GULP1 as a Novel Diagnostic and Predictive Biomarker in Hepatocellular Carcinoma

Hyung Seok Kim[1,*], Jung Hwan Yoon[2,*], Ji Yi Choi[3,4], Moon Gyeong Yoon[3], Geum Ok Baek[3], Minji Kang[3,4], Se Ha Jang[3,4], Won Park[5], Yunjin Go[4,6], Jestlin Tianthing Ng[4,6], Suk Woo Nam[2], Jee-Yeong Jeong[1], Ji Eun Han[3], Hyo Jung Cho[3], Su Bin Lim[6], Soon Sun Kim[3], Jae Youn Cheong[3] and Jung Woo Eun[3]

[1]Department of Biochemistry, College of Medicine, Kosin University; Seo-gu, Busan 49267, South Korea; [2]Department of Pathology, College of Medicine, The Catholic University of Korea; Seocho-gu, Seoul 06591, South Korea; [3]Department of Gastroenterology, Ajou University School of Medicine; 164 Worldcup-ro, Yeongtong-gu, Suwon 16499, South Korea; [4]Department of Biomedical Sciences, Ajou University Graduate School of Medicine; 164 Worldcup-ro, Yeongtong-gu, Suwon 16499, South Korea; [5]Department of Bioscience and Biotechnology, Graduate School, Chungnam National University; 99 Daehak-ro, Yuseong-gu, Daejeon 34134, South Korea; [6]Department of Biochemistry and Molecular Biology, Ajou University School of Medicine; 164 Worldcup-ro, Yeongtong-gu, Suwon 16499, South Korea

*These authors contributed equally to this word

Corresponding author: **Jung Woo Eun**

Department of Gastroenterology, Ajou University School of Medicine; 164 Worldcup-ro, Yeongtong-gu, Suwon 16499, South Korea

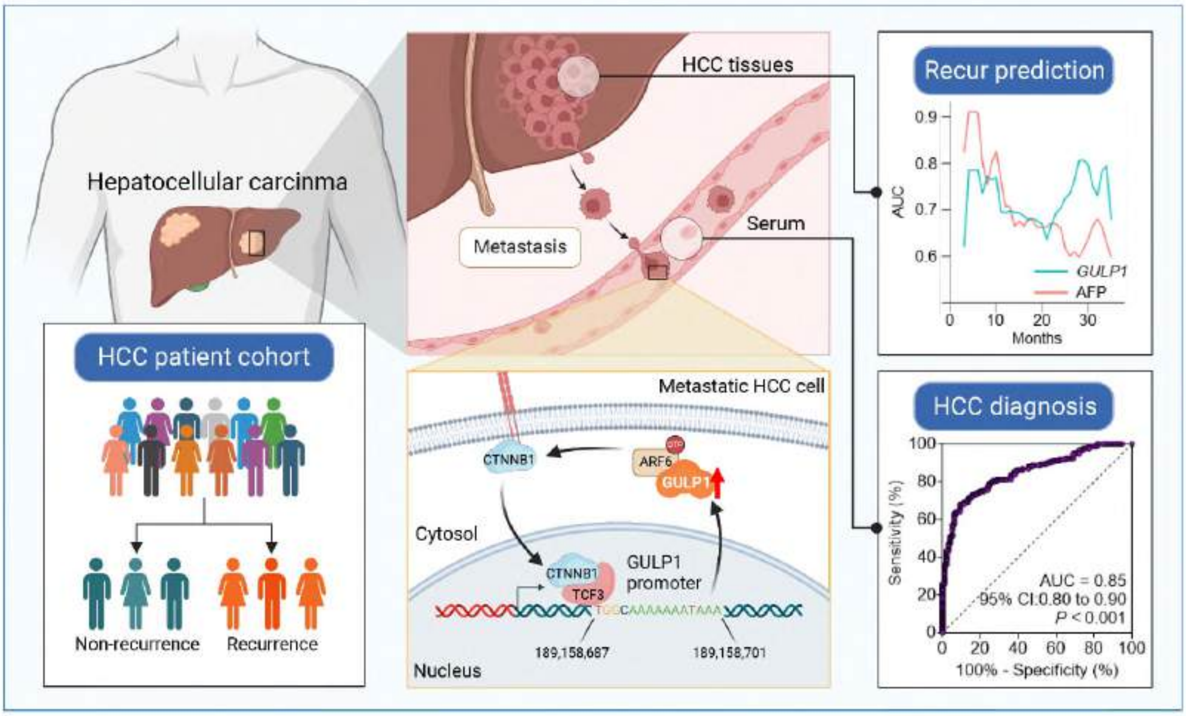
Tel.: 82-31-219-4681, Fax: 82-31-219-4680, E-mail: jetaimebin@gmail.com

List of abbreviations

ADJ_HCC, adjacent non-tumor tissue; AFP, α-fetoprotein; AL, adjacent liver tissue; ANOVA, one-way analysis of variance; ARF6, ADP-ribosylation factor 6; AUC, area under the curve; BrdU, bromodeoxyuridine; CH, chronic hepatitis B virus; CHX, cycloheximide; CI, confidence interval; DAPI, 4′, 6-diamidino-2-phenylindole; DDX5, DEAD-box helicase 5; DEG, differentially expressed gene; ELISA, enzyme-linked immunosorbent assay; EMT, epithelial–mesenchymal transition; FRET, fluorescence resonance energy transfer; GEO, Gene Expression Omnibus; GPC3, glypican-3; GSEA, gene set enrichment analysis; GULP1, GULP PTB domain-containing engulfment adaptor 1; GULP1_OE, GULP1 overexpression; H&E, hematoxylin and eosin; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; HCV, hepatitis C virus; HR, hazard ratio; IF, immunofluorescence; IHC, immunohistochemistry; LC, liver cirrhosis; LIHC, liver hepatocellular carcinoma; MT1, first mutant; MT2, second mutant; MTT, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide; mUICC, modified Union for International Cancer Control; MASH, metabolic dysfunction-associated steatohepatitis; NL, normal liver; NR, non-recurrence; NT, non-tumor; OS, overall survival; PT, primary tumor; qRT-PCR, quantitative reverse transcription polymerase chain reaction; R, recurrence; RFS, recurrence-free survival; ROC, receiver operating characteristics; RS, risk score; RT, recurrent tumor tissue; scRNA-seq, single-cell RNA sequencing; siCTNNB1, CTNNB1-targeting siRNA; siCtrl, scrambled sequence of single interference control RNA; siGULP1,small interfering RNA targeting GULP1; siRNA, small interfering RNA; ssGSEA, single-sample gene set enrichment analysis; TCGA, The Cancer Genome Atlas; TCF3, transcription factor 3; T, tumor; UMAP, uniform manifold approximation and projection; VEGF, vascular endothelial growth factor; WT: wild type.

Study highlights

- GULP1 exhibited predictive accuracy comparable to that of a 15-gene risk score model for HCC recurrence with high clinical application potential and robustness.
- GULP1 was specifically overexpressed in HCC, distinguishing it from other liver conditions and showing significant prognostic and diagnostic value.
- GULP1 promoted tumor growth, EMT, and invasiveness by modulating β-catenin signaling, playing key roles in HCC progression.
- Our findings suggest GULP1 as a promising non-invasive biomarker and therapeutic target for HCC recurrence and progression.



ABSTRACT

Backgrounds/Aims: Hepatocellular carcinoma (HCC) is characterized by high recurrence and mortality, necessitating the identification of reliable biomarkers. In this study, we aimed to identify the predictive gene signatures for HCC recurrence and evaluate the efficiency of GULP PTB domain-containing engulfment adaptor 1 (GULP1) as a predictive and diagnostic marker and therapeutic target for HCC.

Methods: We analyzed genomic datasets from The Cancer Genome Atlas and Gene Expression Omnibus databases via least absolute shrinkage and selection operator Cox regression and 10-fold cross-validation, leading to the development of a 15-gene risk score model, which was validated using three independent datasets. Serum GULP1 and α-fetoprotein levels were assessed to determine the diagnostic accuracy of the model. Using clinical cohorts and patient sera, GULP1 roles were examined, and functional assays *in vitro* and *in vivo* were used to evaluate its effects on cell growth, epithelial–mesenchymal transition (EMT), ADP-ribosylation factor 6 activation, and β-catenin signaling.

Results: Our newly developed risk-score model accurately predicted recurrent HCC in all datasets. Among the 15 genes in the risk score model, *GULP1* was overexpressed in patients with HCC and independently predicted HCC recurrence. Its expression modulation influenced cell growth and EMT, with observed effects on ADP-ribosylation factor 6 activation and β-catenin signaling pathways.

Conclusions: GULP1 is a crucial biomarker for HCC, serving as a non-invasive diagnostic and predictive tool. It also plays key roles in HCC progression. Our findings highlight the potential use of GULP1 in treatment strategies targeting EMT and HCC recurrence to improve the personalized care and patient outcomes.

Keywords: Liver cancer; GULP1; Recurrence; Metastasis; Diagnosis

INTRODUCTION

Hepatocellular carcinoma (HCC) is a prevalent aggressive cancer with high recurrence and, mortality rates and low patient quality of life.[1] Key risk factors for HCC include chronic hepatitis (CH) B and C, excessive alcohol intake, and metabolic disorders.[2, 3] Due to subtle early symptoms, HCC is often diagnosed at advanced stages, complicating its timely treatment and contributing to recurrence rates as high as 80%.[1, 4] Although alpha-fetoprotein (AFP) is widely used in clinical practice as a serum biomarker for HCC, multiple studies, it shows limited sensitivity and specificity.[1] This shortcoming highlights the need for more reliable biomarkers for the early diagnosis, risk stratification, and personalized treatment of HCC.

In this study, we applied advanced machine learning techniques, including least absolute shrinkage and selection operator (LASSO) Cox regression with 10-fold cross-validation, to identify the gene signatures associated with HCC recurrence. Based on the results, we developed a 15-gene risk score (RS) model and identified the GULP PTB domain-containing engulfment adaptor 1 (GULP1) as a potential prognostic marker for HCC recurrence.

GULP1 is the human counterpart of CED6 from *Caenorhabditis elegans*, with both playing a conserved role in cellular engulfment across species.[5] GULP1 facilitates EphB/ephrinB trogocytosis—a process of cell surface material transfer—by collaborating with Tiam2 and is essential for recruiting dynamin, a crucial protein for cellular internalization.[6] Interestingly, although GULP1 is typically recognized as a tumor suppressor, its expression was elevated in HCC, where it promoted β-catenin activity and epithelial–mesenchymal transition (EMT), which are associated with metastasis.[7, 8] Importantly, our data indicate that GULP1 outperforms AFP in detecting early-stage HCC and predicting its recurrence, highlighting it as a more sensitive and specific biomarker than AFP. Furthermore, our investigation highlighted the potential of GULP1 as a biomarker in liquid biopsies, serving as a minimally invasive tool to monitor HCC recurrence and progression, thereby advancing personalized medicine for HCC.

In summary, our study emphasized GULP1's role in recurrence and metastasis, providing insights into the mechanisms behind HCC progression and suggesting promising strategies for early diagnosis and targeted therapy.

MATERIAL AND METHODS

Patients and specimens

HCC and adjacent non-cancerous tissues and blood samples obtained from the Ajou University Hospital (Suwon, South Korea) were used in this study. Tissue samples were collected from 81 patients with HCC who underwent hepatectomy, and 256 blood samples were collected from healthy individuals and patients with CH, cirrhosis, and HCC. To validate the diagnostic performance of GULP1 across various liver disease etiologies, additional patient samples, including hepatitis C virus–induced liver cirrhosis (HCV-LC; n = 30), alcoholic LC (n = 30), alcoholic HCC (n = 30), metabolic dysfunction-associated steatohepatitis-LC (MASH-LC; n = 30), MASH-HCC (n = 30; 5 with confirmed MASH and 25 with unknown origins), cholangiocarcinoma (n = 8), and combined hepatocellular carcinoma and cholangiocarcinoma (n = 22) samples, were obtained from the Human Biobank. The patient demographics and clinical characteristics are presented in Supplementary Tables 1-3.

Collection and analysis of gene expression data for HCC recurrence

We analyzed the gene expression data from three publicly available HCC datasets (GSE14520, GSE114564, and The Cancer Genome Atlas [TCGA] liver hepatocellular carcinoma [LIHC]) to identify the genes associated with HCC recurrence. The patients were divided into non-recurrent and recurrent groups based on their recurrence status within two years post-surgery. Initial differential expression analysis identified the candidate genes across all datasets. Using Elastic Net Cox regression and LASSO Cox regression with cross-validation, we refined these candidates to construct a robust 15-gene recurrence RS model. Detailed patient baseline characteristics are presented in Supplementary Table 4, and additional methodological details, including dataset processing and statistical parameters, are provided in the Supplementary Methods section.

RESULTS

Development of a 15-gene RS model to predict HCC recurrence

In this study, patient selection criteria included patients who underwent surgical liver resection and remained recurrence-free for over two years (non-recurrence [NR] group), as well as patients who experienced recurrence (recurrence [R] group). To identify the differentially expressed genes (DEGs) associated with HCC recurrence, we analyzed three datasets (GSE14520, GSE114564, and TCGA LIHC) with clinical information on patients with HCC who underwent surgical liver resection and were either recurrence-free for over two years or experienced recurrence within two years (Supplementary Table 4). Venn diagram-based analysis identified 50 overlapping DEGs between the R and NR groups (Fig. 1A). To identify the genes potentially associated with recurrence, we conducted Elastic Net Cox regression analysis on the 50 selected genes within TCGA LIHC dataset.

Lambda values, a hyperparameter for model optimization, were examined for each gene (Fig. 1B, left; Supplementary Table 5). Through 10-fold cross-validation, optimal lambda value was determined as 0.1829059, as indicated by the blue line in the graph (Fig. 1B, right panel). From this analysis, 15 genes were identified as independent predictors within the model, and their respective regression coefficients are listed in Supplementary Table 6.

Next, to further refine the RS model, we calculated RSs for each gene by applying univariate Cox regression to the expression values of the 15 genes using TCGA dataset of HCC with recurrence information. Regression coefficient for the RS model was obtained as follows:

RS = (0.03208 × *GULP1*) + (−0.04351 × *LCAT*) + (0.29381 × *PPAT*) + (−0.05648 × *LPXN*) + (0.09575 × *NOP56*) + (−0.03938 × *CD4*) + (0.12171 × *ZC2HC1A*) + (0.42938 × *PPIA*) + (0.08035 × *CST7*) + (−0.04075 × *PRKCQ*) + (0.12525 × *PHF20*) + (0.03416 × *RAB23*) + (0.07323 × *PCDHB6*) + (−0.17598 × *CXCR6*) + (−0.09678 × *SLC4A10*).

Using TCGA model as a reference, a cut-off value was used in each dataset to classify the patients into high- and low-risk groups. Correlation analysis of all datasets consistently demonstrated a negative correlation between patient RS scores and survival (Fig. 1C). Notably, our 15-gene RS model outperformed the existing 7-gene models, showing superior predictive performance and higher net benefits across all datasets. (Supplementary Fig. S1A).[9] Receiver operating characteristics (ROC) analysis further confirmed the predictive capacity of the 15-gene RS model, with area under the curve (AUC) ≥ 0.7, indicating relatively high accuracy in predicting HCC recurrence (Fig. 1D). Additionally, GSEA revealed the significant enrichment of gene sets from Hallmark Collection in the high-risk group, indicating their involvement in specific signaling pathways, such as E2F targets, G2/M checkpoint, mitotic spindle, and MYC targets v1 (Supplementary Fig. S1B). This connection is crucial because MYC influences EMT and recurrence in various cancers, including HCC.[10-12] Notably, Kaplan–Meier curves showed that the low-risk group (RS_Low) exhibited a significantly longer recurrence-free survival (RFS) than the high-risk group (RS_High) in all three datasets (Fig. 1E). Through univariate analysis of the clinical information in each dataset, RS model exhibited a consistent and independent association with the diagnostic outcome in all analyzed datasets (Fig. 1F). Cox regression analysis of RFS in TCGA cohort revealed that the RS model exhibited significant associations in both univariate (hazard ratio [HR]: 2.72; 95% confidence interval [CI]: 2.146–3.443) and multivariate (HR: 2.08; 95% CI: 1.311–3.313) analyses, showing the highest HR compared to other variables (Supplementary Table 7).

Elevated GULP1 expression in HCC highlights its potential as a liver cancer-specific biomarker

Within the 15-gene RS model, *GULP1* emerged as a potent biomarker, showing predictive accuracy comparable to that of the entire model for HCC recurrence. Although the RS model robustness was enhanced by integrating multiple genes, *GULP1* alone was sufficient to stratify patients according to the recurrence risk with similar precision (Supplementary Fig. S2). These findings suggest *GULP1* as a simple and cost-effective biomarker for clinical use that reliably predicts the recurrence risk without the need for complex multigene profiling. Additionally, analysis of *GULP1* in 33 different TCGA cancer types revealed that *GULP1* was generally downregulated in other cancer types (Supplementary Fig. S3A). This finding was contrary to our expectations and prompted us to specifically investigate its role in HCC. We investigated the expression levels of *GULP1* in HCC various stages and conditions using GepLiver DB datasets. This analysis revealed differences in *GULP1* expression levels among the normal, adjacent non-tumor (ADJ_HCC), and HCC liver tissues, with HCC tissues exhibiting markedly elevated *GULP1* levels (Fig. 2A). Quantitative analysis showed that *GULP1* levels were significantly higher in HCC tissues than in normal, viral hepatitis, nonalcoholic fatty liver disease, and cirrhosis tissues, underscoring its potential in distinguishing HCC from other liver conditions (Fig. 2B).

High-throughput analysis of spatial transcriptomics showed a pronounced increase in *GULP1* levels in the malignant hepatocytes, suggesting a distinct spatial distribution of the gene in the tumor microenvironment (Fig. 2C and D; Supplementary Fig. S3B).

scRNA-seq data from GepLiver DB revealed diverse cellular landscapes in various liver cell types. The datasets included various liver-specific and immune cells, providing a comprehensive view of the cell populations present in both normal and diseased liver tissues (Fig. 2E, left; Supplementary Fig. S3C). Within this diverse cellular environment, we specifically examined the *GULP1* expression levels in hepatocytes. The proportion of *GULP1*-positive hepatocytes increased progressively with the progression of liver malignancy. In normal tissues (NT), only 0.82% of hepatocytes were *GULP1*-positive; however, this number increased to 1.91% in ADJ_HCC tissues, 2.9% in non-malignant HCC tissues, and 4.39% in malignant HCC tissues. (Fig. 2E, right). Similarly, analysis of the scAtlasLC database and subsequent quantification of *GULP1*-positive hepatocytes confirmed the increased *GULP1* expression in HCC, emphasizing its significance in liver cancer pathology (Fig. 2F). Mean expression levels of *GULP1* in different cell types in the GSE151530 dataset further confirmed that the hepatocytes were the primary source of *GULP1* expression, highlighting the hepatocyte-specific roles of GULP1 in HCC (Fig. 2G; Supplementary Fig. S3D). Subsequently, hepatocytes were classified into *GULP1*-positive and -negative populations to explore the functional implications of *GULP1* expression (Fig. 2H; Supplementary Fig. S3E). *GULP1*-positive hepatocytes were transcriptionally enriched in the key oncogenic pathways, including the EMT, hypoxia, and KRAS pathways critical for HCC progression (Fig. 2I; Supplementary Fig. S3F and G). These findings suggest that hepatocyte clusters with high *GULP1* expression are transcriptionally aligned with key oncogenic processes.

GULP1 is a prognostic and diagnostic marker for HCC

Considering the predictive strength of GULP1, Kaplan–Meier analysis was conducted to assess the survival differences based on GULP1 expression levels in HCC. Patients in the three datasets, GSE14520, GSE114564, and TCGA LIHC, with high *GULP1* expression levels (GULP1_High) showed worse prognosis than those with low *GULP1* levels (GULP1_Low; Supplementary Fig. S4A–C). To explore the diagnostic and prognostic value of GULP1 in HCC, we examined an independent cohort of 81 patients with HCC who underwent hepatic resection (Supplementary Table 1). Quantitative reverse transcription-polymerase chain reaction (qRT-PCR) analysis revealed significantly increased *GULP1* transcript levels in the tumor (T) samples compared to those in the non-tumor (NT) samples, confirming its elevation in HCC tissues (Supplementary Fig. S4D and E; Fig. 3A, left box plot in the tissue sample panel). In a separate validation cohort, we assessed the protein levels of GULP1 as a diagnostic marker in serum samples (Supplementary Table 2). Serum GULP1 levels were significantly elevated in patients with HCC compared to those in the NT group, confirming its upregulation in HCC (Fig. 3A, left box plot in the blood sample panel). GULP1 concentrations were significantly higher in patients at different HCC stages (mUICC I, I/II, and III/IV) than in those with normal liver (NL), CH, and liver cirrhosis (LC; Fig. 3A, middle scattered dot plot in the blood sample panel). Although the diagnostic power of *GULP1* at the transcriptomic level in tissue samples showed a modest AUC of 0.67 for detecting liver cancer; its diagnostic performance as a serum-based marker was more robust with an AUC of 0.85, demonstrating superior accuracy in distinguishing HCC from the normal liver and early liver diseases (Fig. 3A, ROC curves for tissue and blood sample panels).

Compared to AFP, GULP1 showed significantly higher AUC values in the high-risk group (CH/LC) for all patients with HCC (0.827 for GULP1 and 0.595 for AFP) (Fig. 3B, left panel). In the mUICC III/IV group, GULP1 exhibited a higher AUC (0.877) than AFP (0.766), although the difference was not statistically significant (Fig. 3B, second panel). In the mUICC

I/II group, GULP1 demonstrated a relatively strong diagnostic ability (AUC = 0.816) compared to that of AFP (AUC = 0.556; Fig. 3B, third panel). In the earliest stage of HCC (mUICC I, tumor size ≤ 2 cm), GULP1 showed a high AUC (0.749), whereas the diagnostic accuracy of AFP was significantly lower (AUC = 0.516; Fig. 3B, the fourth panel).

