Identification of Prognostic Biomarkers and Correlation with Immune Infiltrates in Hepatocellular Carcinoma Based on a Competing Endogenous RNA Network

Zhangya Pu 1, Yuanyuan Zhu 2, Xiaofang Wang 1, Yun Zhong 1, Fang Peng 2 *, Yiya Zhang 3 4 *

1 Department of Infectious Diseases, Hunan Key Laboratory of Viral Hepatitis, Xiangya Hospital, Central South University, Changsha, Hunan Province, 410008, China.
2 NHC Key Laboratory of Cancer Proteomics, Xiangya Hospital, Central South University, 87 Xiangya Road, Changsha, Hunan Province, 410008, China.
3 National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, Hunan, China
4 Department of Dermatology, Xiangya Hospital, Changsha, Hunan Province, 410008, China.

* Correspondence
Fang Peng, email: pengfang@csu.edu.cn
Yiya Zhang, email: yiya0108@csu.edu.cn.
Abstract

**Background:** Hepatocellular carcinoma (HCC) is one of the most common malignant tumors worldwide. Recently, competing endogenous RNAs (ceRNA) have revealed a significant role in the progression of HCC. Herein, we aimed to construct a ceRNA network to identify potential biomarkers and illustrate its correlation with immune infiltration in HCC.

**Methods:** RNA sequencing data and clinical traits of HCC patients were downloaded from TCGA. The limma R package was used to identify differentially expressed (DE) RNAs. The predicted prognostic model was established using univariate and multivariate Cox regression. A K-M curve and GEPIA website were utilized for survival analysis. Functional annotation was determined using Enrichr and Reactome. Protein-to-protein network analysis was implemented using SRTNG and Cytoscape. Hub gene expression was validated by Oncomine and the Hunan Protein Atlas database. Immune infiltration was analyzed by TIMMER, and Drugbank was exploited to identify bioactive compounds.

**Results:** The predicted model that was established revealed significant efficacy with 3- and 5-years of the area under ROC at 0.804 and 0.744, respectively. Eleven DEmiRNAs were screened out by a K-M survival analysis. Then, we constructed a ceRNA network, including 56 DElncRNAs, 6 DEmiRNAs, and 28 DEmRNAs. The 28 DEmRNAs were enriched in cancer-related pathways, for example, the TNF signaling pathway. Moreover, six hub genes, CEP55, DEPDC1, KIF23, CLSPN, MYBL2, and RACGAP1, were all overexpressed in HCC tissues and independently correlated with survival rate. Furthermore, expression of hub genes was related to immune cell infiltration in HCC, including B cells, CD8⁺ T cells, CD4⁺ T cells, monocytes, macrophages, neutrophils, and dendritic cells.

**Conclusions:** The findings from this study demonstrate that CEP55, DEPDC1, KIF23, CLSPN, MYBL2, and RACGAP1 are closely associated with prognosis and immune infiltration, representing potential therapeutic targets or prognostic biomarkers in HCC.

**Keywords:** Hepatocellular carcinoma (HCC), competing endogenous RNA network (ceRNA), immune infiltration, prognostic prediction model
Background

Hepatocellular carcinoma (HCC) is one of the most universally malignant tumors in the world, with increasing morbidity and mortality [1, 2]. Currently, it is the fifth most common cancer and the fourth leading cause of cancer-related death worldwide[3, 4]. Tumorigenesis of HCC is correlated with several liver primary diseases, such as viral infections, including Hepatitis B virus (HBV), Hepatitis C virus (HCV), and other kinds of hepatotropic viruses, as well as alcoholic liver diseases, dietary aflatoxin exposure diabetes, and other diseases[5, 6]. Despite continuous improvement in the methods of diagnosis and treatment, HCC remains a global clinical challenge due to its poor prognosis and low rate of 5-year survival[7, 8]. Therefore, individual strategies based on identifying early potential prognostic biomarkers and novel therapeutic targets are urgently needed.

The correlation between protein-coding messenger RNA (mRNA) and noncoding RNA (ncRNA), including long noncoding RNAs (lncRNA) and microRNAs (miRNAs), is complicated and obscure. In 2011, the competing endogenous RNA (ceRNA) hypothesis was elucidated for the first time by Salmena et al, demonstrating that ncRNAs not only directly take part in the regulation of targeted gene expression but also absorb corresponding miRNAs as natural sponges due to their typically containing more than one miRNA response element (MRE) that competes with mRNA[9]. Recently, increasing evidence indicates that the regulatory network comprised of lncRNA-miRNA-mRNA plays an important role in the physiology and development of various tumors, including HCC, gallbladder cancer, gastric cancer, and others[3, 4, 7]. Zhang et al. indicated that lncRNA-correlated ceRNA networks are involved in diverse biological cancer pathways in glioblastoma [10]. Wang et al identified six lncRNAs, including LINC00536 and MIR7-3HG, that have a significant effect on overall survival in breast cancer [11]. Nevertheless, current studies based on ceRNA networks in multiple databases for HCC are insufficient.

