Homologous Recombination Abnormalities Associated With *BRCA1/2* Mutations as Predicted by Machine Learning of Targeted Next-Generation Sequencing Data

Maher Albitar[1], Hong Zhang[1], Andrew Pecora[2], Stanley Waintraub[2], Deena Graham[2], Mira Hellmann[2], Donna McNamara[2], Ahmad Charifa[1], Ivan De Dios[1], Wanlong Ma[1] and Andre Goy[2]

[1]Genomic Testing Cooperative, Irvine, CA, USA. [2]John Theurer Cancer Center at Hackensack University Medical Center, Hackensack, NJ, USA.

ABSTRACT

BACKGROUND: Homologous recombination deficiency (HRD) is the hallmark of breast cancer gene 1/2 (*BRCA1/2*)-mutated tumors and the unique biomarker for predicting response to double-strand break (DSB)–inducing drugs. The demonstration of HRD in tumors with mutations in genes other than *BRCA1/2* is considered the best biomarker of potential response to these DSB-inducer drugs.

OBJECTIVES: We explored the potential of developing a practical approach to predict in any tumor the presence of HRD that is similar to that seen in tumors with *BRCA1/2* mutations using next-generation sequencing (NGS) along with machine learning (ML).

DESIGN: We use copy number alteration (CNA) generated from routine-targeted NGS data along with a modified naïve Bayesian model for the prediction of the presence of HRD.

METHODS: The CNA from NGS of 434 targeted genes was analyzed using CNVkit software to calculate the log2 of CNA changes. The log2 values of various sequencing reads (bins) were used in ML to train the system on predicting tumors with *BRCA1/2* mutations and tumors with abnormalities similar to those detected in *BRCA1/2* mutations.

RESULTS: Using 31 breast or ovarian cancers with *BRCA1/2* mutations and 84 tumors without mutations in any of 12 homologous recombination repair (HRR) genes, the ML demonstrated high sensitivity (90%, 95% confidence interval [CI] = 73%-97.5%) and specificity (98%, 95% CI = 90%-100%). Testing of 114 tumors with mutations in HRR genes other than *BRCA1/2* showed 39% positivity for HRD similar to that seen in *BRCA1/2*. Testing 213 additional wild-type (WT) cancers showed HRD positivity similar to *BRCA1/2* in 32% of cases. Correlation with proportional loss of heterozygosity (LOH) as determined using whole exome sequencing of 51 samples showed 90% (95% CI = 72%-97%) concordance. The approach was also validated in an independent set of 1312 consecutive tumor samples.

CONCLUSIONS: These data demonstrate that CNA when combined with ML can reliably predict the presence of *BRCA1/2* level HRD with high specificity. Using *BRCA1/2* mutant cases as gold standard, this ML can be used to predict HRD in cancers with mutations in other HRR genes as well as in WT tumors.

KEYWORDS: Homologous recombination deficiency, BRCA1, BRCA2, double-strand break, copy number variation, next-generation sequencing, machine learning, prediction, PARP inhibitors, response

Background

The presence of homologous recombination deficiency (HRD) due to DNA double-strand break (DSB) repair deficiency is the hallmark of cancers that carry abnormalities in breast cancer gene 1/2 (*BRCA1/2*) genes.[1-4] The demonstration of HRD has been accepted as a biomarker for response to DSB-inducing drugs, including platinum salts and poly ADP-ribose polymerase inhibitors (PARPis).[5,6] Typically, the presence of a germline mutation in *BRCA1/2* is considered the gold standard biomarker for response to PARPi.[7-13]

Most of the testing for the presence of HRD is based on testing of the presence of mutations in *BRCA1/BRCA2* genes or in other genes involved in homologous recombination repair (HRR), such as *PALB2* or *RAD51*.[14-16] However, the effects of HRD can be demonstrated by the so-called genomic scars detected in the genome of a cancer resulting from the oncogenesis driven by HRD.[14-16] This approach is also reported to be specifically helpful in cases where the gene involved in the HRR is inactivated by a mechanism other than mutations such as methylation or deletion.[15,16] The Food and Drug Administration (FDA) has approved the tests that predict these genomic scars as companion tests for certain PARPi.

The evaluation of these genomic scars is based on assessing chromosomal structural alterations that are typically detected

when *BRCA1/2* genes are mutated and driving oncogenesis. This approach allows for the detection of BRCA-like tumors that may be responsive to DSB-inducing drugs.

Most of the methods used for detecting these scars are based on evaluating loss of heterozygosity (LOH) as well as structural rearrangements. It has been documented that the level of chromosomal aberrations correlates with HRD status. Patients with high LOH (>16%) showed some improved response to PARP inhibitor (rucaparib) as compared with placebo control, but not as good as that seen in the BRCA-mutated group.[17] This suggested that LOH is associated with a higher likelihood of response to DSB-inducing. Subsequent studies added to the level of LOH deletion of stretches larger than 15 Mb but smaller than the whole chromosome,[18] telomeric allelic imbalance (TAI), and large-scale transitions (LSTs). The addition of these measurements improved the prediction of the presence of *BRCA1/2*-associated scars.[17] Telomeric allelic imbalance evaluates if the paternal and maternal alleles are equal; LST evaluates chromosomal aberrations involving large chromosomal regions more than 10 Mb apart.[18] This combination of abnormalities generates a score that is currently used for selecting patients for therapy with DSB-inducer drugs. In a retrospective study of chemotherapy in breast and ovarian cancer, patients with known *BRCA1/2* status were used as control. A score of 42 showed good prediction of *BRCA1/2* mutation, and HRD was a significant predictor of residual cancer burden and pathologic complete response (pCR) when *BRCA1/2* were included, but it was borderline statistically relevant when *BRCA1/2* nonmutated cases were considered.[18] Different approaches have been explored to evaluate HRD including whole-genome sequencing (WGS),[19] comparative genomic hybridization (CGH),[20] and expression profiling[20] and functional assays.[14] Multiple subsequent studies in breast and ovarian cancer have also studied such scores and demonstrated that the presence of BRCA mutation is the best predictor of response to DSB-inducers; having a low HRD score in a non-*BRCA1/2* tumor can be used as an indicator of poor response to PARPi.[21]

