Bioscience Reports (2020) **40** BSR20200124
https://doi.org/10.1042/BSR20200124

PORTLAND PRESS

Research Article

Potential targets and molecular mechanism of miR-331-3p in hepatocellular carcinoma identified by weighted gene coexpression network analysis

Qingjia Chi[1], Xinge Geng[1], Kang Xu[2], Chunli Wang[3] and Han Zhao[4]

[1]Department of Mechanics and Engineering Structure, Wuhan University of Technology, Wuhan 430070, China; [2]Department of Cardiovascular Surgery, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China; [3]'111' Project Laboratory of Biomechanics and Tissue Repair, Bioengineering College, Chongqing University, Chongqing 400044, China; [4]Institute of Biomedical Sciences, School of Medicine, Jianghan University, Wuhan 430056, China

Correspondence: Han Zhao (zhaohan@mail.ustc.edu.cn)

Hepatocellular carcinoma (HCC) is one of the most common malignant tumor. miR-331-3p has been reported relevant to the progression of HCC, but the molecular mechanism of its regulation is still unclear. In the study, we comprehensively studied the role of miR-331-3p in HCC through weighted gene coexpression network analysis (WGCNA) based on The Cancer Genome Atlas (TCGA), Gene Expression Omnibus (GEO) and Oncomine. WGCNA was applied to build gene co-expression networks to examine the correlation between gene sets and clinical characteristics, and to identify potential biomarkers. Five hundred one target genes of miR-331-3p were obtained by overlapping differentially expressed genes (DEGs) from the TCGA database and target genes predicted by miRWalk. The critical turquoise module and its eight key genes were screened by WGCNA. Enrichment analysis was implemented based on the genes in the turquoise module. Moreover, 48 genes with a high degree of connectivity were obtained by protein–protein interaction (PPI) analysis of the genes in the turquoise module. From overlapping genes analyzed by WGCNA and PPI, two hub genes were obtained, namely coatomer protein complex subunit zeta 1 (COPZ1) and elongation factor Tu GTP binding domain containing 2 (EFTUD2). In addition, the expression of both hub genes was also significantly higher in tumor tissues compared with normal tissues, as confirmed by analysis based on TCGA and Oncomine. Both hub genes were correlated with poor prognosis based on TCGA data. Receiver operating characteristic (ROC) curve validated that both hub genes exhibited excellent diagnostic efficiency for normal and tumor tissues.

Introduction

Liver cancer is the fifth and third malignant tumor with morbidity and mortality [1]. In clinical diagnosis, hepatocellular carcinoma (HCC) cannot easily be diagnosed at an early stage, and is often not detected until the late stage of cancer [2,3]. HCC is the most common type of liver cancer, accounting for 75% of liver cancer [4]. Although some progress has been made in diagnosis and treatment strategies, the high metastasis rate and recurrence rate of HCC make it difficult for patients with advanced HCC to be effectively treated [5,6]. Thus, it is meaningful for the treatment to study the underlying molecular and identify novel markers for diagnosis and prognosis.

MicroRNAs (miRNAs) are a class of short, highly conserved, single-stranded non-coding RNAs, each with a length of 18–25 nucleotides [7,8]. MiRNAs play an important part in a variety of biological processes (BPs) by regulating gene expression post-transcriptionally [9,10]. According to recent studies, Let-7, miR-101 and miR-370 are down-regulated in HCC [11–13]. While miR-155, miR-21, miR- 221, miR-146a,

Received: 29 January 2020
Revised: 27 March 2020
Accepted: 01 April 2020

Accepted Manuscript online: 11 June 2020
Version of Record published: 25 June 2020

Bioscience Reports (2020) **40** BSR20200124
https://doi.org/10.1042/BSR20200124

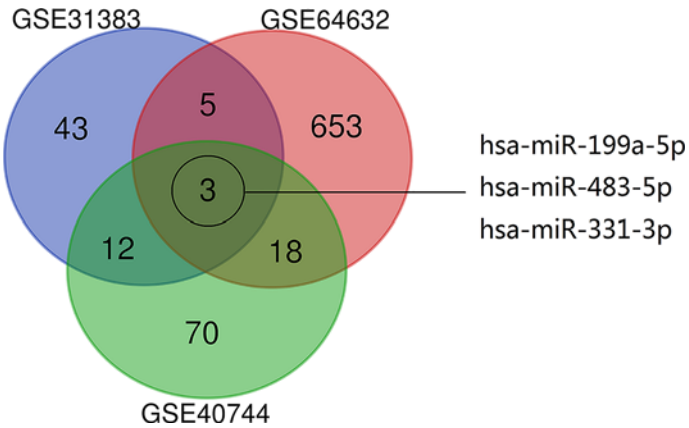


Figure 1. Venn diagram of a differential analysis dataset

Datasets (GSE31383, GSE40744, GSE64632) were selected. Non-coding RNA profiling by array method in the dataset was implemented.

miR-515 and miR-224 are up-regulated in HCC [14–18]. At the same time, these miRNAs have certain diagnostic value and prognostic significance for HCC.

In our previous studies, we combined computational, experimental and bioinformatic methods to investigate the biophysical properties of the nucleic acids [19,20], and the molecular mechanism of inflammation regulation of small molecule drugs [21–23]. miR-331-3p is considered to be an important cancer-related mircoRNA. Chen et al. found that miR-331-3p is an up-regulated micoRNA in pancreatic cancer (PC), while miR-331-3p inhibits suppression of tumorigenicity 7 like (ST7L) and epithelial mesenchymal transition (EMT)-mediated tumor metastasis, thereby promoting PC cell proliferation. miR-331-3p could be used as a potential diagnostic biomarker and drug target [24]. Similarly, Gu et al. found that the recurrence rate of esophageal adenocarcinoma patients with high-expressing serum miR-331-3p was lower, and miR-331-3p could be a potential biomarker for predicting tumor recurrence in patients with esophageal adenocarcinoma [25]. Yang et al. found that miR-331-3p inhibits the development of gastric cancer by targeting MSI 1 and serve as an indicator of gastric cancer prediction and prognosis [26]. Novel targets of miR-331-3p for liver cancer were also revealed. miR-331-3p promotes liver cancer and secondary EMT-mediated metastasis by inhibiting PLPPP-mediated dephosphorylation of protein kinase B (AKT). And miR-331-3p can serve as a new therapeutic target and a potential prognostic biomarker [27]. Similarly, miR-331-3p down-regulates E2F1 to promote the development and metastasis of HCC, indicating the possible application of miR-331-3p in predicting the prognosis and treatment of HCC [28]. There are reports that miR-331-3p expression is affected by viruses in HCC, and hepatitis B virus (HBV) is a typical virus that up-regulates miR-331-3p in HCC cell lines. miR-331-3p reduces von Hippel–Lindau tumor suppressor (VHL) expression by directly targeting its 3′-UT [29].

However, the molecular mechanism of miR-331-3p is not clear in the development of HCC. We aimed to comprehensively studied the role of miR-331-3p in HCC through WGCNA based on TCGA, GEO and Oncomine. WGCNA were applied to build gene co-expression networks to examine the correlation between gene sets and clinical characteristics, and to identify hub genes and critical pathway.

Materials and methods

Differential analysis is performed on three datasets (GSE31383, GSE40744, GSE64632) in the GEO database. These datasets implemented non-coding RNA profiling by microarray technology. Differentially expressed microRNAs (DEMs) were screened ($P<0.05$, $|\log FC|>1$) individually obtained from these datasets were overlapped to get three DEMs (miR-199a-5p, miR-483-5p, miR-331-3p) (Figure 1). The role of miR-199a-5p [30–34] and miR-483-5p [35–39] in HCC has been extensively studied. However, the molecular mechanism of miR-331-3p in HCC is not clear and remains to be explored. Consequently, miR-331-3p was chosen for further study.

First, the expression of miR-331-3p in HCC was obtained by integrating multiple ways, and the prognostic value analysis and comprehensive meta-analysis of miR-331-3p were performed. Then, the HCC-related differentially expressed genes (DEGs) from TCGA were overlapped with the target genes of miR-331-3p predicted from 12 databases to obtain overlapping target genes. Through WGCNA, the overlapping genes and the key genes in the modules that

play an important part in the development of HCC are obtained. Through the gene ontology (GO) enrichment analysis and Kyoto, Encyclopedia of Genes and Genomes (KEGG) pathway analysis of gene sets in important modules to explore its role in the biological process of HCC. Through PPI analysis of the gene sets in the important modules, genes with high gene connectivity are obtained. Hub genes were obtained from key genes and genes with high gene connectivity obtained from PPI analysis. The relationship between the miR-331-3p targeted genes and HCC was explored from the aspects of gene expression and DNA methylation. The working flow chart of the study is shown in Supplementary Figure S1.

TCGA high-throughput data for HCC patients

The batch download mode was used to download the TCGA data and extract the HCC expression data using the TCGA simple download tool-V16 in SangerBox software (http://sangerbox.com/). TCGA Easy Download Tool-V16 is obtained from the TCGA database website (https://portal.gdc.cancer.gov/; accessed in May 2019). A total of 422 mature miRNA expression profile samples were obtained, including 372 HCC tissue samples and 50 paracancerous tissue samples. Further we use SangerBox to combine the expression data in 422 samples and logarithmic conversion to obtain miRNA expression profiles. Finally, high-throughput data of TCGA of HCC patients with miR-331-3p were obtained from miRNA expression profiles.

GEO microarray data screening

The research conducted a keyword search of HCC-related microarray data on the GEO database (Gene Expression Omnibus; https://www.ncbi.nlm.nih.gov/gds/; accessed August 2019). For: (malignant * OR cancer OR tumor OR tumor OR neoplasm * OR carcinoma) AND (hepatocellular OR liver OR hepatic OR HCC) AND (microRNA OR miRNA OR 'micro RNA' OR 'small temporal RNA' OR 'non coding RNA' OR ncRNA OR 'small RNA'). The data retains that meets the following conditions is reserved for further analysis: (1) the data in the dataset are from humans; (2) the dataset has both HCC tissue expression data and healthy or adjacent tissue as control group expression data; (3) the number of samples in the experimental group and the control group is greater than 3; (4) the dataset contains the expression data of miR-331-3p.

Comprehensive meta-analysis

The study used RevMan 5.3 (London, UK) for a comprehensive meta-analysis. Standard mean difference (SMD) and 95% confidence interval (CI) were used to measure continuous results. We use the Mantel–Haenszel formula (fixed effects model) or DerSimonian–Laird formula (random effects model) to summarize SMDs and perform Cochrane's Q test (Chi-square test; Chi^2) and inconsistency (I^2) tests to assess heterogeneity. A random effect model is applied as heterogeneity is shown ($P < 0.05$ or $I^2 > 50\%$). Otherwise, a fixed-effect model is selected. A funnel plot of Egger's test was used to assess publication bias. A significant asymmetry of the funnel plot was determined when $P < 0.1$.

Prognostic value of miR-331-3p

Starbase v3.0 (http://starbase.sysu.edu.cn/index.php) is an open source platform that combines miRNA expression data and prognosis data with various cancers. We determined whether miR-331-3p is an effective prognostic biomarker for HCC by comparing the overall survival (OS) of HCC patients with different expression levels of miR-331-3p.

Target gene prediction of miR-331-3p

GEPIA (http://gepia.cancer-pku.cn/detail.php) is an interactive web server that uses standard processing flow of TCGA and GTEx projects to analyze gene expression data, which contains 9736 tumor and 8587 normal samples. We used GEPIA to access DEGs ($P < 0.05$ and $|\log FC| > 1$) between HCC and paracancerous tissues using its embedded Limma package. There are 12 databases (Microt4, miRWalk, mir-bridge, miRanda, miRDB, miRMap, Pictar2, PITA, MiRNAMap, RNAhybrid, RNA22 and Targetscan) in miRWalk2.0 version (http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/) to predict target genes of miR-331-3p. At least genes coexisted in five database were identified as the DEGs. Then, the DEGs obtained by GEPIA and the target genes predicted by miRWalk2.0 were compared to obtain overlapping genes of the two.

WGCNA explores important modules and key genes of overlapping genes

WGCNA is a systematic biology method to describe the pattern of gene correlation between different samples. It can classify highly synergistically changing gene sets based on the interconnectivity of gene sets and the relationship

Bioscience Reports (2020) **40** BSR20200124
https://doi.org/10.1042/BSR20200124

between gene sets and phenotypes. Associations identify candidate biomarker genes or therapeutic targets. Specific steps are as follows.

Importing data and preprocessing

Based on the overlapping gene set obtained above, by matching HCC-related RNA-seq data and clinical data in the TCGA database, a standardized transcript expression profile of the overlapping genes and clinical phenotypic data of HCC are obtained. The expression profile matrix and phenotypic matrix of overlapping genes were imported into R script, and then the expression profile matrix was clustered to remove outliers and genes.

Define gene expression similarity matrix

Based on the expression profile matrices of overlapping genes, a gene expression similarity matrix S of overlapping genes is calculated.

$$S = [S_{XY}] = [|cor(X, Y)|] = \left[\left| \frac{\sum_n (X_i - E(X))(Y_i - E(Y))}{\sqrt{\sum_n (X_i - E(X))^2} \sqrt{\sum_n (Y_i - E(Y))^2}} \right| \right]$$

Among them, S_{XY} is the similarity between genes X and Y, the absolute value of the Pearson correlation coefficient of vectors X and Y; X is the expression vector of gene X; Y is the expression vector of gene Y; n is the number of samples; $E(X)$ and $E(Y)$ represent the mean of the vectors X and Y, respectively.

Compute adjacency matrix

Select the exponential weighting coefficient β, and the selection of β should satisfy the law of scale-free networks. The similarity matrix S is further transformed into an adjacency matrix A.

$$A = [a_{XY}] = \left[|S_{XY}|^{\beta} \right],$$

where a_{XY} shows the adjacency coefficients of genes X and Y.

Create topological overlap matrix

Based on the adjacency matrix A, a topological overlap matrix TOM is constructed.

$$\mathrm{TOM} = [\omega_{XY}] = \left[\frac{l_{XY} + a_{XY}}{\min\{k_X, k_Y\} + 1 - a_{XY}} \right]$$

$$l_{XY} = \sum_u a_{Xu} a_{uY} \quad k_X = \sum_u a_{Xu} \quad k_Y = \sum_u a_{Yu}$$

The u in l_{XY} indicates the set of genes adjacent to the genes X and Y at the same time; k_X and k_Y the indicate the set of genes adjacent to genes X and Y, respectively.

Building a systematic clustering tree

Calculate the node dissimilarity d^{ω}_{XY} and construct a node dissimilarity matrix *dissTOM*. Based on the matrix *dissTOM*, a dynamic hybrid cutting algorithm is used to identify network modules from the system cluster tree.

$$dissTOM = \left[d^{\omega}_{XY} \right] = [1 - \omega_{XY}]$$

d^{ω}_{XY} indicates the degree of dissimilarity between genes X and Y.

Map gene co-expression networks

Eigenvector genes (module eigengene, ME) of each module is calculated. ME means the overall expression level of the module. Pearson coefficients between modules ME were calculated, and the module ME was clustered using average-linkage hierarchical clustering method. The modules with higher similarity were combined to obtain a co-expression network.

Screening gene modules

Correlate the expression level of the module with the phenotypic data, and calculate the Pearson correlation coefficient MT^T_M between each module's ME and the sample trait feature vector. Select modules that significantly associated with

phenotypic data for downstream analysis. When $P<0.05$, results are considered statistically significant.

$$MT_M^T = cor(ME^M, ST^T)$$

Among them, ME^M indicates the ME vector of the Mth module; ST^T indicates the ST vector of the Tth personality; and MT_M^Tindicates the correlation between the Mth module and the Tth personality.

