BRCA Share: A Collection of Clinical BRCA Gene Variants

Christophe Béroud, Stanley I Letovsky, Corey D Braastad, Sandrine M Caputo, Olivia Beaudoux, Yves Jean Bignon, Brigitte Bressac-De Paillerets, Myriam Bronner, Crystal M Buell, Gwenaelle Collod-Beroud, et al.

To cite this version:

BRCA Share: A Collection of Clinical BRCA Gene Variants

Christophe Béroud,1,2,4 Stanley I. Letovsky,3,1 Corey D. Braastad,4,1 Sandrine M. Caputo,5,1 Olivia Beaudoux,6 Yves Jean Bignon,7 Brigitte Bressac-De Paillerets,8 Myriam Bronner,9 Crystal M. Buell,4 Gwenaelle Collod-Béroud,1 Florence Coulet,10 Nicolas Derive,3 Christina Divincenzo,6 Christopher D. Elzinga,6 Céline Garrec,11 Claude Houdayer,5,12 Izabela Karbassi,4 Sarah Lizard,13 Angela Love,4 Danièle Muller,14 Narasimhan Nagan,2 Camille R. Nery,15 Ghadi Rai,1 Françoise Revillon,16 David Salgado,1 Nicolas Sévenet,17 Olga Sininikova,18,1 Hagay Sobol,19 Dominique Stoppa-Lyonnet,5,12 Christine Toulas,20 Edwin Trautman,3 Dominique Vaur,21 Paul Vilquin,22 Katelyn S. Weymouth,3 Alecia Willis,23 Laboratory Corporation of America Variant Classification Group,2,3,22 Quest Diagnostics Variant Classification Group,3,15,1 UNICANCER Genetic Group BRCA Laboratory Network,6 Marcia Eisenberg,23 and Charles M Strom15,‡

1Aix Marseille Univ, INSERM, GMGF, Marseille, France; 2APHM, Hôpital TIMONE Enfants, Laboratoire de Génétique Moléculaire, Marseille, France; 3Laboratory Corporation of America, Westbrook, Massachusetts; 4Quest Diagnostics, Marlborough, Massachusetts; 5Service de Génétique, Département de Biologie des Tumeurs, Institut Curie, Paris, France; 6CHU et Institut Jean Godinot, Reims, France; 7Centre Jean Perrin, Clermont-Ferrand, France; 8Institut Gustave Roussy, Villejuif, France; 9CHU de Nancy-Brabois, Vandoeuvre-lès-Nancy, France; 10Groupe hospitalier Pitié-Salpêtrière, Assistance Publique-Hôpitaux de Paris, Laboratoire d’Oncogénétique et Angiogénétique moléculaire, Université Pierre et Marie Curie, Paris, France; 11CHU, Institut de Biologie, Hôtel Dieu, Nantes, France; 12Université Paris Descartes, Paris, France; 13CHU de Dijon, Hôpital d’Enfants, Service de Génétique Médicale, Dijon, France; 14Centre Paul Strauss, Strasbourg, France; 15Quest Diagnostics, San Juan Capistrano, California; 16Centre Oscar Lambret, Unité d’Oncologie Moléculaire Humaine, Lille, France; 17Institut Bergonié, Bordeaux, France; 18Hospices Civils de Lyon and Centre Léon Bérard, Lyon, France; 19Institut Paoli-Calmettes, Marseille, France; 20Institut Claudius Régaud, Toulouse, France; 21Laboratoire de biologie et de génétique du cancer, CLCC François Baclesse, INSERM 1079 Centre Normand de Génomique et de Médecine Personnalisée, Caen, France; 22Laboratoire de Biologie Cellulaire et Hormonale (CHU Arnaud de Villeneuve), Montpellier, France; 23Laboratory Corporation of America, Research Triangle Park, North Carolina

For the Next Generation Sequencing special issue

ABSTRACT: As next-generation sequencing increases access to human genetic variation, the challenge of determining clinical significance of variants becomes ever more acute. Germline variants in the BRCA1 and BRCA2 genes can confer substantial lifetime risk of breast and ovarian cancer. Assessment of variant pathogenicity is a vital part of clinical genetic testing for these genes. A database of clinical observations of BRCA variants is a critical resource in that process. This article describes BRCA ShareTM, a database created by a unique international alliance of academic centers and commercial testing laboratories. By integrating the content of the Universal Mutation Database generated by the French Unicancer Genetic Group with the testing results of two large commercial laboratories, Quest Diagnostics and Laboratory Corporation of America (LabCorp), BRCA ShareTM has assembled one of the largest publicly accessible collections of BRCA variants currently available. Although access is available to academic researchers without charge, commercial participants in the project are required to pay a support fee and contribute their data. The fees fund the ongoing curation effort, as well as planned experiments to functionally characterize variants of uncertain significance. BRCA ShareTM databases can therefore be considered as models of successful data sharing between private companies and the academic world.

