Submitted: 18 January, 2021 Accepted: 22 February, 2021 Published: 08 April, 2021 DOI:10.31083/jomh.2021.028

JOMH Journal of Men's Health

Original Research

Identification of immune-related genes for Hepatocellular Carcinoma: a study based on TCGA data

Chun-Bo Li[1], Hui-Feng Wang[1], Zheng-Kai Feng[2], Yu-Bin Fu[2], Jian Zhang[1], Jing-Yi Qin[1,*]

[1]Department of General Surgery, People's Hospital of Henan University of Chinese Medicine, 450003 Henan Province, P. R. China
[2]Department of General Surgery, The first Clinical College of Zhengzhou University, Zhengzhou City, 450003 Henan Province, P. R. China

***Correspondence:** jyqin9653@126.com (Jing-Yi Qin)

Abstract

Background and objective: Hepatocellular Carcinoma (HCC) is frequently diagnosed at the advanced stage and current treatment methods are marginally effective. The immune system is critical for the development of HCC. However, the interplay between the immune system and HCC is not well illustrated. Hence, the aim of our work was to investigate the relationship between HCC and the abnormal immune gene expression to study the potential mechanism.

Material and methods: We downloaded RNA-seq data from TCGA database, identified differentially expressed genes (DEGs) using the edgeR package, and overlapped DEGs and immune-related genes from the InnateDB website to obtain immune-related DEGs. After survival analysis, the immune-related DEGs that were significantly related to prognosis were analyzed by the STRING tool to construct PPI network. The genes in two significant PPI network modules identified by MCODE plugin were investigated by functional enrichment analysis and tumor-infiltration analysis.

Results: 68 immune-related DEGs were found to be significantly associated with prognosis in HCC, and then a PPI network was constructed. The genes in two significant PPI network modules were enriched in 31 biological processes of GO (e.g. regulation of complement activation, extracellular matrix disassembly) and 12 KEGG pathways (e.g. complement and coagulation cascades, cell cycle, proteoglycans in cancer), which were highly involved in HCC. Furthermore, we obtained 7 genes whoes expressions were significantly associated with immune infiltration levels, including E2F1, PLK1, MMP9, CDKN2A, BIRC5, CCNA2 and DCN, and 8 genes whoes copy number variations were significantly related to immune cell infiltrations, including C8A, C8B, E2F1, C6, C7, BIRC5, CCNA2 and CFB.

Conclusions: These findings contribute to understand the mechanism of HCC and provide a direction for further research on the immunotherapy of HCC.

Keywords

Hepatocellular Carcinoma; Immune cells; Immunotherapy; TCGA; GO; KEGG

1. Introduction

Liver cancer is the sixth most frequently diagnosed cancer and the fourth leading cause of cancer death globally [1]. Hepatocellular Carcinoma (HCC) is the most common type of live cancer, accounting for approximately 75%-85% of all cases [1]. The major risk factors for HCC mainly include obesity, type 2 diabetes, smoking, heavy alcohol intake, hepatitis C virus (HCV) and hepatitis B virus (HBV) [2]. For HCC patients diagnosed at the early stage, they are eligible for liver transplantation, surgical resection or local ablation [3]. How-

JOMH
Journal of Men's Health

ever, most patients are frequently diagnosed at the advanced stage and are not suitable for the locoregional therapies [4]. Although transcatheter arterial chemoembolization and the multikinase inhibitor sorafenib have been used to treat the patients with advanced HCC [4, 5], the prognosis in HCC is still poor [6]. It's essential to find other effective treatment for HCC.

A comprehensive understanding of the pathological mechanism of HCC is critical for finding effective treatment methods. The immune system protects the host from exposure to micro-organism, which is considered as a decisive factor in the progression of cancers [7, 8]. The interactions between cancer cells and immune cells can lead to tumor suppression or promotion through the immunoediting procedure that includes elimination, equilibrium and escape phases [9]. In the elimination phase, innate and adaptive cells are able to identify the neoantigens, form tumor-reactive T cells and eliminate tumor cells [10]. In the equilibrium phase, some tumor cells accompanied by epigenetic and genetic changes are resistant the attack of immune cells [11]. In the escape phase, tumor cells evolve to show poor immunogenicity and grow into significant masses, thus evading the immune system [12]. HCC, which is frequently caused by chronic inflammation, is known as a typical immunogenic cancer [13]. Hence, exploring the relationship between HCC and the abnormal immune gene expression will be a potential research direction for understanding the pathological mechanism of HCC.

The present study aimed to identify immune-related genes in HCC and gain insight into understanding the molecular mechanism of HCC. The high-throughput sequencing analysis in our study was applied to understand genomic changes in HCC and identify the differentially expressed genes (DEGs) between HCC and normal liver tissues. Then, we overlapped DEGs and immune-related genes from the InnateDB website to obtain immune-related DEGs. After survival analysis, the immune-related DEGs that were significantly related to prognosis were analyzed by the STRING tool to construct protein-protein interaction (PPI) network. After functional enrichment analysis of the genes in the most significant modules from PPI network, the correlation of hub genes with immune infiltration levels was validated by the TIMER website.

2. Methods

2.1 Data resource

RNA-sequencing (RNA-seq) data in count format were downloaded from The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/) database, including 369 HCC tissues and 50 normal adjacent tissues. The corresponding clinical information was also obtained from TCGA database. The filtering criteria of the above data were as follows: (1) the count format data that were zero in more than half of the number of clinical samples were removed; (2) the samples without clinical information were removed. Those downloaded data are public and available and thus there is no need of ethical approvals.

2.2 Differentially expressed genes analysis

RNA-seq data in count format were normalized and analyzed using edgeR package (Version 3.4; http://www.bioconductor.org/packages/release/bioc/html/edgeR.html) in R language to obtain differentially expressed genes (DEGs) between HCC tissues and normal adjacent tissues. To identify DEGs, Adjust. P. value < 0.01 and $|\log_2 \text{FC (fold change)}| > 1$ were designated as the cut-off criterion. Heatmaps and volcano plots were generated using heatmap package (Version 1.0.12) in R language.

2.3 Immune-related DEGs

InnateDB database (https://www.innatedb.com/) is a publicly available database that includes the genes, proteins, experimentally-verified interactions and signaling pathways associated with the innate immune response of humans and other species. Hence, we downloaded the immune-related genes from the InnateDB database. Next, we overlapped the immune-related genes with DEGs using the Venn diagram tool (http://bioinfogp.cnb.csic.es/tools/venny/) to obtain the immune-related DEGs.

2.4 Prognostic survival analysis

Based on the median value of each of the above immune-related DEGs expression, HCC patients were divided into high-expression and low-expression groups. Survival analysis of those genes between two groups was performed by the "survival" package in R based on the Kaplan-Meier survival analysis method. The log-rank test was used to assess statistical difference and P. value < 0.05 was considered statistically significant.

2.5 Protein-protein interaction network construction and module analysis

Protein-protein interaction (PPI) network of the immune-related DEGs with a significant prognosis was obtained from STRING (Search Tool for the Retrieval of Interacting Genes, version 10.0, https://string-db.org/) and visualized using Cytoscape software (version 3.2.0; http://www.cytoscape.org/). An interaction score of PPI pairs ≥ 0.4 was considered statistically significant. Next, the MCODE (Molecular Complex Detection; Version 1.4.2; http://apps.cytoscape.org/apps/MCODE) plugin in Cytoscape software was used to identify the significant module of the PPI network according to the graph-theoretic clustering algorithm. MCODE scores ≥ 5 was designated as a selection criterion.

2.6 Functional enrichment analysis

For the immune-related DEGs with a significant prognosis and the genes in the significant modules of PPI network, the functional analysis of Gene Ontology (GO) and

