Molecular Therapy
Nucleic Acids
Original Article

A transcriptional signature detects homologous recombination deficiency in pancreatic cancer at the individual level

Shuping Zhuang,[1,2] Tingting Chen,[1,2] Yawei Li,[1,2] Yuquan Wang,[1] Liqiang Ai,[1] Yiding Geng,[1] Min Zou,[1] Kaidong Liu,[1] Huanhuan Xu,[1] Linzhu Wang,[1] Zhangxiang Zhao,[1] Zhiqiang Chang,[1] and Yunyan Gu[1]

[1]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150086, China

Pancreatic cancer (PC) with homologous recombination deficiency (HRD) has been reported to benefit from poly ADP-ribose polymerase (PARP) inhibitors. However, accurate identification of HRD status for PC patients from the transcriptional level is still a great challenge. Here, based on a relative expression ordering (REO)-based algorithm, we developed an HRD signature including 24 gene pairs (24-GPS) using PC transcriptional profiles from The Cancer Genome Atlas (TCGA). HRD samples classified by 24-GPS showed worse overall survival (p = 4.4E-3 for TCGA; p = 1.2E-3 for International Cancer Genome Consortium-Australia cohort; p = 6.4E-2 for GSE17891; p = 7.5E-2 for GSE57495) and higher HRD scores than non-HRD samples (p = 1.4E-4). HRD samples showed highly unstable genomic characteristics and also displayed HRD-related alterations at the epigenomic and proteomic levels. Moreover, HRD cell lines identified by 24-GPS tended to be sensitive to PARP inhibitors (p = 6.6E-2 for olaparib; p = 2.6E-3 for niraparib). Compared with the non-HRD group, the HRD group presented lower immune scores and CD4/CD8 T cell infiltration proportion. Interestingly, PC tumor cells with co-inhibition of PARP-related genes and *ATR* showed reduced survival ability. In conclusion, 24-GPS can robustly identify PC patients with HRD status at the individualized level.

INTRODUCTION

Pancreatic cancer (PC) is a lethal malignancy with a poor prognosis. Only 20% of PC patients are suitable for surgery, while most patients eventually experience recurrence after surgical resection and others receive limited benefit from chemotherapy and radiotherapy.[1,2] Recently, according to the National Comprehensive Cancer Network (NCCN) Clinical Practice Guidelines in Oncology for Pancreatic Adenocarcinoma (Version2, 2021), the poly ADP-ribose polymerase (PARP) inhibitor, olaparib, based on synthetic lethal effect, has been recommended to treat BRCA1/2-defective metastatic PC.[3]

Loss-of-function in BRCA1/2 results in homologous recombination deficiency (HRD) accompanying by highly genomic instability.[4] About 11% ~ 27% PC patients display HRD characteristics,[5] and Park et al. and McIntyre et al. suggested that HRD PC patients showed favorable prognosis under platinum exposure.[6,7] However, PC with HRD displayed a poor prognosis without platinum chemotherapy,[8] suggesting an urgent need to distinguish the HRD sub-population for better treatment. Some popular HRD signatures, such as HRD score,[9] HRDetect score,[10] and mutational signature 3,[11] have been developed based on genomic scars. However, due to the BRCA-mediated phenotype shaped by different tumor types, the designated threshold of these genome signatures applied in ovarian cancer and breast cancer is inapplicable in PC.[12] Moreover, the accuracy of genetic mutation detection varies with the sequencing depth and the quality of the sample, which may fail to capture some hot-spot mutations and thus lead to a high false-negative prediction. Therefore, it is urgent to develop a reliable HRD signature for PC patients.

At present, high-throughput RNA sequencing (RNA-seq) data provides new insights into the detection of HRD at the transcriptome level. Transcriptome reflects the alterations in other omics, such as genome, epigenome, and proteome. Currently, no transcriptional signature of HRD has been established for PC. In epithelial ovarian cancer, the BRCAness profile consisting of 60 genes was previously developed to correlate with responsiveness to platinum and PARP inhibitors.[13] However, quantitative signatures could not be applied to individual patients. Thus, we have proposed a qualitative method named within-sample relative expression ordering (REO) of genes in previous works.[14] The core procedure of REO is to transform the quantitative information of gene pairwise expression to a binary value of gene pairwise. The REO-based signature displays robustness against experimental batch effects with various uncertain factors, including RNA degradation, technical sources, and data

Received 9 July 2021; accepted 6 October 2021;
https://doi.org/10.1016/j.omtn.2021.10.014.

[2]These authors contributed equally

Correspondence: Yunyan Gu, PhD, College of Bioinformatics Science and Technology, Harbin Medical University, 157 Baojian Road, Nangang District, Harbin, China.
E-mail: guyunyan@ems.hrbmu.edu.cn
Correspondence: Zhiqiang Chang, PhD, College of Bioinformatics Science and Technology, Harbin Medical University, 157 Baojian Road, Nangang District, Harbin, China.
E-mail: changzhiqiang@hrbmu.edu.cn

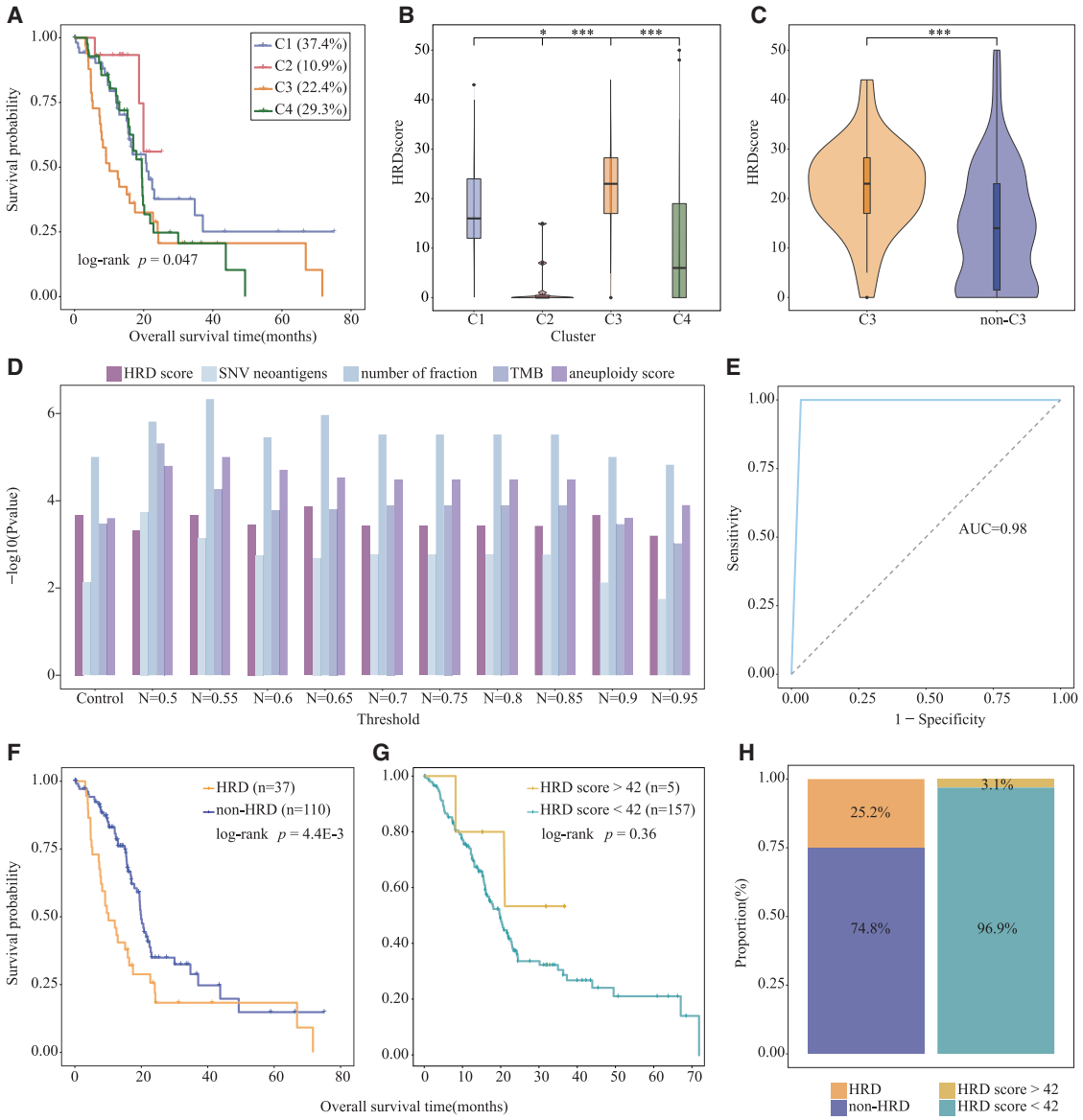

Figure 1. Discovering HRD signature for PC

(A) Kaplan-Meier curve depicts the survival difference among four subgroups. (B) Comparison of the differences of HRD scores among four clusters. (C) Comparing the difference of HRD scores between C3 and non-C3 groups (C1, C2 and C4).

