INTRODUCTION

Hepatocellular carcinoma (HCC) ranks sixth in the list of most commonly occurring solid cancers worldwide and ranks second in the list of most prevalent cause of death from fatal cancer.\(^1\) Hepatitis B or Hepatitis C Virus infection, alcohol drinking, and excessive smoking are the primary causes of HCC.\(^2,3\) Despite emerging evidence in the understanding of molecular mechanisms of HCC and improved therapies for HCC, the average survival time is still short. Regarding the
recent research, over 60% of initial detection of HCC patients in Japan is an early stage with an approximately 40% five-year survival rate and an average survival time of 50 months.6

In the past decade, progress in the genome-wide analysis of mammalian transcriptome has indicated a novel class of transcript, long noncoding RNAs (lncRNAs), which are broadly transcribed in the genome.7 LncRNAs are restricting splicing, while they are involved in RNA activation and stability of miRNA in the cytoplasm.7

Further evidence suggests that the aberrant expressions of lncRNAs have a clinical influence on the diagnosis and prognosis of HCC.8–10 Till now, IncRNA-associated biomarkers for diagnosis of HCC have been reported in many studies. Nevertheless, limited attempts have made to report the IncRNA signature as the prognostic biomarkers for HCC patients.

This study aims to capture the current research hotspots of HCC, based on the analysis of publications related to HCC research from 2013 to 2017, and to identify a novel IncRNA signature for HCC prognosis through the data mining in The Cancer Genome Atlas (TCGA) (http://cancergenome.nih.gov). Through constructing a comprehensive IncRNA expression analyses, we identified a new candidate indicator for the overall survival (OS) prediction in HCC patients.

2 | METHODS AND MATERIALS

2.1 | Source of the literature data and search strategy

Literature was searched from the Science Citation Index-Expanded (SCI-E) of Web of Science (WOS) of Clarivate Analytics on June 30, 2017. The data were collected from the public database, did not involve any interactions with human or animal subjects. Ethical approval was not applicable here.

All searches were conducted on the same day, June 30, 2017, to avoid the bias of daily updating of the database. The following terms were used in search: Title = (“liver cancer*”) OR Title = (“liver neoplasm*”) OR Title = (“Hepatocellular Cancer*”) OR Title = (“Hepatocellular carcinoma*”) OR Title = (“hepatic cancer*”) OR Title = (“hepatic neoplasm*”) OR Title = (“cancer of the liver”) OR Title = (“cancer of liver”) AND Language = English. In this case, only research articles and review articles were included.

2.2 | Literature data collection and analysis method

The data were independently collected from all eligible publications by two authors (Jing Sui and Yan Miao). The txt data were downloaded from WOS, and were imported into VOSviewer 1.6.5 (Leiden University, Leiden, Netherlands) and CiteSpace V (Drexel University, Philadelphia, PA, USA). The data were analyzed objectively. VOSviewer was performed to carry out the cluster analysis of the literature and the hotspot analysis of keywords.

2.3 | TCGA database and patient information

Three hundred and seventy-seven HCC patients’ data were downloaded from TCGA database (up to January 28, 2016). After exclusion criteria: (1) histologic diagnosis ruled out HCC; (2) another malignancy besides HCC. Overall, 317 HCC patients with corresponding clinical features such as race, age, gender, tumor stage, radiation therapy, and residual tumor were included in this study. Moreover, the endpoint in this study was OS. Of these above 317 HCC patients, there were 154 HCC patients with tumor stage I, 78 HCC patients with tumor stage II, 80 HCC patients with tumor stage III, and 5 HCC patients with tumor stage IV. As the data were retrieved from the public database (TCGA database), further ethical approval was not applicable in this study. Data processing procedures also met the policy of TCGA data and human subject protection (http://cancergenome.nih.gov/publications/publicationguidelines).

2.4 | RNA sequence data procession and IncRNA profile mining

The HCC RNA level 3 expression data were downloaded from TCGA database. All the IncRNA sequencing raw reads were postprocessed and normalized using TCGA RNASeqv2 system.11 In this study, IncRNAs with a description from NCBI (https://www.ncbi.nlm.nih.gov/gene/) and Ensemble (http://www.ensembl.org/index.html) would be selected for further study. To identify the differential expression of IncRNAs, patients were divided into HCC four tumor stages, including I, II, III, and IV to compare with adjacent nontumor lung tissues, respectively. The intersection of IncRNAs was selected in the further analysis (Figure 1).

