Identification of a 9-gene signature related to stemness for prognosis in hepatocellular carcinoma basing on weighted gene coexpression network analysis

Xinghong Yao
  Sichuan Cancer Hospital and Institute

Chengquan Lu
  Dalian Medical University

Qidi Pang
  Hospital of Chengdu University of Traditional Chinese Medicine

Ye Zeng (✉ ye@scu.edu.cn)
  Sichuan University  https://orcid.org/0000-0001-5440-6194

Primary research

Keywords: stemness index, mRNAsi, hepatocellular carcinoma, prognosis, microenvironment

DOI: https://doi.org/10.21203/rs.3.rs-210260/v1

License: ☕️ 📧 This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background

Cancer stem cells (CSCs) are heterogeneous with self-renewal and differentiation ability. The mRNA expression-based stemness index (mRNAsi) described the similarity between tumor cells and CSCs, which is positively associated with the poor prognosis of cancer patients. However, the key prognostic genes related to mRNAsi in hepatocellular carcinoma (HCC) remains unclear.

Methods

In this study, RNA sequencing (RNA-seq) data of 374 liver HCC (LIHC) and 50 normal liver tissue samples were downloaded from The Cancer Genome Atlas (TCGA) database, with the corresponding clinical phenotypes. Differential expression of genes between tumor and normal tissues were analyzed. Modules correlated to mRNAsi and hub genes were identified using the weighted gene coexpression network analysis (WGCNA) package. Then, the prognostic genes associated with mRNAsi were identified. The expressions of the prognostic genes in HCC tissues were achieved by the human protein atlas (HPA) database. Gene Ontology (GO), and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of the prognostic genes were performed. Finally, the correlations between hub genes and clinical characteristics, immune subtypes, and tumor microenvironment were analyzed.

Results

A total of 1146 downregulated genes and 5593 upregulated genes were identified in LIHC, compared with normal tissues. A 9-gene signature related to mRNAsi and HCC prognosis including PSMG3, SNRPD1, DTYMK, PIGU, NME1, TXNL4A, IPO4, PES1, and REXO4 was obtained and their expressions were confirmed by the HPA database. High expression of this signature indicates poor prognosis of HCC. Among them, DTYMK and NME1 enriched in pyrimidine metabolism, SNRPD1 and TXNL4A enriched in spliceosome and PIGU enriched in glycosyl phosphatidylinositol (GPI)-anchor biosynthesis pathways. PIGU was an independent prognostic factor of HCC, which was significantly associated with clinical characteristics. High expression of the 9-gene signature was negatively correlated with the stromal cell infiltration, and positively correlated with specific immune subtypes-related to angiogenesis, M1/M2 macrophage polarization, and M2 response.

Conclusion

A 9-gene signature related to mRNAsi and poor prognosis in HCC were identified, which can be used as biomarkers for the diagnosis of HCC and functional mechanism exploration of CSCs in HCC.
Introduction

Primary liver cancer is the fourth most common tumor in the world (Villanueva, 2019). Hepatocellular carcinoma (HCC) takes account of 90% of all cases of primary liver cancer (Chedid et al., 2017), which is the third most fatal cancer in the world (Salem and Lewandowski, 2013). The early onset of HCC is hidden without obvious symptoms and signs, most of HCC has entered the middle and late stages once diagnosed was ensured (Hartke et al., 2017). In recent years, with the continuous progress of comprehensive treatments including surgery and targeted drugs, the overall survival (OS) of patients had improved significantly, but tumor metastasis and recurrence, radiotherapy and chemotherapy resistances lead to a high mortality rate of HCC (Andreou et al., 2018; Kulik and El-Serag, 2019).

The American Association for Cancer Research (AACR) has been defined cancer stem cell (CSC) as Cells that exist in tumor tissue with unlimited self-renewal ability and produce tumor cells with different degrees of differentiation at the same time (Eaves, 2008). Cancer progression includes the progressive loss of differentiation phenotype and the acquisition of progenitor cell and stem cell-like characteristics. The transcriptomic data of all tumor and control samples in the TCGA database were previously acquired by applying a one-class logistic regression machine learning algorithm (OCLR) to obtain mRNA expression-based stemness index (mRNAsi) which is an indicator of their stem cell characteristics in response to gene expression (Malta et al., 2018). It was believed that CSCs are the seed cells of HCC, which may be of endogenous, mainly hepatic origin (Robrecht et al., 1998; Yi and Nan, 2008). CSCs have the properties of mutual transformation with tumor cells, promotion of tumor development, metastasis and recurrence, and resistance to radiotherapy and chemotherapy (Xu et al., 2010; Nassar and Blanpain, 2016). The degree of malignancy of HCC is closely associated with CSC phenotypes (Cheng et al., 2016; Nio et al., 2017). The postoperative recurrence rate of patients with low-differentiated HCC is significantly higher than those with intermediate or high-differentiated HCC.

In recent, weighted gene coexpression network analysis (WGCNA) was used to evaluate the correlation between clinical traits and coexpression gene modules by constructing scale-free topology and find the hub genes related to clinical traits (Langfelder and Horvath, 2008b). By WGCNA, the key genes and pathways for HCC initiation and progression were identified, including GMPS, ACACA, ALB, TGFB1, KRAS, ERBB2, BCL2, EGFR, STAT3, and CD8A, which enriched in response to endoplasmic reticulum stress, cellular response to cadmium and zinc ion, cell locomotion and differentiation, ribonucleoprotein complex and RNA processing, and immune system process, etc. (Zhang et al., 2017). The hub genes such as PLK1 and CDK1 that regulated the tumorigenesis in HCC were identified (Xu et al., 2016; Li et al., 2019a). Hsa-cir-0000517 was identified to be associate with adverse prognosis in HCC (Wang et al., 2019). The modules of immune-related genes were analyzed to explore the immune escape mechanisms of HCC (Li et al., 2019b). It was demonstrated that mRNAsi expression in LIHC increased with the tumor pathological grade (Bai et al., 2020). Twenty-one key genes including DCN, ECM1, HAND2, PTGIS, SFRP1, SRPX, COLEC10, GRP182, ADAMTS7, CD200, CDH11, COL8A1, FAP, LZTS1, MAP1B, NAV1, NOTCH3, OLFML2A, PRR16, TMEM119, and VCAN were reported to involve in angiogenesis, negative regulation of DNA-binding transcription factor activity, apoptosis, and autophagy by identifying stemness-related genes
through WGCNA (Bai et al., 2020). In the present study, we identified a gene signature related to prognosis and mRNAsi in HCC, and analyzed the relationships between those mRNAsi and prognosis-related genes and clinical characteristics, immune subtypes, and tumor microenvironment.

