

Co-Expression Network Revealed the Potential Regulatory Mechanism of Lncrnas in Relapse Hepatocellular Carcinoma.

Yuan Fang

Kunming Medical University First Affiliated Hospital

Yang Yang

Kunming Medical University First Affiliated Hospital

XiaoLi Zhang

Kunming Medical University First Affiliated Hospital

Na Li

Kunming Medical University

Bo Yuan

Kunming Medical University First Affiliated Hospital

Li Jin

Kunming Medical University

Sheng Bao

Kunming Medical University First Affiliated Hospital

MengGe Li

University of Science and Technology of China

Dan Zhao

Shanghai Jiao Tong University School of Medicine

LingRui Li

South China University of Technology

Zhong Zeng

Kunming Medical University First Affiliated Hospital

HanFei Huang (✉ huanghf@ydy.cn)

Kunming Medical University First Affiliated Hospital <https://orcid.org/0000-0002-0852-9596>

Primary research

Keywords: hepatocellular carcinoma, long non-coding RNA, WGCNA, Prognosis, lncRNA-mRNA network

Posted Date: April 12th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-387093/v1>

Abstract

Background:

The mechanistic basis for the relapse of hepatocellular carcinoma (HCC) remains poorly understood. Recent research has highlighted the important roles of long non-coding RNAs (lncRNAs) in HCC. However, there are only a few studies on lncRNAs associated with the relapse of HCC.

Methods:

We analyzed lncRNA and mRNA profiles in the GSE101432 dataset associated with HCC relapse. The differentially expressed lncRNAs and mRNAs were used to construct an lncRNA-mRNA co-expression network.

Weighted gene co-expression network analysis followed by Gene Ontology and Kyoto Encyclopedia of Genes and Genomes enrichment analyses were conducted on the database. Furthermore, correlation and survival analyses were performed using The Cancer Genome Atlas database, and the clinical samples were verified by qRT-PCR.

Results:

In this study, lncRNAs and mRNAs associated with HCC recurrence were identified. Two gene modules were found to be closely linked to HCC relapse. The functional enrichment analysis results of lncRNAs and co-expression mRNAs indicated that they were closely related to the recurrence of HCC. In addition, we verified that the overall survival and recurrence-free survival of these genes in HCC have survival prediction functions. In total, we identified and validated two lncRNAs (LINC00941 and LINC00668) and six mRNAs (LOX, MICB, OTX1, BAIAP2L2, KCTD17, NDUFA4L2) associated with HCC relapse.

Conclusion:

In summary, we identified the key gene modules and central genes associated with recurrent HCC, and constructed lncRNA-mRNA networks related to this cancer type. These results provide a foundation for future basic research on the mechanism of recurrent liver cancer.

1. Introduction

Liver cancer is the sixth most prevalent cancer in world.^[1] Furthermore, the incidence rate of liver cancer showed the greatest increase from 2007 to 2016 at 2–3% per year, although this rate has decreased as compared to that in previous years. For all stages combined, the 5-year relative survival rate is the lowest (18%) for liver cancers.^[2] Although patients receive curative treatment for early hepatocellular carcinoma (HCC), up to 70% of them may experience relapse in the liver after 5 years.^[3] Therefore, we aimed to explore the relapse mechanism of liver tumors and search for prognostic-associated long non-coding RNAs (lncRNAs), which are of great significance as they provide useful information for the diagnosis and prognosis of liver cancer.

lncRNAs are a group of non-coding RNAs that are more than 200 nucleotides in length and have no ability to encode proteins. An expanding body of evidence shows that lncRNAs modulate gene expression at multiple

levels through many complex mechanisms. In addition to their well-established role as regulators of transcription, lncRNAs are also effective regulators of pre-mRNA splicing, mRNA decay, and translation. [4] Portions of lncRNAs are specifically expressed in different tissues and cancers. [5] In addition, lncRNAs are involved in the pathological process of tumors and act via mechanisms, such as cis-regulation at enhancers, chromatin reprogramming, and post-transcriptional regulation of mRNA processing. [6]

The emergence of wide-ranging cancer projects, such as The Cancer Genome Atlas (TCGA) research network project and Gene Expression Omnibus (GEO) has provided an excellent opportunity to characterize lncRNAs in various cancers. Furthermore, bioinformatics analysis has been used to establish lncRNA features with prognostic value: for example, the identification of three lncRNA prognostic markers of ovarian cancer according to genome-wide copy number variation. [7] Additionally, a few survival-related lncRNA signatures have been discovered in malignant tumors of the digestive system, such as gastric cancer [8] and pancreatic cancer. [9] While a few reports have focused on the bioinformatics analysis of liver cancer, [10–12] few studies related to the recurrence of liver cancer can provide us.

In this study, we repurposed and integrated HCC data from TCGA and GEO to identify potential lncRNA signatures for predicting the survival of patients with relapse liver cancer. Combining the correlation analysis of lncRNAs and mRNA expression, it was discovered for the first time that lncRNAs are essential for the diagnosis, treatment, and prognosis of liver cancer relapse.

2. Materials And Methods

2.1 Data downloaded and preprocessing

The expression profile of genes associated with liver cancer (GSE101432) was obtained from the NCBI Gene Expression Omnibus database. Furthermore, the GEPIA 2 database was used to analyze the gene expression profiles from the Cancer Genome Atlas dataset and the Genotype-Tissue Expression projects. [13] Kaplan-Meier Plotter was used to analyze overall survival (OS) and recurrence-free survival (RFS). [14]

2.2 Read alignment and analysis of differentially expressed genes (DEGs)

The clean reads were aligned with the human GRch38 genome using tophat2, allowing up to 4 mismatches. [15] Uniquely mapped reads were ultimately used to calculate the read number and reads per kilobase million (RPKM) for each gene. The gene expression levels were evaluated based on the RPKM. The software edgeR, [16] which is specifically applied to analyze the differential expression of genes was used to screen the RNA-seq data for DEGs. A fold change (FC) ≥ 2 or ≤ 0.5 and a false discovery rate (FDR) ≤ 0.05 were used as cutoffs.

2.3 LncRNA prediction and direction identification

To identify lncRNAs associated with liver cancer relapse, we executed the pipeline described by Liu S et al. [17], which was based on the cufflinks software developed by Trapnell et al. [18]

2.4 Weighted gene co-expression network analysis (WGCNA) and co-expression analysis

The lncRNAs and mRNAs that were differentially expressed between primary tumors, relapsed tumors, and benign adjacent tissue in the dataset were utilized for WGCNA^[19] to reveal gene expression patterns. All expressed genes were considered input data. The eigengenes of each clustering module were utilized as the representative expression pattern of genes in each module. In order to investigate the regulatory relationship between lncRNAs and their target mRNAs, we calculated the Pearson's correlation coefficients (PCCs).

2.5 Functional enrichment analysis

Gene Ontology (GO) terms and KEGG pathways were analyzed to determine the functional categories associated with the identified DEGs using the KOBAS 2.0 server.^[20] Benjamini-Hochberg FDR controlling procedure and hypergeometric test were used to define the enrichment of each term. Reactome pathway analysis was also conducted for functional enrichment analysis of the selected genes sets.

