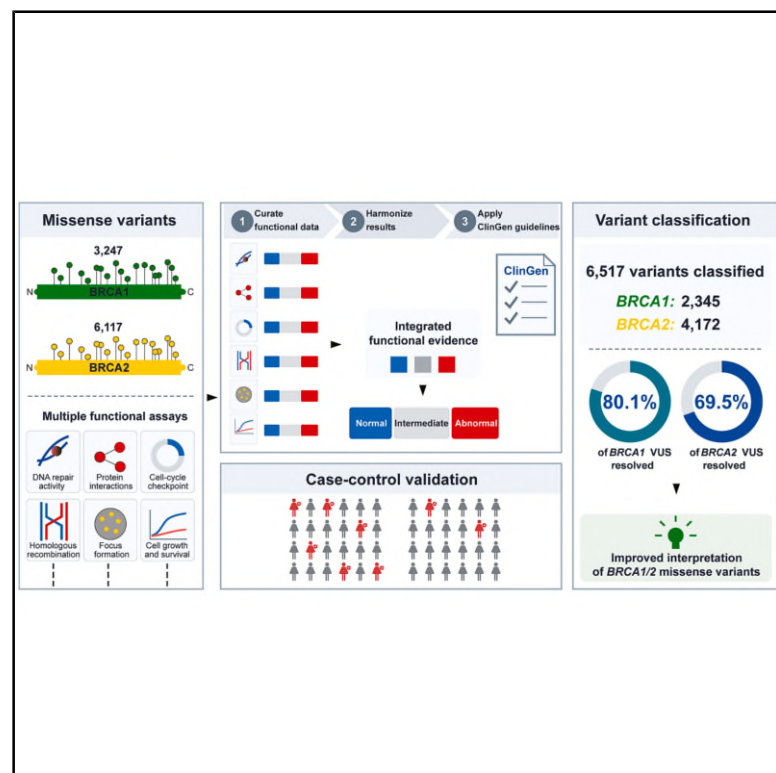


# Large-scale integration of functional assay data for the resolution of germline *BRCA1* and *BRCA2* variants of uncertain significance

## Graphical abstract



## Authors

Paulo C.M. Lyra, Jr., Samuel Brito, Michael Parsons, ..., Fergus J. Couch, Alvaro N.A. Monteiro, the ENIGMA Consortium

## Correspondence

[alvaro.monteiro@moffitt.org](mailto:alvaro.monteiro@moffitt.org)

**By harmonizing functional data for 9,424 *BRCA1* and *BRCA2* missense variants and applying ClinGen guidelines, Lyra, Jr. et al. establish an integrated framework whose pathogenicity assignments track with breast cancer risk. Functional data combined with additional evidence enabled classification of 6,517 missense variants.**

# Large-scale integration of functional assay data for the resolution of germline *BRCA1* and *BRCA2* variants of uncertain significance

Paulo C.M. Lyra, Jr.,<sup>1</sup> Samuel Brito,<sup>1</sup> Michael Parsons,<sup>2</sup> Chunling Hu,<sup>3</sup> Thales C. Nepomuceno,<sup>1</sup> Nam Phuong Nguyen,<sup>1</sup> Lucas Dalcolmo,<sup>1</sup> Marcy E. Richardson,<sup>4</sup> Geise de Oliveira,<sup>5</sup> João da Silva,<sup>5</sup> Laura Caleca,<sup>6</sup> Huaizhi Huang,<sup>3</sup> Sounak Sahu,<sup>7,25</sup> Muthiah Bose,<sup>8</sup> Maria Rossing,<sup>8,9</sup> Phoebe Dace,<sup>10</sup> Gregory M. Findlay,<sup>10</sup> Rehan Villani,<sup>2</sup> Aditi Hazra,<sup>11</sup> Alexandra Martins,<sup>12</sup> Sandrine M. Caputo,<sup>13</sup> Gael A. Millot,<sup>14</sup> Joanne Ngeow Yuen Yie,<sup>15,16</sup> Marcelo A. Carvalho,<sup>17</sup> Melissa S. Cline,<sup>18</sup> Paolo Radice,<sup>6</sup> Romy L.S. Mesman,<sup>19</sup> Valentina Zampiga,<sup>20</sup> Vijai Joseph,<sup>21</sup> Katherine L. Nathanson,<sup>22</sup> Kyriaki Michailidou,<sup>23</sup> Edwin S. Iversen, Jr.,<sup>24</sup> Shyam Sharan,<sup>7</sup> Amanda B. Spurdle,<sup>2</sup> Maaïke P.G. Vreeswijk,<sup>18</sup> Fergus J. Couch,<sup>3</sup> Alvaro N.A. Monteiro,<sup>1,\*</sup> and the ENIGMA Consortium

## Summary

Carriers of *BRCA1* and *BRCA2* pathogenic variants are at a significantly elevated risk for several cancers. Widespread germline genetic testing has led to a significant increase in the detection of variants of uncertain significance (VUSs) in these genes. To assess variant pathogenicity, we curated, harmonized, and integrated functional data for 3,247 *BRCA1* and 6,177 *BRCA2* missense variants derived from *BRCA1/2* functional assays. Functional results were harmonized across studies by converting data into ordinal categorical variables. We applied ClinGen variant interpretation guidelines for each assay to assign evidence for or against pathogenicity. Final evidence criteria for each variant were obtained through the integration of results from different assays, and large case-control studies confirmed that benign and pathogenic variants categorized using the integrated functional framework are associated with low and high risk for breast cancer, respectively. Next, we integrated additional sources of data, leading to resolution of 80.1% of *BRCA1* VUSs and 69.5% of *BRCA2* VUSs and classification of 6,517 (2,345 in *BRCA1* and 4,172 in *BRCA2*) missense variants.

## Introduction

The tumor suppressor genes *BRCA1* and *BRCA2* (*BRCA1/2*) are associated with susceptibility to several types of cancer, including breast, ovarian, prostate, and pancreatic cancers. Germline pathogenic variants (PVs) in these genes that result in a non-existent, unstable, or non-functional protein confer an elevated risk of cancer in comparison with the general population.<sup>1–4</sup> Importantly, it is estimated that a significant fraction of unselected cancer cases in the tissue susceptibility spectrum of *BRCA1/2* is attributable to germline PVs in these genes.<sup>5–7</sup>

In addition to the importance of variant detection for risk stratification, PVs in these genes also predict response

to PARP1 inhibitors, a class of drugs approved for treatment of breast, ovarian, prostate, and pancreatic cancers.<sup>8–16</sup> Given its importance for preventative and therapeutic clinical management, the National Comprehensive Cancer Network guidelines recommend genetic testing including *BRCA1/2* for individuals with personal and/or family history of breast, ovarian, prostate, and pancreatic cancers; cancer surveillance for individuals with *BRCA1/2* PVs; and testing of family members. The presence of germline and somatic PVs in *BRCA1/2* is used to guide therapeutic selection.<sup>17</sup>

The increased uptake of clinical genetic (germline and somatic) testing and results from population-scale sequencing projects has contributed to the identification

<sup>1</sup>Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA; <sup>2</sup>Population Health Program, QIMR Berghofer, Brisbane, QLD 4006, Australia; <sup>3</sup>Mayo Clinic, Rochester, MN, USA; <sup>4</sup>Ambry Genetics, Aliso Viejo, CA, USA; <sup>5</sup>Centro Universitario FAEMA, Ariquemes, Rondonia, Brazil; <sup>6</sup>Department of Experimental Oncology, Fondazione IRCCS Istituto Nazionale dei Tumori, Milano, Italy; <sup>7</sup>Center for Cancer Research, National Cancer Institute, Frederick, MD, USA; <sup>8</sup>Center for Genomic Medicine, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark; <sup>9</sup>Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; <sup>10</sup>The Francis Crick Institute, London, UK; <sup>11</sup>Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; <sup>12</sup>University Rouen Normandie Inserm U1245, 76000 Rouen, France; <sup>13</sup>Department of Genetics, Institut Curie, Paris Sciences and Lettres (PSL) University Paris, France; <sup>14</sup>Institut Pasteur, Université Paris Cité, Paris 75015, France; <sup>15</sup>Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore, Singapore; <sup>16</sup>Cancer Genetics Service, National Cancer Centre Singapore, Singapore, Singapore; <sup>17</sup>Instituto Federal do Rio de Janeiro, Rio de Janeiro 20270-021, Brazil; <sup>18</sup>UC Santa Cruz Genomic Institute, Santa Cruz, CA, USA; <sup>19</sup>Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands; <sup>20</sup>Biosciences Laboratory, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) Dino Amadori, 47014 Meldola, Italy; <sup>21</sup>Memorial Sloan Kettering Cancer Center, New York, New York, NY, USA; <sup>22</sup>Division of Translational Medicine and Human Genetics, Department of Medicine, Abramson Cancer Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; <sup>23</sup>Biostatistics Unit, The Cyprus Institute of Neurology & Genetics, Nicosia, Cyprus; <sup>24</sup>Department of Statistical Science, Duke University, Durham, NC, USA

<sup>25</sup>Present address: Department of Cell Biology, NYU Grossman School of Medicine, New York, NY, USA

\*Correspondence: [alvaro.monteiro@moffitt.org](mailto:alvaro.monteiro@moffitt.org)

<https://doi.org/10.1016/j.ajhg.2026.09.013>

© 2026 The Authors. Published by Elsevier Inc. on behalf of American Society of Human Genetics.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

of over 35,557 unique variants in *BRCA1* and 36,774 in *BRCA2*.<sup>18–20</sup> However, only a small fraction of those variants have been formally classified, posing a significant challenge for the implementation of precision medicine.<sup>21</sup> Nonsense and frameshift changes are typically classified as likely pathogenic/pathogenic (LP/P), whereas synonymous changes that do not affect RNA splicing are classified as likely benign/benign (LB/B). However, the pathogenicity of missense variants, which constitute the largest fraction of reported variants of uncertain significance (VUSs), cannot be inferred from changes in the genetic code.<sup>21</sup> For those variants, classification depends on a combination of population allelic frequency, bioinformatic impact prediction, and clinical information such as segregation analysis or data from case-control studies. However, the frequency of each individual PV in the general population is typically lower than 1/10,000, and segregation analysis is often unavailable. Thus, it is expected that genetic and clinical data will not be available to identify and classify most candidate PVs. Lack of information about missense VUSs constitutes a barrier to the interpretation and, thus, the use of genetic testing results in clinical management. This realization has led to an explosion in functional assays, including large-scale multiplex assays of variant effects (MAVE) experiments to assess the impact on protein function of individual variants.<sup>22–28</sup>

The most recent classification framework, developed by the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP), has incorporated the use of functional data for the classification of variants.<sup>29,30</sup> Recently, our group has shown that integration of functional results from multiple assays is a powerful approach to aid in the classification of *BRCA1* variants.<sup>31</sup> To address continued gaps in variant classification, we have greatly expanded the *BRCA1* dataset and created a new *BRCA2* dataset. We integrate heterogeneous BRCA functional assays with calibrated, clinically validated classification. Here, we collate and harmonize results from 185 individual instances of functional assays for 3,247 missense variants in *BRCA1* and 147 individual instances of functional assays for 6,177 missense variants in *BRCA2* to construct the databases Functional Assay Integration (FYI)-*BRCA1* v.2.0 and FYI-*BRCA2* v.1.0, respectively.

## Material and methods

### General definitions, data extraction, and literature curation

Missense variants are defined here as single-nucleotide changes that result in the substitution of one reference amino acid residue using GenBank accession NCBI reference sequences: *BRCA1*, GenBank: NM\_007294.4 and NP\_009225.1; and *BRCA2*, GenBank: NM\_000059.4 and NP\_000050.3. The determination of functional impact described in this study refers exclusively to effects on protein function and not on RNA splicing, unless stated otherwise.

For simplicity, missense variants are here referred to by their unique amino acid substitution and not by their underlying nucleotide change. Specific instances of a functional assay are called “tracks.” When a publication presented results including data from previous publications to perform a combined analysis, only the latest paper with the combined analysis was used. FYI-*BRCA1* v.2.0 expands the previously published dataset<sup>31</sup> with 17 new studies (Table S1, see the “comments” tab for track descriptions and details). To assemble FYI-*BRCA2* v.1.0, we collated all published functional data for *BRCA2* missense variants from 33 studies (Table S2).

### Universe of missense variants

The *BRCA1* coding region has 11,009 unique possible missense amino acid substitutions resulting from single-nucleotide changes. In the *BRCA2* coding region, there are 20,169 unique missense substitutions. In track 7 (Tables S1 and S2), variants are denoted as “1” if observed (and “0” if not) in BRCA Exchange (v.61, November 9, 2023),<sup>18,19</sup> which include variants observed in clinical testing (ClinVar) or in exome/genome sequencing of large cohorts (gnomAD).<sup>20</sup>

### Harmonization

To harmonize data, we followed the authors’ conclusion of whether a given variant had a significant impact on function. We assigned 0 to a variant that did not lead to a significant impact on the function being tested (functionally normal), 1 to those that showed intermediate activity, and 2 to a variant that had a significant impact on function. We hypothesize that harmonized functional scores 0 and 2 are associated with LB/B and LP/P variants, respectively. If variants representing the same amino acid residue coded by different underlying nucleotides for the codon presented conflicting results within a study, we assigned the intermediate activity 1 code. Therefore, the 1 code includes hypomorphic variants that may represent reduced-penetrance PVs (RPPVs) and variants whose results were indeterminate due to conflicting results within an individual assay. Both groups are assigned a final functional code label of “indeterminate” in Tables S12 and S13 (see “final functional code” column). To distinguish between variants with functional observations indicative of hypomorphism (and presumably associated with RPPVs) and variants with conflicting functional results, we assigned functional evidence in favor of “hypomorph” for the former and indeterminate for the latter in Tables S12 and S13 (see “functional evidence in favor of” column). For details on how each assay’s results were transformed, see Tables S3 and S4.

