ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

ESMO OPEN SCIENCE FOR OPTIMAL CANCER CARE

ORIGINAL ARTICLE

HR-SC—an academic-developed machine learning framework to classify HRD-positive ovarian cancer patients and predict sensitivity to olaparib

L. Beltrame[1†], L. Mannarino[1,2†], A. Sergi[1,3†], A. Velle[4], I. Treilleux[5], S. Pignata[6], L. Paracchini[1,2], P. Harter[7], G. Scambia[8,9‡], F. Perrone[10], A. González-Martin[11,12], R. Berger[13,14], L. Arenare[10], S. Hietanen[15,16], D. Califano[17], S. Derio[18,19], T. Van Gorp[20,21], M. L. Dalessandro[1], K. Fujiwara[22,23], M. Provansal[24,25], D. Lorusso[2,26], P. Buderath[27], M. Masseroli[3], I. Ray-Coquard[25,28,29], E. Pujade-Lauraine[30], C. Romualdi[4], M. D'Incalci[1,2*] & S. Marchini[1]

[1]Laboratory of Cancer Pharmacology, IRCCS Humanitas Research Hospital, Rozzano; [2]Department of Biomedical Sciences, Humanitas University, Pieve Emanuele; [3]Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy; [4]Department of Biology, University of Padova, Padova, Italy; [5]Biopathology Department, Centre Léon Bérard, Lyon, France; [6]Department of Urology and Gynecology, Istituto Nazionale Tumori IRCCS Fondazione G Pascale, Naples, Italy; [7]Department of Gynecology & Gynecologic Oncology, Evang. Kliniken Essen-Mitte, Essen, Germany; [8]Scientific Directorate, Fondazione Policlinico Universitario A Gemelli IRCCS, Rome; [9]Catholic University of Sacred Heart, Rome; [10]Clinical Trial Unit, Istituto Nazionale Tumori IRCCS Fondazione G Pascale, Naples, Italy; [11]Clínica Universidad de Navarra, Madrid; [12]Spanish Ovarian Cancer Research Group (GEICO), Madrid, Spain; [13]Department of Obstetrics and Gynecology, Medical University Innsbruck, Innsbruck; [14]AGO Austria Study Center, Vienna, Austria; [15]Turku University Hospital, Turku, Finland; [16]Nordic Society of Gynaecological Oncology (NSGO-CTU), Copenhagen, Denmark; [17]Microenvironment Molecular Targets Unit, Istituto Nazionale Tumori IRCCS Fondazione G Pascale, Naples; [18]Instituto Europeo di Oncologia, Milan; [19]Mario Negri Gynecologic Oncology (MaNGO) Group, Milan, Italy; [20]University Hospital of Leuven, Leuven Cancer Institute, Leuven; [21]Belgium and Luxemburg Gynaecological Oncology Group (BGOG), Leuven, Belgium; [22]Saitama Medical University International Medical Center, Hidaka; [23]Gynaecological Oncology Trial and Investigation Consortium (GOTIC), Saitama, Japan; [24]Institut Paoli Calmettes, Marseille; [25]Groupe d'Investigateurs Nationaux Pour l'Etude des Cancers Ovariens (GINECO), Paris, France; [26]Gynaecological Oncology Unit, Humanitas Hospital San Pio X, Milan, Italy; [27]Universitätsklinikum Essen, Frauenklinik, Essen, Germany; [28]Centre Léon Bérard, Lyon; [29]University Claude Bernard Lyon I, Lyon; [30]ARCAGY Research, Paris, France

Available online xxx

Background: High-grade serous ovarian cancer (OC) patients with defects in the homologous recombination repair (HRR) pathway benefit from poly (ADP-ribose) polymerase inhibitor (PARPi) maintenance therapy. Clinically approved methods for identifying HRR status suffer from limitations, such as high failure rates and costs, leading to the clinical need for innovative approaches. To this aim, we developed Homologous Recombination Signature Classifier (HR-SC), a machine learning (ML) algorithm that integrates *BRCA1*/*BRCA2* status and copy number signatures, leveraging the availability of OC samples recruited from two international clinical trials, namely PAOLA-1 (dataset A) and MITO16A/MaNGO-OV2 (dataset B).

Patients and methods: 569 DNA samples from datasets A and B were sequenced using a custom library design covering a backbone of structural regions and the full-length sequence of 375 genes. Data were used to train, validate (dataset A), and test (dataset B) HR-SC, using *BRCA1*/*BRCA2* status and a compendium of previously annotated copy number signatures. Lastly, HR-SC was compared with already established approaches to evaluate its predictive and prognostic role.

Results: In dataset A, where the failure rate was 6.4%, HR-SC showed a sensitivity of 92%, a specificity of 94.73%, an accuracy of 93.18%, a positive predictive value (PPV) of 95.83%, and a negative predictive value (NPV) of 90%. In dataset B, where the failure rate was 4%, HR-SC showed a sensitivity of 90.16%, a specificity of 82.86%, an accuracy of 87.5%, a PPV of 90.16%, and an NPV of 82.86%. Univariate and multivariate survival analyses demonstrated its predictive role [progression-free survival (PFS): hazard ratio (HR) $= 0.42$, $P < 0.0001$; overall survival (OS): HR $= 0.63$, $P = 0.036$] and its prognostic role (PFS: HR $= 0.56$, $P = 0.0095$).

Conclusions: The study demonstrates that HR-SC is a novel, clinically feasible solution with a low failure rate for predicting HRR status in OC patients and underscores the importance of leveraging ML approaches for advancing precision oncology in the era of personalized medicine.

Key words: homologous recombination deficiency, machine learning, copy number signatures, ovarian cancer

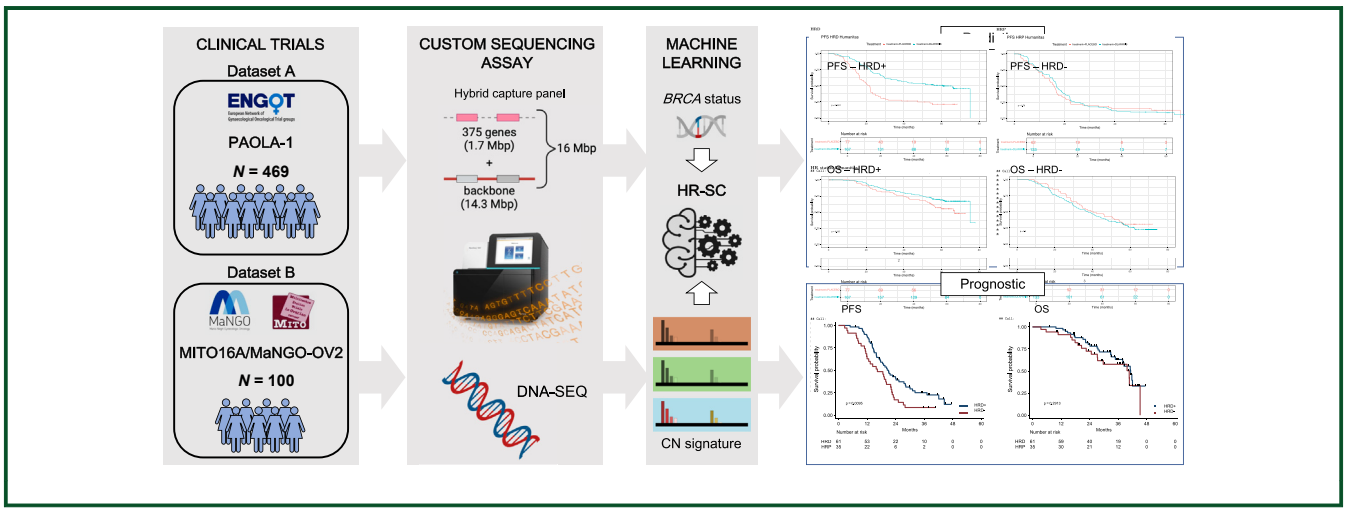
*Correspondence to: Prof. Maurizio D'Incalci, Humanitas Research Hospital, Humanitas University Campus, Building C, Via Rita Levi Montalcini, 4 20072 Pieve Emanuele (MI), Italy. Tel: +39-02-82245259 − 5207

E-mail: maurizio.dincalci@hunimed.eu (M. D'Incalci).