2.6 qRT-PCR

To further verify our results, samples of tumors and matched adjacent tissue were obtained from 5 patients with primary HCC and 3 patients with relapsed liver cancer who underwent tumor resection at the First Affiliated Hospital of Kunming Medical University from July 15, 2019, to December 20, 2020. This study was approved by the Ethics Committee of the First Affiliated Hospital of Kunming Medical University. A signed informed consent form was obtained from each patient or their family members, who were provided a detailed explanation about the study. The total RNA samples were isolated using TRIzol reagent. Next, quantitative real-time reverse transcription-polymerase chain reaction (qRT-PCR) analysis was performed using SYBR Green Master Mix and an ABI Quant Studio 5 system, and changes in gene expression were calculated using the comparative CT approach. Primers used in the study are listed in Table 1.

2.7 Statistical analysis

Principal component analysis (PCA) analysis was employed by R software package factoextra (<https://cloud.r-project.org/package=factoextra>) for effective dimension reduction, pattern recognition, and exploratory visualization of high-dimensional data of the whole genome. After normalizing the reads by TPM (Tags Per Million) of each gene in samples, in house-script (sogen) was utilized for visualization of next-generation sequence (NGS) data and genomic annotations. The pheatmap package (<https://cran.r-project.org/web/packages/pheatmap/index.html>) in R was used to carry the clustering based on Euclidean distance. Student's t-test was used by SPSS22.0 for comparisons between two groups. For results, P value < 0.05 and a false discovery rate (FDR) < 0.25 were saw as statistically significant.

3. Results

3.1 A comprehensive catalog of lncRNA and mRNA genes in primary and relapse tumors and benign adjacent tissue.

Employing the edgeR software package, 4051 lncRNAs and 4968 mRNAs were found to be differentially expressed in the GSE101432 dataset, of which 1218 lncRNAs and 1676 mRNAs were upregulated in the primary tumor samples and 1540 lncRNAs and 1787 mRNAs were upregulated in the relapse tumor samples (Fig. 1).

3.2 Identification of differentially expressed (DE lncRNAs) and mRNAs (DE mRNAs) in the recurrence of liver cancer

Figure 2A shows a Venn diagram that illustrates the overlap between the relapse and primary tumors. Box plots for gene expression data were created to assess the distribution of DE lncRNAs (Fig. 2B). The analysis of the data revealed that a higher number of lncRNAs in the two groups were upregulated (68%). Next, principal component analysis (PCA) was performed to compare the DE lncRNAs and DE mRNAs between the primary tumor and relapse tumor groups (Fig. 2C, 2D). It can be clearly seen that lncRNA expression differs between the primary tumor and the relapse tumor groups. Therefore, lncRNAs could be used as important markers for liver cancer relapse. In addition, a Venn diagram was generated to identify the genes that were similar between all the groups. The intersection of combined co-upregulated and co-downregulated lncRNAs was identified between primary tumors and relapse tumors (co-upregulated = 508, co-downregulated = 216) (Fig. 2E).

3.3 Co-expression network and gene enrichment analysis

We used scatter plots to determine the co-expression of DE lncRNAs with DE mRNAs in relapse tumor samples (Fig. 3A). The top 10 Gene Ontology (GO) terms of the DE mRNAs co-expressed with the specific upregulated lncRNAs were identified using the KOBAS 2.0 server (Fig. 3B). These processes included homophilic cell adhesion, cartilage condensation, and cell-cell junction organization. We then constructed a co-expression network between specific upregulated lncRNAs and co-expressed DE mRNAs, involving the top 10 GO terms (Fig. 3C). We obtained the top three DE lncRNAs and co-expressed mRNAs which were specific up-regulated in relapse tumor and which were expressed in the primary and relapse tumors and benign adjacent tissue (Fig. 3D).

3.4 Identification of weighted gene co-expression network analysis (WGCNA) modules

We performed WGCNA analysis on all DE lncRNAs and DE mRNAs to obtain module-trait associations (Fig. 4A). Here, we focused on modules related to liver cancer relapse, including the pink, black, green-yellow, and yellow modules ($R > 0.5$). When comparing the fold change in lncRNAs and mRNAs with that in the control group in the four modules that were positively correlated in relapse liver cancer, all the genes were found to be upregulated in the four modules (Fig. 4B). We performed GO functional enrichment analysis on mRNAs co-expressed with lncRNAs in the four modules. The first 15 GO terms in the yellow and black modules were related to cell proliferation, differentiation, and survival. Therefore, the yellow and black modules were selected for further analysis. We generated an integrated lncRNA-mRNA regulatory network based on the hub genes identified within the yellow- and black-module hub genes (Fig. 4C, 4D).

3.5 Identification of hub lncRNAs based on key modules

Using the results of the co-expression network and WGCNA analyses, we constructed a bar graph of the expression level of selected lncRNAs from the GSE101432 dataset (Fig. 5A). Therefore, we selected RP11-334A14.8, RP4-738P11.4, TRBV6-6, LINC00668, and LINC00941, which showed higher relapse-tumor group than in the primary-tumor group for qRT-PCR. Taking into account TCGA database correlation analysis (Fig. 5B) and lncRNA qRT-PCR results, mRNA was selected for clinical sample verification (Fig. 5B).

3.6 Validation of lncRNA and regulated gene expression network in clinical samples

To directly validate a subset of these bioinformatics results, we assessed lncRNA and mRNA expression levels in a distinct cohort of the primary liver tumor, relapse liver tumor, and matched normal tissue samples via qRT-PCR (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$) (Fig. 6). Table 2 listed clinical samples relevant information. At the same time, according to the results of qRT-PCR, the expression levels of LOX, OTX1, MICB, NDUFA4L2, BAIAP2L2 and KCTD17 in recurrent HCC group were higher than which in primary HCC group, and had statistical significance in recurrent HCC.

3.7 Survival analysis of lncRNAs related to relapse liver cancer

The Kaplan-Meier plotter was used to assess the relationship between these genes and liver hepatocellular carcinoma patient survival, leading to the identification of prognosis-related lncRNAs (Fig. 7). The OS and RFS of LINC00668, LINC00941 and their co-expressed mRNA in HCC were statistically significant.

4. Discussions

Liver cancer is the second most common cause of cancer-related deaths.^[21] Surgical resection is the most effective first-line treatment for specific patients. The 5-year recurrence-free survival (RFS) rate after partial hepatectomy is only 48.4%.^[22, 23] For HCC, most studies that have published lncRNAs signatures associated with prognosis have focused on primary liver cancer, while few have focused on relapse liver tumors and thus, finding effective biomarkers for HCC is crucial.

Data mining strategies can be used to explore significant biological phenotypes associated with high-dimensional datasets. TCGA and GEO databases, with large-scale genomic analyses, can evaluate the molecular features associated with cancer outcomes. Recent developments in next-generation sequencing technologies have greatly expanded our knowledge on non-coding RNAs, and these non-coding RNAs are considerably more abundant than mRNAs.^[24] Many studies have revealed the role of lncRNAs in cancer development, indicating their potential as novel biomarkers for cancer diagnosis and prognosis.^[25-27] To better understand the molecular markers of relapse HCC, we comprehensively analyzed the database and identified lncRNAs that can be used to predict overall survival (OS) and RFS. Herein, we first retrieved GSE101432 dataset containing data pertaining to the HCC primary and relapse tumors, and matched non-tumor tissues from the GEO database, following which we analyzed the differential expression patterns of lncRNA and mRNA.

