When Two plus Four Does Not Equal Six: Combining Computational and Functional Evidence Towards Classification of 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 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
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 individual who are

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

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

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 data set were

134 tabulated and a frequentist odds ratio was calculated for this group of variants (21.6, 95% CI:

135 16.9-27.7). To evaluate the prediction of benign variants we used an alternative OR ceiling

136 chosen to match the threshold of moderate risk variants (OR = 2.0), which we calculated to be

137 inflated to 4.4 in our family history enriched case-control dataset [8].

138 The general population frequency of breast cancer that we used in our analysis was

139 0.125, as stated in the 2024 update of Cancer Statistics [9].

140

141 Identification of Exceptionally Conserved Ancestral Residues in BRCA1 / BARD1

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

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

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

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

146 coffee suite of sequence alignment tools [10, 11]. The two resulting alignments were used with

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

148 maintained at the HMMER website [12]. Hits were defined as gene models with high-scoring

149 matches to both the RING and tandem BRCT alignments.

150 At the first round, hits from Placozoa and single celled Holozoans were included, the

151 alignments rebuilt, and then trimmed back to include clear BRCA1 and BARD1 orthologs from

152 *Trichoplax* and *Salpingoeca*. The process was iterated to obtain and include homologs from the
153 Fungi and then reiterated to obtain and include homologs from three additional eukaryotic
154 supergroups: Archaeplastida, SAR, and Discoba.
155 To prepare the final alignments and then define exceptionally conserved ancestral
156 residues (ECARs), we stratified the hits into a group of BARD1 and relatively BARD1-like hits
157 and a second group of BRCA1 and relatively BRCA1-like groups; ambiguous hits were placed in
158 the BRCA1-like group. As before, RING and tBRCT alignments were prepared using 3D-
159 Coffee/Expresso. Then the BRCA1 and BARD1 alignments were aligned to each other by
160 introducing gaps across the whole BRCA1 or whole BARD1 alignment, i.e., without modifying
161 the within-group alignments. The tBRCT alignments were then appended to the RING domain
162 alignments. Finally, ECARs were defined as (a) positions that were invariant in our existing
163 metazoan BRCA1 and BARD1 alignments used with Align-GVGD [13, 14], and either (b) also
164 invariant in the new multi-supergroup alignments or (c) had just one conservative substitution
165 (i.e., Grantham Variation <65) in the new multi-supergoup alignment.
166
167 Functional Assay Results and Computational Predictions
168 Functional assay scores were obtained from three separate assays that interrogate
169 missense variants in various domains of BRCA1. These assays were: the 2018 Findlay et al.
170 saturation genome editing (SGE) assay that interrogated missense and splice junction variants
171 in the RING and BRCT domain of BRCA1 [15], the 2022 Clark et al. mammalian-2-hybrid (M2H)
172 assay performed by our laboratory that evaluated missense variants in the RING domain of
173 BRCA1 [7], and the 2023 Nagy et al. homology-directed repair (HDR) assay that evaluated
174 missense variants in the coiled-coil domain of BRCA1 [16]. Scores for these assays were
175 downloaded from MaveDB (SGE and M2H) [17] or from the publication [16]. Thresholds for
176 functional, indeterminate and non-functional variants were used as defined by the relevant
177 publication.

178 Computational tool scores for BayesDel no-AF [18], REVEL [19], and VEST4 [20] were
179 obtained from the database for Non-synonymous Functional Predictions (dbNSFP) via
180 Ensembl's web-based Variant Effect Predictor (VEP) [21, 22]. Computational scores for MutPred2
181 were obtained by downloading a local copy of the software[23]. Scores for AlphaMissense were
182 downloaded from the original publication [24]. Scores for Align-GVGD (A-GVGD) [25, 26] were
183 obtained from agvgd.hci.utah.edu using the Human to Sea Urchin alignment. Thresholds for
184 pathogenic and benign for AlphaMissense, BayesDel, MutPred2, REVEL, and VEST4 were the
185 supporting pathogenic and supporting benign thresholds from recent calibrations of these tools[5,
186 27] Align-GVGD, predictions with a score of C0 were interpreted as predicted benign and
187 predictions with a score of C65 were interpreted as predicted pathogenic with all other scores
188 being indeterminate.
189
190 Estimate of ACMG Points
191 We interpreted the MLE proportion pathogenic as a posterior probability, PP
192 (represented as *q* in the supplemental information and Equation 1), of the variant being
193 pathogenic and we thus converted this estimate to the appropriate ACMG point via an OP using
194 a prior probability (P_1) of 0.102 and Equations 2 and 3 below. Both equations and the prior
195 probability come from Tavtigian et al. [4]. When assigning ACMG points from the estimated
196 proportion pathogenic, the points that could be awarded were limited to the maximum values
197 prescribed in the ACMG guidelines [2, 4].
198 $$PP = (OP \times P_1) \div [(OP - 1) \times P_1 + 1] \quad (2)$$
199 $$OP = 81^{\frac{ACMG\ points}{6}} \quad (3)$$
200
201
202 Case and Control Count Sources and Odds Ratio Calculations

203 Human subjects' data were obtained from Ambry Genetics and Myriad Genetics. Control
204 counts are from the non-cancer cohort of gnomAD v2 exomes and gnomAD v3 genomes.
205 The Ambry dataset initially included data generated by Ambry Genetics from multi-gene
206 sequencing panels for 165,031 individuals "exempted from review by the Western Institutional
207 Review Board" [28]. Filters were applied to remove those individuals without breast or ovarian
208 cancer, individuals who lacked BRCA1 sequence data, individuals of Ashkenazi Jewish
209 ancestry, and individuals with pathogenic or likely pathogenic variants in *TP53*, *BRCA2*, and
210 *PALB2*. This resulted in a final dataset of 91,367 individuals, of which 2,930 had variants in
211 *BRCA1*. Of these 2,930 individuals, 498 had missense variants that fell within the RING (amino
212 acids 1-100), coiled-coil (amino acids 1,280-1,576), or BRCT domain (amino acids 1,650-
213 1,863).
214 The Myriad dataset initially included data generated by full-sequence BRCA1/2 tests
215 performed by Myriad Genetics from 60,607 individuals. For these, a test request form must have
216 been completed by the ordering health care provider, and the form must have been signed by
217 an appropriate individual indicating that "informed consent has been signed and is on file" [13, 28,
218 29]. Filters were applied to remove those individuals without breast or ovarian cancer, individuals
219 who lacked *BRCA1* sequence data, individuals of Ashkenazi Jewish ancestry, and individuals
220 with pathogenic or likely pathogenic variants in *BRCA2*. This resulted in a final dataset of 35,088
221 individuals, of which 4,005 had variants in *BRCA1*. Of these 4,005 individuals, 524 had
222 missense variants that fell within the RING, coiled-coil, or BRCT domain.
223 The gnomAD dataset is a composite of v2 non-cancer exomes and v3 non-cancer
224 genomes that initially included 192,502 individuals. Filters were applied to remove individuals of
225 Finnish or Ashkenazi Jewish ancestry and to remove common missense variants (allele
226 frequency > 0.000271). This resulted in a final dataset of 169,933 individuals, of which 955 had
227 missense variants that fell within the *BRCA1* RING, coiled-coil, or BRCT domain.