†Co-first authors.

‡Passed away on 20 February 2025.

GRAPHICAL ABSTRACT



INTRODUCTION

Although high-grade serous ovarian cancer (HGS-OC) is a heterogeneous disease, almost 50% of patients share a homologous recombination deficient (HRD-positive) genotype with concomitant sensitivity to both DNA double-strand break-inducing therapeutic agents, such as platinum salts (cDDP) and poly (ADP-ribose) polymerase inhibitors (PARPi). Clinical trials, such as PRIMA1,[1] PAOLA-1,[2] and ATHENA-MONO,[3] demonstrated the predictive value of homologous recombination repair (HRR) analysis in the front-line setting, as cases with an HRD-positive profile can benefit from PARPi treatment and exhibited a significant improvement in terms of progression-free survival (PFS) compared with cases with a functional HRR (homologous recombination proficient, HRD-negative) genotype.

Currently, only two commercial assays have been validated in prospective clinical trials, setting the gold standard for HRR status detection: MyriadMyChoice®CDx (Myriad Genetics, Salt Lake City, UT)[1] and Foundation Focus CDx BRCA LOH (Foundation Medicine, Cambridge, MA)[4,5] assays. Both these assays rely on the identification of pathogenic variants in the *BRCA1/BRCA2* genes, while they employ different approaches to evaluate HRR status: the former relies on a value made by the sum of three different scores [loss of heterozygosity (LOH); large scale transitions (LST); telomeric allelic imbalance (TAI)], while the latter measures a global level of LOH. Although widely spread, these solutions still suffer from many drawbacks such as inconclusive or false-negative results (∼18% in the PAOLA-1 trial) and, ultimately, high costs.[2]

Recently, within the ENGOT initiative, eight European laboratories have been involved in a unique gynecological translational research program to develop an academic HRR test on phase III PAOLA-1 tumor samples[2] to develop new biomarkers for HRD detection aimed at facing the many challenges related to the use of commercial HRD tests. As part of this framework, we have previously developed a streamlined and comprehensive genomic hybrid solution originally labeled as 'Lab-1' assay,[6,7] which achieved a high agreement rate with MyriadMyChoice®CDx of 0.92[7] in samples from the MITO16A/MaNGO-OV2 trial.[7]

The HGS-OC genome is highly unstable and deeply permeated by copy number alterations (CNA).[8] Such alterations do not occur randomly in the genome, but exhibit characteristic and recurrent patterns, 'signatures', indicative of the mutational processes acting upon a cell's DNA. These signatures, commonly referred to as copy number (CN) signatures, are identified by analyzing the frequency and distribution of CNA with whole-genome sequencing (WGS) data. Starting from our previous experience and guided by this evidence, in this work we present the development and the clinical performances of Homologous Recombination Signature Classifier (HR-SC), a machine learning (ML)-based algorithm built on CN signatures derived from hybrid-capture sequencing data from a retrospective and multicentric cohort of 569 samples gathered together from two independent international clinical trials, the PAOLA-1[2] and the MITO16A/MaNGO-OV2.[6]

PATIENTS AND METHODS

Patients and treatments

The PAOLA-1/ENGOT-ov25[2] and MITO16A/MaNGO-OV2[6] trials involved two distinct patient populations, selected to address the two different aims of the study. PAOLA-1 is a randomized, double-blind, international phase III trial that enrolled newly diagnosed serous or endometrioid OC patients [International Federation of Gynecology and Obstetrics (FIGO) stage III or IV] who achieved a complete or partial response to surgery, chemotherapy, or both, resulting in no evidence of disease following first-line platinum-based chemotherapy plus bevacizumab. Patients were further randomly assigned to receive olaparib tablets daily or placebo for up to 24 months. This cohort was used to

evaluate the predictive role of the HRS-SC test. Differently, the MITO16A/MaNGO-OV2 trial was a single-arm, phase IV study that enrolled previously untreated OC patients (FIGO stage IIIB-IV), who received standard-of-care treatment [carboplatin plus paclitaxel][6] combined with bevacizumab (15 mg/kg) on day 1 for six 3-week cycles, followed by bevacizumab monotherapy (15 mg/kg) until progression or unacceptable toxicity, up to a maximum of 22 cycles. This cohort of unselected OC patients was used to assess the prognostic role of the HRS-SC test. Patients enrolled in these trials provided written informed consent to perform additional analyses for research purposes.

Collection, management of specimens, and DNA preparation

A total of 469 formalin-fixed, paraffin embedded (FFPE) OC samples from the PAOLA-1 trial were prepared by investigators from ARCAGY-GINECO and are detailed in Supplementary Table S1, available at https://doi.org/10.1016/j.esmoop.2025.105060 (dataset A). Regarding the PAOLA-1 cohort, the first step for the development of an academic test was a preliminary blind evaluation of the performances of 85 PAOLA-1 harboring *BRCA1/BRCA2* wild type tumors (*BRCA1/BRCA2*-wt) by a comparison with the MyriadMyChoice®CDx dataset followed by a final PFS-based test evaluation on 364 additional patient samples.[9] From the whole population of the MITO16A/MaNGO-OV2 trial, a subset of 100 HGS-OC and endometrioid samples was selected as representative of the enrolled population and processed from the coordinating center as previously reported[7] (dataset B). Cases were randomly selected from the MITO16A/MaNGO-OV2 clinical trial, with matched clinicopathological features representative of the trial population. However, due to technical issues with tissue availability, the dataset was originally enriched with *BRCA1/BRCA2* mutation cases or HRD-positive cases, as originally tested with the Myriad MyChoice® CDx assay.[9] Genomic DNA (gDNA) was purified in two independent pathological units from two to four slides of 5 μm containing >20% of tumor cells using the Qiagen GeneRead DNA FFPE kit (Qiagen, Hilden, Germany). DNA concentration was determined. An aliquot of these samples was sent to Humanitas Research Hospital for analysis. HRR status evaluation was blinded to the MyriadMyChoice®CDx test and clinical information. The institutional review boards of the involved institutions approved the study protocol.

Assay design, library preparation, sequencing, and data analysis

In-house analysis of gDNA was carried out on the same samples isolated from the FFPE tissue block used for the MyriadMyChoice®CDx assay. Assay design, performances, coverage distribution, repeatability, and reproducibility are as published.[7] Libraries were paired-end sequenced at 200× on a NextSeq550 benchtop instrument (Illumina, San Diego, CA). The assay design and more detailed information are available in Supplementary Methods, available at https://doi.org/10.1016/j.esmoop.2025.105060.

CN signatures quantification

CN segments for each sample were collected and assembled, providing for each segment the absolute CN for both the major and minor alleles of heterozygous variants. SigProfilerMatrixGenerator from the SigProfiler software suite[10] was used to build a component-by-sample matrix, which then was processed with nonnegative matrix factorization (NMF[11]) to produce CN signatures. Signatures were fitted to previously published CN signatures by COSMIC.[12] This step was carried out with SigProfilerAssignment (version 0.0.13) from the SigProfiler software suite.[13]

Training set data preparation and ground truth label assignment

As a preparation step for the development of the HR-SC model, CN signature activities for each sample in dataset A were coupled with the respective *BRCA1*/*BRCA2* mutation status. Subsequently, dataset A was partitioned into a training test and a test set using scikit-learn.[14] After merging the CN signature exposures with the *BRCA1/BRCA2* status, we obtained the ground truth labels, i.e. each sample's HRD-positive or HRD-negative status obtained with the previously published assay (Lab-1).

ML model and feature selection

As a first step, we identified the best ML model for our data by carrying out a comprehensive assessment of nine distinct ML models: K-Nearest Neighbors, Support Vector Machine Classifier (SVC), Logistic Regression, Random Forest (RF), Decision Tree, AdaBoost, Gradient Boosting, NuSVC, and Linear SVC. Each model was tested using the CN signatures and the *BRCA1/2* mutation status using the default parameters. The performance of the model was evaluated against the ground truth by computing the F1 score for each model tested. Once the best-performing model was selected, we assessed the influence of the features within the model using SHAP (SHapley Additive exPlanations),[15] excluding features whose impact was minimal along with features that had no biological significance.

