A hypoxia-related signature for clinically predicting diagnosis, prognosis and immune microenvironment of hepatocellular carcinoma patients.

Baohui Zhang
Department of Physiology, School of Life Science, China Medical University,Shenyang 110122,China

Bufu Tang
Department of Radiology, Second Affiliated Hospital, School of Medicine,Zhejiang University,Hangzhou 310058, China.

Jianyao Gao
Department of Radiation Oncology, the First Affiliated Hospital of China Medical University,Shenyang, China.

Jiatong Li
Department of Orthopedics, The First Affiliated Hospital of China Medical University,Shenyang, 110001, China.

Lingming Kong
Department of General Surgery, Shengjing Hospital of China Medical University, Shenyang 110004, China

Ling Qin (qinlingling@yahoo.com)
Department of Physiology, China Medical University,Shenyang, People's Republic of China
https://orcid.org/0000-0001-8717-3833

Research

Keywords: hypoxia, hepatocellular carcinoma, prognostic, diagnostic, immune microenvironment

Posted Date: March 24th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-17783/v1

Version of Record: A version of this preprint was published on September 4th, 2020. See the published version at https://doi.org/10.1186/s12967-020-02492-9.

Abstract

Background Hypoxia plays an indispensable role in the development of hepatocellular carcinoma (HCC). However, there are few studies on the application of hypoxia molecules in the prognosis predicting of HCC. We aimed to identify the hypoxia-related genes in HCC and construct reliable models for diagnosis, prognosis and recurrence of HCC patients as well as exploring the potential mechanism.

Methods Differentially expressed genes (DEGs) analysis was performed using The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) database and four clusters were determined by a consistent clustering analysis. Three DEGs closely related to overall survival（OS）were identified using Cox regression and LASSO analysis and the hypoxia-related signature was developed and validated in TCGA and International Cancer Genome Consortium (ICGC) database. Then the Gene Set Enrichment Analysis (GSEA) was performed to explore signaling pathways regulated by the signature and the CIBERSORT was used for estimating the fractions of immune cell types.

Results A total of 397 hypoxia-related DEGs were detected and three genes (PDSS1, CDCA8 and SLC7A11) were selected to construct a prognosis, recurrence and diagnosis model. Then patients were divided into high- and low-risk groups. Our hypoxia-related signature was significantly associated with worse prognosis and higher recurrence rate. The diagnostic model also accurately distinguished HCC from normal samples and nodules. Furthermore, the hypoxia-related signature could positively regulate immune response and the high-risk group had higher fractions of macrophages, B memory cells and follicle-helper T cells, and exhibited higher expression of immunocheckpoints such as PD1and PDL1.

Conclusions Altogether, our study showed that hypoxia-related signature is a potential biomarker for diagnosis, prognosis and recurrence of HCC, and it provided an immunological perspective for developing personalized therapies.

Background

Hepatocellular carcinoma (HCC) accounts for 85% of liver cancers, and the disease burden of HCC is increasing globally[1]. Although progress on treatment strategies for HCC has been made, the overall 5-year survival rate for HCC patients remains less than 20%[2]. Nowadays, the research of molecular mechanism based on bioinformatics analysis has become one of the most important tools for cancer research[3, 4]. Therefore, it is of great significance to search for molecular markers for early diagnosis, survival prediction and recurrence monitoring of HCC, which can improve patients' stratification and optimize medical intervention. The low rate of early diagnosis and high rate of metastasis and recurrence have considerable impact on the prognosis of HCC patients, which are mainly related to the invasiveness and high proliferative activity of tumor cells[5]. However, the mechanism of tumor progression has not been completely realized.

Hypoxia is an intrinsic characteristic of solid tumors due to the imbalance between the rate of tumor cell proliferation and nutrient supply of vascular[6, 7]. Existing studies have recognized the critical roles

played by hypoxia on tumor angiogenesis, cell proliferation, as well as cell differentiation and apoptosis[8, 9]; Liver is one of the three organs most susceptible to hypoxia and it has been found that hypoxia was involved in the metastasis, poor prognosis and radiation resistance of HCC [10, 11]. Nevertheless, its potential regulatory mechanism remains unclear. In recent years, there is an increasing interest in the tumor microenvironment which immune cells in it play a crucial role in the progression of tumor[12–14]. Previous studies have shown that hypoxia can regulate the status of tumor immune microenvironment, such as promoting the recruitment of innate immune cells and interfering with the differentiation and function of adaptive immune cells[15]. Therefore, further study on the relationship between hypoxia and immunity in HCC is required in order to develop new therapeutic strategies.

Immunocheckpoint inhibition has become an effective and frequently-used way of immunotherapy[16]. As a new feature of cancer, tumor mutation burden (TMB) is defined as the total number of somatic mutations in the genome of tumor cells[17], and high TMB may produce many neoantigens to stimulate the anti-tumor immune response[17, 18]. Clinical data demonstrated that patients with high TMB were more likely to benefit from immunocheckpoint inhibitor therapy[19, 20], which suggesting that TMB should be an appropriate biomarker for assessing the effect of immune treatment.

In this study, we analyzed hypoxia-related genes in HCC by using TCGA and GEO database and constructed a consistent clustering. Then we built the prediction model for diagnosis, recurrence and prognosis of HCC. We also explored the association of hypoxia with immune infiltration and immunocheckpoints in HCC. These findings may make a meaningful contribution to the development of comprehensive therapeutic strategies for HCC patients.

Methods

Identification of differentially expressedgenes (DEGs) between HCC and noncancer tissues

The differentially expressed genes (DEGs) related to hypoxia and HCC were identified with limma, an R package[21]. The DEGs with an absolute log2-fold change (FC) > 1 and an adjusted P value < 0.05 were considered for further analysis.

Acquisition of hypoxia-related genes associated with HCC

The mRNA expression profiles and corresponding clinical information associated with HCC patients were obtained from The Cancer Genome Atlas—Liver Hepatocellular Carcinoma dataset (TCGA-LIHC)(including 370 HCC and 50 normal tissue samples). The mRNA-sequencing data of Human HCC cell lines were obtained from the Gene Expression Omnibus database (GEO), which included GSE59729 (with gene expression profiles of Huh-7 cells under normoxia and hypoxia for 24 hours) and GSE41666 (with gene expression profiles of HepG2 cells exposed to normoxia and hypoxia for 24 hours). A total of 1,401 hypoxia-related DEGs expressed by HepG2 from GSE41666 and 1,279 hypoxia-related DEGs expressed by Huh7 from GSE59729 were matched with HCC-related information obtained from TCGA. The data from TCGA and GEO databases are freely available to the public, and this research also strictly followed

access policies and publication guidelines, therefore this study did not require ethical review and approval from an Ethics Committee.

Classification of molecular subgroups by consistent clustering

The ConsensusClusterPlus package in R software was utilized for the consistent clustering to determine subgroups of HCC samples from TCGA. The Euclidean squared distance metric and the K-means clustering algorithm was used for classifying samples into k clusters with k=2 to k=9. About 80% of the samples were selected in each iteration, and the results were compiled over 100 iterations. The results are presented in the form of heatmaps of the consistency matrix generated by pheatmap R package, and the optimal number of clusters was determined by the consistent cumulative distribution function (CDF) graph and the delta region graph[22]. We considered that the optimal number of clusters should satisfy the following criteria: high consistency of clustering, low coefficient of variation, and no significant increase in the area under the CDF curve. According to the relative non-significant change of the area under the CDF curve, the corresponding number of categories was determined.

Establishment and validation of a prognostic predictive signature

The univariate Cox regression analysis was conducted to identify the prognostic value of the DEGs for OS and genes with a P value <0.05 were considered statistically significant. Subsequently the Least absolute shrinkage and selection operator (LASSO) Cox regression[23] was performed by using the glmnet R package to shrink scope of gene screening, we performed 1,000 substitution samples of the dataset and selected the markers with repeat occurrence frequencies of more than 900. Finally, a multivariate Cox regression analysis was performed to identify highly correlated genes and construct the prognostic gene signature. The regression coefficient (β) was derived from multivariate Cox regression analysis and the Prognosis Index (PI) = (β_{mRNA1}* expression level of $mRNA_1$) + (β_{mRNA2}* expression level of $mRNA_2$) +…+ (β_{mRNAn}* expression level of $mRNA_n$). Based on the optimal cut-off value determined by using X-tile software, patients with survival data were divided into high- and low-risk groups. The Kaplan-Meier survival analysis was used to evaluate the predictive ability of the prognostic model, which was further validated in the ICGC dataset.

Independence of the prognostic gene signature from other clinical characteristics

Univariate and multivariate Cox proportional hazard regression analyses were performed to determine whether the predictive ability of prognostic model was independent of conventional clinical characteristics. A bilateral P value <0.05 was considered statistically significant. The hazard ratio (HR) and 95% confidence intervals were calculated.

Construction and evaluation of a predictive nomogram

All independent prognostic factors were used to build a nomogram[24] in order to evaluate the 1-, 3-, and 5-year survival probability for patients with HCC. The calibration plot was performed for an internal

validation to verify the accuracy. Time-dependent receiver operating characteristic (ROC) analysis was conducted to evaluate the predictive performance of the nomogram. Decision curve analysis (DCA) was performed to assess the clinical net benefit[25].

Gene set enrichment analysis

Gene set enrichment analysis (GSEA)[26]was performed using prognosis index with Clusterprofiler package to identify signaling pathways regulated by the hypoxia-related signature. The correlation coefficients, CI and P-values were calculated using R software. P<0.05 was considered statistically significant.

Estimation of immune cell type fractions

CIBERSORT is a method for characterizing the cell composition from their gene expression profiles and is the most frequently cited tool for estimating and analyzing immune cells infiltration[27]. We utilized CIBERSORT to estimate the fractions of immune cell types between low- and high-risk groups. The sum of all the estimated immune cell type scores is equal to 1 in each sample.

Statistical analysis

All the results are presented as the mean ± standard deviation (SD). Differences between groups were compared by Wilcox test through R software. Different hypoxia subtypes were compared by using the Kruskal–Wallistest.

Results

Identification of DEGs related to hypoxia in HCC

We identified DEGs (|LogFC| >1, P < 0.05) using the mRNA expression profile between HCC and adjacent noncancerous tissues from TCGA database (Figure 1A). Then we matched the differentially expressed mRNA-sequencing data between hypoxia-treated and untreated HCC cell lines in GEO database and obtained 397 DEGs which were related to hypoxia in HCC (Figure 1B-D). By using the Gene Ontology (GO) enrichment and functional analysis, we found that these genes are enriched in DNA replication, cell division, cell cycle and also somatic diversification of immune receptors. (Figure 1E).