KEYWORDS: NGS; BRCA1; BRCA2; breast cancer; ovarian cancer; genetic databases; variant classification

Introduction

Breast cancer is the most common malignancy in women in most countries with 464,000 new cases and 131,000 deaths each year in
Europe and 246,660 new cases and 40,450 deaths in the US [Ferlay et al., 2013]. Globally, it is estimated that more than 1 million women worldwide are diagnosed yearly with breast cancer and that more than 400,000 will die from the disease. Although most cases are sporadic, familial clustering has been reported and it is estimated that approximately 10% are likely to be hereditary. The known pathogenic mutations1 from predisposing genes, including BRCA1 (MIM# 113705) and BRCA2 (MIM# 600185), account for 20%–25% of those familial forms [Pharoah et al., 2002].

As breast cancers represent a significant public health problem, many national programs have been dedicated to cancer prevention and early diagnosis. They use various risk assessment (Gail [Chay et al., 2012], Claus [Fischer et al., 2013] and Tyer-Cozick [Boughery et al., 2010]) or probability models such as the BReast Cancer risk PRediction mOdel (BRCAPRO) [Mazzola et al., 2015]. The average cumulative risks in BRCA1 mutation carriers by the age of 70 are 65% (95% confidence interval 44%–78%) for breast cancer and 39% (18%–54%) for ovarian cancer. The corresponding estimates for BRCA2 mutation carriers are 45% (31%–56%) and 11% (2.4%–19%) [Antoniou et al., 2003]. Women with a high lifetime risk (≥20%) are eligible for genetic counseling and testing mainly for mutations of the BRCA1 [Hall et al., 1990] and BRCA2 [Wooster et al., 1994] cancer susceptibility genes.

Professional organizations in North America and Europe have published clinical practice guidelines with respect to BRCA counseling and testing [Balma˜na et al., 2011; Gadzicki et al., 2011; Graña et al., 2011; National Institute for Health Care Excellence 2013; Cancer Institute NSW 2015; Holter et al., 2015]. The United States National Comprehensive Cancer Network guidelines recommend genetic counseling and/or testing for women that fulfill one of the following criteria: breast cancer at the age of 50 years or younger; a bilateral breast cancer; triple-negative breast cancer (estrogen receptor negative, progesterone receptor negative, Her2Neu negative); breast cancer at any age with close relatives with breast/ovarian/pancreatic cancer; breast cancer with Ashkenazi Jewish ancestry; male breast cancer; known mutation for a breast cancer susceptibility gene within the family, or women with a history of ovarian cancer. These recommendations have resulted in a rapid uptake and utilization of BRCA testing in clinical practice and a consequent large increase in variant classification burdens among laboratories worldwide.

The process of assessing clinical significance of germline variants has been codified in guidelines [Babcock 2015] and formalized in probabilistic models for evidence integration [Goldgar et al., 2004; Lovelock et al., 2007]. Evidence may come from family studies, functional studies, algorithmic predictions, patterns of allelic co-occurrences within individuals, and other sources. Clinical significance is typically assessed per the ACMG/AMP guidelines [Richards et al., 2015], by assigning the variant to one of the following classifications: Pathogenic, Likely Pathogenic, VUS (Variant of Uncertain Significance), Likely Benign, and Benign. Integrating evidence to produce a variant classification currently relies on the judgment of geneticists with diverse areas of expertise who primarily assess the information available from the scientific literature and databases, as well as algorithmic predictions, to arrive at a variant classification. Several studies reporting inter reviewer agreements in classifications among experts with widely differing outcomes ranging from “moderate” (Gross k, 0.52) to “high” (k-alpha, 0.91) have been published. Specifically, although a higher concordance of potentially clinically actionable classifications (Pathogenic or Likely Pathogenic) derived using the ACMG-AMP schemes with those reported in locus-specific databases (LSDBs) and ClinVar have been reported, the concordance rates for classifications deemed as non-actionable (VUS, Likely Benign, and Benign) were lower [Maxwell et al., 2016]. Amendola et al. (2016) reported a 79% concordance when nine participating laboratories classified a set of variants using their internally developed methods and the ACMG-AMP criteria. However, only a 34% concordance for either classification system across laboratories was observed. After consensus discussions and detailed review of the ACMG-AMP criteria, concordance increased to 71%, showing that a common framework can help resolve differences in classifications.

It can require many hours to classify a variant when literature review is required. The results and conclusions often need to be fully vetted with a critical eye toward translation to the mechanism(s) and presentation(s) of the disease and phenotype under consideration. Understandably, this classification adds significant skilled labor cost to the testing process. Fully automated methods for variant classification are not ready for clinical practice. Median time required to curate available lines of evidence and classify variants have been reported to vary between 34 min (range 5–233 min) [Dewey et al., 2014], to 37 min per variant (range 1–175 min) [Amendola et al., 2015]. Furthermore, variant classifications are not static; classifications may change over time as the underlying scientific knowledge changes, requiring a periodic re-evaluation of the literature and other evidence prior to clinical reporting. CLIA laboratory regulations in the United States require each laboratory to be responsible for its own reports, and to update their classifications to reflect the most recent information, which limits the ability to use a paradigm for static classifications (and hence share the cost of classification) across laboratories.