228 Combining these datasets resulted in a final dataset of 126,455 breast and ovarian
229 cases from Ambry and Myriad and 169,933 controls from gnomAD v2 and v3. Ethnicity for this
230 dataset was split into four groups: non-Finnish European, Ashkenazi Jewish, Finnish, and all
231 others. As stated previously, Ashkenazi Jewish and Finnish individuals were removed due to
232 strong founder effects in these populations.
233 After filtering, the data for *BRCA1* were formatted into a table that indicated the ethnicity
234 of the individual, whether an individual was a case or control, how the variant was classified by
235 each functional assay and computational tool, and whether the variant fell at an ECAR. Logistic
236 regression was performed using Stata 17.0 (Statacorp) to calculate frequentist odds ratios.
237 Proportions pathogenic were calculated using a custom R script. Following logistic regression,
238 the Wald test was performed to determine statistical significance between strata in each
239 analysis.
240
241 **RESULTS**:
242 Exceptionally Conserved Ancestral Residues in *BRCA1* / *BARD1*
243 *BRCA1*, *BARD1*, and homologs that have been studied in the Archaeplastida are
244 thought to have evolved from a common ancestor gene with a 5' RING domain and 3' tandem
245 BRCT (tBRCT) domain that was present in an ancient eukaryote before the divergence of
246 Amorphea and Archaeplastida [30]. Protein multiple sequence alignment (PMSA) based
247 hmmsearch [12] identified clear *BRCA1* and *BARD1* orthologs in *Trichoplax* and more distantly
248 related single-celled members of Holozoa, plus homologs (some of which may be true
249 orthologs) in the Holomycota, Archaeplastida, SAR, and Discoba. Although the root of the
250 eukaryotic phylogeny remains unclear [31, 32], the resulting PMSA should be sufficient to identify
251 positions in human *BRCA1* and *BARD1* where the amino acid present in the human sequence
252 has been inherited identical by descent from a common ancestor of the Amorphea and is under
253 constraint in 2-3 additional eukaryotic supergroups. We identified 14 such residues: the 8

254 residues that define the C3HC4 RING motif plus 6 additional residues in the tBRCT domain
255 (Table 1, Figure S1), and hereafter refer to them as Exceptionally Conserved Ancestral
256 Residues (ECARs).

257 Of the 82 possible snMSs at ECARs, 78 resulted in a non-functional assay result via the
258 SGE assay. The four exceptions were C24F, C61R, T1700S (which was not interrogated in the
259 SGE assay), and W1718S. C24F, C61R, and W1718S had indeterminate results in the SGE
260 assay, however, C24F and C61R were scored as loss of function in the M2H assay.

261 Using the pathogenic strong thresholds established in Pejaver et al, we also tabulated
262 the number of substitutions at ECARs that were extreme enough to receive +4 ACMG points
263 from computational tools (Table 1). MutPred2 predicted 0 of the possible ECAR substitutions to
264 be extreme enough to receive +4 ACMG points. BayesDel predicted 56 of the 82 possible
265 substitutions to be extreme enough for +4 ACMG points, and these constituted 56 of the 73
266 such RING and tBRCT +4 results called by BayesDel. AlphaMissense, REVEL, and VEST4 had
267 totals that fell between MutPred2 and BayesDel.

268 A separate analysis was performed using the BayesDel threshold prescribed in the
269 ENIGMA BRCA1 Variant Curation Expert Panel (VCEP) guidelines ($\geq$ 0.28) as well as for Align-
270 GVGD using C65 as a pathogenic prediction threshold[33]. Under these parameters, many more
271 of the ECARs identified in the tBRCT domains were predicted to be pathogenic (Table 1). A
272 summary of computational tool performance at ECAR positions and individual scores for each
273 substitution are available in Supplementary Table 1.

274 Functional Assay Performance

275 In our initial analysis, we used an OR threshold set to that of the OR of nonsense
276 substitutions observed in our case control dataset, which was 21.6 (95% CI: 16.9-27.7). The
277 results of our analysis of functional assay performance using this OR threshold are presented in
278 Table 2. As stated in the Methods section, ACMG points awarded are limited to the maximum

279 number of points outlined in the ACMG guidelines and subsequent transformation to a Bayesian
280 framework.

281 In the M2H assay, which interrogates variants in the RING domain of BRCA1, variants
282 that were functional and were observed in our case control dataset yielded an adjusted OR
283 close to 1 (1.21, 95% CI: 0.85-1.72) indicating a group of variants composed almost entirely of
284 benign variants. The proportion pathogenic for this group was estimated to be 0.01 which
285 translates to -3 points in the ACMG points system. There were insufficient observations for
286 variants that were categorized as indeterminate to compute an OR or a proportion pathogenic.
287 The group of all variants that were non-functional and observed in our dataset had an adjusted
288 OR of 22.6 (95% CI: 12.9-39.5) which is similar to that of nonsense substitutions in our dataset,
289 indicating a group consisting of mostly pathogenic variants. This analytical subset had a
290 proportion pathogenic of 1.00, which would be assigned the full +4 points that are available for
291 non-functional assay results in the ACMG guidelines.

292 In addition to evaluating the strength of evidence of a single evidence category, our
293 model can also test the strength of combined evidence categories. To this end, we stratified the
294 non-functional variants from the M2H assay into two groups: variants that fell at the ECARs we
295 identified from the PMSA, and all other residues (ACMG code: PM1). For variants that were
296 non-functional and fell at an ECAR, the computed OR was 24.8 (95% CI: 13.5-45.6), slightly
297 higher than the group of all non-functional variants and indicating a subset of variants with
298 similar risk to that of truncating variants. The computed proportion pathogenic remained at
299 1.00. Combining the +4 points from functional assay and +2 points available from ECARs
300 (PM1), results in +6 points for this group of variants. Thus, using these two evidence criteria
301 alone, these variants reach the likely pathogenic threshold.

302 The computed OR for non-functional variants that fell at other residues was 9.89 (95%
303 CI: 2.23-43.9), indicating that this group of variants confers less risk than those that are non-
304 functional and fall at an ECAR. The proportion pathogenic for the non-functional variants that

305 fell at other residues was 0.74, sufficient for +4 points from functional assay. In the absence of
306 other evidence, this group would remain VUS.
307 The SGE assay, which evaluated variants across the RING and tBRCT domains of
308 *BRCA1*, had similar results to the M2H assay. However, the group of variants that were
309 functional via SGE yielded an OR close to 1 (0.80, 95% CI: 0.64-1.00) and a proportion
310 pathogenic of 0.00, which would qualify for the full -4 points available from a functional assay
311 result. For variants that were non-functional, the SGE assay had an OR of 16.3 (95% CI: 11.6-
312 27.7) and a proportion pathogenic of 0.88 which would confer the full +4 points for these
313 variants. Stratification of non-functional variants at ECARs versus non-functional variants at
314 other residues again demonstrates that non-functional variants at ECARs have enough strength
315 of evidence to reach the likely pathogenic threshold. The ORs for these groups were 26.2 (95%
316 CI: 14.3-48.1) and 12.0 (95% CI: 7.91-18.1), respectively, with a χ^2 p-value of 0.037. Non-
317 functional variants at ECARs had a proportion pathogenic of 1.00 which equates to +6 points.
318 Variants that are non-functional and fall at other residues have a proportion pathogenic of 0.71,
319 just above the threshold required for +4 points; moreover, the χ^2 result provides evidence that
320 the two groups are indeed distinct.
321 Whereas the M2H and SGE assays largely confirmed the assumptions of strength of
322 evidence in the ACMG guidelines, the results from the HDR assay in the coiled-coil domain of
323 *BRCA1* show a different result. The ORs for the functional and non-functional groups were 1.32
324 (95% CI: 0.76-2.29) and 1.48 (95% CI: 0.87-25.2) respectively, indicating that both groups of
325 variants are largely benign. The group of functional variants had a calculated proportion
326 pathogenic of 0.05 which would only justify -1 point. The group of non-functional variants had a
327 proportion pathogenic of 0.06 which is insufficient to justify evidence in either direction (towards
328 pathogenicity or benignity) and thus 0 points. It is noted that there were few observations in our
329 case-control dataset that both fell within this domain and were included in the functional assay.

