Discover Oncology

Research

Exploring non-coding RNA expression profiles of *AKR1B10P1*, *RP11-465B22.3, WASH8P*, and *NPM1P25* as a predictive model for hepatocellular carcinoma patient survival

Khatereh Firouzi-Farsani[1] · Mina Dehghani-Samani[2] · Razieh Gerami[3] · Reihaneh Sadat Moosavi[1] · Marzieh Gerami[4] · Mohammad Mahdevar[5,6]

Received: 31 October 2024 / Accepted: 23 April 2025
Published online: 15 May 2025

Abstract
The primary aim of the study was to analyze novel long non-coding RNAs (lncRNAs) in hepatocellular carcinoma (HCC) to assess their roles as potential oncogenes and tumor suppressors and to develop a survival prediction model based on their expression levels. Data from The Cancer Genome Atlas, GSE135631, and GSE214846, were utilized to evaluate changes in lncRNA expression in HCC and their associations with patient prognosis. A risk model was created based on lncRNA expression to predict patient mortality. The co-expression network was employed to identify associated pathways, and the results were subsequently validated using the RT-qPCR method. The findings indicated that 14 lncRNAs were down-regulated in HCC, and their increased expression was associated with a favorable prognosis. Additionally, eight lncRNAs were overexpressed and correlated with poorer patient outcomes. The multivariate Cox regression analysis demonstrated that overexpression of *AKR1B10P1, RP11-465B22.3, WASH8P*, and the downregulation of *NPM1P25* could independently predict patient survival, irrespective of clinical variables. The risk score model based on these lncRNAs effectively stratified patients by their mortality risk. Furthermore, the co-expression network analysis revealed that the identified lncRNAs might be involved in various pathways, including fatty acid metabolism, mTOR signaling, glycolysis, angiogenesis, Wnt-β-catenin pathway, and DNA repair. RT-qPCR results validated the significant increase in the expression level of *WASH8P* in cancer specimens compared to normal tissues. Our results unveiled that changes in the expression levels of *AKR1B10P1, RP11-465B22.3, WASH8P,* and *NPM1P25* were significantly and independently associated with prognosis. Moreover, the patient mortality risk model constructed using these four lncRNAs exhibited a robust capacity to accurately predict patients' survival rates.

Keywords Prognosis · Gene expression · Non-coding RNAs · RT-qPCR · Liver cancer

Abbreviations

HCC	Hepatocellular carcinoma
TCGA	The cancer genome atlas

Khatereh Firouzi-Farsani and Mina Dehghani-Samani contributed equally to the paper as the first authors.

✉ Marzieh Gerami, m.gerami@iaushk.ac.ir; m.gerami64@gmail.com | [1]Department of Genetics, Faculty of Basic Sciences, Shahrekord University, Shahrekord, Iran. [2]Department of Bioinformatics, School of Advanced Technologies in Medicine, Isfahan University of Medical Sciences, Isfahan, Iran. [3]Independent Researcher, Graduated of PhD Pharmacology from Department of Basic Sciences, School of Veterinary Medicine, Shiraz University, Shiraz, Iran. [4]Department of Computer Engineering, Shahrekord Branch, Islamic Azad University, Shahrekord, Iran. [5]Department of Genetics and Molecular Biology, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran. [6]Genius Gene, Genetics and Biotechnology Company, Isfahan, Iran.

lncRNA Long non-coding RNA
K–M Kaplan–Meier
ROC Receiver operating characteristic

1 Introduction

Hepatocellular carcinoma (HCC) was one of the most prevalent malignancies worldwide in 2021 and is associated with a high fatality rate in patients [1]. Studies on this cancer have demonstrated that various molecular and environmental factors may be involved in the pathogenesis of this malignancy. Investigations into gene expression levels in HCC have shown that specific genes experience significant changes in their expression patterns, making them promising candidates for use in patient diagnosis, treatment, and prognosis [2]. Therefore, it is essential to identify the genes whose expression alterations are linked to the growth and progression of HCC.

Long non-coding RNAs (lncRNAs) are a class of non-coding RNAs that were initially characterized as non-functional when first discovered. However, it is now well-established that lncRNA can play significant roles in various cellular processes, including proliferation, differentiation, metabolism, and pathways related to cancer stem cells [3–5]. For instance, *PDIA3P1* has been demonstrated to be overexpressed in HCC, influencing the proliferation and apoptosis of HCC cell lines through the TP53 pathway [6]. Furthermore, research has indicated that the expression levels of lncRNAs, such as *DUXAP10* and *PLEKHA8P1*, can serve as promising biomarkers for the prognosis of HCC patients and their impact on proliferation and migration [7, 8]. Some studies show that lncRNAs can be helpful as a model for predicting survival. For example, it has been demonstrated that Cuproptosis-related lncRNA can be a viable model for predicting survival in HCC patients [9]. Another study also found that the expression level of RP11-325L7.2, DKFZP434L187, RP11-100L22.4, DLX2-AS1, and RP11-104L21.3 in HCC can provide a valid model for predicting survival [10]. These and other findings strongly suggest that lncRNA can play a pivotal role in the pathogenesis of HCC.

Numerous studies have indicated that long non-coding RNAs (lncRNAs) play a crucial role in the progression of HCC, yet the specific functions of many lncRNAs remain unclear. This study aimed to investigate alterations in the expression of all lncRNAs in HCC, based on data from The Cancer Genome Atlas (TCGA), and to evaluate their potential as prognostic markers. LncRNAs with potential oncogenic and tumor-suppressive roles were identified and presented. A predictive model for patient survival rates was also developed based on the expression of selected lncRNAs. We also identified potential pathways associated with the identified lncRNAs through co-expression network analysis. Finally, the RT-qPCR method was employed to validate the results by comparing the expression levels of candidate lncRNA in HCC samples and normal tissue.

2 Materials and methods

2.1 Data collection and preprocessing

TCGA transcriptome data was utilized to identify lncRNA potentially associated with HCC progression. To achieve this, raw RNAseq data of HCC, comprising 50 healthy and 374 tumor specimens at various stages, was obtained using the TCGAbiolinks package [11]. Initially, genes with negligible expression levels, defined by the criteria of having a CPM (count per million) less than 10 in more than 70% of the samples, were excluded using the edgeR package [12]. Subsequently, the data was normalized using the TMM (trimmed mean of M values) method and transformed into the log2 scale utilizing the limma package and the voom method [12]. All analyses conducted in this study were based on the normalized expression matrix. Furthermore, the most recent clinical data update from the TCGA database was retrieved and applied in survival and other analyses. The BioMart package was also utilized to annotate all lncRNA from the Ensembl genome browser. We used RNA-seq data from the GSE135631 and GSE214846 datasets for further validation. The GSE135631 dataset included 15 hepatocellular carcinoma (HCC) samples and 15 matched adjacent healthy samples, while the GSE214846 dataset comprised 65 HCC samples and 65 adjacent healthy samples. Similar analytical steps to those applied to the TCGA data were employed for the GSE135631 and GSE214846 datasets.

