RAD51 as a biomarker for homologous recombination deficiency in high-grade serous ovarian carcinoma: robustness and interobserver variability of the RAD51 test

Kramer, C.J.H.; Llop-Guevara, A.; Yaniz-Galende, E.; Pellegrino, B.; Haar, N.T. ter; Herencia-Ropero, A.; ... ; Vreeswijk, M.P.G.

The Journal of Pathology: Clinical Research
J Pathol Clin Res November 2023; **9:** 442–448
Published online 28 July 2023 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/cjp2.336

BRIEF REPORT

RAD51 as a biomarker for homologous recombination deficiency in high-grade serous ovarian carcinoma: robustness and interobserver variability of the RAD51 test

Claire JH Kramer[1], Alba Llop-Guevara[2], Elisa Yaniz-Galende[3], Benedetta Pellegrino[4,5], Natalja T ter Haar[1], Andrea Herencia-Ropero[2], Nicoletta Campanini[4,5], Antonino Musolino[4,5], Tjalling Bosse[1], Alexandra Leary[3], Violeta Serra[2] and Maaike PG Vreeswijk[6*]

[1]*Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands*
[2]*Experimental Therapeutics Group, Vall d'Hebron Institute of Oncology, Barcelona, Spain*
[3]*Department of Oncology, Institute Gustave Roussy, Villejuif, France*
[4]*Department of Medicine and Surgery, University of Parma, Parma, Italy*
[5]*Medical Oncology and Breast Unit, University Hospital of Parma, Parma, Italy*
[6]*Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands*

*Correspondence to: Maaike PG Vreeswijk, Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands.
E-mail: m.p.g.vreeswijk@lumc.nl

Abstract

The RAD51 test is emerging as a promising biomarker for the assessment of functional homologous recombination deficiency (HRD). Yet, the robustness and reproducibility of the immunofluorescence-based RAD51 test, in different academic laboratories, have not been systematically investigated. Therefore, we tested the performance of the RAD51 assay in formalin-fixed paraffin-embedded (FFPE) high-grade serous ovarian carcinoma (HGSOC) samples in four European laboratories. Here, we confirm that subtle differences in staining procedures result in low variability of RAD51 and γH2AX scores. However, substantial variability in RAD51 scoring was observed in some samples, likely due to complicating technical and biological features, such as high RAD51 signal-to-noise ratio and RAD51 heterogeneity. These results support the need to identify and perform additional quality control steps and/or automating image analysis. Altogether, resolving technical issues should be a priority, as identifying tumours with functional HRD is urgently needed to guide the individual treatment of HGSOC patients. Follow-up studies are needed to define the key tissue quality requirements to assess HRD by RAD51 in FFPE tumour samples, as this test could help in guiding the individual treatment of HGSOC patients.

Keywords: analytical validation; biomarker; high-hrade serous ovarian carcinoma; homologous recombination deficiency; interobserver variability; RAD51 test

Received 19 April 2023; Revised 30 May 2023; Accepted 2 July 2023

Conflict of interest statement: VS and ALG report a patent for PCT/EP2018/086759 pending. VS reports grants and personal fees from AstraZeneca. AL reports Honoria/expenses (AstraZeneca, Clovis, GSK, and Medscape CCO), consulting/advisory board (AstraZeneca, Clovis, GSK, MSD, Merck Serono, Ability, Zeptalis, Agenus, Iovance, Blueprint, and Apmonia) and funded research (AstraZeneca, Clovis, GSK, MSD, Ability, Zeptalis, Agenus, Iovance, Sanofi, Roche, OSEimmuno, and BMS).

Introduction

The identification of high-grade serous ovarian carcinoma (HGSOC) patients with homologous recombination deficient (HRD) tumours is important to identify those who are most likely to benefit from treatment with platinum-based chemotherapy or poly (ADP-ribose) polymerase inhibitors (PARPis) [1].

