Title: High-Throughput Functional Evaluation of *BRCA2* Variants of Unknown Significance

Authors: Masachika Ikegami[1,2], Shinji Kohsaka[1], Toshihide Ueno[1], Yukihide Momozawa[3], Kenji Tamura[4], Akihiko Shimomura[4], Noriko Hosoya[5,6], Hiroshi Kobayashi[2], Sakae Tanaka[2], and Hiroyuki Mano[1]

Affiliations:

[1]Division of Cellular Signaling, National Cancer Center Research Institute, Tokyo, Japan

[2]Department of Orthopaedic Surgery, Faculty of Medicine, The University of Tokyo, Tokyo, Japan

[3]Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan

[4]Department of Breast and Medical Oncology, National Cancer Center Hospital, Tokyo, Japan

[5]Department of Medical Genomics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[6]Laboratory of Molecular Radiology, Center for Disease Biology and Integrative Medicine, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan.

Running title: High-throughput functional evaluation of *BRCA2* variants

Keywords: *BRCA2*, poly(ADP-ribose) polymerase (PARP) inhibitor, variants of unknown significance (VUSs), companion diagnosis, Bayesian hierarchical model

Corresponding author:

Shinji Kohsaka and Hiroyuki Mano

5-1-1 Tsukiji, Chuo-ku, Tokyo, Japan, 104-0045

Phone: +81-3-3547-5201

Fax: +81-3-5565-0727

E-mail: skohsaka@ncc.go.jp (SK) or hmano@ncc.go.jp (HM)

Financial support: This study was supported by the World-leading Innovative Graduate Study Program for Life Science and Technology, The University of Tokyo, as part of the WISE Program (Doctoral Program for World-leading Innovative & Smart Education), MEXT, Japan; grants from the Program for Integrated Database of Clinical and Genomic Information under grant number JP18kk0205003, the Leading Advanced Projects for Medical Innovation (LEAP) under grant number JP18am0001001, and the Practical Research for Innovative Cancer Control under grant number JP18ck0106252 from the Japan Agency for Medical Research and Development, AMED; a grant for Endowed Department (Department of Medical Genomics, Graduate School of Medicine, The University of Tokyo) from Eisai Co., Ltd.

Conflict of Interest Statement: The authors certify that no actual or potential conflict of interest in relation to this article exists.

ABSTRACT

Numerous nontruncating missense variants of the *BRCA2* gene have been identified, but there is a lack of convincing evidence, such as familial data, demonstrating their clinical relevance and they thus remain unactionable. To assess the pathogenicity of variants of unknown significance (VUSs) within *BRCA2*, we developed a novel method, the MANO-B method, for high-throughput functional evaluation utilizing *BRCA2*-deficient cells and poly(ADP-ribose) polymerase (PARP) inhibitors. The estimated sensitivity and specificity of this assay compared to those of the International Agency for Research on Cancer (IARC) classification system were 95% and 95%, respectively. We classified the pathogenicity of 186 *BRCA2* VUSs with our original computational pipeline, resulting in the classification of 126 mutations as "neutral/likely neutral", 23 as "intermediate", and 37 as "deleterious/likely deleterious". We further invented a simplified, on-demand annotation system, the Accurate BRCA Companion Diagnostic (ABCD) test, as a companion diagnostic for PARP inhibitors in patients with unknown *BRCA2* VUSs. The ABCD test classification was reproducible and consistent with that of a large-scale MANO-B method.

INTRODUCTION

The *BRCA1* and *BRCA2* (*BRCA*) genes encode key proteins in the homology-directed DNA break repair (HDR) pathway, and their inactivation predisposes individuals to cancer development [1]. Germline loss-of-function variants in *BRCA* markedly increase the risk of early-onset breast and ovarian cancer; in such cases, prophylactic oophorectomy and mastectomy and genetic testing for at-risk relatives must be considered [2, 3, 4]. Tumors with pathogenic mutations within *BRCA* and defective HDR have been shown to be particularly sensitive to platinum-based chemotherapies and PARP inhibitors, the efficacy of which is mediated through synthetic lethality in cancer cells with *BRCA* loss-of-function [5, 6].

The American College of Medical Genetics and Genomics (ACMG) standards and guidelines for the interpretation of sequence variants recommend a process for variant classification based on criteria using population, computational, functional, and segregation data [7]. Family-based studies including a multifactorial model of pathology, a cosegregation profile, and the cooccurrence and family history of cancer may exemplify the most reliable methods for classifying *BRCA2* gene variants [8, 9]. Nonsense or frameshift mutations within the coding exons of *BRCA* markedly alter the structures of the protein products and are presumed to confer loss-of-function. However, the vast majority of missense variants are individually rare in both general populations and cancer patients, and case–control studies may not have sufficient statistical significance to classify these variants as pathogenic or benign [10, 11, 12]. No current *in silico* computational prediction algorithm for missense variants is accurate enough when used alone [13]. Thus, the functional evaluation of missense variants of unknown significance (VUSs) is urgently required to improve the interpretation of variants identified by genetic testing and to support clinical decision-making for their carriers [14].

While thousands of *BRCA1* VUSs have been assessed by recently developed high-throughput functional assays [15, 16, 17], a few hundred *BRCA2* variants have been evaluated by a conventional functional assay for *BRCA2*, the HDR assay [18, 19]. The HDR assay has three issues requiring improvement: (i) the throughput is quite low, (ii) it uses a hamster cell line with transient expression of *BRCA2* cDNA, and most importantly, (iii) it can evaluate only variants in the DNA-binding domain [14, 20]. To overcome these limitations, we propose herein a high-throughput method using a human cell line stably expressing *BRCA2* variants that enables the evaluation of all exonic variants of the *BRCA2* gene.

RESULTS

Stable transduction of *BRCA2* variants into DLD1 *BRCA2* (−/−) cells

The introduction and stable expression of *BRCA2* variants in human cells is technically difficult owing to the relatively large coding sequence of this gene (10.2 kbp) [14]. We addressed this issue by employing a piggyBac transposon vector suitable for the stable introduction of large DNA sequences into the genome [21]. In addition, we utilized a *BRCA2* knockout human colorectal adenocarcinoma cell line, DLD1 *BRCA2* (−/−), which is known to be highly sensitive to PARP inhibition compared to parental cells retaining *BRCA2* [22, 23]. If a *BRCA2* missense variant is further introduced into DLD1 *BRCA2* (−/−) cells, the change in PARP inhibitor sensitivity likely reflects the function of the introduced variant. For instance, the expression of a neutral variant in DLD1 *BRCA2* (−/−) cells should restore HDR and thus resistance to PARP inhibitors.

Initially, a total of 107 *BRCA2* variants were selected from a curated database, the BRCA Exchange [24] (**Supplementary Table 1**). Of these, 32, 10, and 65 variants were classified as benign (Class 1/2), pathogenic (Class 4/5), and VUSs (Class 3), respectively, according to the multifactorial five-tier classification system developed by the International Agency for Research on Cancer (IARC) [10, 11, 18, 25, 26]. These *BRCA2* variant cDNAs were generated by site-directed mutagenesis and were then subcloned into the piggyBac vector containing unique 10 bp DNA barcode sequences. These individual piggyBac plasmids, together with the hyPBase transposase expression vector, were transfected into DLD1 *BRCA2* (−/−) cells [27].

The transduction efficiency for 20 selected variants was evaluated by real-time quantitative reverse transcription PCR (qRT-PCR) and digital droplet PCR (**Supplementary Fig. 1A, B and Supplementary Table 2**). The mRNA expression levels were within the physiological

range of endogenous *BRCA2* with no significant difference among the variants, whereas the copy numbers of the integrated *BRCA2* cDNA were approximately 10. Western blotting demonstrated that the protein expression levels of the 19 BRCA2 variants were generally equal to that of wild-type BRCA2 (**Supplementary Fig. 1C**).

Establishment of MANO-B method

We next performed a cell viability assay to evaluate the drug sensitivity of the *BRCA2* variants. Cells treated with PARP inhibitors (olaparib, niraparib, and rucaparib) or carboplatin (CBDCA) at various concentrations for 6 d were assessed with PrestoBlue reagent (**Supplementary Fig. 2 and Supplementary Table 2**). Wild-type *BRCA2* and benign variants showed greater tolerance to these drugs than pathogenic variants. However, the threshold between the benign and pathogenic variants was unclear. It is important to note that wild-type *BRCA2*-induced cells were significantly more sensitive than parental DLD1 cells; therefore, we should establish a sensitive method in which all comparisons could be made with wild-type-induced cells rather than parental cells. We thus modified the mixed-all-nominated-in-one (MANO) method [28, 29] that was originally developed to enable high-throughput evaluation of VUSs in transforming genes and invented the MANO-*BRCA* (MANO-B) method for the functional evaluation of *BRCA2* variants (**Fig. 1**).

In the MANO-B method, DLD1 *BRCA2* (−/−) cells were transfected with the individual *BRCA2* variant cDNAs and the variant-specific barcodes, mixed, and cultured with or without drugs. Genomic DNA was extracted from each cell mixture after the drug treatments, and the barcode sequences were amplified by PCR and deep sequenced. The barcode read count, which was linearly correlated to the number of viable cells harboring the corresponding variant [28],

141 was normalized to that of the vehicle control to evaluate the relative cell viability. Based on
142 these profiles, the drug sensitivity and pathogenicity of each variant were determined (**Fig. 1**).
143
144 **MANO-B method of 107 *BRCA2* variant analysis using PARP inhibitors and CBDCA**
145 We performed the MANO-B method using the 107 variants with various concentrations of four
146 drugs. The difference in drug sensitivity among the variants was clearly demonstrated at high
147 drug concentrations (**Fig. 2 and Supplementary Table 3**), and the optimal concentrations were
148 thus determined to be 2.0, 0.5, 2.0, and 2.0 μM for olaparib, niraparib, rucaparib, and CBDCA,
149 respectively.
150
151 The relative viability of the cells treated with each drug was first calculated by comparing the
152 viability with that of vehicle-treated cells, then normalized to that of cells expressing wild-type
153 *BRCA2*. All variants were sorted by relative viability at the optimal concentrations. A scatter
154 plot of the relative viability of olaparib and the other three drugs revealed an intermediate group
155 including the P2329L, P2639L, D2913H, and S3291C mutants between the benign and
156 pathogenic groups (**Supplementary Fig. 3 and Supplementary Table 3**). With this
157 intermediate group as a boundary, all known pathogenic variants showed lower relative
158 viability, whereas all but two known benign variants—R2842H and V2908G—showed higher
159 relative viability. The two variants were indeed shown to be hypomorphic by HDR assay in a
160 previous study [18]. In addition, we found that one likely pathogenic variant (N3187K) defined
161 by ClinVar exhibited neutral function as evaluated by the MANO-B method. This discordance
162 between the functional classification and the clinical classification suggests potential errors in
163 the clinical annotations for rare variants.
164

We believe that the discrepancy in the classification of IARC Class 1/2 variants (R2842H and V2908G) that were evaluated as fClass 3 and fClass 4, respectively, using the MANO-B method with niraparib was because of the unreliable evidence which the IARC criteria was based on. The IARC classification is based on epidemiological evidence such as cosegregation data and other family-based genetic analyses [10]. According to the Genome Aggregation Database, both R2842H and V2908G (minor allele frequency = 6.34×10^{-5} and 1.69×10^{-5}) were rare variants. Therefore, it might be difficult to obtain solid evidence for these variants. To assess the pathogenicity of such variants accurately, the classification method must be based on a wide variety of evidence. Therefore, we reinterpreted all the 59 Class 1/2/4/5 variants (37 benign and 22 pathogenic) used in the MANO-B method according to the ACMG 2015 guidelines as previously described (Supplementary Table 1) [26,30]. As a result, 18 variants (31%), including R2842H and V2908G, were classified into VUSs because of insufficient evidence (**Supplementary Table 1**). Furthermore, N3187K, defined as a likely pathogenic variant by ClinVar, is also defined as VUS according to both IARC and ACMG criteria. The criteria of the ACMG guidelines are more conservative and stringent than those of other classification methods because ACMG uses multiple criteria from different points of view and perspectives [7]. This discrepancy among clinical databases implies potential errors in the interpretation of clinical variants. Moreover, the IARC and ACMG classification methods were designed for detecting high-risk pathogenic variants, and therefore, hypomorphic function variants or moderate cancer risk variants could be classified inappropriately [25]. Hence, we regard the discrepancy between functional and epidemiological evidence as reasonable.

To evaluate the evenness of the pooled variant ratio, we collected cell pellets at day 0 (the day the drug treatment was started) to count the barcode reads by MiSeq. Regarding the raw data obtained from batch #1, the minimum, maximum, and median read counts were 778, 13,294,

and 5,763 for N55S, R3385H, and Y3098H, respectively. The distribution of each variant ratio is shown as a histogram (**Supplementary Figure 4A**). A total of 97 variants (90%) were within the range of 0.5–2-fold average. We also calculated the fold change from day 0 to day 12 with DMSO (dimethyl sulfoxide) treatment (**Supplementary Figure 4B**). The majority of cells expressing these variants generally exhibited a uniform growth, and the functions of the induced *BRCA2* variants did not have much impact on cell growth. The unevenness of the variant ratio probably has a small effect on the result of the MANO-B method because this method is based on fold-change calculation and the abundance bias is corrected through analysis.

Evaluation of 244 *BRCA2* variants by applying Bayesian inference to MANO-B method

We expanded the experiment to encompass 244 variants, including 186 VUSs, at the optimal concentration of each drug, and obtained an assay dataset comprising 7,344 individual viability values (**Supplementary Tables 1 and 3**). Using the optimal threshold for receiver operating characteristic (ROC) analysis, the estimated sensitivity and specificity of the MANO-B method for variant pathogenicity compared with those of the IARC classification system were 100% and 95%, respectively (**Supplementary Fig. 5**).

Given that the existence of intermediate variants was suggested, a new nondichotomous mathematical approach was explored to precisely determine the pathogenicity of each variant. Relative viability data were log normalized and compensated using the values of the wild-type and the D2723H variant as standard values for benign and pathogenic variants, respectively. The resulting viability data followed a two-component Gaussian mixture model by the expectation–maximization algorithm (**Supplementary Fig. 6**).

In the following analysis, we adopted the Bayes factor (BF) as the strength of evidence in favor of pathogenicity. To calculate the BF, we utilized a Bayesian hierarchical two-component Gaussian mixture model based on the VarCall model [18, 31, 32] with a noninformative prior probability. In this setting, the BFs were calculated as the probability ratio of a variant being deleterious to it being neutral. A five-tiered functional classification based on the BF was assigned to each variant according to established criteria consistent with the ACMG variant evaluation guidelines as follows [33]: fClass 1 (neutral; $BF \leq 0.003$), fClass 2 (likely neutral; $0.003 < BF \leq 0.053$), fClass 3 (intermediate; $0.053 < BF < 18.7$), fClass 4 (likely deleterious; $18.7 \leq BF < 350$), or fClass 5 (deleterious; $350 \leq BF$). Under this definition, fClass 5 can be regarded as extremely strong evidence of pathogenicity in the framework of the ACMG guidelines.

The function of a variant, v, was estimated by a variant-specific effect, η_v. Using 22 known pathogenic and 37 known benign variants as the training set, the η_v values of benign and pathogenic variants were gradually distributed (**Supplementary Fig. 7**). The sensitivity and specificity of the Bayesian inference for detecting IARC class 4/5 variants were 95% and 95%, respectively. We also analyzed the concordance between the ACMG classification and the MANO-B classification. All the 41 benign/pathogenic variants (22 benign and 19 pathogenic) used in the MANO-B method were clearly divided into two groups (Supplementary Figure 8).

To validate the BF-based classification from the functional perspective, we set two functional classification thresholds determined by the ACMG benign/pathogenic variants to classify the following three components according to a previous report: neutral, intermediate, and deleterious [15, 18]. The higher threshold between neutral and intermediate components is defined as the minimal effect of benign variants according to the ACMG criteria (R18H, $\eta = -1.38$),

240 whereas the lower threshold between intermediate and deleterious components is defined as
241 the maximal effect of pathogenic variants (C2535X truncate variant, $\eta = -2.17$). As a result,
242 all the functional effects of fClass 1/2 variants were above the higher threshold, whereas all
243 those of fClass 4/5 variants were below the lower threshold. Now, fClass 4/5 variants can be
244 regarded as pathogenic because their probability is in the deleterious component and their
245 functional levels are equivalent to those of the known pathogenic variants.
246
247 Notably, the fClass values of the individual variants for the four drugs were concordant (**Fig.**
248 **3 A–C**). The data for 0.5 μM niraparib exhibited the highest resolution in the evaluation of the
249 variants of intermediate function. We assume that the difference in functional estimation
250 among drugs in the MANO-B method was primarily because the drug concentrations of
251 olaparib and rucaparib were not as optimal as that of niraparib. The concentration of 2.0 μM of
252 olaparib, rucaparib, and CBDCA was probably too high for DLD1 cells expressing
253 hypomorphic variants to survive, considering that the cell viability assay with 1 μM of olaparib,
254 rucaparib, and carboplatin (shown in Supplementary Figure 2B) demonstrated broader
255 transitional zones between benign and pathogenic variants. Among the 186 VUSs analyzed in
256 the full dataset with niraparib, 98, 28, 23, 6, and 31 variants were classified as fClass 1 to 5,
257 respectively (**Fig. 3D, Supplementary Table 3**). All variants outside functional domains
258 belonged to fClass 1/2. Most deleterious variants were located in the DNA-binding domain
259 (DBD). Other deleterious variants were W31C/G/L in the transactivation domain (TAD) [34],
260 P2329L in the MEILB2-binding domain (MBD) [35], and S3291C in the C-terminal domain
261 (CTD), which is a key phosphorylation site for BRCA2-RAD51 interaction [22].
262
263 This Bayesian model allowed a combinational classification based on log-normalized relative
264 viability data with prior probability, such as the Align-GVGD classification based on the

evolutionary conservation and biophysical properties of amino acids [36] (**Supplementary Fig. 9**). However, this combinational classification for variants outside the DBD should be interpreted with caution because no reliable prior probability data have been established for this region.

Model validation

The internal validation of our model was confirmed by posterior predictive checks. The posterior predictive distribution of the log-normalized relative viability matched the observed data (**Supplementary Fig. 10A**). Quantile-Quantile (Q-Q) plots showed that the posterior expected standardized residuals of relative viability exhibited a normal distribution (**Supplementary Fig. 10B**). Furthermore, a strong correlation between the MANO-B method and the HDR assay for 24 selected variants was confirmed (**Supplementary Fig. 11 and Supplementary Table 2**). The results of the two assays revealed a small discordance in some variants. The S3291C variant exhibited a relatively high value in the HDR assay compared to that in the MANO-B method, probably because S3291 is the phosphorylation site for CDK1/2 and the HDR assay cannot evaluate its function exactly [37].