Identifying pivot genes

The concept of module membership (MM) is introduced to measure the importance of genes in modules. At the same time, gene significance (GS) is introduced to reflect the degree of association between genes and traits. Key genes are identified using screening conditions of MM>0.8 and GS>0.2.

$$MM^M(X) = cor(X, ME^M)$$

$$GS^T(X) = cor(X, ST^T)$$

$MM^M(X)$ indicates the significance of the gene X in the Mth module; and $GS^T(X)$ indicates the degree of the gene X's prominence in the Tth module.

GO enrichment and KEGG pathway analysis

Based on key genes obtained from WGCNA, GO enrichment and KEGG pathway analysis were performed using DAVID 6.8 (https://david.ncifcrf.gov/). $P<0.05$ is considered to be the criteria of statistical significance in the analysis of GO and KEGG pathway enrichment. Three entries in GO enrichment analysis, namely cell component (CC), biological process (BP) and molecular function (MF), reflect the functional annotation of key genes. Critical pathways involved in the activity of miR-331-3p in HCC were analyzed through KEGG pathway analysis.

PPI network analysis

STRING (https://string-db.org/) is a web server for interactive gene search to generate PPI network. Then the PPI network of key genes collected from WGCNA is obtained from STRING. The nodes and lines in the network graph represent the target genes and their interactions, respectively. For accurate results, nodes in the network with an interaction score of less than 0.4 and nodes not connected to the main network will be deleted. Furthermore, interaction results obtained from STRING were imported into Cytoscape 3.6.1 for visualization and determination of gene connectivity. Gene connectivity is a quantitative indicator to assess the degree of interaction between genes.

Identification and validation of hub genes

The key genes of the important modules obtained by WGCNA are intersected with genes with a connectivity of $\geq$8 obtained by PPI network analysis. The hub gene with the highest correlation with HCC among miR-331-3p was obtained. UALCAN (http://ualcan.path.uab.edu) is an interactive web server with TCGA's RNA-seq data and clinical data for 31 cancer types. It can obtain the expression of a single gene in cancer and its survival curve. Use the TCGA simple download tool-V16 in SangerBox software to download the TCGA data and extract the data of coatomer protein complex subunit zeta 1 (COPZ1) and elongation factor Tu GTP binding domain containing 2 (EFTUD2). SPSS was used to draw the ROC curves of COPZ1 and EFTUD2, respectively. The ROC curve was used to distinguish expression data between HCC tissue and control tissue. The area under curve (AUC) value was used to evaluate the clinical diagnostic value of miR-331-3p in HCC. A larger AUC value indicates the higher diagnostic performance. Oncomine (www.oncomine.org) is currently the largest oncogene chip database and comprehensive data mining platform, with 715 gene expression data collections and 86733 samples of cancer and normal tissues. Oncomine can analyze the differential expression of different genes for universal cancer types and their normal tissues. We differentiated target genes by using Oncomine. Differential analysis results of HCC in different data sets were obtained, and the results with significant research significance were selected ($P<0.05$). DiseaseMeth 2.0 (http://bioinfo.hrbmu.edu.cn/diseasemeth/) is the largest database of DNA methylation status today. We use the methylation level of the hub gene obtained in this site in HCC and normal tissues adjacent to the cancer.

Bioscience Reports (2020) **40** BSR20200124
https://doi.org/10.1042/BSR20200124

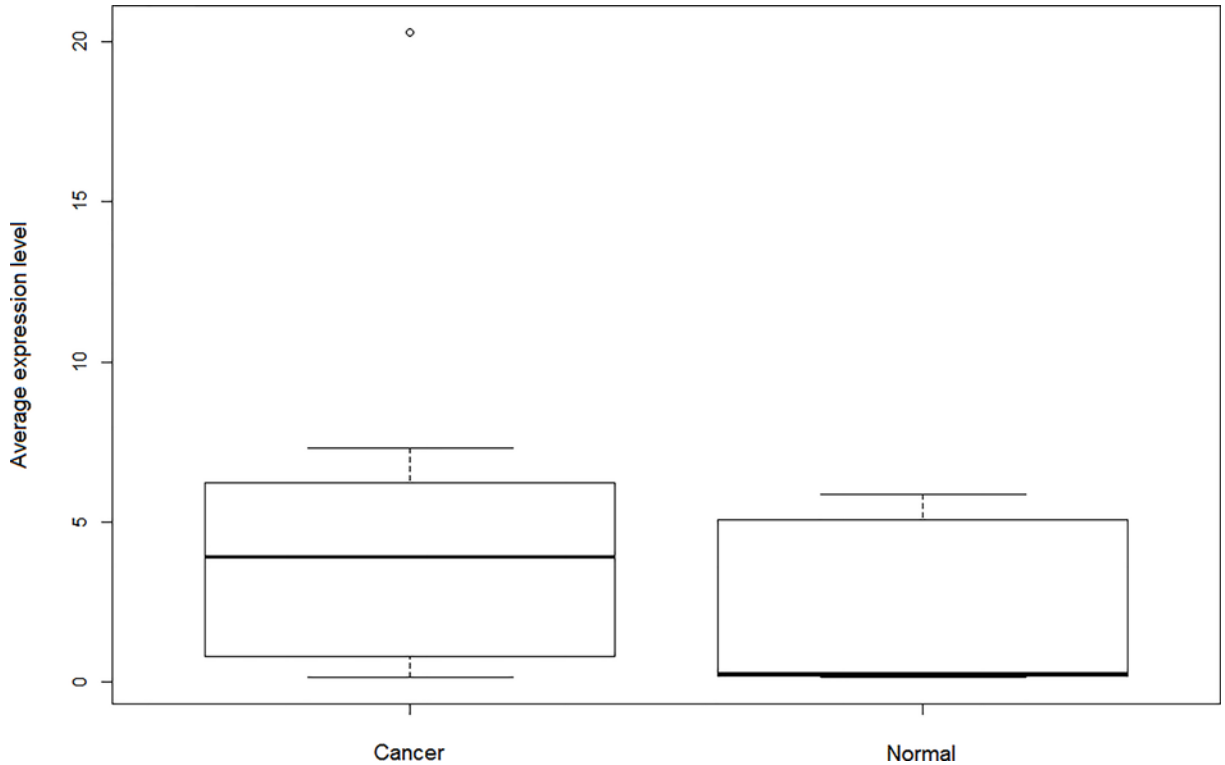


Figure 2. MiR-331-3p expression in HCC and normal tissues in the GEO database

MiR-331-3p expression data in HCC and normal tissues were integrated from the datasets (GSE31383, GSE40744, GSE64632), obtained from the GEO database.

Results

High-throughput data for TCGA in HCC patients

Box plots were drawn from the data in the three datasets (GSE31383, GSE64632, GSE40744) of miR-331-3p differential expression obtained above. The expression of miR-331-3p is higher in cancer tissues than in normal tissues (Figure 2).

GEO data screening

According to the method mentioned above, a keyword search is performed in the GEO database. Save the datasets that meet the requirements and get seven data sets (GSE10694, GSE31383, GSE40744, GSE64632, GSE64989, GSE67882, GSE98269), as shown in Table 1.

Comprehensive meta-analysis

As mentioned previously, RevMan 5.3 software was used to perform a meta-analysis on the included GEO dataset and TCGA dataset, which included 486 HCC and 179 non-cancerous liver tissues. No momentous difference was found between HCC and the control group (SMD = 0.26; 95% CI: 0.17–0.36; $P<0.00001$). The random effects model has significant heterogeneity ($P<0.00001$; $I^2 = 91\%$). The funnel plot does not indicate publication bias. Figure 3 shows the results of the forest plot and funnel plot.

Prognostic significance of miR-331-3p

Based on the results of StarBase v3.0, the curve of the OS rate of HCC patients with high and low expression of miR-331-3p (Figure 4) found that miR-331-3p has a significant prognostic significance for HCC. OS and enrollment time of the patients are demonstrated in Supplementary Table S1.

Table 1 Basic information of the data set studied

Series	Country	Experiment type	Platforms	HCC samples	Healthy samples
GSE10694	China	Non-coding RNA profiling by array	GPL6542	78	88
GSE31383	USA	Non-coding RNA profiling by array	GPL10122	9	10
GSE40744	USA	Non-coding RNA profiling by array	GPL14613	9	7
GSE64632	USA	Non-coding RNA profiling by array	GPL18116	3	3
GSE64989	Germany	Non-coding RNA profiling by array	GPL16384	8	10
GSE67882	India	Non-coding RNA profiling by array	GPL10850	4	8
GSE98269	China	Non-coding RNA profiling by array	GPL20712	3	3
TCGA		miR-seq	Illumina	372	50

Qualified datasets were found from the GEO database and the TCGA database, respectively. (1) The data in the dataset were from humans. (2) The dataset contained HCC tissue expression data and healthy or adjacent tissue control group expression data. The number of samples in both the experimental group and the control group was greater than 3. (4) The data set contained the expression data of miR-331-3p.

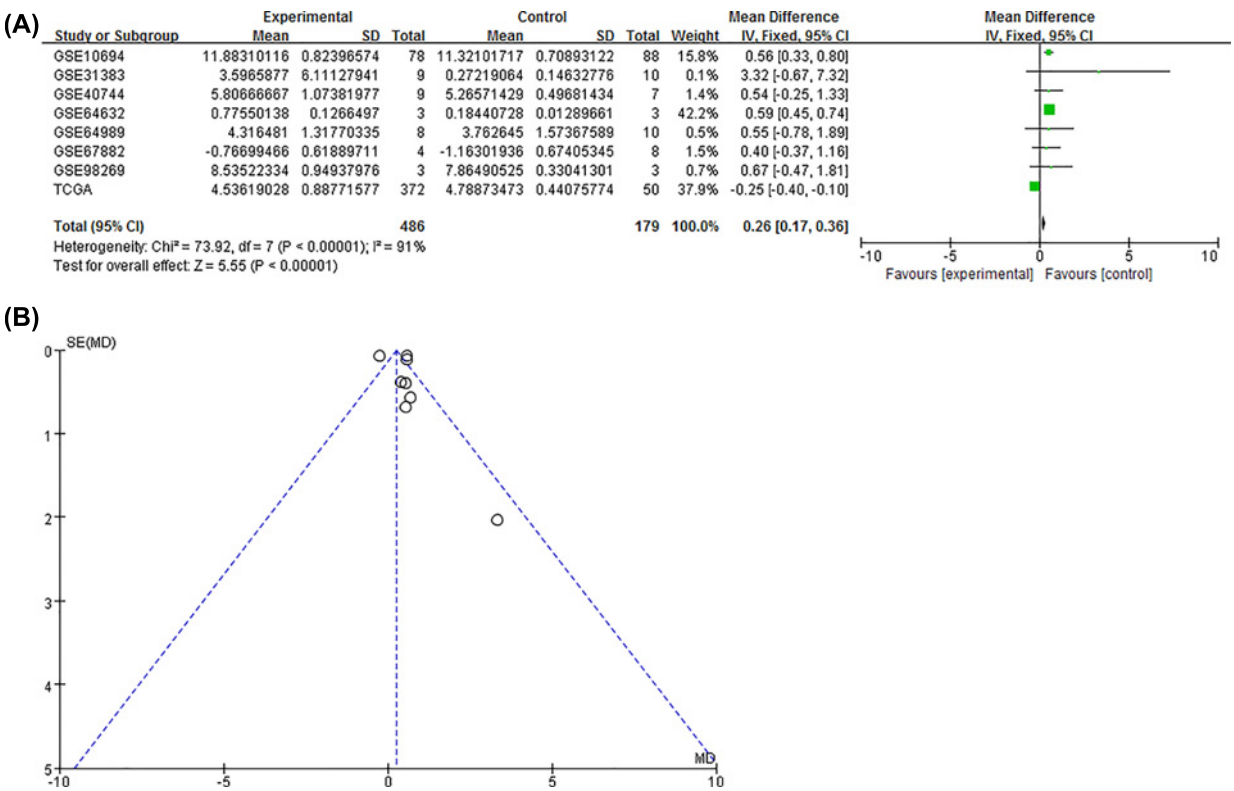


Figure 3. The study performed meta-analysis using RevMan based on data from the GEO and TCGA

(**A**) Meta-analysis of data and corresponding forest plots showed no significant difference between HCC and the control group (SMD = 0.26; 95% CI: 0.17–0.36; $P<0.00001$). The random effects model has significant heterogeneity ($P<0.00001$; $I^2 = 91\%$). (**B**) Funnel plot without publication bias.

miR-331-3p target gene prediction

The first step is to analyze the genetic high-throughput data of the TCGA database through GEPIA. 2206 HCC-related genes were screened out within the difference threshold ($q<0.01$ and $|\log2FC|>1$). Next, we use miRwalk2.0 to predict the potential target genes and retain the genes found in at least three databases. Six thousand two hundred seventy three potential target genes were obtained. In the third step, the genes obtained in the first two steps are overlapped to obtain 501 overlapping genes (Supplementary Figure S2).

Bioscience Reports (2020) **40** BSR20200124
https://doi.org/10.1042/BSR20200124

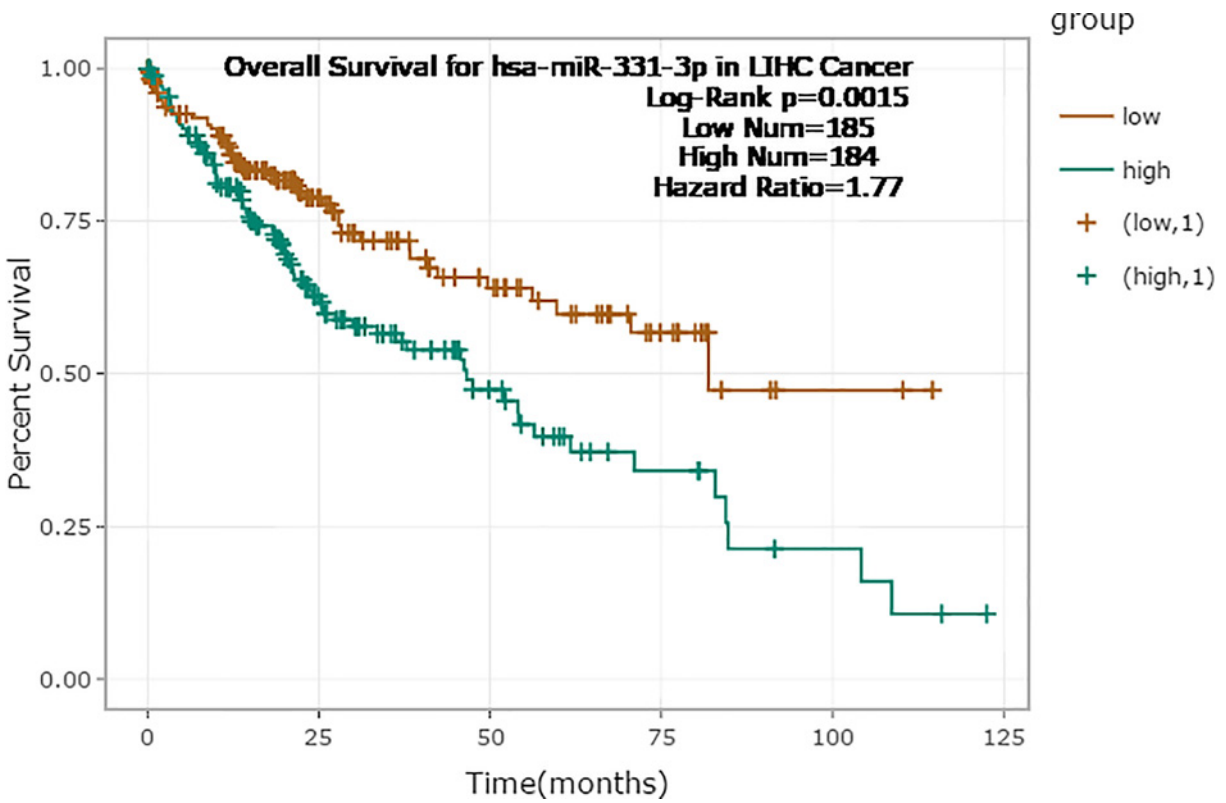


Figure 4. Survival curve of miR-331-3p against HCC based on TCGA data using StarBase v3.0 (median = 21.8, *P* = 0.0015, hazard ratio = 1.77)

WGCNA explores important modules and key genes of overlapping genes

The WGCNA method was used to find the overlapping genes that play an important part in the development of HCC modules and the key genes in the modules. Transcript expression profiles of 501 DEGs corresponding to overlapping genes in the TCGA database were selected and prepared for WGCNA analysis. On account of the gene expression pattern, the optimal soft threshold 4 found by the program was used as the soft threshold. The module selection criteria are cut height of 0.25 and minimum module size of 10. The clustering results show that the overlapping gene set is divided into eight modules. The correlation between the module and clinical trait data showed that the correlation between the turquoise module (correlation coefficient = 0.3, $P = 1e{-}8$) and Grade phenotype data was the most significant. There are 208 genes in the turquoise module (Figure 5). Eight key genes (TRM3 PPMIG PIGU RALY EFTUD2 PYGO2 STIP1 COPZ1) were obtained in the turquoise module (MM>0.8 and GS>0.2).

