Am J Cancer Res 2023;13(3):802-817
www.ajcr.us /ISSN:2156-6976/ajcr0149073

Original Article

The role of the stemness index-associated signature in the analysis of the tumorigenesis of liver cancer patients of different races

Da-Hua Liu[1*], Gui-Min Wen[2*], Chang-Liang Song[3], Ze-Jun Xu[1], Fu Ren[4], Zhen-Ying Zhao[5], Pu Xia[1]

[1]Biological Anthropology Institute, College of Basic Medical Science, Jinzhou Medical University, Jinzhou, Liaoning, P. R. China; [2]Department of Basic Nursing, College of Nursing, Jinzhou Medical University, Jinzhou, Liaoning, P. R. China; [3]Department of Radiotherapy, Center Hospital of Handan, Handan, Hebei, P. R. China; [4]Shenyang Medical College, Shenyang, Liaoning, P. R. China; [5]Department of Pharmacy, Tianjin Union Medical Center, Tianjin, P. R. China. [*]Equal contributors.

Received January 11, 2023; Accepted February 25, 2023; Epub March 15, 2023; Published March 30, 2023

Abstract: Cancer stem cells (CSCs) are a subset of cancer cells with stem cell characteristics. The discovery of CSCs has opened a new era for cancer research. CSCs not only play a critical role in tumorigenesis, but also are responsible for the failure of cancer treatments. Here, we performed weighted gene co-expression network analysis (WGCNA) to identify key stemness genes and prognostic signatures using the data of an Asian liver cancer patient cohort and a White liver cancer patient cohort in The Cancer Genome Atlas (TCGA) database. To compare the difference in tumorigenesis between the Asian patients and the White patients, the prognostic value of the key genes from the Asian patients was evaluated in the White patient cohort and vice versa. We found that some key genes could predict the survival of the patients regardless of race. In addition, the key genes, NUCB2 and KLF4A, were selected from Asian patients and White patients, respectively, for further experimental validation. Knocking down NUCB2 could inhibit the activity of the AKT/mTOR signaling pathway and reverse the epithelial-mesenchymal transition (EMT) in liver cancer cells. We also confirmed that the knockdown of KLF4A suppressed ABCG2 activity and reduced the side population (SP) in liver cancer cells for the first time. Our results suggest that the stemness index is a useful method to identify key genes in tumorigenesis. Compared to the analysis for all patients, applying this index to the analysis of the patients of different races will provide more potential therapeutic targets for cancer treatment.

Keywords: Stemness index, liver cancer, cancer stem cell, tumorigenesis, prognosis

Introduction

Liver cancer is one of the most common malignant tumors and ranks as the second leading cause of cancer-related deaths [1]. Patients with primary liver cancer have high malignancy and a short survival time [2]. However, the etiology and the exact molecular mechanism underlying the development of primary liver cancer are still unclear. The pathogenesis of primary liver cancer is a complex process, which is affected by both environmental and dietary factors [3]. In recent years, cancer stem cell theory has attracted increasing attention [4]. Cancer stem cells (CSCs), a subpopulation of cancer cells, have the ability of self-renewal and can produce heterogeneous cancer cells [5]. It is now clear that cancer stem cells play an important role in tumor progression and in the therapeutic resistance of liver cancer [6].

In 2018, Malta et al. [7] proposed the "stemness index" to quantify tumor cells with stem cell characteristics in tumor tissues using the data from The Cancer Genome Atlas (TCGA) database. Two independent "stemness indexes" were generated by using the innovative one-class logistic regression (OCLR) machine learning algorithm [8]. One is a DNA methylation-based stemness index (mDNAsi), which reflects epigenetic characteristics, while the other is an mRNA expression-based stemness index (mRNAsi), which reflects gene expression [8]. Previous studies have shown that the stem-

ness index correlates with the survival of liver cancer patients [9, 10]. However, to our knowledge, no study has compared the mRNAsi-related genes between Asian liver cancer patients and White liver cancer patients.

In this study, weighted gene co-expression network analysis (WGCNA) was performed to screen key mRNAsi-related genes in Asian and White liver cancer patients. Furthermore, we experimentally validated the functions and mechanisms of two key genes, NUCB2 and KLF4A, selected from Asian and White liver cancer patient cohorts, respectively, in liver cancer cells.

Materials and methods

Data acquisition and preprocessing

The gene expression profiling and the patient information of 345 liver cancer samples and 40 paracancerous samples were downloaded from The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov) [11] (Supplementary Figure 1A). The gene expression data were downloaded as formats "HTSeq-FPKM" (HTSeq-Counts and Fragments Per Kilobase per Million (FPKM)). The downloaded clinical data included the demographics, clinicopathological information, and follow-up data of all patients. The stemness index data for liver cancer were acquired from the study published by Malta et al. [7]. Perl software was utilized to integrate the RNA-seq value of each specimen into a matrix file. Gene symbols were obtained from the Ensembl (http://asia.ensem/index.html) database based on the Ensembl ID and transformed into the matrix configuration file.

Identification of differentially expressed genes (DEGs)

The EdgeR method was applied to identify the DEGs in groups that were divided by age (≤ 65 and > 65), sex (male and female), tumor stage (I-IV), and tumor grades (G1-G3), respectively. A gene with log2 Fold Change (FC) > 1.0 or < -1.0 and a False Discovery Rate (FDR) value < 0.05 was defined as a DEG.

Weighted gene correlation network analysis (WGCNA)

The WGCNA analysis was used to perform co-expression scale-free network analysis and identify gene modules containing strongly correlating genes by the R programming "WGCNA" package. Briefly, DEGs combined with stemness index data were imported into R programming. Based on the correlations of each DEG, the β value was determined to construct scale-free co-expression networks and to detect modules. Furthermore, the relationship between the modules and the mRNAsi score was analyzed.

Key gene analysis

The differential expression of each key gene between the normal and the tumor tissues were visualized in a heatmap. The R programming "corrplot" package was used to analyze and visualize the Pearson correlations among the stemness-related genes based on gene expression levels. The network and the functions of the key genes were predicted and visualized via GeneMANIA (http://genemania.org/) [12].

Cox regression analysis

The R programming "forestplot" package was used to display the results of the univariate and multivariate Cox regression analyses. A nomogram was developed based on the results of the multivariate Cox proportional hazards analysis to predict the 1-, 3-, and 5-year overall survival using the R programming "rms" package.

Survival analysis

The Kaplan-Meier (KM) survival analysis with log-rank test was used to compare the survival difference between the patients with high- and low-expression levels of the key genes using the R programming "survival" and "survminer".

Cell culture and siRNA transfection

SMMC-7721 cells were obtained from the Cell Bank of the Chinese Academy of Sciences (Shanghai, China), and HepG2 cells were obtained from the American Type Culture Collection (Manassas, VA, USA). The cells were maintained at 37°C in a 5% CO_2 incubator in Dulbecco's modified Eagle's medium (DMEM) (Sigma-Aldrich, Carlsbad, CA, USA) containing 10% heat-inactivated fetal bovine serum (FBS) and 1% penicillin-streptomycin (100 IU/ml). The

oligoribonucleotide sequence of the negative control, human NUCB2 siRNA 1, human NUCB2 siRNA 2, human KLF4A siRNA 1, and human KLF4A siRNA 2 were as follows: 5'-GAAUAUGGUUGUUUGAAGAUU-3', 5'-UAUCUUCGCACUUUCCACAGGGUGA-3', 5'-UUGAUUAGCAUAUCUAAAUCUGUGG-3', 5'-ACCUCGCCUUACACAUGAAUU-3', and 5'-GAGAGACCGAGGAGUUCAAUU-3'. These siRNAs were synthesized by Sangon Biotech Co., Ltd (Shanghai, China). For siRNA transfection, a mixture of siRNA and RNAiMax (Invitrogen, Shanghai, China) was added to the cells at a final siRNA concentration of 100 nmol/l following the manufacturer's protocol. The cells were used for experiments 48 h after siRNA transfection.

Quantitative real-time PCR (qPCR)

qPCR was used to validate the knockdown of genes and to measure the changes in EMT. Total RNA was extracted from the cells using TRIzol reagent (Invitrogen) according to the manufacturer's instructions. cDNA was synthesized from 1 μg of total RNA using a high-capacity cDNA reverse transcription kit (Applied Biosystems, CA, USA). qPCR was carried out using the SYBR Green PCR Mix Kit (TaKaRa, Dalian, China) according to the manufacturer's instructions. The ΔΔCT (cycle threshold) method was used to analyze the relative expression levels of target genes. The primers used in this study are summarized in Supplementary Table 1.

MTT assay

A Cytotoxicity Assay Kit was purchased from Beyotime Biotechnology (Shanghai, China). Briefly, cells (1×10^3 cells/well) were plated in 96-well plates 24 h after siRNA transfection and were cultured for 48 h. Then, 20 μl of MTT solution was added to each well for 4 h, and the absorbance was measured at a wavelength of 570 nm by a plate reader.

Side population analysis

Cells (1×10^6 cells/ml) were incubated with 5 μM Hoechst 33342 (Sigma Chemicals, St Louis, MO, USA) at 37°C for 1 h. Control cells were treated with both 5 μM Hoechst 33342 and 500 μM verapamil (an inhibitor of ABC transporter) (Sigma). An FACS Vantage SE cytometer (Becton Dickinson, San Jose, CA, USA) was used to analyze the side population (SP) cells at 402-446 nm for Hoechst blue staining or at 640 nm for Hoechst red staining. Dead cells were recognized by 1 μM DAPI staining.

Western blot analysis

Cells were lysed in RIPA lysis buffer (Beyotime Biotechnology) for 10 minutes, and the lysates were centrifuged at 12000 × g, at 4°C for 15 minutes. Protein lysate (30 μg) was subjected to 8% SDS-PAGE and transferred onto nitrocellulose (NC) membranes (KeyGen, Nanjing, China). The membranes were then incubated with the following primary antibodies overnight at 4°C: NUCB2 (sc-376947), KIF4A (sc-365144), AKT (sc-5298), P-AKT (sc-377556), mTOR (sc-517464), P-mTOR (sc-293133), p21 (sc-6246), ABCG2 (sc-58222), and GAPDH (sc-47724) (Santa Cruz, CA, USA). Each specific antibody binding was detected with horseradish peroxidase (HRP)-conjugated respective secondary antibodies (Beyotime Biotechnology). The signals were developed by ECL-Plus detection reagents (Beyotime Biotechnology) and visualized with X-ray film (Fujifilm, Japan).

