphisms). Given the clinical importance of establishing their status, a nationwide effort to annotate these UVs was launched by laboratories belonging to the French GGC consortium (Groupe Génétique et Cancer), leading to the creation of the UMD-BRCA1/BRCAD databases (http://www.umd.be/BRCA1/ and http://www.umd.be/BRCA2/). These databases have been endorsed by the French National Cancer Institute (INCa) and are designed to collect all variants detected in France, whether causal, neutral or UV. They differ from other BRCA databases in that they contain co-occurrence data for all variants. Using these data, the GGC French consortium has been able to classify certain UVs also contained in other databases. In this article, we report some novel UVs not contained in the BIC database and explore their impact in cancer predisposition based on a structural approach.

INTRODUCTION

BRCA1 and BRCA2 are the two main genes predisposing individuals to breast and/or ovarian cancer. Most monoallelic germline mutations in these genes are point variants or deletions/insertions of a few nucleotides. Large rearrangements (exon loss or duplication, 5' alterations, etc...) have also been described (1,2). These mutations are situated throughout the entire sequence of the two genes. In France, causal mutations in BRCA1 and BRCA2 genes account for about 15% of hereditary breast and ovarian cancers (HBOC) (3).

The pathogenic nature of variants resulting in a premature termination codon (PTC) has been clearly established for those variants leading to nonsense-mediated decay (NMD) (4–7), i.e. BRCA1 proteins with a PTC located upstream from the last 10 carboxy-terminal codons, and BRCA2 proteins with a PTC located upstream from the last 110 carboxy-terminal codons. Some specific missense variants can generate a PTC via abnormal splicing, while others clearly alter the function of BRCA1, including all those occurring in the RING finger and some located in BRC domains (8). The pathogenicity of these variants has been inferred from linkage analysis of high-risk families, functional assays, biochemical studies and/or demonstration of abnormal mRNA processing (8–10).

*To whom correspondence should be addressed. Email: sandrine.caputo@curie.net
†The members of the BRCA GGC Consortium are listed in the Appendix

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© The Author(s) 2011. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
The main difficulty encountered in molecular diagnosis of breast and ovarian cancer predisposition relates to the large number of variants with unclear clinical significance. Variants that do not create a PTC represent a significant proportion of DNA variants and are usually called unclassified variants (UVs). UVs mostly consist of missense and intronic variants and in-frame deletions or insertions. Exhaustive collections of UVs could allow cosegregation studies within pedigrees, and these complementary data could help to define a reliable classification. Whether a given UV contributes to the disease phenotype or is merely a neutral variant (or a neutral polymorphism) has direct clinical implications.

Several approaches can be used to determine pathogenicity, including multidisciplinary analysis of cosegregation within pedigrees and degree of family history of the disease, or by indirect measures, including amino acid conservation, severity of amino acid change and evidence from functional assays, outputs from available bioinformatics tools, tumor pathology (11,12), population genetics and comparisons of allele frequencies (13), lack of co-occurrence with causal mutations (14), RNA and protein function studies (15), evolutionary studies of amino acid conservation (16) and a recently proposed combinatorial approach (17). However, such approaches cannot definitively show whether or not a specific variant is causal.

Mutation databases can help to classify variants by providing a unique source of information. The Breast Cancer Information Core Database (http://research.nhgri.nih.gov/projects/bic/index.shtml: BIC) is a voluntary international collection of variants identified worldwide (18). Other online databases include LOVD (Leiden Open Variation Database)-literature unclassified variant (http://chromium.liacs.nl/LOVD2/cancer/home.php) (19), LOVD-IARC (http://brca.iarc.fr/LOVD/home.php), kConFab (Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer) (20), The Singapore Human Mutation and Polymorphism Database (21), and HGMD (The Human Gene Mutation Database at the Institute of Medical Genetics in Cardiff) (22), NGRL/DmuDb (https://ngrl.manchester.ac.uk/Browser/genes.htm).

