

Research Article

Identification of Long Noncoding RNA Biomarkers for Hepatocellular Carcinoma Using Single-Sample Networks

Xiaoqing Yu¹, Jingsong Zhang², Rui Yang¹ and Chun Li³

¹School of Sciences, Shanghai Institute of Technology, Shanghai 201418, China

²Key Laboratory of Systems Biology, State Key Laboratory of Cell Biology, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, Shanghai 200031, China

³School of Mathematics and Statistics, Hainan Normal University, Haikou 571158, China

Correspondence should be addressed to Xiaoqing Yu; xqyu@sit.edu.cn

Received 21 July 2020; Revised 19 October 2020; Accepted 29 October 2020; Published 16 November 2020

Academic Editor: Zhenguo Zhang

Copyright © 2020 Xiaoqing Yu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective. Many studies have found that long noncoding RNAs (lncRNAs) are differentially expressed in hepatocellular carcinoma (HCC) and closely associated with the occurrence and prognosis of HCC. Since patients with HCC are usually diagnosed in late stages, more effective biomarkers for early diagnosis and prognostic prediction are in urgent need. **Methods.** The RNA-seq data of liver hepatocellular carcinoma (LIHC) were downloaded from The Cancer Genome Atlas (TCGA). Differentially expressed lncRNAs and mRNAs were obtained using the edgeR package. The single-sample networks of the 371 tumor samples were constructed to identify the candidate lncRNA biomarkers. Univariate Cox regression analysis was performed to further select the potential lncRNA biomarkers. By multivariate Cox regression analysis, a 3-lncRNA-based risk score model was established on the training set. Then, the survival prediction ability of the 3-lncRNA-based risk score model was evaluated on the testing set and the entire set. Function enrichment analyses were performed using Metascape. **Results.** Three lncRNAs (RP11-150O12.3, RP11-187E13.1, and RP13-143G15.4) were identified as the potential lncRNA biomarkers for LIHC. The 3-lncRNA-based risk model had a good survival prediction ability for the patients with LIHC. Multivariate Cox regression analysis proved that the 3-lncRNA-based risk score was an independent predictor for the survival prediction of patients with LIHC. Function enrichment analysis indicated that the three lncRNAs may be associated with LIHC via their involvement in many known cancer-associated biological functions. **Conclusion.** This study could provide novel insights to identify lncRNA biomarkers for LIHC at a molecular network level.

1. Introduction

Hepatocellular carcinoma, one of the most common cancers worldwide, is the third leading cause of worldwide mortality for various cancers, and its incidence rate per year remains increasing rapidly [1–3]. The risk factors for HCC include infection with hepatitis B virus (HBV) or hepatitis C virus (HCV), aflatoxin B1 intake, alcohol consumption, nonalcoholic fatty liver disease, and some hereditary diseases [4–6]. Since patients with HCC are usually diagnosed at late stages, when medication is no longer effective, understanding the molecular mechanisms of HCC and identifying biomarkers for early diagnosis and treatment seem to be essential [7, 8].

Long noncoding RNAs (lncRNAs) are non-protein-coding transcripts longer than 200 nucleotides. According to the well-known central dogma of molecular biology, genetic information is stored in protein-coding genes [9, 10], for which noncoding RNA (ncRNAs) have been considered as “junk genes” or “transcriptional noise” for a long time [11]. However, with the development of both experimental technology and computational methods, an increasing number of lncRNAs have been discovered in human transcriptome. Over the past decades, several researches have shown that lncRNAs are involved in almost the whole life cycle of cells through different mechanisms, and they have played diverse and important roles in many fundamental

and critical biological processes, including transcriptional regulation, epigenetic regulation, organ or tissue development, cell differentiation and apoptosis, cell cycle control, metabolic processes, and chromosome dynamics [12–15]. Recently, several lncRNAs have been demonstrated to be associated with the development and survival in patients with different kinds of cancers, including HCC [16–18]. Moreover, many studies have highlighted the molecular mechanism and biological characters of lncRNAs in HCC occurrence and progression, and the result revealed that some lncRNAs can also serve as valuable prognostic predictors for HCC patients [19–21]. Despite precision medicine, which uses molecularly targeted therapy against malignant tumors and speeds up progress toward the discovery of novel molecular targets with the diagnostic and prognostic value [22], the management of patients with HCC remains problematic [23, 24]. Therefore, there is an urgent requirement to identify many more lncRNA biomarkers for HCC.

One key to achieving personalized medicine is to elucidate the molecular mechanisms of individual specific diseases, which generally result from the dysfunction of individual specific molecular network rather than the malfunction of single molecules. With rapid advances in high-throughput technologies, applying molecular networks to analyze human complex disease is attracting increasing wide attention [25, 26]. A molecular network, e.g., a gene regulatory network or a coexpression network, can be generally estimated by correlation coefficients of molecule pairs from expression or sequence data of multiple samples [27, 28]. In recent years, based on biological and clinical data, a number of network-based methods were proposed not only to identify disease modules and pathways but also to elucidate molecular mechanisms of disease development at the network level [29–31]. Many studies have shown that network-based biomarkers are superior to traditional single-molecule biomarkers for accurately characterizing disease states due to their additional information on interactions and networks [28, 30, 32, 33]. In particular, a single-sample network is considered to be reliable for accurately characterizing the specific disease state of an individual. It can be directly used to identify the biomarkers and further elucidate the molecular mechanisms of a disease for individual patients [29].