Statistical analysis

R programming (version 3.6.2) was used for the statistical analyses in this study. The independent t-test was used for normal distribution variables, while the Mann-Whitney U test was used for the non-normally distributed continuous variables. *P* values less than 0.05 were considered statistically significant.

Results

Correlation of the mRNAsi and clinical characteristics in White patients and Asian patients with liver cancer

We first evaluated the correlation between mRNAsi values and the clinical features in White and Asian patients. As shown in **Figures 1**, **2**, no significant difference of the mRNAsi values was found with age (**Figures 1**, **2A**), sex (**Figures 1**, **2B**), G1 to G4 stages (**Figures 1**, **2C**), stage I to stage IV (**Figures 1**, **2D**), and T1 to T4 stages (**Figures 1**, **2E**). However, the mRNAsi scores were significantly different among Asian patients, White patients, and healthy controls (Supplementary Figure 1B). With respect to prognosis, the Asian patients

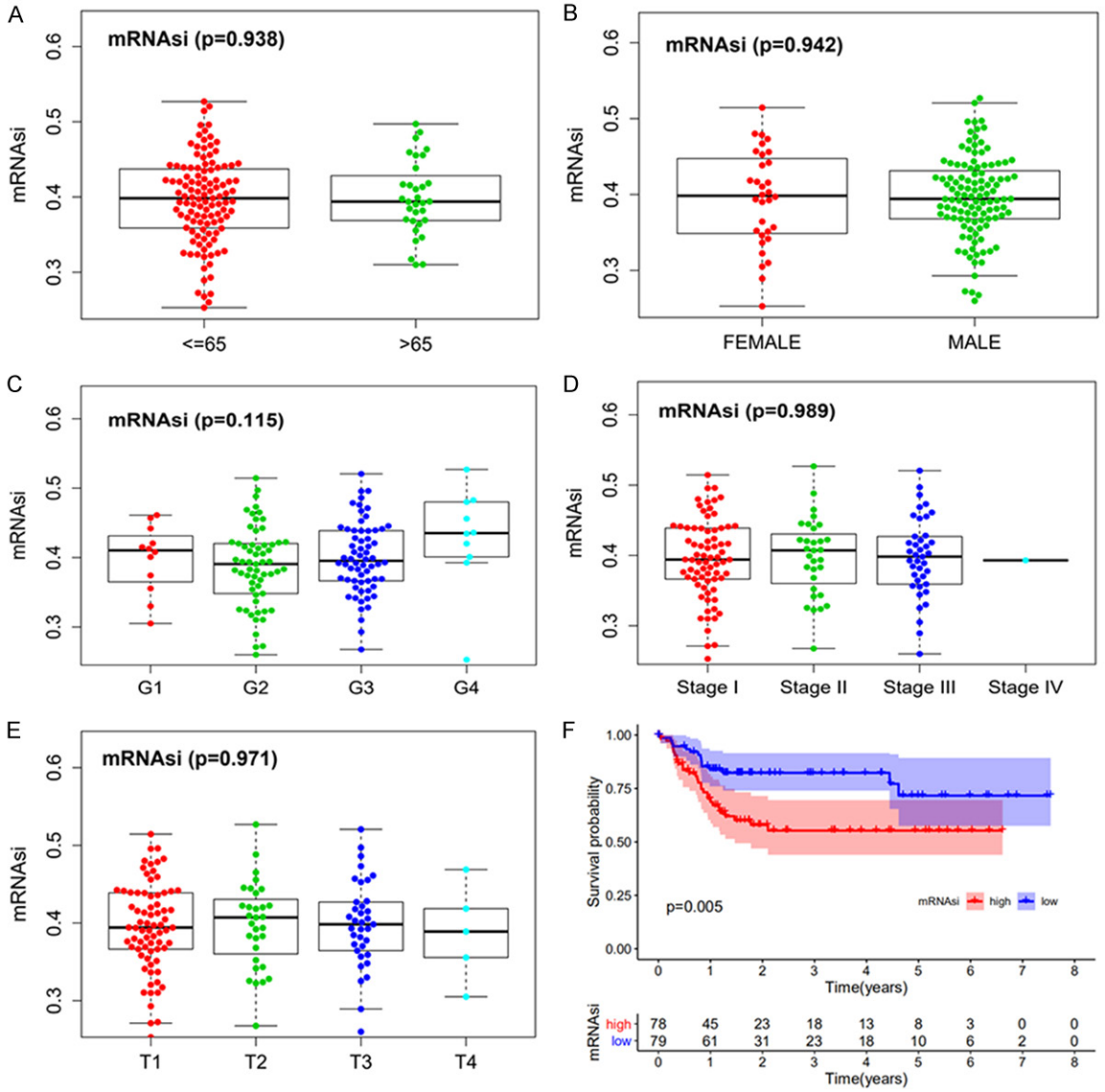


Figure 1. Correlation between the mRNAsi score and the clinical characteristics in Asian patients with liver cancer. Comparison between the mRNAsi score and age (A), sex (B), pathological grade (C), or TNM stage (D, E). (F) Kaplan-Meier survival analysis of the relationship between the mRNAsi score and the survival time in Asian patients.

with a high mRNAsi score showed significantly worse outcomes than those with a low mRNAsi score for overall survival (OS) (P = 0.005) (**Figure 1F**). However, no prognostic difference was observed between the White patients with high and low mRNAsi score (P = 0.112) (**Figure 2F**).

The most significant modules and genes in White patients and Asian patients with liver cancer

WGCNA co-expression network was constructed to classify genes with highly cooperative expression into a gene module in White patients and Asian patients (**Figures 3**, **4A**). Modules were selected for subsequent research due to the highest positive correlation with mRNAsi. Sixteen gene modules and twenty gene modules were obtained from the Asian patient dataset and the White patient dataset, respectively, using scale-free networks (**Figures 3**, **4B**). The grey (R2 = 0.68, P = 7.7e-08), blue (R2 = 0.26, P = 2.5e-12), and green modules (R2 = 0.57, P = 1.1e-25) were found to be associated with the mRNAsi of the Asian patients (**Figure 3C**), whereas the turquoise (R2

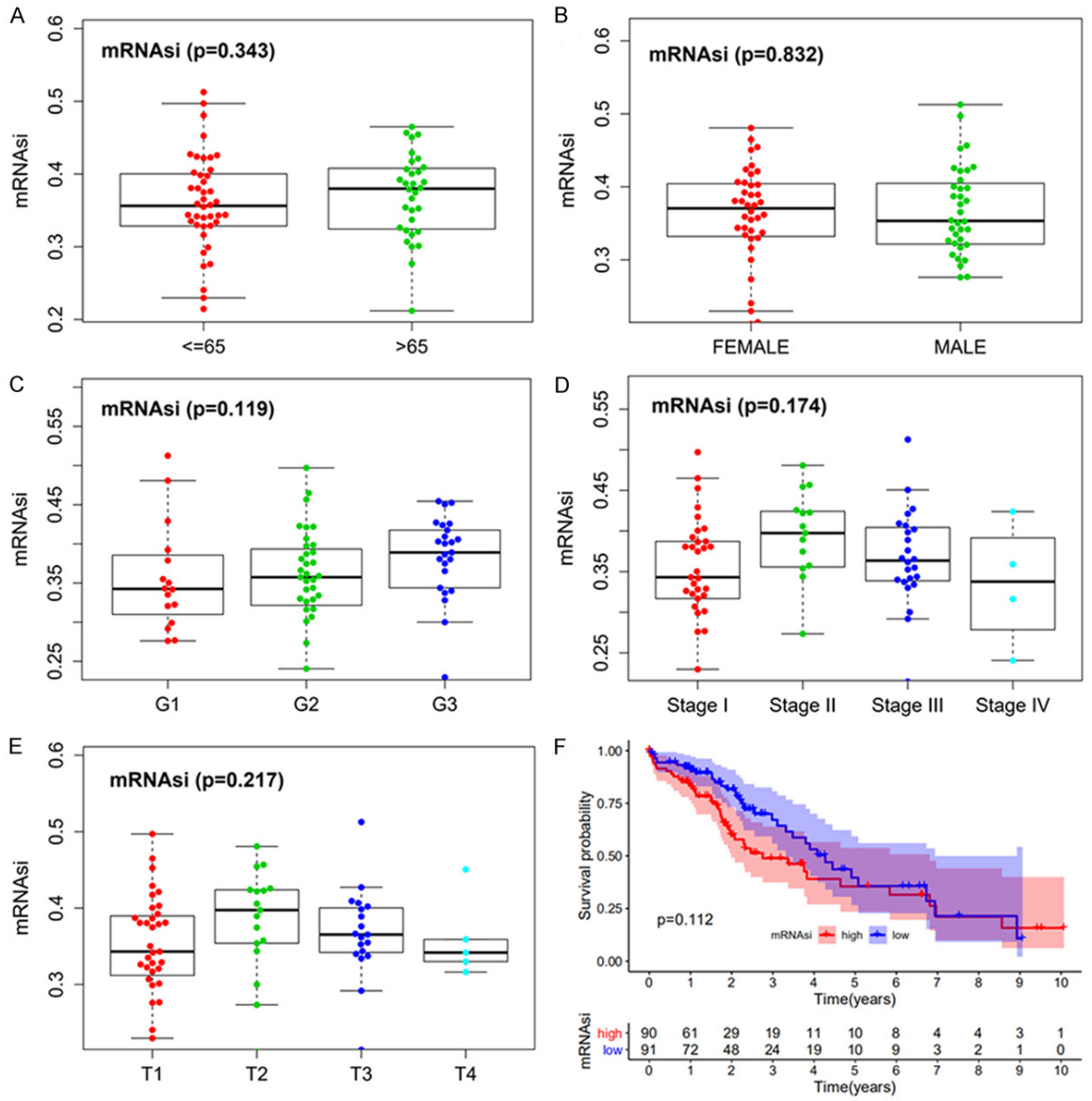


Figure 2. Correlation between the mRNAsi score and the clinical characteristics in White patients with liver cancer. Comparison between the mRNAsi score and age (A), sex (B), pathological grade (C), or TNM stage (D, E). (F) Kaplan-Meier survival analysis of the relationship between the mRNAsi score and the survival time in White patients.