Materials And Methods

Data download and preprocessing

The RNA-Seq data and the corresponding clinical information of 374 liver HCC (LIHC) and 50 normal liver tissue samples were downloaded from The Cancer Genome Atlas TCGA databases (https://www.cancer.gov/) in November 2019. The human.gtf was downloaded from the ensemble website (https://asia.ensembl.org/info/website/upload/gff.html). The mRNAsi was obtained from a previously published study (Malta et al., 2018). Immune subtypes were downloaded from the xena browser (https://xenabrowser.net/datapages/).

Establishment of WGCNA modules and identification of stemness index and prognosis-related genes

The WGCNA R package (Langfelder and Horvath, 2008a) was used to construct a co-expression network. In the first, Limma R package (Ritchie et al., 2015) was used to identify the differentially expressed genes (DEGs) with selection criteria: false discovery rate (FDR) < 0.05, and |log2FC| > 1. After deleting the outliers by setting the abline value to 20000, DEGs were used to construct a weighted co-expression network, which is satisfied scale-free topology (soft threshold $\beta = 9$; scale-free $R^2 = 0.89$). The average linkage hierarchical clustering is established by hierarchical clustering based on topological overlap matrix (TOM)-based dissimilarity. Gene clustering tree is obtained to identify the mRNAsi-related gene module and epigenetically regulated mRNAsi (EREG-mRNAsi)-related gene module (minimum gene number size of 50, $P < 0.05$). Once the modules of interest were selected, hub genes in corresponding modules and gene expression profiles were screened out by thresholds of gene significance (cor. GS) > 0.3, module membership (cor. MM) > 0.7, and $P < 0.05$. Survival analysis of hub genes was performed by the survival (T., 2020) and survminer R package (Kassambara., 2020) with the Kaplan-Meier (KM) method. The genes significantly associated with overall survival (OS) ($P < 0.05$) were screened out and their receiver operating characteristic curves (ROCs) were performed by using the survival ROC R package (J., 2013). The genes in the module with the highest correlation with mRNAsi and prognosis were identified as mRNAsi and prognosis-related genes.

GO and KEGG Analysis

To evaluate the biological functions of mRNAsi and prognosis-related genes, we used the clusterProfiler R package (Yu et al., 2012) for GO functional annotations and KEGG pathway analysis. The threshold values were as follows: $p < 0.05$, and FDR < 0.05.

Gene co-expression analysis
To verify the co-expression of mRNAsi and prognosis-related genes, we used the corrplot R package (Wei and Simko, 2013) to obtain the correlation between the expression levels of each gene, as well as the Person's correlation coefficient.

**Clustering of tumor samples basing on the hub gene-signature, and comparison of OS and clinical characteristics**

To further analyze the role of mRNAsi and prognosis-related genes in HCC, we clustered the tumor samples into two subgroups by ConsensusClusterPlus R package (Wilkerson and Hayes, 2010) according to the expression levels of those hub genes, and performed PCA analysis using the limma R package (Ritchie et al., 2015) and visualized by the ggplot2 R package (Villanueva and Chen, 2019). Then, The difference in OS between the two groups was compared using the survival R package (T., 2020). The differences in clinical characteristics (age, gender, T1-T4, N, M, stage I-IV, and grade) between the two groups were also analyzed. For comparison, t-test was used for numerical variables and the non-parametric test was performed for categorical variables. The survival and forestplot R packages (Lumley., 2020; T., 2020) were used for univariate Cox regression analysis, and the survival R package (T., 2020) was used for multivariate regression analysis.

**Gene set enrichment analyses (GSEA)**

GSEA is a computational method that determines whether an a priori defined set of genes shows statistically significant, concordant differences between two biological states (Subramanian et al., 2005). GSEA v4.0.3 for Windows was used. The gene sets including c5.all.v7.0.symbols.gmt and c2.cp.kegg.v7.0.symbols.gmt were downloaded from the Molecular Signatures Database (http://software.broadinstitute.org/gsea/msigdb/index.jsp). The normalized enrichment score (NES) was acquired by 1000 permutations. A gene set is considered to be significantly enriched when P-value is < 0.05 and false discovery rate (FDR) is < 0.05.

**Relationships between each mRNAsi and prognosis-related gene and clinical characteristics, immune subtypes, and tumor microenvironment**

The correlations between each hub gene in mRNAsi and prognosis-related gene profile and clinical characteristics in HCC were analyzed using the beeswarm R package (Eklund., 2016). The infiltration of immune and stromal cells in HCC tumor was evaluated by immune and stromal scores using the Estimation of STromal and Immune cells in MAliignant Tumor tissues using Expression data (ESTIMATE) algorithm (Yoshihara et al., 2013), and HCC purity was assessed by ESTIMATE score. Correlations between gene panel and these scores were calculated using Spearman's correlation analysis, correlations between gene expression and immune subtypes (Thorsson et al., 2018) in the tumor microenvironment were analyzed using an ANOVA model, and the correlation between HCC purity and mRNAsi-related gene expression was detected by Spearman’s correlation test. Immune subtypes including C1 (wound healing),
C2 (IFN-γ dominant), C3 (inflammatory), C4 (lymphocyte depleted), C5 (immunologically quiet), and C6 (TGF-β dominant) were characterized as previously described (Thorsson et al., 2018).

**Expression of hub genes in HCC in Human Protein Atlas (HPA)**

The protein levels of mRNAsi and prognosis-related genes in HCC were verified by pathological staining images from the HPA database (https://www.proteinatlas.org/).

**Results**

**WGCNA analysis of DEGs basing on mRNAsi**

The bioinformatics analysis was performed to identify the DEGs between 50 normal (non-tumor) samples and 374 LIHC samples obtained from TCGA. As shown in Fig. 1, there are a total of 1146 significantly downregulated and 5593 significantly upregulated genes (Supplemental Table 1). To further analyze the DEGs obtained in the study which the highest correlated with mRNAsi, WGCNA was constructed to identify important gene modules (Fig. 2). After eliminating 22 outliers, the samples were clustered into 325 1-cluster and 27 2-clusters (Fig. 2A). Then, the 1-cluster was selected, and WGCNA analysis was performed on 7667 DEGs, and finally, 13 modules were obtained after merging the similarity modules (Fig. 2B). Among the modules, the grey module was not significantly correlated with EREG-mRNAsi, but other modules were significantly correlated with mRNAsi and EREG-mRNAsi (Fig. 2C). Among the 6 significant negative modules, the magenta module was most significantly correlated with mRNAsi, with a correlation coefficient of -0.71 (P < 0.001). Among the 6 significant positive modules, the blue module was significantly correlated with mRNAsi, with the biggest correlation coefficient of 0.38 (P < 0.001) (Fig. 2C).