The RAD51 test is emerging as one of the most promising markers for the functional assessment of HRD in pathology. The test is based on the (in)ability to accumulate RAD51 protein, a key protein in HR, at sites of DNA double-strand breaks [2–8]. The test can be performed on formalin-fixed paraffin-embedded (FFPE) tumour tissue and comprises co-immunofluorescent staining of RAD51/geminin. The RAD51 test is competitive since it has both a

near-perfect sensitivity to detect *BRCA1/2* pathogenic variants [2–5,7,8] as well as a strong predictive capacity for platinum [4,8,9] and PARPi response [7].

The strong clinical validity of RAD51 as an HRD biomarker has already been established by several European laboratories [2–9]. Nevertheless, the robustness and reproducibility of the RAD51 test, when applied in different research laboratories, have not yet been investigated. Therefore, we aimed to systemically investigate the effect of local analytical variations on RAD51 test results and the interobserver variability by an international exchange of HGSOC samples.

Materials and methods

The evaluation of the RAD51 test, comprising four academic laboratories, was divided into training, evaluation, and testing phases. For all phases, FFPE tumour tissue blocks with high tumour cell percentage from HGSOC specimens were selected by an expert gynecopathologist (training: $n = 12$; testing: $n = 10$) and serial sections were cut. Selected archival cases were derived from routine diagnostics and thus captured variation in diagnostic practice. The study was approved by the local Ethics Committee (nWMO-D4-2022-030).

The design of the training phase was as follows. First, HGSOC sections were stained for γH2AX/geminin in two laboratories, according to local staining procedures [2,3,5]. Next, unstained HGSOC sections were stained for RAD51/geminin in the four academic laboratories, applying local staining protocols (Figure 1A) [5,9,10]. Moreover, HGSOC specimens were stained four-fold, according to a central protocol [2–4], and distributed to participating laboratories for local RAD51-scoring (Figure 1B). Thus, training phase specimens were stained and scored twice, applying both local and central staining protocols.

RAD51 scoring was performed by a local observer according to a predefined scoring methodology. A comprehensive overview of the RAD51 scoring methodology, including RAD51 scoring form, can be found in Supplementary materials and methods and Table S1. In brief, the observer first determined whether the presence and quantity of RAD51 nuclear foci were proportionally (homogenous) or disproportionally (heterogeneous) distributed among different tumour fields. If the latter was the case, it was determined whether this was likely due to technical issues/artefacts (including necrosis or poor fixation) or whether the tumour was 'RAD51 heterogeneous'. Next, ≥100 geminin-positive cells were randomly selected and scored in 3–4 tumour areas. For selected geminin-positive cells, the number of RAD51 foci per nucleus was determined. Any relevant issues regarding tissue or staining quality were noted.

In the evaluation meeting, the RAD51-stained sections were critically re-reviewed, and features were

