UNIVERSITÀ
DI TORINO

Doctoral School in Life and Health Sciences

PhD Program in Complex System for Quantitative Biomedicine
XXXVI Cycle

Homologous Recombination deficiency classification and PARP inhibitor response of reference-free colorectal cancer samples

Author: Giorgio Corti

Tutors: Prof. Alberto Bardelli
Prof. Enzo Medico
Co-Tutor: Prof.ssa Sabrina Arena

Coordinator: Prof. Enzo Medico

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

October 2023

Table of Contents

Abstract

PARP inhibitors (PARPi) are recognized for their ability to induce synthetic lethality in tumors exhibiting homologous recombination deficiency (HRD), commonly referred to as "BRCAness”. 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 HRD from reference-free tumor samples. We initially validated HRDirect using matched breast and colorectal cancer patient samples. Subsequently, we assessed its efficacy in predicting response to the PARP inhibitor (PARPi) olaparib by comparing it with two other commercial assays: AmoyDx HRD by Amoy Diagnostics and the TruSight Oncology (TSO) 500 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 the integration of HRDirect scoring with ATM 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.

Introduction

The Homologous Recombination Repair system

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 (Pilié et al. 2019). A significant proportion of tumors exhibits deficiencies in proteins from one or more DDR repair families, rendering them reliant on intact DDR pathways for survival, with substantial implications at the clinical and therapeutic levels (Curtin et al. 2023). 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/or *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.

Indeed HRD is also called *BRCAness* because it was first linked to BRCA1/2-mutant breast cancers (Turner et al. 2004). To date, this definition still remains but the HR pathway is undoubtedly more generic (Lord et al. 2016) and comprises the loss of function of other genes, such as *RAD51* (and its paralog *RAD51C*) and *ATM* (Durinikova et al. 2022, Tian et al. 2018).

Poly-ADP-ribose polymerase (PARP) inhibitors and Homologous Recombination deficiency

Tumors harboring alterations in the HR pathway often exhibit heightened sensitivity to specific targeted therapies, such as the inhibition of certain polymerases called PARP (poly-ADP-ribose polymerase), as initially evidenced by seminal studies in 2005 demonstrating the synthetic lethality effect of these drugs in BRCA germline deficient cancers (Bryant et al. 2005, Farmer et al. 2005).

PARPs play an important role in different cellular processes, such as transcription, replication, chromatin remodulation and DNA repair (Morales et al. 2014). They belong to a family of related enzymes that catalyzes the transfer of ADP-ribose to target proteins. PARPs are part of complex systems developed by cells during evolution to face very diverse environmental and endogenous genotoxic agents.

The synthetic lethality of PARP inhibition in some cancers and subsequent clinical trials published in 2009-2010 highlighted the profound impact of this exceptional discovery (Fong et al. 2010, Tutt et al. 2010, Fong et al. 2009). 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 (Robson et al. 2019), 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 (Byrum et al. 2019). 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 benefit from FDA-approved PARPi-based therapy or acquiring enrollment eligibility criteria for clinical trials testing novel therapies.

Genomic features of Homologous Recombination Deficiency (HRD)

Although the name *BRCAness* is related to *BRCA1/2* genes, because this phenotype was first discovered and described in BRCA1/2-mutant breast cancers (Turner et al. 2004), more recent studies made this definition more generic (Lord et al. 2016): some triple-negative breast cancer (i.e. not mutated, among other genes, in *BRCA1* or *BRCA2*), share similar clinicopathological and molecular characteristics to BRCA1/2-mutant breast cancers.

It is now well established that the BRCAness phenotype involves a broader spectrum of genes and pathways (Tian et al. 2018). In fact, different studies identified very diverse genomic biomarkers such as high number of telomeric allelic imbalance (TAI, Birkbak et al. 2012), or large chromosomal break between adjacent regions (large state transitions or LST, Popova et al. 2012), or even genomic patterns of loss of heterozygosity (LOH, Abkevich et al. 2012). The occurrence of these markers has been exploited to calculate a “score”, providing a measurement of HRD in tumors: indeed Telli and colleagues combined these three genomic measures of instability and defined the “HRD score” (Telli et al. 2016).

Other genomic features that have been exploited to identify HR defective tumors are mutational signatures (Alexandrov et al. 2013): combinations of single base substitutions (SBS) in a 3-nucleotide context arising from different biological mutagenesis mechanisms. In particular, the etiology of SBS3 is strongly linked to *BRCA1/2* inactivation and thus HRD.

Commercial assays for HRD prediction

Different diagnostic HRD tests have been developed and commercialized in the last years (Pfarr et al. 2024). Some of these have been approved by the U.S. Food and Drug Administration (FDA), such as MyChoice® CDx HRD by Myriad Genetics or FoundationOne® CDx by Foundation Medicine (Pellegrino et al. 2019). FDA-approved assays are routinely used in clinics, but are in development and of increased use in research.

Among the not FDA-approved assays, we can quote AmoyDX by Amoy Diagnostics and the recently developed TruSight Oncology 500 HRD by Illumina. This latter incorporates a proprietary algorithm powered by Myriad Genetic.

All these tests are based on a custom panel DNA sequencing that is analyzed by a patented software, whose specifics are not available, but essentially it exploits the genomic makers described in the previous chapter: LOH, LST and TAI.

Bioinformatic tools for HRD classification

Recently a novel algorithm named HRDetect (Davies et al. 2017) 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. This tool is based on a logistic regression model called LASSO trained on 560 cases of breast cancer.

Some years later, in 2020, another tool was developed. It is called CHORD (Classifier of HOmologous Recombination Deficiency, Nguyen et al. 2020) and it is a genome-wide mutational pan-cancer classifier that can also discriminate between *BRCA1*- and *BRCA2*-subtypes.

Both tools require a number of somatic analyses such as SNVs, Indels, copy-number alterations and structural variants. Exploiting these variations, genomic signatures are calculated and used in the chosen machine learning model. Raw data used in both publications are not public and are available upon request to the authors.

Colorectal cancer (CRC) and HRD

Colorectal cancer (CRC) is currently the second leading cause of cancer death and the third most commonly diagnosed cancer globally (Sung et al. 2021). The majority of CRC patients (65%) are estimated to survive 5 years after cancer diagnosis. The same expectation is reduced to 15% if the fourth (metastatic) stage is considered (National Cancer Institute Surveillance, Epidemiology, and End Results Program 2022).

The treatment of metastatic colorectal cancer (mCRC) has improved over the past fifteen years since the introduction of anti-EGFR targeted therapy, antiangiogenic agents, and the use of intensive triplet chemotherapy regimens based on fluoropyrimidines, oxaliplatin, and irinotecan (Van Cutsem et al. 2016). Although the overall survival of patients with mCRC has been increased by combining and refining

the use of cytotoxics, targeted agents and immunotherapy, the impact of these advances has been incremental rather than transformative.

Consequently, there is an increased need to identify novel therapeutic strategies to improve disease control and to prolong overall survival, especially for mCRC patients who cannot receive targeted agents or immunotherapy.

Extensive molecular profiling of large colorectal cancer datasets has identified specific subsets of cases exhibiting defects in DNA repair pathways. Among these, germline pathogenic variants in the *BRCA1* gene, associated with homologous recombination (HR) repair pathway impairments, are now recognized as a risk factor for colorectal cancer (Oh et al. 2018), especially for early-onset cases (Soyano et al. 2018). Recent studies have revealed that up to 15% of individuals carry genetic defects in HR repair genes, either in their germline or somatic cells (AlDubayan et al. 2018, Knijnenburg et al. 2018). Additionally, in other tumor types like breast and ovarian cancers, alterations in genes such as BRCA, RAD51, and PALB2 lead to the so-called “BRCAness” phenotype (Lord et al. 2016). These BRCAness-associated tumors often exhibit heightened sensitivity to specific DNA damaging agents, including platinum compounds and PARP inhibitors (Bryant et al. 2005, Tutt et al. 2018, Kaufman et al. 2015).

Inspired by these breakthrough findings in breast and ovarian cancers, we recently focused on understanding whether PARPi-based treatment might hold promise for patients with other tumors, particularly CRC. We have demonstrated that approximately 13% of CRC tumors respond to olaparib (Arena et al. 2020), possibly due to defects in HRR genes other than BRCA1/2, such as RAD51C and ATM (Durinikova et al. 2022). Given the limited therapeutic options available for CRC,

identifying CRC patients likely to benefit from PARPi, especially those resistant to immunotherapy or standard-of-care therapies, is of utmost importance.

Results

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

We initially considered the seminal work by Davies and colleagues (Davies et al. 2017) to design an “in house” pipeline (Figure 1, left side) that, by exploiting whole genome sequencing (WGS) data, can classify samples from cancer patients into two distinct classes: HR proficient or deficient (hereafter named as HRP and HRD, respectively) (Figure 1, left side). The initial phase of this workflow entails variant calling, encompassing Single Nucleotide Variations (SNVs), insertions, deletions (indels), and Structural Variants (SVs), achieved through the Cancer Genome Project for WGS pipeline ("cgpwgs") developed by the Wellcome Sanger Institute.

We customized the Docker container holding the cgpwgs pipeline to address unresolved dependencies. 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 (Degasperi et al. 2020) 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").

