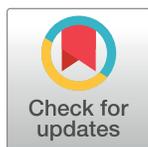


VIEWPOINTS

# BRCA Challenge: BRCA Exchange as a global resource for variants in *BRCA1* and *BRCA2*

Melissa S. Cline<sup>1</sup>✉, Rachel G. Liao<sup>2</sup>✉, Michael T. Parsons<sup>3</sup>✉, Benedict Paten<sup>4</sup>✉, Faisal Alquaddoomi<sup>5,6,7</sup>, Antonis Antoniou<sup>8</sup>, Samantha Baxter<sup>4</sup>, Larry Brody<sup>9</sup>, Robert Cook-Deegan<sup>10</sup>, Amy Coffin<sup>1</sup>, Fergus J. Couch<sup>11</sup>, Brian Craft<sup>1</sup>, Robert Currie<sup>1</sup>, Chloe C. Dlott<sup>1</sup>, Lena Dolman<sup>12</sup>, Johan T. den Dunnen<sup>13</sup>, Stephanie O. M. Dyke<sup>14</sup>, Susan M. Domchek<sup>15</sup>, Douglas Easton<sup>8</sup>, Zachary Fischmann<sup>1</sup>, William D. Foulkes<sup>16</sup>, Judy Garber<sup>17</sup>, David Goldgar<sup>18</sup>, Mary J. Goldman<sup>1</sup>, Peter Goodhand<sup>12</sup>, Steven Harrison<sup>19</sup>, David Haussler<sup>1,4</sup>, Kazuto Kato<sup>20</sup>, Bartha Knoppers<sup>21</sup>, Charles Markello<sup>1,4,22</sup>, Robert Nussbaum<sup>23</sup>, Kenneth Offit<sup>24</sup>, Sharon E. Plon<sup>25,26</sup>, Jem Rashbass<sup>27</sup>, Heidi L. Rehm<sup>28,29</sup>, Mark Robson<sup>24</sup>, Wendy S. Rubinstein<sup>30</sup>, Dominique Stoppa-Lyonnet<sup>31</sup>, Sean Tavtigian<sup>19,32</sup>, Adrian Thorogood<sup>12,33</sup>, Can Zhang<sup>34</sup>, Marc Zimmermann<sup>5,6</sup>, BRCA Challenge Authors<sup>†</sup>, John Burn<sup>35</sup>‡, Stephen Chanock<sup>36</sup>‡, Gunnar Rättsch<sup>5,6,37,38</sup>\*‡, Amanda B. Spurdle<sup>39</sup>\*‡



**1** University of California Santa Cruz Genomics Institute, University of California, Santa Cruz, California, United States of America, **2** Broad Institute, Cambridge, Massachusetts, United States of America, **3** Genetics and Computational Biology Division, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia, **4** Department of Biomolecular Engineering, University of California Santa Cruz, Santa Cruz, California, United States of America, **5** Department of Computer Science, Biomedical Informatics Group Universitätsstrasse, Zürich, Switzerland, **6** Biomedical Informatics, University Hospital Zurich, Zurich, Switzerland, **7** Biocybernetics Laboratory, Computer Science Department, University of California, Los Angeles, California, United States of America, **8** Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom, **9** Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, Bethesda, Maryland, United States of America, **10** School for the Future of Innovation in Society, and Consortium for Science, Policy & Outcomes, Arizona State University, Tempe, Arizona, United States of America, **11** Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, United States of America, **12** The Global Alliance for Genomics and Health, Toronto, Ontario, Canada, **13** Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands, **14** Centre of Genomics and Policy, Faculty of Medicine, McGill University, Montreal, Quebec, Canada, **15** Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, **16** Program in Cancer Genetics, Department of Oncology and Human Genetics, McGill University, Montréal, Quebec, Canada, **17** Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts, United States of America, **18** Huntsman Cancer Institute and Department of Dermatology, University of Utah, Salt Lake City, Utah, United States of America, **19** Partners HealthCare Laboratory for Molecular Medicine and Harvard Medical School, Boston, Massachusetts, United States of America, **20** Graduate School of Medicine, Osaka University, Osaka, Japan, **21** Centre of Genomics and Policy, Faculty of Medicine, Human Genetics, McGill University, Montreal, Québec, Canada, **22** Center for Biomolecular Science & Engineering, University of California, Santa Cruz, California, United States of America, **23** Invitae, San Francisco, California, United States of America, **24** Clinical Genetics Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York, United States of America, **25** Department of Pediatrics, Baylor College of Medicine, Houston, Texas, United States of America, **26** Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, United States of America, **27** National Disease Registration, National Cancer Registration and Analysis Service, Public Health England, London, United Kingdom, **28** Laboratory for Molecular Medicine, Partners Healthcare Personalized Medicine, Cambridge, Massachusetts, United States of America, **29** Department of Pathology, Brigham & Women's Hospital and Harvard Medical School, Boston, Massachusetts, United States of America, **30** CancerLinQ at American Society of Clinical Oncology (ASCO), Alexandria, Virginia, United States of America, **31** Institut Curie, Cancer Genetic Clinic, Paris, France, **32** Department of Oncological Sciences, The University of Utah, Salt Lake City, Utah, United States of America, **33** Centre of Genomics and Policy, McGill University, Montreal, Canada, **34** Department of Computer Science, University of California, Santa Cruz, Santa Cruz, California, United States of America, **35** Institute of Genetic Medicine, Newcastle University, Centre for Life, Newcastle upon Tyne, United Kingdom, **36** Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland, United States of America, **37** Computational Biology Program, Memorial Sloan Kettering Cancer Center, New

OPEN ACCESS

**Citation:** Cline MS, Liao RG, Parsons MT, Paten B, Alquaddoomi F, Antoniou A, et al. (2018) BRCA Challenge: BRCA Exchange as a global resource for variants in *BRCA1* and *BRCA2*. PLoS Genet 14(12): e1007752. <https://doi.org/10.1371/journal.pgen.1007752>

**Editor:** Charis Eng, Cleveland Clinic Genomic Medicine Institute, UNITED STATES

**Published:** December 26, 2018

**Copyright:** This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

**Data Availability Statement:** All classified variants and their publicly available supporting data are available at [brcaexchange.org](https://brcaexchange.org). All other relevant data are within the paper and its Supporting Information files.

**Funding:** Partial Funding from Intramural Research Program, Division of Cancer Epidemiology and Genetics, NCI, Bethesda, MD, Partial Funding from Beau Biden Moonshot Act, ABS is supported by an NHMRC Senior Research Fellowship (ID1061779), AT and BMK are supported by Can-SHARE. Can-SHARE is supported by Genome Quebec, Genome Canada, the government of Canada, the Ministère de l'Économie, Innovation et Exportation du

Québec, and the Canadian Institutes of Health Research (fund #141210), (<http://www.p3g.org/resources/can-share>), BC and MJG are supported by awards 5U24CA180951-05 and 5U24CA210974-02 from the National Cancer Institute's ITCR program (<https://itcr.cancer.gov/>), BC-D is supported by R01 HG 008918 "Building the Medical Information Commons" from the National Human Genome Research Institute (<genome.gov>), BP and MC are supported by grant U54HG007990 from the National Human Genome Research Institute (<genome.gov>), GR is supported by ETH Zurich (core) funding, JB is supported by an educational grant from AstraZeneca, HLR is supported by ClinGen grant U41HG006834 ([www.clinicalgenome.org](http://www.clinicalgenome.org)), MR is supported by the Breast Cancer Research Foundation (<https://www.bcrf.org/>) and an MSKCC Cancer Center Support grant/Core grant P30 CA008748 (<https://www.mskcc.org/>), RGL is supported by grant U24CA210999 from the National Cancer Institute (<cancer.gov>) and grant U01HG008676 from the National Human Genome Research Institute (<genome.gov>), SD is supported by the Bassett Center for BRCA (Philadelphia, PA), SEP is supported by grant 1U41HG009649, and WDF is supported by FDN-148390 from Canadian Institutes of Health Research (<http://www.cihr-irsc.gc.ca>). The funders had no role in the preparation of the article.

**Competing interests:** I have read the journal's policy and have the following conflicts: SEP is a member of the Scientific Advisory Board of Baylor Genetics, and is on the editorial board of PLOS Genetics. MR receives advisory fees, travel fees, and honoraria from AstraZeneca. The other authors have no competing interests to declare.