We identified 1218 lncRNAs and 1676 mRNAs upregulated in the primary tumor samples and 1540 lncRNAs and 1787 mRNAs upregulated in the relapse tumor samples. The results of PCA revealed a significant difference for recurrent liver cancer between lncRNA and mRNA. Additionally, we found that 1540 of the 2261 (68.1%) lncRNAs were upregulated and 508 genes were co-upregulated between the two groups, suggesting upregulated lncRNAs play a more important role in relapse HCC than downregulated lncRNAs.

lncRNA-mRNA co-expression combined with GO enrichment analysis showed that lncRNAs were related to cell adhesion, cell-cell junction organization, adherens junction organization, cell junction assembly, and pattern specification process, suggesting that lncRNAs are related to the development of cancer. In the relapse tumor

group, the co-expression network showed that upregulated lncRNA and co-expressed DE mRNA are involved in processes related to cancer metastasis; this creates conditions to promote the recurrence of liver cancer. CTD-2369P2.5 co-expressed protein coding gene ICAM1 that is associated with liver cancer has a high level of expression, and this expression is regulated by lncRNAs. In HCC, ICR specifically affects cancer stem cell properties of ICAM-1(+) HCC cells and lncRNA ICR contributes to portal vein tumor thrombus development. [28]

WGCNA technology has been leveraged to identify upregulated gene modules associated with HCC recurrence. Additionally, WGCNA has been used to analyze lncRNA and mRNA by clustering genes with similar expression patterns. [29] It considers the expression of all genes evaluated in the experiment to reveal co-expressed gene clusters (modules), which are likely also co-regulated. If some genes are co-expressed in the control group but not expressed in the pathological samples, it can be assumed that the regulatory mechanism has changed, which may be the cause or result of the disease. Therefore, the genes in these modules can play a role in cancer and are therefore considered potential therapeutic targets or diagnostic/prognostic biomarkers. [30] Herein, we employed WGCNA by assessing co-expressed gene networks in the GSE101432 dataset, leading to the identification of 15 gene co-expression modules. Based on the p- and R-values, we chose four modules namely: black, green-yellow, pink, and yellow, and performed GO enrichment analysis on them. Of these four modules, the yellow and black modules were most closely linked to cell proliferation, differentiation, and survival, as well as some transcription-related biological processes; this indicates that there are multiple abnormal cell activities in relapse HCC. These two modules were thus suspected to be crucial regulators of relapse HCC, and were therefore analyzed further.

Given that the co-expression network and WGCNA were found to be closely associated with the relapse of HCC, we next screened for hub genes that showed higher expression and correlation in relapse tumors. We found that the expression of RP11-334A14.8, RP4-738P11.4, TRBV6-6, LINC00668, and LINC00941 was higher in the relapse tumor group than in the primary tumor group. We then assessed the expression of these genes via qRT-PCR in a separate set of primary-HCC patient and relapse-HCC patient clinical samples, which confirmed that LINC00668 and LINC00941 expression levels were increased in both HCC tumor groups, whereas PHACTR2 expression was reduced in these samples relative to control tissue levels. LINC00668 and LINC00941 and their co-expression mRNA were selected using correlation analysis with $R > 0.1$ and $p < 0.05$ as the selection threshold values. Therefore, we selected BAIAP2L2, MDFI, ZNF750, and SLC44A5 as LINC00668 co-expression mRNAs. Besides, GPR160, LOX, NXPH4, OTX1, GMNN, and MICB were LINC00941 co-expression mRNAs. As a result, BAIAP2L2, MDFI, LOX, OTX1, and MICB were highly expressed in both the groups. This means that LINC00941 is co-expressed with LOX, OTX1, and MICB and is related to the recurrence of liver cancer. Furthermore, LINC00668 was found to be co-expressed with BAIA02L2, KCTD17, and NDUFA4L2, which has biological significance for the recurrence of liver cancer.

To the best of our knowledge, LINC00941 has been reported to exhibit protumorigenic and prometastatic behaviors during tumorigenesis, such as colorectal cancer (CRC) and gastric cancer. [31, 32] Wu et al. found that LINC00941 activates the TGF- β /SMAD2/3 axis in metastatic CRC, which provided new insight into the mechanism of metastatic CRC and a novel potential therapeutic target for advanced CRC. This is a potential marker for recurrent liver cancer. As with LINC00941 co-expression mRNAs, expression of OTX1 has been found to be positively correlated with nodal metastasis status ($p = 0.009$) and TNM staging ($p = 0.001$) in HCC tissues. [33] Moreover, overexpression of OTX1 promotes the proliferation, migration, invasion, and tumor angiogenesis

of HCC cells. [34] MICB is highly expressed in HCC, and its expression level is significantly and negatively associated with TNM stages. [35] Among the patients with different stages of hepatitis, asymptomatic individuals have higher MICB serum levels, while liver cirrhosis patients have lower MICB serum levels ($p < 0.0001$) compared to those in other patient groups. [36] The lysyl oxidase (LOX) family members are secreted copper-dependent amine oxidases, which are characterized by catalytic activity that contributes to the remodeling of the cross-linking of the structural extracellular matrix. [37] Umezaki et al. found that high LOX expression was associated with epithelial-mesenchymal transition (EMT) markers and predicted early recurrence and poor survival in patients with HCC. This is supported by our findings, which indicate that LINC00941 and co-expression mRNA are potential biomarkers and therapeutic targets for HCC recurrence.

However, another lncRNA, located at ch18p11.31, may play a pivotal role in the recurrence of liver cancer. In HCC, the molecular mechanism indicated that LINC00668 affects cell division, cell cycle, mitotic nuclear division, and drug metabolizing cytochrome P450 enzymes (all $p \leq 0.05$). [38] However, the co-expression mRNA, BAIAP2L2, was found to be highly expressed in gastric cancer tumors and its expression significantly correlated with tumor diameter, TNM stage, and lymph node metastasis, respectively. [39] KCTD17 is upregulated in the liver tissues of obese mice and nonalcoholic fatty liver disease patients; however, few studies have been published on its role in cancer. [40] NDUFA4L2 can promote cell migration, invasion, proliferation, and EMT of cancer cells under hypoxic conditions. [41]

In terms of survival analysis, whether in OS or RFS, both LINC00941 and LINC00668, as well as the co-expression mRNA, may be representative survival markers. In conclusion, our present research developed two lncRNA signatures for predicting the prognosis of patients with HCC relapse. Our study indicates that the LINC00941 and LINC00668 signatures may serve as candidate prognostic biomarkers and targets for new therapies in human gastric cancer.

While the results of this research are interesting and meaningful, they have some limitations. First, some lncRNAs, such as XLOC and TRBV6-6, lacked relevant information in TCGA or other databases; thus, we were unable to conduct further research on them. Second, the WCGNA method is inherently limited by the criteria used for module selection and the network rejection threshold, which may affect the final research results.

Declarations

Availability of data and materials:

The dataset GSE101432 download form

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE101432>.

Author Contributions:

HanFei Huang MD. and Zhong Zeng MD.: Conceptualization and financial support; Yuan Fang: Database mining, sample collection, preparation of original drafts and chart construction; Yang Yang, XiaoLi Zhang: contributed equally with first author to this work. LingRui Li, Li Na, Bo Yuan, MengGe Li and Dan Zhao: Technical support and manuscript revision. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest:

The authors declare no conflict of interest.

