Weighted Gene Coexpression Network Analysis Reveals the Dynamic Transcriptome Regulation and Prognostic Biomarkers of Hepatocellular Carcinoma

Evolutionary Bioinformatics
Volume 16: 1–10

DOI: 10.1177/1176934320920562
SAGE

Shuping Qu[1,*], Qiuyuan Shi[2,*], Jing Xu[3,*], Wanwan Yi[2] and Hengwei Fan[1]

[1]Department of Hepatic Surgery, The Eastern Hepatobiliary Surgery Hospital, Second Military Medical University, Shanghai, China. [2]Department of Nuclear Medicine, Shanghai Tenth People's Hospital, School of Medicine, Tongji University, Shanghai, China. [3]Department of Interventional Oncology, Shanghai Seventh People's Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China.

ABSTRACT: This study was aimed at revealing the dynamic regulation of mRNAs, long noncoding RNAs (lncRNAs), and microRNAs (miRNAs) in hepatocellular carcinoma (HCC) and to identify HCC biomarkers capable of predicting prognosis. Differentially expressed mRNAs (DEmRNAs), lncRNAs, and miRNAs were acquired by comparing expression profiles of HCC with normal samples, using an expression data set from The Cancer Genome Atlas. Altered biological functions and pathways in HCC were analyzed by subjecting DEmRNAs to Gene Ontology and Kyoto Encyclopedia of Genes and Genomes analysis. Gene modules significantly associated with disease status were identified by weighted gene coexpression network analysis. An lncRNA-mRNA and an miRNA-mRNA coexpression network were constructed for genes in disease-related modules, followed by the identification of prognostic biomarkers using Kaplan-Meier survival analysis. Differential expression and association with the prognosis of 4 miRNAs were verified in independent data sets. A total of 1220 differentially expressed genes were identified between HCC and normal samples. Differentially expressed mRNAs were significantly enriched in functions and pathways related to "plasma membrane structure," "sensory perception," "metabolism," and "cell proliferation." Two disease-associated gene modules were identified. Among genes in lncRNA-mRNA and miRNA-mRNA coexpression networks, 9 DEmRNAs and 7 DEmiRNAs were identified to be potential prognostic biomarkers. MIMAT0000102, MIMAT0003882, and MIMAT0004677 were successfully validated in independent data sets. Our results may advance our understanding of molecular mechanisms underlying HCC. The biomarkers may contribute to diagnosis in future clinical practice.

KEYWORDS: Hepatocellular carcinoma, differential expression analysis, biomarker, prognosis, WGCNA

RECEIVED: January 06, 2020. **ACCEPTED:** March 30, 2020.

TYPE: Original Research

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by special funds for basic scientific research business expenses of central colleges and universities (grant no. 22120180392), Program of Shanghai Academic/Technology Research Leader (grant no. 18XD1403000), Shanghai 2018 "Science and Technology Innovation Action Plan" Science and Technology Support Project in Biomedicine (grant no. 18441903500) and Leading Talents in Shanghai (grant no. 03.05.19005).

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHORS: Wanwan Yi, Department of Nuclear Medicine, Shanghai Tenth People's Hospital, School of Medicine, Tongji University, No. 301, Yanchang Middle Road, Shanghai 200072, China. Email: 18321775325@163.com

Hengwei Fan, Department of Hepatic Surgery, The Eastern Hepatobiliary Surgery Hospital, Second Military Medical University, No. 225, Changhai Road, Shanghai 200438, China. Email: fanhengwei2006@163.com

Background

Hepatocellular carcinoma (HCC) is one of the most frequent and malignant types of liver cancer.[1,2] Approximately 780 000 new cases of HCC are reported each year.[2,3] Although the survival of HCC has been improved as a result of advances in treatment modalities, the prognosis remains unfavorable, with an estimated 5-year survival rate of only 12%.[4] Although image technologies such as magnetic resonance imaging and contrast-enhanced computed tomography can provide useful information for diagnosis and treatment, the performance of imaging technologies remains unsatisfactory in staging and grading HCC.[5] Therefore, suitable molecular predictors are extremely needed for HCC therapy and prognosis prediction.

Cancers generally preceed along with widespread expression alterations of both protein-coding mRNAs and noncoding RNAs (ncRNAs).[6,7] Recent advancements in high-throughput sequencing have allowed for the identification of various HCC biomarkers. For example, it has been reported that integrator complex subunit 6 (*INTS6*) is significantly downregulated and carnitine deficiency-associated gene expressed in ventricle 3 (*CDV3*) is overexpressed in HCC tissues, both of which are indications of unfavorable survival.[8,9]

MicroRNAs (miRNAs) and long noncoding RNAs (lncRNAs) are 2 major types of ncRNAs.[7,10] MicroRNAs are short ncRNAs with ~20 nucleotides in length, whereas lncRNAs are ncRNAs with more than 200 nucleotides in length.[10] Both types of ncRNAs participate in a broad range of biological processes (BPs), such as cell proliferation, differentiation, and apoptosis.[10] They primarily function through regulating gene expression by binding mRNAs at posttranscriptional level.[11] A recently proposed hypothesis indicates that mRNAs and lncRNAs competing for shared miRNAs can serve as competing endogenous RNAs for each other.[11] Accumulating evidence has underlined the key roles of ncRNAs in tumorigenesis and their potential as biomarkers for diagnosis and prognosis prediction. For example, the miRNA hsa-miR-630 plays a tumor-suppressing role in HCC, and low expression level of hsa-miR-630 is associated with poor survival.[12] In addition, the lncRNA *SNHG20* is reported to be overexpressed in HCC tissues and high expression level of *SNHG20* predicts unfavorable survival.[13,14]

To reveal the dynamic regulation of mRNAs, miRNAs, and lncRNAs in HCC and identify prognostic markers, we first

* S.Q., Q.S., and J.X. contributed equally to this work.

comprehensively analyzed the expression data of HCC and identified differentially expressed mRNAs (DEmRNAs), lncRNAs (DElncRNAs), and miRNAs (DEmiRNAs) between HCC and normal samples. Weighted gene coexpression network analysis (WGCNA) was conducted, and an miRNA-mRNA network and an lncRNA-mRNA regulation network were subsequently constructed to delineate the dynamic regulation of transcriptome in HCC. Finally, a combination of topological analysis and survival analysis were further implemented to identify prognostic factors of HCC.

Materials and Methods

Data source and preprocessing

HCC expression data from The Cancer Genome Atlas was downloaded from UCSC Xena (https://xenabrowser.net/datapages/).[15] Gene expression data (IlluminaHiSeq pancan normalized) of 421 samples (371 tumor and 50 normal samples) and miRNA expression data (miRNA mature strand expression RNAseq) of 420 samples (371 tumor and 49 normal samples) were included in the data set. Samples with both gene expression data and miRNA expression data were selected for further study. In total, 414 samples (365 tumor and 49 normal samples) matched the criterion and were used in our study.

Gene expression data were annotated according to the gene transfer format annotation file (Release 26, GRCh38.p10) provided by GENCODE.[16] Genes annotated with "protein-coding" were considered as mRNAs, while those annotated with "lincRNA," "antisense," "sense intronic," "processed transcript," "sense overlapping" "3prime overlapping ncRNA" and "noncoding" were considered as lncRNAs.

