npj | precision oncology

Article

Published in partnership with The Hormel Institute, University of Minnesota

https://doi.org/10.1038/s41698-024-00706-7

Prediction of homologous recombination deficiency identifies colorectal tumors sensitive to PARP inhibition

Giorgio Corti[1,2], Kristi Buzo[1,2], Enrico Berrino[2,3], Martina Miotto[1], Maria Costanza Aquilano[4], Marilena Lentini[1,2], Sara Erika Bellomo[2], Annalisa Lorenzato[1], Alice Bartolini[2], Gianluca Mauri[4,5], Luca Lazzari[5], Mariangela Russo[1], Federica Di Nicolantonio[1,2], Salvatore Siena[4,6], Silvia Marsoni[5], Caterina Marchiò[2,3], Alberto Bardelli[1,5] ✉ & Sabrina Arena[1,2] ✉

The synthetic lethal effect observed with the use of PARP inhibitors (PARPi) with tumors characterized by the loss of key players in the homologous recombination (HR) pathway, commonly referred to as "BRCAness", is maintaining high interest in oncology. While BRCAness is a well-established feature in breast, ovarian, prostate, and pancreatic carcinomas, our recent findings indicate that up to 15% of colorectal cancers (CRC) also harbor defects in the HR pathway, presenting promising opportunities for innovative therapeutic strategies in CRC patients. We developed a new tool called *HRDirect*, which builds upon the HRDetect algorithm and is able to predict HR deficiency (HRD) from reference-free tumor samples. We validated HRDirect using matched breast cancer and CRC patient samples. Subsequently, we assessed its efficacy in predicting response to the PARP inhibitor olaparib by comparing it with two other commercial assays: AmoyDx HRD by Amoy Diagnostics and the TruSight Oncology 500 HRD (TSO500-HRD) panel by Illumina NGS technology. While all three approaches successfully identified the most PARPi-sensitive CRC models, HRDirect demonstrated superior precision in distinguishing resistant models compared to AmoyDX and TSO500-HRD, which exhibited overlapping scores between sensitive and resistant cells. Furthermore, we propose integrating HRDirect scoring with ATM and RAD51C immunohistochemical analysis as part of our "composite biomarker approach" to enhance the identification of HRD tumors, with an immediate translational and clinical impact for CRC personalized treatment.

Background

The DNA damage response (DDR) comprises an intricate network of cellular processes that operate to maintain genomic stability and to prevent the accumulation of DNA mutations, which could predispose to various diseases, including cancer[1]. A significant proportion of tumors exhibit deficiencies in proteins from one or more DDR repair families, rendering them reliant on intact DDR pathways for survival, with substantial implications at clinical and therapeutic levels[2]. Defects in the homologous recombination repair (HRR) pathway can lead to homologous recombination deficiency (HRD), triggering the activation of error-prone DNA repair mechanisms, thereby fostering genomic instability and tumorigenesis. A classic example of HRD is provided by alterations in the *BRCA1* and *BRCA2* genes, which are involved in the HRR pathway and play a crucial role in maintaining genomic integrity. Defects in *BRCA1* and *BRCA2* are associated with an increased risk of developing breast, ovarian, and other cancer types. Tumors harboring these alterations often exhibit heightened sensitivity to targeted therapies, such as poly-ADP-ribose polymerase inhibitors (PARPi), as

[1]Department of Oncology, University of Torino, Torino, Italy. [2]Candiolo Cancer Institute, FPO - IRCCS, Candiolo, Italy. [3]Department of Medical Sciences, University of Torino, Torino, Italy. [4]Department of Hematology, Oncology and Molecular Medicine, Grande Ospedale Metropolitano Niguarda, Milan, Italy. [5]IFOM ETS - The AIRC Institute of Molecular Oncology, Milan, Italy. [6]Department of Oncology and Hemato-Oncology, Università degli Studi di Milano, Milano, Italy. ✉e-mail: alberto.bardelli@unito.it; sabrina.arena@unito.it

initially evidenced by seminal studies in 2005 demonstrating the synthetic lethality effect with PARPi in BRCA germline deficient cancers[3,4]. Subsequent clinical trials published in 2009–2010 underscored the profound impact of this exceptional discovery [5–7]. The clinical significance of HRD status in selecting patients with high-grade serous ovarian cancer or breast cancer for PARPi therapy has been well-documented, demonstrating substantial benefits in HRD-positive tumors compared to HR-proficient (HRP) ones[8], particularly in *BRCA* germline deficient and platinum-sensitive epithelial ovarian cancers.

The identification of clinically relevant DNA repair defects in tumors becomes indeed of primary importance to guide treatment decisions and to tailor personalized therapies for rationally selected patients.

It is noteworthy that HRD may also arise from alterations in other genes within the HRR pathway besides *BRCA* genes[9]. These defects, whether germline or somatic, could serve as potential biomarkers for PARPi sensitivity. Their genetic profiling is of crucial significance and expands to the stratification of patients potentially deriving benefits from FDA-approved PARPi-based therapy or acquiring enrollment eligibility criteria for clinical trials testing novel therapies.

HRD testing, encompassing analysis of mutations in specific genes such as *BRCA1/2* and genomic scars associated with HRD like loss of heterozygosity (LOH), telomeric allelic imbalance (TAI), and large-scale state transitions (LST), has been pivotal in developing numerous FDA-approved companion diagnostic tests, such as those by Myriad Genetics and Foundation Medicine[10]. Targeted gene profiling has been complemented by the emergence of genomic molecular signatures derived from comprehensive analyses of tumor genomes[11,12]. Some of these signatures have been associated with defects in DNA repair pathways, and particularly those emerging from *BRCA1/2* or HRR inactivity, which is predictive for response to PARPi or platinum-based treatment[11–13]. More recently, a novel algorithm named HRDetect[13] has been developed to initially classify breast cancer patients, then extended to ovarian tumors, as either HRD-positive or HRD-negative (i.e., HRP), leveraging various genomic features and mutational signatures associated with HRD.

Inspired by these breakthrough findings in breast and ovarian cancers, our recent focus has been on investigating whether PARPi-based treatment holds promise for patients with other tumors, particularly colorectal cancer (CRC). We have demonstrated that approximately 13% of CRC tumors respond to olaparib[14], possibly due to defects in HRR genes other than *BRCA1/2*, such as *RAD51C* and *ATM*[15]. However, to date most trials who assessed the potential role of olaparib in CRC patients were conducted mostly in molecularly unselected patients who did not significantly benefit from this approach[16,17]. Thus, while being a standard-of-care option in breast, ovarian, and prostate cancers, olaparib is not a standard for CRC patients' treatment. Despite these poor initial results, given the limited therapeutic options available for metastatic CRC, molecularly identifying CRC patients likely to benefit from PARPi, especially those resistant to immunotherapy or standard-of-care therapies, is of utmost importance.

Finally, since CRC is not a BRCA-related disease such as pancreatic, ovarian, breast, or prostate cancer, it is reasonable to hypothesize that different molecular scars should be looked for to identify patients potentially benefitting from this therapeutic strategy.

On these premises and considering that no samples from CRC olaparib-treated patients are currently available, we have leveraged our extensive collection of CRC preclinical models (cell lines and organoids)[14,18,19] to develop and validate a novel approach called HRDirect for identifying and predicting CRCs potentially responsive to olaparib treatment. In addition, HRDirect offers the advantage of calculating the HRD score even in the absence of reference germline genomic DNA (gDNA), which is often unavailable in clinical settings. We have compared this approach with other tools such as HRDetect[13] and other commercially available but not FDA-approved tests to provide an effective approach to aid treatment decision-making, offering more personalized and efficient therapies for CRC patients with HRD-positive tumors.

