

RESEARCH

Open Access



# Single-cell RNA sequencing combined with single-cell genome-wide association study identifies SF3B4 as a hub gene in hepatocellular carcinoma progression

Fujun Ma<sup>1†</sup>, Lihong Kang<sup>2†</sup>, Zhijian Ren<sup>3</sup>, Yang Yang<sup>3</sup>, Tong Shen<sup>3</sup> and Haibo Yu<sup>1\*</sup>

<sup>†</sup>Fujun Ma and Lihong Kang have contributed equally to this work.

\*Correspondence:

Haibo Yu

yhb2101661@163.com

<sup>1</sup>Henan Medical University, Kaifeng, China

<sup>2</sup>Department of Neonatology, Huating First People's Hospital, Pingliang, China

<sup>3</sup>Department of Gastrointestinal Surgery, Xi'an International Medical Center Hospital, Xi'an, China

## Abstract

**Objective** High heterogeneity and poor therapeutic response are major challenges in hepatocellular carcinoma (HCC). This study aimed to identify core hepatocyte subsets and key genes driving HCC progression via a multi-omics approach to inform precise diagnosis and treatment.

**Methods** We analyzed public data from GEO (GSE282701), TCGA-LIHC, and IEU-OpenGWAS (bbj-a-158). scRNA-seq, scPagwas, BayesPrism, and WGCNA were used to identify core HCC cell subsets and genes. SF3B4 expression was validated by Western blot and RT-qPCR in HCC tissues and cell lines. Functional impacts on proliferation and migration were assessed using colony formation, Transwell, and wound healing assays in HepG2 and Huh7 cells.

**Results** Integrated scRNA-seq and scPagwas analysis identified PKHD1<sup>+</sup> hepatocytes as a core HCC subset, showing significantly higher trait relevance scores versus other subtypes and a positive correlation with HCC ( $P < 0.05$ ). BayesPrism quantification in the TCGA-LIHC cohort confirmed that high abundance of PKHD1<sup>+</sup> hepatocytes correlated with poor prognosis ( $P < 0.05$ ), immune microenvironment remodeling (increased CAFs and MDSCs), and distinct somatic mutation profiles (elevated CTNNB1 and reduced TP53 mutation rates). SF3B4 was identified as the key gene associated with this subset via WGCNA and differential expression analysis. SF3B4 was upregulated in HCC tissues and cells, and its knockdown suppressed proliferation and migration in HepG2 and Huh7 cells.

**Conclusion** The PKHD1<sup>+</sup> hepatocyte subset and SF3B4 represent key regulators of HCC malignancy, offering novel potential targets for prognostic assessment and targeted therapy.

**Keywords** Hepatocellular carcinoma, PKHD1<sup>+</sup> hepatocytes, SF3B4



## 1 Background

Hepatocellular Carcinoma (HCC) is one of the most common malignant tumors worldwide, especially in Asia and Africa, where its incidence and mortality rates have remained persistently high. The pathogenesis of HCC is complex and involves multiple types of risk factors, including chronic viral hepatitis (such as hepatitis B virus and hepatitis C virus infections), alcoholic liver disease, non-alcoholic fatty liver disease (NAFLD), and metabolic syndrome [1, 2]. In recent years, with changes in lifestyles worldwide—such as sedentary behavior, high-fat diets, and dietary structure—the incidence of NAFLD-related HCC has shown a significant upward trend. It has become one of the important etiologies of HCC and poses a severe challenge to public health security [2, 3]. Early diagnosis and precise treatment of HCC have long been core challenges in clinical practice and scientific research. Due to the insidious onset of the disease and the lack of specific early clinical manifestations, most patients are diagnosed at an advanced stage. This leads to limited treatment options and poor therapeutic efficacy [4–6]. Currently, early screening and diagnosis of HCC rely on the combined application of imaging examinations and biomarkers. Computed tomography (CT) and magnetic resonance imaging (MRI) can clearly display the characteristic blood supply pattern of tumors—"arterial phase enhancement and venous phase washout"—through dynamic contrast-enhanced scanning. This pattern provides key evidence for distinguishing HCC from other liver lesions such as hepatic hemangiomas and hepatic adenomas [7, 8]. Alpha-fetoprotein (AFP), as a classic HCC biomarker, shows a significant correlation between its substantially elevated levels and both increased tumor malignancy and poor patient prognosis [9]. However, the diagnostic sensitivity and specificity of a single biomarker have limitations. Therefore, the combined detection of multiple biomarkers has become a key strategy to improve diagnostic accuracy.

Current clinical management of liver cancer emphasizes a multidisciplinary treatment model. Treatment strategies need to be formulated comprehensively based on the patient's tumor stage, liver function status, and systemic conditions. These strategies cover a variety of intervention methods, including surgical treatment (hepatectomy, liver transplantation), local ablation therapy (radiofrequency ablation, microwave ablation), transcatheter arterial chemoembolization, systemic targeted therapy (sorafenib, lenvatinib), and immune checkpoint inhibitors [2, 10, 11]. However, the existing treatment system still faces several core limitations. Firstly, chemotherapy resistance mediated by the tumor microenvironment—such as fibrotic barriers formed by cancer-associated fibroblasts and infiltration of immunosuppressive cells—limits the long-term efficacy of local treatments like transcatheter arterial chemoembolization (TACE). Secondly, the population eligible for targeted therapy is relatively narrow; for example, lenvatinib is only effective in some patients with high VEGF expression. Thirdly, the overall response rate to immunotherapy is relatively low. Additionally, for the treatment of advanced hepatocellular carcinoma, although the study by Wang et al. demonstrated that local ± systemic therapy significantly improved the 5-year overall survival (OS) rate, long-term OS remains difficult to enhance [12]. These clinical challenges highlight the urgent need to deeply analyze the molecular mechanisms of liver cancer pathogenesis, develop novel diagnostic biomarkers, and identify innovative therapeutic targets. Additionally, establishing and optimizing early screening systems—such as regular

ultrasound combined with multi-biomarker detection for high-risk populations—is a key entry point to improve the prognostic level of the disease.

Intratumor Heterogeneity (ITH) refers to the significant differences existing among different cells within the same tumor at multiple levels, including genetics, epigenetics, transcriptome, proteome, and immune microenvironment. It is reflected not only in the diversity of biological characteristics of tumor cells themselves but also involves the complex interactions between immune cells (such as CD8<sup>+</sup> T cells and tumor-associated macrophages), stromal cells (such as cancer-associated fibroblasts), and the vascular system in the Tumor Microenvironment (TME) [13, 14]. As one of the core biological characteristics of liver cancer, intratumor heterogeneity plays a decisive role in tumor evolution (e.g., subclonal selection), treatment resistance (e.g., enrichment of targeted drug-resistant subpopulations), and prognostic evaluation. It is also a key reason for the significant individual differences in clinical treatment effects [13, 14]. ITH of liver cancer is a key factor affecting its clinical treatment efficacy and prognostic judgment. As the main pathological subtype of liver cancer, HCC exhibits significant heterogeneous characteristics in terms of histological features, genomic variations, transcriptional regulation, and epigenetic modifications [15]. This multi-level heterogeneity directly leads to differences in the biological behaviors of tumor cells (such as proliferation, invasion, and metastasis abilities) and their responses to treatment, which in turn affects the clinical outcomes of patients [16, 17]. At the epigenetic level, DNA methylation and copy number variation (CNV) are important drivers of liver cancer heterogeneity. Studies have shown that significant changes in DNA methylation and CNV already exist in the early precancerous stage of liver cancer, and these changes exhibit different patterns in regenerative nodules (RNs) and dysplastic nodules (DNs) [18]. For instance, RNs with high-frequency epigenetic changes typically exhibit low CNV. This indicates that different nodules have differences in epigenetic and genetic components, and these differences collectively drive the progression of the disease [19]. Furthermore, epigenetic heterogeneity is also closely associated with the proliferative capacity and clinical characteristics of liver cancer. For example, Ki67 staining shows that nodules with a high epigenetic progression score exhibit stronger proliferative capacity [20]. Single-cell RNA sequencing (scRNA-seq) has demonstrated unique technical advantages and application value in revealing liver cancer heterogeneity. Firstly, through high-throughput sequencing and data analysis, scRNA-seq can classify cells in liver cancer tissues into different subsets and uncover the molecular characteristics of these subsets. For example, studies have identified multiple subtypes in liver cancer cells, and these subtypes exhibit significant differences in gene expression, metabolic pathways, and signaling pathways [21]. These differences may be closely related to tumor invasiveness, metastatic potential, and sensitivity to treatment. Through single-cell analysis, researchers can identify key genes and pathways that drive tumor progression, providing a basis for the development of targeted therapeutic strategies [22]. Secondly, scRNA-seq can also deeply analyze the composition and functional status of immune cells in the liver cancer microenvironment. Immune cells in the liver cancer microenvironment, such as T cells, B cells, and macrophages, play important roles in tumor immune escape and immunotherapy. Through single-cell sequencing, researchers can identify the activation status, functional characteristics of different immune cell subsets, as well as their interactions with cancer cells [21]. Given the high heterogeneity of HCC, current clinical treatment strategies are

gradually moving toward individualized treatment guided by multi-omics integrated analysis [23].

This study adopts a systematic multi-omics integration analysis strategy. By integrating single-cell transcriptome sequencing data, genome-wide association study (GWAS) data, and bulk transcriptome sequencing data, it aims to accurately identify the core hepatocyte subsets in HCC and deeply analyze the mechanisms underlying their association with patients' clinical prognosis, tumor immune microenvironment characteristics, and somatic mutation profiles. Based on the above findings, the study will further screen the key genes regulating HCC progression and verify them through in vitro cell function experiments (such as cell proliferation assays and Transwell invasion assays). This work intends to provide a theoretical basis and experimental foundation for the development of early diagnostic biomarkers and the screening of precise therapeutic targets for liver cancer. The specific research technical route is shown in Fig. 1.