York, New York, United States of America, **38** Swiss Institute for Bioinformatics, Lausanne, Switzerland, **39** Genetics and Computational Biology Division, QIMR Berghofer Medical Research Institute, Herston, Brisbane, Australia

- ☉ These authors contributed equally to this work.
- ‡ Authors JB, SC, GR, and ABS are joint senior authors on this work and contributed equally.
- ¶ BRCA Challenge Authors are provided in the Acknowledgments.
- \* [Gunnar.Ratsch@ratschlab.org](mailto:Gunnar.Ratsch@ratschlab.org) (GR); [Amanda.Spurdle@qimrberghofer.edu.au](mailto:Amanda.Spurdle@qimrberghofer.edu.au) (ABS)

## Abstract

The BRCA Challenge is a long-term data-sharing project initiated within the Global Alliance for Genomics and Health (GA4GH) to aggregate *BRCA1* and *BRCA2* data to support highly collaborative research activities. Its goal is to generate an informed and current understanding of the impact of genetic variation on cancer risk across the iconic cancer predisposition genes, *BRCA1* and *BRCA2*. Initially, reported variants in *BRCA1* and *BRCA2* available from public databases were integrated into a single, newly created site, [www.brcaexchange.org](http://www.brcaexchange.org). The purpose of the BRCA Exchange is to provide the community with a reliable and easily accessible record of variants interpreted for a high-penetrance phenotype. More than 20,000 variants have been aggregated, three times the number found in the next-largest public database at the project's outset, of which approximately 7,250 have expert classifications. The data set is based on shared information from existing clinical databases—Breast Cancer Information Core (BIC), ClinVar, and the Leiden Open Variation Database (LOVD)—as well as population databases, all linked to a single point of access. The BRCA Challenge has brought together the existing international Evidence-based Network for the Interpretation of Germline Mutant Alleles (ENIGMA) consortium expert panel, along with expert clinicians, diagnosticians, researchers, and database providers, all with a common goal of advancing our understanding of *BRCA1* and *BRCA2* variation. Ongoing work includes direct contact with national centers with access to *BRCA1* and *BRCA2* diagnostic data to encourage data sharing, development of methods suitable for extraction of genetic variation at the level of individual laboratory reports, and engagement with participant communities to enable a more comprehensive understanding of the clinical significance of genetic variation in *BRCA1* and *BRCA2*.

### Author summary

The goal of this study and paper has been to develop an international resource to generate an informed and current understanding of the impact of genetic variation on cancer risk across the cancer predisposition genes, *BRCA1* and *BRCA2*. Reported variants in *BRCA1* and *BRCA2* available from public databases were integrated into a single, newly created site, [www.brcaexchange.org](http://www.brcaexchange.org), to provide a reliable and easily accessible record of variants interpreted for a high-penetrance phenotype.

## Introduction

Over 60 years ago, geneticists began to correlate cytogenetic observations with clinical significance. Over time, a succession of new genomic technologies has enabled investigators to

interrogate single-base sequence analyses in clinically informative settings. This led to the emergence of molecular diagnostics, which, in turn, prompted the development of variant databases curated by academic and/or national groups. McKusick's classic textbook evolved into Online Mendelian Inheritance in Man (OMIM; <https://omim.org/>), whereas a number of international efforts were launched to standardize nomenclature and classification—such as the Human Genome Organization (<http://www.hugo-international.org/>), Orphanet ([www.orpha.net](http://www.orpha.net)), the Human Genome Variation Society (HGVS; [www.hgvs.org](http://www.hgvs.org)), and the Human Variome Project (HVP)[1]—as well as reference databases, including the Human Gene Mutation Database (HGMD) (<http://www.hgmd.cf.ac.uk/>), DatabasE of genomiC variation and Phenotype in Humans using Ensembl Resources (DECIPHER; <https://decipher.sanger.ac.uk/>), Leiden Open Variation Database (LOVD; [www.lovd.nl](http://www.lovd.nl)), and ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>). Over the last decade, the expansion of next-generation sequencing and its deployment in routine diagnosis have made variant interpretation a critical rate-limiting factor in the realization of clinical benefit.

At the first partner meeting of the Global Alliance for Genomics and Health (GA4GH) in 2014, it was agreed to focus one of its demonstration projects on the problem of variant capture and interpretation, drawing on both traditional clinical resources and the availability of “big data” to illuminate normal variation. The genes *BRCA1* and *BRCA2* were chosen because of their iconic status in the public consciousness and the scale of routine use of genetic test results to guide major clinical decisions. At the recent launch of the restructured GA4GH, “GA4GH Connect” [2], BRCA Challenge was identified as one of the inaugural driver projects for an international effort to enable more robust genomic data sharing in the future. As of October 2017, it has been relaunched as an independent effort and is working with GA4GH to engage in the strategic development and uptake of data sharing frameworks and standards. From the outset, the BRCA Challenge has convened a spectrum of stakeholders to accurately interpret *BRCA1* and *BRCA2* variation, to make this information freely available worldwide, and to improve the care of patients who present with variants in *BRCA1* or *BRCA2* (Box 1).

Inherited variation in the *BRCA1* and *BRCA2* genes can indicate genetic predisposition to breast, ovarian, and other cancers, but a large proportion of observed variants are not disease associated (“pathogenic”)[3, 4]. In response to the need for an international solution to provide

### Box 1. Specific goals of the BRCA Challenge

1. Share *BRCA1* and *BRCA2* variants publicly.
2. Create an online environment for collaborative variant curation with access to evidence (e.g., phenotypes, family history, genetic data, and functional studies).
3. Create a curated list of *BRCA1* and *BRCA2* variants, interpreted by expert consensus, to serve as a reference for accurate clinical care based on criteria established by the ENIGMA consortium.
4. Address social, ethical, and legal challenges to global data sharing, and engage with patient advocacy organizations from around the world.
5. Create a model system for data sharing for other disease genes.

a comprehensive annotation of *BRCA1* and *BRCA2* variants, we have developed the BRCA Exchange—a single comprehensive portal to display global *BRCA* variation, linking current structures and resources while encouraging deposition of new data. The BRCA Exchange is based on shared data from clinicians, clinical laboratories, and researchers across the world, including existing clinical databases—Breast Cancer Information Core (BIC), ClinVar, and LOVD—linked to a single point of access. An international interpretation community includes the Evidence-based Network for the Interpretation of Germline Mutant Alleles (ENIGMA) consortium [5], established to investigate the clinical significance of variants in breast–ovarian cancer predisposition genes, and which is recognized by Clinical Genome Resource (ClinGen) as an expert panel for interpretation of *BRCA1* and *BRCA2* variants. The purpose of the Exchange is to provide the community with a reliable and easily accessible record of variants interpreted as pathogenic for a high penetrance phenotype; the Exchange also reports on variants deemed “not clinically relevant” (nonpathogenic) and eventually will report on reduced penetrance variants associated with a moderate or modest risk for a spectrum of cancers (such as *BRCA2* p.K3326\*)[6].

### The challenge of interpreting *BRCA1* and *BRCA2* variants

In 1866, the French physician Paul Broca described a cluster of breast cancers in his wife’s family, heralding the concept of familial risk for breast cancer [7]. The concept of a cancer-predisposition gene emerged as a critical driving force for applying the evolving tools of genetic analyses to map cancer genes in families presenting with breast and/or ovarian cancer; this view emerged over decades of studies describing familial clusters, twin studies, and special populations at high risk for breast and ovarian cancer. In 1990, what is now known as *BRCA1* was localized to the long arm of chromosome 17 through analyses of high-risk families [8]. Four years later, positional cloning and mutational analyses identified the gene itself and revealed specific coding variants (predominantly truncating alleles) strongly linked to risk for breast and ovarian cancers within families and high-risk populations [9]. A second major cancer gene for breast and ovarian cancer, *BRCA2*, was localized to the long arm of chromosome 13 in 1994 and cloned and sequenced in 1995 [10–12]. By 1996, genetic testing of these two genes was introduced into clinical practice. Although millions of women, and more recently men, at risk for cancer have been tested for germline genetic variants in *BRCA1* and *BRCA2*, diagnostic sequencing has continued to identify variants with uncertain relevance to cancer risk. This has emerged as a particularly challenging issue for determination of risk for not only breast and ovarian cancer but now aggressive prostate and pancreatic cancers. What has emerged is that whereas some germline variants confer high risk for cancer due to disruption of gene function [13, 14], many *BRCA1* and *BRCA2* variants identified during routine genetic testing are determined to have little or no clinical significance with respect to cancer risk [4, 15, 16].

