



Research article

A deep-learning-based genomic status estimating framework for homologous recombination deficiency detection from low-pass whole genome sequencing

Yang Liu^{a,1}, Xiang Bi^{b,1}, Yang Leng^{c,1}, Dan Chen^c, Juan Wang^c, Youjia Ma^c, Min-Zhe Zhang^{d,***}, Bo-Wei Han^{c,*}, Yalun Li^{b,**}

^a Department of BC Surgery, Harbin Medical University Cancer Hospital, Harbin, China

^b Department of Breast Surgery, Yantai Yuhuangding Hospital, Yantai, Shandong, China

^c Guangdong Jiyin Biotech Co. Ltd, Shenzhen, Guangdong, China

^d GeneGenieDx Corp, 160 E Tasman Dr, San Jose, CA, USA

ARTICLE INFO

Keywords:

Homologous recombination deficiency
Deep learning
Genome sequencing

ABSTRACT

Genome-wide sequencing allows for prediction of clinical treatment responses and outcomes by estimating genomic status. Here, we developed Genomic Status scan (GSscan), a long short-term memory (LSTM)-based deep-learning framework, which utilizes low-pass whole genome sequencing (WGS) data to capture genomic instability-related features. In this study, GSscan directly surveys homologous recombination deficiency (HRD) status independent of other existing biomarkers. In breast cancer, GSscan achieved an AUC of 0.980 in simulated low-pass WGS data, and obtained a higher HRD risk score in clinical BRCA-deficient breast cancer samples ($p = 1.3 \times 10^{-4}$, compared with BRCA-intact samples). In ovarian cancer, GSscan obtained higher HRD risk scores in BRCA-deficient samples in both simulated data and clinical samples ($p = 2.3 \times 10^{-5}$ and $p = 0.039$, respectively, compared with BRCA-intact samples). Moreover, HRD-positive patients predicted by GSscan showed longer progression-free intervals in TCGA datasets ($p = 0.0011$) treated with platinum-based adjuvant chemotherapy, outperforming existing low-pass WGS-based methods. Furthermore, GSscan can accurately predict HRD status using only 1 ng of input DNA and a minimum sequencing coverage of $0.02 \times$, providing a reliable, accessible, and cost-effective approach. In summary, GSscan effectively and accurately detected HRD status, and provide a broadly applicable framework for disease diagnosis and selecting appropriate disease treatment.

1. Introduction

Genomic status reflects the driving forces of tumorigenesis in many cancers types [1–3]. By detecting genomic biomarkers that are associated with the cancer-driving genomic statuses, we can indirectly predict the clinical responses to many cancer-targeting

* Corresponding author.

** Corresponding author.

*** Corresponding author.

E-mail addresses: mzhang@genegeniedx.com (M.-Z. Zhang), yost.han@fapon.com (B.-W. Han), liyulun106@163.com (Y. Li).

¹ Contributed equally.

<https://doi.org/10.1016/j.heliyon.2024.e26121>

Received 11 December 2023; Accepted 7 February 2024

Available online 9 February 2024

2405-8440/Â© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

treatments [4–6]. One of the most commonly used genomic statuses, homologous recombination deficiency (HRD), is deduced by analyzing the homologous recombination repair (HRR) pathway, and could indicate sensitivity to both platinum-based chemotherapy and poly (ADP-ribose) polymerase inhibitors (PARPi) in many cancers, especially in ovarian, breast, prostate, and pancreatic cancer [7–10].

Genetic or epigenetic driven loss of function in BRCA1 and BRCA2 genes are the best-characterized causes of HRD. Therefore, BRCA1/2 alterations have been the most widely used biomarkers for predicting PARPi response in ovarian and breast cancer [11,12]. Beyond BRCA1/2, several methods to predict HRD status have been developed. Such methods detect a set of genomic features associated with HRD, including genome-wide loss of heterozygosity (LOH), telomeric allelic imbalance (TAI), and large-scale state transition (LST), and have been clinically validated to predict PARPi response [13–16]. Although such methods could reflect the genomic instability status caused by HRD, the resulting HRD scores could be influenced by other types of genomic events unrelated to HRD, which could potentially mislead disease treatment [17,18]. In addition, the HRD-score cutoffs of these methods are disease specific and could not be applied to a broader range of cancer types [19,20].

Furthermore, existing HRD detection methods based on LOH, TAI, and LST rely on hybrid capture-based targeted sequencing with a high DNA input requirement, higher cost, and complex lab operation [14–16]. Sequencing artifacts in formalin-fixed paraffin-embedded (FFPE) samples also interfere with the accuracy of HRD detection [21]. To reduce costs and simplify operations, another method called shallowHRD was developed to detect HRD status by counting the number of intra-chromosome arm copy number variation (CNV) breaks [22,23]. Because shallowHRD uses low-pass whole genome sequencing (WGS) data as input, it predicts HRD status robustly at a lower cost in both FFPE and fresh samples [23]. Nevertheless, shallowHRD has several limitations. First, shallowHRD heavily relies on the accuracy of CNV calling algorithm, which is often restricted by low tumor fraction in biopsy samples and

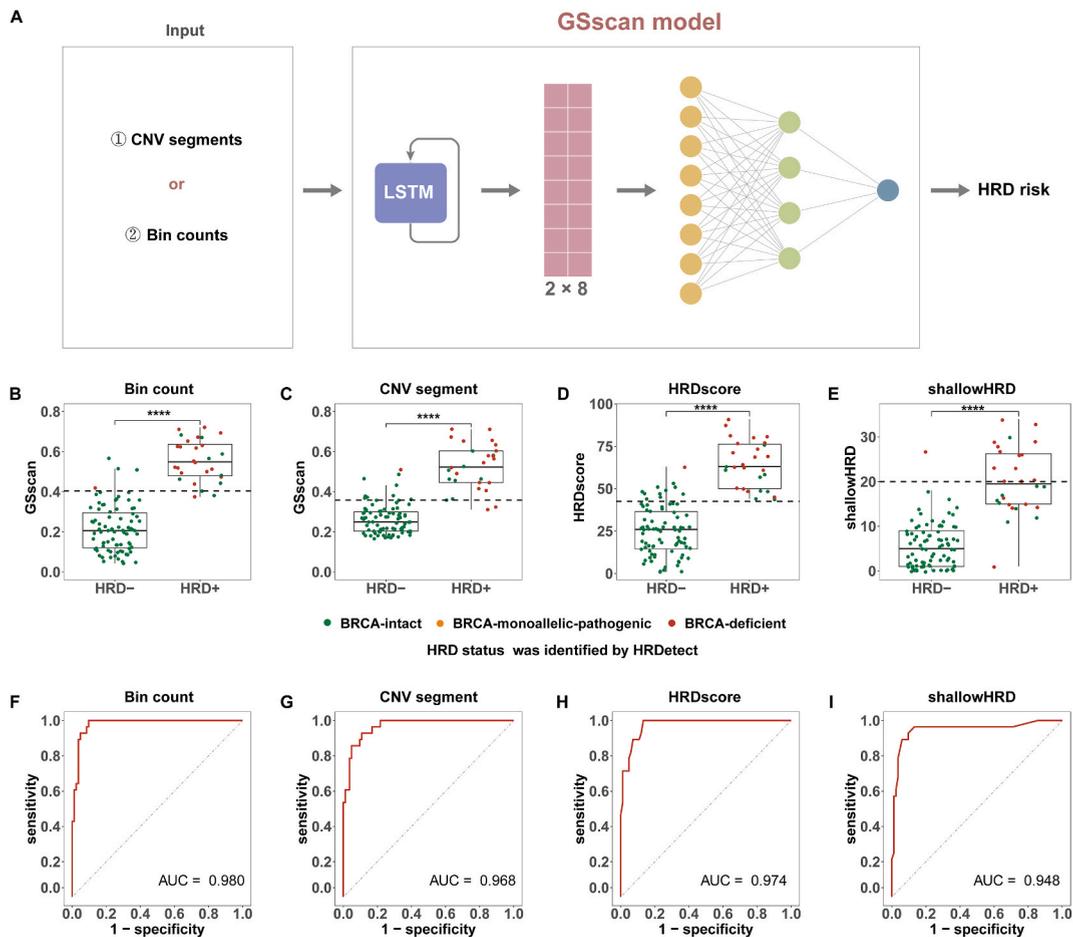


Fig. 1. GSscan model training and validation in simulated low-pass breast cancer WGS data. (A) Overview of the GSscan algorithm. GSscan comprised one LSTM layer with 8 hidden units and two consecutive fully connected layers with 16 and 8 hidden units. Two types of inputs were used for GSscan model training separately. (B–F) HRD prediction between HRD-positive and HRD-negative patients using bin count GSscan model (B), CNV segment model (C), high-depth-based HRDscore (D), and shallowHRD (E), rank-sum test. The dashed line represents the threshold of HRD-positive for each method. GSscan and HRDscore used the threshold determined by the training cohort, and shallowHRD used the threshold provided in the original paper. (F–I) ROC curve of bin count GSscan model (F), CNV segment model (G), high-depth-based HRDscore (H), and shallowHRD (I). ****, $p < 0.0001$, rank-sum test.