(D) Determination of a threshold for 24-GPS to distinguish HRD and non-HRD samples. The x axis represents different thresholds (one control and 10 experimental groups) and the y axis represents the significance (-log(p value)) of the difference between HRD and non-HRD groups at five indexes. (E) AUC represents the efficiency of 24-GPS under the established threshold. (F) Kaplan-Meier curve depicts the survival difference between HRD and non-HRD groups classified by 24-GPS. (G) Kaplan-Meier curve depicts the survival difference between high HRD scores (HRD score > 42) and low HRD scores (HRD score < 42) groups. (H) The histogram displays the proportion of samples divided by the above two grouping methods. *p < 0.05, **p < 0.01, and ***p < 0.001.

normalization.[15,16] Importantly, the advantage of REO is also demonstrated by the ability to perform individualized analysis.[17]

Herein, based on the REO method, we discovered an HRD signature by performing the qualitative analysis of RNA-seq expression profiles for PC from The Cancer Genome Atlas (TCGA). The performance of the signature was validated in several independent datasets.

RESULTS

Discovering HRD signature for PC

To characterize HRD status for PC, four clusters were obtained by the consensus clustering based on the gene expression profiles of DNA damage response (DDR) genes in the TCGA cohort (n = 147) (Figure S1A). Cluster 3 (C3), including 33 samples, displayed the worst overall survival (OS) and the highest HRD scores (Figures 1A and 1B). Besides, C3

presented a remarkable higher genomic instability than the other clusters at the level of number of fraction, single nucleotide variation (SNV) neoantigens, tumor mutation burden (TMB), and aneuploidy score (Wilcoxon rank-sum test; Figures S1B–S1E). Moreover, compared with non-C3 group (C1, C2, and C4), C3 showed a higher score in above HRD-related features (p = 2.1E-4 for HRD score, p = 9.9E-6 for number of fraction, p = 7.5E-3 for SNV neoantigens, p = 3.4E-4 for TMB, p = 2.5E-4 for aneuploidy score; Wilcoxon rank-sum test; Figures 1C and S1F–S1I). These results indicated that C3 harbored a distinct HRD characteristic and thus was considered an HRD-like group for discovering HRD signature.

Using the REO-based algorithm, 156,399 significantly reversal gene pairs were identified and the 1,000 gene pairs with top frequency difference (FD) values were included for the following analysis (see section, "materials and methods"). We discovered a signature composed of 24 gene pairs (24-GPS) by performing a least absolute shrinkage and selection operator (LASSO) model (see section, "materials and methods; Table S2). The cutoff of 65% in 24-GPS was an optimal threshold due to the highest difference in HRD score and a higher difference in other HRD-related features between HRD-like and non-HRD-like groups (p = 1.4E-4 for HRD score, p = 1.08E-6 for number of fraction, p = 2.1E-3 for SNV neoantigens, p = 1.6E-4 for TMB, p = 2.9E-5 for aneuploidy score; Wilcoxon rank-sum test; Figure 1D). As a result, 147 PC samples were classified into 37 HRD (25.2%) and 110 non-HRD (74.8%) with an area under curve (AUC) of 0.98 (Figure 1E). Then, survival analysis revealed a significantly poorer OS in the HRD group compared with non-HRD samples (p = 4.4E-3; log-rank test; Figure 1F). However, according to the previous HRD score cutoff,[9] there was no difference observed in five high-HRD-score cases (HRD score > 42) compared with 157 low-HRD-score cases (HRD score < 42) in OS (p = 0.36; log-rank test; Figure 1G). Compared with the 25.2% HRD samples identified by 24-GPS, the proportion of high HRD scores (3.1%) was much lower than the HRD frequency (11% $\sim$ 27%), as discovered by Golan et al.[5] (Figure 1H).

Notably, four non-C3 samples classified into HRD group by 24-GPS showed a significantly poorer OS (p = 9.5E-4; log-rank test; Figure S2A) and a considerable higher HRD score than the non-HRD group (p = 0.25; Wilcoxon rank-sum test; Figure S2B), which suggests that 24-GPS displays better performance than the clustering method.

Validating 24-GPS in independent datasets

The performance of 24-GPS was validated in independent datasets. For the International Cancer Genome Consortium-Australia (ICGC-AU) cohort, 17 patients classified into the HRD subgroup by 24-GPS displayed a significantly worse OS than 78 patients in the non-HRD group (p = 1.2E-3; log-rank test; Figure 2A), and the HRD subgroup displayed a higher HRDetect score (p = 0.092; Wilcoxon rank-sum test; Figure 2B). However, under the previous HRDetect score cutoff established in ICGC-AU,[10] there was no significant difference in OS between PC patients with high HRDetect scores (HRDetect score $\geq$ 0.7) and low HRDetect scores (HRDetect score < 0.7) (p = 0.37; log-rank test; Figure 2C). In datasets GSE57495 and GSE17891, HRD cases showed poorer OS than non-HRD cases (P = 7.5E-2 for GSE57495, P = 6.4E-2 for GSE17891; log-rank test; Figures 2D and 2E). Additionally, *RAD51* deficiency is one of the most specific cellular hallmarks of homologous recombination (HR) dysfunction,[18] and a significantly higher proportion of *RAD51* deficient cases was found in the HRD group than in the non-HRD group (22.22% versus 5.41%; p = 2.1E-2; Fisher's exact test; Figure 2F).

Concordance analysis for the top 1,000 differentially expressed genes (DEGs) derived from datasets of TCGA, ICGC-AU, and GSE17891 showed the consistent ratio between two DEG lists was over 98.03% (Figure 2G). In total, 80 overlapped DEGs were detected from the three datasets (Figure 2H). Further, Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis showed 80 DEGs participated in some HRD-related pathways, such as cell cycle, DNA replication, and HR (Figure S3A).

Although lacking PARP inhibitor treatment-related data, low expression of *PARP1/2/3* genes could be regarded as a mimic of treatment with PARP inhibitors. Then, PARP-positive expression samples ($PARP_{pos}$) and PARP-negative expression samples ($PARP_{neg}$) samples were defined (see section, "materials and methods") and no obvious difference was observed in the TCGA cohort (p = 0.83; log-rank test; Figure S2C), but a significantly prolonged OS in $PARP_{neg}$ samples was discovered in the ICGC-AU cohort (p = 2.5E-3; log-rank test; Figure S2D).

Moreover, pharmacogenomic data, accessed from the Cancer Therapeutics Response Portal (CTRP), the Genomics of Drug Sensitivity in Cancer database lab version 2 (GDSC2), and CRISPR/Cas9 data were employed to further test whether HRD cells identified by 24-GPS showed vulnerability to PARP inhibition. For PC cell lines treated with olaparib in CTRP, 30 HRD cell lines were classified by 24-GPS and presented a lower AUC score than in eight non-HRD cell lines (p = 6.6E-2; one-sided Wilcoxon rank-sum test; Figure 2I). Similar results were found in GDSC2, and 15 HRD cell lines were more susceptible to niraparib than 16 non-HRD cell lines (p = 2.6E-3; one-sided Wilcoxon rank-sum test; Figure 2J). The mutation landscape analysis discovered that 24-GPS correctly divided all cell lines with *BRCA1/2* mutation into HRD group in GDSC2 (Figure S2E). In addition, three *BRCA1/2* mutation cell lines were rightly divided into HRD group with only one *BRCA1/2* mutation sample misclassified into non-HRD group in CTRP (Figure S2F). Besides, all cell lines with *RAD51C* mutation fell in the HRD group in both CTRP and GDSC2. Interestingly, *PARP1* gene exclusively mutated in non-HRD cases from CTRP. In the CRISPR/Cas9 dataset, HRD cell lines presented lower survival viability than non-HRD cell lines after knocking out *PARP1/2* gene (p = 6.5E-2 for knocking out *PARP1*; p = 0.17 for knocking out *PARP2*; one-sided Wilcoxon rank-sum test; Figures 2K and 2L).

Multi-omics landscape of PC patients with HRD

Integrated analysis of multi-omics data was performed in the TCGA dataset. As shown in Figure 3A, the heatmap depicted the 10 most