Variant databases play a critical role in the classification process, by providing summary information about the state of scientific knowledge, with clinical assessments of the variant by other laboratories, with supporting evidence, and with data on other patients carrying the variant. LSDBs [Claustres et al., 2002] have been established for many disease genes, whereas other databases, including OMIM, ClinVar, LOVD, and ClinVite collect variant information across genes. Recently, network federation protocols such as Beacon [Krol 2015] and Matchmaker exchange [Philippakis et al., 2015] have been established to facilitate peer-to-peer exchange of variant data. Some commercial software products, including GeneInsight Network and Agilent’s Cartagenia, are using a similar approach to enable sharing of variant classifications within their user communities.

In the context of BRCA testing, databases available today include: the Breast Cancer Information Core Database (BIC) [Szabo et al., 2000], LOVD [Fokkema et al., 2011], ClinVar [Landrum et al., 2014], BRCA Exchange [Global Alliance for Genomics and Health 2016], the ARUP databases (www.arup.utah.edu/), and the Universal Mutation Database (UMD)-BRCA1/2 databases [Caputo et al., 2012]. Note that Myriad Genetics, a commercial laboratory, has developed its own database whose quality and content could not be evaluated as it is proprietary and not accessible to the community. These databases have been developed with different goals and have different curation processes, data quality, and content as underlined by Vail et al. [2015]. These authors compared BIC, ClinVar, HGMD (paid version), LOVD, and the UMD databases and concluded that these differences inhibit their wider use in clinical practice.

In order to provide high-quality sustainable BRCA1 and BRCA2 databases, the BRCA ShareTM public/private partnership was co-founded by Quest Diagnostics and the French National Institute

---

1We use the term variant to refer generically to differences from the reference sequence, and mutation to refer to pathogenic and likely pathogenic variants.
of Health and Medical Research (Inserm) in April 2015, with Laboratory Corporation of America Holdings as the first commercial participant. The goal of the initiative is to share clinical, genetic, epidemiological, and biological data on BRCA variants, particularly VUS, in order to improve the quality of laboratory diagnostics to better predict which individuals are at risk of developing hereditary breast and ovarian cancers, and to accelerate research on BRCA gene variants. BRCA ShareTM builds on an efficient data curation process [Caputo et al., 2012] that follows international recommendations [Eccles et al., 2015].

The BRCA ShareTM database now contains over 6,200 total BRCA variants, an increase of nearly 30% compared with the previous UMD-BRCA1/2 databases. Of these variants, 334 are newly identified pathogenic or likely pathogenic, increasing by about 20% the total number of pathogenic or likely pathogenic variants to 1,826.

To our knowledge, BRCA ShareTM is the first example of a successful public/private partnership for LSDBs that ensures free access to the database for research purposes while creating a user group of commercial partners that will collectively endorse running and development costs while also supporting functional tests to rapidly and efficiently classify VUS.

With increased adoption of whole exome and whole genome sequencing in clinical practice, it is believed that many patients might benefit from the indirect discovery of variants in clinically actionable genes. In a study of 1,000 individuals (500 European- and 500 African-descent participants randomly selected from the National Heart, Lung, and Blood Institute Exome Sequencing Project), Dorschner et al. [2013] report that ~3.4% of European-descent adults and ~1.2% of African-descent adults can be expected to have actionable highly penetrant pathogenic or likely pathogenic mutations identified by exome sequencing. Among these, three of the 1,000 participants had a BRCA1/2 pathogenic mutation [Dorschner et al., 2013]. This ratio was confirmed by a subsequent study by Amendola et al. [2015].

With the involvement of the major diagnostic companies in US and reference centers in France, for the first time, all variants will be shared and further annotated by a group of experts [Caputo et al., 2012]. The availability of this database is expected to provide researchers and geneticists rapid access to supporting evidence for classification.

Materials and Methods

The BRCA Share Databases

The BRCA ShareTM databases are derived from the UMD-BRCA1 and UMD-BRCA2 databases [Caputo et al., 2012], based on 20 years' experience of clinical BRCA testing by UGG, and augmented by the results of clinical BRCA testing by Quest and LabCorp. The UMD software [Béroud et al., 2005] has been modified to accommodate new features including a registration process and secure access to the databases under individual login IDs.