Results

Development and validation of the in-house HRDetect pipeline for breast cancer samples

We initially considered the seminal work by Davies and colleagues[13] to design an "in house" pipeline (Fig. 1, left side) that, by exploiting whole genome sequencing (WGS), is able to classify patients as HR proficient or deficient (hereafter referred to as HRP and HRD, respectively). The initial phase of this workflow entails variant calling, encompassing single nucleotide variations (SNVs), insertions, deletions (indels), and structural variants, achieved through the Cancer Genome Project for WGS pipeline ("cgpwgs") developed by the Wellcome Sanger Institute.

This mutational analysis specifically targets somatic variations, necessitating both tumor and healthy reference samples.

Subsequently, the pipeline employs the "HRDetect_pipeline" function within the "signature.tools.lib" R library [20] to predict HRD classification, involving the conversion and filtering of cgpwgs result files to meet the function's specifications.

The HRDetect algorithm yields a score between 0 and 1: samples with a score <0.25 are classified as HR Proficient ("HRP"), those >0.75 as HR Deficient ("HRD"), while scores between 0.25 and 0.75 result in no classification ("NC").

To validate our workflow, we utilized alignment files in BAM format[21] and related predictions from 76 patients in Davies' publication. Our analysis confirmed the reliability of our pipeline (Supplementary Table 1, sensitivity: 0.950, specificity: 1.000).

Recognizing the limitations of mapping data to the dated human genome reference (hg19) without alternative chromosomes and utilizing an outdated aligner function (bwa aln, v.0.5.9), we endeavored to enhance results by remapping data to hg38 using the latest aligner version (bwa mem, v0.7.17) (Li, 2013, https://arxiv.org/abs/1303.3997). Subsequent reclassification of all patients revealed only 4 misclassifications out of 76. Remarkably, of these 4 misclassifications, 3 patients were unclassified ("NC") by Davies and colleagues, resulting in a single misclassification out of 73 patients (Supplementary Table 2; sensitivity: 1.000, specificity: 0.958).

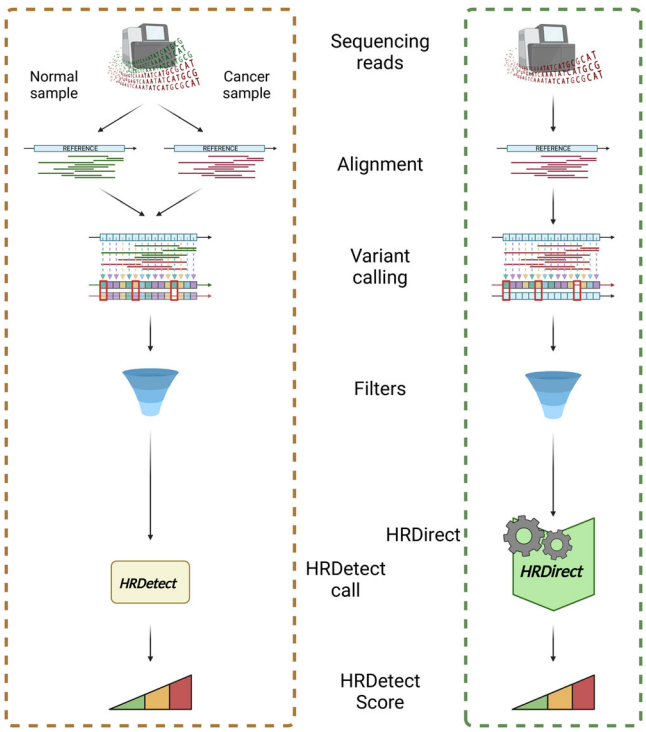


Fig. 1 | HRDetect and HRDirect workflow. Schematic representation of the matched analysis by HRDetect (left, gold dashed box) and the unmatched one by HRDirect using the meta-normal data (right, green dashed box).

Development and validation of the HRDirect pipeline in breast cancer samples

The absence of normal matched samples represents a common challenge in both clinical and preclinical settings, particularly notable in immortalized cell lines lacking paired samples. To address this challenge and concurrently reduce sequencing costs, we refined the HRDetect workflow, introducing an unmatched setting that we termed "HRDirect" (Fig. 1, right side).

Given the necessity of both healthy and tumor data for cgpwgs execution, we generated meta-normal data (see Methods) to substitute for the matched normal sample in the HRDirect workflow.

However, as expected, this substitution gave wrong results when the classification was "HRD" (Supplementary Table 3): only 20% (5 over 25 samples) of the patients were correctly classified. On the contrary, for HRP samples, where there are no genomic scars to be detected, the meta-normal gave perfect classification (48 over 48 patients). We reasoned that some strategy to enrich somatic variants is needed.

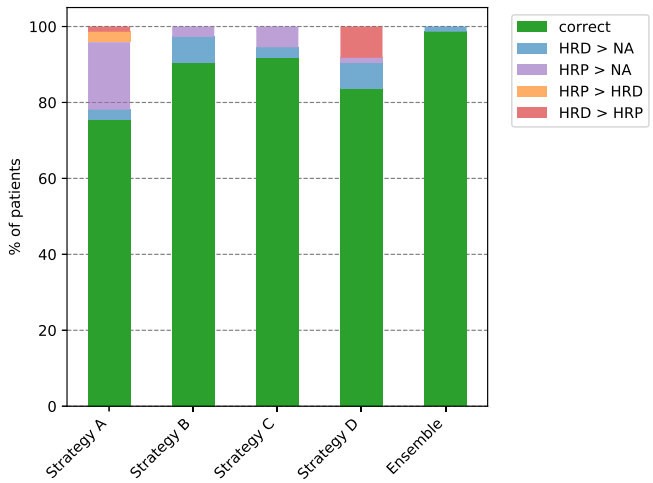


Fig. 2 | Evaluation of prediction accuracy for each strategy. Accuracy was evaluated on the dataset of breast cancer samples ($N = 76$). The outcome of each strategy is divided into the correct (green) and inaccurate classification: the error type is annotated in different colors. "HRD > HRP": expected HRD but classified HRP, "HRP > HRD": expected HRP but classified HRD, "HRP > NC": expected HRP but not classified, "HRD > NC": expected HRP but not classified.

The LASSO logistic regression model used in HRDetect relies on defined properties (features) and a corresponding weight; SNVs and indels features are prevalent and we mainly focused on them. Analysis of variant allele frequencies (VAFs) revealed that somatic mutations are enriched towards lower frequencies in matched comparisons (Supplementary Figs. 1 and 2).

Based on these findings, we developed several approaches to mitigate unwanted germline variations (Supplementary Fig. 3A–D). The initial two strategies, denoted "A" and "B", retained SNVs and indels within specific frequency windows, yielding improvements but with classification errors persisting above 5% (see Fig. 2). Conversely, "strategy C" aimed to replicate the entire desired distribution by sampling variations based on their expected allele frequency occurrence, enhancing enrichment for variants near the peak of the desired distribution (see Supplementary Figs. 1 and 2).

All these approaches are generic and could not be able to catch some very particular patients whose variants' frequencies do not accumulate at the peak of the general distribution.