In this study, we aimed to identify the lncRNA biomarkers for the patients with LIHC based on the RNA-seq data from TCGA. By constructing the single-sample networks for the 371 tumor patients, we obtained three lncRNA biomarkers associated with overall survival (OS) of LIHC patients. Then, a 3-lncRNA-based risk score model was established on the training set, which could effectively predict the OS of LIHC patients. The independence and the predictive ability for survival prediction of the 3-lncRNA-based risk score were validated on the testing set and the entire set.

2. Materials and Methods

2.1. Dataset. By using GDC API (<https://gdc.cancer.gov/developers/gdc-application-programming-interface-api>), RNA sequencing data and clinical information from individuals with liver hepatocellular carcinoma (LIHC) were achieved

from TCGA (<https://portal.gdc.cancer.gov/projects/TCGA-LIHC>), which included 371 tumor samples and 50 normal samples.

2.2. Differential Expression Analysis. EdgeR is a Bioconductor package for differential expression analysis of replicate count data, which was widely used in the differential expression analysis of high-throughput sequencing data. In our study, we extracted the expression profile of mRNAs and lncRNAs from RNA-seq count data, which was normalized using the edgeR package (version 3.22.5). Those mRNAs and lncRNAs with zero expression value in more than 10% samples were discarded. The differentially expressed lncRNAs and mRNAs were calculated by edgeR at a threshold of $FDR < 0.05$ and $|\log_2(\text{fold change})| > 1$.

2.3. Construction of the Single-Sample Networks. In our study, the differentially expressed mRNAs and lncRNAs which had the same gene names were removed. As a result, 3329 differentially expressed mRNAs and 956 differentially expressed lncRNAs for each sample were left for further investigation. A single-sample network was constructed based on statistical perturbation analysis of a single test sample against a group of given reference samples, which can accurately characterize the disease state of an individual or a sample. The more detailed description about the single-sample network can be found in Ref. [34]. In our study, we took 50 normal samples as the reference samples, while 371 tumor samples were the test samples. Firstly, based on the gene expression data of the reference samples, a reference correlation network can be constructed by computing the Pearson correlation coefficient (PCC) between lncRNA-lncRNA and lncRNA-mRNA pairs, which was denoted as N_r . Then, adding a test sample s to the reference samples, another perturbed correlation network was obtained in the same way, which was denoted as N_p . By comparing the difference of the two correlation networks, we can get a single-sample network N_{ssn} for this test sample $s(N_{ssn} = |N_r - N_p|)$. Finally, 371 single-sample networks were obtained in our study.

For convenience, we transformed each single-sample network to an adjacency matrix ΔD , and the element $\Delta D_{i,j}$ represents the ΔPCC of the edge for a pair of molecules in the single-sample network. As the theoretical foundation for this method, if there were obvious differences between the reference samples and the single sample s in terms of the gene expression pattern, adding the tumor sample s to the reference samples would cause significant changes of the PCC on some edges in the perturbed network. We assumed that if a lncRNA might be a key biomarker, the sum of the ΔPCC of the edges linked by the lncRNA would be higher than others. Then, a vector SD was used to represent the sum of ΔPCC of the edges linked by each lncRNA in a single-sample network, which was denoted as

$$SD = (SD_1, SD_2, \dots, SD_i, \dots, SD_{956})^T, \quad (1)$$

where SD_i is the sum of ΔPCC of all the edges linked by the i th lncRNA in a single-sample network and can be calculated by

$$SD_i = \sum_j \Delta D_{ij} (i = 1, 2, \dots, 956, j = 1, 2, \dots, 4285). \quad (2)$$

Consequently, all the 371 single-sample networks can be represented by a matrix $M_{956 \times 371}$:

$$M = \begin{bmatrix} SD_{1,1} & SD_{1,2} & \dots & SD_{1,j} & \dots & SD_{1,371} \\ SD_{2,1} & SD_{2,2} & \dots & SD_{2,j} & \dots & SD_{2,371} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ SD_{i,1} & SD_{i,2} & \dots & SD_{i,j} & \dots & SD_{i,371} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ SD_{956,1} & SD_{956,2} & \dots & SD_{956,j} & \dots & SD_{956,371} \end{bmatrix}, \quad (3)$$

$$\begin{pmatrix} i = 1, 2, \dots, 956 \\ j = 1, 2, \dots, 371 \end{pmatrix},$$

where i denotes the i th lncRNA and j denotes the j th tumor sample.

