





OPEN

DNA damage repair-related gene signature for identifying the immune status and predicting the prognosis of hepatocellular carcinoma

Yongpan Lu^{1,5}, Sen Wang^{2,5}, Tingting Chi³, Yuli Zhao², Huimin Guo⁴, Haizheng Wang² & Li Feng²  

The heterogeneity of hepatocellular carcinoma (HCC) poses a challenge for accurate prognosis prediction. DNA damage repair genes (DDRGs) have an impact on a wide range of malignancies. However, the relevance of these genes in HCC prognosis has received little attention. In this study, we aimed to develop a prognostic signature to identify novel therapy options for HCC. We acquired mRNA expression profiles and clinical data for HCC patients from The Cancer Genome Atlas (TCGA) database. A polygenic prognostic model for HCC was constructed using selection operator Cox analysis and least absolute shrinkage. The model was validated using International Cancer Genome Consortium (ICGC) data. Overall survival (OS) between the high-risk and low-risk groups was compared using Kaplan–Meier analysis. Independent predictors of OS were identified through both univariate and multivariate Cox analyses. To determine immune cell infiltration scores and activity in immune-related pathways, a single-sample gene set enrichment analysis was performed. The protein and mRNA expression levels of the prognostic genes between HCC and normal liver tissues were also examined by immunohistochemistry (IHC), immunofluorescence (IF) and quantitative real-time PCR (qRT-PCR). A novel ten-gene signature (*CHD1L*, *HDAC1*, *KPNA2*, *MUTYH*, *PPP2R5B*, *NEIL3*, *POLR2L*, *RAD54B*, *RUVBL1* and *SPP1*) was established for HCC prognosis prediction. Patients in the high-risk group had worse OS than those in the low-risk group. Receiver operating characteristic curve analysis confirmed the predictive ability of this prognostic gene signature. Multivariate Cox analysis showed that the risk score was an independent predictor of OS. Functional analysis revealed a strong association with cell cycle and antigen binding pathways, and the risk score was highly correlated with tumor grade, tumor stage, and types of immune infiltrate. High expression levels of the prognostic genes were significantly correlated with increased sensitivity of cancer cells to antitumor drugs. IHC, IF and qRT-PCR all indicated that the prognostic genes were highly expressed in HCC relative to normal liver tissue, consistent with the results of bioinformatics analysis. Ten DDRGs were utilized to create a new signature for identifying the immunological state of HCC and predicting prognosis. In addition, blocking these genes could represent a promising treatment.

¹Department of Plastic Surgery, Shandong University of Traditional Chinese Medicine, The First Affiliated Hospital of Shandong First Medical University and Shandong Provincial Qian Foshan Hospital, Jingshi Road, Jinan 250014, Shandong, China. ²Department of Medical Ultrasound, Shandong Medicine and Health Key Laboratory of Abdominal Medical Imaging, The First Affiliated Hospital of Shandong First Medical University and Shandong Provincial Qian Foshan Hospital, Shandong First Medical University, No. 16766, Jingshi Road, Jinan 250014, Shandong, China. ³Department of Acupuncture and Rehabilitation, The Affiliated Qingdao Hai Ci Hospital of Qingdao University (West Hospital Area), Qingdao 266000, Shandong, China. ⁴Department of Medical Ultrasound, Shandong Medicine and Health Key Laboratory of Abdominal Medical Imaging, The First Affiliated Hospital of Shandong First Medical University and Shandong Provincial Qian Foshan Hospital, Jining Medical College, No. 16766, Jingshi Road, Jinan 250014, Shandong, China. ⁵These authors contributed equally: Yongpan Lu and Sen Wang. ✉email: qfsfengli@163.com

Liver cancer is the fourth most frequent cause of cancer-related mortality worldwide and has the sixth highest incidence rate¹. Most primary liver malignancies are hepatocellular carcinomas (HCCs)². The majority of HCCs are triggered by an underlying liver disease, usually alcoholism or hepatitis B or C virus (HBV or HCV) infection³. In China, the 5-year survival rate for patients with HCC, a complicated and diverse illness, is only 14.1% because of the cancer's high incidence of recurrence, which is frequently accompanied by cirrhosis or other comorbidities that make determining the prognosis extremely difficult⁴.

It is well known that the regenerative capacity of the liver is closely linked to the DNA repair process^{5–7}. Both mechanisms are substantially dysregulated after chronic liver injury, and the risk of genetic instability is enhanced^{8,9}. The DNA damage response (DDR) pathways, which coordinate DNA repair, cell cycle arrest, and eventually cell death or senescence, are affected by this imbalance^{10,11}. Hepatic genomic integrity is weakened by abnormalities in DNA repair and related pathways, such as mismatch repair (MMR), homologous recombination (HR), and nonhomologous end joining (NHEJ), which activate hepatocarcinogenesis and result in the development of HCC^{12,13}. Given the significance of liver regeneration and the reaction to DNA damage, they could have an additive influence on prognosis and responsiveness to therapy. An increasing number of studies support the use of many proteins to create complete prognosis scores for malignancies based on DNA damage repair. Some DNA damage repair genes (DDRGs) have been employed to forecast the overall survival (OS) of HCC in addition to serum indicators¹⁴. Recent findings have shown that HCC was categorized into high-repair and low-repair groups based on the expression of DDRGs, and that the high-repair group had a worse prognosis, increased expression of p53 mutation-like genes, and more clinically aggressive features compared to the low-repair group¹⁵. However, additional research is necessary to determine the exact mechanism by which DDR causes cancer in HCC. Additionally, individualized assessments of the prognosis of the disease can be improved and innovated.

We retrieved the mRNA expression profiles and associated clinical information of HCC patients for this investigation from the TCGA database. Then, we created a prognostic profile of differentially expressed genes (DEGs) linked to DNA damage repair, and we used the ICGC databases to confirm the stability and dependability of the model. A functional enrichment analysis was conducted to explore the underlying processes. Additionally, the relationship between immune infiltration and prognostic gene expression was examined. Furthermore, the association between tumor stemness and cancer chemoresistance based on the expression levels of the prognostic genes was investigated. Finally, experimental validation of the protein and mRNA expression levels of the prognostic genes in HCC and normal tissues was performed.

Methods

Data acquisition

To obtain the clinical data and RNA sequencing data for 374 patients with liver cancer, we utilized the TCGA website (<https://portal.gdc.cancer.gov/repository>). Additionally, samples from an additional 231 HCC patients were collected from the ICGC website (<https://dcc.icgc.org/projects/LIRI-JP>). We identified 513 DNA damage repair genes (DDRGs) from the Molecular Signatures database (MSigDB, <http://www.gsea-msigdb.org/gsea>) and prior literature¹⁶ (Supplementary Table S1).

Construction and validation of a prognostic DNA damage repair-related gene signature

The “limma” R package was used to identify DEGs between tumor and normal tissues in the TCGA cohort with a fold change > 1 and a false discovery rate < 0.05. Then, we identified DDRGs with prognostic significance using univariate Cox analysis and the Benjamini and Hochberg (BH) correction approach. To reduce the risk of overfitting, LASSO-penalized Cox regression analysis was employed to construct a prognostic model. The LASSO method can shrink the coefficients of some unimportant features to zero, thus improving the selection of features; and it can produce sparse solutions at multicollinearity, which is suitable for multi-dimensional datasets and multicollinear data; moreover, it can simplify the complexity of the model while maintaining a high prediction accuracy^{17–19}. The “glmnet” R package with the LASSO algorithm was used to choose and compress variables to obtain regression coefficients precisely equal to zero and an understandable model. The independent variable in the LASSO Cox regression analysis was a matrix of standardized expression levels for potential prognostic DEGs, and the dependent variables were OS and the status of patients in the TCGA cohort. We employed ten-fold cross-validation to determine the penalty parameter (λ) for the prognostic model, selecting the value of λ corresponding to the lowest partial likelihood deviation. The DDRG expression levels and their corresponding regression coefficients were used to calculate patient risk scores. The equation was as follows:

$$\text{risk score} = e^{\text{sum}(\text{each gene's expression} \times \text{corresponding coefficient})}$$

Using the median risk scores, patients were classified into high-risk and low-risk groups. The “Rtsne” and “ggplot2” R packages were employed to perform PCA and t-SNE analyses to investigate the distribution of the two distinct risk groups. The “survminer” R package was utilized to perform survival analysis. Furthermore, a time-dependent receiver operating characteristic (ROC) curve analysis was conducted to determine the predictive power of prognostic characteristics. Univariate and multivariate Cox analyses were performed to investigate the independent prognostic significance of the 10 genetic markers.

Functional enrichment analysis

The online websites (<http://www.bioinformatics.com.cn/>) was utilized to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses of the DEGs in high and low risk groups. Additionally, we used the “gsva” package to perform single-sample gene set enrichment analysis (ssGSEA) to determine the infiltration of diverse immune cells and the activity levels of diverse immune-related pathways in each sample. The BH technique was employed to adjust the P value.

Analysis of the tumor microenvironment and immune response

We assessed the level of infiltration of immune and stromal cells in distinct tumor tissues using immune and stromal scores²⁰. To investigate the relationship between the risk score and immune infiltration subtype, we performed a 2-way ANOVA. To measure the stem cell-like characteristics of tumors, we extracted tumor stem cell features from the transcriptome and epigenetics of TCGA tumor samples²¹.

Analysis of chemotherapy sensitivity

We accessed the NCI-60 database through the CellMiner interface (<https://discover.nci.nih.gov/cellminer>), which includes 60 cancer cell lines from 9 different tumor types. Pearson correlation analysis was used to assess the correlation between prognostic gene expression and drug sensitivity. The effectiveness of 217 FDA-approved drugs (Supplementary Table S2) was evaluated using a correlation study.

Analysis of the human protein atlas (HPA) database

The HPA (<http://www.proteinatlas.org/>) database allows researchers to access enormous volumes of proteomic and transcriptome data for individual human cells and tissues²². Protein expression levels of ten genes were verified by immunohistochemistry (IHC) in the HPA database.

Patients and tissue samples

All HCC and healthy liver tissue sections were obtained from the Shandong Provincial Hospital of Traditional Chinese Medicine's Department of Pathology. The Declaration of Helsinki guidelines were followed in the conduct of this investigation. This study was approved by Shandong University of Traditional Chinese Medicine (Approval Number: AF/SC-08/02.0).

HCC mouse model

BALB/c (4–6 weeks) mice weighing 20 ± 3 g were selected for the construction of the HCC model. After one week of acclimatization feeding, 200 μ l of mouse HCC cells (1×10^7 cells/ml resuspended in sterile PBS solution) was injected subcutaneously into the right shoulder of mice. When the tumors had developed to a diameter of approximately 10 mm, the mice were killed with an intraperitoneal injection of 250 mg/kg pentobarbital sodium. Then, the tumor was removed. Controls were normal BALB/c mouse livers without any treatment. Fresh mouse tumor and liver tissues were stored at -80 °C for quantitative real-time PCR (qRT-PCR). All animal experiments in this study were approved by the Animal Experiments Ethics Committee of The First Affiliated Hospital of Shandong First Medical University, approval number: SYDWLS(2021)002.

Immunohistochemistry (IHC) and immunofluorescence (IF) to confirm protein expression differences in prognostic genes between HCC and normal tissues

Tissue slices were deparaffinized, rehydrated, and treated with 0.3% methanol- H_2O_2 solution at room temperature for 20 min to quench endogenous peroxidase activity; the slices were then washed three times with PBS. Antigen retrieval was carried out with EDTA treatment for 20 min; the slices were then washed three times with PBS, blocked with 5% goat serum for 30 min, and incubated with primary antibody overnight. On the second day, the slices were warmed, washed three times with PBS, incubated with secondary antibody for 1 h, washed three times with PBS, and developed with DAB for 4 min. Counterstaining was performed with hematoxylin for 1 min, followed by treatment with hydrochloric acid alcohol differentiation solution for 2 s; hematoxylin was “blued” for 10 min. The slices were dehydrated through graded alcohols, immersed in xylene for 10 min, and mounted with neutral resin.

