



Identification of heterogeneous nuclear ribonucleoprotein as a candidate biomarker for diagnosis and prognosis of hepatocellular carcinoma

Youli Du¹, Xiaoou Ma¹, Dongxu Wang², Yuguang Wang², Tianyu Zhang², Lianjie Bai³, Yunlong Liu⁴, Shaosen Chen⁴

¹Department of Interventional Medicine, The Second Affiliated Hospital of Qiqihar Medical College, Qiqihar, China; ²CT Room of the Second Affiliated Hospital of Qiqihar Medical College, Qiqihar, China; ³The Ultrasound Department of the Second Affiliated Hospital of Qiqihar Medical College, Qiqihar, China; ⁴Department of Oncology, the Second Affiliated Hospital of Qiqihar Medical College, Qiqihar, China

Contributions: (I) Conception and design: Y Du, X Ma; (II) Administrative support: X Ma, D Wang, Y Wang; (III) Provision of study materials or patients: Y Liu, S Chen; (IV) Collection and assembly of data: T Zhang, L Bai; (V) Data analysis and interpretation: Y Du, X Ma, Y Wang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Youli Du. Department of Interventional Medicine, Second Affiliated Hospital of Qiqihar Medical College, No. 333, Bukui Street, Jianhua District, Qiqihar 161006, China. Email: duyouli1236@163.com.

Background: Hepatocellular carcinoma (HCC) is the most common type of liver cancer with a high mortality rate. However, spliceosomal genes are still lacking in the diagnosis and prognosis of HCC.

Methods: Identification of differentially expressed genes (DEGs) was performed using the limma package in R software. Modules highly related to HCC were obtained by weighted gene co-expression network analysis (WGCNA), and the module genes were analyzed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway. The biomarker for diagnosing HCC was determined by receiver operating characteristic (ROC) curve analysis, and the effect of the biomarker in the diagnosis of HCC was evaluated by performing five-fold cross-validation with logistic regression. HCC specimens from preoperatively treated patients were tested for biomarker by real-time quantitative polymerase chain reaction (RT-qPCR). Kaplan-Meier analysis was used to assess the relationship between biomarker and patient survival. The role of biomarker was evaluated using ESTIMATE analysis in the tumor microenvironment.

Results: In this study, 389 DEGs were screened out from three Gene Expression Omnibus (GEO) datasets. We also found that the turquoise module of 123 genes from The Cancer Genome Atlas (TCGA) data was the key module with the highest correlation with HCC traits. Then, 123 genes were analyzed using the KEGG enrichment pathway, and eight genes were found to be most significantly related to the spliceosome pathway. We selected 8 genes and 389 DEGs shared genes, and finally got the only gene, heterogeneous nuclear ribonucleoprotein (hnRNPU). The high expression of hnRNPU was associated with poor prognosis of HCC, and hnRNPU was a biomarker for diagnosing HCC. In the tissues of patients with excellent HCC treatment hnRNPU messenger RNA (mRNA) was lower than in the tissues of patients with poor HCC treatment. High expression of hnRNPU was significantly increased in HCC patients with low stromal ($P < 0.05$), low immune ($P < 0.05$), and low estimation scores ($P < 0.05$), and with high tumor purity ($P < 0.05$) and high malignant progression ($P < 0.05$) of the HCC.

Conclusions: The hnRNPU gene identified in this study may become a new biomarker for the diagnosis and prognosis of HCC.

Keywords: Hepatocellular carcinoma (HCC); heterogeneous nuclear ribonucleoprotein (hnRNPU); molecular biomarker

Submitted Jul 16, 2021. Accepted for publication Sep 02, 2021.

doi: 10.21037/jgo-21-468

View this article at: <https://dx.doi.org/10.21037/jgo-21-468>

Introduction

Hepatocellular carcinoma (HCC) is a life-threatening disease, whose morbidity and mortality rate is increasing (1). According to 2018 Global Cancer Incidence, Mortality, and Prevalence (GLOBOCAN) data, HCC is the sixth most common and malignant tumor in the world. Seventy-five percent to 85% of primary liver cancers are caused by HCC. The incidence of HCC is 4.7%, there are 18.1 million new cases each year, and it has an 8.2% mortality rate out of 9.6 million deaths (2). Due to the lack of sensitive molecular markers for early screening and diagnosis, most patients are already in the middle and late stages at the time of diagnosis. Therefore, it is very important to find promising diagnostic and prognostic markers for HCC.

Alternative splicing (AS) is a post-transcriptional process involving most eukaryotic genes. Alternative messenger RNA (mRNA) transcripts encode structurally or functionally distinct protein isoforms, thereby diversifying the cellular proteome (3). However, there is a disruption of normal AS regulation in cancer cells, leading to cancer-specific RNA transcriptional profiles that further promote proliferation and migration of cancer cells or escape cell death (4,5). In recent studies, large-scale analyses of various solid tumor types have suggested that the patterns of tumor-specific splicing can be attributed to aberrant regulation of splicing factors and include mutations, copy number changes, or alterations in the expression of splicing-regulated genes (6). These discoveries have led to a growing interest in the role of interference splicing factors in cancer development, aiming to provide new therapeutic strategies for cancer treatment (7,8). In HCC, abnormal AS events are also common and lead to the characteristics of HCC (9,10). In addition, multiple AS regulators include *SRSF2* (11), *SRSF3* (12), *hnRNPA2* (13), and *PTBP3* (14) and have been reported to be involved in HCC progress.

Weighted gene co-expression network analysis (WGCNA) is a method for elucidating the interaction of pathogenic genes in cellular processes and is a bioinformatics tool for determining correlations of central genes and modules with clinical features (15). It is available to identify clusters of highly associated genes (modules), summarize these clusters using module characteristic genes or intra-module hub genes, associate modules with each other and with features of external samples using the characteristic gene network method and calculate module membership (MM) metrics (16). Associated networks promote network-based gene screening methods, which can

be used to recognize candidate biomarkers or therapeutic targets. Therefore, we applied WGCNA to The Cancer Genome Atlas (TCGA) database to identify key genes correlated with pathway modules and to provide prognostic markers for HCC.

In this study, we aim to find molecular markers for the clinical application of HCC. Firstly, we screened out 389 differentially expressed genes (DEGs) from three GEO datasets, GSE6764, GSE14520, and GSE60502, which were upregulated together. The WGCNA software package (version 1.69, The R Foundation for Statistical Computing) was used to build a DEGs gene co-expression network from the TCGA database, and specific modules related to the clinical and pathophysiological characteristics of HCC were identified. In addition, the Kyoto Encyclopedia of Genes and Genomes (KEGG) was used to analyze the pathway of module genes. The survival of HCC patients was influenced by heterogeneous nuclear ribonucleoprotein (hnRNPU) gene with the help of Kaplan-Meier analysis. Evaluation of the sensitivity and specificity of hnRNPU for the diagnosis of HCC was analyzed using receiver operating characteristic (ROC) curves. In the tissues of patients with excellent HCC treatment hnRNPU mRNA was lower than in the tissues of patients with poor HCC treatment. Estimate analysis was used to assess the effect of hnRNPU in the development of HCC. Our results indicate that hnRNPU is highly expressed in patients with HCC and predicts a poor prognosis. It is a key gene for the diagnosis of HCC. The high expression of hnRNPU is connected to the progression and tumor purity of HCC. Our objective was to identify specific biomarkers that are closely related to HCC prognosis and to provide insights for the diagnosis and prognosis prediction of HCC.

We present the following article in accordance with the STARD reporting checklist (available at <https://dx.doi.org/10.21037/jgo-21-468>).

Methods

Data collection

Raw counts of clinical features and RNA-sequencing data for HCC samples were obtained from the TCGA data portal (<https://tcga-data.nci.nih.gov/tcga/>). In addition, datasets GSE6764, GSE14520, and GSE60502 of the Gene Chip Transcriptome Array were obtained from the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>). Clinical information includes age,

gender, pathological metastasis (M) stage, pathological node (N) stage, pathological tumor (T) stage, histological grade, mutation count, and fraction genome altered.

Patient selection

Formalin-fixed, paraffin-embedded primary HCC specimens obtained from 40 patients were randomly selected from The Second Affiliated Hospital of Qiqihar Medical College, Qiqihar, Heilongjiang Province, China, between Dec. 2019, and Feb. 2021 to be included in this study. All patients received transcatheter arterial chemoembolization (TACE) after diagnosis. Pre- and postoperative digital subtraction angiograph (DSA) abdominal stem images were collected for the 40 patients for treatment outcome differentiation. Three months after TACE of the liver lobe in patients with HCC, the patient's treatment was determined by reviewing the DSA angiogram showing the depth of tumor staining in the liver lobe. A postoperative contrast that was significantly lighter or almost disappeared compared to the preoperative contrast was considered a good treatment outcome; conversely, a poor treatment outcome was the opposite. Images for each patient were reviewed by two fellowship-trained radiologists with 8 and 10 years of experience, who were blinded to all patient information. HCC specimens from preoperatively treated patients were tested for the hub gene by real-time quantitative polymerase chain reaction (RT-qPCR). All procedures performed in this study involving human participants were in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by institutional ethics board of The Second Affiliated Hospital of Qiqihar Medical College (No.: 2020-01) and informed consent was taken from all the patients.

Analysis of DEGs

We selected datasets of GSE6764, GSE14520, and GSE60502 and downloaded the original files (.CEL files) and platform files. After data processing, the limma package of R software (The R Foundation for Statistical Computing) was used to perform differential expression analysis. We deemed the genes with $P < 0.001$ as DEGs. The DEGs GSE6764, GSE14520, and GSE60502 showed overlapping regions in the Venn diagram and were screened out and matrices were constructed.

Construction of WGCNA

The WGCNA package was installed, and co-expression analysis was performed using the HCC data from the TCGA database. The soft threshold method was used to perform Pearson correlation analysis on the expression profile to determine the strength of the connection between the two transcripts and to construct a weighted network. Average linkage hierarchical clustering of transcripts was performed based on topological overlap of different network connection strength. To obtain the correct number of modules and to clarify gene interactions, we set the minimum number of genes per module to 30 and used a threshold of $P < 0.25$ to merge similar modules.

ROC curve and logistic regression analysis

ROCs curve analysis was performed to evaluate the sensitivity and specificity of the hub gene for HCC diagnosis, and we calculated the value of the area under the curve (AUC) by using the statistical software medcalc (MedCalc Software, Ostend, Belgium). To further evaluate the effect of the hub gene in the diagnosis of HCC, we performed five-fold cross-validation using logistic regression with TCGA data.

Functional enrichment analysis

To predict the potential functions of the genes in the module screened from WGCNA, we performed functional enrichment analysis of module-associated genes to determine significantly enriched KEGG pathway. Then, the hub gene was obtained by overlapping the genes in the most significant pathway with the 389 DEGs.