Despite the low overall prior probability of pathogenicity for a variant of uncertain significance in these two genes [4, 17], such variants nevertheless present a clinical challenge in that they complicate test reporting and medical decision-making for clinicians and patients, and carriers of “equivocal” variants are more likely to over-interpret their result, seek additional management, or in some circumstances, receive overtreatment by clinicians outside clinical genetics [18–20]. There are also *BRCA1* and *BRCA2* variants that appear to be associated with a modest or moderate (but still elevated) risk of breast and ovarian cancer, compared with “traditional” deleterious mutations. The clinical reporting on these variants is often inconsistent and therefore confusing [6, 21–24].

Over decades of study and sequencing at-risk individuals for these two genes and others, it has become clear that there is a need for a consistent variant classification scheme that is

systematically developed, validated, peer-reviewed, and published as authority for use by diverse stakeholders. Previously, the International Agency for Cancer Research (IARC) [25] and the American College of Medical Genetics and Genomics (ACMG) with the Association for Molecular Pathology (AMP) [26] proposed schemes for classification of variants as pathogenic or nonpathogenic, incorporating evidence using quantitative and qualitative approaches. The utility of these classification efforts has been seen in their incorporation into clinical pipelines, but the practice of interpreting variants remains a daunting challenge. It may be difficult for individual testing laboratories to have sufficient clinical or laboratory information to assign classification for rarely encountered variants. Furthermore, the rapid increase in sequence-based testing, particularly among individuals with no personal or family history of cancer, has led to more complex variant interpretation. Many sequencing centers have begun to aggregate observed variants in databases and to classify them using clinical information and variant-level evidence from multiple sources—thereby reducing uncertainty. However, even between large laboratories, discordant reporting of variant interpretations continues due to the balkanization of evidence and differences between curators in utilization of “standard” classification criteria and/or publicly available evidence [27].

The ENIGMA consortium (<http://www.enigmaconsortium.org>) is an international multidisciplinary consortium [5], currently comprising more than 300 listed members from over 190 research groups and/or clinical testing laboratories spanning more than 35 countries. The aim of the ENIGMA consortium is to develop methods for improved classification of variants in *BRCA1* and *BRCA2* and other known or suspected breast cancer predisposition genes to guide their use in clinical testing panels. The National Institutes of Health (NIH)-funded ClinGen (<https://www.clinicalgenome.org/>) has a process to appoint expert panels to provide expert variant curation for specific genes. The ENIGMA consortium was approved as an expert panel by ClinGen to apply its methods to define more precise criteria for annotation of variants in *BRCA1* and *BRCA2*.

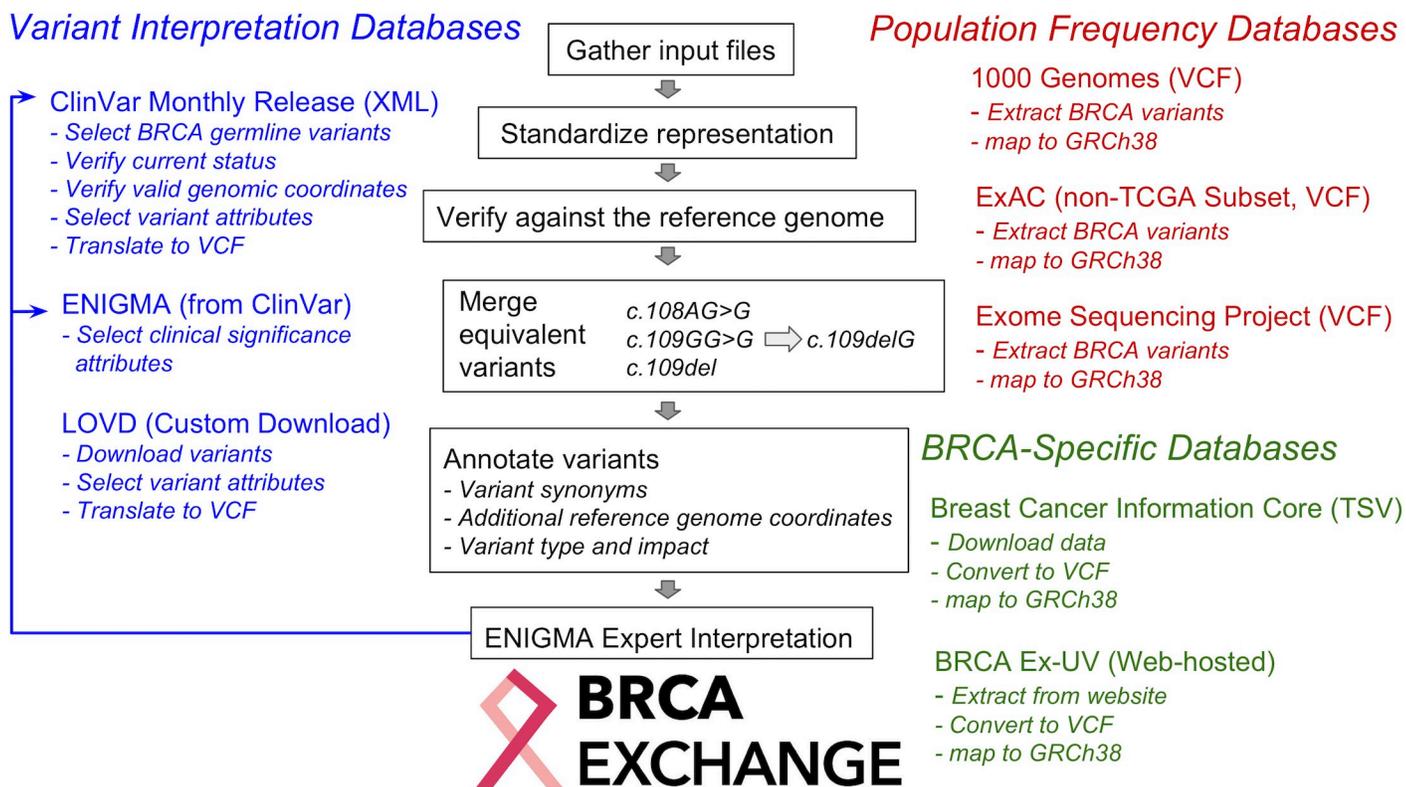
It is anticipated that the number of variants of uncertain clinical significance (VUSs) and differences in interpretation of the same variant between different laboratories should decrease with time as public databases approach a fully comprehensive annotation of variants in familial and population-based settings together with availability of the associated information relevant to variant classification. It is the express purpose of the BRCA Challenge project and the BRCA Exchange website to accelerate this process, using *BRCA1* and *BRCA2* as demonstration genes. In this regard, interpretation of currently identified VUSs can be accelerated by enabling increased sharing of existing variant information and using the additional observations to improve evidence-based classification. Unfortunately, the willingness to share varies among data holders for a variety of reasons, including ethical and legal challenges with respect to sharing patient information, resource limitations, and commercial and national health system models commodifying patient and variant data. Databases often aggregate incompletely curated variant classifications with little or no supporting evidence, further muddying the waters of appropriate interpretation and clinical care.

To date, the scope of *BRCA1* and *BRCA2* testing has been substantially larger than any other cancer predisposition gene, yet the ability to look across the spectrum of publicly available data has not been coordinated. The emergence of multiple public and private databases worldwide, together with proprietary diagnostic testing databases, have retarded efforts to link a large set of well-annotated variants for public interpretation. The rapid rise of next-generation sequencing by academic centers, commercial diagnostic companies, and national programs has created an additional opportunity to collect information—especially variants derived from non-*BRCA1* and *BRCA2* familial cases—and use these to classify more variants. Together, these factors underscore how *BRCA1* and *BRCA2* are a compelling case for sharing

data to advance the precision of variant interpretation, thus enabling clinicians and patients to make important decisions together.

