Original Research *Highlight article*

Intra-tumor heterogeneity and prognostic risk signature for hepatocellular carcinoma based on single-cell analysis

Chengli Liu[1], Meng Pu[2], Yingbo Ma[3], Cheng Wang[2], Linghong Kong[2], Shuhan Zhang[2], Xuying Zhao[4] and Xiaopeng Lian[4]

[1]Department of Hepatobiliary Surgery, Air Force Medical Center, Air Force Clinical College (Air Force Medical Center) of Anhui Medical University, Beijing 100142, China; [2]Department of Hepatobiliary Surgery, Air Force Medical Center, Beijing 100142, China; [3]Department of Hepatobiliary Surgery, Standardized Residency Training Base, Air Force Medical Center, Beijing 100142, China; [4]Department of Hepatobiliary Surgery, Standardized General Surgery Specialists Training Base, Air Force Medical Center, Beijing 100142, China
Corresponding author: Chengli Liu. Email: liuchengli_beijing@163.com

Impact Statement

Intratumoral heterogeneity of hepatocellular carcinoma (HCC) is a major challenge in clinical treatment. In recent years, scRNA-seq has emerged as a significant technique for investigating tumor heterogeneity. In this work, at the single-cell level, we described the immune microenvironment of HCC and identified 25 cell clusters representing 13 HCC cell types. Combining single-cell sequencing analysis with RNA sequencing analysis, a prognostic risk model developed with markers of dendritic cells, hepatocyte, liver bud hepatic cell, and liver progenitor cells enriched in HCC was constructed and could independently perform the survival prediction of HCC patients. This study extended the application of the combination of traditional RNA sequencing and single-cell sequencing to cancer research, providing novel insights into HCC prognosis.

Abstract

Intra-tumor heterogeneity poses a serious challenge in the treatment of cancer, including hepatocellular carcinoma (HCC). Recent developments in single-cell RNA sequencing (scRNA-seq) make it possible to examine the heterogeneity of tumor cells. The Gene Expression Omnibus (GEO) database was retrieved to obtain scRNA-seq data of 13 HCC and 8 para cancer samples, and the cells were clustered using FindNeighbors and FindClusters functions. Cell subsets were defined using the "Enricher" function of the clusterProfiler package. Monocle was used to predict cell developmental trajectory. The LIMMA package included in the R program was utilized to detect differentially expressed genes (DEGs) between HCC and paracancerous tissues. Univariate Cox analysis and Least Absolute and Selection Operator (Lasso) Cox regression analysis were conducted to establish a risk assessment model. Thirteen cell subpopulations were identified from the sequencing data of 64,634 single cells. Four cell subgroups (dendritic cells, hepatocytes, liver bud hepatic cells, and liver progenitor cells) in tumor tissues were highly enriched. Between HCC and para cancer tissues, 3024 DEGs were identified, and 641 were specific markers of four cell subgroups. To develop a prognostic risk model, 9 genes among the 641 genes were selected. In the training and validation sets, the model demonstrated a higher 5-year AUC and independently served as a prognostic marker as confirmed by multivariate and univariate Cox analyses. This study revealed the characteristics of different cell subpopulations of immune cells and tumor cells from the HCC microenvironment. We established a novel nine-gene prognostic model to determine the death risk of HCC patients. The discoveries in this research improved the current knowledge of HCC heterogeneity and may inspire future research.

Keywords: Hepatocellular carcinoma, single-cell RNA sequencing, intra-tumor heterogeneity, prognosis

***Experimental Biology and Medicine* 2022; 247: 1741–1751. DOI: 10.1177/15353702221110659**

Introduction

HCC has been identified as the third major contributor to deaths associated with cancers globally with the majority of the cases being reported from Asia.[1] HCC patients diagnosed early can be treated with curative therapy, including liver transplantation, percutaneous ablation, and surgical resection; however, around 70–80% of HCC patients will develop recurrence after receiving curative therapy.[2] Previous evidence suggests that HCC is immunogenic and comprises infiltrated tumor-specific T cells as well as other immune cells. Through inducing tumor-specific immune responses in cancer cells, immunotherapy offers effective and differentiated tumor cell targeting and improves the postoperative relapse-free survival of HCC patients.[3] Despite these significant advances, owing to significant heterogeneous genomic aberrations and complex immune microenvironment of tumors in HCC, translational immunotherapy for clinical personalized care remains a challenge in precision oncology.[4] Analysis of immune subtypes and the immune

ISSN 1535-3702

microenvironment in HCC may have potential clinical value for personalized immunotherapy.

Transcriptome data analysis could influence decision-making in clinical practice and the development of precise regimes of treatment. However, the general transcriptomic analysis does not apply to the study of cell heterogeneity (i.e. cell subpopulations within major types of cells), specific pathogenic cell populations, rare cell populations, and/or dissecting the cancer microenvironment and clonal evolution.[5] Transcriptomic analysis of a single cell, that is, scRNA-seq, has improved previous understanding of the heterogeneity of cell populations and cellular state diversity, thereby contributing to the profiling of properties of several cell types in tumor and its vicinity.[6] scRNA-seq has been applied to reveal immune cell population in different malignancies and classify several heterogeneous tumors, such as gastric adenocarcinoma,[7] colorectal cancer,[8] breast cancer,[9] and renal clear cell carcinoma[10]. In addition, scRNA-seq could identify key molecular markers for predicting cancer prognosis. Zhang *et al.*[11] constructed a robust wrinkle-associated genes signature with scRNA-seq for stratifying clear cell renal cell carcinoma patients' prognosis. Similarly, scRNA-seq analysis of melanoma samples has also been used for developing a metastasis-associated genes signature to predict patients' prognoses.[12] Although a preliminary classification of HCC using scRNA-seq has been developed,[13] HCC prognostic signature involving more effective genes is still needed.

Previous studies have developed various prognostic signatures from different aspects of HCC prognosis prediction. For instance, Li and his colleagues explored a gene signature incorporating RTN3, SOCS2, and UPB1 for the prognosis of patients with HCC premised on T stage stratification in TCGA-liver hepatocellular carcinoma (LIHC) dataset.[14] Liu *et al.*[15] established a four-gene signature related to metabolism for HCC. Li *et al.*[16] identified seven prognostic genes associated with DNA repair that could significantly classify HCC patients into high- and low-risk subgroups. Long non-coding RNA-associated prognostic signatures were mined using different methodologies.[17–19] However, limited studies used the scRNA-seq approach to evaluate the heterogeneity within HCC tumors and exploit important ligand–receptor interactions. Therefore, this research used the scRNA-seq data of 13 cancer and 8 para cancer tissue samples in GSE149614 to analyze the intra-tumor heterogeneity and cell interactions in HCC. The potential risk models for predicting HCC prognosis were identified through bioinformatics analysis of the transcriptome of HCC patients in the Cancer Genome Atlas (TCGA). Our findings may expand the understanding of the heterogeneity of HCC, providing a potential novel tool for the prognostic prediction of HCC patients.