Paraffin sections were dewaxed in water, and the sections were sequentially placed in environmentally friendly dewaxing solution I for 10 min followed by solution II for 10 min. The sections were then washed with anhydrous ethanol I for 5 min, anhydrous ethanol II for 5 min, anhydrous ethanol III for 5 min, and distilled water. Afterward, the antigen was repaired, and it was finished by natural cooling. The slides were submerged in PBS (pH 7.4) and rinsed three times; each time, they were shaken for five minutes on a decolorizing shaker. After being softly shaken dry, the sections were circled around the tissue with a histochemical pen, and 3% BSA was added dropwise for 30 min. After applying the prepared primary antibody dropwise to the sections, they were incubated overnight at 4 °C in a humidified box. The slides were submerged in PBS (pH 7.4) and cleaned three times; each time, they were shaken for five minutes on a decolorizing shaker. On this shaker, the slides were washed three times in PBS (pH 7.4) for five minutes each time. The slides were observed and recorded using a Nikon Eclipse Ti2 confocal microscope (Nikon Instruments (Shanghai) Co., Ltd., Shanghai, China).

The primary antibodies used in our work were as follows: anti-*CHD1L* antibody (DF8521, Affinity), anti-*HDAC1* antibody (AF6433, Affinity), anti-*KPNA2* antibody (DF6510, Affinity), anti-*MUTYH* antibody (DF6500, Affinity), *NEIL3* polyclonal antibody (YT3031, ImmunoWay), anti-*POLR2L* antibody (DF6960, Affinity), anti-*PPP2R5B* rabbit polyclonal antibody (HA500409, HUABIO), *Rad54B* polyclonal antibody (YT3971, ImmunoWay), anti-*RUVBL1* antibody (DF7961, Affinity), and anti-*SPP1* antibody (AF0227, Affinity).

RNA isolation and qRT-PCR

Total RNA was extracted using RNAiso Plus (9109, Takara, Japan) according to the manufacturer's protocol. Sketch™ RT Master Mix (RR036A, Takara, Japan) was used to synthesize complementary DNA by reverse transcription of RNA into DNA. Quantitative real-time PCR (qRT-PCR) experiments were performed using TB-Green™ Premix™ II (RR820A, Takara, Japan). Primers were designed and synthesized by Servicebio. GAPDH

was used as an endogenous reference gene. Relative gene expression was determined using the $2^{-\Delta\Delta CT}$ method. The sequences of the employed PCR primers are shown below.

GAPDH-F 5'-CCTCGTCCCCTAGACAAAATG-3'
GAPDH-R 5'-TGAGGTCAATGAAGGGGTCGT-3'
HDAC1-F 5'-CACAAAGCCAATGCTGAGGAG-3'
HDAC1-R 5'-CGATGTCCGTCTGCTGCTTAT-3'
KPNA2-F 5'-AGAACCTTTGATGAACCTCCTGA-3'
KPNA2-R 5'-TTTTATCCAAGCCTCCACACTCT-3'
CHD1L-F 5'-GAAGACCTGAGTTTGGGTGATGT-3'
CHD1L-R 5'-CGCTTGCTTTCTTTTCTTTTGC-3'
PPP2R5B-F 5'-GCGGTATTTGGGACCCTCTA-3'
PPP2R5B-R 5'-TTCTGCTGCTCCTGTTGTTTTT-3'
RAD54B-F 5'-CAGACGAGAATCACCAGCGG-3'
RAD54B-R 5'-CCTAAGCCCATTTCATCAGCAA-3'
RUVBL1-F 5'-AGAGCACCACGAAGACGCA-3'
RUVBL1-R 5'-GCAATAGCAAGGGCCAAGG-3'
POLR2L-F 5'-CAAATGGGAAGCCTACCTGG-3'
POLR2L-R 5'-CAGCTTCTCAATCAGGTCCACG-3'
MUTYH-F 5'-ATCGCCTTTGACCAGGTAACC-3'
MUTYH-R 5'-ATGGCAGCTTGATTGAAGTCCC-3'
SPP1-F 5'-TACAGCCTGCACCCAGATCCTAT-3'
SPP1-R 5'-GCTTTCATTGGAATTGCTTGGA-3'
NEIL3-F 5'-GGCTGCTCCAATGAATGCTAA-3'
NEIL3-R 5'-CTCCCGTGGGTTAATCAAGATG-3'

Statistical analysis

R (version 4.0.5) and the appropriate packages were employed for statistical analyses. Statistically significant differences were denoted by a P value of < 0.05. To determine differences between two groups, the Wilcoxon test was applied. Kaplan–Meier curves were used to create survival curves.

Ethics approval

This study were approved by the ethics committee of Shandong University of Traditional Chinese Medicine (AF/SC-08/02.0) and the Animal Experiments Ethics Committee of The First Affiliated Hospital of Shandong First Medical University, approval number: SYDWLS(2021)002. All procedures were carried out in strict accordance with the 1964 Declaration of Helsinki guidelines. All patients involved in this study provided informed consent before the study.

Results

This study's flow chart is shown in Fig. 1. A total of 365 HCC patients from TCGA and 231 HCC patients from ICGC were included in this analysis after several HCC patients with insufficient clinical data were excluded. Supplementary Table S3 provides a summary of the patients' full clinical features.

Identification of prognostic DNA damage repair-associated DEGs in the TCGA cohort

A total of 230 DDRGs were expressed differently in HCC tumor tissue compared to normal tissue; 173 of them were associated with OS, according to a univariate Cox analysis (Fig. 2A). The heatmap showed that 173 DDRGs showed differences in normal and HCC tissues, and almost all of them were highly expressed in HCC (Fig. 2B). As prognostic indications, 173 DDRGs were retained, and the risk ratio for the *RAD54B* gene was 6.846 (95% CI 3.467–13.518, $P < 0.001$, Fig. 2C). Figure 2D displays the association between these genes.

Building prognostic models in the TCGA cohort

To develop a prognostic model, LASSO-Cox regression analysis was applied to analyze the expression profiles of the 173 genes mentioned earlier. Based on the optimal value of λ , markers for 10 genes were identified (Supplementary Fig. S1). The DDRG signature was calculated as follows:

$$\begin{aligned} \text{RiskScore} = & 0.051 \times \text{CHD1L}_{\text{expression}} + 0.018 \times \text{HDAC1}_{\text{expression}} \\ & + 0.151 \times \text{KPNA2}_{\text{expression}} + 0.131 \times \text{MUTYH}_{\text{expression}} \\ & + 0.165 \times \text{NEIL3}_{\text{expression}} + 0.050 \times \text{POLR2L}_{\text{expression}} \\ & + 0.146 \times \text{PPP2R5B}_{\text{expression}} + 0.233 \times \text{RAD54B}_{\text{expression}} \\ & + 0.114 \times \text{RUVBL1}_{\text{expression}} + 0.032 \times \text{SPP1}_{\text{expression}} \end{aligned}$$

Based on median cutoff values, patients were divided into two groups (Fig. 3A). PCA and t-SNE analyses revealed that patients with high and low risk status were concentrated in two areas (Fig. 3E,F). Furthermore, when risk scores increased, HCC patients were more likely to die prematurely, and survival analysis further revealed that high-risk individuals had shorter OS (Fig. 3B,I). The accuracy of the model for predicting the prognosis of HCC was evaluated using time-dependent ROC curves, which yielded AUCs of 0.796, 0.724, and 0.700 for 1-, 2-,

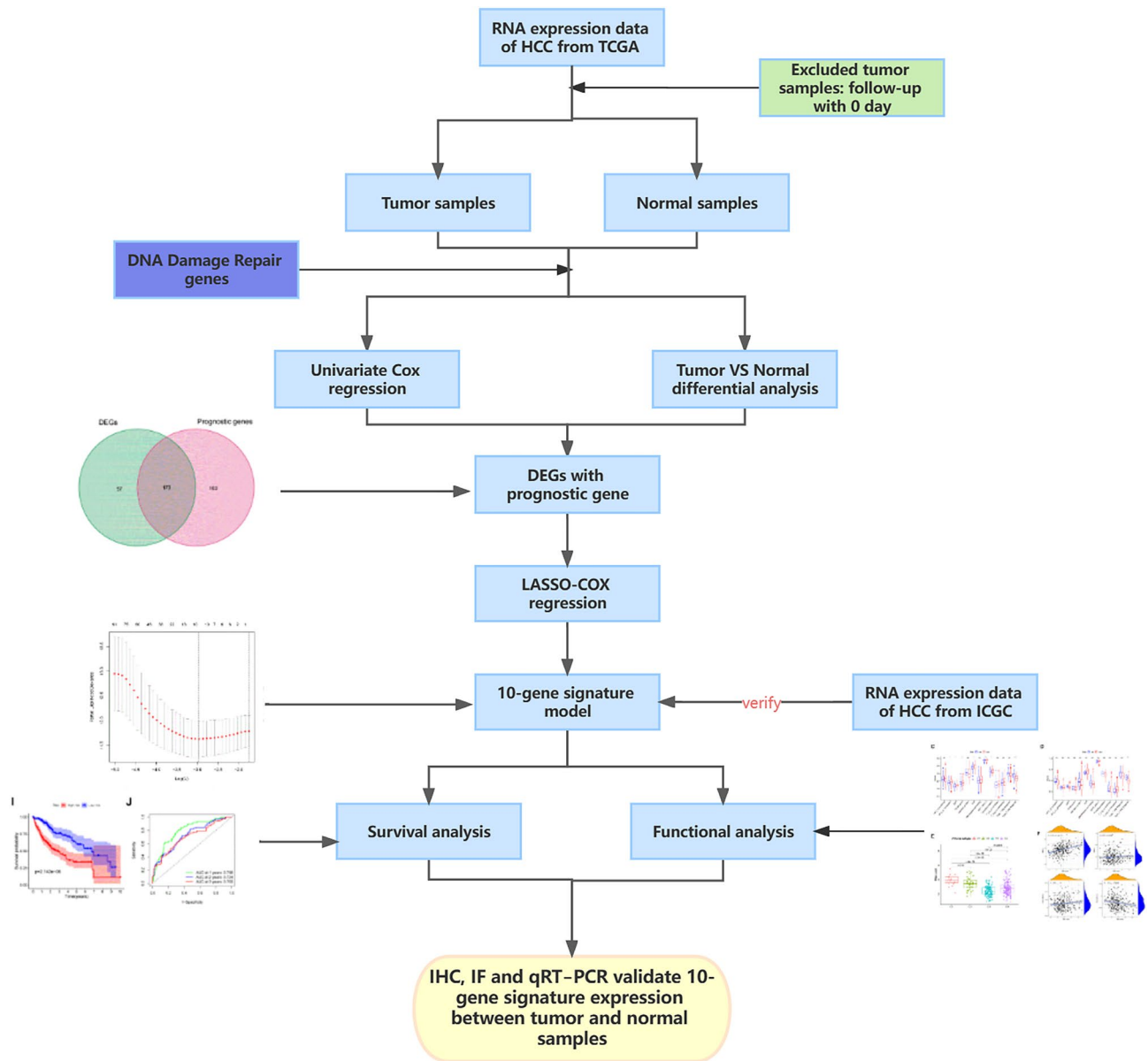


Figure 1. Flow chart of data collection, analysis, and experiments.

and 3-year predictions, respectively (Fig. 3J). To examine the correlation between the ten prognostic genes and prognosis, a prognostic survival study was conducted, which indicated a significant association between high expression of each gene and poor OS (Supplementary Fig. S2, all $P < 0.05$). Additionally, the expression levels of these genes were significantly higher in tumor tissues than in normal tissues (all $P < 0.001$) as demonstrated in Supplementary Fig. S3.

Validation of the 10-gene signature in the ICGC cohort

To test the stability of the model constructed from the TCGA cohort, patients in the ICGC cohort were also categorized into high- and low-risk groups based on the median value of the TCGA cohort, and our model also had strong predictive value. Fewer patients with low-risk than high-risk HCC in the ICGC cohort (Fig. 3C). The high-risk group had a lower survival duration and early mortality in the ICGC cohort (Fig. 3D,K). PCA and t-SNE analyses were conducted to categorize patients into different risk groups, which were divided into two orientations (Fig. 3G,H). In the ICGC queue, this model still had strong predictive ability (Fig. 3L).