Different from the method in Ref. [34], we designed a ranking system to identify the candidate lncRNA biomarker according to the matrix M . Firstly, each column in matrix M was sorted by the value size of SD_{ij} , and the items in the column returned to the corresponding lncRNA in matrix ML . Then, we calculated the frequency of each lncRNA in the top K rows of matrix ML and the top 5% lncRNAs were retained. Finally, in order to improve the effectiveness of our method, we took the intersection of the candidate lncRNAs under different K ($K = 5, 10, 20, 30$) as the final candidate lncRNA biomarkers by SSN. The flowchart of the ranking system is shown in Figure 1.

2.4. Survival Analysis. The differentially expressed lncRNAs, of which the expression level was zero that exceeded 10% of all subjects, were removed from the prognostic analysis. In the meantime, clinicopathological features, including survival information, were also achieved from TCGA. Samples without sufficient clinical data were omitted. Finally, the prognostic analysis included a total of 307 tumor samples with the expression data from 956 lncRNAs. The univariate and multivariate Cox regression analyses were conducted to evaluate the association between the variables and the OS of LIHC patients on the training set, the testing set, and the entire set, and statistical significance was assessed using $p < 0.05$. The Kaplan-Meier plots were employed to observe the prognosis effect. The ROC curve analysis was used to evaluate the prognosis performance for 5-year survival rate. All analyses were performed on the R3.4.3 framework.

2.5. Construction of the Risk Score Model. The 307 patients were randomly divided into two datasets. 154 patients were used as the training set to build a risk score model based on the candidate lncRNA biomarkers, while the other 153 patients were used as the testing set to evaluate the predictive

ability of the risk score model. Firstly, the multivariate Cox regression analysis was used to evaluate the association between the expression of the lncRNA biomarkers and the patient's OS in the training set. Then, a risk score model was built by linear combination of the expression levels of the lncRNA biomarkers weighted by their Cox regression coefficients. The calculation formula of the risk score model was as follows:

$$\text{Risk Score (RS)} = \sum_{i=1}^n (\text{Exp}_i \times \text{Coe}_i), \quad (4)$$

where i represents the total number of the lncRNA biomarkers, Exp indicates the expression profiles of lncRNA, and Coe is the estimated regression coefficient of the i th lncRNA derived from the multivariate Cox analysis. Based on this formula, each patient with LIHC had an RS and the median RS was treated as a cut-off point to divide the patients into the low-risk group and the high-risk group.

2.6. Pathway and Functional Enrichment Analysis. Metascape (<http://metascape.org/gp/index.html>) is a free gene annotation and analysis resource that helps biologists make sense of one or multiple gene lists [35]. To understand the function of the identified lncRNA biomarkers in our study, the Gene Ontology (Go) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was performed by Metascape. A value of $p < 0.01$ was set as the cut-off for significance.

3. Results

3.1. Differentially Expressed lncRNAs and mRNAs in LIHC. We downloaded the expression data of 60483 RNA, from which 14822 lncRNAs and 19814 mRNAs were obtained by the gene type data reported by the genome GRCh38.p13. By calculating with edgeR, a total of 956 lncRNAs and 3329 mRNAs were considered to be differentially expressed between tumor samples and normal samples. The volcano plot of the differentially expressed mRNAs and lncRNAs is shown in Figure 2.

3.2. The Candidate lncRNAs Identified by Single-Sample Networks. In our study, the top 48 lncRNAs (5% of 956 lncRNAs) that had the highest frequency of occurrence in all the 371 tumor samples remained as the candidate lncRNA biomarkers by the ranking system. Furthermore, by taking the intersection of the candidate lncRNAs identified under different K ($K = 5, 10, 20, 30$), 27 lncRNAs were obtained as the final candidate lncRNA biomarkers by the ranking system according to 371 single-sample networks. The detailed differential expression information of these lncRNAs is listed in Table S1.

3.3. Identification of Prognostic lncRNA Biomarkers. Hypothesized that the candidate lncRNA biomarkers selected by SSN might be involved in LIHC development and progression, we examined whether these lncRNAs were associated with the survival of patients with LIHC. The univariate