In order to integrate data from different sources and ensure an optimal curation process, a common template was used for data submission. It contains information related to the individual and related sample(s), including submitter, de-identified subject and family IDs, subject demographic information, disease status of subject and relatives, availability of cell line, tumor, or other physical materials, and so on. Subject and family identifiers are reassigned to yield globally unique anonymous identifiers which can be mapped back to the submitter's anonymous identifiers for future updates. Information related to variants includes the next-generation sequencing (NGS) screening type (whole coding sequence; targeted; single site), sequencing platform, DNA, RNA and protein HGVS names for the variant [den Dunnen et al., 2016], ID of transcript used to derive the names, the variant class (missense, nonsense, frameshift, intronic, rearrangement, insertion, deletion, isosemantic a.k.a. synonymous), the submitter pathogenicity assessment, and supporting evidence collected from the literature, from functional assays and from in silico tools. Variant nomenclature is automatically validated by the UMD software and the HGVS genomic DNA variant name (g.) is generated using the GRCh37 reference sequence. Allele frequencies are automatically collected from dbSNP [Sherry et al., 2001], the Exome Aggregation Consortium including TCGA data (http://evs.gs.washington.edu/EVS/), the 1000 genomes and the 6,500 exomes of the Exome Variant Server (http://evs.gs.washington.edu/EVS/). Finally, a link to the UMD-Predictor system [Salgado et al., 2016] is provided to access the most recent in silico predictions from this system.

The BRCA ShareTM databases are accessible at http://um.d.be/BRCA1/ and http://um.d.be/BRCA2/. They integrate a new Web interface that uses dynamic tabs and forms. It was developed using the JQuery library (http://jquery.com/). New graphical displays were also created, using the D3.js library (http://d3js.org/), featuring dynamic zoom and signals highlights.

As an example of the depth of annotation provided by these databases, consider the BRCA2 c.316+5G>C (IVS3+5G>C) variant, listed as pathogenic in BRCA ShareTM. Figure 1 shows a snippet from BRCA ShareTM for this mutation. It provides brief summaries for descriptions in the published literature, in silico splice predictions, frequency, co-occurrence, and even classifications from other public databases. Supporting evidence for pathogenic classification including references to functional studies describing aberrant splicing, frequency, and co-occurrence in patients are included in the variant record. This allows the investigator to review current data and literature for validation of the variant's classification and not solely rely on the classification from other clinical laboratories.

Two new features have been added on the welcome page to highlight changes in classification status: the "Variants recategorization" and "Variants classification" links. The first one gives access to the list of reclassified variants during the last 6 months, whereas the second gives access to variants that have been classified during the last 6 months. Note that during the registration process, the user can select to receive alerts by email when variant recategorization occurs.

Curation

BRCA Share curation is based on processes developed by the Uni-cancer Genetic Group (UGG) BRCA network that ensure clinical grade quality. It relies on a classification working group which meets regularly to provide expert curation. VUS classification is conducted using a combination of available data: frequencies in the general population, in silico predictions using bioinformatics tools at both splice and protein levels, causality or neutrality scores published in the literature [Goldgar et al., 2004; Easton et al., 2007], functional domain [Millot et al., 2012; Guidugli et al., 2013], co-occurrence with causal mutations in the same gene, co-segregation analyses in French families and published results of functional tests or splicing modeling data as described in Caputo et al. [2012]. Evidence for classification are integrated in a likelihood model previously described [Goldgar et al., 2004]. The classifications displayed in the database are limited to those derived by the UGG. Periodically, discordant classifications among participating laboratories are marked with an asterisk to alert the user toward a closer review.
Figure 1. Classification information for BRCA2 c.316+5G>C (IVS3+5G>C) is present in BRCA Share™ database with annotations of class 5.
Table 1. Current Counts of Records (Variant Occurrences) and Unique Variants in BRCA Share™ by Pathogenicity

<table>
<thead>
<tr>
<th></th>
<th>BRCA1 Records</th>
<th>BRCA1 Unique variants</th>
<th>BRCA2 Records</th>
<th>BRCA2 Unique variants</th>
<th>Combined Records</th>
<th>Combined Unique Variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Neutral (Benign)</td>
<td>70,388</td>
<td>88.2%</td>
<td>114</td>
<td>4.5%</td>
<td>45,221</td>
<td>78.3%</td>
</tr>
<tr>
<td>2. Likely Neutral (Likely Benign)</td>
<td>944</td>
<td>1.2%</td>
<td>142</td>
<td>5.6%</td>
<td>2,487</td>
<td>4.3%</td>
</tr>
<tr>
<td>3. VUS</td>
<td>2,828</td>
<td>3.5%</td>
<td>1,423</td>
<td>56.0%</td>
<td>5,742</td>
<td>9.9%</td>
</tr>
<tr>
<td>4. Likely causal (Likely Pathogenic)</td>
<td>39</td>
<td>0.0%</td>
<td>21</td>
<td>0.8%</td>
<td>46</td>
<td>0.1%</td>
</tr>
<tr>
<td>5. Causal (Pathogenic)</td>
<td>5,626</td>
<td>7.0%</td>
<td>843</td>
<td>33.1%</td>
<td>4,276</td>
<td>7.4%</td>
</tr>
<tr>
<td>Total</td>
<td>79,825</td>
<td>2,543</td>
<td>57,772</td>
<td>3,711</td>
<td>137,597</td>
<td>6,254</td>
</tr>
</tbody>
</table>

Table 2. Current Counts of Records and Unique Variants in BRCA Share™ Broken Down by Variant Type