Development of Accurate BRCA Companion Diagnostic (ABCD) test

To rapidly evaluate the pathogenicity of *BRCA2* VUSs in the clinic, the Accurate BRCA Companion Diagnostic (ABCD) test was developed as a potential companion diagnostic for PARP inhibitor treatment. Newly detected *BRCA2* cDNAs, along with four control variants for benign (wild-type and T2515I) and pathogenic (Y2660D and D2723H) mutations, were generated for assessment by a small-scale MANO-B method. The resulting data correlate with the data from previous MANO-B experiments, enabling batch effect adjustment for pathogenicity determination. When a *BRCA2* VUS is believed to be related to splicing

aberration, its mRNA should be evaluated by RNA sequencing of patient samples [38]. Any aberrant *BRCA2* transcripts detected should also be subjected to the ABCD test (**Fig. 4**).

An experiment was designed to confirm the accuracy of the ABCD test, in which I1929V, D3095E, Y3035S, and S3291C were assessed as if they were novel variants. The classification of neutral and deleterious variants was almost reproducible and consistent with that observed for previous large-scale batches; however, the classification of the D3095E pathogenic variant in batch #1 was underdiagnosed (from fClass5 to fClass4, **Table 1**). One misclassification is crucial in case we judge the clinical application based on the result of a single batch. Because transfection efficiency might cause inconsistent results in the ABCD test, we recognized the importance of multiple independent transfections using different batches of mutants to ensure the accuracy of the assay. Therefore, we combined the results from three batches with different transfections and performed a statistical analysis to annotate the pathogenicity. This approach improved the consistency (4/4, 100%) between the ABCD test and the large-scale MANO-B method, although the validation study remains to be performed in a larger scale.

DISCUSSION

Several functional assays for *BRCA2* variants have been developed to date based on the various functions of BRCA2 protein. However, it should be noted that loss of a certain function does not directly lead to cancer predisposition. For example, an established R3052W pathogenic variant did not show pathogenicity in the centrosome amplification assay [39], whereas G2353R, once annotated as a benign variant, showed a pathogenic response in the spontaneous homologous recombination assay [40]. The results of each functional assay should be carefully interpreted together with clinical genetic data.

It is remarkable that the distribution of the functional parameter η_v for the MANO-B method was continuous from the neutral to the deleterious component and not clearly segregated into two components, in accordance with the HDR assay [18], suggesting the existence of variants of intermediate function [41].

For the nondichotomous classification, a high-resolution quantitative assay together with an appropriate inference method are necessary. In the MANO-B method, all cell clones with individual variants are evenly mixed and cultured in one dish. Since each clone grows competitively under the same culture conditions, including pH, temperature, and drug concentration, we can minimize the technical error and bias and evaluate even small differences in BRCA function among the variants. This is a great advantage over other methods such as the cell viability assay and the HDR assay, which test variants individually.

Recently, Findlay *et al.* reported a robust saturation genome editing method by the use of a haploid cell line, HAP1, that depends on the HDR pathway for growth [15]. In contrast to the MANO-B method, their technology may accurately evaluate the function of variants at the

splicing sites or the promoters of *BRCA* genes. On the other hand, our method uses a diploid cell line and, therefore, may better reflect HDR function in physiological conditions. The MANO-B method is, further, able to directly interrogate the sensitivity of each VUS to different PARP inhibitors. The MANO-B method and the saturation genome editing method may thus be complementary to each other.

While clinical gene testing has identified plenty of *BRCA2* mutations, most VUSs are rare and thus there is a lack of substantial familial data. Recently, the uncommon (but not rare) variants G2508S (IARC Class 4) and Y3035S (IARC Class 2) were reported in a large case–control study showing indecisive cancer predisposition [18, 41]. These two variants demonstrate the clinical utility of the MANO-B method for clinical annotation of VUSs well. G2508S (η_v = −2.42, fClass 4) showed hypomorphism, while Y3035S (η_v = −0.63, fClass 1) showed neutral function, although the possibility of splicing aberration by the c.9104A>C single nucleotide variation was not excluded. We consider that G2508S is a hypomorphic variant because the familial data is consistent with the functional result, whereas the pathogenicity of Y3035S is indeterminate because of the conflict among the data. The clinical annotation of such uncommon variants could be assessed by a combinational approach, although additional clinical data are needed to identify the optimal thresholds for the MANO-B method. It is not yet known how to deal with patients harboring *BRCA2* hypomorphic variants. However, we believe these patients are probably eligible for PARP inhibitor treatment. Judging in combination with *BRCA*ness assessment is a good option to predict the effectiveness of PARP inhibitors.

Overall, the MANO-B method is a robust and scalable analysis approach for evaluating the function of *BRCA2* variants. As we developed a dataset of 244 *BRCA2* variants, including

355 controls, additional independent analysis can be merged for precise evaluation using batch
356 effect compensation. Importantly, this technique is applicable to the high-throughput functional
357 assessment of other tumor suppressor genes, augmenting the functional data for rare variants
358 and identifying the association of these variants with cancer susceptibility.
359
360 Furthermore, we propose the ABCD test as a companion diagnostic for PARP inhibitor
361 treatment in patients newly identified as *BRCA2* VUS carriers. There are two scenarios in
362 which the result of the MANO-B method can be applied to patients in the framework of the
363 ACMG guidelines; one is for predicting the effectiveness of PARP inhibitors for patients with
364 cancer harboring *BRCA2* VUSs, and the other is for considering prophylactic surgery to prevent
365 cancer in patients who at the time are healthy. In the former scenario, clinicians should
366 promptly decide on the usage of PARP inhibitors for patients with cancer harboring *BRCA2*
367 VUSs to prevent tumor progression. Therefore, a rapid evaluation of *BRCA* function using the
368 ABCD test could be a companion diagnostic method for PARP inhibitor treatment, although
369 the clear threshold of fClass as a marker to predict which variant is actually beneficial for PARP
370 inhibitor treatment in the clinical setting should be determined by clinical trials. In the latter
371 scenario, healthy patients with hereditary *BRCA2* VUSs can wait and observe for several years
372 for the tumor to emerge. After receiving genetic test results, approximately half of the women
373 wait more than 12 months before undergoing prophylactic salpingo-oophorectomy [42]. The
374 application of prophylactic surgery should be considered carefully along with genetic
375 counseling because it is irreversible and the patient does not have any life-threatening illness
376 at the moment. The penetration rate of cancer likely depends on the extent of *BRCA* function
377 deficiency; healthy patients with a *BRCA* variant of fClass 3 or 4 might be referred to careful
378 check-up rather than prophylactic surgery.
379

Tumor response to PARP inhibitor therapy has several biomarkers other than *BRCA*1/2 inactivation, such as EMSY amplification, Fanconi anemia pathway inactivation, and other HR gene defects [43]. Therefore, it is a good method to evaluate "*BRCA*ness" by whole-genome sequencing of tumors based on mutation and rearrangement signatures without the knowledge of the precise causative mutations[44, 45]. Nones *et al.* demonstrated that 21 breast cancers from 22 *BRCA2* pathogenic mutation carriers (95%) in the study exhibited "*BRCA*ness," whereas only one was *BRCA*-proficient [45]. In contrast to the test of BRCAness, the ABCD test is a simple functional method for evaluating *BRCA2* variant pathogenicity. Although the ABCD test cannot explain the *BRCA*ness of non-*BRCA1/2* tumors, it is a rapid and precise method that directly evaluates the PARP inhibitor sensitivity of cancers harboring *BRCA2* VUSs. In addition, an advantage of the ABCD test over *BRCA*ness analysis is that the ABCD test can evaluate whether germline *BRCA* variants found in people without cancer are pathogenic variants predisposing to cancer. We believe that both the ABCD test and the whole-genome sequencing of tumors can be used in parallel for the assessment of "*BRCA*ness." As the ABCD test reported in this study was limited to *BRCA* functional analysis, we sought to apply this test to other HR genes in the next study.

There are several limitations regarding the use of the MANO-B method for clinical annotation. The MANO-B method evaluates HDR function of *BRCA2* variants via PARP inhibitor sensitivity and estimates their pathogenicity indirectly. This method does not directly consider other *BRCA2* functions, such as the regulation of the G2–M transition or transcriptional elongation [46, 47]. Although the high concordance of the MANO-B method with the IARC and ACMG classifications suggest that *BRCA2* pathogenicity can be evaluated primarily by investigating HDR activity, we would like to note that BFs for deleteriousness do not have a direct correlation with the likelihood ratios for pathogenicity, as the "neutral" or "deleterious"

405 variants defined by this functional assay are not necessarily “benign” or “pathogenic” variants
406 in the clinical setting. The pathogenicity of a variant should be evaluated in a comprehensive
407 framework such as the ACMG guidelines. Another limitation of our cDNA-based assay is that
408 this approach cannot rule out potential effects on splicing. Importantly, although there are some
409 computational predictors for possible splicing aberrations, either algorithms could not
410 demonstrate adequate reliability for clinical usage, especially for mutations outside of the
411 consensus splice sites [48]. Comprehensive splicing functional assays such as hybrid minigene
412 assay and saturation genome editing technique would reinforce the evidence[15, 49, 50]. It is also
413 recommended to perform RNA-seq of patient blood and tumor to detect aberrantly spliced
414 mRNA. In addition, the MANO-B method was performed in this study using only one cell line,
415 and it is unclear whether the result of the MANO-B method is reproducible in other cell lines.
416 To evaluate the variant functions precisely and robustly, replication studies with multiple cell
417 lines with various homologous recombination statuses are desired.
418
419 We believe that the MANO-B method will lead to a better understanding of cancer biology and
420 provide a new concept of clinical functional annotation to improve cancer diagnosis and
421 treatment.
422

METHODS

Cell lines and culture conditions

Human colorectal adenocarcinoma cell line DLD1 parental cells and homozygous *BRCA2* (-/-) variant cells were purchased from Horizon Discovery, Inc. Cells were cultured in RPMI 1640 medium containing 10% fetal bovine serum and 2 mM L-glutamine at 37 °C in 5% CO_2 and passaged at 70–90% confluency with 0.05% TrypLE Express (all from ThermoFisher Scientific).

Choice of variants

A total of 239 *BRCA2* missense variants identified in the BRCA Exchange database were selected. Sequence nomenclature was based on Human Genome Variant Society (HGVS) and National Center for Biotechnology Information (NCBI) reference NP_000050.2: p.V2466A. We randomly selected 155 variants from functional domains and an additional 89 variants outside functional domains from throughout the coding sequence. Variants in functional domains were selected considering Align-GVGD in silico prediction. Between 14 and 19 VUSs were randomly selected from each of the five Align-GVGD C15–C55 categories, and 62 variants were randomly selected from the C65 category. The proportion of each category was defined as previously reported [18]. An additional 5 nonsense variants thought to be definitely pathogenic were selected as controls (**Supplementary Table 1**). Eight VUSs located in the last 3 exonic bases of the splice donor site and 5 VUSs located in the first 2 exonic bases of the splice acceptor site were possible candidates for disrupting splice consensus sequences. These VUSs were analyzed using Splicing Prediction in Consensus Element (SPiCE) [48], the most accurate in silico splice site prediction algorithm for *BRCA1/2*. Only 6 variants—V159M (c.475G>A), D23Y (c.67G>T), V211I (c.631G>A), R2336P (c.7007G>C), and R2602T

(c.7805G>C)—were predicted to alter canonical splicing, and we assessed the effect of these variants on protein function by the MANO-B method (**Supplementary Table 3**).

Plasmid construction

The piggyBac transposon vector was constructed by inserting a random 10-bp DNA barcode sequence between the CMV promoter and the multiple cloning site of the piggyBac dual promoter vector (PB513B-1; System Biosciences). The GFP sequence in the piggyBac vector was deleted by a site-directed mutagenesis technique for the direct repeat–green fluorescent protein (DR-GFP) assay. The full-length wild-type cDNA of human *BRCA2* was subcloned from pcDNA3 236HSC WT (Addgene plasmid # 16246) into the piggyBac vector. Plasmids encoding *BRCA2* mutations were developed by a site-directed mutagenesis technique with mutation-specific primers designed using an online tool (QuikChange Primer Design; https://www.agilent.com/store/primerDesignProgram.jsp). The primers used for mutagenesis are listed in **Supplementary Table 1**. For western blot analysis, the N-terminal FLAG-tag (DYKDDDDK) was inserted in the piggyBac-*BRCA2* wild-type by site-directed mutagenesis, and then the 19 *BRCA2* variants were also subcloned into the FLAG-*BRCA2*-piggyBac vector. Plasmids were fragmented by an E220 Focused-ultrasonicator (Covaris, Inc.) to an average size of 300 bp, ligated with adaptors, and sequenced using MiSeq (Illumina). All the entire *BRCA2* cDNA sequences, including the ones with point mutations generated by site-directed mutagenesis, the unique barcode sequences, and the piggyBac vector backbone sequences of every plasmid were confirmed by our original pipeline. Next-generation sequencing (NGS) data were aligned by Bowtie2 against the reference sequence composed of wild-type *BRCA2* cDNA and the piggyBac vector and analyzed with IGV 2.4.10 software (Broad Institute). Plasmids harboring undesirable mutations were discarded. Therefore, all plasmids in this study had only target mutations and unique barcode sequences without additional mutations. The

hyperactive piggyBac transposase expression vector (pCMV-hyPBase) was provided by Trust Sanger Institute. pCBASceI and pHPRT-DRGFP were a gift from Maria Jasin (Addgene plasmid # 26477 and # 26476, respectively). The puromycin antibiotic selection marker for pHPRT-DRGFP was replaced with the Sh_ble zeocin resistance gene. To construct the piggyBac-DRGFP vector, the ends of the XhoI-SacI-digested DRGFP reporter fragment and the SpeI-ApaI-digested piggyBac transposon vector were blunted and ligated to each other.

Stable transfection with the piggyBac transposon

The lipid-based method was used to transfect DLD1 cells with the plasmid constructs. Transfections were performed with exactly one unique plasmid per well. This is essential to guarantee that cells were transduced with only one variant. The same clone of each plasmid harboring a variant was used for all independent batches. Twenty-four hours before transfection, Cells were seeded in 96-well plates at a density of 1×10^4 cells/well. Prior to transfection, the culture medium was replaced with 100 µl of fresh medium. The transfection mixture for individual wells comprised 100 ng of the piggyBac transposon vector, 50 ng of pCMV-hyPBase, 0.5 µl of Lipofectamine Stem Transfection Reagent (Thermo Fisher Scientific), and 50 µl of Opti-MEM I Reduced Serum Medium (Thermo Fisher Scientific). The transfection mixture was added to each well after 20 minutes of incubation at room temperature. Two days after transfection, cells were cultured with antibiotics (3 µg/ml puromycin for piggyBac-*BRCA2* and 100 µg/ml zeocin for piggyBac-DRGFP) for 7 days. Puromycin-resistant polyclonal cell populations harboring piggyBac-*BRCA2* were used directly for assays, while a genetically homogeneous cloned cell line harboring piggyBac-DRGFP was generated from a single cell.

Digital droplet PCR

Digital droplet PCR was performed using a QX100 Droplet Digital PCR system (Bio-Rad Laboratories) with *BRCA2* cDNA primers and probe and *BRCA2* intron primers and probe, both at a final concentration of 900 nM primers and 250 nM probe. The sequences are shown in **Supplementary Table 2**. Aqueous droplets with a 20-µl volume containing final concentrations of 1× Droplet Digital PCR Supermix for probes (Bio-Rad Laboratories) and 100 ng of the genomic DNA template were generated. PCR amplification was conducted using a T100 Thermal Cycler under the manufacturer-recommended conditions. After amplification, the digital PCR data were collected and analyzed using a Bio-Rad QX100 droplet reader and QuantaSoft v1.3.2.0 software. Crosshair gating was used to automatically split the data into four quadrants by the software's normal setting. Approximately 15,000 droplets were analyzed per well. The induced *BRCA2* cDNA copy number was calculated based on the observation that DLD1 cells are a pseudodiploid human cell line and the copy number of the endogenous *BRCA2* gene is considered to be 2. The 95% confidence intervals were estimated by a Poisson distribution model.

Real-time quantitative RT-PCR

Freshly recovered DLD1 *BRCA2* (-/-) cells with 20 variants of *BRCA2*, untransduced DLD1 *BRCA2* (-/-) cells, and DLD1 parental cells were pelleted. Total RNA was extracted from these pellets with RNA-Bee reagent following the manufacturer's instructions. DNase I digestion was performed to minimize DNA contamination, and RNA-Bee purification was repeated to inactivate DNase I. The reverse transcription reaction used 1 µg of total RNA from each sample with SuperScript IV VILO reverse transcriptase. The resulting cDNAs were subsequently used for two-step quantitative RT-PCR with Power SYBR Green PCR Master Mix (Thermo Fisher Scientific) on a 7500HT Fast Real-Time PCR System (Thermo Fisher Scientific). The primers specific for *BRCA2* exon 11 and ACTB are shown in **Supplementary Table 2**. The amplified

522 *BRCA2* sequence was present in parental cells but absent from *BRCA2* (-/-) cells; therefore,
523 mRNA detected in the *BRCA2* (-/-) cells was derived from induced cDNA. The relative *BRCA2*
524 expression levels were normalized to the expression levels of the housekeeping gene ACTB.
525 Experiments were performed in technical triplicate, including DLD1 parental cells and DLD1
526 *BRCA2* (-/-) cells transfected with empty vector control. Then, the expression data were further
527 normalized to the corresponding expression levels in DLD1 parental cells. The results are
528 shown as the averages of biological triplicate experiments.
529
530 **Western blotting**
531 Whole-cell lysates from DLD1 *BRCA2* (−/−) cells harboring 20 variants of the N-terminal
532 FLAG-tagged *BRCA2* cDNA and the empty piggyBac vector were prepared with sodium
533 dodecyl sulfate (SDS) sample buffer containing 2-mercaptoethanol (Nacalai tesque). Total
534 protein (70 μg/sample for FLAG-BRCA2 and 10 μg/sample for beta-Actin) obtained from
535 fresh cell lysates was subjected to SDS polyacrylamide gel electrophoresis on 6% (FLAG-
536 BRCA2) or 10% (beta-Actin) polyacrylamide gels. Precision Plus Protein All Blue Standards
537 (Thermo Fisher Scientific) were run as molecular weight markers along with samples.
538 Electrophoresis was performed at 150 V, and the gels were transferred to an Immobilon-P
539 transfer membrane (Millipore) overnight at 50 V. The membranes were blocked with 5%
540 nonfat dry milk in 0.05% Tween 20 containing Tris-buffered saline (TBST) at room
541 temperature for 1 hour, and then incubated with a 500-fold diluted primary anti-FLAG antibody
542 (F3165; Sigma) or a 2,000-fold diluted primary anti-beta-Actin antibody (#4970; Cell
543 Signaling Technology) in TBST at 4 °C for 16 hours. Subsequently, the membranes were
544 incubated with 10,000-fold diluted peroxidase-linked secondary antibody (NA931V for
545 FLAG-BRCA2 and NA934V for beta-Actin; GE Healthcare) at room temperature for 4 hours.

