Identification and Validation of lncRNA-Related Prognostic Signatures in Cholangiocarcinoma

JiaNing Zhang

Department of Hepatic Surgery III, Eastern Hepatobiliary Surgery Hospital, Naval Medical University, Shanghai, China

Fengwei Li

Department of Hepatic Surgery II, Eastern Hepatobiliary Surgery Hospital, Naval Medical University, Shanghai, China

Yihua Huang

Department of Pathology, School of Basic Medical Sciences, Fujian Medical University, Fuzhou, Fujian, China

Hui Xue

Department of Hepatic Surgery II, Eastern Hepatobiliary Surgery Hospital, Naval Medical University, Shanghai, China

Qifei Tao

Department of Hepatic Surgery III, Eastern Hepatobiliary Surgery Hospital, Naval Medical University, Shanghai, China

Guokun Zhuang

Department of Hepatic Surgery II, Eastern Hepatobiliary Surgery Hospital, Naval Medical University, Shanghai, China

Dongyang Ding

Department of Hepatic Surgery III, Eastern Hepatobiliary Surgery Hospital, Naval Medical University, Shanghai, China

Jianbing Xuan

Department of Hepatic Surgery II, Eastern Hepatobiliary Surgery Hospital, Naval Medical University, Shanghai, China

Kui Wang

Department of Hepatic Surgery II, Eastern Hepatobiliary Surgery Hospital, Naval Medical University, Shanghai, China

Weiping Zhou (✉ ehbhwp@126.com)

Eastern Hepatobiliary Surgery Hospital

Research

Keywords: Cholangiocarcinoma, long non-coding RNA, prognosis, MIR4435-2HG, GAPLINC

Posted Date: July 28th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-705273/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background and Aims: Cholangiocarcinoma (CCA), the second most common hepatobiliary cancer, is associated with poor prognosis. Therefore, there is a need to elucidate on the pathogenic mechanisms of CCA. In this study, we aimed at identifying lncRNA-related prognostic signatures for CCA through bioinformatics analysis and further validated their functions in CCA tumorigenesis and progression.

Methods: The RNA-seq data of CCA were downloaded from public databases. Differentially expressed lncRNAs (DElncRNAs) were screened using R packages. Then, candidate OS- and DFS-related DElncRNAs were selected through Kaplan–Meier survival analysis. Furthermore, LASSO regression analyses were performed to establish two prognostic signatures, termed the OS and DFS signatures, respectively. Multivariate COX models and nomograms for overall survival (OS) and disease-free survival (DFS) were established based on OS/DFS signature and clinical data. Hub lncRNAs were identified and enrichment analyses performed to explore their potential functions. Finally, *in vitro* and *in vivo* models were used to validate the effects of the hub lncRNAs in CCA tumorigenesis and progression.

Results: A total of 925 DElncRNAs were selected, from which six candidate OS-related lncRNAs and 15 candidate DFS-related lncRNAs were identified. The OS and DFS signatures were then established using four lncRNAs, respectively. We found that the OS signature and vascular invasion were significant independent risk factors for OS outcomes of CCA, while the DFS signature, vascular invasion and CA19-9 were significant independent risk factors for DFS outcomes of CCA. MiR4435-2HG and GAPLINC were selected as hub lncRNAs because they were included in both OS and DFS signatures. GO and KEGG enrichment analyses revealed that the two hub lncRNAs were involved in CCA tumorigenesis and progression. Finally, we constructed *in vitro* and *in vivo* models and revealed that the lncRNAs, MiR4435-2HG and GAPLINC can prompt CCA proliferation and migration *in vitro* and *in vivo*.

Conclusions: The established OS and DFS signatures, which were based on DElncRNAs, are independent risk factors for OS and DFS of CCA patients, respectively. MiR4435-2HG and GAPLINC were identified as hub lncRNAs. *In vitro* and *in vivo* models revealed that MiR4435-2HG and GAPLINC can prompt CCA progression, which might be novel prognostic biomarkers and therapeutic targets for CCA.

Background

Cholangiocarcinoma (CCA), a malignant tumor arising from biliary epithelial cells, is the second most common hepatobiliary malignancy, accounting for 10–20% of all primary liver cancers^[1]. Based on its anatomical location, it can be classified into intrahepatic (iCCA), perihilar (pCCA) and distal cholangiocarcinoma (dCCA)^[2]. Furthermore, cholangiocarcinoma usually presents with an insidious onset, therefore, most patients are diagnosed at an advanced stage and lose the opportunity for radical surgery^[3]. Therefore, given the poor prognosis, there is a need to elucidate on the molecular mechanisms of CCA.

Long non-coding RNAs (lncRNAs) are special RNA molecules without protein-coding potential and are longer than 200 nucleotides^[4-6]. It has been reported that they are important in tumorigenesis and progression^[4, 7-13]. For example, lncRNA MiR4435-2HG overexpression has been associated with poor prognosis in colorectal cancer (CRC)^[14]. Furthermore, MiR4435-2HG is overexpressed and markedly associated with ovarian cancer and clear cell renal cell carcinoma progression^[15, 16]. Besides, the expression of lncRNA GAPLINC has been shown to promote gastric cancer proliferation and is closely correlated with poor prognosis for gastric cancer patients^[17, 18]. Studies have determined that MiR4435-2HG and GAPLINC are involved in the initiation and progression of various tumors. However, their biological roles in cholangiocarcinoma have not been clearly established.

In this study, different lncRNA expression patterns were examined among appropriately selected HCC cases to identify candidate lncRNA biomarkers based on The Cancer Genome Atlas data (TCGA) and the Gene Expression Omnibus (GEO). The least absolute shrinkage and selection operator (LASSO) algorithm was used in determining key lncRNAs; thereafter, an CCA risk score system was also constructed and the lncRNA signature was validated. Finally, the roles of the target gene were validated *in vitro* and *in vivo*. Our work yielded a signature based on lncRNA expression that can accurately predict CCA prognosis through integrated analysis of genomic data.