2.2 Survival analysis

The clinical data obtained from TCGA were used to categorize patients, allowing survival analysis to evaluate the prognostic significance of lncRNAs on patient outcomes. Patients with missing or very low survival data—specifically those recorded as 'NA,' '0,' or '1' days of life—were excluded from the analysis. Only patients who had a tumor at the time of death were considered among those whose condition was death. After this thorough filtering process, the study cohort was narrowed to 195 survivors and 56 deceased patients. To investigate the relationship between candidate lncRNA expression and patient prognosis, the expression values of the remaining lncRNAs were standardized into Z-scores. These standardized values were then used to evaluate the prognostic potential of the lncRNAs. Univariate Cox regression analysis used this new expression matrix to identify survival-associated lncRNAs. Subsequently, a multivariate Cox regression analysis was conducted to evaluate whether the selected prognostic lncRNAs remained independently significant after adjusting for clinical parameters, including pathological staging, TNM.T, age, gender, and BMI. Kaplan–Meier (K-M) curve analysis was employed to validate the findings. The cancer samples were stratified into two groups, 'high' and 'low,' based on the median expression levels of the candidate lncRNAs.

2.3 Risk score calculation

The multivariate Cox regression results were utilized to compute a risk score model, assessing the potential of lncRNA expressions for predicting the survival rates of patients with HCC. The risk score model was calculated using the formula below:

$$Riskscore = \sum_{j=1}^{n} W_j \times exp_{ij}$$

where Wj represents the multivariate coefficient (beta value) for lncRNA j, Expij denotes the expression value in the Z-score mode of lncRNA j in patient i, and n represents the number of tested lncRNAs. Additionally, the results were corroborated through K-M analysis and scatter plots. The median value of the Risk score was utilized as a threshold to stratify patients into high and low-risk groups. The predictive performance of the survival model was assessed using time-dependent receiver operating characteristic (ROC) curves generated by the timeROC package in R. The area under the ROC curve (AUC) was calculated at 5-year survival time points. Calibration of the model was evaluated using a calibration plot, comparing predicted survival probabilities with observed survival rates at 5 years. The calibration curves were generated using the rms package in R. A model is considered well-calibrated if the predictions closely follow the 45-degree diagonal line.

2.4 Co-expression network and data enrichment

To identify potential pathways associated with the candidate lncRNAs, we conducted a co-expression network analysis using the normalized expression matrix. This involved conducting a correlation test (Pearson) between the expression levels of candidate lncRNA and all other genes. We considered genes significant if they had an R-value exceeding 0.5 and a P-value less than 0.01. These significant genes were then used to construct the co-expression network. Gene enrichment analysis was conducted using the EnrichR database (https://maayanlab.cloud/Enrichr/) and the MSigDB repository.

2.5 Sample collection

The alterations in the expression of the identified lncRNAs were systematically assessed and subsequently validated using tissue specimens sourced from individuals diagnosed with HCC. All tissue samples were obtained from HCC patients referred to Milad Hospital in Isfahan during the period spanning from 2019 to 2021. This comprehensive dataset comprised 20 cancer tissue specimens and corresponding adjacent healthy tissues acquired during surgical procedures, as outlined in Table 1. All experiments were performed according to the relevant guidelines and regulations established by the Ministry of Health and Medical Education of Iran. Approval was granted by the Biomedical Ethics Committee of the

Table 1 Clinical information for HCC samples

Characteristic	Number (N = 20)
Age	
< 50	7
> 50	13
Gender	
Male	14
Female	6
TNM stage	
I	1
II	11
III	6
IV	2
Tumor size	
< 5cm	15
> 5cm	5

Islamic Azad University of Shahrekord (No.IR.IAU.SHK.REC.1402.072). Consent forms were obtained from all candidates involved in the study.

2.6 cDNA synthesis, primer design, and RT-qPCR

All tissue samples underwent a thorough washing procedure involving three successive rinses with phosphate-buffered saline (PBS) to decontaminate and remove necrotic cells. After the washes, RNA extraction was conducted using the TRIzol (Invitrogen) method. Subsequently, a DNase (Sigma) treatment was meticulously carried out to eliminate potential DNA contamination. The complementary DNA (cDNA) synthesis was then performed using the TaKaRa kit per the manufacturer's instructions. To evaluate the expression level of *WASH8P* in HCC and healthy specimens, at first, specific primers (F: 5′-CACAGGTAGACGAGGAC AAG-3′, R: 5′-TGTGTCCATGTCAGAGCAAC-3′) were designed using the primer-blast (https://www.ncbi.nlm.nih.gov/tools/primer-blast) tool. The RT-qPCR method with specific primers and SYBR Green was used for the mentioned gene expression quantification. The expression of each gene in each sample was calculated based on $2^{-\Delta Ct}$, and GAPDH was used as an internal control.

2.7 Statistics and software

All initial TCGA data preprocessing, including differential gene expression analysis, co-expression network construction, and survival analysis, was performed using R (Version 4.0.2). Graphical representations were generated using Prism GraphPad (Version 8). Differential gene expression was determined using the linear model method, with significance levels adjusted for multiple hypothesis testing. The relationship between lncRNA expression and patient survival was examined using Cox regression, with significance assessed using the log-rank test. The co-expression network was constructed using Pearson correlation analysis between candidate lncRNA expression and all genes within the TCGA dataset. Visualization of the co-expression network was achieved using Cytoscape (Version 4). The sensitivity and specificity of candidate gene expression levels in discriminating cancer samples from normal ones were assessed using ROC curves. In all analyses, a significance threshold of $P < 0.05$ was applied.

3 Results

3.1 Significant changes in lncRNA expression in HCC specimens compared to healthy

TCGA data was utilized to identify lncRNA potentially implicated in the development and progression of HCC. The differential gene expression analysis between cancer samples and healthy tissues revealed an increase in the