330 Additionally, many of the variants interrogated by this assay do not appear to be attainable via a
331 single nucleotide substitution to the canonical BRCA1 transcript, which is the type of substitution
332 that is the primary focus of this analysis.

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

340 Combining SGE Functional Results and Computational Predictions

341 To explore the impact of combining functional assay results and computational tool
342 predictions we stratified variants from the SGE assay by computational prediction (pathogenic,
343 indeterminate, and benign) from six computational tools (Align-GVGD, AlphaMissense,
344 BayesDel, MutPred2, REVEL, and VEST4) using thresholds outlined in Pejaver *et al*. or those
345 stated in the Methods section [5]. This analysis was performed only on results from the SGE
346 assay which interrogated both the RING and BRCT domains of *BRCA1* and thus had sufficient
347 observations for stratification.

348 Combining Non-functional Assay Results with Computational Predictions for Substitutions
349 Falling at Non-ECAR Positions

350 The OR for the group of variants that were categorized as non-functional via the SGE
351 assay and did not fall at ECARs was 12.0 (95% CI: 7.91-18.1) with a proportion pathogenic of
352 0.71, just reaching +4 ACMG points, as stated previously. By stratifying these variants by
353 computational prediction, we can evaluate the effect of combining functional assay and
354 computational prediction data toward variant classification (Table 3). It should be noted that if
355 the computational predictions unanimously agreed with the functional assay result, the OR,

356 estimated proportion pathogenic, and ACMG points would remain unchanged after adding in
357 evidence from computational tools. Thus, the true utility of computational predictions in this
358 analysis stems from the ability of the computational tool to accurately disagree with – and
359 thereby stratify – the functional assay result.

360 Of the computational tools included in our analysis, pathogenic predictions from
361 BayesDel and REVEL did not increase the OR or the estimated proportion pathogenic when
362 combined with the non-functional assay data. It is noted that the breakdown of predictions for
363 BayesDel and REVEL are the same. This is not unexpected as both are meta-callers,
364 incorporating predictions from individual callers into a weighted overall score, and share several
365 of the same individual components [18, 19].

366 On the other hand, pathogenic predictions from Align-GVGD, AlphaMissense, MutPred2,
367 and VEST4 did increase the OR and thus the estimated proportion pathogenic from that of the
368 functional assay result, albeit not sufficiently to meet the expected value of +5 ACMG points.

369 Additionally, Align-GVGD, AlphaMissense, MutPred2, and VEST4 decreased the ORs
370 and estimated points for variants that were non-functional via assay and either indeterminate or
371 benign by computational prediction. Notably, Align-GVGD and MutPred2 yielded results in the
372 expected descending proportion pathogenic order from non-functional computationally
373 pathogenic to non-functional computationally benign. For Align-GVGD, the 23 unique variants
374 (106 case-control observations) that were non-functional via assay and predicted pathogenic
375 had an OR of 14.9 (95% CI: 7.97-28.0) and a proportion pathogenic of 0.79, which equates to
376 +4 ACMG points. The 20 unique variants (82 case-control observations) that were non-
377 functional via assay and computationally indeterminate had an OR of 10.9 (95% CI: 5.72-20.6)
378 and a proportion pathogenic of 0.68, which equates to +3 ACMG points. Finally, the 15 unique
379 variants (20 case-control observations) that were non-functional and predicted benign had an
380 OR of 6.80 (95% CI: 2.24-20.6) and a proportion pathogenic of 0.49, which equates to +2

381 ACMG points. Corresponding results for the other computational tools are summarized in Table
382 3.

383 Combining Functionally Normal Assay Results with Computational Predictions for Substitutions
384 Falling at Non-ECAR Positions

385 For the analysis of variants that were functionally normal via SGE and did not fall at
386 ECARs, we used an alternative OR ceiling at the threshold for moderate risk variants (OR = 2 in
387 a general population, which is inflated to approximately 4.4 in our dataset) to evaluate evidence
388 against pathogenicity. For the SGE assay, functional variants had an OR of 0.8 (95% CI: 0.64-
389 1.00) and an estimated proportion pathogenic of 0.00, sufficient for the -4 ACMG points
390 allowable from functional assays. The results of stratification by computational tool are
391 presented in Table 4.

392 Compared to other computational tools, BayesDel and REVEL made few predictions of
393 benignity even though this group of variants were functional via the assay. Interestingly, REVEL
394 did not predict any of the variants in this dataset to be benign and BayesDel only predicted five
395 unique variants to be so. For REVEL, the ORs for variants predicted to be pathogenic or
396 indeterminate remained close to 1.0, 1.08 (95% CI: 0.77-1.52) and 0.49 (95% CI: 0.36-0.68),
397 respectively.

398 Excepting MutPred2 and REVEL, pathogenic predictions from the other computational
399 tools elevated the OR, proportion pathogenic, and ACMG points for the subset of functional
400 variants from the assay (see Table 4). Discordance between functional assay and
401 computational tool is of sufficient magnitude to move variants in these analytical subsets from
402 likely benign (-1 to -6 ACMG points) up to VUS (0-5 ACMG points). Furthermore, predictions of
403 pathogenicity from Align-GVGD, AlphaMissense, BayesDel-VCEP, and VEST4 on functionally
404 normal variants created subsets that were statistically significantly distinct from benign
405 predictions on functionally normal variants.

406 Of particular note is the series of ORs, proportions pathogenic, and ACMG points that
407 result from stratification by Align-GVGD, AlphaMissense, and BayesDel. These tools were the
408 only ones of those evaluated in this study that yielded three appropriately ordered analytical
409 subsets for functional variants from the SGE assay. The ORs for pathogenic, indeterminate,
410 and benign Align-GVGD predictions were 2.53 (95% CI:1.00-6.38), 1.45 (95% CI: 0.93-2.26),
411 and 0.58 (95% CI: 0.44-0.77) respectively. The corresponding proportions pathogenic were
412 0.32, 0.10, and 0.00, corresponding to ACMG points of 1, 0, and -5 respectively.

413 **DISCUSSION**:

414 In the above analysis, we provide an MLE model that links the OR from a case-control
415 analysis to an estimate of the proportion of variants that are pathogenic and thereby to the odds
416 of pathogenicity and corresponding ACMG points. We also defined and identified ECARs within
417 the RING and tBRCT domains of *BRCA1*. We then demonstrated important use cases of our
418 MLE model in evaluating the strength of evidence for several functional assays alone and in
419 combination with ECAR data and computational tool predictions, as summarized in Figure 2.
420 This work takes an important step in the ability to empirically test whether the assertions of the
421 strength of evidence for different evidence types, and the combinations thereof, proposed in the
422 2015 ACMG/AMP guidelines are valid. And it does so in a way that takes account of effect size
423 while avoiding the circularities inherent in analyses that are based on re-call.

