When Two Plus Four Does Not Equal Six: Combining Computational and Functional Evidence to Classify BRCA1 Key Domain Missense Substitutions.

Author List: Scott T. Pew,[1] Madison B. Wiffler,[2] Alun Thomas,[3] Julie L. Boyle,[4,5] Melissa S. Cline,[6] Nicola J. Camp,[5,7] David E. Goldgar,[5,8] Sean V. Tavtigian[1,5*]

Affiliations:

1. Department of Oncological Sciences, University of Utah School of Medicine, Salt Lake City, Utah, 84112, USA

2. Department of Neurobiology, University of Utah School of Medicine, Salt Lake City, Utah, 84112, USA

3. Division of Epidemiology, Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, Utah, 84108, USA

4. Department of Family and Preventive Medicine, University of Utah School of Medicine, Salt Lake City, Utah, 84108, USA

5. Huntsman Cancer Institute, Salt Lake City, Utah, 84108, USA

6. UC Santa Cruz Genomics Institute, Genomics, University of California, Santa Cruz, CA, 95064, USA

7. Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, Utah, 84108, USA

8. Department of Dermatology, University of Utah School of Medicine, Salt Lake City, Utah, 84108, USA

* Correspondence: sean.tavtigian@hci.utah.edu.

ABSTRACT

Classification of genetic variants remains an obstacle to realizing the full potential of clinical genetic sequencing. Because of their ability to interrogate large numbers of variants, multiplexed assays of variant effect (MAVEs) and computational tools are viewed as a critical part of the solution to variant classification uncertainty. However, the (joint) performance of these assays and tools on novel variants has not been established. Transformation of the qualitative classification guidelines developed by the American College of Medical Genetics and Genomics (ACMG) into a quantitative Bayesian point system enables empirical validation of strength of evidence assigned to evidence criteria. Here, we derived a maximum likelihood estimate (MLE) model that converts frequentist odds ratios calculated from case-control data to proportions pathogenic and applied this model to functional assays, alone and in combination with computational tools across several domains of *BRCA1*. Furthermore, we defined exceptionally conserved ancestral residues (ECARs) and interrogated the performance of assays and tools at these residues in *BRCA1*. We found that missense substitutions in *BRCA1* that fall at ECARs are disproportionately likely to be pathogenic with effect sizes similar to that of protein truncating variants. In contrast, for substitutions falling at non-ECAR positions, concordant predictions of pathogenicity from functional assay and computational tool often fail to meet the additive assumptions of strength in ACMG guidelines. Thus, collectively, we conclude that strengths of evidence assigned by expert opinion in the ACMG guidelines are not universally applicable and require empirical validation.

50 **INTRODUCTION**:

51 ClinVar recently celebrated its 10th anniversary. In that time, as clinical sequencing has
52 become more prevalent, the number of variants cataloged has grown from approximately
53 27,000 to well over 2 million. Of these 2 million variants, approximately 1.3 million are single
54 nucleotide missense variants (excluding synonymous and nonsense variants), of which roughly
55 75% remain unclassified (~990,000 variants) [1]. Realizing the full potential of clinical genetic
56 sequencing relies on the ability to accurately and efficiently classify these variants into a realistic
57 continuum from pathogenic to benign. To that end, many multiplexed assays of variant effect
58 (MAVE) and computational tools have been developed. However, the accuracy of these assays
59 and tools, used independently and in combination with one another, towards classifying novel
60 variants identified during sequencing remains to be established.

61 In 2015, the American College of Medical Genetics and Genomics (ACMG) and the
62 Association for Molecular Pathology (AMP) published guidelines that relied on expert opinion to
63 assign strength of evidence to the various evidence criteria used in sequence variant
64 classification [2]. Subsequently, this approach was found to be compatible with a Bayesian point
65 system where the qualitative assessments of evidence strength categories (supporting,
66 moderate, strong, etc.) were converted to quantitative assertions of odds of pathogenicity (OP)
67 [3, 4]. Shifting to the Bayesian point system makes possible objective evaluation of the strength of
68 evidence the various evidence criteria actually provide. For example, it was recently
69 demonstrated that several computational tools exceeded the qualitative threshold of
70 "supporting" evidence in favor of benignity or pathogenicity and were able to provide
71 "moderate", and in some cases "strong" evidence in favor of benignity or pathogenicity [5].

72 In the 2015 ACMG/AMP guidelines, well-established functional assays were qualitatively
73 assigned the ability to provide "strong" evidence in favor of pathogenicity or benignity [6]. In the
74 Bayesian point system, this translates to +4 points (18.7:1 odds in favor of pathogenicity) or -4
75 points (0.053:1 odds in favor of pathogenicity). Yet, limited work has been performed to

76 quantitatively evaluate the strength of evidence that functional assays provide. As these assays
77 are not a direct measure of human health, an important step towards applying a specific
78 functional assay's results to variant classification is validation of the assay's calibration against
79 human subjects' data from the disease in question.

80 Furthermore, the 2015 guidelines provided combining rules that prescribed combinations
81 of different categories of evidence that could move a variant towards classification as
82 pathogenic or benign. Implicit in the original guidelines, and explicitly stated in the subsequent
83 work that transformed the guidelines into a Bayesian framework, is the idea that combining
84 evidence changes the estimate of a sequence variant's probability of pathogenicity. In a
85 previous publication [7], we noted that a majority of variants in the *BRCA1* RING domain could be
86 classified by using computational predictions (ACMG codes: PP3/BP4), whether or not a variant
87 fell at a canonical residue (ACMG code: PM1), functional assay predictions (ACMG codes:
88 PS3/BS3), and minimal human subjects observational data (ACMG codes: PS4, PM2, BS1).
89 This led to our conclusion that validation work is needed to ensure that the strength assigned to
90 these types of evidence was accurate and that addition of Bayesian points associated with
91 these evidence types was the correct mathematical operation.

92 As a step towards an empirical test, we derive a maximum likelihood estimate (MLE)
93 expression for the proportion of variants that are pathogenic within an analytically defined pool.
94 This model transforms odds ratios (OR) from human subjects' case-control data into a
95 proportion pathogenic which can then be used to objectively test the strength of evidence a
96 particular category of evidence provides. Subsequently, we demonstrate two important
97 applications for such a model using case-control data from *BRCA1* as a test case. First, we use
98 our model to determine if evidence from several functional assays developed for *BRCA1*
99 generate the strength of evidence assumed in the 2015 ACMG/AMP guidelines. Second, we
100 explore the impact of combining functional assay, canonical residue (which we rename
101 "exceptionally conserved ancestral residues" (ECARs)), and computational tool data to see if

102 doing so changes the proportion pathogenic of analytical subsets of variants in a way that
103 matches the additivity expectations set forth in the Bayesian adaptation of the 2015
104 classification guidelines (Figure 1). Such an approach allows for the quantitative validation of
105 the qualitative assumptions proposed in variant classification guidelines, a step that is currently
106 lacking for many evidence criteria used in variant classification.

107 **METHODS**:

108 Maximum Likelihood Estimate of Proportion Pathogenic

109 In addition to the assumptions in the Bayesian point system ($OP = 81^{\frac{ACMG\ points}{6}}$, prior
110 probability of pathogenicity = 0.102, points are log odds and therefore additive), this approach
111 relies on several other assumptions:

112 1. The OR of a pool of truncating variants in loss of function susceptibility genes
113 provides a "standard candle" by which one can measure the magnitude of effect of a
114 given analytically defined pool of variants.
115 2. An analytically defined pool of variants with an OR similar to that of a pool of
116 truncating variants consists almost entirely of variants that are pathogenic with an
117 effect size approximately equal to that of the truncating variants.
118 3. An analytically defined pool of variants with an OR near 1.0 consists almost entirely
119 of benign variants.
120 4. Analytically defined pools of variants with an OR between 1.0 and the OR of
121 truncating variants can be modeled as a two-component mixture of benign and
122 pathogenic variants. This assumption is debatable, but importantly, is also testable.

123 The full derivation of the maximum likelihood estimate model is presented in the
124 Supplemental Methods. In brief, our model estimates the proportion of observed variants that
125 are pathogenic, *q*, from the proportion of all variants that are pathogenic, *p*, the number of
126 cases, *n*, the number of controls, *m*, the population frequency of disease among individuals who

127 do not carry a pathogenic variant, δ, and the frequency of disease among individuals who are

128 carriers of a pathogenic variant, δ_1, as shown in Equation 1:

129 $$\hat{q} = \frac{\hat{p}\left[n\hat{\delta}(1-\delta_1) + m(1-\hat{\delta})\delta_1\right]}{\hat{p}\left[n\hat{\delta}(1-\delta_1) + m(1-\hat{\delta})\delta_1\right] + (1-\hat{p})\hat{\delta}(1-\hat{\delta})(n+m)} \quad (1)$$

130

131 In addition to the variables listed above, our MLE model requires input values for an

132 odds ratio threshold at which all variants are assumed to be pathogenic. For the odds ratio

133 threshold in our analysis, truncating variants observed in our case-control dataset were

134 tabulated and frequentist odds ratios for breast and ovarian cancers were calculated for these

135 groups of variants (16.2, 95% CI: 12.7 – 20.8 for breast cancer and 46.4, 95% CI: 35.7 – 60.7

136 for ovarian cancer). To evaluate the prediction of benign variants we used an alternative OR

137 ceiling chosen to match the threshold of moderate-risk variants (OR = 2.0 for breast cancer and

138 OR = 2.9 for ovarian cancer), which was calculated to be inflated to 2.31 and 3.83 in our family

139 history enriched case-control dataset [8, 9]. The general population frequency of cancer used in

140 the analysis was 0.125 for breast cancer, and 0.011 for ovarian cancer [10, 11].

141

142 Identification of Exceptionally Conserved Ancestral Residues in BRCA1 / BARD1

143 To find evolutionarily distant BRCA1/BARD1 homologs, we started with two alignments,

144 one of the RING domains of BRCA1 and BARD1 from *Homo sapiens* and *Strongylocentrotus*

145 *purpuratus* and a second of the corresponding tandem BRCT (tBRCT) domains. These

146 alignments were prepared using the structure-guided 3D-Coffee/Expresso mode from the T-

147 coffee suite of sequence alignment tools [12, 13]. The two resulting alignments were used with

148 hmmsearch to scan phylogenetically defined subsets of the reference proteome database

149 maintained at the HMMER website [14]. Hits were defined as gene models with high-scoring

150 matches to both the RING and tandem BRCT alignments.