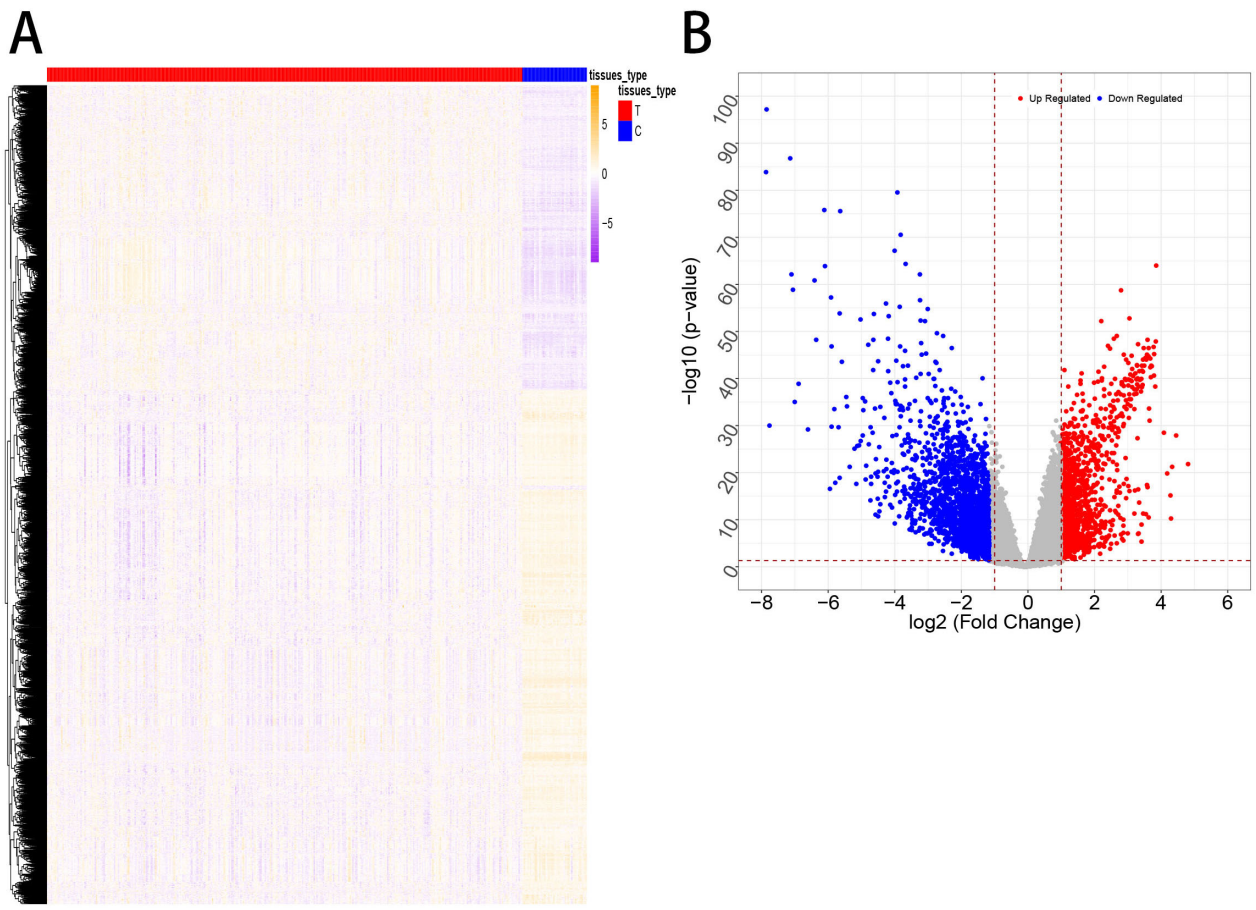

FIG. 1. Differential expression analysis. The expression of differentially expressed genes (DEGs) between HCC tissues and normal adjacent tissues was shown in (A) a heatmap and (B) a Volcano plot.

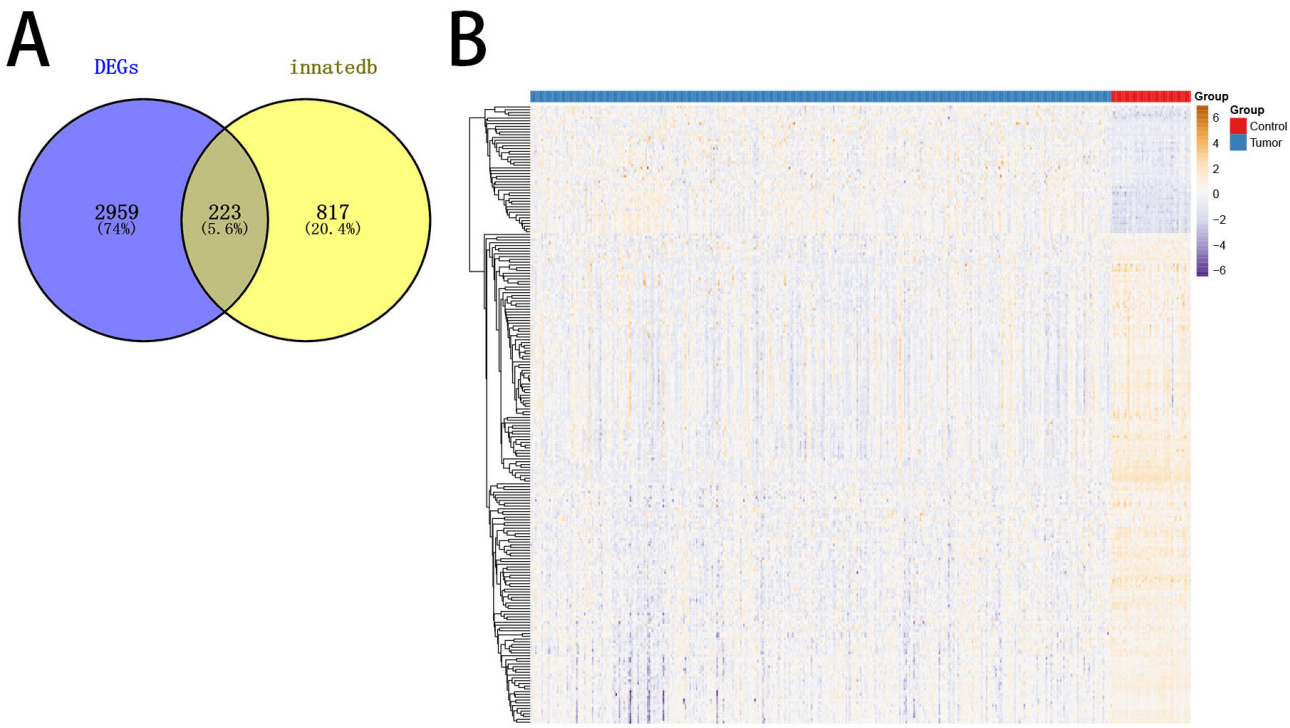

FIG. 2. Immune-related DEGs. (A) Venn diagram of the overlapping of immune-related genes from the InnateDB database with DEGs. (B) The expression of the overlapping genes was shown in a heatmap.

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were performed by the DAVID (Database for Annotation, Visualization and Integrated Discovery) (https://david.ncifcrf.gov/) website that contains a comprehensive set of functional annotation tools. *P.* value < 0.05 and count ≥ 2 were considered as significant enrichment. The top 10 items in biological process (BP) category of GO and KEGG pathways were shown in plot bubble maps. The genes in the significant modules of PPI network were enriched in KEGG pathways, which were selected as hub genes for the following analysis.

2.7 Hub gene expression

The expression of the above hub genes was visualized by the GEPIA (Gene Expression Profiling Interactive Analysis) website (http://gepia.cancer-pku.cn/) and shown as box plots. *P.* value < 0.05 was designated as statistically significant.

2.8 Immunohistochemistry on HCC Tissue

The HPA (human protein atlas; Version 19.3; https://www.proteinatlas.org/) website provides the maps of human proteins in cells, tissues and organs through integration of various omics technologies such as mass spectrometry-based proteomics. Hence, we used the HPA website to analyze the protein expression differences of hub genes between HCC tissues and normal live tissues and downloaded the immunohistochemistry images from the Tissue Atlas and Pathology Atlas in HPA.

2.9 Relationships of hub genes with immune cell infiltration

TIMER (Tumor Immune Estimation Resource; https://cistrome.shinyapps.io/timer/) is a publicly available website for comprehensive analysis of tumor-infiltrating immune cells. TIMER algorithm has been used to assess the abundance of six immune infiltrates, including dendritic cells, neutrophils, macrophages, B cells, CD8+ T cells and CD4+ T cells. Using the Survival module of TIMER, a Cox regression model was constructed based on the abundance of six immune cells in HCC. Then, we used the Gene module in the TIMER website to evaluate the correlation of the hub gene expression with immune cell infiltration level in HCC. Next, SCNA module in the TIMER website was used to compare immune infiltration levels in HCC with different somatic copy number alterations for the hub genes.

3. Results

3.1 Differential expression analysis

RNA-seq data from TCGA database were normalized and analyzed by the edgeR package to obtain DEGs between HCC tissues and normal adjacent liver tissues (cut-off criterion: Adjust. *P.* value < 0.01 and $|\log_2 \text{FC (fold change)}| > 1$). A total of 3182 DEGs were obtained, including 1197 up-regulated DEGs and 1985 down-regulated DEGs, and were listed in a heatmap and a volcano map (Fig. 1).

3.2 Immune-related DEGs

A total of 1040 immune-related genes downloaded from the InnateDB database were overlapped with 3182 DEGs using the Venn diagram tool to obtain 223 immune-related DEGs (Fig. 2A). These overlapping genes were shown in a heatmap (Fig. 2B).

TABLE 1. The information of genes in significant modules

module-a			module-b		
Nodes	Regulation	Degree	Nodes	Regulation	Degree
AHSG	DOWN	15	MMP9	UP	15
C8B	DOWN	14	IGF1	DOWN	12
HRG	DOWN	13	CDKN2A	UP	10
HP	DOWN	13	E2F1	UP	8
CFP	DOWN	12	CCNA2	UP	7
APOA1	DOWN	12	MMP7	DOWN	6
C6	DOWN	11	DCN	DOWN	6
C8A	DOWN	11	PLK1	UP	5
CFB	DOWN	11	BIRC5	UP	4
APOH	DOWN	10			
C5	DOWN	9			
C7	DOWN	9			
APCS	DOWN	8			

3.3 Prognostic survival analysis

To investigate the correlation between the above immune-related DEGs expression and the prognosis of HCC patient, Kaplan-Meier survival analysis with log-rank tests was used to analyze the overall survival (OS). After analysis, 68 immune-related DEGs had a significant effect on prognosis, including HRG, SOCS2, PLK1, etc. (Supplementary Table 1). The top 8 of immune-related DEGs associated with prognosis, including HRG, SOCS2, PLK1, PPARGC1A, BIRC5, DUSP10, WDR62 and SPP1, were listed in Fig. 3. The overall survival time was higher in the group with high expression of HRG, SOCS2, PPARGC1A or DUSP10 than the group with low expression of these genes. In other words, the low expression of HRG, SOCS2, PPARGC1A and DUSP10 was correlated with poor OS in HCC patients. Nevertheless, the high expression of PLK1, BIRC5, WDR62 and SPP1 was correlated with poor OS of HCC patients. The survival analysis of the other 60 immune-related DEGs was shown in Supplementary Fig. 1.