= 0.46, P = 5.6e-47), purple (R2 = 0.86, P = 5e-48), and cyan modules (R2 = 0.84, P = 1e-34) were associated with the mRNAsi of the White patients (**Figure 4C**). Genes in the selected modules with MM > 0.8 and cor. gene GS > 0.6 were considered key genes for further analysis. As observed in the heatmap, 5126 DEGs were identified in Asian patients, of which 4769 were upregulated, while 436 were downregulated (**Figure 5A**). Likewise, 5155 DEGs were identified in White patients, of which 4878 were upregulated, while 277 were downregulated (**Figure 6A**). Based on the expression levels, 14 hub genes were identified in the Asian patients and the White patients, respectively (**Figures 5**, **6B**). The 14 hub genes were also significantly correlated with each other at the transcriptional level (**Figures 5**, **6C**). The functional correlation of the hub genes is shown in **Figures 5**, **6D**. To determine the effects of the key genes in different ethnic groups, we analyzed the prognostic role of the key genes in both White and Asian patients. Asian patients with a lower expression of NUCB2, CSPG5, ZDHHC12, or PHOSPHO2 had a favorable prognosis (P < 0.05) (**Figure 7A**); however, higher

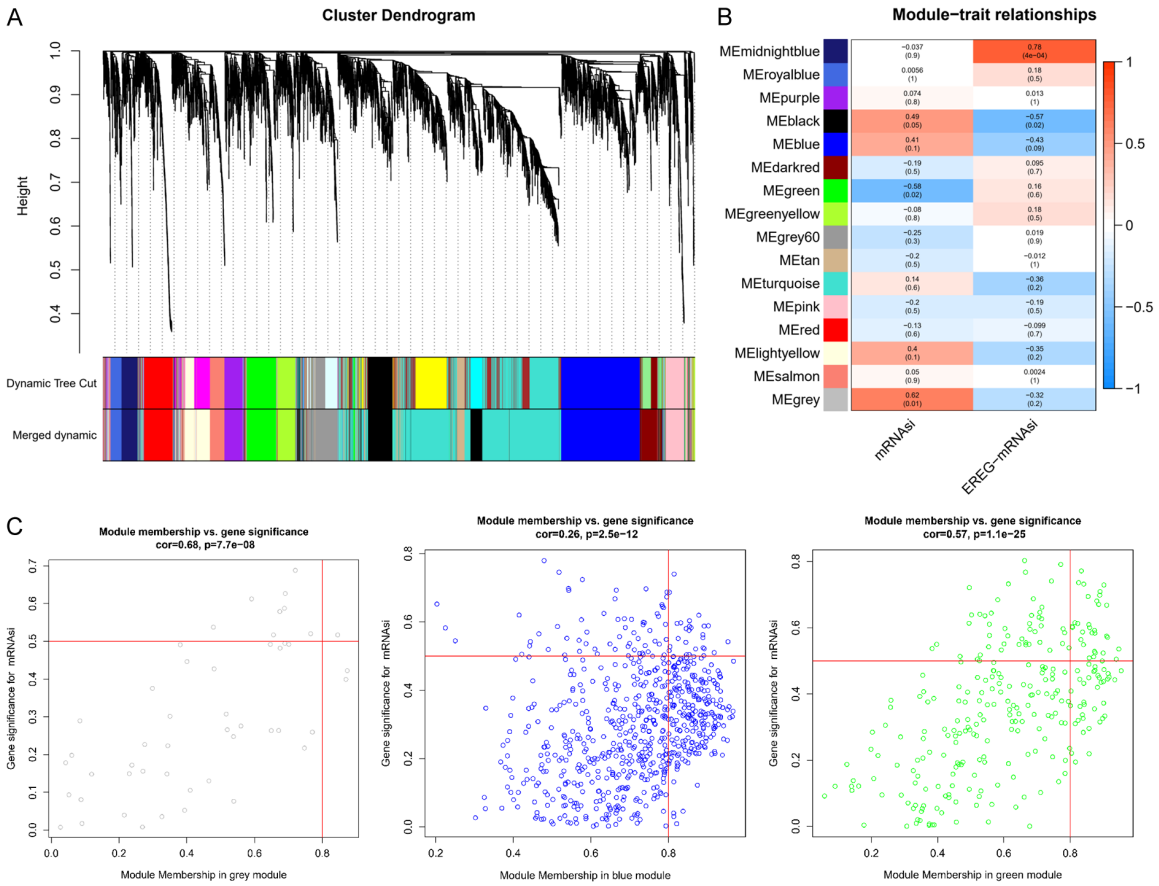


Figure 3. Screening of the key genes related to mRNAsi in Asian patients. A. WGCNA analysis of the DEGs. B. Correlation analysis of the modules and the clinical traits with mRNAsi. C. Scatter plot analysis of the modules in grey, blue, and green.

expression of SLC25A15 indicated a better prognosis for Asian patients (P < 0.05) (**Figure 7A**). Similarly, NEDD4L, RBM3, and PHOSPHO2, which were identified from the Asian patient dataset, were also prognostic markers for White patients (P < 0.05) (**Figure 7B**). Interestingly, the key genes from the White patient dataset, CDCA8, CENPH, PA2G4, and PRIM1, were unfavorable prognostic markers for both White patients (P < 0.05) (**Figure 8A**) and Asian patients (P < 0.05) (**Figure 8B**). Furthermore, univariate Cox regression analysis revealed that NUCB2 (HR = 1.6351; P = 0.0033), CSPG5 (HR = 1.6909; P = 8e-04), ZDHHC12 (HR = 1.9219; P = 0.0066), and PHOSPHO2 (HR = 2.3889; P = 0.0006) were hazard factors for the Asian patients; however, SLC25A15 (HR = 0.7423; P = 0.0073) was a protective factor for these patients (Supplementary Figure 2A). On the other hand, PRIM1 (HR = 1.2656; P = 0.0312), PA2G4 (HR = 1.5257; P = 0.0306), and CDCA8 (HR = 1.2559; P = 0.0128) were hazard factors for the White patients (Supplementary Figure 2D). Nevertheless, these genes didn't show significant influence on Asian patients and White patients when using Multivariate Cox regression analysis (P > 0.05) (Supplementary Figure 2B, 2E). Clinical prediction model showed that pTNM stage and Age were indicators for the outcome of Asian patients and White patients, respectively (Supplementary Figure 2C, 2F).

Based on the results above, we selected NUCB2 from the Asian patient dataset and KLF4A from the White patient dataset for the function analysis (Supplementary Figure 1C).

Validation of the function of NUCB2 and KLF4A in vitro

For the functional validation, we knocked down NUCB2 or KLF4A in HepG2 and SMMC-7721

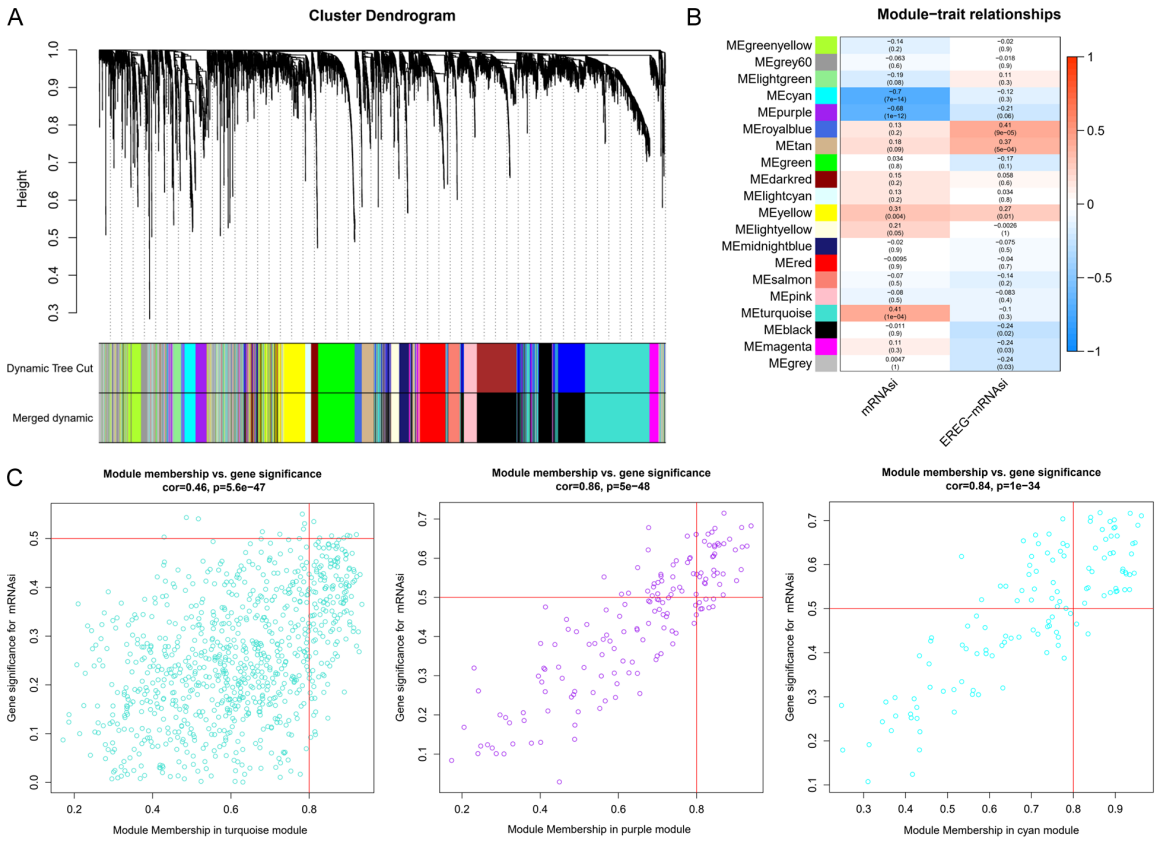


Figure 4. Screening of the key genes related to mRNAsi in White patients. A. WGCNA analysis of the DEGs. B. Correlation analysis of the modules and the clinical traits with mRNAsi. C. Scatter plot analysis of the modules in turquoise, purple, and cyan.

cells using siRNA. qPCR confirmed the successful knockdown of NUCB2 or KLF4A ($P < 0.05$) (**Figure 9A**). In the NUCB2 or KLF4A knockdown cells, E-cadherin was upregulated, while N-cadherin, Vimentin, and Snail were downregulated ($P < 0.05$) (**Figure 9B**). Furthermore, the knockdown of NUCB2 or KLF4A could inhibit the proliferation and induce the apoptosis in HepG2 and SMMC-7721 cells ($P < 0.05$) (**Figure 9C**). Notably, we found P-AKT and P-mTOR levels were downregulated in the NUCB2 knockdown cells, while no significant changes in ERK or MEK level were observed (**Figure 9D**). Moreover, the SP cell fraction comprised 0.89% of the total HepG2 cells and 0.63% of the total SMMC-7721 cells labeled with Hoechst 33342 (**Figure 10A**). Importantly, this population disappeared following the treatment of verapamil (an inhibitor of ABC transporter) or KLF4A knockdown (**Figure 10A**). In the KLF4A knockdown cells, P-AKT and ABCG2 levels were also downregulated (**Figure 10B**).

