

A novel five-lncRNA signature panel improves high-risk survival prediction in patients with cholangiocarcinoma

Xiaozai Xie^{1,2,*}, Yi Wang^{3,*}, Sina Zhang^{3,4,*}, Jialiang Li^{1,2}, Zhengping Yu¹, Xiwei Ding⁵, Longyun Ye¹, Peirong Gong⁶, Qiandong Zhu¹, Junjian Li¹, Ziyang Chen⁷, Xinfei Yao⁷, Zhiyong Du⁸, Qiqiang Zeng^{1,2}, Hanbin Chen⁹, Zhen Yang¹⁰, Gang Chen^{1,2}

¹Department of Hepatobiliary Surgery, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China

²Key Laboratory of Diagnosis and Treatment of Severe Hepato-Pancreatic Diseases of Zhejiang Province, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China

³Department of Epidemiology and Biostatistics, School of Public Health and Management, Wenzhou Medical University, Wenzhou, China

⁴School of Public Health, Inner Mongolia Medical University, Hohhot, China

⁵Department of Gastroenterology, The Affiliated Drum Tower Hospital of Nanjing University Medical School, Nanjing, China

⁶Department of Hepatobiliary Surgery, The Affiliated Cixi Hospital of Wenzhou Medical University, Ningbo, China

⁷Division of Clinical Medicine, First School of Clinical Medicine, Wenzhou Medical University, Wenzhou, China

⁸Department of Hepatobiliary Surgery, Shenzhen People's Hospital, Shenzhen, China

⁹Department of Radiotherapy and Chemotherapy, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China

¹⁰Department of Infectious Diseases, Shandong Provincial Hospital, Jinan, China

*Equal contribution

Correspondence to: Gang Chen, Zhen Yang; **email:** chen.gang@wmu.edu.cn, jingyiyong@sdu.edu.cn

Keywords: cholangiocarcinoma, lncRNA, overall survival, prognosis prediction, risk-stratification

Received: June 19, 2020

Accepted: November 23, 2020

Published: January 20, 2021

Copyright: © 2021 Xie et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Cholangiocarcinoma (CCA) is a fatal disease with dismal survival rates. Long non-coding RNA (lncRNA) expression profiling as potential prognostic biomarkers play critical roles in tumor initiation, development, and poor prognosis. Identifying specific lncRNA to predict the prognosis of CCA patients in the early stages is very important for improving a patient's survival. In the current study, we aimed to establish a novel risk-stratification lncRNA signature panel in CCA. The initial lncRNA discovery was identified in The Cancer Genome Atlas database (TCGA cohort). The Cox regression analysis was used to establish the lncRNA prognostic model and the receiver operating characteristic (ROC) curve analysis was performed to assess the specificity and sensitivity of the model. This was followed by independent validation of the lncRNA signature in the CCA patients from the First Affiliated Hospital of Wenzhou Medical University (WMU cohort). Furthermore, by using the Gene Ontology function and Kyoto Encyclopedia Gene and Genome pathway enrichment analysis, we explored the potential function of prognosis lncRNA. Finally, five lncRNA (HULC; AL359715.5; AC006504.8; AC090114.2; AP00943.4) were screened to establish the predictive model that significantly associated with poor overall survival (HR:4.879;95%CI,1.587-14.996;p=0.006). This five-lncRNA signature model showed excellent accuracy in the TCGA cohort (AUC=0.938), and also robustly predicted survival in the validation WMU cohort (AUC=0.816). Functional enrichment analysis suggested prognostic lncRNA was primarily associated with CCA-related biological processes. Our data established a novel lncRNA signature model for CCA risk-stratification and robust identification of CCA patients with poor molecular genotypes. Moreover, it revealed new molecular mechanisms of CCA.

INTRODUCTION

Cholangiocarcinoma (CCA) is the second most common primary liver cancer after hepatocellular carcinoma, with incidence and mortality rates rising across the world [1, 2]. Despite surgery and liver transplantation being options for patients, the high recurrence rate leads to CCA patients' median survival time of less than one year [3]. Moreover, whether adjuvant therapy after surgical resection is effective, because data about its overall efficacy and survival benefits are limited [4]. Clinicopathological factors of CCA such as grade and stage are strongly associated with prognosis and are also key factors determining the therapeutic regimen. However, even with similar clinical characteristics, the prognosis of CCA patients is significantly different. Therefore, it is important to identify efficient tumor features to help clinicians stratify high-risk patients and tailor personalized treatment regimens for improving treatment outcomes.

With the development of gene sequencing technology, there has been a growing interest in using gene expression signature for risk-stratification of cancer patients. Besides, anti-cancer drugs based on genetic research are developing rapidly [5–7]. Therefore, conducting further studies on CCA-related genes and epigenetic markers such as long noncoding RNA (lncRNA) to guide personalized treatment in order to reduce recurrence and improve survival rate is warranted [8]. In the past decades, lots of evidence has suggested that lncRNA is strongly related to tumor occurrence, metastasis, and prognosis [9–11]. For CCA, studies have confirmed that lncRNA plays a key role in CCA occurrence and progression [12]. For instance, MALAT1 promote CCA cell proliferation and invasion [13], UCA1 affect migration and invasion potential of CCA cells by regulating EMT [14]. Besides, lncRNA such as TUG1 [15], CCAT1 [16], and AFAP1-AS1 [17] could serve as valuable predictive markers for CCA patients prognosis. However, the role and mechanism of lncRNA in the metastasis and recurrence of tumors even in CCA is not completely understood.

In this study, we collected lncRNA expression data and clinical information of CCA patients from two independent database sources to identify and develop a novel lncRNA-based signature panel as an independent predictor, for the prognosis of CCA patients, to guide personalized treatment and hence improve survival. This was achieved by using simple, inexpensive quantitative PCR assays that can be incorporated into the clinical approach. Furthermore, we investigate the possible molecular mechanisms related to this prognostic lncRNA with the occurrence and progress of CCA. We believe this lncRNA-base signature panel

offers an effective platform for risk-stratification in CCA patients, which has great implications in the clinical management of patients suffering from this fatal malignancy.

RESULTS

Establishment of a five-lncRNA signature predictive model from the TCGA cohort

Based on the screening criteria, we obtained 1192 differentially expressed lncRNA, including 744 up-regulated and 448 down-regulated. Among them, 33 lncRNA showed >4-fold decreased expression including HULC, and 51 lncRNA exhibited >4 fold increased expression (Figure 1A). Unsupervised hierarchical cluster analysis showed that the expression of differentially expressed lncRNA distinguished CCA samples from the normal samples (Figure 1B).

Univariate Cox regression analysis assessing the association between 1192 differentially expressed lncRNA and CCA patients' overall survival in TCGA cohort, 10 lncRNA (LINC01336; AP000943.4; AC006504.8; AC090114.2; AC004921.1; AC134682.1; AL449106.1; AL359715.5; AC016876.1 and HULC) were selected for further multivariate Cox regression analysis. In this multivariate Cox regression analysis, five lncRNA (AC006504.8, AC090114.2, AP000943.4, AL359715.5, and HULC) were identified as independent predictors for CCA survival outcomes (Table 1). Among them, the expression of AC006504.8, AC090114.2, and AP000943.4 was up-regulated (Figure 1C–1E); expression of AL359715.5 and HULC was down-regulated (Figure 1F, 1G). Similarly, these 5 lncRNAs also showed similar behaviors in paired carcinoma and paracancerous tissues (Supplementary Figure 1). Based on these five lncRNA expression levels and their corresponding coefficient in multivariate Cox regression analysis, a five-lncRNA signature predictive model for CCA patients prognosis was established as follows: Risk Score = $(0.6542 \times \text{expression level of HULC}) + (-1.2388 \times \text{expression level of AL359715.5}) + (1.3769 \times \text{expression level of AC006504.8}) + (-3.6697 \times \text{expression level of AC090114.2}) + (-1.5165 \times \text{expression level of AP000943.4})$.

Performance evaluation of the five-lncRNA signature model for CCA prognosis in the discovery cohort

We calculated the risk score for each CCA patient in the TCGA cohort based on the five-lncRNA signature model. According to the median of $\log(\text{Risk Score})$, -0.0995 was set as the cutoff level to stratify CCA patients into a high-risk or low-risk group (sensitivity is 82.01%, specificity is 86.02%). Finally, 18 patients

whose log (Risk Score) >0.0995 were assigned to the high-risk group, and 18 patients whose log(Risk Score) <-0.0995 were assigned to the low-risk group (Figure 2A–2C). The KM curves show that patients in the high-risk group had a worse prognosis than those in the low-risk group (median OS is 18.5 months vs. 60 months; log-rank $p=0.002$) (Figure 2D). The univariate Cox regression analysis between the lncRNA-based risk

score and CCA patient survival score showed that high-risk patients have a worse prognosis (HR=6.760; 95%CI, 1.572-29.068; $p=0.008$). The multivariate Cox regression analysis revealed that high-risk score is an independent predictor for CCA patient survival after adjusting the clinical covariate including recurrence status and residual tumor status (HR=4.879; 95%CI, 1.587-14.996; $p=0.006$) (Table 2).

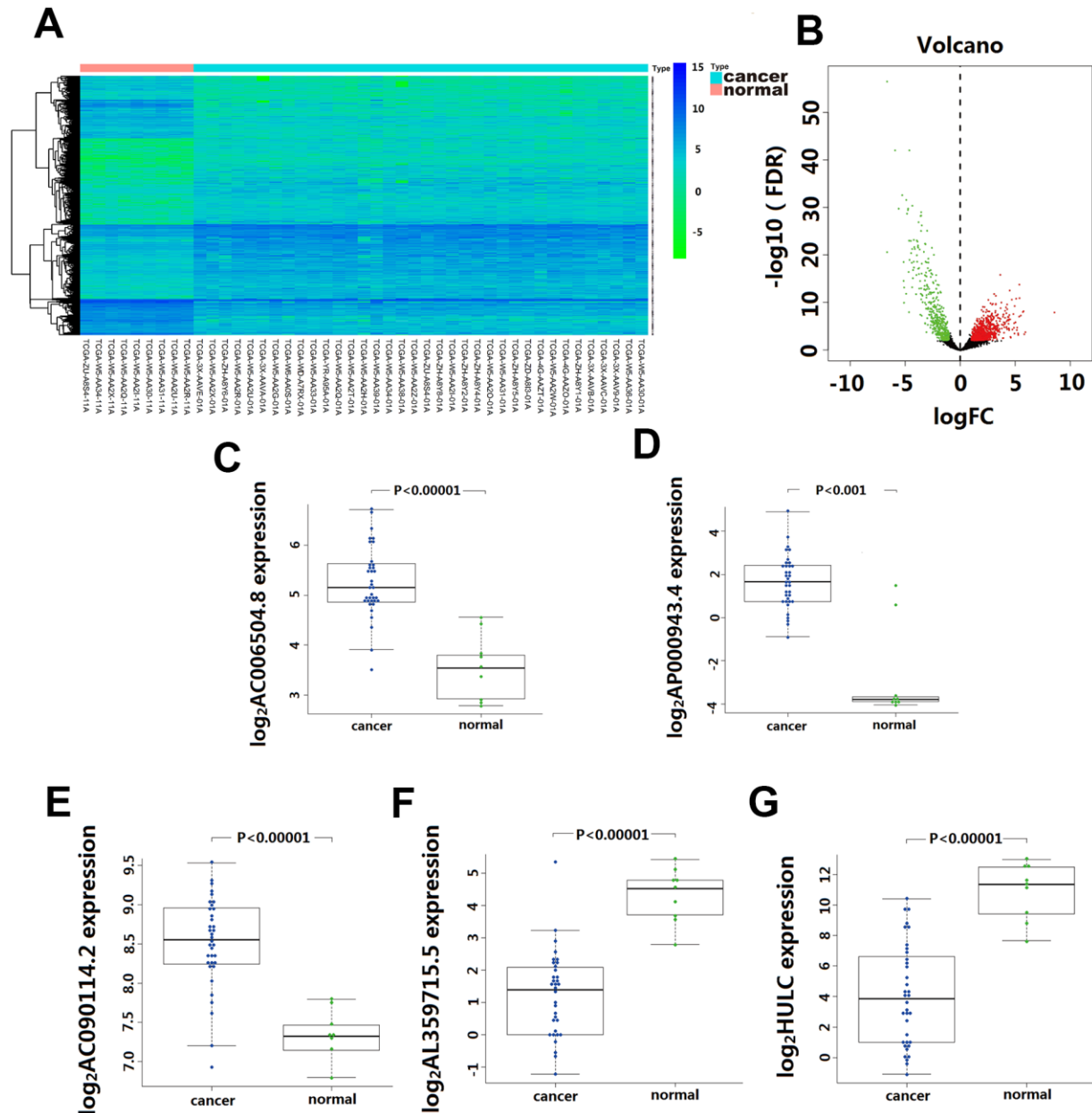


Figure 1. Differentially expressed lncRNA (DElncRNA) between Cholangiocarcinoma and normal tissues. (A) 1192 DElncRNA were detected based on an unsupervised hierarchical clustering heatmap. 744 DElncRNA expression increased, and 448 DElncRNA expression decreased. The indication from green to red represents the expression level from low to high. (B) Volcano plot depicts 1192 DElncRNA (green and red dots) between Cholangiocarcinoma and normal tissues. The X-axis represents log₁₀ fold change; Y-axis represents -log₁₀ of the p -value for each lncRNA. lncRNA with fold change >1 or <-1 and FDR < 0.01 were considered to be differentially expressed between tumor and normal tissues. (C–G) The differentially expressed level of five-lncRNA in 36 CCA tissue and 9 para-carcinoma normal tissue.

