



Overall Survival Signature of 5-Methylcytosine Regulators Related Long Non-Coding RNA in Hepatocellular Carcinoma

Qi Pan¹, Caiyu Yi² and Yijie Zhang^{1*}

¹ Key Laboratory of Organ Transplantation of Liaoning Province, Department of Hepatobiliary Surgery and Organ Transplantation, First Hospital of China Medical University, Shenyang, China, ² China Medical University, Shenyang, China

OPEN ACCESS

Edited by:

Palmiro Poltronieri,
Italian National Research Council, Italy

Reviewed by:

Yuting He,
First Affiliated Hospital of Zhengzhou
University, China
Daniele Vergara,
University of Salento, Italy

*Correspondence:

Yijie Zhang
zhangyijie047314@hotmail.com

Specialty section:

This article was submitted to
Molecular and Cellular Oncology,
a section of the journal
Frontiers in Oncology

Received: 26 February 2022

Accepted: 08 April 2022

Published: 24 May 2022

Citation:

Pan Q, Yi C and Zhang Y (2022),
Overall Survival Signature of
5-Methylcytosine Regulators
Related Long Non-Coding RNA
in Hepatocellular Carcinoma.
Front. Oncol. 12:884377.
doi: 10.3389/fonc.2022.884377

Purpose: Studies reported that 5-methylcytosine (m5C) RNA transferase alters tumor progression; however, studies of m5C-related lncRNA remain lacking. This article intends to study the lncRNA modified by m5C RNA transferase in hepatocellular carcinoma using a combination of computational biology and basic experiments.

Method: We identified 13 m5C RNA transferase-related genes and selected long non-coding RNAs with a Pearson correlation coefficient greater than 0.4. Univariate Cox regression analysis was used to screen m5C RNA transferase lncRNA related to survival phenotype. We divided TCGA-LIHC into two types of m5C RNA using non-negative matrix decomposition. According to WGCNA, the co-expression models of two lncRNA regulation modes were constructed to analyze the characteristic biological processes of the two m5C RNA transferase-related lncRNA gene models. Then, a predictive model of m5C RNA transferase lncRNA was using LASSO regression. Finally, we used cell experiments, transwell experiments, and clone formation experiments to test the relationship between SNHG4 and tumor cell proliferation in Hep-G2 and Hep-3b cells line.

Results: We identified 436 m5C RNA transferase-related lncRNAs. Using univariate Cox regression analysis, 43 prognostic-related lncRNAs were determined according to $P < 0.001$. We divided TCGA-LIHC into two regulation modes of m5C RNA transferase using non-negative matrix factorization. The two regulation modes showed significant differences in overall and disease-free survival. We used LASSO to construct m5c-related lncRNA prognostic signature. Thus, a predictive m5C-lncRNA model was established using four lncRNAs: AC026412.3, AC010969.2, SNHG4, and AP003392.5. The score calculated by the m5C-lncRNA model significantly correlated with the overall survival of hepatocellular carcinoma. The receiver operating characteristic curve and

decision curve analysis verified the accuracy of the predictive model. We observed a more robust immune response in the high-risk score group. The transwell experiments and clone formation experiments suggested that m5C RNA transferase-related lncRNA SNHG4 promotes the proliferation and migration of Hep-G2 and Hep-3b cells line.

Conclusion: Two lncRNA expression patterns regulated by m5C RNA transferase were identified. The difference between the two expression patterns and the survival phenotype in the biological process was pointed out. A 5-methylcytosine RNA methyltransferases-related lncRNA overall survival signature was constructed. These results provide some understanding of the influence of m5C transferase on hepatocellular carcinoma. The prediction model of m5C transferase lncRNA has potential clinical value in managing hepatocellular carcinoma.

Keywords: 5-methylcytosine RNA methyltransferases, long non-coding RNA, weighted gene co-expression network analysis (WGCNA), liver hepatocellular carcinoma, prognosis model

INTRODUCTION

Hepatocellular cancer (HCC) is the sixth most common cause of malignant tumors. In 2020, there will be 900,000 new cases of stem cell cancer worldwide, making HCC the third leading cause of tumor-related death worldwide (1). HCC accounts for nearly 90% of primary liver cancers (2). Because the initial symptoms of HCC are not apparent, many patients are diagnosed with advanced liver cancer, hampering the success of treatment. In recent years, chemoradiotherapy for HCC has benefited patients with progressive disease; however, some patients remain with poor outcomes. Therefore, predicting the outcome of HCC patients by gene sequencing technology can assist clinicians in diagnosis and treatment strategies.

High heterogeneity is a significant feature of HCC. The primary characteristics of high heterogeneity are multiple genomic alterations and epigenetic modifications. Of these, epigenetic modifications are closely associated with tumor progression and metastasis and can be used as targets for cancer treatment. Epigenetics consists of the modification of DNA, RNA, and protein levels. Compared with the relatively limited spectrum of DNA modifications (six types), the abundance of RNA modifications is much higher. Post-transcriptional modification of RNA is an area of intense study. Of the 170 post-transcriptional modifications of RNA discovered to date, 2/3 are methylation modifications, including m1A, m6A, m5C, and m7G (3). Methylation of RNA 5-methylcytosine (m5C) is methylation at the fifth carbon atom of an RNA cytosine. This modification was discovered in rRNA in the 1970s and then successively in transport RNA, messenger RNA, and long non-coding RNA (lncRNA). M5C modification of RNA exists widely in cells and plays an essential role in

regulating gene expression and RNA stability. In addition, m5C methylation is associated with proto-oncogene activation, and m5C modified methyltransferase NSUN2 is differentially expressed in tumor and para cancer tissues.

lncRNA is defined as a DNA transcript with no coding protein action over 200 bp in length (4), first proposed in a study of mouse cDNA sequencing (5). lncRNA is classified as lncRNA, antisense lncRNA, bidirectional lncRNA, intragenic lncRNA, and intergenic lncRNA, depending on its location in the genome (6). RNA methylation of lncRNA has been demonstrated in cancer progression. For example, in HCC, the m6A “writer” METTL3 increases the stability of LINC00958 and promotes cancer progression (7). Similarly, m6A “eraser” ALKBH5 increases the invasion and metastasis of gastric cancer tumor cells by inhibiting the methylation of NEAT1 (8).

In the present study, we analyze 5mC RNA methyltransferase-related lncRNA using computational biology and basic experiments to provide a basis for studying the heterogeneity of HCC.

METHODS

Expression Collection

The gene transcripts and clinical features of the tumor tissues of patients with HCC were obtained from TCGA (<https://cancergenome.nih.gov/>), including 374 samples of HCC tissues and 50 samples of normal adjacent tissues. The clinical characteristics of patients included gender, survival status, survival time, tumor stage, and TNM stage.

Screening for Differential m5C-Related lncRNA

The “EdgeR” program package in RStudio software used applied, and “FDR < 0.1, |log₂FC| > 2” was the standard initially to screen the differentially expressed m5C related lncRNA. The “DESeq2” program package was used to identify differentially expressed m5C-related lncRNA according to “Padj < 0.05 and |log₂FC| > 2.”

Abbreviations: LIHC, Liver hepatocellular carcinoma; M5C, 5-Methylcytosine RNA methyltransferases; NMF, non-negative matrix factorization; WGCNA, Weighted correlation network analysis; LASSO, Least absolute shrinkage and selection operator; GSEA, Gene Set Enrichment Analysis; ESTIMATE, Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression data; SSGSEA, single sample gene set enrichment analysis.

Negative Matrix Factorization (NMF) Clustering of m5C Related lncRNAs Gene Sets

Thirteen m5C-related genes were collected from literature mining (9–20). Based on Pearson coefficient >0.4 and cox coefficient $P < 0.001$. The m5C related lncRNAs were uploaded as **Supplementary Table 1**—the 43 m5C related lncRNA genes for non-negative matrix dimensionality reduction clustering NMF. The non-negative matrix dimensionality reduction method was implemented using the “NMF” R package (21).

Weighted Correlation Network Analysis

A weighted standard expression network was constructed using the R language WGCNA package (22). The pickSoftThreshold function was used to obtain the optimal value of weighting parameters of adjacent parts, which was used as a soft threshold for subsequent network construction. Then, the weighted adjacency matrix was then constructed, and the related gene modules were built using hierarchical clustering based on the dissimilarity measure (1-Tom) of the topological overlap matrix (23). To determine the biological significance of each module, the potential correlation between genes and clinical traits was calculated using the characteristic genes of each module as the main component, and the expression patterns of genes of each module were summarized. Then, the correlation between the module significance and the average gene significance within the module was calculated. Finally, the correlation between the co-expression module and the expression pattern of NMF clustering subtypes was calculated.

LASSO Regression

The LASSO (24) regression algorithm was used to identify genes related to the outcome and survival of hepatocellular cancer patients and construct a risk-scoring model. The model's predictive performance was evaluated by the time-dependent receiver operating characteristic curve (ROC). Kaplan-Meier survival curves were used to compare survival differences of HCC patients between the two groups using the log-rank test.

GSEA

We used GSEA 4.1.0 software with the c2.cp.kegg.v7.0.symbols.gmt dataset in the Molecular Signature Database as the functional gene set to perform GSEA for patients in different risk groups (25). The iterative operations were set to 1000, and other parameters were set to default values.

