

Research Article

Seven Hub Genes Predict the Prognosis of Hepatocellular Carcinoma and the Corresponding Competitive Endogenous RNA Network

Xueqiong Han,¹ Jianxun Lu,¹ Chun Chen,² Yongran Deng,¹ Mingmei Pan,¹ Qigeng Li,¹ Huayun Wu,¹ Zhenlong Li ¹ and Bingqiang Ni ¹

¹Department of Oncology, The Fifth Affiliated Hospital of Guangxi Medical University & the First People's Hospital of Nanning, 89 Qixing Road, Nanning, Guangxi 530022, China

²Department of Cardiology and Endocrinology, The Guangxi Zhuang Autonomous Region Workers' Hospital, Nanning, China

Correspondence should be addressed to Zhenlong Li; zhenlongli@gxzarwh.org and Bingqiang Ni; nbq181@aliyun.com

Xueqiong Han, Jianxun Lu, and Chun Chen contributed equally to this work.

Received 20 June 2022; Revised 15 August 2022; Accepted 18 August 2022; Published 12 October 2022

Academic Editor: Ashok Pandurangan

Copyright © 2022 Xueqiong Han et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Purpose. This study was aimed at identifying hub genes and ceRNA regulatory networks linked to prognosis in hepatocellular carcinoma (HCC) and to identify possible therapeutic targets. **Methods.** Differential expression analyses were performed to detect the differentially expressed genes (DEGs) in the four datasets (GSE76427, GSE6764, GSE62232, and TCGA). The intersected DEmRNAs were identified to explore biological significance by enrichment analysis. We built a competitive endogenous RNA (ceRNA) network of lncRNA-miRNA-mRNA. The mRNAs of the ceRNA network were used to perform Cox and Kaplan-Meier analyses to obtain prognosis-related genes, followed by the selection of genes with an area under the curve >0.8 to generate the random survival forest model and obtain feature genes. Furthermore, the feature genes were subjected to least absolute shrinkage and selection operator (LASSO) and univariate Cox analyses were used to identify the hub genes. Finally, the infiltration status of immune cells in the HCC samples was determined. **Results.** A total of 1923 intersected DEmRNAs were identified in four datasets and involved in cell cycle and carbon metabolism. ceRNA network was created using 10 lncRNAs, 67 miRNAs, and 1,923 mRNAs. LASSO regression model was performed to identify seven hub genes, SOCS2, MYO2, FTCD, ADAMTSL2, TMEM106C, LARS, and KPNA2. Among them, TMEM106C, LARS, and KPNA2 had a poor prognosis. KPNA2 was considered a key gene base on LASSO and Cox analyses and involved in the ceRNA network. T helper 2 cells and T helper cells showed a higher degree of infiltration in HCC. **Conclusion.** The findings revealed seven hub genes implicated in HCC prognosis and immune infiltration. A corresponding ceRNA network may help reveal their potential regulatory mechanism.

1. Introduction

According to 2018 estimates provided by Bray about the incidence of cancer, liver cancer was responsible for 841,080 new cancer cases globally [1] and the incidence rates are expected to increase remarkably by the year 2030 [2]. Hepatocellular carcinoma (HCC) is the most prevalent form of liver cancer, representing 75-85% of all liver cancer cases

[3]. Previous studies confirmed that the main pathogenic factors of HCC are the chronic hepatitis B virus (HBV) or hepatitis C virus (HCV) infection, alcoholic liver disease, and nonalcoholic fatty liver disease [4]. Statistics indicate that 30%-40% of HCC patients are diagnosed in the early stage [5]. Only a few methods for the prognosis and treatment of HCC, these methods have been limited because most HCC patients are diagnosed in advanced stages and

surgically unresectable. Additionally, the methods depend mainly on the tumor stage [6] to reflect the development of tumor cells [7]. Therefore, identifying a sound risk stratification system can effectively treat and improve outcomes.

Competing endogenous RNA (ceRNA), including long noncoding RNA (lncRNA) and circle RNA (circRNA), can combine competitive microRNA (miRNA) and interfere with miRNA binding to messenger RNA (mRNA) to regulate gene expression and play important roles [8]. It has been shown that microRNAs, by virtue of their capacity to interact with many target genes, affect a wide variety of crucial biological processes, including growth, proliferation, and apoptosis of cells [9]. High expression of YKT6 [10] and MTFR2 [11] associated with progression and poor prognosis of HCC. Furthermore, EPHX2 was identified as an independent prognostic biomarker for overall survival of patients with HCC [12]. In summary, ceRNA is a factor that influences the incidence of HCC as well as its progression [13, 14]. The ceRNA was used to understand the interactions of complex genes and identified the potential biomarkers for diagnosing and treating HCC.

In this research, we mainly used the expression patterns of the databases in the Gene Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/geo/>) and the Cancer Genome Atlas (TCGA) (<https://portal.gdc.cancer.gov/>) between HCC and normal samples to perform bioinformatics analysis. The purpose of this study is to construct a potentially competitive endogenous RNA (ceRNA) network to identify the underlying biological mechanisms of HCC. Furthermore, the model classifier and the risk score model were utilized to more precisely identify possible markers associated with prognosis in HCC patients.

2. Material and Methods

2.1. Data Preprocessing. There were a total of 600 HCC and 122 normal samples included within four different datasets (GSE76427, GSE6764, GSE62232, and TCGA datasets). RNA sequencing (RNA-seq) data, microRNA sequencing data, and the corresponding survival data of liver hepatocellular carcinoma (LIHC) patients were obtained from the TCGA database [15]. RNA-seq data and microRNA sequencing data of TCGA contained 369 primary HCC and 50 normal tissues as well as 370 primary HCC and 50 normal tissues, respectively. Corresponding clinicopathological features (age, sex, tumor differentiation degree, TNM stage, survival time, and status) were also obtained from the TCGA database are publicly available. Among these samples, one formalin sample and one relapse sample were excluded. Gene expression data of GSE76427, GSE6764, and GSE62232 were available from the GEO database. GSE76427 comprised 52 adjoining nontumor tissues as normal control and 115 tumor tissues with HCC. HCC patients included in this dataset had a mean age of 63.45 ± 12.63 years, and 93 male and 22 female [16], conducted by GPL10558 platform (Illumina HumanHT-12 V4.0 expression beadchip). GSE6764 was conducted by the GPL570 platform (Affymetrix Human Genome U133 Plus 2.0 Array) and comprised 35 HCC tissues and 10 adjoining nontumor

tissues. HCV infection cases observed in 13 samples from cirrhotic tissues and 17 samples from dysplastic nodules were excluded [17]. GSE62232 containing 81 solid HCC and 10 nontumor liver samples was acquired on the basis of the GPL570 platform (Affymetrix Human Genome U133 Plus 2.0 Array). The individuals of HCC included in this dataset had a mean age of 60.6 ± 13.49 years, and 67 male and 14 female [18]. The “varianceStabilizingTransformation” function of the DESeq2 package [19] was utilized for normalizing the expression patterns of RNA sequencing data and microRNA sequencing from the TCGA dataset. Furthermore, expression profiles of GSE6764 and GSE62232 were normalized using the “RMA” function in the Affy package. The expression profile of GSE76427 was used to normalize by the “lumiExpresso” function in the Lumi R package.

2.2. Identification of Differentially Expressed Genes. By employing the limma package in R, differential expression analysis was carried out for gene expression to find the DEGs between HCC tumor tissues and nontumor liver tissues in GSE76427, GSE6764, and GSE62232 [20]. We also utilized the DESeq2 package [19] to identify the DEGs of TCGA. DEGs of all the four datasets (GSE76427, GSE6764, GSE62232, and TCGA) were deemed to have statistical significance if the adjusted P value was <0.05 . Subsequently, the intersected DEGs of four datasets were identified using the Venn diagrams to obtain consistent expression, including upregulated and downregulated DEGs.

2.3. Gene Ontology (GO) and Pathway Enrichment Analysis. Cellular components (CCs), Biological processes (BPs), and molecular functions (MFs) of intersected DEGs were used to explore the biological significance by performing a GO analysis. Kyoto Encyclopedia of Genes and Genomes (KEGG) was utilized to explore significantly altered pathways enriched in the gene list. GO and KEGG pathways were executed with the help of the clusterProfile package [21]. Gene enrichment in the GO and KEGG pathway with P value <0.05 were judged as significant.

2.4. Gene Set Enrichment Analysis (GSEA). GSEA is a computational approach that enables gene sets to identify genomes excessively increasing or decreasing between biological phenotypes [22]. GSEA analysis was carried out to ascertain the functions premised on the expressed profiles of TCGA using the GSEA software (Version 4.1.0).

