Comprehensive assessment of cleavage and polyadenylation specificity factors in hepatocellular carcinoma: Expression, prognostic significance and immune infiltration analysis

YUXIANG LU[1*], TING WANG[1*], XIULI YAN[2] and HUI ZHANG[1]

[1]Institute of Interdisciplinary Integrative Medicine Research, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, P.R. China; [2]Yueyang Hospital of Integrated Traditional Chinese and Western Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai 200437, P.R. China

Received December 6, 2024; Accepted April 9, 2025

DOI: 10.3892/mco.2025.2855

Abstract. Hepatocellular carcinoma (HCC), a prevalent and highly malignant form of liver cancer, poses significant global health challenges. Previous studies have suggested that alterations in cleavage and polyadenylation specificity factors (CPSFs) play a role in the development and prognosis of HCC. Despite these insights, a thorough evaluation of CPSFs' expression levels, prognostic value and association with immune infiltration in HCC is lacking. To address this gap, the present study conducted a systematic analysis leveraging multiple bioinformatics databases to elucidate the functions of CPSFs in HCC. To comprehensively investigate the role of CPSFs in HCC, a diverse array of bioinformatics tools and publicly accessible datasets were utilized. The present study investigated the gene expression patterns, clinicopathological correlations, and diagnostic and prognostic capabilities of CPSFs. Furthermore, genetic variations, co-expression networks and the role of CPSFs in immune cell infiltration and tumor-related pathways were examined. To elucidate the biological functions of CPSF-associated genes, Gene Ontology and Kyoto Encyclopedia of Genes and Genomes enrichment analyses were integrated. For experimental validation, reverse transcription-quantitative polymerase chain reaction was used to assess gene expression and the Cell Counting Kit-8 assay was utilized to evaluate the effects of CPSFs on HCC cell proliferation. Our analysis offers valuable insights into the molecular mechanisms through which CPSFs contribute to HCC progression. The current findings suggest that CPSFs, particularly CPSF1, CPSF3, CPSF4 and CPSF6, exhibit significant transcriptional upregulation in HCC, with their overexpression closely tied to advanced tumor progression. These CPSFs showed diagnostic and prognostic significance in HCC. Additionally, CPSF expression was associated with immune cell infiltration and activation status. Functional enrichment analysis indicated that CPSF1, CPSF3, CPSF4, CPSF6 and CPSF7 are involved in cancer-related signaling pathways, highlighting their role in tumor immune modulation. Experimental validation demonstrated that the expression of CPSF3 and CPSF7 was notably greater in the HCC cell lines than in the normal liver cells. Knockdown of CPSF3 and CPSF7 inhibited HCC cell proliferation, suggesting their potential oncogenic roles. This research offers an in-depth evaluation of the expression patterns, prognostic relevance and immune modulation-related functions of CPSFs in HCC. The observed upregulation of CPSFs in HCC, coupled with their association with poor clinical outcomes and immune system activation, highlights their potential as prognostic indicators. Nonetheless, additional experimental studies are needed to fully elucidate the molecular mechanisms and clinical significance of CPSFs in HCC.

Correspondence to: Dr Hui Zhang, Institute of Interdisciplinary Integrative Medicine Research, Shanghai University of Traditional Chinese Medicine, 1200 Cailun Road, Pudong New Area, Shanghai 201203, P.R. China
E-mail: zhanghuiman@126.com

Dr Xiuli Yan, Yueyang Hospital of Integrated Traditional Chinese and Western Medicine, Shanghai University of Traditional Chinese Medicine, 110 Ganhe Road, Hongkou, Shanghai 200437, P.R. China
E-mail: yxl9999@126.com

*Contributed equally

Key words: hepatocellular carcinoma, cleavage and polyadenylation specificity factors, prognosis, diagnosis, biomarker

Introduction

Hepatocellular carcinoma (HCC) is recognized as the sixth most common malignancy worldwide and is the fourth leading cause of cancer-related death (1). In the United States, HCC is particularly concerning, men and women with HCC account for ~6 and 4% of all cancer-related deaths, respectively, reflecting a significant disease burden (2). The liver plays a vital role in maintaining physiological homeostasis by performing critical functions (3). These processes are heavily dependent on proteins encoded by hepatocyte-specific genes, which are regulated primarily by liver-enriched transcription factors. Disruption of the regulation of these transcription factors is

closely associated with the pathogenesis of HCC (4). Research has demonstrated that identifying specific gene markers in HCC provides valuable insights for predicting patient outcomes (5). Given the well-established genetic alterations within HCC cells, there is an urgent need to discover novel biomarkers that are strongly linked to the diagnosis, treatment and prognosis of HCC. Such advancements could significantly improve clinical outcomes and enhance the management of HCC. Recently, post-transcriptional modifications have gained considerable attention for their roles in cancer progression. Among these modifications, the cleavage and polyadenylation (CPA) of pre-mRNA are crucial processes that affect mRNA stability, localization and translational efficiency (6). Cleavage and polyadenylation specificity factors (CPSFs) are key components of the CPA machinery, and their dysregulation has been implicated in various malignancies, including HCC. The CPSF complex consists of seven subunits: CPSF1, CPSF2, CPSF3, CPSF4, CPSF5 (NUDT21), CPSF6 and CPSF7. CPSF1 constitutes the largest subunit of the CPSF complex (Murthy and Manley, 1995). Each subunit can also produce different proteins. For example, there are four splice variants named CPSF1-207, CPSF1-210, CPSF1-212 and CPSF1-213 in the CPSF1 subunit, whereas in the CPSF2 subunit, there are three splice variants named CPSF2-201, CPSF2-202 and CPSF2-210.

Despite the known involvement of CPSFs in cancer, a comprehensive assessment of their expression patterns, prognostic significance and roles in immune infiltration within the context of HCC remains unexplored. Understanding the expression and functional relevance of CPSFs could unveil novel molecular targets and prognostic markers, thereby aiding in the development of more effective therapeutic strategies for HCC. Our study focused on a comprehensive evaluation of CPSF expression in HCC tissues relative to that in normal liver tissues. It was aimed to explore their prognostic value and examine their relationship with immune cell infiltration. By leveraging data from various high-throughput platforms and employing advanced bioinformatics approaches, it was aimed to uncover the diverse roles of CPSFs in HCC progression and assess their potential impact on patient prognosis and immune regulation.

Materials and methods

Bioinformatic tools. To comprehensively analyze the role of CPSFs in HCC, multiple bioinformatics platforms and databases that facilitate gene expression analysis, immune correlation, prognostic evaluation and pathway enrichment were employed. Each tool contributed distinct insights into CPSF expression, co-expression patterns, immune cell infiltration, and patient outcomes.

The cBioPortal platform (http://www.cbioportal.org/) was utilized to investigate genetic alterations and identify co-expressed genes associated with CPSFs (7). UALCAN (http://ualcan.path.uab.edu) was used to assess the relationships between CPSF expression and clinicopathological characteristics, as well as gene knockout data, with statistical significance established at $P<0.05$ (8). Data derived from HCCDB15 (http://lifeome.net/database/hccdb/home.html), the same data sourced from The Cancer Genome Atlas (TCGA)-liver hepatocellular carcinoma (LIHC), was employed to support the validation of CPSF expression levels and to perform diagnostic Receiver Operating Characteristic (ROC) curve analyses (9).

In terms of pathway and drug sensitivity analysis, GSCALite (http://bioinfo.life.hust.edu.cn/web/GSCALite/) provided functional elucidation of the role of CPSFs in cancer-associated pathways (10). Furthermore, protein expression levels across normal liver and HCC tissues were evaluated utilizing immunohistochemical data obtained from the Human Protein Atlas (HPA; https://www.proteinatlas.org) (11).

The prognostic impact was determined through Kaplan-Meier Plotter (http://kmplot.com/analysis/), which analyzed overall survival (OS) based on CPSF expression levels (12). To further corroborate the survival trends and differences in mRNA expression between the tumor and normal tissues, GEPIA2 (http://gepia.cancer-pku.cn/index.html) was used, whose data was based on TCGA-LIHC (TCGA, https://portal.gdc.cancer.gov/) (13).

Immune cell infiltration and its association with CPSF expression were examined through TIMER (https://cistrome.shinyapps.io/timer/) (14). The differential expression patterns of CPSFs in the normal, tumor, and metastatic tissues were additionally validated through the use of TNMplot (https://tnmplot.com/analysis/) (15). TISIDB (http://cis.hku.hk/TISIDB/) provided additional immune-related analyses integrating multiple datasets (16).

To identify co-expression networks and functional pathways, FunRich (17), WebGestalt (http://www.webgestalt.org) (18) and LinkedOmics (http://www.linkedomics.org) (19) were applied, enabling enrichment analysis and pathway mapping.

Through these tools, a comprehensive understanding of CPSF expression patterns, their correlations with immune infiltration and their prognostic potential were achieved, offering valuable insights into their roles in HCC progression and treatment strategies.

Analysis of the expression levels of the CPSF complex. Our first step was to understand the expression of CPSFs in both normal liver tissues and HCC tissues to guide our experimental recommendations. First, sample data were collected from the UALCAN database to test our hypothesis. To ensure the reliability of the present findings, the expression of CPSF family members in HCC was also validated through cBioPortal, another database containing genomic data from a variety of cancers. Immunohistochemical images of CPSF proteins in HCC were obtained through the HPA database.

Relationship between tumor markers and the CPSF complex. TNMplot was used to detect the relationship between the CPSF family and tumor markers such as MKI67, proliferating cell nuclear antigen (PCNA) and SRC to determine if the CPSF family can be a new marker of HCC.

Investigation of the expression levels of the CPSF complex and its correlation with clinical features in HCC. In the present study, the correlation between the mRNA expression levels of CPSF1 to CPSF7 and a range of clinical attributes associated with HCC were explored. Using platforms such as UALCAN and TNMplot, it was aimed to gain a deeper understanding of

how these CPSF complex members interact with key clinical factors in patients with HCC, potentially revealing their relevance to disease progression and prognosis.

Prognostic significance of CPSF complex genes in HCC. In the present study, the prognostic implications of CPSF complex genes in HCC were assessed by utilizing tools such as the Kaplan-Meier Plotter and GEPIA. Pertinent data were collected and these tools were employed to produce OS and disease-free survival (DFS) curves utilizing the Kaplan-Meier approach. Moreover, hazard ratios (HR) accompanied by 95% confidence intervals and log-rank P-values were computed to further assess the potential prognostic significance of these genes (20). Alpha-Fetoprotein (AFP) was also incorporated into the survival analysis for both OS and DFS to compare it with our CPSF survival analysis. This comparison aimed to validate the potential of CPSF as a candidate biomarker and therapeutic target for liver cancer. In tumor survival analysis, an HR greater than 1 is considered to be associated with poor prognosis. By evaluating CPSF alongside AFP, it was sought to strengthen the evidence supporting the role of CPSFs in liver cancer prognosis and its possible therapeutic implications.

Immune cell infiltration and its association with CPSF complex members in HCC. In the subsequent phase of our research, focus was addressed on five key members of the CPSF family, which were selected on the basis of preliminary data highlighting a strong correlation between their expression and HCC. It was aimed to elucidate how the expression levels of these CPSF genes correlate with immune cell infiltration in HCC. Through the application of TIMER, it was identified that CPSF1, CPSF3, CPSF4, CPSF6 and CPSF7 displayed notable associations with the presence of six distinct immune infiltrate cell types (B cells, $CD4^+$ T cells, $CD8^+$ T cells, neutrophils, macrophages and dendritic cells). Moreover, the TISIDB database was utilized to corroborate the intricate link between these components of the CPSF complex and the activation state of immune cells within the tumor microenvironment (TME) of HCC.