In this report, we present exhaustive databases of French BRCA1/2 variants (UMD-BRCA1/BRCA2: http://www.umd.be/BRCA1/ and http://www.umd.be/BRCA2/) developed with UMD software (23). These databases have been endorsed by the French Cancer National Institute (INCa) and are available online. A curator collects and compiles information from 16 licensed laboratories belonging to the French GGC consortium (Groupe Génétique et Cancer) and located all over France. UMD-BRCA1/BRCA2 databases centralize all variants (causal, neutral and UVs), allowing identification of co-occurrence data. Data on polymorphisms were not collected. In October 2011, the UMD-BRCA1/BRCA2 databases contained data from a set of 3743 families for BRCA1 and 3,650 families for BRCA2. In this report, we first analyze BRCA1/BRCA2 variants in the French population, and then use our co-occurrence data to classify some UVs as either causal or neutral.

Finally, we describe BRCA1 and BRCA2 UVs and evaluate their pathogenicity, notably from a structural standpoint (Supplementary Data). Therefore, this database should be a tool to classify variant and promote specific studies.

MATERIALS AND METHODS

Software and search engine

The UMD-BRCA1/BRCA2 databases were constructed with an updated version of the Universal Mutation Database generic software (http://www.umd.be) (23,24), which includes an optimized structure to assist and secure data entry and to allow the input of various clinical or biological data. An offline copy of UMD-BRCA1/BRCA2 is continuously edited and updated by the curators; edited copies are regularly deposited on the server. These databases are freely available online. Their integrity is ensured by the original (edited) template offline. Notification of errors in the current version would be gratefully received by the curator author. Database design follows general recommendations for LSDB characteristics and their curation as proposed by the Human Genome Variation Society (HGVS) (25).

Website

The Website is divided into six sections, four of which are dedicated to information on the BRCA1 or BRCA2 genes and numbering of exons and cDNA sequences, information on the protein and the references, while two sections allow the user to search into the database. Hyperlinks are provided to access the reference sequences and data search capabilities. The web pages also contain links to external information including central biological databases, such the Human Genome Database (DB), Pubmed, GeneCard, UniGene, Uniprot, OMIM, HGTV, etc.

Probands

About 24 000 probands were managed at the medical genetics departments of French Comprehensive Cancer Care centers and hospitals between January 1995 and October 2011. Eligibility criteria for familial genetic testing are described in the French consensus statement (26). All probands enrolled in this study gave their written informed consent for specific genetic testing. The probands were the index cases of individual families and were screened for BRCA1 and BRCA2 variants. Records in the French UMD-BRCA1/BRCA2 databases are restricted to index cases of families with clearly causal mutations, neutral variants or UVs. A system of five classes has been used for classification (27). An anonymized and unique sample identifier (sample_ID) is generated for each family.

Variant screening

The full coding sequences and exon–intron junctions of the BRCA1 and BRCA2 genes were screened for variants, based on prescreening (DGGE, SSCP, PTA,
dHPLC, HRM or EMMA) and sequencing. Several large rearrangements were identified by large cDNA sequencing, MLPA (28), QMPSF (29), qPCR (30), qPCR HRM (31), EMMA (32), bar code screening (33) or dedicated array CGH (34).

**UMD-BRCA1/BRCA2 databases**

The UMD software includes an automatic procedure to check for the correct description of the sequence variation at the nucleotide level and to generate the variant name at the protein level. All variant types are documented. All variants were subjected to a curation process and named according to the HGVS nomenclature before integration in the databases. Variants previously described either in publications or in other BRCA1/2 databases were reviewed. Data on 7396 different BRCA1/2 families were collected between 1995 and October 2011 (Table 2). Co-occurrence of each UV with causal mutations was recorded in the databases. Data on polymorphisms were not collected. Families were recorded only once, independently of the number of members tested. No clinical data were collected.

The present version of BRCA1 and BRCA2 is based on GenBank reference sequences NM_007294.2 (mRNA: U14680) and NM_000059.3 (mRNA: U43746), respectively. For BRCA1, this entry includes 119 bases of 5'-untranslated sequence with protein translation starting at nucleotide 120. The reference BRCA2 sequence also includes the 228-base 5'-untranslated region.

**Searching tools**

The UMD knowledgebase system includes many analytical tools (23,24). In addition to previously developed routines, new tools have been specifically developed for the UMD-BRCA1/BRCA2 for in-depth analysis. The user is able to optimize multicriteria search tools to select records from any field.