2.5. ceRNA Network Analysis. We examined the regulated miRNAs using up/downregulated lncRNAs in four datasets by Starbase databases (<http://starbase.sysu.edu.cn/index.php>) while retaining the opposite expression direction of lncRNAs and miRNAs. Among them, we extracted the intersection of the regulated miRNAs of TCGA for the next analysis. Meanwhile, we screened for miRNA-regulated mRNAs on the Targetscan databases (https://www.targetscan.org/vert_72/) while retaining the same expression direction of lncRNAs and mRNAs. Subsequently, we generated the lncRNA-miRNA-mRNA regulatory network

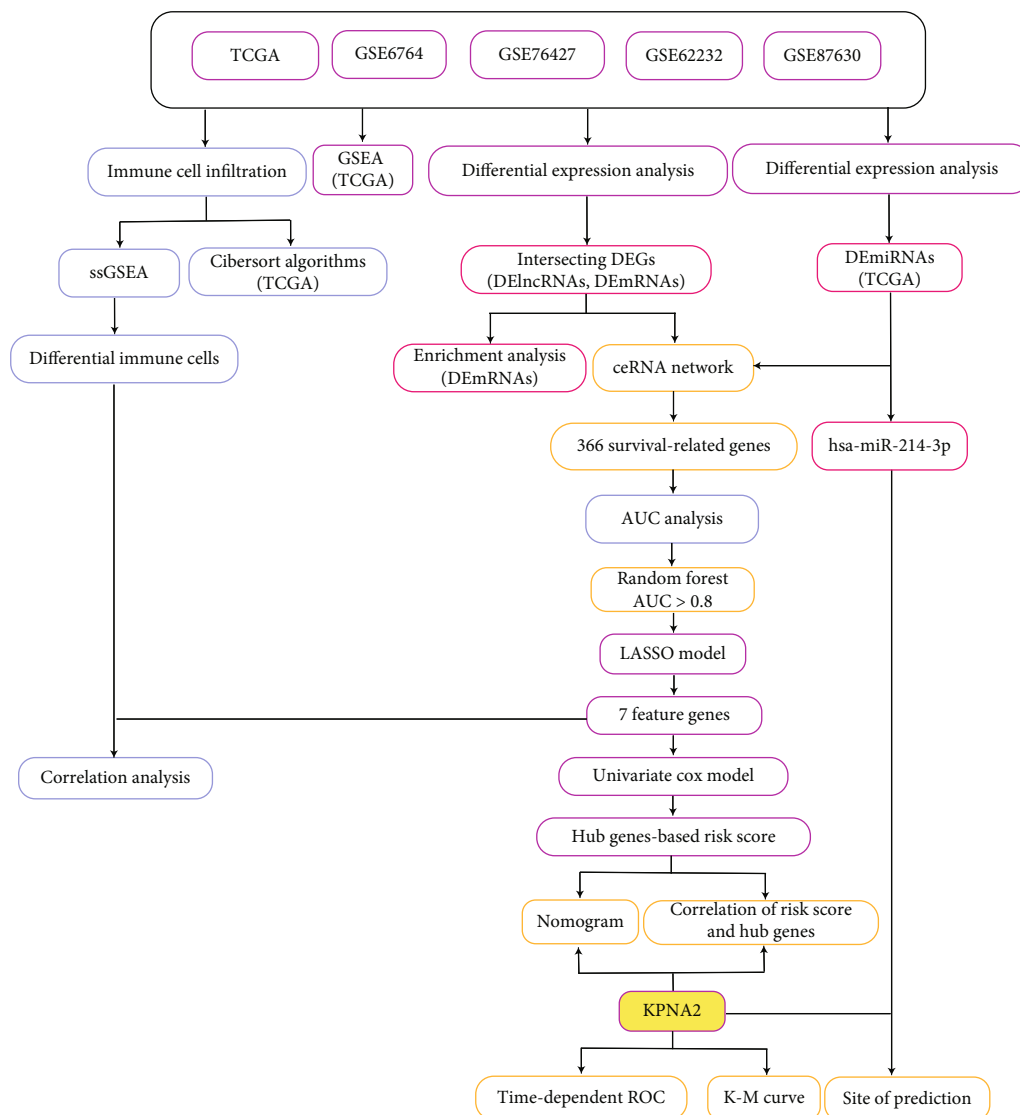


FIGURE 1: The detailed flow chart for this research AUC, area under the curve; DElncRNA, differentially expressed long noncoding RNA; DEG, differentially expressed genes; DEmRNA, differentially expressed messenger RNA; Gene Set Enrichment Analysis; LASSO, least absolute shrinkage, and selection operator; K-M curve, Kaplan-Meier curve; TCGA, The Cancer Genome Atlas; ssGSEA, single sample Gene Set Enrichment Analysis; ROC, receiver operating characteristic curve.

by downloading the binding sites of mRNA, miRNA, and lncRNA using the Starbase and Targetscan databases.

2.6. Establishment of Random Survival Forest Model and Least Absolute Shrinkage and Selection Operator (LASSO) Regression Model. The mRNAs of the ceRNA regulatory network were used to perform Cox and Kaplan-Meier analyses in TCGA to obtain prognosis-related genes. We selected the prognosis-related genes for the area under the curve (AUC) analysis by pROC package. Next, we used prognosis-related genes with an AUC >0.8 to construct the random survival forest model with the help of the RandomSurvivalForest R package. It was determined that genes with a relative importance >0.4 were the ultimate feature genes by examining the link between the error rate and the number of classification trees. Moreover, the feature genes were used to construct

the LASSO model and the hub genes were obtained using “cv.glmnet” function in glmnet R package [23]. Subsequently, a univariate Cox analysis was carried out on the hub genes to determine the prognostic significance utilizing the forestplot R package. The hub genes with a hazard ratio >1 were considered to lead to a poor prognosis.

2.7. Seven Hub Gene-Risk Scores Based on Cox Regression Analysis. After performing a multivariate Cox regression analysis premised on the outcomes of seven hub genes, risk scores were then computed. Following the construction of the risk score model premised on the median risk score, the HCC patients of the TCGA dataset were further classified into high- and low-risk groups, and their overall survival (OS) rates were compared. Furthermore, to determine the impact that hub genes have on the HCC patients’ prognoses,

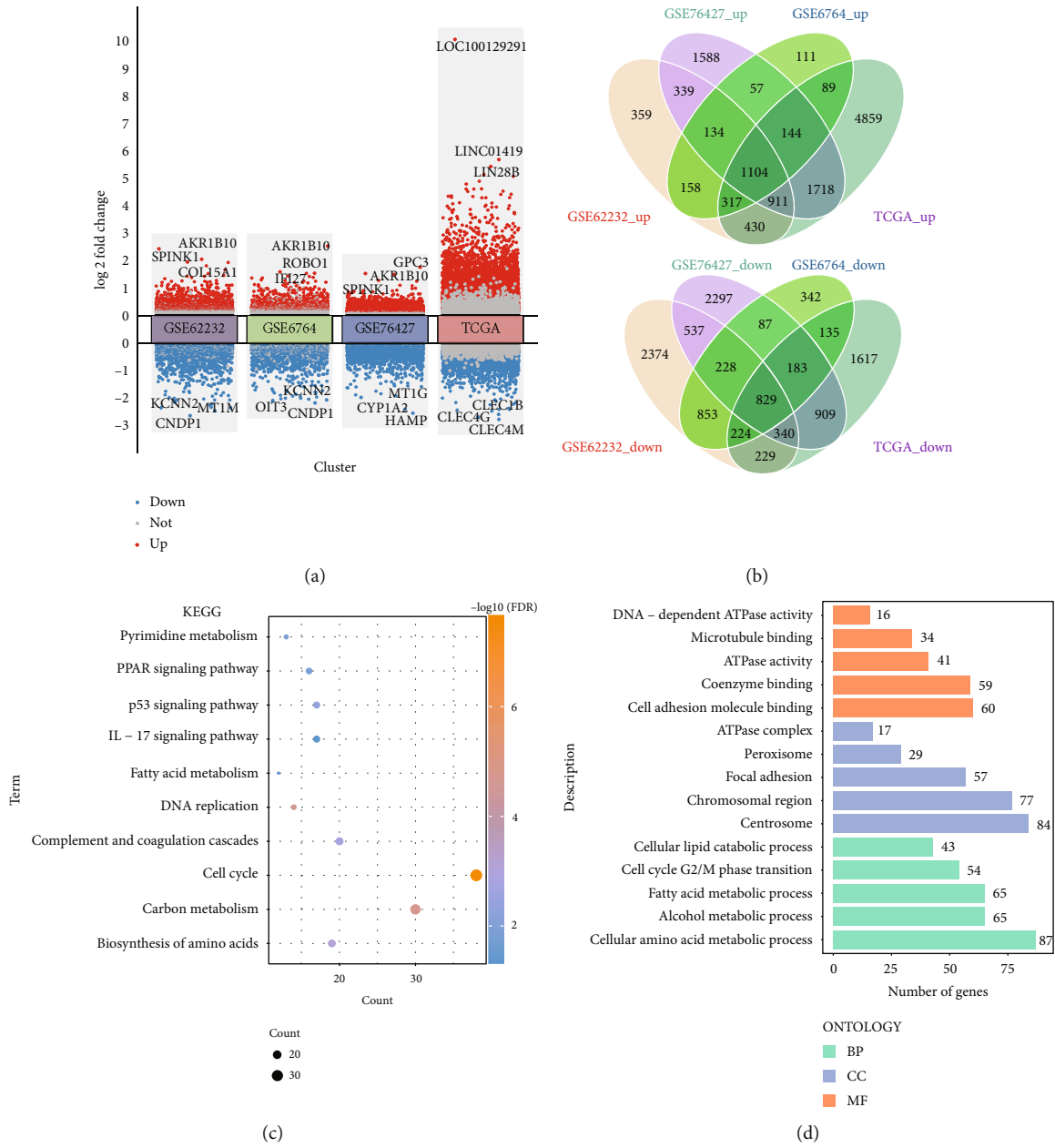


FIGURE 2: Continued.