Funding:

We thank the financial support from the National Natural Science Foundation of China (No. 81960123 and 81760119).

Ethics approval and consent to participate:

This study was approved by the Ethics Committee of the First Affiliated Hospital of Kunming Medical University.

Consent for publication: Not applicable.

References

1. Forner A, Llovet J, Bruix J. Hepatocellular carcinoma. *Lancet*. 2012;379(9822):1245–55.
2. Siegel R, Miller K, Jemal A, *Cancer statistics, 2020*. CA: a cancer journal for clinicians, 2020. **70**(1): p. 7–30.
3. Brown Z, Greten T, Heinrich B, *Adjuvant Treatment of Hepatocellular Carcinoma: Prospect of Immunotherapy*. *Hepatology* (Baltimore, Md.), 2019. **70**(4): p. 1437–1442.
4. Yoon J, Abdelmohsen K, Gorospe M. Posttranscriptional gene regulation by long noncoding RNA. *Journal of molecular biology*. 2013;425(19):3723–30.
5. Huang X, Xiao R, Pan S, et al. Uncovering the roles of long non-coding RNAs in cancer stem cells. *J Hematol Oncol*. 2017;10(1):62.
6. Iyer M, Niknafs Y, Malik R, et al. The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet*. 2015;47(3):199–208.
7. Zheng M, Hu Y, Gou R, et al., *Identification three LncRNA prognostic signature of ovarian cancer based on genome-wide copy number variation*. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie*, 2020. **124**: p. 109810.
8. Ma B, Li Y, Ren Y. Identification of a 6-lncRNA prognostic signature based on microarray re-annotation in gastric cancer. *Cancer medicine*. 2020;9(1):335–49.
9. Zhang H, Zhu M, Du Y, et al. A Panel of 12-lncRNA Signature Predicts Survival of Pancreatic Adenocarcinoma. *J Cancer*. 2019;10(6):1550–9.
10. Sun Y, Zhang F, Wang L, et al. A five lncRNA signature for prognosis prediction in hepatocellular carcinoma. *Mol Med Rep*. 2019;19(6):5237–50.
11. Shi Y, Li Y, Lin J, et al., *The discovery of a novel eight-mRNA-lncRNA signature predicting survival of hepatocellular carcinoma patients*. *Journal of cellular biochemistry*, 2018.
12. Zhao Q, Zhang J, Xu L, et al. Identification of a five-long non-coding RNA signature to improve the prognosis prediction for patients with hepatocellular carcinoma. *World journal of gastroenterology*. 2018;24(30):3426–39.
13. Tang Z, Kang B, Li C, et al. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic acids research*. 2019;47:W556–60.

14. Nagy Á, Lániczky A, Menyhárt O, et al. Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. *Scientific reports*. 2018;8(1):9227.
15. Kim D, Pertea G, Trapnell C, et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology*. 2013;14(4):R36.
16. Robinson M, McCarthy D, Smyth G. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–40.
17. Liu S, Wang Z, Chen D, et al. Annotation and cluster analysis of spatiotemporal- and sex-related lncRNA expression in rhesus macaque brain. *Genome research*. 2017;27(9):1608–20.
18. Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols*. 2012;7(3):562–78.
19. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform*. 2008;9:559.
20. Xie C, Mao X, Huang J, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic acids research*. 2011;39:W316-22.
21. Castelli G, Pelosi E, Testa U. *Liver Cancer: Molecular Characterization, Clonal Evolution and Cancer Stem Cells*. *Cancers*, 2017. 9(9).
22. Lafaro K, Grandhi M, Herman J, et al. The importance of surgical margins in primary malignancies of the liver. *Journal of surgical oncology*. 2016;113(3):296–303.
23. Liu H, Wang Z, Fu S, et al. Randomized clinical trial of chemoembolization plus radiofrequency ablation versus partial hepatectomy for hepatocellular carcinoma within the Milan criteria. *Br J Surg*. 2016;103(4):348–56.
24. Shi X, Sun M, Liu H, et al. Long non-coding RNAs: a new frontier in the study of human diseases. *Cancer letters*. 2013;339(2):159–66.
25. Xiao M, Liu J, Xiang L, et al. MAFG-AS1 promotes tumor progression via regulation of the HuR/PTBP1 axis in bladder urothelial carcinoma. *Clinical translational medicine*. 2020;10(8):e241.
26. Tai Z, Ma J, Ding J, et al. MEG3Aptamer-Functionalized Dendrimer Delivery of Plasmid-Encoding lncRNA Enhances Gene Therapy in Castration-Resistant Prostate Cancer. *Int J Nanomed*. 2020;15:10305–20.
27. Ofoeyeno N, Ekpenyong E, Braconi C. *Pathogenetic Role and Clinical Implications of Regulatory RNAs in Biliary Tract Cancer*. *Cancers*, 2020. 13(1).
28. Guo W, Liu S, Cheng Y, et al. ICAM-1-Related Noncoding RNA in Cancer Stem Cells Maintains ICAM-1 Expression in Hepatocellular Carcinoma. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2016;22(8):2041–50.
29. Weng W, Zhang Z, Huang W, et al. Identification of a competing endogenous RNA network associated with prognosis of pancreatic adenocarcinoma. *Cancer cell international*. 2020;20:231.
30. Giulietti M, Occhipinti G, Righetti A, et al. Emerging Biomarkers in Bladder Cancer Identified by Network Analysis of Transcriptomic Data. *Frontiers in oncology*. 2018;8:450.
31. Wu N, Jiang M, Liu H, et al., *LINC00941 promotes CRC metastasis through preventing SMAD4 protein degradation and activating the TGF-β/SMAD2/3 signaling pathway*. *Cell death and differentiation*, 2020.
32. Liu H, Wu N, Zhang Z, et al. LINC00941 Long Non-coding RNA as a Potential Biomarker Promotes the Proliferation and Metastasis of Gastric Cancer. *Frontiers in genetics*. 2019;10:5.

33. Li H, Miao Q, Xu C, et al. OTX1 Contributes to Hepatocellular Carcinoma Progression by Regulation of ERK/MAPK Pathway. *J Korean Med Sci.* 2016;31(8):1215–23.
34. Hu Z, Li H, Teng F, et al. Long noncoding RNA MAFG-AS1 facilitates the progression of hepatocellular carcinoma via targeting miR-3196/OTX1 axis. *Eur Rev Med Pharmacol Sci.* 2020;24(23):12131–43.
35. Fang L, Gong J, Wang Y, et al. MICA/B expression is inhibited by unfolded protein response and associated with poor prognosis in human hepatocellular carcinoma. *Journal of experimental clinical cancer research: CR.* 2014;33:76.
36. Kohga K, Takehara T, Tatsumi T, et al. Serum levels of soluble major histocompatibility complex (MHC) class I-related chain A in patients with chronic liver diseases and changes during transcatheter arterial embolization for hepatocellular carcinoma. *Cancer Sci.* 2008;99(8):1643–9.
37. Lin H, Li C, Yang Y, et al., *Roles of Lysyl Oxidase Family Members in the Tumor Microenvironment and Progression of Liver Cancer.* *International journal of molecular sciences,* 2020. 21(24).
38. Wang X, Zhou X, Liu J, et al. Genome-wide investigation of the clinical implications and molecular mechanism of long noncoding RNA LINC00668 and protein-coding genes in hepatocellular carcinoma. *Int J Oncol.* 2019;55(4):860–78.
39. Liu J, Shangguan Y, Sun J, et al., *BAIAP2L2 promotes the progression of gastric cancer via AKT/mTOR and Wnt3a/β-catenin signaling pathways.* *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie,* 2020. **129**: p. 110414.
40. Kim K, Ryu D, Dongiovanni P, et al. Degradation of PHLPP2 by KCTD17, via a Glucagon-Dependent Pathway, Promotes Hepatic Steatosis. *Gastroenterology.* 2017;153(6):1568–80.e10.
41. Xu W, Yang R, Zheng H, et al. NDUFA4L2 Regulated by HIF-1α Promotes Metastasis and Epithelial-Mesenchymal Transition of Osteosarcoma Cells Through Inhibiting ROS Production. *Frontiers in cell developmental biology.* 2020;8:515051.

