Check for updates

# Estrogen-related genes influence immune cell infiltration and immunotherapy response in Hepatocellular Carcinoma

Biao Gao[1,2], Yafei Wang[1,2], Chonghui Li[2,3]* and Shichun Lu[2,3]*

[1]Nankai University School of Medicine, Nankai University, Tianjin, China, [2]Faculty of Hepato-Pancreato-Biliary Surgery, Chinese PLA General Hospital, Beijing, China, [3]Institute of Hepatobiliary Surgery, Chinese PLA General Hospital, Beijing, China

**Background:** Immunotherapy has been the first-line treatment option in advanced Hepatocellular Carcinoma(HCC); but now, there are no established molecular markers that can predict immunotherapy response. Estrogen has a crucial role in the development of a variety of liver illnesses, including liver fibrosis, Nonalcoholic fatty liver disease (NAFLD), and HCC. Nonetheless, the significance of estrogen-related genes in HCC immunotherapy and the underlying molecular mechanisms are not yet fully understood.

**Method:** In this study, we constructed a novel estrogen-related gene prognostic signature (ERGPS) by analyzing bulk RNA sequencing data from 365 HCC patients. Based on the median risk score, we divided 365 HCC patients into low- and high-risk groups. Tumor mutation burden (TMB), Microsatellite instability (MSI), T cell receptor (TCR) richness, B cell receptor (BCR) richness, single-nucleotide variants (SNV) Neoantigens, Cancer Testicular Antigens (CTA) scores, and Tumour Immune Dysfunction and Exclusion (TIDE) scores were used to evaluate the magnitude of immunotherapy response. Multiple external datasets validate the validity and robustness of the prognostic signature. Real-time quantitative polymerase chain reaction (qRT-PCR) was used to validate estrogen-related gene overexpression in HCC tissue samples.

**Results:** ERGPS is an independent risk factor affecting the prognosis of HCC patients and is superior to other clinical variables in predicting patient survival and immunotherapy response. Multiple independent external datasets confirmed the superior predictive efficacy of the prognostic signature. The prognostic signature was positively correlated with TMB score, MSI score, TCR richness, BCR richness, SNV Neoantigens score, CTA score, expression levels of immune checkpoint-related genes, and TIDE score. Patients with HCC in the high-risk group identified by the prognostic signature were likely to be more responsive to immunotherapy and more suitable for immunotherapy. qRT-PCR confirmed that estrogen-related genes of the construct signature were highly expressed in HCC tumor tissues.

**Conclusion:** Estrogen-related genes are overexpressed in HCC tissues. Our novel prognostic signature can accurately predict not only the prognosis but also the immunotherapy response of HCC patients. In the future, prognostic signatures will be a useful tool for clinicians to screen patients with HCC who are suitable for immunotherapy.

KEYWORDS

Hepatocellular Cacinoma, immunotherapy, estrogen, MSI (microsatellite instability), TCR (T cell receptor)

# Introduction

Primary liver cancer is the sixth most prevalent malignancy and the third main cause of cancer-related deaths worldwide. Primary liver cancer includes hepatocellular carcinoma (75%-85% of cases), intrahepatic cholangiocarcinoma (10%-15% of cases), and other rare types (1). Due to the insidious onset of the disease, most individuals with liver cancer are already in the advanced stage at the time of diagnosis, and hence lost the time window for radical surgery. In recent years, the breakthrough of immunotherapy research represented by immune checkpoint inhibitors has broken the traditional treatment pattern of liver cancer and opened a new chapter in the treatment of advanced liver cancer (2–5). However, despite the current achievements in immunotherapy for liver cancer, it still faces challenges such as low objective response rates, lack of effective biological markers to predict treatment response, andimmune-related adverse events (irAEs). Meanwhile, only a small percentage of patients benefit from immunotherapy, and even a few patients develop hyperprogressive disease (HPD) after receiving immunotherapy. Therefore, the development of biological markers that can accurately predict the response to immunotherapy and thus assist doctors to screen patients with HCC who are suitable for immunotherapy has become an urgent task.

The tumor microenvironment (TME) plays a vital role in the progression and development of HCC. In addition to tumor cells, the tumor microenvironment of HCC includes stromal cells (such as immune cells, fibroblasts, endothelial cells, etc.), structural components (such as extracellular matrix, etc.), and signaling components (chemokines, cytokines, and growth factors, etc.), which influence tumor immune evasion, response to immunotherapy, and patient prognosis (6–8). As a result, a better understanding of the crosstalk among the components of the HCC immune microenvironment, as well as the identification of potential therapeutic targets from it, can aid in the prediction of immunotherapeutic response, determination of immunotherapeutic efficacy, identification of prognostic markers, and guidance of individualized treatment regimens.

The incidence of HCC shows considerable gender disparities, with a much greater incidence in males than in women, with the ratio usually between 2:1 and 4:1; while female patients have a better prognosis and longer survival time than male patients (9). It is thought that estrogen has a crucial role in the occurrence and progression of HCC (10, 11). Although there is emerging evidence that estrogen is a potentially critical host factor in HCC advancemen, the potential mechanisms of estrogen-related genes in HCC development and immunotherapy are yet unknown.

In this study, to investigate the relationship between estrogen-related genes and HCC, we performed a comprehensive analysis of bulk RNA sequencing data from 365 HCC patients and single-cell sequencing data from 16 HCC patients and constructed a novel prognostic signature based on 3 key estrogen -related genes.TMB, MSI, CTA score, SNV Neoantigens score, mRNA expression levels of immune checkpoint-related genes, and TIDE score were used to evaluate the efficacy of immunotherapy in different risk groups. The prognostic signature showed superiority in predicting both prognosis and response to immunotherapy in HCC patients, and its validity was validated by many independent external data sets. In the future, the prognostic signature will be a useful tool for clinicians to screen populations suitable for immunotherapy.

# Materials and methods

## Data and clinical samples

The bulk RNA sequencing data and clinical information of 424 HCC patients were downloaded from the TCGA database(https://portal.gdc.cancer.gov/), including sequencing information of 365 HCC tissue samples and 59 paracancerous tissue samples as the training set. The bulk RNA sequencing data and clinical information from 231 HCC patients were downloaded from the ICGC database(https://dcc.icgc.org/) as validation sets. As validation sets, sequencing and clinical data from 115 HCC patients were extracted from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). From CellMarker2.0 (http://yikedaxue.slwshop.cn/), 731 estrogen-related genes were extracted. Transcriptomic and matched clinical data were acquired from the IMvigor210 and GSE9106 cohort of anti-PD-L1-treated patients to investigate the predictive utility of ERGPS for immunotherapy response. The clinical features of 365 HCC patients in the training set are shown in Table 1. Figure S1 shows the data analysis process.

TABLE 1 Clinical characteristics of 365 HCC patients in TCGA LIHC.

| Characteristics | Samples (n=365) | Percentage (%) |
|---|---|---|
| Gender | | |
| Female | 120 | 32.9% |
| Male | 245 | 67.1% |
| Age | | |
| >=60 | 200 | 54.8% |
| <60 | 165 | 45.2% |
| Tumor Stage | | |
| Stage I | 180 | 49.3% |
| Stage II | 91 | 25.0% |
| Stage III | 81 | 22.2% |
| Stage IV | 13 | 3.6% |
| M Stage | | |
| M0 | 263 | 72.1% |
| MX | 102 | 28.0% |
| N Stage | | |
| N0 | 248 | 67.9% |
| N1 | 5 | 1.4% |
| NX | 112 | 30.7% |
| Tumor Grade | | |
| G1 | 58 | 15.9% |
| G2 | 176 | 48.2% |
| G3 | 116 | 31.8% |
| G4 | 13 | 3.6% |

## Identification of differentially expressed Estrogen-related genes

To identify differentially expressed genes, we analyzed the difference between 365 HCC samples and 59 paracancarcinoma tissue samples in the training set using the "limma" package of R software and set the criterion as p-value 0.05 and log(FC) > 1 or log (FC) -1. Subsequently, the intersection of differentially expressed genes and estrogen-related genes yielded differentially expressed estrogen-related genes.

## Construction and validation of a prognostic signature based on estrogen-related genes

We first screened the genes influencing the prognosis of HCC patients in the training set with a p-value of 0.05 using univariate Cox analysis of differentially expressed estrogen-related genes. To eliminate overfitting, a least absolute shrinkage and selection operator (LASSO) Cox regression analysis was employed in conjunction with the "glmnet" package. Finally, Cox multivariate

analysis was utilized to screen for independent risk factors affecting the prognosis of HCC patients with screening criterion of p-value < 0.05. Subsequently, we used the selected genes to construct a prognostic signature. Risk score was computed using the following formula: Risk score = (Expi*Coefi). Coefi and Expi respectively represent the risk coefficient and gene expression. According to the median risk score, 365 HCC patients in the training set were divided into high-risk group and low-risk group. The Kaplan-Meier survival curve was used to compare survival differences between the two groups. The receiver operating characteristic curve(ROC) was used to test the predictive power of risk scores for HCC patients. To verify the robustness of the signature, we first used Kaplan-Meier survival analysis to compare survival differences between high-risk and low-risk groups in different clinical subgroups. To validate the efficacy of the prognostic signature, we applied the same algorithm to calculate the risk score of HCC patients in two independent external data sets and utilized the median value of risk groups to classify HCC patients into high-risk and low-risk groups. Kaplan Meier survival analysis was used to compare survival differences between the two groups. ROC curve was used to test the predictive efficacy of prognosis signature for HCC patients.

## Pathways and functional enrichment analysis

In order to further explore the molecular mechanism behind differential genes, the "clusterProfiler" of R software is used for GO function enrichment analysis and KEGG function enrichment analysis. A p-value of less than 0.05 was considered to show substantial enrichment. Software for Gene Set Enrichment Analysis (GSEA) was used to compare the pathway enrichment scores of the high-risk and low-risk groups. The gene expression profile of the two risk samples was used to evaluate the associated pathways and molecular processes, and the reference gene sets "c2.cp.kegg.v7.4.symbols" were obtained from the molecular signatures database(https://www.gsea-msigdb.org/gsea/msigdb).

## Immune cell infiltration analysis and gene sets variation analysis

Based on the principle of linear support vector regression, CIBERSORT deconvolutes the expression matrix of human immune cell subtypes and determines the content of each subtype (12). The R software's CIBERSORT function is used to deconvolute the bulk RNA sequencing data of 365 HCC patients and to quantify the fraction of immune cells in the tumor microenvironment for each patient in the training set. The difference in the infiltration of 22 types of immune cells between the high-risk and low-risk groups was compared. GSVA (Gene Set Variation Analysis) is a nonparametric and unsupervised analysis method, primarily used to evaluate the microarray and transcriptome gene set enrichment results (13). In addition, the GSVA function of the R software is used to compute the enrichment score for each patient's 28 types of immune cells. Additionally, we compared the enrichment scores of the two groups.

## Estimation of stromal and immune scores

The ESTIMATE algorithm can estimate the matrix score and immune score of tumor samples according to transcriptome data, which is used to represent the relative proportion of matrix and immune cells (14, 15). Using the R software's ESTIMATE function, the Stromal score, immunological score, ESTIMATE score, and tumor purity score of each patient in the training set are evaluated. The Wilcoxon t-test was utilized to analyze the difference between the high-risk and low-risk groups in terms of immune cell infiltration score.