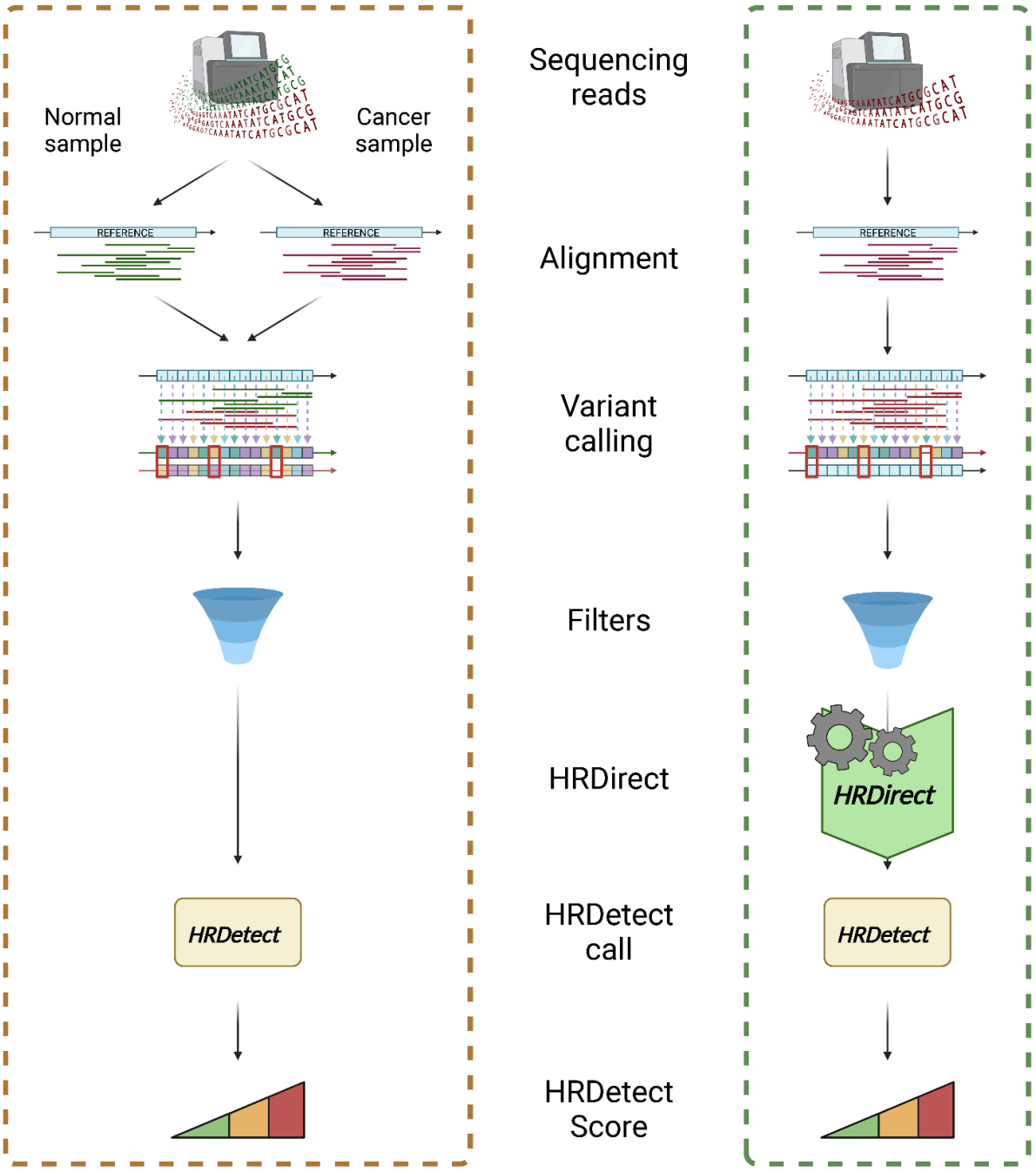

Figure 1. HRDetect and HRDirect workflow. Schematic representation of the matched analysis by HRDetect (left, gold dashed box) and the unmatched one by HRDirect using the meta-normal data (right, green dashed box).

To validate our workflow, we utilized alignment files in BAM format (Li et al. 2009) and related predictions from 76 patients in Davies' publication. Our analysis

confirmed the reliability of our pipeline (Supplementary Table 1, Cohen’s kappa: 0.73 for all samples; Cohen’s kappa: 0.80 for patients not classified as "NC" by Davies).

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 through remapping data to hg38 using the latest aligner version (bwa mem, v0.7.17) (Li 2013). Subsequent reclassification of all patients revealed only 5 misclassifications out of 76 (Supplementary Table 2, Cohen’s kappa: 0.86). Remarkably, of these 4 misclassifications, 3 patients were unclassified ("NC") by Davies and colleagues, resulting in two misclassifications out of 73 patients (Cohen’s kappa: 0.94).

HRDirect: design of an unmatched HRDetect workflow

The absence of normal matched samples presents 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" (Figure 1, right side).

Given the necessity of both healthy and tumor data for cgpwgs execution, we generated a 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 expected classification is “HRD" (Supplementary Table 3): only 20% (5 over 25 samples) of the patients will be correctly classified. On the contrary, for HRP samples, where there are no genomic scars to be detected, the metanormal gave perfect classification (48

over 48 patients). We reasoned that some strategy to enrich for somatic variants is needed.

The LASSO logistic regression model used in HRDetect relies on defined properties (features) and a corresponding weight. Features with a higher weight have a greater impact on the final result (the prediction score).

Feature type	**Weight**
proportion of deletions with microhomology	2.398
number of mutations of substitution signature 3	1.611
number of mutations of rearrangement signature 3	1.153
number of mutations of rearrangement signature 5	0.847
HRD LOH index	0.667
number of mutations of substitution signature 8	0.091

Table 1. Features and weights used in HRDetect algorithm. Description of all features identified by the LASSO logistic regression model used in HRDetect. Features weight is also shown, where a higher number represents the most important ones.

As we can see in Table 1, features related to SNVs and indels are prevalent, so we decided to mainly focus on them. Analysis of variant allele frequencies (VAFs) revealed that somatic mutations are enriched towards lower frequencies in matched comparisons (Figure 2 and 3).

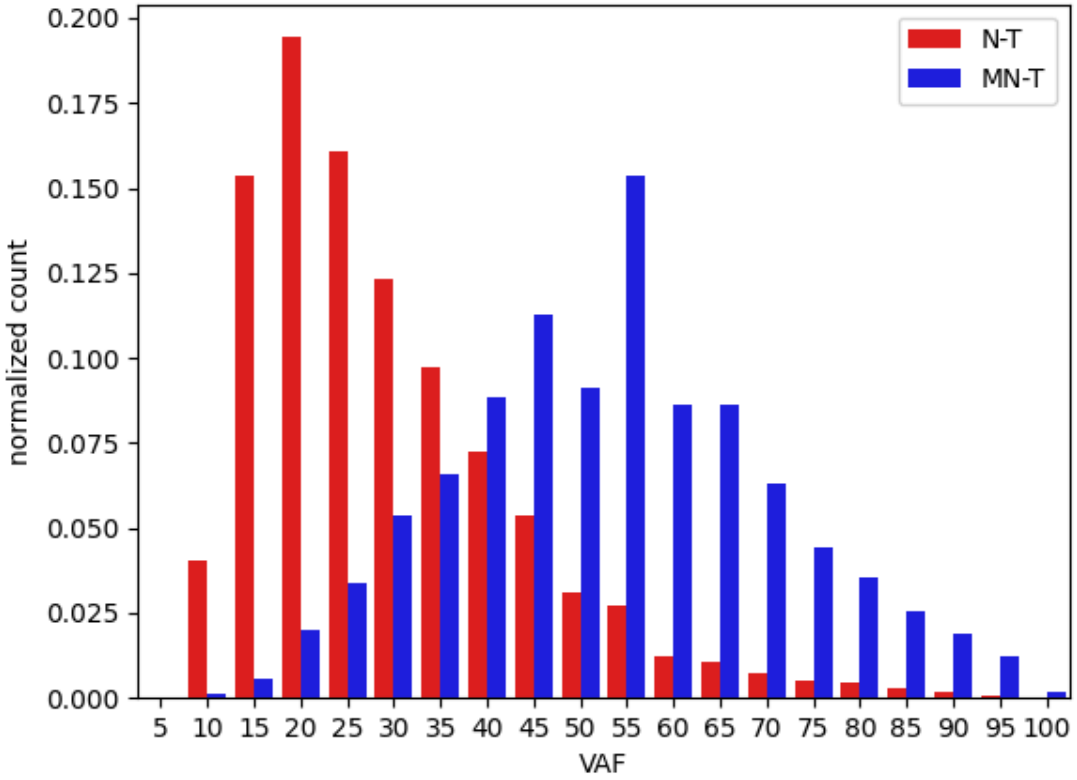

Figure 2. Frequency distribution for somatic and germline single nucleotide variations (SNVs) in a set of samples. Red: the distribution of SNVs variant allele frequencies for a normal/tumor comparison for a set of samples (only somatic variants), blue: the same distribution but with the meta-normal instead of matched normal data (mixed germline/somatic variants).

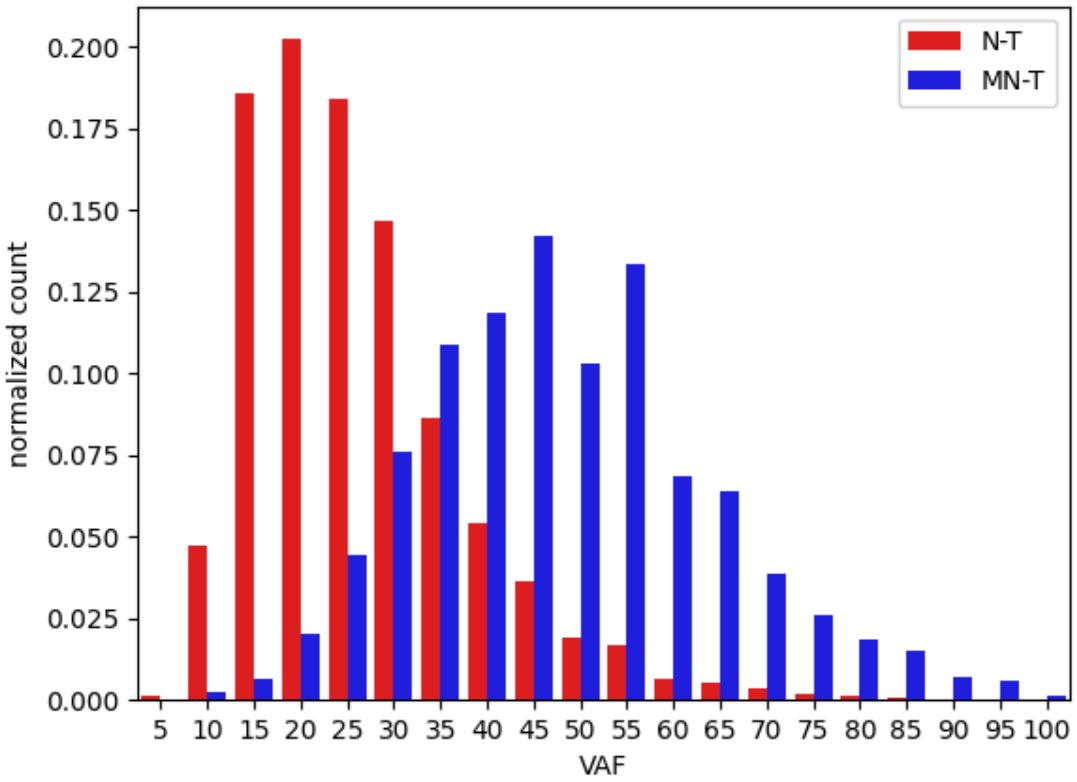

Figure 3. Frequency distribution for somatic and germline insertions/deletions in a set of samples. Red: the distribution of indels variant allele frequencies for a normal/tumor comparison for a set of samples (only somatic variants), blue: the same distribution but with the meta-normal instead of matched normal data (mixed germline/somatic variants).