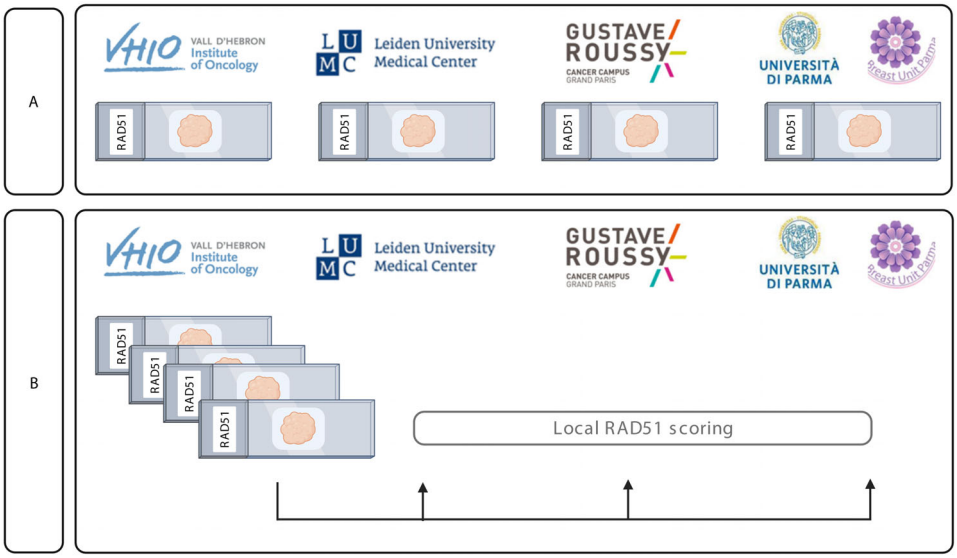

Figure 1. Schematic overview of the design of the training phase of the study. Each of the $n = 12$ HGSOC specimens was stained and scored in the academic laboratories, according to their local RAD51 staining procedures (A: local staining procedure). Additionally, each HGSOC specimen was centrally stained four-fold in one centre and distributed to the remaining laboratories (B: central staining procedure). Thus, each laboratory provided a final RAD51 score for the same locally and centrally stained section.

defined that might underlie the observed variability in RAD51 scores. These features were integrated into an updated RAD51 scoring form.

In the testing phase, the interobserver variability was re-determined in an independent set of HGSOC specimens using the updated RAD51 scoring form (Supplementary materials and methods and Table S2). The RAD51/geminin-staining of testing phase specimens was performed centrally and stained sections were circulated among participating laboratories for local scoring. Now, if RAD51 showed obvious intratumoral heterogeneity, each area was separately scored. Additionally, clear criteria for 'non-evaluability' were predefined.

Final RAD51 and γH2AX scores were calculated using a foci cut-off of 5 or 2, respectively. On a case level, RAD51 scoring variability was defined as the range of RAD51 scores across observers (highest minus lowest score). If two RAD51 scores were provided by one observer in a heterogeneous case, the highest score was prioritised. The final interobserver variability was defined as the median of RAD51 scoring variabilities. Variability in ranges was analysed using the first quartile (Q1), median, and third quartile (Q3). Specific features were reported if noted by ≥3 observers.

Results

First, our goal was to systemically evaluate the robustness of the RAD51 test by analysing the influence of local protocol differences on final test results. The first quality control step of the RAD51 test, γH2AX/geminin co-staining, serves as a proxy for the presence of DNA double-strand breaks. The co-staining was performed in two laboratories with different techniques (immunohistochemistry and immunofluorescence) (Figure 2A) [2,3,5]. The median