Recently, increasing attention has been paid to immunotherapy research in various cancers, especially in advanced stages, including for mesothelioma, HCC, and others. However, the benefits from immunotherapies are diverse in various tumors and are difficult to evaluate due to a lack of trustworthy immune-related biomarkers [12, 13]. Immune-related cells infiltration into the tumor microenvironment (TME) is a key reason leading to immune responses at primary and secondary tumor sites, which is tightly regulated by various mediators, such as chemokines. Several studies have indicated that a variety of different immune cells, including CD4+ and CD8+T-cells, dendritic cells, and tumor-associated macrophages (TAMs), have been identified in different cancers, such as prostate cancer, HCC and others [13, 14]. Chunying et al also demonstrated that CD4+ and CD8+T-cells could be recruited into the TME after CXCR4 inhibition in sorafenib-treated HCC in a mouse model [15]. Therefore, there is an
impending requirement to identify potential predictors related to immune cell infiltration to enhance the efficacy of individual immunotherapeutic treatment in tumors.

Study design is recapitulated in Figure 1. First, differentially expressed RNAs were analyzed in 371 cases of HCC and 50 normal liver tissues from The Cancer Genome Atlas (TCGA). Subsequently, a nomogram predicted model based on 23 miRNAs was established and revealed high performance. Next, we constructed a ceRNA network composed of 56 DELs, 6 DEMs, and 28 DEGs to illustrate preliminary interactions between mRNAs and ncRNAs. DEGs correlating to the ceRNA network were submitted for Gene Ontology (GO) and pathway enrichment analysis to clarify the underlying molecular mechanism in HCC. Finally, six hub genes, including CEP55, DEPDC1, CLSPN, KIF23, MYBL2, and RACGAP1, were identified by protein-to-protein (PPI) analysis and were closely associated with immune infiltration in HCC. In summary, we believe these genes represent potential prognostic markers and immunological therapeutic targets for HCC treatment that should further explored in the future.

Methods

Data acquisition from TCGA database

A total of 371 HCC samples and 50 adjacent normal liver tissues were included in this study. RNA sequencing (RNA-Seq), including lncRNA, mRNA (Illumina HiSeq RNA-Seq platform) and miRNA sequence data (Illumina HiSeq miRNA-Seq platform), were downloaded from TCGA database (https://portal.gdc.cancer.gov/, version 10.1, release time: February 15, 2018), and survival data were manually extracted. The present study conformed to the publication guidelines required by TCGA. The RNA sequence data were annotated based on the Ensemble gene ID. Log2 transformation was performed on all gene expression profiling. Then, Limma package (Version: 3.38.3) in R software (Version:3.5.2) was used to normalize the original data.

Identification of differentially expressed (DE) RNAs

The expression profile of RNA sequencing data retrieved from TCGA was analyzed using the limma package of R software (https://www.r-project.org/) with the criterial of |log2 fold change|> 1 and the adjusted false discovery rate (FDR) of P < 0.05. Screened DE RNAs, including differentially expressed lncRNAs (DELs), differentially expressed miRNAs (DEM), and differentially expressed mRNAs (DEGs), were used for subsequent analysis. Heat maps and volcano plots for DE RNAs were created using the heatmap package of R software.
Univariate and multivariate Cox regression analysis.

Univariate Cox regression was used to screen for potential prognostic miRNAs correlated with overall survival in HCC patients. The least absolute shrinkage and selection operator (LASSO) detects the most influential variables because it analyzes all independent variables simultaneously. According to the principle of a penalty following a regularization path, the coefficients of less influential variables would trend toward zero. The glmnet package was used to perform LASSO algorithm with the criterial of P<0.05. Multivariate Cox regression analysis via survival R package was utilized to establish a prognostic predictive model visualized by nomogram to show the correlation of the expression of DEMs and survival rate of specific HCC patients. The forest plot was created to display the results of multivariate Cox regression using the forestplot package.

Evaluation of miRNA-based clinical predictive model

To evaluate the predictive performance of the prognostic model based on DEMs, first, a calibration curve of 3- and 5-year survival rates was determined to assess agreement between the predictive model and actual survival time. Moreover, the area under the curve (AUC) was calculated according to the time-dependent receiver operating characteristic analysis (ROC). Additionally, the risk score formula was performed to calculate total risk scores for individual patients based on the coefficient for each DEM. The risk score formula was built according to the following method: total risk score = sum of each coefficient×transcriptional expressed value of DEM. Then, HCC patients were divided into high- and low-risk groups by the median risk scores, regarded as the cutoff value. The difference in survival rate between the two groups was also evaluated. The correlation between expression levels of DEMs and OS in HCC patients was calculated by Kaplan-Meier (K-M) survival analysis using the survival package of R software according to the X tile method with a cutoff P-value < 0.05.

Establishment of the ceRNA regulatory network

A co-expressed regulatory network comprised of DELs, DEMs, and DEGs was established to explore the potential functions of these DE RNAs in HCC. The interaction between DEMs and DELs was confirmed using the miRcode database (http://www.mircode.org/), which not only includes putative target sites of miRNAs from the integrated and searchable map but also contains conserved microRNA families annotated by the ENCyclopedia of DNA Elements (ENCODE) [16]. DEM targets were predicted from three databases, including miRDB (http://www.mirdb.org/), miRTarBase (http://mirtarbase.mbc.nctu.edu.tw) and TargetScan (http://www.targetscan.org) [17-19]. Overlapping DEGs were selected for constructing the ceRNA network. Cytoscape (https://cytoscape.org/) software was used to visualize the expression correlation of DE RNAs.
Functional annotation and PPI network analysis

Gene Ontology (GO) analysis of differentially expressed genes, including biological process (BP), molecular function (MF) and cell components (CC), and Kyoto encyclopedia of genes and genomes (KEGG) pathway analysis, were enriched using the Enrich online tool (http://www.enrichnet.org/). The Reactome pathway was determined by the Reactome website (https://reactome.org/). The online STRING (https://string-db.org/) tool was used to construct the protein-to-protein (PPI) interaction network for DEGs involved in the ceRNA network and was visualized by Cytoscape. Hub genes were defined as the top six genes with the highest degree of connections to others via the CytoHubba plug-in of Cytoscape.