RNA extraction and RT-qPCR

Total RNA was extracted from HCC tissues with Trizol Reagent (Invitrogen, Carlsbad, CA, USA), which was reversely transcribed with RT reagent Kit gDNA Eraser (TaKaRa Bio, Kusatsu, Japan). Then, cDNA expression levels were detected using SYBR-Green (TaKaRa) and RT-qPCR analysis with glyceraldehyde 3-phosphate dehydrogenase (GAPDH) as internal reference. The primers were: GAPDH, forward (F): 5'-TGTGTCCTCGTGGATCTGA-3', reverse (R): 5'-CCTGCTTCACCACCTTCTTGA-3';

hnRNPU, F: 5'-GTAGTAGTCA TCCATCTGTA-3', R: 5'-AAGTTAGCGGCAGATCTGTA-3'. PCR amplification was carried out in a formula of three wells. All experiments were repeated three times and genes' relative expression levels were studied with $2^{-\Delta\Delta C_t}$.

Estimate score analysis

The ESTIMATE score is a score to estimate the purity of stromal and immune cells in malignant tumor tissues by using expression data from TCGA samples. We used the estimation algorithm from the website (<https://bioinformatics.mdanderson.org/estimate/>) to calculate the estimated, stromal, and immune scores in HCC.

Statistical analysis

The Kaplan-Meier method was used to calculate survival rates, and log-rank tests were used to determine the significance of differences in survival curves. Statistical analysis was performed using R software (version 3.6.3, The R Foundation for Statistical Computing). Statistical significance was set at $P < 0.05$.

Results

Identification of DEGs in HCC after data integration

To identify DEGs in HCC, we selected the GSE6764 (35 HCC samples and 10 controls), GSE14520 (22 HCC samples and 19 controls), and GSE60502 (18 HCC samples and 18 controls) datasets and downloaded the raw files (.CEL files) and platform files. After data processing, differential expression analysis was performed using the R package software. As shown in the volcano plot, gene expression profiles from the GSE6764 dataset identified 1,510 differential genes in HCC samples compared to normal control tissues, of which 503 genes were expressed upregulated and 1,007 genes were expressed downregulated (Figure 1A). From the GSE14520 dataset, we identified 2,691 DEGs, of which 1,607 genes were expressed upregulated and 1,084 genes were expressed downregulated in HCC (Figure 1B). In GSE60502, 1,686 DEGs were detected, of which 901 genes were upregulated and 785 genes were downregulated in HCC (Figure 1C). Based on the cut-off criterion of $P < 0.001$, we obtained 389 shared differential genes by screening out the overlapping regions of the GSE6764, GSE14520, and GSE60502 genes using

a Venn diagram (Figure 1D, Table S1). Gene expression profiles of paracancer control and tumor samples in TCGA data confirmed the differential expression of these 389 genes (Figure 1E).

Construction of WGCNA and identification of the key module

Using the limma package of R software (The R Foundation for Statistical Computing), differential expression analysis was performed using TCGA data, which included 369 HCC specimens and 50 paracancerous liver specimens. According to the corresponding relationship between the cut-off values and the number of genes, we put genes with $P < 1e^{-5}$ into the co-expression network for analysis. To fully understand the gene expression network during the development of HCC, we analyzed the co-expression network of 12,485 genes using WGCNA. First, to confirm the suitability of all 361 HCC samples for network analysis, a dendrogram of the samples and the corresponding clinical characteristics were analyzed. One sample was removed, and the remaining 360 that passed the threshold were included in the study. Figure 2A shows a hierarchical clustering dendrogram of the 360 HCC samples in TCGA, with the clinical characteristics of gender, variant genomes, survival [overall survival (OS) vs. disease-free survival (DFS)] status, and survival (OS vs. DFS) time shown at the bottom. An important parameter affecting the independence and average connectivity of the co-expression modules is the power value. From this, we filtered to $\beta = 9$ (scale-free $R^2 = 0.900$) for subsequent analysis (Figure 2B, 2C). Then, we used WGCNA to construct a gene co-expression network based on hierarchical clustering of computed dissimilarity points to obtain 17 modules (Figure 2D). Subsequently, we used feature genes as representative profiles to quantify module similarity by feature gene correlation. The network heat map represents the topological overlap matrix (TOM) of 12,485 genes in WGCNA, with darker colors representing higher overlap and lighter colors representing lower overlap (Figure 2E). In the module and clinical feature diagram (Figure 2F), each row corresponds to a modularity, and each column represents a clinical feature. Each cell contains correlations and P values. Finally, we found that the turquoise module was the key module with the highest correlation with all types of HCC traits ($R^2 = -0.12$, $P = 0.02$ with gender; $R^2 = 0.29$, $P = 3e^{-8}$ with fraction genome altered; $R^2 = -0.19$, $P = 3e^{-4}$ with OS time; $R^2 = 0.11$, $P = 0.04$ with OS status; $R^2 = -0.19$, $P = 4e^{-4}$ with DFS time; $R^2 = 0.15$, $P = 0.004$ with

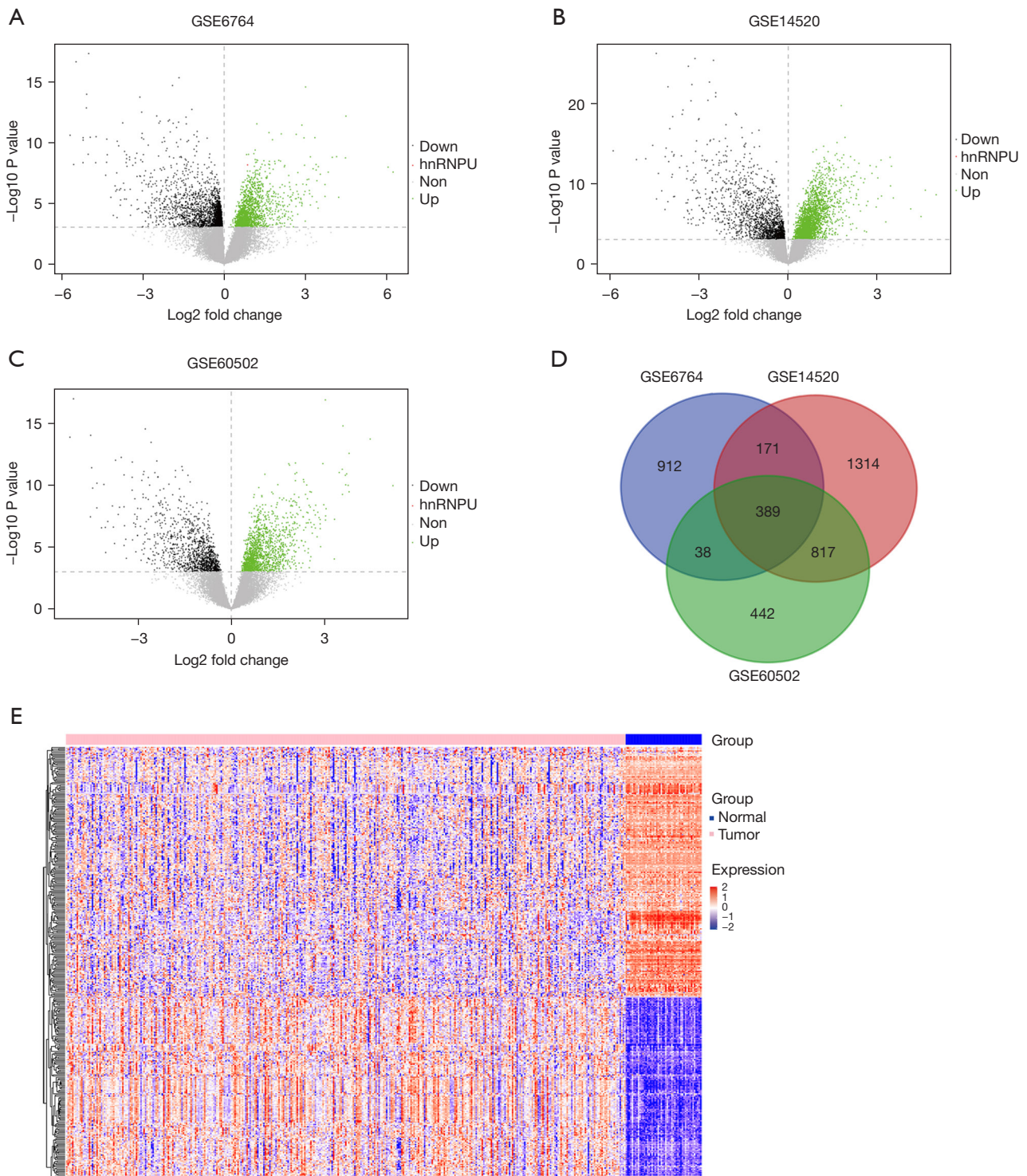


Figure 1 Differential gene expression in GSE6764, GSE14520, and GSE60502 datasets. (A-C) Volcano graphs of gene expression profiles of GSE6764, GSE14520, and GSE60502 datasets, where red represents hnRNPU, black represents down-regulated genes, green represents up-regulated genes and gray is non-significantly expressed genes, $P < 0.001$; (D) the Venn diagram of 389 DEGs among GSE6764, GSE14520 and GSE60502; (E) the clustering analysis of 389 DEGs with HCC and normal specimens in TCGA database. hnRNPU, heterogeneous nuclear ribonucleoprotein; DEGs, differentially expressed genes; HCC, hepatocellular carcinoma; TCGA, The Cancer Genome Atlas.

disease-free status). The scatter plots of gene significant (GS) vs. MM were drawn for the turquoise module, which is the most significant feature of the genomically altered fraction of HCC patients (correlation =0.072, $P=2.5e^{-6}$). Then, we further selected 123 genes as the module-related genes under the condition of MM >0.80 and GS >0.20 (Figure 2G; Table S2).

KEGG pathway enrichment analysis and identification of hnRNPU

To predict the potential functions of genes in the turquoise modules screened out from WGCNA, we performed functional enrichment analysis of module-associated genes to identify significantly enriched pathways. Pathway enrichment analysis using KEGG was performed using the 123 module-related genes and the 10 significant enrichment terms were displayed (Figure 3A,3B). The results demonstrated that the module-related genes were significantly associated with spliceosome, RNA transport, mismatch repair, homologous recombination, and cell cycle. Among them, spliceosome was the most significant pathway. The only hub gene hnRNPU was obtained by overlapping the spliceosome-related genes with 389 DEGs screened out previously (Figure 3C). When the expression of hnRNPU was linked to survival information, it turned out to be a risk factor. In all, HCC patients with higher hnRNPU expression tended to have poorer survival (Figure 3D).