In addition, the clinical information of the liver hepatocellular carcinoma (LIHC) samples involved in our study was also downloaded in March 2019. The clinical information included age, sex, histological grade of the tumor, tumor-node-metastasis (TNM) stage, overall survival (OS), and OS status.

Differential expression analysis

The expression data matrix of 365 tumors and 49 normal samples were normalized by betaqn method of *R*. Based on the normalized data set, differential expression analysis between the tumor and normal samples was performed using limma package (version 3.10.3, http://bioconductor.org/packages/release/bioc/html/limma.html)[17] of *R* and the significance was tested by unpaired *t* test. The resulting *P* value of each gene was further adjusted by Benjamini and Hochberg (BH) method to achieve the corresponding adj. *P* value. The selection criteria of DEmRNAs, DEmiRNAs, and DElncRNAs were $|\log \text{fold change (FC)}| > 1$ and adj. *P* value $< .05$.

Functional annotation of DEmRNAs

Gene ontology (GO) terms enriched by DEmRNAs were analyzed by The Database for Annotation, Visualization and Integrated Discovery (DAVID) (version 6.7, https://david-d.ncifcrf.gov/).[18] The resulting terms included BPs, cellular components (CCs) and molecular functions (MFs). Kyoto Encyclopedia of Genes and Genomes (KEGG)[19] pathways enriched by the DEmRNAs were analyzed by Gene Set Enrichment Analysis (GSEA, version 3.0, http://software.broadinstitute.org/gsea/index.jsp).[20] The selection criterion of GO and KEGG terms was set as adj. *P* value $< .05$.

Weighted gene coexpression network analysis

As a method of systems biology, WGCNA describes the gene correlation patterns across different samples.[21] It could also be used to identify potential biomarkers based on the correlations among gene modules and correlations between modules and phenotype.[22] The WGCNA package (version 1.61, https://cran.r-project.org/web/packages/WGCNA/)[21] of *R* was used to construct a weighted coexpression network, identify genes and gene modules, and calculate topological features. Specifically, the expression values of DEmRNAs, DEmiRNAs, and DElncRNAs were extracted from the expression data set and the clinical information was integrated. The data matrix was normalized using betaqn method of *R* and was used as input. Then the input data matrix was preprocessed by removing genes with median absolute deviation (MAD) beyond 75th percentile and MAD less than .01. In addition, missing values were removed from the data. Pearson coefficient was first calculated for each pair of genes, followed by defining adjacency functions and subsequently clustering genes into different modules. The minimum module size was set as 50. In addition, the correlations between phenotype and gene modules were also calculated and phenotype-related modules were thus obtained.

LncRNA/miRNA-mRNA coexpression network

To construct an lncRNA-mRNA network and an miRNA-mRNA coexpression network for genes in disease-related modules, the Pearson correlation coefficient of each pair of mRNA and lncRNA/miRNA was calculated and adjusted using the corr. test method from the psych package in *R* (version 1.8.12, https://cran.r-project.org/web/packages/psych/index.html).[23] The parameters were *ci* = *F* and adjust = "BH". The selection criteria for lncRNA-mRNA and miRNA-mRNA pairs were set as adj. *P* value $< .05$ and $r \geqslant .8$ or $\leqslant -0.5$. Based on the resulting lncRNA-mRNA and miRNA-mRNA pairs, coexpression networks were constructed using Cytoscape (version 3.7.0, https://cytoscape.org/release_notes_3_7_0.html).[24]

Survival analysis

The prognosis-associated information including OS and OS status was collected. Genes with high connectivity in coexpression networks were used as candidate genes for survival analysis using Kaplan-Meier method.[25] Specifically, the correlation between each gene and prognosis was analyzed by stratifying

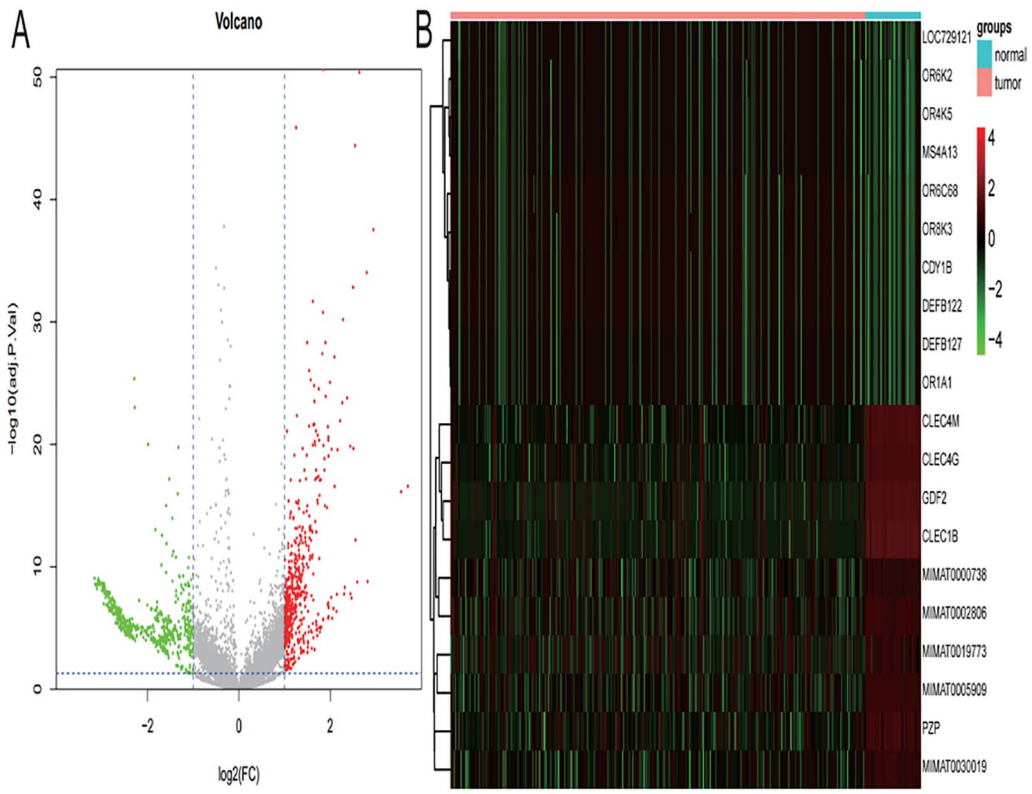

Figure 1. Screening and hierarchical clustering analysis of differentially expressed genes: (A) Volcano plot of differentially expressed genes between tumor and normal samples. The horizontal red dash line indicated where adj. *P* value = .05 and the vertical red dash lines indicated where|logFC| = 1.0. Red and green dots indicated significantly upregulated and downregulated genes, respectively. (B) Two-way hierarchical clustering analysis of top 10 upregulated and downregulated genes. Tumor and normal samples were shown as red and cyan along the horizontal bar and genes were shown along the vertical bar. Upregulated genes were shown as red and downregulated genes were shown as green.

samples into high- and low-expression groups, based on the median expression value of the corresponding gene. The association between genes and prognosis were then analyzed by Kaplan-Meier survival plotting. Meanwhile, *P* value was calculated using log-rank test and genes with $P < 0.01$ were considered as potential prognosis-associated genes.

