Shallow Whole Genome Sequencing (sWGS) pipeline for the assessment of Homologous Recombination Deficiency (HRD) score in Hereditary Breast and Ovarian Cancer (HBOC) patients

Giovanni Luca Scaglione[1,2], Carmela Nardelli[1,3], Mario Setaro[1], Filippo Russo[1,3], Ettore Domenico Capoluongo[3,4]
[1] Oncogenomics Laboratory, CEINGE - Biotecnologie Avanzate - Franco Salvatore, Naples, Italy
[2] Istituto Dermopatico dell' Immacolata, IDI-IRCCS, Rome, Italy
[3] Department of Molecular Medicine and Medical Biotechnologies, University of studies of Naples Federico II, Naples, Italy
[4] Hospital for Emergency Cannizzaro, Catania, Italy

Questo lavoro è stato in parte presentato al 53° Cogresso Nazionale SIBioC, 11-13 Ottobre 2021, Virtual Edition, quale Comunicazione Orale

ABSTRACT

Introduction: the Homologous Recombination Repair (HRR) system is essential for DNA repair and genomic stability. Homologous Recombination Deficiency (HRD) arises upon the inactivation of several genes involved in HRR, and it is commonly observed in breast and ovarian cancer. The detection of HRD is a valuable tool of clinical relevance, indicative of sensitivity to agnostic therapy and DNA damaging reagents.
Methods: shallow Whole Genome Sequencing (sWGS) was performed on 26 ovarian cancer (OC) samples divided into training (n=13) and test group (n=13). The training set included 7 somatic and 6 germline carrying BRCA+ve (n=8) and BRCA-ve (n=5) status. In particular, the BRCA+ve were 3/7 somatic and 5/6 germline samples; the BRCA-ve were 4/7somatic and 1/6 germline. All samples were prepared using the KAPA HyperPlus Library Preparation kit (Roche), pooled and sequenced on NextSeq550 Dx (Illumina). Large-Scale Transitions (LST) profiles were calculated using WGS data at low coverage (<1x) using different sliding window sizes spanning 10-1 000 Kbases. The HRD score was estimated using our bioinformatics pipeline.
Results: the BRCA status was assessed in 22/26 OC samples, and the HRD score was evaluated in all somatic samples. In the test group, 7 BRCA+ve samples were classified as HRD+ve, 2 BRCA-ve were classified as HRD-ve, and 4 BRCA unknown samples were predicted as HRD+ve in two cases and HRD-ve in the other half.
Discussion: germline BRCA1/2 mutation status is currently the main genetic biomarker of HRD, but it has its drawbacks: HRD can be driven purely by somatic events. A bioinformatics pipeline to evaluate the HRR system status in breast and ovarian cancer has been completed; it is based on sWGS to support therapeutics and follow up strategies.

Key words: *Homologous Recombination Deficiency (HRD), Shallow Whole Genome Sequencing (sWGS), Hereditary Breast and Ovarian Cancer Syndrome (HBOC)*

INTRODUCTION

DNA Damage Repair (DDR) pathways act to protect the genome of human cells from mutational insults. Cancer development is a multifactorial process and comprises key events such as the inability to control cellular replication, angiogenic infiltration, invasion activation, and metastasis formation (1). The genomic instability is due to genetic mutations affecting the DNA directly or to errors in the DNA damage repair system. One of the most frequently altered DDR pathways in ovarian and breast cancer is the homologous recombination (HR) system (2). This is a double-strand DNA breaks repairing mechanism, a relatively error-free mechanism because it is based on the presence of intact sister chromatids that guide the repair of the damaged portion of DNA (3). This high-fidelity DNA double-strand break (DSB) repair system involves numerous genes, including *BRCA1* and *BRCA2* (4). HR pathway deficiency (HRD) arises upon inactivation of genes; among these, *BRCA1/2*, *RAD51C* or *PALB2* are associated with several tumors that are characterized by the presence of the genomic scars related to HRD-tumors (3):

- Large-Scale State Transitions (LSTs), chromosomal breaks between adjacent regions of at least 10 Mbase;

Corrisponding Author: Giovanni Luca Scaglione, Via dei Monti di Creta, 104, Rome, Italy, E.mail: g.scaglione@idi.it

Riceved: 01.02.2022 Revised: 23.02.2022 Accepted: 08.04.2022

Published on-line: 22.04.2022 DOI: 10.19186/BC_2022.023

- Telomeric Allelic Imbalance (TAI), defined as the number of regions with an allelic imbalance that extends to the sub-telomere but does not cross the centromere;
- Loss Of Heterozygosity (LOH) indicates when one of the two alleles originally present in the cell is lost.