2.5 | Construction of the IncRNA-based prognostic signature and Statistical analysis

The expression profile of each IncRNA was normalized by log2-transformed for further statistical analysis. However, the differently expressed IncRNAs that were 0 in more than 10% of all data were eliminated. The univariate Cox proportional hazards regression was used to evaluate the association between the differently expressed IncRNAs with OS of HCC patients (P-value <.05). Then, the multivariate Cox proportional hazards regression was used to identify the prognostic value of these independent IncRNA biomarkers.
Meanwhile, the prognostic IncRNA signature (the risk score model) was constructed based on a combination of the expression profiles of each prognostic IncRNAs, weighted by their estimated regression coefficients in the multivariate Cox proportional hazards regression analysis as follows:

\[
\text{risk score} = \exp\text{IncRNA}_1 \beta_{\text{IncRNA}_1} + \exp\text{IncRNA}_2 \beta_{\text{IncRNA}_2} + \cdots + \exp\text{IncRNA}_n \beta_{\text{IncRNA}_n}.
\]

The Kaplan-Meier survival curves were performed to present the difference in OS between high-risk score group and low-risk score group. The statistical significance was examined by the log-rank test. The univariate and multivariate Cox proportional hazards regressions for OS were conducted for individual clinical features with the IncRNA signature. The hazard ratio (HR) and 95% confidence intervals (CI) were calculated in this study. The prognostic performance at five years was accessed using time-dependent receiver operating characteristic (ROC) curves.

2.6 | Functional enrichment analysis

To investigate the biological feature of these above four IncRNAs in IncRNA signature, we identified the genes that highly correlated with these above four IncRNAs expression (Pearson |R| > 0.5) in TCGA database. Pathways and biological processes were predicted using functional enrichment analysis of Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) in the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (https://david.ncifcrf.gov/) Bioinformatics Resources 6.8. The P-value <.05 and FDR <0.05 were considered to be significant. Subsequently, the protein-protein interaction (PPI) network was constructed with the coexpressed genes via STRING (https://string-db.org/).

3 | RESULTS

3.1 | Cluster analysis and hotspot analysis on HCC research

A total of 1792 papers met the search criteria. These papers were analyzed by VOSviewer and divided into three clusters: “Patients Related Study,” “Expression Related Study,” and “Cell Related Study.” The cluster analysis demonstrated that the dominant fields of HCC include three research directions (Figure 2A).

Keywords used in the 1792 papers were extracted and analyzed by VOSviewer. As shown in Figure 2B, VOSviewer applied colors to keywords. The color of an item was determined by the frequency of occurrence, where by default colors range from blue (low frequency) to green (median frequency).
frequency) to red (high frequency). Keywords with high frequency were captured and considered as the hotspots in this field. From the literature analysis, we found hot keywords, including Hepatocellular carcinoma, prognosis, and biomarker. Thus, we confirmed that the current research hotspot of HCC is to identify a prognostic biomarker for HCC.

Furthermore, CiteSpace V was performed to capture the keywords with the most energetic citation bursts that identified as research frontiers over time. The top one research frontier of HCC research was “long noncoding RNA” (Figure 3). We realized a keyword “long noncoding RNA” appeared and grew rapidly. Considering this, our team determined the final research objective that was to discover a lncRNA-related prognostic biomarker for HCC. Based on this destination, we proceeded to the next step of lncRNA-related data mining. Here, we chose The Cancer Genome Atlas (TCGA) as a data source for both clinical information and bio-information.

3.2 | Patient characteristics

There were 317 HCC patients included in this study downloaded from TCGA dataset. Based on American Joint Committee on
Cancer (AJCC) TNM stage, the HCC patients were divided into stage I, stage II, stage III and stage IV, four groups. The age of all HCC patients was $58.019 \pm 13.509$ years. The OS time was $813.108 \pm 747.979$ days, 106 of 317 (33.438%) HCC patients died.

### 3.3 Identification of differentially expressed lncRNAs

We performed differential expression analysis by comparing the expression of 1081 lncRNAs in HCC and adjacent nontumor liver tissues. Fold change $>2$ or $<0.5$, $P$-value $<.05$ and FDR $<0.05$ were set up to identify significantly differentially expressed lncRNAs. Three hundred and seventeen differentially expressed lncRNAs were selected for further analysis, including 181 lncRNAs in stage I, 222 lncRNAs in stage II, 234 lncRNAs in stage III, and 165 lncRNAs in stage IV. We combined these four groups of 317 differentially expressed lncRNAs together, and 90 lncRNAs were identified stability

---

**FIGURE 3** The keywords with the strongest citation bursts of publications on Hepatocellular carcinoma research

**FIGURE 4** Venn diagram analysis of differentially expressed lncRNA in Hepatocellular carcinoma. Each ellipse represents a tumor stage group. The RNA in the middle represents significantly and consistently differentially expressed in four groups.
FIGURE 5  The differential expression of intersected lncRNAs in Hepatocellular carcinoma. A heatmap is showing the differentially expressed lncRNAs.
differentially expressed in all of the HCC tumor stages via two methods (Figures 4 and 5). The differentially expressed lncRNAs in different tumor stages were shown in Table S1.