Identification of hnRNPU for diagnosing and treating HCC

We performed ROC curve analysis of hnRNPU in the GSE6764, GSE14520, and GSE60502 datasets to evaluate the sensitivity and specificity of hnRNPU for the diagnosis of HCC. Figure 4A-4C shows ROC curves of hnRNPU in the three GEO datasets, with mean AUCs of 0.957, 0.868, and 0.870, indicating good sensitivity and specificity. Furthermore, five-fold crossvalidation with logistic regression was performed to assess the effectiveness of hnRNPU in diagnosing HCC with TCGA data, and the average of AUC was 0.928 ± 0.1 (Figure 4D). The mean values of accuracy, precision, recall, and F1-score were 0.853, 0.998, 0.844, and 0.912, respectively (Figure 4E). These results show that hnRNPU can effectively distinguish HCC tissues from normal tissues, thus demonstrating that its expression has good predictive ability for tumors. Further we did a study on the treatment outcome of clinical

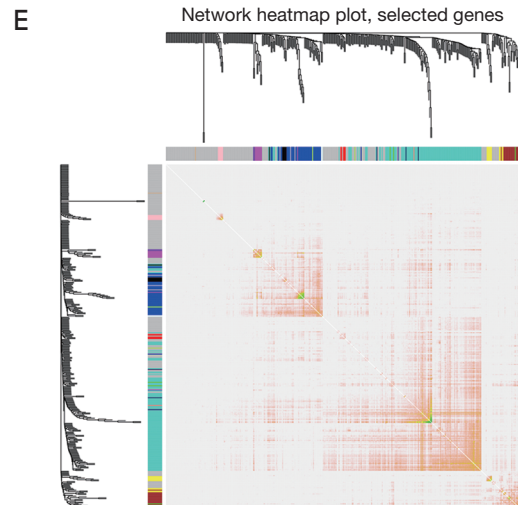
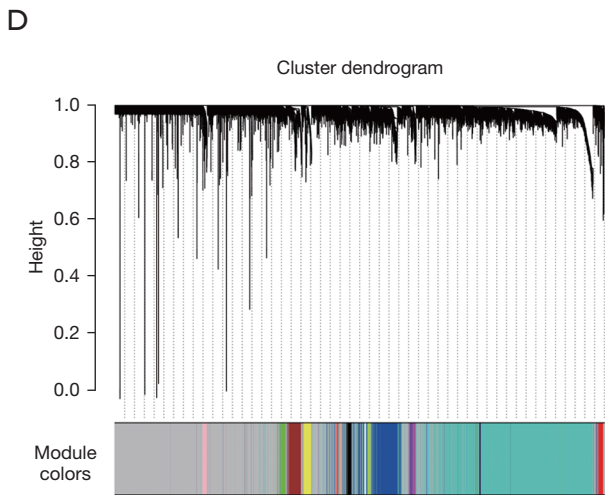
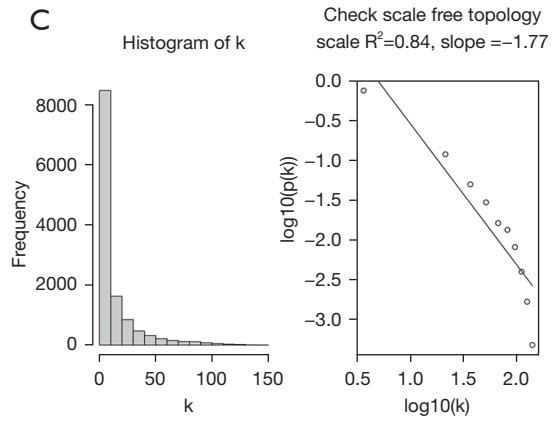
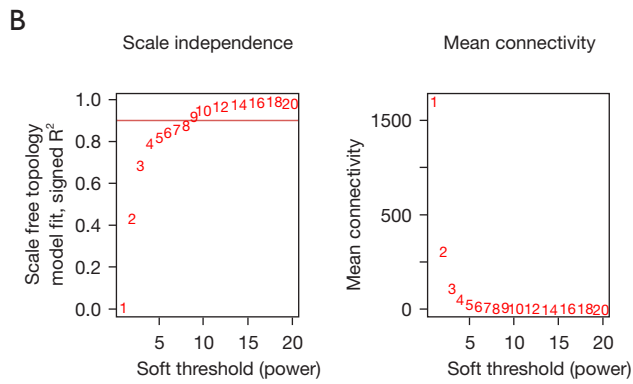
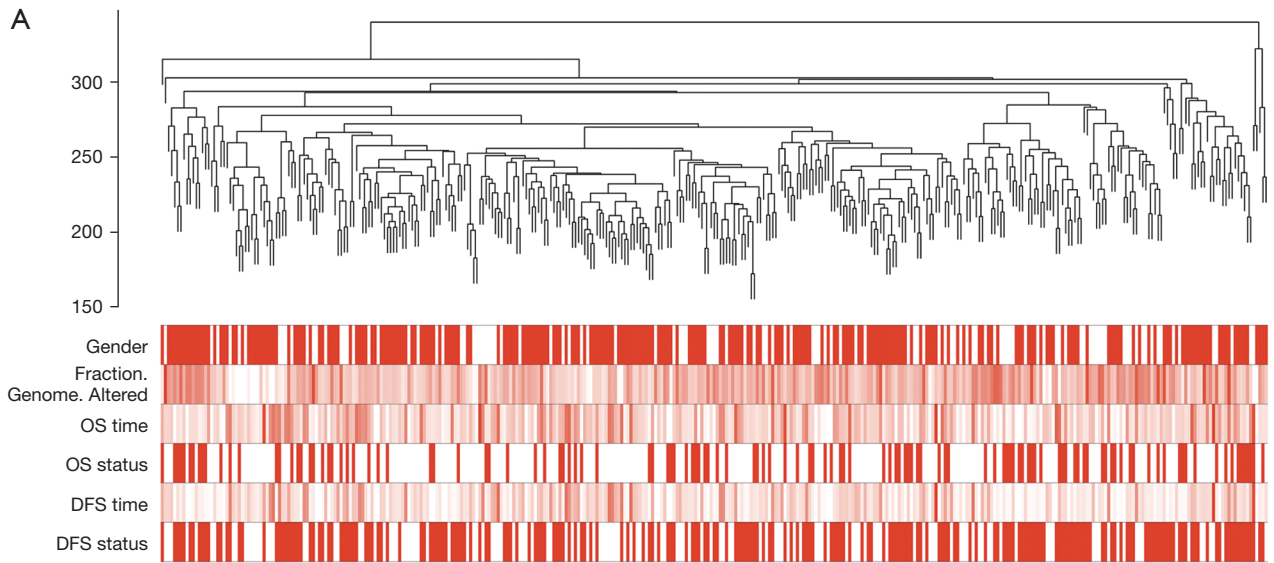
patients after hepatic artery chemoembolization. Based on preoperative and postoperative DSA abdominal stem angiography images, we screened the treatment outcome of 40 patients, of whom 17 had a poor postoperative treatment outcome (Figure 5A,5B), and 23 were well treated (Figure 5C,5D). Subsequently, we examined the expression of hnRNPU in the preoperative diagnostic tissues of well treated and poorly treated patients using RT-qPCR and found that the expression of hnRNPU was significantly lower in well treated patients than in poorly treated patients (Figure 5E), indicating that patients with HCC containing low hnRNPU are better treated postoperatively. DFS of HCC patients in the TCGA database was analyzed using a Kaplan-Meier plot. HCC patients with higher hnRNPU expression tended to have poorer survival (Figure 5F).

Overexpression of hnRNPU correlates with HCC progression and tumor purity

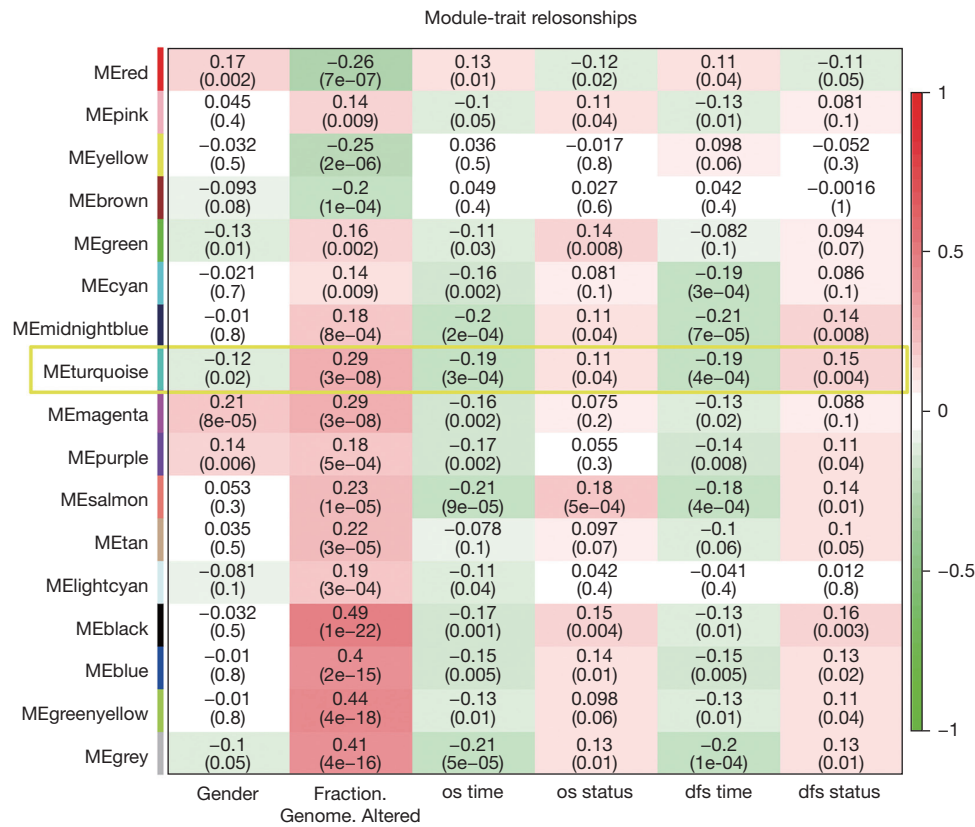
To evaluate the effect of hnRNPU in the development of HCC, we analyzed its expression using data from the TCGA database and found that the expression level of hnRNPU was significantly higher in the corresponding high-score HCC tissues compared to the low stromal score (median value), low immune score (median value), and low estimated score (median value) groups (Figure 6A-6C). In addition, hnRNPU expression level increased with tumor purity (Figure 6D). Further studies showed that the expression level of hnRNPU gradually improved along with HCC malignant progression (indicated by histologic grades; Figure 6E). All of the above results suggested that hnRNPU was expressed at different levels in HCC patients with different characteristics and different progression stages. Therefore, hnRNPU is a key gene in cancer research.