Prequeried data can be displayed by several Searching Tools in ‘Mutations’ such as ‘Type and number of mutations’ or ‘Mutation by exon/intron’ (Supplementary Figure S1A). Alternatively, the option ‘I found a mutation’ provides access to a quick search (nucleotide or amino acid position), ‘Free search’ provides access to an advanced search and a customized search interface searching several items (sample ID, amino acid position, biological significance, etc). A simple click on hyperlinks provides access to a list of interesting variants (Supplementary Data).

**Variant analysis**

For co-occurrence analysis, a UV in trans with two different causal mutations, was considered to be unlikely to cause disease.

The impact of missense substitutions in the BRCA1/2 genes was predicted by using three online programs: Align-GVGD (35), SIFT (36) and PolyPhen (37), which predict the effect of missense substitutions on gene function, as well as the new UMD-Predictor tool included in the UMD package (17).

For structural analysis, we focused on some novel variants not contained in the BIC database and located in a conserved region or in a previously reported functional domain. Regions and domains of BRCA1 or BRCA2 are described in Supplementary Data (Supplementary Figure S2).

**RESULTS**

**UMD-BRCA1/BRCA2 databases**

In UMD-BRCA1, a total of 1391 distinct BRCA1 variants comprising 586 causal mutations (42%), 5 likely causal mutations (0.4%), 22 likely neutral variants (1.6%), 74 neutral variants (5%) and 704 UVs (51%) have been identified in 3743 French families (1 index case harbors two different causal BRCA1 mutations and 81 cases harbor one causal mutation associated with at least one UV) (Tables 1 and 2). In UMD-BRCA2, a total of 1704 distinct BRCA2 variants were reported in 3650 families, with 567 different causal mutations (33.1%), 17 likely neutral variants (1%), 30 neutral variants (1.8%) and 1092 UVs (64.1%) (1 index case harbors two different causal BRCA2 mutations and 129 cases harbor one causal mutation and at least one UV) (Tables 1 and 2). Three index cases were double heterozygotes for causal BRCA1 and BRCA2 mutations. Most variants were found in a unique family (517 of 704 BRCA1 UVs and 738 of 1092 BRCA2 UVs) (Table 1). A total of 3743 BRCA1 families (2636 causal mutations and 948 UVs) and 3650 BRCA2 families (1576 causal mutations and 1775 UVs) were found to have at least one mutation (Table 2).

The exon distribution of the UVs, neutral variants and causal mutations in the UMD-BRCA1/BRCA2 databases was also assessed and no statistically significant differences were found between the UMD-BRCA1/BRCA2 and BIC databases (see Supplementary Data). Based on the ratio between UMD-BRCA1/BRCA2 variants and BIC databases (see Supplementary Data). Based on the ratio between UMD-BRCA1/BRCA2 variants and BIC databases (see Supplementary Data). Based on the ratio between UMD-BRCA1/BRCA2 variants and BIC databases (see Supplementary Data). Based on the ratio between UMD-BRCA1/BRCA2 variants and BIC databases (see Supplementary Data). Based on the ratio between UMD-BRCA1/BRCA2 variants and BIC databases (see Supplementary Data). Based on the ratio between UMD-BRCA1/BRCA2 variants and BIC databases (see Supplementary Data). Based on the ratio between UMD-BRCA1/BRCA2 variants and BIC databases (see Supplementary Data). Based on the ratio between UMD-BRCA1/BRCA2 variants and BIC databases (see Supplementary Data). Based on the ratio between UMD-BRCA1/BRCA2 variants and BIC databases (see Supplementary Data). Based on the ratio between UMD-BRCA1/BRCA2 variants and BIC databases (see Supplementary Data). Based on the ratio between UMD-BRCA1/BRCA2 variants and BIC databases (see Supplementary Data).