<table>
<thead>
<tr>
<th></th>
<th>BRCA1 Records</th>
<th>BRCA1 Unique mutations</th>
<th>BRCA2 Records</th>
<th>BRCA2 Unique mutations</th>
<th>Combined Records</th>
<th>Combined Unique mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>79,825</td>
<td>2,543</td>
<td>57,772</td>
<td>3,711</td>
<td>137,597</td>
<td>6,254</td>
</tr>
<tr>
<td>Large rearrangements</td>
<td>458</td>
<td>0.6%</td>
<td>79</td>
<td>3.1%</td>
<td>59</td>
<td>0.1%</td>
</tr>
<tr>
<td>(&gt;1 exon)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>37</td>
<td>0.1%</td>
</tr>
<tr>
<td>Deletions</td>
<td>372</td>
<td>0.5%</td>
<td>58</td>
<td>2.3%</td>
<td>22</td>
<td>0.0%</td>
</tr>
<tr>
<td>Insertions</td>
<td>86</td>
<td>0.1%</td>
<td>21</td>
<td>0.8%</td>
<td>8</td>
<td>0.2%</td>
</tr>
<tr>
<td>Small deletions and insertions</td>
<td>3,145</td>
<td>3.9%</td>
<td>531</td>
<td>20.9%</td>
<td>2,538</td>
<td>4.4%</td>
</tr>
<tr>
<td>Small deletions</td>
<td>2,234</td>
<td>2.8%</td>
<td>389</td>
<td>15.3%</td>
<td>2,538</td>
<td>4.4%</td>
</tr>
<tr>
<td>Small insertions</td>
<td>911</td>
<td>1.1%</td>
<td>142</td>
<td>5.6%</td>
<td>172</td>
<td>4.6%</td>
</tr>
<tr>
<td>Point mutations</td>
<td>73,420</td>
<td>92.0%</td>
<td>1,264</td>
<td>49.7%</td>
<td>2,195</td>
<td>59.1%</td>
</tr>
<tr>
<td>Missense</td>
<td>72,227</td>
<td>90.5%</td>
<td>1,068</td>
<td>42.0%</td>
<td>1,964</td>
<td>52.9%</td>
</tr>
<tr>
<td>Nonsense</td>
<td>1,193</td>
<td>1.5%</td>
<td>196</td>
<td>7.7%</td>
<td>333</td>
<td>6.2%</td>
</tr>
<tr>
<td>Intronic mutations</td>
<td>2,593</td>
<td>3.2%</td>
<td>628</td>
<td>24.7%</td>
<td>16,103</td>
<td>27.9%</td>
</tr>
<tr>
<td>Splice sites</td>
<td>890</td>
<td>1.0%</td>
<td>167</td>
<td>6.6%</td>
<td>827</td>
<td>14.2%</td>
</tr>
<tr>
<td>(&lt;10 bp from exon)</td>
<td>1,784</td>
<td>2.2%</td>
<td>461</td>
<td>18.1%</td>
<td>15,276</td>
<td>26.4%</td>
</tr>
<tr>
<td>Indels</td>
<td>209</td>
<td>0.3%</td>
<td>41</td>
<td>1.6%</td>
<td>100</td>
<td>2.0%</td>
</tr>
</tbody>
</table>

Results

Usage and Content Statistics

The BRCA Share databases have been accessible online since July 2015. In this first year, the number of registrants has grown regularly to reach 1,119 users and the sites have been queried on average 17,000 times/month.

As a result of the establishment of the BRCA Share™ effort and the addition of clinical variants from Quest and LabCorp, the number of BRCA1 and BRCA2 variants has grown from UMD’s total of 4,838 at the time BRCA Share was launched, to its current total of 6,254 unique variants, an increase of 29.3%.

The current count of records and variants are shown in Tables 1 and 2, and broken down by pathogenicity assessment and variant type, respectively. The distribution by variation type and pathogenicity assessment is available in Supp. Table S1.

Overlap and Classification Agreement with other Collections

The Venn diagrams in Figure 2(A and B) show the counts of BRCA1 and BRCA2, variants, respectively, shared with ARUP (www.arup.utah.edu/), ClinVar [Landrum et al., 2014], the ENIGMA collection [Spurdle et al., 2012], and LOVD [Fokkema et al., 2011]. The BRCA Share™ collection currently is the second largest, behind ClinVar; both contain a large number of variants not present in the other. Note that ARUP only reports class 5 variants; its overlap with other systems would be more favorable if limited to these. Discordant classifications (2 or more classes difference) within ClinVar were treated as VUS for the purposes of comparison. Classification comparisons between the two largest sites, BRCA Share™ and ClinVar, are shown in Figure 3. Seventy-four percent of the variants are classified identically between these two sites. None of the discordances are of the clinically actionable type where a variant goes from likely pathogenic/pathogenic to likely benign/benign. Total VUS discrepancies are shown in the margins. The largest discordance categories are for variants classified as class 2 (Likely Benign) in ClinVar versus class 3 (VUS) in BRCA Share™ and class 5 (Pathogenic) in BRCA Share™ versus class 3 (VUS) in ClinVar. In addition, the comparison of variants classified as class 4/5 in BRCA Share™, ARUP, and ClinVar reveals that BRCA Share™ and ARUP are more in agreement with each other (499/546; 91.4%) than ARUP and ClinVar (400/1,062; 37.7%), possibly reflecting small differences in classification criteria.