### Development of a public portal for *BRCA1* and *BRCA2* variants: BRCA Exchange

The BRCA Challenge has developed an open-access web portal, the BRCA Exchange (<http://brcaexchange.org>), a resource to display *BRCA1* and *BRCA2* variants drawn from global sources and to enable *BRCA1* and *BRCA2* variants to be expert-reviewed, interpreted, classified, and aggregated in an integrated data system (Fig 1). The publicly accessible display of these classifications, with supporting evidence, facilitates accurate understanding of the clinical relevance of any individual *BRCA1* or *BRCA2* variant. The portal is fully open access and enables easy download of variants and variant-level evidence by anyone with an internet connection. The global focus of the portal enables coordination and collaboration with relevant investigators and databases from around the world in a supranational consortium, thus



**Fig 1. Data flow in the BRCA Exchange pipeline.** BRCA Exchange combines information from major public sources to offer a comprehensive view of *BRCA* variants from a single web portal. It combines variant information from ClinVar, ENIGMA (as the ClinGen expert panel on *BRCA* variation) and LOVD (blue); population frequency data from 1000 Genomes, ExAC, and the ESP (red); and *BRCA*-specific information from the BIC on *BRCA 1* and *2* Ex-UV (green). Each month, BRCA Exchange collects variant data from these sources and translates them into a consistent representation. It verifies that all variants are consistent in the reference bases with the reference human genome and discards any variant data that are inconsistent with the genome. Next, it identifies functionally equivalent variants, in which two or more variants might produce the same alternative allele despite distinct representation, and merges equivalencies to generate a set of distinct *BRCA* variants. It gathers annotations, such as functional impact terms and alternative variant names. Finally, it compares the new variant data to the previous month to identify any variants that are new or updated. These data are shared publicly at [brcaexchange.org](http://brcaexchange.org). The ENIGMA consortium analyzes these aggregated data to determine the clinical significance of unreviewed variants and deposits these interpretations in ClinVar, completing the cycle of information. BIC, Breast Cancer Information Core; ENIGMA, Evidence-based Network for the Interpretation of Germline Mutant Alleles; ESP, Exome Sequencing Project; ExAC, Exome Aggregation Consortium; *BRCA* Ex-UV, *BRCA1* and *BRCA2* Ex-UV; GRCh38, Genome Reference Consortium Human Build 38 Organism; LOVD, Leiden Open Variation Database; TCGA, The Cancer Genome Atlas; TSV, Tab Separated Values; VCF, Variant Format Call; XML, Extensible Markup Language.

<https://doi.org/10.1371/journal.pgen.1007752.g001>

creating the most comprehensive source of variation in *BRCA1* and *BRCA2*, including high-quality variant classifications made by ENIGMA.

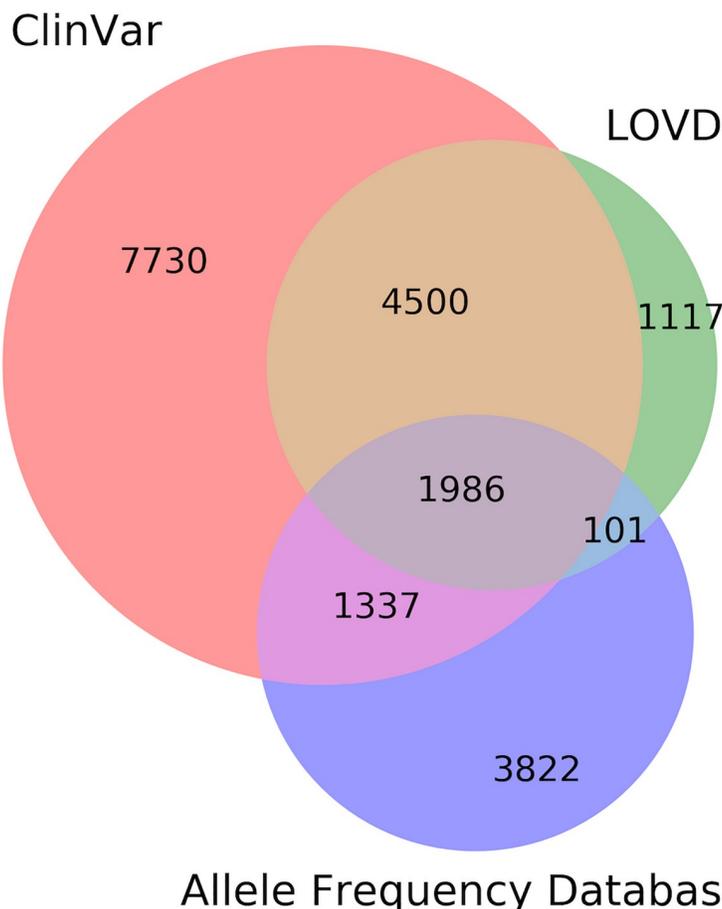
To ensure broad usage of the BRCA Exchange, features have been added to maximize utility of the data. The default view displays only expert-curated variant classifications, in order to support layperson engagement with the data. Users who wish to access more extensive data can click through a disclaimer and arrive at a version of the portal that displays the full database, inclusive of classifications made by LOVD and ClinVar submitters and curators other than ENIGMA, with potential interpretation conflicts. Data quality is ensured by the back-end pipeline, which merges equivalent variants and filters out variants with erroneous nomenclature terms or reference bases that are inconsistent with the current reference genome. These data can be downloaded in full and are available for third-party use. The Exchange data are also versioned with each new data upload, in line with regulations for clinical-grade databases.

The *BRCA1* and *BRCA2* variant data set comprises data from participating groups and major academic data sources, including ClinVar (with diverse submitters such as academic groups, commercial and hospital laboratories, and consortia such as the Consortium of Investigators of Modifiers of *BRCA1/2* [CIMBA] [28, 29]), BIC (also represented in ClinVar) [30], Exome Aggregation Consortium (ExAC)/Genome Aggregation Database (gnomAD) [16], the 1000 Genomes Project, [31] the *BRCA1* and *BRCA2* Ex-UV Database displaying classifications based on multifactorial likelihood analysis [32], LOVD [33], NHLBI GO Exome Sequencing Project (ESP) [34], and the ENIGMA consortium [5]. Currently, the BRCA Exchange displays over 20,000 unique *BRCA1* and *BRCA2* variants; more than 6,100 of those variants are displayed with expert classifications from the ENIGMA Consortium and to date approximately 3,700 have been deemed pathogenic. Variant numbers currently contributed by ClinVar, LOVD, and the three population frequency databases (ExAC, 1000 Genomes, and ESP) are shown in Fig 2, with unique and overlapping variants across those data repositories also indicated. Currently, the BRCA Exchange is the most comprehensive, publicly available representation of *BRCA1* and *BRCA2* variant information.

Ongoing work is focused on integration of new data and further annotation of the large set of VUSs; it is estimated that perhaps 80% to 85% of VUSs will ultimately be judged as non-pathogenic [4, 16], but large-scale aggregation of data is required to support this assertion [3–5, 17, 35]. To ensure the fidelity of the aggregated data, we filter the input data to ensure that the variants are consistent with the reference genome and are expressed in correct HGVS nomenclature. We address duplication in the genomic data by identifying and merging variants that produce equivalent allele strings. In some cases, we receive data on the same variants from multiple sources. For example, we receive variant classification data directly from ENIGMA and BIC and indirectly from these sources via ClinVar. Although these submissions may represent the same variants, they do not provide exactly the same information for each variant, and we adapt for this with separate tiles to show the variant attributes specific to each data source.

We have not yet addressed the issue of noise in the clinical phenotype or other sequence-unrelated data. One reason is because we intend to show data “as is” from various sources. To reduce the impact of this noise, the ENIGMA-curated classifications include high-level statements to support the assertions of the classifications. Moving forward, as we begin collecting individual-level data (see Box 2), it may be possible to assign value judgment on data depending on the source (e.g., recording discordances with expert-curated variants). Although it will likely be impractical and unethical to review actual clinical records for the most part, as an example, we may ask submitters for generic details on the source of their information and note these accordingly (e.g., self-reported cancer, cancer registry confirmed cancer, and cancer diagnosis extracted from pathology report provided to submitter).

## Number of Variants per Contributor



**Fig 2. Venn diagram showing variants per large contributing data repositories, with variant overlap between repositories indicated.** Allele frequency databases currently include ExAC, 1000 Genomes, and ESP. ESP, Exome Sequencing Project; ExAC, Exome Aggregation Consortium; LOVD, Leiden Open Variation Database.

<https://doi.org/10.1371/journal.pgen.1007752.g002>

We are also exploring ways to avoid duplication of individual-level data because we plan to move to collection of such data. It is likely that ethical constraints regarding access to identifying information will require specific ethically approved protocols to allow curator access to identifying information in a secure environment that will allow for crosschecks, such as similarity in pedigree structure, ages, and types of cancer diagnoses. Access to additional accompanying genetic data—in particular, common variation—has utility to detect duplicated or closely related individuals. Complex pseudoanonymization protocols could prove useful to access deidentified data with need for individual-level consent (see below), but the issues around detecting and adjusting for potentially duplicated data points will remain.

