

# Functional evaluation and clinical classification of *BRCA2* variants

<https://doi.org/10.1038/s41586-024-08388-8>

Received: 10 December 2023

Accepted: 12 November 2024

Published online: 8 January 2025

Open access

 Check for updates

Huaizhi Huang<sup>1,2,3,27</sup>, Chunling Hu<sup>1,27</sup>✉, Jie Na<sup>4</sup>, Steven N. Hart<sup>1</sup>, Rohan David Gnanaolivu<sup>4</sup>, Mohamed Abozaid<sup>1</sup>, Tara Rao<sup>1</sup>, Yohannes A. Tecleab<sup>4</sup>, CARRIERS Consortium\*, Tina Pesaran<sup>5</sup>, Paulo Cilas Morais Lyra<sup>6</sup>, Rachid Karam<sup>5</sup>, Siddhartha Yadav<sup>7</sup>, Katherine L. Nathanson<sup>8</sup>, Susan M. Domchek<sup>8</sup>, Miguel de la Hoya<sup>9</sup>, Mark Robson<sup>10</sup>, Miika Mehine<sup>11,12</sup>, Chaitanya Bandlamudi<sup>12</sup>, Diana Mandelker<sup>11</sup>, Alvaro N. A. Monteiro<sup>6</sup>, Edwin S. Iversen<sup>13</sup>, Nicholas Boddicker<sup>4</sup>, Wenan Chen<sup>4</sup>, Marcy E. Richardson<sup>5</sup> & Fergus J. Couch<sup>1,4</sup>✉

Germline *BRCA2* loss-of-function variants, which can be identified through clinical genetic testing, predispose to several cancers<sup>1–5</sup>. However, variants of uncertain significance limit the clinical utility of test results. Thus, there is a need for functional characterization and clinical classification of all *BRCA2* variants to facilitate the clinical management of individuals with these variants. Here we analysed all possible single-nucleotide variants from exons 15 to 26 that encode the *BRCA2* DNA-binding domain hotspot for pathogenic missense variants. To enable this, we used saturation genome editing CRISPR–Cas9-based knock-in endogenous targeting of human haploid HAP1 cells<sup>6</sup>. The assay was calibrated relative to nonsense and silent variants and was validated using pathogenic and benign standards from ClinVar and results from a homology-directed repair functional assay<sup>7</sup>. Variants (6,959 out of 6,960 evaluated) were assigned to seven categories of pathogenicity based on a VarCall Bayesian model<sup>8</sup>. Single-nucleotide variants that encode loss-of-function missense variants were associated with increased risks of breast cancer and ovarian cancer. The functional assay results were integrated into models from ClinGen, the American College of Medical Genetics and Genomics, and the Association for Molecular Pathology<sup>9</sup> for clinical classification of *BRCA2* variants. Using this approach, 91% were classified as pathogenic or likely pathogenic or as benign or likely benign. These classified variants can be used to improve clinical management of individuals with a *BRCA2* variant.

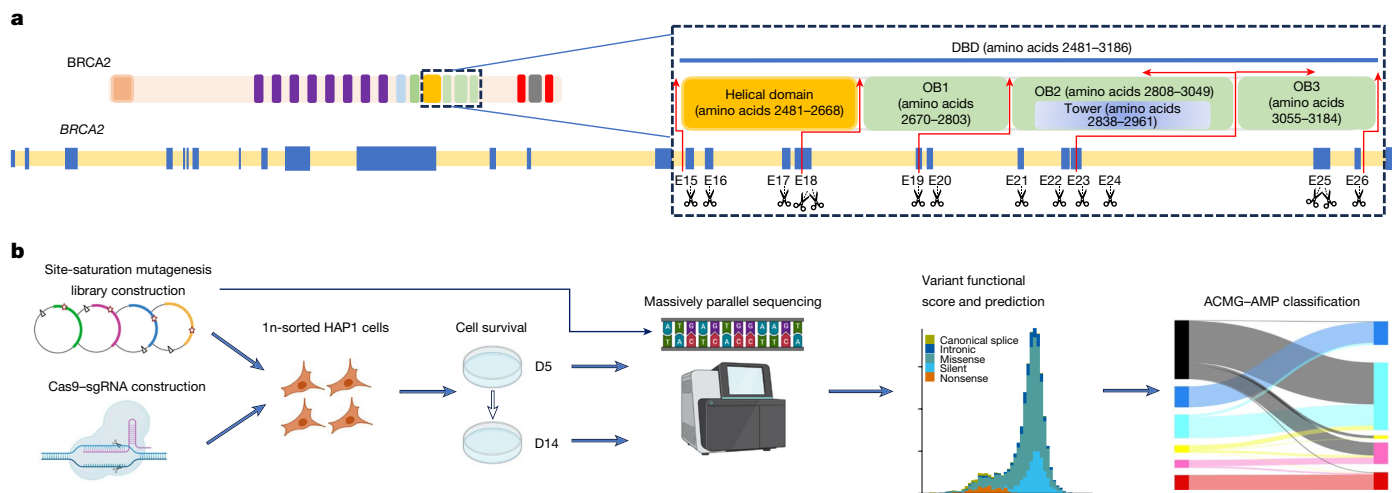
*BRCA2* is an established clinically actionable cancer predisposition gene<sup>5</sup> and has been widely used to test for hereditary cancer risk. In particular, *BRCA2* loss-of-function pathogenic variants are associated with a 69% lifetime risk of developing breast cancer<sup>2</sup> and a 15% risk of developing ovarian cancer<sup>4</sup>. The risk of developing pancreatic cancer or prostate cancer is also substantially increased<sup>1,3</sup>. Pathogenic variants are now used for the clinical management of carriers through prevention, screening and cancer treatment. However, the interpretation and classification of more than 5,000 individual *BRCA2* variants currently classified on ClinVar<sup>10</sup> as variants of uncertain significance (VUS) has not been possible. These predominantly missense and intronic alterations cannot be effectively utilized for clinical care. Thus, there is a need for large-scale characterization and classification of *BRCA2* variants.

Recently, guidelines from the American College of Medical Genetics (ACMG) and the Association for Molecular Pathology (AMP) that

incorporate multiple sources of evidence, including variant frequency in populations, in silico-sequence-based prediction and functional data, among others, have been utilized by clinical testing groups and the ClinGen *BRCA1* and *BRCA2* (*BRCA1/2*) variant curation expert panel (VCEP) for the classification of variants<sup>9</sup>. However, the classification of variants as pathogenic or likely pathogenic using these models is heavily dependent on the results of functional assays. Although functional data from a homology-directed repair (HDR) assay for missense variants of the *BRCA2* DNA-binding domain (DBD) has been integrated into an ACMG–AMP framework<sup>7,11,12</sup>, this and other low-throughput functional assays have not substantially resolved the VUS issue. By contrast, multiplex assay of variant effect (MAVE) experiments enable the functional characterization of large numbers of variants<sup>13</sup>. Using cell-based selection and deep sequencing to link genotype to phenotype, many variants can be functionally characterized and compared with results of known

<sup>1</sup>Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA. <sup>2</sup>Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN, USA. <sup>3</sup>Graduate School of Biomedical Sciences, Mayo Clinic, Rochester, MN, USA. <sup>4</sup>Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA. <sup>5</sup>Ambry Genetics, Aliso Viejo, CA, USA. <sup>6</sup>Department of Epidemiology, H. Lee Moffitt Cancer Center, Tampa, FL, USA. <sup>7</sup>Department of Oncology, Mayo Clinic, Rochester, MN, USA. <sup>8</sup>Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA. <sup>9</sup>Molecular Oncology Laboratory, Hospital Clínico San Carlos, Instituto de Investigación Sanitaria del Hospital Clínico San Carlos, Madrid, Spain. <sup>10</sup>Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA. <sup>11</sup>Department of Pathology and Laboratory Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA. <sup>12</sup>Marie-Josée and Henry R. Kravis Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York, New York, USA. <sup>13</sup>Department of Statistical Science, Duke University, Raleigh Durham, NC, USA. <sup>27</sup>These authors contributed equally: Huaizhi Huang, Chunling Hu. \*A list of authors and their affiliations appears at the end of the paper.

✉e-mail: Hu.chunling@mayo.edu; couch.fergus@mayo.edu



**Fig. 1 | Schematic overview of the SGE MAVE of all SNVs in the BRCA2 DBD.**

**a**, Design of the SGE experiment and the targeted regions. All possible SNVs were introduced and assessed in exon 15 (E15) to E26 encoding the BRCA2 DBD domain, along with 10 bp of adjacent intronic nucleotides for each exon. E18 and E25 were divided into 2 regions, which resulted in a total of 14 target regions. **b**, Schematic of the SGE workflow. In each target region, a SNV library that contained all possible SNVs was transfected with a corresponding

Cas9–sgRNA construct into HAP1 haploid cells. gDNA was extracted at D5 and D14 after transfection, and the target region was amplified and barcoded for targeted gDNA sequencing. SNV abundance was evaluated and normalized to generate functional scores for all SNVs. An ACMG–AMP classification model was applied to formally classify SNVs based on the results of the MAVE functional assays and other evidence. The schematics in this figure were created using BioRender (credit: C.H., <https://BioRender.com/u10b291>; 2024).

pathogenic and benign standards, as shown for the *BRCA1* and *MSH2* cancer predisposition genes<sup>6,14,15</sup>. MAVE studies of *BRCA2* have been limited to proof-of-principle efforts that have focused on relatively small regions of *BRCA2* (refs. 16–18) and have lacked validation. Here we use a CRISPR–Cas9 knock-in-based saturation genome editing (SGE) approach to evaluate the functional consequences of all possible single-nucleotide variants (SNVs) in *BRCA2* exons 15–26 encoding the BRCA2DBD, which is the sole location of known pathogenic missense variants in this gene. The results are combined with other sources of genetic and clinical evidence in a *BRCA2* ClinGen–ACMG–AMP model for the classification of variants as pathogenic or benign and for the development of a comprehensive reference for the clinical management of individuals with these variants.

## SGE of *BRCA2*

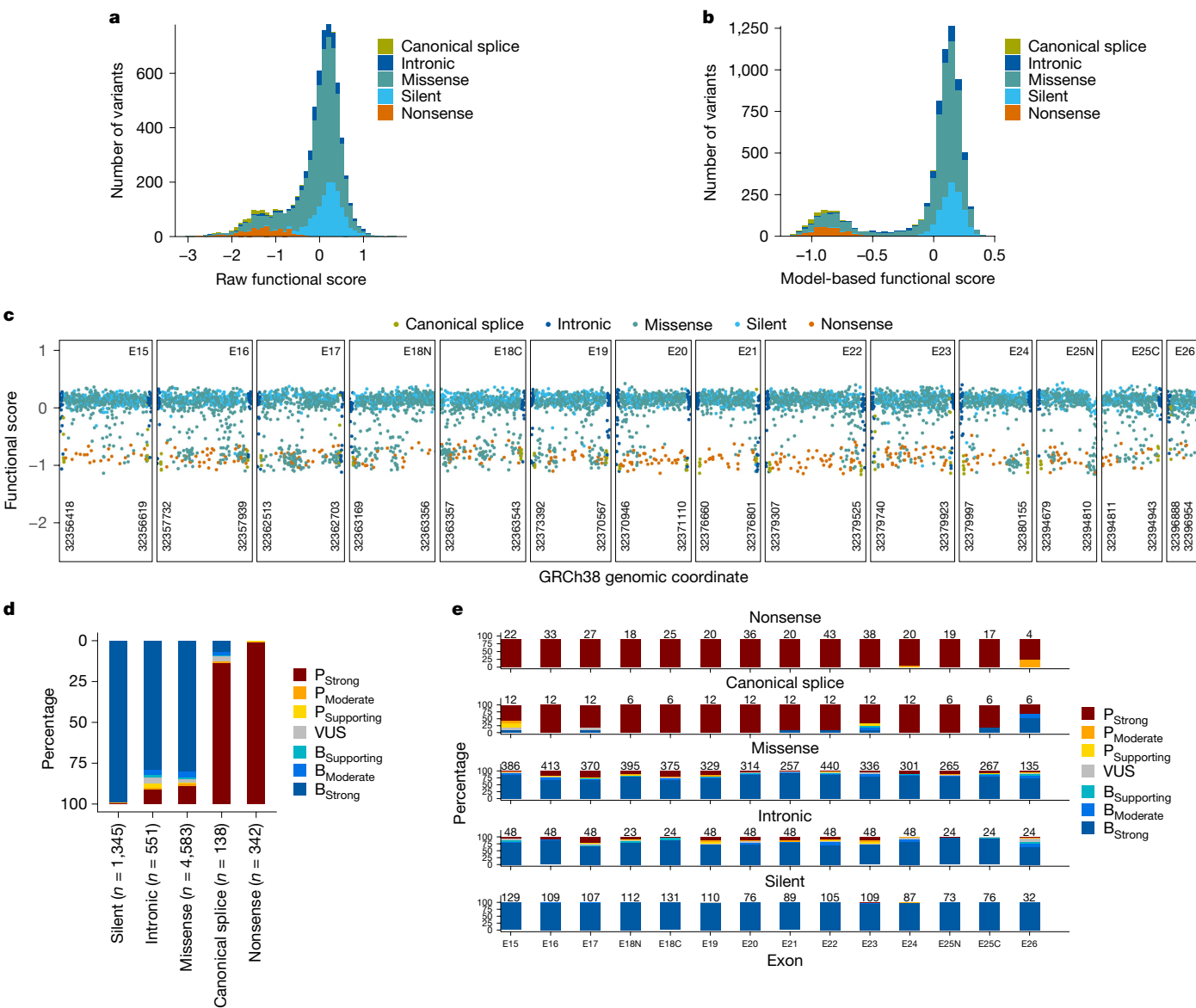
SGE of exons 15–26 of *BRCA2* (MANE transcript ENST00000380152.8; hg38, 32356418–32396954) was performed in the haploid human HAP1 cell line to insert all possible SNVs into the endogenous *BRCA2* gene and to assess the functional impact on cell viability. This approach was based on the essentiality of *BRCA2* in HAP1 cells<sup>19,20</sup> (Supplementary Fig. 1). Individual coding exons together with 10 bp of adjacent intronic nucleotides (exons 18 and 25, which were divided into 2 regions) were selected as SGE target regions (Fig. 1a). Site-saturation mutagenesis libraries that contained 6,959 out of all 6,960 (99.9%) possible SNVs in the 14 target regions were generated by site-directed mutagenesis using NNN-tailed PCR primers (Fig. 1b and Supplementary Tables 1 and 2). An efficient single guide RNA (sgRNA) for each target region was cloned into a sgRNA–Cas9 construct and co-transfected with library plasmids into HAP1 cells in triplicate experiments. gDNA samples from day 0 (D0), D5 and D14 were collected and subjected to amplicon-based deep paired-end sequencing to estimate individual SNV counts at each time point (Fig. 1b and Supplementary Tables 1 and 2). The average sequencing depth for each variant was 3,505 reads for D0 library replicates, 3,948 reads for D5 replicates and 3,810 reads for D14 replicates.