Using the hypoxia-related genes for the consistent clustering of HCC molecular subgroups

Consistent clustering of 397 hypoxia-related DEGs were constructed by using the ConsensusClusterPlus R software package. The average clustering consistency and inter-cluster variation coefficient of each cluster number were calculated and the optimal cluster number was determined by using CDF. As shown in Figure 2A, the clustering outcoming was stable when k=4. We further analyzed CDF delta area curve and found that the area under the CDF curve tended to be stable after 4 clusters (Figure 2B). The item-Consensus Plot also showed that the sample classification was relatively stable when the clustering

number was selected as 4 (Figure 2C). Finally, we built a consensus matrix graph which 397 DEGs were assigned to 4 clusters in order to evaluate the composition and quantity of clustering more intuitively (Figure 2D). The heatmap of 397 hypoxia-related DEGs in 4 clusters was shown in Figure 2E.

The results from Kaplan–Meier plot showed the significant differences in survival probability and recurrence rate among these 4 subgroups. Compared to the other three clusters, the samples in cluster-2 had the worst prognosis and the highest recurrence rate (Figure 3A-B). We further analyzed the distribution of AFP, gender, degree of vascular infiltration, TNM stage, pathological grade, and age in these 4 subgroups (Figure 3C). Samples in cluster-4 were associated with high AFP expression level, undifferentiated tumor cells and lymphatic metastasis while cluster-3 showed high incidence of distant metastasis; cluster-2 had a higher degree of vascular invasion and more tumor cells with low differentiation. Moreover, most of patients in cluster-2 were male and generally aged between 65 and 70 years. It is worth noting that patients in cluster-2 showed the highest TMB than other three clusters (Figure3D-E), suggesting a benefit of immunotherapy.

Construction and validation of a hypoxia-related prognosis signature with good performance

We performed a univariate Cox regression and found 291 DEGs significantly related to OS of HCC patients ($P<0.01$). Then a Lasso-penalized Cox analysis was performed to further shrink the scope of gene screening. The penalty parameter was established through 10-fold cross-validation. We selected 11 DEGs, which appeared over 900 times of a total of 1000 repetitions (Figure S1). Finally, we performed a multivariate Cox regression and selected three genes (PDSS1, SLC7A11, CDCA8) to build a prognostic model as follows: the prognostic index (PI) = (0.337 * expression level of PDSS1) + (0.383* expression level of SLC7A11) + (0.356* expression level of CDCA8). The optimal cut-off value of 2.296 for the risk store was produced using X-tile software and patients with survival time from TCGA were divided into a high- and low-risk group. The K-M curve showed that the OS of the high-risk group was significantly poorer than that of low-risk group ($P<0.001$, HR=4.76) (Figure 4A), and the high-risk group had higher expression of prognostic genes compared with the low-risk group (Figure 4C). The area under the time-dependent ROC curves (AUCs) for 0.5-, 1-, 3- and 5-year overall survival (OS) were 0.76, 0.78, 0.7 and 0.7, respectively, indicating a good predictive performance of this prognostic model (Figure 4D).

We further validated the prediction ability of this prognostic signature using HCC samples from ICGC database. Consistent with above results, HCC patients were divided into a high- and low-risk group with an optimal cut-off value of 18.812 and patients in the high-risk group had poorer survival probability than the low-risk group ($P<0.001$, HR=5.26) (Figure 4B). The risk-score distribution and gene expression were distinctly different (Figure 4E) and the AUCs of the three-gene prognostic model were 0.68, 0. 75, 0.77 and 0.77 for the 0.5-, 1-, 3- and 4-year survival times (Figure 4F). Taken together, our prognostic model showed a high specificity and sensitivity.

Evaluating the independent role of prognostic signature and building a predictive nomogram for OS prediction in the HCC cohort from TCGA

Univariate and multivariate Cox regression analysis were used to evaluate whether the predictive value of the prognostic model was independent of other traditional clinical characteristics. The results showed that the TNM stage (P<0.05, HR=1.828) and the risk score (P<0.05, HR=1.683) were independent prognostic factors for OS (Figure 5A). Then we built a predictive nomogram which may be helpful to accurately predict a certain clinical outcome (Figure 5B). Each level of independent factors was assigned one score and a total score was calculated by summing up the scores in each individual. The survival probability for the individuals at 1-, 3-, and 5- year was obtained through the function conversion relationship of total scores. The calibration plot for internal validation of the nomogram showed better consistency between the predicted OS outcomes and actual observations (Figure 5C-E). The C-index was 0.54, 0.65 and 0.66 for the TNM stage, the prognostic model and the nomogram (95%CI: 0.58–0.73), further indicating that our nomogram had a higher predicting consistency. The AUCs of the nomogram at 1-, 3- and 5- year OS were 0.672, 0.684 and 0.675, which were better than the models with single independent factors (Figure 5F-H). The DCA was used to evaluate guiding significance of these models for clinical application and the results showed that the combined model was the best for predicting the OS (Figure 5I-K).

Evaluation of the hypoxia-related genes for predicting the recurrence of HCC patients

TCGA-LIHC cohort with release-free survival (RFS) information and recurrent status of HCC patients was utilized as a training set for an independent evaluation, and the HCC cohort from GSE14520 was used as a validation set. Based on these three hypoxia-related genes, we constructed a recurrence signature by using the regression coefficient (β') of multivariate Cox proportional hazards. The prognostic index (PI) = (0.060 * expression level of PDSS1) + (0.045* expression level of SLC7A11) + (0.041* expression level of CDCA8). In both training and validation set, patients were divided into a high- and low-risk group based on the risk score of 0.953 and 1.247. From the results of Kaplan–Meier survival analysis, patients in high-risk group had significantly higher recurrence rate than the low-risk group. (Figure 6A-B). The distribution of risk score and gene expression was examined (Figure 6C, E), and we also performed ROC analysis to evaluate the predictive accuracy of our recurrence model (Figure 6D, F). All these results indicated a reliable predictive ability of our hypoxia-related recurrence model.

Building a nomogram for predicting recurrent probability of HCC patients and evaluating its predictive performance

We performed a univariate and multivariate Cox regression analysis and screened out three independent factors related to the recurrence of HCC (P<0.05) (including the age, the TNM stage and the risk score of our recurrence signature) (Figure 7A). The nomogram for recurrence prediction was built by integrating these three factors (Figure 7B) and the calibration plot of the nomogram showed a consistency between the prediction and observation (Figure 7C-E). The C-index was 0.62, 0.56, 0.63 and 0.71 for the age, TNM stage, the prognostic model and the nomogram (95%CI: 0.64–0.78). From the results of ROC analysis in Figure 7F-H, the AUCs of nomogram at 1-, 3-, 5-year was 0.746, 0.741, 0.717, respectively, which was obviously higher than other models with single independent factors. The DCA curves showed that the

combined model obtained a higher net benefit (Figure 7I-K). These results indicated that our recurrent nomogram performed a good sensitivity and specificity of HCC recurrence prediction and could provide clinicians with more specific guidelines.

Establishment of a diagnostic model based on hypoxia-related genes in HCC

As the diagnosis is of great importance for proper management of patients, we further analyzed whether hypoxia-related genes also contribute to more accurate diagnosis of HCC. A diagnostic model based on these three hypoxia-related genes was constructed by using a stepwise logistic regression method. The diagnostic score was finally identified as follows: logit (P = HCC) = 1.171 + (-0.571) × PDSS1 expression level + (-1.019) × SLC7A11 expression level + (-2.037) × CDCA8 expression level. In TCGA cohort with 50 normal samples paired 50 HCC samples, our diagnostic model achieved a sensitivity of 94% and a specificity of 92% (Figure 8A). We also utilized ICGC cohort with 190 normal samples paired 219 HCC samples as a validation set, and the diagnostic model obtained a sensitivity of 90% and a specificity of 94% (Figure 8B). As shown in ROC analysis (Figure 8C-D), the AUCs of our model reached 0.986 and 0.962 in TCGA and ICGC cohort, indicating a satisfactory accuracy of prediction. We correlated the predictive outcomes with the corresponding gene expression, the expression of PDSS1, SLC7A11, CDCA8 was upregulated in samples which were predicted the type of tumor (Figure 8E-F).

Liver nodule is a kind of hepatic hyperplasia caused by various factors. It is indistinguishable from the early stage of liver cancer, and the corresponding treatment methods are different. We aimed to establish a diagnostic model by using a stepwise logistic regression method to better distinguish liver cancer from hepatic nodules. The diagnostic score was identified as follows: logit (P = HCC) = -45.308 + 0.628 × PDSS1 expression level + 8.452 × SLC7A11 expression level + 4.047 × CDCA8 expression level. We tested the diagnostic performance of the model in two databases, GSE6764 and GSE89377 cohort. One achieved a sensitivity of 88.57% and a specificity of 82.35%, the other one achieved a sensitivity of 87.5% and a specificity of 77.27% (Figure 9A-B). The AUCs for GSE6764 and GSE89377 were 0.934 and 0.935 (Figure 9C-D). The gene expression corresponding with the predicted results was shown in Figure 9E-F. These data further confirmed that the diagnostic model was a novel predictive tool with high accuracy and potential clinical value.

Validation of the expression and genetic alterations and independent prognostic analysis for genes

We detected genetic alterations of the three genes from cBioportal database[28] and found that PDSS1, SLC7A11 and CDCA8 possessed genetic alterations of 9%, 3% and 5% (Figure 10A). These results helped explain that the abnormal gene expression may be attributable to genetic alterations. To further confirm the expression level of each gene in HCC, we used TCGA database containing 50 tumor and 50 normal samples. We found all the three genes were highly expressed in HCC compared with in normal liver tissues (Figure10B-D). Moreover, by analyzing gene expression in GSE6764 cohort, we found that the expression levels of PDSS1, CDCA8 and SLC7A11 were significantly higher in tumor tissue than those in liver nodules (Figure10E-G). We also attempted to explore the interaction between each two genes. As

shown in Figure10H-J, there was a sort of synergy between CDCA8 and PDSS1 as well as SLC7A11 ($P<0.05$).

Kaplan-Meier Plotter database[29] was used in order to analyze the effect of single gene on HCC prognosis. The results showed that the high-expression level of PDSS1, CDCA8 or SLC7A11 was separately related to a shorter overall survival time (Figure11A-C). In addition, the progression-free survival (PFS) analysis, which can better reflect tumor progression and predict clinical benefits, also showed an association between higher expression level of a single gene and faster disease progression (Figure11D-F). To achieve a better understanding of the functional characteristics of three genes, we performed Gene set enrichment analysis, which showed that some immune-related pathways, such as JAK–STAT3 signaling, TNF-NF-kappa B signaling, were highly active in the high-risk group (Figure11G-I).

Comparison of the immune microenvironment between high- and low-risk groups

Tumor immune cell infiltration refers that the immune cells move from the blood to the tumor tissue. The immune cells in tumors are closely related to clinical outcomes and they are most likely to serve as drug targets to improve survival rate [30]. Since these three genes have been found to enriched in some immune pathways, we then analyzed the relationship between hypoxia-related genes and immune cell infiltration as well as immune checkpoints in HCC (Figure12A-B). Patients in the high-risk group had higher ratios of M0 macrophages, memory B cells and follicular helper T cells than those in the low-risk group ($P <0.05$) (Figure12D-F). Moreover, we found that the expression levels of TIM3, B7H3, CTLA4, PD1 and PDL1 in the high-risk group were obviously higher than those in the low-risk group ($P <0.05$) (Figure12C, G-K). Our findings lead us to conclude that tumor immune microenvironment may be responsible for the prognosis of HCC patients with high expression of hypoxia-related genes.