424 We analyzed the performance of several computational tools at the ECARs that were
425 identified through our multi-supergroup PMSA. While some tools, especially AlphaMissense,
426 BayesDel, and REVEL, seemed to be “aware” of the ECARs in the RING domain, far fewer
427 substitutions at ECARs in the tBRCT domain had high enough scores to receive +4 ACMG
428 points. Additionally, there was variability between tools in the number of ECARs that received
429 extreme enough scores to award +4 ACMG points (0 substitutions at ECARs received +4 from
430 MutPred2, while 56 did from BayesDel), even though most computational tools incorporate
431 evolutionary conservation into their predictions. This discrepancy may arise due to limitations in

432 creating deep, multi-supergroup PMSAs using currently available automated alignments. By
433 manually curating our alignment, we were able to include sequences that contained inserts of
434 various sizes between the RING and tBRCT domains that would likely be discarded by
435 automated approaches.

436 From our analysis of the strength of evidence that functional assays provide, we
437 conclude that functional assays can indeed reach a quantitative threshold that aligns with the
438 recommendations outlined in the ACMG guidelines, in both the benign and pathogenic
439 directions. However, the strength of evidence can differ in the pathogenic and benign directions
440 and not all assays validate as a predictor of pathogenicity. For example, variants with functional
441 results from the M2H assay in the RING domain of BRCA1 fell just short of the OP threshold for
442 -4 points, while non-functional results provided sufficient evidence for +4 points. Furthermore,
443 our structurally informed analysis of predicted loss of function substitutions in the core alpha
444 helix of the *BRCA1* coiled coil domain (amino acids 1,400 – 1,418) weigh against the hypothesis
445 that missense substitutions in this domain can be pathogenic. Thus, this study provides further
446 evidence that careful evaluation of functional assays is required prior to their use in clinical
447 settings and that +/- 4 points should not be assigned carte blanche [6].

448 For substitutions that fell at ECAR positions, we found that predictions of pathogenicity
449 from functional assay are additive in a manner consistent with ACMG guidelines (i.e +4 points
450 from PS3 and +2 points from PM1 = +6). Furthermore, these variants had an effect size similar
451 to that of truncating variants. However, for non-functional substitutions that fell at non-ECAR
452 positions, we found that concordant predictions of pathogenicity from functional assay and
453 computational tool are not additive as proposed in the ACMG guidelines (i.e +4 points from PS3
454 and +1 point from PP3 = +5). Specifically, we found that none of the tools included in this
455 analysis were able to raise the proportion pathogenic as expected and these pools of variants
456 only provided enough strength of evidence for +4 points. However, for functionally normal
457 variants in the assay, concordant predictions in the benign direction from several computational

458 tools did meet the ACMG combining expectations (i.e -4 points from BS3 and -1 point from BP4
459 = -5).

460 Conflicting predictions between functional assay and computational tools were able to
461 modulate effect size, estimated proportions pathogenic and ACMG points for the resulting
462 analytical subsets. For substitutions that were non-functional via the SGE assay and fell at non-
463 ECAR positions, benign or indeterminate predictions from several computational tools were
464 found to deprecate the evidence from the functional assay. For example, after stratification by
465 Align-GVGD, less than half (23/58) of the observed variants retained a sufficient proportion
466 pathogenic for +4 ACMG points. For substitutions that were functional via the SGE assay,
467 pathogenic or indeterminate predictions from several computational tools created statistically
468 different effect sizes after stratification.

469 While the ACMG expectations were not met for all instances of functional assay and
470 computational tool combinations, we note that two tools, Align-GVGD and MutPred2, were able
471 to produce appropriately ordered series of ORs, proportions pathogenic, and ACMG points for
472 predicted pathogenic variants. On the benign side, only Align-GVGD and AlphaMissense yielded
473 the appropriately ordered series of ORs, proportions pathogenic, and ACMG points. Thus Align-
474 GVGD was the only computational tool that was able to appropriately stratify non-functional and
475 functional results in both the pathogenic and benign directions. While Align-GVGD was not
476 explicitly trained on *BRCA1*, it was optimized for this gene. Therefore, future studies will need
477 to interrogate the performance of this and other tools on a diverse set of genes.

478 This study focused on functional assay results individually and in combination with
479 ECARs and computational tool predictions, however, this method can be applied to any
480 classifier or combination of classifiers. We chose *BRCA1* as a test case, in part, because of the
481 robust case-control dataset that is available for this gene. In instances where case-control data
482 may be more limited, a similar analysis could be conducted by pooling a defined set of genes
483 and evaluating the ability of the classifier in question to generate ordered series of ORs with

484 appropriate effect sizes. Effective evaluation of variant effect predictors will increasingly rely on
485 availability of case data to match the publicly available control data that can obtained from
486 sources such as gnomAD, the UK Biobank, and the All of Us Projects.

487 Additionally, this analysis provides a way of evaluating the strength of evidence that a
488 particular classifier, or combinations of classifiers, provides in relation to an analytical subset of
489 variants (i.e predicted benign, predicted pathogenic). It does not evaluate the pathogenicity of
490 an individual variant. Thus, even though an assay's predictions or combinations of an assay
491 and computational tool perform well in this analysis, analytical subsets of variants may contain
492 individual variants that differ in their unique role in disease from the group at large.

493 Beyond the findings from this study, further considerations may need to be adapted over
494 time and depending on use case. For example, our analysis looks at two possible OR
495 thresholds that one could use, that of nonsense variants in our dataset (OR = 21.62) and the
496 ENIGMA-defined threshold for moderate risk variants in our dataset (OR = 4.4). While the
497 resolution of which threshold to use is beyond the scope of this work; we note that careful
498 selection of a standard candle by which to interpret subsets of variants will be critical.

499 These findings are important to bear in mind as assays and other tools are validated for
500 use in variant classification. Often, functional assays and computational tools are validated by
501 re-call, using the raw classifications of variants to assess the performance of the new assay or
502 tool. It is known that this strategy leads to overly confident evaluations of variant effect
503 classifiers due to circularities inherent in training and validation datasets [34]. Furthermore, it may
504 be the case that variants that are already securely classified are different from variants that
505 remain to be classified and may have other characteristics, such as disproportionately falling at
506 ECARs, that potentially inflate the strength of evidence the new assay or tool naively seems to
507 provide. Therefore, it is important to conduct analyses that assess effect size as a validation
508 step – and be prepared to modify or reject calibrations that do not reach the required effect size.

509 This study provides a means to do this without the need for re-call, provided there are sufficient
510 case-control or family history data available.

511 In 2018, we proposed a bidirectional feedback model wherein empirical measurements
512 of strength of evidence steadily improve the rigor of variant classification [4] thereby providing
513 stepping stones towards clinical validity. This work provides a practical model that can be used
514 to achieve that end. Our MLE model assumes an underlying two-component mixture model in
515 which stratification would yield subsets of variants with effect size similar to that of truncating
516 variants (for loss of function susceptibility genes) and a subset with an effect size similar to
517 benign variants. However, this Platonic Form of clear-cut analytical subsets may not be what is
518 observed in practice. Indeed, in our analysis, variants that had non-functional assay results and
519 did not fall at ECARs had ORs above the benign threshold but below the OR for truncating
520 variants. Future work, therefore, should focus on strategies to further resolve variant
521 stratification on the continuum from benign to pathogenic.