151 At the first round, hits from Placozoa and single celled Holozoans were included, the
152 alignments rebuilt, and then trimmed back to include clear *BRCA1* and *BARD1* orthologs from
153 *Trichoplax* and *Salpingoeca*. The process was iterated to obtain and include homologs from the
154 Fungi and then reiterated to obtain and include homologs from three additional eukaryotic
155 supergroups: Archaeplastida, SAR, and Discoba.
156 To prepare the final alignments and then define exceptionally conserved ancestral
157 residues (ECARs), we stratified the hits into a group of *BARD1* and relatively *BARD1*-like hits
158 and a second group of *BRCA1* and relatively *BRCA1*-like groups; ambiguous hits were placed in
159 the *BRCA1*-like group. As before, RING and tBRCT alignments were prepared using 3D-
160 Coffee/Expresso. Then the BRCA1 and BARD1 alignments were aligned to each other by
161 introducing gaps across the whole BRCA1 or whole BARD1 alignment, i.e., without modifying
162 the within-group alignments. The tBRCT alignments were then appended to the RING domain
163 alignments. Finally, ECARs were defined as (a) positions that were invariant in our existing
164 metazoan BRCA1 and BARD1 alignments used with Align-GVGD [15, 16], and either (b) also
165 invariant in the new multi-supergroup alignments or (c) had just one conservative substitution
166 (i.e., Grantham Variation <65) in the new multi-supergoup alignment.
167
168 Functional Assay Results and Computational Predictions
169 Functional assay scores were obtained from three separate assays that interrogate
170 missense variants in various domains of *BRCA1*. These assays were: the 2018 Findlay et al.
171 saturation genome editing (SGE) assay that interrogated missense and splice junction variants
172 in the RING and tBRCT domain of *BRCA1* [17], the 2022 Clark et al. mammalian-2-hybrid (M2H)
173 assay performed by our laboratory that evaluated missense variants in the RING domain of
174 *BRCA1* [7], and the 2023 Nagy et al. homology-directed repair (HDR) assay that evaluated
175 missense variants in the coiled-coil domain of *BRCA1* [18]. The M2H and SGE assays were
176 selected due to their nearly comprehensive evaluation of all possible snMS variants in the RING

177 and/or tBRCT domains of BRCA1. The HDR assay was selected because it interrogated the
178 coiled-coil domain which is considered a clinically import functional domain[19]. Scores for these
179 assays were downloaded from MaveDB (SGE and M2H) [20] or from the publication [18].
180 Thresholds for functional, indeterminate and non-functional variants were used as defined by
181 the relevant publication. We used the calibrated scores for each assay in our analysis because
182 these are the thresholds that are likely to be used during clinical variant classification.
183 Computational tool scores for BayesDel no-AF [21], REVEL [22], and VEST4 [23] were
184 obtained from the database for Non-synonymous Functional Predictions (dbNSFP) via
185 Ensembl's web-based Variant Effect Predictor (VEP) [24, 25]. Computational scores for MutPred2
186 were obtained by downloading a local copy of the software[26]. Scores for AlphaMissense were
187 downloaded from the original publication [27]. Scores for Align-GVGD (A-GVGD) [28, 29] were
188 obtained from agvgd.hci.utah.edu using the Human to Sea Urchin alignment. Thresholds for
189 pathogenic and benign for AlphaMissense, BayesDel, MutPred2, REVEL, and VEST4 were the
190 supporting pathogenic and supporting benign thresholds from recent calibrations of these tools[5,
191 30]. Align-GVGD, predictions with a score of C0 were interpreted as predicted benign,
192 intermediate scores as indeterminate, and scores of C65 were interpreted as predicted
193 pathogenic. AlphaMissense, BayesDel, MutPred2, REVEL, and VEST4 were selected because
194 they demonstrated the ability to provide strong evidence in favor of pathogenicity in recent
195 calibration studies[5, 30] and some of these tools are often used by Variant Curation Expert Panels
196 (VCEP). While the scales for each tool may differ, the individual thresholds used in this work are
197 those established in Pejaver et al. to be sufficient for supporting evidence for both benign and
198 pathogenic variants in the ACMG classification framework. A-GVGD has previously been
199 demonstrated to perform well in *BRCA1*[15].
200
201 <u>Estimate of ACMG Points</u>

202 We interpreted the MLE proportion pathogenic as a posterior probability, PP
203 (represented as *q* in the supplemental text and Equation 1), of the variants in the analytical
204 subset being pathogenic and we thus converted this estimate to the appropriate ACMG point
205 value via an OP using a prior probability (P_1) of 0.102[4]. Proportions pathogenic were calculated
206 separately for each cancer type using our MLE model, converted to odds of pathogenicity using
207 Equation 2 below, and then combined using a weighted geometric mean based on the number
208 of individuals in the breast cancer case control group and ovarian cancer case control group.
209 This new estimate of odds of pathogenicity was subsequently converted back to a proportion
210 pathogenic to display in Tables 2 – 4 and converted to ACMG points using Equation 3[4]. When
211 assigning ACMG points from the estimated proportion pathogenic, the points that could be
212 awarded were limited to the maximum values prescribed in the ACMG guidelines unless
213 otherwise noted [2, 4].
214
215
216 $$PP = (OP \times P_1) \div [(OP - 1) \times P_1 + 1] \quad (2)$$
217 $$OP = 81^{\frac{ACMG\ points}{6}} \quad (3)$$
218
219
220 Case and Control Count Sources and Odds Ratio Calculations
221 Human subjects' data were obtained from Ambry Genetics and Myriad Genetics. Control
222 counts are from the non-cancer cohort of gnomAD v2 exomes and gnomAD v3 genomes[31, 32].
223 Ethnicity for this dataset was split into four groups: non-Finnish European (NFE), Ashkenazi
224 Jewish, Finnish, and all others. Ashkenazi Jewish and Finnish individuals were removed due to
225 strong founder effects in these populations. Individuals from Ambry and Myriad were considered
226 breast cancer cases if they had breast cancer and did not have ovarian cancer. Individuals from

227 Ambry and Myriad were considered ovarian cancer cases if they either had ovarian cancer or
228 ovarian and breast cancer.

229 The Ambry dataset initially included data generated by Ambry Genetics from multi-gene
230 sequencing panels for 165,031 individuals "exempted from review by the Western Institutional
231 Review Board" [33]. Filters were applied to remove those individuals without breast or ovarian
232 cancer, individuals who lacked *BRCA1* sequence data, individuals of Ashkenazi Jewish or
233 Finnish ancestry, and individuals with pathogenic or likely pathogenic variants in *TP53*, *BRCA2*,
234 and *PALB2*. This resulted in a final dataset of 98,111 individuals, 66,672 (68%) of which were of
235 NFE ethnicity and 31,439 (32%) of other ethnicity. Of individuals of NFE ethnicity, 57,431 (86%)
236 had breast cancer and 9,241 (14%) had ovarian cancer. Of individuals of other ethnicity, 28,114
237 (89%) had breast cancer and 3,325 (11%) had ovarian cancer. 2,970 individuals had variants in
238 *BRCA1*: 1,266 (43%) individuals of NFE ethnicity who had breast cancer, 452 (15%) individuals
239 of NFE ethnicity who had ovarian cancer, 1,013 (34%) individuals of other ethnicity who had
240 breast cancer, and 239 (8%) individuals of other ethnicity who had ovarian cancer. 477
241 individuals had missense variants that fell within the RING (amino acids 1-100), coiled-coil
242 (amino acids 1,280-1,576), or tBRCT domain (amino acids 1,650-1,863): 118 (25%) individuals
243 of NFE ethnicity who had breast cancer, 155 (32%) individuals of other ethnicity who had breast
244 cancer, 123 (26%) individuals of NFE ethnicity who had ovarian cancer, and 81(17%) individuals
245 of other ethnicity who had ovarian cancer.

246 The Myriad dataset initially included data generated by full-sequence BRCA1/2 tests
247 performed by Myriad Genetics from 60,607 individuals. For these, a test request form must have
248 been completed by the ordering health care provider, and the form must have been signed by
249 an appropriate individual indicating that "informed consent has been signed and is on file" [15, 33,
250 34]. Filters were applied to remove those individuals without breast or ovarian cancer, individuals
251 who lacked *BRCA1* sequence data, individuals of Ashkenazi Jewish or Finnish ancestry, and
252 individuals with pathogenic or likely pathogenic variants in *BRCA2*. This resulted in a final

253 dataset of 35,088 individuals, 25,653 (73%) of which were of NFE ethnicity and 9,435 (27%) of
254 other ethnicity. Of individuals of NFE ethnicity, 22,504 (88%) had breast cancer and 3,149 (12%)
255 had ovarian cancer. Of individuals of other ethnicity, 8,347 (88%) had breast cancer and 1,088
256 (12%) had ovarian cancer. 4,005 individuals had variants in *BRCA1*: 2,231(56%) individuals of
257 NFE ethnicity who had breast cancer, 639 (16%) individuals of NFE ethnicity who had ovarian
258 cancer, 904 (22%) individuals of other ethnicity who had breast cancer, and 231 (6%) individuals
259 of other ethnicity who had ovarian cancer. 524 had missense variants that fell within the RING,
260 coiled-coil, or tBRCT domain: 289 (55%) individuals of NFE ethnicity who had breast cancer, 83
261 (16%) individuals of NFE ethnicity who had ovarian cancer, 124 (24%) individuals of other
262 ethnicity who had ovarian cancer, and 28 (5%) individuals of other ethnicity who had ovarian
263 cancer.

264 The gnomAD dataset is a composite of non-overlapping v2 non-cancer exomes and v3
265 non-cancer genomes that initially included 192,502 individuals. Filters were applied to remove
266 individuals of Finnish or Ashkenazi Jewish ancestry and to remove common missense variants
267 (allele frequency > 0.000271). This resulted in a final dataset of 169,933 individuals: 84,396
268 (50%) individuals of NFE ethnicity and 85,537 (50%) individuals of other ethnicity. 644
269 individuals had missense variants that fell within the *BRCA1* RING, coiled-coil, or tBRCT
270 domain: 264 (41%) individuals of NFE ethnicity and 380 (59%) individuals of other ethnicity.

271 Combining these datasets resulted in a final dataset of 133,199 breast and ovarian
272 cases from Ambry and Myriad and 169,933 controls from gnomAD v2 and v3.

273 After filtering, the data for *BRCA1* were formatted into a table that indicated the ethnicity
274 of the individual, whether an individual was a case or control, how the variant was classified by
275 each functional assay and computational tool, and whether the variant fell at an ECAR.

276 Logistic regression was performed using R studio to calculate frequentist odds ratios for
277 each stratum (functional, indeterminate, loss-of-function (LOF) all, LOF-ECAR, LOF-other) of
278 the functional assays included in our analysis. Odds ratios were calculated similarly for strata

279 that resulted in combining SGE functional assay results and computational tool predictions.
280 Proportions pathogenic were calculated using a custom R script. Following logistic regression,
281 the Wald test was performed to determine statistical significance between strata in each
282 analysis.

283 To determine if combining computational tool predictions to functional assay predictions
284 was statistically significant, multivariate logistic regression was performed using combinations of
285 the SGE functional assay result (functional or LOF) and computational tool predictions (benign
286 or pathogenic) for each tool analyzed in our study.