546 The target proteins were visualized by an enhanced chemiluminescence reagent, SuperSignal
547 West Femto (GE Healthcare).
548
549 **Cell viability assay**
550 DLD1 *BRCA2* (-/-) cells harboring 20 variants of *BRCA2* cDNA and empty vector,
551 untransduced DLD1 *BRCA2* (-/-) cells, and DLD1 parental cells were seeded in 96-well plates
552 at a density of 2.0×10^3 cells/well with 100 µl of medium/well, and each drug was added at
553 various concentrations: olaparib (50 nM–5 µM, Selleckchem), niraparib (10 nM–1 µM,
554 Selleckchem), rucaparib (50 nM–5 µM, LC Laboratories), and CBDCA (50 nM–5 µM,
555 Selleckchem). DMSO (Nacalai Tesque) was added to a final concentration of 0.01%
556 (volume/volume) to wells without drugs. Ten microliters of PrestoBlue cell viability reagent
557 (Thermo Fisher Scientific) was added to each well 144 hours after exposure to these drugs, and
558 fluorescence intensity was measured with a 2030 ARVO X3 microplate reader (PerkinElmer)
559 (excitation; 530 nm, emission; 590 nm) [51]. Wells without cells were assessed as the negative
560 controls, and survival data were graphically analyzed as a sigmoid curve by GraphPad Prism
561 software v8.02 for Mac (GraphPad Software, Inc).
562
563 **MANO-B method**
564 The original MANO method is a high-throughput functional assay previously reported by our
565 laboratory [28, 29]. Individually established DLD1 *BRCA2* (-/-) cells with stable *BRCA2*
566 expression were mixed equally and cultured competitively. Cells were seeded at a density of
567 1×10^4 cells/cm^2 on day 0. During the 12-day treatment with PARP inhibitors (olaparib,
568 niraparib, and rucaparib), CBDCA or DMSO, cells were passaged once, on day 6. On day 12,
569 genomic DNA was isolated from cell lysates with a QIAamp DNA Mini Kit (Qiagen). Barcode
570 sequences were amplified by PCR using 300 ng of genomic DNA (100,000-genome and

500,000-barcode equivalent, by the weight of 6 pg of genomic DNA per cell and an average copy number of integrated cDNA of 10 copies per cell) and primers containing index sequences for deep sequencing (**Supplementary Table 3**). The quantity and quality of the obtained libraries were evaluated using a Qubit 2.0 fluorometer (Thermo Fisher Scientific) and an Agilent 2200 TapeStation system. Libraries were deep sequenced on the MiSeq sequencer using a Reagent Kit V2 (300 cycles), and the number of each barcode in each variant was counted. Variants were tested in at least two independent batches in technical triplicate. In every experiment, the wild-type and the D2723H variant were included as the benign and pathogenic controls, respectively. In each independent batch, cells were transfected with the plasmids independently on each occasion because transfection efficiency has a large impact on genomic integration and gene expression, which could affect the result of the MANO-B method. A single clone of each plasmid harboring each desired mutation was used for all batches. To remove the concern that a plasmid harbored an additional undesired mutation, we had the entire sequence of each plasmid checked (including the vector backbone) rather than using two independent clones. In batch #1, 107 variants and one empty vector were evaluated. In batches #2 and #3, 244 variants and one empty vector were evaluated. The specific variants are shown in **Supplementary Table 1**.

Model assessment

The expectation–maximization algorithm was used to determine the appropriate Gaussian mixture model for the data from the MANO-B method with the "mclust" package for the R language [52]. The model with the highest Bayesian information criterion (BIC) was adopted.

Bayesian hierarchical model analysis

595 Barcode counts obtained by the MANO-B method were analyzed using a Bayesian hierarchical
596 mixed model for variant function, which was a modified version of the VarCall model, as
597 previously described [32]. The fold change in the barcode count for variants treated with each
598 drug relative to that for variants treated with the DMSO control was calculated. These values
599 were log normalized to the standard values of $\log_{10}(1.0)$ for wild-type and $\log_{10}(0.003)$ for
600 D2723H as anchors. The remaining 242 variants were unlabeled for pathogenicity in this
601 analysis. We hypothesized that the functional variant-specific effect η_v followed a 2-
602 component Gaussian mixture distribution. The prior probability of the variants' pathogenicity
603 inside key domains was noninformative or based on the Align-GVGD classification obtained
604 from the HCI Breast Cancer Genes Prior probabilities website
605 (http://priors.hci.utah.edu/PRIORS) and the BRCA Exchange website
606 (https://brcaexchange.org). The Align-GVGD classification (C0–C65) was defined on the basis
607 of evolutionary conservation of the protein sequence from *Strongylocentrotus purpuratus* to
608 *Homo sapiens*. Key functional domains annotated by the HCI Breast Cancer Genes Prior
609 probabilities website were as follows: the PALB2 interaction domain (amino acid residues 10–
610 40), DNA-binding domain (2481–3186), and TR2 RAD51-binding domain (3269–3305). The
611 prior probability outside key domains was estimated at 0.02 by the HCI and the BRCA
612 Exchange, independent of the Align-GVGD classification.
613 The equations and parameters of the model are as follows.

$$\prod_{v} \Pr(f_v|D_v, X_v, \theta) = \prod_{\{v:D_v=D\}} \Pr(f_v|D_v = D, X_v, \theta) \prod_{\{v:D_v=N\}} \Pr(f_v|D_v = N, X_v, \theta) \prod_{\{v:D_v \text{ is Unknown}\}} \left[\pi_{a(v,D)} \Pr(f_v|D_v = D, X_v, \theta) + \pi_{a(v,N)} \Pr(f_v|D_v = N, X_v, \theta)\right]$$

614

615 The terms in the equation above are addressed below.

$\Pr(\text{Data} \mid \text{Parameters})$: likelihood of Data observation with Parameters

v: variant index

f_v: measurements of functional experiments for v

D_v: $\begin{cases} D_v = D\ (v = \text{deleterious}) \\ D_v = N\ (v = \text{neutral}) \end{cases}$

X_v: batch and experimental indices for each measurement

θ: model parameters

$\pi_{a(v,D/N)}$: prior probability that v is deleterious/neutral

The true distributions of the data and parameters were estimated by the formula below.

$$\Pr(f_{v^*}|D_{v^*}, X_{v^*}, \theta) = \prod_{\{(v,b,e):v=v^*\}} \Pr(f_v|D_v, b, e, \theta)$$

We established some constraints and weakly informative prior distributions as described below.

b: batch index

e: experimental index

β_b: batch-specific random intercept effect

τ_b: batch-specific random slope effect

κ_1: center of the distribution β_b

κ_2: center of the distribution τ_b

λ_1: standard deviation of distribution β_b

λ_2: standard deviation of distribution τ_b

η_v: variant-specific random effect

η_{del}: center of the deleterious variants' η_v distribution

η_{neu}: center of the neutral variants' η_v distribution

σ_1: standard deviation of the deleterious variants' η_v distribution

σ_2: standard deviation of the neutral variants' η_v distribution

ψ: residual error

Key domains: $\begin{cases} \text{PALB2 interaction domain (amino acid residues 10-40)} \\ \text{DNA-binding domain (2481-3186)} \\ \text{TR2 RAD51-binding domain (3269-3305)} \end{cases}$

The terms in the formulae above are addressed below.

$$
\begin{aligned}
f_v &\sim \text{Normal}(\beta_b + \tau_b\eta_v, \tau_b\psi) \\
\beta_b &\sim \text{Normal}(\kappa_1, \lambda_1) \\
\tau_b &\sim \text{Normal}(\kappa_2, \lambda_2), \tau_b > 0 \\
\eta_v \,|D_v = D &\sim \text{Normal}(\eta_{\text{del}}, \sigma_1) \quad \{v : D_v = D \wedge v \neq \text{D2723H}\} \\
\eta_v \,|D_v = N &\sim \text{Normal}(\eta_{\text{neu}}, \sigma_2) \quad \{v : D_v = N \wedge v \neq \text{WT}\} \\
\eta_{\text{WT}} &= \log_{10}(1.0) \\
\eta_{\text{D2723H}} &= \log_{10}(0.003) \\
\eta_{\text{nue}} &= \text{estimated value by mclust package with the training data set} \\
\eta_{\text{del}} &\sim \text{Normal}(\eta_{\text{D2723H}}, 5) \qquad \text{(training data set)} \\
\eta_{\text{del}} &= \text{as } \beta_b + \tau_b\eta_{\text{del}} \text{ value is the same in the training data set} \quad \text{(full analysis)} \\
\kappa_1 &\sim \text{Normal}(0, 5) \\
\kappa_2 &\sim \text{Normal}(1, 5) \\
\psi, \sigma_1, \sigma_2, \lambda_1, \lambda_2 &\overset{iid}{\sim} \text{HalfNormal}(0, 5) \\
D_v &\sim \text{Bernoulli}\left(\pi_{a(v)}\right) \quad \{v : D_v \text{ is unknown}\} \\
\pi_{a(v)} &\sim \text{Beta}(1, 1) \quad \text{(noninformative prior probability)} \\
&\textit{or} \\
&\text{Align-GVGD-based prior probability} \\
&\begin{cases}
\pi_{a(v)} \sim \text{Beta}(15.00, 3.48) & \{v \in C65, \text{inside key domains}\} \\
\pi_{a(v)} \sim \text{Beta}(5.38, 2.57) & \{v \in C35, C45, C55, \text{inside key domains}\} \\
\pi_{a(v)} \sim \text{Beta}(3.76, 9.00) & \{v \in C15, C25, \text{inside key domains}\} \\
\pi_{a(v)} \sim \text{Beta}(1.43, 73.1) & \{v \in C0, \text{inside key domains}\} \\
\pi_{a(v)} \sim \text{Beta}(1.64, 120.44) & \{v \in \text{outside key domains}\} \\
\pi_{a(v)} \sim \text{Beta}(387, 1.07) & \{v \in \text{nonsense variants}\}
\end{cases}
\end{aligned}
$$

In the above model, the fixed value η_{neu} was arbitrarily estimated by the mclust package with the training set data. In the full analysis, η_{del} was fixed such that the value of $\beta_b + \tau_b\eta_{del}$ was the same as that in the training data set. Normal(μ, σ) denotes the normal distribution with mean μ and standard deviation σ; X ~ HalfNormal(0,5) indicates that X is normally distributed with a mean of 0 and a standard deviation of 5 but is constrained to be nonnegative. The parameters of the Beta distributions were determined such that their 2.5th and 97.5th percentiles fit the upper and lower 95% confidence limits of the Align-GVGD prediction for each class [36]. The ranges of the normal distributions for the parameters' prior distributions were set to adequately exceed the range of values expected for those parameters. The half-Cauchy distribution was thought to be superior to other weakly informative prior distributions for variance component parameters [53], but the Cauchy distribution might be too broad for our

model because very large values could not occur. Thus, a half-normal distribution was adopted for variance component parameters.

Inference

A Hamiltonian Monte Carlo algorithm was used to perform inference for the above model. The source code written in the R [54] and Stan [55] languages is available (https://github.com/MANO-B/Bayes). We ran 4 chains of samplers, including 1,500 warmup iterations followed by 3,500 sampling iterations, and every sampling iteration was adopted. The chain passed the Gelman and Rubin's convergence diagnostic, as the R-hat values were less than 1.1 for all parameters [56], indicating convergence across the 4 chains initiated from disparate starting values, and the simulated posterior values were drawn from the true posterior for each parameter. The marginal posterior means and credible intervals (CIs) of the model parameters, the BF and posterior probabilities of the pathogenic categories for each variant were calculated.

Posterior Predictive Checks

To validate the model's correctness, the fitness of estimation to observed data from 244 variants was statistically checked. QQ plots of the posterior expected standardized residuals of the log-normalized relative viability were generated using the "car" package for the R language [57]. Simultaneous 95% CIs are plotted as green dotted lines. To confirm the match between the posterior predictive distribution of f_v and the observed data, we randomly drew 512 parameter samples from the trace. Then, we generated 612 random values per variant from a normal distribution specified by the values of η, β, τ, and ψ for each sample, accounting for the batch-to-batch ratio. A posterior predictive density curve of the log-normalized relative viability data was computed from 612 generated data sets containing 512 samples each.

659 **HDR assay**

660 Twenty-four *BRCA2* variants encoded individually in the piggyBac vector and an I-SceI
661 expression vector, pCBASce, were cotransfected into DLD1 *BRCA2* (-/-) cells harboring the
662 DRGFP sequence, according to a previous report [18, 58]. Twenty-four hours before transfection,
663 cells were seeded in 12-well plates at a density of 1×10^5 cells/well. Prior to transfection, the
664 culture medium was replaced with 1 ml of fresh medium. The transfection mixture for
665 individual wells was composed of 1 μg of the *BRCA2* expression vector, 500 ng of the
666 pCBASce vector, 4 μl of Lipofectamine Stem Transfection Reagent, and 250 μl of Opti-MEM
667 I Reduced Serum Medium. The transfection mixture was added to each well after 20 minutes
668 of incubation at room temperature. I-SceI-induced DNA double-strand breaks in the DRGFP
669 sequence were subsequently repaired by homologous recombination, and intact GFP-
670 expressing cells were produced. Four days after transfection, cells were collected, and GFP-
671 positive cells were counted by fluorescence-activated cell sorting (FACS) using a BD
672 FACSCanto II (BD Bioscience). Every cell count was normalized and rescaled relative to a 1:5
673 ratio of D2723H:wild-type. All variants were analyzed in biological duplicate and technical
674 triplicate.

675

676 **Statistical Analysis**

677 All analyses were performed via the R language. The data are the means ± SD, means ± , or
678 means ± 95% CI, as stated in the figure legends. Differences between variants were determined
679 by the Kruskal-Wallis rank sum test. $p < 0.05$ was considered statistically significant.

680 **Acknowledgments**

681 We thank A. Maruyama for technical assistance and N. Tanabe and T. Yoshida for collecting

682 data.

683

684 **Authors' Contributions**

685 **Conception and design**: S. Kohsaka and H. Mano

686 **Development of methodology:** M. Ikegami, S. Kohsaka, and H. Mano

687 **Acquisition of data (provided animals, acquired and managed patients, provided facilities,**

688 **etc.):** M. Ikegami

689 **Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational**

690 **analysis):** M. Ikegami and T. Ueno

691 **Writing, review, and/or revision of the manuscript:** M. Ikegami, S. Kohsaka, and H. Mano

692 **Administrative, technical, or material support (i.e., reporting or organizing data,**

693 **constructing databases):** Y. Momozawa, K. Tamura, A. Shimomura, and N. Hosoya

694 **Study supervision:** H. Kobayashi and S. Tanaka

Figure Legends

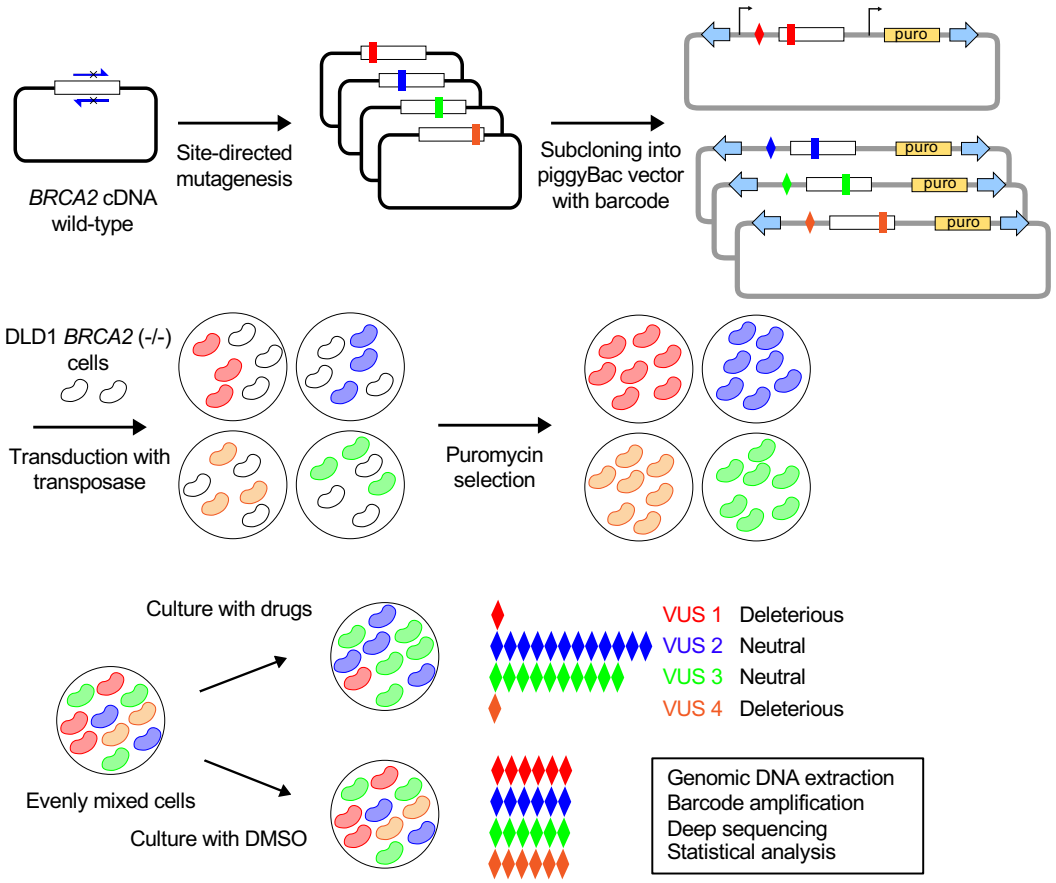


Figure 1. Schematic representation of the MANO-B method. *BRCA2* variants of interest were generated by site-directed mutagenesis and subcloned into the piggyBac transposon vector with a unique barcode sequence, followed by sequence verification. DLD1 *BRCA2* (−/−) cells were individually cotransfected with the *BRCA2* constructs and transposase expression vector. After puromycin selection of stable transfectants, equal numbers of DLD1 cells were mixed and cultured with PARP inhibitors, CBDCA, or DMSO for 12 d. The barcode sequences were amplified by PCR with genomic DNA extracted from the cell mixture and deep sequenced to quantify their relative abundances. The ratio of each barcode abundance was normalized to that of the DMSO-treated control. The relative viability data were analyzed by Bayesian inference, and the pathogenicity of each variant was then classified.

