

ORIGINAL ARTICLE

Comprehensive analysis of partial epithelial mesenchymal transition-related genes in hepatocellular carcinoma

Yu Lei | Wei Yan | Zhuoying Lin | Jingmei Liu | Dean Tian | Ping Han 

Department of Gastroenterology, Tongji Hospital of Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

Correspondence

Dean Tian and Ping Han, Department of Gastroenterology, Tongji Hospital of Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, Hubei Province, China.
Emails: datian@tjh.tjmu.edu.cn (DT); hanzhouping@163.com (PH)

Funding information

National Natural Science Foundation of China, Grant/Award Number: 81702396, 81772607 and 81800547

Abstract

Increasing evidence has revealed that cancer cells undergoing an intermediate state, partial epithelial mesenchymal transition (p-EMT), tend to metastasize rather than complete EMT. We performed a comprehensive analysis of E-cadherin and 25 p-EMT-related genes in HCC to explore the roles and regulatory mechanisms of them in HCC. We analysed E-cadherin and 25 p-EMT-related genes in HCC and constructed an mRNA-miRNA-lncRNA ceRNA subnetwork containing p-EMT-related genes by bioinformatic approaches. IHC was used to identify the protein expression of key p-EMT-related genes, P4HA2, ITGA5, MMP9, MT1X and SPP1. Complete EMT is not necessary for HCC progression. Overexpression of P4HA2, ITGA5, MMP9, SPP1 and down-regulation of MT1X were found in HCC tissues, which were significantly associated with poor prognosis of HCC patients. By means of stepwise reverse prediction and validation from mRNA to lncRNA, an mRNA-miRNA-lncRNA ceRNA subnetwork correlated with HCC prognosis was identified by expression and survival analysis. This study implied that key p-EMT-related genes P4HA2, ITGA5, MMP9, MT1X, SPP1 could be prognostic biomarkers and potential targets of therapy for HCC patients. We constructed an mRNA-miRNA-lncRNA subnetwork containing p-EMT-related genes successfully, among which each component might be utilized as a prognostic biomarker of HCC.

KEYWORDS

competing endogenous RNA, hepatocellular carcinoma, partial epithelial mesenchymal transition, prognosis

1 | INTRODUCTION

Hepatocellular carcinoma (HCC) is one of the most common cancers worldwide and listed as the third leading cause of cancer-related mortality.¹ The long-term prognosis of patients with HCC after hepatectomy remains poor because of frequent metastasis.² Though efforts have been made in investigating the mechanisms of the HCC

initiation and progression, detailed mechanisms of HCC pathogenesis remain largely unclear. Promising biomarkers of HCC urgently need to be identified for HCC diagnosis and therapy improvement.

It is widely acknowledged that epithelial mesenchymal transition (EMT) plays a vital role in invasion and metastasis in diverse types of cancer, including HCC.^{3,4} Cancer cells during EMT lose their epithelial features and gain fully mesenchymal phenotypes,

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Journal of Cellular and Molecular Medicine* published by Foundation for Cellular and Molecular Medicine and John Wiley & Sons Ltd.

known as complete EMT.⁵ Down-regulation of epithelial markers such as E-cadherin (also named CDH1), and up-regulation of mesenchymal markers such as vimentin are considered as hallmarks of EMT and growing studies have shown identify the importance of EMT in HCC progression.^{6,7} Recently, an intermediate state of EMT named partial-EMT(p-EMT), during which cancer cells exhibit both mesenchymal and epithelial features, has attracted more and more attention. Recent studies have found that metastatic cancer cells express a certain level of E-cadherin, and cancer cells in p-EMT state have an intense trend to adapt to the metastatic microenvironment a higher risk of metastasis.⁸ Our understanding of p-EMT programme was enhanced using sophisticated techniques such as lineage tracing in patient-derived xenografts, genetically engineered mouse models (GEMM), single-cell RNA sequence, fluorescent-activated cell sorting (FACS).⁹ A single-cell transcriptomic analysis of head and neck squamous cell carcinoma (HNSCC) has identified several representative genes in p-EMT programme in malignancy including 15 common p-EMT-related genes (SERPINE1, TGFB1, MMP10, LAMC2, P4HA2, PDPN, ITGA5, LAMA3, CDH13, TNC, MMP2, EMP3, INHBA, LAMB3 and VIM) and 10 variable p-EMT-related genes (THBS2, CXCL13, FN1, MMP3, MMP9, RAB25, MT1X, GPX3, SPP1 and MXD1).¹⁰ The 25 p-EMT-related genes are predominantly abundant in extracellular matrix and cell membrane. Previous studies have identified the presence of p-EMT programme in HCC cells.^{11,12} However, the nature of p-EMT-related genes in HCC remains unknown. Although several hallmarks, such as p53 and WNT signalling, were reported to be closely correlated with HCC, little evidence demonstrated their direct role as markers of EMT process. Due to the lack of using sophisticated techniques in p-EMT programme in HCC, we performed analysis of 25 p-EMT-related genes identified in HNSCC previously, all of which were confirmed to be involved in EMT programme.

A competing endogenous RNA (ceRNA) hypothesized that non-coding RNA (ncRNA), including long non-coding RNA (lncRNA), can gain cross-talk with mRNAs by competitively binding to shared miRNAs and then form a regulatory network was proposed by Salmena et al¹³ Accumulating evidence has indicated that the lncRNA-miRNA-mRNA ceRNA network may play an important role in cancer metastasis including HCC.¹⁴ Recently, a lncRNA-miRNA-mRNA ceRNA network associated with diagnosis and prognosis of HCC has been established.¹⁵ Nevertheless, understanding of mRNA-miRNA-lncRNA ceRNA networks containing p-EMT-related genes correlated significantly with prognosis of HCC remain extremely limited and need to be explored.

In our study, we performed a comprehensive analysis of E-cadherin in HCC. Then, we analysed the expression patterns, prognostic values, gene mutations, mutual interactions,

correlations with each other of 25 p-EMT-related genes and identified key genes after comprehensive consideration of expression and prognostic roles of p-EMT-related genes using online databases. We selected key upstream miRNAs and lncRNAs regulating the key genes to construct an mRNA-miRNA-lncRNA ceRNA subnetwork associated with prognosis of HCC. Furthermore, we identified protein expression levels of the selected key p-EMT-related genes in HCC tissues and adjacent nontumour liver tissues by immunohistochemistry (IHC) staining. They may also serve as potential biomarkers for diagnostic and therapeutic improvement of HCC in the future.

2 | MATERIAL AND METHODS

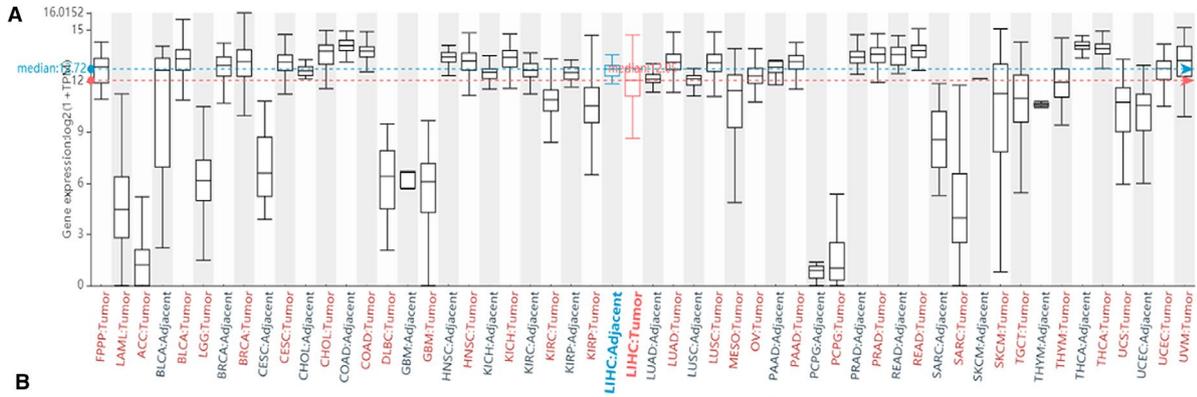
2.1 | HCCDB database analysis

HCCDB (<http://lifeome.net/database/hccdb/home.html>) is an integrative molecular database of HCC containing 15 public HCC datasets, such as Gene Expression Omnibus (GEO), Liver Hepatocellular Carcinoma Project of The Cancer Genome Atlas (TCGA-LIHC) and Liver Cancer-RIKEN, JP Project from International Cancer Genome Consortium (ICGC-LIRI-JP).¹⁶ HCCDB database was utilized to analyse the expression patterns of E-cadherin and 25 p-EMT-related genes in HCC (Figure 1 & Figures S9-S33), as well as other types of cancers in TCGA. Prognostic value of E-cadherin was also explored by HCCDB.