These structural aberrations could be measured to provide a powerful score reflecting genomic instability in cancer cells (5). Marquard et al. (6) have extensively studied Cancer Genome Atlas (TCGA) samples to evaluate the potential clinical significance of these genomics scars in several tumor types. These Authors reported that TAI and LSTs generally correlate with each other (Spearman's $\rho = 0.87$), with $\rho > 0.7$ in 8 of 15 cancer types. LOH shows a lower level of correlation with both TAI ($\rho = 0.81$) and LST ($\rho = 0.73$). TAI and LST measure overlapping genomic scars and give high scores to more similar subpopulations of tumors. In contrast, the lower correlation to LOH suggests that it is counting genomic scars of a somewhat separate tumor subpopulation. Moreover, it is shown that the high correlation between TAI and LST is not caused by an overlap in the individual events counted. Instead, the high correlation indicated that the underlying repair defect generates multiple complex genomic aberration patterns. The different methods rely on capturing different non-overlapping aspects of these patterns to estimate the overall DNA repair competency (6).

Recognizing that these distinctive genomic scars have an important clinical impact because the HRD tumors respond more efficiently to targeted therapy with poly ADP-ribose polymerase inhibitors (PARPi) and DNA damaging reagents (7,8). Identifying these alterations at the somatic level and the germline could thus provide major support in the therapeutic choice.

Based on all these considerations, our study aimed to perform a shallow sequencing protocol followed by the identification and enumeration of one specific genomic scar, namely LSTs, to outline an HRD profile, trying to address HRD-related target therapies as widely indicated, wherever applicable, by the current guidelines for agnostic therapy.

METHODS

Sample preparation

DNA samples from 26 women affected by ovarian carcinoma were collected after obtaining written informed consent and included in this study to evaluate the workflow. The samples were divided into two subsets as training and test groups. This study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the University of Naples Federico II. Informed consent was obtained from all subjects involved in the study. The patients expressed their informed consent and authorization to anonymously use their biological material for diagnostic and scientific research purposes. The training group consisted of 13 DNA samples extracted from formalin-fixed paraffin-embedded (FFPE) ovarian carcinoma tissues. Among these, ten somatic samples were previously genotyped for *BRCA1/2* genes by an external center and then enrolled by our laboratory for shallow sequencing. The other 3 samples were genotyped in our laboratory using the 12 genes panel for *HBOC* genes (Devyser AB, Stockholm, Sweden). The training group included both germline (n=6) and somatic (n=7) samples previously genotyped for both *BRCA1/2* and *HBOC* genes. The somatic samples were 4 BRCA-ve and 3 BRCA+ve. The germline samples were 1 BRCA-ve sample and 5 BRCA+ve. The somatic DNA of 19 samples was isolated using the QIAamp DNA FFPE tissue kit (Qiagen, Hilden, Germany). The 6 germline samples were extracted using the MagPurix Blood DNA Extraction kit 200 (Zinexts Life Science - Taiwan). The quality of the nucleic acid was estimated using the TapeStation system (Agilent Technologies, Santa Clara, CA, USA) followed by quantification on Qubit® Fluorometer 4.0 (Invitrogen Co., Life Sciences, Carlsbad, USA), by using both Qubit dsDNA HS (High Sensitivity) or BR (Broad Range) assays for FFPE and germline respectively.

Library preparation

The DNA libraries were prepared using the KAPA HyperPlus kit (Roche Sequencing Solutions, Pleasanton, CA, US). According to the manufacturer's protocol, the DNA libraries were prepared at a final concentration of 50 ng/µL. Quality and integrity checks of the libraries were performed with the TapeStation system. Finally, the library concentration was measured using a high sensitivity Qubit dsDNA HS assay kit on Qubit® Fluorometer 4.0, after the equimolar pooling step of all samples included in the Next Generation Sequencing (NGS) run. The pool was loaded to a final concentration of about 1.2 pM and 2% Phix. The parallel sequencing step was performed using the NextSeq500/550 Mid output kit (300 cycles) on the Illumina NextSeq550 Dx (Illumina, San Diego, CA, USA) platform RUO mode.