GO enrichment and KEGG pathway analysis of genes in the turquoise module

We perform GO analysis and KEGG pathway analysis on 208 genes in the turquoise module through the DAVID 6.8. Through GO analysis, 55 annotations were found to be enriched in BPs, cellular components (CCs) and MFs (Figure 6). GO analysis shows that in terms of BPs, genes are significantly enriched in certain cellular processes, such as cell division, mitotic nuclear division, etc. In terms of CCs, they are significantly enriched in the following cell groups: cytosol, extracellular exosome, etc. In terms of molecular function, genes are significantly enriched in protein binding (Table 2). KEGG pathway analysis yielded a total of nine enriched pathways (Figure 7). KEGG pathway analysis displays that the target genes were extensively involved in viral carcinogenesis, protein processing in endoplasmic reticulum and cell cycle (Table 3).

PPI network analysis of genes in Turquoise module

A total of 208 genes in the turquoise module to construct a PPI network by STRING. The nodes and lines in the figure represent genes and interactions between genes, respectively. The results of the PPI network obtained from STRING

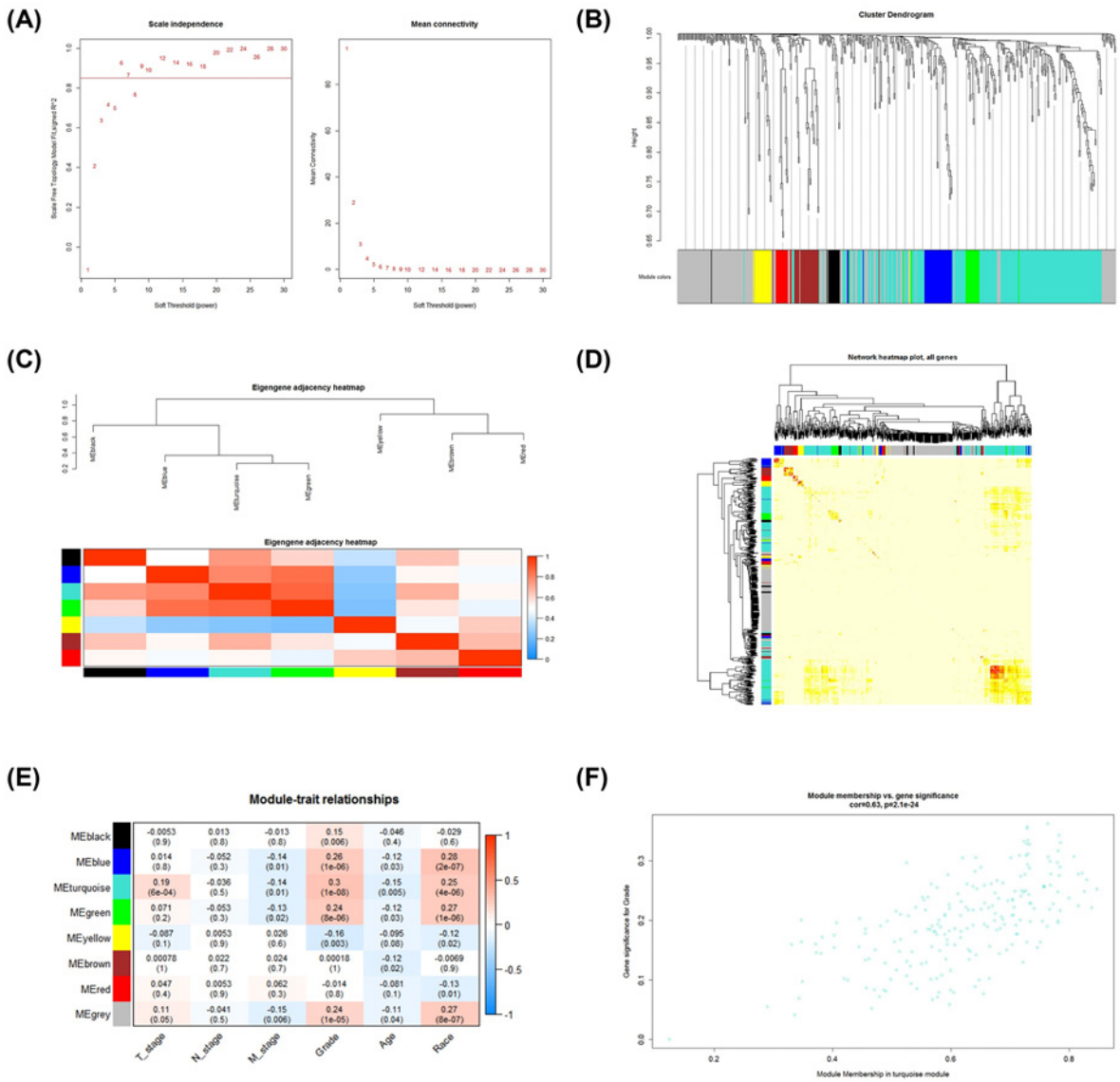


Figure 5. Exploring important modules related to target genes and clinical features through WGCNA

(**A**) Analyze the scale-free fitting index (left) and average connectivity (right) of various soft threshold weights. (**B**) Treemap of all DEGs clustered based on dissimilarity measures. (**C**) Clustering of eigen genes in the module. (**D**) Correlation between genes. (**E**) Heatmap of the correlation between modular feature genes and clinical features. Each unit corresponds to a correlation coefficient and a *P* value. (**F**) Scatter plot of module eigen genes in turquoise module.

were imported into Cytoscape 3.6.1 for further visualization, and 42 genes were screened according to the degree of gene connectivity degree ≥8. The PPI network is shown in Supplementary Figure S3.

Identification and validation of hub genes

The key genes of the significant module obtained by WGCNA above and the genes with the degree of connectivity of degree ≥8 obtained by PPI network analysis are intersected to obtain two hub genes, namely COPZ1 and EFTUD2. The Venn diagram is shown in Supplementary Figure S4. The expression of both genes in HCC tissues was higher than in normal tissues, as revealed by UALCAN. At the same time, they have significant prognostic significance for HCC and have high confidence ($P<0.05$) (Figure 8). ROC curves of COPZ1 (AUC = 0.973, $P<0.001$) and EFTUD2 (AUC = 0.959, $P<0.001$) were drawn from the TCGA database. COPZ1 and EFTUD2 have high diagnostic value

Bioscience Reports (2020) **40** BSR20200124
https://doi.org/10.1042/BSR20200124

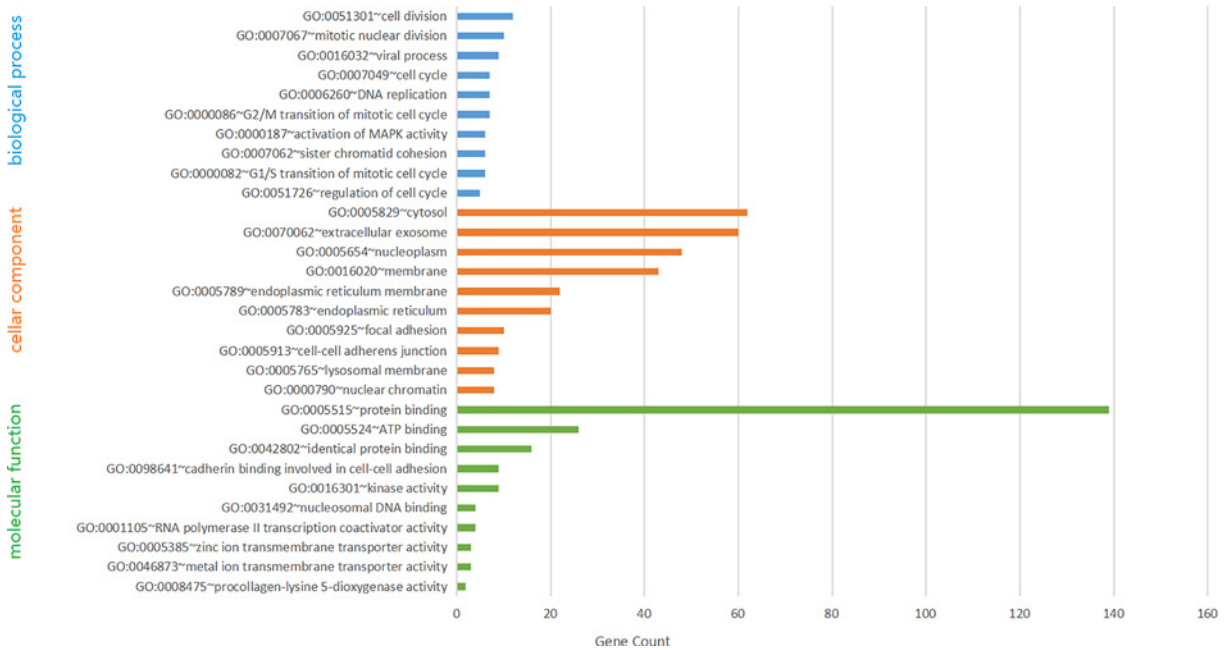


Figure 6. GO enrichment analysis chart, which is divided into biological process, CC and molecular function, and each part is arranged in descending order according to the first 10 terms of Count value ($P<0.05$)

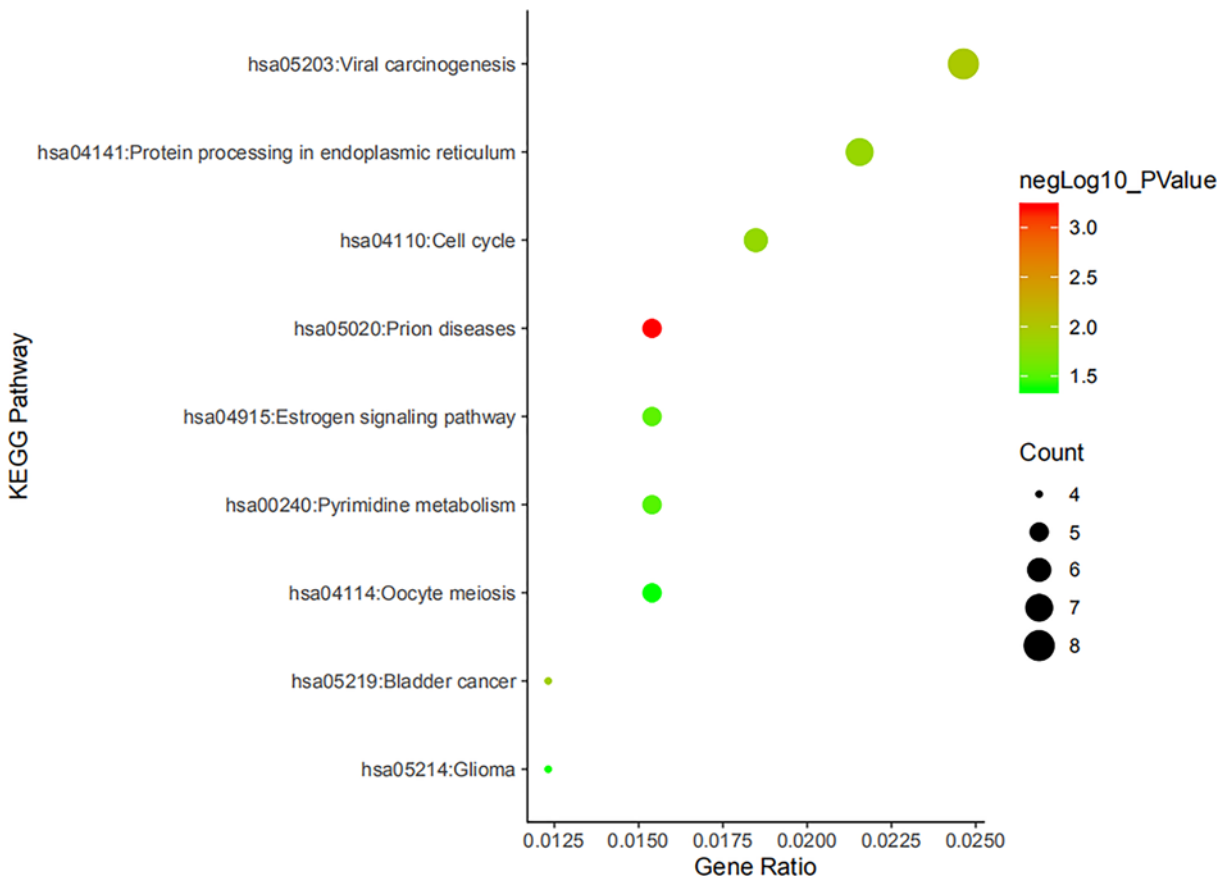


Figure 7. KEGG enrichment analysis

The chart is arranged in descending order of Count value ($P<0.05$)

Table 2 GO enrichment analysis table (divided into GOTERM_BP_DIRECT, GOTERM_CC_DIRECT, GOTERM_CC_DIRECT three groups, *P*<0.05, each group is ranked according to the Count value from high to low, and the first 10 data of each group are taken)