Pathway and drug sensitivity assessment of CPSF. GSCALite was used to perform a simple pathway analysis on selected members of the CPSF complex, followed by Spearman correlation analysis to investigate their relationship with drug sensitivity. Drug sensitivity information was obtained from two primary databases: The Genomics of Drug Sensitivity in Cancer (https://www.cancerrxgene.org/) and the Cancer Therapeutics Response Portal (https://portals.broadinstitute.org/ctrp/).

Comprehensive analysis of CPSF gene co-expression and pathway enrichment. LinkedOmics (http://linkedomics.org/login.php) serves as an extensive online resource offering multi-dimensional data encompassing 32 cancer types sourced from TCGA. To investigate the co-expression relationships of the CPSF genes, Spearman correlation analysis was conducted utilizing the 'LinkFinder' feature of LinkedOmics in conjunction with the cBioPortal platform. Our analysis then narrowed down the top 200 co-expressed genes that were consistently identified across both LinkedOmics and cBioPortal. Next, a deeper functional analysis was executed through Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment, employing the WebGestalt and FunRich tools. Furthermore, the STRING database (https://string-db.org/) was utilized to examine the interactions and correlations within the CPSF gene network, with the findings visualized using Cytoscape (https://cytoscape.org/).

Analysis of gene knockout data in HCC cell lines. The knockout database (DepMap built-in UALCAN) can be used for molecular mechanistic studies of disease. Knockout technology allows the nature of a specific gene to be determined and its effect on the body to be studied. Gene knockout data were collected for selected CPSF family members in liver cancer cell lines, which are expected to have important functions in influencing the process of liver cancer according to previous research. If the growth of this cancer is suppressed by the loss of CPSF-related genes, these genes may play a critical role in the development of HCC.

Cell culture. Human liver cancer cell lines, which included HUH7, JHH7 and PLC/PRF/5, were procured from the Chinese Academy of Sciences Cell Bank (Shanghai, China). The THLE-2 human liver epithelial cell line was procured from Procell Life Science & Technology Co., Ltd. (Wuhan, China). These cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin (all from Gibco; Thermo Fisher Scientific, Inc.) in a humidified incubator at 37˚C and 5% CO_2.

RNA extraction and reverse transcription-quantitative PCR (RT-qPCR). Total RNA was extracted from the cultured cells with TRIzol reagent (Invitrogen; Thermo Fisher Scientific, Inc.). Complementary DNA (cDNA) synthesis was carried out using a reverse transcription kit (Applied Biosystems; Thermo Fisher Scientific, Inc.) according to the manufacturer's protocol. The diluted cDNA was then analyzed by qPCR using SYBR® Green Realtime PCR Master Mix (Toyobo Life Science). The 10-μl reaction mixture consisted of 5 μl SYBR Green Master Mix, 0.5 μl each of forward and reverse primers (10 μM), and 4 μl diluted cDNA. RT-qPCR was performed under the following thermocycling conditions: Initial denaturation at 95˚C for 30 sec, followed by 40 cycles of 95˚C for 5 sec, 56˚C for 15 sec, and 72˚C for 10 sec. The target gene expression levels were normalized to the GAPDH mRNA levels. Relative expression quantification was performed using the $2^{-\Delta\Delta Cq}$ method (21). The primer sequences utilized in the present study are listed in Table I.

Cell proliferation assay. HUH7 cells were seeded into 96-well culture plates at a density of $1x10^4$ cells per well and incubated in the aforementioned DMEM under conditions of 37˚C and 5% CO_2 for 24 h, after which the cells were allowed to adhere. The adhered cells were subsequently transfected with negative control small interfering RNA (siRNA), CPSF3-siRNA, CPSF6-siRNA, or CPSF7-siRNA (Shanghai GenePharma Co., Ltd.) at a concentration of 50 nM using the Lipofectamine 3000 reagent (Invitrogen; Thermo Fisher Scientific, Inc.) according to the manufacturer's instruc-

Table I. Primer sequences of genes.

Gene name	Primer sequence (5'-3')
CPSF1	F: CGCAGCTCTACGTGTACCG
	R: GGACATGACGTTGCCAAAGAA
CPSF2	F: GGAAGCATGTTCACCAGATTGA
	R: CTTTCCGACAGCATACGGGA
CPSF3	F: ATGTCTGCGATTCCTGCTGAG
	R: ATCCCACAGTCGAGCATTATTTT
CPSF4	F: TGTCCGTTTCGCCACATCAG
	R: TTGGTCATGTCATACTCATGCAG
CPSF5	F: GGTCACTCAGTTCGGCAACAA
	R: CTCATGCGCTGAAATCTGGC
CPSF6	F: GGCGTGGACCACATAGACATT
	R: CCATGTAATCTCGGTCTTCTGGG
CPSF7	F: ACAACAAGACCCCTGCAATTC
	R: ACTCCACCACATCATAGACTCC
GAPDH	F: TGCACCACCAACTGCTTAGC
	R: GGCATGGACTGTGGTCATGAG

F, forward; R, reverse; CPSF, cleavage and polyadenylation specificity factor.

tions. Following transfection, cells were cultured for 24, 48, or 72 h under standard conditions. At predetermined time points, the original medium was removed, and 20 μl of Cell Counting Kit-8 (CCK-8; ApexBio Technology LLC) reagent was added to each well, ensuring that no air bubbles formed. The cells were then incubated with a CCK-8-containing medium for 4 h, after which the color development of the culture was observed. The absorbance value at 450 nm (A value) was measured for each well using an enzyme-linked immunosorbent assay (ELISA) reader. The cell survival rate (%) (indicative of cell viability) was calculated using the following formula: [(A value of the experimental group-A value of the blank group)/(A value of the control group-A value of the blank group)] x100%, with the survival rate reflecting the proliferation status of the cells in each group. Sequences of the siRNAs used is in Table II.

Statistical analysis. All bioinformatics analyses were conducted using the corresponding database websites. Statistical evaluations of the experimental data were performed using GraphPad Prism 8.0 software (GraphPad Software Inc.; Dotmatics). The results are presented as the mean ± standard deviation (SD). For comparisons among multiple groups, pairwise Student's t-tests were conducted to determine statistically significant differences. $P<0.05$ was considered to indicate a statistically significant difference. ROC curve analysis was conducted to assess the potential of CPSF members as diagnostic biomarkers for patients with HCC via data from TCGA and/or the Genotype-Tissue Expression project. The raw mRNA expression data for CPSFs were obtained from UCSC XENA (https://xenabrowser.net/datapages/). All the data were subsequently $\log_2$ (x+1) transformed.

Table II. Sense and antisense sequences of siRNA.

siRNA	Sequence (5'-3')
Control	Sense: UUCUCCGAACGUGUCACGU
	Antisense: ACGTGACACGTTCGGAGAA
CPSF3	Sense: GCAGACGACAUGCUGUAUA
	Antisense: UAUACAGCAUGUCGUCUGC
CPSF6	Sense: GGUGUUGGAUCUGAAGCAU
	Antisense: AUGCUUCAGAUCCAACACC
CPSF7	Sense: CAGUGGCCUGCGUAAUAGA
	Antisense: UCUAUUACGCAGGCCACUGTA

siRNA, small interfering RNA; CPSF, cleavage and polyadenylation specificity factor.

Results

Expression levels of the CPSF complex members in patients with HCC. To explore the expression levels of the CPSF complex members in patients with HCC, the present study utilized the GEPIA2 database to compare the mRNA levels of CPSF complex members in HCC tissues and normal liver tissues. The analysis revealed the significant upregulation of CPSF1, CPSF3, CPSF4 and CPSF6 expression levels in HCC tissues compared with normal liver tissues ($P<0.01$) (Fig. 1A).

To ensure the validity of our findings, the data were cross-referenced with the cBioPortal database (RNA Seq V2) to examine the genetic alterations of CPSFs in HCC tissues (Fig. 1B-D). The analysis revealed that CPSF1, CPSF3 and CPSF4 were markedly overexpressed in HCC samples, whereas CPSF5, CPSF6 and CPSF7 exhibited variable expression levels, and CPSF2 was mostly downregulated. The percentages of the genetic changes for CPSF1, CPSF2, CPSF3, CPSF4, CPSF6 and CPSF7 were 32, 11, 10, 9, 5 and 6%, respectively. Notably, more than 50% of the patients with HCC presented genetic alterations in all CPSF genes.

Next, the immunohistochemical staining patterns of CPSF proteins in HCC and normal liver tissues were examined by querying the HPA database. According to the HPA data, the CPSF1 and CPSF4 proteins were undetectable in HCC tissues. Compared with that in normal liver tissue, the expression of the CPSF2 protein in HCC tissues was robust. The CPSF3 protein was present at moderate levels in both normal liver and HCC tissues. However, both the CPSF6 and CPSF7 proteins slightly increased in HCC tissues (Fig. 1E).

Association between CPSF expression and clinicopathological features in HCC. UALCAN database was employed to investigate the associations between the mRNA expression levels of individual CPSF1 to CPSF7 members and various clinicopathological features in HCC, including tumor grade, tumor stage and nodal metastasis. The present analysis revealed a clear trend of increasing CPSF expression levels with increasing cancer grade and stage across all CPSF family members (Fig. 2A and B). Additionally, when nodal metastasis was examined, a corresponding increase in CPSF expression with the presence of metastasis in the lymph nodes