## Functional analysis of variant effects

Replicate-level variant frequencies at each time point (D0, D5 and D14) based on the ratio of variant read counts to total reads were

calculated. Variant position-dependent effects were adjusted using replicate-level generalized additive models with target-region-specific adaptive splines<sup>21</sup>. The log<sub>2</sub>-transformed fold change (LFC) values of D14 to D0 ratios were calculated as the raw functional scores for the 6,959 (99.9%) SNVs (Fig. 2a, Extended Data Table 1 and Supplementary Table 3). A VarCall model<sup>8</sup>, a class of Bayesian hierarchical model that embeds a Gaussian two-component mixture model, was applied to the position-adjusted LFC values of D14 and D0 ratios. Each variant was assigned an indicator of pathogenicity status: deterministically if known and probabilistically if unknown. In detail, nonsense variants were assumed to be pathogenic, whereas silent variants, except for variants with known or predicted splice effects, were assumed to be benign. The method we used adjusted for batch effects by including replicate data of targeted region location and scale random effects and *t*-distributed error terms to allow for outliers. A Markov chain Monte Carlo (MCMC) algorithm<sup>22</sup> was used to obtain adjusted mean functional scores for the 6,959 SNVs (Fig. 2b and Supplementary Table 3). Using a prior probability of pathogenicity of 0.2, based on an AlphaMissense prediction that 22.7% of missense variants in the BRCA2DBD are likely pathogenic, a posterior probability of pathogenicity and a Bayes factor for each variant were calculated. Based on the ClinGen-specified Bayesian interpretation of the ACMG–AMP guidelines<sup>23</sup>, posterior probability thresholds for the functional PS3/BS3 criteria for the following strength of evidence categories were assigned: pathogenic strong ( $P_{\text{Strong}}$ ),  $P_{\text{Moderate}}$  and  $P_{\text{Supporting}}$ ; benign strong ( $B_{\text{Strong}}$ ),  $B_{\text{Moderate}}$  and  $B_{\text{Supporting}}$ ; and VUS (Fig. 2b, Extended Data Table 2, Supplementary Table 3 and Supplementary Fig. 2). Full details of the VarCall model analysis are available in the Supplementary Information.

The VarCall model was validated using 206 known pathogenic and 335 known benign variants, including 70 missense variants from ClinVar with consistent findings from at least two ClinGen-approved testing laboratories or from the *BRCA1/2* VCEP. This analysis showed >99% sensitivity and specificity for pathogenic and benign categories when including nonsense and silent variants, and 94% sensitivity and 95% specificity when comparing with ClinVar missense variants only (Table 1). Similarly, validation using 417 missense variants evaluated using a well-calibrated HDR functional assay achieved 93% sensitivity and 95% specificity (Table 1). Seven out of 122 (5.8%) HDR functionally abnormal missense variants were in the BRCA2 MAVE benign



**Fig. 2 | Functional annotation of BRCA2 SNVs.** **a**, Distribution of raw functional scores of 6,959 SNVs coloured by variant type. **b**, Distribution of adjusted functional scores for all variants from the VarCall model. **c**, Model-based functional score distribution by variant type in each exon. Colour indicates

variant type. **d**, Bar chart illustrating the percentage of each variant type in each of the seven functional categories. Colour indicates functional categories. **e**, Bar chart illustrating the percentage of SNVs by functional category in 14 target regions. Colour indicates functional categories.

categories, whereas 14 out of 295 (4.8%) of HDR functionally normal missense variants were in the MAVE pathogenic categories (Table 1). Finally, 14 pathogenic and 57 benign missense standards identified by the Evidence-based Network for the Interpretation of Germline Mutant Alleles (ENIGMA) consortium and by the ClinGen *BRCA1/2*VCEP produced 93% sensitivity and 96% specificity. Moreover, only 2 out of 57 (3.5%) of the ENIGMA-classified benign missense variants were in the MAVE  $P_{\text{Moderate}}$  category (Table 1 and Supplementary Table 4a–d).

The combined benign ( $B_{\text{Strong}}$ ,  $B_{\text{Moderate}}$  and  $B_{\text{Supporting}}$ ) and combined pathogenic ( $P_{\text{Strong}}$ ,  $P_{\text{Moderate}}$  and  $P_{\text{Supporting}}$ ) categories accounted for 81.6% and 16.6% of variants, respectively, with 1.8% remaining as VUS. Specifically, 5,430 (78%) variants, including 3,661 missense, 1,326 silent, 434 intronic, 9 canonical splice SNVs and 0 nonsense variants, were  $B_{\text{Strong}}$ . By contrast, 1,021 (14.7%), including 502 missense, 339 nonsense, 119 canonical splice, 50 intronic and 11 silent SNVs, were  $P_{\text{Strong}}$  (Fig. 2c and Extended Data Table 2). All nonsense-encoding variants were in the  $P_{\text{Strong}}$ ,  $P_{\text{Moderate}}$  and  $P_{\text{Supporting}}$  categories. Among the missense variants, 3,879 (84.6%) were in the benign categories and 611 (13.3%) were in the

pathogenic categories. Among the 138 variants in +1/2 and –1/2 canonical splice-site positions, 121 (87.7%) were in the pathogenic category, which indicated the presence of aberrant splicing effects. Moreover, 69 (12.5%) intronic SNVs and 13 (1%) silent variants were in the pathogenic categories (Fig. 2c–e, Extended Data Table 2 and Supplementary Table 3). Thus, the MAVE study revealed a large number of variants that may influence RNA splicing. A further 1,329 (98.8%) silent variants were in the benign categories.

### Correlation with DBD architecture

To gain insights into the mechanisms by which the SNVs disrupt BRCA2 activity, the location and influence of the  $P_{\text{Strong}}$  missense-induced changes on protein structure were evaluated.  $P_{\text{Strong}}$  missense variants were enriched in the helical domain and the OB1 domain (15.3% and 16.4%, respectively). These variants were less common in the OB2 and OB3 domains (8.7% and 10.9%, respectively;  $P = 5.9 \times 10^{-6}$ ) and were infrequent (1.8%) in the tower domain (Extended Data Table 3). Moreover, among

**Table 1 | Validation of the BRCA2 functional assay**

		BRCA2 MAVE functional categories <sup>a</sup>							Sensitivity <sup>b</sup>	Specificity <sup>c</sup>
		P <sub>Strong</sub>	P <sub>Moderate</sub>	P <sub>Supporting</sub>	VUS	B <sub>Supporting</sub>	B <sub>Moderate</sub>	B <sub>Strong</sub>		
ClinVar standards	Pathogenic (n=206)	199 (97%)	2 (1%)	3 (1.5%)	1 (0.5%)	0	0	1 (0.5%)	99%	99%
	Benign (n=335)	2 (0.6%)	1 (0.3%)	0	1 (0.3%)	0	2 (0.6%)	329 (98.2%)		
ClinVar missense standards	Pathogenic (n=33)	28 (85%)	0	3 (9%)	1 (3%)	0	0	1 (3%)	94%	95%
	Benign (n=37)	1 (3%)	1 (3%)	0	0	0	2 (5.5%)	33 (89%)		
HDR assay for missense variants <sup>d</sup>	HDR abnormal (n=122)	103 (84%)	7 (6%)	3 (2.5%)	2 (1.6%)	0	1 (0.8%)	6 (5%)	93%	95%
	HDR normal (n=295)	7 (2.4%)	4 (1.4%)	3 (1.0%)	5 (1.7%)	6 (2.0%)	10 (3.4%)	260 (88%)		
ENIGMA, ClinGen and VCEP missense standards	Pathogenic (n=14)	12 (86%)	0	1 (7%)	1 (7%)	0	0	0	93%	96%
	Benign (n=57)	0	2 (3.5%)	0	1 (1.8%)	0	2 (3.5%)	52 (91%)		

<sup>a</sup>MAVE functional categories were based on the VarCall model.

<sup>b</sup>Sensitivity was calculated as the proportion of pathogenic standards identified as P<sub>Strong</sub>, P<sub>Moderate</sub> or P<sub>Supporting</sub> variants.

<sup>c</sup>Specificity was calculated as the proportion of benign standards identified as B<sub>Strong</sub>, B<sub>Moderate</sub>, B<sub>Supporting</sub> or VUS.

<sup>d</sup>Standardized V-C8 BRCA2<sup>-/-</sup> HDR assay<sup>7</sup>.

Note that all variants observed to influence splicing were excluded in the validation datasets.

the 423 P<sub>Strong</sub> missense variants, 154 (36.4%) were in the helical domain, 125 (29.6%) in the OB1 domain and 83 (19.6%) in the OB3 domain. By contrast, only 45 (10.6%) were in the OB2 domain and 13 (3.1%) in the tower domain (Extended Data Table 3 and Extended Data Fig. 1a). These findings are consistent with the HDR assay results for 462 DBD missense variants, which showed an enrichment for functionally abnormal missense alterations in the helical, OB1 and OB3 domains<sup>7</sup>. Identification of 13 P<sub>Strong</sub> missense variants in the tower domain, in which no pathogenic or non-functional missense variants were previously known, confirmed that this domain is required for normal BRCA2 function and established that it is not a cold spot for inactivating or potentially cancer-predisposing variants. P<sub>Strong</sub> missense alterations were observed in 26 out of 50 (52%) DSS1-interacting residues and in 2 out of 17 (12%) single-stranded DNA-interacting residues. This result indicates that DSS1-mediated stability is important for BRCA2 homologous recombination repair activity. It was also noted that 261 out of 423 (62%) P<sub>Strong</sub> missense alterations resulted in changes in amino acid charge or the loss or gain of proline residues (Extended Data Fig. 1a and Supplementary Table 3). Many residues in the BRCA2 DBD are highly conserved from pufferfish to *Homo sapiens*. At least one P<sub>Strong</sub> missense variant was observed in 103 (48.6%) perfectly conserved residues, with 45 (44%) of these in the helical domain and 30 (32%) in the OB1 domain. P<sub>Strong</sub> variants were also observed in 71 (31.6%) highly conserved residues and 39 (15.5%) in poorly conserved residues (Extended Data Table 3 and Extended Data Fig. 1b). Approximately 75% of the residues with P<sub>Strong</sub> variants were located in  $\alpha$ -helices and  $\beta$ -sheet structures needed to maintain essential three-dimensional folding of BRCA2 (Extended Data Fig. 1c–g).

### Comparisons with functional predictors

Several functional assays that assessed the influence of BRCA2 missense variants on protein function are used by the ClinGen BRCA1/2 VCEP for clinical classification of variants. BRCA2 MAVE data were strongly correlated with results from a cell-based HDR assay ( $P = 1.6 \times 10^{-52}$ ; Table 1 and Fig. 3a), and effectively distinguished between class 4–5 (functionally abnormal) and either class 3 (uncertain) ( $P = 4.8 \times 10^{-7}$ ) or class 1–2 (functionally normal) ( $P = 2.0 \times 10^{-11}$ ) variants from an olaparib PARP inhibitor response assay<sup>24</sup> (Fig. 3b). The MAVE data also effectively discriminated between non-functional and functional ( $P = 3.4 \times 10^{-8}$ ) or uncertain ( $P = 1.3 \times 10^{-6}$ ) variants in an endogenously targeted prime-editing study of exons 15 and 17 (ref. 16) (Fig. 3c) and between non-functional and functional ( $P = 1.1 \times 10^{-4}$ ) missense variants in a small embryonic stem cell complementation assay<sup>25</sup> (Fig. 3d). Thus, data from the MAVE analysis are highly consistent with results from several other small-scale functional assays (Fig. 3a–d and Supplementary Table 5a–d).

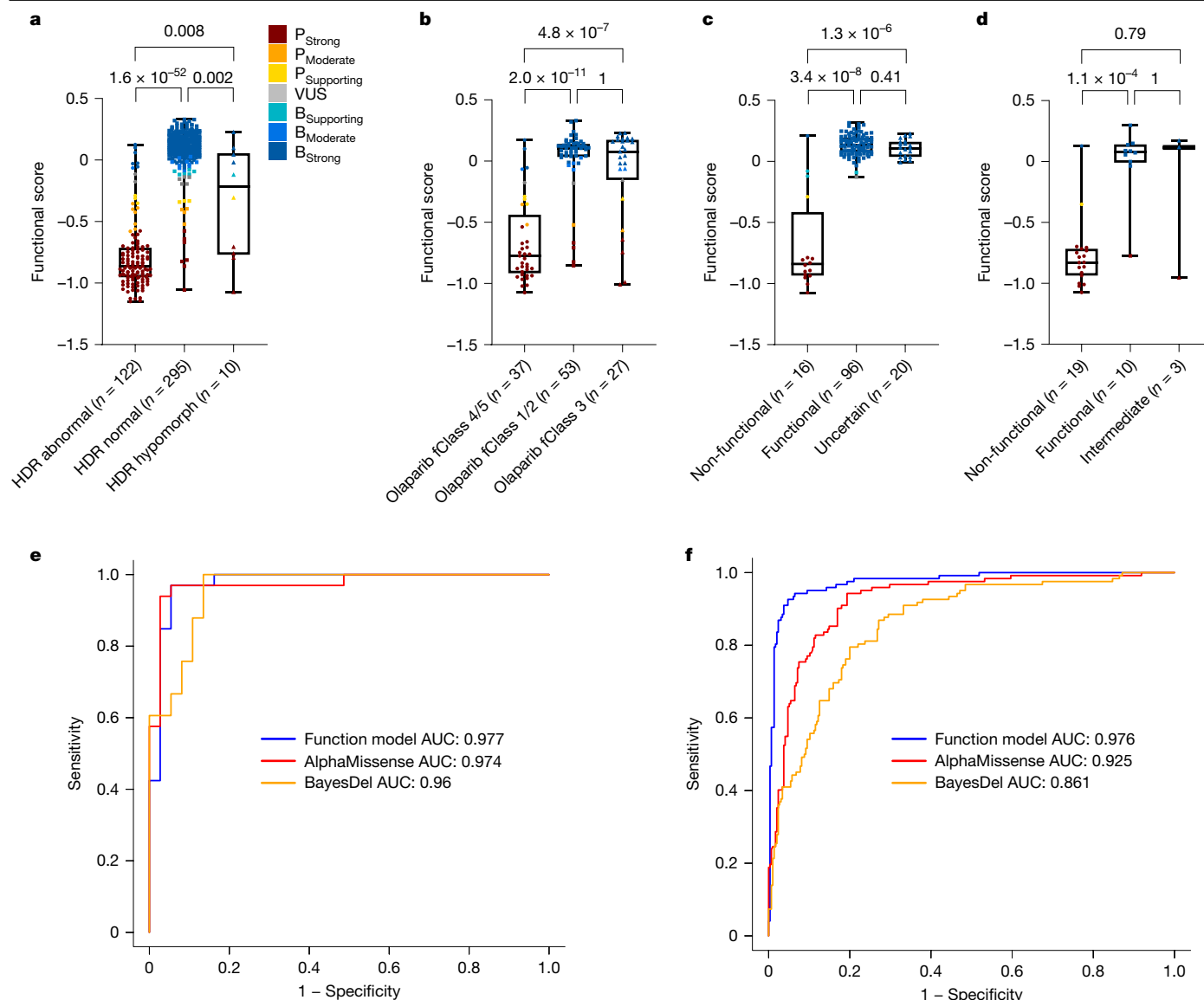
Notably, the MAVE data showed that class 3, uncertain or intermediate variants from these assays are predominantly B<sub>Strong</sub>, B<sub>Moderate</sub> or B<sub>Supporting</sub>.