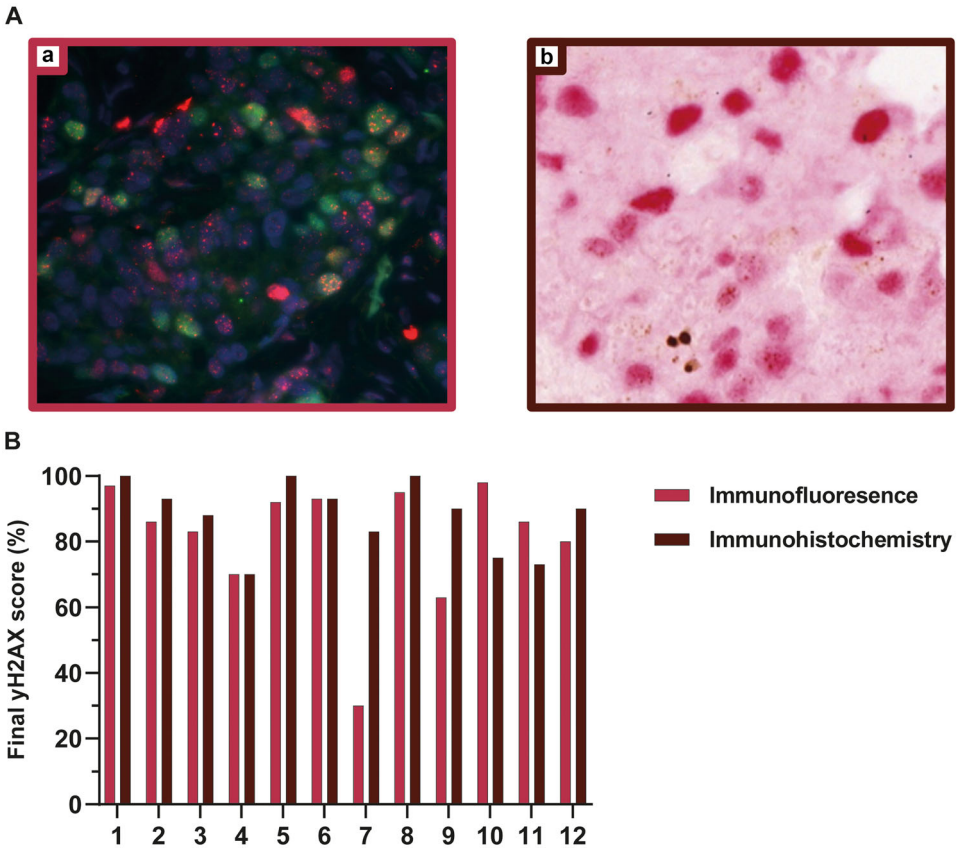

Figure 2. Analytical validation and interobserver agreement of final γH2AX scores. (A) Illustration of (a) immunofluorescence and (b) immunohistochemistry-based γH2AX/geminin-staining of an high-grade serous ovarian carcinoma (HGSOC) specimen (case ID 10). Geminin-positive nuclei of tumour cells [green in (a) and pink in (b)] show γH2AX foci [red in (a) and brown in (b)]. (B) Concordance of γH2AX scores of HGSOC specimens in the training phase, determined by immunofluorescence or immunohistochemical staining. γH2AX scores were defined as the percentage of geminin-positive cells that had at least two γH2AX foci in the nucleus.

range in γH2AX-scores between the two centres was 8% (Q1: 4%, Q3: 21%) (Figure 2B), indicating high concordance of scores for this crucial quality control step.

Next, we investigated the effect of local staining protocol differences on final RAD51 scores. The fact that all HGSOC specimens were stained according to the local and central staining protocol allowed for a side-by-side comparison of RAD51 scores. On a case level, the median difference in RAD51 scores between locally and centrally stained sections was 5% (Q1: 1%, Q3: 9.5%), demonstrating that local protocol differences were trivial and did not affect final RAD51 scores.

In the initial training phase, the interobserver variability was assessed using centrally stained sections. Figure 3A illustrates the RAD51 scores of all four observers per training cohort case, ranked from largest (25%) to smallest (0%) variability in RAD51 scores across observers. The median variability in RAD51 scores between observers was 13.5% (Q1: 9%, Q3: 19%).

For 6 out of 12 cases, ≥3 observers noted complicating features, including weak geminin (case ID 6) or high RAD51 signal-to-noise ratio (case IDs 2 and 11). In cases with these features, RAD51 scoring variability was higher (median: 18.5%) compared to cases without these features (median: 10.5%).

Next, an on-site evaluation meeting was organised to re-evaluate training cohort cases to identify features that may underlie the observed RAD51 scoring variability. Taken together, technical and biological features were defined that may have impacted RAD51 scoring accuracy, such as weak geminin expression, high RAD51 signal-to-noise ratio, and RAD51 heterogeneity (Figure 4). These technical and biological features were incorporated into a refined RAD51 scoring form, including a more