Serum GULP1 and AFP levels showed distinct positivity rates across healthy subjects and CH, LC, and HCC groups. Although GULP1 demonstrated a lower incidence than AFP in the non-HCC groups (i.e., NL, CH, and LC), it exhibited a higher positivity rate in the HCC group (Fig. 3C, left panel). Detection rates of AFP and GULP1 in all 145 patients with liver cancer were 44 and 70%, respectively. When both markers were used together, HCC positivity rate increased to 81% (Fig. 3C, right panel).

To investigate the variation in GULP1 expression levels across different etiologies of HCC, we analyzed the serum GULP1 levels in cohorts stratified by the hepatitis B virus (HBV), hepatitis C virus (HCV), alcohol, and MASH (Supplementary Table 3). GULP1 levels were significantly higher in HCC than in LC across all tested etiologies (Supplementary Fig. S5A–E). In HBV-induced liver disease, GULP1 levels were significantly higher in HCC than in LC ($P < 0.001$), with an AUC of 0.871, surpassing the diagnostic performance of AFP (AUC = 0.707; $P = 0.006$; Supplementary Fig. S5A). In HCV-, alcoholic-, and MASH-related liver diseases, GULP1 levels were consistently and significantly elevated in HCC compared to those in LC. Although ROC analysis indicated that GULP1 demonstrated superior diagnostic performance over AFP for these etiologies, the observed differences were not statistically significant (Supplementary Fig. S5B–D). We further evaluated the diagnostic potential of GULP1 in non-HCC liver tumors, including cholangiocarcinoma and combined hepatocellular-cholangiocarcinoma. GULP1 levels were significantly higher in non-HCC liver tumors than in non-tumor samples (NL and LC; $P < 0.001$). ROC analysis revealed that GULP1 exhibited an

AUC of 0.833 in distinguishing the non-tumor samples (n = 145) from the non-HCC liver tumors (n = 30), markedly outperforming AFP (AUC = 0.540; P < 0.0001; Supplementary Fig. S5E). Comparing LC to non-HCC liver tumors, GULP1 showed an AUC of 0.790, consistently surpassing AFP performance (AUC = 0.524; P = 0.002; Supplementary Fig. S5F).

Considering the strong association between GULP1 and the 15-gene RS model for predicting recurrent HCC, we further analyzed its potential as a recurrence biomarker. qRT-PCR analysis of tissue samples revealed significantly increased *GULP1* expression levels in R versus NR cases (Fig. 3D, left box plot in the tissue sample panel). Time-dependent AUROC curves were plotted to evaluate the ability of GULP1 to predict recurrence over time. Time-dependent AUC-based C-index for GULP1 in tissue samples was 0.745 (95% CI: 0.611–0.859), surpassing that of AFP 0.663 (95% CI: 0.489–0.815). Statistical comparison of mean time-dependent AUCs confirmed the superior predictive performance of GULP1 ($P = 8.82 \times 10^{-18}$; Fig. 3D, time-dependent AUC curve in the tissue samples panel). High serum GULP1 levels were significantly associated with R, further supporting its potential as a biomarker for liver cancer recurrence (Fig. 3D, left box plot in the blood sample panel). Time-dependent AUROC analysis revealed that GULP1 exhibited a slightly higher predictive capacity than AFP, with a C-index of 0.726 (95% CI: 0.623–0.826) vs. 0.703 (95% CI: 0.576–0.814), although the difference was not statistically significant (P = 0.188; Fig. 3D, time-dependent AUC curve in the blood samples panel). Western blotting analysis confirmed this upregulation, showing elevated GULP1 protein levels in RT vs. PT tissues, underscoring its involvement in recurrence mechanisms (Fig. 3E). Furthermore, tissue (left) and serum (right)-based RFS evaluations revealed that high GULP1 expression was significantly associated with a poor patient prognosis (Fig. 3F).

GULP1 promotes tumor growth, proliferation and invasiveness in HCC cells

To investigate the roles of GULP1 in tumor growth, proliferation, and invasion in HCC, we analyzed its expression levels in various liver cancer cell lines. Endogenous GULP1 levels were the highest in PLC/PRF/5 and Huh-7 cells, leading to their selection for subsequent experiments (Supplementary Fig. S6A). *GULP1* knockdown (siGULP1 group) significantly reduced the cell growth, proliferation, and clonogenic capacity. These effects were partially rescued by GULP1 overexpression (GULP1_OE; Fig. 4A). To further validate the oncogenic potential of GULP1 *in vivo*, *GULP1*-suppressed Huh-7 cells were subcutaneously injected into female BALB/c nude mice. *GULP1*-depleted cells exhibited a significantly lower growth rate than the negative control cells (Fig. 4B).

Furthermore, immunohistochemical (IHC) analysis of xenograft tumor tissue sections revealed reduced Ki-67 and PCNA expression levels in the *GULP1*-depleted group (Fig. 4C). These findings highlight the roles of GULP1 in promoting tumor growth and proliferation. However, as GULP1 was upregulated in recurrent HCC and affected RFS, we examined its impact on metastasis and invasion, which are critical phenotypes influencing HCC recurrence, using a wound-healing assay to observe the effect of GULP1 on the migration of liver cancer cells. *GULP1* knockdown significantly reduced the wound-healing capacity of these cells, whereas *GULP1* re-expression effectively restored their migratory potential (Fig. 4D). Transwell invasion assay showed a similar effect of *GULP1* expression on cell invasiveness (Fig. 4E, left panel). Three-dimensional sphere cultures showed that *GULP1* modulation significantly affected sphere formation and cell outgrowth (Fig. 4E, right panel), suggesting that GULP1 contributes to cell invasion.

In a subcutaneous xenograft model, *GULP1* knockdown significantly affected the expression levels of various markers. Specifically, expression levels of epithelial markers (E-cadherin and zonula occludens-1) were upregulated, whereas those of mesenchymal markers (vimentin, fibronectin, and slug) and angioinvasion markers (CD31 and vascular endothelial growth factor [VEGF]) were downregulated upon *GULP1* suppression (Fig. 4F).

Validation of GULP1 roles in promoting HCC recurrence and metastasis *in vivo*

To validate the effect of GULP1 on HCC recurrence *in vivo*, GULP1-suppressed Hepa1-6 cells were orthotopically injected into the mouse liver and recurrent tumors were resected for evaluation (Fig. 5A). IHC analysis revealed a significant increase in GULP1 expression levels, particularly in RT tissues, compared to those in NL and PT tissues, highlighting its role in tumor recurrence and progression (Fig. 5B; Supplementary Fig. S6B). However, GULP1 suppression significantly reduced the number, size, and weight of recurrent tumors (Fig. 5C).

IHC analysis also demonstrated changes in the expression levels of key markers in the *GULP1*-suppressed group (Fig. 5D). Specifically, decreased expression of the mesenchymal marker (vimentin) indicated reduced tumor invasion and metastasis. Increase in the levels of epithelial markers, such as E-cadherin, indicated the reversal of EMT, a critical step in cancer metastasis. Significant changes were observed in PT, with more pronounced alterations in RT than in the normal tissues, indicating the substantial impacts of GULP1 suppression on recurrent tumors (Fig. 5D).

The above-mentioned findings were further validated using a lung metastasis model of *ras*-transformed NIH-3T3 cells. GULP1 expression was regulated followed by tail vein injection, further supporting the metastatic role of GULP1 in promoting tumor recurrence (Supplementary Fig. S6C). Upon resection, metastatic nodules were significantly reduced in the GULP1-suppressed group (Fig. 5E, left panel). Hematoxylin and eosin (H&E) staining showed small metastatic lesions in the lungs of the GULP1-suppressed group, reinforcing the reduction in metastasis (Fig. 5E, right panel). Additionally, analysis of RNA-seq data of the biopsies of patients with liver cancer (GSE164359) revealed higher GULP1 expression levels in the RT than in the adjacent liver (AL) and PT (Fig. 5F).

Notably, no significant differences in body weight were observed in the three *in vivo* models after GULP1 knockdown (Supplementary Fig. S6D), indicating the potential of GULP1-targeting therapeutics for recurrent and metastatic HCC.

Mechanistic role of GULP1 in modulating β-catenin signaling in HCC

DEAD-box helicase 5 directly interacts with β-catenin, facilitating its nuclear translocation and transactivation. This interaction is associated with alterations in the expression of GULP1, a protein associated with neuroblastoma progression.[13] In our initial assessment of liver cancer cell lines, β-catenin suppression considerably decreased the GULP1 protein levels (Fig. 6A and B). Interestingly, modulation of GULP1 expression significantly influenced β-catenin localization, as observed via fluorescence microscopy and supported by quantitative analysis, indicating a strong correlation between GULP1 levels and β-catenin subcellular distribution (Fig. 6C and D).

To further elucidate the mechanisms underlying GULP1-mediated modulation of β-catenin translocation, we investigated its interaction with ADP-ribosylation factor 6 (ARF6)-GTP based on the proposed role of GULP1 in stabilizing ARF6 activity.[14] Consistently, our results revealed that changes in ARF6-GTP were barely affected by the reduction in GULP1 levels (Fig. 6E). However, in the presence of cycloheximide (CHX), a protein synthesis inhibitor, GULP1 depletion significantly decreased the ARF6-GTP levels. Remarkably, GULP1 re-expression under these conditions successfully restored the ARF6-GTP levels, confirming the role of GULP1 in stabilizing ARF6 activity (Fig. 6E). Direct interaction between GULP1 and

ARF6 was further validated via fluorescence resonance energy transfer-based assay. Sm-Bit-tagged ARF6 and Lg-Bit-tagged GULP1 were used to measure the luminescence signals indicative of their interactions (Supplementary Fig. S7A). Indeed, luminescence signals were significantly reduced when GULP1 was depleted, confirming the dependence of the signals on GULP1 presence. In contrast, luminescence signals were successfully rescued when GULP1 was overexpressed in the depleted state (Supplementary Fig. S7B). When ARF6 expression was depleted, luminescence signals decreased, confirming the direct interaction between GULP1 and ARF6 (Supplementary Fig. S7C). Luminescence signals were significantly reduced upon β-catenin depletion (siCTNNB1 group), indicating that β-catenin activity is essential for ARF6–GULP1 interaction (Supplementary Fig. S7D). These findings demonstrate the dependency of GULP1-mediated β-catenin translocation on ARF6 activation, as knockdown of ARF6 suppressed β-catenin localization (Fig. 6F). Furthermore, ARF6 downregulation was accompanied by a decrease in GULP1 expression, reinforcing their relationship (Supplementary Fig. S7E).

Bioinformatics analyses identified a transcription factor 3 (TCF3)-binding motif within the *GULP1* promoter region (chr 2:189,158,687-189,158,701), suggesting direct regulation by TCF3 (Supplementary Fig. S7F). Luciferase promoter mutation reporter assay was performed to confirm the binding of TCF3 to the *GULP*1 promoter. Mutations in the TCF3 binding site significantly reduced the luciferase activity, suggesting that TCF3 directly binds to and regulates the *GULP1* promoter (Fig. 6G).

ARF6 knockdown decreased the binding affinity of β-catenin and TCF3 to the GULP1 promoter, as demonstrated by the chromatin immunoprecipitation assays (Fig. 6H; Supplementary Fig. S7G). Modulation of GULP1 expression further altered the TCF3 binding affinity to the *GULP1* promoter region (Fig. 6I), indicating that these proteins form a complex involving direct binding between β-catenin and TCF3 at the promoter region of GULP1.

GULP1 regulated β-catenin interactions with key adhesion molecules, including N-cadherin and E-cadherin (Fig. 6J). This regulatory capacity of GULP1 extended to the expression of downstream β-catenin targets, such as SRY-box transcription factor 9, c-Myc, and fibronectin, indicating a positive feedback mechanism in the β-catenin signaling pathway that significantly influences EMT and HCC progression (Fig. 6K). GULP1 suppression reduced β-catenin expression in the *in vivo* experiment (Supplementary Fig. S7H). Further analysis of the GSE164359 dataset revealed that several key genes, including *MKI67*, *CTNNB1*, *VIM*, *VEGFA*, *FN1*, and *SOX9*, exhibited elevated expression levels in RT, similar to the *GULP1* expression levels (Supplementary Fig. S7I). Correlation analysis revealed a moderately positive relationship between GULP1 and these gene expression levels, suggesting that GULP1 influences these oncogenic pathways in recurrent HCC (Supplementary Fig. S7J).

In summary, our study identified GULP1 as a β-catenin signaling modulator in HCC. By stabilizing ARF6-GTP, GULP1 facilitated β-catenin localization and transcriptional activation via TCF3 binding at its promoter, thereby driving oncogenic processes, such as EMT and tumor progression (Fig. 7).

DISCUSSION

Various guidelines have been established for primary liver cancer; however, management of recurrent liver cancer remains challenging despite various treatment options, such as hepatectomy, radiotherapy, transplantation, and systemic therapy.[15, 16] Despite surgical resection being the preferred treatment, three-year recurrence rate of HCC remains high.[17] For small recurrent HCC cases following surgical resection, radiofrequency ablation results in overall survival and RFS rates comparable to those observed with repeated surgical resection.[18]

Despite advancements in medical techniques, prediction of recurrent HCC remains challenging. Although such predictions can improve the patient survival rates by enabling proactive measures, accurate HCC recurrence prediction models are lacking. Conventional HCC markers, such as AFP, often exhibit low sensitivity and specificity in diagnosing various liver cancer types; therefore, many studies are exploring novel serum biomarkers, such as glypican-3, and multigene biomarker panels for liver cancer.[19, 20] The near absence of reliable recurrence predictors underscores the importance of our study in this field.

Thorough understanding of EMT, a hallmark of recurrence, is pivotal to predict HCC recurrence and its detrimental effect on the patient overall survival.[21] EMT involves a transition, in which epithelial cells gain mesenchymal features and exhibit enhanced migration and invasion.[22-24] This transition is accompanied by changes in cell morphology, reduced cell–cell adhesion, and increased migration and invasion. Several markers implicated in EMT are recognized as key players in HCC recurrence, including transcription factors Snail, Slug, and Twist.[25] Downregulation of the levels of epithelial markers, such as E-cadherin and cytokeratin, and upregulation of the levels of mesenchymal markers, including N-cadherin, vimentin, and fibronectin, are commonly observed during EMT.[26] Specifically, Wnt/β-catenin pathway is also implicated in EMT regulation, as aberrant activation of β-catenin drives EMT and promotes HCC progression and recurrence.[27-29]

Building on these, this study explored the oncogenic role of GULP1 in HCC metastasis. Interestingly, GULP1 predominantly acts as a tumor suppressor in many cancer types. For instance, its reduced expression in ovarian cancer, attributed to epigenetic silencing via genomic methylation, is associated with advanced disease stages and unfavorable prognosis.[7]

In urothelial carcinoma, downregulation of GULP1 levels induces cell growth, predominantly via activation of the NRF2–KEAP1 signaling pathway.[8] However, our findings revealed its different roles in HCC, where it functioned as an oncogene. Specifically, GULP1 inhibition reduced β-catenin activity, while its overexpression enhanced it. This modulation was achieved via ARF6 activation. The role of GULP1 in oncogenesis is thus context-dependent, akin to other cancer-associated genes with dual functions, such as ARID1A and HDAC6.[30, 31] While GULP1 promotes β-catenin activation and EMT in HCC—facilitating tumor progression and recurrence—it suppresses tumorigenesis in ovarian and urothelial. These divergent roles underscore the importance of the tumor microenvironment and pathway-specific interactions in defining GULP1's function. This not only broadens our understanding of the complex associations among GULP1, β-catenin signaling, and EMT but also highlights the pivotal role of EMT in cancer metastasis.[32] Additionally, experimental results further suggest that GULP1 functions as an oncogene. First, we demonstrated its elevation in both tissue and serum of HCC patients, as well as its association with recurrence through time-dependent AUROC analyses. Second, functional assays *in vitro* revealed that GULP1 knockdown dampens tumor cell proliferation, migratory capacity, and invasive behavior, while rescue experiments reversed these effects. Third, *in vivo* models established GULP1's involvement in tumorigenesis and metastasis—particularly in recurrent settings—suggesting its pivotal role in both the initiation and progression of HCC.

We anticipate that GULP1 serves as an oncogenic factor in HCC through diverse mechanisms. Notably, our GSEA findings indicate that GULP1 is linked not only to the Wnt pathway but also to other key signaling cascades such as NOTCH and HEDGEHOG (Supplementary Fig. S8A-D). These pathways are all well-documented to play crucial roles in HCC onset and progression, suggesting that GULP1 may integrate multiple oncogenic signals. Future research efforts should focus on elucidating these interactions to provide a more comprehensive understanding of GULP1-driven hepatocarcinogenesis.

Nevertheless, we acknowledge that GULP1 expression may also fluctuate due to liver inflammation, cirrhosis, and coexisting treatments. These factors highlight the need for additional large-scale, multicenter validation studies to refine assay cutoffs and adjust for confounding clinical variables. While our mechanistic data pinpoint GULP1's role in ARF6-mediated β-catenin activation, further targeted approaches could more definitively validate these interactions in HCC recurrence. In addition, its modest predictive power in serum underscores the need for further studies to refine detection thresholds, validate assay methods to enhance overall diagnostic accuracy and clinical utility.

Looking ahead, we propose future studies to (1) explore combinational biomarker panels—including GULP1 with AFP or other emergent markers—to enhance predictive accuracy, (2) systematically investigate how GULP1 cooperates with Wnt, Notch, and Hedgehog pathways, and (3) develop therapeutic interventions targeting GULP1-driven ARF6–β-catenin signaling in recurrent HCC. By addressing these gaps, we believe GULP1 may evolve into not only a valuable clinical biomarker but also a therapeutic target for controlling HCC progression and recurrence.

Overall, this study revealed GULP1 as a key biomarker for HCC, acting both as a non-invasive diagnostic tool and an oncogene driving HCC progression. Our findings highlight its potential for improving HCC treatment strategies, particularly for developing therapeutics targeting EMT and cancer cell growth. However, further research and clinical validation are crucial to fully harness the potential of GULP1 and substantially advance HCC treatment.

ACKNOWLEDGEMENTS

This work was supported by grants from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute funded by the Ministry of Health and Welfare, Republic of Korea (HR21C1003), Bio and Medical Technology Development Program of the National Research Foundation (NRF-2022R1A2C2092422 and RS-2023-00210847), and Commercialization Promotion Agency for R&D Outcomes (COMPA) funded by the Ministry of Science and ICT (RS-2024-00422549).

Biospecimens and data used in this study were provided by the Biobank of the Ajou University Hospital, a member of the Korea Biobank Network. Additionally, we would like to thank all members of MOAGEN (Daejeon, South Korea) for assistance and guidance with bioinformatics analysis. We would like to thank Editage (www.editage.com) for English language editing. For this study, a graphical abstract (*JZ27K8N3P9*; https://biorender.com/w63d148*)* was created using BioRender (www.biorender.com). ChatGPT (version 4o) was used for error checking and language refinement during the preparation of this manuscript.

FOOTNOTES

Authors' contributions

H.S.K. and J.H.Y. contributed equally to this study. Conceptualization: H.J.C., S.S.K., J.W.E.; Data curation: H.S.K., J.H.Y., M.G.Y., G.O.B., J.E.H.; Methodology: M.G.Y., G.O.B., J.-Y.J., S.W.N.; Investigation: : H.S.K., J.H.Y., W.P., Y.G., J.T.N., S.B.L.; Resources: M.K., S.W.N., J.E.H., H.J.C., S.S.K., J.Y.C.; Visualization: G.O.B., J.C., S.H.J., J.E.H.; Funding acquisition: J.Y.C., H.S.K., J.W.E.; Project administration: S.S.K., J.W.E.; Supervision: J.Y.C.; Writing – original draft: : H.S.K., J.H.Y., H.J.C., J.W.E.; Writing – review & editing: S.S.K., J.Y.C., J.W.E. All authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Vogel A, Meyer T, Sapisochin G, Salem R, Saborowski A. Hepatocellular carcinoma. Lancet. 2022;400:1345-62.

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. Hepatol Res. 2019;49:1114-20.

3. Huang DQ, Mathurin P, Cortez-Pinto H, Loomba R. Global epidemiology of alcohol-associated cirrhosis and HCC: trends, projections and risk factors. Nat Rev Gastroenterol Hepatol. 2023;20:37-49.

4. Bruix J, Takayama T, Mazzaferro V, Chau GY, Yang J, Kudo M, et al. Adjuvant sorafenib for hepatocellular carcinoma after resection or ablation (STORM): a phase 3, randomised, double-blind, placebo-controlled trial. Lancet Oncol. 2015;16:1344-54.

5. Beyer AS, von Einem B, Schwanzar D, Keller IE, Hellrung A, Thal DR, et al. Engulfment adapter PTB domain containing 1 interacts with and affects processing of the amyloid-beta precursor protein. Neurobiol Aging. 2012;33:732-43.

6. Gong J, Gaitanos TN, Luu O, Huang Y, Gaitanos L, Lindner J, et al. Gulp1 controls Eph/ephrin trogocytosis and is important for cell rearrangements during development. J Cell Biol. 2019;218:3455-71.

7. Maldonado L, Brait M, Izumchenko E, Begum S, Chatterjee A, Sen T, et al. Integrated transcriptomic and epigenomic analysis of ovarian cancer reveals epigenetically silenced GULP1. Cancer Lett. 2018;433:242-51.

8. Hayashi M, Guida E, Inokawa Y, Goldberg R, Reis LO, Ooki A, et al. GULP1 regulates the NRF2-KEAP1 signaling axis in urothelial carcinoma. Sci Signal. 2020;13.

9. Du B, Wang F, Jarad B, Wang Z, Zhang Y. A novel signature based on microvascular invasion predicts the recurrence of HCC. J Transl Med. 2020;18:272.