### Reference panel

To assess the specificity and sensitivity of each track, we used a stringent reference panel that combines data from the ENIGMA consortium and ClinVar.<sup>32,33</sup> The ENIGMA reference panels comprise 297 *BRCA1* and 277 *BRCA2* missense variants assigned to IARC classes by a multifactorial model.<sup>32</sup> We collapsed classes 1 with 2 and 4 with 5 and transformed them into binary categories: 0 to a variant that was B (IARC class 1) or LB (IARC class 2) and 2 to a variant that was P (IARC class 5) or LP (IARC class 4). The ENIGMA panel includes variants classified using a rigorous multifactorial statistical model<sup>34,35</sup> but may suffer from bias toward variants with stronger effects on risk. To mitigate this bias and expand the panel, we selected germline missense variants with ClinVar classifications of LB/B or LP/P. Variants

with a conflict between assertions or from a single submitter who did not provide supporting information were excluded. The ClinVar panels were merged with the ENIGMA panels, resulting in non-redundant panels of 572 *BRCA1* and 396 *BRCA2* variants (Tables S5 and S6). Only variants for which classification was achieved without consideration of functional data were included. For ENIGMA panel variants, this was accomplished by excluding variants that had a functional category assigned in Table S1 in Parsons et al.<sup>32</sup> For ClinVar variants, we removed variants for which functional data were available as indicated (i.e., “Functional data are available for this variant”) in ClinVar records. We excluded (1) variants with effects on splicing as the basis for classification, (2) start-loss variants, and (3) variants in the *BRCA1* p.Cys1787Ser/p.Gly1788Asp haplotype (results for the single p.Cys1787Ser substitution are reported in Tables S1–S20). After exclusions, the final non-redundant ENIGMA+ClinVar [E+C] reference panels comprised 548 *BRCA1* and 385 *BRCA2* reference missense variants (Tables S5 and S6).

## Odds of pathogenicity calculations and credentialing

We estimated the odds of pathogenicity (OddsPath) following the recommendations from the Clinical Genome (ClinGen) Resource Sequence Variant Interpretation Working Group<sup>30</sup> with one modification. To avoid posterior probabilities of 0 or 1, or infinite odds, we added 0.5 to each set for the purpose of OddsPath calculation. For each track, the OddsPath for functionally impaired variants and for functionally normal variants were calculated and used to determine the corresponding level of evidence strength equivalence according to the Bayesian adaptation of the ACMG/AMP variant interpretation guidelines (Tables S7 and S8).<sup>30,36</sup>

Next, each track was credentialled to assign an associated level of evidence strength to each variant tested according to the thresholds of OddsPath.<sup>30,36</sup> The labels PS3\_variable (PS3, PS3\_moderate, or PS3\_supporting) and BS3\_variable (BS3, BS3\_moderate, or BS3\_supporting) were assigned to variants scored as having significant functional impact or with no functional impact, respectively. For example, a track credentialled to assign PS3\_moderate and BS3\_supporting codes assigned a PS3\_moderate code to all variants tested that scored as having a significant impact on function (score = 2) and a BS3\_supporting code to all variants that scored as having no functional impact (score = 0).

## Integration of ACMG/AMP functional evidence label

Variants were assigned an ACMG/AMP evidence strength label based on the integrated functional data according to the flow chart in Figure S1. Variant functional scores (0, 1, or 2) were computed, and variants were labeled “concordant” when all tracks’ scores were the same or “discordant” when at least one track’s score conflicted with another track’s score. Variants tested only once were, by definition, concordant.

To obtain the final evidence labels, variants tested multiple times that received concordant labels were assigned the strongest label obtained as final evidence. For variants with discordant labels, we previously proposed that ratios of 3:1 or greater and 1:3 or smaller ( $1.5 \leq \log_2 [\text{ratio benign/path}] \leq -1.5$ ) constitute preponderance of evidence and are assigned the strongest evidence label among the tracks providing the results.<sup>31</sup> Variants that did not reach these thresholds or for which the highest count

comes from hypomorph or indeterminate labels remain unassigned and are denoted as indeterminate (Figure S1).

Because RPPVs are not accounted for in the current ACMG/AMP framework, tracks were not credentialled to assign “intermediate” labels to hypomorphic variants. For the purpose of identifying candidate RPPVs for future analysis, variants with concordant (or preponderance of evidence of) 1 scores were assigned a provisional label of hypomorph to indicate evidence for hypomorphism, but the integrated ACMG evidence criteria label remained as indeterminate.

## Performance assessment of the integrative approach

To estimate the performance of our approach, we conducted a stratified k-fold cross-validation ( $k = 10$  for *BRCA1* and  $k = 5$  for *BRCA2*), preserving the relative proportions of pathogenic and benign reference variants across folds. In each fold, 1/10 (*BRCA1*) or 1/5 (*BRCA2*) of reference variants were randomly removed before track credentialing (out-of-sample set [OOS]). Track credentialing was then obtained using the remaining 9/10 (*BRCA1*) or 4/5 (*BRCA2*) reference variants. Reference variants in the OOS were then classified according to the new credentialing: BS3\_variable labels were considered functional evidence for being benign and PS3\_variable labels were considered functional evidence of being pathogenic. We then calculated specificity, sensitivity, and 95% confidence intervals (CIs) to summarize the conditional probabilities, and the Matthews correlation coefficient (MCC) to summarize the joint distribution of true and false positives/negatives. “No call” (when a variant has discordant scores that are not resolved by preponderance of evidence) was excluded from specificity and sensitivity computation, and no call rates are reported separately for each fold (Table S11).

## Assessment of amino acid substitutions

For analysis of amino acid substitutions, amino acid residues were assigned to the following classes: hydrophobic (Val, Ile, Leu, Ala, Trp, Tyr, Phe, and Met), polar negative (Ser and Thr), polar positive (Asn and Gln), charged positive (Lys, Arg, and His), charged negative (Asp and Glu), and special cases (Pro, Cys, and Gly). Pairwise comparisons between groups of the substitutions and their associated Grantham distance<sup>37</sup> or side-chain volume difference were performed using Wilcoxon rank-sum tests.

## ACMG/AMP-compliant point system

In this system, supporting strength is assigned  $-1$  (BS3\_supporting) or  $+1$  (PS3\_supporting), moderate strength is assigned  $-2$  (BS3\_moderate) or  $+2$  (PS3\_moderate), and strong evidence is assigned  $-4$  (BS3) or  $+4$  (PS3). To prevent classification into a clinical pathogenic category ( $\geq 6$  points, LP) using a single source of evidence, totals from combinations of multiple assays to arrive at the ACMG/AMP functional points were capped at  $-4$  or  $+4$ . Variants with a final integrated classification of LB/B ( $\leq -2$ ) with a SpliceAI score  $> 0.2$  were flagged.

We used EVE (evolutionary model of variant effect) as an *in silico* predictor to provide computational evidence for the classification of variants.<sup>38</sup> EVE is an unsupervised deep generative model that estimates the effects of amino acid substitutions from evolutionary sequence variation without using clinical variant labels during training. Precomputed predictions were used without retraining or calibration against the *BRCA*

reference variants. Prediction archives were retrieved through the EVE application programming interface using a dedicated artifact-generation script available in the project repository ([https://github.com/paulocilasjr/BRCA\\_integration-code/blob/main/dataset/eve/](https://github.com/paulocilasjr/BRCA_integration-code/blob/main/dataset/eve/)). For each missense substitution, we used the EVE ASM score and the corresponding author-defined classification at 75% retained coverage. EVE scores range from 0 to 1, with higher values indicating greater predicted pathogenicity. EVE classifications were incorporated into the PP3/BP4 component of the point-based ACMG/AMP framework as  $-1$ ,  $0$ , or  $+1$  points to variants with benign, uncertain, and pathogenic predictions, respectively. Variants without an EVE prediction were assigned  $0$  points. Predictions were available for 99.11% of *BRCA1* and 90.68% of *BRCA2* missense variants included in this study.

### Case-control studies

Breast cancer risks associated with individual *BRCA1* and *BRCA2* missense variants and with aggregated variants assigned final functional labels of (1) PS3\_variable (PS3\_supporting, PS3\_moderate, and PS3), (2) indeterminate, or (3) BS3\_variable labels were estimated by performing case-control comparisons. We conducted two case-control comparisons between affected individuals from the Ambry Genetics female breast cancer clinical cohort and (1) control subjects from the All of Us (AoU) Research Program<sup>39</sup> using logistic regression adjusted for age and ancestry and (2) control subjects from gnomAD v.4.1,<sup>20</sup> excluding samples from the UK Biobank (UKBB),<sup>40</sup> and one population-based study comparing affected individuals and control subjects from the CARRIERS<sup>2</sup> and BRIDGES<sup>4</sup> studies combined. Analyses other than Ambry Genetics versus AoU were done by Fisher's exact test with no adjustment for age and ancestry. For the Ambry Genetics breast cancer clinical cohort, breast cancer cases and associated clinical phenotypes were collected from individuals who underwent cancer genetic testing by Ambry Genetics from 2016 to 2021. Variants with an allele frequency of  $>0.001\%$  were excluded from the analyses. We excluded from the PS3 categories ENIGMA benign variants (*BRCA1*: p.Ile1275Val; *BRCA2*: p.Thr1354Met, p.Glu1593Asp, and p.Leu2396Phe) and *BRCA2*: p.Asn2622Ser, which is enriched in African American (AA) individuals. This study complies with all relevant ethical regulations and was approved by the Mayo Clinic Institutional Review Board (IRB 21-008216). The analysis of the Ambry Genetics clinical-testing cohort was deemed exempt from review by the Western IRB.

### Selection of variants found in individuals of African/AA ancestry

We used gnomAD v.4.1<sup>20</sup> to identify missense variants in *BRCA1* and *BRCA2* with ClinVar Germline classifications of "uncertain significance" or "conflicting classifications of pathogenicity." Only variants that had at least 1 allele count in AA individuals were retained.

### Visualization and data and code availability

Data for all tracks are included as [supplemental information](#) (Tables S1–S20). Visualization for Sankey diagrams was obtained through Sankeymatic (<https://sankeymatic.com/>). Datasets and codes for queries to generate summary statistics and variant calls are available through GitHub (<https://github.com/FunctionalAssayIntegration/BRCA1-BRCA2-Large-integration>).

## Results

### Composition of BRCA functional datasets

We curated and harmonized data to build two datasets of functional information: FYI-BRCA1 v.2.0 expands the previously published dataset with a total of 185 functional tracks and data on 3,247 *BRCA1* missense variants, of which 2,896 are VUSs. These variants represent 46.9% (2,592/5,526) of all reported missense variants and 29.5% (3,247/11,009) of all possible single-nucleotide missense changes (Figure 1A; Table S9). The second, FYI-BRCA2 v.1.0, contains functional data for 6,177 variants tested in 147 functional tracks, of which 6,001 are VUSs representing 30.3% of all possible single-nucleotide missense changes (Figure 1B; Table S9).

Coverage of variants in defined functional domains of *BRCA1* (i.e., RING finger and BRCT) and *BRCA2* (C-terminal DNA-binding domain), where most known pathogenic missense variants occur, is greater than 97% (Figure 1, insets). Furthermore, 16.7% of *BRCA1* and 20.7% of *BRCA2* possible missense variants have been assessed by more than one track (Table S9). Over 90% of the tracks can be considered low-to-mid throughput ( $<100$  variants tested) (Table S10), but high-throughput tracks contribute to 74.3% of *BRCA1* and 90% of *BRCA2* aggregate variant tests.

In summary, these datasets contain functional data from 22,489 test results (9,279 for *BRCA1* and 13,210 for *BRCA2*) for 3,247 missense variants in *BRCA1* and 6,177 in *BRCA2* (total = 9,424), tested in a wide variety of functional assays such as protein-protein binding; focus formation; protein expression, stability, folding, and localization; transcription activation; sensitivity; recombination/repair; proliferation; cell migration; chromosome/mitotic apparatus; cell viability; catalytic activity; RNA expression; and cell cycle checkpoint<sup>29,30</sup> and with very high coverage of variants in functionally important domains.

### Accuracy and credentialing of functional tracks

Using the reference panels, we determined the sensitivity and specificity of all functional tracks (Tables S7 and S8). For *BRCA1*, 18.9% (35/185) of tracks achieved specificity and sensitivity of 1.0, and 33.5% (62/185) achieved  $\geq 0.8$ . For *BRCA2*, 13.6% (20/147) of tracks achieved specificity and sensitivity of 1.0, and 19.7% (29/147) achieved  $\geq 0.8$ , although in some cases the CIs were wide (Figure S2).

For *BRCA1* variants, 42.7% (79/185) of tracks were credentialled to assign PS3\_variable evidence labels (PS3, PS3\_moderate, or PS3\_supporting) to variants with a functional impact; the remaining tracks either did not test reference PVs or had low ( $<0.48$ ) or indeterminate ( $0.48\text{--}2.1$ ) OddsPath. For variants that were functionally normal, 55.1% (102/185) of tracks were credentialled to assign BS3\_variable evidence labels (BS3, BS3\_moderate, or BS3\_supporting). For *BRCA2*, fewer tracks were credentialled to assign PS3\_variable (23.1%) and BS3\_variable (17.0%) evidence labels (Tables S7 and S8).