insufficient sequencing depth [23]. Second, HRD causes different kinds of instabilities in the genome [24]. ShallowHRD only counts CNV breaks, while discarding much other information of genomic events, which might cause false negatives in some cases. Moreover, the relationship between HRD status predicted by shallowHRD and clinical treatment response has not been validated.

To capture the genomic features that are directly linked to HRD status, we developed GSscan (Genomic Status scan), a long short-term memory (LSTM)-based deep learning framework, which uses the read distribution of WGS for genomic status capturing. We applied GSscan framework to optimize for the HRD predicting performance based on low-pass WGS. Evaluated on simulated low-pass WGS data and actual low-pass WGS data generated from FFPE samples from breast and ovarian cancer patients, GSscan outperforms shallowHRD in terms of predicting HRD status and patient prognosis. Thus, GSscan provides a novel deep-learning based approach which directly survey a broad range of genomic events and could be adopted to predict HRD status and treatment response with high accuracy, ease of operation, and acceptable cost.

2. Methods

2.1. Overview of the GSscan algorithm

GSscan is a long short-term memory (LSTM) based neural network model. It is comprised of one LSTM layer with 8 hidden units and two consecutive fully connected layers with 16 and 8 hidden units. The sigmoid activation function is used to get the final output. This LSTM model is applied separately to each individual chromosome of a sample to predict a chromosome feature score. Then sample level HRD risk score is obtained by taking the average of 22 autosome feature scores. GSscan is implemented using the Pytorch framework. The diagram of GSscan is in Fig. 1A.

2.2. Patients and sample collection

FFPE tumor tissues from 68 breast cancer patients (clinical breast cancer cohort) and 61 ovarian cancer patients (clinical ovarian cancer cohort) with known BRCA1/2 alteration status were retrospectively collected at Harbin Medical University Cancer Hospital, China (Fig. S1). For each specimen, tumor fractions were estimated by pathology reviews, and samples with tumor fraction <5% were excluded for further analysis. Progression-free interval (PFI) was defined as the interval from the end of platinum-based chemotherapy to objective tumor progression or death, “platinum-sensitive (Pt-sensitive)” with a PFI of at least 6 months and “platinum-resistant (Pt-resistant)” if the PFI was less than 6 months. This study was approved by the ethics committee of Harbin Medical University Cancer Hospital, China (Approval No. KY2022-26). All participants provided written informed consent. Detailed sample information is presented in Table S1 and S2.

2.3. DNA extraction and sequencing

As previously described [23], total genomic DNA was extracted from FFPE tissues using the GeneRead DNA FFPE Kit (Qiagen) following the manufacturer’s standard protocol. By default, 50 ng of total DNA was used for library preparation. Libraries for WGS were prepared using the NEBNext Ultra II FS DNA Library Prep Kit for Illumina (NEB). For each sample, a minimum of 2.5 billion bases of raw data was generated using the Illumina Novaseq 6000 platform.

2.4. Data preprocessing

After filtering out low-quality reads using fastp [25], the high-quality reads were aligned to the hg38 human reference genome using BWA [26]. PCR duplicates were identified and removed using Sambamba [27]. CNV segments for each sample were generated using HMMcopy. As the read distribution across the chromosome is associated with genomic status, we split chromosomes into non-overlapping regions (referred to as “genomic bins”) and calculated the read counts of each genomic bin (named “bin counts” for short) to quantify the read distributions. In detail, we evenly tiled the hg38 autosomes into non-overlapping 100 kb bins, and bins with GC content higher than 80% or lower than 20%, or mappability lower than 80% were excluded. Moreover, bins overlapped with Duke blacklisted regions [28] (downloaded from the UCSC Genome Browser) or centromere regions were also excluded. After GC content adjustment using Loess regression and mappability adjustment using Lowess regression, read counts of 10-neighboring bins were merged into non-overlapped 1 Mb bins to reduce the noise introduced by low sequencing depth. To test whether the 1 M bins are large enough to reduce the errors caused by noise, we constructed sequencing libraries for three samples in triplicate. The results from three repetitions, both in terms of bin profiles and final GSscan scores, are highly consistent (Fig. S2), implying the bin size of 1 M is sufficient for GSscan. Finally, Z-score normalized read counts of 1 Mb bins were used as bin counts for further analysis. The pipeline of bin count generation is shown schematically in Fig. S3, and an example of the spectrum of bin profiles is shown in Fig. S4.

2.5. BRCA status and HRD status classification

The classification of somatic and germline BRCA1/2 mutations’ pathogenicity was determined according to the guidelines set by the American College of Medical Genetics and Genomics (ACMG). Samples without any pathogenic BRCA1/2 alterations were categorized as BRCA-intact. Samples with biallelic pathogenic alterations or with monoallelic pathogenic alteration accompanied by a heterozygous deletion of BRCA1 or BRCA2 were classified as BRCA-deficient. The remaining samples were categorized as BRCA-

monoallelic-pathogenic. For high-depth WGS data, we estimated allele-specific copy numbers using PureCN [16,29], and calculated LOH, TAI, and LST scores using ScarHRD [30]. For low-pass WGS data (including simulated data and data generated from clinical samples), shallowHRD and GSscan were used to predict HRD status [29]. For shallowHRD, we used the HRD-positive threshold provided in the original paper [22]. For breast cancer samples with high-depth WGS data, we labeled the HRD status using HRDetect, because HRDetect is highly accurate in breast cancer [31]. For each sample, the HRD scores determined by each method were listed in Table S3.

2.6. Public datasets

For breast cancer, high-depth WGS are from the EGA database (Training cohort: EGAD00001001322, public breast cancer cohort: EGAD00001002129 and EGAD00001002122). For ovarian cancer, high-depth WGS are from EGA database (Public ovarian cancer cohort: EGAD00001003227) and TCGA database (TCGA ovarian cancer cohort) (ref). Enrolled cohorts were described in Fig. S1. Simulated low-pass WGS data were down-sampled from high-depth WGS using Sambamba [27].

2.7. Training of GSscan

Two kinds of inputs, namely CNV segments and bin counts, were used to generate two GSscan models separately (Fig. 1A). To train GSscan, we degraded patient-level data instances to chromosome-level data instances to increase the sample size. All chromosomes would have positive labels if they were from an HRD-positive patient, and negative labels if from an HRD-negative patient. GSscan was trained to predict the chromosome level labels. Binary cross entropy (BCE) was used as the target optimization function and was optimized using stochastic gradient descent (SGD). The model was trained with a batch size of 1. The training stops if it reaches 500 epochs or if validation loss does not improve for 5 epochs. The threshold of each model was determined using training data at the patient-level, and the threshold was 0.361 for the CNV segment model and 0.404 for the bin count model (Fig. S5).

2.8. Statistical analysis

The statistical analyses were conducted using R (v4.1.1). Wilcoxon rank-sum test was utilized to compare variables between two groups, while the Kruskal-Wallis rank-sum test was employed to compare variables among three groups using the ggsingif packages (0.6.3) in R. ROC curve analyses were performed using the pROC package (v1.18.0) in R to calculate the area under the curve (AUC), sensitivity, and specificity. The best cutoff point was determined based on the maximum sum of sensitivity and specificity. Spearman's correlation coefficient was employed to assess the correlation between two datasets. Kaplan-Meier survival curves were analyzed using the log-rank test with the survival (v3.2-11) and survminer (v0.4.9) packages. A significance level of $p < 0.05$ was considered statistically significant.

