Splicing factor proline- and glutamine-rich is a prognostic biomarker and correlated with clinical pathologic features and immune infiltrates in hepatocellular carcinoma

Chao-Ran Zhu\textsuperscript{1*}, Shi-Chen Zhu\textsuperscript{2}

\textsuperscript{1}Clinical and Basic Medical Sciences College, Shandong First Medical University, Jinan 250021, China. \textsuperscript{2}Department of Surgery, Jinan Changqing District Traditional Chinese Medicine Hospital, Jinan 250021, China.

*Corresponding to: Chao-Ran Zhu, Clinical and Basic Medical Sciences College, Shandong First Medical University, No. 6699 Qingdao Road, Huaiyin District, Jinan 250021, China. E-mail: zrz20010713@outlook.com.

Author contributions
Chao-Ran Zhu Designed the research plan, conducted data analysis, and completed the manuscript writing. Shi-Chen Zhu provided helpful suggestions and improvements for the manuscript.

Competing interests
The authors declare no conflicts of interest.

Acknowledgments
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Peer review information
Medical Data Mining thanks all anonymous reviewers for their contribution to the peer review of this paper.

Abbreviations
HCC, hepatocellular carcinoma; SFPQ, splicing factor proline- and glutamine-rich; TCGA, The Cancer Genome Atlas; FDR, false discovery rate; DEGs, differentially expressed genes; GSEA, gene set enrichment analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; PPI, protein-protein interaction; CIBERSORT, cell-type identification by estimating relative subsets of RNA transcripts; ssGSEA, single-sample gene set enrichment analysis; TME, tumor microenvironment; IncRNA, long non-coding RNAs; OS, overall survival.

Abstract
Background: Hepatocellular carcinoma (HCC) is the fourth leading cause of cancer-related deaths globally. Splicing factor proline- and glutamine-rich (SFPQ) is a multifunctional protein that controls various biological functions. As a potential therapeutic target and a promising prognostic indicator, the potential effects and processes of SFPQ in HCC require further investigation. Methods: The RNA sequencing data were obtained from the Gene Expression Omnibus, International Cancer Genome Consortium, and The Cancer Genome Atlas databases to analyze SFPQ expression and differentially expressed genes (DEGs). We utilized the LinkedOmics database to identify co-expressed genes. A Venn diagram was constructed to determine the overlapping genes between the DEGs and the co-expressed genes. Functional enrichment analysis was performed on the overlapping genes and DEGs. Furthermore, our study involved functional enrichment analysis, a protein-protein interaction network analysis, and an analysis of immune cell infiltration. The eBioPortal and Tumor Immune Single-cell Hub were utilized to investigate the genetic alterations of SFPQ and the single-cell transcriptome visualization of the tumor microenvironment. A ceRNA network was established with the assistance of the ENCORI website. Finally, we elucidated the clinical significance of SFPQ in HCC by employing Kaplan-Meier survival analysis, univariate and multivariate Cox regression, and prognostic nomogram models. Results: The expression of SFPQ in HCC tissues was significantly elevated compared to normal tissues. GSEA results indicated that increased expression of SFPQ was associated with pathways related to HCC. The ceRNA network, including SFPQ, hsa-mir-101-3p, AC0223043.4, AC124798.1, AC145207.5, and GSEC, was constructed with the assistance of ENCORI. High SFPQ expression was related to a poor prognosis in HCC and its subtypes. Univariate and multivariate Cox regression analysis showed that elevated SFPQ expression is an independent predictive factor. Conclusions: The overexpression of SFPQ may serve as a potential prognostic biomarker, indicating a poor prognosis in HCC.

Keywords: biomarker; splicing factor proline- and glutamine-rich; prognosis; immune infiltration; hepatic cell carcinoma

Shandong Medical University, Jinan 250021, China. E-mail: zrz20010713@outlook.com.

Author contributions
Chao-Ran Zhu Designed the research plan, conducted data analysis, and completed the manuscript writing. Shi-Chen Zhu provided helpful suggestions and improvements for the manuscript.

Competing interests
The authors declare no conflicts of interest.

Acknowledgments
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Peer review information
Medical Data Mining thanks all anonymous reviewers for their contribution to the peer review of this paper.

Abbreviations
HCC, hepatocellular carcinoma; SFPQ, splicing factor proline- and glutamine-rich; TCGA, The Cancer Genome Atlas; FDR, false discovery rate; DEGs, differentially expressed genes; GSEA, gene set enrichment analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; PPI, protein-protein interaction; CIBERSORT, cell-type identification by estimating relative subsets of RNA transcripts; ssGSEA, single-sample gene set enrichment analysis; TME, tumor microenvironment; IncRNA, long non-coding RNAs; OS, overall survival.

