Large-scale analyses identify a cluster of novel long noncoding RNAs as potential competitive endogenous RNAs in progression of hepatocellular carcinoma

Mengjia Song[1,2,*], Ailin Zhong[4,*], Jieying Yang[1,2,*], Junyi He[1,2,*], Shaoyan Cheng[3], Jianxiong Zeng[1,2], Yue Huang[1,2], Qiuzhong Pan[1,2], Jingjing Zhao[1,2], Ziqi Zhou[1,2], Qian Zhu[1,2], Yan Tang[1,2], Hao Chen[1,2], Chaopin Yang[1,2], Yuan Liu[1,2], Xiaocong Mo[1,2], Desheng Weng [1,2], Jian-Chuan Xia[1,2]

[1]Department of Biotherapy, Sun Yat-sen University Cancer Center, Guangzhou, P. R. China
[2]Collaborative Innovation Center for Cancer Medicine, State Key Laboratory of Oncology in South China, Sun Yat-sen University Cancer Center, Guangzhou, China
[3]Biotherapy Center, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China
[4]Office of International Exchange and Cooperation, Guangzhou University of Chinese Medicine, Guangzhou, People's Republic of China
*Equal contribution

Correspondence to: Jian-Chuan Xia, Desheng Weng; email: xiajch@sysucc.org.cn, wengdsh@sysucc.org.cn
Keywords: CeRNA, lncRNA, hepatocellular carcinoma, progression
Received: July 23, 2019   Accepted: November 8, 2019   Published: November 23, 2019

ABSTRACT

The abnormal expression of noncoding RNAs has attracted increasing interest in the field of hepatocellular carcinoma progression. However, the underlying molecular mechanisms mediated by noncoding RNAs in these processes are unclear. Here, we obtained the expression profiles of long noncoding RNAs, microRNAs, and mRNAs from the Gene Expression Omnibus database and identified hepatocarcinogenesis-specific differentially expressed transcripts. Next, we identified significant Gene Ontology and pathway terms that the differentially expressed transcripts involved in. Using functional analysis and target prediction, we constructed a hepatocellular carcinoma-associated deregulated competitive endogenous RNA network to reveal the potential mechanisms underlying tumor progression. By analyzing The Cancer Genome Atlas dataset, six key long noncoding RNAs showed significant association with overall survival as well as strong correlation with some microRNAs and mRNAs in the competitive endogenous RNA network. We further validated the above results and determined their diagnostic and prognostic value in clinical samples. Importantly, by large-scale analyses, we identified a cluster of long noncoding RNAs, *GBAP1*, *MCM3AP-AS1*, *SLC16A1-AS1*, *C3P1*, *DIO3OS*, and *HNF4A-AS1* as candidate biomarkers for the diagnosis and prognosis of hepatocellular carcinoma, which will improve our understanding of competitive endogenous RNA-mediated regulatory mechanisms underlying hepatocellular carcinoma development and will provide novel therapeutic targets in the future.

INTRODUCTION

Hepatocellular carcinoma (HCC) is one of the fatal malignant tumors worldwide, especially in Asia [1]. Despite advances in prevention, detection, diagnosis, and treatment of HCC in recent years, its incidence and mortality are increasing significantly every year, and the five-year overall survival (OS) rate is only 18% [2]. The development of HCC is a complex multistep process involving a series of molecular pathogeneses and multiple factors; therefore, there is an urgent need to identify more detailed mechanisms to improve the

outcome of patients with HCC [3]. Recent genomic researches have identified many oncogenes as indispensable factors involved in the development of various types of cancer, including HCC, which provide more advanced diagnostic approaches and new targets for cancer treatment [4, 5].

Long non-coding RNAs (lncRNAs) comprise transcripts with a length of over 200 nucleotides, which are characterized by low expression in cancer, high expression in tissues, and cell-specificity [6]. Despite their inability to encode proteins, lncRNAs regulate the expression of many mRNAs by acting in cis and in trans. Cis-acting lncRNAs influence the expression and chromatin state of nearby genes via altering their transcription, recruiting regulatory factors, and splicing of the lncRNA. They also rely on DNA elements within the lncRNA promoter or gene locus. However, trans-acting lncRNAs leave the site of transcription and execute an array of functions throughout the cell by influencing nuclear structure and organization, as well as regulating the behavior of proteins and other RNA molecules [7]. As a result, the dysregulation of lncRNA expression affects cellular homeostasis, which might lead to cancer initiation and progression [6, 8]. MicroRNAs (miRNAs) are a type of small non-coding RNAs composed of 21–22 nucleotides. They exert their biological effects by silencing genes post-transcriptionally via binding to miRNA response elements (MREs) in the target mRNA [9]. Recently, the theory of a competing endogenous RNA (ceRNA) regulation network in cancer has been proposed [10]. This hypothesis states that ceRNAs harbor MREs and bind to miRNAs in competition with their target mRNAs, leading to blockade of the silencing effect of miRNAs on their target mRNAs. Accumulating studies have confirmed that lncRNAs act as sponges to sequester and bind miRNAs in competition with mRNAs [11]. Therefore, lncRNA can be considered as a kind of ceRNA that regulates transcript expression. This theory has been proposed in different types of cancer. For example, Liang et al. constructed a lncRNA-mediated ceRNA network for mesenchymal ovarian cancer and showed that lncRNA *PTAR* acted as a ceRNA to promote epithelial-mesenchymal transition (EMT), invasion, and metastasis by competitively binding to miR-101-3p to regulate *ZEB1* expression [12]. Another study identified transforming growth factor-beta (TGF-β) promoted tumor invasion and metastasis by downregulating *EPB41L4A-AS2*, a novel lncRNA functioning as a ceRNA in head and neck squamous cell carcinoma [13]. In HCC, Sui et al. reported that lncRNA *GIHCG* promoted HCC cell proliferation, migration, and invasion *in vitro* and xenograft growth and metastasis *in vivo* depending on its silencing of miR-200b/a/429 [14]. Ren et al. proposed the TP53-

miR-215-PCAT-1-CRKL axis as an important regulatory pathway inhibiting tumor cell proliferation, migration, and invasion in HCC [15]. However, whether the ceRNA network mediates HCC initiation and progression remains unclear.

The present study aimed to reveal the potential regulatory mechanisms involved in HCC development by constructing an mRNA-miRNA-lncRNA interaction network. We comprehensively analyzed the regulatory network of HCC-related genes, miRNAs, and lncRNAs based on the gene expression profile, including differential expression profiles analysis, gene ontology (GO) enrichment, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis, and signal regulation network (signal-net) construction. Based on these results and target transcripts prediction, we built a ceRNA network to indicate the key mRNA-miRNA-lncRNA interactions. Furthermore, a protein regulation network was used to clarify the regulatory interactions among these transcripts in the ceRNA network. For validation, we performed Pearson correlation analysis and survival analysis for the mRNAs, miRNAs, and lncRNAs identified in the ceRNA network in HCC tissues from The Cancer Genome Atlas (TCGA) database and clinical samples, respectively. The identified ceRNA network might provide potential biomarkers for predicting the prognosis and novel therapeutic targets for the treatment of patients with HCC.

RESULTS

Differential and clustering analysis

The workflow chart was shown in Supplementary Figure 1. We first identified 11036 differentially expressed genes (DEGs) from three profiles, GSE29721, GSE40367, and GSE62232, consisting of 3949 (35.8%) upregulated and 7087 (64.2%) downregulated genes. The top 5 DEGs in the HCC samples compared with normal samples from different profiles were shown in Table 1 according to their fold change (FC) values. Besides, 3826 lncRNAs were selected as differentially expressed lncRNAs (DELs) from the same three profiles, including 615 (16.1%) upregulated and 3211 (83.9%) downregulated lncRNAs. From another two profiles, GSE36915 and GSE74618, 206 differentially expressed miRNAs (DEMs) were selected, including 80 (38.8%) upregulated and 126 (61.2%) downregulated miRNAs. The top 10 DELs and DEMs in the HCC samples in contrast to normal samples from different profiles were shown in Supplementary Tables 1 and 2, respectively. The results of hierarchical cluster analyses were shown in heatmaps for the expression level changes of DEGs (Figure 1A), DELs (Figure 1B), and DEMs (Figure 1C) between HCC and normal tissues.

Table 1. The top 5 upregulated and downregulated DEGs in different expression profiles.

Profile	Gene symbol	Gene ID	style	Fold change	*P*-value
GSE29721	CCL20	6364	up	30.93373	0.000163
	HJURP	55355	up	23.79123	0.00002
	SPINK1	6690	up	22.69303	0.000635
	SULT1C2	6819	up	20.0141	0.0001927
	REG3A	5068	up	18.2934	0.005302
	MT1M	4499	down	-30.4032	0.001046
	TTC36	143941	down	-28.8462	0.000253
	THRSP	7069	down	-26.6209	0.0003503
	HAMP	57817	down	-26.0083	0.001316
	CLEC4M	10332	down	-22.2482	0.000288
GSE29721	CCL20	6364	up	106.6007	0.000019
	SPINK1	6690	up	47.69886	0.000022
	GABBR1	2550	up	43.14953	0.000018
	KIF20A	10112	up	36.35204	0.000019
	ELOVL7	79993	up	34.84073	0.00002
	CXCL14	9547	down	-132.845	0.0000185
	CYP1A2	1544	down	-101.348	0.000018
	CNDP1	84735	down	-87.8993	0.000018
	MT1M	4499	down	-62.8525	0.00002
	MME	4311	down	-60.0369	0.0000265
GSE62232	RPS4Y1	6192	up	90.17847	<0.00001
	AKR1B10	57016	up	34.80468	<0.00001
	CCL20	6364	up	26.06746	0.000001
	TOP2A	7153	up	20.6209	9.933E-05
	SPINK1	6690	up	18.16443	0.001901
	MT1M	4499	down	-75.7478	<0.00001
	CNDP1	84735	down	-70.9896	<0.00001
	CLEC4G	339390	down	-69.8899	<0.00001
	CXCL14	9547	down	-67.7081	3.333E-06
	CLEC4M	10332	down	-67.278	<0.00001

Intersection analysis

Next, we selected 1016 intersecting DEGs and 116 intersecting DELs among the three profiles to perform further analysis. Twenty-one DEMs were also selected between the two profiles GSE36915 and GSE74618. The intersections were displayed as Venn diagrams (Supplementary Figure 2A–2C).

GO and pathway analysis

According to the intersecting DEGs, we identified 1803 upregulated DEGs-related GO terms and 6339 downregulated DEGs-related GO terms using functional enrichment analysis. The plots of the top 25 upregulated and downregulated GO enrichment terms were shown in Figure 2A and 2B. The results demonstrated that the upregulated DEGs mainly participated in cell division, sister chromatid cohesion, mitotic spindle organization, DNA replication, and mitotic cell cycle; while the downregulated DEGs were closely associated with the oxidation-reduction process, xenobiotic metabolic process, and the epoxygenase P450 pathway.

We subsequently distinguished 221 pathways among the upregulated DEGs and 274 pathways among the downregulated DEGs using KEGG pathway analysis based on the intersecting DEGs. The plots of the top 25 pathways for the upregulated and downregulated DEGs were shown in Figure 2C and 2D. The most obviously upregulated pathways were the cell cycle, the p53 signaling pathway, and DNA replication. Pathways

dramatically enriched in downregulated genes were metabolic pathways, chemical carcinogenesis, and retinol metabolism. Taken together, these dramatically changed GO terms and pathways might be involved in HCC initiation and progression.

Signal-net analysis

To clarify the interactions among different genes and their products, we further constructed a gene signal-net using 1016 intersecting DEGs based on KEGG database according to the network biology theory. Signal-net could display the relationships between different gene groups and identify upstream and downstream molecules by obtaining gene interactions in multiple pathways. The network we constructed included 97 interacting DEGs, consisting of 23 upregulated genes and 74 downregulated genes (Supplementary Figure 3). According to their regulation degree, several DEGs were considered as the hub genes that exerted the most significant regulatory function in the network. For example, *PRKAA2*, which held the highest degree, had the strongest interactions with other genes in the network. The pathway information from the KEGG database indicated that *PRKAA2* and its related genes were involved in the AMPK/PI3K/AKT signaling
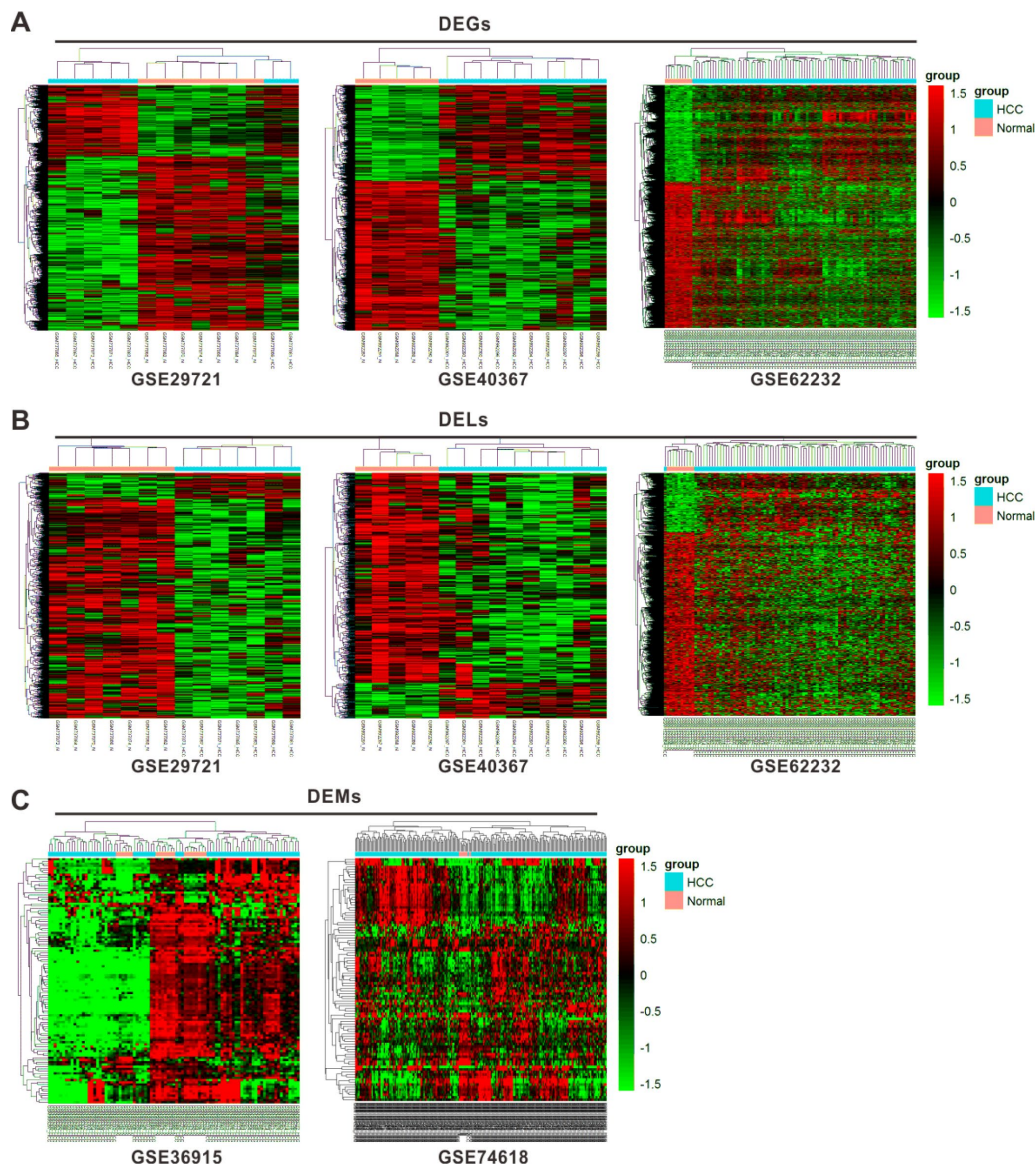

Figure 1. Cluster analysis of differentially expressed profiles. Hierarchical cluster dendrogram of DEGs (**A**) and DELs (**B**) identified in GES29721, GSE40367 and GSE62232 as well as DEMs (**C**) identified in GSE36915 and GSE74618. The rows showed DEGs, DELs, and DEMs, while the columns showed paired samples. The pink part represents normal samples and the blue part represents HCC samples. The left vertical axis shows clusters of DEGs, DELs and DEMs, while the above horizontal axis shows clusters of samples. Red represents high expression and green represents low expression.

pathways, a representative signaling pathway involved in the development of HCC [16]. Besides, several closely connected hub genes, including *ITGA6*, *ITGA2*, and *ITGB3*, are the main cellular adhesion receptors belonging to the integrins family and extensively implicated in multiple steps from cancer initiation to metastasis by acting as signaling molecules, mechano-transducers, and critical components of the cell migration machinery [17]. Therefore, DEGs with a high regulation degree in this signal-net, including *PRKAA2*, *PLCB1*, and several genes encoding the integrins family members might play crucial roles in the regulation of HCC development.

CeRNA network

The above results demonstrated that some intersecting DEGs might play a crucial role in HCC development by participating in important processes, such as cell cycle, p53 signaling pathway, and metabolic pathways. We next investigated the underlying molecular mechanisms regulating these processes. LncRNAs have been reported to promote tumor progression through various types of gene regulatory mechanisms, such as epigenetic and transcriptional regulation and serving as ceRNAs for miRNAs [10, 11, 14, 15, 18–20]. Therefore, we constructed a ceRNA regulatory network to reveal the unknown mechanisms driving HCC development.