Independent prognostic value of the 10-gene signature

Using univariate and multivariate Cox analyses, we identified the risk score as an independent prognostic factor for OS. Univariate Cox analysis revealed a significant relationship between risk scores and OS (TCGA cohort: HR = 4.587, 95% CI 3.066–6.861, $P < 0.001$; ICGC cohort: HR: 3.458, 95% CI 1.923–6.218, $P < 0.001$) (Fig. 4A,B). Multivariate Cox analysis showed that the risk score remained an independent predictor of OS after controlling

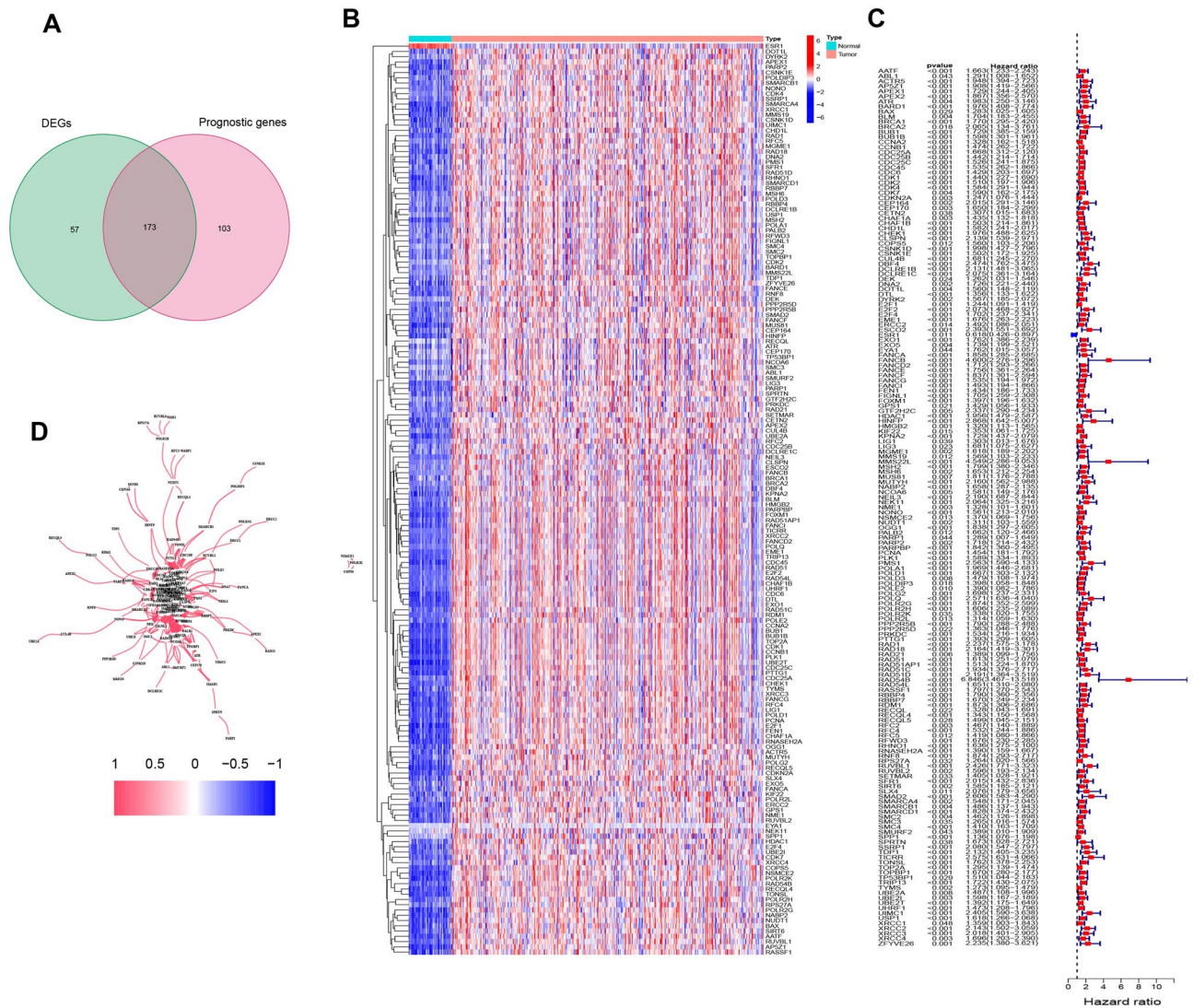


Figure 2. Identification of the candidate DDEGs in the TCGA cohort. **(A)** Venn diagram to identify DEGs between HCC tissues and adjacent normal tissues. **(B)** Expression of 173 overlapping genes between HCC tissues and adjacent normal tissues. **(C)** Forest plots showing the results of the correlation with 173 overlapping genes and OS. **(D)** The correlation network of candidate genes.

for other confounders (TCGA cohort: HR = 3.976, 95% CI 2.626–6.018, $P < 0.001$; ICGC cohort: HR: 2.455, 95% CI 1.343–4.488, $P = 0.004$) (Fig. 4C,D). Furthermore, we discovered that the stage of HCC was an independent predictor of OS (TCGA cohort: HR = 1.493, 95% CI 1.201–1.856, $P < 0.001$; ICGC cohort: HR: 1.938, 95% CI 1.338–2.805, $P < 0.001$). The risk score and clinicopathological characteristics therefore provided strong predictive value for HCC.

Risk scores and clinical characteristics of prognostic models

We found significantly higher risk scores for grades 3–4 compared to grades 1–2 in HCC (Supplementary Fig. S4D). In both the TCGA database and the ICGC database, risk scores were significantly higher for tumors in stages III–IV compared to tumors in stages I–II (as data about the grade of HCC were not available for the ICGC dataset) (Supplementary Fig. S4C,G). Furthermore, we obtained the relationship between the expression of prognostic genes and HCC staging and grading through the online website (<http://vip.sangerbox.com/home.html>). The results revealed that the expression of prognostic genes was significantly higher in tumor grades 3–4 than in tumor grades 1–2, except for *POLR2L* and *PPP2R5B* ($P < 0.05$, Supplementary Fig. S5). In terms of tumor stage, stage III or stage II had higher gene expression than stage I. Except for the genes *CHD1L* and *POLR2L*, there was no difference in the genes in each stage (stage IV was not compared because of the small sample size) ($P < 0.05$, Supplementary Fig. S6).

Analysis of immune status and tumor microenvironment

We comprehensively analyzed the relationship between risk scores and immune status. In the TCGA cohort, the high-risk group had significantly higher levels of components associated with the antigen presentation

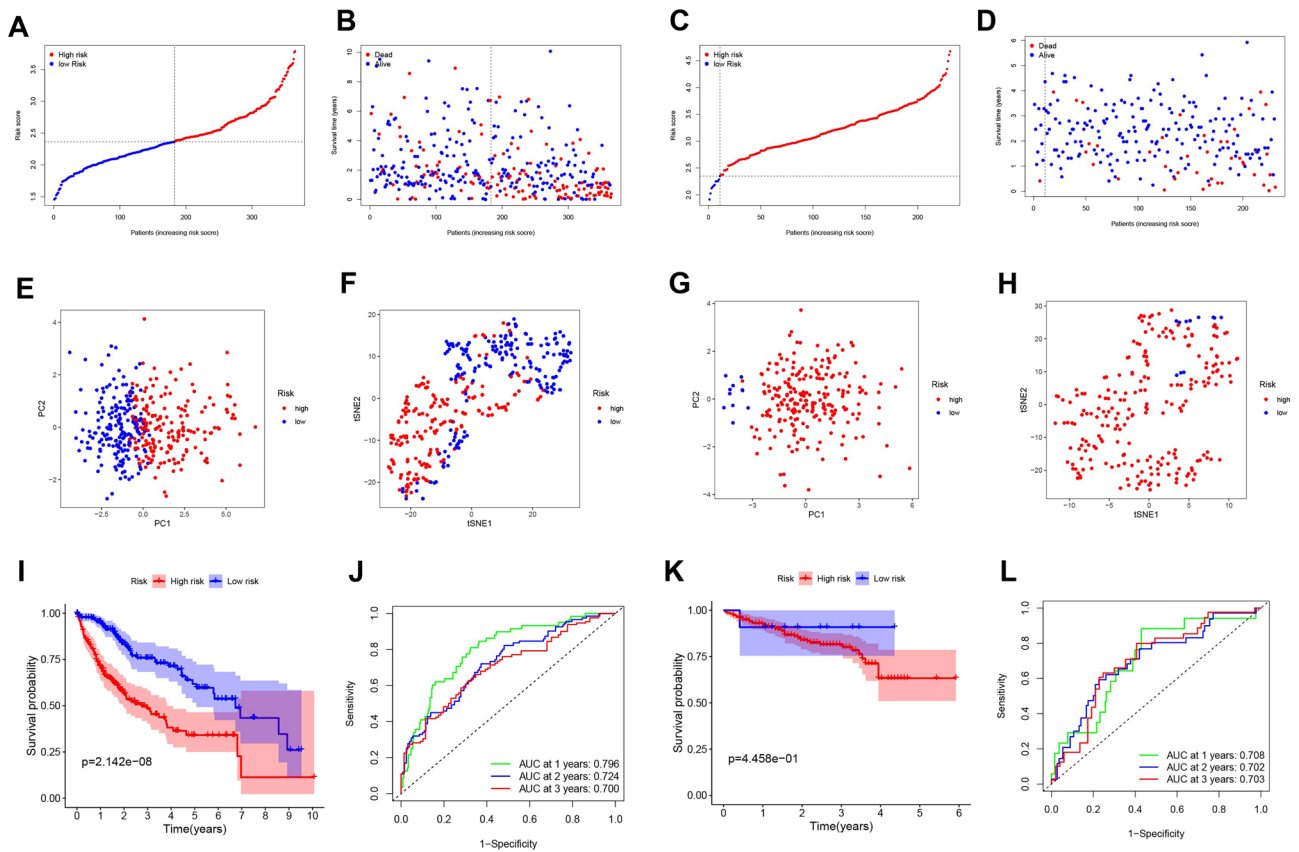


Figure 3. The prognosis analysis of the 10-gene signature model in the TCGA cohort and the ICGC cohort. TCGA cohort (A,B,E,F,I,J), ICGC cohort (C,D,G,H,K,L). (A,C) The median value and distribution of the risk scores. (B,D) The distribution of OS status. (E,G) PCA plot. (F,H) t-SNE analysis. (I,K) Kaplan–Meier curves for OS of patients in the high- and low-risk groups. (J,L) AUC time-independent ROC curves for OS.

pathway, including aDCs, APC costimulation, HLA, and MHC class I (all adjusted P values < 0.05, as shown in Fig. 5A,C). This finding suggests a potential association between increased risk scores and altered immune activity, specifically within the antigen presentation pathway. Compared to those in the low-risk group, individuals in the high-risk group exhibited increased fractions of Tfh cells, Treg cells, Th 1 cells and Th 2 cells indicating alterations in T-cell regulation between the two groups. Furthermore, the high-risk group exhibited higher CCR, and checkpoint, macrophage activity scores, but the reverse activity was found for the type II IFN response score (adjusted $P < 0.05$). The outcomes of the comparisons between the two risk categories in the ICGC cohort were comparable to those in the TCGA cohort (adjusted $P < 0.05$, Fig. 5B,D).

We next examined the correlation between risk scores and immune infiltration. Six immune infiltration subtypes associated with human tumor promotion and tumor suppression were identified: C1 (wound healing), C2 (INF-g dominant), C3 (inflammatory), C4 (lymphocyte depleted), C5 (immunologically quiet) and C6 (TGF- β dominant)^{23,24}. Two immune categories—C5 and C6—were excluded from this study due to the limited number of HCC samples available for these subtypes. Our investigation revealed a significant correlation between risk scores and immune infiltration of HCC in the TCGA database, with high-risk scores being strongly associated with the C1 subtype and low-risk scores demonstrating a significant link with the C3 subtype (Fig. 5E).