Discussion

mRNAsi is a cancer stemness score to describe the similarity between tumor cells and stem cells, which can be used to quantify the CSCs in tumors [13]. In recent years, WGCNA has become a widely used and effective method for cancer-related gene screening [14]. By analyzing the data in the liver cancer dataset of the TCGA database, previous studies have found several liver cancer-related genes that could serve as potential therapeutic targets [9, 10, 15-17]. However, there is no comparative study on the key genes involved in tumorigenesis between different races using this screening method. Hence, in this study, we used the mRNAsi index to identify the key genes during the tumorigenesis of liver cancer in White patients and Asian patients, as well as validated the functions of two key genes *in vitro*.

Based on the results of the WGCNA analysis, 14 key genes were identified in Asian patients

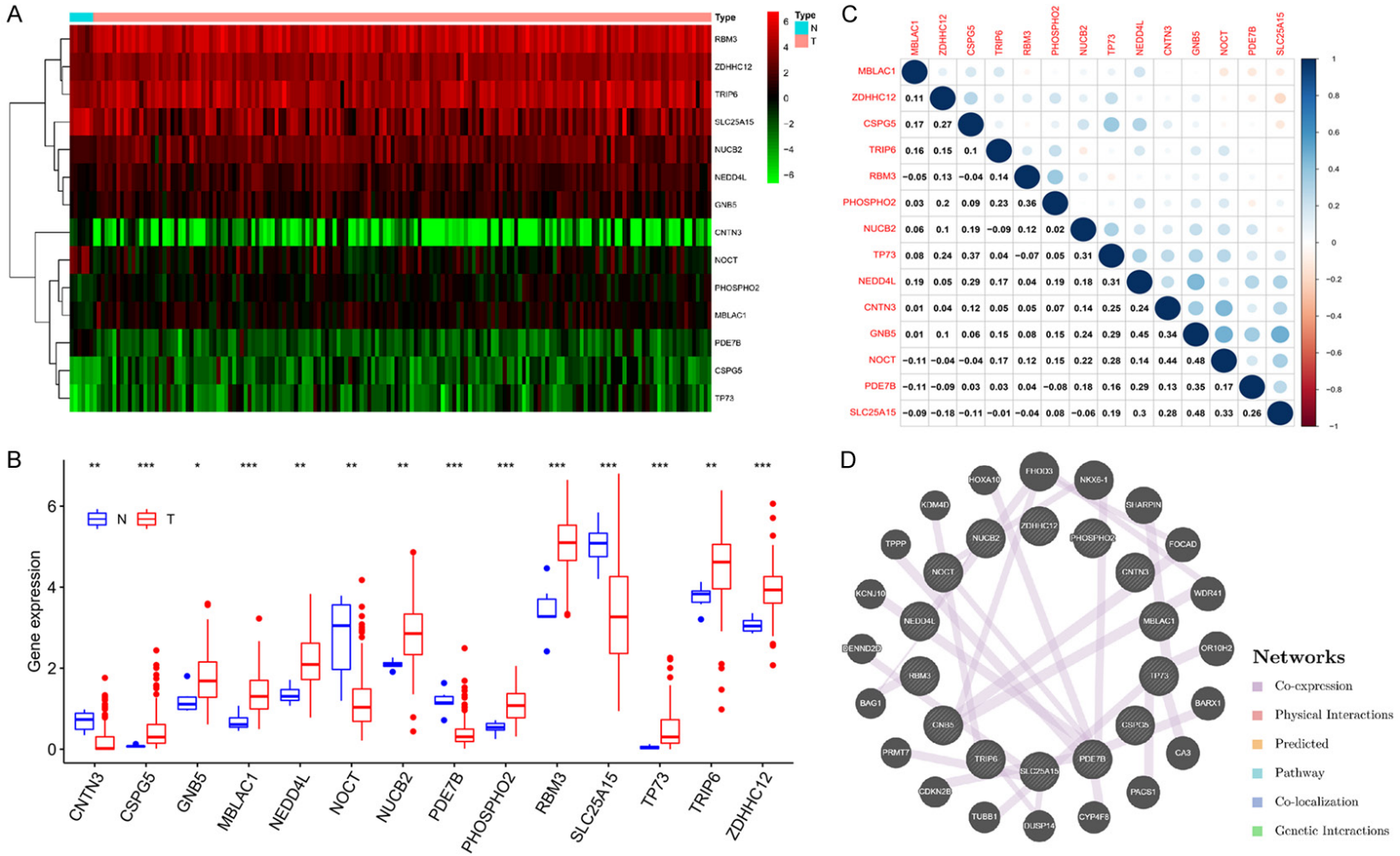


Figure 5. Co-expression network of the key genes in Asian patients. A. Heatmap of the key genes. B. The expression levels of the key genes in normal and cancer samples. C. Correlation analysis of the key genes at transcription-level among the three modules. D. Functional correlation analysis of the key genes among the three modules.

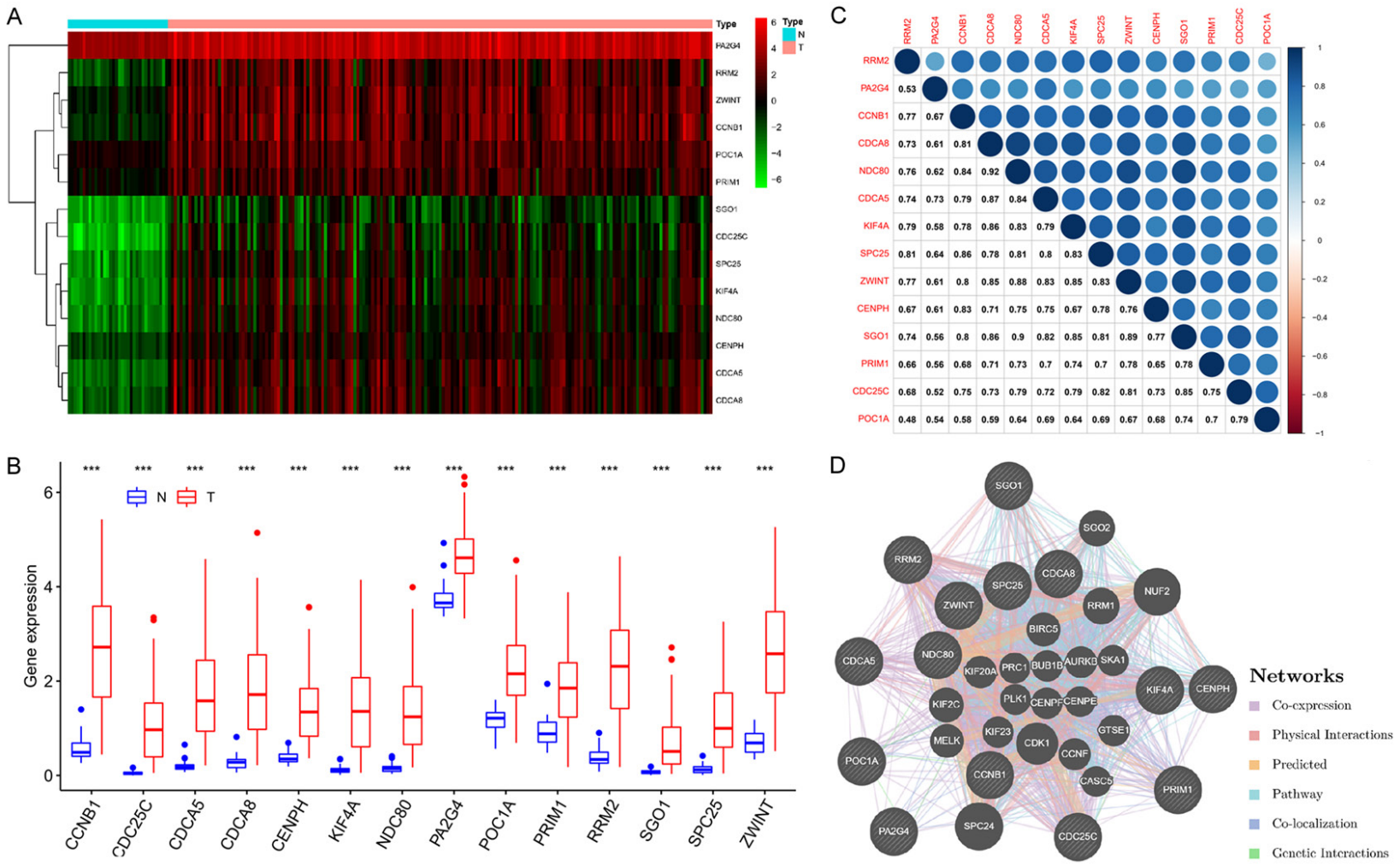


Figure 6. Co-expression network of the key genes in White patients. A. Heatmap of the key genes. B. The expression levels of the key genes in normal and cancer samples. C. Correlation analysis of the key genes at transcription-level among the three modules. D. Functional correlation analysis of the key genes among the three modules.