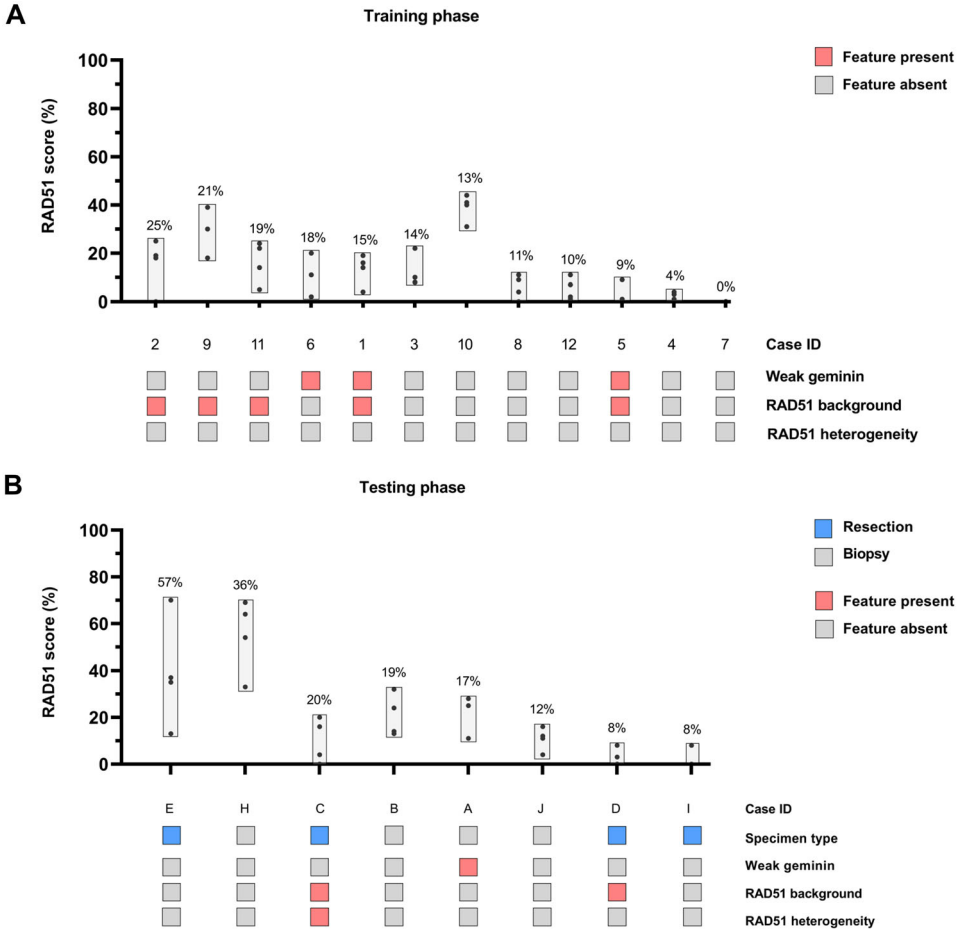

Figure 3. RAD51 scores of individual observers per (A) training phase and (B) testing phase cases, sorted from largest to smallest RAD51 scoring variability. Box plots indicate the variability of RAD51 score across observers. Case ID and staining features in RAD51 scores are shown below the graphs. The presence or absence of weak geminin, RAD51 background, or RAD51 heterogeneity is described if this was noted by at least three of the four observers.

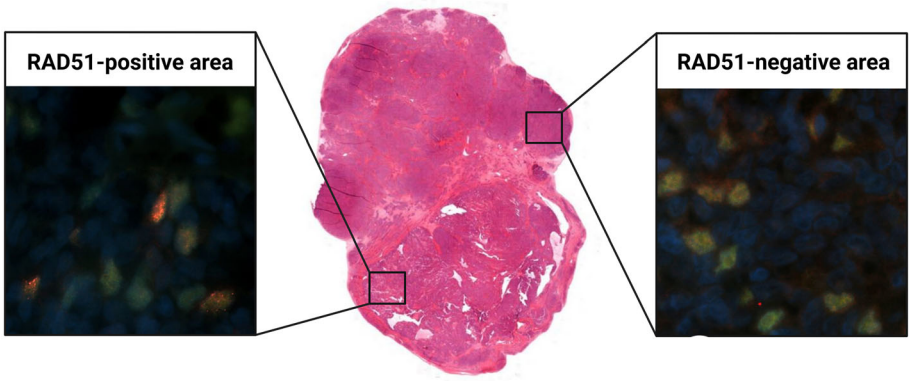

Figure 4. Example of a high-grade serous ovarian carcinoma (HGSOC) with RAD51 heterogeneity. A representative H&E image is shown. The boxes indicate RAD51-positive (left lower corner) and RAD51-negative (right upper corner) tumour areas with the representative RAD51/geminin-staining. This highlights that the RAD51 test enables the identification of hHRD heterogeneity.

in-depth assessment of RAD51 heterogeneity and clear-cut criteria for 'non-evaluability'.