Tables

Table 1 Primers for Validated Genes

Gene	lncRNA ID/GENE ID	Prime sequence (5'-3')	
		Forward	Reverse
RP11-334A14.8	NONHSAT003288.2	AGGGAACCCCATTCTTCCAA	AGAGGTCTCAAGGCACAGAT
LINC00668	NONHSAT057143.2	GGGTCCAAGGGATCTGCAAG	CCGCCCAAGATTCTCTAGC
CTD-2369P2.5	NONHSAT061043.2	GCACTTTCCCAAGGGTACAA	CCCTTGACCCTAACCTCTGA
LINC00941	NONHSAT027514.2	CATGTACGTTGCTATCCAGGC	CTCCTTAATGTCACGCACGAT
AF124730.4	NONHSAT081627.2	GCCACACAAATGTTCAACCAC	GAGCAGAAGGACTCAGTGTG
RP13-143G15.4	NONHSAT115109.2	ACCCTTGCAACCTTTCTCAG	TGGCCACCTTGTAGGTTAGA
RP11-973F15.2	NONHSAT128496.2	GACAGCCCCCAAAGAACC	ATCCCCGGACTGCTGC
RP5-967N21.11	NONHSAT188398.1	TGGTATGTTTGTGTTCCATTTGC	AATGCGAGCACAAGATAATTTTT
RASAL1	8437	CAGCTCCCTGAATGTTTCGC	TCCTCATCCAGCACGTAGAAG
NDUFA4L2	56901	CCTGAGCCCCAATGACCAATA	TCTGGCCGGTCCTTCTTCA
KCTD17	79734	AGTCGGACCGGGATGAGAC	CCATGCCGGAGGAAGTTCA
BAIAP2L2	80115	GCCGAGGTCTACTTCAGTGC	GCTGGGTGTCAGACATCTGC
MDFI	4188	CCACCGGAAGTTGCAGACA	GGAACTCGCAGAACAGGCA
SLC44A5	204962	GACTTTGGTGATCCAAGGACAT	GTCTGTAGGATAGGCTGCTCT
GPR160	26996	CCAGCCATCTACCAAAGCCTG	GCCAGTAACTCTGAATGCTGACA
LOX	4015	GCCGACCAAGATATTCCTGGG	GCAGGTCATAGTGGCTAAACTC
NXPH4	11247	AAGGTCTTCGGACGGCCTA	GCAGCGAAAACCTTGAGGGTAT
OTX1	5013	GCGTCGTCGCTGAGTACAC	ACATGGGATAAGAGGCTGCTG
GMNN	51053	GCCCTGGGGTTATTGTCCC	AGGCCTTTCTCCGTTTTTCT
MICB	4277	TCTTCGTTACAACCTCATGGTG	TCCCAGGTCTTAGCTCCCAG

Table 2 Clinical information

	Sample
Primary Tumor	5
Relapse Tumor	3
Age, median, (mean)	54
Gender, male	4 (50%)
Length of hospital stay (days) (mean)	10-14 (12.875)
Liver cirrhosis	4 (50%)
Classification of liver cancer	
Hepatocellular carcinoma	8 (100%)
Intrahepatic cholangiocarcinoma	0
Tumor diameter	
<2cm	1 (12.5%)
2cm-5cm	5 (62.5%)
>5cm	2 (25.0%)
HBV	
Positive	4 (50%)
Negative	4 (50%)
AFP	
<400	4 (50%)
>400	4 (50%)
Tumor location	
Left liver	5 (62.5%)
Right liver	3 (37.5%)
Tumor recurrence time (months)	18,27,76
Outcome	
Cure	1 (12.5%)
Improve	6 (75%)
Death	1 (12.5%)