Discussion

Hepatocellular carcinoma (HCC) is one of the leading causes of cancer-related death in the world, and the development of HCC is a complicated process influenced by various factors[31]. Though some progresses have been made in the treatments of HCC, such as surgical resection, microwave ablation and liver transplantation, the prognosis of HCC patients remains poor[32–34]. In recent years, high-throughput sequencing and data analysis have gradually become more significant tools for biomedical research, which can identify biomarkers for prognosis predicting, recurrence monitoring as well as clinical stratification[3, 35, 36]. Therefore, it is urgent to apply to HCC and explore key targets for the treatment.

Hypoxia is a prominent characteristic of malignant tumors, especially in HCC[37]. It was demonstrated in several studies that hypoxia was involved in the aggressive development of HCC[9]. Nevertheless, due to the multiple roles of hypoxia, the specific role of hypoxia in the development of liver cancer remains unclear [38]. In this study, we identified three hypoxia-related genes (PDSS1, CDCA8 and SLC7A11) closely relating to HCC. CDCA8, involving in protein metabolism and mitosis, has been demonstrated to participate in malignant progression of tumor cells and lead to poor prognosis in liver, stomach and lung cancer[39, 40]. SLC7A11 (also known as xCT) plays a critical role in maintaining redox homeostasis and

has been confirmed to be associated with the prognosis of HCC[41]. PDSS1 is involved in coenzyme Q biosynthesis, but little is known about the relationship between PDSSI and cancer[42]. Our results showed that this three-gene signature was an independent factor affecting the prognosis of HCC and the model had a better predictive performance on both prognosis and recurrence. What's more, the diagnostic model based on these three genes had a high sensitivity and specificity, and could help distinguish HCC from dysplastic nodules. Consensus Clustering is a common method for classification of cancer subtypes. We divided the samples into 4 clusters according to the hypoxia-related DEGs dataset of HCC and compared the differences among clusters. It should be pointed out that cluster-2 had a higher TMB, indicating that patients in cluster-2 were more likely to benefit from immunocheckpoint inhibitor therapy[43].

Much work so far has focused on the role of hypoxia in regulating the immune response in tumors. Hypoxia can interfere with the differentiation and function of immune cells through regulating the expression of co-stimulating receptors and the types of cytokines[44, 45]. The immune system is able to recognize and eliminate tumor cells through innate and adaptive mechanisms. However, the tumor microenvironment could suppress this anti-tumor response through a number of inhibitory pathways which were known as immunocheckpoints[46]. Our results of GSEA indicated that hypoxia-related signature could positively regulate some immune signaling pathways. The high-risk group based on the expression level of hypoxia-related genes had a higher infiltration proportion of macrophages, B memory cells and follicle-assisted T cell, as well as higher expression levels of immune checkpoints. These evidence for the association between hypoxia and immunity highlighted the importance of immunotherapy for HCC patients with high expression level of three hypoxia-related genes.

However, some limitations of this study should be noted. First, the process of adjusting the weight of regression coefficient in LASSO might ignore some important factors contributing to HCC prognosis. Second, our nomogram did not perform external validation as there was a lack of specific clinical data in ICGC database. Moreover, our retrospective findings need to be further validated in prospective research. Finally, the complex interaction between tumor cells and immune cells in hypoxic environments remains to be further explored.

Conclusion

In summary, we identified the hypoxia-related DEGs between HCC and normal tissues and clustered HCC samples into 4 subgroups. We established the diagnosis, prognosis and recurrence models based on three hypoxia-related genes, which performed favorable diagnosis and prediction performance for HCC. Finally, we identified higher proportions of immune cell infiltration and immunocheckpoint expression in the high-risk group, which may be more sensitive to benefit from immunotherapy.

Abbreviations

HCC Hepatocellular carcinoma

TCGA The Cancer Genome Atlas

GEO Gene Expression Omnibus

ICGC International Cancer Genome Consortium

DEGs Differentially expressed genes

OS Overall survival

ROC Receiver operating characteristic

AUC Area under the ROC curve

DCA Decision curve analysis

PD1 Programmed cell death-1

PDL1 Programmed cell death 1 ligand 1

GSEA Gene Set Enrichment Analysis

TMB Tumor mutation burden

CDF Cumulative distribution function

LASSO Least absolute shrinkage and selection operator

PI Prognosis Index

HR Hazard ratio

GO Gene Ontology

RFS Release-free survival

PFS Progression-free survival

Declarations

Acknowledgments

Not applicable.

Authors' contributions

Conception and design: BZ and BT. Acquisition of data: BZ and BT. Analysis and interpretation of data: BZ and BT. Drafting the article: BZ, BT and LQ. Critically revising the article: all authors. Reviewed submitted version of manuscript: all authors. Study supervision: LQ.

Funding

None.

Availability of data and materials

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

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests

Contributor Information

Ling Qin, Email: qinlingling@yahoo.com.

Baohui Zhang, Email: 18102489912@163.com

Bufu Tang, Email: tangbufu@hotmail.com

Jianyao Gao. Email: gaojygao@163.com

Jiatong Li, Email: jtli1995@163.com

Lingming Kong, Email: konglingmingklm@163.com

References

1. El-Serag HB: **Hepatocellular carcinoma.** *N Engl J Med* 2011, **365:**1118-1127.
2. Nault JC, Villanueva A: **Biomarkers for hepatobiliary cancers.** *Hepatology* 2020.
3. Kyrochristos ID, Ziogas DE, Roukos DH: **Dynamic genome and transcriptional network-based biomarkers and drugs: precision in breast cancer therapy.** *Medicinal research reviews* 2019, **39:**1205-

1227.

4. Yin F, Shu L, Liu X, Li T, Peng T, Nan Y, Li S, Zeng X, Qiu X: **Microarray-based identification of genes associated with cancer progression and prognosis in hepatocellular carcinoma.** *J Exp Clin Cancer Res* 2016, **35:**127.
5. Schafer DF, Sorrell MF: **Hepatocellular carcinoma.** *Lancet* 1999, **353:**1253-1257.
6. Palazón A, Aragonés J, Morales-Kastresana A, de Landázuri MO, Melero I: **Molecular pathways: hypoxia response in immune cells fighting or promoting cancer.** *Clinical cancer research : an official journal of the American Association for Cancer Research* 2012, **18:**1207-1213.
7. Gray LH, Conger AD, Ebert M, Hornsey S, Scott OC: **The concentration of oxygen dissolved in tissues at the time of irradiation as a factor in radiotherapy.** *Br J Radiol* 1953, **26:**638-648.
8. Nishida N, Kudo M: **Oxidative stress and epigenetic instability in human hepatocarcinogenesis.** *Digestive diseases (Basel, Switzerland)* 2013, **31:**447-453.
9. Wu X-Z, Xie G-R, Chen D: **Hypoxia and hepatocellular carcinoma: The therapeutic target for hepatocellular carcinoma.** *Journal of gastroenterology and hepatology* 2007, **22:**1178-1182.
10. Erler JT, Giaccia AJ: **Lysyl oxidase mediates hypoxic control of metastasis.** *Cancer research* 2006, **66:**10238-10241.
11. Graham K, Unger E: **Overcoming tumor hypoxia as a barrier to radiotherapy, chemotherapy and immunotherapy in cancer treatment.** *International journal of nanomedicine* 2018, **13:**6049-6058.
12. Tlsty TD, Coussens LM: **Tumor stroma and regulation of cancer development.** *Annual review of pathology* 2006, **1:**119-150.
13. Mantovani A, Allavena P, Sica A, Balkwill F: **Cancer-related inflammation.** *Nature* 2008, **454:**436-444.
14. Aggarwal BB, Shishodia S, Sandur SK, Pandey MK, Sethi G: **Inflammation and cancer: how hot is the link?** *Biochemical pharmacology* 2006, **72:**1605-1621.
15. Palazon A, Goldrath AW, Nizet V, Johnson RS: **HIF transcription factors, inflammation, and immunity.** *Immunity* 2014, **41:**518-528.
16. Chen Y-P, Zhang Y, Lv J-W, Li Y-Q, Wang Y-Q, He Q-M, Yang X-J, Sun Y, Mao Y-P, Yun J-P, et al: **Genomic Analysis of Tumor Microenvironment Immune Types across 14 Solid Cancer Types: Immunotherapeutic Implications.** *Theranostics* 2017, **7:**3585-3594.
17. Schumacher TN, Schreiber RD: **Neoantigens in cancer immunotherapy.** *Science (New York, NY)* 2015, **348:**69-74.
18. Chabanon RM, Pedrero M, Lefebvre C, Marabelle A, Soria J-C, Postel-Vinay S: **Mutational Landscape and Sensitivity to Immune Checkpoint Blockers.** *Clinical cancer research : an official journal of the American Association for Cancer Research* 2016, **22:**4309-4321.
19. Wang X, Li M: **Correlate tumor mutation burden with immune signatures in human cancers.** *BMC immunology* 2019, **20:**4.
20. Chalmers ZR, Connelly CF, Fabrizio D, Gay L, Ali SM, Ennis R, Schrock A, Campbell B, Shlien A, Chmielecki J, et al: **Analysis of 100,000 human cancer genomes reveals the landscape of tumor**

mutational burden. *Genome medicine* 2017, **9:**34.