Neoplastic stemness may be assessed using DNA stemness scores (DNAss) and RNA stemness scores (RNAss) based on DNA methylation patterns and mRNA expression, respectively²⁵. Correlation analysis was then conducted to examine whether risk scores were associated with tumor stem cells and the immune microenvironment. The findings revealed that risk scores were not significantly correlated with DNAss or immune scores but instead showed a positive and significant correlation with RNAss and a negative correlation with stromal scores ($P < 0.001$, Fig. 5F).

Analyses of immune-related genes

An important regulator of cancer immune evasion is the PD-1/PD-L1 pathway. The expression levels of immunological checkpoints such as PD-1 and PD-L1 are crucial markers for personalized immunotherapy. As expected, the expression levels of PD-1, PD-L1, CTLA4 and LAG3 were found to be significantly higher in the high-risk group than in the low-risk group, with statistically significant differences observed (Fig. 6A–D). Furthermore, our analysis showed that the expression levels of these immune checkpoints were positively correlated with the risk score (Fig. 6E–H).

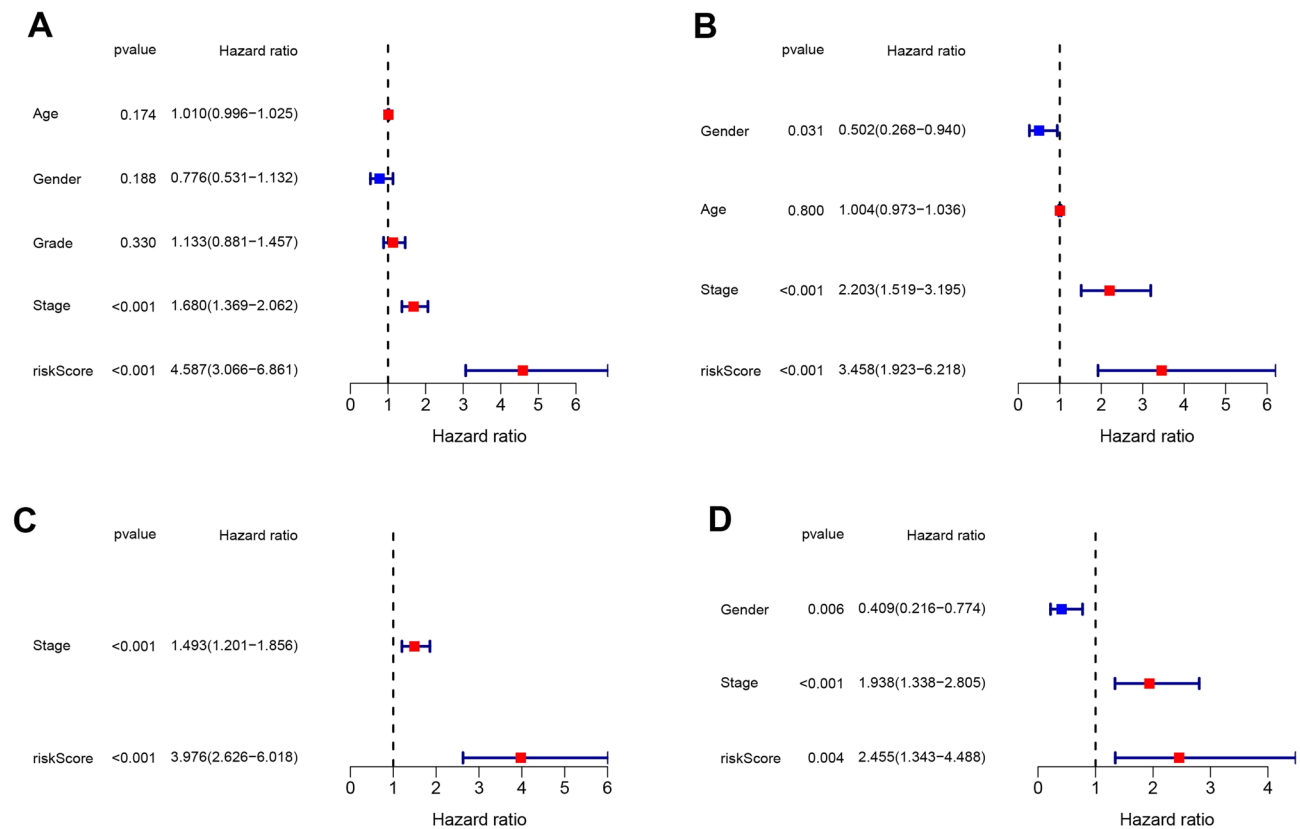


Figure 4. Independence validation of the risk score in the TCGA and ICGA cohorts. (A,B) TCGA cohorts, (C,D) ICGC cohorts.

Analyses of biological function and pathway enrichment

KEGG functional enrichment studies were performed utilizing DEGs in high and low risk groups to clarify the underlying pathways that were related to the risk score. DEGs were shown to be abundant in the cell cycle, phagosome, DNA replication, human T-cell leukemia virus 1 infection, and HIF-1 signaling pathways (Fig. 7D, Supplementary Table S4). GO functional enrichment studies were also carried out. The results demonstrated that the DEGs virtually mapped to immune-related GO keywords such as antigen binding, organelle fission, chromosomal region, and tubulin binding (Fig. 7A–C, Supplementary Table S5), indicating that changes in survival across subgroups may be connected to the immunological state of patients.

Prognostic gene expression and cancer cell sensitivity to chemotherapy

We investigated the association between prognostic gene expression levels and drug sensitivity in NCI-60 cell lines. Our analysis revealed that each prognostic gene was significantly associated with specific chemotherapeutic drug sensitivities ($P < 0.01$) (Fig. 8, Supplementary Table S6). For instance, higher levels of *CHD1L*, *HDAC1*, *MUTYH*, *NEIL3*, *POLR2L*, *RUVBL1*, and *SPPI* expression in cancer cells have been linked to higher levels of drug resistance to nelarabine, acridine, chlorambucil, dexrazoxane, cladribine, and other drugs. On the other hand, elevated *PPP2R5B* expression has been found to be associated with increased drug sensitivity of cancer cells to several chemotherapeutic agents, including oxaliplatin.

Validation of prognostic gene expression in HCC and normal tissues

According to the HPA database, the protein level of *MUTYH* was higher in normal tissues than in tumor tissues, and the protein expression levels of the seven genes (data for *NEIL3* and *POLR2L* were lacking and therefore not presented) were significantly higher in tumor samples than in normal samples (Fig. 9). Additionally, we further verified these 10 prognostic genes by IHC, IF, and qRT-PCR to confirm their expression in the protein and mRNA of hepatocellular carcinoma and normal liver tissue. The results of IHC showed that except for the expression of *NEIL3*, which was not statistically significant in normal liver tissues versus HCC tissues, the other 9 genes matched our bioinformatic predictions, and they were significantly overexpressed in HCC tissues compared to normal liver tissues (Fig. 10). As shown in Fig. 11, the expression of all 10 genes according to IF was significantly higher in HCC tissues than in normal liver tissues, which is consistent with our bioinformatics results. We hypothesized that the sample size might be too small, resulting in the expression of *NEIL3* not being statistically significant according to IHC. Therefore, to increase the credibility of the results, we constructed a mouse HCC model to further verify whether there is a difference in the expression of these 10 genes at the RNA level in mouse HCC tissues and normal mouse liver tissues by qRT-PCR. As shown in Fig. 12, at the mRNA level,

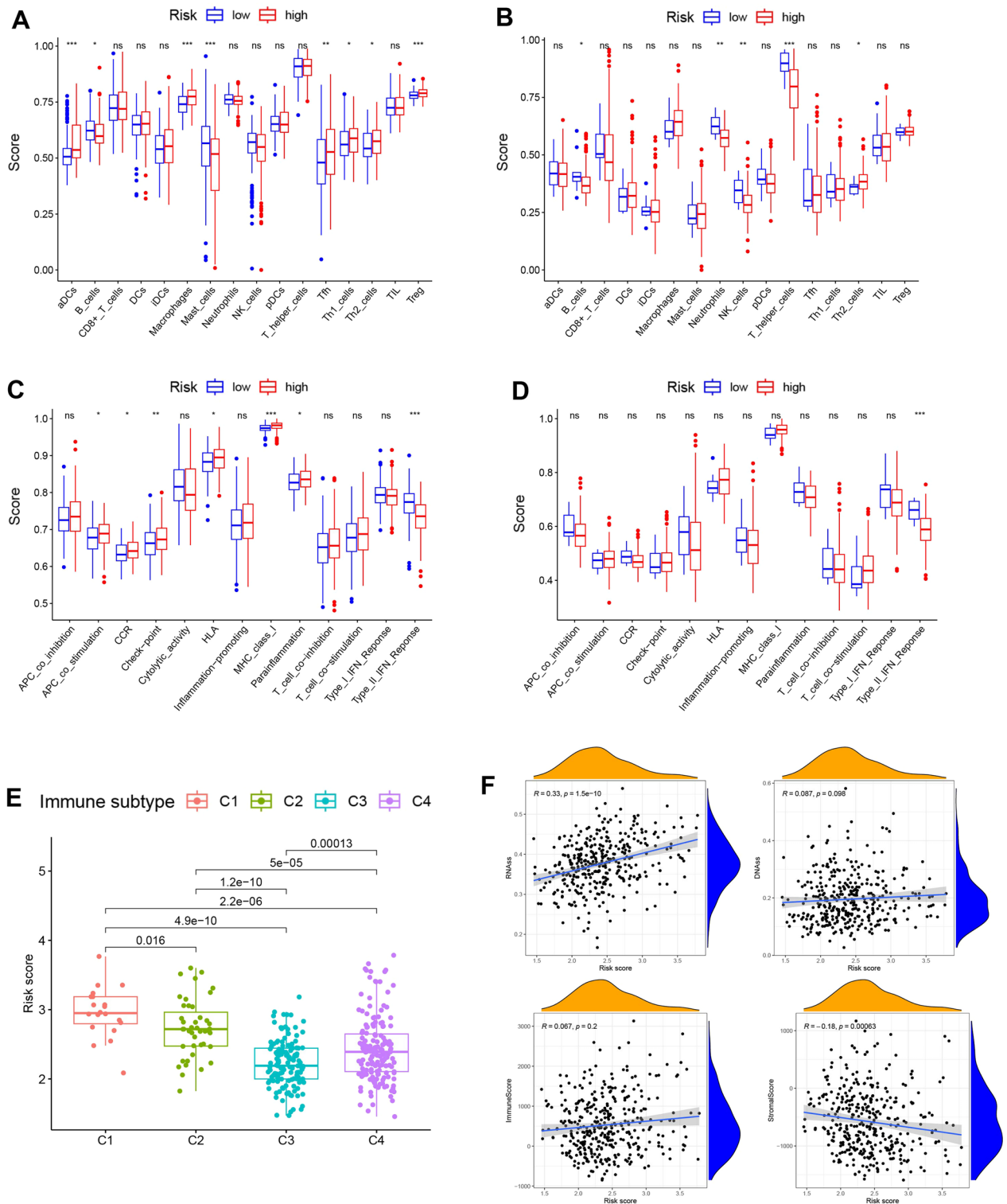


Figure 5. Relationship between immune status and risk scores and the tumor microenvironment between different risk groups. (A,C) TCGA cohort, (B,D) ICGC cohort. (A,B) The scores of 16 immune cells and (C,D) 13 immune-related functions are shown in boxplots. (E) Comparison of the risk score in different immune infiltration subtypes. (F) The association between risk score and RNAss, DNAss, stromal score and immune score. P values are shown as follows: *ns* not significant; **P* < 0.05; ***P* < 0.01, ****P* < 0.001.