Category	Term	Count	*P* value
GOTERM_BP_DIRECT	GO:0051301~cell division	12	0.001766834
GOTERM_BP_DIRECT	GO:0007067~mitotic nuclear division	10	0.001760521
GOTERM_BP_DIRECT	GO:0016032~viral process	9	0.017941637
GOTERM_BP_DIRECT	GO:0000086~G2/M transition of mitotic cell cycle	7	0.004120868
GOTERM_BP_DIRECT	GO:0006260~DNA replication	7	0.007530052
GOTERM_BP_DIRECT	GO:0007049~cell cycle	7	0.033261067
GOTERM_BP_DIRECT	GO:0000082~G1/S transition of mitotic cell cycle	6	0.005413891
GOTERM_BP_DIRECT	GO:0007062~sister chromatid cohesion	6	0.005640417
GOTERM_BP_DIRECT	GO:0000187~activation of MAPK activity	6	0.00661362
GOTERM_BP_DIRECT	GO:0051726~regulation of cell cycle	5	0.048344183
GOTERM_CC_DIRECT	GO:0005829~cytosol	62	7.53E−06
GOTERM_CC_DIRECT	GO:0070062~extracellular exosome	60	1.43E−07
GOTERM_CC_DIRECT	GO:0005654~nucleoplasm	48	8.85E−04
GOTERM_CC_DIRECT	GO:0016020~membrane	43	1.39E−04
GOTERM_CC_DIRECT	GO:0005789~endoplasmic reticulum membrane	22	4.08E−04
GOTERM_CC_DIRECT	GO:0005783~endoplasmic reticulum	20	0.001556487
GOTERM_CC_DIRECT	GO:0005925~focal adhesion	10	0.025871509
GOTERM_CC_DIRECT	GO:0005913~cell-cell adherens junction	9	0.023871692
GOTERM_CC_DIRECT	GO:0000790~nuclear chromatin	8	0.004950709
GOTERM_CC_DIRECT	GO:0005765~lysosomal membrane	8	0.029022075
GOTERM_MF_DIRECT	GO:0005515~protein binding	139	3.87E−08
GOTERM_MF_DIRECT	GO:0005524~ATP binding	26	0.036592117
GOTERM_MF_DIRECT	GO:0042802~identical protein binding	16	0.026280507
GOTERM_MF_DIRECT	GO:0016301~kinase activity	9	0.006798774
GOTERM_MF_DIRECT	GO:0098641~cadherin binding involved in cell-cell adhesion	9	0.019123621
GOTERM_MF_DIRECT	GO:0001105~RNA polymerase II transcription coactivator activity	4	0.008714107
GOTERM_MF_DIRECT	GO:0031492~nucleosomal DNA binding	4	0.015792542
GOTERM_MF_DIRECT	GO:0046873~metal ion transmembrane transporter activity	3	0.006750044
GOTERM_MF_DIRECT	GO:0005385~zinc ion transmembrane transporter activity	3	0.026101256
GOTERM_MF_DIRECT	GO:0008475~procollagen-lysine 5-dioxygenase activity	2	0.034083931

Table 3 KEGG path analysis table (ranked from highest to lowest according to the Count value)

Category	Term	Count	*P* value
KEGG_PATHWAY	hsa05203:Viral carcinogenesis	8	0.010157189
KEGG_PATHWAY	hsa04141:Protein processing in endoplasmic reticulum	7	0.014300805
KEGG_PATHWAY	hsa04110:Cell cycle	6	0.015123404
KEGG_PATHWAY	hsa05020:Prion diseases	5	6.32E−04
KEGG_PATHWAY	hsa04915:Estrogen signaling pathway	5	0.028902908
KEGG_PATHWAY	hsa00240:Pyrimidine metabolism	5	0.030805553
KEGG_PATHWAY	hsa04114:Oocyte meiosis	5	0.041413445
KEGG_PATHWAY	hsa05219:Bladder cancer	4	0.012156376
KEGG_PATHWAY	hsa05214:Glioma	4	0.040779829

(AUC>0.75), and both have statistical significance ($P<0.05$) (Figure 9). The gene expression and DNA copy number of the hub gene in HCC were obtained through the Oncomine 4.5 database, and data with high reliability were selected ($P<0.05$).The gene expression and DNA copy number of COPZ1 and EFTUD2 were significantly higher in liver cancer tissues than in normal tissues, and the Over-expression Gene Rank and DNA Copy Number Gain Gene Rank of all gene expressions are in the top 30% (Figure 10).The DNA methylation status of COPZ1 and EFTUD2 obtained from DiseaseMeth 2.0 in normal tissues was higher than that in cancer tissues (Supplementary Figure S5).

Bioscience Reports (2020) **40** BSR20200124
https://doi.org/10.1042/BSR20200124

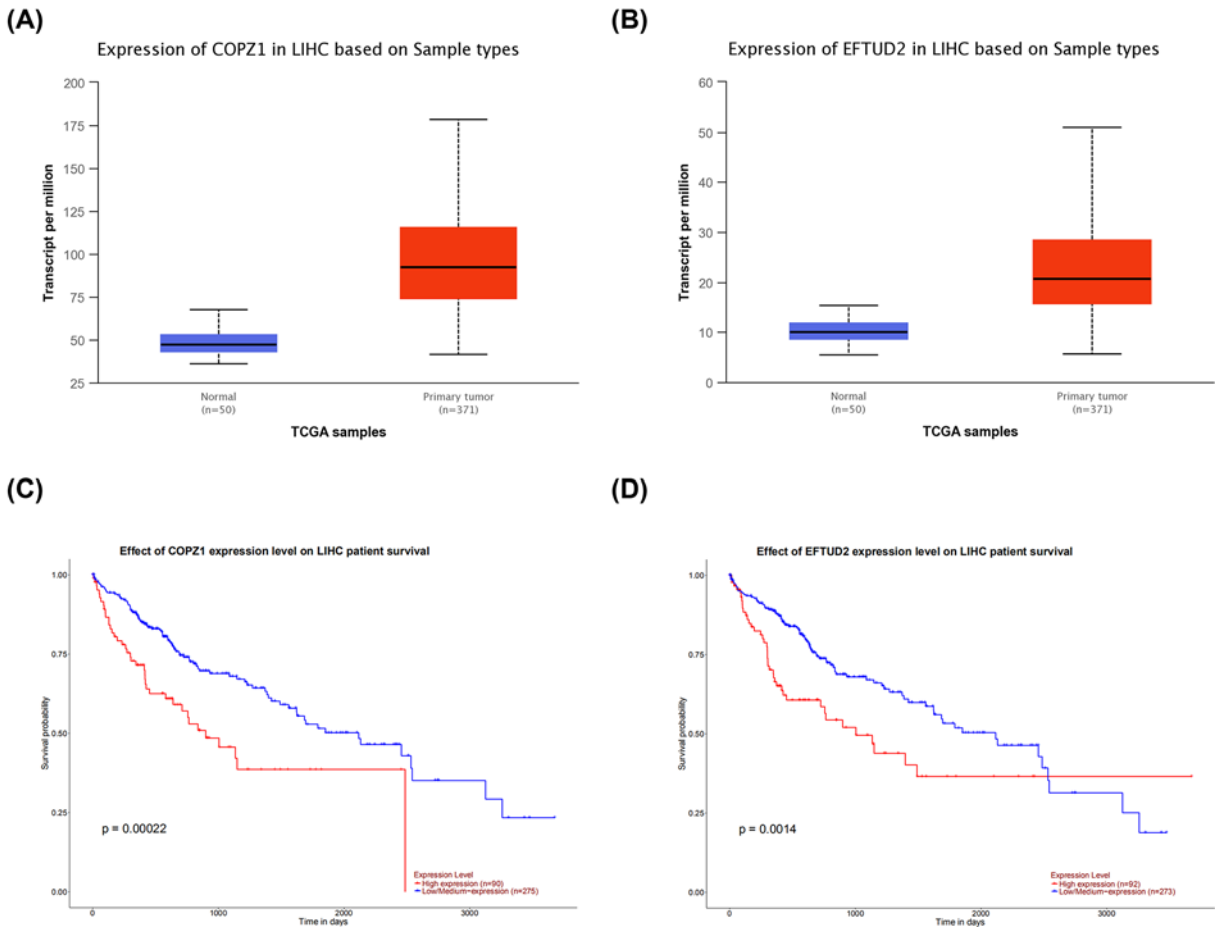


Figure 8. Box plot and survival curve of gene expression in HCC obtained by UALCAN

(**A**) Differential expression of COPZ1 in normal tissues and liver cancer tissues (P = 1.624E−12). (**B**) Differential expression of EFTUD2 in normal tissues and liver cancer tissues (P = 1.625E−12). (**C**) The effect of differential expression of COPZ1 on survival rate of patients with liver cancer (P = 0.00022). (**D**) The effect of differential expression of EFTUD2 on survival rate of patients with liver cancer (P = 0.0014).

Discussion

In the study, miR-331-3p was up-regulated in HCC from the GEO database. A large amount of data on miR-331-3p expression in HCC were collected by the GEO and TCGA databases. Then we performed meta-analysis based on GEO microarray and TCGA-based RNA-seq data to explore the diagnostic value of miR-331-3p in HCC. The prognostic significance of miR-331-3p in liver cancer is further confirmed by the study of the OS rate of HCC patients. We superimpose the 2206 HCC genes obtained by differential analysis of the TCGA database and the 6273 predicted target genes to obtain 501 overlapping genes. Important modules were further obtained through WGCNA analysis of overlapping genes, and eight key genes were screened out (TRM3 PPMIG PIGU RALY EFTUD2 PYGO2 STIP1 COPZ1). We performed PPI network analysis on genes in important modules, and selected 42 genes with a degree of connectivity of degree $\geq$8. The key genes in the important module and the genes with high gene connectivity obtained from the PPI network analysis are overlapped to obtain the hub gene (COPZ1 EFTUD2).

Although many studies have found that miR-331-3p is related to cancers such as colorectal cancer and, research on the regulation of miR-331-3p in HCC is still limited [40]. Chang et al. analyzed miRNA expression profiles and found that miR-331-3p can inhibit the expression of PH domain and leucine-rich repeat protein phosphatase (PHLPP) -mediated protein kinase B (AKT) and promote the proliferation of cancer cells [27]. Chen et al. found that miR-331-3p was significantly up-regulated in HCC through real-time PCR, which is consistent with our results [41]. As for miR-331-3p downstream genes, Cao et al. experimentally found that up-regulation of miR-331-3p inhibits the

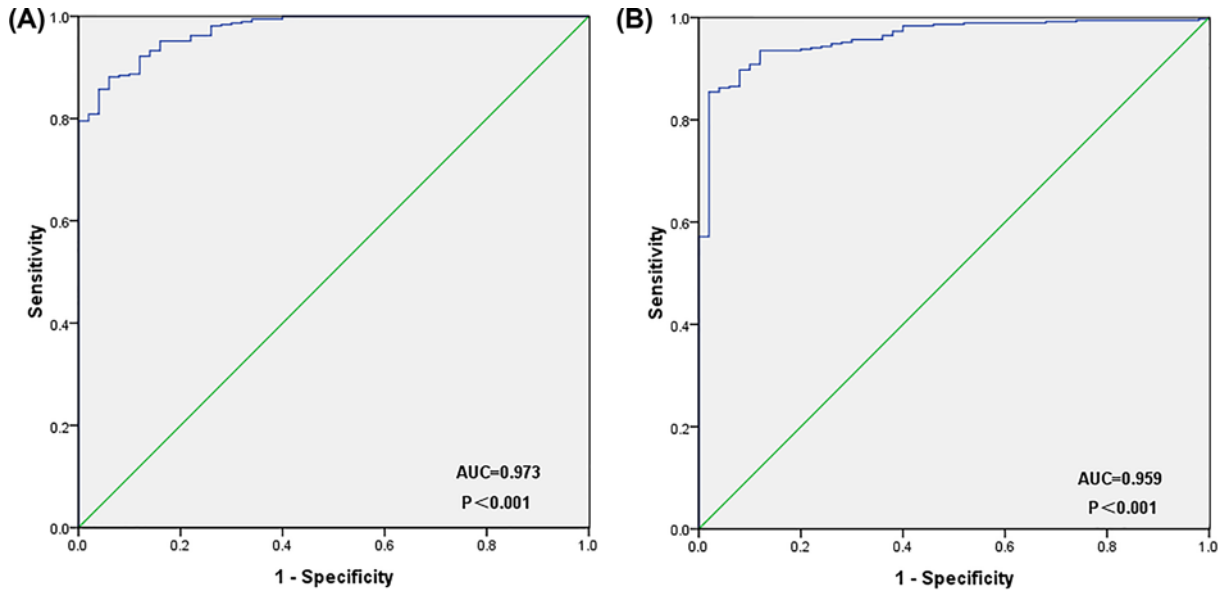


Figure 9. ROC curve of hub genes

(**A**) COPZ1 (AUC = 0.973, $P<0.001$). (**B**) EFTUD2 (AUC = 0.959, $P<0.001$). Both COPZ1 and EFTUD2 have diagnostic significance (AUC>0.75) and statistical significance ($P<0.05$).

expression of inhibitor of growth family member 5 (ING5), and then promote the proliferation of HCC cells [42]. Cao et al. used qRT-PCR to find that HBV up-regulates the expression of miR-331-3p in HCC cell lines. And miR-331-3p directly inhibits VHL expression on 3′-UTR [29]. miR-331-3p was suggested as a possible prognostic marker in HCC [43]. Moreover miR-331-3p had been used as a critical marker for early detection of high-risk HCC patients [44]. Studies have shown that miR-331-3p has clinical significance in HCC and is considered a possible prognostic marker [45,46], which is consistent with the conclusions of our study. And the diagnostic potential of serum miR-331-3p has also been confirmed [47]. ROC curve obtained from the expression data of miR-331-3p in TCGA (AUC = 0.594, $P<0.05$) is shown in Supplementary Figure S6. It shows that miR-331-3p is of diagnostic significance.

Pathway analysis indicates that miR-331-3p is engaged in the inflammatory response of HCC through viral, endoplasmic reticulum and cell cycle pathways. Lots of investigations have shown that miR-331-3p is related to inflammatory process of HCC [48]. Results have shown that miR-331-3p is up-regulated by HBV and promotes HCC cell proliferation by inhibiting ING5 expression [42]. Recent experiments have shown that miR-331-3p can be a serum biomarker for early hepatitis C virus associated hepatocellular carcinoma [44]. The endoplasmic reticulum is an important organelle responsible for various functions, and there have been many studies on its relationship with cancer [49], including that endoplasmic reticulum stress is closely related to tumor treatment [50]. Research by Su et al. also verified that galangin inhibits HCC proliferation by inducing endoplasmic reticulum stress [51]. For the cell cycle, studies have shown that miR-138 inhibits HCC through the cyclin D3 (CCND3) gene regulating the cell cycle of HCC [52].