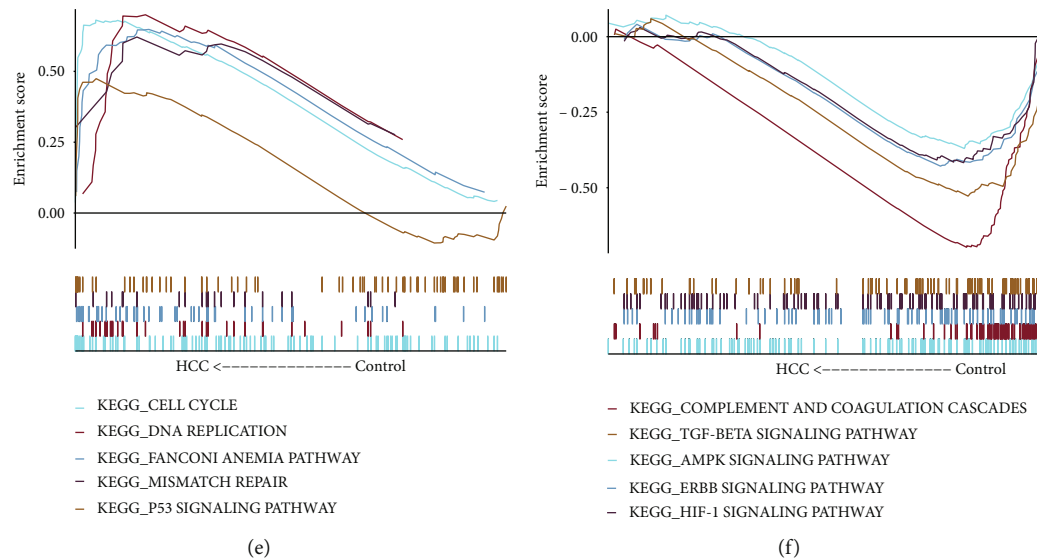


FIGURE 2: Enrichment analysis of differentially expressed genes (DEGs). (a) DEGs of GSE62232, GSE6764, GSE76427, and TCGA datasets between HCC and controls. Red denotes upmodulated DEGs, whereas blue denotes downmodulated DEGs. (b) DEGs in the same direction across all four datasets are shown by a Venn diagram. (c) Intersected DEGs of four datasets were involved in pathways by enrichment analysis. (d) Intersected DEGs of four datasets were involved in biological process, molecular function, and cellular component. (e, f) Gene set enrichment analysis illustrated the head and tail of five pathways enriched in HCC patients. KEGG, Kyoto Encyclopedia of Genes and Genomes; HCC, hepatocellular carcinoma; CC, cellular component; MF, molecular function; BP, biological process.

a nomogram was developed with the use of the rms package of the R program. With the help of the ggstatsplot package, correlations between hub genes and risk scores were derived.

2.8. Identification of Key Gene. To further illustrate the expression of hub genes between HCC and control samples in the TCGA dataset, we created a heat map and a violin plot. Above all, time-dependent receiver operating characteristic curve (ROC) analysis was conducted using the “survivalROC” R package to evaluate the prediction accuracy of the 1-, 3-, 5- year of the key gene. The Kaplan–Meier survival curve was utilized to make a comparison between the high- and low-risk groups for the survival of the key gene. Furthermore, P adjust value <0.05 was selected as the criterion for determining differentially expressed miRNAs (DEmiRs), after which we plotted the associated ceRNA regulatory network of key genes.

2.9. Immune Infiltration Analysis. Single-sample gene set enrichment analysis (ssGSEA) calculated the degree of immune cell infiltration in the HCC patients of 24 immune cell types using marker gene sets [24]. Using the limma R program, variations in the types of immune cells seen between HCC and control samples were computed Radar and scatter plots show correlation plots of risk score of immune cells and hub gene. Second, we adopted the Pearson correlation to determine the link between the seven hub genes and the immune cell types. Additionally, The CIBERSORT algorithm was utilized to execute an analysis of the infiltration levels of 22 different types of immune cells in HCC samples taken from the TCGA dataset.

2.10. Statistical Analysis. The analyses of the present study were conducted utilizing the Bioinforcloud platform (<http://www.bioinforcloud.org.cn>).

3. Results

Workflow of the present study (Figure 1).

3.1. Biological Function of DEGs between HCC and Controls. To obtain dysfunctional genes associated with HCC, we identified DEGs between HCC and controls (Figure 2(a)). In total, 9,366 DEGs were identified in GSE62232, 11,405 DEGs in GSE76427, 4,995 DEGs in GSE7696, and 14,038 DEGs in TCGA. A total of 1,933 intersected DEGs were detected in four datasets, including 10 lncRNA and 1,923 mRNA. Among them, 1,104 were upregulated and 829 downregulated DEGs in HCC and controls (Figure 2(b)). Intersected DEGs were involved in the cell cycle and carbon metabolism (Figure 2(c)). Additionally, intersected DEGs were involved in 1,584 BP, 210 CC, and 213 MF (Figure 2(d)). GSEA showed that genes of the TCGA dataset were positively linked to DNA replication and the cell cycle (Figure 2(e)) and were inversely linked to the ERBB signaling pathway and HIF-1 signaling pathway (Figure 2(f)).

3.2. Identification of Hub Genes Related to Prognosis. The ceRNA regulatory network was constructed using the 366 prognosis-related mRNAs that were obtained from Kaplan–Meier and Cox survival analyses (Table S1). Based on AUC analysis, 366 prognosis-related mRNAs were analyzed to determine their possible involvement in the GSE76427 and TCGA datasets. As depicted in Figure 3(a), 211 genes with