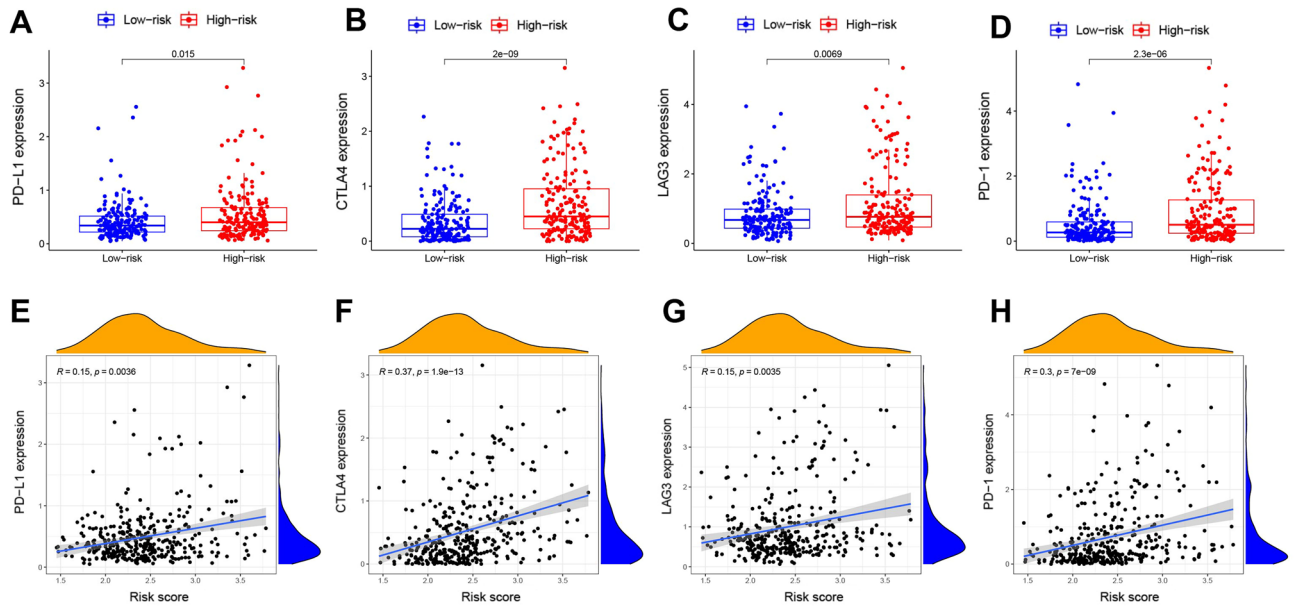


Figure 6. Comparison of the expression levels of PD-L1, CTLA4, LAG3 and PD-1 between different risk groups and correlation analysis between the risk score and the expression levels of PD-L1, CTLA4, LAG3 and PD-1. (A,E) PD-L1. (B,F) CTLA4. (C,G) LAG3. (D,H) PD-1.

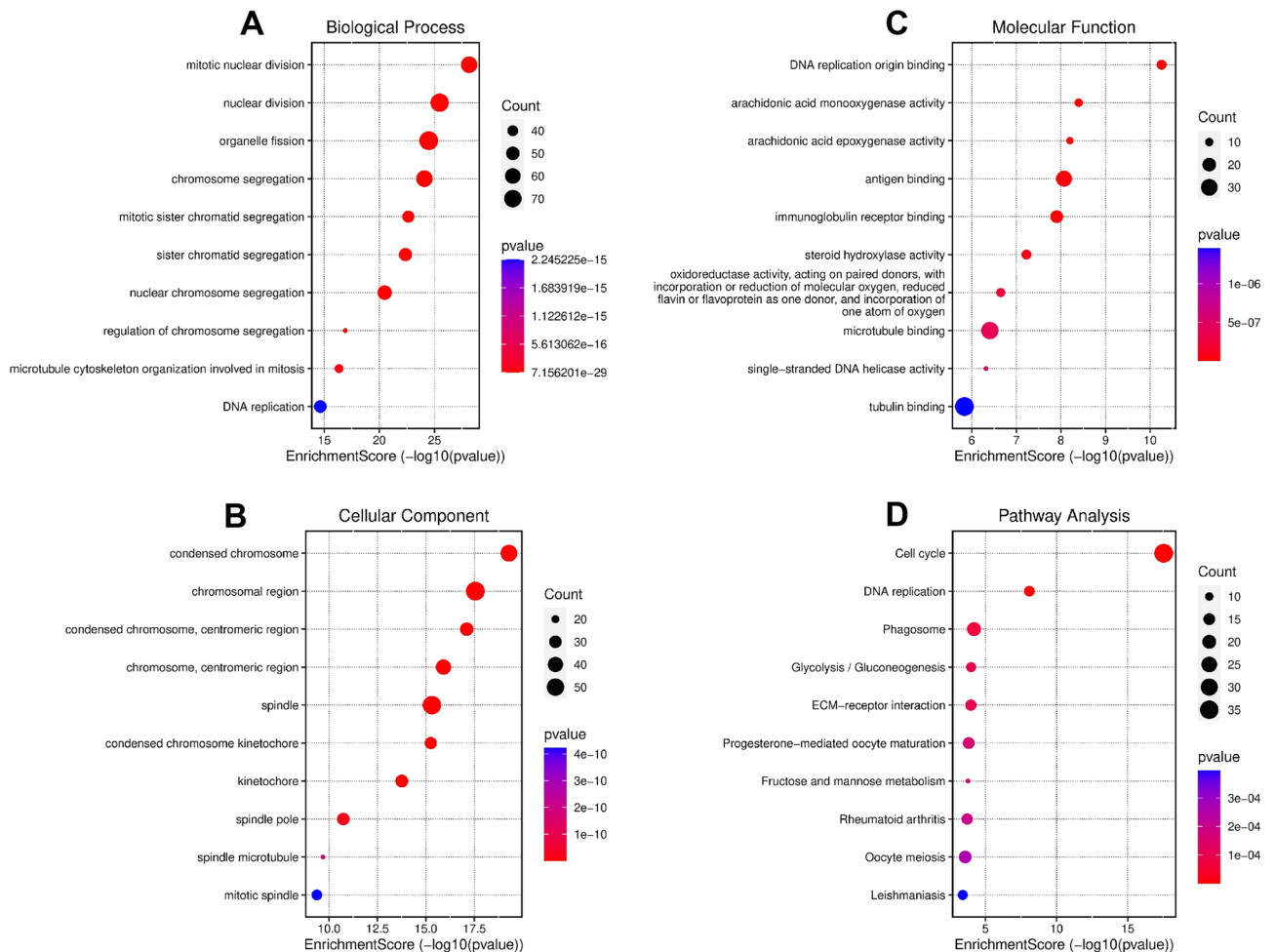


Figure 7. Analysis of Gene Ontology (A–C) and KEGG pathway enrichment (D) in the DEGs in the TCGA database.

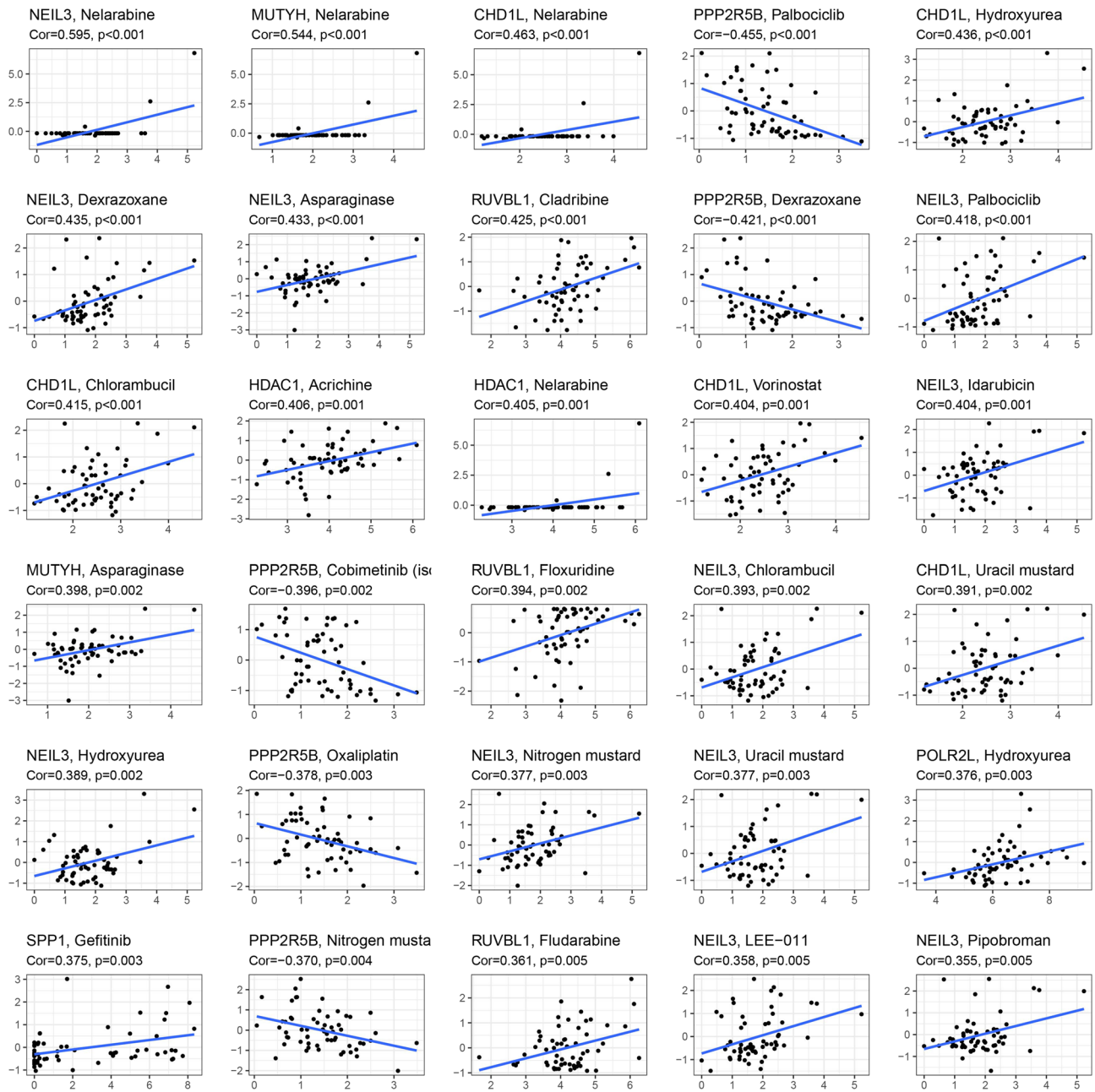


Figure 8. Scatter plot of the relationship between prognostic gene expression and drug sensitivity.

these 10 genes were significantly expressed in HCC tissues and expressed at lower levels in normal liver tissues. These observations further confirm our conclusion that these genes are of high clinical significance as potential biomarkers for the diagnosis and prognosis of HCC.

Discussion

With the establishment of next-generation sequencing technologies and the era of precision medicine approaches, a variety of treatments for HCC have been introduced. However, the small number of useful biomarkers often prevents the early diagnosis of HCC and prediction of the outcome of treatment. Several studies have demonstrated that novel serum biomarkers, such as a combination of circulating tumor cells, circulating nucleic acids and retinol and retinal panels, exhibit excellent prognostic accuracy in HCC^{26–28}. Additionally, if DNA damage is not properly repaired and continues to accumulate, it might cause hepatocytes to change malignantly and ultimately result in HCC²⁹. Therefore, DNA damage and repair (DDR) is a crucial molecular process for the emergence and progression of HCC, and more research on this process will establish the groundwork for an all-encompassing approach to treating HCC. However, the use of a gene profile associated with DDR as a prognostic indicator for HCC has not been documented. Our gene signature has higher accuracy in predicting prognosis, earlier studies have shown that gene signatures related to inflammatory, ferroptosis, immunity, m6A