The Proportion of Infiltrating Immune Cells in HCC

We used six methods to evaluate the relative proportion of immune infiltrating cells in the immune microenvironment, namely CIBERSORT (26, 27), EPIC (28), quanTIseq (29), MCPcounter (30), XCELL (31), and TIMER (32) algorithms to evaluate the immune response of different risk scores. We used Heatmap to analyze the differences in immune responses using the various algorithms.

The Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression data (ESTIMATE) is an algorithmic tool. The detailed algorithm is shown in **Supplementary File 2**.

The Correlation Between Risk Score and Immune Inflammation Response

We selected several classic immune-related sub-gene sets, including primary histocompatibility complex class II, lymphocyte-specific kinase, hematopoietic cell kinase, immunoglobulin G, signal transduction, and activation transcription 1, costimulatory molecule, interferon, and TNF gene sets (33). Genes with concentrations are displayed in **Supplementary Table 3**. We analyzed the association between risk scores and the genes associated with immune responses.

Cell Culture

The Hep-G2 and Hep-3b cell line was provided by the Shanghai Cell Bank of the Chinese Academy of Sciences. Cells were cultured in a complete DMEM medium containing 10% fetal bovine serum and placed in an incubator at 37°C and 5% CO₂. Cells were seeded in 6-well plates at 4×10^5 cells per well, and we observed cell fusion after culturing overnight for subsequent experiments.

Cell Transfection Experiment

We selected Hep-G2 and Hep-3b cells in the logarithmic growth phase, trypsinized them, and seeded them in 6-well plates. After adherence, according to the lentivirus packaging manual, we transfected the cells with a multiplicity of infection of 10. After 24 hours, we added two μ l of polybrene at a final concentration of 5 μ g/ml for screening for 1–2 weeks, incubated at 37°C, and changed the medium once according to cell status 8–12 hours. We transferred the successfully transfected cells from each group to the cell flask and continued culturing to obtain stable cells. Cells were grouped as follows: Si-NC group, si1-SNHG4 group, and si2-SNHG4 group. Si-SNHG4F: GTCAGCGAGCGAA CCAATTGGC; R: CCGATCGGCAGCCGCGCGCA.

RT-PCR Detection of SNHG4 Gene Expression

We extracted the total RNA from each group of cells after transfection and reverse transcribed the RNA into cDNA according to kit instructions. We designed the primer sequence and used the cDNA containing the amplified sequence as a template for PCR reaction. After the response, the results of each group were recorded, and GAPDH was used as an internal control to compare and analyze the expression of SNHG4 in each group. SNHG4 F: CCGCCGATAGGAGCGACACCCC AAC, R: AACCATCGAGCGGGGGCTCTCGCAA.

Clone Formation Experiment to Observe the Effect of SNHG4 Gene on the Proliferation of Hep-G2 and Hep-3b Tumor Cell Line

After the cells were transfected, we transferred cell suspensions to 1.5 mL Eppendorf tubes, mixed and diluted, and inoculated 6-well plates at 20,000/well. We changed the medium once every three days and cloned for about ten days to observe the

formation of cloning groups. The medium was then aspirated, and the cells were washed and fixed in 4% paraformaldehyde for 15 min, followed by staining with 0.1% crystal violet for 15 min. Finally, cells were washed, dried, and photographed to count clonal cell clusters and perform statistical analysis.

Transwell Method to Observe the Effect of SNHG4 Gene on the Migration of Hep-G2 and Hep-3b Cells

We added 200 μ L of HepG2 cell line suspension (1×10^4 cells) to the upper chamber of the Transwell chambers. The experiment was divided into regular cell group (si-NC), SNHG4 gene knockdown 1 group (si1-SNHG4), and SNHG4 gene knockdown 2 Group (si2-SNHG4). Cells were placed in a 37 °C incubator for 24 hours, after which the upper chamber was removed and washed with PBS three times. Cells were then fixed with paraformaldehyde for 20 minutes, stained with 0.1% crystal violet for 30 minutes after air-drying, and we randomly selected five fields under the microscope for counting. The number of cells and the ratio of the number of penetrating cells between the experimental and control groups represent cell migration changes. The Hep-3b cells line was tested using the HepG2 cell line.

Western Blotting

First, we used precooled RIPA buffer containing protease inhibitor (Thermo Scientific, USA) (Beyotime, Shanghai, China) to extract total protein from cells. Equivalent amounts of protein samples were isolated with 4-12% SDS-Page (GenScript, Nanjing, China) and then transferred to 0.45 μ m PVDF membrane (Millipore, USA). The membrane was sealed with TBST containing 5% skim milk for two h and incubated with primary antibody at four °C overnight. After washing with TBST 3 times, the antibody was coupled with HRP and incubated for one h at room temperature. Immunoblots were detected by an imaging system (Bio-Rad, USA) using an enhanced chemiluminescence detection kit (Servicebio, Wuhan, China). Western blots were performed using an imaging system (Bio-RAD, USA) using an enhanced chemiluminescence assay kit (Servicebio, Wuhan, China). The primary antibodies consisted of beta-catenin (Proteintech, 51067-2-AP), cyclin D1 (Cell Signaling Technology, 55506S), and GAPDH (Cell Signaling Technology, 5174S). The above antibodies are used in accordance with manufacturer's agreement and instructions

RESULTS

The Molecular Subtypes of lncRNA Regulated by m5C in HCC Based on the NMF Classification Method

The flow chart of the article is shown in **Figure 1A**. The *coxph* function in R was used to evaluate the predictive value of m5C-regulated lncRNA. According to the Pearson correlation coefficient greater than 0.4, we identified 436 m5C-related lncRNAs (**Figures 1B, C**). Then, according to the standard of

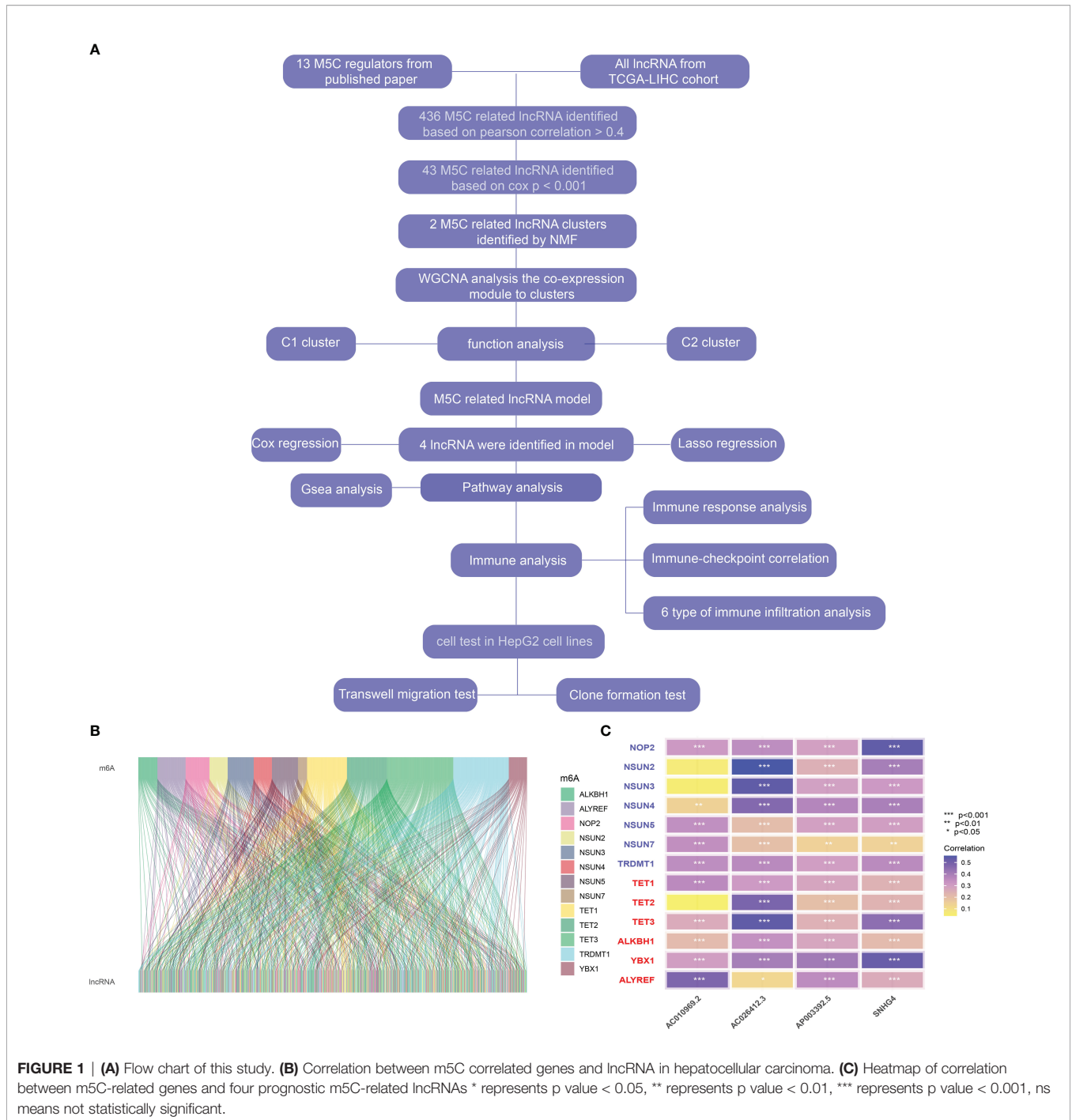
single-factor Cox regression $P < 0.01$, we obtained 436 cancer outcome-related m5C-regulated lncRNA genes. The significance and risk ratio of m5C-regulated lncRNA significant genes are shown in **Figure 2A**. We then performed non-NMF on these prognostic-related hepatocellular cancer lncRNA-related genes using 50 iterations. We conducted nine clusters; the number of collections k was 2-10, and the minimum sample of each group was set to 10 using the 'NMF' R package. According to three parameters (cophenetic, dispersion, and silhouette), we choose the ideal cluster group to be 2 (**Figures 2B, C**). We found that patients with different lncRNA gene expression patterns showed differences in overall survival and disease survival rates (**Figure 2D**; log-rank $p = 0.01$).