Table 1. Five lncRNA significantly correlated with the overall survival of cholangiocarcinoma.

Gene name	Ensemble ID	Chromosome	HR*	Coefficient*	P-value*
HULC	ENSG00000251164.1	chr6:8652137- 8653846	1.9236	3.42	0.00062
AC006504.8	ENSG00000281468.1	chr19:27802838- 27803472	3.9626	2.85	0.00443
AC090114.2	ENSG00000273270.1	chr7:128524016- 128531069	0.0255	-3.28	0.00104
AL359715.5	ENSG00000279022.1	chr6:80440730- 80441172	0.2897	-2.47	0.01366
AP000943.4	ENSG00000280167.1	chr11:94559018- 94559374	0.2195	-3.01	0.00264

Statistics derived from univariate Cox proportional hazards regression analysis in 36 patients in TCGA cohort; HR, Hazard ratio.

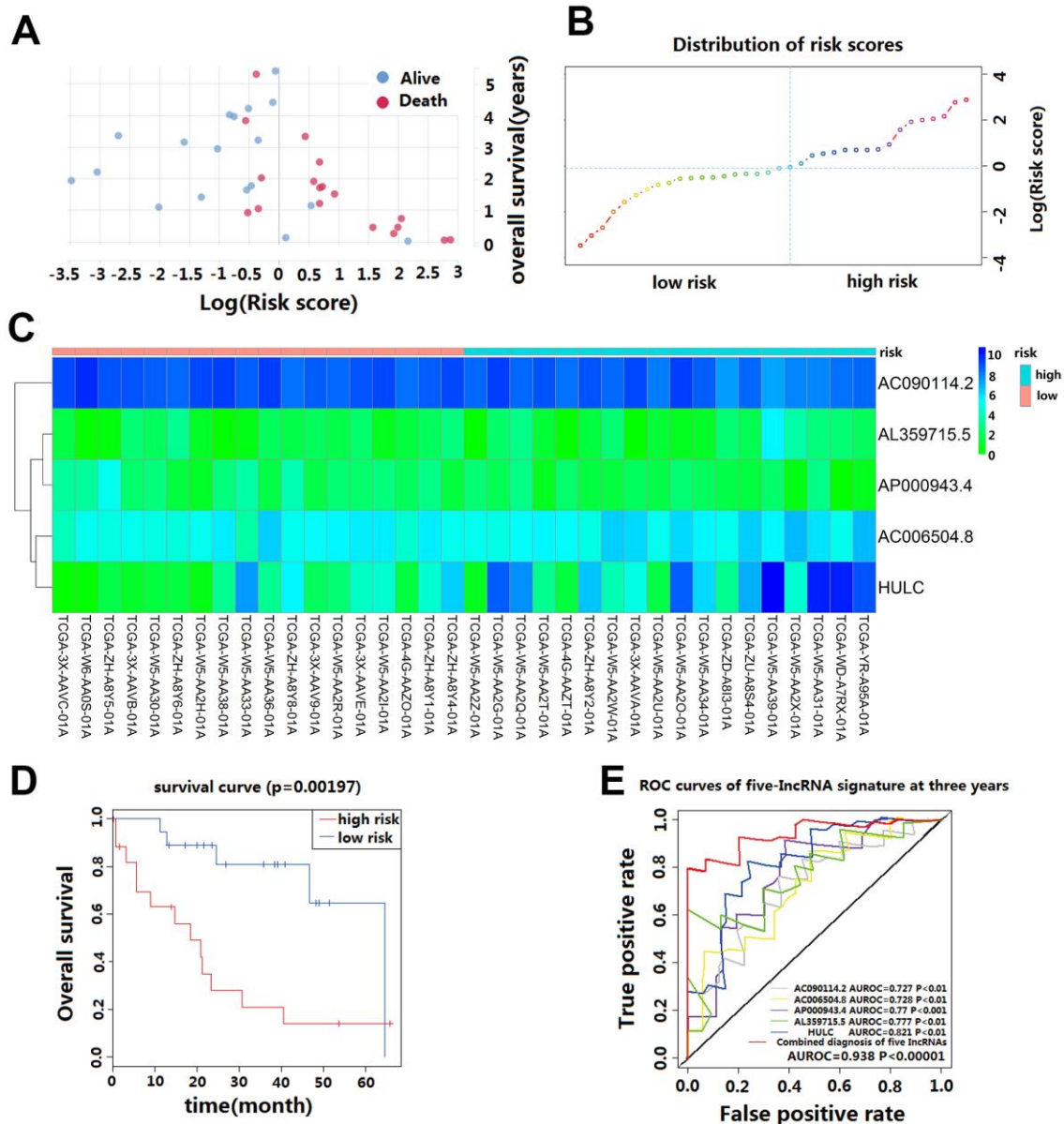


Figure 2. Prognostic evaluation of the five-lncRNA signature in CCA patients in TCGA. (A) The survival status and duration of CCA patients; (B) lncRNA risk score distribution; the blue dashed line indicates dividing the patient into a low-risk group and a high-risk group with a median value as a cut off value. (C) Heatmap of five lncRNA expressions in CCA patients. (D) KM curves based on OS outcomes for risk cutoffs with a *p*-value of less than 0.01 for the log-rank test. (E) Time-dependent ROC curve analysis was predicted by five lncRNA for 3-year survival, and each was performed separately.

Table 2. Univariate and multivariate Cox regression analysis of cholangiocarcinoma patients in the TCGA cohort.

Variables	Univariate analysis		Multivariate analysis	
	HR (95% CI)	P-value	HR (95% CI)	P-value
Age				
(≥60/<60)	1.600 (0.414-6.177)	0.494		
Gender				
(Male/Female)	1.000 (0.269-3.724)	1.000		
Grade				
(G3-G4/G1-G2)	1.000 (0.269-3.724)	1.000		
Stage				
(III-IV/I-II)	1.923 (0.383-9.646)	0.691		
T				
(T2-T3/T1)	1.300 (0.313-5.393)	0.717		
N				
(N1/N0)	2.400 (0.339-16.968)	0.625		
M				
(M1/M0)	1.731 (0.249-12.011)	0.656		
Primary pathology residual tumor				
(R1/R0)	2.000 (0.288-13.910)	0.639		
Family cancer history				
(Yes/No)	0.375 (0.085-1.646)	0.188		
Recurrence				
(Yes/No)	17.500 (3.312-92.475)	<0.001	7.145 (1.481-37.477)	0.014
BMI				
(>24.9/≤24.9)	0.571 (0.130-2.514)	0.571		
History hepatoma risk factors				
(Yes/No)	1.257 (0.333-4.742)	0.735		
Cancer status				
(With tumor/Tumor free)	24.375 (3.822-155.448)	<0.001	1.845 (0.306-11.142)	0.502
Postoperative radiotherapy				
(Yes/No)	0.917 (0.208-4.048)	1.000		
Primary histological type				
(Intrahepatic/Other)	1.000 (0.173-5.772)	1.000		
Five-lncRNA risk score				
(High/Low)	6.760 (1.572-29.068)	0.008	4.879 (1.587-14.996)	0.006

HR: Hazard Ratio; CI: Confidence Interval.

The AUC value and F1 score were calculated to assess the performance of the five-lncRNA signature model. Among the down-regulated lncRNA, the AUC for AL359715.5 and HULC were 0.777 and 0.821, respectively. The AUC for up-regulated lncRNA AC090114.2, AC006504.8, and AP000943.4 were 0.727, 0.728 and 0.77, respectively. For the whole model, the AUC was 0.938 (Figure 2E), and the F1 score was 0.7222. These results indicate that the five-lncRNA signature model has a high predictive value for CCA prognosis.

Performance validation of the five-lncRNA signature model in the validation cohort

We validated the prediction ability of the five-lncRNA signature model in the WMU cohort to assess the

robustness of the model for survival prediction in patients with CCA. Based on the five-lncRNA model and cutoff point derived from the TCGA cohort, patients in the WMU cohort were divided into high-risk group (n=54) and low-risk group(n=36). We compared the KM curve between these two groups and found that the OS of patients in the high-risk group was worse than those in the low-risk group ($p<0.001$, Figure 3B). The AUC value was 0.816 in the WMU cohort (Figure 3C).

Stratified survival analysis of five-lncRNA model in the TCGA cohort

Further stratified analysis showed that the five-lncRNA signature model could accurately distinguish the prognosis between the high-risk group and the low-risk

group in young patients <60 years ($n=14$, $p=0.032$, Figure 4E) and older patients ≥ 60 years ($n=22$, $p=0.037$, Figure 4A). Similarly, stratifying the patients based on the stage of disease, revealed that the five-lncRNA signature model has good discriminatory ability for earlier-stage patients ($n=28$, $p=0.007$, Figure 4F) and advanced-stage patients ($n=8$, $p=0.028$, Figure 4B). For patients with or without recurrence, the five-lncRNA model could divide patients into the high-risk or low-risk groups in those with recurrence ($n=19$, $p=0.005$, Figure 4C) and without recurrence ($n=17$, $p=0.02$, Figure 4G). Moreover, the five-lncRNA signature model could separate the high-risk group and low-risk group for patients with tumors ($n=19$, $p=0.008$, Figure 4D) and those who were tumor-free ($n=15$, $p=0.18$, Figure 4H). Multivariate Cox regression analysis combined with stratification analysis showed that there was no significant difference in OS between the high-risk and low-risk groups with five-LncRNA markers in tumor-free patients, and it this suggests that patients in

the early stages of tumor development may benefit significantly from these prognostic biomarkers.