Validation of prognosis-associated miRNAs in independent data set

The prognosis-associated miRNAs were further validated in independent data sets. The expression data set of HCC was searched in the Gene Expression Omnibus database. The differential expression of miRNAs was validated in HCC and normal tissues and the association of miRNAs with prognosis was validated in HCC and matched adjacent tissues. Finally, GSE36915, including miRNA expression data of 68 HCC and 21 nontumor liver tissues were selected for validation of differential expression. The data set was deposited by Shih et al[26] and was based on the platform of GPL8179 (Illumina Human v2 MicroRNA expression beadchip). GSE31384, including 166 pairs of HCC and matched noncancerous liver tissues, were selected for prognosis validation. This data set was deposited by Wei et al[27] and was based on the platform of GPL14140 (CapitalBio custom Human microRNA array).

The expression data of miRNAs in GSE36915 were downloaded and normalized by betaqn method in *R*. Then, unpaired *t* test was used to compare the difference of miRNAs between tumor group and normal group. $P < .05$ was regarded as significant level.

The expression data of miRNAs in GSE31384 were downloaded. Samples were divided into high-expression group and low-expression group based on the median expression value of each miRNA. Kaplan-Meier survival plotting and log-rank test were conducted. $P < .05$ was regarded as significant level.

Results

DEmRNAs, DEmiRNAs, and DElncRNAs

The integrated data set contained expression data of 19 187 mRNAs, 713 lncRNAs, and 2172 miRNAs. Differential expression analysis showed that a total of 541 genes (397 DEmRNAs, 37 DElncRNAs and 107 DEmiRNAs) were upregulated and 679 genes (395 DEmRNAs, 241 DElncRNAs, and 43 DEmiRNAs) were downregulated (Figure 1A; Table 1). Two-way hierarchical clustering was then performed based on the expression levels of the top 10 upregulated and downregulated genes. As a result, the tumor and normal samples could be completely separated (Figure 1B).

Functional annotation of DEmRNAs

To reveal the functions and pathways involved in HCC development and progression, functional annotation was performed based on the DEmRNAs identified above. GO analysis using DAVID showed that DEmRNAs were significantly enriched in terms related to plasma membrane structure and sensory perception, such as "GO:00016021~integral to membrane," "GO:0007186~G-protein-coupled receptor protein signaling pathway," "GO:0007608~sensery perception of smell" and "GO:0007606~sensory perception of chemical stimulus" (Figure 2). Meanwhile, KEGG analysis using GSEA showed that DEmRNAs were enriched in 93 terms, including 90 upregulated pathways, normalized enrichment score (NES) >0, and 3 downregulated pathways (NES < 0). These pathways were closely related to metabolism and cell proliferation, such as "valine leucine and isoleucine degradation," "tryptophan metabolism," "fatty acid degradation," "DNA replication," and "p53 signaling pathway" (Figure 3).

Table 1. Statistical data of genes.

	MRNA	LNCRNA	MIRNA	TOTAL
Total number	19817	713	2172	22702
Upregulated	397	37	107	541
Downregulated	395	241	43	679

Abbreviations: lncRNA, long noncoding RNA; miRNA, microRNA.

Weighted gene coexpression network analysis

The expression data of the DEmRNAs, DEmiRNAs, and DElncRNAs and the clinical information of samples were extracted and used as input for WGCNA. A total of 848 genes were identified by WGCNA and were clustered into 3 modules (Figure 4A). The modules were designated as module (ME)turquoise, MEgrey, and MEblue module, each contained 334, 259, and 255 genes, respectively. Overall, genes belong to the same module showed strong correlations (Figure 4B). Besides, genes in MEturquoise showed strong correlations with genes in MEblue, whereas genes in MEgray showed weaker correlations with genes in other modules (Figure 4B). In addition, genes in MEturquoise and MEblue module showed high intraconnectivity and interconnectivity degrees, whereas genes in MEgray module located at the margin of the network and showed low connectivity (Figure 4C). Module-phenotype correlation analysis showed that there was almost no significant correlation between each module and sex or vital

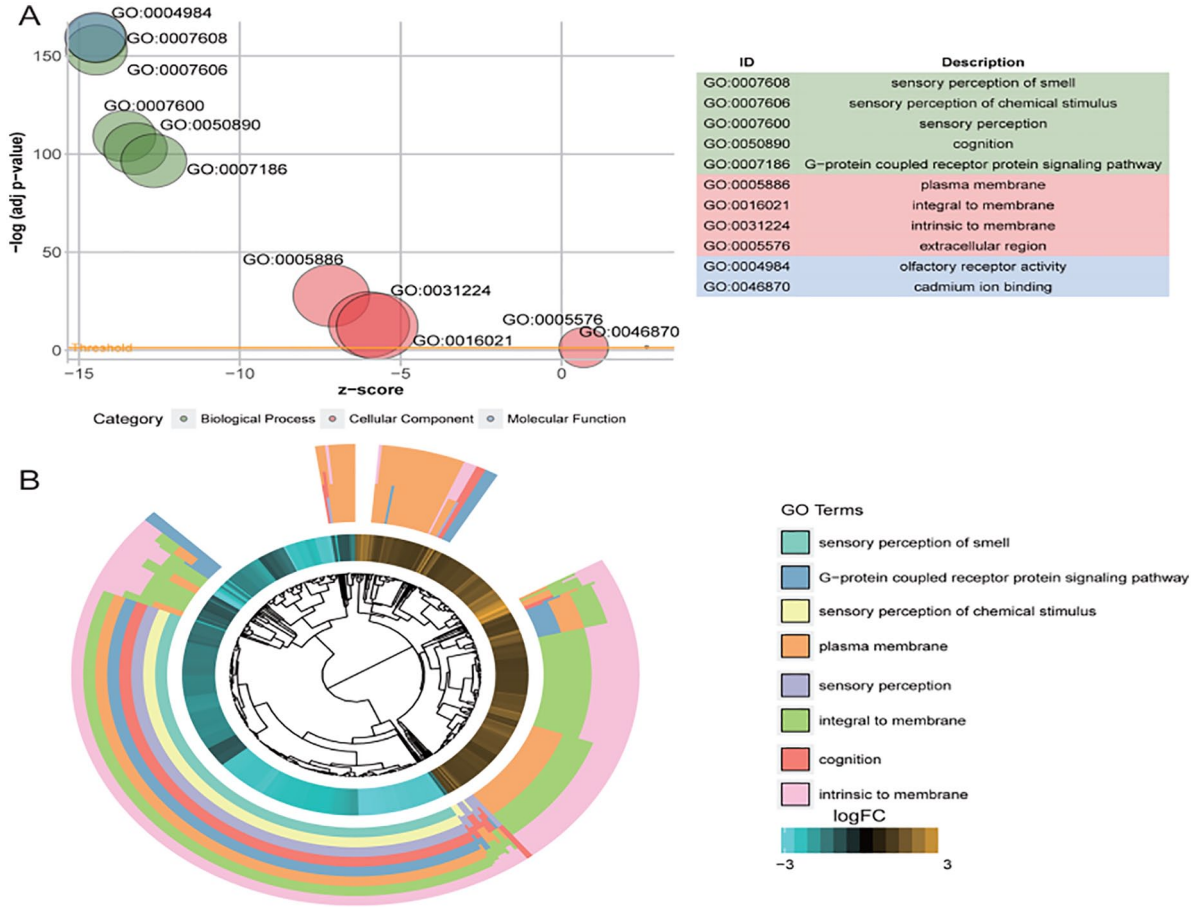

Figure 2. Gene ontology (GO) analysis: (A) The most significant GO biological processes, cellular components and molecular functions enriched by differentially expressed mRNAs. (B) Phylogenetic tree constructed using genes enriched in GO terms in (A). GO indicates gene ontology.