WGCNA Gene Co-Expression Network Analysis to Identify the Biological Characteristics of Different lncRNA Groups

We included the protein-coding genes and clinical samples in HCC into a WGCNA input file. In the subsequent investigation, we followed the omics cluster analysis to include samples with similar expression patterns. According to the cut-off value of 10000 and the β -value setting at 5, the gene in the smallest module is set to 30, and 18 co-expression modules are finally obtained (**Figures 3A, B**). The C1 lncRNA feature group strongly correlates with the brown module (**Figure 3C**; $Cor = 0.58$). We enriched the genes in the brown module that were associated with greater than 0.4 with the C1 group and found that the genes in the brown module were involved in the biological processes of oxidative phosphorylation and ATP metabolic process (**Figure 3D**). The lncRNA feature group of the C2 group had the strongest correlation with the yellow module (**Figure 3C**; $Cor = 0.28$). We enriched the genes in the yellow module that correlate greater than 0.4 with the C2 group. We found that yellow genes in the module were involved in the small molecule catabolic process, carboxylic acid catabolic process, and cellular amino acid metabolic process (**Figure 3D**).

Construction of HCC Outcome Model Based on lncRNA-Related Prognostic Genes

First, we randomly divided the entire TCGA-LIHC queue into training and validation sets. We arranged them in ascending order according to the ID of the sample and used SPSS to assign a random number to each sample for classification. The classification results satisfy the following criteria: 1) the two groups were similar in age distribution, clinical staging, follow-up time, and patient mortality; and 2) the gene expression profiles of the two randomized data sets were similar. Then we used LASSO regression to construct the lncRNA-related outcome model. First, we used the 43 prognostic-related m5C-related lncRNAs obtained above as input data and regression based on the overall survival rate as clinical follow-up data. This number of genes is not conducive to clinical detection. Therefore, to reduce the range of m5C-related lncRNAs while maintaining high accuracy, the R package *glmnet* was used to perform LASSO regression analysis with the trajectory of each independent



variable (**Figure 4A**). As the lambda increased, the independent coefficients also gradually increased, and the same was obtained for the independent coefficients. Three-fold cross-validation was used to build the model and analyze the confidence interval under each lambda. Finally, we constructed a predictive risk model containing the four genes. RiskScore = 0.75 * expAC026412.3 + 0.13 * expAC010969.2 + 0.15 * expSNHG4 + 0.33 * expAP003392.5 We calculated the RiskScore according to the expression level of the gene, and obtained the RiskScore

distribution of the sample (**Figure 4B**). The death rate of the high-risk samples was significantly higher than that of the samples with a low-risk score, indicating that the samples with high RiskScore had a worse outcome. We divided the RiskScore into high- and low-risk groups and drew Kaplan-Meier curves; there was a significant difference between the two (**Figure 4C**). We used ROC to classify RiskScore. We analyzed the 1.3 and 5-year forecast classification efficiency. The 5-year AUC area was 0.612, the 3-year AUC area was 0.636, and the 1-year AUC area was 0.746. Finally,

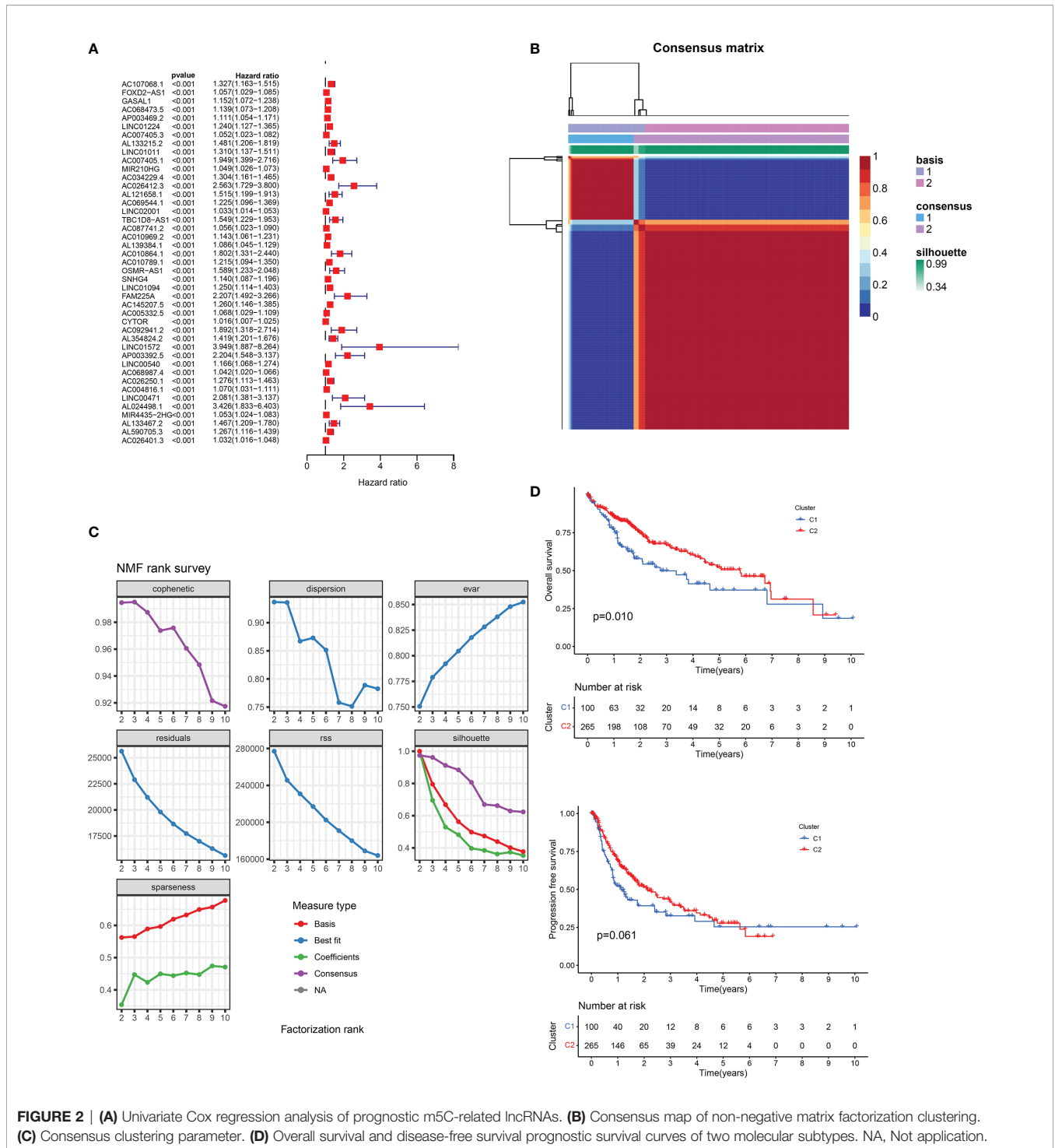


FIGURE 2 | (A) Univariate Cox regression analysis of prognostic m5C-related lncRNAs. **(B)** Consensus map of non-negative matrix factorization clustering. **(C)** Consensus clustering parameter. **(D)** Overall survival and disease-free survival prognostic survival curves of two molecular subtypes. NA, Not application.

the variables in the model were used as independent prognostic factors to assess patient risk (Figure 4D).

Evaluation of Model Results

We drew Kaplan-Meier curves for risk scores in the training and validation sets and found a significant difference between the high- and low-risk groups in the training and validation sets (Figure 5).