3.4 PPI network construction and module analysis

The STRING tool was used to establish the PPI network among the above obtained 68 DEGs. As shown in Fig. 4A, the PPI network included 57 nodes and 191 edges. Next, two significant PPI network modules identified by MCODE plugin contained module-a and module-b. Module-a (MCODE score = 7.667) had 13 nodes and 46 edges (Fig. 4B); Module-b (MCODE scroe = 5.5) had 9 nodes and 22 edges (Fig. 4B).

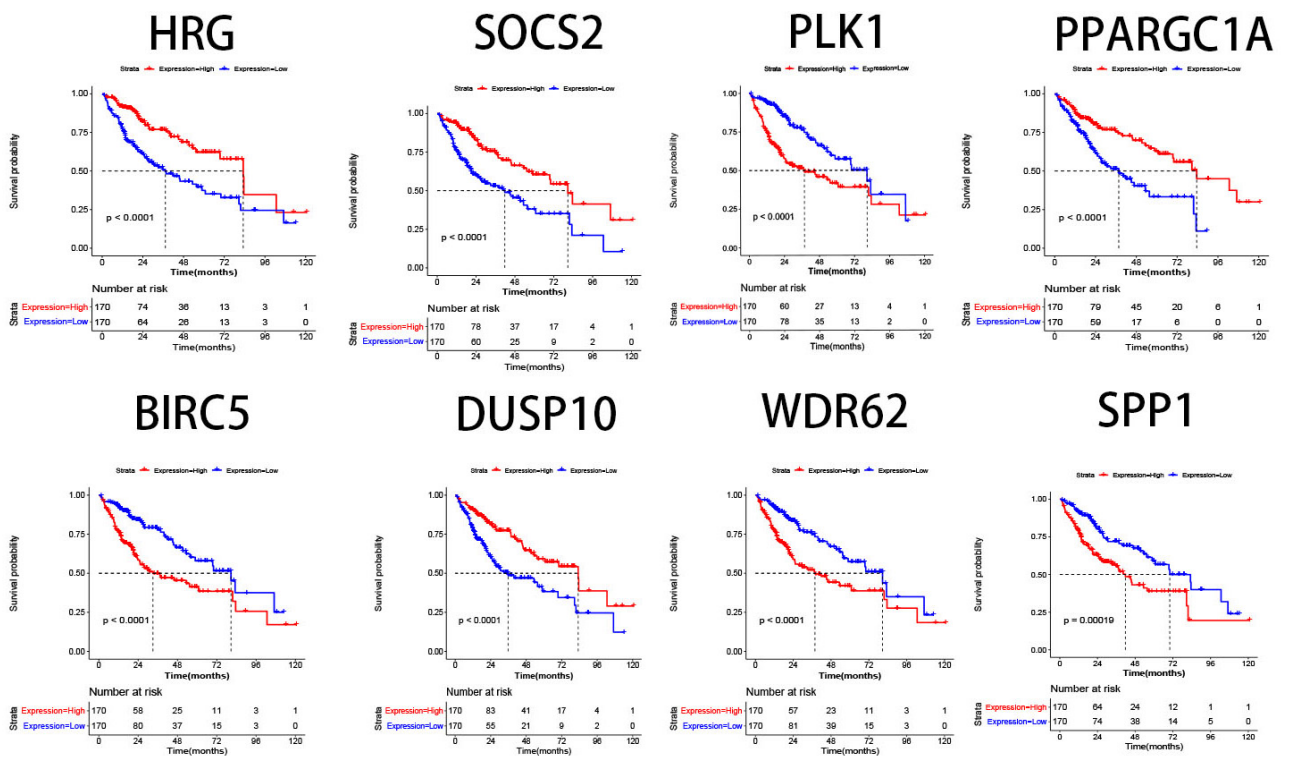

FIG. 3. Prognostic survival analysis of the immune-related DEGs. The overall survival (OS) curves of the top 8 immune-related DEGs with significant prognosis, including HRG, SOCS2, PLK1, PPARGC1A, BIRC5, DUSP10, WDR62 and SPP1, were plotted by Kaplan-Meier survival analysis.

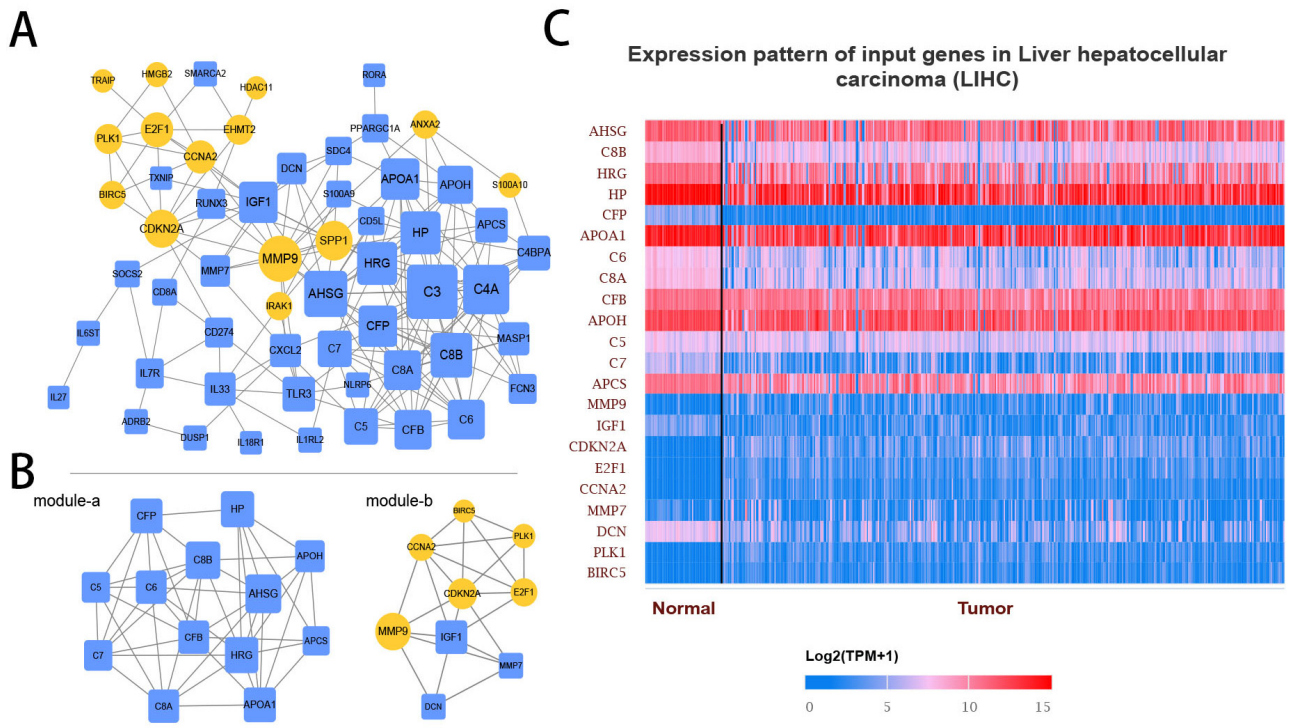

FIG. 4. Protein-protein network (PPI) network construction and module analysis. (A) PPI network of 68 immune-related DEGs with significant prognosis was constructed by the STRING tool. (B) Two significant modules of PPI network were identified by MCODE plugin in Cytoscape software. (C) The expression of a total of 22 genes in two significant modules was listed in a heatmap. The blue square node represents the down-regulated genes and the yellow circle node represents the up-regulated genes. The size of node is positively correlated with the degree value.

Two significant modules had 22 genes, and their information and expression were listed in Table 1 and Fig. 4C, respectively.