More recent study (SWOG S9313 phase 3 study)[22] has demonstrated that disease-free survival (DFS) was better for triple-negative breast cancer patients with HRD in general as compared with patients without HRD. Unfortunately, the design of this study did not allow for determining the predictive value of the HRD score by itself.[22]

Overall, the currently used assays for predicting HRD are useful and accepted by FDA as companion tests. However, some studies in both high-grade serous ovarian or endometrial cancers and triple-negative breast cancer suggested that value of the HRD score is of limited clinical value.[22-24]

We rationalized that the key for predicting the presence or absence of HRD is to compare genomic abnormalities of tumors with those *BRCA1/2* mutation-positive tumors. We

used copy number alteration (CNA) abnormalities detected in *BRCA1/2* mutation-positive cases along with a machine learning (ML) to build a model for predicting HRD. In this model, we demonstrate very high sensitivity in predicting cases with *BRCA1/2* mutations and in predicting cases with similar abnormalities. Although there is overlap between our approach and prior approaches, the use of ML may improve the approach.

Methods

Patient samples

Formalin-fixed, paraffin-embedded (FFPE) cancer samples were sequenced using a targeted next-generation sequencing (NGS) panel of 434 genes. This included 31 patients with breast or ovarian cancer with confirmed *BRCA1/BRCA2* mutations, 84 cancer samples with no evidence of mutations in *BRCA1/2* or any HRR genes, 114 cancers with mutations in one of the genes involved in HRR, 213 additional breast or ovarian samples wild-type (WT) for HRR genes, and 51 random samples tested with the targeted sequencing panel and with whole exome sequencing (WES; Table 1). The 31 *BRCA1/2*-mutated samples used for establishing the ML included 22 (74%) with breast cancer and 9 (26%) with ovarian cancer whereas the negative samples included 28 (33%) with breast cancer, 18 (22%) with lung cancer, 27 (32%) with ovarian cancer, 8 (10%) with pancreatic cancer, and 3 patients (3%) with prostate cancer. The HRR genes that were considered included *PALB2*, *CDK12*, *RAD50*, *RAD51*, *RAD51C*, *RAD54L*, *MRE11A*, *NBM*, *ATM*, *ATR*, *FANCA*, and *FANCC*. Cases were considered DSB mutant if the mutation was heterozygous or homozygous. In addition, after developing the ML and in a prospective testing, we analyzed HRD in 1312 consecutive solid tumor samples from various tissue including breast, lung, colorectal, head and neck, ovary, skin, pancreas, and others. Homologous recombination deficiency results were correlated with *BRCA1/2* and other genes involved in HRR in these tumors. The tumor tissue was macrodissected from slides, and only samples with tumor percentage at 30% or greater were included. Our validations showed that all chromosomal structural abnormalities are adequately captured when tumor fraction is ⩾30%.

Targeted next-generation sequencing and copy number variation evaluation

The DNA from FFPE was extracted using FormaPure and KingFisher Flex. The extracted DNA from FFPE was sequenced using 100 ng of DNA. Library for targeted 434 gene sequencing is based on Single Primer Extension (SPE) chemistry. The 434 gene panel was a custom panel included genes reported to play a role in the oncogenesis of various types of solid tumors. The DNA sequencing includes all coding exons of the 434 genes. For each exon, approximately 50 intronic

Table 1. List of samples used for training and validating the machine learning algorithm.

	NUMBER	HRD
BRCA1/2-mutated ovarian/breast cases	31	98%
No mutation in HRR genes (WT) ovarian/breast cases	84	5%
Ovarian/breast samples with mutations in genes other than *BRCA1/2*	114	39%
Ovarian breast cases with no mutation in HRR genes (WT)	213	32%
WES cases	51	57%
Prospective validation; consecutive solid tumors[a]	1312	6%
Breast	136	24%
Lung	307	1%
Colorectal	310	1%
Head and neck	20	0%
Ovarian	124	26%
Esophageal	31	0%
Pancreas	90	7%
Brain	145	0%
Other	149	1%

Abbreviations: BRCA1/2, breast cancer gene 1/2; HRD, homologous recombination deficiency; HRR, homologous recombination repair; WES, whole exome sequencing; WT, wild-type.
The HRD results are also shown.
[a]Of these, HRD was detected in 97% of 58 cases that were in breast, ovarian, pancreas, and prostate and had *BRCA1/2* mutations.

nucleotides were also sequenced. Genomic DNA samples were end repaired and A-tailed, and then added unique medical identifiers (UMIs) and sample index. Target enrichment is performed post-UMI assignment to ensure that DNA molecules containing UMIs are sufficiently enriched in the sequenced library. For enrichment, ligated DNA molecules were subjected to several cycles of targeted polymerase chain reaction (PCR) using 1 region-specific primer and 1 universal primer complementary to the adapter. A universal PCR was ultimately carried out to amplify the library and add platform-specific adapter sequences and additional sample indices. Normal tissue was sequenced and used for the CNVkit algorithm. The sequencing was conducted using the Illumina NovaSeq 6000 or NextSeq 550 instruments. The *BRCA1/2* genomic point mutations, deletions, or duplication alterations were defined as pathogenic, or variants of unknown significance mainly based on ClinVar database. Accuracy was confirmed by manual inspection of BAM files.