We analyzed the prediction classification efficiency of risk scores 1, 3, and 5 years in the training and validation sets (Figure 5). The 5-year AUC area in the training set was 0.629, the 3-year AUC area was 0.658, and the 1-year AUC area was 0.771. In the verification set, the 5-year AUC area was 0.578, the 3-year AUC area was 0.608, and the 1-year AUC area was 0.692. We included risk scores into different subgroups, such as age, stage, and others. We grouped

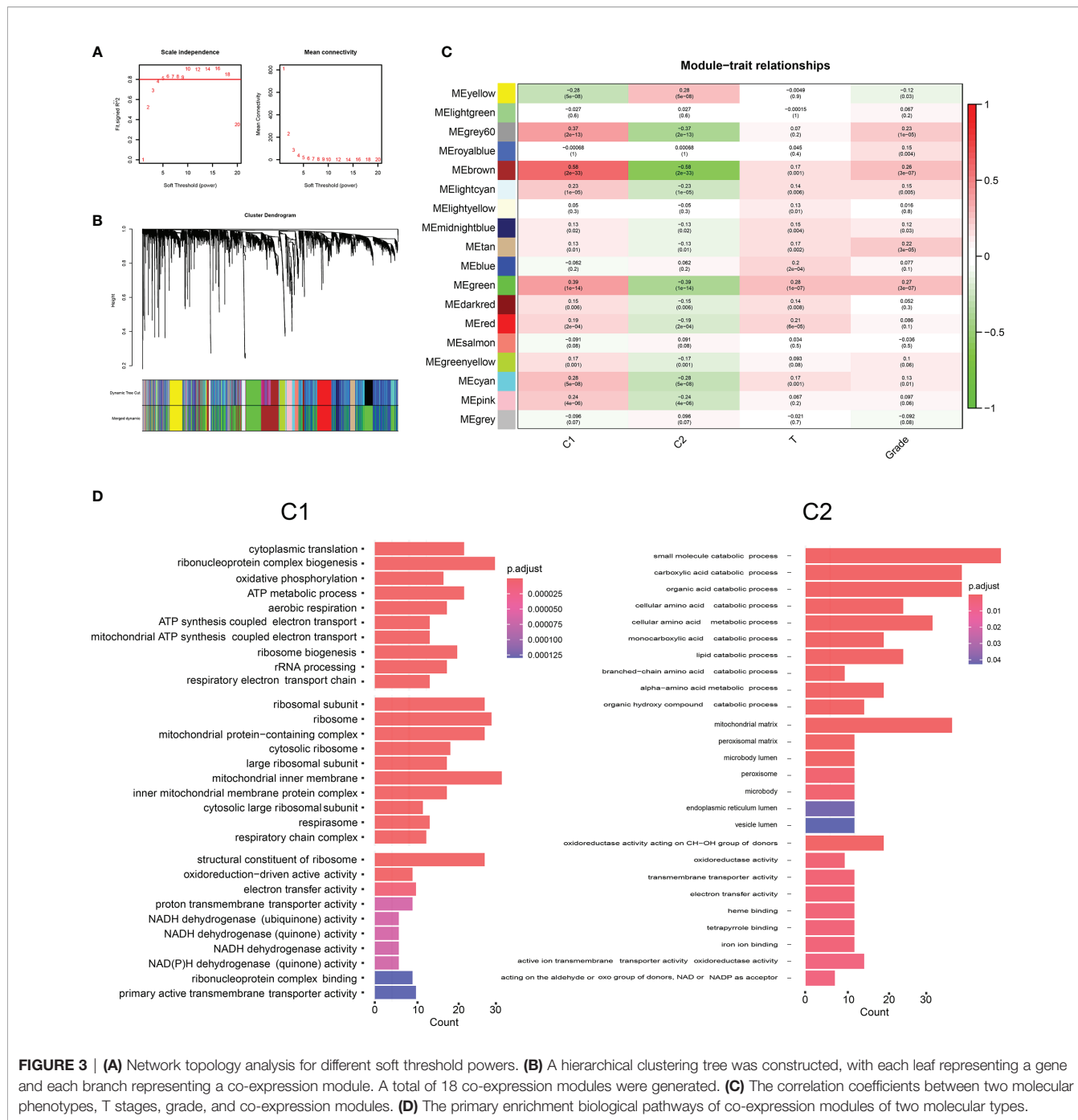
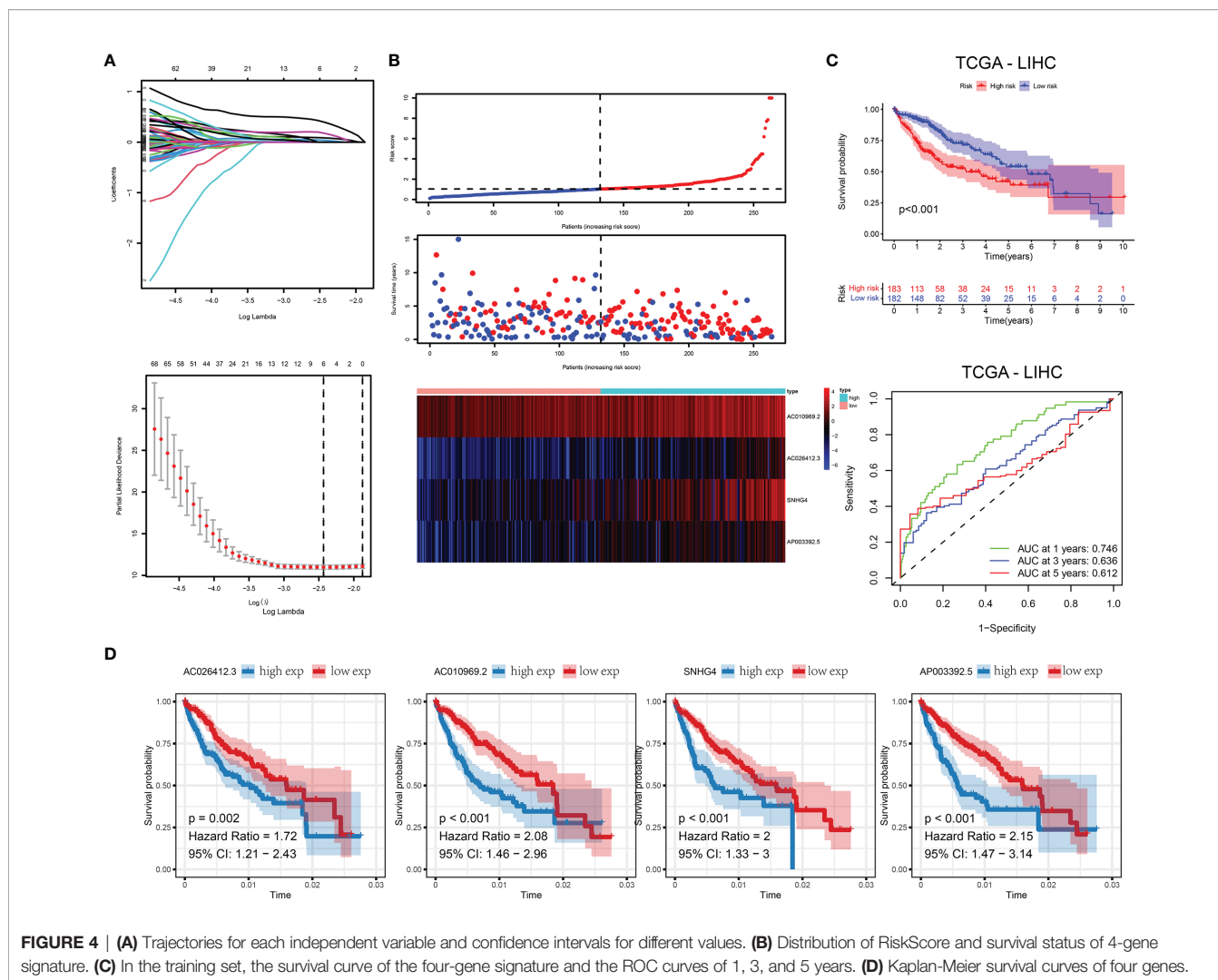


FIGURE 3 | (A) Network topology analysis for different soft threshold powers. **(B)** A hierarchical clustering tree was constructed, with each leaf representing a gene and each branch representing a co-expression module. A total of 18 co-expression modules were generated. **(C)** The correlation coefficients between two molecular phenotypes, T stages, grade, and co-expression modules. **(D)** The primary enrichment biological pathways of co-expression modules of two molecular types.

them according to subgroup indicators to evaluate the prognostic assessment ability of risk scores in the various subgroups (Figure 6A). The risk score distinguished patients with different outcomes in the whole cohort and patients in groups with characteristics such as age, stage, and others (Figure 6A). We then compared the area under the AUC curve of the nomogram, RiskScore, age, and staging and found that the area under the curve of the risk score in the training set was the largest, with the AUC area in the training set 0.749 (Figure 6B).

GSEA Analysis

We performed GSEA analysis in high- and low-risk patients to determine pathways related to the patient’s prognostic risk. As shown in Figure 6C, in patients with high-risk scores of HCC, cell cycle, cytokine-cytokine-receptor interaction, ECM receptor interaction, and other tumor-related pathways were enriched. In patients with low-risk scores of HCC, butanoate metabolism, fatty-acid metabolism, and tryptophan metabolism were enriched in several tumor metabolism-related pathways.



The Relationship Between Risk Score and Immune Microenvironment

The risk score positively correlated with inflammatory and immune responses. These reactions were induced by hematopoietic cell kinases, immunoglobulin G, interferon, lymphocyte-specific kinase, primary histocompatibility complex class I, major histocompatibility complex class II, and activator of transcription 1. Patients with higher risk scores had more clustered immune-inflammatory responses (Figure 7A).

The Relationship Between Risk Score and Immune Infiltration

The relationship between the level of immune cell infiltration and risk score evaluated based on the six methods of CIBERSORT, EPIC, quanTIseq, MCPcounter, XCELL, and TIMER is shown in Figure 7B. We found significant differences in the level of infiltration of macrophages and CD8⁺ T cells in different RiskScore groups.