Recognizing the diversity of patient variants, we proposed "strategy D", which attempts to delineate somatic distribution by considering that germline distribution is typically bell-shaped and nearly symmetrical. This approach identifies somatic variations as those exceeding the bell-shaped portion of the total distribution (see Methods).

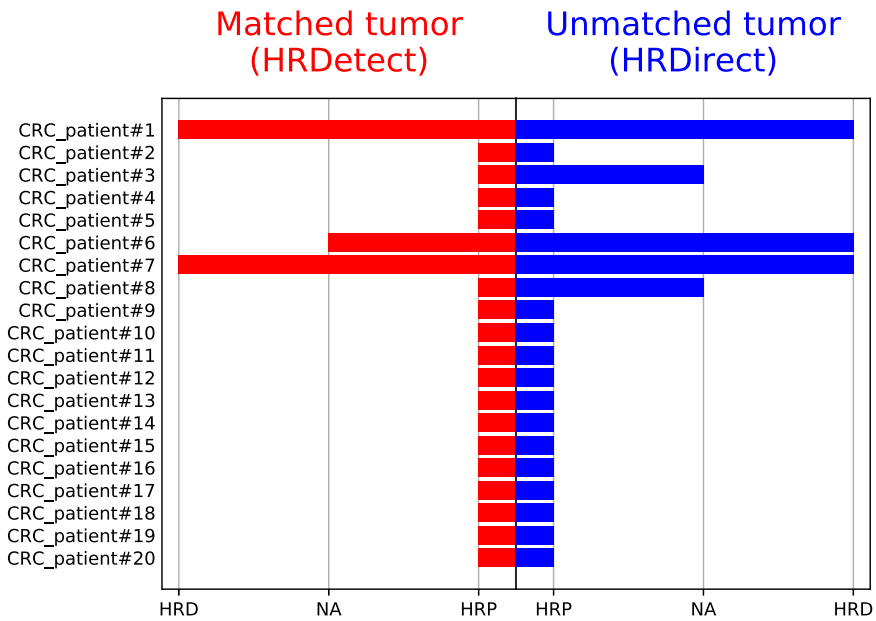
These strategies were integrated into an intermediate step between variant calling and the final scoring. While none of these approaches individually achieved accurate patient HR status predictions within a small margin of error (Fig. 2), combining results using a majority vote rule, termed the "ensemble" strategy, proved highly effective. This approach successfully captured the heterogeneity across the breast cancer patient dataset used, as published by Davies and colleagues, resulting in correct classification of 99% of samples (Fig. 2).

Each strategy is applied to the data and HRDetect is called, and then all the resulting scores are merged in the ensemble procedure to get the final classification: we call this whole step "HRDirect" (Fig. 1, right side).

Testing HRDetect and HRDirect in patient-derived colorectal cancer organoids

Having confirmed the efficacy of our unmatched method on Davies' breast cancer dataset, we turned our focus to our extensive collection of preclinical CRC samples, as CRC is our specific tumor type of interest, to assess potential tissue-specific considerations for the algorithm.

Fig. 3 | Predictions by HRDetect and HRDirect on CRC organoids. On the left, the matched analysis by HRDetect on 20 samples; on the right, the same analysis but with HRDirect (using the meta-normal data instead of the matched normal sample).



Given the current unavailability of fresh tissue from CRC patients treated with PARPi, we utilized a large collection of established patient-derived organoids (PDOs) as surrogates, since PDOs closely mimic the complex structure and genotype of the corresponding patient's tumor of origin. We selected 20 PDOs with available matched normal and tumor genomic DNA.

Following Whole Genome Sequencing on all matched samples, our HRDetect pipeline predicted BRCAness, identifying two HRD (10%; CRC_patient #1 and #7), one undefined (CRC_patient #6), and seventeen HRP (85%) samples (Fig. 3, left red bars). Despite the limited patient sample size, the observed percentage of HRD-classified CRC organoids aligns with findings from other studies investigating germline or somatic genetic defects in HR repair genes[14,22–24].

Remarkably, when applying the HRDirect workflow to the same samples using meta-normal data instead of matched normals, we observed high concordance with matched predictions. Specifically, the two HRD organoids identified by HRDetect (CRC patient #1 and #7) were correctly classified by HRDirect, while CRC_patient #6 (previously undefined by HRDetect) was classified as HR deficient by HRDirect (Fig. 3).

To ascertain whether HRD prediction correlates with sensitivity to PARP inhibitors, we selected for olaparib treatment nine nicely growing CRC organoids, whose histology resembled the patient's tumor they were derived from (Fig. 4).

Notably, all three HRD-predicted organoids (patient #1, #6, and #7) exhibited sensitivity to PARP blockade (Fig. 5). Previous data suggested that patient #1 might be sensitive to olaparib due to the absence of RAD51C expression, a RAD51 paralog involved in HR-mediated DNA repair[15]. Leveraging our "composite biomarker" approach we previously defined[15], we assessed the expression of ATM and RAD51C in CRC patient #6 and #7 via immunohistochemistry (Supplementary Fig. 4). While CRC patient #7 showed very low expression of both biomarkers (Supplementary Fig. 4A), as also confirmed by western blot analysis (Supplementary Fig. 4B), patient #6 exhibited normal expression. Of note, when considering the putative foci-like staining pattern in the nucleus, this was hardly visible, likely due to the untreated status of the preclinical models and their assumed HRD phenotype. Interestingly, CRC patient #3 and #8, with undefined ("NC") HRDirect scores, demonstrated modest sensitivity to olaparib, suggesting the involvement of other partially penetrant HRD mechanisms influencing PARPi sensitivity to a lesser extent.

Finally, also CRC patient #2 resulted responsive to olaparib, despite being predicted HR proficient by both HRDetect and HRDirect tests, implying involvement of mechanisms not strictly HRD-related but possibly related to replication stress (RS) sensitivity[15] (Supplementary Fig. 5) or not captured by the assays. To better understand whether CRC patient #2 could be more sensitive to RS, we performed analysis on ATM and DNA-PK, as HR and NHEJ key players, RPA as a protein involved in replication stress response, and H2AX, a biomarker of DNA damage upon treatment with either ceralasertib (an ATR inhibitor) or hydroxyurea (HU, a potent RS inducer). We observed that in patient #2, drug treatment was able to activate both ATM and DNA-PK, suggesting a higher sensitivity to RS in this patient with respect to patient #5. Consistently with our previous publication[15], we found that basal levels of pRPA were lower in RS-sensitive patient #2, and both pRPA32-S4S8 and pRPA32-S33 increased more in ATRi and HU-treated RS-sensitive patient #2 organoids. In addition, the same pattern was visible for γH2AX, confirming the higher sensitivity to RS of patient #2 with respect to patient #5.