287

288 **RESULTS**:

289 Exceptionally Conserved Ancestral Residues in *BRCA1* / *BARD1*

290 *BRCA1*, *BARD1*, and homologs that have been studied in the Archaeplastida are
291 thought to have evolved from a common ancestor gene with a 5' RING domain and 3' tandem
292 BRCT (tBRCT) domain that was present in an ancient eukaryote before the divergence of
293 Amorphea and Archaeplastida [35]. Protein multiple sequence alignment (PMSA) based
294 hmmsearch [14] identified clear *BRCA1* and *BARD1* orthologs in *Trichoplax* and more distantly
295 related single-celled members of Holozoa, plus homologs (some of which may be true
296 orthologs) in the Holomycota, Archaeplastida, SAR, and Discoba. Although the root of the
297 eukaryotic phylogeny remains unclear [36, 37], the resulting PMSA should be sufficient to identify
298 positions in human *BRCA1* and *BARD1* where the amino acid present in the two human
299 sequences has been inherited identical by descent from a common ancestor gene and is under
300 constraint in the Amorphea and 2-3 additional eukaryotic supergroups. We identified 14 such
301 residues: the 8 residues that define the C3HC4 RING motif plus 6 additional residues in the
302 tBRCT domain (Table 1, Figure S1), and hereafter refer to them as Exceptionally Conserved
303 Ancestral Residues (ECARs).

304 Of the 82 possible snMSs at ECARs, 78 resulted in a non-functional assay result via the
305 SGE assay. The four exceptions were C24F, C61R, T1700S (which was not interrogated in the
306 SGE assay), and W1718S. C24F, C61R, and W1718S had indeterminate results in the SGE
307 assay, however, C24F and C61R were scored as loss of function in the M2H assay.
308 Using the pathogenic strong thresholds established in Pejaver et al, we also tabulated
309 the number of substitutions at ECARs that were extreme enough to receive +4 ACMG points
310 from computational tools (Table 1). MutPred2 predicted 0 of the possible ECAR substitutions to
311 be extreme enough to receive +4 ACMG points. BayesDel predicted 56 of the 82 possible
312 substitutions to be extreme enough for +4 ACMG points, and these constituted 56 of the 73
313 such RING and tBRCT +4 results called by BayesDel. AlphaMissense, REVEL, and VEST4 had
314 totals that fell between MutPred2 and BayesDel.
315 A separate analysis was performed using the BayesDel threshold prescribed in the
316 ENIGMA BRCA1 VCEP (henceforth Bayes-Del VCEP) guidelines (≥ 0.28) as well as for Align-
317 GVGD using C65 as a pathogenic prediction threshold[19]. Under these parameters, which only
318 generate predictions of supporting benign, indeterminate, or supporting pathiogenic, many more
319 of the ECARs identified in the tBRCT domains were predicted to be pathogenic (Table 1). A
320 summary of computational tool performance at ECAR positions and individual scores for each
321 substitution are available in Table S1 and the supplemental spreadsheet.
322 Functional Assay Performance
323 To analyze the performance of functional assays, an OR threshold set to that of the OR
324 of nonsense substitutions observed in our case control dataset for each cancer type was used,
325 which was 16.2 (95% CI: 12.7 – 20.7) for breast cancer and 46.4 (95% CI: 35.7 – 60.7) for
326 ovarian cancer. The combined proportion pathogenic of the breast cancer and ovarian cancer
327 analysis and subsequent ACMG points of functional assay performance using the nonsense
328 substitution OR thresholds are presented in Table 2. ORs and proportions pathogenic by
329 cancer type are presented in the supplemental text. As stated in the Methods section, ACMG

330 points awarded are limited to the maximum number of points outlined in the ACMG guidelines
331 and subsequent transformation to a Bayesian framework.
332 In the M2H assay, which interrogates variants in the RING domain of BRCA1, variants
333 that were functional and were observed in our case control dataset yielded an adjusted OR
334 close to 1 (1.42, 95% CI: 1.01 – 1.99 for breast cancer and 2.45, 95% CI: 1.41 – 4.24 for
335 ovarian cancer) indicating a group of variants composed almost entirely of benign variants. The
336 combined proportion pathogenic for this group was estimated to be 0.04 which translates to -1
337 points in the ACMG points system. There were insufficient observations for variants that were
338 categorized as indeterminate to compute an OR or a proportion pathogenic. The group of all
339 variants that were non-functional and observed in our dataset had an adjusted OR of 19.9 (95%
340 CI: 11.3 – 35.0) for breast cancer and 40.0 (95% CI: 21.5 – 73.1) for ovarian cancer which is
341 similar to that of nonsense substitutions for each cancer type in our dataset, indicating a group
342 consisting of mostly pathogenic variants. This analytical subset had a combined proportion
343 pathogenic of 0.99, which would be assigned the full +4 points that are available for non-
344 functional assay results in the ACMG guidelines.
345 In addition to evaluating the strength of evidence of a single evidence category, our
346 model can also test the strength of combined evidence categories. To this end, the non-
347 functional variants from the M2H assay were stratified into two groups: variants that fell at the
348 ECARs identified from the PMSA (ACMG code: PM1), and all other residues. For variants that
349 were non-functional and fell at an ECAR, the computed OR was 22.0 (95% CI: 12.0 – 40.6) for
350 breast cancer and 40.8 (95% CI: 21.1 – 79.0) for ovarian cancer, slightly higher than the group
351 of all non-functional variants and indicating a subset of variants with similar risk to that of
352 truncating variants. The computed combined proportion pathogenic was 1.00. Combining the
353 +4 points from functional assay and +2 points available from ECARs (PM1), results in +6 points
354 for this group of variants. Thus, using these two evidence criteria alone, these variants reach
355 the Likely Pathogenic threshold.

356 The computed OR for non-functional variants that fell at other residues was 7.82 (95%
357 CI: 1.68 – 36.3) for breast cancer and 32.9 (95% CI: 6.37 – 170.1) for ovarian cancer, indicating
358 that this group of variants confers less risk than those that are non-functional and fall at an
359 ECAR. The combined proportion pathogenic for the non-functional variants that fell at other
360 residues was 0.70, sufficient for +4 points from functional assay. In the absence of other
361 evidence, this group would remain variants of uncertain significance (VUS).
362 The SGE assay, which evaluated variants across the RING and tBRCT domains of
363 *BRCA1*, had similar results to the M2H assay. The group of variants that were functional via
364 SGE yielded an OR close to 1 (1.11, 95% CI: 0.91 – 1.36 for breast cancer and 1.34, 95% CI:
365 0.90 – 2.01 for ovarian cancer) and a combined proportion pathogenic of 0.00, which would
366 qualify for the full -4 points available from a functional assay result. For variants that were non-
367 functional, the SGE assay had an OR of 15.7 (95% CI: 11.1 – 22.0) for breast cancer and 35.1
368 (95% CI: 24.0 – 51.2) for ovarian cancer and a combined proportion pathogenic of 0.92 which
369 would confer the full +4 points for these variants. Stratification of non-functional variants at
370 ECARs versus non-functional variants at other residues again demonstrates that non-functional
371 variants at ECARs have enough strength of evidence to reach the Likely Pathogenic threshold.
372 The ORs for non-functional variants at ECARs were 23.6 (95% CI: 12.9 – 43.3) for breast
373 cancer and 42.3 (95% CI: 21.9 – 81.7) for ovarian cancer with a combined proportion
374 pathogenic of 1.00, sufficient for +6 ACMG points. The ORs for non-functional variants at non-
375 ECAR positions were 12.2 (95% CI: 8.07 – 18.5) for breast cancer and 31.7 (95% CI: 19.9 –
376 50.5) for ovarian cancer with a combined proportion pathogenic of 0.76, which qualifies for +4
377 ACMG points. Moreover, the risk estimates for non-functional variants at ECAR versus non-
378 ECAR positions were distinct from each other, P=0.031 in a two-sided 3x2 FET test.
379 Whereas the M2H and SGE assays largely confirmed the assumptions of strength of
380 evidence in the ACMG guidelines, the results from the HDR assay in the coiled-coil domain of
381 *BRCA1* produce a different result. The ORs for the functional group of variants were 1.87 (95%

382 CI: 1.11 – 3.15) for breast cancer and 2.04 (95% CI: 0.77 – 5.39) for ovarian cancer with a
383 combined proportion pathogenic of 0.08 translating to 0 ACMG points. The variants that were
384 non-functional via the HDR assay had an OR of 1.58 (95% CI: 0.09 – 26.8) for breast cancer
385 and insufficient observations to calculate an OR for ovarian cancer, with an identical proportion
386 pathogenic of 0.08 and 0 ACMG points. It is noted that there were few observations in our
387 case-control dataset that both fell within this domain and were included in the functional assay.
388 Additionally, many of the variants interrogated by this assay do not appear to be attainable via a
389 single nucleotide substitution to the canonical *BRCA1* transcript, which is the type of substitution
390 that is the primary focus of this analysis.

391 Due to these limitations, a separate analysis was performed focusing on structurally
392 informed predictions of damage to coiled-coil interactions at the alpha helix spanning amino
393 acids 1,400-1,418, i.e., non-conservative substitutions at the helix *a* and *d* positions plus any
394 substitutions to proline falling in the same interval. With only one observation in a case
395 (p.L1407R) against four control observations (one each for p.M1400K, p.L1404R, p.L1407H,
396 and p.L1414P), the observational evidence weighs against the hypothesis that missense
397 substitutions in this domain are pathogenic (Table 2, final line).

398 Combining SGE Functional Results and Computational Predictions

399 To explore the impact of combining functional assay results and computational tool
400 predictions variants from the SGE assay were stratified by computational prediction
401 (pathogenic, indeterminate, and benign) from seven computational tools (Align-GVGD,
402 AlphaMissense, BayesDel, BayesDel-VCEP, MutPred2, REVEL, and VEST4) using thresholds
403 outlined in Pejaver *et al*.[5, 30] or those stated in the Methods section. This analysis was
404 performed only on results from the SGE assay which interrogated both the RING and tBRCT
405 domains of *BRCA1* and thus had sufficient observations for stratification. To focus on the impact
406 of combining functional assays and computational tool predictions on variant classification, the
407 combined proportions pathogenic and subsequent ACMG points are presented in Tables 3 and

408 4. ORs and proportions pathogenic for each stratum and cancer type are presented in the
409 supplemental text (Tables S3 – S6. Multivariate logistic regression was also performed to
410 determine if adding computational tool predictions to functional assay results was statistically
411 significant (Table S7).

412 Combining Assay Results of Non-functional with Computational Predictions for Substitutions
413 Falling at Non-ECAR Positions

414 The proportion pathogenic for the group of variants that were categorized as non-
415 functional via the SGE assay and did not fall at ECARs was 0.76, corresponding to +4 ACMG
416 points, as stated previously. The ORs for this group of variants for breast and ovarian cancer are
417 12.2 (95% C.I. 8.07 – 18.5) and 31.7 (19.9 – 50.5) respectively. Stratifying these variants by
418 computational prediction allows evaluation of the effect of combining functional assay and
419 computational prediction data toward variant classification (Table 3). It should be noted that if
420 the computational predictions unanimously agreed with the functional assay result, the OR,
421 estimated proportion pathogenic, and ACMG points would remain unchanged after adding in
422 evidence from computational tools. Thus, any utility of computational predictions in this analysis
423 emerges from the ability of the computational tool to accurately disagree with – and thereby
424 stratify – the functional assay result.