Correlation of hub genes and immune infiltration analysis

TIMMER is a comprehensive online database used for systematic analysis of the correlation of immune infiltration and gene markers of interest in 32 cancer types from TCGA (https://cistrome.shinyapps.io/timer/). The abundance of tumor-infiltrating immune cells (TIICs) from gene profiles is evaluated based on the statistical method of deconvolution published previously [20, 21]. We analyzed expressed levels of hub genes in various tumors and the relationship between the expression of hub genes and immune infiltration, including B cells, CD4+ T cells, CD8+ T cells, neutrophils, macrophages, and dendritic cells. Furthermore, the correlation of hub gene expression and gene markers associated with tumor immune infiltration of monocytes, tumor associated macrophages (TAMs), M1 and M2 macrophages was performed via correlation modules. These gene markers were identified in previous studies [22-24]. The strength of the correlation was evaluated by Spearman’s algorithm divided into five levels: very weak (0.00-0.19), weak (0.20-0.39), moderate (0.40-0.59), strong (0.60-0.79) and very strong (0.80-1.0). The cutoff criteria for statistical significance was P-value <0.05. In addition, the SCNA module was used to identify differences in tumor infiltration of hub genes in HCC with different somatic copy number alterations.

Results

Identification of differentially expressed RNAs

A total of 371 HCC samples and 50 nontumor tissues were included in this study. Differentially expressed RNAs were analyzed using the limma R package with a screening cutoff threshold of |log2-fold change| > 1 and an adjusted P-value < 0.05. 1999 DEGs, 251 DEMs, and 1092 DELs were identified as differentially expressed RNAs. Among them, there were 1794 DEGs, 229 DEMs, and 1034 DELs upregulated. Moreover, differential expression of DEGs, DEMs and DELs is displayed by hierarchical clustering and volcano plots (Figure 2).
Establishment of miRNA-based prognostic predictive model in HCC

A total of 42 DEMs survival-related miRNAs were screened by univariate Cox regression from 251 DEMs identified in the HCC cohort (Table S1, S2). Next, these miRNAs were included in the LASSO analysis to calculate the corresponding coefficients. Twenty-three DEMs significantly correlated with survival were selected out (Figure S1, Table S3). A simple-to-use nomogram predictive model was established to describe correlation of the expression of each miRNA and the 3- and 5-year overall survival rate of HCC patients based on multivariate Cox regression (Figure 3, 4). Meanwhile, the 3- and 5-year calibration curves were drawn, showing good consistency between predicted survival probability and the actual survival rate (Figure 5 A-B). The ROC curve also exhibited great reliability for the nomogram prediction model in discriminating tumors from normal tissues with the area under curve (AUC) of 3- and 5-year being 0.804 and 0.744, respectively (Figure 5C). Furthermore, the risk score (RS) of each patient in the HCC cohort was calculated, and the patients were subsequently divided into high-risk and low-risk groups according to the mean RS. K-M analysis indicated that patients in the high-risk group exhibited decreased survival compared to the low-risk group, with a log-rank P-value <0.05 (Figure 5D).

Construction of the lncRNA-miRNA-mRNA regulatory network

K-M analysis performed for 23 DEMs significantly related with OS revealed that 11 DEMs, including 8 that were upregulated, were independently statistically significant (Table S4). Results showed that HCC patients with highly expressed DEMs had a shorter survival time than those with lower expression with a P-value < 0.05 (Figure 6). Recently, increasing evidence has indicated that miRNAs play a significant role in the development and metastasis of tumors. To reveal potential signaling pathways regulated by these DEMs in HCC, the DIANA-miRPath database was exploited and revealed enrichment of cancer-related signaling pathways, such as PI3K-AKT, NF-kappa B, VEGF and others (Figure 7).

The lncRNA targeted by the 11 DEMs was screened based on the interactions with 1092 DELs aforementioned. Six DEMs were targeted by 56 DELs according to the miRcode database. Next, targets of the 6 DEMs were predicted using miRTarBas, miRDB, and TargetScan databases. The overlapping 574 mRNAs predicted in all three databases were further intersected with 1999 DEGs identified in the HCC cohort, and only 28 DEGs existed in both groups (Figure S2, Table S5). The representative interactions among 56 DELs, 6 DEMs, and 28 DEGs are summarized in Table 1, and the ceRNA regulatory network, including gene nodes and preliminary interactions, was visualized using Cytoscape software (Figure 8).

Function and pathway enrichment analysis of DEGs involved in the ceRNA network
GO and pathway enrichment analyses were performed to elucidate the functions of 28 DEGs in the ceRNA network correlated with the progression of HCC. Functional annotation of biological process (BP), cellular component (CC) and molecular function (MF), as well as KEGG pathway enrichment, were performed on the Enrichr comprehensive database, in which the top 10 highly enriched items for BP, CC, MF and KEGG pathway are shown based on a P-value <0.05 (Figure 9 A-D). Notably, all top 10 items were closely related to cancer-related pathways, such as TNF signaling, breast cancer, small cell lung cancer and others. Moreover, Reactome pathway analysis was also developed to identify possible metabolic pathways in which the 28 DEGs are involved. A total of 28 pathways were identified, and the top 15 highly enriched pathways are presented in Figure 9E.