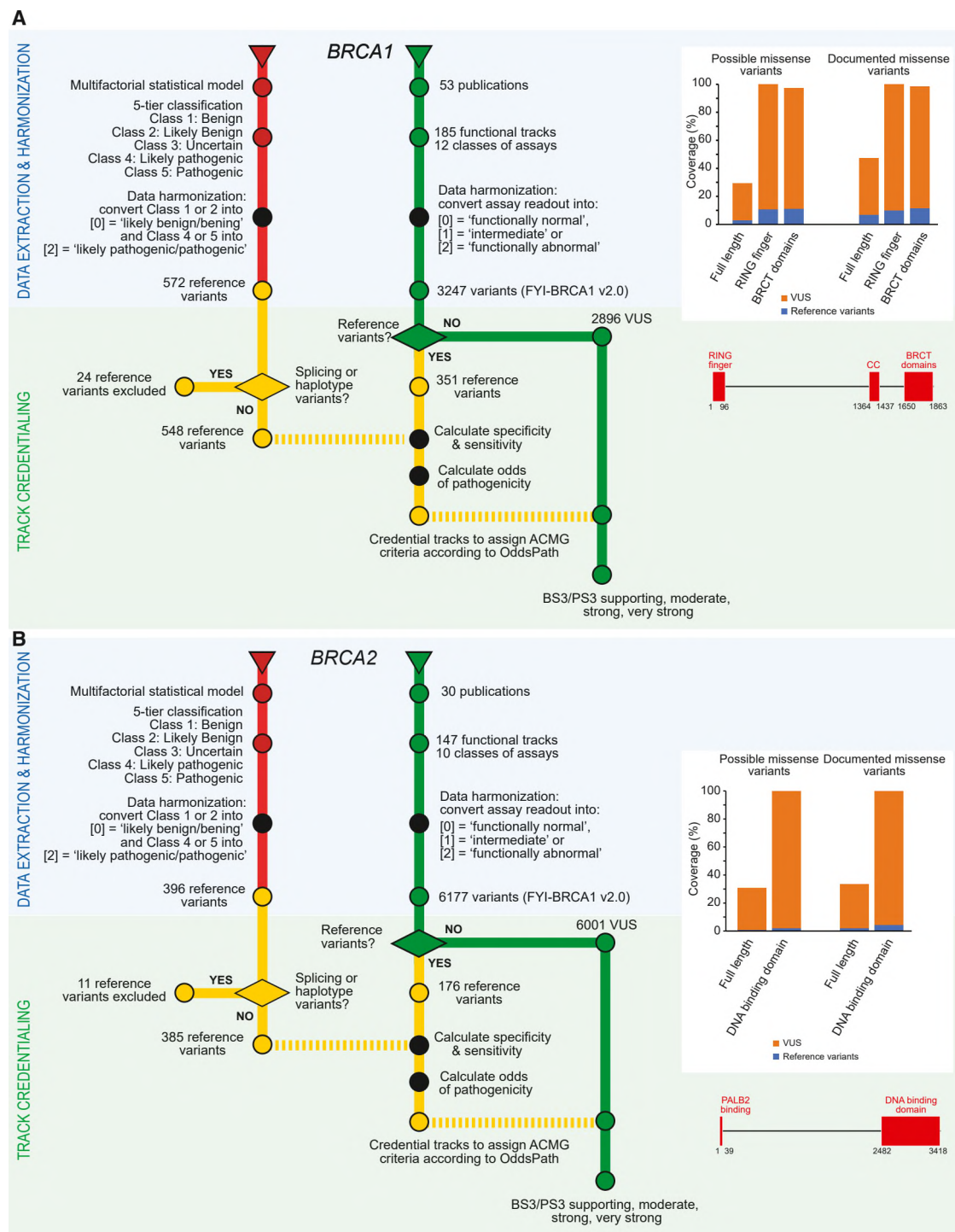


**Figure 1. Overview of study design for classification of *BRCA* missense variants**

“Subway chart” illustrating the processes of data collection, curation, and harmonization of functional data (green line); harmonization of reference variant classification (red line); and data analysis and evidence criteria assignment (yellow line) for *BRCA1* (A) and *BRCA2* (B) variants. Three stages (data extraction and harmonization, track credentialing, and final classification) are indicated by blue-, green-, and orange-shaded boxes. The percentage of coverage of possible and documented missense variants is shown for full length and for the *BRCA1* RING and BRCT functional domains and the *BRCA2* C-terminal DNA-binding domain (insets).

## Performance assessment of the integrative approach

To provide an initial estimate of the performance of our approach, we conducted a k-fold cross-validation in which

tracks were re-credentialled within each fold (Table S11). For *BRCA1* ( $k = 10$ ), the average sensitivity was 0.98 (0.71–1.00; 95% CI), the average specificity was 0.99 (0.82–1.00; 95% CI), and the MCC was 0.97. For *BRCA2* ( $k = 5$ ; due to a

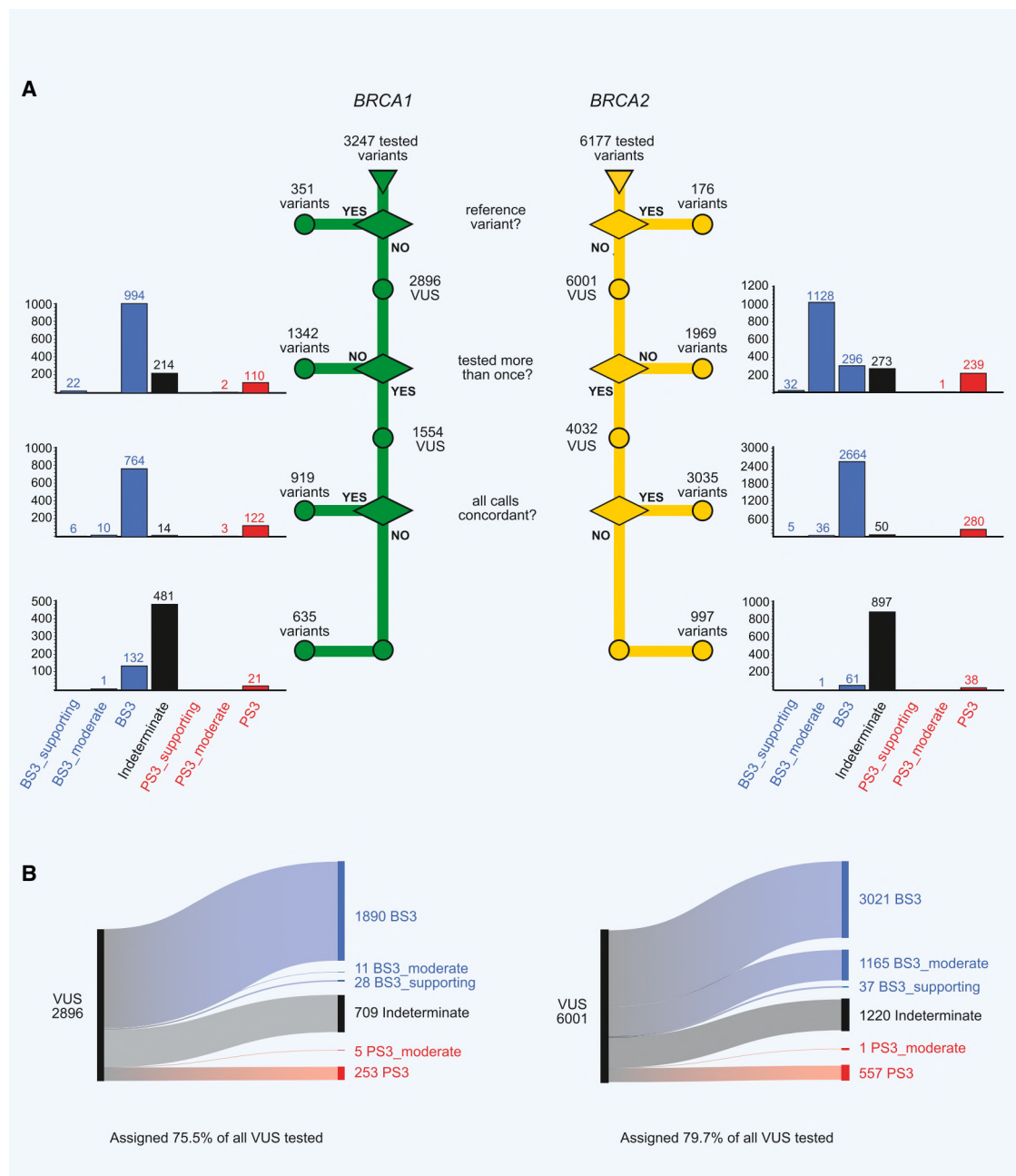


**Figure 2. Process of assigning functional code labels to variants**

(A) Subway chart illustrating the stages of evidence criteria assignment by the integration analysis to VUSs for *BRCA1* (green line) and *BRCA2* (yellow line).

(B) Sankey diagrams indicate how VUSs were resolved (assigned an ACMG evidence strength code) after applying the functional data integration. VUS\* denotes VUSs that were tested only in non-credentialed tracks.

smaller number of reference variants), the average sensitivity was 1 (0.33–1.00; 95% CI), the average specificity was 0.97 (0.83–0.99; 95% CI), and the MCC was 0.8. The integrative approach displayed a strong performance when classifying *BRCA1* and *BRCA2* missense variants according to the available functional information.

### Assignment of evidence strength labels to VUSs

From the set of *BRCA1* variants tested (FYI-*BRCA1* v2.0), we removed reference variants and identified 1,342 VUSs

that had been tested only once and assigned BS3\_variable labels to 75.7% of VUSs and PS3\_variable labels to 8.3% of VUSs (Table S12; Figure 2A). From FYI-*BRCA2* v.1.0, 1,969 VUSs were tested only once; we assigned BS3\_variable codes to 73.9% of VUSs, and 12.2% were assigned PS3\_variable codes (Table S13; Figure 2A).

Next, we identified 919 *BRCA1* and 3,035 *BRCA2* VUSs that had been tested more than once with concordant scores and assigned the strongest of all codes assigned to the variant, with 84.9% of *BRCA1* and 89.1% of *BRCA2*

variants assigned BS3\_variable codes and 13.6% of *BRCA1* and 10.9% of *BRCA2* variants assigned PS3\_variable codes (Figure 2A).

There were 635 *BRCA1* and 997 *BRCA2* VUSs with discordant results (Tables S12 and S13). For variants with discordant calls, we have previously proposed a preponderance-of-evidence principle in which ratios of 3:1 or greater and 1:3 or smaller ( $1.5 \leq \log_2 [\text{ratio benign/path}] \leq -1.5$ ) constitute preponderance of evidence. While 24.3% (*BRCA1*) and 9.6% (*BRCA2*) of VUSs were resolved by the preponderance-of-evidence rule, the majority of VUSs with discordant results remained unresolved. Interestingly, *BRCA* variants with discordant results were more likely to have had at least one hypomorph observation than all VUSs tested (*BRCA1*: 0.4 versus 0.11,  $X^2 = 257.916$ ,  $p < 0.00001$ ; *BRCA2*: 0.32 versus 0.1,  $X^2 = 440.1673$ ,  $p < 0.00001$ ).

At the end of the process, 75.5% of all *BRCA1* and 79.7% of all *BRCA2* VUSs tested received PS3\_variable and BS3\_variable ACMG codes according to the strength of functional evidence (Figure 2B).

### Reference variants with conflicting results

We examined the set of reference variants with functional labels conflicting with their classification to identify potential causes of discordant results or limitations of our approach. Of 140 PVs (IARC classes 4 and 5), only two (p.Arg71Thr and p.Gly1803Ala) were incorrectly assigned as benign (<1.5%). Variant p.Arg71Thr is pathogenic due to a splicing defect, and two tracks only evaluated the protein function from cDNA. To prevent this type of misclassification during the integration of functional data with other sources of information, variants with a final classification of LB/B with a SpliceAI<sup>41</sup> score > 0.2 were flagged (see below). In the case of p.Gly1803Ala, many of the tracks used in classification were based on structural integrity of the BRCT domain, and presumably, the variant does not substantially affect the structure despite being functionally compromised. Of 211 benign variants (IARC classes 1 and 2), only two (p.Ile1275Val and p.His1680Gln) were incorrectly assigned (<1%) as pathogenic. Notably, all four tracks (T153–T156) that contribute to the misclassification of p.Ile1275Val are from the same study (Table S12).

For *BRCA2*, 176 reference variants were evaluated. No reference PV (IARC classes 4 and 5) was incorrectly assigned, whereas seven reference benign variants (p.Thr774Ala, p.Thr1087Ile, p.Thr1354Met, p.Asp1911Gly, p.Leu2396Ser, p.Arg2784Gln, and p.Ile3207Thr) were incorrectly assigned. Notably, five of these incorrect assignments were made by the same track (T142) (Table S13).

### Amino acid substitutions, protein modular domains, and structural motifs

We hypothesized that variants assigned to different functional categories (functional impact, normal function, or hypomorph) would display specific amino acid substitution patterns. First, we determined the frequency and

enrichment ratio (compared with all variants tested) for all possible amino acid substitutions in each functional category (Tables S14 and S15). We then focused on the top ten most enriched and most depleted substitutions. *BRCA1* and *BRCA2* variants having functional impact (PS3\_variable) were enriched in amino acid substitutions that changed amino acid class (hydrophobic, polar, charged, and special cases), particularly from hydrophobic to other classes, and depleted in variants with substitutions that remained in the same or related (e.g., charged and polar) class (Figure 3A). Conversely, *BRCA1* and *BRCA2* variants scoring as functionally normal (BS3\_variable) were enriched in amino acid substitutions that remained in the same or related amino acid class and depleted in substitutions that changed class (Figure 3B).

*BRCA1* and *BRCA2* variants that received the provisional label of hypomorph had a pattern of substitution enrichment and depletion similar to variants with functional impact (Figure 3C). However, the enrichment pattern differed between *BRCA1* and *BRCA2*: hypomorph *BRCA1* variants were enriched for cross-class substitutions originating in hydrophobic residues, whereas hypomorph *BRCA2* variants were enriched for cross-class substitutions originating in polar or charged residues (Figure 3C). As expected, there was considerable overlap between (1) enriched substitutions in functional impact assignments and depleted substitutions in normal function assignments and (2) enriched substitutions in normal function assignments and depleted substitutions in functional impact assignments (Figure S3).