JOMH
Journal of Men's Health

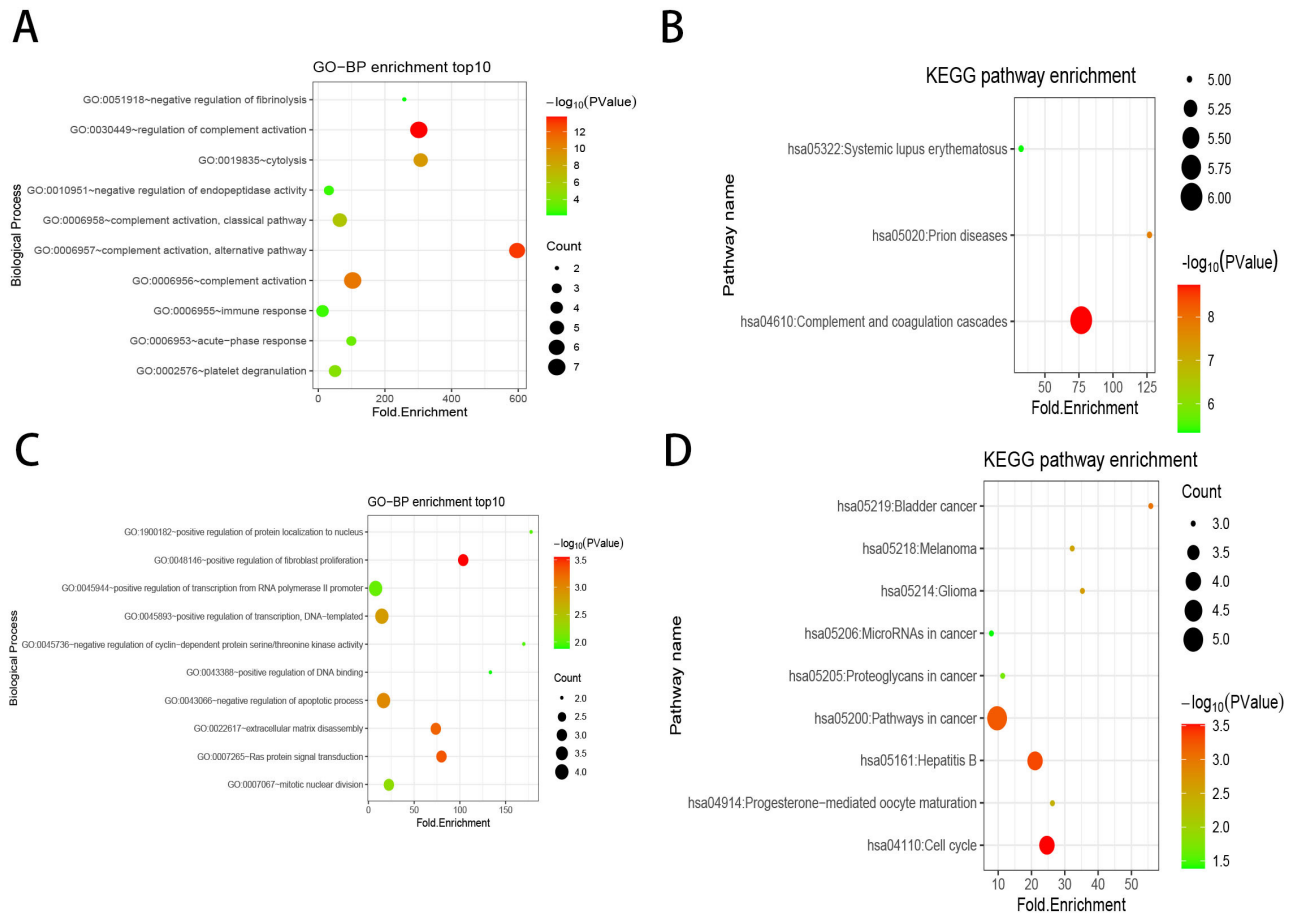

FIG. 5. Functional enrichment analysis. (A, B) Enriched Gene Ontology (GO)-biological process (BP) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways of genes in module-a. (C, D) Enriched GO-BP and KEGG pathways of genes in module-b.

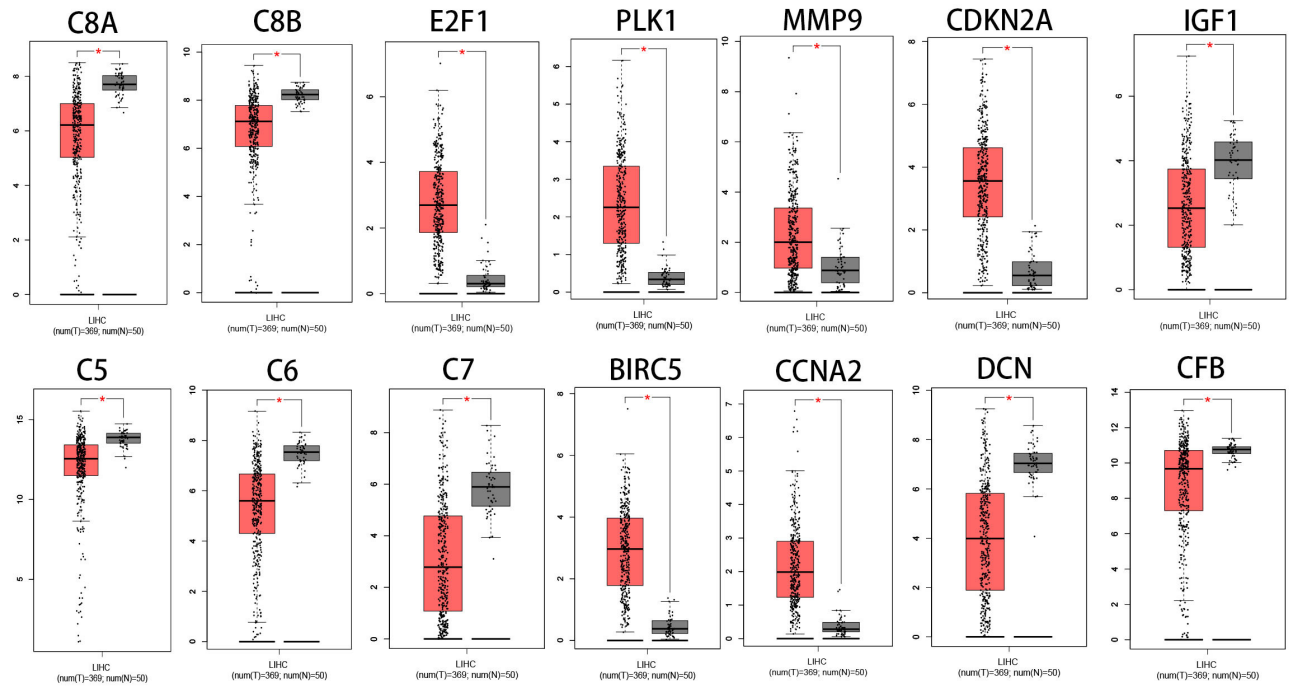

FIG. 6. Hub gene expression. A total of 14 genes from module-a and module-b enriched in KEGG pathways, including C8A, C8B, E2F1, PLK1, MMP9, CDKN2A, IGF1, C5, C6, C7, BIRC5, CCNA2, DCN and CFB, which were designated as hub genes. The expression of the hub genes was validated by the GEPIA websites and presented on box plots.

TABLE 2. Enriched GO-biological process and KEGG pathways of differentially expressed genes in module-a

Category	Term	count	*P*-value	Genes
BP	GO:0030449~regulation of complement activation	7	1.75E-14	C8A, CFP, C8B, C7, CFB, C6, C5
BP	GO:0006957~complement activation, alternative pathway	6	9.14E-14	C8A, CFP, C8B, C7, CFB, C5
BP	GO:0006956~complement activation	7	1.46E-11	C8A, CFP, C8B, C7, CFB, C6, C5
BP	GO:0019835~cytolysis	5	8.89E-10	C8A, C8B, C7, C6, C5
BP	GO:0006958~complement activation, classical pathway	5	5.43E-07	C8A, C8B, C7, C6, C5
BP	GO:0002576~platelet degranulation	4	4.74E-05	APOA1, APOH, HRG, AHSG
BP	GO:0006953~acute-phase response	3	3.42E-04	APCS, HP, AHSG
BP	GO:0006955~immune response	4	0.00291	C8A, CFP, C8B, C7
BP	GO:0010951~negative regulation of endopeptidase activity	3	0.00324	C5, HRG, AHSG
BP	GO:0051918~negative regulation of fibrinolysis	2	0.00713	APOH, HRG
BP	GO:0051006~positive regulation of lipoprotein lipase activity	2	0.00713	APOA1, APOH
BP	GO:0016525~negative regulation of angiogenesis	2	0.04343	APOH, HRG
KEGG	hsa04610: Complement and coagulation cascades	6	1.81E-09	C8A, C8B, C7, CFB, C6, C5
KEGG	hsa05020: Prion diseases	5	1.72E-08	C8A, C8B, C7, C6, C5
KEGG	hsa05322: Systemic lupus erythematosus	5	4.61E-06	C8A, C8B, C7, C6, C5

BP, biological process; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

TABLE 3. Enriched GO-biological process and KEGG pathways of differentially expressed genes in module-b

Category	Term	count	*P*-value	Genes
BP	GO:0048146~positive regulation of fibroblast proliferation	3	2.81E-04	E2F1, IGF1, CCNA2
BP	GO:0007265~Ras protein signal transduction	3	4.72E-04	CDKN2A, IGF1, CCNA2
BP	GO:0022617~extracellular matrix disassembly	3	5.56E-04	MMP9, MMP7, DCN
BP	GO:0043066~negative regulation of apoptotic process	4	1.00E-03	PLK1, MMP9, IGF1, BIRC5
BP	GO:0045893~positive regulation of transcription, DNA-templated	4	0.001431	E2F1, CDKN2A, IGF1, CCNA2
BP	GO:0007067~mitotic nuclear division	3	0.005736	PLK1, BIRC5, CCNA2
BP	GO:0045944~positive regulation of transcription from RNA polymerase II promoter	4	0.008917	E2F1, CDKN2A, IGF1, DCN
BP	GO:1900182~positive regulation of protein localization to nucleus	2	0.009963	CDKN2A, PLK1
BP	GO:0045736~negative regulation of cyclin-dependent protein serine/threonine kinase activity	2	0.010435	CDKN2A, PLK1
BP	GO:0043388~positive regulation of DNA binding	2	0.013265	MMP9, IGF1
BP	GO:0031648~protein destabilization	2	0.016557	CDKN2A, PLK1
BP	GO:0007346~regulation of mitotic cell cycle	2	0.018902	PLK1, BIRC5
BP	GO:0045892~negative regulation of transcription, DNA-templated	3	0.021913	E2F1, CDKN2A, BIRC5
BP	GO:0000910~cytokinesis	2	0.022645	PLK1, BIRC5
BP	GO:0030574~collagen catabolic process	2	0.030093	MMP9, MMP7
BP	GO:0014068~positive regulation of phosphatidylinositol 3-kinase signaling	2	0.030557	IGF1, DCN
BP	GO:0071456~cellular response to hypoxia	2	0.044840	E2F1, CCNA2
BP	GO:0006469~negative regulation of protein kinase activity	2	0.046213	CDKN2A, DCN
BP	GO:0007062~sister chromatid cohesion	2	0.048040	PLK1, BIRC5
KEGG	hsa04110: Cell cycle	4	3.00E-04	E2F1, CDKN2A, PLK1, CCNA2
KEGG	hsa05161: Hepatitis B	4	4.75E-04	E2F1, MMP9, BIRC5, CCNA2
KEGG	hsa05200: Pathways in cancer	5	6.11E-04	E2F1, CDKN2A, MMP9, IGF1, BIRC5
KEGG	hsa05219: Bladder cancer	3	9.49E-04	E2F1, CDKN2A, MMP9
KEGG	hsa05214: Glioma	3	0.002373	E2F1, CDKN2A, IGF1
KEGG	hsa05218: Melanoma	3	0.002825	E2F1, CDKN2A, IGF1
KEGG	hsa04914: Progesterone-mediated oocyte maturation	3	0.004214	PLK1, IGF1, CCNA2
KEGG	hsa05205: Proteoglycans in cancer	3	0.020982	MMP9, IGF1, DCN
KEGG	hsa05206: MicroRNAs in cancer	3	0.040858	E2F1, CDKN2A, MMP9