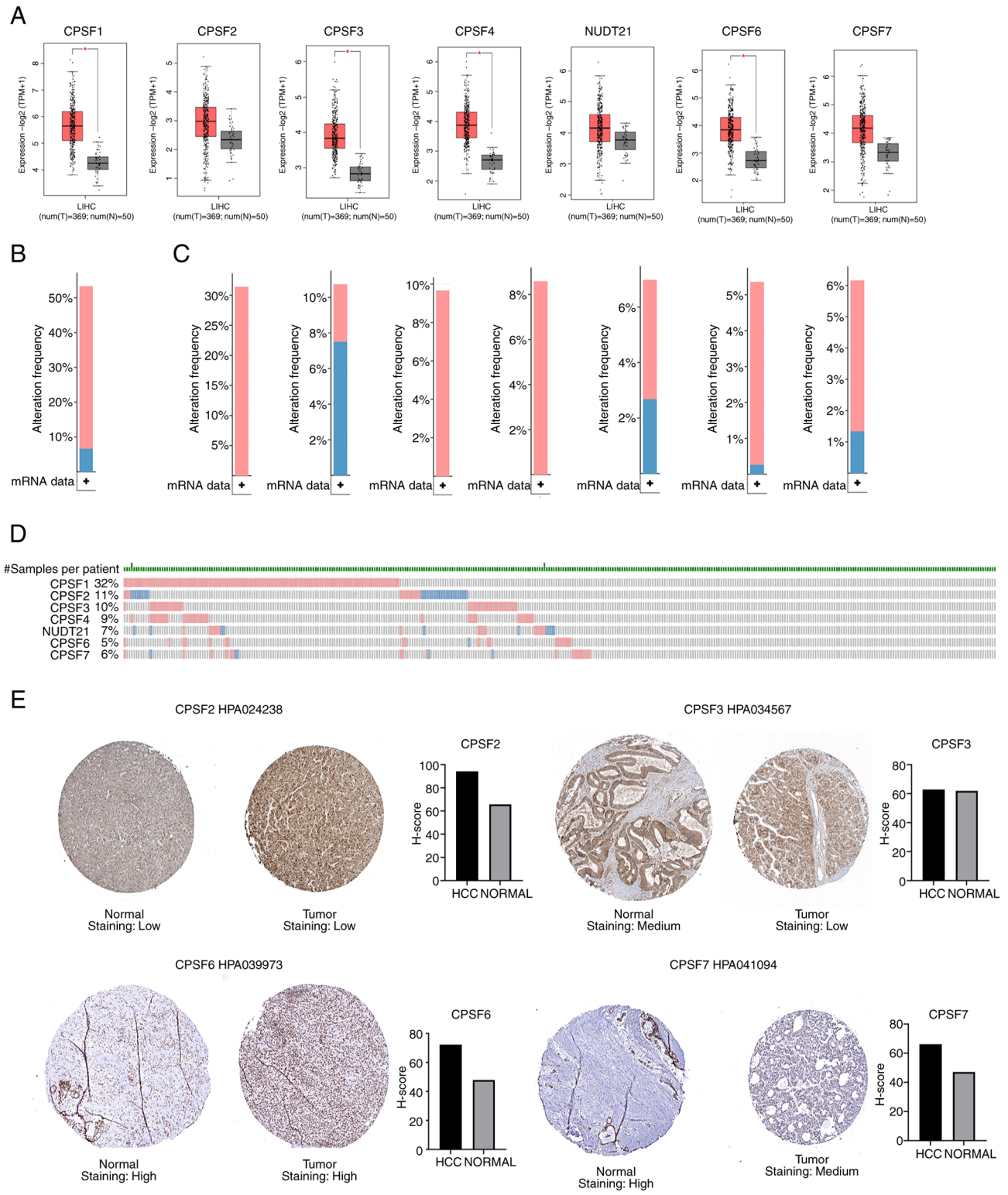

Figure 1. Expression of CPSF complex members at the transcriptional level in patients with HCC. (A) mRNA expression of CPSF complex members in HCC tissues (GEPIA2). (B) Expression of all CPSFs in tissues of patients with HCC . Red represents high expression, and blue represents low expression. (C and D) Expression of CPSFs in tissues of patients with HCC. In panel C, from left to right are CPSF1, CPSF2, CPSF3, CPSF4, NUDT21 (CPSF5), CPSF6 and CPSF7. (E) Representative images of immunohistochemical staining for CPSF proteins in the HPA database. The left image shows normal liver tissue, and the right image shows HCC tissue. Bar plot showing the H-scores (Data S2) of the corresponding immunohistochemically stained images. CPSF, cleavage and polyadenylation specificity factor; HCC, hepatocellular carcinoma; HPA, human protein atlas.

was observed for specific CPSF complex members (CPSF1, CPSF4, NUDT21, CPSF6 and CPSF7, as other CPSFs were not available in the TNMplot database) (Fig. 2C).

Furthermore, TCGA data was leveraged to generate ROC curves (Fig. 2D) to assess the diagnostic value of CPSFs in distinguishing HCC tissues (n=351) from non-HCC tissues

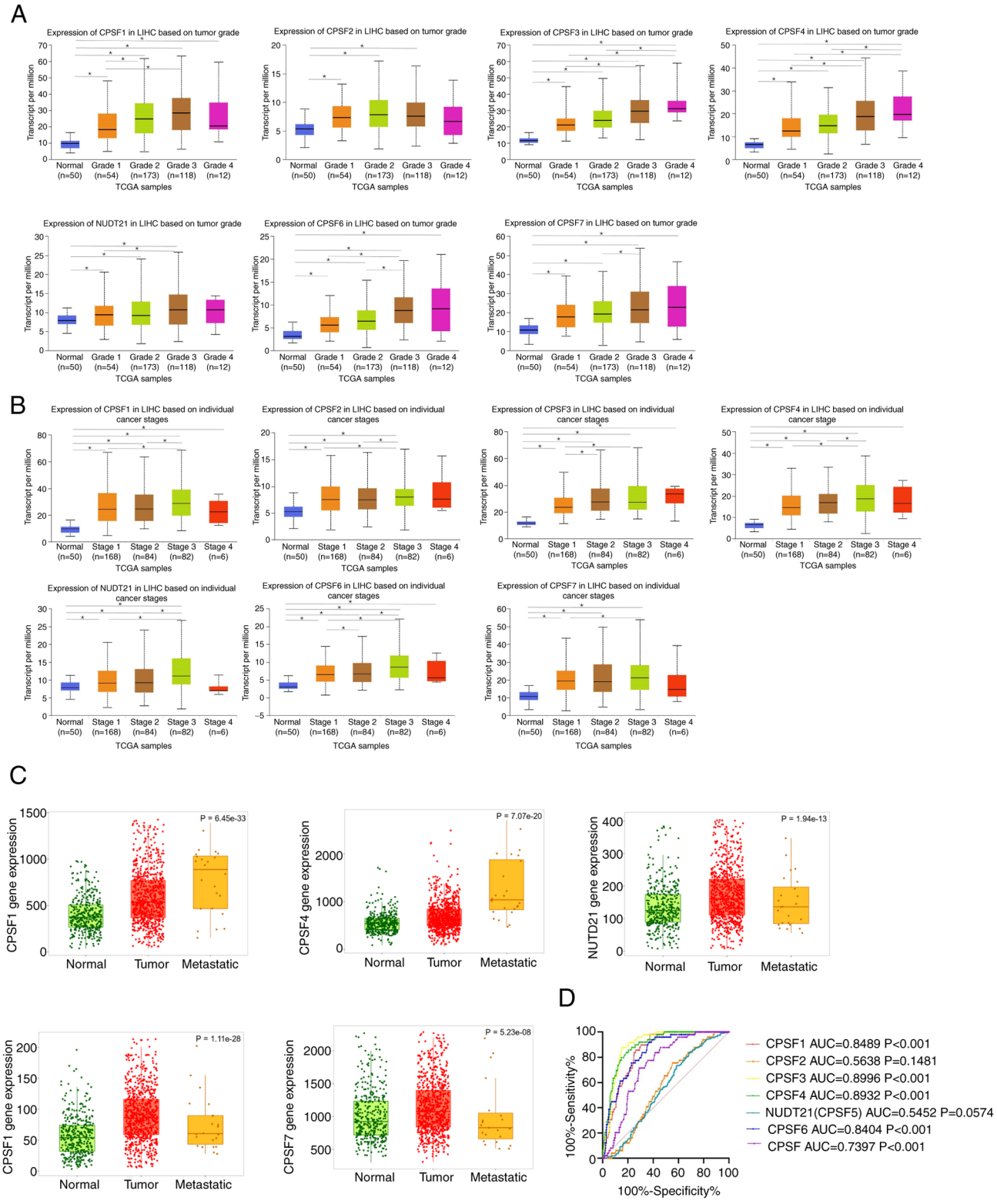

Figure 2. Correlation between different CPSF expression levels and clinicopathological parameters according to UALCAN. (A) Relationships between different CPSF expression levels and the tumor grade of patients with HCC. (B) Relationships between different CPSF expression levels and the tumor stage of patients with HCC. (C) Relationships between different CPSF expression levels and lymph node metastasis status in patients with HCC. N0; no regional lymph node metastasis, N1; 1-3 cases of axillary lymph node metastasis in UALCAN and TNMplot. *P<0.05. (D) ROC curves of different CPSF complex members in the TCGA-HCC cohort. CPSF, cleavage and polyadenylation specificity factor; HCC, hepatocellular carcinoma; TCGA, The Cancer Genome Atlas; LIHC, liver hepatocellular carcinoma.

(n=49). The analysis revealed a significant correlation between most CPSFs and HCC, as indicated by the substantial area under the curve (AUC). Specifically, the AUCs for CPSF1, CPSF3, CPSF4, CPSF6 and CPSF7 in the TCGA HCC cohort were 0.8489, P<0.001; 0.8996, P<0.001; 0.8932, P<0.001; 0.8404, P<0.001; and 0.7397, P<0.001, respectively.

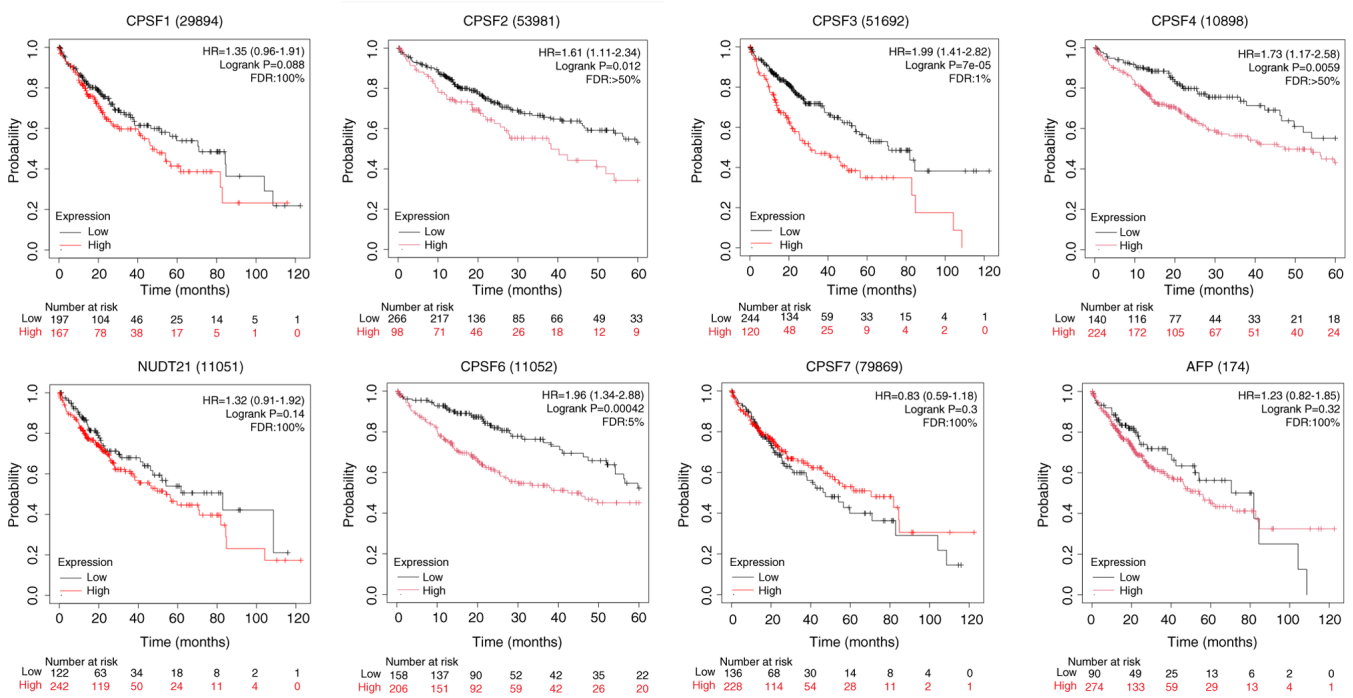

Figure 3. Prognostic value of CPSFs and AFP in patients with HCC according to OS curves (Kaplan-Meier plotter). The OS curves revealed that patients with hepatocellular carcinoma with low expression of CPSF1, CPSF3, NUDT21 and CPSF6 are more likely to survive for a relatively longer period of time, which indicates that these genes are closely linked to a favorable prognosis when compared with AFP. CPSF, cleavage and polyadenylation specificity factor; OS, overall survival; FDR, false discovery rate; HR, hazard ratio.

In summary, dysregulated expression of CPSFs is implicated in the progression of liver cells in HCC and has distinctive diagnostic value in this context.

Prognostic analysis findings of CPSF complex members in HCC. The capabilities of GEPIA and Kaplan-Meier Plotter were leveraged to explore the relationship between expression levels and prognostic significance. By collecting data and generating OS and DFS curves using the Kaplan-Meier method, it was observed that patients with HCC with low expression levels of CPSFs tended to have longer survival times, suggesting a favorable prognosis compared with AFP patients according to HR values (Fig. 3). Similarly, the DFS curves generated from GEPIA, which assessed the prognostic impact of different CPSF expression levels in patients with HCC, indicated that low expression of CPSFs was slightly correlated with DFS compared with AFP (Fig. 4).