Bioinformatics analysis

The performance of the NGS run was evaluated with the Illumina Run Manager installed on the instrument. Quality checks of sequencing data were performed with MultiQC software (9). The Fastq files were aligned to the GRCh37/hg19 assembly of the human reference genome using BWA-MEM algorithm (10). Moreover, supplementary and duplicate reads were filtered with Samtools and PicardTools' MarkDuplicates, respectively. To detect biases in the sequencing and/or mapping steps, all Binary Alignment Map (BAM) files were evaluated with the platform-independent tool Qualimap (11). Circular binary segmentation on the aligned files was performed using different window sizes using the QDNAseq tool (12,13). Downstream analysis of our bioinformatics tool performed data normalization and dynamic coverage calculation throughout the entire genome (WGS data) with very low depth in several fixed bins. The algorithm scans for copy number alterations in each chromosome in a range of preset windows bins. The available range is

10-1 000Kb. After reads normalization, genomic regions affected by LSTs are filtered and collected. An HRD score is assigned to each sample in 6 different windows bins at this stage. Finally, the segmentation profiles are processed, and the number of LSTs calculated in each genome window is integrated in a unique result for the HRD classification. Three possible outcomes are available: HRD negative, HRD positive, or Undetermined when the algorithm cannot classify the HRD status.

RISULTS

The NGS protocol for shallow sequencing was optimized to preserve somatic DNA available in very small amounts, as it was extracted from FFPE samples. At the same time, we tried to avoid unnecessary high read coverage, thus allowing more samples per run to be loaded. The first run was loaded with 16 samples on the NextSeq 550 Dx using the Mid output kit described in the Methods section. In the subsequent runs, we fixed the number of samples per run to 50 to preserve coverage to be not lower than 0.4x. Shallow WGS was performed on both DNA isolated from FFPE sections and the blood samples (germline). The average coverage for each sample was expected to be lower than 1x. Therefore, the entire protocol was optimized to obtain a final coverage between 0.4x and 0.6x of genomic fraction covered. This was a sufficient criterion to allow a good performance of the algorithm. Once the samples were sequenced and quality checks performed on the fastq, raw reads were aligned against the hg19 human reference. The coverage profiles were then inspected with Integrative Genomics Viewer (14), as shown in Figure 1. The picture shows coverage data in a small genomic region for 3 randomly selected samples sequenced by our sWGS protocol. This tool makes possible to highlight the discontinuity of the genomic fractions covered as expected in the shallow NGS mode. For each sample, the overall coverage profile was evaluated and inspected. A detailed view of the coverage profile is shown in Figure 2. Here the coverage across the entire reference is shown as a continuous red line through all chromosomes in consecutive order from left to right. The genome fraction coverage for all the samples

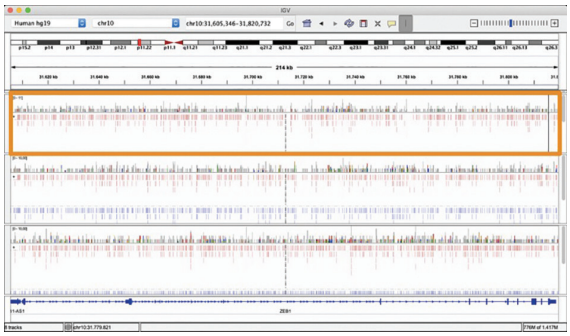

Figure 1
Integrative Genomics Viewer (IGV) profiles of a genetic spot (chr10:31,605,346-31,820,732) sequenced in 3 shallow Whole Genome Sequencing (sWGS) samples sequenced on Illumina NextSeq550DX. The figure shows coverage data in a small genomic region of 3 random samples sequenced using the shallow Next Generation Sequencing (sNGS) protocol. The coverage profiles are examined with Integrative Genomics Viewer (Broad Institute, University of California). The orange box highlights the overall coverage profile in a single sample reflecting the discontinuity of the genomic fractions covered (as foreseen by shallow Whole Genome Sequencing modality).