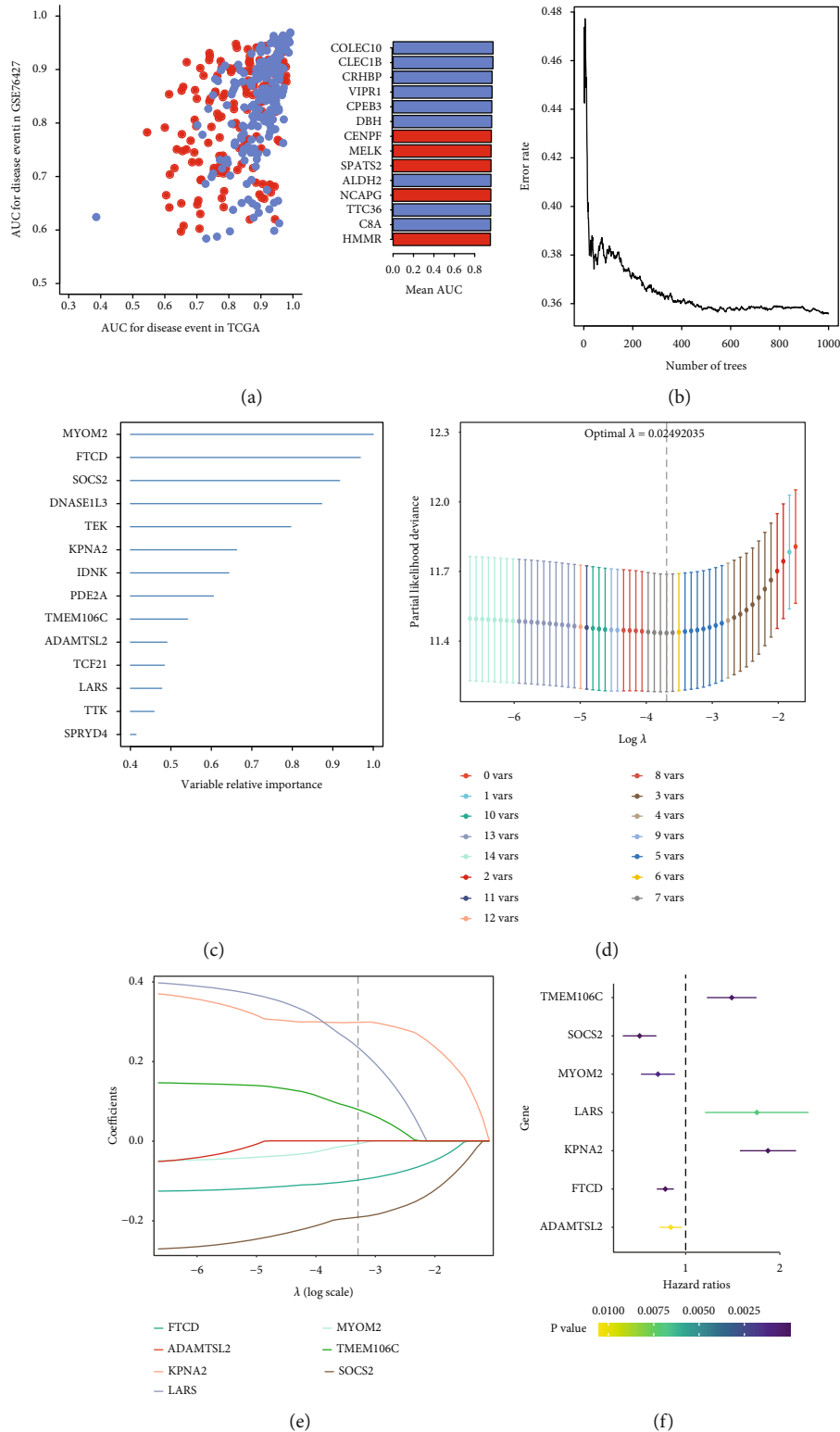
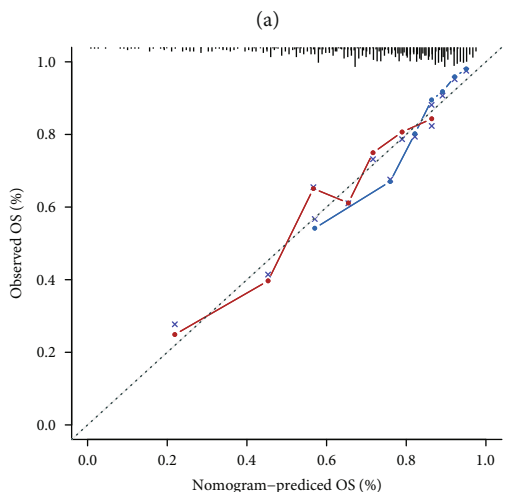
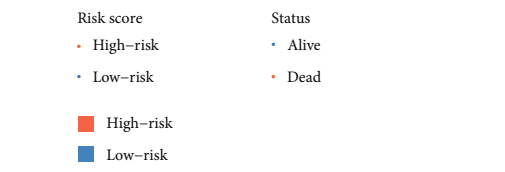
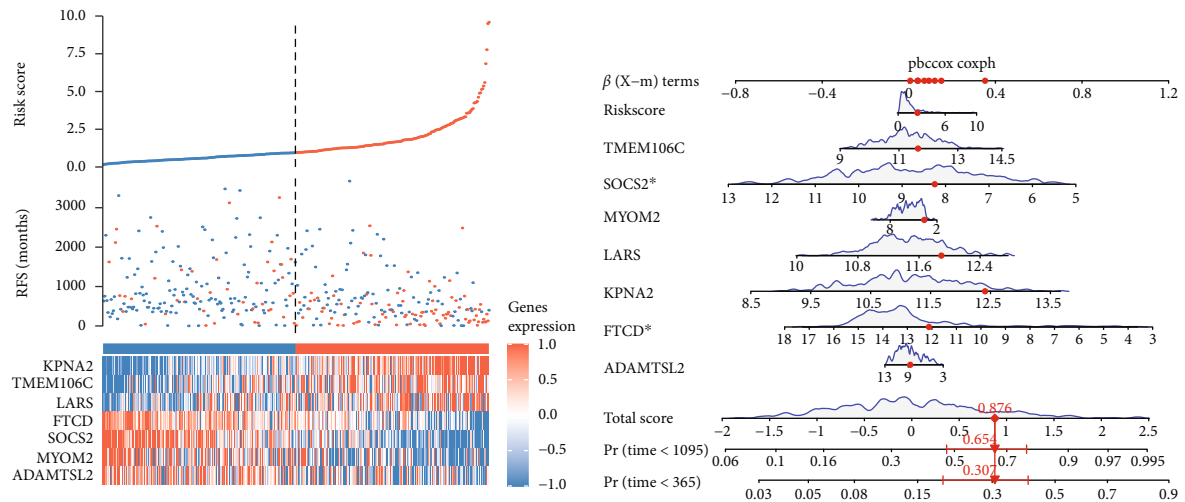
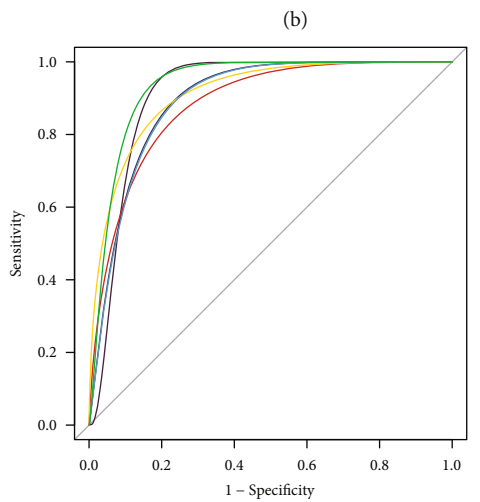


FIGURE 3: Identification of hub genes by least absolute shrinkage and selection operator (LASSO) and univariate Cox models. (a) The top 14 genes are shown with AUC>0.80 in TCGA and GSE76427. Genes that have been upregulated are shown in red, whereas genes that have been downregulated are shown in blue. (b) The correlation between the number of trees and the error rate. (c) A random forest model was constructed to determine the importance of the 14 genes in an order. (d) LASSO coefficient profiles of the 7 feature genes. (e) 10-fold cross-validation of parameter selection in the LASSO analysis. (f) Univariate analysis of feature genes. The feature genes with hazard ratios >1 are poor prognosis genes, and hazard ratios <1 are protective genes. TCGA, The Cancer Genome Atlas; AUC, area under the curve.



n = 363 d = 130 p = 8, 50 subjects per group
 gray: ideal
 X - resampling optimism added, B = 1000 based on observed-predicted

— 1-year
 — 3-year



— LARS, AUC = 0.91
 — ADAMTSL2, AUC = 0.9
 — TMEM106C, AUC = 0.9
 — KPNA2, AUC = 0.88
 — SOCS2, AUC = 0.91
 — FTCD, AUC = 0.9
 — MYOM2, AUC = 0.9

(c)

(d)

FIGURE 4: Continued.

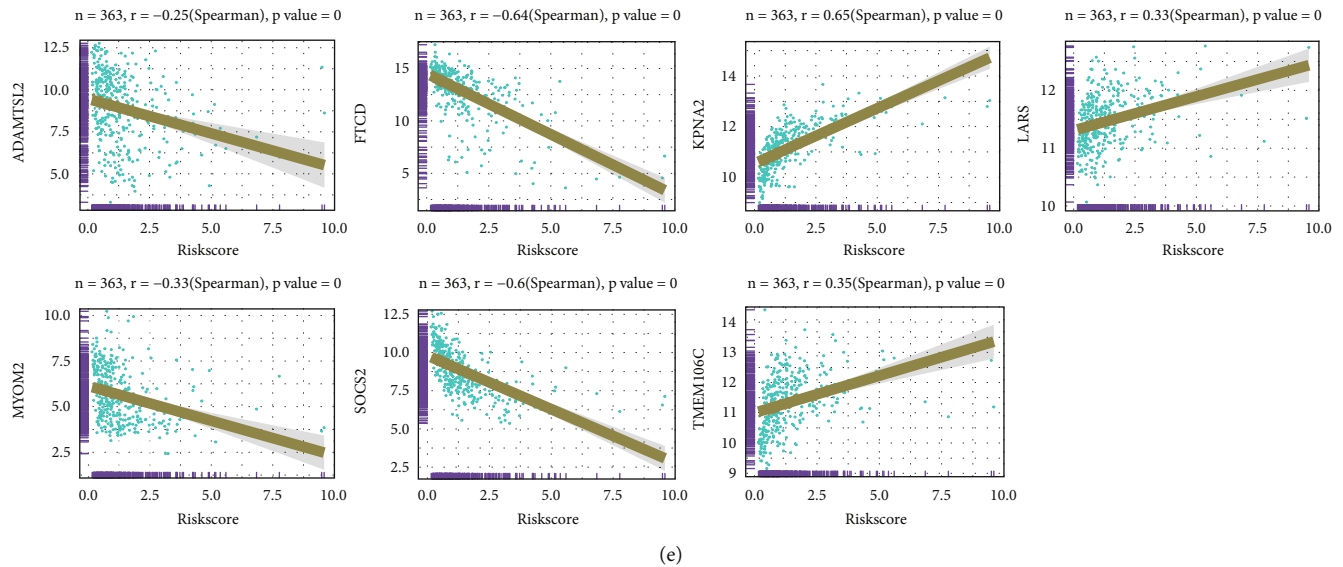


FIGURE 4: Seven feature gene-risk scores based on Cox analyses. (a) Expression, risk score, and survival status of seven genes in HCC patients of the TCGA dataset. (b) Nomogram for the prediction of 1- and 3-year overall survival (OS) for HCC patients with seven feature genes. (c) Calibration curves for seven feature genes and 1-, 3-year OS in the validation set (TCGA). (d) AUCs value of seven feature genes. (e) Correlation of risk score and seven hub genes. OS, overall survival; AUC, area under the curve.

an AUC of <0.80 were found in both datasets. Figures 3(b) and 3(c) provide an orderly representation of the association between the error rate for the data, the number of classification trees, and the importance of the 14 genes. LASSO regression model was performed to identify 7 hub genes, SOCS2, MYOM2, FTCD, ADAMTSL2, TMEM106C, LARS, and KPNA2 (Figures 3(e) and 3(f)). In univariate Cox regression analysis, TMEM106C, LARS, and KPNA2 had a poor prognosis (Figure 3(f)).