2 Methods

2.1 Cell culture and tissue sample acquisition

The hepatocellular carcinoma (HCC) cell lines HepG2 (RRID: CVCL\_0027) and normal liver cell line LO2 (RRID: CVCL\_1D04) were purchased from Shanghai Enzyme

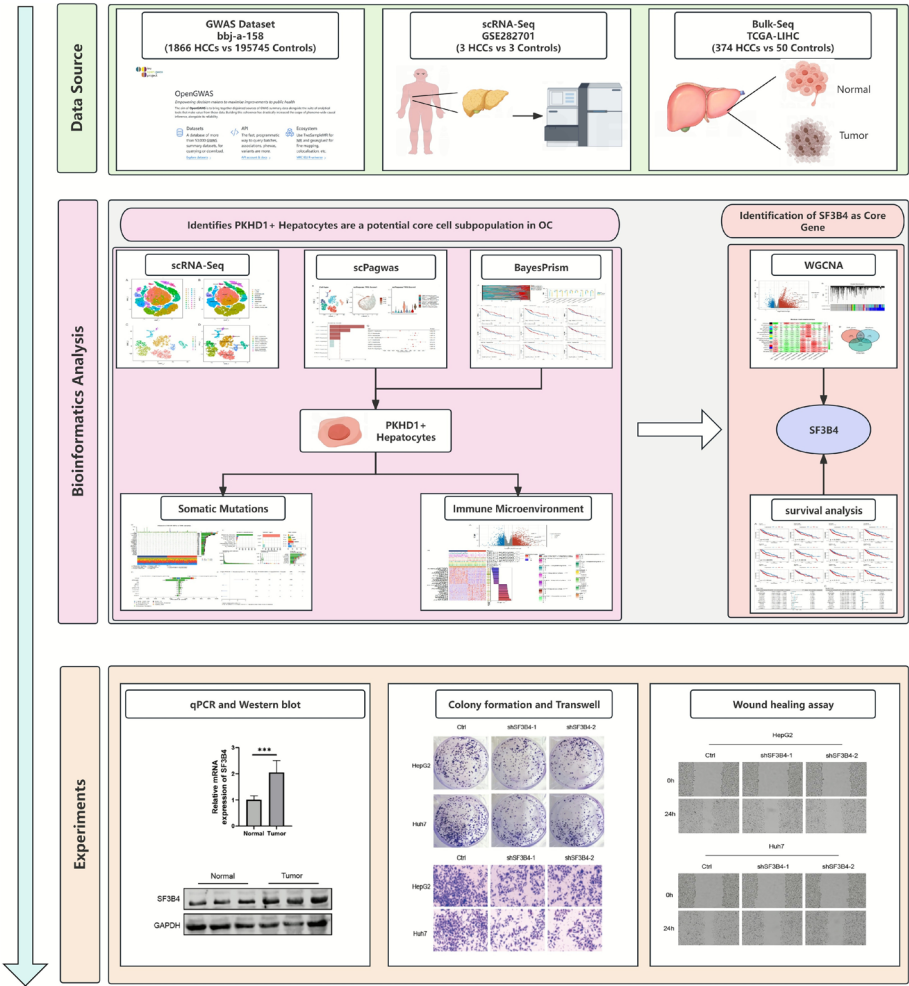


Fig. 1 Technical workflow of this study

Research Biotechnology Co., Ltd. (Shanghai, China), while the HCC cell line Huh-7 (RRID: CVCL\_0336) was obtained from Wuhan Procell Life Science & Technology Co., Ltd. The HEK-293 T cell line was provided by Abcell. The HepG2 cell line was cultured in high-glucose DMEM medium (Gibco, USA) supplemented with 1% penicillin/streptomycin (Beyotime Biotechnology, Shanghai, China) and 10% fetal bovine serum (Procell, Wuhan, China). The LO2, HEK-293 T, and Huh-7 cell lines were cultured in MEM medium (Gibco, USA) with 1% penicillin/streptomycin (Beyotime Biotechnology, Shanghai, China) and 10% fetal bovine serum (Procell, Wuhan, China). All cell lines were maintained in a constant temperature incubator at 37 °C with 5% CO<sub>2</sub>. Normal liver tissue (n = 10) and HCC tumor tissue (n = 10) samples were collected from Xi'an International Medical Center. The experimental protocol of this study was approved by the Ethics Committee of Xi'an International Medical Center, and informed consent was obtained from all participants.

## 2.2 Transfection

All plasmids were purchased from Synbio Technologies (Suzhou, China). For lentiviral packaging, the pXPax2 plasmid, which encodes essential viral packaging components, and the pMD2.G plasmid, which provides the vesicular stomatitis virus glycoprotein (VSV-G) envelope for pseudotyping, were co-transfected with the target transfer plasmid (directed against the SF3B4 gene) into HEK293T cells to produce lentiviral particles. Following lentiviral transduction, the target cells (HepG2 and Huh7 cells) underwent puromycin selection. The optimal puromycin concentration was predetermined via a kill curve assay to ensure complete elimination of non-transduced cells. The transduced HepG2 and Huh7 cells were cultured in puromycin-containing medium for 7–14 days to establish stable polyclonal cell populations. Finally, successful integration of the target gene into the cell genome was verified by measuring SF3B4 transcript levels using RT-qPCR and/or detecting SF3B4 protein expression via Western blot analysis.

## 2.3 Quantitative Real-Time PCR (RT-qPCR)

Total RNA was extracted from HepG2 and Huh7 cells (including the siSF3B4-2 knock-down group, siSF3B4-1 knockdown group, and Ctrl control group) and HCC tissue samples. For cell samples, total RNA was extracted directly using TRIzol reagent (Invitrogen, USA) according to the manufacturer's instructions. For HCC tissue samples, the tissues were first thoroughly homogenized by grinding in liquid nitrogen or using a tissue homogenizer, followed by RNA extraction with TRIzol reagent (Invitrogen, USA), strictly adhering to the manufacturer's protocol. Subsequently, reverse transcription was performed using a Reverse Transcription Kit (TOYOBO, Osaka, Japan) with the extracted total RNA from both cells and tissues as the template to synthesize complementary DNA (cDNA).

RT-qPCR analysis was carried out using the ChamQ SYBR qPCR Master Mix kit (Vazyme, Nanjing, China) on a LightCycler 96 real-time PCR system (Roche, Germany). The expression levels of the target genes in all cell and HCC tissue samples were normalized to the endogenous control gene Beta-actin rRNA. All primers were purchased from Sangon Biotech (Shanghai, China). The relative expression level of SF3B4 mRNA, normalized to Beta-actin rRNA, was calculated using the  $2^{-\Delta\Delta CT}$  method for unified

analysis of both cell and tissue sample data. Detailed information for all primers is provided in Table 1.

2.4 Western Blot

To detect the expression levels of SF3B4 protein in human hepatocellular carcinoma (HCC) tissue samples and in HepG2 and Huh7 cell lines, sample lysates were prepared as follows. Cell samples were directly lysed on ice for 30 min using RIPA lysis buffer (supplemented with protease inhibitors) with repeated pipetting to ensure complete lysis. HCC tissue samples were first ground into a powder in liquid nitrogen, then homogenized on ice in RIPA lysis buffer (containing protease inhibitors) and lysed for an additional 30 min. After lysis, the samples were centrifuged at 12,000 ×g and 4 °C for 15 min, and the resulting supernatant was collected as the total protein extract. The protein concentration of all samples was measured using a Nanodrop One spectrophotometer (Thermo Fisher, USA) to standardize the loading amount.

The quantified protein samples were mixed with 5 × SDS loading buffer at a 4:1 ratio and denatured at 95 °C for 10 min. Subsequently, 30 µg of protein per lane was loaded and separated by electrophoresis on a 10% SDS-PAGE gel (80 V for the stacking gel and 120 V for the separation gel). After electrophoresis, proteins were transferred from the gel to a nitrocellulose membrane (Pall, Port Washington, NY, USA) using the wet transfer method under constant current of 300 mA for 90 min in an ice bath. Following transfer, the membrane was blocked with 5% skim milk (prepared in TBST buffer) at room temperature for 2 h to prevent non-specific binding.

After blocking, the membrane was incubated overnight at 4 °C with the following primary antibodies: anti-SF3B4 antibody (Proteinch, Cat# 85,663–5-RR, dilution 1:5000) and anti-GAPDH antibody (Abways, Cat# Ab0037, dilution 1:10,000) as the loading control. The next day, the membrane was washed three times for 10 min each with TBST buffer at room temperature on a shaker. It was then incubated with HRP-conjugated goat anti-rabbit or goat anti-mouse secondary antibody (dilution 1:5000) at room temperature for 1 h. After incubation, the membrane was washed again three times for 10 min each with TBST buffer. Finally, the blots were developed, and images were captured using a gel imaging system for band intensity analysis.

2.5 Colony formation assay

In the colony formation assay, HepG2 and Huh7 cells in the logarithmic growth phase were first trypsinized, collected, and washed twice with PBS. The cells were then resuspended in DMEM medium supplemented with 10% fetal bovine serum (FBS), and the cell density was adjusted using a hemocytometer. Subsequently, HepG2 cells were seeded at a density of 1 × 10<sup>3</sup> cells per well and Huh7 cells at 1.2 × 10<sup>3</sup> cells per well into 6-well plates. Each well contained 2 mL of DMEM medium with 10% FBS. The plates were gently shaken to ensure even cell distribution and then placed in a humidified

**Table 1** The primer information for qPCR in this study

| Gene name  | Primer name | Primer sequence         |
|------------|-------------|-------------------------|
| SF3B4      | Primer F    | GAGGCCCTCTCCCTCAGTAA    |
|            | Primer R    | TTTGCCCCAAGGAGCTACAG    |
| Beta Actin | Primer F    | GAGAGGGAACCTCGTGC GTGAC |
|            | Primer R    | CATCTGCTGGAAGGTGGACA    |

incubator at 37 °C with 5% CO<sub>2</sub>. The culture medium was replaced with fresh medium every three days to maintain optimal growth conditions.

After 14 days of continuous culture, when visible cell colonies had formed, the assay was terminated. The medium was carefully aspirated from each well, and the cells were gently washed twice with PBS to remove residual medium. Next, the cells were fixed with 4% paraformaldehyde for 30 min at room temperature. After fixation, the fixative was removed, and the wells were washed twice with PBS. The cells were then stained with 0.1% crystal violet solution (Solarbio, Beijing, China) for 20 min at room temperature, protected from light. Following staining, the plates were rinsed gently with deionized water until excess dye was removed and air-dried in an inverted position. Finally, colonies containing more than 50 cells were observed and counted under a standard light microscope. The number of colonies and the colony formation rate were calculated for each group to evaluate the impact of SF3B4 knockdown on the clonogenic ability of HepG2 and Huh7 cells.

## 2.6 Transwell migration assay

The Transwell migration assay was performed using 24-well Transwell chambers (Corning, NY, USA, Cat. #3422). HepG2 and Huh7 cells in the logarithmic growth phase were first trypsinized, collected, and washed twice with PBS. The cells were resuspended in serum-free DMEM medium, and the cell density was adjusted to  $1.5 \times 10^5$  cells/mL using a hemocytometer.

Subsequently, 0.2 mL of the cell suspension (containing  $3 \times 10^4$  cells per well) was added to the upper chamber of the Transwell insert. The lower chamber was filled with 0.6 mL of DMEM medium supplemented with 10% FBS, which served as a chemoattractant to induce cell migration. The Transwell chamber was carefully placed into the 24-well plate to avoid air bubbles and incubated at 37 °C with 5% CO<sub>2</sub> for 36 h.

After incubation, the Transwell chamber was removed. The medium in the upper chamber was discarded, and the chamber was gently washed twice with PBS. Non-migrated cells on the upper surface of the membrane were carefully wiped off using a sterile cotton swab. The cells on the lower surface of the membrane were then fixed with 4% paraformaldehyde for 30 min at room temperature. After fixation, the fixative was removed, and the membrane was washed twice with PBS. Staining was performed with 0.1% crystal violet solution (Solarbio, Beijing, China) for 20 min at room temperature, protected from light. After staining, the chamber was rinsed gently with deionized water to remove excess dye and air-dried in an inverted position.

Finally, the migrated cells on the lower side of the membrane were observed and photographed under an optical microscope (20× objective). Cells in five randomly selected fields were counted, and the average number of migrated cells per group was calculated to evaluate the effect of SF3B4 knockdown on the migratory ability of HepG2 and Huh7 cells.