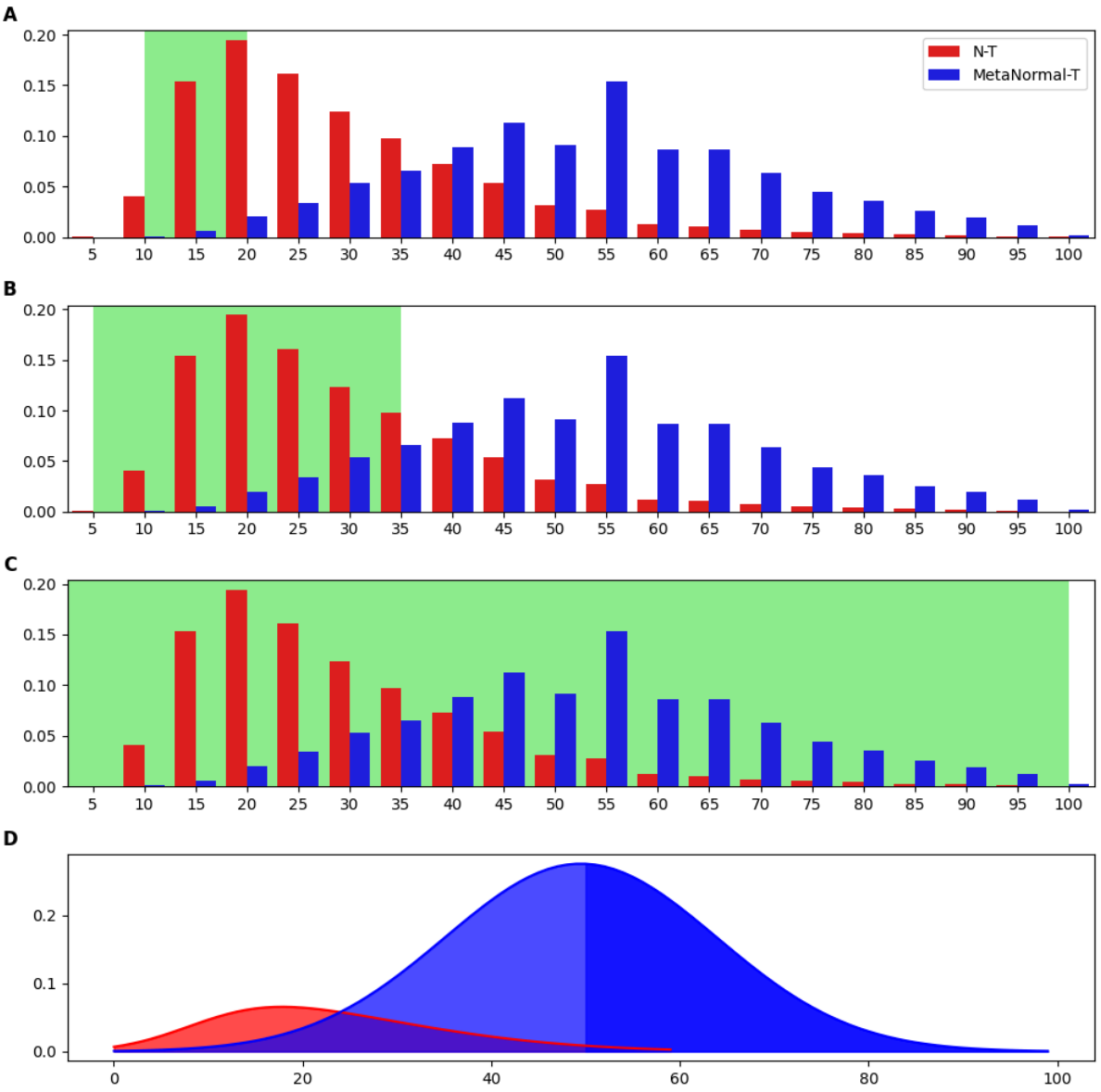

Figure 4. Strategies used to enrich for somatic variations in a variant call analysis (SNVs or In/Dels) with meta-normal data instead of matched normal. Here are the 4 strategies developed in order to filter the vast majority of germline variants and retain the somatic ones. An unmatched comparison results in a mix of germline and somatic variations: the former ”dilutes” the latter and the resulting variant allele frequency (VAF) distribution is quite indistinguishable from a ”germline-only” one (blue). In a matched analysis, conversely, the VAF distribution (red) is very different. A and B: only variations with a VAF falling in the green range are retained (”strategy A” and “strategy B”); C, variations of any VAF are sampled with weight depending on the occurrence of that particular VAF in the red distribution (”strategy C”); D, draft of ”strategy D”: the unmatched distribution is the sum of the red and blue distributions in the sketch; since the germline VAFs distribution is essentially bell-shaped and specular, and the somatic distribution is not enriched for frequencies larger than 50% (dark blue), this part of the unmatched distribution can be considered as pure germline; the exceeding fraction (red) of the mirrored dark blue part (light blue) of that distribution is enriched for somatic variants.

Based on these findings, we developed several approaches to mitigate unwanted germline variations (Figure 4 A-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% (Figure 5). 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 (Figure 4 C).

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 HRDetect scoring. While none of these approaches individually achieved accurate patient HR status predictions within a small margin of error (Figure 5), 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 et al., resulting in correct classification of 99% of samples (Figure 5).

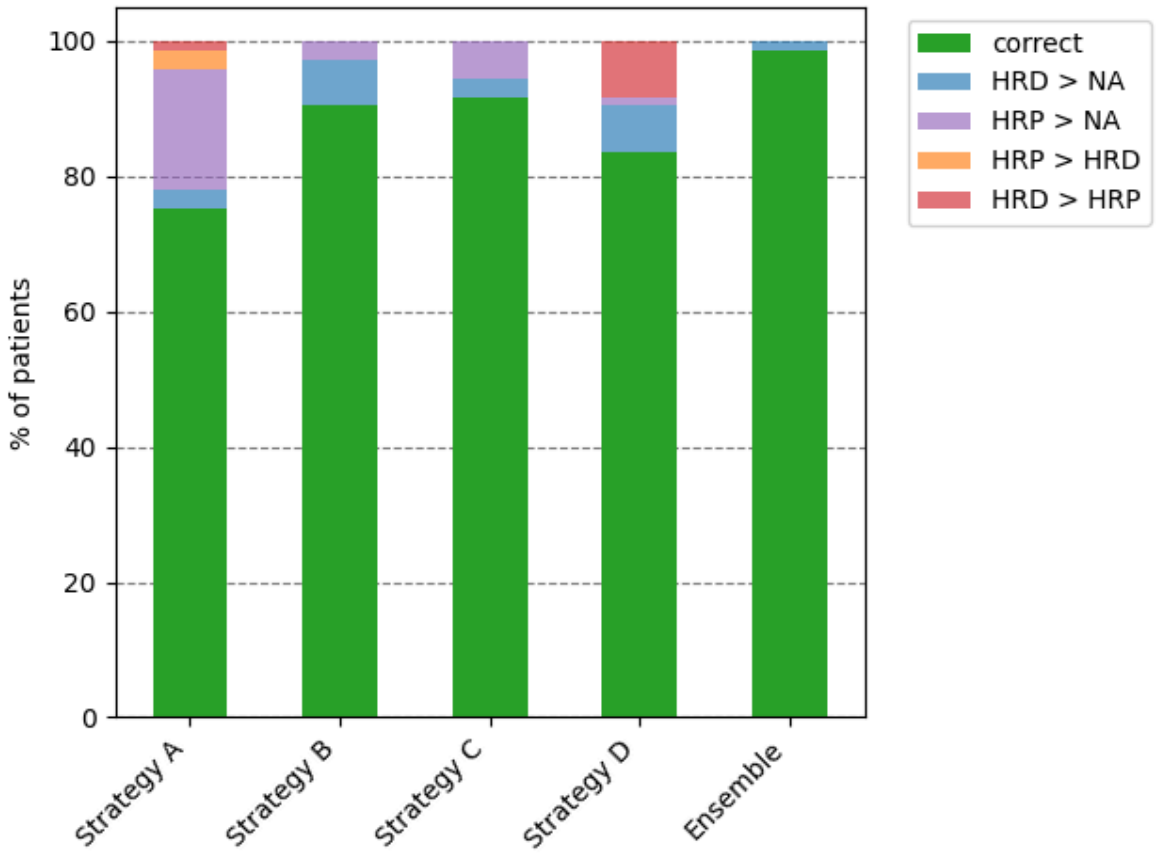

Figure 5. Evaluation of prediction accuracy for each strategy. Accuracy was evaluated on the dataset of breast cancer samples (N=76). The outcome of each strategy is divided into the correct (green) and inaccurate classification: the error type is annotated in different colors. “HRD>HRP”: expected HRD but classified HRP, “HRP>HRD”: expected HRP but classified HRD, “HRP>NC”: expected HRP but not classified, “HRD>NC”: expected HRP but not classified.

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.

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 (Figure 6, 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 (Arena et al. 2020, AlDubayan et al. 2018, Knijnenburg et al. 2018, Heeke et al. 2018).

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 (Figure 6).