**Construction of PPI network and identification of hub genes**

To further investigate the function of 28 DEGs associated with the ceRNA network at the protein level, we established a protein-to-protein interaction (PPI) network composed of 109 nodes and 218 degrees to visualize detailed interactions (Figure 10A). Considering the significance of hub genes in the ceRNA network, the CytoHubba plugin in Cytoscape software was exploited to identify hub genes by evaluating the number of degrees and connections. Finally, six hub genes, CEP55, DEPDC1, MYBL2, RACGAP1, CLSPN, and KIF23, were identified, which were all upregulated in HCC cohort (Figure 10 B). The filled color of nodes from red to yellow indicates the degree of connectivity of hub genes with others gradually decreases. GO and pathway enrichment, including KEGG and Reactome analyses, were also performed (Figure S3). Additionally, the sub ceRNA network, including 46 DELs, 3 DEMs (hsa-mir-30d, hasa-mir-195, has-mir-301a), and 6 hub genes, was built to delineate correlations among the DELs, DEMs and hub genes (Figure 10 C, Table S6).

**Validation of survival analysis and expression of hub genes**

The correlation between expression of hub genes and OS and recurrence-free survival (RFS) was assessed using the GEPIA website. HCC patients with high expression levels of hub genes had a lower survival rate with respect to both OS and RFS. Among these, CEP55 had the highest prognostic P-value (0.00033 and 0.00063 correlated to overall and recurrence-free survival rates, respectively (Figure 11)). mRNA levels of hub genes in various tumor and normal tissues were examined using the TIMMER database. Results indicated that expression of hub genes was higher in various tumors than in corresponding normal tissues in breast cancer, colorectal cancer, gastric cancer and others (Figure S4). Next, we validated transcriptional expression of hub genes in another HCC cohort from the Oncomine database, which demonstrated overexpression of genes in the tumor group compared to nontumor tissues as well with cutoff P-value < 0.01 (Figure 12). Moreover, immunohistochemical data from the
Human Protein Atlas was used to verify protein expression of hub genes. Data for CLSPN was lacking in the database, and expression of DEPDC1 in both HCC and normal samples was not detected. However, the staining intensity or the range of positive areas of CEP55, KIF23, MYBL2, and RACGAP1 was higher in tumor samples (medium or high levels) than in nontumor tissue (Figure 13).

**Correlation of hub gene expression and immune infiltration**

Recently, increasing evidence has demonstrated that tumor-infiltrating lymphocytes play a significant role in predicting lymph node status and survival in tumors. Therefore, we investigated whether expression of hub genes was related to immune-infiltrating levels in HCC using the TIMMER database. Results revealed that six hub genes were significantly related to immune infiltration of B cells, CD4+ T cells, CD8+ T cells, macrophages, neutrophils, and dendritic cells in HCC, with statistical P-values of <0.0001. However, expression of CLSPN (cor = 0.08, P = 1.39e-01), CEP55 (cor = 0.018, P = 7.33e-01) and MYBL2 (cor = 0.098, P = 6.97e-02) had no significant correlation with tumor purity (Figure 14). Next, the relationship between six hub genes’ expression and gene markers of tumor-associated macrophages (TAMs) (CD68, IL10), monocytes (CD86, CSF1R), M1 (IRF5, PTSG2) and M2 (CD163, VSIG4, MS4A4A) macrophages was investigated via the correlation module in TIMMER. All included gene markers have been reported in previous studies. Results revealed that expression levels of hub genes were significantly associated with most immune markers, except for MYBL2 and RACGAP1, which had no distinct correlation with gene markers PTGS2, CD163, VSIG4, or MS4A4A (Table 2). These findings strongly suggest that hub genes are correlated with immune infiltration in HCC. Additionally, whether immune infiltrating levels of each immune subset are related to differential copy number of hub genes was also analyzed. No significant relationship was observed between most immune cells and hub genes (Figure S5).

**Identification of bioactive compounds targeting hub genes.**

Finally, we predicted potential bioactive compounds targeting hub genes using the Drugbank database, which is a comprehensive, freely accessible, online database including both drugs and drug target information. A total of 15 compounds targeting hub genes were identified, including CEP55 (Irdabions, CEP-9722, CEP-1347, CEP-37440, Cefapirin), CLSPN (Calfactant, Calusterone) and MYBL2 (Clotrimazole, Propafenone, Letrozole, Sildenafil, Ranitidine, Valproic acid, Esomeprazole, Pregabalin) (Table 3). Except for Calfactant, the 3D chemical structure of the other compounds is presented in Figure S6. These results could provide new insight into potential novel therapeutic targets for HCC in the future.
Discussion