**Identification of mRNAsi and prognosis-related genes**

The KM analysis of the DEGs was performed and a total of 245 of the DEGs had significantly correlated with the 3-year, 5-year survival of HCC patients (Supplemental Table 2).

To evaluate the relationship between mRNAsi and the survival of HCC patients, we compared the prognostic-related DEGs and genes in all modules. There were not prognostic-related genes in the magenta module. Most prognostic-related DEGs were also appeared in the turquoise (55 DEGs) and blue (50 DEGs) modules. However, the correlation coefficient between the turquoise module and mRNAsi was only 0.12 (Fig. 2C). So, we subsequently analyzed the mRNAsi and prognosis-related DEGs in the blue module.

With cor. GS > 0.3, cor. MM > 0.7, and P-value < 0.05, a total of 51 hub genes were obtained in the blue module (Fig. 2D, Supplemental Table 3) and 49 genes were obtained in the magenta module (Fig. 2E). The genes in the magenta module were not considered due to their poor prognostic value. Among the hub genes in the blue module, 9 hub genes were prognostic-related genes, including PSMG3, SNRPD1,
DTYMK, PIGU, NME1, TXNL4A, IPO4, PES1, and REXO4 (Fig. 3). A better prediction efficiency of 1 year, 3 years, and 5 years of the 9 hub genes was obtained.

**Verification of expression of mRNAsi and prognosis-related genes in HCC samples**

Basing on the expression of those hub genes, the HCC and normal tissue samples were clustered. All the 9 hub genes were significantly upregulated (Fig. 4A, P < 0.001). The expression of 9 genes was positively correlated, and the correlation coefficient was greater than 0.5 (Fig. 4B, P < 0.001). NME1, DTYMK, SNRPD1 and TXNL4A in a PPI network, and REXO4, PES1 and IPO4 in another PPI network (Fig. 4C). PIGU and PSMG3 were separate with other hub genes, respectively (Fig. 4C).

**Identified two clusters of HCC samples base on mRNAsi and prognosis-related genes expression**

We clustered the tumor samples according to stemness index and prognosis-related gene expression. The clustering results showed that skew of cumulative distribution function (CDF) towards 0 at k = 2 (Fig. 5A), and areas of CDF at k = 2 and k = 3 were bigger than other cluster numbers (k) (Fig. 5B). However, the robustness at k = 3 is poor (Fig. 5C), suggesting that the tumor samples divided into two clusters was the optimal choice. To determine whether our classification was correct, we analyzed the clustering of the two clusters of samples using principal component analysis (PCA), the distribution of the two subgroups was significantly different from each other (Fig. 5D).

**Compared with the clinical outcomes and clinicopathological features between two sample subgroups**

In order to further analyze the differences between the HCC samples of the 2 clusters, we used KM survival analysis and one-way ANOVA to analyze the differences in survival and clinicopathological characteristics of the two clusters of patients. The results showed that the KM survival curves of OS of the two clusters were significantly different, and the survival rates of cluster 1 patients within 9 years were significantly greater than that of cluster 2 patients (Fig. 6A, P < 0.01). Consistent with the survival, the prognosis of cluster 1 was better than that of cluster 2 (alive vs. dead, P < 0.05; Fig. 6B). Comparing the differences in hub genes expression and clinicopathological characteristics between the two subgroups, the expression of mRNAsi and prognosis-related genes in cluster 2 was significantly higher than that in cluster 1, and the T, stage, and grade between the two subgroups were significantly different (Fig. 6B, P < 0.05). The results suggest that the abnormally high expression of mRNAsi and prognosis-related genes is closely related to clinicopathological characteristics and prognosis.

**Functional annotation of mRNAsi and prognosis-related genes**
To further understand the biological function of the mRNAsi and prognosis-related genes, we performed GO and KEGG enrichment analysis (Fig. 7 and Table 1). GO analysis includes three parts of biological process (BP), cellular component (CC), and molecular function (MF). BP analysis showed that the top 10 enriched terms were pyrimidine nucleoside triphosphate biosynthetic process, ribonucleoprotein complex biogenesis, pyrimidine nucleoside triphosphate metabolic process, nucleobase-containing small molecule interconversion, pyrimidine nucleotide biosynthetic process, pyrimidine-containing compound biosynthetic process, pyrimidine nucleotide metabolic process, spliceosomal complex assembly, pyrimidine-containing compound metabolic process, and nucleoside diphosphate metabolic process. The significantly enriched BP terms were all associated with nucleic acid metabolic processes, and the major enriched genes were DTYMK and NME1 (Fig. 7A and Table 1).
Table 1  
The enriched GO terms and pathways with associated hub mRNAs and prognosis-related genes.

| Go     | Description                                                                 | q-value | Gene ID                      |
|--------|-----------------------------------------------------------------------------|---------|------------------------------|
| BP     | pyrimidine nucleoside triphosphate biosynthetic process                      | 0.001   | DTYMK/NME1                   |
|        | ribonucleoprotein complex biogenesis                                         | 0.001   | SNRPD1/TXNL4A/PES1/REXO4     |
|        | pyrimidine nucleoside triphosphate metabolic process                        | 0.001   | DTYMK/NME1                   |
|        | nucleobase-containing small molecule interconversion                        | 0.001   | DTYMK/NME1                   |
|        | pyrimidine nucleotide biosynthetic process                                   | 0.003   | DTYMK/NME1                   |
|        | pyrimidine-containing compound biosynthetic process                         | 0.004   | DTYMK/NME1                   |
|        | pyrimidine nucleotide metabolic process                                      | 0.004   | DTYMK/NME1                   |
|        | spliceosomal complex assembly                                                | 0.006   | SNRPD1/TXNL4A                |
|        | pyrimidine-containing compound metabolic process                            | 0.008   | DTYMK/NME1                   |
|        | nucleoside diphosphate metabolic process                                    | 0.010   | DTYMK/NME1                   |
| CC     | U5 snRNP                                                                    | 0.001   | SNRPD1/TXNL4A                |
|        | U2-type spliceosomal complex                                                 | 0.005   | SNRPD1/TXNL4A                |
|        | spliceosomal snRNP complex                                                   | 0.005   | SNRPD1/TXNL4A                |
|        | small nuclear ribonucleoprotein complex                                      | 0.005   | SNRPD1/TXNL4A                |
|        | Sm-like protein family complex                                               | 0.005   | SNRPD1/TXNL4A                |
|        | spliceosomal complex                                                         | 0.009   | SNRPD1/TXNL4A                |
|        | methylosome                                                                  | 0.015   | SNRPD1                       |
|        | U4 snRNP                                                                    | 0.015   | SNRPD1                       |
|        | nuclear membrane                                                             | 0.016   | TXNL4A/IPO4                  |
|        | SMN-Sm protein complex                                                       | 0.016   | SNRPD1                       |
| MF     | ribonucleoprotein complex binding                                            | 0.000   | SNRPD1/NME1/PES1             |
|        | nucleoside diphosphate kinase activity                                       | 0.000   | DTYMK/NME1                   |
| Go | Description                                                                 | q-value | Gene ID  |
|----|-----------------------------------------------------------------------------|---------|----------|
|    | phosphotransferase activity, phosphate group as acceptor                    | 0.001   | DTYMK/NME1 |
|    | nucleobase-containing compound kinase activity                              | 0.001   | DTYMK/NME1 |
|    | nuclease activity                                                           | 0.010   | NME1/REXO4 |
|    | intermediate filament binding                                                | 0.015   | NME1     |
|    | ribosomal small subunit binding                                              | 0.015   | NME1     |
|    | glycolipid binding                                                          | 0.016   | PIGU     |
|    | nucleoside monophosphate kinase activity                                    | 0.016   | DTYMK     |
|    | gamma-tubulin binding                                                       | 0.016   | NME1     |
| KEGG Pathway | Description                                                                 | q-value | Gene ID  |
|    | Pyrimidine metabolism                                                       | 0.001   | DTYMK/NME1 |
|    | Spliceosome                                                                 | 0.003   | SNRPD1/TXNL4A |
|    | Glycosylphosphatidylinositol (GPI)-anchor biosynthesis                       | 0.011   | PIGU     |