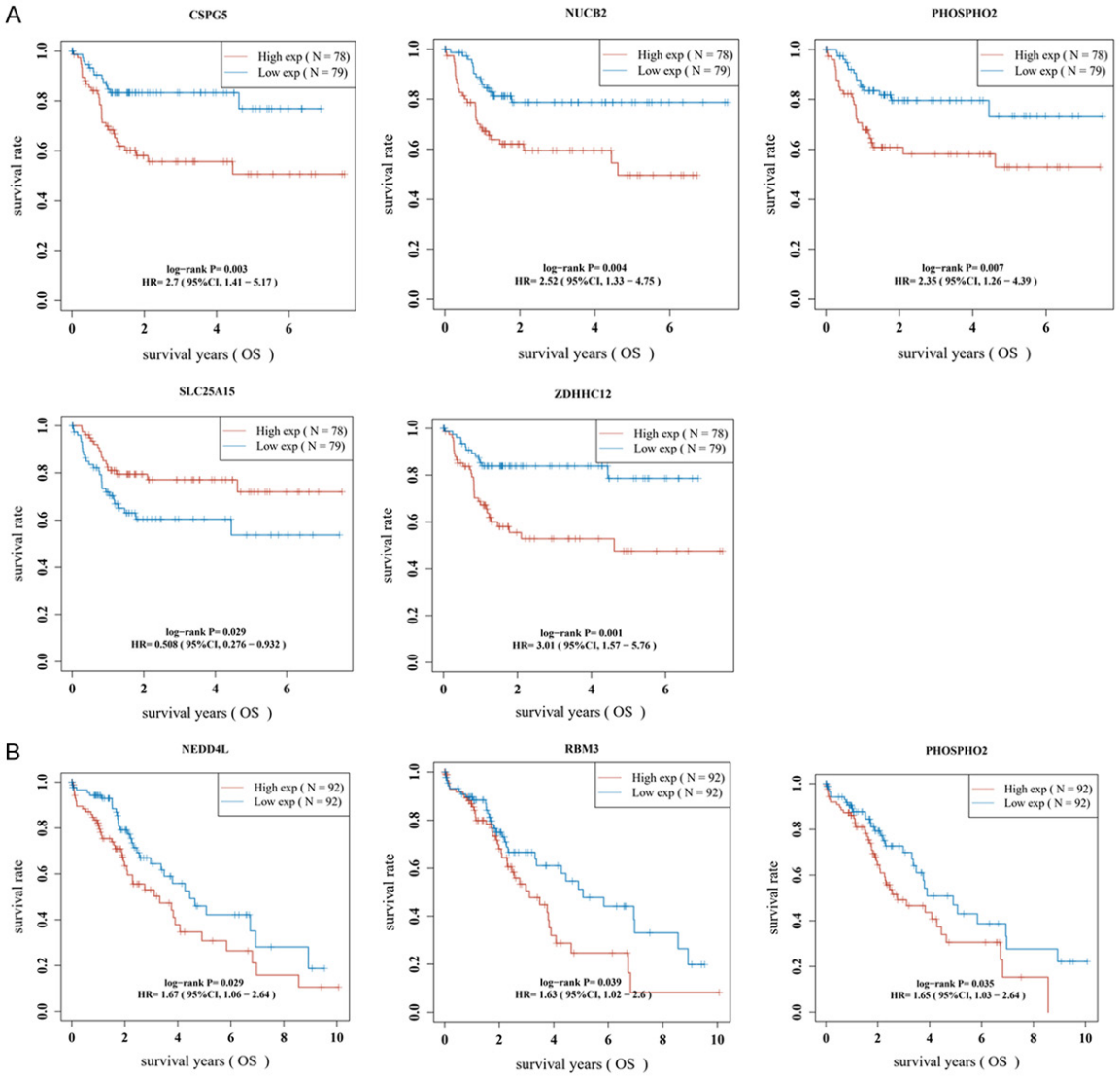


Figure 7. The overall survival of patients according to the key genes identified from the Asian patient cohort. A. The key genes indicating the prognosis of Asian patients. B. The key genes indicating the prognosis of White patients.

and White patients, respectively. Among them, several key genes, such as CENPH, could serve as prognostic markers for both White and Asian patients. CENPH has been reported to play an important role in the connection between the kinetochore and microtubule [18], and dysregulation of CENPH will cause abnormal of chromosome separation [18]. In addition, CENPH is involved in the proliferation and apoptosis of hepatocellular carcinoma (HCC) cells through the mitochondrial apoptotic pathway [19]. In line with these finding, CENPH expression was higher in HCC tumor tissues than in adjacent noncancerous tissues [20], and HCC patients with low CENPH expression had a longer survival time [20]. CENPH was identified as key gene in both White and Asian patients, suggesting that tumorigenesis is similar in different races.

To experimentally validate the function of key genes we identified, we selected two key genes, NUCB2 and KLF4A, from White and Asian patients, respectively, for further investigation. The NUCB2 protein is composed of 420 amino acids [21] and can be split into three fragments with glycogen invertase: nesfatin-1, nesfatin-2, and nesfatin-3 [22]. NUCB2 is closely related to the occurrence, progression, and prognosis of various malignant tumors, such as gastric can-

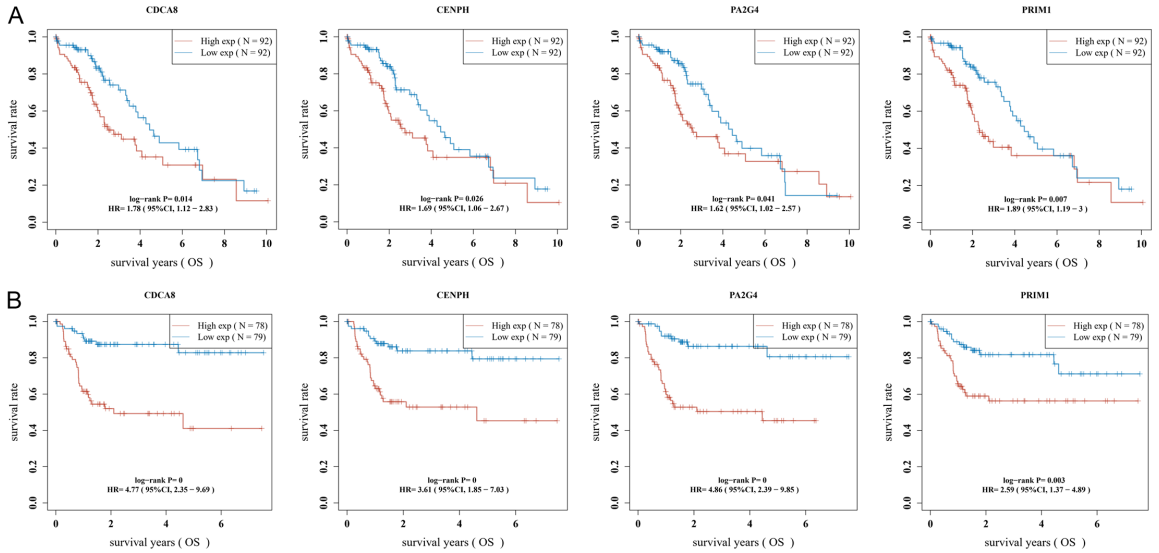


Figure 8. Overall survival of patients according to the key genes identified from the White patient cohort. A. The key genes indicating the prognosis of White patients. B. The key genes indicating the prognosis of Asian patients.

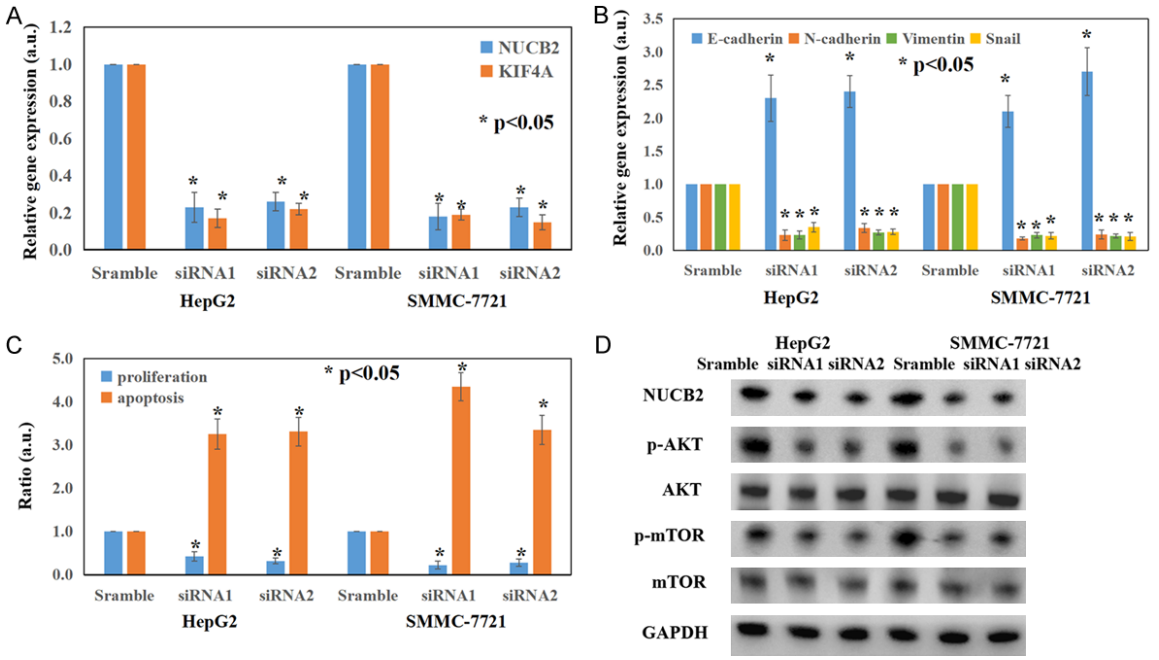


Figure 9. The effects of NUCB2 on liver cancer cells. A. The NUCB2 and KIF4A mRNA levels in HepG2 and SMMC-7721 after NUCB2 knockdown. B. The mRNA levels of EMT markers in HepG2 and SMMC-7721 cells after NUCB2 knockdown. C. The proliferation and apoptosis of HepG2 and SMMC-7721 cells with NUCB2 knockdown. D. p-AKT and p-mTOR expression levels in HepG2 and SMMC-7721 cells after NUCB2 knockdown.

cer [23]. It also promotes the proliferation, migration, and invasion of breast cancer cells, bladder cancer cells, and endometrial cancer cells [24-26]. In this study, we found knockdown of NUCB2 could inhibit the proliferation and EMT as well as induce the apoptosis of liver cancer cells. Previous studies also showed that NUCB2 enhanced EMT in gastric carcinoma, renal cell carcinoma, and colorectal cancer [27-29]. In addition, the AKT/mTOR signaling pathway was inhibited in NUCB2 knockdown liver cancer cells. Although no previous studies have