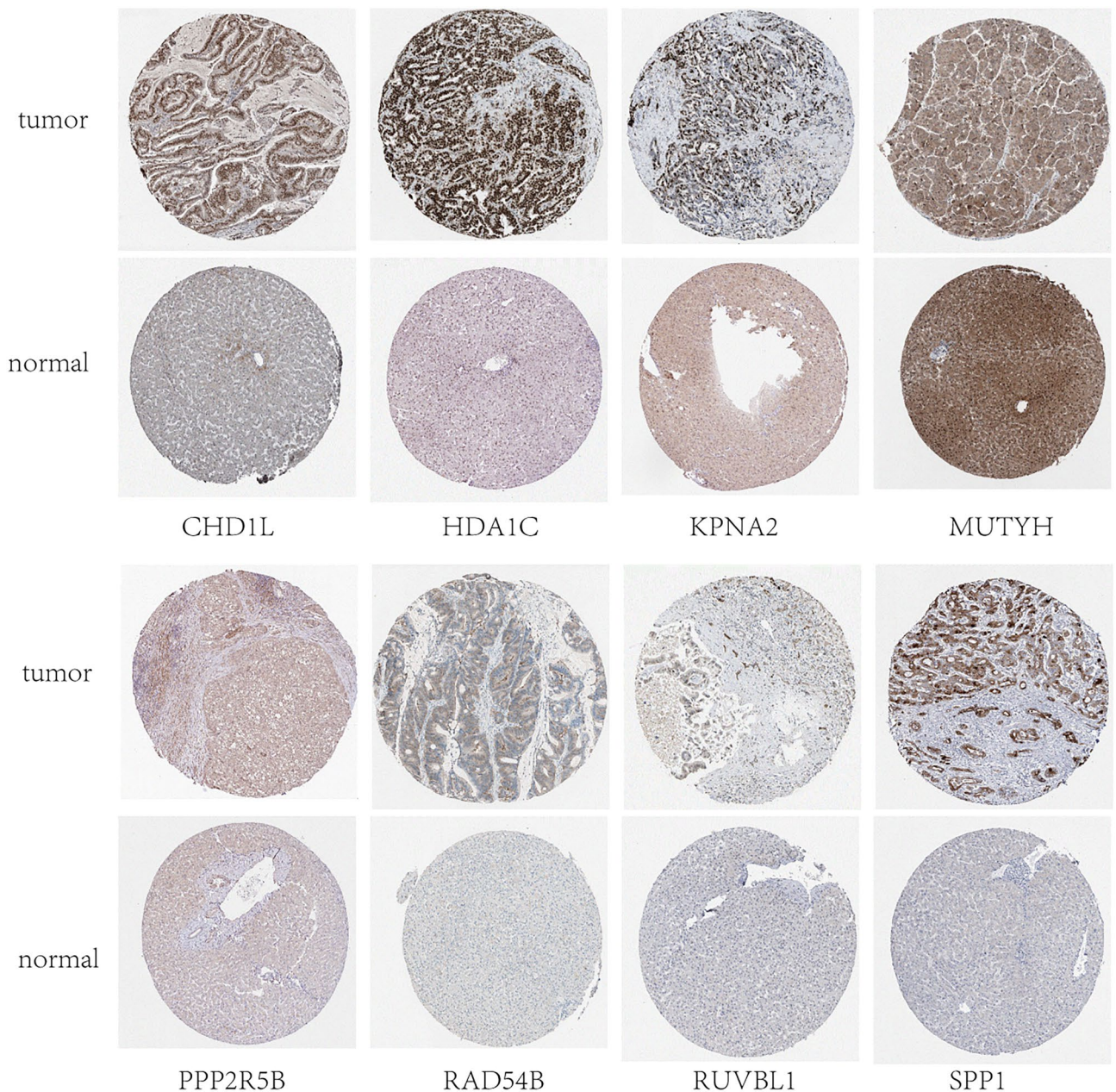


Figure 9. Protein expression of the eight genes in normal and HCC tissues based on Human Protein Atlas immunohistochemistry.

and energy metabolism in predicting 3-year OS in HCC only the immune-related gene signature had an AUC of > 0.70 in the experimental group, and the AUC of our DDRG signature was greater than 0.70 in both experimental and validation groups^{28,30–33}. The DDRG signature created in our work exhibits greater benefits than the gene signature above, in addition to having high predictive performance for HCC prognosis. The algorithm is capable of categorizing immune checkpoint genes and tumor resistance genes into high- and low-expression groups. Additionally, several studies have demonstrated a significant correlation between risk scores and drug resistance to various chemotherapeutic agents.

From the TCGA cohort, we identified DEGs through a screening process. Further analysis using univariate Cox regression revealed 173 DEGs that were significantly associated with OS. By utilizing LASSO regression analysis, we developed a prognostic model that incorporated 10 DDRGs. We then evaluated this model using the ICGC cohort. Our analysis revealed that the high-risk group was significantly associated with a shorter OS, advanced TNM stage, and higher tumor grade. Moreover, independent prognostic analysis confirmed that the risk score was an independent predictor of OS, this is similar to previous findings that high repair groups tend to be accompanied by a worse prognosis¹⁵.

The prognostic model established in this study consisted of 10 DDRGs (*CHD1L*, *HDAC1*, *KPNA2*, *MUTYH*, *PPP2R5B*, *NEIL3*, *POLR2L*, *RAD54B*, *RUVBL1* and *SPP1*). A poor prognosis was nearly always linked to these

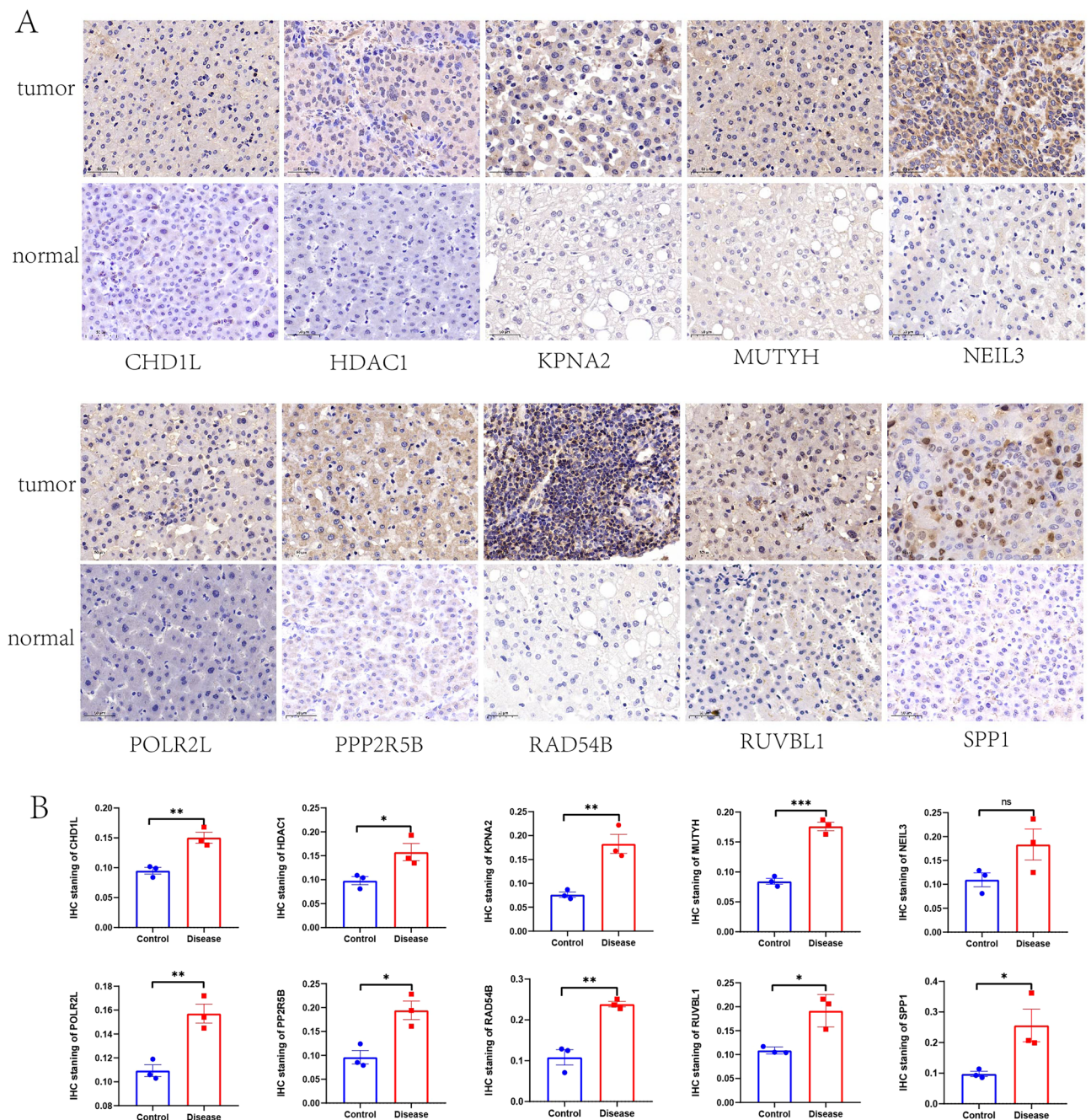


Figure 10. Protein expression levels of 10 prognostic genes in human HCC and normal liver tissues measured by IHC (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, NS not significant).

genes. *CHD1L* is a newly discovered oncogene that has been linked to cancer via apoptosis inhibition, G1/S transition, and uncontrolled cell proliferation³⁴. *HDAC1*-mediated inhibition of hepatocyte markers is a critical stage in the formation of hepatoblastoma, laying the groundwork for the development of treatments for aggressive hepatoblastoma by inhibiting *HDAC1* activity³⁵. *KPNA2*, a nuclear transporter family member, has been recently shown to promote tumor growth and progression by involvement in cell differentiation, proliferation, apoptosis, immunological response, and viral infection³⁶. *MUTYH* has been shown to be possibly associated with hepatocarcinogenesis in nonalcoholic steatohepatitis in a mouse model³⁷. The role of *PPP2R5B* in liver cancer remains unclear. Some data suggest that deleting the gene causes paclitaxel sensitivity in cervical cancer and that this sensitivity shift is related to apoptosis³⁸. Hui-huang Lai et al. demonstrated that *NEIL3* activated the BRAF/MEK/ERK/TWIST pathway-mediated EMT and therapeutic resistances, leading to HCC progression³⁹. Single-cell RNA sequencing has revealed that *POLR2L* may contribute to the development of HCC through cell cycle-related pathways⁴⁰. *RAD54B* has been investigated as a predictive marker for HCC patients, and it may play a significant role in the development of HCC through cancer cell DNA amplification⁴¹. *RUVBL1* is an AAA + ATPase whose expression is associated with a poor prognosis in HCC. On the other hand, the involvement of *RUVBL1* in the

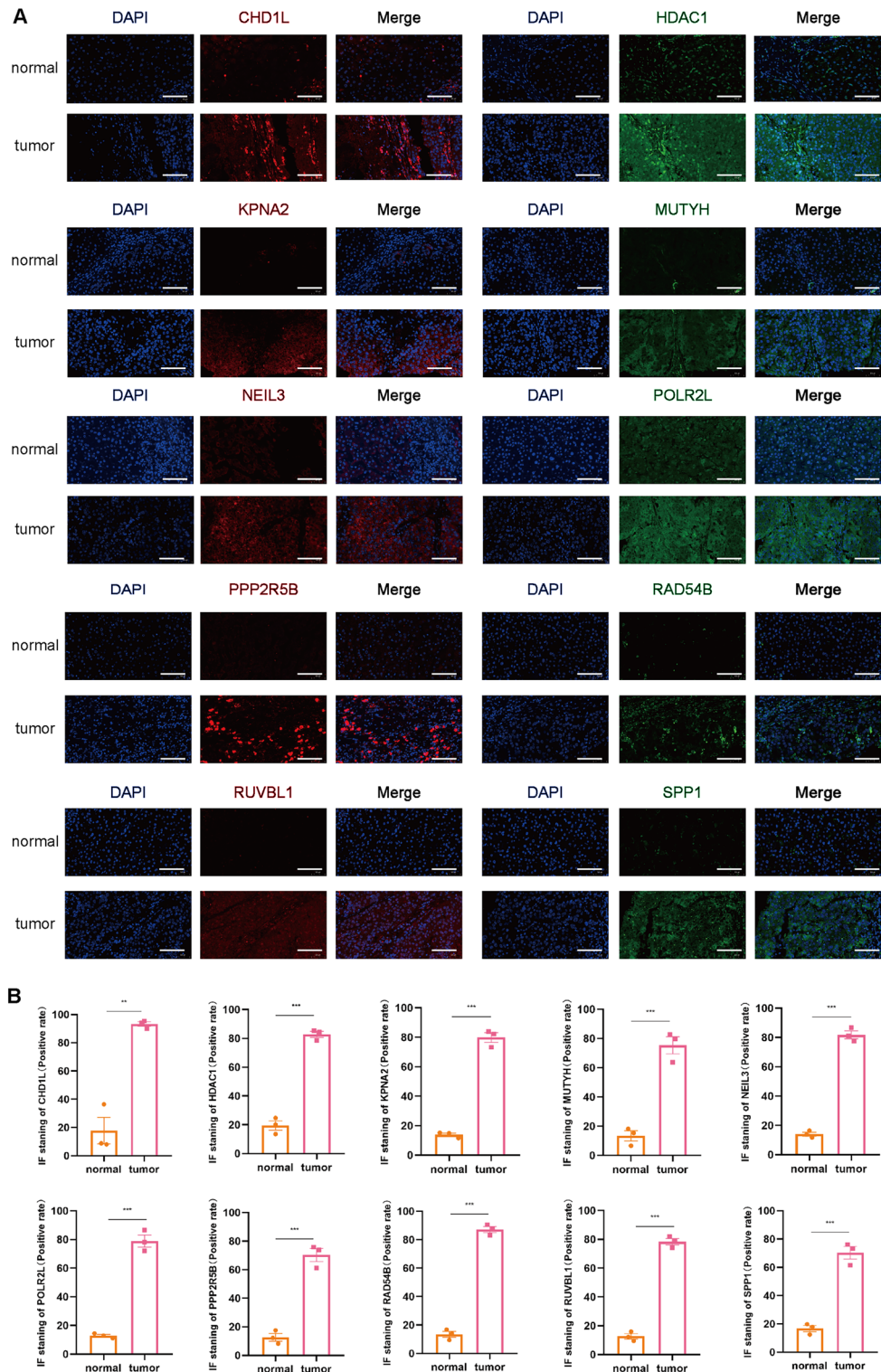


Figure 11. Protein expression levels of 10 prognostic genes in human HCC and normal liver tissues measured by IF (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$).