CC analysis showed that the top 10 enriched terms were U5 snRNP, U2-type spliceosomal complex, spliceosomal snRNP complex, small nuclear ribonucleoprotein complex, Sm-like protein family complex, spliceosomal complex, methylosome, U4 snRNP, nuclear membrane, and SMN-Sm protein complex. The CC significantly enriched terms were all associated with nucleic acid proteins, and the major enriched genes were SNRPD1 and TXNL4A (Fig. 7A and Table 1).

MF analysis showed that the top 10 enriched terms were ribonucleoprotein complex binding, nucleoside diphosphate kinase activity, phosphotransferase activity, phosphate group as acceptor, nucleobase-containing compound kinase activity, nuclease activity, intermediate filament binding, ribosomal small subunit binding, glycolipid binding, nucleoside monophosphate kinase activity, and gamma-tubulin binding. The MF significantly enriched terms were all related to nucleic acid processing, and the major enriched genes were DTYMK and NME1 (Fig. 7A and Table 1).

The KEGG pathway analysis showed that DTYMK and NME1 enriched in pyrimidine metabolism, SNRPD1 and TXNL4A enriched in spliceosome, and PIGU enriched in glycosyl phosphatidylinositol (GPI)-anchor biosynthesis pathways (Fig. 7B and Table 1).

**Prognostic value of mRNAsi and prognosis-related genes**
To better understand the prognostic role of mRNAsi and prognosis-related genes in HCC, we performed univariate Cox regression analysis based on the expression status of each hub gene, with the patient's prognosis as the outcome variable, in the analysis we assigned a risk score to each patient, patients were divided into high- and low-risk subgroups (Supplemental Table 4). The univariate Cox regression analysis results showed that high expression of PES1 (HR = 1.669, 95%CI = 1.270–2.192), PSMG3 (HR = 1.332, 95%CI = 1.086–1.633), TXNL4A (HR = 1.709, 95%CI = 1.289–2.267), DTYMK (HR = 1.679, 95%CI = 1.345–2.097), IPO4 (HR = 1.916, 95%CI = 1.285–2.858), SNRPD1 (HR = 1.583, 95%CI = 1.233–2.032), PIGU (HR = 2.006, 95%CI = 1.513–2.660), REXO4 (HR = 1.905, 95%CI = 1.393–2.606), NME1 (HR = 1.358, 95%CI = 1.113–1.658) had a poor prognosis in HCC (Fig. 8A).

As all genes were significant in univariate cox regression analysis, we included all genes in the multifactorial regression model for calculation, which showed that PIGU was an independent prognostic factor (HR = 2.01, 95% CI = 1.51–2.66). For the results of multifactorial regression analysis, we applied KM survival analysis to establish a survival model for the difference in OS between the high-risk and low-risk subgroups, and the results showed that there was a significant difference in the survival of patients in the high-risk and low-risk subgroups at different times (Fig. 8B, P < 0.001), and ROC for model validation showed that the prediction efficiency of the model was 0.672 (Fig. 8C), suggesting that the model has a good predictive role.

**Gene set enrichment analysis (GSEA) delineates biological pathways basing on the risk score**

We performed GSEA analysis based on the risk scores and found that in the high-risk subgroup, GO analysis were predominantly enriched in RNA catabolic process, regulation of cell cycle phase transition, cell cycle G2-M phase transition, mRNA cis splicing via spliceosome, non-canonical wnt signaling pathway, rRNA catabolic process, negative regulation of response to DNA damage stimulus, and regulation of intrinsic apoptotic signaling pathway (Fig. 9A and Supplemental Table 5). KEGG analysis was predominantly enriched in spliceosome, RNA degradation, pyrimidine metabolism, DNA replication, n glycan biosynthesis, cell cycle, RNA polymerase, pathways in cancer, wnt signaling pathway, mismatch repair, etc. (Fig. 9B and Supplemental Table 5).

In the low-risk subgroup, the GO analysis enriched in cellular amino acid catabolic process, alpha amino acid catabolic process, aspartate family amino acid catabolic process, aromatic amino acid family catabolic process, branched chain amino acid catabolic process, glutamate metabolic process, sulfur amino acid metabolic process, monooxycylic acid metabolic process, alpha amino acid metabolic process, etc. (Fig. 9C and Supplemental Table 5). The KEGG analysis enriched in tryptophan metabolism, valine leucine and isoleucine degradation, glycine serine and threonine metabolism, arginine and proline metabolism, tyrosine metabolism, phenylalanine metabolism, alanine aspartate and glutamate metabolism, pyruvate metabolism, and starch and sucrose metabolism (Fig. 9D and Supplemental Table 5).
Relationship between mRNAsi and prognosis-related genes and clinical characteristics

Further analysis of the correlation between 9 hub genes and clinical characteristics (Fig. 10) revealed that PSMG3 was significantly associated with grade (Fig. 10A), SNRPD1 was significantly associated with age and grade (Figs. 10B-C). DTYMK was significantly associated with age, stage, grade, and T (Figs. 10D-G). PIGU was significantly associated with gender, stage, grade, and T (Figs. 10H-K). IPO was significantly associated with grade (Fig. 10L). NME1 was significantly associated with stage, grade, and T (Figs. 10M-O). TXNL4A and REXO4 were significantly associated with stage, grade, and T (Figs. 10P-U). It was not observed significant association between PES1 and the clinical characteristics.