2.2 | Patients and IHC assay

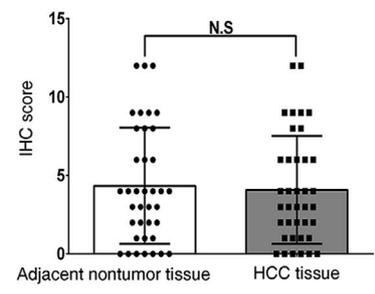
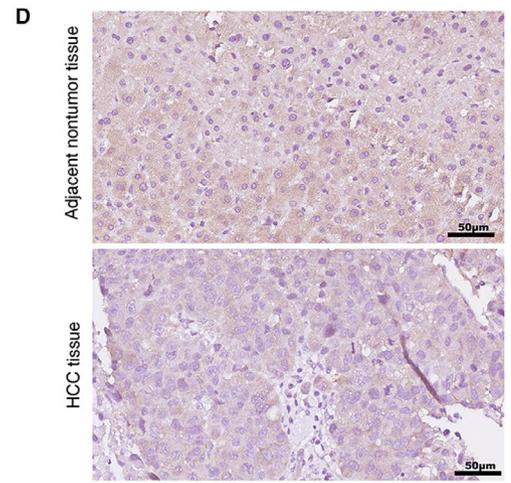
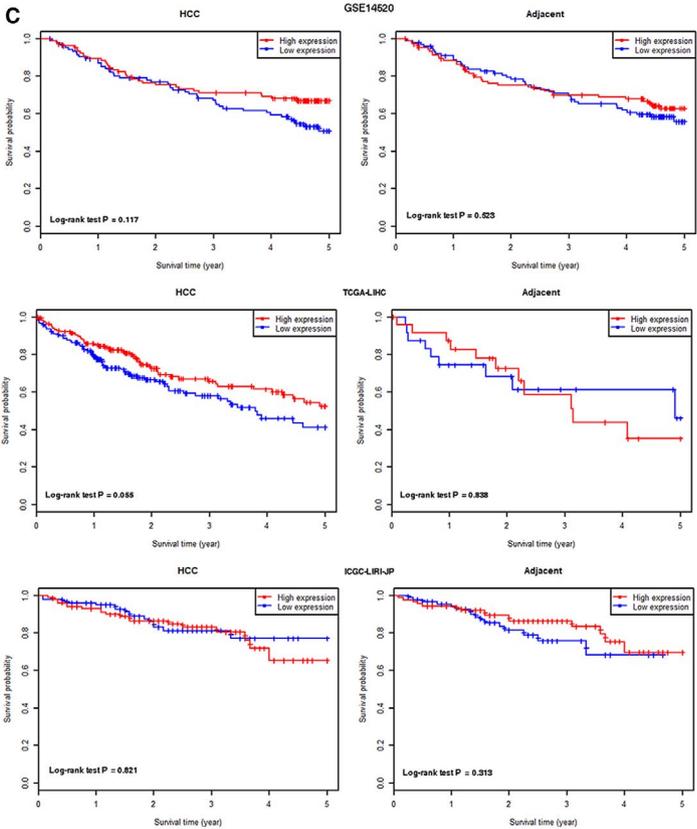
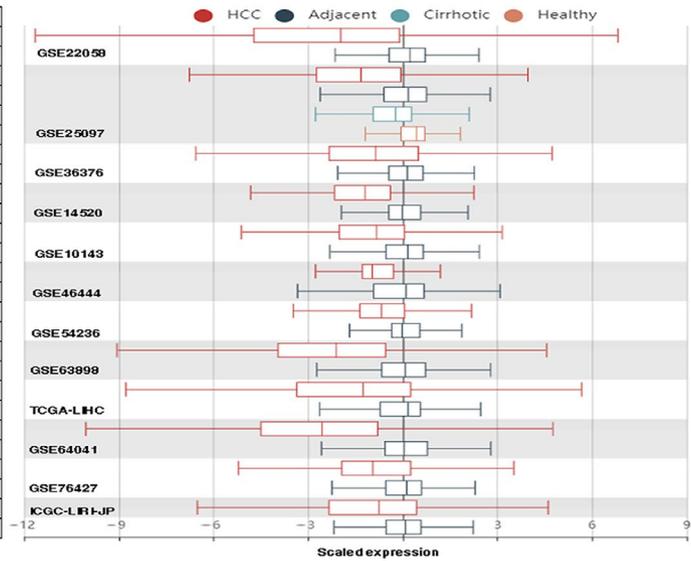
HCC samples and the corresponding adjacent nontumour liver tissues were collected from adult patients with HCC who underwent curative resection at the Tongji Hospital of Tongji Medical College (Wuhan, China). A preoperative clinical diagnosis of HCC was based on the diagnostic criteria of the American Association for the Study of Liver Diseases and haematoxylin & eosin staining of the samples was also performed. All of the samples were selected with distinctive pathologic diagnosis and none of the patients received any preoperative chemotherapy or radiotherapy. These tissues were stained for E-cadherin (Cell Signaling Technology, 3195), MMP10 (Abclonal, A3033), P4HA2 (Abcam, ab233197), ITGA5 (Abcam, ab150361), MMP9 (Cell Signaling Technology, 13667), MT1X (Proteintech, 17172-1-AP) and SPP1 (Abclonal, A1499) expression. Furthermore, negative control for IHC staining was shown in Figure S8. IHC staining was performed using the Dako Envision Plus System (Dako) according to the manufacturer's instructions. The IHC staining intensity was scored as 0 (negative);

FIGURE 1 Comprehensive analysis of E-cadherin in HCC. A, The mRNA expression levels of E-cadherin across multiple cancer types in TCGA. B, Chart and plot showing the expression of E-cadherin in HCC tissues and the adjacent normal tissues in HCCDB. C, The correlation between mRNA expression levels of E-cadherin and survival of HCC patients in three datasets. D, Representative IHC images and IHC score of E-cadherin in HCC tissues and adjacent nontumour tissues



B

Dataset	P-value	Type	Nums	Mean	STD	IQR
GSE22058	8.57E-10	HCC	100	11.94	1.178	1.221
		Adjacent	97	12.76	0.2642	0.3015
GSE25097	1.40E-21	HCC	268	8.06	4.145	6.077
		Adjacent	243	11.01	2.265	3.051
		Cirrhotic	40	10.18	2.118	2.759
		Healthy	6	11.24	2.298	1.7
GSE36376	1.57E-11	HCC	240	9.728	1.088	1.52
		Adjacent	193	10.29	0.5384	0.5837
GSE14520	1.23E-20	HCC	225	6.82	0.7763	0.8675
		Adjacent	220	7.427	0.4904	0.4921
GSE10143	4.16E-06	HCC	80	13.31	0.6759	0.748
		Adjacent	82	13.72	0.3616	0.429
GSE46444	2.56E-05	HCC	88	6.047	0.7736	1.081
		Adjacent	48	6.845	1.091	1.751
GSE54236	1.20E-05	HCC	81	12.04	0.9243	1.167
		Adjacent	80	12.67	0.8258	0.7357
GSE63898	2.49E-26	HCC	228	6.962	0.7908	0.8921
		Adjacent	168	7.622	0.2622	0.3615
TCGA-LIHC	2.68E-15	HCC	351	11.74	1.594	1.67
		Adjacent	49	12.65	0.4625	0.59
GSE64041	6.37E-10	HCC	60	9.882	0.9091	1.017
		Adjacent	60	10.76	0.2748	0.3687
GSE76427	5.9E-06	HCC	115	10.84	1.464	1.335
		Adjacent	52	11.6	0.6118	0.695
ICGC-LIRI-JP	6.77E-10	HCC	212	4.496	1.073	1.435
		Adjacent	177	5.026	0.5162	0.57



1 (weak); 2 (medium); 3 (strong). The percentage of positive cells was scored from 0 to 4 (0%, 1%-25%, 26%-50%, 51%-75%, 76%-100%). Overall score ranging from 0 to 12 was calculated by multiplying the above two scores, resulting in a negative (0-3) staining or a positive (4-12) staining for each example.

2.3 | UALCAN database analysis

UALCAN (<http://ualcan.path.uab.edu>) is a newly developed interactive web resource for 31 cancer types from TCGA database.¹⁷ In this study, UALCAN database, containing 371 primary HCC samples and 50 normal samples, was used to analyse mRNA expression levels of p-EMT-related genes in HCC. In addition, correlations between genes and clinical characteristics were investigated in UALCAN database. Transcript per million <1 was excluded due to the extremely low value. *P* value <.05 was considered as statistically significant.

2.4 | Oncomine database analysis

Oncomine (<https://www.oncomine.org/>) is an integrated online cancer microarray database.¹⁸ In this study, Oncomine was used to analyse mRNA expression levels of the 25 p-EMT-related genes in Roessler's data of HCC.¹⁹ Difference of mRNA expression was compared by Students' *t* test. Cut-off of *P* value and fold change were as following: *P* value: .01, fold change: 1.5.

2.5 | Kaplan-Meier Plotter analysis

The correlation between mRNA expression levels of p-EMT-related genes and prognosis of HCC patients was evaluated using Kaplan-Meier Plotter (www.kmplot.com), an online database containing information about the effects of 54 675 genes on survival in more than 20 types of cancers.²⁰ Each p-EMT-related gene was first entered into 'Liver cancer' item in this database. HCC patients were divided into high and low expression group according to median values of mRNA expression and then Kaplan-Meier overall survival curves were generated. Significant difference was considered when logrank *P* value <.05.

2.6 | Functional enrichment analysis

Database for Annotation, Visualization and Integrated Discovery (<https://david.ncifcrf.gov/>)²¹ was utilized to perform Gene Ontology (GO) functional annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of the 25 p-EMT-related genes. Information about the enriched GO terms and KEGG pathways with *P* < .05 was downloaded from the webpage and visualized using Python software.

2.7 | The cBioPortal database analysis

The cBioPortal (www.cbioportal.org) is an online open website resource capable to assess multidimensional cancer genomics data.²² In this study, we analysed the genomic profiles of the 25 p-EMT-related genes, which contained mutations, putative copy-number alterations from GISTIC and mRNA Expression z-Scores (RNASeq V2 RSEM) with a z-score threshold ± 1.8 . The correlations of p-EMT-related genes with each other were analysed via the cBioPortal online tool and visualized by using ggcorplot package of R software. Pearson's correction was included.