HCC is one of the most common malignant tumors in the world. In recent years, the prevalence of HCC is gradually increasing, especially in nontraditional high incidence areas, such as the United States and Europe. Most HCC patients are likely to be diagnosed at an advanced stage since HCC is asymptomatic at early stages, and effective biomarkers for early diagnosis and prognostic prediction are lacking [25-27]. Currently, the main treatments for HCC include radiofrequency ablation, surgical resection, immunotherapy, and liver implantation, among others. However, the clinical efficacy of treatments for specific patients is not satisfactory due to poor therapeutic targets, tumor immune escape and complications, leading to a low 5-year survival rate [7, 14, 28]. In past decades, studies focused on immunotherapy for various cancers have obtained meaningful breakthroughs, especially in melanoma, non-small cell lung cancer, and others. Different types of immune cells infiltrate into the tumor microenvironment, which is a crucial reason for effective immune responses [12, 13]. It has been reported that immune cells, including natural killer cells, CD4+ and CD8+ T-cells, TAMs, and dendritic cells, and others, are detected in cancer tissues, including HCC. Furthermore, these immune infiltrating cells are regulated by various mediators, such as chemokines. However, the mechanism of immune infiltration in the development of tumors is not completely understood [23, 24]. Therefore, it is urgent and significant to elucidate the molecular mechanism and to identify immune-related signatures in HCC, which will aid in identifying new therapeutic targets and prognostic markers to increase the clinical efficacy and 5-year survival rate of HCC patients.

Currently, the ceRNA hypothesis of crosstalk between ncRNAs and mRNAs has received much attention and is considered a new measure of gene regulation at the posttranscriptional level, which provides new insight into revealing mechanisms of tumorigenesis and identifying potential diagnostic and prognostic biomarkers. A growing number of published studies have demonstrated that many predictive signatures are detected in various tumors based on ceRNA network analysis [29-33]. MiRNAs, included in ncRNAs, are identified to be evolutionarily conserved, with an average length of 22-nt, and may bind to the 3’untranslated region (3’UTR) of the targeted mRNAs according to the principle of complementary base pairing. An increasing body of evidence has demonstrated that dysregulated miRNAs play a crucial role in the initiation, progression, and therapy of various tumors [4, 28, 34]. Baolei et al revealed that miRNA-124 is a negative regulator of HCC with respect to proliferation and invasion by downregulating lncRNA-UCA1[35]. Baltruskeviciene E et al found that downregulated expression of miRNA-148a and miRNA-625-3p is related to tumor budding in colorectal cancer, and EMT was considered a possible molecular mechanism [36].

In the present study, a prognostic predictive model was established based on 23
DEMs and exhibited great performance with the area under ROC, which was 0.804 for 3-year and 0.744 for 5-year survival. Moreover, we constructed a ceRNA network, including 56 DELs, 6 DEMs (hsa-mir-9-1, hsa-mir-9-2, hsa-mir-30d, hsa-mir-139, hsa-mir-195, hsa-mir-301a), and 28 DEGs, which identified several potential prognostic signatures for HCC. Functional annotation revealed that DEGs were related to mitotic spindle assembly, G1/S-specific transcription, and TNF signaling pathway, among others. U Lehmann et al reported that hypermethylation of hsa-mir-9-1 is related to the development of breast cancer. Patients with pre-invasive intraductal lesions were detected by hypermethylated hsa-mir-9-1 [37]. Notably, hsa-mir-195 belongs to the miR-195 family and is located on chromosome 17p13.1, and is correlated with proliferation and angiogenesis in prostate tumors by downregulating expression of the PRR11 gene [38]. In M.K. Sannigrahi et al’s research, they indicated that hsa-mir-139 was downregulated by HPV-16, leading to activation of HPV-16 oncogenic pathways and carcinogenesis of HPV-16 induced cervical and head and neck cancers [39]. Hsa-mir-301a was found to play an oncogenic role in the occurrence and development of laryngeal squamous cell carcinoma (LSCC) by directly targeting the tumor suppressor gene Smad44 and downregulating its expression. Furthermore, it participates in the process of epithelial-mesenchymal transition (EMT) [40]. Through analysis of SNP-array data generated from 8 medulloblastoma cell lines, Yuan Lu et al found that hsa-mir-30d is overexpressed, however, the potentially involved biological processes and molecular mechanisms are still not clarified [41].

In this study, six hub genes (including CEP55, DEPDC1, CLSPN, KIF23, MYBL2, and RACGAP1) were identified after PPI network analysis, all of which are upregulated in HCC samples. Except for RACGAP1, the hub genes were reported to be involved in various tumors. Furthermore, CEP55 promotes proliferation and invasion in osteosarcoma by regulating the AKT signaling pathway. In another study, data showed that CEPP55 facilitates the EMT process in renal cells and participates in the AKT pathway [42, 43]. DEPDC1 has been identified in various cancers, including HCC, breast cancer, and prostate cancer, among others and is positively involved with multiple tumorous biological processes, including proliferation, invasion, and angiogenesis by activating different signaling pathways, such as CCL20/CCR6, E2F et al [44-46]. According to previous reports, CLSPN is overexpressed in renal cell carcinoma (RCC) as assessed by immunohistochemistry in 95 RCC cases. They also found that patients with higher T grade, advanced tumor stage, vein invasion and inferior prognosis were likely to exhibit stronger CLSPN staining. In addition, CLSPN activates the AKT signaling pathway and is co-expressed with several known tumor-related genes, such as programmed death ligand-1, epidermal growth factor receptor et al [47]. Previous studies have reported that KIF23 is a kinesin-like motor protein that plays a significant role in cytokinesis and has two splice variants, KIFV1 and KIFV2, which are overexpressed in HCC samples but were not detected in nontumor tissues.
HCC patients with positive identified KIF V1 had a better overall 5-year survival rate than those with no KIF V1. However, there was no significant correlation between expression of KIF V2 and overall survival in patients [48, 49]. These discoveries by Xiaotong et al conflicted with results in this study, indicating that additional investigations should be performed to further elucidate the role of KIF23 in the tumorigenesis of HCC [48]. MYBL2 (MYB proto-oncogene like 2) is included in the family of MYB transcription factors and was overexpressed in breast cancer. Jianlin et al revealed that overexpressed MYBL2 in breast cancer promotes growth and metastasis, which was inhibited by miR-143-3p [50].