Abstract
Background: Hepatocellular carcinoma (HCC) is the fourth leading cause of cancer-related deaths globally. Splicing factor proline- and glutamine-rich (SFPQ) is a multifunctional protein that controls various biological functions. As a potential therapeutic target and a promising prognostic indicator, the potential effects and processes of SFPQ in HCC require further investigation. Methods: The RNA sequencing data were obtained from the Gene Expression Omnibus, International Cancer Genome Consortium, and The Cancer Genome Atlas databases to analyze SFPQ expression and differentially expressed genes (DEGs). We utilized the LinkedOmics database to identify co-expressed genes. A Venn diagram was constructed to determine the overlapping genes between the DEGs and the co-expressed genes. Functional enrichment analysis was performed on the overlapping genes and DEGs. Furthermore, our study involved functional enrichment analysis, a protein-protein interaction network analysis, and an analysis of immune cell infiltration. The eBioPortal and Tumor Immune Single-cell Hub were utilized to investigate the genetic alterations of SFPQ and the single-cell transcriptome visualization of the tumor microenvironment. A ceRNA network was established with the assistance of the ENCORI website. Finally, we elucidated the clinical significance of SFPQ in HCC by employing Kaplan-Meier survival analysis, univariate and multivariate Cox regression, and prognostic nomogram models. Results: The expression of SFPQ in HCC tissues was significantly elevated compared to normal tissues. GSEA results indicated that increased expression of SFPQ was associated with pathways related to HCC. The ceRNA network, including SFPQ, hsa-mir-101-3p, AC0223043.4, AC124798.1, AC145207.5, and GSEC, was constructed with the assistance of ENCORI. High SFPQ expression was related to a poor prognosis in HCC and its subtypes. Univariate and multivariate Cox regression analysis showed that elevated SFPQ expression is an independent predictive factor. Conclusions: The overexpression of SFPQ may serve as a potential prognostic biomarker, indicating a poor prognosis in HCC.

Keywords: biomarker; splicing factor proline- and glutamine-rich; prognosis; immune infiltration; hepatic cell carcinoma

Shandong Medical University, Jinan 250021, China. E-mail: zrz20010713@outlook.com.
Introduction

Hepatocellular carcinoma (HCC) is considered one of the most common types of primary liver cancers [1]. In the last several decades, HCC-related mortality has increased more rapidly than deaths associated with any other type of cancer [2]. The majority of people diagnosed with liver cancer are in the middle or late stages of the illness, and their prognosis is poor [3]. Therefore, the search for a more reliable and potent diagnostic biomarker for HCC is necessary.

Splicing factor proline- and glutamine-rich (SFPQ) is a multifunctional protein implicated in the development of various cancers. Previous studies have shown that SFPQ plays a crucial role in post-transcriptional gene silencing [4], DNA double-strand break repair [5], and homologous recombination repair of DNA damage [6]. SFPQ is upregulated in lung cancer mesenchymal stem cells and plays a role in enhancing mesenchymal stem cell proliferation, chemical resistance, and invasion capabilities [7]. The expression and subcellular localization of SFPQ represent distinctive features that can serve as a diagnostic and treatment-monitoring marker in lung cancer [8]. Additionally, loss of SFPQ/polypyrimidine tract-binding protein-associated splicing factor leads to apoptosis in BRAFV600E-driven colorectal cancer cells [9]. SFPQ is also a significant driver of melanoma, likely attributed to SFPQ-RNA interactions that enhance the expression of numerous oncogenic transcripts [10]. SFPQ/p54nrb protects cells from platinum-induced death, ultimately contributing to chemoresistance in epithelial ovarian cancer [11]. However, the prognostic significance and tumor immunology of aberrant SFPQ expression in HCC require further investigation.

This research aims to investigate and comprehend the association between SFPQ expression and clinicopathological features and its prognostic relevance. This study delves deep into the expression levels of SFPQ and investigates the potential mechanisms of SFPQ through enrichment analysis, immune infiltration, single-cell sequencing analysis, and DNA methylation analysis. Additionally, the establishment of ceRNA network provides a better understanding of the function and regulatory mechanism of SFPQ in the organism. This work provides unique insights into the processes underlying HCC carcinogenesis and identifies SFPQ as a potential biomarker for the diagnosis and prognosis of HCC.

Materials and methods

Source of data and preprocessing

We extracted data from The Cancer Genome Atlas (TCGA) database, which included 419 samples, consisting of 50 para-carcinoma tissues and 369 tumor tissues. For the purpose of this research, TCGA RNA sequencing data were converted from Fragments Per Kilobase Million format to transcripts per million. It is important to note that all written informed consents were obtained before data collection, as the TCGA database is accessible to the public in accordance with specific regulations. Additionally, gene expression profiles were sourced from the GSE36376, GSE64041, and HCCDB18 databases in Gene Expression Omnibus and International Cancer Genome Consortium. The Human Protein Atlas website presents immunohistochemical data on the expression of SFPQ in both normal and tumor tissues.

Co-expression genes and differentially expressed genes in HCC

SFPQ co-expression genes were identified using the LinkedOmics database through Spearman test. The criteria for co-expression genes were set as false discovery rate (FDR) \(< 0.05\), \(P \!< \!0.05\), and \(|\text{correlation}\| \geq 0.3\). Furthermore, patients with HCC from the TCGA database were categorized into high and low SFPQ expression groups based on the median score of SFPQ expression. Differentially expressed genes (DEGs) analysis between these groups was conducted using the R package DESeq2 [12]. The criteria for DEGs were defined as an adjusted \(P\) value \(< \!0.05\) and \(|\log(2\text{-fold change})| \geq 1\). The overlapping genes between DEGs and co-expressed genes were identified with the assistance of the Draw Venn diagram.