2.8 | Protein-protein interaction (PPI) network

The PPI interaction network between the p-EMT-related genes was constructed by Search Tool for the Retrieval of Interacting Genes (STRING) database (<http://string-db.org/>)²³ and visualized by Cytoscape software, an open source platform for visualizing complex networks.²⁴

2.9 | Prediction of miRNAs and lncRNAs

We predicted the upstream miRNAs of the six key p-EMT-related genes by utilizing miRTarbase (<http://mirtarbase.mbc.nctu.edu.tw/php/index.php>), an online database containing more than thousands of miRNA-target interactions validated experimentally by reporter assay, Western blot, microarray and next-generation sequencing experiments.²⁵ For the credibility of predicted results, only the miRNA-target interactions validated by reporter assay were selected for further analysis. The upstream candidate lncRNAs interacted with key miRNAs were predicted by using the miRNet database (<https://www.mirnet.ca/>), an online platform providing comprehensive analyses of miRNA-target interactions. Selection criteria were 'Organism-H.sapiens' and 'target type-lncRNAs'. The expression levels and prognostic values of predicted miRNAs and lncRNAs were analysed between HCC tissues and normal tissues using starBase v3.0 (<http://starbase.sysu.edu.cn/index.php>), an online database providing information on differential expression, survival and coexpression analysis of RNAs data from the TCGA projects.²⁶ *P* value <.05 was considered as statistically significant. We drew the Venn diagram using VENNY 2.1.0 (<http://bioinfogp.cnb.csic.es/tools/venny/index.html>), an interactive online tool.

2.10 | Correlation analysis

We explored the correlations of mRNA-miRNA, miRNA-lncRNA and mRNA-lncRNA pairs in HCC using starBase v3.0 database and *P* value <.05 was considered as statistically significant.

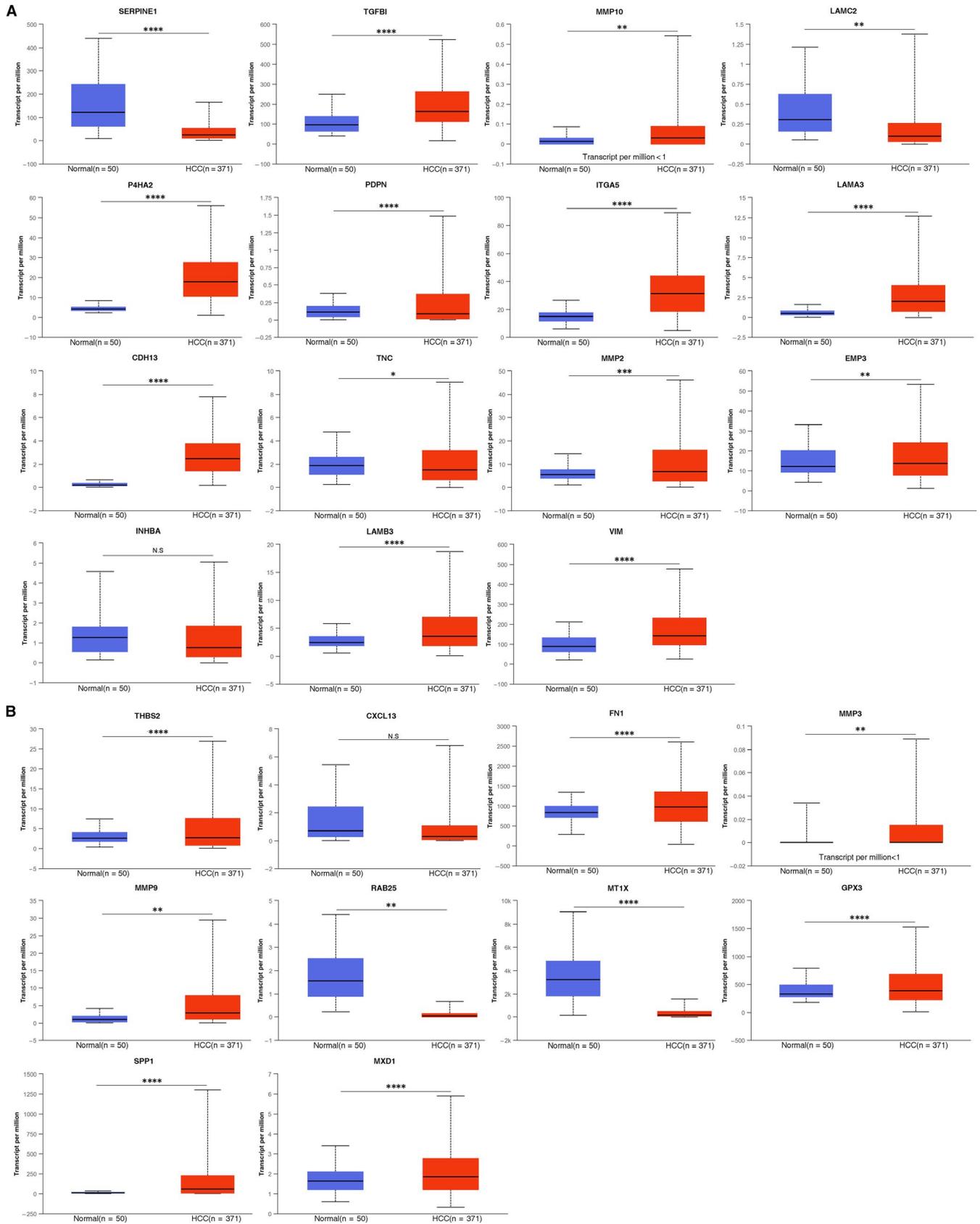


FIGURE 2 The mRNA expression of 25 p-EMT-related genes in HCC. The mRNA expression levels of 25 p-EMT-related genes in HCC

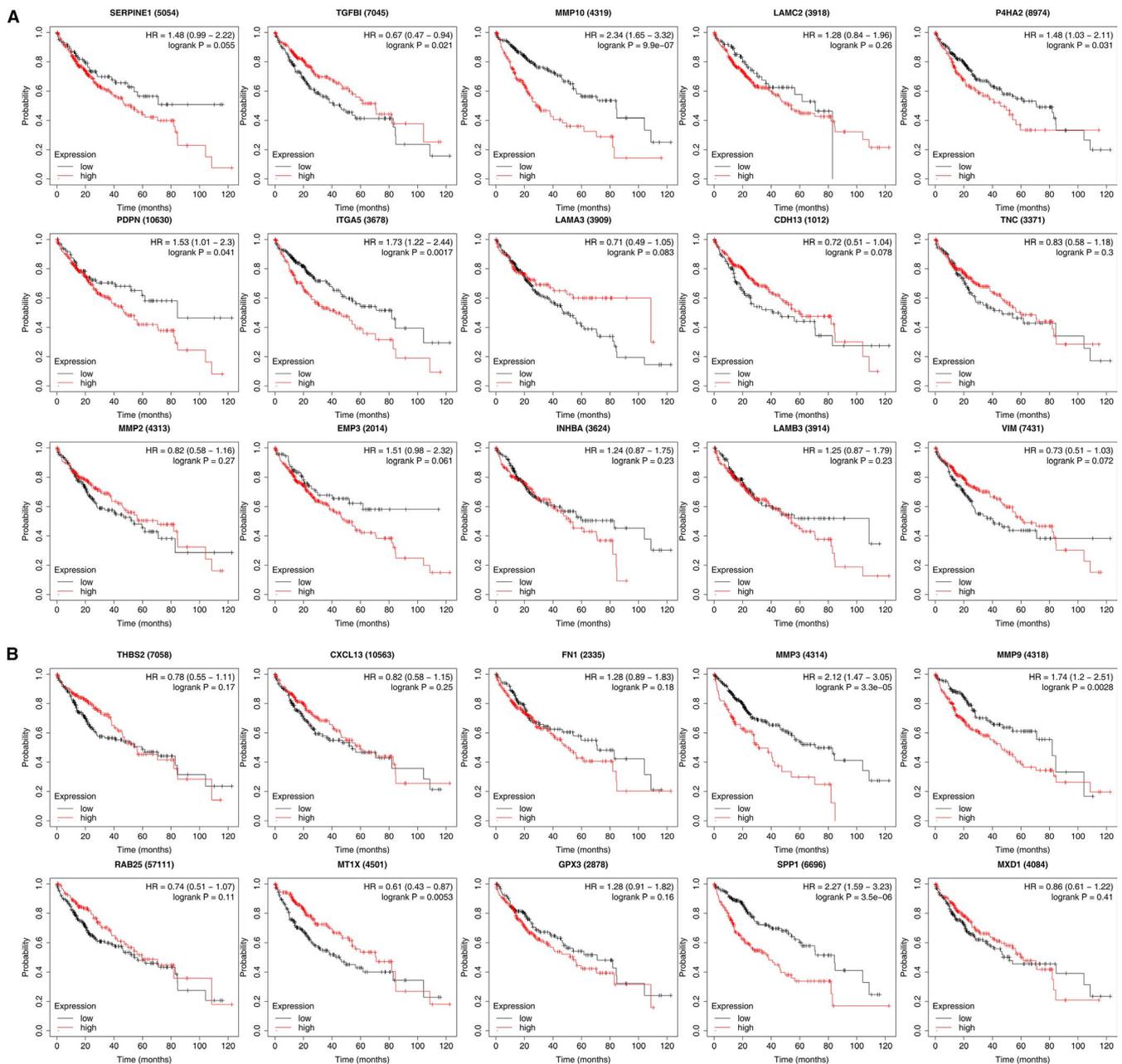


FIGURE 3 Prognostic value of mRNA expression levels of distinct p-EMT-related genes in HCC patients. The correlation between mRNA expression levels of distinct p-EMT-related genes and overall survival of HCC patients