Using CNVkit for copy number detection

The CNVkit software was implemented to evaluate CNA in the analyzed samples.[18] Briefly, the software takes advantage of both on-target and off-target sequencing reads, compares binned read depths in on-target and off-target regions to pooled normal reference, and estimates the copy number at various resolutions.

Using machine learning model for classifying samples

The log2 of the normalized data of various segments (bins) of the 434 sequenced genes generated by CNVkit (total 26 940) was used in the ML approach for predicting the presence or absence of *BRCA1/BRCA2* mutations. This was achieved through 2 separate steps. In the first step, we selected the bins that distinguished between *BRCA1/2* mutated and unmutated and ranked them. In the second step, we used naïve Bayes with selected combination of bins to distinguish between the 2 groups.

To avoid overfitting, we applied a modified version of naïve Bayes (Geometric Mean Naïve Bayes [GMNB]). The conditional independence assumption of the naïve Bayes is selected because it is useful tool to resolve a very high-dimensional problem with a limited sample size. Estimating the correlations between bins would be counterproductive. The naïve Bayes approach has a small number of parameters and hence, a lower capacity as a learning system, which will help address the overfitting problem according to statistical learning theory. We used the GMNB method to address the numeric underflow issue of standard naïve Bayes when

applied to a high-dimensional problem. In GMNB, we applied the geometric mean to the conditional probabilities. The method is documented in a separate article.[25] We show that the geometric mean is essentially the only operation that will preserve the conditional independence of naïve Bayes and will not cause underflow.

In the selection and ranking of the specific bins, we used 2 criteria to facilitate effective and stable selection. The first criterion is a performance-based measure for selecting relevant bins capable of discriminating different classes. This method uses cross-validation to obtain a realistic performance measure. The second criterion is a stability measure based on statistical significance tests (2-tailed *t*-test significance) to ensure the robustness and stability of the selected bin.

The first criterion is a direct performance measure by the cross-validation errors

$$d = \sum_{c=1}^{k} 1 - \frac{error_c}{n_c}$$

The cross-validation errors are from the individual genes independently. Because it is a 1-variable classifier, almost any simple classifier would give similar results.

The second measure is a *P* value obtained with analysis of variance (ANOVA; or equivalently *t*-test for 2-class problems). For a data set with sample size *n* and *k* classes, the ANOVA coefficient is defined as

$$F = \frac{MSB}{MSW}$$

where MSB is the mean sum of squares between the groups, and MSW is the mean sum of squares within the groups. The ANOVA coefficient F follows the *F*-distribution with degrees of freedom $k-1, n-k$. The *P* value can be obtained from the F statistic. The minimum threshold for d/k (where *k* is the number of classes) would be 0.5. If $d/k \leqslant 0.5$, then there would be no evidence that the bin has any power in distinguishing the classes. Such a bin should be eliminated. We used a canonical 5% *t*-test *P* value in this process.

The selected and ranked bins using the above 2 criteria were then used to distinguishing between *BRCA1/2*-positive and *BRCA1/2*-negative cases with *k*-fold cross-validation procedure (with *k* = 12). The *k*-fold was applied during the bin selection stage only. A naïve Bayesian classifier was constructed on the training of *k* – 1 subsets and tested on the other testing subset. We applied GMNB as the classifier to predict specific class. The classifier (GMNB) over the filtered bins is not used during the filtering process. The classifier used in the first measure for filtering is a trivial one with only 1 input variable. This process does not repeat iteratively.

The training and testing subsets are then rotated, and the average of the classification errors is used to measure the relevancy of the bin. The classification system is trained with the selected subset of most relevant segments of the genes. The processes of bins selection and cancer classification are applied iteratively to obtain an optimal classification system and a subset of bins relevant to the specific class. The code for the ML will be available on direct request to the corresponding author.

Whole exome sequencing and LOH calculation

For the exome library preparation, 50 ng of DNA of each sample was used with Nextera Rapid Capture Expanded Exome (Illumina), according to the manufacturer's recommendations. Quantified DNA library was loaded on flow cell for subsequent cluster generation. Samples were paired-end sequenced on Illumina NextSeq 500 High-Output Kit—300 cycles (Illumina). Copy number variation (CNV) and LOH percentage were calculated from the CNVkit segmented format (.cns) data and the GATK SNP format variant call format (VCF). Each segment has a weighted mean log2 value that is then used to calculate percent-LOH. As each segment is of a different size, the weights differ and thus this affects the total calculation. Percent-LOH is calculated as the sum of all segment weights with a detected LOH divided by the sum of all segment weights. This provides a simple method of assessing LOH in the samples.

