Original Article
Exploring potential key genes and pathways associatedwith hepatocellular carcinoma prognosis through bioinformatics analysis, followed by experimental validation

Xi Chen[1], Jianhua Zhao[1], Jiaming Shu[1], Xueming Ying[1], Salman Khan[2], Sara Sarfaraz[3], Reza Mirzaeiebrahimabadi[4], Majid Alhomrani[5,6], Abdulhakeem S Alamri[5,6], Naif ALSuhaymi[7]

[1]Department of Oncology, Jingdezhen First People's Hospital, Jindezhen 333000, Jiangxi, China; [2]DHQ Teaching Hospital, GMC, Dikah, Pakistan; [3]Department of Bioinformatics, Faculty of Biomedical and Life Sciences, Kohsar University Murree, Pakistan; [4]The First Affiliated Hospital of Zhengzhou University, Zhengzhou University, Zhengzhou, Henan, China; [5]Department of Clinical Laboratories Sciences, The Faculty of Applied Medical Sciences, Taif University, Taif, Saudi Arabia; [6]Research Centre for Health Sciences, Taif University, Taif Saudi Arabia; [7]Department of Emergency Medical Services, Faculty of Health Sciences AlQunfudah, Umm Al-Qura University, Mekkah, Saudi Arabia

Abstract: Background: Liver Hepatocellular Carcinoma (LIHC) is a prevalent and aggressive liver cancer with limited therapeutic options. Identifying key genes involved in LIHC can enhance our understanding of its molecular mechanisms and aid in the development of targeted therapies. This study aims to identify differentially expressed genes (DEGs) and key hub genes in LIHC using bioinformatics approaches and experimental validation. Method: We analyzed two LIHC-related datasets, GSE84598 and GSE19665, from the Gene Expression Omnibus (GEO) database to identify DEGs. Differential expression analysis was performed using the limma package in R to identify DEGs between cancerous and non-cancerous liver tissues. A Protein-Protein Interaction (PPI) network was constructed using STRING to determine key hub genes. Further validation of these hub genes was conducted through UALCAN, OncoDB, and the Human Protein Atlas (HPA) databases for mRNA and protein expression levels. Promoter methylation and mutational analyses were performed using cBioPortal. Kaplan-Meier survival analysis assessed the impact of hub gene expression on patient survival. Correlations with immune cell abundance and drug sensitivity were explored using GSCA. Finally, AURKA was knocked down in HepG2 cells, and cell proliferation, colony formation, and wound healing assays were performed. Results: Analysis identified 180 DEGs, with four key hub genes, including AURKA, BUB1B, CCNA2, and PTTG1 showing significant overexpression and hypomethylation in LIHC tissues. AURKA knockdown in HepG2 cells led to decreased cell proliferation, reduced colony formation, and impaired wound healing, confirming its role in LIHC progression. These hub genes were also hypomethylated and their elevated expression correlated with poor overall survival. Conclusion: AURKA, BUB1B, CCNA2, and PTTG1 are crucial for LIHC pathogenesis and may serve as potential biomarkers or therapeutic targets. Our findings provide new insights into LIHC mechanisms and suggest promising avenues for future research and therapeutic development.

Keywords: LIHC, DEGs, hub genes, prognosis, treatment

Introduction

Liver hepatocellular carcinoma (LIHC) is the most prevalent form of primary liver cancer, and it remains a major global health challenge [1, 2]. As of 2023, LIHC accounts for approximately 905,000 new cases and 830,000 deaths annually, making it one of the leading causes of cancer-related mortality worldwide [3]. The high mortality rate associated with LIHC is largely due to the aggressive nature of the disease, late-stage diagnosis, and limited effective therapeutic options [1, 4]. Understanding the molecular mechanisms underlying

LIHC is critical for improving early detection and developing more effective treatments. LIHC typically develops in the context of chronic liver disease, with several key risk factors contributing to its onset [5, 6]. Chronic hepatitis B virus (HBV) and hepatitis C virus (HCV) infections are the leading causes of LIHC, accounting for approximately 80% of cases worldwide [7, 8]. These infections lead to chronic liver inflammation, fibrosis, and eventually cirrhosis, creating a conducive environment for malignant transformation [9]. In addition to viral hepatitis, other significant risk factors include excessive alcohol consumption, non-alcoholic fatty liver disease (NAFLD), and exposure to aflatoxins, which are toxic compounds produced by certain molds in food [10].

Over the years, extensive research has focused on identifying biomarkers that could serve as diagnostic, prognostic, or therapeutic targets in LIHC. Several biomarkers have been reported, including alpha-fetoprotein (AFP), which is commonly used for LIHC diagnosis, though its sensitivity and specificity are suboptimal [11, 12]. Other biomarkers such as glypican-3 (GPC3), osteopontin (OPN), and des-gamma-carboxy prothrombin (DCP) have shown promise in improving diagnostic accuracy [13]. Additionally, molecular markers like telomerase reverse transcriptase (TERT) mutations, and alterations in tumor suppressor genes such as TP53, have been linked to LIHC progression and prognosis [14].

Despite these advances, the heterogeneity of LIHC at the molecular level poses challenges in identifying universally applicable biomarkers. This has led to an increasing interest in utilizing gene expression profiling to discover novel biomarkers and potential therapeutic targets. High-throughput datasets, including GSE55092 [15] and GSE47197 [16] have been pivotal in this regard, offering insights into the differentially expressed genes (DEGs) associated with LIHC.

In our study, we conducted a thorough analysis of the GSE55092 and GSE47197 datasets to identify hub genes that play central roles in the molecular networks of LIHC. These hub genes, due to their critical positions within gene interaction networks, are hypothesized to be key drivers of LIHC pathogenesis. To validate their functional relevance, we performed a series of in vitro experiments, aiming to confirm their roles as potential biomarkers or therapeutic targets. This integrated approach not only enhances our understanding of LIHC but also contributes to the ongoing efforts to identify more effective strategies for early detection and treatment of this devastating disease.

Methodology

Dataset acquisition, differentially expressed genes (DEGs), and hub genes identification

Two datasets related to LIHC, including GSE84598 and GSE19665 were obtained from the Gene Expression Omnibus (GEO) database [17]. These datasets were selected based on their comprehensive expression profiles. The raw data from GSE84598 and GSE19665 were preprocessed for quality control and normalization. The preprocessing steps included the removal of low-quality probes and normalization using the RMA algorithm, ensuring that the data were suitable for downstream differential expression analysis. Differential expression analysis was performed on the processed datasets to identify DEGs between cancerous and non-cancerous liver tissues. The analysis was conducted using the limma package in R, where a p-value < 0.05 and |log2 fold change| > 1 were set as the criteria for significant differential expression. The DEGs identified in each dataset were then compared to identify common DEGs across the two datasets using Venn analysis [18]. The common DEGs identified from GSE84598 and GSE19665 were further analyzed to construct a Protein-Protein Interaction (PPI) network using STRING database [19]. The interaction data were obtained from the STRING database, with a confidence score > 0.7 considered for significant interactions. The PPI network was visualized using Cytoscape software (version 3.10.2). Hub genes within the PPI network were identified using the degree method [20] using Cytoscape. Top nodes with a high degree of connectivity were considered as hub genes, indicating their potential central role in the network and possible significance in LIHC pathology.

Expression analysis of the hub genes in LIHC using data from The Cancer Genome Atlas (TCGA)

UALCAN, OncoDB, and the Human Protein Atlas (HPA) are essential databases for cancer re-

search. UALCAN provides a comprehensive resource for analyzing cancer OMICS data, allowing researchers to explore gene expression, survival analysis, and promoter methylation in various cancers [21]. OncoDB offers a rich database for studying oncogenes and tumor suppressor genes, facilitating insights into cancer biology and potential therapeutic targets [22]. The Human Protein Atlas (HPA) focuses on protein expression profiles across different tissues, including cancerous and normal samples, aiding in the identification of biomarkers and understanding protein function in cancer [23]. Together, these databases offer powerful tools for understanding cancer mechanisms, identifying potential biomarkers, and exploring therapeutic targets, making them invaluable resources in oncology research. In the current work, UALCAN and OncoDB databases were used to analyze mRNA expression while the HPA database was used to analyze protein expression levels of the hub genes in normal and LIHC tissue samples.

Promoter methylation analysis of the hub genes in LIHC using data from TCGA

In this study, UALCAN [21] and OncoDB [22] databases were utilized for promoter methylation analysis of the hub genes in LIHC.

Mutational analysis of the hub genes

cBioPortal is a comprehensive, open-access database designed for exploring multidimensional cancer genomics data [24]. It provides tools for visualizing and analyzing large-scale datasets, including mutations, copy number alterations, and gene expression. Widely used in cancer research, cBioPortal aids in identifying genetic alterations and understanding their clinical implications across various cancer types. In our work, the cBioPortal database was utilized to analyze genetic mutations in hub genes across LIHC samples.

Survival analysis of the hub genes

Kaplan Meier (KM) plotter is an online tool designed for assessing the prognostic value of genes in various cancers using KM survival plots [25]. It allows researchers to explore the correlation between gene expression and patient survival outcomes by analyzing publicly available datasets. This tool supports multiple cancer types and provides valuable insights

into potential biomarkers for prognosis and treatment strategies. Herein, we used the KM plotter to perform the survival analysis of the hub genes in LIHC patients.