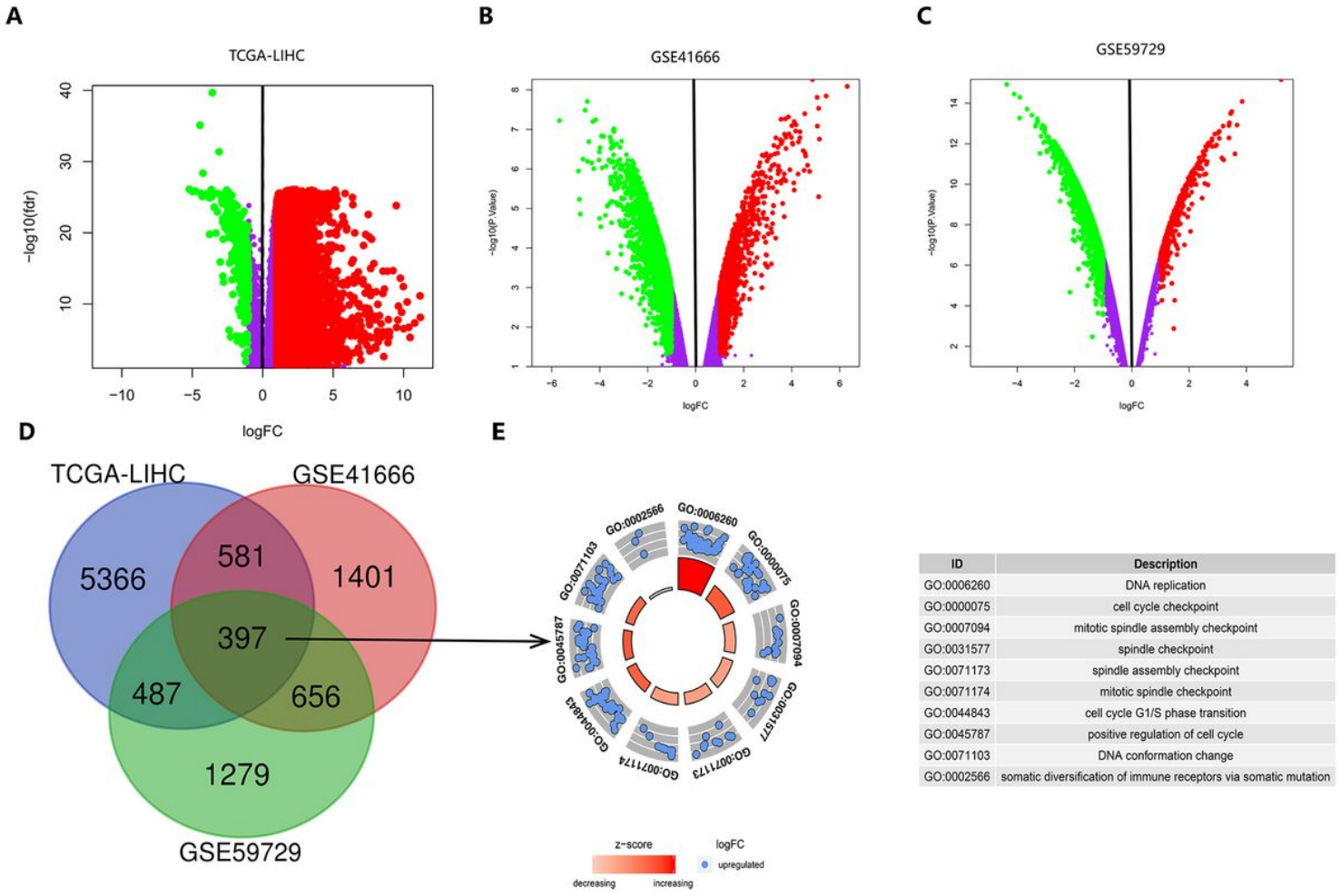
21. Diboun I, Wernisch L, Orengo CA, Koltzenburg M: **Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma.** *BMC Genomics* 2006, **7:**252.
22. Wilkerson MD, Hayes DN: **ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking.** *Bioinformatics* 2010, **26:**1572-1573.
23. Tibshirani R: **The lasso method for variable selection in the Cox model.** *Stat Med* 1997, **16:**385-395.
24. Iasonos A, Schrag D, Raj GV, Panageas KS: **How to build and interpret a nomogram for cancer prognosis.** *J Clin Oncol* 2008, **26:**1364-1370.
25. Vickers AJ, Cronin AM, Elkin EB, Gonen M: **Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers.** *BMC Med Inform Decis Mak* 2008, **8:**53.
26. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *Proc Natl Acad Sci U S A* 2005, **102:**15545-15550.
27. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA: **Robust enumeration of cell subsets from tissue expression profiles.** *Nat Methods* 2015, **12:**453-457.
28. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, et al: **The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data.** *Cancer Discov* 2012, **2:**401-404.
29. Hou GX, Liu P, Yang J, Wen S: **Mining expression and prognosis of topoisomerase isoforms in non-small-cell lung cancer by using Oncomine and Kaplan-Meier plotter.** *PLoS One* 2017, **12:**e0174515.
30. Galuppini F, Dal Pozzo CA, Deckert J, Loupakis F, Fassan M, Baffa R: **Tumor mutation burden: from comprehensive mutational screening to the clinic.** *Cancer Cell Int* 2019, **19:**209.
31. El-Serag HB, Rudolph KL: **Hepatocellular carcinoma: epidemiology and molecular carcinogenesis.** *Gastroenterology* 2007, **132:**2557-2576.
32. Chen X, Li B, He W, Wei Y-G, Du Z-G, Jiang L: **Mesohepatectomy versus extended hemihepatectomy for centrally located hepatocellular carcinoma.** *Hepatobiliary & pancreatic diseases international : HBPD INT* 2014, **13:**264-270.
33. Sangiovanni A, Colombo M: **Treatment of hepatocellular carcinoma: beyond international guidelines.** *Liver international : official journal of the International Association for the Study of the Liver* 2016, **36 Suppl 1:**124-129.
34. Marquardt JU, Galle PR, Teufel A: **Molecular diagnosis and therapy of hepatocellular carcinoma (HCC): an emerging field for advanced technologies.** *Journal of hepatology* 2012, **56:**267-275.
35. Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for transcriptomics.** *Nature reviews Genetics* 2009, **10:**57-63.
36. Xiao M, Liu L, Zhang S, Yang X, Wang Ya: **Cancer stem cell biomarkers for head and neck squamous cell carcinoma: A bioinformatic analysis.** *Oncology reports* 2018, **40:**3843-3851.

37. Wilson WR, Hay MP: **Targeting hypoxia in cancer therapy.** *Nat Rev Cancer* 2011, **11:**393-410.
38. Xiong XX, Qiu XY, Hu DX, Chen XQ: **Advances in Hypoxia-Mediated Mechanisms in Hepatocellular Carcinoma.** *Mol Pharmacol* 2017, **92:**246-255.
39. Yan H, Li Z, Shen Q, Wang Q, Tian J, Jiang Q, Gao L: **Aberrant expression of cell cycle and material metabolism related genes contributes to hepatocellular carcinoma occurrence.** *Pathol Res Pract* 2017, **213:**316-321.
40. Ci C, Tang B, Lyu D, Liu W, Qiang D, Ji X, Qiu X, Chen L, Ding W: **Overexpression of CDCA8 promotes the malignant progression of cutaneous melanoma and leads to poor prognosis.** *Int J Mol Med* 2019, **43:**404-412.
41. Koppula P, Zhang Y, Zhuang L, Gan B: **Amino acid transporter SLC7A11/xCT at the crossroads of regulating redox homeostasis and nutrient dependency of cancer.** *Cancer Commun (Lond)* 2018, **38:**12.
42. Mollet J, Giurgea I, Schlemmer D, Dallner G, Chretien D, Delahodde A, Bacq D, de Lonlay P, Munnich A, Rotig A: **Prenyldiphosphate synthase, subunit 1 (PDSS1) and OH-benzoate polyprenyltransferase (COQ2) mutations in ubiquinone deficiency and oxidative phosphorylation disorders.** *J Clin Invest* 2007, **117:**765-772.
43. Steuer CE, Ramalingam SS: **Tumor Mutation Burden: Leading Immunotherapy to the Era of Precision Medicine?** *J Clin Oncol* 2018, **36:**631-632.
44. Mamlouk S, Wielockx B: **Hypoxia-inducible factors as key regulators of tumor inflammation.** *Int J Cancer* 2013, **132:**2721-2729.
45. Nishida N, Kudo M: **Oncogenic Signal and Tumor Microenvironment in Hepatocellular Carcinoma.** *Oncology* 2017, **93 Suppl 1:**160-164.
46. Pardoll DM: **The blockade of immune checkpoints in cancer immunotherapy.** *Nat Rev Cancer* 2012, **12:**252-264.

Supplemental Information Note

Figure S1. Identification of key hypoxia-related genes closely related to the prognosis of HCC. A-B LASSO-penalized Cox regression. The dataset was subsampled 1000 times and chose the genes repeated >900 times.

Figures



ID	Description
GO:0006260	DNA replication
GO:0000075	cell cycle checkpoint
GO:0007094	mitotic spindle assembly checkpoint
GO:0031577	spindle checkpoint
GO:0071173	spindle assembly checkpoint
GO:0071174	mitotic spindle checkpoint
GO:0044843	cell cycle G1/S phase transition
GO:0045787	positive regulation of cell cycle
GO:0071103	DNA conformation change
GO:0002566	somatic diversification of immune receptors via somatic mutation

Figure 1

Volcano plot showing the differentially expressed hypoxia-related genes of HCC in different databases. A Gene expression levels in the TCGA database. B Gene expression profiles of Huh-7 cells under normoxia and hypoxia for 24 hours in the GSE41666 cohort. C Gene expression profiles of HepG2 cells under normoxia and hypoxia for 24 hours in the GSE59729 cohort. D Common differentially expressed genes between the TCGA and GEO databases. E Gene Ontology (GO) analysis of 397 hypoxia-related DEGs in HCC.