3.4 Prognostic signature construction

Based on these 165 differentially expressed lncRNAs and clinical features in 317 HCC patients from TCGA database, 18 lncRNAs significantly associated with OS ($P < .05$) were identified by the univariate Cox regression model in Table 1. Afterward, the multivariate Cox proportional hazards regression analysis was used to calculate the interrelated relationship among 18 lncRNAs with OS, and only four lncRNAs exhibited a significant prognostic value for HCC, including LINC00261, TRELM3P, GBP1P1 and CDKN2B-AS1 (Table 2 and Figure 6).

The risk score for predicting prognostic value was constructed with the formula:

$$\text{Risk score} = \text{Exp}_{\text{LINC00261}} \times (-0.511) + \text{Exp}_{\text{TRELM3P}} \times (0.671) + \text{Exp}_{\text{GBP1P1}} \times (-0.554) + \text{Exp}_{\text{CDKN2B-AS1}} \times (0.447).$$

Based on the risk score model, HCC patients were classified as low-risk score or high-risk score patients via the median risk score as the cutoff value, which divided into the low-risk score group (n = 159) and high-risk score group (n = 158) (Figure 7). K-M curves confirmed that the survival time of patients in the low-risk score group was 929.698 ± 773.779 days, predominantly longer than that of the high-risk score group (695.032 ± 703.854 days, $P = .002$, Figure 8A). Furthermore, the risk score could largely predict the 5-year survival of

### Table 1 Prognostic value of the differentially expressed lncRNAs by univariate Cox regression analysis

| LncRNA     | Estimate | StdErr | ChiSq  | $P$ | HR (95%CI)     |
|------------|----------|--------|--------|-----|---------------|
| AADACP1    | −0.539   | 0.199  | 7.335  | .007* | 0.58 (0.395-0.862) |
| C3P1       | −0.451   | 0.197  | 5.237  | .022* | 0.637 (0.433-0.937) |
| CDKN2B-AS1 | 0.522    | 0.198  | 6.957  | .008* | 1.686 (1.144-2.486) |
| DHR54-AS1  | −0.536   | 0.198  | 7.343  | .007* | 0.585 (0.397-0.862) |
| FOXD2-AS1  | 0.459    | 0.197  | 5.449  | .020* | 1.583 (1.076-2.329) |
| GBP1P1     | −0.546   | 0.199  | 7.558  | .006* | 0.579 (0.392-0.855) |
| GOLGA2P    | 0.444    | 0.196  | 5.123  | .024* | 1.559 (1.061-2.290) |
| GVINP1     | −0.394   | 0.197  | 4.014  | .045* | 0.675 (0.459-0.992) |
| LINC00152  | 0.636    | 0.200  | 10.119 | .001* | 1.889 (1.277-2.796) |
| LINC00261  | −0.604   | 0.200  | 9.144  | .002* | 0.547 (0.370-0.809) |
| LINC01018  | −0.398   | 0.197  | 4.089  | .043* | 0.672 (0.457-0.988) |
| LINC01554  | −0.450   | 0.199  | 5.110  | .024* | 0.638 (0.432-0.942) |
| LOC645166  | 0.507    | 0.198  | 6.563  | .010* | 1.660 (1.126-2.445) |
| MAFG-AS1   | 0.423    | 0.196  | 4.645  | .031* | 1.526 (1.039-2.241) |
| MEIS3P1    | −0.474   | 0.197  | 5.820  | .016* | 1.522 (1.039-2.241) |
| PLGLA      | −0.497   | 0.200  | 6.202  | .013* | 0.608 (0.411-0.899) |
| TREML3P    | 0.795    | 0.203  | 15.314 | <.001* | 2.214 (1.487-3.296) |
| TSPEAR-AS2 | −0.498   | 0.197  | 6.394  | .011* | 0.608 (0.413-0.894) |

Bold font represents a statistically significant p-value.

* $P < .05$.

### Table 2 Prognostic value of the differentially expressed lncRNAs by multivariate Cox regression analysis

| LncRNA     | Estimate | StdErr | ChiSq  | $P$ | HR (95%CI)     |
|------------|----------|--------|--------|-----|---------------|
| LINC00261  | −0.511   | 0.203  | 6.332  | .012* | 0.600 (0.403-0.893) |
| TREML3P    | 0.671    | 0.206  | 10.638 | .001* | 1.957 (1.307-2.930) |
| GBP1P1     | −0.554   | 0.200  | 7.671  | .006* | 0.575 (0.388-0.850) |
| CDKN2B-AS1 | 0.447    | 0.200  | 5.005  | .025* | 1.564 (1.057-2.314) |

HR, hazard ratio; CI, confidence interval.