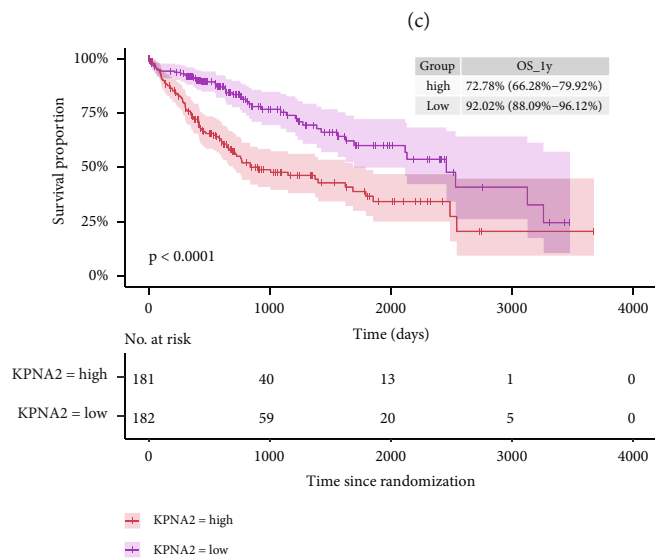
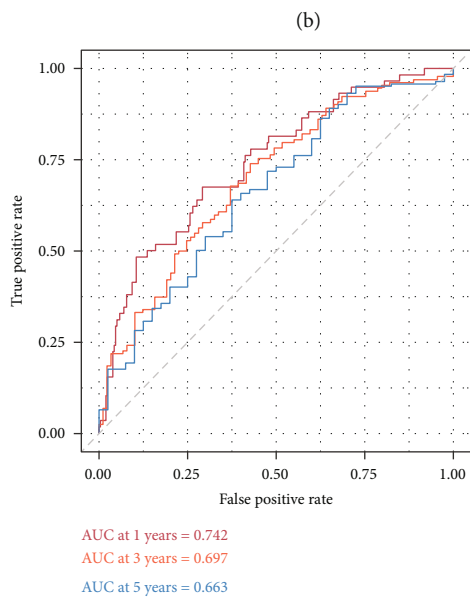
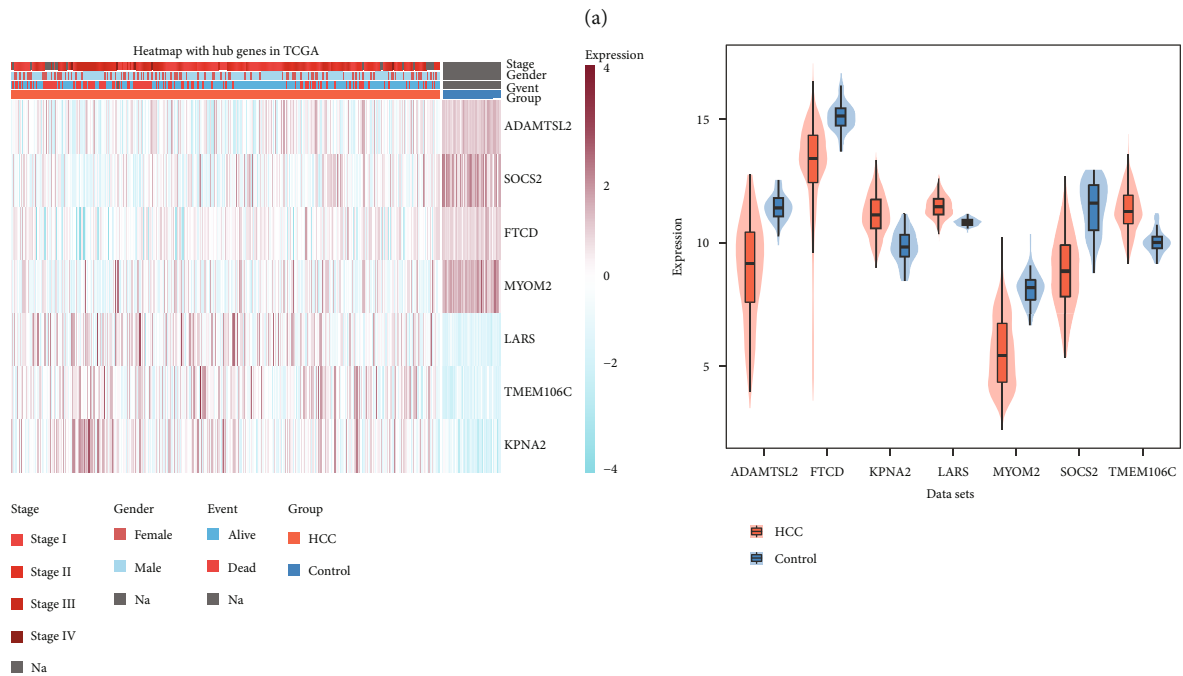
3.3. Construction of Hub Genes and Calculation of Risk Score in HCC. The distributions of the risk scores, RFS and hub gene expression of the 369 patients in the TCGA dataset are shown in Figure 4(a). The hub genes were incorporated into a nomogram model to predict the HCC patients' prognoses (Figure 4(b)). The calibration curve demonstrated excellent agreement between the observed and predicted OS over 1 and 3 years in the TCGA cohort (Figure 4(c)). The AUC analysis of seven hub genes as illustrated in Figure 4(d) and the relevant findings illustrated that the seven hub genes had a better diagnostic power in the prognostic model. We additionally examined the link between hub genes and risk scores and found that KPNA2, LARS, and TMEM106C had a positive link to risk scores, whereas an inverse correlation was observed between ADAMTSL2, MYOM2, FTCD, SOCS2, and risk scores (Figure 4(e)).

3.4. KPNA2 as a Key Gene in Prognosis for HCC. In four datasets (GSE76427, GSE6764, GSE62232, and TCGA), the results suggested that hub genes of AUCs value, adjusted P value, and fold change were depicted in Figure 5(a). A heat map showed the expression level of hub genes, stage, gender, event, and groups in TCGA (Figure 5(b)). In comparison with controls, ADAMTSL2, FTCD, KPNA2, LARS, and

TMEM106C were found to be expressed at a high level in HCC (Figure 5(c)). Time-dependent ROC survival analysis was employed to examine the prognosis of KPNA2 and the findings illustrated that the AUCs values over 1, 3, and 5 years were all greater than 0.66 (Figure 5(d)). Moreover, the predicted 1-year survival time for HCC patients indicated that KPNA2 had considerably improved OS (Figure 5(e)). The volcano plot showed a total of 901 DE miRNAs, comprising 730 upmodulated miRNAs and 171 downmodulated miRNAs (Figure 5(f)). Ultimately, Figure 5(g) depicts the obtained binding sites within the lncRNA HCP5-KPNA2-miR-214-3p as ceRNA regulatory network. Above all, KPNA2 performs an essential function in the ceRNA regulatory network, making it a key gene in the prognosis and fundamental biological processes involving HCC.

3.5. Estimation of Infiltrating Immune Cells in HCC. T helper (Th) 2 cells, T helper cells, and plasmacytoid dendritic cells (pDCs) all have higher degrees of infiltration between HCC and controls in four datasets (Figure 6(a)). Figure 6(b) depicts the association between the median risk score and the types of immune cells. Among them, CD8 T cells, Th17 cells, and DC exhibited a substantially positive correlation, whereas Th2 cells exhibited a significantly negative correlation (Figure 6(c)). By determining the Pearson correlation between the hub genes and the 24 different types of immune cells, we observed that KPNA2 and Th2 cells had a considerably high association in HCC samples (Figure 6(d)). To further evaluate the proportion of immune cells for TCGA, the findings indicated that HCC samples were extensively infiltrated by Macrophages M2 (Figure 6(e)).