**Characterization of French causal mutations in the UMD-BRCA1/BRCA2 databases**

The spectrum of causal mutations in the BRCA1 and BRCA2 genes involves the entire sequence (Figure 1). The BRCA1 exons most frequently mutated are exons 2, 5 and 20 and the BRCA2 exons most frequently mutated are exons 8, 13 and 23 (Figure 1).
The most prevalent causal BRCA1 mutations identified in the French population were: the French founder frameshift c.3481_3491del (3600del11, exon 11) from North-Eastern France (137 records) (38,39) and c.5266dup (5382insC, exon 20), a founder Ashkenazi Jewish mutation (138 records). This causal mutation was found 48 times in Ile-de-France and 27 times in North-Eastern France. Six other frequent recurrent causal BRCA1 mutations were observed more than 8 times (Table 3): all had been recorded early in the BIC database (October 2011), except for c.5128G>T (9 records), which was found 8 times in North-Eastern France (39).

We identified eight frequent causal BRCA2 mutations (Table 3), all of which were contained in the BIC database (August 2011). The most prevalent causal BRCA2 mutation in the French population is c.2808_2811delACAA, a Western European mutation (41 records) (40). All of these eight recurrent causal BRCA2 mutations were frequently detected in the Paris region.

In comparison with the BIC database, the UMD database also identified causal mutations from the Ashkenazi Jewish population (26.88% for BIC-BRCA1 versus 6.14% UMD-BRCA1, 9.49% for BIC-BRCA2 versus 0.6% for UMD-BRCA2), but also several frequent causal mutations from other European ethnic populations.

**Characterization of French UVs in the UMD-BRCA1/BRCA2 databases**

UVs were also detected throughout the two genes. The most frequently mutated exons are 17, 18 and 19 for BRCA1 and 15, 20 and 26 for BRCA2 (Figure 1). As expected, numerous variants (at least 51% of distinct variants in BRCA1 and 64.1% of distinct variants in BRCA2) could not be classified as causal or neutral. These UVs were found in BRCA1 in 1018 families and in BRCA2 in 1944 families. In BRCA1, 704 distinct UVs were identified in one or several families: 948 families with one BRCA1 UV only, 81 with a UV associated with a causal mutation and 70 with UVs only (Tables 1 and 2).

Most UVs were observed only once (517/704 in BRCA1 and 738/1092 in BRCA2) (Table 1), while some were recurrent, including BRCA1 c.5075-53C>T (IVS17-53C>T, SNP rs8176258) with 66 entries, c.5277+48_59dup (IVS20+48_59dup, 47 entries) and BRCA2 c.7008-62A>G (IVS13-62A>G, 69 entries), c.9257-16T>C (IVS24-16T>C, SNP rs11571818, 50 entries) and c.6100C>T (p.Arg2034Cys, exon 11, SNP rs1799954, 85 entries).
Most UVs identified in this study were either missense changes (45% in BRCA1 and 53.58% in BRCA2) or silent variants (11.5% in BRCA1 and 15.38% in BRCA2) (Figure 2). Small deletions or insertions of various lengths were also observed (2.6% in BRCA1 and 2.19% in BRCA2) (Figures 2), as well as intronic variants (40.9% in BRCA1 and 15.38% in BRCA2) (Figure 2). Intronic variants lie outside intron–exon boundaries and their potential impact on the splicing process is difficult to predict and need specific RNA analysis (Figure 2).

GGC classification of some UVs already reported in other databases

Co-occurrence data on a UV in trans of a known causal mutation (the UV and the causal mutation are on opposite chromosomes) can provide a powerful means of ruling out pathogenicity (14). Identification of co-occurrence of a UV and a causal mutation requires further investigations such as cloning to exclude the in cis position of these two variants (41). Without these additional analyses, there is still a very weak probability that the two causal mutations are situated in cis, which is why the following hypothesis can be used: if a UV coexists with at least two different causal mutations in two different families, this UV can be considered to be always in trans with respect to the causal mutation. To support the co-occurrence results, we also used an integrated approach combining the Align-GVGD, SIFT and PolyPhen programs, published data, and the UMD-Predictor tool, allowing 3 BRCA2 UVs to be classified as neutral and polymorphism (Table 4). These BRCA2 UVs, described in the BIC database, were reported in our databases as co-occurring with causal mutations: c.6100C>T [p.Arg2034Cys, exon 11, 137 records, 16 in trans (9 distinct causal mutations)], c.7008-62A>G (IVS13-62A>G, 106 records, 6 distinct causal mutations in trans), and c.9257-16T>C [IVS24-16T>C, 55 records, 8 in trans (4 distinct causal mutations)]. This suggests that these variants do not sufficiently alter BRCA2 function to predispose to cancer. The status of the BRCA2 c.6100C>T variant differed between the algorithms: it was predicted to be pathogenic by the UMD-Predictor tool and neutral by SIFT, Align-GVGD, PolyPhen and co-occurrence analysis. Moreover, c.6100C>T has been described as a polymorphism in a Spanish population of cancer families and patients (42). We classified this variant as a neutral
Table 3. Geographical distribution of founder causal mutations in BRCA1 and BRCA2 genes