522 The combination of functional assays and computational tools have been heralded as the path
523 forward to resolve variant classification uncertainty due to their ability to predict the outcomes of
524 large sets of variants [7, 35, 36]. However, as shown in this study, current assays, tools, and their
525 combinations may not meet the quantitative strength assigned in classification guidelines or
526 published calibration studies. Additional refinement, careful calibration, and rigorous validation
527 of variant classifiers is critical to ensure that those who use these tools are aware of the strength
528 of evidence they provide.

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.

537 **REFERENCES:**

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

542 (2) Richards, S.; Aziz, N.; Bale, S.; Bick, D.; Das, S.; Gastier-Foster, J.; Grody, W. W.; Hegde,
543 M.; Lyon, E.; Spector, E.; et al. Standards and guidelines for the interpretation of sequence
544 variants: a joint consensus recommendation of the American College of Medical Genetics and
545 Genomics and the Association for Molecular Pathology. *Genet Med* **2015**, *17* (5), 405-424. DOI:
546 10.1038/gim.2015.30 From NLM Medline.

547 (3) Tavtigian, S. V.; Greenblatt, M. S.; Harrison, S. M.; Nussbaum, R. L.; Prabhu, S. A.; Boucher,
548 K. M.; Biesecker, L. G.; ClinGen Sequence Variant Interpretation Working, G. Modeling the
549 ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genet Med*
550 **2018**, *20* (9), 1054-1060. DOI: 10.1038/gim.2017.210 From NLM Medline.

551 (4) Tavtigian, S. V.; Harrison, S. M.; Boucher, K. M.; Biesecker, L. G. Fitting a naturally scaled
552 point system to the ACMG/AMP variant classification guidelines. *Hum Mutat* **2020**, *41* (10),
553 1734-1737. DOI: 10.1002/humu.24088 From NLM Medline.

554 (5) Pejaver, V.; Byrne, A. B.; Feng, B. J.; Pagel, K. A.; Mooney, S. D.; Karchin, R.; O'Donnell-
555 Luria, A.; Harrison, S. M.; Tavtigian, S. V.; Greenblatt, M. S.; et al. Calibration of computational
556 tools for missense variant pathogenicity classification and ClinGen recommendations for
557 PP3/BP4 criteria. *Am J Hum Genet* **2022**, *109* (12), 2163-2177. DOI:
558 10.1016/j.ajhg.2022.10.013 From NLM Medline.

559 (6) Brnich, S. E.; Abou Tayoun, A. N.; Couch, F. J.; Cutting, G. R.; Greenblatt, M. S.; Heinen, C.
560 D.; Kanavy, D. M.; Luo, X.; McNulty, S. M.; Starita, L. M.; et al. Recommendations for application

561 of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant
562 interpretation framework. *Genome Med* **2019**, *12* (1), 3. DOI: 10.1186/s13073-019-0690-2 From
563 NLM Medline.

564 (7) Clark, K. A.; Paquette, A.; Tao, K.; Bell, R.; Boyle, J. L.; Rosenthal, J.; Snow, A. K.; Stark, A.
565 W.; Thompson, B. A.; Unger, J.; et al. Comprehensive evaluation and efficient classification of
566 BRCA1 RING domain missense substitutions. *Am J Hum Genet* **2022**, *109* (6), 1153-1174. DOI:
567 10.1016/j.ajhg.2022.05.004 From NLM Medline.

568 (8) Spurdle, A. B.; Greville-Heygate, S.; Antoniou, A. C.; Brown, M.; Burke, L.; de la Hoya, M.;
569 Domchek, S.; Dork, T.; Firth, H. V.; Monteiro, A. N.; et al. Towards controlled terminology for
570 reporting germline cancer susceptibility variants: an ENIGMA report. *J Med Genet* **2019**, *56* (6),
571 347-357. DOI: 10.1136/jmedgenet-2018-105872 From NLM Medline.

572 (9) Siegel, R. L.; Giaquinto, A. N.; Jemal, A. Cancer statistics, 2024. *CA Cancer J Clin* **2024**, *74*
573 (1), 12-49. DOI: 10.3322/caac.21820 From NLM Medline.

574 (10) Wallace, I. M.; O'Sullivan, O.; Higgins, D. G.; Notredame, C. M-Coffee: combining multiple
575 sequence alignment methods with T-Coffee. *Nucleic Acids Res* **2006**, *34* (6), 1692-1699. DOI:
576 10.1093/nar/gkl091 From NLM Medline.

577 (11) Armougom, F.; Moretti, S.; Poirot, O.; Audic, S.; Dumas, P.; Schaeli, B.; Keduas, V.;
578 Notredame, C. Expresso: automatic incorporation of structural information in multiple sequence
579 alignments using 3D-Coffee. *Nucleic Acids Res* **2006**, *34* (Web Server issue), W604-608. DOI:
580 10.1093/nar/gkl092 From NLM Medline.

581 (12) Potter, S. C.; Luciani, A.; Eddy, S. R.; Park, Y.; Lopez, R.; Finn, R. D. HMMER web server:
582 2018 update. *Nucleic Acids Res* **2018**, *46* (W1), W200-W204. DOI: 10.1093/nar/gky448 From
583 NLM Medline.

584 (13) Tavtigian, S. V.; Byrnes, G. B.; Goldgar, D. E.; Thomas, A. Classification of rare missense
585 substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications. *Hum*
586 *Mutat* **2008**, *29* (11), 1342-1354. DOI: 10.1002/humu.20896 From NLM Medline.

587 (14) Young, E. L.; Feng, B. J.; Stark, A. W.; Damiola, F.; Durand, G.; Forey, N.; Francy, T. C.;
588 Gammon, A.; Kohlmann, W. K.; Kaphingst, K. A.; et al. Multigene testing of moderate-risk
589 genes: be mindful of the missense. *J Med Genet* **2016**, *53* (6), 366-376. DOI:
590 10.1136/jmedgenet-2015-103398 From NLM Medline.

591 (15) Findlay, G. M.; Daza, R. M.; Martin, B.; Zhang, M. D.; Leith, A. P.; Gasperini, M.; Janizek, J.
592 D.; Huang, X.; Starita, L. M.; Shendure, J. Accurate classification of BRCA1 variants with
593 saturation genome editing. *Nature* **2018**, *562* (7726), 217-222. DOI: 10.1038/s41586-018-0461-
594 z From NLM Medline.

595 (16) Nagy, G.; Diabate, M.; Banerjee, T.; Adamovich, A. I.; Smith, N.; Jeon, H.; Dhar, S.; Liu, W.;
596 Burgess, K.; Chung, D.; et al. Multiplexed assay of variant effect reveals residues of functional
597 importance in the BRCA1 coiled-coil and serine cluster domains. *PLoS One* **2023**, *18* (11),
598 e0293422. DOI: 10.1371/journal.pone.0293422 From NLM Medline.

599 (17) Esposito, D.; Weile, J.; Shendure, J.; Starita, L. M.; Papenfuss, A. T.; Roth, F. P.; Fowler, D.
600 M.; Rubin, A. F. MaveDB: an open-source platform to distribute and interpret data from
601 multiplexed assays of variant effect. *Genome Biol* **2019**, *20* (1), 223. DOI: 10.1186/s13059-019-
602 1845-6 From NLM Medline.