Identifying the functions of the five-lncRNA signature model

Co-expression analysis showed significant co-expression of 1429 DPCGs, 1440 DPCGs, 300 DPCGs, 495 DPCGs, and 552 PCGs with HULC, AL359715.5, AP000943.4, AC006504.8, AC090114.2, respectively. Functional enrichment analysis indicated that 72 GO biological processes (BP) terms, 21 GO cellular components (CC) terms, and 35 GO molecular functions (MF) terms were enriched for HULC-related DPCGs. Biological processes were mainly involved in the oxidation-reduction process, xenobiotic metabolic process, metabolic process; cellular components were mainly involved in extracellular exosome, mitochondrial matrix, blood microparticle; molecular functions were mainly involved in oxidoreductase activity, electron

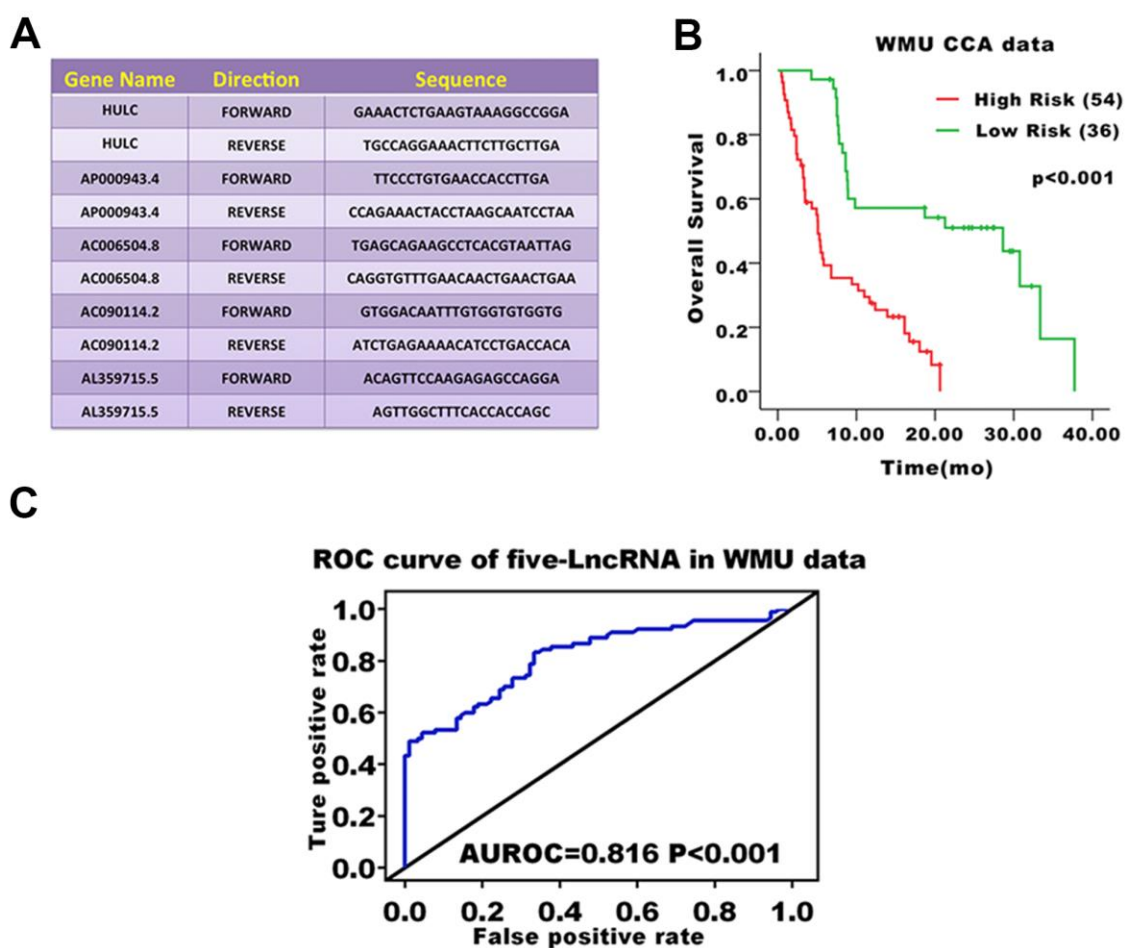


Figure 3. Prognostic evaluation of the five-lncRNA signature in CCA patients in the WMU cohort. (A) Primer sequence of five-lncRNA markers; (B) KM curve analysis of OS validated the prognostic differences between high and low-risk groups in the WMU cohort; (C) ROC curve analysis of 3-year survival validated the reliability of five-lncRNA model.

carrier activity, monooxygenase activity (Supplementary Figure 2A, 2C). There was significant enrichment of 60 KEGG pathways in HULC-associated DPCGs, including leucine, isoleucine and valine degradation, complement and coagulation cascades, fatty acid degradation, carbon metabolism and chemical carcinogenesis (Supplementary Figure 2B, 2C).

47 GO BP terms, 11 GO CC terms, and 33 GO MF terms were enriched for AL359715.5-related DPCGs, whose

biological processes were mainly associated with drug metabolic process, lipid metabolic process, lipoprotein metabolic process; cellular components were mainly associated with organelle membrane, mitochondrion, peroxisome; molecular functions were mainly associated with iron ion binding, heme binding, cholesterol transporter activity (Supplementary Figure 3A, 3C).

56 KEGG pathways were enriched for AL359715.5-related DPCGs, which were primarily linked to Drug

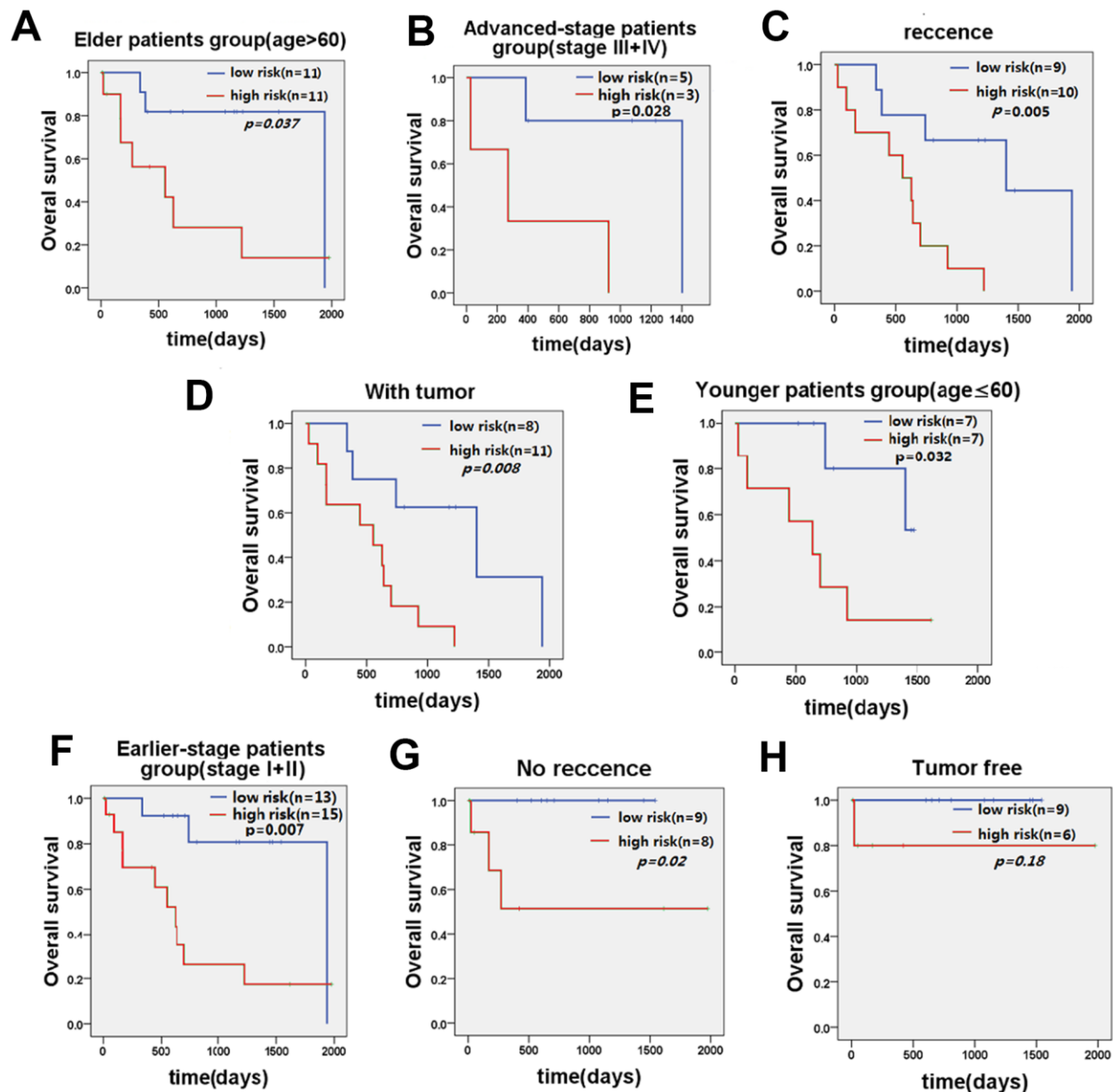


Figure 4. KM curve of OS of patients stratified by age, stage, recurrence, and current tumor status by five-lncRNA signature. (A) KM curves of the elder patients' group and (E) KM curves in the younger patients' group; (B) KM curves in the advanced-stage patients' group and (F) KM curves in the earlier-stage patients' group. (C) KM curves in the recurrence patients' group and (G) KM curves in the no recurrence patients' group. (D) KM curves in the with tumor patients' group and (H) KM curves in the tumor-free patients' group.

metabolism - cytochrome P450, Metabolism of xenobiotics by cytochrome P450, Retinol metabolism, Peroxisome, and Cholesterol metabolism (Supplementary Figure 3B, 3C).

16 GO BP terms, 18 GO CC terms and 4 GO MF terms were enriched for AC006504.8-related DPCGs, whose biological processes were mainly involved in cell division, sister chromatid cohesion, mitotic nuclear division; cellular components were mainly involved in condensed chromosome kinetochore, midbody, nucleoplasm; molecular functions were mainly involved in protein binding, ATP binding, and cadherin binding involved in cell-cell adhesion (Supplementary Figure 4A, 4C). It was a significant enrichment of 7 KEGG pathways in AC006504.8-related DPCGs, including DNA replication, cell cycle, the p53 signaling pathway, Fanconi anemia pathway and progesterone-mediated oocyte maturation (Supplementary Figure 4B, 4C).

7 GO BP terms, 2 GO CC terms and 1 GO MF terms were raised in AC090114.2-associated DPCGs, whose biological processes were mainly related to sister

chromatid cohesion, DNA replication initiation, G1/S transition of mitotic cell cycle; cellular components were mainly associated with cytosol and MCM complex; molecular functions were mainly associated with protein binding (Supplementary Figure 5A, 5C). 4 KEGG pathways were raised in AC090114.2-related DPCGs, which were mainly related to DNA replication, cell cycle, cellular senescence and oocyte meiosis (Supplementary Figure 5B, 5C).

1 GO CC terms and 1 GO MF terms could be found for DPCGs related to AP000943.4, whose cellular components were associated with cytoskeleton; molecular functions were associated with extracellular matrix organization and involved a KEGG pathway for Human papillomavirus infection (Data not shown).

Functional evaluation of common DPCGs for five-lncRNA signature model

The intersection of the DPCGs corresponding to the five-lncRNA signature model showed that 171 DPCGs were shared by this five-lncRNA signature (Figure 5A).

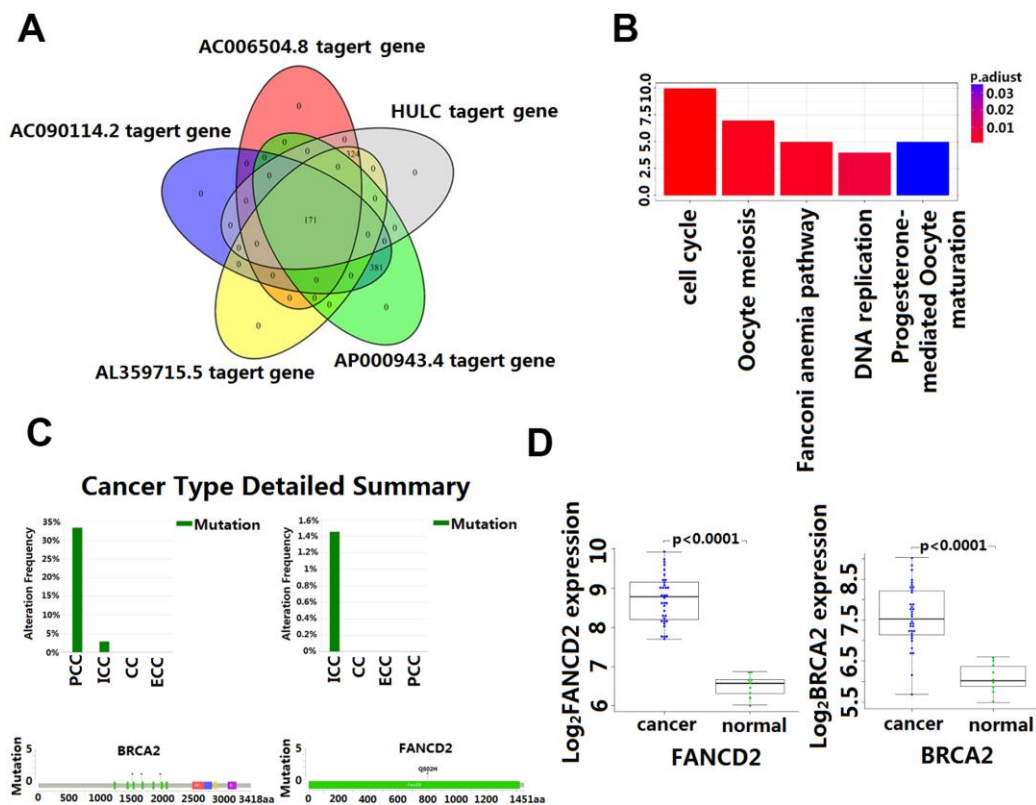


Figure 5. Enrichment and analysis of five lncRNA in the presence of common DPCGs. (A) Venn diagram showing 171 common DPCGs of the five-lncRNA. (B) The KEGG pathways were significantly associated with the enrichment of 171 common protein-coding genes co-expressed with five-lncRNA. The ordinate is the number of DPCGs that is enriched to the target gene. (C) Mutation of FANCD1 and FADCD2 genes in cholangiocarcinoma (from the cbiportal database <http://www.cbiportal.org/>). (D) Expression of FANCD1 and FADCD2 genes in cholangiocarcinoma.