BP, biological process; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

3.5 Functional enrichment analysis

To investigate the potential functions of genes in the significant modules, the DAVID tool was used to analyze the BP category of GO and KEGG pathways. In module-a, genes were significantly enriched in 12 of BP terms such as regulation of complement activation and complement activation, alternative pathway (Table 2 and Fig. 5A), and a total of 3 KEGG pathways were found to be significant, including Complement and coagulation cascades, Prion diseases and Systemic lupus erythematosus (Table 2 and Fig. 5B). In module-b, a total of 19 significant BP (e.g. positive regulation of fibroblast proliferation, Ras protein signal transduction, extracellular matrix disassembly) and 9 significant KEGG

JOMH
Journal of Men's Health

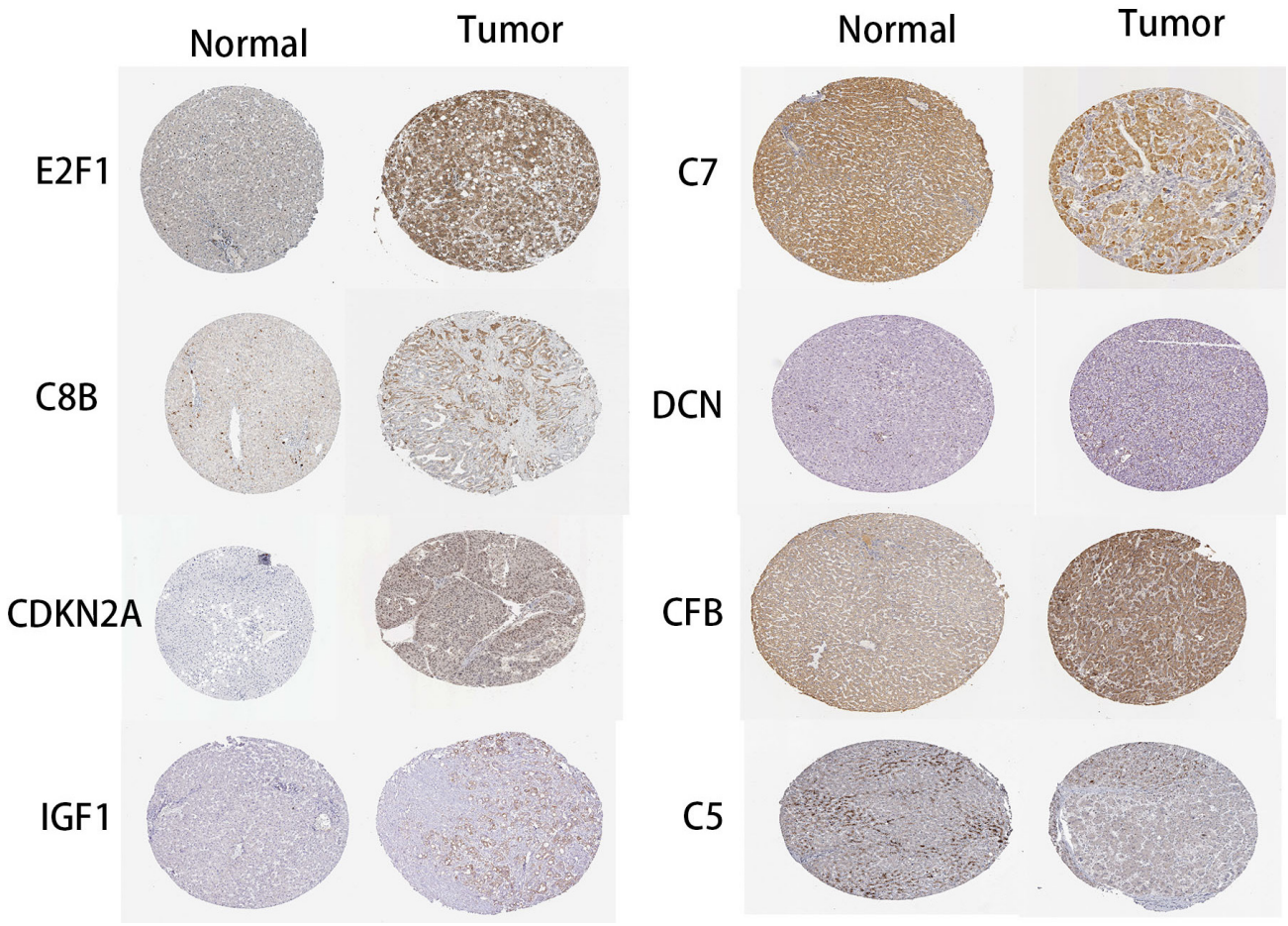

FIG. 7. Immunohistochemistry on HCC Tissue. The Representative protein expression of the eight hub genes (E2F1, C8B, CDKN2A, IGF1, C7, DCN, CFB and C5). The data are from the Human Protein Atlas (http://www.proteinatlas.org).

pathways (e.g. Cell cycle, Hepatitis B, Pathways in cancer, Proteoglycans in cancer) were found (Table 3 and Fig. 5C-5D). In module-a and module-b, a total of 14 genes were enriched in KEGG pathways, including C8A, C8B, E2F1, PLK1, MMP9, CDKN2A, IGF1, C5, C6, C7, BIRC5, CCNA2, DCN and CFB, which were selected as hub genes for the following analysis.

3.6 Hub gene expression

The mRNA expression levels of the hub genes were validated by the GEPIA websites. The results confirmed that the expression levels of E2F1, PLK1, MMP9, CDKN2A, BIRC5 and CCNA2 were significantly increased in HCC tissues compared with normal liver tissues (Fig. 6). Meanwhile, C8A, C8B, IGF1, C5, C6, C7, DCN and CFB showed lower expression levels in HCC tissues compared with normal liver tissues (Fig. 6).

3.7 Immunohistochemistry on HCC Tissue

The protein expression of hub genes in HCC was validated by the HPA database. According to the existing data in the HPA database, we obtained the immunohistochemistry results of 8 genes, as shown in Fig. 7. The increased protein expression of E2F1 and CDKN2A as well as the decreased protein expression of C7 and C5 in HCC tissues were consistent with their mRNA expression trends. The increase in protein expression of C8B, IGF1, DCN and CFB in HCC tissues was contrary to the trends of their mRNA expression, which might be affected by post-transcriptional processes.

3.8 Relationships of hub genes with immune cell infiltration

We used the TIMER web tool to further investigate the relationship between the prognosis and the infiltration of six immune cell and the expression of 14 hub genes. Cox analysis results revealed that lower infiltration levels of B cells (HR = 0, $P = 0.001$), CD8+ T cells (HR = 0.002, $P = 0.015$) and higher infiltration levels of dendritic cells (HR = 979.013, $P = 0.001$) were risk factors for prognosis in HCC patients (Table 4). Besides, lower expression levels of IGF1 (HR = 0.816, $P = 0.012$) and C5 (HR = 0.782, $P = 0.029$) were risk factors for prognosis in HCC patients (Table 4). Meanwhile, the TIMER web was also used to perform Kaplan-Meier survival analysis on six immune cells and 14 hub genes. The results indicated that the infiltration levels of six immune cells had no significant correlation with OS (Supplementary Fig. 2). There was no correlation between OS and the expression of C8A, IGF1, DCN and CFB. The results further confirmed that the expression levels of C8B, E2F1, PLK1, MMP9, CDKN2A, C5,