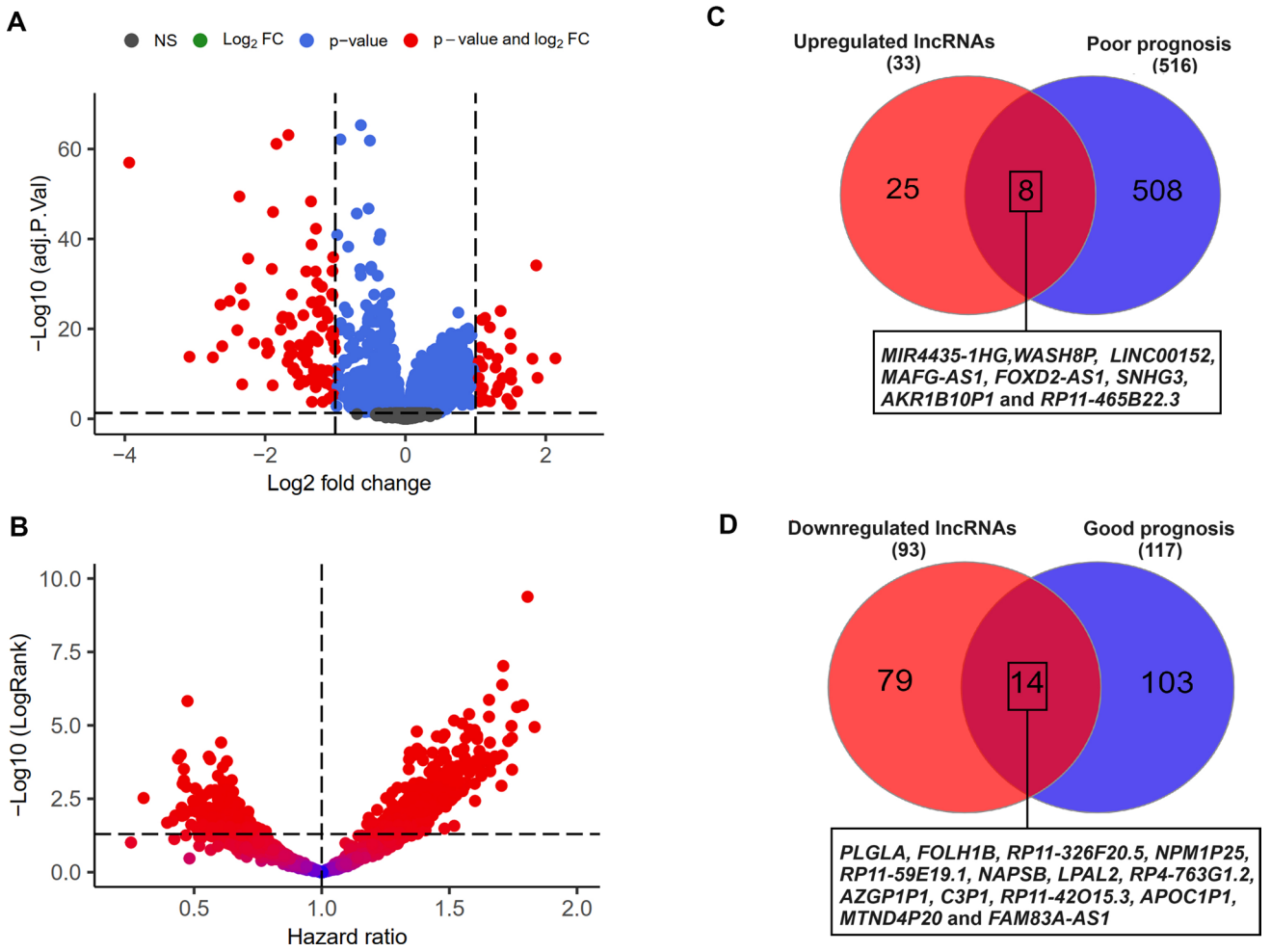

Fig. 1 Differential expression patterns of numerous lncRNAs in HCC Samples versus normal tissues. **A** Based on TCGA expression data, a volcano plot exhibiting differential expression of lncRNAs between HCC and healthy samples is shown. **B** The association of all lncRNAs expressed in HCC samples with patients' prognosis. Expression levels of 516 and 117 lncRNAs were associated with poor and good prognoses in patients, respectively. **C** Eight lncRNAs have shown both overexpression and poor prognostic patterns. **D** Fourteen lncRNAs were identified with downregulation and favorable prognostic characteristics

expression of 33 lncRNAs, while 93 lncRNAs exhibited decreased expression (Fig. 1A, |logFC| > 1 and FDR < 0.01). Also, the impact of the expressed lncRNAs on patients' prognoses was assessed through a univariate Cox regression analysis utilizing TCGA data. The results revealed that the expression of 516 lncRNAs was correlated with a poorer prognosis in patients, with a hazard ratio (HR) greater than one and a log-rank less than 0.05, as illustrated in Fig. 1B. Similarly, the expression of 117 lncRNAs was associated with a good prognosis in patients (HR < 1 and log-rank < 0.05, Fig. 1B). Subsequently, common lncRNAs were identified between the two previous analyses. As presented in Fig. 1C, eight lncRNAs, specifically *MIR4435-1HG, LINC00152, MAFG-AS1, FOXD2-AS1, SNHG3, WASH8P, AKR1B10P1*, and *RP11-465B22.3*, exhibited significant upregulation in HCC and were linked to a poor prognosis in patients. Conversely, a subset of 14 lncRNAs—*PLGLA, FOLH1B, RP11-326F20.5, NPM1P25, RP11-59E19.1, NAPSB, LPAL2, RP4-763G1.2, AZGP1P1, C3P1, RP11-42O15.3, APOC1P1, MTND4P20,* and *FAM83A-AS1*—exhibited significant downregulation in cancer samples. It demonstrated a strong association with favorable patient prognosis (Fig. 1D). The upregulated lncRNAs may act as potential oncogenes, while the downregulated lncRNAs could function as tumor suppressor genes in HCC. For further analysis, the identified lncRNAs were considered potential candidate genes.

3.2 Utilizing *AKR1B10P1, RP11-465B22.3, WASH8P*, and *NPM1P25* expression as a model for predicting patient survival rates

The previous findings confirmed the significant dysregulation of 24 lncRNAs and their association with patient prognosis. A multivariate Cox regression analysis was then conducted to evaluate their potential as independent predictors of patient survival. The outcomes of the multivariate test indicated that the overexpression of *AKR1B10P1*, *RP11-465B22.3*, and *WASH8P* was independently linked to a poorer prognosis among HCC patients (Table 2, HR > 1 and logRank < 0.05). Higher expression levels of *NPM1P25* were found to be independently linked to a better prognosis in patients (Table 2, HR < 1 and logRank < 0.05). Hence, it is plausible to suggest that the expression levels of these four lncRNAs may serve as suitable prognostic markers in HCC. As detailed in the Materials and Methods section, we constructed a survival prediction model (risk score model) utilizing the expression levels of these four lncRNAs, stratifying the patients into high-risk and low-risk groups. The results demonstrated that the expression of *AKR1B10P1, RP11-465B22.3, WASH8P,* and *NPM1P25* could significantly predict the risk of mortality in patients (Fig. 2A). Additionally, the K-M results revealed a significantly lower survival rate in the high-risk group compared to the low-risk group (Fig. 2B, logRank < 0.01). Additionally, an investigation was conducted to explore the relationship between the expression levels of all four lncRNAs and patient survival, aiming to support the findings. Elevated expression of *AKR1B10P1, RP11-465B22.3,* and *WASH8P* was significantly correlated with an unfavorable prognosis, whereas heightened expression of *NPM1P25* was notably associated with reduced patient mortality (Fig. 2C–F, logRank < 0.05). We performed a time-dependent ROC curve analysis to evaluate

Table 2 The multivariate Cox regression analysis was performed for the 24 candidate lncRNAs to explore their associations with patient prognosis while considering the influence of clinical parameters