The common KEGG pathway is cell cycle, DNA replication, oocyte meiosis, Fanconi anemia pathway, and progesterone-mediated oocyte maturation (Figure 5B and Supplementary Table 1).

GSEA between the high-risk group and low-risk group

Through GSEA analysis, we clarified the significant difference in survival between the high-risk and low-risk groups. The results showed significant enrichment of markers including the "complement pathway" in the high-risk group. Pathways including IL-2 Receptor Beta Chain in T cell Activation, Keratinocyte Differentiation, T cell receptor pathway, and Neurotrophin signaling pathway were enriched in the low-risk group [18], many of these being closely related to the occurrence and development of cancer [19] (Figure 6 and Supplementary Table 2).

DISCUSSION

Currently, the molecular genotype for a variety of tumors (breast cancer, gastric cancer, and colorectal cancer) has been applied in a clinical setting. Some

molecular genotypes are not only used to predict the prognosis but also to select the best therapy target [20]. The comprehensive study of the mechanism has led to the discovery of many kinds of targeted drugs used in the treatment of these diseases [21]. However, for CCA, there are relatively few studies on prognostic molecular markers. Therefore, establishing a molecular prediction model in CCA for guiding personalized treatment and predicting prognosis is particularly urgent. In this study, we established a prediction model based on five lncRNA for the prognosis of CCA and validate its reliability in an independent clinical center biobank. The molecular mechanism of these five lncRNA was further explored by the signal pathway analysis.

There is growing evidence that lncRNA plays a key role in transcription and post-transcriptional regulation of gene expression [22–24] as well as in different cells and developmental processes [25–27]. Experimental evidence indicates that abnormal expression of lncRNA is relative to the onset of various diseases including gastric cancer, breast cancer, HCC, lung cancer, and CCA [28–30]. Recent reports indicate that oxidative stress up-regulates the dysfunction of lncRNA H19 and HULC, and then modulates CCA cell migration and

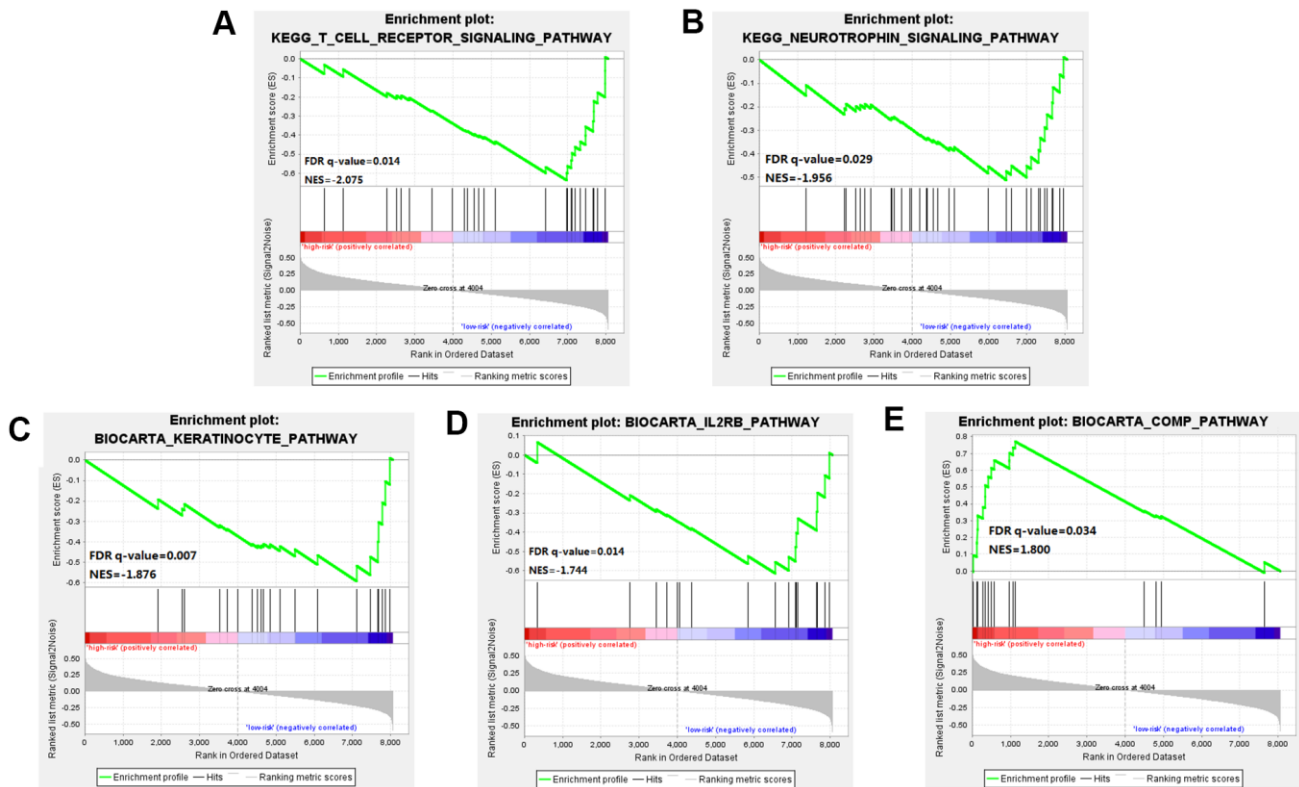


Figure 6. Gene Set Enrichment Analysis (GSEA) was performed between the high risk score group and the low-risk score group. (A–D) Pathways including IL-2 Receptor Beta Chain in T cell Activation, Keratinocyte Differentiation, T cell receptor pathway, and Neurotrophin signaling pathway were enriched in the low-risk group. (E) The results showed significant enrichment of markers including the "complement pathway" in the high-risk group.

invasion through ceRNA targeting IL-6 and CXCR4 [31]. Similarly, the lncRNA CPS1-IT1 is up-regulated in intrahepatic CCA. Conversely, knockdown of CPS1 and/or CPS1-IT1 reduced the proliferation and increased apoptosis of ICC-9810 cells [30]. By comparing the expression of AFAP1-AS1 in CCA tissues and paired adjacent tissues and analyzing the relationship between AFAP1-AS1 expression and the clinical features of CCA, it was discovered that AFAP1-AS1 is significantly associated with the malignant degree and poor prognosis of CCA. Studies have shown that knockdown AFAP1-AS1 inhibits tumor growth *in vivo* and inhibits cell proliferation and invasion *in vitro* [32]. Other studies have found that certain lncRNA play a critical role in the metastasis and malignant progression of CCA. It has been reported that some lncRNA increased in the tissues of patients with advanced CCA and lymph node metastasis, and through inhibition and overexpression in lncRNA experiments, it was found that this overexpression of certain lncRNA may promote the growth and metastasis of CCA through some miRNA (miRNA-200c, miR-296-5p, et al.) [31]. Another study has found that lncRNA-DANCR can bind to EZH2 and regulate histone methylation FBP1 promoter expression, which regulates the growth and migration of CCA cells [33].

Although the study of the lncRNA function has attracted more and more attention and a great number of lncRNA have been identified in the human genome, the function of most lncRNA has not been fully revealed. Functional annotation of the gene encoding the lncRNA-associated co-expressed protein is a viable method for finding the biological characteristics of lncRNA [31]. By extension, annotation of lncRNA function through co-expressed genes was reported to be effective [34]. In this study, GO and KEGG enrichment analysis was used to identify co-expressed mRNAs of the five lncRNA to speculate on the functions of the predictive lncRNA. Our data revealed that the HULC and AL359715.5 participated in a number of biological processes that were most relevant to the cholesterol and fatty acid metabolism which is reported to be responsible for the growth and accelerated development of CCA [34, 35]. Also, of interest is the identification of the complement and coagulation cascades that are involved in many physiological and pathological processes, including those in the inflammatory process which, once dysregulated become an important factor in tumorigenesis [36]. In this study, we found that AC006504.8 was enriched in the p53 signaling pathway. The molecular epidemiological analysis revealed that p53 is mutated in almost all kinds of tumors, and approximately 5% of patients with colorectal cancer, lung cancer, melanoma, sarcoma, head and neck cancer, leukemia, esophageal cancer, ovarian cancer, testicular cancer, and cervical cancer have been

found to have p53 mutations [37, 38]. Of significance to this study is the amount of research that has indicated p53 inactivation plays a key role in the occurrence and development of CCA [39]. The mechanisms by which AC006504.8 is involved in CCA are probably related to cell cycle and DNA replication. The 171 DPCGs intersected by the five-lncRNA signature were enriched in the function of the Fanconi anemia (FA) pathway. Fanconi anemia is a recessive genetic disorder characterized by congenital malformation, bone marrow failure, and high susceptibility to cancers [36, 40]. It is a cancer susceptibility gene involved in the repairing of genomic damage and maintaining genomic stability [41]. Recent evidence indicates that genetic instability is a key factor in the metastasis and recurrence of malignant tumors. Many studies have shown that mutations and abnormal expression of the FANCD1 and FANCD2, two major genes in the Fanconi anemia pathway, are significantly associated with poor prognosis of CCA [42]. Our study also showed that FANCD1 and FANCD2 mutated to different degrees in CCA (Figure 5C), and their expression in CCA and matched paracarcinoma tissues was also significantly different (Figure 5D). These results would seem to suggest that the predictive five-lncRNA may mediate the development and progression of CCA via DPCG interactions in biological processes related to cancer. However, more experimental studies are needed to further explain the potential roles of these lncRNA in CCA. To our knowledge, four out of the five lncRNA biomarker functions have never been reported. Therefore, we postulate that further investigation of the function of the lncRNA will contribute to early diagnosis and provide a clinical basis for the development of new prognostic factors in CCA.

In summary, we systematically studied the lncRNA expression profiles of CCA patients and their corresponding clinical information and found five-lncRNA (HULC, AP000943.4, AC006504.8, AC090114.2, AL359715.5) signature showing the risk scoring model in this study was an excellent way to classify patients with different survival outcomes. Also, the five lncRNA molecular biomarkers performed very well in predicting 3-year survival in patients with CCA, which could be an independent predictor of survival prognosis, and which could provide novel insights into the molecular mechanisms of CCA tumorigenesis and development. However, this study has some limitations. Firstly, our five-lncRNA signature model was only tested and validated in the TCGA and WMU cohort. If possible, it should be validated in other independent cohorts. Secondly, our research only investigated the biological function of predictive lncRNA by computational methods. It should be supplemented with *in vitro* cell research and animal experiments *in vivo*. Combining

these data will help to unravel the mechanism of lncRNA involved in CCA tumorigenesis.