In addition to the aggregation of large variant repositories, the BRCA Exchange works with data sets specific to local populations around the world, especially national initiatives such as the Brazilian Initiative on Precision Medicine (BIPMed), to support data submission. All national governments have been approached to encourage participation in the BRCA Challenge by the United Nations Education, Science and Culture Organization (UNESCO), coordinated by Global Variome ([www.humanvariomeproject.org](http://www.humanvariomeproject.org))—an associated nongovernmental

## Box 2. Future plans for BRCA Challenge

### Toward a comprehensive annotation of genetic variants of *BRCA1* and *BRCA2*

1. Establish consistency and collaboration in classifications across testing organizations and local databases
2. Consistent information for BRCA families internationally
3. Global representation of data deposition
4. Use collated individual-level data for developing new classification algorithms
5. Develop safe display solutions for case-level data

### Advances for technical development

1. Define a standardized model for international data sharing
2. Develop APIs for distinct shareholders—both internationally and with respect to laboratory and healthcare providers
3. Pseudonymization software to accelerate annotation of remaining VUS

### Outreach

1. Educate country-specific funding bodies on the need to develop a model for federated funding of data-sharing initiatives
2. Full engagement of patient advocacy

organization, a member of GA4GH, and a partial sponsor of the BRCA Challenge. When possible, it is recommended that these projects structure data elements for inclusion using the LOVD open-source software or by direct upload to ClinVar, but it is also possible for these variant sources to add additional data elements to the Exchange directly in order to capture the most exhaustive evidence available.

The BRCA Exchange has initiated engagement with patient and advocacy groups, primarily focusing on getting input from key patient advocates and stakeholders with plans to raise awareness of the importance of data sharing among global patient and advocate communities and to encourage patient and advocate groups to engage with data holders to promote sharing of their data. A further goal is to develop tools to make the database meaningful and useful to patients and advocates, and to that end, a pilot effort recently released a mobile phone application for simple access to the variant database and push notifications when classifications are updated for variants of interest.

## Technical framework for BRCA Exchange

The BRCA Exchange software has been developed as open-source code found on <https://github.com/BRCAChallenge/brca-exchange/>. Technical development has been led by researchers at University of California, Santa Cruz, and ETH Zurich and hosted on the Microsoft Azure cloud. The central component of the BRCA Exchange software is the data analysis pipeline (Fig 1). Once per month, the automated pipeline downloads variant data from the ClinVar [36] and LOVD [33] variation repositories, the BIC [30] and BRCA Ex-UV [32] locus-specific repositories, and the ExAC [37], ESP [34], and 1000 Genomes [31] population genetics repositories. Additionally, the ENIGMA consortium [5] sends new variant interpretations to BRCA Exchange when submitted to ClinVar.

After collecting these data, BRCA Exchange determines the GRCh38 (Genome Reference Consortium Human Build 38 Organism) genomic coordinates of each variant with the HGVS python library [38] and CrossMap [39]. The pipeline identifies and removes variants with erroneous genomic representations, such as HGVS strings that are inconsistent with the HGVS standard and cannot be parsed, or variants for which the reference allele is no longer consistent with the reference genome. In the most recent release (version 21; July 2018), the pipeline detected and removed 1,330 erroneous variants. The pipeline then identifies and merges variants that are genomically equivalent by deriving an alternative allele string for each variant that consists of the variant plus flanking genomic bases, comparing these allele strings and merging any variants that yield equivalent allele strings. With this strategy, the pipeline detects equivalent variants that show no apparent similarity in their HGVS strings or genomic coordinates, such as complex indels that produce the same alternative allele string through differing combinations of insertions and deletions. Based on this strategy, sets of variants from the input repositories are merged into one set of distinct variants. In the most recent release, the pipeline merged 1,773 pairs of equivalent variants. BRCA Exchange publishes a Docker container on Dockerhub (<https://dockerhub.com>) to repeat these processing steps.

The pipeline also computes *in silico* predictions for each variant to indicate whether the variant is likely to disrupt splicing, introduce a premature stop codon, and/or impact a key domain or functional residue (see annotation below). These data are stored in a PostgreSQL database, in which they serve the BRCA Exchange web portal, mobile application, and application programming interface (API).

As part of this monthly update, BRCA Exchange tracks the update history of each variant. The update history is displayed in the variant details pages and includes dates of variant data updates, indications of the specific data updated, and links to download previous versions of the variant data. In addition to summarizing the changes for each variant, BRCA Exchange tracks and displays changes to ClinVar and LOVD for each variant to indicate how and when a clinical or functional interpretation has been updated by its submitter. Users can access versions of the variant data from any past date, in line with regulations for clinical-grade databases.

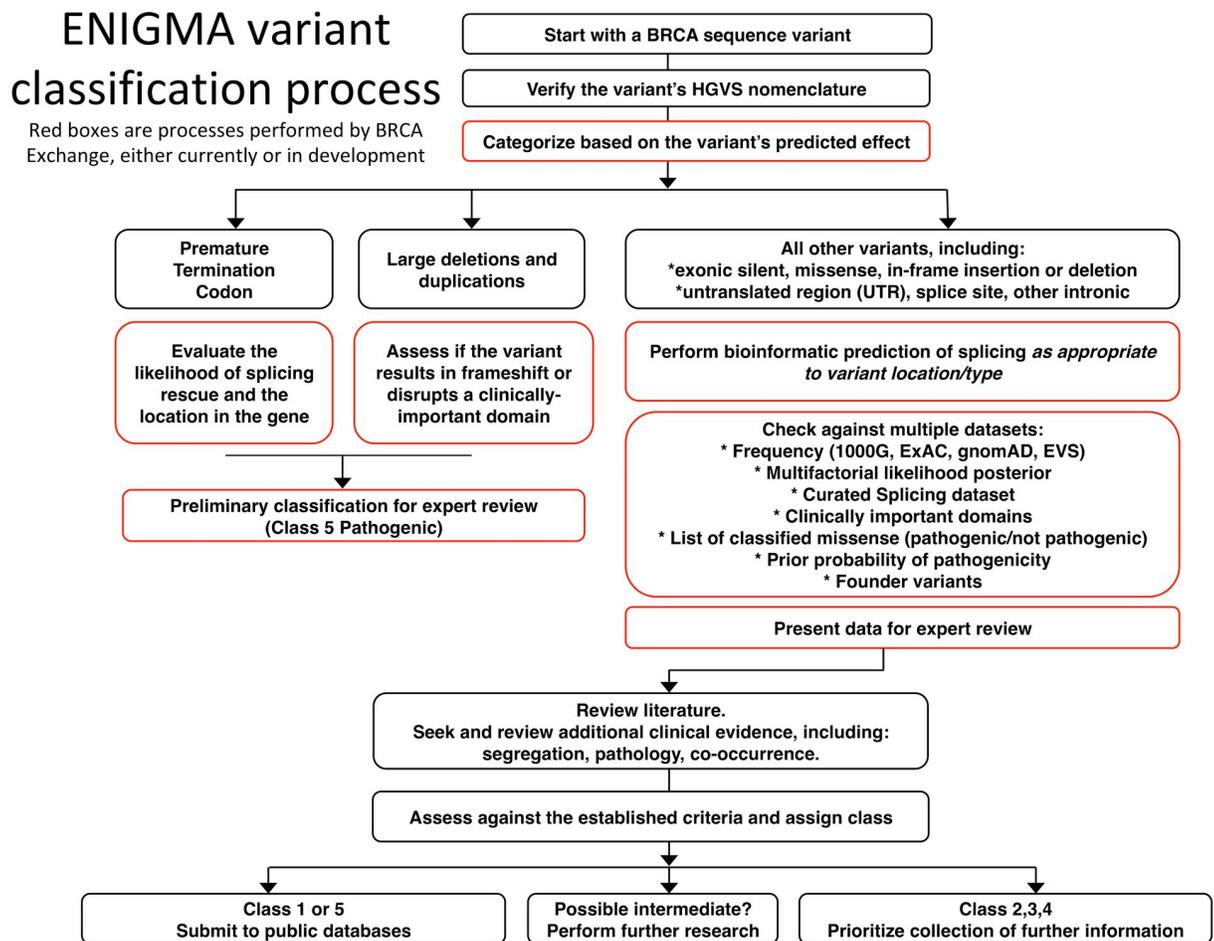
The variant data can be downloaded via the API or web portal. Under the data use policy, these data can be freely queried, downloaded, reused, and redistributed. The web portal offers downloads of the latest variant data and all previous data releases since the launch of the project. Each data release is distributed as a tarball that contains not just the final variant data but also the intermediate output files and log files that were generated by the pipeline execution, to increase the transparency and repeatability of these analysis steps. By using the BRCA Exchange Docker container published at Dockerhub, a user can replicate the variant merging and validation portions of the pipeline from this previous data release data for all releases subsequent to February 2018.