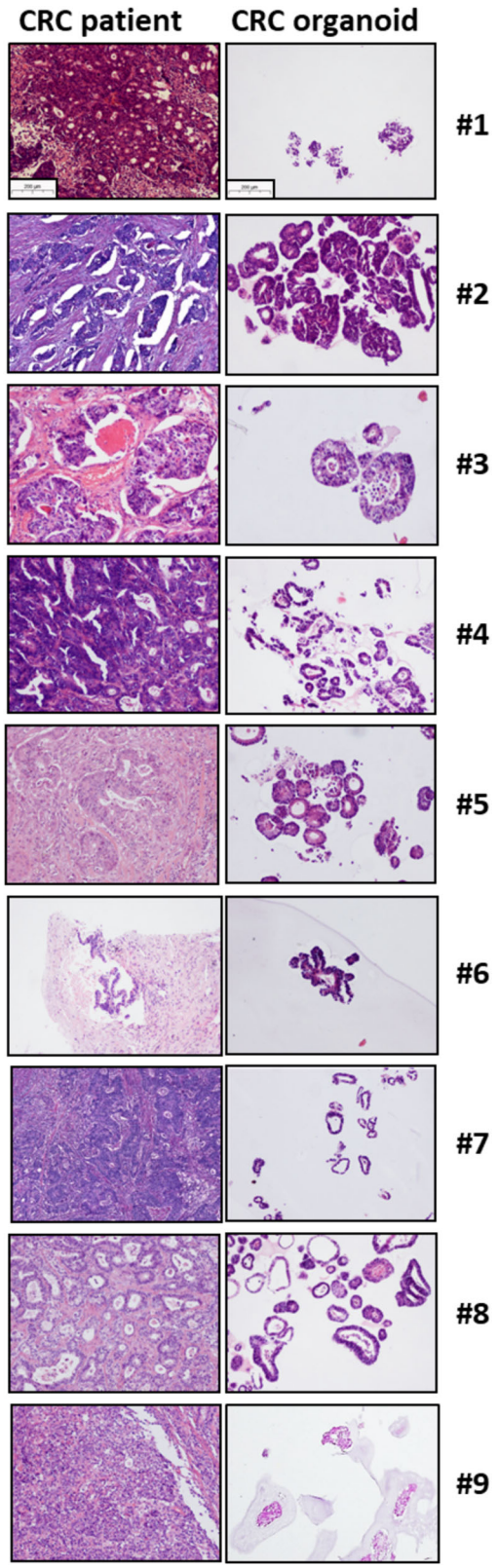


Fig. 4 | Histological staining of colorectal cancer patient's tumor tissues and organoid sections. Hematoxylin and Eosin (H&E) staining of CRC patient's tumor tissues and derived organoids. Scale bar 200 μm; magnification ×10 for all samples.

In-house parallel testing of HRDirect and commercial tests in CRC cell lines

Encouraged by these promising results with organoids, we expanded our analysis to a broader panel of models, focusing on microsatellite stable (MSS) CRC cell lines. Leveraging a subset of 31 cell lines previously screened for olaparib sensitivity[14], comprising 11 sensitive and 20 resistant lines, we applied the HRDirect pipeline since these cell lines lack matched normal DNA. The HRDirect predicted four cell lines as HRD, consistently aligning with heightened sensitivity to PARP blockade (Fig. 6, left panel).

To benchmark our findings against commercially available assays for HRD assessment, we evaluated the IVD assay AmoyDX by Amoy

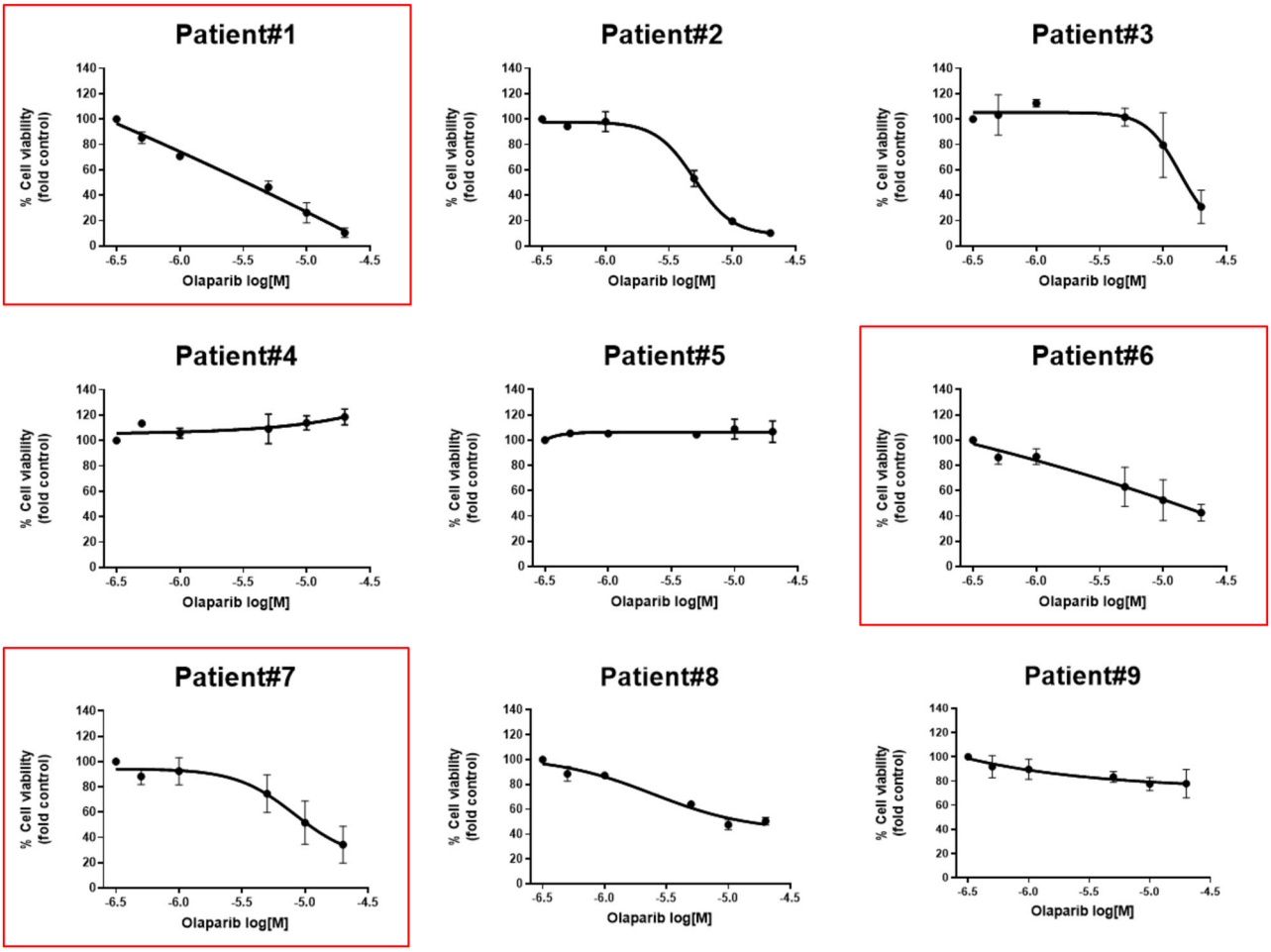


Fig. 5 | Pharmacologic testing of organoids derived from patients with colorectal cancer. Nine CRC organoids were tested with olaparib in a 7-day-long viability assay. The results at the endpoint are normalized to control wells containing DMSO vehicle. Data about patients #1, #2, #3, #4, and #5 were reproduced from the reference Arena et al., Clin Cancer Res 2020, for the purpose of clarity of the Figure. Results represent mean ± SD (at least 2 biological replicates). Red squares indicate organoids that resulted HRD by HRDirect analysis.