Results
High sensitivity and specificity in predicting the presence of BRCA1/2 mutations

Using log2 normalized copy number of the 26 940 segments (bins) of the 434 sequenced genes in ML model, we explored the potential of distinguishing between *BRCA1/2* mutated samples and *BRCA1/2* unmutated cases. We used 31 cases with confirmed *BRCA1/2* mutation and 84 cases confirmed negative for mutations in *BRCA1/2* or any of the genes implicated in HRR. These control cases had the expected CAN that are typically seen in solid tumors without any specific selection. The automated ML system that we developed selected 15 000 markers (bins) (Supplemental Table 1) for best separation between *BRCA1/2*-positive and *BRCA1/2*-negative cases. The receiver-operating characteristic curve showed AUC (area under the curve) of 0.984 (Figure 1). Using smaller numbers of markers showed significantly less prediction, although using a larger number of markers did not change the prediction significantly. Based on using a cut-off point of the ML of 0.486, the sensitivity was 90% and specificity was 98% (Table 2). The selected cut-off point emphasizes specificity over sensitivity. However, the cut-off point can be changed if clinical trials with outcome data suggest a better optimal cut-off point. The actual log2 copy ratio by CNVkit of 1-positive and 1-negative example is shown in Figure 2.
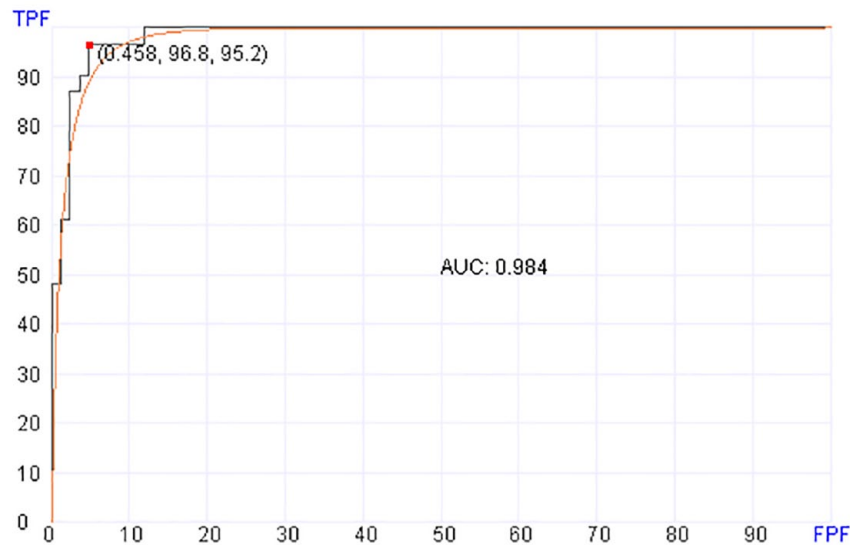
Figure 1. Receiver-operating characteristic (ROC) curve for prediction of HRD in samples with *BRCA1/2* mutations. The area under the curve (AUC) of 0.984 is obtained using 31 samples with pathogenic *BRCA1/2* mutations and 84 samples with no evidence of mutations in any of the HRR genes. BRCA1/2 indicates breast cancer gene 1/2; HRD, homologous recombination deficiency; FPF, false-positive fraction (specificity); TPF, true-positive fraction (sensitivity).

Table 2. Sensitivity and specificity of detecting the presence of HRD associated with *BRCA1/2* mutations using copy number alteration (CNA) and machine learning algorithm.

		95% CONFIDENCE INTERVAL
Sensitivity	97%	81.5%-99.8%
Specificity	95%	87.6%-98.5%
PPV	88%	71.6%-96.2%
NPV	99%	92.4%-99.9%

Abbreviations: BRCA1/2, breast cancer gene 1/2; HRD, homologous recombination deficiency; NPV, negative predictive value; PPV, positive predictive value.

Predicting homologous recombination deficiency in cancers with mutations in various homologous recombination repair genes as compared with wild–type cancers

To explore the value of the developed ML model after training to predict *BRCA1/2* positive tumors, we tested 124 ovarian/breast cancers without mutations in any of the HRR genes and 114 cancers with mutations in one of the genes involved in HRR. These HRR genes-positive cases included cancers with mutations in *ATM* (N = 36), *ATR* (N = 17), *CDK12* (N = 14), Fanconi anemia genes (N = 16), *NBN* (N = 12), *RAD50* (N = 9), *RAD51B* (N = 1), *RAD51* (N = 6), and *RAD54L* (N = 4). The ML classified 44 of the 114 samples (39%) as having structural genomic abnormalities similar to those detected in *BRCA1/2*-positive cases, implying high positivity for HRD. These HRD-positive cases had mutations in *ATM* (N = 13), *CDK12* (N = 5), Fanconi genes (N = 6), *ATR* (N = 26), *NBN* (N = 6), *RAD51* (N = 3), *RAD54L* (N = 1), and *RAD50* (N = 4). All these genes

have been reported to be associated with HRD phenotype.[16] Testing 213 random cancer cases without HRR mutations showed 68 cancer (32%) positive scores similar to those seen in *BRCA1/2* cases (Figure 3). There was significant difference in score between Mut-positive cases with genes other than *BRCA1/2* and *BRCA1/2*-positive cases ($P < .0001$) (Kruskal-Wallis ANOVA test), but there was no difference between Mut-positive cases and WT cancers ($P = .47$) (Kruskal-Wallis ANOVA test). This suggests that cancers with mutations in HRR genes other than *BRCA1/2* are heterogeneous, and overall more similar to WT cases than to *BRCA1/2*-mutant cancers.

Correlation with loss of heterozygosity

To correlate with other methods used in predicting HRD, we compared the HRD prediction using our CNV/ML method with LOH data resulting from WES in 51 selected tumor samples. As the LOH is calculated from WES rather than whole genome, we used a 9% LOH as cut-off between positive and negative LOH.[20] The LOH positive cases (>9%) were 29 and negative cases (LOH ⩽ 9%) were 22. Of the positive cases, 26 (90%) (95% confidence interval [CI] = 72%-97%) were also positive by our CNV/ML method, and of the negative cases, 16 (73%) (95% CI = 50%-88%) were negative.