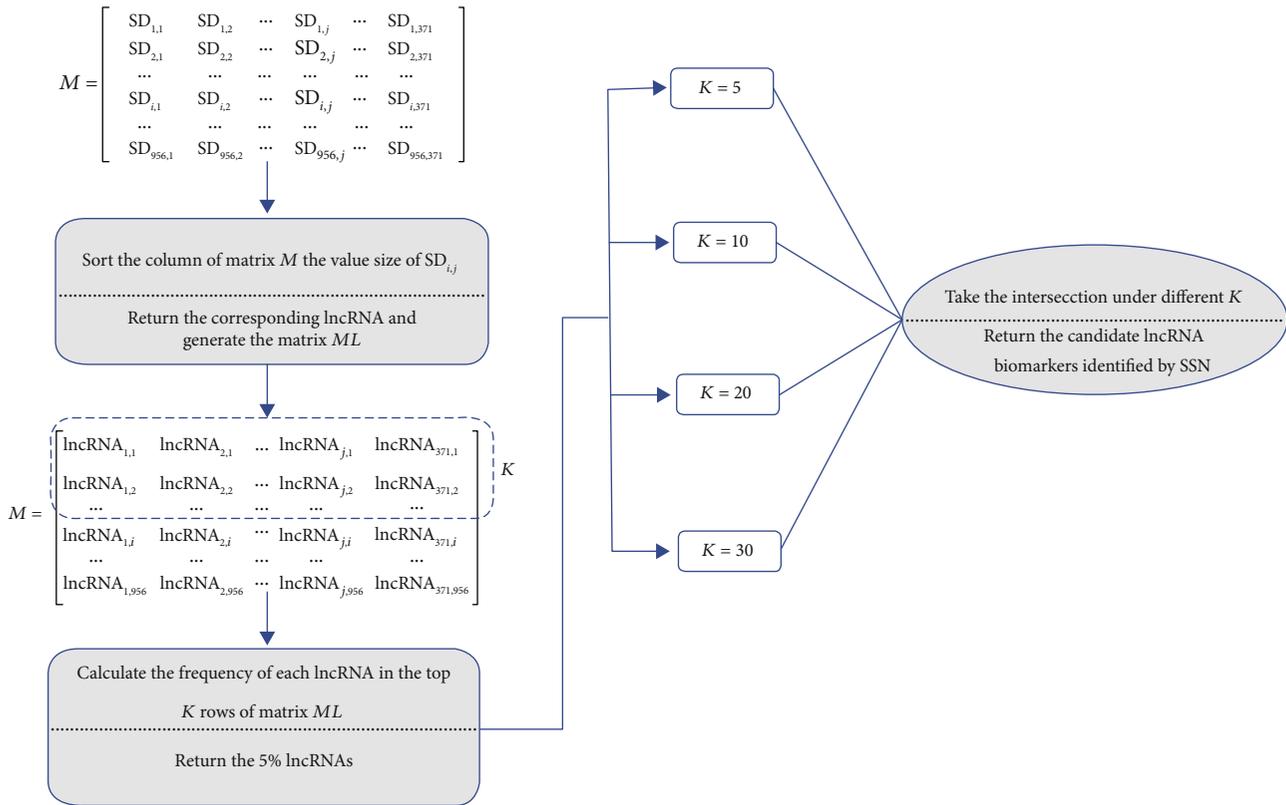


FIGURE 1: The ranking system for identifying the candidate lncRNA biomarkers by SSN.

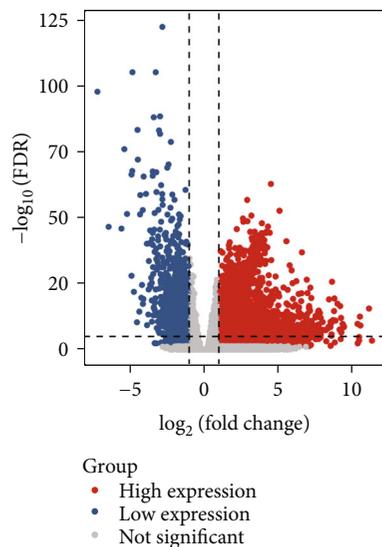


FIGURE 2: The volcano plot of the differentially expressed mRNAs and lncRNAs. Red indicates high expression, blue indicates low expression, and grey indicates nondifferential expression. The x-axis represents a log₂FC and the y-axis represents the -log₁₀(FDR).

survival analysis was performed based on the expression value of these lncRNAs and the OS of the LIHC patients. The follow-up period of these patients involved was 5 years (1825 days). According to the expression of the three lncRNA biomarkers, 307 patients were divided into high expression

group ($n = 154$) and low expression group ($n = 153$). The results showed that of all the 27 lncRNAs identified by SSN, three lncRNAs (RP11-150O12.3, RP11-187E13.1, and RP13-143G15.4) had significant prognostic differences ($p < 0.05$). The Kaplan-Meier plots employed to observe the prognosis effect of the three lncRNAs are presented in Figure 3. The detailed description of the three lncRNAs is shown in Table 1.

3.4. Construction of Risk Score Model in the Training Set. By using the multivariate Cox regression analysis in the training set, based on the expression value of the three lncRNAs and the regression coefficients, a 3-lncRNA-based risk score model was constructed as follows: risk score (RS) = 0.5324 * expression level of RP11 – 150O12.3 + 0.7105 * expression level of RP13 – 143G15.4 + 0.3738 * expression level of RP11 – 187E13.1. Based on the above risk score formula, the RS of each patient in the training set was calculated. The 154 patients in the training set were divided into two groups of low-risk group ($n = 77$) and high-risk group ($n = 77$) according to the median point of the risk score. The risk distribution and vital status of 154 patients and the expression heat map of the 3 lncRNAs are presented in Figure 4. The Kaplan-Meier survival analysis displayed that the survival of LIHC patients in the high-risk group was significantly shorter than that in the low-risk group (Figure 5(a), $p = 0.04$). The median survival of patients in the high-risk group and the low-risk group was 3.72 years and 4.89 years, respectively. To assess the prognostic performance of the risk