10. Liu M, Casimiro MC, Wang C, Shirley LA, Jiao X, Katiyar S, et al. p21CIP1 attenuates

Ras- and c-Myc-dependent breast tumor epithelial mesenchymal transition and cancer stem cell-like gene expression in vivo. Proc Natl Acad Sci U S A. 2009;106:19035-9.

11. Yang L, Wang M, Wang Y, Zhu Y, Wang J, Wu M, et al. LINC00460-FUS-MYC feedback loop drives breast cancer metastasis and doxorubicin resistance. Oncogene. 2024;43:1249-62.

12. Liu F, Liao Z, Zhang Z. MYC in liver cancer: mechanisms and targeted therapy opportunities. Oncogene. 2023;42:3303-18.

13. Zhao X, Li D, Yang F, Lian H, Wang J, Wang X, et al. Long Noncoding RNA NHEG1 Drives beta-Catenin Transactivation and Neuroblastoma Progression through Interacting with DDX5. Mol Ther. 2020;28:946-62.

14. Ma Z, Nie Z, Luo R, Casanova JE, Ravichandran KS. Regulation of Arf6 and ACAP1 signaling by the PTB-domain-containing adaptor protein GULP. Curr Biol. 2007;17:722-7.

15. Saraswat VA, Pandey G, Shetty S. Treatment algorithms for managing hepatocellular carcinoma. J Clin Exp Hepatol. 2014;4:S80-9.

16. Zhou J, Sun HC, Wang Z, Cong WM, Wang JH, Zeng MS, et al. Guidelines for Diagnosis and Treatment of Primary Liver Cancer in China (2017 Edition). Liver Cancer. 2018;7:235-60.

17. Gosalia A, Martin P, Jones P. Advances and Future Directions in the Treatment of Hepatocellular Carcinoma. Gastroenterol Hepatol (N Y) 2017;13:398-410.

18. Li JK, Liu XH, Cui H, Xie XH. Radiofrequency ablation vs. surgical resection for resectable hepatocellular carcinoma: A systematic review and meta-analysis. Mol Clin Oncol. 2020;12:15-22.

19. Qiao SS, Cui ZQ, Gong L, Han H, Chen PC, Guo LM, et al. Simultaneous measurements of serum AFP, GPC-3 and HCCR for diagnosing hepatocellular carcinoma. Hepatogastroenterology. 2011;58:1718-24.

20. Kaur H, Dhall A, Kumar R, Raghava GPS. Identification of Platform-Independent Diagnostic Biomarker Panel for Hepatocellular Carcinoma Using Large-Scale

Transcriptomics Data. Front Genet. 2019;10:1306.

21. van Zijl F, Zulehner G, Petz M, Schneller D, Kornauth C, Hau M, et al. Epithelial-mesenchymal transition in hepatocellular carcinoma. Future Oncol. 2009;5:1169-79.

22. Brabletz T, Kalluri R, Nieto MA, Weinberg RA. EMT in cancer. Nat Rev Cancer. 2018;18:128-34.

23. Ribatti D, Tamma R, Annese T. Epithelial-Mesenchymal Transition in Cancer: A Historical Overview. Transl Oncol. 2020;13:100773.

24. Huang Y, Hong W, Wei X. The molecular mechanisms and therapeutic strategies of EMT in tumor progression and metastasis. J Hematol Oncol. 2022;15:129.

25. Yang MH, Chen CL, Chau GY, Chiou SH, Su CW, Chou TY, et al. Comprehensive analysis of the independent effect of twist and snail in promoting metastasis of hepatocellular carcinoma. Hepatology. 2009;50:1464-74.

26. Zeisberg M, Neilson EG. Biomarkers for epithelial-mesenchymal transitions. J Clin Invest. 2009;119:1429-37.

27. Deldar Abad Paskeh M, Mirzaei S, Ashrafizadeh M, Zarrabi A, Sethi G. Wnt/beta-Catenin Signaling as a Driver of Hepatocellular Carcinoma Progression: An Emphasis on Molecular Pathways. J Hepatocell Carcinoma. 2021;8:1415-44.

28. Zhang Q, Bai X, Chen W, Ma T, Hu Q, Liang C, et al. Wnt/beta-catenin signaling enhances hypoxia-induced epithelial-mesenchymal transition in hepatocellular carcinoma via crosstalk with hif-1alpha signaling. Carcinogenesis. 2013;34:962-73.

29. Yang M, Li SN, Anjum KM, Gui LX, Zhu SS, Liu J, et al. A double-negative feedback loop between Wnt-beta-catenin signaling and HNF4alpha regulates epithelial-mesenchymal transition in hepatocellular carcinoma. J Cell Sci. 2013;126:5692-703.

30. Sun X, Wang SC, Wei Y, Luo X, Jia Y, Li L, et al. Arid1a Has Context-Dependent Oncogenic and Tumor Suppressor Functions in Liver Cancer. Cancer Cell. 2018;33:151-2.

31. Jung KH, Noh JH, Kim JK, Eun JW, Bae HJ, Chang YG, et al. Histone deacetylase 6

functions as a tumor suppressor by activating c-Jun NH2-terminal kinase-mediated beclin 1-dependent autophagic cell death in liver cancer. Hepatology. 2012;56:644-57.

32. Leung RWH, Lee TKW. Wnt/beta-Catenin Signaling as a Driver of Stemness and Metabolic Reprogramming in Hepatocellular Carcinoma. Cancers (Basel). 2022;14.

FIGURE LEGENDS

Figure 1. Identification of differentially expressed genes (DEGs) in recurrent hepatocellular carcinoma (HCC), and development of a risk score (RS) model. (A) Flow chart of the selection of genes associated with recurrent HCC. Patient selection criteria included patients who underwent surgical liver resection and remained recurrence-free for over two years (non-recurrence [NR]) and those who experienced recurrence (R). (B) Left: Coefficient profiles of correlated DEGs in least absolute shrinkage and selection operator (LASSO) analysis. Right: LASSO model parameter adjustment with 10-fold cross-validation. Blue dashed line indicates the optimal lambda value cutoff (0.1829059). (C) Correlation analysis of RS and patient survival using three HCC datasets: GSE14520, GSE114564, and The Cancer Genome Atlas (TCGA) liver hepatocellular carcinoma (LIHC). (D) Predictive potency determined via receiver operating characteristic (ROC) analysis. (E) Risk-based analysis of recurrence-free survival (RFS) in the GSE14520, GSE114564, and TCGA LIHC datasets using the Kaplan–Meier approach. (F) Forest plots of univariate Cox regression analyses of clinical parameters affecting RFS. HCC differentiation was defined using the Edmondson grade scale. Statistical significance was determined via Cox proportional hazards regression analysis and log-rank tests for survival data. *$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$.

Figure 1

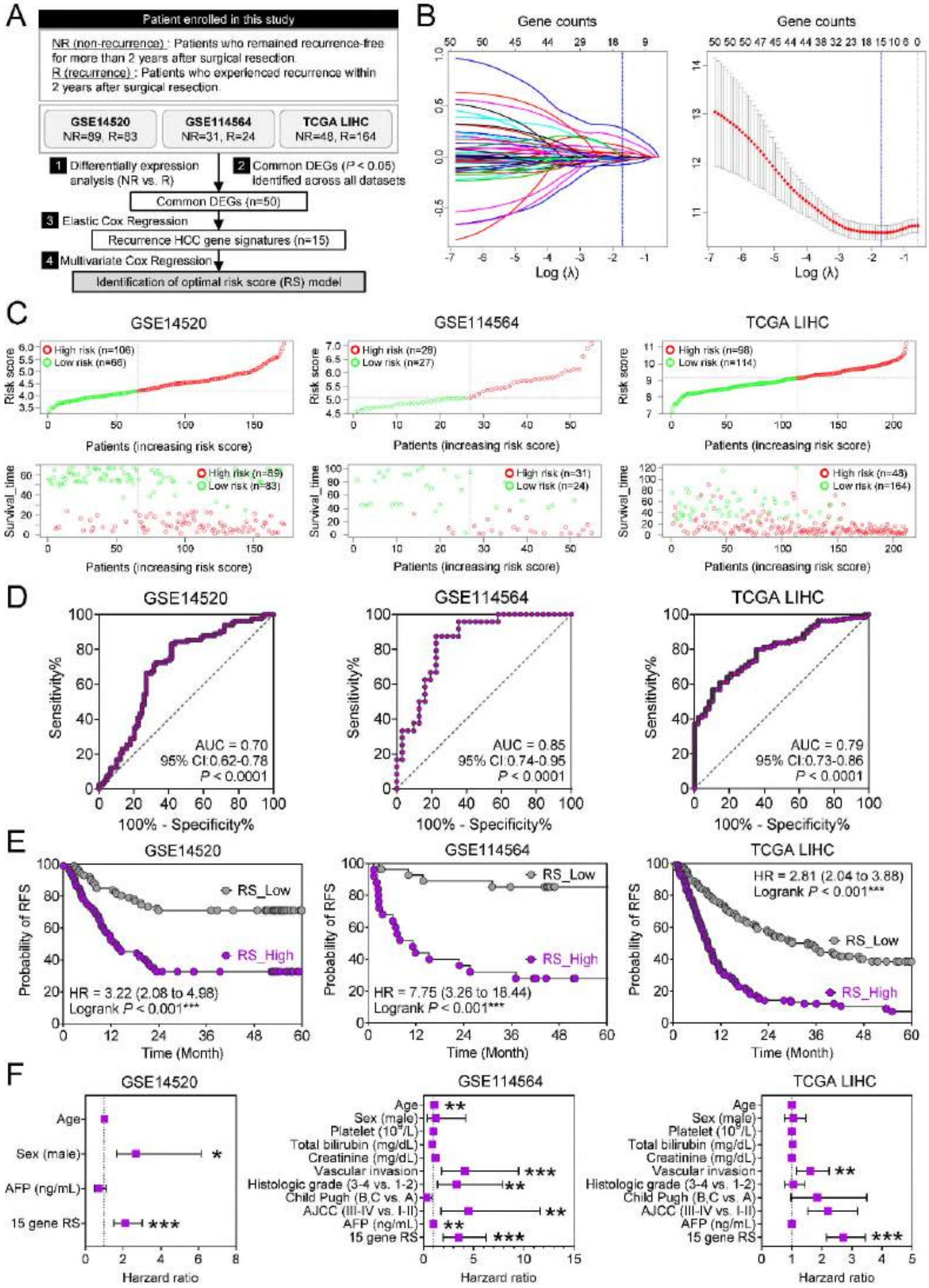

A
Patient enrolled in this study
NR (non-recurrence) : Patients who remained recurrence-free for more than 2 years after surgical resection.
R (recurrence) : Patients who experienced recurrence within 2 years after surgical resection.
GSE14520 NR=89, R=83
GSE114564 NR=31, R=24
TCGA LIHC NR=48, R=164
1 Differentially expression analysis (NR vs. R)
2 Common DEGs (P < 0.05) identified across all datasets
Common DEGs (n=50)
3 Elastic Cox Regression
Recurrence HCC gene signatures (n=15)
4 Multivariate Cox Regression
Identification of optimal risk score (RS) model
B
Gene counts
Log (λ)
C
GSE14520
GSE114564
TCGA LIHC
High risk (n=106)
Low risk (n=66)
High risk (n=28)
Low risk (n=27)
High risk (n=98)
Low risk (n=114)
Risk score
Patients (increasing risk score)
High risk (n=89)
Low risk (n=83)
High risk (n=31)
Low risk (n=24)
High risk (n=48)
Low risk (n=164)
Survival_time
D
GSE14520
AUC = 0.70
95% CI:0.62-0.78
P < 0.0001
GSE114564
AUC = 0.85
95% CI:0.74-0.95
P < 0.0001
TCGA LIHC
AUC = 0.79
95% CI:0.73-0.86
P < 0.0001
Sensitivity%
100% - Specificity%
E
GSE14520
RS_Low
RS_High
HR = 3.22 (2.08 to 4.98)
Logrank P < 0.001***
GSE114564
HR = 7.75 (3.26 to 18.44)
Logrank P < 0.001***
TCGA LIHC
HR = 2.81 (2.04 to 3.88)
Logrank P < 0.001***
Probability of RFS
Time (Month)
F
GSE14520
Age
Sex (male)
AFP (ng/mL)
15 gene RS
GSE114564
Platelet (10⁹/L)
Total bilirubin (mg/dL)
Creatinine (mg/dL)
Vascular invasion
Histologic grade (3-4 vs. 1-2)
Child Pugh (B,C vs. A)
AJCC (III-IV vs. I-II)
TCGA LIHC
Harzard ratio

Figure 2. *GULP1* is a HCC progression indicator. (A) Heatmap of GULP1 expression levels (Z-Score of log2(TPM+1)) in the normal, adjacent non-HCC (ADJ_HCC), and HCC tissues across 22 datasets from GepLiver DB. (B) Comparison of GULP1 expression levels across different liver phenotypes using data from GepLiver DB. Normal, n=362; viral hepatitis, n=180; nonalcoholic fatty liver disease (NAFLD), n=503; cirrhosis, n=63; HCC, n=724. (C) Spatial transcriptomic analysis of the tumor tissues. *GULP1* expression levels were notably higher in the malignant hepatocytes (purple) of the same tissue, with minimal to no expression observed in the non-malignant hepatocytes (yellow). P9T and P10T indicate the unique patient IDs. (D) Proportion of GULP1-positive cells (%) across all analyzed tumor tissues from spatial transcriptomics. The labels (P1T, P2T, etc.) indicate patient IDs. (E) Uniform manifold approximation and projection (UMAP) plot of an integrated liver single-cell RNA-sequencing (scRNA-seq) dataset from GepLiver DB. Left: Cells colored by major cell type, indicating the distribution of cell types (e.g., hepatocytes, cholangiocytes, and immune cells) across different phenotypes. Right: Proportion of *GULP1*-positive cells (%) among hepatocytes in the normal, ADJ_HCC, and HCC tissues. (F) Left: UMAP plot of scRNA-seq data from the scAtlasLC database. Right: Proportion of *GULP1*-positive cells (%) among non-malignant and malignant hepatocytes from the scAtlasLC database. (G) Mean expression levels of *GULP1* in different liver-associated cell types in the GSE151530 dataset. (H) UMAP plot showing the *GULP1*-positive (+) and *GULP1*-negative (−) hepatocytes among the analyzed 18,539 hepatocytes. (I) Enrichment score plots of the hallmark pathways with *GULP1*-positive hepatocytes derived via single-sample gene set enrichment analysis (ssGSEA).

Figure 2

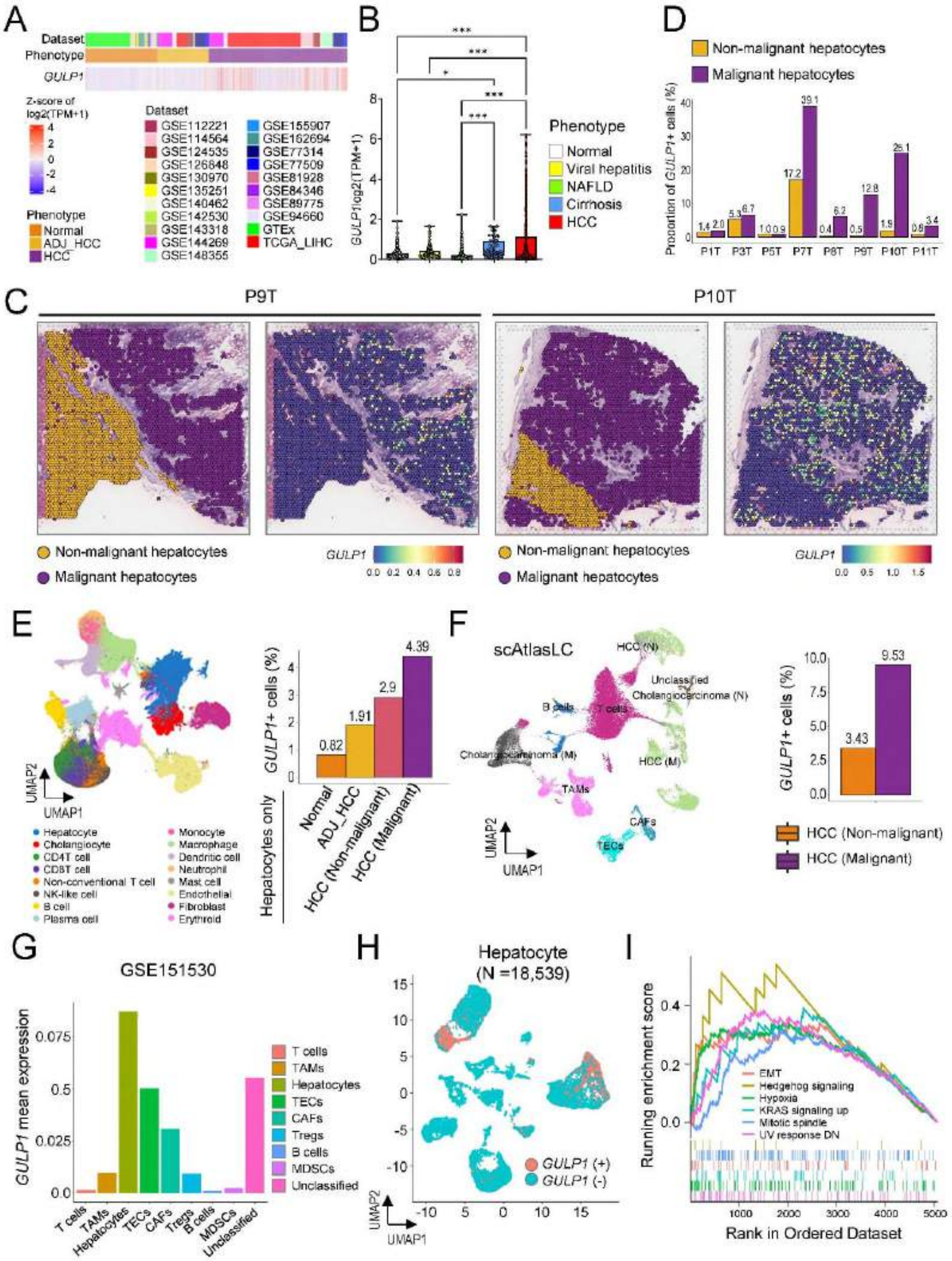

A
Dataset
Phenotype
GULP1
Z-score of log2(TPM+1)
4
2
0
-2
-4
Phenotype
Normal
ADJ_HCC
HCC
Dataset
GSE112221
GSE114564
GSE124535
GSE126848
GSE130970
GSE135251
GSE140462
GSE142530
GSE143318
GSE144269
GSE148355
GSE155907
GSE162694
GSE77314
GSE77509
GSE81928
GSE84346
GSE89775
GSE94660
GTEx
TCGA_LIHC
B
GULP1 log2(TPM+1)
0
2
4
6
8
***
***
*
***
***
Phenotype
Normal
Viral hepatitis
NAFLD
Cirrhosis
HCC
D
Non-malignant hepatocytes
Malignant hepatocytes
Proportion of GULP1+ cells (%)
0
10
20
30
40
1.4
2.0
5.3
6.7
1.0
0.9
17.2
39.1
0.4
6.2
0.5
12.8
1.9
25.1
0.8
3.4
P1T
P3T
P5T
P7T
P8T
P9T
P10T
P11T
C
P9T
P10T
Non-malignant hepatocytes
Malignant hepatocytes
GULP1
0.0
0.2
0.4
0.6
0.8
Non-malignant hepatocytes
Malignant hepatocytes
GULP1
0.0
0.5
1.0
1.5
E
UMAP2
UMAP1
Hepatocyte
Cholangiocyte
CD4T cell
CD8T cell
Non-conventional T cell
NK-like cell
B cell
Plasma cell
Monocyte
Macrophage
Dendritic cell
Neutrophil
Mast cell
Endothelial
Fibroblast
Erythroid
GULP1+ cells (%)
0
1
2
3
4
0.82
1.91
2.9
4.39
Normal
ADJ_HCC
HCC (Non-malignant)
HCC (Malignant)
Hepatocytes only
F
scAtlasLC
HCC (N)
Unclassified
Cholangiocarcinoma (N)
B cells
T cells
Cholangiocarcinoma (M)
HCC (M)
TAMs
CAFs
TECs
UMAP2
UMAP1
GULP1+ cells (%)
0.0
2.5
5.0
7.5
10.0
3.43
9.53
HCC (Non-malignant)
HCC (Malignant)
G
GSE151530
GULP1 mean expression
0.0
0.025
0.5
0.075
T cells
TAMs
Hepatocytes
TECs
CAFs
Tregs
B cells
MDSCs
Unclassified
H
Hepatocyte
(N =18,539)
15
10
5
0
-5
-10
-10
5
0
10
15
UMAP2
UMAP1
GULP1 (+)
GULP1 (-)
I
Running enrichment score
0.0
0.2
0.4
EMT
Hedgehog signaling
Hypoxia
KRAS signaling up
Mitotic spindle
UV response DN
1000
2000
3000
4000
5000
Rank in Ordered Dataset

Figure 3. Clinical significance of GULP1 expression in HCC. (A) GULP1 levels in the paired tissue (left) and blood (right) samples. GULP1 levels were significantly elevated in the HCC tissues and serum samples of patients with HCC compared to those in the non-tumor (NT) groups. ROC analysis showed the high diagnostic accuracy of serum GULP1 in distinguishing HCC from NT and early liver disease cases. (B) ROC analysis results of GULP1 and AFP in high-risk liver disease groups (CH/LC) and across mUICC stages. (C) Left: Comparison of the serum GULP1 and AFP positivity rates among different groups (NL, chronic hepatitis [CH], liver cirrhosis [LC], and HCC). Right: Positivity rates of AFP, GULP1, and their combination in patients with liver cancer. (D) Comparison of GULP1 levels in recurrent (R) vs. non-recurrent (NR) cases using the tissue (left panels) and serum (right panels) samples. GULP1 levels are elevated in R cases, showing superior time-dependent predictive performance in tissues and comparable but slightly higher performance than AFP in the serum samples for recurrence prediction. (E) Left: Western blotting analysis of GULP1 protein levels in the primary tumor tissues (PTs) and recurrent tumor tissues (RTs) of three patients. #P01, #P02, and #P03 indicate the unique IDs of patients with HCC. Right: Densitometry analysis shows the increased GULP1 levels in recurrent tumors. (F) RFS analysis based on GULP1 expression in the tissue (left) and serum (right) samples. High GULP1 expression was associated with significantly poor prognosis in both tissue and serum evaluations in the validation cohort. Statistical significance was determined via unpaired *t*-tests for within-group comparisons and one-way analysis of variance (ANOVA) for multi-group comparisons. ROC curve analysis was performed for diagnostic evaluations. $^{*}P < 0.05$, $^{**}P < 0.01$, and $^{***}P < 0.001$. Data are represented as the mean ± standard error of the mean (SEM).