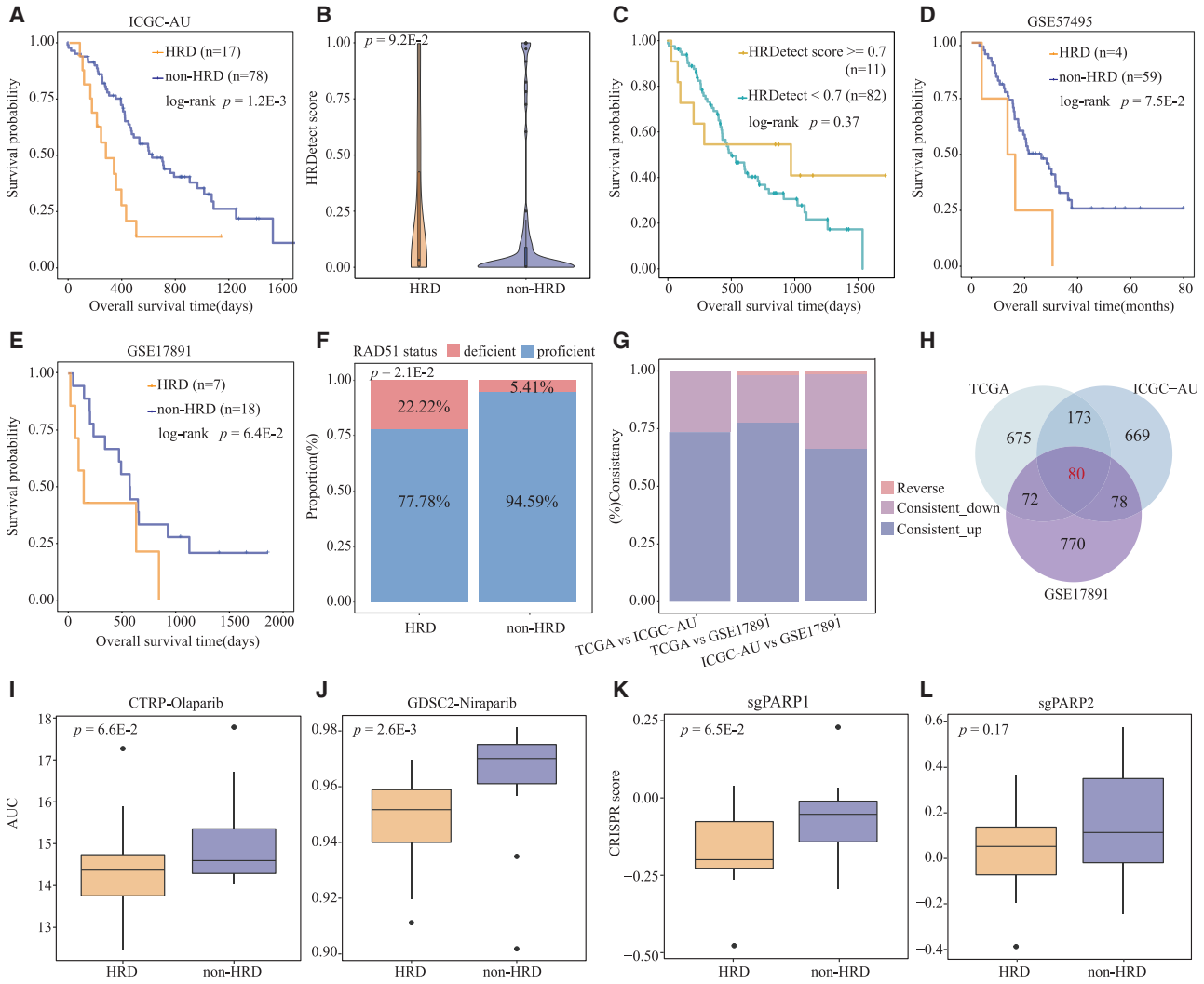

Figure 2. Verification for the performance of 24-GPS

(A) Kaplan-Meier curve depicts the survival difference between HRD and non-HRD groups in the ICGC-AU. (B) Box plot displays the difference of HRDetect score between HRD and non-HRD groups in ICGC-AU cohort. (C) Kaplan-Meier curve depicts the survival difference between high HRDetect scores group and low HRDetect scores group. (D and E) Kaplan-Meier curve depicts the survival difference between HRD and non-HRD groups in dataset GSE57495 (D) and GSE17891 (E), respectively. (F) The histogram displays the proportion of different *RAD51* status (deficient or proficient) in HRD and non-HRD samples. (G) Concordance analysis for DEGs derived from different datasets. The x axis represents the combination of any two data cohorts in TCGA, ICGC-AU, and GSE17891, and the y axis represents the consistent ratio of two datasets. (H) Venn diagram depicts the overlapped quantity of DEGs among the three data cohorts. (I and J) Box plot shows the differences in the response of PC cell lines to olaparib in CTRP (I) and to niraparib in GDSC2 (J). (K and L) Box plot shows the differences in tumor cell survival ability after separately knocking out *PARP1* (K) or *PARP2* (L) genes. SgPARP1/2, single-guide PARP1/2.

mutated genes, DDR genes, and genes with differential mutation frequency. Notably, two typical markers for HRD, *BRCA1* and *BRCA2* genes, were found mutated only in HRD samples. Meanwhile, some other DDR genes were also found exclusively mutated in HRD group, such as *ATRX* and *FANCA*. In addition, 10 genes showed significantly higher mutation frequency in HRD samples compared with non-HRD samples ($p < 0.05$; one-sided Fisher's exact test). Among them, *FAT4* is recurrently mutated in several types of human cancers, including PC, and suppresses tumor growth via activation of Hippo signaling.[19] The expression of *FBN1* is considered to be positively correlated with tumor-related immune infiltration.[20] Together, these results demonstrated HRD PC harbors a specific genetic mutation spectrum.

In addition to the somatic mutation spectrum, we also compared copy number aberration (CNA) profiles between two subgroups. Compared with non-HRD cases, a large number of genes with significantly higher frequencies of CNA in HRD cases were observed (false

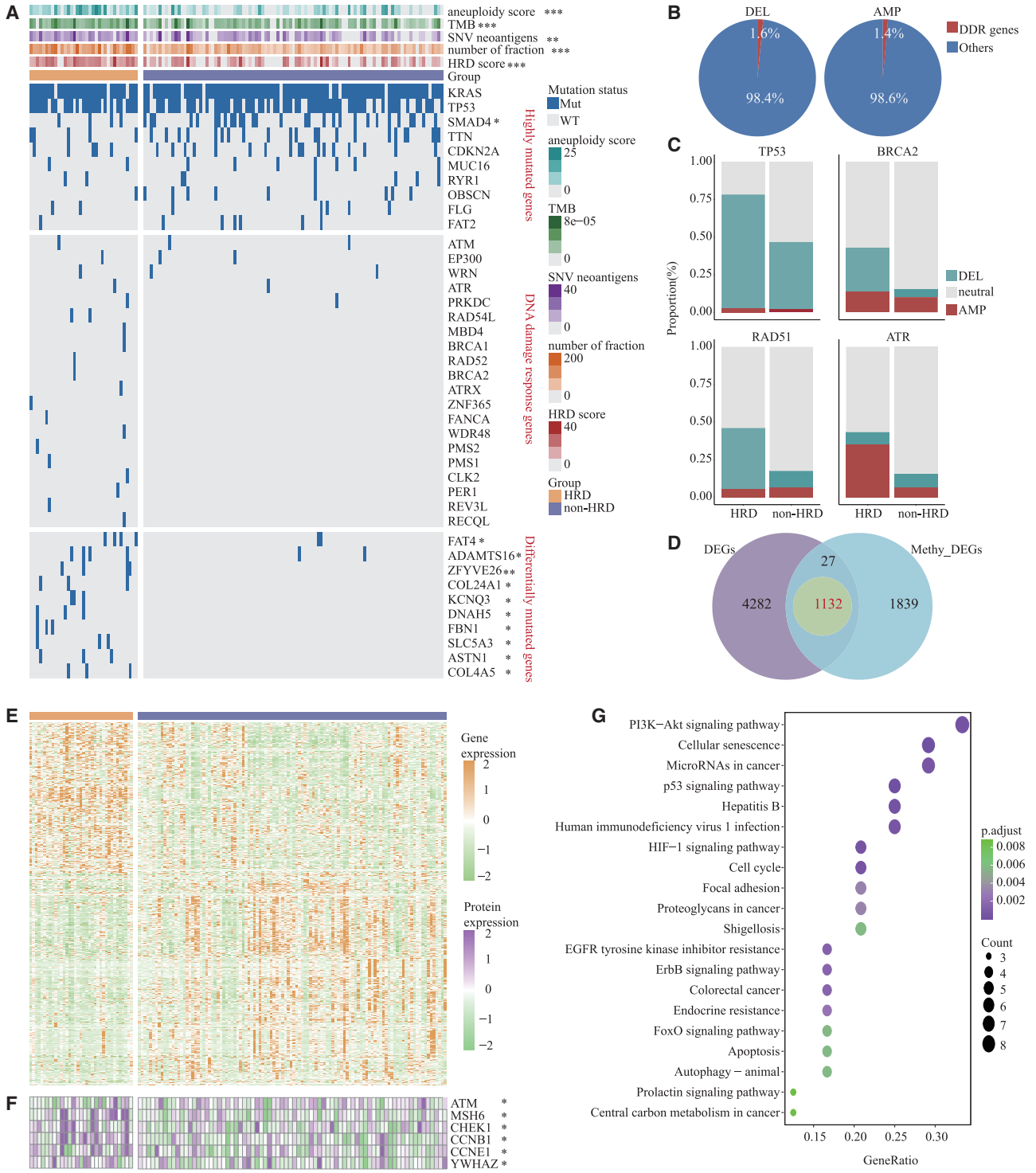

Figure 3. Multi-omics analysis in HRD

(A) Top: differences in the distribution of aneuploidy score, TMB, SNV neoantigens, number of fraction, and HRD score between HRD and non-HRD groups in TCGA. Bottom: the heatmap comprehensively depicts the mutation landscape of highly mutated genes, DDR genes, and differential mutation genes. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$. (B) Pie charts depict the proportion of DEL and AMP of DDR genes in all altered genes. (C) The histograms depict the proportion of different CNA statuses (DEL, neutral, or AMP) of *TP53*, *BRCA2*, *RAD51*, and *ATR* genes in HRD and non-HRD samples. (D) Venn diagram depicts the overlap of transcriptome and methylation differential genes. (E) The heatmap displays the expression of genes with consistent differential expression at the level of transcriptome and epigenome. (F) Distribution of the expression of ATM, MSH6, CHEK1, CCNB1, CCNE1, and YWHAZ proteins in HRD and non-HRD cases. (G) The dot plot exhibits the top 20 significant enriched pathways from KEGG.

discovery rate [FDR] < 0.05; one-sided Fisher's exact test). CNA events of DDR genes accounted for 1.6% deletion (DEL) and 1.4% amplification (AMP), respectively (Figure 3B). Notably, *TP53*, *BRCA2*, and *RAD51* genes had higher frequencies of deletions in the HRD group, while a higher frequency of amplification for *ATR* gene was observed (Figure 3C).