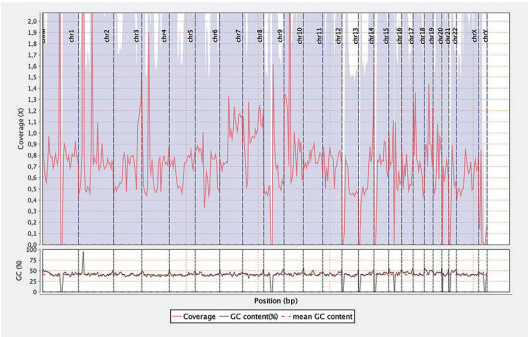

Figure 2
Coverage Across Human Reference. Overview of a coverage profile using Qualimap tool (11) of a sample analyzed with our shallow Whole Genome Sequencing (sWGS) protocol. From left to right, the average coverage across the entire reference is shown as a continuous red line through all chromosomes (dashed horizontal lines).

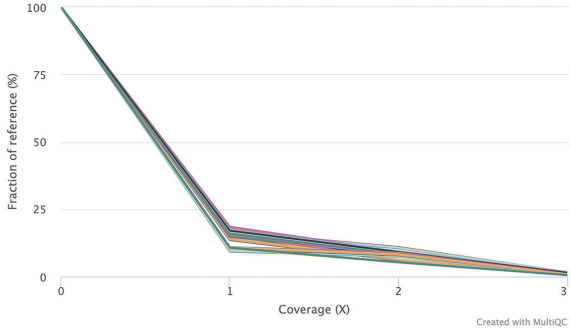

Figure 3
Genome Fraction Coverage. Aligning all the Binary Alignment Map (BAM) files, it is possible to analyze the genomic instability as Large-Scale Transitions. As shown in the figure, comparing the results in terms of Coverage (X) and Fraction of reference (%), there is a strong similarity between all the samples (each represented with different colored lines) in terms of the covered genome fraction. The plot was extracted from Qualimap report (11).

is shown in Figure 3. This is a visual representation of how much reference has been sequenced with at least a given coverage rate (9) and remarks the high similarity among different samples in terms of fraction of genome covered.

All the aligned BAM files were then processed to study the genomic instability in terms of large-scale transitions. The R package QDNAseq (13) was used to investigate these chromosomal aberrations. Finally, the HRD prediction was performed using our bioinformatics pipeline. Training samples (n=13) were used to set the thresholds for the HRD scoring algorithm. In the training set, one "undetermined" sample, for whom HRD prediction was not measured, was intentionally included to evaluate the performance of our classifier in this possible scenario. For each model, three values are calculated and reported as: negative (indicated with "0"), potentially positive (indicated with "1") or positive sample (indicated with "2"). In Table 1, the classification score coupled with the model's predictions for both the training and the test group is shown. In the training group, we analyzed 13 known samples: 1 undetermined sample, 10 negative samples, both somatic and germline, and 2 positive samples with confirmed HRD status. The algorithm performed well in all the 13 cases in the test group, as no one was classified as undetermined. We reported 4 samples as negative and 9 samples as positive for HRD.

Interestingly, these results were successively compared with *BRCA* status when available. Indeed, we found 10 out of 13 samples matching the *BRCA* status and HRD score. Specifically, 8 out of 9 sBRCA+ve samples were also HRD positive, and 2 sBRCA-ve samples were found as HRD negative. The remaining 3 samples were not genotyped before the sWGS protocol for *BRCA,* so we could not compare these results. As expected, all the germline samples in the training group were HRD negative.

DISCUSSION

Carcinogenesis is the sum of several events provoking the accumulation of many malignant features in cells, such as the ability of unlimited replication, the induction of angiogenesis, and the activation of invasion and metastasis. These events are mainly caused by genomic instability arising from genetic mutations (16). Generally, DNA damage response mechanisms, also called DNA Damage Repair pathways (DDRs), can protect the genome from mutational stimuli. Among these, the two pivotal repair mechanisms against DNA double-strand breaks (DSB) are the homologous recombination repair (HRR) and non-homologous end-joining (NHEJ). One of the most frequently impaired DDR pathways in cancer cells is the homologous recombination repair (HRR). The loss of the HRR activity, namely Homologous Recombination Repair Deficiency or HRD, is caused by somatic or germline mutations in those genes that are principally involved in the repair of DSBs, especially *BRCA1/2* or *PALB2*. The HRD appears to be one of the most dangerous mechanisms involved in cancer development. When this repairing mechanism is lost, the cell primarily adopts a NHEJ pathway that is an error-prone system that increases genomic instability in cancerous cells (17). In literature, both *BRCA1* and *BRCA2* mutations are described as the most associated with pathogenic alteration in HRR (18).

