

Original Research

A Novel Liver-Specific Pseudogene Biomarker, BMS1P8, for Diagnosis and Prognosis in Hepatocellular Carcinoma

Hyung Seok Kim^{1,†}, Ji Yi Choi^{2,3,†}, Geum Ok Baek², Moon Gyeong Yoon², Se Ha Jang^{2,3}, Ji Eun Han², Soon Sun Kim², Jae Youn Cheong², Jee-Yeong Jeong^{1,*}, Jung Woo Eun^{2,*}

¹Department of Biochemistry, Kosin University College of Medicine, 49267 Busan, Republic of Korea

²Department of Gastroenterology, Ajou University School of Medicine, 16499 Suwon, Republic of Korea

³Department of Biomedical Sciences, Ajou University Graduate School of Medicine, 16499 Suwon, Republic of Korea

*Correspondence: jyjeong@kosin.ac.kr (Jee-Yeong Jeong); jetaimebin@gmail.com (Jung Woo Eun)

†These authors contributed equally.

Academic Editor: Chen Li

Submitted: 19 May 2025 Revised: 5 July 2025 Accepted: 15 July 2025 Published: 29 July 2025

Abstract

Background: Hepatocellular carcinoma (HCC) is the leading cause of cancer-related mortality worldwide. Despite advances in therapeutic approaches, the lack of effective biomarkers continues to limit early detection and prognostic evaluation. Pseudogenes, once considered nonfunctional, have emerged as regulators of biological processes in tumors and as potential biomarkers. This study aimed to identify and validate BMS1 Pseudogene 8 (*BMSIP8*) as a liver-specific, clinically relevant diagnostic and prognostic biomarker in HCC. **Methods:** A comprehensive survey of pseudogene expression across different stages of liver disease was performed and validated using clinical HCC samples. Correlation, enrichment, and competing endogenous RNA (ceRNA) analyses integrating matched microRNA (miRNA)-seq and mRNA-seq were used to explore the functional networks surrounding *BMSIP8*. Public RNA-seq datasets (GSE114564, The Cancer Genome Atlas-Liver Hepatocellular Carcinoma (TCGA_LIHC)) were used to delineate differentially expressed pseudogenes, and 98 paired tumor and non-tumor tissues were assessed using quantitative reverse transcription polymerase chain reaction. Diagnostic and prognostic performances were evaluated using receiver operating characteristic curves and Kaplan–Meier statistics. **Results:** *BMSIP8* was markedly upregulated in HCC and was overexpressed in 25 other cancer types. Receiver operating characteristics analysis yielded an area under the curve of 0.81, underscoring the diagnostic utility. High *BMSIP8* expression and enrichment of cell cycle pathways were associated with poor survival. ceRNA screening revealed an inverse *BMSIP8*–miR-30c-2-3p correlation and concordant *NME/NM23* nucleoside diphosphate kinase 6 (*NME6*) upregulation, with the *BMSIP8*/miR-30c-2-3p/*NME6* triad further stratifying patient outcomes. **Conclusion:** Our findings highlight *BMSIP8* as a novel liver-specific biomarker with substantial diagnostic and prognostic value in HCC. Its diagnostic utility suggests its potential application in early detection and personalized treatment strategies, contributing to improved patient outcomes.

Keywords: hepatocellular carcinoma; pseudogene; biomarkers; diagnosis; prognosis

1. Introduction

Hepatocellular carcinoma (HCC) is one of the most common and lethal malignancies worldwide, representing a major global health burden owing to its high incidence and mortality rates [1]. It typically develops in the context of chronic liver disease, including hepatitis B virus (HBV) or hepatitis C virus (HCV) infection, alcoholic liver disease, and non-alcoholic steatohepatitis (NASH) [2–4]. Despite advances in the management of early-stage HCC with surgical resection, transplantation, and local ablative therapies, most cases are diagnosed at an advanced stage, contributing to poor long-term outcomes [5].

Thus, the early and accurate detection of HCC is imperative to improve patient survival. Although several diagnostic modalities, including imaging and serological tests, exist, these approaches have inherent limitations. Alpha-fetoprotein (AFP), the most commonly used serum

biomarker, lacks the sensitivity and specificity required for reliable screening in the general population [6]. This shortcoming underscores the urgent need for more effective molecular biomarkers that can reliably detect HCC at an early stage and provide prognostic insights.

Recently, pseudogenes have attracted increasing interest as potential cancer biomarkers. Originally viewed as nonfunctional remnants of gene duplication or retrotransposition events, pseudogenes can be transcribed and exhibit regulatory functions like non-coding RNAs [7]. Growing evidence suggests that pseudogenes are involved in diverse biological processes, including cell cycle regulation, signal transduction, and epigenetic control [8]. For example, the phosphatase and tensin homolog pseudogene 1 (*PTENP1*) has gained attention for its capacity to regulate the tumor-suppressor gene *PTEN* by functioning as a microRNA (miRNA) decoy, thereby influencing cancer cell



proliferation and survival in several malignancies [9–11]. Similarly, POU class 5 homeobox 1B (*POU5F1B*), a pseudogene of *POU5F1/OCT4*, is upregulated in gastric cancer and contributes to oncogenic behaviors [12]. Above these and other studies have underscored how pseudogenes can actively shape tumor biology and serve as potential diagnostic or prognostic markers across diverse cancer types.

In this study, we conducted a comprehensive analysis of pseudogene expression across multiple stages of liver disease and HCC, using publicly available RNA-seq datasets (GSE114564 and The Cancer Genome Atlas - Liver Hepatocellular Carcinoma (TCGA_LIHC)) and clinical samples. Our findings revealed that BMS1 Pseudogene 8 (*BMSIP8*) is highly upregulated in HCC with strong diagnostic performance and potential prognostic relevance. We further investigated its functional relationships using correlation analyses and pathway enrichment, which implicated *BMSIP8* in cell cycle regulation and underscored its importance as a candidate biomarker for HCC. Moreover, competing endogenous RNA (ceRNA) analysis suggested that *BMSIP8* may act as a molecular sponge for tumor-suppressive miR-30c-2-3p, thereby reducing the post-transcriptional repression of the oncogenic effector NME/NM23 nucleoside diphosphate kinase 6 (*NME6*). These findings provide a foundation for future studies to validate *BMSIP8*'s clinical utility and explore its mechanism of action in hepatocarcinogenesis.

2. Materials and Methods

2.1 Expression and Prognosis in Public Omics Data

We developed the GSE114564 dataset (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE114564>) by integrating RNA-seq data from previously published HCC-related studies, capturing a range of phenotypes, including normal liver tissues and various HCC subtypes [13–15]. This comprehensive dataset, comprising 39,864 genes and 7913 pseudogenes, was used to characterize the progression from normal liver (NL) to advanced HCC (aHCC), encompassing intermediate stages such as chronic hepatitis (CH), liver cirrhosis (LC), dysplastic nodules (DN), and early-stage HCC (eHCC). Differential expression analysis was performed to identify pseudogenes significantly upregulated ($\log_2$ fold change [FC] ≥ 0.5 and $*p < 0.05$) during the progression from NL to HCC.

Expression levels in non-tumor (NT) versus tumor (T) samples were further assessed using TCGA_LIHC dataset ([https://xenabrowser.net/datapages/?cohort=GDC%20TCGA%20Liver%20Cancer%20\(LIHC\)&removeHub=https%3A%2F%2Fxcna.treehouse.gi.ucsc.edu%3A443](https://xenabrowser.net/datapages/?cohort=GDC%20TCGA%20Liver%20Cancer%20(LIHC)&removeHub=https%3A%2F%2Fxcna.treehouse.gi.ucsc.edu%3A443)), followed by overall survival (OS) and disease-free survival (DFS) analyses to evaluate prognostic relevance. Here, OS was defined as the time from HCC diagnosis to death from any cause, and DFS was defined as the time from curative treatment to disease recurrence. Additionally, pan-cancer expression data from the TCGA database were examined

to assess whether *BMSIP8* expression was specific to liver cancer or also present in other malignancies. These data were downloaded from the Genomic Data Commons (GDC) hub of UCSC Xena (<https://xena.ucsc.edu/>) to ensure comprehensive coverage and accessibility [16].

2.2 Quantitative Reverse Transcription PCR (qRT-PCR)

Total RNA was extracted from frozen tissues using QIAzol Reagent (Qiagen, Cat# 79306, Hilden, Germany), following the manufacturer's instructions. cDNA was synthesized from 500 ng of total RNA using 5× PrimeScript™ RT Master Mix (Takara Bio, Cat# RR036A, Shiga, Japan) under the following conditions: 37 °C for 15 min, 85 °C for 5 s, and 4 °C. qRT-PCR was performed with amfiSure qGreen Q-PCR Master Mix (GenDEPOT, Barker, Cat# Q5602, TX, USA) on a CFX Connect Real-Time PCR Detection System (Bio-Rad Laboratories, Hercules, CA, USA). Expression levels were normalized to hydroxymethylbilane synthase (*HMBS*) as an internal control. The amplification *BMSIP8* primer sequences were 5'-GCACATTCCAAAAGCCTTGC-3' (forward) and 5'-TGTGCACCATACTCAGTGCA-3' (reverse); and the *HMBS* primer sequences were 5'-ACGGCTCAGATAGCATACAAGAG-3' (forward) and 5'-GTTACGAGCAGTGATGCCTACC-3' (reverse). The PCR conditions were as follows: 95 °C for 2 min, 40 cycles of 95 °C for 15 s, 58 °C for 34 s, and 72 °C for 30 s, followed by a dissociation stage at 95 °C for 10 s, 65 °C for 5 s, and 95 °C for 5 s. The relative standard curve method ($2^{-\Delta\Delta C_t}$) was used to determine the relative expression. All experiments were performed at least three times.

2.3 Clinical Sample Collection

To validate the *BMSIP8* expression patterns identified from public omics data, 98 paired HCC and corresponding non-cancerous liver tissues were obtained from the Biobank of Ajou University Hospital (Suwon, South Korea). qRT-PCR analysis was carried out as described above. Demographic and clinical information, including age, sex, etiology of liver disease, body mass index (BMI), platelet count, serum albumin, total bilirubin, international normalized ratio (INR), creatinine, sodium, aspartate aminotransferase (AST), alanine aminotransferase (ALT), AFP, protein induced by vitamin K absence-II (PIVKA-II), hemoglobin, glucose, total cholesterol levels, and the presence of ascites, was recorded (Table 1).

2.4 Enrichment Analysis Using Databases

Gene Ontology (GO) analyses of the enriched genes were performed using the enrichGO function in the R clusterProfiler package (v3.18.1, Bioconductor; <https://bioconductor.org/packages/clusterProfiler>). Enrichment analysis of the MSigDB Hallmark 2020 database (Broad Institute, Cambridge, MA, USA; <https://www.gsea-msigdb.org/gsea/msigdb>) was conducted using the enrichr function in the

Table 1. Clinical characteristics of the HCC cohort.