Bold font represents a statistically significant p-value

* $P < .05$.  

$\text{Risk score} = \text{Exp}_{\text{LINC00261}} \times (-0.511) + \text{Exp}_{\text{TREML3P}} \times (0.671) + \text{Exp}_{\text{GBP1P1}} \times (-0.554) + \text{Exp}_{\text{CDKN2B-AS1}} \times (0.447).$
FIGURE 6  Four differentially expressed lncRNAs (LINC00261, TREML3P, GBP1P1, and CDKN2B-AS1). A, Kaplan-Meier curves showing the relationship between these four lncRNAs and overall survival. The patients were divided into over- and underexpression groups by the mean lncRNAs level; B, ROC curves of the four lncRNAs to distinguish HCC tissue from adjacent normal tissues.
HCC patients, as the area under ROC curve (AUC) was 0.709 (Figure 8B).

The expression pattern of these four differentially expressed lncRNAs in the HCC and adjacent normal tissues, low-risk score and high-risk score groups were shown in Figure 9.

**3.5 Correlation between lncRNA signature and clinical characteristics**

We examined the association of four-lncRNA signature (risk score) with clinical features in HCC patients used the univariate and multivariate Cox proportional hazard regression...
analysis. The univariate Cox proportional hazards regression showed that gender, TNM stage, T stage, M stage, Neoplasm cancer (person neoplasm cancer status), BMI and history of Hepatocellular Carcinoma risk factors (Hist hepato carc fact) could predict poorer survival of HCC patients in Table 3 (\(P < .05\)). Meanwhile, the multivariate Cox proportional hazards regression showed Neoplasm cancer (\(P = .002\)) and risk score (\(P < .001\)) could predict as an independent prognostic indicator of HCC (Table 3).

In this study, the K-M curves of these clinical features were shown in Figure 10A. Moreover, it synthetically presented that the risk score conferred a prognostic value for predicting patients’ status of tumor stage (AUC = 0.603, \(P = .002\)) and Neoplasm cancer (AUC = 0.586, \(P = .001\)) (Figure 10B).

### 3.6 Functional assessment of the four-lncRNA signature

There were 626 genes identified in TCGA database coexpressed with these four lncRNAs (LINC00261, TREL3M3P, GBP1P1, and CDKN2B-AS1) (\(|R|>0.5\)), including 424 genes with LINC00261, 36 genes with TREL3M3P, 132 genes with GBP1P1, and 31 genes with CDKN2B-AS1, respectively (Table S2). It revealed enrichment of 628 GO Terms and 131 Pathways (\(P\)-value <.05 and an enrichment score of >1.5; Table S3). It was found that the top GO biological process of coexpressed genes was small molecule metabolic process (GO: 0044281) and cellular nitrogen compound metabolic process (GO: 0034641) (Table 4 and Figure 11A). After the pathway analysis, the coexpressed genes were mainly
| Variables                  | Patient N = 317 | Univariate analysis | Multivariate analysis |
|----------------------------|-----------------|---------------------|-----------------------|
|                            |                 | HR (95% CI)         | P                     | HR (95% CI)         | P                     |
| Race                       |                 |                     |                       |                     |
| Asian                      | 151             | 1 (reference)       |                       |                     |
| Black                      | 14              | 1.890 (0.746-4.793) | .180                  |                     |
| White                      | 141             | 1.138 (0.757-1.710) | .535                  |                     |
| Gender                     |                 |                     |                       |                     |
| Female                     | 99              | 1 (reference)       |                       | 1 (reference)       |                       |
| Male                       | 217             | 0.657 (0.445-0.969) | .034*                 | 1.354 (0.698-2.626) | .370                  |
| Age                        |                 |                     |                       |                     |
| <=55                       | 119             | 1 (reference)       |                       |                     |
| >55                        | 197             | 1.102 (0.739-1.644) | .634                  |                     |
| TNM stage                  |                 |                     |                       |                     |
| I                          | 154             | 1 (reference)       |                       | 1 (reference)       |                       |
| II                         | 77              | 1.339 (0.799-2.244) | .268                  | 1.636 (0.802-3.338) | .176                  |
| III                        | 80              | 2.592 (1.668-4.028) | <.001*                | 2.714 (1.467-5.019) | .001*                 |
| IV                         | 5               | 5.499 (1.689-17.901)| .005*                 |                     |
| T stage                    |                 |                     |                       |                     |
| T1                         | 156             | 1 (reference)       |                       | 1 (reference)       |                       |
| T2                         | 79              | 1.294 (0.774-2.163) | .325                  | 0.000 (0.000-1.640E58) | .908                  |
| T3                         | 71              | 2.461 (1.565-3.869) | <.001*                | 0.448 (0.051-3.955) | .470                  |
| T4                         | 10              | 5.040 (2.231-11.384)| <.001*                | 0.617 (0.061-6.210) | .682                  |
| N stage                    |                 |                     |                       |                     |
| N0                         | 243             | 1 (reference)       |                       |                     |
| N1                         | 1               | 0.049 (0.000-4.654E32) | .940                  |                     |
| M stage                    |                 |                     |                       |                     |
| M0                         | 248             | 1 (reference)       |                       |                     |
| M1                         | 4               | 3.960 (1.243-12.617)| .020*                 |                     |
| Radiation therapy          |                 |                     |                       |                     |
| No                         | 288             | 1 (reference)       |                       |                     |
| Yes                        | 8               | 1.074 (0.340-3.397) | .903                  |                     |
| Neoplasm cancer            |                 |                     |                       |                     |
| Tumor free                 | 174             | 1 (reference)       |                       | 1 (reference)       |                       |
| With tumor                 | 126             | 2.498 (1.643-3.798) | <.001*                | 2.432 (1.386-4.267) | .002*                 |
| Residual tumor             |                 |                     |                       |                     |
| R0                         | 280             | 1 (reference)       |                       |                     |
| R1 + R2                    | 11              | 1.038 (0.328-3.284) | .949                  |                     |
| Fibrosis ishak score       |                 |                     |                       |                     |
| No fibrosis                | 62              | 1 (reference)       |                       |                     |
| Portal fibrosis            | 28              | 0.861 (0.365-2.035) | .734                  |                     |
| Fibrous speta              | 24              | 0.896 (0.362-2.219) | .813                  |                     |
| Nodular formation and      | 8               | 0.841 (0.196-3.603) | .816                  |                     |
| incomplete cirrhosis       |                 |                     |                       |                     |
| Established cirrhosis      | 57              | 0.807 (0.420-1.552) | .521                  |                     |
enriched in Metabolic pathways and “Valine, leucine and isoleucine degradation” (Table 4 and Figure 11B). For the construction of the protein-protein interaction (PPI) network, there were 470 genes in the PPI network, which were regarded as hub genes (Figure 12).