## 2.7 Wound healing assay

The wound healing assay was used to evaluate the migratory ability of HepG2 and Huh7 cells. Cells from each group in the logarithmic growth phase were first trypsinized with 0.25% trypsin, collected, and resuspended in DMEM medium (Gibco, USA) containing 10% FBS. The cell density was adjusted to  $1 \times 10^6$  cells/mL, and 2 mL of the cell

suspension was seeded into each well of a 6-well plate. The plate was then incubated at 37 °C with 5% CO<sub>2</sub> until the cells reached 90–95% confluence.

After the cells formed a confluent monolayer, a sterile 200 µL pipette tip was used to create a straight, uniform scratch vertically through the center of each well. The plate was gently washed 2–3 times with PBS to remove dislodged cells and debris. Subsequently, 2 mL of serum-free DMEM medium was added to each well to eliminate the influence of growth factors present in serum, thereby assessing only cell migration. The plate was returned to the incubator for continued culture.

Images of the scratched area were captured at 0 h and 24 h using an inverted optical microscope (Olympus, Japan). Three fixed fields of view per well were selected and marked to ensure the same regions were observed at each time point. The average wound width at each time point was measured using ImageJ software. The cell migration rate was calculated using the following formula: Migration rate (%) = [(0 h wound width - wound width at a given time point) / 0 h wound width] × 100%. By comparing the migration rates among the different groups, the impact of SF3B4 knockdown on the migratory ability of HepG2 and Huh7 cells was analyzed.

## 2.8 Bioinformatics analysis

The dataset GSE282701 was downloaded from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). It contains single-cell transcriptome data from 3 paracancerous tissues and 3 HCC tissues. Transcriptome data, somatic mutation data, and clinical data for 374 HCC samples and 50 paracancerous control tissues from the TCGA-LIHC project were downloaded from The Cancer Genome Atlas (TCGA) database. Samples with incomplete survival information or a survival time of less than 30 days were excluded during survival analysis and model construction. GWAS summary data for liver cancer (ID: bbj-a-158) were downloaded from IEU-OpenGWAS (<https://gwas.mrcieu.ac.uk/>), which included 1,866 liver cancer cases and 195,745 healthy controls.

The scRNA-seq dataset GSE282701 was analyzed using the standard workflow in the R package "Seurat". Cells with fewer than 200 or more than 6000 detected genes, and those where mitochondrial genes accounted for more than 10% of the counts, were filtered out. The R package "harmony" was used to reduce batch effects between samples. The top 2000 highly variable genes were identified using the FindVariableFeatures function. Principal Component Analysis (PCA) was employed for dimensionality reduction. Marker genes were identified using the FindMarkers function with its default parameters. Cell subpopulations were annotated using the "SingleR" package.

scPagwas employs a multi-gene regression model to prioritize a set of trait-associated genes and identify trait-relevant cell subpopulations by integrating pathway activity scores derived from scRNA-seq data with GWAS summary statistics. In this study, the R package "scPagwas" was used to identify key cell subpopulations in liver cancer.

BayesPrism, a cutting-edge Bayesian model-based method, effectively integrates scRNA-seq data as a reference to deconvolve bulk RNA-seq data, enabling the inference of posterior distributions for cell type proportions and gene expression. This study utilized the R package "BayesPrism" to project the cell subtypes identified from the HCC scRNA-seq data onto the bulk RNA-seq data from the TCGA-LIHC project and to score each cell subtype.

Limma (Linear Models for Microarray Data) is a method for differential expression analysis based on generalized linear models. Here, we used the R package limma (version 3.40.6) to perform differential analysis to identify differentially expressed genes (DEGs) between different comparison groups and the control group. Genes with an absolute fold change  $> 2$  and an adjusted p-value  $< 0.05$  were defined as DEGs. The results were visualized using the R package "ggplot2", presented as volcano plots and heatmaps.

The R package WGCNA was used to construct co-expression networks. First, sample clustering was performed to check for potential outliers. Second, an automatic network construction function was used to build the co-expression network. The soft-thresholding power  $\beta$  was calculated using the pickSoftThresholdfunction to which the co-expression similarity was raised to calculate adjacency. Third, hierarchical clustering and the dynamic tree cut function were applied to detect modules. Fourth, gene significance and module membership were calculated, and modules were correlated with the trait (cell abundance). Gene information from relevant modules was extracted for further analysis.

## 2.9 Statistical analysis

For comparisons between two groups, either the Student's t-test or the Mann–Whitney U test was used, depending on whether the data followed a normal distribution. Comparisons among more than two groups were performed using the Kruskal–Wallis test. Correlation analyses were conducted using Spearman's rank correlation. Survival analysis was carried out using the Kaplan–Meier method with the log-rank test, along with univariate Cox proportional hazards regression. All statistical analyses were performed using R software. A P-value of less than 0.05 was considered statistically significant ( $P < 0.05$ ,  $P < 0.005$ ,  $P < 0.001$ , ns:  $P > 0.05$ ).

## 3 Results

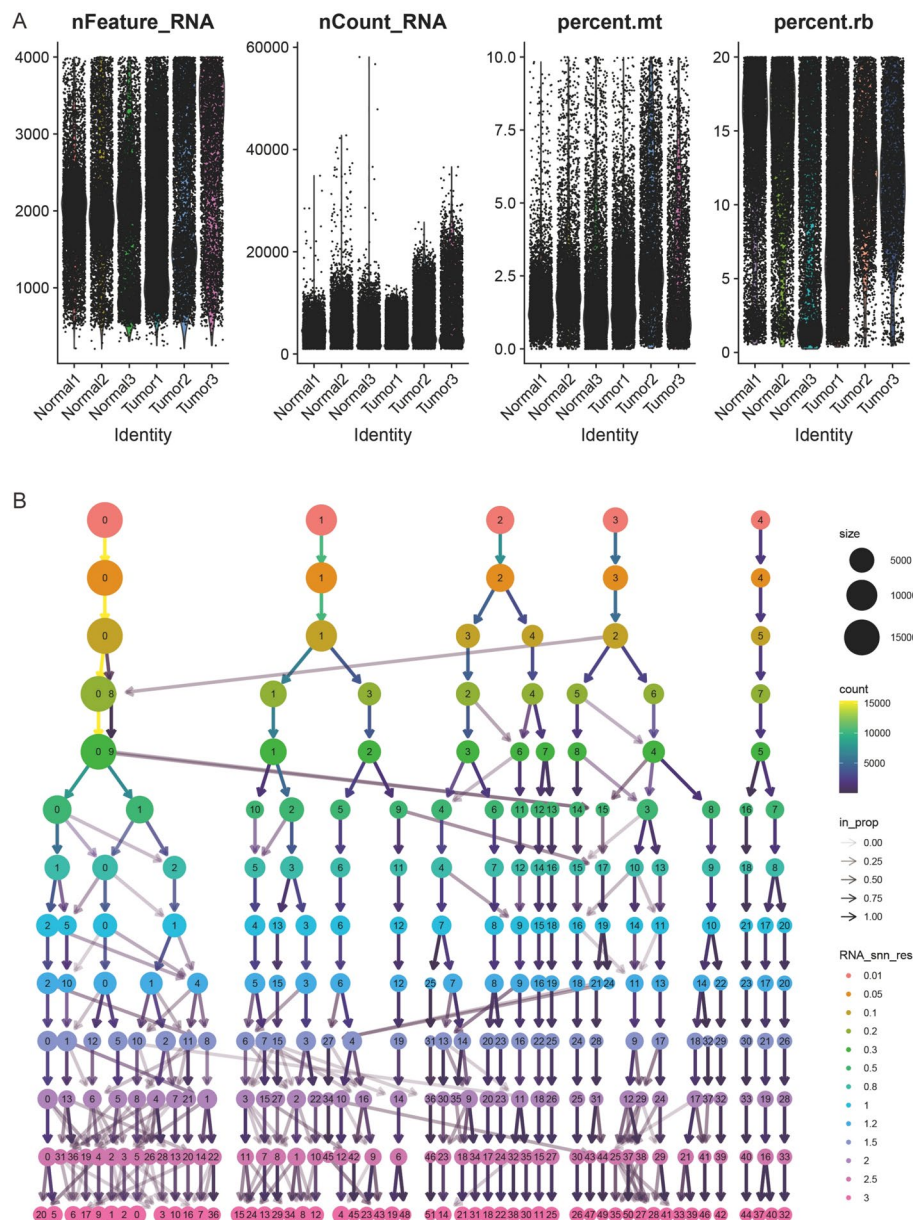
### 3.1 scRNA-seq analysis identifies heterogeneity of hepatocytes in HCC

To better understand the immune microenvironment landscape of hepatocellular carcinoma (HCC) at the single-cell level, we performed an analysis of public scRNA-seq data (Fig. 2). A total of 22 distinct cell clusters were identified across the control and HCC groups (Fig. 3A). Following annotation, these clusters were classified into 10 major cell subtypes: natural killer (NK) cells, T cells, endothelial cells, monocytes, macrophages, neutrophils, hepatocytes, B cells, tissue stem cells, and smooth muscle cells (Fig. 3B).

Subsequently, hepatocytes were extracted for further sub-clustering analysis. This analysis identified 9 distinct hepatocyte subpopulations across the two groups (Fig. 3C). Based on the expression of established marker genes, these subpopulations were annotated and designated as follows: FBP1-positive hepatocytes, ABCB11-positive hepatocytes, DLK1-positive hepatocytes, CYP3A4-positive hepatocytes, PKHD1-positive hepatocytes, CD74-positive hepatocytes, TUBA1B-positive hepatocytes, CCL4-positive hepatocytes, and FKBP5-positive hepatocytes (Fig. 3D).