MATERIALS AND METHODS

Screening differentially expressed lncRNA in CCA patients in the discovery cohort

Pre-processed level 3 RNA sequencing count data and relative clinical information for CCA patients were obtained from the Genomic Data Commons Data Portal database (TCGA, <https://portal.gdc.cancer.gov/projects/TCGA-CHOL>) [43]. Ultimately, 60,483 RNA-ENSG_ID expression profiles in 36 CCA patients were included for further analysis as the discovery cohort (TCGA cohort).

The Gencode.v27.long_noncoding_RNAs.gtf compressed file was downloaded from the GENCODE database (release 27) (<http://www.genecodegenes.org/>) and the transformed data (antisense, lincRNA, and sense_intronic) was determined as lncRNA [44]. It was then filtered by removing the exon-expressing lncRNA from any known coding gene by GENCODE-based gene annotation. A total of 59,264 lncRNA-ENSG_IDs, and RNA-ENSG_ID and lncRNA-ENSG_ID were intersected, 13,126 lncRNA-Gene_names were obtained (Supplementary Figure 6). Samples in which the lncRNA with RPKM expression value is 0 more than 20% were removed. Finally, 3651 lncRNA were kept for further analysis. Furthermore, we downloaded Homo_sapiens.GRCh38.91.chr.gtf zip file from Ensembl genome browser 91 (ftp://ftp.ensembl.org/pub/release-91/gtf/homo_sapiens) to obtain 57,000 RNA-Gene_symbol expression profiles, then used “mygene” package in R to find the RNA with human entrezgene symbol (GeneID) in the 57,000 RNAs.

Differentially expressed genes (DEGs) between 36 CCA tissues (30 iCCA tissues, 6 pCCA/dCCA tissues) and 9 normal tissues with the $|\log_2 \text{fold change} (\log_2 \text{FC})| \geq 1$ and false discovery rate (FDR) < 0.01 were considered as selection criteria for subsequent analysis. At the same time, differentially expressed analysis was performed on the RNAs identified with GeneID.

Establishment and performance evaluation of a lncRNA-based prediction model in the discovery cohort

Firstly, we assessed the association between differentially expressed lncRNA and the overall survival (OS) of CCA patients in the TCGA cohort by univariate Cox proportional hazards regression analysis. lncRNA that were statistically significant were included in further multivariate Cox regression analysis. Then, using the

multivariate Cox regression analysis, we identified independent lncRNA predictors. Furthermore, the corresponding coefficients of each lncRNA in the Cox regression model were obtained to calculate the risk score [45]. Finally, a lncRNA-based prediction model to evaluate the CCA patient survival outcomes was established as below:

$$\text{lncRNA-based Risk Score} = \sum_{i=1}^N (\text{Coe}_i \times \text{EV}_i)$$

In this formula, N represents the number of prognostic lncRNA, Coe_i is the coefficient of the lncRNA_i in the multivariate Cox regression analysis, EV_i represents the expression level of the lncRNA_i.

We used the lncRNA-based model to calculate the risk score for each CCA patients in the TCGA cohort. Setting the median value of log (Risk Score) as a threshold, CCA patients were divided into high-risk and low-risk groups and the difference of OS between these two groups was compared. Then univariate and multivariate Cox regression analysis was performed to determine whether the lncRNA signature was an independent predictor variable of other clinicopathologic features for survival outcomes. Further stratification analysis was conducted on clinicopathologic characteristics, which were statistically significant in a multivariate Cox regression model to determine the lncRNA signature model's predictive capacity within the same clinical features. We calculated the area under the time-dependent receiver operating characteristic (ROC) curve (AUC) within a 3-year survival period to evaluate the sensitivity and specificity of the lncRNA model to predict survival outcomes.

Verification of lncRNA signature model for survival prediction in the validation cohort

We performed validation of discovered lncRNA in fresh frozen tissues from 90 CCA patients who underwent surgery between November 2012 and December 2015 at the First Affiliated Hospital of Wenzhou Medical University (WMU cohort). The patient inclusion criteria were as follows: (1) pathological diagnosed with primary CCA; (2) had completed clinicopathological and follow-up monitoring; (3) had no anti-tumor treatment before this surgical resection. Exclusion criteria were as follows: (1) previous radiofrequency or other anti-tumor treatment before surgery; (2) patients who were lost to follow-up after surgery. The study was approved by the institutional review boards of First Affiliated Hospital of Wenzhou Medical University and written informed consent was obtained from all patients. Patients demographics and clinicopathological characteristics are shown in Supplementary Table 3.

The lncRNA expression of primary CCA tumor fresh frozen samples was assessed by real-time quantitative PCR. RNAeasy mini kit (Qiagen, CA, USA) was used to extract the total RNA. High-Capacity cDNA Reverse Transcription kit from Applied Biosystems (Grand Island, NY, USA) was used to synthesize cDNA from 2 µg of total RNA. Semi-quantitative detection of mRNA was performed using an ABI 7300 RT-PCR system. Relative quantification of mRNA levels was performed with 18S ribosomal RNA as an internal reference gene and data from the $\Delta\Delta C_t$ method. The statistical mean and standard error were determined by the ΔC_t value. All data were independently inputted three times. The primer sequences used in the present study are shown in Figure 4A.

Through the same lncRNA signature model and cutoff level derived from the discovery cohort, patients in the WMU cohort were divided into high-risk and low-risk groups. Then, we investigated the performance of the lncRNA signature model in the WMU cohort.

Co-expression and functional enrichment analysis

The Spearman correlation coefficient was calculated between lncRNA expression level and the differentially expressed protein-coding genes (DPCGs). The DPCGs with Spearman correlation coefficient greater than 0.50 were considered to be lncRNA-related DPCGs. Gene Ontology (GO), Cell Component (CC), Biological Process (BP), Molecular Function (MF), Cell Component (CC) and Kyoto Encyclopedia Gene and Genome (KEGG) pathway enrichment assays were performed using the Database for Annotation, Visualization, and Integrated Discovery (DAVID, version 6.8, <https://david.ncifcrf.gov/>) [46]. Gene Set Enrichment Analysis (GSEA 3.0) analyses were carried out to elucidate the survival difference between high-risk and low-risk groups. GSEA analyses were implemented with java software GSEA (<http://software.broadinstitute.org/gsea/index.jsp>) [18]. GO (BP, MF, CC) terms, KEGG pathways, and GSEA analyses with adjusted *P* value or a false discovery rate (FDR) less than 0.05 were considered statistically significant.

Statistical analysis

The continuous variables were expressed as mean \pm standard deviation or median (quartile range), and categorical variables were presented as frequencies (percentages). A Chi-square test was used to compare the differences between independent groups. Cox proportional hazard regression analysis was conducted to evaluate the association of lncRNA signature in predicting overall survival in CCA patients. Kaplan-Meier (KM) analysis was used to determine survival outcomes. The

median values were used as a cut-off level to plot the KM curves, and the log-rank test was performed to evaluate the statistical significance. The results of the stepwise multivariate Cox regression analysis of the AIC (Akaike Information Criterion, assessing the goodness of fit of a statistical model) test yielded a predictive model with optimal interpretation and information effectiveness. A linear correlation model was performed to evaluate the relationships between the variables and the Pearson correlation coefficient or Spearman rank correlation coefficient was used to present the result. Unless otherwise indicated, all statistical tests were two-sided and a *p*-value < 0.05 was considered as statistically significant. All data analysis was performed with R (version 3.3.3; <http://www.r-project.org/>). The differential expression of the lncRNA profile was estimated by the R “edgeR” package. Unsupervised hierarchical clustering analysis was accomplished by the R “pheatmap” package and represented as a volcano plot. KM survival analysis and Cox proportional hazard regression analysis was performed by the R “survival” package. The AUC value was calculated by the R “Survival ROC” package. The R “clusterProfiler”, “pathview” and “venn” package were used to find the common DPCGs of lncRNAs for KEGG pathways.

Data sharing and data accessibility

The data used to support the findings of this study are available from the corresponding author upon request.

AUTHOR CONTRIBUTIONS

X.X., Y.W., S.Z., and G.C. designed the research. J.L., Y.W., Z.Y., X.D., L.Y., P.G., C.H., and Q.Z. acquired and analyzed the data. X.X., Y.W., G.C., J.L., Z.C., Z.D., Q.Z., and Z.Y. contributed to the writing of the manuscript. Finally, all authors have reviewed and approved the final submitted manuscript.

CONFLICTS OF INTEREST

The authors declare that there is no conflicts of interest.

FUNDING

Research Foundation of the National Health Commission of China-Major Medical and Health Technology Project for Zhejiang Province, Grant/Award Number: WKJ-ZJ-1706; Natural Science Foundation of Zhejiang Province, Grant/Award Number: LY17H160047; Public Projects of Zhejiang Province, Grant/Award Numbers: 2016C37127, 2018C37114; National Natural Science Foundation of China, Grant/Award Numbers: 81470868, 81772628,