In the testing phase, the RAD51 test was performed on an independent cohort of HGSOC specimens. Now, we enriched for biopsy specimens to get insight into the impact of size of tumour tissue on RAD51 scores. Since the training phase demonstrated that local staining protocol differences do not affect final RAD51 scores, HGSOC sections of the testing cohort were stained only centrally and stained sections were circulated among academic laboratories.

Two HGSOC cases (case IDs F and G) were non-evaluable based on low γH2AX-scores by ≥3 observers and were excluded from further analyses. Regarding RAD51 scoring, the median range of RAD51 scoring variability was 18% (Q1: 9%, Q3: 32%; Figure 3B).

In three cases (case IDs C, A, and D), previously defined complicating features seemed to hamper scoring accuracy, although RAD51 variability was not higher in cases with (median: 17%) compared to those without quality issues (median: 19%).

Finally, RAD51 scoring variability was not higher in large resection specimens (median: 14%) than in smaller biopsy specimens (median: 18%), indicating that the variability was not explained by HGSOC tissue size.

Discussion

To ensure the successful clinical implementation of the RAD51 test, it is imperative to maximise the robustness of the test and to minimise interobserver variability in RAD51 assessments. Hence, the main focus of this multi-centre study was to optimise these crucial aspects. Therefore, we aimed to systematically address this requirement in this multi-centre study. In this RAD51 biomarker assessment, we confirmed that subtle differences in staining procedures resulted in low variability of RAD51 and γH2AX scores, demonstrating the robustness of this biomarker. In addition, we demonstrated that, in some cases, substantial differences in RAD51 scores were reported by observers. This can be particularly challenging when the scoring variability is observed surrounding established HRD thresholds. Currently, establishing HRD thresholds remains challenging due to the lack of a clear 'gold standard' for defining HRD, which further complicates the validation and calibration of HRD tests. This complexity is exemplified by the recent modification of the MyChoice HRD threshold aimed at enhancing PARPi sensitivity prediction [11]. Hence, in the context of this study, we have chosen not to emphasise threshold-related aspects. Overall, the most variability was seen in cases for which complicating technical and biological features were reported, although this did not completely explain substantial scoring variability.

Regarding biological features, RAD51 heterogeneity may lead to scoring variability as the tumour areas selected for quantification are reader-dependent. This underlines the importance of accurately screening the entire specimen and selecting multiple representative tumour tissue areas for RAD51 scoring. Notably, it is crucial to acknowledge the presence of HR-proficient tumour cells, as these cells have the potential to drive acquired treatment resistance. Automated image analysis may facilitate RAD51 scoring in the context of high tumour heterogeneity.

In terms of technical challenges, RAD51 scoring accuracy was potentially affected by weak geminin and high RAD51 signal-to-noise ratio. The latter might

be explained by the poor-to-moderate fixation quality of large routine diagnostic HGSOC specimens. Further studies are ongoing to dissect the fixation variables that impact RAD51 scoring.

'Real-life' variability is complex to control in diagnostics. This variability may be circumvented by limiting the RAD51 test to smaller, well-fixed biopsy specimens, although we did not find a clear relationship between scoring variability and tissue size. An alternative would be to apply more strict quality criteria to ensure the evaluability of surgical samples. Even though we acknowledge that the 'non-evaluable' category will diminish the number of cases with a final RAD51 score, it is likely that it will positively contribute to the tests' reproducibility.

We acknowledge that this biomarker assessment did not yield adequate statistical power to provide strong conclusions on features impacting interobserver variability. Still, our data give valuable insight into the robustness and pitfalls within the RAD51 test. Next, this performance assessment of RAD51 is limited to HGSOC and does not reflect the robustness of RAD51 tests in other carcinomas, in which promising results have been published [2–5,7,10,12,13].

Altogether, previous studies have established that the RAD51 test is an attractive, low-cost alternative to genomic-based HRD tests [2–5,7–9]. Once described challenges in RAD51 scoring have been tackled, it is conceivable that the RAD51-based HRD test will play a crucial role in guiding individualised HGSOC treatment.