Gene	TCGA AUC	TCGA logFold Change	TCGA p adjust	GSE76427 AUC	GSE76427 logFold Change	GSE76427 p adjust	GSE62232 AUC	GSE62232 logFold Change	GSE62232 p adjust	GSE6764 AUC	GSE6764 logfold change	GSE6764 p adjust
ADAMTSL2	0.8972358	-1.51 (rank: 03)	3.141662e-09	0.8933110	-1.88 (rank: 01)	1.710110e-16	0.8382716	-0.63 (rank: 07)	4.578539e-03	0.7885714	-0.42 (rank: 07)	4.185487e-02
FTCD	0.8968022	-1.39 (rank: 06)	2.860642e-09	0.9132107	-1.54 (rank: 03)	7.726937e-18	0.9641975	-1.59 (rank: 02)	4.777393e-04	0.7600000	-0.98 (rank: 03)	1.485547e-02
KPNA2	0.8829268	1.41 (rank: 05)	1.177452e-28	0.9408027	0.77 (rank: 06)	1.941301e-20	0.9592593	1.46 (rank: 03)	7.176953e-07	0.8360000	1.61 (rank: 01)	3.135729e-05
LARS	0.9142547	0.69 (rank: 07)	3.485848e-25	0.8655518	0.59 (rank: 07)	5.578071e-14	0.9901235	0.82 (rank: 05)	3.199658e-08	0.8800000	0.57 (rank: 06)	1.606274e-03
MYOM2	0.9402710	-2.06 (rank: 02)	4.677377e-16	0.9560564	-1.70 (rank: 02)	1.946919e-32	0.9444444	-0.80 (rank: 06)	1.390891e-06	0.8628571	-0.73 (rank: 05)	8.686889e-07
SOCS2	0.9147425	-2.19 (rank: 01)	4.233325e-25	0.8839465	-1.53 (rank: 04)	3.799142e-18	0.9370370	-1.93 (rank: 01)	1.373581e-07	0.8285714	-1.44 (rank: 02)	6.238659e-03
TMEM106C	0.8986992	1.50 (rank: 04)	4.330979e-31	0.8494147	1.01 (rank: 05)	1.868355e-12	0.9617264	0.86 (rank: 04)	1.179064e-05	0.9514286	0.76 (rank: 04)	1.409836e-04



(d)

(e)

FIGURE 5: Continued.

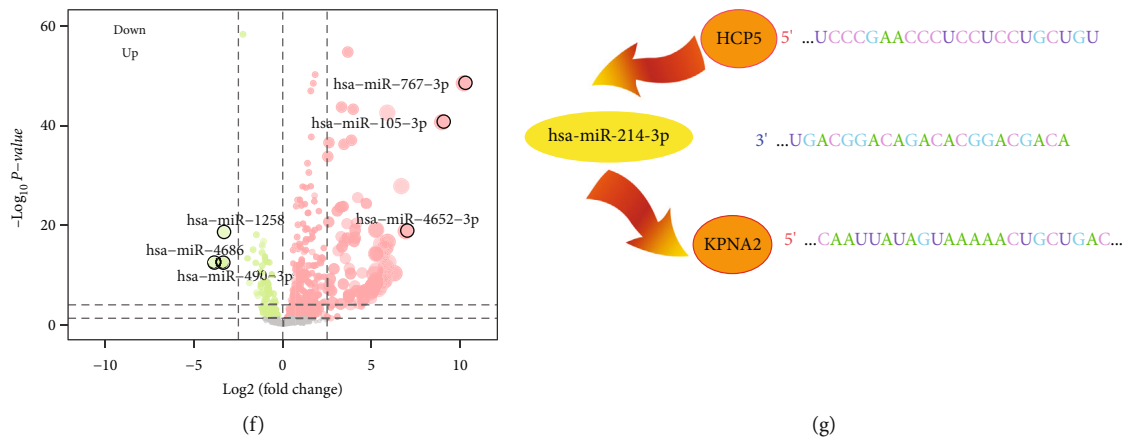


FIGURE 5: Expression of hub genes and construction of ceRNA regulated network. (a) The log-fold change, AUC, and adjusted P value of seven genes in TCGA, GSE76427, GSE62232, and GSE6764. (b) Stage, gender, event, and expressed levels were shown in HCC and control in HCC. (c) Violin plots illustrated the expression of seven hub genes. The thick black bar in the middle indicates the interquartile range, and the black line extending from it represents the 95% confidence interval. (d) Time-dependent receiver operating characteristic curve analysis displayed the AUC values over 1, 3, and 5 years. (e) Estimating the survival time of the KPNA2d by Kaplan-Meier survival curve. (f) Differentially expressed miRNAs were identified between controls and HCC samples in TCGA. Red represents upregulated miRNAs, whereas green represents downregulated miRNAs. (g) Bind sites of HCP5/has-miR-214-3p/KPNA2. Orange indicates upregulated, and yellow indicates down-regulated. HCC, hepatocellular carcinoma; AUC, area under the curve; TCGA, The Cancer Genome Atlas.

4. Discussion

HCC is a malignancy with a high death rate and an unfavorable prognosis, necessitating novel diagnostic and therapeutic markers [25]. In this research, we determined the seven hub genes of HCC prognosis as vital biomarkers, which can be used to improve outcomes of patients with HCC. We constructed a ceRNA regulatory network to explore the biological mechanism, including lncRNA HCP5-hsa-miR-214-3p-KPNA2. Furthermore, we also found a high degree of immune cell infiltration with seven hub genes.

Intersected DEGs of four datasets were involved in the cell cycle and carbon metabolism. Previous research has shown that cell cycle regulation inhibits the proliferative ability of HCC cells [26]. Importantly, GSEA showed that genes of the TCGA dataset were enriched in the cell cycle and carbon metabolism. Studies have shown that tumor cells often regulate the genes of the cell cycle producing damage and inactivation of this pathway may be involved in tumor development [27]. Furthermore, high expression of miR-452-5p performs an integral function in the progression of HCC through carbon metabolism [28].

Seven hub genes (SOCS2, MYOM2, FTCD, ADAMTSL2, TMEM106C, LARS, and KPNA2) involved in the process of the HCC prognosis were identified. SOCS2 was associated with distinct stages that indicated poor survival outcomes for patients with HCC [29]. Previous studies have shown that FTCD is a protective factor in HCC development and prognosis [30]. Overexpression of TMEM106C [31] and KPNA2 [32] predicted an unsatisfactory prognosis in HCC patients. Down-regulated MYOM2 was observed in a majority of clinical cases of breast cancer [33], however, the role of MYOM2, ADAMTSL2, and LARS in HCC development and prognosis remains unclear.

In this present research, in the TCGA liver cancer cohorts, a high level of KPNA2 expression accurately predicted the 1-, 3-, and 5-year survival times, with AUCs of 0.742, 0.697, and 0.663, correspondingly. A risk score and nomogram model also indicated that high-expressed KPNA2 led to an unfavorable prognosis. There was also a strong association between the risk score and KPNA2, which suggests that KPNA2 could be a crucial biological marker in determining the prognosis of patients with HCC. Recently, an increasing number of studies have demonstrated that lncRNA and miRNA primarily mediated posttranscriptional regulation, and that dysregulation of this process has been linked to many malignancies [34]. Interestingly, through targeting KPNA2, the lncRNA HCP5 served as a ceRNA, which had the effect of adversely modulating the expression of miR-214-3p. Downmodulation of lncRNA HCP5 in HCC tissues, when contrasted to normal samples, could affect the proliferation, metastatic and invasive, while the relevant mechanism of HCC still needs to be elucidated [35]. miR-214-3p is shown to modulate cell growth, metastasis, and apoptosis in HCC cells, endometrial cancer cells, and retinoblastoma cells by directly targeting certain genes linc00665 [36], ATWIST1 [37], BCB1, and XIAP [38]. In general, this ceRNA regulatory network further helps us understand the regulatory mechanisms of these genes in HCC.

ssGSEA indicated that Th2 cells and plasmacytoid dendritic cells (pDCs) all exhibited a higher degree of infiltration between HCC and controls. An intratumoral infiltration of pDCs is predictive of an unfavorable prognosis among patients who undergo curative resection for HCC; pDCs exist in numerous primary as well as metastatic human neoplasms [39]. Furthermore, the significantly positive correlation between Th2 cells and KPNA2 that Th2 cells were linked to HCC patient survival [40]. Th17 cells, dendritic