REFERENCES

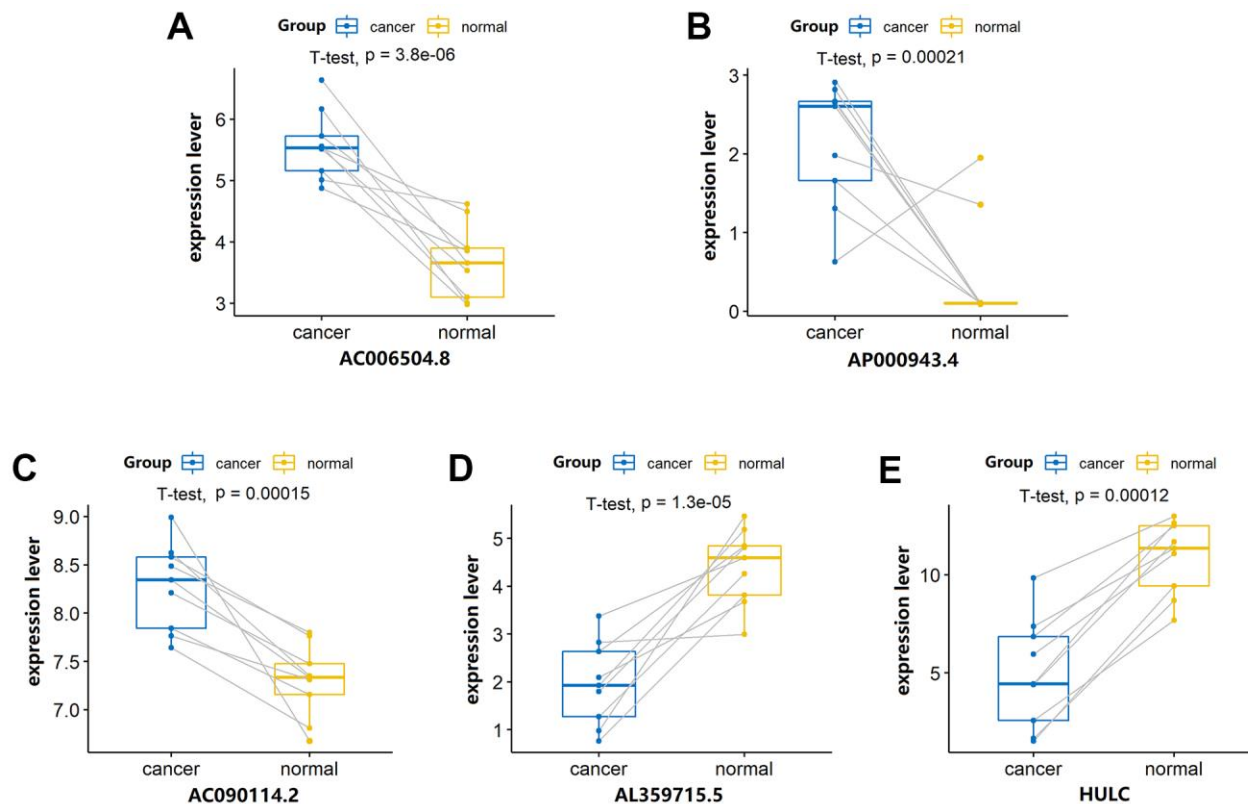
1. Khan SA, Davidson BR, Goldin RD, Heaton N, Karani J, Pereira SP, Rosenberg WM, Tait P, Taylor-Robinson SD, Thillainayagam AV, Thomas HC, Wasan H, and British Society of Gastroenterology. Guidelines for the diagnosis and treatment of cholangiocarcinoma: an update. *Gut*. 2012; 61:1657–69. <https://doi.org/10.1136/gutjnl-2011-301748> PMID:22895392
2. Shaib Y, El-Serag HB. The epidemiology of cholangiocarcinoma. *Semin Liver Dis*. 2004; 24:115–25. <https://doi.org/10.1055/s-2004-828889> PMID:15192785
3. Valle J, Wasan H, Palmer DH, Cunningham D, Anthony A, Maraveyas A, Madhusudan S, Iveson T, Hughes S, Pereira SP, Roughton M, Bridgewater J, and ABC-02 Trial Investigators. Cisplatin plus gemcitabine versus gemcitabine for biliary tract cancer. *N Engl J Med*. 2010; 362:1273–81. <https://doi.org/10.1056/NEJMoa0908721> PMID:20375404
4. Guro H, Kim JW, Choi Y, Cho JY, Yoon YS, Han HS. Multidisciplinary management of intrahepatic cholangiocarcinoma: current approaches. *Surg Oncol*. 2017; 26:146–52. <https://doi.org/10.1016/j.suronc.2017.03.001> PMID:28577720
5. Kou Y, Koag MC, Lee S. N7 methylation alters hydrogen-bonding patterns of guanine in duplex DNA. *J Am Chem Soc*. 2015; 137:14067–70. <https://doi.org/10.1021/jacs.5b10172> PMID:26517568
6. Kou Y, Koag MC, Cheun Y, Shin A, Lee S. Application of hypoidite-mediated aminyl radical cyclization to synthesis of solasodine acetate. *Steroids*. 2012; 77:1069–74. <https://doi.org/10.1016/j.steroids.2012.05.002> PMID:22583912
7. Wen R, Umeano AC, Kou Y, Xu J, Farooqi AA. Nanoparticle systems for cancer vaccine. *Nanomedicine (Lond)*. 2019; 14:627–48. <https://doi.org/10.2217/nnm-2018-0147> PMID:30806568
8. Shen FZ, Zhang BY, Feng YJ, Jia ZX, An B, Liu CC, Deng XY, Kulkarni AD, Lu Y. Current research in perineural invasion of cholangiocarcinoma. *J Exp Clin Cancer Res*. 2010; 29:24. <https://doi.org/10.1186/1756-9966-29-24> PMID:20219134
9. Spizzo R, Almeida MI, Colombatti A, Calin GA. Long non-coding RNAs and cancer: a new frontier of translational research? *Oncogene*. 2012; 31:4577–87. <https://doi.org/10.1038/onc.2011.621> PMID:22266873
10. Zeng JH, Liang L, He RQ, Tang RX, Cai XY, Chen JQ, Luo DZ, Chen G. Comprehensive investigation of a novel differentially expressed lncRNA expression profile signature to assess the survival of patients with colorectal adenocarcinoma. *Oncotarget*. 2017; 8:16811–28. <https://doi.org/10.18632/oncotarget.15161> PMID:28187432
11. Zhou M, Xu W, Yue X, Zhao H, Wang Z, Shi H, Cheng L, Sun J. Relapse-related long non-coding RNA signature to improve prognosis prediction of lung adenocarcinoma. *Oncotarget*. 2016; 7:29720–38. <https://doi.org/10.18632/oncotarget.8825> PMID:27105492
12. Zhang F, Wan M, Xu Y, Li Z, Kang P, Jiang X, Wang Y, Wang Z, Zhong X, Li C, Cui Y. Transcriptome analysis reveals dysregulated long non-coding RNAs and mRNAs associated with extrahepatic cholangiocarcinoma progression. *Oncol Lett*. 2017; 14:6079–84. <https://doi.org/10.3892/ol.2017.6987> PMID:29113249
13. Wang C, Mao ZP, Wang L, Wu GH, Zhang FH, Wang DY, Shi JL. Long non-coding RNA MALAT1 promotes cholangiocarcinoma cell proliferation and invasion by activating PI3K/Akt pathway. *Neoplasma*. 2017; 64:725–31. https://doi.org/10.4149/neo_2017_510 PMID:28592124
14. Xu Y, Yao Y, Leng K, Li Z, Qin W, Zhong X, Kang P, Wan M, Jiang X, Cui Y. Long non-coding RNA UCA1 indicates an unfavorable prognosis and promotes tumorigenesis via regulating AKT/GSK-3 β signaling pathway in cholangiocarcinoma. *Oncotarget*. 2017; 8:96203–14. <https://doi.org/10.18632/oncotarget.21884> PMID:29221199
15. Xu Y, Leng K, Li Z, Zhang F, Zhong X, Kang P, Jiang X, Cui Y. The prognostic potential and carcinogenesis of long non-coding RNA TUG1 in human cholangiocarcinoma. *Oncotarget*. 2017; 8:65823–35. <https://doi.org/10.18632/oncotarget.19502> PMID:29029475
16. Jiang XM, Li ZL, Li JL, Zheng WY, Li XH, Cui YF, Sun DJ. LncRNA CCAT1 as the unfavorable prognostic biomarker for cholangiocarcinoma. *Eur Rev Med Pharmacol Sci*. 2017; 21:1242–47. PMID:28387907
17. Shi X, Zhang H, Wang M, Xu X, Zhao Y, He R, Zhang M, Zhou M, Li X, Peng F, Shi C, Shen M, Wang X, et al.

- LncRNA AFAP1-AS1 promotes growth and metastasis of cholangiocarcinoma cells. *Oncotarget*. 2017; 8:58394–404.
<https://doi.org/10.18632/oncotarget.16880>
PMID:28938565
18. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005; 102:15545–50.
<https://doi.org/10.1073/pnas.0506580102>
PMID:16199517
 19. Janelle V, Langlois MP, Tarrab E, Lapierre P, Poliquin L, Lamarre A. Transient complement inhibition promotes a tumor-specific immune response through the implication of natural killer cells. *Cancer Immunol Res*. 2014; 2:200–06.
<https://doi.org/10.1158/2326-6066.CIR-13-0173>
PMID:24778316
 20. Sia D, Villanueva A, Friedman SL, Llovet JM. Liver cancer cell of origin, molecular class, and effects on patient prognosis. *Gastroenterology*. 2017; 152:745–61.
<https://doi.org/10.1053/j.gastro.2016.11.048>
PMID:28043904
 21. Liu B, Guo H, Zhang J, Xue J, Yang Y, Qin T, Xu J, Guo Q, Zhang D, Qian W, Li B, Hou S, Dai J, et al. In-depth characterization of a pro-antibody-drug conjugate by LC-MS. *Mol Pharm*. 2016; 13:2702–10.
<https://doi.org/10.1021/acs.molpharmaceut.6b00280>
PMID:27377124
 22. Singh P, Patel T. Advances in the diagnosis, evaluation and management of cholangiocarcinoma. *Curr Opin Gastroenterol*. 2006; 22:294–99.
<https://doi.org/10.1097/01.mog.0000218967.60633.64>
PMID:16550045
 23. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet*. 2009; 10:155–59.
<https://doi.org/10.1038/nrg2521>
PMID:19188922
 24. Hawkins PG, Morris KV. Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5. *Transcription*. 2010; 1:165–75.
<https://doi.org/10.4161/trns.1.3.13332>
PMID:21151833
 25. Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, Freier SM, Bennett CF, Sharma A, Bubulya PA, Blencowe BJ, Prasanth SG, Prasanth KV. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell*. 2010; 39:925–38.
<https://doi.org/10.1016/j.molcel.2010.08.011>
PMID:20797886
 26. Evans JR, Feng FY, Chinnaiyan AM. The bright side of dark matter: lncRNAs in cancer. *J Clin Invest*. 2016; 126:2775–82.
<https://doi.org/10.1172/JCI84421> PMID:27479746
 27. Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet*. 2016; 17:47–62.
<https://doi.org/10.1038/nrg.2015.10> PMID:26666209
 28. Wang F, Yang H, Deng Z, Su Y, Fang Q, Yin Z. HOX antisense lincRNA HOXA-AS2 promotes tumorigenesis of hepatocellular carcinoma. *Cell Physiol Biochem*. 2016; 40:287–96.
<https://doi.org/10.1159/000452545> PMID:27855366
 29. Wang WT, Ye H, Wei PP, Han BW, He B, Chen ZH, Chen YQ. LncRNAs H19 and HULC, activated by oxidative stress, promote cell migration and invasion in cholangiocarcinoma through a ceRNA manner. *J Hematol Oncol*. 2016; 9:117.
<https://doi.org/10.1186/s13045-016-0348-0>
PMID:27809873
 30. Ma SL, Li AJ, Hu ZY, Shang FS, Wu MC. Co-expression of the carbamoyl-phosphate synthase 1 gene and its long non-coding RNA correlates with poor prognosis of patients with intrahepatic cholangiocarcinoma. *Mol Med Rep*. 2015; 12:7915–26.
<https://doi.org/10.3892/mmr.2015.4435>
PMID:26499888
 31. Ma H, Hao Y, Dong X, Gong Q, Chen J, Zhang J, Tian W. Molecular mechanisms and function prediction of long noncoding RNA. *ScientificWorldJournal*. 2012; 2012:541786.
<https://doi.org/10.1100/2012/541786> PMID:23319885
 32. Lu X, Zhou C, Li R, Deng Y, Zhao L, Zhai W. Long noncoding RNA AFAP1-AS1 promoted tumor growth and invasion in cholangiocarcinoma. *Cell Physiol Biochem*. 2017; 42:222–30.
<https://doi.org/10.1159/000477319> PMID:28535506
 33. Han L, Zhang EB, Yin DD, Kong R, Xu TP, Chen WM, Xia R, Shu YQ, De W. Low expression of long noncoding RNA PANDAR predicts a poor prognosis of non-small cell lung cancer and affects cell apoptosis by regulating Bcl-2. *Cell Death Dis*. 2015; 6:e1665.
<https://doi.org/10.1038/cddis.2015.30>
PMID:25719249
 34. Samudio I, Harmancey R, Fiegl M, Kantarjian H, Konopleva M, Korchin B, Kaluarachchi K, Bornmann W, Duvvuri S, Taegtmeier H, Andreeff M. Pharmacologic inhibition of fatty acid oxidation sensitizes human leukemia cells to apoptosis induction. *J Clin Invest*. 2010; 120:142–56.