Intersection datasets were acquired between the DEGs involved in the significant enriched GO terms and KEGG pathways with $P < 0.05$ and false discovery rate (FDR) < 0.05, which included 393 DEGs. We predicted 213 target genes and 85 target lncRNAs that might be regulated by the 21 intersected DEMs among the 393 DEGs and the 116 intersecting DELs, respectively. The key lncRNA-miRNA and miRNA-mRNA pairs were shown in Supplementary Tables 3 and 4, respectively. According to the constructed association and the theory of
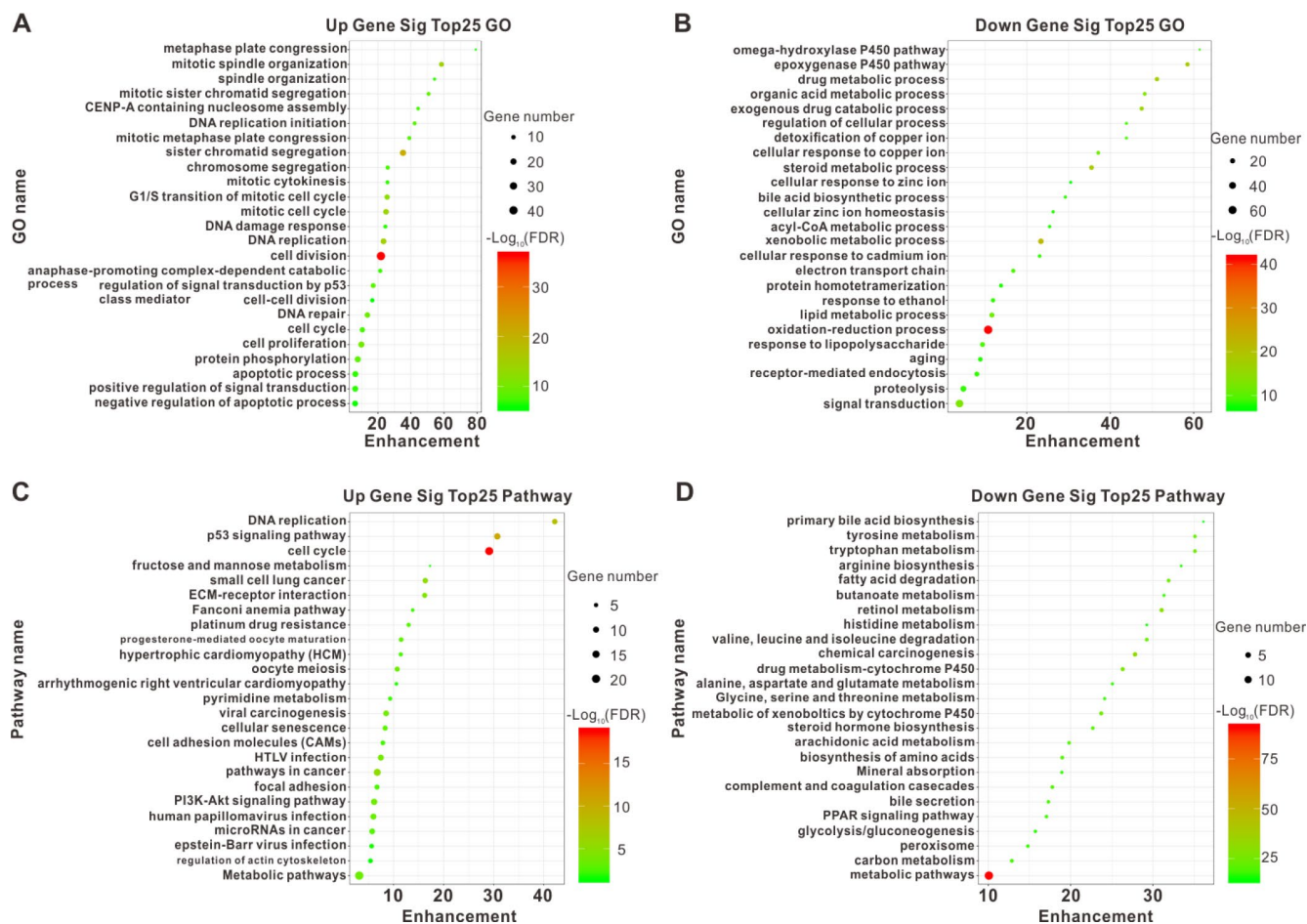

Figure 2. Top25 enrichment of GO and KEGG pathway analyses for upregulated and downregulated DEGs. The upregulated (**A**) and downregulated (**B**) DEGs enriched in GO categories. The upregulated (**C**) and downregulated (**D**) DEGs enriched in different pathways. The horizontal axis represents the enrichment score of DEGs. The vertical axis represents different GO categories and pathways. The bubble size indicates the number of genes in each category and pathway, and different colors correspond to different log (FDR) values.

ceRNAs, we chose the negatively correlated lncRNA-miRNA and miRNA-mRNA pairs to build an mRNA-miRNA-lncRNA network using Cytoscape v3.0 (Figure 3). As a result, 59 lncRNAs were identified as ceRNAs that interacted with 17 essential miRNAs, and subsequently indirectly regulated 63 coding mRNAs involved in HCC development.

Relying on the constructed relationships in the ceRNA network, we inferred that lncRNAs might indirectly participate in several significant KEGG pathways by serving as ceRNAs of mRNAs, such as the PI3K-AKT signaling pathway enriched by *ITGA2*, *PRKAA2*, and *CDK1*, the cell cycle enriched by *MCM5*, *CDC6*, and *CHEK1*, as well as ECM-receptor interactions enriched by *ITGA6* and *ITGA2*. Similarly, we also inferred that lncRNAs participated in some significant biological processes, including cell division enriched by *BIRC5* and *CDCA5*, as well as cell proliferation enriched by *CD34* and *ITGA2*. Indeed, some miRNA-mRNA and lncRNA-miRNA pairs in this network have been verified to play a part in cancer development in previous studies. For instance, Jafarzadeh et al. provided experimental evidences for hsa-miR-497-5p as a negative regulator of SMAD3, which was a key modulator of the TGF-β signaling pathway during carcinogenesis [21]. Among the genes interacting with hsa-miR-497-5p in this network, *BIRC5*, a well-known cancer-related gene encoding survivin, has been proven to be upregulated by TGF-β to modulate the cell cycle and apoptosis in various types of cancer [22, 23]. Therefore, the regulatory relationship between hsa-miR-497-5p and *BIRC5* presented in the ceRNA network was justifiable. Besides, for lncRNAs, *MCM3AP-AS1* has been reported to directly bind to miR-194-5p and act as ceRNA, which subsequently facilitated the expression of miR-194-5p's target gene *FOXA1* in HCC cells, thus promoting HCC cell proliferation, colony formation, and cell cycle progression [24]. Another lncRNA identified in this network, *HAND2-AS1*, has been proven to increase cell migration of HCC cell lines [25].

PPI (protein-protein interaction) network analysis

Subsequently, we built a protein-protein interaction network predicting the interaction among the proteins encoded by the 51 DEGs in the ceRNA network (Supplementary Figure 4). In the network, *ESR1*, *IGF1*, *BIRC5*, and *CD34* had higher degrees (17, 15, 12, and 10, respectively) (Supplementary Table 5). The genes encoding these proteins have been confirmed to be associated with HCC progression [26–29]. Combined with the functional analysis results, we found that the interacting DEGs in the ceRNA network were mainly enriched in cell proliferation, cell adhesion, the PI3K-Akt signaling pathway, the p53 signaling pathway, pathways in cancer, and metabolic pathways.
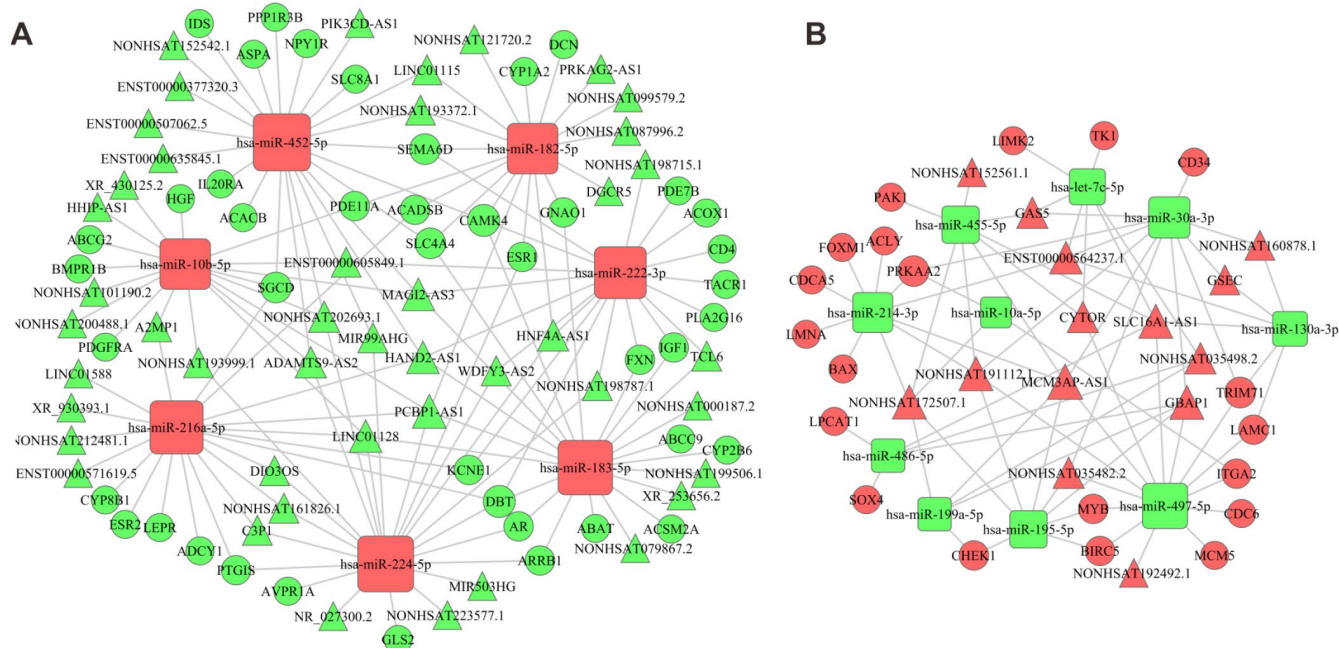

Figure 3. The ceRNA network. (**A**) Network constructed by upregulated miRNAs, downregulated lncRNAs and downregulated mRNAs. (**B**) Network constructed by downregulated miRNAs, upregulated lncRNAs and upregulated mRNAs. Red diamonds represent upregulated miRNAs, red balls, upregulated mRNAs, and red triangles, upregulated lncRNAs. Green diamonds represent downregulated miRNAs, green balls, downregulated mRNAs, green triangles, downregulated lncRNAs.

Survival analysis and expression validation for the ceRNA network in TCGA dataset

In the Gene Expression Omnibus (GEO) datasets, we identified an extensive and comprehensive lncRNA-miRNA-mRNA regulation network in HCC development. To further clarify their expression and prognostic value, we performed Kaplan-Meier survival analysis for all the DEGs, DELs, and DEMs in the ceRNA network in patients with HCC from TCGA. The results showed that 6 lncRNAs (Figure 4), 23 mRNAs, and 4 miRNAs had a significant impact on OS (Supplementary Figure 5) and a consistent expression pattern (Supplementary Figure 6) in the GEO datasets. Among them, 3 lncRNAs (*GBAP1*, *MCM3AP-AS1*, and *SLC16A1-AS1*) (Figure 4A–4C, 4G–4I), 15 mRNAs (*ACLY*, *BAX*, *BIRC5*, *CDC6*, *CDCA5*, *CHEK1*, *FOXM1*, *ITGA2*, LAMC1, *MCM5*, *MYB*, *PAK1*,
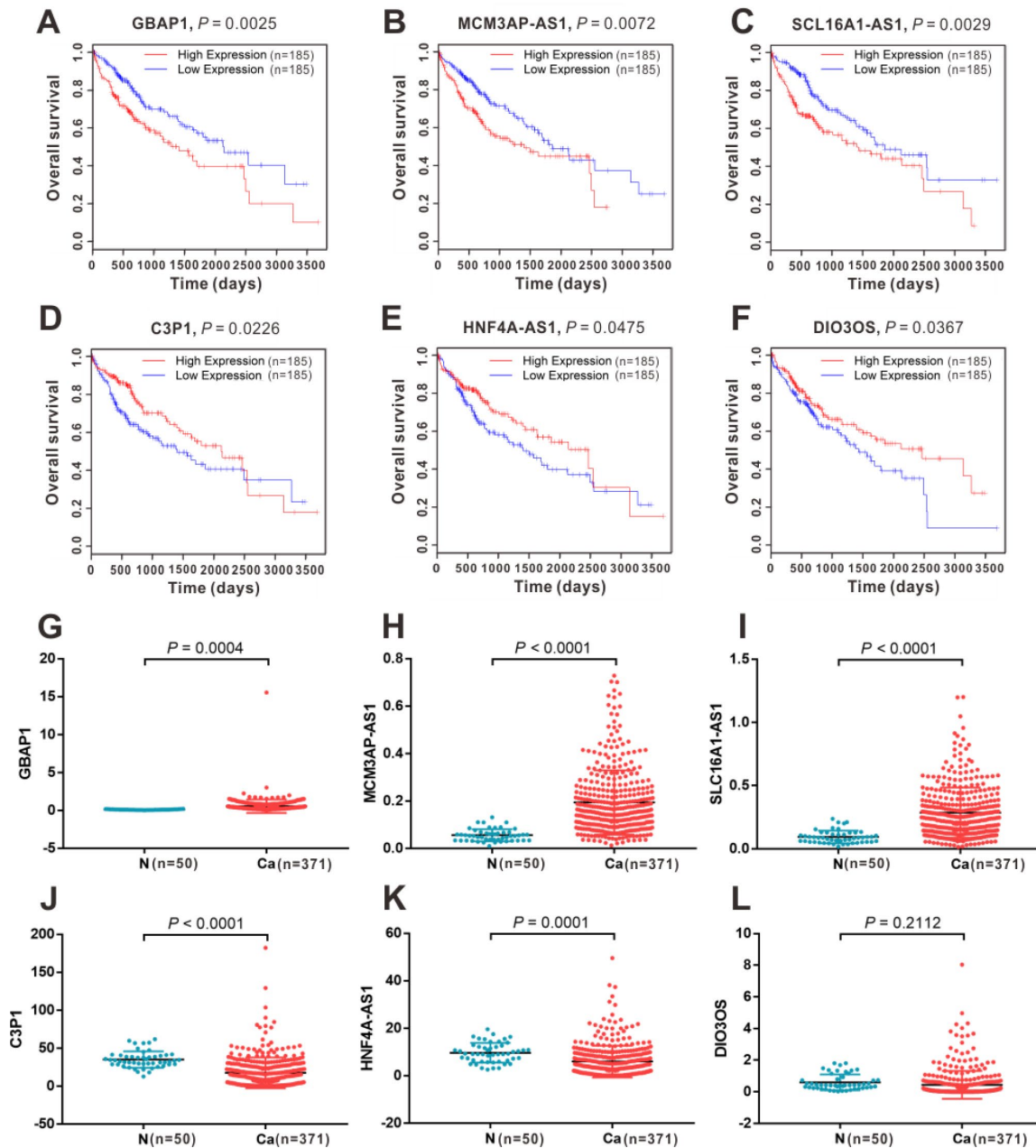

Figure 4. The expression and survival significance of the six key lncRNAs in the ceRNA network in HCC TCGA database. Kaplan-Meier survival curves showed significant OS differences between high- and low-expression of *GBAP1* (**A**), *MCM3AP-AS1* (**B**), *SLC16A1-AS1* (**C**), *C3P1*(**D**), *HNF4A-AS1* (**E**) and *DIO3OS* (**F**) in TCGA dtabase. Expression levels of *GBAP1* (**G**), *MCM3AP-AS1* (**H**), *SLC16A1-AS1* (**I**), *C3P1*(**D**), *HNF4A-AS1* (**E**) and *DIO3OS* (**F**) in normal and HCC tissues from TCGA database. N, normal tissue; Ca, cancer tissue.

PRKAA2, *SOX4*, and *TK1*) (Supplementary Figures 5Aa–5Ao, 6Aa–6Ao) and 3 miRNAs (hsa-miR-10b-5p, hsa-miR-183-5p, and hsa-miR-222-3p) (Supplementary Figures 5Aa–5Ao, 6Ba–6Bc) were identified as pro-tumor factors because of their high expression in cancer tissues and their correlation with shorter OS in patients with HCC. In contrast, another three lncRNAs (*C3P1*, *DIO3OS*, *and HNF4A-AS1*) (Figure 4D–4F, 4J–4L), eight mRNAs (*ABAT*, *ACSM2A*, *ASPA*, *CAMK4*, *CYP8B1*, *ESR1*, *IGF1*, and *PDE7B*) (Supplementary Figure 5Ca–5Ch, 6Ca–6Ch) and one miRNA (hsa-let-7c-5p) (Supplementary Figures 5D, 6D) showed low expression in cancer tissues and correlated with longer OS, implying that these transcripts might be protective factors in HCC.