Next, comparisons between the MAVE results and in silico prediction methods were performed using the MAVE P<sub>Strong</sub>, P<sub>Moderate</sub> and P<sub>Supporting</sub> categories and the B<sub>Strong</sub>, B<sub>Moderate</sub> and B<sub>Supporting</sub> categories. The Align-GVGD model class C65 (likely non-functional) category<sup>8,26</sup> demonstrated moderate sensitivity (41%) and high specificity (91%) compared with the MAVE results. The AlphaMissense deep-learning model<sup>27</sup> also produced moderate sensitivity (74%) and specificity (84%) for the likely pathogenic score threshold ( $>0.564$ ). The BayesDel predictor, which is currently used by the ClinGen BRCA1/2 VCEP for the curation of BRCA1/2 variants, produced moderate sensitivity (73%) and specificity (83%) when using the ClinGen-specified P<sub>Strong</sub> BayesDel predictor, but moderate sensitivity (43%) and high specificity (95%) when using the BRCA1/2 VCEP pathogenic threshold (Extended Data Table 4 and Supplementary Table 6a,b). The BRCA2 MAVE, AlphaMissense and BayesDel data produced area under the receiver/operator curve (AUC) values of  $>0.96$  based on 70 ClinVar-classified missense variants ( $n = 70$ ). However, when comparing with HDR-characterized variants, the AUC for BRCA2 MAVE data (0.98) was better than for AlphaMissense (0.93) or BayesDel (0.86) data (Fig. 3e,f).

### Cancer risks for variant categories

To understand the contributions of the characterized variants to cancer risk, associations between combined variants in functional P<sub>Strong</sub>, P<sub>Moderate</sub> and P<sub>Supporting</sub> categories and B<sub>Strong</sub>, B<sub>Moderate</sub> and B<sub>Supporting</sub> categories and breast cancer and ovarian cancer were evaluated in case-control studies. Specifically, the frequencies of missense variants in breast cancer cases in women who received hereditary cancer genetic testing by Ambry Genetics from 2012 to 2021 and the frequency in reference controls (women) from gnomAD v.4, excluding the UK Biobank<sup>28</sup>, were compared according to functional category. The P<sub>Strong</sub>-only missense variants (odds ratio (OR) = 4.45, 95% confidence interval (CI) = 3.30–6.13) and the combined P<sub>Strong</sub>, P<sub>Moderate</sub> and P<sub>Supporting</sub> missense variants (OR = 4.34, 95% CI = 3.27–5.85) produced high risks (OR  $> 4.0$ ) for breast cancer. By contrast, B<sub>Strong</sub>, B<sub>Moderate</sub> and B<sub>Supporting</sub> missense variants were not associated with clinically relevant (OR  $> 2$ ) increased breast cancer risk (OR = 0.78, 95% CI = 0.71–0.85) (Table 2). This P<sub>Strong</sub>, P<sub>Moderate</sub> and P<sub>Supporting</sub> missense OR was attenuated compared with P<sub>Strong</sub>, P<sub>Moderate</sub> and P<sub>Supporting</sub> nonsense variants (OR = 5.65, 95% CI = 3.98–8.28). Pathogenic missense variants designated by the ENIGMA expert panel (OR = 5.9, 95% CI = 3.08–12.74) and DBD protein-truncating variants (OR = 6.68, 95% CI = 5.19–8.74) (Table 2) also had higher ORs than the P<sub>Strong</sub>, P<sub>Moderate</sub> and P<sub>Supporting</sub> missense variants. However, when restricting to variants with a





**Fig. 3 | Comparison of *BRCA2*MAVE data with data from functional assays and in silico predictors. a–d**, Boxplots showing functional scores of SNVs encoding missense variants compared with a *BRCA2*<sup>-/-</sup> V-C8 HDR assay (**a**), a DLD1 *BRCA2*<sup>-/-</sup> olaparib sensitivity assay (**b**), a prime-editing-based haploid cell-survival assay (**c**) and a mouse *Brca2*<sup>-/-</sup> embryonic stem cell complementation assay (**d**). The numbers of variants of each type resulting from the individual assays are shown. Functionally abnormal variants have significantly lower functional scores than functionally normal variants in **a** ( $P = 1.6 \times 10^{-52}$ ), **b** ( $P = 2.0 \times 10^{-11}$ ), **c** ( $P = 3.4 \times 10^{-8}$ ) and **d** ( $P = 1.1 \times 10^{-4}$ ), using two-sided Mann–Whitney–Wilcoxon tests. *P* values

for all comparisons are shown. Boxes represent the interquartile range, the horizontal line is the median functional score, and whiskers show maximum and minimum values. Variants are shown as points and coloured by the functional strength of the evidence category. **e**, Comparison of the AUC values between MAVE and two in silico predictors (AlphaMissense and BayesDel) using ClinVar-classified missense standards ( $n = 70$ ). **f**, Comparison of the AUC values between MAVE and two in silico predictors (AlphaMissense and BayesDel) using missense variants characterized using a well-calibrated HDR assay ( $n = 417$ ).

posterior probability of pathogenicity  $\geq 95\%$  within the  $P_{\text{Strong}}$  category, 76% (380 out of 502) of missense variants had risks similar to nonsense variants (OR = 5.09, 95% CI = 3.62–7.35). The risks were further increased in the 60% (299 out of 502) of variants with a posterior probability of pathogenicity  $\geq 99\%$  (OR = 5.38, 95% CI = 3.69–8.15). Moderate (OR = 2–4) to high risks of breast cancer were also observed using the non-cancer gnomAD v.2.1 and v.3.1 control reference dataset in place of the gnomAD v.4 dataset (Supplementary Table 7).  $P_{\text{Strong}}$ ,  $P_{\text{Moderate}}$  and  $P_{\text{Supporting}}$  missense variants in women who identified as African American (OR = 3.34, 95% CI = 1.59–7.13) also showed moderate-to-high risks of breast cancer (Supplementary Table 7). Additional analyses using case-control data from the CARRIERS and BRIDGES population-based breast cancer studies<sup>2,29</sup> and from the UK Biobank (www.ukbiobank.ac.uk) produced similar findings. However, the ORs were attenuated owing

to the population-based nature of the cases and controls (Table 2 and Supplementary Table 7). In the population-based studies, it was notable that the variants with posterior probability of pathogenicity  $\geq 99\%$  (299 out of 502; 60%) in the  $P_{\text{Strong}}$  category were associated with high risks of breast cancer (OR = 4.19, 95% CI = 2.23–7.89).  $P_{\text{Strong}}$ ,  $P_{\text{Moderate}}$  and  $P_{\text{Supporting}}$  missense variants were also associated with substantially increased risks of ovarian cancer (OR = 7.76, 95% CI = 5.34–11.29), which were attenuated relative to nonsense variants (Table 2). However, the  $P_{\text{Strong}}$  variants with posterior probability of pathogenicity  $\geq 95\%$  (OR = 9.32, 95% CI = 5.98–14.65) had similar risks of ovarian cancer to the nonsense variants (OR = 9.13, 95% CI = 5.74–14.63). Lifetime risks for breast cancer and ovarian cancer were estimated using ORs from the current study and from rates of disease reported by the Surveillance, Epidemiology, and End Results (SEER) registry. The  $P_{\text{Strong}}$  missense variants conferred

**Table 2 | Associations between variants in the BRCA2 DBD and risk of breast cancer and ovarian cancer**

	Case		Control		OR (95% CI) <sup>f</sup>	Pvalue <sup>f</sup>
	No. of variants	No. tested	No. of variants	No. tested		
Breast cancer clinical testing cohort <sup>a</sup>						
MAVE missense P <sub>Strong</sub> <sup>b</sup>	236	197,659	53	181,964	4.45 (3.30–6.13)	1.85×10 <sup>-20</sup>
MAVE missense P <sub>Strong</sub>						
≥99% Probability	168	197,659	32	181,964	5.38 (3.69–8.15)	7.41×10 <sup>-17</sup>
≥95% Probability	201	197,659	40	181,964	5.09 (3.62–7.35)	1.71×10 <sup>-19</sup>
MAVE missense P <sub>SMS</sub>	261	197,659	60	181,964	4.34 (3.27–5.85)	4.35×10 <sup>-22</sup>
MAVE missense B <sub>Strong</sub>	920	197,659	1,155	181,964	0.78 (0.71–0.85)	7.14×10 <sup>-8</sup>
MAVE missense B <sub>SMS</sub>	981	197,659	1,232	181,964	0.78 (0.71–0.85)	1.49×10 <sup>-8</sup>
ACMG missense P/LP	210	197,659	29	181,964	6.96 (4.77–10.56)	5.87×10 <sup>-21</sup>
ACMG missense B/LB	970	197,659	1,234	181,964	0.77 (0.70–0.83)	2.33×10 <sup>-9</sup>
ENIGMA missense P	58	197,659	8	181,964	5.9 (3.08–12.74)	1.32×10 <sup>-6</sup>
MAVE P <sub>SMS</sub> (no ENIGMA P) <sup>c</sup>	203	197,659	52	181,964	4.03 (2.96–5.61)	5.92×10 <sup>-17</sup>
MAVE nonsense	205	197,659	34	181,964	5.65 (3.98–8.28)	1.23×10 <sup>-19</sup>
Protein truncating <sup>d</sup>	469	197,659	63	181,964	6.68 (5.19–8.74)	1.91×10 <sup>-44</sup>
Breast cancer population-based cohort <sup>a</sup>						
MAVE missense P <sub>Strong</sub>	63	81,073	26	83,247	2.49 (1.55–3.94)	4.67×10 <sup>-5</sup>
MAVE missense P <sub>Strong</sub>						
≥99% Probability	49	81,073	12	83,247	4.19 (2.23–7.89)	9.77×10 <sup>-7</sup>
≥98% Probability	54	81,073	16	83,247	3.47 (1.98–6.06)	3.13×10 <sup>-6</sup>
≥95% Probability	57	81,073	20	83,247	2.93 (1.76–4.87)	1.70×10 <sup>-5</sup>
MAVE missense P <sub>SMS</sub>	76	81,073	42	83,247	1.86 (1.26–2.72)	1.19×10 <sup>-3</sup>
MAVE missense B <sub>Strong</sub>	503	81,073	508	83,247	1.02 (0.90–1.15)	0.801
MAVE missense B <sub>SMS</sub>	544	81,073	551	83,247	1.01 (0.90–1.14)	0.832
ACMG missense P/LP	53	81,073	16	83,247	3.4 (1.93–6.17)	4.93×10 <sup>-6</sup>
ACMG missense B/LB	542	81,073	554	83,247	1 (0.89–1.13)	0.952
ENIGMA missense P	26	81,073	5	83,247	5.34 (2.03–14.59)	9.52×10 <sup>-5</sup>
MAVE nonsense	90	81,073	15	83,247	6.16 (3.58–11.15)	7.05×10 <sup>-15</sup>
Protein truncating	179	81,073	25	83,247	7.36 (4.82–11.43)	4.87×10 <sup>-31</sup>
Ovarian cancer clinical testing cohort <sup>a</sup>						
MAVE missense P <sub>Strong</sub>	51	24,981	53	181,964	7.91 (5.32–11.79)	2.45×10 <sup>-23</sup>
MAVE missense P <sub>Strong</sub>						
≥99% Probability	36	24,981	32	181,964	9.93 (6.03–15.56)	3.24×10 <sup>-19</sup>
≥95% Probability	44	24,981	40	181,964	9.32 (5.98–14.65)	1.05×10 <sup>-22</sup>
MAVE missense P <sub>SMS</sub>	57	24,981	60	181,964	7.76 (5.34–11.29)	1.39×10 <sup>-25</sup>
MAVE missense B <sub>Strong</sub>	135	24,981	1,155	181,964	0.91 (0.76–1.09)	0.362
MAVE missense B <sub>SMS</sub>	148	24,981	1,232	181,964	0.93 (0.78–1.10)	0.463
ACMG missense P/LP	43	24,981	29	181,964	11.69 (7.27–19.17)	2.28×10 <sup>-22</sup>
ACMG missense B/LB	147	24,981	1,234	181,964	0.92 (0.77–1.09)	0.405
ENIGMA missense P	11	24,981	8	181,964	8.27 (3.47–20.03)	3.17×10 <sup>-6</sup>
MAVE P <sub>SMS</sub> (no ENIGMA P) <sup>c</sup>	46	24,981	52	181,964	7.65 (5.05–11.59)	6.02×10 <sup>-21</sup>
MAVE nonsense P	40	24,981	34	181,964	9.13 (5.74–14.63)	9.93×10 <sup>-20</sup>
Protein truncating	85	24,981	63	181,964	9.35 (6.79–12.93)	1.51×10 <sup>-40</sup>

<sup>a</sup>The clinical testing cohort included cases from patients with breast cancer and from patients with ovarian cancer (all had cancer genetic testing by Ambry Genetics), and controls from gnomAD v.4 (women), excluding the UK Biobank.

<sup>b</sup>MAVE missense variants excluded variants enriched in subpopulation (>75%) or with an allele frequency of >0.001 in any population.

<sup>c</sup>MAVE missense pathogenic category variants (P<sub>Strong</sub>, P<sub>Moderate</sub> and P<sub>Supporting</sub> (P<sub>SMS</sub>)) without ENIGMA-designated missense pathogenic variants.

<sup>d</sup>Protein-truncating variants included frameshift and nonsense variants in the BRCA2 DBD.

<sup>e</sup>Cases and controls were from CARRIERS<sup>2</sup> and BRIDGES<sup>29</sup> studies.

<sup>f</sup>For the clinical cohort, ORs were calculated using weighted logistic regression with control populations weighted for the relative frequency of different races and ethnicities in the cases; Pvalues were adjusted for multiple testing using the Benjamini–Hochberg method. For the population-based cohort, ORs were calculated using Fisher’s exact tests.

Note that ovarian cancer included malignant epithelial tumours of the ovary and fallopian tube.

an estimated lifetime risk of 41% and 11% up to age 80 years for breast cancer and ovarian cancer, respectively, which was similar to the 52% and 12% risks, respectively, for DBD protein-truncating variants (Extended Data Fig. 2). All data shown are provided with the explicit written consent of the study participants following approval from the institutional review boards.