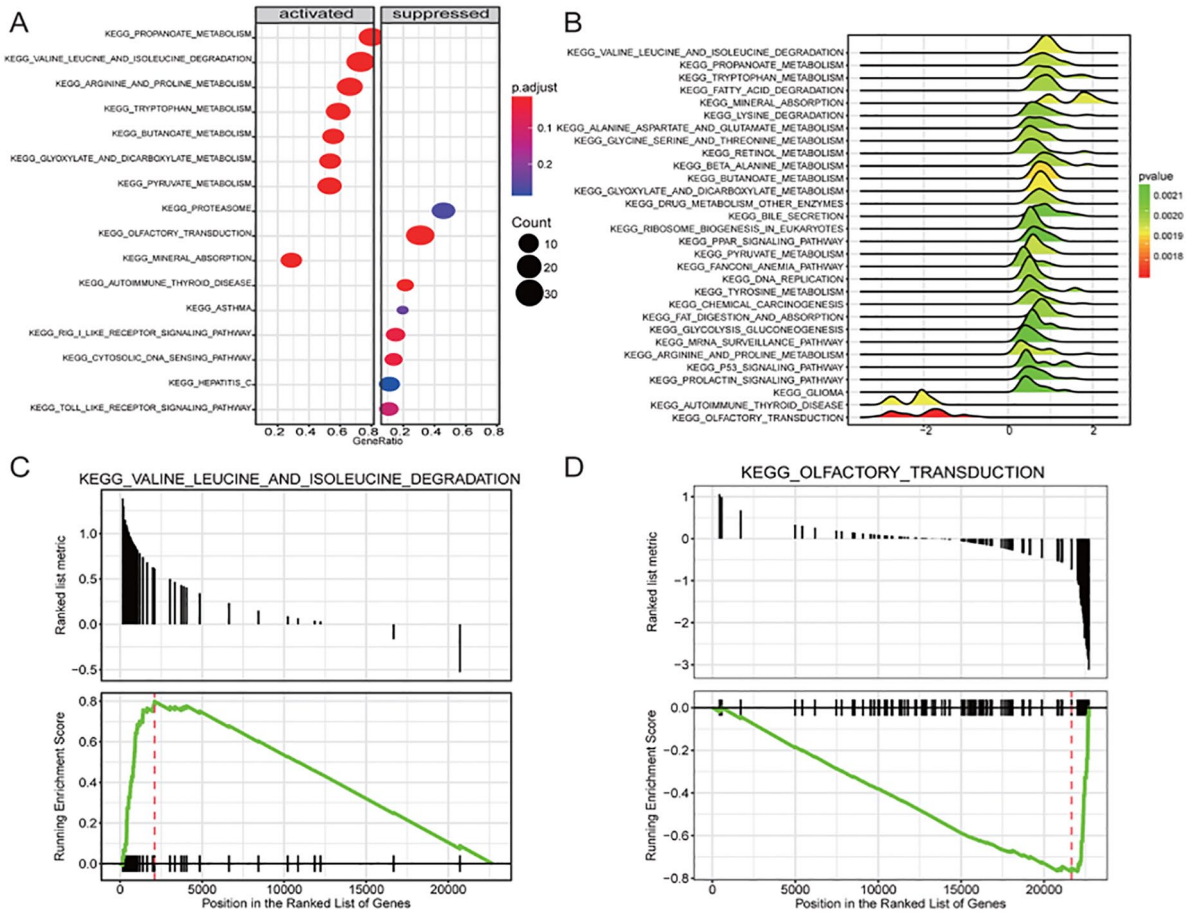

Figure 3. Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis: (A) Dotplot of the most significant KEGG pathways. The number of genes enriched by KEGG terms was indicated by circle size. The significance was indicated by color from blue (low) to red (high). (B) Joyplot of the most significant KEGG pathways. The significance was indicated by color from green (low) to red (high). (C and D) Enrichment plot of "aline leucine and isoleucine degradation" (C) and "olfactory transduction" (D). KEGG indicates Kyoto Encyclopedia of Genes and Genomes.

status (Figure 4D). In contrast, MEturquoise and MEblue modules were correlated with disease status with statistical significance (Figure 4D).

lncRNA/miRNA-mRNA coexpression network

According to WGCNA results, genes in MEturquoise and MEblue modules were selected for further analysis. The correlation coefficient of each pair of lncRNA-mRNA or miRNA-mRNA and the corresponding statistical significance were calculated. Long ncRNA-mRNA and miRNA-mRNA pairs with significant correlations (adj. P value $<.05$ and $r \geqslant .8$ or $\leqslant -0.5$) were selected to construct an lncRNA-mRNA and an miRNA-mRNA coexpression networks. The lncRNA-mRNA coexpression network contained 193 nodes (165 mRNAs and 28 lncRNAs) and 2208 edges (Figure 5A). Nodes with high connectivity included small nucleolar RNA C/D Box 9 (*SNORD9*), corticotropin-releasing hormone binding protein (*CRHBP*), C-type lectin domain family 1 member B (*CLEC1B*), growth differentiation factor 2 (*GDF2*) and olfactory receptor family 5 subfamily L member 2 (*OR5L2*). The network consisted of 1 major subnetwork and 3 minor subnetworks. The major subnetwork included 25 lncRNAs. A subnetwork contained these lncRNAs and coexpressed mRNAs were extracted and shown in Figure 5B.

The miRNA-mRNA coexpression network contained 184 nodes (168 mRNAs and 16 miRNAs) and 1374 edges (Figure 5C). MiRNAs in the major subnetwork and coexpressed mRNAs were extracted to construct a new subnetwork (Figure 5D). Nodes with high connectivity degree included MIMAT0002806 (hsa-miR-490-3p), MIMAT0030021 (hsa-miR-7706), MIMAT0019880 (hsa-miR-4746-5p), and MIMAT0004556 (hsa-miR-10b-3p).

Survival analysis

Top 20 mRNAs in MEturquoise and MEblue modules were selected according to connectivity degrees and were used as candidate genes for Kaplan-Meier survival analysis. Other candidate genes included the 28 lncRNAs and 16 miRNAs in MEturquoise and MEblue modules. A total of 16 genes ($P<.01$), including 9 mRNAs and 7 miRNAs, were identified as potential prognostic factors (Figure 6). These genes were beta-1,4-galactosyltransferase 3 (*B4GALT3*), complement component 7 (*C7*), *CRHBP*,