According to the interactions in the ceRNA network, we inferred that the aberrant expression of DEGs might be indirectly regulated by six lncRNAs (*GBAP1*, *MCM3AP-AS1*, *SLC16A1-AS1*, *C3P1*, *DIO3OS*, and *HNF4A-AS1*). To further determine the association of the identified lncRNAs with survival time, we conducted univariate and multivariate Cox regression model analyses based on the

clinical characteristics in TCGA dataset. As shown in Table 2, in the univariate analysis, all six lncRNAs could be incorporated in the COX regression model. Multivariate analysis demonstrated that the expression levels of *lncRNAs C3P1*, *DIO3OS*, and *SLC16A1-AS1* were independent prognostic factors for OS in patients with HCC. Thus, it was reasonable to infer that these lncRNAs might be crucial factors to predict the prognosis of patients with HCC.

Correlation analysis in TCGA dataset

Since the main aim of this study was identifying the clinical noteworthy lncRNAs and uncovering the regulatory role of lncRNAs as potential ceRNAs to mediate downstream/upstream RNAs functions/associations, we selected the six lncRNAs with great survival significance in the ceRNA network for further study. To validate the regulatory role of the six lncRNAs as ceRNAs, we analyzed the correlation with their associated miRNAs and mRNAs in the ceRNA network based on TCGA data containing 366 patients
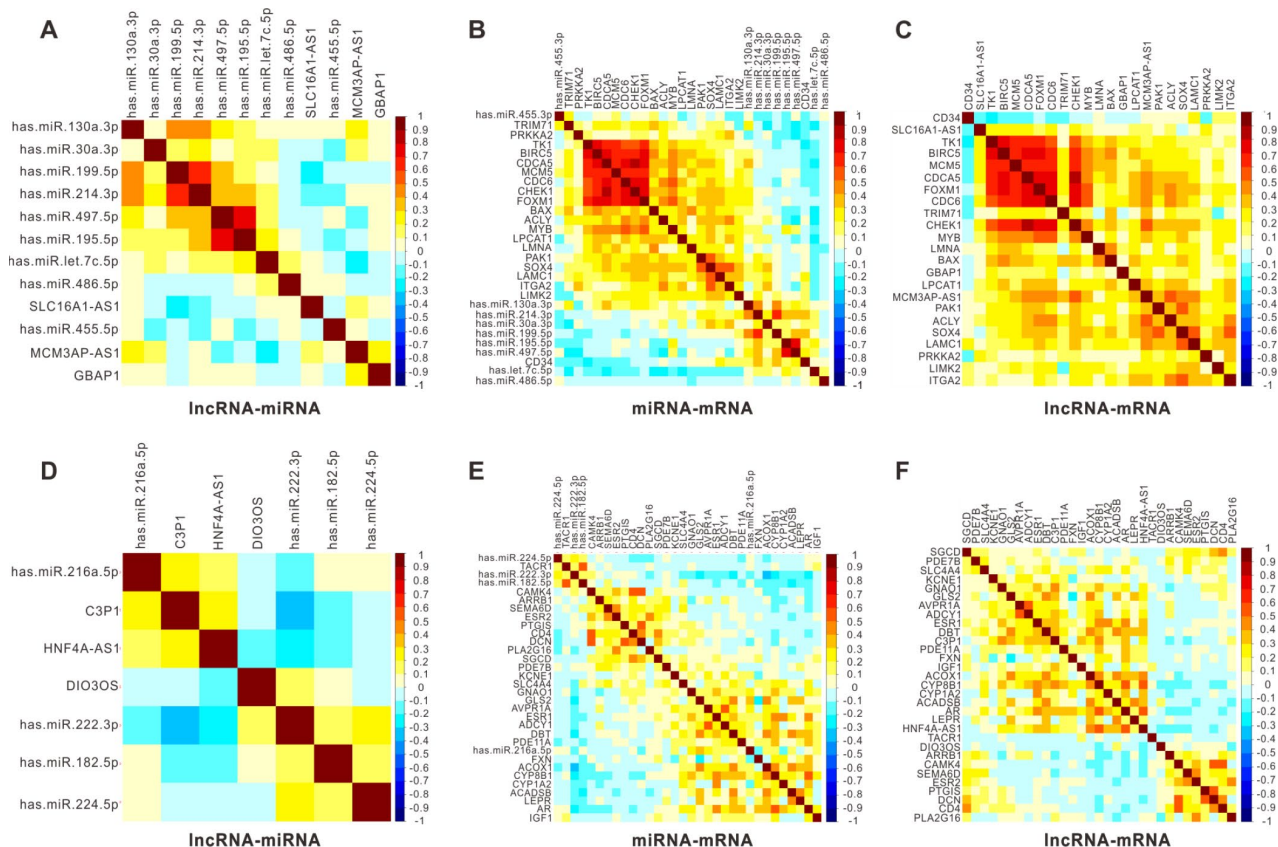

Figure 5. Pearson correlation analysis for six key lncRNAs with their associated miRNA and mRNAs in 366 HCC patients from TCGA. Pearson correlograms of lncRNA-miRNA paris (**A**), lncRNA-mRNA pairs (**B**) and mRNA-miRNA pairs (**C**) in upregulated lncRNAs including *GBAP1*, *MCM3AP-AS1* and *SLC16A1-AS1*. lncRNA-miRNA paris (**D**), lncRNA-mRNA paris (**E**) and mRNA-miRNA paris (**F**) in downregulated lncRNAs including *C3P1*, *HNF4A-AS1* and *DIO3OS* by Pearson correlation analysis. The correlation coefficient R value ranges from -1 to 1, the color of which changes from blue to brown.

with HCC. As shown in Figure 5, a considerable number of lncRNA-miRNA pairs showed a significant negative correlation for both upregulated (Figure 5A) and downregulated (Figure 5D) lncRNAs. Some of the six lncRNAs-associated miRNA-mRNA pairs also demonstrated a negative correlation (Figure 5B and 5E). Moreover, almost all six lncRNAs-associated lncRNA-mRNA pairs, for either upregulated (Figure 5C) or downregulated (Figure 5F) lncRNAs, demonstrated a tight positive correlation in TCGA database. Together, these results indicated the proposed six lncRNAs, *GBAP1*, *MCM3AP-AS1*, *SLC16A1-AS1 C3P1*, *DIO3OS*, and *HNF4A-AS1*, had a partially negative correlation with their associated miRNAs, as well as a partially positive correlation with their associated mRNAs, further confirming that the six might function as ceRNAs to regulate the expression of mRNAs and miRNAs in the progression of HCC.

Expression validation and survival analysis for the six lncRNAs in clinical samples

To further verify the significance of the six lncRNAs mentioned in section 2.8, we analyzed their expression levels in tumor tissues and adjacent non-tumor tissues

from 158 diagnosed patients with HCC using quantitative real-time reverse transcriptase-polymerase chain reaction (RT-PCR). As shown in Figure 6A–6F, all six lncRNAs were differentially expressed in tumor tissues and normal tissues. Three lncRNAs (*GBAP1*, *MCM3AP-AS*1, and *SLC16A1-AS1*) were upregulated in HCC tissues (Figure 6A–6C), while the other three lncRNAs (*C3P1*, *HNF4A-AS1*, and *DIO3OS*) were downregulated in HCC tissues (Figure 6D–6F). These results were consistent with the previous bioinformatic analysis. Furthermore, based on the clinical information, we analyzed the relationship between the expression levels of these DELs and OS, progression-free survival (PFS) and distant metastasis-free survival (DmFS) in these patients, which was demonstrated using Kaplan-Meier curves in Figure 6G–6X. Consistent with the results in TCGA, the upregulated lncRNAs, including *GBAP1*, *MCM3AP-AS1*, and *SLC16A1-AS1*, were associated with worse OS (Figure 6G–6I), PFS (Figure 6M–6O) and DmFS (S-U), thus identifying them as pro-tumor factors. Inversely, the downregulated lncRNAs, including *C3P1*, *HNF4A-AS1*, and *DIO3OS*, were correlated with the better OS (Figure 6J–6L), PFS (Figure 6P–6R) and DmFS (Figure 6V–6X), and thus represented protective factors.
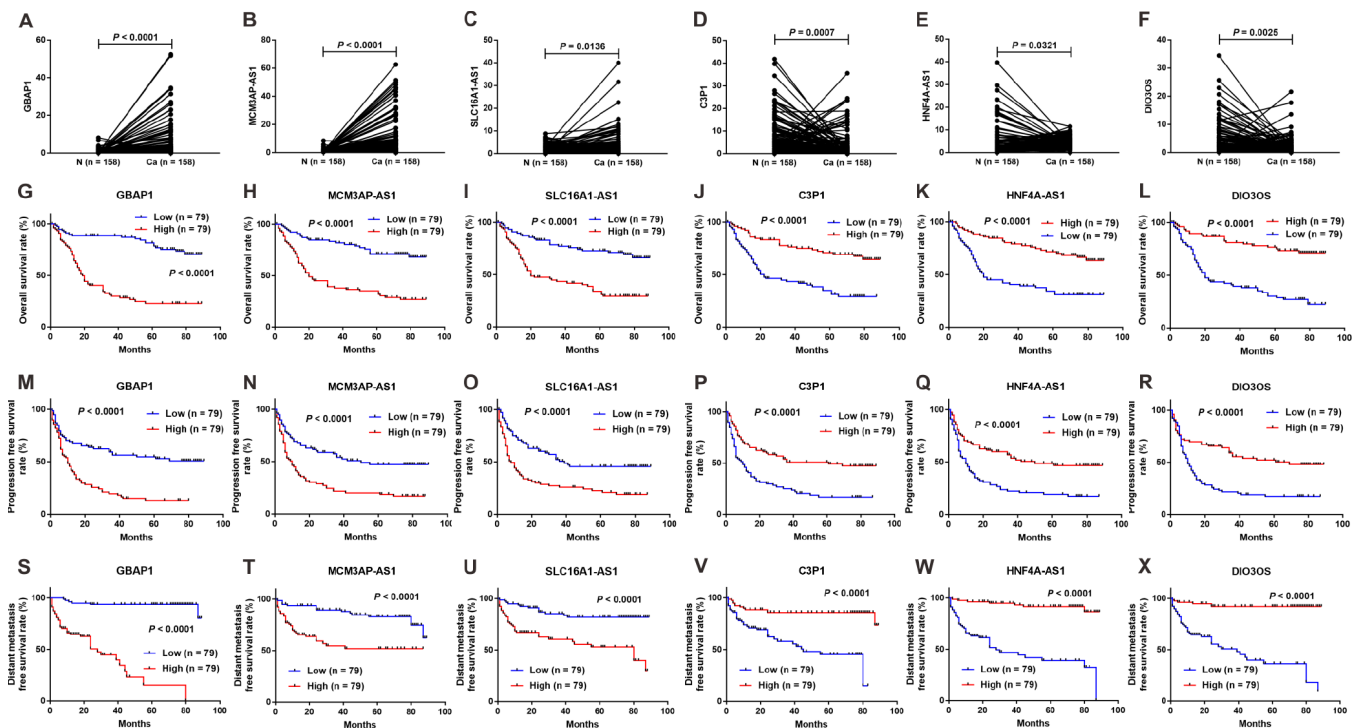

Figure 6. Expression and survival sifnificance for the six lncRNAs in 158 clinical samples. Expression levels of lncRNAs *GBAP1* (**A**), *MCM3AP-AS1* (**B**), *SLC16A1-AS1* (**C**), *C3P1* (**D**), *HNF4A-AS1* (**E**), and *DIO3OS* (**F**) in 158 HCC tissues and adjacent normal tissues. Kaplan-Meier survival curves showed significant OS differences between high- and low-expression of lncRNAs including *GBAP1* (**G**), *MCM3AP-AS1* (**H**), *SLC16A1-AS1* (**I**), *C3P1*(**J**), *HNF4A-AS1* (**K**), and *DIO3OS* (**L**), and significant PFS differences including *GBAP1* (**M**), *MCM3AP-AS1* (**N**), *SLC16A1-AS1* (**O**), *C3P1* (**P**), *HNF4A-AS1* (**Q**) and *DIO3OS* (**R**), as well as significant DmFS differences including *GBAP1* (**S**), *MCM3AP-AS1* (**T**), *SLC16A1-AS1* (**U**), *C3P1* (**V**), *HNF4A-AS1* (**W**) and *DIO3OS* (**X**). N =158, Ca = 158; N, normal tissue; Ca, cancer tissue.

Table 2. Univariate and multivariate analysis of OS in 371 HCC patients from TCGA.

Variable	Univariate cox		Multivariate cox	
	P-value	HR（95%CI）	*P*-value	HR（95%CI）
Age				
≥50 or<50	0.649857	1.11(0.72-1.70)		
Gender				
Male or female	0.259970	1.23(0.86-1.75)		
HBsAg				
Positive or negative	**0.000018**	2.36(1.59-3.49)	**0.000310**	2.18(1.43-3.33)
Liver cirrhosis				
Yes or no	**0.002210**	0.54(0.37-0.80)	**0.030828**	0.64(0.42-0.96)
TNM stage				
III/IV or I/II	**0.000011**	2.23(1.56-3.19)	**0.002199**	1.80(1.24-2.62)
AFP				
≥400ng/ml or<400ng/ml	0.328566	0.79(0.50-1.26)		
Histological differentiation				
Poor or well	0.538730	1.12(0.78-1.61)		
Vascular invasion				
Yes or no	0.866489	1.03(0.70-1.52)		
C3P1 expression				
High or low	**0.023555**	0.67(0.47- 0.95)	**0.017161**	0.64(0.45-0.92)
HNF4A-AS1 expression				
High or low	**0.048667**	0.70(0.50-0.10)	0.123098	
DIO3OS expression				
High or low	**0.037804**	0.69(0.49-0.98)	**0.027333**	0.67(0.47-0.96)
MCM3AP-AS1 expression				
High or low	**0.007812**	1.61(1.13-2.28)	0.384940	
GBAP1 expression				
High or low	**0.003447**	1.68(1.19-2.38)	0.068636	1.40(0.97-2.01)
SLC16A1-AS1 expression				
High or low	**0.003193**	1.69(1.19-2.39)	**0.011557**	1.59(1.11-2.29)

Importantly, combined with the clinical characteristics, we constructed univariate and multivariate Cox proportional hazards regression models for OS, PFS, and DmFS. As shown in Table 3, multivariate Cox's regression analysis revealed that *DIO3OS, MCM3AP-AS*1, and *SLC16A1-AS1* expression levels were independent prognostic factors for OS. *C3P1, MCM3AP-AS1*, and *SLC16A1-AS1* expression levels were independent prognostic factors for PFS, and *HNF4A-AS1, MCM3AP-AS*1, *GBAP1*, and *SLC16A1-AS1* expression levels were independent prognostic factors for DmFS. These data indicated that the proposed six lncRNAs were clinical noteworthy biomarkers for predicting the prognosis and metastasis for patients with HCC.

Correlation validation for the six lncRNAs in clinical samples

To further understand the links of the six lncRNAs with miRNAs or mRNAs in the ceRNA network, we selected lncRNAs associated miRNAs, as well as specific mRNAs with survival significance in TCGA dataset of HCC from the ceRNA network. And their expression levels were assessed by RT-PCR in 158 clinical samples. By correlation analysis based on RT-PCR results, a significantly negative correlation was found between a majority of lncRNA-miRNA pairs for both upregulated (Figure 7A) and downregulated (Figure 7D) lncRNAs. In accordance with TCGA dataset, most of the lncRNAs-associated miRNA-mRNA pairs demonstrated a negative correlation in clinical samples (Figure 7B and 7E), and the positive correlation in a great part of lncRNA-mRNA pairs was also observed (Figure 7C and 7F). Furthermore, we analyzed the expression pattern of the proposed ceRNA network in samples from different stages (early = TNM I or II; late = III) based on the results of RT-PCR, including the six lncRNAs and associated miRNAs and specific mRNAs. As shown in Figure 7G, we found the upregulated lncRNA *GBAP1*, *MCM3AP-AS1* and *SLC16A1-AS1*

Table 3. Univariate and multivariate analysis of OS and PFS in 158 HCC patients.