	Univariate			Multivariate			
	HR	*P* value	95% CI	HR	*P* value	95% CI	Beta value
Pathological stage (stage I, II vs. stage III, IV)	**7.1**	**0.00001**	**4.4–11.6**	**6.9**	**0.0002**	**3.21–10.4**	**0.71**
TNM.T (T1,2 vs. T3,4)	2.6	0.002	1.6–4.1	2.1	0.07	0.94–1.7	0.35
Age (> 60 vs. < 60)	0.87	0.45	0.76–1.18	–	–	–	–
Gender (female vs. male)	0.93	0.67	0.65–1.3	–	–	–	–
BMI (> 30 vs. < 30)	1.07	0.15	0.91–1.1	–	–	–	–
PLGLA expression (high vs. low)	0.66	0.01	0.48–0.91	0.95	0.88	0.52–1.73	−0.04
FOLH1B expression (high vs. low)	0.65	0.03	0.45–0.98	0.71	0.09	0.48–1.06	−0.33
RP11-326F20.5 expression (high vs. low)	0.67	0.01	0.46–0.94	0.93	0.66	0.63–2.1	0.12
***NPM1P25* expression (high vs. low)**	**0.54**	**0.001**	**0.38–0.79**	**0.65**	**0.03**	**0.42–0.97**	**−0.42**
RP11-59E19.1 expression (high vs. low)	0.69	0.02	0.49–0.96	0.73	0.15	0.42–1.14	−0.36
NAPSB expression (high vs. low)	0.73	0.04	0.54–1	0.71	0.16	0.45–1.23	−0.2
LPAL2 expression (high vs. low)	0.64	0.001	0.48–0.84	0.98	0.91	0.58–1.6	−0.06
RP4-763G1.2 expression (high vs. low)	0.64	0.008	0.46–0.89	0.8	0.34	0.42–1.15	−0.35
AZGP1P1 expression (high vs. low)	0.62	0.006	0.44–0.88	0.89	0.63	0.57–1.39	−0.13
C3P1 expression (high vs. low)	0.65	0.01	0.46–0.91	0.81	0.39	0.48–1.32	−0.23
RP11-42O15.3 expression (high vs. low)	0.63	0.02	0.42–0.94	0.81	0.45	0.51–1.21	−0.2
APOC1P1 expression (high vs. low)	0.66	0.007	0.49–0.9	0.86	0.61	0.53–1.51	−0.14
MTND4P20 expression (high vs. low)	0.71	0.03	0.64–0.91	0.94	0.45	0.78–1.23	−0.11
FAM83A-AS1 expression (high vs. low)	0.67	0.01	0.54–0.88	0.82	0.14	0.69–1.25	−0.26
MIR4435-1HG expression (high vs. low)	1.39	0.006	1.1–1.79	1.2	0.43	0.75–2.02	0.21
LINC00152 expression (high vs. low)	1.43	0.01	1.04–1.99	1.03	0.84	0.7–1.55	0.03
MAFG-AS1 expression (high vs. low)	1.5	0.004	1.14–2.1	1.29	0.28	0.8–2.07	0.25
FOXD2-AS1 expression (high vs. low)	1.3	0.02	1.03–1.66	0.96	0.85	0.64–1.24	−0.03
SNHG3 expression (high vs. low)	1.55	0.0001	1.21–1.97	1.08	0.72	0.71–1.65	0.08
***WASH8P* expression (high vs. low)**	**1.46**	**0.01**	**1.07–2.13**	**1.47**	**0.01**	**0.95–2.31**	**0.39**
***AKR1B10P1* expression (high vs. low)**	**1.38**	**0.01**	**1.06–1.38**	**1.4**	**0.03**	**1.2–2.3**	**0.39**
***RP11-465B22.3* expression (high vs. low)**	**1.44**	**0.02**	**1.04–1.99**	**1.52**	**0.02**	**1.09–2.42**	**0.44**

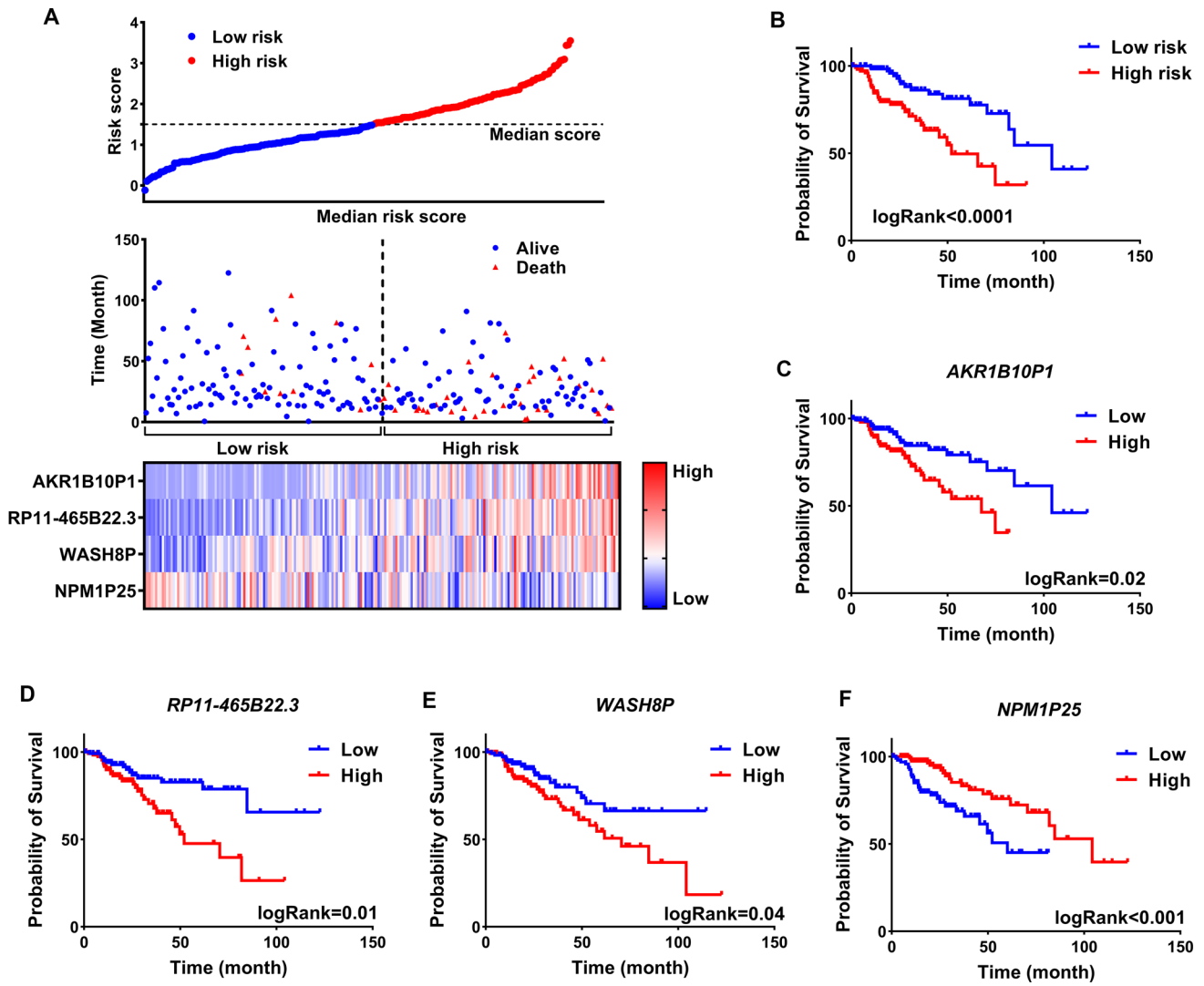

Fig. 2 The utility of *AKR1B10P1*, *RP11-465B22.3*, *WASH8P*, and *NPM1P25* expression as a robust predictive model for mortality rates in HCC patients. **A** Based on the computed risk score model, the scatter plot and heatmap illustrate the expression of *AKR1B10P1*, *RP11-465B22.3*, *WASH8P*, and *NPM1P25* in cancer samples and their association with patient mortality. **B** The K-M survival curve, when comparing high-risk and low-risk patient groups, illustrates that individuals with elevated expression of *AKR1B10P1*, *RP11-465B22.3*, *WASH8P*, and decreased expression of *NPM1P25*, are associated with a higher mortality rate. **C–F** The K-M diagram for four survival model genes is depicted. The median expression of each gene serves as the cut-off

the model's discriminative ability. At 5 years, the AUC was 0.76 ($P < 0.0001$, Fig. 3A). These results suggest that the model maintains substantial predictive accuracy over time. The calibration of the prognostic model was evaluated to assess how well the predicted survival probabilities matched the observed survival outcomes at different time points. Calibration curves were plotted for 5-year survival predictions, comparing the predicted probabilities with the actual observed survival rates. At the 5-year survival time point, the calibration plot showed that the predicted probabilities closely followed the 45-degree diagonal line, indicating excellent calibration (Fig. 3B). These results suggest that the expression levels of *AKR1B10P1*, *RP11-465B22.3*, *WASH8P*, and *NPM1P25* could independently serve as prognostic factors for the survival of HCC patients, irrespective of clinical characteristics.