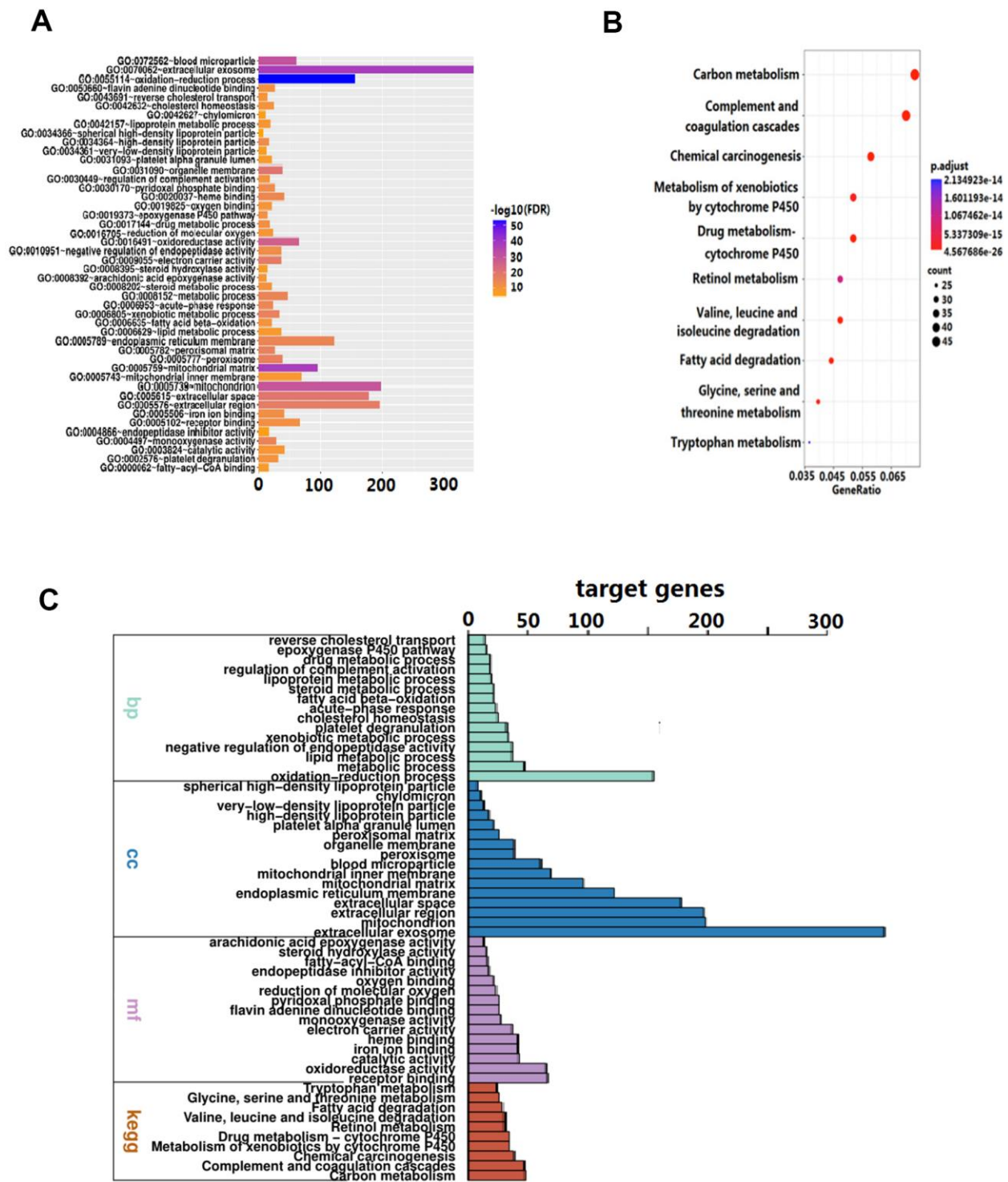
- <https://doi.org/10.1172/JCI38942>
PMID:20038799
35. Iyengar NM, Hudis CA, Dannenberg AJ. Obesity and cancer: local and systemic mechanisms. *Annu Rev Med.* 2015; 66:297–309.
<https://doi.org/10.1146/annurev-med-050913-022228>
PMID:25587653
36. Vundinti BR, Korgaonkar S, Ghosh K. Incidence of malignancy and clonal chromosomal abnormalities in Fanconi anemia. *Indian J Cancer.* 2010; 47:397–99.
<https://doi.org/10.4103/0019-509X.73575>
PMID:21131752
37. Robles AI, Harris CC. Clinical outcomes and correlates of TP53 mutations and cancer. *Cold Spring Harb Perspect Biol.* 2010; 2:a001016.
<https://doi.org/10.1101/cshperspect.a001016>
PMID:20300207
38. Khan SA, Thomas HC, Toledano MB, Cox IJ, Taylor-Robinson SD. P53 mutations in human cholangiocarcinoma: a review. *Liver Int.* 2005; 25:704–16.
<https://doi.org/10.1111/j.1478-3231.2005.01106.x>
PMID:15998419
39. Scheffner M, Werness BA, Huibregtse JM, Levine AJ, Howley PM. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell.* 1990; 63:1129–36.
[https://doi.org/10.1016/0092-8674\(90\)90409-8](https://doi.org/10.1016/0092-8674(90)90409-8)
PMID:2175676
40. Mirchandani KD, D'Andrea AD. The Fanconi anemia/BRCA pathway: a coordinator of cross-link repair. *Exp Cell Res.* 2006; 312:2647–53.
<https://doi.org/10.1016/j.yexcr.2006.06.014>
PMID:16859679
41. Masserot C, Peffault de Latour R, Rocha V, Leblanc T, Rigolet A, Pascal F, Janin A, Soulier J, Gluckman E, Socié G. Head and neck squamous cell carcinoma in 13 patients with Fanconi anemia after hematopoietic stem cell transplantation. *Cancer.* 2008; 113:3315–22.
<https://doi.org/10.1002/cncr.23954>
PMID:18831513
42. Jacquemont C, Taniguchi T. The Fanconi anemia pathway and ubiquitin. *BMC Biochem.* 2007 (Suppl 1); 8:S10.
<https://doi.org/10.1186/1471-2091-8-S1-S10>
PMID:18047734
43. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature.* 2011; 474:609–15.
<https://doi.org/10.1038/nature10166>
PMID:21720365
44. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, Barnes I, Bignell A, Boychenko V, et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* 2012; 22:1760–74.
<https://doi.org/10.1101/gr.135350.111>
PMID:22955987
45. Zhou M, Zhao H, Wang Z, Cheng L, Yang L, Shi H, Yang H, Sun J. Identification and validation of potential prognostic lncRNA biomarkers for predicting survival in patients with multiple myeloma. *J Exp Clin Cancer Res.* 2015; 34:102.
<https://doi.org/10.1186/s13046-015-0219-5>
PMID:26362431
46. Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009; 4:44–57.
<https://doi.org/10.1038/nprot.2008.211>
PMID:19131956

SUPPLEMENTARY MATERIALS

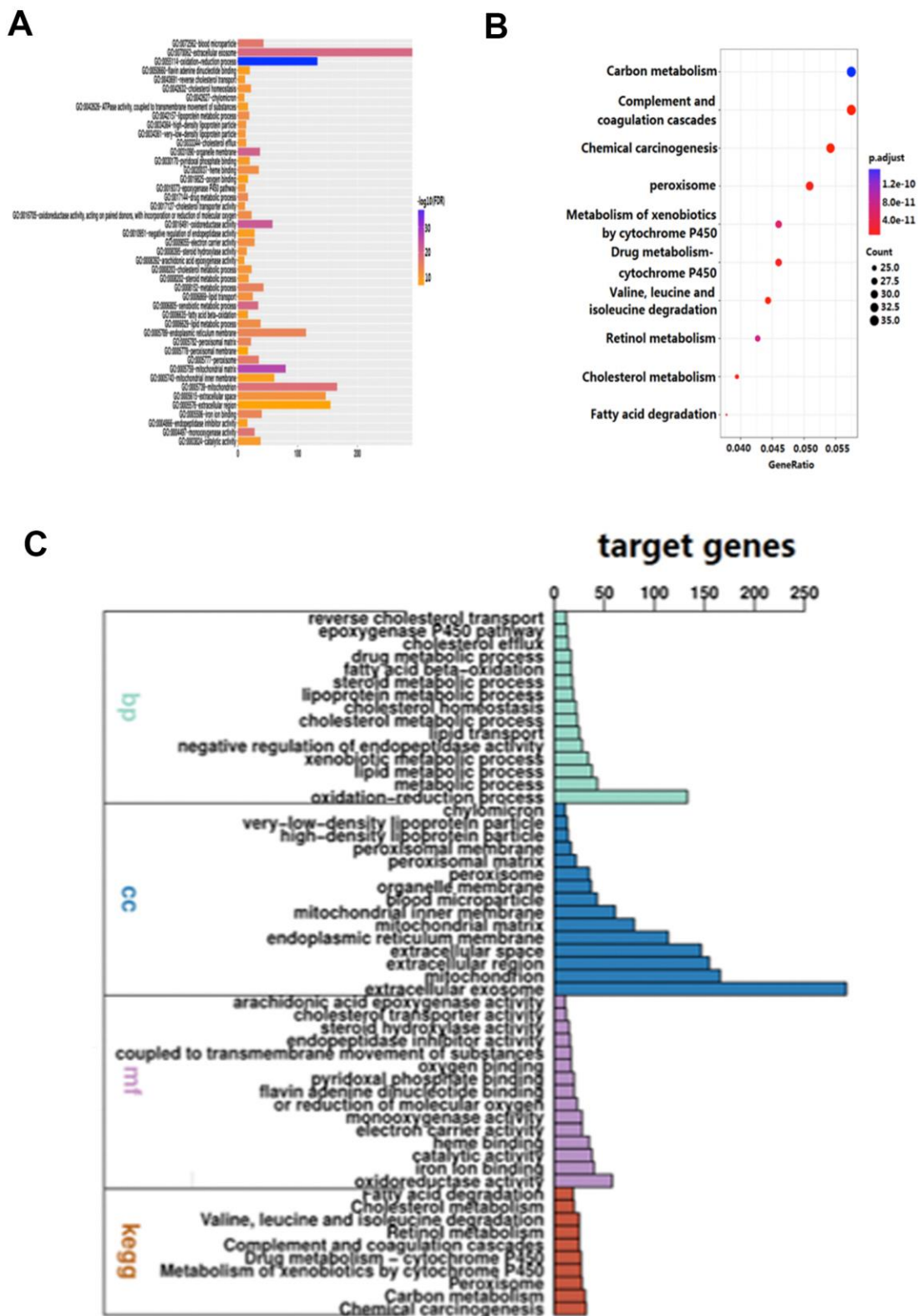
Supplementary Figures



Supplementary Figure 1. The differentially expressed level of five-lncRNA in paired carcinoma and paracancerous tissues. (A–C) AC006504.8, AP000943.4 and AC090114.2 expressed higher in cancer tissues. (D, E) AL359715.5 and HULC expressed lower in paracancerous tissues.

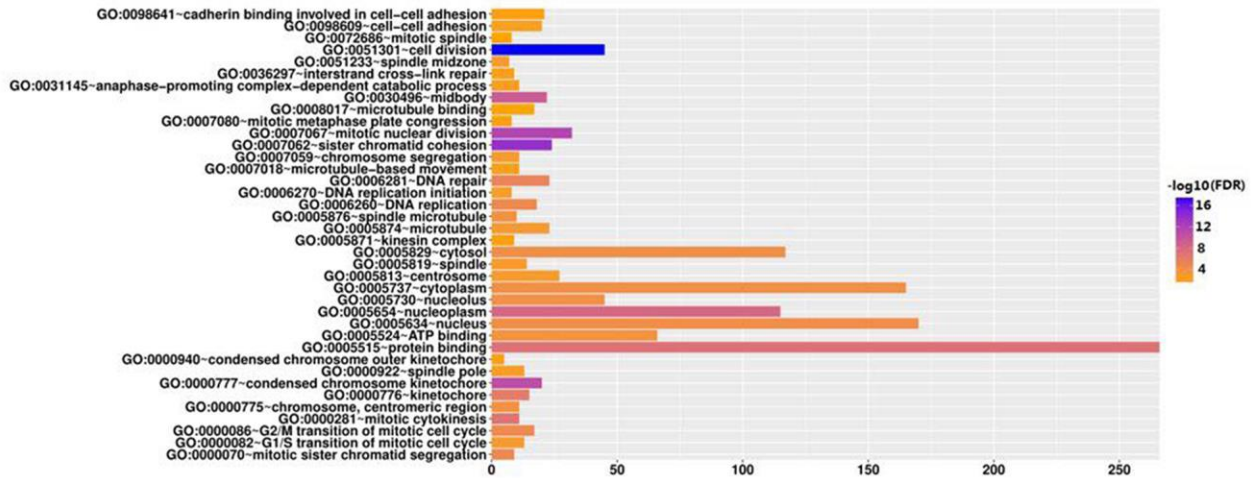


Supplementary Figure 2. Protein-coding genes co-expressed with HULC lncRNA enriched by significant KEGG pathways and GO-based functional enrichment analysis. (A) The top 15 biological processes (BP), top 15 cellular components (CC) and top 15 molecular functions (MF) enriched among 1429 DPCG targets of HULC lncRNA. (B) Top 10 KEGG pathways for HULC lncRNA-related DPCGs. (C) Number of target genes enriched (the horizontal axis represented the corresponding -Log10 (FDR) of different paths and the vertical axis denoted the path name; bp: Biology Process; cc: Cellular Components; mf: Molecular Function).