425 For four of the computational tools evaluated, Align-GVGD, AlphaMissense MutPred2,
426 and VEST4, the benign and/or indeterminate prediction yielded a lower proportion pathogenic
427 and lower ACMG point score than the overall pool of SGE non-ECAR, non-functional
428 substitutions. For Align-GVGD and AlphaMissense, this effect of accurate disagreement was
429 strong enough that concordant prediction of pathogenicity from the SGE assay and
430 computational tool resulted in a higher proportion pathogenic and higher ACMG point score
431 (+5). Align-GVGD was the only computational tool that successfully stratified the group of non-
432 functional assay results into three corresponding strata that matched the ACMG expectations.
433 The corresponding ORs for concordant predictions of pathogenicity were 13.9 (95% C.I. 7.39 –

434 26.1) for breast cancer and 40.2 (95% C.I. 20.1 – 80.3) for ovarian cancer. When Align-GVGD
435 disagreed with the functional assay and predicted variant to be benign the ORs were 6.06 (95%
436 C.I. 1.98 – 18.5) and 22.1 (95% C.I. 6.58 – 74.2) for breast and ovarian cancer respectively.
437 In contrast, BayesDel and REVEL produced an unexpected reverse stratification, with a
438 higher proportion pathogenic and higher ACMG point score in their indeterminate or benign
439 prediction than in their pathogenic prediction. Correlation between these two - even in failure - is
440 not unexpected as both are meta-callers, incorporating predictions from individual callers into a
441 weighted overall score, and share several of the same individual components [21, 22].
442 Combining Assay Results of Functional with Computational Predictions for Substitutions Falling
443 at Non-ECAR Positions
444 For the analysis of variants that were functionally normal via SGE and did not fall at
445 ECARs, an alternative OR ceiling was set at the threshold for moderate-risk variants to evaluate
446 evidence against pathogenicity (OR = 2.0 for breast cancer and 2.9 for ovarian cancer in a
447 general population[8, 9], which are inflated to 2.31 and 3.83 respectively in our dataset). For the
448 SGE assay, functional variants had an estimated proportion pathogenic of 0.00, sufficient for the
449 -4 ACMG points allowable from functional assays. The OR for breast cancer for this group of
450 variants was 1.11 (95% C.I. 0.91 – 1.36) and 1.34 (95% C.I. 0.90 – 2.01) for ovarian cancer.
451 The results of stratification by computational tool are presented in Table 4; as in Table 3, the
452 utility of computational predictions in this analysis emerges from the ability of a computational
453 tool to accurately disagree with – and thereby stratify – the functional assay result. In this
454 analysis, all of the computational tools were to some extent successful, with their predictions of
455 pathogenicity yielding a higher proportion pathogenic and higher ACMG point score than the
456 overall pool of SGE functional substitutions.
457 However, BayesDel and REVEL made very few predictions of benignity even though this
458 group of variants were functional in the SGE assay and had ensemble odds ratios and
459 proportions pathogenic consistent with the vast majority actually being benign. On the other

460 hand, for Align-GVGD, AlphaMissense, BayesDel (with and without the VCEP calibration), and
461 VEST4 the effect of discordant prediction of benignity from the SGE assay and pathogenicity by
462 the computational tool was a substantial pool of missense substitutions with an ACMG point
463 score of +1 or higher – though it must be held in mind that these positive scores represent
464 modest evidence in favor of pathogenicity as moderate-risk variants, not as high-risk variants.
465 This is evidenced by the magnitude of the ORs generated for the group of variants predicted to
466 be pathogenic by Align-GVGD: 2.71 (95% C.I. 1.09 – 6.72) for breast cancer and 3.73 (95% C.I.
467 0.77 – 18.1) for ovarian cancer. Corresponding ORs for the other tools evaluated are presented
468 in the supplemental text (Tables S5 – S6).

469 Finally, results from multivariate logistic regression substantiate the observed trend in
470 proportions pathogenic and ACMG points provided by computational tool stratification.
471 Predictions from three tools, Align-GVGD, AlphaMissense, and VEST4 were statistically
472 significant (p-values low enough to overcome the intrinsic multiple test) when added to
473 functional assay results. Predictions from BayesDel, BayesDel-VCEP, MutPred2, and REVEL,
474 did not improve the regression model compared to a model that included only the functional
475 assay (Table S7). Note that while adding the results of some computational tools as main effects
476 significantly improved the ability of the model to predict pathogenicity, adding interaction terms
477 between the assay result and the computational tool prediction did not. This indicates that the
478 predictive power of the computational tool is the same regardless of the assay result (wild-type
479 of LOF). Additionally, these logistic regressions did not impose an order on the predictions of the
480 computational tool (pathogenic or benign), however, in general we see that the resulting
481 coefficients reflect our expectations that pathogenic predictions are more indicative of
482 pathogenicity than are benign predictions. Thus, we conclude that Align-GVGD and
483 Alphamissense had the best-balanced performance of variant stratification by providing
484 statistical significance to regression models and meeting the ACMG additivity expectations, as
485 summarized in Table 5.

486 Example VUS Evaluations

487 To explore the impact that joint calibration of the SGE functional assay and
488 computational tool prediction might have on variant classification, we work through several
489 examples from ClinVar and/or gnomAD in Table 6. For this analysis, only substitutions with
490 concordant predictions from Align-GVGD and AlphaMissense are considered as these tools
491 performed the best in our analysis thus far. The substitutions selected include one that reaches
492 Likely Pathogenic, three that cover the VUS range from Hot to Cold[38], and one that reaches
493 Likely Benign. For two of these, p.Asp1840Ala and p.Arg1737Ile, the results from naïve
494 summation of functional and computational points scores would be Likely Benign, whereas the
495 results from joint calibration of functional and computational evidence are VUS.

496 **DISCUSSION**:

497 Here, we derived a maximum likelihood estimate (MLE) model that links the OR for an
498 analytically defined pool of sequence variants within a case-control analysis to an estimate of
499 the proportion of variants within that pool that are pathogenic. Bayes rule is then used to
500 determine the odds of pathogenicity associated with the analysis that defined the pool and
501 thence ACMG points. We also defined Exceptionally Conserved Ancestral Residues (ECARs),
502 a concept that fits within the ACMG PM1 evidence criterion, and then identified ECARs within
503 the RING and tBRCT domains of *BRCA1*. The MLE model was then used to evaluate the
504 strength of evidence for several functional assays alone, in combination with ECAR data, and in
505 combination with computational tool predictions, as summarized in Figure 2. This work takes an
506 important step in the ability to empirically test whether assertions of the strength of evidence for
507 different evidence types, and combinations thereof, proposed in the 2015 ACMG/AMP
508 guidelines are valid. And it does so in a way that takes account of effect size while avoiding
509 circularities inherent in analyses that are based on re-call.

510 Results from three *BRCA1* functional assays were examined: a saturation genome
511 editing (SGE) assay that examined both the RING and tBRCT domain, a mammalian 2-hybrid

512 (M2H) assay that examined the RING domain, and a homology directed repair (HDR) assay that
513 examined the coiled-coil domain through which BRCA1 interacts with PALB2. With enormous
514 differences in the case-control ORs calculated for non-functional versus functional missense
515 substitutions, the SGE and M2H assays both provide significant evidence towards sequence
516 variant classification, which will be discussed in greater detail below. In contrast, there was no
517 OR-based evidence that loss of function in the coiled-coil HDR assay predicted increase risk of
518 breast or ovarian cancer; moreover, null results from a structurally informed analysis of
519 predicted loss of function substitutions in the core alpha helix of the coiled-coil domain (amino
520 acids 1,400 – 1,418) weighs against the hypothesis that missense substitutions in this domain
521 can be pathogenic. Therefore, careful evaluation of functional assays is required prior to their
522 use in clinical settings and +/- 4 ACMG points should not be assigned carte blanche[6].
523 Using an evolutionarily deep multi-supergroup protein multiple sequence alignment of
524 BRCA1, BARD1, and related RING-tBRCT proteins, we identified 14 ECARs in *BRCA1* (and
525 *BARD1*). Stratification of ECAR substitutions by *BRCA1* MAVE functional assay results,
526 followed by analysis through the MLE model, led to three interesting observations. (1) Of 82
527 possible missense substitutions at an ECAR, 78 had loss of function in the SGE; the remaining
528 four were either indeterminate or not assessed by the SGE assay, and 2 of these had loss of
529 function in the M2H assay. (2) Loss of function ECAR substitutions had an OR very similar to
530 that of *BRCA1* nonsense substitutions; this is an example of the expected additivity between
531 evidence criteria, consistent with +6 ACMG points for the combination. (3) The 326 non-
532 functional substitutions at non-ECAR positions also have an elevated OR, but significantly lower
533 than that observed for substitutions at ECAR positions, corresponding to +4 points.
534 With the recent publication of ACMG points-based calibration of several computational
535 tools, including AlphaMissense, it is of interest to examine results from combining computational
536 tool evidence with functional assay evidence. Since ECARs fit within the PM1 evidence
537 criterion, yet should ideally be captured by computational tools, we avoided double counting

538 ECAR evidence by applying the computational tools to functional assay results at the non-
539 ECAR RING and tBRCT positions.

540 For the set of 326 non-ECAR RING and tBRCT substitutions with SGE assay loss-of-
541 function, the expected result is that computational tool evidence in favor of pathogenicity will
542 result in an analytic subset with a higher OR, adding an ACMG point(s) to the score of the input
543 data set. Computational tool evidence in favor of benignity should then result in an analytic
544 subset with a lower OR, subtracting an ACMG point(s) from the score of the input data set. Of
545 the six computational tools tested, only two – Align-GVGD and AlphaMissense – generated this
546 bidirectional pattern. Two other tools – MutPred2 and VEST4 – generated negative points with
547 their indeterminate and/or evidence of benignity outputs, but did not generate positive points
548 with their evidence of pathogenicity. It should be noted that for these non-ECARs substitutions,
549 concordant evidence of pathogenicity from the functional assay and either Align-GVGD or
550 AlphaMissense resulted in a score of +5 points, meaning that additional evidence in favor of
551 pathogenicity – perhaps from co-segregation or tumor features as was the case with *BRCA1*
552 p.Pro34Ser[7] – would be required for the variants to reach ≥ +6 points and thus Likely
553 Pathogenic. Under the computational tool score thresholds established in Pejvaer et al.[5, 30] for
554 Pathogenic Supporting (+1 ACMG point), most tools added 0 ACMG points when combined with
555 loss of function results (except for AlphaMissense). When these thresholds were changed to
556 that of Pathogenic Moderate (+2 ACMG points), no tool added +2 ACMG points on top of the
557 functional assay result (Table S8), which is another example of departure from simple additivity
558 assumed in the ACMG guidelines.

