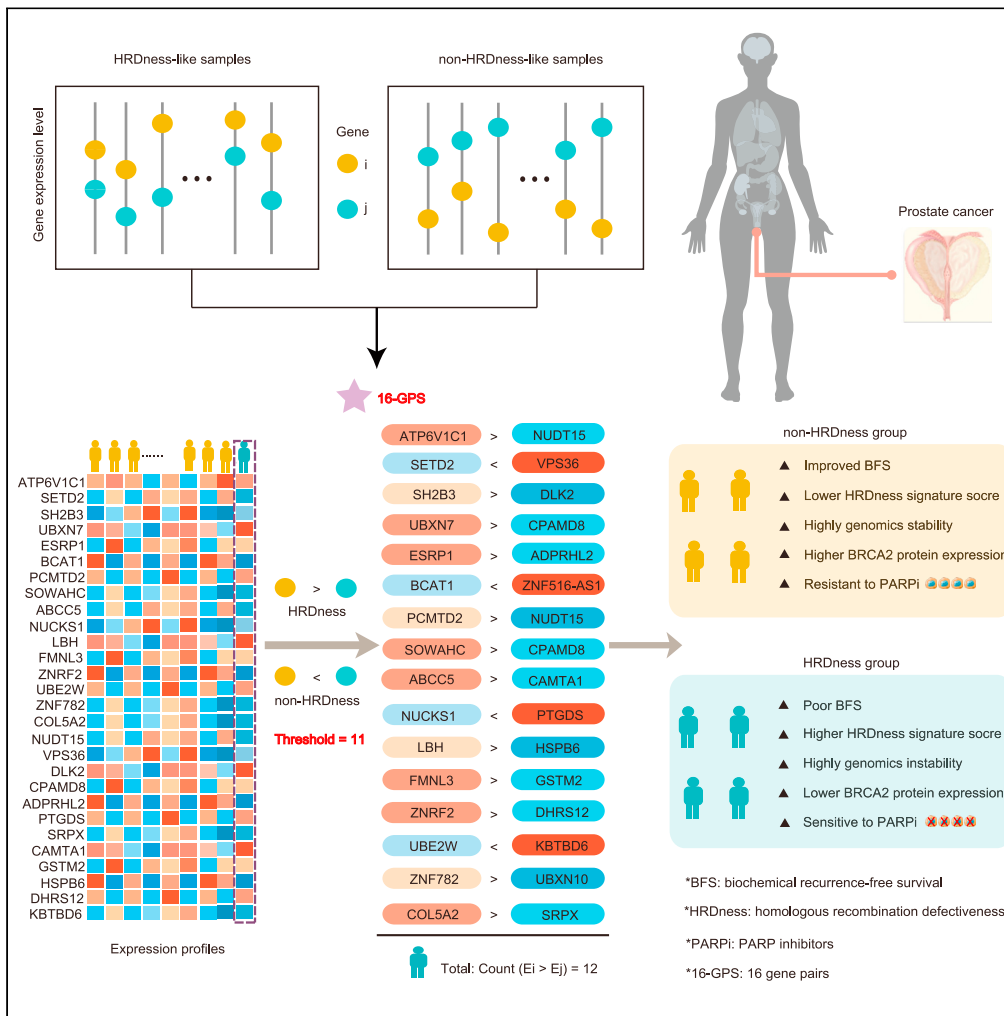


Article

Discovering a qualitative transcriptional signature of homologous recombination defectiveness for prostate cancer



Yawei Li,
Zhangxiang Zhao,
Liqiang Ai, ..., Min
Zou, Yunyan Gu,
Xia Li

lixia@hrbmu.edu.cn (X.L.)
guyunyan@ems.hrbmu.edu.cn
(Y.G.)

Highlights

16 gene pairs (16-GPS)
could predict HRDness for
prostate cancer at
individual level

HRDness samples
classified by 16-GPS
showed HRD molecular
and clinical features

HRDness cells classified
by 16-GPS tend to be
sensitive to PARPi



Article

Discovering a qualitative transcriptional signature of homologous recombination defectiveness for prostate cancer

Yawei Li,^{1,3} Zhangxiang Zhao,^{1,3} Liqiang Ai,¹ Yuquan Wang,¹ Kaidong Liu,¹ Bo Chen,¹ Tingting Chen,¹ Shuping Zhuang,¹ Huanhuan Xu,¹ Min Zou,¹ Yunyan Gu,^{1,4,*} and Xia Li^{2,*}

SUMMARY

The discovery of homologous recombination deficiency (HRD) biomarkers in prostate cancer is important for patients who will benefit from poly ADP-ribose polymerase inhibitor (PARPi). Here, we developed a transcriptional homologous recombination defectiveness (HRDness) signature, comprising 16 gene pairs (16-GPS), for prostate cancer by a relative expression ordering (REO)-based discovery procedure. Subsequently, two newly subtypes classified by 16-GPS showed a higher significance level in various clinicopathological and HRD features than subtypes obtained by other methods, such as HRDetect. HRDness subtype also displayed more aggressive features and higher genomics scores than non-HRDness in three independent datasets. HRDness prostate cancer cells were more sensitive to PARPi than non-HRDness. Moreover, the HRDness samples showed distinct multi-omics characteristics related to homologous recombination repair function loss. Overall, the newly proposed qualitative signature can robustly determine the HRD status for prostate cancer at the personalized level, and especially be an auxiliary tool for PARPi treatment strategy.

INTRODUCTION

Prostate cancer is the second most common solid tumor in males (Siegel et al., 2020). Approximately 20–30% of patients suffer from disease recurrence after receiving radical resection, radiation, or androgen deprivation therapy (ADT), leading to a progression of the disease to metastatic castration-resistant prostate cancer (mCRPC), and finally, 90% of patients succumb to the disease without effective treatments (Kodama et al., 2020; Lalonde et al., 2014). Recently, based on the synthetic lethal effect, poly ADP-ribose polymerase inhibitor (PARPi) has been approved by the Food and Drug Administration for the treatment of mCRPC with BRCA1/2 or other homologous recombination repair (HRR) gene mutations (Rescigno et al., 2018). Importantly, patients with HRR gene alterations also carried a higher Gleason score and prostate specific antigen (PSA) level, and a worse prognosis similar to BRCA alteration, indicating a homologous recombination defectiveness (HRDness) phenotype beyond narrow BRCA-related way (Evans et al., 2016; Marshall et al., 2019a; Piliie et al., 2019). Thus, there is an urgent need to identify the homologous recombination deficiency (HRD) phenotype to extend the utilization of PARPi for tumors with BRCA-deficiency as well as HR-deficiency, thereby expanding the demographic of patients who may benefit from PARPi treatment (Weston et al., 2010; Mendes-Pereira et al., 2009).

Approximately 20–25% mCRPC and 5–10% of primary prostate cancer cases exhibited mutations in the HRR genes (Criscuolo et al., 2019; Handy et al., 2018). Owing to fewer hotspot mutations and the requirement of high tumor purity, detection of genomic variations presents a high false-negative rate. Besides, several presently popular BRCAness and HR-deficiency signatures, named HRD score and HRDetect score, have been developed based on the genomic scars and applied in prostate, breast, and ovarian cancers with the same threshold (Sztupinski et al., 2020; Telli et al., 2016; Davies et al., 2017). However, it is irrational to use the same threshold for classifying different tumor types, considering the tissue-specific background for HRD signature (Jonsson et al., 2019). Thus, in view of the aforementioned limitations, our study aims to discover a robust HRDness classification signature for prostate cancer.

¹Department of Systems Biology, College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, 150081, China

²Department of Bioinformatics, College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, 150081, China

³These authors contributed equally

⁴Lead contact

*Correspondence: lixia@hrbmu.edu.cn (X.L.), guyunyan@ems.hrbmu.edu.cn (Y.G.)

<https://doi.org/10.1016/j.isci.2021.103135>



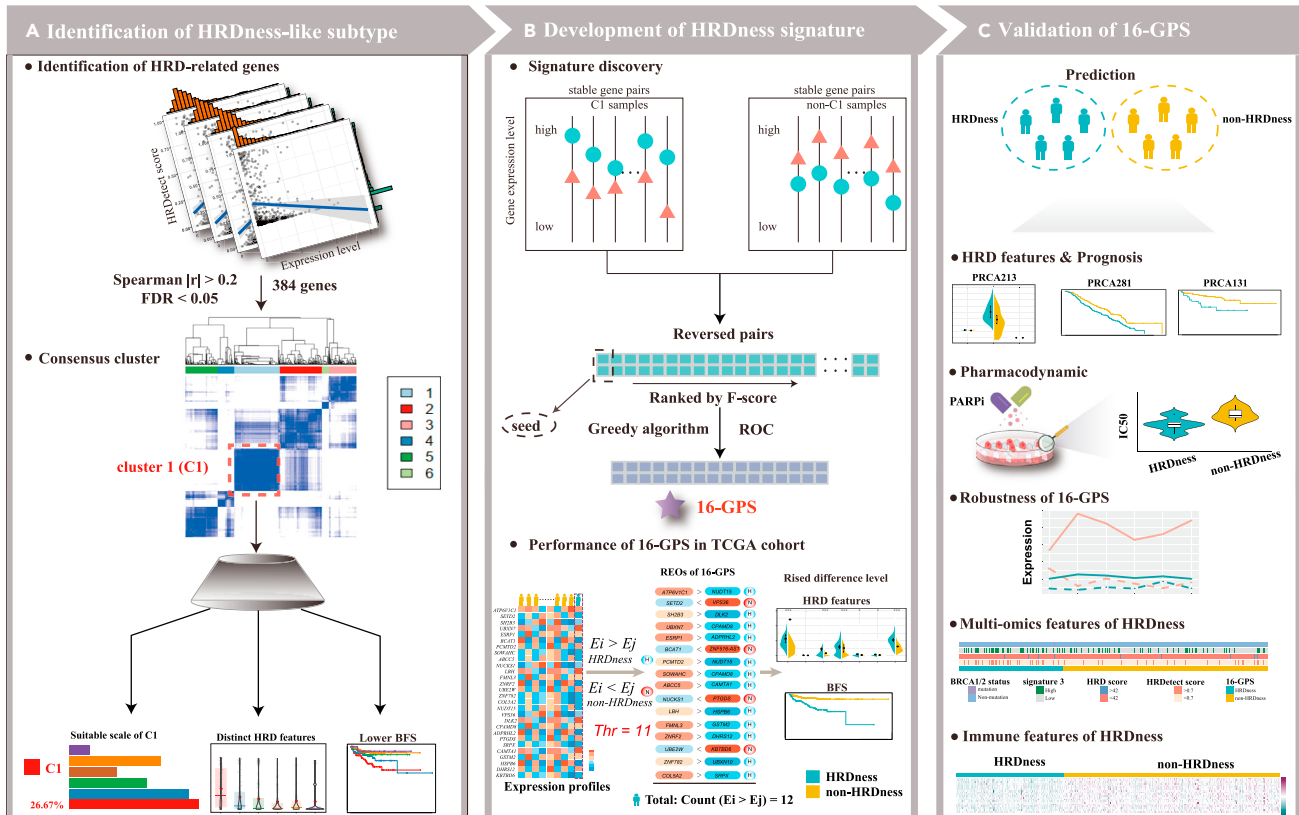


Figure 1. Workflow of this study

(A) Identifying HRDetect score-related genes and then HRDness-like subgroups (C1) was defined by these genes through consensus clustering method. (B) based on gene pairs with stable REO patterns in C1 and significantly reverse in non-C1, a qualitative signature consisting of 16 gene pairs (16-GPS) was developed by performing a forward selection approach. (C) Validation of 16-GPS in independent datasets.