Methods

Data collection and preprocessing

The collected CCA RNA-seq data and corresponding clinical data included the TCGA-CHOL dataset which was downloaded from The Cancer Genome Atlas (TCGA) database (<https://portal.gdc.cancer.gov/>) and the GSE107943 dataset which was downloaded from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). The TCGA-CHOL dataset included 36 tumor and 9 non-tumor tissue samples while the GSE107943 dataset included 30 tumor and 27 non-tumor tissue samples. Batch effects from the two databases were corrected by adjusting expression values using the ComBat function (sva R package^[19]). Furthermore, principal component analysis (PCA) was performed to visualize the RNA-seq data before and after correcting the batch effect. Relative clinical data were downloaded and combined as well, and missing values were imputed through mice R package (<https://CRAN.R-project.org/package=mice>). Tumor samples with OS < 1 month or without the corresponding clinical data were removed, 61 tumor and 36 non-tumor samples were included in the bioinformatics analysis.

Differentially expressed lncRNAs

DESeq2^[20] and edgeR^[21] R packages were used to determine differential expression levels of lncRNAs (DElncRNA) between tumor and non-tumor tissues in TCGA-CHOL or GSE107943 datasets. The DElncRNAs criteria was set as FDR < 0.05 and |Log₂fold change| > 1.3. Then, four DElncRNA lists, including GSE107943_DESeq2 list, GSE107943_edgeR list, TCGA-CHOL_DESeq2 list, and TCGA-

CHOL_edgeR list were obtained after our analyses. To ensure plausibility, DElncRNA within at least three DElncRNA lists were selected to be “real” DElncRNA.

Construction of lncRNA related OS/DFS signature

Based on the median TPM value, the obtained DElncRNAs were divided into two groups. The CCA DFS-related or OS-related lncRNAs were selected using univariate Cox regression analysis (survival R package: <https://CRAN.R-project.org/package=survival>). Kaplan–Meier analysis was used to calculate survival differences in DFS or OS between high- and low-expression groups. LASSO regression analysis was used to further compress the number of CCA DFS-related or OS-related lncRNAs, and the most significant prognostic lncRNAs were left to comprise OS or DFS signatures, which was calculated based on the final LASSO model as follows:

$$\text{OS/DFS signature} = \sum_{i=1}^n \text{Expi} \times \text{Coei}$$

Whereby; n represents the number of lncRNAs included in the OS or DFS signature, Expi represents the TPM value of each lncRNA and Coei represents the coefficient generated by LASSO regression. Furthermore, the Receiver Operating Characteristic (ROC) curve and the area under the ROC curve (AUC) were calculated to evaluate the predictive ability of the OS/DFS signature to CCA patients’ death or recurrence. Moreover, patients were allocated into high- or low-risk groups based on their OS/DFS signatures, and the most appropriate cut-off was determined using the survminer R package (<https://CRAN.R-project.org/package=survminer>). Kaplan–Meier analysis was used to calculate survival differences in DFS or OS between high-risk and low-risk groups. Finally, lncRNAs that were included in both OS or DFS signatures were selected as hub lncRNAs.

Construction of multivariate COX models and OS / DFS nomogram

To establish the prognostic values of the OS/DFS signatures, multivariate survival analysis was performed using the COX proportional hazards regression model based on OS/DFS signatures and other clinical characteristics. Then, the nomogram was established based on the result of multivariate survival analysis. Concordance index was used to validate model discrimination of the established nomogram.

GO and KEGG pathway enrichment analyses

To establish the biological functions of our hub lncRNAs, relevant mRNAs were selected through Pearson correlation ($|r| > 0.5$, $p < 0.05$). Furthermore, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed to investigate their functional and pathway annotation. $p \leq 0.05$ was considered statistically significant.

Cell culture and transfection

HCCC-9810 and RBE cells were purchased from Cell Bank of Chinese Academy of Sciences. The cells were cultured in 1640 medium supplemented with 10% fetal bovine serum (FBS) and incubated at 37°C in

a 5% CO₂ atmosphere. When cells grew to a density of 80%, they were seeded in 6-well plates. Plasmids containing Sh-RNA molecules directed against target lncRNA or scrambled Sh-RNA were transfected into cells using the Lipofectamine 3000 reagent (Invitrogen, Carlsbad, CA, USA), according to the manufacturer's instructions. For gene knockdown, the psi-LVRU6GP vectors containing shRNAs targeting MIR4435-2HG from GeneCopoeia™ (China, Shanghai) were used. Then, 50-nM vector with shRNA or negative control was transfected into cells. The two cell lines were characterized by Genetic Testing Biotechnology Corporation (Suzhou, China) using short tandem repeat (STR) markers.

Cell clone formation experiment

Logarithmically growing HCCC-9810 and RBE cells were seeded in 6-well plates, incubated, and the medium changed after every 24 h. Incubation was stopped after cell clones could be seen using the naked eye. Paraformaldehyde was added to fix the clones, after which they were dried in air after taking images.

Cell proliferation assay

The CCA cells (5000 cells per well) were seeded in 96 well plates overnight. At 0, 24, 48, 72, and 96 h post-transfection, 10 µL Cell Counting Kit-8 solution (Dojindo, Shanghai) was added into each well, followed by incubation for 2 h at 37 °C. Then, absorbance was determined at a wavelength of 450 nm to measure cell proliferation.

Cell invasion and migration assay

The transwell chamber (Millipore) with 50 µL of Matrigel (0.2 µg/µL) was used to perform the invasion assay. Cells were added to the upper chamber at a concentration of 3×10^4 /mL with serum-free medium, while about 500 µL of the medium containing 10% FBS was added to the lower chamber. Then, the transwell chamber was incubated at 37°C for 48 h after which cells were stained using crystal violet for 15 min and counted. The same procedures were repeated to evaluate cell migration ability, except that the trans-chamber was not coated with Matrigel.