Relationship between mRNAsi and prognosis-related genes and immune microenvironment

The relationships between 9 hub genes and immune subtypes were analyzed (Fig. 11). It was demonstrated that patients with C3 and C5 immune subtypes had better survival than those with C4 (Tamborero et al., 2018). No HCC patients were characterized to the C5 immune subtype (Fig. 11A). All the 9 hub genes were highly expressed in LIHC samples, and their higher expressions were correlated with C1, C2, and C4, indicated they promote angiogenesis, M1/M2 macrophage polarization, and M2 response, thus associated with prognosis of HCC.

In the microenvironment analysis (Fig. 11B), the 9 hub genes were negatively correlated with the stroma (stromal score) (P < 0.001), and SNRPD1 and TXNL4 were positively correlated with immune infiltrate (immune score) (P < 0.001). There was a negative correlation between NME1 and tumor purity (ESTIMATE score) (P < 0.05).

The mRNAsi and prognosis-related genes Expression in HCC Tissue

IHC staining (HPA databases) confirmed that PSMG3, SNRPD1, DTYMK, PIGU, NME1, IPO4, PES1, and REXO4 levels were upregulated in HCC tissues, compared with normal tissues (Fig. 12).

Discussion

The origin of hepatoma stem cells is still unclear, and it is generally believed that genetic mutation of normal hepatic stem cells (Oishi et al., 2014) and various factors-induced mutation of some mature hepatocytes to form hepatoma stem cells via de-differentiate and regain the ability to continue to proliferate and differentiate (Bu and Cao, 2012; Karakasiliotis and Mavromara, 2015). Some hub genes and pathways associated with the cancer stem cell characteristics in LIHC were previously identified using WGCNA analysis of transcriptome stemness index (Bai et al., 2020). In this study, we found that PSMG3, SNRPD1, DTYMK, PIGU, NME1, TXNL4A, IPO4, PES1, and REXO4 are potential novel biomarkers for prognosis of HCC and functional mechanism exploration of the CSCs in HCC. Those mRNAsi- and
prognosis-related genes in HCC was significantly higher than that in normal liver tissues. The expression of those hub genes was confirmed by HPA databases, except TXNL4A. The mutation of REXO4 was implied in prolactinoma (Melo et al., 2016). The prognosis of HCC patients with overexpression of those genes was poorer, showing the 9-gene signature is a credible prognostic indicator of HCC. In addition, the nine genes were negatively correlated with stroma in the microenvironment. SNRPD1 and TXNL4 were positively correlated with immune cell infiltration, and NME1 was negatively correlated with the tumor purity, suggesting that those genes may affect the purity of tumor cells and mediate the microenvironment stimulus.

Proteasome assembly chaperone 3 (PSMG3) is significantly correlated with tumor grade. However, its antisense strand PSMG3-AS1 that is a lncRNA partially complementary to PSMG3, was found to promote the progression of breast, liver, and lung cancers by increasing the stability of PSMG3(Cui et al., 2020; Yue et al., 2020; Zhang et al., 2020). PPI analysis showed that PSMG3 not interacted with other identified hub genes, its function in HCC should be investigated in the future.

Small nuclear ribonucleoprotein Sm D1 (SNRPD1) is a novel gene signature for prognosis prediction in ovarian cancer (Bao et al., 2020). Overexpression of SNRPD1 is significantly associated with poor survival of lung adenocarcinomas (Yi et al., 2020). It was demonstrated that SNRPD1 is associated with mesenchymal stem cell osteogenic differentiation by WGCNA (Yang et al., 2019). We found that SNRPD1 is significantly enriched in the spliceosome pathway, which might contribute to maintaining the stem cell characteristics of liver CSCs.

Deoxythymidylate kinase (DTYMK) is involved in dTTP biosynthesis, which is part of pyrimidine metabolism. LKB1 mutations in human non-cellular lung cancer exhibit a positive effect on DTYMK sensitivity, suggesting that DTYMK is a potential therapeutic target for lung cancer (Liu et al., 2013). In low-differentiated liver cancer, pyrimidine metabolism rate-limiting enzyme is a signature gene of cancer stemness, which is associated with poor prognosis (Yeh et al., 2017), which is consistent with our results.

Nucleoside diphosphate kinase A (NME1) is a member of the NME gene family (Puts et al., 2018). Extracellular NME has a role in promoting stem cell pluripotency and inducing central nervous system development as well as neuroprotective effects (Romani et al., 2018). Low expression of NME1 is associated with poor prognosis, low survival rate, and lymph node infiltration in breast cancer, ovarian cancer, and melanoma (Hartsough and Steeg, 2000). However, overexpression of NME1 is a poor prognostic signature for HCC (Yang et al., 2017; Hindupur et al., 2018). Functional analysis revealed that NME1 enrichment is associated with pyrimidine metabolism.

Thioredoxin like 4A (TXNL4A) is a pre-mRNA splicing factor gene, which is upregulated in a maternal HF diet caused hepatic hypermethylation in male mice (Seki et al., 2017), and endometrial cancer (Wang et al., 2020). The role of TXNL4A in the transformation of cells in HCC tumors to stem cells remains unclear.

Importin-4 (IPO4) acts as a nuclear transport receptor for nuclear protein import. IPO4 was found to be involved in poor prognosis in primary gastric cancer (Xu et al., 2019). IPO4-mediated CEBPD nuclear
import can increase cervical cancer chemosensitivity by inhibiting PRKDC-driven DNA damage repair, thereby increasing cervical cancer incidence (Zhou et al., 2020). However, IPO4 has not been reported in hepatocellular carcinoma. We found that IPO4 is associated with the grade in HCC. IPO4 might inhibit DNA damage repair of HCC cells, which should be studied in the future.

Pescadillo ribosomal biogenesis factor 1 (PES1) is involved in ribosome generation and embryonic development (Lapik et al., 2004; Hölzel et al., 2005). Overexpression of PES1 promotes the development of breast cancer, ovarian cancer, and gastric cancer (Thomas and Gustafsson, 2012; Li et al., 2013; Li et al., 2016), and is reported to be associated with poor prognosis in HCC (Fan et al., 2018; Fu et al., 2019). Cell growth of liver CSCs was regulated by CD44 via a miR-105-5p/PSE1 signaling pathway (Wei et al., 2019). Our study showed that PES1 was not significantly correlated with clinical stages and grades. The relationship between CSCs and clinical stages and grades should be carefully considered.