3 | RESULTS

3.1 | Complete EMT may not be necessary for HCC metastasis

E-cadherin has been considered as a hall mark of EMT.³ Firstly, we performed a comprehensive analysis of E-cadherin as a marker for complete EMT in HCC. We evaluated the expression patterns of E-cadherin across multiple cancer types in TCGA and the differential expression level of E-cadherin in HCC was not very obvious compared with many other cancers (Figure 1A). Analysis of 12 HCC cohorts in the HCCDB database revealed that mRNA expression level of E-cadherin was lower in HCC tissues than in adjacent

nontumour tissues in 8 of 12 cohorts (Figure 1B), and no difference of E-cadherin expression was observed in the other cohorts. Intriguingly, prognostic values of E-cadherin for HCC were not significant in all of GEO, TCGA and ICGC-LIRI-JP datasets (Figure 1C). IHC staining showed that E-cadherin was mainly localized in the cell membrane. There was no significant difference between the protein expression level of E-cadherin in HCC tissues and adjacent nontumour tissues (Figure 1D). Moreover, E-cadherin expression might have no correlations with clinical features of HCC, including individual cancer stages, tumour grade and nodal metastasis status (Figure S1). Considering the results above, we thought the complete EMT might not be necessary for HCC progression and concentrated our attention on the role of p-EMT-related genes in HCC.

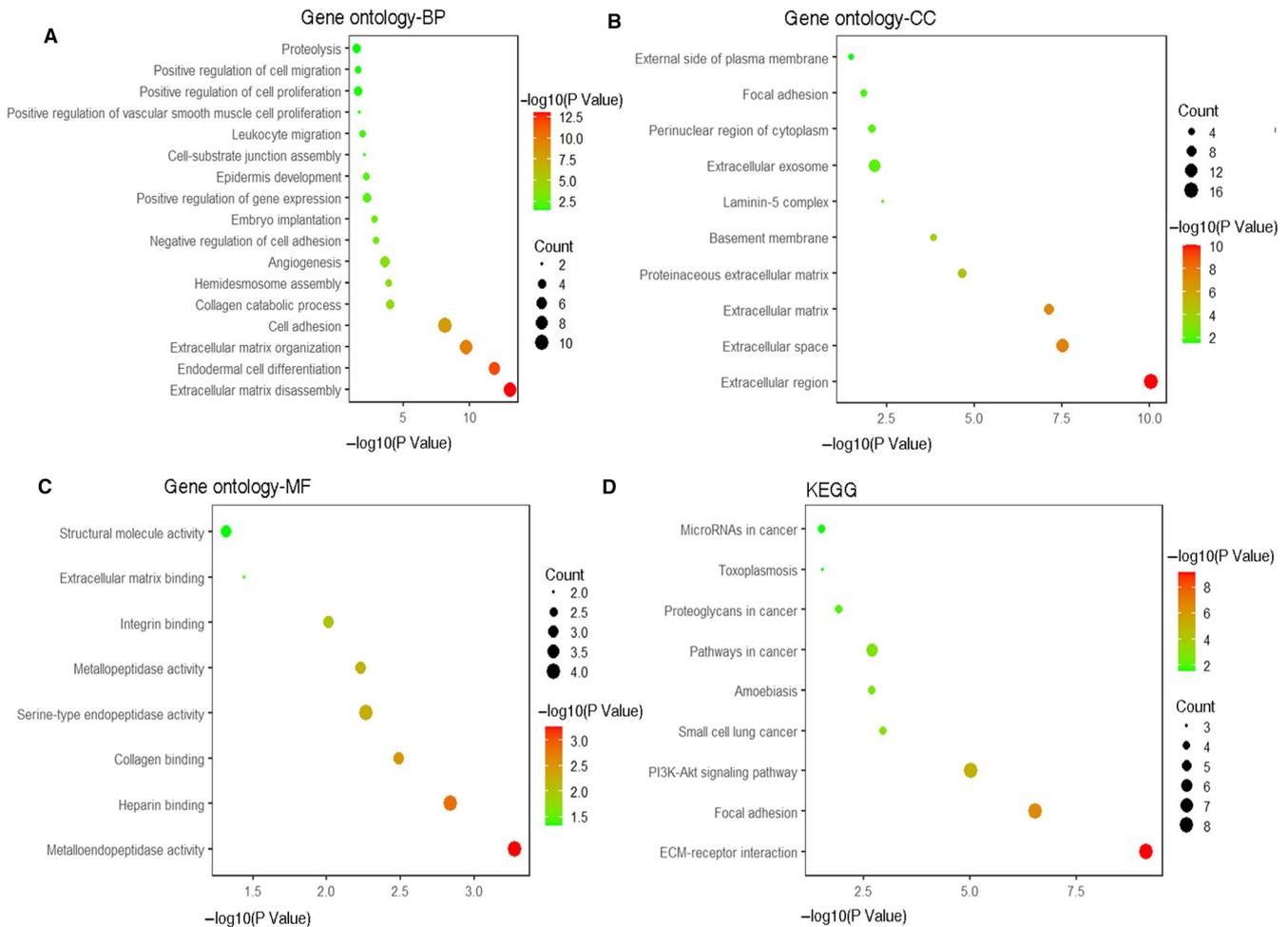


FIGURE 4 Functional enrichment analysis of 25 p-EMT-related genes. A, Gene ontology analysis of biological process for 25 p-EMT-related genes. B, Gene ontology analysis of cellular component for 25 p-EMT-related genes. C, Gene ontology analysis of molecular function for 25 p-EMT-related genes. D, KEGG analysis of p-EMT-related genes

3.2 | Different mRNA expression levels of p-EMT-related genes in HCC patients

We used UALCAN to compare the mRNA expression of p-EMT-related genes in HCC tissues and normal liver tissues (Figure 2 and Figure S2). The expression patterns of these genes were further confirmed the results in Oncomine and GEPIA (Figures S3 and S34). Among common p-EMT-related genes, the mRNA expression levels of P4HA2, ITGA5, LAMA3, CDH13, LAMB3, VIM were found significantly higher in HCC tissues than in normal tissues in all databases. Up-regulation of variable p-EMT-related genes MMP9 and SPP1 was observed in HCC tissues, while MT1X was remarkably down-regulated in HCC tissues.

3.3 | Correlation between the mRNA expression levels of p-EMT-related genes and prognosis of HCC patients

To determine the role of p-EMT-related genes in prognosis of HCC patients, we analysed the correlation between the mRNA expression levels of p-EMT-related genes and the survival of patients with

HCC by using Kaplan-Meier plotter. The results showed that the increased mRNA expression levels of MMP10, P4HA2, PDPN, ITGA5, MMP3, MMP9 and SPP1 were significantly associated with worse overall survival of HCC patients ($P < .05$). However, HCC patients with higher mRNA expression levels of TGFBI and MT1X showed better prognosis (Figure 3).

We selected common p-EMT-related genes P4HA2, ITGA5 and variable p-EMT-related genes MMP9, MT1X, SPP1 as key genes according to the criterion of converse expression pattern and survival prognosis. The mRNA expression levels of them all were remarkably correlated with patients' individual cancer stages (Figure S4).

3.4 | Functional enrichment analysis for p-EMT-related genes

In addition to biological processes identified already, the p-EMT programme may be associated with some other biological processes and pathways. We performed functional enrichment analysis to explore the particular biological functions of the 25 p-EMT related genes. Consistent with previous studies, the results showed that these