Model optimization

The optimal parameter combination for the selected model was determined by carrying out a 'grid search', an exhaustive search of all parameter combinations possible for the model, and then selecting the parameter set that yields the best performance against the ground truth. To ensure the robustness of this approach we employed a stratified k-fold cross-validation, splitting the training data into 10 smaller data sets ('folds'), dividing them into training and test sets, and performing an evaluation of the model. As in the case of model selection, the F1 score was used as the metric to evaluate the best parameters.

Model validation

For a final assessment of the robustness and capabilities of the model, we tested it against a completely independent dataset (dataset B), which was used as a validation set. We assessed the model's performance in predicting the HRR status, employing specific metrics such as the F1 score and Matthews correlation coefficient (MCC). Subsequently, we extended our assessment to include the model's proficiency in stratifying patients, leveraging clinical data from the dataset, and evaluating the survival outcomes in terms of PFS. The predictive value of the HRD-positive/HRD-negative classification was evaluated on dataset A, calculated in terms of PFS between patients treated with olaparib versus placebo. PFS was defined as the time from registration to documented progression according to RECIST criteria, death from any cause, or last follow-up date. Survival curves were calculated using the Kaplan−Meier method and compared using a log-rank test. Hazard ratios (HRs) were estimated using the Cox regression model. The model exhibiting the best association with PFS was chosen as the final model for HR-SC. In the last step, we aimed to investigate the prognostic role of the HR-SC algorithm by combining the HRD-positive/HRD-negative values with survival curves derived from cases enrolled in dataset B. Similarly to the analysis carried out on dataset A, we investigated the HR-SC model prognostic value in terms of PFS, by comparing HRD-positive and HRD-negative samples. HRs were estimated using the Cox regression model. In the multivariate models, the following covariates were added: age ($<$65 versus $\geq$65 years); Eastern Cooperative Oncology Group (ECOG) performance status (PS; 0 versus 1-2); residual disease ($\leq$1 cm versus none and $>$1 cm not operated versus none); and FIGO stage (III versus IV).

Statistics

Both the performances of the Lab-1 assay versus the gold standard (MyriadMyChoice®CDx test) as well as of the ML models were assessed using specific parameters including specificity, accuracy, positive predictive value (PPV), negative predictive value (NPV), recall (sensitivity), F1 score, K-Cohen, MCC, and area under the curve (AUC). R-squared statistic was used to compare the results of the Lab-1 test with the MyriadMyChoice®CDx test.

RESULTS

Cohort description and workflow of the study

As part of the ENGOT initiative and being directly involved in the translational programs of the MITO and MaNGO groups, we had access to a unique and retrospective collection of 569 DNA samples purified from FFPE tumor biopsies taken from OC cases, naive to chemotherapy and enrolled in two international clinical trials: the PAOLA-1 (referred to as dataset A, 469 samples) and the MITO16A/MaNGO-OV2 trial (referred to as dataset B, 100 samples). Patients' clinical and demographic characteristics were as previously reported,[2,6,7] and are detailed in Supplementary Table S1, available at https://doi.org/10.1016/j.esmoop.2025.105060. Overall, samples were from HGS-OC tumors (95.6%, 544/569), and were endometrioid (2.98%, 17/569), or undifferentiated (0.87%, 5/569). For three OC patients (0.52%, 3/569), the histological type could not be determined by the pathologists.

Figure 1 summarizes the workflow of the study. To identify cases as HRD-positive or HRD-negative, all samples enrolled in the study were originally analyzed according to the MyriadMyChoice®CDx test (considered from now onwards as the external benchmark)[2,4] and further profiled with a previously 'in-house'-developed hybrid capture-based solution (originally identified as Lab-1 workflow,[7] Supplementary Figure S1, available at https://doi.org/10.1016/j.esmoop.2025.105060). Independent confirmation of the *BRCA1/BRCA2* status and the HRD score was not available until the experiments were fully completed, as part of the policy set by the ENGOT initiative. Considering dataset A, the failure rate for the external benchmark workflow was 9.38% (44/469 samples) and 6.4% (30/469 samples) for the internal benchmark. Next, to establish a novel metric to assess HRR status, we reasoned that since the genome of OC is deeply permeated by somatic copy number aberrations (SCNAs), analysis of CN signatures could be informative. We developed a novel HRR data-driven algorithm, HR-SC, to derive HRR status using an ML approach to select those CN signatures able to correctly stratify HRD-positive and HRD-negative cases. Dataset A was used to assess the predictive value of HR-SC stratification, particularly to evaluate whether patients with an HRD-positive genotype significantly benefit from PARPi treatments, and whether those with an HRD-negative genotype have little or no benefit. Dataset B was used to assess the prognostic role of HR-SC classification, specifically whether patients labeled as HRD-positive survive longer in terms of both PFS and OS than those labeled as HRD-negative. Finally, the concordance of HRD-positive/HRD-negative classification based on CN signatures was compared with the classification based on internal and external benchmark metrics.

In conclusion, the graphical summary in Figure 1 illustrates that the unique 'compendium' of gDNA from two international clinical trials, providing a valuable opportunity to develop a novel academic approach for detecting HRD-positive/HRD-negative patients and to compare its analytical and clinical performance with conventional benchmarks in terms of both predictive and prognostic value.

Definition of the internal benchmark

The analytical and clinical performances of Lab-1 workflow were initially evaluated in dataset A and compared with the external benchmark. As reported in Supplementary Table S1, available at https://doi.org/10.1016/j.esmoop.2025.105060, for both of the two assays the HRD-positive/HRD-negative pattern was obtained by combining the *BRCA1/BRCA2* mutational status with the genomic instability score (GIS) value.