Functional enrichment analysis and gene set enrichment analysis (GSEA)

Cluster profile 3.6.0 was implemented to conduct gene ontology enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis [13]. Functional enrichment analysis was performed on overlapping genes. Furthermore, GSEA was performed to investigate the biological distinction between patients with high and low SFPQ expression [14]. Enrichment results satisfying \(P_{\text{adj}} \!< \!0.05\) and FDR Q value \(< \!0.25\) were considered statistically significant.

Construction of protein-protein interaction (PPI) network

To investigate SFPQ-related proteins, PPI networks associated with SFPQ have been independently constructed using the STRING database and GeneMANIA database. The software Cytoscape v3.6.1 with the MCODE plugin was utilized to visualize the PPI network and identify hub genes.

Immune infiltration analysis in HCC

The cell-type identification by estimating relative subsets of RNA transcripts (CIBERSORT) and single-sample GSEA (ssGSEA) algorithms from the GSVA package in R were utilized to analyze immune infiltration in HCC samples [15]. The Wilcoxon rank-sum test was then employed to investigate the correlation between immune cell levels and the expression of SFPQ.

Genetic alterations and DNA methylation analysis

We utilized the cbioPortal to investigate the SFPQ genetic alterations in TCGA, Firehose Legacy dataset [16]. Correlation analysis was performed using the MEXPRESS database to further investigate the association between SFPQ expression and DNA methylation [17, 18].

Single-cell sequencing analysis

Tumor Immune Single-cell Hub was utilized to delve into the heterogeneity of the tumor microenvironment (TME) and visualize the single-cell transcriptome of the TME [19]. Based on four datasets (LIHC_GSE140228_10X, LIHC_GSE140228_Smartseq2, GSE166635, and GSE98638), the expression of SFPQ across multiple cell types in the TME was investigated.

Construction of ceRNA network

The ENCORI website, as a database and web platform, offers various computational tools for analyzing the interactions between SFPQ and non-coding RNA, including microRNAs and long non-coding RNAs (lncRNAs). Cytoscape was utilized to construct lncRNA-miRNA-mRNA networks.

The generation of prognostic model

Univariate and multivariate Cox regression analyses were performed to determine whether SFPQ could function as an independent prognostic factor. The clinical parameters, including T stage and tumor status, were involved. In addition, the rms package and survival package constructed a nomogram and calibration plot for predicting 1-year, 3-year, and 5-year overall survival (OS).

Statistical analysis

Statistical analyses and graphics were performed and presented with R (version 3.6.3). The Wilcoxon rank-sum test was utilized to determine the statistical significance of SFPQ expression between normal and tumor tissues. Survival analysis was performed using the Kaplan-Meier method and the log-rank test, employing the median level of SFPQ expression as the cutoff value. Univariate and multivariate Cox regression, as well as prognostic nomogram models, were utilized to unveil the clinical significance of SFPQ in HCC.

Results

Elevated SFPQ expression in tumor samples compared to normal...
tissues
To investigate the variations in SFPQ expression between tumor and normal tissues, we initiated the study by evaluating the expression of SFPQ in the TCGA’s pan-cancer data utilizing the TIMER database (Figure 1). The expression of SFPQ was elevated in eleven cancer types compared to normal tissue, including breast invasive carcinoma, cholangiocarcinoma, colon adenocarcinoma, esophageal carcinoma, head and neck squamous cell carcinoma, liver hepatocellular carcinoma, lung squamous cell carcinoma, prostate adenocarcinoma, rectum adenocarcinoma, stomach adenocarcinoma, and uterine corpus endometrial carcinoma. Conversely, SFPQ expression is low in kidney chromophobe and thyroid carcinoma. This data indicates that the mRNA expression of SFPQ is abnormally expressed in different cancer types. Figure 1 illustrates a significant upregulation of SFPQ expression in HCC tissues compared to normal tissues ($P < 0.001$). Additionally, the upregulation of SFPQ expression was verified using three datasets, including GSE36376, GSE64041, and HCCDB18. The HPA database provides the immunohistochemistry results of SFPQ in HCC tissues and normal tissues, further validating the elevated protein level of SFPQ in HCC tissues. In addition, the HPA database also helps us compare the 20 genes with the highest significance associated with an unfavorable prognosis in liver cancer. As shown in Table 1, the expression level of SFPQ ranks fourth among the expression levels of 20 genes.

![Figure 1](https://example.com/figure1.png)

**Figure 1** SFPQ is overexpressed in HCC. (A) Expression of SFPQ in normal tissues vs pan-cancer samples. (B) Based on TCGA data, differential expression of unpaired difference analysis of SFPQ mRNA expression in HCC. (C–E) Verification of the expression of SFPQ in HCC relative to normal tissues in the GSE36376, GSE64041, and ICGC datasets. (F & G) Immunohistochemistry results of SFPQ in HCC tissues and normal tissues. HCC, hepatocellular carcinoma; SFPQ, splicing factor proline- and glutamine-rich; TCGA, The Cancer Genome Atlas; ICGC, International Cancer Genome Consortium.