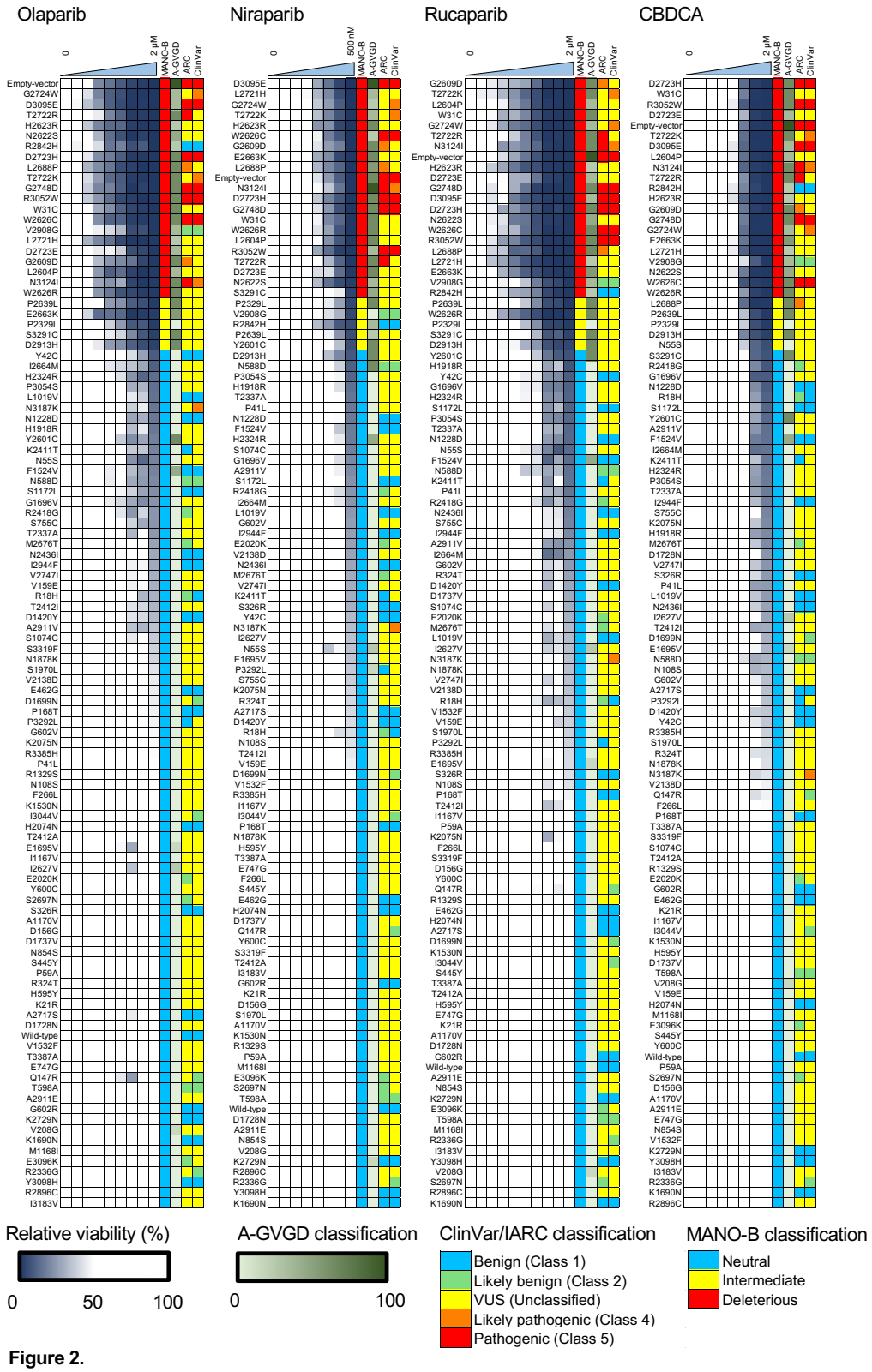


Figure 2.

Figure 2. Comparative results of the MANO-B method with the IARC, ClinVar, and Align-GVGD classifications for 107 variants. DLD1 *BRCA2* (−/−) cells transfected separately with

the 107 *BRCA2* variants or the empty vector were treated with PARP inhibitors (olaparib, niraparib, and rucaparib), CBDCA, or DMSO at the indicated concentrations. The relative viability of the cells treated with each drug was calculated based on the viability of vehicle-treated cells, then normalized to that of cells expressing wild-type *BRCA2*. All variants were ordered by relative viability at the optimal concentration. The classification of the MANO-B method was based on the relative viability at the optimal concentration of each drug and defined as neutral (relative viability of 1–82), intermediate (83–87), and deleterious (88–108). The IARC, ClinVar, and Align-GVGD classifications are color-coded in each column. The ClinVar classifications of five variants (V159E, M1168I, K1530N, G1696V, and A2911E) without assertion criteria were considered as VUSs.

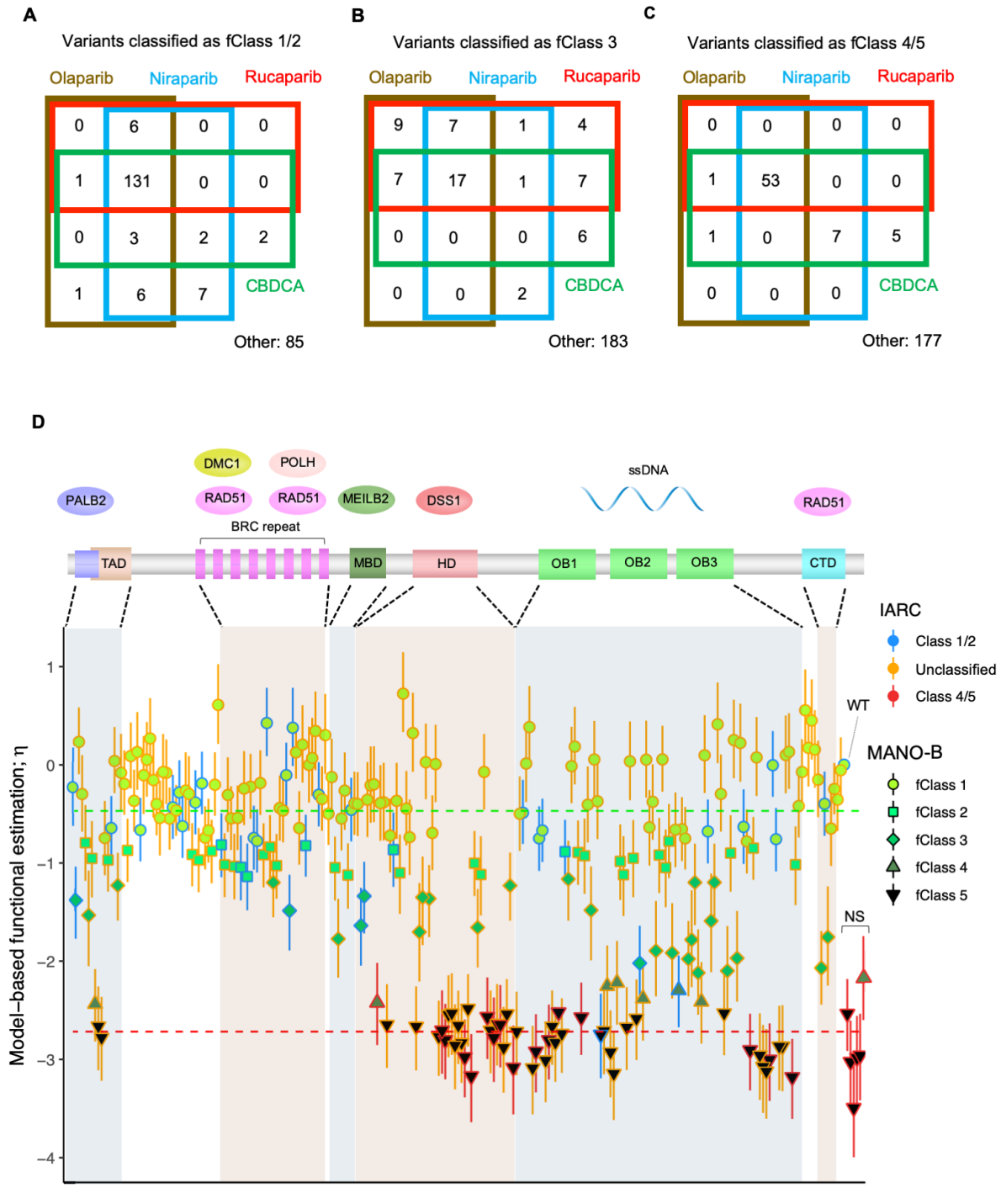


Figure 3. Evaluation of 244 *BRCA2* variants by applying Bayesian inference to the MANO-B method. **A-C,** Functional classification by the MANO-B method with four drugs. A total of 244 *BRCA2* variants were classified by Bayesian inference. The number of variants classified for each drug as fClass 1/2 (**A**), fClass3 (**B**), and fClass4/5 (**C**) is shown. No variant was

726 classified as fClass1/2 with one drug and fClass 4/5 with another drug. **D,** Bar plots of
727 functional variant effects and functional diagnosis of 244 *BRCA2* variants mapped against the
728 *BRCA2* full-length sequence and domains. The key domains of *BRCA2* consist of a PALB2
729 interaction domain encompassing amino acids (a.a.) 10–40, a transactivation domain (TAD)
730 encompassing a.a. 18–105, a RAD51-binding domain including eight BRC repeats
731 encompassing a.a. 1008–2082, an MEILB2-binding domain (MBD) encompassing a.a. 2117–
732 2339, a DNA-binding domain encompassing a.a. 2402–3186 and containing a helical domain
733 (HD) encompassing a.a. 2402–2669, oligonucleotide/oligosaccharide-binding domains (OBs)
734 (OB) encompassing a.a. 2670–2803, 2809–3048, and 3056–3102, and a C-terminal RAD51-
735 binding domain (CTD) encompassing a.a. 3270–3305. The function of each variant was
736 estimated as a variant-specific effect value, η_v, based on a two-component model. Data of
737 niraparib are shown as a representative case. The upper and lower components correspond to
738 neutral function and loss-of-function, respectively. The dotted lines indicate the median value
739 of each component. The classification of each variant based on the Bayes factor (BF) is
740 indicated by the color and shape of the lines and plots, as shown in the legend. Variant functions
741 were classified into five classes by the MANO-B method: fClass 1 (neutral) ($BF \leq 0.003$),
742 fClass 2 (likely neutral) ($0.003 < BF \leq 0.053$), fClass 3 (intermediate) ($0.053 < BF < 18.7$),
743 fClass 4 (likely deleterious) ($18.7 \leq BF < 350$), and fClass 5 (deleterious) ($350 \leq BF$). Open
744 error bars, 95% CIs. WT, wild-type; NS, nonsense variants.

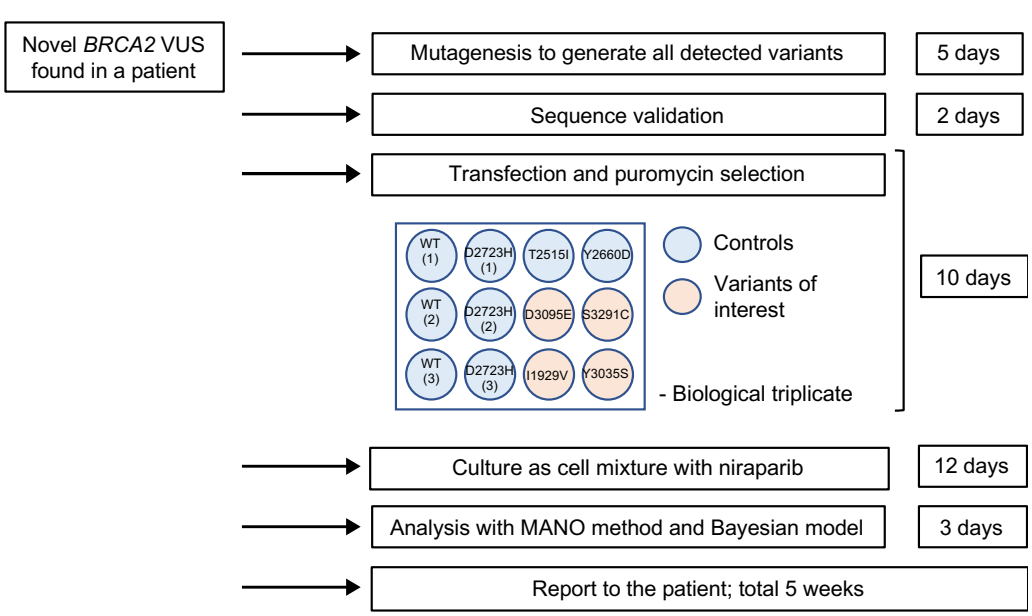


Figure 4. The Accurate BRCA Companion Diagnostic (ABCD) test for the clinical functional annotation of novel *BRCA2* VUSs. Schematic representation of the ABCD test. *BRCA2* mutations were investigated by genomic DNA sequencing. The effect of mutations on mRNA splicing was evaluated by RNA sequencing to detect the aberrant *BRCA2* transcripts. All detected cDNAs, including four control variants for neutral (wild-type and T2515I) and deleterious (Y2660D and D2723H) mutations and four variants of interest, were generated for assessment by the small-scale MANO-B method in a 12-well plate. Technical and biological triplicate experiments using niraparib at 0.5 μM are recommended. Data analysis was performed with data from previous experiments, enabling batch effect adjustment. The experimental steps and required times are indicated.

Variant	IARC classification	ABCD test	η			BF	fClass
			Mean	95% CI			
				LL	UL		
I1929V	Class 1	Previous batches	0.37	-0.02	0.79	1.05×10^{-5}	fClass 1
		New batch-1	0.19	-0.33	0.73	1.02×10^{-5}	fClass 1
		New batch-2	-0.23	-0.73	0.28	1.63×10^{-4}	fClass 1
		New batch-3	0.08	-0.43	0.60	3.71×10^{-5}	fClass 1
		3 batches combined	0.11	-0.23	0.44	1.40×10^{-5}	fClass 1
Y3035S	Class 2	Previous batches	-0.63	-1.01	-0.25	1.11×10^{-3}	fClass 1
		New batch-1	-0.55	-1.05	-0.04	1.12×10^{-3}	fClass 1
		New batch-2	-0.53	-1.04	-0.02	9.57×10^{-4}	fClass 1
		New batch-3	-0.36	-0.87	0.16	3.01×10^{-4}	fClass 1
		3 batches combined	-0.47	-0.79	-0.15	2.79×10^{-4}	fClass 1
D3095E	Class 5	Previous batches	-3.00	-3.41	-2.62	1.33×10^{4}	fClass 5
		New batch-1	-2.31	-2.80	-1.78	4.45×10^{1}	fClass 4
		New batch-2	-2.63	-3.12	-2.15	5.75×10^{2}	fClass 5
		New batch-3	-2.69	-3.20	-2.19	9.28×10^{2}	fClass 5
		3 batches combined	-2.52	-2.85	-2.19	7.85×10^{2}	fClass 5
S3291C	Unclassified	Previous batches	-2.07	-2.43	-1.69	1.15×10^{1}	fClass 3
		New batch-1	-1.06	-1.70	-0.53	6.15×10^{-2}	fClass 3
		New batch-2	-2.06	-2.60	-1.34	7.05×10^{0}	fClass 3
		New batch-3	-1.35	-2.12	-0.74	3.17×10^{-1}	fClass 3
		3 batches combined	-1.46	-1.86	-1.11	3.39×10^{-1}	fClass 3

BF, Bayes factor; LL, lower limit; UL, upper limit; CI, credible interval.

765

766 **Table 1.** Results of an experimental ABCD test. The classification of the four variants was
767 reproducible and consistent with that observed for previous large-scale batches. BF, Bayes
768 factor; LL, lower limit; UL, upper limit; CI, credible interval.

References

1. Wooster R, Weber BL. Breast and ovarian cancer. *N Engl J Med* **348**, 2339-2347 (2003).

2. Finch AP, *et al.* Impact of oophorectomy on cancer incidence and mortality in women with a BRCA1 or BRCA2 mutation. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **32**, 1547-1553 (2014).

3. Heemskerk-Gerritsen BA, *et al.* Prophylactic mastectomy in BRCA1/2 mutation carriers and women at risk of hereditary breast cancer: long-term experiences at the Rotterdam Family Cancer Clinic. *Ann Surg Oncol* **14**, 3335-3344 (2007).

4. Paluch-Shimon S, *et al.* Prevention and screening in BRCA mutation carriers and other breast/ovarian hereditary cancer syndromes: ESMO Clinical Practice Guidelines for cancer prevention and screening. *Ann Oncol* **27**, v103-v110 (2016).

5. Farmer H, *et al.* Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**, 917-921 (2005).

6. Bryant HE, *et al.* Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* **434**, 913-917 (2005).

7. Richards S, *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405-424 (2015).

8. Plon SE, *et al.* Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum Mutat* **29**, 1282-1291 (2008).

9. Tavtigian SV, Greenblatt MS, Goldgar DE, Boffetta P, Group IUGVW. Assessing pathogenicity: overview of results from the IARC Unclassified Genetic Variants Working Group. *Hum Mutat* **29**, 1261-1264 (2008).

10. Easton DF*, et al.* A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes. *American journal of human genetics* **81**, 873-883 (2007).

11. Lindor NM*, et al.* A review of a multifactorial probability-based model for classification of BRCA1 and BRCA2 variants of uncertain significance (VUS). *Hum Mutat* **33**, 8-21 (2012).

12. Momozawa Y*, et al.* Germline pathogenic variants of 11 breast cancer genes in 7,051 Japanese patients and 11,241 controls. *Nat Commun* **9**, 4083 (2018).

13. Ernst C*, et al.* Performance of in silico prediction tools for the classification of rare BRCA1/2 missense variants in clinical diagnostics. *BMC Med Genomics* **11**, 35 (2018).

14. Toland AE, Andreassen PR. DNA repair-related functional assays for the classification of BRCA1 and BRCA2 variants: a critical review and needs assessment. *J Med Genet* **54**, 721-731 (2017).

15. Findlay GM*, et al.* Accurate classification of BRCA1 variants with saturation genome editing. *Nature* **562**, 217-222 (2018).

16. Starita LM*, et al.* Massively Parallel Functional Analysis of BRCA1 RING Domain Variants. *Genetics* **200**, 413-422 (2015).

17. Starita LM*, et al.* A Multiplex Homology-Directed DNA Repair Assay Reveals the Impact of More Than 1,000 BRCA1 Missense Substitution

Variants on Protein Function. *American journal of human genetics* **103**, 498-508 (2018).

18. Guidugli L*, et al.* Assessment of the Clinical Relevance of BRCA2 Missense Variants by Functional and Computational Approaches. *American journal of human genetics*, (2018).

19. Hart SN*, et al.* Comprehensive annotation of BRCA1 and BRCA2 missense variants by functionally validated sequence-based computational prediction models. *Genet Med*, (2018).

20. Guidugli L*, et al.* Functional assays for analysis of variants of uncertain significance in BRCA2. *Hum Mutat* **35**, 151-164 (2014).

21. Ding S, Wu X, Li G, Han M, Zhuang Y, Xu T. Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell* **122**, 473-483 (2005).

22. Hucl T, Rago C, Gallmeier E, Brody JR, Gorospe M, Kern SE. A syngeneic variance library for functional annotation of human variation: application to BRCA2. *Cancer research* **68**, 5023-5030 (2008).

23. Drean A*, et al.* Modeling Therapy Resistance in BRCA1/2-Mutant Cancers. *Mol Cancer Ther* **16**, 2022-2034 (2017).

24. Cline MS*, et al.* BRCA Challenge: BRCA Exchange as a global resource for variants in BRCA1 and BRCA2. *PLoS Genet* **14**, e1007752 (2018).