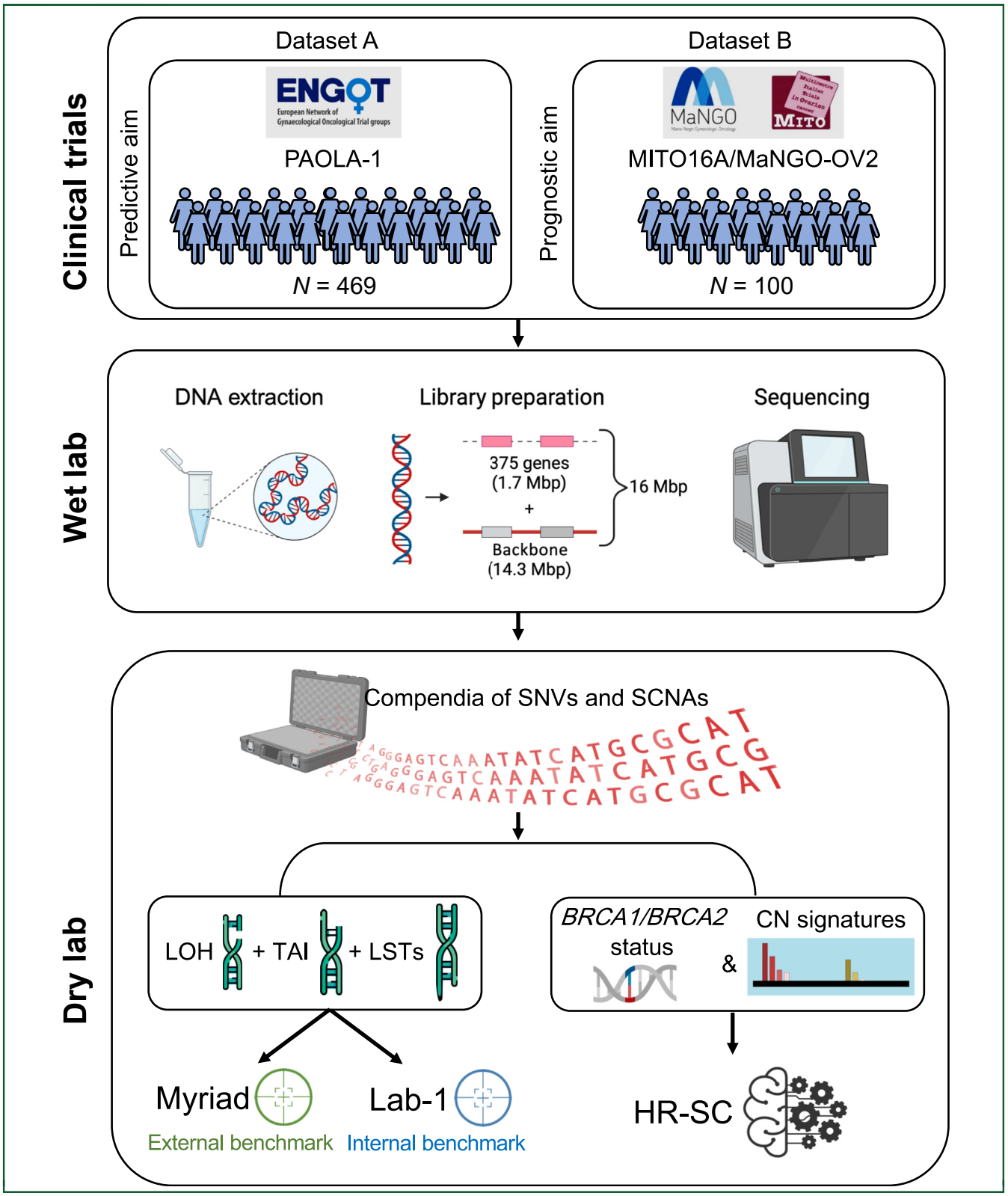


Figure 1. Workflow for the development of Homologous Recombination Signature Classifier (HR-SC) algorithm. The figure summarizes the main experimental (wet lab) and analytical (dry lab) steps in which the study has been organized. Samples used in this work come from two randomized international clinical trials, the PAOLA-1 (dataset A), and the MITO16A/MaNGO-OV2 (dataset B). The first was used to assess the HR-SC predictive role, while the second the prognostic value. Patients' DNA was analyzed with an 'in-house'-designed hybrid-capture sequencing solution covering 375 full-length genes and a backbone of genomic regions from which 'compendia' of single nucleotide variants (SNVs) and copy number (CN) aberrations have been derived. HRD-positive/HRD-negative labels were defined (dataset A) or were available (dataset B) through the external benchmark (MyriadMyChoice®CDx) or the internal benchmark (Lab-1). *BRCA1/BRCA2* status and CN signatures were the basis for the development of a new homologous recombination deficiency (HRD) test called HR-SC based on machine learning approaches. Partially created with BioRender.com.

LOH, loss of heterozygosity; LST, large-scale transition; SCNA, somatic copy number aberrations; TAI, telomeric allelic imbalance.

https://doi.org/10.1016/j.esmoop.2025.105060

Among the 469 cases analyzed with the external benchmark, the *BRCA1/BRCA2* status was available for 98.5% of cases, the GIS value was assigned to 87.84% of cases, and the final HRD-positive/HRD-negative call was determined in 90.61% of cases (Supplementary Table S2, available at https://doi.org/10.1016/j.esmoop.2025.105060). In the case of Lab-1 analysis, the *BRCA1/BRCA2* status, GIS, and HRD-positive/HRD-negative status were available for 438 out of 469 cases (93.39%, Supplementary Table S2, available at https://doi.org/10.1016/j.esmoop.2025.105060). Additional information is reported in Supplementary Table S1, available at https://doi.org/10.1016/j.esmoop.2025.105060 As detailed in Supplementary Figure S1 and Table S2, available at https://doi.org/10.1016/j.esmoop.2025.105060, 85.07% of cases ($n = 399$) were successfully tested with both assays.

Lab-1 identified *BRCA1/BRCA2* mutations in 29.82% ($n = 119$), HRD-positive in 59.39% ($n = 237$), and HRD-negative in 40.60% ($n = 162$) of cases (Supplementary Tables S3A and B, available at https://doi.org/10.1016/j.esmoop.2025.105060). The Lab-1 performances against the external benchmark in *BRCA1/BRCA2* mutation calling were sensitivity 83.60% (95% CI 80.06% to 87.31%); specificity 99.61% (95% CI 99% to 100%); and Cohen's kappa score 0.86 (95% CI 0.82-0.89) (Supplementary Table S3A, available at https://doi.org/10.1016/j.esmoop.2025.105060). When comparing the HRR status determined by the Lab-1 test with the external benchmark HRR status, sensitivity was 93.85% (95% CI 91.49% to 96.20%), specificity 86.54% (95% CI 83.20% to 89.89%), and Cohen's kappa score 0.81 (95% CI 0.77-0.85) (Supplementary Table S3B, available at https://doi.org/10.1016/j.esmoop.2025.105060). Lastly, the correlation between the two tests in the identification of HRD-positive/HRD-negative cases was 0.88 ($P < 0.0001$).

With regards to the clinical performance of the Lab-1 workflow, Kaplan—Meier curves showed that HRD-positive cases receiving olaparib had significantly better PFS (HR = 0.44, 95% CI 0.31-0.62, $P < 0.0001$) compared with those receiving placebo, while the PFS curves for the HRD-negative cases were comparable between the two treatments (HR = 0.97, 95% CI 0.677-1.39, $P = 0.88$; Supplementary Figure S2, available at https://doi.org/10.1016/j.esmoop.2025.105060). With regards to overall survival (OS), HRD-positive cases receiving olaparib had significantly better survival than those receiving placebo (HR = 0.63, 95% CI 0.41-0.97, $P = 0.036$), while there was no significant difference in OS among HRD-negative patients receiving either treatment (HR = 1.161, 95% CI 0.78-1.73, $P = 0.46$).

The results were comparable with the data from the external benchmark tested on the same 399 samples. Kaplan—Meier showed that HRD-positive cases receiving olaparib had a better PFS (HR = 0.42; 95% CI 0.29-0.62, $P < 0.0001$) compared with those receiving placebo, while the PFS curves for the HRD-negative cases were comparable between the two treatments (HR = 0.96, 95% CI 0.67-1.38, $P = 0.83$) (Supplementary Figure S3, available at https://doi.org/10.1016/j.esmoop.2025.105060). Additionally, HRD-positive cases receiving olaparib had significantly better OS than those receiving placebo (HR = 0.59, 95% CI 0.38-0.90, $P = 0.016$), while there was no significant difference in OS among HRD-negative patients receiving either treatment (HR = 1.19, 95% CI 0.79-1.77, $P = 0.41$) (Supplementary Figure S3, available at https://doi.org/10.1016/j.esmoop.2025.105060).

These results support the notion that our 'in-house'-developed hybrid-capture solution was robust enough to correctly interrogate the genome for the presence of genomic alterations due to defects in the HRR pathway and thus could be used to develop novel approaches to identifying HRD-positive/HRD-negative OC cases. From this point onwards, Lab-1 workflow was referred to as the internal benchmark of the study.

Being consistent with previous findings, raw data from dataset A (439 of 469 cases) were merged with the genomic profiles of cases previously profiled with the same hybrid solution from dataset B (97 of 100 cases),[7] thus generating a unique 'compendium' of genomic rearrangements and single nucleotide variants across a total of 536 samples enrolled into two international clinical trials, worthy for downstream analysis (Supplementary Table S4 and Figure S1, available at https://doi.org/10.1016/j.esmoop.2025.105060).

CN signatures as descriptors of HGS-OC genomic architecture

Since CN signatures represent characteristic patterns of genomic alterations that occur throughout the genome during tumor evolution, they could be used to identify the consequences of defects in the HRR pathway. The computational framework employed to extract CN signatures[12] identified and quantified the 24 COSMIC CN signatures (version 3.3). The biological characteristics of these signatures are as previously reported.[12] As shown in Supplementary Figure S4, available at https://doi.org/10.1016/j.esmoop.2025.105060, the most prevalent CN signatures in both dataset A and B are CN9, indicative of focal loss of heterozygosity (fLOH) and chromosomal instability, and CN17, which reflects the HRD-positive phenotype and tandem duplications. CN1 is associated with diploid genomes, while CN2 is related to tetraploidy. Both of these signatures are without a known etiology. These findings suggest that the protocol for calling CN signatures is robust and capable of identifying features relevant to the biological issues under investigation.