## Clinical classification of SNVs

Functional data for SNVs must be integrated into classification models to determine the clinical relevance of each variant. Here the ClinGen *BRCA1/2* VCEP classification framework, adapted for point scoring<sup>30</sup>, was applied to the MAVE SNVs. As noted above, thresholds for  $P_{\text{Strong}}$ ,  $P_{\text{Moderate}}$ ,  $P_{\text{Supporting}}$ ,  $B_{\text{Strong}}$ ,  $B_{\text{Moderate}}$  and  $B_{\text{Supporting}}$  functional categories under the PS3/BS3 code were determined on the basis of the Bayesian interpretation of the ACMG–AMP guidelines. The  $P_{\text{Strong}}$  and  $B_{\text{Strong}}$  categories were capped at +4 or –4 points to avoid classification by functional evidence alone.  $P_{\text{Moderate}}$  and  $B_{\text{Moderate}}$  were assigned +2 or –2 points, and the  $P_{\text{Supporting}}$  and  $B_{\text{Supporting}}$  categories were assigned +1 or –1 points (Fig. 4a and Supplementary Table 8). The points for each variant derived from each VCEP code, including the function-based PS3/BS3 code, were combined and variants were classified as pathogenic (P) ( $\geq 10$  points), likely pathogenic (LP) (6 to 9 points), uncertain/VUS (–1 to 5 points), likely benign (LB) (–6 to –2 points) or benign (B) ( $\leq -7$  points) (Supplementary Table 8). Overall, among all the SNVs, 5,566 were classified as B/LB, 785 as P/LP and 608 as VUS. Among the nonsense SNVs, 3 were classified as LP and 339 were classified as pathogenic. Among the 4,583 missense SNVs, 261 were classified as LP/P, 3,786 were LB/B and 536 remained as VUS when using the *BRCA1/2* VCEP rules (Fig. 4a and Extended Data Table 5). Notably, the LP/P-classified missense variants were associated with a high risk of breast cancer (OR = 6.96, 95% CI = 4.77–10.56), whereas the LB/B-classified missense variants were not associated with increased breast cancer risk (OR = 0.77, 95% CI = 0.70–0.83) (Table 2). Among the 138 canonical splice sites, 23 were classified as LP and 105 as pathogenic. Overall, 43 out of 48 canonical splice-site variants with available mRNA assay data and PVS1 (RNA)-weighted points were classified as  $P_{\text{Strong}}$  and  $P_{\text{Moderate}}$  in the MAVE assay (Supplementary Table 8). Four canonical splice-site variants in the +2 position that were designated  $B_{\text{Strong}}$  and  $B_{\text{Moderate}}$  in the *BRCA2* functional analysis were attributed PVS1<sub>NA</sub> by the *BRCA1/2* VCEP and 0 points and were classified as LB (Supplementary Table 8).

To evaluate the impact of results from the *BRCA2* functional study on variant classification, comparisons were made with the classification results from both ClinVar and ENIGMA. Of the 5,589 SNVs classified as B/LB, ClinVar and ENIGMA accounted for 724 (13.0%) and the *BRCA2* functional study accounted exclusively for 4,865 (87.0%). Among 793 classified as P/LP, ClinVar and ENIGMA accounted for 396 (49.9%) and the functional study accounted exclusively for 397 (50.1%) (Fig. 4b, Extended Data Table 5 and Supplementary Table 8). Moreover, of the 322 SNVs with discordant classifications in ClinVar (P/LP versus VUS, or B/LB versus VUS), 290 (90.0%) were classified as B/LB or P/LP and 32 remained as VUS when incorporating the *BRCA2* functional data into the *BRCA1/2* VCEP model. In an effort to compare results from the current functional study and a parallel mouse embryonic stem cell survival assay<sup>31</sup>, the functional data from both studies were incorporated into the *BRCA1/2* VCEP classification model. Concordance was 87%, with only 1% ( $n = 60$ ) of variants assigned to conflicting classification categories (Extended Data Fig. 3). Notably, classifications in ClinVar for 5 of these 60 conflicting SNVs (c.8168A>C, c.8976A>C, c.8982A>T, c.8995C>G and c.9005A>G) and HDR results for 10 out of 12 missense SNVs evaluated (c.7634T>G, c.7679T>C, c.7796A>C, c.7823C>T, c.7904A>G, c.8060T>G, c.8168A>C, c.8588A>G, c.8594T>C, c.9272T>G but not c.7967T>G or c.8300C>T) (Supplementary Table 3) were consistent with results from the current MAVE study.

## Phenotypic characteristics for SNVs

The mean age of breast cancer diagnosis among women with  $P_{\text{Strong}}$  or  $P_{\text{Moderate}}$  SNVs from the population-based CARRIERS study was 56 years, which was significantly younger than the mean age at diagnosis of 61 years for women with  $B_{\text{Strong}}$  or  $B_{\text{Moderate}}$  SNVs ( $P < 0.001$ ) (Extended Data Table 6). A similar significant difference was observed in the

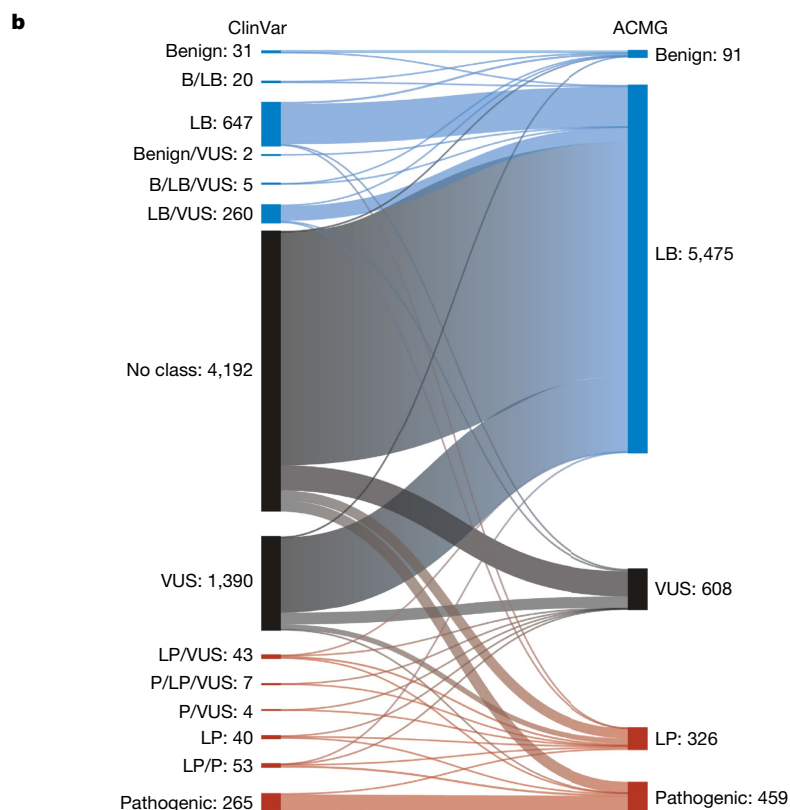
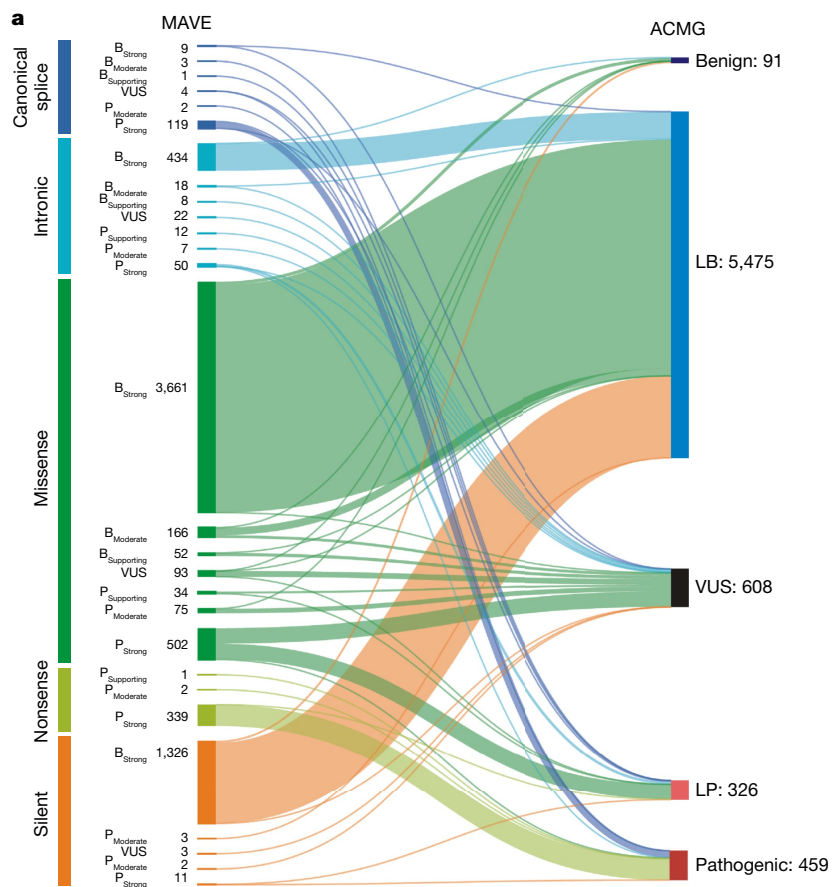
clinical testing cohort ( $P < 0.001$ ), even though the clinical cohort was enriched for onset disease at a young age (Extended Data Table 6). Similarly, a significant difference was observed for missense SNVs in the clinical cohort ( $P = 0.039$ ) and for SNVs classified as LP/P compared with LB/B using the *BRCA1/2* VCEP classification model (Supplementary Table 9). A significant difference in family history of breast cancer, defined as any first-degree or second-degree relative with disease, was observed for individuals with  $P_{\text{Strong}}$  or  $P_{\text{Moderate}}$  SNVs compared with  $B_{\text{Strong}}$  or  $B_{\text{Moderate}}$  SNVs ( $P < 0.001$ ) and SNVs classified as LP/P and LB/B in the clinical testing cohort. Similar trends were observed in the population-based study (Extended Data Table 6 and Supplementary Table 9).

Loss of heterozygosity (LOH) of *BRCA2* in tumours was evaluated to assess whether  $P_{\text{Strong}}$  and  $P_{\text{Moderate}}$  variants in *BRCA2* may be drivers of tumour development. LOH at *BRCA2* was evaluated in 50,000 breast tumour, ovarian tumour, prostate tumour and pancreatic tumour samples with >40% tumour content and had been sequenced using a cancer gene panel in the integrated mutation profiling of actionable cancer targets (IMPACT) study<sup>32</sup>. LOH was detected in 22 out of 26 (85%) tumours with *BRCA2*  $P_{\text{Strong}}$  SNVs and in 23 out of 29 (79%) tumours associated with  $P_{\text{Strong}}$  and  $P_{\text{Moderate}}$  SNVs (Extended Data Table 7 and Supplementary Table 10). By contrast, LOH was observed in 58 out of 233 (25%) tumours with  $B_{\text{Strong}}$  variants, which was significantly different to the inactivating variants ( $P = 3.1 \times 10^{-9}$ ) (Extended Data Table 7 and Supplementary Table 10). Thus,  $P_{\text{Strong}}$  and  $P_{\text{Moderate}}$  variants seem to enrich for loss of the wild-type *BRCA2* allele and inactivation of *BRCA2*, a result consistent with a role for these SNVs as drivers of tumour formation.

## Discussion

The functional evaluation of variants in *BRCA2* has been an active area of research. This is because of the high risks of several cancers (breast, ovarian, prostate, pancreatic and cholangiocarcinoma) associated with inactivating variants in *BRCA2*, the large number of VUS in *BRCA2* that may only be clinically classified after the inclusion of functional evidence and the insights into *BRCA2* function and biology that can be gained from such studies. However, so far, only 557 missense variants in the *BRCA2* DBD have been evaluated through well-established functional assays<sup>7,8,11,12,24,25</sup>. The substantial number of identified variants with clinical uncertainty has necessitated more rapid functional characterization. Here a SGE study of human haploid cells was used to functionally evaluate the effects of all *BRCA2* SNVs in the exons encoding the *BRCA2* DBD pathogenic missense variant hotspot on *BRCA2* activity, as measured by cell viability. Functional scores were obtained for 6,959 SNVs (99% of all possible SNVs) from 12 coding exons and 23 flanking intronic sequences. Although more than 600 DBD SNVs have previously been evaluated using other functional assays, the current study established a sequence–function map for nearly all possible SNVs in the *BRCA2* DBD. Variants were each assigned a probability of pathogenicity in a Bayesian VarCall model. Thresholds for the PS3/BS3 rule (variant effect on protein function) from the ClinGen–ACMG–AMP variant classification guidelines<sup>23</sup>, based on the Bayesian interpretation of these rules, placed variants into seven categories related to the strength of evidence of pathogenicity. The direct assignment of a posterior probability and a strength of evidence of pathogenicity for each variant in the functional study represents a significant advancement in characterization of variants in *BRCA2* (similar to a previous study<sup>14</sup> of missense variants in the RING domain of *BRCA1*). That is, previous approaches focused on the sensitivity and specificity of the functional assay<sup>33</sup> and the grouping of variants into non-functional, uncertain and functional categories, whereas in the VarCall approach, each individual variant is independently assessed.

Notably, the functional data do not directly determine the clinical relevance of any variants. This can currently only be achieved by incorporating the functional data into ClinGen–ACMG–AMP classification



**Fig. 4 | Clinical classification of *BRCA2* SNVs. a**, Sankey plot illustrating the clinical classification of SNVs after integration of *BRCA2* MAVE functional data into the ClinGen *BRCA1/2* VCEP ACMG-AMP classification framework. The numbers of SNVs for functional categories in each variant type are shown in the left-hand MAVE column. The numbers of variants in the classification category are shown in the right-hand ACMG column. **b**, Sankey plot illustrating the changes in variant classification status in ClinVar before (left) and after (right) incorporating *BRCA2* MAVE functional results into the *BRCA1/2* VCEP ACMG-AMP classification framework.

models. For this purpose, the functional data under the ClinGen-ACMG-AMP PS3/BS3 rule was capped at +4 points for pathogenicity and -4 for benign level based on the P<sub>Strong</sub> and B<sub>Strong</sub> categories, respectively, to avoid classification of variants with functional evidence

alone (+6 points is sufficient for a LP classification). These PS3/BS3 points were then combined with point scores from other genetic and clinical data for variant classification under the ClinGen-ACMG-AMP *BRCA1/2* VCEP rules. The outcome was that 261 missense SNVs and 785



of all SNVs were classified as P/LP, whereas 3,786 missense and 5,566 of all SNVs were classified as B/LB. Although 536 missense and 608 SNVs remained as VUS, it seems likely that many of these variants will be classified as P/LP or B/LB in the future following the addition of data from other sources to the now available functional data.

Although 1,120 *BRCA2* DBD SNVs had previously been classified by ClinVar as P/LP ( $n = 396$ ) or B/LB ( $n = 724$ ), the functional data increased this number to 6,382 classified SNVs. Thus, the functional study accounted for 82% of all classifications, which represents a substantial improvement for VUS and is anticipated to have important implications for the many carriers of these germline variants. Individuals with P/LP variants may now qualify for enhanced mammography and MRI screening and for surgical prevention through prophylactic mastectomy or oophorectomy to reduce the possibility of cancer development. Furthermore, carriers may be eligible for treatment of breast, ovarian and potentially other cancers, such as prostate and pancreatic, with PARP inhibitors in the adjuvant and/or metastatic setting. In addition, family members of those with P/LP variants may benefit from testing and preventive measures and screening before the onset of cancer. Moreover, those with B/LB variants can benefit from the knowledge that the variant that they carry is probably not a cancer predisposing allele.