Xenograft model in athymic mice

Five-week-old male BALB/c nu/nu mice were obtained from Shanghai Experimental Animal Center, Chinese Academy of Science. Mice were maintained under a pathogen-free condition and treated in accordance with the institutional animal welfare guidelines of the Chinese Academy of Science. For the assay to assess tumorigenicity, suspended HCCC-9810 cells (5×10^6) were subcutaneously injected into each nude mice following the AAALAC criterion. Tumor volumes (cm³) were recorded after every week from the 7th day. Finally, at the end of 4 weeks, the mice were sacrificed with 120 µl 10% chloral hydrate (Sinopharm Chemical Reagent Co., Ltd., Shanghai, China) at the experimental endpoint, and the tumours were harvested and weighed.

Statistical analysis

Normally distributed continuous data were presented as the mean $\pm$ standard deviation (SD) and compared using the student's t test. Non-normally distributed continuous data were presented as medians and interquartile range (IQR) and compared using the Mann-Whitney U test. Categorical variables were presented as percentages and compared using the χ^2 test. Statistical tests were performed using R (Version 3.6.3). $p \leq 0.05$ was considered statistically significant.

Results

DElncRNAs identification

Differentially expressed lncRNAs (DElncRNAs) between tumor and non-tumor tissues in the GSE107943 (**Table S1**) and TCGA-CHOL (**Table S2**) datasets were identified. The “edgeR” and “DESeq2” packages of R were used to identify DElncRNAs using a criteria of $FDR < 0.05$ and $|\text{Log}_2\text{fold change}| > 1.3$ (**Table S3**). Clinic-pathologic characteristics are presented in Tables 1 and **S4**. The abundance of male, Child-Plugh B or vascular invasion positive CCA patients in the GSE107943 dataset were significantly high than those in the TCGA-CHOL dataset ($p < 0.05$). There were no significant differences between the two datasets in other characteristics, including age, AJCC Stage, CA19-9, recurrence, and death ($p > 0.05$). Volcano plots and heatmaps of DElncRNAs are presented in Fig. 1A-D. Then, a venn diagram was used to reveal the overlapped DElncRNAs in the two datasets. To ensure reliability, DElncRNAs that appeared at least three times were considered real DElncRNAs. Finally, a total of 925 DElncRNAs were identified and used in the downstream analysis (Fig. 1E-F).

Table 1
Clinicopathologic features of patients in TCGA-CHOL and GSE107943 datasets.

Characteristic	TCGA-CHOL (n = 32)	GSE107943 (n = 29)	p value
Age	65(57,72)	65(59,74)	0.639
Gender			0.002
Female	19(59.38%)	6(20.69%)	
Male	13(40.62%)	23(79.31%)	
Child-Plugh Class			< 0.001
A	27(84.38%)	12(41.38%)	
B	5(15.62%)	17(58.62%)	
AJCC Stage			0.765
I	18(56.25%)	14(48.28%)	
II	8(25.00%)	6(48.28%)	
III	1(3.13%)	1(3.45%)	
IV	5(15.63%)	8(27.59%)	
Vascular invasion			0.010
Negative	28(87.5%)	17(58.62%)	
Positive	4(12.5%)	12(41.38%)	
CA19-9	36.6(26.5,136)	30.3(18.0,128.5)	0.563
Recurrence			0.189
No	14(43.75%)	8(27.59%)	
Yes	18(56.25%)	21(72.41%)	
Death			0.359
No	17(53.13%)	12(41.38%)	
Yes	15(46.87%)	17(58.62%)	

Construction of a DFS signature with four lncRNAs

The RNA-seq data from GSE107943 and TCGA-CHOL datasets were combined and the batch effect corrected using the sva package in R. PCA was performed to visualize sample distribution before and

after correcting the batch effect. As shown in **Figure S1 A**, there was a significant batch effect before correction, which was reduced after our correction (**Figure S1 B**).

To determine whether the 925 DEIncRNAs are associated with DFS of CCA patients, we performed univariate COX analysis and filtered out 15 DFS-related DEIncRNAs. Then, Kaplan-Meier analysis was performed to validate the influence of each IncRNA on CCA patients' DFS. **Figure S2** shows that all 15 DFS-related DEIncRNAs were associated with DFS of CCA patients ($p < 0.05$). The most significant DFS-related DEIncRNAs were selected through LASSO regression analysis (**Figure S1 C-D**), and 3-fold cross-validation performed. Four DEIncRNAs were identified, including HRAT92, PTGES2-AS1, GAPLINC and MIR4435-2HG. Then, the DFS signature was calculated as follows:

$$\text{DFS signature} = (-0.0001117420 \times \text{HRAT92}) + (0.0002521861 \times \text{PTGES2 - AS1}) + (0.0003073142 \times \text{GAPLINC}) + (0.0001889755 \times \text{MIR4435 - 2HG})$$

The DFS signatures for all patients were calculated and arranged in an ascending order (Fig. 2A). Then, the CCA patients were allocated into low-risk and high-risk groups using a cut-off value of 0.21579145. Besides, a high mortality rate and worse DFS were found in the high-risk group (Fig. 2B). Expression levels of the four IncRNAs are shown in Fig. 2C. As the DFS signature increased, expression levels of HRAT92 decreased, while those of GAPLINC, MIR4435-2HG, and PTGES2-AS1 increased. Furthermore, the ROC curve was used to evaluate the predictive value of the DFS signature for CCA recurrence (Fig. 2D). The DFS signature was found to have a strong predictive value for CCA recurrence (1-, 2-, 3-year recurrence AUC 0.805, 0.854, 0.822). Finally, Kaplan-Meier analysis revealed that the high-risk group had worse DFS outcomes than the low-risk group ($p < 0.0001$, Fig. 2E). These findings imply that the DFS signature, based on the four IncRNAs is a significant prognostic factor for DFS and has a strong predictive value for CCA recurrence in patients.