Validation using independent set of samples

Prospective testing of independent set of solid tumors from various types including (Table 1) from lung, melanoma, colo-rectal, ovary, brain, breast head and neck, and others showed detection of HRD in 81 out of 1312 samples (6%). Germline and somatic *BRCA1/2* mutations were detected 117 (9%) of these cases. Of these *BRCA1/2* mutant tumors but HRD
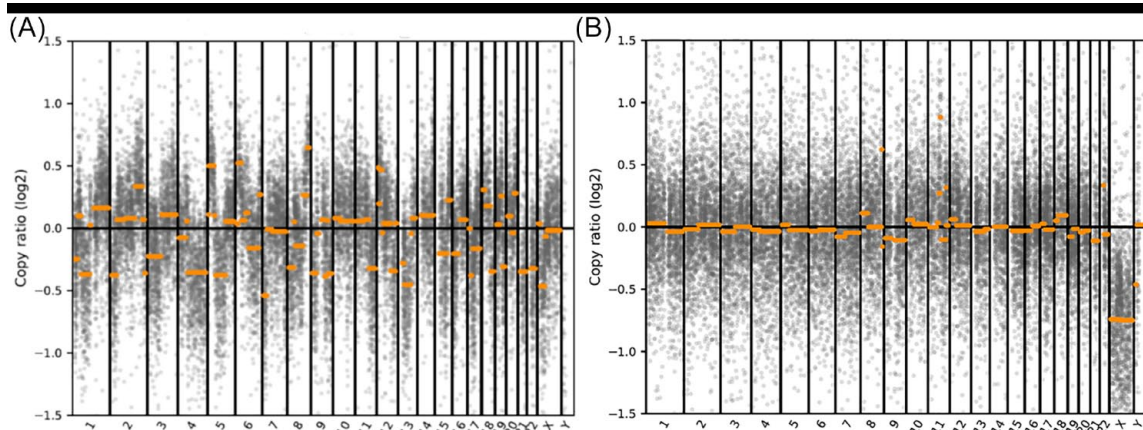
Figure 2. Representative example plots of log2 copy ratio of sequence segments (bins) in 1 example of breast cancer with BRCA mutation (A) and breast cancer without BRCA mutation (B).
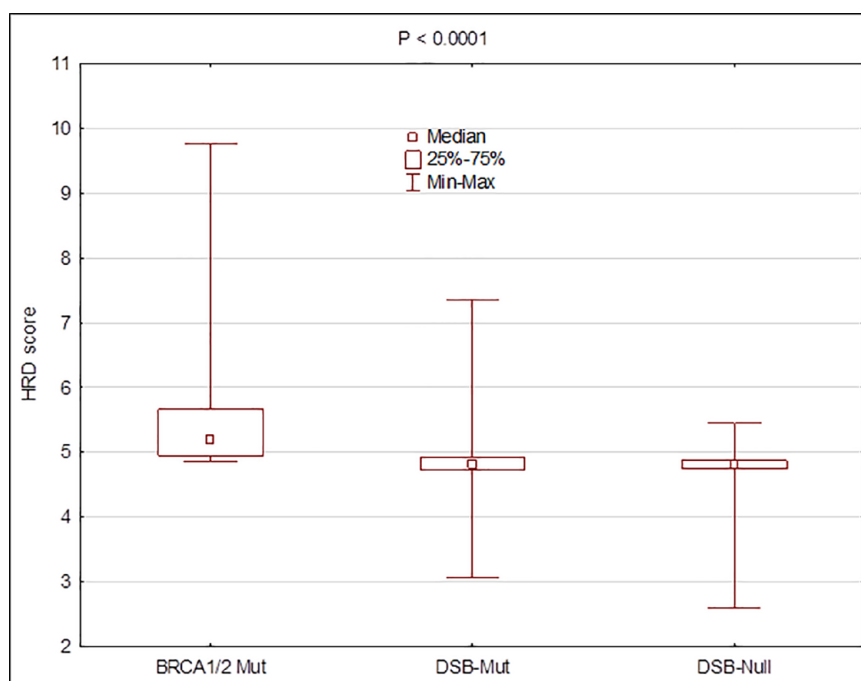BRCA indicates breast cancer gene.

Figure 3. Box plots showing the HRD scores obtained by machine learning for samples with *BRCA1/2* mutations (No. 31), with HRR mutations other than *BRCA1/2* (No. 84), and with no mutation in any HRR genes (WT) (No. 114). There was significant difference ($P < .0001$) (Kruskal-Wallis ANOVA test) between *BRCA1/2* mutant and other HRR-mutant group as well as between HRR-mutant (other than *BRCA1/2*) and the WT group. The score is based on the ROC curve shown in Figure 1.
ANOVA indicates analysis of variance; BRCA1/2, breast cancer gene 1/2; HRD, homologous recombination deficiency; HRR, homologous recombination repair; ROC, receiver-operating characteristic; WT, wild-type.

negative, 59 were from tumors other than ovarian, breast, pancreas, or prostate. This suggests that the HRD phenotype is not necessarily associated with mutations in *BRCA1/2* in tumors other than breast, ovarian, and pancreas. Only 2 cases of the 58 (3%) with *BRCA1/2* mutations from ovarian, breast, pancreas, and prostate cancers showed negative HRD. One of these 2 cases had a somatic BRCA1 mutation that is detected in subclone at variant allele frequency of 7%. The second case had a missense mutation (S2670L) in BRCA2 that is likely not pathogenic. Of the remaining negative cases for *BRCA1/2*

mutations, 25 (2%) cases had HRD positivity, most of which had mutations in genes other than *BRCA1/2* that are involved in double-strand DNA repair.