Discussion

Most genes in humans are first transcribed into pre-mRNA, containing non-coding sequences (introns) and coding sequences (exons), which are then processed by spliceosome complexes to remove introns before producing mature mRNA (17). Multiple proteins are allowed to translate from a single gene transcript by this approach (18). As a result, the human population has about 20,000 human genes that can encode at least 100,000 different proteins (19). Selective splicing produces cellular proteins that are selectively expressed in a tissue-specific and time-dependent manner and participate in a variety of regulatory pathways including



F



G

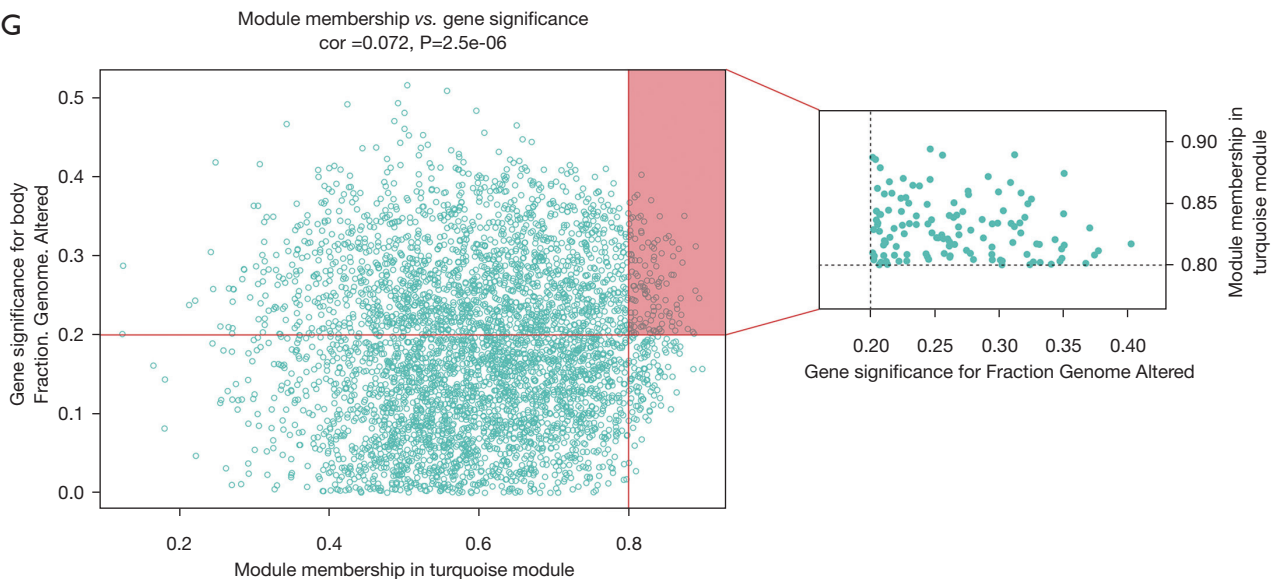
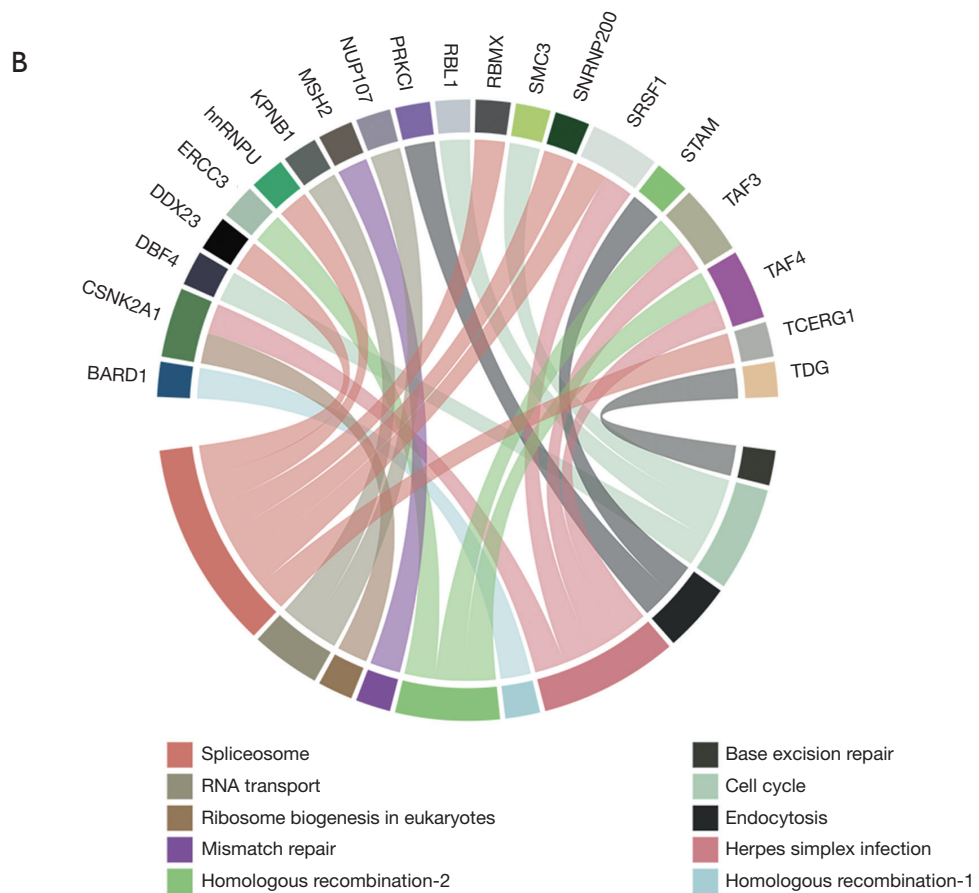
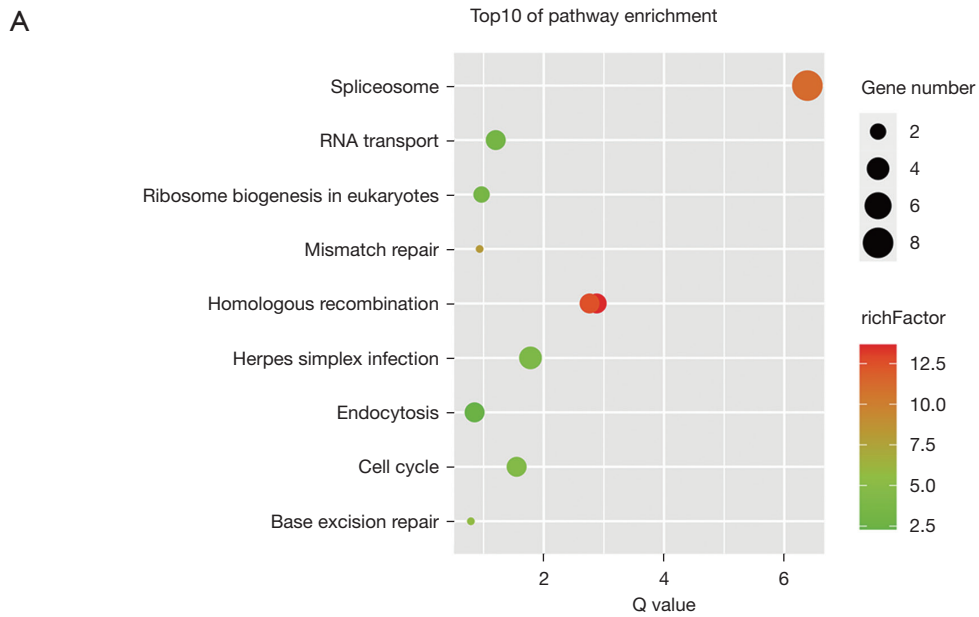


Figure 2 Construction of weighted co-expression network and identification of key modules. (A) Hierarchical clustering dendrogram of 360 HCC samples in TCGA; (B,C) analysis of scale-free fit indices and mean connectivity for different soft thresholds; (D) hierarchical clustering dendrograms of heterogeneous genes based on topological overlap. Modules are branches of the clustering tree; (E) the network heat map represents the TOM of 12,485 genes in WGCNA; (F) correlation between modularity and clinical features; (G) scatter plot of 123 genes in turquoise modules. HCC, hepatocellular carcinoma; TCGA, The Cancer Genome Atlas; TOM, topological overlap matrix; WGCNA, weighted co-expression network analysis.



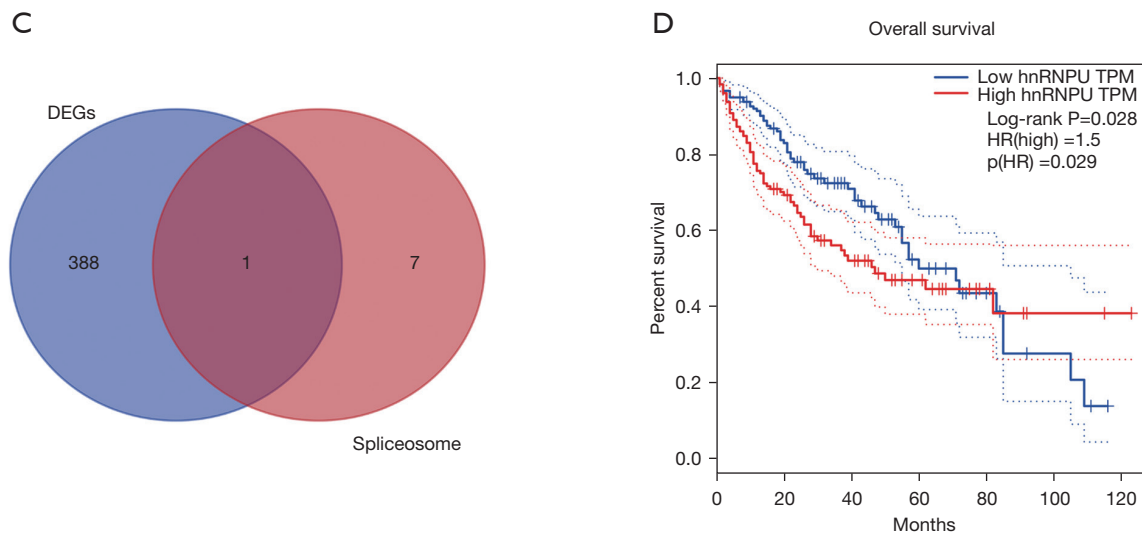


Figure 3 KEGG pathway enrichment analysis and identification of hnRNPU. (A) Bubble plots of the top 10 significantly enriched pathways; (B) KEGG pathway enrichment analysis of 123 genes; (C) 389 DEGs in the three GEO datasets were overlapped with eight genes associated with spliceosomes in the KEGG database to obtain the unique gene hnRNPU; (D) OS of HCC patients in the TCGA database was analyzed by Kaplan-Meier plot. KEGG, Kyoto Encyclopedia of Genes and Genomes; hnRNPU, heterogeneous nuclear ribonucleoprotein; DEGs, differentially expressed genes; GEO, Gene Expression Omnibus; OS, overall survival; HCC, hepatocellular carcinoma; TCGA, The Cancer Genome Atlas.

cell cycle control, differentiation, and apoptosis (20).