## Mutation analysis and microsatellite instability

To examine the difference between the tumor cell mutation gene and TMB between high-risk and low-risk groups, we obtained the mutation data of 365 HCC patients in the training set from the TCGA database and analyzed the mutation data using the "maftools" package (16) of R software. TMB and MSI can be utilized as useful biomarkers of immune checkpoint inhibitors, as demonstrated in numerous solid malignancies (17, 18). To analyze the difference between the high-risk group and the low-risk group in response to immunotherapy, we evaluated the differences in tumor mutation genes, TMB and MSI, between the two groups.

## Prediction and validation of immunotherapy response

Immunotherapy is playing an increasingly important role in the treatment of advanced hepatocellular carcinoma. The mRNA expression levels of immune checkpoint-associated genes are the basis of immunotherapy. To identify the degree of immunotherapy response in different risk groups, we analyzed the expression differences of mRNA of immune checkpoint-related genes in high-risk and low-risk groups. Richness and the Shannon Diversity Index were used to describe the diversity of the TCR repertoire. Richness measures the number of unique TCRs in the sample, while the Shannon diversity index reflects the relative abundance of different TCRs. BCR is a B lymphocyte receptor and BCR richness is a combination of the various BCR isoforms produced in an individual. TCR richness and BCR richness can be utilized to determine the intensity of immunotherapy response, according to previous research. Neoantigens are novel antigens created by tumor cells. The quantity of tumor neoantigens corresponds with the effectiveness of PD-1/PD-L1 antibodies as immunotherapy improves the immune system's ability to recognize and destroy neoantigens on the surface of tumor cells. The cancer testicular antigens (CTA) score is used to evaluate tumor immunogenicity, which indirectly reflects the intensity of the immunotherapy response. In order to assess disparities in immunotherapy response, we examined the differences in TCR, BCR, SNV, and CTA scores between the high-risk and low-risk groups. Tumor immune dysfunction and rejection (TIDE) consists of two major mechanisms through the expression of characteristic genes: the

induction of T cell dysfunction in tumors with high CTL infiltration and the blocking of T cell infiltration in tumors with low CTL infiltration to construct mathematical prediction models that can also serve as new biomarkers to predict the response to immune check inhibitors. We computed TIDE scores for each HCC patient and evaluated the distributional differences between the high-risk and low-risk groups in the training set.

## Independent external immunotherapy data for validation of immunotherapy response

To effectively confirm the differences in immunotherapy response among risk groups, we obtained comprehensive bulk RNA sequencing data and clinical information from two external immunotherapy data cohorts (IMvigor210,GSE9106). We calculated risk scores for patients receiving immunotherapy in both immunotherapy datasets using the same formula as in the training set, and classified patients into high-risk and low-risk groups based on the median risk score. The proportion of patients who achieved CR/PR(CR: complete response; PR: partial response) or SD/PD(SD: stable disease;PD: progressive disease) following immunotherapy was compared between the two groups to identify the differential in response to immunotherapy between risk groups.

## Drug sensitivity prediction

The pRRophetic package of R software was used to predict the sensitivity (IC50 values) of 138 drugs in the GDSC database in combination with model gene expression data, and the sensitivity of HCC patients to drug therapy was assessed by IC50 values. The differences in IC50 values between the risk groups were compared, and the Wilcoxon test was performed to identify drugs with significant differences between the two groups.

## qRT-PCR confirmed the overexpression of three estrogen-related genes in HCC tissues

To evaluate the robustness of the signature in clinical samples, we gathered tumor and paraneoplastic tissue samples from 12 patients with HCC verified by postoperative pathology. We used qRT-PCR to validate the mRNA expression levels of the three estrogen-related genes used to construct the prognostic signature.

## Statistical analysis

Statistical tests were performed by R statistical computing software (R version 4.1.2). Comparisons between two groups were estimated for significance by the nonparametric Wilcoxon test, and in multiple comparisons by the Kruskal-Wallis test. Categorical data were tested by chi-square test and trend chi-square test. K-M analysis of overall survival(OS) between different subgroups was performed followed by log-rank test. Univariate and multivariate Cox regression analyses were used to screen for independent risk factors affecting the

prognosis of HCC patients. A significance criterion of P < 0.05 was selected.

# Results

## Identification of differentially expressed estrogen-related genes

A total of 8904 differential genes were identified by differential analysis of bulk RNA sequencing data from 365 HCC and 59 paraneoplastic tissue samples in the training set, with the set criteria of p-value of < 0.05 and a |log2 (fold change) of |> 1 (Figure 1A). The intersection of 8904 differential genes with 731 estrogen-related genes resulted in the identification of 245 differential estrogen-related genes (Figure 1B). To investigate the signaling pathways associated with differential estrogen-related genes, we did GO and KEGG functional enrichment analysis on these genes. The results of GO functional enrichment analysis showed that these differential estrogen-related genes were mainly enriched based on signaling receptor activator activity, receptor-ligand activity, and DNA (Figure 1C). KEGG functional enrichment analysis showed that these differential estrogen-related genes were mainly enriched in PI3K-Akt signaling pathway, MAPK signaling pathway, and Ras signaling pathway(Figure 1D). The PI3K-Akt signaling pathway, MAPK signaling pathway, and Ras signaling pathway play important roles in the development and progression of various tumors, and these results suggest that estrogen-related genes may play role in HCC through these signaling pathways.

## Construction and validation of estrogen-related genes prognostic signature

We analyzed 245 differential estrogen-related genes using univariate Cox regression analysis and screened 99 genes that had an impact on the prognosis of HCC patients with a set criterion of p<0.05. Then, LASSO regression and Cox proportional hazard model studies were utilized to determine the optimal model in the training dataset (Figures 2A, B). In addition, multivariate Cox regression analysis was utilized to analyze six differential estrogen-related genes, resulting in the identification of three genes that may serve as independent risk factors influencing the prognosis of HCC patients (Figure 2C). We determined each patient's risk score using the following formula: Risk Score = 3.365e-04*AKR1B15 mRNA expression plus 6.65e-04*KCTD6 mRNA expression plus 1.41e-04*KPNA2 mRNA expression. Based on the median risk score, 365 HCC patients were divided into high-risk group or low-risk group in the training set. As shown in the heatmap, the mRNA expression levels of 3 estrogen-related genes were significantly higher in the high-risk group than in the low-risk group; the mortality rate of patients gradually increased with increasing risk scores(Figure 2D). The Kaplan-Meier survival analysis revealed that the overall survival time of patients in the high-risk group in the training set was considerably inferior to that of patients in the low-risk group (Figure 2E). To assess the predictive accuracy of the risk signature, time-dependent area under the ROC curves for OS was calculated,

and the 1-, 3-, and 5-year AUC values were 0.78, 0.72, and 0.73 (Figure 2F), respectively.

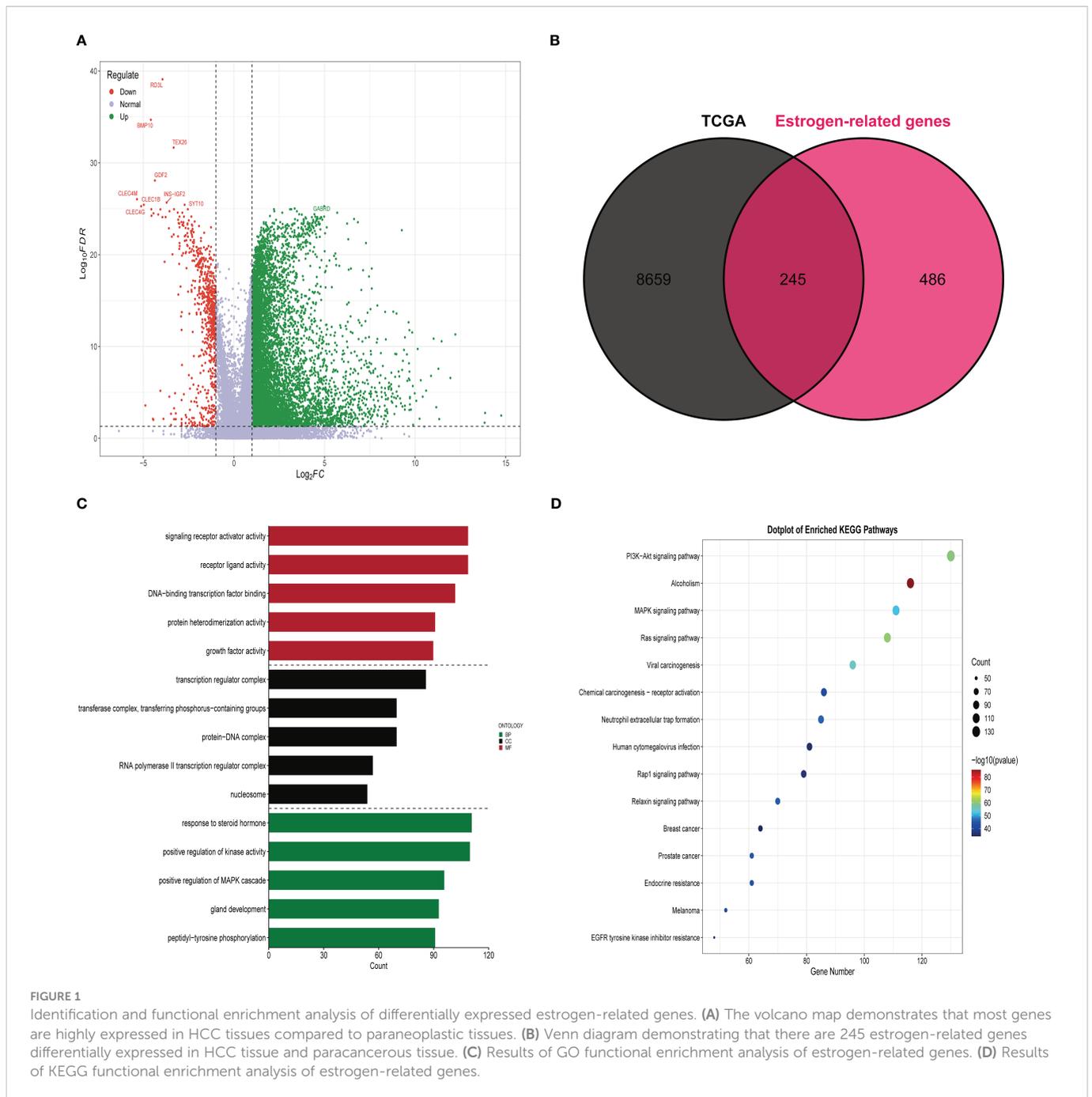## Validation of the ERGPS in different clinical subgroups

We first evaluated the relationship between ERGPS and clinical characteristics including age, sex, N stage, M stage, tumor grade, tumor stage, microvascular invasion(MVI), and survival status. Patients with higher tumor grade and tumor stage had higher risk scores. Also, patients who died had a higher risk score compared to those who survived(Figure S2). The predictive value of ERGPS was first evaluated in TCGA LIHC patients of different age, gender, microvascular invasion, tumor grade and tumor stage. The results revealed that high-risk score was significantly correlated with an inferior prognosis in the young(P<0.001, Figure S3A), old(P<0.001, Figure S3B), male(P<0.001, Figure S3C) or female(P=0.04, Figure S3D), MVI negative(P=0.024, Figure S3E), MVI positive(P=0.031, Figure S3F) or low tumor grade(P=0.033, Figure S4A), high tumor grade (P= 0.007, Figure S4B), early stage(P= 0.004, Figure S4C) or advanced stage(P= 0.005, Figure S4D).