Using the methylation profiles in the TCGA cohort, 2,998 genes showed significant differential methylation between HRD and non-HRD groups (FDR < 0.05; Wilcoxon rank-sum test). As shown in Figure 3D, 1,159 of them were overlapped with previously identified DEGs, and the correlation ratio reached 97% (1,132/1,159; over-expressed and hypomethylated genes, under-expressed and hypermethylated genes), which did not happen by chance (p < 2.2E-16; binomial distribution test). Then, the heatmap depicted the differences between HRD and non-HRD groups at the transcriptome and epigenome levels (Figures 3E and S3B).

Finally, we identified 30 proteins with significant differential expression in HRD samples compared with non-HRD samples (p < 0.05; one-sided Wilcoxon rank-sum test). Compared with non-HRD cases, HRD displayed higher expression of several cell cycle proteins (CHEK1, CCNB1, CCNE1 and YWHAZ) and remarkably lower expression of ATM protein (Figure 3F). Functional enrichment analysis of 30 proteins demonstrated that multiple HRD-related pathways are dysregulated, such as PI3K-Akt signaling pathway, P53 signaling pathway, and cell cycle pathway (FDR < 0.05; Figure 3G).

HRD PC showed inhibitory immune microenvironment

We investigated the infiltration differences of 24 immune cells predicted by ImmuneCellAI between HRD and non-HRD groups in the TCGA cohort. The results showed significantly higher infiltration proportion of Tr1 (p = 3.7E-4), MAIT (p = 3.9E-3), Gamma delta (p = 3.1E-2), CD4 T (p = 2.5E-3), and CD8 T cells (p = 2.5E-3) in the HRD group, while there was a lower infiltration proportion of CD8-naive (p = 4.9E-2), effect memory (p = 5.0E-3), and monocyte (p = 6.7E-3) cells in the HRD group (Figure 4A). Notably, CD4 and CD8 T cells are positive regulators of immunity, and lower infiltration may indicate immunosuppression. Besides, immune checkpoint inhibitor response-predicted results indicated that HRD cases might be insensitive to immunotherapy (p = 5.3E-4; one-sided Fisher's exact test; Figure 4A). Next, a significantly lower immune score was observed in the HRD group when compared with non-HRD group using Estimation of STromal and Immune cells in MAlignant Tumours using Expression data (ESTIMATE) method (p = 2.3E-4; Wilcoxon rank-sum test; Figure 4A). In summary, these results demonstrated that PC samples with HRD display lower immune cell infiltration.

Finally, we compared the expression level of 20 immune checkpoint genes between HRD and non-HRD groups. As a result, most immune checkpoint genes were found to be significantly downregulated in the HRD group, such as *CD274* (*PD-L1*) and *CTLA4* genes. Significantly, only *CD276* (*B7-H3*) gene expression was significantly upregulated in HRD samples compared with non-HRD samples (p = 0.01; Wilcoxon rank-sum test; Figure 4B). Comparisons over T cell receptor (TCR) richness and TCR Shannon showed that the HRD group had a significantly lower TCR diversity than the non-HRD group, indicating a weaker immune response in these cases (p = 1.6E-3 for TCR richness, p = 6.6E-4 for TCR Shannon; Wilcoxon rank-sum test; Figures 4C and 4D).

Genes in 24-GPS positively regulate HRD process via perturbing cell cycle pathway

Eighteen genes in 24-GPS were remarkably upregulated in the HRD group compared with the non-HRD group and were significantly enriched with cell cycle pathways and so forth (p = 5.2E-6; Figures S3C and S3D). Investigating the correlations among 18 genes and HRD-related genes (80 DEGs) in a protein-protein interaction (PPI) network from the Pathway Commons database, we found 76 significantly correlated gene pairs with PPI interactions, which could not be expected in random experiments and constructed as HRD network (p = 4.0E-3; see section, "materials and methods; Figures 5A and S3E). In the HRD network, all the PPI interactions were significantly positively correlated except for two negative interactions. In addition, some cell cycle pathway genes, such as *CDC45* and *E2F1*, had a high degree in the HRD network. In addition, *CCNB1* correlated with many genes in 24-GPS and might act as a key factor regulated by the cell cycle pathway.

Interestingly, *E2F1*, a member of the transcription factor *E2F* family, had a high degree in the HRD network, and it was found to be significantly over-expressed in HRD samples compared with non-HRD samples in the TCGA cohort (p = 2.2E-15; Wilcoxon rank-sum test). Twenty-two target genes of *E2F1* showed significant correlations with *E2F1* in HRD cases rather than non-HRD cases (Figure 5B). Specifically, *CHD4* and *NONO* genes were oncogenes, *CTCF* and *E2F3* acted as transcription factors, and these four functional genes were significantly negatively correlated with *E2F1* in the HRD group (R = -0.36 and p =0.028 for *CHD4*, R = -0.34 and p = 0.039 for *NONO*, R = -0.34 and p = 0.039 for *CTCF*, R = -0.35 and p = 0.032 for *E2F3*; Spearman rank correlation). Among them, *CHD4*, *CTCF*, and *E2F3* showed weak negative correlation with *E2F1* in the non-HRD group (R = -0.045 for *CHD4*, R = -0.16 for *CTCF*, R = -0.18 for *E2F3*; Spearman rank correlation) and were found to be down-expressed in the HRD group compared with the non-HRD group (p = 0.096 for *CHD4*, p = 0.052 for *CTCF*, p = 0.17 for *E2F3*; Wilcoxon rank-sum test; Figure 5C). However, *NONO*, whose expression was positively correlated with *E2F1* in the non-HRD group (R = 0.046; Spearman rank correlation), was significantly over-expressed in the HRD group compared with the non-HRD group (p = 5.0E-3; Wilcoxon rank-sum test; Figure 5C), indicating that *E2F1* upregulates its expression during HRD.

Predicting the combination therapy with PARP inhibitors

Because that deregulation of the cell cycle pathway in the HRD group was found at both transcriptomic and proteomic levels, and considering that *ATR* is a member of cell cycle pathway with the