Correlation analysis of the hub genes with immune cell abundance and drug sensitivity

GSCA (Gene Set Cancer Analysis) is a powerful bioinformatics tool used to analyze gene sets in the context of cancer [26]. It integrates genomic, transcriptomic, and epigenetic data, enabling researchers to explore gene expression patterns, mutations, and pathway activities, aiding in the identification of key cancer-related genes and therapeutic targets. In this work, the GSCA database was utilized to explore the correlation of the hub genes with immune cells abundance and drug sensitivity in LIHC.

Gene enrichment analysis

DAVID (Database for Annotation, Visualization, and Integrated Discovery) is a bioinformatics tool designed to analyze large gene or protein lists [27]. It helps researchers interpret the biological meaning behind extensive datasets by providing functional annotations, gene ontology classifications, and pathway analysis. Herein, DAVID was utilized to perform gene enrichment analysis of the hub genes.

Cell culture

In our work, we purchased the following 9 LIHC cell lines: HepG2, Huh7, SNU-182, SNU-387, SNU-449, SNU-475, PLC/PRF/5, Hep3B, and MHCC97H, along with 5 normal liver cell lines: THLE-2, THLE-3, MIHA, PH5CH, and LO2 from the American Type Culture Collection (ATCC), USA. The cell lines were cultured under standard conditions at 37°C in a humidified atmosphere with 5% CO_2. We used Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin. The medium was changed every 2-3 days to maintain optimal growth conditions.

RNA extraction and real-time quantitative reverse transcription polymerase chain reaction (RT-qPCR)

After culturing cell lines, we extracted total RNA using the Thermo Fisher PureLink RNA Mini Kit (Cat. No. 12183018A), following the manufac-

Table 1. Primer sequences for AURKA, BUB1B, CCNA2, PTTG1, and GAPDH

Gene	Forward primer	Reverse primer
AURKA	TCCATCTTCCAGGAGGACCA	TCCAAGGCTCCAGAGATCCA
BUB1B	GTGGAAGAGACTGCACAACAGC	TCAGACGCTTGCTGATGGCTCT
CCNA2	AAGACGAGACGGGTTGCAC	CATGAATGGTGAACGCAGGC
PTTG1	GGGGTCTGGACCTTCAATCA	TTGTTTGAGGGGTCCCTTGG
GAPDH	GTCTTCCTGGGCAAGCAGTA	CTGGACAGAAACCCCACTTC

turer's protocol. The RNA concentration and purity were measured using a Nanodrop spectrophotometer, ensuring that the A260/A280 ratio was between 1.8 and 2.0 for high-quality RNA. For cDNA synthesis, we used the Thermo Fisher High-Capacity cDNA Reverse Transcription Kit (Cat. No. 4368814). We utilized 1 µg of RNA as input in a 20 µL reaction, which included random primers and reverse transcriptase. The reaction was performed in a thermal cycler with the following conditions: 25°C for 10 minutes, 37°C for 120 minutes, and 85°C for 5 minutes. The synthesized cDNA was then diluted 1:10 for subsequent RT-qPCR analysis. RT-qPCR was performed using the Thermo Fisher PowerUp SYBR Green Master Mix (Cat. No. A25742) on an Applied Biosystems QuantStudio 3 Real-Time PCR System. Each 20 µL reaction contained 2 µL of diluted cDNA, 10 µL of PowerUp SYBR Green Master Mix, 1 µL of forward and reverse primers (0.5 µM each), and 7 µL of nuclease-free water. Melting curve analysis was performed at the end of the PCR run to confirm the specificity of the amplified products. We analyzed the expression levels of AURKA, BUB1B, CCNA2, and PTTG1, with GAPDH serving as the housekeeping gene. Each sample was run in triplicate, and relative gene expression was calculated using the $2^{-\Delta\Delta Ct}$ method, normalizing the target gene expression to GAPDH and comparing it to the control group (normal liver cell lines). The primer sequences for the amplified genes were listed in **Table 1**.

AURKA knockdown in HepG2 cells

AURKA was knocked down in the HepG2 cell line using Thermo Fisher Silencer Select siRNA targeting AURKA (Cat. No. 4392420) for this purpose. HepG2 cells were seeded in 6-well plates at a density of 2 × 10^5 cells per well and allowed to reach 60-70% confluency. Transfection was performed using Lipofecta-

mine RNAiMAX Transfection Reagent (Cat. No. 13778150), following the manufacturer's protocol. Briefly, 10 nM of siRNA was mixed with Lipofectamine RNAiMAX in Opti-MEM Reduced Serum Medium (Cat. No. 31985062) and incubated for 15 minutes at room temperature. The siRNA-Lipofectamine complex was then added to the cells, and the medium was changed after 6 hours. The cells were incubated for 48 hours before proceeding with further analyses.

RT-qPCR analysis post-knockdown

RNA was extracted from the siRNA-treated HepG2 cells using the Thermo Fisher PureLink RNA Mini Kit (Cat. No. 12183018A). The cDNA synthesis and RT-qPCR were carried out as described previously using the Thermo Fisher High-Capacity cDNA Reverse Transcription Kit (Cat. No. 4368814) and PowerUp SYBR Green Master Mix (Cat. No. A25742).

Cell proliferation assay

Cell proliferation was assessed using the Thermo Fisher AlamarBlue Cell Viability Reagent (Cat. No. DAL1025). HepG2 cells transfected with AURKA siRNA were seeded in 96-well plates at a density of 5 × 10^3 cells per well. At 24, 48, and 72 hours post-transfection, 10 µL of AlamarBlue reagent was added to each well and incubated for 4 hours at 37°C.

Colony formation assay

Colony formation ability was evaluated by seeding 500 transfected HepG2 cells in 6-well plates. The cells were cultured for 10-14 days in a humidified atmosphere at 37°C with 5% CO_2, allowing for colony growth. The colonies were then fixed with 4% paraformaldehyde and stained with 0.5% crystal violet. The number of colonies containing more than 50 cells was counted under a microscope, and the results were compared between the AURKA knockdown and control groups.

Wound healing assay

The wound healing assay was performed to assess cell migration. HepG2 cells were seed-
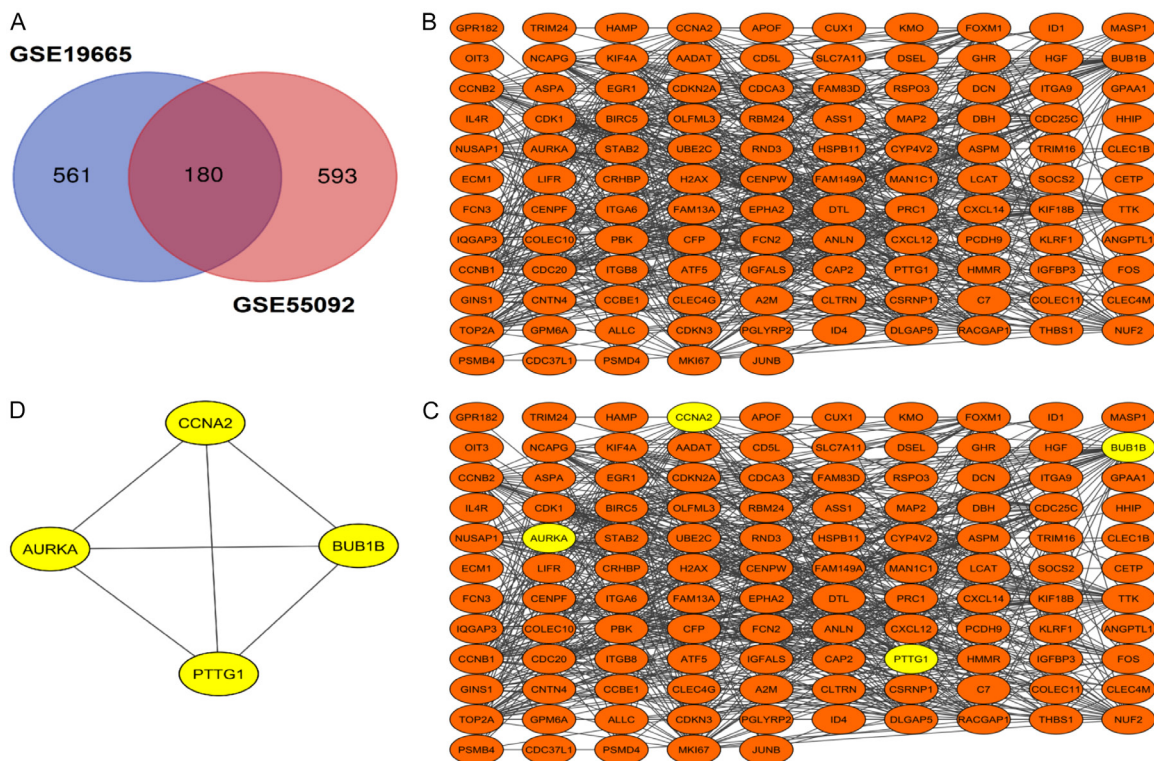
Figure 1. Identification of hub genes in liver hepatocellular carcinoma (LIHC) using differential expression analysis and network analysis. A. Venn diagram showing the overlap of differentially expressed genes (DEGs) between the GSE19665 and GSE55092 datasets. B. Protein-protein interaction (PPI) network of DEGs derived from the GSE19665 and GSE55092 datasets. C. PPI network of DEGs and hub genes derived from the GSE19665 and GSE55092 datasets. D. Network of the four most significant hub genes identified from both datasets.

ed in 6-well plates and grown to 90% confluency. A straight scratch was made across the cell monolayer using a sterile 200 μL pipette tip. The cells were then washed with PBS to remove debris and incubated in serum-free medium. Images of the wound were captured at 0 and 24 hours post-scratch using a microscope. The wound area was measured using ImageJ software, and the percentage of wound closure was calculated to compare the migratory capacity between AURKA knockdown and control cells.