Construction of an OS signature with four IncRNAs

Univariate COX regression (**Figure S3**) and LASSO regression (**Figure S1E-F**) analyses were sequentially performed to identify the most significant OS-related DEIncRNAs. Four IncRNAs, including MIR4435-2HG, LOC100129215, AC123595.1, and GAPLINC were screened to establish the OS signature. The DFS signature was calculated as:

$$\text{OS signature} = (0.002887391 \times \text{LOC100129215}) + (0.000241777 \times \text{MIR4435-2HG}) + (0.00482334 \times \text{AC123595.1}) + (0.000893947 \times \text{GAPLINC})$$

All CCA patients were allocated into high- and low-risk groups based on the OS signature, using a cutoff value of 0.106651484. The OS signature and survival time for all CCA patients are in Fig. 3A-B. The four IncRNAs included in the OS signature were highly expressed in the high-risk group, implying that they were associated with worse OS outcomes for CCA patients (Fig. 3C). Similarly, the predictive value of the OS signature for CCA patients' death was also strong (1-, 2-, 3-year death AUC 0.827, 0.773, 0.838).

Finally, Kaplan–Meier analysis revealed that the high-risk group had worse OS outcomes than the low-risk group ($p = 0.0003$, Fig. 3E)

Establishment of nomogram for predicting CCA patients' DFS and OS outcomes

To elucidate on the prognostic value of the DFS/OS signature, univariate and multivariate COX proportional hazards regression analyses were sequentially performed (Fig. 4A, Fig. 5A). Multivariate COX regression analysis revealed that the DFS/OS signature and vascular invasion were independent risk factors for CCA patients' DFS and OS outcomes. In addition, CA19-9 was also found to be an independent risk factor for DFS. Then, based on the results of multivariate COX regressions analyses, we established nomograms to predict CCA patients' 1-, 2-, and 3-year recurrence-free or survival (Fig. 4B, Fig. 5B). The C-indices of our nomograms for DFS and OS prediction were 0.87 and 0.77, respectively, implying good predictive values (Fig. 4C, Fig. 5C). Therefore, the DFS and OS nomograms exhibited a high prognostic value.

Identification of prognostic hub lncRNAs and enrichment analysis

We established the DFS/OS signature using DElncRNAs. Since MIR4435-2HG and GAPLINC were both included in the DFS/OS signature, they were selected as our prognostic hub lncRNAs. Then, the relevant mRNAs were identified (**Table S5**). Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed to evaluate their potential biological functions. It was found that MIR4435-2HG was associated with focal adhesion, Hippo signaling pathway, PI3K-Akt signaling pathway, and TGF – beta signaling pathway among others, implying that MIR4435-2HG is highly involved in tumorigenesis and progression (**Figure S4A-B, Table S5, S6**). Moreover, GAPLINC was associated with histone acetylation and methylation, protein methylation, and chromatin binding among others, implying that GAPLINC may be associated with epigenetic modifications of chromosomes (**Figure S4C-D, Table S5, S6**). Our enrichment results revealed that MIR4435-2HG and GAPLINC are involved in various biological processes, which are highly correlated with tumorigenesis and progression.

The hub lncRNAs enhanced CCA proliferation, invasion and migration abilities *in vitro*

To elucidate on the effects of the two hub lncRNAs in CCA tumorigenesis and progression, we performed loss-of-function assays *in vitro*. Transwell assays were performed to identify their effects on CCA cell invasion and migration while the CCK8 assays and colony formation assays were used to verify their effects on CCA cell proliferation. We found that MIR4435-2HG or GAPLINC knockdown significantly inhibited cell proliferation, invasion and migration (Fig. 6A-D, Fig. 7A-D). These findings imply that MIR4435-2HG and GAPLINC are crucial genes in CCA progression.

The hub lncRNAs enhanced CCA proliferation abilities *in vivo*

To confirm our *in vitro* findings, *in vivo* assays were also performed. It was found that the average tumor volume in the Sh-MIR4435-2HG group was significantly smaller compared to that in the Sh-NC group (4-week Sh-MIR4435-2HG vs. Sh-NC: 0.050 ± 0.032 vs. 0.190 ± 0.008 , $p < 0.001$) (Fig. 6E). Similarly, the Sh-GAPLINC group also had a smaller average tumor volume than the Sh-NC group (4-week Sh-GAPLINC vs. Sh-NC: 0.040 ± 0.018 vs. 0.208 ± 0.033 , $p < 0.001$) (Fig. 7E). Besides, tumor weights of the Sh-MIR4435-2HG and Sh-GAPLINC groups were significantly smaller than those of the control group (4-week Sh-MIR4435-2HG vs. Sh-NC: 0.167 ± 0.029 vs. 0.422 ± 0.070 , $p < 0.05$; 4-week Sh-GAPLINC vs. Sh-NC: 0.170 ± 0.045 vs. 0.465 ± 0.050 , $p < 0.001$). These findings imply that inhibition of MIR4435-2HG or GAPLINC might suppress CCA proliferation *in vivo*.

Discussion

CCA, the second most common primary hepatobiliary cancer, originates from the bile duct epithelium^[22–25]. Due to a lack of a specific tumor marker, diagnosis of CCA at an early stage is difficult^[26–28]. In addition to its insidious onset, CCA is also characterized by high malignancy and early recurrence^[23–25, 28]. Studies have reported that lncRNAs are crucial regulators of diverse biological processes of mRNA and play essential roles in tumorigenesis^[29–32]. Therefore, in this study, we aimed at identifying prognostic-related DElncRNAs in CCA and validated their effects on CCA tumorigenesis and progression.

It has been documented that lncRNAs can be used as prognostic biomarkers for various malignancies^[33–35]. The occurrence and development of CCA is closely associated with some tumor-related lncRNAs, including PVT1^[36], H19^[37], NEAT1^[38] and HATOIR^[39]. These tumor-related lncRNAs have been shown to affect CCA proliferation, apoptosis, invasion, metastasis and EMT. However, identification and validation of hub lncRNAs associated with CCA prognosis, based on high-throughput transcriptome sequencing has not been performed.