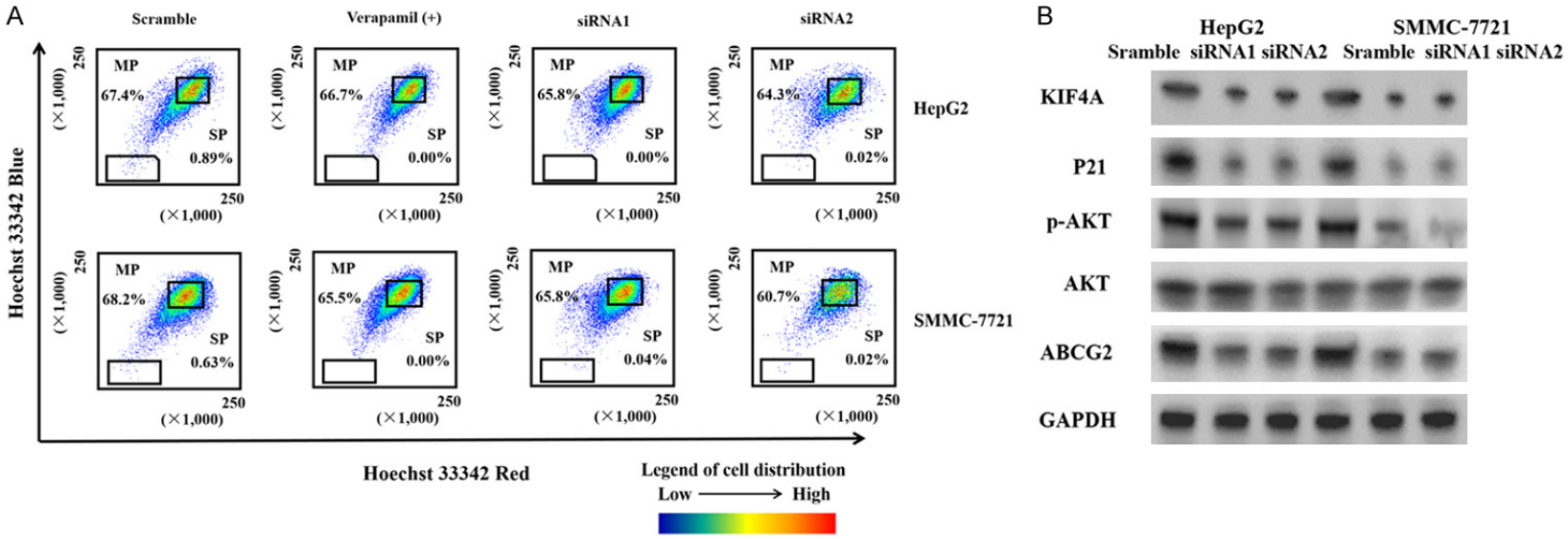


Figure 10. The effect of KIF4A on liver cancer cells. A. The percentage of the side population in HepG2 and SMMC-7721 cells after KIF4A knockdown. B. P21, p-AKT, and ABCG2 expression levels in HepG2 and SMMC-7721 cells after KIF4A knockdown.

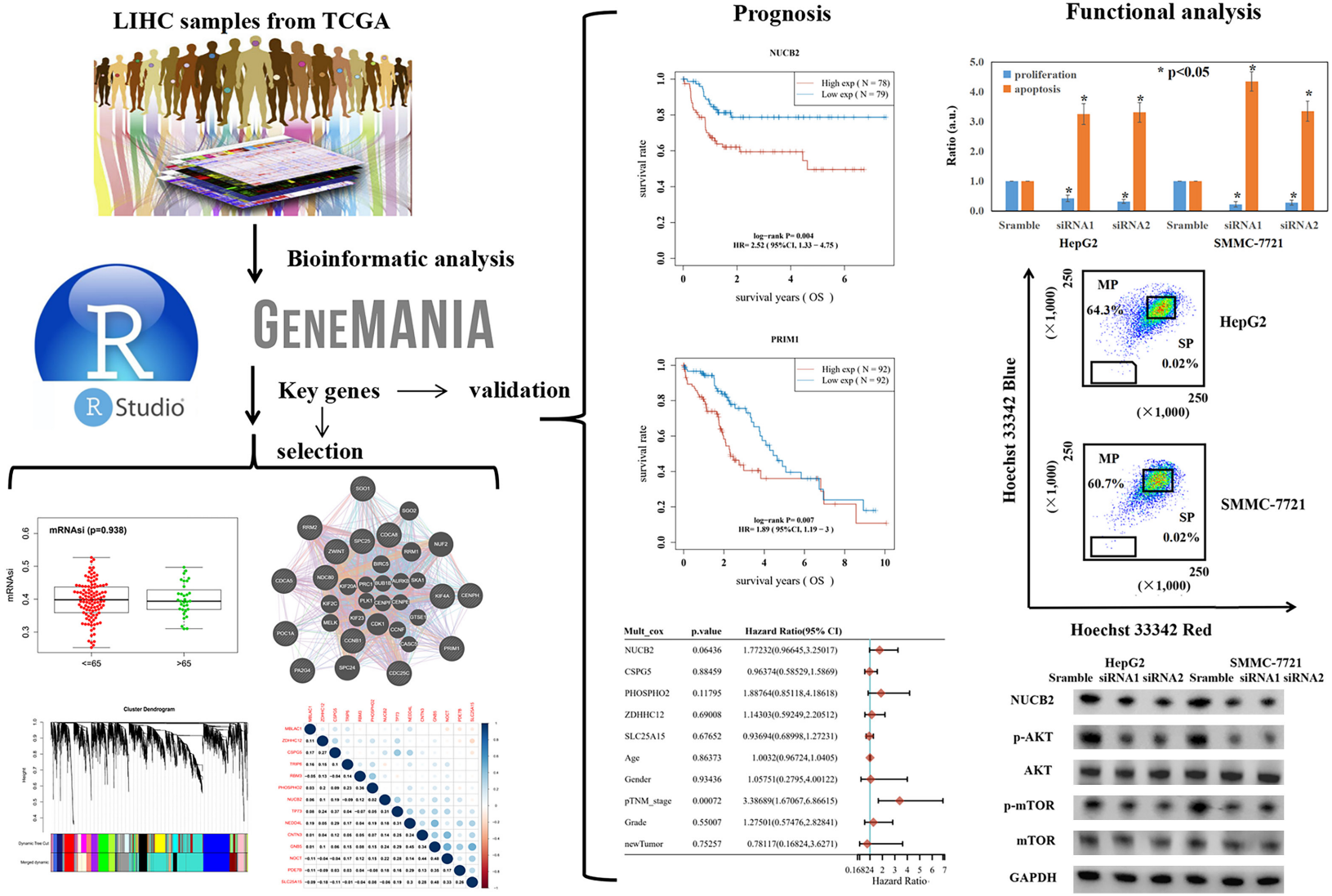


Figure 11. Flow chart of the study design and the functional validation of DEGs selected using "stemness index" from the Asian and White patient cohorts in TCGA database.

shown the interconnection between NUCB2 and the AKT/mTOR signaling pathway in cancer cells, nesfatin-1-activated PI3K/Akt/mTOR signaling was confirmed in trophoblast cells, vascular smooth muscle cells, and ovarian epithelial carcinoma cells [30-32]. KLF4, also known as epidermal zinc finger factor, is an important transcription factor of the Sp1/Kruppel like factor family [33]. KLF4 has three major protein domains: the nuclear localization domain, DNA binding domain, and transcriptional regulation domain [33]. KLF4 can form transcription factor complexes with Sox2, Oct4, and Nanog to regulate downstream genes, thereby maintaining the stemness of cells [34]. In addition, KLF4 can induce non-liver CSCs (EpCAM-/CD133-) in the CSC-like phenotype (EpCAM+/CD133+) through directly binding to the promoter of EpCAM and therefore upregulating its expression [35]. In this study, we found knockdown of KLF4A downregulated ABCG2 expression in liver cancer cells, which directly led to the decrease in the proportion of SP in liver cancer cells. SP cells refer to the subpopulations of the cells screened by the Hoechst 33342 exclusion experiment and have the characteristics of cancer stem cells [36]. The ABCG2 transporter is related to the drug resistance in cancer, as it extrudes chemotherapy drugs out of tumor cells, leading to the failure of cancer therapy [37]. Previous studies have shown both KLF4A and ABCG2 regulate the stemness of cancer cells and normal cells [38-40]. Our study confirmed KLF4A was a regulator of ABCG2 in liver cancer cells for the first time; however, the mechanism of it function is still unknown. We hypothesize that KLF4A, as a transcription factor, can bind the promoter of ABCG2 to upregulate ABCG2 expression level. We will determine the mechanisms in a future study.

Conclusions

In this study, we confirmed that mRNAsi is a reliable index to analyze the stemness genes related to the prognosis of liver cancer, regardless of ethnic differences (**Figure 11**). The key stemness genes identified from the White or Asian patient datasets can affect the occurrence, development, and prognosis of liver cancer. In addition, the oncogenic functions of two key genes, NUCB2 and KLF4A, were validated experimentally. Nonetheless, there are several limitations in this study. First, the functions and mechanisms of the key genes identified in this study should be further validated *in vitro* and *in vivo*. Second, liver cancer samples should be collected from both White and Asian patients to cross-examine the functions of the key genes during hepatocarcinogenesis. Despite these limitations, the stemness index is still a useful method to analyze the subtypes of liver cancer and to identify more key genes for our future study.

Acknowledgements

This study was supported by the National Natural Scientific Foundation of China (No. 81972784), the Natural Scientific Foundation of Liaoning (No. 2019-MS-146), the Scientific Research Fund of the Liaoning Provincial Department of Education (No. JYTQN2020039), and the “Double First-Class” Disciplinary Construction Project of Jinzhou Medical University.

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Pu Xia, Biological Anthropology Institute, College of Basic Medical Science, Jinzhou Medical University, Jinzhou, Liaoning, P. R. China. E-mail: xiapu@jzmu.edu.cn

References

[1] Llovet JM, Kelley RK, Villanueva A, Singal AG, Pikarsky E, Roayaie S, Lencioni R, Koike K, Zucman-Rossi J and Finn RS. Hepatocellular carcinoma. Nat Rev Dis Primers 2021; 7: 6.