UALCAN was used to obtain COPZ1, EFTUD2 gene expression was up-regulated in HCC, and COPZ1, EFTUD2 have prognostic significance for HCC. The ROC curve shows that COPZ1 and EFTUD2 have an ideal diagnostic performance. Oncomine further confirmed that COPZ1 and EFTUD2 gene expression was up-regulated in HCC. The DNA methylation levels of COPZ1 and EFTUD2 obtained through DiseaseMeth were down-regulated in HCC, which is consistent with the results of the up-regulation in gene expression above. However, studies on COPZ1 and EFTUD2 in HCC are still few, but some studies have shown that COPZ1 and EFTUD2 are closely related to cancer development. COPZ1 is involved in inflammatory, intracellular traffic, autophagy and lipid homeostasis. And COPZ1 also shows some function in abortive autophagy, endoplasmic reticulum stress, unfolded protein response and cell apoptosis [53]. Research by Anania et al. demonstrates the key role of COPZ1 in thyroid tumor cell viability, suggesting that it may be considered an attractive target for new treatments for thyroid cancer [54]. Oliver's study elucidates the mechanism by which cancer cells emit apoptosis signals when COPZ1 is depleted. Depletion of COPZ1 can lead to loss of cancer-specific COPI function and subsequent paralysis of the Golgi apparatus. It also shows that COPZ1 is a significant target in the treatment of cancer [55]. EFTUD2, which is a spliceosome protein, was also involved in

Bioscience Reports (2020) **40** BSR20200124
https://doi.org/10.1042/BSR20200124

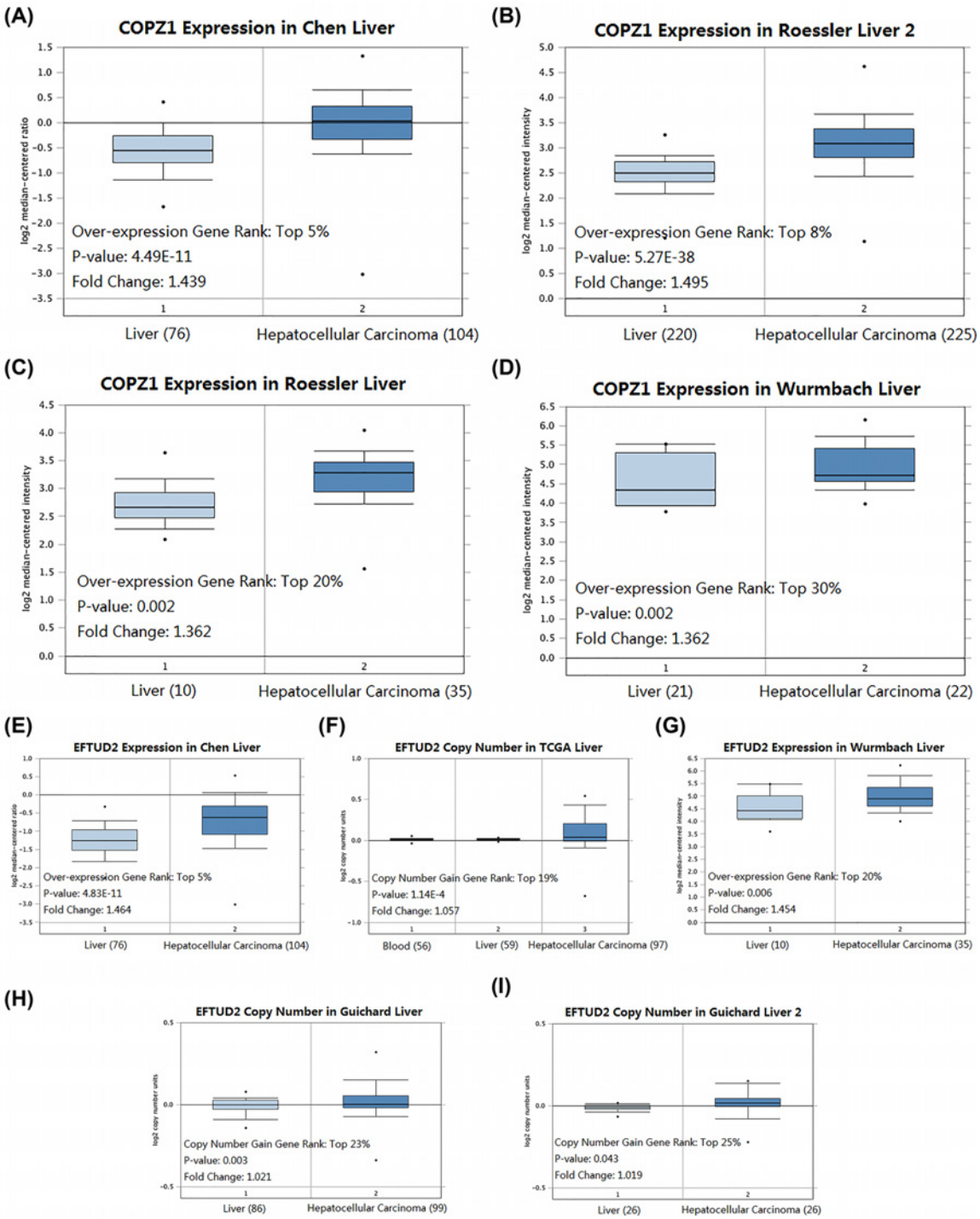


Figure 10. Gene expression in HCC and normal tissues from the Oncomine 4.5 database

(**A**) Expression of COPZ1 in different databases: (**A**) Chen liver; (**B**) Roessler liver; 2 (**C**) Roessler liver; (**D**) Wurmbach liver. The selected databases were all up-regulated in HCC and had statistical significance ($P<0.05$). (**B**) Expressions of EFTUD2 in different databases: (**E**) Chen liver; (**F**) TCGA liver; (**G**) Wurmbach liver; (**H**) Guichard liver; (**I**) Guichard liver 2. Expressions of these genes were all up-regulated in HCC and had statistical significance for the selected databases ($P<0.05$).

innate immunity and cell apoptosis [56]. EFTUD2 regulates RIG-I and MDA5 through mRNA splicing. EFTUD2 will be an interesting target to study the splicing mechanism by which EFTUD2 regulate on MyD88, RIG-I and MDA5 [57]. Studies by Sato et al. indicate that EFTUD2 plays an important part in the development of breast cancer. It is proved that depletion of SNW domain containing 1 (SNW1) and its related factor EFTUD2 can induce breast cancer cell apoptosis. In addition, expression of the SNW1 or EFTUD2 deletion construct can inhibit the association of endogenous proteins, thereby significantly increasing the number of apoptosis cells [58]. Later studies have shown that EFTUD2 accelerates the development of colitis-related tumors. EFTUD2 is constantly overexpressed in colon tissue and infiltrating macrophages. EFTUD2's myeloid-specific knockout significantly inhibits chronic intestinal inflammation and tumorigenesis, which is related to the reduction in the production of inflammatory cytokines and tumorigenic factors [59]. Zhu et al. have shown that EFTUD2 mainly restricts HCV infection through a retinoic acid-inducible gene 1 (RIG-I)/melanoma differentiation-associated protein 5 (MDA5)-mediated pathway independent of JAK-STAT. And they suggested a potential antiviral pathway [60]. And experiments show that EFTUD2 inhibits HBV infection by up-regulating the expression of RIG-I [61], which is consistent with our description of the pathway above.

Conclusion and perspectives

In the study, we comprehensively studied the role of miR-331-3p in HCC through weighted gene coexpression network analysis (WGCNA) based on TCGA, GEO and Oncomine. WGCNA were applied to build gene co-expression networks to examine the correlation between gene sets and clinical characteristics, and to identify hub genes and critical pathways. miR-331-3p is upregulated in HCC and demonstrates good prognosis and diagnostic performance for HCC based on the GEO and TCGA data sets. HCC-related genes obtained from the TCGA database overlapped with miR-331-3p potential genes, and 501 target genes for miR-331-3p related to HCC were obtained. Based on overlapping genes, the critical turquoise module and its eight key genes were screened by WGCNA. Genes in the turquoise module were analyzed for enrichment analysis to explore their role in the BPs of HCC. Based on the genes in the turquoise module, 48 genes with the degree of gene connectivity ≥ 8 were obtained by PPI analysis. Both genes (COPZ1 EFTUD2) were obtained by overlapping the key genes and those obtained by PPI analysis. The relationship between miR-331-3p targeting COPZ1 and EFTUD2 and HCC was explored from gene expression and DNA methylation. The study of miR-331-3p may be helpful for understanding the genetic level of HCC progression and revealing its potential molecular mechanisms and regulatory networks.

There are still some problems to be solved in the study. First of all, the data of existing databases for bioinformatic analysis and data mining were applied in the study. We hope to verify our conclusions experimentally in the future work. Second, there are few studies on the role of COPZ1 and EFTUD2 in HCC. The role of both genes in HCC was required to be studied in detail. Third, the diagnostic performance of miRNA-331-3p could be better presented if the expression level from serum samples was determined of HCC and analyzed. Fourth, the infiltration of immune cells may be related to the progress of HCC. Taking immune infiltration into consideration will help us to further explore the role of miR-331-3p in HCC.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

The study was supported by the National Natural Science Foundation of China [grant number 11602181]; the Fundamental Research Funds for the Central Universities [grant number WUT: 2018IB005], the Open Project of the State Key Laboratory of Trauma, Burn and Combined Injury, Army Medical University [grant number SKLKF201606]; the Visiting Scholar Foundation of Key Laboratory of Biorheological Science and Technology (Chongqing University), Ministry of Education [grant numbers CQKLBST-2018-006, CQKLBST-2018-009].

Author Contribution

Q.C. wrote the manuscript; X.G. prepared some figures and tables; K.X. and C.W. participated in result interpretation; and H.Z. supervised the whole project.

Abbreviations

AUC, area under curve; BP, biological process; CC, cellular component; CCND3, cyclin D3; CI, confidence interval; COPZ1, coatomer protein complex subunit zeta 1; DEG, differentially expressed gene; DEM, differentially expressed microRNA; EFTUD2, elongation factor Tu GTP binding domain containing 2; EMT, epithelial mesenchymal transition; GEO, Gene Expression

Omnibus; GO, gene ontology; GS, gene significance; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; HCV, hepatitis C virus; ING5, inhibitor of growth family member 5; KEGG, Kyoto, Encyclopedia of Genes and Genomes; MDA5, melanoma differentiation-associated protein 5; ME, module eigengene; MF, molecular functions; miRNA, microRNA; MM, module membership; PC, pancreatic cancer; PPI, protein–protein interaction; RIG-I, retinoic acid-inducible gene 1; ROC, receiver operating characteristic; SNW1, SNW domain containing 1; SMD, standard mean difference; ST7L, suppression of tumorigenicity 7 like; TCGA, The Cancer Genome Atlas; VHL, von Hippel–Lindau tumor suppressor; WGCNA, weighted gene coexpression network analysis.

References

1 Farazi, P.A. and DePinho, R.A. (2006) Hepatocellular carcinoma pathogenesis: from genes to environment. *Nat. Rev. Cancer* **6**, 674, https://doi.org/10.1038/nrc1934
2 Schwabe, R.F. and Wang, T.C. (2011) Targeting liver cancer: first steps toward a miRacle? *Cancer Cell* **20**, 698–699, https://doi.org/10.1016/j.ccr.2011.11.021
3 Wang, P., Xu, M., Toffanin, S., Li, Y., Llovet, J.M. and Russell, D.W. (2012) Induction of hepatocellular carcinoma by in vivo gene targeting. *Proc. Natl. Acad. Sci.* **109**, 11264–11269, https://doi.org/10.1073/pnas.1117032109
4 Kim, J.U., Shariff, M.I., Crossey, M.M., Gomez-Romero, M., Holmes, E., Cox, I.J. et al. (2016) Hepatocellular carcinoma: review of disease and tumor biomarkers. *World J. Hepatol.* **8**, 471, https://doi.org/10.4254/wjh.v8.i10.471
5 Baek, K.K., Kim, J., Uhm, J.E., Park, S.H., Lee, J., Park, J.O. et al. (2011) Prognostic factors in patients with advanced hepatocellular carcinoma treated with sorafenib: a retrospective comparison with previously known prognostic models. *Oncology* **80**, 167–174, https://doi.org/10.1159/000327591
6 Qiu, L., Li, H., Fu, S., Chen, X. and Lu, L. (2018) Surface markers of liver cancer stem cells and innovative targeted-therapy strategies for HCC. *Oncol. Lett.* **15**, 2039–2048
7 Hammond, S.M. (2015) An overview of microRNAs. *Adv. Drug. Deliv. Rev.* **87**, 3–14, https://doi.org/10.1016/j.addr.2015.05.001
8 Chen, L. (2016) The biogenesis and emerging roles of circular RNAs. *Nat. Rev. Mol. Cell Biol.* **17**, 205, https://doi.org/10.1038/nrm.2015.32
9 Danish, S. and Maha, A. (2011) MicroRNAs in development and disease. *Physiol. Rev.* **91**, 827–887
10 Bogucka-Kocka, A., Zalewski, D.P., Ruszel, K.P., Stępniewski, A., Galkowski, D., Bogucki, J. et al. (2019) Dysregulation of microRNA regulatory network in lower extremities arterial disease. *Front. Genet.* **10**, 1200, https://doi.org/10.3389/fgene.2019.01200
11 Wang, Y., Lu, Y., Toh, S.T., Sung, W., Tan, P., Chow, P. et al. (2010) Lethal-7 is down-regulated by the hepatitis B virus x protein and targets signal transducer and activator of transcription 3. *J. Hepatol.* **53**, 57–66, https://doi.org/10.1016/j.jhep.2009.12.043
12 Li, C., Deng, M., Hu, J., Li, X., Chen, L., Ju, Y. et al. (2016) Chronic inflammation contributes to the development of hepatocellular carcinoma by decreasing miR-122 levels. *Oncotarget* **7**, 17021, https://doi.org/10.18632/oncotarget.7740
13 Sun, G., Hou, Y.B., Jia, H.Y., Bi, X.H., Yu, L. and Chen, D.J. (2016) MiR-370 promotes cell death of liver cancer cells by Akt/FoxO3a signalling pathway. *Eur. Rev. Med. Pharmacol. Sci.* **20**, 2011–2019
14 Qiu, X., Dong, S., Qiao, F., Lu, S., Song, Y., Lao, Y. et al. (2013) HBx-mediated miR-21 upregulation represses tumor-suppressor function of PDCD4 in hepatocellular carcinoma. *Oncogene* **32**, 3296, https://doi.org/10.1038/onc.2013.150
15 Rong, M., Chen, G. and Dang, Y. (2013) Increased miR-221 expression in hepatocellular carcinoma tissues and its role in enhancing cell growth and inhibiting apoptosis in vitro. *BMC Cancer* **13**, 21, https://doi.org/10.1186/1471-2407-13-21
16 Sun, X., Zhang, J., Hou, Z., Han, Q., Zhang, C. and Tian, Z. (2015) miR-146a is directly regulated by STAT3 in human hepatocellular carcinoma cells and involved in anti-tumor immune suppression. *Cell Cycle* **14**, 243–252, https://doi.org/10.4161/15384101.2014.977112
17 Yao, J., Liang, L., Huang, S., Ding, J., Tan, N., Zhao, Y. et al. (2010) MicroRNA-30d promotes tumor invasion and metastasis by targeting Galphai2 in hepatocellular carcinoma. *Hepatology* **51**, 846–856
18 Zhuang, L. and Meng, Z. (2015) Serum miR-224 reflects stage of hepatocellular carcinoma and predicts survival. *Biomed. Res. Int.* **2015**, https://doi.org/10.1155/2015/731781
19 Chi, Q. and Jiang, J. (2012) A bead-spring model and mean field theory based re-calculation reveals uncertainty of rouse-type dna dynamics in dilute solution. *Biomed. Eng. Appl. Basis Commun.* **24**, 355–364
20 Chi, Q., Wang, G. and Jiang, J. (2013) The persistence length and length per base of single-stranded DNA obtained from fluorescence correlation spectroscopy measurements using mean field theory. *Physica A* **392**, 1072–1079, https://doi.org/10.1016/j.physa.2012.09.022
21 Xu, K., Sha, Y., Wang, S., Chi, Q., Liu, Y., Wang, C. et al. (2019) Effects of Bakuchiol on chondrocyte proliferation via the PI3K-Akt and ERK1/2 pathways mediated by the estrogen receptor for promotion of the regeneration of knee articular cartilage defects. *Cell Prolif.*, https://doi.org/10.1111/cpr.12666
22 Xu, K., Zhou, T., Huang, Y., Chi, Q., Shi, J., Zhu, P. et al. (2018) Anthraquinone emodin inhibits tumor necrosis factor alpha-induced calcification of human aortic valve interstitial cells via the NF-κB pathway. *Front. Pharmacol.* **9**, https://doi.org/10.3389/fphar.2018.01328
23 Xu, K., Huang, Y., Zhou, T., Wang, C., Chi, Q., Shi, J. et al. (2019) Nobiletin exhibits potent inhibition on tumor necrosis factor alpha-induced calcification of human aortic valve interstitial cells via targeting ABCG2 and AKR1B1. *Phytother. Res.*, https://doi.org/10.1002/ptr.6360
24 Chen, X., Luo, H., Li, X., Tian, X., Peng, B., Liu, S. et al. (2018) miR-331-3p functions as an oncogene by targeting ST7L in pancreatic cancer. *Carcinogenesis* **39**, 1006–1015, https://doi.org/10.1093/carcin/bgy074
25 Gu, J., Zhang, J., Zheng, L., Ajani, J.A., Wu, X. and Ye, Y. (2018) Serum miR-331-3p predicts tumor recurrence in esophageal adenocarcinoma. *Sci. Rep.* **8**, 14006, https://doi.org/10.1038/s41598-018-32282-9
26 Yang, L., Song, G., Zhai, X., Wang, L., Liu, Q. and Zhou, M. (2019) MicroRNA-331 inhibits development of gastric cancer through targeting musashi1. *World J. Gastrointest. Oncol.* **11**, 705, https://doi.org/10.4251/wjgo.v11.i9.705