Part of the difference in amino acid substitutions driving functional assignment of pathogenic, benign, or hypomorph can be ascribed to physicochemical characteristics as described by Grantham distances and by side-chain volume differences (Figures 3D and 3E).<sup>37</sup>

Most variants receiving assignments of functional impact (PS3\_variable) were part of recognized functional domains and were likely to disrupt secondary structures within these regions. In *BRCA1*, 99% of functional impact assignments were present either in the RING finger or the BRCT domains, although these regions have been more extensively tested (Figures 4A and 4B).

In *BRCA2*, 67.2% of functional impact assignments were located in the PALB2 binding region or the C-terminal DNA-binding domain (Figures 4A and 4B). Although pathogenic reference variants are restricted to functional domains (Figure 4C), 190 variants showed functional impact outside of these domains, but only one is spliceogenic. Most of these variants were scored by the same assay (track T142), but only 11 of them (5.8%) are predicted by EVE to be pathogenic, and five are benign reference variants, raising the possibility that these variants were incorrectly scored as functionally compromised.

Even within known functional domains, some secondary structures appear to be relatively tolerant to substitutions. For example, within the *BRCA1* BRCT, loops in intervening seq 6 (aa 1726–1729), the  $\beta$ '2 sheet (aa

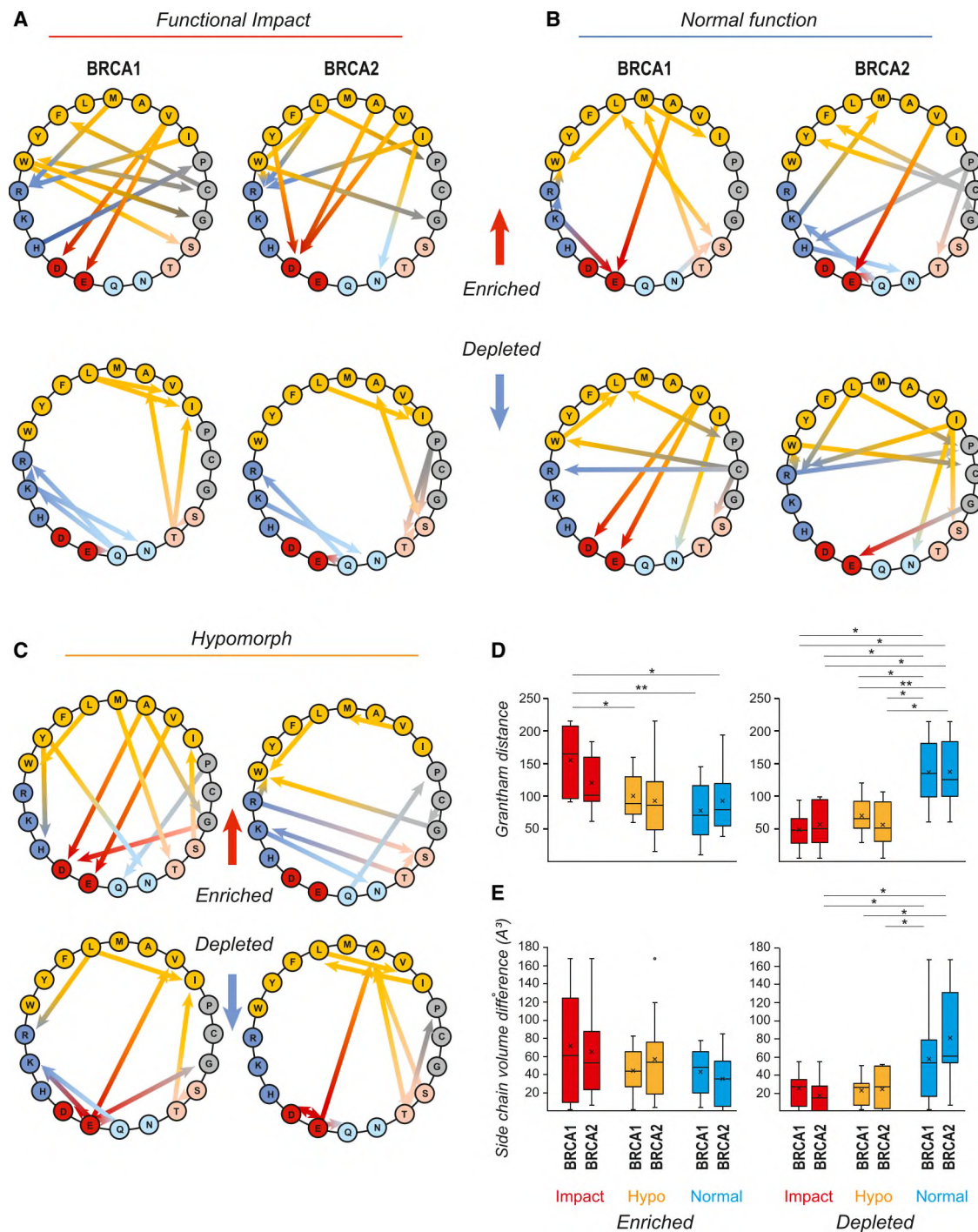


**Figure 3. Landscape of amino acid substitutions in *BRCA1* and *BRCA2* stratified by final assignment**

(A–C) Amino acid substitutions in variants assigned as functional impact (A), normal function (B), and hypomorph (C). Amino acids (one-letter code) are colored according to class: hydrophobic, yellow; polar negative, pink; polar positive, light blue; negatively charged, red; positively charged, dark blue; and special cases, gray. Arrows indicate the top ten enriched and depleted substitutions by enrichment ratio variant.

(D and E) Distribution of Grantham distance (D) and side-chain volume difference (E) in each set. \* $p \leq 0.05$  and \*\* $p \leq 0.01$ .

1790–1792), and the loop in intervening seq 12 (aa 1793–1804) had 104 variants tested, yet no variant with significant functional impact was found (Table S16). The  $\beta$ 2 sheet in the first BRCT, which corresponds spatially to  $\beta$ '2 in the second BRCT domain, is also relatively tolerant

to changes, consistent with their small contribution to the packing of BRCT domains.<sup>42</sup>

In the *BRCA2* C-terminal DNA-binding domain, nine secondary structures located in the helical, OB1, and OB2 domains were sensitive to substitution, with a  $\geq 0.25$  fraction

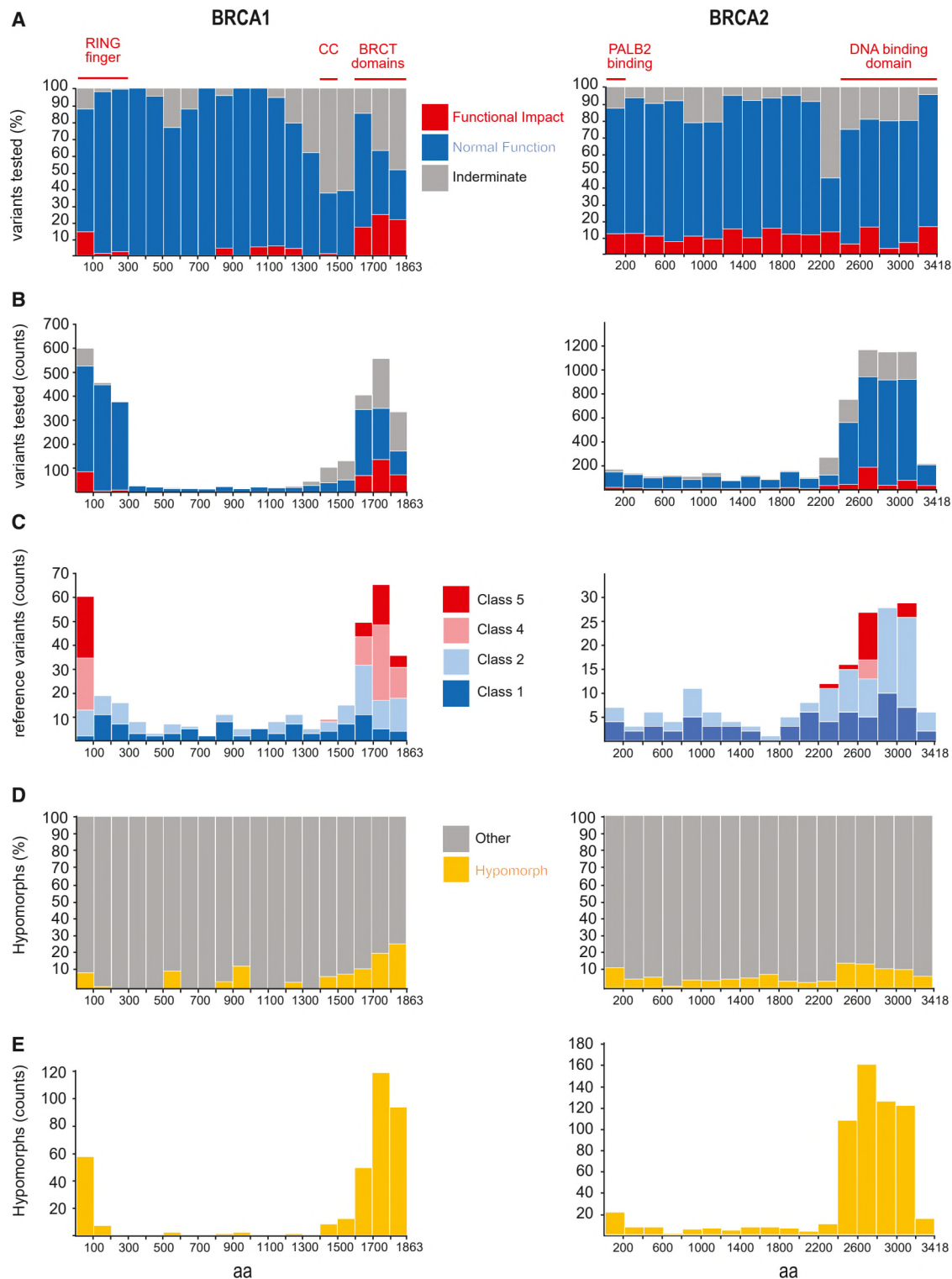


**Figure 4. Location of variants**

(A and B) Fraction (A) and absolute counts (B) of final assignments by 100- (*BRCA1*) or 200- (*BRCA2*) amino acid windows. The RING finger, the coiled-coil (CC), and the BRCT domains are indicated by red lines. Red, functional impact; blue, normal function; and gray, indeterminate.

(C) Location of reference variants annotated by IARC class. Red, class 5 (pathogenic); pink, class 4 (likely pathogenic); light blue, class 2 (likely benign); and dark blue (benign).

(D and E) Fraction (D) and absolute counts (E) of final hypomorph assignments. Yellow, hypomorphs, and gray, other.

of variants showing functional impact. Surprisingly, structures in the Tower domain were very tolerant to substitution (Table S16). To date, no LP/P reference variant has been identified in this region.

### Hypomorphic variants

We hypothesized that variants for which the functional results were intermediate (score = 1) and therefore were considered hypomorphic represent clinically relevant RPPVs.<sup>43–45</sup> Likewise, it is possible that RPPVs manifest themselves as generating discordant results across different assays.

We identified 350 *BRCA1* and 616 *BRCA2* variants that had at least one track scoring as intermediate (hypomorph observation set) (Figure 4D). In addition, 526 *BRCA1* and 924 *BRCA2* missense variants had unresolved discordant calls (unresolved discordant set) (Tables S12 and S13).

Further, we hypothesized that variants found in individuals with Fanconi anemia (FA) also represent RPPVs. To determine the extent to which these datasets contain RPPVs, we estimated the likelihood that the five *BRCA1* and 28 *BRCA2* RPPVs previously identified in individuals with FA<sup>45</sup> would be found in different variant sets (i.e., unresolved discordant, indeterminate, functional impact, normal function, and hypomorph observation) compared to the set of all variants (Table S17). *BRCA1* variants found in individuals with FA were enriched in the unresolved discordant, intermediate, and hypomorph observation sets. *BRCA2* FA variants were enriched in the unresolved discordant, indeterminate, functional impact, and hypomorph observation sets (Table S17).

### Resolution of variants with discordant scores with an ACMG/AMP-compatible point system

We compared the resolution of discordant variants using two approaches. In approach 1, the ACMG codes result from integrating the functional data and the preponderance-of-evidence rule (as described above) (Tables S12 and S13), and in approach 2, the functional data are integrated using the ACMG/AMP-compatible point system<sup>36,46</sup> (Tables S18 and S19).

We used the functional integration via the two approaches to resolve discordance in functional data for reference variants and assess their out-of-sample performance by comparing the results with the reference variant disposition (i.e., truth set) via cross-validation. For both genes, application of the point system to generate pathogenicity assertions based on the functional integration improved the resolution of discordant results. The rate of resolution (assignment of a non-indeterminate label) improved from 66.6% to 92.1% in *BRCA1* reference variants with discordant functional results and from 49.1% to 98.1% in *BRCA2* reference variants (Figures 5A and 5B). This improvement is accompanied by an increase in consequential (from benign to pathogenic or from pathogenic to benign) mislabeling from 0.8% to 5.5% in *BRCA1* and from 1.9% to 7.5% in *BRCA2*.

Despite high rates of resolution of reference variants, discordance is high (40.9% for *BRCA1* and 24.7% for *BRCA2*) among VUSs, and the preponderance-of-evidence approach resolves only a minority of variants (24.3% in *BRCA1* and 10% in *BRCA2*). A pairwise comparison between code- and point-based ACMG functional assertions for VUSs that received discordant scores also showed an improvement in resolution from 24.3% to 63.6% in *BRCA1* and from 10% to 66.4% in *BRCA2* (Figure 5C). Importantly, we did not observe any consequential change when moving from code-based to point-based integration. Taken together, these data show that moving to a point-based system to resolve variants with discordant results improves the overall resolution.