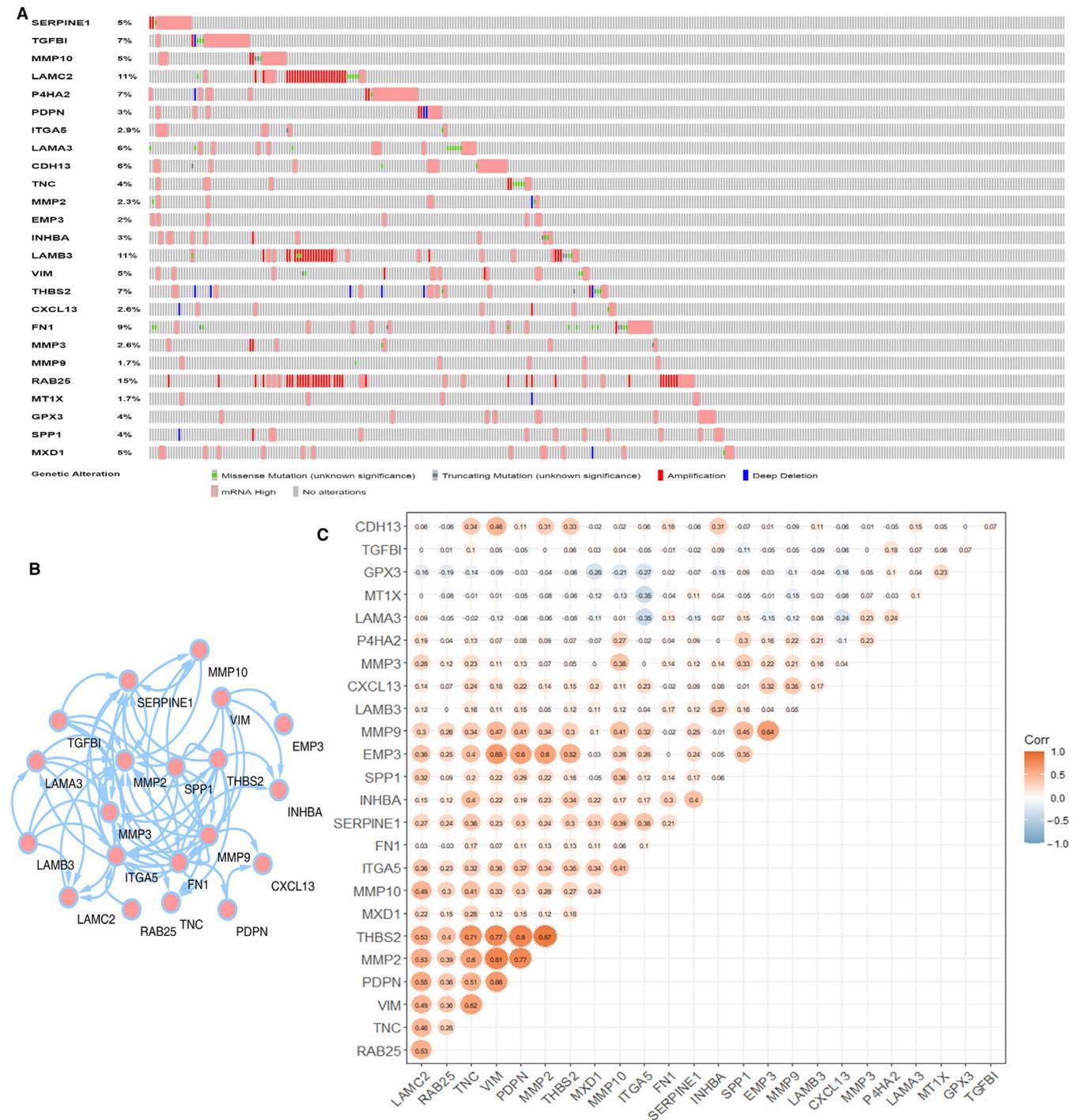


FIGURE 5 Analysis of genetic mutations, mutual interaction and correlations with each other in 25 p-EMT-related genes. A, Genetic mutations analysis of 25 p-EMT-related genes in HCC. B, The PPI network for 25 p-EMT-related genes. C, Correlations of different p-EMT-related genes with each other in HCC

genes were mainly enriched in extracellular matrix disassembly and organization. Moreover, KEGG pathways including ECM-receptor interaction, focal adhesion were closely related to p-EMT-related genes. Besides the common functions, these genes were also found to have close relationship with endodermal cell differentiation, leucocyte migration, PI3K-Akt signalling pathway, and several pathways in cancer, which indicated the implication of 25 p-EMT-related genes in cancer (Figure 4A-D).

3.5 | Genetic mutations in p-EMT-related genes and PPI network construction

The alterations of p-EMT-related genes were explored by using cBioPortal online tool for liver hepatocellular carcinoma (TCGA, PanCancer Atlas). The 25 p-EMT-related genes were altered in 222 samples of 348 patients with HCC (64%). More than two types of alterations of these genes were detected in HCC samples

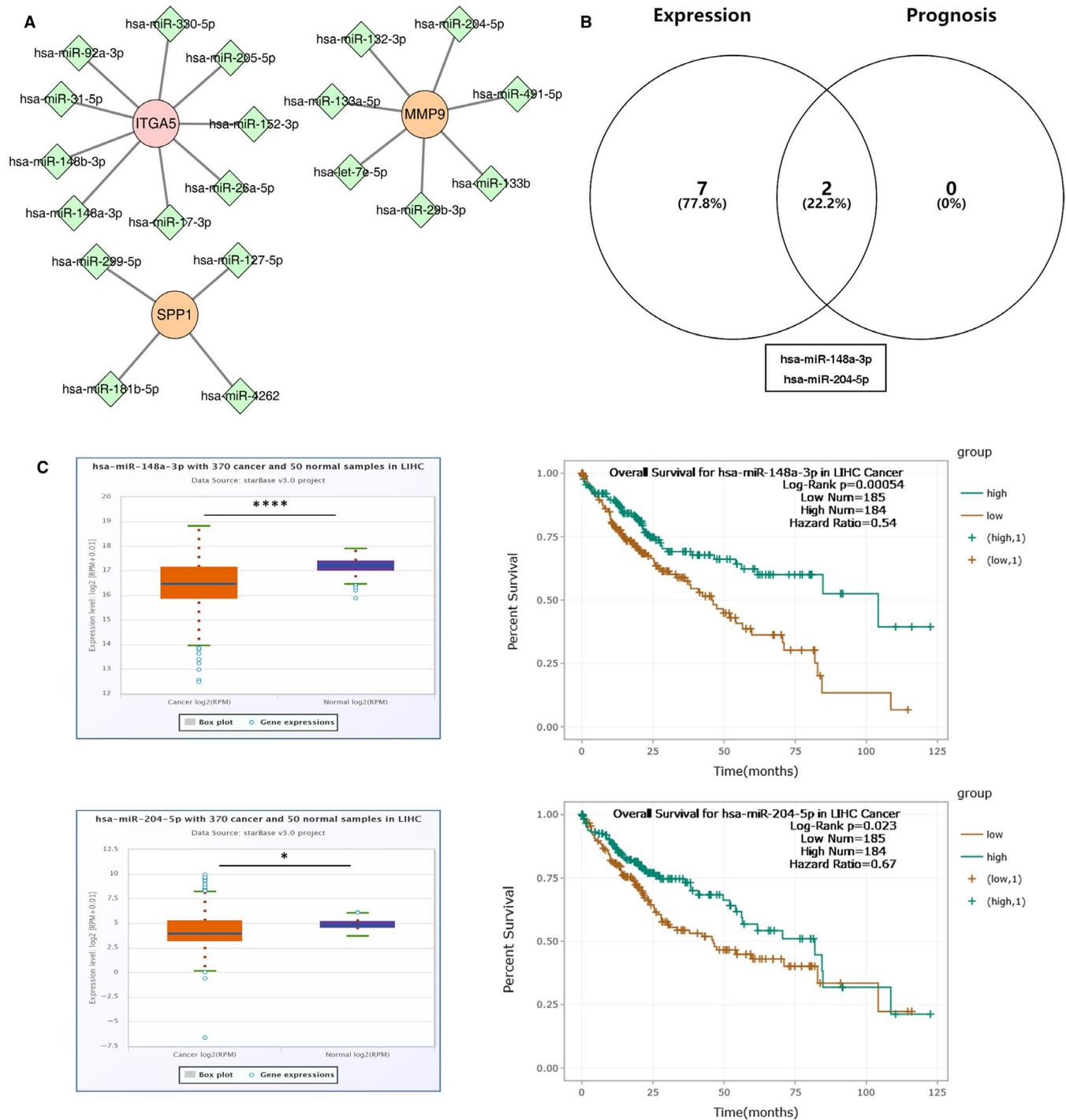


FIGURE 6 Prediction of key upstream miRNAs. A, The mRNA-miRNA network containing selected key p-EMT-related genes. B, Identification of the key upstream miRNAs with low expression and poor prognosis values in the predicted candidate miRNAs. C, The expression and survival analysis of miR-148a-3p and miR-204-5p in HCC

including mutation, amplification, depletion and mRNA alterations (Figure 5A). To figure out the mutual interactions of p-EMT-related genes, we utilized Cytoscape and String to construct a PPI network (Figure 5B). The result showed that 20 of p-EMT-related genes might have interactions with others. In addition, correlations of different p-EMT-related genes expressions with each

other were showed in the correlation heatmap and Pearson's correction was included (Figure 5C). The result demonstrated a low to high correlation among several p-EMT-related genes. For example, THBS2 expression was observed to be closely correlated with MMP2, while there was no correlation between P4HA2 and INHBA expression.