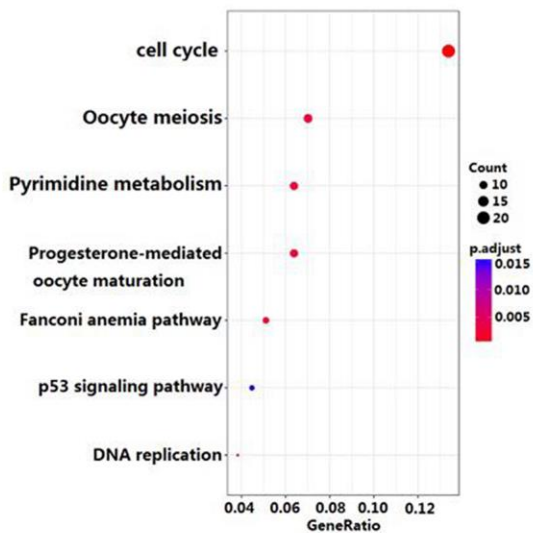


Supplementary Figure 3. Protein-coding genes co-expressed with AL359715.5 lncRNA enriched by significant KEGG pathways and GO-based functional enrichment analysis. (A) The top 15 biological processes (BP), top 15 cellular components (CC) and top 15 molecular functions (MF) enriched among 1440 DPCG targets of AL359715.5 lncRNA. (B) Top 10 KEGG pathways for AL359715.5 lncRNA-related DPCGs. (C) Number of target genes enriched (the horizontal axis represented the corresponding $-\log_{10}$ (FDR) of different paths and the vertical axis denoted the path name; bp: Biology Process; cc: Cellular Components; mf: Molecular Function).

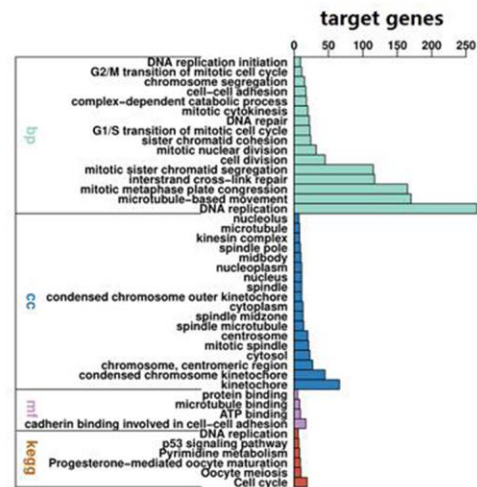
A



B

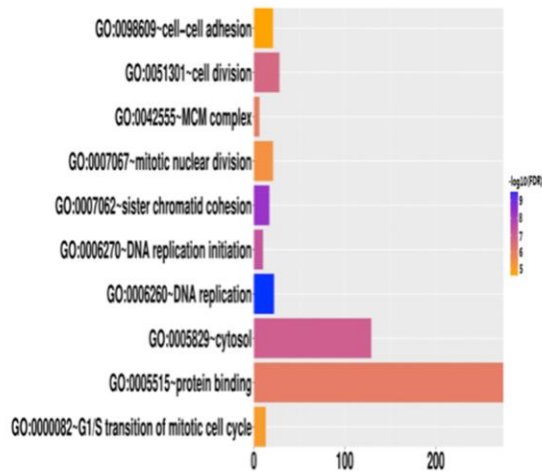


C

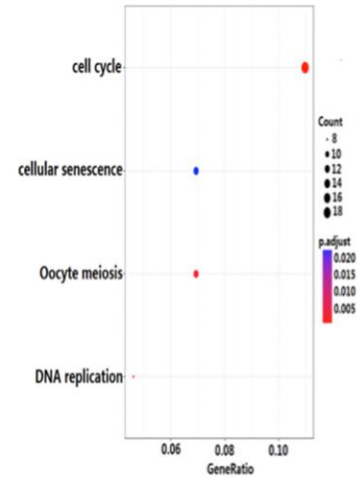


Supplementary Figure 4. Protein-coding genes co-expressed with AC006504.8 lncRNA enriched by significant KEGG pathways and GO-based functional enrichment analysis. (A) The 38 GO processes enriched among 495 DPCG targets of AC006504.8 lncRNA. (B) The 6 KEGG pathways for AC006504.8 lncRNA-related DPCGs. (C) Number of target genes enriched (the horizontal axis represented the respective $-\log_{10}(\text{FDR})$ of different paths and the vertical axis denoted the path name; bp: Biology Process, cc: Cellular Components; mf: Molecular Function).

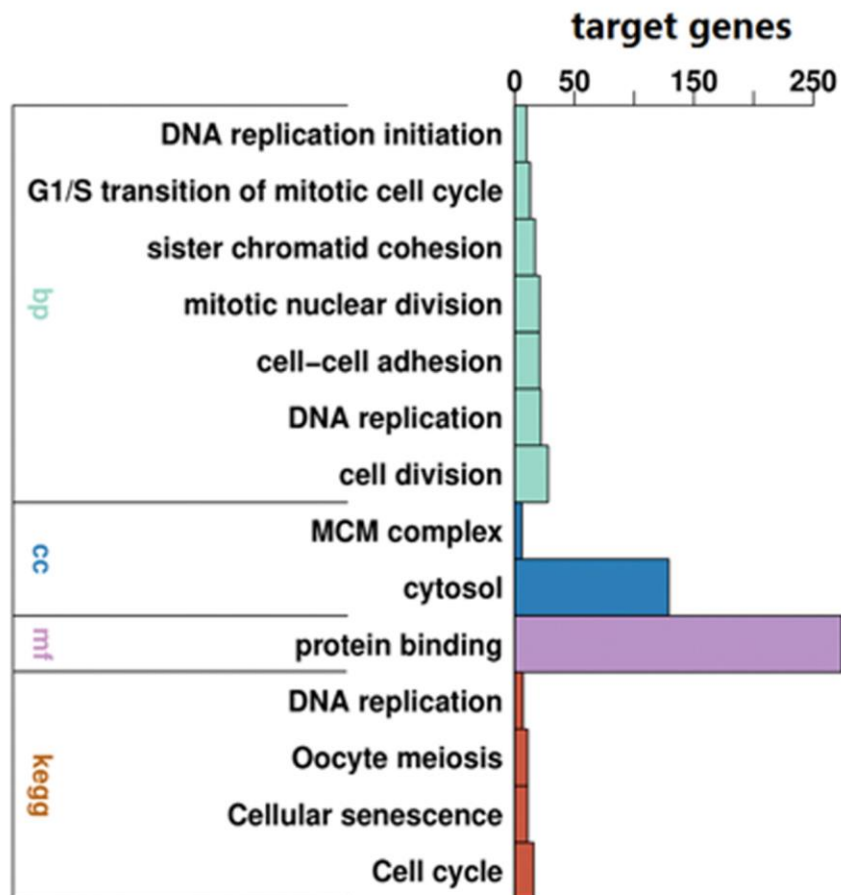
A



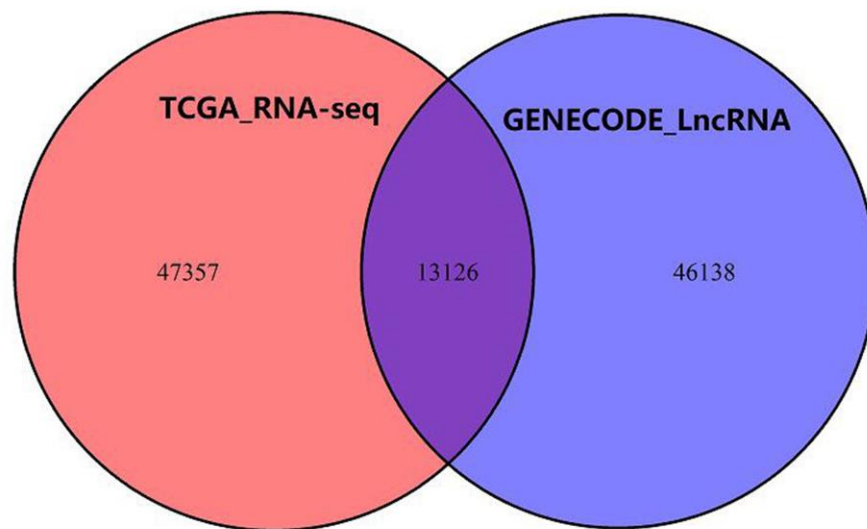
B



C



Supplementary Figure 5. Protein-coding genes co-expressed with AC090114.2 lncRNA enriched by significant KEGG pathways and GO-based functional enrichment analysis. (A) The 10 GO processes enriched among 552 DPCG targets of AC090114.2 lncRNA. (B) The 4 KEGG pathways for AC090114.2 lncRNA-related DPCGs. (C) Number of target genes enriched (the horizontal axis represented the respective $-\log_{10}$ (FDR) of different paths and the vertical axis denoted the path name; bp: Biology Process, cc: Cellular Components, mf: Molecular Function).



Supplementary Figure 6. After 60483 RNA-ENSG_ID and 59264 lncRNA-ENSG_ID were intersected to obtain 13126 lncRNA-Gene_name for this study.

Supplementary Tables

Supplementary Table 1. KEGG pathway analysis of predicted targets from the 5-lncRNA.

Pathway	Count	Genes	FDR
hsa04110:Cell cycle	10	BUB1; CCNB2; CDC20; CDC45; MAD2L1; MCM4; ORC6; PKMYT1; PLK1; PTTG1	P<0.001
hsa04114:Oocyte meiosis	7	BUB1; CCNB2; CDC20; PKMYT1; PLK1; PTTG1; MAD2L1	0.002990543
hsa03460:Fanconi anemia pathway	5	BLM; FAAP24; FANCG; FANCI; RAD51	0.002990543
hsa03030:DNA replication	4	FEN1; MCM4; POLA2; RFC4	0.005179535
hsa04914:Progesterone-mediated oocyte maturation	5	PKMYT1; PLK1; BUB1; CCNB2; MAD2L1	0.027335116

Supplementary Table 2. Results of high- and low-risk score groups gene set enrichment analysis (GSEA) analysis of the predicted core target genes.

Risk score groups	Pathway	Core enrichment (YES) genes	p-value	FDR
Low risk	KEGG_T_CELL_RECEPTOR_SIGNALING_PATHWAY	LAT; MAPK3; JUN; IL10; CTLA4; FOS; NFKBIE; CBL; RAF1; CARD11; CHP2; CD4; CD3G; RASGRP1; CSF2	<0.001	0.014
	KEGG_NEUROTROPHIN_SIGNALING_PATHWAY	MAPK10; NTF4; NFKBIE; RAF1; RIPK2; YWHAZ; MAP3K5; SH2B2; MAPK3; JUN; BCL2; NTRK3	<0.001	0.029
	BIOCARTA_KERATINOCYTE_PATHWAY	RAF1; MAP3K5; MAPK3; JUN; TNFRSF1B; BCL2; FOS	0.002	0.007
	BIOCARTA_IL2RB_PATHWAY	SYK; SOCS3; CBL; RAF1; JAK3; IL2RG; MAPK3; BCL2; FOS	0.008	0.014
High risk	BIOCARTA_COMP_PATHWAY	C3; C4B; MASP2; C9; C6; C5; C2; C4A; C8A; CFB; C1S; C1R; MBL2; MASP1	0.010	0.034

Supplementary Table 3. The clinicopathological characteristics of the TCGA cohort and WMU cohort.

Clinicopathological characteristics	Discovery cohort (n=36)	Validation cohort (n=90)	χ^2	p-value
Age			0.59	0.441
<60	9	33		
≥ 60	22	57		
Gender			0.20	0.652
Male	20	46		
Female	16	44		
Stage			0.79	0.375
Stage I-II	23	59		
Stage III-IV	8	31		
Grade			0.31	0.578
Grade 1-2	10	40		
Grade 3-4	19	50		
Relative family cancer history			0.02	0.900
Yes	19	43		
No	10	24		
Primary pathology residual tumor			1.6	0.205
R1	4	18		
R0	25	53		
BMI			1.21	0.271
>25	23	57		
≤ 24.9	8	33		
Recurrence			0.60	0.440
Yes	15	52		
No	10	24		
History Hepatoma risk factors			1.82	0.177
Yes	14	45		
No	21	39		
Cancer status			1.57	0.21
With tumor	17	64		
Tumor free	12	26		
Primary pathology histological type			0.66	0.416
Intrahepatic	25	66		
Other	6	24		
Postoperative radiotherapy			0.47	0.495
Yes	8	24		
No	20	43		
Five lncRNA Risk score			1.05	0.306
High	18	54		
Low	18	36		