Figure 3

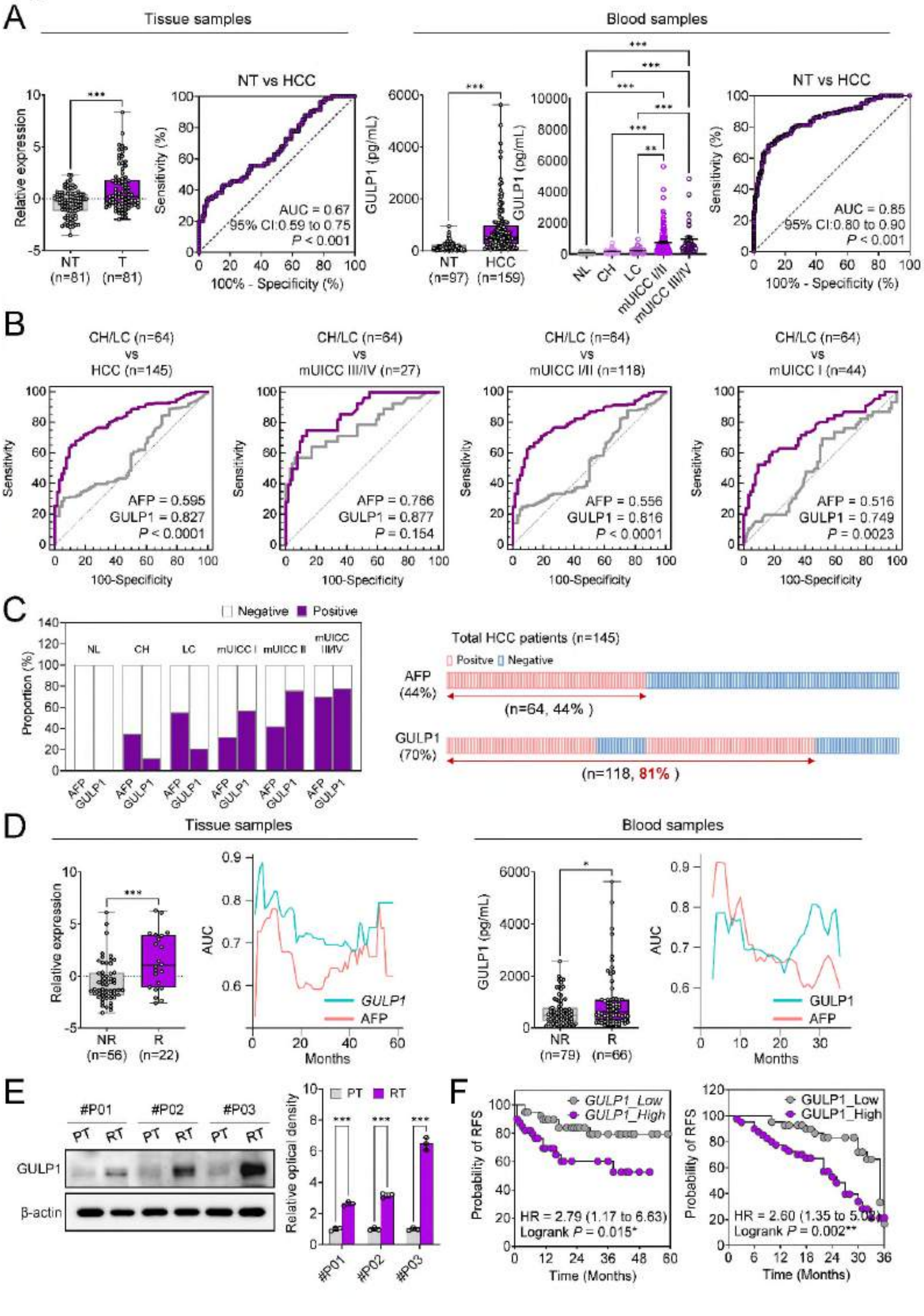

A
Tissue samples
Blood samples
Relative expression
***
NT (n=81)
T (n=81)
NT vs HCC
Sensitivity (%)
AUC = 0.67
95% CI:0.59 to 0.75
P < 0.001
100% - Specificity (%)
GULP1 (pg/mL)
***
NT (n=97)
HCC (n=159)
***
**
NL
CH
LC
mUICC I/II
mUICC III/IV
NT vs HCC
AUC = 0.85
95% CI:0.80 to 0.90
P < 0.001
B
CH/LC (n=64) vs HCC (n=145)
AFP = 0.595
GULP1 = 0.827
P < 0.0001
CH/LC (n=64) vs mUICC III/IV (n=27)
AFP = 0.766
GULP1 = 0.877
P = 0.154
CH/LC (n=64) vs mUICC I/II (n=118)
AFP = 0.556
GULP1 = 0.816
P < 0.0001
CH/LC (n=64) vs mUICC I (n=44)
AFP = 0.516
GULP1 = 0.749
P = 0.0023
Sensitivity
100-Specificity
C
Negative
Positive
Proportion (%)
NL
CH
LC
mUICC I
mUICC II
mUICC III/IV
AFP
GULP1
Total HCC patients (n=145)
Positve
Negative
AFP (44%)
(n=64, 44%)
GULP1 (70%)
(n=118, 81%)
D
Tissue samples
Relative expression
***
NR (n=56)
R (n=22)
AUC
GULP1
AFP
Months
Blood samples
GULP1 (pg/mL)
*
NR (n=79)
R (n=66)
AUC
GULP1
AFP
Months
E
#P01
#P02
#P03
PT
RT
GULP1
β-actin
Relative optical density
PT
RT
***
#P01
#P02
#P03
F
GULP1_Low
GULP1_High
Probability of RFS
HR = 2.79 (1.17 to 6.63)
Logrank P = 0.015*
Time (Months)
GULP1_Low
GULP1_High
Probability of RFS
HR = 2.60 (1.35 to 5.03)
Logrank P = 0.002**
Time (Months)

Figure 4. Targeted inactivation of GULP1 suppressed the tumorigenic potential of liver cancer cells. (A) Effects of GULP1 modulation on HCC cell growth, proliferation, and colony formation. Left: 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay showed the regulated cell growth of PLC/PRF/5 and Huh-7 cells upon the alteration of GULP1 expression. Middle: 5-Bromo-2′-deoxyuridine (BrdU) incorporation assay indicated the regulated proliferation of *GULP1*-modulated cells. Right: Clonogenic assay revealed the significantly lower colony formation capacity of siGULP1-treated cells compared to that of the controls. siCtrl, scrambled sequence of single interference control RNA; siGULP1, small interfering RNA targeting GULP1; GULP1_OE, GULP1 overexpression (B) Left: Subcutaneous xenograft tumor growth assay revealed that the *GULP1*-depleted Huh-7 cells exhibited significantly lower growth rates than the control cells. Right: Tumor weight differences in xenografts with *GULP1* knockdown. (C) Left: Histopathological examination of tumor sections via hematoxylin and eosin (H&E) staining and immunohistochemistry (IHC) for GULP1, Ki67, and proliferating cell nuclear antigen (PCNA). Right: Inter-group comparative analysis of IHC staining intensity for GULP1, Ki67, and PCNA. (D) Scratch wound-healing assay revealed the effect of GULP1 on cell migration. Left: Representative images of wound closure in PLC/PRF/5 and Huh-7 cells at 24-h post-scratch. Right: Quantification of wound closure, showing significantly regulated migration of GULP1-modulated cells. (E) Effects of GULP1 on cell invasion and sphere formation. Left: Invasion assay showed the altered invasiveness of PLC/PRF/5 and Huh-7 cells after GULP1 modulation. Right: Sphere formation assay revealed the significant changes in sphere formation and cell outgrowth upon altered GULP1 expression, with enlarged images highlighting the reduced lamellipodia (black arrows). (F) IHC images showing the expression patterns in HCC tissues derived from the subcutaneous xenografts injected with *GULP1*-depleted Huh7 cells.

Statistical significance was determined via unpaired *t*-tests for comparisons between two groups. *P < 0.05, **P < 0.01, and ***P < 0.001.

Figure 4

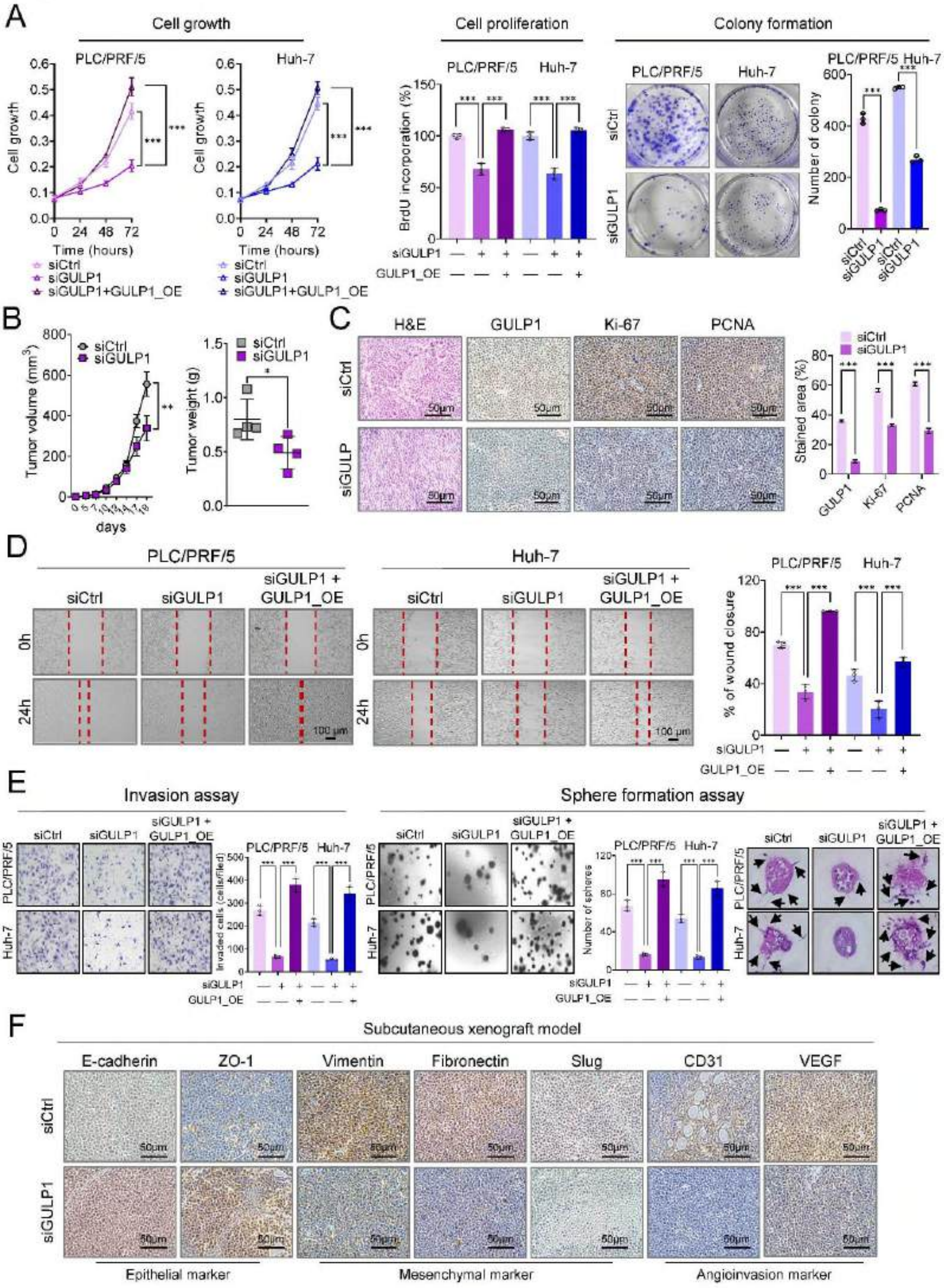

A
Cell growth
PLC/PRF/5
Huh-7
Cell growth
Time (hours)
siCtrl
siGULP1
siGULP1+GULP1_OE
Cell proliferation
BrdU incorporation (%)
siGULP1
GULP1_OE
Colony formation
siCtrl
siGULP1
Number of colony
B
Tumor volume (mm^3)
days
Tumor weight (g)
C
H&E
GULP1
Ki-67
PCNA
siGULP
Stained area (%)
D
0h
24h
100 μm
siGULP1 + GULP1_OE
% of wound closure
E
Invasion assay
Invaded cells (cells/filed)
Sphere formation assay
Number of spheres
F
Subcutaneous xenograft model
E-cadherin
ZO-1
Vimentin
Fibronectin
Slug
CD31
VEGF
50μm
Epithelial marker
Mesenchymal marker
Angioinvasion marker

Figure 5. Validation of GULP1 effects on epithelial–mesenchymal transition (EMT) and HCC recurrence *in vivo.* (A) Schematic representation of the experiment illustrating the orthotopic injection of *GULP1*-suppressed Hepa1-6 cells into the mouse liver. (B) IHC analysis showing the significant increase in GULP1 expression levels, particularly in the recurrence tumors compared to those in the normal liver and primary tumors. Magnifications: 200× and 400×. (C) Left: Representative images of the livers with tumor burden from different groups (siCtrl and siGULP1). Right: Corresponding bar graphs on the right show the calculated nodule volume and weight. (D) Top: IHC analysis of various markers in the NT, primary tumor (PT), and recurrent tumor (RT) tissues. Bottom: Quantification of stained GULP1, Ki-67, E-cadherin, vimentin, and vascular endothelial growth factor (VEGF) expression levels. Proliferation marker: Ki-67. Epithelial marker: E-cadherin. Mesenchymal marker: Vimentin. Angioinvasion marker: VEGF. (E) Lung metastasis model using *ras*-transformed NIH-3T3 cells with siCtrl or siGULP1 treatment. Left: Lung nodules (arrows) are significantly reduced in the siGULP1-treated tissues. Right: H&E-stained images at 40× and 100× magnification show the dispersed HCC cells (arrows) in siCtrl-treated lungs, whereas siGULP1-treated lungs show improved morphology. Scale bars = 50 µm. (F) Analysis of RNA-seq data from the GSE164359 dataset revealed the higher GULP1 levels in RT samples than in the adjacent liver tissue (AL) and PT samples. Statistical significance was determined via unpaired *t*-tests for comparisons between two groups. Data are represented as the mean ± SEM. $^{*}P < 0.05$, $^{**}P < 0.01$, and $^{***}P < 0.001$.

Figure 5

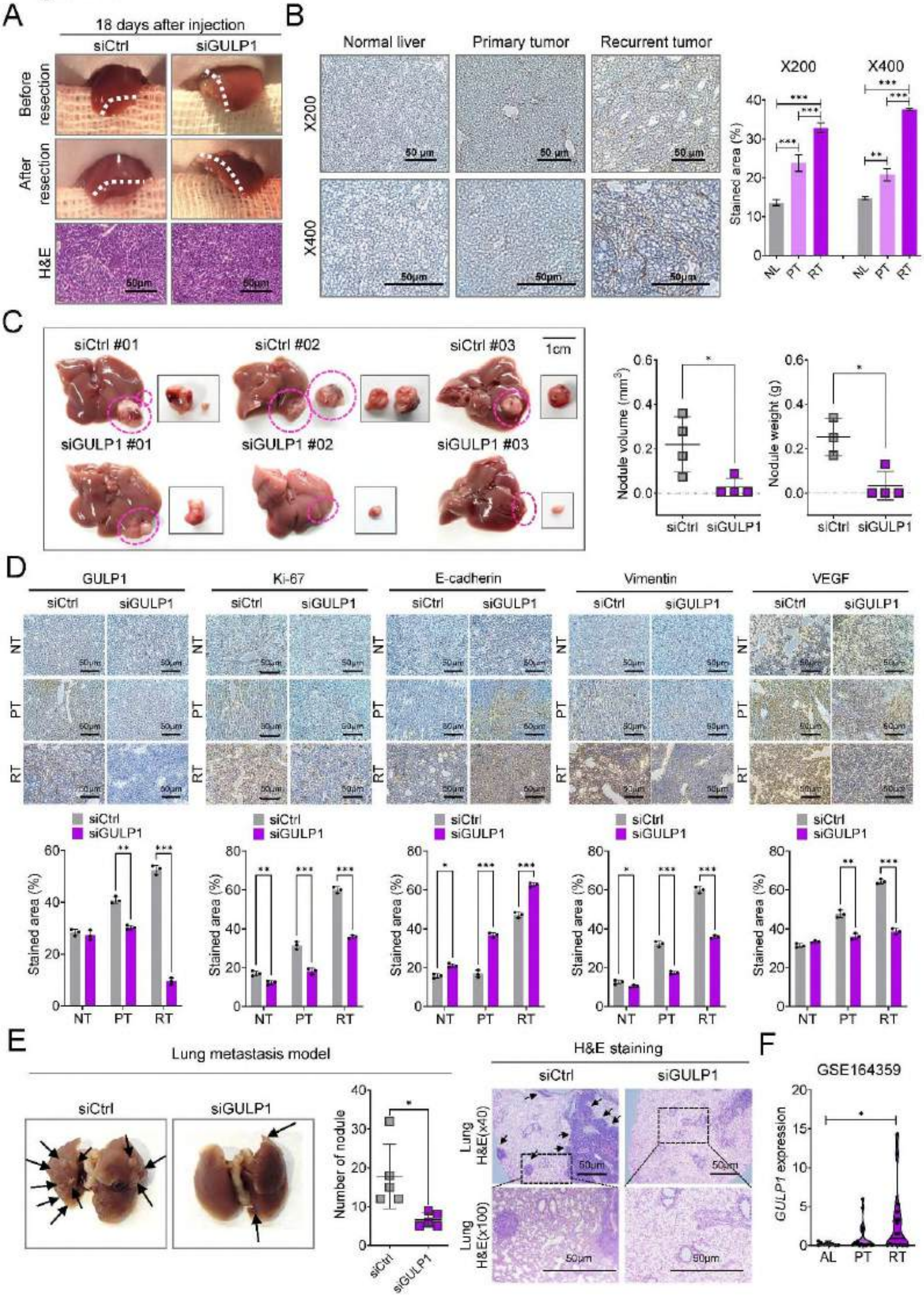

A
18 days after injection
siCtrl
siGULP1
Before resection
After resection
H&E
50μm
B
Normal liver
Primary tumor
Recurrent tumor
X200
X400
50 μm
X200
X400
Stained area (%)
0
10
20
30
40
NL
PT
RT
***
**
C
siCtrl #01
siCtrl #02
siCtrl #03
1cm
siGULP1 #01
siGULP1 #02
siGULP1 #03
Nodule volume (mm³)
Nodule weight (g)
0.0
0.2
0.4
0.6
*
siCtrl
siGULP1
D
GULP1
Ki-67
E-cadherin
Vimentin
VEGF
siCtrl
siGULP1
NT
PT
RT
50μm
Stained area (%)
0
20
40
60
80
*
**
***
E
Lung metastasis model
siCtrl
siGULP1
Number of nodule
0
10
20
30
40
H&E staining
Lung H&E(x40)
Lung H&E(x100)
50μm
F
GSE164359
GULP1 expression
0
5
10
15
20
AL
PT
RT

Figure 6. Mechanistic role of GULP1 in β-catenin signaling pathway regulation in HCC. (A) Western blotting analysis revealed that β-catenin suppression considerably decreased the GULP1 protein levels in the HCC cell lines, Huh-7 and PLC/PRF/5. The cells were treated with the CTNNB1-targeting siRNA (siCTNNB1) or siCtrl. (B) Quantification of GULP1 protein levels via enzyme-linked immunosorbent assay (ELISA) after β-catenin suppression in HCC cell lines. Data indicate the significant reduction in GULP1 expression upon β-catenin knockdown. (C) Immunofluorescence (IF) microscopy images illustrating β-catenin nuclear translocation influenced by GULP1 expression in Huh-7 and PLC/PRF/5 cells. siGULP1 or GULP1 overexpressing vector (GULP1_OE) treatment was performed, with nuclei stained with 4′, 6-diamidino-2-phenylindole (DAPI). (D) Quantitative representation of β-catenin translocation normalized to that of histone H3, corresponding to the microscopy findings. Data indicate that GULP1 modulates β-catenin localization. (E) ELISA of the modulation of ARF6-GTP levels in response to changes in GULP1 levels under conditions where protein synthesis was inhibited by cycloheximide (CHX). Expression of ARF6-GTP was normalized to that of total ARF6. siGULP1 and GULP1_OE treatments showed that GULP1 stabilized ARF6-GTP. (F) Visualization of β-catenin nuclear translocation via IF microscopy (left), with accompanying quantitative analysis (right). Suppression of β-catenin localization upon ARF6 knockdown using the ARF6-targeting siRNA (siARF6) supported the critical role of ARF6 in GULP1-mediated β-catenin translocation. (G) Luciferase activity in Huh7 cells transfected with the wild-type (WT) GULP1 promoter construct, first mutant (MT1) with a deletion from +1304 to +2330, and second mutant (MT2) with a sequence alteration at +1006 (CCCGCATCCT). WT construct showed significantly higher activity than MT1 and MT2. (H) Chromatin immunoprecipitation (ChIP) assays showing the decreased binding affinity of β-catenin to the GULP1 promoter region upon ARF6 knockdown in Huh-7 and PLC/PRF/5 cells.