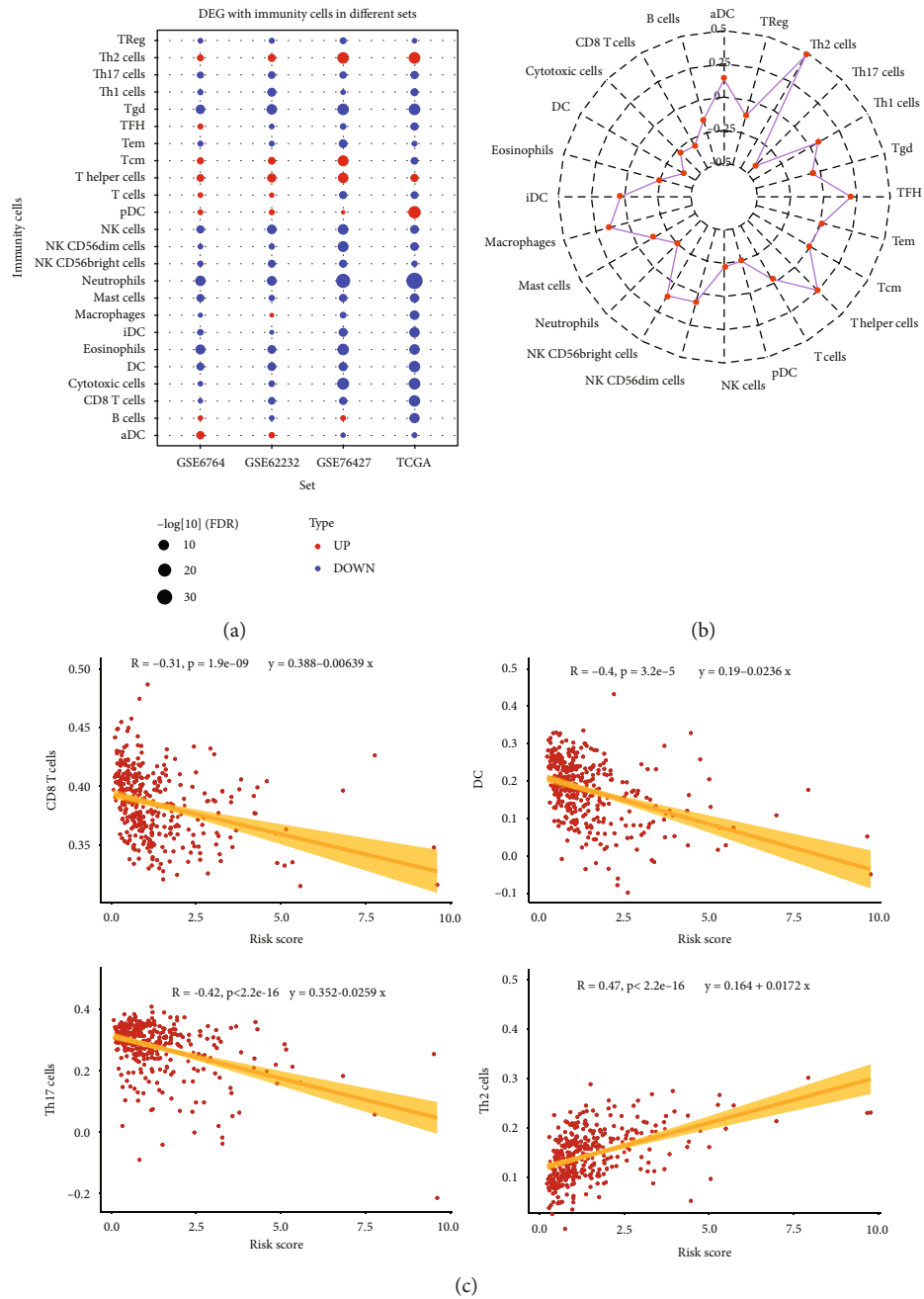


FIGURE 6: Continued.

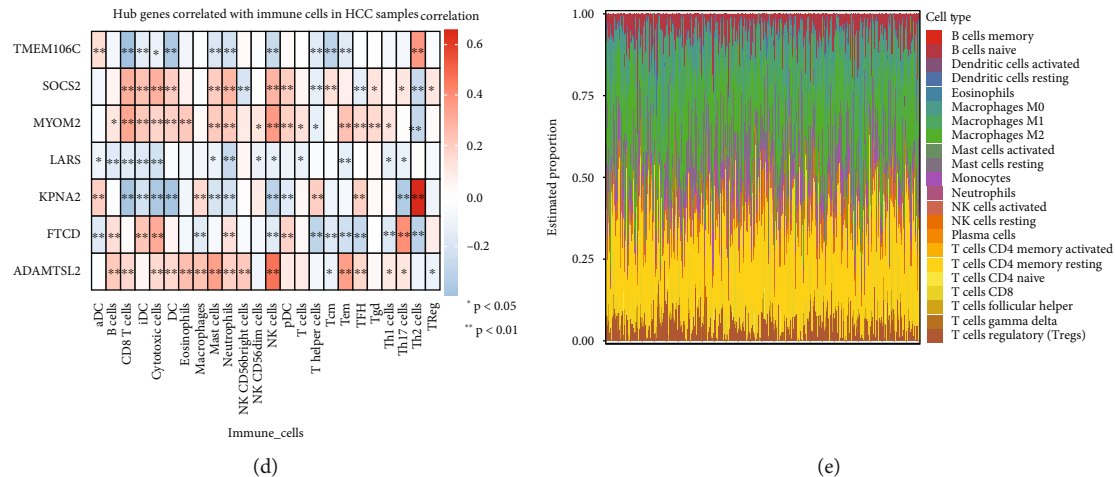


FIGURE 6: Immune cell infiltration in HCC. (a) Single sample gene set enrichment analysis was used to estimate the infiltration of immune cells. Red indicates high infiltration and blue indicates low infiltration. (b) Radar plot shows the correlation between 24 immune cell types and risk score of seven hub genes. (c) Correlation scatter plots shows the most significant infiltrating cells and risk score. (d) Correlation with seven hub genes and immune cells in HCC samples. (e) Estimated proportions of 22 immune cell types. DEGs, differentially expressed genes; HCC, hepatocellular carcinoma.

cells (DCs), and CD8 T cells were positively linked to the risk score of prognosis. The levels of Th17 cells were substantially elevated in tumors of patients with HCC [41]. DCs and CD8+ T cells have increased infiltration levels and are associated with relapse, compared with primary HCC [42]. The seven hub genes were defined as biomarkers for OS and we constructed a high immune cell infiltration model to predict the HCC patients' prognoses.

Despite the new findings at the level of bioinformatics analysis, understanding of the prognosis and immune-related biomarkers are still limited. Firstly, the markers associated with HCC currently lack sufficient sensitivity and specificity. Secondly, molecular and animal experiments are needed to verify the biomarkers and apply the biomarkers from preclinical studies in clinical practice.

5. Conclusion

SOCS2, MYOM2, FTCD, ADAMTSL2, TMEM106C, LARS, and KPNA2 are vital biomarkers and involved in the process of the HCC prognosis and immune infiltration.

Data Availability

The datasets (GSE76427, GSE6764, GSE62232) supporting the conclusions of this article are available in the Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>) and the Cancer Genome Atlas <https://portal.gdc.cancer.gov/> database.

Conflicts of Interest

The authors declare that they have no competing interest.

Authors' Contributions

Xueqiong Han was responsible for manuscript writing, Jianxun Lu was responsible for manuscript writing and proofreading, Yongran Deng was responsible for data collection, Mingmei Pan was responsible for data analysis, Qigeng Li was responsible for data collation, Huayun Wu was responsible for picture typesetting, Zhenlong Li was responsible for the construction of ideas, Bingqiang Ni was responsible for overall proofreading and review. Xueqiong Han, Jianxun Lu and Chun Chen contributed equally to this work and share first authorship.

Acknowledgments

The authors would like to thank Shaowen Mo and Qiong Song for assisting with bioinformatics analysis on the BioInforCloud platform.

Supplementary Materials

Supplementary Table 1 366 prognosis-related mRNA list. (*Supplementary Materials*)

References

- [1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.
- [2] P. C. Valery, M. Laversanne, P. J. Clark, J. L. Petrick, K. A. McGlynn, and F. Bray, "Projections of primary liver cancer to 2030 in 30 countries worldwide," *Hepatology*, vol. 67, no. 2, pp. 600–611, 2018.