27 Chang, R.M., Yang, H., Fang, F., Xu, J.F. and Yang, L.Y. (2014) MicroRNA-331-;3p promotes proliferation and metastasis of hepatocellular carcinoma by targeting PH domain and leucine-rich repeat protein phosphatase. *Hepatology* **60**, 1251–1263
28 Jin, W., Zhong, N., Wang, L., Yu, J., Yin, F. and Zhang, K. (2019) MiR-331-3p inhibition of the hepatocellular carcinoma (HCC) Bel-7402 cell line by down-regulation of E2F1. *J. Nanosci. Nanotechnol.* **19**, 5476–5482, https://doi.org/10.1166/jnn.2019.16535
29 Cao, Y., Zhang, J., Xiong, D., Wang, D., Wu, T., Huang, A. et al. (2015) Hsa-miR-331-3p inhibits VHL expression by directly targeting its mRNA 3′-UTR in HCC cell lines. *Acta Biochim. Pol.* **62**, https://doi.org/10.18388/abp.2014˙779
30 Xu, N., Zhang, J., Shen, C., Luo, Y., Xia, L., Xue, F. et al. (2012) Cisplatin-induced downregulation of miR-199a-5p increases drug resistance by activating autophagy in HCC cell. *Biochem. Biophys. Res. Commun.* **423**, 826–831, https://doi.org/10.1016/j.bbrc.2012.06.048
31 Shen, Q., Cicinnati, V., Zhang, X., Iacob, S., Weber, F., Sotiropoulos, G. et al. (2010) Role of microRNA-199a-5p and discoidin domain receptor 1 in human hepatocellular carcinoma invasion. *Mol. Cancer* **9**, 227, https://doi.org/10.1186/1476-4598-9-227
32 Li, B., He, L., Zuo, D., He, W., Wang, Y., Zhang, Y. et al. (2017) Mutual regulation of MiR-199a-5p and HIF-1α modulates the Warburg effect in hepatocellular carcinoma. *J. Cancer* **8**, 940–949, https://doi.org/10.7150/jca.17496
33 Liu, L., Lu, L., Zheng, A., Xie, J., Xue, Q., Wang, F. et al. (2017) MiR-199a-5p and let-7c cooperatively inhibit migration and invasion by targeting MAP4K3 in hepatocellular carcinoma. *Oncotarget* **8**, 13666–13677, https://doi.org/10.18632/oncotarget.14623
34 Huang, G., Shan, H., Li, D., Zhou, B. and Pang, P. (2017) MiR-199a-5p suppresses tumorigenesis by targeting clathrin heavy chain in hepatocellular carcinoma. *Cell Biochem. Funct.* **35**, https://doi.org/10.1002/cbf.3252
35 Shen, J., Wang, A., Wang, Q., Gurvich, I., Siegel, A., Remotti, H. et al. (2013) Exploration of genome-wide circulating microRNA in hepatocellular carcinoma (HCC): MiR-483-5p as a potential biomarker. *Cancer Epidemiol. Biomarkers Prev.* **22**, 2364–2373, https://doi.org/10.1158/1055-9965.EPI-13-0237
36 Zhang, Z., Ge, S., Wang, X., Yuan, Q., Yan, Q., Ye, H. et al. (2012) Serum miR-483-5p as a potential biomarker to detect hepatocellular carcinoma. *Hepatol. Int.* **7**, 199–207, https://doi.org/10.1007/s12072-012-9341-z
37 Tang, S., Chen, Y., Feng, S., Yi, T., Liu, X., Qiang, L. et al. (2017) MiR-483-5p promotes IGF-II transcription and is associated with poor prognosis of hepatocellular carcinoma. *Oncotarget* **8**, https://doi.org/10.18632/oncotarget.21737
38 Lu, X., Di, C., Gu, X., Ding, J., Zhao, Y., Zhao, Q. et al. (2018) Predicting value of ALCAM as a target gene of microRNA-483-5p in patients with early recurrence in hepatocellular carcinoma. *Front. Pharmacol.* **8**, 973, https://doi.org/10.3389/fphar.2017.00973
39 Gailhouste, L., Liew, L.C., Yasukawa, K., Hatada, I., Tanaka, Y., Kato, T. et al. (2019) MEG3-derived miR-493-5p overcomes the oncogenic feature of IGF2-miR-483 loss of imprinting in hepatic cancer cells. *Cell Death. Dis.* **10**, 1–16, https://doi.org/10.1038/s41419-019-1788-6
40 Shee, S.M.E., Koh, R.Y., Voon, K.G.L., Chye, S.M., Othman, I. and Ng, K.Y. (2019) The roles of microRNA-331 family in cancers. *J. Cancer Res. Pract.* **6**, 1
41 Chen, L., Chu, F., Cao, Y., Shao, J. and Wang, F. (2015) Serum miR-182 and miR-331-3p as diagnostic and prognostic markers in patients with hepatocellular carcinoma. *Tumor Biol.* **36**, 7439–7447, https://doi.org/10.1007/s13277-015-3430-2
42 Cao, Y., Chen, J., Wang, D., Peng, H., Tan, X., Xiong, D. et al. (2015) Upregulated in hepatitis B virus-associated hepatocellular carcinoma cells, miR-331-3p promotes proliferation of hepatocellular carcinoma cells by targeting ING5. *Oncotarget* **6**, 38093, https://doi.org/10.18632/oncotarget.5642
43 Reichl, P. and Mikulits, W. (2016) Accuracy of novel diagnostic biomarkers for hepatocellular carcinoma: an update for clinicians. *Oncol. Rep.* **36**, 613–625, https://doi.org/10.3892/or.2016.4842
44 Sun, Q., Li, J., Jin, B., Wang, T. and Gu, J. (2019) Evaluation of miR-331-3p and miR-23b-3p as serum biomarkers for hepatitis c virus-related hepatocellular carcinoma at early stage. *Clin. Res. Hepatol. Gastroenterol.* 21–28, https://doi.org/10.1016/j.clinre.2019.03.011
45 Sukata, T., Sumida, K., Kushida, M., Ogata, K., Miyata, K., Yabushita, S. et al. (2011) Circulating microRNAs, possible indicators of progress of rat hepatocarcinogenesis from early stages. *Toxicol. Lett.* **200**, 46–52, https://doi.org/10.1016/j.toxlet.2010.10.013
46 Chang, R., Yang, H., Fang, F., Xu, J. and Yue, Y. (2014) MicroRNA-331-3p promotes proliferation and metastasis of hepatocellular carcinoma by targeting PH domain and leucine-rich repeat protein phosphatase. *Hepatology* **60**, https://doi.org/10.1002/hep.27221
47 Chen, L., Chu, F., Cao, Y., Shao, J. and Wang, F. (2015) Serum miR-182 and miR-331-3p as diagnostic and prognostic markers in patients with hepatocellular carcinoma. *Tumour Biol.* **36**, https://doi.org/10.1007/s13277-015-3430-2
48 Zhang, Z., Liu, X., Wang, D., Teng, M., Niu, L., Huang, A. et al. (2011) Hepatitis B virus and hepatocellular carcinoma at the miRNA level. *World J. Gastroenterol.* **17**, 3353, https://doi.org/10.3748/wjg.v17.i28.3353
49 Moenner, M., Pluquet, O., Bouchecareilh, M. and Chevet, E. (2007) Integrated Endoplasmic Reticulum Stress Responses in Cancer. *Cancer Res.* **67**, 10631, https://doi.org/10.1158/0008-5472.CAN-07-1705
50 Yadav, R.K., Chae, S., Kim, H. and Chae, H.J. (2014) Endoplasmic reticulum stress and cancer. *J. Cancer Prevention* **19**, 75–88, https://doi.org/10.15430/JCP.2014.19.2.75
51 Su, L., Chen, X., Wu, J., Lin, B., Zhang, H., Lan, L. et al. (2013) Galangin inhibits proliferation of hepatocellular carcinoma cells by inducing endoplasmic reticulum stress. *Food Chem. Toxicol.* **62**, 810–816, https://doi.org/10.1016/j.fct.2013.10.019
52 Wang, W., Zhao, L., Tan, Y., Ren, H. and Qi, Z. (2012) MiR-138 induces cell cycle arrest by targeting cyclin D3 in hepatocellular carcinoma. *Carcinogenesis* **33**, 1113–1120, https://doi.org/10.1093/carcin/bgs113
53 Anania, M.C., Cetti, E., Lecis, D., Todoerti, K., Gulino, A., Mauro, G. et al. (2017) Targeting COPZ1 non-oncogene addiction counteracts the viability of thyroid tumor cells. *Cancer Lett.* **410**, 201–211, https://doi.org/10.1016/j.canlet.2017.09.024
54 Anania, M.C., Cetti, E., Lecis, D., Todoerti, K., Gulino, A., Mauro, G. et al. (2017) Targeting COPZ1 non-oncogene addiction counteracts the viability of thyroid tumor cells. *Cancer Lett.* **410**, 201–211, https://doi.org/10.1016/j.canlet.2017.09.024
55 Oliver, D.J. (2016) High-throughput sequencing-based approaches for the identification and development of molecular targets for cancer therapy.

56 De Arras, L., Laws, R., Leach, S.M., Pontis, K., Freedman, J.H., Schwartz, D.A. et al. (2014) Comparative genomics RNAi screen identifies Eftud2 as a novel regulator of innate immunity. *Genetics* **197**, 485–496, https://doi.org/10.1534/genetics.113.160499

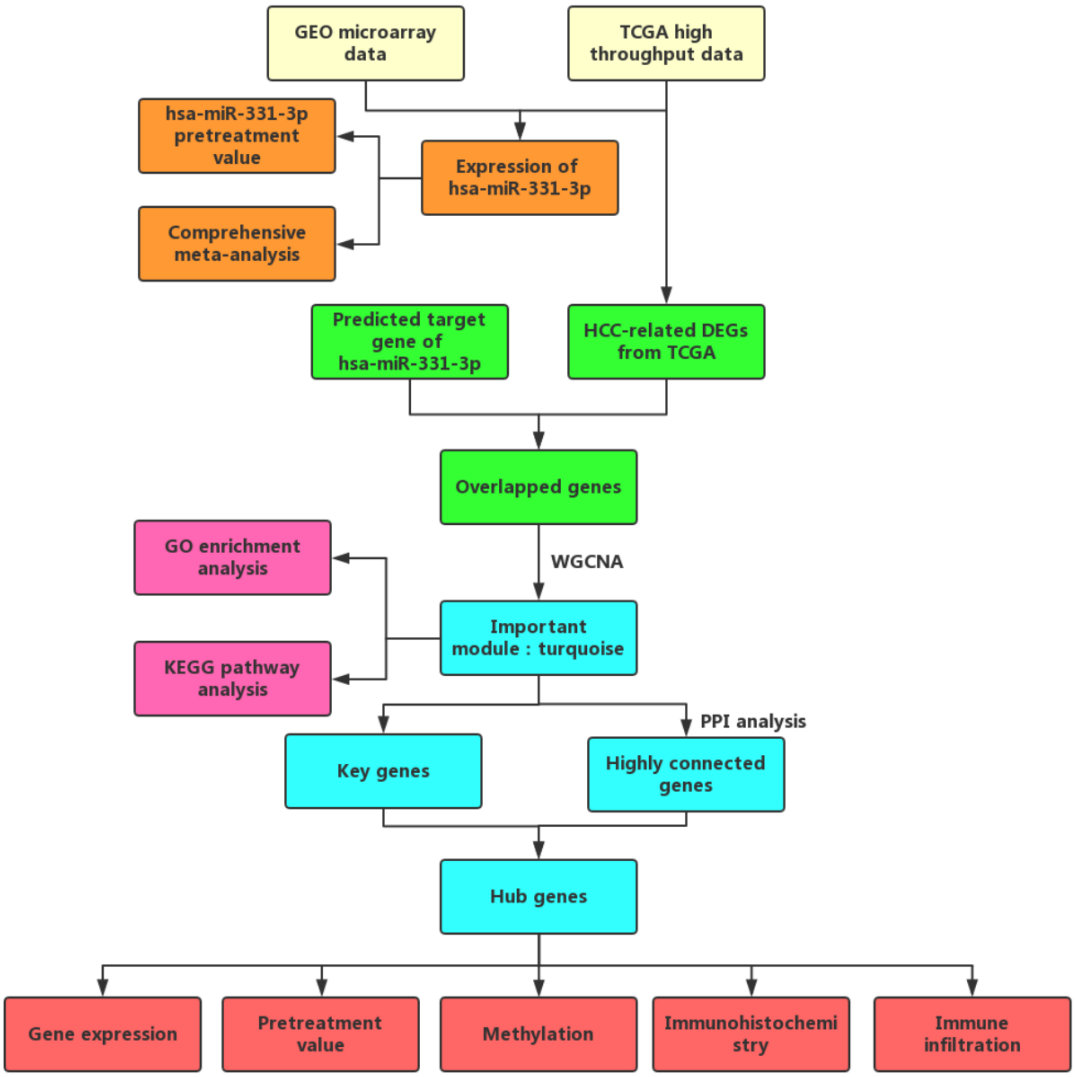
57 Zhu, C., Xiao, F. and Lin, W. (2015) EFTUD2 on innate immunity. *Oncotarget* 32313–32314, https://doi.org/10.18632/oncotarget.5863

58 Sato, N., Maeda, M., Sugiyama, M., Ito, S., Hyodo, T., Masuda, A. et al. (2015) Inhibition of SNW 1 association with spliceosomal proteins promotes apoptosis in breast cancer cells. *Cancer Med.* **4**, 268–277, https://doi.org/10.1002/cam4.366

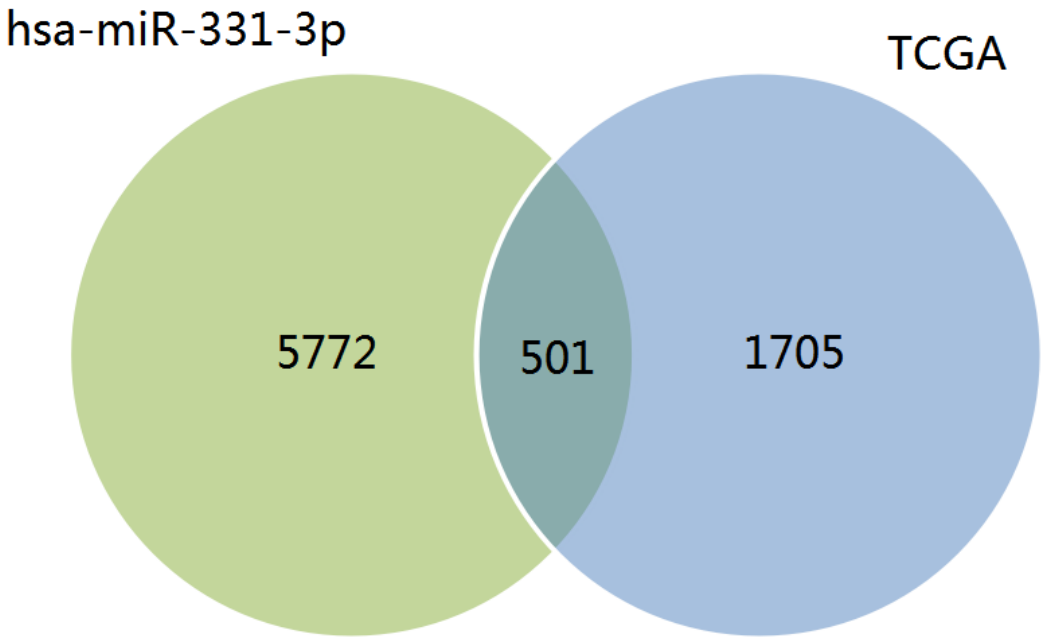
59 Lv, Z., Wang, Z., Luo, L., Chen, Y., Han, G., Wang, R. et al. (2019) Spliceosome protein Eftud2 promotes colitis-associated tumorigenesis by modulating inflammatory response of macrophage. *Mucosal Immunol.* **12**, 1164–1173, https://doi.org/10.1038/s41385-019-0184-y