Construction of the HR-SC algorithm

To build a novel, robust, accurate, and reproducible HRR classification algorithm using an ML approach, we partitioned datasets A and B in different ways. Dataset A was used to construct the model and was divided into training and test sets, comprising 80% ($n = 351$) and 20% ($n = 88$) of the total samples, respectively. Survival analysis was focused on assessing the predictive role of the HR-SC algorithm in olaparib treatment. Differently, since cases from

dataset B were exposed to chemotherapy (carboplatin + paclitaxel) plus the antiangiogenic agent bevacizumab, the prognostic role the HR-SC algorithm was assessed. Following ML best practices, we undertook three pivotal steps: (i) feature selection; (ii) model construction and refinement; and (iii) model validation.

Feature selection

The feature selection step identifies the most relevant genomic characteristics from a larger set of potential inputs. In the training set, we used this approach to extract the most important variables for determining HRR status from the previously identified 24 CN signatures. Additionally, due to their significant biological role in HRR activity, we included *BRCA1/BRCA2* status (Supplementary Table S1, available at https://doi.org/10.1016/j.esmoop.2025.105060), resulting in a total of 25 features to be tested. The feature selection process was conducted using the RF ML algorithm, which was chosen as the best among nine different ML models (Supplementary Table S5, available at https://doi.org/10.1016/j.esmoop.2025.105060) based on the F1 score statistic. This process identified a subset of 16 features—*BRCA1/BRCA2* mutation status, CN17, CN13, CN2, CN9, CN8, CN1, CN14, CN10, CN12, CN3, CN11, CN7, CN4, CN6, and CN16—as the most relevant for HRR status classification without affecting the performance of the algorithm (Supplementary Figures S5 and S6, respectively, available at https://doi.org/10.1016/j.esmoop.2025.105060).

Model construction and refinement

After the feature selection process, both the training and test sets were utilized for model fine-tuning. Compared with the internal benchmark, the fine-tuned HR-SC model showed a sensitivity of 92.42% (95% CI 89.16% to 94.83%), a specificity of 91.95% (95% CI 88.59% to 94.42%), an accuracy of 90.07% (95% CI 86.94% to 93.19%), and a Cohen's kappa score of 0.83 (95% CI 0.79-0.86) on the training set (Supplementary Table S6, available at https://doi.org/10.1016/j.esmoop.2025.105060). In the test set evaluation, the model demonstrated a sensitivity of 92% (95% CI 87.66% to 98.33%), a specificity of 94.73% (95% CI 90.07% to 99.45%), and an accuracy of 93.18% (95% CI 87.92% to 98.45%). The Cohen's kappa score was 0.86 (95% CI 0.78-0.93) (Supplementary Table S6, available at https://doi.org/10.1016/j.esmoop.2025.105060). In this configuration, HR-SC was unable to determine the HRR status for 30 of 469 samples, corresponding to a failure rate of 6.4% (Supplementary Table S1, available at https://doi.org/10.1016/j.esmoop.2025.105060). In conclusion, the HR-SC algorithm is an RF ML-based model to infer the HRD-positive/HRD-negative status of OC cases, that relies on 15 CN signatures and the *BRCA1/BRCA2* status.

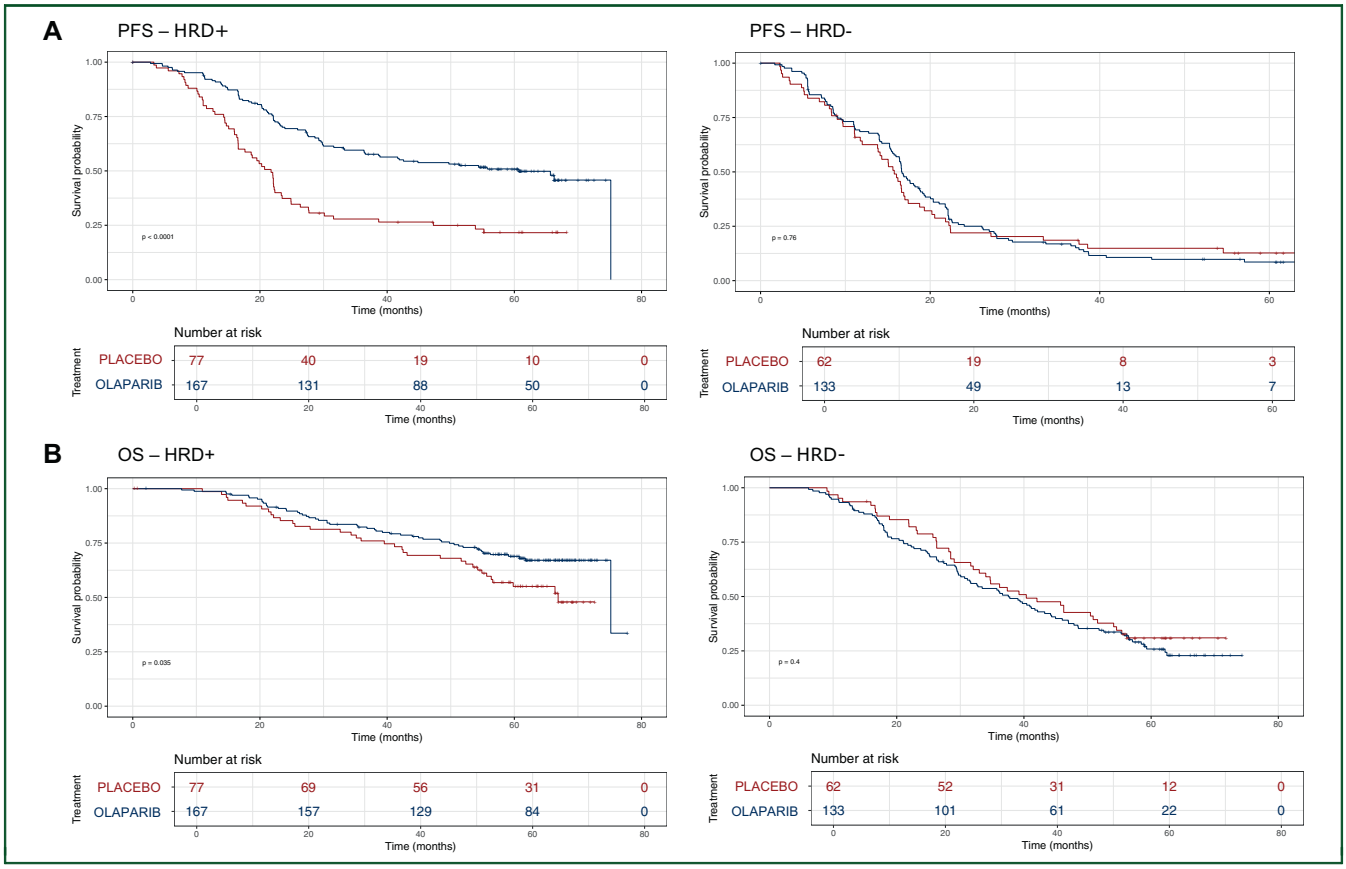


Figure 2. Kaplan–Meier (KM) curves and statistical analysis based on Homologous Recombination Signature Classifier (HR-SC) classification on dataset A. KM curves for progression-free survival (PFS) (A) and overall survival (OS) (B) to compare olaparib-treated versus untreated samples stratified as homologous recombination deficiency (HRD)-positive or HRD-negative according to the HR-SC algorithm. Survival is reported in months (x-axis).

Table 1. Independent predictive role of HR-SC on dataset A

	PFS			OS		
	HR	95% CI	*P* value	HR	95% CI	*P* value
HRD-positive versus HRD-negative	0.36	0.25-0.51	<0.0001	0.39	0.26-0.59	<0.0001
BRCA1/BRCA2 status	0.57	0.37-0.89	0.013	0.53	0.30-0.92	0.0243

Multivariate analysis of PFS and OS in ovarian cancer samples from dataset A, considering cases as HRD-positive or HRD-negative according to the HRS-SC algorithm and *BRCA1/BRCA2*.

CI, confidence interval; HR, hazard ratio; HRD-positive, homologous recombination deficiency; HRD-negative, homologous recombination proficiency.