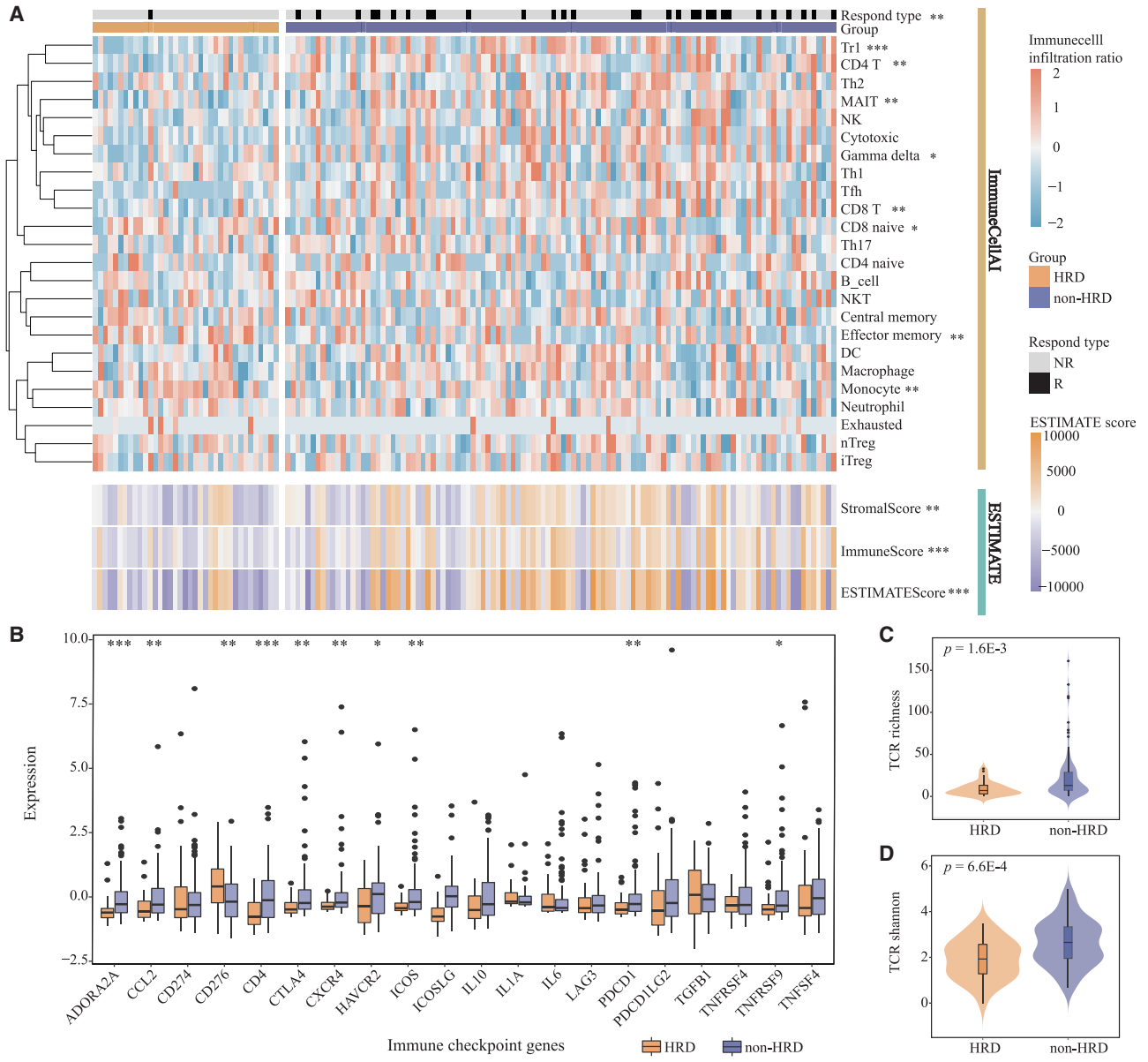

Figure 4. Immune landscape in HRD

(A) Top: annotation bars show the prediction of the response status of TCGA cases to immune checkpoint inhibitors. R, response; NR, not response. Center: the heatmap displays the infiltration differences of 24 immune cells between HRD and non-HRD groups. Bottom: the differences of StromalScore, ImmuneScore, and ESTIMATE score were evaluated in two groups. (B) Box plots display the comparison of expression of immune checkpoint genes between HRD and non-HRD samples. The x axis represents 20 immune checkpoint genes and the y axis exhibits expression value. (C and D) Differences of TCR richness (C) and TCR Shannon (D) between HRD and non-HRD groups. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

target drug, we investigated the co-inhibition effect of *ATR* and PARP genes. First, we observed the tendency of higher expression of *ATR* in HRD cases compared with non-HRD cases in the training cohort (p = 0.25; one-sided Wilcoxon rank-sum test; Figure 6A). Although without significance, PC samples with *ATR* and PARP co-low expression displayed longer OS than other PC samples (p = 0.40; log-rank test; Figure 6B), where low expression was defined as expression level lower than the median value. Additionally, based on 13 HRD cell lines identified earlier, seven cell lines with *ATR* expression higher than their median level were defined as hiATR samples and the remaining six cell lines were classified as the loATR group. Comparative analysis discovered significantly lower survival viability in the loATR group compared with the hiATR group when knocking out the *PARP2* gene (p = 0.037;

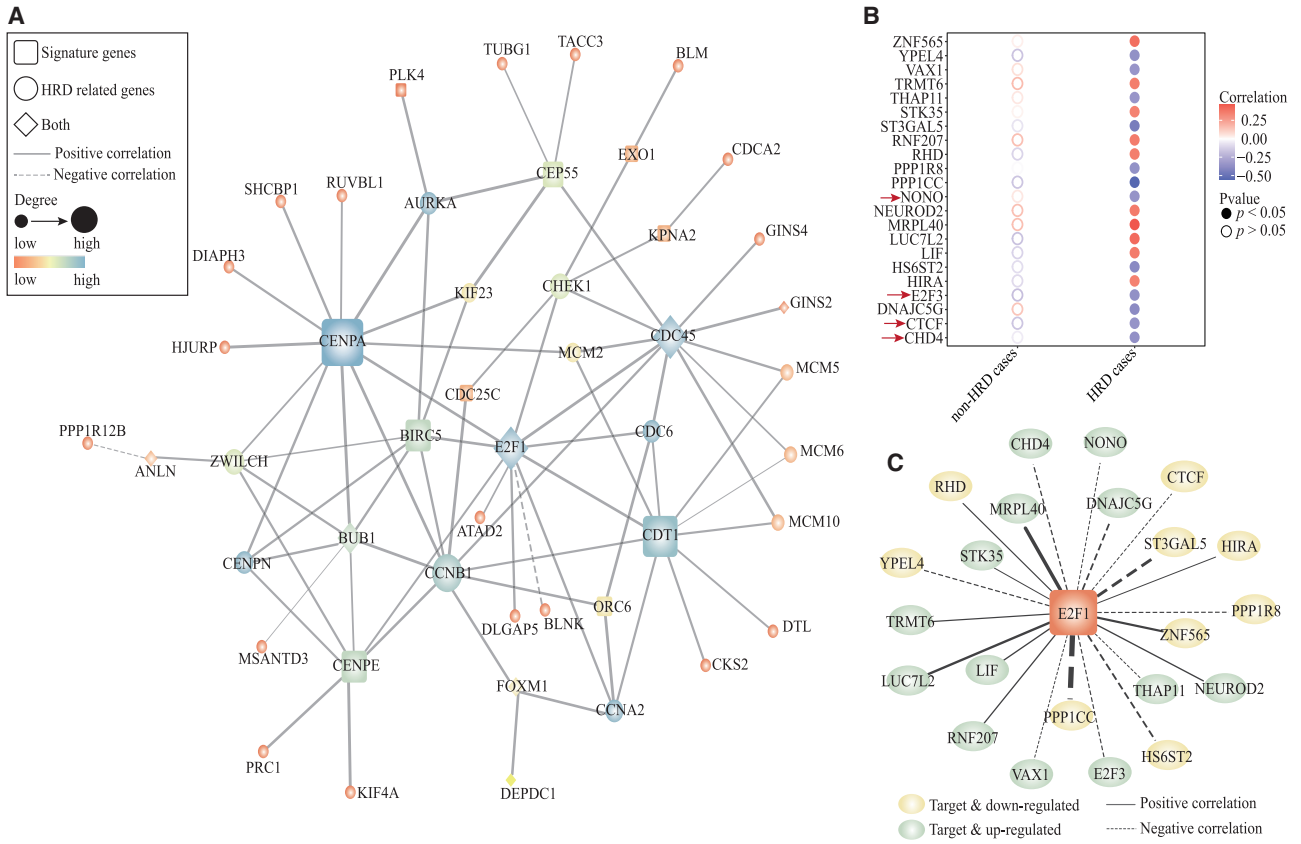

Figure 5. Functional analysis of genes in 24-GPS

(A) PPI network of genes in 24-GPS and HRD-related genes. (B) Comparison of the correlations of *E2F1* transcription factor expression and target genes in HRD and non-HRD groups. Genes of interest are indicated by red arrows. (C) Regulatory network of *E2F1*. The thickness of the edge is positively related to the strength of the expression correlation.

one-sided Wilcoxon rank-sum test; Figure 6C). However, the above difference was not found in the cell lines with knocked down *PARP1* or *PARP3* genes, respectively (p = 0.69 for *PARP1*; p = 0.58 for *PARP3*; one-sided Wilcoxon rank-sum test; Figure 6C).

To unearth more feasible combination strategies with PARP inhibitors, we investigated another 451 cell cycle genes derived from the work of Lundberg et al.[21] Finally, *CDCA2* and *HJURP* genes, which were both significantly upregulated in the HRD group ($p < 0.05$; Wilcoxon rank-sum test), were screened out. Notably, both of them were risk factors for poor prognosis of PC as reported.[22,23] As shown in Figure 6C, PC cells with low expression of *CDCA2* or *HJURP* exhibited a remarkable lower survival ability than those cells with high expression (p = 0.73 for sgPARP1 and p = 0.31 for sgPARP2 and p = 0.017 for sgPARP3 in *CDCA2*, p = 0.42 for sgPARP1 and p = 0.47 for sgPARP2 and p = 0.037 for sgPARP3 in *HJURP*; one-sided Wilcoxon rank-sum test; Figure 6D). Further, a significantly higher expression of *HJURP* was found in HRD cells compared with non-HRD cells, and a similar result was discovered in *CDCA2* (p = 0.021 for *HIURP*, p = 0.12 for *CDCA2*; Wilcoxon rank-sum test; Figure 6E). These results revealed the potential for co-inhibition of the cell cycle pathway and PARP genes.

DISCUSSION

In this work, we developed a transcriptional qualitative signature, 24-GPS, to detect HRD for PC. HRD samples classified by 24-GPS had a significantly unfavorable prognosis without platinum treatment and harbored a higher HRD score compared with non-HRD samples. The HRD group showed genome instability, such as high TMB, greater number of fractions, and more frequent genomic alterations of DDR genes, including mutations and CNA events. Interestingly, the cell cycle pathway was activated in the HRD group at the transcriptome and proteome levels. Furthermore, immune analysis found that PC samples with HRD exhibited lower CD4 and CD8 T cell infiltration, lower immune scores, and under-expression of immune checkpoint genes, which suggests that HRD patients might be not suitable for immunosuppressive therapy. However, our results suggested HRD PC could benefit from the combination of PARP inhibitors and ATR inhibitors.