3. Results

3.1. Read distribution-based GSscan could detect HRD status in simulated low-pass breast cancer WGS data

To train the low-pass WGS-based GSscan models, we collected 80 breast cancer high-depth WGS data (the training cohort) with known HRD statuses identified by HRDetect, which was well-validated in breast cancer [31,32]. We down-sampled all high-depth data to $0.5 \times$, generated bin counts to represent read distribution patterns, and subsequently trained the GSscan model using bin counts across each chromosome (Fig. S3, detailed in Method). Because most of the existing HRD biomarkers are based on CNV, we also trained a CNV segment-based GSscan model (Fig. 1A). For subsequent analysis, we determined the best cutoff of each model based on training cohort (Fig. S5).

Although both bin count model and CNV segment model achieve AUCs of 1.00 at sample-level in training cohort, the bin count model performed better than CNV segment model at chromosome-level (AUC = 0.891 vs. 0.828, $p = 2.9 \times 10^{-9}$, DeLong's test, Fig. S6). Notably, for the CNV segment model, the chromosome-level AUC decreases with the decrease in chromosome length. Correspondingly, the variance of chromosome-level AUCs is lower in the bin count model (standard deviation = 0.061 in the bin count model and 0.104 in the CNV segment model, Fig. S6), suggesting bin count-based GSscan is more stable across chromosomes with different lengths, thus performs better at the chromosome level.

3.2. Validation of GSscan's performance in public breast cancer cohort

To evaluate the performance of GSscan, we simulated low-depth whole-genome sequencing data ($0.5 \times$) using 111 breast cancer samples (the public breast cancer cohort) with known HRD status and calculated HRD risk scores using GSscan. In this cohort, all samples were either BRCA-deficient or BRCA-intact. As expected, HRD-positive samples had significantly increased GSscan HRD risk scores in the bin count and CNV segment models ($p = 3.8 \times 10^{-14}$ and 1.5×10^{-13} , respectively, rank-sum test, Fig. 1B and C) compared with HRD-negative samples. At chromosome level, the bin count model achieved higher AUC than the CNV segment model (AUC = 0.770 vs. 0.740, $p = 0.017$, DeLong's test, Fig. S7), and at the sample-level, the bin count model performed better (AUC = 0.980, with a sensitivity of 0.893 and a specificity of 0.952, Fig. 1F) when compared with the CNV segment model (AUC = 0.968 with a sensitivity of 0.857 and a specificity of 0.904, Fig. 1G). Considering the bin count model performs better and doesn't rely on the third-

party CNV calling software, we selected the bin count model as the final GSscan model for all further discussions. Using this GSscan model and its corresponding threshold, 19 out of 20 (95%) BRCA-deficient patients were identified as HRD-positive (Fig. 1B), suggesting GSscan could accurately detect BRCAness in HRD patients with BRCA1/2 alteration.

It's worth noting that GSscan has comparable performance with existing HRD detection methods. We compared GSscan with HRDscore (the mean of LOH, TAI, and LST scores) [15], which is generated from high-depth sequencing and widely used in clinical practice, and found the AUCs of the two methods were close (AUC = 0.980 vs. 0.974, $p = 0.7235$, Delong's test, Fig. 1H). Moreover, when compared with shallowHRD, which is an existing method for HRD detection from low-pass WGS sequencing, although not statistically significant, GSscan achieved a slightly higher AUC than shallowHRD (AUC = 0.980 vs. 0.948, $p = 0.3215$, Delong's test, Fig. 1I).

3.3. GSscan's performance in public ovarian cancer cohort

As GSscan was trained in breast cancer, we investigated whether GSscan could be directly apply to other cancer types without any modification. We simulated low-pass data from 63 ovarian cancer cases (the public ovarian cancer cohort). As the ovarian cancer cohort lacked HRD status labels, HRDscores generated from high-depth WGS were used to label the samples as "HRD-high" (HRDscore \geq the median HRDscore of all samples) or HRD-low (HRDscore < the median HRDscore). Both GSscan and shallowHRD scores are significantly elevated in HRD-high samples compared with HRD-low samples using preset threshold ($p = 2.3 \times 10^{-5}$ and 2.0×10^{-7} , respectively, Fig. 2A and B), suggesting GSscan could accurately predict HRD status in ovarian cancer with a threshold consistent with that used in breast cancer.

In addition, we generated LOH, TAI, LST, HRDscore, shallowHRD and GSscan scores in the same ovarian cancer cohort. In this cohort, HRDscore highly correlates with LST, TAI and LOH as expected, since it is the average of these three (Spearman's correlation coefficient $r = 0.943, 0.889$, and 0.744 , respectively, Fig. 2C) [15]. ShallowHRD is highly correlated with HRDscore, and LST in particular (Spearman's correlation coefficient $r = 0.841$, and 0.877 , respectively, Fig. 2A), by which we suspect that the captured underlying biological events of shallowHRD and LST are mostly similar [22]. GSscan is also highly correlated with HRDscores (Spearman's correlation coefficient $r = 0.693$, Fig. 2C) as well as shallowHRD scores (Spearman's correlation coefficient $r = 0.777$, Fig. 2C), albeit not as significantly correlated as the aforementioned pairs, suggesting that GSscan may actually capture different, or a wider range of underlying biological events compared with existing methods.

3.4. Evaluation of GSscan on different sequencing depths

Sequencing depth is a crucial parameter in next-generation sequencing (NGS)-based tests, where higher sequencing depth improves the robustness of read distribution but also increases the cost of testing. Our above results demonstrated that GSscan can accurately predict HRD status in low-depth data at $0.5 \times$ coverage. To evaluate its performance in even lower sequencing depths, we mimicked data at lower depths ($0.2 \times, 0.1 \times, 0.05 \times, 0.02 \times, 0.01 \times$), and predicted HRD status using the GSscan model trained on $0.5 \times$ WGS.

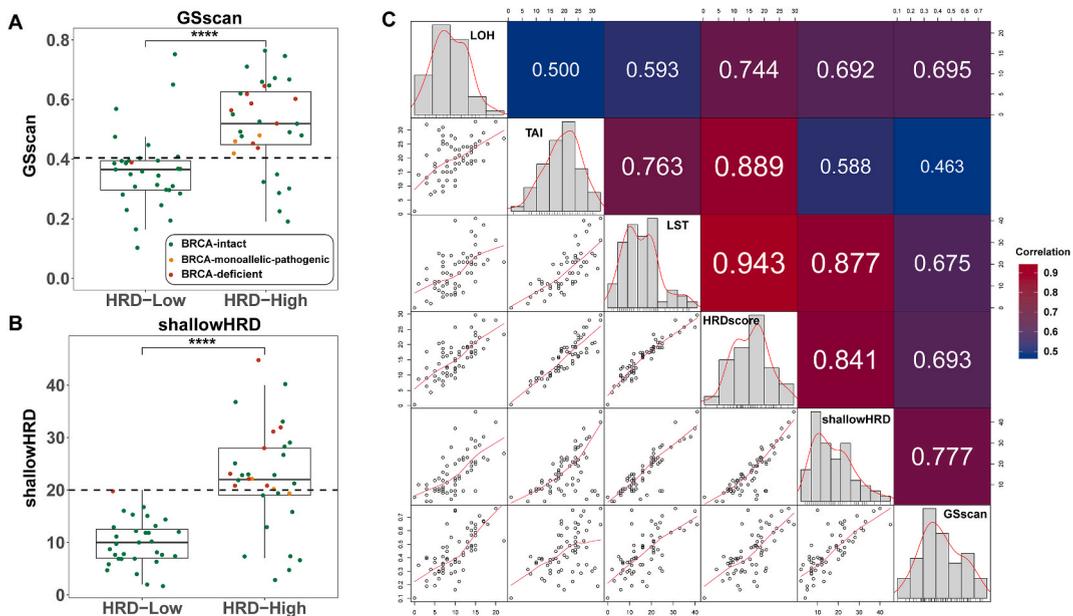


Fig. 2. Performance of GSscan on low-pass ovarian cancer WGS data. (A–B) GSscan and shallowHRD between HRD-low and HRD-high patients, rank-sum test. The dashed line represents the threshold of HRD-positive for each method. ****, $p < 0.0001$, rank-sum test. (C) The correlations between GSscan scores and other commonly used HRD features.