Variables	HCC (n = 98)
Age (years), mean ± SD (range)	56 ± 10.2 (29–75)
Male sex, n (%)	74 (75.5)
Etiology, n (%)	—
HBV	91 (92.9)
HCV	4 (4.1)
Alcohol	2 (2)
HCV + Alcohol	1 (1)
Cirrhosis, n (%)	71 (72.4)
Ascites, n (%)	15 (15.3)
BMI (kg/m ²), mean ± SD (range)	24.1 ± 3.8 (16.6–36.0)
Platelet, ×10 ⁹ /L, mean ± SD (range)	178.4 ± 68.5 (58–534)
Albumin (g/dL), mean ± SD (range)	4.5 ± 0.6 (0.3–5.4)
Total bilirubin (mg/dL), mean ± SD (range)	0.8 ± 1.3 (0.1–12.9)
INR, mean ± SD (range)	1.1 ± 0.1 (0.91–1.94)
Creatinine (mg/dL), mean ± SD (range)	0.92 ± 0.2 (0.5–2.01)
Sodium (mmol/L), mean ± SD (range)	139.3 ± 2.2 (130–143)
AST (U/L), mean ± SD (range)	43.5 ± 57.8 (15–541)
ALT (U/L), mean ± SD (range)	40.5 ± 42.1 (8–332)
AFP (ng/mL), mean ± SD (range)	3185.2 ± 9989.2 (1.3–60,500)
PIVKA-II (mAU/mL), mean ± SD (range)	8711.3 ± 34,086.5 (13–300,000)
Hemoglobin (g/dL), mean ± SD (range)	14.4 ± 1.6 (8.2–17.1)
Glucose (mg/dL), mean ± SD (range)	123.0 ± 53.5 (72–411)
Total cholesterol (mg/dL), mean ± SD (range)	170.4 ± 41.0 (97–289)

AFP, alpha-fetoprotein; ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; HCV, hepatitis C virus; INR, international normalized ratio; PIVKA-II, protein induced by vitamin K absence-II.

R enrichR package (v3.0; <https://cran.r-project.org/package=enrichR>). Additional pathway analyses, including the Kyoto Encyclopedia of Genes and Genomes (KEGG) 2021 Human (Kanehisa Laboratories, Kyoto, Japan; <https://www.genome.jp/kegg/>) and Reactome Pathways 2024 (Ontario Institute for Cancer Research, Toronto, Canada; <https://reactome.org>), were also carried out with clusterProfiler to identify biological processes and signaling pathways potentially associated with *BMSIP8*.

2.5 Risk Score Calculation for the 7-Gene Signature

A composite risk score was calculated by standardizing expression levels of the seven genes (*BMSIP8*, *CCNB2*, *CDC20*, *CDC45*, *ESPL1*, *PLK1*, and *PTTG1*) using Z-score transformation across the TCGA_LIHC cohort. Z-scores were computed for each gene as follows:

$$Z = \frac{(X - \mu)}{\sigma}$$

where X is the individual gene expression, μ is the mean expression, and σ is the standard deviation across the cohort.

The individual risk score for each patient was then defined as the arithmetic mean of the Z-scores of the seven genes:

$$\text{Risk score} = \left(\frac{1}{7}\right) \sum_{i=1}^7 Z_i$$

This composite metric integrates the combined expression pattern into a single prognostic variable. Receiver operating characteristic (ROC) curve analysis was subsequently performed using these risk scores to evaluate the diagnostic and prognostic performance of the 7-gene signature.

2.6 miRNA Expression Profiling and ceRNA Network Construction

Raw mature-miRNA read counts for TCGA_LIHC were downloaded from the GDC using the TCGAAbiolinks pipeline (v2.30.1, Bioconductor; <https://bioconductor.org/packages/TCGAbiolinks>) [17,18]. Counts were filtered to retain miRNAs expressed at ≥ 1 count per million (CPM) in ≥ 30% of samples, then TMM-normalized and transformed to log₂-CPM with the edgeR package (v4.2, Bioconductor;

<https://bioconductor.org/packages/edgeR>). Differential expression between non-tumor (NT) and tumor (T) tissues was assessed with limma-voom (v3.58, Bioconductor; <https://bioconductor.org/packages/limma>); miRNAs with $|\log_2 \text{FC}| \geq 0.5$ and false discovery rate (FDR) < 0.05 were considered significantly deregulated.

For ceRNA screening, Pearson correlation coefficients (r) were calculated between *BMSIP8* and each miRNA across matched mRNA- and miRNA-seq tumor samples. Candidate tumor-suppressive miRNAs were defined as (i) downregulated in tumors ($\log_2 \text{FC} \leq -0.5$, FDR < 0.05) and (ii) inversely correlated with *BMSIP8* ($r \leq -0.20$, $*p < 0.05$). Putative mRNA targets of each candidate miRNA were retrieved from miRDB (<https://mirdb.org/>) with a target score ≥ 70 [19]. Predicted targets were filtered to keep transcripts that were (i) upregulated in tumors ($\log_2 \text{FC} \geq 0.5$, FDR < 0.05) and (ii) positively correlated with *BMSIP8* ($r \geq 0.20$, $*p < 0.05$) while showing a negative correlation with the cognate miRNA ($r \leq -0.20$, $*p < 0.05$).

2.7 Statistical Analysis

All results are expressed as mean $\pm$ standard deviation (SD). Unpaired Student's *t*-tests (GraphPad Software, version 10.0, San Diego, CA, USA) were used to determine differences between groups. Kaplan–Meier survival curves were generated for both OS and DFS, with significance assessed via the log-rank test. ROC curve analyses were conducted using MedCalc (MedCalc Software Ltd., version 22.0, Ostend, Belgium), providing the area under the curve (AUC), 95% confidence intervals (CIs), sensitivity, and specificity. Statistical significance was defined as $p < 0.05$. All experiments were repeated at least three times.

3. Results

3.1 Comprehensive Analysis of Pseudogene Expression in Liver Disease and HCC Progression

Using the multistage liver disease and cancer dataset GSE114564, RNA sequencing analysis was conducted to systematically evaluate gene expression changes across different stages of liver disease and cancer progression. A total of 39,864 genes were analyzed, classified into eight categories based on their coding potential: protein-coding genes (46.20%), long non-coding RNA (lncRNA) genes (13.09%), pseudogenes (19.85%), antisense transcripts (10.35%), miscellaneous RNA (2.58%), sense intronic RNA (2.16%), and processed transcripts (1.10%) (Fig. 1A, left pie chart). Among these, 7913 pseudogenes were identified and further categorized based on their biogenesis and characteristics. The majority, 71.06%, were processed pseudogenes, which are retro-transposed copies of functional genes that have lost their coding potential. The second-largest group, 11.03%, included unprocessed pseudogenes that retain introns and resemble their parent genes. Additionally, 6.46% were transcribed unprocessed

pseudogenes, and 4.40% were transcribed processed pseudogenes. Smaller fractions included 3.48% generic pseudogenes, 1.58% IG_V pseudogenes, and 1.36% unitary pseudogenes. The remaining 0.63% were classified as etc., encompassing pseudogenes with ambiguous or uncommon features (Fig. 1A, right pie chart).

Based on these classifications, the hepatic tissue samples were categorized into six groups to represent the different stages of liver disease and cancer progression: NL, CH, LC, DN, eHCC, and aHCC (Fig. 1B). Heatmap analysis of 171 differentially expressed pseudogenes revealed dynamic expression patterns across these stages, with 98 pseudogenes significantly upregulated in advanced HCC (Fig. 1B). Notably, *BMSIP8*, *RP11-390F10.3*, *RP11-443P15.2*, and *ZNF192P1* showed consistent upregulation during disease progression, highlighting their potential role in driving HCC development (Fig. 1C). Although *PTGES3P1* clustered with other upregulated pseudogenes in early- and late-stage HCC (Fig. 1C), further analysis indicated it may not be a robust liver cancer-specific biomarker. *PTGES3P1* showed elevated expression in the chronic hepatitis (CH) group, suggesting non-specific upregulation. Analysis of variance (ANOVA) with Tukey's multiple comparisons test revealed significant differences only between normal liver and aHCC and between LC and aHCC, without consistent stepwise increases in HCC progression. Moreover, ROC analysis yielded an AUC below 0.7 for distinguishing tumor from non-tumor tissues, indicating limited diagnostic potential. Based on these findings, we concluded that *PTGES3P1* does not meet the criteria for a promising HCC-specific biomarker (**Supplementary Fig. 1A**). These findings underscore the importance of pseudogene expression in the molecular landscape of liver cancer progression.

3.2 Diagnostic and Prognostic Significance of the Candidate Pseudogenes in HCC Progression

The expression patterns of four pseudogenes, *BMSIP8*, *RP11-390F10.3*, *RP11-443P15.2*, and *ZNF192P1*, were analyzed across six stages of liver disease progression (NL, CH, LC, DN, eHCC, and aHCC) using the GSE114564 dataset. Among these, *BMSIP8*, *RP11-443P15.2*, and *ZNF192P1* exhibited progressive and statistically significant increases in expression as the disease advanced, reaching peak levels in advanced HCC. In contrast, *RP11-390F10.3* showed a less pronounced and statistically non-significant increase across the stages (Fig. 2A, left panels for each pseudogene).

The diagnostic potential was assessed through ROC curve analysis. Among the four pseudogenes, *BMSIP8* showed strong diagnostic performance with an AUC of 0.81 (95% CI: 0.73–0.89, $p < 0.0001$), and *RP11-443P15.2* exhibited the highest AUC value of 0.84 (95% CI: 0.77–0.92, $p < 0.0001$). *ZNF192P1* also demonstrated strong diagnostic capability with an AUC of 0.81 (95% CI: 0.73–0.89, $p < 0.0001$), while *RP11-390F10.3* had moderate diagnos-