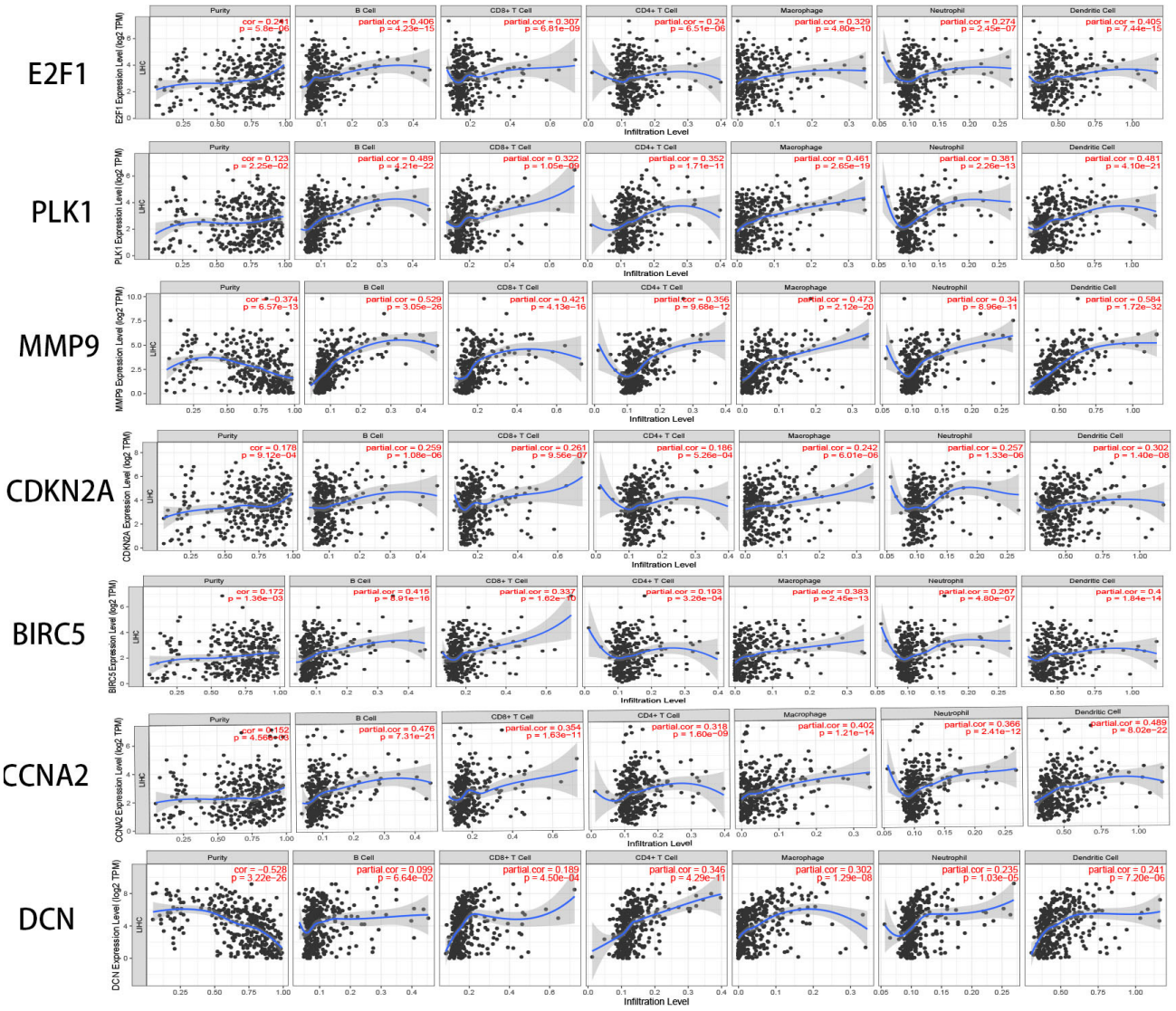

FIG. 8. The correlation of hub gene expression with the infiltration levels of immune cells. After the analysis by the TIMER web tool, there were 7 hub genes that had a significant correlation with immune cells, including E2F1, PLK1, MMP9, CDKN2A, BIRC5, CCNA2 and DCN.

C6, C7, BIRC5 and CCNA2 had the significant correlation with OS (Supplementary Fig. 2). The high expression of E2F1, PLK1, MMP9, CDKN2A, BIRC5 and CCNA2 was correlated with poor OS. The low expression of C8B, C5, C6 and C7 was correlated with poor OS in HCC patients.

To study the relationship of the above 14 hub genes with immune cell infiltration, we firstly used the TIMER web tool to analyze the correlation between hub gene expression and the infiltration levels of immune cells in HCC. The top 7 hub genes with significant correlation were shown in Fig. 8. The results indicated that E2F1 expression was highly correlated with B cells (r = 0.406, *P* = 4.23e-15) and dendritic cells (r = 0.405, *P* = 7.44e-15) infiltration, and its expression was weakly correlated with CD8+ T cells (r = 0.307, *P* = 6.18e-09), CD4+ T cells (r = 0.24, *P* = 6.51e-06), macrophages (r = 0.329, *P* = 4.8e-10) and neutrophils (r = 0.274, *P* = 2.45e-07) infiltration. PLK1 levels were strongly correlated with B cells (r = 0.489, *P* = 4.21e-22), macrophages (r = 0.461, *P* = 2.65e-19) and dendritic cells (r = 0.481, *P* = 4.10e-21) infiltration, and its levels were weakly correlated with CD8+ T cells (r = 0.322, *P* = 1.05e-09), CD4+ T cells (r = 0.352, *P* = 1.71e-11), and neutrophils (r = 0.381, *P* = 2.26e-13) infiltration. MMP9 expression was strongly correlated with B cells (r = 0.529, *P* = 3.05e-26), CD8+ T cells (r = 0.421, *P* = 4.13e-16), macrophages (r = 0.473, *P* = 2.12e-20) and dendritic cells (r = 0.584, *P* = 1.72e-32) infiltration, and weakly correlated with CD4+ T cells (r = 0.356, *P* = 9.68e-12), and neutrophils (r = 0.34, *P* = 8.96e-11) infiltration. CDKN2A was weakly correlated with B cells (r = 0.259, *P* = 1.08e-06), CD8+ T cells (r = 0.261, *P* = 9.56e-07), CD4+ T cells (r = 0.186, *P* = 5.26e-04), macrophages (r = 0.242, *P* = 6.01e-06), neutrophils (r = 0.257, *P* = 1.33e-06) or dendritic cells (r = 0.302, *P* = 1.4e-08) infiltration. BIRC5 expression was strongly associated with B cells (r = 0.415, *P* = 8.91e-16) and dendritic cells (r = 0.4, *P* = 1.84e-14) infiltration, while its expression was weakly associated with CD8+ T cells (r = 0.337, *P* = 1.62e-10), macrophages (r = 0.383, *P* = 2.45e-13) and neutrophils (r = 0.267, *P* = 4.8e-07) infiltration. CCNA2 levels were

JOMH
Journal of Men's Health

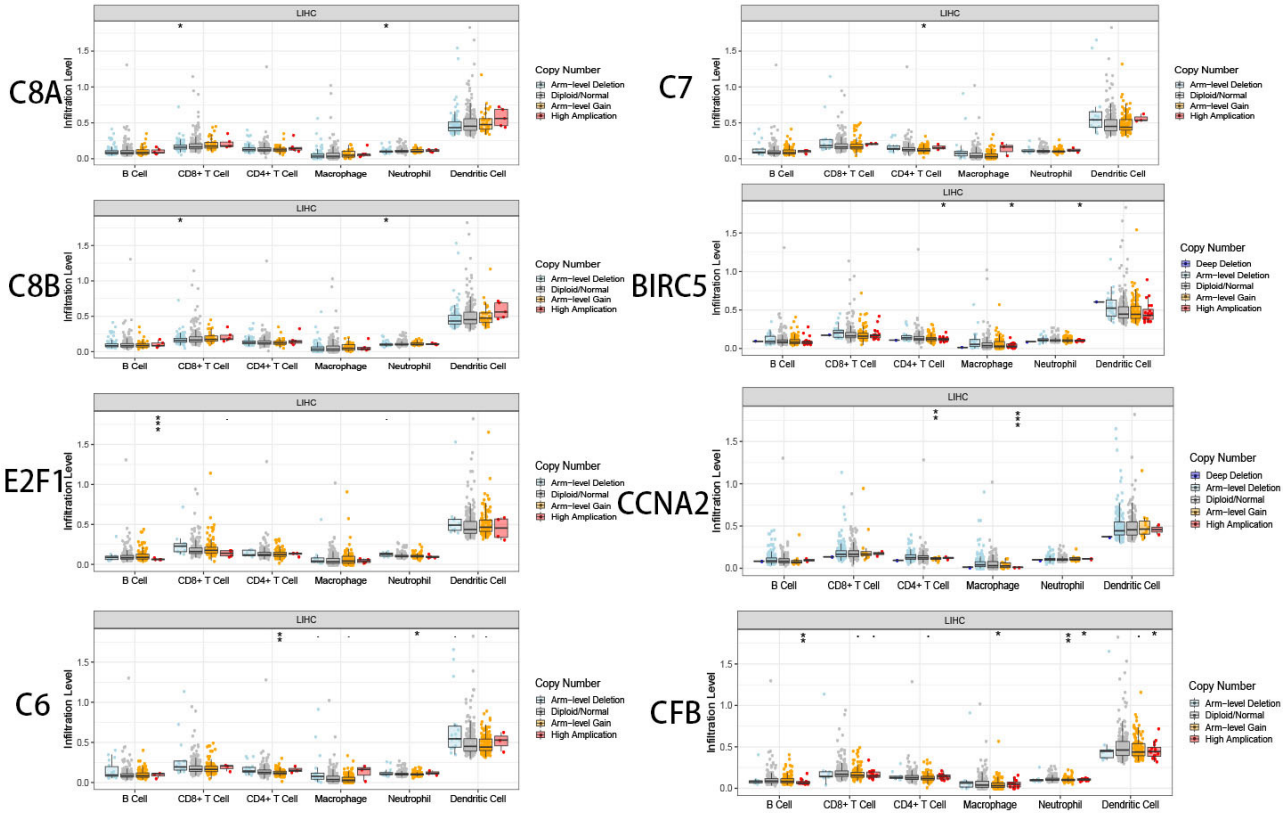

FIG. 9. The relationship between somatic copy number alterations of hub genes and HCC infiltration levels. After the analysis by the TIMER web tool, somatic copy number alterations of 8 hub genes were observed to associate with the infiltration levels of six immune cells. The 8 hub genes include C8A, C8B, E2F1, C6, C7, BIRC5, CCNA2 and CFB.