The Effect of Knocking Out the SNHG4 Gene on the Clone Formation Ability of Hep G2 and Hep-3b Cells

The above paper constructed a predictive scoring model based on M5C methylation-related long non-coding RNA. The predictive scoring gene model contained four long non-coding RNAs: AC026412.3, AC010969.2, SNHG4, and AP003392.5. SNHG4 has been extensively studied in several cancers. Long Non-Coding RNA SNHG4 was a biomarker in Non-Small Cell Lung Cancer in colorectal cancer (34, 35). However, there are not enough studies on the effect of SNHG4 on liver cancer, and we used a cell assay to analyze the impact of SNHG4 in liver cell carcinoma. Clone formation experiments showed that the number of clones formed by Hep G2 and Hep-3b cell lines in the si-SNHG4 and si2-SNHG4 groups after culture and staining was significantly lower than that of the si-NC group ($P < 0.05$) (Figures 8A, B).

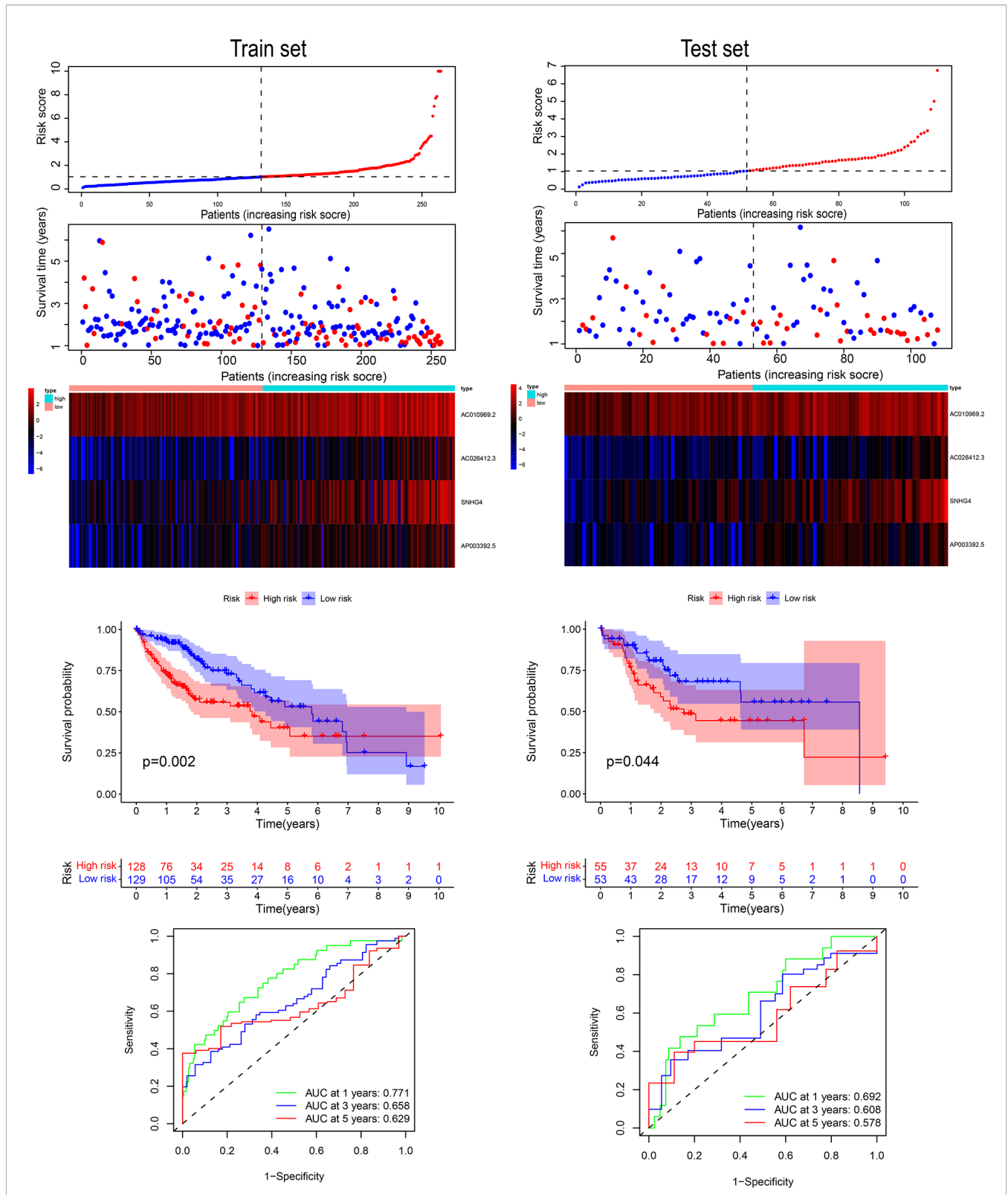


FIGURE 5 | The four-gene signature’s risk score distribution and survival status in the training and validation sets. ROC curves of 1, 3, and 5 years, and survival curves between the two risk groups based on the four-gene signature classification.

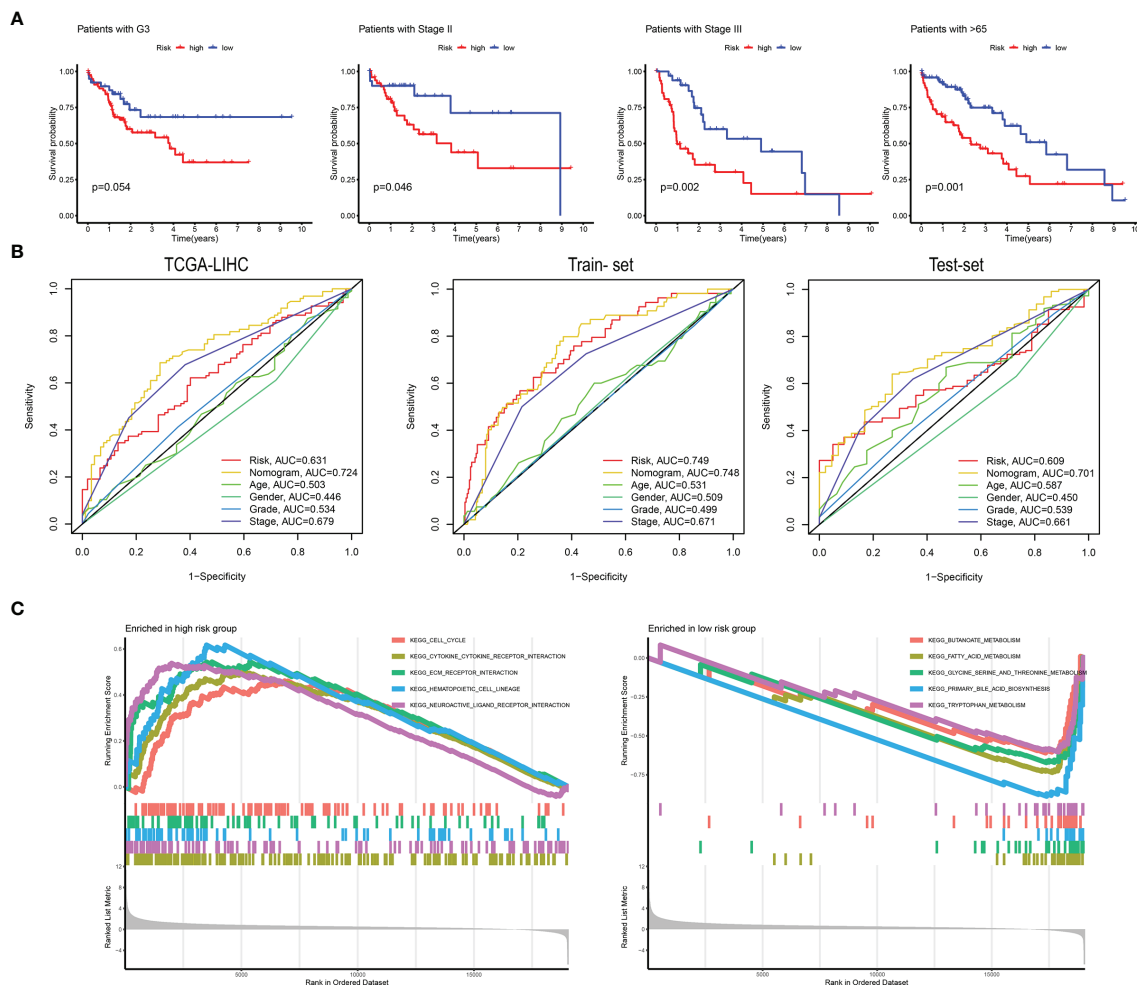


FIGURE 6 | (A) The high-risk and low-risk groups were divided according to G3, Stage II, Stage III, and 65 years. Survival curves were plotted. **(B)** ROC curves are grouped by risk, nomogram, age, sex, grade, and stage in TCGA-LIHC, training set, and validation set. **(C)** The main KEGG pathways are enriched in the high- and low-risk groups.

Transwell Assay

Compared with the si-NC control group, the number of Hep-G2 and Hep-3b cells passing through the Transwell chamber in the Si1-SNHG4 and si2-SNHG4 groups was significantly lower, suggesting SNHG4 promotes the migration of Hep-G2 and Hep-3b cells ($P < 0.05$) (Figures 8C, D).