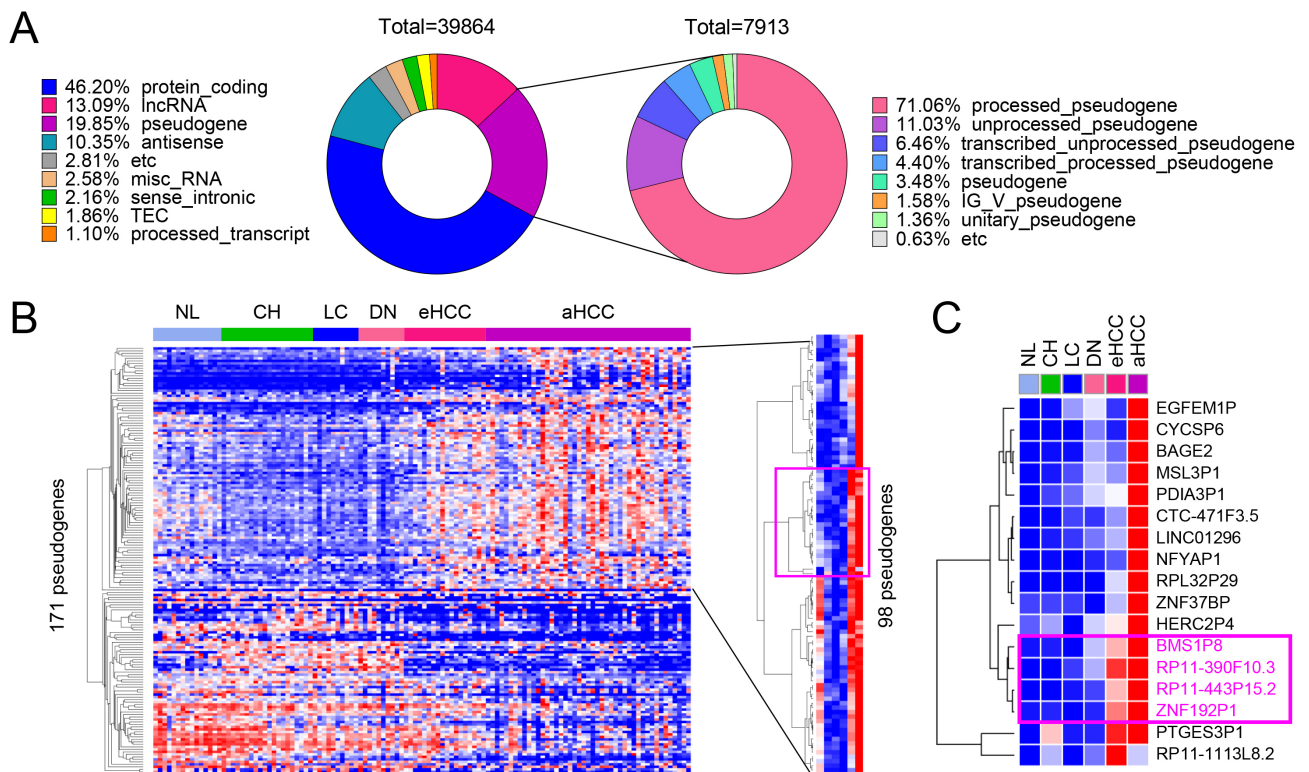


Fig. 1. Identification of differentially expressed pseudogenes in HCC. (A) Gene composition of the GSE114564 dataset, including 39,864 genes, with 7913 pseudogenes. (B) Heatmap of 171 pseudogenes showing differential expression across liver disease stages including non-cancerous liver (NL, normal liver; CH, chronic hepatitis; LC, liver cirrhosis; DN, dysplastic nodule) and tumor (eHCC, early-stage HCC; aHCC, advanced HCC) tissues, with 98 pseudogenes significantly upregulated during HCC progression. (C) Key upregulated pseudogenes, including *BMS1P8*, *RP11-390F10.3*, *RP11-443P15.2*, and *ZNF192P1*, identified as potential HCC diagnostic markers.

tic potential (AUC = 0.66, 95% CI: 0.56–0.76, $p = 0.003$) (Fig. 2A, right panels for each pseudogene).

To validate tumor-specific expression, these pseudogenes were analyzed in T ($n = 371$) versus NT ($n = 50$) tissues using the TCGA_LIHC dataset. *BMS1P8* and *RP11-443P15.2* were significantly upregulated in tumor tissues, with fold changes (FCs) of 2.1 and 2.3, respectively ($p < 0.001$ for both). *ZNF192P1* showed a mild but statistically significant increase (FC = 1.1, $p < 0.001$), whereas *RP11-390F10.3* did not show significant differential expression (Fig. 2B).

Kaplan–Meier survival analysis revealed the prognostic significance of *BMS1P8* and *RP11-443P15.2*. High expression of *BMS1P8* was significantly associated with poorer OS, with a hazard ratio (HR) of 1.64 (95% CI: 1.16–2.31, $p = 0.005$) (Fig. 2C, left). While *RP11-443P15.2* showed a trend toward worse survival, its result was not statistically significant (HR = 1.36, 95% CI: 0.96–1.91, $p = 0.08$) (Fig. 2C, right). These findings highlight *BMS1P8* as a strong candidate for further investigation owing to its diagnostic and prognostic value in liver disease and HCC progression.

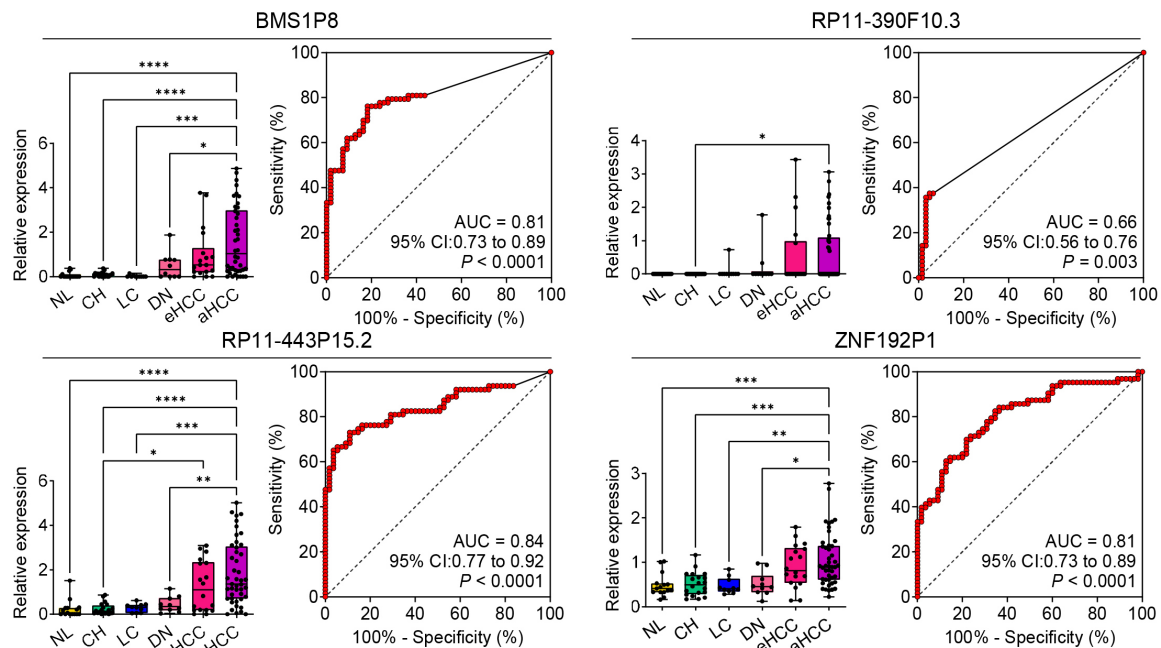
3.3 Comparative Analysis of BMS1 Pseudogenes in HCC and Identification of BMS1P8 as a Leading Diagnostic Candidate

To evaluate the diagnostic potential of BMS1-derived pseudogenes in HCC, matched T and NT tissue pairs ($n = 50$) were examined to identify the pseudogenes that were significantly upregulated in cancerous liver tissues. Among the 17 tested genes, eight, including *BMS1P1*, *BMS1P2*, *BMS1P4*, *BMS1P8*, *BMS1P10*, *BMS1P16*, *BMS1P17*, and *BMS1P20*, exhibited markedly higher expression levels in tumor tissues than those in their NT counterparts (Fig. 3A).

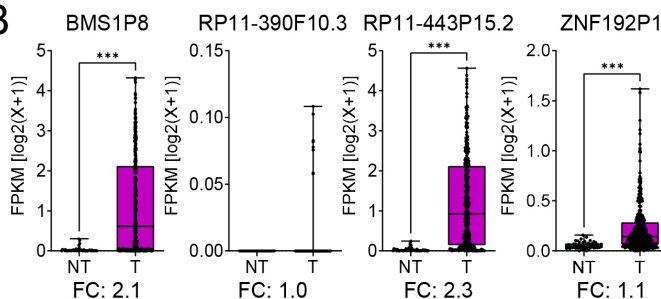
ROC curve analyses were then performed to assess the diagnostic performance of each pseudogene. While several BMS1 pseudogenes demonstrated moderate diagnostic capabilities (AUC = 0.58–0.74), both *BMS1P8* and *BMS1P20* showed particularly high AUC values of 0.80 (95% CI: 0.71–0.89) and 0.84 (95% CI: 0.76–0.92), respectively, indicating strong potential for HCC detection (Fig. 3B).

Although *BMS1P20* appeared to be a strong candidate based on the preliminary findings, further validation using a dataset encompassing multistage liver disease and HCC progression (GSE114564) revealed limited diagnostic relevance. *BMS1P20* did not exhibit a statistically sig-

A



B



C

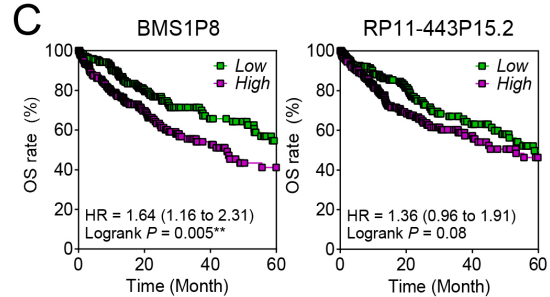


Fig. 2. Diagnostic and prognostic significance of four pseudogenes in hepatocellular carcinoma (HCC) progression. (A) Relative expression levels of *BMS1P8*, *RP11-390F10.3*, *RP11-443P15.2*, and *ZNF192P1* in NL, CH, LC, DN, eHCC, and aHCC, based on the GSE114564 dataset (the left panels for each pseudogene). Receiver operating characteristic (ROC) curves demonstrate the diagnostic performance of each pseudogene in distinguishing HCC from non-tumor liver tissue, with area under the curve (AUC) values and 95% confidence intervals (CIs) (the right panels for each pseudogene). (B) Validation of tumor-specific expression for the four pseudogenes in paired non-tumor (NT) and tumor (T) tissues from The Cancer Genome Atlas - Liver Hepatocellular Carcinoma (TCGA_LIHC) dataset (n = 421). The y-axis represents fragments per kilobase of transcript per million mapped reads (FPKM) on a $\log_2(x + 1)$ scale, highlighting fold changes (FC). (C) Kaplan-Meier overall survival (OS) analyses for *BMS1P8* and *RP11-443P15.2* in the TCGA_LIHC dataset. High-expression groups (purple) show worse survival compared to low-expression groups (green). Hazard ratios (HRs), 95% CIs, and log-rank *p* values are presented. Statistically significant differences were determined using the log-rank test; **p* < 0.05, ***p* < 0.01, ****p* < 0.001, *****p* < 0.0001. Data are shown as mean $\pm$ SD.

nificant differential expression between non-cancerous and cancerous tissues, and its AUC value for distinguishing these groups was relatively low (AUC = 0.60, *p* = 0.06) (**Supplementary Fig. 1**). In contrast, *BMS1P8* consistently demonstrated robust diagnostic performance across both datasets and maintained significant differences in expression between the T and NT samples (Fig. 2A). Collectively, these findings highlight *BMS1P8* as the most promising BMS1 pseudogene marker for HCC diagnosis, underscoring its potential utility for early detection and guiding future investigations into the clinical implications of pseudogene-based biomarkers.

3.4 *BMS1P8* is Overexpressed in HCC and Exhibits Liver-Specific Diagnostic Potential

To investigate the potential cancer-specific expression of *BMS1P8*, we initially performed a broad pan-cancer survey using TCGA database, spanning 25 different tumor types. *BMS1P8* expression was largely undetectable or remained at very low levels in most cancer types. In contrast, LIHC samples exhibited a pronounced increase in *BMS1P8* expression relative to NT tissues (Fig. 4A). This discrepancy underscores the possibility that *BMS1P8* plays a role in the oncogenic processes of the liver rather than in a broad spectrum of malignancies.