### Evaluating capping of functional data ACMG points

The application of BS3 and PS3 points is capped at  $-4$  and  $+4$ , respectively, to temper classification from a single source (i.e., functional data) (Tables S18 and S19). Limiting the computation of points from multiple tracks to  $-4$  for benign variants and  $+4$  for PVs penalizes 41.1% (1,334/3,247) of *BRCA1* variants and 52.8% (3,261/6,177) of *BRCA2* variants that received over four points (*BRCA1* LP/P range:  $+5$  to  $+82$ ; LB/B range:  $-5$  to  $-108$ ; *BRCA2* LP/P range:  $+6$  to  $+52$ ; and LB/B range:  $-5$  to  $-39$ ). This cap conflates variants that receive BS3\_variable or PS3\_variable labels with concordant results from a single assay with those that receive labels supported by results from multiple assays. Because individual assays have limitations (e.g., variants with functional impact due to (1) splicing defects will not be detected in cDNA-based assays or (2) temperature-sensitive effects may not be detected in yeast-based assays and variants with no functional impact may show functional defects due to technical errors or incorrect thresholds), variants tested multiple times with concordant results are less likely to be incorrectly labeled. To retain the information without violating the cap, we annotate variants that receive final LP/P or LB/B classifications supported by multiple concordant assays (Tables S18 and S19, see column notes).

To assess the rate of potential incorrect clinical classification if the cap was not applied, we determined how many reference variants would have been incorrectly classified. For *BRCA1*, one LB/B variant would have been incorrectly classified as LP/P (p.Ile1275Val, 16 points), whereas four LP/P (p.Arg71Thr,  $-8$  points; p.Met1411Thr,  $-10$  points; p.Ala1823Thr,  $-8$  points; and p.Gly1803Ala,  $-12$  points) would have been incorrectly classified as benign, totaling a 1.4% (5/351) consequential misclassification rate. For *BRCA2*, only two LB/B variants (p.Arg2784Gln, 23 points, and p.Val2908Gly, 8 points) were incorrectly classified as pathogenic, resulting in a 1.1% (2/176) consequential misclassification rate, within the acceptable error rates in the IARC (5%) and ACMG (10%) frameworks.

Taken together, the analysis of this small dataset (reference variants only) suggests that variants reaching  $> +23$  or  $< -12$  could be classified using only functional data.

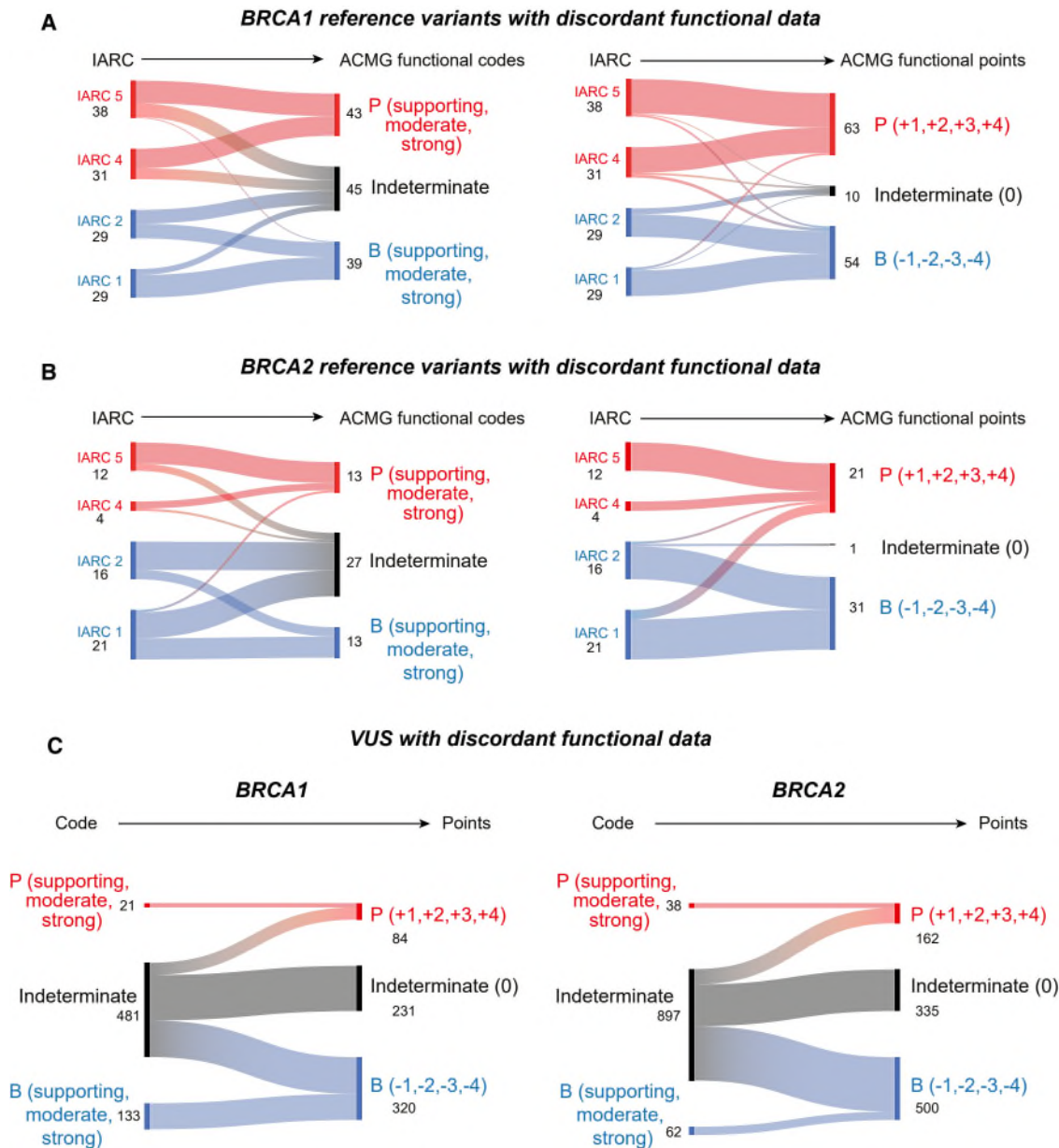


**Figure 5. Comparison of ACMG code-base and point-base integration on the resolution of variants with discordant results** (A and B) Resolution of discordant results for *BRCA1* (A) and *BRCA2* (B) reference variants. (C) Resolution of discordant results for *BRCA1* and *BRCA2* VUSs. Codes favoring pathogenicity (PS3\_supporting, PS3\_moderate, and PS3\_strong) or non-pathogenicity (BS3\_supporting, BS3\_moderate, and BS3\_strong) were collapsed into pathogenic (P) or benign (B) classes, respectively. Points favoring pathogenicity (+1, +2, +3, and +4) or non-pathogenicity (−1, −2, −3, and −4) were collapsed into P or B classes, respectively.

### Cancer risks associated with functional categories

Next, we determined the association of variants, aggregated by final functional code labels (BS3\_variable, indeterminate, and PS3\_variable), with breast cancer risk. We used female individuals affected by breast cancer who received hereditary cancer genetic testing by Ambry Genetics and female control subjects from AoU or gnomAD 4.1.

*BRCA1* PS3\_variable missense variants were associated with high risks of breast cancer (odds ratio [OR] > 4.8) in all case-control analyses, in line with those obtained for premature protein termination variants (PTVs) or known

PVs using the same datasets (Table 1). Variants that remained indeterminate after functional integration displayed moderately increased risks ( $1.8 < \text{OR} < 3.4$ ), suggesting that they may still contain PVs. In contrast, BS3\_variable missense variants were not associated with significantly increased risks of breast cancer (Table 1). The Ambry Genetics versus AoU comparison displayed a significantly reduced OR compared with the other comparisons, suggesting selection bias perhaps due to ascertainment differences and highlighting the limitations of pooled analyses.

**Table 1. Associations between variants categorized by integration of functional data and breast cancer risk**

|                                                           | OR    | <i>p</i> value | 95% CI     | Case AC | Case AN | Case freq | Control AC | Control AN | Control freq | Adjusted <i>p</i> value |
|-----------------------------------------------------------|-------|----------------|------------|---------|---------|-----------|------------|------------|--------------|-------------------------|
| <b><i>BRCA1 PS3</i></b>                                   |       |                |            |         |         |           |            |            |              |                         |
| Ambry (Cs) x AoU (Ct)                                     | 9.81  | 0.000          | 6.21–16.39 | 166     | 514,288 | 0.03%     | 19         | 385,856    | 0.00%        | 1.19e-31                |
| Ambry (Cs) x gnomAD v.4.1 (Ct)                            | 4.87  | 1.08e-12       | 3.22–7.74  | 191     | 631,308 | 0.03%     | 28         | 363,940    | 0.01%        | 1.73e-12                |
| CARRIERS+BRIDGES (Cs x Ct)                                | 10.27 | 6.58e-09       | 3.78–31.42 | 40      | 162,146 | 0.02%     | 4          | 166,494    | 0.00%        | 1.05e-07                |
| <b><i>BRCA1 indeterminate (VUS)</i></b>                   |       |                |            |         |         |           |            |            |              |                         |
| Ambry (Cs) x AoU (Ct)                                     | 1.87  | 4.56e-10       | 1.53–2.28  | 302     | 514,288 | 0.06%     | 165        | 385,856    | 0.04%        | 4.56e-10                |
| Ambry (Cs) x gnomAD v.4.1 (Ct)                            | 3.38  | 8.48e-20       | 2.62–4.42  | 384     | 631,308 | 0.06%     | 70         | 363,940    | 0.02%        | 1.70e-19                |
| CARRIERS+BRIDGES (Cs x Ct)                                | 1.75  | 3.89e-06       | 1.38–2.22  | 182     | 162,146 | 0.11%     | 107        | 166,494    | 0.06%        | 3.11e-05                |
| <b><i>BRCA1 BS3</i></b>                                   |       |                |            |         |         |           |            |            |              |                         |
| Ambry (Cs) x AoU (Ct)                                     | 0.69  | 3.07e-08       | 0.60–0.79  | 469     | 514,288 | 0.09%     | 536        | 385,856    | 0.14%        | 3.07e-08                |
| Ambry (Cs) x gnomAD v.4.1 (Ct)                            | 1.28  | 7.78e-04       | 1.11–1.48  | 605     | 631,308 | 0.10%     | 283        | 363,940    | 0.08%        | 8.89e-04                |
| CARRIERS+BRIDGES (Cs x Ct)                                | 1.05  | 0.118          | 0.99–1.11  | 2,309   | 162,146 | 1.42%     | 2,264      | 166,494    | 1.36%        | 0.189                   |
| <b><i>BRCA1 reference</i></b>                             |       |                |            |         |         |           |            |            |              |                         |
| PTV Ambry (Cs) x AoU (Ct)                                 | 9.21  | 0.000          | 8.14–10.46 | 2,532   | 514,288 | 0.49%     | 287        | 385,856    | 0.07%        | 0.000                   |
| PV CARRIERS+BRIDGES (Cs x Ct)                             | 8.58  | 1.28e-141      | 6.91–10.67 | 790     | 162,146 | 0.49%     | 95         | 166,494    | 0.06%        | 0.000                   |
| <b><i>BRCA2 PS3</i></b>                                   |       |                |            |         |         |           |            |            |              |                         |
| Ambry (Cs) x AoU (Ct)                                     | 2.81  | 0.000          | 2.30–3.44  | 435     | 514,316 | 0.08%     | 140        | 385,856    | 0.04%        | 6.12e-27                |
| Ambry (Cs) x gnomAD v.4.1 (Ct)                            | 2.56  | 1.9e-20        | 2.10–3.13  | 531     | 631,336 | 0.08%     | 120        | 362,999    | 0.03%        | 5.08e-20                |
| CARRIERS+BRIDGES (Cs x Ct)                                | 1.62  | 7.03e-04       | 1.22–2.14  | 129     | 162,146 | 0.08%     | 82         | 166,494    | 0.05%        | 0.002                   |
| <b><i>BRCA2 PS3 restricted to DBD domain variants</i></b> |       |                |            |         |         |           |            |            |              |                         |
| Ambry (Cs) x AoU (Ct)                                     | 5.39  | 0.000          | 4.06–7.29  | 338     | 514,136 | 0.07%     | 56         | 385,856    | 0.01%        | 2.50e-39                |
| Ambry (Cs) x gnomAD v.4.1 (Ct)                            | 5.20  | 5.80e-26       | 3.87–7.16  | 410     | 631,336 | 0.06%     | 46         | 362,999    | 0.01%        | 5.80e-26                |
| CARRIERS+BRIDGES (Cs x Ct)                                | 2.55  | 4.50e-06       | 1.68–3.93  | 77      | 162,146 | 0.05%     | 31         | 166,494    | 0.02%        | 1.35e-05                |
| <b><i>BRCA2 indeterminate (VUS)</i></b>                   |       |                |            |         |         |           |            |            |              |                         |
| Ambry (Cs) x AoU (Ct)                                     | 1.15  | 0.0936         | 0.98–1.35  | 408     | 514,316 | 0.08%     | 283        | 385,856    | 0.07%        | 0.0936                  |
| Ambry (Cs) x gnomAD v.4.1 (Ct)                            | 1.31  | 6.11e-04       | 1.13–1.54  | 516     | 631,336 | 0.08%     | 269        | 362,999    | 0.07%        | 8.15e-04                |
| CARRIERS+BRIDGES (Cs x Ct)                                | 1.03  | 0.403          | 0.96–1.11  | 1,530   | 162,146 | 0.94%     | 1,524      | 166,494    | 0.92%        | 0.496                   |
| <b><i>BRCA2 BS3</i></b>                                   |       |                |            |         |         |           |            |            |              |                         |