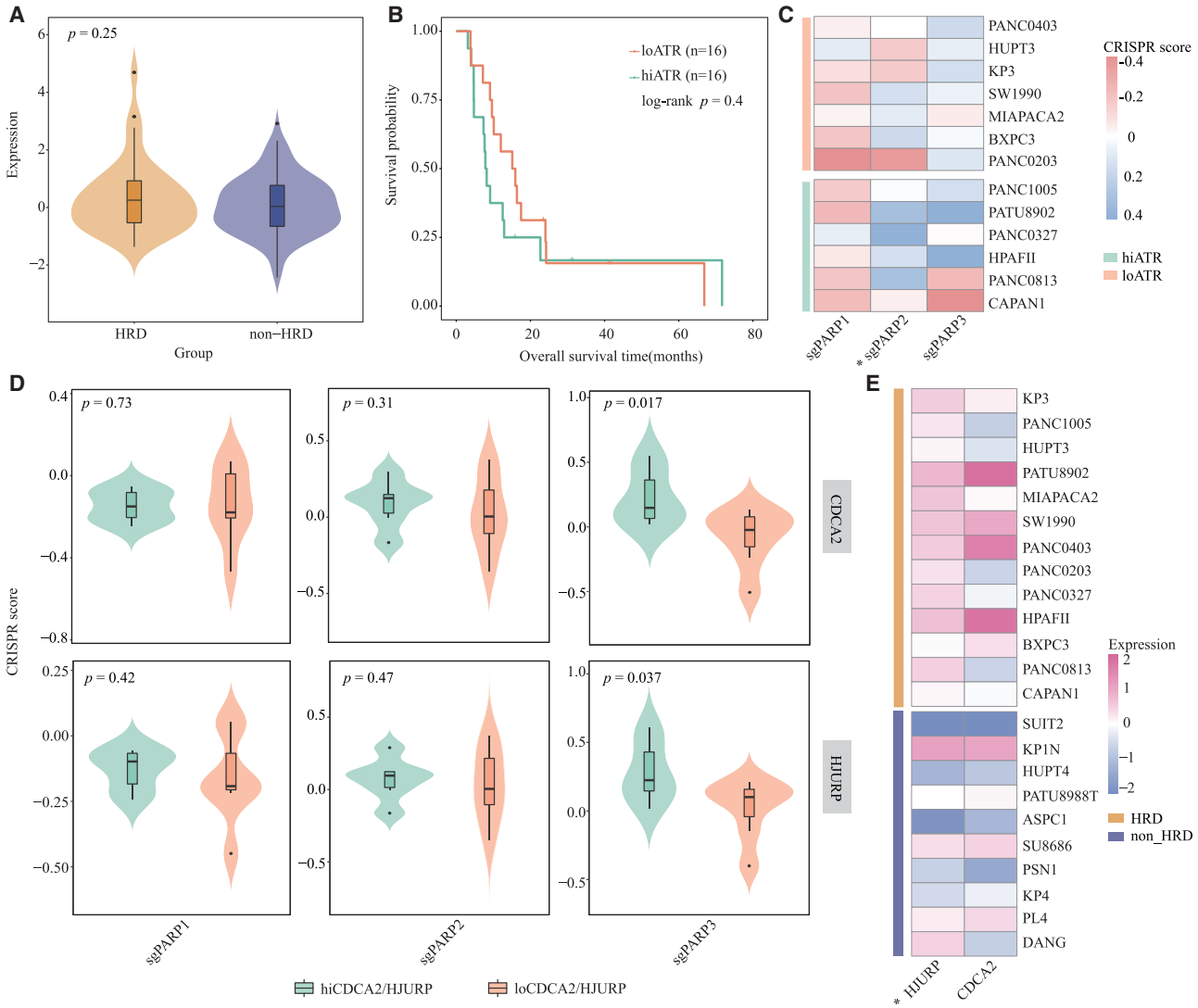

Figure 6. Exploration of the combination of PARP inhibitors and the inhibition of cell cycle pathway in HRD

(A) Box plot displays the expression difference of *ATR* between HRD and non-HRD groups in the TCGA cohort. (B) Kaplan-Meier curve depicts survival difference in samples with high *ATR* expression (hiATR) and low expression (loATR) in $PARP_{neg}$ cases in the TCGA cohort. (C) The differences of CRISPR score between hiATR and loATR samples after knocking out *PARP1*, *PARP2*, and *PARP3* genes, respectively. (D) Box plot displays the difference of tumor cell survival ability between low and high expression of *CDCA2* and *HJURP* genes after knocking out *PARP1*, *PARP2*, and *PARP3* genes, respectively. (E) Distribution of expression of *CDCA2* and *HIURP* genes.

Although HRD also could be induced by some alterations of certain HR genes beyond *BRCA1/2* genes,[6] no gold standard gene panels for HRD detection had been proposed in PC. Besides, for instance, a clinical study showed that patients with germline *BRCA1* mutation underwent disease progression after receiving FOLFIRINOX treatment, but still benefited from olaparib.[24] A similar finding was found in another case, which reported that a patient with a novel somatic *BRCA2* point mutation (p. I2315T) also experienced a progressive disease after 6 weeks of first-line chemotherapy and finally acquired an improved progression-free survival after olaparib treatment.[25] These two case reports indicated that patients with germline or somatic *BRCA1/2* mutation display poor prognosis before treatment with PARP inhibitors, highlighting the great importance of determining the HRD sub-population before clinical treatment. In general, identifying patients who may benefit from treatment with PARP inhibitors is important and challenging.

Because that the success of the unsupervised clustering method in previous research on determining specific subtypes in PC,[26] we performed unsupervised clustering to identify a subgroup with HRD. However, the clustering method could not be applied to individual PC patients. Thus, we developed 24-GPS based on the REO algorithm using transcriptome data to determine HRD status in individuals. The 24-GPS could probably be applied in clinical practice and because (1)

Table1. Detailed information about cohorts

Item	Training	Validation		
Data	TCGA	ICGC-AU	GSE17891	GSE57495
Platform	Illumina HiSeq	GPL10558	GPL570	GPL15048
All samples	147	95	27	63
Subtype	PDAC	PDAC	PDAC	PDAC
OS time	0 ~ 75.07 (months)	1 ~ 1715 (days)	20 ~ 1858 (days)	2.93 ~79.83 (days)
Age (years)	35 ~ 85	34 ~ 90	44 ~ 86	–
Gender				
Male	79 (53.7%)	50 (52.6%)	–	–
Female	68 (46.3%)	45 (47.4%)		
Grade				
G1	21 (14.3%)		4 (14.8%)	
G2	84 (57.1%)	–	12 (44.4%)	–
G3	41 (27.9%)		11 (40.7%)	
G4	1 (0.7%)		–	
Stage				
I	12(8.2%)	9 (6.1%)	3 (11.1%)	13 (20.6%)
II	128(87.1%)	79 (53.7%)	19 (70.4%)	50 (79.4%)
III	3(2%)	1(0.68%)	–	–
IV	3(2%)	6(4.1%)	–	–
Chemotherapy	Yes	No	No	No

PDAC, pancreatic ductal adenocarcinoma

the pairwise relative transcript abundance measurements of genes could be acquired using a low-throughput detection technique, such as real-time PCR; (2) this qualitative signature is relatively robust against experimental and technical variations; (3) 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.

Regardless of the successful application of immune checkpoint inhibitors in melanoma, they are rarely effective on PC.[27] According to the NCCN Clinical Practice Guidelines in Oncology for Pancreatic Adenocarcinoma (Version3, 2019.6), only different mismatch repair or microsatellite Instability-high patients, accounting for only 1% in PC, were suggested to treat pembrolizumab (anti-PD-L1). Beyond single-agent checkpoint inhibitors in PC, recent research has pointed out that PARP inhibitors and immune checkpoint inhibitors represent a rational combination at present, and several I/II clinical trials are ongoing, such as the combination of niraparib with the anti-CTLA-4 ipilimumab or the anti-PD-L1 nivolumab in advanced PC (NCT03404960). However, our results suggested that HRD PC patients might not be suitable for immune checkpoint inhibitor treatment, although a significantly higher TMB was discovered in this group. High TMB has been shown to predict response to immune checkpoint inhibitors across multiple cancer types.[28,29] Notably, beyond the TMB level, some other factors could influence the response rate to immune checkpoint inhibitors, such as the types of mutation, binding affinity for class I MHC proteins, and T cell receptor recognition.[30,31] Thus, although HRD PC patients harbored high TMB and SNV neoantigens, these patients displayed lower TCR diversity (TCR richness and TCR Shannon), which may lead to a weaker immune response.

Interestingly, without PARP inhibitors or platinum-based treatment, HRD PC cases classified by 24-GPS displayed worse OS compared with non-HRD PC cases. Although HRD is an important characteristic in cancer, heterogeneity does exist among different cancer types. Ovarian cancer and breast cancer patients with HRD show better prognosis to platinum-based treatment, while, in PC and prostate cancer, HRD may act as a prognostic risk factor.[32]

There were some limitations in our work. Lacking transcriptomics data of PC with PARP inhibitors treatment hindered a more direct validation for 24-GPS, so we employed both pharmacodynamic and CRISPR/Cas9 data to test the efficiency of 24-GPS in predicting the HRD population and susceptibility to PARP inhibitors. In addition, due to limited research on *CDCA2* and *HJURP* inhibitors, we could not verify the co-inhibition effect with PARP inhibitors using public data. Finally, considering these findings were obtained by bioinformatics analysis, it is worthwhile to further verify the clinical applications of 24-GPS and validate the combination of co-inhibition of PARP and cell cycle pathway in cell line and animal models.