**BRCA1**

<table>
<thead>
<tr>
<th>Mutation</th>
<th>IDF</th>
<th>NW</th>
<th>North</th>
<th>NE</th>
<th>SE</th>
<th>SW</th>
<th>Center</th>
<th>UMD database</th>
<th>BIC database</th>
</tr>
</thead>
<tbody>
<tr>
<td>c.68_69delAG (185delAG)</td>
<td>28</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>14</td>
<td>1</td>
<td>5</td>
<td>56 (1.77%)</td>
<td>1980 (17.5%)</td>
</tr>
<tr>
<td>c.3839_3843delCTCAG (3958del5ins4)</td>
<td>12</td>
<td>4</td>
<td>2</td>
<td>94</td>
<td>–</td>
<td>3</td>
<td>3</td>
<td>41 (1.29%)</td>
<td>3 (0.03%)</td>
</tr>
<tr>
<td>c.3481_3491del11 (3600del11)</td>
<td>18</td>
<td>4</td>
<td>13</td>
<td>94</td>
<td>12</td>
<td>4</td>
<td>2</td>
<td>137 (4.34%)</td>
<td>137 (4.34%)</td>
</tr>
<tr>
<td>c.3841C&gt;T (3960C&gt;T)</td>
<td>6</td>
<td>2</td>
<td>–</td>
<td>4</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>31 (0.98%)</td>
<td>31 (0.98%)</td>
</tr>
<tr>
<td>c.4327C&gt;T (4446C&gt;T)</td>
<td>17</td>
<td>25</td>
<td>–</td>
<td>1</td>
<td>–</td>
<td>2</td>
<td>2</td>
<td>8 (0.29%)</td>
<td>48 (1.52%)</td>
</tr>
<tr>
<td>c.5030_5033delCTAA (5149delCTAA)</td>
<td>13</td>
<td>3</td>
<td>–</td>
<td>1</td>
<td>–</td>
<td>–</td>
<td>2</td>
<td>32 (1.01%)</td>
<td>32 (1.01%)</td>
</tr>
<tr>
<td>c.5128G&gt;T (5247G&gt;T)</td>
<td>3</td>
<td>7</td>
<td>–</td>
<td>5</td>
<td>1</td>
<td>–</td>
<td>7</td>
<td>9 (0.29%)</td>
<td>9 (0.29%)</td>
</tr>
<tr>
<td>c.5266dup (5382insC)</td>
<td>13</td>
<td>–</td>
<td>–</td>
<td>1</td>
<td>8</td>
<td>27</td>
<td>14</td>
<td>138 (4.37%)</td>
<td>138 (4.37%)</td>
</tr>
</tbody>
</table>