HRD is a genomic event. Changes in the genome can affect gene expression through transcription factor dependent and independent transcriptomic regulation (Ladan et al., 2021; Peng et al., 2014). However, transcriptome signature for HRD in prostate cancer has rarely been investigated. Regular transcriptional signatures based on the quantitative expression measurements of multiple key genes lack robustness because of the marked variation in expression levels of the same gene across samples, and this variation is inevitably induced by different experimental conditions (Guan et al., 2018). Thus, we initiated a relative expression ordering (REO)-based discovery frame to overcome these limitations. The advantage of REOs is that individual patients can be classified by the within-sample REOs of genes without a complex scoring process. The REO-based biomarkers are relatively robust against batch or normalization effects with different technical sources and poor sample preparation or quality, which are likely responsible for most of the false-negative results in detecting gene mutations, and could be applied in individual cancer samples (Qi et al., 2016; Guo et al., 2018).

In the present study, using transcriptomic quantitative measurement of prostate cancer from The Cancer Genome Atlas (TCGA) database, we developed a qualitative REO-based HRDness classification signature composed of 16 gene pairs (16-GPS) for prostate cancer.

RESULTS

Identifying HRDness signature by a REO-based discovering workflow

Briefly, the procedure of identifying HRDness signature is shown in the flow chart (Figures 1A–1C). Firstly, we identified 384 genes whose expression levels were significantly correlated with the HRDetect score in the PRCA495 dataset (spearman's rank correlation, $|r| > 0.2$, FDR < 0.05). Then, consensus clustering

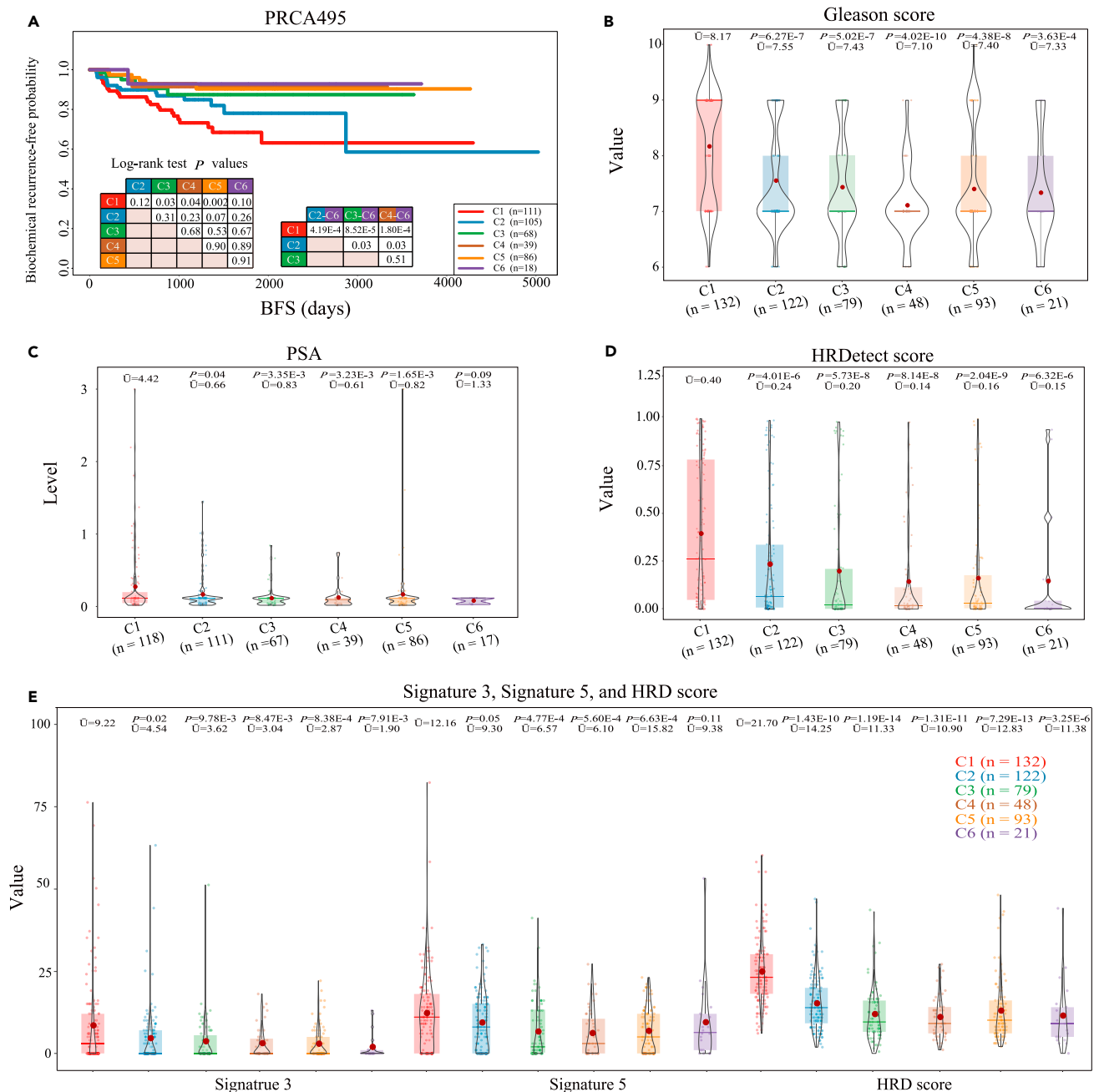


Figure 2. HRD-related clusters of prostate cancer

(A) Biochemical recurrence survival for subjects according to molecular clusters. *p* values were determined by log-rank test.

(B–E) Violin plots illustrate the distribution of pathological malignant features and genomic signature scores, including Gleason score (B), PSA level (C), HRDetect score (D), signature 3, signature 5, and HRD score (E) among the six clusters. Each dot represents an individual sample. Horizontal lines represent the median. *p* values were determined by a one-sided Wilcoxon rank-sum test.

divided the PRCA495 samples into six clusters ($k = 6$, Figure 1A). A total of 111 samples from the cluster 1 (C1) group had a significantly poorer biochemical recurrence-free survival (BFS) than the 316 samples from the non-C1 group ($p = 4.19 \times 10^{-4}$; hazard ratio, HR = 2.47; 95% confidence interval, CI: 1.47–4.14; Figure 2A). In addition, within the samples in clusters C1–C6 (132, 122, 79, 48, 93, and 21 samples, respectively), C1 carried the most aggressive pathological features, including Gleason score (Figure 2B) and PSA level (Figure 2C). More importantly, the C1 cluster maintained the highest HRD-related scores, including

HRDetect score (Figure 2D), signature 3 (Figure 2E), signature 5 (Figure 2E), and HRD score (Figure 2E), among the six clusters. Therefore, based on an appropriate proportion of sample size, i.e., 26.67% (132/495), and distinct HRDness information, the C1 subtype was used as the HRDness-like subtype for training the transcriptional qualitative signature.

The quantitative gene expression values of pairwise genes were transformed to binary values for discovering qualitative HRDness signature (Figure 1B). First, we identified 256,099,836 highly stable gene pairs in 132 samples from C1 (see Method details) and 306,023,521 significantly stable gene pairs in 363 non-C1 samples (binomial distribution test, FDR < 0.01), wherein REOs of 1,100 candidate gene pairs were stable in C1, but reversed in non-C1 samples. Then, for each candidate gene pair, we calculated the F-score to evaluate the prediction power of the gene pair. Next, using the gene pair *ATP6V1C1-NUDT15*, with the max F-score 0.71 as the seed, we finally obtained a set of 16-GPS that reached the highest F-score (0.92) and the largest Area Under Curve (AUC, 95.50%), with the threshold of 11.5, by performing forward selection and AUC of the receiver operating characteristic (AUC-ROC) analysis (Figure 3A). A sample was classified into the HRDness subtype if at least 12 gene pairs voted for HRDness; else, the sample was classified into the non-HRDness subtype.

16-GPS displaying improved performance than HRDetect score and HRD score

Comparative analysis with a one-sided Wilcoxon rank-sum test suggested that HRDness subtype from PRCA495, classified by 16-GPS, showed a significantly higher level of HRDness genomic and clinical features, including signature 3 ($p = 8.22 \times 10^{-6}$, Figure 3B), signature 5 ($p = 2.00 \times 10^{-4}$, Figure 3B), HRD score ($p < 2.20 \times 10^{-16}$, Figure 3B), HRDetect score ($p = 7.92 \times 10^{-14}$, Figure 3C), Gleason score ($p = 6.00 \times 10^{-16}$, Figure 3D), and PSA value ($p = 3.51 \times 10^{-7}$, Figure 3E), than the non-HRDness subtype. Particularly, significance of the difference between HRDness and non-HRDness was increased upon comparing the difference between C1 and non-C1 at the levels of signature 3 ($p = 8.22 \times 10^{-6}$ vs. $p = 1.87 \times 10^{-4}$, Figure 3B), HRDetect score ($p = 7.92 \times 10^{-14}$ vs. $p = 2.61 \times 10^{-13}$, Figure 3C), Gleason score ($p = 6.00 \times 10^{-16}$ vs. $p = 2.88 \times 10^{-13}$, Figure 3D), and PSA ($p = 3.51 \times 10^{-7}$ vs. $p = 3.90 \times 10^{-4}$, Figure 3E). Similarly, the difference in BFS between subtypes classified by 16-GPS ($p = 7.06 \times 10^{-8}$, HR = 3.88, 95% CI: 2.28–6.59, Figure 3F) was greater than that in consensus clustering subtypes ($p = 4.19 \times 10^{-4}$, Figure 2A).