The functional study was validated through three independent datasets: ClinVar pathogenic and benign variants; orthogonal HDR assay functionally abnormal and normal variants, and nonsense and silent variants. Overall, the VarCall model resulted in only approximately 5% miscategorization of the standards in each of the validation sets. Although this result raises the possibility of error in the ACMG-AMP-ClinGen clinical classification of *BRCA2* SNVs, the need for multiple sources of evidence for formal classification minimizes the likelihood of a misclassification. However, as other functional studies are completed, consistency between the studies for each variant will be useful for overcoming any study-specific errors. Indeed, 87% concordance for variant classification using the ClinGen-ACMG *BRCA1/2* VCEP model was observed between the current *BRCA2* functional study and a parallel *BRCA2* DBD MAVE study of cell survival in embryonic stem cells<sup>31</sup>. In a separate effort to further validate the MAVE findings, the IMPACT tumour sequencing dataset from the Memorial Sloan Kettering Cancer Center was used to assess whether functionally pathogenic variants displayed LOH at the *BRCA2* locus, as should be observed for a driver mutation<sup>32</sup>. Indeed, 85% of  $P_{\text{Strong}}$  variants, but only 25% of  $B_{\text{Strong}}$  variants, identified in the IMPACT study showed *BRCA2* LOH, which indicated strong enrichment for loss of the wild-type second *BRCA2* allele in the tumours with functionally  $P_{\text{Strong}}$  SNVs.

Case-control association analyses confirmed that  $P_{\text{Strong}}$ -only SNVs and combined  $P_{\text{Strong}}$ ,  $P_{\text{Moderate}}$  and  $P_{\text{Supporting}}$  SNVs were associated with an increased risk of breast cancer in a clinical cohort of high-risk individuals, in individuals in population-based studies and in African American individuals. These SNVs were also associated with an increased risk of ovarian cancer in a clinical high-risk population. Although publicly available reference controls were used for the clinical high-risk analysis, the consistency of the findings confirmed the increased risk of developing cancer. The similar effects observed in the various populations suggest that these variants will confer increased risk in all populations. It was noted that the  $P_{\text{Strong}}$ -only and  $P_{\text{Strong}}$ ,  $P_{\text{Moderate}}$  and  $P_{\text{Supporting}}$  missense SNVs were associated with lower risks than nonsense variants for both breast cancer and ovarian cancer. However, 380 out of 502 (76%) variants with posterior probabilities of pathogenicity  $\geq 95\%$  in the clinical cohort and 299 out of 502 (60%) with probabilities  $\geq 95\%$  in the population-based cohorts were associated with high risks ( $OR > 4.0$ ) of breast cancer similar to the nonsense variants. The remaining 24–40% of missense variants were associated with attenuated moderate risks of breast cancer or ovarian cancer. This attenuation suggests that many missense variants have reduced effects on function and reduced risks of cancer and/or that the attenuation in part results from intrinsic

variability in the functional data. Future studies of *BRCA2* SNVs are needed to verify the reduced risks for subsets of variants and/or the existence of reduced penetrance variants, which may require modified approaches to risk counselling and patient management.

The MAVE study had several limitations. The small level of error in functional evaluation may still result in some improperly classified variants. Additional studies and comparisons with other functional assay datasets are anticipated to resolve some of the residual VUS and to confirm the results obtained from haploid HAP1 cells. Although RNA studies were not conducted as part of this study, several SNVs in canonical splice sites, intronic regions or with high SpliceAI scores were shown to be functionally pathogenic, which suggests that the variants result in aberrant RNA splicing and protein truncation. Further studies of these variants, which are beyond the scope of the current study, will establish whether the effects are through aberrant splicing.

In summary, SNVs in the *BRCA2* exons encoding the DBD mutation hotspot were characterized for effects on *BRCA2* activity using a cell-survival assay. The production of functional maps for 99% of all SNVs enabled the separation of nucleotide-level and protein-level functional aberrations and led to the clinical classification of more than 6,000 individual variants. These data will prove useful in the future, through integration with other datasets, for the characterization and classification of all variants in this genetic location in individuals from all racial and ethnic backgrounds and for all *BRCA2*-associated forms of cancer.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-024-08388-8>.

- Agalliu, I. et al. Germline mutations in the *BRCA2* gene and susceptibility to hereditary prostate cancer. *Clin. Cancer Res.* **13**, 839–843 (2007).
- Hu, C. et al. A population-based study of genes previously implicated in breast cancer. *N. Engl. J. Med.* **384**, 440–451 (2021).
- Hu, C. et al. Association between inherited germline mutations in cancer predisposition genes and risk of pancreatic cancer. *JAMA* **319**, 2401–2409 (2018).
- Kotsopoulos, J. et al. Germline mutations in 12 genes and risk of ovarian cancer in three population-based cohorts. *Cancer Epidemiol. Biomarkers Prev.* **32**, 1402–1410 (2023).
- Tavtigian, S. V. et al. The complete *BRCA2* gene and mutations in chromosome 13q-linked kindreds. *Nat. Genet.* **12**, 333–337 (1996).
- Findlay, G. M. et al. Accurate classification of *BRCA1* variants with saturation genome editing. *Nature* **562**, 217–222 (2018).
- Hu, C. et al. Functional analysis and clinical classification of 462 germline *BRCA2* missense variants affecting the DNA binding domain. *Am. J. Hum. Genet.* **111**, 584–593 (2024).
- Guidugli, L. et al. Assessment of the clinical relevance of *BRCA2* missense variants by functional and computational approaches. *Am. J. Hum. Genet.* **102**, 233–248 (2018).
- Richards, S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).
- Landrum, M. J. et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062–D1067 (2018).
- Hu, C. et al. Classification of *BRCA2* variants of uncertain significance (VUS) using an ACMG/AMP model incorporating a homology-directed repair (HDR) functional assay. *Clin. Cancer Res.* **28**, 3742–3751 (2022).
- Richardson, M. E. et al. Strong functional data for pathogenicity or neutrality classify *BRCA2* DNA-binding-domain variants of uncertain significance. *Am. J. Hum. Genet.* **108**, 458–468 (2021).
- Weile, J. & Roth, F. P. Multiplexed assays of variant effects contribute to a growing genotype-phenotype atlas. *Hum. Genet.* **137**, 665–678 (2018).
- Clark, K. A. et al. Comprehensive evaluation and efficient classification of *BRCA1* RING domain missense substitutions. *Am. J. Hum. Genet.* **109**, 1153–1174 (2022).
- Jia, X. et al. Massively parallel functional testing of *MSH2* missense variants conferring Lynch syndrome risk. *Am. J. Hum. Genet.* **108**, 163–175 (2021).
- Erwood, S. et al. Saturation variant interpretation using CRISPR prime editing. *Nat. Biotechnol.* **40**, 885–895 (2022).
- Li, H. et al. Functional annotation of variants of the *BRCA2* gene via locally haploid human pluripotent stem cells. *Nat. Biomed. Eng.* <https://doi.org/10.1038/s41551-023-01065-7> (2023).

18. Sahu, S. et al. Saturation genome editing of 11 codons and exon 13 of *BRCA2* coupled with chemotherapeutic drug response accurately determines pathogenicity of variants. *PLoS Genet.* **19**, e1010940 (2023).
19. Blomen, V. A. et al. Gene essentiality and synthetic lethality in haploid human cells. *Science* **350**, 1092–1096 (2015).
20. Hart, T. et al. Evaluation and design of genome-wide CRISPR/SpCas9 knockout screens. *Genetics* **197**, 2719–2727 (2017).
21. Wood, S. N. *Generalized Additive Models: An Introduction with R* 2nd edn (Chapman and Hall/CRC, 2017).
22. The R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2016).
23. Tavtigian, S. V., Harrison, S. M., Boucher, K. M. & Biesecker, L. G. Fitting a naturally scaled point system to the ACMG/AMP variant classification guidelines. *Hum. Mutat.* **41**, 1734–1737 (2020).
24. Ikegami, M. et al. High-throughput functional evaluation of *BRCA2* variants of unknown significance. *Nat. Commun.* **11**, 2573 (2020).
25. Biswas, K. et al. A computational model for classification of *BRCA2* variants using mouse embryonic stem cell-based functional assays. *NPJ Genom. Med.* **5**, 52 (2020).
26. Tavtigian, S. V. et al. Comprehensive statistical study of 452 *BRCA1* missense substitutions with classification of eight recurrent substitutions as neutral. *J. Med. Genet.* **43**, 295–305 (2006).
27. Cheng, J. et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* **381**, eadg7492 (2023).
28. Goodrich, J. et al. gnomAD v4: building scalable frameworks to process and quality control 730,913 exomes and 76,156 genomes. *Annu. Meeting Am. Soc. Hum. Genet.*, 247 (Washington DC, 2023).
29. Breast Cancer Association Consortium, et al. Breast cancer risk genes—association analysis in more than 113,000 women. *N. Engl. J. Med.* **384**, 428–439 (2021).
30. Tavtigian, S. V. et al. Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genet. Med.* **20**, 1054–1060 (2018).
31. Sahu, S. et al. Saturation genome editing-based clinical classification of *BRCA2* variants. *Nature* <https://doi.org/10.1038/s41586-024-08349-1> (2024).
32. Cheng, D. T. et al. Memorial Sloan Kettering—integrated mutation profiling of actionable cancer targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *J. Mol. Diagn.* **17**, 251–264 (2015).
33. Brnich, S. E. et al. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med.* **12**, 3 (2019).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

#### CARRIERS Consortium

**Christine B. Ambrosone<sup>14</sup>, Song Yao<sup>14</sup>, Amy Trentham-Dietz<sup>15</sup>, A. Heather Eliassen<sup>16</sup>, Lauren R. Teras<sup>17</sup>, Alpa Patel<sup>17</sup>, Christopher A. Haiman<sup>18</sup>, Esther M. John<sup>19</sup>, Elena Martinez<sup>20</sup>, James V. Lacey<sup>21</sup>, Dale P. Sandler<sup>22</sup>, Clarice R. Weinberg<sup>22</sup>, Julie R. Palmer<sup>23</sup>, Celine M. Vachon<sup>4</sup>, Janet E. Olson<sup>4</sup>, Kathryn E. Ruddy<sup>7</sup>, Hoda Anton-Culver<sup>24</sup>, Jeffrey N. Weitzel<sup>25</sup> & Peter Kraft<sup>26</sup>**

<sup>14</sup>Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA. <sup>15</sup>University of Wisconsin–Madison, Madison, WI, USA. <sup>16</sup>Harvard T. H. Chan School of Public Health, Boston, MA, USA. <sup>17</sup>Department of Population Science, American Cancer Society, Atlanta, GA, USA. <sup>18</sup>Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. <sup>19</sup>Stanford University School of Medicine, Stanford, CA, USA. <sup>20</sup>University of California San Diego, San Diego, CA, USA. <sup>21</sup>Beckman Research Institute, City of Hope Cancer Center, Duarte, CA, USA. <sup>22</sup>National Institute of Environmental Health Sciences, Durham, NC, USA. <sup>23</sup>Boston University School of Medicine and Slone Epidemiology Center, Boston, MA, USA. <sup>24</sup>University of California Irvine, Irvine, CA, USA. <sup>25</sup>University of Kansas Medical Center, Kansas City, KS, USA. <sup>26</sup>Division of Genetic Epidemiology, National Cancer Institute, National Institutes of Health, Rockville, MD, USA.

## Methods

### Cell line and reagents

HAP1 cells (Horizon Discovery) were maintained in IMDM with 10% FBS and 1% penicillin–streptomycin. For haploidy sorting,  $1 \times 10^{-7}$  HAP1 cells were resuspended in  $5 \text{ mg ml}^{-1}$  Hoechst 34580 (BD, 565877) and sorted at  $4^\circ\text{C}$ . HAP1 cells were transfected using Turbofectin 8.0 (Origene). All oligonucleotides and primers were synthesized by Integrated DNA Technologies.

### Generation of site-saturation mutagenesis libraries and Cas9–sgRNA plasmids

Exons 15–26 encoding the BRCA2 DBD, and adjacent upstream and downstream 10-bp intronic regions flanking each exon, were selected for SGE. Exons 18 and 25 were split into amino-terminal-targeted and carboxy-terminal-targeted regions because of their large exon size, which resulted in a total of 14 SGE target regions. Multiple sgRNAs were designed using the Benchling design tool. sgRNA-annealed oligonucleotides were ligated into pSpCas9(BB)-2A-Puro (PX459 v.2.0) (Addgene, 62988) following BbsI (New England Biolabs, R0539L) digestion to create a Cas9–sgRNA co-expression construct for each individual SGE. For each SGE, 600–1,000 bp homologous arms upstream and downstream of the target region were amplified from wild-type HAP1 gDNA and cloned into a BamHI-HF-digested pUC19 vector using a NEBuilder HiFi DNA assembly Cloning kit. Cloned plasmid backbones were subjected to site-saturation mutagenesis by inverse PCR<sup>34</sup> using mutagenized codon NNN primers for all possible nucleotide changes at each amino-acid position. A protospacer protection edit encoding a silent mutation was introduced by site-directed mutagenesis into the protospacer adjacent motif site or the sgRNA recognition site of each target region to prevent re-cutting by the Cas9–sgRNA after successful editing. Furthermore, a single 3-nucleotide mutation was introduced into the introns of each homologous arm to facilitate specific reamplification of the targeted DNA.

### CRISPR–Cas9 SGE

Multiple sgRNAs with predicted high editing efficiencies in HAP1 cells were evaluated in SGE experiments of each target region and the optimal sgRNAs were selected (Supplementary Table 1). In each SGE experiment, 5 million haploid-sorted HAP1 cells were co-transfected with 4 mg of the target-specific variant library and 16 mg of the Cas9–sgRNA targeting construct. Cells were selected in puromycin ( $1 \text{ mg ml}^{-1}$ ) for 3 days. Cells were collected at D0, D5 (24 h after puromycin selection) and D14 after transfection, and gDNA was extracted using a Monarch Genomic DNA Purification kit (New England Biolabs, T3010L). Target regions were amplified by PCR to add barcodes for multiplexing. All PCR reactions were performed in 50  $\mu\text{l}$  reactions using Q5 High-Fidelity 2 $\times$  master mix (New England Biolabs, M0492L). Primers for gDNA amplification are provided in Supplementary Table 2. All reactions were cleaned and concentrated using Ampure XP beads before sequencing for 150 cycles on an Illumina MiSeq (approximately 5 million reads per run) or NextSeq (approximately 30 million reads per run) instrument. Base calls were performed using the instrument control software and further processed using a customized algorithm.

### Sequencing data processing

FASTQ files of sequenced samples from Illumina MiSeq or NextSeq assays were trimmed for adapter sequences using cutadapt (v.3.5). SeqPrep (v.1.2) converted the paired-end reads into single reads. The single reads were aligned to the human reference genome (GRCh38) utilizing bwa-mem (v.0.7.17). Following alignment, the custom-developed tool CountReads was used for DNA-sequencing data analyses, with a particular focus on the identification and characterization of mutations. CountReads included the preparation of reference amino acid and DNA sequences, validation of sequencing data integrity and precise trimming of reads to relevant regions. The method also differentiated

between variant types and confirmed the presence of specific variants and aggregated and reported variant data. CountReads produced a variant call format (VCF) file, which was annotated using CAVA<sup>35</sup>. The SpliceAI tool (v.1.3.1)<sup>36</sup> was utilized to evaluate splicing effects associated with all observed SNVs.