60 Zhu, C., Xiao, F., Hong, J., Wang, K., Liu, X., Cai, D. et al. (2015) EFTUD2 is a novel innate immune regulator restricting hepatitis C virus infection through the RIG-I/MDA5 pathway. *J. Virol.* **89**, 6608, https://doi.org/10.1128/JVI.00364-15

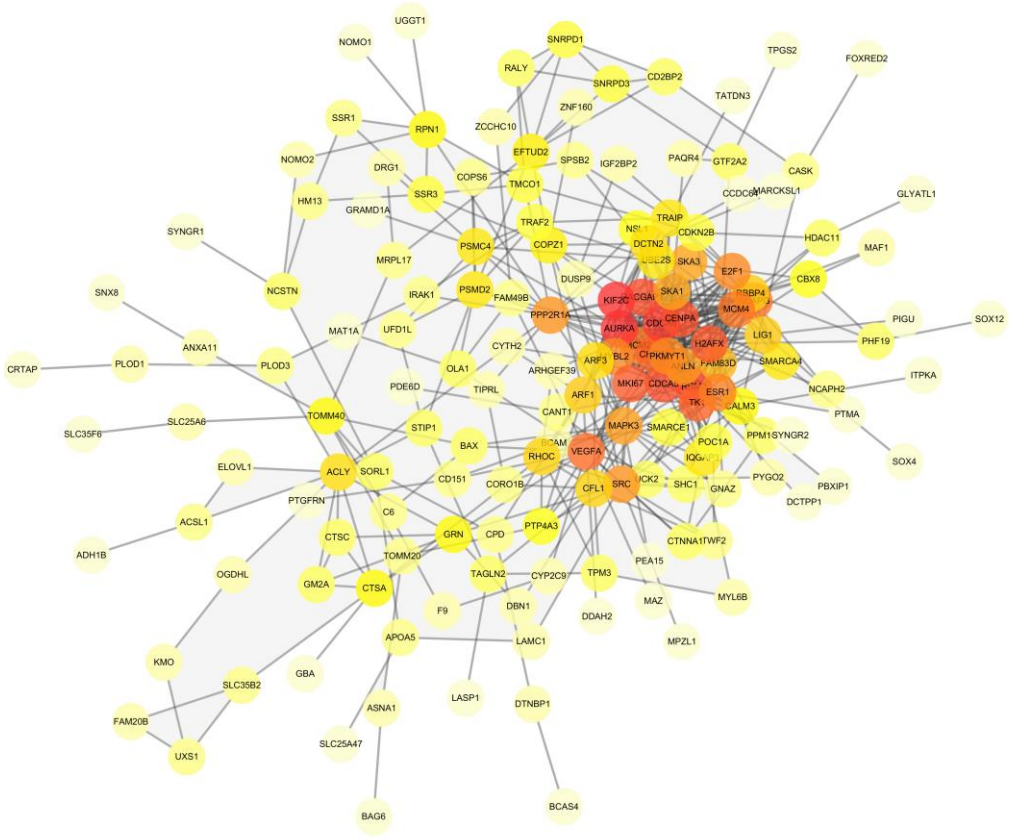
61 Zhu, C., Xiao, F., Wang, Q., Zhu, T., Kong, L. and Li, J. (2018) EFTUD2 restricts hepatitis B virus infection by regulating RIG-I expression. *J. Hepatol.* **68**, S776–S777, https://doi.org/10.1016/S0168-8278(18)31821-X



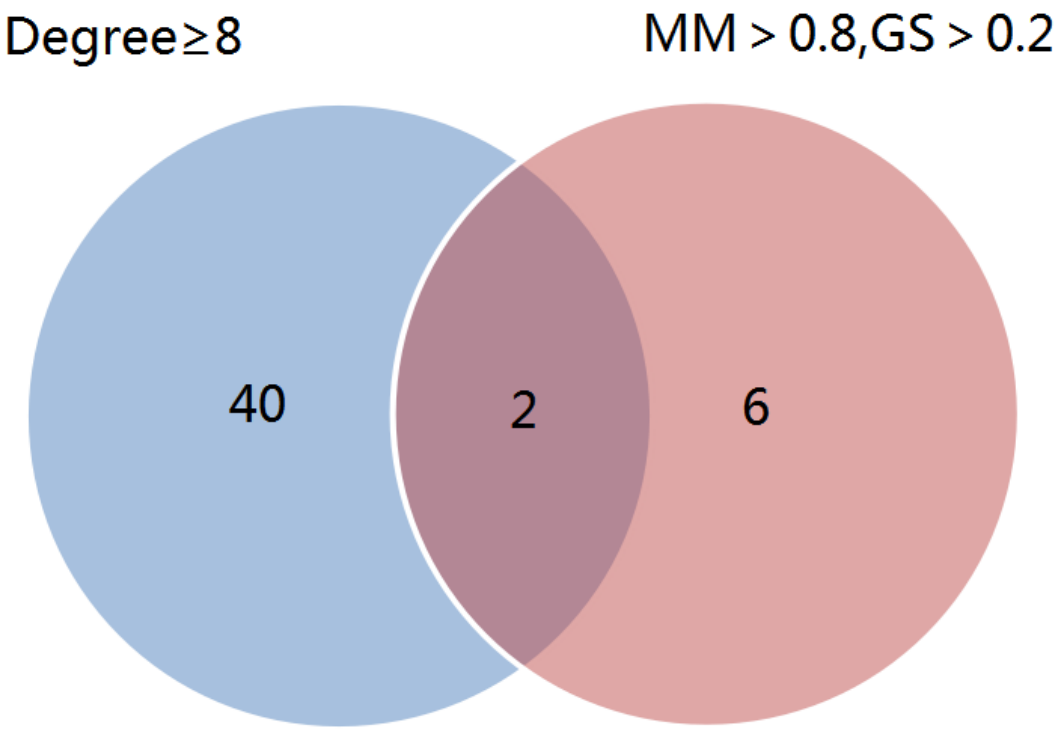
FigureS1. Work flow diagram.



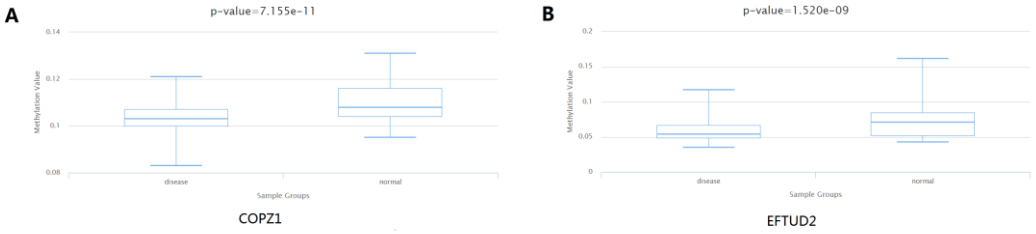
FigureS2. Venn plot of overlapping gene overlaps of miR-331-3p predicted by HCC-related DEGs and miRwalk2.0 in TCGA. Green: the miR-331-3p overlapping gene predicted by miRwalk2.0; blue :HCC-related DEGs in TCGA.Finally, 501 overlapping genes were obtained.



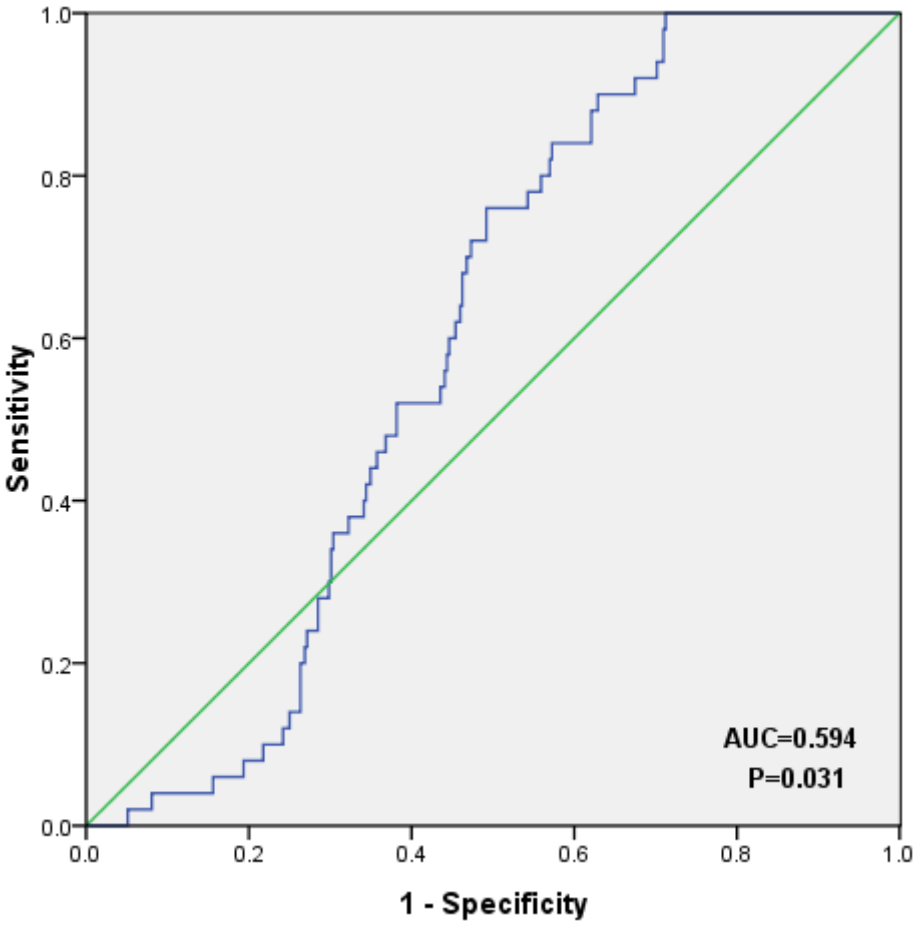
FigureS3. PPI network visualized by Cytoscape. The color of the node changes from yellow to red according to the Degree value.



FigureS4. Venn plot for hub genes. Hub genes were obtained from target genes in important modules (Module Membership> 0.8, Gene Significance> 0.2) and PPI analysis (Degree ≥ 8) of genes in important modules. Blue: PPI analysis (Degree ≥ 8) of genes in important modules; red: the target genes in important modules (Module Membership> 0.8, Gene Significance> 0.2).



FigureS5. Methylation analysis of the hub gene. The methylation levels of A.COPZ1 (p = 7.155e-11) and B.EFTUD2 (p = 1.520e-09) in HCC and normal tissues surrounding the tumor were obtained using DiseaseMeth 2.0.



FigureS6.ROC curve of miR-331-3p (AUC=0.594，p < 0.05).

Table S1. Overall Survival (OS)，yearofformcompletion，yearofinitialpathologicdiagnosis of HCC patients in the TCGA database.

	OS	Yearofformcompletion	Yearofinitialpathologicdiagnosis
TCGA-2Y-A9GT	1624	2014	2006
TCGA-BC-A10U	837	2011	2000
TCGA-2Y-A9GU	1939	2014	2009
TCGA-CC-A3MC	363	2012	2011
TCGA-G3-A25X	1779	2011	2007
TCGA-G3-A7M8	430	2014	2013
TCGA-2Y-A9GW	1271	2014	2007
TCGA-DD-AACM	1769	2014	2009
TCGA-DD-A39X	1694	2012	1998
TCGA-MR-A8JO	330	2013	2013
TCGA-ZP-A9CZ	706	2014	2011
TCGA-DD-AAD3	1295	2014	2011
TCGA-ZS-A9CF	2412	2014	2008
TCGA-G3-A7M6	632	2014	2013
TCGA-2Y-A9H5	555	2014	2010
TCGA-DD-AAE2	638	2014	2013
TCGA-XR-A8TF	693	2015	2012
TCGA-G3-AAV7	361	2014	2013
TCGA-5R-AA1D	449	2014	2013
TCGA-G3-A25S	416	2011	2009
TCGA-DD-AACE	2184	2014	2008
TCGA-WX-AA46	756	2014	2012
TCGA-XR-A8TG	898	2015	2012
TCGA-ED-A66Y	296	2013	2013
TCGA-DD-A115	2542	2010	2002
TCGA-LG-A9QC	425	2014	2013
TCGA-DD-A73E	44	2013	2012

TCGA-RC-A7SH	468	2014	2012
TCGA-EP-A12J	570	2011	2010
TCGA-UB-A7MF	214	2014	2013
TCGA-2Y-A9H9	697	2014	2011
TCGA-5C-A9VH	322	2014	2013
TCGA-G3-A25V	860	2011	2010
TCGA-ED-A66X	406	2013	2013
TCGA-DD-AADV	574	2014	2013
TCGA-DD-AAW2	1855	2014	2009
TCGA-WQ-AB4B	395	2014	2013
TCGA-ED-A7PX	6	2014	2013
TCGA-DD-A4NJ	928	2013	2011
TCGA-ZP-A9D4	395	2014	2012
TCGA-BC-A110	2116	2011	2003
TCGA-2Y-A9GY	757	2014	2007
TCGA-DD-AADO	453	2014	2013
TCGA-DD-A1EG	1372	2013	2007
TCGA-DD-AAE0	555	2014	2013
TCGA-DD-A4NB	989	2012	2011
TCGA-DD-AADK	1049	2014	2011
TCGA-MI-A75G	698	2013	2012
TCGA-FV-A2QR	581	2012	2006
TCGA-DD-AAE7	644	2014	2012
TCGA-2Y-A9H7	1168	2014	2011
TCGA-CC-A5UE	272	2014	2012
TCGA-K7-A5RG	519	2013	2012
TCGA-ED-A4XI	819	2013	2012
TCGA-RC-A7S9	640	2014	2012
TCGA-DD-AACY	1450	2014	2010
TCGA-FV-A3I1	247	2012	2011
TCGA-DD-A11C	662	2010	2006
TCGA-DD-AADD	1231	2014	2011