Figures

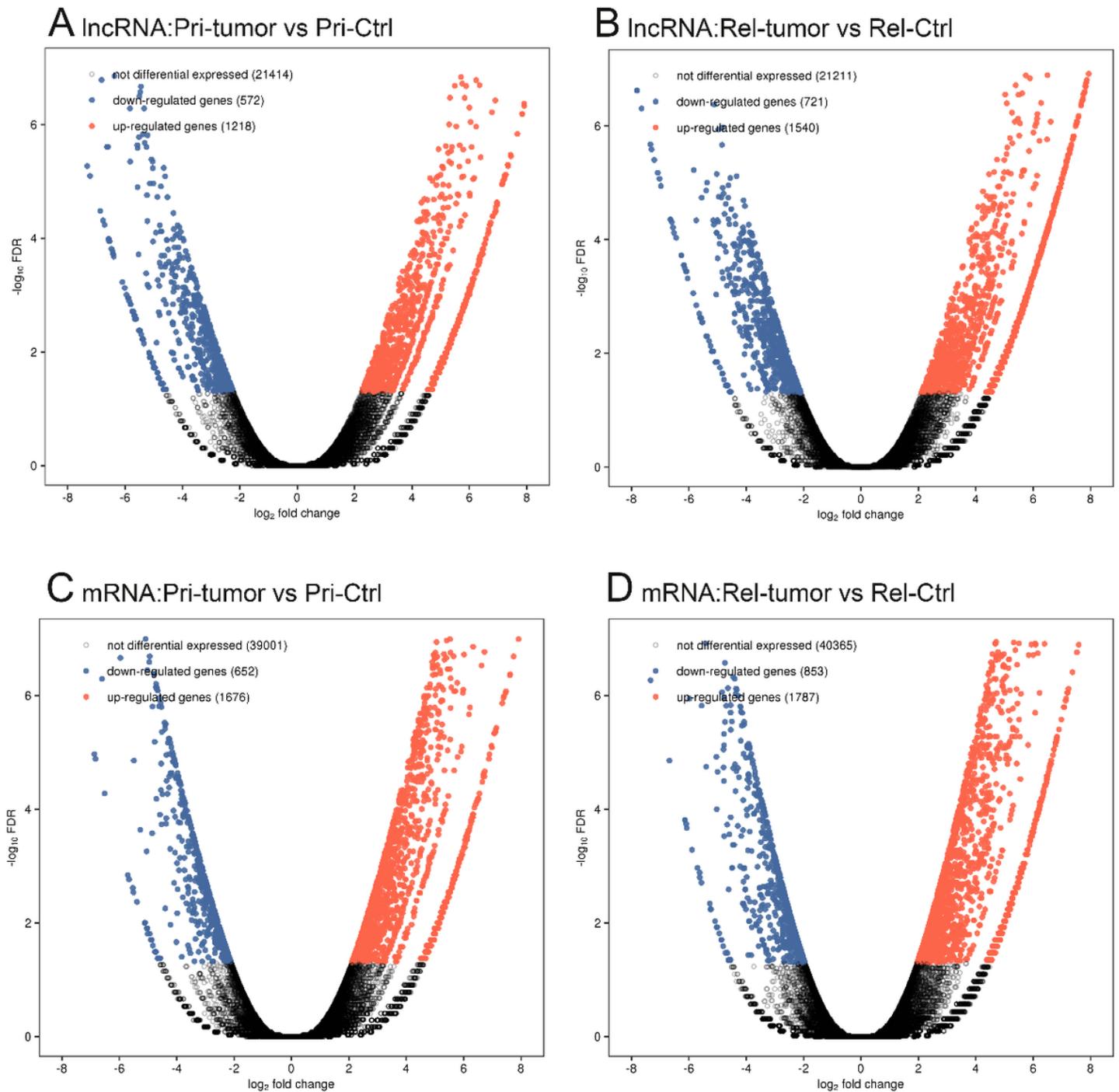


Figure 1

A comprehensive catalog of lncRNA genes in primary tumor, relapse tumor, benign adjacent tissue. A: Volcano plot of the lncRNAs in primary tumor and benign adjacent tissue. B: Volcano plot of the lncRNAs in relapse tumor and benign adjacent tissue. C: Volcano plot of the mRNAs in primary tumor and benign adjacent tissue. D: Volcano plot of the mRNAs in relapse tumor and benign adjacent tissue.

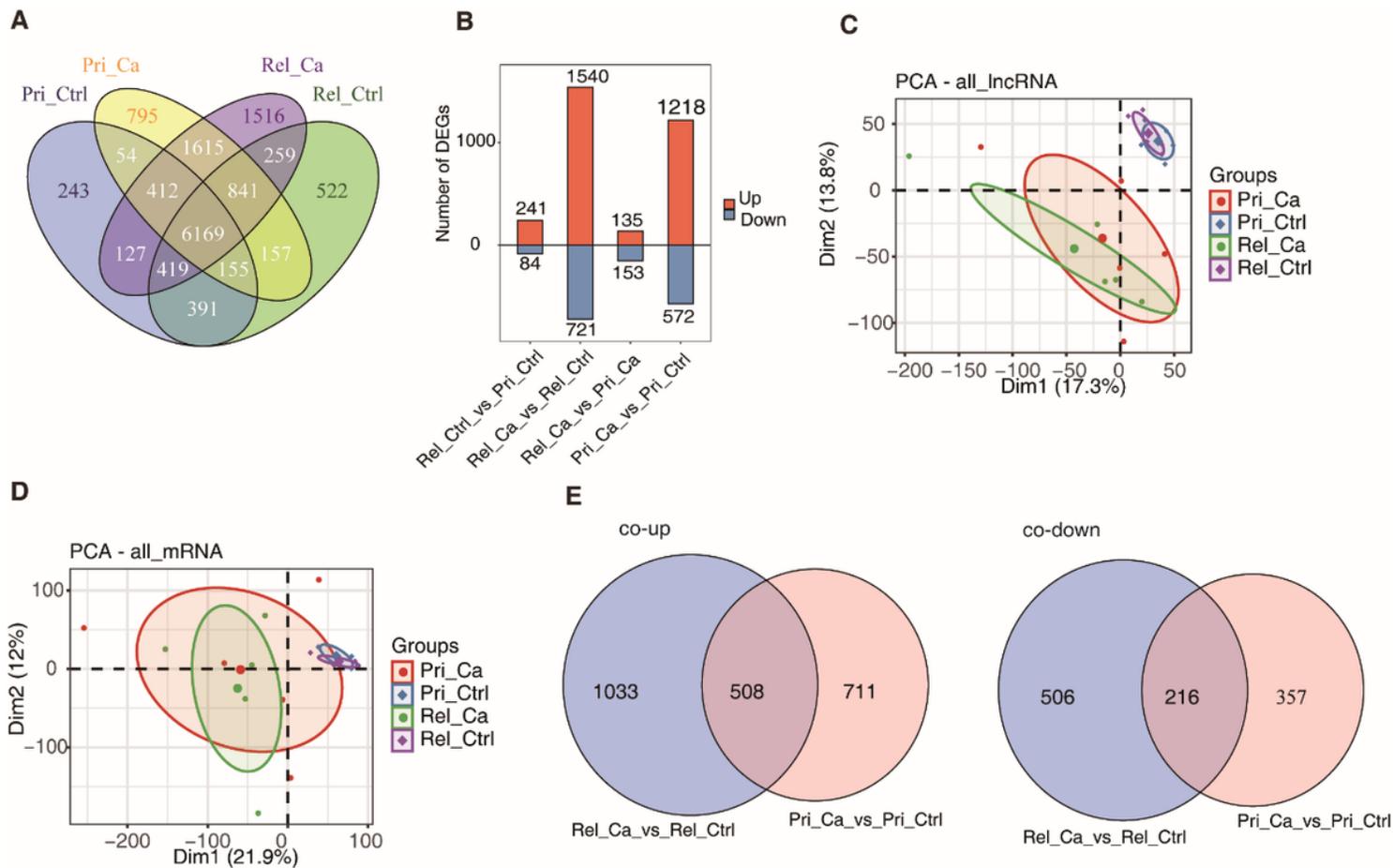


Figure 2

Analysis of differential expression of lncRNA and mRNA in HCC and adjacent tissues. A. Venn diagram of detected lncRNA genes in primary tumor, relapse tumor, benign adjacent tissue. At least two samples with RPKM \geq 0.2 was considered to be detected in the group. B. The number of differentially expressed (DE) lncRNAs among different groups. Bar plot showing the number of up-regulated and down-regulated DE lncRNAs. C-D. Principal component analysis (PCA) of primary tumor, relapse tumor, benign adjacent tissue based on normalized lncRNAs (C) and mRNAs (D) expression level. The samples were grouped by disease state and the ellipse for each group is the confidence ellipse. E. Venn diagram of detected co-up-regulated (left panel) and co-down-regulated (right panel) lncRNAs in primary tumor or relapse tumor compared with benign adjacent liver samples.