### 4 | DISCUSSION

Hepatocellular carcinoma (HCC) is one of the deadliest malignancies with the high global mortality. Most HCC patients were diagnosed in the advanced stages of tumor progression (stage III and stage IV). However, HCC patients in the same stage might exhibit different prognosis outcome, owning to differences in various biomarkers, which are still being discovered. The novel biomarkers for early diagnosis, therapeutic process monitoring, and prognostic evaluation might increase the survival rate for HCC. Accumulating evidence suggested that IncRNAs might play major role in tumorigenesis, metastasis, development and the prognosis of HCC. The large-scale genome analyses have revealed the molecular characteristics associated with HCC OS. However, most studies focused on miRNA, mRNA, gene, and protein expression. With knowledge...
The functional role of lncRNAs in tumorigenesis and development also represents a significant untapped resource for HCC prognosis. In the present study, to identify lncRNAs significantly related to the OS of HCC, HCC data were analyzed on HCC patients TNM stage with clinical features from the TCGA database in groups. After the univariate and multivariate Cox proportional hazards regression, a total of four HCC OS-related lncRNAs were identified as significant prognostic value for HCC survival. Then, the signature (risk score) was set by combining these above four lncRNAs and found that this four-lncRNA signature could independently predict OS in HCC patients. The advantage of this study is a combination of clinical features and TCGA data to assess the survival of HCC patients by setting the lncRNA-related risk score.

Wang et al. also identified a four-lncRNA signature (RP11-322E11.5, RP11-150O12.3, AC093609.1, CTC-297N7.9) which might be an independent prognostic
biomarker for the prediction of HCC patient survival. However, compared with previous study, we used more stringent screening criteria. Firstly, we used different classification regarding the clinical information extracted from TCGA datasets. Secondly, we screened the lncRNAs which were not described in NCBI and Emsemble, the left lncRNAs were considered to have potential clinical significance for further validation. Then, the differently