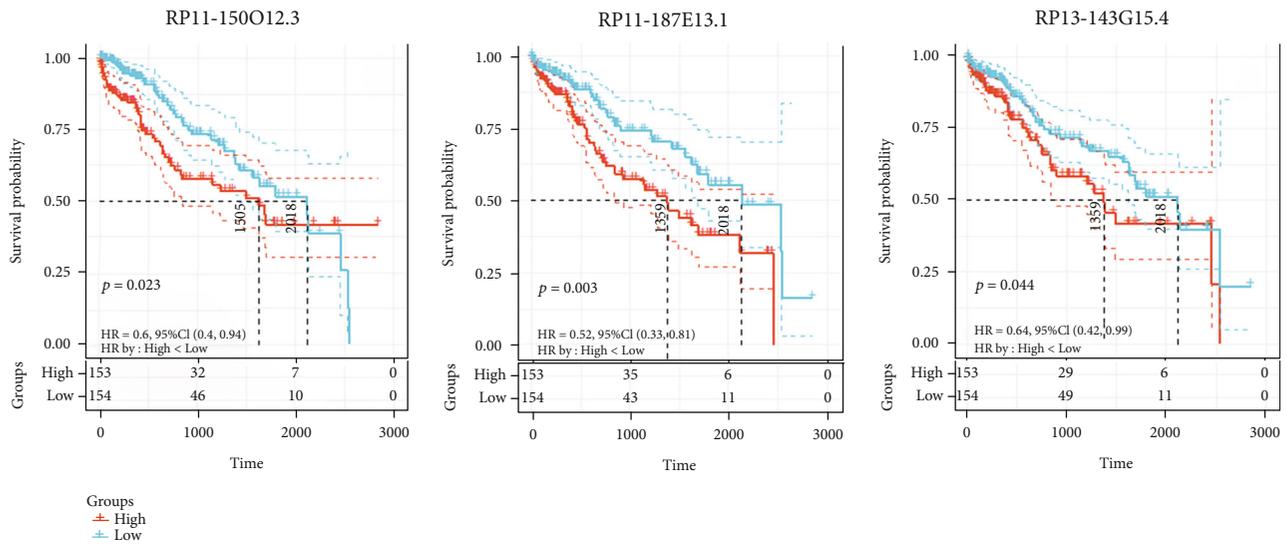


FIGURE 3: Kaplan-Meier curves of the survival probability of the three lncRNAs.

TABLE 1: The three lncRNAs significantly associated with OS of LIHC patients.

No.	Gene symbol	Ensemble ID	Chromosome	Hazard ratio	p value
1	RP11-150O12.3	ENSG00000254290.1	Chr.8:37,597,480-37,599,858	0.61 (0.4-0.94)	0.023
2	RP11-187E13.1	ENSG00000258474.1	Chr.14:31,944,853-31,950,382	0.52 (0.33-0.81)	0.003
3	RP13-14G15.4	ENSG00000237596.2	Chr.6:136415853-136546733	0.64 (0.42-0.99)	0.044

score, a 5-year ROC curve analysis was performed. The result showed that the 3-lncRNA-based risk score model had a good prediction efficiency in the training set (Figure 5(d), AUC = 0.706).

3.5. Evaluation of the Risk Score Model in the Testing Set and the Entire Set. To assess the robustness of the 3-lncRNA-based risk score model in OS prediction for LIHC patients, we further examined it in the testing set and the entire set. The same 3-lncRNA-based risk score model and cut-off point derived from the training set were used to divide the testing set and the entire set into the high-risk group and the low-risk group ($n = 77$ and 77 , $n = 154$ and 153 , respectively). In the testing set, the Kaplan-Meier curve showed that the survival of LIHC patients in the low-risk group exhibited a longer OS as compared to those in the high-risk group (Figure 5(b), $p = 0.0037$). The median survival of patients in the high-risk group and the low-risk group was 2.29 years and 6.69 years, respectively. In the entire set, a similar result was shown that patients in the high-risk group exhibited a shorter OS as compared to those in the low-risk group (Figure 5(c), $p = 0.0011$). The median survival of patients in the high-risk group and the low-risk group was 3.39 years and 5.52 years, respectively. The AUC of 5-year ROC curve in the testing set and the entire set was 0.704 and 0.664, implying a good prognostic capacity of the 3-lncRNA-based risk score (Figures 5(e) and 5(f)).

3.6. Independence of the 3-lncRNA-Based Risk Score for Survival Prediction. To evaluate the independent prediction

performance of the three lncRNAs when considering the traditional clinical factors, univariate and multivariate Cox regression analyses were conducted on the training set, the testing set, and the entire set. For each dataset, the explanatory variables included 3-lncRNA-based risk score, age, sex, weight, grade, and tumor stage. In the multivariate Cox regression analysis, the above factors were regarded as covariates, while the OS was used as a dependent variable. The results showed that the 3-lncRNA-based risk score was significantly associated with the OS of patients (HR = 0.416, 95% CI [0.207-0.836], $p = 0.014$ for the training set; HR = 0.459, 95% CI [0.225-0.936], $p = 0.032$ for the testing set; HR = 0.498, 95% CI [0.313-0.792], $p = 0.003$ for the entire set), which revealed that the 3-lncRNA-based risk score was an independent predictor of OS for LIHC patients. More detailed results are provided in Table 2.