### 3.2 scPagwas analysis identifies PKHD1 + hepatocytes as a core hepatocyte subpopulation in HCC associated with patient prognosis

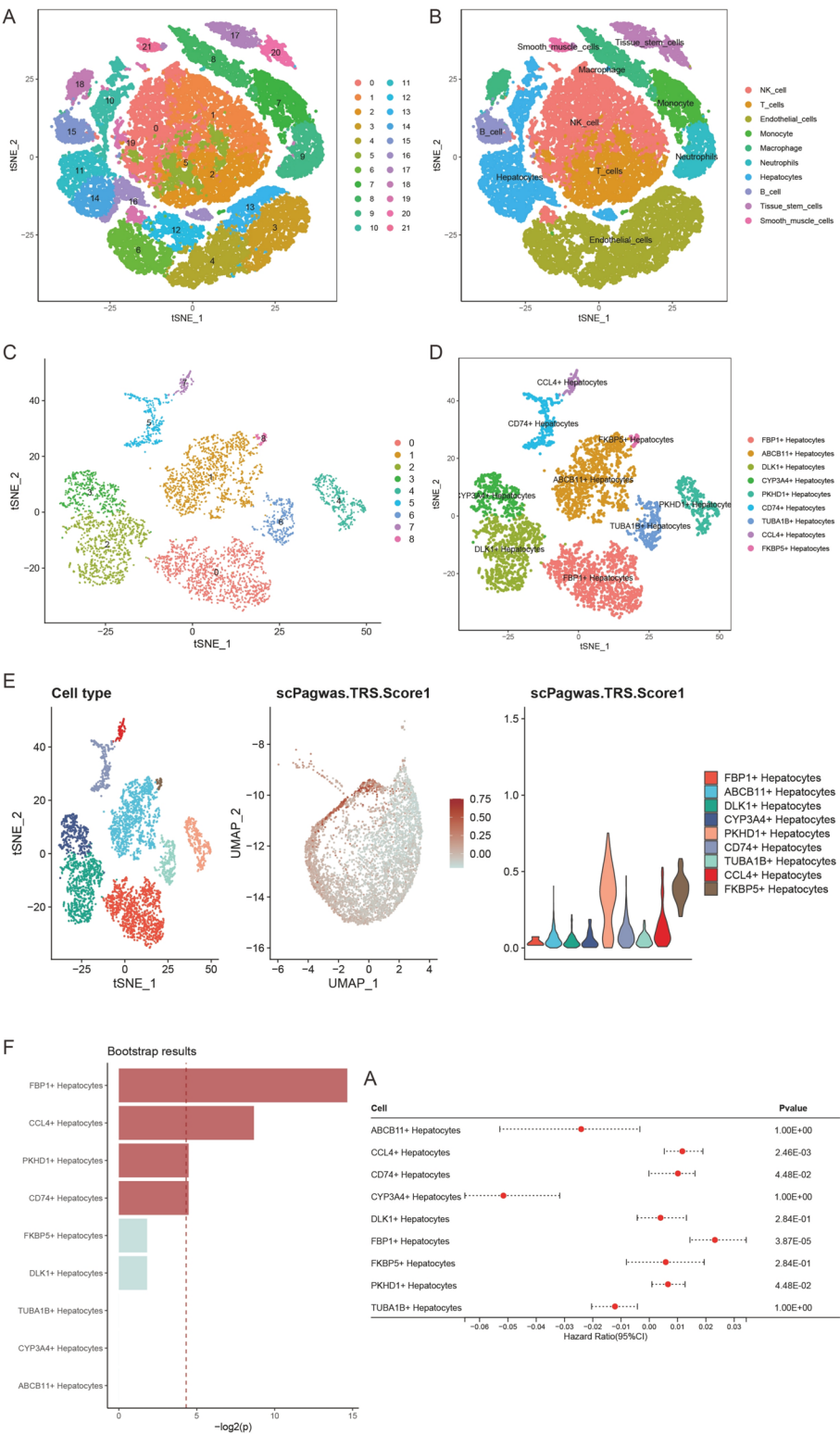
Subsequently, we performed an scPagwas analysis by integrating GWAS summary data to calculate the trait relevance score (TRS) for each epithelial cell subtype. The results revealed that among these nine hepatocyte subtypes, PKHD1 + hepatocytes exhibited a



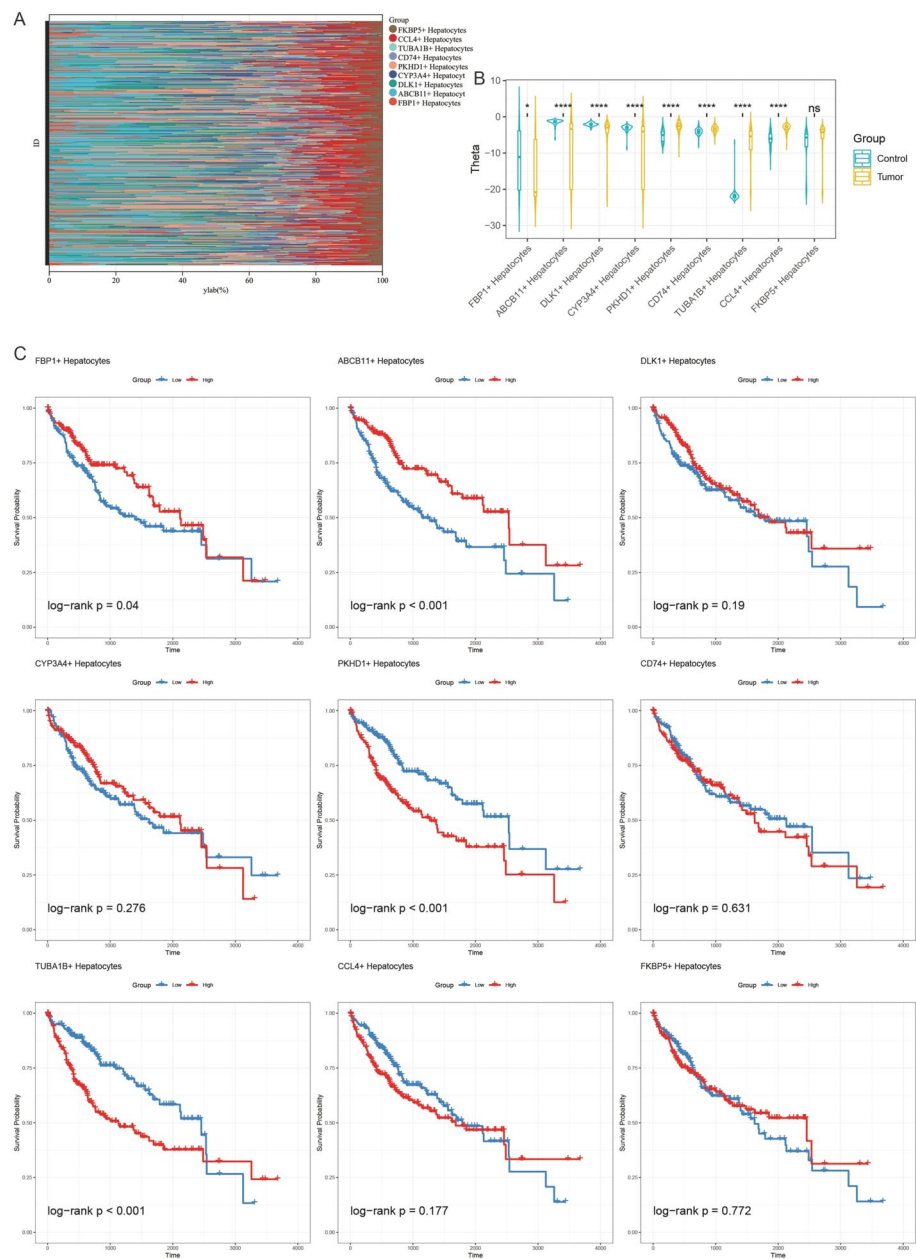
**Fig. 2** Single-cell transcriptome analysis of hepatocellular carcinoma (HCC). **A** Quality control plot. **B** Resolution parameter selection

significantly higher TRS (Fig. 3E). Bootstrap analysis further confirmed a positive association between PKHD1 + hepatocytes and HCC (Fig. 3F and G,  $P < 0.05$ ). These findings suggest that PKHD1 + hepatocytes represent a potential core cell subpopulation in HCC.

We then quantified the proportions of the nine hepatocyte subtypes, identified from the scRNA-seq data, within the TCGA-LIHC dataset using the BayesPrism algorithm (Fig. 4A). Subsequently, we divided them into high- and low-hepatocyte score groups based on the median score of each hepatocyte subtype. The results showed that the proportions of FBP1-positive hepatocytes, ABCB11-positive hepatocytes, DLK1-positive hepatocytes, and CYP3A4-positive hepatocytes were decreased in the HCC group. Conversely, the proportions of PKHD1-positive hepatocytes, CD74-positive hepatocytes,



**Fig. 3** Single-cell RNA sequencing analysis reveals cellular heterogeneity in HCC. **A** t-SNE plot of cell clusters from the control and HCC groups, identifying 22 distinct clusters. **B** Annotation of cell subpopulations into 10 subtypes using SingleR. **C** t-SNE plot of hepatocyte subclustering from control and HCC groups, identifying 9 hepatocyte clusters. **D** Naming of the 9 hepatocyte subtypes based on marker gene expression. **E** Bar plot of trait relevance scores (TRS) for each hepatocyte subtype, showing PKHD1+ hepatocytes have the highest TRS. **F–G** Bootstrap validation results demonstrating a positive correlation between PKHD1+ hepatocytes and HCC ( $P < 0.05$ )



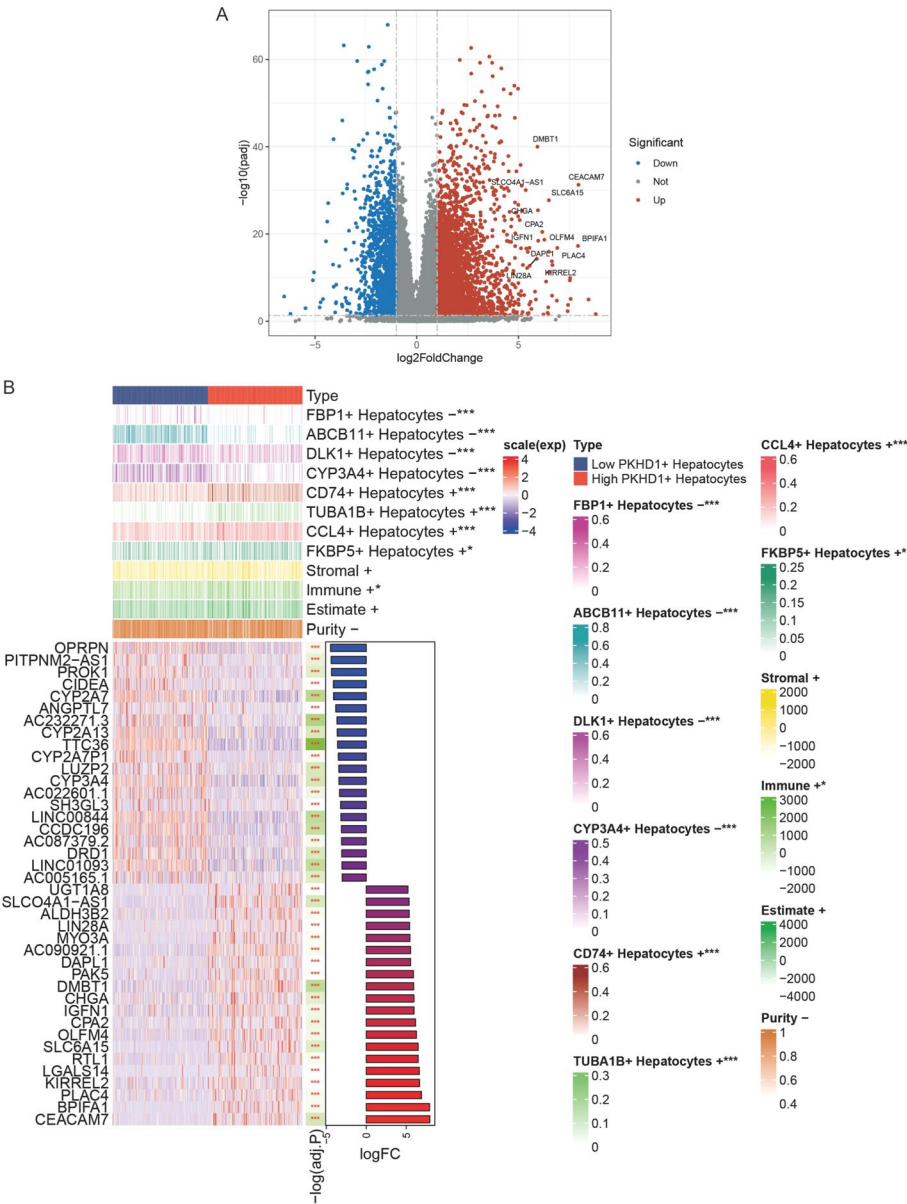
**Fig. 4** Identification of PKHD1<sup>+</sup> hepatocytes as a core cellular subpopulation in HCC via integrated scPagwas and BayesPrism analysis. **A** Heatmap of the abundances of 9 hepatocyte subtypes in the TCGA-LIHC cohort, quantified using BayesPrism. **B** Box plots comparing the abundances of the 9 hepatocyte subtypes between HCC and paracancerous tissues. **C** Kaplan–Meier overall survival (OS) curves based on the abundances of the 9 hepatocyte subtypes

TUBA1B-positive hepatocytes, and CCL4-positive hepatocytes were increased in the HCC group (Fig. 4B,  $P < 0.05$ ).