Materials and methods

Data acquisition and preprocessing

See Figure S1 for the workflow of this study. We retrieved the scRNA-seq data of 21 samples (10 primary tumors; 1 metastatic lymph node; 2 portal vein tumor thrombus; and 8 para cancer tissues) in the GEO database in NCBI (https://www.ncbi.nlm.nih.gov/geo/). The FPKM data of the samples (containing 371 HCC and 50 para cancer tissues) from TCGA were converted to TPM data format. Complete survival follow-up records of 366 samples were included in the analysis. HCCDB18 cohort with HCC samples was acquired from the hepatocellular carcinoma database (HCCDB, http://lifeome.net/database/hccdb/home.html),[20] of which normal samples and those with missing data on survival status or time were excluded.

scRNA-seq data dimension reduction and unsupervised clustering

Transcriptomic sequencing data of cell samples from HCC patients from the GSE149614 cohort were obtained and converted into Seurat objects using the "Seurat" package[21] in R software. The PercentageFeatureSet function was used to calculate the proportion of mitochondria and rRNA. Quality filtering was performed to screen cells having >500 expressed genes or $>30\%$ mitochondrial counts. After quality filtering, to identify the genes with highly variable expressions, the FindVariableFeatures function in Seurat was employed. Normalization of data was achieved through log-normalization, and the ScaleData function was used to scale genes. Next, principal component analysis (PCA) of the linear dimension reduced data was performed, with dim = 50 and resolution = 0.1 as the set parameters. The cells were clustered using FindClusters and FindNeighbors algorithms. 2D uniform manifold approximation and projection (UMAP) visualized the results.

Defining cell subsets

From the official website of CellMarker,[22] cell marker genes in humans (http://biocc.hrbmu.edu.cn/CellMarker/) were downloaded, and the corresponding liver tissue specimen data were obtained. Simultaneously, cell subsets were defined using the Enricher function in the clusterProfiler package.[23]

Pseudotime trajectory analysis of cell subpopulations

To analyze the alterations in immune cell type distributions during tumor development, Monocle[24] was used to investigate the developmental trajectory of each cell subpopulation. Branching of cell trajectories often occurs as a result of differential gene expression patterns in cells. During development, as cells make their fate choices, the evolving trajectory would branch off, resulting in one developmental lineage following one path and another, a second path. The trajectories were visualized in a 2D T-distributed Neighbor Embedding (tSNE) graph.

Pathway enrichment analysis

Marker genes in each subgroup were identified, extracted, and input into the WebGestaltR package for Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis.[25] We adopted a false discovery rate (FDR) <0.05 to determine the significant key pathways, and the enriched pathways were presented using a bubble map, with a larger enrichment ratio positively correlating to a stronger correlation between genes and pathways.

Intercellular communication networks

CellChat[26] is a program that allows for the drawing of quantitative inferences and assessment of intercellular communication networks premised on the scRNA-seq data. CellChat identifies and analyzes complex interactions between subpopulations of cells. We followed the standard process of loading standardized data onto the CellChat platform. After the CellChat object was created, the assumed ligand–receptor interaction pair was identified through cellChatdb. human using the default parameters and visualized by the "circlize" tool.

Establishment and validation of the prognostic risk model for HCC

Differential analysis of the expression spectrum matrix of HCC and para-carcinoma tissues was performed using the limma package. Screening parameters were set as FDR < 0.05 and | log2 fold change (FC) | >1. Differentially expressed genes (DEGs) that overlapped with marker genes of subpopulations were identified. For the commonly shared genes, univariate Cox proportional risk regression analysis was executed with the aid of the survival coxph function in the survival R package (https://mran.microsoft.com/web/packages/survival/index.html), and $P < 0.05$ was employed to identify the genes considerably linked to HCC survival. Next, the Least Absolute Shrinkage and Selection Operator (Lasso) Cox regression analysis was conducted with the help of the R package glmnet,[27] and the weighted coefficients were calculated for developing a cancer risk assessment model. The calculation was as follows: Risk score = Σ ($\beta_i \times \mathrm{Exp}_i$), here, β_i signifies the weighted coefficient of genes whereas Exp_i signifies the gene expression levels. The efficiency of the risk assessment model for predicting prognosis was assessed using receiver operating characteristic (ROC) curves created through the use of the R software package timeROC.[28] Using the risk prediction model, each sample was given a risk score, which was subsequently converted into a Z score. HCC samples were finally classified into two groups of opposite risks (0 was the truncation value). Log-rank tests and Kaplan–Meier curves were utilized to analyze the results of the survival comparisons.

Statistical analysis

R Studio packages (version 3.6.3) were utilized to execute all analyses of statistical data. Evaluation of the prognostic significance of risk prediction models included both univariate and multivariate COX regression analyses. The threshold for statistical significance was established at $P < 0.05$.

Results

Cell classification in HCC and paired paracancerous tissues based on scRNA-seq data

The quality control diagram showed the percentages of mitochondrial genes, the unique molecular identifiers (UMIs) number, and rRNA counts before and after quality control. The results indicated a high-quality control of analysis samples (Figure S2). A total of 64,634 single cells were subjected to the quality filtration process. Cell numbers were assigned to each sample before- and after-mass filtration, as shown in Figure S3. Figure S4 showed the top 20 genes with highly variable expressions among a total of 2000. PCA combined with ElbowPlot demonstrated that most real signals were captured in the first 30 PCs (Figure S5). A total of 25 cell clusters were obtained by subsequent cell clustering. The use of UMAP allowed for the visualization of the cell distribution in 21 different tissues, including 26,771 single cells in normal and 37,863 single cells in tumor tissues (Figure 1(A)). In both the HCC and the para cancer tissues, we discovered the presence of several enriched cell clusters. On counting the abundances of 25 cell clusters in each sample, we found the highest abundance in normal tissues in subgroup 0, while cell clusters 1, 2, and 5 were abundant in tumor tissues (Figure 1(B)). In addition, we employed the FindAllMarkers function to filter each cell cluster consisting of gene markers (|log2(FC)| = 0.5, Minpct = 0.1). The heat map presented the top 5 marker gene expressions in each cell cluster (Figure S6A). According to known markers in the CellMarker database, 25 cell clusters were annotated to a total of 13 cell types (Table 1, Figure S6B), namely liver bud hepatic cells, B cells, cytotoxic CD4+ T cells, endothelial cells, dendritic cells, exhausted CD8+ T cells, hepatocytes, Kupffer cells, liver progenitor cells, memory B cells, mucosal-associated invariant T cells, myofibroblasts, and regulatory T cells (Tregs).