(Continued on next page)

**Table 1. Continued**

|                                   | OR   | p value   | 95% CI    | Case AC | Case AN | Case freq | Control AC | Control AN | Control freq | Adjusted p value |
|-----------------------------------|------|-----------|-----------|---------|---------|-----------|------------|------------|--------------|------------------|
| Ambry (Cs) x AoU (Ct)             | 1.05 | 0.459     | 0.93–1.19 | 624     | 514,316 | 0.12%     | 540        | 385,856    | 0.14%        | 0.459            |
| Ambry (Cs) x gnomAD v.4.1 (Ct)    | 1.19 | 0.005     | 1.05–1.35 | 788     | 631,336 | 0.12%     | 398        | 362,999    | 0.11%        | 0.005            |
| CARRIERS+BRIDGES (Cs x Ct)        | 0.95 | 0.038     | 0.90–1.00 | 3,060   | 162,146 | 1.89%     | 3,309      | 166,494    | 1.99%        | 0.076            |
| <b>BRCA2 reference</b>            |      |           |           |         |         |           |            |            |              |                  |
| PTV Ambry (Cs) x AoU (Ct)         | 5.38 | 0.000     | 4.94–5.86 | 3,906   | 514,316 | 0.76%     | 663        | 385,856    | 0.17%        | 0.000            |
| PV CARRIERS+BRIDGES (Cs x Ct)     | 5.68 | 1.18e-166 | 4.90–6.58 | 1,171   | 162,146 | 0.72%     | 213        | 166,494    | 0.13%        | 0.000            |
| MAVE PV Ambry (Cs) x gnomAD v.4.1 | 5.55 | 1.64e-46  | 4.14–7.53 | 566     | 758,236 | 0.07%     | 49         | 363,928    | 0.01%        | 0.000            |

p values marked as 0.000 are for extremely low values. OR, odds ratio; 95% CI, 95% confidence interval; Case AC, allele count in affected individuals; Case AN, total number of alleles in sample; Case Freq, allele frequency in affected individuals; Control AC, allele count in control subjects; Control AN, total number of alleles in control subjects; Control Freq, frequency in control subjects; PS3, includes PS3 only (as there were no events for PS3\_moderate or PS3\_supporting, ENIGMA-designated benign variants were excluded from PS3 OR calculation [*BRCA1*: p.Ile1275Val; *BRCA2*: p.Thr1354Met, p.Glu1593Asp, and p.Leu2396Phe], and recurrent variant p.Asn2622Ser, which was enriched in African American individuals, was also excluded from *BRCA2* PS3 calculation); (Cs), affected individuals; AoU, All of Us; (Ct), control subjects; BS3, includes BS3\_supporting, BS3\_moderate, and BS3; *BRCA1/2* reference, ORs associated with variants with disposition obtained independently of functional data or from multiplex assays of variant effects (MAVEs); PTV, premature termination variants; PV, known pathogenic variants; PV, missense variants scoring as pathogenic in *BRCA2* MAVE.<sup>28</sup>

*BRCA2* PS3\_variable missense variants were associated with elevated risks of breast cancer (OR > 2.5) in both case-control analyses, albeit attenuated compared with those obtained for PTVs or known PVs using the same datasets, suggesting that this category may improperly contain variants with no clinically relevant impact on function. Interestingly, when we limited our analysis to the C-terminal DNA-binding domain, ORs in the Ambry Genetics case-control studies were larger than 5, consistent with those obtained for PTVs or known PVs using the same datasets (Table 1). These results suggested that the attenuation in OR was due to variants incorrectly labeled as PS3\_variable in the regions outside the C-terminal DNA-binding domain.

A slightly but significantly elevated OR for BS3\_variable variants in the Ambry Genetic versus gnomAD case-control analysis raises the possibility that this category may contain variants with clinically relevant impact on function, as variants with a significant impact on function may be missed by specific assays. On the other hand, similar to what was found for BS3\_variable variants in *BRCA1*, there was no evidence for association with risk for *BRCA2* BS3\_variable variants in the AoU comparison and the CARRIERS+BRIDGES population-based case-control analyses (Table 1).

Taken together, these analyses confirm that BS3\_variable and PS3\_variable variants categorized using the integrated functional framework are associated with population average and increased risk for breast cancer, respectively. Population-based case-control studies, albeit with far fewer affected individuals, were consistent with the results from other case-control comparisons (Table 1). These results pro-

vide further support for the validity of the integration of functional data.

### Integration of functional data with other sources of information to classify variants

To formally classify variants, we integrated the functional data with other available sources of information (Tables S18 and S19). We used EVE as the *in silico* predictor and, following the ACMG guidelines for PP3 and BP4 (*in silico* predictions), we assigned +1 and –1 (supporting evidence) to variants considered pathogenic and benign, respectively. For PM2 (absent/rare from control subjects in an ethnically matched cohort population sample), we assigned +1 (supporting evidence) for variants absent from gnomAD 4.1. For PP1 and BS4 (co-segregation or lack of co-segregation with disease in multiple affected family members), we integrated previously published data and assigned points according to the likelihood ratios in the original studies.<sup>32,47</sup> For BA1 (stand-alone allele frequency), we assigned –8 points for variants with gnomAD filtering allele frequency (FAF) ≥ 0.001. For BS1 (allele frequency greater than expected for disease), variants with FAF ≥ 0.0001 were assigned –4, and variants with FAF ≥ 0.00002 and <0.0001 were assigned –1. No VUS achieved the stringent threshold of OR ≥ 4 (with the lower 95% CI excluding 2.0 and with  $p \leq 0.05$ ) in case-control analyses; therefore, we did not include points for the PS4 code.<sup>48</sup>

The total points from all codes were computed for each variant, and variants were classified as P (≥10 points), LP (6–9 points), LB (–6 to –2 points), or B (≤–7 points). Variants scoring between –1 and 5 points were labeled as VUSs. In *BRCA1*, 2,896 VUSs for which functional data

were available were assessed, and the integration of points corresponding to ACMG codes led to ClinVar-compliant classifications of 2,085 variants as LB/B and 260 as LP/P, and 551 remained VUSs, representing a resolution rate of 80.1% (Figure 6A). In *BRCA2*, 6,001 VUSs were assessed and resulted in 3,849 LB/B and 323 LP/P, and 1,829 remained VUSs, with a resolution rate of 69.5% (Figure 6B).

Finally, we evaluated the impact of the integrated classification on variants present in a population underrepresented in genomic studies. Of 328 *BRCA1* variants identified in individuals of AA ancestry in gnomAD, 84 were VUSs or conflicting classifications in ClinVar and were functionally tested (Table S20). Incorporating the integrated functional classification led to the classification of 71 variants (84.5% resolution), all going to the LB/B group (Figure S4). For the 215 *BRCA2* VUSs or conflicting classifications functionally assessed, 76.3% were resolved (Figure S4), identifying six LP variants (p.Ser2616Tyr, p.Ile2672Thr, p.Gly2724Glu, p.Trp2788Leu, p.Tyr3006Asp, and p.Glu3167Asp). All LP variants were very rare (allele frequency  $< 1.3 \times 10^{-6}$ ;  $< 1.4 \times 10^{-5}$  in AA individuals), five of which were only found in AA individuals (except p.Trp2788Leu) (Table S20).

## Discussion

A significant challenge for the implementation of precision medicine in oncogenomics lies in the accurate interpretation of germline and somatic genetic data. For carriers of variants in *BRCA1* and *BRCA2*, the ability to interpret germline genetic data has implications for risk stratification, prevention, and treatment for several types of cancer.

We reasoned that experimental data could be harnessed to address the challenges of variant classification for cancer susceptibility genes. Here, we systematically collected, curated, and harmonized data from functional assays testing 9,424 missense variants in *BRCA1* and *BRCA2* published in the last 28 years. We leveraged these data to (1) determine the strength of evidence for each assay using defined reference set variants following ClinGen recommendations,<sup>30,36</sup> (2) assign evidence strength based on all relevant calibrated assays for each variant, (3) determine the code/points applicable for each variant based on the integrated data, (4) generate a curated resource of functional data, and (5) integrate functional data with other sources of information to classify 6,517 *BRCA1* and *BRCA2* variants as LB/B or LP/P.

First, we focused on developing and refining rules to conduct the integration of functional data from a large number of independent published functional assays. Analysis of reference variants with discordant results revealed a small percentage of incorrect assignment after integration ( $< 2\%$ ). These incorrect assignments highlighted the need to account for RNA splicing defects,<sup>48</sup> and splice prediction should always be checked before

assuming that the functional effect is due to the missense substitution. More importantly, we demonstrate that the functional integration framework represented here displays a strong performance compatible with the requirements of ClinGen guidelines. Furthermore, variant categories based on functional assignment align well with risk of breast cancer in large case-control studies, confirming its validity for clinical use in variant classification.

The integrated functional data analysis conducted herein also provides a detailed view of regions, domains, and motifs important for function and how structural-based metrics such as Grantham distance and differences in side-chain volume contribute to the discrimination between functional impact, hypomorphic, and normal function assignments. For *BRCA1*, 99% of functional impact assignments were in the RING finger or the BRCT domains, in agreement with the notion that other regions in the protein constitute “cold spots” and are unlikely to have any pathogenic missense variant.<sup>49</sup> An important caveat is that functional assessments have been targeted in the RING and BRCT regions, creating a bias in variant testing by location. *BRCA2* presented a relatively lower rate (67%) of variants with functional impact located in the known functional regions. However, a vast majority of the functional impact assignments in the *BRCA2* cold spots were derived from a single assay and not corroborated by structural predictions. This suggests potential issues with random errors of specific assays rather than previously undetected functionally important regions. Since the assay was credentialed, it suggests that the OddsPath metric presents some shortcomings. The data support the use of the Dines et al. cold spot information in variant classification.<sup>49</sup> Our analyses also indicate that some specific structures within “hot spots,” such as the Tower domain in the *BRCA2* C-terminal DNA-binding domain, are unlikely to harbor pathogenic missense variants.

As an alternative to evidence criteria codes, a point-based system derived from the Bayesian formulation of the ACMG/AMP guidelines has been proposed.<sup>36,46</sup> As the points are proportional to  $\log(\text{odds})$  rather than to the OddsPath, they are additive. Our data show that moving from codes to points facilitates resolution of variants with discordant results. Importantly, ACMG/AMP recommends that a variant should not be clinically classified using a single type of evidence, in particular when classifying LP and PVs.<sup>29</sup> That tenet, which forms the basis for capping the functional scores, presupposes that functional assays are not independent and should be considered a single type of source of evidence. However, for *BRCA1* and *BRCA2*, some tests in the vast array of assays exploring different activities may be sufficiently independent of one another. Therefore, the combination of results in this way may be a principled approach to the natural accumulation of evidence. A further refinement of this model would be to develop a matrix of interdependencies across tracks, with contributions to adjusted final scores. In addition, one could also weigh contributions of

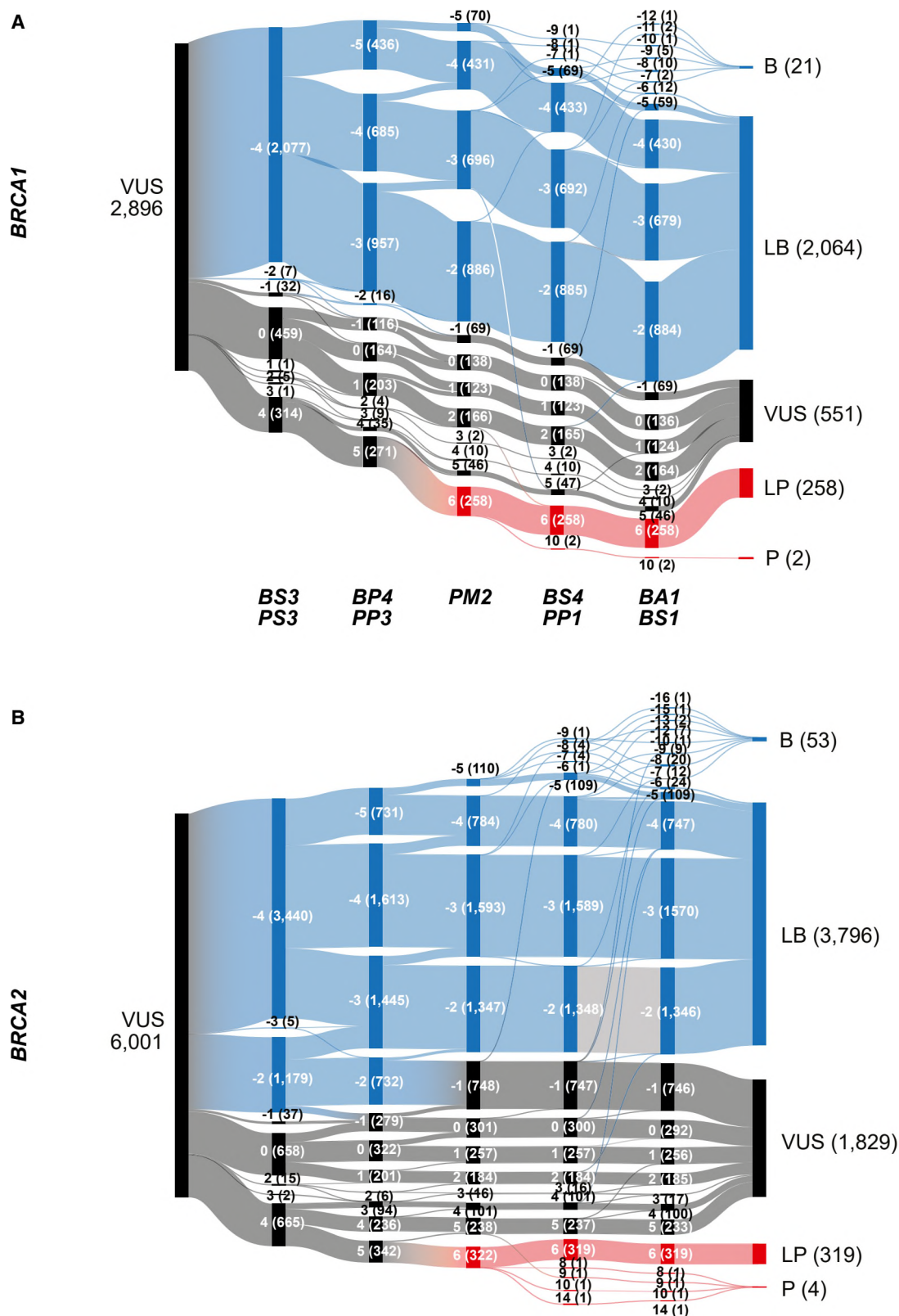


**Figure 6. ACMG-compliant classification**

Sankey diagram showing integration of applicable ACMG codes and points to resolve variants of uncertain significance in *BRCA1* (A) and *BRCA2* (B). Refer to [Tables S19](#) and [S20](#). Stepwise application of ACMG codes BS3/PS3, BP4/PP3, PM2, and BA1/BS1 and points according to the strength of evidence (supporting  $\pm 1$ , moderate  $\pm 2$ , strong  $\pm 4$ , and very strong  $\pm 8$ ). The total points from all codes considered and variants were classified as pathogenic (P) ( $\geq 10$  points), likely pathogenic (LP) (6–9 points), likely benign (LB) (–6 to –2 points), or benign (B) ( $\leq -7$  points). Variants scoring between –1 and 5 points remained VUSs.

different assays according to their uncertainty (judged by the CIs of sensitivity and specificity of each assay).