Pairwise comparisons among BRCA Share™, ClinVar, and ARUP show that BRCA Share™ and ClinVar agree on 72% of classifications, BRCA Share™ and ARUP on 81%, and ARUP and ClinVar on 60% of shared variants. Twenty four percent of variants classified as VUS by BRCA Share™ are classified as some other category by ClinVar, whereas 19% of variants classified as VUS by ClinVar are classified otherwise by BRCA Share™.

Variant Frequencies

The BRCA Share™ database affords an opportunity to characterize the spectrum of BRCA variation in some detail, to gain understanding the relationships between frequency, type, and classification. Although population frequency is not directly available from the BRCA Share™ collection, it can be estimated from the database. This collection incorporates a number of biases of sampling and reporting, including emphasis on affected subjects and...
Figure 2. Venn diagram showing sharing of (A) BRCA1 and (B) BRCA2 variants between collections.

Co-occurrences

Co-occurrence of a VUS with a known disease-causing mutation, either from the same or the other BRCA gene, can provide evidence to classify the variant, especially if it is reported in multiple patients [Goldgar et al., 2008; Cherbal et al., 2012; Santos et al., 2014]. Co-occurrence data are used in two ways: in a Posterior Probability calculation [Goldgar et al., 2004], as it is one of its elements, and as a standalone criterion for classification. Indeed, if a given variant co-occurs with at least two different pathogenic variants in the same gene or is demonstrated co-occurring in trans with a pathogenic variant of the same gene, it is classified as Neutral. This criterion
requires that the patient has been examined by a clinical geneticist to exclude Fanconi-like for BRCA2.

Since the BRCA Share databases preserve patient-centric information, they allow co-occurrences to be readily queried, as shown in Figure 5. During curation, such evidence can provide arguments against pathogenicity, as illustrated in Table 3. For each BRCA1 VUS from a specific sample, the table displays co-occurring pathogenic or likely pathogenic BRCA1 mutations. We can distinguish cases in which the same pair of variants co-occur repeatedly, as illustrated for BRCA2 c.324T>C (p.Asn108Asn), which is specifically associated with the c.2612C>A (p.Ser871X) pathogenic mutation. This might suggest a common origin of the patients or the presence of both variants on the same allele. By contrast, the BRCA2 c.2350A>G (p.Met784Val) variant was reported in two patients from different laboratories. Each of them also harbors a different pathogenic BRCA2 mutation: c.6952C>T (p.Arg2318X) and c.5771_5774delTTCA (p.Ile1924ArgfsX38). This indicates that those mutations are probably in trans and that these patients are unrelated. Additionally, this is a strong evidence to consider the BRCA2 c.2350A>G (p.Met784Val) variant as non-pathogenic. Thus, the co-occurrence data facilitate weighting of evidence based upon the number of samples displaying a co-occurrence with another pathogenic variant.

Classification Discordances

The initial combined dataset contained 687 variants that were shared between two or more of the three participating laboratories. 67% (N = 457) of which were in BRCA1 and 33% (N = 230) were in BRCA2. Classifications were identical across the laboratories for 57% of the shared variants, whereas 69% were concordant when Likely Benign is grouped with Benign and Likely Pathogenic with Pathogenic. The remaining 31% included variants that crossed over from a VUS to a non-VUS category between laboratories. No drastic differences in classifications (i.e., Benign/Likely Benign vs. Pathogenic/Likely Pathogenic) with a potential to adversely impact patient outcomes were identified. Concordance rates were higher in BRCA1 than BRCA2, for example, exact concordances were 76% and 53%, respectively. We emphasize that these are per variant discordance rates and not per patient; since the discordances involve rare variants, per patient discordance rates would be far lower.

A process to resolve these discordances is underway, involving periodic review of the classification evidence by all the contributing laboratories. Initial review of 148 BRCA1 discordances revealed that most were attributable to changes in the available evidence since the original classification; after review the number was significantly reduced to 37, or 8.1% of the BRCA1 shared variants. Of these 37 discordances, 17 were attributable to differences in the approach toward classification of rare synonymous variants in the absence of other supportive evidence, with some groups classifying these as VUS and others as Likely Benign. The remaining 20 variants (4.4%) are all missense (18) or intronic (2), and are still under review, as are the BRCA2 variants. BRCA Share displays a primary classification for each variant, though submitter classifications are accessible. The primary classification displayed is from the French Consortium (UGG) if available or else the closest to VUS of the submitted variants is displayed. Discordant classifications among participating laboratories are marked with an asterisk.