In summary, we developed 24-GPS based on REO, which could detect the HRD status of PC patients at the individual level. HRD PC samples classified by 24-GPS showed distinct HRD characteristics at the multi-omics levels and in clinical factors. 24-GPS has potential implications for selecting PC patients who may benefit from PARP inhibitors in the clinic.

MATERIALS AND METHODS

Data sources

The gene expression profiles of PC patients were downloaded from the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo), cBioportal (https://www.cbioportal.org/), and the ICGC (https://icgc.org/). PC samples from TCGA were used as the training cohort to discover the HRD signature. For samples documented in ICGC-AU, GSE17891 and GSE57495 datasets were used to validate the signature. The multidimensional data of PC patients were derived from cBioportal and TCGA (https://portal.gdc.cancer.gov). In this study, only PC patients with clinical data were included. Detailed information for these datasets is provided in Table 1.

Data processing

For RNA-seq data, the expression of each gene was normalized by the fragment per kilobase million method and transformed by Z score. The illumina Human Methylation 450 array detected 16,826 genes and the reverse-phase protein arrays (RPPA) for 157 selected proteins. The silent mutation was excluded in our work. PC CNA, including AMP, neutral, and DEL, was obtained from cBioportal.

The gene expression profiles from the GEO database have been processed in the following processes: if a gene was mapped to multiple probes, the expression amount of the gene was generated by averaging. Probes that failed to map to any gene ID or map to more than one gene ID were removed.

Cell lines datasets

PC cell line data, including drug toxicity data and knockout data, were accessed from the GDSC database (https://www.cancerrxgene.org/), the CTRP database (http://portals.broadinstitute.org/ctrp/), and the work of Dwane et al.[33] Detailed information for cell lines is provided in Table S1.

Unsupervised clustering

A panel of 247 DDR genes, also detected in the expression profile of the TCGA dataset, was collected from the KEGG database and literature.[34,35] Based on the 247 DDR gene expression profiles, we performed consensus clustering analysis in the k = 1 ∼ 10 (the pre-assigned number of clusters) using the hierarchical clustering algorithm. The clustering process was performed by the R package ConsensusClusterPlus.[36] Multiple HRD-related features were compared among clusters using Wilcoxon rank-sum test.

Discovering a REO-based HRD signature

The REO-based HRD signature-discovering procedure included three main parts: (1) using E_i and E_j to represent the expression value of a gene pair (gene $_i$ and gene $_j$, respectively). We screened out a panel of gene pairs, which had a stable expression pattern of $E_i > E_j$ in more than 95% of HRD-like samples and significantly reversed in non-HRD-like samples examined by Fisher's exact test with 5% FDR control. (2) Then we ranked these gene pairs in descending order according to FD value (Equations 1 and 2), and the top 1,000 gene pairs were used for further signature discovery. (3) Finally, we implemented the LASSO model to further optimize the 1,000 gene pairs. A set of 24 gene pairs with regression coefficient greater than 0 was obtained. The process was carried out by the R package glmnet.[37]

$$\boldsymbol{pij}(\boldsymbol{c}) = \boldsymbol{P}\left(\boldsymbol{E_i} > \boldsymbol{E_j} \middle| \boldsymbol{c}\right), \boldsymbol{c} = 1, 2, \quad \text{Equation 1}$$

Equation 1 calculates the probabilities of observing $E_i > E_j$ in HRD-like and non-HRD-like group, respectively.

$$\boldsymbol{FD}(\boldsymbol{ij}) = \boldsymbol{pij}(1) - \boldsymbol{pij}(2), \quad \text{Equation 2}$$

Equation 2 calculates the FD value of gene pair (gene $_i$, gene $_j$).

Definition of HRD samples

A patient was classified into the HRD group if at least 65% of the 24-GPS voted for HRD, otherwise the non-HRD group. A cell line was classified into the HRD group only if at least 23 gene pairs voted for HRD, otherwise the non-HRD group.

HRD-related features score

HRD score, number of fraction, SNV neoantigens, and aneuploidy score were derived from research by Thorsson et al.[38] HRDetect score was acquired from Davies et al.[10] TMB was defined by mutations per million bases and calculated by the R package maftools.

The definition of RAD51-defective samples

We ranked the protein expression level of *RAD51* gene from high to low across all samples and defined the bottom 10% as *RAD51* deficient, with the others regarded as *RAD51* proficient. Fisher's exact test was used to examine whether *RAD51*-defective cases were significantly enriched in the HRD group.

Defining $PARP_{neg}$ and $PARP_{pos}$ samples

PC samples whose expression of any *PARP1/2/3* gene was lower than median level were defined as$PARP_{neg}$, otherwise they were $PARP_{pos}$.

Multi-omics analysis

Fisher's exact test was performed to detect genes with significantly differential mutations or significantly higher-frequency CNA between HRD and non-HRD groups. Differential expression and methylation genes were identified by a two-sided Wilcoxon rank-sum test with FDR values less than 0.05 adjusted by Benjamini and Hochberg (BH). One-sided Wilcoxon rank-sum test was used to identify proteins with significant DEGs in HRD samples compared with non-HRD samples.

Immune analysis

The relative fraction of 24 immune cells and immune checkpoint inhibitor response prediction were estimated using the ImmuneCellAI method based on TCGA RNA-seq data.[39] Then, the overall immune score and stroma score were calculated by the R package estimate.[40] Twenty known immune checkpoint genes were acquired from Mak et al.[41] Finally, TCR diversity scores, TCR richness, and TCR Shannon were obtained from Thorsson et al.[38] Wilcoxon rank-sum test was used to compare the differences of immune cell infiltration abundance, immune score, immune checkpoint gene expression, and two TCR-related features between HRD and non-HRD groups. The difference in immune checkpoint inhibitor response prediction between the two groups was examined by Fisher's exact test.

Functional enrichment analysis

Functional enrichment analysis of HRD-related DEGs and differentially expressed proteins was performed by the R package clusterProfiler based on KEGG. A hypergeometric test was used to test the significance of enrichment results with FDR value adjusted by BH.

PPI network analysis

The HRD network consisting of interactions between HRD-related genes and 18 genes in 24-GPS was constructed using pathway common database (https://www.pathwaycommons.org). The HRD network was visualized by the Cytoscape tool (version 3.7.0). Spearman rank correlation was used to calculate the correlation of interactions in the HRD network.

Then we set up 1,000 random networks to test the biological significance of the HRD network. The process was performed as follows: (1) in each experiment, we randomly extracted nodes with the same amount of HRD-related genes and 18 genes, and calculated the number of interactions between them. (2) We counted the number of times that the interactions in random experiments exceeded those in the HRD network, and divided it by 1,000 to get the statistical p value. If the p value was less than 0.05, the HRD network was considered biologically significant.

Survival analysis

OS was estimated by Kaplan-Meier curves and tested by log-rank test with p value less than 0.05.

Statistical analysis

All statistical analyses were carried out using R software version 4.0.2 (http://www.r-project.org/).

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.omtn.2021.10.014.

ACKNOWLEDGMENTS

This research was funded by grants from the National Natural Science Foundation of China (grant number 61673143), the Outstanding Youth Foundation of Heilongjiang Province of China (grant number YQ2021H005), and the Postdoctoral Scientific Research Developmental Fund (grant number LBH-Q16166).