Remarkably, even at a coverage depth as low as $0.02 \times$, GSscan achieved an AUC of 0.969 (sensitivity of 0.857, and specificity of 0.976, Fig. 3A) and a high correlation with $0.5 \times$ data (Spearman's correlation coefficient $r = 0.916$, Fig. 3B) in the public breast cancer cohort. Similarly, in $0.02 \times$ coverage data, GSscan achieved a high correlation with $0.5 \times$ data in the public ovarian cancer cohort (Spearman's correlation coefficient $r = 0.942$, Fig. S8). These results suggest that GSscan could achieve acceptable performance even on ultra-low-depth sequencing data.

3.5. Performance of GSscan in clinical breast and ovarian cancer cohorts

To further investigate the performance of GSscan in clinical samples, we collected breast and ovarian cancer tissue samples with known BRCA status. The samples with tumor fraction greater than 30% that meet the criteria of HRD detection guidelines were selected for performance evaluation [15,33,34]. As we speculate GSscan captured relatively stable features, we used the same thresholds of GSscan among all cohorts. Among 50 breast cancer samples, the median GSscan score was significantly higher among BRCA-deficient samples than BRCA-intact samples ($p = 1.3 \times 10^{-5}$, rank-sum test), and all BRCA-deficient samples were identified as HRD-positive (Fig. 4A). Additionally, 30.6% (11 out of 36) BRCA-intact patients were identified as HRD-positive, which is similar to previous studies (ref). Similarly, among 47 ovarian cancer samples with tumor fraction greater than 30%, the GSscan scores of BRCA-deficient samples were significantly higher than BRCA-intact samples ($p = 0.039$, rank-sum test, Fig. 4B). 13 out of 14 (92.9%) BRCA-deficient patients were identified as HRD-positive, and 17 out of 27 (62.9%) BRCA-intact patients were identified as HRD-positive. Consistent with previous studies, in BRCA-intact and BRCA-monoallelic-pathogenic groups, the HRD-positive fractions are higher in the ovarian cancer cohort than in the breast cancer cohort [16,35,36]. Moreover, even among samples with tumor fraction between 5 and 30%, 5 out of 7 (71.4%) BRCA-deficient breast cancer patients and 4 out of 5 (80%) BRCA-deficient ovarian cancer patients were unequivocally identified as HRD-positive, suggesting GSscan could identify HRD-positive samples even when the tumor fraction is relatively low (Fig. S9).

We analyzed the same batch of samples using shallowHRD. Surprisingly, although the shallowHRD score is higher in BRCA-deficient breast cancer samples ($p = 0.041$, rank-sum test, Fig. 4C), the default threshold seems not suitable for this cohort [22] (dashed line in Fig. 4C). Besides, using shallowHRD, there was no significant difference between the BRCA-deficient and BRCA-intact groups in ovarian cancer ($p = 0.27$, rank-sum test, Fig. 4D).

For patients unable to undergo tumor resection, only limited amounts of DNA could be obtained from the biopsy tissues. We examined the robustness of GSscan with respect to low amount of DNA input in a subset of clinical samples. Rather than using 50 ng DNA input as recommended, we tested the performance of GSscan with 1 ng, 5 ng, and 10 ng of DNA inputs. Even using 1 ng DNA input, the GSscan scores were comparable to that generated from 50 ng input DNA (Pearson correlation coefficient $r = 0.829$, Fig. S10).

3.6. Association of GSscan with platinum-based chemotherapy outcomes

Since HRD status directly convey sensitivity to platinum-based therapy [37–41], if GSscan can accurately identifying HRD status, it should be able to predicts platinum-based chemotherapy outcomes. To evaluate the applicability of GSscan on such treatment response, WGS data from 43 TCGA ovarian cancer patients who had received prior platinum-based adjuvant chemotherapy with PFI data available were analyzed. We simulated low-pass WGS data and evaluated HRD status using GSscan, a significantly longer PFI was observed in GSscan HRD-positive patients than those with GSscan HRD-negative (median 568 vs 315 days, $p = 0.0011$, Log-rank test, Fig. 5A). And all three Pt-resistant patients were identified as HRD-negative (Fig. S11A), suggesting GSscan could serve as an independent prognosis marker. Moreover, in patients without BRCA1/2 mutation, HRD-positive patients still had longer PFI ($p = 0.0033$, Fig. S12A), indicating that GSscan has the potential to assist in drug selection for BRCA-intact patients. Additionally, a considerable

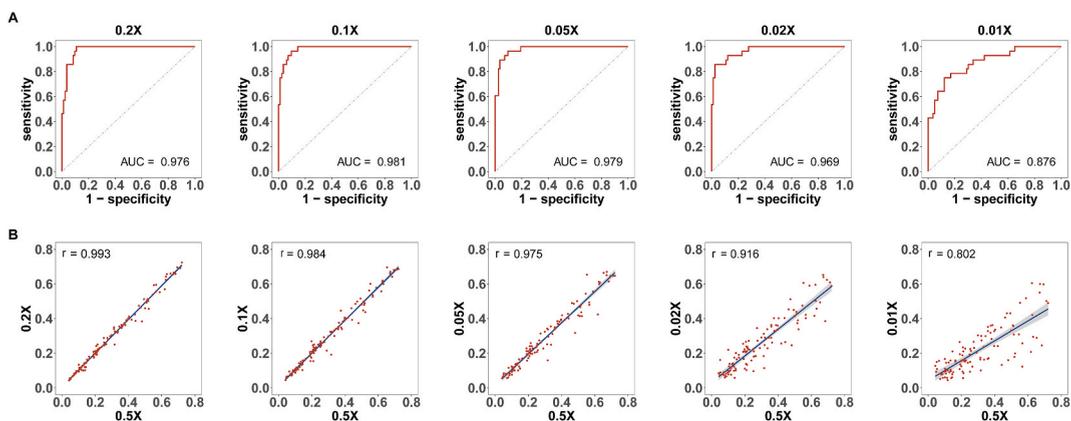


Fig. 3. Performance of GSscan on different sequencing depths. (A) ROC curve of GSscan score on sequencing depth of $0.2 \times$, $0.1 \times$, $0.05 \times$, $0.02 \times$, and $0.01 \times$. (B) Correlation between GSscan score generated by $0.5 \times$ data and GSscan score generated by sequencing depth of $0.2 \times$, $0.1 \times$, $0.05 \times$, $0.02 \times$, and $0.01 \times$.