Survival analysis indicated that the proportions of FBP1-positive hepatocytes, ABCB11-positive hepatocytes, PKHD1-positive hepatocytes, and TUBA1B-positive hepatocytes were associated with HCC patient prognosis. Specifically, a higher proportion of FBP1-positive hepatocytes was associated with better patient prognosis, whereas higher proportions of ABCB11-positive hepatocytes, PKHD1-positive hepatocytes, and TUBA1B-positive hepatocytes were associated with poorer prognosis (Fig. 4C,  $P < 0.05$ ).

3.3 PKHD1 +hepatocyte proportion is associated with the HCC immune microenvironment and somatic mutations

We subsequently performed differential gene expression analysis between HCC samples with high and low PKHD1 +hepatocyte scores. The results revealed 3,650 upregulated genes and 1,479 downregulated genes in the high-score group compared to the low-score group (Fig. 5A). Further analysis showed that, relative to the low-score group, the high-score group exhibited significantly lower proportions of FBP1-positive hepatocytes, ABCB11-positive hepatocytes, DLK1-positive hepatocytes, and CYP3A4-positive hepatocytes, but significantly higher proportions of CD74-positive hepatocytes, TUBA1B-positive hepatocytes, CCL4-positive hepatocytes, and FKBP5-positive hepatocytes (Fig. 5B). Assessment of the HCC immune microenvironment using the ESTIMATE



**Fig. 5** Association analysis between PKHD1+ hepatocytes and the HCC immune microenvironment. **A** Volcano plot of differentially expressed genes between high and low PKHD1+ hepatocyte score groups. **B** Comparison of hepatocyte subtype abundances and immune scores between the two groups

algorithm indicated a significantly higher Immune Score in the high PKHD1 + hepatocyte score group compared to the low-score group (Fig. 5B).

We further investigated the association between PKHD1 + hepatocytes and the immune microenvironment in HCC patients. Compared to the low PKHD1 + hepatocyte abundance group, the high-abundance group showed increased scores for CAFs (Cancer-Associated Fibroblasts), CD8 + T cells, MDSCs (Myeloid-Derived Suppressor Cells), and IFN- $\gamma$ , but decreased Dysfunction, MSI (Microsatellite Instability), and TIDE (Tumor Immune Dysfunction and Exclusion) scores (Fig. 6A), suggesting a significant link between PKHD1 + hepatocytes and the immune microenvironment.

Somatic mutation analysis was performed on the available mutation data from 335 samples. TP53 was the most frequently mutated gene in HCC, with a mutation rate of 28%, followed by CTNNB1 (25%), with Missense\_Mutation being the predominant mutation type for both (Fig. 6B, C). The frequencies of TP53, TG, IDH1, and TAF1L mutations were lower in the high PKHD1 + hepatocyte abundance group compared to the low-abundance group, whereas the CTNNB1 mutation frequency was higher in the high-abundance group (Fig. 6D and E).

### 3.4 Identification of SF3B4 as a core gene associated with PKHD1 + hepatocytes

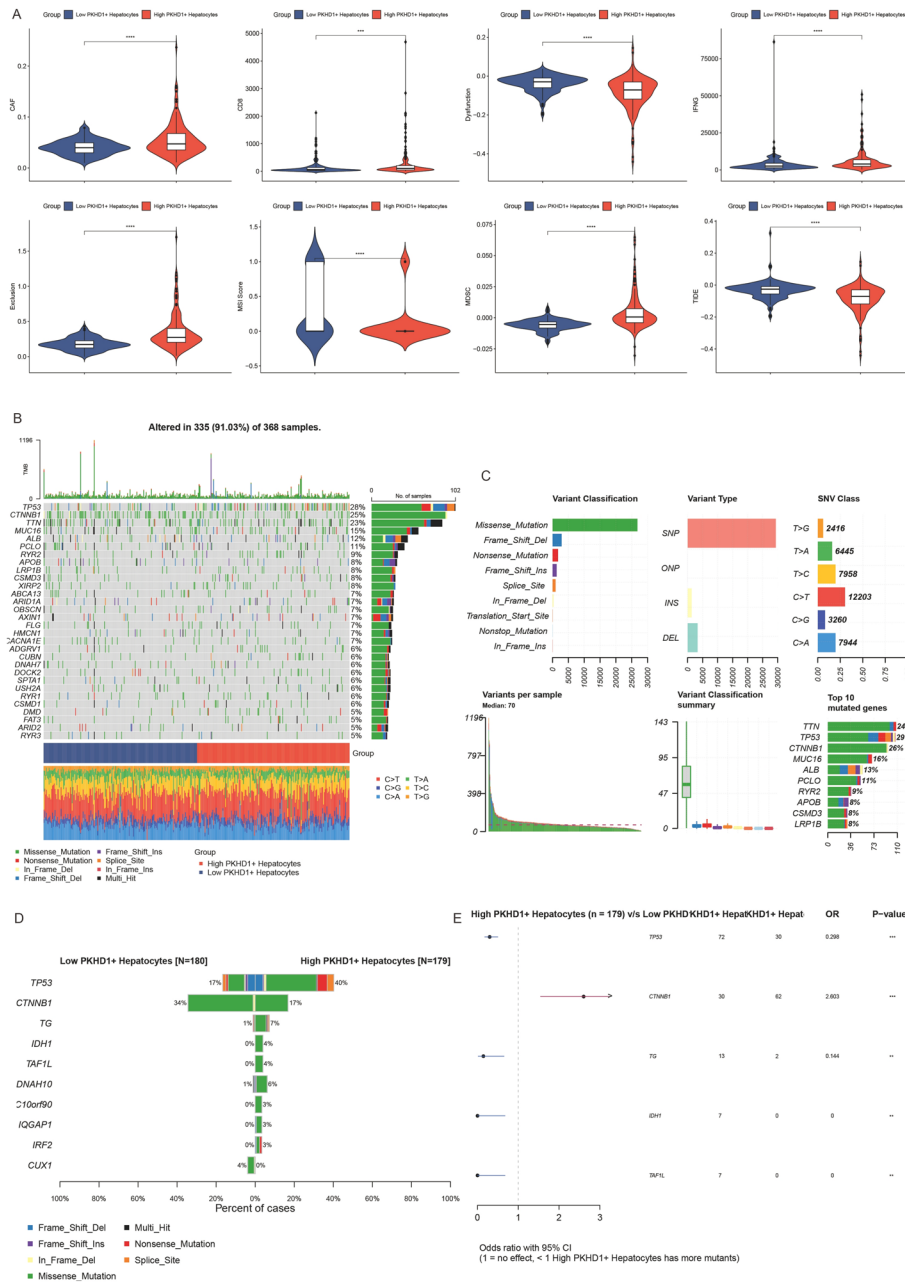
Given the potential important role of PKHD1 + hepatocytes in HCC, this study aimed to identify core genes associated with this cell population. Differential expression analysis revealed 8,636 upregulated genes and 1,659 downregulated genes in HCC tissues compared to adjacent non-tumor tissues (Fig. 7A). Weighted Gene Co-expression Network Analysis (WGCNA) clustered genes from the TCGA-LIHC dataset into 16 co-expression modules based on their expression profiles (Fig. 7B, C). Correlation analysis indicated that the Blue module was significantly positively correlated with both the PKHD1 + hepatocyte proportion ( $\text{cor} = 0.66$ ) and HCC status ( $\text{cor} = 0.38$ ) (Fig. 7C,  $p < 0.05$ ).

The Blue module contained 2,876 genes. Taking the intersection of these 2,876 genes, the 8,636 genes upregulated in HCC tissues from the bulk transcriptome analysis, and the marker genes specific to PKHD1 + hepatocytes from the single-cell data yielded 12 candidate genes: GLS, DDX39B, SF3B4, ACBD6, SUCO, ANXA2, SIPA1L3, CBFA2T2, SOX4, SOX9, DNAH14, and LAMC1 (Fig. 7D).

Survival analysis showed that high expression of GLS, SF3B4, ACBD6, SUCO, ANXA2, CBFA2T2, SOX4, DNAH14, and LAMC1 was associated with poor prognosis in HCC patients (Fig. 8A). Univariate Cox analysis indicated that 11 of these 12 genes (all except DDX39B) were risk factors for poor patient prognosis (Fig. 8B). Subsequent multivariate Cox analysis of these 12 genes demonstrated that SF3B4 and ACBD6 were independent prognostic factors. Notably, after adjustment in the multivariate model, ACBD6 transitioned from being a high-risk factor to a low-risk factor (Fig. 8C). Therefore, we selected SF3B4 as the primary focus for further investigation in this study.