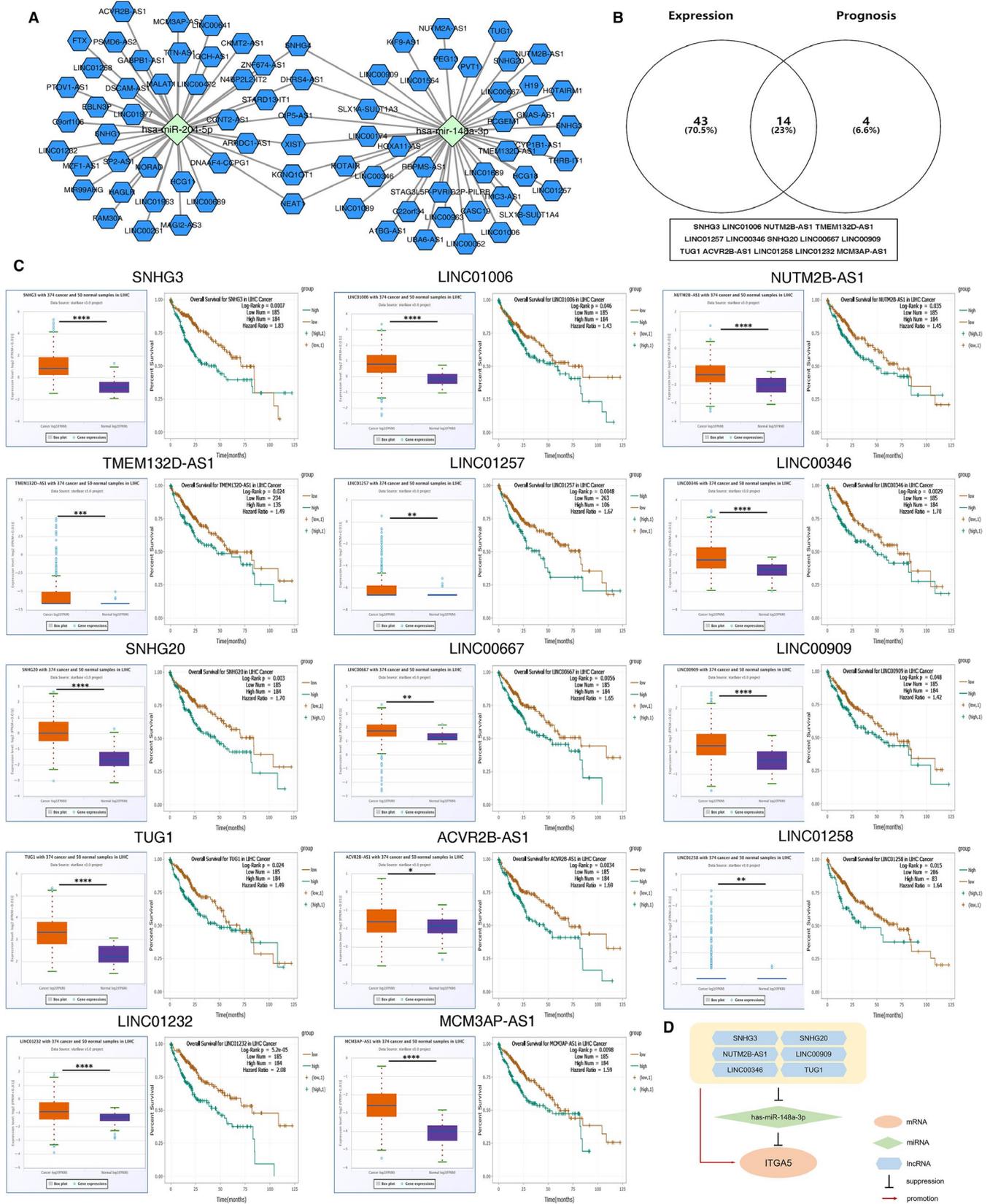


FIGURE 7 Prediction of key upstream lncRNAs. **A**, The miRNA-lncRNA network containing selected key upstream miRNAs. **B**, Identification of the key upstream lncRNAs with high expression and poor prognosis values in the predicted candidate lncRNAs. **C**, The expression and survival analysis of 14 selected candidate lncRNAs in HCC. **D**, The potential mRNA-miRNA-lncRNA regulatory subnetwork associated with prognosis of HCC

3.6 | Identification of key upstream miRNAs

Then, we sought to establish an mRNA-miRNA-lncRNA competing endogenous RNA subnetwork associated with p-EMT-related genes in HCC, among which each component was markedly correlated with HCC prognosis. In order to screen key miRNAs regulating these five key p-EMT-related genes, we predicted upstream miRNA of the five key genes by using miRTarBase, an experimentally validated miRNA-target interactions database. To improve the credibility of the predicted results, only miRNA-target interactions validated by reporter assays were included in our study. A total of 20 miRNAs were predicted to regulate the three key p-EMT-related genes (ITGA5, MMP9 and SPP1), all of which were up-regulated in HCC tissues (Figure 6A). There was no miRNA noticed to target P4HA2 and MT1X in this database. Given the inverse regulatory relationship between miRNAs and their target genes, we explored the expression and prognostic values of the predicted miRNA in HCC patients by using starBase v3.0. The result showed that seven miRNAs were significantly down-regulated in HCC tissues, and only two of them (has-miR-148a-3p and has-miR-204-5p) were associated with poor prognosis in HCC patients (Figure 6B). The expression boxplot and survival curve of the two candidate miRNAs were shown in Figure 6C, and they were selected as key miRNAs for next analysis.

3.7 | Identification of key upstream lncRNAs

It is well known that lncRNA can act as sponge to competitively bind to miRNA, regulating expression of target genes. Hence, we predicted key upstream lncRNAs that potentially bound to the two key miRNAs by miRNet, an online database for miRNA-associated studies. A total of 80 lncRNAs were discovered in the database for the two miRNAs has-miR-148a-3p and has-miR-204-5p (Figure 7A). The ceRNA hypothesis believes that lncRNA could weaken miRNA activity to up-regulate expression of miRNA-related target genes.¹³ Based on the hypothesis, there should be a negative correlation between lncRNA and miRNA. Thus, we analysed the expression patterns and survival curves of these lncRNAs in HCC using starBase v3.0. Compared with in normal tissues, only 57 out of 80 lncRNAs were dramatically up-regulated in HCC tissues. Survival analysis showed that patients with high expression of 18 lncRNAs had poor prognosis (Figure 6B). Combining expression and survival analysis results, 14 lncRNAs were screened as candidate lncRNAs for the ceRNA network (Figure 7B-C).

3.8 | Construction of the mRNA-miRNA-lncRNA subnetwork associated with prognosis in HCC

According to the ceRNA hypothesis, the key eligible lncRNAs in ceRNA subnetwork should be negatively correlated with miRNA and meanwhile positively correlated with mRNA. Thus, we analysed the correlation between the key p-EMT-related genes mRNA

and candidate miRNAs, miRNAs and candidate lncRNAs as well as mRNA and lncRNAs. As a result, one of two mRNA-miRNA pair (ITGA5- miR-148a-3p), six of 14 miRNA-lncRNA pairs (miR-148a-3p-SNHG3, miR-148a-3p-NUTM2B-AS1, miR-148a-3p-LINC00346, miR-148a-3p-SNHG20, miR-148a-3p-LINC00909 and miR-148a-3p-TUG1) and six of 14 mRNA-lncRNA pairs (ITGA5-SNHG3, ITGA5-NUTM2B-AS1, ITGA5-LINC00346, ITGA5-SNHG20, ITGA5-LINC00909 and ITGA5-TUG1) were fitted with the ceRNA mechanism (Supplementary Table S1 and Figures S5 and S6). Taken these three levels into consideration, we constructed an mRNA-miRNA-lncRNA subnetwork correlated with p-EMT-related genes, and each component of which had associations with prognosis of HCC (Figure 7D). The subnetwork might be developed as promising diagnostic biomarkers or therapeutic targets concerning p-EMT programme for HCC.

3.9 | Identification of the protein expression levels of key p-EMT-related genes in HCC

In addition to mRNA expression analysis, we further observed the protein expression levels of the five key p-EMT-related genes in HCC tissues and corresponding adjacent nontumour tissues using IHC assay. The results showed that most of the five key genes were mainly localized in the extracellular region, cytoplasm and cell membranes. Consistent with previous results, the protein expression levels of P4HA2, ITGA5, MMP9 and SPP1 were higher in HCC tissues than in adjacent nontumour tissues, while MT1X showed the opposite results in protein expression level. Representative IHC images and IHC score of the five key p-EMT-related genes were presented in Figure 8.

4 | DISCUSSION

Increasing evidence has identified the vital role of EMT in cancer progression, in which E-cadherin dysregulation was observed in many types of cancers,^{6,27} including HCC. In our study, we found that E-cadherin expression was lower in HCC tissues in few cohorts and no significant difference was observed in the others. Analysis of prognostic values of E-cadherin and its association with clinical features in HCC showed that E-cadherin may be not necessary for malignancy of HCC, indicating the EMT model may not be the most suitable for HCC progression.

During HCC progression, EMT represents the first step of HCC metastasis, where epithelial cells in the primary tumour lose cell-cell adhesion and gain the capacity of migration and invasion into blood vessels. Upon intravasation, these cells remain in the bloodstream until they colonize distant organs to seed micrometastasis. Then, these cells undergo a reversal of EMT-MET to restore their epithelial features and form secondary tumours, thereby completing the cascade of metastasis and invasion.²⁸ The intermediate p-EMT programme in the process has achieved more and more