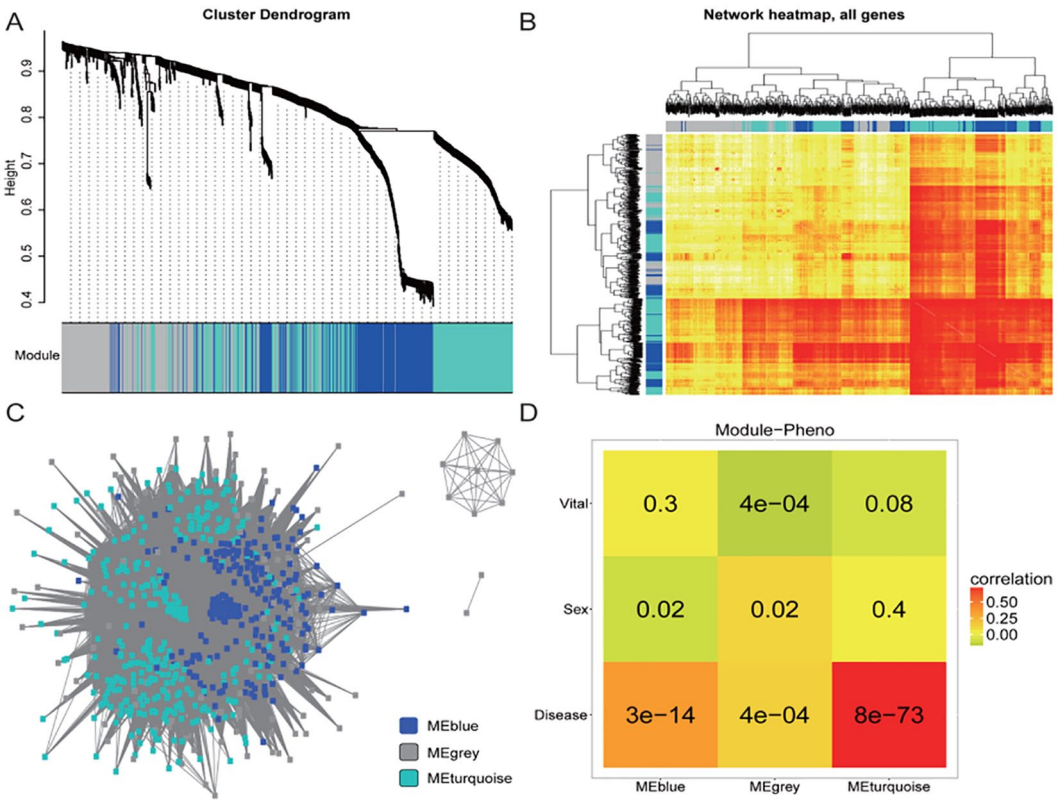

Figure 4. Weighted gene coexpression network analysis (WGCNA): (A) Gene dendrogram derived from hierarchical clustering. Three different modules were indicated by gray, blue, and turquoise underneath the dendrogram. (B) Hierarchical clustering analysis. Coexpression level was indicated by color temperature. (C) Coexpression network. genes in different modules were indicated by dots with the corresponding colors. (D) Module-phenotype relationships. Correlation coefficients were indicated by color from green (low) to red (high), and *P* values were shown in corresponding grids. WGCNA indicates weighted gene coexpression network analysis.

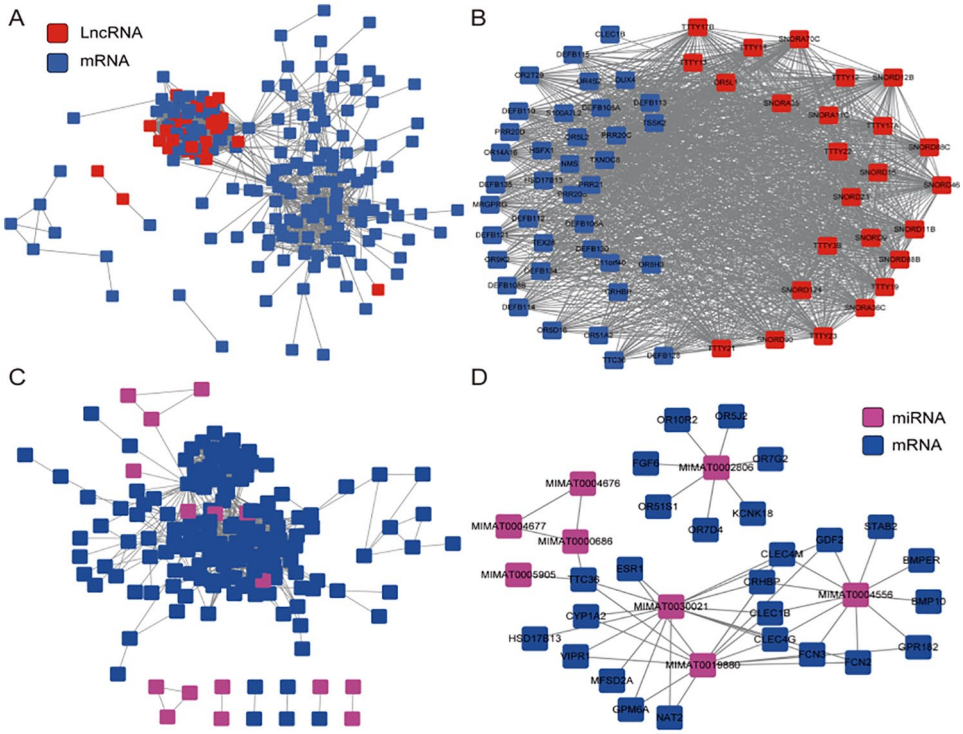

Figure 5. LncRNA-mRNA and miRNA-mRNA coexpression network. (A) The lncRNA-mRNA coexpession network. (B) A network consisting of the 25 lncRNAs from the major subnetwork of (A) and coexpressed mRNAs. (C) The miRNA-mRNA coexpession network. (D) A network consisting of miRNAs from the major subnetwork of (C) and coexpressed mRNAs. lncRNA indicates long noncoding RNA; miRNA, microRNA.