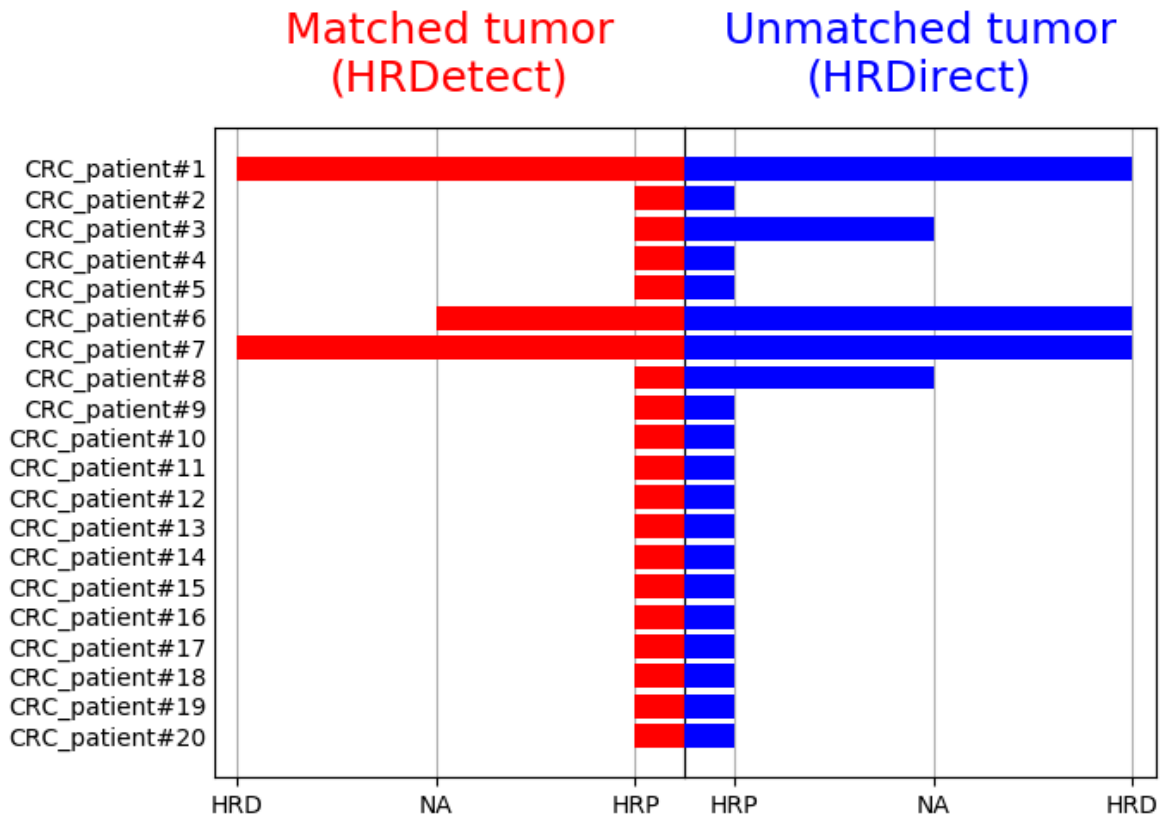

Figure 6. Predictions by HRDetect and HRDirect on CRC organoids. On the left, the matched analysis by HRDetect on 20 samples; on the right, the same analysis but with HRDirect (using the meta-normal data instead of the matched normal sample).

To ascertain whether HRD prediction correlates with sensitivity to PARP inhibitors, we selected nine nicely growing CRC organoids for olaparib treatment. Notably, all three HRD-predicted organoids (patient#1, #6, and #7) exhibited sensitivity to PARP blockade (Figure 7). 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. Leveraging our "composite biomarker" approach we previously defined (Durinikova et al. 2022), we assessed the expression of ATM and RAD51C in CRC patient#6 and #7 via immunohistochemistry. While CRC patient#7 showed very low expression of both biomarkers (Figure 8 A), as also confirmed by western blot analysis (Figure 8 B), patient#6 exhibited normal expression.

Interestingly, CRC patient#3 and #8, with undefined ("NC") HRDirect scores, demonstrated modest sensitivity to olaparib, suggesting involvement of other partially penetrant HRD mechanisms influencing PARPi sensitivity to a lesser extent.

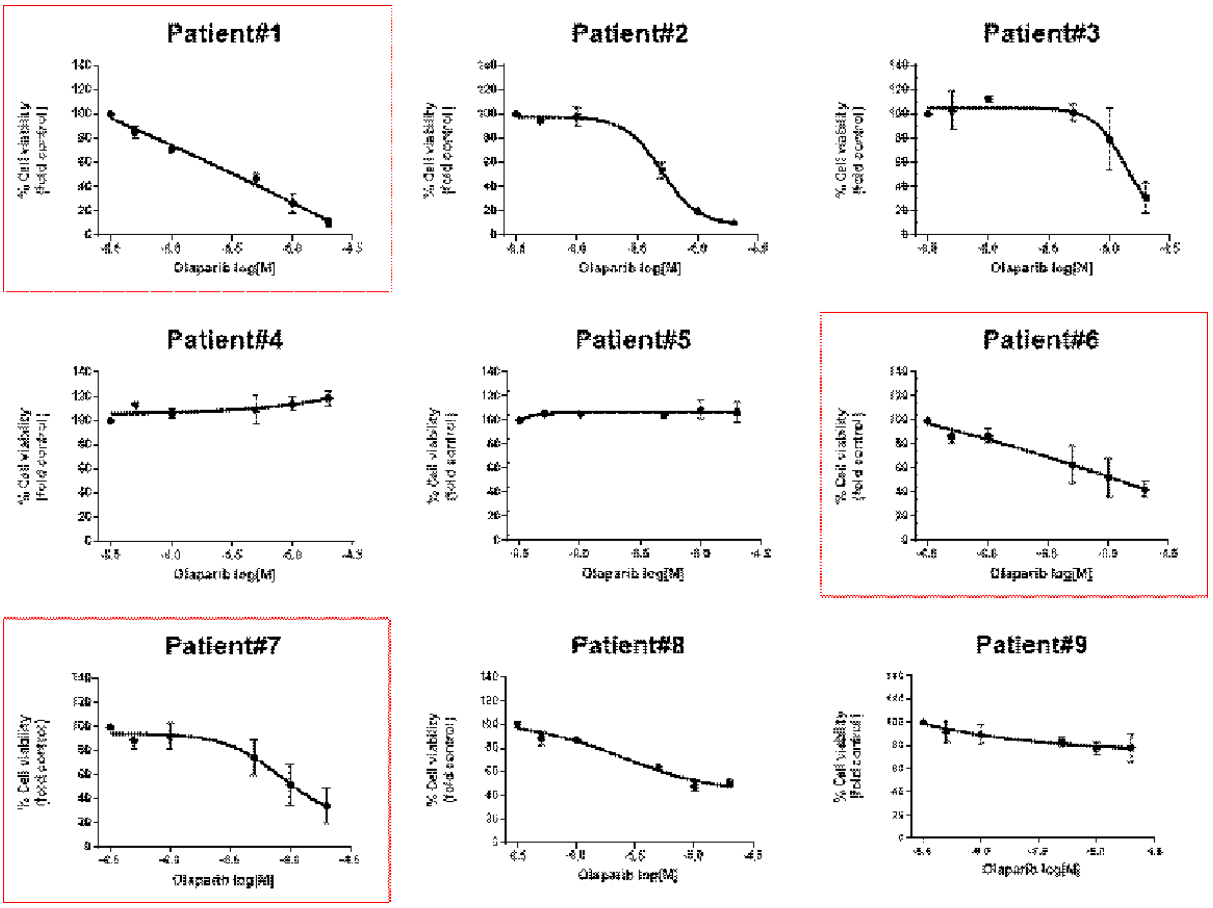

Figure 7. Pharmacologic testing of organoids derived from patients with colorectal cancer. Nine CRC organoids were tested with olaparib in a 7-day-long viability assay. The results at the endpoint are normalized to control wells containing DMSO vehicle. Data about patient #1, #2, #3, #4, #5 were reproduced 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.

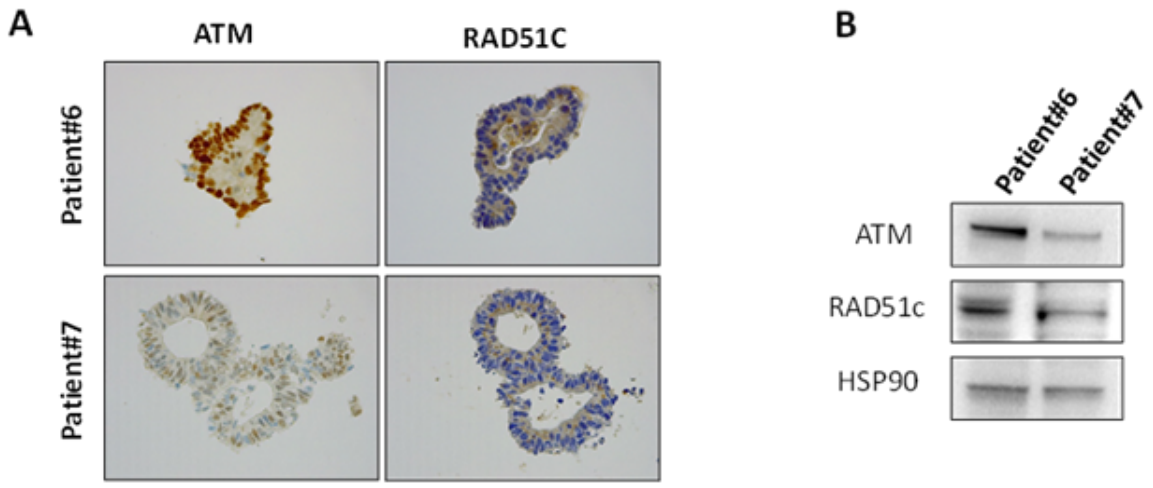

Figure 8. ATM and RAD51C expression analysis in patient-derived organoids. A, IHC detection of RAD51C and ATM proteins in CRC PDOs' cytoclots. B, western blot analysis of CRC PDOs. HSP90 was used as a loading control.

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 sensitivity (Durinikova et al. 2022) or not captured by the assays.

In-house testing of HRDirect and HRD commercial assays in CRC cell lines

Encouraged by 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 (Arena et al. 2020), comprising 11 sensitive and 20 resistant lines, we applied the HRDirect algorithm since these cell lines lack matched normal DNA. The HRDirect predicted four cell lines as HRD, consistently aligning with heightened sensitivity to PARP blockade (Figure 9, left panel).