3.7. Pathway and Functional Enrichment Analysis. We were also interested in the molecular mechanisms of the three lncRNAs. Unfortunately, little publication was found on the functional mechanism of these lncRNAs. Finally, we performed Pearson correlation analyses between the three lncRNAs and protein-coding genes based on their expression levels in TCGA LIHC cohort. The protein-coding genes that correlated with at least 1 of the three lncRNAs (Pearson coefficient > 0.45, $p < 0.01$) were considered to be significantly correlated coexpressed genes of the three lncRNAs, which are listed in Table 3. Then, pathway and process enrichment analyses were carried out with the following ontology sources in Metascape: KEGG pathways

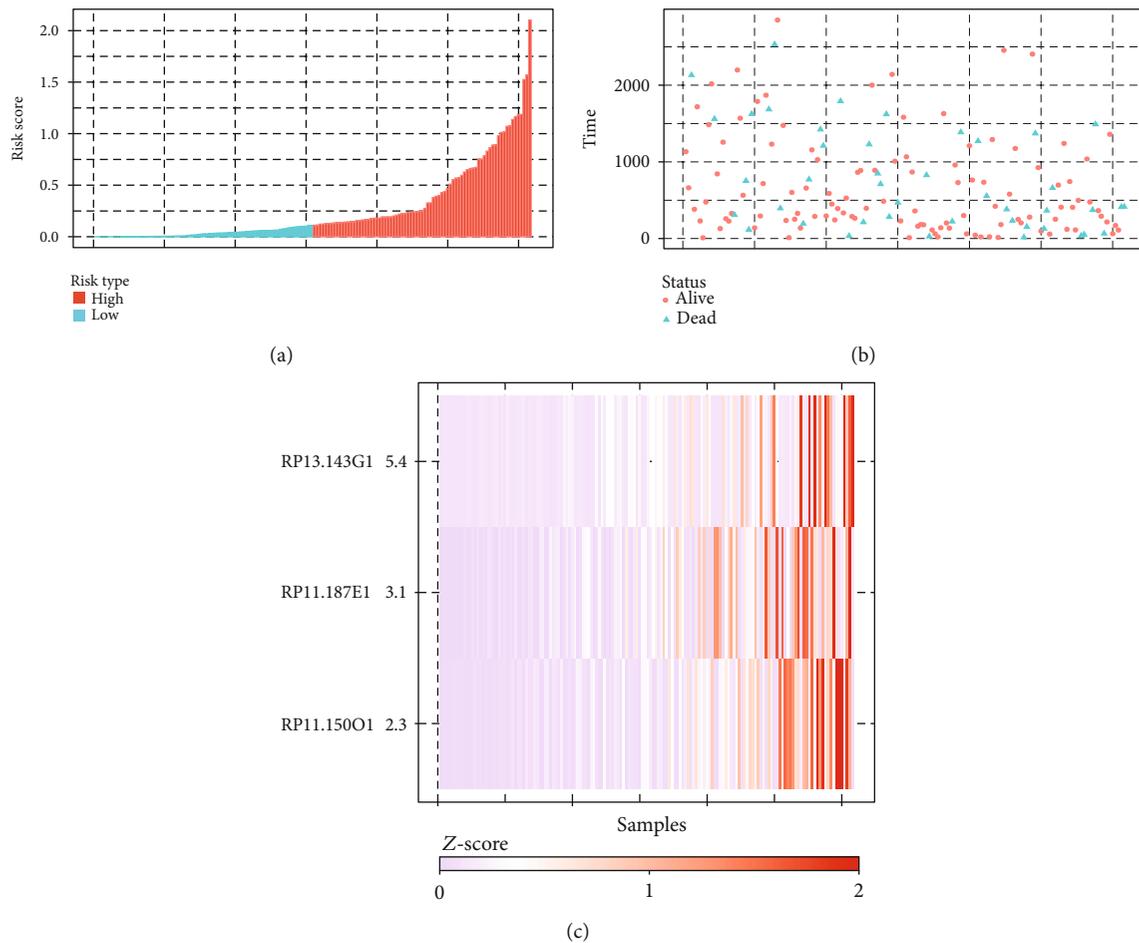


FIGURE 4: Construction of the 3-lncRNA-based risk score model. (a, b) The risk score distribution and survival status of the patients in the training set. (c) Heat map of the expression levels of the three lncRNAs.

and GO biological processes. All genes in the genome were used as the enrichment background. Terms with $p < 0.01$, minimum count = 3, and enrichment factor > 1.5 were collected and grouped into clusters based on membership similarities. We found that the coexpressed genes of three lncRNAs were mainly associated with terms related to endothelial cell migration, ncRNA metabolic process, cell cycle arrest, etc. (Figure 6). The retrieved data is also provided in Table S2.