(I) Modulation of TCF3 binding affinity to the GULP1 promoter region in response to changes in GULP1 expression. ChIP assay was performed to measure the binding activity. (J) Immunoprecipitation revealed β-catenin nuclear translocation and its enhanced interaction with the key adhesion molecules, N-cadherin and E-cadherin, following GULP1 modulation in Huh-7 and PLC/PRF/5 cells. (K) ELISA showing the expression levels of downstream β-catenin targets, such as SRY-box transcription factor 2 (SOX9), c-Myc, and fibronectin, in response to altered GULP1 expression in Huh-7 and PLC/PRF/5 cells. These results indicate that GULP1 influences the expression levels of these targets via its regulatory effects on β-catenin signaling. Unpaired *t*-tests were used to determine the statistical significance. *P < 0.05, **P < 0.01, and ***P < 0.001. Data are represented as the mean ± SEM.

Figure 6

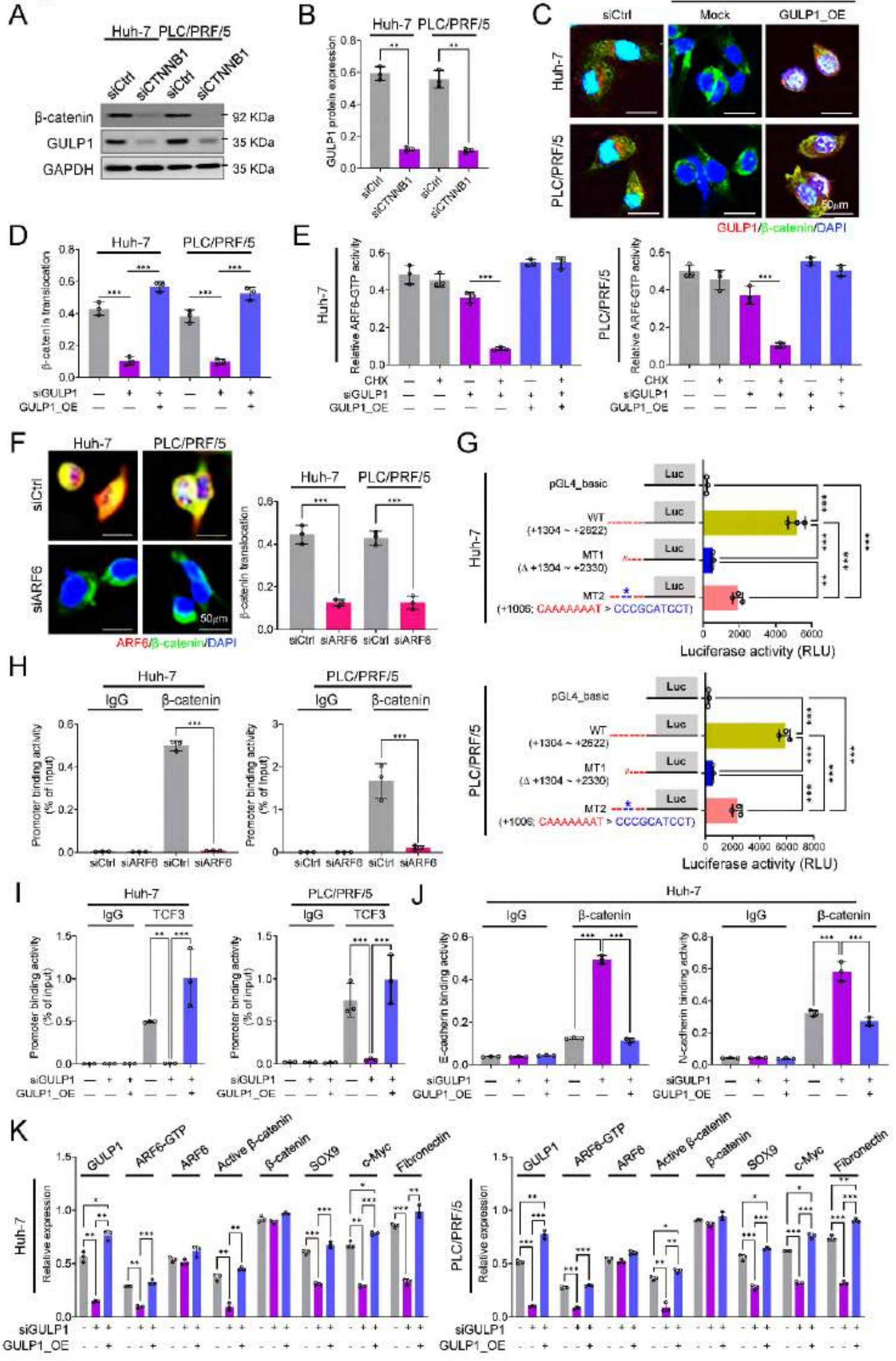
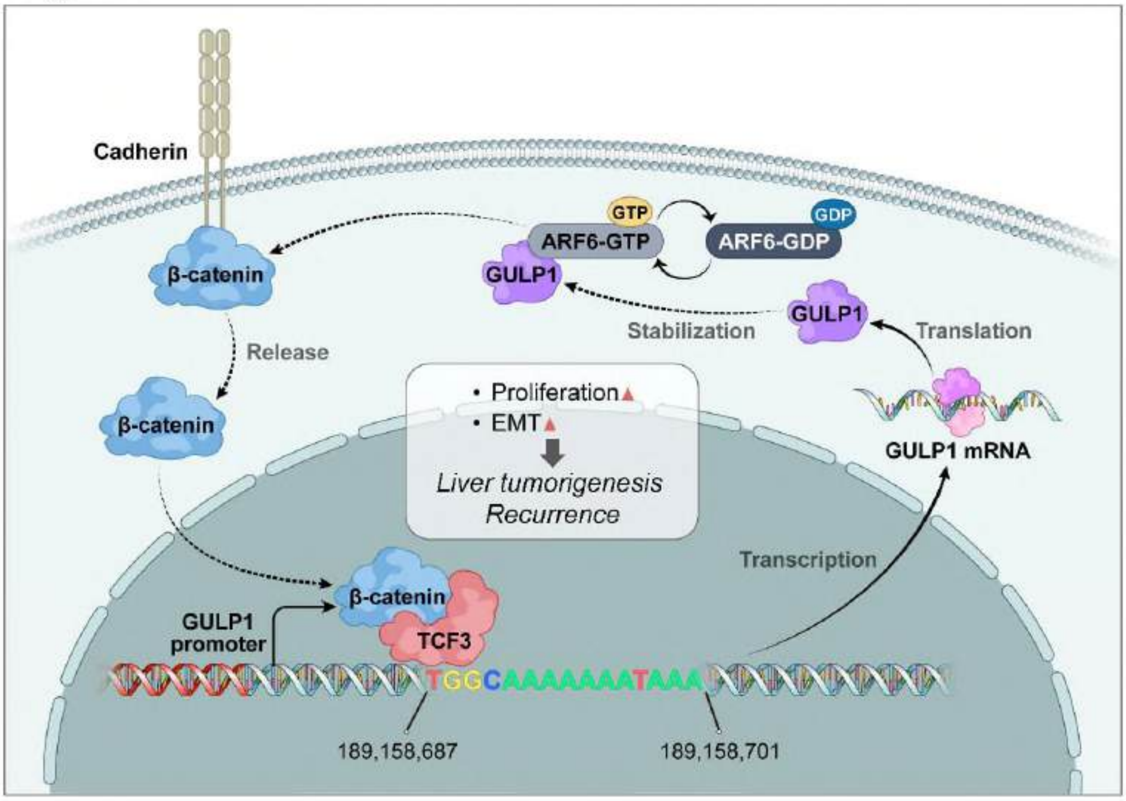

A
Huh-7
PLC/PRF/5
siCtrl
siCTNNB1
siCtrl
siCTNNB1
β-catenin
GULP1
GAPDH
92 KDa
35 KDa
35 KDa
B
Huh-7
PLC/PRF/5
GULP1 protein expression
0.0
0.2
0.4
0.6
0.8
**
**
siCtrl
siCTNNB1
siCtrl
siCTNNB1
C
siGULP1
siCtrl
Mock
GULP1_OE
Huh-7
PLC/PRF/5
50μm
GULP1/β-catenin/DAPI
D
Huh-7
PLC/PRF/5
β-catenin translocation
0.0
0.2
0.4
0.6
0.8
***
siGULP1
GULP1_OE
E
Huh-7
Relative ARF6-GTP activity
PLC/PRF/5
Relative ARF6-GTP activity
0.0
0.2
0.4
0.6
***
CHX
siGULP1
GULP1_OE
F
Huh-7
PLC/PRF/5
siCtrl
siARF6
50μm
ARF6/β-catenin/DAPI
Huh-7
PLC/PRF/5
β-catenin translocation
0.0
0.2
0.4
0.6
***
siCtrl
siARF6
siCtrl
siARF6
G
Huh-7
pGL4_basic
WT
(+1304 ~ +2622)
MT1
(Δ +1304 ~ +2330)
MT2
(+1006; CAAAAAAAT > CCCGCATCCT)
Luc
0
2000
4000
6000
Luciferase activity (RLU)
***
**
PLC/PRF/5
pGL4_basic
WT
(+1304 ~ +2622)
MT1
(Δ +1304 ~ +2330)
MT2
(+1006; CAAAAAAAT > CCCGCATCCT)
Luc
0
2000
4000
6000
8000
Luciferase activity (RLU)
H
Huh-7
IgG
β-catenin
Promoter binding activity (% of input)
0.0
0.2
0.4
0.6
***
siCtrl
siARF6
siCtrl
siARF6
PLC/PRF/5
IgG
β-catenin
Promoter binding activity (% of input)
0
1
2
3
I
Huh-7
IgG
TCF3
Promoter binding activity (% of input)
0.0
0.5
1.0
1.5
**
***
PLC/PRF/5
IgG
TCF3
siGULP1
GULP1_OE
J
Huh-7
IgG
β-catenin
E-cadherin binding activity
N-cadherin binding activity
0.0
0.2
0.4
0.6
0.8
***
siGULP1
GULP1_OE
K
GULP1
ARF6-GTP
ARF6
Active β-catenin
β-catenin
SOX9
c-Myc
Fibronectin
Huh-7
PLC/PRF/5
Relative expression
0.0
0.5
1.0
1.5
*
**
***
siGULP1
GULP1_OE

Figure 7. Schematic diagram of the GULP1–β-catenin co-regulatory mechanism in HCC. GULP1 functions as a crucial modulator in the β-catenin signaling axis in HCC that is critical for the stabilization of ARF6-GTP, which further influences the cellular distribution of β-catenin. This stabilization facilitates β-catenin release and subsequent nuclear translocation, where it binds to TCF3 at the GULP1 promoter region, impacting gene transcription. Downstream effects of this binding include enhanced cell proliferation, EMT, and tumor recurrence, which contribute to liver tumorigenesis.

Figure 7



SUPPLEMENTARY METHODS

Identification and validation of the risk score (RS) model for hepatocellular carcinoma (HCC) recurrence using TCGA LIHC

Regularization was applied to create generalized models by applying a penalty to the estimated coefficients, reducing overfitting, and aiding variable selection. For the Elastic Net methods combining both lasso and ridge, the regularization parameter lambda (λ) was chosen using the "cv.glmnet" function with the training cohort based on the common differentially expressed genes (DEGs) through cross-validation. To ensure consistent results, the "set.seed()" function was used to fix the random number generation with the mixing parameter alpha (α) set to a default value of 0.5, allowing a more flexible alternative to the lasso procedure. A list of potential prognostic genes was obtained based on optimal lambda values. These potential prognostic genes, identified using Elastic Net regression by weighting the L1 and L2 penalty functions, were then entered into a multivariate Cox bidirectional stepwise regression model. Finally, a multivariate Cox regression model with regression coefficients was constructed based on the gene expression and patient survival data.

The RS for each patient was calculated using the following formula:

$$\mathrm{RS} = \sum_{i}^{n} [\mathrm{Coefficiency\ (i)} \times \mathrm{Expression}\ (i)]$$

The "coef" derived from the multivariate Cox regression represents the regression coefficient of the gene, and "expression" indicates the gene expression in terms of log2 (TPM + 1). To calculate the RS for each patient, the gene expression level and the risk coefficient of each gene were used with the "predict" function from the "stats" package in R (version 4.2.3; R Foundation for Statistical Computing, Vienna, Austria). In TCGA cohort, patients were divided into high-risk and low-risk groups based on the optimal cut-off value of the RS derived from the "surv_cutpoint" function in the "survminer" package in R. This cut-off value was also used to stratify patients into high- and low-risk groups.

Differential expression analysis

To compare gene expression between the non-recurrent and recurrent groups in each dataset, we applied different normalization methods based on the data type. The GSE14520 dataset, retrieved from the GEO microarray data, was normalized using the R limma package. For the Next Generation Sequencing (NGS) datasets (GSE114564 and TCGA LIHC), Quantile Normalization was used to ensure consistency in expression values. Differential expression analysis was then performed using the limma package on the normalized data. A statistical cutoff of $P < 0.05$ was set to identify DEGs between the recurrent (R) and non-recurrent (NR) groups. Since expression values between microarray and NGS data are represented differently, we converted the expression values to official gene symbols, enabling the identification of common DEGs across the three datasets. Additionally, to assess the clinical utility of the identified DEGs, decision curve analysis (DCA) was conducted using the rmda package and the stdca.R tool. This allowed for an evaluation of the net benefit of the nomogram across various threshold probabilities.[1, 2]

GSEA (gene set enrichment analysis) for high-risk recurrence group

To investigate the functions and pathways of DEGs between the high-risk and low-risk groups in the TCGA cohort, GSEA was performed using the clusterProfiler package in R (version 4.6.2). Gene set permutations were set at 1,000 for each analysis. Differences with an false discovery value (FDR) value of < 0.05 were considered statistically significant.

Bulk and spatial transcriptomics data analysis

To analyze bulk RNA expression of *GULP1* in normal vs adjacent HCC vs HCC tissue, expression data were downloaded from the GepLiver DB (GepLiver database: http://www.gepliver.org/) on June 2023.[3] Spatial transcriptomics data were downloaded from Mendeley Data (skrx2fz79n).[4] Data analysis was performed on R (version 4.2.3). Heatmap was plotted with ComplexHeatmap package (version 2.14). Boxplots and barplots were plotted with ggplot2 (version3.4.3) and ggparl (version 0.0.1). Spatial transcriptomics data was analyzed using Seurat (version 4.3.0).[5] For bulk-level RNA-seq analysis, *GULP1* expression in respective groups were tested for Normality with Shapiro-Wilk Normality Test. Comparison of non-normally distributed groups were performed with Wilcoxon Rank Sum or Kruskal-Wallis Rank Sum Test and Dunn's test. Statistical significance was set at $P < 0.05$.

Single-cell RNA sequencing (scRNA-seq) data analysis

scRNA-seq data were downloaded from GepLiver DB, comprising a total of 17 scRNA-seq datasets [3]. For downstream analysis, only human samples from HCC, adjacent HCC and normal tissues were selected. The Single-cell Atlas in Liver Cancer (scAtlasLC: https://scatlaslc.ccr.cancer.gov/) data were obtained from the NCBI Gene Expression Omnibus (GSE151530) [6]. The classification of malignant and non-malignant hepatocytes was conducted using inferCNV (https://github.com/broadinstitute/inferCNV). To verify the presence of *GULP1* in HCC, we extracted the count matrix for *GULP1* using the Seurat R package (v.4.0.0) in R (v.4.3.1).[5] Any instances with a count of 0 were categorized as 'GULP1 negative,' while those with a count of 1 or more were considered 'GULP1 positive.' Subsequently, we calculated the proportions of *GULP1* positive cells by dividing the number of *GULP1* positive cells by the total number of cells within each phenotype. To visualize the *GULP1* transcript counts for each cell, we utilized ggplot2 (v.3.4.3) to display *GULP1* expression according to phenotype.[7]

Group comparisons were conducted using a student's t-test with ggpubr (v.0.6.0) (https://rpkgs.datanovia.com/ggpubr/), and statistical significance was determined with a threshold of $P < 0.05$.

Cell culture and transfection

Hep3B, Huh-7, PLC/PRF/5, SNU368, SNU398, SNU423 and Hepa1-6 cell lines were acquired from the Korean Cell Line Bank (KCLB, Seoul, South Korea), whereas *ras*-transformed NIH-3T3 cells were purchased from American Type Culture Collection (ATCC, Bethesda, MD, USA). Cells were cultured in RPMI-1640 or DMEM (GenDEPOT, Barker, TX, USA) with 10% fetal bovine serum (Invitrogen, Waltham, MA, USA) and 100 units/mL penicillin-streptomycin (GenDEPOT). The THLE-2 immortalized normal hepatocytes were obtained from the American Type Culture Collection (ATCC) and cultured in bronchial epithelial cell growth medium (Lonza, Walkersville, MD, USA) supplemented with 10% fetal bovine serum, 5 ng/mL epidermal growth factor (Sigma-Aldrich, St. Louis, MO, USA), 70 ng/mL phosphoethanolamine (Sigma-Aldrich), and antibiotics. Cells were grown in a humidified incubator with 5% CO_2 at 37°C.

For transfection, small interfering RNA molecules (siRNA) and negative control RNA duplexes were synthesized by Bioneer (Daejeon, South Korea) and Genolution (Seoul, South Korea). The RNA duplexes were transfected at a concentration of 100 nM. The sequences for siRNAs were as follows: Control siRNA, 5'-UUCUCCGAACGUGUCACGUUU-3'; *GULP1* siRNA, 5'-CUGCAUAAGGACUACUCUU-3'; CTNNB1 siRNA, 5'-CCACAGCUCCUUCUCUGAGUGGUAA-3'; ARF6 siRNA, 5'-GCGACCACUAUGAUAAUAU-3'. All transfection procedures were conducted with Lipofectamine 2000 transfection reagent (Invitrogen) according to the manufacturer's instructions.

RNA extraction and (quantitative reverse transcription PCR) qRT-PCR

Total RNA was isolated from frozen tissues and HCC cell lines using QIAzol reagent (Qiagen, Hilden, Germany), according to the manufacturer's instructions. Complementary DNA (cDNA) was synthesized from 500 ng of total RNA using 5X PrimeScript™ RT Master Mix (Takara Bio, Shiga, Japan). qRT-PCR was performed using amfiSure qGreen Q-PCR Master Mix (GenDEPOT) and monitored in real-time using a CFX Connect Real-Time PCR Detection System (Bio-Rad Laboratories, Hercules, CA, USA). The cycling conditions were as follows: 95°C for 2 min, 40 cycles of 95°C for 15 s, 58−60°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.

Western blotting

Proteins from whole cell lysates were prepared with radioinnumoprecipitation (RIPA) buffer containing Halt™ Protease Inhibitor Cocktail (Thermo Fisher Scientific, Waltham, MA, USA). Total proteins were separated by SDS-PAGE and transferred to a polyvinylidene difluoride (PVDF) membrane (Merck Millipore, Burlington, MA, USA). The membrane was blocked with blocking buffer (5% skim milk in Tris-buffered saline, 0.1% Tween-20) and incubated with the following antibodies: anti-GULP1 (Novus biologicals, Centennial, CO, USA), anti-ARF6-GTP (NewEastBiosciences, PA, USA), anti-ARF6 (Cell Signaling Technology, Danvers, MA, USA), anti-active β-catenin (Cell Signaling Technology), anti-β-catenin (Cell Signaling Technology), anti-TCF3 (Cell Signaling Technology), anti-ß-actin (Abcam, Cambridge, UK), and anti-GAPDH (Santa Cruz Biotechnology, Santa Cruz, CA, USA). Chemiluminescence signals were detected using Clarity™ Western ECL Substrate (Bio-Rad Laboratories) and visualized using ChemiDoc™ (Bio-Rad Laboratories). The membrane band optical density was quantified using ImageJ software version 1.49 (Laboratory for Optical and Computational Instrumentation, Madison, WI, USA).

Measurement of cell viability and proliferation

Cells were seeded in 24-well plates at a density of 7×10^4 cells/well and transfected with scrambled sequence of single interference control RNA (siCtrl) or *GULP1*-targeting siRNA (siGULP1). After transfection, 50 μL of 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT; 1 mg/mL, Merck KGaA, Darmstadt, Germany) was added to each well of the plate and incubated for 1 h at 37°C in the dark. The supernatant was carefully aspirated, and 500 μL of dimethyl sulfoxide (DMSO; Ducksan, Gwangju, South Korea) was added, following which optical density values at 560 nm were measured using a Promega Glomax microplate reader (Promega, Madison, Wisconsin, USA).

To analyze cell proliferation, 5-bromo-2′-deoxyuridine (BrdU) incorporation was determined by estimating the DNA uptake of BrdU. Cells were plated into 24-well plates at 5×10^4 cells/well for 24 h and measured using a BrdU kit (Roche, Indianapolis, IN, USA) according to the manufacturer’s instructions.

Clonogenic proliferation assay

Cells were transfected with siCtrl or siGULP1 in 60 mm dishes for 48 h, reseeded in 6-well plates (2, 4, or 6×10^3 cells/well), and incubated at 37°C in a CO_2 incubator for 10 days. Cells were washed with phosphate-buffered saline (PBS) and fixed with 1% paraformaldehyde for 30 min at room temperature. The fixed cells were stained with 0.5% crystal violet overnight at room temperature, and colonies were counted using ImageJ software.

Wound healing assay

HCC cells were transfected with siCtrl or siGULP1 in 60 mm dishes for 48 h, reseeded in 6-well plates (1.5 × 10^6 cells/well), and incubated at 37°C in a CO_2 incubator. After overnight incubation, cell monolayers were scraped with a sterile 1000 μL micropipette tip. Photos of the scrape line were taken at 0 and 24 or 48 h. Each experiment was repeated three times. The image analysis was performed via ImageJ software.