initial stages and growth of HCC is uncertain⁴². *SPP1*, which is a member of the SIBLING family, has been shown in some research to be overexpressed in numerous malignancies and can serve as a prognostic indicator⁴³.

To gain insight into the interaction between the risk score and immune components, we examined the function of the risk score in the immune infiltration type. Our findings revealed that high risk scores were significantly associated with C1, whereas low risk scores were clearly associated with C3. These results suggest that C1 promotes tumorigenesis and progression, while C3 functions as a protective factor. This finding demonstrated

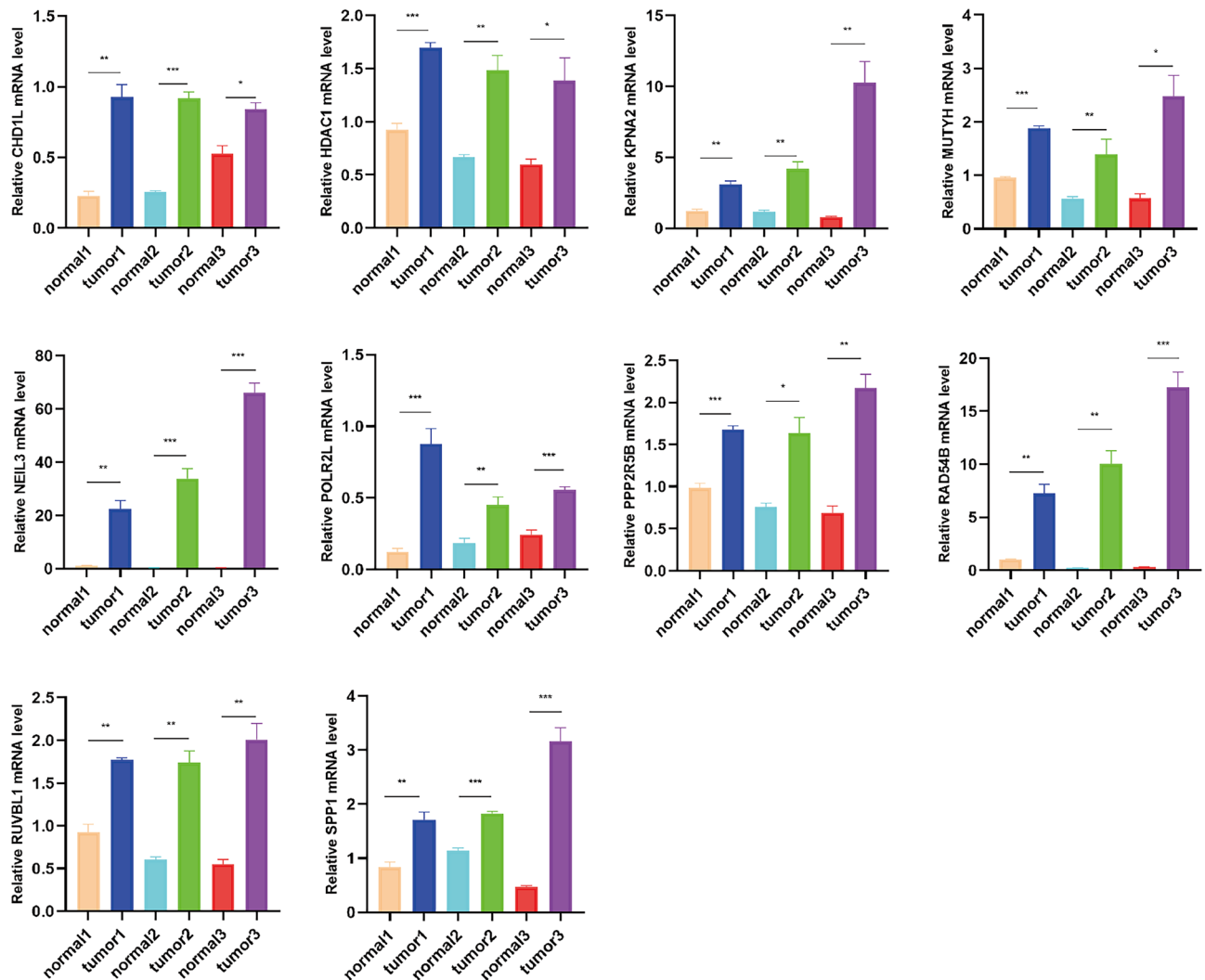


Figure 12. mRNA expression levels of 10 prognostic genes in mouse HCC and normal mouse liver tissues measured by qRT-PCR (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$).

that increased cytotoxicity might prevent tumor onset and growth⁴⁴. Regarding the association between risk scores and clinical features, there was a significant correlation between high-risk scores and tumor grade 3–4 or tumor stage III–IV. These findings suggest that high risk scores are strongly associated with poor prognosis.

Despite the findings linking these genes to DNA damage repair, further research is necessary to determine their impact on the prognosis of HCC patients. Additionally, tumor-related signaling pathways, such as the cell cycle, phagosome, antigen binding, and organelle fission, were significantly enriched in the GO and KEGG analyses. The constant activation of these pathways has been associated with HCC, indicating potential novel treatment targets^{45,46}. Furthermore, the high-risk group exhibited a higher number of macrophages and Treg cells. Previous research has revealed that an increase in tumor-associated macrophages and Treg cells is associated with poor prognosis in patients with HCC, likely due to their role in immune invasion^{47–49}. Immunotherapy for cancer targeting immune checkpoints (e.g., anti-PD-L1 antibodies) has shown clinical activity in a variety of cancer types⁵⁰. Immunotherapy based on immune checkpoint inhibitors has shown significant progress in the treatment of HCC. The expression of PD-1 and CTLA4 receptors inhibits the antitumor immunological response of T cells. As a result, the tumor can evade the body's natural defenses and continue to grow. However, the use of immune checkpoint inhibitors has demonstrated significant therapeutic benefits in prolonging survival and improving quality of life for HCC patients⁵¹. Our study uncovered a strong correlation between the risk score and the expression of both PD-L1 and PD-1, whereby markedly higher scores were evident in the high-risk group than in the low-risk group. Our prognostic model therefore holds significant potential in facilitating accurate predictions of immune checkpoint expression levels and aiding in treatment decision-making. Furthermore, we found a distinct correlation between a high-risk score and a marked decrease in type II interferon (IFN) response activity. High levels of type II IFN response activity are central to tumor immune surveillance, as they stimulate antitumor immunity and support tumor elimination^{44,52,53}. Additionally, our study identified increases in T-follicular helper (Tfh) cells, regulatory T (Treg) cells, T-helper 1 (Th1) cells, T-helper 2 (Th2) cells, and both T-cell costimulatory and T-cell coinhibitory activities within the high-risk group, which all point to a disruption

in immune regulation within this cohort. Thus, we postulated that the diminished antitumor immunity was likely responsible for the poor prognosis found in the high-risk group.

Using NCI-60 cell line data, numerous prognostic genes were found to have increased expression associated with increased resistance to many FDA-approved chemotherapeutic agents. Of course, various prognostic genes were also correlated with increased sensitivity to a few drugs. In fact, immuno-oncology preclinical work has shown that combination immunotherapies including LAG-3 and CTLA4 blockade may have a synergistic effect on the anticancer immune response⁵⁴. Based on these findings, several prognostic genes could serve as potential therapeutic targets to mitigate medication resistance or enhance susceptibility to adjuvant therapy.

In conclusion, the 10 DDRGs related with HCC prognosis that we analysed using the LASSO method offer fresh perspectives on how to treat HCC on an individual basis. We further confirmed by IHC, IF, and qRT-PCR that these 10 genes were differentially expressed in HCC tissues and normal tissues both at the protein level and mRNA level. All of them were highly expressed, and the experimentally verified expression trends were consistent with the findings of the bioinformatics analysis, which strengthened the validity of our findings. Furthermore, only a limited amount of research has been done on these 10 genes in HCC, and through our experiments, we have discovered that some of these genes are highly expressed in HCC for the first time. As a result, our prognostic model and experimental findings can offer fresh perspectives and a theoretical framework for future studies on HCC.

Conclusion

In summary, our research identified a novel predictive signature comprising ten DDRGs. This signature was independently associated with OS and demonstrated utility in functional analysis, analysis of the tumor microenvironment, and prediction of medication sensitivity, providing novel insights for prognostic prediction in HCC. In addition, we verified by IHC, IF and qRT-PCR that the protein and mRNA expression of these ten DDRGs were all high in HCC and low in normal liver tissues. The particular underlying pathways connecting DDRGs and tumor immunity in HCC remain unknown, and our findings provide new directions for further investigation.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 31 July 2023; Accepted: 26 October 2023