Submit a manuscript: https://www.tmrjournals.com/mdm
Table 1 The 20 genes with highest significance associated with an unfavorable prognosis in liver cancer

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
<th>mRNA (cancer)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCT4</td>
<td>Chaperonin containing TCP1 subunit 4</td>
<td>40.9</td>
<td>9.37e-12</td>
</tr>
<tr>
<td>CCT5</td>
<td>Chaperonin containing TCP1 subunit 5</td>
<td>20.6</td>
<td>6.57e-11</td>
</tr>
<tr>
<td>KPNA2</td>
<td>Karyopherin subunit alpha 2</td>
<td>19.8</td>
<td>2.12e-10</td>
</tr>
<tr>
<td>SFPQ</td>
<td>Splicing factor proline and glutamine rich</td>
<td>18.7</td>
<td>7.91e-13</td>
</tr>
<tr>
<td>PSMD1</td>
<td>Proteasome 26S subunit, non-ATPase 1</td>
<td>17.6</td>
<td>8.63e-12</td>
</tr>
<tr>
<td>G6PD</td>
<td>Glucose-6-phosphate dehydrogenase</td>
<td>13.5</td>
<td>9.45e-13</td>
</tr>
<tr>
<td>YARS1</td>
<td>Tyrosyl-tRNA synthetase 1</td>
<td>9.30</td>
<td>6.97e-11</td>
</tr>
<tr>
<td>SNX5</td>
<td>Sorting nexin 5</td>
<td>8.30</td>
<td>2.30e-10</td>
</tr>
<tr>
<td>KDM1A</td>
<td>Lysine demethylase 1A</td>
<td>7.00</td>
<td>4.47e-12</td>
</tr>
<tr>
<td>GTPBP4</td>
<td>GTP binding protein 4</td>
<td>5.60</td>
<td>8.85e-12</td>
</tr>
<tr>
<td>INTS13</td>
<td>Integrator complex subunit 13</td>
<td>4.70</td>
<td>1.17e-10</td>
</tr>
<tr>
<td>CAD</td>
<td>Carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase</td>
<td>4.20</td>
<td>1.72e-11</td>
</tr>
<tr>
<td>CDC8</td>
<td>Cell division cycle associated 8</td>
<td>3.80</td>
<td>4.93e-10</td>
</tr>
<tr>
<td>TRMT6</td>
<td>tRNA methyltransferase 6 non-catalytic subunit</td>
<td>3.60</td>
<td>4.70e-12</td>
</tr>
<tr>
<td>HILPDA</td>
<td>Hypoxia inducible lipid droplet associated</td>
<td>3.20</td>
<td>4.30e-11</td>
</tr>
<tr>
<td>KIF20A</td>
<td>Kinesin family member 20A</td>
<td>2.90</td>
<td>2.14e-12</td>
</tr>
<tr>
<td>HDAC2</td>
<td>Histone deacetylase 2</td>
<td>2.60</td>
<td>5.39e-10</td>
</tr>
<tr>
<td>GTSE1</td>
<td>G2 and S-phase expressed 1</td>
<td>1.40</td>
<td>1.90e-10</td>
</tr>
<tr>
<td>TTK</td>
<td>TTK protein kinase</td>
<td>1.20</td>
<td>3.84e-10</td>
</tr>
<tr>
<td>RBM28</td>
<td>RNA binding motif protein 29</td>
<td>1.10</td>
<td>2.72e-11</td>
</tr>
</tbody>
</table>

Identification of co-expression genes and DEGs in HCC associated with SFPQ
We employed the LinkedOmics database to identify co-expression genes of SFPQ. Applying the criteria FDR < 0.05, P < 0.05, and |corporation| ≥ 0.3 as thresholds, we identified 3185 co-expression genes. We divided 419 HCC patients into high- and low-SFPQ expression groups based on the median SFPQ expression value. We identified 2036 DEGs (1625 up-regulated and 411 down-regulated) in the high- and low-SFPQ expression groups using a threshold of absolute log-fold change > 1 and P < 0.05 (Figure 2). 713 overlapping genes between DEGs and co-expression genes were identified. The overlapping genes were then subjected to enrichment analysis for further investigation.

Functional annotation of SFPQ-associated differentially expressed genes in HCC
Gene ontology and KEGG enrichment analyses were utilized to investigate the functional significance of the DEGs in high and low SFPQ expression groups (Figure 2). As for biological process, nuclear division, chromosome segregation, nuclear chromosome segregation, sister chromatid segregation, and mitotic sister chromatid segregation were the most significant enrichments, accompanied by the notable enrichments in the spindle, chromosomal region, condensed chromosome, and mitotic spindle the cellular component level. Molecular function included oxidoreductase activity, monoxygenase activity, microtubule motor activity, and steroid hydroxylase activity. Additionally, the KEGG pathway analysis indicated that SFPQ DEGs are involved in the cell cycle, complement, and coagulation cascades, metabolism of xenobiotics by cytochrome P450, and retinol metabolism. Subsequently, GSEA analysis revealed significant differences in multiple liver cancer-associated pathways between the high and low SFPQ expression groups (Figure 3).