| Category           | Term                                      | No. of genes | –lgP |
|--------------------|-------------------------------------------|--------------|------|
| Go term            | Small molecule metabolic process          | 134          | 87.035 |
|                    | Cellular nitrogen compound metabolic process | 32           | 27.871 |
|                    | Immune response                           | 39           | 26.456 |
|                    | Cellular lipid metabolic process           | 26           | 22.824 |
|                    | Xenobiotic metabolic process               | 25           | 22.295 |
|                    | Innate immune response                    | 36           | 16.620 |
|                    | T-cell receptor signaling pathway         | 17           | 15.850 |
|                    | Fatty acid beta-oxidation                 | 13           | 15.568 |
|                    | Signal transduction                       | 45           | 14.248 |
|                    | Bile acid metabolic process               | 12           | 14.211 |
|                    | Blood coagulation                         | 28           | 12.113 |
|                    | Defense response to virus                 | 17           | 11.978 |
|                    | T-cell costimulation                      | 13           | 11.930 |
|                    | Fatty acid metabolic process              | 12           | 11.644 |
|                    | Transmembrane transport                   | 29           | 11.348 |
|                    | Antigen processing via MHC class II       | 7            | 11.208 |
|                    | Interferon-gamma-mediated signaling pathway | 12           | 10.823 |
|                    | Epoxygenase P450 pathway                  | 7            | 10.773 |
|                    | Branched-chain amino acid catabolic process | 8            | 10.660 |
|                    | Drug metabolic process                    | 9            | 10.649 |

| KEGG pathways      | Term                                      | No. of genes | –lgP |
|--------------------|-------------------------------------------|--------------|------|
|                    | Metabolic pathways                        | 116          | 72.555 |
|                    | “Valine, leucine and isoleucine degradation” | 21           | 28.145 |
|                    | Fatty acid degradation                    | 18           | 23.249 |
|                    | Propanoate metabolism                     | 14           | 18.589 |
|                    | Antigen processing and presentation       | 18           | 18.187 |
|                    | Peroxisome                                | 18           | 17.552 |
|                    | Carbon metabolism                         | 19           | 16.314 |
|                    | PPAR signaling pathway                    | 16           | 16.109 |
|                    | Complement and coagulation cascades       | 16           | 16.109 |
|                    | Butanoate metabolism                      | 12           | 16.031 |
|                    | Influenza A                               | 20           | 13.807 |
|                    | Retinol metabolism                        | 14           | 13.615 |
|                    | Graft-versus-host disease                 | 12           | 13.432 |
|                    | Beta-Alanine metabolism                   | 11           | 13.376 |
|                    | Staphylococcus aureus infection           | 13           | 13.211 |
|                    | T-cell receptor signaling pathway         | 16           | 13.113 |
|                    | Fatty acid metabolism                     | 12           | 12.515 |
|                    | Systemic lupus erythematosus              | 17           | 12.501 |
|                    | Allograft rejection                       | 11           | 12.396 |
|                    | Herpes simplex infection                  | 19           | 12.311 |
expressed that were 0 in more than 10% of all data were eliminated. Finally, we used “FDR < 0.05 and P < .05” as the inclusion criteria. Therefore, the standards for bioinformatics analysis are more rigorous in our work, compared to the work in previous study. Thus, the number of candidate lncRNAs for further analyses is different in both studies. Other studies found novel biomarkers via different classification methods. Herein, it was reported in the

**FIGURE 11** Top 20 enrichment of GO terms and KEGG pathways for coexpressed mRNAs

**FIGURE 12** The map represents the protein-protein interaction (PPI) network of coexpressed genes
present study that expression of four novel lncRNAs could also become a novel independent prognostic signature for HCC patients.

Accumulating evidence has presented that a series of lncRNAs could act as tumor suppressors or oncogenes in HCC. However, the roles of most lncRNAs in HCC remain largely unknown. Hu et al. found overexpressed SVUGP2 could suppress cell proliferation and suppresses the invasion ability of HCC cell lines in vitro, and tumor growth in vivo. SchLAH was found downregulated in HCC with significantly correlated with shorter overall survival of HCC patients. Moreover, HOTAIR and HOTTIP were also upregulated in HCC indicating a poorer prognosis and reduced overall survival.

Among these above four lncRNAs in the risk score, decreased LINC00261 was identified associated with poor prognosis and metastasis in Gastric Cancer (GC). Moreover, LINC00261 was found related to cell growth, migration, cell proliferation, and cell apoptosis in endometriosis and choriocarcinoma. Furthermore, multivariate analyses revealed that expression of CDKN2B-AS1 could be an independent predictor for OS ($P = .036$) in GC. The other two lncRNAs (TRELM3P and GBP1P1) were not reported till now.

Moreover, we identified the genes that strongly related with these above four lncRNAs expression in HCC dataset from TCGA database. The relevant genes were mainly enriched in metabolic pathways, “Valine, leucine and isoleucine degradation,” cellular nitrogen compound metabolic process and small molecule metabolic process. However, there is no study as of yet investigated the biological and clinical function of those above four lncRNAs in HCC, there is still many research that needs to be accomplished.