Table 1. Cell type of each subgroup.

cell_type	seraut_cluster
CD4+ cytotoxic T cell	0
Kupffer cell	1
Liver progenitor cell	2
Endothelial cell	3
Mucosal-associated invariant T cell	4
Liver bud hepatic cell	5
Regulatory T (Treg) cell	6
Liver bud hepatic cell	7
Myofibroblast	8
Kupffer cell	9
Liver bud hepatic cell	10
Hepatocyte	11
Kupffer cell	12
Memory B cell	13
Liver bud hepatic cell	14
Liver bud hepatic cell	15
Liver bud hepatic cell	16
B cell	17
Dendritic cell	18
Liver bud hepatic cell	19
Regulatory T (Treg) cell	20
Liver progenitor cell	21
Dendritic cell	22
Liver bud hepatic cell	23
Exhausted CD8+ T cell	24

Identification of highly enriched cell subsets and characterization of cell differentiation trajectories in tumor tissues

On analyzing the differences in the proportion of 13 types of cells between tumor and para cancer samples, significant heterogeneity in the tumor microenvironment between these

Table 2. Differences in the number of 13 cell types between tumor and paracancer samples (Fisher test).

Cell_name	T_celltype	N_cell type	P value	FC
B cell	459	760	1.26E-49	0.419989172
CD4+ cytotoxic T cell	739	14,616	0	0.016554501
Dendritic cell	1052	118	5.22E-127	6.455088729
Endothelial cell	1855	1888	2.38E-30	0.678962305
Exhausted CD8+ T cell	57	15	0.000284218	2.689329736
Hepatocyte	1805	0	0	Inf
Kupffer cell	8413	5103	1.80E-22	1.212994547
Liver bud hepatic cell	11,409	1199	0	9.198176601
Liver progenitor cell	3948	303	0	10.16866219
Memory B cell	1270	418	1.36E-47	2.188061303
Mucosal-associated invariant T cell	2203	1231	7.77E-12	1.281728373
Myofibroblast	1747	372	3.49E-126	3.432715387
Regulatory T (Treg) cell	2906	748	1.44E-167	2.892126064

FC: fold change.

samples was observed. Dendritic cells, hepatocytes, liver bud hepatic cells, and liver progenitor cells were highly enriched in tumor tissues (Table 2). Using Monocle to predict state trajectories, we reasonably inferred dynamic immune states and cellular transitions. The four cell subpopulations were differentiated into three branches, with each representing a different state (Figure S7A). Dendritic cells and liver progenitor cells were at the initial developmental trajectory, while hepatocytes were at the end (Figure S7B). The Seraut_cluster locus diagram of four subgroups showed that dendritic cells were in Clusters 18 and 22. Cluster 18 was in state 1 and 2 branches, while subgroup 22 was in the state 2 branch only. Cluster 11 was enriched in hepatocytes. There were eight cell clusters (5, 7, 10, 14, 15, 16, 19, and 23) in liver bud hepatic cells. Clusters 5, 10, 14, and 15 were dominant in the terminal state. Liver progenitor cells were annotated in Clusters 2 and 21, whereas 2 was in state 3, and 21 was in states 1 and 3 (Figure S7C, Figure S8).

Intercellular and molecular interaction networks of four cell subpopulations

To study the functions of dendritic cells, hepatocytes, liver bud hepatic cells, and liver progenitor cell subgroups, the corresponding markers of each subgroup were extracted for further KEGG analysis. Figure S9A displayed the heat map of the top three enriched pathways and the top 50 genes specifically expressed in each subgroup. Dendritic cells were associated with tyrosine metabolism, infection by *Staphylococcus aureus*, and retinol metabolism. The hepatocyte marker genes were mainly enriched in proximal tubule bicarbonate reclamation, protein export, and pentose and glucuronate interconversions. The specific genes in liver progenitor cells were annotated in mineral absorption, glycine, metabolism of serine and threonine, complement, and coagulation cascades. Liver bud hepatic cells played an important role in butanoate metabolism, asthma, and allograft rejection. As scRNA-seq can indicate the cellular interactions, owing to the integration of ligand and receptor information, therefore, CellChat was used to construct the interaction network of 13 cell subpopulations, and the changes in the number and intensities of ligand–receptor interactions were shown in Figure S9B. The cell-ligand–receptor interaction network of the cell subgroups comprising dendritic cells, hepatocytes, liver bud hepatic cells, and liver progenitor cells was developed. The results showed that the interactions among cells were complex; for instance, the communication between dendritic cells and the other three types of cells was realized through 50 ligand–receptor interaction pairs. Hepatocytes, liver bud hepatic cells, and liver progenitor cells supported the cellular communication through 64, 50, and 57 ligand–receptor crosslinks, respectively (Figure S9C). MIF – (CD74 + CXCR4) was the most potent ligand–receptor pair interaction between hepatocytes and B cells, and it also performed an integral function in the interaction between hepatocytes and other cell types, including Kupffer cells, dendritic cells, memory B cells, mucosal-associated invariant T cell, cytotoxic CD4+ T cells, and Treg cells. Furthermore, these interactions were strong. In addition, MIF – (CD74 + CXCR4) showed the strongest binding upon liver progenitor cell interaction with Kupffer cells, B cells, dendritic cells, mucosal-associated invariant T cells, cytotoxic CD4+ T cells, memory B cells, and regulatory T (Tregs) cells (Figure S9D).

Development and verification of the nine-gene signature prognosis risk model

The differential analysis identified 3024 DEGs (2529 high-expressed and 492 genes with lower expressions in HCC) between HCC tissues and para cancer tissues. Among them, 641 were specific markers of four subgroups (Figure S10). Premised on the selection criteria of the univariate Cox proportional regression analysis on the 3024 DEGs, 266 genes were considerably linked to the HCC patients' survival (Table S1). For TCGA specimens, LASSO modeling was performed to select 20 prognostic genes with multiple variables through ten-fold cross-validation (Figure 2(A) and (B)). To obtain the optimal model, nine genes (GTPBP4, TXN, ERBB3, PPP1R1A, CYP2C9, CENPU, SCGN, CD4, and SEMA7A, Table S2) were selected for developing the prognostic risk model by the stepAIC method. Protein–protein interaction (PPI) analysis on these nine prognostic genes revealed that five of them (GTPBP4, TXN, ERBB3, CD4, and SEMA7A) were directly or indirectly interacted within a PPI network