AUTHOR CONTRIBUTIONS

Z.C. and Y.G. conceived the original idea and supervised the study. S.Z. carried out the data processing and analysis and wrote the manuscript. Y.W., L.A., Y.G., and M.Z. carried out the visualization. K.L., H.X., L.W., and Z.Z. carried out the data curation; and T.C. and Y.L. edited the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Ward, E., DeSantis, C., Robbins, A., Kohler, B., and Jemal, A. (2014). Childhood and adolescent cancer statistics, 2014. CA Cancer J. Clin. *64*, 83–103.
2. Gillen, S., Schuster, T., Meyer Zum Buschenfelde, C., Friess, H., and Kleeff, J. (2010). Preoperative/neoadjuvant therapy in pancreatic cancer: a systematic review and meta-analysis of response and resection percentages. PLoS Med. *7*, e1000267.
3. Tempero, M.A., Malafa, M.P., Al-Hawary, M., Behrman, S.W., Benson, A.B., Cardin, D.B., Chiorean, E.G., Chung, V., Czito, B., Del Chiaro, M., et al. (2021). Pancreatic Adenocarcinoma, Version 2.2021, NCCN Clinical Practice Guidelines in Oncology. J. Natl. Compr. Canc. Netw. *19*, 439–457.
4. Singh, H.M., Bailey, P., Hubschmann, D., Berger, A.K., Neoptolemos, J.P., Jager, D., Siveke, J., and Springfeld, C. (2021). Poly(ADP-ribose) polymerase inhibition in pancreatic cancer. Genes Chromosomes Cancer *60*, 373–384.
5. Golan, T., O'Kane, G.M., Denroche, R.E., Raitses-Gurevich, M., Grant, R.C., Holter, S., Wang, Y., Zhang, A., Jang, G.H., Stossel, C., et al. (2021). Genomic features and classification of homologous recombination deficient pancreatic ductal adenocarcinoma. Gastroenterology *160*, 2119–2132 e2119.
6. Park, W., Chen, J., Chou, J.F., Varghese, A.M., Yu, K.H., Wong, W., Capanu, M., Balachandran, V., McIntyre, C.A., El Dika, I., et al. (2020). Genomic methods identify homologous recombination deficiency in pancreas adenocarcinoma and optimize treatment selection. Clin. Cancer Res. *26*, 3239–3247.
7. McIntyre, C.A., Lawrence, S.A., Richards, A.L., Chou, J.F., Wong, W., Capanu, M., Berger, M.F., Donoghue, M.T.A., Yu, K.H., Varghese, A.M., et al. (2020). Alterations in driver genes are predictive of survival in patients with resected pancreatic ductal adenocarcinoma. Cancer *126*, 3939–3949.
8. Pokataev, I., Fedyanin, M., Polyanskaya, E., Popova, A., Agafonova, J., Menshikova, S., Tryakin, A., Rumyantsev, A., and Tjulandin, S. (2020). Efficacy of platinum-based chemotherapy and prognosis of patients with pancreatic cancer with homologous recombination deficiency: comparative analysis of published clinical studies. ESMO Open *5*, e000578.
9. 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.
10. Davies, H., Glodzik, D., Morganella, 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.
11. 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.
12. 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.
13. Konstantinopoulos, P.A., Spentzos, D., Karlan, B.Y., Taniguchi, T., Fountzilas, E., Francoeur, N., Levine, D.A., and Cannistra, S.A. (2010). Gene expression profile of BRCAness that correlates with responsiveness to chemotherapy and with outcome in patients with epithelial ovarian cancer. J. Clin. Oncol. *28*, 3555–3561.
14. Wang, H., Sun, Q., Zhao, W., Qi, L., Gu, Y., Li, P., Zhang, M., Li, Y., Liu, S.L., and Guo, Z. (2015). Individual-level analysis of differential expression of genes and pathways for personalized medicine. Bioinformatics *31*, 62–68.
15. 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.
16. Patil, P., Bachant-Winner, P.O., Haibe-Kains, B., and Leek, J.T. (2015). Test set bias affects reproducibility of gene signatures. Bioinformatics *31*, 2318–2323.
17. Guan, Q., Chen, R., Yan, H., Cai, H., Guo, Y., Li, M., Li, X., Tong, M., Ao, L., Li, H., et al. (2016). Differential expression analysis for individual cancer samples based on robust within-sample relative gene expression orderings across multiple profiling platforms. Oncotarget *7*, 68909–68920.
18. Michels, J., Vitale, I., Saparbaev, M., Castedo, M., and Kroemer, G. (2014). Predictive biomarkers for cancer therapy with PARP inhibitors. Oncogene *33*, 3894–3907.
19. Katoh, M. (2012). Function and cancer genomics of FAT family genes (review). Int. J. Oncol. *41*, 1913–1918.
20. Luan, H., Zhang, C., Zhang, T., He, Y., Su, Y., and Zhou, L. (2020). Identification of key prognostic biomarker and its correlation with immune infiltrates in pancreatic ductal adenocarcinoma. Dis Markers *2020*, 8825997.
21. Lundberg, A., Lindstrom, L.S., Parker, J.S., Loverli, E., Perou, C.M., Bergh, J., and Tobin, N.P. (2020). A pan-cancer analysis of the frequency of DNA alterations across cell cycle activity levels. Oncogene *39*, 5430–5440.
22. Xing, C., Wang, Z., Zhu, Y., Zhang, C., Liu, M., Hu, X., Chen, W., and Du, Y. (2021). Integrate analysis of the promote function of cell division cycle-associated protein family to pancreatic adenocarcinoma. Int. J. Med. Sci. *18*, 672–684.
23. Wang, C.J., Li, X., Shi, P., Ding, H.Y., Liu, Y.P., Li, T., Lin, P.P., Wang, Y.S., Zhang, G.Q., and Cao, Y. (2020). Holliday junction recognition protein promotes pancreatic cancer growth and metastasis via modulation of the MDM2/p53 signaling. Cell Death Dis. *11*, 386.

24. Pimenta, J.R., Ueda, S.K.N., and Peixoto, R.D. (2020). Excellent response to olaparib in a patient with metastatic pancreatic adenocarcinoma with germline BRCA1 mutation after progression on FOLFIRINOX: case report and literature review. Case Rep. Oncol. *13*, 904–910.

25. Wang, D., Guan, R., Tao, Q., Liu, S., Yu, M., and Li, X. (2021). A novel somatic BRCA2 point mutation in a metastatic pancreatic cancer patient: a case report. BMC Med. Genomics *14*, 6.

26. Bailey, P., Chang, D.K., Nones, K., Johns, A.L., Patch, A.M., Gingras, M.C., Miller, D.K., Christ, A.N., Bruxner, T.J., Quinn, M.C., et al. (2016). Genomic analyses identify molecular subtypes of pancreatic cancer. Nature *531*, 47–52.

27. Li, J., Byrne, K.T., Yan, F., Yamazoe, T., Chen, Z., Baslan, T., Richman, L.P., Lin, J.H., Sun, Y.H., Rech, A.J., et al. (2018). Tumor cell-intrinsic factors underlie heterogeneity of immune cell infiltration and response to immunotherapy. Immunity *49*, 178–193 e177.

28. Buttner, R., Longshore, J.W., Lopez-Rios, F., Merkelbach-Bruse, S., Normanno, N., Rouleau, E., and Penault-Llorca, F. (2019). Implementing TMB measurement in clinical practice: considerations on assay requirements. ESMO Open *4*, e000442.

29. Samstein, R.M., Lee, C.H., Shoushtari, A.N., Hellmann, M.D., Shen, R., Janjigian, Y.Y., Barron, D.A., Zehir, A., Jordan, E.J., Omuro, A., et al. (2019). Tumor mutational load predicts survival after immunotherapy across multiple cancer types. Nat. Genet. *51*, 202–206.

30. Duffy, M.J., and Crown, J. (2019). Biomarkers for predicting response to immunotherapy with immune checkpoint inhibitors in cancer patients. Clin. Chem. *65*, 1228–1238.

31. Luksza, M., Riaz, N., Makarov, V., Balachandran, V.P., Hellmann, M.D., Solovyov, A., Rizvi, N.A., Merghoub, T., Levine, A.J., Chan, T.A., et al. (2017). A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy. Nature *551*, 517–520.

32. Knijnenburg, T.A., Wang, L., Zimmermann, M.T., Chambwe, N., Gao, G.F., Cherniack, A.D., Fan, H., Shen, H., Way, G.P., Greene, C.S., et al. (2018). Genomic and molecular landscape of DNA damage repair deficiency across The Cancer Genome Atlas. Cell Rep. *23*, 239–254 e236.

33. Dwane, L., Behan, F.M., Goncalves, E., Lightfoot, H., Yang, W., van der Meer, D., Shepherd, R., Pignatelli, M., Iorio, F., and Garnett, M.J. (2021). Project Score database: a resource for investigating cancer cell dependencies and prioritizing therapeutic targets. Nucleic Acids Res. *49*, D1365–D1372.

34. Liu, C., Rohart, F., Simpson, P.T., Khanna, K.K., Ragan, M.A., and Le Cao, K.A. (2016). Integrating multi-omics data to dissect mechanisms of DNA repair dysregulation in breast cancer. Sci. Rep. *6*, 34000.

35. Stefansson, O.A., Villanueva, A., Vidal, A., Marti, L., and Esteller, M. (2012). Re: A DNA repair pathway-focused score for prediction of outcomes in ovarian cancer treated with platinum-based chemotherapy. J. Natl. Cancer Inst. *104*, 1514, author reply 1514-1515.

36. Wilkerson, M.D., and Hayes, D.N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics *26*, 1572–1573.

37. Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. *33*, 1–22.

38. Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Ou Yang, T.H., Porta-Pardo, E., Gao, G.F., Plaisier, C.L., Eddy, J.A., et al. (2018). The immune landscape of cancer. Immunity *48*, 812–830 e814.

39. Miao, Y.R., Zhang, Q., Lei, Q., Luo, M., Xie, G.Y., Wang, H., and Guo, A.Y. (2020). ImmuCellAI: A unique method for comprehensive T-cell subsets abundance prediction and its application in cancer immunotherapy. Adv. Sci. (Weinh) *7*, 1902880.

40. Yoshihara, K., Shahmoradgoli, M., Martinez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Trevino, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. Nat. Commun. *4*, 2612.

41. Mak, M.P., Tong, P., Diao, L., Cardnell, R.J., Gibbons, D.L., William, W.N., Skoulidis, F., Parra, E.R., Rodriguez-Canales, J., Wistuba, , II, et al. (2016). A patient-derived, pan-cancer EMT signature identifies global molecular alterations and immune target enrichment following epithelial-to-mesenchymal transition. Clin. Cancer Res. *22*, 609–620.