To benchmark our findings against commercially available assays for HRD assessment, we evaluated the IVD assay AmoyDX by Amoy Diagnostics, and the recently developed TruSightTM Oncology 500 HRD (TSO500-HRD) platform by Illumina NGS technology (Figure 9, central and right panels).

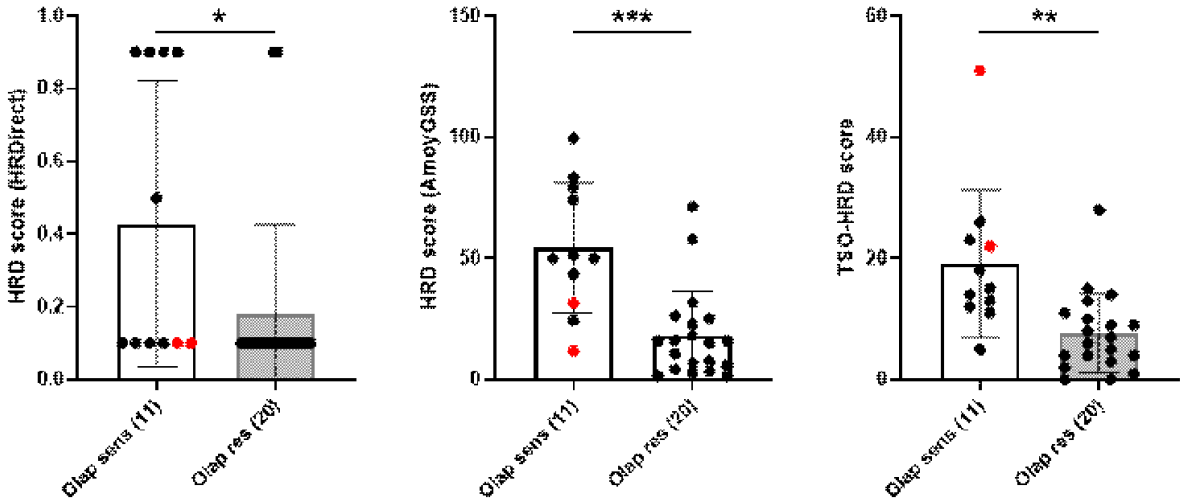

Figure 9. Comparison between three different HRD test scoring methods. HRD scores reported from the analysis of 31 cell models and calculated by HRDirect, AmoyDX and TSO-HRD tests. For technical details, see main text and methods. Red dots represent samples lacking ATM expression as shown in reference Durinikova et al. 2022. Statistical significance: *, P < 0.05; **, P < 0.01; ***, P < 0.001 (2way ANOVA test).

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 (Durinikova et al. 2022), 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 (Minten et al. 2019).

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 replication stress (RS) pathways (Pilié et al. 2019).

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 of prostate and pancreatic cancer (Kamel e al. 2018). However, while this stratification primarily relies on BRCA1/2 mutational status, HR deficiency (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 few of them are commercially available, clinically validated, FDA-approved and used for clinical decision making (Mangogna et al. 2023), 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 more 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 (Davies et al. 2017), 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 (Cohen’s kappa: 0.86 and 0.89 respectively). with published data in both matched and unmatched settings.

Extending our analysis to colorectal cancer (CRC), HRDirect demonstrated high concordance with HRDetect. We exploited our collection of CRC patient-derived organoids (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 were predicted to be HRD in both matched and 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 a 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 response to olaparib (Arena et al. 2020). We focused on microsatellite stable (MSS) CRC, whose treatment and long-term response to current therapies still represents 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 TSO-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 (Durinikova et al. 2022) and that includes parallel expression analysis of HRR

biomarkers as RAD51C and ATM that are not captured by the HRDirect algorithm, 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.

Conclusions and future perspectives

The HR status, and in particular HRD, has been evaluated in many tumor types, such as ovarian, breast, prostate and pancreatic cancers. For these tumors, different genomic assays, either commercial or open-source, have been developed and validated with accuracy. All these approaches rely on the detection of mutational signatures and structural variants analyzing whole genome or targeted sequencing data. Identification of HRD tumors is of significant relevance because of the predictive value of response to a specific class of drugs, called PARP inhibitors.

Recently also CRC has been included into the list of tumors with HR deficiency, or BRCAness affected, albeit with a lower incidence.

Our results reveal that the prediction of the HR status of a patient by using the tumor data only instead of the previous normal/tumor matched data is feasible, allowing the classification of a broader range of cases when the normal tissue is missing, a frequent issue in the clinics and in the management of preclinical models such as cell lines and patient-derived organoids. In addition, our new pipeline HRDirect allows the limitation of the confounding impact on the classification of germline variations, which are the vast majority in an unmatched setting.

One limitation of both our open-source pipeline HRDirect and commercially available assays, such as AmoyDx HRD or TruSight Oncology 500 HRD, is the inability to identify as HRD those patients that are effectively HR deficient because of ATM loss, an alteration that dies not trigger detectable genomic scars.

In conclusion, we suggest to identify HRD tumors by HRDirect that, as an open-source tool, allows the collection of multiple levels of genomic information, and to perform it with parallel IHC analysis to identify multiple biomarkers of HR deficiency such as ATM. This analysis, when validated in a larger cohort of CRC patients, might provide an immediate and clinical impact for the treatment of colorectal tumors, especially those that are refractory to commonly used therapies.

Supplementary Tables

Supplementary Table 1:

#Patient	Davies score	observed	HR_class_Davies	HR_class_obs
PD22251a	0.002538573672	0.0002728802	P	P
PD22355a	0.9999328577	0.9999705	D	D
PD22358a	0.9169706398	0.6545219	D	NC
PD22359a	0.001753778715	0.0004988132	P	P
PD22360a	0.9747501969	0.7590769	D	D
PD22361a	0.0003891565291	0.0000414314	P	P
PD22362a	0.002184709154	0.001108328	P	P
PD22363a	0.9723387388	0.8274132	D	D
PD22364a	0.003467522582	0.002734244	P	P
PD22365a	0.006155799811	0.01857085	P	P
PD22366a	0.999533145	0.9887145	D	D
PD23550a	0.004741235687	0.0002885829	P	P
PD23554a	0.9963515185	0.9373199	D	D
PD23558a	0.9945728039	0.9257712	D	D
PD23559a	0.1953678552	0.2685341	P	NC
PD23561a	0.00000627904768	0.00000141229	P	P
PD23562a	0.9998238401	0.9973681	D	D
PD23563a	0.9993416955	0.9936378	D	D
PD23564a	0.00001498101632	0.000003909538	P	P
PD23565a	0.03508900279	0.02881097	P	P
PD23566a	0.9998741697	0.9976837	D	D
PD23567a	0.9983035308	0.9819094	D	D
PD23570a	0.007059955803	0.0005820124	P	P
PD23577a	0.9999134262	0.9990128	D	D
PD23578a	0.9940894141	0.9256367	D	D
PD23579a	0.00001205121118	0.000003531773	P	P
PD24182a	0.9996258969	0.9969335	D	D
PD24186a	0.778758571	0.146323	D	P
PD24189a	0.000008817211782	0.000007344771	P	P
PD24190a	0.01409196844	0.001021789	P	P
PD24191a	0.8995646295	0.2993892	D	NC
PD24193a	0.000005640938725	0.000001579533	P	P
PD24195a	0.0009129459784	0.00001006585	P	P
PD24196a	0.3752606152	0.06197837	NC	P
PD24197a	0.9978130445	0.9791919	D	D
PD24199a	0.009595035402	0.002658904	P	P
PD24200a	0.0271942336	0.001035352	P	P
PD24201a	0.9968961849	0.9878134	D	D
PD24202a	0.9997049834	0.9808006	D	D
PD24204a	0.00006259021472	0.0008046256	P	P
PD24205a	0.9969882864	0.9566222	D	D
PD24206a	0.0006936979798	0.00003811698	P	P
PD24207a	0.02974777694	0.0003737172	P	P
PD24208a	0.01860428757	0.003389219	P	P

PD24209a	0.06915469361	0.002217942	P	P
PD24212a	0.9986556874	0.9964623	D	D
PD24214a	0.002986862016	0.000104998	P	P
PD24215a	0.6086091375	0.7976588	NC	D
PD24216a	0.005538125974	0.003095702	P	P
PD24217a	0.000382986196	0.000006109153	P	P
PD24218a	0.004741735794	0.001770323	P	P
PD24219a	NC	0.0004757906	NC	P
PD24220a	0.001235276456	0.000489168	P	P
PD24221a	0.007817531418	0.01096516	P	P
PD24223a	0.001293771854	0.00003928684	P	P
PD24224a	0.000644756871	0.0005730174	P	P
PD24225a	0.001291907985	0.0001931868	P	P
PD24302a	0.0001567061103	0.0008816913	P	P
PD24303a	0.8696888186	0.5437647	D	NC
PD24304a	0.8739754243	0.5793976	D	NC
PD24306a	0.9660179381	0.7238305	D	NC
PD24307a	0.0005661291349	0.00002716791	P	P
PD24314a	0.003963887816	0.0007513831	P	P
PD24318a	0.02158143765	0.003402058	P	P
PD24320a	0.0002094823758	0.00001118715	P	P
PD24322a	0.0009132615787	0.0006661009	P	P
PD24325a	0.1706384084	0.007348217	P	P
PD24326a	0.0007915688169	0.0000313853	P	P
PD24327a	0.0009995479536	0.00003272696	P	P
PD24329a	0.0002021011997	0.0006728405	P	P
PD24332a	0.002130785586	0.001756998	P	P
PD24333a	0.00233084177	0.002233376	P	P
PD24335a	0.008266288999	0.003648168	P	P
PD24336a	0.0006158844917	0.0001726117	P	P
PD24337a	0.9999800604	0.9999409	D	D

Supplementary Table 2:

#Patient	Davies score	observed	HR_class_Davies	HR_class_obs
PD22036a	0.0003475195706	0.001717023	P	P
PD22251a	0.002538573672	0.002300439	P	P
PD22355a	0.9999328577	0.9999897	D	D
PD22358a	0.9169706398	0.9970982	D	D
PD22359a	0.001753778715	0.0001682815	P	P
PD22360a	0.9747501969	0.9991703	D	D
PD22361a	0.0003891565291	0.0001374701	P	P
PD22362a	0.002184709154	0.0006028422	P	P
PD22363a	0.9723387388	0.9957364	D	D
PD22364a	0.003467522582	0.0004501091	P	P
PD22365a	0.006155799811	0.00861013	P	P
PD22366a	0.999533145	0.9998236	D	D
PD23550a	0.004741235687	0.001194263	P	P
PD23554a	0.9963515185	0.9998951	D	D
PD23558a	0.9945728039	0.9968476	D	D
PD23559a	0.1953678552	0.9216509	P	D
PD23561a	0.00000627904768	0.000001226871	P	P
PD23562a	0.9998238401	0.9994764	D	D
PD23563a	0.9993416955	0.9995987	D	D
PD23564a	0.00001498101632	0.000003336251	P	P
PD23565a	0.03508900279	0.004464632	P	P
PD23566a	0.9998741697	0.9998067	D	D
PD23567a	0.9983035308	0.9994113	D	D
PD23570a	0.007059955803	0.001667251	P	P
PD23577a	0.9999134262	0.9998253	D	D
PD23578a	0.9940894141	0.9993901	D	D
PD23579a	0.00001205121118	0.00000326497	P	P
PD24182a	0.9996258969	0.9998843	D	D
PD24186a	0.778758571	0.9631645	D	D
PD24189a	0.000008817211782	0.000005184506	P	P
PD24190a	0.01409196844	0.0002567859	P	P
PD24191a	0.8995646295	0.9260632	D	D
PD24193a	0.000005640938725	0.000001814321	P	P
PD24195a	0.0009129459784	0.001403952	P	P
PD24196a	0.3752606152	0.8918445	NC	D
PD24197a	0.9978130445	0.9998481	D	D
PD24199a	0.009595035402	0.005657342	P	P
PD24200a	0.0271942336	0.9903061	P	D
PD24201a	0.9968961849	0.999132	D	D
PD24202a	0.9997049834	0.9986693	D	D
PD24204a	0.00006259021472	0.0001751756	P	P
PD24205a	0.9969882864	0.9954829	D	D
PD24206a	0.0006936979798	0.001306471	P	P
PD24207a	0.02974777694	0.001958925	P	P
PD24208a	0.01860428757	0.009047675	P	P
PD24209a	0.06915469361	0.01050596	P	P

PD24212a	0.9986556874	0.9995559	D	D
PD24214a	0.002986862016	0.0007555976	P	P
PD24215a	0.6086091375	0.8827086	NC	D
PD24216a	0.005538125974	0.007401395	P	P
PD24217a	0.000382986196	0.00006932183	P	P
PD24218a	0.004741735794	0.003926461	P	P
PD24219a	NC	0.0001266929	NC	P
PD24220a	0.001235276456	0.0001481508	P	P
PD24221a	0.007817531418	0.1816094	P	P
PD24223a	0.001293771854	0.0004429787	P	P
PD24224a	0.000644756871	0.0006908266	P	P
PD24225a	0.001291907985	0.0008109631	P	P
PD24302a	0.0001567061103	0.0002647536	P	P
PD24303a	0.8696888186	0.998782	D	D
PD24304a	0.8739754243	0.9965404	D	D
PD24306a	0.9660179381	0.9964732	D	D
PD24307a	0.0005661291349	0.0003133627	P	P
PD24314a	0.003963887816	0.009931973	P	P
PD24318a	0.02158143765	0.0008804143	P	P
PD24320a	0.0002094823758	0.00001166255	P	P
PD24322a	0.0009132615787	0.002868998	P	P
PD24325a	0.1706384084	0.02157705	P	P
PD24326a	0.0007915688169	0.00008128636	P	P
PD24327a	0.0009995479536	0.0001568124	P	P
PD24329a	0.0002021011997	0.0001259568	P	P
PD24332a	0.002130785586	0.0002691386	P	P
PD24333a	0.00233084177	0.01384881	P	P
PD24335a	0.008266288999	0.005307105	P	P
PD24336a	0.0006158844917	0.0003981453	P	P
PD24337a	0.9999800604	0.9999823	D	D

Supplementary Table 3:

#Patient	Davies score	observed with metanormal **without** HRDirect	HR_class_Davies	HR_class_observed
PD22036a	0.0003475195706	0.0290589	P	P
PD22251a	0.002538573672	0.0153526	P	P
PD22355a	0.9999328577	0.887577	D	D
PD22358a	0.9169706398	0.102229	D	P
PD22359a	0.001753778715	0.038851	P	P
PD22360a	0.9747501969	0.274668	D	NC
PD22361a	0.0003891565291	0.0247964	P	P
PD22362a	0.002184709154	0.0399398	P	P
PD22363a	0.9723387388	0.11283	D	P
PD22364a	0.003467522582	0.0278286	P	P
PD22365a	0.006155799811	0.0255322	P	P
PD22366a	0.999533145	0.655348	D	NC
PD23550a	0.004741235687	0.0319526	P	P
PD23554a	0.9963515185	0.718769	D	NC
PD23558a	0.9945728039	0.170533	D	P
PD23559a	0.1953678552	0.0546315	P	P
PD23561a	0.00000627904768	0.0169551	P	P
PD23562a	0.9998238401	0.815889	D	D
PD23563a	0.9993416955	0.432633	D	NC
PD23564a	0.00001498101632	0.000336858	P	P
PD23565a	0.03508900279	0.0956286	P	P
PD23566a	0.9998741697	0.501728	D	NC
PD23567a	0.9983035308	0.672668	D	NC
PD23570a	0.007059955803	0.0419873	P	P
PD23577a	0.9999134262	0.823179	D	D
PD23578a	0.9940894141	0.620447	D	NC
PD23579a	0.00001205121118	0.00017393	P	P
PD24182a	0.9996258969	0.484437	D	NC
PD24186a	0.778758571	0.0579696	D	P
PD24189a	0.000008817211782	0.000353399	P	P
PD24190a	0.01409196844	0.0475738	P	P
PD24191a	0.8995646295	0.194261	D	P
PD24193a	0.000005640938725	0.00839089	P	P
PD24195a	0.0009129459784	0.137448	P	P
PD24196a	0.3752606152	0.0973581	NC	P
PD24197a	0.9978130445	0.636173	D	NC
PD24199a	0.009595035402	0.0245687	P	P
PD24200a	0.0271942336	0.15089	P	P
PD24201a	0.9968961849	0.751817	D	D
PD24202a	0.9997049834	0.723072	D	NC
PD24204a	0.00006259021472	0.0445226	P	P
PD24205a	0.9969882864	0.113053	D	P
PD24206a	0.0006936979798	0.0797985	P	P
PD24207a	0.02974777694	0.0255814	P	P

PD24208a	0.01860428757	0.144683	P	P
PD24209a	0.06915469361	0.0699585	P	P
PD24212a	0.9986556874	0.199464	D	P
PD24214a	0.002986862016	0.0243975	P	P
PD24215a	0.6086091375	0.0433253	NC	P
PD24216a	0.005538125974	0.0618186	P	P
PD24217a	0.000382986196	0.020874	P	P
PD24218a	0.004741735794	0.0310872	P	P
PD24219a	NC	0.0484603	NC	P
PD24220a	0.001235276456	0.0721387	P	P
PD24221a	0.007817531418	0.0511289	P	P
PD24223a	0.001293771854	0.0428675	P	P
PD24224a	0.000644756871	0.0448988	P	P
PD24225a	0.001291907985	0.0364603	P	P
PD24302a	0.0001567061103	0.0404194	P	P
PD24303a	0.8696888186	0.395617	D	NC
PD24304a	0.8739754243	0.2298	D	P
PD24306a	0.9660179381	0.205989	D	P
PD24307a	0.0005661291349	0.0173359	P	P
PD24314a	0.003963887816	0.0353762	P	P
PD24318a	0.02158143765	0.0451665	P	P
PD24320a	0.0002094823758	0.0146918	P	P
PD24322a	0.0009132615787	0.0401941	P	P
PD24325a	0.1706384084	0.058887	P	P
PD24326a	0.0007915688169	0.0156699	P	P
PD24327a	0.0009995479536	0.0491703	P	P
PD24329a	0.0002021011997	0.0327167	P	P
PD24332a	0.002130785586	0.0281926	P	P
PD24333a	0.00233084177	0.0376883	P	P
PD24335a	0.008266288999	0.0339594	P	P
PD24336a	0.0006158844917	0.0524107	P	P
PD24337a	0.9999800604	0.860287	D	D

Materials and 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. 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 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 500bp and 130ul AFA tubes. 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.

Design of HRDetect pipeline: from alignment to prediction

With the aim to reproduce as faithfully as possible the original results by Davies and colleagues (Davies et al. 2017), we develop a pipeline that is able to generate an HRDetect score starting from alignment data (Figure 1). Following the description in that paper, our workflow is composed of these steps:

1. annotation of BAM files
2. variant calling: SNV, Indel, copy number alteration and structural variant
3. filtering
4. HRDetect method calling.

Alignment files must be compliant with Cancer Genome Project pipeline that requires them to have ReadGroup 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 simple custom script was coded to annotate properly the raw BAM as specified.

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 Docker and the corresponding Dockerfile was downloaded from https://github.com/cancerit/dockstore-cgpwgs. This file has been modified in order to fix dependencies.

This variant calling step always requires two samples as input: one healthy and one tumoral. The reason is that only somatic variations are needed. To improve the significance of variants, we filtered them retaining only the most confident ones with

a custom script. Raw variations results were converted to match the format requested by the “HRDetect_pipeline” function in “signature.tools.lib” R library (Degasperi et al. 2020) with another custom script. This library has been specifically developed by the authors of Davies’ paper.

HRDirect pipeline for unmatched analyses

Starting from the previous pipeline, we developed a new workflow able to be independent from normal or healthy data. To make limited changes to the HRDetect workflow, in particular to the structure of calling the *cgpwgs* pipeline, we needed a normal sample that could be used for any unmatched patient to be analyzed. So we generated a meta-normal alignment data using 14 healthy WGS from the Davies breast dataset. These data were merged and subsampled to achieve a median total depth of 40x (typical depth for a WGS data) and resulted in a new single BAM file. This will be used instead of the matched (and possibly missing) normal sample as input for the *cgpwgs* workflow.

Four methods have been developed in order to enrich real somatic variants (Figure 4). The first two methods took into account only SNVs or indels whose VAF is in the range 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 (Figure 2 and 3).