According to data reported in several studies (like the PAOLA-1 study) (1), HRD positivity was approved as a new biomarker for the stratification of patients with high-grade epithelial ovarian cancer due to excellent results in maintenance therapy obtained from the combined administration of Olaparib and Bevacizumab (19). Furthermore, the use of new-generation drugs that inhibit the activity of poly-ADP-ribose-polymerase (PARPi) has shown very high efficacy, especially in *BRCA*-mutated tumors. The PARP1 and PARP2 enzymes are involved in the repair mechanisms of single-stranded DNA damage by adding long poly-ADP-ribose chains. Cells with HRD take advantage of NHEJ repair mechanisms more frequently, thus leading to an accumulation of damage at the genomic level. For this reason, PARP inhibition is particularly effective if the cell has a deficiency in the HRR system (20). Initially, PARPi were reserved for germline *BRCA1* and *BRCA2* positive patients, but today it is possible to extend the list of HRD-tumors in which these two genes are not mutated (20). This means that the identification of genomic instability, even if not linked to *BRCA* alterations, should be used as a biomarker in response to PARP inhibition.

One of the main mechanisms of carcinogenesis consists of a type of structural alteration at the chromosomal level, also known as Copy Number Variations (CNVs) (21,22). The gold standard for distinguishing CNVs has been committed to using a genome-wide comparative genomic hybridization array (aCGH) (23). Array-CGH is currently the most widely used technique. It compares the hybridization intensity of a DNA sample and a matched reference DNA, using oligonucleotide probes on a slide (23). This technique requires high-quality input DNA (13), thus limiting its application to somatic samples. The DNA extracted from FFPE tissue could be degraded, and formalin fixation could induce nucleotide changes. So, the use of FFPE-samples leads to an increase in the background noise with a consequently altered interpretation of the results (24,25).

An alternative technique is the detection of CNVs by means of shallow whole-genome sequencing (sWGS), a sequencing method based on the number of aligned reads within chromosomal windows. This sequencing protocol requires a very low genomic coverage (0.1-1.0x) (12), and it has been demonstrated that the diagnostic and clinical significance of sWGS is comparable to the use of aCGH (26,27). Furthermore, this approach has three advantages: cost reduction, reference-free analysis, and lower analytical background noise (13). These features result from accurate computational corrections coupled with analyses implemented in well-established bioinformatics workflows and alignment algorithms (28).

As mentioned above, it is essential to define specific genomic signatures that can reflect the genomic instability of tumor cells. Among the three types of genomic scars

Table 1
Homologous Recombination Deficiency (HRD) score prediction using the six integrated models included in our bioinformatics pipeline

Sample ID	Sample Type	BRCA status	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	HRD Prediction
Training_01	Somatic	Negative	0	2	0	0	1	0	Undetermined
Training_02	Somatic	Positive	0	0	0	0	0	0	Negative
Training_03	Somatic	Negative	0	0	0	0	0	0	Negative
Training_04	Somatic	Positive	2	2	2	2	2	2	Positive
Training_05	Somatic	Negative	0	0	0	0	0	0	Negative
Training_06	Somatic	Positive	1	2	2	2	2	2	Positive
Training_07	Somatic	Negative	0	0	0	0	0	0	Negative
Training_08	Germline	Positive	0	0	0	0	0	0	Negative
Training_09	Germline	Positive	0	0	0	0	0	0	Negative
Training_10	Germline	Negative	0	0	0	0	0	0	Negative
Training_11	Germline	Positive	0	0	0	0	0	0	Negative
Training_12	Germline	Positive	0	0	0	0	0	0	Negative
Training_13	Germline	Positive	0	0	0	0	0	0	Negative
Test_01	Somatic	NA	0	0	0	0	0	0	Negative
Test_02	Somatic	NA	0	1	1	1	1	1	Positive*
Test_03	Somatic	NA	0	0	0	0	0	0	Negative
Test_04	Somatic	Positive	2	2	2	2	2	2	Positive
Test_05	Somatic	Positive	1	2	2	2	2	2	Positive
Test_06	Somatic	NA	1	1	1	2	1	1	Positive*
Test_07	Somatic	Positive	2	2	2	2	2	2	Positive
Test_08	Somatic	Positive	2	2	2	2	2	2	Positive
Test_09	Somatic	Positive	2	2	2	2	2	2	Positive
Test_10	Somatic	Positive	2	1	2	1	2	2	Positive*
Test_11	Somatic	Negative	0	0	0	0	0	0	Negative
Test_12	Somatic	Negative	1	0	0	0	0	0	Negative
Test_13	Somatic	Positive	0	2	2	2	2	2	Positive