### Functional read count process

The log<sub>2</sub> ratio between the frequency of D14 and D0 read counts was used to measure the depletion or enrichment effect for each variant. The comparison between experimental D0 and D5 was used for positional adjustment using a Loess transformation<sup>6</sup>. Variants with under-represented read counts ( $<10$ ) at D0 and D5 were excluded from further analysis. log<sub>2</sub> ratios of variants were linearly scaled within each exon across replicate experiments relative to median silent and median nonsense SNV values. For each variant, the average score was calculated from all non-missing values among replicates. Linear scaling was used to normalize scores across exons using median synonymous and nonsense values, similar to the within exon normalization. After completion of all data cleaning and quality control, a raw functional score was available for 6,959 SNVs (Supplementary Table 3).

### VarCall model for assessment of evidence of pathogenicity

Replicate-level variant frequencies were computed at each assay time point (D0, D5 and D14) by dividing the variant read count by the replicate total for each exon. To remove positional bias, the positional effect was estimated using the ratio between D0 and D5 read counts, using replicate-level generalized additive models with exon-specific adaptive splines<sup>21</sup>. The VarCall model<sup>37</sup> was applied to the positionally adjusted log ratio of the D14 and D0 read counts. VarCall is a class of Bayesian hierarchical model with context-specific measurement models that embed a Gaussian two-component mixture model for the variant effects. The formulation used here is based on a previous analysis of BRCA2 variants<sup>8</sup>. Variants were each assigned a binary indicator of pathogenicity status: deterministically if assumed known and probabilistically if not. Silent variants were assumed benign and nonsense variants pathogenic. The measurement model adjusted for batching by including replicate by exon-level location and scale random effects and included *t*-distributed error terms to allow for outliers. The JAGS language<sup>38</sup> was used to specify and fit the VarCall model using a MCMC algorithm. All related computations were carried out in the R programming language<sup>22</sup>. A prior probability of pathogenicity of 0.2 for variants in the DNA-binding region was used based on a predicted frequency of 0.23 for pathogenic variants in this region by AlphaMissense. Using the MCMC output, the Bayes factor in favour of pathogenicity for each variant was computed. The thresholds for the Bayes factor based on strength of evidence of pathogenicity or benign level ( $P_{\text{Strong}}$ ,  $P_{\text{Moderate}}$  or  $P_{\text{Supporting}}$ ,  $VUS$ ,  $B_{\text{Strong}}$ ,  $B_{\text{Moderate}}$  or  $B_{\text{Supporting}}$ ) were derived from the Bayesian interpretation of the ACMG–AMP guidelines<sup>23</sup>. Full details of the analysis are available in the Supplementary Methods.

### Three-dimensional structural modelling

BRCA2 functionally  $P_{\text{Strong}}$  missense alterations were mapped in the DBD using PyMol software. The Protein Data Bank source file (identifier 1MJE) was downloaded from the NCBI Molecular Modeling Database. Three-dimensional structural modelling was based on the crystal structure of a BRCA2–DSS1–ssDNA complex<sup>39</sup>.

### Multi-species amino-acid sequence conservation and in silico pathogenicity prediction

BRCA2 amino-acid sequences were obtained from Align-GVGD (<http://agvgd.hci.utah.edu/>). Sequence alignments were performed using ten species: *Homo sapiens*, *Pan troglodytes*, *Macaca mulatta*, *Rattus norvegicus*, *Canis familiaris*, *Bos taurus*, *Monodelphis domestica*, *Gallus gallus*, *Xenopus laevis* and *Tetraodon nigroviridis*. Sequence conservation analyses were performed on amino-acid residues that contained

BRCA2 DBD functionally pathogenic variants. Align-GVGD<sup>26</sup>, AlphaMissense<sup>27</sup> and Bayes-De1<sup>40</sup> were used for in silico pathogenicity prediction.

### Study populations

Breast cancer and ovarian cancer cases and associated clinical phenotypes were collected from individuals receiving cancer genetic testing by Ambry Genetics. Publicly available reference controls were women from gnomAD (v.2.1, v.3.1 and v.4 excluding the UK Biobank). Matching case-control data for breast cancer were also available from the CARRIERS and BRIDGES population-based breast cancer studies<sup>2,29</sup>, and breast cancer case-control data from the UK Biobank ([www.ukbiobank.ac.uk](http://www.ukbiobank.ac.uk)). Variants with an allele frequency of >0.001 were excluded from the analyses.

### Comparison with other BRCA2 functional assays

SGE functional results were compared with those from other studies, including a BRCA2-deficient cell-based HDR assay<sup>7</sup>, a BRCA2-deficient cell line-based drug assay<sup>24</sup>, a prime-editing-based SGE study<sup>16</sup> and a mouse embryonic-stem-cell-based functional analysis<sup>25</sup>.

### ACMG-AMP framework for classification of BRCA2 DBD variants

The ACMG-AMP rule-based framework combines evidence from population, computational and predictive, segregation, functional, and other data, with each contributing source weighted as very strong (PVS1), strong (PS1, PS2, PS3 and PS4), moderate (PM1, PM2, PM3, PM4, PM5 and PM6) or supporting (PP1, PP2, PP3, PP4 and PP5) evidence for pathogenic effects, or stand-alone (BA1), strong (BS1, BS2, BS3 and BS4) or supporting (BP1, BP2, BP3, BP4, BP5, BP6 and BP7) for benign effects. The combined data produce variant classifications of benign, LB, pathogenic, LP and VUS<sup>9</sup>. In this study, ACMG-AMP scoring rules established by the ClinGen BRCA1/2 VCEP were used for clinical classification of BRCA2 DBD SNVs. The BRCA2 functional data were integrated into the ClinGen-ACMG-AMP BRCA1/2 VCEP classification model under the PS3/BS3 rule. The values for functional evidence were capped at +4 and -4 on the log scale to avoid LP or LB classification with functional evidence alone. The study was approved by the Western Institutional Review Board, which exempted review of the clinical testing cohort, and by the Mayo Clinic Institutional Review Board (21-008216). Detailed ACMG-AMP criteria used in this study are provided in the Supplementary Methods.

### Tumour LOH analysis

LOH status for breast, ovarian, pancreatic, and prostate cancer tumours carrying germline BRCA2 DBD variants was acquired from tumour-normal paired sequencing using the IMPACT dataset<sup>32</sup>. The FACETS algorithm<sup>41</sup> was used to determine LOH from matched tumour-normal pairs. Only tumour samples with >40% tumour content were included in the analysis.

### Statistical analysis

Associations between variant classification groups in BRCA2 and the risk of breast cancer or ovarian cancer were performed for women who received genetic testing from Ambry Genetics and for women without cancer in gnomAD (v.2.1, v.3.1 and v.4 (excluding UK Biobank, from v.4)) using weighted logistic regression of control populations and weighting for the relative frequencies of different races and ethnicities in the cases. Associations in the population-based CARRIERS and BRIDGES matched breast cancer cases and unaffected women (as controls) and for UK Biobank breast cancer cases and controls were performed using Fisher's exact test. Phenotypic comparisons between cases with functionally pathogenic and benign variants were conducted using Student's *t*-test for quantitative variables and a Chi-squared test for qualitative variables. Lifetime absolute risks of breast cancer or ovarian cancer (malignant epithelial tumours of the ovary or fallopian tube) up to age 80 years were estimated for different classification groups by incorporating OR estimates with age-specific breast cancer or ovarian cancer incidence rates (restricted to individuals who identified as non-Hispanic white)

from the SEER Program of the National Cancer Institute, accounting for all-cause mortality rates<sup>2</sup>. One-way analysis of variance tests were conducted to compare the functional score differences of functional categories from other BRCA2 functional assays. Fisher's exact tests were used in tumour LOH analysis. All analyses were performed with R software (v.4.2.2) and all tests were two-sided. SGE data in bar graphs or scatter plots are presented as means from replicate experiments.

### Ethics statement

All data shown in this paper are provided with the explicit written consent of the study participants following approval from the institutional review boards.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

All data presented in the article and/or in the Supplementary Methods are available in the article or from the Gene Expression Omnibus (identifier GSE270424).

### Code availability

All related code for the VarCall model and for statistical analysis is available from GitHub ([https://github.com/najemayo/Couch\\_SGE\\_BRCA2](https://github.com/najemayo/Couch_SGE_BRCA2)).

34. Jain, P. C. & Varadarajan, R. A rapid, efficient, and economical inverse polymerase chain reaction-based method for generating a site saturation mutant library. *Anal. Biochem.* **449**, 90–98 (2014).
35. Münz, M. et al. CSN and CAVA: variant annotation tools for rapid, robust next-generation sequencing analysis in the clinical setting. *Genome Med.* **7**, 76 (2015).
36. Jaganathan, K. et al. Predicting splicing from primary sequence with deep learning. *Cell* **176**, 535–548 (2019).
37. Iversen, E. S. Jr., Couch, F. J., Goldgar, D. E., Tavtigian, S. V. & Monteiro, A. N. A computational method to classify variants of uncertain significance using functional assay data with application to BRCA1. *Cancer Epidemiol. Biomarkers Prev.* **20**, 1078–1088 (2011).
38. Plummer, M. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. *Proc. 3rd Int. Workshop on Distributed Statistical Computing* 1–10 (Vienna, 2003).
39. Yang, H. et al. BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. *Science* **297**, 1837–1848 (2002).
40. Pejaver, V. et al. Calibration of computational tools for missense variant pathogenicity classification and ClinGen recommendations for PP3/BP4 criteria. *Am. J. Hum. Genet.* **109**, 2163–2177 (2022).
41. Shen, R. & Seshan, V. E. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res.* **44**, e131 (2016).

**Acknowledgements** This study was funded in part by NIH grants R35CA253187 and R01CA225662, a Specialized Program of Research Excellence (SPORE) in Breast Cancer to the Mayo Clinic (P50CA116201) and the Breast Cancer Research Foundation (BCRF). Other sources of support included a Women's Cancer Program Pilot Project Award at Mayo Clinic (to S.Y.), a Halt Cancer at X National Research Grant (to S.Y.), a Paul Calabresi Award in Clinical Translation Research at the Mayo Clinic (K12CA090628) (to S.Y.), a Career Development Award from the Conquer Cancer Foundation (to S.Y.), a Sarasota Innovation Fund/Moffitt Foundation (to A.N.A.M.), a Precision Prevention BCRF award (to A.N.A.M.) and the Sigrid Jusélius Foundation (to M.M.). This research has been conducted in part using the UK Biobank Resource under Application Number 65898.

**Author contributions** H.H. and C.H. performed all SGE experiments and co-wrote the paper. J.N., S.N.H., T.R., M.A., R.D.G., Y.A.T., N.B., W.C., E.S.I. and F.J.C. designed and performed data analyses and co-wrote the paper. M.R., M.M., C.B. and D.M. analysed the IMPACT tumour data. P.C.M.L. and A.N.A.M. generated figures and co-wrote the paper. M.d.I.H. conducted splice analyses. S.Y., S.M.D. and K.L.N. performed clinical interpretation of functional results. T.P., R.K. and M.E.R. collected and performed analyses of clinical data. Members of the CARRIERS consortium contributed case-control association analyses. All authors contributed to writing of the paper.

**Competing interests** T.P., R.K. and M.R. are all employees of Ambry Genetics. All other authors declare no conflicts of interest.

### Additional information

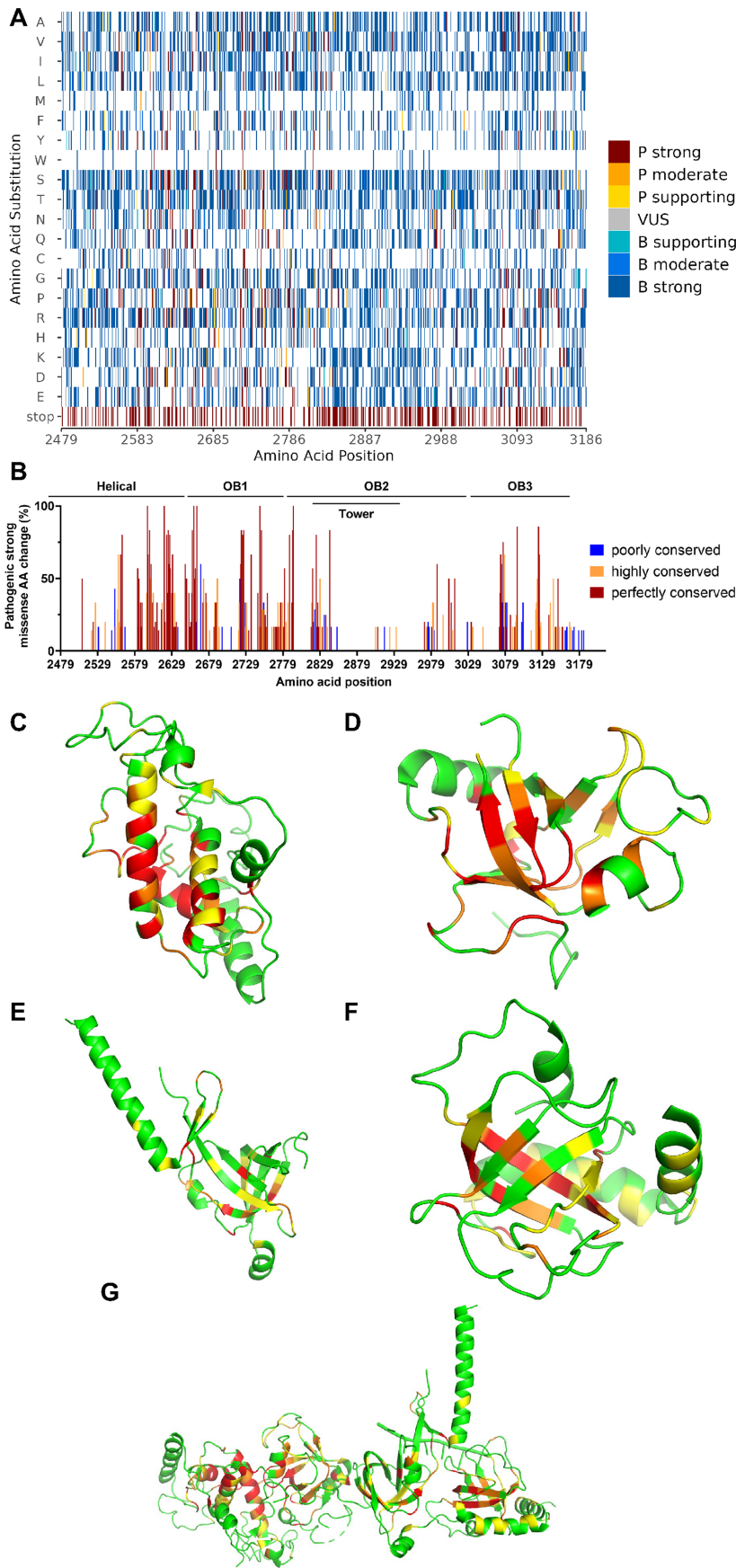
**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41586-024-08388-8>.