Transwell assay

Cell migration was assessed using Transwell chambers (Corning, New York, USA) without Matrigel, while invasion was measured using Transwell chambers coated with 50 μL of Matrigel (Corning); 5 × 10^4 cells were serum-starved for 12 h, resuspended in 100 μL serum-free medium and added to the upper compartment of the chamber, while the bottom chamber was filled with medium supplemented with 10% fetal bovine serum (FBS). After incubation at 37°C in a 5% CO_2 humidified atmosphere for 24 h, the chambers were analyzed for migration and invasion. The experiments were independently repeated thrice.

Three-dimensional spheroid culture and invasion assays

To form spheroids using PLC/PRF/5 and Huh-7 cells, a Matrigel-based 3D sandwich culture system was employed. Initially, sterile 13 mm coverslips treated with HNO_3 were placed in each well of a 24-well plate (Greiner Bio-One GmbH, Kremsmünster, Austria). The plate was chilled and kept on a cooling pad to maintain a cool temperature during the application of Matrigel, ensuring even polymerization of the Matrigel across the plate. For optimization of Matrigel concentration, Matrigel was diluted in ice-cold serum-free RPMI-1640 media (Merck

KGaA), with concentrations of 3.5, 5.0, and 6.5 mg/mL. The Matrigel solutions (50 μL) were drop-casted onto the coverslips, and the plates were incubated at 37°C with 5% CO_2 for 1 h to allow for Matrigel polymerization. PLC/PRF/5 and Huh-7 cells were then trypsinized, and 500 cells were suspended in 30 μL of RPMI-1640 media supplemented with 10% FBS, penicillin/streptomycin, and L-glutamine. The cell suspension was seeded onto each coverslip on top of the Matrigel and incubated for 1 h to allow for cell attachment in an incubator at 37°C with 5% CO_2. After removing the excess media, a top layer of 30 μL of Matrigel (5.0 mg/mL) was applied and allowed to polymerize for an additional hour in the incubator. The wells were subsequently filled with 600 μL of RPMI-1640media, and the plates were returned to the incubator. The morphology of the colonies was evaluated using an EVOS 5000 microscopy system (Invitrogen).

Animal experiments

Five-week-old female BALB/c nude mice, obtained from ORIENT BIO Inc. (Seongnam, Korea), and six-week-old athymic female nude mice, purchased from Koatech (Pyeongtaek, South Korea), were used. The mice were housed in individually ventilated cages in a pathogen-free environment and given one week to acclimate before being used in the study.

The orthotopic model was prepared by transfecting mouse Hepa1-6 cells with either siGULP1 or siCtrl. After transfection, 2×10^6 of these cells were suspended in 10 μL of a Matrigel mixture (Corning) and serum-free DMEM. The cell mixture was then injected orthotopically into the liver of five-week-old female BALB/c nude mice. The procedure involved making a small incision in the abdomen under anesthesia, exposing the liver, and injecting the cells directly into the liver tissue. The incision was then closed using surgical sutures. The mice were monitored post-operation for any signs of distress or complications. The development and growth of the tumor in the liver were regularly tracked using non-invasive imaging techniques. Tumors were allowed to grow until they reached a size where the mice showed signs of discomfort or until it was ethically appropriate to euthanize them for further examination of the tumor.

In the subcutaneous xenograft model, after siGULP1 or siCtrl transfection in Huh-7 cells, 5×10^5 cells in 0.1 mL volume of Matrigel/serum-free DMEM mix were injected subcutaneously into the flanks of BALB/c nude mice. Tumor diameter was measured three times a week for 19 days in the tumor-bearing mice using a digital caliper. The tumor volumes were calculated according to the formula $0.52 \times \text{length} \times \text{width}^2$. The mice were humanely euthanized on day 19 using carbon dioxide. Post-euthanasia confirmation was carried out by checking for the absence of reflexes. Tumors were removed for weighting and histological analysis. The tumor tissues were collected after removal, immediately flash-frozen in liquid nitrogen for RNA and protein extraction and placed in 10% neutral buffered formalin for Hematoxylin and eosin (H&E) staining and Immunohistochemistry (IHC) analysis.

For the lung metastasis assay, a tail vein injection model was used. *ras*-transformed NIH-3T3 cells transfected with siGULP1 or siCtrl (3×10^5 cells) were mixed with 0.2 mL serum-free DMEM and injected into the athymic nude mice through tail veins using a 29-gauge BD Ultra-Fine Needle insulin syringe. The mice were sacrificed 14 days after cell inoculation, and their lungs were collected to assess the number of peritoneal nodules. The summary of the *in vivo* experiments can be found in Supplementary Table 8.

IHC

Tumors were harvested, fixed in 10% neutral buffered formalin, and embedded in paraffin blocks before being cut into 5-μm sections and deparaffinized. One section was stained with

H&E and another was evaluated using IHC.

For IHC, sections were deparaffinized in xylene, hydrated in graded alcohol, and incubated with primary antibodies overnight at 4°C. After being washed three times, the sections were incubated with secondary antibodies for 1 h and then a peroxidase substrate until the desired stain intensity developed (Supplementary Table 9).

Immunofluorescence (IF)

Nuclear translocation of β-catenin was identified via staining for IF using anti-β-catenin antibody (1:100, Cell Signaling Technology). Cells were fixed in 3% paraformaldehyde for 20 min at room temperature and permeabilized in 0.5% Triton X-100 on ice for 7 min. The cells were washed in PBS with 0.5% normal goat serum and incubated with anti-β-catenin antibody. AlexaFluor-488-conjugated rabbit IgG antibody was used as required. Nuclear DNA was stained with 4′, 6-diamidino-2-phenylindole (DAPI), and cells were analyzed using a EVOS M5000 fluorescent microscope (Invitrogen) to visualize the endogenous level of proteins under study.

Dual Luciferase Reporter Assay

To investigate the role of TCF3 in the regulation of *GULP1* gene transcription, specific promoter regions of the *GULP1* gene were cloned into the pGL4.10[luc2] luciferase reporter vector (Promega). The *GULP1* promoter fragment spanning nucleotides +1304 to +2622 relative to the transcription start site was amplified by PCR and inserted into the pGL4.10 vector using NheI and HindIII restriction sites. The deletion mutant, which lacked the region from +1304 to +2330, was created by PCR amplification of the remaining promoter segments and subsequent ligation into the vector. A point mutation was introduced at position +1006, altering the sequence to CCCGCATCCT through site-directed mutagenesis utilizing mutagenic primers. All constructs were verified via Sanger sequencing to confirm the presence of desired mutations and the integrity of the sequences. Huh7 and PLC/PRF/5 cells, cultured in RPMI-1640 supplemented with 10% FBS and 1% penicillin-streptomycin, were seeded in 24-well plates at a density of 1×10^5 cells per well. Transient transfections were performed using Lipofectamine 2000 (Invitrogen), where each well received 500 ng of the respective pGL4.10 construct (wild-type, deletion, or point mutation) and 50 ng of the pGL4.74[hRluc/TK] vector (Renilla luciferase control vector) to normalize for transfection efficiency.

48 hours post-transfection, cells were lysed with Passive Lysis Buffer (Promega), and luciferase activity was assessed using the Dual-Luciferase Reporter Assay System (Promega) following the manufacturer's instructions. Firefly luciferase activity from the pGL4.10 constructs was normalized to Renilla luciferase activity from the pGL4.74 control vector to control for variations in transfection efficiency. Relative luciferase activity, calculated as the ratio of Firefly to Renilla luciferase signals, was used to determine the impact of TCF3 on *GULP1* promoter activity. Comparisons were made between the wild-type promoter and the mutated constructs (both deletion and point mutation) to evaluate TCF3's regulatory effect. Each experiment was conducted in triplicate, and data are presented as mean ± standard deviation. The sequences for primers were as follows: Wild type (+1304 ~ +2622) forward (F), 5’-GCGGCTAGCGTTCTCTCTATTCTGAGGCTCCTG-3’; reverse (R), 5’-GCGAAGCTTTGAGAATACGTAATTGAATTATTCAATTTA-3’; Mutant 1 (Del +1304 ~ +2330) F, 5’-GCGGCTAGCGAGTGCAGTTGGATATGACCGAGG-3’; R, 5’-GCGAAGCTTTGAGAATACGTAATTGAATTATTCAATTTA-3’; Mutant 2 (Site direct; +1006 ~ +1014; CAAAAAAAT > CCCGCATCCT) F, 5’-GTGGCTAAGATCAGTCTTCTTGGCCCGCATCCTAAAGAAAAA-3’; R, 5’-

TTTTTCTTTTAGGATGCGGGCCAAGAAGACTGATCTTAGCCAC-3’.

Lumit immunoassay

To investigate the protein-protein interaction between GULP1 and ARF6, we employed the Lumit Immunoassay Kit from Promega. Following the manufacturer's instructions, we labeled GULP1 antibody with Sm-BiT and ARF6 antibody with Lg-BiT using the HaloTag technology provided in the kit. The labeling process involved a two-step reaction: first, the amine-reactive HaloTag Succinimidyl Ester (O4) Ligand was conjugated to primary amines on lysine residues of the antibodies, and then the labeled antibodies were incubated with HaloTag-Sm-BiT or HaloTag-Lg-BiT fusion proteins to create covalent conjugates. After labeling, we performed the homogeneous bioluminescent immunoassay in a 96-well plate. The labeled antibodies were mixed with cell lysates containing the target proteins, and upon binding of GULP1 and ARF6, the Sm-BiT and Lg-BiT subunits came into close proximity, reconstituting the NanoBiT enzyme (Supplementary Fig. S7A). We then added the Lumit Detection Reagent, which contains the substrate furimazine, and measured the resulting luminescence signal using a plate-reading luminometer. The intensity of the luminescent signal was proportional to the extent of GULP1-ARF6 interaction, allowing us to quantitatively assess their binding capacity.

Co-immunoprecipitation (co-IP)

Transfected cells were washed with PBS and lysed at 4°C using lysis buffer composed of PBS (pH 7.2) containing 1.0% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 10 mmol/L NaF, 1.0 mmol/L $NaVO_4$, and a 1.0% protease inhibitor cocktail (Sigma-Aldrich), following a previously described protocol.[8] Equal protein aliquots (1.0 mg) were subjected to immunoprecipitation with 2.0 μg of specific antibodies against β-catenin (Cell Signaling

Technology) or GULP1, using protein A/G-agarose (Santa Cruz Biotechnology). The immunoprecipitated proteins were separated on 10% SDS-polyacrylamide gels and transferred to polyvinylidene fluoride membranes (Bio-Rad Laboratories). After blocking with PBS containing 0.1% Tween 20 (PBS-T) and 5% non-fat dry milk (Sigma-Aldrich) for 1 h, the membranes were incubated with antibodies against anti-β-catenin (Cell Signaling Technology), anti-E-cadherin (Cell Signaling Technology), anti-N-cadherin (Cell Signaling Technology), anti-GULP1 (Santa Cruz Biotechnology), and anti-ARF6 (Cell Signaling Technology) at a dilution of 1:1,000. Following washing with PBS-T, the membranes were incubated with horseradish peroxidase-conjugated anti-mouse (Sigma-Aldrich) or anti-rabbit immunoglobulin G (IgG) antibody (Sigma-Aldrich) at a dilution of 1:5,000 for 1 h at room temperature. Protein bands were visualized using a luminol-based enhanced chemiluminescence plus Western blotting detection system (Amersham Biosciences, Buckinghamshire, UK). The identification of immunoreactive bands was achieved by co-migration with pre-stained protein size markers (Fermentas, Glen Burnie, MD, USA).

Chromatin immunoprecipitation (ChIP)

ChIP assays were performed using the Thermo Scientific Pierce Agarose ChIP kit (Thermo Fisher Scientific). Briefly, cells were cross-linked with 1% formaldehyde and collected into lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl, pH 8.0, 1× protease inhibitor mixture). Cell lysates were digested with micrococcal nuclease, followed by immunoprecipitating with anti-TCF3 (Cell Signaling Technology) or anti-β-catenin (Cell Signaling Technology) antibody. Immunoprecipitation with a normal rabbit IgG (Thermo Fisher Scientific) was used as a negative control. After incubation with the protein A/G Plus agarose resin, immunoprecipitants were washed and then heated at 65°C for 1.5 h to reverse the formaldehyde cross-linking. DNA fragments were purified with the DNA clean-up column and reagents included in the Pierce agarose ChIP kit. The expression of gene promoter and enhancer were quantified by SYBR green-based real-time quantitative PCR. The primer sequence for *GULP1* promoter follows; Forward 5'-ACACCACAGTACCAGATTCAGT-3' and Reverse 5'-GTCAGTTCAGGGAAGGGGAA-3'

Enzyme-linked immunosorbent assay (ELISA)

Immediately after sample collection, the serum extracted from the blood of both healthy individuals and liver disease patients was stored at -80°C. For the assay, the samples were retrieved from storage and carefully thawed at 4°C. Following the thawing process, we measured the concentration of the GULP1 protein within the serum of each cohort. The assay was conducted using an ELISA kit (HUFI04000; Assay Genie, Ireland) strictly in accordance with the manufacturer's instructions. Serum samples were diluted 1:10 before measurement to ensure optimal detection within the assay's range. The assay's performance was validated with a control curve yielding an R-squared value of 0.9967, indicating excellent linearity. The mean optical density (OD) of the blank was 0.141 with a standard deviation of 0.005, demonstrating low background and high precision. The Hillslope of 0.8283 suggested appropriate assay kinetics. Notably, the GULP1 ELISA kit boasted a remarkably low Limit of Detection (LOD) of 0.02 pg/mL, ensuring high sensitivity for accurate protein quantification.

Protein levels were quantified by indirect ELISA and the plates coated with cell lysates diluted in carbonate buffer (34 mM Na_2CO_3, 100 mM $NaHCO_3$, pH 9.5) were incubated at 4 °C overnight. Washed with PBS-T thrice and blocked with 1% Bovine Serum Albumin (BSA, Sigma-aldrich) 1 h at RT. Washed and incubated at 37 °C for 1 h with 100 µL of anti-GULP1 (Novus biologicals), anti-ARF6-GTP (NewEastBiosciences), anti-ARF6, anti-active β-catenin, anti-β-catenin, anti-SOX9, anti-c-Myc, anti-Fibronectin (Cell Signaling Technology), and anti-GAPDH (Santa Cruz Biotechnology) antibody, respectively. Washed and incubated at 37 °C for 1 h with 100 μL of rabbit anti-mice and goat anti-rabbit (1:10000) IgG/HRP conjugate, respectively. Washed and developed with 3,3',5,5'-Tetramethylbenzidine (TMB) solution for 20 min. 50 μL of stop solution (1 N H_2SO_4) was added per well and absorbance was measured at 450 nm on ELISA reader. Protein levels were normalized to the GAPDH control for each plate and samples.

Statistical Analysis

The statistical significance of the differences between two groups was assessed via the paired Student's t-test or unpaired Welch's t-test using GraphPad Prism (version 9.0; GraphPad Software, San Diego, CA, USA). For multiple comparisons among three groups, one-way analysis of variance (ANOVA) with Tukey's post-hoc analysis was used. Kaplan–Meier survival curves were constructed to assess the significance of the prognostic power between the two patient groups. Significant differences between survival curves were determined using the log-rank test. Receiver operating characteristics (ROC) and Cox proportional hazard regression analyses were performed using IBM SPSS software (IBM SPSS Statistics for Windows, version 22.0, released 2013; Armonk, NY, USA). The cut-off values for the diagnosis of liver cancer were established at 20 ng/mL for serum α-fetoprotein (AFP) and 340 pg/mL for serum GULP1. All experiments were performed at least thrice. Statistical significance was set at $P < 0.05$.

Ethics approval and consent to participate

All experiments were performed in accordance with the Declaration of Helsinki, and the study was approved by the Institutional Review Board of Ajou University Hospital (AJIRB-BMR-KSP-16-365, AJIRB-BMR-SMP-17-189, AJOUIRB-KSP-2019-417, AJOUIRB-EX-2022-389 and AJOUIRB-EX-2024-332). Anonymous serum samples and clinical data were provided by the Ajou Human Bio-Resource Bank; the requirement for informed consent was waived.

All animals were cared for in accordance with the Guide for the Care and Use of Laboratory Animals and experiments were approved by the Ethics Committee for Laboratory Animal Research Center of Ajou University Medical Center (IACUC_2022-0049).

REFERENCES

1. Brown M. rmda: Risk model decision analysis (R package version 1) 2018.

2. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26:565-74.

3. Li Z, Zhang H, Li Q, Feng W, Jia X, Zhou R, et al. GepLiver: an integrative liver expression atlas spanning developmental stages and liver disease phases. Sci Data. 2023;10:376.

4. Qi J, Sun H, Zhang Y, Wang Z, Xun Z, Li Z, et al. Single-cell and spatial analysis reveal interaction of FAP(+) fibroblasts and SPP1(+) macrophages in colorectal cancer. Nat Commun. 2022;13:1742.

5. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. Nat Biotechnol. 2015;33:495-502.

6. Ma L, Wang L, Khatib SA, Chang CW, Heinrich S, Dominguez DA, et al. Single-cell atlas of tumor cell evolution in response to therapy in hepatocellular carcinoma and intrahepatic cholangiocarcinoma. J Hepatol. 2021;75:1397-408.

7. Wickham H. ggplot2: Elegant Graphics for Data Analysis (3e): New York: Springer-Verlag; 2023.

8. Yoon JH, Eun JW, Choi WS, Kim O, Nam SW, Lee JY, Park WS. NKX6.3 Is a Transcription Factor for Wnt/beta-catenin and Rho-GTPase Signaling-Related Genes to Suppress Gastric Cancer Progression. EBioMedicine. 2016;9:97-109.

Supplementary Table 1. Clinicopathological characteristics of 81 patients who underwent hepatectomy.

Variables	Total cohort (n=81)
Age (years), mean ± SD	56.0 ± 10.3
Male sex, n (%)	64 (79)
AST, IU/mL	50.6 ± 76.1
ALT, IU/mL	44.9 ± 53.6
Platelet, x10^9/L	176.0 ± 64.1
AFP (ng/mL), mean ± SD	2523.8 ± 8092.2
Albumin (g/L), mean ± SD	4.5 ± 0.7
Bilirubin (mg/dL), mean ± SD	0.8 ± 1.0
Sodium (mmol/L), mean ± SD	106.5 ± 59.4
Etiology, n (%)	
HBV	76 (93.8)
HCV	5 (6.2)
Alcohol	1 (1.2)
Others	0 (0)
BCLC stage, n (%)	
0	34 (42.0)
A	27 (33.3)
B	2 (2.5)
C	18 (22.2)
D	0 (0)
Modified UICC stage, n (%)	
I	22 (27.2)
II	31 (38.3)
III	22 (27.2)
IV	6 (7.4)
Recurrence, n (%)	24 (29.6)
Vascular invasion, n (%)	40 (49.4)

AST, aspartate aminotransferase; ALT, alanine aminotransferase; AFP, α-fetoprotein; HBV, hepatitis B virus; HCV, hepatitis C virus; BCLC, Barcelona Clinic Liver Cancer staging system; UICC, The Union for International Cancer Control.

Supplementary Table 2. Clinicopathological characteristics of 256 patients for evaluation of GULP1 as a blood marker.

Variables	Total cohort (n=256)			
	NL (n=30)	CH (n=34)	LC (n=33)	HCC (n=159)
Age (years), mean ± SD	34.6 ± 7.5	46.5 ± 10.7	53.1 ± 10.5	55.8 ± 9.2
Male sex, n (%)	4 (13.3)	20 (58.8)	20 (60.6)	119 (82.1)
AST, IU/mL	16.6 ± 3.7	54.3 ± 49.1	80.7 ± 99.1	55.3 ± 63.9
ALT, IU/mL	13.7 ± 7.8	65.7 ± 72.9	77.0 ± 99.4	42.2 ± 38.7
Platelet, x10^9/L	294.7 ± 35.8	175.2 ± 52.1	129.4 ± 77.8	163.3 ± 73.8
AFP (ng/mL), mean ± SD	1.7 ± 0.7	26.4 ± 32.4	65.8 ± 133.0	2433.5 ± 10224.8
Albumin (g/L), mean ± SD		4.4 ± 0.9	4.0 ± 0.5	4.3 ± 0.5
Bilirubin (mg/dL), mean ± SD		0.8 ± 0.4	1.1 ± 1.0	0.9 ± 2.5
INR, mean ± SD		0.4 ± 0.6	1.0 ± 0.5	1.1 ± 0.1
Etiology, n (%)				
HBV				132 (91.0)
HCV				11 (7.6)
Alcohol				2 (1.4)
Others				0 (0)
Modified UICC stage, n (%)				
I				43 (29.7)
II				74 (51.0)
III				18 (12.4)
IV				10 (12.4)
Vascular invasion, n (%)				36 (24.8)
Recurrence, n (%)				66 (45.5)

AST, aspartate aminotransferase; ALT, alanine aminotransferase; AFP, α-fetoprotein; HBV, hepatitis B virus; HCV, hepatitis C virus; INR, international normalized ratio; UICC, The Union for International Cancer Control; NL, normal liver; CH, chronic hepatitis; LC, liver cirrhosis; HCC, hepatocellular carcinoma.

Supplementary Table 3. Clinicopathological characteristics of 302 patients for the etiological evaluation of GULP1.