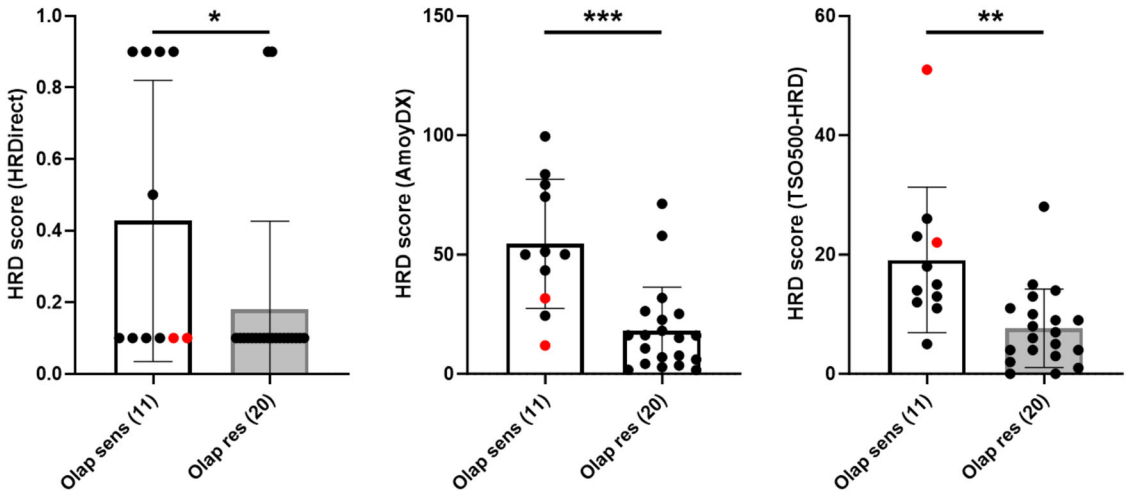


Fig. 6 | Comparison between three different HRD test scoring methods. HRD scores reported from the analysis of 31 cell models (11 olaparib-sensitive and 20 olaparib-resistant) and calculated by HRDirect, AmoyDX and TSO500-HRD tests. For technical details, see main text and methods. Red dots represent samples lacking ATM expression, as shown in reference Durinikova et al., Clin Cancer Res, 2022. Statistical significance: $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$ (2way ANOVA test).

Diagnostics, and the recently developed TruSightTM Oncology 500 HRD (TSO500-HRD) platform by Illumina NGS technology (Fig. 6, central and right panels).

Both assays effectively discriminated between sensitive and resistant cell line groups, with the four previously identified sensitive cell lines by HRDirect also ranking prominently in the other two assays, particularly notable in the AmoyDX test. However, it is noteworthy that cell lines characterized by ATM loss, a recognized biomarker of olaparib response[15], were not detected by the HRDirect and AmoyDX tests, but were classified as sensitive by the TSO500-HRD test. Conversely, in the TSO500-HRD assay, the score differentiation between sensitive and resistant cell lines was narrower, posing challenges in distinguishing between the two groups in comparison with the other two assays.

Discussion

Functional DNA repair is crucial for maintaining genome integrity and ensuring cell survival affected daily by endogenous or exogenous damage. The impairment of this repair mechanism has gathered increasing interest, both in research and clinical contexts, as it presents valuable therapeutic targets with significant implications[25]. Particularly, in CRC, the HRD predictive role of olaparib or other DDR targeting agents remains to be elucidated since few data are available from clinical trials[16,17,26].

The concept of "synthetic lethality" has led to the development of numerous drugs targeting DNA repair deficiencies in cancer cells. This progress began with the observation of PARP inhibitors' efficacy in *BRCA* mutant cells and has extended to inhibitors of non-homologous end joining (NHEJ) and RS pathways[1].

The homologous recombination (HR)-based system has so far been the most well-studied and characterized DDR pathway. The stratification of patients based on their tumor HR proficiency or deficiency has facilitated the targeted use of PARP inhibitors for the treatment of breast, ovarian, and, more recently, prostate and pancreatic cancer[27]. However, while this stratification primarily relies on *BRCA1/2* mutational status, HRD, or "BRCAness," can also result from the inactivation of other HR pathway genes. Since HRD retains both predictive and prognostic values, various companies have developed HRD tests based on HR mutations or genomic scar detection, but only a few of them are commercially available, clinically validated, FDA-approved, and used for clinical decision-making[28], such as myChoice CDx (Myriad Genetics) and FoundationOne CDx (Foundation Medicine). While these tests provide HRD prediction with good levels of sensibility, none of them represents a gold standard for HRD assessment, especially considering some HR-proficient patients' sensitivity to PARP inhibition. This implies that these applications might require additional clinical validation in order to be used in different clinical settings. Moreover, their closed-source algorithms limit additional genomic insights from tumor tissue.

Recently, an open-source bioinformatic tool, HRDetect[13], was developed to predict BRCAness by applying a particular machine learning model (called LASSO) to a set of predefined mutational signatures, extracted from WGS analysis of matched normal/tumor samples from breast cancer patients. Building upon HRDetect, we introduce HRDirect, which addresses a key limitation by predicting HR deficiency even in samples lacking matched "normal" data, crucial in clinical settings where normal counterparts are often unavailable.

HRDirect utilizes variant allele frequencies of single nucleotide variations and insertions/deletions to filter out germline mutations, which constitute the majority of variants in an unmatched comparison. We validated HRDirect on breast cancer WGS data, achieving strong agreement with published data in both matched and unmatched settings.

Extending our analysis to CRC, HRDirect demonstrated high concordance with HRDetect. We exploited our collection of CRC PDOs and tested 20 normal/tumor pairs. The general concordance between HRDetect and HRDirect resulted high (HRD true positive rate: 1.0 (2/2), HRP true positive rate: 0.88 (15/17)). In particular, two PDOs (patient #1 and patient #7) were predicted to be HRD in both matched and unmatched analysis. Moreover, one sample (patient #6), unclassified by the HRDetect workflow, resulted HRD by HRDirect, and the response to olaparib (Fig. 5) was in line with the unmatched analysis. All the other PDOs were classified as HRP by HRDetect and correctly not predicted as HRD by HRDirect. Notably, HRDirect accurately identified PDOs sensitive to olaparib treatment, highlighting its utility in identifying PARP inhibitor-responsive CRC tumors.

To further evaluate the predictive power of HRDirect, we directly compared it with two commercially available IVD assays: AmoyDX HRD focus panel by Amoy Diagnostics and the recently developed platform TruSightTM Oncology 500 HRD by Illumina. While the former can only provide the mutational profile of *BRCA1* and *BRCA2* genes and a genomic stability score (GSS), the latter can provide both an HRD score powered by Myriad Genetics and sequencing information on more than 500 genes, including HRR causal variants and genomic signatures such as HRD, tumor mutational burden (TMB), and microsatellite instability status (MSI). We compared the HRD scores obtained from the three different methods applied in a subset of CRC cell lines with known responses to olaparib[14]. We focused on MSS CRC, whose treatment and long-term response to current therapies still represent a clinical unmet need, especially for those cancers bearing *KRAS* and *BRAF* alterations. We observed that all three tests were able to distinguish the cell lines according to their response to olaparib with predicted HRD within the sensitive group. Although HRDirect was able to distinguish between the two groups (sensitive vs resistant) with lower statistical power, we must acknowledge the fact that HRDirect score is in effect a label and not a continuous number. In consideration of this, HRDirect was the test that most precisely indicated the resistant models, while AmoyDX and especially TSO500-HRD provided scores that still overlapped among sensitive and resistant cells, making the clinical application not straightforward.

Overall, considering the "composite biomarker approach" that we recently developed[15] and which includes parallel expression analysis of HR biomarkers such as RAD51C and ATM, which is not captured by the HRDirect pipeline, we believe that the combination of this approach together with the HRDirect test might provide the most accurate prediction for HRD tumor stratification, with immediate clinical implications.