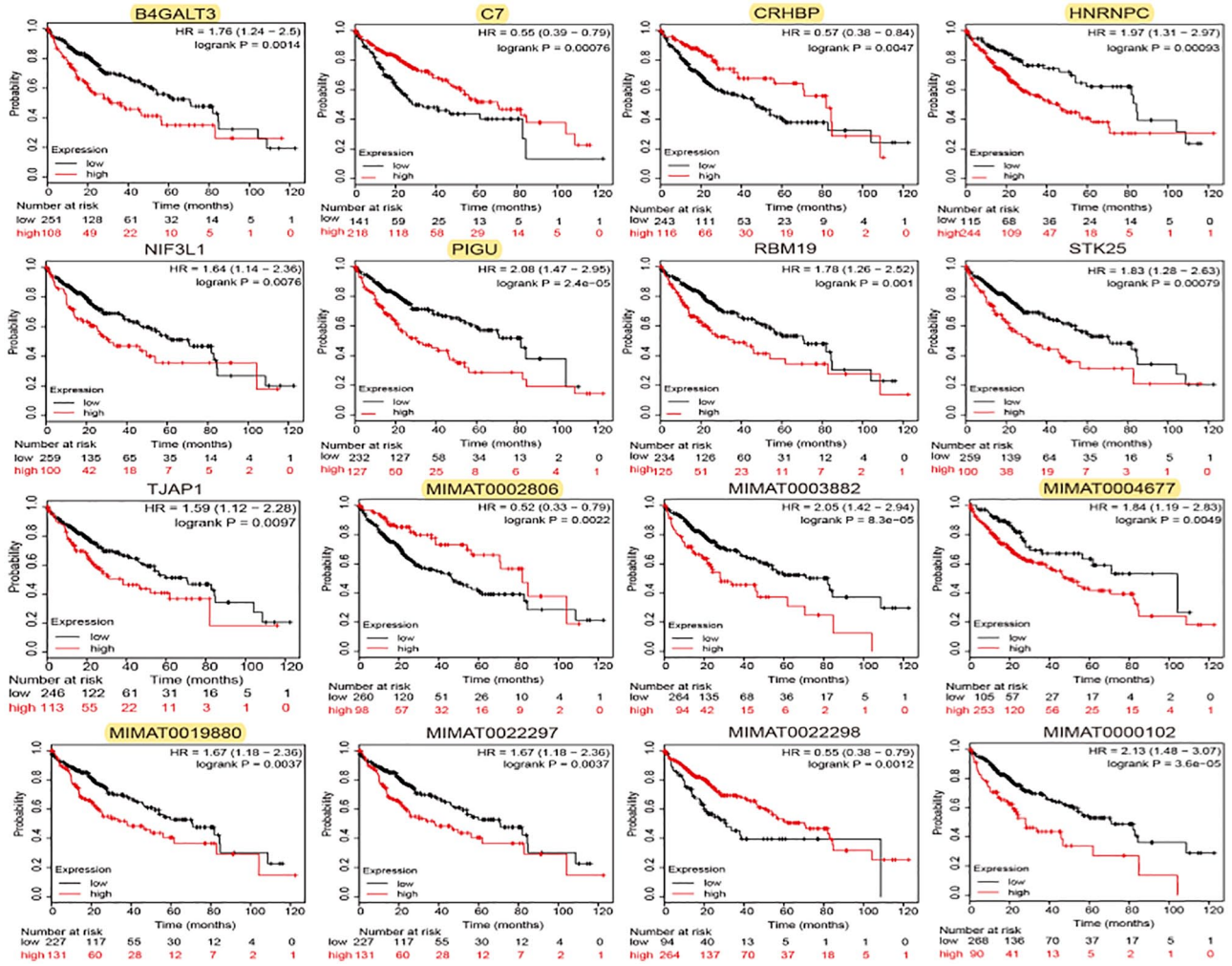

Figure 6. Kaplan-Meier survival analysis of candidate genes.
Red and blue lines indicated patient groups with expression levels above and below median value, respectively.

heterogeneous nuclear ribonucleoprotein C1/C2 (*HNRNPC*), NGG1 interacting factor 3 like 1 (*NIF3L1*), phosphatidylinositol glycan anchor biosynthesis class U (*PIGU*), RNA-binding motif protein 19 (*RBM19*), serine/threonine kinase 25 (*STK25*), tight junction-associated protein 1 (*TJAP1*), MIMAT0002806, MIMAT0003882 (hsa-miR-767-5p), MIMAT0004677 (hsa-miR-34c-3p), MIMAT0019880, MIMAT0022297 (hsa-miR-5589-5p), MIMAT0022298 (hsa-miR-5589-3p), and MIMAT0000102 (hsa-miR-105-5p). High expression was associated with better survival for *C7, CRHBP*, MIMAT0002806, and MIMAT0022298, and high expression was associated worse prognosis for the remaining 12 genes (Figure 6).

Validation of prognosis-associated miRNAs in an independent data set

GSE36915 and GSE31384 were subjected for validation of differential expression and association with prognosis of the 7 miRNAs, respectively. Three miRNAs, including MIMAT0019880, MIMAT0022297, and MIMAT0022298 were not included in the 2 data sets. Therefore, validation was conducted for MIMAT0000102, MIMAT0002806, MIMAT0003882, and MIMAT0004677. As shown in Figure 7A, the expression levels of MIMAT0000102, MIMAT0003882, and MIMAT0004677 were significantly decreased in HCC compared with normal tissues ($P < .05$). This result is consistent with the previous analysis. The expression level of MIMAT0002806 was higher in HCC than that in normal tissues, which was in line with previous analysis; however, the difference did not reach to significance level ($P > .05$).

As shown in Figure 7B, all 4 miRNAs were significantly associated with prognosis ($P < .05$). The expression of MIMAT0000102, MIMAT0003882, and MIMAT0004677 was negatively associated with prognosis and the expression of MIMAT0002806 was positively associated with prognosis. This result was in agreement with the previous analysis.

Discussion

In our study, we comprehensively analyzed the expression data of HCC and identified DEmRNAs, DElncRNAs, and DEmiRNAs between tumor and normal samples. Based on WGCNA and coexpression analysis, an miRNA-mRNA

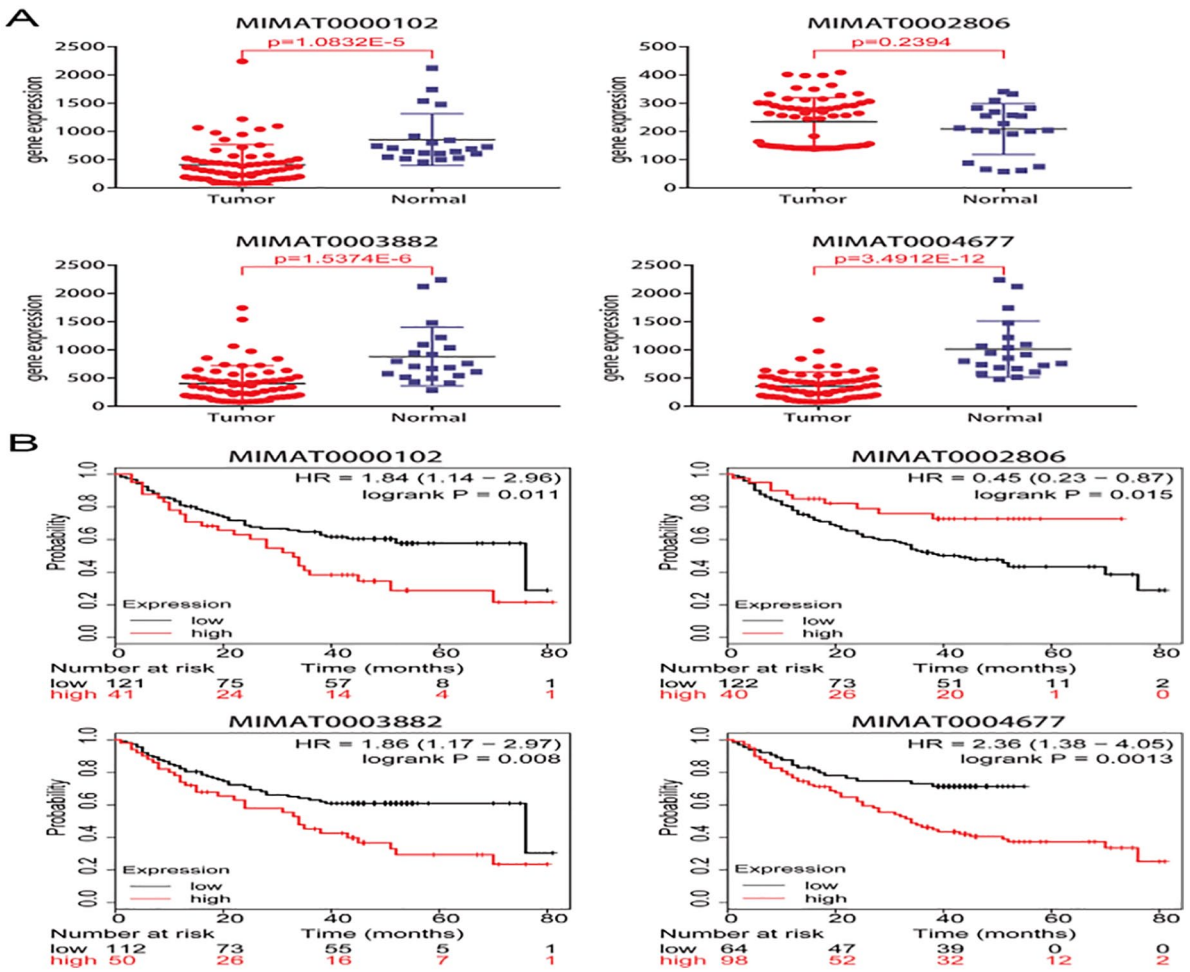

Figure 7. Validation of prognosis-associated miRNAs in an independent data set: (A) The differential expression level of 4 miRNAs in GSE36915 and (B) Kaplan-Meier survival analysis of 4 miRNAs in GSE31384. miRNA indicates microRNA; HR indicates hazard ratio.
Red and black lines indicated patient groups with expression levels above and below median value, respectively.

network and an lncRNA-mRNA regulation network were constructed. Following topological analysis and survival analysis, we further identified 9 DEmRNAs and 7 DEmiRNAs that were potential prognostic biomarkers. MIMAT0000102, MIMAT0003882, and MIMAT0004677 were successfully validated in independent data sets.