Variable	Univariate cox						Multivariate cox					
	Overall survival		Progression free survival		Distant metastasis-free survival		Overall survival		Progression free survival		Distant metastasis-free survival	
	P-value	HR (95%CI)	*P*-value	HR (95%CI)	*P*-value	HR (95%CI)	*P*-value	HR (95%CI)	*P*-value	HR (95%CI)	*P*-value	HR (95%CI)
Age												
≥50 or<50	0.8201	1.05 (0.67-1.66)	0.1448	1.34 (0.91-1.97)	0.6109	1.16 (0.65-2.06)						
Gender												
Male or female	0.6688	1.13 (0.64-1.99)	0.9348	1.02 (0.64-1.62)	0.7827	0.91 (0.46-1.79)						
HBsAg												
Positive or negative	0.6781	1.17 (0.56-2.43)	0.3082	1.39 (0.74-2.59)	0.7134	0.83 (0.30-2.30)						
Liver cirrhosis												
Yes or no	0.9817	1.01 (0.64-1.58)	0.3427	0.83 (0.56-1.23)	0.2657	0.72 (0.40-1.29)						
Tumor size												
≥5cm or<5cm	0.2149	0.75 (0.47-1.19)	0.1353	0.74 (0.49-1.10)	0.0723	0.57 (0.31-1.05)					**0.0422**	2.08 (1.03-4.23)
TNM stage												
III or I/II	**0.0306**	1.66 (1.05-2.64)	0.1796	1.32 (0.88-1.97)	**<0.0001**	9.06 (4.59-17.90)					**0.0009**	3.92 (1.75-8.80)
Tumor encapsulation												
Yes or no	**0.0021**	0.47 (0.29-0.76)	0.3202	0.82 (0.55-1.22)	0.0584	0.55 (0.30-1.02)						
Tumor number												
Multiple or single	**0.0009**	2.20 (1.38-3.52)	**0.0115**	1.66 (1.12-2.46)	0.1019	1.63 (0.91-2.93)	**0.0113**	1.90 (1.16-3.11)	**0.0495**	1.49 (1.00-2.23)		
AFP												
≥400ng/ml or <400ng/ml	0.3263	1.25 (0.80-1.97)	0.6073	1.11 (0.75-1.63)	0.0152	2.09 (1.15-3.79)						
Histological differentiation												
Poor or well	0.0923	1.77 (0.91-3.45)	0.1932	1.43 (0.83-2.49)	0.1072	2.15 (0.85-5.44)						
Vascular invasion												
Yes or no	0.0785	1.69 (0.94-3.01)	**0.0033**	2.09 (1.28-3.42)	**<0.0001**	5.97 (3.26-10.95)					**0.0100**	2.42 (1.24-4.75)
C3P1 expression												
High or low	**<0.0001**	0.32 (0.20-0.52)	**<0.0001**	0.43 (0.29-0.65)	**<0.0001**	0.20 (0.10-0.40)	0.1021	-	**0.0374**	0.61 (0.38-0.97)	0.5810	-
HNF4A-AS1 expression												
High or low	**<0.0001**	0.36 (0.23-0.58)	**0.0002**	0.47 (0.32-0.69)	**<0.0001**	0.08 (0.04-0.18)	0.4931	-	0.9147	-	**0.0052**	0.30 (0.13-0.70)
DIO3OS expression												
High or low	**<0.0001**	0.16 (0.09-0.28)	**<0.0001**	0.34 (0.22-0.52)	**<0.0001**	0.17 (0.08-0.35)	**0.0042**	2.14 (1.27-3.61)	0.0865	0.64 (0.38-1.07)	0.3061	-

	P	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)
MCM3AP-AS1 expression												
High or low	<0.0001	3.75 (2.27-6.17)	<0.0001	2.38 (1.59-3.55)	<0.0001	3.54 (1.88-6.64)	0.0042	2.14 (1.27-3.61)	0.0092	1.74 (1.15-2.65)	0.0092	2.77 (1.29-5.98)
GBAP1 expression												
High or low	<0.0001	5.31 (3.14-8.99)	<0.0001	2.67 (1.77-4.02)	<0.0001	23.38 (8.74-62.56)	0.0625	-	0.4362	-	0.0005	6.83 (2.32-20.12)
SLC16A1-AS1 expression												
High or low	<0.0001	3.75 (2.27-6.17)	<0.0001	2.38 (1.59-3.55)	<0.0001	3.54 (1.88-6.64)	0.0001	2.79 (1.65-4.70)	0.0053	1.82 (1.19-2.78)	0.0156	2.36 (1.18-4.73)

were highly expressed in late-stage samples compared with early-stage samples. Inverse and similar expression patterns were observed in their associated miRNAs and mRNAs, respectively (Figure 7G). In contrast, the downregulated lncRNA *C3P1*, *DIO3OS*, and *HNF4A-AS1* were low expressed in late-stage samples compared to early-stage samples (Figure 7H). Inverse and similar expression patterns were also observed in their associated miRNAs and mRNAs, respectively (Figure 7H). Taken together, these data further verified that the six lncRNAs might play an important role as ceRNAs in regulating the expression of miRNAs and mRNAs. Moreover, upregulation of *GBAP1*, *MCM3AP-AS1*, and *SLC16A1-AS1* and downregulation of *C3P1*, *DIO3OS*, and *HNF4A-AS1* might be involved in regulating the progression of HCC from early to late stage.

Receiver operating characteristic (ROC) curves analysis for the key lncRNAs in clinical samples

To assess the discriminatory ability of the six key lncRNAs in HCC, ROC curve analyses were conducted in 158 patients with HCC and the areas under the curve (AUCs) were calculated. As shown in Figure 8A–8F, in the assessment of OS, the AUCs of five lncRNAs (*GBAP1*, *MCM3AP-AS1*, *SLC16A1-AS1*, *C3P1*, and *DIO3OS*) were more than 0.75, and the AUCs of *HNF4A-AS1* was 0.737. In the assessment of PFS, the AUCs of all the six lncRNAs were greater than 0.75 (Figure 8G–8L). Additionally, five lncRNAs (*GBAP1*, *MCM3AP-AS1*, *SLC16A1-AS1*, *HNF4A-AS1* and *DIO3OS*) had a good performance in diagnosing DmFS with AUCs above 0.75, and the AUCs of *C3P1* was 0.712 (Figure 8M–8R). These results suggested that the six lncRNAs had good sensitivity and specificity to predict survival and distant metastasis in patients with HCC.

DISCUSSION

In recent decades, the diagnosis and treatment of HCC have improved dramatically. Although patients with HCC have benefited from multiple options that improve their survival, regardless of the cancer stage at diagnosis, the survival time of patients with HCC remains limited [30]. Apart from traditional surgical resection, ablation, and systematic chemotherapy, the targeted molecular therapy has become a promising therapeutic option for patients with advanced-stage disease or patients who transitioned into it after other therapies failed [1, 2]. However, the efficacy of this improved therapy is still limited to certain patients. Thus, there is an urgent need to discover more molecules that drive HCC tumorigenesis and progression to develop new and effective therapeutic targets for HCC.

Non-coding RNAs have raised considerable research interest due to their regulation of the transcription of protein-coding genes to accelerate cancer progression. As typical non-coding RNAs, lncRNAs and miRNAs have been extensively implicated in the oncogenesis of a variety of cancers [31], including HCC [14, 15]. In the present study, we firstly analyzed five expression profiles to identify differentially expressed genes, lncRNAs, and miRNAs between HCC samples and normal samples based on data from the GEO database. For the identified DEGs, GO functional enrichment analysis showed that the upregulated DEGs were most significantly associated with cell division, a process that allows cells to proliferate persistently, while the downregulated DEGs were mainly involved in the oxidation-reduction process. Consistently, it has been reported that the oxidant production, such as H_2O_2, is elevated in the tumor microenvironment because of the imbalance between oxidation and reduction. Some oxidants can trigger cell growth and immune cell dysfunction in different kinds of cancer types, including HCC [32–34]. KEGG pathway analysis demonstrated that the upregulated DEGs were most strongly implicated in the cell cycle, and downregulated DEGs mainly participated in metabolic pathways. Abnormalities of these pathways were considered as two essential hallmarks of cancer [35]. The signal-net analysis demonstrated how these DEGs influenced each other, which suggested that some hub genes exerted central regulatory functions, such as *PRKAA2*, *ITGB3*, *PLCB1*, and *ITGA2*,

providing further scientific clues to study and treat HCC. Consistent with our results, Zhang et al. confirmed that the upregulated expression of *ITGB3* mediated the expression of *MMP2* by activating the FAK/PI3K/AKT signaling pathway, contributing to the enhancement of metastatic potential of residual cancer in the HCCLM3 model after insufficient radiofrequency ablation [36]. Similarly, *ITGB3* and *PLCB1* were also confirmed to facilitate HCC progression by enhancing the adhesion and proliferation of tumor cells. However, contradictory results were observed in some other cancer types, such as gastric and prostate cancer, which indicated that *PRKAA2* was a protective gene in cancer development by upregulating the expression of hypoxia-inducible factor-1α and hepatocyte nuclear factor 4α [37, 38]. More experimental evidence is required to verify the role of *PRKAA2* in the development of HCC.

MiRNAs mediate the expression of transcripts by binding to MREs of their target mRNAs. CeRNAs are a group of non-coding transcripts that maintain the
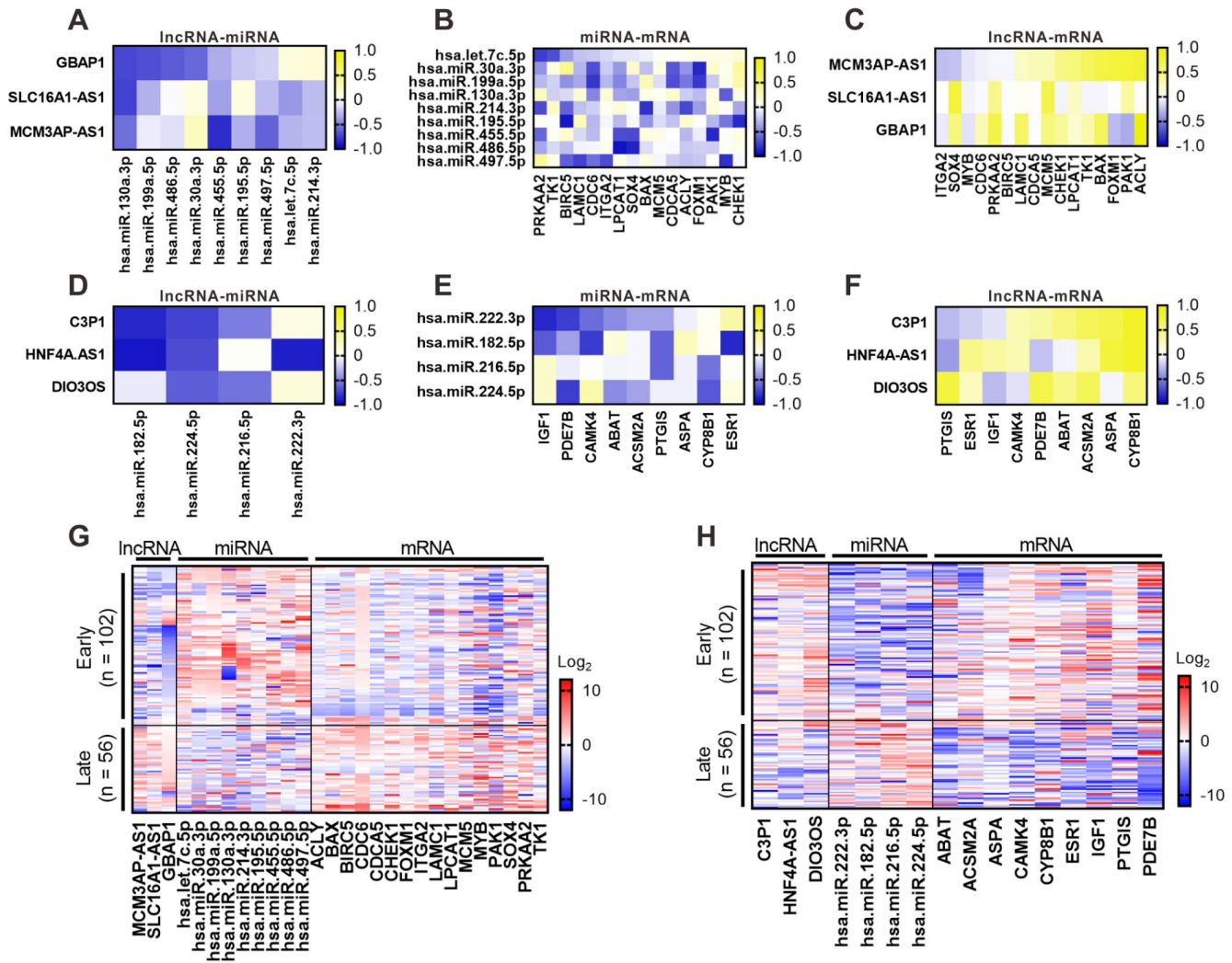
Figure 7. Correlation validation for the six lncRNAs with their associated miRNA and mRNAs in 158 clinical samples of HCC. RT-PCR was performed to detect the expression of the six lncRNAs associated miRNA and mRNAs in 158 clinical samples. Pearson correlation analysis for six key lncRNAs with their associated miRNA and mRNAs was performed based on RT-PCR results. Pearson correlograms of lncRNA-miRNA paris (**A**), lncRNA-mRNA pairs (**B**) and mRNA-miRNA pairs (**C**) in upregulated lncRNAs including *GBAP1*, *MCM3AP-AS1* and *SLC16A1-AS1*. lncRNA-miRNA paris (**D**), lncRNA-mRNA paris (**E**) and mRNA-miRNA paris (**F**) in downregulated lncRNAs including *C3P1*, *HNF4A-AS1* and *DIO3OS* by Pearson correlation analysis. The correlation coefficient R value ranges from -1 to 1, the color of which changes from blue to yellow. 158 clinical samples were diveded into early and late groups according to TMN stage (early = TNM I or II; late = III), the expression of the six lncRNAs associated miRNA and mRNAs was shown as heatmaps. (**G**) Expression of the upregulated lncRNA *GBAP1*, *MCM3AP-AS1* and *SLC16A1-AS1* and their ssociated miRNA and mRNAs. (**H**) Expression of th edownregulated lncRNA *C3P1*, *DIO3OS*, and *HNF4A-AS1* and their ssociated miRNA and mRNAs. Log$_2$(P) value was shown as color ranging from blue to red.

balance between miRNAs and their target genes [10]. These targets include pseudogenes, protein-coding genes, and lncRNAs. The ceRNA hypothesis proposes that these transcripts can act as bona fide miRNA competitors owing to the presence of MREs or high-sequence homology. Therefore, ceRNAs actively compete with their ancestral protein-coding genes for the same pool of miRNAs. The consequence of competition for miRNAs is a decrease in their activity to its targets. Therefore, except for the conventional miRNA-RNA regulation, a reversed RNA-miRNA regulation relationship also exists, in which coding and noncoding RNA targets can exert crosstalk through their ability to compete for miRNA binding [10, 39, 40]. In this study, we constructed an mRNA-miRNA-lncRNA interaction network (ceRNA network) to show the interactive regulation relationships among the DEGs, DELs, and DEMs.

In this network, lncRNAs functioned as ceRNAs to sequester miRNAs and regulate mRNA transcripts containing shared MREs. The correlation analysis demonstrated a significant negative correlation in a majority of lncRNA-miRNA and miRNA-mRNA pairs, in line with the theoretical silencing effect between miRNAs and its target transcripts. Hence, the

constructed mRNA-miRNA-lncRNA interaction network was a ceRNA crosstalk network. It reminded us of the potential regulatory mechanisms for transcripts involved in HCC development. Indeed, accumulating evidence pointed to the function of ceRNA network in regulating tumor progression. LncRNA *PVT1* significantly promoted autophagy and subsequent proliferation of tumor cells through acting as a ceRNA to target *ATG3* by sponging microRNA-365 in HCC [41]. Also, lncRNA *FAL1* was found to promote the proliferation and migration of HCC cells by acting as a ceRNA of miR-1236 [42]. Similarly, ceRNA networks also serve as important regulatory mechanisms to accelerate cancer progression in other cancer types. For example, lncRNA *PTAR* promoted EMT and invasion- metastasis in serous ovarian cancer by competitively binding to miR-101-3p to regulate *ZEB1* expression [12]. LncRNA-*KRTAP5-AS1* and lncRNA-*TUBB2A* could act as ceRNAs to affect the function of Claudin-4, which reinforces proliferation, invasion, and EMT in gastric cancer [43].