Variables	Total cohort (n=302)				
	NL (n=32)	LC (n=120)	HCC (n=120)	CCC (n=8)	cHCC-CCC (n=22)
Age (years), mean ± SD	33.5 ± 8.3	60.1 ± 10.5	66 ± 10.7	65.4 ± 10.6	59.8 ± 12.9
Male sex, n (%)	4 (13)	70 (58)	92 (77)	6 (75)	19 (86)
Etiology, n (%)					
HBV		30 (25)	30 (25)		
HCV		30 (25)	30 (25)		
Alcohol		30 (25)	30 (25)		
MASH		30 (25)			
MASH / unkown			30 (25)		
AST, U/L	16.7 ± 3.9	50.7 ± 57.2	49.9 ± 40.2	21.8 ± 3.2	51 ± 58.1
ALT, U/L	13.4 ± 6.4	40.1 ± 59	42.4 ± 63.6	19.4 ± 6.6	61.8 ± 105
Platelet, x10^9/L	247 ± 43.2	144.8 ± 65.1	170 ± 84.6	230.6 ± 97	190.4 ± 68.2
Albumin (g/dL), mean ± SD		4.6 ± 3.7	4.2 ± 0.6	4.6 ± 0.3	4.5 ± 0.2
Total bilirubin (mg/dL), mean ± SD		1.1 ± 1.9	1 ± 1.6	0.7 ± 0.4	0.9 ± 1.7
AFP (ng/mL), mean ± SD	2.1 ± 1.4	24 ± 75.9	2366.5 ± 9463.8	2.3 ± 0.6	2840.4 ± 12583.9

AST, aspartate aminotransferase; ALT, alanine aminotransferase AFP, α-fetoprotein; HBV, hepatitis B virus; HCV, hepatitis C virus; MASH, metabolic dysfunction-associated steatohepatitis; NL, normal liver; LC, liver cirrhosis; HCC, hepatocellular carcinoma; CCC, cholangiocarcinoma; cHCC-CCC, combined hepatocellular carcinoma and cholangiocarcinoma.

Supplementary Table 4. Baseline characteristics of the patients in three cohort.

Variables	Number (%) of Participants		
	GSE14520 (n=172)	GSE114564 (n=55)	TCGA LIHC (n=212)
Age, mean (SD), year	50.88(10.75)	54.98(9.02)	59.13(13.22)
Sex, male	148(86.04%)	47(85.45%)	141(66.50%)
Platelet count, mean (SD), 10^9/L		104.29(72.67)	216.22(75.10)
Total bilirubin, mean (SD), mg/dL		2.84(3.06)	0.94(1.66)
Creatinine, mean (SD), mg/dL		1.22(1.29)	2.60(10.83)
Serum AFP, mean (SD), ng/mL		3874(23780.15)	19922(160629.4)
Serum AFP >300 ng/ml	81(47.09%)		
Vascular invasion, micro-macro		19(34.54%)	66(31.13%)
Histologic grade, G3-G4		9(16.36%)	83(39.15%)
Child Pugh, B-C		23(41.81%)	11(5.18 %)
AJCC, III-IV		8(14.54%)	59(27.83 %)
Recurrence status (NR/R)	89/83	31/24	48/164

SD, standard deviation; AFP, α-fetoprotein; AJCC, The American Joint Committee on Cancer;

NR, non-recurrence; R, recurrence.

Supplementary Table 5. The lambda values for the 50 selected genes in Elastic Net Cox Regression analysis.

* Supplementary Table 5 was submitted as a separate Excel file.

Supplementary Table 6. The regression coefficients of 15-gene signatures based on the optimal lambda value.

Genes	Coefficiency
GULP1	0.03846
LCAT	-0.0279
PPAT	0.09907
LPXN	-0.0423
NOP56	0.17944
CD4	-0.01382
ZC2HC1A	0.06148
PPIA	0.27442
CST7	-0.01075
PRKCQ	-0.0344
PHF20	0.0052
RAB23	0.00434
PCDHB6	0.01753
CXCR6	-0.07688
SLC4A10	-0.0375

Supplementary Table 7. Cox regression analyses of the variables associated with disease free survival in TCGA cohort.

Variables	Univariate Coefficiency	Univariate HR (95% CI)	Univariate *P*-value	Multivariate Coefficiency	Multivariate HR (95% CI)	Multivariate *P*-value
Age	-0.0005	0.99 (0.988-1.011)	0.922			
Sex (male)	0.0586	1.06 (0.764-1.471)	0.726			
Platelet count (10^9/L)	5.15e-07	1.00 (1.00-1.00)	0.672			
Total bilirubin (mg/dL)	0.0356	1.04 (0.954-1.125)	0.398			
Creatinine (mg/dL)	0.0004	1.00 (0.986 - 1.015)	0.954			
Serum AFP (ng/mL)	-1.16e-06	1.00 (1.00-1.00)	0.259			
Vascular invasion	0.4824	1.62 (1.168-2.246)	0.004	0.2548	1.29 (0.835-1.994)	0.25126
Histologic grade (3-4 vs. 1-2)	0.0513	1.05 (0.765-1.448)	0.753			
Child Pugh (B, C vs. A)	0.6144	1.85 (0.978 - 3.494)	0.059			
AJCC (III-IV vs. I-II)	0.7935	2.21 (1.535-3.185)	2.03e-05	0.2216	1.25 (0.825-1.888)	0.294
15-gene RS	**1.0000**	**2.72 (2.146-3.443)**	**2e-16**	**0.7341**	**2.08 (1.311-3.313)**	**0.002**
Du et al. RS (7-genes)	0.7111	2.04 (1.704-2.433)	5.12e-15	0.3797	1.46 (1.147-1.862)	0.002

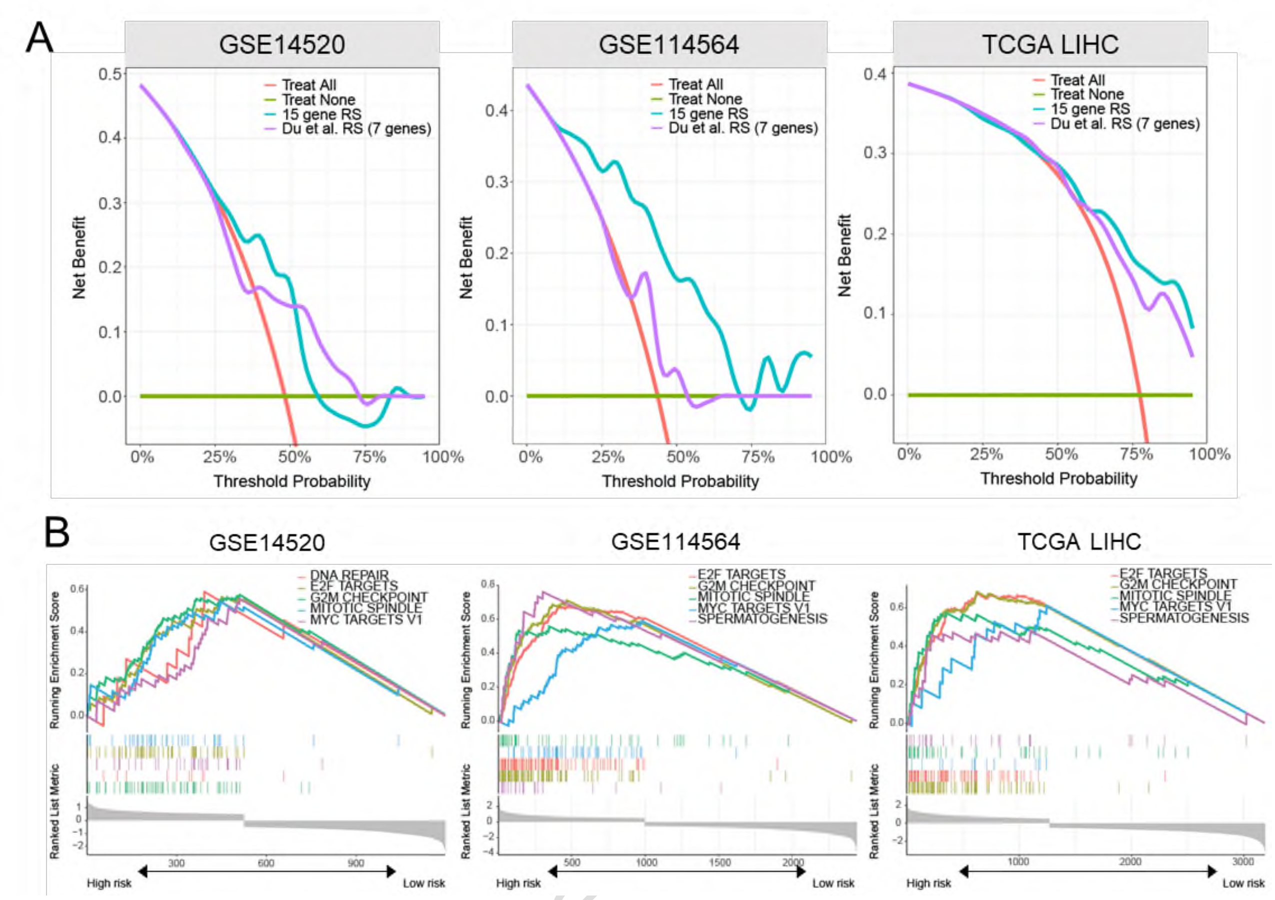
HR, hazard ratio; AFP, α-fetoprotein; AJCC, The American Joint Committee on Cancer; RS, risk score.

Supplementary Table 8. The experimental designs for the *in vivo* studies.

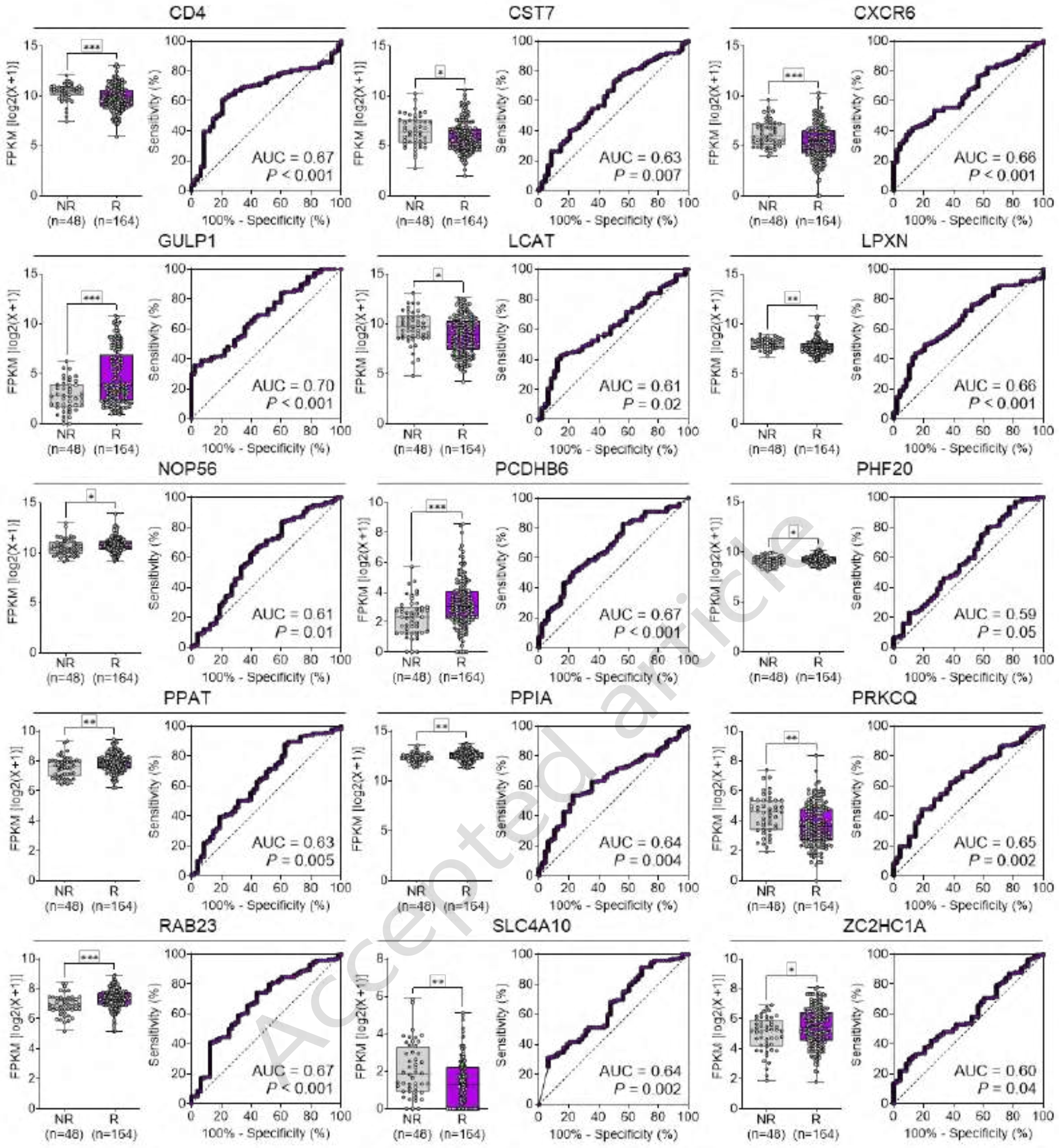
Model (mouse count per group)	Cell number (Cell line)	Injection volume (Matrigel mixing ratio)	Detail (experimental parameters)
Subcutaneous xenograft model (n = 4)	5×10^5 (Huh-7)	100 µL (Matrigel at 20% v/v)	Sacrificed 19 day after cell injection. (Tumor volume, body weight, tumor weight, tumor IHC)
Tail vein model (n = 5)	3×10^5 (*ras*-NIH-3T3)	200 µL	Sacrificed 14 day after cell injection. (Number of nodules, body weight, lung H&E)
Orthotopic xenograft model (n = 4)	2×10^6 (Hepa1-6)	10 µL (Matrigel at 50% v/v)	Tumor resection performed 18 days after cell injection and sacrificed 14 days after tumor resection. (Nodule volume, nodule weight, tumor IHC)

Supplementary Table 9. The list of primary antibodies for IHC staining.

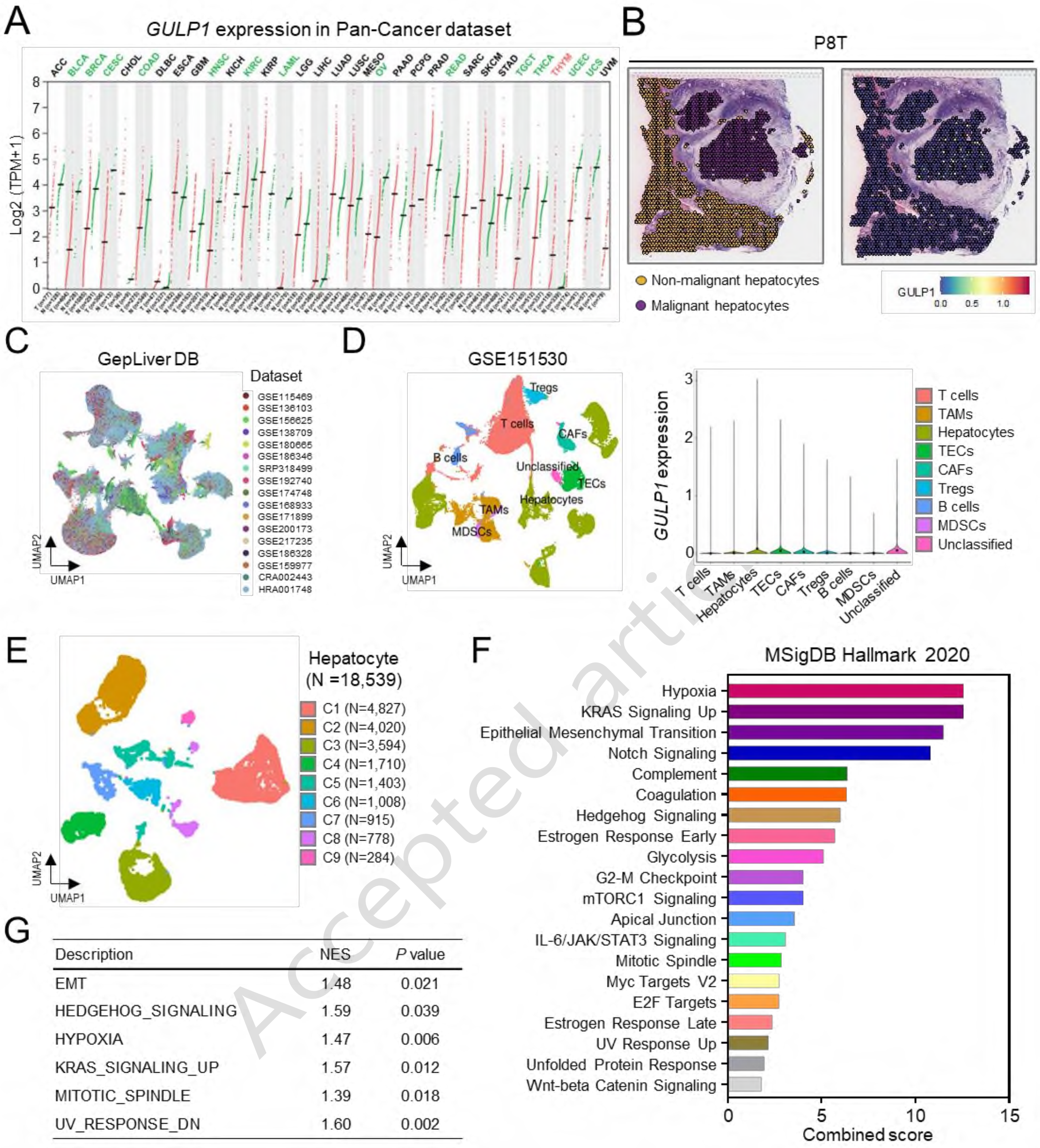
Antibodies	Species	Dilution	Manufacturer
GULP1	Rabbit	1:200	Novus biologicals
Ki-67	Rabbit	1:500	Abcam
PCNA	Mouse	1:10000	Abcam
β-catenin	Rabbit	1:200	Cell Signaling Technology
SLUG	Rabbit	1:100	Cell signaling Technology
E-cadherin	Mouse	1:100	BD bioscience
Vimentin	Rabbit	1:500	GeneTex
ZO-1	Mouse	1:200	Thermo Fisher Scientific
Fibronectin	Mouse	1:100	Santa Cruz Biotechnology
CD31	Rabbit	1:100	Abcam
VEGF	Mouse	1:100	Santa Cruz Biotechnology



Supplementary Figure 1. Evaluating predictive models for hepatocellular carcinoma (HCC) recurrence and their clinical impact. (A) Decision curve analysis (DCA) comparing the net benefits of the 15-gene risk score (RS) model and the existing 7-gene model by Du et al. across different threshold probabilities in three datasets, indicating that the 15-gene RS model outperforms the 7-gene model. (B) Biological and functional pathway analysis between the low- and high-risk groups performed using gene set enrichment analysis (GSEA).

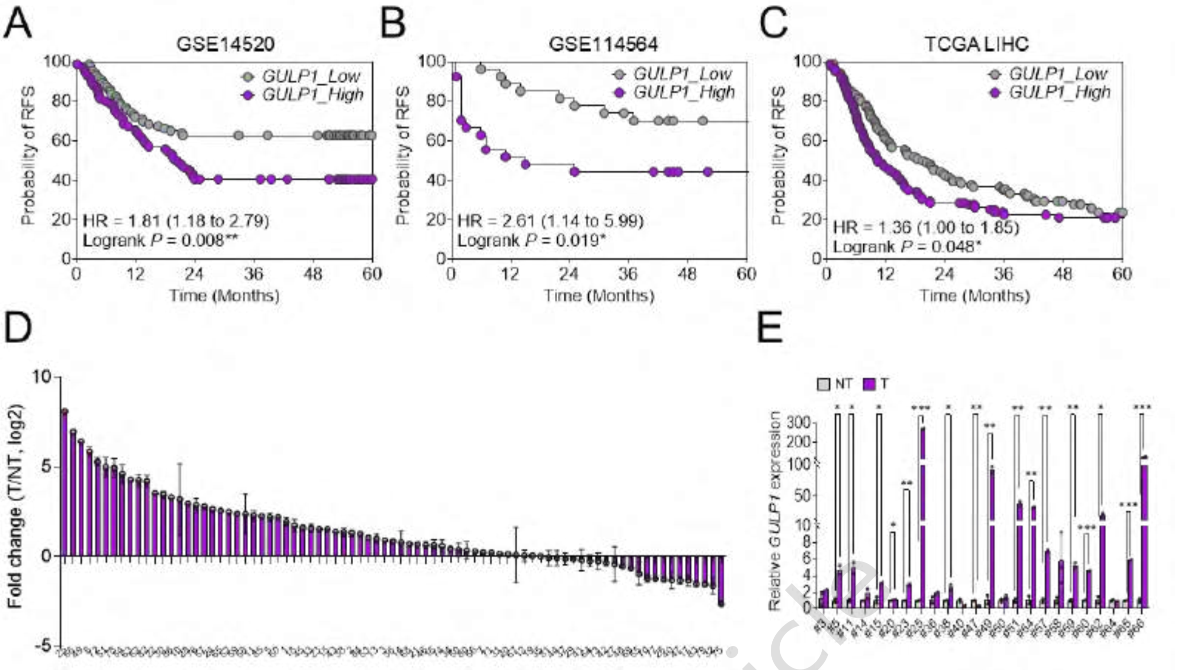


Supplementary Figure 2. Expression levels and diagnostic performance of *GULP1* and other genes. Box plots and ROC curves for the 15 genes identified in the recurrence risk score (RS) model, comparing expression levels between non-recurrence (NR) and recurrence (R) groups. Among these genes, *GULP1* showed the highest expression in the R group and demonstrates superior sensitivity and specificity for recurrence diagnosis, as indicated by the ROC analysis.