25. Parsons MT*, et al.* Large scale multifactorial likelihood quantitative analysis of BRCA1 and BRCA2 variants: An ENIGMA resource to support clinical variant classification. *Hum Mutat* **40**, 1557-1578 (2019).

868 26. Lee JS, *et al.* Reclassification of BRCA1 and BRCA2 variants of uncertain
869 significance: a multifactorial analysis of multicentre prospective cohort. *J*
870 *Med Genet* **55**, 794-802 (2018).
871
872 27. Yusa K, Zhou L, Li MA, Bradley A, Craig NL. A hyperactive piggyBac
873 transposase for mammalian applications. *Proc Natl Acad Sci U S A* **108**,
874 1531-1536 (2011).
875
876 28. Kohsaka S, *et al.* A method of high-throughput functional evaluation of
877 EGFR gene variants of unknown significance in cancer. *Science*
878 *translational medicine* **9**, (2017).
879
880 29. Nagano M, *et al.* High-throughput functional evaluation of variants of
881 unknown significance in ERBB2. *Clin Cancer Res* **24**, 5112-5122 (2018).
882
883 30. So MK, *et al.* Reinterpretation of BRCA1 and BRCA2 variants of
884 uncertain significance in patients with hereditary breast/ovarian cancer
885 using the ACMG/AMP 2015 guidelines. *Breast Cancer*, (2019).
886
887 31. Woods NT, *et al.* Functional assays provide a robust tool for the clinical
888 annotation of genetic variants of uncertain significance. *NPJ Genom Med*
889 **1**, (2016).
890
891 32. Iversen ES, Jr., Couch FJ, Goldgar DE, Tavtigian SV, Monteiro AN. A
892 computational method to classify variants of uncertain significance using
893 functional assay data with application to BRCA1. *Cancer Epidemiol*
894 *Biomarkers Prev* **20**, 1078-1088 (2011).
895
896 33. Tavtigian SV, *et al.* Modeling the ACMG/AMP variant classification
897 guidelines as a Bayesian classification framework. *Genet Med* **20**, 1054-
898 1060 (2018).
899
900 34. Xia B, *et al.* Control of BRCA2 cellular and clinical functions by a nuclear
901 partner, PALB2. *Mol Cell* **22**, 719-729 (2006).

35. Zhang J, Fujiwara Y, Yamamoto S, Shibuya H. A meiosis-specific BRCA2 binding protein recruits recombinases to DNA double-strand breaks to ensure homologous recombination. *Nat Commun* **10**, 722 (2019).

36. Tavtigian SV, Byrnes GB, Goldgar DE, Thomas A. Classification of rare missense substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications. *Hum Mutat* **29**, 1342-1354 (2008).

37. Esashi F*, et al.* CDK-dependent phosphorylation of BRCA2 as a regulatory mechanism for recombinational repair. *Nature* **434**, 598-604 (2005).

38. Farber-Katz S*, et al.* Quantitative analysis of BRCA1 and BRCA2 germline splicing variants using a novel RNA-massively parallel sequencing assay. *Front Oncol* **8**, 286 (2018).

39. Farrugia DJ*, et al.* Functional assays for classification of BRCA2 variants of uncertain significance. *Cancer research* **68**, 3523-3531 (2008).

40. Balia C, Galli A, Caligo MA. Effect of the overexpression of BRCA2 unclassified missense variants on spontaneous homologous recombination in human cells. *Breast Cancer Res Treat* **129**, 1001-1009 (2011).

41. Shimelis H*, et al.* BRCA2 hypomorphic missense variants confer moderate risks of breast cancer. *Cancer research* **77**, 2789-2799 (2017).

42. Bradbury AR*, et al.* Uptake and timing of bilateral prophylactic salpingo-oophorectomy among BRCA1 and BRCA2 mutation carriers. *Genet Med* **10**, 161-166 (2008).

43. Lim D, Ngeow J. Evaluation of the methods to identify patients who may benefit from PARP inhibitor use. *Endocr Relat Cancer* **23**, R267-285 (2016).

936 44. Davies H, *et al.* HRDetect is a predictor of BRCA1 and BRCA2 deficiency
937 based on mutational signatures. *Nature Medicine* **23**, 517-525 (2017).
938
939 45. Nones K, *et al.* Whole-genome sequencing reveals clinically relevant
940 insights into the aetiology of familial breast cancers. *Ann Oncol* **30**, 1071-
941 1079 (2019).
942
943 46. Daniels MJ, Wang Y, Lee M, Venkitaraman AR. Abnormal cytokinesis in
944 cells deficient in the breast cancer susceptibility protein BRCA2. *Science*
945 *(New York, NY)* **306**, 876-879 (2004).
946
947 47. Shivji MKK, Renaudin X, Williams CH, Venkitaraman AR. BRCA2
948 regulates transcription elongation by RNA polymerase II to prevent R-loop
949 accumulation. *Cell Rep* **22**, 1031-1039 (2018).
950
951 48. Leman R, *et al.* Novel diagnostic tool for prediction of variant
952 spliceogenicity derived from a set of 395 combined in silico/in vitro studies:
953 an international collaborative effort. *Nucleic Acids Res* **46**, 7913-7923
954 (2018).
955
956 49. Acedo A, Hernandez-Moro C, Curiel-Garcia A, Diez-Gomez B, Velasco EA.
957 Functional classification of BRCA2 DNA variants by splicing assays in a
958 large minigene with 9 exons. *Hum Mutat* **36**, 210-221 (2015).
959
960 50. Acedo A, *et al.* Comprehensive splicing functional analysis of DNA
961 variants of the BRCA2 gene by hybrid minigenes. *Breast cancer research :*
962 *BCR* **14**, R87 (2012).
963
964 51. Lall N, Henley-Smith CJ, De Canha MN, Oosthuizen CB, Berrington D.
965 Viability reagent, PrestoBlue, in comparison with other available
966 reagents, utilized in cytotoxicity and antimicrobial assays. *Int J Microbiol*
967 **2013**, 420601 (2013).
968

52. Scrucca L, Fop M, Murphy TB, Raftery AE. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R journal* **8**, 289-317 (2016).

53. Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Anal*, 515-534 (2006).

54. Ihaka R, Gentleman R. R: a language for data analysis and graphics. *J Comp Graph Stat* **5**, 299-314 (1996).

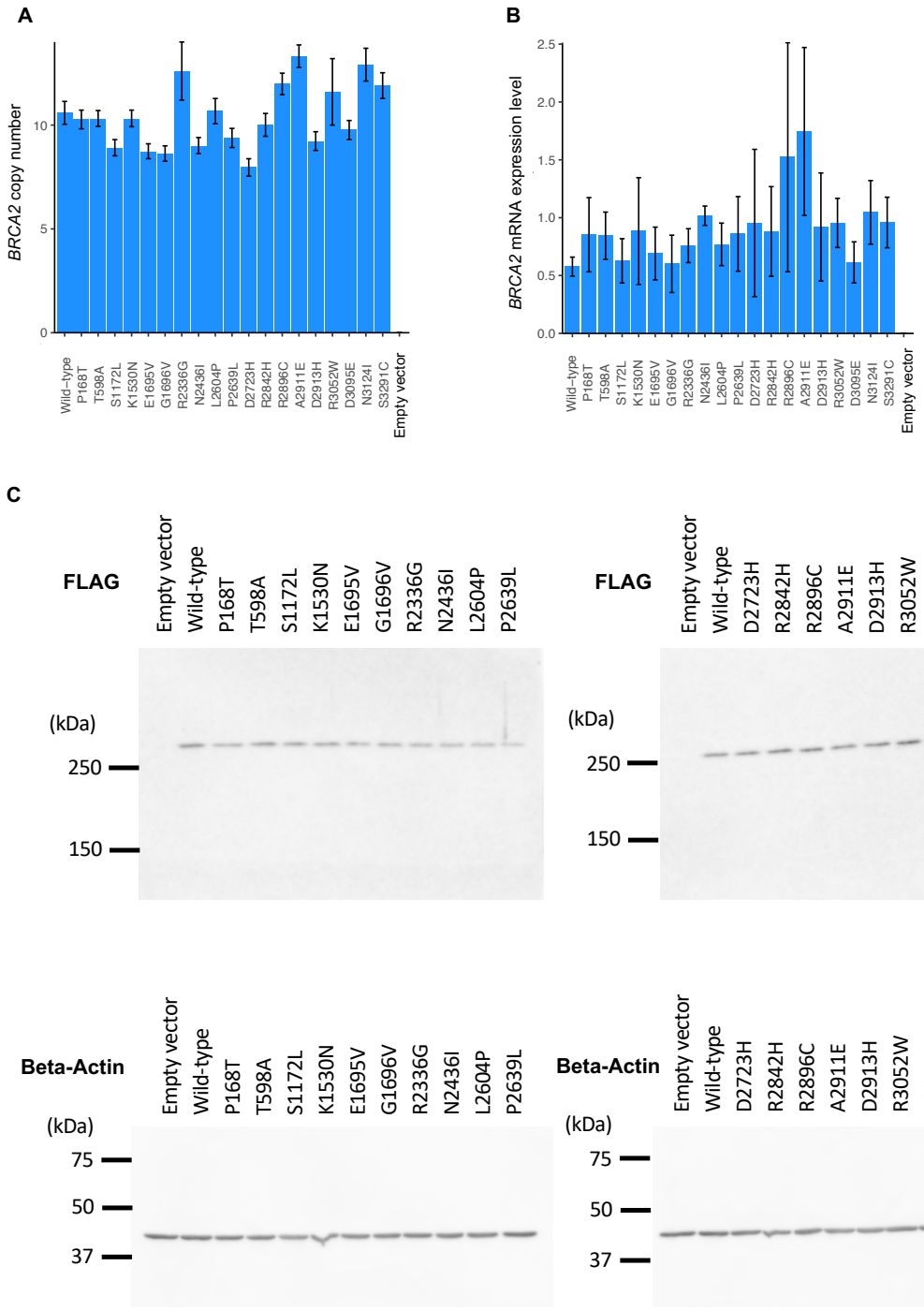
55. Gelman A, Lee D, Guo J. Stan: A probabilistic programming language for Bayesian inference and optimization. *J Educ Behav Stat* **40**, 530-543 (2015).

56. Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. *J Comput Graph Stat* **7**, 434-455 (1997).

57. Fox J, Weisberg S. *An R Companion to Applied Regression*, Second Edition edn. Sage Publications (2011).

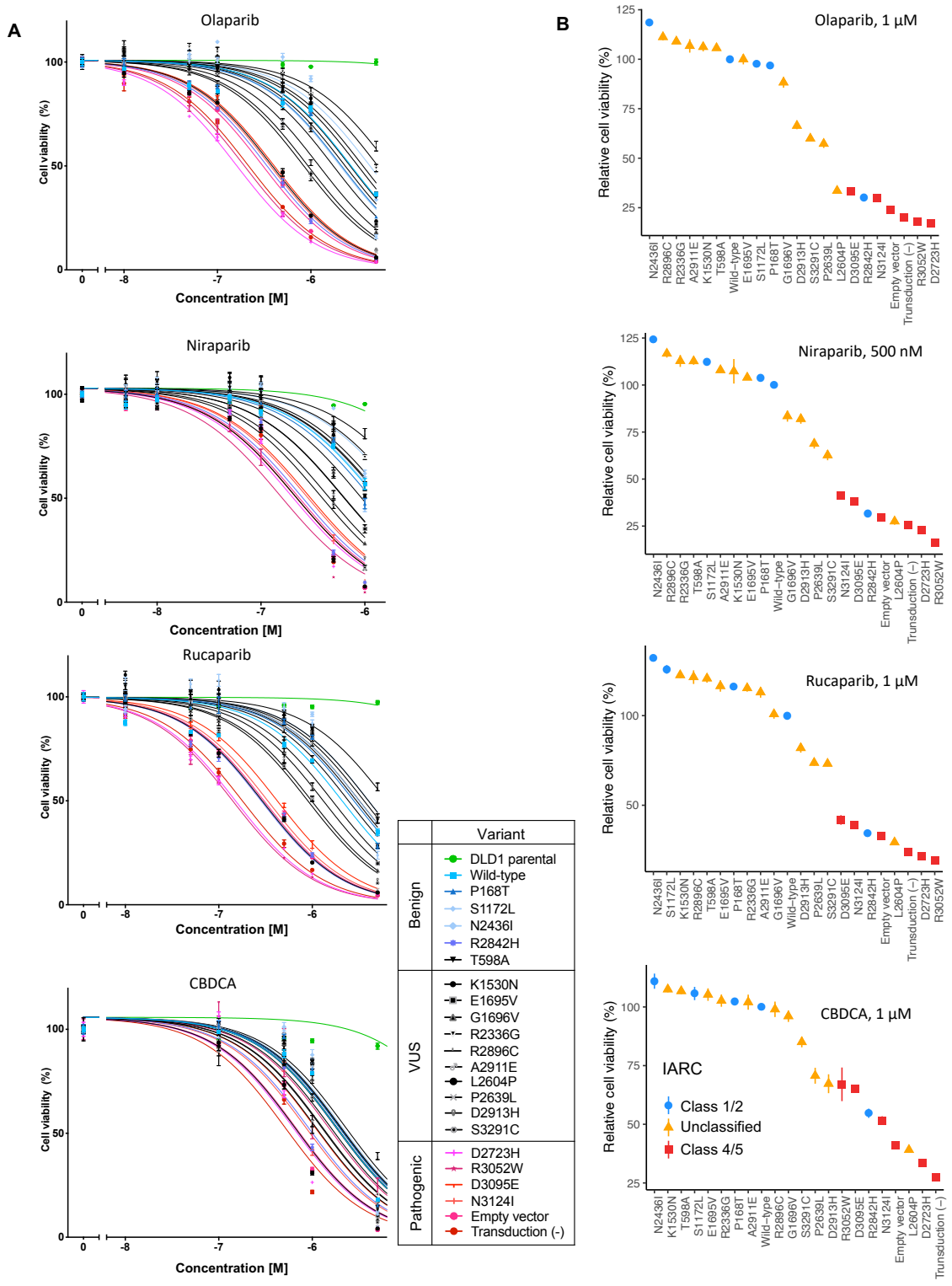
58. Guidugli L*, et al.* A classification model for BRCA2 DNA binding domain missense variants based on homology-directed repair activity. *Cancer research* **73**, 265-275 (2013).

994 **Supporting Information**



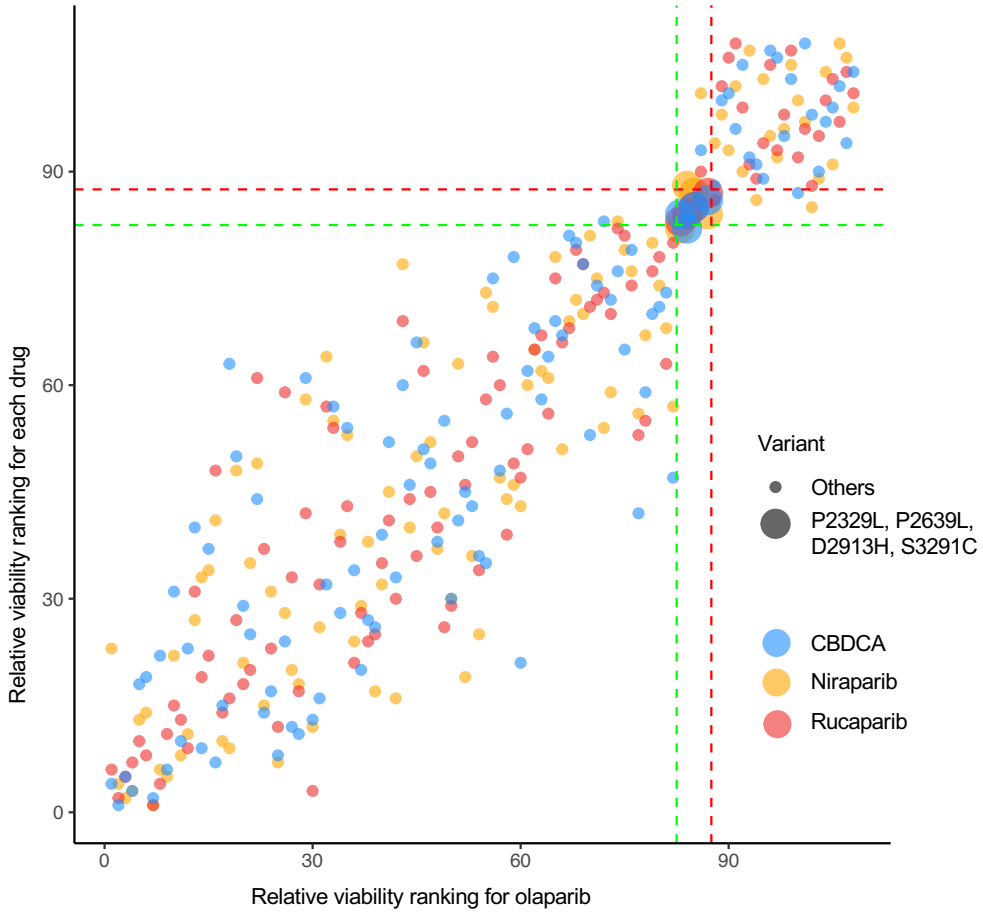
Supplementary Figure 1: Establishment of DLD1 *BRCA2* (-/-) cells expressing *BRCA2* variants. **A,** The induced *BRCA2* cDNA copy number of each variant was assessed by digital droplet PCR. Error bars, 95% confidence intervals. **B,** Real-time RT-PCR analysis of *BRCA2* mRNA expression in DLD1 parental cells and DLD1 *BRCA2* (-/-) cells. The *BRCA2* expression levels were normalized to those of ACTB and to those in DLD1 parental cells. Experiments

1024 were performed in technical triplicate. Data are shown as the average of biological triplicates.
1025 No significant difference between variants was observed (Kruskal-Wallis rank sum test, p =
1026 0.37). Error bars, SD (n = 3). **C,** Immunoblot analysis of DLD1 *BRCA2* (−/−) cells stably
1027 expressing the FLAG-tagged *BRCA2* variants. Cell lysates prepared from each cell were
1028 immunoblotted with antibodies against FLAG or beta-Actin. The protein expression levels of
1029 the 19 BRCA2 variants were generally equal to that of the wild-type BRCA2.