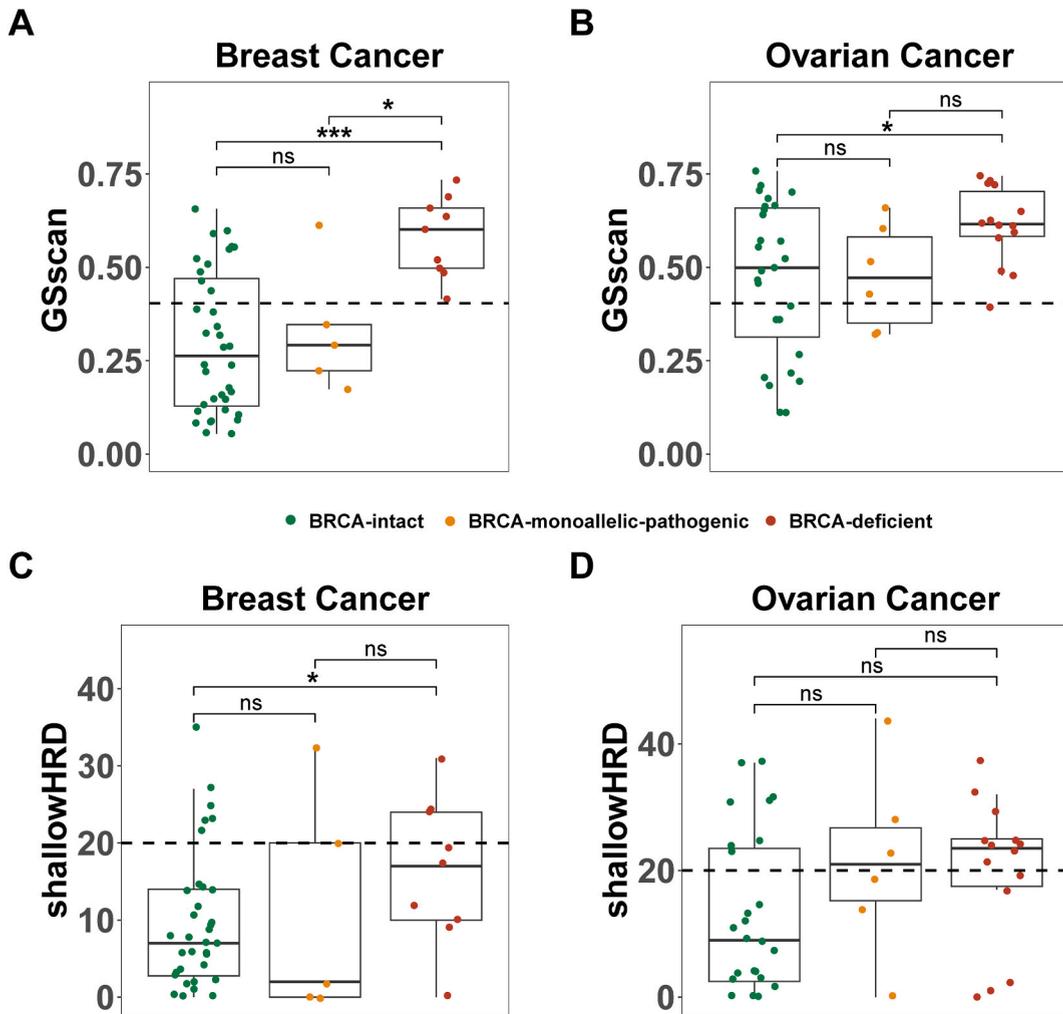


Fig. 4. GSscan score in clinical cohorts. (A–B) GSscan scores among patients with BRCA-intact, BRCA-monoallelic-pathogenic, and BRCA-deficient in breast and ovarian cancer. (C–D) ShallowHRD scores among patients with BRCA-intact, BRCA-monoallelic-pathogenic, and BRCA-deficient in breast and ovarian cancer. The dashed lines represent the threshold of HRD-positive. *, $p < 0.05$, **, $p < 0.01$, ***, $p < 0.001$, rank-sum test.

trend toward significance was observed in PFI between GSscan HRD-positive and GSscan HRD-negative in our clinical ovarian cancer cohort ($p = 0.085$, Log-rank test, Fig. 5C). Contrastively, there was no significant difference in PFI between groups with distinct HRD status defined by shallowHRD in both TCGA and clinical ovarian cancer cohorts (TCGA ovarian cancer cohort, $p = 0.77$, Fig. 5B; clinical ovarian cancer cohort, $p = 0.15$, Log-rank test, Fig. 5D).

4. Discussion

Genomic status reflects crucial characteristics of almost all human cancers and has the potential to predict treatment response, long-term prognosis, minimal residual disease et al. [4–6,24,42,43]. In this study, we developed a deep-learning based framework, GSscan, to capture a wide range of genomic features and predict genomic status using low-pass WGS. Therefore, GSscan could be directly adopted to survey HRD status and detects HRD without reliance on existing biomarkers. In breast and ovarian cancer patients, GSscan generated comparable results to existing methods and yielded HRD-positive rates in line with epidemiological expectations [16,34]. Moreover, GSscan is associated with the prognosis of ovarian cancer patients with platinum-based adjuvant chemotherapy, implying GSscan could be served as the predictor of the response to platinum-based or PARPi therapy. Our findings demonstrated both the accuracy and robustness of HRD detection by GSscan, and provide evidence for its clinical relevance.

Existing HRD detection methods can be mainly categorized into several approaches: large genomic aberration based (LOH, TAI, LST, and shallowHRD) [15,22], mutational signature based (Signature 3, SigMA) [44,45], and combination of both features (HRDetect, CHORD) [31,35]. Except for shallowHRD, these approaches need high-depth WGS or hybrid capture-based targeted sequencing, which is either technically complex or highly costly [15,31,35,44,45]. Because GSscan is based on genome-wide coverage profiles generated by low-pass WGS, it provides HRD detection with moderate cost and simple operation. Moreover, we demonstrated

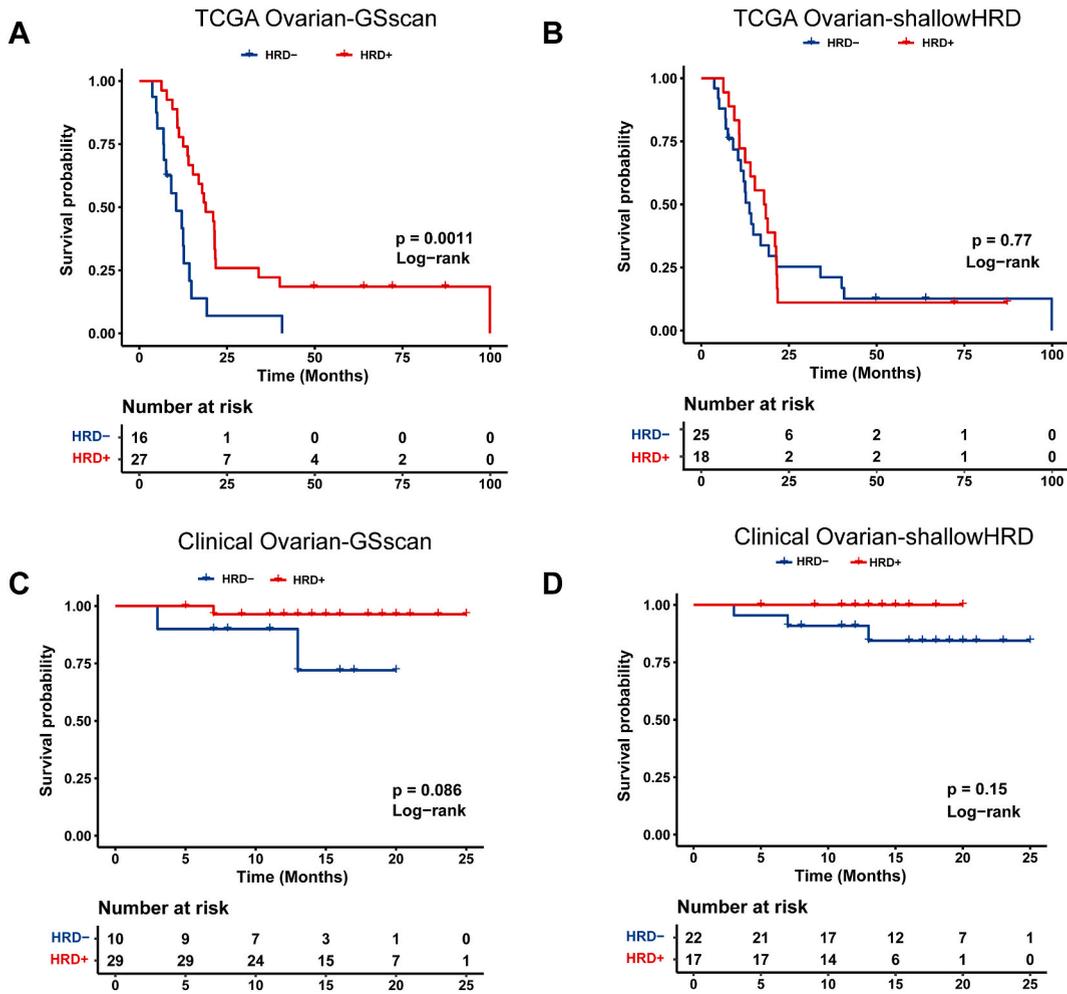


Fig. 5. GSscan in ovarian cancer patients received prior platinum-based adjuvant chemotherapy. (A–B) Kaplan–Meier curves for HRD-positive and HRD-negative patients identified by GSscan or shallowHRD in TCGA ovarian cancer. (C–D) Kaplan–Meier curves for HRD-positive and HRD-negative patients identified by GSscan or shallowHRD in clinical ovarian cancer cohort.