### 3.5 SF3B4 is upregulated in HCC and promotes HCC progression

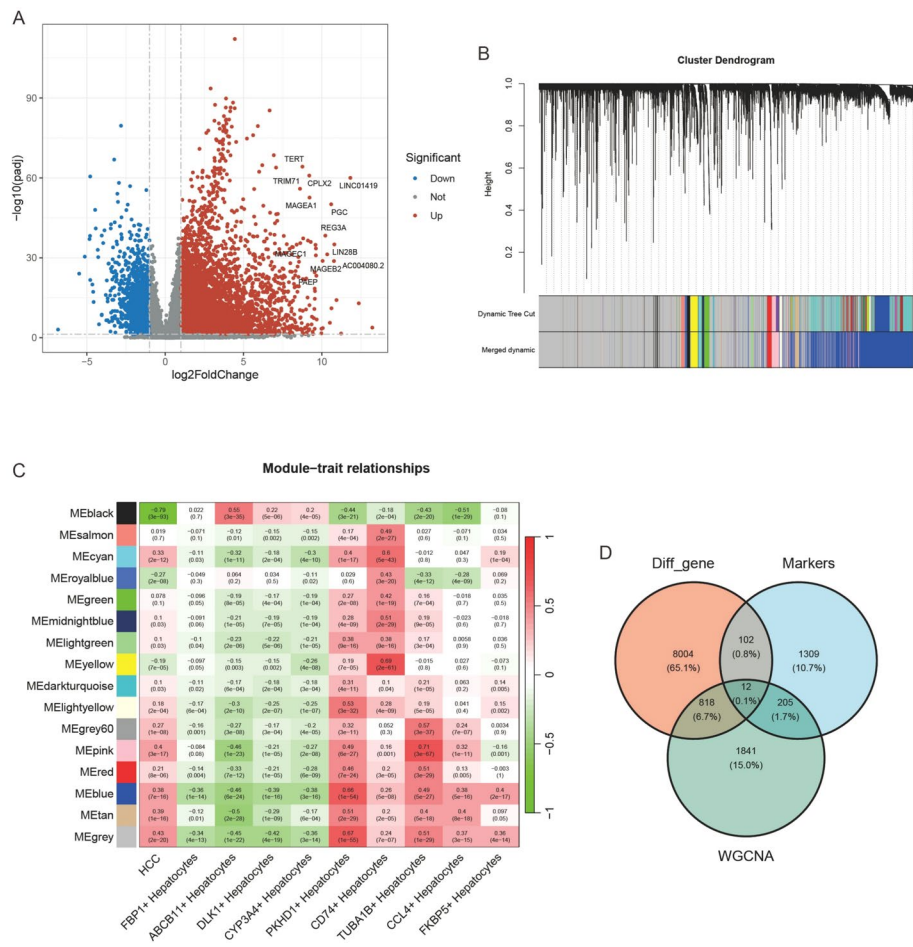
The Mann–Whitney U test was employed to compare the expression levels of SF3B4 in HCC and adjacent non-tumor tissues within the TCGA-LIHC dataset. The results indicated a significant upregulation of SF3B4 in the HCC group (Fig. 9A). Consistent with



**Fig. 6** Association analysis of PKHD1<sup>+</sup> hepatocytes with the immune microenvironment and somatic mutations in HCC. **A** Comparison of immune cell infiltration scores between high and low PKHD1<sup>+</sup> hepatocyte abundance groups. **B, C** Landscape of somatic mutations in the TCGA-LIHC cohort. **D, E** Comparison of somatic mutation frequencies between high and low PKHD1<sup>+</sup> hepatocyte abundance groups

this, results from qPCR and Western blot analysis confirmed the significant upregulation of SF3B4 at both the mRNA and protein levels in HCC tissues (Fig. 9B, C).

scRNA-seq analysis indicated that SF3B4 expression was predominantly localized to PKHD1-positive hepatocytes (Fig. 9D, E). To further investigate the impact of SF3B4 on HCC progression, we established stable cell lines with SF3B4 knockdown (Fig. 10A, B). The colony formation assay demonstrated that downregulation of SF3B4 expression led to a decrease in the proliferative capacity of HCC cells compared to the control group (Fig. 10C). Furthermore, Transwell migration and wound healing assays showed



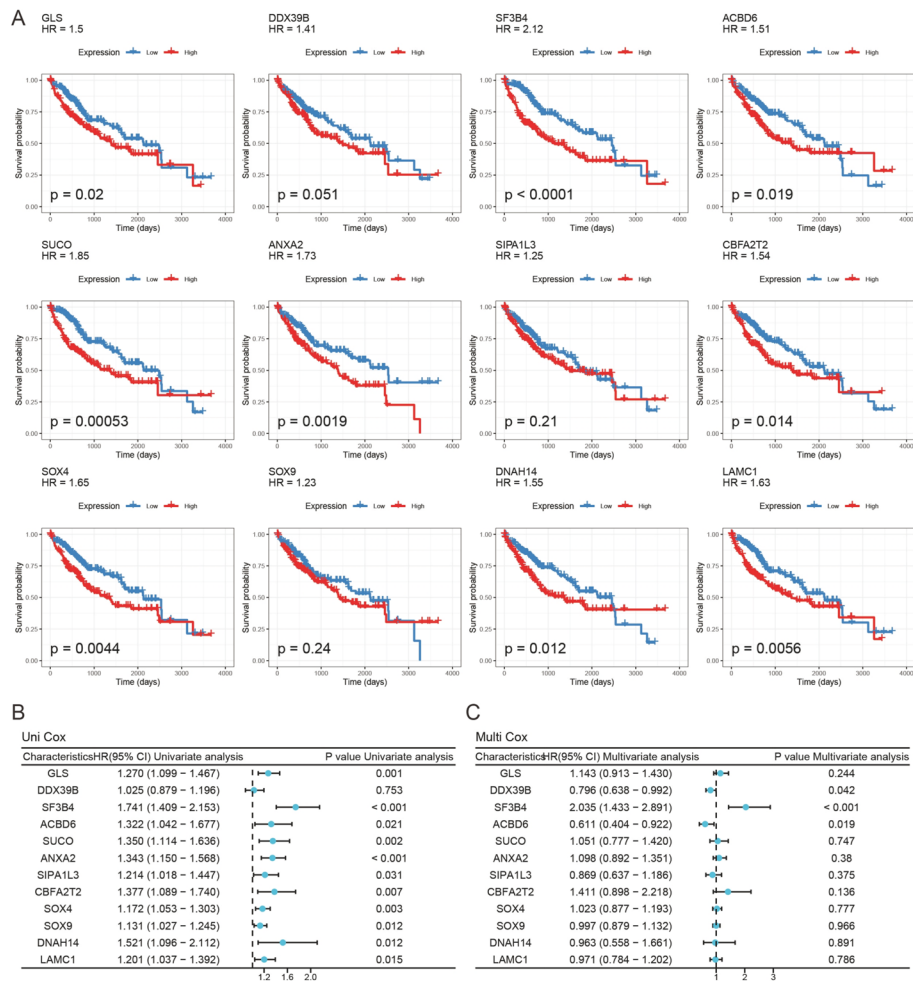
**Fig. 7** Screening of PKHD1\* hepatocyte-associated genes via differential expression analysis and WGCNA. **A** Volcano plot of differentially expressed genes between HCC and paracancerous tissues. **B** Cluster dendrogram of co-expression modules from WGCNA. **C** Heatmap of module-trait associations. **D** Venn diagram showing the intersection of genes from the Blue module, HCC-upregulated DEGs, and PKHD1\* hepatocyte marker genes

that suppressing SF3B4 expression resulted in reduced migratory ability of HCC cells (Fig. 10D, E). Collectively, these results indicate that SF3B4 is upregulated in HCC and promotes its progression.

#### 4 Discussion

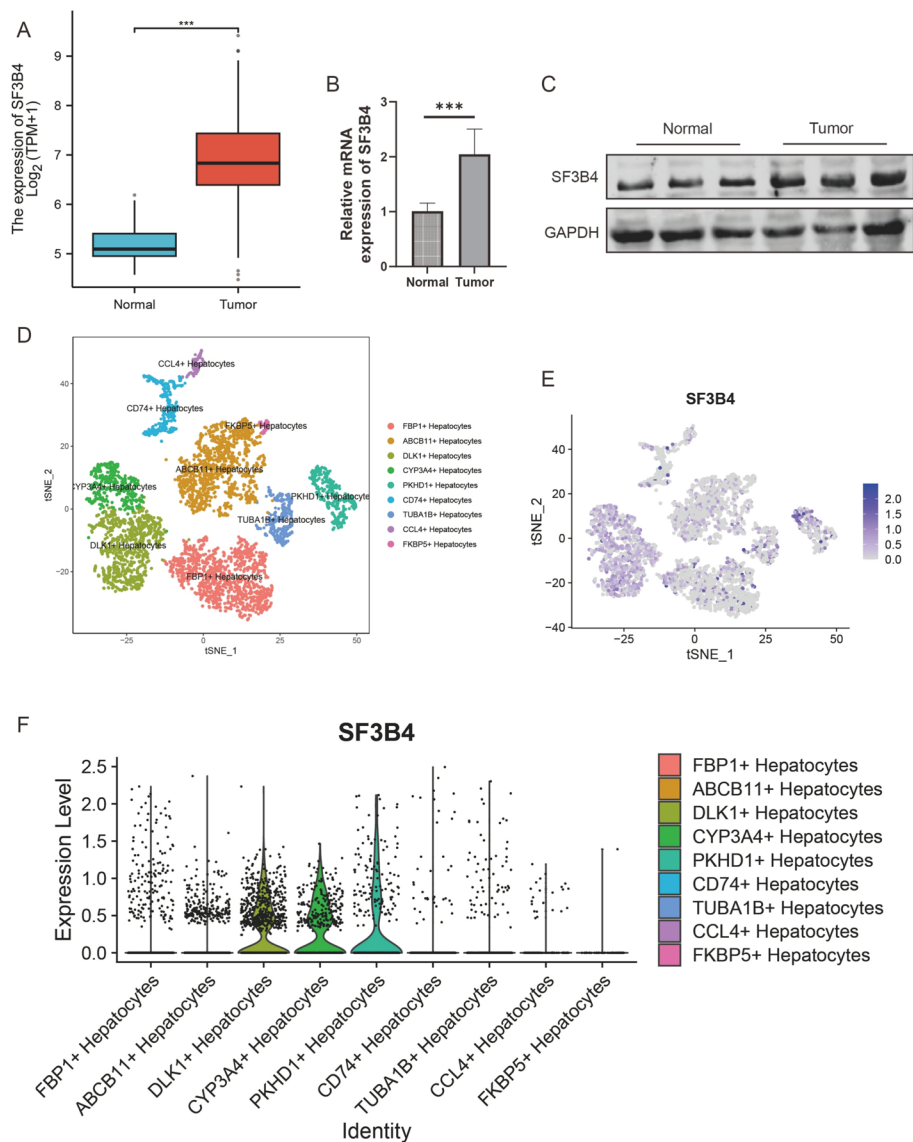
By integrating single-cell transcriptome sequencing, GWAS data, and bulk transcriptome data, this study systematically identified, for the first time, the PKHD1+ hepatocyte subpopulation as a core regulatory subset within the heterogeneity of HCC. We further discovered that the SF3B4 gene serves as an independent prognostic factor associated with this subpopulation. Subsequent in vitro experiments confirmed that SF3B4 promotes HCC progression by enhancing tumor cell proliferation and migration. These findings not only address a gap in the functional annotation and regulatory mechanisms of hepatocyte subtypes in liver cancer but also provide a new theoretical foundation and a potential therapeutic target for the precise subtyping, prognostic assessment, and targeted therapy of HCC.