Figure 4. A: ExAC aggregate frequency versus BRCA Share occurrence count, colored by pathogenicity, with log–log regression lines based on benign variants only (solid line) and all variants (dashed line). The asymmetric distribution of pathogenic variants reflects the pathogenic bias of the BRCA Share collection compared with the ExAC. The solid regression line was used to estimate population frequencies for BRCA Share variants in B–D; collection biases imply that frequencies of pathogenic variants are likely overestimated in BRCA Share. B: Classifications of unique variants in BRCA Share binned by log10 estimated frequency, showing the predominance of VUS and pathogenic among rare variants. C–E: proportion of variants of each type broken down pathogenicity group (likely grouped with definite).
Figure 5. Example of co-occurrence information provided for patients harboring the BRCA1 c.4644G>A mutation. This mutation has been reported in nine patients, two of them also harboring a pathogenic mutation in the BRCA1 gene: c.5260G>T (p.Glu1754X) for sample –1526 and c.342_343delTC (p.Pro115X) for patient—86EE54FQ. Also note the availability of in silico predictions above the co-occurrence table.

Discussion

Since its inception a little over a year ago, the BRCA Share™ initiative has significantly augmented the UMD dataset collected by the UGG, by leveraging the clinical sample flow of two large laboratories. The combination of US and European clinical samples has likely increased the ethnic diversity of the BRCA Share™ sample, while still leaving large portions of human diversity undersampled.

The fact that BRCA Share™ and ClinVar, the two largest collection sites, each have a substantial number of variants not present in the other, indicates that no existing collection has yet to uncover all the variants in the human population.

Classification agreement within the participating BRCA Share™ groups was generally higher than previous reports, elsewhere, with most discordances being attributable either to incorporation of newer data or likely versus VUS differences.

The BRCA Share™ funding strategy, in which academic researchers have free access, while commercial labs pay support fees, may provide a sustainable model in a time when public funding for biological databases is under stress [Check Hayden 2016; Kaiser 2016]. Public collections such as ClinVar and GA4GH provide an alternative model with fewer restrictions on access, but they rely on government funding. ClinVar has historically functioned as a submission archive, without requiring resolution of conflicting
pathogenicity classifications between different submitters, instead showing all submitted classifications for each variant. Submitters are encouraged to provide classification evidence but are not required to. Consequently, the quality of data curation can vary significantly from one variant to the next. ClinVar has implemented a star rating system to provide an indication of annotation quality. The NIH-funded ClinGen curation effort, in collaboration with ClinVar [https://www.clinicalgenome.org/data-sharing/clinvar/] is encouraging the formation of expert panels to resolve discordances with the intention to create over time a higher quality subset of ClinVar content that will be more readily usable in the clinic.

GA4GH [Global Alliance for Genomics and Health 2016] has recently announced the creation of BRCAExchange [http://brcaexchange.org/], a consolidated database that integrates all publicly available datasets on BRCA gene variants. The quality assurance and funding models for BRCAExchange are still under development. BRCA Share, by contrast, builds on the curation model developed by the French National Working Group on BRCA VUS [Grandval et al., 2013]. No such collection currently exists for BRCA1. Options for which VUSes to characterize and when include:

- Per patient, in which a functional study is performed as part of or as a follow up to the BRCA test of a particular patient specimen with the intent of improving the report for that patient [e.g., Loke et al. 2015]. This model is more appropriate for clinical labs than for a collection.
- Prioritized list, in which variants in a database are prioritized for functional studies according to some cost/benefit criteria.
- Exhaustive, in which all possible variants are generated via high-throughput mutagenesis and characterized in a high-throughput assay [Bouwman et al. 2013; Guidugli et al. 2013; Findlay et al. 2014; Starita et al. 2015; Sun et al. 2016].

An additional dimension for the design of functional studies is whether they require live patient cells or extracted DNA, or whether an in silico description of the variant will suffice. The high-throughput methods tend to require only in silico descriptions. Prioritized lists are compatible with both approaches but must be coupled to a cell, tissue or DNA bank, such as kConfab (http://www.kconfab.org/). No such collection currently exists for BRCA Share™.

We are piloting the prioritized list model using BRCA Share™ data. One obvious prioritization criterion is frequency: given a fixed capacity for doing functional studies, choosing more frequent variants will impact more patients. Since most VUS are rare, the impact of a functional study for a single variant on the total per patient VUS rate will necessarily be minimal; many VUS must be classified to have an impact. Moreover, under current ACMG Guidelines [Richards et al., 2015] functional studies can provide “strong” but not “very strong” evidence, which is insufficient to reclassify a variant as likely pathogenic or benign without 1–2 additional pieces of moderate strength evidence, or 2–3 pieces of supporting evidence. Probabilistic approaches to evidence integration may provide stronger weighting to functional studies based on demonstrated predictive value, but there is not yet a consensus on how to do this [Grandval et al., 2013].