Creating protein interaction network
The PPI network of the target protein was constructed using the STRING database and visualized through Cytoscape v3.6.1 software with the MCODE plugin to identify hub genes. As illustrated in Figure 4, the PPI network displayed 141 nodes and 5224 edges, with a score of 74.629. Furthermore, GeneMANIA was utilized to generate and visualize the PPI network. The GeneMANIA analysis revealed 20 genes that are associated with SFPQ (Figure 4), including NONO, CHD3, FSPC1, RAD51D, RAD51, KDM1A, SNRPA, YLPM1, HES1, RPA3, U2AF2, HNRNPM, MATR3, MKI67, TOP1, GSK3B, SND1, LMO7, PINK1, TNF. The functions of SFPQ and these genes are primarily associated with the regulation of the apoptotic signaling pathway, recombinational repair, negative regulation of the apoptotic signaling pathway, regulation of oxidative stress-induced intrinsic apoptotic signaling pathway, telomere organization, regulation of oxidative stress-induced cell death, and DNA recombination.

Association between SFPQ and immunological infiltration
CIBERSORT and ssGSEA were utilized to investigate immune infiltration of SFPQ expression. As shown in Figure 5, the stacked bar chart of 22 immune cell types was generated by CIBERSORT analysis. The analysis of immune infiltration in the samples revealed that the high SFPQ expression group exhibited significantly higher proportions of B cell memory, plasma cells, T cell CD4 memory activated, T cells follicular helper, macrophages M0, and neutrophils as determined by CIBERSORT analysis. Furthermore, ssGSEA analysis indicated that the high SFPQ expression group showed elevated levels of activated CD4 T cell, activated dendritic cell, effector memory CD4 T cell, natural
killer cell, plasmacytoid dendritic cell, T follicular helper, and type 2 T helper cell. These observations suggest that high SFPQ expression may be associated with increased T cell immune infiltration and activation, potentially implying a role for SFPQ in regulating T cell immune response.

Figure 2 Enrichment analysis of SFPQ in HCC. (A) co-expression genes of SFPQ. (B) DEGs volcano plot. The blue and red spots represent strongly down-regulated and up-regulated DEGs, respectively. (C) Overlapping genes between DEGs and co-expressed genes. (D–G) GO enrichment analysis and KEGG enrichment analysis. HCC, hepatocellular carcinoma; SFPQ, splicing factor proline- and glutamine-rich; DEGs, differentially expressed genes; KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, gene ontology; BP, biological process; CC, cellular component; MF, molecular function; PPAR, peroxisome proliferator-activated receptor; NAD, nicotinamide adenine dinucleotide; NADP, nicotinamide adenine dinucleotide phosphate.

Figure 3 Enrichment analysis of SFPQ in HCC. Gene set enrichment analysis of SFPQ in HCC. HCC, hepatocellular carcinoma; SFPQ, splicing factor proline- and glutamine-rich; FDR, false discovery rate; NES, normalized enrichment score.
Figure 4 Protein-protein interaction comprehensive analysis of SFPQ. (A) PPI network for SFPQ from STRING database. (B) Hub genes identified by MCODE plugin. (C) PPI network for SFPQ from GeneMANIA database. SFPQ, splicing factor proline- and glutamine-rich; PPI, protein-protein interaction; STRING, The Search Tool for the Retrieval of Interacting Genes; MCODE, Molecular Complex Detection.

Figure 5 Analysis of the relationship between SFPQ expression and immune infiltration in HCC. (A) Stacked bar chart of 22 immune cell types. (B & C) Differential enrichment scores of 22 immune cell types in SFPQ high/low expression groups as determined by CIBERSORT and ssGSEA analysis. HCC, hepatocellular carcinoma; SFPQ, splicing factor proline- and glutamine-rich; ssGSEA, single-sample gene set enrichment analysis; MDSC, myeloid-derived suppressor cell.
Relationship between immunoregulatory checkpoints and the expression of SFPQ

The TISIDB database investigated the association between SFPQ and immunoinhibitors, immunostimulators, and major histocompatibility complex molecules (Figure 6). In addition, the association between 4 immunoinhibitors (CTLA4, KDR, LAG3, PDCD1), 4 immunostimulators (CD80, ICOSLG, IL6R, TNFRSF9), and 4 major histocompatibility complex molecules (B2M, HLA-C, HLA-DOB, HLA-E) with SFPQ was further investigated. The findings revealed that the expression of CTLA4, LAG3, PDCD1, CD80, TNFRSF9, and HLA-DOB positively associated with SFPQ expression. On the contrary, KDR, ICOSLG, IL6R, B2M, HLA-C, and HLA-E exhibited an inverse association with SFPQ expression.