A more formal model-based approach would involve fitting to the data a hierarchical model that includes groupings associated with variant, assay type, lab, and experimental replicate or batch, as available. The resulting partial pooling, or statistical shrinkage, would naturally adapt to and account for structural dependencies present in the data; variation in the operating characteristics of the various assays would likewise be accounted for by the inclusion of appropriate random-effects distributions. The resource presented in the current study provides a basis for advanced simulations of changes in rules of classification.

Functional data for *BRCA1* and *BRCA2* are consistent with the existence of hypomorphic variants in those genes. The data also provide insight into the relationship between level of functional impact and risk, as variants from individuals with FA, presumably RPPVs, were enriched in the set of variants scoring as intermediate (hypomorph observation set). As risk estimates are extremely difficult to obtain for individual variants, functional data will be instrumental to identify RPPVs.<sup>43,50</sup> RPPVs in *TP53* and *OTC* were enriched in the hypomorphic variant sets.<sup>51,52</sup> Thus, functional data integration can help in the identification of RPPVs, but functional assays should be calibrated with clinically relevant RPPVs. Once clinically calibrated, these variants can then be associated with risk-appropriate recommendations.<sup>43,50</sup> For example, an RPPV associated with a less than 20% lifetime risk of breast cancer may trigger increased surveillance but not justify offering prophylactic surgery.

To classify variants with existing functional data using the ACMG/AMP-compliant framework, we combined the functional integration (BS3/PS3 codes) with other available evidence (PP3/BP4, PM2, PP1/BS4, BA1, and BS1 codes). This approach led to a high-resolution rate (>80%) of VUSs, illustrating the power of functional integration to accelerate clinically meaningful classifications. Moreover, the approach was also impactful for the classification of variants found in AA individuals. Overall resolution rates of variants with discordant functional results using the point system were markedly higher than those using the code system (63.6% versus 24.3% for *BRCA1* and 66.4% versus 10% for *BRCA2*). However, a considerable fraction of missense VUSs (19.9% in *BRCA1* and 30.5% in *BRCA2*) lack or have conflicting data from multiple sources and are still unresolved. Improvements in analytical methodology, comprehensiveness of MAVes, better and more diverse estimates of allele frequency, and capture of associated clinical data are likely to close this gap in the near future.

Our study has several limitations. First, individual tracks display wide heterogeneity in measurements and design. For example, tracks have data on a variable number of variants (from a few variants to several thousand), with different numbers of experimental replicates and use var-

iable cutoffs to distinguish normal from abnormal function. Also, variants tested only once may achieve classifications that have the same weight as those tested multiple times. These issues are mitigated by harmonization and calibration of functional results and using the point-based system to integrate results. Some of these issues could be resolved using quantitative models adapted to integration.<sup>23,25,53</sup> However, challenges in accessing raw data still exist. Finally, models should also consider the full spectrum of risk to accommodate variants with reduced penetrance. However, there are currently no established missense references for RPPVs, although a few have been proposed.<sup>43–45,54</sup> Despite these limitations, the current study provides a framework for the clinical application of functional data.

In summary, we curated a large dataset of functional assays for missense variants of *BRCA1* and *BRCA2*. Using this dataset, we developed a data-driven ACMG/AMP-compliant framework for data integration. This integration framework showed strong performance, assessed using a rigorously curated panel of reference variants, and its clinical validity was supported by large case-control studies. Because this framework is based on general functional, structural, and logical attributes, this model can, with the necessary adjustments, be generalized to other proteins of clinical relevance with existing functional data.

## Data and code availability

Data for all tracks are included as [supplemental information](#) (Tables S1–S20). Datasets and codes for queries to generate summary statistics and variant calls are available through GitHub (<https://github.com/FunctionalAssayIntegration/BRCA1-BRCA2-Large-integration>). Ambry Genetics supports open data sharing. All data generated or analyzed during this study are included in this published article and its [supplemental information](#), except for detailed patient data that would compromise HIPAA-compliant safeguards applied to protect patient privacy.

## Acknowledgments

This work was supported by a Precision Prevention grant from the Breast Cancer Research Foundation (BCRF) (A.N.A.M.); the BRCA Research & Cure Alliance (CureBRCA) Foundation (K.L.N., A.N.A.M., and F.J.C.); the Sarasota Innovation Funds/Moffitt Foundation (A.N.A.M.); awards National Institutes of Health (NIH) R35CA253187 (F.J.C.), R01CA225662 (F.J.C.), P50CA116201 (F.J.C.), U24CA302705 (M.S.C.), and U24CA258058 (F.J.C. and A.B.S.); the BCRF (F.J.C. and K.L.N.); the Bassett Center for BRCA (K.L.N.); the Gray Foundation (K.L.N.); Dutch Society Grant 12754 (M.P.G.V.); Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (M.A.C.); Conselho Nacional de Desenvolvimento Científico e Tecnológico (M.A.C.); Fundação do Câncer (Programa de Oncobiologia) (M.A.C.); and Italian Association for Cancer Research AIRC IG 2018 Id.22093 (L.C. and P.R.). This research was supported in part by the Intramural Research Program of the NIH (S. Sahu). The contributions of the NIH authors were made as part of their official duties as NIH federal employees, are in compliance with agency policy requirements, and

are considered works of the United States government. However, the findings and conclusions presented in this paper are those of the authors and do not necessarily reflect the views of the NIH or the US Department of Health and Human Services. T.N. is the recipient of the Ann and Sol Schreiber Mentored Investigator Award from the Ovarian Cancer Research Alliance. The authors thank Tanisha Taneja for help with variant curation.

## Declaration of interests

M.E.R. is a full-time, salaried employee of Ambry Genetics.

## Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors used OpenAI's GPT-5.5 and Codex to develop and refine code to perform calculations and generate tables and figures. The authors reviewed and edited the outputs as needed and take full responsibility for the content of the published article.

## Supplemental information

Supplemental information can be found online at <https://doi.org/10.1016/j.ajhg.2026.09.013>.

## Web resources

GenBank, <https://www.ncbi.nlm.nih.gov/genbank/>  
BRCA Exchange, <https://brcaexchange.org/>  
gnomAD, <https://gnomad.broadinstitute.org/>  
ClinVar, <https://www.ncbi.nlm.nih.gov/clinvar/>  
All of Us Research Program, <https://allofus.nih.gov/>  
UK Biobank, <https://www.ukbiobank.ac.uk/>

Received: May 4, 2026

Accepted: September 21, 2026

## References

- Li, S., Silvestri, V., Leslie, G., Rebbeck, T.R., Neuhausen, S.L., Hopper, J.L., Nielsen, H.R., Lee, A., Yang, X., McGuffog, L., et al. (2022). Cancer Risks Associated With *BRCA1* and *BRCA2* Pathogenic Variants. *J. Clin. Oncol.* **40**, 1529–1541. <https://doi.org/10.1200/JCO.21.02112>.
- Hu, C., Hart, S.N., Gnanaolivu, R., Huang, H., Lee, K.Y., Na, J., Gao, C., Lilyquist, J., Yadav, S., Boddicker, N.J., et al. (2021). A Population-Based Study of Genes Previously Implicated in Breast Cancer. *N. Engl. J. Med.* **384**, 440–451. <https://doi.org/10.1056/NEJMoa2005936>.
- Hu, C., Hart, S.N., Polley, E.C., Gnanaolivu, R., Shimelis, H., Lee, K.Y., Lilyquist, J., Na, J., Moore, R., Antwi, S.O., et al. (2018). Association Between Inherited Germline Mutations in Cancer Predisposition Genes and Risk of Pancreatic Cancer. *JAMA* **319**, 2401–2409. <https://doi.org/10.1001/jama.2018.6228>.
- Breast Cancer Association Consortium, Dorling, L., Carvalho, S., Allen, J., González-Neira, A., Luccarini, C., Wahlström, C., Pooley, K.A., Parsons, M.T., Fortuno, C., et al. (2021). Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. *N. Engl. J. Med.* **384**, 428–439. <https://doi.org/10.1056/NEJMoa1913948>.
- Winter, C., Nilsson, M.P., Olsson, E., George, A.M., Chen, Y., Kvist, A., Törngren, T., Vallon-Christersson, J., Hegardt, C., Häkkinen, J., et al. (2016). Targeted sequencing of *BRCA1* and *BRCA2* across a large unselected breast cancer cohort suggests that one-third of mutations are somatic. *Ann. Oncol.* **27**, 1532–1538. <https://doi.org/10.1093/annonc/mdw209>.
- Gonzalez-Angulo, A.M., Timms, K.M., Liu, S., Chen, H., Litton, J.K., Potter, J., Lanchbury, J.S., Stemke-Hale, K., Hennesy, B.T., Arun, B.K., et al. (2011). Incidence and outcome of *BRCA* mutations in unselected patients with triple receptor-negative breast cancer. *Clin. Cancer Res.* **17**, 1082–1089. <https://doi.org/10.1158/1078-0432.CCR-10-2560>.
- Zhang, S., Royer, R., Li, S., McLaughlin, J.R., Rosen, B., Risch, H.A., Fan, I., Bradley, L., Shaw, P.A., and Narod, S.A. (2011). Frequencies of *BRCA1* and *BRCA2* mutations among 1,342 unselected patients with invasive ovarian cancer. *Gynecol. Oncol.* **121**, 353–357. <https://doi.org/10.1016/j.ygyno.2011.01.020>.
- Bryant, H.E., Schultz, N., Thomas, H.D., Parker, K.M., Flower, D., Lopez, E., Kyle, S., Meuth, M., Curtin, N.J., and Helleday, T. (2005). Specific killing of *BRCA2*-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* **434**, 913–917. <https://doi.org/10.1038/nature03443>.
- Farmer, H., McCabe, N., Lord, C.J., Tutt, A.N.J., Johnson, D.A., Richardson, T.B., Santarosa, M., Dillon, K.J., Hickson, I., Knights, C., et al. (2005). Targeting the DNA repair defect in *BRCA* mutant cells as a therapeutic strategy. *Nature* **434**, 917–921. <https://doi.org/10.1038/nature03445>.
- Fong, P.C., Boss, D.S., Yap, T.A., Tutt, A., Wu, P., Mergui-Roelvink, M., Mortimer, P., Swaisland, H., Lau, A., O'Connor, M.J., et al. (2009). Inhibition of poly(ADP-ribose) polymerase in tumors from *BRCA* mutation carriers. *N. Engl. J. Med.* **361**, 123–134. <https://doi.org/10.1056/NEJMoa0900212>.
- Matulonis, U.A., Penson, R.T., Domchek, S.M., Kaufman, B., Shapira-Frommer, R., Audeh, M.W., Kaye, S., Molife, L.R., Gelmon, K.A., Robertson, J.D., et al. (2016). Olaparib monotherapy in patients with advanced relapsed ovarian cancer and a germline *BRCA1/2* mutation: a multistudy analysis of response rates and safety. *Ann. Oncol.* **27**, 1013–1019. <https://doi.org/10.1093/annonc/mdw133>.
- Kennedy, R.D., and D'Andrea, A.D. (2006). DNA repair pathways in clinical practice: lessons from pediatric cancer susceptibility syndromes. *J. Clin. Oncol.* **24**, 3799–3808. <https://doi.org/10.1200/JCO.2005.05.4171>.
- Ciccio, A., and Elledge, S.J. (2010). The DNA damage response: making it safe to play with knives. *Mol. Cell* **40**, 179–204. <https://doi.org/10.1016/j.molcel.2010.09.019>.
- Fong, P.C., Yap, T.A., Boss, D.S., Carden, C.P., Mergui-Roelvink, M., Gourley, C., De Greve, J., Lubinski, J., Shanley, S., Messiou, C., et al. (2010). Poly(ADP)-ribose polymerase inhibition: frequent durable responses in *BRCA* carrier ovarian cancer correlating with platinum-free interval. *J. Clin. Oncol.* **28**, 2512–2519. <https://doi.org/10.1200/JCO.2009.26.9589>.
- Kaufman, B., Shapira-Frommer, R., Schmutzler, R.K., Audeh, M.W., Friedlander, M., Balmaña, J., Mitchell, G., Fried, G., Stemmer, S.M., Hubert, A., et al. (2015). Olaparib monotherapy in patients with advanced cancer and a germline *BRCA1/2* mutation. *J. Clin. Oncol.* **33**, 244–250. <https://doi.org/10.1200/JCO.2014.56.2728>.