## External validation of the robustness of the ERGPS in two independent cohorts

To validate the robustness of ERGPS, we included two independent cohorts in this study. We calculated risk scores for each patient in 2 independent cohorts using the same formula and divided HCC patients into high- or low-risk groups based on the median risk score. Kaplan–Meier analysis demonstrated that the high-risk group had an inferior prognosis than the low-risk group in both of these 2 independent cohorts, namely GSE76427 (Figure 3A, HR=0.411,95%CI:0.1668-1.0130,p=0.046) and ICGC (Figure3B, HR=2.397, 95%CI:1.262-4.555,p=0.00758). In addition, we evaluated the cumulative predictive value of the training set (TCGA) and the two validation sets for the prognosis of HCC patients by conducting a prognostic meta-analysis utilizing the random effects model R software meta. The results of the meta-analysis indicated that ERGPS was an extremely accurate predictor of prognosis for HCC patients (Figure 3C, HR = 1.13, 95% CI = 0.45-2.05, p<0.05).

## Identification of ERGPS as an independent prognostic factor

To assess the impact of ERGPS on the prognosis of HCC patients, we included clinical characteristics including age, gender, N stage, M stage, tumor grade, and tumor stage in the training set for analysis. Univariate Cox regression analysis showed an effect of tumor stage and risk score on the prognosis of HCC patients (p<0.05). The results of multivariate Cox regression analysis showed that tumor stage and risk score were independent risk factors affecting the prognosis of HCC patients (p<0.05) (Table 2). We compared the predictive efficacy of risk scores and clinical characteristics for the prognosis of HCC patients in the

**FIGURE 1**

Identification and functional enrichment analysis of differentially expressed estrogen-related genes. **(A)** The volcano map demonstrates that most genes are highly expressed in HCC tissues compared to paraneoplastic tissues. **(B)** Venn diagram demonstrating that there are 245 estrogen-related genes differentially expressed in HCC tissue and paracancerous tissue. **(C)** Results of GO functional enrichment analysis of estrogen-related genes. **(D)** Results of KEGG functional enrichment analysis of estrogen-related genes.

training set using ROC curves. The results revealed that the area under the curve for risk scores was greater than that for clinical features (Figure 4A), indicating that risk scores are more predictive for HCC patients than clinical characteristics. To forecast the prognosis of HCC patients more intuitively and precisely, we analyzed the tumor grade and risk score using the random forest method and created a nomogram to predict the prognosis of HCC patients (Figure 4B). The ROC curve demonstrated the excellent predictive power of the nomogram for overall survival time at 1, 3, and 5 years in HCC patients (Figure 4C). Calibration curve indicates the great accuracy of the nomogram in predicting HCC patient survival at 1, 3, and 5 years (Figure 4D). The decision curve confirmed that the nomogram, constructed from the tumor stage and risk score, significantly outperformed the other factors (Figure 4E).

## Gene set enrichment analysis and functional enrichment analysis

To further examine the molecular mechanism underlying ERGPS, we performed the differential analysis of bulk RNA sequencing data from high-risk and low-risk patients to identify differentially expressed genes, using the following criteria: p-value 0.05, log(FC) > 1 or log(FC) < -1. In total, we obtained 3723 differentially expressed genes, of which 3590 were overexpressed in the high-risk group and 133 in the low-risk group (Figure S5A). GO functional enrichment analysis of overexpressed genes in the two groups showed that the overexpressed genes in the high-risk group were mainly enriched in the passive transmembrane transporter activity, channel activity, and ion channel activity signaling pathways
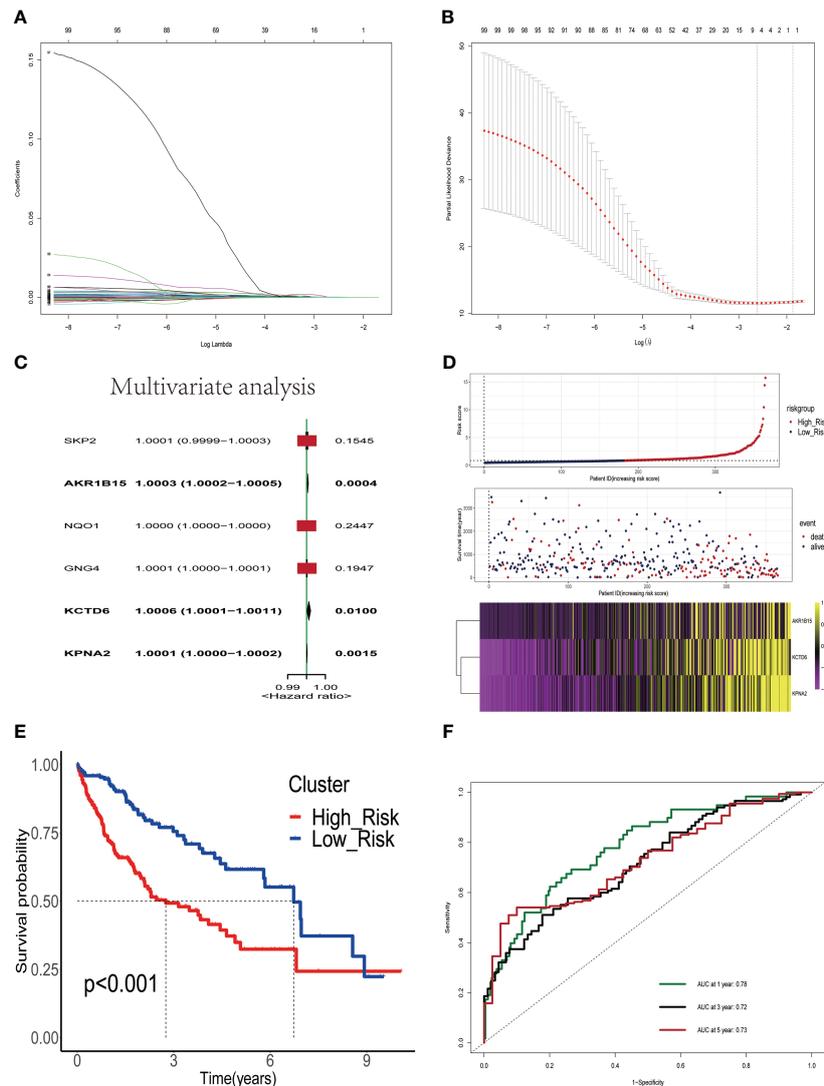
**FIGURE 2**

Construction and validation of a prognostic signature based on estrogen-related genes. **(A)** The coefficients of genes calculated by multivariate Cox regression using LASSO. **(B)** The partial likelihood deviance of genes. **(C)** Results of multivariate analysis of six differentially expressed estrogen-related genes. **(D)** The association of risk scores with survival status and gene expression in HCC patients. **(E)** Kaplan-Meier curves were used to compare the overall survival of HCC patients between the high-risk and low-risk groups. **(F)** ROC curves of the prognostic signature for predicting the risk of death at 1, 3, and 5 years.

(Figure S5B), while the low-risk group was mainly enriched in the tetrapyrrole binding, heme binding and receptor-ligand activity signaling pathways(Figure S5C). GSEA enrichment analysis showed that the high-risk group was mainly enriched in the KEGG_CELL_CYCLE, KEGG_RNA_DEGRADATION, and KEGG_OOCYTE_MEIOSIS signaling pathways, while the low-risk group was mainly enriched in the KEGG_PRIMARY_BILE_ ACID_BIOSYNTHESIS, KEGG_COMPLEMENT_AND_ COAGULATION_CASCADES, and KEGG_LINOLEIC_ACID_ METABOLISM signaling pathways(Figure S5D). The enrichment results of the HALLMARK pathway showed that E2F_TARGETS, G2M_CHECKPOINT, and DNA_REPAIR were mainly enriched in high-risk group, while COAGULATION was mainly enriched in low-risk group(Figures S6A–D), suggesting that estrogen in tumor microenvironment was involved in the regulation of cell cycle.

## Association between ERGPS and immune cell infiltration

The microenvironment of a tumor plays an essential role in the development and progression of the disease. We inferred the proportion of 22 immune cell types in the tumor microenvironment of 365 HCC patients using the CIBERSOR function of the R software. In addition, we evaluated the fraction of immune cells infiltrating the tumor microenvironment between high-risk and low-risk groups. In comparison to the low-risk group, the high-risk group had a higher proportion of immune cells with immunosuppressive functions, such as Tregs, Macrophages, Dendritic cells resting, and Neutrophils, and a lower proportion of immune cells with immunocidal functions, such as T cells CD8 and T cells CD4 memory resting (Figure 5A). To more comprehensively

TABLE 2  Results of univariate and multivariate Cox regression analyses of clinical characteristics and risk scores in the TCGA LIHC dataset.

| Characteristics | Univariate analysis | | Multivariate analysis | |
|---|---|---|---|---|
| | Hazard ratio (95% CI) | P value | Hazard ratio (95% CI) | P value |
| Age | 1.011 (0.997-1.026) | 0.125 | | |
| Gender | | | | |
| Female | Reference | | | |
| Male | 0.792 (0.545-1.150) | 0.221 | | |
| G stage | | | | |
| G1+G2 | Reference | | | |
| G3+G4 | 1.186 (0.819-1.716) | 0.366 | | |
| M stage | | | | |
| M0 | Reference | | | |
| MX | 1.392 (0.915-2.117) | 0.122 | | |
| N stage | | | | |
| N0 | Reference | | | |
| NX | 1.228 (0.845-1.961) | 0.239 | | |
| Stage | | | | |
| StageI+StageII | Reference | | | |
| Stage III+Stage IV | 2.409 (1.664-3.487) | <0.001 | 2.268 (1.566- 3.285) | <0.001 |
| RiskScore | 1.266 (1.184-1.355) | <0.001 | 1.268 (1.178-1.363) | <0.001 |

assess the differences in the infiltration of immune cells in the tumor microenvironment of 365 HCC patients, we used the GSVA function of R software to calculate the percentage of infiltration of 28 immune cell types. The results showed a higher proportion of Activated_CD4_T_cell, Activated_dendritic_cell, Memory_B_cell, and Type_2_T_helper_cell in the high-risk group compared to the low-risk group, while Activated_CD8_T_cell, CD56bright_natural_ killer_cell, Effector_memeory_CD8_T_cell, and Type_1_T_helper_ cell have a lower percentage(Figure 5B). These results suggest the presence of a immunosuppressive microenvironment in patients in the high-risk group, where the killer function of immune cells is suppressed, which partly explains the poorer prognosis of patients in the high-risk group.

We then evaluated the relationship between StromalScore, ImmuneScore, ESTIMATEScore, TumorPurity, and the ERGPS, and the results showed that ESTIMATEScore and StromalScore were higher in the low-risk group, whereas TumorPurity was higher in the high-risk group (Figures 5C–F).