Pathway analysis and drug sensitivity evaluation of CPSFs in HCC. To gain deeper insight into the role of CPSFs in HCC and identify genes closely linked to cancer progression, the GSCALite database was utilized to investigate the pathways involving CPSFs in HCC (Fig. 5A). Our pathway analysis revealed a significant association between CPSF complex members and critical tumor-related signaling pathways in HCC, such as the DNA damage response, epithelial-mesenchymal transition (EMT), RAS/mitogen-activated protein kinases (MAPK) pathway and receptor tyrosine kinase (RTK) pathway (Fig. 5B). These findings underscore the potential importance of CPSFs in HCC, encouraging further exploration of their co-expressed genes to elucidate the underlying mechanisms of cancer development. Furthermore, drug sensitivity analysis was performed using GSCALite. Correlational analysis between selected CPSF complex members and drug sensitivity revealed a predominantly negative correlation, indicating that increased expression of most CPSF complex members was associated with increased drug sensitivity. Conversely, a positive correlation suggests that elevated gene expression leads to drug resistance. Notably, CPSF1, CPSF3, CPSF4 and CPSF6 were found to significantly influence drug sensitivity (Fig. 5C).

Phenotypic assessment of CPSF1, CPSF3, CPSF4, CPSF6 and CPSF7 in HCC. Upon comprehensive literature review, two well-established tumor biomarkers (MKI67 and PCNA) were selected as reference indicators. Correlation analysis via the TNMplot's Spearman method was employed to evaluate the degree of association between gene pairs based on their slopes (Fig. 6A). Notably, owing to the limited data available for CPSF2 and CPSF3 in the database, our analysis focused exclusively on CPSF1, CPSF4, CPSF5 (NUDT21), CPSF6 and CPSF7. The findings revealed a predominantly positive correlation between genes within the CPSF family and the selected proto-oncogenes in HCC tissues, with the exception of CPSF5, which exhibited a negative correlation with MKI67 with a minor slope. These results support the hypothesis that the CPSF gene family is intricately linked to HCC and holds promise as a potential diagnostic marker warranting further investigation.

Moreover, CPSF1, CPSF3, CPSF4, CPSF6 and CPSF7 were identified as genes that are closely associated with HCC development. These genes not only have significantly elevated expression levels in HCC, but also present distinct

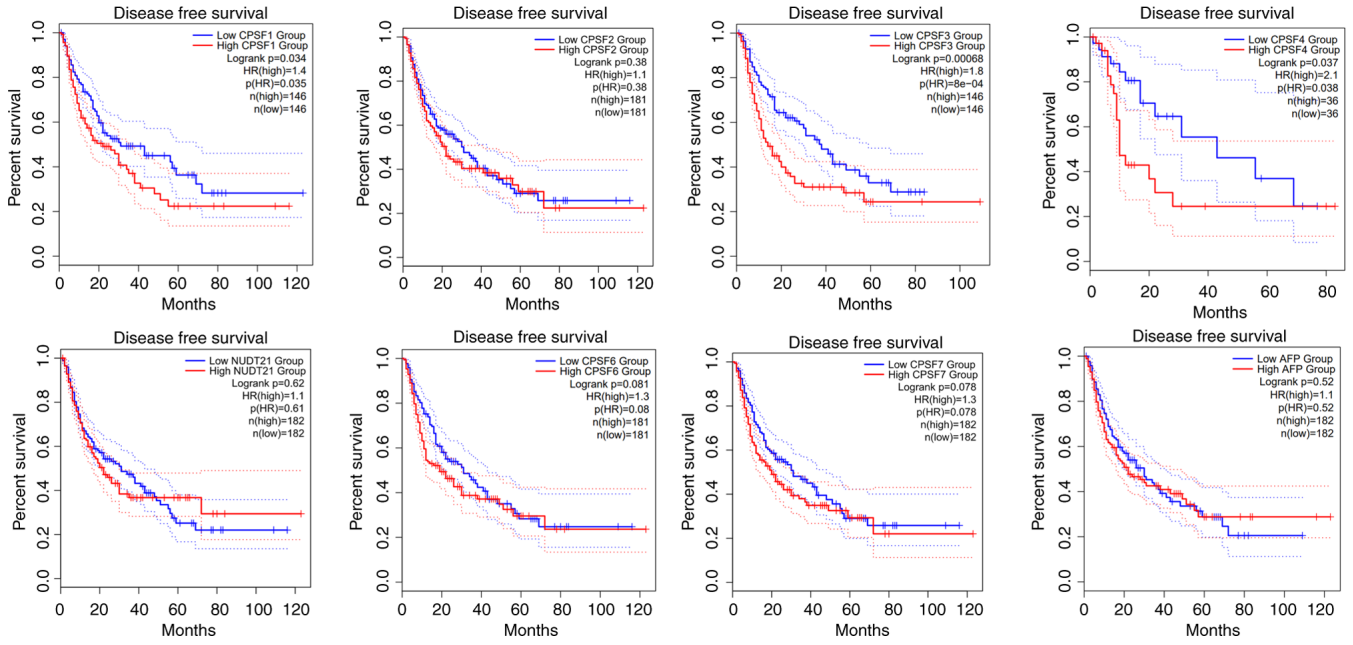

Figure 4. Prognostic value of different expression levels of CPSFs in patients with HCC according to the DFS curve (GEPIA). Compared with AFP, low expression levels of CPSF1, CPSF3 and CPSF4 are more likely to be correlated with DFS in patients with HCC. CPSF, cleavage and polyadenylation specificity factor; HCC, hepatocellular carcinoma; DFS, disease-free survival; HR, hazard ratio.

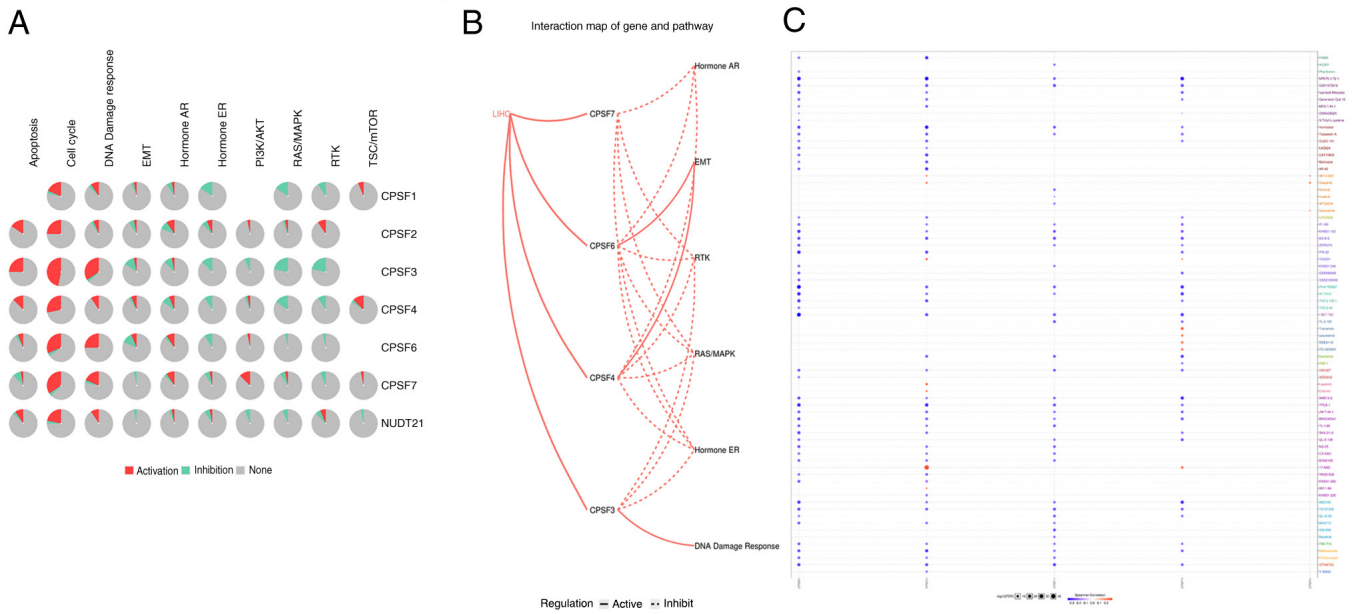

Figure 5. Comprehensive easy pathway and drug sensitivity analysis of CPSF complex members through GSCALite. (A) Relationships between pathways and CPSF complex members and (B) interactions between pathways and CPSF complex members. (C) Drug sensitivity analysis of CPSF1, CPSF3, CPSF4, CPSF6 and CPSF7 in hepatocellular carcinoma was also conducted. Spearman correlation was used to evaluate the drug resistance of the gene set derived from the Genomics of Drug Sensitivity in Cancer/Cancer Therapeutics Response Portal IC_{50} drug data. High expression of genes was found to be positively correlated with drug resistance, whereas low expression was associated with decreased resistance to treatment. CPSF, cleavage and polyadenylation specificity factor.

P-values that enable discrimination across different tumor grades. Importantly, their positive correlation with MKI67 and PCNA-recognized proto-oncogenes further supports their relevance as tumor markers. Additionally, CPSF3, CPSF4, CPSF6 and CPSF7 have been implicated in various tumor-related signaling pathways in HCC. Furthermore, the satisfactory ROC curve results (Data S1), with high AUC values, further bolster the significance of these genes.

Subsequent analysis utilizing the DepMap database revealed that the knockout of CPSF1, CPSF3, CPSF4 and CPSF6 via CRISPR/Cas9 significantly inhibited the proliferation of HCC cell lines, whereas CPSF7 did not demonstrate comparable efficacy (Fig. 6B).

Immune cell infiltration patterns in HCC in relation to the expression levels of CPSFs. Tumor inflammation and immune