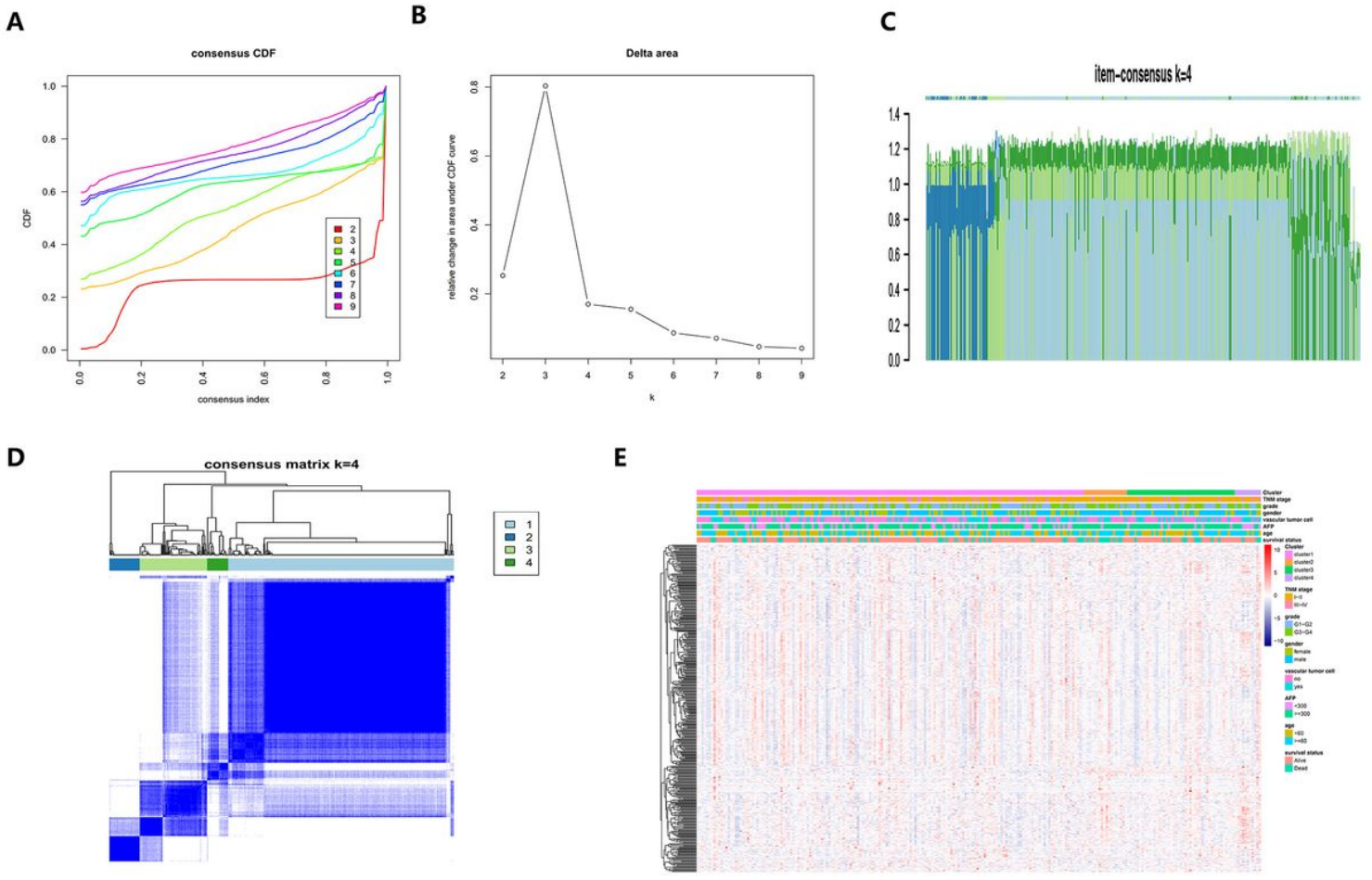


Figure 2

Consensus clustering of HCC molecular subgroups based on hypoxia-related DEGs. A Cumulative distribution function (CDF) curve. B CDF Delta area curve, which indicates the relative change in the area under the CDF curve for each category number k compared with k-1. The horizontal axis represents the number k and the vertical axis represents the relative change in the area under the CDF curve. C The Item-Consensus Plot for k=4. The vertical axis represents item-consensus values and each bar represents each sample. D The heatmap corresponding to the consensus matrix for k=4 obtained by applying consensus clustering. The rows and columns of the matrix represent samples. The values of the consistency matrix are shown in white to dark blue from 0 to 1, which represent the degree of consensus. E The heatmap of 397 hypoxia-related genes in 4 clusters.

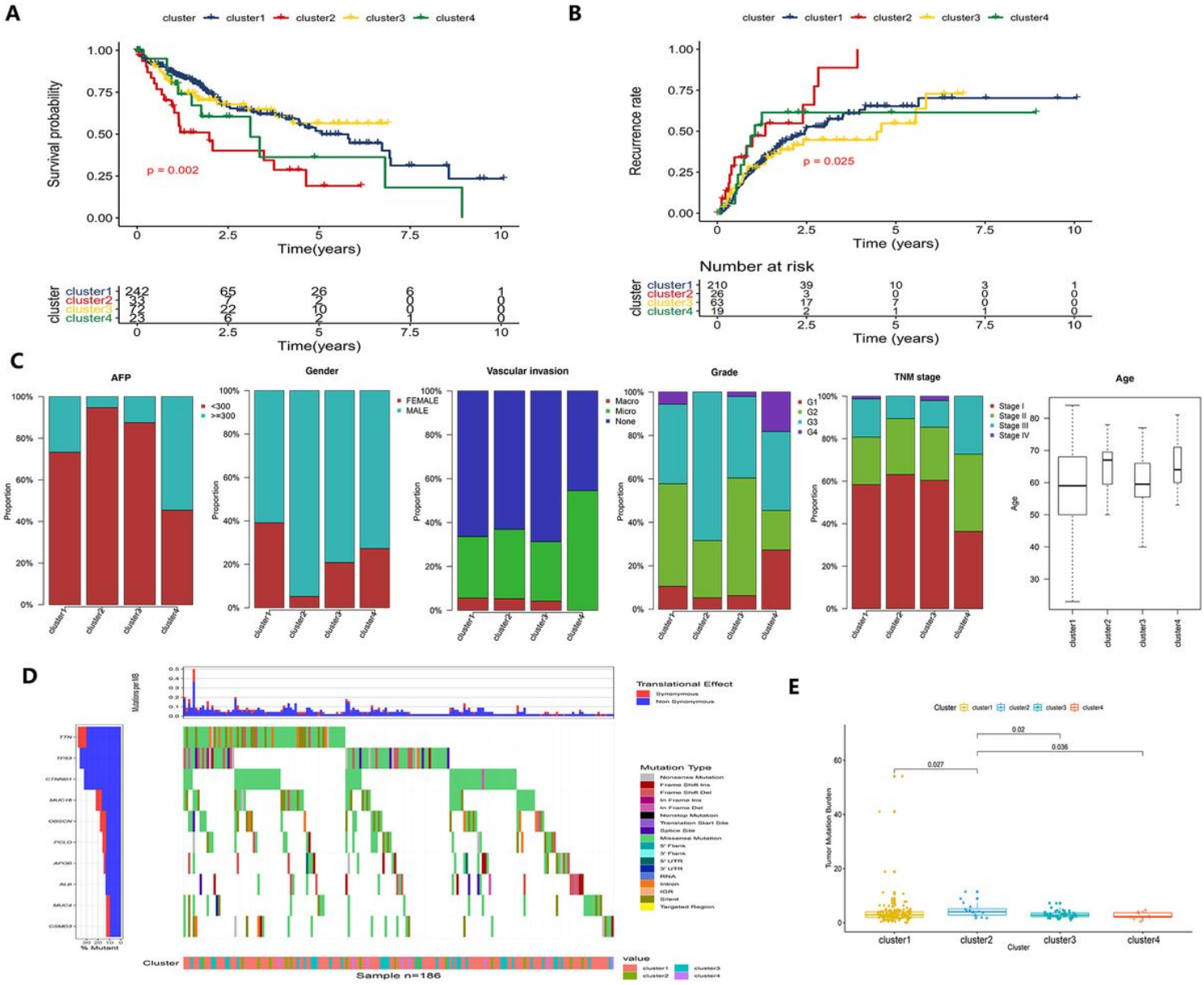


Figure 3

Characterization of different features of hypoxia-related DEGs clustering. A-B K-M survival curves showed the differences of overall survival and recurrence rate among the 4 clusters. C Proportion of other clinical characteristics in 4 clusters. D-E The differences of TMB among 4 clusters.

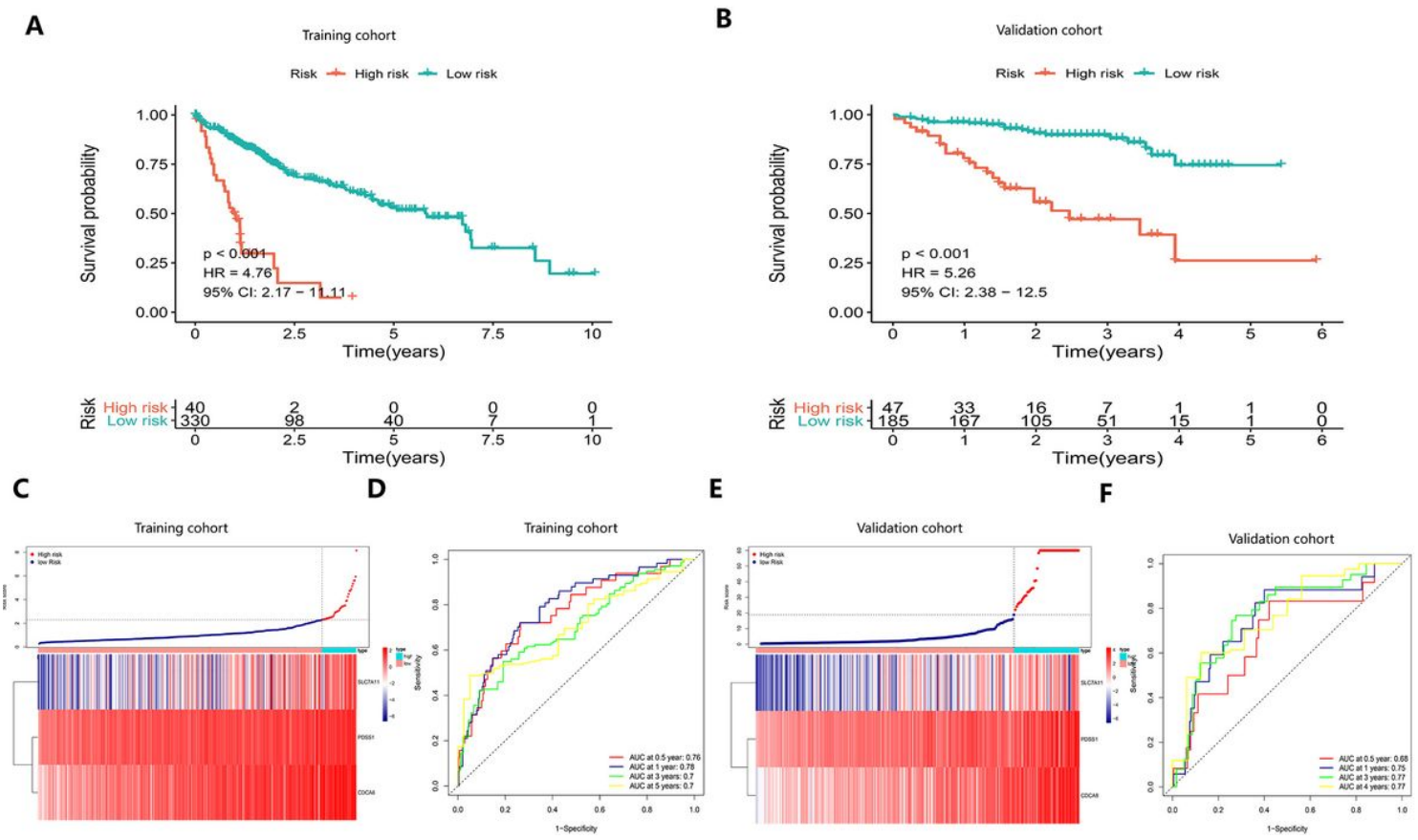


Figure 4

Kaplan–Meier analysis, risk score analysis, time-dependent ROC analysis for a prognosis model based the three-gene signature in HCC. A-B K-M survival curve of high- and low-risk in TCGA cohort and ICGC cohort. C and E Distribution of risk scores of HCC patients with different gene expression levels. D and F Time-dependent ROC analysis for OS prediction in TCGA and ICGC cohort.