A total of 41 patients in the non-C1 cluster were classified into the HRDness group by 16-GPS, which were termed as HRDness & non-C1 patients. Further, eight patients in C1 were classified into the non-HRDness group, and were termed as non-HRDness & C1. The 35 patients with PSA information in the HRDness & non-C1 group showed greater PSA levels than the eight patients in the non-HRDness & C1 group ($p = 0.07$, Figure S2A). No significant results were observed when comparing with other features (Figures S2B–S2D), which was mainly because of the small size of the non-HRDness & C1 group. Notably, although we obtained no significant results with a huge HR value ($p = 0.12$, HR = 26.52×10^7 , 95% CI: 0-Inf, Figure 3G), seven out of eight patients (one without BFS record) had no disease relapse with intact BRCA. In contrast, when comparing the HRDness & non-C1 group with the non-HRDness subtype with a greater patient pool defined by 16-GPS, we found that the HRDness and non-C1 group showed a higher level of HRDness features, such as PSA level ($p = 2.46 \times 10^{-4}$, Figure S2A), HRDetect score ($p = 8.17 \times 10^{-3}$, Figure S2B), Gleason score ($p = 6.92 \times 10^{-5}$, Figure S2C), signature 3 ($p = 0.03$, Figure S2D), signature 5 ($p = 0.17$, Figure S2D), and HRD score ($p = 4.83 \times 10^{-5}$, Figure S2D), than the non-HRDness group. Moreover, these HRDness & non-C1 patients had poorer BFS than the non-HRDness patients ($p = 4.31 \times 10^{-6}$, HR = 4.87, 95% CI: 2.31–10.29, Figure 3H). Directly, when applying other genomic signatures, HRD score (cutoff = 42) and HRDetect score (cutoff = 0.70), into the dataset PRCA495, the predicted HRD_{pos} and HRD_{neg} had no BFS difference (HRD score, $p = 0.59$, HR = 1.72, 95% CI: 0.24–12.45; HRDetect score, $p = 0.10$, HR = 1.68, 95% CI: 0.90–3.11, Figures 3I and 3J). These results suggest that 16-GPS displays better performance compared with HRD score, HRDetect score, and even HRDetect score-based consensus clustering.

16-GPS showing robust classification in independent datasets

In the independent validation dataset of PRCA213, 12 and 54 patients with HRD feature information were respectively divided into the HRDness and non-HRDness subtypes by 16-GPS. Patients in the HRDness group showed higher levels of HRD score ($p = 0.01$, Figure 4A) and HRDetect score ($p = 0.16$, Figure 4B) than the patients in the non-HRDness group. In the testing dataset PRCA281, a classification rule of “winner

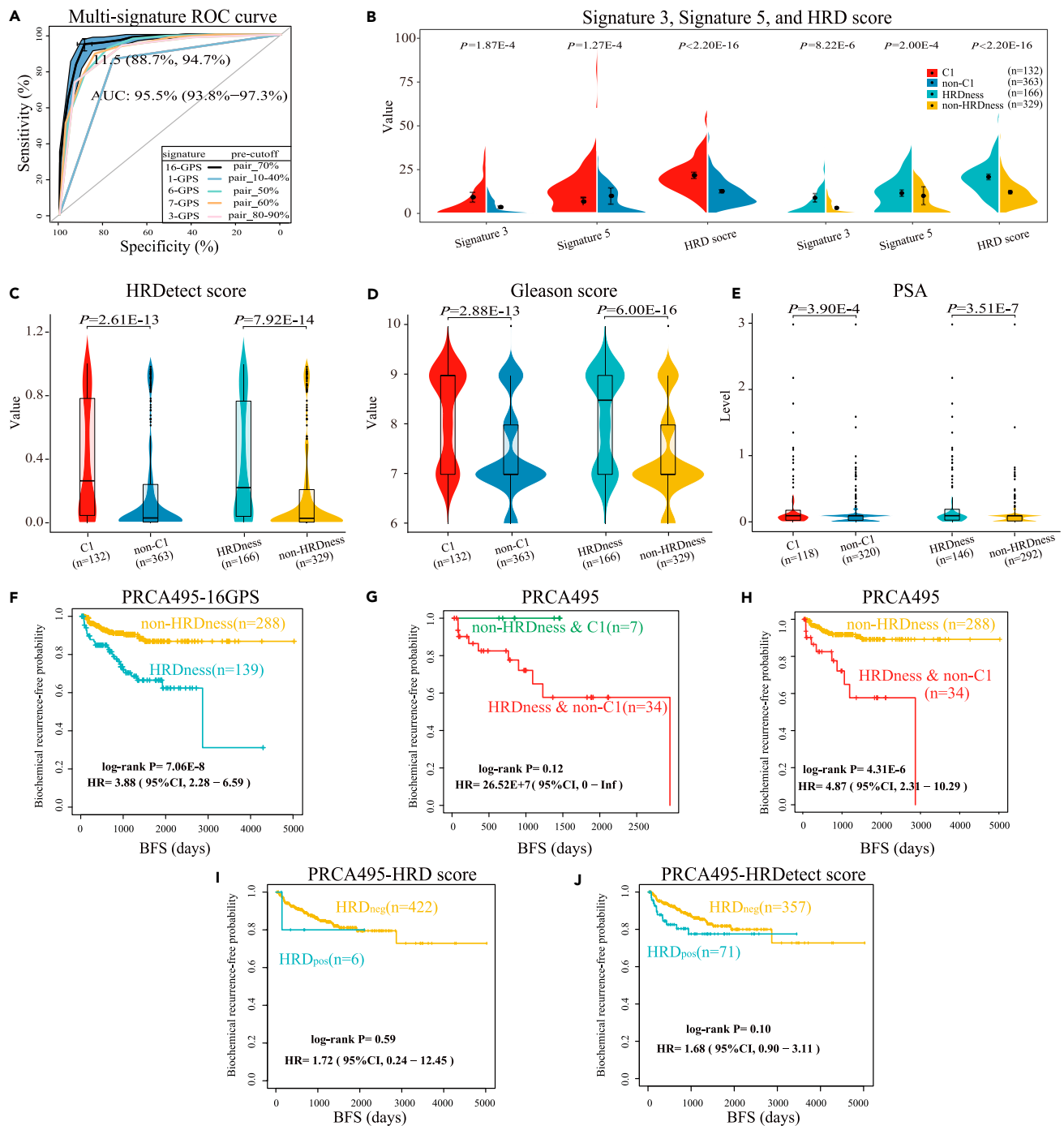


Figure 3. Performance of 16-GPS in the PRCA495 dataset

(A) The area under receiver characteristic operating curves for confirming signature and threshold. AUC was highest at a threshold of 11.5, and the signature was marked as 16-GPS.

(B–E) The distribution of pathological malignant features and genomic signature scores among the C1, non-C1, HRDness, and non-HRDness subgroups in the PRCA495 dataset, including signature 3, signature 5, and HRD score (B), HRDetect score (C), Gleason score (D), and PSA level (E). p values were determined by a one-sided Wilcoxon rank-sum test.

(F) Kaplan–Meier survival analysis of HRDness and non-HRDness patients in the PRCA495 dataset.

(G) Kaplan–Meier survival analysis of HRDness & non-C1 and non-HRDness & C1 patients in the PRCA495 dataset.

(H) Kaplan–Meier survival analysis of HRDness & non-C1 and non-HRDness patients in the PRCA495 dataset.

(I) Kaplan–Meier survival analysis of HRD_{pos} and HRD_{neg} patients classified by HRD score in the PRCA495 dataset.

(J) Kaplan–Meier survival analysis of HRD_{pos} and HRD_{neg} patients classified by HRDetect score in the PRCA495 dataset. p values marked in (F–G) were determined by log-rank test.

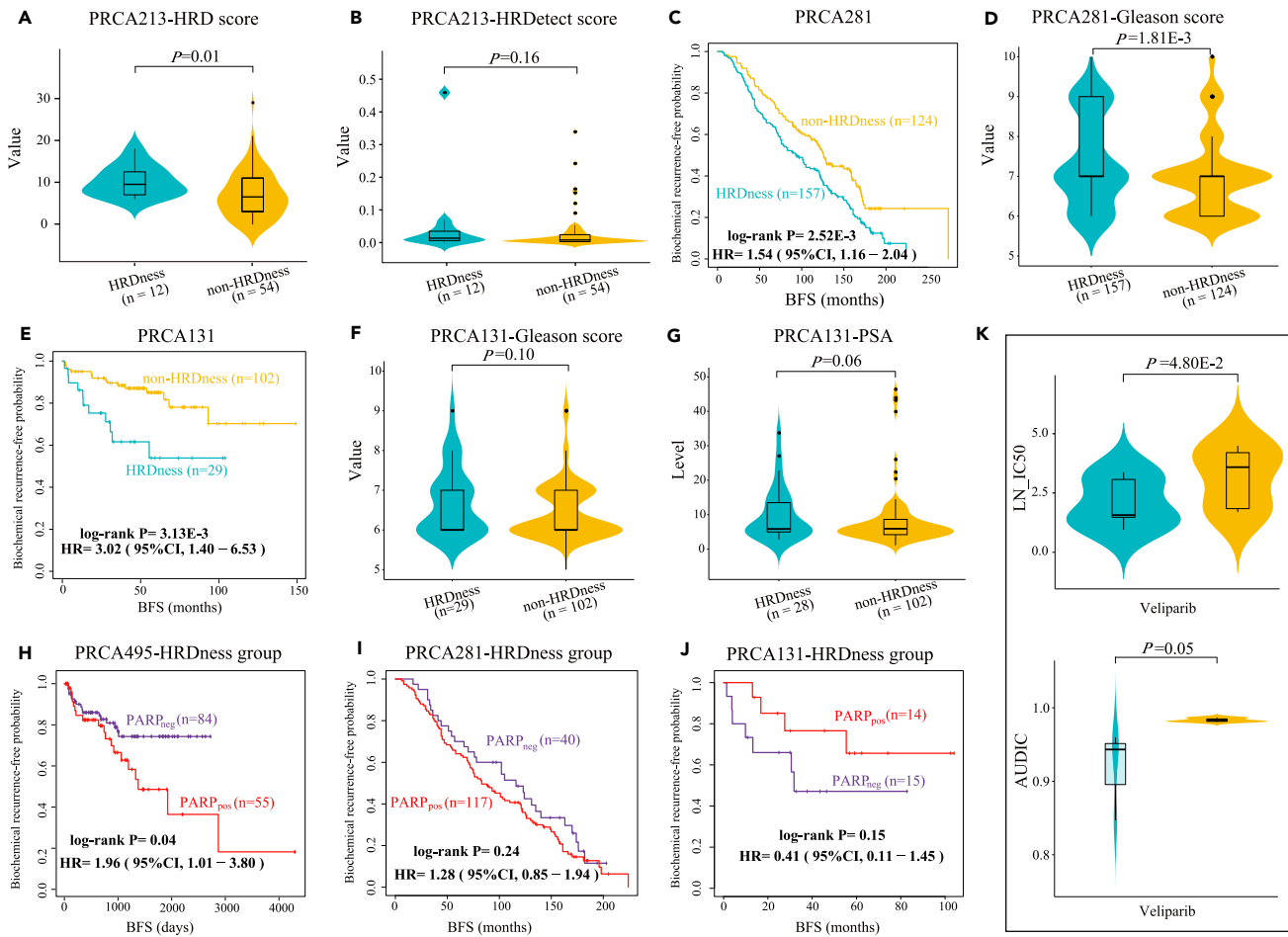


Figure 4. Performance of 16-GPS in validation datasets

(A and B) HRD and HRDetect scores in different groups classified by 16-GPS from the PRCA213 dataset. p values were determined by a one-sided Wilcoxon rank-sum test.