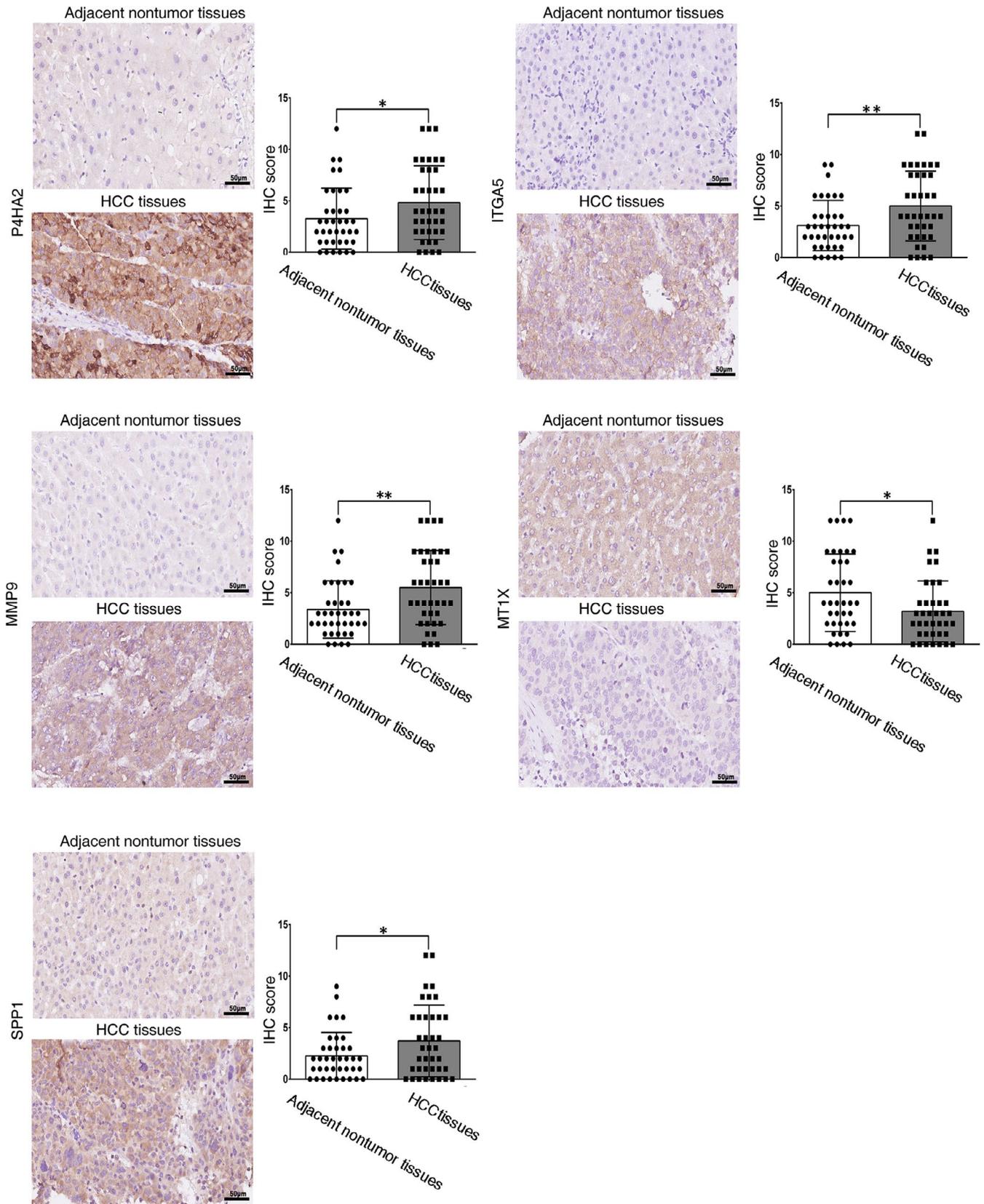


FIGURE 8 IHC assays of P4HA2, ITGA5, MMP9, MT1X and SPP1 in HCC tissues and adjacent nontumour tissues

attention because of its importance. Cancer cells undergoing p-EMT state function like cancer cells with mesenchymal state, yet they do not lose epithelial characteristics completely.²⁹ Recently, it has

been recognized that the ability of cancer cells with p-EMT state, rather than complete EMT, possesses higher metastatic risks.³⁰ In addition, p-EMT programme may be linked to functions of cancer

stem cells and progenitor cell, resulting in multiple cell phenotypes through self-renewal or asymmetric divisions.³¹ During this process of p-EMT, multiple relevant genes altered in HCC, which was implicated with HCC progression. Although several p-EMT-related genes have been confirmed to be potential targets of HCC, the distinct roles of the identified 25 p-EMT-related genes in HCC and regulatory ceRNA networks remain to be elucidated. This study is the first time to explore the expression, prognostic values, functional enrichment analysis of 25 p-EMT-related genes in HCC and identify five key p-EMT-related genes according to the criterion of converse expression and survival. Moreover, an mRNA-miRNA-lncRNA ceRNA network regulating the key p-EMT-related genes correlated with prognosis of HCC was also constructed. The workflow was shown in Figure S7. We hope that our findings will contribute to available knowledge about improvement of therapy strategy and prognosis accuracy for patients with HCC.

In this study, we explored the expression patterns and prognosis values of 25 p-EMT-related genes and identified five key p-EMT-related genes associated with prognosis of HCC, including two common p-EMT-related genes P4HA2, ITGA5 and three variable p-EMT-related genes MMP9, MT1X, SPP1. Genetic mutation analysis showed that amplification was the most common type of genetic mutations in HCC patients and the number of patients with RAB25 mutation was the highest. PPI network and correlation analysis of the 25 p-EMT-related genes suggested that most of these genes had close relationships with each other. Functional enrichment analysis revealed that the 25 p-EMT-related genes are enriched in malignant behaviours associated with cancers, including angiogenesis, cell adhesion, positive regulation of cell migration and proliferation. Indeed, previous studies have reported the involvement of these genes in malignancy. P4HA2, an essential enzyme during collagen formation, has been reported to be an indicator of many types of cancers progression. *In vitro* and *in vivo* experiments are used to clarify the mechanisms of aspirin in the suppression of HCC targeting abnormal collagen deposition associated with P4HA2³² and HBx is capable to promote hepatocarcinogenesis through miR-30e targeting P4HA2 mRNA.³³ ITGA5 belongs to the integrin alpha chain family and interacts with the beta 1 subunit to form a fibronectin receptor.³⁴ ITGA5 can be considered as a mesenchymal marker, overexpression of which is correlated with HCC proliferation, differentiation, invasion and metastasis.^{35,36} MMP9 has been considered as biomarker for prognosis of many types of solid tumours, including HCC.³⁷ Transforming growth factor (TGF)- β signalling plays an important role at the early stage of EMT programme in HCC development and MMP9 is a key target of the tumour-progressive TGF- β signalling, which might be a potential target for HCC therapy.³⁸ The protein encoded by SPP1 can be secreted into serum and is related with the attachment of osteoclasts to mineralized bone matrix.³⁹ SPP1 has been screened as a molecule for HCC diagnosis and prognosis.⁴⁰ Our results showed that the four p-EMT-related genes P4HA2, ITGA5, MMP9 and SPP1 were overexpressed in HCC tissues compared with in normal liver tissues, and the mRNA expression levels of them all were associated with prognosis of patients

with HCC. However, MT1X was identified as a tumour suppressor involved in HCC progression and metastasis.⁴¹ The expression and survival analysis showed that MT1X mRNA expression was higher in normal tissues and correlated with better prognosis of HCC patients.

Considering the importance of ceRNA networks in cancer, we constructed a novel ceRNA subnetwork associated with prognosis of HCC, containing the key p-EMT-related genes we selected, to explore the regulatory mechanisms of p-EMT-related genes. Therefore, candidate miRNAs and lncRNAs binding to potential miRNAs were subsequently predicted. 20 miRNAs of ITGA5, MMP9 and SPP1 were first identified using miRTarBase database. Given the action mechanism of miRNA on mRNA, we identified the ITGA5-miR-148a-3p pair for further analysis by performing correlation analysis for these mRNA-miRNA interactions in HCC using starBase v3.0. Low expression of miR-148a-3p was an independent risk factor for overall survival for HCC.⁴² *In vitro* experiment showed that miR-148a-3p inhibited the growth of HCV-infected HCC cells by targeting c-Jun mRNA.⁴³ Then, lncRNAs binding to miR-148a-3p were predicted by miRNet databases. Based on the ceRNA hypothesis, we predicted six upstream lncRNAs (SNHG3, NUTM2B-AS1, LINC00346, SNHG20, LINC00909 and TUG1), which were remarkably up-regulated in HCC and associated with poor prognosis of HCC. Most of these lncRNAs have been reported the oncogenic roles of lncRNAs in HCC. For example, SNHG3 promotes EMT and sorafenib resistance via activating miR-128/CD151 axis in HCC.⁴⁴ LINC00346 promotes HCC development via JAK-STAT3 signalling pathway activation.⁴⁵ Overexpression of SNHG20 has been reported to be an indicator of poor prognosis of HCC.⁴⁶ In addition, TUG1 is overexpressed in HCC and enhances cell proliferation and tumorigenicity by epigenetically inhibition of Kruppel-like factor 2 (KLF2) transcription.⁴⁷ In consequence, a novel mRNA-miRNA-lncRNA regulatory subnetwork associated with prognosis of HCC was constructed successfully. Interestingly, we found that some interactions in this network were identified in previous studies, which further confirmed the reliability of our results.⁴⁸ In the end, six pairs of subnetwork (SNHG3/SNHG20/NUTM2B-AS1/LINC00909/LINC00346/TUG1-miR-148a-3p-ITGA5) were acceptable, which might be utilized to be prognostic biomarkers for HCC.

We further identified the protein expression levels of the five key p-EMT-related genes in HCC tissues and corresponding adjacent nontumour tissues using IHC staining. Consistent with results above, the expression of P4HA2, ITGA5, MMP9 and SPP1 was higher in HCC tissues than in adjacent nontumour tissues in protein expression level. Nevertheless, the protein expression level of MT1X was lower in HCC tissues compared with nontumour liver tissues.

In conclusion, we systematically analysed the expression patterns, prognostic values, genetic mutations of p-EMT-related genes as well as their correlation with each other in HCC, and identified five key p-EMT-related genes in HCC. Moreover, a novel mRNA-miRNA-lncRNA ceRNA subnetwork containing p-EMT-related genes associated with HCC was constructed, each component of which possessed high prognostic value for HCC. More *in vitro* and *in vivo* experiments should be conducted to confirm the role of the subnetwork and p-EMT programme in HCC. Our

findings suggested potential therapeutic and prognostic values of p-EMT-related genes as well as their ceRNA regulatory subnetwork for the improvement of survival and prognostic accuracy in HCC.

ACKNOWLEDGEMENTS

The work was supported by grants from the National Natural Science Foundation of China No.81702396, No.81772607 and No.81800547.

CONFLICT OF INTEREST

The authors report no conflicts of interest in this work.