Figure 6 Relationship between SFPQ expression and immune checkpoints. (A) Correlation between immunoinhibitors and SFPQ expression. (B) Correlation between immunostimulator and SFPQ expression. (C) Correlation between MHC molecule and SFPQ expression. (D–G) Association between SFPQ and CTLA4, KDR, LAG3, and PDCD1. (H–K) Association between SFPQ and CD80, ICOSLG, IL6R, and TNFRSF9. (L–O) Association between SFPQ and B2M, HLA-C, HLA-DOB, HLA-E. SFPQ, splicing factor proline- and glutamine-rich; MHC, major histocompatibility complex.
Genetic alterations and DNA methylation analysis

The cbioPortal was utilized to investigate the genetic alterations of SFPQ. Genetic alterations of SFPQ occurred in 1.6% of the HCC patients, including mutations, amplification, and deep deletion. In addition, SFPQ was found to undergo missense mutation and frameshift mutation insertion, specifically including the substitution of isoleucine for methionine 579, threonine replacing lysine 421, and replacement of lysine for asparagine 511 at resulting in the appearance of a premature stop codon, causing early termination of protein synthesis. The MEXPRESS was utilized to investigate the correlation between SFPQ and DNA methylation. 4 CpG sites (cg24719193 (r = −0.264, P < 0.001), cg23207673 (r = −0.307, P < 0.001), cg24344221 (r = −0.108, P < 0.05), and cg20800296 (r = −0.129, P < 0.01)) showed a negative correlation with SFPQ expression (Figure 7).

Single-cell sequencing analysis

On the basis of four datasets (LIHC_GSE140228_10X, LIHC_GSE140228_Smartseq2, GSE146115 and GSE166635), the single-cell sequencing analysis showed that SFPQ was mainly expressed within Tprolif cells, CD8+ T cells (CD8T), B cells, Mono/Macro. The findings suggest that SFPQ may play a crucial role in TME of HCC (Figure 8).

Construction of ceRNA network

ENCORI was utilized to predict and analyze the interactions between miRNA, mRNA, and lncRNA, which are utilized to construct ceRNA networks. As shown in Figure 9, hsa-miR-101-3p expression exhibited a significant inverse association with SFPQ in HCC (P < 0.001). In addition, hsa-miR-101-3p expression was significantly increased in HCC tissues compared to adjacent normal tissues (P < 0.001). Furthermore, patients with high expression of hsa-miR-101-3p exhibited a worse prognosis than those with low expression (P < 0.001). These findings indicate that hsa-miR-101-3p may play a role in regulating HCC progression by targeting SFPQ.

ENCORI was utilized to identify lncRNAs that may potentially regulate hsa-miR-101-3p. The selection criteria for IncRNAs include a positive correlation with SFPQ and a negative correlation with hsa-miR-101-3p. Additionally, the IncRNA should be upregulated in HCC, with patients exhibiting high expression of this IncRNA showing survival rates. AC023043.4, AC124798.1, AC145207.5, and GSEC were found to be negatively correlated with hsa-miR-101-3p. Conversely, AC023043.4, AC124798.1, AC145207.5, and GSEC were positively correlated with SFPQ and highly expressed in HCC tissues. Moreover, patients with high expression of AC023043.4, AC124798.1, AC145207.5, and GSEC had a poorer prognosis than those with low expression (P < 0.001). Finally, as shown in Figure 10, the ceRNA network was constructed using Cytoscape.

Figure 7 Genetic features of SFPQ. (A) Genetic alterations of SFPQ. (B) Missense mutation and frameshift mutation insertion of SFPQ. (C) Genetic alteration frequency of SFPQ. (D) DNA methylation analysis of SFPQ. SFPQ, splicing factor proline- and glutamine-rich.
Figure 8 Expression of SFPQ in TME. (A) Heatmap of SFPQ expression in different clusters of cells. (B–I) Single-cell cluster map of SFPQ in four datasets. SFPQ, splicing factor proline- and glutamine-rich; TME, tumor microenvironment.
Figure 9 **hsa-miR-101-3p in HCC.** (A) Association between SFPQ expression and hsa-miR-101-3p expression. (B) The expression of hsa-miR-101-3p in HCC relative to normal tissues. (C) Kaplan-Meier survival curves of hsa-miR-101-3p in HCC. HCC, hepatocellular carcinoma; SFPQ, splicing factor proline- and glutamine-rich.

Figure 10 **Construction of ceRNA network.** (A–D) Association between AC023043.4, AC124798.1, AC145207.5, and GSEC expression and hsa-miR-101-3p expression. (E–H) Association between AC023043.4, AC124798.1, AC145207.5, and GSEC expression and SFPQ expression. (I–L) The expression of AC023043.4, AC124798.1, AC145207.5, and GSEC in HCC relative to normal tissues. (M–P) Kaplan-Meier survival curves of AC023043.4, AC124798.1, AC145207.5, and GSEC in HCC. (Q) ceRNA network constructed by Cytoscape. HCC, hepatocellular carcinoma; SFPQ, splicing factor proline- and glutamine-rich.
Predictive value of SFPQ for HCC prognosis

By using the Kaplan-Meier method, SFPQ expression and HCC prognosis were calculated. SFPQ-positive patients had a significantly worse prognosis than those with SFPQ-negative patients. As shown in Figure 11, the increased expression of SFPQ was associated with poor OS (P < 0.001). In addition, high expression of SFPQ was associated with T stages 1, 2, and 3 (P < 0.001), N0 (P < 0.001), M0 (P < 0.001), with tumor (P = 0.04), male (P < 0.001), age > 60 (P < 0.001), Histologic grade G2, G3 (P < 0.001) (Figure 11).