A third strategy was evaluated, in which variations are not filtered based on their value (i.e. range filter) but they 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 (Figure 5). Notably, the outcome of HRDirect is no more a number but a classification label.

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) and the PROFILING protocol (001-IRCC-00IIS-10). 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. PDOs #1, #2, #3, #4, #5, #11 and #12 were previously established and characterized as described in (Arena et al. 2020) and (Lorenzato et al. 2020).

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 (Arena, Corti et al. 2020).

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 (Arena et al. 2020). Olaparib response data for patients 1 to 5 are retrieved from previous publication (Arena et al. 2020), while patients 6 to 9 have been *de novo* screened.

Immunohistochemistry (IHC) 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, pattern and intensity of membranous, cytoplasmic and nuclear staining were recorded; for ATM presence or lack of nuclear staining was recorded.

Western blot analysis

Organoids were lysed 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 (Thermo Fisher 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). Detection of the chemiluminescent signal was performed with the ChemiDoc Imaging System (Bio-Rad). The following primary antibodies were used for Western blotting: anti-HSP90 (ABCAM, ab2928; 1:1,000) anti-RAD51C (Santa Cruz Biotechnology SC-56214; 1:1,000) anti-ATM (Cell Signaling Technology, 2873S; 1:1,000).

Homologous recombination Deficiency commercial targeted assays for HRD score analysis

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, small indels and the assessment of microsatellite instability (MSI) status (120 loci), tumor mutational burden (TMB), and copy number (CN) 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, sum of three independent scars: LOH, TAI and LST (Telli et al. 2016). Briefly, following the manufacturers protocol, 150 ng of genomic DNA were used as starter material to generate libraries, while 80 ng of post-fragmentation material were 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 were used for library preparation, and then sequenced on 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 genomic scar score (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.

Acknowledgements

At the end of this journey I am very happy to thank all the people that made this achievement possible.

First of all, I would like to thank Prof. Alberto Bardelli for always believing in me despite my personality: he made it possible for me to undertake this doctorate even if I was no longer at the right age to do so.

I also thank Prof. Sabrina Arena for having led me to the finish line even in the moments when I was inclined to lose myself.

I also thank all my colleagues for all the suggestions and curiosities they showed while the project was taking shape: in particular I remember Giuseppe Rospo among the bioinformaticians and Giovanni Germano and Annalisa Lorenzato among the biologists.

This doctorate did not only have a scientific aspect, but also a human aspect, of introspection, of questioning my abilities. For this reason, I am very grateful to my wife Valentina for accompanying me on this journey, and to my son Matteo who fills the days (not all!) with joy and lightheartedness that is always needed.

A thought also goes to my super parents Roberto and Isa, and my sister Mariella who, among other more important things, made an effort to understand what I was doing.

I thank and remember, even without naming them all, my friends, my PhD fellows and the researchers of the Candiolo Cancer Institute with whom I have often exchanged ideas and visions.

References

Abkevich V, Timms KM, Hennessy BT, Potter J, Carey MS, Meyer LA, Smith-McCune K, Broaddus R, Lu KH, Chen J, Tran TV, Williams D, Iliev D, Jammulapati S, FitzGerald LM, Krivak T, DeLoia JA, Gutin A, Mills GB, Lanchbury JS. Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. Br J Cancer. 2012 Nov 6;107(10):1776-82. doi: 10.1038/bjc.2012.451. Epub 2012 Oct 9. PMID: 23047548; PMCID: PMC3493866.

AlDubayan SH, Giannakis M, Moore ND, Han GC, Reardon B, Hamada T, et al. Inherited DNA-Repair Defects in Colorectal Cancer. Am J Hum Genet. 2018;102(3):401-14.

Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500(7463):415-21.

Arena S, Corti G, Durinikova E, Montone M, Reilly NM, Russo M, et al. A Subset of Colorectal Cancers with Cross-Sensitivity to Olaparib and Oxaliplatin. Clin Cancer Res. 2020;26(6):1372-84.

Birkbak NJ, Wang ZC, Kim JY, Eklund AC, Li Q, Tian R, Bowman-Colin C, Li Y, Greene-Colozzi A, Iglehart JD, Tung N, Ryan PD, Garber JE, Silver DP, Szallasi Z, Richardson AL. Telomeric allelic imbalance indicates defective DNA repair and sensitivity to DNA-damaging agents. Cancer Discov. 2012 Apr;2(4):366-375. doi: 10.1158/2159-8290.CD-11-0206. Epub 2012 Mar 22. Erratum in: Cancer Discov. 2013 Aug;3(8):952. PMID: 22576213; PMCID: PMC3806629.

Bryant HE, Schultz N, Thomas HD, Parker KM, Flower D, Lopez E, et al. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. Nature. 2005;434(7035):913-7.

Byrum AK, Vindigni A, Mosammaparast N. Defining and Modulating 'BRCAness'. Trends Cell Biol. 2019;29(9):740-51.

Curtin NJ. Targeting the DNA damage response for cancer therapy. Biochemical Society Transactions. 2023;51(1):207-21.

Davies H, Glodzik D, Morganella S, Yates LR, Staaf J, Zou X, et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. Nat Med. 2017;23(4):517-25.

Degasperi A, Amarante TD, Czarnecki J, Shooter S, Zou X, Glodzik D, et al. A practical framework and online tool for mutational signature analyses show inter-tissue variation and driver dependencies. Nat Cancer. 2020;1(2):249-63.

Durinikova E, Reilly NM, Buzo K, Mariella E, Chilà R, Lorenzato A, et al. Targeting the DNA Damage Response Pathways and Replication Stress in Colorectal Cancer. Clin Cancer Res. 2022;28(17):3874-89.

Farmer H, McCabe N, Lord CJ, Tutt AN, Johnson DA, Richardson TB, et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature. 2005;434(7035):917-21.

Fong PC, Boss DS, Yap TA, Tutt A, Wu P, Mergui-Roelvink M, et al. Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. N Engl J Med. 2009;361(2):123-34.

Fong PC, Yap TA, Boss DS, Carden CP, Mergui-Roelvink M, Gourley C, et al. Poly(ADP)-ribose polymerase inhibition: frequent durable responses in BRCA carrier ovarian cancer correlating with platinum-free interval. J Clin Oncol. 2010;28(15):2512-9.

Heeke AL, Pishvaian MJ, Lynce F, Xiu J, Brody JR, Chen WJ, et al. Prevalence of Homologous Recombination-Related Gene Mutations Across Multiple Cancer Types. JCO Precis Oncol. 2018;2018.

Kamel D, Gray C, Walia JS, Kumar V. PARP Inhibitor Drugs in the Treatment of Breast, Ovarian, Prostate and Pancreatic Cancers: An Update of Clinical Trials. Curr Drug Targets. 2018;19(1):21-37.

Kaufman B, Shapira-Frommer R, Schmutzler RK, Audeh MW, Friedlander M, Balmaña J, Mitchell G, Fried G, Stemmer SM, Hubert A, Rosengarten O, Steiner M, Loman N, Bowen K, Fielding A, Domchek SM. Olaparib monotherapy in patients with advanced cancer and a germline BRCA1/2 mutation. J Clin Oncol. 2015 Jan 20;33(3):244-50. doi: 10.1200/JCO.2014.56.2728. Epub 2014 Nov 3. PMID: 25366685; PMCID: PMC6057749.

Knijnenburg TA, Wang L, Zimmermann MT, Chambwe N, Gao GF, Cherniack AD, et al. Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. Cell Rep. 2018;23(1):239-54.e6.

Lazzari L, Corti G, Picco G, Isella C, Montone M, Arcella P, et al. Patient-derived xenografts and matched cell lines identify pharmacogenomic vulnerabilities in colorectal cancer. Clinical Cancer Research; 2019.

Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, Lu S, Kemberling H, Wilt C, Luber BS, Wong F, Azad NS, Rucki AA, Laheru D, Donehower R, Zaheer A, Fisher GA, Crocenzi TS, Lee JJ, Greten TF, Duffy AG, Ciombor KK, Eyring AD, Lam BH, Joe A, Kang SP, Holdhoff M, Danilova L, Cope L, Meyer C, Zhou S, Goldberg RM, Armstrong DK, Bever KM, Fader AN, Taube J, Housseau F, Spetzler D, Xiao N, Pardoll DM, Papadopoulos N, Kinzler KW, Eshleman JR, Vogelstein B, Anders RA, Diaz LA Jr. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. Science. 2017 Jul 28;357(6349):409-413. doi: 10.1126/science.aan6733. Epub 2017 Jun 8. PMID: 28596308; PMCID: PMC5576142.

Li H Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM https://arxiv.org/abs/1303.3997, 2013.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.

Lonardi S, Sobrero A, Rosati G, Di Bartolomeo M, Ronzoni M, Aprile G, Massida B, Scartozzi M, Banzi M, Zampino MG, Pasini F, Marchetti P, Cantore M, Zaniboni A, Rimassa L, Ciuffreda L, Ferrari D, Barni S, Zagonel V, Maiello E, Rulli E, Labianca R; TOSCA (Three or Six Colon Adjuvant) Investigators. Phase III trial comparing 3-6 months of adjuvant FOLFOX4/XELOX in stage II-III colon cancer: safety and compliance in the TOSCA trial. Ann Oncol. 2016 Nov;27(11):2074-2081. doi: 10.1093/annonc/mdw404. Epub 2016 Aug 29. Erratum in: Ann Oncol. 2017 Dec 1;28(12):3110. PMID: 27573560.

Lord CJ, Ashworth A. BRCAness revisited. Nat Rev Cancer. 2016 Feb;16(2):110-20. doi: 10.1038/nrc.2015.21. Epub 2016 Jan 18. PMID: 26775620.

Lorenzato A, Magrì A, Matafora V, Audrito V, Arcella P, Lazzari L, Montone M, Lamba S, Deaglio S, Siena S, Bertotti A, Trusolino L, Bachi A, Di Nicolantonio F, Bardelli A, Arena S. Vitamin C Restricts the Emergence of Acquired Resistance to EGFR-Targeted Therapies in Colorectal Cancer. Cancers (Basel). 2020 Mar 14;12(3):685. doi: 10.3390/cancers12030685. PMID: 32183295; PMCID: PMC7140052.