**BRCA2**

<table>
<thead>
<tr>
<th>Mutation</th>
<th>IDF</th>
<th>NW</th>
<th>North</th>
<th>NE</th>
<th>SE</th>
<th>SW</th>
<th>Center</th>
<th>UMD database</th>
<th>BIC database</th>
</tr>
</thead>
<tbody>
<tr>
<td>c.771_775delTCAA (999del5)</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>–</td>
<td>10 (0.32%)</td>
<td>20 (0.64%)</td>
</tr>
<tr>
<td>c.1310_1313delAAGA (1538del4)</td>
<td>8</td>
<td>5</td>
<td>–</td>
<td>3</td>
<td>6</td>
<td>2</td>
<td>–</td>
<td>20 (0.64%)</td>
<td>41 (1.3%)</td>
</tr>
<tr>
<td>c.2808_2811delACAA (3036del4)</td>
<td>21</td>
<td>–</td>
<td>–</td>
<td>7</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>41 (1.3%)</td>
<td>17 (0.54%)</td>
</tr>
<tr>
<td>c.5645C&gt;A (5873C&gt;A)</td>
<td>8</td>
<td>1</td>
<td>–</td>
<td>2</td>
<td>2</td>
<td>–</td>
<td>2</td>
<td>17 (0.54%)</td>
<td>19 (0.6%)</td>
</tr>
<tr>
<td>c.5948delTT (6174delT)</td>
<td>13</td>
<td>2</td>
<td>–</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>19 (0.6%)</td>
<td>15 (0.48%)</td>
</tr>
<tr>
<td>c.6275_6276delTT (6503delTT)</td>
<td>8</td>
<td>3</td>
<td>–</td>
<td>3</td>
<td>3</td>
<td>–</td>
<td>1</td>
<td>15 (0.48%)</td>
<td>18 (0.57%)</td>
</tr>
<tr>
<td>c.6644_6647delACTC (6872del4)</td>
<td>4</td>
<td>2</td>
<td>–</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>–</td>
<td>25 (0.80%)</td>
<td>20 (0.18%)</td>
</tr>
<tr>
<td>c.9026_9030delATCAT (9254del5)</td>
<td>14</td>
<td>–</td>
<td>–</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>25 (0.80%)</td>
<td>10 (0.09%)</td>
</tr>
</tbody>
</table>

Numbers of families with founder mutations throughout the country (Polymorphisms are excluded); IDF: Paris; NW: Caen, Nantes; North: Lille; NE: Nancy, Reims, Strasbourg; SE: Montpellier, Marseille, Nice; SW: Bordeaux, Toulouse; Center: Lyon, Clermont-Ferrand. *χ² test.
variant. Our classification is reported in the UMD-BRCA1/BRCA2 databases. When it possible, a structural approach is used to help classify the UVs (see example in Supplementary Data: A structural approach to novel UVs).

**DISCUSSION**

The UMD-BRCA1/2 databases provide an overview of the BRCA1/2 variants in the French population

We examined BRCA1/2 UVs collected by 16 licensed laboratories located throughout France and belonging to the French GGC consortium ‘Groupe Génétique et Cancer’, which is designed to validate diagnostic procedures and strategies. The 16 licensed laboratories automatically send all results of families tested in France with a BRCA1/2 variant. Every 3 months, an update is performed by all laboratories. The UMD-BRCA1/BRCA2 databases have been approved by the French National Cancer Institute (INCa).

Among 24,000 families tested, 2,733 BRCA1 and 1,707 BRCA2 families with one causal mutation were identified, as well as 1018 BRCA1 and 1,944 BRCA2 families with at least one UV (Table 2). About 95% of BRCA1/2 variants (causal mutations and UVs) were observed only once.

Certain exons showed an excess of variants. Interestingly, in BRCA1, these exons correspond to the RING-finger domain and the first BRCT domain, while in BRCA2, the most frequently mutated exons are not associated with resolved structural domains (Supplementary Tables S1A and S1B).
As in other populations, some variants are recurrent in the French population (40,43–45). Sixteen recurrent causal mutations in \( \text{BRCA1} \) and \( \text{BRCA2} \) were reported more than eight times each. Some UVs were also recurrent. Three were most frequent in \( \text{BRCA2} \), namely c.7008-62A>G (106 families), c.9257-16T>C (55 families) and c.6100C>T (137 families). These three variants could be geographic polymorphisms, or specific to the French population. Detection of c.7008-62A>G and c.9257-16T>C could also be variable according to the positions of the primers used to amplify the region and the frequency differs from one database to another due to the selection of PCR primers.