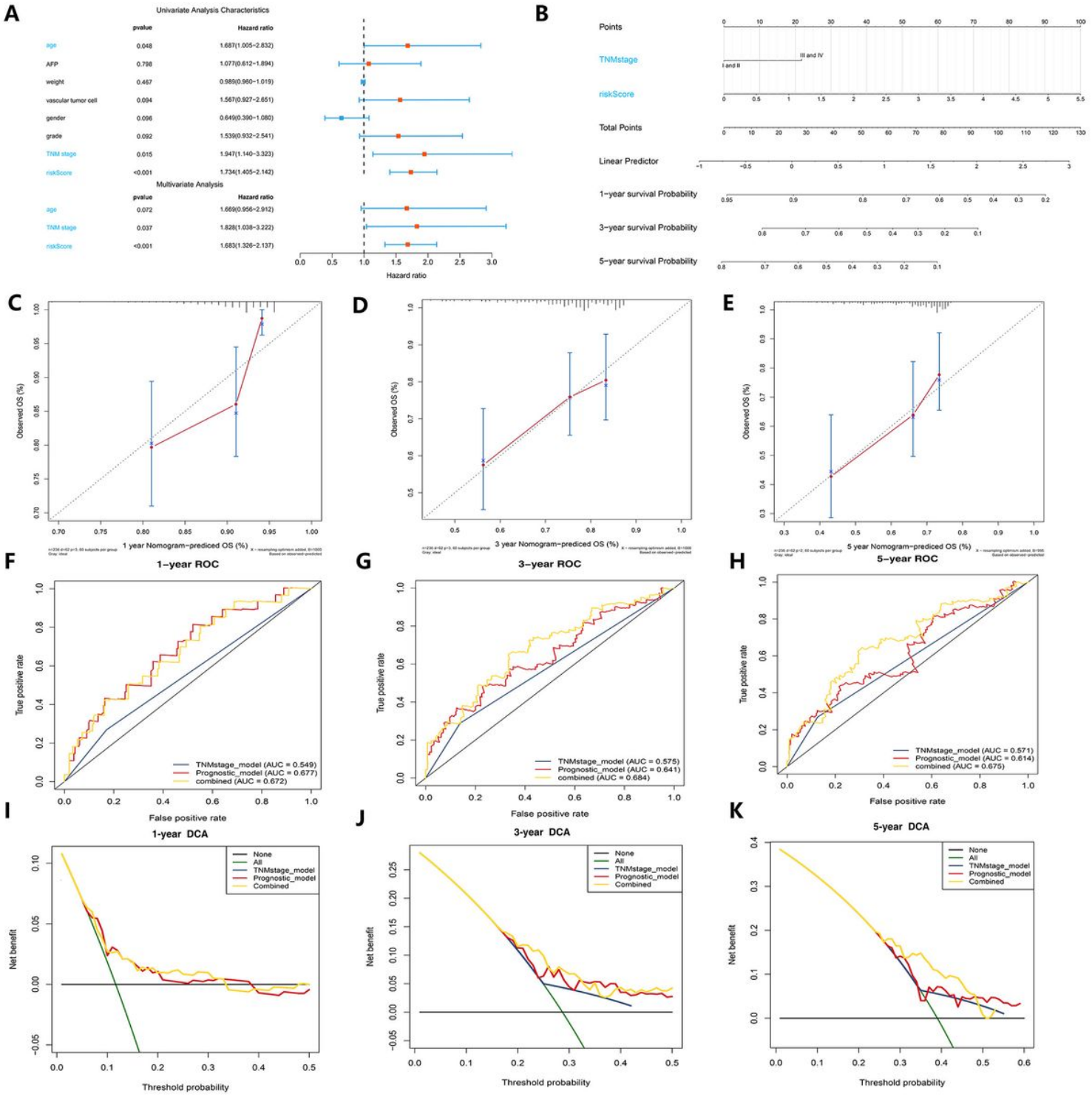


Figure 5

Construction of the nomogram predicting overall survival for HCC patients in the TCGA cohort. A Forrest plot of the univariate and multivariate association of the prognostic model and clinicopathological characteristics with overall survival. B The nomogram was built based on two independent prognostic factors for predicting OS in HCC patients at 1-, 3-, and 5- year. C-E The calibration plot for internal validation of the nomogram. F-H Time-dependent ROC curves of the nomogram for 1-,3- and 5-year overall survival in HCC to evaluate the predictive performance of the nomogram. I-K DCA curves of the

nomogram for 1-,3- and 5-year overall survival in HCC to evaluate the clinical decision-making benefits of the nomogram.

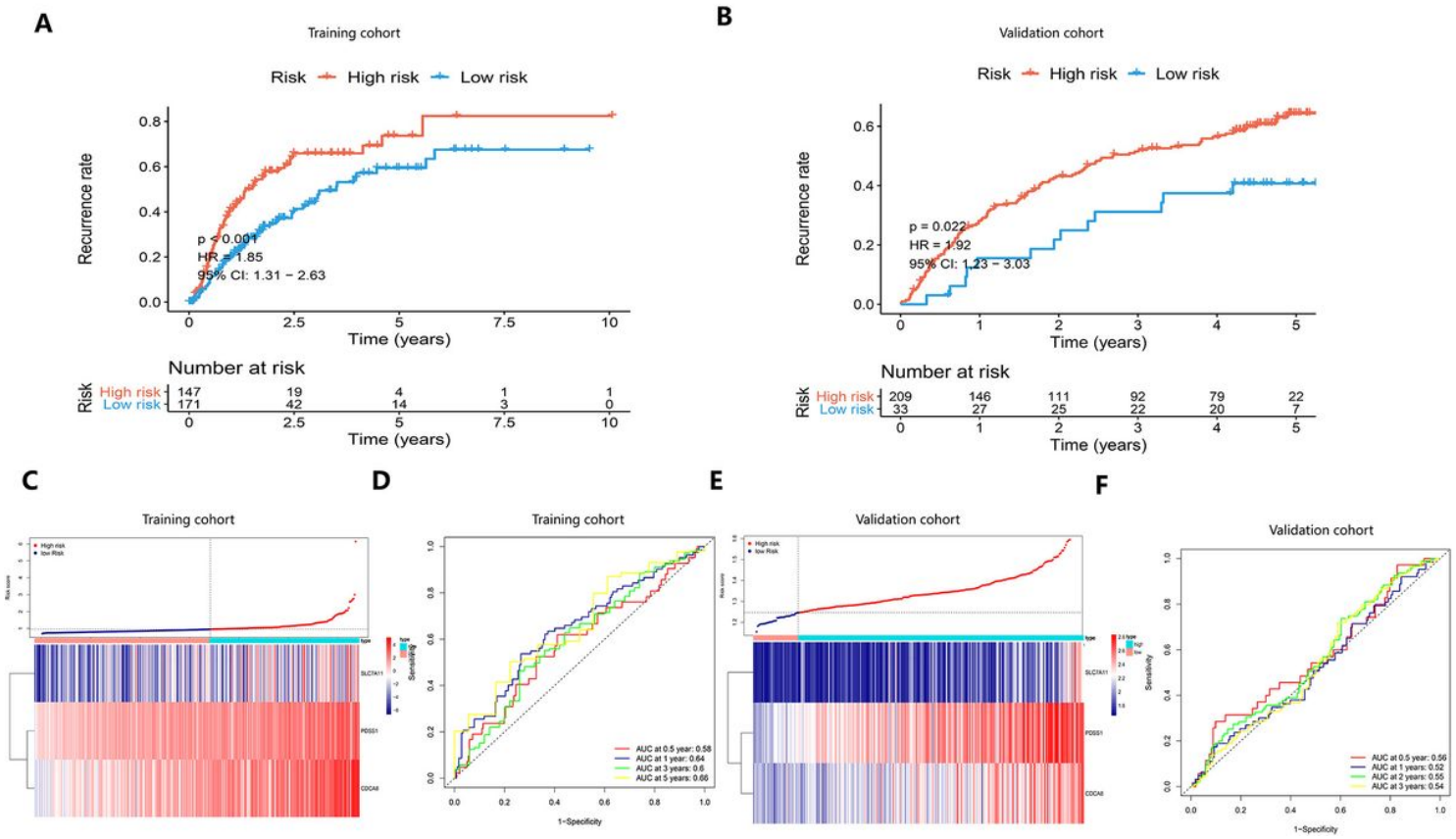


Figure 6

Kaplan–Meier analysis, risk score analysis, time-dependent ROC analysis for the recurrence model based on three-gene signature in HCC. A-B The recurrence rates of high- and low-risk group in TCGA and ICGC cohort. C and E Distribution of risk scores of HCC patients with different gene expression levels. D and F Time-dependent ROC analysis for recurrence prediction in TCGA and ICGC cohort.