559 Under the logic of both the IARC/ENIGMA/InSiGHT and the Bayesian interpretation of
560 the ACMG/AMP sequence variant classification systems, evidence against pathogenicity has
561 been interpreted as evidence in favor of benignity. At the same time, both ENIGMA (Evidence-
562 based Network for the Interpretation of Germline Mutant Alleles) and the BRCA1/2 VCEP
563 (Variant Classification Expert Panel) recognize the existence of moderate-risk susceptibility

genes and moderate-risk variants of high-risk susceptibility genes (also called Reduced Penetrance Pathogenic Variants, RPPV)[8, 19, 39]. To reduce the likelihood of generating evidence in favor of benignity for missense substitutions that might eventually be classified as RPPVs, we made the potentially controversial choice to estimate proportions pathogenic and ACMG points for variants with normal function in the functional assays using moderate-risk breast and ovarian cancer thresholds.

For the set of 1,284 RING and tBRCT missense substitutions that were functional in the SGE assay, the ORs for breast and ovarian cancer were near 1.0, consistent with -4 ACMG points, even if calculated with the moderate-risk threshold. Stratification of these with the set of six computational tools revealed two trends plus an analytic difficulty. First, two of the computational tools – BayesDel and REVEL – returned scores in favor of benignity for only a very small fraction of these missense substitutions: 3% by BayesDel and <1% by REVEL. We interpret this as a failure within these two tools' algorithms. Second, three of the tools – Align-GVGD, AlphaMissense, and VEST4 – generated a bidirectional stratification, with higher ORs and ACMG point scores associated with computational evidence of pathogenicity and lower ORs and point scores associated with evidence of benignity. We interpret this as a subset of the computational tools adding useful evidence towards variant classification on top of the functional assay result. The analytic difficulty is that modest changes in the OR produced large changes in the ACMG point scores; when the SGE assay and computational tool were discordant, combining the two produced point scores that were beyond the range expected from simple additivity.

The individual sequence variants evaluated in Table 6 provide examples for many of the trends described above. For example, functional assay loss of function at an ECAR, with concordant computational evidence, (p.Thr1700Ile) results in Likely Pathogenic. On the other hand, loss of function at a non-ECAR position with concordant computational evidence in favor of pathogenicity (p.Ile15Thr) results in a hot VUS, and further evidence of pathogenicity would

590 be required to reach Likely Pathogenic. In either scenario, evidence in favor of benignity that
591 might arise from human subjects data or other evidence could move variants with these
592 evidence combinations from Likely Pathogenic and hot VUS to hot or cool VUS respectively. At
593 the other end of the spectrum, and using the moderate-risk OR threshold for calibration to
594 reduce the likelihood of classifying RPPVs as likely benign, an assay result of functional with
595 concordant computational evidence of benignity (p.Ala1823Val) results in Likely Benign. But
596 discordant results between a calibrated functional assay and calibrated computational tool
597 (p.Asp1840Ala, p.Arg1737Ile) result in cool to cold VUS – and a concern that the computational
598 tools need to be accurately calibrated in order to avoid this classification result for too many
599 missense substitutions.
600 Like most studies, this one had several limitations. A challenge to the ECARs
601 component of the approach pursued here is variable quality of assembly and subsequent
602 annotation of reference genomes. Thus, searches against available databases may give a
603 misleading view of which organisms outside of Holozoa encode genes with both RING and
604 tBRCT domains. Furthermore, this study was conducted on *BRCA1*, a gene that probably
605 shares ancestry with *BARD1*, these two representing Holozoan examples of a gene family that
606 has undergone a modest level of evolutionary diversification. It is not known if this method will
607 be applicable to genes which are members of gene families that have undergone wide
608 diversification or that are unique to Holozoa. Much of the subsequent results depend on case-
609 control estimates of OR and downstream calculations. However, carrier data were too sparse,
610 and the racial/ethnic categorizations between our source data too different for us to make
611 detailed adjustment for race/ethnicity, let alone other potentially relevant variables. These
612 concerns may impact the precision of our results, especially at effect sizes that are relevant to
613 distinctions between benign and moderate-risk missense substitutions. Finally, repeated use of
614 the MLE model, especially applying it at two distinctly different OR thresholds, may have

615 overstressed its assumption of a 2-component mixture. This concern may also impact the
616 precision of our results.

617 The MLE model coupled with a preliminary definition of ECARs allowed us to tease out
618 the observation that substitutions falling at the 14 *BRCA1*/*BARD1* ECAR residues contribute
619 disproportionate evidence of pathogenicity for missense substitutions in *BRCA1*, leading to a
620 hypothesis that will need to be tested with evidence from other genes. With substitutions falling
621 at ECAR positions set aside, Align-GVGD and AlphaMissense emerged as the two strongest
622 computational tools for prediction of *BRCA1* missense substitution pathogenicity. These two
623 tools also performed well for prediction of *BRCA1* missense substitution benignity. Since
624 AlphaMissense is a proteome-wide tool whereas Align-GVGD is not, this observation leads to a
625 recommendation in favor of the use of AlphaMissense – but also a recommendation that one of
626 the Meta-predictor computational tools be upgraded to include explicit recognition of ECARs, an
627 Align-GVGD like scoring component, and AlphaMissense itself.

628 The MLE model also enabled joint evaluation of functional assay and computational tool
629 outputs with separate OR thresholds for evidence in favor of pathogenicity and evidence in favor
630 of benignity. In both instances, discordant results from functional assay and computational tool
631 created analytical subsets of variants that may harbor variants that are actually RPPV, which
632 otherwise may have been incorrectly classified as high-risk Likely Pathogenic or Likely Benign.
633 On the benign end of the spectrum, this analysis was only possible because the BRCA1/2
634 VCEP has a well-defined moderate-risk threshold. As the findings from this study are applied to
635 other health outcomes, a critical effort will be defining similar moderate-risk thresholds for the
636 relevant set of genes and diseases. Doing so will enable the analytical detection of the
637 biological reality that disease risk is not an all or nothing outcome but rather falls on a continuum
638 from benign to pathogenic.

639

640

641

642

ACKNOWLEDGEMENTS:

We would like to thank Marcy Richardson, Tina Pesaran, Colin Young, and Steven Harrison from Ambry Genetics for thoughtful discussion and database support. We would also like to thank AC Tan, Jay Gertz, and Pinar Bayrak-Toydemir for thoughtful reading and comments on the manuscript. This work was funded in part by NIH NCI R01CA264971, to SVT. STP is a recipient of the NIH NCI T32CA265782.

DATA AND CODE AVAILABILITY:

Simple observational data tables are available from S.V.T. R code for the main analysis is available from A.T.

DECLARATION OF INTERESTS:

The authors declare no competing interests.

REFERENCES:

(1) Landrum, M. J.; Lee, J. M.; Riley, G. R.; Jang, W.; Rubinstein, W. S.; Church, D. M.; Maglott, D. R. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* **2014**, *42* (Database issue), D980-985. DOI: 10.1093/nar/gkt1113 From NLM Medline.

(2) Richards, S.; Aziz, N.; Bale, S.; Bick, D.; Das, S.; Gastier-Foster, J.; Grody, W. W.; Hegde, M.; Lyon, E.; Spector, E.; 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* **2015**, *17* (5), 405-424. DOI: 10.1038/gim.2015.30 From NLM Medline.

(3) Tavtigian, S. V.; Greenblatt, M. S.; Harrison, S. M.; Nussbaum, R. L.; Prabhu, S. A.; Boucher, K. M.; Biesecker, L. G.; ClinGen Sequence Variant Interpretation Working, G. Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genet Med* **2018**, *20* (9), 1054-1060. DOI: 10.1038/gim.2017.210 From NLM Medline.

(4) 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* **2020**, *41* (10), 1734-1737. DOI: 10.1002/humu.24088 From NLM Medline.

(5) Pejaver, V.; Byrne, A. B.; Feng, B. J.; Pagel, K. A.; Mooney, S. D.; Karchin, R.; O'Donnell-Luria, A.; Harrison, S. M.; Tavtigian, S. V.; Greenblatt, M. S.; et al. Calibration of computational tools for missense variant pathogenicity classification and ClinGen recommendations for PP3/BP4 criteria. *Am J Hum Genet* **2022**, *109* (12), 2163-2177. DOI: 10.1016/j.ajhg.2022.10.013 From NLM Medline.

(6) Brnich, S. E.; Abou Tayoun, A. N.; Couch, F. J.; Cutting, G. R.; Greenblatt, M. S.; Heinen, C. D.; Kanavy, D. M.; Luo, X.; McNulty, S. M.; Starita, L. M.; et al. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med* **2019**, *12* (1), 3. DOI: 10.1186/s13073-019-0690-2 From NLM Medline.

(7) Clark, K. A.; Paquette, A.; Tao, K.; Bell, R.; Boyle, J. L.; Rosenthal, J.; Snow, A. K.; Stark, A. W.; Thompson, B. A.; Unger, J.; et al. Comprehensive evaluation and efficient classification of BRCA1 RING domain missense substitutions. *Am J Hum Genet* **2022**, *109* (6), 1153-1174. DOI: 10.1016/j.ajhg.2022.05.004 From NLM Medline.

(8) Spurdle, A. B.; Greville-Heygate, S.; Antoniou, A. C.; Brown, M.; Burke, L.; de la Hoya, M.; Domchek, S.; Dork, T.; Firth, H. V.; Monteiro, A. N.; et al. Towards controlled terminology for reporting germline cancer susceptibility variants: an ENIGMA report. *J Med Genet* **2019**, *56* (6), 347-357. DOI: 10.1136/jmedgenet-2018-105872 From NLM Medline.

(9) Yang, X.; Wu, Y.; Ficorella, L.; Wilcox, N.; Dennis, J.; Tyrer, J.; Carver, T.; Pashayan, N.; Tischkowitz, M.; Pharoah, P. D. P.; et al. Validation of the BOADICEA model for epithelial tubo-ovarian cancer risk prediction in UK Biobank. *Br J Cancer* **2024**, *131* (9), 1473-1479. DOI: 10.1038/s41416-024-02851-z From NLM Medline.

(10) Siegel, R. L.; Giaquinto, A. N.; Jemal, A. Cancer statistics, 2024. *CA Cancer J Clin* **2024**, *74* (1), 12-49. DOI: 10.3322/caac.21820 From NLM Medline.

(11) Institute, N. C. Cancer Stat Facts.