Moreover, considering the wealth of genomic information provided by WGS, regardless of HR status, an integrated approach combining HRDirect and composite biomarker analysis should offer comprehensive genetic profiling of tumor samples at a lower cost compared to existing commercial tests.

Translationally, the opportunity to identify HR (or, more in general, DDR) deficient CRC patients could be crucial for broadening the therapeutic armory, particularly for MSS CRC patients, for whom treatment options are still limited. Our HRDirect composite biomarker approach stands as a novel opportunity to design a proof-of-concept phase II clinical trial in which olaparib, or potentially also other DDR targeting or DNA damaging (i.e., chemotherapy) agents, could be administered to molecularly selected CRC patients and better stratify those that might benefit from these therapies.

Methods

Whole genome sequencing

Genomic DNA (gDNA) was extracted from both cell lines and human organoids by means of Maxwell® RSC Blood DNA Kit AS1400 (Promega, Madison, WI, USA).

Starting from 500 ng of cell line-derived gDNA, Next Generation Sequencing (NGS) libraries were prepared in-house by means of Nextera DNA Flex Library Prep kit (Illumina Inc., San Diego, CA, USA), according to manufacturer's protocol. The quality of libraries was checked with Qubit™ dsDNA Quantification Assay Kits (ThermoFisher Scientific, Waltham, MA USA), while DNA fragments' size distribution was assessed using the 2100 Bioanalyzer with a High-Sensitivity DNA assay kit (Agilent Technologies, Santa Clara, CA, USA). Equal amounts of final DNA libraries were pooled and sequenced on NovaSeq 6000 (Illumina Inc., San Diego, CA, USA) as paired-end 150 bp reads. In the case of organoid samples, 2ug of gDNA have

been used as starting material for TruSeq DNA PCR-Free Library Prep Kit (Illumina Inc., San Diego, CA, USA), in order to generate WGS data according to manufacturer's protocol. DNA fragmentation step has been obtained by using M220 Focused-ultrasonicator (Covaris LLC, Woburn, MA, USA) with settings for 500 bp and 130ul AFA tubes. The quality of libraries was checked with Qubit™ dsDNA Quantification Assay Kits (ThermoFisher Scientific, Waltham, MA USA), while DNA fragments' size distribution was assessed using the 2100 Bioanalyzer with a High-Sensitivity DNA assay kit (Agilent Technologies, Santa Clara, CA). Equal amounts of final DNA libraries were pooled and sequenced on NovaSeq 6000 (Illumina Inc., San Diego, CA, USA) as paired-end 150 bp reads.

HRDetect pipeline

With the aim to reproduce the original results by Davies et al. (Davies, Glodzik et al. 2017), we develop a pipeline that is able to generate an HRDetect score starting from alignment data (Fig. 1). This workflow is composed of these steps: annotation of BAM files, variant calling, filtering and finally HRDetect method calling. Alignment files must be compliant with Cancer Genome Project pipeline that requires them to have Read-Group identifier set ("RG:") on every row along with several information in the BAM header:

-read group ("RG" line)
-genome assembly identifier ("AS" field in "SQ" line)
-species ("SP" filed in "SQ" line)
-sample name ("SM" filed in "RG" line).

A raw BAM input file is annotated as requested by a custom script.

The official Cancer Genome Project for WGS pipeline (also called "cgpwgs") by Wellcome Sanger Institute was used for variational analysis; this workflow is implemented in a Docker image and a corresponding Dockerfile is hosted as "dockstore-cgpwgs" on GitHub website (https://github.com/cancerit/dockstore-cgpwgs). Some of the dependencies in the Dockerfile were broken, so it was then modified in order to fix them and have a running container. This step always requires two samples as input: one healthy and one tumoral. Custom scripts were used to filter variants retaining only the most confident ones. Raw variations results were converted to match the format requested by the "HRDetect_pipeline" function in "signature.tools.lib" R library [20] with a custom script.

HRDirect pipeline

Starting from the previous pipeline, we developed a new workflow able to be independent of normal or healthy data. To make limited changes to the HRDetect workflow, we generated a meta-normal sample using 14 healthy WGS from the Davies breast dataset. These data were merged and subsampled to achieve a median total depth of 40× (typical depth for WGS data), resulting in a new single BAM file. This will be used instead of the matched (and possibly missing) normal sample as input for the workflow. Four methods have been developed in order to enrich real somatic variants (Supplementary Fig. 3). The first two methods took into account only SNVs or indels whose VAF is in the range of 10–20% (strategy A) or 5–35% (strategy B) respectively. These values come from the observation of the distribution of VAFs for matched comparison (hereafter reference distribution), both in SNVs and indels (Supplementary Figs. 1 and 2).

A third strategy was evaluated, in which variations are not filtered based on their value (i.e., range filter), but are sampled proportionally to the reference distribution, trying to reproduce this latter (strategy C): most represented VAF are more sampled. This approach had similar results to strategy B.

Strategy D is not based on the reference distribution but on the sample data. Germline and somatic variations are mixed, but the formers are the vast majority; since the germline VAFs distribution is essentially bell-shaped and specular and the somatic distribution is not enriched for frequencies larger than 50%, this part of the real distribution can be considered as pure germline. Therefore, we can hypothesize that the germline shape is the mirror of the right part of the total distribution. The exceeding fraction is then the somatic one. This is the core of strategy D.

Based on each strategies' results, the HRDirect workflow implements an "ensemble" approach collecting all the classifications using a majority vote rule and outputting the final prediction (Fig. 2). Notably, the outcome of HRDirect is no more a number but a classification label.

Ethics, organoid culture, and drug screening

Tumor samples were obtained from patients treated at Niguarda Cancer Center, (Milano, Italy) and IRCCS (Candiolo, Turin, Italy). All patients signed a dedicated informed consent in accordance with guidelines of the ALFAOMEGA Master Observational Trial (NCT04120935, IFOM-CPO003/2018/PO002) and the PROFILING protocol (001-IRCC-00IIS-10, 6.0 version, dated 24th August 2015).

The study was conducted in accordance with the Declaration of Helsinki and under the approval of the local Independent Ethical Committees of each participating center (for ALFAOMEGA study: Ethical Committee Niguarda Cancer Center Milano Area 3, decision n. 617-122018 dated 13/12/20218 and Ethical Committee IRCCS Candiolo, decision n. CE IRCCS 102/2021 dated 25/03/2021, for Profiling study: Ethical Committee IRCCS Candiolo, decision n. CE IRCCS 225/2015 dated 10/09/2015). PDOs #1, #2, #3, #4, #5, #11, and #12 were previously established and characterized as described in ref. 14,29,30.

Organoids from CRC patients #7, #8, #9, #10, #13, #14, #15, #16, #17, #18, #20 were established directly from tissue biopsy obtained at the time of surgery, while organoids from patients #6, #19 were generated from patient-derived xenografts (PDXs) models following procedures described in full details in ref. 14.

To generate PDXs, tumor specimens were subcutaneously implanted in 7-week-old NOD-SCID mice (Charles River Laboratory). All animal procedures were approved by the Ethical Committee of the Candiolo Cancer Institute and by the Italian Ministry of Health.

Organoids from patients #1-#9 were tested with olaparib in a 7-day-long viability assay as described in ref. 14. Olaparib response data for patients 1 to 5 are retrieved from previous publication[14], while patients 6 to 9 have been de novo screened.