We also validated our results in TCGA data and clinical samples. Using survival analysis, COX regression analysis, and ROC analysis, we demonstrated that a
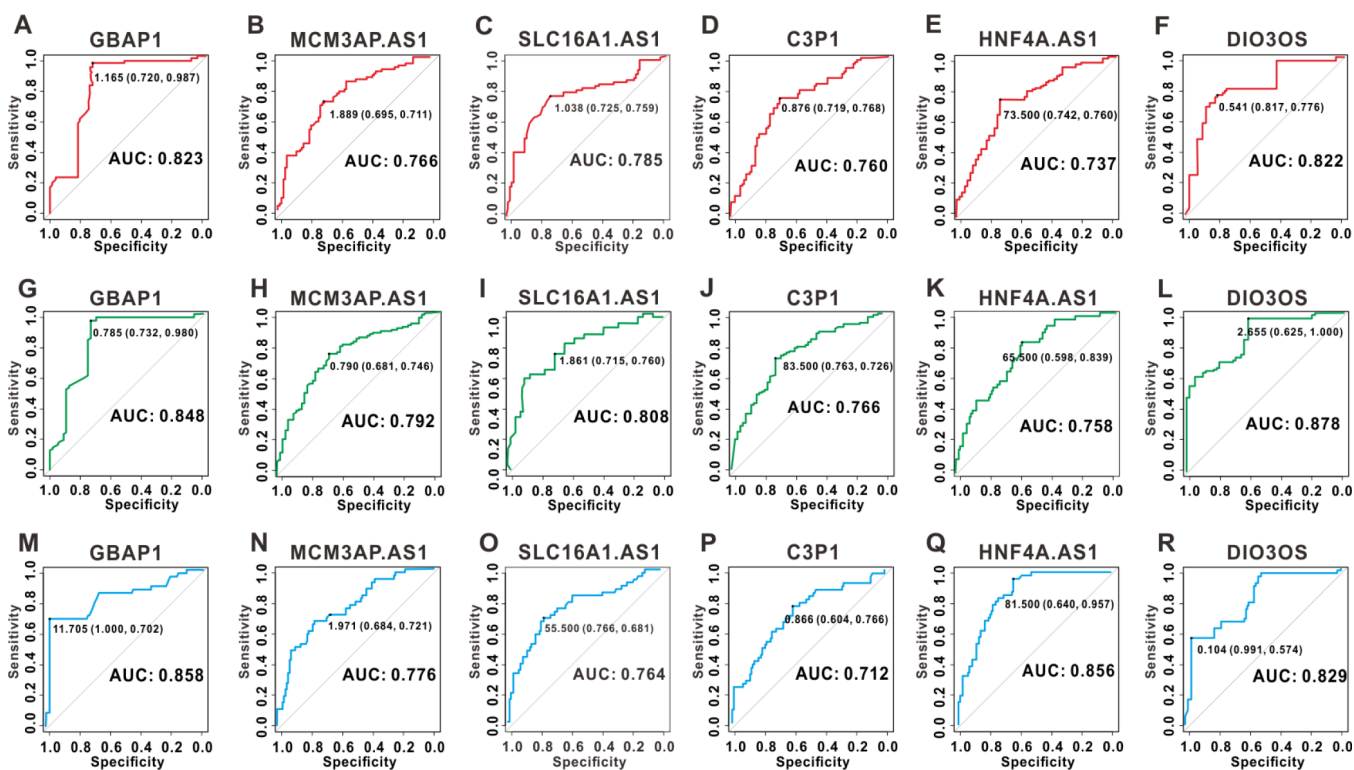

Figure 8. ROC curves for the six key lncRNAs in 158 clinical samples. ROC curves of lncRNA *GBAP1*, *MCM3AP-AS1*, *SLC16A1-AS1*, *C3P1*, *HNF4A-AS1* and *DIO3OS* for OS (**A–F**), PFS (**G–L**), and DmFS (**M–R**) respectively. The AUCs under binomial exact confidence interval was calculated to generate the ROC curve.

proportion of DELs, DEMs, and DEGs in the ceRNA network had a significant impact on the OS of patients with HCC, presenting promising clinical and translational significance. Among them, *BAX*, *BIRC5*, *ITGA2*, and *MYB* were identified as significantly correlated with the survival of patients with HCC, which was in line with the results of previous studies [28, 44–46], while *PRKAA2*, *PTGIS*, *PDE7B*, *CYP8B1*, *CHEK1*, *CAMK4*, *ASPA*, and *ACSM2A* were identified to predict the prognosis of patients with HCC for the first time. Also, we found six lncRNAs and five miRNAs that might be valuable predictive factors for the survival of patients with HCC in both TCGA and clinical samples. LncRNA *MCM3AP-AS1* has shown the same result in the survival analysis in another group containing 80 HCC patients, which was also verified to facilitate the proliferation and suppress the apoptosis of HCC cells by acting as ceRNA of miR-194-5p targeting *FOXA1* [37]. Notably, the other five lncRNAs (*C3P1*, *DIO3OS*, *GBAP1*, *SLC16A1-AS1*, and *HNF4A-AS1*) and all five miRNAs were identified to be related to the prognosis of patients with HCC for the first time. Also, the results of correlation validation of the six lncRNAs with their associated miRNAs and mRNAs in TCGA were mostly consistent with the results in the constructed ceRNA network. This indicated that the network was reliable and that the six identified lncRNAs might play a role in the mechanisms of HCC progression. Therefore, they might have promising prognostic and therapeutic values in patients with HCC. However, there are also a few limitations of our study. For example, we demonstrated their relationship in the ceRNA network at the RNA level but in-depth work will be needed to verify their function using experimental data. Moreover, the data used were obtained from the GEO database, rather than directly from patients with HCC. Therefore, we need to perform a series of verification studies in a large-scale cohort of patients and at multiple centers to confirm these results.

In conclusion, the present study performed large-scale analyses and highlighted the complex crosstalk involving lncRNA-miRNA-mRNA networks and the important roles of lncRNAs as ceRNAs in HCC development. We identified a cluster of lncRNAs (*GBAP1*, *MCM3AP-AS1*, *SLC16A1-AS1*, *C3P1*, *DIO3OS*, and *HNF4A-AS1*) as potential ceRNAs that regulate HCC carcinogenesis and progression, defining them as specific biomarkers to diagnose HCC and predict the prognosis and metastasis of patients with HCC. These findings will improve our understanding of the ceRNAs' regulatory mechanisms in HCC development and contribute to the identification of potential targets for the clinical diagnosis and treatment of patients with HCC.

MATERIALS AND METHODS

Patients and samples

Tissue samples, including HCC tumor tissues and adjacent non-cancerous tissues (n = 158), were obtained from patients who had resection of primary HCC in the Cancer Center of Sun Yat-sen University between 2005 and 2008. None of these patients had received pre-operative chemotherapy or radiotherapy. After resection, matched fresh tissues were immersed immediately in RNAlater® (Ambion, Austin, TX, USA), kept overnight at 4 °C, and then stored at −80 °C until RNA isolation before qRT-PCR detection. Follow-up was performed in our outpatient department and involved clinical and laboratory examinations every three months for the first two years, every six months during the third to fifth years, and annually for an additional five years or until death, whichever occurred first. Follow-up periods for survivors ranged from 2 to 87 months, with a median follow-up of 41 months. OS and PFS were used as measures of prognosis. Written informed consent was obtained from each patient, and the Ethics Committee of Sun Yat-Sen University Cancer Center approved the study protocol.

GEO gene expression datasets

HCC gene expression data were obtained from the GEO database (http://www.ncbi.nlm.nih.gov/geo/), including GSE29721, GSE40367, and GSE62232. The three profiles included a total of 120 samples, consisting of 98 cancerous and 22 normal samples. The mRNA expression profiles were obtained using an Affymetrix Human Genome U133 Plus 2.0 Array. According to the annotation file, we chose the probe sets and BLAST searched them in the NCBI RefSeq database to identify the probe sets of noncoding RNAs with a length of over 200 bp. We then used these probe sets to acquire the expression data of lncRNAs from the three profiles. The miRNA expression data were also obtained from the GEO databases. Two profiles, GSE36915 and GSE74618, were selected that contained 286 cancerous and 31 normal samples.

Differential and clustering analysis

Based on the expression data from the datasets, a random variance model (RVM) t-test was applied to filter the DEGs, DELs, and DEMs among cancerous and normal samples. After significance analysis and FDR analysis, we selected the DEGs, and DELs from GSE29721, GSE40367, and GSE62232 according to the *P*-value threshold. $P < 0.05$ and FC > 2 was considered as significant difference. DEMs were selected using a threshold of *P*-value < 0.05, FDR < 0.05, and FC > 1.5.

Hierarchical cluster analysis was performed and a cluster dendrogram was constructed to demonstrate distinct characterizations of screened DEGs, DELs, and DEMs between the cancerous and normal tissues. Furthermore, we selected the intersecting DEGs, DELs, and DEMs among the different profiles.

GO and KEGG pathway analysis

Gene functional enrichment analysis was used to predict the biological functions of the intersected DEGs according to the GO database with $P < 0.05$ and FDR <0.05. The potentially involved signaling pathways were identified using the KEGG pathways analysis program (http://www.genome.jp/kegg/tool/map_pathway1.html).

Two-side Fisher's exact test and a χ^2 test were used to classify the GO categories, and the FDR [47] was calculated to correct the P-value: The smaller the FDR, the smaller the error in judging the P-value. The FDR was defined as:

$$FDR = 1 - N_k / T, \qquad (1)$$

where N_k referred to the number of Fisher's test P-values less than the test P-values. We computed P-values for the GO categories of all the DEGs. Within the significant categories, the enrichment, Re, was calculated as:

$$Re = (n_f / n)/(N_f / N), \qquad (2)$$

where "n_f" was the number of flagged genes within the particular category, "n" was the total number of genes within the same category, "N_f" was the number of flagged genes in the entire microarray, and "N" was the total number of genes in the microarray [48]. For KEGG pathway analysis, we again used Fisher's exact test and χ^2 test to select the significant pathways, and the threshold of significance was defined P-value <0.05 and FDR < 0.05. The Re value was calculated using equation (2) [49–51].

Signal-net analysis of intersecting DEGs

Based on the KEGG database analysis, a gene-gene interaction network of DEGs was constructed to demonstrate the regulatory relationships among the DEGs identified in the intersection analysis among different expression profiles of mRNAs. The networks were stored and presented as graphs, where nodes were mainly genes (or proteins or compounds) and edges represented the type relationships between the nodes, e.g. activation or phosphorylation. We investigated the

nature of networks using tools implemented in the R software.

The important nodes were identified computationally. To this end, we used the connectivity (also known as degree), which was defined as the sum of connection strengths with the other network genes:

$$K_i = \sum_{u \neq i} a_{ui} \qquad (3)$$

In the gene networks, the connectivity measured how a gene correlated with all other network genes. For a gene in the network, the number of source genes of a gene was called the indegree of the gene and the number of target genes of a gene was its outdegree. The character of a gene was described using betweenness centrality measures reflecting the importance of a node in a graph relative to other nodes. For a graph G: (V, E) with n vertices, the relative betweenness centrality $C'_B(v)$ is defined by:

$$C'_B(v) = \frac{2}{n^2 - 3n + 2} \sum_{\substack{s \neq v \neq t \in V \\ s \neq t}} \frac{\sigma_{st}(v)}{\sigma_{st}}, \qquad (4)$$

where σ_{st} is the number of shortest paths from s to t, and $\sigma_{st}(v)$ is the number of shortest paths from s to t that pass through a vertex v [52–56].

Target transcripts of DEMs prediction and ceRNA network construction

Based on the functional DEGs identified in the GO and KEGG pathway analysis, we constructed the intersection datasets between the DEGs involved in the significant enriched GO terms and pathways with $P < 0.05$ and FDR < 0.05. Ultimately, 393 DEGs were selected, comprising 78 upregulated DEGs and 315 downregulated DEGs. Combining the intersecting DELs and DEMs, we predicted the targeted sponge lncRNAs using the miRanda tool (http://www.microrna.org/microrna/home.do), and the target mRNAs using Targetscan (http://www.targetscan.org/) and miRWalk (http://129.206.7.150/). For each pair of miRNA-mRNA or mirRNA-lncRNA, we conducted Pearson correlation analysis and chose the significantly correlated pairs [57]. In particular, we summarized the intersecting targeted mRNAs discerned using miRanda and miRWalk.

Then, we chose the miRNAs that negatively regulated the expression levels of lncRNAs and mRNAs to construct a ceRNA network according to the normalized signal intensity of the expression of specific mRNAs and lncRNAs.

Protein regulation network analysis

The STRING online database tool (https://string-db.org/cgi/input.pl) was used to construct a PPI network of the DEGs identified in the ceRNA network. The interacting pairs with a confidence score greater than 0.4 were considered as significant and were retained. The proteins encoded by the hub genes were screened according to the degree of the nodes. The degree represents the number of interaction partners and was calculated using Perl code.

Survival analysis

We screened the HCC RNA-Seq TCGA datasets containing survival information and selected 370 patients with HCC as a dataset to analyze the relationship between the expression level of lncRNAs and mRNAs and OS. Another dataset that contained 371 patients with HCC was selected to analyze the relationship between the expression level of miRNAs and OS. In the clinical samples, we also investigated the correlation of lncRNAs with OS and PFS for 158 patients with HCC. According to the expression level of mRNAs, lncRNAs, and miRNAs in TCGA or our clinical samples, we classified them into two groups: High expression and low expression. Survival curves were displayed using Kaplan–Meier plots. The Wilcoxon log-rank test was used to analyze the survival difference between the high and low groups. All survival analyses in TCGA were conducted using the R package, Survival.

Correlation analysis

To verify the correlation of expression among the miRNAs, lncRNAs, and mRNAs identified in the ceRNA network, we chose the lncRNAs that had significance for the prognosis of patients with HCC and their associated-mRNAs and miRNAs. Pearson correlation analyses were performed among them based on the RNA-Seq data of 366 patients with HCC searched in TCGA database (https://cancergenome.nih. gov/). The correlogram was constructed using the R package, Corrplot.

RNA extraction and qRT-PCR

Total RNA was isolated from tissues using the TRIzol reagent (Invitrogen Corporation, Waltham, MA, USA) according to the manufacturer's instructions. The concentration and purity of the RNA were evaluated using a NanoDrop 2000 instrument (Thermo Scientific, Waltham, MA, USA). For mRNA, the first-strand cDNA was synthesized from total RNA using a GoScript Reverse Transcription System (Promega, Madison, WI, USA). *GAPDH* was used as an endogenous control for normalization. For miRNAs, the first-strand cDNA was synthesized from total RNA using a riboSCRIPT™ Reverse Transcription Kit (RIBOBIO, GuangzZhou, China). *U6* was used as an endogenous control for normalization. QPCR was performed using GoTaq qPCR Master Mix (Promega).

ROC curve analysis

ROC analyses were performed using the pROC package in the R language, based on data from 158 clinical samples. The diagnostic ability of the prediction model was evaluated by calculating the area under a ROC curve. The ROC curve was used for classifier evaluation and was drawn by plotting sensitivity against the false-positive rate. The AUC under a binomial exact confidence interval was calculated to generate the ROC curve.

Statistical analysis

Data are shown as the mean ± s.d. and analyzed using Student's t-test. A paired t-test was used for paired samples. Statistical analyses were performed using GraphPad Prism 7 and SPSS software (GraphPad Software, La Jolla, CA, USA). A value of $P<0.05$ was considered statistically significant.

Abbreviations

AUC: area under the curve; ceRNA: competitive endogenous RNA; DEGs: differentially expressed genes; DELs: differentially expressed lncRNAs; DEMs: differentially expressed miRNAs; EMT: epithelial-mesenchymal transition; FC: Fold Change; FDR: false discovery rate; GEO: Gene Expression Omnibus; GO: Gene Ontology; HCC: hepatocellular carcinoma; KEGG: Kyoto Encyclopedia of Genes and Genomes; lncRNA: long noncoding RNA; miRNA: microRNAs; MRE: miRNA response elements; OS: overall survival; PFS: progression-free survival; qRT-PCR: quantitative reverse transcriptase-polymerase chain reaction; ROC: Receiver operating characteristic; signal-net: signal regulation network; TCGA: the Cancer Genome Atlas; TGF-β: transforming growth factor-beta.

AUTHOR CONTRIBUTIONS

Conception and design: M. Song, A. Zhong, J. Yang, J. Xia, D. Weng; Development of methodology: J. Yang, J. He, M. Song, A. Zhong, S. Cheng; Acquisition of data: M. Song, J. Yang, J. He, S. Cheng, J. Zeng, Y. Huang; Analysis and interpretation of data: M. Song, J. Yang, J. He, A. Zhong, S Cheng, J. Zeng, Y. Huang, Q. Pan, J. Zhao, Z. Zhou; Writing, review, and/or revision of the manuscript: J. Yang, M. Song, A. Zhong, J. He; Administrative, technical, or material support: M. Song,

A. Zhong, J. Yang, Q Zhu, H. Chen, Y. Tang, C. Yang, Y. Liu, X. Mo.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interests.

FINDING

REFERENCES

1. El-Serag HB. Hepatocellular carcinoma. N Engl J Med. 2011; 365:1118–27.
https://doi.org/10.1056/NEJMra1001683
PMID:21992124

2. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. Lancet. 2018; 391:1301–14.
https://doi.org/10.1016/S0140-6736(18)30010-2
PMID:29307467

3. Forner A, Llovet JM, Bruix J. Hepatocellular carcinoma. Lancet. 2012; 379:1245–55.
https://doi.org/10.1016/S0140-6736(11)61347-0
PMID:22353262

4. Agarwal R, Narayan J, Bhattacharyya A, Saraswat M, Tomar AK. Gene expression profiling, pathway analysis and subtype classification reveal molecular heterogeneity in hepatocellular carcinoma and suggest subtype specific therapeutic targets. Cancer Genet. 2017; 216-217:37–51.
https://doi.org/10.1016/j.cancergen.2017.06.002
PMID:29025594

5. Lu C, Chen H, Shan Z, Yang L. Identification of differentially expressed genes between lung adenocarcinoma and lung squamous cell carcinoma by gene expression profiling. Mol Med Rep. 2016; 14:1483–90.
https://doi.org/10.3892/mmr.2016.5420
PMID:27356570

6. Wang J, Zhang X, Chen W, Hu X, Li J, Liu C. Regulatory roles of long non-coding RNAs implicated in cancer hallmarks. Int J Cancer. 2019.
https://doi.org/10.1002/ijc.32277.