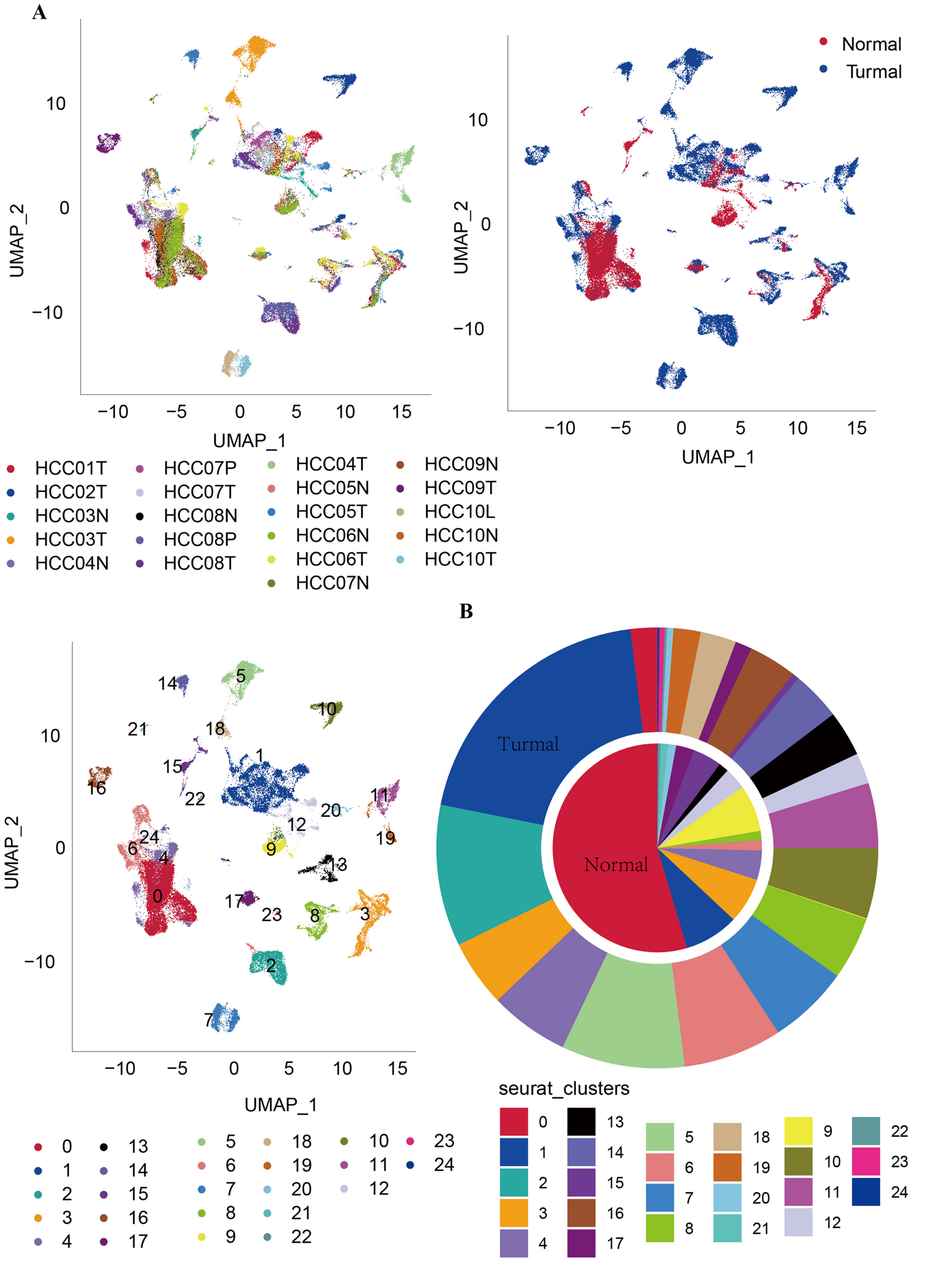


Figure 1. Clustering analysis of single-cell RNA-seq data for HCC. (A) UMAP of 64,634 cells, color-coded correspondingly as a patient (top panel), sample type (middle panel), and cell cluster (right panel). (B) In each sample, the abundance of 25 cell clusters. (A color version of this figure is available in the online journal.)

(Figure S11), suggesting that these five genes may play a central role for HCC prognosis.

Each sample's risk score was obtained by multiplying the expression levels of the nine genes using their correlation coefficients derived from the multivariate analysis. The Kaplan–Meier analysis showed lower rates of survival in the TCGA-high-risk HCC cohort as opposed to the low-risk HCC patients at various time points (Figure 2(C)). For the 5-year OS prediction in TCGA-LIHC cohorts, the AUC value of the ROC curve was found to be 0.86 (Figure 2(D)). In the validation cohort HCCDB18, patients diagnosed with HCC who were classified as having a low risk experienced considerably better survival in contrast with those classified as having a high risk (Figure 2(E)). Consequently, the AUC value for 5-year OS was found to be 0.71 in the validation set (Figure 2(F)).

Risk model as an independent prognostic marker for HCC

The association of risk score with clinical parameters including gender, AJCC stage (I, II, III, and IV), M stage (M0 and M1), N stage (N0 and N1), T stage (T1, T2, T3, and T4), and tumor grade (G1, G2, G3, and G4) was analyzed for the HCC samples from TCGA, which were grouped to calculate the differences in risk scores across various groups. The findings illustrated no remarkable differences in risk scores between the samples grouped basis of M stage, N stage, and gender (Figure 3(A), (C), and (D)). However, considerable differences were found in the risk scores for the samples grouped according to the T stage, with a higher T stage indicating a greater risk score (Figure 3(B)). Patients were grouped according to the AJCC stage to assess the correlation according to the risk scores. The findings illustrated a substantial correlation between the risk scores and the AJCC stage (Figure 3(E)). For patients grouped based on grade, it was found that their risk scores increased in direct proportion to the grade (Figure 3(F)). Multivariate and univariate Cox analyses were performed using the aforementioned six clinical parameters and risk scores to establish whether or not the nine-gene risk model may be used independently for clinical application. Among all seven clinical factors, we discovered that the risk score was the only one that independently served as a prognostic predictor (Figure 3(G) and (H)).

Relationship between risk scores and pathways

By employing gene set variation analysis (GSVA), we compared the high- and low-risk groups to determine the difference in pathway enrichment. After calculating the scores for each individual pathway, we investigated how closely those values correlated with the risk scores. By selecting pathways whose correlations are greater than 0.4, we discovered a total of 17 pathways with significant inverse correlation and nine pathways with significant positive correlation with the samples' risk scores (Figure 4(A)). Cluster analysis based on the accumulation scores of the 26 KEGG pathways demonstrated that nine pathways, including those involving non-homologous DNA, end joining, cell cycle, oocyte meiosis, spliceosome, nucleotide excision repair, base excision repair, mismatch repair, DNA replication, and homologous recombination, were significantly enriched with the increase of the risk score (Figure 4(B)). Thus, these pathways might have a possible involvement in the malignant advancement of HCC.