## Relationship between the ERGPS and tumor mutation burden and microsatellite instability

Interestingly, we found that the frequency of somatic mutations in the high-risk group was 91.62%(Figure 6A), whereas in the low-risk group it was 84.75%(Figure 6B). TP53 was the gene with the highest mutation frequency in the high-risk group, whereas CTNNB1 had the

highest mutation frequency in the low-risk group. The findings of a comparison of the difference in TMB scores between the two groups revealed that the high-risk group had a higher TMB score(p<0.05) (Figure 6C). Subsequently, we explored the relationship between ERGPS and MSI scores, and the results showed that the high-risk group had higher MSI scores compared with the low-risk group (p<0.05)(Figure 6D). These results suggest that the high frequency of TP53 gene mutations may be partly responsible for the poorer prognosis of HCC patients in the high-risk group. Meanwhile, patients in the high-risk group had a higher frequency of somatic mutations, a higher TMB score, and a higher MSI score, suggesting that patients in the high-risk group may be more responsive to immunotherapy and more suitable for immunotherapy.

## Prediction of immunotherapy response

Immunotherapy acts against tumors by acting on tumor cells or immune checkpoint-associated proteins on the surface of immune cells. Therefore, the mRNA expression levels of immune checkpoint-related genes have an important role in the immunotherapy of tumors. We first compared the differences in mRNA expression levels of immune checkpoint-associated genes commonly used in most solid tumors between the two groups. The results showed that the mRNA expression levels of these immune checkpoint-associated genes were elevated in the high-risk group(Figure 7A, Figure S7). The relationship between ERGPS and mRNA expression levels of immune checkpoint-related genes (PD-L1, PDCD1LG2, PDCD1, TIGIT, TIM-3, CTLA4)
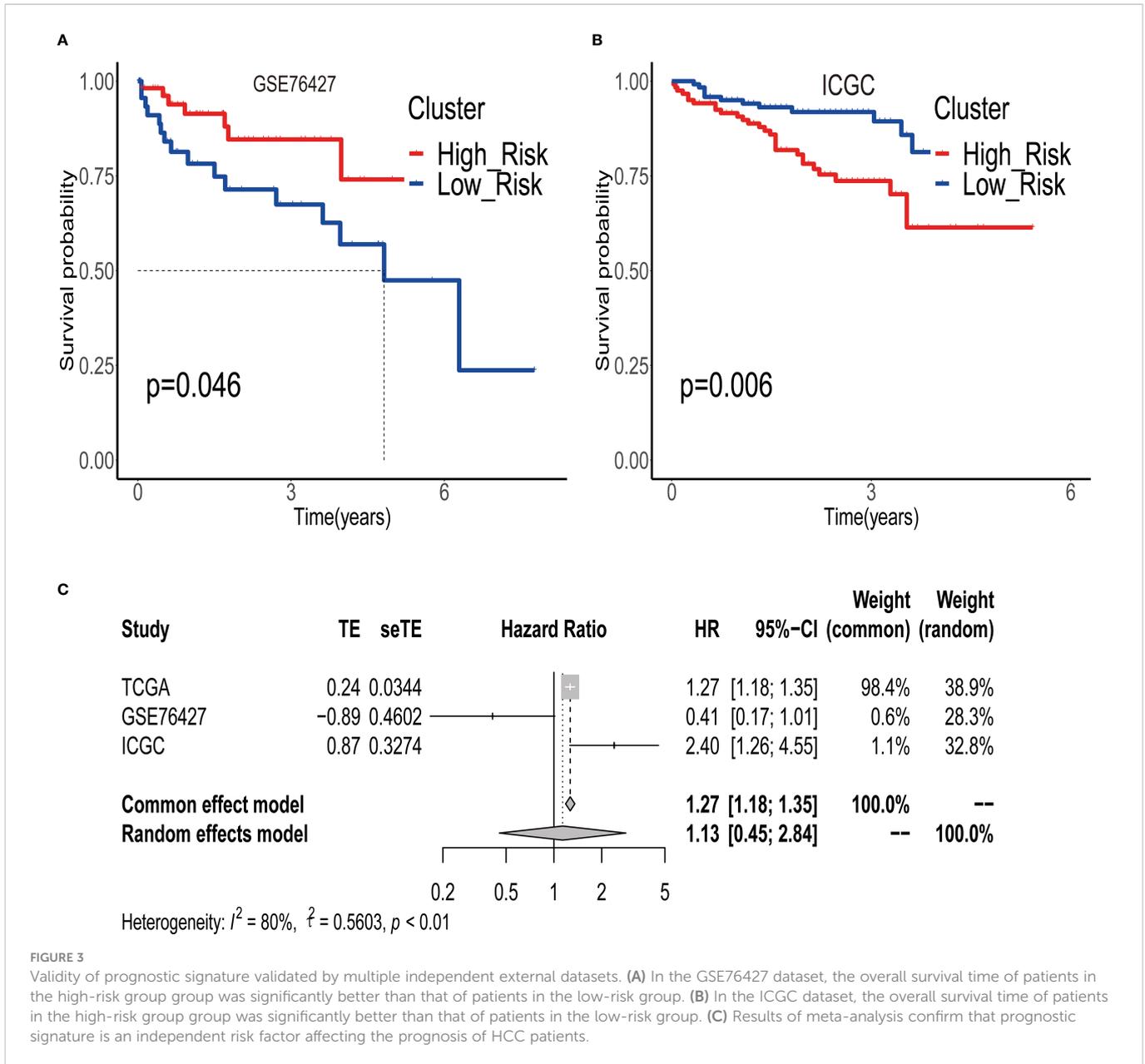
**FIGURE 3**

Validity of prognostic signature validated by multiple independent external datasets. **(A)** In the GSE76427 dataset, the overall survival time of patients in the high-risk group group was significantly better than that of patients in the low-risk group. **(B)** In the ICGC dataset, the overall survival time of patients in the high-risk group group was significantly better than that of patients in the low-risk group. **(C)** Results of meta-analysis confirm that prognostic signature is an independent risk factor affecting the prognosis of HCC patients.

commonly used in liver cancer immunotherapy was investigated further, and the results indicated that mRNA expression levels of these immune checkpoint-related genes were elevated in the high-risk group compared to the low-risk group (Figures 7B–G). These findings imply that immunosuppressive microenvironment exists in the high-risk group and that immunotherapy can reverse this immunosuppressive state so that patients can benefit. In previous research, T cell receptors (TCR) and B cell receptors (BCR) were responsible for recognizing antigens presented by MHC, and rearrangement analysis of TCR and BCR has proven to be an effective biomarker for stratifying and monitoring immunotherapy patients (19–22). To further evaluate the link between ERGPS and immunotherapy response, we evaluated the differences between the two groups' TCR, BCR, SNV, and CTA scores. Intriguingly, we discovered

that the high-risk group had higher TCR, BCR, SNV, and CTA scores than the low-risk group(Figures 8A–D), and these findings suggest that high-risk patients may be more susceptible to immunotherapy. The TIDE score, a recently discovered computational technique for modeling tumor immune evasion, is a more reliable biomarker for predicting immunotherapy response than TMB or PD-L1 expression (23). Further examination of the association between ERGPS and TIDE scores revealed that the TIDE scores and Dysfunction scores were considerably lower in the high-risk group compared to the low-risk group (Figures 8E, F). Since patients with a lower TIDE score are more likely to have a lesser chance of antitumor immune escape, immune checkpoint blockade (ICB) treatment has a better response rate (24). These data imply that immunotherapy was more effective for HCC patients with higher risk scores in the training set.
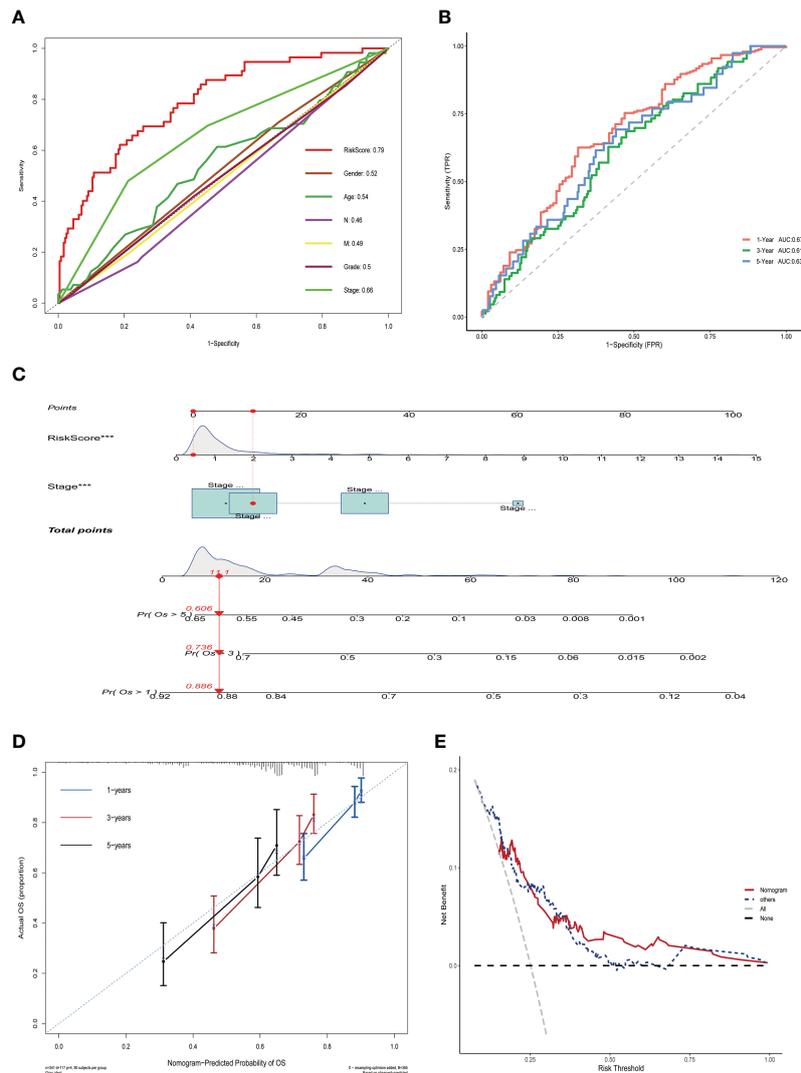
**FIGURE 4**
Nomogram construction and validation in TCGA LIHC dataset. **(A)** ROC curves showing the comparative results of risk scores and other clinical factors for prognostic prediction.**(B)** ROC curves showing the results of nomogram for predicting 1-, 3-, and 5-year survival in HCC patients.**(C)** Nomogram constructed from Stage and risk score.**(D)** Calibration curves show strong predictive power of nomogram for 1-, 3-, and 5-year survival in HCC patients. **(E)** Decision curves show that the nomogram model outperforms other models.

## Independent external immunotherapy data for validation

To further explore the value of ERGPS in predicting immunotherapy response, two independent immunotherapy data (IMvigor210, GSE91061) were included in the study. The IMvigor210 dataset has 298 patients treated with anti-PD-L1 and the GSE91061 dataset contains 51 patients treated with anti-PD-L1 or Anti-CTLA4; both datasets include entire transcriptome sequencing data and matched clinical information. We calculated risk scores for each patient in both immunotherapy datasets using the same formula as in the training set, and divided patients receiving immunotherapy into low- and high-risk groups based on the median value of the risk scores. Intriguingly, we discovered a link between risk scores and immunotherapy response, with higher risk scores among patients with CR/PR than SD/PD (Figures 9A, B). Comparing the proportion of patients with CR/PR between the high-risk and low-risk groups, we

determined that the proportion of patients with CR/PR was greater in the high-risk group than in the low-risk group in both datasets (Figures 9C, D). These validate the important value of ERGPS in predicting immunotherapy response in oncology patients.