In this study, we combined TCGA-CHOL and GSE107943 datasets and screened a total of 925 DElncRNAs. Then, the DElncRNA related DFS/OS signature was established. The established DFS/OS signature was highly correlated with CCA patients' DFS/OS. Combining clinical characteristics and the DFS/OS signature, multivariate Cox regression analysis was performed, and nomograms were established based on the results of multivariate Cox regression models. We found that the DFS/OS signature was an independent risk factor for CCA patients' DFS/OS. MIR4435-2HG and GAPLINC were selected as our hub lncRNAs, as they were included in both DFS and OS signatures.

MIR4435-2HG has been shown to play important roles in the progression of diverse malignancies^[14, 40–45]. It has been reported that MIR4435-2HG is overexpressed in non-small cell lung carcinoma (NSCLC), and its knockdown inhibited NSCLC cell proliferation and migration abilities^[42]. In lung cancer, MIR4435-2HG was found to be upregulated, and that it enhanced lung cancer cell proliferation and invasion^[41]. In colorectal cancer (CRC), MIR4435-2HG was associated with poor prognosis of CRC patients, and it promoted CRC growth and metastasis through the miR-206/YAP1 axis^[44]. Furthermore, Xuanlin Shen et

al. reported that MIR4435-2HG is overexpressed in hepatocellular carcinoma (HCC) tissues and is associated with the poor prognosis of HCC patients. Further studies determined that MIR4435-2HG can enhance HCC cell proliferation and invasion *in vitro*^[45]. We found that overexpression of MIR4435-2HG was associated with worse prognostic outcomes for CCA patients and that it may be involved in multiple processes of tumorigenesis. Furthermore, MIR4435-2HG was found to act as an oncogene in CCA, which enhances CCA proliferation, invasion and migration.

GAPLINC was initially reported in gastric cancer, and is therefore referred to as, gastric adenocarcinoma predictive long intergenic non-coding RNA (GAPLINC)^[18]. It has been reported that GAPLINC is a negative regulator of inflammation in mice^[46]. A study on human umbilical vein endothelial cells (HUVECs) revealed that GAPLINC can increase HUVECs angiogenesis under hypoxia conditions, implying that GAPLINC may play a critical role in tumor angiogenesis^[47]. It has also been reported that GAPLINC regulates rheumatoid arthritis fibroblast-like synoviocytes (RA-FLSs) proliferation, invasion and migration, and promotes RA-FLSs tumor-like behaviors^[48]. In gastric cancer, GAPLINC was shown to promote cell proliferation and migration by regulating CD44 expression^[18]. Pro-oncogenic effects of GAPLINC in other cancers, including CRC^[49], NSCLC^[50] and bladder cancer^[51] have also been reported. We identified GAPLINC as our hub lncRNA, and it was determined that GAPLINC is a risk factor for CCA and a predictor for poor prognosis. Furthermore, enrichment analysis revealed that GAPLINC may be involved in epigenetic modifications of chromosomes in CCA. Loss- and gain-function assays determined that GAPLINC promoted CCA proliferation and migration *in vitro* and *in vivo*.

Compared to existing studies on lncRNA-based prognostic signatures in other tumors^[4, 52, 53], our study has some strengths that should be noted. First, to ensure a sufficient sample size for our bioinformatics analysis, we combined two independent datasets. Second, necessary preprocessing was performed to remove batch effects to ensure reliability of our data as the systematic bias caused by the batch effect was devastating^[54, 55]. Third, essential experimental validations were indispensable. Loss- and gain-function assays *in vitro* and *in vivo* were performed to validate the potential functions of our hub lncRNAs. However, this study also has some limitations. First, although we identified hub lncRNAs and validated their effects in tumorigenesis and progression, the specific signaling pathways regulated by the hub lncRNAs were not elucidated, therefore, more experiments are needed for further exploration. In addition, our research was at the stage of experimental exploration, there is a long way to go before true clinical translation for cancer treatment.

Conclusions

The established OS and DFS signatures, which were based on DElncRNAs, are independent risk factors for OS and DFS of CCA patients, respectively. MIR4435-2HG and GAPLINC were identified as hub lncRNAs. *In vitro* and *in vivo* models revealed that MiR4435-2HG and GAPLINC can prompt CCA progression, which might be novel prognostic biomarkers and therapeutic targets for CCA.

Abbreviations

CCA

Cholangiocarcinoma

iCCA

Intrahepatic cholangiocarcinoma

pCCA

Perihilar cholangiocarcinoma

dCCA

Distal cholangiocarcinoma

lncRNAs

Long non-coding RNAs

DElncRNAs

Differentially expressed lncRNAs

GAPLINC

Gastric adenocarcinoma predictive long intergenic non-coding RNA

OS

Overall survival

DFS

Disease-free survival

CRC

Colorectal cancer

TCGA

The Cancer Genome Atlas data

GEO

Gene Expression Omnibus

LASSO

Least absolute shrinkage and selection operator

GO

Gene Ontology

FBS

Fetal bovine serum

SD

Standard deviation

IQR

Interquartile range

NSCLC

Non-small cell lung carcinoma

HCC

Hepatocellular carcinoma

HUVECs

Human umbilical vein endothelial cells

Declarations

Availability of data and materials

The datasets generated and/or analyzed during the current study are available on TCGA (<https://portal.gdc.cancer.gov/>) and the GSE107943 dataset which was downloaded from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). The clinico-pathological data for the 87 patients are available in Additional file 1 (Supplementary Table 1 and Table 2).

Acknowledgements

Not applicable.

Funding

This work was supported by the Clinical Research Plan of SHDC (No. SHDC2020CR2038B and SHDC2020CR007), MengChao talents cultivation program, collar talents climbing program (Kui Wang), Naval Medical University, "Strong sea" innovation team program (Kui Wang)

Author information

Affiliations

Department of Hepatic Surgery I, Eastern Hepatobiliary Surgery Hospital, Naval Medical University, Shanghai, China

Weiping Zhou, JiaNing Zhang, QifeiTao, Dongyang Ding

Department of Hepatic Surgery II, Eastern Hepatobiliary Surgery Hospital, Naval Medical University, Shanghai, China

Kui Wang, Fengwei Li, Hui Xue, Guokun Zhuang, Jianbing Xuan

Department of Pathology, School of Basic Medical Sciences, Fujian Medical University, Fuzhou, Fujian, China.