Published online: 03 November 2023

References

- Sung, H. *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249. <https://doi.org/10.3322/caac.21660> (2021).
- Villanueva, A. Hepatocellular carcinoma. *N. Engl. J. Med.* **380**, 1450–1462. <https://doi.org/10.1056/NEJMra1713263> (2019).
- Yang, J. D. *et al.* A global view of hepatocellular carcinoma: Trends, risk, prevention and management. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 589–604. <https://doi.org/10.1038/s41575-019-0186-y> (2019).
- Allemani, C. *et al.* Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): Analysis of individual records for 7513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet* **391**, 1023–1075. [https://doi.org/10.1016/S0140-6736\(17\)33326-3](https://doi.org/10.1016/S0140-6736(17)33326-3) (2018).
- Yang, S. Q., Lin, H. Z., Mandal, A. K., Huang, J. & Diehl, A. M. Disrupted signaling and inhibited regeneration in obese mice with fatty livers: Implications for nonalcoholic fatty liver disease pathophysiology. *Hepatology* **34**, 694–706. <https://doi.org/10.1053/jhep.2001.28054> (2001).
- Riehle, K. J. *et al.* Regulation of liver regeneration and hepatocarcinogenesis by suppressor of cytokine signaling 3. *J. Exp. Med.* **205**, 91–103. <https://doi.org/10.1084/jem.20070820> (2008).
- Forbes, S. J. & Newsome, P. N. Liver regeneration—Mechanisms and models to clinical application. *Nat. Rev. Gastroenterol. Hepatol.* **13**, 473–485. <https://doi.org/10.1038/nrgastro.2016.97> (2016).
- Caldez, M. J., Bjorklund, M. & Kaldis, P. Cell cycle regulation in NAFLD: When imbalanced metabolism limits cell division. *Hepatol. Int.* **14**, 463–474. <https://doi.org/10.1007/s12072-020-10066-6> (2020).
- Yang, S. F. *et al.* Involvement of DNA damage response pathways in hepatocellular carcinoma. *Biomed Res. Int.* **2014**, 153867. <https://doi.org/10.1155/2014/153867> (2014).
- Fortini, P., Ferretti, C. & Dogliotti, E. The response to DNA damage during differentiation: Pathways and consequences. *Mutat. Res.* **743–744**, 160–168. <https://doi.org/10.1016/j.mrfmmm.2013.03.004> (2013).
- Lee, Y. H., Kuo, C. Y., Stark, J. M., Shih, H. M. & Ann, D. K. HP1 promotes tumor suppressor BRCA1 functions during the DNA damage response. *Nucleic Acids Res.* **41**, 5784–5798. <https://doi.org/10.1093/nar/gkt231> (2013).
- Chung, Y. L. & Wu, M. L. Promyelocytic leukaemia protein links DNA damage response and repair to hepatitis B virus-related hepatocarcinogenesis. *J. Pathol.* **230**, 377–387. <https://doi.org/10.1002/path.4195> (2013).
- Chang, H. H. Y., Pannunzio, N. R., Adachi, N. & Lieber, M. R. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell Biol.* **18**, 495–506. <https://doi.org/10.1038/nrm.2017.48> (2017).
- Huo, J., Fan, X., Qi, B. & Sun, P. A five-gene signature associated with DNA damage repair molecular subtype predict overall survival for hepatocellular carcinoma. *Front. Genet.* **13**, 771819. <https://doi.org/10.3389/fgene.2022.771819> (2022).
- Smith, M. A. *et al.* DNA damage repair classifier defines distinct groups in hepatocellular carcinoma. *Cancers* <https://doi.org/10.3390/cancers14174282> (2022).
- Yang, X., Wang, G., Gu, R., Xu, X. & Zhu, G. A signature of tumor DNA repair genes associated with the prognosis of surgically-resected lung adenocarcinoma. *PeerJ* **8**, e10418. <https://doi.org/10.7717/peerj.10418> (2020).
- Simon, N., Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for Cox's proportional hazards model via coordinate descent. *J. Stat. Softw.* **39**, 1–13. <https://doi.org/10.18637/jss.v039.i05> (2011).
- Tibshirani, R. The lasso method for variable selection in the Cox model. *Stat. Med.* **16**, 385–395. [https://doi.org/10.1002/\(sici\)1097-0258\(19970228\)16:4%3c385::aid-sim380%3e3.0.co;2-3](https://doi.org/10.1002/(sici)1097-0258(19970228)16:4%3c385::aid-sim380%3e3.0.co;2-3) (1997).
- Simon, N., Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for Cox's proportional hazards model via coordinate descent. *J. Stat. Softw.* <https://doi.org/10.18637/jss.v039.i05> (2011).

20. Yoshihara, K. *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612. <https://doi.org/10.1038/ncomms3612> (2013).
21. Dib, L., San-Jose, L. M., Ducrest, A. L., Salamin, N. & Roulin, A. Selection on the major color gene melanocortin-1-receptor shaped the evolution of the melanocortin system genes. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms18122618> (2017).
22. Thul, P. J. & Lindskog, C. The human protein atlas: A spatial map of the human proteome. *Protein Sci.* **27**, 233–244. <https://doi.org/10.1002/pro.3307> (2018).
23. Tamborero, D. *et al.* A pan-cancer landscape of interactions between solid tumors and infiltrating immune cell populations. *Clin. Cancer Res.* **24**, 3717–3728. <https://doi.org/10.1158/1078-0432.CCR-17-3509> (2018).
24. Desgrosellier, J. S. & Cheresch, D. A. Integrins in cancer: Biological implications and therapeutic opportunities. *Nat. Rev. Cancer* **10**, 9–22. <https://doi.org/10.1038/nrc2748> (2010).
25. Malta, T. M. *et al.* Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell* **173**, 338–354 e315. <https://doi.org/10.1016/j.cell.2018.03.034> (2018).
26. Trevisan Franca de Lima, L. *et al.* The use of minimally invasive biomarkers for the diagnosis and prognosis of hepatocellular carcinoma. *Biochim. Biophys. Acta Rev. Cancer* **1874**, 188451. <https://doi.org/10.1016/j.bbcan.2020.188451> (2020).
27. Han, J. *et al.* Tissue and serum metabolomic phenotyping for diagnosis and prognosis of hepatocellular carcinoma. *Int. J. Cancer* **146**, 1741–1753. <https://doi.org/10.1002/ijc.32599> (2020).
28. Lin, Z., Xu, Q., Miao, D. & Yu, F. An inflammatory response-related gene signature can impact the immune status and predict the prognosis of hepatocellular carcinoma. *Front. Oncol.* **11**, 644416. <https://doi.org/10.3389/fonc.2021.644416> (2021).
29. Gillman, R., Lopes Floro, K., Wankell, M. & Hebbard, L. The role of DNA damage and repair in liver cancer. *Biochim. Biophys. Acta Rev. Cancer* **1875**, 188493. <https://doi.org/10.1016/j.bbcan.2020.188493> (2021).
30. Liang, J. Y. *et al.* A novel ferroptosis-related gene signature for overall survival prediction in patients with hepatocellular carcinoma. *Int. J. Biol. Sci.* **16**, 2430–2441. <https://doi.org/10.7150/ijbs.45050> (2020).
31. Dai, Y. *et al.* An immune-related gene signature for predicting survival and immunotherapy efficacy in hepatocellular carcinoma. *Cancer Immunol. Immunother.* **70**, 967–979. <https://doi.org/10.1007/s00262-020-02743-0> (2021).
32. Li, Z., Li, F., Peng, Y., Fang, J. & Zhou, J. Identification of three m6A-related mRNAs signature and risk score for the prognostication of hepatocellular carcinoma. *Cancer Med.* **9**, 1877–1889. <https://doi.org/10.1002/cam4.2833> (2020).
33. Chen, Q. *et al.* Identification of energy metabolism genes for the prediction of survival in hepatocellular carcinoma. *Front. Oncol.* **10**, 1210. <https://doi.org/10.3389/fonc.2020.01210> (2020).
34. Cheng, W., Su, Y. & Xu, F. CHD1L: A novel oncogene. *Mol. Cancer* **12**, 170. <https://doi.org/10.1186/1476-4598-12-170> (2013).
35. Rivas, M. *et al.* HDAC1-dependent repression of markers of hepatocytes and P21 is involved in development of pediatric liver cancer. *Cell. Mol. Gastroenterol. Hepatol.* **12**, 1669–1682. <https://doi.org/10.1016/j.jcmgh.2021.06.026> (2021).
36. Han, Y. & Wang, X. The emerging roles of KPNA2 in cancer. *Life Sci.* **241**, 117140. <https://doi.org/10.1016/j.lfs.2019.117140> (2020).
37. Sakamoto, H. *et al.* MUTYH is associated with hepatocarcinogenesis in a non-alcoholic steatohepatitis mouse model. *Sci. Rep.* **11**, 3599. <https://doi.org/10.1038/s41598-021-83138-8> (2021).
38. Wei, W., He, Y. & Wu, Y. M. Identification of genes associated with SiHa cell sensitivity to paclitaxel by CRISPR-Cas9 knockout screening. *Int. J. Clin. Exp. Pathol.* **11**, 1972–1978 (2018).
39. Lai, H. H. *et al.* NEIL3 promotes hepatoma epithelial–mesenchymal transition by activating the BRAF/MEK/ERK/TWIST signaling pathway. *J. Pathol.* **258**, 339–352. <https://doi.org/10.1002/path.6001> (2022).
40. Yao, F. *et al.* Single-cell RNA sequencing reveals the role of phosphorylation-related genes in hepatocellular carcinoma stem cells. *Front. Cell Dev. Biol.* **9**, 734287. <https://doi.org/10.3389/fcell.2021.734287> (2021).
41. Feng, S. *et al.* Amplification of RAD54B promotes progression of hepatocellular carcinoma via activating the Wnt/beta-catenin signaling. *Transl. Oncol.* **14**, 101124. <https://doi.org/10.1016/j.tranon.2021.101124> (2021).
42. Mello, T. *et al.* Liver haploinsufficiency of RuvBL1 causes hepatic insulin resistance and enhances hepatocellular carcinoma progression. *Int. J. Cancer* **146**, 3410–3422. <https://doi.org/10.1002/ijc.32787> (2020).
43. Zhao, K., Ma, Z. & Zhang, W. Comprehensive analysis to identify SPP1 as a prognostic biomarker in cervical cancer. *Front. Genet.* **12**, 732822. <https://doi.org/10.3389/fgene.2021.732822> (2021).
44. Zou, W. *et al.* Identification of inflammatory response-related gene signature associated with immune status and prognosis of lung adenocarcinoma. *Front. Bioeng. Biotechnol.* **9**, 772206. <https://doi.org/10.3389/fbioe.2021.772206> (2021).
45. Greenbaum, L. E. Cell cycle regulation and hepatocarcinogenesis. *Cancer Biol. Ther.* **3**, 1200–1207. <https://doi.org/10.4161/cbt.3.12.1392> (2004).
46. Liu, Y. L. *et al.* Autophagy potentiates the anti-cancer effects of the histone deacetylase inhibitors in hepatocellular carcinoma. *Autophagy* **6**, 1057–1065. <https://doi.org/10.4161/auto.6.8.13365> (2010).
47. Zhou, S. L. *et al.* Tumor-associated neutrophils recruit macrophages and T-regulatory cells to promote progression of hepatocellular carcinoma and resistance to sorafenib. *Gastroenterology* **150**, 1646–1658 e1617. <https://doi.org/10.1053/j.gastro.2016.02.040> (2016).
48. Zhang, Q. *et al.* Landscape and dynamics of single immune cells in hepatocellular carcinoma. *Cell* **179**, 829–845 e820. <https://doi.org/10.1016/j.cell.2019.10.003> (2019).
49. Fu, J. *et al.* Increased regulatory T cells correlate with CD8 T-cell impairment and poor survival in hepatocellular carcinoma patients. *Gastroenterology* **132**, 2328–2339. <https://doi.org/10.1053/j.gastro.2007.03.102> (2007).
50. Chinai, J. M. *et al.* New immunotherapies targeting the PD-1 pathway. *Trends Pharmacol. Sci.* **36**, 587–595. <https://doi.org/10.1016/j.tips.2015.06.005> (2015).
51. Giannini, E. G. *et al.* Overview of immune checkpoint inhibitors therapy for hepatocellular carcinoma, and the ITA.LI.CA cohort derived estimate of amenability rate to immune checkpoint inhibitors in clinical practice. *Cancers (Basel)* <https://doi.org/10.3390/cancers11111689> (2019).
52. Street, S. E., Trapani, J. A., MacGregor, D. & Smyth, M. J. Suppression of lymphoma and epithelial malignancies effected by interferon gamma. *J. Exp. Med.* **196**, 129–134. <https://doi.org/10.1084/jem.20020063> (2002).
53. Shankaran, V. *et al.* IFN γ and lymphocytes prevent primary tumour development and shape tumour immunogenicity. *Nature* **410**, 1107–1111. <https://doi.org/10.1038/35074122> (2001).
54. Lecocq, Q., Keyaerts, M., Devoogdt, N. & Breckpot, K. The next-generation immune checkpoint LAG-3 and its therapeutic potential in oncology: Third time's a charm. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms22010075> (2020).

Acknowledgements

With the consent of all authors, we would like to extend special thanks to Dr. Yixin Zhang (Shandong First Medical University) for her important suggestions on this paper. We would also like to give special thanks to the administrators of the TCGA database and the ICGC database as well as all open public websites.

Author contributions

This work was designed by L.F. and Y.P.L., S.W. and Y.P.L. authored the text after analyzing the data. T.T.C. and H.Z.W. assisted in data collection and text revision. Y.L.Z. and H.M.G. contributed to the data analysis. The final manuscript was reviewed and approved by all writers. The publishing of this paper has the approval of all authors.

Funding

This work was supported by the National Natural Science Foundation of China (No. 31900889).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-45999-z>.

Correspondence and requests for materials should be addressed to L.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023