(C) Kaplan–Meier survival analysis of groups classified by 16-GPS in the PRCA281 dataset. p value were determined by log-rank test.

(D) Gleason score in groups classified by 16-GPS in the PRCA281 dataset. p value were determined by a one-sided Wilcoxon rank-sum test.

(E) Kaplan–Meier survival analysis of groups classified by 16-GPS in the PRCA131 dataset. p value were determined by log-rank test.

(F and G) Gleason score and PSA level in groups classified by 16-GPS from the PRCA131 dataset. p values were determined by a one-sided Wilcoxon rank-sum test.

(H–J) Kaplan–Meier survival analysis of PARP_{neg} and PARP_{pos} samples in HRDness group from the PRCA495, PRCA281, and PRCA131 datasets. p values were determined by log-rank test.

(K) LN_IC50 and AUDIC in cells classified by 16-GPS. p values were determined by a one-sided Wilcoxon rank-sum test.

takes all” was applied, as only one gene pair overlapped with 16-GPS. A total of 157 patients in the HRDness group showed poorer BFS compared with that in the non-HRDness counterparts ($p = 2.52 \times 10^{-3}$, $HR = 1.54$, 95% CI: 1.16–2.04, Figure 4C), and the former showed a significantly higher Gleason score (one-sided Wilcoxon rank-sum test, $p = 1.81 \times 10^{-3}$, Figure 4D). Similarly, for the 131 samples from the PRCA131 dataset, 29 patients classified into the HRDness group by 16-GPS had a significantly lower BFS than 102 patients classified into the non-HRDness group ($p = 3.13 \times 10^{-3}$, $HR = 3.02$, 95% CI: 1.40–6.53, Figure 4E), and the HRDness patients displayed higher Gleason scores ($p = 0.10$, Figure 4F) and PSA levels ($p = 0.06$, Figure 4G) through a one-sided Wilcoxon rank-sum test.

On the other hand, patients were classified into the PARP_{neg} and PARP_{pos} (high PARP expression) subgroups by the StepMiner algorithm (see Method details). As expected, for the 139 HRDness patients from the PRCA495 dataset with expression profile, 84 PARP_{neg} tumors showed greater BFS than 55 PARP_{pos} tumors ($p = 0.04$, $HR = 1.96$, 95% CI: 1.01–3.80, Figure 4H). A similar tendency was observed when analyzing

the 157 HRDness patients from PRCA281 ($p = 0.24$, HR = 1.28, 95% CI: 0.85–1.94, Figure 4I). Unfortunately, we obtained an opposite result in PRCA131, where 15 PARP_{neg} patients had a poor BFS than 14 PARP_{pos} patients ($p = 0.15$, HR = 0.41, 95% CI: 0.11–1.45, Figure 4J).

Further, using the gene expression profiles of prostate cancer cell lines treated by five PARPi types from the Genomics of Drug Sensitivity in Cancer (GDSC), three and four cell lines were stratified by 16-GPS into the HRDness and non-HRDness groups, respectively. The HRDness cells tended to be sensitive to veliparib by comparing the natural log of the fitted half maximal inhibitory concentration (LN_IC50, $p = 4.80 \times 10^{-2}$, Figure 4K) and area under the drug inhibition curve (AUDIC, $p = 0.05$, Figure 4K) using one-sided Wilcoxon rank-sum test. HRDness cells also displayed the same tendency of sensitivity for the drugs, including olaparib, rucaparib, and talazoparib, when comparing the LN_IC50 (Figure S3A) and AUDIC (Figure S3B) with non-HRDness cells.

Robustness of 16-GPS against different tumor sampling sites and partial RNA degradation

Here, we proved that REO-based qualitative 16-GPS was not susceptible to different tumor sampling positions or RNA degradation in prostate cancer. 16-GPS was applied to classify six paired pre-ADT and post-ADT samples from the PRCA12-ADT dataset, and all samples were classified into the non-HRDness group. Although the quantitative expression values of 26 genes in 16-GPS varied considerably between paired samples, the REOs in 16-GPS were stable enough to stratify paired patients into the same group (Figure S4). Similarly, the signature resulted in a classification where six paired fresh-frozen/formalin-fixed paraffin-embedded (FF/FFPE) samples from PRCA12-FF were assorted into the non-HRDness group. Relative expression ranks of 16-GPS were highly robust against partial RNA degradation in FFPE subjects, and consistent between pairwise samples (Figure S5).

16-GPS characterizing distinct multi-omics features in HRDness prostate cancer

Multi-omics analysis was conducted in the PRCA495 dataset. A total of 61 genes showed significantly different mutation frequencies between the HRDness and non-HRDness groups after filtering the mutation type of silent mutation (Fisher's exact test, $p < 0.05$, Figure 5A), and the HRDness samples showed significantly higher tumor mutational burden (TMB) than the non-HRDness samples (one-sided Wilcoxon rank-sum test, $p < 2.20 \times 10^{-16}$, Figure 5B). Further, the histogram showed a high frequency of copy number aberrations (CNA) in the HRDness samples (Figure 5C). Specifically, focusing on the genomic region, 613 lesions were identified with significantly different CNA frequencies between these two groups (Fisher's exact test, $p < 0.05$), of which, 612 of 613 genomic lesions had significantly higher frequencies of copy number gain or loss in the HRDness group, which was coincidentally higher than expected (binomial distribution test, $p < 2.20 \times 10^{-16}$).

We identified 2,281, 4,641, and 10,185 differentially expressed genes (DEGs) between HRDness and non-HRDness samples from the PRCA131, PRCA213, and PRCA495 datasets, respectively (Wilcoxon rank-sum test, FDR < 0.01). Only 15 DEGs were identified between the HRDness and non-HRDness groups from the PRCA281 dataset because of the small detected gene scale (Wilcoxon rank-sum test, FDR < 0.01). Concordance analysis for DEGs showed that the consistency of HRDness-related DEGs between both datasets, excluding the PRCA281 dataset, was over or equal to 92.23% (Figure 5D), indicating significantly reproducible and consistent transcriptional differences between these two groups.

By comparing the methylation profiles between HRDness and non-HRDness samples in PRCA495, we detected 2,800 differentially hypermethylated and 2,552 hypomethylated genes (Wilcoxon rank-sum test, FDR < 0.01). As shown in Figure 5D, 104 of the 2,800 hypermethylated genes and 196 of the 2,552 hypomethylated genes overlapped with 831 unique and consistent deregulation direction DEGs, respectively, which were obtained by overlapping the identified DEGs in the three datasets. Further, 98.08% (102/104) hypermethylated genes were concordantly under-expressed in the HRDness group. Similarly, 98.98% (194/196) hypomethylated genes were concordantly overexpressed in the HRDness group, which could not be coincidental (binomial distribution test, $p < 2.20 \times 10^{-16}$).

Further, we also identified 24 differentially expressed proteins by comparing the proteomic data of these two groups (Wilcoxon rank-sum test, FDR < 0.01). These genes were enriched in some HRD- and/or immune-related pathways (hypergeometric distribution model, $p < 0.05$, Table S1). Importantly, two key proteins, BRCA2 and MRE11A, which are essential for HR (Li et al., 2019; D'Andrea and Grompe, 2003), were

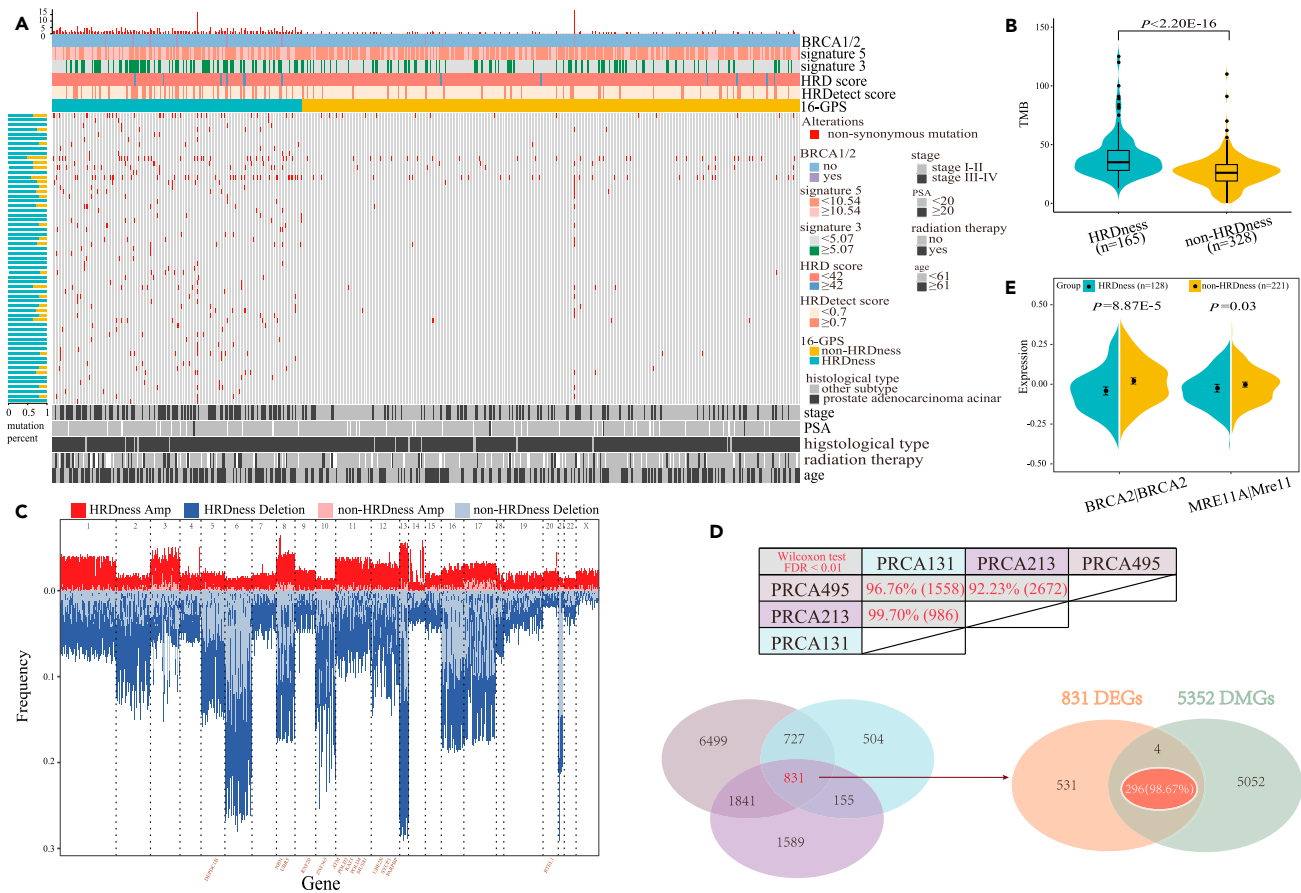


Figure 5. The distinct multi-omics characteristics of the HRDness and non-HRDness samples

(A) OncoPrint showing 61 genes with significantly different mutation frequencies (rows) across subjects (columns) split by 16-GPS. The left shows the mutation percentage, and the top shows the overall number of mutations. Molecular subgroups divided by BRCA mutation, signature 5, signature 3, HRD score, and HRDetect score across subjects are shown at the top. Demographic and clinical annotations are provided at the bottom.