Statistics

All statistical analyses were performed using GraphPad Prism (version 10.3.1). For comparisons between the two groups, an unpaired two-tailed Student's t-test was used. For comparisons involving more than two groups, one-way ANOVA followed by Tukey's post-hoc test was applied to assess significant differences. Receiver operating characteristics (ROC) curve analysis was used to assess the diagnostic

abilities. The threshold for statistical significance was set at *p*-value < 0.05 or *p*-value.

Results

Identification of DEGs and hub genes in LIHC

Two gene expression profiles (GSE84598 and GSE19665) were analyzed using the limma package in R, yielding the top 1000 DEGs from each dataset. Subsequently, Venn diagram analysis identified 180 common DEGs across the two datasets (**Figure 1A**). A PPI network for these overlapping DEGs was constructed with STRING (**Figure 1B**). Node degree calculations were performed to assess the significance of DEGs within the PPI network, highlighting AURKA, BUB1B, CCNA2, and PTTG1 as key hub genes (**Figure 1C**, **1D**).

Expression analysis of the hub genes across LIHC and normal control samples

Figure 2 illustrates the expression levels and staining intensities of four hub genes (AURKA,
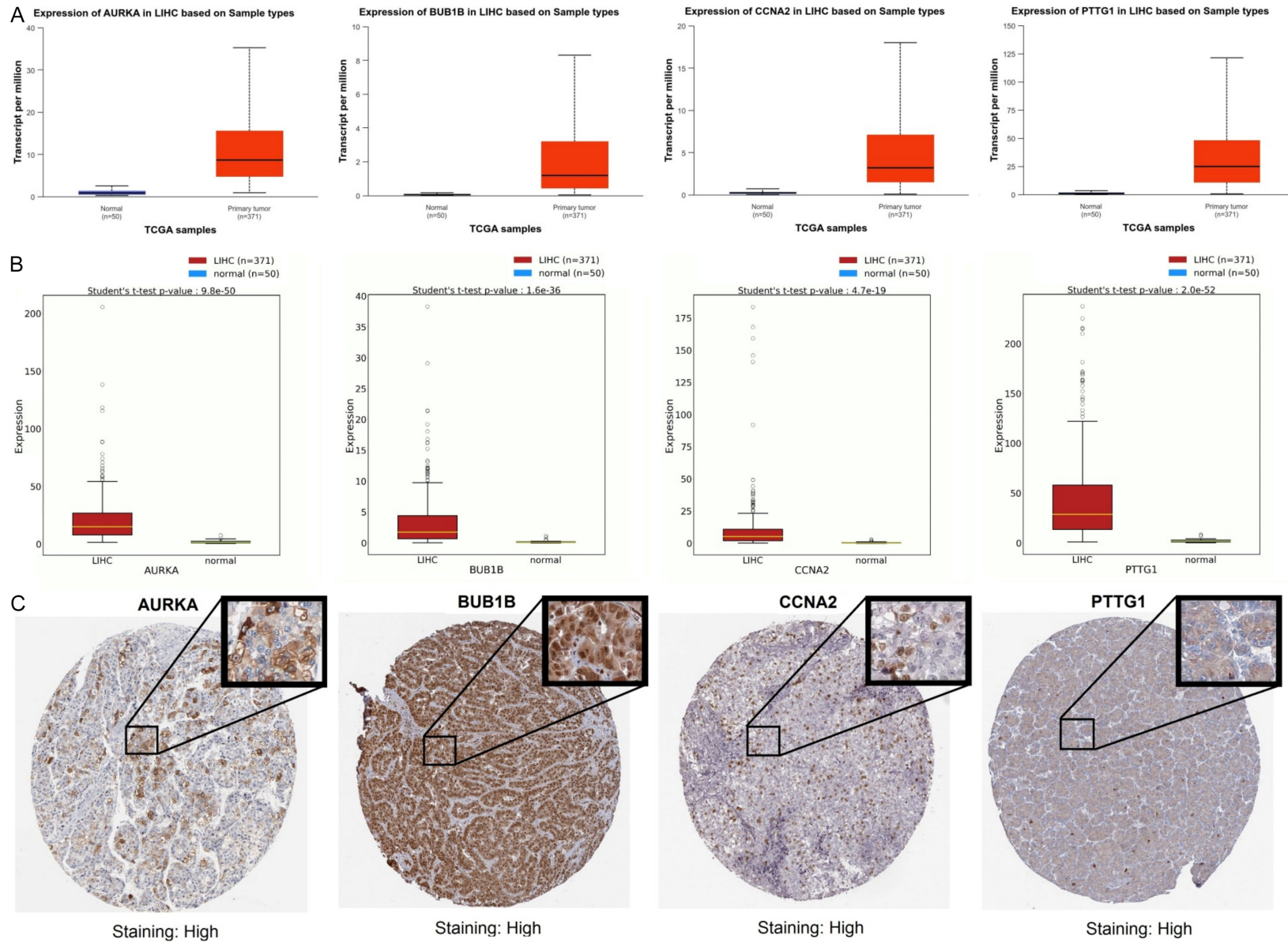
Figure 2. Analysis of hub gene mRNA and protein expression in liver hepatocellular carcinoma (LIHC). A. Box plots sourced from UALCAN displaying the transcript per million (TPM) levels of AURKA, BUB1B, CCNA2, and PTTG1 in LIHC primary tumors (n = 371) and normal tissues (n = 50). B. Box plots sourced from OncoDB showing the expression levels of AURKA, BUB1B, CCNA2, and PTTG1 in LIHC (n = 371) and normal tissues (n = 50). C. Immunohistochemistry images sourced from the Human Protein Atlas (HPA) showing high protein staining for AURKA, BUB1B, CCNA2, and PTTG1 in LIHC tissues. *P*-value < 0.05.

BUB1B, CCNA2, and PTTG1) in LIHC tissue samples based on data from TCGA. In **Figure 2A**, the box plots display the transcript per million (TPM) values for these genes in normal tissues compared to LIHC tissues via the UALCAN database. The expression levels of all four hub genes (AURKA, BUB1B, CCNA2, and PTTG1) are significantly (*p*-value < 0.05) higher in LIHC compared to normal tissues. **Figure 2B** further quantifies these differences using the OncoDB database, which confirms significant (*p*-value < 0.05) overexpression of AURKA, BUB1B, CCNA2, and PTTG1 in LIHC tissues. **Figure 2C** presents immunohistochemical staining images for these four genes from the HPA database, showing high staining intensity in LIHC tissues. This suggests that the proteins encoded by these genes are overexpressed in LIHC cells, consistent with the high transcript levels observed in the TCGA data. Furthermore, **Figure 3A** displays violin plots from the GEPIA2 database that illustrate the expression levels of hub genes across different cancer stages (I to IV). For AURKA, BUB1B, and CCNA2, the expression generally increases with advancing stages, indicating a potential correlation between higher expression levels and disease progression. PTTG1 shows a similar pattern, though with more variation in Stage III.