Then, to explore the clinical utility of the HR-SC algorithm, we tested its predictive role by integrating the molecular HRD-positive/HRD-negative stratification with clinical data from the PAOLA-1 trial.[2] Kaplan–Meier curves depicted in Figure 2A show that HRD-positive cases predicted by the HR-SC algorithm had significantly better PFS when treated with olaparib compared with placebo (HR = 0.42, 95% CI 0.30-0.60, $P < 0.0001$). No significant differences were observed for HRD-negative samples between olaparib and placebo ($P = 0.76$). Moreover, the Cox proportional hazard model demonstrated a significant interaction between the HRD-positive status and olaparib treatment (HR 0.45, 95% CI 0.28-0.72, $P = 0.0009$). OS in HRD-positive patients treated with olaparib was significantly improved compared with placebo (HR 0.63, 95% CI 0.41-0.97, $P = 0.035$). In contrast there was no significant difference in OS among HRD-negative patients receiving either treatment ($P = 0.4$; Figure 2B). Multivariate analysis confirmed the independent predictive role of HR-SC, in terms of both PFS (HR 0.36, 95% CI 0.25-0.51, $P < 0.0001$) and OS (HR 0.39, 95% CI 0.26-0.59, $P < 0.0001$) (Table 1).

HR-SC validation and prognostic evaluation of HRR status

We subsequently applied HR-SC stratification to cases enrolled in dataset B (Supplementary Table S1, available at https://doi.org/10.1016/j.esmoop.2025.105060) to assess both its analytical performances and prognostic role. Compared with the internal benchmark, HR-SC reported a sensitivity of 90.16% (95% CI 84.21% to 96.12%), a specificity of 82.86% (95% CI 75.32% to 90.40%), an accuracy of 87.50% (95% CI 80.88% to 94.12%), and a Cohen's kappa score of 0.73 (95% CI 0.64-0.82) (Supplementary Table S7, available at https://doi.org/10.1016/j.esmoop.2025.105060). In 4 of 100 samples, HR-SC failed to call an HRR status, corresponding to a failure rate of 4%, (Supplementary Table S1, available at https://doi.org/10.1016/j.esmoop.2025.105060).

Subsequently, from a clinical perspective, we evaluated the prognostic value of HR-SC by analyzing the risk of relapse and survival of HRD-positive and HRD-negative groups of patients treated with chemotherapy (carboplatin + paclitaxel) and the antiangiogenic agent bevacizumab. As shown in Figure 3A, HRD-positive and HRD-negative samples had significantly different PFS in both univariate (HR 0.56, 95% CI 0.35-0.90, $P = 0.0095$) and multivariate analysis (HR 0.53, 95% CI 0.33-0.85, $P = 0.009$, Table 2). Univariate analysis did not report differences in terms of OS (HR 0.71, 95% CI 0.38-1.33, $P = 0.2913$) (Figure 3B).

Comparison with the external benchmark

Given that the external benchmark is the current gold-standard assay for the determination of HRR status, we

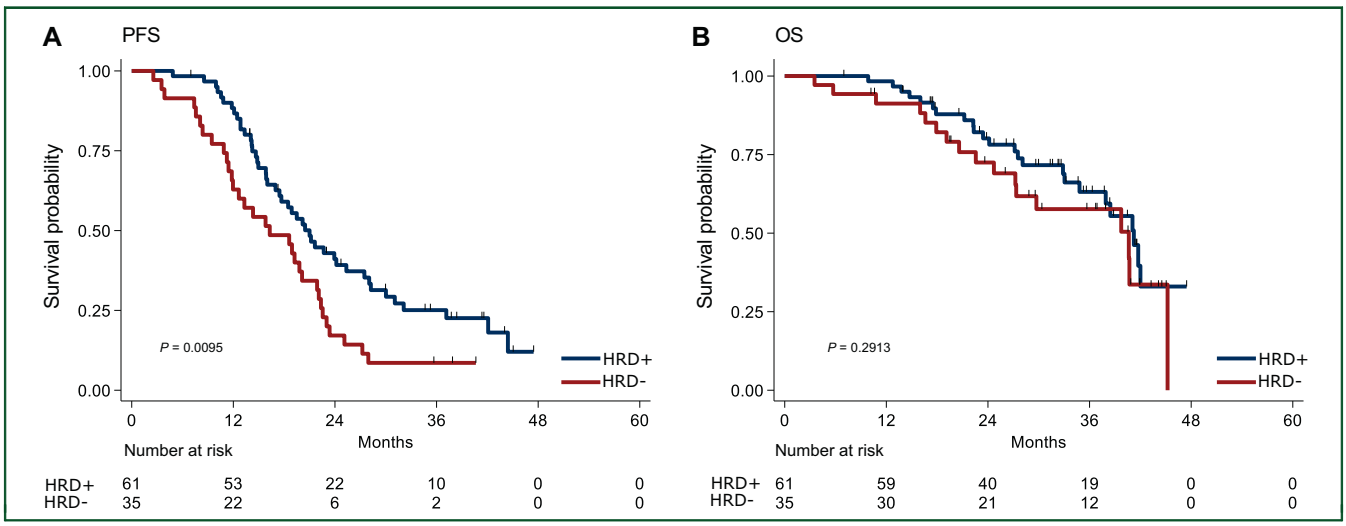


Figure 3. Kaplan–Meier (KM) curves and statistical analysis based on Homologous Recombination Signature Classifier (HR-SC) classification on dataset B. KM curves of progression-free survival (PFS) (A) and overall survival (OS) (B) in dataset B according to the homologous recombination repair status as determined by HR-SC. Survival is reported in months (x-axis).

HRD+, homologous recombination deficiency; HRD–, homologous recombination proficiency.

Table 2. Independent prognostic role of HR-SC on dataset B

	PFS			
	HR	*P*	95% CI	
HRD-positive versus HRD-negative	0.53	0.009	0.33	0.85
Age (<65 years versus ≥65 years)	0.72	0.228	0.43	1.22
PS: 0 versus 1-2	1.13	0.705	0.6	2.10
Residual tumor (≤1 cm versus none)	1.16	0.628	0.63	2.14
Residual tumor (>1 cm/not operated versus none)	2.03	0.025	1.09	3.74
FIGO stage: III versus IV	2.37	0.003	1.34	4.20

Multivariate analysis of progression-free survival (PFS) in ovarian cancer samples from dataset B, considering cases as HRD-positive or HRD-negative according to the HRS-SC classifier.
CI, confidence interval; FIGO, International Federation of Gynecology and Obstetrics; HR, hazard ratio; HRD-positive, homologous recombination deficiency; HRD-negative, homologous recombination proficiency; PS, Eastern Cooperative Oncology Group performance status.

compared the HR-SC performances with the external benchmark calls on both datasets A and B. In dataset A, 399/439 samples had HRR status available for both assays (Supplementary Table S2; Supplementary Figure S3, available at https://doi.org/10.1016/j.esmoop.2025.105060). The analytical evaluation of the assay using the external benchmark as ground truth yielded a sensitivity of 89.03% (95% CI 85.96% to 92.10%), and a specificity of 82.50% (95% CI 78.72% to 86.18%), with a good agreement between the two calls, supported by a Cohen's kappa score of 0.72 (95% CI 0.675-0.763) (Table 3). For dataset B, results from both assays were available for 91/96 samples (Supplementary Figure S3, available at https://doi.org/10.1016/j.esmoop.2025.105060). The comparison between the two tests achieved a sensitivity of 90.74% (95% CI 84.78% to 96.66%) and a specificity of 72.97% (95% CI 63.84% to 82.09%), while the Cohen's kappa score was 0.65 (95% CI 0.55-0.75) (Table 3). These results sustain the comparable performance of HR-SC to the external benchmark in assessing the HRR status.

DISCUSSION

In this study, we report the development and the clinical performances of HR-SC, an academic ML-based assay that encapsulates *BRCA1/BRCA2* status and 15 COSMIC CN signatures within a single framework, using sequencing data with low complexity, thus ensuring a short turnaround time, a low cost per sample, and a lower failure rate compared with the external benchmark. Consequently, HR-SC emerges as a viable tool for integration into routine clinical practice, helping clinicians in the selection of the most appropriate maintenance strategy for newly diagnosed HGS-OC, if any: PARPi alone, bevacizumab alone, or a combination of PARPi and bevacizumab.