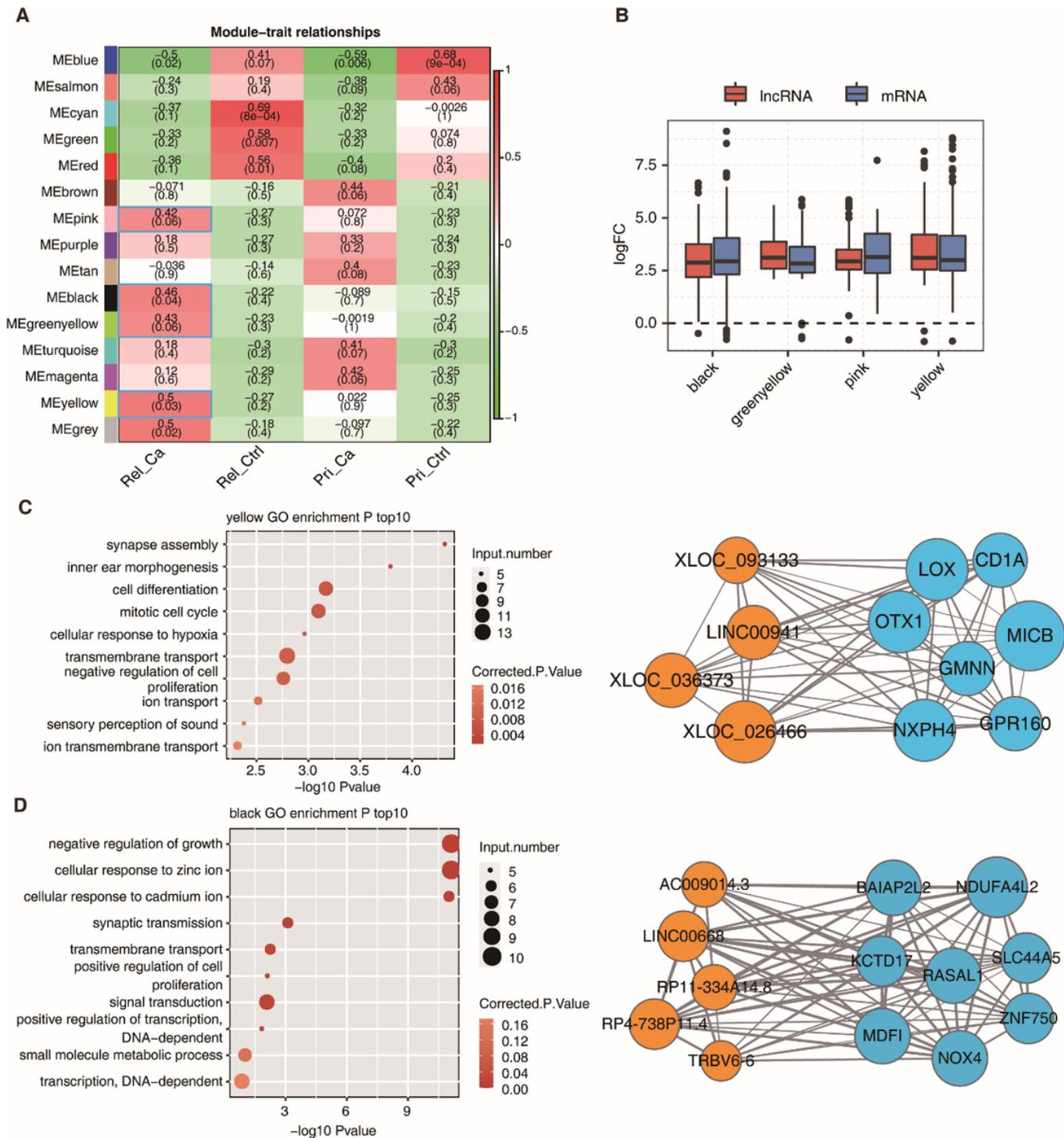


Figure 4

WGCNA analysis of all expressed lncRNAs and mRNAs. A. Module-trait associations as computed by an LME model with all factors on the x axis used as covariates. All Pearson's correlation value and p values are displayed. Module-trait associations with R and correlation > 0.4 were framed with blue box. B. Boxplot showing expression fold change of mRNAs and lncRNAs from the four relapse associated modules. C-D. Module plots displaying the top hub genes along with the GO terms enrichment of each of the three modules, MEyellow (C), MEblack (D). Orange circles indicate lncRNAs and lightblue circles indicate mRNAs.

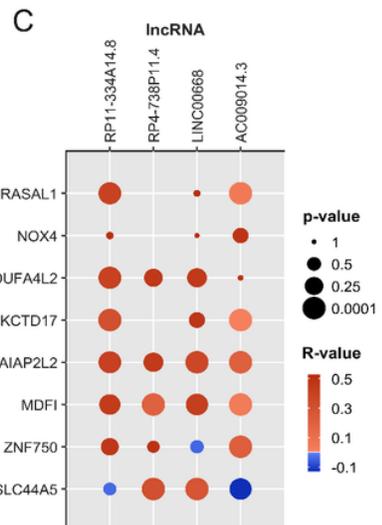
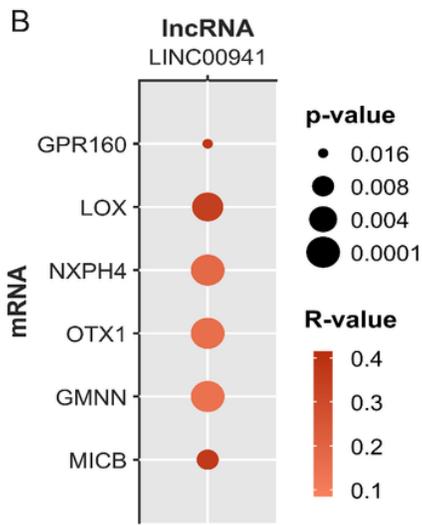
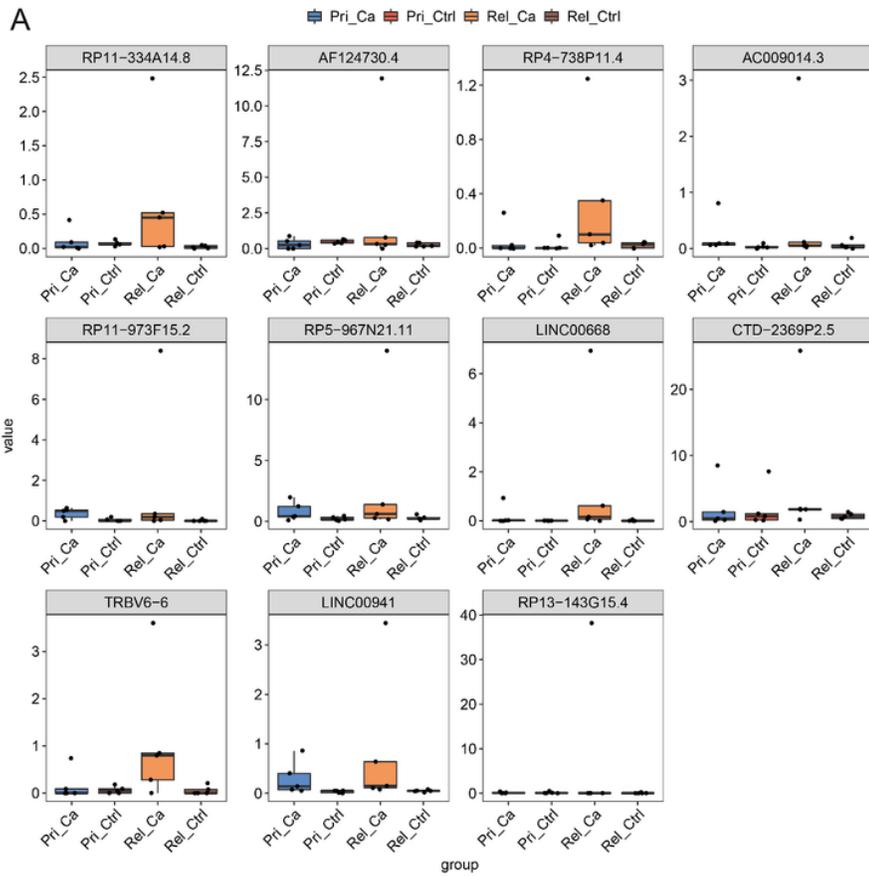


Figure 5

Survival analysis of genes which were selected using TCGA database. A. The results of co-expression network and WGCNA selected the expression level of lncRNA in GSE101432. B-C. The bubble chart was used to show the correlation analysis between lncRNA and co-expressed mRNA.