Promoter methylation analysis of the hub genes across LIHC and normal control samples
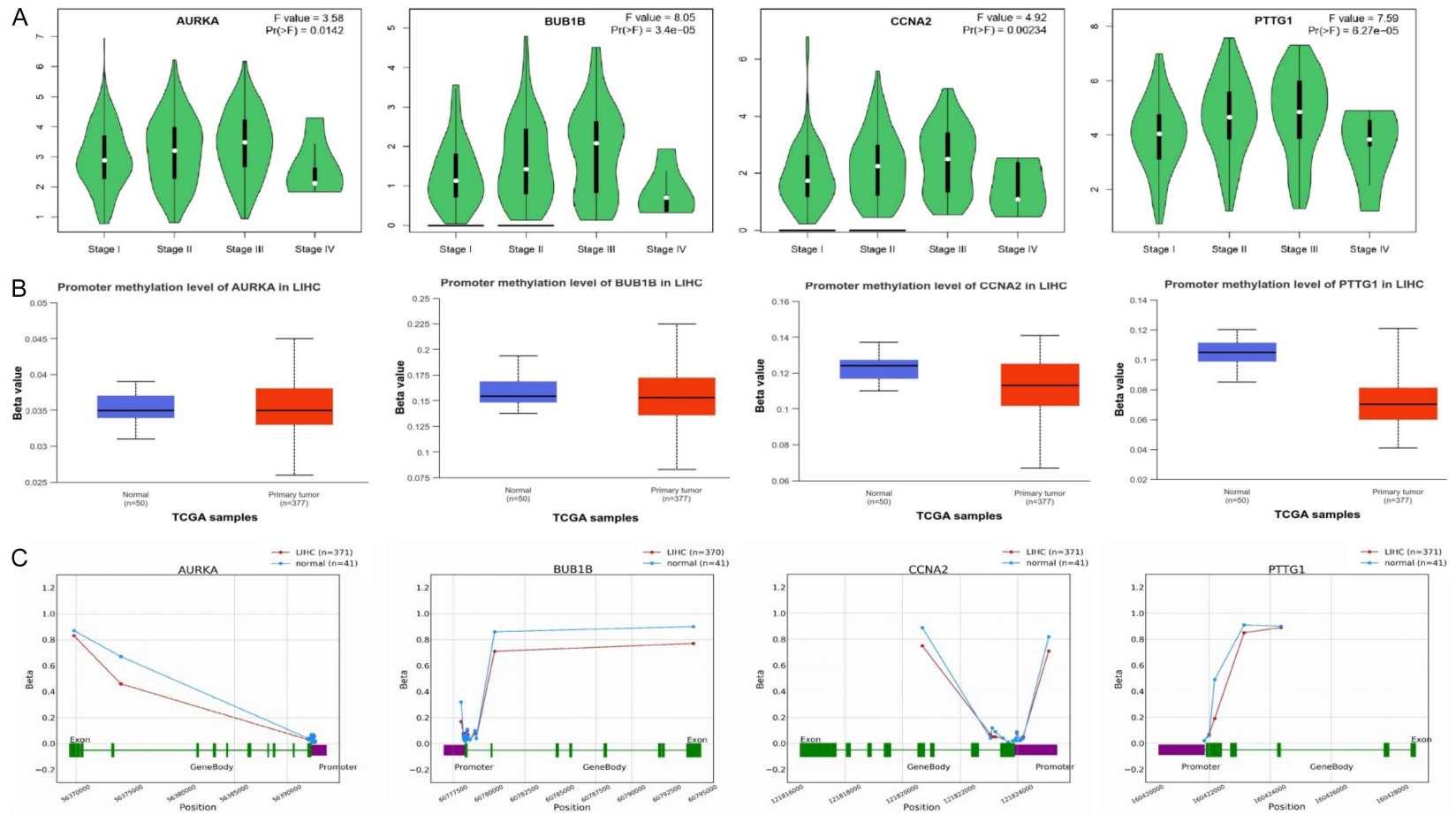
Promoter methylation analysis of the hub genes was conducted in regard to LIHC based on the data from TCGA using UALCAN and the OncoDB database. **Figure 3B** focuses on the promoter methylation levels of these genes in normal versus LIHC samples. The box plots indicate that promoter methylation levels are significantly (*p*-value < 0.05) reduced in LIHC tissues compared to normal tissues for AURKA, BUB1B, CCNA2, and PTTG1. **Figure 3C** presents more detailed methylation data, showing beta values across various positions within the promoter and gene body regions for AURKA, BUB1B,

CCNA2, and PTTG1. In each case, the promoter regions exhibit significantly (*p*-value < 0.05) lower methylation levels in tumor tissues (red line) compared to normal tissues (blue line). The decrease in methylation within the promoter regions of these genes in LIHC tissues supports the observation that hypomethylation is associated with their increased expression.

Mutational and survival analysis of the hub genes in LIHC

Mutational analysis of the hub genes in LIHC was conducted using the cBioPortal database. **Figure 4A** provides an overview of the mutation frequency of hub genes across 364 LIHC samples. AURKA is altered in 1% of the samples, specifically through missense mutations, whereas BUB1B and CCNA2 showed missense mutations in less than 1% of the samples and PTTG1 showed no detectable mutations in the cohort analyzed (**Figure 4A**). **Figure 4B** delves into the classification and types of variants found in AURKA, BUB1B, and CCNA2. All observed variants are classified as missense mutations and are single nucleotide polymorphisms (SNPs) (**Figure 4B**). The SNV (single nucleotide variant) class breakdown indicates that the most common base change is C>T, accounting for three instances, followed by other transitions and transversions (T>G, T>A, T>C) (**Figure 4B**). Furthermore, **Figure 4C** presents KM survival curves from KM plotter comparing overall survival in LIHC patients with high versus low expression levels of AURKA, BUB1B, CCNA2, and PTTG1. For all four genes, high expression correlates with significantly worse overall survival, as indicated by the hazard ratios (HR) and log-rank *p*-values: AURKA: HR = 1.77 (1.25-2.5), P = 0.0011, BUB1B: HR = 2.01 (1.42-2.86), P = 6.6e-05, CCNA2: HR = 1.92 (1.36-2.72), P = 0.00018, and PTTG1: HR = 2.14 (1.51-3.02), P = 1.1e-05. These data collectively suggest that genetic mutations in hub genes are relatively rare in LIHC, and their elevated expression levels are strongly associated with poor prognosis in LIHC.

Figure 3. Analysis of stage-specific expression and promoter methylation of hub genes in liver hepatocellular carcinoma (LIHC). A. Violin plots sourced from GEPIA2 showing the stage-specific expression levels of AURKA, BUB1B, CCNA2, and PTTG1 in LIHC patients across different stages (I-IV). B. Box plots sourced from UALCAN displaying the promoter methylation levels of AURKA, BUB1B, CCNA2, and PTTG1 in LIHC primary tumors (n = 377) compared to normal tissues (n = 50). C. Line graphs sourced from OncoDB showing the beta values of DNA methylation sites in the promoter and gene body regions of AURKA, BUB1B, CCNA2, and PTTG1 in LIHC (n = 371) compared to normal tissues (n = 41). P-value < 0.05.
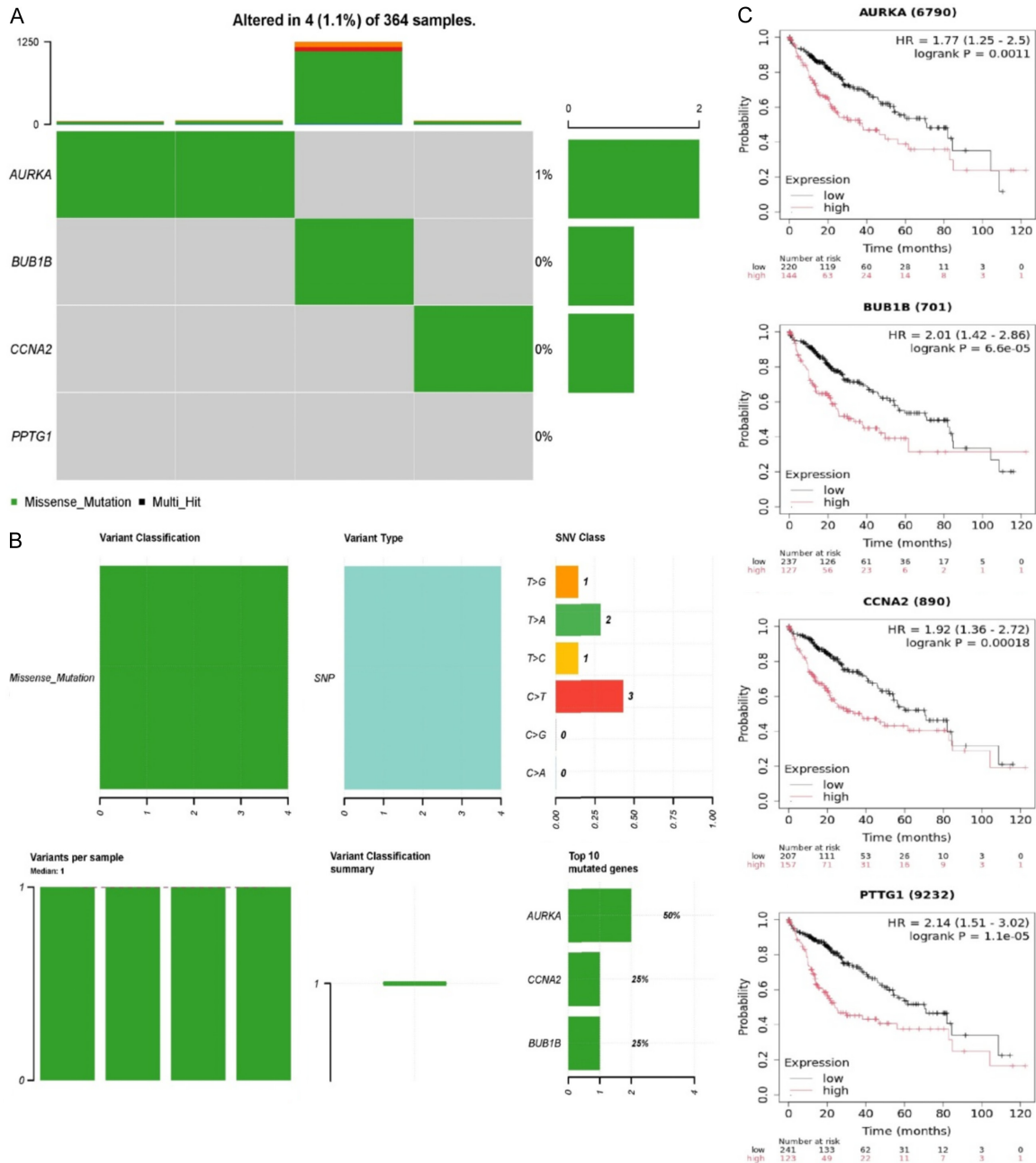
Figure 4. Mutation landscape and prognostic impact of hub genes in liver hepatocellular carcinoma (LIHC). A. The oncoprint shows the mutation frequency in AURKA, BUB1B, CCNA2, and PPTG1 across 364 LIHC samples using cBioPortal. B. Detailed breakdown of the mutation types and classifications in the same cohort. C. Kaplan-Meier survival plots were generated using the KM Plotter tool, illustrating the prognostic significance of high vs. low expression of AURKA, BUB1B, CCNA2, and PPTG1 in cancer patients. *P*-value < 0.05.