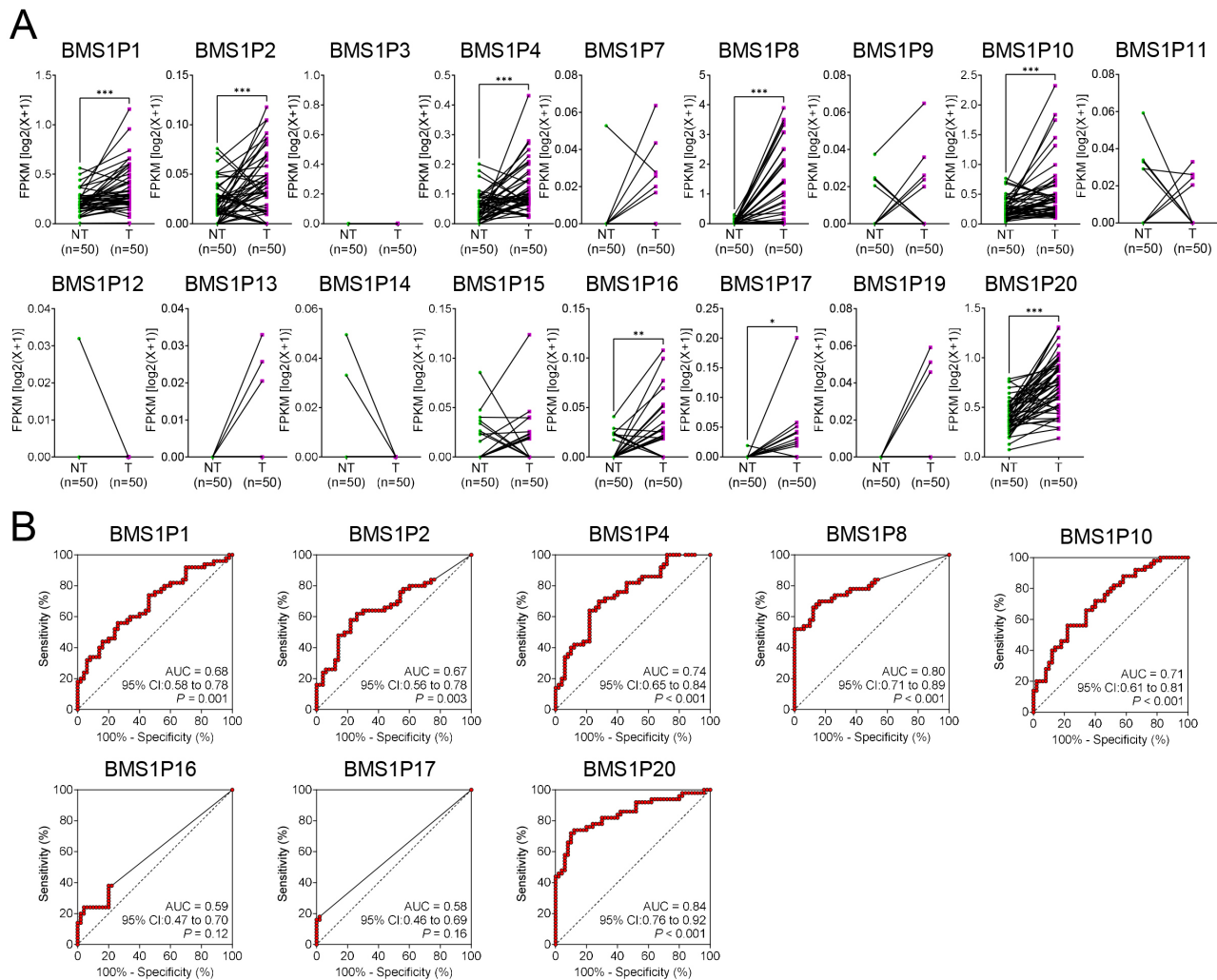


Fig. 3. Comparative expression and diagnostic evaluation of *BMS1* pseudogenes in matched HCC and non-tumor liver tissues. (A) Expression patterns of 17 *BMS1* pseudogenes in paired non-tumor (NT) and tumor (T) samples ($n = 50$). Each connected line represents an individual patient sample. The y-axis shows fragments per kilobase of transcript per million mapped reads (FPKM) on a $\log_2(x + 1)$ scale. (B) Receiver operating characteristic (ROC) curves evaluating the diagnostic performance of the eight most upregulated *BMS1* pseudogenes (*BMS1P1*, *BMS1P2*, *BMS1P4*, *BMS1P8*, *BMS1P10*, *BMS1P16*, *BMS1P17*, and *BMS1P20*) in distinguishing tumor from non-tumor tissues. Area under the curve (AUC), 95% confidence interval (CI), and corresponding p values are shown for each pseudogene. Higher AUC values indicate stronger diagnostic potential. Statistical significance levels (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$) are indicated where applicable.

To confirm these pan-cancer observations at the clinical level, paired tumor and NT liver tissues were collected from 98 patients with HCC undergoing hepatectomy. The relevant clinical information is detailed in Table 1. qRT-PCR revealed that 80 of 98 (82%) patient samples demonstrated significantly elevated *BMS1P8* expression (Fig. 4B), thus reinforcing the findings from both the TCGA and GSE114564 datasets. The high proportion of overexpressing cases suggests that *BMS1P8* may be functionally relevant to HCC pathogenesis or tumor progression.

In alignment with these expression data, *BMS1P8* exhibited robust diagnostic performance in distinguishing HCC tissues from non-tumor tissues. Specifically, ROC

curve analysis yielded an AUC of 0.81 (95% CI: 0.74–0.89, $p < 0.001$) (Fig. 4C), indicating a high degree of sensitivity and specificity. This diagnostic potential complements the observed overexpression patterns, indicating that *BMS1P8* may serve as a practical biomarker for the early detection or clinical monitoring of HCC.

3.5 *BMS1P8* May Regulate HCC Prognosis Through Interactions With Cell Cycle-Related Genes

To elucidate the functional importance of *BMS1P8* in HCC, correlation analysis in the TCGA_LIHC dataset identified 1784 genes ($|r| \geq 0.2$), including 1175 protein-coding genes, that were associated with *BMS1P8*. Pathway enrichment analysis using EnrichR and referencing the MSigDB

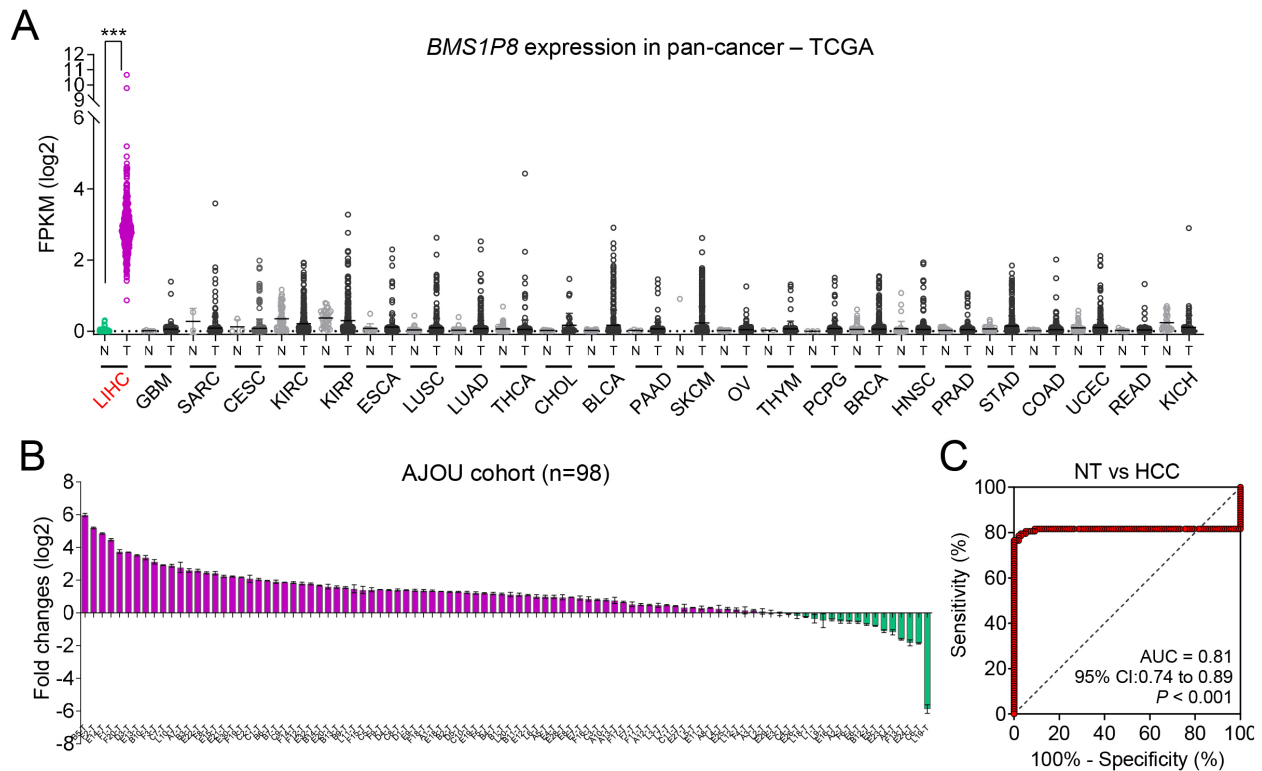


Fig. 4. *BMS1P8* exhibits liver-specific overexpression and robust diagnostic potential in an independent HCC cohort. (A) Pan-cancer analysis of *BMS1P8* expression across 25 tumor types in the TCGA dataset. The y-axis indicates FPKM on a $\log_2$ scale, highlighting that *BMS1P8* is predominantly overexpressed in LIHC compared with other malignancies. (B) Fold change ($\log_2$) of *BMS1P8* expression in 98 paired HCC and non-tumor liver tissues from the Ajou University cohort. Bars above the x-axis indicate samples with upregulated *BMS1P8*, while those below represent downregulation. (C) ROC curve evaluating *BMS1P8* as a diagnostic marker for distinguishing HCC from NT tissues in the Ajou University cohort. The area under the curve (AUC) is 0.81 (95% CI: 0.74–0.89, $p < 0.001$), underscoring its potential clinical utility. Data are shown as mean $\pm$ SD. LIHC, Liver hepatocellular carcinoma; GBM, Glioblastoma multiforme; SARC, Sarcoma; CESC, Cervical squamous cell carcinoma and endocervical adenocarcinoma; KIRC, Kidney renal clear cell carcinoma; KIRP, Kidney renal papillary cell carcinoma; ESCA, Esophageal carcinoma; LUSC, Lung squamous cell carcinoma; LUAD, Lung adenocarcinoma; THCA, Thyroid carcinoma; CHOL, Cholangiocarcinoma; BLCA, Bladder urothelial carcinoma; PAAD, Pancreatic adenocarcinoma; SKCM, Skin cutaneous melanoma; OV, Ovarian serous cystadenocarcinoma; THYM, Thymoma; PCPG, Pheochromocytoma and paraganglioma; BRCA, Breast invasive carcinoma; HNSC, Head and neck squamous cell carcinoma; PRAD, Prostate adenocarcinoma; STAD, Stomach adenocarcinoma; COAD, Colon adenocarcinoma; UCEC, Uterine corpus endometrial carcinoma; READ, Rectum adenocarcinoma; KICH, Kidney chromophobe.

Hallmark 2020, KEGG 2021 Human, and Reactome Pathways 2024 databases consistently highlighted the G2–M checkpoint and cell cycle pathways among the top-ranked categories (Fig. 5A–C). This overarching pattern suggests that *BMS1P8* orchestrates pivotal cell-cycle processes involved in HCC progression.