3.3 Elevated *WASH8P* expression in HCC samples: unveiling its potential as a diagnostic biomarker

Data from the GSE135631 and GSE214846 studies were utilized to validate the results further. As shown in Fig. 3C, D, the expression levels of *AKR1B10P1*, *RP11-465B22.3*, and *WASH8P* were higher in tumor samples compared to healthy samples (FDR < 0.01). In contrast, *NPM1P25* showed decreased expression (Fig. 3C, D, FDR < 0.05). To confirm

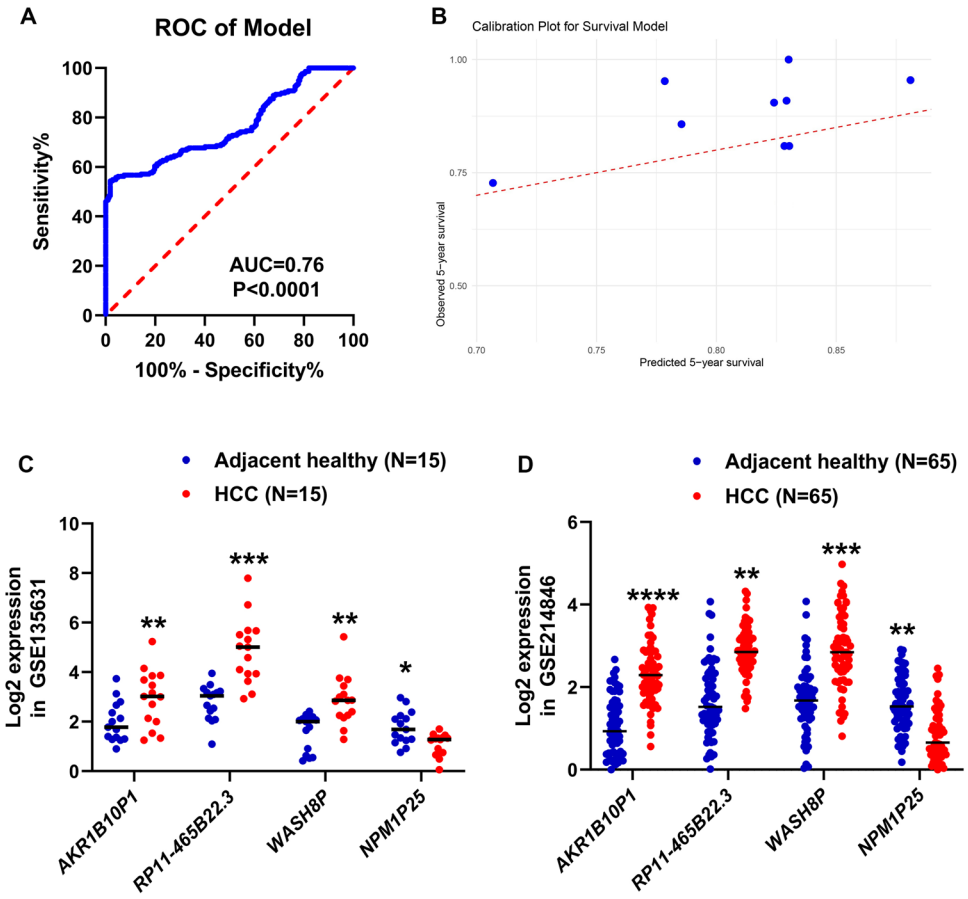

Fig. 3 Increased expression of candidate lncRNAs in other studies associated with HCC. **A** The results of the ROC curve for survival prediction based on the model results are shown. False positive and true positive values were used to draw the curve. **B** The scatter plot of the model calibration is shown. All samples were divided into nine equal groups to check the model calibration. **C**, **D** Other studies show the results of candidate lncRNAs' expression changes, including GSE135631 and GSE214846 (*FDR < 0.05; **FDR < 0.01; ***FDR < 0.001; and ****FDR < 0.0001)

the findings obtained from the in silico analyses, the expression level of *WASH8P* in HCC samples was compared to healthy samples using RT-qPCR. The preceding findings demonstrated a significant upregulation in the expression level of *WASH8P* in HCC samples when compared to healthy samples based on TCGA data (Fig. 4A). Furthermore, the results of the ROC curve analysis indicated that the expression level of this lncRNA could effectively differentiate between cancer samples and healthy samples in TCGA data (Fig. 4B, AUC = 0.77 and $P < 0.0001$). Our investigations into the expression level of *WASH8P* in HCC revealed a significant increase in its expression in cancer samples compared to healthy samples (Fig. 4C, $P < 0.01$). Additionally, our data showed that its expression level could perfectly discriminate between cancer and healthy samples (Fig. 4D, AUC = 0.9 and $P < 0.0001$).

3.4 Significant correlation of oncogenic genes with candidate lncRNAs

A co-expression network was constructed to better understand the pathways linked to the lncRNAs identified in the previous steps. The outcomes of the co-expression network analysis indicated that a total of 136 genes displayed a substantial correlation with the expression levels of *AKR1B10P1*, *RP11-465B22.3*, and *WASH8P*, meeting the criterion of a correlation coefficient greater than 0.5 and $P < 0.0001$ (Fig. 5A–C). It was also found that these 136 genes are involved in pathways related to cancer malignancy, including reactive oxygen species, mTORC1 signaling, glycolysis, angiogenesis, and Wnt-beta catenin (Fig. 5D, FDR < 0.05). Conversely, 54 genes exhibited a strong and significant correlation with the expression level of *NPM1P25* (Fig. 6A, R > 0.5A and $P < 0.01$). The enrichment analysis revealed that specific genes within the *NPM1P25* network are associated with metabolism-related pathways, including xenobiotic metabolism, fatty acid metabolism, and DNA repair (Fig. 6B, FDR < 0.05). The findings suggest that the candidate lncRNAs may play a crucial role in developing HCC by influencing the pathways mentioned earlier, highlighting their potential importance in the disease's progression.