Hematoxylin & Eosin staining of patients' tumor tissue and organoids

The tumor specimens were fixed with 10% buffered formalin for 24 hours at room temperature. After rinsing with running water, the specimens were transferred to the Tissue Processor Donatello Series 2 Diapath (Diapath, Martinengo (BG), Italy) for dehydration. The blocks were embedded into paraffin in the Embedding Module Canova Diapath (Diapath, Martinengo (BG), Italy). The FFPE blocks were sectioned at 3 µm using manual rotary microtome Leica RM 2255 (Leica Biosystems, Frankfurt, Germany) and were stained with Hematoxylin & Eosin solution.

Immunohistochemical staining of CRC organoids

Three micron-thick sections were cut from formalin-fixed paraffin-embedded cell blocks of PDOs and stained with antibodies raised against RAD51C (rabbit polyclonal antibody, E185, Life Technologies, Thernofisher) and ATM (rabbit monoclonal antibody Y170, Abcam). The protocol for RAD51C was optimized on the Leica BOND staining system (Leica Biosystems), whereas the protocol for ATM was optimized on an automated immunostainer Ventana Benchmark ULTRA (Ventana Roche). Positive and negative controls were included in each immunohistochemical run. For RAD51C, the pattern and intensity of membranous, cytoplasmic, and nuclear staining were recorded; for ATM, the presence or lack of nuclear staining was recorded.

Western blot analysis on organoids

Organoids were enzymatically dissociated using TrypLE Express Enzyme for 10 to 20 minutes at 37 °C to obtain single-cell suspensions and seeded at a density of 140.000 cells per well in 6-well plates precoated

with basement membrane extract (BME; Cultrex BME Type 2; Amsbio) overlayed with 2 mL of growth media containing 2% BME. The treatment with drugs started on day 4 after seeding when formed growing organoids were visible. Organoids were treated in fresh 2 mL medium without BME with 1 µM ATRi ceralasertib for 24 hours and 2.5 mmol/L HU for 4 hours. Collection of the organoids was performed using a Cell dissociation Solution. Complete removal of BME was conducted following several washes with cold PBS. Organoids were subsequently lysed in using boiling SDS buffer [50 mmol/L Tris-HCl (pH 7.5), 150 mmol/L NaCl, and 1% SDS] to extract total cellular proteins, quantified by the BCA Protein Assay Reagent kit (ThermoFisher Scientific), and prepared using LDS and Reducing Agent (Invitrogen). Western blot analysis was performed with Enhanced Chemiluminescence System (GE Healthcare) and peroxidase-conjugated secondary antibodies (Amersham). The following primary antibodies were used for Western blotting: anti-phospho-RPA32 (Ser33; Bethyl Laboratories A300-246A; 1:1000), anti-RPA32 (Abcam AB252861; 1:2000), anti-RPA32 (S4/S8; Bethyl Laboratories A300-245A; 1:3000), anti-phospho-Histone H2AX (Ser139; Cell Signaling Technology, 80312S; 1:1000), anti-H2AX (Cell Signaling Technology, 7631S; 1:1000), anti-phospho-DNA-PK (Ser2056; Cell Signaling Technology, 68716S; 1:1000), anti-DNA-PK (Cell Signaling Technology, 12311S; 1:1000), anti-ATM (Cell Signaling Technology, 2873S; 1:1000); anti- phospho-ATM (Ser1981; Cell Signaling Technology, 4526 s; 1:1000) and anti-Vinculin (MERCK 05-386; 1:3000). Detection of the chemiluminescent signal was performed with ChemiDoc Imaging System (Bio-Rad).

The following primary antibodies were used for Western blotting in supplementary Fig.4: anti-HSP90 (ABCAM, ab2928; 1:1000), anti-RAD51C (Santa Cruz Biotechnology SC-56214; 1:1000), anti-ATM (Cell Signaling Technology, 2873S; 1:1000).

HRD score by commercial targeted assays

First, DNA extracted from 31 cell lines was subjected to targeted sequencing using the TruSight™ Oncology 500 HRD panel (TSO500-HRD; Illumina Inc., San Diego, CA, USA). The panel covers 533 genes for a total sequenced size of 1.94 Mb, allowing the detection of SNVs and small indels, and the assessment of MSI status (120 loci), TMB, and copy number for 59 genes. TruSight Oncology 500 HRD also includes probes specifically designed to assess genomic scars taking advantage of the Myriad Genetics Genomic Instability Score (GIS) algorithm to enable HRD evaluation. GIS score is an a-dimensional value ranging from 0 to 100, the sum of three independent scars: LOH, TAI, and LST (PMC6773427). Briefly, following the manufacturer's protocol, 150 ng of genomic DNA was used as starter material to generate libraries, while 80 ng of post-fragmentation material was used for the subsequent steps. Final libraries were sequenced on the Novaseq 6000 instrument (Illumina, San Diego, California, USA) to reach a minimum of 500× read depth. Raw data were then processed on a local DRAGEN™ server v3 by the DRAGEN™ TruSight Oncology 500 v2 Analysis Software which incorporates a proprietary GIS algorithm powered by Myriad Genetics for HRD assessment. Details about DRAGEN™ pipelines were reported in the user-manual https://support.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/trusight/trusight-oncology-500/200019138_01_DRAGEN-trusight-oncology-500-analysis-software-v2_1-local-user-guide.pdf.

HRD status was also evaluated with the CE-IVD AmoyDx HRD Focus Panel provided by AmoyDx (AmoyDx, Xiamen, China), according to the manufacturer's instructions. Briefly, 120 ng of DNA was used for library preparation and then sequenced on an Illumina Novaseq 6000 instrument (Illumina, San Diego, California, USA). Raw data were analyzed using the AmoyDx NGS Data Analysis System-ANDAS Software to estimate a GSS. The GSS by AmoyDx returns a score between 0 and 100, based on the evaluation of the same genomic scars identified by the GIS (PMID: 36191839), encompassing a trademarked genomic region with the proprietary AmoyDX algorithm.

Data availability

The Breast Whole Genome Sequencing dataset was downloaded from the European Genome-phenome Archive (EGA) under the study EGAS00001001178, dataset EGAD00001001322. Whole Genome Sequencing data for CRC cell lines are available at ENA website (https://www.ebi.ac.uk/ena/browser/home) with accession PRJEB71992. WGS for patient-derived organoids is available at EGA archive (https://ega-archive.org) under accession EGAD50000000617. All the HRDirect code is available at https://github.com/joecorti/HRDirect.