A Venn diagram analysis focusing on cell cycle-related genes correlated with *BMS1P8* across the three databases uncovered six overlapping genes, such as *CCNB2*, *CDC20*, *CDC45*, *ESPL1*, *PLK1*, and *PTTG1* (Fig. 5D). Subsequent validation within the TCGA_LIHC dataset confirmed significant positive correlations ($r > 0.2$) between *BMS1P8* and each of these genes (Fig. 5E), reinforcing the notion that *BMS1P8* functions as a central node in cell cycle regulation.

To assess prognostic implications, *BMS1P8* was combined with the six correlated cell cycle genes to form a 7-gene signature (7 sigs), comprising *BMS1P8* plus the six correlated genes including *CCNB2*, *CDC20*, *CDC45*, *ESPL1*, *PLK1*, and *PTTG1*. Kaplan–Meier analyses of OS and DFS revealed significantly worse outcomes in patients exhibiting high expression of this signature (Fig. 5F). Specifically, OS analysis yielded a log-rank $p = 6.2 \times 10^{-6}$ with a HR = 3, while DFS analysis produced a log-rank $p = 9.2 \times 10^{-5}$ and an HR = 2.3. Collectively, these results underscore the critical role of *BMS1P8* in modulating cell cycle-associated pathways and highlight the potential of this 7-gene signature as a robust prognostic marker for HCC. To further evaluate the diagnostic capability of the 7-gene signature, ROC curve analyses were performed in

two clinically relevant comparisons. When comparing all available non-tumor samples (NT, $n = 50$) with all tumor samples (T, $n = 371$) in the TCGA_LIHC cohort, the signature achieved an AUC of 0.92 (95% CI: 0.89–0.95, $p < 0.0001$). Additionally, analysis of 50 paired NT and T tissues from the same patients demonstrated an AUC of 0.98 (95% CI: 0.95–1.00, $p < 0.0001$), indicating strong discriminatory performance even in matched samples (Fig. 5G). When comparing the 7-gene signature with *BMSIP8* alone, the 7-gene signature showed significantly improved diagnostic performance, with a higher AUC of 0.92 compared to *BMSIP8*'s AUC of 0.80 (Fig. 3B). Collectively, these results underscore the critical role of *BMSIP8* in modulating cell cycle-associated pathways and highlight the potential of this 7-gene signature as both a robust prognostic marker and a promising diagnostic biomarker panel for HCC.

3.6 *BMSIP8* Functions as a ceRNA Modulating the miR-30c-2-3p–*NME6* Axis in HCC

Because pseudogenes, like lncRNAs, are well known to function as ceRNAs modulating miRNA availability and downstream gene expression, we investigated whether *BMSIP8* might engage in a ceRNA regulatory network influencing HCC progression [10]. This analysis aimed to explore a potential mechanistic link by identifying miRNAs that could interact with *BMSIP8* and affect expression of relevant oncogenic targets. Following the analytical workflow (Fig. 6A), miRNAs sharing complementary sequences with *BMSIP8* were identified using BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) and miRNA-target prediction from miRDB. Among the seven downregulated miRNAs that showed negative correlations with *BMSIP8* in HCC (Supplementary Table 1, Supplementary Fig. 2), we focused on hsa-miR-30c-2-3p, which showed the most significantly reduced expression in HCC and the strongest negative correlation with *BMSIP8*. miRDB prediction yielded 267 target genes with a target score ≥ 70 , and among them, 46 genes were found to be upregulated in tumors and positively correlated with *BMSIP8* in the TCGA_LIHC cohort (Supplementary Table 2). Among these, the only gene that showed both a significant negative correlation with hsa-miR-30c-2-3p ($r \leq -0.2$, $p < 0.05$) and a significant positive correlation with *BMSIP8* ($r \geq 0.3$, $p < 0.05$) was *NME6* (Fig. 6B). Expression analysis showed that hsa-miR-30c-2-3p was significantly downregulated in HCC ($p < 0.0001$), with a diagnostic AUC of 0.79 (95% CI: 0.75–0.84), while *NME6* was markedly upregulated ($p < 0.0001$) with an AUC of 0.97 (95% CI: 0.95–0.99) (Fig. 6C). Furthermore, to validate these findings in an independent in-house cohort, we analyzed RNA-seq and miRNA-seq data from the KOSIN cohort. Consistent with TCGA results, hsa-miR-30c-2-3p expression was significantly reduced in HCC ($p = 0.004$), yielding a diagnostic AUC of 0.66 (95% CI: 0.56–0.76), while *NME6* expression was significantly elevated ($p = 0.01$) with a diagnostic AUC of 0.64 (95% CI: 0.53–

0.74) (Fig. 6D). Correlation analysis in the TCGA_LIHC dataset revealed significant inverse associations between miR-30c-2-3p and both *NME6* ($r = -0.26$, $p < 0.001$) and *BMSIP8* ($r = -0.29$, $p < 0.001$), along with a positive correlation between *BMSIP8* and *NME6* ($r = 0.33$, $p < 0.001$). These findings suggest a potential ceRNA regulatory network among the three molecules. To independently validate these associations, we performed correlation analysis using RNA-seq and miRNA-seq data from the KOSIN cohort, which confirmed similar trends: miR-30c-2-3p was inversely correlated with both *NME6* ($r = -0.36$, $p < 0.001$) and *BMSIP8* ($r = -0.20$, $p = 0.03$), while *BMSIP8* showed a significant positive correlation with *NME6* ($r = 0.29$, $p < 0.001$) (Fig. 6E). Although correlation coefficients in the KOSIN cohort were comparable to those observed in the TCGA dataset, this additional analysis provides robust, independent support for the proposed ceRNA axis involving *BMSIP8*, miR-30c-2-3p, and *NME6* in HCC. In survival analysis, high expression of *NME6* was significantly associated with worse OS (HR = 2.02, 95% CI: 1.41–2.89, $p < 0.001$), while low expression of miR-30c-2-3p showed a non-significant trend toward poor prognosis (HR = 1.27, $p = 0.19$) (Fig. 6F). Notably, patients with high expression of both *BMSIP8* and *NME6* showed markedly poorer OS and DFS compared to other groups (OS: HR = 2.75, 95% CI: 1.74–4.34; DFS: HR = 2.28, 95% CI: 1.53–3.39, both $p < 0.001$) (Fig. 6G). A mechanistic model summarizing these findings is presented in Fig. 7, where in normal hepatocytes, miR-30c-2-3p is sufficiently expressed to repress *NME6* mRNA via RNA-induced silencing complex (RISC) complex formation, whereas in HCC, increased *BMSIP8* sequesters miR-30c-2-3p, preventing RISC-mediated repression and thereby releasing *NME6* from post-transcriptional inhibition, potentially contributing to tumor progression.

4. Discussion

Pseudogenes are traditionally considered junk DNA or nonfunctional remnants arising from gene duplication or retrotransposition events [20]. Their sequences typically harbor premature stop codons or frame shifts, preventing the production of functional proteins [21]. However, with the advent of high-throughput technologies, including next-generation RNA sequencing, single-cell transcriptomics, and clustered regularly interspaced short palindromic repeat-based functional screens, our understanding of these genetic elements has evolved substantially [22,23]. This paradigm shift has been supported by emerging research data: many pseudogenes are transcriptionally active and can exert regulatory functions analogous to non-coding RNAs, influencing fundamental cellular processes including proliferation, apoptosis, and metastasis [8]. For instance, the pseudogene-expressed lncRNA small ubiquitin-like modifier 1 pseudogene 3 (*SUMO1P3*) has been identified as upregulated in gastric cancer tissues compared with adjacent nontumorous tissues, and