Yihua Huang

Contributions

FWL and JNZ performed research, analyzed data, wrote manuscript, performed research, YHH wrote manuscript, QT, DD, HX, GZ, and JBX performed research, WPZ, KW, designed research, analyzed data,

edited manuscript. All authors read and approved the final manuscript.

Corresponding author

Correspondence to Weiping Zhou (the main corresponding author), or Kui Wang.

Ethics declarations

Ethics approval and consent to participate

The breeding of the mice was done in agreement with governmental laws.

Consent for publication

Not applicable.

Competing interests

Not applicable.

References

1. Jusakul A, Cutcutache I, Yong CH, et al. Whole-Genome and Epigenomic Landscapes of Etiologically Distinct Subtypes of Cholangiocarcinoma[J]. *Cancer Discov.* 2017;7(10):1116–35.
2. Hezel A, Deshpande V, Zhu A. Genetics of biliary tract cancers and emerging targeted therapies[J]. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology.* 2010;28(21):3531–40.
3. Nagino M. Surgical Treatment of Perihilar Cholangiocarcinoma: Resection or Transplant?[J]. *Ann Surg.* 2018;267(5):806–7.
4. Lin JZ, Lin N, Zhao WJ. Identification and validation of a six-lncRNA prognostic signature with its ceRNA networks and candidate drugs in lower-grade gliomas[J]. *Genomics.* 2020;112(5):2990–3002.
5. Cabili M, Trapnell C, Goff L, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses[J]. *Genes Dev.* 2011;25(18):1915–27.
6. Statello L, Guo C, Chen L, et al. Gene regulation by long non-coding RNAs and its biological functions[J]. *Nat Rev Mol Cell Biol.* 2021;22(2):96–118.
7. Li F, Bai L, Li S, et al. Construction and evaluation of a prognosis lncRNA model for hepatocellular carcinoma[J]. *J Cell Biochem.* 2020. 10.1002/jcb.29608.
8. Wang Y, Li L, Zhao Z, et al. L1 and epithelial cell adhesion molecules associated with gastric cancer progression and prognosis in examination of specimens from 601 patients[J]. *32: Journal of experimental & clinical cancer research: CR;* 2013. p. 66.
9. Amelio I, Bernassola F, Candi E. Emerging roles of long non-coding RNAs in breast cancer biology and management[J]. *Sem Cancer Biol.* 2020. 10.1016/j.semcancer.2020.06.019.

10. Su M, Xiao Y, Ma J, et al. Long non-coding RNAs in esophageal cancer: molecular mechanisms, functions, and potential applications[J]. *J Hematol Oncol*. 2018;11(1):118.
11. Qi P, Zhou X, Du X. Circulating long non-coding RNAs in cancer: current status and future perspectives[J]. *Mol Cancer*. 2016;15(1):39.
12. Parasramka M, Maji S, Matsuda A, et al. Long non-coding RNAs as novel targets for therapy in hepatocellular carcinoma[J]. *Pharmacol Ther*. 2016;161:67–78.
13. Jiang C, Li X, Zhao H, et al. Long non-coding RNAs: potential new biomarkers for predicting tumor invasion and metastasis[J]. *Mol Cancer*. 2016;15(1):62.
14. Ouyang W, Ren L, Liu G, et al. LncRNA MIR4435-2HG predicts poor prognosis in patients with colorectal cancer[J]. *PeerJ*. 2019;7:e6683.
15. Zhu L, Wang A, Gao M, et al. LncRNA MIR4435-2HG triggers ovarian cancer progression by regulating miR-128-3p/CKD14 axis[J]. *Cancer cell international*. 2020;20:145.
16. Zhu K, Miao C, Tian Y, et al. lncRNA MIR4435-2HG promoted clear cell renal cell carcinoma malignant progression via miR-513a-5p/KLF6 axis[J]. *J Cell Mol Med*. 2020;24(17):10013–26.
17. Diao L, Wang S, Sun Z. Long noncoding RNA GAPLINC promotes gastric cancer cell proliferation by acting as a molecular sponge of miR-378 to modulate MAPK1 expression[J]. *OncoTargets therapy*. 2018;11:2797–804.
18. Hu Y, Wang J, Qian J, et al. Long noncoding RNA GAPLINC regulates CD44-dependent cell invasiveness and associates with poor prognosis of gastric cancer[J]. *Cancer research*. 2014;74(23):6890–902.
19. Gagnon-Bartsch JA, Speed TP. Using control genes to correct for unwanted variation in microarray data[J]. *Biostatistics*. 2012;13(3):539–52.
20. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2[J]. *Genome Biol*. 2014;15(12):550.
21. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data[J]. *Bioinformatics*. 2010;26(1):139–40.
22. Li S, Qian H, Peng Y, et al. Differentiating peripheral cholangiocarcinoma in stages T1N0M0 and T2N0M0 from hepatic hypovascular nodules using dynamic contrast-enhanced MRI[J]. *Scientific reports*. 2017;7(1):8084.
23. Rizvi S, Gores G. Emerging molecular therapeutic targets for cholangiocarcinoma[J]. *Journal of hepatology*. 2017;67(3):632–44.
24. Banales J, Marin J, Lamarca A, et al. Cholangiocarcinoma 2020: the next horizon in mechanisms and management[J]. *Nat Rev Gastroenterol Hepatol*. 2020;17(9):557–88.
25. Patel T. Cholangiocarcinoma—controversies and challenges[J]. *Nat Rev Gastroenterol Hepatol*. 2011;8(4):189–200.
26. Zhou X, Wang J, Tang M, et al. Hepatocellular carcinoma with hilar bile duct tumor thrombus versus hilar Cholangiocarcinoma on enhanced computed tomography: a diagnostic challenge[J]. *BMC*