- [3] S. Mittal and H. B. El-Serag, "Epidemiology of hepatocellular carcinoma: consider the population," *Journal of Clinical Gastroenterology*, vol. 47, Supplement 1, pp. S2–S6, 2013.
- [4] F. Kanwal and A. G. Singal, "Surveillance for hepatocellular carcinoma: current best practice and future direction," *Gastroenterology*, vol. 157, no. 1, pp. 54–64, 2019.
- [5] I. C. Lee, H. J. Lei, G. Y. Chau et al., "Predictors of long-term recurrence and survival after resection of HBV-related hepatocellular carcinoma: the role of HBsAg," *American Journal of Cancer Research*, vol. 11, no. 7, pp. 3711–3725, 2021.
- [6] C. Liu, L. G. Duan, W. S. Lu et al., "Prognosis evaluation in patients with hepatocellular carcinoma after hepatectomy: comparison of BCLC, TNM and Hangzhou criteria staging systems," *PLoS One*, vol. 9, no. 8, article e103228, 2014.
- [7] L. Zhou, J. A. Rui, W. X. Zhou, S. B. Wang, S. G. Chen, and Q. Qu, "Edmondson-Steiner grade: a crucial predictor of recurrence and survival in hepatocellular carcinoma without microvascular invasion," *Pathology, Research and Practice*, vol. 213, no. 7, pp. 824–830, 2017.
- [8] L. Salmena, L. Poliseno, Y. Tay, L. Kats, and P. P. Pandolfi, "A ceRNA Hypothesis: the rosetta stone of a hidden RNA language?," *Cell*, vol. 146, no. 3, pp. 353–358, 2011.
- [9] W. R. Farid, Q. Pan, A. J. van der Meer et al., "Hepatocyte-derived microRNAs as serum biomarkers of hepatic injury and rejection after liver transplantation," *Liver Transplantation*, vol. 18, no. 3, pp. 290–297, 2012.
- [10] J. Z. Xu, J. J. Jiang, H. J. Xu, X. D. Sun, Z. C. Liu, and Z. M. Hu, "High expression of YKT6 associated with progression and poor prognosis of hepatocellular carcinoma," *Scandinavian Journal of Gastroenterology*, vol. 56, no. 11, pp. 1349–1354, 2021.
- [11] D. Li, Y. Ji, J. Guo, and Q. Guo, "Upregulated expression of MTFR2 as a novel biomarker predicts poor prognosis in hepatocellular carcinoma by bioinformatics analysis," *Future Oncology*, vol. 17, no. 24, pp. 3187–3201, 2021.
- [12] K. Zhan, Y. Bai, S. Liao et al., "Identification and validation of EPHX2 as a prognostic biomarker in hepatocellular carcinoma," *Molecular Medicine Reports*, vol. 24, no. 3, 2021.
- [13] B. Li, R. Mao, C. Liu, W. Zhang, Y. Tang, and Z. Guo, "LncRNA FAL1 promotes cell proliferation and migration by acting as a ceRNA of miR-1236 in hepatocellular carcinoma cells," *Life Sciences*, vol. 197, pp. 122–129, 2018.
- [14] W. Y. Hu, H. Y. Wei, K. M. Li, R. B. Wang, X. Q. Xu, and R. Feng, "LINC00511 as a ceRNA promotes cell malignant behaviors and correlates with prognosis of hepatocellular carcinoma patients by modulating miR-195/EYA1 axis," *Biomedicine & Pharmacotherapy*, vol. 121, article 109642, 2020.
- [15] K. Tomczak, P. Czerwinska, and M. Wiznerowicz, "Review The cancer genome atlas (TCGA): an immeasurable source of knowledge," *Contemporary Oncology/Współczesna Onkologia*, vol. 2015, no. 1A, pp. 68–77, 2015.
- [16] O. V. Grinchuk, S. P. Yenamandra, R. Iyer et al., "Tumor-adjacent tissue co-expression profile analysis reveals pro-oncogenic ribosomal gene signature for prognosis of resectable hepatocellular carcinoma," *Molecular Oncology*, vol. 12, no. 1, pp. 89–113, 2018.
- [17] E. Wurmbach, Y. B. Chen, G. Khitrov et al., "Genome-wide molecular profiles of HCV-induced dysplasia and hepatocellular carcinoma," *Hepatology*, vol. 45, no. 4, pp. 938–947, 2007.
- [18] K. Schulze, S. Imbeaud, E. Letouzé et al., "Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets," *Nature Genetics*, vol. 47, no. 5, pp. 505–511, 2015.
- [19] M. I. Love, W. Huber, and S. Anders, "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2," *Genome Biology*, vol. 15, no. 12, p. 550, 2014.
- [20] M. E. Ritchie, B. Phipson, D. Wu et al., "Limma powers differential expression analyses for RNA-sequencing and microarray studies," *Nucleic Acids Research*, vol. 43, no. 7, p. e47, 2015.
- [21] G. Yu, L. G. Wang, Y. Han, and Q. Y. He, "clusterprofiler: an R package for comparing biological themes among gene clusters," *OMICS*, vol. 16, no. 5, pp. 284–287, 2012.
- [22] A. Subramanian, P. Tamayo, V. K. Mootha et al., "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 43, pp. 15545–15550, 2005.
- [23] S. Engebretsen and J. Bohlin, "Statistical predictions with glmnet," *Epigenetics*, vol. 11, no. 1, p. 123, 2019.
- [24] G. Bindea, B. Mlecnik, M. Tosolini et al., "Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer," *Immunity*, vol. 39, no. 4, pp. 782–795, 2013.
- [25] S. Shen, Y. Lin, X. Yuan et al., "Biomarker MicroRNAs for diagnosis, prognosis and treatment of hepatocellular carcinoma: a functional survey and comparison," *Scientific Reports*, vol. 6, no. 1, p. 38311, 2016.
- [26] Y. Zhao, C. Zhu, Q. Chang et al., "MiR-424-5p regulates cell cycle and inhibits proliferation of hepatocellular carcinoma cells by targeting E2F7," *PLoS One*, vol. 15, no. 11, article e0242179, 2020.
- [27] C. J. Sherr, "Cancer cell cycles," *Science*, vol. 274, no. 5293, pp. 1672–1677, 1996.
- [28] M. H. Rong, K. T. Cai, H. P. Lu et al., "Overexpression of MiR-452-5p in hepatocellular carcinoma tissues and its prospective signaling pathways," *International Journal of Clinical and Experimental Pathology*, vol. 12, no. 11, pp. 4041–4056, 2019.
- [29] J. Liu, Z. Liu, W. Li, and S. Zhang, "SOCS2 is a potential prognostic marker that suppresses the viability of hepatocellular carcinoma cells," *Oncology Letters*, vol. 21, no. 5, p. 399, 2021.
- [30] J. Liu, S. Q. Zhang, J. Chen et al., "Identifying prognostic significance of RCL1 and four-gene signature as novel potential biomarkers in HCC patients," *Journal of Oncology*, vol. 2021, Article ID 5574150, 20 pages, 2021.
- [31] X. Luo, G. Han, R. Lu et al., "Transmembrane protein 106C promotes the development of hepatocellular carcinoma," *Journal of Cellular Biochemistry*, vol. 121, no. 11, pp. 4484–4495, 2020.
- [32] X. Guo, Z. Wang, J. Zhang et al., "Upregulated KPNA2 promotes hepatocellular carcinoma progression and indicates prognostic significance across human cancer types," *Acta biochimica et Biophysica Sinica*, vol. 51, no. 3, pp. 285–292, 2019.
- [33] F. Yamamoto and M. Yamamoto, "Identification of genes that exhibit changes in expression on the 8p chromosomal arm by the systematic multiplex RT-PCR (SM RT-PCR) and DNA microarray hybridization methods," *Gene Expression*, vol. 14, no. 4, pp. 217–227, 2008.
- [34] Y. Li, X. Liu, X. Cui et al., "LncRNA PRADX-mediated recruitment of PRC2/DDX5 complex suppresses UBXN1 expression and activates NF- κ B activity, promoting tumorigenesis," *Theranostics*, vol. 11, no. 9, pp. 4516–4530, 2021.

- [35] Y. Zhou, K. Li, T. Dai et al., “Long non-coding RNA HCP5 functions as a sponge of miR-29b-3p and promotes cell growth and metastasis in hepatocellular carcinoma through upregulating DNMT3A,” *Aging (Albany NY)*, vol. 13, no. 12, pp. 16267–16286, 2021.
- [36] H. Wan, Y. Tian, J. Zhao, and X. Su, “LINC00665 targets miR-214-3p/MAPK1 Axis to accelerate hepatocellular carcinoma growth and Warburg effect,” *Journal of Oncology*, vol. 2021, Article ID 9046798, 14 pages, 2021.
- [37] Y. Y. Fang, M. R. Tan, J. Zhou et al., “miR-214-3p inhibits epithelial-to-mesenchymal transition and metastasis of endometrial cancer cells by targeting TWIST1,” *Oncotargets and Therapy*, vol. 12, pp. 9449–9458, 2019.
- [38] L. Yang, L. Zhang, L. Lu, and Y. Wang, “miR-214-3p regulates multi-drug resistance and apoptosis in retinoblastoma cells by targeting ABCB1 and XIAP,” *Oncotargets and Therapy*, vol. 13, pp. 803–811, 2020.
- [39] Z. J. Zhou, H. Y. Xin, J. Li, Z. Q. Hu, C. B. Luo, and S. L. Zhou, “Intratumoral plasmacytoid dendritic cells as a poor prognostic factor for hepatocellular carcinoma following curative resection,” *Cancer Immunology, Immunotherapy*, vol. 68, no. 8, pp. 1223–1233, 2019.
- [40] F. Foerster, M. Hess, A. Gerhold-Ay et al., “The immune contexture of hepatocellular carcinoma predicts clinical outcome,” *Scientific Reports*, vol. 8, no. 1, p. 5351, 2018.
- [41] J. Yan, X. L. Liu, G. Xiao et al., “Prevalence and clinical relevance of T-helper cells, Th17 and Th1, in hepatitis B virus-related hepatocellular carcinoma,” *PLoS One*, vol. 9, no. 5, article e96080, 2014.
- [42] Y. Sun, L. Wu, Y. Zhong et al., “Single-cell landscape of the ecosystem in early-relapse hepatocellular carcinoma,” *Cell*, vol. 184, no. 2, pp. 404–421.e16, 2021.