Additionally, in our study, we found that all six hub genes were positively correlated with immune cell infiltration of B cells, CD4+ T cells, CD8+ T cells, macrophages, neutrophils and dendritic cells in HCC and are associated with most of the genetic markers of monocytes, TAMs, M1 and M2 macrophages. However, until now, correlation of these six hub genes and immune infiltration in tumors has not been reported. Through an integrated bioinformatics analysis based on a ceRNA network, we identified several new biomarkers related to immune infiltration in HCC, which may represent potential prognostic and therapeutic targets. However, furthermore in-depth experimental studies are required to verify the function and elaborate on the underlying mechanisms in HCC. Finally, based on analysis of the Drugbank database, we found a total of 15 bioactive compounds targeting CEP55, CLSPN, and MYBL2, providing some new directions for drug development for HCC treatments in the future.

In conclusion, a simple-to-use nomogram predictive model was established based on miRNAs revealed great performance. We also constructed a ceRNA regulatory network to better understand the interactions between mRNAs and ncRNAs in HCC. Moreover, six hub genes were identified through PPI network analysis, all of which are overexpressed in HCC and are associated with survival. In addition, expression of hub genes was closely correlated with immune infiltration in HCC. We believe these genes may be involved in the development of HCC and may represent potential prognostic biomarkers and individual therapeutic targets. However, further biological and molecular experiments are required.

**Abbreviations**

lncRNAs: long non-coding RNAs; ceRNAs: competing endogenous RNAs; HCC: hepatocellular carcinoma; RFS: recurrence-free survival; OS: overall survival; TCGA: the Cancer Genome Atlas; DElncRNAs: differentially expressed lncRNA; DEMiRNAs: differentially expressed miRNA; DEMiRNAs: differentially expressed mRNA; MREs: miRNA-response elements; GO: gene ontology; ROC: receiver operating characteristic; BP: biological process; CC: cellular component; MF: molecular function; EMT: Epithelial-mesenchymal transformation; RS: risk score.
Authors’ contributions
Fang Peng and Yiya Zhang designed the experiments and revised the manuscript. Zhangya Pu analyzed the data and wrote their manuscript. Xiangfang Wang, Yun Zhong and Yuangyuan Zhu searched the data and do some help to analyze the data. All authors read and consent the final manuscript.

Author details
1 Department of Infectious Diseases, Hunan Key Laboratory of Viral Hepatitis, Xiangya Hospital, Central South University, Changsha, Hunan Province, 410008, China.
2 NHC Key Laboratory of Cancer Proteomics, Xiangya Hospital, Central South University, 87 Xiangya Road, Changsha, Hunan Province, 410008, China.
3 National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, Hunan, China
4 Department of Dermatology, Xiangya Hospital, Changsha, Hunan Province, 410008, China.

Acknowledgements
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Availability of data and materials
The authors declare that the data supporting the findings in the present study could be found within the article.

consent for publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

Funding
None
References

1. Li S, Huang Y, Huang Y, Fu Y, Tang D, Kang R, Zhou R, Fan XG: The long non-coding RNA TP73-AS1 modulates HCC cell proliferation through miR-200a-dependent HMGB1/RAGE regulation. J Exp Clin Cancer Res 2017, 36(1):51.

2. Shi L, Peng F, Tao Y, Fan X, Li N: Roles of long noncoding RNAs in hepatocellular carcinoma. Virus Res 2016, 223:131-139.

3. Gu X, Li H, Sha L, Zhao W: Construction and Comprehensive Analyses of a Competing Endogenous RNA Network in Tumor-Node-Metastasis Stage I Hepatocellular Carcinoma. Biomed Res Int 2020, 2020:5831064.

4. Long J, Bai Y, Yang X, Lin J, Yang X, Wang D, He L, Zheng Y, Zhao H: Construction and comprehensive analysis of a ceRNA network to reveal potential prognostic biomarkers for hepatocellular carcinoma. Cancer Cell Int 2019, 19:90.

5. Fu S, Wang J, Hu X, Zhou RR, Fu Y, Tang D, Kang R, Huang Y, Sun L, Li N et al: Crosstalk between hepatitis B virus X and high-mobility group box 1 facilitates autophagy in hepatocytes. Mol Oncol 2018, 12(3):322-338.

6. Li Y, Fu Y, Hu X, Sun L, Tang D, Li N, Peng F, Fan XG: The HBx-CTTN interaction promotes cell proliferation and migration of hepatocellular carcinoma via CREB1. Cell Death Dis 2019, 10(6):405.

7. Yue C, Ren Y, Ge H, Liang C, Xu Y, Li G, Wu J: Comprehensive analysis of potential prognostic genes for the construction of a competing endogenous RNA regulatory network in hepatocellular carcinoma. Onco Targets Ther 2019, 12:561-576.

8. Zhang R, Jiang YY, Xiao K, Huang XQ, Wang J, Chen SY: Candidate lncRNA-miRNA-mRNA network in predicting hepatocarcinogenesis with cirrhosis: an integrated bioinformatics analysis. J Cancer Res Clin Oncol 2020, 146(1):87-96.

9. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP: A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell 2011, 146(3):353-358.

10. Zhang K, Li Q, Kang X, Wang Y, Wang S: Identification and functional characterization of IncRNAs acting as ceRNA involved in the malignant progression of glioblastoma multiforme. Oncol Rep 2016, 36(5):2911-2925.

11. Wang JJ, Huang YQ, Song W, Li YF, Wang H, Wang WJ, Huang M: Comprehensive analysis of the IncRNA-associated competing endogenous RNA network in breast cancer. Oncol Rep 2019, 42(6):2572-2582.

12. Chen J, Wang Z, Wang W, Ren S, Xue J, Zhong L, Jiang T, Wei H, Zhang C: SYT16 is a prognostic biomarker and correlated with immune infiltrates in glioma: A study based on TCGA data. Int Immunopharmacol 2020, 84:106490.

13. Pan JH, Zhou H, Cooper L, Huang JL, Zhu SB, Zhao XX, Ding H, Pan YL, Rong L: LAYN Is a Prognostic Biomarker and Correlated With Immune Infiltrates in Gastric and Colon Cancers. Front Immunol 2019, 10:6.

14. Zhang Y, Zhang L, Xu Y, Wu X, Zhou Y, Mo J: Immune-related long noncoding RNA signature for predicting survival and immune checkpoint blockade in hepatocellular carcinoma. J Cell Physiol 2020.

15. Chen Y, Ramjiwan RR, Reiberger T, Ng MR, Hato T, Huang Y, Ochiiai H, Kitahara S, Unan EC, Reddy TP et al: CXCR4 inhibition in tumor microenvironment facilitates
anti-programmed death receptor-1 immunotherapy in sorafenib-treated hepatocellular carcinoma in mice. Hepatology 2015, 61(5):1591-1602.

16. Jeggari A, Marks DS, Larsson E: miRcode: a map of putative microRNA target sites in the long non-coding transcriptome. Bioinformatics 2012, 28(15):2062-2063.

17. Agarwal V, Bell GW, Nam JW, Bartel DP: Predicting effective microRNA target sites in mammalian mRNAs. Elife 2015, 4.

18. Chou CH, Shrestha S, Yang CD, Chang NW, Lin YL, Liao KW, Huang WC, Sun TH, Tu SJ, Lee WH et al: miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. Nucleic Acids Res 2018, 46(D1):D296-D302.

19. Wong N, Wang X: miRDB: an online resource for microRNA target prediction and functional annotations. Nucleic Acids Res 2015, 43(Database issue):D146-152.

20. Aran D, Sirota M, Butte AJ: Systematic pan-cancer analysis of tumour purity. Nat Commun 2015, 6:8971.

21. Li B, Severson E, Pignon JC, Zhao H, Li T, Novak J, Jiang P, Shen H, Aster JC, Rodig S et al: Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. Genome Biol 2016, 17(1):174.

22. Danaher P, Warren S, Dennis L, D'Amico L, White A, Disis ML, Geller MA, Odunsi K, Beechem J, Fling SP: Gene expression markers of Tumor Infiltrating Leukocytes. J Immunother Cancer 2017, 5:18.

23. Siemers NO, Holloway JL, Chang H, Chasalow SD, Ross-MacDonald PB, Voliva CF, Szustakowski JD: Genome-wide association analysis identifies genetic correlates of immune infiltrates in solid tumors. PLoS One 2017, 12(7):e0179726.

24. Sousa S, Maatta J: The role of tumour-associated macrophages in bone metastasis. J Bone Oncol 2016, 5(3):135-138.

25. Lei B, Zhou J, Xuan X, Tian Z, Zhang M, Gao W, Lin Y, Ni B, Pang H, Fan W: Circular RNA expression profiles of peripheral blood mononuclear cells in hepatocellular carcinoma patients by sequence analysis. Cancer Med 2019, 8(4):1423-1433.

26. Li Y, Ma B, Yin Z, Liu P, Liu J, Li J, Wang F, Chen H: Competing endogenous RNA network and prognostic nomograms for hepatocellular carcinoma patients who underwent R0 resection. J Cell Physiol 2019, 234(11):20342-20353.

27. Liu J, Li W, Zhang J, Ma Z, Wu X, Tang L: Identification of key genes and long non-coding RNA associated ceRNA networks in hepatocellular carcinoma. PeerJ 2019, 7:e8021.

28. Zhang Z, Tang D, Wang B, Wang Z, Liu M: Analysis of miRNA-mRNA regulatory network revealed key genes induced by aflatoxin B1 exposure in primary human hepatocytes. Mol Genet Genomic Med 2019, 7(11):e971.

29. Wang M, Mao C, Ouyang L, Liu Y, Lai W, Liu N, Shi Y, Chen L, Xiao D, Yu F et al: Long noncoding RNA LINC00336 inhibits ferroptosis in lung cancer by functioning as a competing endogenous RNA. Cell Death Differ 2019, 26(11):2329-2343.

30. Wu ZH, Cai F, Zhong Y: Comprehensive Analysis of the Expression and Prognosis for GBPs in Head and neck squamous cell carcinoma. Sci Rep 2020, 10(1):6085.