The high heterogeneity of HCC is a critical bottleneck leading to poor treatment response and unfavorable prognosis. As the core functional cells within the tumor



**Fig. 8** Prognostic analysis of the 12 candidate genes. **A** Kaplan–Meier OS curves for the 12 genes. **B** Forest plot of univariate Cox regression analysis for the candidate genes. **C** Forest plot of multivariate Cox regression analysis for the candidate genes

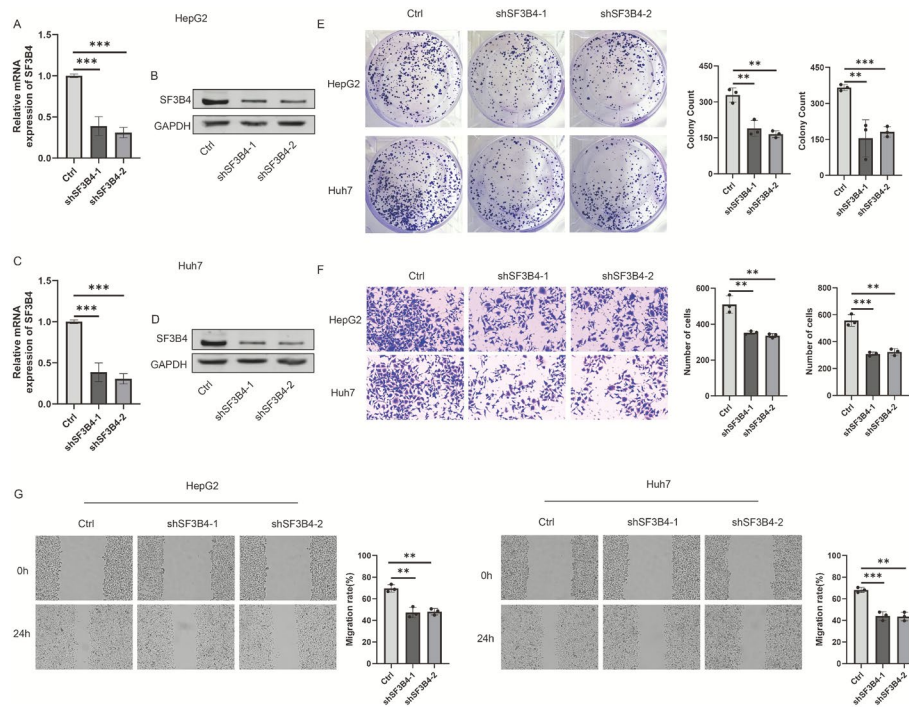
tissue, the diversity of hepatocyte subtypes and their regulatory roles in disease progression have been a major focus of research [24]. Existing studies utilizing single-cell transcriptomics have revealed the presence of multiple hepatocyte subtypes with distinct metabolic features and proliferative capacities within HCC tissues. However, these investigations have primarily focused on characterizing the molecular profiles of these subtypes and have yet to identify a "functional core subtype" with direct relevance to clinical prognosis and the immune microenvironment [25, 26]. The protein fibrocystin/polyductin, encoded by the PKHD1 (Polycystic Kidney and Hepatic Disease 1) gene, plays a critical role in various cellular functions, particularly in the development and functional maintenance of the kidneys and liver. In recent years, studies have revealed that mutations in PKHD1 are not only associated with polycystic kidney disease but also play significant roles in the initiation and progression of various cancers [27]. In colorectal cancer (CRC), PKHD1 is identified as one of the frequently mutated genes, alongside APC, TP53, and KRAS. Studies have shown that the expression of PKHD1 is regulated by the Wnt/ $\beta$ -catenin signaling pathway, which is frequently dysregulated in CRC progression. Analysis of 3702 colon cancer samples revealed that PKHD1 mutations are significantly associated with increased tumor mutational burden (TMB),



**Fig. 9** SF3B4 is upregulated in HCC. **A** Box plot showing SF3B4 expression in the TCGA-LIHC cohort. **B, C** Validation of SF3B4 upregulation at the mRNA (**B**) and protein (**C**) levels. **D** t-SNE plot showing SF3B4 expression across hepatocyte subtypes. **E** Violin plots showing SF3B4 expression across hepatocyte subtypes

elevated microsatellite instability (MSI), and reduced chromosomal instability (CIN) scores [28]. Furthermore, PKHD1 has also been shown to be associated with tumor progression in intrahepatic cholangiocarcinoma (ICC). For instance, knockdown of PKHD1 significantly enhanced the proliferation, migration, and invasion abilities of ICC cells by activating the Notch1 pathway and the expression of epithelial-mesenchymal transition (EMT)-related proteins [29]. These findings reveal the complex regulatory mechanisms of PKHD1 in various cancers, providing a theoretical foundation for developing targeted therapeutic strategies against PKHD1 and its associated pathways.

Through scPagwas analysis integrating single-cell data with GWAS-based genetic susceptibility information, this study revealed for the first time that the PKHD1<sup>+</sup> hepatocyte subpopulation exhibits a significantly higher TRS (Trait Relevance Score) compared to other hepatocyte subtypes and shows a positive correlation with HCC. Further



**Fig. 10** In vitro functional assays for SF3B4. **A, B** Validation of SF3B4 knockdown efficiency by qPCR (**A**) and Western blot (**B**) in HepG2 cells. **C, D** Validation of SF3B4 knockdown efficiency by qPCR (**C**) and Western blot (**D**) in Huh7 cells. **E** Colony formation assay results. **F** Transwell migration assay results. **G** Wound healing assay results and quantification of migration rates

quantification in the TCGA-LIHC cohort using BayesPrism confirmed that a high abundance of this subtype is significantly associated with poorer patient prognosis ( $P < 0.05$ ). Moreover, variations in its abundance are closely linked to immune microenvironment remodeling—specifically, increased infiltration of cancer-associated fibroblasts (CAFs) and myeloid-derived suppressor cells (MDSCs) was observed in the high-abundance group. This suggests that PKHD1<sup>+</sup>hepatocytes may contribute to unfavorable prognosis by attenuating anti-tumor immune activity [30–32]. This finding transcends the limitations of traditional studies that merely describe subtype differences, defining PKHD1<sup>+</sup> hepatocytes as a "prognosis-associated core subtype" and providing a cell-level marker for prognostic stratification in HCC. Furthermore, this study revealed a significant association between the abundance of PKHD1<sup>+</sup>hepatocytes and the somatic mutation profile in liver cancer: the high-abundance group showed a decreased frequency of TP53 mutations but an increased frequency of CTNNB1 mutations. TP53 is the most frequently mutated tumor suppressor gene in HCC. Its mutations not only impair its tumor-suppressive function but may also confer gain-of-function (GOF) properties that promote cancer progression [33]. This association suggests that the enrichment of PKHD1<sup>+</sup> hepatocytes may be linked to the mutational evolutionary pathway of hepatocellular carcinoma—potentially by suppressing TP53 mutation-associated malignant proliferation pathways or by maintaining a "low-malignancy phenotype" in hepatocytes through CTNNB1-regulated differentiation pathways. This provides a new perspective for deciphering the interactive mechanisms between HCC mutations and the evolution of cellular subtypes. After establishing the central role of the PKHD1<sup>+</sup> hepatocyte subpopulation, this study employed WGCNA co-expression analysis, differential gene

intersection screening, and multivariate Cox regression to ultimately identify SF3B4 as the core regulatory gene associated with this subtype. As a core subunit of the U2-type spliceosome, SF3B4 plays a key role in regulating gene expression and alternative splicing [34]. In HCC, the expression level of SF3B4 is significantly upregulated and is closely associated with tumor proliferation and metastasis [35]. Studies have revealed that SF3B4 promotes HCC cell proliferation and inhibits apoptosis through its interaction with TRIM28 and SETD5, indicating that SF3B4 functions not merely as a splicing factor in HCC but also facilitates tumor progression by regulating gene expression and alternative splicing events [36]. Furthermore, SF3B4 shows potential for the early diagnosis of HCC. Integrated analysis of transcriptomic and clinicopathological data from multi-stage HCC tissues identified SF3B4 as a biomarker for early-stage HCC. It can be detected in precancerous lesions, and its diagnostic performance surpasses that of commonly used HCC diagnostic markers [37]. This finding provides a new direction for the early detection of HCC and may improve patient survival rates. In terms of treatment, the overexpression of SF3B4 is closely associated with HCC cell survival and tumor progression. Through genome-wide CRISPR screening, SF3B4 has been identified as an essential survival gene in HCC, and the splicing landscape it regulates reveals its critical role in HCC cell survival and tumor progression [38]. Moreover, the expression of SF3B4 is associated with increased DNA copy number, and this overexpression is significantly correlated with poor prognosis in HCC, suggesting that SF3B4 may promote HCC progression through DNA copy number amplification [39]. This study provides multidimensional evidence confirming that SF3B4 is upregulated in HCC tissues and promotes HCC progression. Unlike previous research, these findings establish for the first time a regulatory link between SF3B4 and a core hepatocyte subtype (PKHD1<sup>+</sup>) in liver cancer, and validate its clinical value as a dual "gene-cell" target—serving not only as a molecular marker for prognostic assessment but also as a potential therapeutic target for intervening in the malignant phenotype associated with PKHD1<sup>+</sup> hepatocytes.

The findings of this study hold significant implications for clinical practice in HCC. Firstly, in terms of prognostic assessment, the combination of PKHD1<sup>+</sup> hepatocyte abundance and SF3B4 expression levels can serve as a "cellular-molecular" dual biomarker, addressing the limitations of traditional markers like AFP and imaging evaluations. For instance, in AFP-negative cases or early-stage micro-HCC, detecting SF3B4 expression and PKHD1<sup>+</sup> hepatocyte proportions could refine prognostic stratification. It is worth mentioning that combining SF3B4, PKHD1<sup>+</sup> hepatocytes with other imaging indicators, such as microvascular invasion, may serve as a better strategy for future prognostic prediction [40]. Secondly, regarding treatment strategies, the pro-oncogenic function of SF3B4 suggests its potential as a novel therapeutic target. Future research could explore combining SF3B4 inhibitors with existing therapies (e.g., TACE, PD-1 inhibitors), particularly for high-risk patients with high SF3B4 expression and low PKHD1<sup>+</sup> hepatocyte abundance. Such strategies may suppress malignant phenotypes and enhance responses to immunotherapy.

However, this study has limitations: (1) In vitro validation was limited to two HCC cell lines (HepG2 and Huh7), lacking verification in other subtype-specific models. The absence of in vivo experiments (e.g., xenograft tumor formation in nude mice) necessitates further animal studies to confirm SF3B4's tumor-promoting role; (2) The clinical tissue sample size was relatively small, potentially introducing bias. Expanding

multi-center cohorts is critical to validate SF3B4's expression patterns and prognostic value; (3) The specific molecular pathways through which SF3B4 regulates PKHD1<sup>+</sup> hepatocyte functions and HCC progression remain unclear. Future research needs to further explore the regulatory mechanisms of SF3B4 expression in the PKHD1<sup>+</sup> subpopulation. Elucidating these mechanisms will provide precise targets for future interventions.

## 5 Conclusion

This study, focusing on the heterogeneity of HCC, employed an integrated multi-omics strategy to identify for the first time the PKHD1<sup>+</sup> hepatocyte subpopulation as a prognosis-associated core subtype. We established SF3B4 as an independent prognostic factor and oncogene linked to this subtype, and preliminarily elucidated its associations with the immune microenvironment and somatic mutations. These findings not only enhance the understanding of the "cell–gene–immune" regulatory network underlying HCC heterogeneity but also provide translatable targets for precise prognostic assessment and targeted therapy, offering new research directions to overcome current therapeutic bottlenecks in HCC.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1007/s12672-026-04501-7>.

Additional file 1.

## Author contributions

Fujun Ma and Lihong Kang contributed equally to this work. Fujun Ma led the study design, data analysis, and manuscript drafting. Lihong Kang performed bioinformatics analyses including scRNA-seq and scPagwas integration. Zhijian Ren and Yang Yang conducted experimental validations (RT-qPCR, Western blot, functional assays). Tong Shen assisted with data curation and statistical analysis. Haibo Yu supervised the project, provided resources, and revised the manuscript. All authors reviewed and approved the final manuscript.