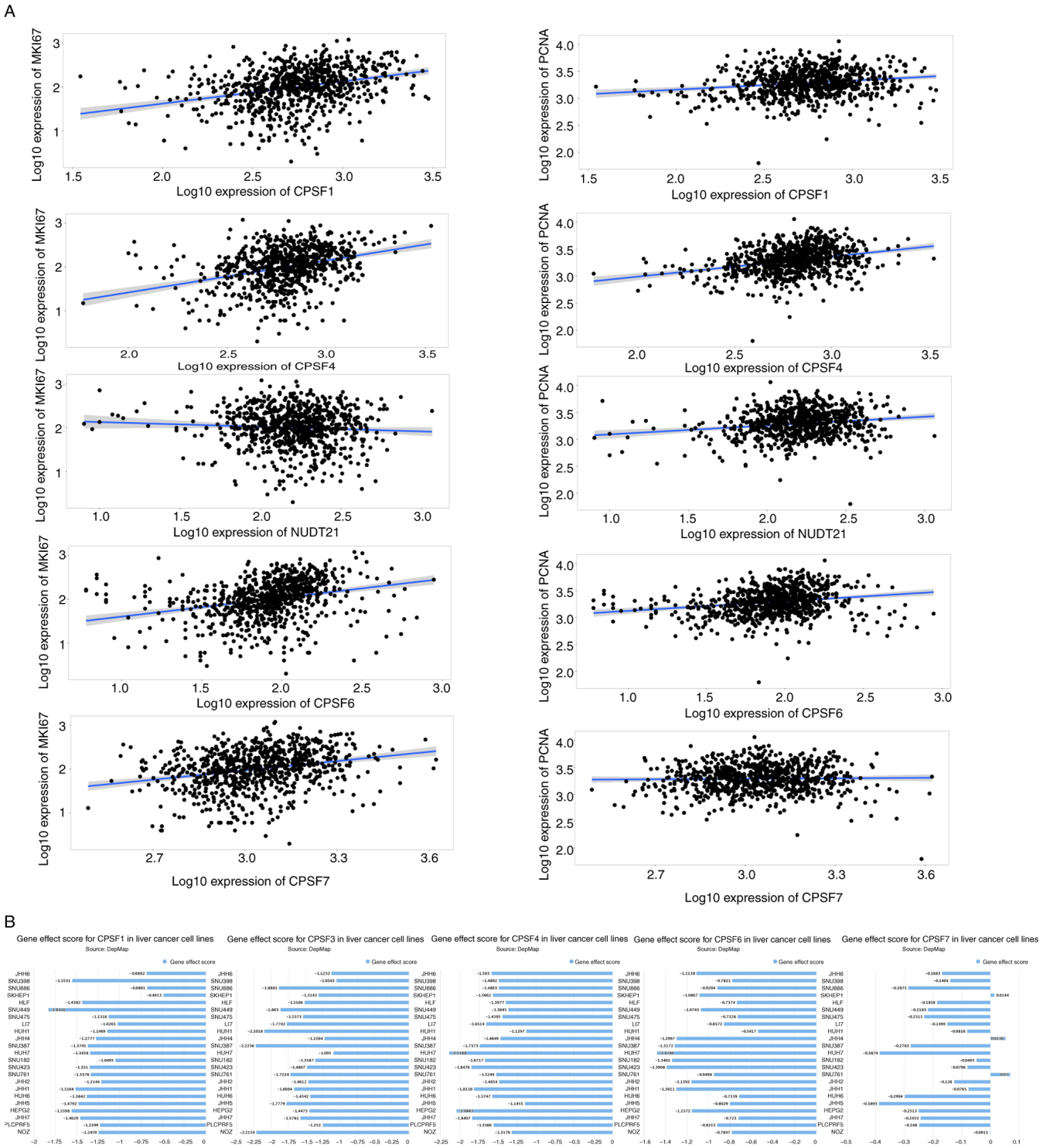

Figure 6. Phenotypic evaluation of CPSF complex family members. (A) The correlation between CPSFs and tumor markers in the TNM plot. A positive slope indicates a positive correlation, and a negative slope indicates a negative correlation. (B) Knockout of CPSF1, CPSF3, CPSF4 and CPSF6 using CRISPR/Cas9 could significantly suppress the proliferation of hepatocellular carcinoma cell lines, whereas CPSF7 could not, according to an analysis of the DepMap database. CPSF, cleavage and polyadenylation specificity factor.

cell infiltration are crucial determinants of cancer prognosis. Within the TME, infiltrating immune cells are categorized as either pro-tumorigenic or anti-tumorigenic, depending on their functional roles. During tumor progression, the immunosuppressive activities of these immune cells often increase, which can hinder the effectiveness of cytotoxic immune cells such as $CD8^+$ T cells and natural killer (NK) cells. This process contributes to the creation of an immune-suppressive microenvironment that facilitates tumor growth and progression.

To improve understanding of the relationship between CPSF expression and immune cell infiltration, the TIMER database was utilized. Our analysis revealed that the expression of CPSF1, CPSF3, CPSF4, CPSF6 and CPSF7 was slightly positively correlated with the infiltration of several

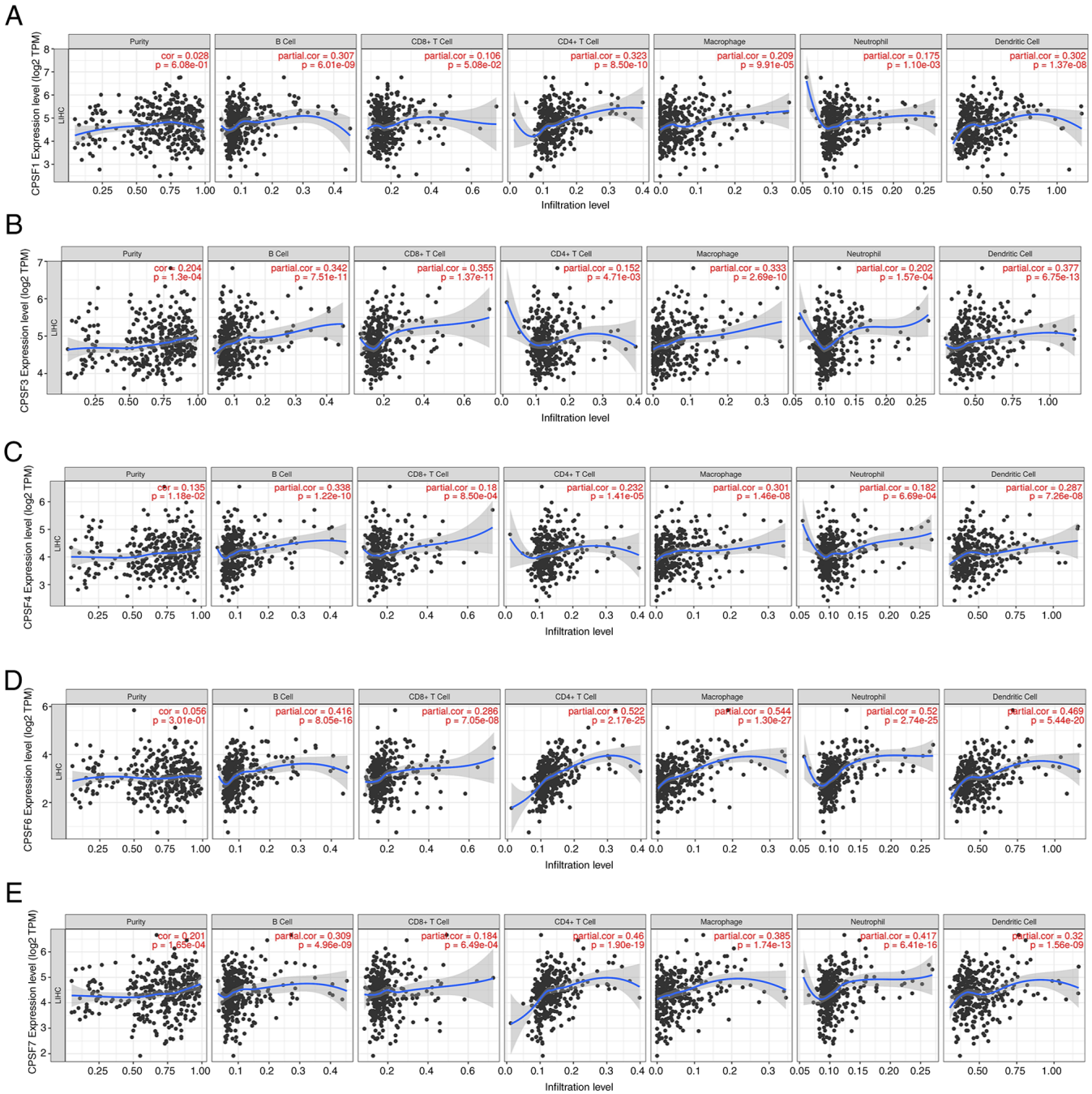

Figure 7. Correlation between differentially expressed CPSFs and immune cell infiltration (TIMER). The horizontal coordinate represents the degree of immune infiltration of specific immune cells, and the vertical coordinate represents the expression of the CPSF gene. (A-E) CPSF1, CPSF3, CPSF4, CPSF6 and CPSF7 expression was positively correlated with the infiltration of B cells, $CD8^+$ T cells, $CD4^+$ T cells, macrophages, neutrophils and dendritic cells. CPSF, cleavage and polyadenylation specificity factor.

immune cell types, including B cells, $CD8^+$ T cells, $CD4^+$ T cells, macrophages, neutrophils and dendritic cells, although the degree of correlation varied (Fig. 7A-E). Interestingly, these CPSF genes, which are typically associated with poor prognosis, demonstrated a positive correlation with immune cell infiltration, especially with $CD8^+$ T cells-an indicator often associated with favorable clinical outcomes. This apparent contradiction prompted us to consider the activation status and heterogeneity of the immune cell subpopulations involved.

To explore this, immune infiltration data from the TISIDB database was analyzed, focusing on functionally distinct immune cell subgroups in HCC. The results from this analysis revealed a generally negative correlation between CPSF expression and immune infiltration, highlighting a reduced presence of activated immune cells such as B cells, NK cells, and macrophages (Fig. 8). These findings suggest that CPSF expression may contribute to the establishment of an immune-suppressive microenvironment in HCC. Among the CPSFs, CPSF6 and CPSF7 emerged as critical factors, demonstrating significant negative correlations with all analyzed immune cell types. These findings suggest that CPSFs play a role in immune evasion mechanisms.

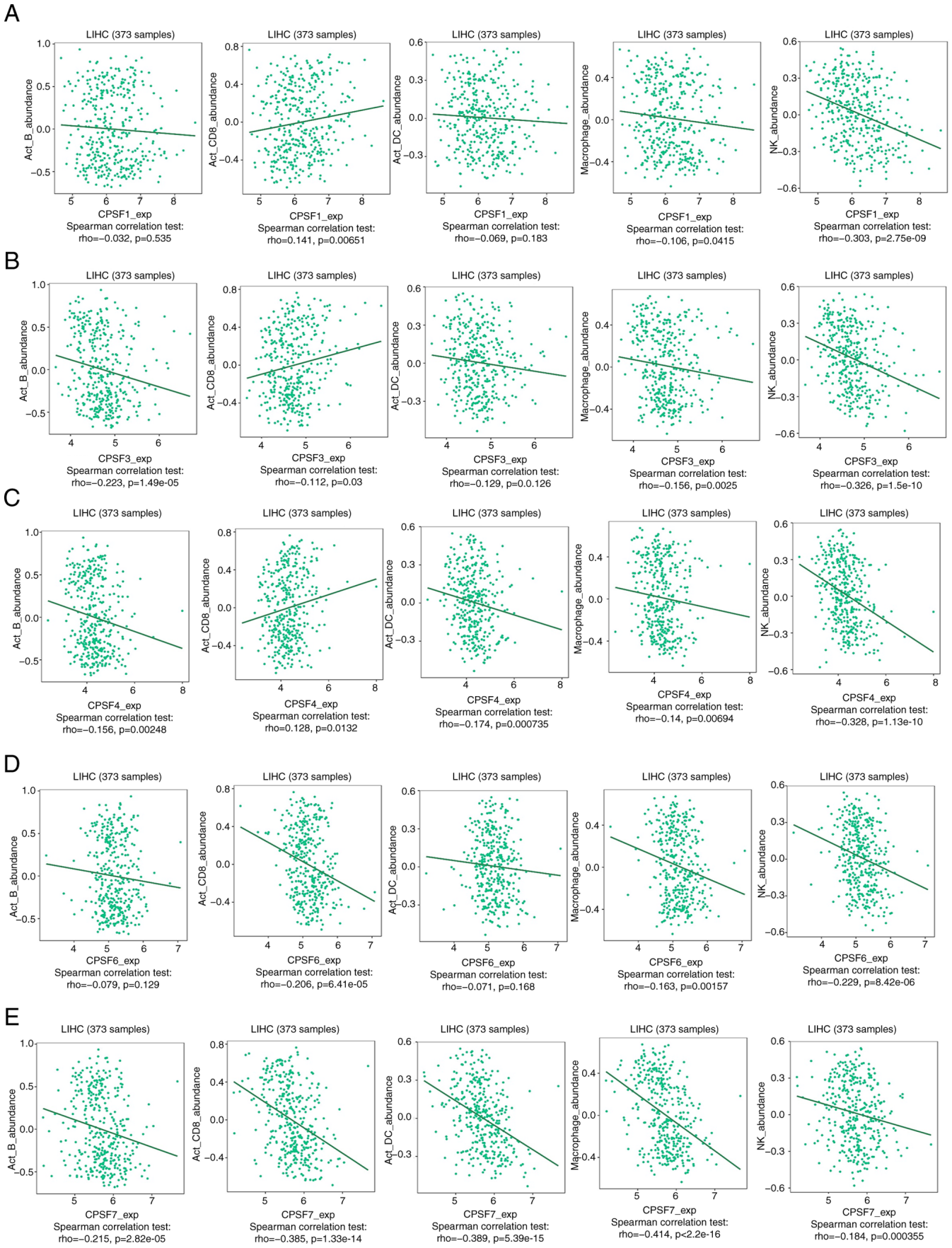

Figure 8. Association of the expression of CPSFs with immune infiltration level in HCC from TISIDB. The horizontal and vertical coordinates represent the expression of the CPSF gene and the abundance of immune cells, respectively. (A-E) Scatter plot of the correlation between the expression level of CPSFs and the infiltration of immune cells, including Act_B cells, Act_CD8 T cells, NK cells, Act_DC and macrophages (Spearman correlation, n=373). These plots show that in HCC, CPSF is strongly negatively correlated with activated immune cells, especially CPSF6 and CPSF7. CPSF, cleavage and polyadenylation specificity factor; HCC, hepatocellular carcinoma; Act, activated; NK, natural killer cells; DC, dendritic cells; LIHC, liver hepatocellular carcinoma.