TCGA-FV-A496	10	2013	2012
TCGA-DD-A3A1	233	2012	2001
TCGA-DD-A4NE	660	2012	2011
TCGA-MI-A75E	507	2013	2013
TCGA-2Y-A9H0	3675	2014	2004
TCGA-2Y-A9H2	1731	2014	2010
TCGA-BC-A217	1397	2011	2010
TCGA-DD-AADG	1145	2014	2011
TCGA-DD-AACG	469	2014	2008
TCGA-DD-AACK	9	2014	2008
TCGA-NI-A8LF	799	2014	2012
TCGA-DD-AAD1	564	2014	2010
TCGA-DD-AADN	898	2014	2012
TCGA-ZP-A9CY	782	2014	2011
TCGA-G3-A5SJ	698	2013	2012
TCGA-BC-A3KF	8	2012	2011
TCGA-FV-A3I0	848	2012	2011
TCGA-DD-AAEB	478	2014	2012
TCGA-2Y-A9GZ	848	2014	2007
TCGA-CC-A8HS	300	2014	2013
TCGA-DD-A1EB	2017	2011	2001
TCGA-DD-AADY	555	2014	2013
TCGA-DD-AAVP	2752	2014	2007
TCGA-G3-A5SK	744	2013	2012
TCGA-RC-A7SF	579	2014	2012
TCGA-G3-AAV3	412	2014	2013
TCGA-KR-A7K8	906	2014	2012
TCGA-ED-A627	423	2013	2012
TCGA-BC-A10Z	34	2011	2003
TCGA-EP-A3RK	363	2012	2012
TCGA-UB-A7ME	486	2014	2013
TCGA-ED-A7XP	400	2013	2013

TCGA-UB-A7MA	848	2014	2013
TCGA-5R-AAAM	46	2014	2013
TCGA-DD-AACP	415	2014	2009
TCGA-FV-A2QQ	729	2012	2011
TCGA-DD-AACJ	2102	2014	2008
TCGA-HP-A5MZ	91	2013	Not Available
TCGA-CC-A8HT	140	2014	2013
TCGA-CC-5259	250	2012	2010
TCGA-DD-A4NK	1210	2013	2004
TCGA-RC-A7SK	472	2014	2013
TCGA-DD-AAEE	810	2014	2012
TCGA-YA-A8S7	412	2014	2012
TCGA-BC-A10W	91	2011	2002
TCGA-CC-A5UC	347	2014	2012
TCGA-DD-A1EI	183	2011	2008
TCGA-DD-AAD2	658	2014	2010
TCGA-DD-AAD6	672	2014	2011
TCGA-XR-A8TC	1339	2015	2011
TCGA-UB-A7MB	601	2014	2013
TCGA-DD-A4NA	1008	2012	2009
TCGA-DD-A4NG	802	2013	2011
TCGA-2Y-A9GV	2532	2014	2007
TCGA-DD-AAD0	137	2014	2010
TCGA-DD-AAE9	722	2014	2012
TCGA-BC-4073	849	2010	2009
TCGA-DD-AACF	365	2014	2008
TCGA-DD-A11A	79	2010	2004
TCGA-DD-A73A	728	2013	2012
TCGA-4R-AA8I	262	2014	2013
TCGA-DD-AAW0	2015	2014	2009
TCGA-BW-A5NP	0	2014	2010
TCGA-DD-A3A2	2131	2012	1998

TCGA-DD-A4NO	2245	2013	2007
TCGA-KR-A7K2	829	2014	2013
TCGA-BC-A10Y	711	2011	2002
TCGA-CC-A9FW	248	2014	2013
TCGA-DD-AACA	2301	2014	2008
TCGA-ZP-A9D0	1091	2014	2011
TCGA-PD-A5DF	639	2013	2007
TCGA-DD-AAVS	1823	2014	2007
TCGA-G3-A7M5	447	2014	2013
TCGA-DD-A1E9	2759	2011	2002
TCGA-CC-A123	219	2011	2010
TCGA-G3-AAV1	359	2014	2013
TCGA-BC-A69H	444	2013	2013
TCGA-BC-A3KG	680	2012	2011
TCGA-DD-AAW1	1989	2014	2009
TCGA-WX-AA44	615	2014	2011
TCGA-DD-AACH	195	2014	2008
TCGA-DD-A73B	283	2013	2012
TCGA-DD-AAW3	1633	2014	2009
TCGA-DD-AAVV	2455	2014	2007
TCGA-RC-A6M4	22	2015	2010
TCGA-RC-A7SB	588	2014	2012
TCGA-BW-A5NO	20	2014	2010
TCGA-DD-A73C	701	2013	2012
TCGA-DD-AADA	1233	2014	2011
TCGA-FV-A4ZQ	12	2013	2012
TCGA-2Y-A9H4	1452	2014	2010
TCGA-DD-AADB	1242	2014	2011
TCGA-DD-AACL	107	2014	2008
TCGA-DD-AAEK	1067	2014	2010
TCGA-DD-A73G	3478	2013	2005
TCGA-DD-AACS	1804	2014	2009

TCGA-G3-AAV2	372	2014	2013
TCGA-DD-AADS	474	2014	2013
TCGA-BC-A5W4	547	2013	2012
TCGA-MR-A520	229	2012	2012
TCGA-XR-A8TE	925	2015	2012
TCGA-DD-A114	1149	2010	2005
TCGA-DD-AAVR	2513	2014	2007
TCGA-RC-A6M3	0	2015	2009
TCGA-3K-AAZ8	396	2014	2013
TCGA-G3-A5SM	520	2013	2012
TCGA-O8-A75V	538	2013	2013
TCGA-ZS-A9CD	1386	2014	2010
TCGA-2Y-A9H1	1229	2014	2010
TCGA-2Y-A9HB	260	2014	2012
TCGA-DD-AACO	1876	2014	2009
TCGA-G3-AAV5	354	2014	2013
TCGA-LG-A9QD	366	2014	2013
TCGA-ZP-A9D2	765	2014	2012
TCGA-DD-AAVW	2317	2014	2008
TCGA-DD-AAEI	1531	2014	2010
TCGA-CC-A9FU	0	2014	2013
TCGA-DD-AACZ	171	2014	2010
TCGA-5C-AAPD	20	2014	2013
TCGA-2Y-A9H8	633	2014	2011
TCGA-DD-AA3A	410	2014	2009
TCGA-BD-A2L6	1363	2012	2011
TCGA-DD-AADM	12	2014	2012
TCGA-DD-A3A3	535	2012	1999
TCGA-DD-A1EJ	1005	2011	2008
TCGA-UB-AA0V	314	2014	2013
TCGA-DD-AAC8	16	2014	2008
TCGA-DD-A3A4	612	2012	1998

TCGA-GJ-A6C0	31	2013	2012
TCGA-CC-A7II	399	2014	2013
TCGA-EP-A3JL	303	2012	2011
TCGA-DD-A73D	693	2013	2012
TCGA-BC-A10R	308	2011	1999
TCGA-DD-AADL	636	2014	2012
TCGA-BW-A5NQ	0	2014	2010
TCGA-DD-AAC9	347	2014	2008
TCGA-DD-A4NQ	373	2013	2007
TCGA-G3-A7M9	56	2014	2013
TCGA-DD-A4ND	2746	2012	2007
TCGA-5C-A9VG	328	2014	2013
TCGA-5R-AA1C	520	2014	2013
TCGA-ED-A82E	408	2014	2013
TCGA-G3-AAV4	27	2014	2012
TCGA-CC-A7IK	262	2014	2013
TCGA-G3-A3CK	585	2012	2011
TCGA-CC-A8HV	279	2014	2013
TCGA-BC-A10T	837	2011	2000
TCGA-G3-A3CG	673	2012	2011
TCGA-DD-AACT	1562	2014	2010
TCGA-DD-A73F	1085	2013	2010
TCGA-G3-A25Z	655	2011	2010
TCGA-CC-A3MB	315	2012	2011
TCGA-DD-A1EK	558	2013	2008
TCGA-DD-A116	1622	2010	2002
TCGA-DD-AAE1	552	2014	2013
TCGA-DD-A1EL	415	2011	2009
TCGA-2Y-A9H3	1516	2014	2010
TCGA-DD-AADJ	1066	2014	2011
TCGA-RG-A7D4	1098	2014	2012
TCGA-DD-AAVU	2202	2014	2008

TCGA-G3-A25W	935	2011	2010
TCGA-K7-A5RF	631	2013	2012
TCGA-G3-A7M7	361	2014	2013
TCGA-BD-A3ER	1115	2012	2011
TCGA-FV-A3R3	366	2012	2012
TCGA-DD-AAD8	1219	2014	2011
TCGA-UB-AA0U	327	2014	2013
TCGA-CC-5262	103	2011	2010
TCGA-CC-A5UD	304	2014	2012
TCGA-DD-AACV	1531	2014	2010
TCGA-CC-A8HU	344	2014	2013
TCGA-MI-A75C	291	2013	2012
TCGA-DD-A4NP	3308	2013	2005
TCGA-MI-A75I	630	2013	2011
TCGA-MI-A75H	747	2013	2012
TCGA-DD-AACU	1567	2014	2010
TCGA-ED-A459	910	2013	2012
TCGA-BC-4072	1490	2010	2005
TCGA-BC-A8YO	562	2014	2013
TCGA-DD-AADF	115	2014	2011
TCGA-DD-A119	223	2010	2003
TCGA-DD-A4NI	816	2013	2011
TCGA-WQ-A9G7	30	2014	2013
TCGA-G3-AAUZ	480	2014	2013
TCGA-DD-AACW	1424	2014	2010
TCGA-G3-A25Y	452	2011	2006
TCGA-FV-A23B	1852	2011	2005
TCGA-DD-AACB	2324	2014	2008
TCGA-ZP-A9D1	21	2014	2012
TCGA-LG-A6GG	387	2013	2013
TCGA-DD-AACD	381	2014	2008
TCGA-G3-A3CJ	594	2012	2011

TCGA-G3-A5SL	621	2013	2012
TCGA-ED-A5KG	854	2013	2012
TCGA-KR-A7K0	65	2014	2013
TCGA-DD-AADE	1202	2014	2011
TCGA-CC-A7IE	217	2014	2012
TCGA-XR-A8TD	1030	2015	2012
TCGA-2Y-A9H6	357	2014	2011
TCGA-BC-A69I	387	2013	2013
TCGA-DD-A39V	643	2012	2000
TCGA-DD-AAVQ	2728	2014	2007
TCGA-ED-A8O6	56	2014	2013
TCGA-G3-A3CI	180	2012	2011
TCGA-CC-5261	97	2011	2010
TCGA-EP-A2KB	596	2011	2011
TCGA-RC-A6M5	15	2015	2012
TCGA-DD-A11D	1560	2010	2006
TCGA-G3-A25U	1636	2011	2007
TCGA-G3-A25T	1553	2011	2007
TCGA-KR-A7K7	951	2014	2012
TCGA-DD-AADU	554	2014	2013
TCGA-WX-AA47	556	2014	2012
TCGA-DD-A39Y	171	2012	2002
TCGA-DD-AAD5	1345	2014	2011
TCGA-DD-A113	2425	2010	2006
TCGA-EP-A2KC	19	2011	2011
TCGA-DD-A1EF	394	2011	2007
TCGA-DD-AACC	1685	2014	2008
TCGA-ES-A2HS	688	2011	2008
TCGA-CC-A7IL	278	2014	2013
TCGA-DD-A3A9	931	2012	2001
TCGA-DD-AACI	1618	2014	2008
TCGA-DD-AAVY	1970	2014	2009

TCGA-DD-AAVX	1718	2014	2008
TCGA-RC-A6M6	9	2015	2011
TCGA-CC-A9FS	211	2014	2013
TCGA-G3-AAV6	65	2014	2013
TCGA-CC-5264	102	2011	2010
TCGA-NI-A4U2	1791	2013	2005
TCGA-CC-5263	129	2011	2010
TCGA-K7-A6G5	512	2013	2013
TCGA-DD-A39W	827	2012	1995
TCGA-CC-A7IJ	382	2014	2013
TCGA-BC-A10Q	1135	2011	1998
TCGA-ZP-A9CV	1088	2014	2011
TCGA-ZS-A9CG	341	2014	2013
TCGA-DD-A3A6	3258	2012	1997
TCGA-DD-AACQ	432	2014	2009
TCGA-DD-A4NN	899	2013	2004
TCGA-ED-A7PZ	6	2014	2013
TCGA-DD-AAVZ	1900	2014	2009
TCGA-WJ-A86L	345	2014	2013
TCGA-FV-A495	1	2013	2012
TCGA-DD-AAE6	141	2014	2012
TCGA-GJ-A9DB	67	2014	2013
TCGA-BC-A10S	1423	2011	1999
TCGA-DD-AADR	2028	2014	2009
TCGA-DD-A3A7	419	2013	1998
TCGA-CC-A7IF	649	2014	2013
TCGA-CC-A9FV	0	2014	2013
TCGA-DD-AAED	763	2014	2012
TCGA-DD-AADQ	436	2014	2013
TCGA-DD-AAE4	608	2014	2013
TCGA-GJ-A3OU	879	2015	2011
TCGA-BC-A216	1351	2011	2008

TCGA-BC-A112	153	2011	2009
TCGA-FV-A3R2	194	2012	2005
TCGA-DD-A4NR	9	2013	2005
TCGA-G3-A5SI	768	2013	2011
TCGA-DD-A39Z	601	2012	1996
TCGA-CC-5258	129	2011	2010
TCGA-DD-A3A8	11	2012	2001
TCGA-DD-AADI	1085	2014	2011
TCGA-DD-A3A5	3125	2012	1996
TCGA-DD-AACX	170	2014	2010
TCGA-2Y-A9HA	36	2014	2012
TCGA-DD-AADW	587	2014	2013
TCGA-K7-AAU7	359	2014	2013
TCGA-BC-A10X	770	2011	2002
TCGA-DD-AADC	425	2014	2011
TCGA-EP-A26S	608	2011	2011
TCGA-DD-AACN	1302	2014	2009
TCGA-DD-AAEA	575	2014	2012
TCGA-CC-A3M9	300	2012	2011
TCGA-DD-A11B	14	2010	2005
TCGA-DD-A4NF	942	2013	2011
TCGA-DD-A1EH	1495	2011	2008
TCGA-CC-5260	87	2011	2010
TCGA-ED-A7XO	427	2013	2013
TCGA-2Y-A9GX	2442	2014	2007
TCGA-DD-AAEG	719	2014	2012
TCGA-CC-A3MA	303	2012	2011
TCGA-ED-A8O5	406	2014	2013
TCGA-CC-A7IH	365	2014	2013
TCGA-QA-A7B7	94	2013	2013
TCGA-ES-A2HT	438	2011	2007
TCGA-DD-A3A0	785	2012	2001

TCGA-G3-A3CH	780	2012	2010
TCGA-UB-A7MD	52	2014	2013
TCGA-DD-AADP	458	2014	2013
TCGA-DD-AAEH	784	2014	2012
TCGA-DD-A1EA	2415	2011	2002
TCGA-DD-A4NH	917	2013	2011
TCGA-DD-A1EE	349	2011	2006
TCGA-DD-A1ED	2301	2011	2006
TCGA-DD-A4NL	1711	2013	2007
TCGA-HP-A5N0	752	2014	Not Available
TCGA-2Y-A9GS	724	2014	2006
TCGA-BD-A3EP	409	2012	2010
TCGA-G3-AAV0	476	2014	2013
TCGA-DD-A4NS	2456	2013	2005
TCGA-DD-A118	3437	2010	2003
TCGA-EP-A2KA	627	2011	2011
TCGA-ZS-A9CE	1241	2014	2011
TCGA-ED-A7PY	390	2014	2013
TCGA-G3-A6UC	671	2013	2012
TCGA-FV-A4ZP	2486	2013	2005
TCGA-UB-A7MC	500	2014	2013
TCGA-DD-A1EC	602	2011	2004
TCGA-T1-A6J8	23	2014	2013
TCGA-DD-AAE8	664	2014	2012
TCGA-DD-A4NV	2398	2013	2007
TCGA-CC-A7IG	299	2014	2013
TCGA-DD-AAE3	566	2014	2013
TCGA-CC-A1HT	101	2011	2010
TCGA-ED-A97K	6	2014	2013