Mangogna A, Munari G, Pepe F, Maffii E, Giampaolino P, Ricci G, et al. Homologous Recombination Deficiency in Ovarian Cancer: from the Biological Rationale to Current Diagnostic Approaches. J Pers Med. 2023;13(2).

Mauri G, Arena S, Siena S, Bardelli A, Sartore-Bianchi A. The DNA damage response pathway as a land of therapeutic opportunities for colorectal cancer. Ann Oncol. 2020 Sep;31(9):1135-1147. doi: 10.1016/j.annonc.2020.05.027. Epub 2020 Jun 5. PMID: 32512040.

Medico E, Russo M, Picco G, Cancelliere C, Valtorta E, Corti G, et al. The molecular landscape of colorectal cancer cell lines unveils clinically actionable kinase targets. Nat Commun. 2015;6:7002.

Minten EV, Yu DS. DNA Repair: Translation to the Clinic. Clin Oncol (R Coll Radiol). 2019;31(5):303-10.

Morales J, Li L, Fattah FJ, Dong Y, Bey EA, Patel M, Gao J, Boothman DA. Review of poly (ADP-ribose) polymerase (PARP) mechanisms of action and rationale for targeting in cancer and other diseases. Crit Rev Eukaryot Gene Expr. 2014;24(1):15-28. doi: 10.1615/critreveukaryotgeneexpr.2013006875. PMID: 24579667; PMCID: PMC4806654.

Moretto R, Elliott A, Zhang J, Arai H, Germani MM, Conca V, Xiu J, Stafford P, Oberley M, Abraham J, Spetzler D, Rossini D, Antoniotti C, Marshall J, Shields A, Lopes G, Lonardi S, Pietrantonio F, Tomasello G, Passardi A, Tamburini E, Santini D, Aprile G, Masi G, Falcone A, Lenz HJ, Korn M, Cremolini C. Homologous Recombination Deficiency Alterations in Colorectal Cancer: Clinical, Molecular, and Prognostic Implications. J Natl Cancer Inst. 2022 Feb 7;114(2):271-279. doi: 10.1093/jnci/djab169. PMID: 34469533; PMCID: PMC8826505.

National Cancer Institute Surveillance, Epidemiology, and End Results Program. Cancer stat facts: colorectal cancer. Accessed September, 2022. https://seer.cancer.gov/statfacts/html/colorect.html.

Nguyen, L., W. M. Martens, J., Van Hoeck, A. *et al.* Pan-cancer landscape of homologous recombination deficiency. *Nat Commun* **11**, 5584 (2020). doi:10.1038/s41467-020-19406-4

Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature. 2016;534(7605):47-54.

Oh M, McBride A, Yun S, Bhattacharjee S, Slack M, Martin JR, Jeter J, Abraham I. BRCA1 and BRCA2 Gene Mutations and Colorectal Cancer Risk: Systematic Review and Meta-analysis. J Natl Cancer Inst. 2018 Nov 1;110(11):1178-1189. doi: 10.1093/jnci/djy148. PMID: 30380096.

Overman MJ, Lonardi S, Wong KYM, Lenz HJ, Gelsomino F, Aglietta M, Morse MA, Van Cutsem E, McDermott R, Hill A, Sawyer MB, Hendlisz A, Neyns B, Svrcek M, Moss RA, Ledeine JM, Cao ZA, Kamble S, Kopetz S, André T. Durable Clinical Benefit With Nivolumab Plus Ipilimumab in DNA Mismatch Repair-Deficient/Microsatellite Instability-High Metastatic Colorectal Cancer. J Clin Oncol. 2018 Mar 10;36(8):773-779. doi: 10.1200/JCO.2017.76.9901. Epub 2018 Jan 20. PMID: 29355075.

Pellegrino B, Mateo J, Serra V, Balmana J. Controversies in oncology: are genomic tests quantifying homologous recombination repair deficiency (HRD) useful for treatment decision making? ESMO Open. 4. England2019. p. e000480.

Pfarr N, von Schwarzenberg K, Zocholl D, Merkelbach-Bruse S, Siemanowski J, Mayr EM, Herold S, Kleo K, Heukamp LC, Willing EM, Menzel M, Lehmann U, Bartels S, Chakraborty S, Baretton G, Demes MC, Döring C, Kazdal D, Budczies J, Rad R, Wild P, Christinat Y, McKee T, Schirmacher P, Horst D, Büttner R, Stenzinger A, Sehouli J, Vollbrecht C, Hummel M, Braicu EI, Weichert W; German HRD assay Harmonization Consortium. High Concordance of Different Assays in the Determination of Homologous Recombination Deficiency-Associated Genomic Instability in Ovarian Cancer. JCO Precis Oncol. 2024 Mar;8:e2300348. doi: 10.1200/PO.23.00348. PMID: 38513168; PMCID: PMC10965219.

Pilié PG, Tang C, Mills GB, Yap TA. State-of-the-art strategies for targeting the DNA damage response in cancer. Nat Rev Clin Oncol. 2019;16(2):81-104.

Popova T, Manié E, Rieunier G, Caux-Moncoutier V, Tirapo C, Dubois T, Delattre O, Sigal-Zafrani B, Bollet M, Longy M, Houdayer C, Sastre-Garau X, Vincent-Salomon A, Stoppa-Lyonnet D, Stern MH. Ploidy and large-scale genomic instability consistently identify basal-like breast carcinomas with BRCA1/2 inactivation. Cancer Res. 2012 Nov 1;72(21):5454-62. doi: 10.1158/0008-5472.CAN-12-1470. Epub 2012 Aug 29. PMID: 22933060.

Robson ME, Tung N, Conte P, Im SA, Senkus E, Xu B, 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. Annals of Oncology. 2019;30(4):558-66.

Sartore-Bianchi A, Trusolino L, Martino C, Bencardino K, Lonardi S, Bergamo F, Zagonel V, Leone F, Depetris I, Martinelli E, Troiani T, Ciardiello F, Racca P, Bertotti A, Siravegna G, Torri V, Amatu A, Ghezzi S, Marrapese G, Palmeri L, Valtorta E,

Cassingena A, Lauricella C, Vanzulli A, Regge D, Veronese S, Comoglio PM, Bardelli A, Marsoni S, Siena S. Dual-targeted therapy with trastuzumab and lapatinib in treatment-refractory, KRAS codon 12/13 wild-type, HER2-positive metastatic colorectal cancer (HERACLES): a proof-of-concept, multicentre, open-label, phase 2 trial. Lancet Oncol. 2016 Jun;17(6):738-746. doi: 10.1016/S1470-2045(16)00150-9. Epub 2016 Apr 20. Erratum in: Lancet Oncol. 2016 Oct;17 (10):e420. PMID: 27108243.

Soyano AE, Baldeo C, Kasi PM. BRCA Mutation and Its Association With Colorectal Cancer. Clin Colorectal Cancer. 2018 Dec;17(4):e647-e650. doi: 10.1016/j.clcc.2018.06.006. Epub 2018 Jul 3. PMID: 30033118.

Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209-49

Telli ML, Timms KM, Reid J, Hennessy B, Mills GB, Jensen KC, Szallasi Z, Barry WT, Winer EP, Tung NM, Isakoff SJ, Ryan PD, Greene-Colozzi A, Gutin A, Sangale Z, Iliev D, Neff C, Abkevich V, Jones JT, Lanchbury JS, Hartman AR, Garber JE, Ford JM, Silver DP, Richardson AL. Homologous Recombination Deficiency (HRD) Score Predicts Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with Triple-Negative Breast Cancer. Clin Cancer Res. 2016 Aug 1;22(15):3764-73. doi: 10.1158/1078-0432.CCR-15-2477. Epub 2016 Mar 8. PMID: 26957554; PMCID: PMC6773427.

Tian T, Shan L, Yang W, Zhou X, Shui R. Evaluation of the BRCAness phenotype and its correlations with clinicopathological features in triple-negative breast cancers. Hum Pathol. 2019 Feb;84:231-238. doi: 10.1016/j.humpath.2018.10.004. Epub 2018 Oct 16. PMID: 30339969.

Turner N, Tutt A, Ashworth A. Hallmarks of ‘BRCAness’ in sporadic cancers. Nat Rev Cancer 2004;4:814-9.

Tutt A, Robson M, Garber JE, Domchek SM, Audeh MW, Weitzel JN, 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. 2010;376(9737):235-44.

Tutt A, Tovey H, Cheang MCU, Kernaghan S, Kilburn L, Gazinska P, Owen J, Abraham J, Barrett S, Barrett-Lee P, Brown R, Chan S, Dowsett M, Flanagan JM, Fox L, Grigoriadis A, Gutin A, Harper-Wynne C, Hatton MQ, Hoadley KA, Parikh J, Parker P, Perou CM, Roylance R, Shah V, Shaw A, Smith IE, Timms KM, Wardley AM, Wilson G, Gillett C, Lanchbury JS, Ashworth A, Rahman N, Harries M, Ellis P, Pinder SE, Bliss JM. Carboplatin in BRCA1/2-mutated and triple-negative breast cancer BRCAness subgroups: the TNT Trial. Nat Med. 2018 May;24(5):628-637. doi: 10.1038/s41591-018-0009-7. Epub 2018 Apr 30. PMID: 29713086; PMCID: PMC6372067.

Van Cutsem E, Cervantes A, Adam R, Sobrero A, Van Krieken JH, Aderka D, Aranda Aguilar E, Bardelli A, Benson A, Bodoky G, Ciardiello F, D'Hoore A, Diaz-Rubio E, Douillard JY, Ducreux M, Falcone A, Grothey A, Gruenberger T, Haustermans K, Heinemann V, Hoff P, Köhne CH, Labianca R, Laurent-Puig P, Ma B, Maughan T,

Muro K, Normanno N, Österlund P, Oyen WJ, Papamichael D, Pentheroudakis G, Pfeiffer P, Price TJ, Punt C, Ricke J, Roth A, Salazar R, Scheithauer W, Schmoll HJ, Tabernero J, Taïeb J, Tejpar S, Wasan H, Yoshino T, Zaanan A, Arnold D. ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. Ann Oncol. 2016 Aug;27(8):1386-422. doi: 10.1093/annonc/mdw235. Epub 2016 Jul 5. PMID: 27380959.

[1]

[1] Some of the figures in this thesis were produced using Biorender®.