In recent years, both profit and nonprofit initiatives have made significant strides in developing novel genomic biomarkers predictive of PARPi treatment efficacy. Three distinct philosophies for HRR deficiency detection have emerged globally. The first, reliant on gene panel analysis, has encountered limitations during clinical validation.[9] In line with this, the panel of 375 genes concomitantly sequenced in this study did not provide any additional information on the mutational status of genes directly or indirectly involved in the mechanism of DNA repair and damage sensing that could be exploited for prognostic or predictive purposes. The second approach, focusing on whole-genome aberration analysis, has gained traction, particularly with the advent of low-pass WGS. Thirdly, a functional assay targeting *RAD51* has been proposed as an alternative measure for HRR status.[16]

Although these two strategies have traditionally been viewed as mutually exclusive, a novel 'comprehensive approach' combining genomic and functional assays for HRD-positive assessment has recently been evaluated in the MITO16A/MaNGO-OV-2 trial, offering a new perspective on enhancing patient stratification in OC.[17] Genomic tests provide a static measure of accumulated DNA damage, while functional assays, such as the *RAD51* test, offer

Table 3. Analytical performances of HR-SC compared with the external benchmark

Dataset A				**Dataset B**			
External benchmark	HR-SC		Total	External benchmark	HR-SC		Total
	HRD-positive	HRD−			HRD-positive	HRD−	
HRD-positive	203	25	228	HRD-positive	49	5	54
HRD-negative	233	141	171	HRD-negative	10	27	37
Total	233	166	399	Total	59	32	91
	Estimation	**95% CI**			**Estimation**	**95% CI**	
Sensitivity	89.03%	85.96% to 92.10%		Sensitivity	90.74%	84.78% to 96.66%	
Specificity	82.5%	78.72% to 86.18%		Specificity	72.97%	63.84% to 82.09%	
Accuracy	86.2%	82.83% to 89.59%		Accuracy	83.51%	75.89% to 95.11%	
PPV	87.1%	83.83% to 90.41%		PPV	83.05%	75.34% to 90.75%	
NPV	84.9%	81.434% to 88.4%		NPV	84.37%	76.19% to 91.83%	
F1 score	88.1%	85.03% to 90.93%		F1 score	86.72%	79.75% to 93.69%	
K-Cohen	0.72	0.67-0.76		K-Cohen	0.65	0.55-0.75	
MCC	0.72	0.67-0.76		MCC	0.65	0.55-0.73	
AUC	0.85	0.80-0.87		AUC	0.81	0.72-0.88	

The analytical performance of HR-SC was compared with the external benchmark on both datasets A and B, respectively.
AUC, area under the curve; CI, confidence interval; HRD-negative, homologous recombination proficiency; HRD-positive, homologous recombination deficiency; MCC, Matthews correlation coefficient; NPV, negative predictive value; PPV, positive predictive value.

real-time insights into HRR activity. This integration captures both inherited and dynamic aspects of HRD-positive status, refining patient identification for targeted therapies like PARP inhibitors.[17]

The uniqueness of the HR-SC solution lies in the light of a hybrid capture-based approach (16 Mbp in size), encompassing a backbone surveying the entire genome (14.3 Mbp in size) and a set of probes (1.7 Mbp in size) targeting the full-length sequencing of 375 genes implicated in various DNA repair pathways (HRR, nonhomologous end joining, base excision repair, mismatch repair), cell signaling, cell cycle regulation, and PARPi resistance. HR-SC boasts three distinct features setting it apart from previous solutions: (i) derivation of HRD-positive/HRD-negative labels through CN signatures rather than gene panels or genomic scar metrics; (ii) utilization of an ML-based approach; (iii) extraction of CN signatures from a dataset derived from a targeted probe solution, a departure from the traditional WGS or whole-exome sequencing approaches. The last point is extremely important for the potential clinical implementation of the HR-SC in diverse hospital settings. Unlike CN signatures derived from WGS datasets (such as those published by COSMIC), HR-SC leverages a more flexible and compatible hybrid capture-based next-generation sequencing (NGS) panel. This approach optimizes computational algorithms to identify recurrent patterns of SCNA alterations across tumor samples, providing valuable insights into HRR status in OC.

Despite the plethora of studies demonstrating the biological value of CN signature analysis and in particular their relationship with HRR status in HGS-OC,[12,18,19] the clinical utility of CN signatures has been limited till now. Recurrent patterns of CNAs across tumor samples are identified through sophisticated computational algorithms that analyze large-scale genomic data, such as WGS, that, due to their high cost, long turnaround time, and deep coverage, prevented them from being implemented in the vast majority of clinical centers.[18,20,21] On the other hand, HR-SC has been developed on a scalable genomic dataset obtained from a hybrid capture-based NGS panel in which probes have been previously tailored to survey the mutational profile of a selection of 375 genes with concomitant analysis of customized regions affected by SCNAs.[7] This solution has many advantages: (i) in a single experiment, we can provide information on genomic alterations that affect the structure of the genome, as well as the sequence of genes, such as *BRCA1/BRCA2*, *RAD51* and the genes belonging to the *RAD51* family, *PALB2*, and a selection of other genes involved in DNA repair or PARPi resistance, such as *TP53BP1*[22]; (ii) it reduces the size of data storage at the selected coverage (200×); (iii) a turnaround time compatible with the clinical need; (iv) a low cost per sample as almost 24 libraries can be barcoded in a single run on a benchtop sequencer. The use of this low-complexity genome dataset required a priori optimization of the current algorithms to call CN signatures congruent (in terms of number and biological functions) with those previously identified.[11] Within the 24 identified CN signatures, CN1 refers to the diploid signature,[12] CN2 is linked to tetraploidy,[12] and CN3 to octoploidy.[12] CN4-CN8 are signatures associated with chromothripsis,[12,23] CN9-CN12 to fLOH[14], while CN13-CN16 are associated with chromosomal LOH (cLOH).[24-26] Both CN9 and CN17 are strongly associated with promoter hypermethylation of *BRCA1*.[12] These findings support the use of the COSMIC CN signatures with our alternative sequencing approach.

In recent years, ML has emerged as a powerful tool for extracting meaningful features from complex genomic data, particularly in characterizing processes like HRR. ML offers advantages such as a comprehensive analysis of SCNAs, capturing subtle patterns, such as focal and broad-scale events, or iterative refinement through training on labeled datasets to detect relationships in the data that may not be evident through traditional statistical methods. By training ML models on labeled datasets, ML identifies patterns associated with HRD-positive status and generalizes that knowledge to make predictions on unseen data.

It is important to note that ML approaches for measuring CN signatures for HRD-positive status are not fully developed and have certain limitations. One challenge is the need for high-quality, curated datasets for training ML models. To address this challenge we used two independent datasets of tumor biopsies, naive to chemotherapy, that derive from international clinical trials, with well-curated clinical records. ML-based approaches allow for iterative refinement. As more data become available, the models can be updated and retrained to refine their accuracy and effectiveness in predicting HRR status from CN signatures. Although the optimal correlation with the external benchmark is important for the validation of a new test, a correlation with patient outcome in prospective studies is more relevant and mandatory for the future development of the HR-SC test. To overcome this limitation, HR-SC has been included in two prospective clinical trials such as IOlanTHe (clinical trial NCT06121401) and the MITO 35 initiative (clinical trial NCT05255471), in which the clinical relevance of HR-SC classification with the patients' outcome is much important than with both the internal and external benchmark.

CONCLUSIONS

In conclusion, we have shown that HR-SC is a viable approach for the determination of HRD-positive or HRD-negative status in OC patients, with good predictive and prognostic roles, along with a low failure rate. This ML approach to measure CN signatures for HRR status holds great potential to advance our understanding of genomic instability in cancer and to improve patient care. By leveraging the power of ML algorithms, we can extract valuable insights from complex genomic data, leading to more precise HRR status assessment and better treatment strategies.