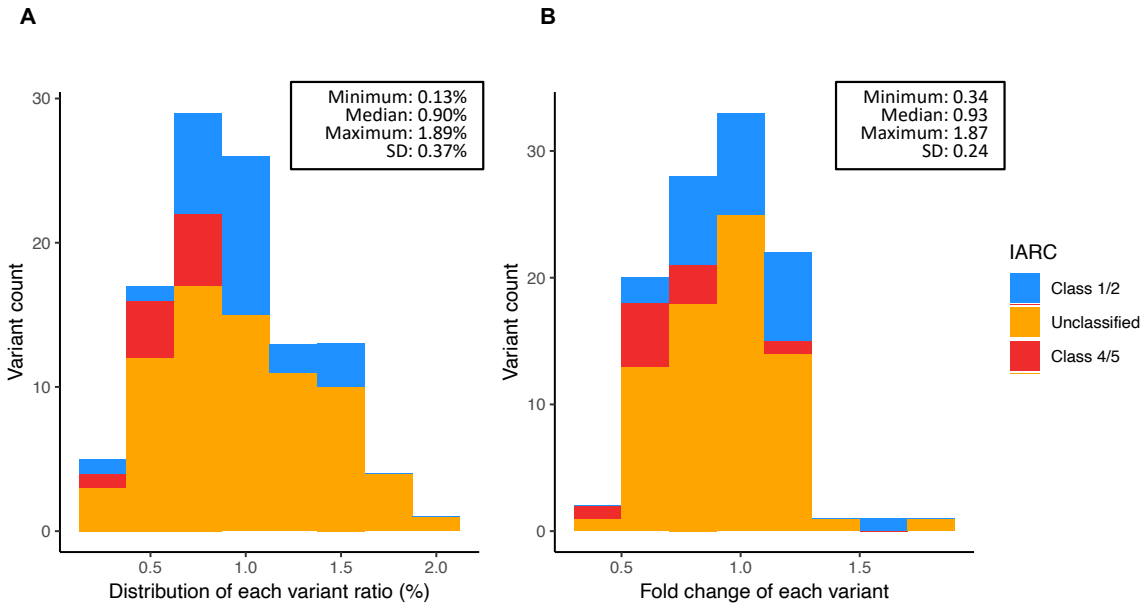
Supplementary Figure 2: Sensitivities of cells harboring *BRCA2* variants to PARP inhibitors and CBDCA. DLD1 parental cells, DLD1 *BRCA2* (-/-) cells harboring cDNA for one of 20 *BRCA2* variants or empty vector, or untransduced DLD1 *BRCA2* (-/-) cells were treated with the indicated concentrations of drugs for 144 hours. **A,** Cell viability was measured using

PrestoBlue cell viability reagent. The sensitivity to the drugs was in concordance with the IARC classification. IARC class 1/2 (benign), blue; unclassified (VUS), black; class 4/5 (pathogenic), red. Error bars, SEM (n = 5). **B,** Relative viability of cells harboring *BRCA2* variants relative to those harboring the wild-type *BRCA2* with the indicated concentrations of drugs. The IARC classification of each variant is indicated by the color and shape of the lines and plots. Error bars, SEM (n = 5).



1042

1043 **Supplementary Figure 3:** Relative viability of the 107 *BRCA2* variants at the highest drug
1044 concentration by the MANO-B method. The relative viability of each variant for niraparib,
1045 rucaparib, and CBDCA are plotted against that for olaparib. Neutral variants showing a high
1046 relative viability are located in the lower left area, whereas deleterious variants are located in
1047 the upper right area. Three components are shown: neutral (relative viability of 1–82),
1048 intermediate (83–87), and deleterious (88–108). P2329L, P2639L, D2913H, and S3291C,
1049 thought to be intermediate variants, are represented by larger circles.



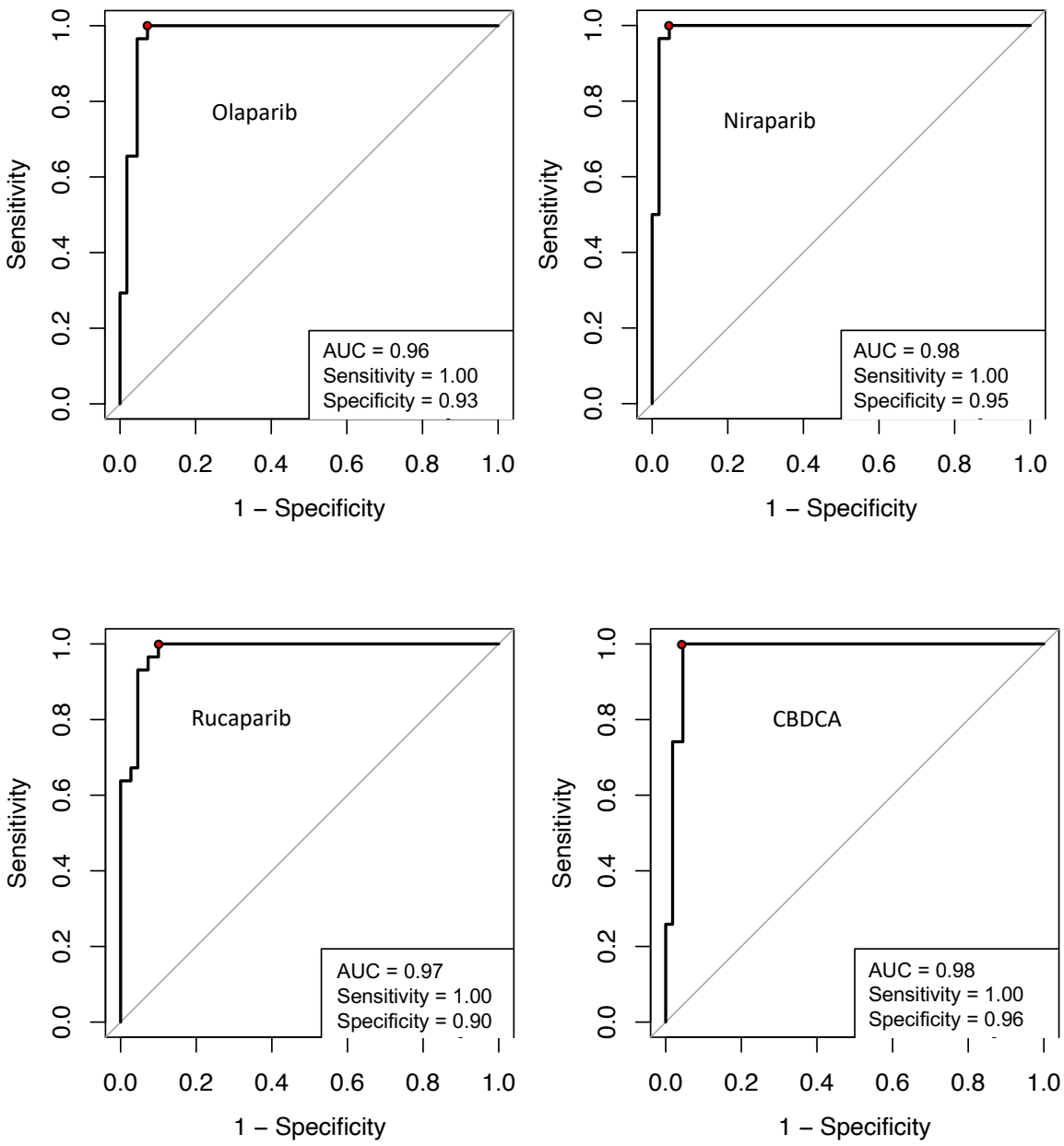
1050

1051 **Supplementary Figure 4:** Histograms of variant distribution in a batch of the MANO-B

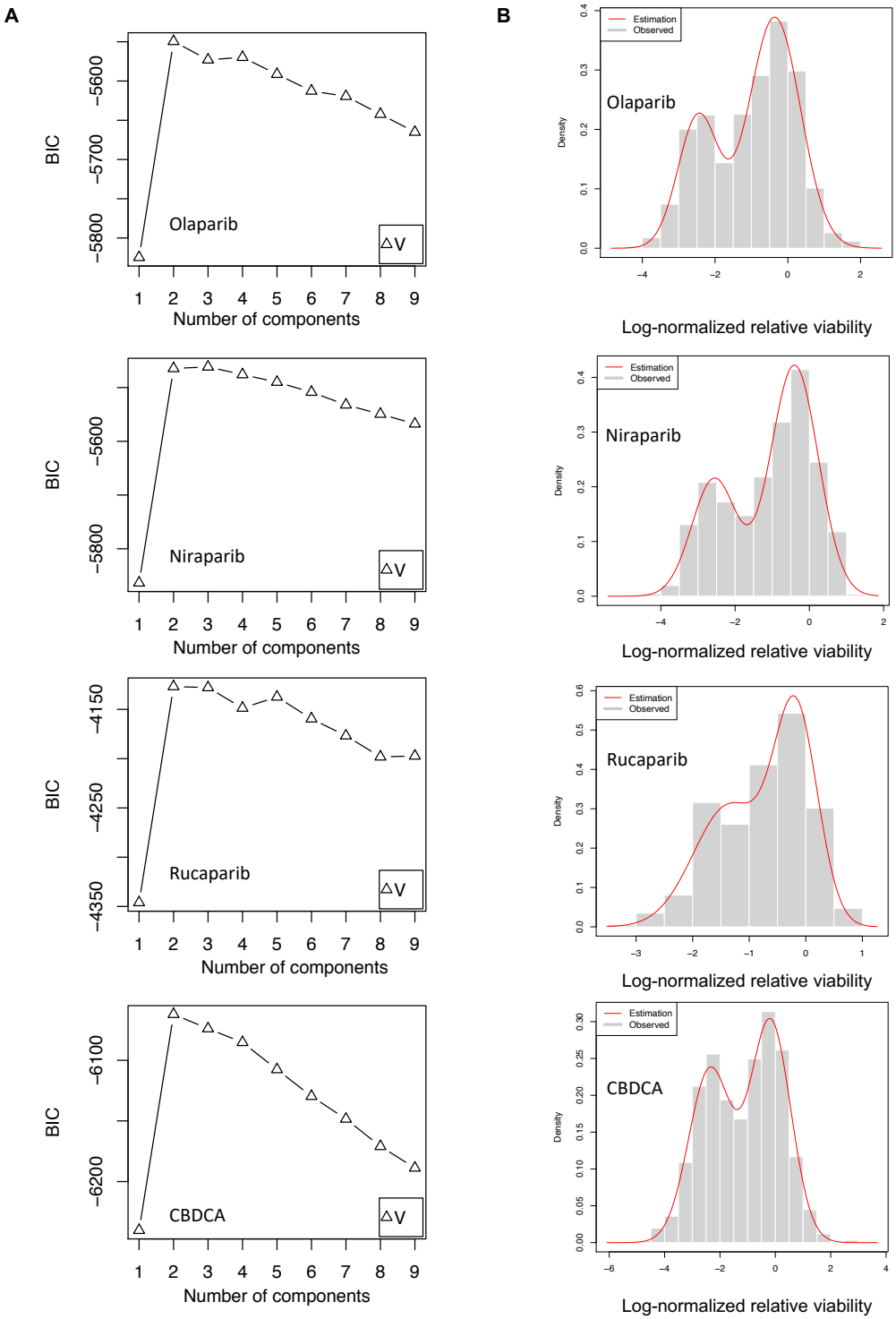
1052 method. **A,** The distribution of each variant ratio at day 0 is shown. **B,** The fold change of each

1053 variant from day 0 to day 12 with DMSO treatment is shown. SD, standard deviation.

1054

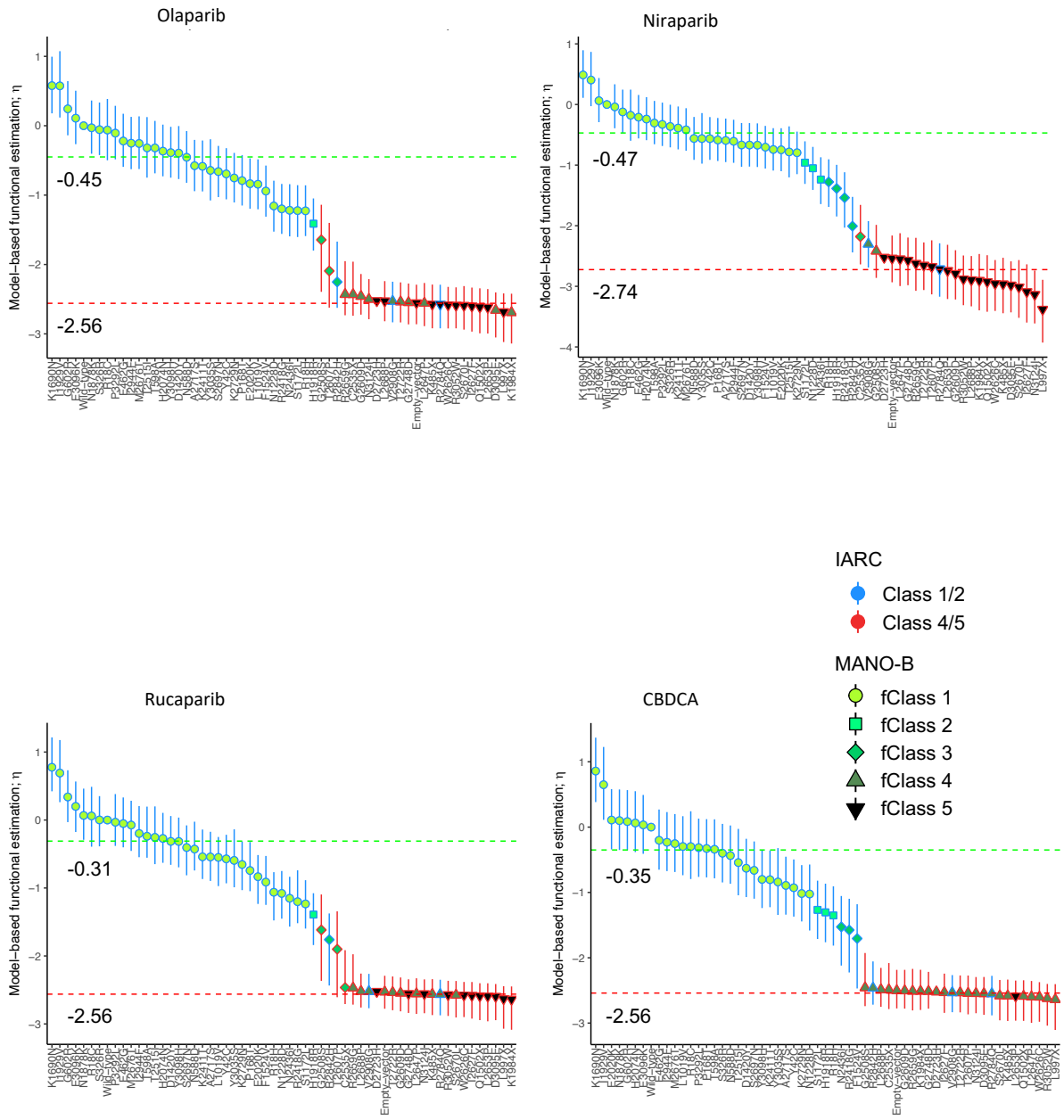


Supplementary Figure 5: ROC curves for data obtained from 37 known benign and 22 known pathogenic *BRCA2* variants. All established benign (IARC class 1/2) and pathogenic (IARC class 4/5) variants were extracted from the 244 variants. Potential hypomorphic variants, R2784Q, R2842H, and G2908V, were classified as pathogenic at the indicated threshold, and all other variants were correctly classified.

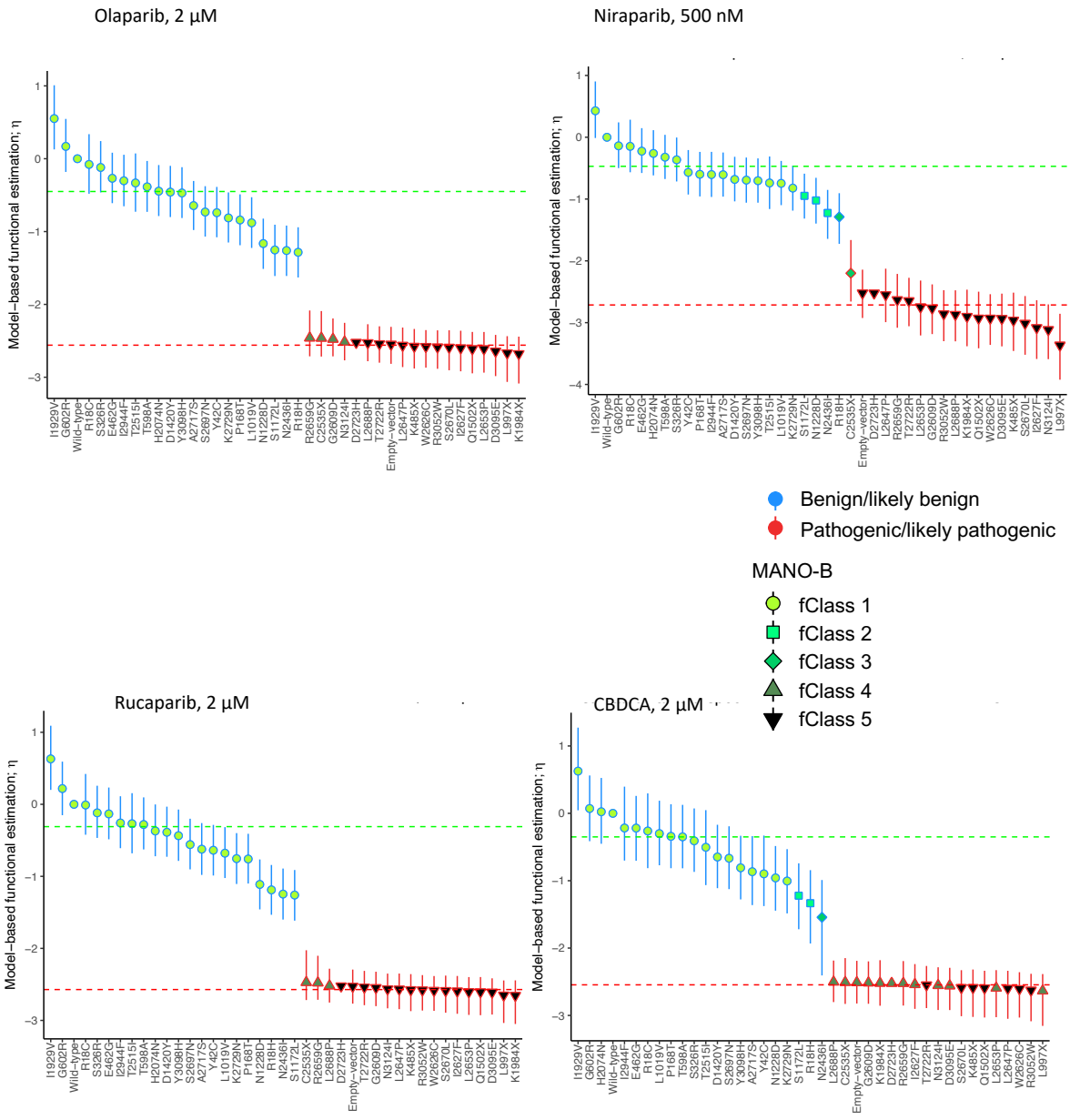


Supplementary Figure 6: Model estimation for data obtained from 244 *BRCA2* variants and an empty vector. **A,** Expectation–maximization algorithm to determine the appropriate model for the data. According to the Bayesian information criterion (BIC), a two-component Gaussian mixture model was the appropriate model. **B,** The estimated two-component Gaussian mixture