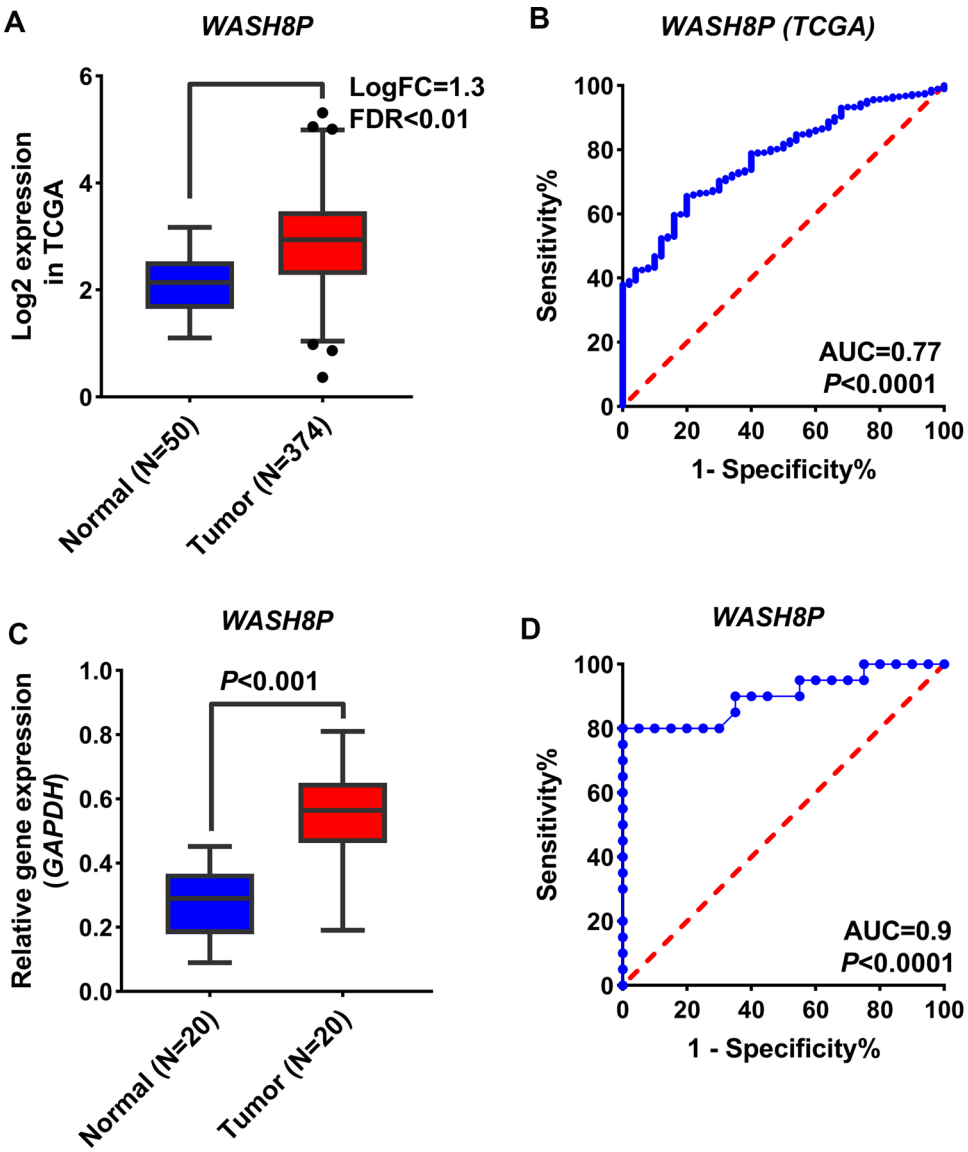

Fig. 4 *WASH8P* higher expression level in HCC patients. **A** Based on TCGA data, WASH8P expression levels in HCC specimens are shown to be higher than those in healthy ones. **B** The ROC curve analysis results were presented to assess the differentiation of healthy samples from cancer samples based on *WASH8P* expression. **C** The expression level of *WASH8P* in cancer samples showed a significant increase compared to healthy by RT-qPCR. **D** A ROC curve was constructed to assess the WASH8P expression levels in cancer samples as compared to normal samples, utilizing RT-qPCR data

4 Discussion

In 2020, HCC was one of the most prevalent cancers globally [13]. Despite extensive efforts to understand the factors contributing to cancer development and treatment at both clinical and fundamental levels, the patient mortality rate remains alarmingly high. Research has revealed that long non-coding RNAs (lncRNAs), initially perceived as insignificant, may play a pivotal role in the pathogenesis of various cancers, including HCC [14]. Moreover, they have been demonstrated to be overexpressed in cancer cells, making them potential targets for therapeutic interventions and promising candidates for precise diagnosis and prognostic markers [15]. This study rigorously assessed the dynamic alterations in the expression profiles of every lncRNA within the context of HCC. It delved into the intricate web of connections between their expression patterns, the prognosis of cancer, and their functional implications in cellular pathways. Considering the updating of clinical data and the discovery of new lncRNAs, identifying survival prediction models based on gene expression can be crucial.

Our results showed that the expression levels of *PLGLA, FOLH1B, RP11-326F20.5, NPM1P25, RP11-59E19.1, NAPSB, LPAL2, RP4-763G1.2, AZGP1P1, C3P1, RP11-42O15.3, APOC1P1, MTND4P20,* and *FAM83A-AS1* were significantly reduced in HCC samples, and this reduction was associated with improved patient prognosis. Research findings have unequivocally demonstrated that *PLGLA* expression undergoes a significant decrease in HCC, strongly correlated

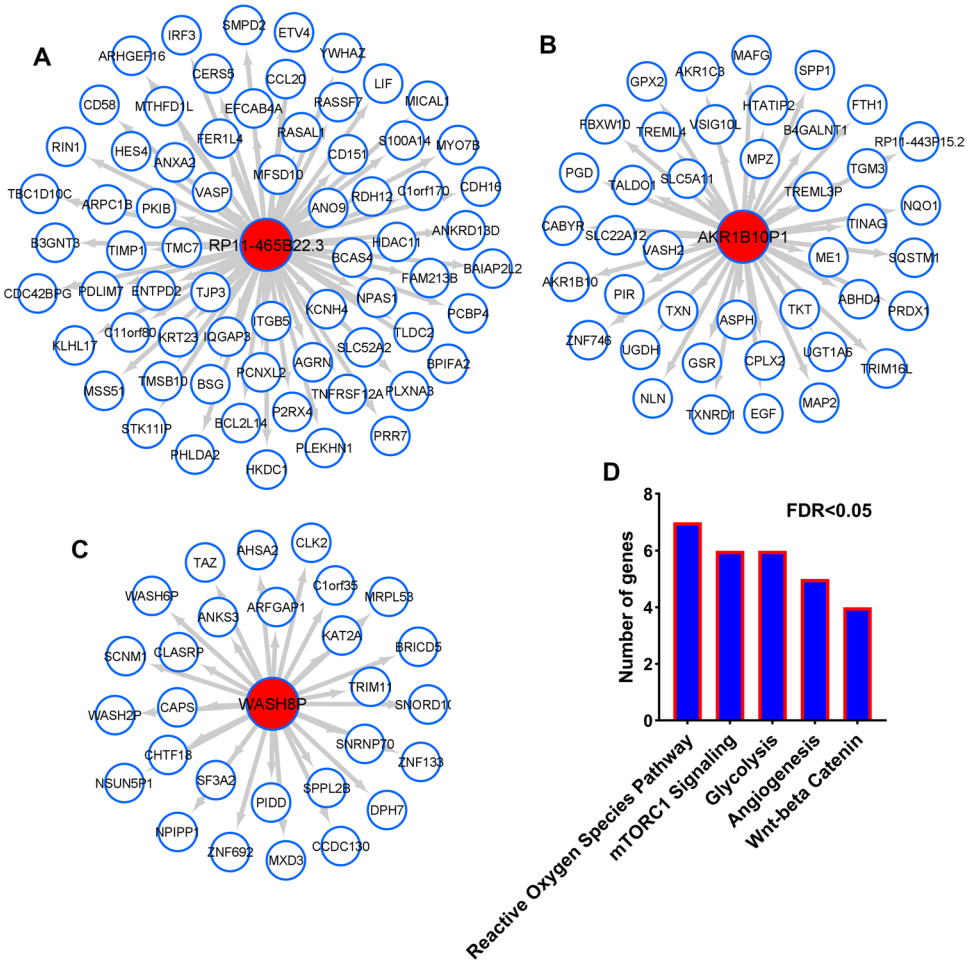

Fig. 5 Elevated *AKR1B10P1*, *RP11-465B22.3,* and *WASH8P* expression linked to key oncogenic pathways. **A–C** The co-expression network is shown for the lncRNAs AKR1B10P1, RP11-465B22.3, and WASH8P, which are both overexpressed and associated with poor prognostic outcomes. **B** The genes within the co-expression network were found to be associated with reactive oxygen species, mTORC1 signaling, glycolysis, angiogenesis, and the Wnt-beta catenin pathway, as per the MSigDB data