Author contributions

TB, AL, VS and MPGV conceptualised the study. TB, AL, VS and MPGV designed the methodology. CJHK, ALG, EY and BP contributed to the formal analysis and investigation of data. TB provided study resources. CJHK, ALG, EY and BP contributed to data curation. CJHK, ALG, TB, VS and MPGV contributed to writing the original draft. EY, BP, NTH, AH, NC and AM were involved in writing, reviewing and editing the manuscript. CJHK, ALG, EY and BP were involved in visualization. NC, AM, TB, AL, VS and MPGV supervised the study. NTH contributed to project administration. AL, TB, VS and MPGV were involved in funding acquisition.

Acknowledgements

This work was supported by the Dutch Cancer Society (KWF) (grant: 12995 to TB and MV) and by ERA PerMed (grant ERAPERMED2019-215 to VS). VS received funding from Instituto de Salud Carlos III (CPII19/00033). ALG received funding from Asociación Española Contra el Cáncer (AECC) (INVES20095LLOP) and AHR from Generalitat de Catalunya (PERIS SLT017/20/000081).

Data availability statement

Due to privacy restrictions, data are only available from the corresponding author upon reasonable request.

References

1. Miller RE, Leary A, Scott CL, *et al*. ESMO recommendations on predictive biomarker testing for homologous recombination deficiency and PARP inhibitor benefit in ovarian cancer. *Ann Oncol* 2020; **31**: 1606–1622.
2. Castroviejo-Bermejo M, Cruz C, Llop-Guevara A, *et al*. A RAD51 assay feasible in routine tumor samples calls PARP inhibitor response beyond BRCA mutation. *EMBO Mol Med* 2018; **10**: e9172.
3. Cruz C, Castroviejo-Bermejo M, Gutiérrez-Enríquez S, *et al*. RAD51 foci as a functional biomarker of homologous recombination repair and PARP inhibitor resistance in germline BRCA-mutated breast cancer. *Ann Oncol* 2018; **29**: 1203–1210.
4. Llop-Guevara A, Loibl S, Villacampa G, *et al*. Association of RAD51 with homologous recombination deficiency (HRD) and clinical outcomes in untreated triple-negative breast cancer (TNBC): analysis of the GeparSixto randomized clinical trial. *Ann Oncol* 2021; **32**: 1590–1596.
5. van Wijk LM, Kramer CJH, Vermeulen S, *et al*. The RAD51-FFPE test; calibration of a functional homologous recombination deficiency test on diagnostic endometrial and ovarian tumor blocks. *Cancers* 2021; **13**: 2994.
6. van Wijk LM, Nilas AB, Vrieling H, *et al*. RAD51 as a functional biomarker for homologous recombination deficiency in cancer: a promising addition to the HRD toolbox? *Expert Rev Mol Diagn* 2022; **22**: 185–199.
7. Eikesdal HP, Yndestad S, Elzawahry A, *et al*. Olaparib monotherapy as primary treatment in unselected triple negative breast cancer. *Ann Oncol* 2021; **32**: 240–249.
8. Blanc-Durand F, Yaniz E, Genestie C, *et al*. Evaluation of a RAD51 functional assay in advanced ovarian cancer, a GINECO/GINEGEPS study. *J Clin Oncol* 2021; **39**: 5513.
9. Blanc-Durand F, Yaniz-Galende E, Llop-Guevara A, *et al*. A RAD51 functional assay as a candidate test for homologous recombination deficiency in ovarian cancer. *Gynecol Oncol* 2023; **171**: 106–113.

10. 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**: 1646–1657.
11. Coleman RL, Fleming GF, Brady MF, *et al*. Veliparib with first-line chemotherapy and as maintenance therapy in ovarian cancer. *N Engl J Med* 2019; **381**: 2403–2415.
12. Carreira S, Porta N, Arce-Gallego S, *et al*. Biomarkers associating with PARP inhibitor benefit in prostate cancer in the TOPARP-B trial. *Cancer Discov* 2021; **11**: 2812–2827.
13. 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**: 100585.

SUPPLEMENTARY MATERIAL ONLINE

Supplementary materials and methods

Table S1. Predefined RAD51 scoring form used in the training phase

Table S2. RAD51 scoring form used in the testing phase