Abnormal splicing may lead to the production of abnormal mRNA isomers that encode mutant proteins, which have increased or decreased function and participate in cell transformation and the development and metastasis of cancer (21). The functions of spliceosome complexes and splicing modulators have been widely studied in cancer (22). In particular, splicing regulators such as snRNPs, hnRNPs, and SR proteins have been proven to play the role of carcinogenic or tumor suppressor proteins in diverse types of cancer (23,24), including breast (25) and lung cancer (26). There are several mechanisms underlying the aberrant splicing process in human cancers. Firstly, overexpression of uncontrolled splicing factors may lead to aberrant splicing incidents in tumors. The upregulation of SRSFs and hnRNPs has been demonstrated to be triggered by gene rearrangement and by copy number variation in multiple cancer types (27). For example, *HNRNPA2B1* gene amplification was identified in glioblastoma, and its copy number was negatively correlated with patient survival (28). Secondly, splice-regulated transcriptome processing changes represent a further mechanism for abnormal maturation of precursor mRNA. For example, the knockdown experiments of S6K2 kinase, which phosphorylates serine 6 residues of

hnRNP A1 protein, led to increased synthesis of PKM2 isoforms and improved glycolysis in the colorectal cancer cells (29). Furthermore, recurrent somatic mutations in genes encoding splicing factors have also been documented to affect the splicing process in cancer. For example, the point mutation and deletion of the *HNRNPK* gene, resulting in downregulation of *hnRNPK*, has been considered to play a role in the progression of acute myeloid leukemia (30). Therefore, abnormal splicing factors may directly lead to the occurrence of tumors or even promote the progression of tumors.

hnRNPU is an essential splicing regulator belonging to the hnRNP family, which is an RNA-binding protein family (RBPs) consisting of 20 major RBPs (31). Most hnRNPs, including *hnRNP A1/A2*, *hnRNPB1/B2*, *hnRNP E*, *hnRNP J*, and *hnRNP K*, are localized in the nucleus and shuttle to the cytoplasm through binding after the formation of homologous and allogeneic complexes (32,33). On the other hand, *hnRNPC* and *hnRNPU* have nuclear retention sequences that inhibit the transfer to the cytoplasm (34), and which contribute to a variety of aspects of nucleic acid metabolism, including selective splicing, mRNA stabilization, and regulation of transcription and translation. This study focused on potential cancer genes

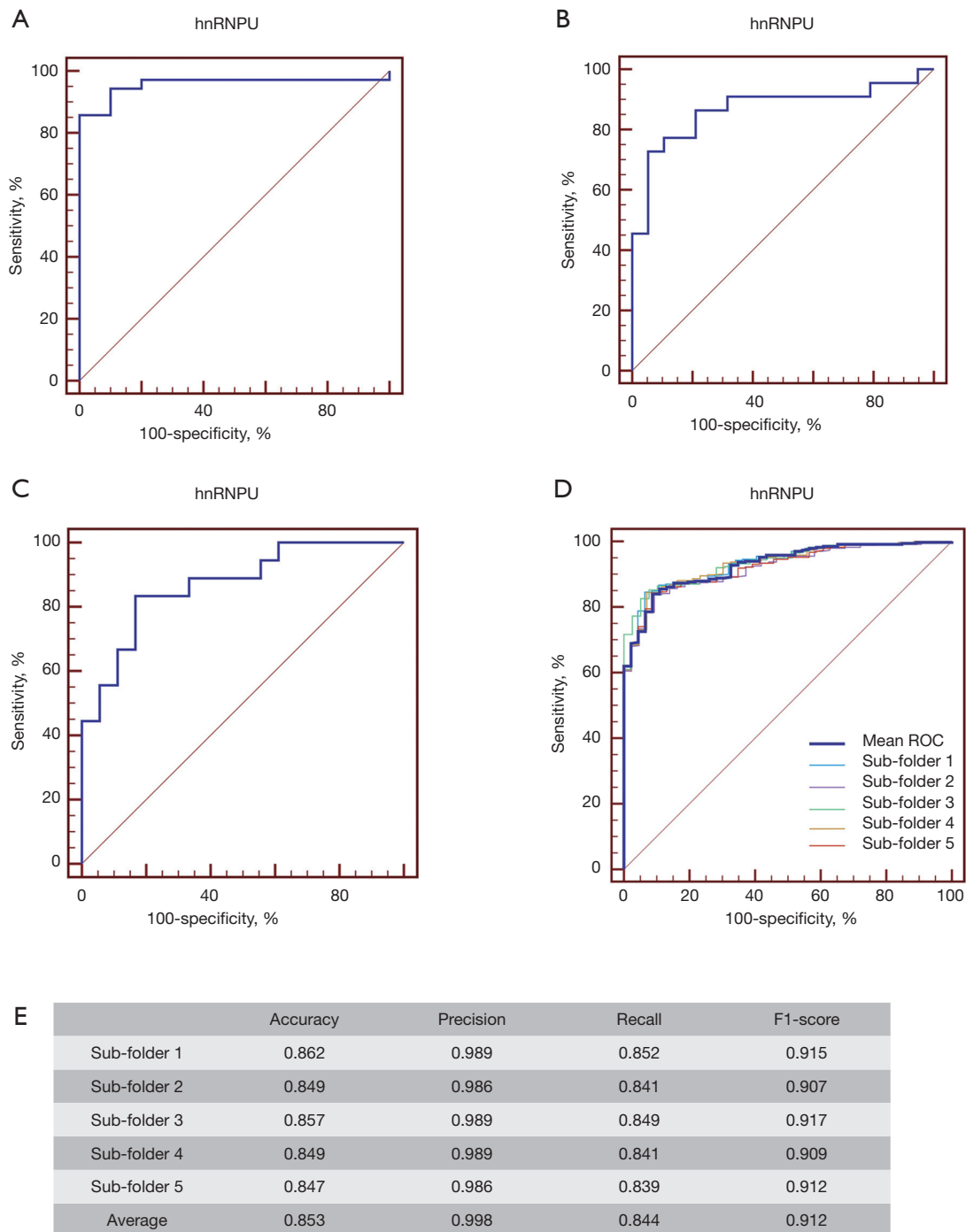


Figure 4 hnRNPU expression can be used as a diagnostic biomarker for HCC. (A-C) ROC curves of hnRNPU in GSE6764, GSE14520 and GSE60502 datasets; (D) ROC curves of hnRNPU in five-fold cross validation; (E) ROC evaluation of five validations parameter table. hnRNPU, heterogeneous nuclear ribonucleoprotein; HCC, hepatocellular carcinoma; ROC, receiver operating characteristic.

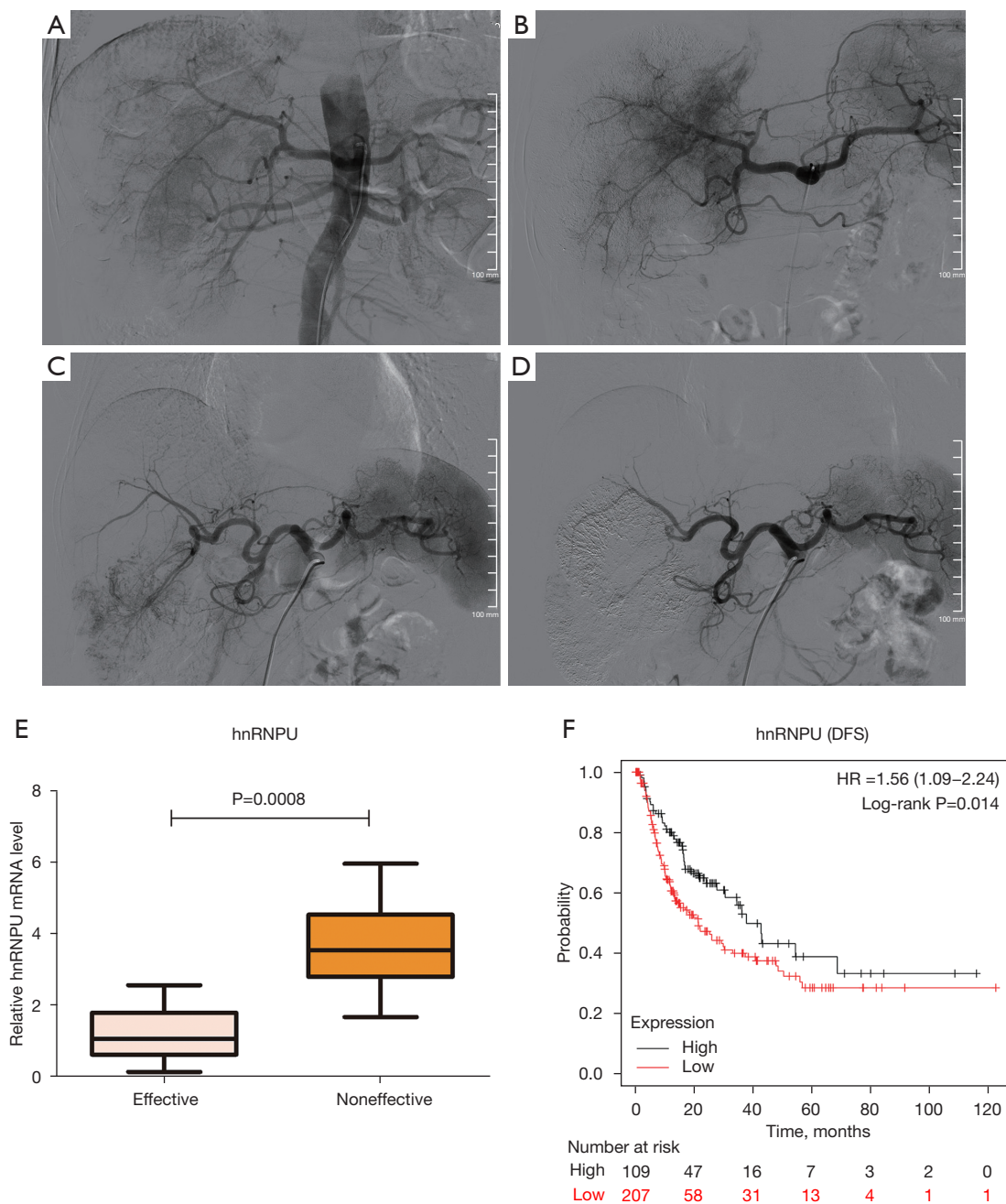


Figure 5 Expression of hnRNPU predicts postoperative treatment of HCC. (A) Patient A had a preoperative DSA trichocardiography of the abdomen, showing tumor staining in the right lobe of the liver; (B) 3 months after hepatic arterial chemoembolization, DSA peritoneal trunk imaging was performed for Patient A, and some tumor staining disappeared, but the tumor staining was still visible, indicating that the treatment effect was not good; (C) Patient B had a preoperative DSA trichocardiography of the abdomen, showing tumor staining in the right lobe of the liver; (D) 3 months after hepatic arterial chemoembolization for Patient B, DSA celiac trunk radiography showed that the tumor staining had basically disappeared, indicating that the treatment effect was good; (E) expression of hnRNPU in the tissues of well-treated and poorly treated patients after surgery; (F) DFS of HCC patients in the TCGA database was analyzed by Kaplan-Meier plot. hnRNPU, heterogeneous nuclear ribonucleoprotein; HCC, hepatocellular carcinoma; DSA, digital subtraction angiograph; DFS, disease-free survival; TCGA, The Cancer Genome Atlas.