Discussion

Intratumoral heterogeneity contributes to the non-responsiveness of most cancer types to current therapies.[29] Intratumor heterogeneity can only be fully characterized at the single-cell level.[30] Single-cell techniques reveal intratumoral heterogeneity through epigenomics, proteomics, transcriptomics, and genomics, of the cell constituents and their spatial distributions, and this has greatly improved the development of both basic and translational cancer research.[31] In this study, 25 cell clusters were identified by analyzing the scRNA-seq data from 64,634 single cells, which were subsequently grouped into 13 cell types.

We found that dendritic cells, hepatocytes, liver bud hepatic cells, and liver progenitor cells were highly enriched in the HCC tissues. They were divided into three branches, with each representing a different state. The differentiation of dendritic cells, the most effective antigen-presenting cell type in immune response, was generally impaired during tumor development.[32] Therefore, dendritic cells were only involved in the initial states of the estimated differentiation pathway. Notably, in addition to dendritic cells, aberrant metabolism processes were observed in liver bud hepatocytes, liver progenitor cells, and hepatocytes.

Among the genes specifically expressed in highly enriched cell subsets in HCC tissues, 641 were differentially expressed between HCC and para cancer tissues, which may contribute significantly to the overall prognosis of patients with HCC. Using classical univariate Cox regression analysis and Lasso modeling, we identified nine HCC prognostic markers and established a gene signature. Some of these markers played important roles in the onset and advancement of cancer. GTPBP4 is oncogenic in HCC[33] and lung adenocarcinoma.[34] Up-modulation of TXN is linked to unfavorable HCC prognosis and promotes HCC metastasis *in vitro* and *in vivo*.[35] ERBB3 is often abnormally activated in many human cancers, and inhibition of ERBB3 signaling is important to overcome therapeutic resistance.[36] PPP1R1A mediates tumorigenesis and metastasis in Ewing sarcoma, and its consumption leads to a substantial reduction in the cell migration and oncogenic transformation *in vitro* and the xenograft tumor growth and metastasis in mouse models.[37] CYP2C9 performs an instrumental function in DNA methylation as well as iron metabolism in HCC and is considerably linked to the HCC patients' prognoses.[38] High CENPU levels are substantially linked to the absence of distant metastasis and overall survival. *In vitro* cell experiments show that CENPU knockout inhibits vascular endothelial growth factor A production in triple-negative breast cancer cells, and significantly reduces tube formation as well as the migratory ability of endothelial cells in the human umbilical veins *in vitro*.[39] SCGN has lower expression in colorectal cancer cells, and it drives cell apoptosis and attenuates cell migration and invasion.[40] SEMA7A has been previously identified as a new target to block the advancement of breast tumors.[41]

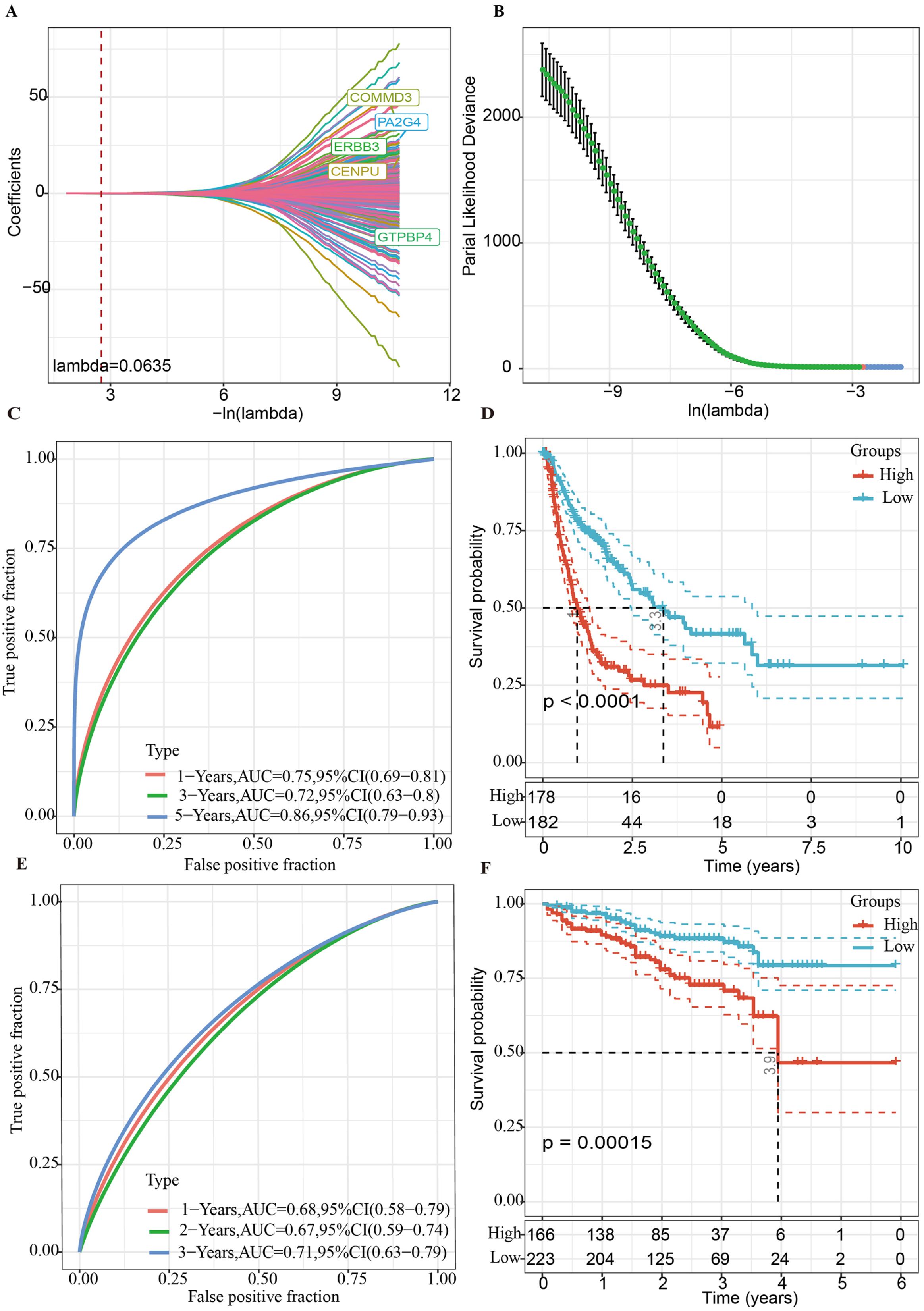
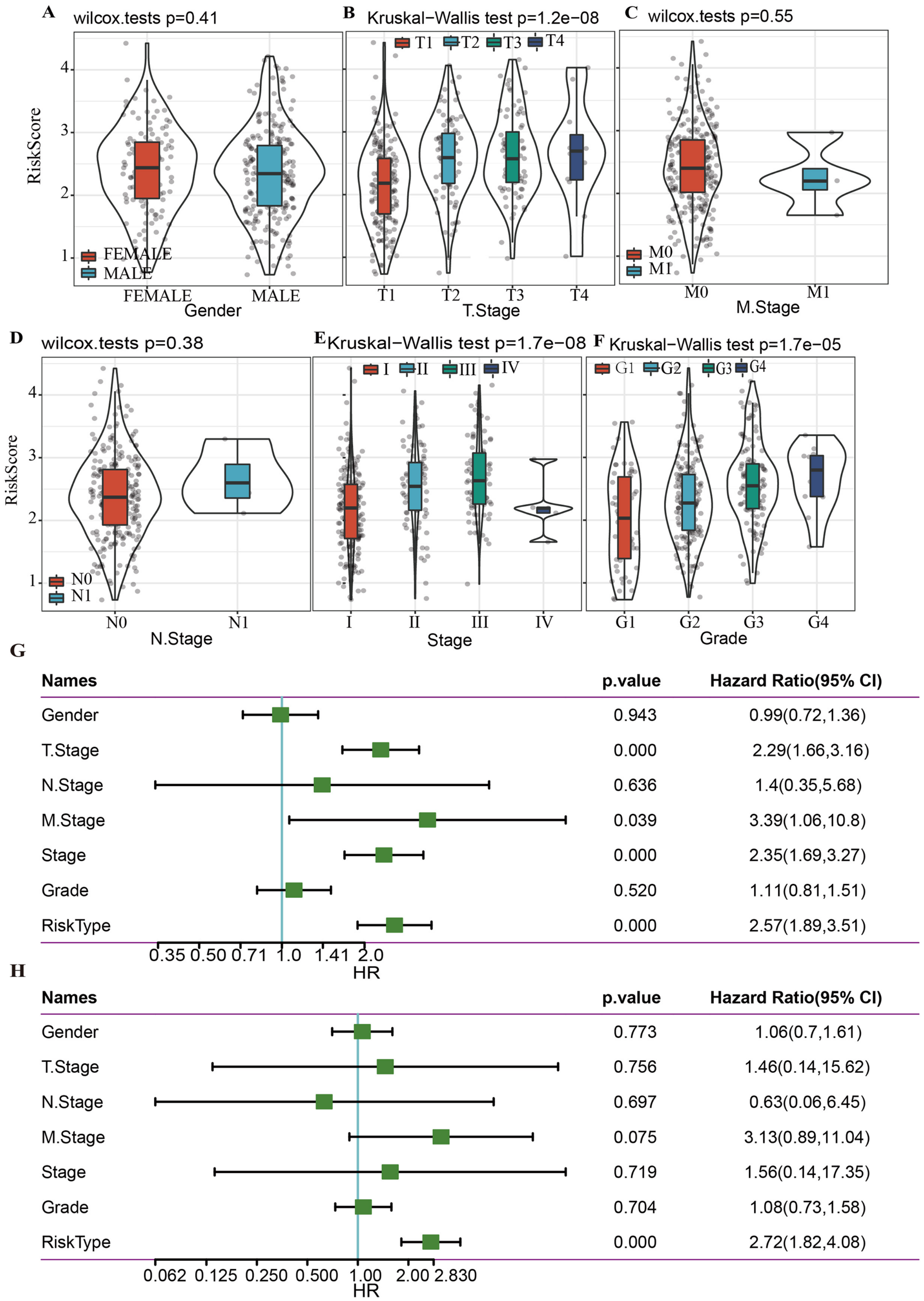