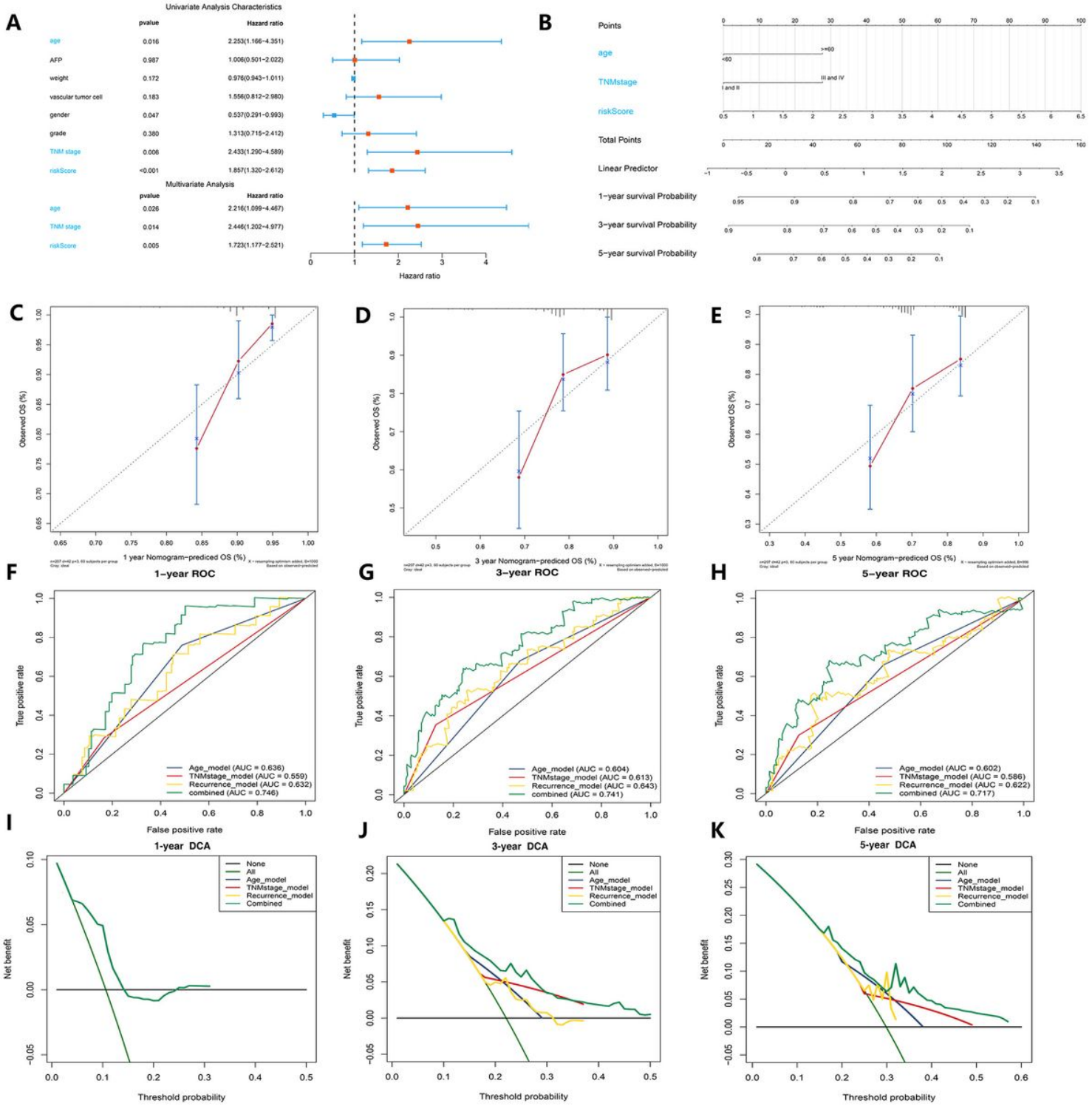
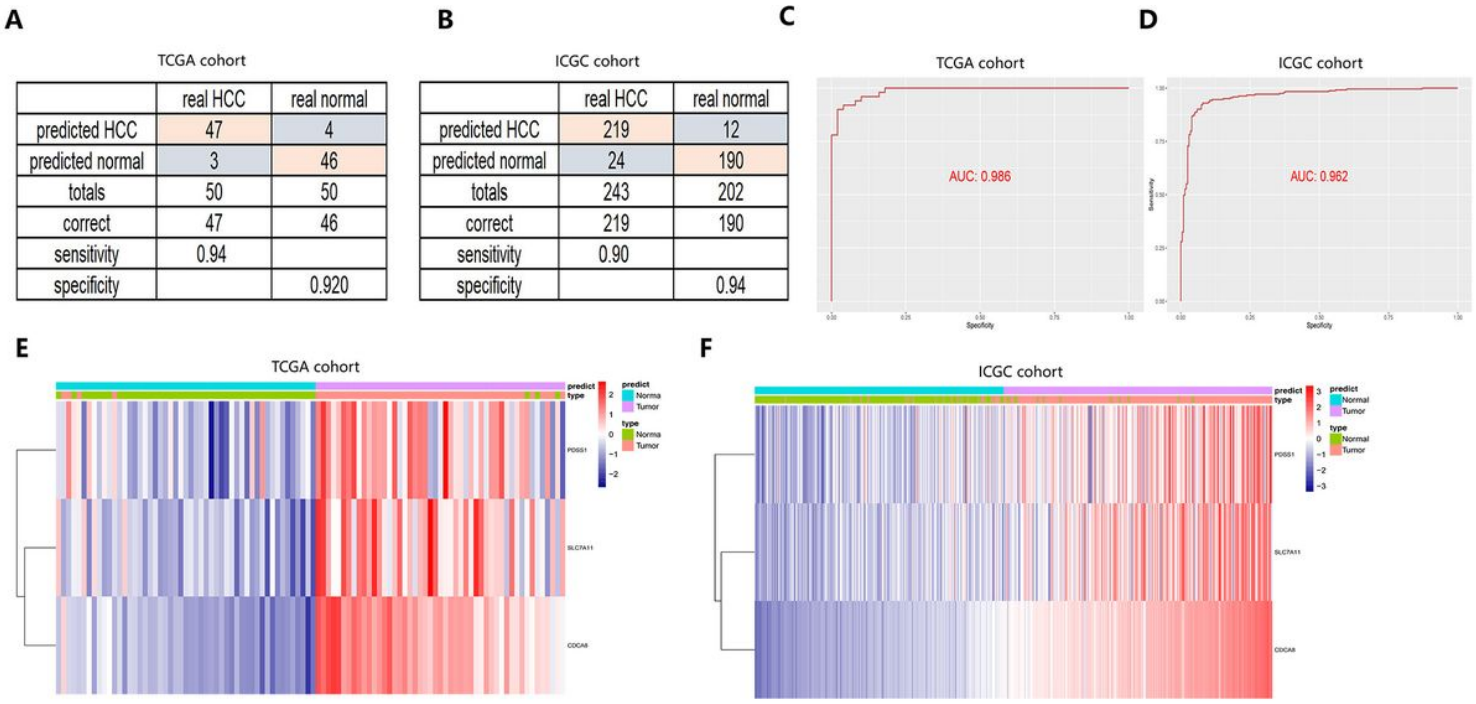


Figure 7

Construction of a recurrence nomogram for HCC patients in the TCGA cohort. A Forrest plot of the univariate and multivariate association of the risk-score model and clinicopathological characteristics with overall survival. B The nomogram was built based on three independent prognostic factors for predicting the recurrence in HCC patients at 1-, 3-, and 5- year. C-E The calibration plot for internal validation of the nomogram. F-H Time-dependent ROC curves of the nomogram for 1-,3- and 5-year recurrence prediction in HCC to evaluate the predictive performance of the nomogram. I-K DCA curves of

the nomogram for 1-,3- and 5-year recurrence prediction in HCC to evaluate the clinical decision-making benefits of the nomogram.

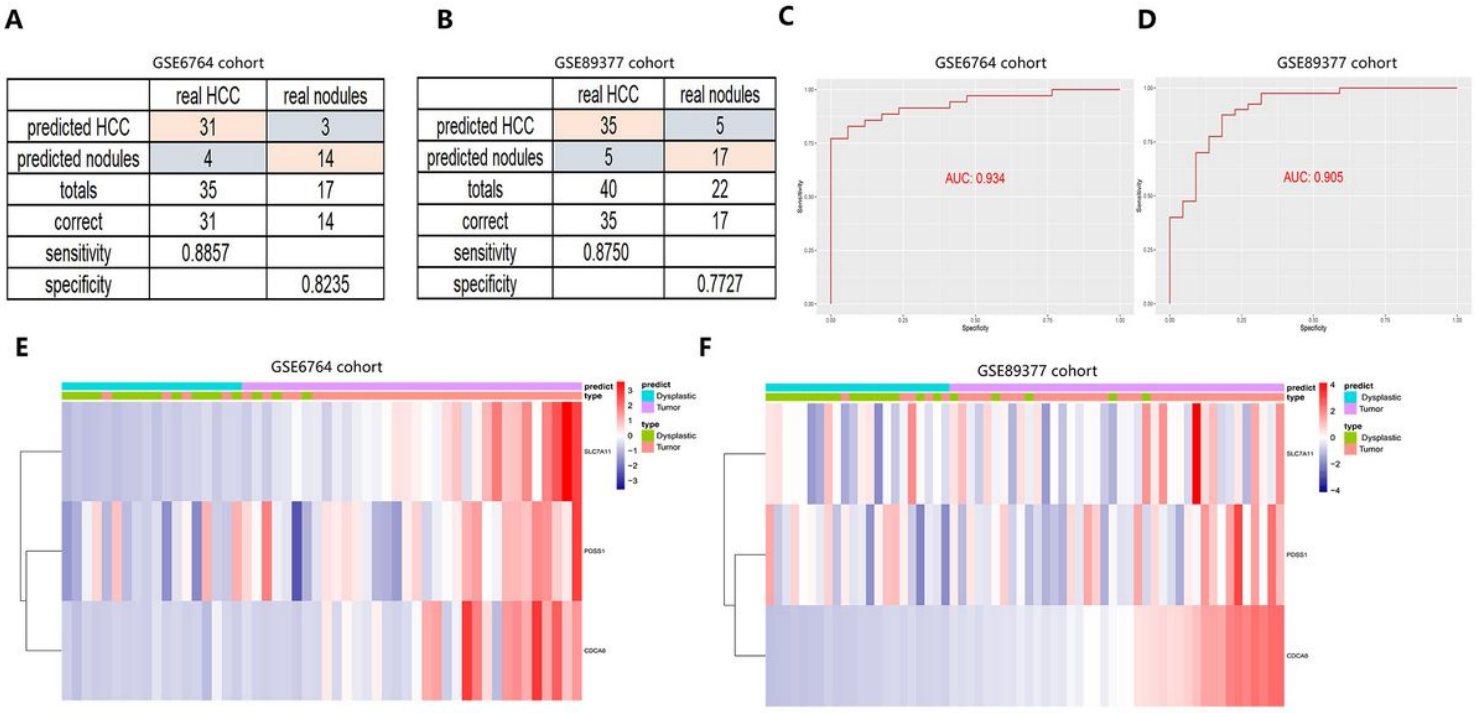


	real HCC	real normal
predicted HCC	47	4
predicted normal	3	46
totals	50	50
correct	47	46
sensitivity	0.94	
specificity		0.920

	real HCC	real normal
predicted HCC	219	12
predicted normal	24	190
totals	243	202
correct	219	190
sensitivity	0.90	
specificity		0.94

Figure 8

Building a diagnostic model for distinguishing HCC from normal samples. A-B sensitivity and specificity validation of the diagnostic model in the TCGA and ICGC cohort. C-D ROC curves for evaluating the predictive performance of the diagnostic model. E-F The association between the predictive type of diagnostic model and the distribution of gene expression.



	real HCC	real nodules
predicted HCC	31	3
predicted nodules	4	14
totals	35	17
correct	31	14
sensitivity	0.8857	
specificity		0.8235

	real HCC	real nodules
predicted HCC	35	5
predicted nodules	5	17
totals	40	22
correct	35	17
sensitivity	0.8750	
specificity		0.7727

Figure 9

Building a diagnostic model for distinguishing HCC from dysplastic nodules. A-B sensitivity and specificity validation of the diagnostic model in the GSE6764 and GSE89377 cohort. C-D ROC curves for evaluating the predictive performance of the diagnostic model. E-F The association between the predictive type of diagnostic model and the distribution of gene expression.