## Funding

This study did not receive any funding.

## Data availability

The datasets analysed during the current study are available in the IEU-Open GWAS (<https://opengwas.io/datasets/bbj-a-158>), GEO database (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE282701>) and TCGA database (<https://www.cancer.gov/ccg/research/genome-sequencing/tcga>).

## Declarations

### Ethics approval and consent to participate

The experimental protocol of this study was approved by the Ethics Committee of Henan Medical University in accordance with the World Medical Association Declaration of Helsinki and the relevant ethical guidelines and regulations for medical research involving human participants. All procedures involving human tissue samples were conducted strictly following the approved protocol and ethical standards of the aforementioned committee. Informed written consent was obtained from all participants prior to sample collection, and all personal identifying information of participants was de-identified to ensure privacy protection.

### Competing interests

The authors declare no competing interests.

Received: 3 November 2025 / Accepted: 20 January 2026

Published online: 14 February 2026

## References

1. Wang SY, Yin L, Wang C, Ma MP. Atypical magnetic resonance imaging features and differential diagnosis of hepatocellular carcinoma. *J Int Med Res.* 2020;48(10):300060520943415.
2. Vogel A, Meyer T, Sapich G, Salem R, Saborowski A. Hepatocellular carcinoma. *Lancet.* 2022;400(10360):1345–62.

3. Llovet JM, Kelley RK, Villanueva A, Singal AG, Pikarsky E, Roayaie S, et al. Hepatocellular carcinoma. *Nat Rev Dis Primers*. 2021;7(1):6.
4. Feng T, Yang X, Wang Q, Liu X. Hepatocellular carcinoma-derived exosomes: key players in intercellular communication within the tumor microenvironment. *Sichuan Da Xue Xue Bao Yi Xue Ban*. 2024;55(1):6–12.
5. Li YK, Wu S, Wu YS, Zhang WH, Wang Y, Li YH, et al. Portal venous and hepatic arterial coefficients predict post-hepatectomy overall and recurrence-free survival in patients with hepatocellular carcinoma: a retrospective study. *J Hepatocell Carcinoma*. 2024;11:1389–402.
6. Xiang J, Li Y, Mei S, Ou Z, Wang L, Ke Y, et al. Novel diagnostic and therapeutic strategies based on PANoptosis for hepatocellular carcinoma. *Cancer Biol Med*. 2025;22(8):928–39.
7. Lin MT, Wang CC, Cheng YF, Eng HL, Yen YH, Tsai MC, et al. Comprehensive comparison of multiple-detector computed tomography and dynamic magnetic resonance imaging in the diagnosis of hepatocellular carcinoma with varying degrees of fibrosis. *PLoS ONE*. 2016;11(11):e0166157.
8. Gao N, Wang D, Ma X, Lv F, Ren X. Contrast-enhanced US and contrast-enhanced CT for diagnosis of focal liver lesions in liver transplant recipients: a comparative study. *Iliver*. 2025;4(1):100147.
9. Yu B, Ma W. Biomarker discovery in hepatocellular carcinoma (HCC) for personalized treatment and enhanced prognosis. *Cytokine Growth Factor Rev*. 2024;79:29–38.
10. Moris D, Martinino A, Schiltz S, Allen PJ, Barbas A, Sudan D, et al. Advances in the treatment of hepatocellular carcinoma: an overview of the current and evolving therapeutic landscape for clinicians. *CA Cancer J Clin*. 2025. <https://doi.org/10.3322/caac.70018>.
11. Llovet JM, Castet F, Heikenwalder M, Maini MK, Mazzaferro V, Pinato DJ, et al. Immunotherapies for hepatocellular carcinoma. *Nat Rev Clin Oncol*. 2022;19(3):151–72.
12. Wang QB, Li J, Zhang ZJ, Li YK, Liang YB, Chen XM, et al. The effectiveness and safety of therapies for hepatocellular carcinoma with tumor thrombus in the hepatic vein, inferior vena cava and/or right atrium: a systematic review and single-arm meta-analysis. *Expert Rev Anticancer Ther*. 2025;25(5):561–70.
13. Kwon H, Kang E, Kim S, Baek Y, Bark I, Cho J. Predicting prognosis prior to the combination of atezolizumab and bevacizumab on unresectable HCC: analysis and comparison of tumor heterogeneity at CT and Gd-EOB-DTPA hepatobiliary MR imaging. *Medicine (Baltimore)*. 2024;103(49):e40769.
14. Sharma A, Merritt E, Hu X, Cruz A, Jiang C, Sarkodie H, et al. Non-genetic intra-tumor heterogeneity is a major predictor of phenotypic heterogeneity and ongoing evolutionary dynamics in lung tumors. *Cell Rep*. 2019;29(8):2164–2174.e2165.
15. Hladý RA, Zhou D, Puszky W, Roberts LR, Liu C, Robertson KD. Initiation of aberrant DNA methylation patterns and heterogeneity in precancerous lesions of human hepatocellular cancer. *Epigenetics*. 2017;12(3):215–25.
16. Villanueva A. Hepatocellular carcinoma. *N Engl J Med*. 2019;380(15):1450–62.
17. Shen Y, Bai X, Zhang Q, Liang X, Jin X, Zhao Z, et al. Oncolytic virus VG161 in refractory hepatocellular carcinoma. *Nature*. 2025;641(8062):503–11.
18. Ogul A, Birsengul I, Kelle AP, Solmaz AA, Coskun Y, Yetisir AE, et al. Importance of 18F-fluorodeoxyglucose positron emission tomography/computed tomography heterogeneity indices in non-small cell lung cancer. *Rev Assoc Med Bras (1992)*. 2025;71(5):e20241408.
19. Helal Tel A, Radwan NA, Shaker M. Extrahepatic metastases as initial manifestations of hepatocellular carcinoma: an Egyptian experience. *Diagn Pathol*. 2015;10:82.
20. Silva M, Coelho R, Rios E, Gomes S, Carneiro F, Macedo G. Breast metastasis from a combined hepatocellular-cholangiocarcinoma. *ACG Case Rep J*. 2019;6(4):e00057.
21. Boxer E, Feigin N, Tschernichovsky R, Darnell NG, Greenwald AR, Hoefflin R, et al. Emerging clinical applications of single-cell RNA sequencing in oncology. *Nat Rev Clin Oncol*. 2025;22(5):315–26.
22. Van de Sande B, Lee JS, Mutasa-Gottgens E, Naughton B, Bacon W, Manning J, et al. Applications of single-cell RNA sequencing in drug discovery and development. *Nat Rev Drug Discov*. 2023;22(6):496–520.
23. Wang K, Liu M, Wang HW, Jin KM, Yan XL, Bao Q, et al. Mutated DNA damage repair pathways are prognostic and chemosensitivity markers for resected colorectal cancer liver metastases. *Front Oncol*. 2021;11:643375.
24. Chan LK, Tsui YM, Ho DW, Ng IO. Cellular heterogeneity and plasticity in liver cancer. *Semin Cancer Biol*. 2022;82:134–49.
25. Ye M, Li X, Chen L, Mo S, Liu J, Huang T, et al. A high-throughput sequencing data-based classifier reveals the metabolic heterogeneity of hepatocellular carcinoma. *Cancers (Basel)*. 2023. <https://doi.org/10.3390/cancers15030592>.
26. Wang X, Li P, Ji H, Xu Z, Xing H. Single-cell transcriptomics reveals over-activated reactive oxygen species pathway in hepatocytes in the development of hepatocellular carcinoma. *Sci Rep*. 2024;14(1):29809.
27. Zheng C, Quan R, Xia EJ, Bhandari A, Zhang X. Original tumour suppressor gene polycystic kidney and hepatic disease 1-like 1 is associated with thyroid cancer cell progression. *Oncol Lett*. 2019;18(3):3227–35.
28. Han L, Gong F, Wu X, Tang W, Bao H, Wang Y, et al. Comprehensive characterization of PKHD1 mutation in human colon cancer. *Cancer Med*. 2024;13(1):e6796.
29. Shang T, Chen X, Xue H, Wu Y, Lin S, Zhu Y. The PKHD1 gene inhibits tumor proliferation and invasion in intrahepatic cholangiocarcinoma by activating the Notch pathway. *Int J Med Sci*. 2024;21(14):2655–63.
30. Matsuda M, Seki E. Hepatic stellate cell-macrophage crosstalk in liver fibrosis and carcinogenesis. *Semin Liver Dis*. 2020;40(3):307–20.
31. Kaps L, Schuppan D. Targeting cancer associated fibroblasts in liver fibrosis and liver cancer using nanocarriers. *Cells*. 2020. <https://doi.org/10.3390/cells9092027>.
32. Xie Y, Zhang Y, Wei X, Zhou C, Huang Y, Zhu X, et al. Jianpi huayu decoction attenuates the immunosuppressive status of H(22) hepatocellular carcinoma-bearing mice: by targeting myeloid-derived suppressor cells. *Front Pharmacol*. 2020;11:16.
33. Wang L, Yan K, He X, Zhu H, Song J, Chen S, et al. LRP1B or TP53 mutations are associated with higher tumor mutational burden and worse survival in hepatocellular carcinoma. *J Cancer*. 2021;12(1):217–23.
34. Wu T, Xiao Z, Su B, Yan Z, Zhao Y, Huang C, et al. Splicing factor 3b subunit 4 (SF3b4) is mediated by EP300 and CREBBP to promote colorectal cancer (CRC) proliferation by enhancing autophagy. *Am J Cancer Res*. 2025;15(6):2826–42.
35. Liu Z, Li W, Pang Y, Zhou Z, Liu S, Cheng K, et al. SF3B4 is regulated by microRNA-133b and promotes cell proliferation and metastasis in hepatocellular carcinoma. *EBioMedicine*. 2018;38:57–68.
36. Huang H, Fang Y, Li Z, Qu S, Yuan B, Gan K, et al. SF3B4 regulates proliferation and apoptosis in hepatocellular carcinoma via alternative splicing and interaction with TRIM28 and SETD5. *J Transl Med*. 2025;23(1):441.

37. Shen Q, Nam SW. SF3B4 as an early-stage diagnostic marker and driver of hepatocellular carcinoma. *BMB Rep.* 2018;51(2):57–8.
38. Guo Y, Xu M, Xue H, Ding X, Wong AM, Lin N, et al. Genome-wide CRISPR screen identifies splicing factor SF3B4 in driving hepatocellular carcinoma. *Sci Adv.* 2025;11(41):eadw7181.
39. Iguchi T, Komatsu H, Masuda T, Nambara S, Kidogami S, Ogawa Y, et al. Increased copy number of the gene encoding SF3B4 indicates poor prognosis in hepatocellular carcinoma. *Anticancer Res.* 2016;36(5):2139–44.
40. Wang QB, Luo WL, Li YK, Li J, Yang ZS, Zhao K, et al. Tumor compression of the hepatic or portal vein predicts the presence of microvascular invasion and satellite nodules in hepatocellular carcinoma: a retrospective study. *J Hepatocell Carcinoma.* 2025;12:2055–67.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.