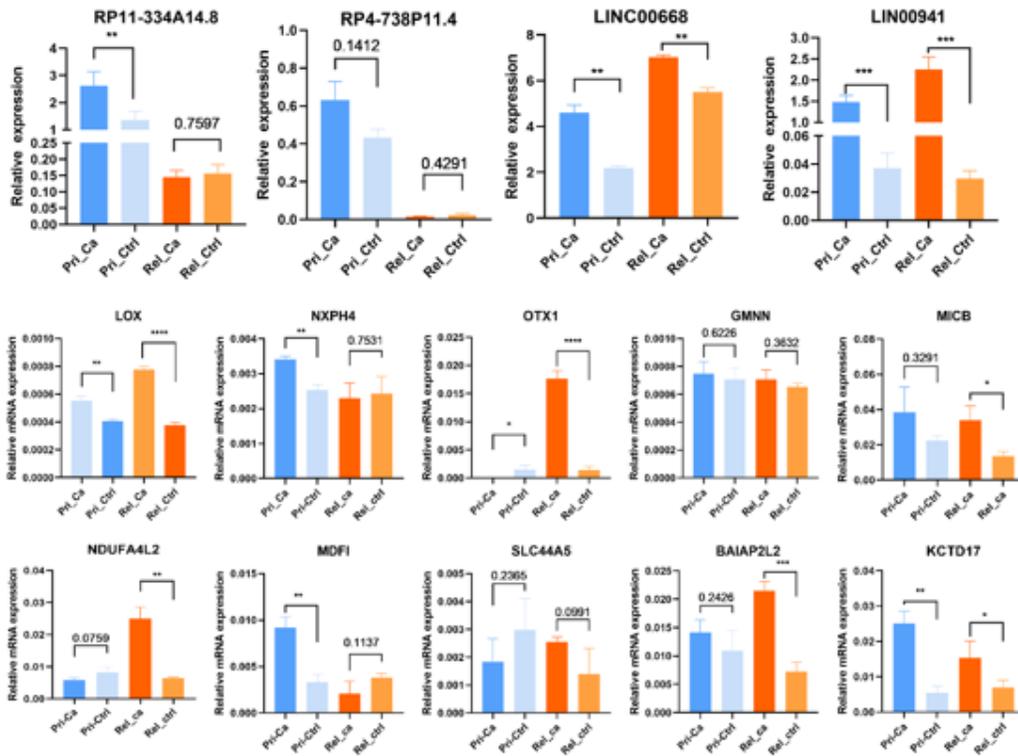


Figure 6

Relative expression levels of lncRNAs and mRNAs in clinical samples.

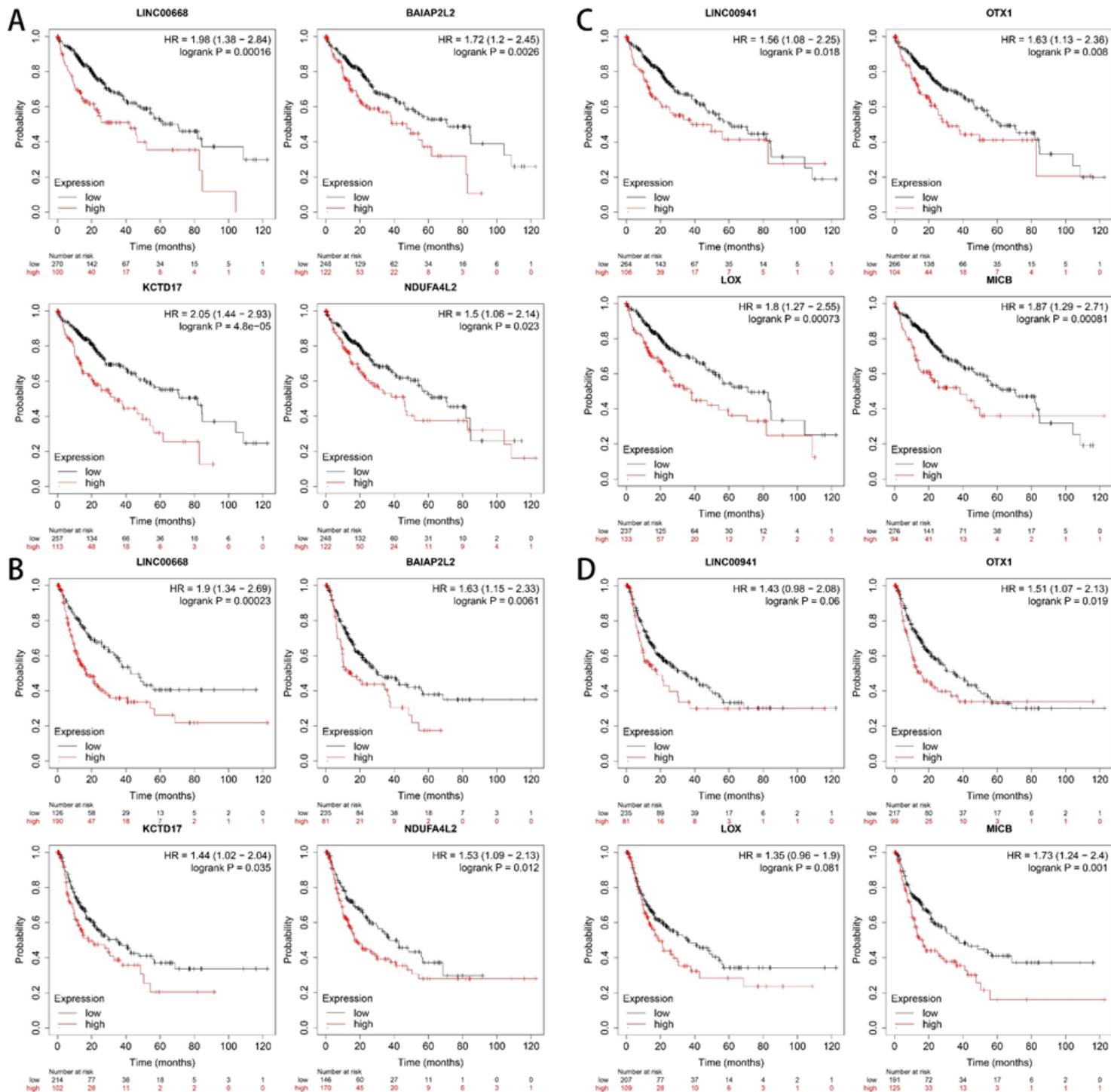


Figure 7

Survival analysis of hub genes in the key modules using the Kaplan-Meier Plotte A. OS of LINC00668 and co-expressed mRNA in HCC B. RFS of LINC00668 and co-expressed mRNA in HCC C. OS of LINC00941 and co-expressed mRNA in HCC D. RFS of LINC00941 and co-expressed mRNA in HCC