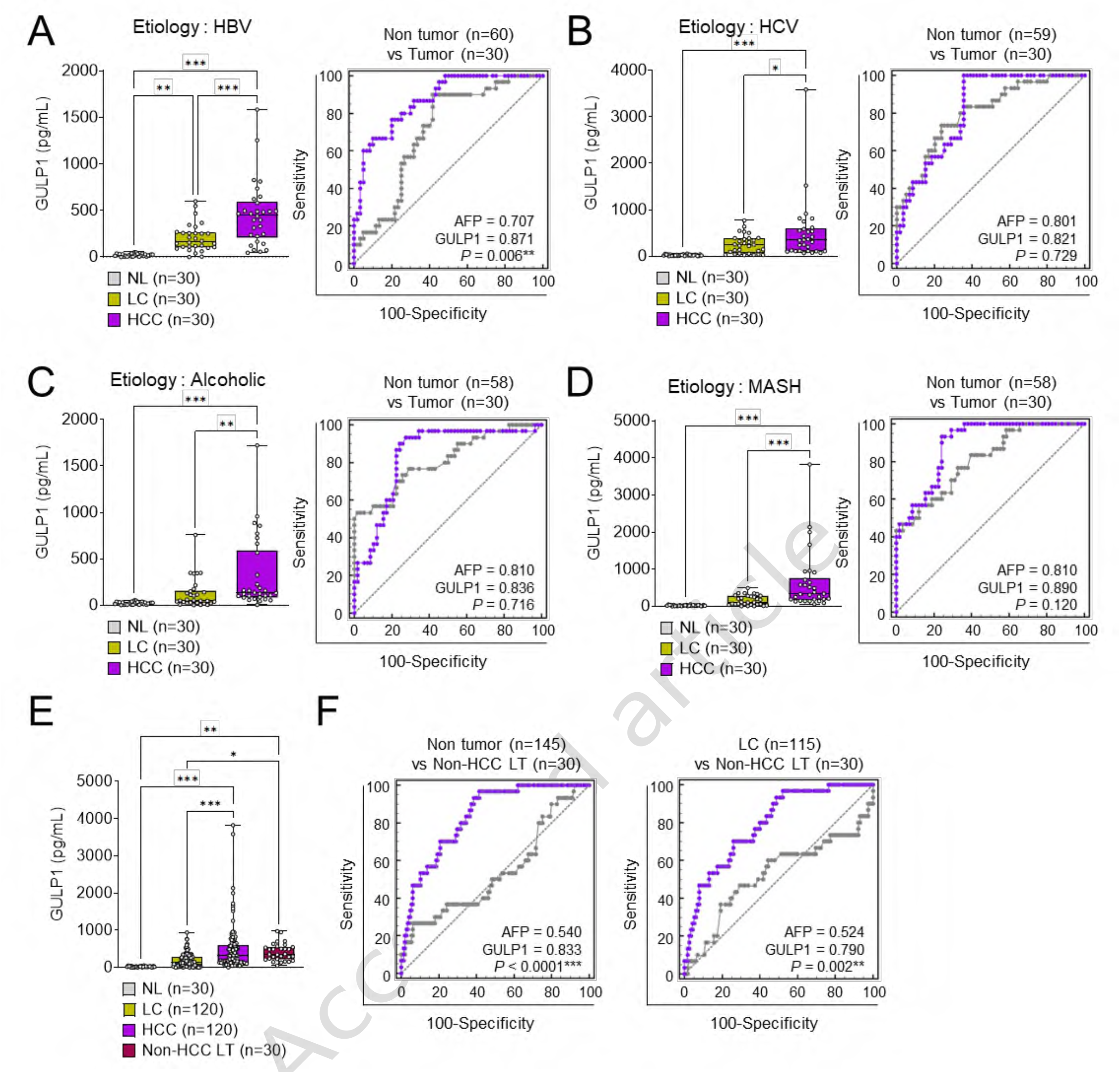


Description	NES	*P* value
EMT	1.48	0.021
HEDGEHOG_SIGNALING	1.59	0.039
HYPOXIA	1.47	0.006
KRAS_SIGNALING_UP	1.57	0.012
MITOTIC_SPINDLE	1.39	0.018
UV_RESPONSE_DN	1.60	0.002

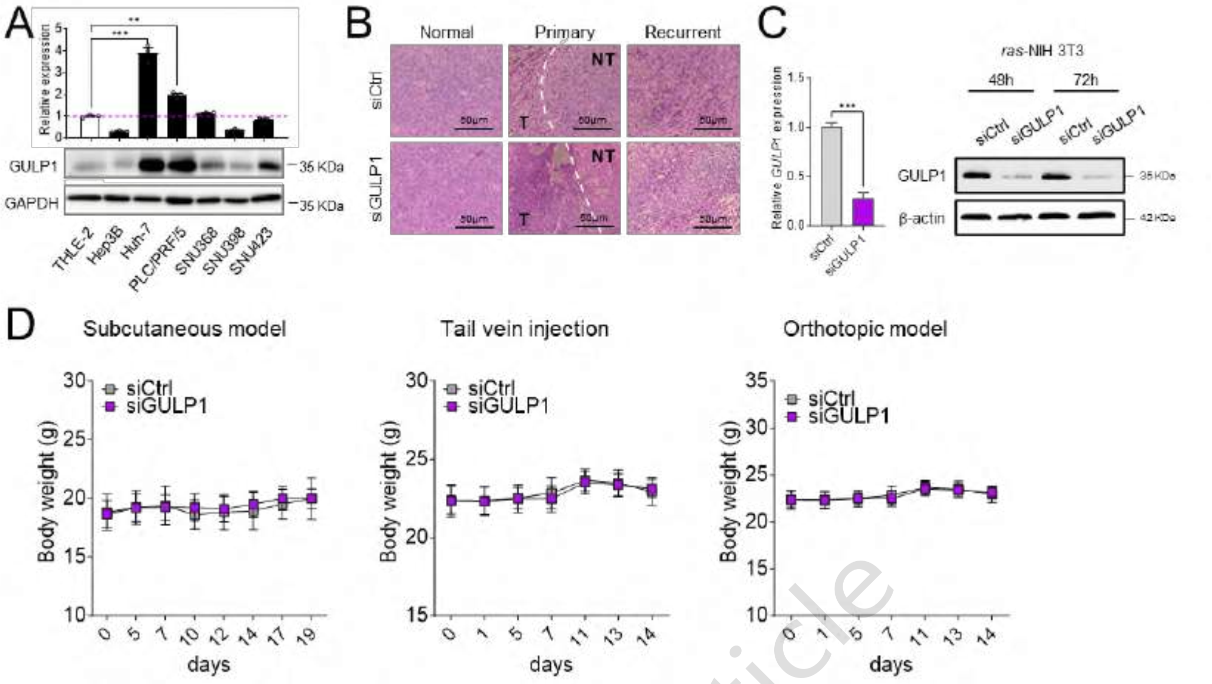
Supplementary Figure 3. *GULP1* as an indicator for HCC aggressiveness. (A) The expression level of *GULP1* was analyzed across a wide array of cancer types in the TCGA dataset. *GULP1* is significantly overexpressed in HCC compared to its expression in other cancer types, where it appears predominantly downregulated. The data points for individual samples are plotted, with median expression levels indicated by horizontal black lines. TPM, transcripts per million. (B) Spatial transcriptomic analysis. Left: representative spatial transcriptomic spots colored by non-malignant (yellow) and malignant (purple) hepatocytes from tumor tissues. Right: *GULP1* expression in spatial sections. (C) UMAP plot displaying cell type clusters across various liver cancer-related datasets, with distinct cell types marked by different colors, including hepatocytes, T cells, B cells, macrophages, endothelial cells, and fibroblasts. (D) Single-cell transcriptomic data (GSE151530) revealing GULP1 expression across various cell types, including T cells, TAMs (tumor-associated macrophages), hepatocytes, TECs (tumor endothelial cells), Tregs (regulatory T cells), B cells, and MDSCs (myeloid-derived suppressor cells). The bar chart on the right confirms that hepatocytes predominantly express GULP1. (E) UMAP-based sub-clustering of 18,539 hepatocytes into nine groups (C1–C9). (F) Hallmark pathway enrichment analysis (MSigDB Hallmark 2020) ranking gene sets enriched in GULP1 (+) samples. (G) Summary table of selected hallmark pathways with corresponding normalized enrichment scores (NES) and p-values, indicating the statistical significance of GULP1-linked oncogenic signaling in HCC.



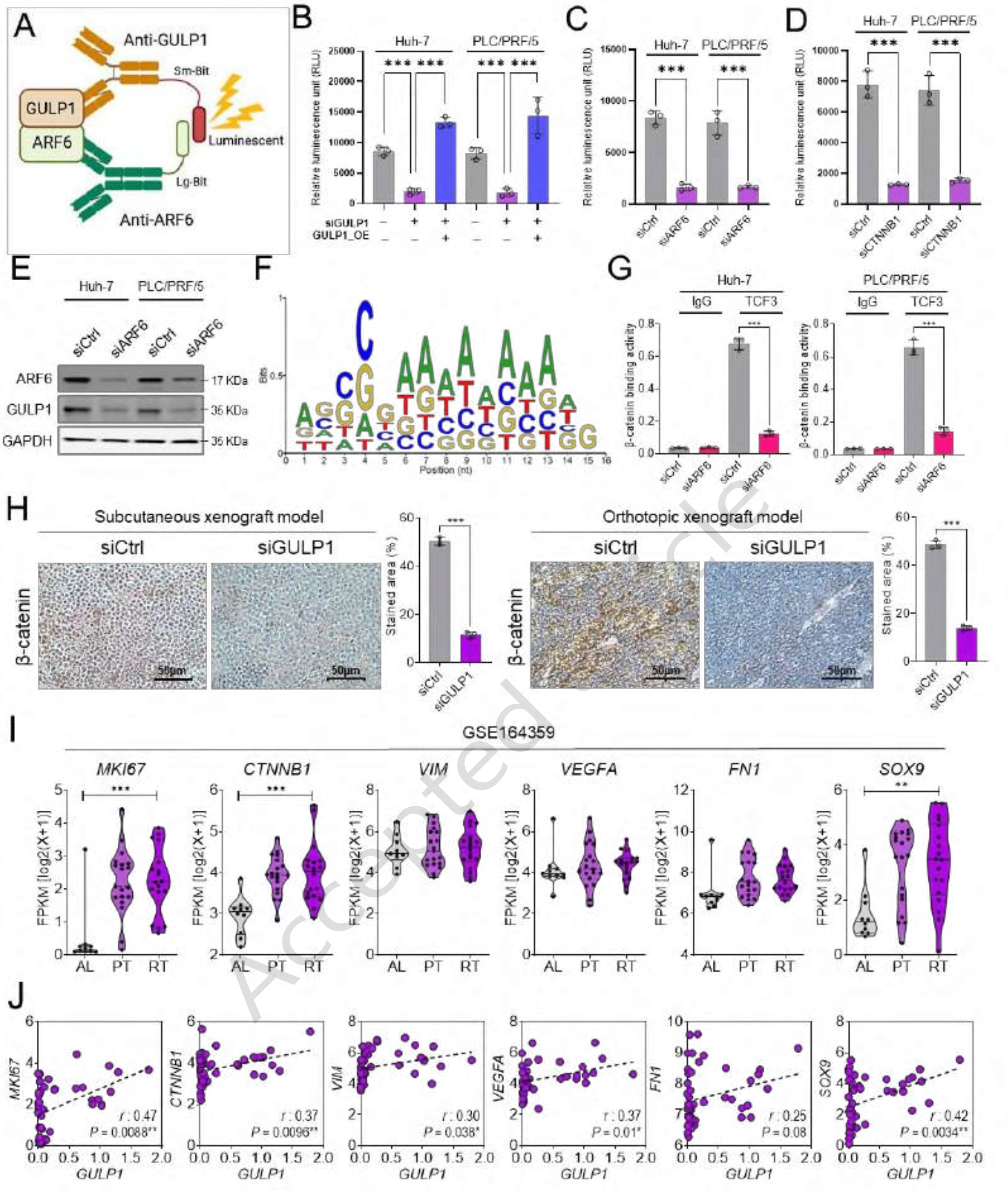
Supplementary Figure 4. GULP1 expression correlates with poor recurrence-free survival (RFS) in HCC and is elevated in recurrent tissues. (A-C) Kaplan–Meier curves showing RFS for HCC patients, stratified by high (purple) and low (gray) GULP1 expression levels. Patients with high GULP1 expression demonstrate significantly poorer RFS outcomes. Analysis was conducted across three datasets: (A) GSE14520, (B) GSE114564, and (C) TCGA LIHC, all providing recurrence information. (D) Comparative mRNA expression analysis of *GULP1* using qRT-PCR in 81 pairs of normal and HCC liver tissue samples, demonstrating significantly higher expression in HCC tissues. The x-axis numbers represent the unique patient IDs. (E) qRT-PCR analysis in a subset of 24 patients with recurrent HCC, showing elevated levels of GULP1 in recurrent tumor tissues compared to matched adjacent non-tumor (NT) tissues. The x-axis numbers represent the unique patient IDs.



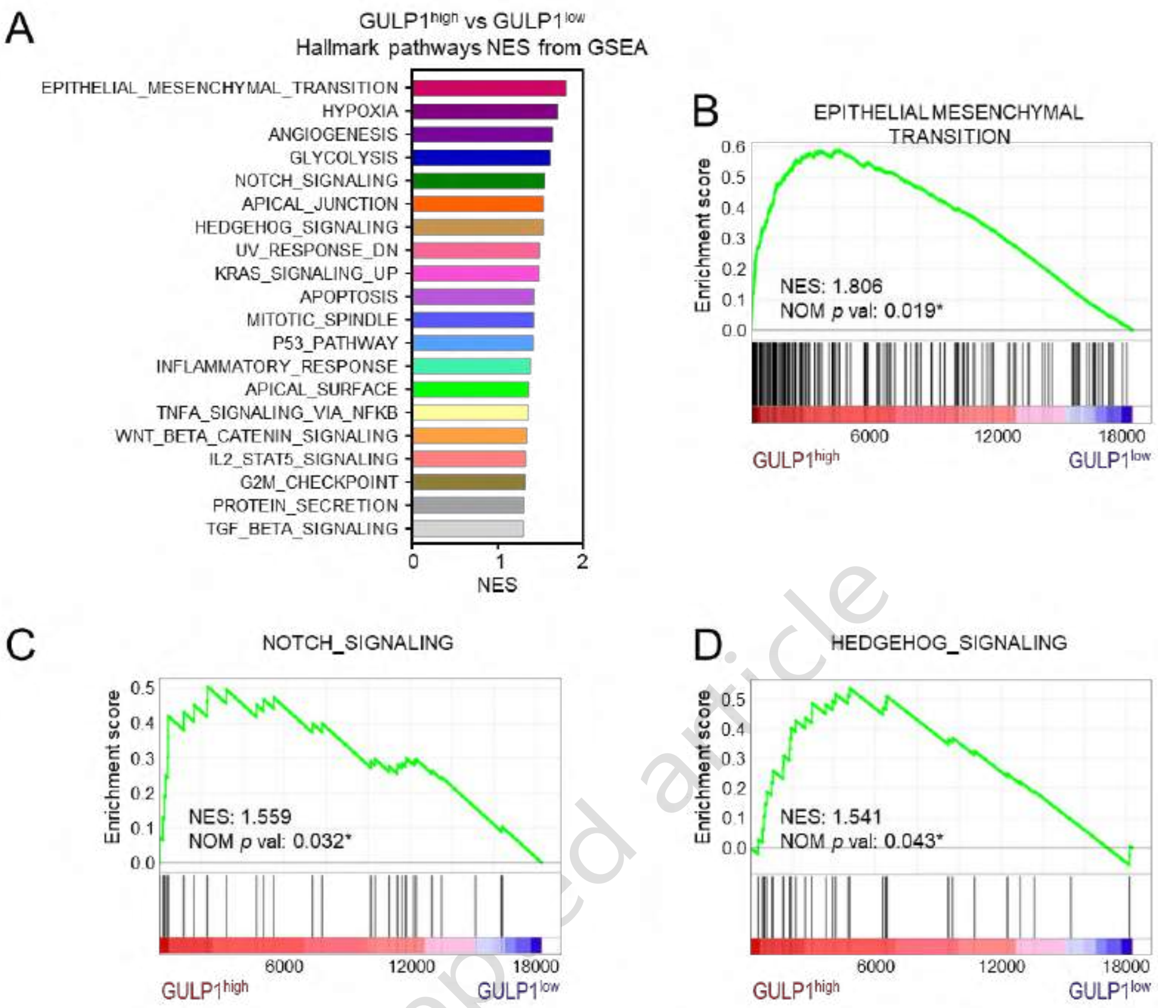
Supplementary Figure 5. GULP1 expression and diagnostic performance across liver disease etiologies. (A) Hepatitis B virus (HBV) cohort: Serum GULP1 measurements in liver cirrhosis (LC; n = 30) and HCC (n = 30). GULP1 levels were notably higher in HCC than LC (*P* < 0.001), yielding an AUC of 0.871, which exceeds that of AFP (AUC = 0.707; *P* = 0.006). NL (normal liver) refers to healthy control samples included for baseline comparison. (B) Hepatitis C virus (HCV) cohort: Serum GULP1 levels were assessed in LC (n = 30) versus HCC (n = 30). GULP1 was substantially upregulated in HCC relative to LC (*P* < 0.001), achieving an AUC of 0.821 compared to AFP's 0.801 (*P* = 0.729). (C) Alcoholic group: Serum GULP1 levels in alcoholic LC (n = 30) and alcoholic HCC (n = 30). GULP1 expression was significantly elevated in HCC versus LC (*P* < 0.001), with an AUC of 0.836 surpassing AFP's 0.810 (*P* = 0.716). (D) Metabolic dysfunction-associated steatohepatitis (MASH) group: Serum GULP1 levels in LC (n = 30) compared to HCC (n = 30). GULP1 was markedly increased in HCC (*P* < 0.001), generating an AUC of 0.890, higher than AFP's 0.810 (*P* = 0.120). (E) Comparison of serum GULP1 concentrations among normal liver (NL), liver cirrhosis (LC), hepatocellular carcinoma (HCC), and non-HCC liver tumors (Non-HCC LT). Statistical significance was determined via one-way ANOVA followed by multiple comparison tests. *P < 0.05, **P < 0.01, and ***P < 0.001. Data are represented as the mean ± standard deviation (SD). (F) Left: Non-tumor versus non–HCC liver tumor comparison: Serum GULP1 levels for non-tumor samples (NL and LC, n = 145) and non–HCC liver tumors (n = 30). GULP1 was significantly higher in non–HCC liver tumors compared to non-tumor samples (*P* < 0.001), providing an AUC of 0.833, which exceeds AFP's 0.540 (*P* < 0.0001). Right: Serum GULP1 levels and ROC analysis for LC (n = 115) and non-HCC liver tumors (n = 30). GULP1 levels were significantly higher in non-HCC liver tumors compared to LC (P < 0.001), with an AUC of 0.790, outperforming AFP's AUC of 0.524 (P = 0.002).



Supplementary Figure 6. GULP1 expression and its validation *in vivo* HCC models. (A) Relative protein abundance of GULP1 in healthy liver cell lines versus HCC cell lines, determined by Western blot analysis. The results indicate a marked increase in GULP1 protein levels in HCC cell lines, particularly in PLC/PRF/5 and Huh-7 cells. (B) Immunohistochemistry (IHC) analysis of orthotopically injected, GULP1-suppressed Hepa1-6 cells in mouse liver. The stained sections demonstrate a reduction in GULP1 expression in tumor (T) tissues compared to non-tumor (NT) tissues. (C) Validation of GULP1 suppression in ras-transformed NIH-3T3 cells. The bar graph shows the results of qRT-PCR confirming reduction in GULP1 expression (left). The Western blot analysis corroborates the suppression of GULP1 levels (right). (D) Comparative analysis of body weight upon GULP1 suppression in different *in vivo* models. Statistical significance was determined using unpaired t-tests, with ***P<0.001 indicating P<0.001. Data are shown as mean ± SEM.



Supplementary Figure 7. Direct modulation of GULP1 by ARF6 and β-catenin in HCC. (A) Schematic diagram of the FRET (Förster resonance energy transfer) assay used to detect ARF6–GULP1 interaction. ARF6 and GULP1 were each tagged with either Sm-Bit (small fragment of luciferase) or Lg-Bit (large fragment of luciferase). When the two proteins come into close proximity, these luciferase fragments reconstitute into an active enzyme that emits a luminescent signal, indicating direct protein–protein binding. (B) FRET-based quantification of luminescence signals under varying GULP1 expression levels in HCC cells. After transfection with *GULP1*-targeting siRNA (siGULP1) or *GULP1*-overexpressing vector (GULP_OE), luminescence was measured to assess changes in ARF6–GULP1 interaction. (C) FRET-based evaluation of ARF6 knockdown (siARF6). Luminescence was markedly lower upon ARF6 suppression. (D) Assessment of ARF6–GULP1 interaction following β-catenin (*CTNNB1*) depletion. Cells were transfected with siCTNNB1, and luminescence signals were measured to determine any impact on the ARF6–GULP1 complex. (E) Western blot analysis confirmed a decrease in GULP1 expression following ARF6 suppression. (F) Sequence logos representing the TCF3 transcription factor consensus DNA binding sites; the y-axis indicates the information amount at each motif position. (G) The binding activity of TCF3 at the *GULP1* promoter regions, quantified with the chromatin immunoprecipitation (ChIP) assays, and presented in bar graphs. (H) Immunohistochemistry (IHC) of β-catenin in subcutaneous and orthotopic HCC xenograft models, comparing negative control (siCtrl) and siGULP1 groups. Bar graphs show significantly reduced β-catenin staining with GULP1 suppression. (I) Violin plots from the GSE164359 dataset illustrating elevated GULP1 expression in recurrent tumor (RT) samples relative to primary tumor (PT) and adjacent liver tissue (AL). (J) Pearson correlation analysis from the GSE164359 dataset, revealing significant positive correlations between GULP1 and its downstream targets (*CTNNB1*, *FN1*, *SOX9*, etc.).



Supplementary Figure 8. GULP1-driven oncogenic pathways in HCC. (A) Bar chart of the top 20 enriched Hallmark gene sets in the GULP1^{high} subgroup (top quartile of GULP1 expression) compared to the GULP1^{low} subgroup (bottom quartile) in the TCGA LIHC dataset (n = 371). Rankings are based on normalized enrichment score (NES). (B) GSEA enrichment plot for EPITHELIAL_MESENCHYMAL_TRANSITION (NES = 1.806, NOM *p* = 0.019). (C) GSEA enrichment plot for NOTCH_SIGNALING (NES = 1.559, NOM *p* = 0.032). (D) GSEA enrichment plot for HEDGEHOG_SIGNALING (NES = 1.541, NOM *p* = 0.043).