[2] Forner A, Reig M and Bruix J. Hepatocellular carcinoma. Lancet 2018; 391: 1301-1314.

[3] Yang JD, Hainaut P, Gores GJ, Amadou A, Plymoth A and Roberts LR. A global view of hepatocellular carcinoma: trends, risk, prevention and management. Nat Rev Gastroenterol Hepatol 2019; 16: 589-604.

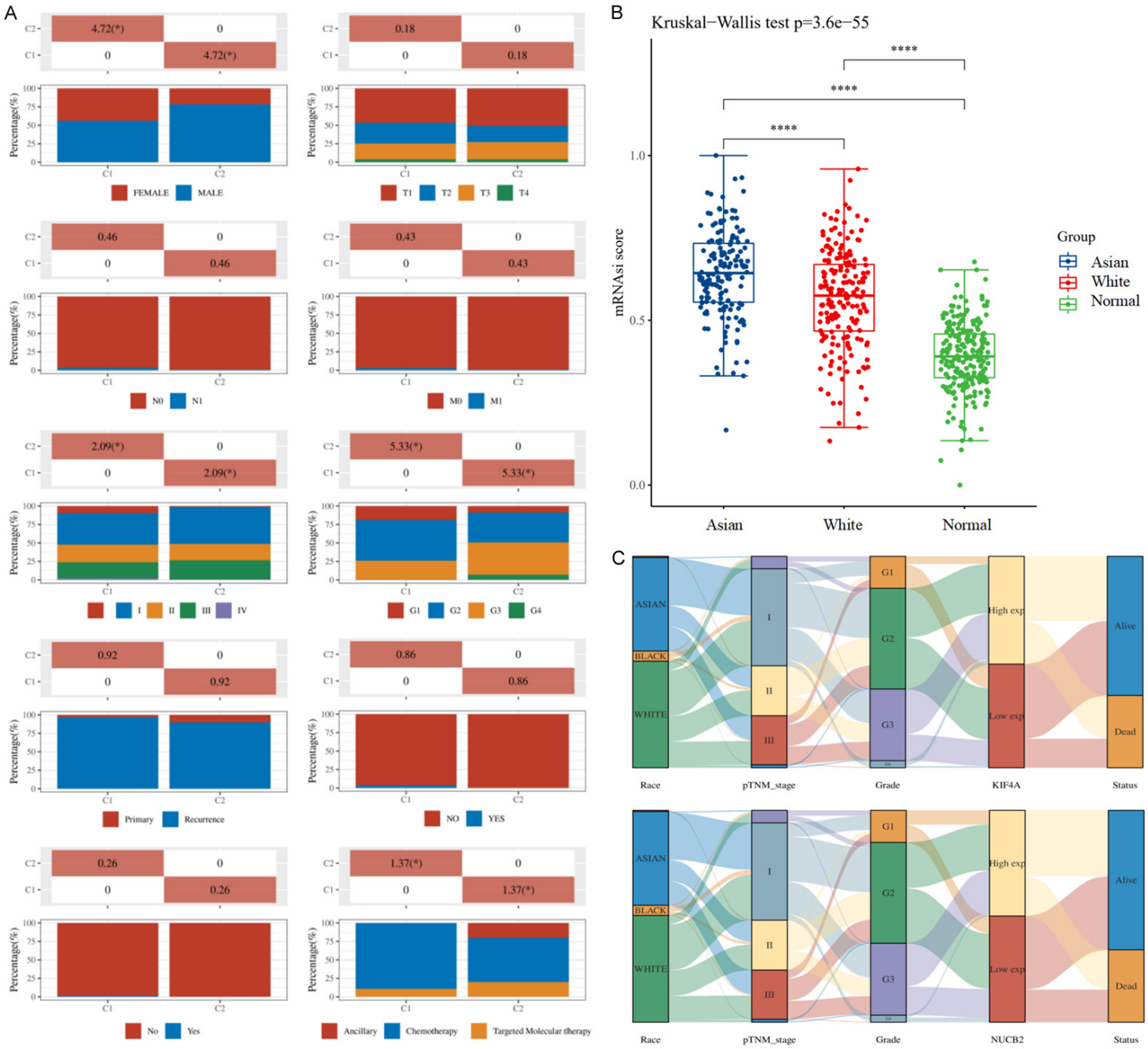
[4] Clarke MF, Dick JE, Dirks PB, Eaves CJ, Jamieson CH, Jones DL, Visvader J, Weissman IL and Wahl GM. Cancer stem cells--perspectives on current status and future directions: AACR workshop on cancer stem cells. Cancer Res 2006; 66: 9339-9344.

[5] Clevers H. The cancer stem cell: premises, promises, and challenges. Nat Med 2011; 17: 313-319.

[6] Fang X, Yan Q, Liu S and Guan XY. Cancer stem cells in hepatocellular carcinoma: intrinsic and

extrinsic molecular mechanisms in stemness regulation. Int J Mol Sci 2022; 23: 12327.
[7] Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, Kamińska B, Huelsken J, Omberg L, Gevaert O, Colaprico A, Czerwińska P, Mazurek S, Mishra L, Heyn H, Krasnitz A, Godwin AK, Lazar AJ, Stuart JM, Hoadley KA, Laird PW, Noushmehr H and Wiznerowicz M; Cancer Genome Atlas Research Network. Machine learning identifies stemness features associated with oncogenic dedifferentiation. Cell 2018; 173: 338-354, e15.
[8] Lian H, Han YP, Zhang YC, Zhao Y, Yan S, Li QF, Wang BC, Wang JJ, Meng W, Yang J, Wang QH, Mao WW and Ma J. Integrative analysis of gene expression and DNA methylation through one-class logistic regression machine learning identifies stemness features in medulloblastoma. Mol Oncol 2019; 13: 2227-2245.
[9] Zhang G, Zhang K, Zhao Y, Yang Q and Lv X. A novel stemness-hypoxia-related signature for prognostic stratification and immunotherapy response in hepatocellular carcinoma. BMC Cancer 2022; 22: 1103.
[10] Mai H, Xie H, Luo M, Hou J, Chen J, Hou J and Jiang DK. Implications of stemness features in 1059 hepatocellular carcinoma patients from five cohorts: prognosis, treatment response, and identification of potential compounds. Cancers (Basel) 2022; 14: 563.
[11] Love MI, Huber W and Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014; 15: 550.
[12] Franz M, Rodriguez H, Lopes C, Zuberi K, Montojo J, Bader GD and Morris Q. GeneMANIA update 2018. Nucleic Acids Res 2018; 46: W60-W64.
[13] Sokolov A, Paull EO and Stuart JM. One-class detection of cell states in tumor subtypes. Pac Symp Biocomput 2016; 21: 405-16.
[14] Shen S, Xing J, Liu X and Zhang Y. A hypoxia-related prognostic model predicts overall survival and treatment response in hepatocellular carcinoma. Biosci Rep 2022; 42: BSR20221089.
[15] Yuan J, Liu Z, Wu Z, Yan L, Yang J and Shi Y. A novel medication decision gene signature predicts response to individualized therapy and prognosis outcomes in hepatocellular carcinoma patients. Front Immunol 2022; 13: 990571.
[16] 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.
[17] Han T, Liu Y, Chen Y, Chen T, Li Y, Li Q and Zhao M. Identification of the mechanism of matrine combined with glycyrrhizin for hepatocellular carcinoma treatment through network pharmacology and bioinformatics analysis. Oxid Med Cell Longev 2022; 2022: 2663758.
[18] Hoischen C, Yavas S, Wohland T and Diekmann S. CENP-C/H/I/K/M/T/W/N/L and hMis12 but not CENP-S/X participate in complex formation in the nucleoplasm of living human interphase cells outside centromeres. PLoS One 2018; 13: e0192572.
[19] Lu G, Hou H, Lu X, Ke X, Wang X, Zhang D, Zhao Y, Zhang J, Ren M and He S. CENP-H regulates the cell growth of human hepatocellular carcinoma cells through the mitochondrial apoptotic pathway. Oncol Rep 2017; 37: 3484-3492.
[20] Lu G, Shan T, He S, Ren M, Zhu M, Hu Y, Lu X and Zhang D. Overexpression of CENP-H as a novel prognostic biomarker for human hepatocellular carcinoma progression and patient survival. Oncol Rep 2013; 30: 2238-2244.
[21] Zhou Q, Liu Y, Feng R and Zhang W. NUCB2: roles in physiology and pathology. J Physiol Biochem 2022; 78: 603-617.
[22] Lenda R, Padjasek M, Krężel A, Ożyhar A and Bystranowska D. Does one plus one always equal two? Structural differences between nesfatin-1, -2, and nesfatin-1/2. Cell Commun Signal 2022; 20: 163.
[23] Altan B, Kaira K, Okada S, Saito T, Yamada E, Bao H, Bao P, Takahashi K, Yokobori T, Tetsunari O, Nishiyama M and Yamada M. High expression of nucleobindin 2 is associated with poor prognosis in gastric cancer. Tumour Biol 2017; 39: 1010428317703817.
[24] Kmiecik A, Ratajczak-Wielgomas K, Grzegrzółka J, Romanowicz H, Smolarz B and Dziegiel P. Expression of NUCB2/NESF-1 in breast cancer cells. Int J Mol Sci 2022; 23: 9177.
[25] Liu GM, Xu ZQ and Ma HS. Nesfatin-1/nucleobindin-2 is a potent prognostic marker and enhances cell proliferation, migration, and invasion in bladder cancer. Dis Markers 2018; 2018: 4272064.
[26] Takagi K, Miki Y, Tanaka S, Hashimoto C, Watanabe M, Sasano H, Ito K and Suzuki T. Nucleobindin 2 (NUCB2) in human endometrial carcinoma: a potent prognostic factor associated with cell proliferation and migration. Endocr J 2016; 63: 287-299.
[27] Ren L, Bao D, Wang L, Xu Q, Xu Y and Shi Z. Nucleobindin-2/nesfatin-1 enhances the cell proliferation, migration, invasion and epithelial-mesenchymal transition in gastric carcinoma. J Cell Mol Med 2022; 26: 4986-4994.
[28] Tao R, Niu WB, Dou PH, Ni SB, Yu YP, Cai LC, Wang XY, Li SY, Zhang C and Luo ZG. Nucleobindin-2 enhances the epithelial-mesenchy-