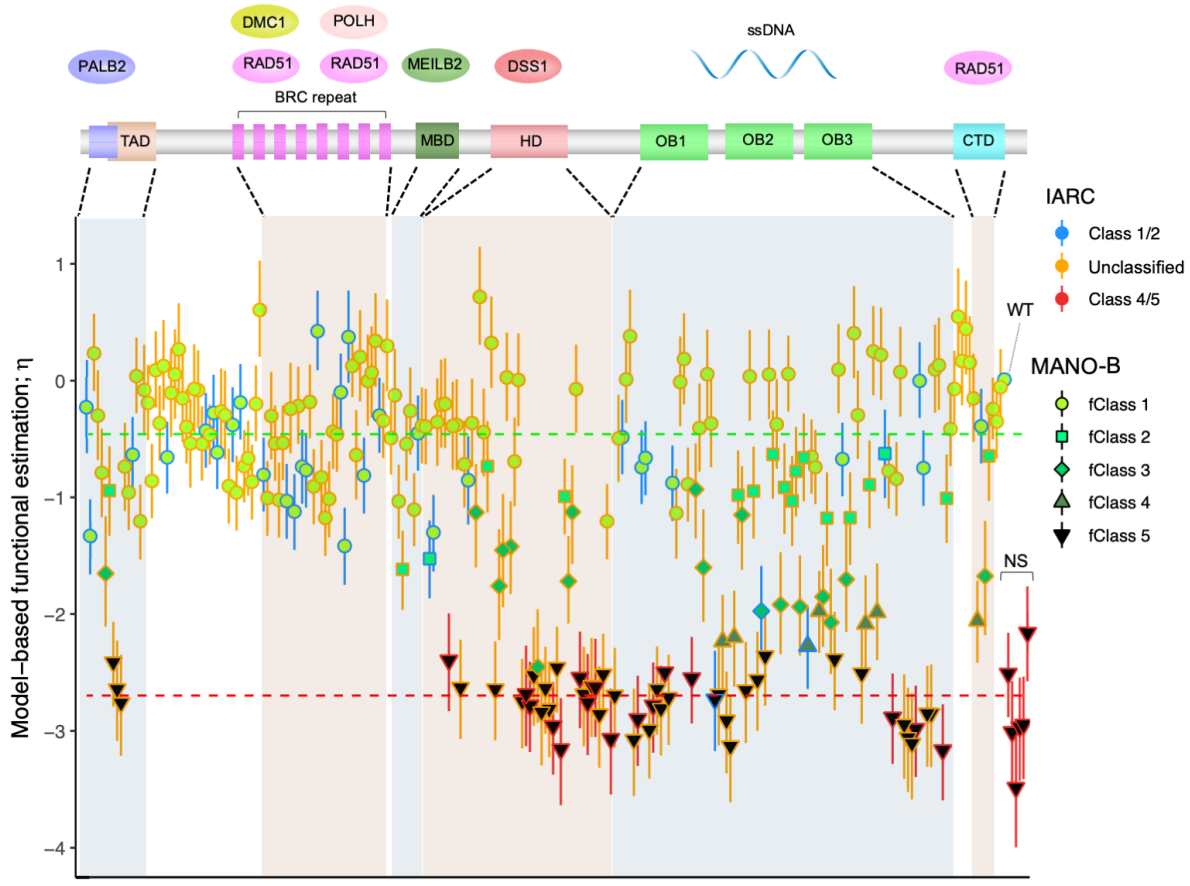
1066 distribution density curve (red line) fit the distribution of the observed log-normalized relative
1067 viability data. The left component corresponds to deleterious variants, whereas the right
1068 component corresponds to neutral variants.



Supplementary Figure 7: Estimated function and classification for the training set comprising known benign and pathogenic *BRCA2* variants. The bar plots show the functional variant-specific effect η. The classification of each variant is indicated by the color and shape of the lines and plots, as shown in the legend. The η values of the wild-type and the D2723H variant are anchored to $\log_{10}(1.0)$ and $\log_{10}(0.003)$, respectively. The dotted lines indicate the center of the neutral and deleterious distributions. Open error bars, 95% CIs.



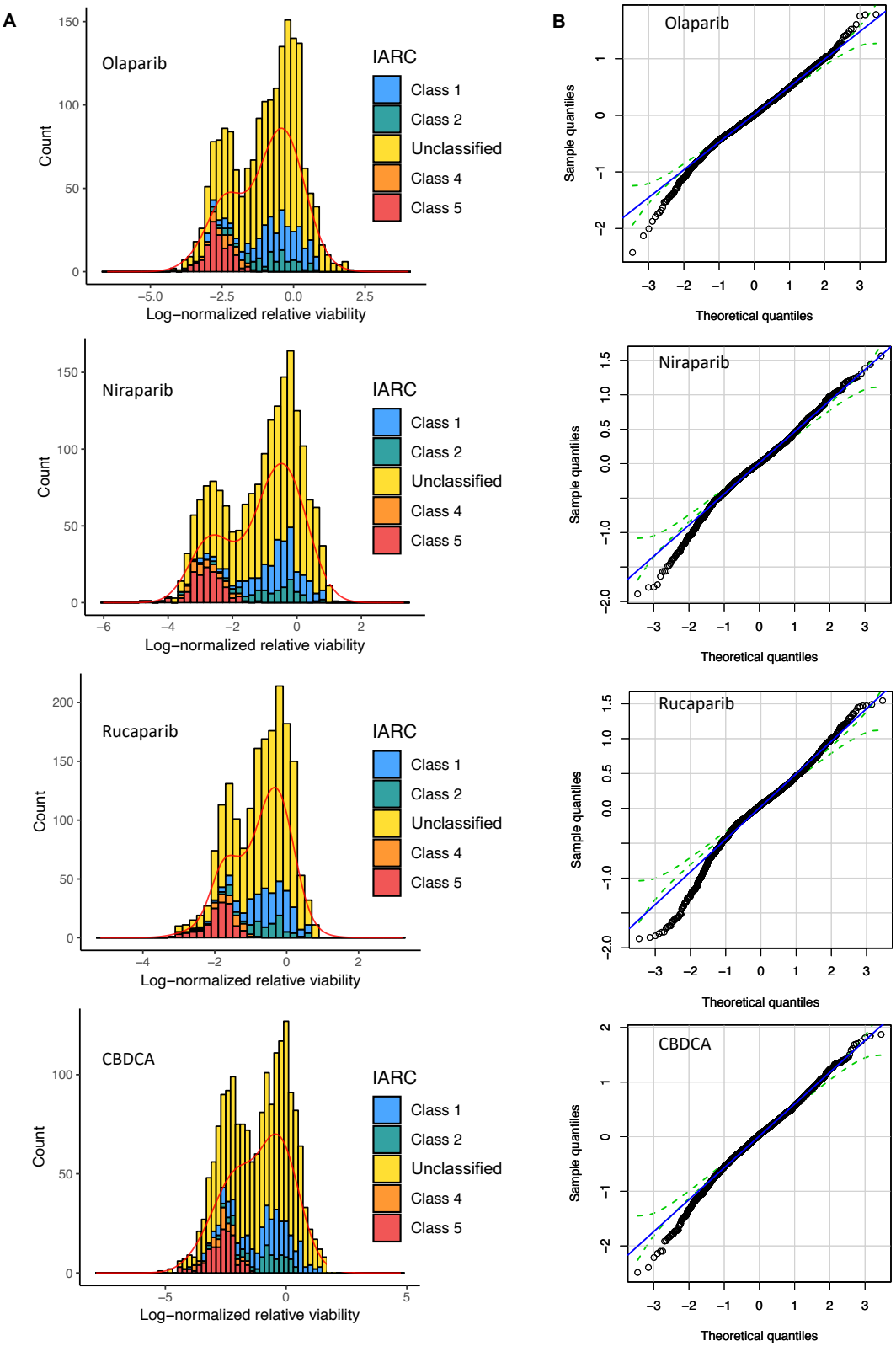
Supplementary Figure 8: Estimated function and classification for all the established benign and pathogenic *BRCA2* variants defined by the ACMG guidelines criteria. The barplots show the functional variant-specific effect η. The classification of each variant is indicated by the color and shape of the lines and plots, as shown in the legend. The η values of the wild-type and the D2723H variant are anchored to $\log_{10}(1.0)$ and $\log_{10}(0.003)$, respectively. The dotted lines indicate the center of the neutral and deleterious distributions. Open error bars, 95% CIs.



1084

1085 **Supplementary Figure 9:** Bar plots of functional variant effects and functional diagnosis of
1086 244 *BRCA2* variants analyzed with Align-GVGD-based prior probability. The bar plots show
1087 the functional variant-specific effect η. Bayesian inference was performed with prior
1088 probability values according to the Align-GVGD algorithm. The classification of each variant
1089 is indicated by the color and shape of the lines and plots, as shown in the legend. The η_v values
1090 of the wild-type and the D2723H variant are anchored to $\log_{10}(1.0)$ and $\log_{10}(0.003)$,
1091 respectively. Dotted lines indicate the median value of the neutral and deleterious distributions.
1092 Open error bars, 95% CIs. WT, wild-type; NS, nonsense variants.

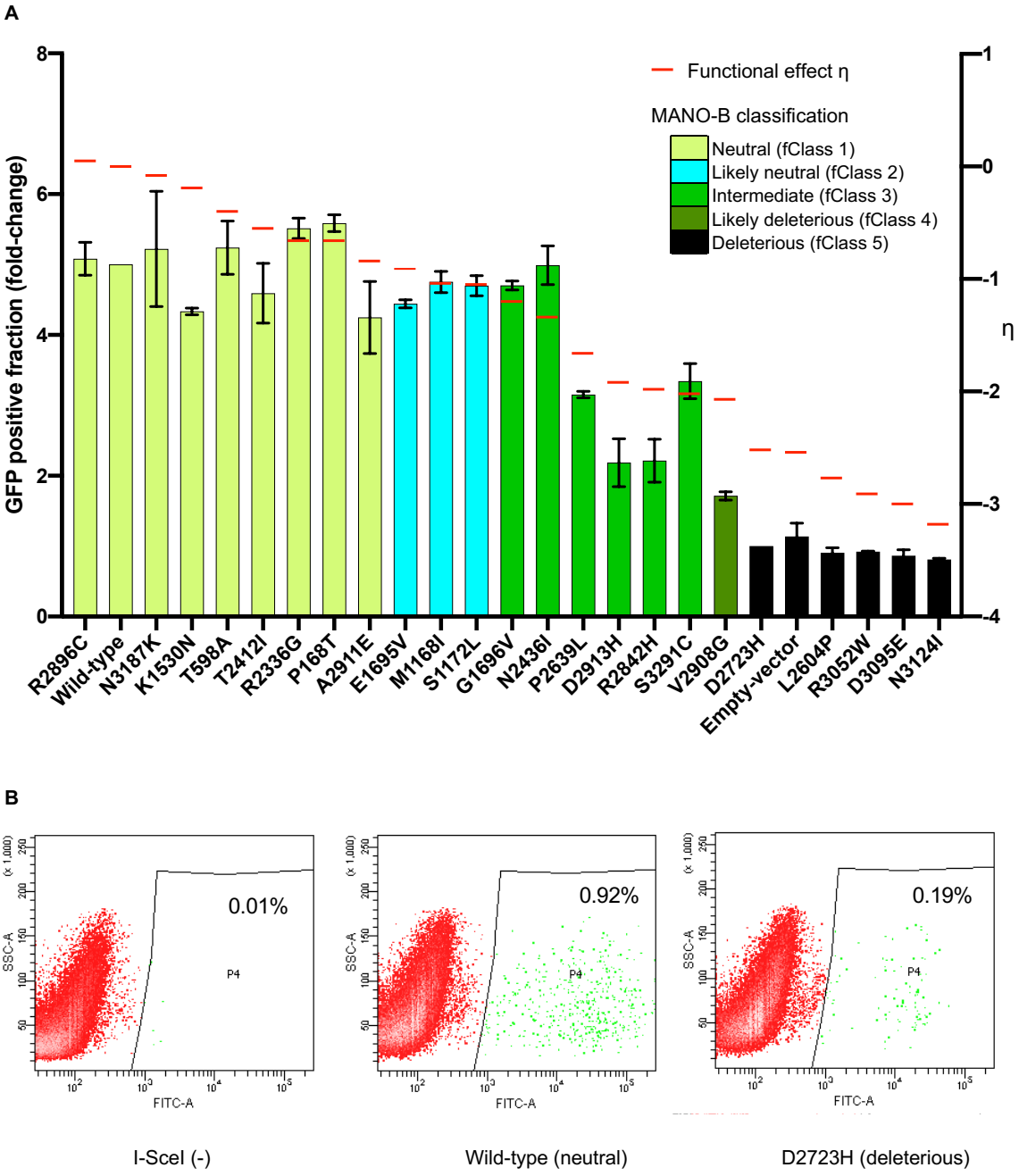
1093



Supplementary Figure 10: Posterior predictive checks of model parameters. **A,** Posterior predictive density curves of the log-normalized relative viability values for the 244 variants fit

1097 to the histogram of observed data. **B,** Normal Q-Q plots of the expected standardized residuals

1098 of the log-normalized relative viability.



Supplementary Figure 11: Homology-directed repair assay of 24 *BRCA2* variants. **A,** MANO-B classifications exhibited good correlation with the results of the HDR assay. *BRCA2* variant expression vectors and an I-SceI expression vector were cotransfected into DLD1 *BRCA2* (−/−) cells harboring the DRGFP sequence, and GFP-positive cells were counted by FACS 4 d after transfection. The GFP-positive fraction was normalized and rescaled relative to a 1:5 ratio of D2723H:wild-type. All variants were analyzed in biological

duplicate and technical triplicate. The colors of the bar graphs indicate the MANO-B classification of each variant. The heights of the bar graphs indicate the GFP-positive fractions (HDR assay). Red lines show the functional effects of each variant (η_v). Error bars, s.d. (n = 2); ηv, functional variant-specific effect. **B,** Gating for GFP-positive cells.

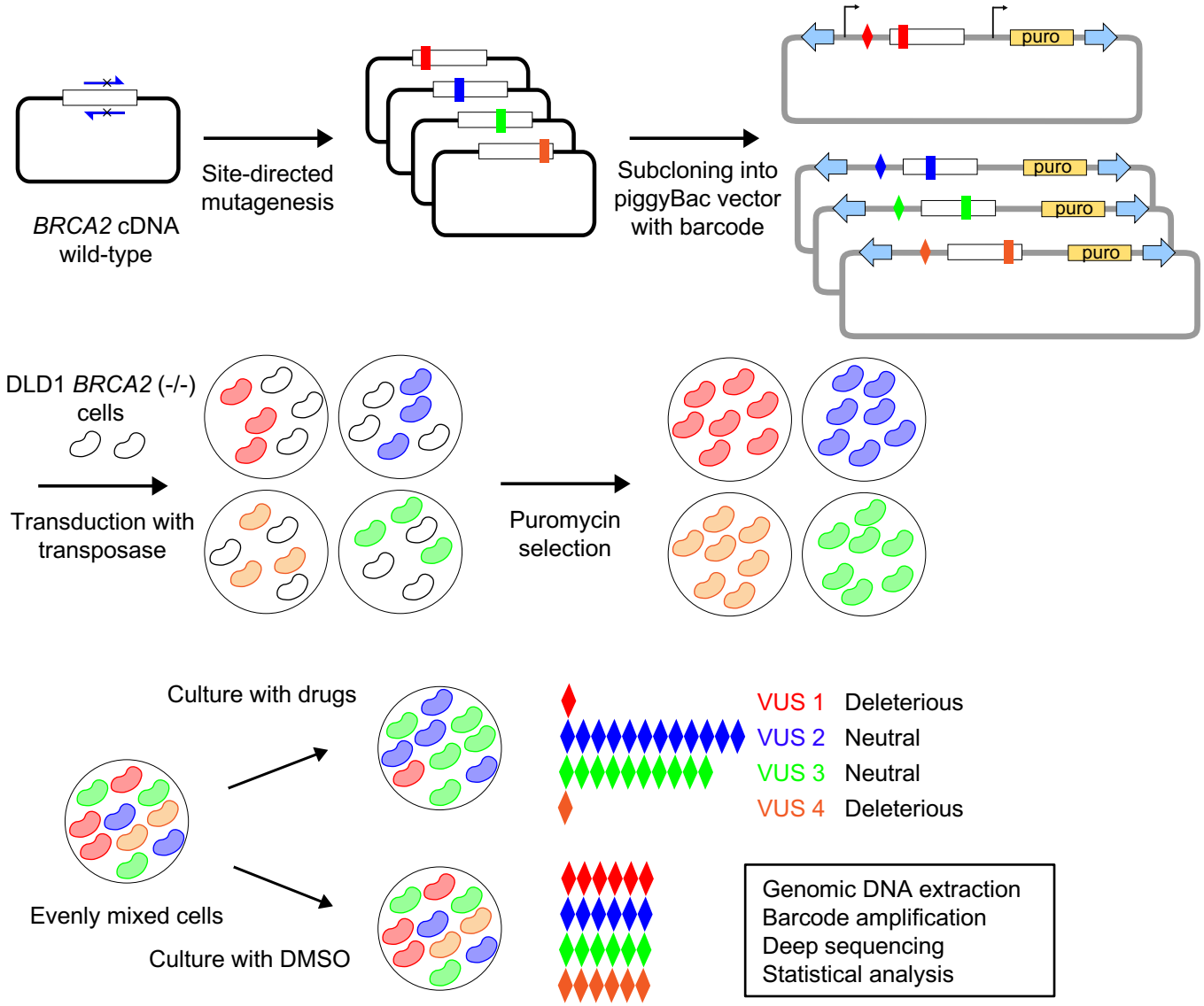

BRCA2 cDNA
wild-type
Site-directed
mutagenesis
Subcloning into
piggyBac vector
with barcode
puro
puro
puro
DLD1 BRCA2 (-/-)
cells
Transduction with
transposase
Puromycin
selection
Culture with drugs
Evenly mixed cells
Culture with DMSO
VUS 1 Deleterious
VUS 2 Neutral
VUS 3 Neutral
VUS 4 Deleterious
Genomic DNA extraction
Barcode amplification
Deep sequencing
Statistical analysis

Figure 1.