(B) TMB levels between groups classified by 16-GPS. p value were determined by a one-sided Wilcoxon rank-sum test.

(C) Percentage of samples (Y axis) with CNA (amplification or deletion) along the exome, separately shown for the HRDness and non-HRDness subtypes.

(D) Concordance between DEGs detected from the different datasets (upper half), and the Venn diagram displaying the overlap between the consensus DEGs in the three datasets and differentially methylated genes (DMGs) detected from the PRCA495 dataset.

(E) Protein expression levels of two HRR genes between the HRDness and non-HRDness samples. p values were determined by a one-sided Wilcoxon rank-sum test.

significantly under-expressed in HRDness compared with that in the non-HRDness group (one-sided Wilcoxon rank-sum test, *BRCA2*: $p = 8.87 \times 10^{-5}$; *MRE11A*: $p = 0.03$, Figure 5E). Besides, the HRDness group suffered genomic instability and epigenomic alterations, including *ATM* mutation, *POLD4* and *NBN* copy number gain, *DEPDC1B* and *H2AFX* loss, *UBR5* hypomethylation, and *FANCE* hypermethylation, which were enriched in HRR genes.

HRDness prostate cancer presenting elevated immune infiltration

The 296 consensus genes, which represent consistent alterations in the transcriptome and epigenome levels caused by HRDness, not only participated in the HR-related pathways, but also some typically immune-related pathways, such as "Endocytosis," "Autophagy," and "Hippo signaling pathway" (hypergeometric distribution model, $p < 0.05$, Figure 6A). Therefore, considering emerging evidence to suggest that HRD may influence the immune system (Pellegrino et al., 2020), we next investigated the main infiltrating immune cells, including B cells, CD4 T cells, CD8+ T cells, neutrophils, macrophages, and dendritic cells, between the HRDness and non-HRDness groups from the PRCA495 dataset in the Tumor Immune Estimation Resource (TIMER) database. The violin plot showed a significantly increased fraction of 5/6 immune cells in the HRDness group than that in the non-HRDness group (Figure 6B). Surprisingly, two important genes regulating immune cells, *IRF3*

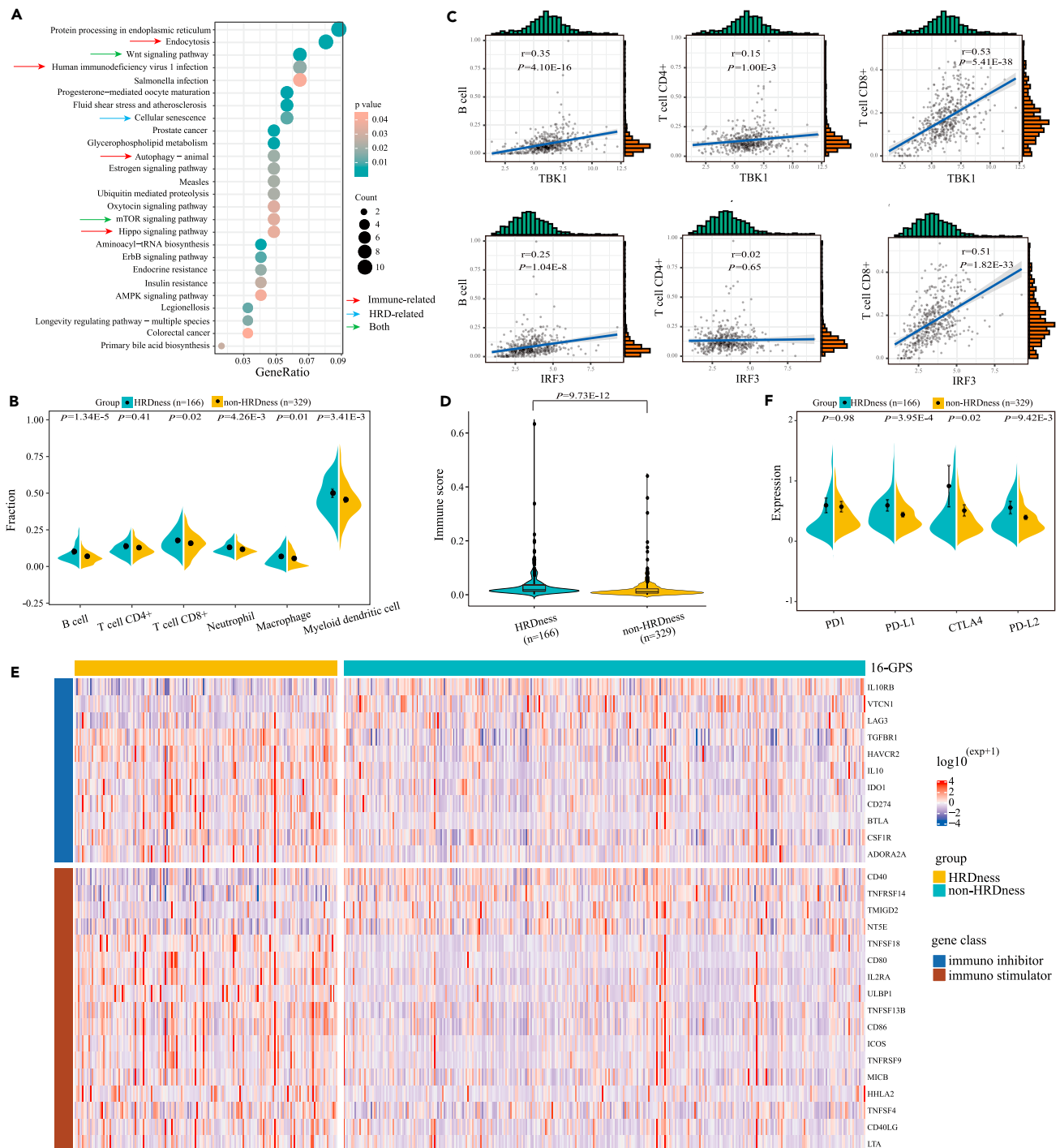


Figure 6. Immune infiltration analysis of HRDness tumors

(A) Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways significantly enriched with the 296 consensus genes were identified. Red, blue, and green arrows represent pathways related to immunity, HRD, and both, respectively (hypergeometric distribution model, $p < 0.05$).

(B) Abundance of six main infiltrating immune cells between the groups from the PRCA495 dataset. p values were determined by a one-sided Wilcoxon rank-sum test.

(C) Relationship between the two key genes belonging to the cGAS-STING pathway and fractions of three types of immune cells.

(D) Distribution of immune scores in groups classified by 16-GPS from the PRCA495 dataset. p value were determined by a one-sided Wilcoxon rank-sum test.

(E) Heatmap displaying the distribution of log-transformed expression levels for immune-stimulators and immune-inhibitors with significant differences between the groups in the PRCA495 dataset.

(F) Expression levels of four immunotherapeutic targets, viz., *PD1*, *PD-L1*, *CTLA4*, and *PD-L2*, between the two groups classified by 16-GPS from the PRCA495 dataset. p values were determined by a one-sided Wilcoxon rank-sum test.