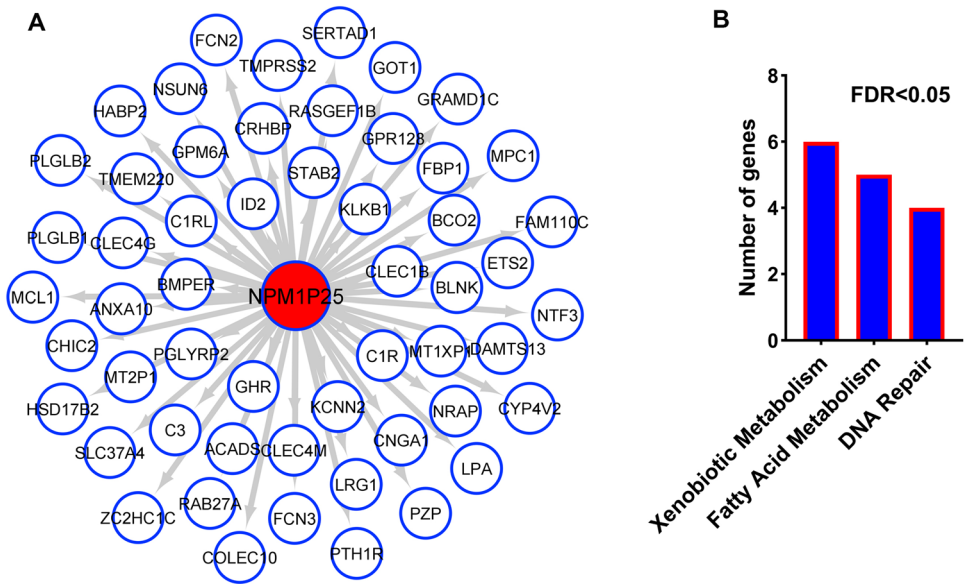

Fig. 6 *NPM1P25* expression in HCC Shows co-expression with genes associated with metabolic pathways. **A** A co-expression network is visualized for NPM1P25. **B** The co-expression network pertaining to *NPM1P25* revealed gene involvement in pathways such as fatty acid metabolism, DNA repair, and xenobiotic metabolism. Enrichment results are based on MSigDB data

with an adverse prognosis for patients. Moreover, the expression levels of *PLGLA* can influence the migration, cell proliferation, and invasive potential of HCC cell lines, thus establishing its function as a tumor suppressor [16]. Furthermore, there is compelling evidence to suggest that the expression level of *NAPSB* experiences a marked

decrease in HCC, which can be closely associated with the mortality rate of patients. Additionally, the expression level of this lncRNA has been shown to influence drug sensitivity and resistance in the context of HCC [17]. Studies have reported that *LPAL2* exhibits a marked decrease in expression in HCC; this downregulation is consistently linked to the survival rates of patients. *LPAL2* has been shown to act as a tumor suppressor in HCC by regulating the expression of genes related to metastasis [18]. Previous reports have indicated a decrease in *C3P1* levels in HCC, suggesting a potential correlation with patients' survival rates [19]. Furthermore, the findings suggest a decreased expression of *MTND4P20* and its likely correlation with HCC [20]. The other lncRNAs mentioned have been less thoroughly studied; therefore, this is the first report on their influence regarding cancer progression and prognosis. According to the co-expression network, downregulated lncRNAs were involved in pathways such as DNA repair, fatty acid metabolism, and xenobiotic metabolism. Fatty liver and fatty acid metabolism disorders can activate signaling pathways in the liver that are linked to oncogenes. Inflammation is a side effect of these diseases, raising the risk of HCC [21, 22]. Consequently, the downregulated lncRNA observed in response to metabolic alterations in fatty acids may play a pivotal role in the pathogenesis of HCC. Nevertheless, further in vitro validation is imperative to elucidate the underlying mechanism.

Our studies have revealed a significant increase in the expression levels of several lncRNAs, namely, *MIR4435-1HG, LINC00152, MAFG-AS1, FOXD2-AS1, SNHG3, WASH8P, AKR1B10P1,* and *RP11-465B22.3*, in HCC samples compared to normal tissues. This upregulation is associated with a poor prognosis among the patients. Studies show that *AKR1B10P1* expression is dramatically enhanced in HCC and is related to metastatic pathways [23]. Furthermore, the heightened expression of MIR4435-1HG in HCC and its association with patient mortality rates have been documented [24]. Several investigations have indicated an elevation in *LINC00152* expression in HCC, highlighting its role in the disease's pathogenesis [25, 26]. Conversely, evidence demonstrates that *MAFG-AS1* can enhance malignancy in HCC by promoting cell proliferation and invasion [27]. The oncogenic role of *FOXD2-AS1* and *SNHG3* in HCC has also been mentioned [28, 29]. This study marks the first instance of our observation that, aside from the lncRNAs mentioned above, *WASH8P* and *RP11-465B22.3* may also assume an oncogenic role in HCC and exhibit potential associations with patients' survival rates. Furthermore, the ex vivo findings of this study have shown a substantial increase in the expression of *WASH8P* in cancer samples compared to healthy samples. The co-expression network results also indicate that the mentioned overexpressed lncRNAs were associated with the main pathways of cancer cells, such as reactive oxygen species, mTORC1 signaling, glycolysis, angiogenesis, and Wnt-beta catenin. Extensive studies show that these pathways play a significant role in the development and progression of HCC [30, 31]. Although the findings of this study are based on in silico analyses, in-vitor confirmations are needed to gain a deeper understanding of the molecular mechanisms underlying the lncRNA stated.

The multivariate Cox regression analysis highlights that the expression of *AKR1B10P1, RP11-465B22.3, WASH8P*, and *NPM1P25* can serve as independent prognostic biomarkers for HCC. Risk scores based on these lncRNAs effectively predict mortality risk in HCC patients, suggesting their potential as targets for diagnosis and treatment. One of the analytical limitations in this study is the lack of artificial intelligence tools due to the small number of samples, which could help improve the model by increasing the number of samples.

5 Conclusion

The findings of this study reveal significant changes in the expression of various lncRNAs in HCC, which are closely associated with patient prognosis. Furthermore, these changes exhibit associations with genes that play pivotal roles in the core pathways of cancer cells. Our results have also established that the expression levels of *AKR1B10P1, RP11-465B22.3, WASH8P*, and *NPM1P25* possess the potential to accurately predict patient mortality rates, making them valuable prognostic biomarkers within the context of cancer. This study comprehensively explores lncRNA expression and their roles as oncogenes and tumor suppressors.

Author contributions The design and conceptualization of study and methodology was done by M.G. and M.M. Data mining, sample collection, formal analysis and investigation was performed by K.F.F., M.D.S., R.G. and R.S.M, Supervision, validation and visualization was done by K.F.F., M.D.S., and M.G., Interpretation of the obtained information was done by K.F.F., M.D.S., M.M. and M.G. The manuscript was written by K.F.F. and M.D.S. Review, editing and approved by M.G. and M.M. All authors read and approved the final manuscript.