GSscan can adapt to WGS data with low DNA input amount, thus having the potential to be applied to plasma cell-free DNA and minimal biopsy samples. Hence GSscan could provide accurate HRD tests for patients whose health conditions does not allow for surgery.

The calculation of existing HRD scores relies on pre-determined length cutoff of genomic events: the LOH score is the number of LOH regions longer than 15 Mb, the TAI score only counts the allelic imbalance region longer than 11 Mb, and the LST score only considers the regions longer than 10 Mb [15]. Similarly, shallowHRD uses 10 Mb length as a filter of independent CNV [22]. Although the HRD scores generated using these cutoffs were highly predictive of HRD status, the biological basis underlying is not well understood. Thus, the optimal length cutoff of different chromosomes or genomic regions may differ for these different methods. Additionally, the calculation of these scores is based on CNV, which is heavily influenced by the parameters and cutoffs of the CNV calling algorithm [23,34]. Because GSscan uses a deep-learning based framework, it automatically learns the best parameters and is not limited by the accuracy of CNV calling algorithm [23,34]. Additionally, we have demonstrated that in ovarian cancer, the same threshold as determined in the breast cancer training cohort could still apply, implying GSscan captured HRD status features that may be universal to multiple cancer types. Thus, GSscan directly captures genomic status features independent of existing HRD markers and provides room for further improvement.

Genomic status provides insight into the biological mechanisms involved in tumorigenesis and has shown their applicability in cancer [3,6,24,46]. Because our deep-learning based GSscan can not only fit with HRD status but also be trained to predict other genomic statuses, GSscan has the potential to provide companion diagnostic features and predict treatment responses for other cancer targeting drugs and treatments. Other than those features based on somatic point mutation (which requires high-depth sequencing), several genomic features have proven successful in oncological management and targeted therapies [6,24,43,46–49]. For example, microsatellite instability caused by mismatch repair deficiency and CNV burden could indicate treatment response to immune checkpoint inhibitors [6,47–49], copy number-based signatures could indicate prognosis in serous ovarian cancer and pancreatic

adenocarcinoma [43,46], and chromosomal copy number heterogeneity could predict patient survival across multiple cancer types [24]. Because many of such genomic features can be deduced by the distribution of sequencing reads, our deep-learning framework has the potential to capture these genomic features using only low-pass WGS data. Moreover, there is a growing number of drugs and treatments that directly target different forms of genomic instability, and the response might be predictable using specific genomic features [5,50,51]. Therefore, our GSscan framework might provide the opportunity to predict the response to some of these drugs or treatments, and provides overall guidance for multiple therapies in a single test, which enables the provision of personalized disease management in the future.

Compared with existing HRD detection methods, GSscan not only offers accurate HRD status assessment and treatment prognosis evaluation but also enhances cost-effectiveness, usability, and result stability through the utilization of low-pass WGS. Furthermore, GSscan exhibits enhanced scalability in comparison to existing methods. It has the potential for utilization in liquid biopsies, expansion to capture additional genomic status features, and the ability to predict the response of other drugs or treatments. Therefore, GSscan is the only existing method capable of simultaneously predicting the response of multiple drugs and treatments in a single cost-effective test. Meanwhile, GSscan has several potential limitations. The deep learning GSscan procedure can have several other potential sources of mistakes and misrecognition, including FFPE sequencing noise, different sequencing platforms, changes in sequencing and library preparation reagents, and et al. Additionally, shallow sequencing may also introduce some limitations. It is worth noting that 6 cases were misclassified in the validation cohort of breast cancer, where HRD status was identified using HRDetect. In cases where the diagnostics are not consistent between GSscan and HRDetect, there is no significant difference between GSscan and HRDscore, which is based on large-scale genomic variations. However, the GSscan scores differs greatly from Signature 3, which is based on single-nucleotide variations. Therefore, for a small subset of HRD patients with inconsistencies between single-nucleotide-level and large fragment-level variations, shallow sequencing-based methods might be inconsistent with the existing approaches due to their inability to capture single-nucleotide information. Moreover, there is still a significant need for an effective verification of therapy response, especially for the HRD-positive patients without BRCA1/2 mutation. To tackle this challenge, the utilization of patient-derived xenograft (PDX) models and patient-derived organoids, along with in vitro drug sensitivity assays, will aid in further validating the accuracy of HRD prediction. And further clinical trials would verify the effectiveness of HRD test for clinical application.

In conclusion, using low-pass WGS, we established a genomic feature capturing and status detection framework, GSscan, to predict HRD status and treatment response. Independent of existing HRD biomarkers, GSscan could accurately and directly predict HRD status and treatment response in breast and ovarian cancer. Further tuning GSscan on other treatment-associated genomic statuses might increase the scope of the applications of GSscan, and shed light on guiding various treatment decisions.

Ethics statement

This study was approved by the ethics committee of Harbin Medical University Cancer Hospital, China (Approval No. KY2022-26). All participants provided written informed consent.

CRedit authorship contribution statement

Yang Liu: Writing – original draft, Methodology, Conceptualization. **Xiang Bi:** Investigation, Formal analysis, Data curation. **Yang Leng:** Visualization, Formal analysis. **Dan Chen:** Data curation. **Juan Wang:** Methodology. **Youjia Ma:** Visualization, Software. **Min-Zhe Zhang:** Writing – review & editing, Conceptualization. **Bo-Wei Han:** Writing – original draft, Validation, Supervision, Methodology. **Yalun Li:** Writing – review & editing, Supervision, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Yang Leng, Dan Chen, Juan Wang, Youjia Ma, and Bo-Wei Han are employees Guangdong Jiyin Biotech Co. Ltd. Min-Zhe Zhang is an employee of GeneGenieDx Corp. All other authors declare no conflict of interest.

Acknowledgments

We thank Min-Zhe Zhang for accessing and analyzing the public data, and Yang Leng and Youjia Ma for analyzing the low-pass WGS generated from clinical samples. The work was supported by Heilongjiang Provincial Health Commission Scientific Research Foundation 2019-059.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2024.e26121>.