To assess the quality of the resulting data, we have compared the expert interpretations from ENIGMA and consensus classification based on available ClinVar, LOVD, BIC, and ENIGMA classifications with the variant interpretations from BRCA Share (S1 Table) [40]. BRCA Share is an independent repository of variants from the Universal Mutation Database (UMD) and the testing labs Quest Diagnostics and Laboratory Corporation of America (LabCorp). The contents of BRCA Share are available to academic researchers free of charge and to commercial partners under a subscription fee. We have not yet been able to integrate BRCA Share with the BRCA Exchange variants due to these differences in the fee structure (the contents of BRCA Exchange are freely available to all) but we would welcome the opportunity to do so in the future should this issue be surmounted. As revealed in the comparison, the repositories are largely concordant (S1 Table). Because the content of these repositories was developed independently, this concordance reflects favorably on the quality of the data. Each repository contains some variants that are not seen in the other, and in many cases, offers clinical interpretations of these variants. This point underscores the value of data sharing, and the potential for additional data sources to provide new knowledge.

### Annotation framework: Variant curation and interpretation

The BRCA Exchange seeks to comprehensively gather or compute data for authoritative variant curation using classification approaches developed by the ENIGMA consortium and to display these classifications with supporting rationale. The detailed criteria developed by ENIGMA for *BRCA1* and *BRCA2* variant classification have been informed by clinically directed research and are publicly available (<http://www.enigmaconsortium.org>). These criteria will evolve as new information, including findings from ongoing research, accrues. The BRCA Exchange resource will incorporate data from many sources, including next generation sequence data from diagnostic laboratories, and data sets relevant for the ENIGMA variant classification process that have been collated from publicly available and /or other unpublished information and curated for accuracy. The key differences of the ENIGMA *BRCA1* and *BRCA2* classification criteria to generic classification criteria include the following: both qualitative and quantitative (statistically derived) approaches can be applied; variant interpretation considers a catalog of clinically important protein domains, derived by overlaying extensive knowledge of protein function and clinical effect of missense alterations; knowledge of naturally occurring mRNA transcripts has been used to inform the interpretation of possible spliceogenic variants; bioinformatic prediction of variant pathogenicity due to alteration at the level of protein (missense or in-frame deletion) or mRNA transcription (donor, acceptor, other exonic/intronic sequences like splice enhancers/silencers, and exonic de novo donor site) have been clinically calibrated for *BRCA1* and *BRCA2* sequences [16, 32] (Fig 3). It is notable that reference to population data sets is a key initial step used to parse variants classified as benign or likely benign because of frequency incompatible with high risk.

Automated processes to prioritize variants within the BRCA Exchange for streamlined classification and to support the continued curation of variants are under active development, including calculation of missense alteration prediction scores and prior probability of pathogenic as well as reference to knowledge of clinically important functional domains. In addition, variant annotation has been expanded to include automated calculation of splicing prediction scores and secondary annotation of predicted effects on mRNA splicing. This will include reference to basic variant position and/or effect annotation to define if a variant leads to gain or loss of splicing function, assessment of the score relationship to nearest donor or acceptor, as well as the predicted effect of likely mRNA aberration, e.g., out-of-frame versus in-frame. These will be made available via the BRCA Exchange and integrated into future versions of the automated variant annotation process.



**Fig 3. ENIGMA variant classification process.** ENIGMA, Evidence-based Network for the Interpretation of Germline Mutant Alleles; ExAC, Exome Aggregation Consortium; gnomAD, Genome Aggregation Database; HGVS, Human Genome Variation Society; ins-del, insertion-deletions; UTR, untranslated region; 1000G, 1,000 Genomes Project.

<https://doi.org/10.1371/journal.pgen.1007752.g003>

### Ethical and regulatory considerations

In collaboration with the GA4GH Regulatory and Ethics Work Stream, the BRCA Exchange has considered the importance of local as well as global issues surrounding patient data sharing. Recognizing that barriers to data sharing are more than technical, the BRCA Challenge not only explores technical solutions such as container methods but also receives support and insight from a highly active Regulatory and Ethics Foundational Work Stream as part of its relationship as a Driver Project of GA4GH. At the same time, the national context can inform the use of data and protections required according to local laws and standards, adding complexity to the BRCA Exchange’s goal to aggregate diverse multinational resources. However, it is possible to develop shared solutions to enable annotated data to be shared widely and better inform the public as to the significance of specific variants in *BRCA1* and *BRCA2*—some of which could have widely differing allele frequencies based on population histories.

Sharing can result in extensive benefits through improved clinical care for patients, but it also brings concerns that public data may be misinterpreted by patients or providers or misused to the detriment of patients [41]. The responsibility to ensure data quality and responsible use within an online, large-scale data-sharing and curation community such as the BRCA Challenge needs to be carefully laid out, delegated, and clarified through data submission and

data access and use agreements; for BRCA Exchange, these have been drafted with support from GA4GH to ensure maximal uptake from global contributors. BRCA Exchange is also preparing materials for genetic counselors as well as nonspecialist physicians and advocacy and support organizations, as well as considering engagement strategies with such organizations to enable their participation in design, governance, and oversight. To this end, engagement with patient advocacy groups has been a central goal—to ensure the community has an accurate and useful understanding of the public resources. All data were obtained from third-party sources and no additional ethics approval was required.

As the call for global data sharing increases, it is critical for the BRCA Challenge to develop strategies to enhance the ability to share variant annotation—and eventually primary data—through safe and protected mechanisms. One such mechanism is a container software architecture, in which data can be queried in its home institution, with anonymized and aggregated results exported. Additionally, the BRCA Challenge supports the use of pseudonymization software to enable a more comprehensive understanding the clinical significance of genetic variation in *BRCA1* and *BRCA2*, which, in turn, should lead to reclassification of a fraction of the current VUSs based on new informative data drawn from disparate sources, including diagnostic, clinical, and national nodes.

For example, a BRCA Challenge case-level national study is underway involving the National Cancer Registration and Analysis Service (NCRAS) in Public Health England that assembles, in near real time, the world's most detailed clinical data set on all 530,000 cancers diagnosed by the National Health Service (NHS) each year in the 56 million people of England and Wales. To date, 14 of the country's 17 genomic laboratories responsible for BRCA testing have fully engaged and have uploaded *BRCA1* and *BRCA2* variant data from over 30,000 patients using in-house open-source pseudonymization software; the irreversible pseudonymized reports are matched using NHS Number plus fuzzy matching based on other identifiers. The pseudonym allows anonymous records derived from the same individual to be linked without the identity of the individual being revealed. Individual clinical data is also encrypted with its own key and only matching patients' clinical data is decrypted and made available to the NCRAS's cancer registration database. Data is collected by the NCRAS under Regulations 2 and 5 of Section 251 of the NHS Act 2006. The Regulation provides the legal gateway to set aside the common-law duty of confidence to allow data to be collected without direct consent to support healthcare functions. This is validated by the local "Caldicott" data guardian at each center. This approach pioneers the hypothesis-free linkage of molecular variants with electronic patient records to build an unbiased prospective knowledge base of cancer predisposition.

## Future directions

The BRCA Challenge represents an important step in the sharing and collaborative annotation of *BRCA1* and *BRCA2* variants to advance human health globally. The urgent need to develop stable and comprehensive solutions for sharing data in cancer predisposition and other disease syndromes has been articulated by many, especially in response to the avalanche of clinical testing underway across the globe [42–45]. Because much of this is not necessarily driven by the classical paradigms based on familial risk, the BRCA Challenge is positioned to provide a stable, global solution that accounts for important differences with respect to academic, commercial, and national interests through the development of tools such as the BRCA Exchange and its corresponding mobile application. The BRCA Challenge is positioned to bridge gaps between academic approaches to variant classification—often conducted in familial collections—with population-based next-generation sequence data. The inclusion of this latter data source has begun to alter how and in what context risk for cancer could be estimated.

For the BRCA Challenge, the first development phase of a mobile application—as part of the BRCA Exchange—is complete, allowing individuals to track altered classification of variants. In turn, this information should be useful for clinicians and patients faced with decisions that will continue to be driven by recommending bodies and the critical relationship between patient and care provider. Engagement with patient advocacy groups as well as national programs should provide guidance to refine mobile applications to optimally fit the needs of many different constituents worldwide. The commitment of the BRCA Challenge to global data sharing—both for *BRCA1* and *BRCA2* variants as well as creating a standard applicable to other genes—can only succeed with involvement of the full spectrum of stakeholders.