Phosphatidylinositol glycan anchor biosynthesis class U protein (PIGU) is involved in the recognition of either the GPI attachment signal or the lipid portion of GPI (Hong et al., 2003). PIGU overexpression was found to express differentially in different stages of HCC and could be used for prognostic stratification of HCC patients (Cao et al., 2019; Sarathi and Palaniappan, 2019). We found that PIGU is associated with gender, grade, stage, and T, and is enriched in spliceosomes (El Marabti and Younis, 2018), pyrimidine metabolism (Yeh et al., 2017), DNA repair (Gerson et al., 2006; Mauger-Saccà et al., 2012) and Notch signaling pathway, etc. The Notch pathway promotes the tumor stem cell properties of CD90+ cells in HCC (Luo et al., 2016). Thus, PIGU might play an important role in HCC CSC stemness.

To sum up, our study found a series of stemness index and prognosis-related genes in HCC. These genes can be used as biomarkers to assist in the diagnosis of tumor stem cells in hepatocellular carcinoma. Among these genes, PIGU can serve as an independent prognostic factor for the survival of HCC patients. All the 9 hub genes were highly expressed in LIHC samples, and their higher expressions were correlated with C1, C2, and C4, indicated they promote angiogenesis, M1/M2 macrophage polarization, and M2 response, thus associated with prognosis of HCC. Angiogenesis is critical for tumor development and therapy resistance (Zeng et al., 2019; Zeng and M.Fu, 2020). Further function and mechanism investigations need to be taken out for validating the roles of those hub genes.

**Declarations**

**ETHICS APPROVAL AND CONSENT TO PARTICIPATE**

Not applicable.

**DATA AVAILABILITY STATEMENT**

The datasets supporting the conclusions of this article are included within the article and its additional files.

**COMPETING INTERESTS**
The authors declare no conflict of interest.

**FUNDING**

Supported by the key research and development projects in Sichuan Province (2021YFS0188).

**AUTHORS’S CONTRIBUTIONS**

ZY contributed to the conception and design of the work. YXH, LCQ, and PQD performed the analysis. All authors work on the manuscript.

**ACKNOWLEDGEMENTS**

Not applicable.

**References**

Andreou, A., Struecker, B., Raschzok, N., Krenzien, F., Haber, P., Wabitsch, S., et al. (2018). Minimal-invasive versus open hepatectomy for hepatocellular carcinoma: Comparison of postoperative outcomes and long-term survivals using propensity score matching analysis. *Surg Oncol* 27(4), 751-758.

Bai, K.H., He, S.Y., Shu, L.L., Wang, W.D., Lin, S.Y., Zhang, Q.Y., et al. (2020). Identification of cancer stem cell characteristics in liver hepatocellular carcinoma by WGCNA analysis of transcriptome stemness index. *Cancer Med* 9(12), 4290-4298. doi: 10.1002/cam4.3047.

Bao, M., Zhang, L., and Hu, Y. (2020). Novel gene signatures for prognosis prediction in ovarian cancer. *J Cell Mol Med*. doi: 10.1111/jcmm.15601.

Bu, Y., and Cao, D. (2012). The origin of cancer stem cells. *Front Biosci (Schol Ed)* 4, 819-830. doi: 10.2741/s302.

Cao, J., Wang, P., Chen, J., and He, X. (2019). PIGU overexpression adds value to TNM staging in the prognostic stratification of patients with hepatocellular carcinoma. *Hum Pathol* 83, 90-99. doi: 10.1016/j.humpath.2018.08.013.

Chedid, M.F., Kruel, C.R.P., Pinto, M.A., Grezzana-Filho, T.J.M., Leipnitz, I., Kruel, C.D.P., et al. (2017). Hepatocellular Carcinoma: Diagnosis and Operative Management. *Arq Bras Cir Dig* 30(4), 272-278.

Cheng, Z., Li, X., and Ding, J. (2016). Characteristics of liver cancer stem cells and clinical correlations. *Cancer Lett* 379(2), 230-238. doi: 10.1016/j.canlet.2015.07.041.

Cui, Y., Fan, Y., Zhao, G., Zhang, Q., Bao, Y., Cui, Y., et al. (2020). Novel IncRNA PSMG3-AS1 functions as a miR-143-3p sponge to increase the proliferation and migration of breast cancer cells. *Oncol Rep* 43(1), 229-239. doi: 10.3892/or.2019.7390.
Eaves, C.J. (2008). Cancer stem cells: Here, there, everywhere? *Nature* 456(7222), 581-582. doi: 10.1038/456581a.

Eklund., A. (2016). beeswarm: The Bee Swarm Plot, an Alternative to Stripchart. R package version 0.2.3. [https://CRAN.R-project.org/package=beeswarm](https://CRAN.R-project.org/package=beeswarm).

El Marabti, E., and Younis, I. (2018). The Cancer Spliceome: Reprograming of Alternative Splicing in Cancer. *Front Mol Biosci* 5, 80. doi: 10.3389/fmolb.2018.00080.

Fan, P., Wang, B., Meng, Z., Zhao, J., and Jin, X. (2018). PES1 is transcriptionally regulated by BRD4 and promotes cell proliferation and glycolysis in hepatocellular carcinoma. *Int J Biochem Cell Biol* 104, 1-8. doi: 10.1016/j.biocel.2018.08.014.

Fu, Z., Jiao, Y., Li, Y.Q., Ke, J.J., Xu, Y.H., Jia, B.X., et al. (2019). PES1 In Liver Cancer: A Prognostic Biomarker With Tumorigenic Roles. *Cancer Manag Res* 11, 9641-9653. doi: 10.2147/cmar.s226471.

Gerson, S.L., Reese, J., and Kenyon, J. (2006). DNA repair in stem cell maintenance and conversion to cancer stem cells. *Ernst Schering Found Symp Proc* (5), 231-244. doi: 10.1007/2789_2007_053.

Hartke, J., Johnson, M., and Ghabril, M. (2017). The diagnosis and treatment of hepatocellular carcinoma. *Semin Diagn Pathol* 34(2), 153-159.

Hartsough, M.T., and Steeg, P.S. (2000). Nm23/nucleoside diphosphate kinase in human cancers. *J Bioenerg Biomembr* 32(3), 301-308. doi: 10.1023/a:1005597231776.

Hindupur, S.K., Colombi, M., Fuhs, S.R., Matter, M.S., Guri, Y., Adam, K., et al. (2018). The protein histidine phosphatase LHPP is a tumour suppressor. *Nature* 555(7698), 678-682. doi: 10.1038/nature26140.

Hölzel, M., Rohrmoser, M., Schlee, M., Grimm, T., Harasim, T., Malamoussi, A., et al. (2005). Mammalian WDR12 is a novel member of the Pes1-Bop1 complex and is required for ribosome biogenesis and cell proliferation. *J Cell Biol* 170(3), 367-378. doi: 10.1083/jcb.200501141.