7. Kopp F, Mendell JT. Functional Classification and Experimental Dissection of Long Noncoding RNAs. Cell. 2018; 172:393–407.
https://doi.org/10.1016/j.cell.2018.01.011
PMID:29373828

8. Esposito R, Bosch N, Lanzós A, Polidori T, Pulido-Quetglas C, Johnson R. Hacking the Cancer Genome: Profiling Therapeutically Actionable Long Non-coding RNAs Using CRISPR-Cas9 Screening. Cancer Cell. 2019; 35:545–57.
https://doi.org/10.1016/j.ccell.2019.01.019
PMID:30827888

9. Hausser J, Zavolan M. Identification and consequences of miRNA-target interactions—beyond repression of gene expression. Nat Rev Genet. 2014; 15:599–612.
https://doi.org/10.1038/nrg3765
PMID:25022902

10. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell. 2011; 146:353–58.
https://doi.org/10.1016/j.cell.2011.07.014
PMID:21802130

11. Song X, Cao G, Jing L, Lin S, Wang X, Zhang J, Wang M, Liu W, Lv C. Analysing the relationship between lncRNA and protein-coding gene and the role of lncRNA as ceRNA in pulmonary fibrosis. J Cell Mol Med. 2014; 18:991–1003.
https://doi.org/10.1111/jcmm.12243 PMID:24702795

12. Liang H, Yu T, Han Y, Jiang H, Wang C, You T, Zhao X, Shan H, Yang R, Yang L, Shan H, Gu Y. LncRNA PTAR promotes EMT and invasion-metastasis in serous ovarian cancer by competitively binding miR-101-3p to regulate ZEB1 expression. Mol Cancer. 2018; 17:119.
https://doi.org/10.1186/s12943-018-0870-5
PMID:30098599

13. Huang T, Huang W, Lu H, Zhang BY, Ma J, Zhao D, Wang YJ, Yu DH, He X. Identification and validation a TGF-β-associated long non-coding RNA of head and neck squamous cell carcinoma by bioinformatics method. J Transl Med. 2018; 16:46.
https://doi.org/10.1186/s12967-018-1418-6
PMID:29490660

14. Sui CJ, Zhou YM, Shen WF, Dai BH, Lu JJ, Zhang MF, Yang JM. Long noncoding RNA GIHCG promotes hepatocellular carcinoma progression through epigenetically regulating miR-200b/a/429. J Mol Med (Berl). 2016; 94:1281–96.
https://doi.org/10.1007/s00109-016-1442-z
PMID:27380494

15. Ren Y, Shang J, Li J, Liu W, Zhang Z, Yuan J, Yang M. The long noncoding RNA PCAT-1 links the microRNA miR-215 to oncogene CRKL-mediated signaling in

hepatocellular carcinoma. J Biol Chem. 2017; 292:17939–49.
https://doi.org/10.1074/jbc.M116.773978
PMID:28887306

16. Zhou X, Chen J, Chen L, Feng X, Liu Z, Hu L, Zeng Z, Jia X, Liang M, Shi B, Yi G, Liu J. Negative regulation of Sirtuin 1 by AMP-activated protein kinase promotes metformin-induced senescence in hepatocellular carcinoma xenografts. Cancer Lett. 2017; 411:1–11.
https://doi.org/10.1016/j.canlet.2017.09.027
PMID:28964787

17. Hamidi H, Ivaska J. Every step of the way: integrins in cancer progression and metastasis. Nat Rev Cancer. 2018; 18:533–48.
https://doi.org/10.1038/s41568-018-0038-z
PMID:30002479

18. Li CY, Liang GY, Yao WZ, Sui J, Shen X, Zhang YQ, Peng H, Hong WW, Ye YC, Zhang ZY, Zhang WH, Yin LH, Pu YP. Integrated analysis of long non-coding RNA competing interactions reveals the potential role in progression of human gastric cancer. Int J Oncol. 2016; 48:1965–76.
https://doi.org/10.3892/ijo.2016.3407 PMID:26935047

19. Xu S, Sui J, Yang S, Liu Y, Wang Y, Liang G. Integrative analysis of competing endogenous RNA network focusing on long noncoding RNA associated with progression of cutaneous melanoma. Cancer Med. 2018; 7:1019–29.
https://doi.org/10.1002/cam4.1315 PMID:29522273

20. You X, Zhao Y, Sui J, Shi X, Sun Y, Xu J, Liang G, Xu Q, Yao Y. Integrated analysis of long noncoding RNA interactions reveals the potential role in progression of human papillary thyroid cancer. Cancer Med. 2018; 7:5394–410.
https://doi.org/10.1002/cam4.1721 PMID:30318850

21. Jafarzadeh M, Soltani BM, Dokanehiifard S, Kay M, Aghdami N, Hosseinkhani S. Experimental evidences for hsa-miR-497-5p as a negative regulator of SMAD3 gene expression. Gene. 2016; 586:216–21.
https://doi.org/10.1016/j.gene.2016.04.003
PMID:27063509

22. Martini E, Schneider E, Neufert C, Neurath MF, Becker C. Survivin is a guardian of the intestinal stem cell niche and its expression is regulated by TGF-β. Cell Cycle. 2016; 15:2875–81.
https://doi.org/10.1080/15384101.2016.1231286
PMID:27715398

23. Lee J, Choi JH, Joo CK. TGF-β1 regulates cell fate during epithelial-mesenchymal transition by upregulating survivin. Cell Death Dis. 2013; 4:e714.
https://doi.org/10.1038/cddis.2013.244
PMID:23828577

24. Wang Y, Yang L, Chen T, Liu X, Guo Y, Zhu Q, Tong X, Yang W, Xu Q, Huang D, Tu K. A novel lncRNA MCM3AP-AS1 promotes the growth of hepatocellular carcinoma by targeting miR-194-5p/FOXA1 axis. Mol Cancer. 2019; 18:28.
https://doi.org/10.1186/s12943-019-0957-7
PMID:30782188

25. Yang Y, Chen L, Gu J, Zhang H, Yuan J, Lian Q, Lv G, Wang S, Wu Y, Yang YT, Wang D, Liu Y, Tang J, et al. Recurrently deregulated lncRNAs in hepatocellular carcinoma. Nat Commun. 2017; 8:14421.
https://doi.org/10.1038/ncomms14421
PMID:28194035

26. Dou CY, Fan YC, Cao CJ, Yang Y, Wang K. Sera DNA Methylation of CDH1, DNMT3b and ESR1 Promoters as Biomarker for the Early Diagnosis of Hepatitis B Virus-Related Hepatocellular Carcinoma. Dig Dis Sci. 2016; 61:1130–38.
https://doi.org/10.1007/s10620-015-3975-3
PMID:26660680

27. D'Alessandro R, Refolo MG, Lippolis C, Carella N, Messa C, Cavallini A, Carr BI. Strong enhancement by IGF1-R antagonists of hepatocellular carcinoma cell migration inhibition by Sorafenib and/or vitamin K1. Cell Oncol (Dordr). 2018; 41:283–96.
https://doi.org/10.1007/s13402-018-0370-z
PMID:29470830

28. Su C. Survivin in survival of hepatocellular carcinoma. Cancer Lett. 2016; 379:184–90.
https://doi.org/10.1016/j.canlet.2015.06.016
PMID:26118774

29. Cui DJ, Wu Y, Wen DH. CD34, PCNA and CK19 expressions in AFP- hepatocellular carcinoma. Eur Rev Med Pharmacol Sci. 2018; 22:5200–05.
PMID:30178842

30. Villanueva A. Hepatocellular Carcinoma. N Engl J Med. 2019; 380:1450–62.
https://doi.org/10.1056/NEJMra1713263
PMID:30970190

31. Anastasiadou E, Jacob LS, Slack FJ. Non-coding RNA networks in cancer. Nat Rev Cancer. 2018; 18:5–18.
https://doi.org/10.1038/nrc.2017.99 PMID:29170536

32. Langford TF, Huang BK, Lim JB, Moon SJ, Sikes HD. Monitoring the action of redox-directed cancer therapeutics using a human peroxiredoxin-2-based probe. Nat Commun. 2018; 9:3145.
https://doi.org/10.1038/s41467-018-05557-y
PMID:30087344

33. Lee D, Xu IM, Chiu DK, Leibold J, Tse AP, Bao MH, Yuen VW, Chan CY, Lai RK, Chin DW, Chan DF, Cheung TT, Chok SH, et al. Induction of Oxidative Stress Through Inhibition of Thioredoxin Reductase 1 Is an Effective

Therapeutic Approach for Hepatocellular Carcinoma. Hepatology. 2019; 69:1768–86.
https://doi.org/10.1002/hep.30467 PMID:30561826

34. Chen X, Song M, Zhang B, Zhang Y. Reactive Oxygen Species Regulate T Cell Immune Response in the Tumor Microenvironment. Oxid Med Cell Longev. 2016; 2016:1580967.
https://doi.org/10.1155/2016/1580967 PMID:27547291

35. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011; 144:646–74.
https://doi.org/10.1016/j.cell.2011.02.013 PMID:21376230

36. Zhang N, Ma D, Wang L, Zhu X, Pan Q, Zhao Y, Zhu W, Zhou J, Wang L, Chai Z, Ao J, Sun H, Tang Z. Insufficient Radiofrequency Ablation Treated Hepatocellular Carcinoma Cells Promote Metastasis by Up-Regulation ITGB3. J Cancer. 2017; 8:3742–54.
https://doi.org/10.7150/jca.20816 PMID:29151962

37. Kim YH, Liang H, Liu X, Lee JS, Cho JY, Cheong JH, Kim H, Li M, Downey TJ, Dyer MD, Sun Y, Sun J, Beasley EM, et al. AMPKα modulation in cancer progression: multilayer integrative analysis of the whole transcriptome in Asian gastric cancer. Cancer Res. 2012; 72:2512–21.
https://doi.org/10.1158/0008-5472.CAN-11-3870 PMID:22434430

38. Vila IK, Yao Y, Kim G, Xia W, Kim H, Kim SJ, Park MK, Hwang JP, González-Billalabeitia E, Hung MC, Song SJ, Song MS. A UBE2O-AMPKα2 Axis that Promotes Tumor Initiation and Progression Offers Opportunities for Therapy. Cancer Cell. 2017; 31:208–24.
https://doi.org/10.1016/j.ccell.2017.01.003 PMID:28162974

39. Karreth FA, Tay Y, Perna D, Ala U, Tan SM, Rust AG, DeNicola G, Webster KA, Weiss D, Perez-Mancera PA, Krauthammer M, Halaban R, Provero P, et al. In vivo identification of tumor- suppressive PTEN ceRNAs in an oncogenic BRAF-induced mouse model of melanoma. Cell. 2011; 147:382–95.
https://doi.org/10.1016/j.cell.2011.09.032 PMID:22000016

40. Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature. 2010; 465:1033–38.
https://doi.org/10.1038/nature09144 PMID:20577206

41. Yang L, Peng X, Jin H, Liu J. Long non-coding RNA PVT1 promotes autophagy as ceRNA to target ATG3 by sponging microRNA-365 in hepatocellular carcinoma. Gene. 2019; 697:94–102.

https://doi.org/10.1016/j.gene.2019.02.036 PMID:30794914

42. Li B, Mao R, Liu C, Zhang W, Tang Y, Guo Z. LncRNA FAL1 promotes cell proliferation and migration by acting as a CeRNA of miR-1236 in hepatocellular carcinoma cells. Life Sci. 2018; 197:122–29.
https://doi.org/10.1016/j.lfs.2018.02.006 PMID:29421439

43. Song YX, Sun JX, Zhao JH, Yang YC, Shi JX, Wu ZH, Chen XW, Gao P, Miao ZF, Wang ZN. Non-coding RNAs participate in the regulatory network of CLDN4 via ceRNA mediated miRNA evasion. Nat Commun. 2017; 8:289.
https://doi.org/10.1038/s41467-017-00304-1 PMID:28819095

44. Garcia EJ, Lawson D, Cotsonis G, Cohen C. Hepatocellular carcinoma and markers of apoptosis (bcl-2, bax, bcl-x): prognostic significance. Appl Immunohistochem Mol Morphol. 2002; 10:210–17.
https://doi.org/10.1097/00129039-200209000-00004 PMID:12373145

45. Wong KF, Liu AM, Hong W, Xu Z, Luk JM. Integrin α2β1 inhibits MST1 kinase phosphorylation and activates Yes-associated protein oncogenic signaling in hepatocellular carcinoma. Oncotarget. 2016; 7: 77683–95.
https://doi.org/10.18632/oncotarget.12760 PMID:27765911

46. Guan Z, Cheng W, Huang D, Wei A. High MYBL2 expression and transcription regulatory activity is associated with poor overall survival in patients with hepatocellular carcinoma. Curr Res Transl Med. 2018; 66:27–32.
https://doi.org/10.1016/j.retram.2017.11.002 PMID:29274707

47. Dupuy D, Bertin N, Hidalgo CA, Venkatesan K, Tu D, Lee D, Rosenberg J, Svrzikapa N, Blanc A, Carnec A, Carvunis AR, Pulak R, Shingles J, et al. Genome-scale analysis of in vivo spatiotemporal promoter activity in Caenorhabditis elegans. Nat Biotechnol. 2007; 25: 663–68.
https://doi.org/10.1038/nbt1305 PMID:17486083

48. Schlitt T, Palin K, Rung J, Dietmann S, Lappe M, Ukkonen E, Brazma A. From gene networks to gene function. Genome Res. 2003; 13:2568–76.
https://doi.org/10.1101/gr.1111403 PMID:14656964

49. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. Nucleic Acids Res. 2004; 32:D277–80.
https://doi.org/10.1093/nar/gkh063 PMID:14681412

50. Yi M, Horton JD, Cohen JC, Hobbs HH, Stephens RM. WholePathwayScope: a comprehensive pathway-

based analysis tool for high-throughput data. BMC Bioinformatics. 2006; 7:30.
https://doi.org/10.1186/1471-2105-7-30
PMID:16423281

51. Draghici S, Khatri P, Tarca AL, Amin K, Done A, Voichita C, Georgescu C, Romero R. A systems biology approach for pathway level analysis. Genome Res. 2007; 17:1537–45.
https://doi.org/10.1101/gr.6202607
PMID:17785539

52. Jansen R, Greenbaum D, Gerstein M. Relating whole-genome expression data with protein-protein interactions. Genome Res. 2002; 12:37–46.
https://doi.org/10.1101/gr.205602
PMID:11779829

53. Binder H, Schumacher M. Comment on 'Network-constrained regularization and variable selection for analysis of genomic data'. Bioinformatics. 2008; 24:2566–68.
https://doi.org/10.1093/bioinformatics/btn412
PMID:18682424

54. Wei Z, Li H. A Markov random field model for network-based analysis of genomic data. Bioinformatics. 2007; 23:1537–44.
https://doi.org/10.1093/bioinformatics/btm129
PMID:17483504

55. Zhang JD, Wiemann S. KEGGgraph: a graph approach to KEGG PATHWAY in R and bioconductor. Bioinformatics. 2009; 25:1470–71.
https://doi.org/10.1093/bioinformatics/btp167
PMID:19307239

56. Spirin V, Mirny LA. Protein complexes and functional modules in molecular networks. Proc Natl Acad Sci USA. 2003; 100:12123–28.
https://doi.org/10.1073/pnas.2032324100
PMID:14517352

57. Prieto C, Risueño A, Fontanillo C, De las Rivas J. Human gene coexpression landscape: confident network derived from tissue transcriptomic profiles. PLoS One. 2008; 3:e3911.
https://doi.org/10.1371/journal.pone.0003911
PMID:19081792