31. Xiao Y: Construction of a circRNA-miRNA-mRNA network to explore the pathogenesis and treatment of pancreatic ductal adenocarcinoma. J Cell Biochem 2020,
Xiong DD, Dang YW, Lin P, Wen DY, He RQ, Luo DZ, Feng ZB, Chen G: A circRNA-miRNA-mRNA network identification for exploring underlying pathogenesis and therapy strategy of hepatocellular carcinoma. *J Transl Med* 2018, 16(1):220.

Xu F, Zhao Y, Qin G, Huan Y, Li L, Gao W: Comprehensive analysis of competing endogenous RNA networks associated with cholangiocarcinoma. *Exp Ther Med* 2019, 18(5):4103-4112.

Lou W, Liu J, Ding B, Chen D, Xu L, Ding J, Jiang D, Zhou L, Zheng S, Fan W: Identification of potential miRNA-mRNA regulatory network contributing to pathogenesis of HBV-related HCC. *J Transl Med* 2019, 17(1):7.

Zhao B, Lu Y, Cao X, Zhu W, Kong L, Ji H, Zhang F, Lin X, Guan Q, Ou K et al: MiRNA-124 inhibits the proliferation, migration and invasion of cancer cell in hepatocellular carcinoma by downregulating IncRNA-UCA1. *Onco Targets Ther* 2019, 12:4509-4516.

Baltruskeviciene E, Schveigert D, Stankevicius V, Mickys U, Zvirblis T, Bublevic J, Suziedelis K, Aleknavicius E: Down-regulation of miRNA-148a and miRNA-625-3p in colorectal cancer is associated with tumor budding. *BMC Cancer* 2017, 17(1):607.

Lehmann U, Hasemeier B, Christgen M, Muller M, Romermann D, Langer F, Kreipe H: Epigenetic inactivation of microRNA gene hsa-mir-9-1 in human breast cancer. *J Pathol* 2008, 214(1):17-24.

Cai C, He H, Duan X, Wu W, Mai Z, Zhang T, Fan J, Deng T, Zhong W, Liu Y et al: miR-195 inhibits cell proliferation and angiogenesis in human prostate cancer by downregulating PRR11 expression. *Oncol Rep* 2018, 39(4):1658-1670.

Sannigrahi MK, Sharma R, Singh V, Panda NK, Rattan V, Khullar M: Role of Host miRNA Hsa-miR-139-3p in HPV-16-Induced Carcinomas. *Clin Cancer Res* 2017, 23(14):3884-3895.

Lu Y, Gao W, Zhang C, Wen S, Huangfu H, Kang J, Wang B: Hsa-miR-301a-3p Acts as an Oncogene in Laryngeal Squamous Cell Carcinoma via Target Regulation of Smad4. *J Cancer* 2015, 6(12):1260-1275.

Lu Y, Ryan SL, Elliott DJ, Bignell GR, Futreal PA, Ellison DW, Bailey S, Clifford SC: Amplification and overexpression of Hsa-miR-30b, Hsa-miR-30d and KHDRBS3 at 8q24.22-q24.23 in medulloblastoma. *PLoS One* 2009, 4(7):e6159.

Chen H, Zhu D, Zheng Z, Cai Y, Chen Z, Xie W: CEP55 promotes epithelial-mesenchymal transition in renal cell carcinoma through PI3K/AKT/mTOR pathway. *Clin Transl Oncol* 2019, 21(7):939-949.

Xu L, Xia C, Sheng F, Sun Q, Xiong J, Wang S: CEP55 promotes the proliferation and invasion of tumor cells via the AKT signalling pathway in osteosarcoma. *Carcinogenesis* 2018, 39(4):623-631.

Guo W, Li H, Liu H, Ma X, Yang S, Wang Z: DEPDC1 drives hepatocellular carcinoma cell proliferation, invasion and angiogenesis by regulating the CCL20/CCR6 signaling pathway. *Oncol Rep* 2019, 42(3):1075-1089.

Huang L, Chen K, Cai ZP, Chen FC, Shen HY, Zhao WH, Yang SJ, Chen XB, Tang GX, Lin X: DEPDC1 promotes cell proliferation and tumor growth via activation of E2F signaling in prostate cancer. *Biochem Biophys Res Commun* 2017, 490(3):707-712.
46. Zhao H, Yu M, Sui L, Gong B, Zhou B, Chen J, Gong Z, Hao C: High Expression of DEPDC1 Promotes Malignant Phenotypes of Breast Cancer Cells and Predicts Poor Prognosis in Patients With Breast Cancer. *Front Oncol* 2019, 9:262.

47. Kobayashi G, Sentani K, Babasaki T, Sekino Y, Shigematsu Y, Hayashi T, Oue N, Teishima J, Matsubara A, Sasaki N et al: Claspin overexpression is associated with high-grade histology and poor prognosis in renal cell carcinoma. *Cancer Sci* 2020, 111(3):1020-1027.

48. Sun X, Jin Z, Song X, Wang J, Li Y, Qian X, zhang Y, Yin Y: Evaluation of KIF23 variant 1 expression and relevance as a novel prognostic factor in patients with hepatocellular carcinoma. *BMC Cancer* 2015, 15:961.

49. Wei B, Kong W, Mou X, Wang S: Comprehensive analysis of tumor immune infiltration associated with endogenous competitive RNA networks in lung adenocarcinoma. *Pathol Res Pract* 2019, 215(1):159-170.

50. Chen J, Chen X: MYBL2 Is Targeted by miR-143-3p and Regulates Breast Cancer Cell Proliferation and Apoptosis. *Oncol Res* 2018, 26(6):913-922.