Knocking SNHG4 Affects Wnt Signaling Pathways

After pathway enrichment analysis in the TCGA-LIHC cohort, SNHG4 was found to be closely associated with liver cancer progression. To verify the specific effect mechanism of SNHG4 on liver cancer, hepatocyte carcinoma cell lines of HEP-G2 and HEP-3B cells with SNHG4 knockdown cell lines were used for Western blotting analysis. The results showed that the expression level of cyclin D1 and β -catenin protein in the SNHG4 knockout group was significantly lower than that in the negative control group (NC) (Figures 9A, B). Three repeated experiments

demonstrated that knocking down SNHG4 down-regulated the WNT signaling pathway and affected the expression of cyclin D1.

DISCUSSION

The mechanism of m5C methylation modification of lncRNA is unclear; therefore, we attempted to comprehensively analyze lncRNA related to m5C methylation modification using computational biology. The m5C methylation modification of RNA is dynamically regulated by methyltransferase and demethyltransferase. Under the action of methyltransferase, RNA undergoes m5C modification and then combines with recognition protein to exert specific biological functions (3). The methyltransferases modified by m5C include NSUN1, NSUN2, NSUN3, NSUN4, NSUN5, NSUN7, and DNMT2. The point is a structurally conserved cysteine residue that catalyzes m5C methylation in various types of RNA with the

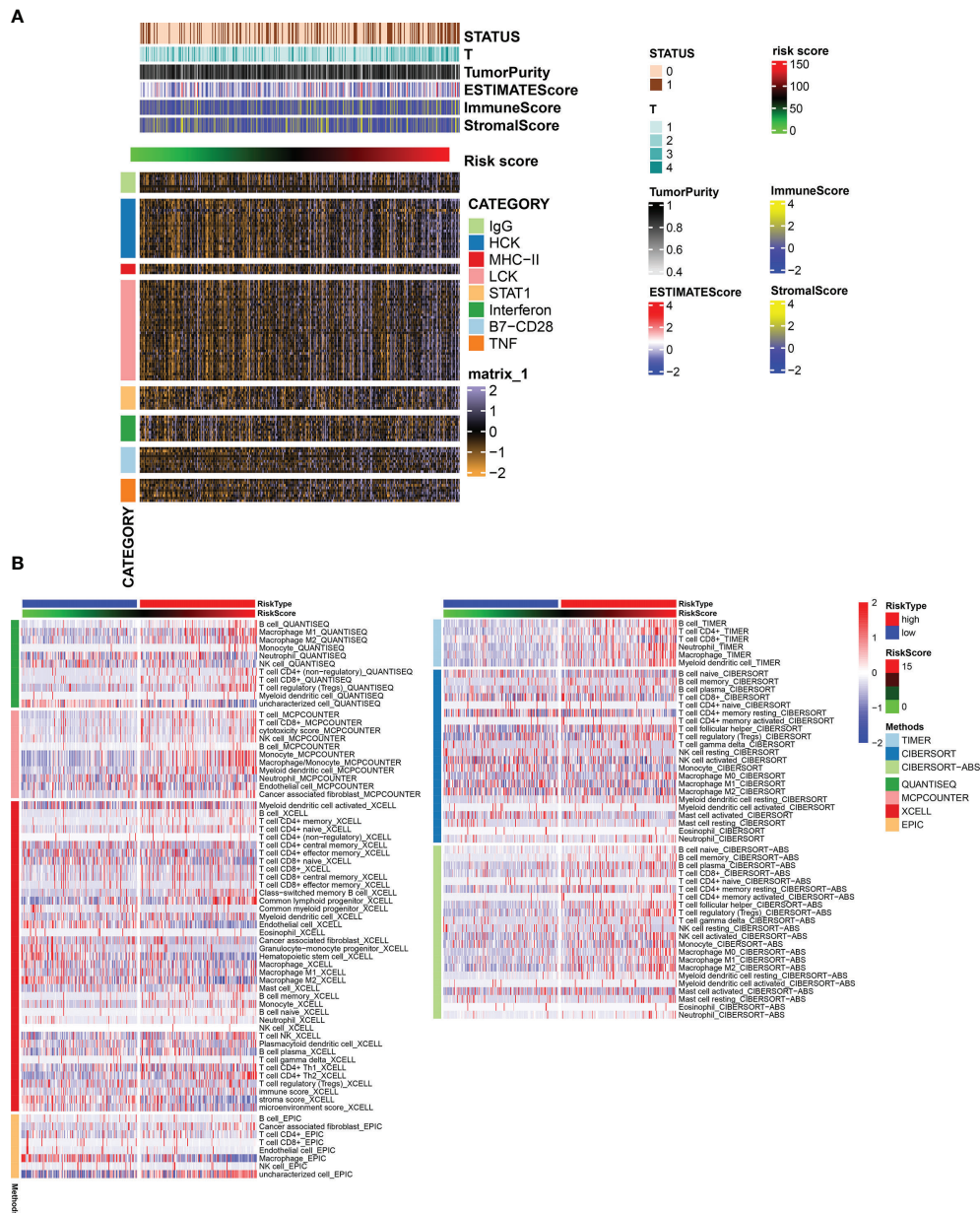


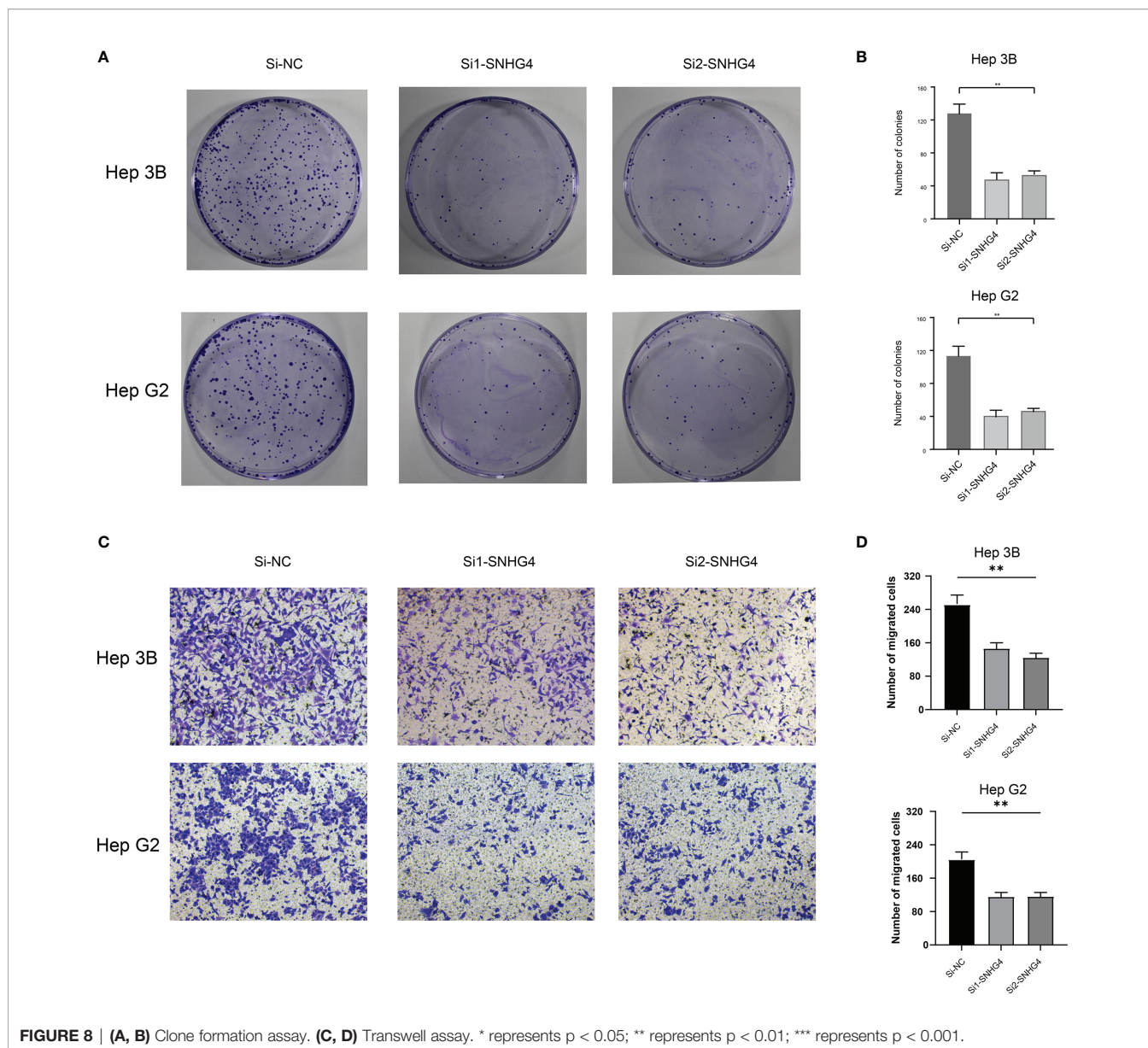
FIGURE 7 | (A) Heatmap for genesets associated with immune and inflammation. **(B)** Heatmap for immune responses based on TIMER, CIBERSORT, CIBERSORT-ABS, QUANTISEQ, MCPCOUNTER, XCELL, and EPIC algorithms among high- and low-risk groups.

help of the methylated donor S-adenosine-L-methionine. Therefore, we focused on the lncRNA, regulated by the aforementioned transferases.