Figure 2. Construction and verification of the prognosis risk model composed of nine genes. (A, B) LASSO analysis based on the glmnet package was used for identifying the prognostic genes in the TCGC-LIHC dataset, and the λ values were determined according to the partial likelihood deviance after 10-fold cross-validation. (C) TCGA-LIHC cohort was used for survival analysis for comparing the prognosis of patients in the high- and low-risk groups. (D) ROC curves of the risk model in the TCGA dataset. (E) Risk score Kaplan–Meier survival curve for risk score in the HCCDB18 cohort. (F) The ROC curves of the risk model in the HCCDB18 cohort. (A color version of this figure is available in the online journal.)



G

Names	p.value	Hazard Ratio(95% CI)
Gender	0.943	0.99(0.72,1.36)
T.Stage	0.000	2.29(1.66,3.16)
N.Stage	0.636	1.4(0.35,5.68)
M.Stage	0.039	3.39(1.06,10.8)
Stage	0.000	2.35(1.69,3.27)
Grade	0.520	1.11(0.81,1.51)
RiskType	0.000	2.57(1.89,3.51)

H

Names	p.value	Hazard Ratio(95% CI)
Gender	0.773	1.06(0.7,1.61)
T.Stage	0.756	1.46(0.14,15.62)
N.Stage	0.697	0.63(0.06,6.45)
M.Stage	0.075	3.13(0.89,11.04)
Stage	0.719	1.56(0.14,17.35)
Grade	0.704	1.08(0.73,1.58)
RiskType	0.000	2.72(1.82,4.08)

Figure 3. Relationship between the risk scores of different subgroups with various clinical factors. (A) Risk score differences among HCC samples grouped by gender. (B) Association of risk score with T stage. (C) Comparison of risk scores of HCC patients grouped by N stage. (D) Correlation of M stage and risk score. (E) Differences in risk scores between different AJCC stages. (F) Changes in risk score with tumor grades. (G) In the TCGA-LIHC database, the forest plot for the univariate Cox analysis assessed the association of clinical factors with patient prognosis and the risk score. (H) Multivariate Cox regression analysis identified variables with independent prognostic significance according to the clinical characteristics and risk scores. (A color version of this figure is available in the online journal.)