16. Kim, G., Ison, G., McKee, A.E., Zhang, H., Tang, S., Gwise, T., Sridhara, R., Lee, E., Tzou, A., Philip, R., et al. (2015). FDA Approval Summary: Olaparib Monotherapy in Patients with Deleterious Germline BRCA-Mutated Advanced Ovarian Cancer Treated with Three or More Lines of Chemotherapy. *Clin. Cancer Res.* 21, 4257–4261. <https://doi.org/10.1158/1078-0432.CCR-15-0887>.
17. Daly, M.B., Pal, T., Maxwell, K.N., Churpek, J., Kohlmann, W., AlHilli, Z., Arun, B., Buys, S.S., Cheng, H., Domchek, S.M., et al. (2023). NCCN Guidelines(R) Insights: Genetic/Familial High-Risk Assessment: Breast, Ovarian, and Pancreatic, Version 2.2024. *J. Natl. Compr. Cancer Netw.* 21, 1000–1010. <https://doi.org/10.6004/jnccn.2023.0051>.
18. Cline, M.S., Liao, R.G., Parsons, M.T., Paten, B., Alquadoomi, F., Antoniou, A., Baxter, S., Brody, L., Cook-Deegan, R., Coffin, A., et al. (2018). BRCA Challenge: BRCA Exchange as a global resource for variants in BRCA1 and BRCA2. *PLoS Genet.* 14, e1007752. <https://doi.org/10.1371/journal.pgen.1007752>.
19. BRCA Exchange. <https://brcaexchange.org/>.
20. gnomAD - Genome Aggregation Database. (2024). <https://gnomad.broadinstitute.org/>.
21. Monteiro, A.N., Bouwman, P., Kousholt, A.N., Eccles, D.M., Millot, G.A., Masson, J.Y., Schmidt, M.K., Sharan, S.K., Scully, R., Wiesmüller, L., et al. (2020). Variants of uncertain clinical significance in hereditary breast and ovarian cancer genes: best practices in functional analysis for clinical annotation. *J. Med. Genet.* 57, 509–518. <https://doi.org/10.1136/jmedgenet-2019-106368>.
22. Findlay, G.M., Boyle, E.A., Hause, R.J., Klein, J.C., and Shendure, J. (2014). Saturation editing of genomic regions by multiplex homology-directed repair. *Nature* 513, 120–123. <https://doi.org/10.1038/nature13695>.
23. Findlay, G.M., Daza, R.M., Martin, B., Zhang, M.D., Leith, A.P., Gasperini, M., Janizek, J.D., Huang, X., Starita, L.M., and Shendure, J. (2018). Accurate classification of BRCA1 variants with saturation genome editing. *Nature* 562, 217–222. <https://doi.org/10.1038/s41586-018-0461-z>.
24. Clark, K.A., Paquette, A., Tao, K., Bell, R., Boyle, J.L., Rosenthal, J., Snow, A.K., Stark, A.W., Thompson, B.A., Unger, J., et al. (2022). Comprehensive evaluation and efficient classification of BRCA1 RING domain missense substitutions. *Am. J. Hum. Genet.* 109, 1153–1174. <https://doi.org/10.1016/j.ajhg.2022.05.004>.
25. Starita, L.M., Islam, M.M., Banerjee, T., Adamovich, A.I., Gullingsrud, J., Fields, S., Shendure, J., and Parvin, J.D. (2018). A Multiplex Homology-Directed DNA Repair Assay Reveals the Impact of More Than 1,000 BRCA1 Missense Substitution Variants on Protein Function. *Am. J. Hum. Genet.* 103, 498–508. <https://doi.org/10.1016/j.ajhg.2018.07.016>.
26. Diabate, M., Islam, M.M., Nagy, G., Banerjee, T., Dhar, S., Smith, N., Adamovich, A.I., Starita, L.M., and Parvin, J.D. (2023). DNA repair function scores for 2172 variants in the BRCA1 amino-terminus. *PLoS Genet.* 19, e1010739. <https://doi.org/10.1371/journal.pgen.1010739>.
27. Sahu, S., Galloux, M., Southon, E., Caylor, D., Sullivan, T., Arnaud, M., Zanti, M., Geh, J., Chari, R., Michailidou, K., et al. (2025). Saturation genome editing-based clinical classification of BRCA2 variants. *Nature* 638, 538–545. <https://doi.org/10.1038/s41586-024-08349-1>.
28. Huang, H., Hu, C., Na, J., Hart, S.N., Gnanaolivu, R.D., Abozaid, M., Rao, T., Tecleab, Y.A., et al.; CARRIERS Consortium, and Pesaran, T. (2025). Functional evaluation and clinical classification of BRCA2 variants. *Nature* 638, 528–537. <https://doi.org/10.1038/s41586-024-08388-8>.
29. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., 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. *Genetics in medicine. Genet. Med.* 17, 405–424. <https://doi.org/10.1038/gim.2015.30>.
30. Brnich, S.E., Abou Tayoun, A.N., Couch, F.J., Cutting, G.R., Greenblatt, M.S., Heinen, C.D., Kanavy, D.M., Luo, X., McNulty, S.M., Starita, L.M., et al. (2019). Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med.* 12, 3. <https://doi.org/10.1186/s13073-019-0690-2>.
31. Lyra, P.C.M., Jr., Nepomuceno, T.C., de Souza, M.L.M., Machado, G.F., Veloso, M.F., Henriques, T.B., Dos Santos, D.Z., Ribeiro, I.G., Ribeiro, R.S., Jr., Rangel, L.B.A., et al. (2021). Integration of functional assay data results provides strong evidence for classification of hundreds of BRCA1 variants of uncertain significance. *Genet. Med.* 23, 306–315. <https://doi.org/10.1038/s41436-020-00991-0>.
32. Parsons, M.T., Tadini, E., Li, H., Hahnen, E., Wappenschmidt, B., Feliubadaló, L., Aalfs, C.M., Agata, S., Aittomäki, K., Alducci, E., et al. (2019). Large scale multifactorial likelihood quantitative analysis of BRCA1 and BRCA2 variants: An ENIGMA resource to support clinical variant classification. *Hum. Mutat.* 40, 1557–1578. <https://doi.org/10.1002/humu.23818>.
33. ClinVar. (2023). <https://www.ncbi.nlm.nih.gov/clinvar/>.
34. Easton, D.F., Deffenbaugh, A.M., Pruss, D., Frye, C., Wenstrup, R.J., Allen-Brady, K., Tavtigian, S.V., Monteiro, A.N.A., Iversen, E.S., Couch, F.J., and Goldgar, D.E. (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, 873–883. <https://doi.org/10.1086/521032>.
35. Goldgar, D.E., Easton, D.F., Deffenbaugh, A.M., Monteiro, A.N.A., Tavtigian, S.V., Couch, F.J., and Breast Cancer Information Core BIC Steering Committee (2004). Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. *Am. J. Hum. Genet.* 75, 535–544. <https://doi.org/10.1086/424388>.
36. Tavtigian, S.V., Greenblatt, M.S., Harrison, S.M., Nussbaum, R.L., Prabhu, S.A., Boucher, K.M., Biesecker, L.G.; and ClinGen Sequence Variant Interpretation Working Group ClinGen SVI (2018). Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genet. Med.* 20, 1054–1060. <https://doi.org/10.1038/gim.2017.210>.
37. Grantham, R. (1974). Amino acid difference formula to help explain protein evolution. *Science* 185, 862–864. <https://doi.org/10.1126/science.185.4154.862>.
38. Frazer, J., Notin, P., Dias, M., Gomez, A., Min, J.K., Brock, K., Gal, Y., and Marks, D.S. (2021). Disease variant prediction with deep generative models of evolutionary data. *Nature* 599, 91–95. <https://doi.org/10.1038/s41586-021-04043-8>.
39. All of Us Research Program. (2025). U.S (Department of Health and Human Services). <https://allofus.nih.gov/>.
40. UK Biobank (2025). <https://www.ukbiobank.ac.uk/>.

41. Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Darbandi, S.F., Knowles, D., Li, Y.I., Kosmicki, J.A., Arbelaez, J., Cui, W., Schwartz, G.B., et al. (2019). Predicting Splicing from Primary Sequence with Deep Learning. *Cell* 176, 535–548.e24. <https://doi.org/10.1016/j.cell.2018.12.015>.
42. Williams, R.S., Green, R., and Glover, J.N. (2001). Crystal structure of the BRCT repeat region from the breast cancer-associated protein BRCA1. *Nat. Struct. Biol.* 8, 838–842. <https://doi.org/10.1038/nsb1001-838>.
43. Pal, T., Mundt, E., Richardson, M.E., Chao, E., Pesaran, T., Slavin, T.P., Couch, F.J., and Monteiro, A.N.A. (2024). Reduced penetrance BRCA1 and BRCA2 pathogenic variants in clinical germline genetic testing. *npj Precis. Oncol.* 8, 247. <https://doi.org/10.1038/s41698-024-00741-4>.
44. Domchek, S.M., Tang, J., Stopfer, J., Lilli, D.R., Hamel, N., Tischkowitz, M., Monteiro, A.N.A., Messick, T.E., Powers, J., Yonker, A., et al. (2013). Biallelic deleterious BRCA1 mutations in a woman with early-onset ovarian cancer. *Cancer Discov.* 3, 399–405. <https://doi.org/10.1158/2159-8290.CD-12-0421>.
45. Johnatty, S.E., Tudini, E., Parsons, M.T., Michailidou, K., Zanti, M., Canson, D.M., Davidson, A.L., Berger, T., Rosti, R.O., Kratz, C.P., et al. (2025). BRCA1-BRCA2-and PALB2-related Fanconi anemia: Scope to expand disease phenotypic features and predict breast cancer risk in heterozygotes. *Am. J. Hum. Genet.* 112, 2902–2921. <https://doi.org/10.1016/j.ajhg.2025.10.007>.
46. Tavtigian, S.V., Harrison, S.M., Boucher, K.M., and Biesecker, L.G. (2020). Fitting a naturally scaled point system to the ACMG/AMP variant classification guidelines. *Hum. Mutat.* 41, 1734–1737. <https://doi.org/10.1002/humu.24088>.
47. Caputo, S.M., Golmard, L., Léone, M., Damiola, F., Guillaud-Bataille, M., Revillion, F., Rouleau, E., Derive, N., Buisson, A., Basset, N., et al. (2021). Classification of 101 BRCA1 and BRCA2 variants of uncertain significance by cosegregation study: A powerful approach. *Am. J. Hum. Genet.* 108, 1907–1923. <https://doi.org/10.1016/j.ajhg.2021.09.003>.
48. Parsons, M.T., de la Hoya, M., Richardson, M.E., Tudini, E., Anderson, M., Berkofsky-Fessler, W., Caputo, S.M., Chan, R.C., Cline, M.S., Feng, B.J., et al. (2024). Evidence-based recommendations for gene-specific ACMG/AMP variant classification from the ClinGen ENIGMA BRCA1 and BRCA2 Variant Curation Expert Panel. *Am. J. Hum. Genet.* 111, 2044–2058. <https://doi.org/10.1016/j.ajhg.2024.07.013>.
49. Dines, J.N., Shirts, B.H., Slavin, T.P., Walsh, T., King, M.C., Fowler, D.M., and Pritchard, C.C. (2020). Systematic misclassification of missense variants in BRCA1 and BRCA2 "cold-spots". *Genetics in medicine. Genet. Med.* 22, 825–830. <https://doi.org/10.1038/s41436-019-0740-6>.
50. Pal, T., Christopher, J., Astiazaran-Symonds, E., Foulkes, W.D., James, P., Klugman, S., Kurian, A., Mak, J., Monteiro, A.N., Robson, M., et al. (2026). Consideration of inherited cancer risk on a continuum: An international and multidisciplinary perspective: A points to consider statement of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.* 28, 101659. <https://doi.org/10.1016/j.jgim.2025.101659>.
51. Lo, R.S., Cromie, G.A., Tang, M., Teng, K., Owens, K., Sirr, A., Kutz, J.N., Morizono, H., Caldovic, L., Ah Mew, N., et al. (2023). The functional impact of 1,570 individual amino acid substitutions in human OTC. *Am. J. Hum. Genet.* 110, 863–879. <https://doi.org/10.1016/j.ajhg.2023.03.019>.
52. Fortunato, C., Richardson, M.E., Pesaran, T., McGoldrick, K., James, P.A., and Spurdle, A.B. (2025). Characteristics predicting reduced penetrance variants in the high-risk cancer predisposition gene TP53. *HGG Adv.* 6, 100484. <https://doi.org/10.1016/j.xhgg.2025.100484>.
53. Iversen, E.S., Couch, F.J., Goldgar, D.E., Tavtigian, S.V., and Monteiro, A.N.A. (2011). A Computational Method to Classify Variants of Uncertain Significance Using Functional Assay Data with Application to BRCA1. *Cancer Epidemiol. Biomarkers Prev.* 20, 1078–1088. <https://doi.org/10.1158/1055-9965.epi-10-1214>.
54. Lovelock, P.K., Spurdle, A.B., Mok, M.T.S., Farrugia, D.J., Lakhani, S.R., Healey, S., Arnold, S., Buchanan, D., et al.; kConFab Investigators, and Couch, F.J. (2007). Identification of BRCA1 missense substitutions that confer partial functional activity: potential moderate risk variants? *Breast Cancer Res.* 9, R82. <https://doi.org/10.1186/bcr1826>.