Funding Not available.

Data availability All pertinent data backing the results of this study is contained in the manuscript and accompanying supplementary files. Any additional questions can be directed to the corresponding author.

Declarations

Ethics approval and consent to participate All experiments were conducted according to the pertinent guidelines and regulations of the Ministry of Health and Medical Education of Iran. The Biomedical Ethics Committee of the Islamic Azad University of Shahrekord (No.IR.IAU.SHK.REC.1402.072) approved all procedures.

Informed consent Informed consent was obtained from all individual participants included in the study.

Consent for publication All authors support the submission of the present manuscript to this journal.

Competing interests The authors declare no competing interests.

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2021. CA Cancer J Clin. 2021;71(1):7–33.
2. Yang H, Zhang X, Cai X, Wen D, Ye Z, Liang L, et al. From big data to diagnosis and prognosis: gene expression signatures in liver hepatocellular carcinoma. PeerJ. 2017;5:e3089.
3. Chen S, Zhu J, Wang F, Guan Z, Ge Y, Yang X, et al. LncRNAs and their role in cancer stem cells. Oncotarget. 2017;8(66):110685.
4. Huarte M. The emerging role of lncRNAs in cancer. Nat Med. 2015;21(11):1253–61.
5. Mahdevar M, Vatandoost J, Forootan FS, Kiani-Esfahani A, Esmaeili M, Peymani M, et al. Steroid receptor RNA activator gene footprint in the progression and drug resistance of colorectal cancer through oxidative phosphorylation pathway. Life Sci. 2021;285:119950.
6. Kong Y, Zhang L, Huang Y, He T, Zhang L, Zhao X, et al. Pseudogene PDIA3P1 promotes cell proliferation, migration and invasion, and suppresses apoptosis in hepatocellular carcinoma by regulating the p53 pathway. Cancer Lett. 2017;407:76–83.
7. Han K, Li C, Zhang X, Shang L. DUXAP10 inhibition attenuates the proliferation and metastasis of hepatocellular carcinoma cells by regulation of the Wnt/β-catenin and PI3K/Akt signaling pathways. Biosci Rep. 2019;39(5):BSR20181457.
8. Lee J, Hwang J-H, Chun H, Woo W, Oh S, Choi J, et al. PLEKHA8P1 promotes tumor progression and indicates poor prognosis of liver cancer. Int J Mol Sci. 2021;22(14):7614.
9. Zhang G, Sun J, Zhang X. A novel cuproptosis-related LncRNA signature to predict prognosis in hepatocellular carcinoma. Sci Rep. 2022;12(1):11325.
10. Sun Y, Zhang F, Wang L, Song X, Jing J, Zhang F, et al. A five lncRNA signature for prognosis prediction in hepatocellular carcinoma. Mol Med Rep. 2019;19(6):5237–50.
11. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. 2016;44(8):e71.
12. Law CW, Alhamdoosh M, Su S, Dong X, Tian L, Smyth GK, et al. RNA-seq analysis is easy as 1–2–3 with limma, Glimma and edgeR. F1000Research. 2016;5.
13. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209–49.
14. Xiao-Jie L, Ai-Mei G, Li-Juan J, Jiang X. Pseudogene in cancer: real functions and promising signature. J Med Genet. 2015;52(1):17–24.
15. Sisu C. Pseudogenes as biomarkers and therapeutic targets in human cancers. Pseudogenes: Pseudogenes: Functions andProtocols. 2021. p. 319–37.
16. Bao L, Li P, Zhao H, Chen L, Wang Y, Liang S, et al. Pseudogene PLGLA exerts anti-tumor effects on hepatocellular carcinoma through modulating miR-324–3p/GLYATL1 axis. Digest Liver Dis. 2021;54(7):918-26.
17. Ning Y-M, Lin K, Liu X-P, Ding Y, Jiang X, Zhang Z, et al. NAPSB as a predictive marker for prognosis and therapy associated with an immune-hot tumor microenvironment in hepatocellular carcinoma. BMC Gastroenterol. 2022;22(1):392.
18. Lin Y-H, Liu Y-C, Chen C-Y, Chi H-C, Wu M-H, Huang P-S, et al. LPAL2 suppresses tumor growth and metastasis of hepatocellular carcinoma by modulating MMP9 expression. Cells. 2022;11(16):2610.
19. Liu J, Chen L, Pan J, Chen M, Zhou J, Zhou F, et al. Comprehensive analysis of key lncRNAs in HCV-positive hepatocellular carcinoma. Arch Med Sci AMS. 2021;17(1):142.
20. Xu S, Dong K, Gao R, Yang Y, Zhou Y, Luo C, et al. Cuproptosis-related signature for clinical prognosis and immunotherapy sensitivity in hepatocellular carcinoma. J Cancer Res Clin Oncol. 2023;149(13):12249–63.

21. Reddy JK, Sambasiva RM. Lipid metabolism and liver inflammation. II. Fatty liver disease and fatty acid oxidation. Am J Physiol-Gastrointest Liver Physiol. 2006;290(5):G852–8.
22. Wang M, Han J, Xing H, Zhang H, Li Z, Liang L, et al. Dysregulated fatty acid metabolism in hepatocellular carcinoma. Hepat Oncol. 2016;3(4):241–51.
23. Hao F, Fei X, Ren X, Xi Xiao J, Chen Y, Wang J. Pseudogene AKR1B10P1 enhances tumorigenicity and regulates epithelial-mesenchymal transition in hepatocellular carcinoma via stabilizing SOX4. J Cellul Mol Med. 2020;24(20):11779–90.
24. Zhong C, Xie Z, Zeng L, Yuan C, Duan S. MIR4435–2HG is a potential pan-cancer biomarker for diagnosis and prognosis. Front Immunol. 2022;13:855078.
25. Deng X, Zhao X, Liang X, Chen R, Pan Y, Liang J. Linc00152 promotes cancer progression in hepatitis B virus-associated hepatocellular carcinoma. Biomed Pharmacother. 2017;90:100–8.
26. Li J, Wang X, Tang J, Jiang R, Zhang W, Ji J, et al. HULC and Linc00152 act as novel biomarkers in predicting diagnosis of hepatocellular carcinoma. Cellul Physiol Biochem. 2015;37(2):687–96.
27. Ouyang H, Zhang L, Xie Z, Ma S. Long noncoding RNA MAFG-AS1 promotes proliferation, migration and invasion of hepatocellular carcinoma cells through downregulation of miR-6852. Exp Therap Med. 2019;18(4):2547–53.
28. Chang Y, Zhang J, Zhou C, Qiu G, Wang G, Wang S, et al. Long non-coding RNA FOXD2-AS1 plays an oncogenic role in hepatocellular carcinoma by targeting miR-206. Oncol Rep. 2018;40(6):3625–34.
29. Zhang T, Cao C, Wu D, Liu L. SNHG3 correlates with malignant status and poor prognosis in hepatocellular carcinoma. Tumor Biol. 2016;37:2379–85.
30. Huang Y-L, Ning G, Chen L-B, Lian Y-F, Gu Y-R, Wang J-L, et al. Promising diagnostic and prognostic value of E2Fs in human hepatocellular carcinoma. Cancer Manag Res. 2019;11:1725.
31. Matter MS, Decaens T, Andersen JB, Thorgeirsson SS. Targeting the mTOR pathway in hepatocellular carcinoma: current state and future trends. J Hepatolgy. 2014;60(4):855–65.