603 (18) Feng, B. J. PERCH: A Unified Framework for Disease Gene Prioritization. *Hum Mutat* **2017**,
604 *38* (3), 243-251. DOI: 10.1002/humu.23158 From NLM Medline.

605 (19) Ioannidis, N. M.; Rothstein, J. H.; Pejaver, V.; Middha, S.; McDonnell, S. K.; Baheti, S.;
606 Musolf, A.; Li, Q.; Holzinger, E.; Karyadi, D.; et al. REVEL: An Ensemble Method for Predicting

607 the Pathogenicity of Rare Missense Variants. *Am J Hum Genet* **2016**, *99* (4), 877-885. DOI:
608 10.1016/j.ajhg.2016.08.016 From NLM Medline.

609 (20) Carter, H.; Douville, C.; Stenson, P. D.; Cooper, D. N.; Karchin, R. Identifying Mendelian
610 disease genes with the variant effect scoring tool. *BMC Genomics* **2013**, *14 Suppl 3* (Suppl 3),
611 S3. DOI: 10.1186/1471-2164-14-S3-S3 From NLM Medline.

612 (21) Martin, F. J.; Amode, M. R.; Aneja, A.; Austine-Orimoloye, O.; Azov, A. G.; Barnes, I.;
613 Becker, A.; Bennett, R.; Berry, A.; Bhai, J.; et al. Ensembl 2023. *Nucleic Acids Res* **2023**, *51*
614 (D1), D933-D941. DOI: 10.1093/nar/gkac958 From NLM Medline.

615 (22) Liu, X.; Li, C.; Mou, C.; Dong, Y.; Tu, Y. dbNSFP v4: a comprehensive database of
616 transcript-specific functional predictions and annotations for human nonsynonymous and splice-
617 site SNVs. *Genome Med* **2020**, *12* (1), 103. DOI: 10.1186/s13073-020-00803-9 From NLM
618 Medline.

619 (23) Pejaver, V.; Urresti, J.; Lugo-Martinez, J.; Pagel, K. A.; Lin, G. N.; Nam, H. J.; Mort, M.;
620 Cooper, D. N.; Sebat, J.; Iakoucheva, L. M.; et al. Inferring the molecular and phenotypic impact
621 of amino acid variants with MutPred2. *Nat Commun* **2020**, *11* (1), 5918. DOI: 10.1038/s41467-
622 020-19669-x From NLM Medline.

623 (24) Cheng, J.; Novati, G.; Pan, J.; Bycroft, C.; Zemgulyte, A.; Applebaum, T.; Pritzel, A.; Wong,
624 L. H.; Zielinski, M.; Sargeant, T.; et al. Accurate proteome-wide missense variant effect
625 prediction with AlphaMissense. *Science* **2023**, *381* (6664), eadg7492. DOI:
626 10.1126/science.adg7492 From NLM Medline.

627 (25) Tavtigian, S. V.; Deffenbaugh, A. M.; Yin, L.; Judkins, T.; Scholl, T.; Samollow, P. B.; de
628 Silva, D.; Zharkikh, A.; Thomas, A. Comprehensive statistical study of 452 BRCA1 missense
629 substitutions with classification of eight recurrent substitutions as neutral. *J Med Genet* **2006**, *43*
630 (4), 295-305. DOI: 10.1136/jmg.2005.033878 From NLM Medline.

631 (26) Mathe, E.; Olivier, M.; Kato, S.; Ishioka, C.; Hainaut, P.; Tavtigian, S. V. Computational
632 approaches for predicting the biological effect of p53 missense mutations: a comparison of three
633 sequence analysis based methods. *Nucleic Acids Res* **2006**, *34* (5), 1317-1325. DOI:
634 10.1093/nar/gkj518 From NLM Medline.

635 (27) Bergquist, T.; Stenton, S. L.; Nadeau, E. A. W.; Byrne, A. B.; Greenblatt, M. S.; Harrison, S.
636 M.; Tavtigian, S. V.; O'Donnell-Luria, A.; Biesecker, L. G.; Radivojac, P.; et al. Calibration of
637 additional computational tools expands ClinGen recommendation options for variant
638 classification with PP3/BP4 criteria. *bioRxiv* **2024**. DOI: 10.1101/2024.09.17.611902 From NLM
639 PubMed-not-MEDLINE.

640 (28) Li, H.; LaDuca, H.; Pesaran, T.; Chao, E. C.; Dolinsky, J. S.; Parsons, M.; Spurdle, A. B.;
641 Polley, E. C.; Shimelis, H.; Hart, S. N.; et al. Classification of variants of uncertain significance in
642 BRCA1 and BRCA2 using personal and family history of cancer from individuals in a large
643 hereditary cancer multigene panel testing cohort. *Genet Med* **2020**, *22* (4), 701-708. DOI:
644 10.1038/s41436-019-0729-1 From NLM Medline.

645 (29) Easton, D. F.; Deffenbaugh, A. M.; Pruss, D.; Frye, C.; Wenstrup, R. J.; Allen-Brady, K.;
646 Tavtigian, S. V.; Monteiro, A. N.; Iversen, E. S.; Couch, F. J.; et al. A systematic genetic
647 assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and
648 BRCA2 breast cancer-predisposition genes. *Am J Hum Genet* **2007**, *81* (5), 873-883. DOI:
649 10.1086/521032 From NLM Medline.

650 (30) Trapp, O.; Seeliger, K.; Puchta, H. Homologs of breast cancer genes in plants. *Front Plant*
651 *Sci* **2011**, *2*, 19. DOI: 10.3389/fpls.2011.00019 From NLM PubMed-not-MEDLINE.

652 (31) Keeling, P. J.; Burki, F. Progress towards the Tree of Eukaryotes. *Curr Biol* **2019**, *29* (16),
653 R808-R817. DOI: 10.1016/j.cub.2019.07.031 From NLM Medline.

654 (32) Al Jewari, C.; Baldauf, S. L. An excavate root for the eukaryote tree of life. *Sci Adv* **2023**, *9*
655 (17), eade4973. DOI: 10.1126/sciadv.ade4973 From NLM Medline.

656 (33) Parsons, M. T.; de la Hoya, M.; Richardson, M. E.; Tudini, E.; Anderson, M.; Berkofsky-
657 Fessler, W.; Caputo, S. M.; Chan, R. C.; Cline, M. S.; Feng, B. J.; et al. Evidence-based
658 recommendations for gene-specific ACMG/AMP variant classification from the ClinGen ENIGMA
659 BRCA1 and BRCA2 Variant Curation Expert Panel. *Am J Hum Genet* **2024**, *111* (9), 2044-2058.
660 DOI: 10.1016/j.ajhg.2024.07.013 From NLM Medline.

661 (34) Grimm, D. G.; Azencott, C. A.; Aicheler, F.; Gieraths, U.; MacArthur, D. G.; Samocha, K. E.;
662 Cooper, D. N.; Stenson, P. D.; Daly, M. J.; Smoller, J. W.; et al. The evaluation of tools used to
663 predict the impact of missense variants is hindered by two types of circularity. *Hum Mutat* **2015**,
664 *36* (5), 513-523. DOI: 10.1002/humu.22768 From NLM Medline.