## Analysis of the correlation between ERGPS and common chemotherapeutics

To evaluate ERGPS in the clinic for HCC treatment, we attempted to explore associations between risk scores and the efficacy of administering common chemotherapeutics. A greater risk score was associated with a lower IC50 for anticancer medicines such as Midostaurin, Salubrinal, Tipifarnib, Etoposide,Embelin, and Doxorubicin, according to our study (Figure S8). These results suggest that patients in the high-risk category may be more responsive to these anticancer drugs than patients in the low-risk group.
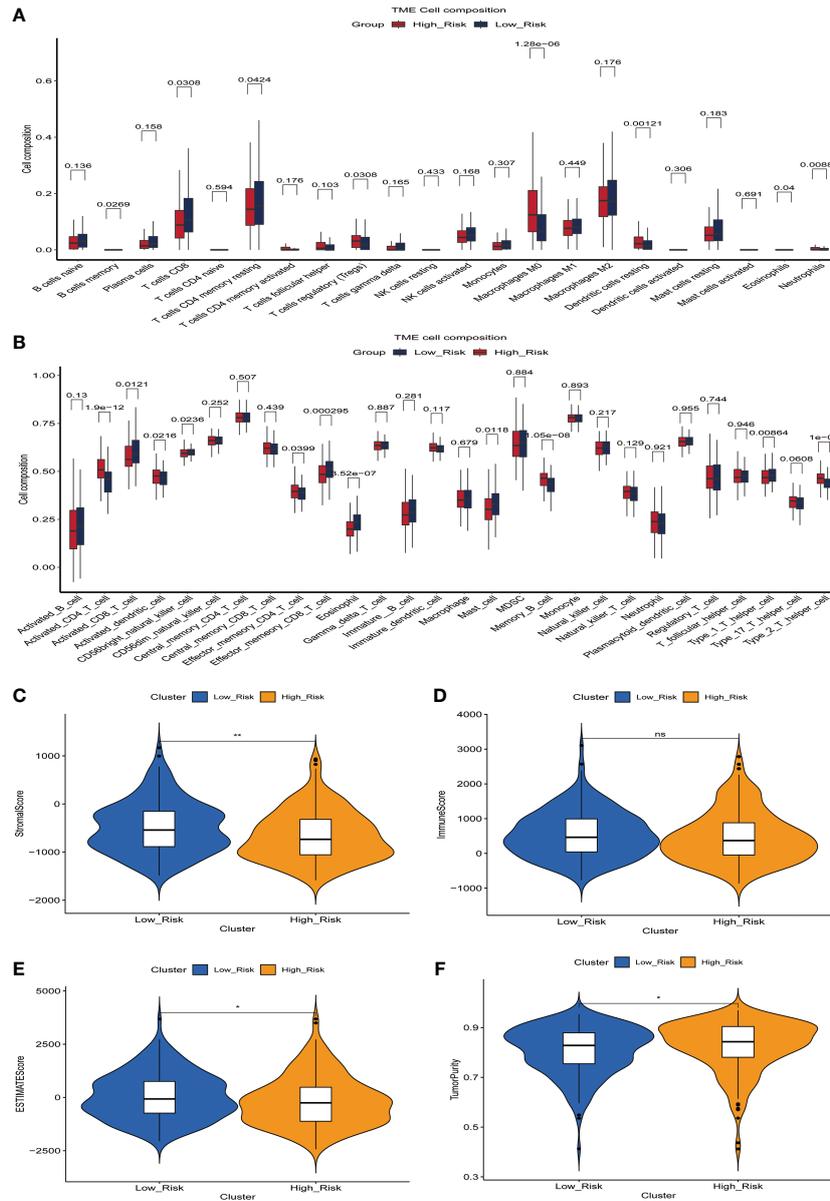
**FIGURE 5**

The relationship between prognostic signature and the proportion of immune cell infiltration in the tumor microenvironment. **(A)** Differences in the proportion of infiltrating immune cells between the high-risk and low-risk groups. **(B)** Differences in the proportion of 28 immune cell infiltrates in the tumor microenvironment between high- and low-risk groups. **(C-F)** Relationship between prognostic signature and StromalScore **(C)**, ImmuneScore **(D)**, ESTIMATEScore **(E)** and TumorPurity **(F)** in the training set.

## qRT-PCR confirmed overexpression of three estrogen-related gene in HCC tissues

To further validate the robustness of the prognostic signature, we collected tumor tissue samples and matched paraneoplastic samples from 12 patients with HCC confirmed by postoperative pathology. qRT-PCR results confirmed that the mRNA expression levels of all these three estrogen-related genes(AKR1B15, KCTD6, KPNA2) in ERGPS were significantly higher in HCC tissue than in paraneoplastic tissue(Figures 9E–G). These results confirmed the robustness of ERGPS.

## Discussion

In recent years, tyrosine kinase inhibitors (TKI), exemplified by sorafenib, have been able to extend the survival of a subset of HCC patients, but systemic therapy still confronts numerous obstacles (25, 26). With the unveiling of clinical studies of programmed death receptor 1 (PD-1) and cytotoxic T-lymphocyte-associated protein 4 (CTLA4) inhibitors, immune checkpoint inhibitor (ICI)-based immunotherapy has shown encouraging efficacy in multiple cancer types (27–29). Despite significant advancements in the use of immunotherapy for HCC, only a small percentage of patients benefit from it due to factors such as
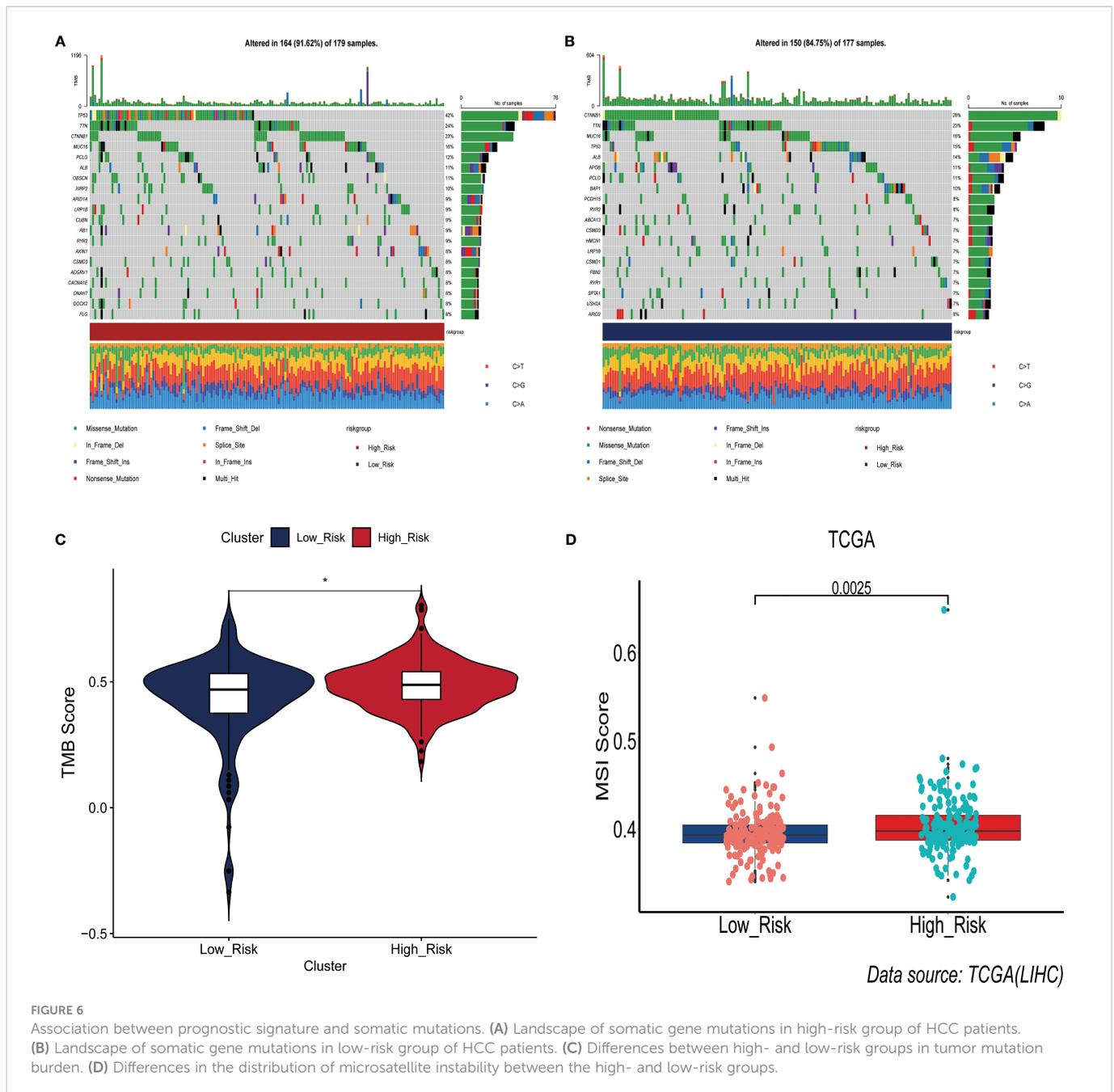
**FIGURE 6**
Association between prognostic signature and somatic mutations. **(A)** Landscape of somatic gene mutations in high-risk group of HCC patients. **(B)** Landscape of somatic gene mutations in low-risk group of HCC patients. **(C)** Differences between high- and low-risk groups in tumor mutation burden. **(D)** Differences in the distribution of microsatellite instability between the high- and low-risk groups.

decreased tumor immunogenicity, antigen-presenting mutations, missing mutations in signaling pathways, and decreased immune cell function in the tumor microenvironment (2, 30, 31). Some immunotherapy recipients exhibit no response or significant immune-related side effects, and a small number of immunotherapy recipients even develop HPD (32, 33). Therefore, there is an urgent need for effectively predictive biomarkers to assist in screening potential populations that could benefit from immunotherapy, further to guide the rational application of immune drugs in clinical practice, improve patient response rates, and reduce the risk of immunotoxicity in patients.