Discussion

Clinical studies have suggested that response to DSB-inducing agents is best in cancers that have genomic abnormalities dictated by *BRCA1/2* mutations.[1-7] These genomic abnormalities are typically manifested by structural chromosomal abnormalities resulting from HRD including CNVs,

translocations, and LOH. Tumors with *BRCA1/2* are considered the gold standard for these abnormalities. Tumors with abnormalities similar to those seen in cases with *BRCA1/2* are currently classified as eligible for treatment with DSB-inducer agents. The chromosomal structural abnormalities historically were measured using approaches involving evaluation of LOH, TAI, and LST.[14] The results of these measurements are combined giving a score that is used to determine which tumor might respond to DSB-inducer therapy. Multiple studies demonstrated that this approach is useful for selecting patients, but there is room to improve on current methods and more accurately select patients.

With the advances in NGS, chromosomal structural abnormalities can be measured and quantified. Kim et al[26] measured HRD using the same approach (LOH, TAI, and LST) using NGS and WES and demonstrated that HRD-high tumors had significantly ($P=.003$) higher pCR rates and higher near-pCR rates ($P=.049$) compared with those of the HRD-low tumors. High score was detected in tumors with germline mutations in HRR genes, but not in somatic mutations. Eeckhoutte et al[27] used a shallow WGS for measuring large-scale genomic alteration and predicting HRD. Whole-genome sequencing and mutation profile were also used by Davies et al[28] for detecting HRD using a supervised lasso logistic regression model.

Here, we present data for predicting HRD based on using routinely generated sequencing data from targeted molecular profiling used for the detection of various clinically relevant mutations in solid tumors, measuring tumor mutation burden and microsatellite instability (MSI). The assay is specifically designed to be cost-effective and amenable for adaptation in routine clinical laboratories. The panel included targeted coding regions of 434 genes. We used the normalized log2 ratio of sequenced fragments (bins) generated by CNVkit software in an ML model to develop a model for classifying tumors with *BRCA1/2* mutations vs tumors without *BRCA1/2* mutations. Practically, this approach quantifies gains and losses of various DNA fragments in the genomic areas covered by these 434 genes, and then uses an ML model for classifying cases. As shown in Figure 1, this approach allows us to distinguish between *BRCA1/2*-positive and *BRCA1/2*-negative cases with high sensitivity and specificity (AUC = 98.4%). Accuracy (sensitivity and specificity) can be estimated from the receiver-operating characteristic (ROC) curve. We selected a cut-off that provided sensitivity of 90% and specificity of 98%. The demonstration that CNA abnormalities are adequate to distinguish tumors with *BRCA1/2* mutations from *BRCA1/2*-null cases indicates that CNV changes when used in ML model reflect the biology that drives the specific neoplastic process drive by *BRCA1/2* mutations. The use of highly sophisticated ML model for first selecting the proper changes markers (bins) and then using these markers for comparing with the findings in typical *BRCA1/2*-positive cases in automated fashion is crucial for the success of this approach. However, this approach

success depends on the tumor fraction in the analyzed samples. Only samples with 30% or greater tumor fraction are accepted for such an approach.

We show that 39% of cases with mutations in HRR genes other than *BRCA1/2* can be classified as having genomic structural abnormalities similar to those seen in *BRCA1/2*-positive cases, whereas 32% of WT ovarian or breast cancers had HRD similar to that seen in the *BRCA1/2* cases. This confirms that HRR genes vary in the level of deficiency in homologous recombination they cause. As we use HRD caused by BRCA1/2 as the gold standard, it is expected that any scar caused by other HRR genes that does not meet the *BRCA1/2* standard level will be considered negative for HRD. On the contrary, HRD can be seen in some cancers possibly due to multigene effects, and these cases can be HRD positive despite lack of mutations in specific HRR genes. There was significant difference in the HRD score ($P<.0001$) (Kruskal-Wallis ANOVA test) between *BRCA1/2*-positive cancers and cases with mutations in HRR genes other BRCA1/2 as well as WT cases. There was no significant difference ($P=.47$) (Kruskal-Wallis ANOVA test) between Mut-positive and WT cases (Figure 3). This suggests that mutations in HRR genes other than *BRCA1/2* are associated with increased tendency to have HRD, but HRD can also be seen in WT cases. This may explain the reason that clinical trials failed to show significant improvement in outcome in cancers with mutations in HRR genes other than *BRCA1/2* when treated with PARPi.[29] Testing random consecutive cancer samples that included lung, colon, brain, sarcoma, and others in addition to ovarian and breast showed HRD only in only 6% of cases. The low number of cases with HRD in this group of samples most likely reflects the high stringency (specificity) of our approach in predicting HRD.

Using LOH for comparison with our methodology for detecting HRD, we show 90% concordance in detecting HRD-positive cases and 73% concordance with HRD-negative cases. The high concordance with HRD-positive cases suggests similar sensitivity. The relatively less robust concordance with HRD-negative cases is likely due difference in the approaches. Homologous recombination deficiency testing based on LOH may not be the gold standard for predicting HRD. Loss of heterozygosity has not shown high level of prediction HRD and response to PARPi.[22] Irrespective, clinical trials with clinical outcome data are needed to better understand the biological reasons for such difference and for determining which approach is more accurate in predicting response to PARPi. We speculate that by focusing on predicting similarity to *BRCA1/2* biology, our approach is more accurate in predicting HRD than LOH scoring system which uses an arbitrary cut-off point to distinguish between HRD-positive and HRD-negative cases.[22]

Methylation of *BRCA1/2* has been reported in significant number of breast and ovarian cancers, and this methylation is reported to be associated with positive HRD.[30-33] This may explain that we detected HRD in cases without mutations. The