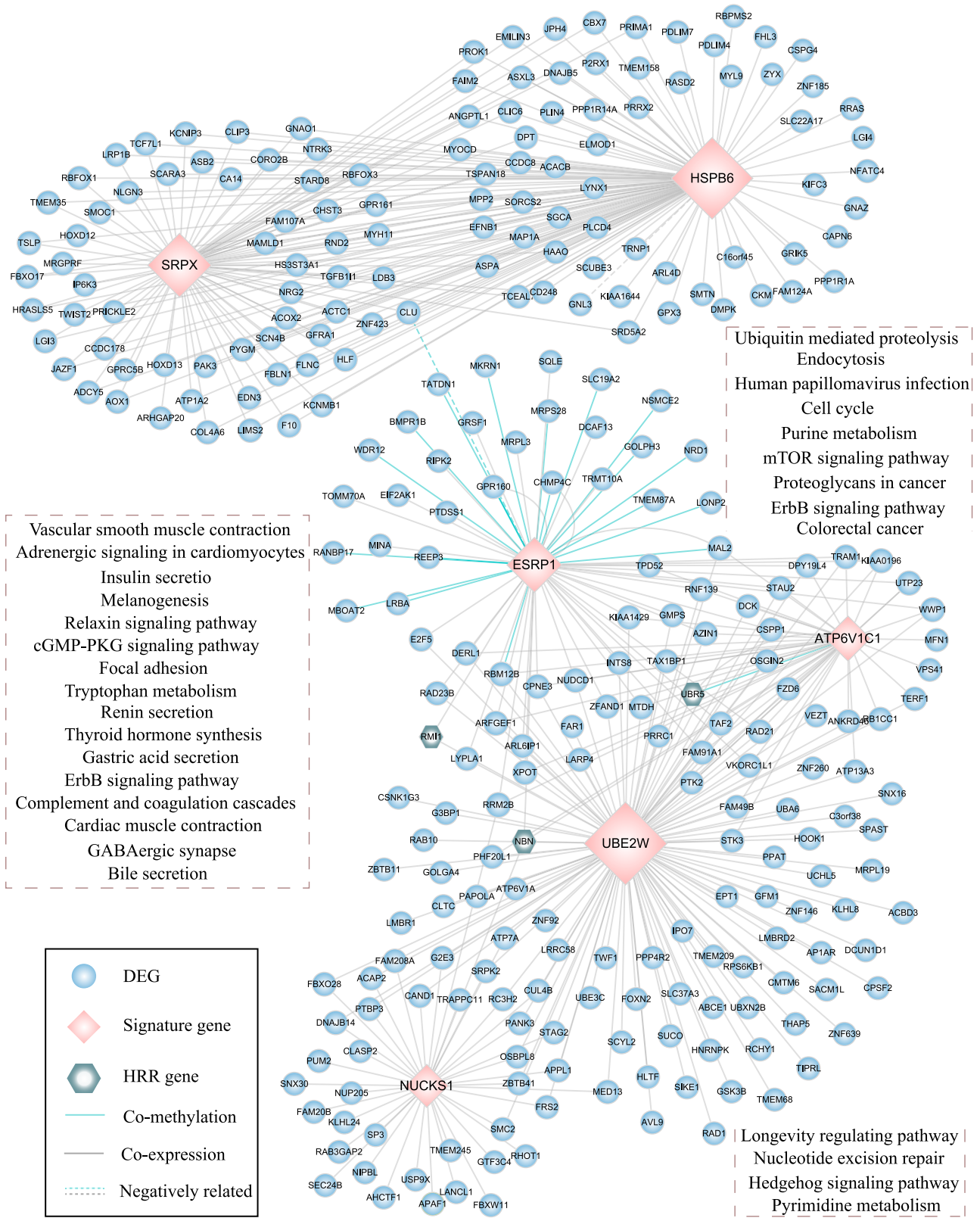


Figure 7. The regulatory network of signature genes for HRD

The nodes (circle and hexagon) represent DEGs of co-expression or co-methylation with six differentially expressed signature genes (rhombus); the solid and dashed lines represent positive and negative relationships, respectively. KEGG pathways significantly enriched with DEGs connected with *SRPX* and *ATP6V1CA* are shown on the left and right sides, respectively. The pathways enriched with the HRD-related DEGs regulated by each of the remaining four signatures are displayed in Table S2.

and *TBK1*, were upregulated in the HRDness group in more than two datasets. Results suggest that the overexpression may contribute to B cell and CD8+ T cell activation via the STING-TBK1-IRF3 mechanism as reported previously (Figure 6C) (Zheng et al., 2020).

Further, the HRDness group exhibited a significantly increased immune score (one-sided Wilcoxon rank-sum test, $p = 9.73 \times 10^{-12}$, Figure 6D), which was likely responsible for the activation of immune-stimulators in the HRDness samples (Figure 6E). In addition, when localizing immunotherapeutic targets, higher *PD-L1*, *PD-L2*, *PD1*, and *CTLA4* expressions were noticed in the HRDness group (Figure 6F). These results demonstrated that HRDness patients may benefit from immune checkpoint inhibitors (ICIs).

Genes in 16-GPS playing a central role in regulating the HRD-related pathways

There were six overlaps with consistent deregulation directions between signature genes and 831 unique DEGs. Co-expression or co-methylation among these six signature genes and DEGs were used to construct the modulation network of HRDness. As shown in Figure 7, the 82 DEGs regulated by *SRPX* were significantly enriched in genetic instability and DNA damage repair (DDR)-related biological pathway, namely “cGmp-pkg signaling pathway,” and an immune-related “ErbB signaling pathway” ($p < 0.05$, hypergeometric distribution model). Similarly, using the hypergeometric distribution model and 121 DEGs linked with *UBE2W* ($p < 0.05$), some HRD- and/or immune-related pathways were also observed, such as “cell cycle” and the “mTOR signaling pathway.” Focusing on HRR genes, we found one key HRR gene, viz., *NBN*, which was simultaneously connected to several signature genes, including *ATP6V1C1*, *ESRP1*, *UBE2W*, and *NUCKS1*, and was upregulated in the HRDness group, consistent with the results of previous research (Min et al., 2020); the result may be due to the significantly higher frequency of copy number gain of *NBN* in HRDness tumors (Fisher’s exact test, $p < 0.0001$). Likewise, another HRR gene, viz., *UBR5*, was linked with *ATP6V1C1* and *UBE2W*. In addition, *UBR5* was over-expressed in the HRD group, which likely resulted from hypomethylation of this gene. As expected, these results suggest that the signature genes driving some HRD-related transcriptome alterations matched upstream-specific variations.

DISCUSSION

In this study, we derived a transcriptional qualitative signature, termed as 16-GPS, whose within-sample REOs could be used to demonstrate the HRD status of prostate cancer patients. To our knowledge, this was the first study to establish an HRDness transcriptional signature in prostate cancer for expanding patients who may benefit from PARPi treatment. As expected, the proportion of HRDness samples in the largest dataset, PRCA495, was slightly higher than the reported scale of HRR gene mutation (33.54% vs. 25%) (Criscuolo et al., 2019; Handy et al., 2018). In agreement with previous reports, the HRDness samples showed poor BFS, along with a high Gleason score and PSA level. Moreover, when 16-GPS was applied to the cell line data, we found that HRDness cells were more sensitive to PARPi (Evans et al., 2016; Marshall et al., 2019a). Further, analysis of genomic data demonstrated that HRDness prostate cancers classified by 16-GPS displayed higher frequencies of mutation and CNA than the non-HRDness samples. HRD-associated alterations at the epigenomic level were consistent with transcriptional differences obtained between the HRDness and non-HRDness groups, suggesting that DNA methylation alterations of the CpG loci in these DEGs play key roles in promoting HRD. Further functional enrichment analysis showed that these consistent DEGs were enriched with some HRD- and/or immune-related pathways. Subsequently, HRDness samples were found with increased immune infiltration and expression of the immune checkpoint genes, suggesting that the HRDness tumors may respond to ICIs. Finally, in addition to some pathways related to HRD and/or immune system, we found that several HRR genes, such as *NBN*, *UBR5*, and *RMI1*, were regulated by genes involved in the 16-GPS, which were also regulated by genomic and epigenomic alterations.

Presently, pre-clinical trials showed various PARPi response rates with different HRR gene mutations. Moreover, some cases without such mutations also responded to PARPi in prostate cancer (Marshall et al., 2019b; Mateo et al., 2015). This suggests that these genomic testing panels are not broad enough

in illustrating the molecular characteristics of HRD tumors and/or to catch the mechanisms of PARPi vulnerability outside of the HRR gene mutations. There are two assumptions: (i) these tumors indeed exhibited the intact HRR gene, but shared a HRDness phenotype; (ii) current genomic testing methods have strict requirements for tumor materials, leading to false-negative results. Therefore, a more accurate biomarker for determining tumor HRD status is needed. Taking studies on breast and ovarian cancers as a reference, Zsofia *et al.* established the same HRD signatures in prostate cancer (Sztupinski *et al.*, 2020; Telli *et al.*, 2016). However, setting a reliable cutoff for these HRD signatures induced by tissue-specific backgrounds is challenging (Jonsson *et al.*, 2019). Even signature 3, the basis of HRDetect discovery, lacks a discrete threshold to determine the BRCA states of tumors (Alexandrov *et al.*, 2013). Besides, a transcriptomics-based study created a DDR pathway signature, which mainly focused on predicting the prognosis of high-risk prostate cancer, rather than defining the DDR status (Evans *et al.*, 2016).

Sokolova *et al.* shared a unique insight that it may be more reliable to detect the downstream consequences of HR function loss for accurately predicting the response to PARPi (Sokolova *et al.*, 2020). Therefore, considering the important role of the transcriptome as a “bridge” from genome to proteome and current high-throughput assays for gene expression detection, we developed a transcriptional qualitative signature based on the within-sample REOs rather than traditional quantitative biomarkers. Here, we chose HRDetect score rather than the HRD score to identify HRD-associated genes, because the former genomic signature showed a more accurate characterization of HRD than the latter in the prostate cancer whole-exome sequencing data from TCGA (Sztupinski *et al.*, 2020). Strong genomic and proteomic evidence was obtained, which indicated that the HRDness tumors classified by 16-GPS were associated with enhancing genomic instability and decreasing HRR protein expression. More importantly, the REO-based signature probably meet great application in clinical practice because: (i) the pairwise relative transcript abundance measurements of genes could be acquired using a low-throughput detection technique, such as real-time PCR; (ii) these qualitative signatures are relatively robust against experimental and technical variations; and (iii) the qualitative signature can be applied at an individual level, without pre-collecting a set of samples for setting a reliable cutoff or for data standardization (Guan *et al.*, 2018; Bhuva *et al.*, 2020).

In summary, we developed 16-GPS based on REOs, which could be applied to define the HRD status of prostate cancer at an individual level. The HRDness samples classified by the 16-GPS showed distinct HRD characteristics at the multi-omics level and in clinical factors. Therefore, it is worthwhile to further verify the clinical applications of 16-GPS, which may reduce unnecessary medical expenses as well as cytotoxicity of some patients who represent PARPi-therapy-irrelevant low risk of death because of the presence of HRD phenotype.

Limitations of study

There are a few limitations in this study. We were unable to collect prostate cancer samples with PARPi treatment to validate the inference that HRDness patients determined by 16-GPS could be benefited from PARPi. However, based on the expression of PARP genes, the prostate cancer samples were divided into PARP_{neg} and PARP_{pos} groups, and the PARP_{neg} prostate cancer group was suggested to have a loss of PARP function. As expected, PARP_{neg} patients with HRDness showing higher BFS were recognized rather than the PARP_{pos} patients in the PRCA495 and PRCA281 datasets. However, a disagreement result was found in PRCA131 dataset, which may be caused by small sample size and low expression of PARP genes may be compensated by other genes or pathways in the HR deficient background and further promote cancer cell development (Ghosh *et al.*, 2016; Ceccaldi *et al.*, 2015). Besides, because of the absence of data with prostate cancer or cell lines treated by ICIs, we were only able to obtain some indirect results for the sensitivity of HRDness to ICIs, including the enlarged immune infiltration, high TMB, and expression of immunotherapeutic targets in HRDness samples. It is necessary to collect ICIs treatment data to explore the relationship between HRDness and ICIs in the future.

STAR★ METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability

- Data and code availability
- **METHOD DETAILS**
 - Data processing
 - Mutational signature, HRD, and HRDetect score
 - Unsupervised clustering
 - Development of the REO-based signature
 - Defining PARP_{neg} and PARP_{pos} tumors
 - Pathway enrichment analysis
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - Analysis of the multi-omics data
 - Analysis of immunity features
 - Network analysis
 - Survival analysis
 - Statistical analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2021.103135>.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China [grant numbers: 61873075, 61673143], the National Key R&D Program of China [grant number: 2018YFC2000100], the Postdoctoral Scientific Research Developmental Fund [grant number: LBH-Q16166], the Fundamental Research Funds for the Heilongjiang Provincial Universities [grant number: 31041180010] and the Outstanding Youth Foundation of Heilongjiang Province of China [grant number: YQ2021H005].