As a potent endogenous antioxidant, estrogen contributes to the development of a variety of liver disorders, including liver fibrosis, NAFLD, and hepatocellular carcinoma (34–36). In addition to being a consequence of disease progression, abnormal estrogen levels may also play a role in the pathophysiology of the development of chronic

liver disorders. Also, estrogen has non-reproductive effects as a regulator of the immune system, growth, neurological function, and metabolism. In particular, impaired expression and function of estrogen receptors in the liver are closely associated with obesity and liver-related metabolic dysfunction (37). A growing number of studies have established that estrogen signaling plays an immunosuppressive role in the tumor microenvironment, hence promoting tumor cell migration and invasion (38–40). Chronic liver disease and liver cancer are more prevalent in men (41), while men with cirrhosis and HCC have increased serum estrogen levels (42, 43). In addition, serum estrogen levels are elevated in patients undergoing surgical hepatectomy, suggesting the importance of estrogen regulation in the liver regeneration process. The mechanisms by which the liver senses and responds to estrogen to affect liver growth and cancer formation remain undetermined. In
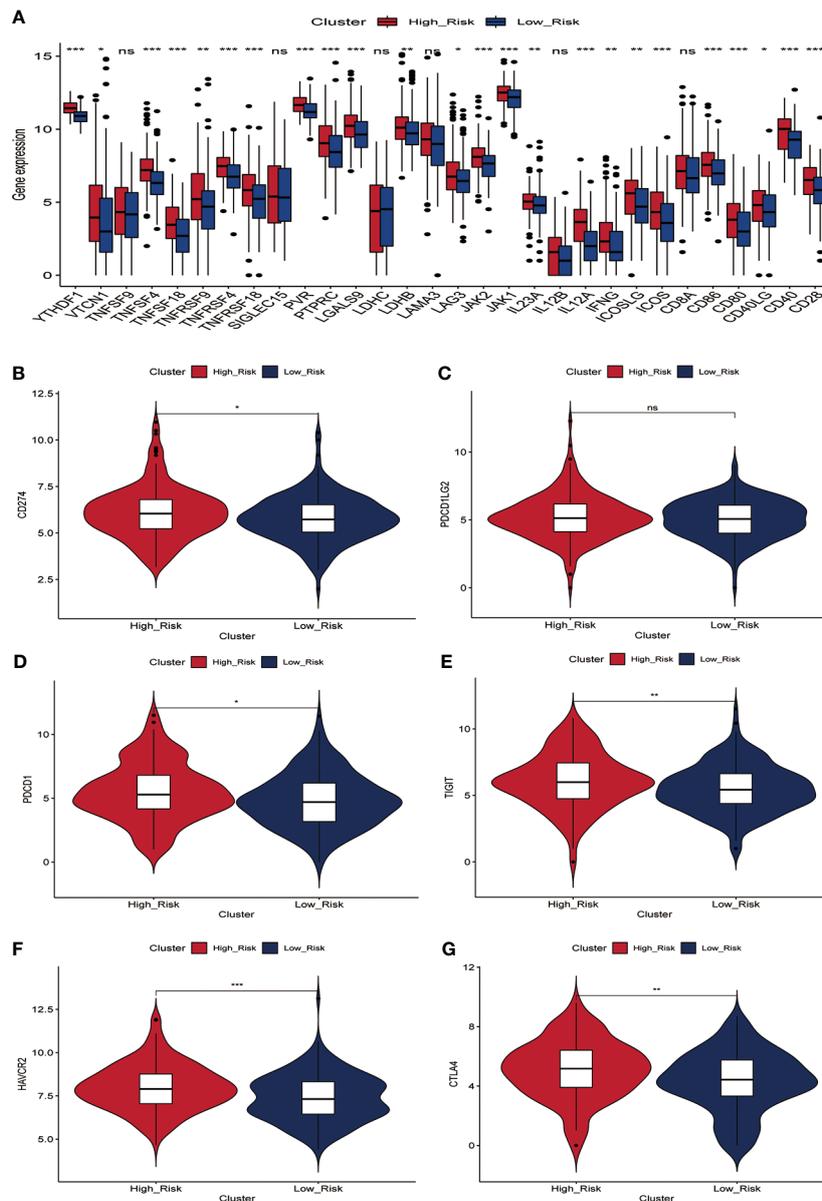
**FIGURE 7**
Relationship between prognostic signature and mRNA expression levels of immune checkpoint-associated genes in the training set. **(A)** The mRNA expression levels of most immune checkpoint-related genes were significantly higher in the high-risk group.**(B-G)** Differences in mRNA expression levels of six immune check-related genes commonly used in HCC immunotherapy between high- and low-risk groups, including PD-L1 **(B)**, PD-L2 **(C)**, PD1 **(D)**, TIGIT **(E)**, TIM-3 **(F)** and CTLA4 **(G)**.

this study, through a comprehensive analysis of transcriptome data from 365 HCC patients, we developed a novel estrogen-related gene prognostic signature (ERGPS) in the current study. To study the association between estrogen-related genes and immunotherapy, we undertook a complete profiling in terms of mRNA expression levels of immune cell infiltration, somatic mutations, TMB, MSI, TCR, BCR, SNV, TIDE, and immune checkpoint-related genes. In the microenvironment of HCC tumors, estrogen promotes the infiltration of immune cells with immunosuppressive functions while rejecting the invasion of immune cells with killing functions. In the meantime, the expression of estrogen-related genes may increase the mRNA expression of immune checkpoint-related genes. These findings indicate that estrogen has a significant role in

tumor cell growth, invasion, and immune cell evasion. Consequently, estrogen-related genes have the potential to serve as new immunotherapeutic targets in HCC.

In this study, ERGPS includes three estrogen-related genes (AKR1B15, KCTD6, and KPNA2), all of which are connected with the prognosis of HCC patients and are involved in estrogen production. AKR1B15 (Aldo-Keto Reductase Family 1 Member B15) is a Protein Coding gene (44). Its related pathways are Metabolism and Metabolism of steroid hormones. Studies have shown that AKR1B10 can induce a variety of cancers, such as HCC (45), non-small cell lung cancer (46), and pancreatic cancer (47), and is a promising potential cancer target. KCTD6 (Potassium Channel Tetramerization Domain Containing 6) is a Protein Coding gene.
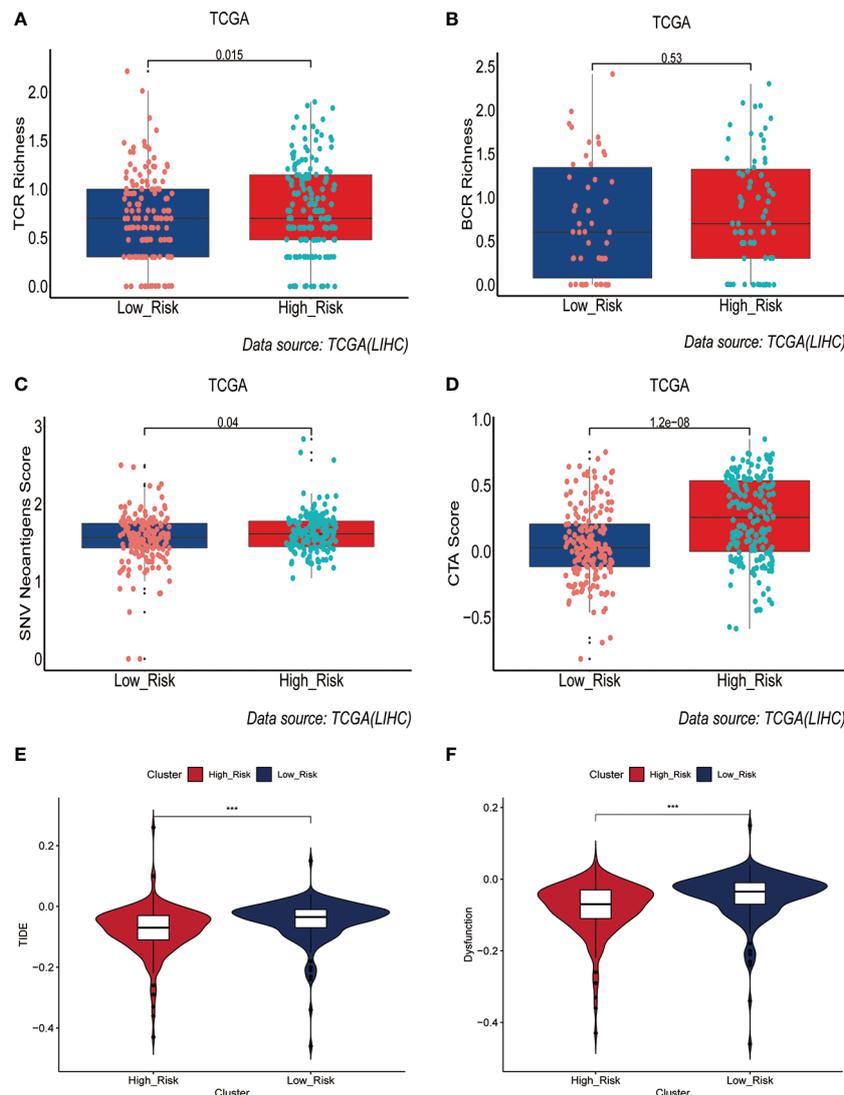
**FIGURE 8**

Prediction of immunotherapy response by prognostic signature in the training set. **(A-D)** Differences in TCR richness **(A)**, BCR richness **(B)**, SNV Neoantigens **(C)** and CTA scores **(D)** between the high-risk and low-risk groups. **(E-F)** TIDE scores **(E)** and Dysfunction scores **(F)** were used to assess immunotherapy response.

Diseases associated with KCTD6 include Medulloblastoma. Among its related pathways are Sweet Taste Signaling and Class I MHC-mediated antigen processing and presentation (48). There are currently few investigations on the correlation between KCTD6 and cancer. We validated KCTD6 overexpression in clinical samples using RT-PCR, and survival analysis revealed that the overall survival time of HCC patients in the overexpression KCTD6 group was significantly worse than in the low expression group, indicating that KCTD6 overexpression may promote tumor progression in HCC. Karyopherin α2 (KPNA2) belongs to the karyopherin family, which plays a crucial role in nucleocytoplasmic transport (49). Guo et al (50) verified that the overexpression of KPNA2 can accelerate the proliferation and migration of tumor cells and is related to a poor prognosis in patients with HCC. These findings demonstrate that overexpression of estrogen-related genes can increase tumor cell proliferation and invasion, and that these genes are prospective immunotherapy targets for HCC.