mal transition in renal cell carcinoma. Oncol Lett 2020; 19: 3653-3664.
[29] Kan JY, Yen MC, Wang JY, Wu DC, Chiu YJ, Ho YW and Kuo PL. Nesfatin-1/nucleobindin-2 enhances cell migration, invasion, and epithelial-mesenchymal transition via LKB1/AMPK/TORC1/ZEB1 pathways in colon cancer. Oncotarget 2016; 7: 31336-31349.
[30] Li T, Wei S, Fan C, Tang D and Luo D. Nesfatin-1 promotes proliferation, migration and invasion of HTR-8/SVneo trophoblast cells and inhibits oxidative stress via activation of PI3K/AKT/mTOR and AKT/GSK3β pathway. Reprod Sci 2021; 28: 550-561.
[31] Lu QB, Wang HP, Tang ZH, Cheng H, Du Q, Wang YB, Feng WB, Li KX, Cai WW, Qiu LY and Sun HJ. Nesfatin-1 functions as a switch for phenotype transformation and proliferation of VSMCs in hypertensive vascular remodeling. Biochim Biophys Acta Mol Basis Dis 2018; 1864: 2154-2168.
[32] Xu Y, Pang X, Dong M, Wen F and Zhang Y. Nesfatin-1 inhibits ovarian epithelial carcinoma cell proliferation *in vitro*. Biochem Biophys Res Commun 2013; 440: 467-472.
[33] Wang H, Yang L, Jamaluddin MS and Boyd DD. The Kruppel-like KLF4 transcription factor, a novel regulator of urokinase receptor expression, drives synthesis of this binding site in colonic crypt luminal surface epithelial cells. J Biol Chem 2004; 279: 22674-22683.
[34] Lee S, Wottrich S and Bonavida B. Crosstalks between Raf-kinase inhibitor protein and cancer stem cell transcription factors (Oct4, KLF4, Sox2, Nanog). Tumour Biol 2017; 39: 1010428317692253.
[35] Karagonlar ZF, Akbari S, Karabicici M, Sahin E, Avci ST, Ersoy N, Ates KE, Balli T, Karacicek B, Kaplan KN, Celiker C, Atabey N and Erdal E. A novel function for KLF4 in modulating the dedifferentiation of EpCAM-/CD133-nonstem cells into EpCAM+/CD133+ liver cancer stem cells in HCC cell line HuH7. Cells 2020; 9: 1198.
[36] Wu C, Wei Q, Utomo V, Nadesan P, Whetstone H, Kandel R, Wunder JS and Alman BA. Side population cells isolated from mesenchymal neoplasms have tumor initiating potential. Cancer Res 2007; 67: 8216-8222.
[37] Kukal S, Guin D, Rawat C, Bora S, Mishra MK, Sharma P, Paul PR, Kanojia N, Grewal GK, Kukreti S, Saso L and Kukreti R. Multidrug efflux transporter ABCG2: expression and regulation. Cell Mol Life Sci 2021; 78: 6887-6939.
[38] Kumar A, Xu Y, Yang E and Du Y. Stemness and regenerative potential of corneal stromal stem cells and their secretome after long-term storage: implications for ocular regeneration. Invest Ophthalmol Vis Sci 2018; 59: 3728-3738.
[39] Duz MB and Karatas OF. Differential expression of ABCB1, ABCG2, and KLF4 as putative indicators for paclitaxel resistance in human epithelial type 2 cells. Mol Biol Rep 2021; 48: 1393-1400.
[40] Park J, Kim SK, Hallis SP, Choi BH and Kwak MK. Role of CD133/NRF2 axis in the development of colon cancer stem cell-like properties. Front Oncol 2022; 11: 808300.

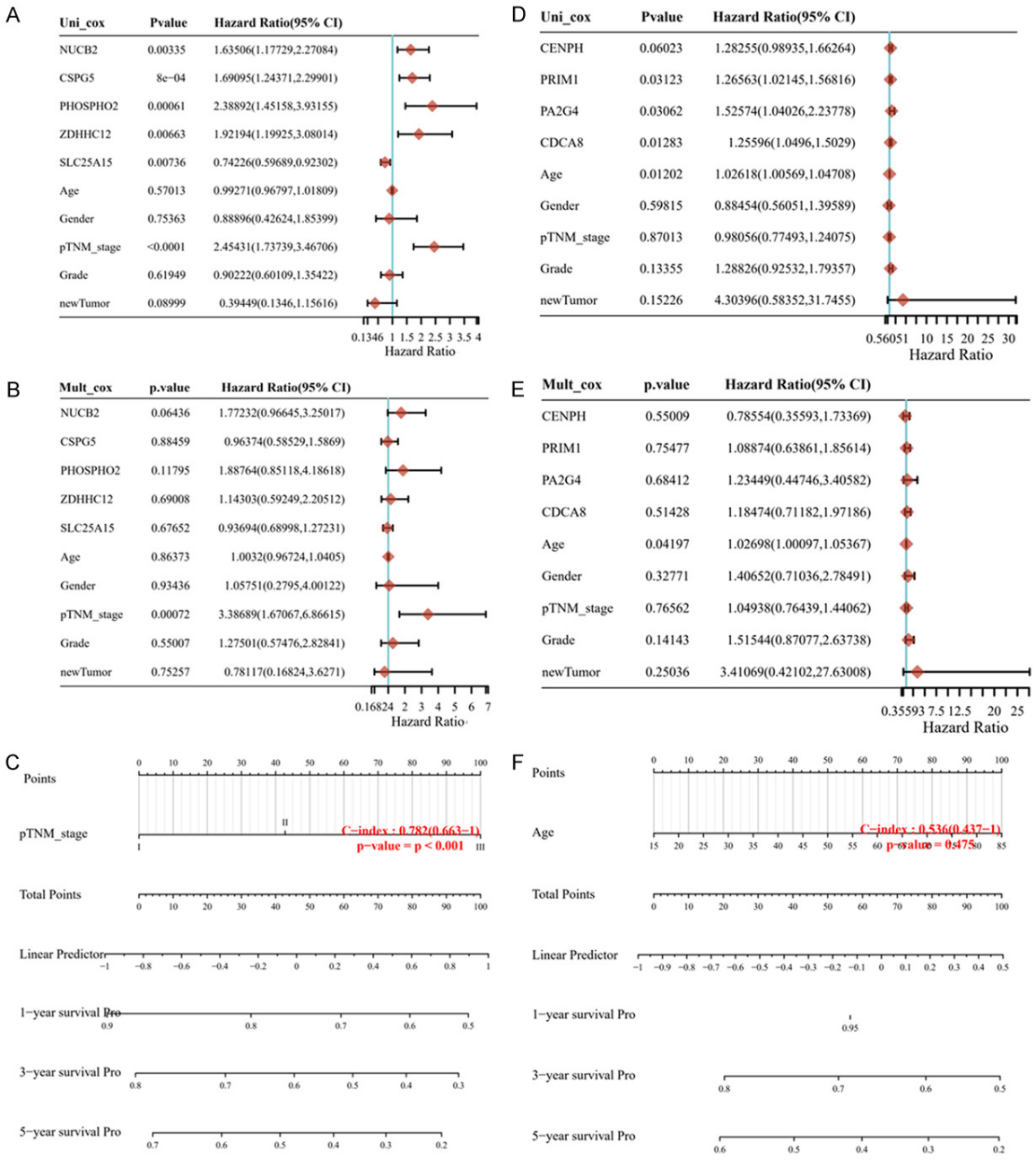

A
4.72(*)
0.18
0.46
0.43
2.09(*)
5.33(*)
0.92
0.86
0.26
1.37(*)
Percentage(%)
FEMALE
MALE
T1
T2
T3
T4
N0
N1
M0
M1
I
II
III
IV
G1
G2
G3
G4
Primary
Recurrence
NO
YES
No
Yes
Ancillary
Chemotherapy
Targeted Molecular therapy
B
Kruskal−Wallis test p=3.6e−55
mRNAsi score
Asian
White
Normal
Group
C
ASIAN
BLACK
WHITE
High exp
Low exp
Alive
Dead
Race
pTNM_stage
Grade
KIF4A
NUCB2
Status

Supplementary Figure 1. A. Clinical data of patients with liver cancer in TCGA database. B. mRNAsi score in Asian patients, White patients, and healthy controls. C. Distribution of KLF4A and NUCB2 in race, pTNM stage, grade, and the survival status of liver cancer patients.

Supplementary Table 1. Primers used in qPCR

Accession Number	Gene Symbol	Primers (5'-3')
NM_005013.4	NUCB2	F: TCTTGGAGCCAGATAGCTGG
		R: AGCTTCTGAGCCTCCAGTTG
NM_004235.6	KLF4A	F: TCGGACCACCTCGCCTTACA
		R: CTGGGCTCCTTCCCTCATCG
NM_131820.1	E-cadherin	F: AGCCATGTACGTTGCTATCC
		R: CGTAGCACAGCTTCTCCTTAAT
NM_031333.1	N-cadherin	F: GAGATCCTACTGGACGGTTCG
		R: TCTTGGCGAATGATCTTAGGA
NM_003380.5	Vimentin	F: GGCTCGTCACCTTCGTGAAT
		R: GAGAAATCCTGCTCTCCTCGC
NM_005985.4	Snai1	F: TGCCCTCAAGATGCACATCCGA
		R: GGGACAGGAGAAGGGCTTCTC
NM_002046.5	GAPDH	F: TGTTCCTACCCCCAATGTGT
		R: TGTGAGGGAGATGCTCAGTG

F: forward; R: reverse.



Supplementary Figure 2. A. Univariate Cox analysis of Hazard ratio of key genes from Asian patient cohort. B. Multivariate Cox analysis of Hazard ratio of key genes from Asian patient cohort. C. The nomogram for predicting the probabilities of 1-, 3-, and 5-year overall survival (OS) in Asian patients. D. Univariate Cox analysis of Hazard ratio of key genes from White patient cohort. E. Multivariate Cox analysis of Hazard ratio of key genes from White patient cohort. F. The nomogram for predicting probabilities of 1-, 3-, and 5-year OS in White patients.