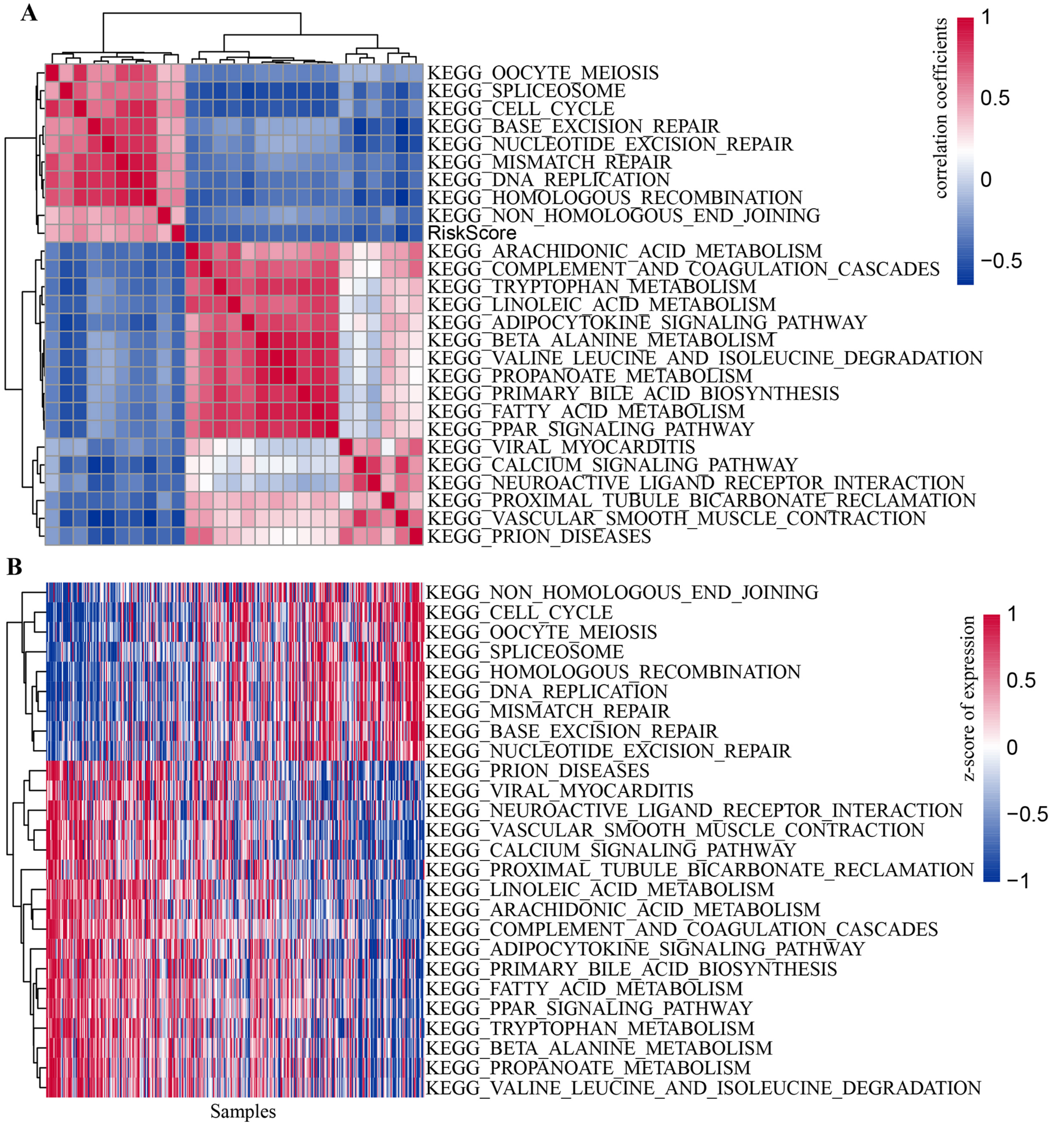


Figure 4. GSVA of low- and high-risk score groups in the TCGA-LIHC database. (A) Clustering correlation coefficients for KEGG pathways and risk score larger than 0.4. (B) Heatmap for the contribution of single-sample gene set enrichment analysis (ssGSEA) scores to hallmarks in the low- and high-risk groups. (A color version of this figure is available in the online journal.)

In summary, almost all the nine genes in the risk assessment model were associated with the malignant biological progression of cancer. Therefore, we reasonably speculated that the nine-gene signature had great potential in anticipating the clinical prognosis of patients with HCC.

The nine-gene prognostic model could be used to evaluate HCC prognostic survival by calculating the risk score of HCC samples and showed higher 5-year AUC values in the training and validation sets. In addition, the risk score independently served as a prognostic marker, and HCC samples at advanced AJCC stage, T stage, and clinical grade had significantly higher risk scores. Non-homologous DNA end joining, cell cycle, oocyte meiosis, spliceosome, nucleotide excision repair, base excision repair, mismatch repair, DNA replication, and homologous recombination were considerably enriched with the increase in risk score.

In conclusion, we integrated the scRNA-seq data and multi-omic data to evaluate the characteristics of different cell subsets of immune and tumor cells in the HCC microenvironment and established a novel nine-gene prognostic model predicated specifically on the gene expression levels of important subsets. The prognostic model can be applied to determine the death risk of HCC patients. This study may contribute to a deeper understanding of underlying heterogeneity in HCC and offer a potential new tool for the establishment of individualized treatment for HCC.

AUTHORS' CONTRIBUTIONS

The design of the research, the assessment of the studies, the analysis of the data, and the evaluation of the manuscript were all collaborative efforts involving all the authors. CL and MP conceived of and developed the research. The literature research was performed by YM, CW, and XZ. LK, SZ, and XL performed the data analysis and interpretation. CL drafted the original version of the manuscript and edited and revised it. The manuscript was reviewed and approved by all authors.

DECLARATION OF CONFLICTING INTERESTS

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

FUNDING

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Clinical research project of Air Force Medical Center (grant number 2021LC013) and the Youth Science and Technology Support Project of Air Force Medical Center (grant number 2020KTC14).