These findings of the present study may have substantial clinical significance. However, the limitations should be taken into consideration in the present study. Firstly, only 1801 human lncRNAs were identified, which would be selected with a description from NCBI and Ensemble for further study. The prognostic-related lncRNAs identified here might not represent all the lncRNAs, which were potentially related to HCC OS. Secondly, the mean time of follow-up in the model was 813.108 days. Thus, the further study with the longer follow-up time is warranted. Thirdly, the role of these four lncRNAs in HCC is still unknown; in vivo and in vitro experiments should be investigated in the further study.

In conclusion, by synthetically analyzing the HCC lncRNA expression profiles in TCGA database, we identified a four-lncRNA signature, which could act as an indicator for HCC patient outcome and could be a potential independent biomarker for prognosis prediction of HCC. Future functional investigations are required to explore the mechanisms underlying the roles of these lncRNAs in HCC.

ACKNOWLEDGMENTS

The present study was supported by the National Natural Science Foundation of China (81673132 and 81472939), the Scientific Research Foundation of Graduate School of Southeast University (YBJJ1796) and the Foundation of Nanjing Medical University (2017NJMUZD140), Key Research and Development Project of Jiangsu Province (Social Development) (BE2015719, BE2017694). We thank Donglin Cheng who supports technical assistance in the project.

CONFLICT OF INTEREST

The authors declared that they had no competing interests.

ORCID

Jing Sui http://orcid.org/0000-0001-8223-9872
Siyi Xu http://orcid.org/0000-0002-5147-2644

REFERENCES

1. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015;136:E359‐E386.
2. El-Serag HB, Rudolph KL. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. Gastroenterology. 2007;132:2557-2576.
3. Crail KJ, Spittal PM, Patel SH, et al. Prevalence and incidence of hepatitis C virus infection among Aboriginal young people who use drugs: results from the Cedar Project. Open Med. 2009;3:e220-e227.
4. Gores GJ, Lieberman D. Good news-bad news: current status of GI cancers. Gastroenterology. 2016;151:13-16.
5. Carninci P, Kasukawa T, Katayama S, et al. The transciptional landscape of the mammalian genome. Science. 2005;309:1559-1563.
6. Kapranov P, Cheng J, Dike S, et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science. 2007;316:1484-1488.
7. Qi P, Du X. The long non-coding RNAs, a new cancer diagnostic and therapeutic gold mine. Mod Pathol. 2013;26:155-165.
8. Jiang X, Liu W. Long noncoding RNA highly upregulated in liver cancer activates p53-p21 pathway and promotes nasopharyngeal carcinoma cell growth. DNA Cell Biol. 2017;36:596-602.
9. Zhu XT, Yuan JH, Zhu TT, Li YY, Cheng XY. Long noncoding RNA glypican 3 (GPC3) antisense transcript 1 promotes hepatocellular carcinoma progression via epigenetically activating GPC3. FEBS J. 2016;283:3739-3754.
10. Shi F, Xiao F, Ding P, Qin H, Huang R. Long noncoding RNA highly up-regulated in liver cancer predicts unfavorable outcome and regulates metastasis by MMPs in triple-negative breast cancer. Arch Med Res. 2016;47:446-453.
11. Sui J, Li YH, Zhang YQ, et al. Integrated analysis of long non-coding RNAs associated ceRNA network reveals potential
IncRNA biomarkers in human lung adenocarcinoma. Int J Oncol. 2016;49:2023-2036.

12. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics. 2000;56:337-344.

13. Sho S, Court CM, Winograd P, Russell MM, Tomlinson JS. A prognostic mutation panel for predicting cancer recurrence in stages II and III colorectal cancer. J Surg Oncol. 2017;116:996-1004.

14. Martinez-Canales S, Cifuentes F, De Rodas L, et al. Transcriptomic immunologic signature associated with favorable clinical outcome in basal-like breast tumors. PLoS ONE. 2017;12:e0175128.

15. Liu Z, Borlak J, Tong W. Deciphering miRNA transcription factor feed-forward loops to identify drug repurposing candidates for cystic fibrosis. Genome Med. 2014;6:94.

16. Xu S, Ao J, Gu H, et al. IL-22 impedes the proliferation of hepatocellular carcinoma. Oncotarget. 2017;8:3018-3028.

17. Wu Y, Zheng S, Yao J, et al. Decreased expression of proteasome subunit alpha 8 is associated with poor prognosis in hepatocellular carcinoma. J Hepatol. 2014;61:840-849.

18. Miao R, Luo H, Zhou H, et al. Identification of prognostic biomarkers in hepatitis B virus-related hepatocellular carcinoma and stratification by integrative multi-omics analysis. J Hepatol. 2014;61:840-849.

19. Wang J, Wang H, Zhang Y, et al. Mutual inhibition between YAP1 and SRSF1 maintains long non-coding RNA Malat1-induced tumourigenesis in liver cancer. Cell Signal. 2014;26:1048-1059.