Enrichment analysis of genes co-expressed with CPSFs in HCC. In the present study, the roles of CPSF1, CPSF3, CPSF4, CPSF6 and CPSF7 as key research targets were investigated. Using the LinkedOmics and cBioPortal databases, the top 200 genes positively correlated with these CPSFs were identified, and the overlapping genes across the datasets were determined (Fig. 9A). The overlapping gene sets were further analyzed using GO and KEGG enrichment analyses to explore the functional roles and associated pathways of CPSFs. The GO and KEGG enrichment analyses (Fig. 9B-F) were conducted using the WebGestalt platform and the FunRich analysis tool, respectively. GO analysis, which encompasses cellular component, molecular function and biological process terms, revealed that the genes co-expressed with CPSF complex members were primarily involved in processes related to metabolism, biological regulation and cellular organization. Specifically, in terms of the cellular component, the genes were significantly active in the nucleus, protein complexes, membrane-bounded compartments and cytosol. The molecular function analysis revealed key functions critical for cancer progression, including protein binding, nucleic acid binding and ion binding, as well as roles in hydrolase activity and nucleotide binding.

Pathway analysis revealed that the genes co-expressed with CPSF complex members were involved in several crucial cancer-related signaling pathways, underscoring their potential value as tumor biomarkers. For example, the KEGG analysis of the CPSF1-coexpressed genes highlighted enrichment in pathways such as the ataxia telangiectasia mutated (ATM) pathway, the tumor necrosis factor-related apoptosis-inducing ligand (TRAIL) signaling pathway, the mTOR signaling pathway, the TNFα/NF-kB pathway, the interferon (IFN)-γ pathway and metabolism. Similarly, the CPSF3-coexpressed genes were associated with pathways related to TRAIL signaling, mTOR signaling, Polo-like kinase 1 (PLK-1) signaling, Class I phosphatidylinositol 3-kinase signaling, the insulin-like growth factor 1 (IGF1) pathway, the IFN-γ pathway and cell cycle regulation. Further analyses revealed correlations between CPSF4 and gene expression, mTOR signaling, IL-3-mediated events, the IGF1 pathway, the IFN-γ pathway, TRAIL signaling, metabolism and cell cycle regulation. Similarly, CPSF6 and CPSF7 were linked to various signaling events and pathways critical for cancer development, highlighting their potential roles in tumor progression and providing valuable insights into their significance as tumor biomarkers.

Expression of CPSF family genes in tumor cell lines and the cell proliferation assay. To validate the results obtained from the bioinformatics analysis, RT-qPCR was performed on three HCC cell lines, with the THLE-2 cell line serving as a normal control. These findings indicated substantial upregulation of the majority of CPSF genes in the HCC cell lines, namely, HUH7, JHH7 and PLC/PRF/5, compared with those in the THLE-2 cell line. Specifically, significant increases in the expression levels of CPSF1, CPSF3, CPSF4, CPSF6 and CPSF7 were observed in the HCC cell lines (Fig. 10).

Based on the RT-qPCR results and the bioinformatics analysis aforementioned, further investigations into the functional roles of CPSF3, CPSF6 and CPSF7 in the HUH7 cell line were conducted using a CCK-8 assay. The knockdown efficiency was validated through RT-qPCR analysis (Fig. 11A). The results clearly demonstrated that CPSF3 siRNA and CPSF7 siRNA significantly slowed the proliferation of HUH7 cells, with cell survival rates of ~88 and 85%, respectively, at 72 h ($P<0.05$) (Fig. 11B).

Discussion

HCC is a highly aggressive malignancy, that is distinguished by intricate molecular mechanisms and substantial interindividual variability. Identifying prevalent genetic aberrations in HCC is pivotal for advancing effective targeted therapeutic strategies (22). The CPSF is an essential component of the 3' end processing machinery responsible for polyadenylated mRNA synthesis in metazoans. CPSF interacts with the polyadenylation signal sequence (AAUAAA), ensuring precise sequence recognition during both pre-mRNA cleavage and polyadenylation processes, and facilitates the cleavage of pre-mRNA. It has been previously revealed that CPSF1 promotes the proliferation of ovarian cancer cells *in vitro*, indicating that its inhibition may offer a novel therapeutic avenue for ovarian cancer (23). Additionally, CPSF3 has been implicated in the prognosis and recurrence of lung adenocarcinoma (LUAD) (24). Despite these findings, the comprehensive role of CPSFs in HCC remains elusive. In the present study, the potential implications of CPSFs in HCC pathogenesis were systematically investigated through various analytical approaches.

Our first step was to understand the expression of CPSFs in both normal liver tissues and HCC tissues. Different public databases were used, including the GEPIA2 database and cBioPortal, to analyze and evaluate the data for increased accuracy. In the GEPIA2 database, it was found that CPSF1/2/3/4/5/6/7 were more highly expressed in HCC tissues than in normal liver tissues, especially CPSF1, CPSF3, CPSF4 and CPSF6 ($P<0.05$). The cBioPortal data revealed that most CPSF complex members (over 50%) presented gene alterations in patients with HCC and that the expression of CPSF1/3/4/5/6/7 increased, while the expression of CPSF2 decreased. Therefore, it was hypothesized that CPSF complex members may be potential diagnostic markers of HCC and have diagnostic value for liver cancer.

The associations between the expression levels of CPSF complex components and specific clinical tumor indicators were subsequently explored. The present findings revealed that the elevated expression of CPSFs was significantly correlated with more advanced HCC stage, higher tumor grade and lymph node metastasis. Notably, several CPSF family members, including CPSF1, CPSF3, CPSF4, CPSF6 and CPSF7, demonstrated favorable P-values that effectively distinguished between different tumor grades. In addition, most of the CPSF complex members had a positive relationship with several typical tumor markers, such as PCNA and MKI67. Genes such as PCNA have been shown to be valuable prognostic molecular markers for liver cancer because they contribute to the cell cycle (25). MKI67 plays a crucial role in the TME and innate immunity. MKI67 has the potential to serve as a prognostic biomarker, especially in HCC. The upper level of MKI67 increases immune cell infiltration, involving various subtypes, such as functional T cells, $CD4^+$ T cells and $CD8^+$ T cells. Furthermore, MKI67 is closely associated with T cell exhaustion, which contributes to the exacerbation of T cells within LIHC (26). As CPSFs are posi-

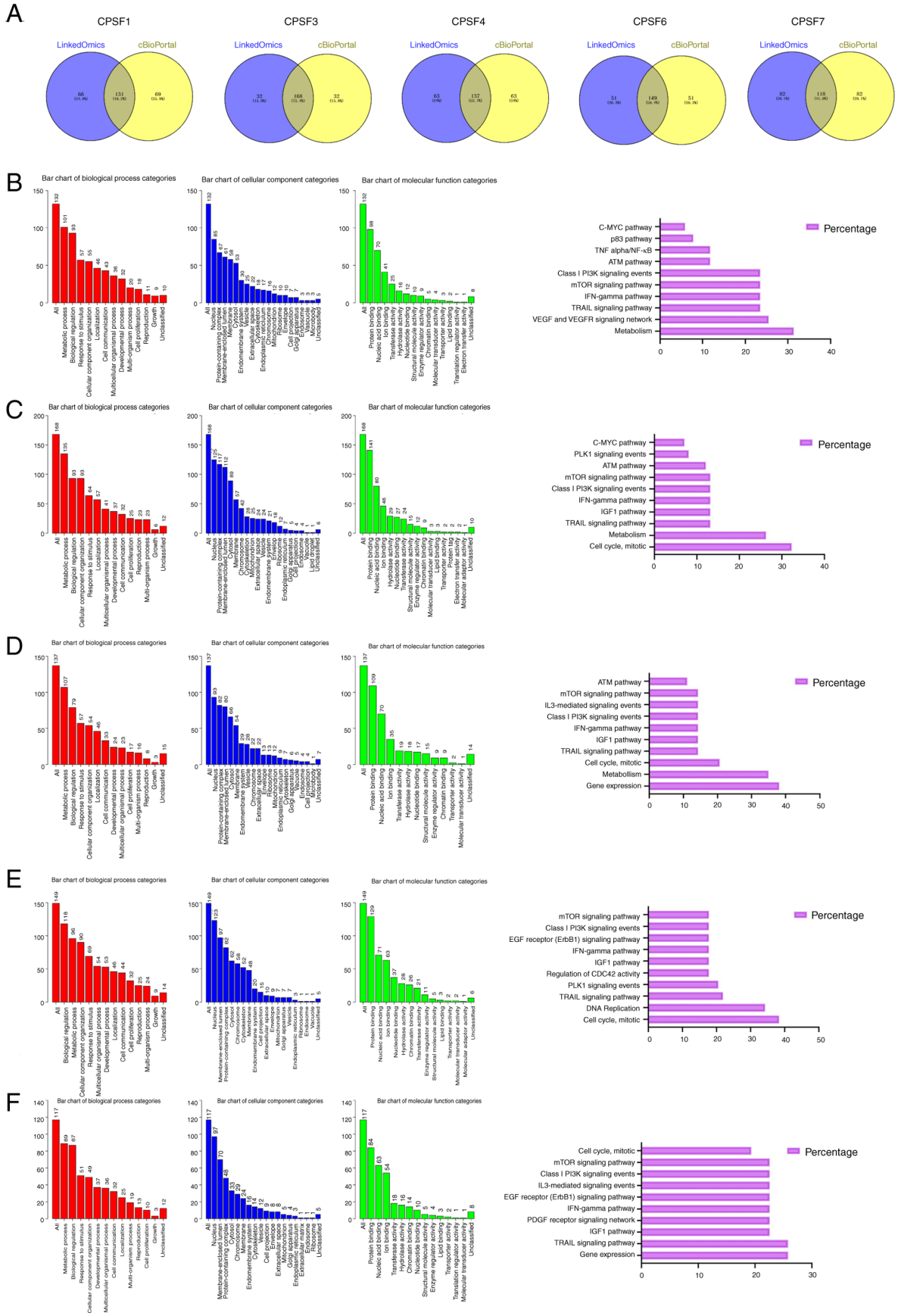

Figure 9. Comprehensive bioinformatics analysis of selected CPSFs and individual co-expressed genes in hepatocellular carcinoma. (A) Venn diagrams were used to screen the reliable co-expressed genes of CPSFs by using the LinkedOmics database and the cBioPortal database. Gene Ontology functional enrichment analysis via WebGestalt predicted the main functions of CPSFs and their 100 related genes, including biological processes, cellular components and molecular functions. (B-F) Kyoto Encyclopedia of Genes and Genomes pathway analysis of selected CPSFs (panel B corresponds to CPSF1, panel C corresponds to CPSF3, panel D corresponds to CPSF4, panel E corresponds to CPSF6 and panel F corresponds to CPSF7) and their 50 related genes through the FunRich tool. CPSF, cleavage and polyadenylation specificity factor.