(12) Wallace, I. M.; O'Sullivan, O.; Higgins, D. G.; Notredame, C. M-Coffee: combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Res* **2006**, *34* (6), 1692-1699. DOI: 10.1093/nar/gkl091 From NLM Medline.
(13) Armougom, F.; Moretti, S.; Poirot, O.; Audic, S.; Dumas, P.; Schaeli, B.; Keduas, V.; Notredame, C. Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res* **2006**, *34* (Web Server issue), W604-608. DOI: 10.1093/nar/gkl092 From NLM Medline.
(14) Potter, S. C.; Luciani, A.; Eddy, S. R.; Park, Y.; Lopez, R.; Finn, R. D. HMMER web server: 2018 update. *Nucleic Acids Res* **2018**, *46* (W1), W200-W204. DOI: 10.1093/nar/gky448 From NLM Medline.
(15) Tavtigian, S. V.; Byrnes, G. B.; Goldgar, D. E.; Thomas, A. Classification of rare missense substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications. *Hum Mutat* **2008**, *29* (11), 1342-1354. DOI: 10.1002/humu.20896 From NLM Medline.
(16) Young, E. L.; Feng, B. J.; Stark, A. W.; Damiola, F.; Durand, G.; Forey, N.; Francy, T. C.; Gammon, A.; Kohlmann, W. K.; Kaphingst, K. A.; et al. Multigene testing of moderate-risk genes: be mindful of the missense. *J Med Genet* **2016**, *53* (6), 366-376. DOI: 10.1136/jmedgenet-2015-103398 From NLM Medline.
(17) Findlay, G. M.; Daza, R. M.; Martin, B.; Zhang, M. D.; Leith, A. P.; Gasperini, M.; Janizek, J. D.; Huang, X.; Starita, L. M.; Shendure, J. Accurate classification of BRCA1 variants with saturation genome editing. *Nature* **2018**, *562* (7726), 217-222. DOI: 10.1038/s41586-018-0461-z From NLM Medline.
(18) Nagy, G.; Diabate, M.; Banerjee, T.; Adamovich, A. I.; Smith, N.; Jeon, H.; Dhar, S.; Liu, W.; Burgess, K.; Chung, D.; et al. Multiplexed assay of variant effect reveals residues of functional importance in the BRCA1 coiled-coil and serine cluster domains. *PLoS One* **2023**, *18* (11), e0293422. DOI: 10.1371/journal.pone.0293422 From NLM Medline.
(19) Parsons, M. T.; de la Hoya, M.; Richardson, M. E.; Tudini, E.; Anderson, M.; Berkofsky-Fessler, W.; Caputo, S. M.; Chan, R. C.; Cline, M. S.; Feng, B. J.; et al. Evidence-based recommendations for gene-specific ACMG/AMP variant classification from the ClinGen ENIGMA BRCA1 and BRCA2 Variant Curation Expert Panel. *Am J Hum Genet* **2024**, *111* (9), 2044-2058. DOI: 10.1016/j.ajhg.2024.07.013 From NLM Medline.
(20) Esposito, D.; Weile, J.; Shendure, J.; Starita, L. M.; Papenfuss, A. T.; Roth, F. P.; Fowler, D. M.; Rubin, A. F. MaveDB: an open-source platform to distribute and interpret data from multiplexed assays of variant effect. *Genome Biol* **2019**, *20* (1), 223. DOI: 10.1186/s13059-019-1845-6 From NLM Medline.
(21) Feng, B. J. PERCH: A Unified Framework for Disease Gene Prioritization. *Hum Mutat* **2017**, *38* (3), 243-251. DOI: 10.1002/humu.23158 From NLM Medline.
(22) Ioannidis, N. M.; Rothstein, J. H.; Pejaver, V.; Middha, S.; McDonnell, S. K.; Baheti, S.; Musolf, A.; Li, Q.; Holzinger, E.; Karyadi, D.; et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet* **2016**, *99* (4), 877-885. DOI: 10.1016/j.ajhg.2016.08.016 From NLM Medline.
(23) Carter, H.; Douville, C.; Stenson, P. D.; Cooper, D. N.; Karchin, R. Identifying Mendelian disease genes with the variant effect scoring tool. *BMC Genomics* **2013**, *14 Suppl 3* (Suppl 3), S3. DOI: 10.1186/1471-2164-14-S3-S3 From NLM Medline.

(24) Martin, F. J.; Amode, M. R.; Aneja, A.; Austine-Orimoloye, O.; Azov, A. G.; Barnes, I.; Becker, A.; Bennett, R.; Berry, A.; Bhai, J.; et al. Ensembl 2023. *Nucleic Acids Res* **2023**, *51* (D1), D933-D941. DOI: 10.1093/nar/gkac958 From NLM Medline.
(25) Liu, X.; Li, C.; Mou, C.; Dong, Y.; Tu, Y. dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Genome Med* **2020**, *12* (1), 103. DOI: 10.1186/s13073-020-00803-9 From NLM Medline.
(26) Pejaver, V.; Urresti, J.; Lugo-Martinez, J.; Pagel, K. A.; Lin, G. N.; Nam, H. J.; Mort, M.; Cooper, D. N.; Sebat, J.; Iakoucheva, L. M.; et al. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nat Commun* **2020**, *11* (1), 5918. DOI: 10.1038/s41467-020-19669-x From NLM Medline.
(27) Cheng, J.; Novati, G.; Pan, J.; Bycroft, C.; Zemgulyte, A.; Applebaum, T.; Pritzel, A.; Wong, L. H.; Zielinski, M.; Sargeant, T.; et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* **2023**, *381* (6664), eadg7492. DOI: 10.1126/science.adg7492 From NLM Medline.
(28) Tavtigian, S. V.; Deffenbaugh, A. M.; Yin, L.; Judkins, T.; Scholl, T.; Samollow, P. B.; de Silva, D.; Zharkikh, A.; Thomas, A. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *J Med Genet* **2006**, *43* (4), 295-305. DOI: 10.1136/jmg.2005.033878 From NLM Medline.
(29) Mathe, E.; Olivier, M.; Kato, S.; Ishioka, C.; Hainaut, P.; Tavtigian, S. V. Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods. *Nucleic Acids Res* **2006**, *34* (5), 1317-1325. DOI: 10.1093/nar/gkj518 From NLM Medline.
(30) Bergquist, T.; Stenton, S. L.; Nadeau, E. A. W.; Byrne, A. B.; Greenblatt, M. S.; Harrison, S. M.; Tavtigian, S. V.; O'Donnell-Luria, A.; Biesecker, L. G.; Radivojac, P.; et al. Calibration of additional computational tools expands ClinGen recommendation options for variant classification with PP3/BP4 criteria. *bioRxiv* **2024**. DOI: 10.1101/2024.09.17.611902 From NLM PubMed-not-MEDLINE.
(31) Chen, S.; Francioli, L. C.; Goodrich, J. K.; Collins, R. L.; Kanai, M.; Wang, Q.; Alföldi, J.; Watts, N. A.; Vittal, C.; Gauthier, L. D.; et al. A genomic mutational constraint map using variation in 76,156 human genomes. *Nature* **2024**, *625* (7993), 92-100. DOI: 10.1038/s41586-023-06045-0.
(32) Karczewski, K. J.; Francioli, L. C.; Tiao, G.; Cummings, B. B.; Alföldi, J.; Wang, Q.; Collins, R. L.; Laricchia, K. M.; Ganna, A.; Birnbaum, D. P.; et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **2020**, *581* (7809), 434-443. DOI: 10.1038/s41586-020-2308-7.
(33) Li, H.; LaDuca, H.; Pesaran, T.; Chao, E. C.; Dolinsky, J. S.; Parsons, M.; Spurdle, A. B.; Polley, E. C.; Shimelis, H.; Hart, S. N.; et al. Classification of variants of uncertain significance in BRCA1 and BRCA2 using personal and family history of cancer from individuals in a large hereditary cancer multigene panel testing cohort. *Genet Med* **2020**, *22* (4), 701-708. DOI: 10.1038/s41436-019-0729-1 From NLM Medline.
(34) Easton, D. F.; Deffenbaugh, A. M.; Pruss, D.; Frye, C.; Wenstrup, R. J.; Allen-Brady, K.; Tavtigian, S. V.; Monteiro, A. N.; Iversen, E. S.; Couch, F. J.; et al. A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and

790 BRCA2 breast cancer-predisposition genes. *Am J Hum Genet* **2007**, *81* (5), 873-883. DOI:
791 10.1086/521032 From NLM Medline.
792 (35) Trapp, O.; Seeliger, K.; Puchta, H. Homologs of breast cancer genes in plants. *Front*
793 *Plant Sci* **2011**, *2*, 19. DOI: 10.3389/fpls.2011.00019 From NLM PubMed-not-MEDLINE.
794 (36) Keeling, P. J.; Burki, F. Progress towards the Tree of Eukaryotes. *Curr Biol* **2019**, *29* (16),
795 R808-R817. DOI: 10.1016/j.cub.2019.07.031 From NLM Medline.
796 (37) Al Jewari, C.; Baldauf, S. L. An excavate root for the eukaryote tree of life. *Sci Adv* **2023**,
797 *9* (17), eade4973. DOI: 10.1126/sciadv.ade4973 From NLM Medline.
798 (38) Garrett, A.; Durkie, M.; Callaway, A.; Burghel, G. J.; Robinson, R.; Drummond, J.; Torr,
799 B.; Cubuk, C.; Berry, I. R.; Wallace, A. J.; et al. Combining evidence for and against
800 pathogenicity for variants in cancer susceptibility genes: CanVIG-UK consensus
801 recommendations. *J Med Genet* **2021**, *58* (5), 297-304. DOI: 10.1136/jmedgenet-2020-
802 107248 From NLM Medline.
803 (39) Pal, T.; Mundt, E.; Richardson, M. E.; Chao, E.; Pesaran, T.; Slavin, T. P.; Couch, F. J.;
804 Monteiro, A. N. A. Reduced penetrance BRCA1 and BRCA2 pathogenic variants in clinical
805 germline genetic testing. *NPJ Precis Oncol* **2024**, *8* (1), 247. DOI: 10.1038/s41698-024-
806 00741-4 From NLM PubMed-not-MEDLINE.
807
808
809
810
811
812
813
814

815

816 **FIGURE LEGENDS**

817 **Figure 1**. Description of Study Workflow

818 Functional assay results were combined with other data types (Exceptionally Conserved
819 Ancestral Residue (ECAR) or computational tool prediction) to evaluate the additivity
820 assumptions in the ACMG guidelines. Table 1: Tabulation of the number of functionally abnormal
821 variants (red line) that fall at ECAR positions. Table 2: A case control analysis and conversion to
822 ACMG points is then performed on these variants and results. Table 3: Functionally abnormal
823 variants that do not fall at ECAR positions are combined with predictions from six computational
824 tools, followed by a case control analysis and conversion to ACMG points. Table 4: Functionally
825 normal variants (blue line) from the assay are combined with computational tool predictions from
826 the same six tools, followed by a case control analysis and conversion to ACMG points.

827

828 **Figure 2**. Graphical Summary of Results

829 The matrix represents the main combinations of evidence evaluated in this study. The number in
830 the row or column headers represent the Bayesian point value assigned to each evidence type in
831 the ACMG guidelines. Each box has three lines: Expectation, Observed, and Percent of calls.
832 "Expectation" is the Bayesian points expected according to the ACMG guidelines. The
833 combination of ECAR, MAVE loss of function, and computational tool pathogenic was capped at
834 ≥+6 points. Otherwise expected points per category are additive as described in the ACMG
835 guidelines. The Observed line represents the Bayesian point value that was observed using the
836 MLE model in this study. The Percent all calls line represents the range, as a percent, of all
837 possible substitutions that were observed for a particular combination of evidence for the
838 computational tools evaluated. Blue filled boxes indicate that the combination of evidence met or
839 exceeded the ACMG guideline expectations in this study. Peach filled boxes indicate that the
840 combination of evidence departed from the expected additivity of the ACMG guideline in this

study. The bold black outline highlights that the assumption that MAVE LOF and computational predictions of pathogenicity are additive is not met in this study; moreover, this combination had a substantial number of observations. White filled boxes represent combinations of evidence with little to no data. Dark grey filled boxes represent combinations of evidence with results that may be beyond the dynamic range of the MLE model.