TABLE 4. Multivariate Cox regression analyses of immune cells and hub genes with overall survival

Variables	coef	HR	95% CI_l	95% CI_u	*P*-value	sig
Purity	0.27	1.31	0.375	4.57	0.672	
B_cell	-12.125	0	0	0.009	0.001	**
CD8_Tcell	-6.423	0.002	0	0.285	0.015	*
CD4_Tcell	-4.109	0.016	0	15.548	0.24	
Macrophage	4.228	68.575	0.512	9177.904	0.091	
Neutrophil	2.99	19.89	0	2646283.1	0.619	
Dendritic	6.887	979.013	17.245	55579.414	0.001	**
C8A	0.092	1.097	0.904	1.33	0.35	
E2F1	-0.146	0.864	0.64	1.167	0.341	
PLK1	0.274	1.315	0.914	1.893	0.14	
MMP9	-0.008	0.992	0.855	1.15	0.915	
C8B	-0.05	0.951	0.757	1.194	0.665	
CDKN2A	-0.045	0.956	0.824	1.108	0.549	
IGF1	-0.204	0.816	0.695	0.957	0.012	*
C7	-0.046	0.955	0.761	1.2	0.693	
BIRC5	0.177	1.193	0.884	1.61	0.248	
CCNA2	0.098	1.103	0.838	1.452	0.485	
DCN	-0.049	0.952	0.765	1.187	0.664	
CFB	0.135	1.144	0.919	1.425	0.228	
C6	0.005	1.005	0.864	1.17	0.947	
C5	-0.246	0.782	0.627	0.975	0.029	*

strongly correlated with B cells (r = 0.476, P = 7.31e-21), macrophages (r = 0.402, P = 1.21e-14) and dendritic cells (r = 0.489, P = 8.02e-22) infiltration, whereas its levels were weakly correlated with CD8+ T cells (r = 0.354, P = 1.63e-11), CD4+ T cells (r = 0.318, P = 1.6e-09), and neutrophils (r = 0.366, P = 2.41e-12) infiltration. There was a weak correlation between DCN expression and CD4+ T cells (r = 0.346, P = 4.29e-11), macrophages (r = 0.302, P = 1.29e-08), neutrophils (r = 0.235, P = 1.03e-05) and dendritic cells (r = 0.241, P = 7.2e-06) infiltration.

Furthermore, copy number variation analysis indicated that somatic copy number alterations of 8 hub genes were associated with HCC infiltration levels. As shown in Fig. 9, arm-level deletion of C8A was associated with CD8+ T cells ($P < 0.05$) and neutrophils ($P < 0.05$) infiltration. Arm-level deletion of C8B was involved in CD8+ T cells ($P < 0.05$) and neutrophils ($P < 0.05$) infiltration. High amplification of E2F1 was associated with B cells ($P < 0.001$) infiltration. Arm-level gain of C6 was related with CD4+ T cells ($P < 0.01$) and neutrophils ($P < 0.05$) infiltration. Arm-level gain of C7 was associated with CD4+ T cells ($P < 0.05$) infiltration. High amplification of BIRC5 was involved in CD4+ T cells ($P < 0.05$), macrophages ($P < 0.05$), neutrophils ($P < 0.05$) infiltration. Arm-level gain of CCNA2 was associated with CD4+ T cells ($P < 0.01$) infiltration, and its high amplification was involved in macrophages ($P < 0.001$) infiltration. High amplification of CFB was related with B cells ($P < 0.01$), neutrophils ($P < 0.05$) and dendritic cells ($P < 0.05$) infiltration, and its arm-level gain was associated

with macrophages ($P < 0.05$) and neutrophils ($P < 0.01$) infiltration.

4. Discussion

Most patients with HCC are diagnosed at the advanced stage and current treatment methods are marginally effective [4]. The overall survival of HCC patients with advanced stage is less than one year [13]. Exploring the mechanism of HCC is important for finding effective and novel treatments. Accumulating evidence suggests that the immune system is critical for regulating the progression of cancers [12, 14]. Thus, investigating the interplay between the immune system and HCC is essential for understanding the mechanism of HCC. Our work studied this issue from the perspective of the relationship between HCC and the abnormal immune gene expression.

In the present study, we firstly downloaded RNA-seq data from TCGA database, identified DEGs using the edgeR package, and overlapped DEGs with immune-related genes from the InnateDB website to obtain immune-related DEGs. The overlapping genes were analyzed by survival analysis, and the immune-related DEGs significantly related to prognosis were analyzed by the STRING tool to construct PPI network. Meanwhile, two significant modules in the PPI network were identified by MCODE plugin. There were 22 genes in two significant modules, and these genes were enriched by GO and KEGG pathways analysis. GO enrichment analysis results showed that the genes (CFP, HP, C5, C6, C8B, AHSG, APOH, C7, CFB, HRG, APCS, C8A and APOA1) in the module-a were mainly enriched in 12 biological processes such as regulation of complement activation and complement activation, alternative pathway. The complement system in plasma and cells is a complex working organization of proteins, receptors and regulators, which cooperates to offer innate immune defense against infection [15]. However, dysregulation of complement activation causes local and/or system inflammation, leading to tissue damages [16]. As HCC is an inflammation-related cancer [17], the genes in the module-a could be associated with the progression of HCC. The genes (BIRC5, CCNA2, PLK1, MMP9, CDKN2A, E2F1, IGF1, DCN and MMP7) in the module-b were mainly enriched in 19 biological processes such as positive regulation of fibroblast proliferation, Ras protein signal transduction and extracellular matrix disassembly. Activated fibroblasts are critical for supporting the growth, motility and invasion of cancers, and the interaction of activated fibroblasts with cancers is usually affected by inflammation [18]. Thus, the genes in the module-b could also be related to the development with HCC. Furthermore, after KEGG pathways analysis, a total of 3 KEGG pathways in the module-a were found to be significant, including complement and coagulation cascades, Prion diseases and Systemic lupus erythematosus, and 9 significant KEGG pathways in the module-b were identified, including cell cycle, Hepatitis B pathways in cancer, bladder cancer, glioma, melanoma, progesterone-mediated oocyte maturation, proteoglycans in cancer, and microRNAs in cancer.

Additionally, a total of 14 genes from two significant modules enriched in KEGG pathways, including C8A, C8B, E2F1, PLK1, MMP9, CDKN2A, IGF1, C5, C6, C7, BIRC5, CCNA2, DCN and CFB were selected as hub genes for the following analysis. The GEPIA website was used to confirm the increase in mRNA expression of E2F1, PLK1, MMP9, CDKN2A, BIRC5 and CCNA2 and the decrease in mRNA expression of C8A, C8B, IGF1, C5, C6, C7, DCN and CFB in HCC tissues compared with the normal live tissues. The increased protein expression of E2F1 and CDKN2A as well as the decreased protein expression of C7 and C5 in HCC tissues were consistent with their mRNA expression trends, validated by the HPA online tool. Next, the TIMER web tool was used to investigate the relationship between prognosis and six immune cell infiltration levels and 14 hub genes expression. The results further indicated that the aberrantly high expressions of E2F1, PLK1, MMP9, CDKN2A, BIRC5 and CCNA2 were correlated with poor OS, and the aberrantly low expressions of C8B, C5, C6 and C7 were correlated with poor OS in HCC patients. In the module-a, C8A, C8B, C5, C6, and C7 were enriched in the complement and coagulation cascades KEGG pathway. Under normal conditions, activation of the complement and coagulation cascades can protect against invading pathogens and moderate tissue damage, while the over-amplification of these cascades destroys the imbalance and causes tissue damage [19]. Thus, the relationship between the complement and coagulation cascades pathway and the low expression of C8A, C8B, C5, C6 and C7 is a potential research direction for studying the mechanism of HCC. In the module-b, E2F1, PLK1, CDKN2A and CCNA2 were enriched in the cell cycle KEGG pathway. In addition, MMP9, IGF1 and DCN were enriched in the proteoglycans in cancer pathway. Previous reports indicate that the mutated genes directly mediate cell cycle and cause uncontrolled cell proliferation in tumors [20]. Proteoglycans interact with plenty of regulatory molecules and signaling pathways to influence various cellular processes, thereby affecting the development of liver cancer [21]. E2F1 can act as a proliferative and/or apoptotic factor in HCC [22]. Overexpression of PLK1 (Polo-like kinase 1) has been found in many cancers such as HCC and its dysfunction promotes the progression of cancers [23, 24]. MMP9 (Matrix metallopeptidase 9) is involvement with tumor aggressiveness [25]. CDKN2A functions as a tumor suppressor in various cancers [26], but its role in HCC has not been fully studied. BIRC5 also named survivin is a key target for cancer treatment [27]. CCNA2, as an important prognostic biomarker, is frequently observed to be overexpressed in different cancers [28]. The aberrantly high expression of E2F1, PLK1, CDKN2A and CCNA2 may regulate cell cycle to control cell proliferation in HCC. The abnormal expression of MMP9, IGF1 and DCN can be associated with the proteoglycans in cancer pathway to affect the progression of HCC. Therefore, it is necessary to further investigate the interaction between the above key pathways and genes in HCC using many basic biological experiments in the future.