**Correspondence and requests for materials** should be addressed to Chunling Hu or Fergus J. Couch.

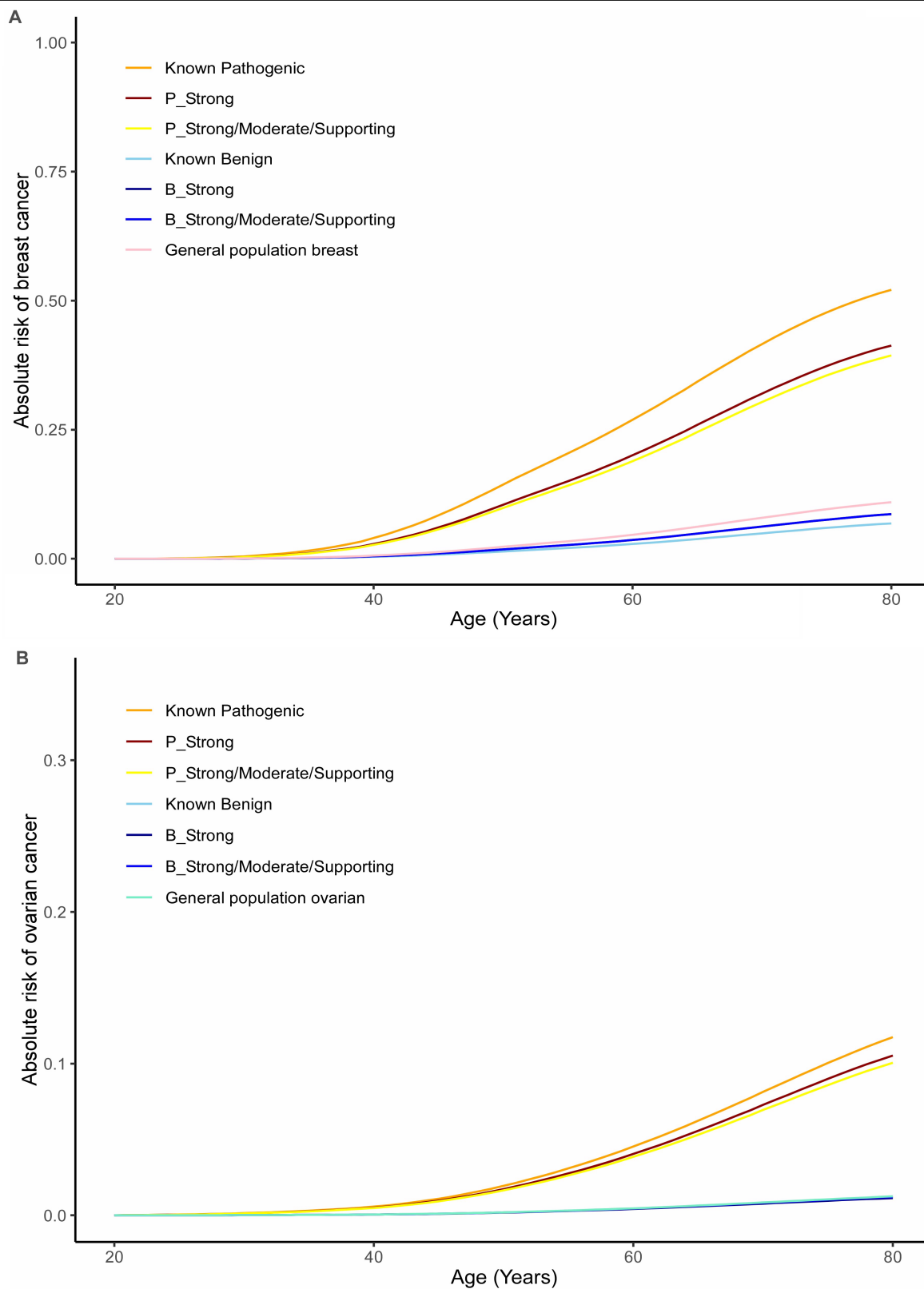
**Peer review information** Nature thanks Yongsu Kim, Sean Tavtigian and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.



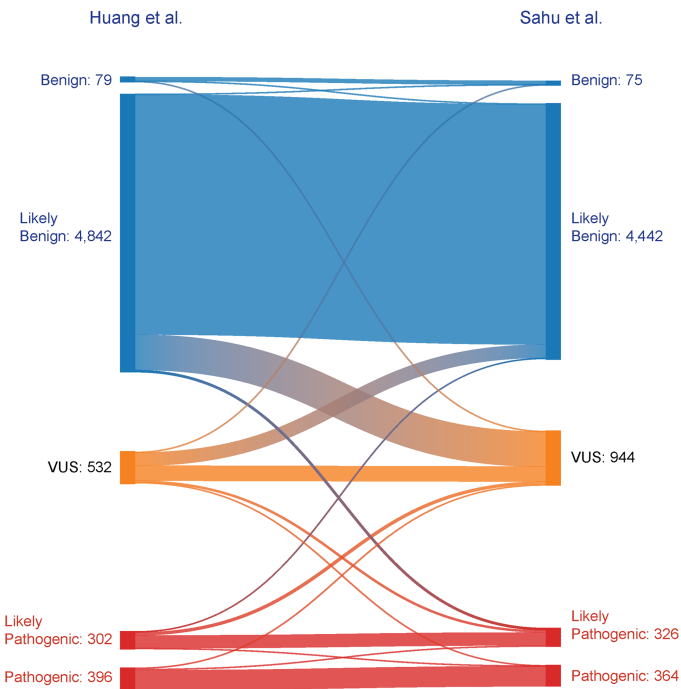


**Extended Data Fig. 1 | Functional effects of SNVs on the BRCA2 protein.** **A**, Heatmap of functional categories (colour) for all possible amino acid substitutions encoded by SNVs. **B**, Cross-species sequence conservation from pufferfish to *Homo sapiens* ( $n = 10$ ) relative to frequency of P\_Strong missense variants (perfectly conserved: 100% identity across 10 species; highly conserved: 80% or 90% identity; poorly conserved:  $\leq 70\%$  identity). **C–G**, BRCA2 3-dimensional protein ribbon diagrams showing the frequency of P\_Strong missense variants encoded by SNVs at each amino acid in the Helical (**C**), OB1 (**D**), OB2 (**E**), OB3 (**F**) domains, and the BRCA2-DSS1-ssDNA complex (PBD 1MJJE) (**G**). Colour denotes the frequency of P\_Strong missense alterations (green: 0%; yellow:  $< 25\%$ ; orange: 25–49.9%; red:  $\geq 50\%$ ). The subdomains were oriented to maximize views of the functionally pathogenic missense alterations. The BRCA2-DSS1-ssDNA complex is shown from N-terminus (left) to C-terminus (right).



**Extended Data Fig. 2 | Lifetime risks of breast and ovarian cancer associated with categories of pathogenic and benign variants. A,B.** Lifetime risk estimates for breast cancer (**A**) and ovarian cancer (**B**) associated with categories (P\_Strong, P\_Strong/Moderate/Supporting, B\_Strong, B\_Strong/Moderate/Supporting) of BRCA2 DNA binding domain SNVs from the BRCA2

MAVE study. Standards included known pathogenic (all protein truncating alterations), known benign (benign variants established by the ClinGen *BRCA1/2* VCEP), general population breast/ovarian (age related risks of these cancers from the general SEER registry).



**Extended Data Fig. 3 | Comparisons of variant classifications from two MAVE studies.** Sankey plot of ClinGen/ACMG/AMP BRCA1/2 VCEP-based classification of commonly evaluated BRCA2 DBD SNVs from two independent MAVE studies (our study of HAP1 cells (Huang et al.), and Sahu et al.'s study of ES cells<sup>31</sup>).

Extended Data Table 1 | SNV type by target region

Target regions	SNVs	Missense SNVs	Nonsense SNVs	Silent SNVs	Canonical splice SNVs	Intronic SNVs
E15	597	386	22	129	12	48
E16	615	413	33	109	12	48
E17	564	370	27	107	12	48
E18N	554	395	18	112	6	23
E18C	561	375	25	131	6	24
E19	519	329	20	110	12	48
E20	486	314	36	76	12	48
E21	426	257	20	89	12	48
E22	648	440	43	105	12	48
E23	543	336	38	109	12	48
E24	468	301	20	87	12	48
E25N	387	265	19	73	6	24
E25C	390	267	17	76	6	24
E26 (to aa3186)	201	135	4	32	6	24
Total	6959	4583	342	1345	138	551

Variants in Protospacer Protection Edit (PPE) sites that were not SNVs after site-mutagenesis were excluded from counting in both possible SNVs and SNVs with functional score.  
SNV: single nucleotide variant.



Extended Data Table 2 | Summary of MAVE results in functional categories based on the VarCall model

BRCA2 functional categories based on VarCall model by variant type																
Nonsense (n=342) Canonical splice (n=138) Missense (n=4583) Intronic (n=551) Silent (n=1345) Total (n=6959)	P strong	P moderate	P supporting		VUS		B supporting		B moderate		B strong		P Combined		B Combined	
	339 (99.1%)	2 (0.6%)	1 (0.3%)		0 (0%)		0 (0%)		0 (0%)		0 (0%)		342 (100%)		0 (0%)	
	119 (86.2%)	2 (1.4%)	0 (0%)		4 (2.9%)		1 (0.7%)		3 (2.2%)		9 (6.5%)		121 (87.7%)		13 (9.4%)	
	502 (10.9%)	75 (1.6%)	34 (0.7%)		93 (2.0%)		52 (1.1%)		166 (3.6%)		3661 (79.9%)		611 (13.3%)		3879 (84.6%)	
	50 (9.1%)	7 (1.3%)	12 (2.2%)		22 (4.0%)		8 (1.5%)		18 (3.3%)		434 (78.8%)		69 (12.5%)		460 (83.5%)	
	11 (0.8%)	2 (0.1%)	0 (0%)		3 (0.2%)		0 (0%)		3 (0.2%)		1326 (98.6%)		13 (1.0%)		1329 (98.8%)	
	1021 (14.7%)	88 (1.3%)	47 (0.7%)		122 (1.8%)		61 (0.9%)		190 (2.7%)		5430 (78.0%)		1156 (16.6%)		5681 (81.6%)	
BRCA2 functional categories by variant type from each targeted region																
		E15	E16	E17	E18N	E18C	E19	E20	E21	E22	E23	E24	E25N	E25C	E26	Total
P strong	Total	42	117	138	74	115	93	67	46	76	75	64	61	42	11	1021
	Canonical splice	7	12	10	6	6	12	12	11	11	8	12	6	4	2	119
	Intronic	3	4	10	1	1	6	5	7	5	6	0	1	0	1	50
	Missense	10	67	91	49	82	53	14	7	17	19	32	35	21	5	502
	Nonsense	22	33	26	18	25	20	36	20	43	38	19	19	17	3	339
	Silent	0	1	1	0	1	2	0	1	0	4	1	0	0	0	11
P moderate	Total	7	4	9	7	8	12	6	4	4	6	6	4	6	5	88
	Canonical splice	1	0	0	0	0	0	0	0	0	0	0	0	1	0	2
	Intronic	0	0	1	1	0	0	1	2	1	1	0	0	0	0	7
	Missense	6	4	7	6	8	12	5	2	3	5	4	4	5	4	75
	Nonsense	0	0	1	0	0	0	0	0	0	0	0	0	0	1	2
	Silent	0	0	0	0	0	0	0	0	0	0	2	0	0	0	2
P supporting	Total	6	8	2	8	4	5	1	0	3	3	3	0	2	2	47
	Intronic	0	0	1	1	0	3	1	0	0	3	2	0	0	1	12
	Missense	6	8	1	7	4	2	0	0	3	0	0	0	2	1	34
	Nonsense	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
VUS	Total	13	18	4	9	8	11	10	4	11	13	5	6	7	3	122
	Canonical splice	2	0	1	0	0	0	0	0	0	1	0	0	0	0	4
	Intronic	4	1	2	0	0	3	4	0	4	2	1	0	0	1	22
	Missense	6	17	1	9	8	7	6	4	7	9	4	6	7	2	93
	Silent	1	0	0	0	0	1	0	0	0	1	0	0	0	0	3
B supporting	Total	7	5	8	6	2	4	2	1	8	7	3	1	2	5	61
	Canonical splice	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	Intronic	0	0	2	1	0	1	0	0	0	1	1	0	1	1	8
	Missense	6	5	6	5	2	3	2	1	8	6	2	1	1	4	52
B moderate	Total	16	24	11	6	17	9	12	5	19	23	10	5	19	14	190
	Canonical splice	0	0	0	0	0	0	0	0	0	2	0	0	0	1	3
	Intronic	3	1	0	1	2	0	1	0	2	1	3	0	1	3	18
	Missense	13	22	11	5	15	9	11	5	17	20	7	5	16	10	166
	Silent	0	1	0	0	0	0	0	0	0	0	0	0	2	0	3
B strong	Total	506	439	392	444	407	385	388	366	527	416	377	310	312	161	5430
	Canonical splice	1	0	1	0	0	0	0	1	1	1	0	0	1	3	9
	Intronic	38	42	32	18	21	35	36	39	36	34	41	23	22	17	434
	Missense	339	290	253	314	256	243	276	238	385	277	252	214	215	109	3661
	Silent	128	107	106	112	130	107	76	88	105	104	84	73	74	32	1326
	Total	597	615	564	554	561	519	486	426	648	543	468	387	390	201	6959

Summary of VarCall model strength of evidence results by variant type and by target region. P: Pathogenic; B: Benign.

**Extended Data Table 3 | BRCA2 DBD subdomain specific multi-species amino acid residue conservation and comparison with MAVE functional results**

BRCA2 subdomain							
	Helical	OB1	OB2	Tower	OB3	Subdomain linker region	Total
<b>All residues</b>	187	134	111	124	130	22	708
Evaluated residues*	181	132	103	123	128	21	688
<b>All missense AA&amp;</b>	1008	764	520	708	757	107	3864
P_Strong AA/missense AA (%)	154/1008 (15.3%)	125/764 (16.4%)	45/520 (8.7%)	13/708 (1.8%)	83/757 (10.9%)	3/107 (2.8%)	423/3864 (10.9%)
P_Strong AA+/total (%)	154/423 (36.4%)	125/423 (29.6%)	45/423 (10.6%)	13/423 (3.1%)	83/423 (19.6%)	3/423 (0.7%)	423/423 (100%)
<b>Perfectly conserved residues</b>	77	47	35	28	19	6	212
Residues with P_Strong AA/residue (%)\$	45/77 (58.4%)	30/47 (63.8%)	12/35 (34.3%)	2/28 (7.1%)	14/19 (73.7%)	1/6 (16.7%)	103/212 (48.6%)
Residues with P_Strong AA/total (%)!	45/103 (44%)	30/103 (32%)	12/103 (11.7%)	2/103 (1.9%)	14/103 (13.6%)	1/103 (1.0%)	103/103 (100%)
P_Strong AA/total P_Strong AA (%)#	116/154 (75%)	73/125 (58%)	23/45 (51%)	6/13 (46%)	35/83 (42%)	1/3 (33%)	254/423 (60%)
<b>Highly conserved residues</b>	50	47	26	49	48	5	225
Residues with P_Strong AA/residue (%)\$	15/50 (30%)	24/47 (51.1%)	9/26 (34.6%)	4/49 (8.2%)	19/48 (39.6%)	0/5 (0%)	71/225 (31.6%)
Residues with P_Strong AA/total (%)!	15/71 (21.1%)	24/71 (33.8%)	9/71 (12.7%)	4/71 (5.6%)	19/71 (26.8%)	0/71 (0%)	71/71 (100%)
P_strong AA/total P_Strong AA (%)#	27/154 (18%)	43/125 (35%)	12/45 (27%)	4/13 (31%)	29/83 (35%)	0/3 (0%)	115/423 (27%)
<b>Poorly conserved residues</b>	54	38	42	46	61	10	251
Residues with P_Strong AA/residue (%)\$	7/54 (13%)	5/38 (13.2%)	9/42 (21.4%)	3/46 (6.5%)	13/61 (21.3%)	2/10 (20%)	39/251 (15.5%)
Residues with P_Strong AA/total (%)!	7/39 (17.9%)	5/39 (12.8%)	9/39 (23.1%)	3/39 (7.7%)	13/39 (33.3%)	2/39 (5.1%)	39/39 (100%)
P_Strong AA/total P_Strong AA (%)#	11/154 (7%)	9/125 (7%)	10/45 (22%)	3/13 (23%)	19/83 (23%)	2/3 (67%)	54/423 (13%)

\*: Residues containing the Protospacer Protection Edit (PPE) sites were excluded.