In total, 792 DEmRNAs were identified. According to our GO analysis, DEmRNAs were mainly enriched in CCs related to plasma membrane structure and BPs related to sensory perception. The membrane structure is an essential component for intercellular communication. Deregulation of membrane structure is frequently occurred in cancers and contributes to cancer cell proliferation and metastasis. Besides, it has also been reported that taste and smell functions are also deregulated in reported in cancer patients.[28] Differential expressions of olfactory receptors, such as *OR5L2* between HCC and normal samples were identified. Olfactory receptor contains more than 380 family members and is the largest gene family in human.[29] In recent years, ectopic expression of olfactory receptors was found to be related to many physiological processes, such as would healing, cytokinesis, and myocardial function.[30-32] Besides, dysregulation of olfactory receptors was reported to regulate cell proliferation, apoptosis, and migration in several cancers.[33-35] A recent study demonstrated that olfactory receptors could be used as biomarkers in human breast cancer tissues.[36] In addition to alterations in intercellular communication and sensory perception, cellular metabolism and cell cycle are also frequently deregulated in cancers.[37,38] Consistent with this, our KEGG analysis showed that DEmRNAs were significantly enriched in various KEGG pathways related to the metabolism of amino acids, carbohydrates, and fatty acids, indicating universal metabolic remodeling in HCC. Meanwhile, DEmRNAs were also significantly enriched in KEGG pathways such as DNA replication and p53 signaling, suggesting that the cell cycle of HCC may be out of control to a large extent. Taken together, our results indicated a wide range of biological functions and pathways in HCC were dysregulated.

To analyze the dynamic regulation of lncRNAs, miRNAs, and mRNAs, we constructed an lncRNA-mRNA coexpression network and an miRNA-mRNA coexpression network. Generally, nodes with high connectivity degree are correlated with important biological functions.[39,40] In the 2 coexpression networks, *SNORD9, CRHBP, CLEC1B, GDF2, OR5L2*, MIMAT0002806, MIMAT0030021, MIMAT0019880, and

MIMAT0004556 were genes with high connectivity, indicating that they may play important roles in the development and progression of LIHC.

Our subsequent survival analysis confirmed that 3 of the hub genes, *CRHBP*, MIMAT0002806, and MIMAT0019880, were prognosis-associated genes, further supporting their essential roles in HCC. Patients with low expression of *CRHBP* and MIMAT0002806 had a worse prognosis than those with high expression levels, whereas patients with low expression of MIMAT0019880 had a better prognosis. The expression of *CRHBP* is significantly suppressed in HCC and reduced *CRHBP* expression is proposed as a negative predictor of clinical prognosis.[41,42] MIMAT0002806 is upregulated in HCC and may promote cancer progression by enhancing cell proliferation and stimulating epithelial to mesenchymal transition.[43] Few clues have been reported about whether MIMAT0019880 is associated with HCC. However, based on our results, we speculated that MIMAT0019880 may also be potential prognostic indicators of HCC.

In addition to *CRHBP*, MIMAT0002806, and MIMAT0019880, 13 other genes, including *B4GALT3, C7, PIGU, HNRNPC*, and MIMAT0004677 (hsa-miR-34c-3p), were also identified to be prognosis-associated genes. *B4GALT3* encodes a protein-mediating glycosylation, and suppression of *B4GALT3* may promote nuclear factor-κB phosphorylation in HCC, thereby enhancing cell motility and cancer progression.[44] Higher serum C7 protein level of HCC has been proposed as an indicator of poor response to transarterial chemoembolization.[45] Both *PIGU* and *HNRNPC* have been confirmed to be overexpressed in HCC and the high expression level of these 2 genes is associated with unfavorable prognosis.[46,47] In addition, MIMAT0004677 plays an oncogenic role in HCC through promoting cancer progression by directly targeting NCKAP1 and low expression of MIMAT0004677 is an indicator of favorable survival.[48] Consequently, *B4GALT3, C7, PIGU, HNRNPC*, and MIMAT0004677 may be potential prognostic indicators of HCC.

The main advantage of our study was that a combination of WGCNA, topological analysis, and survival analysis was implemented to identify HCC prognostic factors. Several prognostic factors including *CRHBP*, MIMAT0002806, and MIMAT0019880 were eventually identified. Among them, MIMAT0019880 was a novel HCC biomarker, which has never been proposed as an indicator of HCC prognosis. However, independent data sets are still needed to verify the prognostic significance of the biomarkers identified in our study. In addition, experimental investigations should also be conducted to provide an insight into the underlying roles of the prognostic biomarkers in HCC.

Conclusion

In conclusion, we analyzed the dynamic transcriptome regulation in HCC and identified 9 DEmRNAs and 7 DEmiRNAs that might be potential prognostic biomarkers for HCC. MIMAT0000102, MIMAT0003882, and MIMAT0004677 were successfully validated in independent data sets. These genes correlated significantly with prognosis and might contribute to HCC diagnosis and treatment in future clinical practice if implemented into decision-making processes.

Author Contributions

SPQ, QYS and JX analyzed the data and drafted the manuscript. WWY and HWF designed the study and obtained the fundings. All authors reviewed the final version of the manuscript.