Supplementary Figures
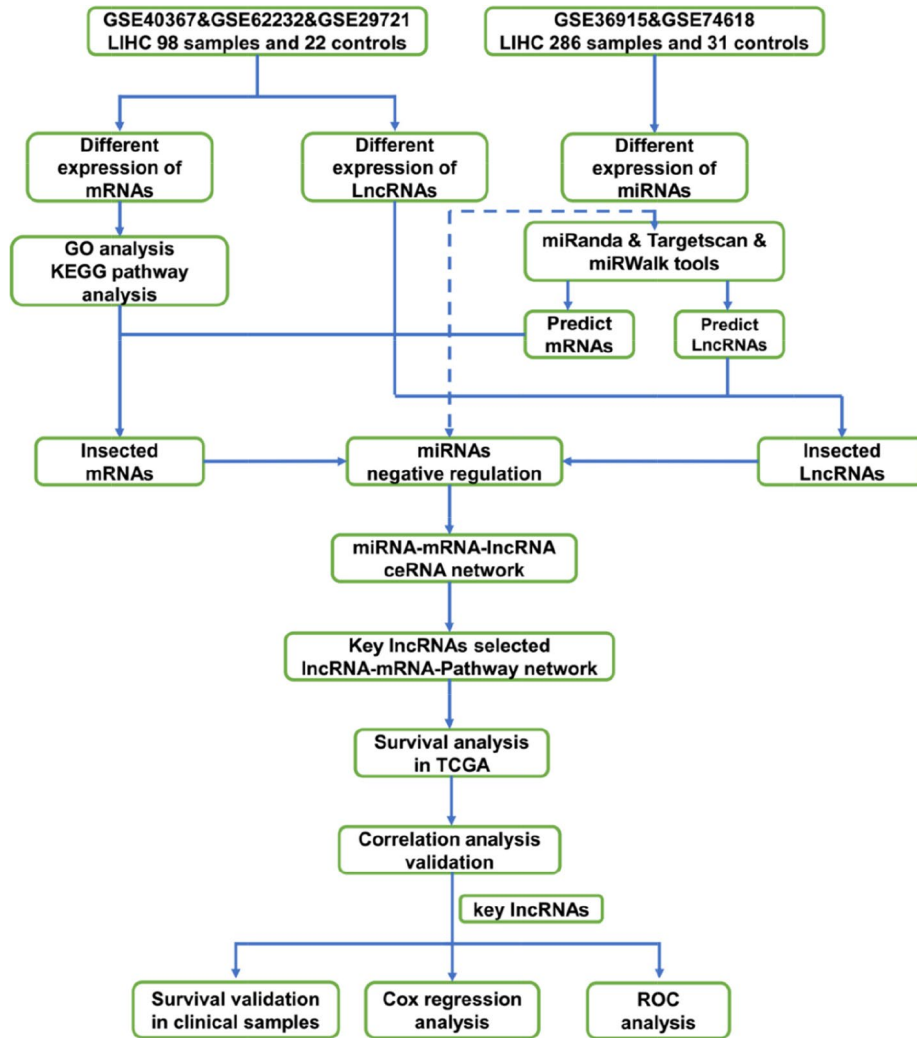
Supplementary Figure 1. Bioinformatics analysis flowchart.
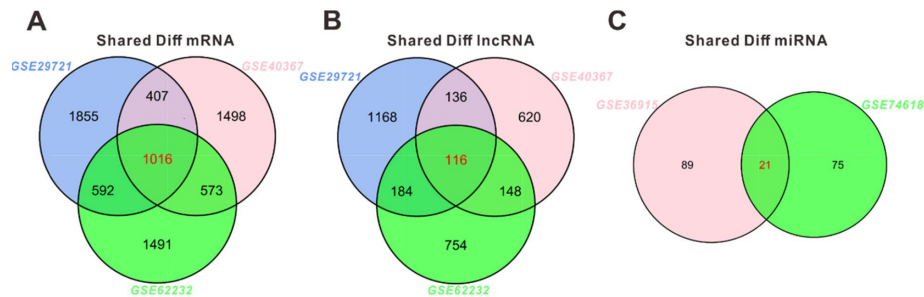
Supplementary Figure 2. Venn diagram analysis of the intersected DEGs, DELs and DEMs among different expression profiles. (**A**) The 1016 intersected DEGs obtained from GSE29721, GSE40367 and GSE62232. (**B**) The 116 intersected DELs obtained from GSE29721, GSE40367 and GSE62232. (**C**) The 21 intersected DEMs obtained from GSE36915 and GSE74618. Diff, Different.
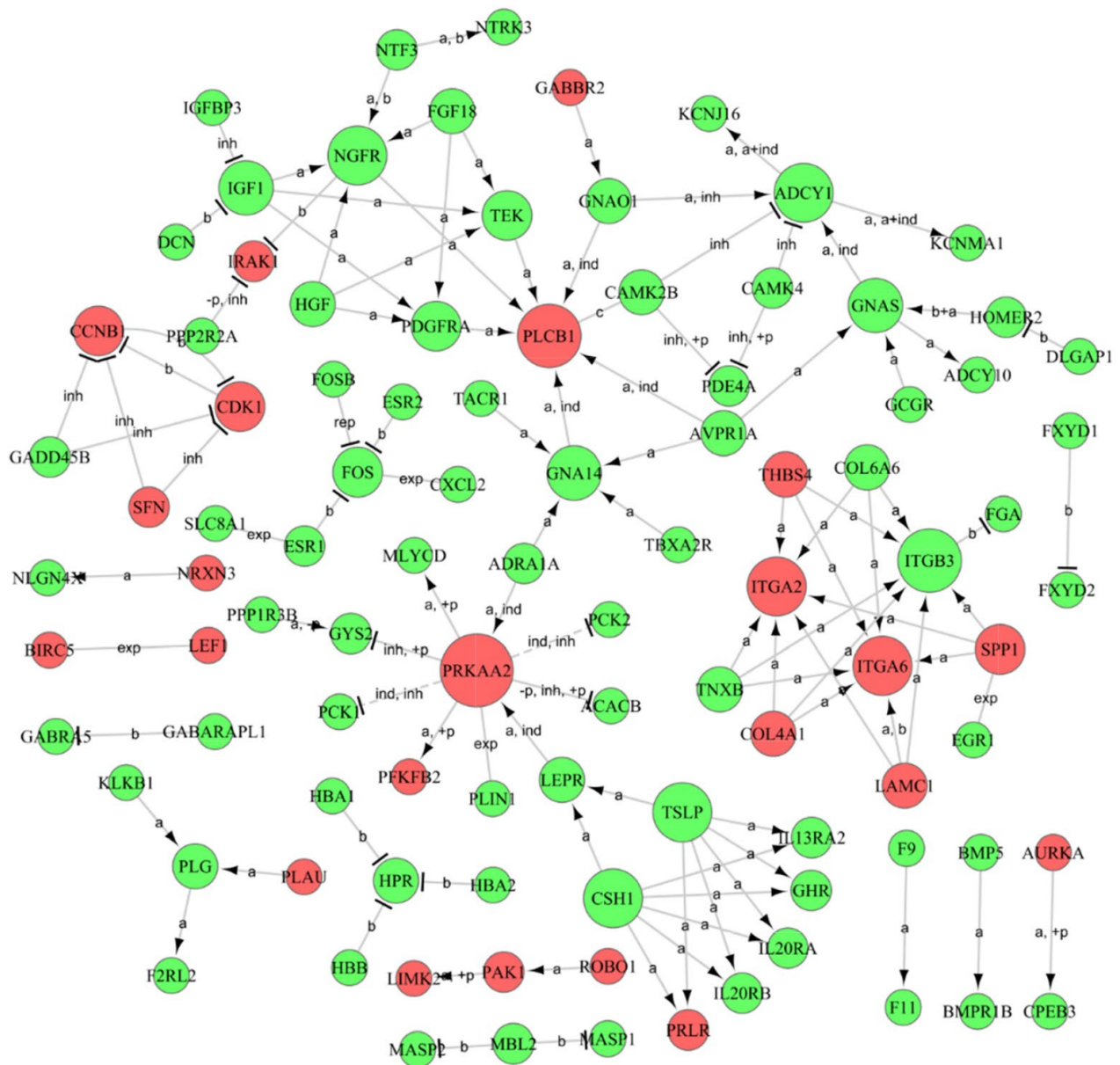
Supplementary Figure 3. Signal network of differentially expressed mRNAs genes. Red cycle nodes represent upregulated genes, and green cycle nodes represent downregulated genes. Edges represent interactions between genes (arrowheads represent targets). Interaction types: a, activation; b, binding/association; c, compound; exp, expression; ind, indirect effect; inh, inhibition; +p, phosphorylation; -p, dephosphorylation.
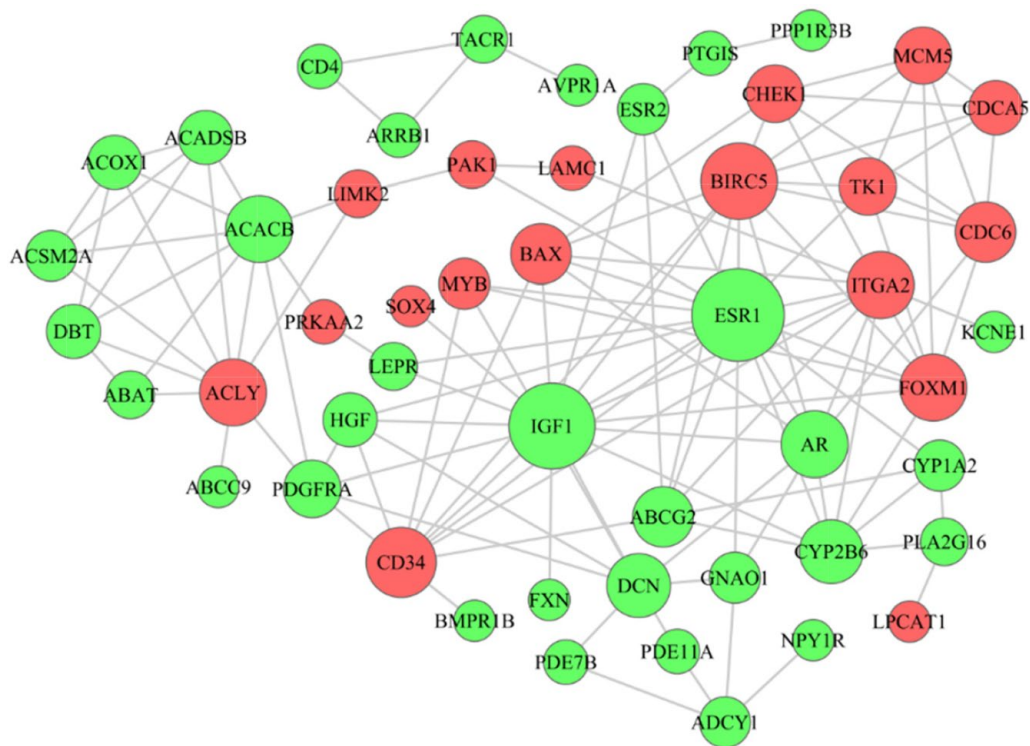
Supplementary Figure 4. PPI network contrasted by the differentially expressed genes identified in the ceRNA network. Red dots represent up-regulated genes and green dots represent down-regulated genes. The degree of the dots represents the ability to interact with other genes.
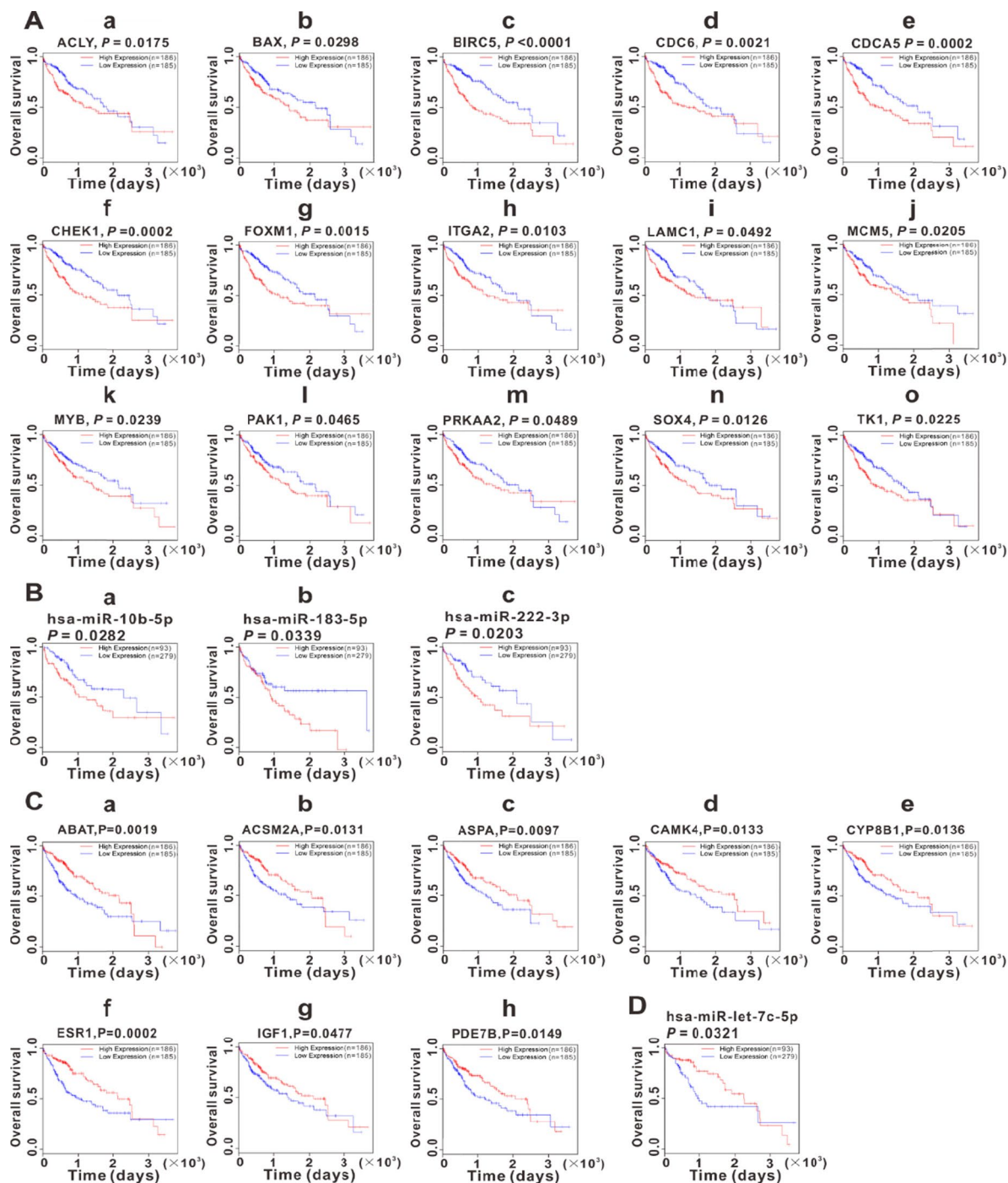
Supplementary Figure 5. Survival significance of the DEGs and DEMs in ceRNA network in HCC TCGA database. The Kaplan-Meier survival curves showed the significant correlation of DEGs and DEMs with OS time in HCC patients. The upregulated DEGs suggesting shorter OS were ACLY, BAX, BIRC5, CDC6, CDCA5, CHEK1, FOXM1, ITGA2, LAMC1, MCM5, MYB, PAK1, PRKAA2, SOX4, TK1 (**Aa–Ao**). The upregulated DEMs suggesting shorter OS were hsa-miR-10b-5p, hsa-miR-183-5p, and hsa-miR-222-3p (**Ba–Bc**). The upregulated DEGs suggesting longer OS were ABAT, ACSM2A, ASPA, CAMK4, CYP8B1, ESR1, IGF1, and PDE7B (**Ca–Ch**). The upregulated DEMs suggesting longer OS were hsa-let-7c-5p (**D**).
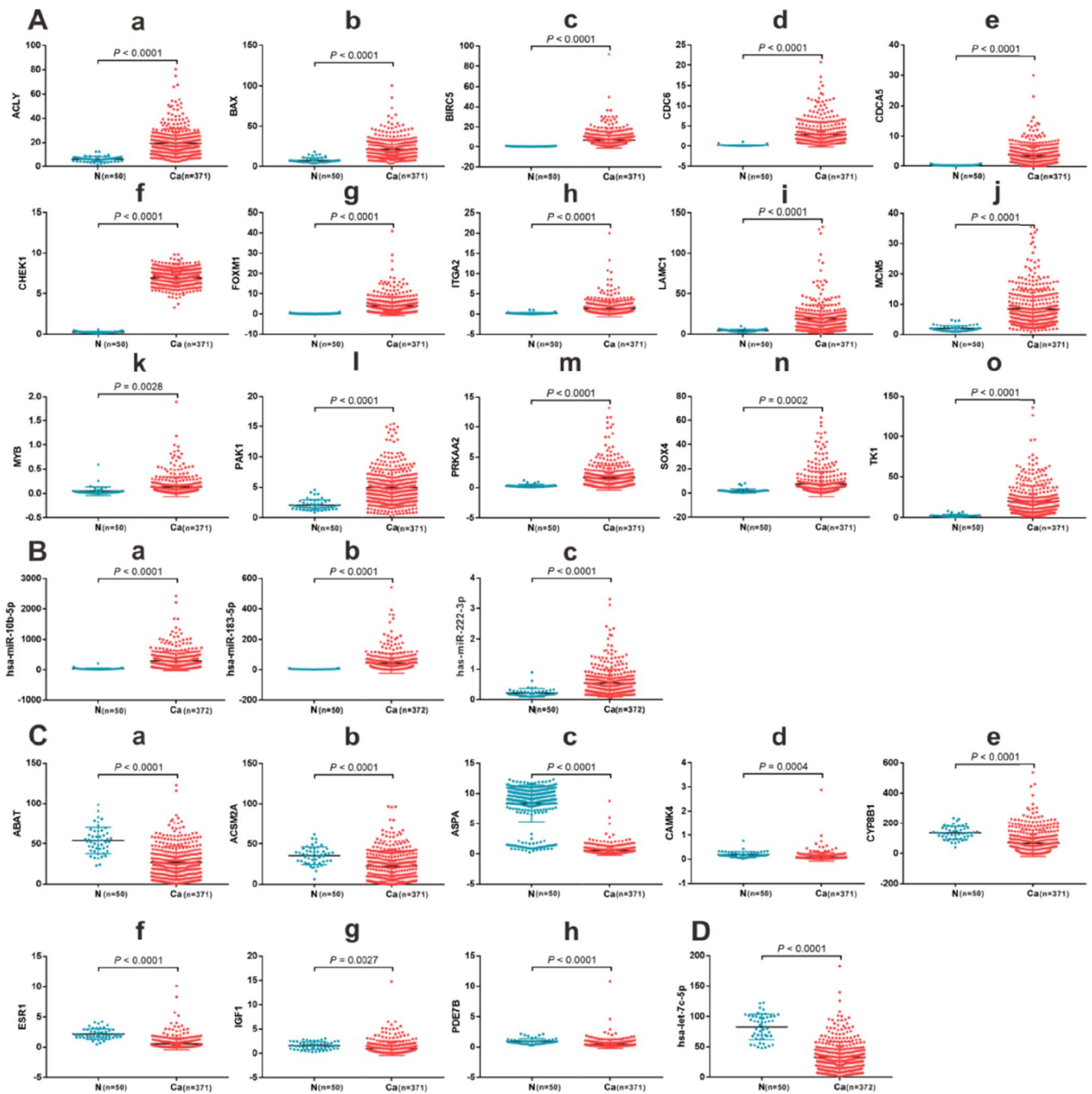
Supplementary Figure 6. The expression level in HCC TCGA database of the DEGs and DEMs in the ceRNA network. Upregulated DEGs in HCC tissues were ACLY, BAX, BIRC5, CDC6, CDCA5, CHEK1, FOXM1, ITGA2, LAMC1, MCM5, MYB, PAK1, PRKAA2, SOX4, TK1 (**Aa–Ao**). Upregulated DEMs in HCC tissues were hsa-miR-10b-5p, hsa-miR-183-5p, and hsa-miR-222-3p (**Ba–Bc**). Downregulated DEGs in HCC tissues were ABAT, ACSM2A, ASPA, CAMK4, CYP8B1, ESR1, IGF1, and PDE7B (**Ca–Ch**). Downregulated DEMs in HCC tissues were hsa-let-7c-5p (**D**).