ORCID ID

Chengli Liu https://orcid.org/0000-0002-9654-7486

SUPPLEMENTAL MATERIAL

Supplemental material for this article is available online.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021;**71**:209–49
2. Lee SK, Lee SW, Jang JW, Bae SH, Choi JY, Yoon SK. Immunological markers, prognostic factors and challenges following curative treatments for hepatocellular carcinoma. *Int J Mol Sci* 2021;**22**:10271
3. Lawal G, Xiao Y, Rahnemai-Azar AA, Tsilimigras DI, Kuang M, Bakopoulos A, Pawlik TM. The immunology of hepatocellular carcinoma. *Vaccines (Basel)* 2021;**9**:1184
4. Wu Y, Liu Z, Xu X. Molecular subtyping of hepatocellular carcinoma: a step toward precision medicine. *Cancer Commun (Lond)* 2020;**40**:681–93
5. Saviano A, Henderson NC, Baumert TF. Single-cell genomics and spatial transcriptomics: discovery of novel cell states and cellular interactions in liver physiology and disease biology. *J Hepatol* 2020; **73**:1219–30
6. Ding S, Chen X, Shen K. Single-cell RNA sequencing in breast cancer: understanding tumor heterogeneity and paving roads to individualized therapy. *Cancer Commun (Lond)* 2020;**40**:329–44
7. Zhang M, Hu S, Min M, Ni Y, Lu Z, Sun X, Wu J, Liu B, Ying X, Liu Y. Dissecting transcriptional heterogeneity in primary gastric adenocarcinoma by single cell RNA sequencing. *Gut* 2021;**70**:464–75
8. Li H, Courtois ET, Sengupta D, Tan Y, Chen KH, Goh JJL, Kong SL, Chua C, Hon LK, Tan WS, Wong M, Choi PJ, Wee LJK, Hillmer AM, Tan IB, Robson P, Prabhakar S. Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat Genet* 2017;**49**:708–18
9. Chung W, Eum HH, Lee HO, Lee KM, Lee HB, Kim KT, Ryu HS, Kim S, Lee JE, Park YH, Kan Z, Han W, Park WY. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nat Commun* 2017;**8**:15081
10. Hu J, Chen Z, Bao L, Zhou L, Hou Y, Liu L, Xiong M, Zhang Y, Wang B, Tao Z, Chen K. Single-cell transcriptome analysis reveals intratumoral heterogeneity in ccRCC, which results in different clinical outcomes. *Mol Ther* 2020;**28**:1658–72
11. Zhang C, He H, Hu X, Liu A, Huang D, Xu Y, Chen L, Xu D. Development and validation of a metastasis-associated prognostic signature based on single-cell RNA-seq in clear cell renal cell carcinoma. *Aging (Albany NY)* 2019;**11**:10183–202
12. Wan Q, Liu C, Liu C, Liu W, Wang X, Wang Z. Discovery and validation of a metastasis-related prognostic and diagnostic biomarker for melanoma based on single cell and gene expression datasets. *Front Oncol* 2020;**10**:585980
13. Zhang Q, Lou Y, Yang J, Wang J, Feng J, Zhao Y, Wang L, Huang X, Fu Q, Ye M, Zhang X, Chen Y, Ma C, Ge H, Wang J, Wu J, Wei T, Chen Q, Wu J, Yu C, Xiao Y, Feng X, Guo G, Liang T, Bai X. Integrated multiomic analysis reveals comprehensive tumour heterogeneity and novel immunophenotypic classification in hepatocellular carcinomas. *Gut* 2019;**68**:2019–31
14. Li B, Feng W, Luo O, Xu T, Cao Y, Wu H, Yu D, Ding Y. Development and validation of a three-gene prognostic signature for patients with hepatocellular carcinoma. *Sci Rep* 2017;**7**:5517
15. Liu GM, Xie WX, Zhang CY, Xu JW. Identification of a four-gene metabolic signature predicting overall survival for hepatocellular carcinoma. *J Cell Physiol* 2020;**235**:1624–36
16. Li N, Zhao L, Guo C, Liu C, Liu Y. Identification of a novel DNA repair-related prognostic signature predicting survival of patients with hepatocellular carcinoma. *Cancer Manag Res* 2019;**11**:7473–84
17. Li W, Chen QF, Huang T, Wu P, Shen L, Huang ZL. Identification and validation of a prognostic lncRNA signature for hepatocellular carcinoma. *Front Oncol* 2020;**10**:780
18. Yan J, Zhou C, Guo K, Li Q, Wang Z. A novel seven-lncRNA signature for prognosis prediction in hepatocellular carcinoma. *J Cell Biochem* 2019;**120**:213–23
19. Zhao QJ, Zhang J, Xu L, Liu FF. Identification of a five-long non-coding RNA signature to improve the prognosis prediction for patients with hepatocellular carcinoma. *World J Gastroenterol* 2018;**24**:3426–39
20. Lian Q, Wang S, Zhang G, Wang D, Luo G, Tang J, Chen L, Gu J. HCCDB: a database of hepatocellular carcinoma expression atlas. *Genom Proteom Bioinf* 2018;**16**:269–75
21. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 2018;**36**:411–20
22. Zhang X, Lan Y, Xu J, Quan F, Zhao E, Deng C, Luo T, Xu L, Liao G, Yan M, Ping Y, Li F, Shi A, Bai J, Zhao T, Li X, Xiao Y. CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Research* 2019;**47**:D721–128
23. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012;**16**:284–7
24. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, Trapnell C. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* 2017;**14**:979–82
25. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Research* 2019;**47**:W199–205
26. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, Myung P, Plikus MV, Nie Q. Inference and analysis of cell-cell communication using CellChat. *Nat Commun* 2021;**12**:1088
27. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010;**33**:1–22
28. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Statistics in Medicine* 2013;**32**:5381–97

29. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol* 2018;**15**:81–94
30. Rantalainen M. Application of single-cell sequencing in human cancer. *Brief Funct Genomics* 2018;**17**:273–82
31. Pan D, Jia D. Application of single-cell multi-omics in dissecting cancer cell plasticity and tumor heterogeneity. *Front Mol Biosci* 2021;**8**:757024
32. Mion F, Tonon S, Valeri V, Pucillo CE. Message in a bottle from the tumor microenvironment: tumor-educated DCs instruct B cells to participate in immunosuppression. *Cell Mol Immunol* 2017;**14**:730–2
33. Chen J, Zhang J, Zhang Z. Upregulation of GTPBP4 promotes the proliferation of liver cancer cells. *J Oncol* 2021;**2021**:1049104
34. Zhang Z, Wang J, Mao J, Li F, Chen W, Wang W. Determining the clinical value and critical pathway of GTPBP4 in lung adenocarcinoma using a bioinformatics strategy: a study based on datasets from the cancer genome atlas. *Biomed Res Int* 2020;**2020**:5171242
35. Cao MQ, You AB, Cui W, Zhang S, Guo ZG, Chen L, Zhu XD, Zhang W, Zhu XL, Guo H, Deng DJ, Sun HC, Zhang T. Cross talk between oxidative stress and hypoxia via thioredoxin and HIF-2alpha drives metastasis of hepatocellular carcinoma. *FASEB J* 2020; **34**:5892–905
36. Ma J, Lyu H, Huang J, Liu B. Targeting of erbB3 receptor to overcome resistance in cancer treatment. *Mol Cancer* 2014;**13**:105
37. Luo W, Xu C, Ayello J, Dela Cruz F, Rosenblum JM, Lessnick SL, Cairo MS. Protein phosphatase 1 regulatory subunit 1A in Ewing sarcoma tumorigenesis and metastasis. *Oncogene* 2018;**37**:798–809
38. Shuaichen L, Guangyi W. Bioinformatic analysis reveals CYP2C9 as a potential prognostic marker for HCC and liver cancer cell lines suitable for its mechanism study. *Cell Mol Biol (Noisy-le-grand)* 2018;**64**:70–4
39. Pan T, Zhou D, Shi Z, Qiu Y, Zhou G, Liu J, Yang Q, Cao L, Zhang J. Centromere protein U (CENPU) enhances angiogenesis in triple-negative breast cancer by inhibiting ubiquitin-proteasomal degradation of COX-2. *Cancer Lett* 2020;**482**:102–11
40. Yang XY, Liu QR, Wu LM, Zheng XL, Ma C, Na RS. Overexpression of secretagogin promotes cell apoptosis and inhibits migration and invasion of human SW480 human colorectal cancer cells. *Biomed Pharmacother* 2018;**101**:342–7
41. Black SA, Nelson AC, Gurule NJ, Futscher BW, Lyons TR. Semaphorin 7a exerts pleiotropic effects to promote breast tumor progression. *Oncogene* 2016;**35**:5170–8

(Received February 16, 2022, Accepted June 7, 2022)