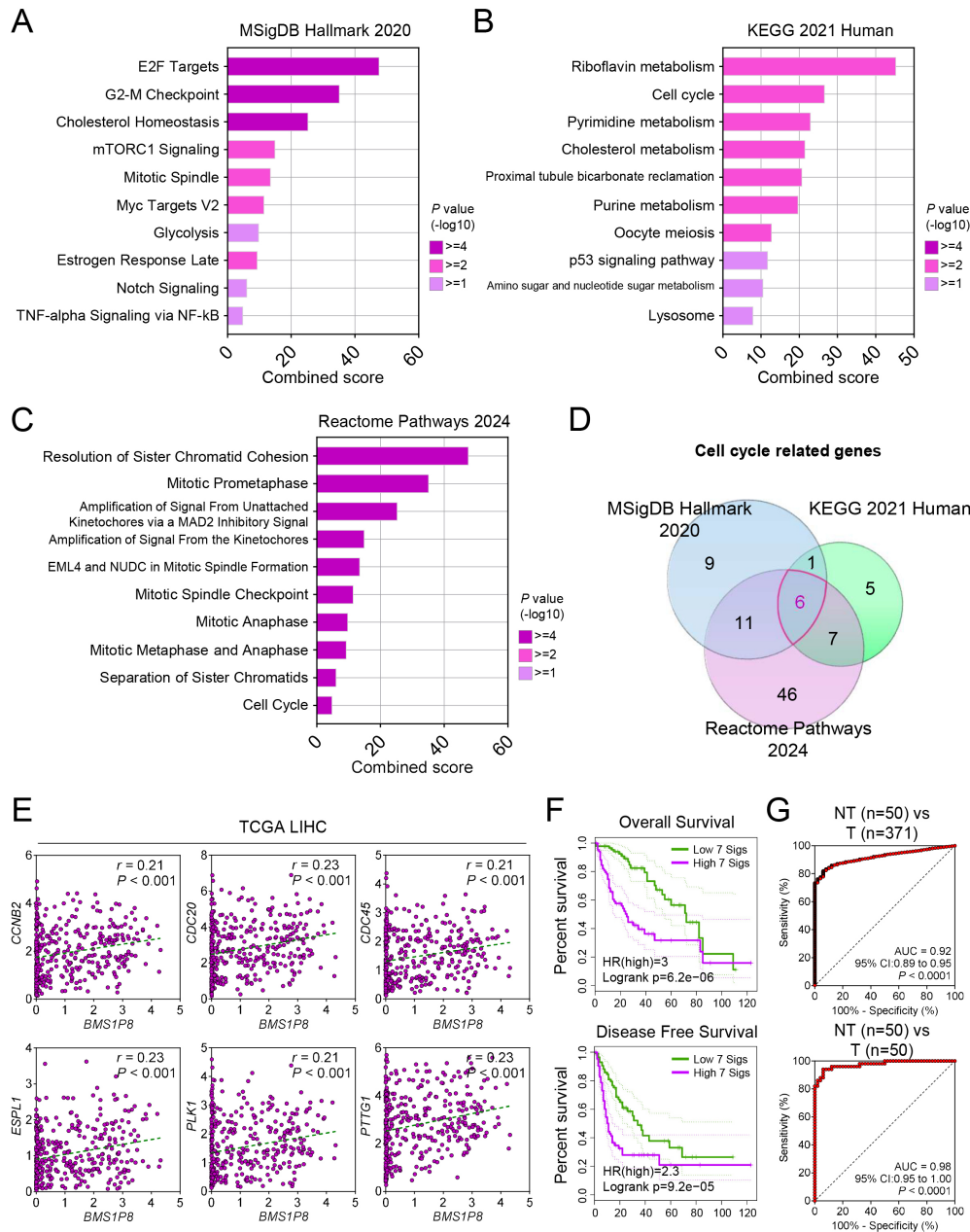


Fig. 5. *BMS1P8* and its cell cycle-related 7-gene signature predict prognosis and accurately distinguish HCC tissues. Pathway enrichment analyses of *BMS1P8*-correlated genes ($|r| \geq 0.2$) using three different databases: (A) MSigDB Hallmark 2020, (B) KEGG 2021 Human, and (C) Reactome Pathways 2024. Bar graphs depict the top enriched pathways based on combined scores and adjusted p values. Of note, G2-M checkpoint and cell cycle processes consistently emerge among the highest-ranked categories. (D) Venn diagram analysis illustrating the overlap of cell cycle-associated genes correlated with *BMS1P8* across the three databases. Six genes, including *CCNB2*, *CDC20*, *CDC45*, *ESPL1*, *PLK1*, and *PTTG1*, are commonly identified. (E) Scatter plots showing significant positive correlations ($r > 0.2$, $p < 0.001$) between *BMS1P8* and the six overlapping cell cycle genes in the TCGA_LIHC dataset, suggesting *BMS1P8* as a regulatory node in cell cycle control. (F) Kaplan–Meier curves for overall survival (OS) and disease-free survival (DFS) in patients stratified by a 7-gene signature (comprising *BMS1P8* plus the six correlated genes including *CCNB2*, *CDC20*, *CDC45*, *ESPL1*, *PLK1*, and *PTTG1*). High 7-gene signature expression (purple) is associated with markedly poorer OS and DFS. Log-rank p values and hazard ratios (HR) are presented, highlighting the prognostic relevance of *BMS1P8* and its cell cycle-related partners. (G) Receiver operating characteristic (ROC) curves illustrating the diagnostic accuracy of the 7-gene signature, calculated as the mean of standardized expression (Z-scores) of *BMS1P8* and six correlated cell cycle genes (*CCNB2*, *CDC20*, *CDC45*, *ESPL1*, *PLK1*, *PTTG1*). (Top) ROC curve comparing all non-tumor samples (NT, $n = 50$) to all tumor samples (T, $n = 371$) in the TCGA_LIHC cohort. (Bottom) ROC curve for 50 paired NT and T tissues from the same patients.

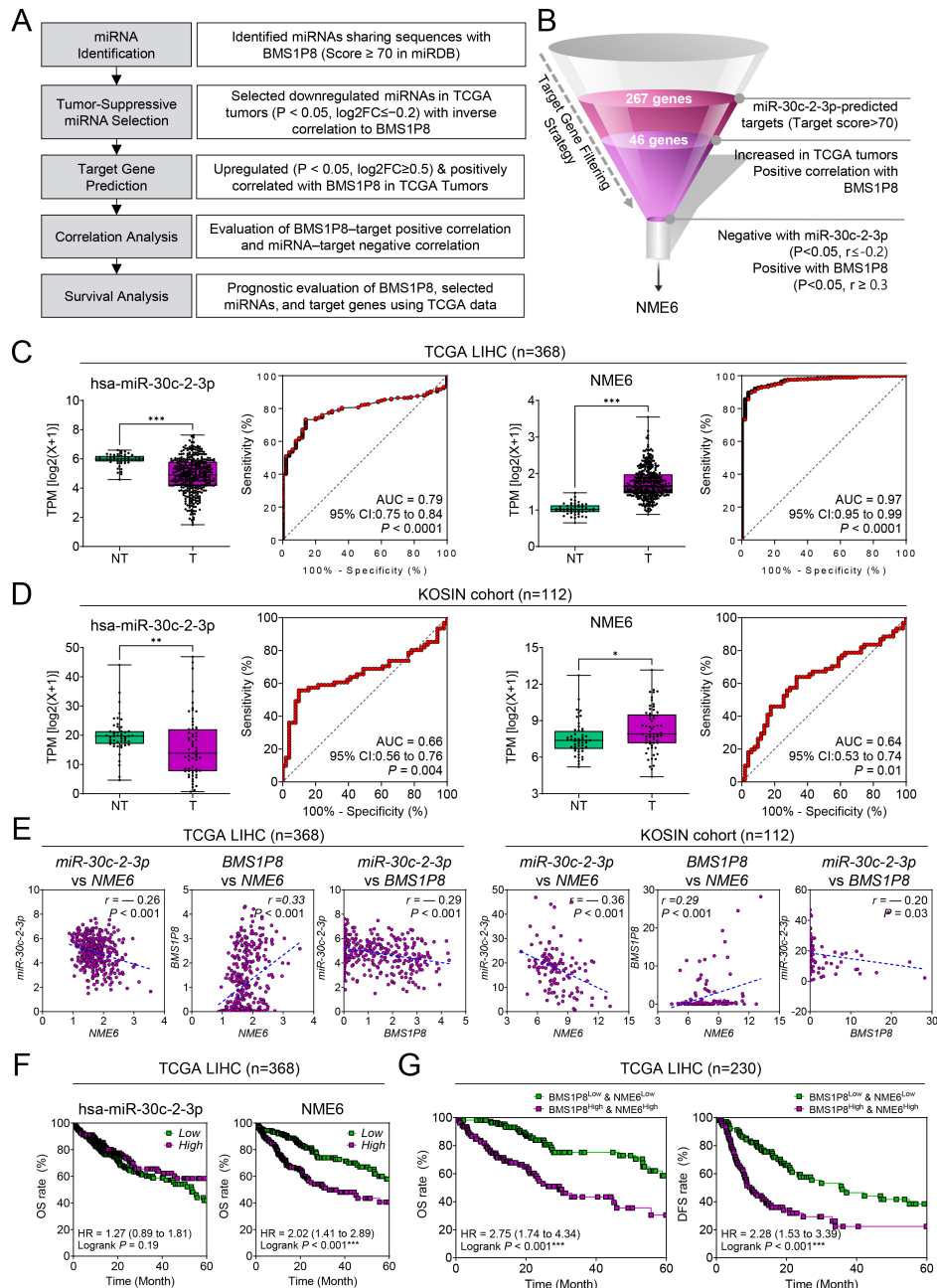


Fig. 6. A *BMS1P8*–miR-30c-2-3p–*NME6* competing-endogenous-RNA (ceRNA) axis in HCC. (A) Analytical workflow showing the five-step *in-silico* pipeline that identified miR-30c-2-3p and its target *NME6* within the *BMS1P8* ceRNA network. (B) Funnel diagram summarizing progressive filtering of miR-30c-2-3p targets: 267 predicted targets → 46 targets upregulated in tumors and positively correlated with *BMS1P8* → *NME6*, the only transcript also inversely correlated with miR-30c-2-3p ($r = -0.2$, $p < 0.05$). (C) Box plot (left) and receiver operating characteristic (ROC) curve (right) for miR-30c-2-3p and *NME6* expression in TCGA cohort. (D) Box plot (left) and ROC curve (right) for miR-30c-2-3p and *NME6* expression in KOSIN cohort. Box plots show log₂-transformed transcripts per million (TPM) levels comparing NT and T tissues; ROC curves show sensitivity versus 1 – specificity for distinguishing T from NT samples. (E) Pair-wise Pearson correlation scatterplots in both TCGA ($n = 368$) and KOSIN ($n = 112$) cohort: miR-30c-2-3p versus *BMS1P8*, *BMS1P8* versus *NME6*, and miR-30c-2-3p versus *NME6*; correlation coefficients (r) and two-tailed p values are indicated. (F) Kaplan–Meier overall survival (OS) curves comparing low- versus high-expression groups (median split) for miR-30c-2-3p and *NME6*; hazard ratio (HR) with 95 % CI and log-rank p values are shown. (G) Kaplan–Meier curves for OS (left) and disease-free survival (DFS, right) stratified by combined *BMS1P8* and *NME6* expression status (both^{Low}, either^{High}, both^{High}). Statistical comparisons of expression used unpaired Student's t -tests; survival differences used log-rank tests. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Data are presented as mean $\pm$ SD unless otherwise stated.

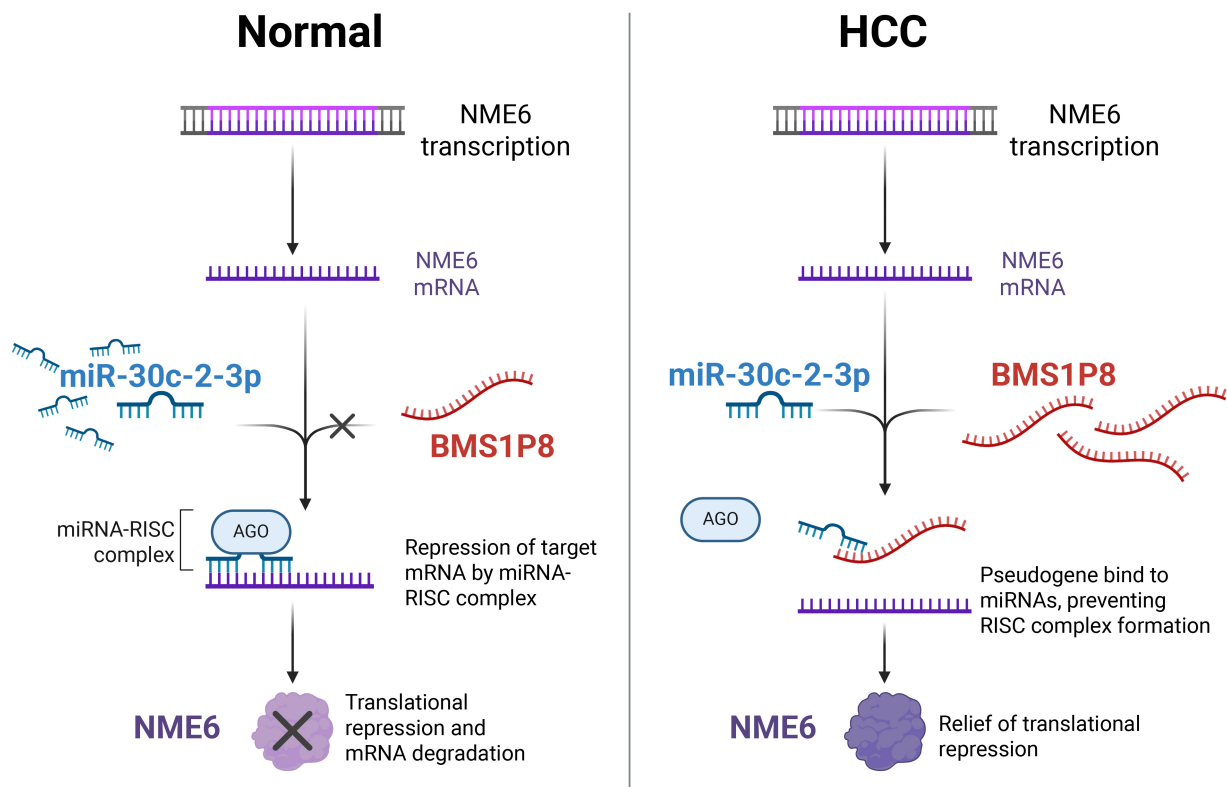


Fig. 7. Proposed mechanistic model of the *BMS1P8*–miR-30c-2-3p–*NME6* ceRNA axis in HCC. In normal hepatocytes (left), miR-30c-2-3p binds to *NME6* mRNA, recruiting the RNA-induced silencing complex (RISC) complex and leading to translational repression and mRNA degradation of *NME6*. In HCC cells (right), elevated *BMS1P8* sequesters miR-30c-2-3p, preventing RISC formation on *NME6* mRNA, thereby relieving translational repression and resulting in increased *NME6* expression. The figure was created using BioRender.com (Agreement number: MG28JOSDOF). The original figure source is available at <https://BioRender.com/5bss1pr>.