32% HRD positivity we detected in our testing of WT ovarian and breast cancer cases is likely driven in part by BRCA1/2 methylation. The high specificity of our ML in detecting *BRCA1/2* mutations may overcome the problem of basing eligibility for PARPi based on methylation alone. Our ML detects HRD solely based on detecting genomic scars that are strictly caused by *BRCA1/2* mutation and may be more robust than the conventional methods which rely on using a specific cut-off for detecting genomic scars. The clinical relevance of promoter methylation remains controversial, and it has been reported that cases with promoter methylation are able to adaptively lose methylation.[31]

In summary, using our approach, cancers can be classified into 2 groups: (1) high HRD score including *BRCA1/2*-positive and cases with score similar to *BRCA1/2*-positive irrespective if they have mutations in HRR genes or not and (2) negative HRD score including cancers with HRD score lower than that seen in cases with *BRCA1/2* genes. This suggests that high-score cancers can be considered eligible for treatment with DSB-inducing drugs. This assumption requires confirmation by clinical studies. However, the demonstrated high sensitivity and specificity of this approach in predicting *BRCA1/2*-associated genomic abnormalities suggests that this approach has high potential for predicting response to PARPi. Furthermore, this approach can predict increased susceptibility to homologous recombination due to causes other than mutations, such as methylation, deletion in HRR genes, multigenic factors, and others. The weakness of our study is lack of clinical correlation with treatment and outcome. Head-to-head comparison of our approach with other approaches in detecting HRD and comparing the prediction of HRD of each methodology with clinical outcome may advance this field significantly.

Declarations

Ethics Approval And Consent To Participate

The work was performed with approved WCG-IRB (Western IRB [WIRB], Copernicus Group IRB [CGIRB], Midlands IRB [MLIRB], New England IRB [NEIRB], and Aspire IRB) waiver for consent by Western Institutional Review Board (WCG IRB # 1-1476184-1).

Consent For Publication

This manuscript does not include any individual person's data.

Author Contributions

Maher Albitar: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Writing—original draft.
Hong Zhang: Conceptualization; Formal analysis; Software; Validation; Writing—review & editing.
Andrew Pecora: Data curation; Investigation; Resources; Writing—review & editing.
Stanley Waintraub: Conceptualization; Investigation; Resources; Writing—review & editing.
Deena Graham: Conceptualization; Data curation; Resources; Writing—review & editing.
Mira Hellmann: Conceptualization; Investigation; Resources; Writing—review & editing.
Donna McNamara: Conceptualization; Investigation; Resources; Writing—review & editing.
Ahmad Charifa: Conceptualization; Data curation; Formal analysis; Investigation; Resources; Writing—review & editing.
Ivan De Dios: Data curation; Investigation; Methodology; Writing—review & editing.
Wanlong Ma: Data curation; Methodology; Writing—review & editing.
Andre Goy: Conceptualization; Investigation; Resources; Writing—review & editing.

Availability of Data and Materials

The data sets used and/or analyzed during this study are available from the corresponding author.

SUPPLEMENTAL MATERIAL

Supplemental material for this article is available online.

REFERENCES

1. Tarsounas M, Sung P. The antitumorigenic roles of BRCA1-BARD1 in DNA repair and replication. *Nat Rev Mol Cell Biol*. 2020;21:284-299. doi:10.1038/s41580-020-0218-z.
2. Sharma R, Lewis S, Wlodarski MW. DNA repair syndromes and cancer: insights into genetics and phenotype patterns. *Front Pediatr*. 2020;8:570084. doi:10.3389/fped.2020.570084.
3. Jensen RB, Rothenberg E. Preserving genome integrity in human cells via DNA double-strand break repair. *Mol Biol Cell*. 2020;31:859-865. doi:10.1091/mbc.E18-10-0668.
4. Iliakis G, Wang H, Perrault AR, et al. Mechanisms of DNA double strand break repair and chromosome aberration formation. *Cytogenet Genome Res*. 2004;104:14-20. doi:10.1159/000077461.
5. Hodgson DR, Dougherty BA, Lai Z, et al. Candidate biomarkers of PARP inhibitor sensitivity in ovarian cancer beyond the BRCA genes. *Br J Cancer*. 2018;119:1401-1409. doi:10.1038/s41416-018-0274-8.
6. Ray Chaudhuri A, Nussenzweig A. The multifaceted roles of PARP1 in DNA repair and chromatin remodelling. *Nat Rev Mol Cell Biol*. 2017;18:610-621. doi:10.1038/nrm.2017.53.
7. Abkevich V, Timms KM, Hennessy BT, et al. Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. *Br J Cancer*. 2012;107:1776-1782. doi:10.1038/bjc.2012.451.
8. Kadouri L, Rottenberg Y, Zick A, et al. Homologous recombination in lung cancer, germline and somatic mutations, clinical and phenotype characterization. *Lung Cancer*. 2019;137:48-51. doi:10.1016/j.lungcan.2019.09.008.
9. Foote JR, Secord AA, Liang MI, et al. Targeted composite value-based endpoints in platinum-sensitive recurrent ovarian cancer. *Gynecol Oncol*. 2019;152:445-451. doi:10.1016/j.ygyno.2018.11.028.
10. Li Y, Zhang X, Gao Y, et al. Development of a genomic signatures-based predictor of initial platinum-resistance in advanced high-grade serous ovarian cancer patients. *Front Oncol*. 2021;10:625866. doi:10.3389/fonc.2020.625866.
11. Da Costa AABA, Do Canto LM, Larsen SJ, et al. Genomic profiling in ovarian cancer retreated with platinum based chemotherapy presented homologous recombination deficiency and copy number imbalances of CCNE1 and RB1 genes. *BMC Cancer*. 2019;19:422. doi:10.1186/s12885-019-5622-4.
12. Ledermann JA, Drew Y, Kristeleit RS. Homologous recombination deficiency and ovarian cancer. *Eur J Cancer*. 2016;60:49-58. doi:10.1016/j.ejca.2016.03.005.
13. Wong W, Raufi AG, Safyan RA, Bates SE, Manji GA. BRCA mutations in pancreas cancer: spectrum, current management, challenges and future prospects. *Cancer Manag Res*. 2020;12:2731-2742. doi:10.2147/CMAR.S211151.