Hong, Y., Ohishi, K., Kang, J.Y., Tanaka, S., Inoue, N., Nishimura, J., et al. (2003). Human PIG-U and yeast Cdc91p are the fifth subunit of GPI transamidase that attaches GPI-anchors to proteins. *Mol Biol Cell* 14(5), 1780-1789. doi: 10.1091/mbc.e02-12-0794.

J., P. (2013). Heagerty and packaging by Paramita Saha-Chaudhuri. survivalROC: Time-dependent ROC curve estimation from censored survival data. R package version 1.0.3. [https://CRAN.R-project.org/package=survivalROC](https://CRAN.R-project.org/package=survivalROC).

Karakasiliotis, I., and Mavromara, P. (2015). Hepatocellular carcinoma: from hepatocyte to liver cancer stem cell. *Front Physiol* 6, 154. doi: 10.3389/fphys.2015.00154.
Kassambara., A. (2020). Marcin Kosinski and Przemyslaw Biecek (2020). survminer: Drawing Survival Curves using 'ggplot2'. R package version 0.4.8. https://CRAN.R-project.org/package=survminer.

Kulik, L., and El-Serag, H.B. (2019). Epidemiology and Management of Hepatocellular Carcinoma. *Gastroenterology* 156(2), 477-491.

Langfelder, P, and Horvath, S. (2008a). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9(559), 1471-2105.

Langfelder, P, and Horvath, S. (2008b). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559. doi: 10.1186/1471-2105-9-559.

Lapik, Y.R., Fernandes, C.J., Lau, L.F., and Pestov, D.G. (2004). Physical and functional interaction between Pes1 and Bop1 in mammalian ribosome biogenesis. *Mol Cell* 15(1), 17-29. doi: 10.1016/j.molcel.2004.05.020.

Li, B., Pu, K., and Wu, X. (2019a). Identifying novel biomarkers in hepatocellular carcinoma by weighted gene co-expression network analysis. *J Cell Biochem*. doi: 10.1002/jcb.28420.

Li, J., Zhou, X., Lan, X., Zeng, G., Jiang, X., and Huang, Z. (2016). Repression of PES1 expression inhibits growth of gastric cancer. *Tumour Biol* 37(3), 3043-3049. doi: 10.1007/s13277-015-4069-8.

Li, J., Zhuang, Q., Lan, X., Zeng, G., Jiang, X., and Huang, Z. (2013). PES1 differentially regulates the expression of ERα and ERβ in ovarian cancer. *IUBMB Life* 65(12), 1017-1025. doi: 10.1002/iub.1228.

Li, W., Wang, H., Ma, Z., Zhang, J., Ou-Yang, W., Qi, Y., et al. (2019b). Multi-omics Analysis of Microenvironment Characteristics and Immune Escape Mechanisms of Hepatocellular Carcinoma. *Front Oncol* 9, 1019. doi: 10.3389/fonc.2019.01019.

Liu, Y., Marks, K., Cowley, G.S., Carretero, J., Liu, Q., Nieland, T.J., et al. (2013). Metabolic and functional genomic studies identify deoxythymidylate kinase as a target in LKB1-mutant lung cancer. *Cancer Discov* 3(8), 870-879. doi: 10.1158/2159-8290.cd-13-0015.

Lumley., M.G.a.T. (2020). forestplot: Advanced Forest Plot Using 'grid' Graphics. R package version 1.10.1. https://CRAN.R-project.org/package=forestplot.

Luo, J., Wang, P, Wang, R., Wang, J., Liu, M., Xiong, S., et al. (2016). The Notch pathway promotes the cancer stem cell characteristics of CD90+ cells in hepatocellular carcinoma. *Oncotarget* 7(8), 9525-9537. doi: 10.18632/oncotarget.6672.

Malta, T.M., Sokolov, A., Gentles, A.J., Burzykowski, T., Poisson, L., Weinstein, J.N., et al. (2018). Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell* 173(2), 338-354.e315. doi: 10.1016/j.cell.2018.03.034.
Maugeri-Saccà, M., Bartucci, M., and De Maria, R. (2012). DNA damage repair pathways in cancer stem cells. *Mol Cancer Ther* 11(8), 1627-1636. doi: 10.1158/1535-7163.mct-11-1040.

Melo, F.M., Couto, P.P., Bale, A.E., Bastos-Rodrigues, L., Passos, F.M., Lisboa, R.G., et al. (2016). Whole-exome identifies RXRG and TH germline variants in familial isolated prolactinoma. *Cancer Genet* 209(6), 251-257. doi: 10.1016/j.cancergen.2016.05.065.

Nassar, D., and Blanpain, C. (2016). Cancer Stem Cells: Basic Concepts and Therapeutic Implications. *Annu Rev Pathol* 11, 47-76. doi: 10.1146/annurev-pathol-012615-044438.

Nio, K., Yamashita, T., and Kaneko, S. (2017). The evolving concept of liver cancer stem cells. *Mol Cancer* 16(1), 4. doi: 10.1186/s12943-016-0572-9.

Oishi, N., Yamashita, T., and Kaneko, S. (2014). Molecular biology of liver cancer stem cells. *Liver Cancer* 3(2), 71-84. doi: 10.1159/000343863.

Puts, G.S., Leonard, M.K., Pamidimukkala, N.V., Snyder, D.E., and Kaetzel, D.M. (2018). Nuclear functions of NME proteins. *Lab Invest* 98(2), 211-218. doi: 10.1038/labinvest.2017.109.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43(7), 20.

Robrechts, C., De Vos, R., Van den Heuvel, M., Van Cutsem, E., Van Damme, B., Desmet, V., et al. (1998). Primary liver tumour of intermediate (hepatocyte-bile duct cell) phenotype: a progenitor cell tumour? *Liver* 18(4), 288-293.

Romani, P., Igunesti, M., Gargiulo, G., Hsu, T., and Cavaliere, V. (2018). Extracellular NME proteins: a player or a bystander? *Lab Invest* 98(2), 248-257. doi: 10.1038/labinvest.2017.102.

Salem, R., and Lewandowski, R.J. (2013). Chemoembolization and Radioembolization for Hepatocellular Carcinoma. *Clinical Gastroenterology and Hepatology* 11(6), 604-611.

Sarathi, A., and Palaniappan, A. (2019). Novel significant stage-specific differentially expressed genes in hepatocellular carcinoma. *BMC Cancer* 19(1), 663. doi: 10.1186/s12885-019-5838-3.