its expression associated with tumor size, differentiation, lymphatic metastasis, and invasion [24]. In parallel, recent work has delineated additional pseudogene-driven circuits that shape the key hallmarks of HCC. Double homeobox A pseudogene 8 (*DUXAP8*) sponges miR-490-5p, releasing budding uninhibited by benzimidazoles 1 (BUB1) and intensifying phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit beta (PI3K)/AKT serine/threonine kinase 1 (AKT)-driven proliferation, while misato family member 2, pseudogene (*MSTO2P*) simultaneously boosts E-cadherin and activates the PI3K/AKT/mechanistic target of rapamycin kinase (mTOR) axis to sustain tumor growth [25,26]. Also, methyltransferase 3, N⁶-adenosine-methyltransferase complex catalytic subunit (METTL3)-mediated N⁶-methyladenosine modification stabilizes glucosylceramidase beta 1 like, pseudogene (*GBAP1*); the resultant *GBAP1* overexpression sequesters miR-22-3p, up-regulates bone morphogenetic protein receptor type 1A (*BMPRIA*), and activates both BMP/Smad family member and PI3K/AKT cascades, promoting hepatocarcinogenesis [27,28]. Small nuclear ribonucleoprotein polypeptide F pseudogene 1 (*SNRPFPI*) is markedly up-regulated in HCC, associates with poor prognosis, and promotes proliferation, motility, and apoptosis resistance by sponging

the tumor-suppressive miR-126-5p [29], whereas oncogenic ubiquitin conjugating enzyme E2 M pseudogene 1 (*UBE2MPI*) promotes proliferation and apoptosis resistance by sponging miR-145-5p to de-repress regulator of G protein signaling 3 (RGS3) [30]. Collectively, these mechanistic insights—highlighting the emerging theme that pseudogene exert their influence primarily through miRNA-sponge activity—have driven the development of multi-pseudogene prognostic signatures that outperform conventional clinicopathologic variables, thereby reinforcing the regulatory and clinical relevance of pseudogenes in HCC and broader cancer biology.

In this study, we identified *BMS1P8* as a liver cancer-specific pseudogene with diagnostic and prognostic implications. By conducting a comprehensive analysis of RNA-seq datasets and validating the findings in clinical samples, we observed that *BMS1P8* expression was minimal in most other tumor types but prominently upregulated in HCC. Notably, *BMS1P8* remained barely detectable in CH and LC but rose modestly in DN and surged in both early- and advanced-stage HCC (Fig. 2A), indicating that its induction is tumor-specific rather than a generic response to chronic liver injury. The robust diagnostic performance of *BMS1P8* underscores its potential as a biomarker for early

detection or patient stratification. Moreover, our correlation and pathway enrichment analysis implicated *BMSIP8* in cell cycle regulation, a critical pathway often dysregulated in HCC progression. These insights were supported by the positive associations between *BMSIP8* expression and multiple cell cycle-related genes as well as the adverse prognostic outcomes linked to a *BMSIP8*-based 7-gene signature including *CCNB2*, *CDC20*, *CDC45*, *ESPL1*, *PLK1*, and *PTTG1*. Although the threshold of correlation coefficient ($r \geq |0.2|$) applied in this study may appear relatively low, it reflects a practical consideration when analyzing pseudogenes or non-coding RNAs with inherently low basal expression levels and high variability in large-scale datasets like TCGA. Prior study of ceRNA networks has also demonstrated that modest correlations ($r = 0.2-0.3$) can reveal biologically meaningful interactions, especially when these findings are further supported by independent pathway enrichment results [31]. This integrated approach provides additional confidence in the biological relevance of *BMSIP8*-cell cycle associations despite the moderate r values. Additionally, our ceRNA analysis suggested that *BMSIP8* sponge the tumor-suppressive miR-30c-2-3p, thereby derepressing *NME6*. This interaction could disrupt the normal regulatory network of miR-30c-2-3p, thereby promoting the expression of *NME6*, which is involved in tumor progression. Also, this axis was tightly linked to poorer OS and DFS in TCGA, suggesting a dual contribution of *BMSIP8* to HCC progression through both cell-cycle promotion and post-transcriptional regulation of oncogenic transcripts.

NME/NM23 nucleoside diphosphate kinase 6 (*NME6*) is a member of the NME gene family, which plays important roles in cellular processes such as nucleoside diphosphate kinase activity, maintenance of nucleotide pools, and regulation of cell proliferation and differentiation [32,33]. While several NME family members, such as *NME1* and *NME2*, have been extensively studied as metastasis suppressors in various cancers, the specific biological functions of *NME6* remain less well characterized [34]. Nonetheless, emerging studies have reported that *NME6* expression is elevated in multiple malignancies, including breast, colorectal, and lung cancer, where it has been associated with increased tumor cell proliferation, enhanced metastatic potential, and poorer clinical outcomes [35–37]. Although direct experimental research on *NME6* in HCC is limited, data mining analyses have indicated that *NME6* expression is elevated in HCC tissues and inversely correlated with patient survival, suggesting a potential oncogenic role for *NME6* in liver cancer progression [38]. In our analysis, we observed significant upregulation of *NME6* in HCC tumor tissues compared to non-tumor samples, accompanied by a strong association with poor prognosis. Furthermore, our ceRNA network analysis indicated that *BMSIP8* may sequester miR-30c-2-3p, leading to derepression of *NME6*, thereby implicating this axis as a

potential driver of HCC progression through dysregulation of cell cycle-related and oncogenic pathways.

One of the most notable findings is the liver specificity of *BMSIP8*, which helps distinguish it from other pseudogenes that may be broadly upregulated across multiple malignancies. This tissue specificity may allow *BMSIP8* to serve as a more targeted biomarker for HCC, potentially reducing false-positive results that can occur with conventional markers such as AFP [6]. Furthermore, the strong prognostic value observed for *BMSIP8* supports its clinical utility not only in diagnosing HCC but also in risk stratification and treatment decision-making. Despite these promising results, our study has certain limitations. Although the correlation and enrichment data suggest a functional role for *BMSIP8* in cell cycle regulation, we did not perform *in vitro* or *in vivo* functional assays to validate the mechanistic underpinnings of how *BMSIP8* might drive tumorigenesis.

In conclusion, our findings demonstrate the diagnostic and prognostic importance of *BMSIP8* in HCC and highlight its potential tissue specificity. Moreover, our identification of links between *BMSIP8* expression and cell cycle-related pathways provides a foundation for understanding its role in liver cancer progression. Although this study derived key insights primarily from tissue-based RNA analysis, we recognize that tissue sampling is invasive and may limit clinical applicability. However, with the recent advancements in liquid biopsy technologies, including extracellular vesicle (EV)-based RNA analysis, there is potential for *BMSIP8* to be detected in serum or plasma-derived EVs. This raises the possibility of applying *BMSIP8* as a non-invasive biomarker in clinical practice. Therefore, further studies evaluating the detectability of *BMSIP8* in patient blood samples and its correlation with tissue expression will be essential. Future mechanistic studies of this pseudogene could lead to improved early diagnosis and therapeutic strategies for HCC patients.

5. Conclusion

This study identified *BMSIP8* as a liver-specific pseudogene biomarker with a strong diagnostic and prognostic value in HCC. Its distinct upregulation in liver cancer compared with other malignancies underscores its potential clinical utility for early detection and patient stratification. Furthermore, the correlation between *BMSIP8* and cell cycle-related pathways highlight its relevance to disease progression. Our results also revealed a potential *BMSIP8*/miR-30c-2-3p/*NME6* ceRNA circuit that may amplify oncogenic signaling in HCC. Elucidating and targeting this newly defined axis could broaden the therapeutic possibilities for *BMSIP8*. These findings not only enhance our understanding of the molecular landscape of HCC but also provide a foundation for future translational research aimed at integrating *BMSIP8* into diagnostic workflows and exploring its potential as a target for therapeutic intervention.

Abbreviations

AFP, alpha-fetoprotein; ALT, alanine aminotransferase; AST, aspartate aminotransferase; AUC, area under curve; aHCC, advanced hepatocellular carcinoma; BMI, body mass index; cDNA, complementary deoxyribonucleic acid; CH, chronic hepatitis; CI, confidence interval; DFS, disease-free survival; DN, dysplastic nodule; eHCC, early staged hepatocellular carcinoma; FC, fold changes; GO, gene ontology; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; HCV, hepatitis C virus; HMBS, hydroxymethylbilane synthase; HR, hazard ratio; INR, international normalized ratio; LC, liver cirrhosis; LIHC, liver hepatocellular carcinoma; lncRNA, long non-coding RNA; MSigDB, Molecular Signatures Database; NASH, non-alcoholic steatohepatitis; NL, normal liver; NT, non-tumor; OS, overall survival; PIVKA-II, prothrombin-induced by vitamin K absence or antagonist-II; qRT-PCR, quantitative real-time polymerase chain reaction; RNA-seq, RNA-sequencing; ROC, receiver operating characteristics; SD, standard deviation; T, tumor tissue; TCGA, The Cancer Genome Atlas.

Availability of Data and Materials

Data supporting the findings of this study are presented in the paper and supplementary files. All other data are available from the corresponding authors upon reasonable request.

Author Contributions

Conceptualization: HSK, JWE; data curation: HSK, JYChoi, GOB, SSK, JYCheong; methodology: MGY, JYJ; investigation: SHJ, JEJ, JWE; resources: SSK, JYCheong; visualization: HSK, JYChoi; funding acquisition: HSK, JYCheong, JYJ, JWE; project administration: JYJ, JWE; supervision: JYJ, JWE; writing—original draft: HSK, JWE; writing – review & editing: JYCheong, JWE. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

The study was carried out in accordance with the guidelines of the Declaration of Helsinki. This study was approved by the Institutional Review Board of the Ajou University Hospital (Suwon, South Korea) (approval number: AJOURB-EX-2024-332 and AJOURB-EX-2024-389). The requirement for informed consent was waived.

Acknowledgment

The biospecimens and data used in this study were provided by the Biobank of Ajou University Hospital, a member of the Korea Biobank Network. In addition, we thank all the members of MOAGEN (Daejeon, South Ko-

rea) for providing bioinformatics analysis and guidance. Fig. 7 was created using BioRender.com (Agreement number: MG28JOSDOF). The original figure source is available at <https://BioRender.com/5bss1pr>.

Funding

This work was supported by the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare, Republic of Korea (grant number HR21C1003); by the National Research Foundation of Korea (NRF), funded by the Ministry of Science and ICT (MSIT), Republic of Korea (grant numbers NRF-2022R1A2C1092155, RS-2022-NR070489, RS-2023-00210847, RS-2025-00562556 and RS-2025-00521818).

Conflict of Interest

The authors declare no conflict of interest.

Declaration of AI and AI-Assisted Technologies in the Writing Process

ChatGPT (version 1o pro) was used for error checking and language refinement during the manuscript preparation. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at <https://doi.org/10.31083/FBL41684>.