Figure 1

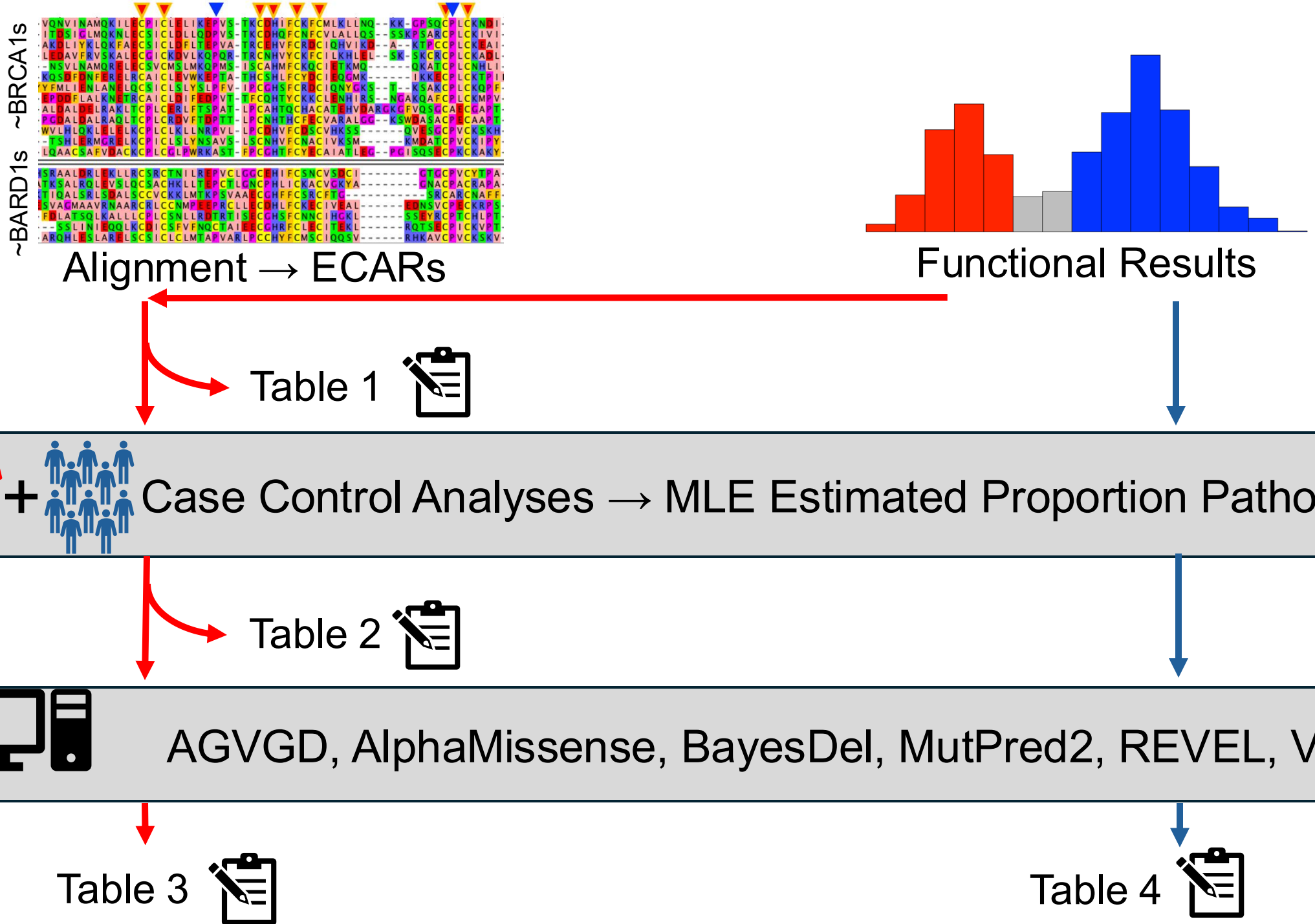
~BRCA1s
~BARD1s
Alignment → ECARs
Functional Results
Table 1
Case Control Analyses → MLE Estimated Proportion Pathogenic
Table 2
AGVGD, AlphaMissense, BayesDel, MutPred2, REVEL, VEST4
Table 3
Table 4

Figure 2

		Substitution at ECAR (+2)		Substitution at non-ECAR (+0)	
		MAVE LOF (+4)	MAVE Functional (-4)	MAVE LOF (+4)	MAVE Functional (-4)
Computational Prediction	Pathogenic (+1)	Expectation: ≥+6 ACMG Points Observed: ≥+6 Percent of calls: (2.67% - 4.25%)	Expectation: -1 ACMG Points Observed: NA Percent of calls: (0.00%)	Expectation: +5 ACMG Points Observed: +4 to +5 Percent of calls: (6.42% - 15.24%)	Expectation: -3 ACMG Points Observed: 0 to +2 Percent of calls: (3.32% - 22.65%)
	Benign (-1)	Expectation: +5 ACMG Points Observed: NA Percent of calls: (0.00% - 0.11%)	Expectation: -3 ACMG Points Observed: NA Percent of calls: (0.00%)	Expectation: +3 ACMG Points Observed: +1 to +5 Percent of calls: (0.00% - 4.90%)	Expectation: -5 ACMG Points Observed: ≤-5 Percent of calls: (0.16% - 54.98%)

= Meets or exceeds expectations

= Departs from expected additivity

= Scant Observations

= Limited sensitivity between individual data points

Table 1. Predictions from SGE Functional Assay and Computational Tools at ECAR Positions.

			Number of +4 Calls made by:					Other Models:	
Residue	Total Number of Possible snMS	LOF in SGE	AlphaMissense	BayesDel	MutPred2	REVEL	VEST4	BayesDel-VCEP[a]	A-GVGD
Cys24	6	5	6	5	0	5	1	6	6
Cys27	6	6	4	5	0	5	0	6	6
Cys39	6	6	5	6	0	4	2	6	6
His41	7	7	2	6	0	3	0	7	5
Cys44	6	6	5	6	0	4	0	6	6
Cys47	6	6	6	6	0	2	3	6	6
Cys61	6	5	5	6	0	4	2	6	6
Cys64	6	6	4	6	0	5	4	6	6
Thr1685	5	5	0	0	0	0	0	3	1
His1686	7	7	0	0	0	0	0	6	5
Thr1700	6	5	0	0	0	0	0	6	3
Trp1718	5	4	1	5	0	2	0	5	4
Arg1753	5	5	0	0	0	0	0	5	4
Trp1837	5	5	2	5	0	0	0	5	4
Totals at ECAR positions	82	78[b]	41	56	0	34	12	79	68
Total ECAR snMS with LOF in SGE	78	78	39	53	0	32	11	75	65
Totals with LOF via SGE assay	404	404	52	67	0	33	11	301	183
Totals in RING/tBRCT domains	1,840	404	60	73	1	35	12	530	267
Totals across BRCA1	11,009	NA	60	122	1	45	12	846	314

ECAR = Exceptionally Conserved Ancestral Residue
snMS = Single Nucleotide Missense Substitution
LOF = Loss of Function
SGE = Saturation Genome Editing
VCEP = Variant Curation Expert Panel

a. VCEP thresholds divide ‘BayesDel no AF’ scores into supporting benign, supporting pathogenic, and indeterminate scores. Scores $\geq$ 0.28 are supporting pathogenic.
b. 4 variants at ECAR positions were not LOF. C24F, C61R, and W1718S had indeterminate results. T1700S was not interrogated in the assay.

Table 2. Proportions Pathogenic, and ACMG Point Estimates of Functional Assay Analytical Subsets and Non-functional Substitutions Falling at ECAR Positions. OR Threshold equal to that of nonsense variants in breast and ovarian cancer groups[a]

	Possible SNS[b]	Observed OR: Breast Cancer	Observed OR: Ovarian Cancer	M.L.E Proportion Pathogenic[c]	Estimated ACMG Point[d]	Expected ACMG Point[e]	Meets ACMG Expectations
Nonsense Substitutions	800	16.2	46.4	1 (Reference)	+8	+8	Yes
M2H Assay Categories							
Functional	452	1.42[h]	2.45[h]	0.04	-1	-4	No
Indeterminate	17	NA	NA	NA	NA	NA	
Non-Functional – All	100	19.9[h]	40.0[h]	0.99	≥+4	+4	Yes
Non-functional – ECAR	50	22.0	40.8	1.00	≥+6	+6	Yes
Non-functional – Other	50	7.82	32.9	0.70	+4	+4	Yes
SGE Assay Categories							
Functional	1,185	1.11[h]	1.34[h]	0.00	≤-4	-4	Yes
Indeterminate	148	2.17	5.29	0.12	0	0	Yes
Non-Functional – All	413	15.7	35.1[h]	0.92	≥+4	+4	Yes
Non-functional – ECAR	89	23.6	42.3	1.00	≥+6	+6	Yes
Non-functional – Other	324	12.2	31.7	0.76	+4	+4	Yes
HDR Assay Categories							
Functional	152	1.87	2.04	0.08	0	-4	No
Indeterminate	15	NA	NA	NA	NA	NA	
Non-Functional – All	16	1.58	NA	0.08	0	+4	No
A Priori Damaging[f]	22[g]	0.46	NA	0.00	-4	+2	No

ACMG = American College of Medical Genetics and Genomics
ECAR = Exceptionally Conserved Ancestral Residue
SNS = Single Nucleotide Substitution. Row 1 is the total of SNS that result in stop codons. All other rows represent single nucleotide missense substitutions.
Obs. = Observations
OR = Odds Ratio
C.I. = Confidence Interval
M.L.E = Maximum Likelihood Estimate
NA = Not Applicable

M2H = Mammilian-2-Hybrid
SGE = Saturation Genome Editing
ECAR = Exceptionally Conserved Ancestral Residue
HDR = Homology Directed Repair
a. Nonsense OR for breast cancer = 16.2, Nonsense OR for ovarian cancer = 46.4.
b. Restricted to those snMS interrogated in the assay indicated.
c. The OR threshold used for this calculation is that of the nonsense variants observed in our case/control dataset.
d. The maximum allowed points are capped as specified in ACMG variant classification guidelines per evidence category/combination of evidence categories.
e. The expected points from combining evidence categories according to ACMG guidelines.
f. This subset interrogates the substitutions to Proline and other non-conservative amino acid substitutions that fell at the a,d positions of the alpha helix in the coiled-coil domain (M1400, Q1401, H1402, L1404, L1407, Q1408, Q1409, M1411, A1412, L1414, A1416, L1418), regardless of whether they were interrogated in the assay. Only Q1401P was included in the assay and had no clinical or gnomAD observations in our dataset.
g. This is the total number of possible substitutions listed in footnote e. irrespective of inclusion in the HDR assay of the coiled coil domain.
h. Indicates that these ORs are statistically significant from one another with a p-value less than 0.0001.