AUTHOR CONTRIBUTION

Yu Lei: Investigation (equal); Methodology (equal); Visualization (equal); Writing-original draft (equal). **Wei Yan:** Funding acquisition (supporting); Visualization (supporting). **Zhuoying Lin:** Methodology (supporting). **Jingmei Liu:** Methodology (supporting). **De-An Tian:** Funding acquisition (supporting); Visualization (lead). **Han Ping:** Funding acquisition (lead); Writing-original draft (lead); Writing-review & editing (lead).

ETHICS AND CONSENT APPROVAL

This study was approved by the Academic Committee of Tongji Medical College, Huazhong University of Science and Technology. Written informed consent from the patients was obtained.

DATA AVAILABILITY STATEMENT

All the data in our study can be accessed from the online database.

ORCID

Ping Han  <https://orcid.org/0000-0003-3736-6340>

REFERENCES

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68:394-424.
- Forner A, Reig M, Bruix J. Hepatocellular carcinoma. *Lancet*. 2018;391(10127):1301-1314.
- Diepenbruck M, Christofori G. Epithelial-mesenchymal transition (EMT) and metastasis: yes, no, maybe? *Curr Opin Cell Biol*. 2016;43:7-13.
- Simeone P, Trerotola M, Franck J, et al. The multiverse nature of epithelial to mesenchymal transition. *Semin Cancer Biol*. 2019;58:1-10.
- Ye X, Weinberg RA. Epithelial-mesenchymal plasticity: a central regulator of cancer progression. *Trends Cell Biol*. 2015;25(11):675-686.
- Nieto MA, Huang RY, Jackson RA, Thiery JP. EMT: 2016. *Cell*. 2016;166(1):21-45.
- Han P, Fu Y, Luo M, et al. BVES inhibition triggers epithelial-mesenchymal transition in human hepatocellular carcinoma. *Dig Dis Sci*. 2014;59(5):992-1000.
- Aiello NM, Kang Y. Context-dependent EMT programs in cancer metastasis. *J Exp Med*. 2019;216(5):1016-1026.
- Saxena K, Jolly MK, Balamurugan K. Hypoxia, partial EMT and collective migration: emerging culprits in metastasis. *Transl Oncol*. 2020;13(11):100845.
- Puram SV, Tirosh I, Parkh AS, et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell*. 2017;171(7):1611-1624.e24.
- Karaosmanoglu O, Banerjee S, Sivas H. Identification of biomarkers associated with partial epithelial to mesenchymal transition in the secretome of slug over-expressing hepatocellular carcinoma cells. *Cell Oncol*. 2018;41(4):439-453.
- Chen J, Zhu H, Liu Q, et al. DEPTOR induces a partial epithelial-to-mesenchymal transition and metastasis via autocrine TGFbeta1 signaling and is associated with poor prognosis in hepatocellular carcinoma. *J Exp Clin Cancer Res*. 2019;38(1):273.
- 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.
- Wang H, Huo X, Yang XR, et al. STAT3-mediated upregulation of lncRNA HOXD-AS1 as a ceRNA facilitates liver cancer metastasis by regulating SOX4. *Mol Cancer*. 2017;16(1):136.
- Zhang J, Lou W. A key mRNA-miRNA-lncRNA competing endogenous RNA triple sub-network linked to diagnosis and prognosis of hepatocellular carcinoma. *Front Oncol*. 2020;10:340.
- Lian Q, Wang S, Zhang G, et al. HCCDB: a database of hepatocellular carcinoma expression Atlas. *Genomics Proteomics Bioinformatics*. 2018;16(4):269-275.
- Chandrashekar DS, Bashel B, Balasubramanya SAH, et al. UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia*. 2017;19(8):649-658.
- Rhodes DR, Yu J, Shanker K, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia*. 2004;6(1):1-6.
- Roessler S, Jia HL, Budhu A, et al. A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. *Cancer Res*. 2010;70(24):10202-10212.
- Gyorffy B, Lanczky A, Eklund AC, et al. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res Treat*. 2010;123(3):725-731.
- Sherman BT, da Huang W, Tan Q, et al. DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis. *BMC Bioinformatics*. 2007;8:426.
- Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6(269):pl1.
- Szklarczyk D, Morris JH, Cook H, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res*. 2017;45(D1):D362-D368.
- Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498-2504.
- Chou CH, Shrestha S, Yang CD, et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res*. 2018;46(D1):D296-D302.
- Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res*. 2014;42(Database issue):D92-D97.
- Xin Y, Cummins B, Gedeon T. Multistability in the epithelial-mesenchymal transition network. *BMC Bioinformatics*. 2020;21(1):71.
- Jolly MK, Boareto M, Huang B, et al. Implications of the hybrid epithelial/mesenchymal phenotype in metastasis. *Front Oncol*. 2015;5:155.
- Kisoda S, Shao W, Fujiwara N, et al. Prognostic value of partial-EMT-related genes in head and neck squamous cell carcinoma by a bioinformatic analysis. *Oral Dis*. 2020;26:1149-1156.

30. Parikh AS, Puram SV, Faquin WC, et al. Immunohistochemical quantification of partial-EMT in oral cavity squamous cell carcinoma primary tumors is associated with nodal metastasis. *Oral Oncol*. 2019;99:104458.
31. Saitoh M. Involvement of partial EMT in cancer progression. *J Biochem*. 2018;164(4):257-264.
32. Wang T, Fu X, Jin T, et al. Aspirin targets P4HA2 through inhibiting NF-kappaB and LMCD1-AS1/let-7g to inhibit tumour growth and collagen deposition in hepatocellular carcinoma. *EBioMedicine*. 2019;45:168-180.
33. Feng GX, Li J, Yang Z, et al. Hepatitis B virus X protein promotes the development of liver fibrosis and hepatoma through downregulation of miR-30e targeting P4HA2 mRNA. *Oncogene*. 2017;36(50):6895-6905.
34. Juliano RL, Reddig P, Alahari S, Edin M, Howe A, Aplin A. Integrin regulation of cell signalling and motility. *Biochem Soc Trans*. 2004;32(Pt3):443-446.
35. Yao M, Zhou X, Zha X. Expression of the integrin alpha 5 subunit in hepatocellular carcinoma. *Zhonghua yi xue za zhi*. 1997;77(5):382-384.
36. Huang RY, Kuay KT, Tan TZ, et al. Functional relevance of a six mesenchymal gene signature in epithelial-mesenchymal transition (EMT) reversal by the triple angiokinase inhibitor, nintedanib (BIBF1120). *Oncotarget*. 2015;6(26):22098-22113.
37. Gobin E, Bagwell K, Wagner J, et al. A pan-cancer perspective of matrix metalloproteases (MMP) gene expression profile and their diagnostic/prognostic potential. *BMC Cancer*. 2019;19(1):581.
38. Reichl P, Dengler M, van Zijl F, et al. Axl activates autocrine transforming growth factor-beta signaling in hepatocellular carcinoma. *Hepatology*. 2015;61(3):930-941.
39. Hao C, Cui Y, Owen S, Li W, Cheng S, Jiang WG. Human osteopontin: potential clinical applications in cancer (Review). *Int J Mol Med*. 2017;39(6):1327-1337.
40. Long J, Zhang L, Wan X, et al. A four-gene-based prognostic model predicts overall survival in patients with hepatocellular carcinoma. *J Cell Mol Med*. 2018;22(12):5928-5938.
41. Liu Z, Ye Q, Wu L, et al. Metallothionein 1 family profiling identifies MT1X as a tumor suppressor involved in the progression and metastatic capacity of hepatocellular carcinoma. *Mol Carcinogenesis*. 2018;57(11):1435-1444.
42. Song SK, Jung WY, Park SK, Chung CW, Park Y. Significantly different expression levels of microRNAs associated with vascular invasion in hepatocellular carcinoma and their prognostic significance after surgical resection. *PLoS One*. 2019;14(9):e0216847.
43. Deng Y, Wang J, Huang M, Xu G, Wei W, Qin H. Inhibition of miR-148a-3p resists hepatocellular carcinoma progress of hepatitis C virus infection through suppressing c-Jun and MAPK pathway. *J Cell Mol Med*. 2019;23(2):1415-1426.
44. Zhang PF, Wang F, Wu J, et al. LncRNA SNHG3 induces EMT and sorafenib resistance by modulating the miR-128/CD151 pathway in hepatocellular carcinoma. *J Cell Physiol*. 2019;234(3):2788-2794.
45. Yin YZ, Zheng WH, Zhang X, Chen YH, Tuo YH. LINC00346 promotes hepatocellular carcinoma progression via activating the JAK-STAT3 signaling pathway. *J Cell Biochem*. 2020;121(1):735-742.
46. Zhang D, Cao C, Liu L, Wu D. Up-regulation of LncRNA SNHG20 predicts poor prognosis in hepatocellular carcinoma. *J Cancer*. 2016;7(5):608-617.
47. Huang MD, Chen WM, Qi FZ, et al. Long non-coding RNA TUG1 is up-regulated in hepatocellular carcinoma and promotes cell growth and apoptosis by epigenetically silencing of KLF2. *Mol Cancer*. 2015;14:165.
48. Tang Y, Yang P, Zhu Y, Su Y. LncRNA TUG1 contributes to ESCC progression via regulating miR-148a-3p/MCL-1/Wnt/beta-catenin axis in vitro. *Thorac Cancer*. 2020;11(1):82-94.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Lei Y, Yan W, Lin Z, Liu J, Tian D, Han P. Comprehensive analysis of partial epithelial mesenchymal transition-related genes in hepatocellular carcinoma. *J Cell Mol Med*. 2021;25:448-462. <https://doi.org/10.1111/jcmm.16099>