4. Discussion

HCC is the most common type of liver cancer, accounting for about 80% of the tumours in this organ. Although there are advances in the diagnosis and treatment of HCC, the OS time of patients with HCC remains poor. Recently, molecular biomarkers have been identified throughout the various clinical stages and pathological tissue types, and multiple studies have revealed that differential expression of lncRNAs played critical roles in cancer development, indicating their potential as novel biomarkers for cancer diagnosis and prognosis [36, 37]. However, until now, only a few lncRNAs had been experimentally verified for HCC. Thus, it is necessary to identify more potential lncRNA biomarkers for HCC.

In the present study, we analyzed the expression profiles and clinical information of LIHC patients from TCGA database. The differentially expressed lncRNAs and mRNAs were screened to construct the single-sample networks for 371 LIHC patients. According to the 371 single-sample networks, 27 candidate lncRNAs were selected by the ranking system designed for identifying the lncRNA biomarkers from the single-sample networks. Furthermore, by using the univariate Cox regression, three lncRNAs (RP11-150O12.3, RP11-187E13.1, and RP13-143G15.4) were selected as the lncRNA biomarkers associated with patients' OS. Next, multivariate Cox regression was performed in the training set to establish a 3-lncRNA-based risk score model based on the expressions and relative contributions in the multivariate Cox regression model. The 3-lncRNA-based risk score had a good prognosis prediction ability for OS of patients with LIHC, which was tested in the testing set and the entire set. Moreover, we performed multivariate Cox regression analyses using the 3-lncRNA-based risk score and other clinical data in the training set, the testing set, and the entire set. The result showed that the 3-lncRNA-based risk score was an independent predictor for the OS of LIHC patients.

The results of function enrichment analysis showed that these KEGG pathways and functional categories of the three