As the project progresses, next steps will include collaboration in data aggregation and interpretation with more global data generators and data holders, continued technical development, and increased engagement with patients and patient advocates from around the world to further this important work (Box 2). With the development of the BRCA Exchange, our dedicated team has delivered a solution to improve data sharing and variant classification in the *BRCA1* and *BRCA2* genes and thus to improve clinical care for clinicians, patients, and their families. The BRCA Challenge serves as a demonstration project to bring together clinicians, researchers, data scientists, patients, and advocates to work toward a new model of data sharing. Its purpose is to create a comprehensive data-sharing and curation system that could be easily applied to the more than 114 other cancer predisposition genes identified to date [46], and ultimately, to all genes underlying recognized disease phenotypes, currently estimated to be around 3,000. In conclusion, the BRCA Challenge is a model for systematically engaging broad communities and resources for sharing data across genes and eventually regulatory regions of major clinical importance.

## Supporting information

**S1 Table. Comparison of BRCA Exchange to BRCA Share.** The BRCA Share [40] repository contains variants from the UMD and the testing labs Quest Diagnostics and Laboratory Corporation of America (LabCorp). The contents of BRCA Share are available to academic researchers free of charge and to commercial partners under a subscription fee. To assess ENIGMA's concordance with BRCA Share, we analyzed 5,452 variants—most of the BRCA Share set—by running them through the BRCA Exchange pipeline. BRCA Share contained 3,938 variants not assessed by ENIGMA, and BRCA Exchange contained 4,640 variants that were assessed by ENIGMA but not found in BRCA Share. In addition, we considered a consensus classification based on available ClinVar, LOVD, BIC, and ENIGMA classifications. For these consensus classifications, there were 2,228 variants in BRCA Share with no reference assertion in BRCA Exchange and 13,341 with reference assertions in BRCA Exchange but no assessment in BRCA Share. We compared the clinical significance assessments of the BRCA Share variants with the 6,154 assessments from ENIGMA and the 16,565 BRCA Exchange consensus reference interpretations. Overall, the clinical assessments showed a strong similarity. “Most Neutral” or “Likely Neutral” variants in BRCA Share were assessed as “Benign” or “Likely Benign” in BRCA Exchange, and most “Causal” or “Likely Causal” variants in BRCA Share were assessed as “Pathogenic” or “Likely Pathogenic” in BRCA Exchange. The largest differences were in the VUS counts. BRCA Share contained 310 variants designated as VUS that were assessed by ENIGMA—296 as “Benign” or “Likely Benign” and 14 as “Pathogenic.” BRCA Share also contained 473 variants designated VUS that were assessed in BRCA Exchange consensus—431 as “Benign” or “Likely Benign,” and 42 as “Pathogenic” or “Likely Pathogenic” (there are also 1,177 variants listed as VUS in both BRCA Share and BRCA Exchange consensus). In comparison to the number of variant classifications analyzed (5,552

from BRCA Share, 6,154 from ENIGMA, 16,565 from BRCA Exchange), this represents a high rate of concordance, which speaks favorably to the overall quality of the public variation data. Furthermore, the amount of new variant information in BRCA Share, beyond that which is available in BRCA Exchange, illustrates the potential gains of expanding data-sharing efforts. BIC, Breast Cancer Information Core; ENIGMA, Evidence-based Network for the Interpretation of Germline Mutant Alleles; LOVD, Leiden Open Variation Database; UMD, Universal Mutation Database; VUS, variants of uncertain significance. (XLSX)

## Acknowledgments

### BRCA Challenge Authors:

Gaia Andreoletti, Dixie Baker, Steven Brenner, Matthew Brush, Sandrine Caputo, Laurent Castera, Fiona Cunningham, Miguel de la Hoya, Mark Diekhans, Jill Dolinsky, Selina Dwight, Diana Eccles, Bingjian Feng, Marc Fiume, Paul Flicek, Pascale Gaudet, Encarna Gomez Garcia, Melissa Haendel, Max Haeussler, Eric Hahnen, Claude Houdayer, Sarah Hunt, Paul James, Matt Lebo, Jennifer Lee, Jordan Lerner-Ellis, Mike Lin, Steve Lincoln, Adriana Malheiro, Arjen Mesenkamp, Alvaro Monteiro, E. Natzijl-Visser, Joanne Ngeow, Kathryn North, Helen Parkinson, Justin Paschall, George Patrinos, Bette Phimister, Paolo Radice, Irene Rainville, Matthew Rasmussen, George Riley, Etienne Rouleau, Rita Schmutzler, Kent Shefchek, Heidi Sofia, Melissa Southey, J. Stuart, Joseph Thomas, Amanda Toland, Rebecca Truty, Clare Turnbull, Dominique Vaur, Maikke P.G. Vreeswijk, Logan Walker, Michael Walsh, Barbara Wappenschmidt, J. Weitzel, Matt Wright, Vadim Zalunin, Alexander Zaranek, Daniel Zerbino, Alicia Zhou, Jingchun Zhou, Justin Zook

The authors Melissa S. Cline, Benedict Paten, David Goldgar, Sean Tavtigian, Gunnar Rättsch, and Amanda B. Spurdle are GA4GH Driver Project Champions for the BRCA Challenge.

## References

1. Burn J, Watson M (2016) The Human Variome Project. *Hum Mutat* 37: 505–7. <https://doi.org/10.1002/humu.22986> PMID: 26987309
2. Global Alliance for Genomic & Health. GA4GH Connect: A 5-Year Strategic Plan. [cited Mar 20 2018]. In: GA4GH Connect. Available from: <https://www.ga4gh.org/docs/GA4GH-Connect-A-5-year-Strategic-Plan.pdf>.
3. Abkevich V, Zharkikh A, Deffenbaugh AM, Frank D, Chen Y, Shattuck D, et al. (2004) Analysis of missense variation in human BRCA1 in the context of interspecific sequence variation. *J Med Genet* 41: 492–507. <https://doi.org/10.1136/jmg.2003.015867> PMID: 15235020
4. Easton DF, Deffenbaugh AM, Pruss D, Frye C, Wenstrup RJ, Allen-Brady K, et al. (2007) A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes. *Am J Hum Genet* 81: 873–83. <https://doi.org/10.1086/521032> PMID: 17924331
5. Spurdle AB, Healey S, Devereau A, Hogervorst FB, Monteiro AN, Nathanson KL, et al. (2012) ENIGMA—evidence-based network for the interpretation of germline mutant alleles: an international initiative to evaluate risk and clinical significance associated with sequence variation in BRCA1 and BRCA2 genes. *Hum Mutat* 33: 2–7. <https://doi.org/10.1002/humu.21628> PMID: 21990146
6. Meeks HD, Song H, Michailidou K, Bolla MK, Dennis J, Wang Q, et al. (2016) BRCA2 Polymorphic Stop Codon K3326X and the Risk of Breast, Prostate, and Ovarian Cancers. *J Natl Cancer Inst* 108.
7. Broca P. *Traité des tumeurs*. Paris: P. Asselin; 1866–1869.
8. Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, et al. (1990) Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* 250: 1684–9. PMID: 2270482
9. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, et al. (1994) A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 266: 66–71. PMID: 7545954