20. Fang TT, Sun XJ, Chen J, et al. Long non-coding RNAs are differentially expressed in hepatocellular carcinoma cell lines with differing metastatic potential. Asian Pac J Cancer Prev. 2014;15:10513-10524.

21. Xiong H, Li B, He J, Zeng Y, Zhang Y, He F. IncRNA HULC promotes the growth of hepatocellular carcinoma cells via stabilizing COX-2 protein. Biochem Biophys Res Commun. 2017;490:693-699.

22. Yang X, Xie X, Xiao YF, et al. The emergence of long non-coding RNAs in the tumorigenesis of hepatocellular carcinoma. Cancer Lett. 2015;360:119-124.

23. Wang J, Zhang SM, Wu JM, et al. Mastermind-like transcriptional coactivator 1 overexpression predicts poor prognosis in human with hepatocellular carcinoma. Ann Clin Lab Sci. 2016;46:502-507.

24. Shao Y, Gu W, Ning Z, Song X, Pei H, Jiang J. Evaluating the prognostic value of microRNA-203 in solid tumors based on a meta-analysis and the cancer genome atlas (TCGA) datasets. Cell Physiol Biochem. 2017;41:1468-1480.

25. Lin SB, Zhou L, Liang ZY, Zhou WX, Jin Y. Expression of GRK2 and IGF1R in hepatocellular carcinoma: clinicopathological and prognostic significance. J Clin Pathol. 2017;70:754-759.

26. Wang J, Zhou Y, Fei X, et al. Integrative bioinformatics analysis identifies ROBO1 as a potential therapeutic target modified by miR-218 in hepatocellular carcinoma. Oncotarget. 2017;8:61327-61337.

27. Hou G, Dong C, Dong Z, et al. Upregulate KIF4A enhances proliferation, invasion of hepatocellular carcinoma and indicates poor prognosis across human cancer types. Sci Rep. 2017;7:4148.

28. Qin Y, Xu SQ, Pan DB, et al. Silencing of WW2 inhibits adhesion, invasion, and migration in liver cancer cells. Tumour Biol. 2016;37:6787-6799.

29. Li Y, Jiang Z, Xu L, Yao H, Guo J, Ding X. Stability analysis of liver cancer-related microRNAs. Acta Biochim Biophys Sin (Shanghai). 2011;43:69-78.

30. Niu HX, Du T, Xu ZF, Zhang XK, Wang RG. Role of wild type p53 and double suicide genes in interventional therapy of liver cancer in rabbits. Acta Cir Bras. 2012;27:522-528.

31. Wang Z, Wu Q, Feng S, Zhao Y, Tao C. Identification of four prognostic LncRNAs for survival prediction of patients with hepatocellular carcinoma. PeerJ. 2017;5:e3575.

32. Hu J, Song C, Duan B, et al. LncRNA-SVUGP2 suppresses progression of hepatocellular carcinoma. Oncotarget. 2017;8:97835-97850.

33. Ge Z, Cheng Z, Yang X, et al. Long noncoding RNA SchLAH suppresses metastasis of hepatocellular carcinoma through interacting with fused in sarcoma. Cancer Sci. 2017;108:653-662.

34. Yang Z, Zhou L, Wu LM, et al. Overexpression of long non-coding RNA HOTAIR predicts tumor recurrence in hepatocellular carcinoma patients following liver transplantation. Ann Surg Oncol. 2011;18:1243-1250.

35. Ishibashi M, Kogo R, Shibata K, et al. Clinical significance of the expression of long non-coding RNA HOTAIR in primary hepatocellular carcinoma. Oncol Rep. 2013;29:946-950.

36. Quagliata L, Matter MS, Piscuoglio S, et al. Long noncoding RNA HOTIp/HOXA13 expression is associated with disease progression and predicts outcome in hepatocellular carcinoma patients. Hepatolology. 2014;59:911-923.

37. Fan Y, Wang YF, Su HF, et al. Decreased expression of the long noncoding RNA LINC00261 indicate poor prognosis in gastric cancer and suppress gastric cancer metastasis by affecting the epithelial-mesenchymal transition. J Hematol Oncol. 2016;9:57.

38. Sha L, Huang L, Luo X, et al. Long non-coding RNA LINC00261 inhibits cell growth and migration in endometriosis. J Obstet Gynaecol Res. 2017;43:1563-1569.

39. Wang Y, Xue K, Guan Y, et al. Long noncoding RNA LINC00261 inhibits cell growth and migration in endometriosis. J Obstet Gynaecol Res. 2017;43:1563-1569.

40. Wang Y, Xue K, Guan Y, et al. Long noncoding RNA LINC00261 inhibits cell growth and migration in endometriosis. J Obstet Gynaecol Res. 2017;43:1563-1569.

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section at the end of the article.