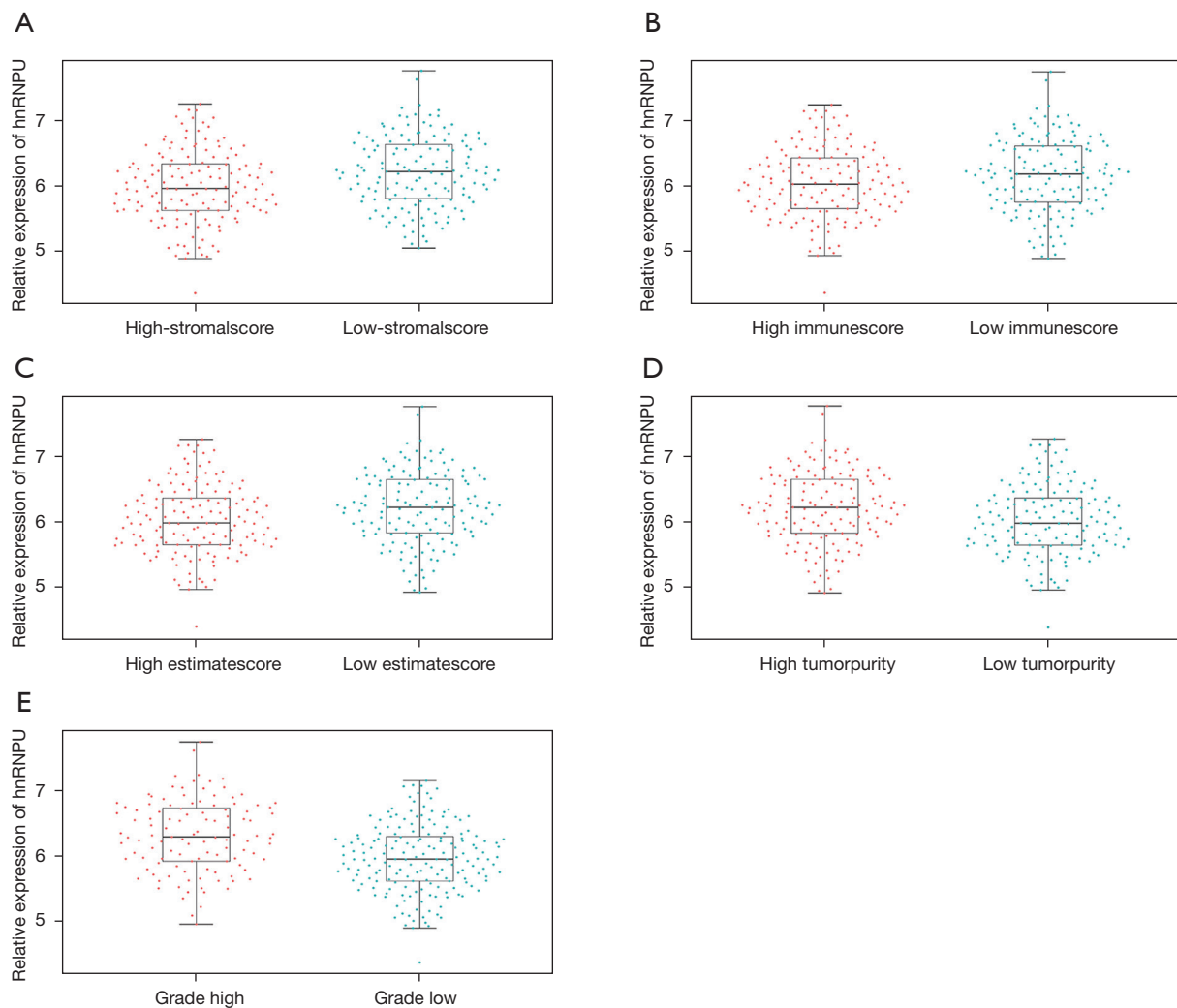


Figure 6 Expression of hnRNPU mRNA in clinical tissue samples of HCC. (A-C) The expression of hnRNPU in the interstitial scoring group, immune scoring group and assessment scoring group; (D) the expression of hnRNPU in tumor purity; (E) the expression of hnRNPU in histological grading of clinical samples. hnRNPU, heterogeneous nuclear ribonucleoprotein; mRNA, messenger RNA; HCC, hepatocellular carcinoma.

with diagnostic and prognostic value. We screened out 389 DEGs from three GEO datasets. At the same time, module analysis of HCC data in TGCA was performed using WGNCA to obtain blue-green modules significantly related to HCC. Then, enrichment analysis of module-related genes was performed using KEGG pathway, and 123 module-related genes were found to be significantly associated with spliceosome, RNA transport, mismatch repair, homologous recombination, and cell cycle, among which the spliceosome was the most significant pathway. We selected eight spliceosomal genes and 389 DEGs shared

genes and obtained the only gene, hnRNPU.

hnRNPU is a 120 kDa versatile protein which regulates the pre-mRNA splicing process via direct binding to target genes, or protein/protein interaction-mediated splicing (35,36). Furthermore, the actin-hnRNPU complex is a critical regulator of the beginning stage of transcriptional activation in eukaryotic cells (37). The hnRNPU and c-Myc regulatory loop have combined effects on the proliferation and self-renewal of HCC and promote its progression (38). Long non-coding RNA H19 also combines with hnRNPU and represses RNA polymerase II-mediated transcription by

destroying the actin-hnRNPU complex (39). In the present study, we found that the splicing regulator hnRNPU was significantly and highly expressed in HCC tissues and was significantly linked to poor prognostic survival in HCC patients ($P < 0.05$).

Liver cancer is an aggressive tumor, which frequently occurs in patients with underlying chronic liver disease such as chronic hepatitis B virus infection and cirrhosis. Therefore, early diagnosis of liver cancer plays an important role in controlling disease progression and prolonging survival time. The main clinical screening methods for HCC are histopathology, imaging techniques, and alpha-fetoprotein (AFP), which is a major plasma protein produced by the yolk sac and the liver (40). But liver biopsy is invasive and may increase the risk of puncture metastasis; imaging techniques usually only detect tumors larger than 1 cm in diameter (41); and AFP is the most widely used noninvasive biomarker in clinical practice, but it lacks sufficient sensitivity and specificity (42). In this study, we performed ROC curve analysis on GSE6764, GSE14520, and GSE60502 datasets to evaluate the sensitivity and specificity of hnRNPU for HCC diagnosis. The analysis results suggested that the average AUC of the three GEO datasets was 0.957, 0.868, and 0.870, indicating good sensitivity and specificity. hnRNPU had a good diagnostic value for differentiating HCC patients from normal control samples, suggesting that hnRNPU is a potential biomarker for the diagnosis of HCC. Interestingly, we also found that hnRNPU mRNA was lower in the tissues of patients with excellent HCC treatment than in the tissues of patients with poor HCC treatment ($P < 0.05$). The result demonstrated that hnRNPU can predict the therapeutic effect of HCC.

In HCC, the tumor microenvironment plays a critical role as both a positively and a negatively regulating factor of tumor signaling. The infiltrating stromal cells and immune cells play a key role in tumor development, and a comprehensive understanding of them can provide important perspectives on tumor progression and prognosis. In our work, the expression of hnRNPU was further verified in the TCGA database, and the ESTIMATE analysis demonstrated that hnRNPU was highly expressed in tumor tissues compared to normal liver tissues, particularly in patients with low stromal, low immune, and low estimation scores. Moreover, the expression level of hnRNPU was positively correlated with tumor purity. Further studies revealed that the level of hnRNPU expression gradually increased with the malignant progression of HCC ($P < 0.05$). Therefore, we speculate that hnRNPU may play an

essential role in the pathogenesis and progression of HCC. The expression of hnRNPU was negatively correlated with the stromal immune score, suggesting that hnRNPU may be one of the tumor microenvironment-related genes affecting the recruitment of infiltrating stromal cells and immune cells in the tumor microenvironment of HCC. The relationship between the tumor microenvironment and hnRNPU may influence the efficacy of radiotherapy. However, the network of interactions between hnRNPU and infiltrating stromal cells and immune cells needs to be further investigated.

In summary, our highlight is the clinical application of hnRNPU. We conducted a study on hnRNPU in the clinical diagnosis and prognosis of therapeutic efficacy of HCC and found that hnRNPU is a very significant gene. Zhang *et al.* conducted an in-depth study on the molecular mechanism of this gene inside tumors and found that hnRNPU promotes tumor progression by regulating the proliferation of hepatocellular carcinoma cells through the downstream gene c-Myc (38), again suggesting that hnRNPU is a key gene. In our future study, we will explore the molecular mechanism of hnRNPU in HCC in depth.

Acknowledgments

Funding: This study was funded by a Qiqihar Academy of Medical Sciences Project Grant (QMSI2020M-10).

Footnote

Reporting Checklist: The authors have completed the STARD reporting checklist. Available at <https://dx.doi.org/10.21037/jgo-21-468>

Data Sharing Statement: Available at <https://dx.doi.org/10.21037/jgo-21-468>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://dx.doi.org/10.21037/jgo-21-468>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. All procedures performed in this study involving human participants were in accordance with the Declaration of Helsinki (as revised