- Cancer. 2020;20(1):54.
27. Su Z, Liu G, Fang T, et al. Expression and prognostic value of glutamate dehydrogenase in extrahepatic cholangiocarcinoma patients[J]. American journal of translational research. 2017;9(5):2106–18.
 28. Razumilava N, Gores G. Cholangiocarcinoma[J]. Lancet. 2014;383(9935):2168–79.
 29. Merdrignac A, Papoutsoglou P, Coulouarn C. Long Noncoding RNAs in Cholangiocarcinoma[J]. Hepatology. 2021;73(3):1213–26.
 30. Yang Y, Deng X, Li Q, et al. Emerging roles of long noncoding RNAs in cholangiocarcinoma: Advances and challenges[J]. Cancer communications (London England). 2020;40(12):655–80.
 31. Li J, Huang L, Li Z, et al. Functions and roles of long noncoding RNA in cholangiocarcinoma[J]. Journal of cellular physiology. 2019;234(10):17113–26.
 32. Salati M, Braconi C. Noncoding. RNA in Cholangiocarcinoma[J]. Semin Liver Dis. 2019;39(1):13–25.
 33. McCabe E, Rasmussen T. lncRNA involvement in cancer stem cell function and epithelial-mesenchymal transitions[J]. Sem Cancer Biol. 2020. 10.1016/j.semcancer.2020.12.012.
 34. Loewen G, Jayawickramarajah J, Zhuo Y, et al. Functions of lncRNA HOTAIR in lung cancer[J]. J Hematol Oncol. 2014;7:90.
 35. Li X, Jin F, Li Y. A novel autophagy-related lncRNA prognostic risk model for breast cancer[J]. J Cell Mol Med. 2021;25(1):4–14.
 36. Yu Y, Zhang M, Liu J, et al. Long Non-coding RNA PVT1 Promotes Cell Proliferation and Migration by Silencing ANGPTL4 Expression in Cholangiocarcinoma[J]. Molecular therapy. Nucleic acids. 2018;13:503–13.
 37. Xu Y, Wang Z, Jiang X, et al. Overexpression of long noncoding RNA H19 indicates a poor prognosis for cholangiocarcinoma and promotes cell migration and invasion by affecting epithelial-mesenchymal transition[J]. 92: Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie; 2017. pp. 17–23.
 38. Zhang C, Li J, Tian F, et al. Long Noncoding RNA NEAT1 Promotes Growth and Metastasis of Cholangiocarcinoma Cells[J]. Oncology research. 2018;26(6):879–88.
 39. Qin W, Kang P, Xu Y, et al. Long non-coding RNA HOTAIR promotes tumorigenesis and forecasts a poor prognosis in cholangiocarcinoma[J]. Scientific reports. 2018;8(1):12176.
 40. Zhu L, Wang A, Gao M, et al. lncRNA MIR4435-2HG triggers ovarian cancer progression by regulating miR-128-3p/CKD14 axis[J]. Cancer Cell Int. 2020;20:145.
 41. Qian H, Chen L, Huang J, et al. The lncRNA MIR4435-2HG promotes lung cancer progression by activating beta-catenin signalling[J]. J Mol Med (Berl). 2018;96(8):753–64.
 42. Yang M, He X, Huang X, et al. lncRNA MIR4435-2HG-mediated upregulation of TGF- β 1 promotes migration and proliferation of nonsmall cell lung cancer cells[J]. Environmental toxicology. 2020;35(5):582–90.

43. Zhang H, Meng H, Huang X, et al. lncRNA MIR4435-2HG promotes cancer cell migration and invasion in prostate carcinoma by upregulating TGF- β 1[J]. *Oncology letters*. 2019;18(4):4016–21.
44. Dong X, Yang Z, Yang H, et al. Long Non-coding RNA MIR4435-2HG Promotes Colorectal Cancer Proliferation and Metastasis Through miR-206/YAP1 Axis[J]. *Front Oncol*. 2020;10:160.
45. Shen X, Ding Y, Lu F, et al. Long noncoding RNA MIR4435-2HG promotes hepatocellular carcinoma proliferation and metastasis through the miR-22-3p/YWHAZ axis[J]. *Am J Transl Res*. 2020;12(10):6381–94.
46. Vollmers AC, Covarrubias S, Kuang D, et al. A conserved long noncoding RNA, GAPLINC, modulates the immune response during endotoxic shock[J]. *Proc Natl Acad Sci U S A*, 2021, 118(7).
47. He Y, Wu Z, Qiu C, et al. Long non-coding RNA GAPLINC promotes angiogenesis by regulating miR-211 under hypoxia in human umbilical vein endothelial cells[J]. *J Cell Mol Med*. 2019;23(12):8090–100.
48. Mo BY, Guo XH, Yang MR, et al. Long Non-Coding RNA GAPLINC Promotes Tumor-Like Biologic Behaviors of Fibroblast-Like Synoviocytes as MicroRNA Sponging in Rheumatoid Arthritis Patients[J]. *Front Immunol*. 2018;9:702.
49. Luo Y, Ouyang J, Zhou D, et al. Long Noncoding RNA GAPLINC Promotes Cells Migration and Invasion in Colorectal Cancer Cell by Regulating miR-34a/c-MET Signal Pathway[J]. *Digestive diseases sciences*. 2018;63(4):890–9.
50. Zhao J, Wang C, Liu S, et al. TGF- β 1 mediates lncRNA GAPLINC expression to promote the migration and invasion of non-small cell lung cancer[J]. *OncoTargets therapy*. 2019;12:6175–80.
51. Zheng Z, Zhu D, Zhao F, et al. Upregulated GAPLINC predicts a poor prognosis in bladder cancer patients and promotes tumor proliferation and invasion[J]. *Oncology letters*. 2018;15(5):6770–6.
52. Guo Q, Wang J, Gao Y, et al. Dynamic TF-lncRNA Regulatory Networks Revealed Prognostic Signatures in the Development of Ovarian Cancer[J]. *Front Bioeng Biotechnol*. 2020;8:460.
53. Li W, Chen QF, Huang T, et al. Identification and Validation of a Prognostic lncRNA Signature for Hepatocellular Carcinoma[J]. *Front Oncol*. 2020;10:780.
54. Leek JT. svaseq: removing batch effects and other unwanted noise from sequencing data[J]. *Nucleic Acids Res*, 2014, 42(21).
55. Leigh DM, Lischer HEL, Grossen C, et al. Batch effects in a multiyear sequencing study: False biological trends due to changes in read lengths[J]. *Mol Ecol Resour*. 2018;18(4):778–88.