Received: 4 March 2024; Accepted: 9 September 2024;
Published online: 14 October 2024

References

1. Pilié, P. G. et al. State-of-the-art strategies for targeting the DNA damage response in cancer. *Nat. Rev. Clin. Oncol.* **16**, 81–104 (2019).
2. Curtin, N. J. Targeting the DNA damage response for cancer therapy. *Biochem. Soc. Trans.* **51**, 207–221 (2023).
3. Bryant, H. E. et al. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* **434**, 913–917 (2005).
4. Farmer, H. et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**, 917–921 (2005).
5. Fong, P. C. et al. Poly(ADP)-ribose polymerase inhibition: frequent durable responses in BRCA carrier ovarian cancer correlating with platinum-free interval. *J. Clin. Oncol.* **28**, 2512–2519 (2010).
6. Tutt, A. et al. Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and advanced breast cancer: a proof-of-concept trial. *Lancet* **376**, 235–244 (2010).
7. Fong, P. C. et al. Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N. Engl. J. Med.* **361**, 123–134 (2009).
8. Robson, M. E. et al. OlympiAD final overall survival and tolerability results: olaparib versus chemotherapy treatment of physician's choice in patients with a germline BRCA mutation and HER2-negative metastatic breast cancer. *Ann. Oncol.* **30**, 558–566 (2019).
9. Byrum, A. K., Vindigni, A. & Mosammaparast, N. Defining and modulating 'BRCAness. *Trends Cell Biol.* **29**, 740–751 (2019).
10. Pellegrino, B. et al. Controversies in oncology: are genomic tests quantifying homologous recombination repair deficiency (HRD) useful for treatment decision making? *ESMO Open* **4**, e000480 (2019).
11. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
12. Nik-Zainal, S. et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).
13. Davies, H. et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat. Med.* **23**, 517–525 (2017).
14. Arena, S. et al. A subset of colorectal cancers with cross-sensitivity to olaparib and oxaliplatin. *Clin. Cancer Res.* **26**, 1372–1384 (2020).
15. Durinikova, E. et al. Targeting the DNA damage response pathways and replication stress in colorectal cancer. *Clin. Cancer Res.* **28**, 3874–3889 (2022).
16. Behl, D. et al. Olaparib (O) in patients (pts) with colorectal cancer (CRC) with ATM mutation (mut): results from the Targeted Agent and Profiling Utilization Registry (TAPUR) study. *J. Clin. Oncol.* **41**, 122–122 (2023).
17. Takashima, A. et al. Olaparib with or without bevacizumab versus bevacizumab plus a fluoropyrimidine as maintenance therapy in advanced colorectal cancer: the randomized phase 3 LYNK-003 study. *Eur. J. Cancer*, **205**. (2024).
18. Lazzari, L. et al. Patient-derived xenografts and matched cell lines identify pharmacogenomic vulnerabilities in colorectal cancer. *Clin. Cancer Res.* **25**, 6243–6259 (2019).

19. Medico, E. et al. The molecular landscape of colorectal cancer cell lines unveils clinically actionable kinase targets. *Nat. Commun.* **6**, 7002 (2015).
20. Degasperi, A. et al. A practical framework and online tool for mutational signature analyses show inter-tissue variation and driver dependencies. *Nat. Cancer* **1**, 249–263 (2020).
21. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
22. AlDubayan, S. H. et al. Inherited DNA-repair defects in colorectal cancer. *Am. J. Hum. Genet.* **102**, 401–414 (2018).
23. Knijnenburg, T. A. et al. Genomic and molecular landscape of DNA damage repair deficiency across The Cancer Genome Atlas. *Cell Rep.* **23**, 239–254.e6 (2018).
24. Heeke, A. L. et al. Prevalence of homologous recombination-related gene mutations across multiple cancer types. *JCO Precis. Oncol.* **2018,** 17.00286 (2018).
25. Minten, E. V. & Yu, D. S. DNA repair: translation to the clinic. *Clin. Oncol.* **31**, 303–310 (2019).
26. Phase II trial of pembrolizumab and olaparib in homologous-recombination deficient (HRD) advanced colorectal cancer (CRC), S. Merk and S.A. Dohme España, Editors. (2022).
27. Kamel, D. et al. PARP inhibitor drugs in the treatment of breast, ovarian, prostate and pancreatic cancers: an update of clinical trials. *Curr. Drug Targets* **19**, 21–37 (2018).
28. Mangogna, A. et al. Homologous recombination deficiency in ovarian cancer: from the biological rationale to current diagnostic approaches. *J. Pers. Med.* **13**, 284 (2023).
29. Lorenzato, A. et al. Vitamin C restricts the emergence of acquired resistance to EGFR-targeted therapies in colorectal cancer. *Cancers (Basel)* **12**, 685 (2020).
30. Magrì, A. et al. High-dose vitamin C enhances cancer immunotherapy. *Sci. Transl. Med.* **12**, eaay8707 (2020).

Acknowledgements

The authors would like to thank Gaia Stievano and Giuseppe Rospo for constructive data discussion and Gaia Grasso for helping with the graphics. This work was supported by AIRC under BRIDGE 2022-ID 27321 P.I. Sabrina Arena (S.A.); FPRC 5×1000 MIUR 2018 INSIDE, FPRC 5×1000 Ministero della Salute 2021 EmaGen and FPRC 5x1000 Ministero della Salute 2022 CARESS to S.A.; Italian Ministry of Health, Ricerca Corrente 2024 to S.A.; MUR Dipartimento di Eccellenza 2023-2027 14586 DIORAMA to S.A.; Ricerca Locale 2022 and 2023 from Department of Oncology, University of Torino to S.A. and M.R.; MIUR Progetti di Ricerca di Rilevante Interesse Nazionale (PRIN) grant 2022LSFKWE to S.A.; Prin 2022 PNRR finanziato dall'Unione Europea NextGenerationEU M4 C2 I.1.1.-P2022E3BTH to S.A.; European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (TARGET, grant agreement n. 101020342) (A.B.); AIRC under 5 per Mille 2018–ID. 21091 program—P.I. Bardelli Alberto (A.B.), G.L. Marchiò Caterina (C.M.), G.L. Siena Salvatore (S.S.), G.L. Marsoni Silvia (S.M.), G.L. Federica Di Nicolantonio (F.D.N.); AIRC under IG 2023 - ID. 28922 project—P.I. Bardelli Alberto (A.B.); International Accelerator Award, ACRCelerate, jointly funded by Cancer Research UK (A26825 and A28223), FC AECC (GEACC18004-TAB) and AIRC (22795) (A.B.); IMI contract no. 101007937 PERSIST-SEQ (A.B.); AIRC under MFAG 2021-ID 26439 project—P.I. Mariangela Russo (M.R.); Fondazione Oncologia Niguarda Onlus to S.S.; G.M. is a PhD students within the European School of Molecular Medicine (SEMM).

Author contributions

G.C., A.B., and S.A. conceived the study; G.C. designed the pipeline; E.B., G.C., S.E.B. performed bioinformatics analyses; A.L., C.M., F.D.N., G.C., G.M., K.B., M.M., M.R., S.A. analyzed data; K.B., M.L., M.M. performed experiments with preclinical models; M.C.A. performed tissue assessment, staining and imaging of clinical and preclinical samples; A.Bartolini, E.B. performed sequencing; A.B., C.M., L.L., S.M., S.S., and S.A. provided resources; A.B. and S.A. supervised the work; G.C. and S.A. wrote the original manuscript; all authors revised and approved the final manuscript.

Competing interests

S.A. reports personal fees from MSD Italia and a patent (international PCT patent application no. WO 2023/199255 and Italian patent application No. 102022000007535) outside the submitted work. C.M. reports personal consultancy fees from Roche, Bayer, AstraZeneca, Daiichi Sankyo, and Novartis outside the scope of the present work. C.M. also reports speaker fees from Veracyte and Daiichi Sankyo. A.B. declares the following competing financial interests: receipt of grants/research supports from Neophore, AstraZeneca, Boehringer; receipt of honoraria or consultation fees from Guardant Health; stock shareholder: Neophore, Kither Biotech; member of the SAB of Neophore. S.S. is an advisory board member for Agenus, AstraZeneca, Bayer, BMS, CheckmAb, Daiichi Sankyo, Guardant Health, Merck, Novartis, Roche-Genentech, and Seagen. G.M. received honoraria from COR2ED.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41698-024-00706-7.

Correspondence and requests for materials should be addressed to Alberto Bardelli or Sabrina Arena.