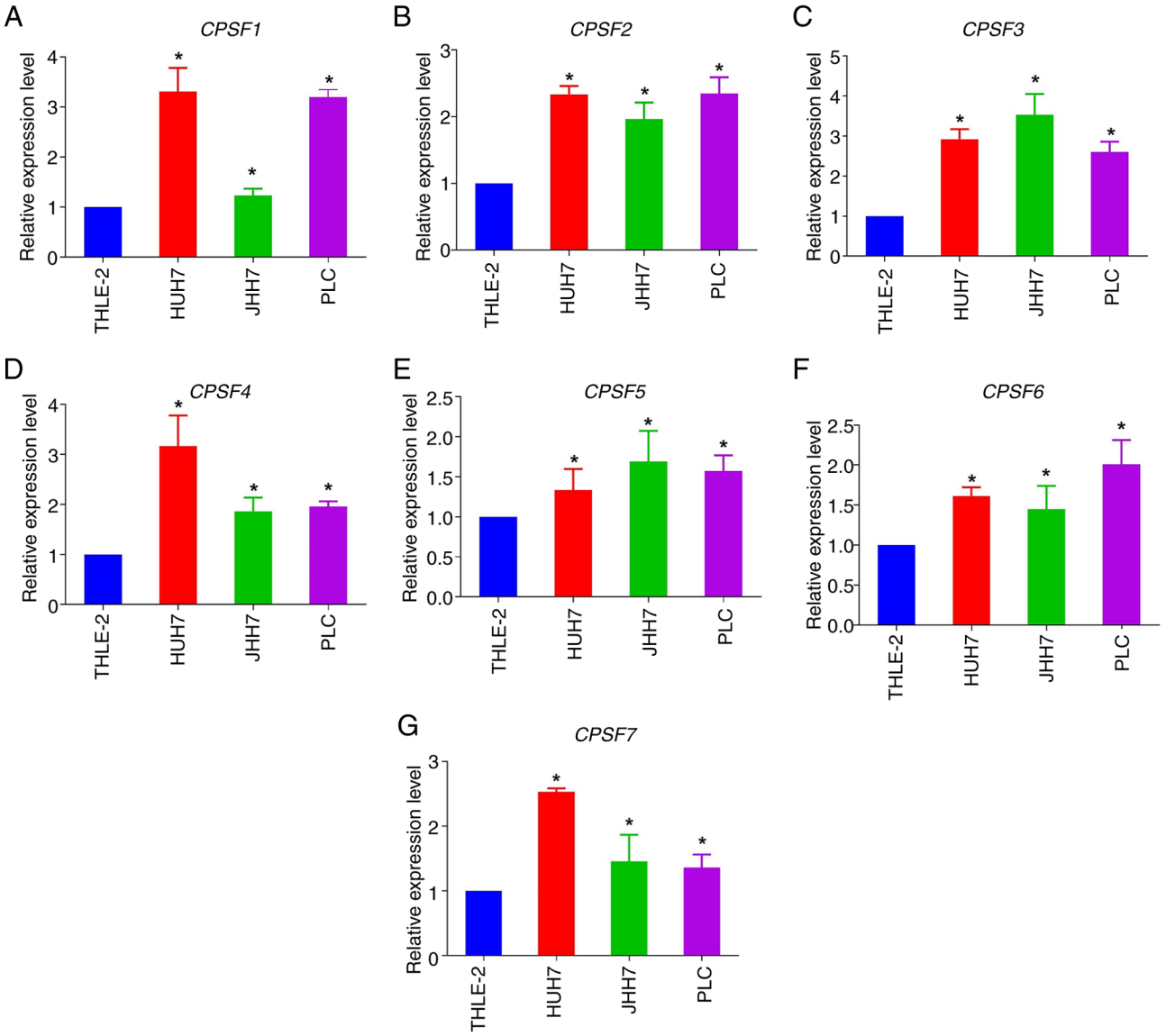

Figure 10. Relative expression levels of CPSF1-7 measured by RT-qPCR. (A-G) The expression levels of CPSF1-7 in THLE-2 cells and 3 hepatocellular carcinoma cell lines (HUH7, JHH7 and PLC/PRF/5) were detected by RT-qPCR. *P<0.05, compared with the normal human liver cell line THLE-2. CPSF, cleavage and polyadenylation specificity factor; RT-qPCR, reverse transcription-quantitative PCR.

tively correlated with MKI67, it provides a theoretical basis to illustrate the increase in the expression level of related immune cells with elevated expression levels of CPSFs (CPSF1, CPSF4, CPSF6 and CPSF7). These conclusions also support the great reliability of CPSFs as biomarkers for HCC diagnosis and prognosis. The gene knockout data of CPSF1, CPSF3, CPSF4, CPSF6 and CPSF7, which are considered the most typical biomarkers of CPSFs because they are closely related to HCC, all strongly increased the proliferation of HCC cell lines except for CPSF7.

The current analysis of OS and DFS confirmed that only CPSF3 was significantly associated with both prognosis and recurrence in patients with LUAD, indicating its potential as a promising biomarker. Additionally, ROC curve analysis suggested that CPSF3 could serve as a diagnostic biomarker to distinguish between the two histological subtypes of non-small cell lung cancer (24). Based on these findings, the prognostic significance of CPSF family members in HCC was investigated. By utilizing Kaplan-Meier Plotter and GEPIA, OS and DFS curves were generated for each CPSF complex member as well as the known HCC biomarker AFP via the Kaplan-Meier method. The results clearly demonstrated that increased expression levels of certain CPSF members were associated with significantly poorer prognosis in patients with HCC. These findings highlight the need for further in-depth research to explore their potential as prognostic biomarkers in HCC.

To further elucidate the relationships between HCC and cleavage and CPSFs, the associations between CPSF expression levels and immune cell infiltration in HCC were examined. Our tumor immune infiltration analysis revealed that CPSF expression in HCC was positively correlated with immune cell infiltration. However, it was hypothesized that this positive correlation might be misleading because of the inactivated state of immune cells. To test this hypothesis, the TISIDB database was used to perform Spearman correlation tests between selected CPSFs and activated immune cells. The results indicated that CPSFs exhibited an inverse relationship with the infiltration

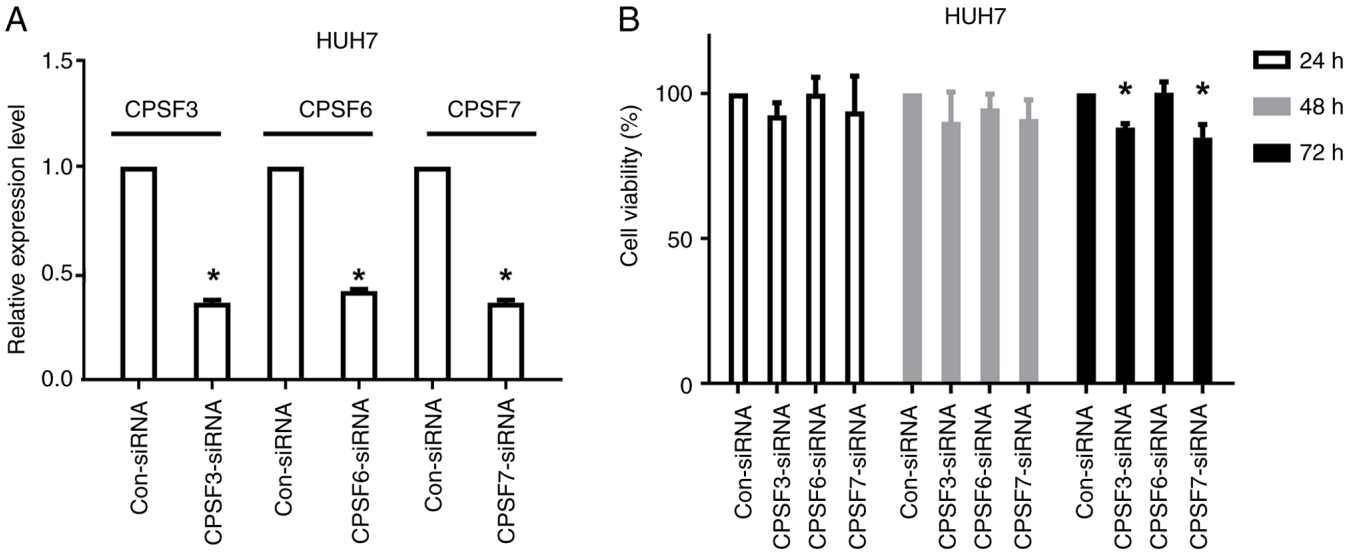

Figure 11. Cell viability of the HUH7 cell line detected by Cell Counting Kit-8 assay. (A) The relative gene expression levels of CPSF3, CPSF6 and CPSF7 were determined to ensure that the siRNAs successfully inhibited these genes. (B) HUH7 cell viability and changes in each group. *P<0.05. CPSF, cleavage and polyadenylation specificity factor; siRNA, small interfering RNA.

of several immune cell types, particularly activated B cells, NK cells and macrophages. The contrasting results observed between the TIMER and TISIDB databases (positive correlations in TIMER and negative correlations in TISIDB) require further investigation. These discrepancies could arise from differences in immune cell subpopulations or their activation states, which may offer insights into how CPSFs contribute to immune evasion and the creation of an immunosuppressive TME in HCC. Moreover, the authors propose that CPSFs may play a role in immune evasion mechanisms, possibly through pathways involving PD-L1, TGF-β, or myeloid-derived suppressor cells (MDSCs). PD-L1, an immune checkpoint molecule, can inhibit the activity of cytotoxic T cells by binding to PD-1, promoting immune evasion. Additionally, TGF-β, a cytokine known for its immune-suppressive effects, promotes the differentiation of regulatory T cells (Tregs) and inhibits the activation of NK cells and cytotoxic T lymphocytes. Furthermore, MDSCs, which accumulate in the TME, inhibit T cell function and promote the expansion of Tregs. Additionally, the HPA database was utilized to generate immunohistochemical images of CPSF proteins in both HCC and normal liver tissues. This direct visualization revealed distinct expression patterns. While CPSF1 and CPSF4 were not detected, CPSF2 and CPSF7 were strongly expressed in normal liver tissues but weakly expressed in HCC tissues. CPSF5 exhibited weak expression in normal liver tissues and moderate expression in HCC tissues. CPSF6 was highly expressed in both normal and tumor tissues, whereas CPSF3 was moderately expressed in both normal liver and HCC tissues. These findings provide strong evidence that HCC can significantly impact the expression of CPSFs and their proteins. Collectively, these observations underscore the significant influence of CPSF1, CPSF3, CPSF4, CPSF6 and CPSF7 on the TME, thereby affecting the occurrence and development of HCC. This makes them valuable candidates as diagnostic and prognostic markers.