The UMD-BRCA1/BRCA2 databases and pooled data from the French GGC are sufficiently powerful for co-occurrence studies. If homozygous and compound heterozygous variants in \( \text{BRCA1} \) are assumed to be embryonically lethal (46–48), while those in \( \text{BRCA2} \) are responsible for Fanconi anemia (49–51), several variants can be classified as neutral based on a single observation. One patient reported in the UMD-BRCA1/BRCA2 databases harbors two different causal \( \text{BRCA1} \) mutations in \( \text{cis} \) (Bezieau, S. and Delnatte, C. unpublished data). Another patient harbors two different causal \( \text{BRCA2} \) mutations. These two causal mutations result in a breast/ovarian cancer syndrome but also in glioblastoma (Sevenet, N., Longy, M. et al. unpublished data).

A double heterozygote with a causal \( \text{BRCA1} \) mutation and a causal \( \text{BRCA2} \) mutation did not appear to present a more severe phenotype (52). We also identified three families that included double heterozygotes with a causal mutation in \( \text{BRCA1} \) and a causal mutation in \( \text{BRCA2} \). Coexistence of causal mutations in \( \text{BRCA1} \) and \( \text{BRCA2} \) is unusual in non-Ashkenazi populations. Two of these three families did not harbor the Ashkenazi founder mutations (unpublished data) (52), but these three families presented a breast cancer and other cancers syndrome.

UVs classified by GGC

Most information used to determine the pathogenicity of an UV is indirect, such as segregation data, loss of heterozygosity, absence of the UV in an unaffected subject or in \textit{in silico} prediction based on the conservation, position and nature of amino acid change or the impact on the mRNA.

We used co-occurrence data and predictive algorithms to classify three \( \text{BRCA2} \) variants (Table 4). The French consortium uses these data to classify UVs in the UMD-BRCA1/BRCA2 databases. For example, like Goldgar et al., we classified the \( \text{BRCA2} \) c.6100C>T (p.Arg2034Cys) variant as a neutral variant (53). In our database, this c.6100C>T variant was associated with eight distinct causal mutations strongly suggesting that it is not causal. However, the status of the \( \text{BRCA2} \) c.6100C>T variant differed between the algorithms: SIFT score was 0.07, corresponding to borderline pathogenicity (0.05). The UMD Predictor\textsuperscript{*} tool contained the score SIFT but also other scores. The final conclusion of neutrality was based on a combination of prediction programs with co-occurrence data. We discuss the contribution of the 3D structure in Supplementary Data.

CONCLUSION

\( \text{BRCA1} \) and \( \text{BRCA2} \) are the two genes principally involved in predisposition to breast and ovarian cancer. Classification of UVs of \( \text{BRCA1} \) and \( \text{BRCA2} \) has major implications for hereditary forms of these malignancies. Most variants identified to date cannot yet be definitively classified as causal or neutral. The French UMD-BRCA1/BRCA2 databases, which collect all variants (neutral, causal and UVs), differ from other BRCA databases by the fact that they contain co-occurrence data for all variants, allowing the GGC consortium to classify a number of UVs. In the future, co-occurrence data and structural approaches should help to classify further unclassified \( \text{BRCA1}/\text{BRCA2} \) variants.

Table 4. Variants classified by GGC

<table>
<thead>
<tr>
<th>Variant (cDNA level)</th>
<th>p.Arg2034Cys</th>
<th>c.7008-62A&gt;G</th>
<th>c.9257-16T&gt;C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reported in UMD-BRCA1/BRCA2</td>
<td>137</td>
<td>106</td>
<td>55</td>
</tr>
<tr>
<td>Co-occurrence with a causal mutation</td>
<td>16 (9)</td>
<td>6</td>
<td>8 (4)</td>
</tr>
<tr>
<td>dbSNP</td>
<td>Rs1 799 954</td>
<td>-</td>
<td>Rs1 571 818</td>
</tr>
<tr>
<td>Align-GVGD</td>
<td>GV = 257.44, GD = 11.33, C0 (less likely)</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>PolyPhen</td>
<td>0.423 (benign)</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>SIFT</td>
<td>0.07 (no deleterious)</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>UMD-Predictor\textsuperscript{*}</td>
<td>82 (pathogenic)</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Literature data</td>
<td>Polymorphism</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Classification UMD-BRCA1/BRCA2 database</td>
<td>UV (97)</td>
<td>UV (2)</td>
<td>UV (6)</td>
</tr>
<tr>
<td>Classification BIC database</td>
<td>Polymorphism</td>
<td>Neutral</td>
<td>Neutral/Polymorphism</td>
</tr>
</tbody>
</table>