Table 3. Proportions Pathogenic, and ACMG Point Estimates of Non-functional, Non-ECAR Assay Results Combined with Computational Predictions. OR Threshold equal to that of nonsense variants in breast and ovarian cancer groups[a].

Computational Tool Prediction	Possible snMS[b]	Observed OR: Breast Cancer	Observed OR: Ovarian Cancer	M.L.E Proportion Pathogenic[c]	Change in Proportion Pathogenic[d]	Estimated ACMG Point[e]	Expected ACMG Point[f]	Meets or Exceeds ACMG Expectations
A-GVGD								
Pathogenic	120	13.9	40.2	0.84	+0.08	+5	+5	Yes
Indeterminate	120	12.5	25.4	0.73	-0.03	+4	+4	Yes
Benign	91	6.06	22.1	0.56	-0.20	+3	+3	Yes
AlphaMissense								
Pathogenic	232	15.4	34.4	0.85	+0.09	+5	+5	Yes
Indeterminate	80	1.86	9.87	0.17	-0.59	0	+4	Yes
Benign	13	NA	NA	NA	NA	NA	+3	NA
BayesDel								
Pathogenic	284	12	30.1	0.75	-0.01	+4	+5	No
Indeterminate	46	15.1	52.7	0.96	+0.20	≥+5	+4	No
Benign	0	NA	NA	NA	NA	NA	+3	NA
BayesDel - VCEP								
Pathogenic	231	12.5	32.6	0.77	+0.01	+4	+5	No
Indeterminate	45	5.51	6.83	0.42	-0.34	+2	+4	Yes
Benign	55	16.2	52.7	0.98	+0.22	≥+5	+3	No
MutPred2								
Pathogenic	142	13.3	31.7	0.79	+0.03	+4	+5	No
Indeterminate	159	11.3	32	0.75	-0.01	+4	+4	Yes
Benign	29	3.7	11.54	0.29	-0.47	+1	+3	Yes
REVEL								
Pathogenic	285	11.8	29.8	0.74	-0.02	+4	+5	No
Indeterminate	46	19.6	60.1	1.00	+0.24	≥+5	+4	No
Benign	0	NA	NA	NA	NA	NA	+3	NA
VEST4								
Pathogenic	213	13.5	33.8	0.81	+0.05	+4	+5	No
Indeterminate	108	6.46	20.1	0.53	-0.23	+3	+4	Yes
Benign	10	NA	NA	NA	NA	NA	+3	NA

ACMG = American College of Medical Genetics and Genomics

ECAR = Exceptionally Conserved Ancestral Residue
snMS = Single Nucleotide Missense Substitution
Obs. = Observations
OR = Odds Ratio
C.I. = Confidence Interval
M.L.E = Maximum Likelihood Estimate
NA = Not Applicable
VCEP = Variant Curation Expert Panel
a. Nonsense OR for breast cancer = 16.2, Nonsense OR for ovarian cancer = 46.4.
b. Restricted to those snMS interrogated in the assay indicated.
c. The OR threshold used for this calculation is that of the nonsense variants observed in our case/control dataset.
d. The starting proportion pathogenic is that of the Non-ECAR, Non-functional group of variants from the SGE assay, 0.76.
e.The maximum allowed points are capped as specified in ACMG variant classification guidelines per evidence category/combination of evidence categories.
f. The expected points from combining evidence categories according to ACMG guidelines.

Table 4. Proportions Pathogenic, and ACMG Point Estimates of Functionally Normal Assay Results Combined with Computational Predictions. OR threshold equal to that of moderate risk variants[a].

Computational Tool Prediction	Possible snMS[b]	Observed OR: Breast Cancer	Observed OR: Ovarian Cancer	M.L.E Proportion Pathogenic	Change in Proportion Pathogenic	Estimated ACMG Point[c]	Expected ACMG Point[d]	Meets or Exceeds ACMG Expectations
A-GVGD								
Pathogenic	61	2.71[e]	3.73	0.83	+0.83	≥+2	-3	Yes
Indeterminate	212	1.76	1.79	0.42	+0.42	+2	-4	Yes
Benign	912	0.88[e]	1.09	0.00	0.00	≤-5	-5	Yes
AlphaMissense								
Pathogenic	122	2.12[e]	1.69	0.38	+0.38	+2	-3	Yes
Indeterminate	586	1.31	1.42	0.14	+0.14	0	-4	Yes
Benign	477	0.71[e]	1.13	0.00	0.00	≤-5	-5	Yes
BayesDel								
Pathogenic	379	1.69	1.89	0.35	+0.35	+2	-3	Yes
Indeterminate	769	0.79	0.98	0.00	0.00	≤-4	-4	Yes
Benign	37	1.67	1.97	0.28	+0.28	+1	-5	No
BayesDel - VCEP								
Pathogenic	152	1.53	1.90	0.32	+0.32	+1	-3	Yes
Indeterminate	181	1.59	1.79	0.29	+0.29	+1	-4	Yes
Benign	852	0.89	1.07	0.00	0.00	≤-5	-5	Yes
MutPred2								
Pathogenic	86	1.49	NA	0.04	+0.04	-1	-3	Yes
Indeterminate	399	1.35	1.43	0.21	+0.21	+1	-4	Yes
Benign	697	0.95	1.28	0.00	0.00	≤-5	-5	Yes
REVEL								
Pathogenic	413	1.38	1.45	0.16	+0.16	0	-3	Yes
Indeterminate	769	0.95	1.23	0.00	+0.00	≤-4	-4	Yes
Benign	3	NA	NA	NA	NA	NA	-5	NA
VEST4								
Pathogenic	116	3.43[e]	3.94[e]	1.00	+1.00	≥+2	-3	Yes
Indeterminate	763	1.18	1.40	0.13	+0.13	0	-4	Yes
Benign	306	0.57[e]	0.66[e]	0.00	0.00	≤-5	-5	Yes

ACMG = American College of Medical Genetics and Genomics
snMS = Single Nucleotide Missense Substitution

Obs. = Observations
OR = Odds Ratio
C.I. = Confidence Interval
M.L.E = Maximum Likelihood Estimate
NA = Not Applicable
VCEP = Variant Curation Expert Panel

a. The OR threshold used for this calculation is that of moderate risk variants (OR = 2 for breast cancer and OR = 2.9 for ovarian cancer) adjusted for observational inflation in our data to OR = 2.31 and OR = 3.83 respectively.
b. Restricted to those snMS interrogated in the assay indicated.
c. Positive points should be interpreted in the context of moderate risk variants. Points in favor of pathogenicity are capped at +2 to reflect this. Points in favor of benignity are capped at the lower end specified in ACMG variant classification guidelines per evidence category/combination of evidence categories.
d. The expected points from combining evidence categories according to ACMG guidelines.
e. These ORs are statistically significant from one another with a p-value less than 0.05 according to a Wald test.

Table 5. Summary of Combining SGE Functional Assay Results with Computational Tool Predictions Compared to ACMG Expectations.

Computational Tool	Addition of Computational Tool Prediction was Statistically Significant in Logistic Regression Model	Met Expectations When Combined With non-ECAR LOF Results	Met Expectations When Combined With non-ECAR Functionally Normal Results	Met Overall Expectations
AGVGD	Yes	Yes	Yes	Yes
AlphaMissense	Yes	Yes	Yes	Yes
BayesDel	No	No	No	No
BayesDel – VCEP	Yes	No	Yes	No
MutPred2	No	No	Yes	No
REVEL	No	No	No	No
VEST4	Yes	No	Yes	No

Table 6. Examples of Differences Between Naïve ACMG Addition and Jointly Calibrated Addition of Conserved Domain Data, Functional Assay Results, and Computational Tool Predictions in High-throughput Variant Classification.

BRCA1 Variant	c.5099C>T p.Thr1700Ile		c.44T>C p.Ile15Thr		c.5519A>C p.Asp1840Ala		c.5210G>T p.Arg1737Ile		c.5468C>T p.Ala1823Val	
ClinVar Overall Classification	Conflicting Ambry: LP LabCorp: Uncertain		Uncertain Color: Uncertain LabCorp: Uncertain		Uncertain Invitae: Uncertain		Uncertain Invitae: Uncertain		Not observed in ClinVar One observation in gnomAD	
Naïve ECAR Position (PM1)	Yes	+2	No	+0	No	+0	No	+0	No	+0
Naïve SGE Functional Assay Result (PS3/BS3)	LOF	+4	LOF	+4	WT	-4	WT	-4	WT	-4
Naïve Computational Tool Prediction (PP3/BP4)[a]	PP3	+1	PP3	+1	PP3	+1	Indeterminate	+0	BP4	-1
Sum of Addition of Naïve ACMG Points: ECAR + Functional Assay + Computational Tool[b]	+7		+5		-3		-4		-5	
Jointly Calibrated ACMG Points: ECAR + Functional Assay + Computational Tool[b]	+7		+5		≥+2 if using A-GVGD		+2 if using A-GVGD		≤-5	
					+2 if using AlphaMissense		+0 if using AlphaMissense			
Additional Evidence Available for High-throughput Classification										
Novel Missense Change at an Amino Acid Residue Where a Different Missense Change is Pathogenic (PM5)	+0[c]		+0[c]		+0[c]		+0[c]		+0[d]	
Case/Control Observations (PS4)[e]	+0		+0		+0[f]		+0		+0[g]	
Classification Based on Jointly Calibrated ACMG Points	LP		VUS (Hot)[h]		VUS (Cool/Cold)[h]		VUS (Cold)[h]		LB	

a. Points from computational tool predictions are limited to +1/-1 to avoid double counting of evidence contained in PM1.

b. Variants evaluated here had concordant computational tool results from A-GVGD and AlphaMissense.

c. No other variant at these locations has an overall classification of Likely Pathogenic or Pathogenic in ClinVar
d. Other variants at this amino acid residue but different nucleotide positions are classified as Likely Pathogenic or Pathogenic in ClinVar due to their impact on splicing. SpliceAI does not predict the c.5468C>T change to impact splicing[1].
e. ACMG points were calculated using the PS4 LR calculator from Rowlands et al. using the default OR threshold of 5. Although all of these variants are reported in either ClinVar or gnomAD, there were 0 case or control observations of these variants in our dataset except where notated (see f,g).[2]
f. There was one clinical observation of this missense substitution in the Myriad dataset used in Clark et al. but in the dataset used in the present analysis, which was strictly reduced to index cases with breast and/or ovarian cancer, there were no clinical observations[3].
g. There was 1 observation of this missense substitution in gnomAD.
h. Temperature ranking from Cancer Variant Interpretation Group UK Consensus Recommendations[4]