In this study, the prognosis signature proved to be a powerful prediction tool for the prognosis of HCC patients in the training set and verification set. The outstanding predicted accuracy of the prognostic signature motivated us to investigate the underlying molecular pathways. Initially, we did GO and KEGG analyses to investigate the enrichment pathways of genes closely connected with ERGPS. The results indicated that these genes were primarily enriched in the CELL CYCLE, OOCYTE MEIOSIS signaling pathway. Therefore, the poorer prognosis of individuals with high-risk scores may be partially attributable to the aberrant regulation of the cell cycle, which is closely linked to tumor proliferation and progression. Tumor-infiltrating immune cells play an important role in the tumor microenvironment and greatly influence the prognosis of HCC patients. We used the CIBERSORT algorithm of the R software to calculate the proportion of immune cells infiltrating the tumor microenvironment in the training set for each patient. Interestingly, we found that the prognostic signature was positively
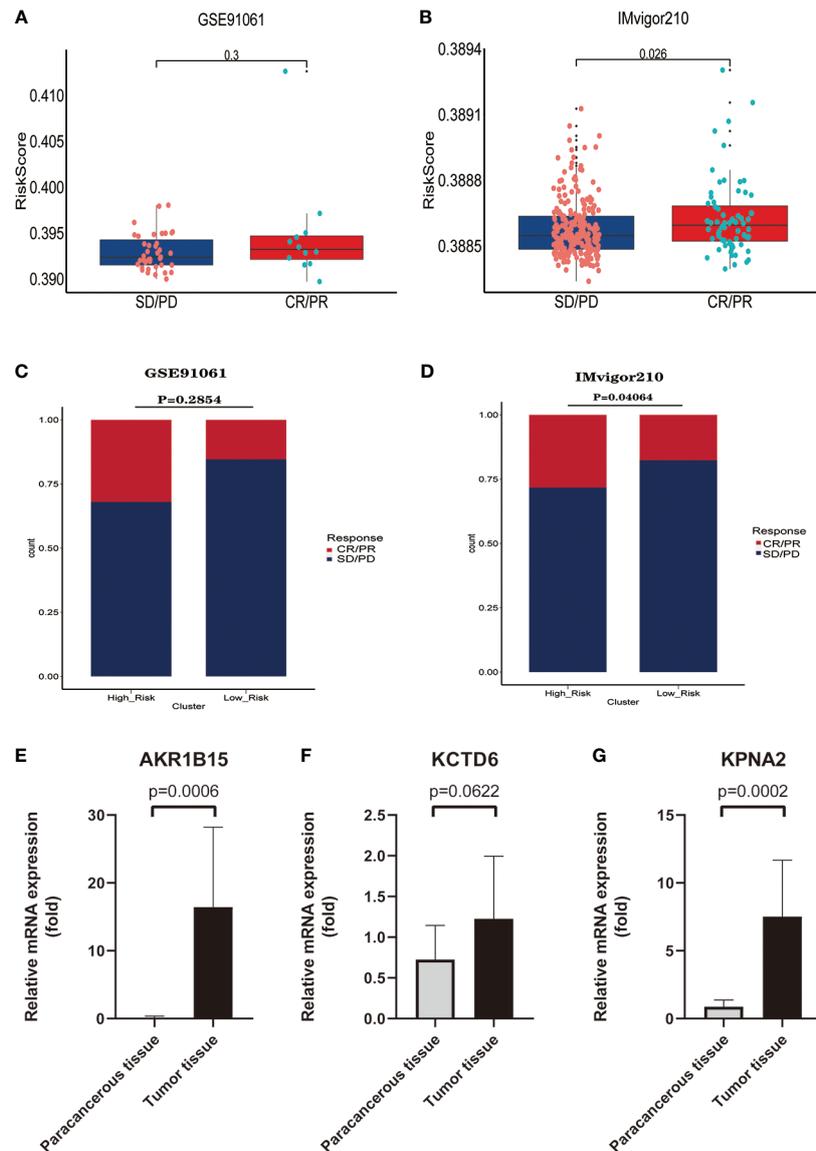
FIGURE 9
Independent external immunotherapy data and RT-qPCR techniques were used to validate the validity and robustness of the prognostic signature. **(A)** In the GSE91061 dataset, patients who responded to immunotherapy (CR/PR) had a higher median risk score than patients in the low-risk group.**(B)** In the IMvigor210 dataset, patients who responded to immunotherapy (CR/PR) had significantly higher risk scores than patients in the low-risk group.**(C)** In the GSE91061 dataset, the percentage of patients who responded to immunotherapy (CR/PR) was higher in the high-risk group than in the low-risk group. **(D)** In the IMvigor210 dataset, the percentage of patients who responded to immunotherapy (CR/PR) was significantly higher in the high-risk group than in the low-risk group.**(E-G)** qRT-PCR results confirmed that the three estrogen-related genes used to construct the prognostic signature were overexpressed in HCC tissues, including AKR1B15 **(E)**,KCTD6 **(F)** and KPNA2 **(G)**.

correlated with the proportion of cells with immunosuppressive functions, such as Treg cells, dendritic cells, and macrophages. These findings indicate that overexpression of estrogen-related genes may recruit these cells to the tumor microenvironment, thereby exerting an immunosuppressive impact and allowing tumor cells to evade immune cell surveillance. In conclusion, with all the above findings, we infer that the powerful predictive power of ERGPS may lie in cell cycle dysregulation and immunosuppressive microenvironment. Therefore, inhibition of estrogen-related gene overexpression may inhibit the convergence of immunosuppressive immune cells to the tumor microenvironment and thus exert an anti-tumor effect.

The success of immunotherapy for HCC is related to many factors, such as the immunogenicity of tumor cells, mRNA expression levels of immune checkpoint-related genes, and the functionality of tumor-infiltrating T cells. The mRNA expression level of immune checkpoint-related genes is a well-recognized biological marker for determining the response to immunotherapy. We first explored the relationship between the mRNA expression levels of immune checkpoint-related genes and ERGPS, and interestingly, we found that ERGPS was positively correlated with the mRNA expression levels of most immune checkpoint-related genes. The mRNA expression levels of immune checkpoint-related genes were significantly higher in the high-risk group than in the low-

risk group, especially for immune checkpoint-related genes commonly used in liver cancer immunotherapy, including (PD-L1, PDCD1LG2, PDCD1, TIGIT, TIM-3, CTLA4). These results further suggest that immune cells in the tumor microenvironment of patients in the high-risk group form an immunosuppressive microenvironment through immune checkpoint interactions with immunomodulatory cells with immunosuppressive functions, allowing tumor cells to evade the immune system; immunotherapy can block this interaction to reverse immunosuppressive state. Patients in the high-risk group are consequently more likely to respond to immunotherapy than those in the low-risk group. TMB and MSI have been demonstrated to be valuable tools for predicting immunotherapy response in a range of solid cancers. In this study, we investigated the relationship between ERGPS, TMB, and MSI. The results indicated that ERGPS was positively correlated with TMB and MSI, and that the TMB and MSI scores in the high-risk group were significantly higher than those in the low-risk group, indicating that tumor cells in the high-risk group had a higher immunogenicity. SNV and CTA levels can partially reflect the immunogenicity of tumor cells. In this study, patients in the high-risk group had higher SNV and CTA scores than those in the low-risk group, indicating that the high-risk group's tumor cells were more immunogenic. TCR and BCR are unique molecules on the surface of T and B cells, respectively, which recognize MHC-presented antigens. Studies of TCR, and BCR profiles in multiple cancer types have shown that TCR and BCR profiles can be used as predictive biomarkers of response to treatment with CTLA-4 or PD-1 inhibitors ([19], [20], [22]). We evaluated the differences in TCR richness and BCR richness between the high-risk and low-risk groups and found that the high-risk group had significantly higher TCR abundance and BCR richness than the low-risk group, which reflected greater functionality of T cells and B cells in recognizing antigens and killing tumor cells in high-risk HCC patients. In addition, the TIDE prediction score can predict the effectiveness of immunotherapy for a wide range of cancers. In our study, we observed that patients in the high-risk group had lower TIDE scores compared to the low-risk group, suggesting that patients with HCC in the high-risk group had a better response to immunotherapy. To further validate our conclusions, we used two independent external immunotherapy data (Imvigor210, GSE91061) for validation. The results revealed a larger proportion of immunotherapy responders in the high-risk group, indicating that high-risk patients are more likely to benefit from immunotherapy. With more confirmation, ERGPS has the potential to serve as a reliable biomarker for predicting immunotherapy response in HCC.

Despite the encouraging results, the study has some limitations. First, our study was conducted based on a public database and lacks validated experiments to validate the molecular mechanisms behind the prognostic signature. Secondly, the validity of our prognostic siganture was not validated with a large number of clinical samples. Finally, all the mechanistic analyses in this study are descriptive in this study. A large number of future studies are needed for validation.

In conclusion, a three-gene signature based on estrogen-related genes was identified and validated to have powerful performance in predicting prognosis and immunotherapy in HCC patients. It may provide a deeper understanding and new insights into developing novel immunotherapies for HCC. Also, it can be used as a prognostic biomarker for individualized prediction of clinical decisions and to help screen appropriate patients who may benefit from immunotherapy.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

## Author contributions

## Acknowledgments

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2023.1114717/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Flowchart of the data analysis procedures.

SUPPLEMENTARY FIGURE 2
The relationship between risk scores and clinical characteristics in the training set. (A−H) Relationship between risk score and clinical characteristics, including age (A), gender (B), N stage (C), M stage (D), tumor grade (E), tumor stage(F), MVI (G) and survival status (H).

SUPPLEMENTARY FIGURE 3

Validation of prognostic signature in clinical subgroups. **(A-F)** The validity of the prognostic signature was validated in different clinical subgroups, including young **(A)**, old **(B)**, male **(C)**, female **(D)**, MVI negative **(E)** and MVI positive **(F)**.

SUPPLEMENTARY FIGURE 4

The predictive efficacy of prognostic signature was verified in different clinical subgroups. **(A-D)** Kaplan-Meier curves showed that in different clinical subgroups, such as low tumor grade **(A)**, high tumor grade **(B)**, early stage **(C)** and advanced stage **(D)**, the overall survival time of low-risk group was significantly better than that of high-risk group.

SUPPLEMENTARY FIGURE 5

Exploration of the molecular mechanisms behind the prognostic signature. **(A)** The volcano map demonstrates that most genes are highly expressed in the high-risk group. **(B)** Results of GO functional enrichment analysis of genes highly expressed in the high-risk group. **(C)** Results of GO functional enrichment analysis of genes highly expressed in the low-risk group. **(D)** Results of GSEA KEGG functional enrichment analysis in the high-risk and low-risk groups.

SUPPLEMENTARY FIGURE 6

Results of HALLMARKER pathway enrichment analysis in the high-risk and low-risk groups. **(A-C)** The enrichment results of HALLMARKER pathway showed that E2F_TARGETS **(A)**, G2M_CHECKPOINT **(B)** and DNA_REPAIR **(C)** were mainly enriched in high-risk group, while COAGULATION **(D)** was mainly enriched in low-risk group

SUPPLEMENTARY FIGURE 7

Heatmap shows that most immune checkpoint-associated genes are overexpressed in high-risk groups.

SUPPLEMENTARY FIGURE 8

Drug sensitivity prediction in the training set. **(A-F)** The results of drug sensitivity analysis showed a higher risk score was related to lower IC50 among antitumor drugs, such as Midostaurin **(A)**, Salubrinal **(B)**, Tipifarnib **(C)**, Etoposide **(D)**, Embelin **(E)** and Doxorubicin **(F)**.