The enriched deleterious group consisted of those substitutions with Align-GVGD scores of GV and GD < 62 and SIFT scores < 0.05. The enriched neutral group consisted of those substitutions with Align-GVGD scores of GV > 62 and GD > 0 and SIFT scores > 0.05. PolyPhen results for each variant were classified as benign (score ≤ 0.5), possibly damaging (0.5 ≤ score < 2), probably damaging (score > 2), and unknown. UMD predictor\textsuperscript{*} results for each missense were classified as polymorphism (score < 0.5), probable polymorphism (0.5 ≤ score ≤ 64), probably pathogenic (65 ≤ score ≤ 74) and pathogenic (score > 74). ND: not determined because algorithms only predict the impact of missense mutations.

\textsuperscript{*}For co-occurrence studies. If homozygous and compound heterozygous variants in \( \text{BRCA1} \) are assumed to be embryonically lethal (46–48), while those in \( \text{BRCA2} \) are responsible for Fanconi anemia (49–51), several variants can be classified as neutral based on a single observation. One patient reported in the UMD-BRCA1/BRCA2 databases harbors two different causal \( \text{BRCA1} \) mutations in \( \text{cis} \) (Bezieau, S. and Delnatte, C. unpublished data). Another patient harbors two different causal \( \text{BRCA2} \) mutations. These two causal mutations result in a breast/ovarian cancer syndrome but also in glioblastoma (Sevenet, N., Longy, M. et al. unpublished data). A double heterozygote with a causal \( \text{BRCA1} \) mutation and a causal \( \text{BRCA2} \) mutation did not appear to present a more severe phenotype (52). We also identified three families that included double heterozygotes with a causal mutation in \( \text{BRCA1} \) and a causal mutation in \( \text{BRCA2} \). Coexistence of causal mutations in \( \text{BRCA1} \) and \( \text{BRCA2} \) is unusual in non-Ashkenazi populations. Two of these three families did not harbor the Ashkenazi founder mutations (unpublished data) (52), but these three families presented a breast cancer and other cancers syndrome.

UVs classified by GGC

Most information used to determine the pathogenicity of an UV is indirect, such as segregation data, loss of heterozygosity, absence of the UV in an unaffected subject or in \textit{in silico} prediction based on the conservation, position and nature of amino acid change or the impact on the mRNA.

We used co-occurrence data and predictive algorithms to classify three \( \text{BRCA2} \) variants (Table 4). The French consortium uses these data to classify UVs in the UMD-BRCA1/BRCA2 databases. For example, like Goldgar et al., we classified the \( \text{BRCA2} \) c.6100C>T (p.Arg2034Cys) variant as a neutral variant (53). In our database, this c.6100C>T variant was associated with eight distinct causal mutations strongly suggesting that it is not causal. However, the status of the \( \text{BRCA2} \) c.6100C>T variant differed between the algorithms: SIFT score was 0.07, corresponding to borderline pathogenicity (0.05). The UMD Predictor\textsuperscript{*} tool contained the score SIFT but also other scores. The final conclusion of neutrality was based on a combination of prediction programs with co-occurrence data. We discuss the contribution of the 3D structure in Supplementary Data.
SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–2, Supplementary Figures 1–7, Supplementary References [8,17,23,27,54–61].

ACKNOWLEDGMENTS

The authors thank the French oncogeneticists, the GGC leads by Dr Catherine Noguès and probands for their cooperation.

FUNDING

French National Cancer Institute (INCa) (also to S.C. and L.B.), Development of the UMD tool used to construct the UMD-BRCA1/BRCA2 databases received funding from the European Community’s Seventh Framework Program (FP7/2007-2013) under grant agreement no. 200754—the GEN2PHEN project; Association d’Aide à la Recherche Cancérologique de Saint Cloud (ARCs). Funding for open access charge: Institut Curie—Hospital René Huguenin.

Conflict of interest statement. None declared.

REFERENCES


**APPENDIX**