It is believed that m5C methyltransferase regulates lncRNA in liver cancer. The role of m5C methylation in the occurrence and progression of cancer has been identified in liver cancer, including mRNA, microRNA, lncRNA, and other types of RNA. The m5C modification of RNA plays an essential regulatory role in the occurrence and progression of tumors (13). The m5C methyltransferase NSUN4 recognizes protein ALYREF associated with liver cancer outcomes. A study found

that expression levels of H19 lncRNA in cancer tissues were significantly higher than those in adjacent tissues. Other studies found that this effect was due to the m5C modification on H19 lncRNA mediated by NSUN2, which increases the stability of H19 lncRNA. H19 lncRNA with m5C change specifically binds to G3BP1 protein, further leading to the accumulation of oncoprotein and promoting the occurrence and progression of liver cancer (36).

We identified SNHG4 as an m5C methylation modification lncRNA in the present study. SNHG4 encodes small nucleolar RNA host gene 4. Some lncRNAs encode small nuclear RNA host



genes. In recent years, many studies have found that the abnormal expression of snRNAs may play the role of oncogenes in the development of tumors. For example, Chen found that SNHG8 was upregulated in non-small cell lung cancer (37). Other investigators found that the expression trend of SNHG8 in glioma and liver cancer was consistent with these studies (38–40).

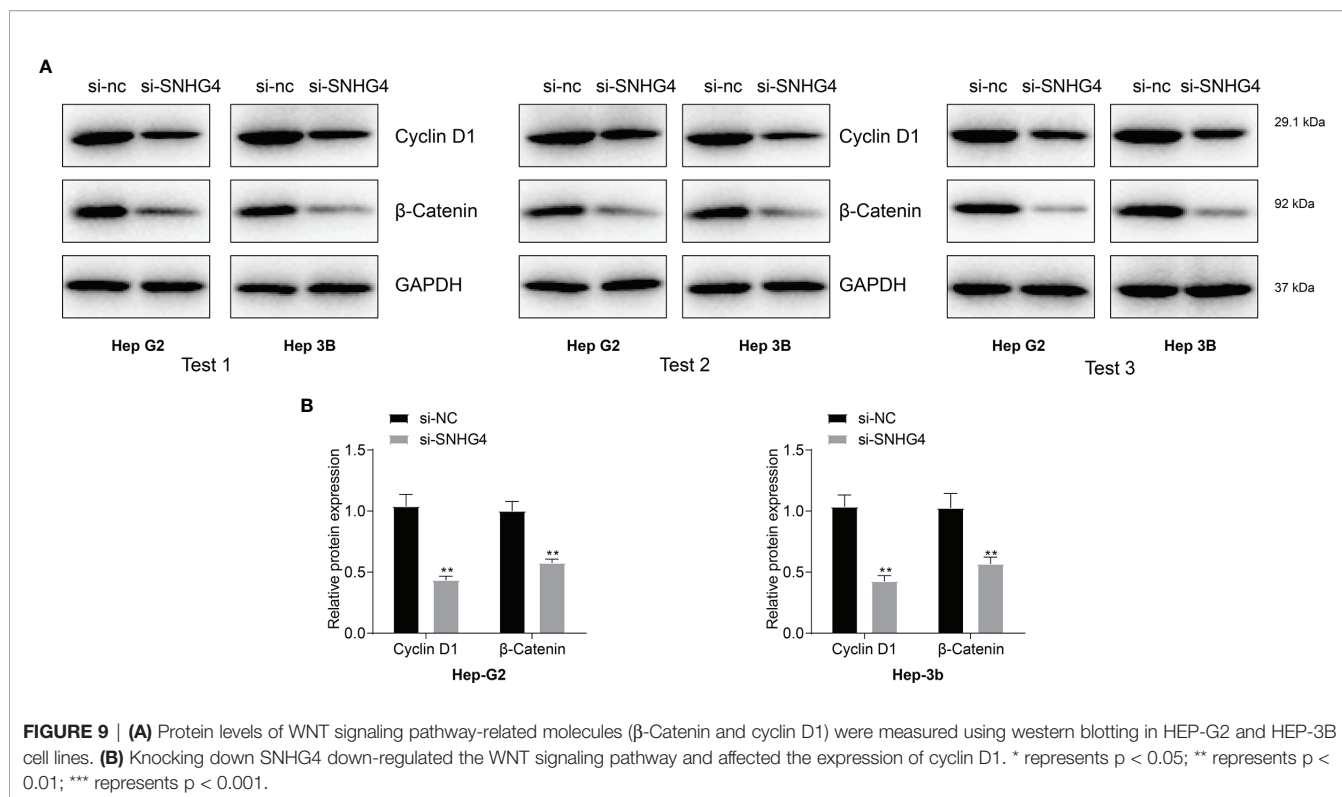
There are few in-depth studies discussing the predictive value of SNHG4. SNHG1, SNHG3, and SNHG20 are predictive biomarkers for neuroblastoma (41), ovarian cancer (42), and colorectal cancer (43), respectively. Zhu et al. conducted a bioinformatics analysis of lncRNA and found that SNHG4 may be a valuable prognostic marker in HCC (40). In the present study, we reached the same conclusion that the expression of SNHG4 was an independent predictor of poor outcomes in HCC. We further studied the predictive value of

SNHG4 in the subgroups and found its limitations in women and young patients, which may help direct precision therapy.

There are some limitations to our paper. We only analyzed lncRNAs associated with m5C transferase in TCGA; more sequencing cohorts are needed to validate our findings. This paper only conducted a comprehensive analysis of m5C related lncRNA and did not include a complete regulatory mechanism study.

CONCLUSION

We immediately identified 436 m5C transferase-related long non-coding RNAs and 43 prognostic-related lncRNAs related to m5C transferase. Four lncRNA were determined by LASSO regression to reduce the screening range further. Finally, we found that SNHG4



was significantly associated with the protein-coding gene of m5C methyltransferase. Cell experiments showed that knocking down SNHG4 affected the proliferation and migration of HCC. This comprehensive analysis of lncRNA regulated by m5C transferase provides a basis for future research on the methylation regulation of long-chain non-coding RNA.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: The datasets TCGA-LIHC for this study can be found at <http://cancergenome.nih.gov/>.

REFERENCES

- Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring Tumor Purity and Stromal and Immune Cell Admixture From Expression Data. *Nat Commun* (2013) 4:2612. doi: 10.1038/ncomms3612
- Forner A, Reig M, Bruix J. Hepatocellular Carcinoma. *Lancet* (2018) 391(10127):1301–14. doi: 10.1016/S0140-6736(18)30010-2
- Trixl L, Lusser A. The Dynamic RNA Modification 5-Methylcytosine and its Emerging Role as an Epitranscriptomic Mark. *Wiley Interdiscip Rev RNA* (2019) 10(1):e1510. doi: 10.1002/wrna.1510
- Mercer TR, Dinger ME, Mattick JS. Long non-Coding RNAs: Insights Into Functions. *Nat Rev Genet* (2009) 10(3):155–9. doi: 10.1038/nrg2521
- Hirota K, Miyoshi T, Kugou K, Hoffman CS, Shibata T, Ohta K. Stepwise Chromatin Remodeling by a Cascade of Transcription Initiation of Non-Coding RNAs. *Nature* (2008) 456(7218):130–4. doi: 10.1038/nature07348
- Li J, Tian H, Yang J, Gong Z. Long Noncoding RNAs Regulate Cell Growth, Proliferation, and Apoptosis. *DNA Cell Biol* (2016) 35(9):459–70. doi: 10.1089/dna.2015.3187
- Zuo X, Chen Z, Gao W, Zhang Y, Wang J, Wang J, et al. M6A-Mediated Upregulation of LINC00958 Increases Lipogenesis and Acts as a Nanotherapeutic Target in Hepatocellular Carcinoma. *Hematol Oncol* (2020) 13(1):5. doi: 10.1186/s13045-019-0839-x
- Dong F, Qin X, Wang B, Li Q, Hu J, Cheng X, et al. ALKBH5 Facilitates Hypoxia-Induced Paraspeckle Assembly and IL8 Secretion to Generate an Immunosuppressive Tumor Microenvironment. *Cancer Res* (2021) 81(23):5876–88. doi: 10.1158/0008-5472.CAN-21-1456
- Sato K, Tahata K, Akimoto K. Five Genes Associated With Survival in Patients With Lower-Grade Gliomas Were Identified by Information-Theoretical Analysis. *Anticancer Res* (2020) 40(5):2777–85. doi: 10.21873/anticancer.14250
- Blanco S, Bandiera R, Popis M, Hussain S, Lombard P, Aleksic J, et al. RNA Methyltransferase NSUN2 Promotes Gastric Cancer Cell Proliferation by

AUTHOR CONTRIBUTIONS

QP wrote the manuscript. CY and YZ revised the manuscript. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2022.884377/full#supplementary-material>