&: Missense changes with known splicing or potential splicing effects (SpliceAI > 0.2) excluded.

+: P\_Strong functional category was defined by the VarCall model.

\$: "Residues with P\_Strong AA/residue (%)": Of the 77 perfectly conserved residues in the helical subdomain, 45 residues contain at least one P-strong AA change.!: "Residues with P\_Strong AA/total (%)": Of all perfectly conserved residues with at least one P-strong AA change (n=103), how many residues are located within each subdomain.

#: "P-Strong AA/ total P\_Strong AA (%)": Of all 154 P-strong AA changes found in the helical subdomain, how many are perfectly conserved.

DBD: DNA binding domain; MAVE: multiplex assays of variant effect; AA: amino acid.

Perfectly conserved residues were defined as 100% residue sequence similarity across 10 species (pufferfish to homo sapiens).

Highly conserved residues were defined as 80-90% residue sequence similarity across 10 species (pufferfish to homo sapiens).

Poorly conserved residues were defined as ≤70% residue sequence similarity across 10 species (pufferfish to homo sapiens).

Extended Data Table 4 | Comparison of in silico prediction models with BRCA2 DBD MAVE functional results

In silico prediction	MAVE functional categories							sensitivity	specificity
	B Strong	B Moderate	B Supporting	VUS	P Supporting	P Moderate	P Strong		
<u>AGVGD</u>									
Class C0	2419	120	24	58	16	36	136	0.71	0.57
Class C15	257	11	4	12	5	1	26		
Class C25	185	11	1	5	2	5	31		
Class C35	160	14	8	4	2	4	34		
Class C45	38	3	0	0	2	4	22		
Class C55	96	3	3	5	1	3	28		
Class C65	271	17	13	18	7	19	195	0.41	0.91
<u>AlphaMissense</u>									
ambiguous	411	26	5	16	7	8	51		
Likely benign (<0.34)	2729	103	24	41	12	36	78	0.75	0.68
Likely pathogenic (>0.564)	521	37	23	36	15	31	373	0.74	0.84
<u>BayesDel (BRCA1/2 VCEP)</u>									
Pathogenic (>=0.3)	149	11	7	18	7	17	218	0.43	0.95
VUS (0.18 to 0.3)	237	17	3	10	6	9	97		
Benign (<=0.18)	3275	138	42	65	21	49	187	0.89	0.46
<u>BayesDel (ClinGen)</u>									
P strong (>=0.5)	3	0	0	2	0	0	19	0.73	0.83
P moderate (0.27 to 0.5)	187	15	7	19	9	18	216		
P supporting (0.13 to 0.27)	374	25	9	17	6	13	131		
VUS (-0.18 to 0.13)	2257	106	28	50	16	36	115		
B supporting (-0.36 to -0.18)	762	19	8	3	2	8	18		
B moderate (<=-0.36)	78	1	0	2	1	0	3	0.23	0.93

The sensitivity was calculated using MAVE functional results as ground truth and comparing the MAVE P\_Strong evidence thresholds with predicted pathogenic/likely pathogenic class (i.e. C65 from AGVGD, likely pathogenic from AlphaMissense, pathogenic from BayesDel, and P Strong/Moderate/Supporting for BayesDel with ClinGen thresholds), or comparing the MAVE B\_Strong evidence thresholds with predicted benign/likely benign class (i.e. C0 from AGVGD, likely benign from AlphaMissense, benign from BayesDel, and B Moderate/Supporting for BayesDel with ClinGen thresholds). Formula for the sensitivity calculation is shown below:  
(MAVE P\_strong and predicted P/LP) / (all MAVE P\_strong) or (MAVE B\_strong and predicted B/LB) / (all MAVE B\_strong)

The specificity was calculated using MAVE functional results as ground truth and comparing the MAVE non-P\_Strong evidence thresholds with predicted non-pathogenic/non-likely pathogenic class, or comparing the MAVE non-B\_Strong evidence thresholds with predicted non-benign/non-likely benign class. Formula for the specificity calculation is shown below:  
(MAVE non-P\_strong and predicted non-P/LP) / (all MAVE non-P\_strong) or (MAVE non-B\_strong and predicted non-B/LB) / (all MAVE non-B\_strong)

**Extended Data Table 5 | Summary of *BRCA2* MAVE ClinGen/ACMG/AMP *BRCA1/2* VCEP rules classification**

<b>DBD SNVs</b>	ClinVar+ ENIGMA	MAVE ACMG/AMP classification	ClinVar+ ENIGMA variants after applying MAVE ACMG/AMP	MAVE ACMG/AMP classified not in ClinVar or ENIGMA	Classified in MAVE data from all sources	ClinVar+ ENIGMA not in MAVE data	Total classified
B/LB	724	5566	2112	3474	5569	20	5589
P/LP	396	785	502	291	785	8	793
VUS	1761	608	267	362	605	21	626
Total	2881	6959	2881	4127	6959	49	7008

<b>DBD Missense</b>	ClinVar+ ENIGMA	MAVE ACMG/AMP classification	ClinVar+ ENIGMA variants after applying MAVE ACMG/AMP	MAVE ACMG/AMP classified not in ClinVar or ENIGMA	Classified in MAVE data from all sources	ClinVar+ ENIGMA not in MAVE	Total classified
B/LB	79	3786	1386	2405	3788	3	3791
P/LP	66	261	145	121	261	5	266
VUS	1623	536	237	318	534	21	555
Total	1768	4583	1768	2844	4583	29	4612

MAVE: multiplex assays of variant effect; ACMG/AMP: American College of Medical Genetics/ Association for Molecular Pathology; DBD: DNA binding domain; SNVs: single nucleotide variant; ENIGMA: Evidence-based Network for the Interpretation of Germline Mutant Alleles; B: benign; LB: likely benign; P: pathogenic; LP: likely pathogenic; VUS: variant of uncertain significance. Note: all variants evaluated were restricted to *BRCA2* DBD SNVs including adjacent intronic +/-10nt region of each exon. ClinVar VUS refers to variants identified as VUS by at least one reporting group in ClinVar, unless classified as B/LB or P/LP by the *BRCA1/2* VCEP.



# Article

**Extended Data Table 6 | Phenotype of carriers of BRCA2 DBD MAVS benign and pathogenic strong/moderate variants**

	Benign	Pathogenic strong/moderate	Total	p value
<b>Population-based cohort</b>				
Number of variant carriers	1122	117	1239	
Age at diagnosis (breast)				2.00X10 <sup>-5</sup>
Mean (SD)	61.34 (12.03)	56.32 (12.50)	60.87 (12.16)	
Median	62.25	56.10	62	
Range	25.00 - 91.00	21.00 - 88.50	21.00 - 91.00	
ER status				0.352
Negative	139 (16.8%)	15 (21.1%)	154 (17.1%)	
Positive	689 (83.2%)	56 (78.9%)	745 (82.9%)	
Family History of cancer (Breast)				0.146
No	908 (79.6%)	88 (73.9%)	996 (79.1%)	
Yes	232 (20.4%)	31 (26.1%)	263 (20.9%)	
Family History of cancer (Ovarian)				0.635
No	1093 (95.9%)	113 (95.0%)	1206 (95.8%)	
Yes	47 (4.1%)	6 (5.0%)	53 (4.2%)	
Variant type				1.29X10 <sup>-84</sup>
Canonical splice	0 (0.0%)	6 (5.0%)	6 (0.5%)	
Intronic	86 (7.5%)	0 (0.0%)	86 (6.8%)	
Missense	926 (81.2%)	79 (66.4%)	1005 (79.8%)	
Silent	128 (11.2%)	1 (0.8%)	129 (10.2%)	
nonsense	0 (0.0%)	33 (27.7%)	33 (2.6%)	
Ethnicity				0.056
African American	267 (23.4%)	19 (16.0%)	286 (22.7%)	
White	737 (64.6%)	90 (75.6%)	827 (65.7%)	
Others	136 (11.9%)	10 (8.4%)	146 (11.6%)	
<b>Clinical cohort</b>				
Number of variant carriers	7095	297	7392	
Age at diagnosis (breast)				2.70X10 <sup>-5</sup>
Mean (SD)	48.75 (11.39)	45.93 (10.41)	48.64 (11.36)	
Median	47	46	47	
Range	18.00 - 88.00	22.00 - 75.00	18.00 - 88.00	
ER status				0.792
Negative	1378 (27.2%)	56 (28.0%)	1434 (27.2%)	
Positive	3697 (72.8%)	144 (72.0%)	3841 (72.8%)	
Family History of cancer (Breast)				2.63X10 <sup>-9</sup>
No	1461 (42.0%)	61 (23.3%)	1522 (40.7%)	
Yes	2016 (58.0%)	201 (76.7%)	2217 (59.3%)	
Variant type				9.99X10 <sup>-20</sup>
Canonical splice	0 (0.0%)	42 (14.1%)	42 (0.6%)	
Intronic	58 (0.8%)	10 (3.4%)	68 (0.9%)	
Missense	4276 (60.3%)	132 (44.4%)	4408 (59.6%)	
Silent	2761 (38.9%)	10 (3.4%)	2771 (37.5%)	
nonsense	0 (0.0%)	103 (34.7%)	103 (1.4%)	
Ethnicity				0.061
African American	1607 (22.6%)	53 (17.8%)	1660 (22.5%)	
Alaskan native	1 (0.0%)	0 (0.0%)	1 (0.0%)	
Ashkenazi Jewish	184 (2.6%)	1 (0.3%)	185 (2.5%)	
Asian	276 (3.9%)	15 (5.1%)	291 (3.9%)	
Caucasian	3411 (48.1%)	166 (55.9%)	3577 (48.4%)	
Hispanic	595 (8.4%)	29 (9.8%)	624 (8.4%)	
Middle eastern	42 (0.6%)	2 (0.7%)	44 (0.6%)	
Mixed ethnicity	505 (7.1%)	18 (6.1%)	523 (7.1%)	
Native American	16 (0.2%)	0 (0.0%)	16 (0.2%)	
Other	27 (0.4%)	0 (0.0%)	27 (0.4%)	
Unknown	431 (6.1%)	13 (4.4%)	444 (6.0%)	

Chi-Square tests were used for categorical variables; Kruskal Wallis test was used for continuous variable. Both tests are two-sided. No adjustments were made for multiple comparisons.

Extended Data Table 7 | Comparison of tumor biallelic status with MAVE functional results and ACMG classifications

	Number of tumors	Number of tumor biallelic loss	Tumor LOH rate (%)
MAVE functional categories*			
MAVE P Strong	26	22	85
MAVE P Strong & Moderate	29	23	79
MAVE B Strong	233	58	25
MAVE missense P Strong	6	4	67
MAVE missense P Strong & Moderate	9	5	56
ACMG classifications#			
ACMG P & LP	25	22	88
ACMG B & LB	249	60	24
ACMG missense P & LP	7	5	71

\*: Tumor biallelic status was obtained from the Memorial Sloan Kettering Cancer Center (MSKCC) – IMPACT dataset.  
#: Recurrent variants and variants with conflicting interpretation of pathogenicity were excluded from the analysis.  
MAVE: multiplexed assay of variant effect; ACMG: American College of Medical Genetics; P: pathogenic; B: benign; LP: likely pathogenic; LB: likely benign.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
<input type="checkbox"/>	<input checked="" type="checkbox"/> The exact sample size ( <i>n</i> ) for each experimental group/condition, given as a discrete number and unit of measurement
<input type="checkbox"/>	<input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
<input type="checkbox"/>	<input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
<input type="checkbox"/>	<input checked="" type="checkbox"/> A description of all covariates tested
<input type="checkbox"/>	<input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
<input type="checkbox"/>	<input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
<input type="checkbox"/>	<input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i> ) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
<input type="checkbox"/>	<input checked="" type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
<input type="checkbox"/>	<input checked="" type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
<input type="checkbox"/>	<input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i> ), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	Pymol 4.6 was used to obtain protein structure data
Data analysis	We used the VarCall Bayesian two component mixture model (Iversen CEBP 2011). The custom code is included in Supplementary Materials and was deposited to github (GitHub - najiemayo/Couch_SGE_BRCA2).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

- All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
  - A description of any restrictions on data availability
  - For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All data presented in the article and/or in the Supplementary Materials and Methods are available in the article or in GEO (GSE270424), all related code was deposited in Github (GitHub - najiemayo/Couch\_SGE\_BRCA2). The databases used in the study include GRCh38.

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	Analyses were restricted to females because breast and ovarian cancer are predominantly diseases in women
Reporting on race, ethnicity, or other socially relevant groupings	Analyses were inclusive of all races and ethnicities. Sub-analyses were restricted to non-Finn Europeans and African Americans to show associations with breast cancer in these populations. The numbers of cases and controls from other populations were insufficient for analyses.
Population characteristics	Women with cancer from the CARRIERS population-based cohort comprised of 8 contributing nested case-control studies from US based cohorts. Women from the BRIDGES population-based cohort. Women with and without breast cancer from the UK Biobank. Unaffected controls from the same cohorts were matched 1:1 in 5 year categories. Breast and ovarian cancer patients undergoing hereditary cancer testing that were selected using NCCN criteria to qualify for testing. GnomAD V4 and GnomAD 2.1 and 3.1 public reference controls. Covariates for patients include age or age at diagnosis of breast or ovarian cancer, and race/ethnicity, family history of breast or ovarian cancer, estrogen receptor status of breast tumors.
Recruitment	N/A
Ethics oversight	Mayo Clinic IRB approval (21-008216). Waiver of approval from the Western IRB for the Ambry Genetics dataset

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculations were used because this is an observational study. The largest available studies were used.
Data exclusions	Genes with variant counts of less than 5 were excluded from odds ratio calculations. Variants with frequencies in case or control populations of >0.001 were excluded.
Replication	Experiments were performed in triplicate.
Randomization	Samples and patients were not randomized. Data were organized by cancer case and control status.
Blinding	Blinding is not relevant to this study. Variant functional effects did not rely on any previous data. Statistical analyses were conducted based on case-control status.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems		Methods	
n/a	Involved in the study	n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies	<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines	<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology	<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants		

## Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	Human leukemia cell line purchased from Horizon Discovery Inc.
Authentication	All cell lines tested negative for mycoplasma
Mycoplasma contamination	Negative
Commonly misidentified lines (See <a href="#">ICLAC</a> register)	N/A

## Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A