Correlation analysis of the hub genes with immune cell abundance and drug sensitivity in LIHC

Correlations of the hub genes with immune cell abundance were explored using the GSCA database. **Figure 5A** shows the correlation between the expression of PTTG1, CCNA2, AURKA, and BUB1B with immune cell abundance in LIHC. Notably, all four genes (PTTG1, CCNA2, AURKA, and BUB1B) show strong positive correlations with the infiltration of macrophages and neutrophils (**Figure 5A**). These genes also appear to correlate negatively with other immune cells, such as B cells and T cells, though these correlations are less significant (**Figure 5B**). Fur-
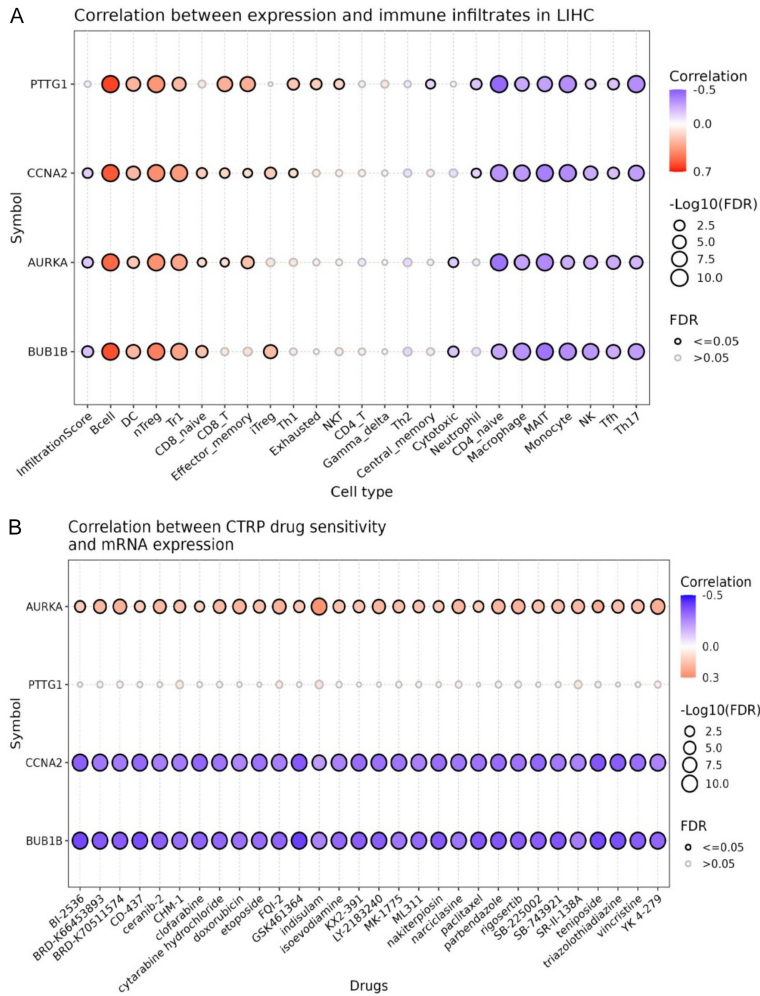
Figure 5. Correlation between hub gene expression and immune infiltration & drug sensitivity in liver hepatocellular carcinoma (LIHC). A. Correlation analysis between the expression of hub genes (PTTG1, CCNA2, AURKA, BUB1B) and immune cell infiltration in LIHC via the GSCA database. B. Correlation analysis between mRNA expression of the hub genes and drug sensitivity via the GSCA database. *P*-value < 0.05.

thermore, **Figure 5B** illustrates the correlation between hub genes and drug sensitivity in LIHC using the GSCA database (**Figure 5B**). AURKA and PTTG1 show positive correlations with a range of drugs, suggesting that higher expression of these genes might be associated with increased drug resistance. In contrast, CCNA2 and BUB1B exhibit strong negative correlations with many drugs, indicating that their higher expression could be associated with drug sensitivity in LIHC (**Figure 5B**).

Gene enrichment analysis of the hub genes

Gene enrichment analysis of the hub genes was conducted using the DAVID tool. **Figure 6A**

illustrates the enrichment of various cellular components. The "Germinal vesicle" and "Mitotic checkpoint complex" show the highest fold enrichment, suggesting that these components are particularly significant in the context of hub genes. Other highly enriched components include the "Cyclin A2-CDK2 complex" and "Chromosome passenger complex", all of which play crucial roles in cell division and mitosis. **Figure 6B** focuses on the molecular functions enriched in hub genes. The most significantly enriched function is "Histone serine kinase activity", followed by "Histone kin-ase activity" and "Cyclin-dependent protein serine/threonine kinase regulator activity". These functions are involved in cell cycle regulation and chromatin modification, which are critical for maintaining genomic stability during cell division. The presence of "Protein kinase activity" and "Kinase binding" further emphasizes the importance of phosphorylation events in the regulation of these biological processes. **Figure 6C** showcases the enrichment of various biological processes. The most significant processes include "Histone phosphoryla-

tion" and "Negative regulation of sister chromatid segregation", indicating a strong focus on processes that ensure proper chromosome segregation and stability during cell division. Other significantly enriched processes include "Meiotic nuclear division" and "Mitotic cell cycle", further underscoring the central role of these processes in cell proliferation and division. **Figure 6D** highlights the enriched pathways associated with hub genes. The most enriched pathway is the "Cell cycle", followed by "Progesterone-mediated oocyte maturation" and "Oocyte meiosis". These pathways are crucial for the regulation of cell division and reproductive processes. Interestingly, pathways re-
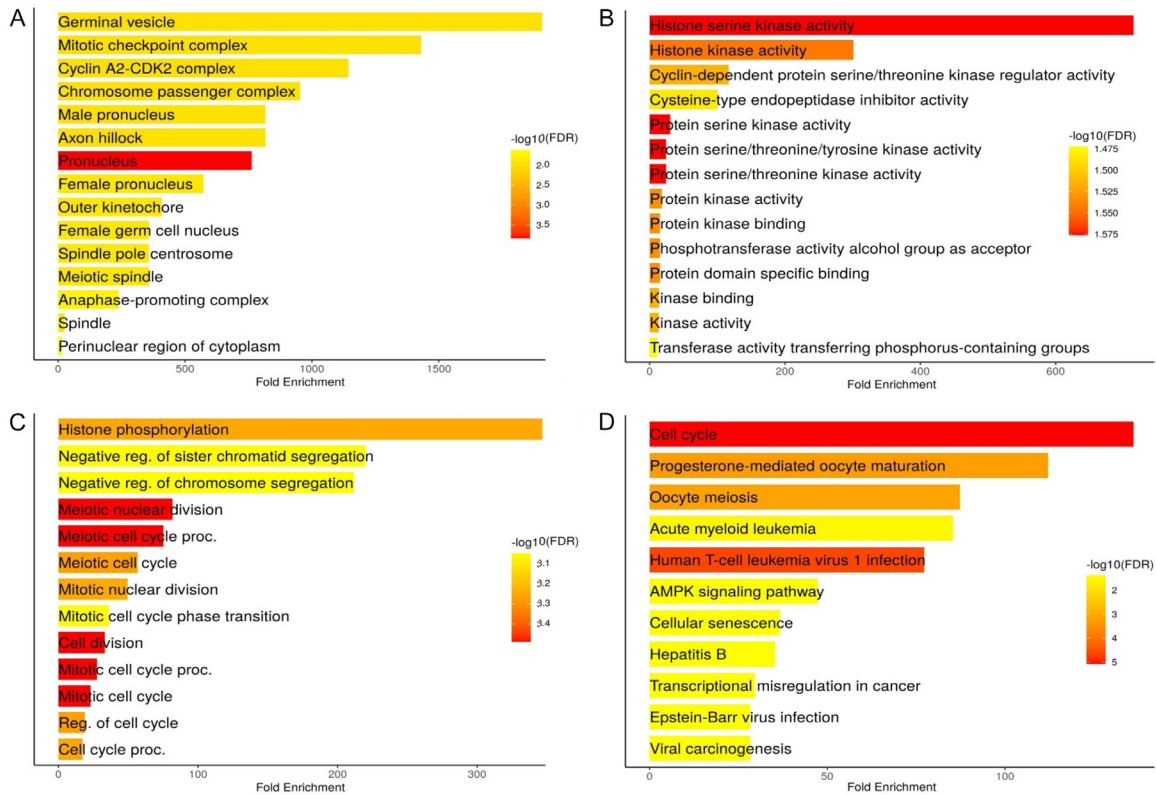
Figure 6. Gene ontology and pathway enrichment analysis of hub genes. A. Cellular component (CC) enrichment analysis. B. Molecular function (MF) enrichment analysis. C. Biological process (BP) enrichment analysis. D. Pathway enrichment analysis. *P*-value < 0.05.

lated to cancers, such as "Acute myeloid leukemia" and "Human T-cell leukemia virus 1 infection", are also significantly enriched, suggesting a potential link between these genes and oncogenic processes. Additionally, viral infection-related pathways like "Epstein-Barr virus infection" and "Hepatitis B" are also enriched, indicating a potential interaction between these genes and viral mechanisms.