References

- [1] Vogel A, Meyer T, Sapisochin G, Salem R, Saborowski A. Hepatocellular carcinoma. *Lancet* (London, England). 2022; 400: 1345–1362. [https://doi.org/10.1016/S0140-6736\(22\)01200-4](https://doi.org/10.1016/S0140-6736(22)01200-4).
- [2] Mashiba T, Joko K, Kurosaki M, Ochi H, Hasebe C, Akahane T, *et al.* Real-world efficacy of elbasvir and grazoprevir for hepatitis C virus (genotype 1): A nationwide, multicenter study by the Japanese Red Cross Hospital Liver Study Group. *Hepatology Research: the Official Journal of the Japan Society of Hepatology*. 2019; 49: 1114–1120. <https://doi.org/10.1111/hepr.13362>.
- [3] Huang DQ, Mathurin P, Cortez-Pinto H, Loomba R. Global epidemiology of alcohol-associated cirrhosis and HCC: trends, projections and risk factors. *Nature Reviews. Gastroenterology & Hepatology*. 2023; 20: 37–49. <https://doi.org/10.1038/s41575-022-00688-6>.
- [4] Wei S, Zhang Y, Kang BE, Park W, Guo H, Nam S, *et al.* CDKN2 expression is a potential biomarker for T cell exhaustion in hepatocellular carcinoma. *BMB Reports*. 2024; 57: 287–292. <https://doi.org/10.5483/BMBRep.2023-0214>.
- [5] Gosalia AJ, Martin P, Jones PD. Advances and Future Directions in the Treatment of Hepatocellular Carcinoma. *Gastroenterology & Hepatology*. 2017; 13: 398–410.
- [6] Gupta S, Bent S, Kohlwes J. Test characteristics of alpha-fetoprotein for detecting hepatocellular carcinoma in patients with hepatitis C. A systematic review and critical analysis. *An-*

- nals of Internal Medicine. 2003; 139: 46–50. <https://doi.org/10.7326/0003-4819-139-1-200307010-00012>.
- [7] Hu X, Yang L, Mo YY. Role of Pseudogenes in Tumorigenesis. *Cancers*. 2018; 10: 256. <https://doi.org/10.3390/cancer10080256>.
 - [8] Roberts TC, Morris KV. Not so pseudo anymore: pseudogenes as therapeutic targets. *Pharmacogenomics*. 2013; 14: 2023–2034. <https://doi.org/10.2217/pgs.13.172>.
 - [9] Li RK, Gao J, Guo LH, Huang GQ, Luo WH. PTENP1 acts as a ceRNA to regulate PTEN by sponging miR-19b and explores the biological role of PTENP1 in breast cancer. *Cancer Gene Therapy*. 2017; 24: 309–315. <https://doi.org/10.1038/cgt.2017.29>.
 - [10] Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature*. 2010; 465: 1033–1038. <https://doi.org/10.1038/nature09144>.
 - [11] Tay Y, Kats L, Salmena L, Weiss D, Tan SM, Ala U, *et al.* Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell*. 2011; 147: 344–357. <https://doi.org/10.1016/j.cell.2011.09.029>.
 - [12] Hayashi H, Arai T, Togashi Y, Kato H, Fujita Y, De Velasco MA, *et al.* The OCT4 pseudogene POU5F1B is amplified and promotes an aggressive phenotype in gastric cancer. *Oncogene*. 2015; 34: 199–208. <https://doi.org/10.1038/onc.2013.547>.
 - [13] Kim HS, Na MJ, Son KH, Yang HD, Kim SY, Shin E, *et al.* ADAR1-dependent miR-3144-3p editing simultaneously induces MSI2 expression and suppresses SLC38A4 expression in liver cancer. *Experimental & Molecular Medicine*. 2023; 55: 95–107. <https://doi.org/10.1038/s12276-022-00916-8>.
 - [14] Eun JW, Yoon JH, Ahn HR, Kim S, Kim YB, Lim SB, *et al.* Cancer-associated fibroblast-derived secreted phosphoprotein 1 contributes to resistance of hepatocellular carcinoma to sorafenib and lenvatinib. *Cancer Communications (London, England)*. 2023; 43: 455–479. <https://doi.org/10.1002/cac2.12414>.
 - [15] Son JA, Weon JH, Baek GO, Ahn HR, Choi JY, Yoon MG, *et al.* Circulating small extracellular vesicle-derived splicing factor 3b subunit 4 as a non-invasive diagnostic biomarker of early hepatocellular carcinoma. *Journal of Experimental & Clinical Cancer Research*. 2023; 42: 288. <https://doi.org/10.1186/s13046-023-02867-y>.
 - [16] Goldman MJ, Craft B, Hastie M, Repecka K, McDade F, Kamath A, *et al.* Visualizing and interpreting cancer genomics data via the Xena platform. *Nature Biotechnology*. 2020; 38: 675–678. <https://doi.org/10.1038/s41587-020-0546-8>.
 - [17] Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, *et al.* TCGAAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Research*. 2016; 44: e71. <https://doi.org/10.1093/nar/gkv1507>.
 - [18] Mounir M, Lucchetta M, Silva TC, Olsen C, Bontempi G, Chen X, *et al.* New functionalities in the TCGAAbiolinks package for the study and integration of cancer data from GDC and GTEx. *PLoS Computational Biology*. 2019; 15: e1006701. <https://doi.org/10.1371/journal.pcbi.1006701>.
 - [19] Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Research*. 2020; 48: D127–D131. <https://doi.org/10.1093/nar/gkz757>.
 - [20] Robicheau BM, Susko E, Harrigan AM, Snyder M. Ribosomal RNA Genes Contribute to the Formation of Pseudogenes and Junk DNA in the Human Genome. *Genome Biology and Evolution*. 2017; 9: 380–397. <https://doi.org/10.1093/gbe/evw307>.
 - [21] Pink RC, Wicks K, Caley DP, Punch EK, Jacobs L, Carter DRF. Pseudogenes: pseudo-functional or key regulators in health and disease? *RNA (New York, N.Y.)*. 2011; 17: 792–798. <https://doi.org/10.1261/rna.2658311>.
 - [22] Park CS, Habib O, Lee Y, Hur JK. Applications of CRISPR technologies to the development of gene and cell therapy. *BMB Reports*. 2024; 57: 2–11. <https://doi.org/10.5483/BMBR.ep.2023-0221>.
 - [23] Kalyana-Sundaram S, Kumar-Sinha C, Shankar S, Robinson DR, Wu YM, Cao X, *et al.* Expressed pseudogenes in the transcriptional landscape of human cancers. *Cell*. 2012; 149: 1622–1634. <https://doi.org/10.1016/j.cell.2012.04.041>.
 - [24] Mei D, Song H, Wang K, Lou Y, Sun W, Liu Z, *et al.* Up-regulation of SUMO1 pseudogene 3 (SUMO1P3) in gastric cancer and its clinical association. *Medical Oncology (Northwood, London, England)*. 2013; 30: 709. <https://doi.org/10.1007/s12032-013-0709-2>.
 - [25] Zhang H, Chu K, Zheng C, Ren L, Tian R. Pseudogene *DUXAP8* Promotes Cell Proliferation and Migration of Hepatocellular Carcinoma by Sponging MiR-490-5p to Induce BUB1 Expression. *Frontiers in Genetics*. 2020; 11: 666. <https://doi.org/10.3389/fgene.2020.00666>.
 - [26] Yan L, Yue C, Xu Y, Jiang X, Zhang L, Wu J. Prognostic Value and Molecular Regulatory Mechanism of MSTO2P in Hepatocellular Carcinoma: A Comprehensive Study Based on Bioinformatics, Clinical Analysis and in vitro Validation. *Oncotargets and Therapy*. 2020; 13: 2583–2598. <https://doi.org/10.2147/OTT.S245741>.
 - [27] Liu R, Yin G, Tuo H, Guo Y, Zhu Y, Zhang L, *et al.* METTL3-induced lncRNA GBAP1 promotes hepatocellular carcinoma progression by activating BMP/SMAD pathway. *Biology Direct*. 2023; 18: 53. <https://doi.org/10.1186/s13062-023-00409-2>.
 - [28] Chen R, Zhao M, An Y, Liu D, Tang Q. GBAP1 functions as a tumor promoter in hepatocellular carcinoma via the PI3K/AKT pathway. *BMC Cancer*. 2023; 23: 628. <https://doi.org/10.1186/s12885-023-11107-7>.
 - [29] Wang N, Guo S, Hao F, Zhang Y, Chen Y, Fei X, *et al.* Pseudogene SNRFP1 derived long non-coding RNA facilitates hepatocellular carcinoma progress in vitro by sponging tumor-suppressive miR-126-5p. *Scientific Reports*. 2022; 12: 21867. <https://doi.org/10.1038/s41598-022-24597-5>.
 - [30] Hao F, Wang N, Gui H, Zhang Y, Wu Z, Wang J. Pseudogene UBE2MP1 derived transcript enhances *in vitro* cell proliferation and apoptosis resistance of hepatocellular carcinoma cells through miR-145-5p/RGS3 axis. *Aging*. 2022; 14: 7906–7925. <https://doi.org/10.18632/aging.204319>.
 - [31] Hsiao YW, Wang L, Lu TP. ceRNAR: An R package for identification and analysis of ceRNA-miRNA triplets. *PLoS Computational Biology*. 2022; 18: e1010497. <https://doi.org/10.1371/journal.pcbi.1010497>.
 - [32] Schlattner U. The Complex Functions of the NME Family-A Matter of Location and Molecular Activity. *International Journal of Molecular Sciences*. 2021; 22: 13083. <https://doi.org/10.3390/ijms222313083>.
 - [33] Proust B, Radić M, Vidaček NŠ, Cottet C, Attia S, Lamarche F, *et al.* NME6 is a phosphotransfer-inactive, monomeric NME/NDPK family member and functions in complexes at the interface of mitochondrial inner membrane and matrix. *Cell & Bioscience*. 2021; 11: 195. <https://doi.org/10.1186/s13578-021-00707-0>.
 - [34] Wu H, Huang X, Chen S, Li S, Feng J, Zou Xu X, *et al.* Comprehensive analysis of the NME gene family functions in breast cancer. *Translational Cancer Research*. 2020; 9: 6369–6382. <https://doi.org/10.21037/tcr-20-1712>.
 - [35] Proust B, Horvat A, Tadjian A, Vlačić I, Herak Bosnar M. Mitochondrial NME6 Influences Basic Cellular Processes in Tumor Cells In Vitro. *International Journal of Molecular Sciences*. 2024; 25: 9580. <https://doi.org/10.3390/ijms25179580>.
 - [36] Kim S, Kang M, Jeong S, Kim J, Kim KO, Lee WS, *et al.* Elucidating prognostic significance of purine metabolism in col-

orectal cancer through integrating data from transcriptomic, immunohistochemical, and single-cell RNA sequencing analysis. *Molecular Oncology*. 2025; 10.1002/1878-0261.70010. <https://doi.org/10.1002/1878-0261.70010>.

- [37] Luo L, Li Y, Zhang L, Yang L. NME6 as a potential biomarker and therapeutic target involved in immune infiltration for lung adenocarcinoma. *Technology and Health Care: Official Journal*

of the European Society for Engineering and Medicine. 2024; 32: 2277–2291. <https://doi.org/10.3233/THC-231058>.

- [38] Ren X, Rong Z, Liu X, Gao J, Xu X, Zi Y, *et al.* The Protein Kinase Activity of NME7 Activates Wnt/ β -Catenin Signaling to Promote One-Carbon Metabolism in Hepatocellular Carcinoma. *Cancer Research*. 2022; 82: 60–74. <https://doi.org/10.1158/0008-5472.CAN-21-1020>.