Our study indicated that the infiltration levels of B cells, CD8+ T cells and dendritic cells as well as the expression of IGF1 and C5 were risk factors for prognosis in HCC patients. It inspired us to study the relationship of hub genes with immune response in HCC. Firstly, we assessed the correlation of hub genes with immune cell infiltration levels and found that 7 hub genes were significantly associated with immune cell infiltration levels, including E2F1, PLK1, MMP9, CDKN2A, BIRC5, CCNA2 and DCN. The infiltration levels of B cells and dendritic cells were significantly correlated with the expression of E2F1, PLK1, MMP9, BIRC5 and CCNA2. The infiltration levels of macrophages were significantly correlated with the levels of PLK1, MMP9, CCNA2. The infiltration levels of CD8+ T cells were significantly correlated with the expression of MMP9. Next, we also investigated the relationships between copy number variations of hub genes and immune cell infiltrations and found that 8 hub genes were significantly related to immune cell infiltrations. Arm-level deletion of C8A and C8B was significantly associated with CD8+ T cells and neutrophils infiltration. Arm level of gain of C6, C7 and CCNA2 was involved in CD4+ T cells infiltration. High amplification of BIRC5 and CCNA2 was associated with macrophages infiltration, and BIRC5 and CFB amplification was relevant with neutrophils infiltration. Immune cells have diverse function in different stages of HCC progression [29] and play contradictory roles in tumor microenvironment such as inhibiting the tumor progression or supporting the tumor growth [30, 31]. For instance, macrophages can be activated via inflammatory signals and then eliminate precancerous senescent cells together with CD4+ T cells [32]. However, tumor-associated macrophages secrete different cytokines, such as IL-1β, IL-6, TNF-α and VEGF, to promote tumor proliferation, migration and invasion [33, 34]. Therefore, the relationships between the above 7 hub genes and immune cells can provide new insight for studying the mechanism of HCC.

In conclusion, our study firstly selected TCGA for obtaining the RNA-seq data because it is a publicly available and comprehensive database and contains many different types of cancer samples such as HCC. Then, the DEGs were identified by differential expression analysis and overlapped with immune-related genes to obtain immune-related DEGs. These DEGs were successively identified by prognostic survival analysis, PPI network construction, functional enrichment and tumor-infiltration analysis. Finally, our work identified 14 immune-related genes which were significantly related to prognosis of HCC and validated to be involved in the progression of HCC through some potential biological processes (e.g. regulation of complement activation, extracellular matrix disassembly) and KEGG pathways (e.g. complement and coagulation cascades, cell cycle, proteoglycans in cancer). Moreover, the results suggested the relationship between the genes and immune cell infiltration levels, thereby providing a direction for further research on the immunotherapy of HCC.

Author contributions

CBL and HFW designed the study, supervised the data collection, ZKF analyzed the data, interpreted the data, YBF, JZ and JYQ prepare the manuscript for publication and reviewed the draft of the manuscript. All authors have read and approved the manuscript.

Ethics approval and consent to participate

Those data are public and available and thus there is no need of ethical approvals and consent to participate in this study.

Acknowledgment

Thanks to all the peer reviewers and editors for their opinions and suggestions.

Funding

This research did not receive any specific grant from funding agencies.

Conflict of interest

The authors state that there are no conflicts of interest to disclose.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at https://....

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

References

[1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a Cancer Journal for Clinicians. 2018; 68: 394-424.

[2] Llovet JM, Zucman-Rossi J, Pikarsky E, Sangro B, Schwartz M, Sherman M, *et al.* Hepatocellular carcinoma. Nature Reviews Disease Primers. 2016; 2: 16018.

[3] Liu C, Chen K, Chen P. Treatment of liver cancer. Cold Spring Harbor Perspectives in Medicine. 2015; 5: a021535.

[4] Wang H, Lu Z, Zhao X. Tumorigenesis, diagnosis, and therapeutic potential of exosomes in liver cancer. Journal of Hematology & Oncology. 2019; 12: 133.

[5] Anwanwan D, Singh SK, Singh S, Saikam V, Singh R. Challenges in liver cancer and possible treatment approaches. Biochimica et Biophysica Acta - Reviews on Cancer. 2020; 1873: 188314.

[6] Yim SY, Kang SH, Shin JH, Jeong YS, Sohn BH, Um SH, *et al.* Low ARID1A expression is associated with poor prognosis in hepatocellular carcinoma. Cells. 2020; 9: 2002.

[7] Sugie T. Immunotherapy for metastatic breast cancer. Chinese Clinical Oncology. 2018; 7: 28.

[8] Zhang H, Wu D, Jin M. GCDH contributes to better outcome and

acts on chemoresistance and immune exclusion in cervical cancer. European Journal of Gynaecological Oncology. 2019; 40: 831-838.
[9] Wedekind MF, Denton NL, Chen C, Cripe TP. Pediatric cancer immunotherapy: opportunities and challenges. Pediatric Drugs. 2018; 20: 395-408.
[10] Schreiber RD, Old LJ, Smyth MJ. Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. Science. 2011; 331: 1565-1570.
[11] Vesely MD, Kershaw MH, Schreiber RD, Smyth MJ. Natural innate and adaptive immunity to cancer. Annual Review of Immunology. 2011; 29: 235-271.
[12] Mittal D, Gubin MM, Schreiber RD, Smyth MJ. New insights into cancer immunoediting and its three component phases-elimination, equilibrium and escape. Current Opinion in Immunology. 2014; 27: 16-25.
[13] Pardee AD, Butterfield LH. Immunotherapy of hepatocellular carcinoma: Unique challenges and clinical opportunities. Oncoimmunology. 2012; 1: 48-55.
[14] Li S, Yang F, Ren X. Immunotherapy for hepatocellular carcinoma. Drug Discoveries & Therapeutics. 2015; 9: 363-371.
[15] Morgan BP, Harris CL. Complement, a target for therapy in inflammatory and degenerative diseases. Nature Reviews Drug Discovery. 2015; 14: 857-877.
[16] Bajic G, Degn SE, Thiel S, Andersen GR. Complement activation, regulation, and molecular basis for complement-related diseases. The EMBO Journal. 2015; 34: 2735-2757.
[17] Bishayee A. The role of inflammation and liver cancer. Advances in Experimental Medicine and Biology. 2014; 816: 401-435.
[18] Kuzet S, Gaggioli C. Fibroblast activation in cancer: when seed fertilizes soil. Cell and Tissue Research. 2016; 365: 607-619.
[19] Satyam A, Graef ER, Lapchak PH, Tsokos MG, Dalle Lucca JJ, Tsokos GC. Complement and coagulation cascades in trauma. Acute Medicine & Surgery. 2019; 6: 329-335.
[20] Sherr CJ. Cancer cell cycles. Science. 1996; 274: 1672-1677.
[21] Baghy K, Tátrai P, Regős E, Kovalszky I. Proteoglycans in liver cancer. World Journal of Gastroenterology. 2016; 22: 379-393.
[22] Farra R, Grassi G, Tonon F, Abrami M, Grassi M, Pozzato G, *et al.* The role of the transcription factor E2F1 in hepatocellular carcinoma. Current Drug Delivery. 2017; 14: 272-281.
[23] Liu Z, Sun Q, Wang X. PLK1, a potential target for cancer therapy. Translational Oncology. 2017; 10: 22-32.
[24] Xu L, Zhu Y, Shao J, Chen M, Yan H, Li G, *et al.* Dasatinib synergises with irinotecan to suppress hepatocellular carcinoma via inhibiting the protein synthesis of PLK1. British Journal of Cancer. 2017; 116: 1027-1036.
[25] Niu H, Li F, Wang Q, Ye Z, Chen Q, Lin Y. High expression level of MMP9 is associated with poor prognosis in patients with clear cell renal carcinoma. PeerJ. 2018; 6: e5050.
[26] Padhi SS, Roy S, Kar M, Saha A, Roy S, Adhya A, *et al.* Role of CDKN2a/p16 expression in the prognostication of oral squamous cell carcinoma. Oral Oncology. 2017; 73: 27-35.
[27] Li F, Aljahdali I, Ling X. Cancer therapeutics using survivin BIRC5 as a target: what can we do after over two decades of study? Journal of Experimental & Clinical Cancer Research. 2019; 38: 368.
[28] Gao T, Han Y, Yu L, Ao S, Li Z, Ji J. CCNA2 is a prognostic biomarker for ER+ breast cancer and tamoxifen resistance. PLoS ONE. 2014; 9: e91771.
[29] Zhang Q, Lou Y, Bai X, Liang T. Immunometabolism: a novel perspective of liver cancer microenvironment and its influence on tumor progression. World Journal of Gastroenterology. 2018; 24: 3500-3512.
[30] Domingues P, González-Tablas M, Otero Á, Pascual D, Miranda D, Ruiz L, *et al.* Tumor infiltrating immune cells in gliomas and meningiomas. Brain, Behavior, and Immunity. 2016; 53: 1-15.
[31] He J, LZ, L. Wang, S. Zhuang. Effects of regulatory T cells, natural killer cells, and natural killer T cells on immunosuppression therapy in patients with recurrent embryo implantation failure. Clinical and Experimental Obstetrics & Gynecology. 2019; 46: 606-610.
[32] Eggert T, Wolter K, Ji J, Ma C, Yevsa T, Klotz S, *et al.* Distinct functions of senescence-associated immune responses in liver tumor surveillance and tumor progression. Cancer Cell. 2016; 30: 533-547.
[33] Yeung OWH, Lo C, Ling C, Qi X, Geng W, Li C, *et al.* Alternatively activated (M2) macrophages promote tumour growth and invasiveness in hepatocellular carcinoma. Journal of Hepatology. 2015; 62: 607-616.
[34] Capece D, Fischietti M, Verzella D, Gaggiano A, Cicciarelli G, Tessitore A, *et al.* The inflammatory microenvironment in hepatocellular carcinoma: a pivotal role for tumor-associated macrophages. BioMed Research International. 2013; 2013: 187204.