*Validation of the hub gene expression across LIHC cell lines*

The expression of hub genes across 9 LIHC and 5 normal control cell lines was validated using RT-qPCR. As depicted in **Figure 7A**. The box plots revealed that AURKA, BUB1B, CCNA2, and PTTG1 genes exhibit a marked (*p*-value < 0.001) upregulation in the LIHC cell lines group as compared to the normal cell lines group. Furthermore, **Figure 7B** presents ROC curves for these genes, with each curve showing an Area Under the Curve (AUC) of 1.00. This indicates that AURKA, BUB1B, CCNA2, and PTTG1

have perfect discrimination between LIHC and normal individuals, highlighting their potential as highly effective biomarkers of LIHC.

*AURKA gene knockdown and functional assays*

Finally, the AURKA gene was knocked down in HepG2 cells to analyze its effects on cell proliferation, colony formation, and wound healing abilities. **Figure 8A** confirms that the knockdown of AURKA was successful, as indicated by the significant (*p*-value < 0.01) reduction in AURKA expression in the si-AURKA-HepG2 cells compared to the control (Ctrl-HepG2) cells via RT-qPCR. Cell proliferation assay revealed that the reduction in AURKA expression leads to a noticeable (*p*-value < 0.01) decrease in cell proliferation, as shown in **Figure 8B**. Furthermore, the colony formation assay demonstrates that the si-AURKA-HepG2 cells formed significantly (*p*-value < 0.01) fewer colonies than the control cells (**Figure 8C**, **8D**). Interestingly, the data from the wound healing assay (**Figure 8E-G**) revealed a contrasting
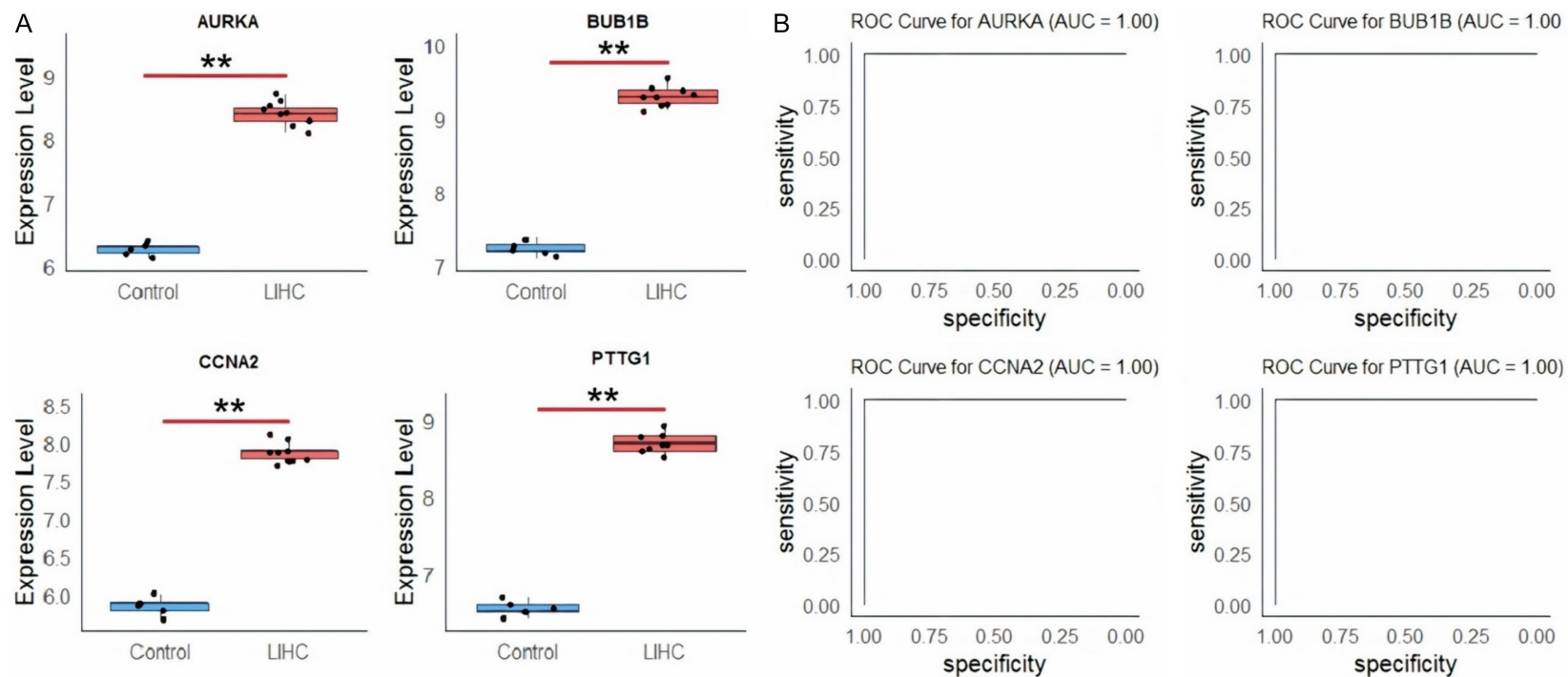
Figure 7. Differential expression and diagnostic accuracy of hub genes in liver hepatocellular carcinoma (LIHC) and normal control cell lines. A. Box plots showing the expression levels of AURKA, BUB1B, CCNA2, and PTTG1 in LIHC cell lines compared to normal control cell lines. B. Receiver Operating Characteristic (ROC) curves demonstrating the diagnostic accuracy of AURKA, BUB1B, CCNA2, and PTTG1 in distinguishing LIHC from normal individuals. *P*-value < 0.05.
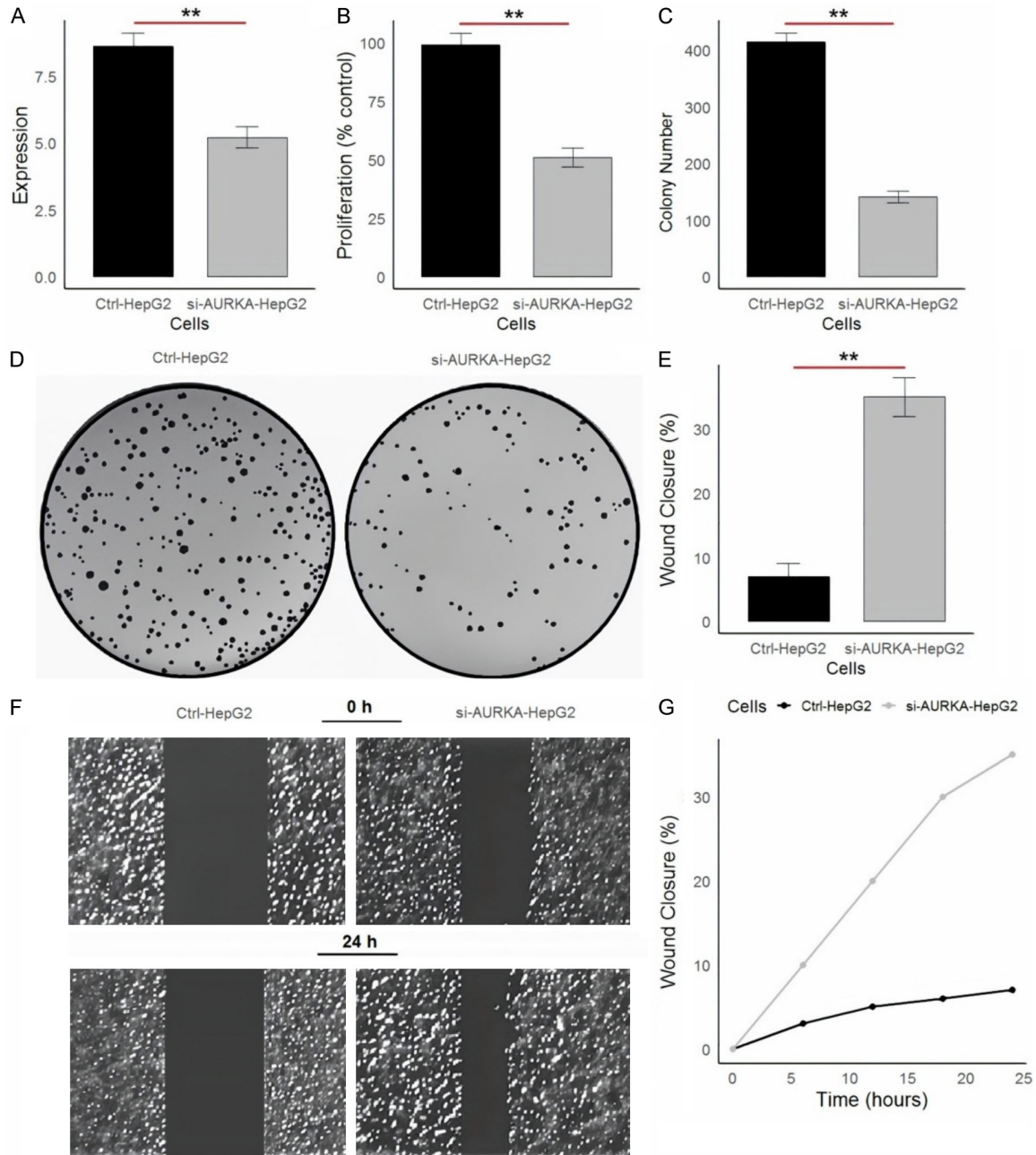
Figure 8. Effects of SURKS knockdown on HepG2 cell proliferation, colony formation, and wound healing. A. Relative expression of AURKA in si-AURKA-transfected HepG2 cells (si-AURKA-HepG2) compared to control HepG2 cells (Ctrl-HepG2). B. Proliferation of si-AURKA-HepG2 cells compared to Ctrl-HepG2 cells. C. Count of colonies formed by si-AURKA-transfected HepG2 cells (si-AURKA-HepG2) compared to control HepG2 cells (Ctrl-HepG2). D. Representative images of colony formation assays in Ctrl-HepG2 and si-AURKA-HepG2 cells. E. Quantification of wound healing closure abilities of Ctrl-HepG2 and si-AURKA-HepG2 cells. F. Representative images of wound healing assays at 0 hour and 24 hours for Ctrl-HepG2 and si-AURKA-HepG2 cells. The wound closure is significantly impaired in si-AURKA-HepG2 cells. G. Quantitative analysis of wound closure over time. $**P$-value < 0.01.