Univariate Cox regression was conducted to identify factors influencing OS in HCC patients (Table 2). The analysis revealed that T stage (T1 & T2 VS. T3 & T4, P < 0.001), M stage (M0 VS. M1, P = 0.001), tumor status (tumor free VS. with tumor, P < 0.001), and SFPQ (low VS. high, P = 0.02) were predictive factors for worse OS. In the subsequent multivariate Cox regression, T stage, M stage, tumor status, and SFPQ were included. These findings indicated that T stage (P = 0.001), tumor status (P < 0.001), and SFPQ (P = 0.003) were independent predictors of overall survival in HCC patients.

Prognostic model of SFPQ in HCC

The nomogram model was constructed to provide prognostic predictions for patients with HCC. The nomogram included T stage, tumor status, and SFPQ. Besides, we evaluated the nomogram’s prediction efficacy using time-dependent receiver operating characteristic and calibration curves. The 1-year, 3-year, and 5-year areas under the curves were correspondingly 0.743, 0.65, and 0.604, respectively (Figure 12).

Prediction of small molecule drugs

CTD database was utilized to screen for the correlation between SFPQ and potential drugs. As shown in Figure 13, 13 chemicals exhibited a significant inverse association with SFPQ expression, and 12 chemicals exhibited a significant positive association with SFPQ expression. In addition, the reference count of selected chemicals was visualized by R package ggplot2. PubChem and partcommunity (https://b2b.partcommunity.com/community/) displayed the 3d structures of the top four chemicals (bisphenol A, nanotubes carbon, ethyl estradiol, and acrolein) with the highest reference count.

![Figure 11 Kaplan-Meier survival curves.](image1)

(A) Kaplan-Meier analysis comparing OS, (B–H) Curves of Kaplan-Meier for HCC patients with T stage 1, 2, and 3, N0, M0, with tumor, male, age > 60, and histologic grade G2, G3. HCC, hepatocellular carcinoma; OS, overall survival.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total (N)</th>
<th>Univariate analysis</th>
<th>Multivariate analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Hazard ratio (95% CI)</td>
<td>P value</td>
</tr>
<tr>
<td>T stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1 &amp; T2</td>
<td>416</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>T3 &amp; T4</td>
<td>308</td>
<td>2.323 (1.699–3.175)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>N stage</td>
<td>108</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>N0</td>
<td>286</td>
<td>1.975 (0.624–6.253)</td>
<td>0.247</td>
</tr>
<tr>
<td>M stage</td>
<td>281</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>M0</td>
<td>5</td>
<td>1.175 (0.624–2.253)</td>
<td>0.65</td>
</tr>
<tr>
<td>M1</td>
<td>302</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Tumor status</td>
<td>298</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Tumor free</td>
<td>4</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>With tumor</td>
<td>324</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>178</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>146</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>418</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>277</td>
<td>Reference</td>
<td></td>
</tr>
</tbody>
</table>

HCC, hepatocellular carcinoma; CI, confidence interval.

Submit a manuscript: https://www.tmrjournals.com/mdm
Table 2 Univariate Cox regression and multivariate Cox regression in HCC (continued)

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total (N)</th>
<th>Univariate analysis</th>
<th>Multivariate analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Hazard ratio (95% CI)</td>
<td>P value</td>
</tr>
<tr>
<td>&gt; 60</td>
<td>228</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>≤ 60</td>
<td>190</td>
<td>0.850 (0.623–1.161)</td>
<td>0.307</td>
</tr>
<tr>
<td>BMI</td>
<td>371</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>≤ 25</td>
<td>193</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 25</td>
<td>178</td>
<td>0.874 (0.625–1.224)</td>
<td>0.434</td>
</tr>
<tr>
<td>Histologic grade</td>
<td>410</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>G1 &amp; G2</td>
<td>263</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G3 &amp; G4</td>
<td>147</td>
<td>1.134 (0.822–1.564)</td>
<td>0.445</td>
</tr>
<tr>
<td>Race</td>
<td>404</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>163</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>White &amp; Black or African American</td>
<td>241</td>
<td>1.392 (0.989–1.960)</td>
<td>0.058</td>
</tr>
<tr>
<td>Albumin</td>
<td>341</td>
<td>≤ 3.5</td>
<td>0.874 (0.625–1.224)</td>
</tr>
<tr>
<td>OS event</td>
<td>416</td>
<td>0.735 (0.494–1.094)</td>
<td>0.129</td>
</tr>
<tr>
<td>Alive</td>
<td>252</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Dead</td>
<td>164</td>
<td>1508803158.124 (0.000–Inf)</td>
<td>0.992</td>
</tr>
<tr>
<td>SFPQ</td>
<td>418</td>
<td>high</td>
<td>0.695 (0.511–0.945)</td>
</tr>
<tr>
<td>low</td>
<td>212</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

HCC, hepatocellular carcinoma; BMI, body mass index; SFPQ, splicing factor proline- and glutamine-rich; CI, confidence interval; OS, overall survival.