- Repressing p57Kip2 in an M5c-Dependent Manner. *Cell Death Dis* (2020) 11(4):270. doi: 10.1038/s41419-020-2487-z
11. Cheng JX, Chen L, Li Y, Cloe A, Yue M, Wei J, et al. Epigenetic Loss of RNA-Methyltransferase NSUN5 in Glioma Targets Ribosomes to Drive a Stress Adaptive Translational Program. *Acta Neuropathol* (2019) 138(6):1053–74. doi: 10.1038/s41467-018-03513-4
 12. García MG, Carella A, Urdinguio RG, Bayón GF, Lopez V, Tejedor JR, et al. Novel Long Noncoding RNA NMR Promotes Tumor Progression via NSUN2 and BPTF in Esophageal Squamous Cell Carcinoma. *Cancer Lett* (2018) 430:57–66. doi: 10.1016/j.canlet.2018.05.013
 13. He Y, Yu X, Li J, Zhang Q, Zheng Q, Guo W. Role of M(5)C-Related Regulatory Genes in the Diagnosis and Prognosis of Hepatocellular Carcinoma. *Am J Transl Res* (2020) 12(3):912–22.
 14. Chen X, Li A, Sun BF, Yang Y, Han YN, Yuan X, et al. Epigenetic Dysregulation of TET2 in Human Glioblastoma. *Oncotarget* (2018) 9(40):25922–34. doi: 10.18632/oncotarget.25406
 15. Gao Y, Wang Z, Zhu Y, Zhu Q, Yang Y, Jin Y, et al. NOP2/Sun RNA Methyltransferase 2 Promotes Tumor Progression via its Interacting Partner RPL6 in Gallbladder Carcinoma. *Cancer Sci* (2019) 110(11):3510–9. doi: 10.1111/cas.14190
 16. Janin M, Ortiz-Barahona V, de Moura MC, Martínez-Cardús A, Llinás-Arias P, Soler M, et al. RNA Cytosine Methylation and Methyltransferases Mediate Chromatin Organization and 5-Azacytidine Response and Resistance in Leukaemia. *Nat Commun* (2018) 9(1):1163. doi: 10.1038/s41467-018-03513-4
 17. Carella A, Tejedor JR, García MG, Urdinguio RG, Bayón GF, Sierra M, et al. 5-Methylcytosine Promotes Pathogenesis of Bladder Cancer Through Stabilizing mRNAs. *Nat Cell Biol* (2019) 21(8):978–90. doi: 10.1038/s41556-019-0361-y
 18. He Y, Yu X, Li J, Zhang Q, Zheng Q, Guo W, et al. Epigenetic Downregulation of TET3 Reduces Genome-Wide 5hmC Levels and Promotes Glioblastoma Tumorigenesis. *Int J Cancer* (2020) 146(2):373–87. doi: 10.1002/ijc.32520
 19. Mei L, Shen C, Miao R, Wang JZ, Cao MD, Zhang YS, et al. Stem Cell Function and Stress Response are Controlled by Protein Synthesis. *Nature* (2016) 534(7607):335–40. doi: 10.1038/nature18282
 20. Sato K, Tahata K, Akimoto K. Family-Based Exome-Wide Assessment of Maternal Genetic Effects on Susceptibility to Childhood B-Cell Acute Lymphoblastic Leukemia in Hispanics. *Cancer* (2016) 122(23):3697–704. doi: 10.1002/cncr.30241
 21. Gaujoux R, Seoighe C. A Flexible R Package for Nonnegative Matrix Factorization. *BMC Bioinf* (2010) 11:367. doi: 10.1186/1471-2105-11-367
 22. Langfelder P, Horvath S. WGCNA: An R Package for Weighted Correlation Network Analysis. *BMC Bioinf* (2008) 9:559. doi: 10.1186/1471-2105-9-559
 23. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási AL. Hierarchical Organization of Modularity in Metabolic Networks. *Science* (2002) 297(5586):1551–5. doi: 10.1126/science.1073374
 24. Tibshirani R. The Lasso Method for Variable Selection in the Cox Model. *Stat Med* (1997) 16(4):385–95. doi: 10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3
 25. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. *Proc Natl Acad Sci USA* (2005) 102(43):15545–50. doi: 10.1073/pnas.0506580102
 26. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining Cell Type Abundance and Expression From Bulk Tissues With Digital Cytometry. *Nat Biotechnol* (2019) 37(7):773–82. doi: 10.1038/s41587-019-0114-2
 27. Steen CB, Liu CL, Alizadeh AA, Newman AM. Profiling Cell Type Abundance and Expression in Bulk Tissues With CIBERSORTx. *Methods Mol Biol* (2020) 2117:135–57. doi: 10.1007/978-1-0716-0301-7_7
 28. Racle J, de Jonge K, Baumgaertner P, Speiser DE, Gfeller D. Simultaneous Enumeration of Cancer and Immune Cell Types From Bulk Tumor Gene Expression Data. *Elife* (2017) 6. doi: 10.7554/eLife.26476
 29. Finotello F, Mayer C, Plattner C, Laschober G, Rieder D, Hackl H, et al. Molecular and Pharmacological Modulators of the Tumor Immune Contexture Revealed by Deconvolution of RNA-Seq Data. *Genome Med* (2019) 11(1):34. doi: 10.1186/s13073-019-0655-5
 30. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the Population Abundance of Tissue-Infiltrating Immune and Stromal Cell Populations Using Gene Expression. *Genome Biol* (2016) 17(1):218. doi: 10.1186/s13059-016-1070-5
 31. Aran D, Hu Z, Butte AJ. Xcell: Digitally Portraying the Tissue Cellular Heterogeneity Landscape. *Genome Biol* (2017) 18(1):220. doi: 10.1186/s13059-017-1349-1
 32. Li B, Severson E, Pignon JC, Zhao H, Li T, Novak J, et al. Comprehensive Analyses of Tumor Immunity: Implications for Cancer Immunotherapy. *Genome Biol* (2016) 17(1):174. doi: 10.1186/s13059-016-1028-7
 33. Rody A, Holtrich U, Pusztai L, Liedtke C, Gaetje R, Ruckhaeberle E, et al. T-Cell Metagene Predicts a Favorable Prognosis in Estrogen Receptor-Negative and HER2-Positive Breast Cancers. *Breast Cancer Res* (2009) 11(2):R15. doi: 10.1186/bcr2234
 34. Wang F, Quan Q. The Long non-Coding RNA SNHG4/microRNA-Let-7e/KDM3A/p21 Pathway is Involved in the Development of non-Small Cell Lung Cancer. *Mol Ther Oncolytics* (2021) 20:634–45. doi: 10.1016/j.jomto.2020.12.010
 35. Zhou Z, Tan F, Pei Q, Li C, Zhou Y, Li Y, et al. lncRNA SNHG4 Modulates Colorectal Cancer Cell Cycle and Cell Proliferation Through Regulating miR-590-3p/CDK1 Axis. *Aging (Albany NY)* (2021) 13:9838–58. doi: 10.18632/aging.202737
 36. Sun Z, Xue S, Zhang M, Xu H, Hu X, Chen S, et al. Aberrant NSUN2-Mediated M(5)C Modification of H19 lncRNA is Associated With Poor Differentiation of Hepatocellular Carcinoma. *Oncogene* (2020) 39(45):6906–19. doi: 10.1038/s41388-020-01475-w
 37. Chen C, Zhang Z, Li J, Sun Y. SNHG8 is Identified as a Key Regulator in non-Small-Cell Lung Cancer Progression Sponging to miR-542-3p by Targeting CCND1/CDK6. *Onco Targets Ther* (2018) 11:6081–90. doi: 10.2147/OTT.S170482
 38. Dong J, Teng F, Guo W, Yang J, Ding G, Fu Z. lncRNA SNHG8 Promotes the Tumorigenesis and Metastasis by Sponging miR-149-5p and Predicts Tumor Recurrence in Hepatocellular Carcinoma. *Cell Physiol Biochem* (2018) 51(5):2262–74. doi: 10.1159/000495871
 39. Meng Q, Yang BY, Liu B, Yang JX, Sun Y. Long non-Coding RNA SNHG6 Promotes Glioma Tumorigenesis by Sponging miR-101-3p. *Int J Biol Markers* (2018) 33(2):148–55. doi: 10.1177/1724600817747524
 40. Zhu Q, Yang H, Cheng P, Han Q. Bioinformatic Analysis of the Prognostic Value of the lncRNAs Encoding snoRNAs in Hepatocellular Carcinoma. *Biofactors* (2019) 45(2):244–52. doi: 10.1002/biof.1478
 41. Sahu D, Hsu CL, Lin CC, Yang TW, Hsu WM, Ho SY, et al. Co-Expression Analysis Identifies Long Noncoding RNA SNHG1 as a Novel Predictor for Event-Free Survival in Neuroblastoma. *Oncotarget* (2016) 7(36):58022–37. doi: 10.18632/oncotarget.11158
 42. Hong L, Chen W, Wu D, Wang Y. Upregulation of SNHG3 Expression Associated With Poor Prognosis and Enhances Malignant Progression of Ovarian Cancer. *Cancer biomark* (2018) 22(3):367–74. doi: 10.3233/CBM-170710
 43. Li C, Zhou L, He J, Fang XQ, Zhu SW, Xiong MM, et al. Increased Long Noncoding RNA SNHG20 Predicts Poor Prognosis in Colorectal Cancer. *BMC Cancer* (2016) 16:655. doi: 10.1186/s12885-016-2719-x
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.
- Copyright © 2022 Pan, Yi and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.