References

- [1] D. Hanahan, Hallmarks of cancer: new dimensions, *Cancer Discov.* 12 (2022) 31–46, <https://doi.org/10.1158/2159-8290.CD-21-1059>.
- [2] D. Hanahan, R.A. Weinberg, Hallmarks of cancer: the next generation, *Cell* 144 (2011) 646–674, <https://doi.org/10.1016/j.cell.2011.02.013>.
- [3] E. Wang, N. Zaman, S. McGee, J.S. Milanese, A. Masoudi-Nejad, M. O'Connor-McCourt, Predictive genomics: a cancer hallmark network framework for predicting tumor clinical phenotypes using genome sequencing data, *Semin. Cancer Biol.* 30 (2015) 4–12, <https://doi.org/10.1016/j.semcancer.2014.04.002>.
- [4] O.M. Sieber, K. Heinemann, I.P. Tomlinson, Genomic instability—the engine of tumorigenesis? *Nat. Rev. Cancer* 3 (2003) 701–708, <https://doi.org/10.1038/nrc1170>.
- [5] S.R. Hengel, M.A. Spies, M. Spies, Small-molecule inhibitors targeting DNA repair and DNA repair deficiency in Research and cancer therapy, *Cell Chem. Biol.* 24 (2017) 1101–1119, <https://doi.org/10.1016/j.chembiol.2017.08.027>.
- [6] C. Luchini, F. Bibeau, M.J.L. Ligtenberg, N. Singh, A. Nottegar, T. Bosse, R. Miller, N. Riaz, J.Y. Douillard, F. Andre, A. Scarpa, ESMO recommendations on microsatellite instability testing for immunotherapy in cancer, and its relationship with PD-1/PD-L1 expression and tumour mutational burden: a systematic review-based approach, *Ann. Oncol.* 30 (2019) 1232–1243, <https://doi.org/10.1093/annonc/mdz116>.
- [7] W.D. den Brok, K.A. Schrader, S. Sun, A.V. Tinker, E.Y. Zhao, S. Aparicio, K.A. Gelmon, Homologous recombination deficiency in breast cancer: a clinical review, *JCO Precis Oncol* 1 (2017) 1–13, <https://doi.org/10.1200/PO.16.00031>.
- [8] R.E. Miller, A. Leary, C.L. Scott, V. Serra, C.J. Lord, D. Bowtell, D.K. Chang, D.W. Garsed, J. Jonkers, J.A. Ledermann, et al., ESMO recommendations on predictive biomarker testing for homologous recombination deficiency and PARP inhibitor benefit in ovarian cancer, *Ann. Oncol.* 31 (2020) 1606–1622, <https://doi.org/10.1016/j.annonc.2020.08.2102>.
- [9] J.A. Ledermann, Y. Drew, R.S. Kristeleit, Homologous recombination deficiency and ovarian cancer, *Eur. J. Cancer* 60 (2016) 49–58, <https://doi.org/10.1016/j.ejca.2016.03.005>.
- [10] I. Vergote, A. Gonzalez-Martin, I. Ray-Coquard, P. Harter, N. Colombo, P. Pujol, D. Lorusso, M.R. Mirza, B. Brasiuniene, R. Madry, et al., European experts consensus: BRCA/homologous recombination deficiency testing in first-line ovarian cancer, *Ann. Oncol.* 33 (2022) 276–287, <https://doi.org/10.1016/j.annonc.2021.11.013>.
- [11] P. Pujol, M. Barberis, P. Beer, E. Friedman, J.M. Piulats, E.D. Capoluongo, J. Garcia Foncillas, I. Ray-Coquard, F. Penault-Llorca, W.D. Foulkes, et al., Clinical practice guidelines for BRCA1 and BRCA2 genetic testing, *Eur. J. Cancer* 146 (2021) 30–47, <https://doi.org/10.1016/j.ejca.2020.12.023>.
- [12] S. Paluch-Shimon, F. Cardoso, C. Sessa, J. Balmana, M.J. Cardoso, F. Gilbert, E. Senkus, E.G. Committee, Prevention and screening in BRCA mutation carriers and other breast/ovarian hereditary cancer syndromes: ESMO Clinical Practice Guidelines for cancer prevention and screening, *Ann. Oncol.* 27 (2016) v103–v110, <https://doi.org/10.1093/annonc/mdw327>.
- [13] M. Toh, J. Ngeow, Homologous recombination deficiency: cancer predispositions and treatment implications, *Oncol.* 26 (2021) e1526–e1537, <https://doi.org/10.1002/onco.13829>.
- [14] M.L. Telli, K.M. Timms, J. Reid, B. Hennessy, G.B. Mills, K.C. Jensen, Z. Szallasi, W.T. Barry, E.P. Winer, N.M. Tung, et al., Homologous recombination deficiency (HRD) score predicts response to platinum-containing neoadjuvant chemotherapy in patients with triple-negative breast cancer, *Clin. Cancer Res.* 22 (2016) 3764–3773, <https://doi.org/10.1158/1078-0432.CCR-15-2477>.
- [15] K.M. Timms, V. Abkevich, E. Hughes, C. Neff, J. Reid, B. Morris, S. Kalva, J. Potter, T.V. Tran, J. Chen, et al., Association of BRCA1/2 defects with genomic scores predictive of DNA damage repair deficiency among breast cancer subtypes, *Breast Cancer Res.* 16 (2014) 475, <https://doi.org/10.1186/s13058-014-0475-x>.
- [16] H. Wen, Z. Feng, Y. Ma, R. Liu, Q. Ou, Q. Guo, Y. Shen, X. Wu, Y. Shao, H. Bao, X. Wu, Homologous recombination deficiency in diverse cancer types and its correlation with platinum chemotherapy efficiency in ovarian cancer, *BMC Cancer* 22 (2022) 550, <https://doi.org/10.1186/s12885-022-09602-4>.
- [17] S. Thiagalangam, R.L. Foy, K.H. Cheng, H.J. Lee, A. Thiagalangam, J.F. Ponte, Loss of heterozygosity as a predictor to map tumor suppressor genes in cancer: molecular basis of its occurrence, *Curr. Opin. Oncol.* 14 (2002) 65–72, <https://doi.org/10.1097/00001622-200201000-00012>.
- [18] C.D. Steele, A. Abbasi, S.M.A. Islam, A.L. Bowes, A. Khandekar, K. Haase, S. Hames-Pathi, D. Ajayi, A. Verfaillie, P. Dhami, et al., Signatures of copy number alterations in human cancer, *Nature* 606 (2022) 984–991, <https://doi.org/10.1038/s41586-022-04738-6>.
- [19] M.D. Stewart, D. Merino Vega, R.C. Arend, J.F. Baden, O. Barbash, N. Beaubier, G. Collins, T. French, N. Ghahramani, P. Hinson, et al., Homologous recombination deficiency: concepts, definitions, and assays, *Oncol.* 27 (2022) 167–174, <https://doi.org/10.1093/oncolo/oyab053>.
- [20] H. Takaya, H. Nakai, S. Takamatsu, M. Mandai, N. Matsumura, Homologous recombination deficiency status-based classification of high-grade serous ovarian carcinoma, *Sci. Rep.* 10 (2020) 2757, <https://doi.org/10.1038/s41598-020-59671-3>.
- [21] H. Do, A. Dobrovic, Sequence artifacts in DNA from formalin-fixed tissues: causes and strategies for minimization, *Clin. Chem.* 61 (2015) 64–71, <https://doi.org/10.1373/clinchem.2014.223040>.
- [22] A. Eekhoutte, A. Houy, E. Manie, M. Reverdy, I. Bieche, E. Marangoni, O. Goundiam, A. Vincent-Salomon, D. Stoppa-Lyonnet, F.C. Bidard, et al., ShallowHRD: detection of homologous recombination deficiency from shallow whole genome sequencing, *Bioinformatics* 36 (2020) 3888–3889, <https://doi.org/10.1093/bioinformatics/btaa261>.
- [23] Y. Liu, Y. Li, M.Z. Zhang, D. Chen, Y. Leng, J. Wang, B.W. Han, J. Wang, Homologous recombination deficiency prediction using low-pass whole genome sequencing in breast cancer, *Cancer Genet* 272 (273) (2023) 35–40, <https://doi.org/10.1016/j.cancergen.2023.02.001>.
- [24] E. van Dijk, T. van den Bosch, K.J. Lenos, K. El Makrini, L.E. Nijman, H.F.B. van Essen, N. Lansu, M. Boekhout, J.H. Hageman, R.C. Fitzgerald, et al., Chromosomal copy number heterogeneity predicts survival rates across cancers, *Nat. Commun.* 12 (2021) 3188, <https://doi.org/10.1038/s41467-021-23384-6>.
- [25] S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor, *Bioinformatics* 34 (2018) i884–i890, <https://doi.org/10.1093/bioinformatics/bty560>.
- [26] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform, *Bioinformatics* 25 (2009) 1754–1760, <https://doi.org/10.1093/bioinformatics/btp324>.
- [27] A. Tarasov, A.J. Vilella, E. Cuppen, I.J. Nijman, P. Sambamba Prins, Fast processing of NGS alignment formats, *Bioinformatics* 31 (2015) 2032–2034, <https://doi.org/10.1093/bioinformatics/btv098>.
- [28] H.M. Amemiya, A. Kundaje, A.P. Boyle, The ENCODE blacklist: identification of problematic regions of the genome, *Sci. Rep.* 9 (2019) 9354, <https://doi.org/10.1038/s41598-019-45839-z>.
- [29] M. Riemer, A.P. Singh, A.R. Brannon, K. Yu, C.D. Campbell, D.Y. Chiang, M.P. Morrissey, C.N. Pure, Copy number calling and SNV classification using targeted short read sequencing, *Source Code Biol. Med.* 11 (2016) 13, <https://doi.org/10.1186/s13029-016-0060-z>.
- [30] Z. Sztupinski, M. Dioisy, M. Krzystanek, L. Reiniger, I. Csabai, F. Favero, N.J. Birkbak, A.C. Eklund, A. Syed, Z. Szallasi, Migrating the SNP array-based homologous recombination deficiency measures to next generation sequencing data of breast cancer, *NPJ Breast Cancer* 4 (2018) 16, <https://doi.org/10.1038/s41523-018-0066-6>.
- [31] H. Davies, D. Glodzik, S. Morganello, L.R. Yates, J. Staaf, X. Zou, M. Ramakrishna, S. Martin, S. Boyault, A.M. Sieuwerts, et al., HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures, *Nat. Med.* 23 (2017) 517–525, <https://doi.org/10.1038/nm.4292>.
- [32] E.Y. Zhao, Y. Shen, E. Pleasance, K. Kasaian, S. Leelakumari, M. Jones, P. Bose, C. Ch'ng, C. Reisle, P. Eirew, et al., Homologous recombination deficiency and platinum-based therapy outcomes in advanced breast cancer, *Clin. Cancer Res.* 23 (2017) 7521–7530, <https://doi.org/10.1158/1078-0432.CCR-17-1941>.
- [33] C. Fumagalli, I. Betella, A. Ranghiero, E. Guerini-Rocco, G. Bonaldo, A. Rappa, D. Vacirca, N. Colombo, M. Barberis, In-house testing for homologous recombination repair deficiency (HRD) testing in ovarian carcinoma: a feasibility study comparing AmoyDx HRD Focus panel with Myriad myChoiceCDx assay, *Pathologica* 114 (2022) 288–294, <https://doi.org/10.32074/1591-951X-791>.
- [34] D. Chen, M. Shao, P. Meng, C. Wang, Q. Li, Y. Cai, C. Song, X. Wang, T. Shi, GSA: an independent development algorithm for calling copy number and detecting homologous recombination deficiency (HRD) from target capture sequencing, *BMC Bioinf.* 22 (2021) 562, <https://doi.org/10.1186/s12859-021-04487-9>.
- [35] L. Nguyen, W.M.M. J. A. Van Hoeck, E. Cuppen, Pan-cancer landscape of homologous recombination deficiency, *Nat. Commun.* 11 (2020) 5584, <https://doi.org/10.1038/s41467-020-19406-4>.