Figure 12 Model for predicting the prognosis of SFPQ in LIHC. (A) Nomogram for estimating the 1-, 3-, and 5-year OS probabilities for LIHC. (B) Calibration plot illustrates the nomogram prediction of 1-, 3-, and 5-year total LIHC patient survival rates. (C) Time-dependent ROC curves for the prediction of 1-, 3-, and 5-year OS survival. SFPQ, splicing factor proline- and glutamine-rich; LIHC, liver hepatocellular carcinoma; OS, overall survival; ROC, receiver operating characteristic; AUC, area under the curve.
Discussion

HCC is one of the most frequent types of primary liver cancer [20]. HCC is asymptomatic until cancer has reached an advanced stage, making early diagnosis through monitoring crucial for minimizing HCC-related mortality [21].

In this research, the expression of SFPQ in various cancers was compared. SFPQ was found to be considerably up-regulated in eleven cancers, including breast invasive carcinoma, cholangiocarcinoma, colon adenocarcinoma, esophageal carcinoma, head and neck squamous cell carcinoma, liver hepatocellular carcinoma, lung squamous cell carcinoma, prostate adenocarcinoma, rectum adenocarcinoma, stomach adenocarcinoma, and uterine corpus endometrial carcinoma. Based on GSEA, the research delved into the functional implications of SFPQ and the potential mechanisms underpinning its influence on HCC development and metastasis. The study unveiled that pathways highly enriched in the group with the high SFPQ expression were associated with HCC-related pathways.

Gibersort and ssGSEA analyses showed that elevated SFPQ expression was significantly correlated with T cell CD4 memory activated, T cell follicular helper, and type 2 T helper cell. In single-cell analysis, SFPQ was highly expressed in Tprolf and CDRT cell clusters. These observations suggest a potential association between elevated SFPQ expression and enhanced T cell immune infiltration and activation, indicating a potential role for SFPQ in regulating T cell immune response.

Prior studies have indicated that miRNAs and IncRNAs regulate cancer progression through multiple mechanisms [22–24]. We utilized ENCORI to identify miRNAs and IncRNAs that potentially control HCC by regulating SFPQ. Hsa-miR-101-3p was downregulated in HCC tissues, negatively correlated with SFPQ expression, and associated with a worse prognosis in HCC patients who exhibited low hsa-miR-101-3p expression. Prior studies have shown that miR-101-3p inhibits Beclin-1-mediated autophagy and determines the sensitivity of HCC cells to oxaliplatin [25]. Hepatitis B virus has been reported to downregulate miR-101-3p expression and promote HCC cell proliferation and migration by targeting Rab5a [26].

Additionally, IncRNAs have been identified to reduce miRNA regulatory effects on target mRNAs [27]. We identified four IncRNAs, including AC023043.4, AC124798.1, AC145207.5, and GSEC, with the assistance of ENCORI. GSEC has been shown to promote the progression of triple negative breast cancer [28], lung adenocarcinoma [29], colon cancer [30], osteosarcoma [31], and HCC [32, 33]. The GSEC/hsa-miR-101-3p/SFPQ axis was detected as a putative regulatory pathway in HCC. Further studies are necessary to confirm the role of this potential regulatory pathway in HCC.

We identified a correlation between SFPQ expression and OS using Kaplan-Meier survival analysis. Elevated SFPQ expression was associated with a poorer prognosis in patients with T stage 1, 2, and 3, N0, M0, with tumor, male, age > 60, histologic grade G2, G3. In addition, the prognostic value of SFPQ in HCC patients was investigated. Through multivariate Cox regression analysis, T stage, tumor status, and high SFPQ expression were the independent predictive variables for OS deterioration. To predict the 1-year, 3-year, and 5-year OS of HCC, the nomogram prognostic model based on SFPQ expression level, T stage, and tumor status was further developed. The nomogram’s predicting capacity was presented using time-dependent receiver operating characteristic curves and calibration plots.

This study offers a comprehensive explanation of the significance of SFPQ in HCC. The ceRNA network identified in this study sheds light on the complex regulatory interactions among SFPQ and its downstream targets, unveiling a novel layer of understanding in the molecular pathways underlying cancer pathogenesis. The implications of this study extend beyond the individual role of SFPQ to reveal a broader network of regulatory relationships that contribute to the multifaceted nature of cancer biology.

Limitations

Our limitation of the study is the reliance on public databases without experimental validation. The primary reliance on existing datasets and bioinformatics analyses may introduce inherent limitations, as the findings are based on computational predictions and not experimental confirmation. Future research efforts should consider complementing bioinformatics analyses with wet experiments to validate the findings and strengthen the conclusions drawn from the study.

Conclusions

SFPQ is speculated to be a possible indicator of poor prognosis in HCC.
This investigation illuminates SFPQ as a potential treatment target for HCC.

References