Training samples (n=13) were used to set the thresholds for Homologous Recombination Deficiency (HRD) scoring algorithm. Notably, sample 'training_01' was intentionally included to evaluate performance in case of undetermined samples in our classifier. The three available outcomes calculated by each model are color coded: white cells: negative sample (0); light gray cells: potentially positive sample (1); dark gray cells: positive sample (2).
"NA": BRCA status not available.
The Homologous Recombination Deficiency (HRD) prediction (rightmost column) is obtained by combining direct annotation and Homologous Recombination Deficiency (HRD) score calculated by means of the ensemble learning method (machine learning module) included in our pipeline. When strong evidence of positive Homologous Recombination Deficiency (HRD) score was not available, the prediction is reported with a 'starred' text (positive).*

associated with HRD-tumors (3), namely LSTs, TAI and LOH, in this study, we applied and evaluated a bioinformatic pipeline based on LST metrics only that has shown promising performances, despite the absence of TAI and LOH measurements. This scenario analyzed 13 samples (to train our algorithm) and 13 germline and somatic test samples sequenced with sWGS workflow. HRD prediction comes from the combination of direct annotation and HRD score calculated with a machine learning module included in our pipeline. We reported "positive" or "negative" patients for HRD in the presence of strong evidence; otherwise, the prediction was flagged with a star ("positive*"). As shown in Table 1, HRD status is reported in association with the models' predictions, confirming how the algorithm performed well in all the 13 tested samples. Indeed, no "HRD undetermined" predictions were reported. In our study, 7 previously genotyped subjects as BRCA positive (see Table 1) were classified as HRD positive by our bioinformatics pipeline. Specifically, 6 cases had strong evidence of positivity. In contrast, 1 case was reported as a borderline according to the prediction criteria included in our tool, and each sample is independently classified according to the level of agreement of the 6 build-in independent models. Strong evidence is assigned when at least 5 models return the same value. So, for the following samples: Test_04, Test_7, Test_08 and Test_09, we had strong evidence of positivity as the 6 models always assigned "2" as a result. Samples Test_05 and Test_13 were collected up to 5 times '2' only but still considered strongly positive. On the other hand, sample Test_10 has collected 4 times '2' and 2 times '1'; therefore, it has been classified as borderline. To classify as negative samples, we must collect 5 or more zeros, as for samples Test_03 and Test_11. Notably, these 2 patients were BRCA negative.

Moreover, 4 samples with no *BRCA* information were also analyzed with our pipeline. Test_01 and Test_03 were classified as HRD negative with 6 out 6 zeroes as a model outcome. Additionally, our software classified Test_02 and Test_06 as borderline as they collected 5 times '1' as an outcome.

Finally, there is evidence supporting the possibility of studying relevant genomic aberrations starting from sWGS protocols (13) that reduce costs in the field of molecular diagnostics and allow sequencing of low-quality DNA, such as the one deriving from paraffin-fixed samples. In this scenario, our pipeline is a robust, lightweight, and reliable front-runner diagnostic strategy tailored to identify genomic instability in a *BRCA*-independent manner. Nevertheless, we are aware of the limitations due to the small sample size. Therefore, to enforce our pipeline as a putative diagnostic tool, we are planning to increase the number of samples and the tumor types that should be evaluated with our workflow. Furthermore, our efforts are pointed to define a valid statistical framework to fully comply with the clinical and experimental application of this process. In conclusion, despite some amendable flaws can be recognized, we are herein proposing our pipeline as a further diagnostics tool for patient's care in HBOC syndrome.