Supplementary Tables

Supplementary Table 1. The top 10 upregulated and downregulated DELs in different expression profiles.

Profile	lncRNA	style	Fold Change	*P*-value	FDR
GSE29721	NONHSAT207339.1	up	13.563082	0.000629	0.158919
	NR_027687.1	up	10.552491	0.000232	0.097955
	NONHSAT142613.2	up	7.639081	0.00386	0.314951
	ENST00000389897.3	up	7.171353	0.000096	0.060247
	NONHSAT035498.2	up	6.544641	0.012906	0.433058
	LINC00348	up	6.469905	0.006573	0.385467
	NONHSAT165875.1	up	6.469905	0.006573	0.385467
	LINC01186	up	6.409546	0.00174	0.216378
	NONHSAT223306.1	up	6.409546	0.00174	0.216378
	SLC16A1-AS1	up	6.250394	0.001614	0.216378
	XR_430125.2	down	-30.4032	0.00042	0.062794
	LINC00844	down	-28.8462	0.002705	0.062794
	NONHSAT013506.2	down	-26.6209	0.002705	0.062794
	LINC01093	down	-26.0083	0.000273	0.062794
	NONHSAT099579.2	down	-22.2482	0.000273	0.062794
	A2MP1	down	-9.702393	0.000356	0.062794
	NONHSAT161826.1	down	-9.702393	0.000356	0.062794
	LOC105373764	down	-9.685478	0.00005	0.062794
	NONHSAT183166.1	down	-9.685478	0.00005	0.062794
	C3P1	down	-9.527459	0.001042	0.062794
GSE40367	NONHSAT150376.1	up	10.09334	0.002556	0.103234
	NONHSAT073632.2	up	9.626051	0.000564	0.05498
	ENST00000624682.1	up	8.135566	0.000031	0.003617
	ENST00000602609.1	up	6.688906	0.000346	0.02906
	NONHSAT093645.2	up	6.679915	0.009827	0.306808
	ENST00000602609.1	up	6.688906	0.000346	0.02906
	NONHSAT093645.2	up	6.679915	0.009827	0.306808
	OVOS2	up	6.635638	0.000953	0.075262
	NMRAL2P	up	6.128219	0.001682	0.103234
	NR_033752.2	up	6.128219	0.001682	0.103234
	NONHSAT024268.2	down	-15.590468	0.00014	0.004823
	NONHSAT121724.2	down	-13.397352	0.000076	0.004823
	XR_926922.1	down	-9.912165	0.000151	0.004823
	MAGI2-AS3	down	-9.341953333	0.000204667	0.007230667
	ENST00000562300.5	down	-9.189726	0.000515	0.014854
	LOC105370832	down	-9.189726	0.000515	0.014854
	LOC157273	down	-9.082558	0.000457	0.014854
	NONHSAT215396.1	down	-9.082558	0.000457	0.014854
	ENST00000417354.2	down	-8.714519	0.000317	0.011311
	FAM3D-AS1	down	-8.685381	0.000166	0.004823
GSE62232	NONHSAT214730.1	up	2.000594	0.000705	0.001847
	ENST00000588226.5	up	2.01129	0.000159	0.000548
	PARD6G-AS1	up	2.01129	0.000159	0.000548

NR_001578.1	up	2.014546	0.002505	0.005106
TDH	up	2.014546	0.002505	0.005106
NONHSAT177356.1	up	2.016526	0.00104	0.002528
LOC105371453	up	2.018036	0.000004	0.000026
XR_917739.1	up	2.018036	0.000004	0.000026
NONHSAT066299.2	up	2.03482	0.010096	0.015225
LOC105375172	up	2.035126	0.032603	0.036968
NONHSAT101639.2	down	-2.000372	0.000026	0.000122
PART1	down	-2.000372	0.000026	0.000122
LINC-PINT	down	-2.001105	0.039153	0.042336
NONHSAT213991.1	down	-2.001105	0.039153	0.042336
NR_110825.1	down	-2.001473	0.002129	0.004477
NONHSAT092521.2	down	-2.0025	0.000014	0.000072
LOC439933	down	-2.003334	0.038805	0.042034
NONHSAT150653.1	down	-2.003633	0.000053	0.000222
ENST00000578662.1	down	-2.00697	0.041058	0.043781
LOC101927539	down	-2.00697	0.041058	0.043781

Supplementary Table 2. The top 10 upregulated and downregulated DEMs in different expression profiles.

Profile	miRNA	style	Fold Change	*P*-value	FDR
GSE36915	hsa-miR-551b-3p	up	2.732694	0.000198	0.000553
	hsa-miR-96-5p	up	2.589663	0.000087	0.000269
	hsa-miR-10b-5p	up	2.469124	0.015412	0.014288
	hsa-miR-224-5p	up	2.435419	0.010374	0.010986
	hsa-miR-182-5p	up	2.327596	0.005243	0.006487
	hsa-miR-452-3p	up	2.301487	0.000233	0.000636
	hsa-miR-135a-5p	up	2.250943	0.009832	0.010825
	hsa-miR-515-5p	up	2.211741	0.001409	0.002642
	hsa-miR-520h	up	2.163997	0.002988	0.004427
	hsa-miR-501-5p	up	1.98213	0.003605	0.00509
	hsa-miR-203a-3p	down	-3.338125	0.000001	0.000005
	hsa-miR-375	down	-3.151798	0.00187	0.003151
	hsa-miR-1-3p	down	-2.831155	0.000031	0.000118
	hsa-miR-483-3p	down	-2.735831	0.000022	0.000092
	hsa-miR-144-3p	down	-2.731645	0.000047	0.00016
	hsa-miR-10a-5p	down	-2.715267	0.000311	0.00079
	hsa-miR-429	down	-2.536476	0.002207	0.003531
	hsa-miR-542-3p	down	-2.536104	0.000013	0.000062
	hsa-miR-30a-3p	down	-2.524737	0.000117	0.00035
	hsa-miR-200a-3p	down	-2.484562	0.000003	0.00002
GSE74618	hsa-miR-21-5p	up	4.65618	0.000011	0.000181
	hsa-miR-34a-5p	up	4.629343	0.000376	0.003044
	hsa-miR-155-5p	up	3.04827	0.00063	0.004483
	hsa-miR-155-5p	up	3.04827	0.00063	0.004483
	hsa-miR-532-5p	up	2.823342	0.000956	0.005882
	hsa-miR-4322	up	2.560282	0.00006	0.00075
	hsa-miR-502-3p	up	2.26894	0.000063	0.000762
	hsa-miR-1202	up	2.185917	0.001181	0.007133
	hsa-miR-1307-3p	up	2.185861	0.000303	0.002566
	hsa-miR-501-3p	up	2.140183	0.00004	0.000546
	hsa-miR-375	down	-4.040495	0.001517	0.008667
	hsa-miR-486-5p	down	-3.431948	0.000095	0.001033
	hsa-miR-378a-5p	down	-3.396752	0.000018	0.000272
	hsa-miR-497-5p	down	-2.592723	0.000066	0.000779
	hsa-miR-125b-2-3p	down	-2.469376	0.000417	0.003297
	hsa-miR-30a-3p	down	-2.173553	0.00001	0.00017
	hsa-miR-505-3p	down	-2.106977	0.000289	0.002497
	hsa-miR-122-3p	down	-2.074589	0.012504	0.043968
	hsa-miR-424-3p	down	-1.988536	0.001084	0.006604
	hsa-miR-192-3p	down	-1.956065	0.000069	0.000793

Supplementary Table 3. miRNAs and targeted intersected mRNAs in HCC.

miRNA	Gene symbol
hsa-let-7c-5p	ACOX1, AGXT2, ALDH6A1, C7,CYP8B1, DCN, DMD, EPHA7, ESR2, ETNK2, FXN, IDS，IGF1, LIMK2, MASP1, PPP2R2A ,SGCD, SLC10A1, TK1, VSIG4
hsa-miR-10a-5p	ACOX1, ADCY1, ALDH6A1, CAMK2B, ITGA2, MAT1A, PRKAA2, SGCD, TACR1
hsa-miR-10b-5p	ABCG2, BMPR1B, HGF, PDE11A, PDGFRA, PRKAA2, SGCD
hsa-miR-130a-3p	CAMK4，ESR1，FXN，GNAO1，LAMC1，PDGFRA，TRIM71
hsa-miR-182-5p	ACADSB，CAMK4，COL4A1，CYP1A2，DCN，ESR1，GNAO1，PDE11A，SLC4A4
hsa-miR-183-5p	ABAT，ABCC9，ACSM2A，AR，ARRB1，CAMK4，CYP2B6，DBT，FXN，GNAO1，I GF1，SLC4A4，STMN1
hsa-miR-195-5p	AR，BIRC5，CHEK1，GLS2，MASP1，MME，MYB，SEMA6D，SLC4A4
hsa-miR-199a-5p	AADAT，ABCG2，ADH4，AR，CBS，CHEK1，EPHA7，HOGA1，PCK1，PDE11A，PT GIS，SLC8A1
hsa-miR-203a-3p	FHL1，IDS，LAMC1，PLA2R1，PLCB1，PPP1R3B，SGCD
hsa-miR-214-3p	ACLY，ACOX1，ADCY1，APOA5，AR，BAAT，BAX，BCKDHB，CAMK2B，CDCA5 ，CYP2C19，CYP8B1，FOXM1，GNAO1，GPT2，IGF1，KMO，LMNA，MASP1，MAS P2，NGFR，PDE4A，PRKAA2，SEMA6D，SGCD，SOCS2，TACR1
hsa-miR-216a-5p	ACADSB，ADCY1，BCAT1，CYP8B1，DBT，ESR2，KCNE1，LEPR，PTGIS，SGCD
hsa-miR-222-3p	ACOX1，CAMK4，CD4，COL4A1，DBT，FXN，IGF1，KCNE1，PDE7B，PLA2G16，P RKAA2，SEMA6D，TACR1，TRIM71
hsa-miR-224-5p	AR，ARRB1，AVPR1A，DBT，ESR1，GLS2，PLCB1，PRKAA2，PTGIS
hsa-miR-30a-3p	ADCY1，AR，ARRB1，CAMK4，CD34，ESR1，HPGD，KCNE1，PDE11A，PRKAA2，SGCD，SLC1A1，ST6GAL2，TRIM71，USP25
hsa-miR-375	ADCY10，ADH1B，ADH4，AR，CYP2A6，LEPR，NCAM1，PDE2A，PLCB1，PRLR，S LC8A1，SMAD6，SPAM1，TACR1，TK1
hsa-miR-452-5p	ACACB，ASPA，CAMK4，IDS，IL20RA，NPY1R，PPP1R3B，PRKAA2，SEMA6D，SG CD，SLC4A4，SLC8A1
hsa-miR-455-5p	BAAT，CAMK4，HMGCL，HOGA1，KMO，MASP1，PAK1，VSIG4
hsa-miR-486-5p	ASS1，INMT，KMO，LPCAT1，SOX4
hsa-miR-497-5p	ADCY1，ALDH6A1，BIRC5，CAMK4，CDC6，ESR2，ITGA2，KCNK5，LAMC1，MAS P1，MCM5，MME，MYB，SGCD，TRIM71
hsa-miR-505-3p	ABAT，ACOX1，ACVR1C，ADCY1
hsa-miR-99a-5p	ACAA2，ADCY1，AR，CYP1A1，DAPK2，DBT，ESR2，FHL1，GHR，KCNMA1，LEP R，MASP1，PDE11A，PTGS2，SLC16A10，SLC4A4，TF，TNXB，TRIM71

Supplementary Table 4. miRNAs and targeted intersected lncRNAs in HCC.

lncRNA	miRNA
MCM3AP-AS1	hsa-miR-455-5p，hsa-miR-214-3p，hsa-miR-497-5p，hsa-miR-497-5p，hsa-miR-199a-5p，hsa-miR-30a-3p，hsa-miR-182-5p，hsa-miR-195-5p
GBAP1	hsa-miR-486-5p，hsa-miR-497-5p，hsa-miR-199a-5p，hsa-miR-182-5p，hsa-miR-224-5p，hsa-miR-195-5p，hsa-miR-30a-3p
SLC16A1-AS1	hsa-let-7c-5p，hsa-miR-497-5p，hsa-miR-183-5p，hsa-miR-130a-3p，hsa-miR-222-3p，hsa-miR-455-5p，hsa-miR-486-5p，hsa-miR-505-3p
NONHSAT172507.1	hsa-miR-10a-5p，hsa-miR-214-3p，hsa-miR-10b-5p，hsa-miR-505-3p，hsa-miR-199a-5p，hsa-miR-182-5p，hsa-miR-195-5p，hsa-miR-222-3p，hsa-miR-10a-5p，hsa-miR-455-5p
NONHSAT191112.1	hsa-miR-455-5p，hsa-miR-214-3p，hsa-miR-497-5p，hsa-miR-30a-3p，hsa-miR-182-5p，hsa-miR-195-5p
PCBP1-AS1	hsa-miR-455-5p，hsa-let-7c-5p，hsa-miR-497-5p，hsa-miR-199a-5p，hsa-miR-183-5p，hsa-miR-224-5p，hsa-let-7c-5p，hsa-miR-497-5p，hsa-miR-214-3p，hsa-miR-216a-5p，hsa-miR-203a-3p，hsa-miR-497-5p，hsa-miR-30a-3p，hsa-miR-10b-5p，hsa-miR-182-5p，hsa-miR-224-5p，hsa-miR-505-3p，hsa-miR-195-5p，hsa-miR-10a-5p，hsa-miR-216a-5p，hsa-miR-30a-3p
LINC01128	hsa-miR-214-3p，hsa-miR-224-5p，hsa-miR-183-5p，hsa-miR-99a-5p，hsa-miR-199a-5p，hsa-miR-497-5p，hsa-miR-216a-5p，hsa-miR-10b-5p，hsa-miR-505-3p，hsa-miR-195-5p，hsa-miR-30a-3p，hsa-miR-203a-3p，hsa-miR-497-5p，hsa-miR-10a-5p
HAND2-AS1	hsa-miR-216a-5p，hsa-miR-222-3p，hsa-miR-183-5p，hsa-let-7c-5p，hsa-miR-452-5p，hsa-miR-455-5p
MAGI2-AS3	hsa-let-7c-5p，hsa-miR-10a-5p，hsa-let-7c-5p，hsa-miR-30a-3p，hsa-miR-10b-5p，hsa-let-7c-5p，hsa-miR-183-5p，hsa-miR-195-5p，hsa-miR-222-3p，hsa-miR-455-5p，hsa-miR-452-5p，hsa-miR-130a-3p，hsa-miR-203a-3p
C3P1	hsa-miR-216a-5p，hsa-miR-224-5p，hsa-miR-224-5p
NONHSAT035498.2	hsa-let-7c-5p，hsa-miR-182-5p，hsa-miR-486-5p，hsa-miR-199a-5p，hsa-miR-452-5p
GAS5	hsa-let-7c-5p，hsa-miR-455-5p，hsa-miR-30a-3p，hsa-miR-452-5p
CYTOR	hsa-let-7c-5p，hsa-miR-455-5p，hsa-miR-182-5p，hsa-miR-497-5p
DIO3OS	hsa-miR-455-5p，hsa-miR-214-3p，hsa-let-7c-5p，hsa-miR-214-3p，hsa-miR-216a-5p，hsa-miR-224-5p，
HNF4A-AS1	hsa-miR-182-5p，hsa-miR-224-5p，hsa-miR-195-5p，hsa-miR-222-3p
NONHSAT202693.1	hsa-miR-452-5p，hsa-miR-224-5p，hsa-let-7c-5p，hsa-miR-10a-5p，hsa-miR-10b-5p，hsa-miR-30a-3p，hsa-miR-214-3p

Supplementary Table 5. Top 20 degree of genes in PPI network.

Gene symbol	Biotype	style	Degree
ESR1	mRNA	down	17
IGF1	mRNA	down	15
BIRC5	mRNA	up	12
CD34	mRNA	up	10
FOXM1	mRNA	up	9
ITGA2	mRNA	up	9
AR	mRNA	down	9
ACLY	mRNA	up	9
ACACB	mRNA	down	9
DCN	mRNA	down	8
CYP2B6	mRNA	down	8
ABCG2	mRNA	down	7
CDC6	mRNA	up	7
BAX	mRNA	up	7
CHEK1	mRNA	up	6
TK1	mRNA	up	6
PDGFRA	mRNA	down	6
MCM5	mRNA	up	6
DBT	mRNA	down	5
HGF	mRNA	down	5