AUTHOR CONTRIBUTIONS

XL., YG., and YL. conceived the idea. XL., YG., and YL. conceived and designed the experiments and wrote the manuscript. YL., ZZ., and LA. designed the experiments. YL., ZZ., LA., YW., KL., TC. and BC. performed the experiments and analyzed the data. TC., SZ., HX., and MZ. helped in interpreting the results and writing. All authors approved the final version.

DECLARATION OF INTERESTS

The authors have declared no conflicts of interest. One or more of the authors of this paper received support from a program designed to increase minority representation in science.

Received: June 3, 2021

Revised: July 28, 2021

Accepted: September 12, 2021

Published: October 22, 2021

REFERENCES

- Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Borresen-Dale, A.L., et al. (2013). Signatures of mutational processes in human cancer. *Nature* 500, 415–421. <https://doi.org/10.1038/nature12477>.
- Bhuvan, D.D., Cursons, J., and Davis, M.J. (2020). Stable gene expression for normalisation and single-sample scoring. *Nucleic Acids Res.* 48, e113. <https://doi.org/10.1093/nar/gkaa802>.
- Ceccaldi, R., Liu, J.C., Amunugama, R., Hajdu, I., Primack, B., Petalcorin, M.I., O'Connor, K.W., Konstantinopoulos, P.A., elledge, S.J., Boulton, S.J., et al. (2015). Homologous-recombination-deficient tumours are dependent on Poltheta-mediated repair. *Nature* 518, 258–262. <https://doi.org/10.1038/nature14184>.
- Criscuolo, D., Morra, F., Giannella, R., Cerrato, A., and Celetti, A. (2019). Identification of novel biomarkers of homologous recombination defect in DNA repair to predict sensitivity of prostate cancer cells to PARP-inhibitors. *Int. J. Mol. Sci.* 20, 3100. <https://doi.org/10.3390/ijms20123100>.
- D'Andrea, A.D., and Grompe, M. (2003). The Fanconi anaemia/BRCA pathway. *Nat. Rev. Cancer* 3, 23–34. <https://doi.org/10.1038/nrc970>.
- Davies, H., Glodzik, D., Morganello, S., Yates, L.R., Staaf, J., Zou, X., Ramakrishna, M., Martin, S., Boyault, S., Sieuwerts, A.M., et al. (2017). HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat. Med.* 23, 517–525. <https://doi.org/10.1038/nm.4292>.
- Evans, J.R., Zhao, S.G., Chang, S.L., Tomlins, S.A., Erho, N., Sboner, A., Schiewer, M.J., Spratt, D.E., Kothari, V., Klein, E.A., et al. (2016). Patient-level DNA damage and repair pathway profiles and prognosis after prostatectomy for high-risk prostate cancer. *JAMA Oncol.* 2, 471–480. <https://doi.org/10.1001/jamaoncol.2015.4955>.
- Ghosh, R., Roy, S., Kamyab, J., Danzter, F., and Franco, S. (2016). Common and unique genetic interactions of the poly(ADP-ribose) polymerases PARP1 and PARP2 with DNA double-strand break repair pathways. *DNA Repair (Amst)* 45, 56–62. <https://doi.org/10.1016/j.dnarep.2016.06.001>.
- Guan, Q., Yan, H., Chen, Y., Zheng, B., Cai, H., He, J., Song, K., Guo, Y., Ao, L., Liu, H., et al. (2018). Quantitative or qualitative transcriptional diagnostic signatures? A case study for colorectal

- cancer. *BMC Genom.* 19, 99. <https://doi.org/10.1186/s12864-018-4446-y>.
- Guo, Y., Jiang, W., Ao, L., Song, K., Chen, H., Guan, Q., Gao, Q., Cheng, J., Liu, H., Wang, X., et al. (2018). A qualitative signature for predicting pathological response to neoadjuvant chemoradiation in locally advanced rectal cancers. *Radiother. Oncol.* 129, 149–153. <https://doi.org/10.1016/j.radonc.2018.01.010>.
- Handy, C., Baras, A.S., Lotan, T.L., and Antonarakis, E.S. (2018). Prevalence of homologous recombination deficiency (HRD) mutations in localized prostate cancer according to Gleason grade: implications for neoadjuvant clinical trial design. *J. Clin. Oncol.* 36, 5062. https://doi.org/10.1200/JCO.2018.36.15_suppl.5062.
- Jonsson, P., Bandlamudi, C., Cheng, M.L., Srinivasan, P., Chavan, S.S., Friedman, N.D., Rosen, E.Y., Richards, A.L., Bouvier, N., Selcuklu, S.D., et al. (2019). Tumour lineage shapes BRCA-mediated phenotypes. *Nature* 571, 576–579. <https://doi.org/10.1038/s41586-019-1382-1>.
- Kodama, H., Koie, T., Oikawa, M., Narita, T., Tanaka, T., Noro, D., Iwamura, H., Tobisawa, Y., Yoneyama, T., Hashimoto, Y., et al. (2020). Castration-resistant prostate cancer without metastasis at presentation may achieve cancer-specific survival in patients who underwent prior radical prostatectomy. *Int. Urol. Nephrol.* 52, 671–679. <https://doi.org/10.1007/s11255-019-02339-3>.
- Ladan, M.M., van Gent, D.C., and Jager, A. (2021). Homologous recombination deficiency testing for BRCA-like tumors: the road to clinical validation. *Cancers* 13, 1004. <https://doi.org/10.3390/cancers13051004>.
- Lalonde, E., Ishkanian, A.S., Sykes, J., Fraser, M., Ross-Adams, H., Erho, N., Dunning, M.J., Halim, S., Lamb, A.D., Moon, N.C., et al. (2014). Tumour genomic and microenvironmental heterogeneity for integrated prediction of 5-year biochemical recurrence of prostate cancer: a retrospective cohort study. *Lancet Oncol.* 15, 1521–1532. [https://doi.org/10.1016/S1470-2045\(14\)71021-6](https://doi.org/10.1016/S1470-2045(14)71021-6).
- Li, Y., Shen, Y., Jin, K., Wen, Z., Cao, W., Wu, B., Wen, R., Tian, L., Berry, G.J., Goronzy, J.J., et al. (2019). The DNA repair nuclease MRE11A functions as a mitochondrial protector and prevents T cell pyroptosis and tissue inflammation. *Cell Metab.* 30, 477–492.e6. <https://doi.org/10.1016/j.cmet.2019.06.016>.
- Marshall, C.H., Fu, W., Wang, H., Baras, A.S., Lotan, T.L., and Antonarakis, E.S. (2019a). Prevalence of DNA repair gene mutations in localized prostate cancer according to clinical and pathologic features: association of Gleason score and tumor stage. *Prostate Cancer Prostatic Dis.* 22, 59–65. <https://doi.org/10.1038/s41391-018-0086-1>.
- Marshall, C.H., Sokolova, A.O., McNatty, A.L., Cheng, H.H., Eisenberger, M.A., Bryce, A.H., Schweizer, M.T., and Antonarakis, E.S. (2019b). Differential response to olaparib treatment among men with metastatic castration-resistant prostate cancer harboring BRCA1 or BRCA2 versus ATM mutations. *Eur. Urol.* 76, 452–458. <https://doi.org/10.1016/j.eururo.2019.02.002>.
- Mateo, J., Carreira, S., Sandhu, S., Miranda, S., Mossop, H., Perez-Lopez, R., Nava Rodrigues, D., Robinson, D., Omlin, A., Tunariu, N., et al. (2015). DNA-repair defects and olaparib in metastatic prostate cancer. *N. Engl. J. Med.* 373, 1697–1708. <https://doi.org/10.1056/NEJMoa1506859>.
- Mendes-Pereira, A.M., Martin, S.A., Brough, R., Mccarthy, A., Taylor, J.R., Kim, J.S., Waldman, T., Lord, C.J., and Ashworth, A. (2009). Synthetic lethal targeting of PTEN mutant cells with PARP inhibitors. *EMBO Mol. Med.* 1, 315–322. <https://doi.org/10.1002/emmm.200900041>.
- Min, A., Kim, K., Jeong, K., Choi, S., Kim, S., Suh, K.J., Lee, K.H., Kim, S., and Im, S.A. (2020). Homologous repair deficiency score for identifying breast cancers with defective DNA damage response. *Sci. Rep.* 10, 12506. <https://doi.org/10.1038/s41598-020-68176-y>.
- Pellegrino, B., Musolino, A., Llop-Guevara, A., Serra, V., de Silva, P., Hlavata, Z., Sangiolo, D., Willard-Gallo, K., and Solinas, C. (2020). Homologous recombination repair deficiency and the immune response in breast cancer: a literature review. *Transl. Oncol.* 13, 410–422. <https://doi.org/10.1016/j.tranon.2019.10.010>.
- Peng, G., Chun-Jen Lin, C., Mo, W., Dai, H., Park, Y.Y., Kim, S.M., Peng, Y., Mo, Q., Siwko, S., Hu, R., et al. (2014). Genome-wide transcriptome profiling of homologous recombination DNA repair. *Nat. Commun.* 5, 3361. <https://doi.org/10.1038/ncomms4361>.
- Pilie, P.G., Tang, C., Mills, G.B., and Yap, T.A. (2019). State-of-the-art strategies for targeting the DNA damage response in cancer. *Nat. Rev. Clin. Oncol.* 16, 81–104. <https://doi.org/10.1038/s41571-018-0114-z>.
- Qi, L., Li, Y., Qin, Y., Shi, G., Li, T., Wang, J., Chen, L., Gu, Y., Zhao, W., and Guo, Z. (2016). An individualised signature for predicting response with concordant survival benefit for lung adenocarcinoma patients receiving platinum-based chemotherapy. *Br. J. Cancer* 115, 1513–1519. <https://doi.org/10.1038/bjc.2016.370>.
- Rescigno, P., Chandler, R., and de Bono, J. (2018). Relevance of poly (ADP-ribose) polymerase inhibitors in prostate cancer. *Curr. Opin. Support Palliat. Care* 12, 339–343. <https://doi.org/10.1097/SPC.0000000000000358>.
- Sahoo, D., Dill, D.L., Tibshirani, R., and Plevritis, S.K. (2007). Extracting binary signals from microarray time-course data. *Nucleic Acids Res.* 35, 3705–3712. <https://doi.org/10.1093/nar/gkm284>.
- Siegel, R.L., Miller, K.D., and Jemal, A. (2020). Cancer statistics, 2020. *CA Cancer J. Clin.* 70, 7–30. <https://doi.org/10.3322/caac.21590>.
- Sokolova, A.O., Yu, E.Y., and Cheng, H.H. (2020). Honing in on PARPi response in prostate cancer: from HR pathway to gene-by-gene granularity. *Clin. Cancer Res.* 26, 2439–2440. <https://doi.org/10.1158/1078-0432.CCR-20-0707>.
- Sztupinszki, Z., Diossy, M., Krzystanek, M., Borcsok, J., Pomerantz, M.M., Tisza, V., Spisak, S., Ruzs, O., Csabai, I., Freedman, M.L., et al. (2020). Detection of molecular signatures of homologous recombination deficiency in prostate cancer with or without BRCA1/2 mutations. *Clin. Cancer Res.* 26, 2673–2680. <https://doi.org/10.1158/1078-0432.CCR-19-2135>.
- Telli, M.L., Timms, K.M., Reid, J., Hennessy, B., Mills, G.B., Jensen, K.C., Szallasi, Z., Barry, W.T., Winer, E.P., Tung, N.M., et al. (2016). Homologous recombination deficiency (HRD) score predicts response to platinum-containing neoadjuvant chemotherapy in patients with triple-negative breast cancer. *Clin. Cancer Res.* 22, 3764–3773. <https://doi.org/10.1158/1078-0432.CCR-15-2477>.
- Weston, V.J., Oldreive, C.E., Skowronska, A., Oscier, D.G., Pratt, G., Dyer, M.J., Smith, G., Powell, J.E., Rudzki, Z., Kearns, P., et al. (2010). The PARP inhibitor olaparib induces significant killing of ATM-deficient lymphoid tumor cells in vitro and in vivo. *Blood* 116, 4578–4587. <https://doi.org/10.1182/blood-2010-01-265769>.
- Zheng, J., Mo, J., Zhu, T., Zhuo, W., Yi, Y., Hu, S., Yin, J., Zhang, W., Zhou, H., and Liu, Z. (2020). Comprehensive elaboration of the cGAS-STING signaling axis in cancer development and immunotherapy. *Mol. Cancer* 19, 133. <https://doi.org/10.1186/s12943-020-01250-1>.