REFERENCES

1. Wagener-Ryczek S, Merkelbach-Bruse S, Siemanowski J. Biomarkers for homologous recombination deficiency in cancer. J Pers Med 2021;11:612.
2. Barzaman K, Karami J, Zarei Z, et al. Breast cancer: Biology, biomarkers, and treatments. Int Immunopharmacol 2020;84:106535.
3. van Wilpe S, Tolmeijer SH, Koornstra RHT, et al. Homologous recombination repair deficiency and implications for tumor immunogenicity. Cancers (Basel) 2021;13:2249.
4. Nguyen L, Martens JW, Van Hoeck A, et al. Pan-cancer landscape of homologous recombination deficiency. Nat Commun 2020;11:5584.
5. De Luca XM, Newell F, Kazakoff SH, et al. Using whole-genome sequencing data to derive the homologous recombination deficiency scores. NPJ Breast Cancer 2020;6:33.
6. Marquard AM, Eklund AC, Joshi T, et al. Pan-cancer analysis of genomic scar signatures associated with homologous recombination deficiency suggests novel indications for existing cancer drugs. Biomark Res 2015;3:9.
7. Marabelle A, Le DT, Ascierto PA, et al. Efficacy of Pembrolizumab in patients with noncolorectal high microsatellite instability/mismatch repair-deficient cancer: results from the phase II KEYNOTE-158 Study. J Clin Oncol 2020;38:1-10.
8. Stover EH, Fuh K, Konstantinopoulos PA, et al. Clinical assays for assessment of homologous recombination DNA repair deficiency. Gynecol Oncol 2020;159:887-98.
9. Ewels P, Magnusson M, Lundin S, et al. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 2016;32:3047-8.
10. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754-60.
11. Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 2015;32:292-4.
12. Smolander J, Khan S, Singaravelu K, et al. Evaluation of tools for identifying large copy number variations from ultra-low-coverage whole-genome sequencing data. BMC Genomics 2021;22:357.
13. Scheinin I, Sie D, Bengtsson H, et al. DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. Genome Res 2014;24:2022-32.
14. Robinson JT, Thorvaldsdóttir H, Wenger AM, et al. Variant review with the integrative genomics viewer. Cancer Res 2017;77:e31-4.
15. Ewels P, Magnusson M, Lundin S, et al. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 2016;32:3047-8.
16. Beggs R, Yang ES. Targeting DNA repair in precision medicine. Adv Protein Chem Struct Biol 2019;115:135-55.
17. Jeggo PA, Pearl LH, Carr AM. DNA repair, genome stability and cancer: a historical perspective. Nat Rev Cancer 2016;16:35-42.
18. Heeke AL, Pishvaian MJ, Lynce F, et al. Prevalence of homologous recombination-related gene mutations across multiple cancer types. JCO Precis Oncol 2018;2018:PO.17.00286.
19. Ray-Coquard I, Pautier P, Pignata S, et al. Olaparib plus Bevacizumab as first-line maintenance in ovarian cancer. N Engl J Med 2019;381:2416-28.
20. Sunada S, Nakanishi A, Miki Y. Crosstalk of DNA double-strand break repair pathways in poly(ADP-ribose)

polymerase inhibitor treatment of breast cancer susceptibility gene 1/2-mutated cancer. Cancer Sci 2018;109:893-9.
21. Pinkel D, Segraves R, Sudar D, et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet 1998;20:207-11.
22. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144:646-74.
23. Mc Sherry EA, Mc Goldrick A, Kay EW, et al. Formalin-fixed paraffin-embedded clinical tissues show spurious copy number changes in array-CGH profiles. Clin Genet 2007;72:441-7.
24. Chin SF, Santonja A, Grzelak M, et al. Shallow whole genome sequencing for robust copy number profiling of formalin-fixed paraffin-embedded breast cancers. Exp Mol Pathol 2018;104:161-9.
25. Koshiba M, Ogawa K, Hamazaki S, et al. The effect of formalin fixation on DNA and the extraction of high-molecular-weight DNA from fixed and embedded tissues. Pathol Res Pract 1993;189:66-72.
26. Hostetter G, Kim SY, Savage S, et al. Random DNA fragmentation allows detection of single-copy, single-exon alterations of copy number by oligonucleotide array CGH in clinical FFPE samples. Nucleic Acids Res 2010;38:e9.
27. Van der Linden M, Raman L, Vander Trappen A, et al. detection of copy number alterations by shallow whole-genome sequencing of formalin-fixed, paraffin-embedded tumor tissue. Arch Pathol Lab Med 2019.
28. Chaubey A, Shenoy S, Mathur A, et al. Low-pass genome sequencing: validation and diagnostic utility from 409 clinical cases of low-pass genome sequencing for the detection of copy number variants to replace constitutional microarray. J Mol Diagn 2020;22:823-40.