# References

1. Sung H, Ferlay J, Siegel RL, Sung MW, Baron AD, Kudo M, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660

2. Llovet JM, Kudo M, Cheng A, Laversanne M, Soerjomataram I, Jemal A, et al. Lenvatinib (len) plus pembrolizumab (pembro) for the first-line treatment of patients (pts) with advanced hepatocellular carcinoma (HCC): Phase 3 LEAP-002 study. *J Clin Oncol* (2019) 37:S4152. doi: 10.1200/JCO.2019.37.15_suppl.TPS4152

3. Llovet JM, Montal R, Sia D, Finn RS. Molecular therapies and precision medicine for hepatocellular carcinoma. *Nat Rev Clin Oncol* (2018) 15(10):599–616. doi: 10.1038/s41571-018-0073-4

4. Cheng A, Finn RS, Qin S, Han K, Ikeda K, Piscaglia F, et al. Phase III trial of lenvatinib (LEN) vs sorafenib (SOR) in first-line treatment of patients (pts) with unresectable hepatocellular carcinoma (uHCC). *J Clin Oncol* (2017) 35:4001. doi: 10.1200/JCO.2017.35.15_suppl.4001

5. Breder VV, Vogel A, Merle P, Finn RS, Galle PR, Zhu AX, et al. IMbrave150: Exploratory efficacy and safety results of hepatocellular carcinoma (HCC) patients (pts) with main trunk and/or contralateral portal vein invasion (Vp4) treated with atezolizumab (atezo) + bevacizumab (bev) versus sorafenib (sor) in a global ph III study. *J Clin Oncol* (2021) 39:4073. doi: 10.1200/JCO.2021.39.15_suppl.4073

6. Hato T, Goyal L, Greten TF, Duda DG, Zhu AX. Immune checkpoint blockade in hepatocellular carcinoma: Current progress and future directions. *Hepatology* (2014) 60(5):1776–82. doi: 10.1002/hep.27246

7. Chen R, Li Q, Xu S, Ye C, Tian T, Jiang Q, et al. Modulation of the tumour microenvironment in hepatocellular carcinoma by tyrosine kinase inhibitors: From modulation to combination therapy targeting the microenvironment. *Cancer Cell Int* (2022) 22(1):73. doi: 10.1186/s12935-021-02435-4

8. Llovet JM, Castet F, Heikenwalder M, Maini MK, Mazzaferro V, Pinato DJ, et al. Immunotherapies for hepatocellular carcinoma. *Nat Rev Clin Oncol* (2022) 19(3):151–72. doi: 10.1038/s41571-021-00573-2

9. Zheng B, Zhu YJ, Wang HY, Chen L. Gender disparity in hepatocellular carcinoma (HCC): multiple underlying mechanisms. *Sci China Life Sci* (2017) 60(6):575–84. doi: 10.1007/s11427-016-9043-9

10. Batmunkh B, Choijookhuu N, Srisowanna N, Byambatsogt U, Synn OP, Noor AM, et al. Estrogen accelerates cell proliferation through estrogen receptor α during rat liver regeneration after partial hepatectomy. *Acta Histochem Cytochem* (2017) 50(1):39–48. doi: 10.1267/ahc.17003

11. Kao TL, Kuan YP, Cheng WC, Chang WC, Jeng LB, Yeh S, et al. Estrogen receptors orchestrate cell growth and differentiation to facilitate liver regeneration. *Theranostics* (2018) 8(10):2672–82. doi: 10.7150/thno.23624

12. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* (2015) 12(5):453–7. doi: 10.1038/nmeth.3337

13. Hänzelmann S, Castelo R, Guinney J. GSVA: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf* (2013) 14:7. doi: 10.1186/1471-2105-14-7

14. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* (2013) 4:2612. doi: 10.1038/ncomms3612

15. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. *Nat Commun* (2015) 6:8971. doi: 10.1038/ncomms9971

16. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res* (2018) 28(11):1747–56. doi: 10.1101/gr.239244.118

17. Fang XN, Fu LW. Predictive efficacy biomarkers of programmed cell death 1/Programmed cell death 1 ligand blockade therapy. *Recent Pat Anticancer Drug Discovery* (2016) 11(2):141–51. doi: 10.2174/1574892811666160226150506

18. Hellmann MD, Ciuleanu TE, Pluzanski A, Lee JS, Otterson GA, Audigier-Valette C, et al. Nivolumab plus ipilimumab in lung cancer with a high tumor mutational burden. *N Engl J Med* (2018) 378(22):2093–104. doi: 10.1056/NEJMoa1801946

19. Page DB, Yuan J, Redmond D, Wen YH, Durack JC, Emerson R, et al. Deep sequencing of T-cell receptor DNA as a biomarker of clonally expanded TILs in breast cancer after immunotherapy. *Cancer Immunol Res* (2016) 4(10):835–44. doi: 10.1158/2326-6066.CIR-16-0013

20. Sims JS, Grinshpun B, Feng Y, Ung TH, Neira JA, Samanamud JL, et al. Diversity and divergence of the glioma-infiltrating T-cell receptor repertoire. *Proc Natl Acad Sci U.S.A.* (2016) 113(25):E3529–37. doi: 10.1073/pnas.1601012113

21. Han Y, Li H, Guan Y, Huang J. Immune repertoire: A potential biomarker and therapeutic for hepatocellular carcinoma. *Cancer Lett* (2016) 379(2):206–12. doi: 10.1016/j.canlet.2015.06.022

22. Hu Q, Hong Y, Qi P, Lu G, Mai X, Xu S, et al. Atlas of breast cancer infiltrated b-lymphocytes revealed by paired single-cell RNA-sequencing and antigen receptor profiling. *Nat Commun* (2021) 12(1):2186. doi: 10.1038/s41467-021-22300-2

23. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med* (2018) 24(10):1550–8. doi: 10.1038/s41591-018-0136-1

24. Wang Q, Li M, Yang M, Yang Y, Song F, Zhang W, et al. Analysis of immune-related signatures of lung adenocarcinoma identified two distinct subtypes: Implications for immune checkpoint blockade therapy. *Aging (Albany NY)* (2020) 12(4):3312–39. doi: 10.18632/aging.102814

25. Huang A, Yang XR, Chung WY, et al. Targeted therapy for hepatocellular carcinoma. *Signal Transduct Target Ther* (2020) 5(1):146. doi: 10.1038/s41392-020-00264-x

26. Keating GM. Sorafenib: A review in hepatocellular carcinoma. *Target Oncol* (2017) 12(2):243–53. doi: 10.1007/s11523-017-0484-7

27. Steven A, Fisher SA, Robinson BW. Immunotherapy for lung cancer. *Respirology* (2016) 21(5):821–33. doi: 10.1111/resp.12789

28. Li K, Zhang A, Li X, et al. Advances in clinical immunotherapy for gastric cancer. *Biochim Biophys Acta Rev Cancer* (2021) 1876(2):188615. doi: 10.1016/j.bbcan.2021.188615

29. Jadoon Y, Siddiqui MA. Immunotherapy in multiple myeloma. *Cancer Treat Res Commun* (2021) 29:100468. doi: 10.1016/j.ctarc.2021.100468

30. Zhu AX, Finn RS, Ikeda M, Sung MW, Baron AD, Kudo M, et al. A phase ib study of lenvatinib (LEN) plus pembrolizumab (PEMBRO) in unresectable hepatocellular carcinoma (uHCC). *J Clin Oncol* (2020) 38:4519. doi: 10.1200/JCO.2020.38.15_suppl.4519

31. Abou-Alfa GK, Chan SL, Kudo M, et al. Phase 3 randomized, open-label, multicenter study of tremelimumab (T) and durvalumab (D) as first-line therapy in patients (pts) with unresectable hepatocellular carcinoma (uHCC): HIMALAYA. *J Clin Oncol* (2022) 40:379. doi: 10.1200/JCO.2022.40.4_suppl.379

32. Yang X, Chen SL, Lin CS, et al. Tyrosine metabolic enzyme HPD is decreased and predicts unfavorable outcomes in hepatocellular carcinoma. *Pathol Res Pract* (2020) 216(11):153153. doi: 10.1016/j.prp.2020.153153

33. Zhang L, Wu L, Chen Q, et al. Predicting hyperprogressive disease in patients with advanced hepatocellular carcinoma treated with anti-programmed cell death 1 therapy. *EClinicalMedicine* (2021) 31:100673. doi: 10.1016/j.eclinm.2020.100673

34. Ezhilarasan D. Critical role of estrogen in the progression of chronic liver diseases. *Hepatobil Pancreat Dis Int* (2020) 19(5):429–34. doi: 10.1016/j.hbpd.2020.03.011

35. Jiang M, Klein M, Zanger UM, et al. Inflammatory regulation of steroid sulfatase: A novel mechanism to control estrogen homeostasis and inflammation in chronic liver disease. *J Hepatol* (2016) 64(1):44–52. doi: 10.1016/j.jhep.2015.07.022

36. Xu L, Yuan Y, Che Z, et al. The hepatoprotective and hepatotoxic roles of sex and sex-related hormones. *Front Immunol* (2022) 13:939631. doi: 10.3389/fimmu.2022.939631

37. El MKT, Abd EBE, Mohamed EM, et al. Relation between sex hormones and hepatocellular carcinoma. *Andrologia* (2016) 48(9):948–55. doi: 10.1111/and.12536

38. Abdelrazek H, Mahmoud M, Tag HM, et al. Soy isoflavones ameliorate metabolic and immunological alterations of ovariectomy in female wistar rats: Antioxidant and estrogen sparing potential. *Oxid Med Cell Longev* (2019) 2019:5713606. doi: 10.1155/2019/5713606

39. Müller ST, Keiler AM, Kräker K, et al. Influence of estrogen on individual exercise motivation and bone protection in ovariectomized rats. *Lab Anim* (2018) 52(5):479–89. doi: 10.1177/0023677218756455

40. Chaturantabut S, Shwartz A, Evason KJ, et al. Estrogen activation of G-Protein-Coupled estrogen receptor 1 regulates phosphoinositide 3-kinase and mTOR signaling to promote liver growth in zebrafish and proliferation of human hepatocytes. *Gastroenterology* (2019) 156(6):1788–804. doi: 10.1053/j.gastro.2019.01.010

41. Sato N, Lindros KO, Baraona E, et al. Sex difference in alcohol-related organ injury. *Alcohol Clin Exp Res* (2001) 25:40S–5S. doi: 10.1097/00000374-200105051-00007

42. Guéchot J, Peigney N, Ballet F, et al. Sex hormone imbalance in male alcoholic cirrhotic patients with and without hepatocellular carcinoma. *Cancer* (1988) 62(4):760–2. doi: 10.1002/1097-0142(19880815)62:4<760::aid-cncr2820620420>3.0.co;2-6

43. Castagnetta LA, Agostara B, Montalto G, et al. Local estrogen formation by nontumoral, cirrhotic, and malignant human liver tissues and cells. *Cancer Res* (2003) 63 (16):5041–5.

44. Weber S, Salabei JK, Möller G, et al. Aldo-keto reductase 1B15 (AKR1B15): A mitochondrial human aldo-keto reductase with activity toward steroids and 3-keto-acyl-CoA conjugates. *J Biol Chem* (2015) 290(10):6531–45. doi: 10.1074/jbc.M114.610121

45. Tsuzura H, Genda T, Sato S, et al. Expression of aldo-keto reductase family 1 member b10 in the early stages of human hepatocarcinogenesis. *Int J Mol Sci* (2014) 15 (4):6556–68. doi: 10.3390/ijms15046556

46. Fukumoto S, Yamauchi N, Moriguchi H, et al. Overexpression of the aldo-keto reductase family protein AKR1B10 is highly correlated with smokers' non-small cell lung carcinomas. *Clin Cancer Res* (2005) 11(5):1776–85. doi: 10.1158/1078-0432.CCR-04-1238

47. Chung YT, Matkowskyj KA, Li H, et al. Overexpression and oncogenic function of aldo-keto reductase family 1B10 (AKR1B10) in pancreatic carcinoma. *Mod Pathol* (2012) 25(5):758–66. doi: 10.1038/modpathol.2011.191

48. Ji AX, Chu A, Nielsen TK, et al. Structural insights into KCTD protein assembly and Cullin3 recognition. *J Mol Biol* (2016) 428(1):92–107. doi: 10.1016/j.jmb.2015.08.019

49. Noetzel E, Rose M, Bornemann J, et al. Nuclear transport receptor karyopherin-α2 promotes malignant breast cancer phenotypes *in vitro*. *Oncogene* (2012) 31(16):2101–14. doi: 10.1038/onc.2011.403

50. Guo X, Wang Z, Zhang J, et al. Upregulated KPNA2 promotes hepatocellular carcinoma progression and indicates prognostic significance across human cancer types. *Acta Biochim Biophys Sin (Shanghai)* (2019) 51(3):285–92. doi: 10.1093/abbs/gmz003