10. Tavtigian SV, Simard J, Rommens J, Couch F, Shattuck-Eidens D, Neuhausen S, et al. (1996) The complete BRCA2 gene and mutations in chromosome 13q-linked kindreds. *Nat Genet* 12: 333–7. <https://doi.org/10.1038/ng0396-333> PMID: 8589730
11. Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, et al. (1995) Identification of the breast cancer susceptibility gene BRCA2. *Nature* 378: 789–92. <https://doi.org/10.1038/378789a0> PMID: 8524414
12. Wooster R, Neuhausen SL, Mangion J, Quirk Y, Ford D, Collins N, et al. (1994) Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science* 265: 2088–90. PMID: 8091231
13. Antoniou A, Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL, et al. (2003) Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. *Am J Hum Genet* 72: 1117–30. <https://doi.org/10.1086/375033> PMID: 12677558
14. Kuchenbaecker KB, Hopper JL, Barnes DR, Phillips KA, Mooij TM, Roos-Blom MJ, et al. (2017) Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. *JAMA* 317: 2402–16. <https://doi.org/10.1001/jama.2017.7112> PMID: 28632866
15. Lindor NM, Guidugli L, Wang X, Vallee MP, Monteiro AN, Tavtigian S, et al. (2012) A review of a multifactorial probability-based model for classification of BRCA1 and BRCA2 variants of uncertain significance (VUS). *Hum Mutat* 33: 8–21. <https://doi.org/10.1002/humu.21627> PMID: 21990134
16. Vallee MP, Franczy TC, Judkins MK, Babikyan D, Lesueur F, Gammon A, et al. (2012) Classification of missense substitutions in the BRCA genes: a database dedicated to Ex-UVs. *Hum Mutat* 33: 22–8. <https://doi.org/10.1002/humu.21629> PMID: 21990165
17. Tavtigian SV, Byrnes GB, Goldgar DE, Thomas A (2008) Classification of rare missense substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications. *Hum Mutat* 29: 1342–54. <https://doi.org/10.1002/humu.20896> PMID: 18951461
18. Kurian AW, Li Y, Hamilton AS, Ward KC, Hawley ST, Morrow M, et al. (2017) Gaps in Incorporating Germline Genetic Testing Into Treatment Decision-Making for Early-Stage Breast Cancer. *J Clin Oncol* 35: 2232–9. <https://doi.org/10.1200/JCO.2016.71.6480> PMID: 28402748
19. Solomon I, Harrington E, Hooker G, Erby L, Axilbund J, Hampel H, et al. (2017) Lynch Syndrome Limbo: Patient Understanding of Variants of Uncertain Significance. *J Genet Couns* 26: 866–77. <https://doi.org/10.1007/s10897-017-0066-y> PMID: 28127677
20. Welsh JL, Hoskin TL, Day CN, Thomas AS, Cogswell JA, Couch FJ, et al. (2017) Clinical Decision-Making in Patients with Variant of Uncertain Significance in BRCA1 or BRCA2 Genes. *Ann Surg Oncol* 24: 3067–72. <https://doi.org/10.1245/s10434-017-5959-3> PMID: 28766224
21. Arbustini E, Sgarella A, Ferrari A, Grasso D, Cassani C, Lucioni M, et al. (2016) RE: BRCA2 Polymorphic Stop Codon K3326X and the Risk of Breast, Prostate, and Ovarian Cancers. *J Natl Cancer Inst* 108.
22. Moghadasi S, Meeks HD, Vreeswijk MP, Janssen LA, Borg A, Ehrencrona H, et al. (2018) The BRCA1 c. 5096G>A p.Arg1699Gln (R1699Q) intermediate risk variant: breast and ovarian cancer risk estimation and recommendations for clinical management from the ENIGMA consortium. *J Med Genet* 55: 15–20. <https://doi.org/10.1136/jmedgenet-2017-104560> PMID: 28490613
23. Shimelis H, Mesman RLS, Von Nicolai C, Ehlen A, Guidugli L, Martin C, et al. (2017) BRCA2 Hypomorphic Missense Variants Confer Moderate Risks of Breast Cancer. *Cancer Res* 77: 2789–99. <https://doi.org/10.1158/0008-5472.CAN-16-2568> PMID: 28283652
24. Spurdle AB, Whaley PJ, Thompson B, Feng B, Healey S, Brown MA, et al. (2012) BRCA1 R1699Q variant displaying ambiguous functional abrogation confers intermediate breast and ovarian cancer risk. *J Med Genet* 49: 525–32. <https://doi.org/10.1136/jmedgenet-2012-101037> PMID: 22889855
25. Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, et al. (2008) Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum Mutat* 29: 1282–91. <https://doi.org/10.1002/humu.20880> PMID: 18951446
26. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 17: 405–24. <https://doi.org/10.1038/gim.2015.30> PMID: 25741868
27. Lebo MS, Zakoor KR, Chun K, Speevak MD, Wayne JS, McCready E, et al. (2017) Data sharing as a national quality improvement program: reporting on BRCA1 and BRCA2 variant-interpretation comparisons through the Canadian Open Genetics Repository (COGR). *Genet Med*.
28. Chenevix-Trench G, Milne RL, Antoniou AC, Couch FJ, Easton DF, Goldgar DE, et al. (2007) An international initiative to identify genetic modifiers of cancer risk in BRCA1 and BRCA2 mutation carriers: the Consortium of Investigators of Modifiers of BRCA1 and BRCA2 (CIMBA). *Breast Cancer Res* 9: 104. <https://doi.org/10.1186/bcr1670> PMID: 17466083

29. Rebbeck TR, Friebel TM, Friedman E, Hamann U, Huo D, Kwong A, et al. (2018) Mutational spectrum in a worldwide study of 29,700 families with BRCA1 or BRCA2 mutations. *Hum Mutat* 39: 593–620. <https://doi.org/10.1002/humu.23406> PMID: 29446198
30. Szabo C, Masiello A, Ryan JF, Brody LC (2000) The breast cancer information core: database design, structure, and scope. *Hum Mutat* 16: 123–31. [https://doi.org/10.1002/1098-1004\(200008\)16:2<123::AID-HUMU4>3.0.CO;2-Y](https://doi.org/10.1002/1098-1004(200008)16:2<123::AID-HUMU4>3.0.CO;2-Y) PMID: 10923033
31. Clarke L, Zheng-Bradley X, Smith R, Kulesha E, Xiao C, Toneva I, et al. (2012) The 1000 Genomes Project: data management and community access. *Nat Methods* 9: 459–62. <https://doi.org/10.1038/nmeth.1974> PMID: 22543379
32. Vallee MP, Di Sera TL, Nix DA, Paquette AM, Parsons MT, Bell R, et al. (2016) Adding In Silico Assessment of Potential Splice Aberration to the Integrated Evaluation of BRCA Gene Unclassified Variants. *Hum Mutat* 37: 627–39. <https://doi.org/10.1002/humu.22973> PMID: 26913838
33. Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT (2011) LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat* 32: 557–63. <https://doi.org/10.1002/humu.21438> PMID: 21520333
34. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, et al. (2012) Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 337: 64–9. <https://doi.org/10.1126/science.1219240> PMID: 22604720
35. Woods NT, Baskin R, Golubeva V, Jhuraney A, De-Gregoriis G, Vaclova T, et al. (2016) Functional assays provide a robust tool for the clinical annotation of genetic variants of uncertain significance. *NPJ Genom Med* 1.
36. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. (2014) ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* 42: D980–5. <https://doi.org/10.1093/nar/gkt1113> PMID: 24234437
37. Karczewski KJ, Weisburd B, Thomas B, Solomonson M, Ruderfer DM, Kavanagh D, et al. (2017) The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic Acids Res* 45: D840–D5. <https://doi.org/10.1093/nar/gkw971> PMID: 27899611
38. Hart RK, Rico R, Hare E, Garcia J, Westbrook J, Fusaro VA (2015) A Python package for parsing, validating, mapping and formatting sequence variants using HGVS nomenclature. *Bioinformatics* 31: 268–70. <https://doi.org/10.1093/bioinformatics/btu630> PMID: 25273102
39. Zhao H, Sun Z, Wang J, Huang H, Kocher JP, Wang L (2014) CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* 30: 1006–7. <https://doi.org/10.1093/bioinformatics/btt730> PMID: 24351709
40. Beroud C, Letovsky SI, Braastad CD, Caputo SM, Beaudoux O, Bignon YJ, et al. (2016) BRCA Share: A Collection of Clinical BRCA Gene Variants. *Hum Mutat* 37: 1318–28. <https://doi.org/10.1002/humu.23113> PMID: 27633797
41. Thorogood A, Cook-Deegan R, Knoppers BM (2017) Public variant databases: liability? *Genet Med* 19: 838–41. <https://doi.org/10.1038/gim.2016.189> PMID: 27977006
42. Domchek SM, Bradbury A, Garber JE, Offit K, Robson ME (2013) Multiplex genetic testing for cancer susceptibility: out on the high wire without a net? *J Clin Oncol* 31: 1267–70. <https://doi.org/10.1200/JCO.2012.46.9403> PMID: 23460708
43. Eliade M, Skrzypski J, Baurand A, Jacquot C, Bertolone G, Loustalot C, et al. (2017) The transfer of multigene panel testing for hereditary breast and ovarian cancer to healthcare: What are the implications for the management of patients and families? *Oncotarget* 8: 1957–71. <https://doi.org/10.18632/oncotarget.12699> PMID: 27779110
44. Graffeo R, Livraghi L, Pagani O, Goldhirsch A, Partridge AH, Garber JE (2016) Time to incorporate germline multigene panel testing into breast and ovarian cancer patient care. *Breast Cancer Res Treat* 160: 393–410. <https://doi.org/10.1007/s10549-016-4003-9> PMID: 27734215
45. Offit K (2017) Multigene Testing for Hereditary Cancer: When, Why, and How. *J Natl Compr Canc Netw* 15: 741–3. PMID: 28515260
46. Rahman N (2014) Realizing the promise of cancer predisposition genes. *Nature* 505: 302–8. <https://doi.org/10.1038/nature12981> PMID: 24429628