effect of AURKA knockdown on cell migration. While AURKA knockdown inhibits cell proliferation, it significantly (p-value < 0.01) enhances the migratory ability of HepG2 cells (**Figure 8E-G**).

Discussion

Liver Hepatocellular Carcinoma (LIHC) is a significant and aggressive form of liver cancer, known for its high mortality rates due to late

diagnosis and limited treatment options [28, 29]. This study aimed to elucidate the roles of key hub genes in LIHC and assess their potential as biomarkers and therapeutic targets. We analyzed gene expression profiles from datasets GSE84598 and GSE19665 to identify DEGs associated with LIHC. From this analysis, 180 common DEGs were identified and used to construct a PP network. Based on the degree of connectivity, up-regulated AURKA, BUB1B, CCNA2, and PTTG1 emerged as key hub genes within this network.

The upregulation of AURKA, BUB1B, CCNA2, and PTTG1 can drive cancer progression through several mechanistic pathways. AURKA is crucial for centrosome maturation and spindle assembly during mitosis; its overexpression can lead to chromosomal instability and aneuploidy, which are hallmarks of cancer [30-32]. BUB1B is involved in the spindle assembly checkpoint, ensuring accurate chromosome segregation; its dysregulation can result in mitotic errors and tumorigenesis [33, 34]. CCNA2 regulates the cell cycle transition from G1 to S phase and its overexpression can promote uncontrolled cell proliferation and resistance to apoptosis [35, 36]. PTTG1 is associated with chromosomal segregation and tumor progression, and its upregulation can contribute to the development of aggressive cancer phenotypes by enhancing cell proliferation and inhibiting cell death [37, 38]. Together, the overexpression of these genes disrupts normal cell cycle regulation and mitotic processes, leading to genomic instability and malignant transformation. Our finding supports previous research that has highlighted these genes as critical in various cancers, including breast, lung, and colorectal cancers [39-42].

Promoter methylation analysis provided further insights into gene regulation. We observed reduced promoter methylation in LIHC tissues compared to normal tissues, suggesting that hypomethylation contributes to the upregulation of these genes. This finding is consistent with previous research showing that hypomethylation of oncogenes often leads to their overexpression and is a common feature in various cancers [43-45].

Mutational analysis showed that alterations in AURKA, BUB1B, and CCNA2 are present in a small fraction of LIHC samples, primarily as missense mutations. Although these genetic alterations are relatively rare, their impact on protein function and cancer progression is significant. Survival analysis using Kaplan-Meier plots indicated that high expression levels of these hub genes correlate with poor overall survival in LIHC patients, reinforcing their role as prognostic markers in cancer.

Correlation analysis revealed that the expression of hub genes is positively correlated with the abundance of macrophages and neutrophils, and negatively correlated with other immune cell types in LIHC. Our observation that these genes are positively correlated with macrophage and neutrophil abundance adds new insights into their impact on the immune microenvironment of LIHC. Additionally, AURKA and PTTG1 showed positive correlations with drug resistance, whereas CCNA2 and BUB1B were associated with drug sensitivity. Our observation that AURKA and PTTG1 are associated with drug resistance is consistent with reports suggesting that these genes can mediate cellular responses to chemotherapy, thus influencing treatment outcomes [46, 47].

Functional assays showing that AURKA knockdown reduces proliferation and colony formation while enhancing migration, which are consistent with earlier research. Studies have previously shown that AURKA's role in cell division and migration is critical for cancer progression [48]. The reduction in cell proliferation and colony formation due to AURKA knockdown echoes findings that AURKA promotes oncogenic cell growth [48]. The enhancement of migratory abilities observed in our study supports reports that AURKA can also influence cancer cell motility, contributing to metastasis.

This study provides valuable insights into the roles of AURKA, BUB1B, CCNA2, and PTTG1 in LIHC, revealing their potential as biomarkers and therapeutic targets. The merit of the study lies in its comprehensive approach, utilizing multiple datasets and functional assays to validate gene expression and elucidate the mechanistic pathways by which these hub genes contribute to cancer progression. The correlation of hub genes with immune cell abundance and drug sensitivity adds depth to the understanding of their role in the tumor microenvironment and treatment response. However, the study's limitations include the reliance on Bioinfor-

matics analyses and cell line models, which may not fully replicate the complexity of in vivo tumor biology. Additionally, the focus on a limited set of genes may overlook other potentially significant factors in LIHC progression. Future research could address these limitations by incorporating more diverse models and exploring additional molecular targets.

Conclusion

In conclusion, our study provides a comprehensive analysis of AURKA, BUB1B, CCNA2, and PTTG1 in LIHC, integrating gene expression, methylation, mutation, and functional data. These findings emphasize the importance of these hub genes in cancer progression and their potential as biomarkers and therapeutic targets, offering new insights into the molecular mechanisms underlying LIHC and paving the way for future research and clinical applications.

Disclosure of conflict of interest

None.

Address correspondence to: Xueming Ying, Department of Oncology, Jingdezhen First People's Hospital, Jindezhen 333000, Jiangxi, China. E-mail: 15607082800@163.com; Salman Khan, DHQ Teaching Hospital, GMC, Dikah, Pakistan. E-mail: salmankhn663@gmail.com

References

[1] Huang PS, Wang LY, Wang YW, Tsai MM, Lin TK, Liao CJ, Yeh CT and Lin KH. Evaluation and application of drug resistance by biomarkers in the clinical treatment of liver cancer. Cells 2023; 12: 869.

[2] Rukmangad A, Deshpande A, Jamthikar A, Gupta D, Bhurane A and Meshram NB. Classification of H&E stained liver histopathology images using ensemble learning techniques for detection of the level of malignancy of hepatocellular carcinoma (HCC). Advances in Artificial Intelligence-Empowered Decision Support Systems: Papers in Honour of Professor John Psarras. Springer; 2024. pp. 89-108.

[3] Li S, Xue P, Diao X, Fan QY, Ye K, Tang XM, Liu J, Huang ZY, Tang QH, Jia CY, Xin R, Lv ZW, Liu JB, Ma YS and Fu D. Identification and validation of functional roles for three MYC-associated genes in hepatocellular carcinoma. J Adv Res 2023; 54: 133-146.

[4] Tang J, Peng X, Xiao D, Liu S, Tao Y and Shu L. Disulfidptosis-related signature predicts prognosis and characterizes the immune microenvironment in hepatocellular carcinoma. Cancer Cell Int 2024; 24: 19.

[5] Yan C, Niu Y, Ma L, Tian L and Ma J. System analysis based on the cuproptosis-related genes identifies LIPT1 as a novel therapy target for liver hepatocellular carcinoma. J Transl Med 2022; 20: 452.

[6] Gao J, Han S, Gu J, Wu C and Mu X. The prognostic and therapeutic role of histone acetylation modification in LIHC development and progression. Medicina (Kaunas) 2023; 59: 1682.

[7] Sarfaraz N, Somarowthu S and Bouchard MJ. The interplay of long noncoding RNAs and hepatitis B virus. J Med Virol 2023; 95: e28058.