The BRCA Share™ team has developed a prioritized list available from the “Statistics” section of the Website using the “Prioritization of VUS for validation” feature. It combines frequency (number of cases), available evidence, and co-occurrence data to produce a prioritization score, a high value means that additional evidence is available and that the addition of a functional study is likely to bring about reclassification by ACMG guidelines. In the future this may

---

Table 3. Selected Examples of Co-occurrence of BRCA1 Pathogenic Mutations in Samples with a VUS BRCA1 Mutation

<table>
<thead>
<tr>
<th>HGVS c.</th>
<th>HGVS p.</th>
<th>Sample ID</th>
<th>Class</th>
<th>Pathogenic BRCA1 mutation</th>
</tr>
</thead>
<tbody>
<tr>
<td>c.28A+G</td>
<td>p.Thr10Ala</td>
<td>04-3625</td>
<td>3</td>
<td>c.2701delC (p.Ala902LeufX2)</td>
</tr>
<tr>
<td>c.324T-C</td>
<td>p.Asn108Asn</td>
<td>00-2102</td>
<td>3</td>
<td>c.2612C-A (p.Asn705GluX11)</td>
</tr>
<tr>
<td>c.324T-C</td>
<td>p.Asn108Asn</td>
<td>01-6588A001</td>
<td>3</td>
<td>c.2612C-A (p.Asn705GluX11)</td>
</tr>
<tr>
<td>c.2320A-G</td>
<td>p.Thr774Ala</td>
<td>16-767</td>
<td>3</td>
<td>c.5576T&gt;5579delTTAA (p.Ile1859LysfsX3)</td>
</tr>
<tr>
<td>c.2350A-G</td>
<td>p.Met784Val</td>
<td>01-8210A001</td>
<td>3</td>
<td>c.6952T&gt;C (p.Arg2318X)</td>
</tr>
<tr>
<td>c.2350A-G</td>
<td>p.Met784Val</td>
<td>02-20417</td>
<td>3</td>
<td>c.5771T&gt;5774delITCA (p.Ile1924ArgfsX38)</td>
</tr>
<tr>
<td>c.2416G&gt;C</td>
<td>p.Asp806His</td>
<td>FYsmpGWHaFaYn</td>
<td>3</td>
<td>c.5786+2A&gt;G (p.Gln2496X)</td>
</tr>
<tr>
<td>c.2780T&gt;C</td>
<td>p.Met922Thr</td>
<td>07-A633</td>
<td>3</td>
<td>c.1310L1313delAAGA (p.Lys437IlefsX22)</td>
</tr>
<tr>
<td>c.2873A&gt;G</td>
<td>p.Asp946Gly</td>
<td>33-008FCF1006/33-4-292</td>
<td>3</td>
<td>c.2701delC (p.Ala902LeufX2)</td>
</tr>
<tr>
<td>c.2919G&gt;A</td>
<td>p.Ser973Ser</td>
<td>12-FK16</td>
<td>3</td>
<td>c.1184G&gt;A (p.Thr395X)</td>
</tr>
<tr>
<td>c.3152T&gt;C</td>
<td>p.Leu1051Ser</td>
<td>01-6295A001</td>
<td>3</td>
<td>c.6665C&gt;G (p.Ser2219X)</td>
</tr>
<tr>
<td>c.3170A&gt;G</td>
<td>p.Lys1057Arg</td>
<td>05-026081</td>
<td>3</td>
<td>c.1745+1G&gt;A (p.741+1G&gt;A)</td>
</tr>
<tr>
<td>c.3304A&gt;T</td>
<td>p.Asn1102Tyr</td>
<td>Ym9AfYaGCdFdsG</td>
<td>3</td>
<td>c.1929delG (p.Arg645GlufsX15)</td>
</tr>
<tr>
<td>c.3539A&gt;G</td>
<td>p.Lys1180Arg</td>
<td>02-16448</td>
<td>3</td>
<td>c.32707+1G&gt;C (p.Arg3752fsX14)</td>
</tr>
</tbody>
</table>

Notes:
- First column: cDNA HGVS nomenclature of the mutation; second column: protein HGVS nomenclature of the mutation; third column: unique sample ID; fourth column: mutation class (3 = VUS); fifth column: HGVS c. and p. nomenclature of BRCA1 pathogenic mutations co-occurring in each sample.

---
be used to prioritize allocation of BRCA Share™ funding for functional studies. Funding of high-throughput studies or construction of reagents to support them is also under consideration.

Such efforts will ultimately provide high-quality databases for actionable genes for which patients might directly benefit from adequate variant annotation during high-throughput sequencing analysis especially in the context of WES and WGS. We believe that such databases will still be needed in the future as they provide accurate annotations as well as the list of available evidence to help the end-user interpret difficult variants. They will exist in parallel to general databases such as ExAC dedicated to variant frequency. BRCA1/2 BRCA Share™ databases could today be considered as models to demonstrate that data sharing is now a reality between private companies and the academic world; that resulting data are shared with research teams as well as commercial partners engaged in the long-term sustainability of such initiatives; that strong efforts are dedicated to reclassify VUS; and finally that such systems might facilitate NGS handling and secondary findings interpretation.

Finally, although BRCA Share™ is currently limited to BRCA1 and 2, there are other clinically important genes that could benefit from a similar alliance between researchers and commercial labs. A useful next step would be to include other important cancer risk genes into a Cancer Risk Share.

Disclosure statement: The authors declare no conflict of interest.

References