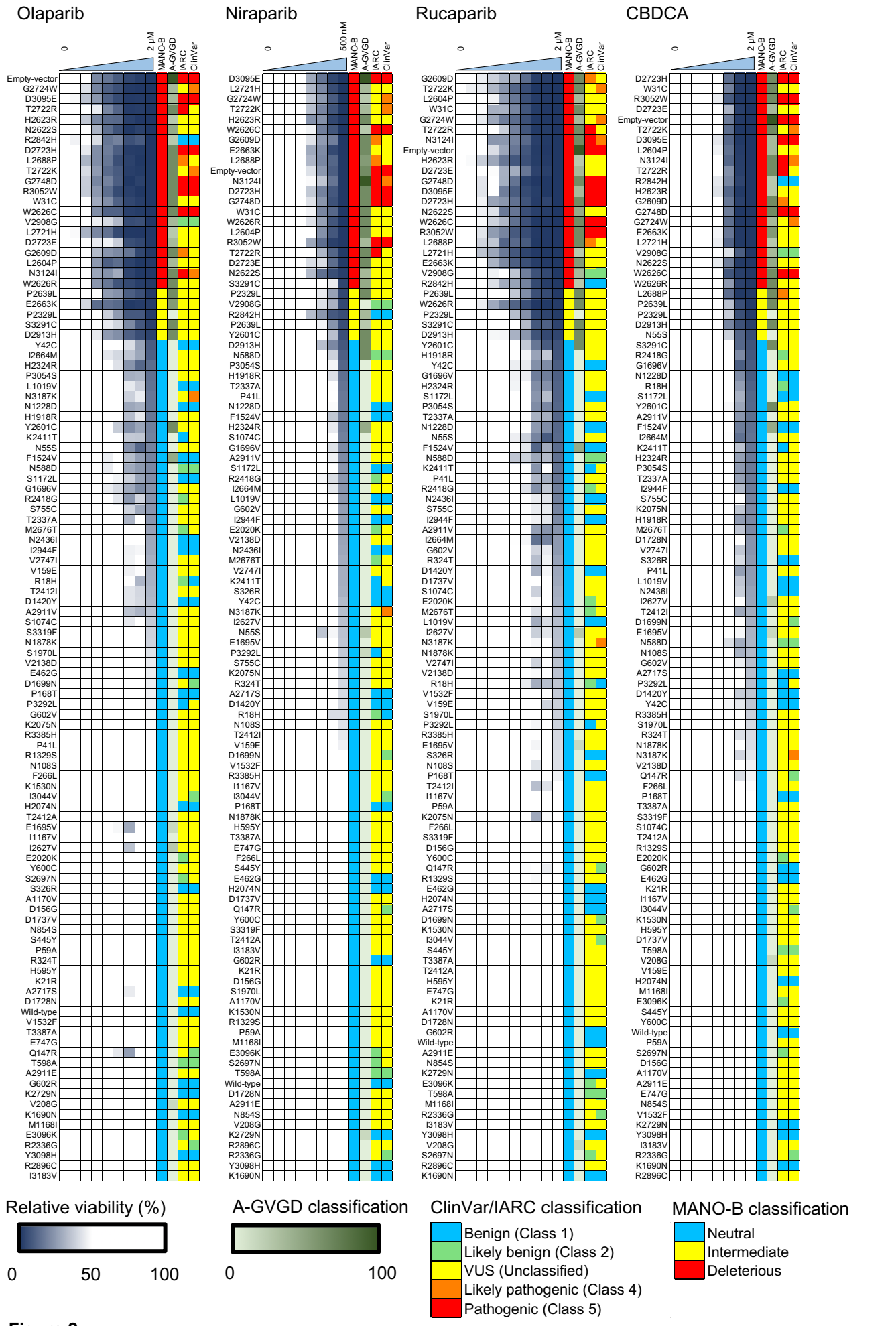

Olaparib
Niraparib
Rucaparib
CBDCA
0
2 µM
500 nM
MANO-B
A-GVGD
IARC
ClinVar
Empty-vector
G2724W
D3095E
T2722R
H2623R
N2622S
R2842H
D2723H
L2688P
T2722K
G2748D
R3052W
W31C
W2626C
V2908G
L2721H
D2723E
G2609D
L2604P
N3124I
W2626R
P2639L
E2663K
P2329L
S3291C
D2913H
Y42C
I2664M
H2324R
P3054S
L1019V
N3187K
N1228D
H1918R
Y2601C
K2411T
N55S
F1524V
N588D
S1172L
G1696V
R2418G
S755C
T2337A
M2676T
N2436I
I2944F
V2747I
V159E
R18H
T2412I
D1420Y
A2911V
S1074C
S3319F
N1878K
S1970L
V2138D
E462G
D1699N
P168T
P3292L
G602V
K2075N
R3385H
P41L
R1329S
N108S
F266L
K1530N
I3044V
H2074N
T2412A
E1695V
I1167V
I2627V
E2020K
Y600C
S2697N
S326R
A1170V
D156G
D1737V
N854S
S445Y
P59A
R324T
H595Y
K21R
A2717S
D1728N
Wild-type
V1532F
T3387A
E747G
Q147R
T598A
A2911E
G602R
K2729N
V208G
K1690N
M1168I
E3096K
R2336G
Y3098H
R2896C
I3183V
D3095E
L2721H
G2724W
T2722K
H2623R
W2626C
G2609D
E2663K
L2688P
Empty-vector
N3124I
D2723H
G2748D
W31C
W2626R
L2604P
R3052W
T2722R
D2723E
N2622S
S3291C
P2329L
V2908G
R2842H
P2639L
Y2601C
D2913H
N588D
P3054S
H1918R
T2337A
P41L
N1228D
F1524V
H2324R
S1074C
G1696V
A2911V
S1172L
R2418G
I2664M
L1019V
G602V
I2944F
E2020K
V2138D
N2436I
M2676T
V2747I
K2411T
S326R
Y42C
N3187K
I2627V
N55S
E1695V
P3292L
S755C
K2075N
R324T
A2717S
D1420Y
R18H
N108S
T2412I
V159E
D1699N
V1532F
R3385H
I1167V
I3044V
P168T
N1878K
H595Y
T3387A
E747G
F266L
S445Y
E462G
H2074N
D1737V
Q147R
Y600C
S3319F
T2412A
I3183V
G602R
K21R
D156G
S1970L
A1170V
K1530N
R1329S
P59A
M1168I
E3096K
S2697N
T598A
Wild-type
D1728N
A2911E
N854S
V208G
K2729N
R2896C
R2336G
Y3098H
K1690N
G2609D
T2722K
L2604P
W31C
G2724W
T2722R
N3124I
Empty-vector
H2623R
D2723E
G2748D
D3095E
D2723H
N2622S
W2626C
R3052W
L2688P
L2721H
E2663K
V2908G
R2842H
P2639L
W2626R
P2329L
S3291C
D2913H
Y2601C
H1918R
Y42C
G1696V
H2324R
S1172L
P3054S
T2337A
N1228D
N55S
F1524V
N588D
K2411T
P41L
R2418G
N2436I
S755C
I2944F
A2911V
I2664M
G602V
R324T
D1420Y
D1737V
S1074C
E2020K
M2676T
L1019V
I2627V
N3187K
N1878K
V2747I
V2138D
R18H
V1532F
V159E
S1970L
P3292L
R3385H
E1695V
S326R
N108S
P168T
T2412I
I1167V
P59A
K2075N
F266L
S3319F
D156G
Y600C
Q147R
R1329S
E462G
H2074N
A2717S
D1699N
K1530N
I3044V
S445Y
T3387A
T2412A
H595Y
E747G
K21R
A1170V
D1728N
G602R
Wild-type
A2911E
N854S
K2729N
E3096K
T598A
M1168I
R2336G
I3183V
Y3098H
V208G
S2697N
R2896C
K1690N
D2723H
W31C
R3052W
D2723E
Empty-vector
T2722K
D3095E
L2604P
N3124I
T2722R
R2842H
H2623R
G2609D
G2748D
G2724W
E2663K
L2721H
V2908G
N2622S
W2626C
W2626R
L2688P
P2639L
P2329L
D2913H
N55S
S3291C
R2418G
G1696V
N1228D
R18H
S1172L
Y2601C
A2911V
F1524V
I2664M
K2411T
H2324R
P3054S
T2337A
I2944F
S755C
K2075N
H1918R
M2676T
D1728N
V2747I
S326R
P41L
L1019V
N2436I
I2627V
T2412I
D1699N
E1695V
N588D
N108S
G602V
A2717S
P3292L
D1420Y
Y42C
R3385H
S1970L
R324T
N1878K
N3187K
V2138D
Q147R
F266L
P168T
T3387A
S3319F
S1074C
T2412A
R1329S
E2020K
G602R
E462G
K21R
I1167V
I3044V
K1530N
H595Y
D1737V
T598A
V208G
V159E
H2074N
M1168I
E3096K
S445Y
Y600C
Wild-type
P59A
S2697N
D156G
A1170V
A2911E
E747G
N854S
V1532F
K2729N
Y3098H
I3183V
R2336G
K1690N
R2896C
Relative viability (%)
0
50
100
A-GVGD classification
0
100
ClinVar/IARC classification
Benign (Class 1)
Likely benign (Class 2)
VUS (Unclassified)
Likely pathogenic (Class 4)
Pathogenic (Class 5)
MANO-B classification
Neutral
Intermediate
Deleterious

Figure 2.

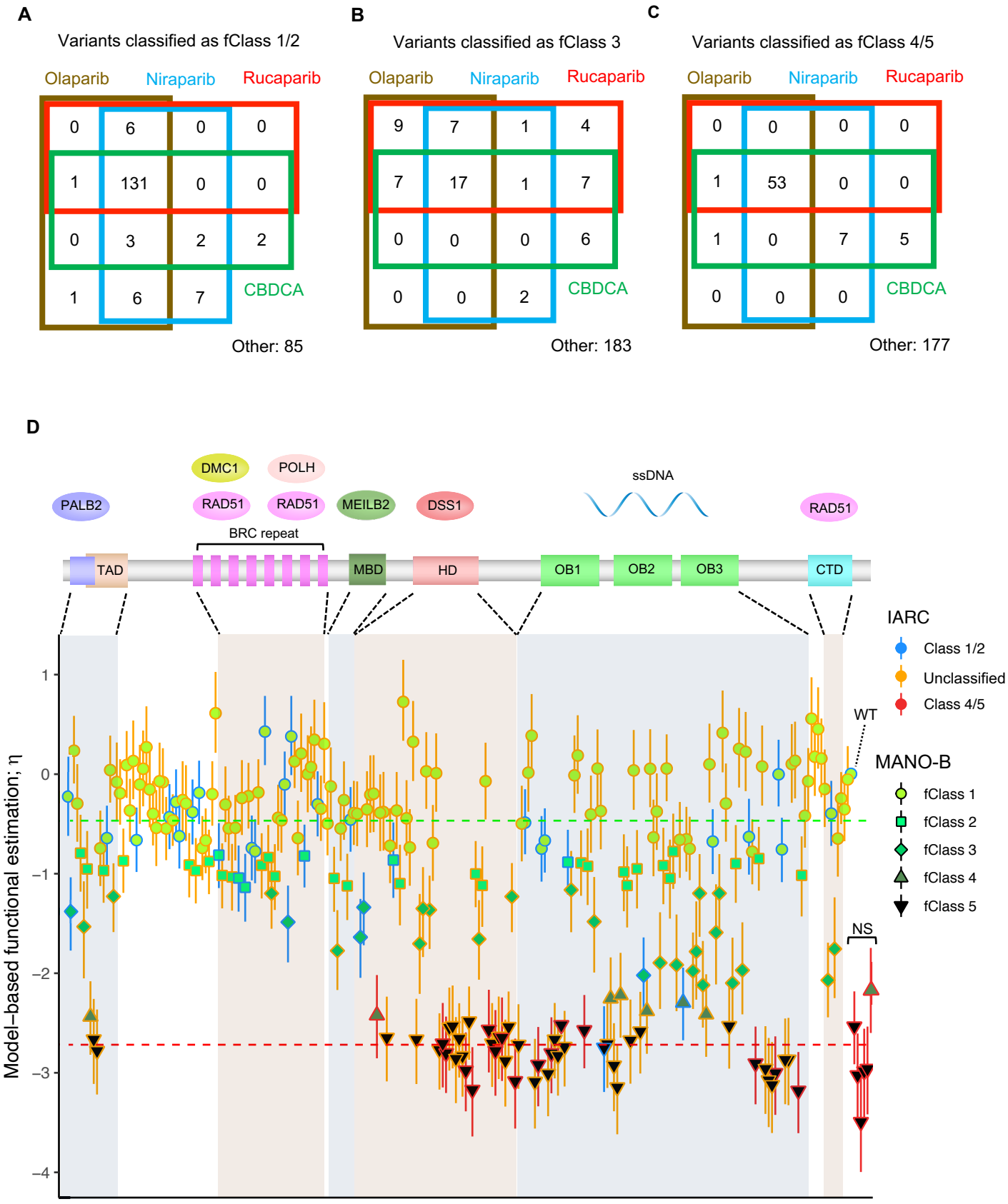

A
Variants classified as fClass 1/2
Olaparib
Niraparib
Rucaparib
0 6 0 0
1 131 0 0
0 3 2 2
1 6 7
CBDCA
Other: 85
B
Variants classified as fClass 3
Olaparib
Niraparib
Rucaparib
9 7 1 4
7 17 1 7
0 0 0 6
0 0 2
CBDCA
Other: 183
C
Variants classified as fClass 4/5
Olaparib
Niraparib
Rucaparib
0 0 0 0
1 53 0 0
1 0 7 5
0 0 0
CBDCA
Other: 177
D
DMC1
POLH
PALB2
RAD51
RAD51
MEILB2
DSS1
ssDNA
RAD51
BRC repeat
TAD
MBD
HD
OB1
OB2
OB3
CTD
IARC
Class 1/2
Unclassified
Class 4/5
WT
MANO-B
fClass 1
fClass 2
fClass 3
fClass 4
fClass 5
NS
Model–based functional estimation; η
1
0
−1
−2
−3
−4

Figure 3.

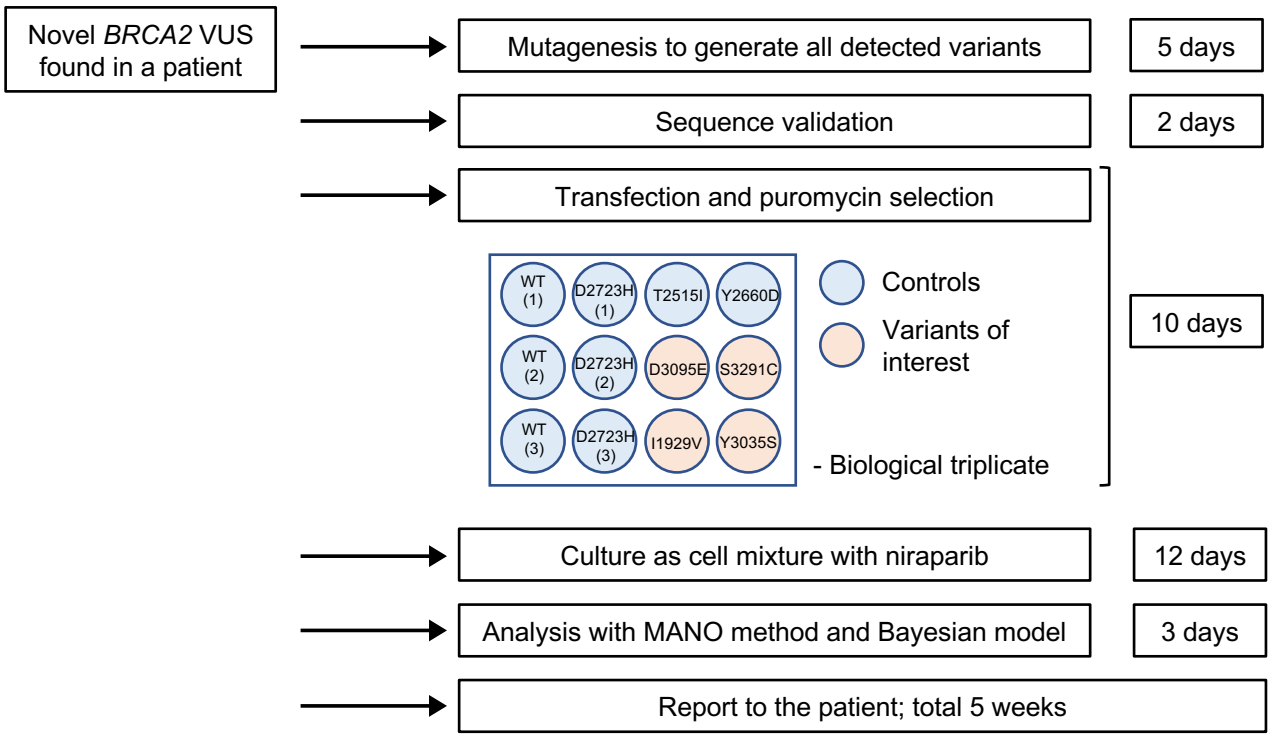

Novel *BRCA2* VUS found in a patient
Mutagenesis to generate all detected variants
5 days
Sequence validation
2 days
Transfection and puromycin selection
WT (1)
D2723H (1)
T2515I
Y2660D
WT (2)
D2723H (2)
D3095E
S3291C
WT (3)
D2723H (3)
I1929V
Y3035S
Controls
Variants of interest
- Biological triplicate
10 days
Culture as cell mixture with niraparib
12 days
Analysis with MANO method and Bayesian model
3 days
Report to the patient; total 5 weeks

Figure 4.

Table 1. Results of an experimental ABCD test.

Variant	IARC classification	ABCD test	η			BF	fClass
			Mean	95% CI			
				LL	UL		
I1929V	Class 1	Previous batches	0.37	-0.02	0.79	1.05×10^{-5}	fClass 1
		New batch-1	0.19	-0.33	0.73	1.02×10^{-5}	fClass 1
		New batch-2	-0.23	-0.73	0.28	1.63×10^{-4}	fClass 1
		New batch-3	0.08	-0.43	0.6	3.71×10^{-5}	fClass 1
		3 batches combined	0.11	-0.23	0.44	1.40×10^{-5}	fClass 1
Y3035S	Class 2	Previous batches	-0.63	-1.01	-0.25	1.11×10^{-3}	fClass 1
		New batch-1	-0.55	-1.05	-0.04	1.12×10^{-3}	fClass 1
		New batch-2	-0.53	-1.04	-0.02	9.57×10^{-4}	fClass 1
		New batch-3	-0.36	-0.87	0.16	3.01×10^{-4}	fClass 1
		3 batches combined	-0.47	-0.79	-0.15	2.79×10^{-4}	fClass 1
D3095E	Class 5	Previous batches	-3	-3.41	-2.62	1.33×10^{4}	fClass 5
		New batch-1	-2.31	-2.8	-1.78	4.45×10^{1}	fClass 4
		New batch-2	-2.63	-3.12	-2.15	5.75×10^{2}	fClass 5
		New batch-3	-2.69	-3.2	-2.19	9.28×10^{2}	fClass 5
		3 batches combined	-2.52	-2.85	-2.19	7.85×10^{2}	fClass 5
S3291C	Unclassified	Previous batches	-2.07	-2.43	-1.69	1.15×10^{1}	fClass 3
		New batch-1	-1.06	-1.7	-0.53	6.15×10^{-2}	fClass 3
		New batch-2	-2.06	-2.6	-1.34	7.05×10^{0}	fClass 3
		New batch-3	-1.35	-2.12	-0.74	3.17×10^{-1}	fClass 3
		3 batches combined	-1.46	-1.86	-1.11	3.39×10^{-1}	fClass 3

BF, Bayes factor; LL, lower limit; UL, upper limit; CI, credible interval.