665 (35) Starita, L. M.; Ahituv, N.; Dunham, M. J.; Kitzman, J. O.; Roth, F. P.; Seelig, G.; Shendure,
666 J.; Fowler, D. M. Variant Interpretation: Functional Assays to the Rescue. *Am J Hum Genet*
667 **2017**, *101* (3), 315-325. DOI: 10.1016/j.ajhg.2017.07.014 From NLM Medline.

668 (36) Fowler, D. M.; Rehm, H. L. Will variants of uncertain significance still exist in 2030? *Am J*
669 *Hum Genet* **2024**, *111* (1), 5-10. DOI: 10.1016/j.ajhg.2023.11.005 From NLM Medline.

670
671
672
673
674
675
676
677

678 **FIGURE LEGENDS**

679 **Figure 1**. Description of study flow

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

689

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

691 The matrix represents the main combinations of evidence evaluated in this study. The number in
692 the row or column headers represent the Bayesian point value assigned to each evidence type in
693 the ACMG guidelines. Each box has three lines: Expectation, Observed, and Percent of calls.
694 "Expectation" is the Bayesian points expected according to the ACMG guidelines. The
695 combination of ECAR, MAVE loss of function, and computational tool pathogenic was capped at
696 ≥+6 points. Otherwise expected points per category are additive as described in the ACMG
697 guidelines. The Observed line represents the Bayesian point value that was observed using the
698 MLE model in this study. The Percent all calls line represents the range, as a percent, of all
699 possible substitutions that were observed for a particular combination of evidence for the
700 computational tools evaluated. Blue filled boxes indicate that the combination of evidence met or
701 exceeded the ACMG guideline expectations in this study. Peach filled boxes indicate that the
702 combination of evidence departed from the expected additivity of the ACMG guideline in this
703 study. The bold black outline highlights that the assumption that MAVE LOF and computational

704 predictions of pathogenicity are additive is not met in this study; moreover, this combination had
705 a substantial number of observations. White filled boxes represent combinations of evidence
706 with little to no data. Dark grey filled boxes represent combinations of evidence with results that
707 may be beyond the dynamic range of the MLE model.
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729

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
C24	6	5	6	5	0	5	1	6	6
C27	6	6	4	5	0	5	0	6	6
C39	6	6	5	6	0	4	2	6	6
H41	7	7	2	6	0	3	0	7	5
C44	6	6	5	6	0	4	0	6	6
C47	6	6	6	6	0	2	3	6	6
C61	6	5	5	6	0	4	2	6	6
C64	6	6	4	6	0	5	4	6	6
T1685	5	5	0	0	0	0	0	3	1
H1686	7	7	0	0	0	0	0	6	5
T1700	6	5	0	0	0	0	0	6	3
W1718	5	4	1	5	0	2	0	5	4
R1753	5	5	0	0	0	0	0	5	4
W1837	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. Odds Ratios, Proportions Pathogenic, and ACMG Point Estimates of Functional Assay Analytical Subsets and Non-functional Substitutions Falling at ECAR Positions.

	Possible snMS[a]	Unique Variant Obs.	Clinical Obs.	gnomAD Obs.	Adjusted OR[b]	95% C.I.	M.L.E Proportion Pathogenic[c]	Estimated ACMG Point[d]	Expected ACMG Point[e]
Nonsense Substitutions	NA	213	1078	68	21.6	16.9 – 27.7	NA	NA	
M2H Assay Categories									
Functional	452	56	58	75	1.21	0.85 – 1.72	0.01	-3	-4
Indeterminate	17	1	14	0	NA	NA	NA	NA	0
Non-Functional – All	100	24	235	13	22.6	12.9 – 39.5	1.00	4	4
Non-functional – ECAR	50	14	220	11	24.8	13.5 – 45.6	1.00	≥ 6	6
Non-functional – Other	50	10	15	2	21.6	2.23 – 43.9	0.74	4	4
SGE Assay Categories									
Functional	1,185	154	121	239	0.80	0.64 – 1.00	0.00	-4	-4
Indeterminate	148	28	36	24	2.51	1.48 – 4.25	0.16	0	0
Non-Functional – All	413	77	413	37	16.3	11.6 – 27.7	0.88	4	4
Non-functional – ECAR	89	19	231	11	26.2[h]	14.3 – 48.1	1.00	≥ 6	6
Non-functional – Other	324	58	182	26	12.0[h]	7.9 – 18.1	0.71	4	4
HDR Assay Categories									
Functional	152	14	26	27	1.32	0.76 – 2.29	0.05	-1	-4
Indeterminate	15	2	0	15	NA	NA	NA	NA	0
Non-Functional – All	16	2	1	1	1.48	0.87 – 25.2	0.06	0	4
A Priori Damaging[f]	22[g]	5	1	4	0.46	0.50 – 4.24	0.00	-4	2

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
M2H = Mammilian-2-Hybrid

SGE = Saturation Genome Editing
ECAR = Exceptionally Conserved Ancestral Residue
HDR = Homology Directed Repair
a. Restricted to those snMS interrogated in the assay indicated.
b. Adjusted for race/ethnicity.
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.

Table 3. Odds Ratios, Proportions Pathogenic, and ACMG Point Estimates of Non-functional, Non-ECAR Assay Results Combined with Computational Predictions.

Computational Tool Prediction	Possible snMS[a]	Unique Variants Obs.	Clinical Obs.	gnomAD Obs.	Adjusted OR[b]	95% C.I.	M.L.E Proportion Pathogenic[c]	Estimated ACMG Point[d]	Expected ACMG Point[e]
A-GVGD									
Pathogenic	120	23	95	11	14.9	7.97 – 28.0	0.79	4	5
Indeterminate	120	20	71	11	10.9	5.72 – 20.6	0.68	3	4
Benign	91	15	16	4	6.80	2.24 – 20.6	0.49	2	3
AlphaMissense									
Pathogenic	232	43	171	19	15.4	9.56 – 24.8	0.81	4	5
Indeterminate	80	14	9	7	2.03	0.86 – 10.4	0.12	1	4
Benign	13	1	2	0	NA	NA	NA	NA	3
BayesDel									
Pathogenic	284	52	175	25	11.9	7.80 – 18.2	0.71	4	5
Indeterminate	46	6	7	1	13.5	1.64 – 110.8	0.71	4	4
Benign	0	0	0	0	NA	NA	NA	NA	3
BayesDel - VCEP									
Pathogenic	231	47	170	23	12.6	8.14 – 19.6	0.73	4	5
Indeterminate	45	5	5	2	3.62	0.68 – 19.2	0.32	1	4
Benign	55	6	7	1	13.5	1.64 – 110.8	0.71	4	3
MutPred2									
Pathogenic	142	30	115	15	13.4	7.82 – 23.7	0.75	4	5
Indeterminate	159	25	65	10	10.6	5.40 – 20.7	0.68	3	4
Benign	29	3	2	1	3.52	0.31 – 40.4	0.24	1	3
REVEL									
Pathogenic	285	52	175	25	11.9	7.80 – 18.3	0.71	4	5
Indeterminate	46	6	7	1	13.5	1.64 – 110.8	0.71	4	4
Benign	0	0	0	0	NA	NA	NA	NA	3
VEST4									
Pathogenic	213	43	167	21	13.5	8.54 – 21.3	0.76	4	5
Indeterminate	108	15	15	5	5.49	1.97 – 15.3	0.38	2	4
Benign	10	0	0	0	NA	NA	NA	NA	3

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. Restricted to those snMS interrogated in the assay indicated.
b. Adjusted for race/ethnicity.
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.