in 2013). The study was approved by institutional ethics board of The Second Affiliated Hospital of Qiqihar Medical College (No.: 2020-01) and informed consent was taken from all the patients.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Omata M, Cheng AL, Kokudo N, et al. Asia-Pacific clinical practice guidelines on the management of hepatocellular carcinoma: a 2017 update. *Hepatol Int* 2017;11:317-70.
2. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394-424.
3. Chabot B, Shkreta L. Defective control of pre-messenger RNA splicing in human disease. *J Cell Biol* 2016;212:13-27.
4. Lokody I. RNA dynamics: destabilizing mRNAs promotes metastasis. *Nat Rev Cancer* 2014;14:578.
5. Climente-González H, Porta-Pardo E, Godzik A, et al. The functional impact of alternative splicing in cancer. *Cell Rep* 2017;20:2215-26.
6. Sebestyén E, Singh B, Miñana B, et al. Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. *Genome Res* 2016;26:732-44.
7. Lee SC, Abdel-Wahab O. Therapeutic targeting of splicing in cancer. *Nat Med* 2016;22:976-86.
8. Salton M, Misteli T. Small molecule modulators of pre-mRNA splicing in cancer therapy. *Trends Mol Med* 2016;22:28-37.
9. Chen H, Gao F, He M, et al. Long-read RNA sequencing identifies alternative splice variants in hepatocellular carcinoma and tumor-specific isoforms. *Hepatology* 2019;70:1011-25.
10. Li S, Hu Z, Zhao Y, et al. Transcriptome-wide analysis reveals the landscape of aberrant alternative splicing events in liver cancer. *Hepatology* 2019;69:359-75.
11. Luo C, Cheng Y, Liu Y, et al. SRSF2 regulates alternative splicing to drive hepatocellular carcinoma development. *Cancer Res* 2017;77:1168-78.
12. Sen S, Langiewicz M, Jumaa H, et al. Deletion of serine/arginine-rich splicing factor 3 in hepatocytes predisposes to hepatocellular carcinoma in mice. *Hepatology* 2015;61:171-83.
13. Shilo A, Ben Hur V, Denichenko P, et al. Splicing factor hnRNP A2 activates the Ras-MAPK-ERK pathway by controlling A-Raf splicing in hepatocellular carcinoma development. *RNA* 2014;20:505-15.
14. Yang X, Qu S, Wang L, et al. PTBP3 splicing factor promotes hepatocellular carcinoma by destroying the splicing balance of NEAT1 and pre-miR-612. *Oncogene* 2018;37:6399-413.
15. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559.
16. Liu J, Jing L, Tu X. Weighted gene co-expression network analysis identifies specific modules and hub genes related to coronary artery disease. *BMC Cardiovasc Disord* 2016;16:54.
17. Shi Y. Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat Rev Mol Cell Biol* 2017;18:655-70.
18. Bush SJ, Chen L, Tovar-Corona JM, et al. Alternative splicing and the evolution of phenotypic novelty. *Philos Trans R Soc Lond B Biol Sci* 2017. doi: 10.1098/rstb.2015.0474.
19. Wang Y, Liu J, Huang BO, et al. Mechanism of alternative splicing and its regulation. *Biomed Rep* 2015;3:152-8.
20. Baralle FE, Giudice J. Alternative splicing as a regulator of development and tissue identity. *Nat Rev Mol Cell Biol* 2017;18:437-51.
21. Oltean S, Bates DO. Hallmarks of alternative splicing in cancer. *Oncogene* 2014;33:5311-8.
22. Li H, Liu J, Shen S, et al. Pan-cancer analysis of alternative splicing regulator heterogeneous nuclear ribonucleoproteins (hnRNPs) family and their prognostic potential. *J Cell Mol Med* 2020;24:11111-9.
23. Dvinge H, Kim E, Abdel-Wahab O, et al. RNA splicing factors as oncoproteins and tumour suppressors. *Nat Rev Cancer* 2016;16:413-30.
24. Cheng Z, Sun Y, Niu X, et al. Gene expression profiling reveals U1 snRNA regulates cancer gene expression. *Oncotarget* 2017;8:112867-74.
25. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses

- with clinical implications. *Proc Natl Acad Sci U S A* 2001;98:10869-74.
26. Köhler J, Schuler M, Gauler TC, et al. Circulating U2 small nuclear RNA fragments as a diagnostic and prognostic biomarker in lung cancer patients. *J Cancer Res Clin Oncol* 2016;142:795-805.
 27. Shilo A, Siegfried Z, Karni R. The role of splicing factors in deregulation of alternative splicing during oncogenesis and tumor progression. *Mol Cell Oncol* 2015;2:e970955.
 28. Golan-Gerstl R, Cohen M, Shilo A, et al. Splicing factor hnRNP A2/B1 regulates tumor suppressor gene splicing and is an oncogenic driver in glioblastoma. *Cancer Res* 2011;71:4464-72.
 29. Sun Y, Luo M, Chang G, et al. Phosphorylation of Ser6 in hnRNPA1 by S6K2 regulates glucose metabolism and cell growth in colorectal cancer. *Oncol Lett* 2017;14:7323-31.
 30. Gallardo M, Lee HJ, Zhang X, et al. hnRNP K Is a Haploinsufficient Tumor Suppressor that Regulates Proliferation and Differentiation Programs in Hematologic Malignancies. *Cancer Cell* 2015;28:486-99.
 31. Geuens T, Bouhy D, Timmerman V. The hnRNP family: insights into their role in health and disease. *Hum Genet* 2016;135:851-67.
 32. Siomi H, Dreyfuss G. A nuclear localization domain in the hnRNP A1 protein. *J Cell Biol* 1995;129:551-60.
 33. Mili S, Shu HJ, Zhao Y, et al. Distinct RNP complexes of shuttling hnRNP proteins with pre-mRNA and mRNA: candidate intermediates in formation and export of mRNA. *Mol Cell Biol* 2001;21:7307-19.
 34. Nakielnny S, Dreyfuss G. The hnRNP C proteins contain a nuclear retention sequence that can override nuclear export signals. *J Cell Biol* 1996;134:1365-73.
 35. Yugami M, Kabe Y, Yamaguchi Y, et al. hnRNP-U enhances the expression of specific genes by stabilizing mRNA. *FEBS Lett* 2007;581:1-7.
 36. Kiledjian M, Dreyfuss G. Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *EMBO J* 1992;11:2655-64.
 37. Kukulev A, Nord Y, Palmberg C, et al. Actin and hnRNP U cooperate for productive transcription by RNA polymerase II. *Nat Struct Mol Biol* 2005;12:238-44.
 38. Zhang B, Wang HY, Zhao DX, et al. The splicing regulatory factor hnRNPU is a novel transcriptional target of c-Myc in hepatocellular carcinoma. *FEBS Lett* 2021;595:68-84.
 39. Bi HS, Yang XY, Yuan JH, et al. H19 inhibits RNA polymerase II-mediated transcription by disrupting the hnRNP U-actin complex. *Biochim Biophys Acta* 2013;1830:4899-906.
 40. Benson AB 3rd, D'Angelica MI, Abbott DE, et al. NCCN Guidelines Insights: Hepatobiliary Cancers, Version 1.2017. *J Natl Compr Canc Netw* 2017;15:563-73.
 41. Roberts LR, Sirlin CB, Zaiem F, et al. Imaging for the diagnosis of hepatocellular carcinoma: A systematic review and meta-analysis. *Hepatology* 2018;67:401-21.
 42. Farinati F, Marino D, De Giorgio M, et al. Diagnostic and prognostic role of alpha-fetoprotein in hepatocellular carcinoma: both or neither? *Am J Gastroenterol* 2006;101:524-32.
- (English Language Editor: B. Meiser)

Cite this article as: Du Y, Ma X, Wang D, Wang Y, Zhang T, Bai L, Liu Y, Chen S. Identification of heterogeneous nuclear ribonucleoprotein as a candidate biomarker for diagnosis and prognosis of hepatocellular carcinoma. *J Gastrointest Oncol* 2021;12(5):2361-2376. doi: 10.21037/jgo-21-468

Table S1 Three hundred and eighty-nine common DEGs

No.	Gene	No.	Gene	No.	Gene	No.	Gene	No.	Gene
1	AACS	79	PIK3C2G	157	HNRNPA1P10	235	SNRPA	313	COLEC10
2	MEFV	80	CLEC4M	158	LOC100506699	236	CS	314	USP21
3	CYP26A1	81	NPY1R	159	SAA4	237	ETS2	316	JUNB
4	BBOX1	82	CYP4F12	160	PLPP3	238	INTS8	317	ALDH8A1
5	ECHS1	83	GCH1	161	SUCO	239	CASC3	318	AZGP1
6	HSPB1	84	MELK	162	KCNN2	240	MT1M	319	PAMR1
7	KLHL2	85	PODXL	163	LIPC	241	C12orf10	320	GNMT
8	CYP39A1	86	PROS1	164	GSTZ1	242	PABPC3	321	CXCL12
9	CHD8	87	ZWINT	165	OLFML3	243	EGR1	322	RRM2
10	C1RL	88	NDC80	166	FXYD1	244	ECM1	323	MDK
11	TPX2	89	GPR182	167	SARAF	245	MPC1	324	ABAT
12	PON3	90	NSUN5	168	FOSB	246	ZCCHC24	325	ADK
13	ZG16	91	ARHGAP10	169	IL1B	247	MEA1	326	HMGCL
14	AGPAT1	92	MRC1	170	MARCO	248	FCN2	327	ACAA1
15	PCOLCE	93	GPM6A	171	CTSA	249	COG2	328	SMG5
16	SORL1	94	PPP4R1	172	DNM1L	250	UBR5	329	EDEM1
17	GINS1	95	PTTG1	173	IGFBP3	251	SLC22A1	330	ZFYVE26
18	ACSM3	96	TARBP1	174	PLVAP	252	IL13RA2	331	BUB1B
19	CYP2C19	97	STK39	175	WDYHV1	253	C11orf71	332	LILRB5
20	ZNF623	98	GOLGA2P7	176	LOC101930591	254	YY1AP1	333	PAIP2B
21	NOL8	99	APOF	177	RRP15	255	ASPA	334	MAT1A
22	KPNA2	100	CKS2	178	CIDEB	256	PCNA	335	ADAMTS13
23	SMYD3	101	BSG	179	LRAT	257	MAN1C1	336	ITIH1
24	LOC727751	102	LOC101929479	180	LOC101927180	258	ATP6AP1	337	NVL
25	GHR	103	SLC38A6	181	NAT1	259	ROBO1	338	CAP2
26	CLEC1B	104	GRAMD1C	182	HGFAC	260	KBTBD11	339	GNA14
27	STARD5	105	CD14	183	MT2A	261	TSNAX	340	IL33
28	KNTC1	106	FAM21A	184	CDA	262	VSIG4	341	LOC100509620
29	GTF2IP4	107	DUSP12	185	HNRNPU	263	RND3	342	HAAO
30	ALDH2	108	FAM20B	186	USP39	264	RDH16	343	PEMT
31	NELFE	109	ATP6V1F	187	PRCC	265	AATF	344	CDK5
32	SHBG	110	CD302	188	LPCAT1	266	CDC20	345	HMMR
33	BCHE	111	FNDC4	189	CYR61	267	UBD	346	LY6E
34	RACGAP1	112	CENPE	190	HNRNPA1L2	268	SUN2	347	PIP
35	GPD1	113	ACADM	191	CCT6A	269	MT1X	348	FCN3

Table S1 (continued)

Table S1 (continued)