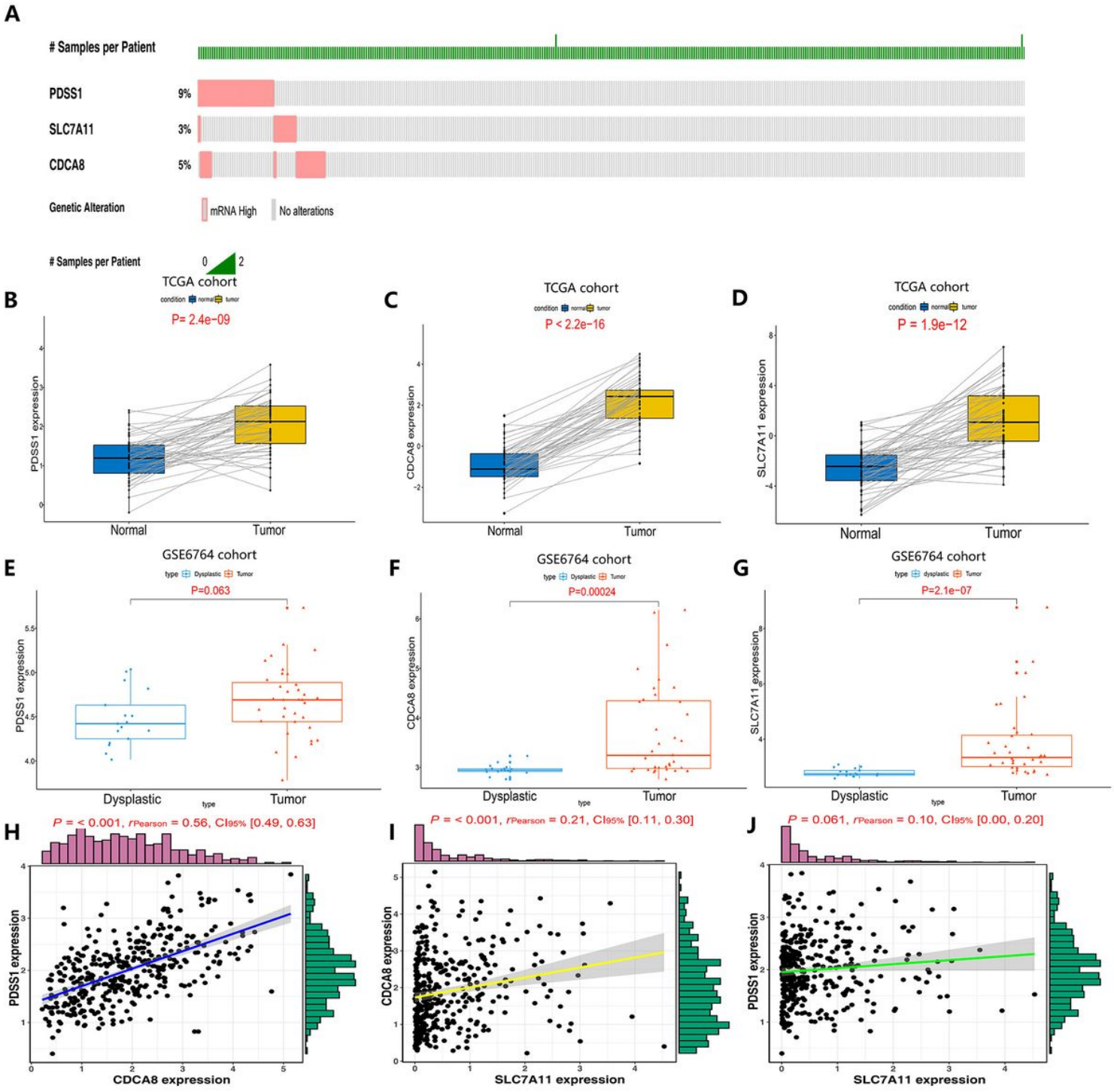


Figure 10

Validation of the expression characteristics of hypoxia-related genes. A Genetic alteration detection of the hypoxia-related genes from the cBioPortal database. B-D The expression level of each gene in TCGA

cohort with 50 HCC samples paired 50 normal samples. E-G The expression level of each gene in GSE6764 cohort with 35 HCC samples paired 17 dysplastic nodules. H-J The correlation analysis between expression levels of different genes in TCGA cohort with 370 HCC samples.

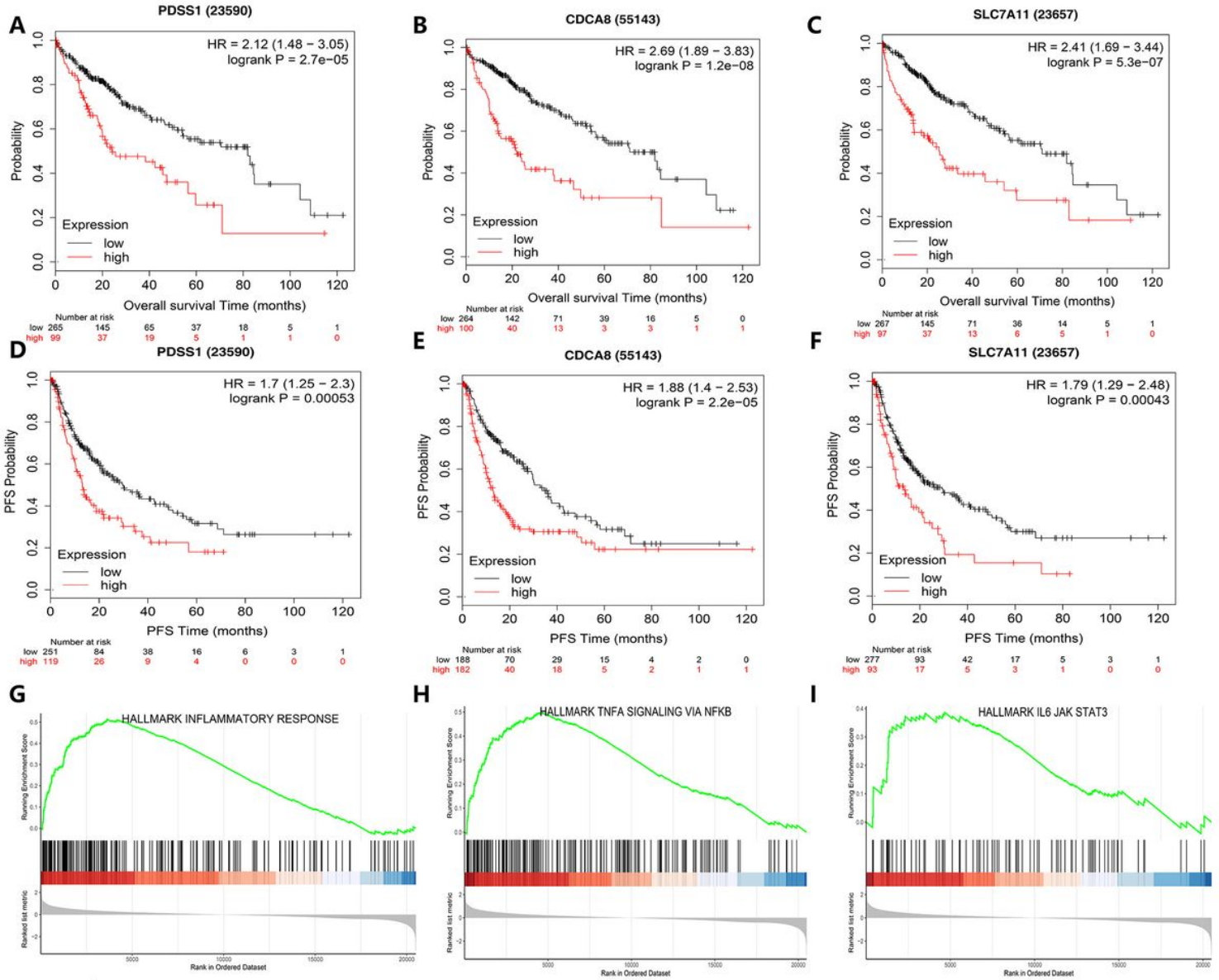


Figure 11

Prediction performance of hypoxia-related genes for OS and Gene Set Enrichment Analyses of the three-gene signature. A-C K-M survival curves for high and low expression levels of each gene. D-F progression-free survival analysis for high and low expression levels of each gene. G-I Three representative Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways in high-risk group via GSEA.

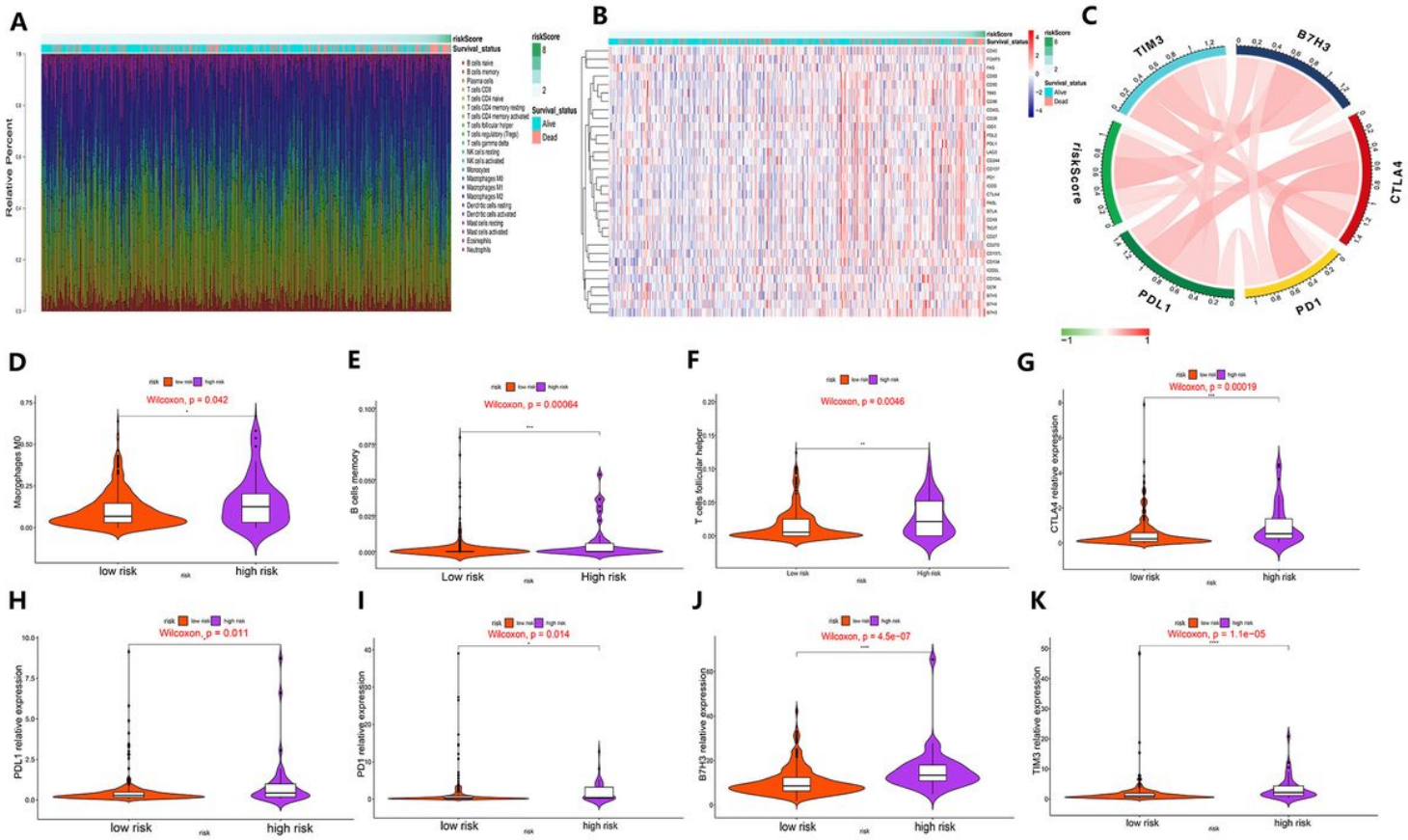


Figure 12

The overview of immune infiltration and expression of immune checkpoints in HCC patients with different risk scores. A The distribution of 22 immune cell types in HCC patients with different risk score. B The expression levels of immune checkpoints in HCC patients with different risk score. C The top 6 immune checkpoints with high expression level in high-risk group. D-F Violin plots showing infiltration fractions of different immune cells in the high- and low-risk groups. G-K Violin plots showing the expression level of immunocheckpoints in high- and low-risk groups.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1.tif