Table 4. Odds Ratios, Proportions Pathogenic, and ACMG Point Estimates of Functionally Normal Assay Results Combined with Computational Predictions.

Computational Tool Prediction	Possible snMS[a]	Unique Variant Obs.	Clinical Obs.	gnomAD Obs.	Adjusted OR[b]	95% C.I.	M.L.E Proportion Pathogenic[c]	Estimated ACMG Point[d]	Expected ACMG Point[e]
A-GVGD									
Pathogenic	61	9	11	8	2.53[f]	1.00 – 6.38	0.32	1	-3
Indeterminate	212	25	40	43	1.45	0.93 – 2.26	0.10	0	-4
Benign	912	120	70	188	0.58[f]	0.44 – 0.77	0.00	-5	-5
AlphaMissense									
Pathogenic	122	19	19	17	1.70[f]	0.87 – 3.32	0.20	1	-3
Indeterminate	586	78	78	115	1.07	0.80 – 1.44	0.00	-4	-4
Benign	477	57	24	107	0.36[f]	0.23 – 0.56	0.00	-5	-5
BayesDel									
Pathogenic	379	54	65	77	1.36	0.97 – 1.90	0.06	0	-3
Indeterminate	769	95	54	155	0.54	0.36 – 0.74	0.00	-4	-4
Benign	37	5	2	7	0.52	0.11 – 2.55	0.00	-5	-5
BayesDel - VCEP									
Pathogenic	152	20	21	28	1.21[f]	0.68 – 2.15	0.00	-4	-3
Indeterminate	181	29	35	45	1.20	0.76 – 0.83	0.02	-3	-4
Benign	852	105	65	166	0.62[f]	0.46 – 0.83	0.00	-5	-5
MutPred2									
Pathogenic	86	6	4	7	1.10	0.32 – 3.81	0.00	-4	-3
Indeterminate	399	57	62	79	1.14	0.81 – 1.61	0.02	-2	-4
Benign	697	91	55	153	0.60	0.44 – 0.82	0.00	-5	-5
REVEL									
Pathogenic	413	52	56	85	1.08	0.77 – 1.52	0.00	-4	-3
Indeterminate	769	102	65	154	0.49	0.36 – 0.68	0.00	-4	-4
Benign	3	0	0	0	NA	NA	NA	NA	-5
VEST4									
Pathogenic	116	14	19	15	2.56[f]	1.29 – 5.07	0.27	1	-3
Indeterminate	763	100	84	139	0.88	0.67 – 1.16	0.00	-4	-4
Benign	306	40	18	85	0.37[f]	0.22 – 0.62	0.00	-5	-5

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. Restricted to those snMS interrogated in the assay indicated.
b. Adjusted for race/ethnicity.
c. The OR threshold used for this calculation is that of moderate risk variants (OR = 2) adjusted for observational inflation in our data to OR = 4.4.
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. Indicates that these ORs are statistically significant from one another. Comparison is made between the predicted benign and predicted pathogenic categories for a given computational tool.

Figure 1

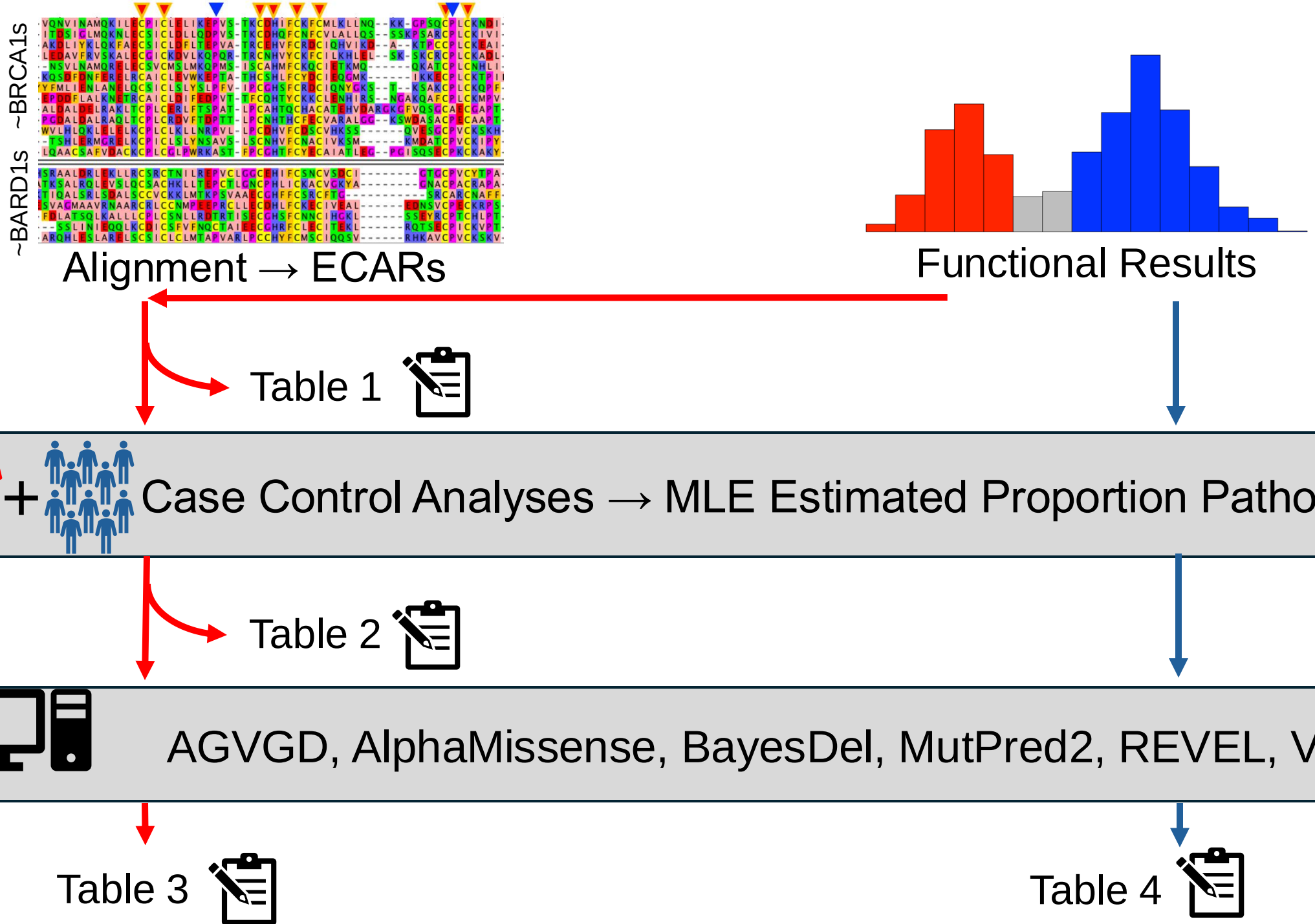
~BRCA1s
~BARD1s
Alignment → ECARs
Functional Results
Table 1
Case Control Analyses → MLE Estimated Proportion Pathogenic
Table 2
AGVGD, AlphaMissense, BayesDel, MutPred2, REVEL, VEST4
Table 3
Table 4

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 Percent of calls: (6.42% - 15.24%)	Expectation: -3 ACMG Points Observed: -4 to +1 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 +4 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 invdividual points