STAR★ METHODS

KEY RESOURCES TABLE

REAGENTE or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
TCGA RNA-seq data (PRCA495)	Genomic data commons application programming interface (GDC API)	https://gdc.cancer.gov/developers/gdc-application-programming-interface-api
PRCA213	International Cancer Genome Consortium	PRAD-CA (https://icgc.org/)
PRCA131	Gene Expression Omnibus (GEO)	GSE21032 (http://www.ncbi.nlm.nih.gov/geo/)
PRCA281	GEO	GSE16560 (http://www.ncbi.nlm.nih.gov/geo/)
PRCA12-ADT	GEO	GSE150368 (http://www.ncbi.nlm.nih.gov/geo/)
PRCA12-FF	ArrayExpress	E-MTAB-2523 (https://www.ebi.ac.uk/arrayexpress/)
Genomics of Drug Sensitivity in Cancer	GDSC	https://www.cancerxgene.org/
Cancer Genome Project	CGP	https://www.nature.com/articles/nature11005#Sec18
TCGA mutation data	GDC API	https://gdc.cancer.gov/developers/gdc-application-programming-interface-api
TCGA CNA data	GDC API	https://gdc.cancer.gov/developers/gdc-application-programming-interface-api
TCGA DNA methylation data	cBioPortal	http://www.cbioportal.org/
TCGA proteomics data	Firehose	https://gdac.broadinstitute.org/
TCGA immune-related data	TIMER	http://timer.cistrome.org/
Software and algorithms		
consensus cluster	R-Bioconductor	https://bioconductor.org/packages/release/bioc/html/ConsensusClusterPlus.html
ComplexHeatmap	R-Bioconductor	https://bioconductor.org/packages/release/bioc/html/ComplexHeatmap.html
clusterProfiler	R-Bioconductor	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
Cytoscape	Graphpic Software	https://cytoscape.org/
R (4.0.2)	The R foundation	https://www.r-project.org/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Yunyan Gu (guyunyan@ems.hrbmu.edu.cn).

Materials availability

This study did not generate new unique materials.

Data and code availability

- This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the [Key resources table](#).
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHOD DETAILS

Data processing

A total of 1144 samples from six datasets were used in this study, as described in the [Key resources table](#). The dataset name is defined by the sample size. For survival analysis, we only used datasets containing more than 50 samples to avoid the bias of small sample size. For the training dataset of PRCA495, the mRNA-seq profiles of level 3 Fragments Per Kilobase Million was extracted. For validation datasets PRCA213, PRCA131, PRCA281 and PRCA12-ADT, which were measured by four different platforms, the gene expression profiles were directly used after probe re-annotation without normalization. The probe-set identifiers (IDs) were mapped to Entrez gene IDs using the corresponding platform information. If multiple probe-sets were mapped to the same gene, the expression value of the gene was defined as their average value, and probe sets that did not map to any Entrez gene ID or mapped to multiple Entrez gene IDs were excluded in our study. For the dataset of PRCA12-FF, whose transcriptional data were derived from Illumina HiSeq 2000, the gene expression profiles of processed Reads Per Kilobase per Million mapped reads were used.

TCGA mutation profiles were extracted according to the following inclusion criteria: (i) non-silent somatic mutation. (ii) At least one mutation called by Muse, Mutect2, Somaticsniper, or VarScan2 methods. Moreover, CNA, epigenomics and proteomics data were directly used in this study without any processed.

LN_{IC50} and AUDIC of five PARPi, viz., veliparib, olaparib, rucaparib, niraparib, and talazoparib, were obtained from GDSC database, and IC₅₀ values of veliparib and olaparib were obtained from CGP database. Data on the seven prostate cancer cell lines used in this study was referenced from the Catalogue of Somatic Mutations in Cancer (<https://cancer.sanger.ac.uk/cosmic>).

Mutational signature, HRD, and HRDetect score

The mutational signature information, including signature 3 and signature 5, HRD score, and HRDetect score for 495 samples from the PRCA495 dataset and 66 samples from the PRCA213 dataset was derived from research by [Sztupinski et al. \(2020\)](#).

Unsupervised clustering

To determine the molecular subclass that is most biologically related to HRD in the RNA-Seq-based dataset ($n = 495$), we performed consensus clustering analysis in the range $k = 2-6$ (the pre-assigned number of clusters) based on the expression profiles of genes whose expressions were significantly correlated with the HRDetect score. Spearman's rank correlation was used for evaluating the relationships between gene expression and HRDetect scores. To avoid overfitting, the clustering data was resampled 1,000 times, each cycle consisting of a randomized selection of 80% samples and genes. Then, the resampled data was converted into a similarity matrix to calculate the probability of two samples clustered together, termed as consensus matrix. Although the lowest proportion of ambiguous clustering was observed at $k = 5$ ([Figure S1A](#)), considering the clinical and genomics characteristics that HRDness sample should carry and the appropriate sample size of HRDness, we chose $k = 6$, and C1 was considered the HRDness-like subclass ([Figure S1B](#)). Clinical factors were compared between all cluster pairs using a one-sided Wilcoxon rank-sum test.

Development of the REO-based signature

Transforming the quantitative expression of a single gene into a binary relation of the within-sample REOs of gene pairs, a binary matrix was obtained by pairwise comparisons of all genes, with $E_i > E_j$ or $E_i < E_j$ on a per sample basis, where E_i and E_j represent the expression values of gene i and gene j , respectively. We identified highly stable gene pairs whose REO pattern was preserved at $\geq 85\%$ of C1 samples, and the significant gene pairs in non-C1 samples were detected using the binomial distribution test with false discovery rate (FDR) < 0.01 . Further, gene pairs whose REO patterns were preserved in HRDness-like samples, but reversed in non-HRDness-like samples, were termed as candidate gene pairs. The p -values were adjusted using the Benjamini-Hochberg (BH) procedure. We classified the samples from the training dataset into the HRDness or non-HRDness subtypes for each candidate gene pair according to their within-sample REO pattern. Subsequently, we evaluated the sensitivity and specificity for each gene pair by calculating the ratio of correctly classified HRDness samples to all HRDness-like samples, and the ratio of correctly classified non-HRDness samples to all non-HRDness-like samples. Finally, from these candidate gene pairs, we

performed a forward selection procedure to search for a signature that achieved the largest F-score value for determining the molecular subclasses of samples, which is defined as follows:

$$F\text{-score} = \frac{2(\text{sensitivity} * \text{specificity})}{(\text{sensitivity} + \text{specificity})}$$

All candidate gene pairs were sorted in descending order according to their F-score value. The gene pair with the largest F-score was chosen as the seed, and the gene pairs that follow were added to the set one at a time, until the F-score value did not increase. Taking 10% as the gradient, we adopted 10-90% gene pairs contained in the signatures as the cutoff. A sample was classified into the HRDness group if more than the pre-defined cutoff of the REOs of gene pairs in the set was voted for HRDness; else, the sample was assigned to the non-HRDness group. Finally, a set of gene pairs with the largest AUC and F-score was selected as the final HRDness classification signature by employing AUC-ROC. Particularly, the classification cutoff correspondingly reduced with the number of matched gene pairs within the sample.

Defining PARP_{neg} and PARP_{pos} tumors

Using the StepMiner algorithm (Sahoo et al., 2007), other than the PRCA281 dataset, in which only *PARP1* overlapped with the detected genes, if any two among *PARP1*, *PARP2*, and *PARP3* expression levels from a sample were smaller than the corresponding StepMiner cutoff, then the sample was classified into the PARP_{neg} group.

Pathway enrichment analysis

Functional enrichment analysis of DEGs and differentially expressed proteins was performed using the R package 'clusterProfiler'. Subsequently, bubble maps of the KEGG pathway were drawn.

QUANTIFICATION AND STATISTICAL ANALYSIS

Analysis of the multi-omics data

Fisher's exact test was used to identify genes or genomic regions with significantly different mutations or CNA frequencies between two groups classified by 16-GPS. TMB was calculated based on somatic nonsynonymous mutations. The Wilcoxon rank-sum test was used to detect DEGs or DMGs between two groups. The BH procedure was used to calculate FDR. Further, protein expression was compared between the two groups using a one-sided Wilcoxon rank-sum test.

Analysis of immunity features

The one-sided Wilcoxon rank-sum test was used to compare the immunity features mentioned above and the expression of the immune checkpoint genes between two groups. Spearman rank correlation was used for evaluating the relationships between the expression of two key genes in the STING pathway and the three main infiltrating immune cells.

Network analysis

Spearman's rank correlation, with $|r| > 0.6$ and $FDR < 0.01$, was used for identifying DEGs that were co-expressed and co-methylated with signature genes based on the expression and methylation profiles of PRCA495.

Survival analysis

The Kaplan-Meier curves of BFS were estimated using the Kaplan-Meier method and compared with the log-rank test.

Statistical analysis

All statistical analyses in this study were performed using R software versions 4.0.2.