Pathway analysis of the CPSF gene family revealed that CPSF3, CPSF4, CPSF6 and CPSF7 are closely associated with EMT, the RAS/MAPK pathway, the RTK pathway and the DNA damage response in HCC. In HCC, the process of EMT is mediated primarily by the Wnt/β-catenin signaling pathway. This pathway plays a key role in modulating the TME and is strongly associated with the progression of tumor metastasis (27). MAPKs are a group of serine-threonine kinases that represent key signaling pathways governing cell proliferation and extend from the cell membrane to the nucleus. Several non-coding RNAs linked to the MAPK pathway play crucial roles in the progression of various cancers, including liver cancer (28). Moreover, CPSFs are associated with the RTK pathway, which is carcinogenic. These findings lay the foundation for further exploration of the relationship between CPSFs and cancer. GO and pathway analyses of the co-expressed genes in the screened CPSF family (CPSF1, CPSF3, CPSF4, CPSF6 and CPSF7) were then further performed via the WebGestalt website and FunRich tool. The results of the GO analysis were mainly related to the nucleus and protein activities, such as metabolism and binding, cytosol and cell communication and localization. As aforementioned, the biogenesis of mammalian mRNAs involves the specific recognition of a hexanucleotide AAUAAA motif in the polyadenylation signals of pre-mRNA transcripts by the CPSF complex (6). As a result, they are fatal in transcription and translation. The sudden change in the gene transcription process would produce irreversible consequences of the upregulation or downregulation of genes. If transcriptional abnormalities occur in cancer-associated genes, tumor development is highly likely. Splicing, a crucial stage of mRNA processing, plays a significant role in regulating gene expression. Abnormal splicing events can be considered early indicators of cancerous transformation in cells (29). Pathway analysis revealed that these genes were enriched mainly in the TNF alpha/NF-kB, PLK-1 signaling events, the IFN-γ pathway, the ATM pathway, and the cell cycle mitotic and TRAIL signaling pathways. To provide further evidence, the effectiveness of Chk1 inhibition was verified in terms of its ability to abolish the G2/M cell cycle checkpoint and induce apoptosis in the adherent cancer

cell line (30). PLK1 is crucial for the initiation, progression and completion of mitosis. Its inactivation can lead to tumorigenesis and accelerate cancer progression. Overexpression of PLK1 has been observed in various human cancers and is linked to poor patient prognosis (31). The progression and development of cancer are influenced by a combination of genetic and epigenetic modifications, with the TME serving as a critical factor in tumor advancement and transformation. This is especially evident through processes such as inflammation promotion. Tumor necrosis factor-alpha (TNF-α) is a prominent pro-inflammatory cytokine present in the TME of patients with cancer and has been demonstrated to play a pro-tumorigenic role in the progression and metastasis of breast cancer (32). TRAIL, a member of the TNF superfamily, can induce apoptosis in tumor or infected cells, further influencing tumor behavior (33). Substantial evidence has shown that endogenously produced IFN plays a crucial role in the regulation of tumor development. It not only promotes protective host responses against tumors but also orchestrates responses within tumors, enabling them to evade immune attacks (34). These findings provide strong evidence for the value of CPSFs in HCC.

The findings of the present study clearly demonstrated that the mRNA expression levels of CPSFs are greater in HCC cell lines than in normal liver cells. Specifically, the expression of CPSF1, CPSF3, CPSF4, CPSF6 and CPSF7 was significantly upregulated, especially in the HUH7 and JHH7 cell lines. In addition, cell proliferation assays indicated that CPSF3 and CPSF7 may influence the viability of HUH7 cells. When the expression of CPSF3 and CPSF7 was knocked down by siRNA, a reduction in cell proliferation was observed. These results offer valuable insights into the potential significance of CPSFs as biomarkers and their impact on the proliferation of tumor cells in HCC. Furthermore, the validation of the current findings through multiple databases, RT-qPCR and cell viability assays strengthened the reliability of our conclusions. However, it is crucial to acknowledge the limitations of the present study. These limitations include the reliance on public databases, the scarcity of further statistical tests according to the lack of raw data, and the need for additional experimental validation to clarify the biological functions of CPSFs in HCC. Due to the nature of publicly available datasets, limitations in accessing raw data were encountered, which restricted our ability to perform additional mechanistic analyses and further hypothesis tests directly within the present study. Moreover, our current validation is based solely on *in vitro* experiments, which just providing preliminary evidence of the functional role of CPSFs. The importance of deeper mechanistic validation and plan to address these limitations in future studies is recognized. Specifically, the authors aim to conduct *ex vivo* and *in vivo* functional experiments, such as knockdown/overexpression models, to further elucidate the downstream regulatory targets of CPSFs in HCC progression. Additionally, the authors plan to perform RNA-sequencing analysis on CPSF-knockdown cell lines to identify key pathways influenced by CPSF dysregulation. These follow-up studies will be crucial in verifying the current findings and providing mechanistic insights into the role of CPSFs in HCC.

In conclusion, the present study offers a thorough examination of the expression patterns, prognostic value and immune infiltration of CPSFs in HCC. The increased expression of CPSFs in HCC, coupled with their correlation with unfavorable prognosis and immune infiltration, highlights their preliminary promise as possible therapeutic targets and prognostic indicators in HCC. Additional studies are needed to further understand and confirm the underlying mechanisms and clinical significance of CPSFs in HCC.

Acknowledgements

Not applicable.

Funding

The present study was supported by the National Natural Science Foundation of China (grant nos. 82474125 and 82074101) and Shanghai Municipal Health Commission (grant no. 202340073).

Availability of data and materials

The data generated in the present study may be requested from the corresponding author.

Authors' contributions

YL and HZ conceived and designed the experiments. YL drafted and finalized the manuscript with the direction of HZ and XY. YL, TW and XY participated in some experiments and data result processing. All authors read and approved the final version of the manuscript. YL and HZ confirm the authenticity of all the raw data.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Villanueva A: Hepatocellular carcinoma. N Engl J Med 380: 1450-1462, 2019.
2. Siegel RL, Miller KD, Wagle NS and Jemal A: Global cancer statistics, 2023. CA Cancer J Clin 73: 17-48, 2023.
3. Pradeep G, Vikram B and Sharma DVHS: Biomarkers to assess liver function in various types of liver diseases. Int J Clin Biomed Res 5:28-31, 2019.
4. Cai SH, Lu SX, Liu LL, Zhang CZ and Yun JP: Increased expression of hepatocyte nuclear factor 4 alpha transcribed by promoter 2 indicates a poor prognosis in hepatocellular carcinoma. Therap Adv Gastroenterol 10: 761-771, 2017.
5. Villanueva A, Hoshida Y, Battiston C, Tovar V, Sia D, Alsinet C, Cornella H, Liberzon A, Kobayashi M, Kumada H, *et al*: Combining clinical, pathology, and gene expression data to predict recurrence of hepatocellular carcinom. Gastroenterology 140: 1501-12.e2, 2011.
6. Clerici M, Faini M, Muckenfuss LM, Aebersold R and Jinek M: Structural basis of AAUAAA polyadenylation signal recognition by the human CPSF comple. Nat Struct Mol Biol 25: 355, 2018.

7. Pallante P, Terracciano L, Carafa V, Schneider S, Zlobec I, Lugli A, Bianco M, Ferraro A, Sacchetti S, Troncone G, *et al:* The loss of the CBX7 gene expression represents an adverse prognostic marker for survival of colon. Eur J Cancer 46: 2304-2313, 2010.
8. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi BVSK and Varambally S: UALCAN: A portal for facilitating tumor subgroup gene expression and survival analyses. Neoplasia 19: 649-658, 2017.
9. Tomczak K, Czerwińska P and Wiznerowicz M: The cancer genome atlas (TCGA): An immeasurable source of knowledge. Contemp Oncol (Pozn) 19: A68-A77, 2015.
10. Liu CJ, Hu FF, Xia MX, Han L, Zhang Q and Guo AY: GSCALite: A web server for gene set cancer analysi. Bioinformatics 34: 3771-3772, 2018.
11. Asplund A, Edqvist PH, Schwenk JM and Pontén F: Antibodies for profiling the human proteome-the human protein atlas as a resource for cancer researc. Proteomics 12: 2067-2077, 2012.
12. Nagy Á, Lánczky A, Menyhárt O and Győrffy B: Validation of MiRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. Sci Rep 8: 9227, 2018.
13. Tang Z, Li C, Kang B, Gao G, Li C and Zhang Z: GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res 45: W98-W102, 2017.
14. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B and Liu XS: TIMER: A web server for comprehensive analysis of tumor-infiltrating immune cells. Cancer Res 77: e108-e110, 2017.
15. Bartha Á and Győrffy B: TNMplot.com: A web tool for the comparison of gene expression in normal, tumor and metastatic tissue. Int J Mol Sci 5: 2622, 2017.
16. Ru B, Wong CN, Tong Y, Zhong JY, Zhong SSW, Wu WC, Chu KC, Wong CY, Lau CY, Chen I, *et al*: TISIDB: An integrated repository portal for tumor-immune system interaction. Bioinformatics 35: 4200-4202, 2019.
17. Pathan M, Keerthikumar S, Ang CS, Gangoda L, Quek CY, Williamson NA, Mouradov D, Sieber OM, Simpson RJ, Salim A, *et al*: FunRich: An open access standalone functional enrichment and interaction network analysis tool. Proteomics 15: 2597-2601, 2015.
18. Liao Y, Wang J, Jaehnig EJ, Shi Z and Zhang B: WebGestalt 2019: Gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res 47: W199-W205, 2019.
19. Vasaikar SV, Straub P, Wang J and Zhang B: LinkedOmics: Analyzing multi-omics data within and across 32 cancer types. Nucleic Acids Res 46: D956-D963, 2018.
20. Menyhárt O, Nagy Á and Győrffy B: Determining consistent prognostic biomarkers of overall survival and vascular invasion in hepatocellular carcinom. R Soc Open Sci 5: 181006, 2018.
21. Livak KJ and Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods 25: 402-408, 2001.
22. Bruix J, Gores GJ and Mazzaferro V: Hepatocellular carcinoma: Clinical frontiers and perspectives. Gut 63: 844-855, 2001.
23. Schönemann L, Kühn U, Martin G, Schäfer P, Gruber AR, Keller W, Zavolan M and Wahle E: Reconstitution of CPSF active in polyadenylation: Recognition of the polyadenylation signal by WDR33. Genes Dev 28: 2381-2393, 2014.
24. Ning Y, Liu W, Guan X, Xie X and Zhang Y: CPSF3 is a promising prognostic biomarker and predicts recurrence of non-small cell lung cancer. Oncol Lett 18: 2835-2844, 2019.
25. Zheng Y, Shi Y, Yu S, Han Y, Kang K, Xu H, Gu H, Sang X, Chen Y and Wang J: GTSE1, CDC20, PCNA, and MCM6 synergistically affect regulations in cell cycle and indicate poor prognosis in liver cance. Ana Cell Pathol (Amst) 30: 1038069, 2019.
26. Wu SY, Liao P, Yan LY, Zhao QY, Xie ZY, Dong J and Sun HT: Correlation of MKI67 with prognosis, immune infiltration, and T cell exhaustion in hepatocellular carcinom. BMC Gastroenterol 21: 416, 2021.
27. Gurzu S, Kobori L, Fodor D and Jung I: Epithelial mesenchymal and endothelial mesenchymal transitions in hepatocellular carcinoma: A revie. Biomed Res Int 2962580, 2019.
28. Wang Q, Feng J and Tang L: Non-coding RNA related to MAPK signaling pathway in liver cancer. Int J Mol Sci 23: 11908, 2019.
29. Coltri PP, Dos Santos MGP and da Silva GHG: Splicing and cancer: Challenges and opportunities. Wiley Interdiscip Rev RNA 10: e1527, 2019.
30. Furusawa Y, Iizumi T, Fujiwara Y, Zhao QL, Tabuchi Y, Nomura T and Kondo T: Inhibition of checkpoint kinase 1 abrogates G2/M checkpoint activation and promotes apoptosis under heat stress. Apoptosis 17: 102-112, 2012.
31. Liu Z, Sun Q and Wang X: PLK1, a potential target for cancer therapy. Transl Oncol 10: 22-32, 2017.
32. Cruceriu D, Baldasici O, Balacescu O and Berindan-Neagoe I: The dual role of tumor necrosis factor-alpha (TNF-α) in breast cancer: Molecular insights and therapeutic approaches. Cell Oncol (Dordr) 43: 1-18, 2020.
33. Cardoso Alves L, Corazza N, Micheau O and Krebs P: The multifaceted role of TRAIL signaling in cancer and immunity. FEBS J 288: 5530-5554, 2021.
34. Ikeda H, Old LJ and Schreiber RD: The roles of IFN gamma in protection against tumor development and cancer immunoediting. Cytokine Growth Factor Rev 13: 95-109, 2002.