ACKNOWLEDGEMENTS

We thank Fondazione Alessandra Bono for partially supporting this study. The authors dedicate this manuscript to Giovanni Scambia. Giovanni was a highly admired and respected surgeon and gynecological oncologist, who dedicated his life to conduct innovative research with the aim of improving patients' treatment. His warmth and humanity were cherished by all who came in contact with him.

FUNDING

This work was supported by the Associazione Italiana per la Ricerca sul Cancro [grant numbers IG 2021-ID 25932 projects to SP (principal investigator), CO-2018-12367051 (Ministero della Salute) to SP (principal investigator), IG 29071 to CR, IG 19997 and IG 30381 to SM], and from Ricerca Corrente [grant L3/13 from Ministero della Salute to SP and PNRR-MAD-2022-12375663 to SP]; the MITO16A/MaNGO-OV2 trial was partially supported by Roche; the Fund for Scientific Research—Flanders [grant number FWO Vlaanderen 18B2921N to TVG (senior clinical investigator)].

DISCLOSURE

TVG has received consulting fees (via institution) from AbbVie, AstraZeneca, BioNTech, Cancer Communications and Consultancy Ltd, Eisai, GSK, ImmunoGen, Incyte, Karyopharm, Merck Sharp & Dohme (MSD)/Merck, OncXerna Therapeutics, Seagen, Tubulis, Zentalis; received corporate sponsored research (via institution) from Amgen, AstraZeneca, Roche; received honoraria for lectures (via institution) from AstraZeneca, Eisai, GSK, ImmunoGen, MSD; and received support for meetings and/or travel from ImmunoGen, MSD, PharmaMar. SP has received honoraria from AstraZeneca, MSD, Roche, GSK, Novartis, and PharmaMar, and research funding from Roche, AstraZeneca, MSD, GSK, and Pfizer. PH has received honoraria from Amgen, AstraZeneca, GSK, Roche, ImmunoGen, Sotio, Stryker, Zai Lab, MSD, Clovis, Miltenyi, Eisai, Mersana, Exscientia, Daiichi Sankyo, and Karyopharm; has participated in advisory board for AstraZeneca, Roche, GSK, Clovis, ImmunoGen, MSD, Miltenyi, Novartis, Eisai, Corcept, and BioNTech; and has received research funding (institional) from AstraZeneca, Roche, GSK, Genmab, Deutsche Forschungsgemeinschaft, European Union, Deutsche Krebsgesellschaft, ImmunoGen, Seagen, Clovis, and Novartis. IRC is the principal investigator of the PAOLA-1 trial (uncompensated). All other authors have declared no conflicts of interests.

DATA SHARING

Design data for the hybrid panel have been previously published[4] and are available from the lead contact upon reasonable request. Sequence file data have been uploaded to the European Genome-Phenome Archive (EGA) and are publicly available as of the date of publication. Processing software and scripts are available at https://github.com/DIncalciLab/HR-SC and are publicly available as of the date of publication. The list of genes in the panel is available at Zenodo 10.5281/zenodo.13120772. The HRR prediction model and the Nextflow workflow are available at Zenodo 10.5281/zenodo.13120772. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

REFERENCES

1. González-Martín A, Pothuri B, Vergote I, et al. Niraparib in patients with newly diagnosed advanced ovarian cancer. *N Engl J Med*. 2019;381(25):2391-2402.
2. Ray-Coquard I, Pautier P, Pignata S, et al. Olaparib plus bevacizumab as first-line maintenance in ovarian cancer. *N Engl J Med*. 2019;381(25):2416-2428.
3. Monk BJ, Parkinson C, Lim MC, et al. A randomized, phase III trial to evaluate rucaparib monotherapy as maintenance treatment in patients with newly diagnosed ovarian cancer (ATHENA-MONO/GOG-3020/ENGOT-ov45). *J Clin Oncol*. 2022;40(34):3952-3964.
4. Moore K, Colombo N, Scambia G, et al. Maintenance olaparib in patients with newly diagnosed advanced ovarian cancer. *N Engl J Med*. 2018;379(26):2495-2505.
5. Oza AM, Tinker AV, Oaknin A, et al. Antitumor activity and safety of the PARP inhibitor rucaparib in patients with high-grade ovarian carcinoma and a germline or somatic BRCA1 or BRCA2 mutation: integrated analysis of data from Study 10 and ARIEL2. *Gynecol Oncol*. 2017;147(2):267-275.
6. Daniele G, Raspagliesi F, Scambia G, et al. Bevacizumab, carboplatin, and paclitaxel in the first line treatment of advanced ovarian cancer patients: the phase IV MITO-16A/MaNGO-OV2A study. *Int J Gynecol Cancer*. 2021;31(6):875-882.
7. Capoluongo ED, Pellegrino B, Arenare L, et al. Alternative academic approaches for testing homologous recombination deficiency in ovarian cancer in the MITO16A/MaNGO-OV2 trial. *ESMO Open*. 2022;7(5):100585.
8. Gerstung M, Jolly C, Leshchiner I, et al. The evolutionary history of 2,658 cancers. *Nature*. 2020;578(7793):122-128.
9. Pujade-Lauraine E, Brown J, Barnicle A, et al. Homologous recombination repair gene mutations to predict olaparib Plus bevacizumab efficacy in the first-line ovarian cancer PAOLA-1/ENGOT-ov25 trial. *JCO Precis Oncol*. 2023;7:e2200258.
10. Khandekar A, Vangara R, Barnes M, et al. Visualizing and exploring patterns of large mutational events with SigProfilerMatrixGenerator. *BMC Genomics*. 2023;24(1):469.
11. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics*. 2010;11(1):367.
12. Steele CD, Abbasi A, Islam SMA, et al. Signatures of copy number alterations in human cancer. *Nature*. 2022;606(7916):984-991.
13. Islam SMA, Díaz-Gay M, Wu Y, et al. Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. *Cell Genom*. 2022;2(11):100179. https://doi.org/10.1016/j.xgen.2022.100179.
14. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12(85):2825-2830.
15. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. 2020;2(1):56-67.
16. Pellegrino B, Herencia-Ropero A, Llop-Guevara A, et al. Preclinical in vivo validation of the RAD51 test for identification of homologous recombination-deficient tumors and patient stratification. *Cancer Res*. 2022;82(8):1646-1657.
17. Pellegrino B, Capoluongo ED, Bagnoli M, et al. Unraveling the complexity of HRD assessment in ovarian cancer by combining genomic and functional approaches: translational analyses of MITO16-MaNGO-OV-2 trial. *ESMO Open*. 2025;10(1):104091.
18. Tao Z, Wang S, Wu C, et al. The repertoire of copy number alteration signatures in human cancer. *Brief Bioinform*. 2023;24(2):bbad053.
19. Macintyre G, Goranova TE, De Silva D, et al. Copy number signatures and mutational processes in ovarian carcinoma. *Nat Genet*. 2018;50(9):1262-1270.

20. Wang LM, Wang P, Chen XM, et al. Thioparib inhibits homologous recombination repair, activates the type I IFN response, and overcomes olaparib resistance. *EMBO Mol Med*. 2023;15(3):e16235.
21. Jacobson DH, Pan S, Fisher J, Secrier M. Multi-scale characterisation of homologous recombination deficiency in breast cancer. *Genome Med*. 2023;15(1):90.
22. Harvey-Jones E, Raghunandan M, Robbez-Masson L, et al. Longitudinal profiling identifies co-occurring BRCA1/2 reversions, TP53BP1, RIF1 and PAXIP1 mutations in PARP inhibitor-resistant advanced breast cancer. *Ann Oncol*. 2024;35(4):364-380.
23. Kim H, Nguyen NP, Turner K, et al. Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers. *Nat Genet*. 2020;52(9):891-897.
24. Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature*. 2013;501(7467):338-345.
25. Zheng S, Cherniack AD, Dewal N, et al. Comprehensive pan-genomic characterization of adrenocortical carcinoma. *Cancer Cell*. 2016;29(5):723-736.
26. Davis CF, Ricketts CJ, Wang M, et al. The somatic genomic landscape of chromophobe renal cell carcinoma. *Cancer Cell*. 2014;26(3):319-330.