14. Pacheco-Barcia V, Muñoz A, Castro E, et al. The homologous recombination deficiency scar in advanced cancer: agnostic targeting of damaged DNA repair. *Cancers*. 2022;14:2950. doi:10.3390/cancers14122950.

15. Moschetta M, George A, Kaye SB, Banerjee S. BRCA somatic mutations and epigenetic BRCA modifications in serous ovarian cancer. *Ann Oncol*. 2016;27:1449-1455. doi:10.1093/annonc/mdw142.

16. Chelariu-Raicu A, Coleman RL. Breast cancer (BRCA) gene testing in ovarian cancer. *Chin Clin Oncol*. 2020;9:63. doi:10.21037/cco-20-4.

17. Coleman RL, Oza AM, Lorusso D, et al. Rucaparib maintenance treatment for recurrent ovarian carcinoma after response to platinum therapy (ARIEL3): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet*. 2017;390:1949-1961. doi:10.1016/S0140-6736(17)32440-6.

18. Stronach EA, Paul J, Timms KM, et al. Biomarker assessment of HR deficiency, tumor BRCA1/2 mutations, and CCNE1 copy number in ovarian cancer: associations with clinical outcome following platinum monotherapy. *Mol Cancer Res*. 2018;16:1103-1111. doi:10.1158/1541-7786.MCR-18-0034.

19. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput Biol*. 2016;12:e1004873. doi:10.1371/journal.pcbi.1004873.

20. Stover EH, Fuh K, Konstantinopoulos PA, Matulonis UA, Liu JF. Clinical assays for assessment of homologous recombination DNA repair deficiency. *Gynecol Oncol*. 2020;159:887-898. doi:10.1016/j.ygyno.2020.09.029.

21. Telli ML, Timms KM, Reid J, et al. Homologous recombination deficiency (HRD) score predicts response to platinum-containing neoadjuvant chemotherapy in patients with triple-negative breast cancer. *Clin Cancer Res*. 2016;22:3764-3773. doi:10.1158/1078-0432.CCR-15-2477.

22. Sharma P, Barlow WE, Godwin AK, et al. Impact of homologous recombination deficiency biomarkers on outcomes in patients with triple-negative breast cancer treated with adjuvant doxorubicin and cyclophosphamide (SWOG S9313). *Ann Oncol*. 2018;29:654-660. doi:10.1093/annonc/mdx821.

23. Washington CR, Moore KN. PARP inhibitors in the treatment of ovarian cancer: a review. *Curr Opin Obstet Gynecol*. 2021;33:1-6. doi:10.1097/GCO.0000000000000675.

24. Li W, Gao L, Yi X, et al. Patient assessment and therapy planning based on homologous recombination repair deficiency [published online ahead of print February 13, 2023]. *Genomics Proteomics Bioinformatics*. doi:10.1016/j.gpb.2023.02.004.

25. Albitar M, Zhang H, Goy A, et al. Determining clinical course of diffuse large B-cell lymphoma using targeted transcriptome and machine learning algorithms. *Blood Cancer J*. 2022;12:25. doi:10.1038/s41408-022-00617-5.

26. Kim SJ, Sota Y, Naoi Y, et al. Determining homologous recombination deficiency scores with whole exome sequencing and their association with responses to neoadjuvant chemotherapy in breast cancer. *Transl Oncol*. 2021;14:100986. doi:10.1016/j.tranon.2020.100986.

27. Eeckhoutte A, Houy A, Manié E, et al. ShallowHRD: detection of homologous recombination deficiency from shallow whole genome sequencing. *Bioinformatics*. 2020;36:3888-3889. doi:10.1093/bioinformatics/btaa261.

28. Davies H, Glodzik D, Morganella S, et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat Med*. 2017;23:517-525. doi:10.1038/nm.4292.

29. Ladan MM, van Gent DC, Jager A. Homologous recombination deficiency testing for BRCA-like tumors: the road to clinical validation. *Cancers*. 2021;13:1004. doi:10.3390/cancers13051004.

30. Tutt A, Tovey H, Cheang MCU, et al. Carboplatin in BRCA1/2-mutated and triple-negative breast cancer BRCAness subgroups: the TNT Trial. *Nat Med*. 2018;24:628-637. doi:10.1038/s41591-018-0009-7.

31. Bücker L, Lehmann L, Kumar P, et al. CDH1 (E-cadherin) gene methylation in human breast cancer: critical appraisal of a long and twisted story. *Cancers*. 2022;14:4377. doi:10.3390/cancers14184377.

32. van Wagensveld L, van Baal JOAM Timmermans M, et al. Homologous recombination deficiency and cyclin E1 amplification are correlated with immune cell infiltration and survival in high-grade serous ovarian cancer. *Cancers*. 2022;14:5965. doi:10.3390/cancers14235965.

33. Menghi F, Banda K, Kumar P, et al. Genomic and epigenomic BRCA alterations predict adaptive resistance and response to platinum-based therapy in patients with triple-negative breast and ovarian carcinomas. *Sci Transl Med*. 2022;14:eabn1926. doi:10.1126/scitranslmed.abn1926.