[8] Wu Z, Dong Z, Luo J, Hu W, Tong Y, Gao X, Yao W, Tian H and Wang X. A comprehensive comparison of molecular and phenotypic profiles between hepatitis B virus (HBV)-infected and non-HBV-infected hepatocellular carcinoma by multi-omics analysis. Genomics 2024; 116: 110831.

[9] Liu C, Dai Q, Ding Q, Wei M and Kong X. Identification of key genes in hepatitis B associated hepatocellular carcinoma based on WGCNA. Infect Agent Cancer 2021; 16: 18.

[10] Mekuria AN, Routledge MN, Gong YY and Sisay M. Aflatoxins as a risk factor for liver cirrhosis: a systematic review and meta-analysis. BMC Pharmacol Toxicol 2020; 21: 39.

[11] Zenlander R. Biomarkers in hepatocellular carcinoma. Inst för medicin, Huddinge/Dept of Medicine, Huddinge; 2024.

[12] Ye J, Ying J, Chen H, Wu Z, Huang C, Zhang C, Chen Z and Chen H. PPIH acts as a potential predictive biomarker for patients with common solid tumors. BMC Cancer 2024; 24: 681.

[13] Yu J, Park R and Kim R. Promising novel biomarkers for hepatocellular carcinoma: diagnostic and prognostic insights. J Hepatocell Carcinoma 2023; 10: 1105-1127.

[14] Wang J, Dai M, Xing X, Wang X, Qin X, Huang T, Fang Z, Fan Y and Xu D. Genomic, epigenomic, and transcriptomic signatures for telomerase complex components: a pan-cancer analysis. Mol Oncol 2023; 17: 150-172.

[15] Melis M, Diaz G, Kleiner DE, Zamboni F, Kabat J, Lai J, Mogavero G, Tice A, Engle RE, Becker S, Brown CR, Hanson JC, Rodriguez-Canales J,

Emmert-Buck M, Govindarajan S, Kew M and Farci P. Viral expression and molecular profiling in liver tissue versus microdissected hepatocytes in hepatitis B virus-associated hepatocellular carcinoma. J Transl Med 2014; 12: 230.

[16] Ji Y, Yin Y and Zhang W. Integrated bioinformatic analysis identifies networks and promising biomarkers for hepatitis B virus-related hepatocellular carcinoma. Int J Genomics 2020; 2020: 2061024.

[17] Clough E and Barrett T. The gene expression omnibus database. Methods Mol Biol 2016; 1418: 93-110.

[18] Jia A, Xu L and Wang Y. Venn diagrams in bioinformatics. Brief Bioinform 2021; 22: bbab108.

[19] Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ and Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 2019; 47: D607-D613.

[20] Li T, Gao X, Han L, Yu J and Li H. Identification of hub genes with prognostic values in gastric cancer by bioinformatics analysis. World J Surg Oncol 2018; 16: 114.

[21] Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi BVSK and Varambally S. UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. Neoplasia 2017; 19: 649-658.

[22] Tang G, Cho M and Wang X. OncoDB: an interactive online database for analysis of gene expression and viral infection in cancer. Nucleic Acids Res 2022; 50: D1334-D1339.

[23] Thul PJ and Lindskog C. The Human Protein Atlas: a spatial map of the human proteome. Protein Sci 2018; 27: 233-244.

[24] Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C and Schultz N. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2012; 2: 401-404.

[25] Lánczky A and Győrffy B. Web-based survival analysis tool tailored for medical research (KMplot): development and implementation. J Med Internet Res 2021; 23: 27633.

[26] Liu CJ, Hu FF, Xie GY, Miao YR, Li XW, Zeng Y and Guo AY. GSCA: an integrated platform for gene set cancer analysis at genomic, pharmacogenomic, and immunogenomic levels. Brief Bioinform 2023; 24: bbac558.

[27] Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, Imamichi T and Chang W. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res 2022; 50: W216-W221.

[28] Sales CBS, Dias RB, de Faro Valverde L, Bomfim LM, Silva LA, de Carvalho NC, Bastos JLA, Tilli TM, Rocha GV, Soares MBP, de Freitas LAR, Gurgel Rocha CA and Bezerra DP. Hedgehog components are overexpressed in a series of liver cancer cases. Sci Rep 2024; 14: 19507.

[29] Natu A, Singh A and Gupta S. Hepatocellular carcinoma: understanding molecular mechanisms for defining potential clinical modalities. World J Hepatol 2021; 13: 1568-1583.

[30] Athwal H, Kochiyanil A, Bhat V, Allan AL and Parsyan A. Centrosomes and associated proteins in pathogenesis and treatment of breast cancer. Front Oncol 2024; 14: 1370565.

[31] Sarı S and Özsoy ER. Aurora Kinases: their role in cancer and cellular processes. Türk Doğa ve Fen Dergisi 2024; 13: 128-139.

[32] Hameed Y and Ejaz S. TP53 lacks tetramerization and N-terminal domains due to novel inactivating mutations detected in leukemia patients. J Cancer Res Ther 2021; 17: 931-937.

[33] Cicirò Y, Ragusa D and Sala A. Expression of the checkpoint kinase BUB1 is a predictor of response to cancer therapies. Sci Rep 2024; 14: 4461.

[34] Dong Y, Wu X, Xu C, Hameed Y, Abdel-Maksoud MA, Almanaa TN, Kotob MH, Al-Qahtani WH, Mahmoud AM, Cho WC and Li C. Prognostic model development and molecular subtypes identification in bladder urothelial cancer by oxidative stress signatures. Aging (Albany NY) 2024; 16: 2591-2616.

[35] Zhang J, Di Y, Zhang B, Li T, Li D and Zhang H. CDK1 and CCNA2 play important roles in oral squamous cell carcinoma. Medicine (Baltimore) 2024; 103: e37831.

[36] Usman M, Hameed Y and Ahmad M. Does human papillomavirus cause human colorectal cancer? Applying Bradford Hill criteria postulates. Ecancermedicalscience 2020; 14: 1107.

[37] Lu Y, Wang D, Chen G, Shan Z and Li D. Exploring the molecular landscape of osteosarcoma through PTTG family genes using a detailed multi-level methodology. Front Genet 2024; 15: 1431668.

[38] Wang L and Liu X. Multi-omics analysis of the oncogenic value of pituitary tumor-transforming gene 1 (PTTG1) in human cancers. Front Biosci (Landmark Ed) 2024; 29: 87.

[39] Toolabi N, Daliri FS, Mokhlesi A and Talkhabi M. Identification of key regulators associated with colon cancer prognosis and pathogenesis. J Cell Commun Signal 2022; 16: 115-127.

[40] Su Q, Li W, Zhang X, Wu R, Zheng K, Zhou T, Dong Y, He Y, Wang D and Ran J. Integrated

bioinformatics analysis for the screening of hub genes and therapeutic drugs in hepatocellular carcinoma. Curr Pharm Biotechnol 2023; 24: 1035-1058.

[41] Yang WX, Pan YY and You CG. CDK1, CCNB1, CDC20, BUB1, MAD2L1, MCM3, BUB1B, MCM2, and RFC4 may be potential therapeutic targets for hepatocellular carcinoma using integrated bioinformatic analysis. Biomed Res Int 2019; 2019: 1245072.

[42] Hu H, Umair M, Khan SA, Sani AI, Iqbal S, Khalid F, Sultan R, Abdel-Maksoud MA, Mubarak A, Dawoud TM, Malik A, Saleh IA, Al Amri AA, Algarzae NK, Kodous AS and Hameed Y. CDCA8, a mitosis-related gene, as a prospective pancancer biomarker: implications for survival prognosis and oncogenic immunology. Am J Transl Res 2024; 16: 432-445.

[43] Höbaus J, Hummel DM, Thiem U, Fetahu IS, Aggarwal A, Müllauer L, Heller G, Egger G, Mesteri I, Baumgartner-Parzer S and Kallay E. Increased copy-number and not DNA hypomethylation causes overexpression of the candidate proto-oncogene CYP24A1 in colorectal cancer. Int J Cancer 2013; 133: 1380-1388.

[44] Van Tongelen A, Loriot A and De Smet C. Oncogenic roles of DNA hypomethylation through the activation of cancer-germline genes. Cancer Lett 2017; 396: 130-137.

[45] Hameed Y. Decoding the significant diagnostic and prognostic importance of maternal embryonic leucine zipper kinase in human cancers through deep integrative analyses. J Cancer Res Ther 2023; 19: 1852-1864.

[46] Caporali S, Alvino E, Levati L, Esposito AI, Ciomei M, Brasca MG, Del Bufalo D, Desideri M, Bonmassar E, Pfeffer U and D'Atri S. Downregulation of the PTTG1 proto-oncogene contributes to the melanoma suppressive effects of the cyclin-dependent kinase inhibitor PHA-848125. Biochem Pharmacol 2012; 84: 598-611.

[47] Gao M, Kong W, Huang Z and Xie Z. Identification of key genes related to lung squamous cell carcinoma using bioinformatics analysis. Int J Mol Sci 2020; 21: 2994.

[48] Wu J, Yang L, Shan Y, Cai C, Wang S and Zhang H. AURKA promotes cell migration and invasion of head and neck squamous cell carcinoma through regulation of the AURKA/Akt/FAK signaling pathway. Oncol Lett 2016; 11: 1889-1894.