ORCID iD

Hengwei Fan https://orcid.org/0000-0001-5367-7806

REFERENCES

1. Maluccio M, Covey AJ. Recent progress in understanding, diagnosing, and treating hepatocellular carcinoma. *CA Cancer J Clin*. 2012;62:394-399.
2. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. *Lancet*. 2018;391:1301-1314.
3. Ghouri YA, Mian I, Rowe JH. Review of hepatocellular carcinoma: epidemiology, etiology, and carcinogenesis. *J Carcinog*. 2017;16:1.
4. Khalaf N, Ying J, Mittal S, et al. Natural history of untreated hepatocellular carcinoma in a US cohort and the role of cancer surveillance. *Clin Gastroenterol Hepatol*. 2017;15:273-281.e271.
5. Zhang YC, Xu Z, Zhang TF, Wang YL. Circulating microRNAs as diagnostic and prognostic tools for hepatocellular carcinoma. *World J Gastroenterol*. 2015;21:9853-9862.
6. Flavahan WA, Gaskell E, Bernstein BE. Epigenetic plasticity and the hallmarks of cancer. *Science*. 2017;357:eaal2380.
7. Liz J, Esteller M. LncRNAs and microRNAs with a role in cancer development. *Biochim Biophys Acta*. 2016;1859:169-176.
8. Lui KY, Zhao H, Qiu C, et al. Integrator complex subunit 6 (INTS6) inhibits hepatocellular carcinoma growth by Wnt pathway and serve as a prognostic marker. *BMC Cancer*. 2017;17:644.
9. Xiao H, Zhou B, Jiang N, et al. The potential value of CDV3 in the prognosis evaluation in Hepatocellular carcinoma. *Genes Dis*. 2018;5:167-171.
10. Esteller M. Non-coding RNAs in human disease. *Nat Rev Genet*. 2011;12:861.
11. Ling H, Fabbri M, Calin GA. MicroRNAs and other non-coding RNAs as targets for anticancer drug development. *Nat Rev Drug Discov*. 2013;12:847-865.
12. Chen WX, Zhang ZG, Ding ZY, et al. MicroRNA-630 suppresses tumor metastasis through the TGF-beta- miR-630-Slug signaling pathway and correlates inversely with poor prognosis in hepatocellular carcinoma. *Oncotarget*. 2016;7:22674-22686.
13. Liu J, Lu C, Xiao M, Jiang F, Qu L, Ni R. Long non-coding RNA SNHG20 predicts a poor prognosis for HCC and promotes cell invasion by regulating the epithelial-to-mesenchymal transition. *Biomed Pharmacother*. 2017;89:857-863.
14. Zhang D, Cao C, Liu L, Wu D. Up-regulation of LncRNA SNHG20 predicts poor prognosis in hepatocellular carcinoma. *J Cancer*. 2016;7:608-617.
15. Haeussler M, Zweig AS, Tyner C, et al. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res*. 2019;47:D853-D858.
16. Harrow J, Frankish A, Gonzalez JM, et al. GENCODE: the reference human genome annotation for the ENCODE project. *Genome Res*. 2012;22:1760-1774.
17. Ritchie ME, Phipson B, Wu D, et al. LIMMA powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43:e47.
18. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4:44-57.
19. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28:27-30.
20. Damian D, Gorfine M. Statistical concerns about the GSEA procedure. *Nat Genet*. 2004;36:663.
21. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
22. Giulietti M, Occhipinti G, Principato G, Piva F. Identification of candidate miRNA biomarkers for pancreatic ductal adenocarcinoma by weighted gene co-expression network analysis. *Cell Oncol (Dordr)*. 2017;40:181-192.

23. Jason M. Psych issues. *JEMS*. 2013;38:14.
24. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13: 2498-2504.
25. Bland JM, Altman DG. Survival probabilities (the Kaplan-Meier method). *BMJ*. 1998;317:1572.
26. Shih TC, Tien YJ, Wen CJ, et al. MicroRNA-214 downregulation contributes to tumor angiogenesis by inducing secretion of the hepatoma-derived growth factor in human hepatoma. *J Hepatol*. 2012;57:584-591.
27. Wei R, Huang GL, Zhang MY, et al. Clinical significance and prognostic value of microRNA expression signatures in hepatocellular carcinoma. *Clin Cancer Res*. 2013;19:4780-4791.
28. Spotten LE, Corish CA, Lorton CM, et al. Subjective and objective taste and smell changes in cancer. *Ann Oncol*. 2017;28:969-984.
29. Firestein S. How the olfactory system makes sense of scents. *Nature*. 2001;413:211-218.
30. Zhang X, Bedigian AV, Wang W, Eggert US. G protein-coupled receptors participate in cytokinesis. *Cytoskeleton (Hoboken)*. 2012;69:810-818.
31. Busse D, Kudella P, Gruning NM, et al. A synthetic sandalwood odorant induces wound-healing processes in human keratinocytes via the olfactory receptor OR2AT4. *J Invest Dermatol*. 2014;134:2823-2832.
32. Jovancevic N, Dendorfer A, Matzkies M, et al. Medium-chain fatty acids modulate myocardial function via a cardiac odorant receptor. *Basic Res Cardiol*. 2017;112:13.
33. Massberg D, Simon A, Haussinger D, et al. Monoterpene (-)-citronellal affects hepatocarcinoma cell signaling via an olfactory receptor. *Arch Biochem Biophys*. 2015;566:100-109.
34. Kalbe B, Schulz VM, Schlimm M, et al. Helional-induced activation of human olfactory receptor 2J3 promotes apoptosis and inhibits proliferation in a non-small-cell lung cancer cell line. *Eur J Cell Biol*. 2017;96:34-46.
35. Weber L, Al-Refae K, Ebbert J, et al. Activation of odorant receptor in colorectal cancer cells leads to inhibition of cell proliferation and apoptosis. *PLoS ONE*. 2017;12:e0172491.
36. Weber L, Massberg D, Becker C, et al. Olfactory receptors as biomarkers in human breast carcinoma tissues. *Front Oncol*. 2018;8:33.
37. Shender V, Arapidi G, Pavlyukov M, et al. The role of intercellular communication in cancer progression. *Russ J Bioorg Chem*. 2018;44:473-480.
38. Pascual G, Dominguez D, Benitah SA. The contributions of cancer cell metabolism to metastasis. *Dis Model Mech*. 2018;11:dmm032920.
39. Yuan L, Chen L, Qian K, et al. Co-expression network analysis identified six hub genes in association with progression and prognosis in human clear cell renal cell carcinoma (ccRCC). *Genom Data*. 2017;14:132-140.
40. Vallabhajosyula RR, Chakravarti D, Lutfeali S, Ray A, Raval A. Identifying hubs in protein interaction networks. *PLoS ONE*. 2009;4:e5344.
41. Xia HB, Wang HJ, Fu LQ, et al. Decreased CRHBP expression is predictive of poor prognosis in patients with hepatocellular carcinoma. *Oncol Lett*. 2018;16: 3681-3689.
42. Ho DW, Kai AK, Ng IO. TCGA whole-transcriptome sequencing data reveals significantly dysregulated genes and signaling pathways in hepatocellular carcinoma. *Front Med*. 2015;9:322-330.
43. Zhang LY, Liu M, Li X, Tang H. miR-490-3p modulates cell growth and epithelial to mesenchymal transition of hepatocellular carcinoma cells by targeting endoplasmic reticulum-Golgi intermediate compartment protein 3 (ERGIC3). *J Biol Chem*. 2013;288:4035-4047.
44. Fang T, Lv H, Lv G, et al. Tumor-derived exosomal miR-1247-3p induces cancer-associated fibroblast activation to foster lung metastasis of liver cancer. *Nat Commun*. 2018;9:191.
45. Yu SJ, Kim H, Min H, et al. Targeted proteomics predicts a sustained complete-response after transarterial chemoembolization and clinical outcomes in patients with hepatocellular carcinoma: a prospective cohort study. *J Proteome Res*. 2017;16:1239-1248.
46. Cao J, Wang P, Chen J, He X. PIGU overexpression adds value to TNM staging in the prognostic stratification of patients with hepatocellular carcinoma. *Hum Pathol*. 2019;83:90-99.
47. Tremblay MP, Armero VE, Allaire A, et al. Global profiling of alternative RNA splicing events provides insights into molecular differences between various types of hepatocellular carcinoma. *BMC Genomics*. 2016;17:683.
48. Xiao CZ, Wei W, Guo ZX, et al. MicroRNA-34c-3p promotes cell proliferation and invasion in hepatocellular carcinoma by regulation of NCKAP1 expression. *J Cancer Res Clin Oncol*. 2017;143:263-273.