No.	Gene	No.	Gene	No.	Gene	No.	Gene	No.	Gene
36	<i>CRHBP</i>	114	<i>MT1E</i>	192	<i>NDEL1</i>	270	<i>RFC4</i>	349	<i>PTPN3</i>
37	<i>ASS1</i>	115	<i>MIR1236</i>	193	<i>SERPINA10</i>	271	<i>SERPINF2</i>	350	<i>GBA3</i>
38	<i>HMGNA4</i>	116	<i>IL10</i>	194	<i>ACSL4</i>	272	<i>PLOD3</i>	351	<i>GMNN</i>
39	<i>BDH2</i>	117	<i>ANXA10</i>	195	<i>TP53BP2</i>	273	<i>MGMT</i>	352	<i>SCRIB</i>
40	<i>PPBP</i>	118	<i>TBCE</i>	196	<i>LOC101930168</i>	274	<i>PBK</i>	353	<i>CYP2B6</i>
41	<i>SLC38A4</i>	119	<i>GYS2</i>	197	<i>ALPL</i>	275	<i>F8</i>	354	<i>KIF20A</i>
42	<i>ACADL</i>	120	<i>NUPL2</i>	198	<i>C1QTNF1</i>	276	<i>SLC28A1</i>	355	<i>HPX</i>
43	<i>DHX16</i>	121	<i>SEC24B</i>	199	<i>GOLPH3L</i>	277	<i>LOC642423</i>	356	<i>ANKRD27</i>
44	<i>CLEC3B</i>	122	<i>PLGLB2</i>	200	<i>C9</i>	278	<i>DROSHA</i>	357	<i>SF3B4</i>
45	<i>GLA</i>	123	<i>LYVE1</i>	201	<i>FANCG</i>	279	<i>RFXANK</i>	358	<i>GTF2IRD1</i>
46	<i>CNOT6</i>	124	<i>PRC1</i>	202	<i>CBFA2T3</i>	280	<i>MOGAT2</i>	359	<i>TTPAL</i>
47	<i>SAC3D1</i>	125	<i>ABCF1</i>	203	<i>FAM50A</i>	281	<i>LAMC1</i>	360	<i>MT1H</i>
48	<i>SRD5A2</i>	126	<i>MUT</i>	204	<i>DONSON</i>	282	<i>ACADS</i>	361	<i>MFSD5</i>
49	<i>PBLD</i>	127	<i>NAT2</i>	205	<i>E2F8</i>	283	<i>EPHX2</i>	362	<i>MMRN1</i>
50	<i>KIF14</i>	128	<i>AGL</i>	206	<i>CXCL14</i>	284	<i>ITGB3BP</i>	363	<i>AKR1C3</i>
51	<i>PXMP2</i>	129	<i>GNE</i>	207	<i>C1R</i>	285	<i>KLRF1</i>	364	<i>VIPR1</i>
52	<i>KLKB1</i>	130	<i>MIR224</i>	208	<i>C8A</i>	286	<i>RBKS</i>	365	<i>PLAC8</i>
53	<i>MT1G</i>	131	<i>PVALB</i>	209	<i>RFX5</i>	287	<i>CD34</i>	366	<i>MIR6866</i>
54	<i>ALG6</i>	132	<i>ACSL1</i>	210	<i>TOPBP1</i>	288	<i>P2RY13</i>	367	<i>CFP</i>
55	<i>RIC3</i>	133	<i>NCOR1</i>	211	<i>KLF11</i>	289	<i>DIRAS3</i>	368	<i>HNRNPA1P33</i>
56	<i>PLSCR4</i>	134	<i>UXS1</i>	212	<i>SLCO1B3</i>	290	<i>RMI1</i>	369	<i>ANKRD55</i>
57	<i>SCAMP3</i>	135	<i>CCT3</i>	213	<i>MCRS1</i>	291	<i>MYD88</i>	370	<i>PTH1R</i>
58	<i>TTC13</i>	136	<i>SLC25A20</i>	214	<i>STAM</i>	292	<i>SKAP1</i>	371	<i>ADAMTSL3</i>
59	<i>MIR452</i>	137	<i>GCKR</i>	215	<i>FCGR2B</i>	293	<i>PDE2A</i>	372	<i>TTK</i>
60	<i>ZFP36</i>	138	<i>NTF3</i>	216	<i>LYRM1</i>	294	<i>UBE2S</i>	373	<i>DDX41</i>
61	<i>DDX39A</i>	139	<i>CDK7</i>	217	<i>MIR6824</i>	295	<i>TBCC</i>	374	<i>ZGPAT</i>
62	<i>CCL14</i>	140	<i>LIG1</i>	218	<i>TMEM267</i>	296	<i>PRG4</i>	375	<i>CDKN3</i>
63	<i>KIF4A</i>	141	<i>DHRS4L2</i>	219	<i>FOLH1B</i>	297	<i>LY75-CD302</i>	376	<i>CYP2J2</i>
64	<i>PGM1</i>	142	<i>FBP1</i>	220	<i>MIR6778</i>	298	<i>CYP1A2</i>	377	<i>NCAPG</i>
65	<i>RCAN1</i>	143	<i>MCC</i>	221	<i>OSTF1</i>	299	<i>ACADSB</i>	378	<i>CUTA</i>
66	<i>THY1</i>	144	<i>CETP</i>	222	<i>COL15A1</i>	300	<i>PXDC1</i>	379	<i>VRK1</i>
67	<i>GTF2IP1</i>	145	<i>ADIRF</i>	223	<i>EZH2</i>	301	<i>MSH5</i>	380	<i>PRODH2</i>
68	<i>OSBPL3</i>	146	<i>MT1HL1</i>	224	<i>PIAS3</i>	302	<i>ATAD2</i>	381	<i>BOLA2B</i>
69	<i>DCAF11</i>	147	<i>C21orf33</i>	225	<i>ARMC1</i>	303	<i>DNMT1</i>	382	<i>GLS2</i>
70	<i>CBX1</i>	148	<i>SRPX</i>	226	<i>MRGBP</i>	304	<i>CDK5RAP1</i>	383	<i>C6</i>

Table S1 (continued)

Table S1 (*continued*)

No.	Gene	No.	Gene	No.	Gene	No.	Gene	No.	Gene
71	<i>PTGS2</i>	149	<i>DLGAP5</i>	227	<i>STEAP3</i>	305	<i>CFI</i>	384	<i>NUSAP1</i>
72	<i>MIR6743</i>	150	<i>RAD51AP1</i>	228	<i>DNASE1L3</i>	306	<i>ETFDH</i>	385	<i>BCKDHB</i>
73	<i>GABRE</i>	151	<i>DHRS4</i>	229	<i>KIF11</i>	307	<i>LPA</i>	386	<i>FAM21C</i>
74	<i>AZGP1P1</i>	152	<i>COL4A1</i>	230	<i>C8orf4</i>	308	<i>CD5L</i>	387	<i>HBA2</i>
75	<i>GCGR</i>	153	<i>ICAM3</i>	231	<i>H3F3AP4</i>	309	<i>C8B</i>	388	<i>HBA1</i>
76	<i>ITGA6</i>	154	<i>HAO2</i>	232	<i>BCAP31</i>	310	<i>STIL</i>	389	<i>AQP7</i>
77	<i>SLC16A2</i>	155	<i>PLGLB1</i>	233	<i>IGFALS</i>	311	<i>UBE2C</i>		
78	<i>RIC8A</i>	156	<i>MT1F</i>	234	<i>GDI1</i>	312	<i>RBM42</i>		

DEGs, differentially expressed genes.

Table S2 One hundred and twenty-third turquoise module-related genes

No.	Gene	No.	Gene	No.	Gene	No.	Gene	No.	Gene
1	<i>AC027237.1</i>	26	<i>CPSF6</i>	51	<i>KHDC4</i>	76	<i>POLA1</i>	101	<i>SUZ12</i>
2	<i>ADNP</i>	27	<i>CSNK2A1</i>	52	<i>KHDRBS1</i>	77	<i>PPHLN1</i>	102	<i>TAF3</i>
3	<i>ANKRD52</i>	28	<i>DBF4</i>	53	<i>KIAA1841</i>	78	<i>PRKCI</i>	103	<i>TAF4</i>
4	<i>ANLN</i>	29	<i>DBF4B</i>	54	<i>KIF18A</i>	79	<i>PRRC2C</i>	104	<i>TARDBP</i>
5	<i>ARHGAP11A</i>	30	<i>DDX23</i>	55	<i>KIF20B</i>	80	<i>QRICH1</i>	105	<i>TASOR</i>
6	<i>ARMC8</i>	31	<i>DDX55</i>	56	<i>KPNB1</i>	81	<i>QSER1</i>	106	<i>TCERG1</i>
7	<i>ATAD5</i>	32	<i>E2F6</i>	57	<i>LARP4B</i>	82	<i>RACGAP1</i>	107	<i>TDG</i>
8	<i>ATXN7L3</i>	33	<i>ECT2</i>	58	<i>LIN9</i>	83	<i>RAD51AP1</i>	108	<i>TET3</i>
9	<i>BARD1</i>	34	<i>ERCC3</i>	59	<i>MASTL</i>	84	<i>RALGAPB</i>	109	<i>TOPBP1</i>
10	<i>BCOR</i>	35	<i>FBXO11</i>	60	<i>MAU2</i>	85	<i>RBL1</i>	110	<i>U2AF2</i>
11	<i>BRPF1</i>	36	<i>GAS2L3</i>	61	<i>MCM8</i>	86	<i>RBMX</i>	111	<i>U2SURP</i>
12	<i>CAPRIN1</i>	37	<i>GGA3</i>	62	<i>MSH2</i>	87	<i>RIF1</i>	112	<i>UBA2</i>
13	<i>CASP2</i>	38	<i>GPSM2</i>	63	<i>MTBP</i>	88	<i>SAP130</i>	113	<i>VEZF1</i>
14	<i>CBFA2T2</i>	39	<i>GTF3C3</i>	64	<i>NCAPD2</i>	89	<i>SASS6</i>	114	<i>WIPF2</i>
15	<i>CBX1</i>	40	<i>HAUS3</i>	65	<i>NCAPG2</i>	90	<i>SGO2</i>	115	<i>XPO1</i>
16	<i>CCDC93</i>	41	<i>HNRNPA2B1</i>	66	<i>NCOA5</i>	91	<i>SMARCC1</i>	116	<i>XRCC2</i>
17	<i>CENPF</i>	42	<i>HNRNPH1</i>	67	<i>NEMP1</i>	92	<i>SMARCD1</i>	117	<i>YEATS2</i>
18	<i>CENPL</i>	43	<i>HNRNPU</i>	68	<i>NEPRO</i>	93	<i>SMC3</i>	118	<i>ZBTB12</i>
19	<i>CEP250</i>	44	<i>HNRNPUL1</i>	69	<i>NRF1</i>	94	<i>SMC4</i>	119	<i>ZNF207</i>
20	<i>CEP295</i>	45	<i>HSPA14</i>	70	<i>NSD2</i>	95	<i>SNRNP200</i>	120	<i>ZNF248</i>
21	<i>CIP2A</i>	46	<i>HUWE1</i>	71	<i>NUP107</i>	96	<i>SPATS2</i>	121	<i>ZNF765</i>
22	<i>CKAP2L</i>	47	<i>ILF3</i>	72	<i>PAXIP1</i>	97	<i>SPDL1</i>	122	<i>ZSCAN25</i>
23	<i>CNNM4</i>	48	<i>IQCB1</i>	73	<i>PHF20</i>	98	<i>SRSF1</i>	123	<i>ZWILCH</i>
24	<i>CNOT6</i>	49	<i>KANSL1</i>	74	<i>PLAGL2</i>	99	<i>STAM</i>		
25	<i>COMMD2</i>	50	<i>KCTD7</i>	75	<i>POGK</i>	100	<i>STIL</i>		