Seki, Y., Suzuki, M., Guo, X., Glenn, A.S., Vuguin, P.M., Fiallo, A., et al. (2017). In Utero Exposure to a High-Fat Diet Programs Hepatic Hypermethylation and Gene Dysregulation and Development of Metabolic Syndrome in Male Mice. *Endocrinology* 158(9), 2860-2872. doi: 10.1210/en.2017-00334.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102(43), 15545-15550.
T., T. (2020). A Package for Survival Analysis in R. R package version 3.2-3. https://CRAN.R-project.org/package=survival 20(13), 2053-2054.

Tamborero, D., Rubio-Perez, C., Muiños, F., Sabarinathan, R., Piulats, J.M., Muntasell, A., et al. (2018). A Pan-cancer Landscape of Interactions between Solid Tumors and Infiltrating Immune Cell Populations. Clin Cancer Res 24(15), 3717-3728. doi: 10.1158/1078-0432.Ccr-17-3509.

Thomas, C., and Gustafsson, J. (2012). Targeting PES1 for restoring the ERα/ERβ ratio in breast cancer. J Clin Invest 122(8), 2771-2773. doi: 10.1172/jci65133.

Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Ou Yang, T.H., et al. (2018). The Immune Landscape of Cancer. Immunity 48(4), 812-830.

Villanueva, A. (2019). Hepatocellular Carcinoma. N Engl J Med 380(15), 1450-1462.

Villanueva, R.A.M., and Chen, Z.J. (2019). ggplot2: Elegant Graphics for Data Analysis (2nd ed.). Measurement Interdisciplinary Research & Perspectives 17(3), 160-167.

Wang, L., Zhou, W., Li, H., Yang, H., and Shan, N. (2020). Clinical Significance, Cellular Function, and Potential Molecular Pathways of CCT7 in Endometrial Cancer. Front Oncol 10, 1468. doi: 10.3389/fonc.2020.01468.

Wang, X., Wang, X., Li, W., Zhang, Q., Chen, J., and Chen, T. (2019). Up-Regulation of hsa_circ_0000517 Predicts Adverse Prognosis of Hepatocellular Carcinoma. Front Oncol 9, 1105. doi: 10.3389/fonc.2019.01105.

Wei, S., Liu, K., He, Q., Gao, Y., and Shen, L. (2019). PES1 is regulated by CD44 in liver cancer stem cells via miR-105-5p. FEBS Lett 593(14), 1777-1786. doi: 10.1002/1873-3468.13459.

Wei, T., and Simko, V. (2013). corrplot: Visualization of a Correlation Matrix. Mmwr Morbidity & Mortality Weekly Report 52(12), 145-151.

Wilkerson, M.D., and Hayes, D.N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics 26(12), 1572-1573.

Xu, X., Zhang, X., Xing, H., Liu, Z., Jia, J., Jin, C., et al. (2019). Importin-4 functions as a driving force in human primary gastric cancer. J Cell Biochem 120(8), 12638-12646. doi: 10.1002/jcb.28530.

Xu, X., Zhou, Y., Miao, R., Chen, W., Qu, K., Pang, Q., et al. (2016). Transcriptional modules related to hepatocellular carcinoma survival: coexpression network analysis. Front Med 10(2), 183-190. doi: 10.1007/s11684-016-0440-4.

Xu, X.L., Xing, B.C., Han, H.B., Zhao, W., Hu, M.H., Xu, Z.L., et al. (2010). The properties of tumor-initiating cells from a hepatocellular carcinoma patient’s primary and recurrent tumor. Carcinogenesis 31(2), 167-
Yang, J., Lv, Z., Huang, J., Zhao, Y., and Li, Y. (2017). High expression of NME1 correlates with progression and poor prognosis in patients of hepatocellular carcinoma. *Int J Clin Exp Pathol* 10(8), 8561-8568.

Yang, W., Xia, Y., Qian, X., Wang, M., Zhang, X., Li, Y., et al. (2019). Co-expression network analysis identified key genes in association with mesenchymal stem cell osteogenic differentiation. *Cell Tissue Res* 378(3), 513-529. doi: 10.1007/s00441-019-03071-1.

Yeh, H.W., Lee, S.S., Chang, C.Y., Hu, C.M., and Jou, Y.S. (2017). Pyrimidine metabolic rate limiting enzymes in poorly-differentiated hepatocellular carcinoma are signature genes of cancer stemness and associated with poor prognosis. *Oncotarget* 8(44), 77734-77751. doi: 10.18632/oncotarget.20774.

Yi, M., Li, T., Qin, S., Yu, S., Chu, Q., Li, A., et al. (2020). Identifying Tumorigenesis and Prognosis-Related Genes of Lung Adenocarcinoma: Based on Weighted Gene Coexpression Network Analysis. *Biomed Res Int* 2020, 4169691. doi: 10.1155/2020/4169691.

Yi, S.Y., and Nan, K.J. (2008). Tumor-initiating stem cells in liver cancer. *Cancer Biol Ther* 7(3), 325-330.

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

Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* 16(5), 284-287. doi: 10.1089/omi.2011.0118.

Yue, N., Ye, M., Zhang, R., and Guo, Y. (2020). MiR-449b-5p targets IncRNA PSMG3-AS1 to suppress cancer cell proliferation in lung adenocarcinoma. *BMC Pulm Med* 20(1), 152. doi: 10.1186/s12890-020-01189-5.

Zeng, Y., and M.Fu, B. (2020). Resistance mechanisms of anti-angiogenic therapy and exosomes-mediated revascularization in cancer. *Front Cell Dev Biol*. doi: 10.3389/fcell.2020.610661.

Zeng, Y., Yao, X., Liu, X., He, X., Li, L., Liu, X., et al. (2019). Anti-angiogenesis triggers exosomes release from endothelial cells to promote tumor vasculogenesis. *J Extracell Vesicles* 8(1), 1629865. doi: 10.1080/20013078.2019.1629865.

Zhang, C., Peng, L., Zhang, Y., Liu, Z., Li, W., Chen, S., et al. (2017). The identification of key genes and pathways in hepatocellular carcinoma by bioinformatics analysis of high-throughput data. *Med Oncol* 34(6), 101. doi: 10.1007/s12032-017-0963-9.

Zhang, J., Huang, J., Chen, W., Hu, Z., and Wang, X. (2020). miR-143-3p Targets IncRNA PSMG3-AS1 to Inhibit the Proliferation of Hepatocellular Carcinoma Cells. *Cancer Manag Res* 12, 6303-6309. doi: 10.2147/cmarr.s242179.
Zhou, Y., Liu, F., Xu, Q., Yang, B., Li, X., Jiang, S., et al. (2020). Inhibiting Importin 4-mediated nuclear import of CEBPD enhances chemosensitivity by repression of PRKDC-driven DNA damage repair in cervical cancer. *Oncogene* 39(34), 5633-5648. doi: 10.1038/s41388-020-1384-3.

**Supplemental Table**

Supplemental Table 5 is not available with this version.