- [36] E. Rempel, K. Kluck, S. Beck, I. Ourailidis, D. Kazdal, O. Neumann, A.L. Volckmar, M. Kirchner, H. Goldschmid, N. Pfarr, et al., Pan-cancer analysis of genomic scar patterns caused by homologous repair deficiency (HRD), *npj Precis. Oncol.* 6 (2022) 36, <https://doi.org/10.1038/s41698-022-00276-6>.
- [37] A. Bhattacharyya, U.S. Ear, B.H. Koller, R.R. Weichselbaum, D.K. Bishop, The breast cancer susceptibility gene BRCA1 is required for subnuclear assembly of Rad51 and survival following treatment with the DNA cross-linking agent cisplatin, *J. Biol. Chem.* 275 (2000) 23899–23903, <https://doi.org/10.1074/jbc.C000276200>.
- [38] W. Park, J. Chen, J.F. Chou, A.M. Varghese, K.H. Yu, W. Wong, M. Capanu, V. Balachandran, C.A. McIntyre, I. El Dika, et al., Genomic methods identify homologous recombination deficiency in pancreas adenocarcinoma and optimize treatment selection, *Clin. Cancer Res.* 26 (2020) 3239–3247, <https://doi.org/10.1158/1078-0432.CCR-20-0418>.
- [39] T. Taniguchi, M. Tischkowitz, N. Ameziane, S.V. Hodgson, C.G. Mathew, H. Joenje, S.C. Mok, A.D. D'Andrea, Disruption of the Fanconi anemia-BRCA pathway in cisplatin-sensitive ovarian tumors, *Nat. Med.* 9 (2003) 568–574, <https://doi.org/10.1038/nm852>.
- [40] S.S. Yuan, S.Y. Lee, G. Chen, M. Song, G.E. Tomlinson, E.Y. Lee, BRCA2 is required for ionizing radiation-induced assembly of Rad51 complex in vivo, *Cancer Res.* 59 (1999) 3547–3551.
- [41] N. Turner, A. Tutt, A. Ashworth, Hallmarks of 'BRCAness' in sporadic cancers, *Nat. Rev. Cancer* 4 (2004) 814–819, <https://doi.org/10.1038/nrc1457>.
- [42] A.O. Giacomelli, X. Yang, R.E. Lintner, J.M. McFarland, M. Duby, J. Kim, T.P. Howard, D.Y. Takeda, S.H. Ly, E. Kim, et al., Mutational processes shape the landscape of TP53 mutations in human cancer, *Nat. Genet.* 50 (2018) 1381–1387, <https://doi.org/10.1038/s41588-018-0204-y>.
- [43] Q. Zhan, C. Wen, Y. Zhao, L. Fang, Y. Jin, Z. Zhang, S. Zou, F. Li, Y. Yang, L. Wu, et al., Identification of copy number variation-driven molecular subtypes informative for prognosis and treatment in pancreatic adenocarcinoma of a Chinese cohort, *EBioMedicine* 74 (2021) 103716, <https://doi.org/10.1016/j.ebiom.2021.103716>.
- [44] P. Polak, J. Kim, L.Z. Braunstein, R. Karlic, N.J. Haradhavala, G. Tiao, D. Rosebrock, D. Livitz, K. Kubler, K.W. Mouw, et al., A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer, *Nat. Genet.* 49 (2017) 1476–1486, <https://doi.org/10.1038/ng.3934>.
- [45] D.C. Gulhan, J.J. Lee, G.E.M. Melloni, I. Cortes-Ciriano, P.J. Park, Detecting the mutational signature of homologous recombination deficiency in clinical samples, *Nat. Genet.* 51 (2019) 912–919, <https://doi.org/10.1038/s41588-019-0390-2>.
- [46] R.P. Graf, R. Eskander, L. Brueggeman, D.G. Stupack, Association of copy number variation signature and survival in patients with serous ovarian cancer, *JAMA Netw. Open* 4 (2021) e2114162, <https://doi.org/10.1001/jamanetworkopen.2021.14162>.
- [47] V. Lee, A. Murphy, D.T. Le, L.A. Diaz Jr., Mismatch repair deficiency and response to immune checkpoint blockade, *Oncol.* 21 (2016) 1200–1211, <https://doi.org/10.1634/theoncologist.2016-0046>.
- [48] X. Yang, Y. Hu, K. Yang, D. Wang, J. Lin, J. Long, F. Xie, J. Mao, J. Bian, M. Guan, et al., Cell-free DNA copy number variations predict efficacy of immune checkpoint inhibitor-based therapy in hepatobiliary cancers, *J Immunother Cancer* 9 (2021), <https://doi.org/10.1136/jitc-2020-001942>.
- [49] X. Zhang, Y. Wang, J. Xiang, P. Zhao, Y. Xun, S. Zhang, N. Xu, Association between plasma somatic copy number variations and response to immunotherapy in patients with programmed death-ligand 1-negative non-small cell lung cancer, *J. Int. Med. Res.* 50 (2022) 3000605221093222, <https://doi.org/10.1177/03000605221093222>.
- [50] L.R. Ferguson, H. Chen, A.R. Collins, M. Connell, G. Damia, S. Dasgupta, M. Malhotra, A.K. Meeker, A. Amedei, A. Amin, et al., Genomic instability in human cancer: molecular insights and opportunities for therapeutic attack and prevention through diet and nutrition, *Semin. Cancer Biol.* 35 (Suppl) (2015) S5–S24, <https://doi.org/10.1016/j.semcancer.2015.03.005>.
- [51] P.H.G. Duijf, D. Nanayakkara, K. Nones, S. Srihari, M. Kalimutho, K.K. Khanna, Mechanisms of genomic instability in breast cancer, *Trends Mol. Med.* 25 (2019) 595–611, <https://doi.org/10.1016/j.molmed.2019.04.004>.