Figures

Figure 1

DElncRNAs identification in TCGA-CHOL and GSE107943 datasets. (A) Volcano plot for DElncRNAs in TCGA-CHOL datasets. Darkred and darkblue points represent high and low expression lncRNAs in CCA tumor tissues. Gray points imply no significant differences ($\text{FDR} < 0.05$ and $|\log_2(\text{fold change})| > 1.3$). (B) Heatmap of DElncRNAs in TCGA-CHOL datasets. TPM values of DElncRNAs were determined. Green and red represent normal and tumor tissues, respectively. (C) Volcano plot of DElncRNAs in GSE107943

datasets. (D) Heatmap of DElncRNAs in GSE107943 datasets. (E) Venn diagram showing overlapped DElncRNAs in the two datasets through different analytic methods. DElncRNA that appeared at least three times were considered real DElncRNA. (F) Histogram of DElncRNAs. The x-axis represents the frequency identified as DElncRNA in the two datasets by two different analytic methods, while the y-axis represents the number of DElncRNAs. Red represents real DElncRNAs used in the following analysis.

Figure 2

Establishment and validation of DFS signatures based on DElncRNAs. (A) Distributions of the DFS signature in CCA patients. Patients were allocated into high and low risk score groups using a cutoff value of 0.21579145. The x-axis represents CCA patients in an ascending order, while the y-axis represents the score of the DFS signature. (B) Dot plot of CCA patients' survival status. The x-axis represents CCA patients, consistent with (A), while the y-axis represents survival time (months). Red and blue dots represent death and alive, respectively. (C) Heatmap of the lncRNAs included in the DFS signature. (D) Receiver operating characteristic (ROC) curves for DFS signature predicting CCA recurrence in patients. The AUCs of 1-year, 2-year and 3-year recurrence were 0.805, 0.854 and 0.822, respectively. (E) Kaplan-Meier survival analysis of the low and high risk group. The DFS of CCA patients in the high risk group was significantly worse than that of patients in the low risk group ($p < 0.0001$).

Figure 3

Establishment and validation of the OS signature based on DElncRNAs. (A) Distributions of the OS signature in CCA patients. Patients were divided into high and low risk score groups, using a cutoff value of 0.106651484. The x-axis represents CCA patients in an ascending order, while the y-axis represents the OS signature score. (B) Dot plot of CCA patients' survival status. The x-axis represents CCA patients, consistent with (A), while the y-axis represents the survival time (months). Red and blue dots represent death and alive, respectively. (C) Heatmap of lncRNAs included in the OS signature. (D) Receiver operating characteristic (ROC) curves for OS signature predicting CCA patients' death. The AUCs of 1-year, 2-year and 3-year recurrence were 0.827, 0.773 and 0.838, respectively. (E) Kaplan-Meier survival analysis of the low and high risk group. The OS outcomes for CCA patients in the high risk group were significantly worse than those of the low risk group ($p = 0.0003$).

Figure 4

Cox hazard regression analysis and nomogram for CCA patients' DFS. (A) Univariate and multivariate Cox hazard regression analyses for CCA patients' DFS. DFS signature, vascular invasion and CA19-9 were independent risk factors for CCA patients' DFS. (B) Nomogram for prediction of CCA patients' DFS, which was established based on multivariate Cox hazard regression analysis. (C) Calibration curve of the nomogram.

Figure 5

Cox hazard regression analysis and nomogram of CCA patients' OS. (A) Univariate and multivariate Cox hazard regression analysis of CCA patients' OS. OS signature, and vascular invasion were independent risk factors for CCA patients' OS. (B) Nomogram for the prediction of CCA patients' OS, which was established based on multivariate Cox hazard regression analysis. (C) Calibration curve of the nomogram.

Figure 6

Functional validation of MIR4435-2HG in CCA tumorigenesis and progression in vitro and in vivo. Transwell assays revealed that invasion and migration abilities of HCCC-9810 (A) and RBE (B) cell lines were significantly suppressed in the Sh-MIR4435-2HG group. CCK-8 and colony formation assays revealed that proliferative abilities of HCCC-9810 (C) and RBE (D) cell lines were significantly suppressed in the Sh-MIR4435-2HG group. (E) Xenograft model of nude mice revealed that the growth of subcutaneous tumors was significantly suppressed in the Sh-MIR4435-2HG group. (The experiment was repeated at least three times, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$)

Figure 7

Functional validation of GAPLINC in CCA tumorigenesis and progression in vitro and in vivo. Transwell assays revealed that invasion and migration abilities of HCCC-9810 (A) and RBE (B) cell lines were significantly suppressed in the Sh-GAPLINC group. The CCK-8 and colony formation assays revealed that proliferative abilities of HCCC-9810 (C) and RBE (D) cell lines were significantly suppressed in the Sh-GAPLINC group. (E) Xenograft models in nude mice revealed that the growth of subcutaneous tumor was significantly suppressed in the Sh-GAPLINC group. (The experiment was repeated at least three times, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigureS1PCA.jpg](#)
- [FigureS2allDFS.jpg](#)
- [FigureS3allOS.jpg](#)
- [FigureS4.jpg](#)
- [TableS1GSE107943rawdata.xlsx](#)
- [TableS2TCGACHOLrawdata.xlsx](#)
- [TableS3DEGs.xlsx](#)
- [TableS4clinicalinformation.xlsx](#)
- [TableS5CorrelationofmRNA.xlsx](#)
- [TableS6Enrichmentresult.xlsx](#)