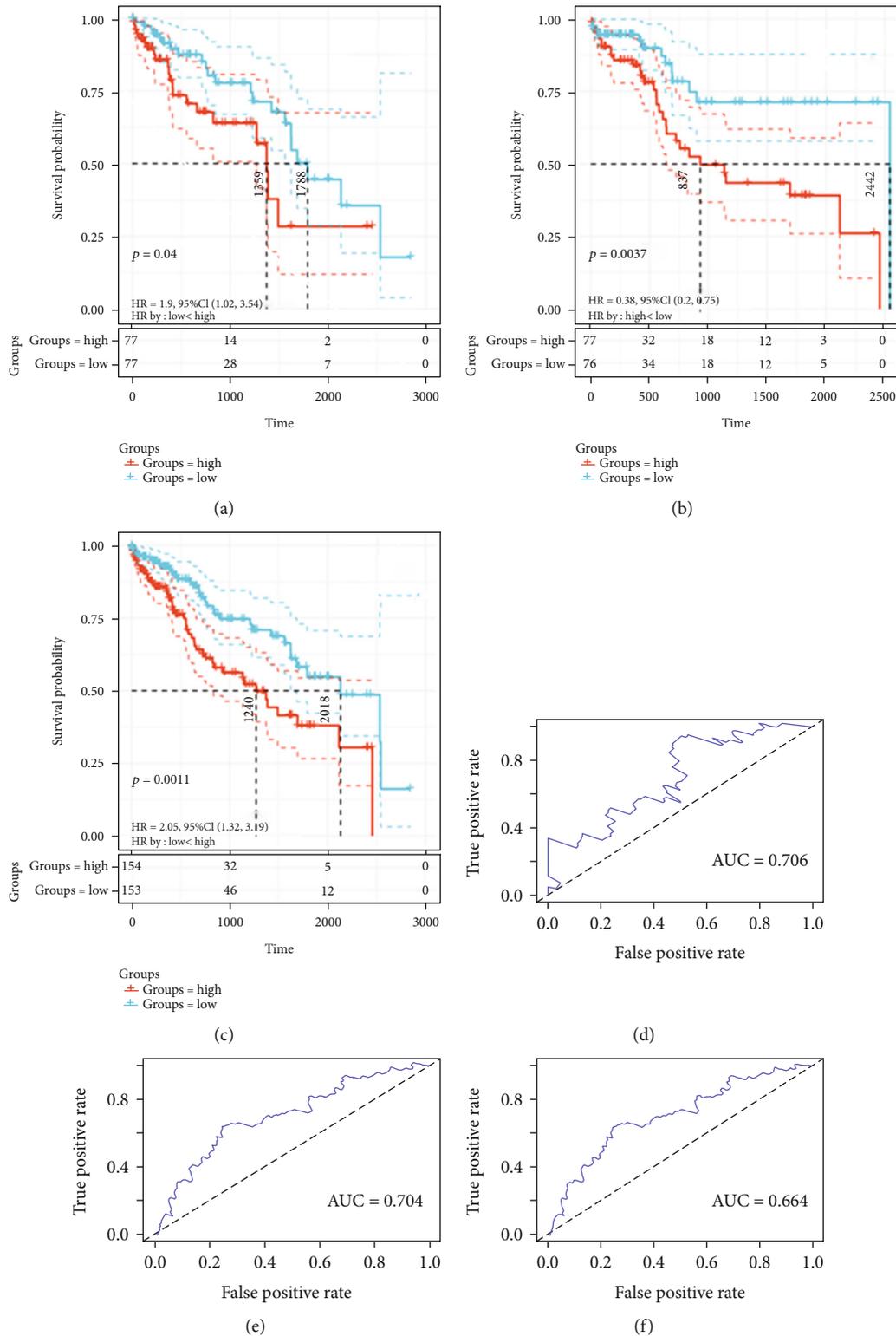


FIGURE 5: Kaplan-Meier survival analysis and ROC curve analysis of the 3-lncRNA-based risk score model. (a-c) Kaplan-Meier survival analysis in the training set, the testing set, and the entire set. (d-f) 5-year ROC curve analysis for survival prediction in the training set, the testing set, and the entire set.

lncRNA biomarkers were all closely associated with tumorigenesis. For instance, the transcription factor p53 is a tumor suppressor, activating downstream targets to trigger cell cycle

arrest and apoptosis [38], and several lncRNAs participate in the p53 regulatory network and serve as p53 regulators or effectors [39]. Moreover, a recent global transcriptome study

TABLE 2: Univariate and multivariate Cox proportional hazards regression analysis.

Variables	Univariate analysis			Multivariate analysis		
	HR	95% CI	<i>p</i> value	HR	95% CI	<i>p</i> value
Training set (<i>n</i> = 154)						
^a Risk score (high vs. low)	0.526	0.282-0.980	0.043	0.416	0.207-0.836	0.014
Age	1.031	1.006-1.058	0.015	1.029	0.998-1.059	0.061
Sex (male vs. female)	0.851	0.462-1.566	0.605	1.077	0.519-2.236	0.841
Weight	0.999	0.983-1.016	0.939	0.997	0.981-1.014	0.772
Grade (I+II vs. III+IV)	0.138	0.312-0.607	0.009	0.352	0.032-1.524	0.290
Stage (I+III vs. II+IV)	0.379	0.143-1.000	0.050	0.730	0.189-2.830	0.670
Testing set (<i>n</i> = 153)						
^a Risk score (high vs. low)	0.384	0.197-0.751	0.003	0.459	0.225-0.936	0.032
Age	1.025	0.998-1.053	0.069	1.023	0.994-1.052	0.114
Sex (male vs. female)	0.605	0.325-1.126	0.113	0.711	0.361-1.397	0.322
Weight	0.999	0.983-1.017	0.971	0.999	0.983-1.016	0.979
Grade (I+II vs. III+IV)	1.242	0.166-9.313	0.833	0.510	0.180-1.441	0.996
Stage (I+III vs. II+IV)	0.478	0.204-1.123	0.090	0.675	0.220-2.068	0.204
Entire set (<i>n</i> = 307)						
^a Risk score (high vs. low)	0.487	0.313-0.756	0.001	0.498	0.313-0.792	0.003
Age	1.028	1.010-1.047	0.002	1.023	1.004-1.043	0.017
Sex (male vs. female)	0.715	0.463-1.105	0.131	0.811	0.506-1.299	0.384
Weight	0.999	0.988-1.011	0.930	0.999	0.988-1.011	0.966
Grade (I+II vs. III+IV)	0.436	0.134-1.418	0.168	0.582	0.116-2.910	0.510
Stage (I+III vs. II+IV)	0.427	0.226-0.804	0.008	0.570	0.256-1.269	0.169

Notes. ^aDerived from the 3-lncRNA-based risk score model.

TABLE 3: The coexpressed genes of three lncRNAs.

Key lncRNA	Coexpressed gene
RP11-150O12.3	CLK1, BAG3, MXD1, GRIK1, TNFSF9, DNAJA4, FAM83A, FOXL1, PPP1R15A, SNORD14E, HSPA6, FOXC2
RP13-143G15.4	ZNF703, TMCC1, NKD1, RNF43, CYB5B
RP11-187E13.1	UFSP1, HN1, C7orf61, LSM8, AP1S1, ELOB, TSR3, PSMG3, COPS9, COPS6, ZNF593, LOC101927420, CLDND2, ST7-AS1, C6orf52, LOC106660606, B9D1, FIS1, PRSS3, S100A13, S100A16, NSUN5, LMNA, LAGE3, ACD, SNHG19, HSPB1, TMEM54, LINC00896, TMSB10, SFN, BRI3, EIF3B, HRAS, MRPL28, MIR6132, NT5C, PTC1, RASSF1-AS1, MINCR, LOC105371849, METTL26, DPM3, BCL7C, NUDT1, AP1M2, MCM7, BOLA2B, JOSD2, LOC400684, PUSL1, ARF4-AS1, UQCC3, POLR2J, NAA10, LOC101928659, PGP, S100P, POP7, LAMTOR4

identified that sixteen p53 target lncRNAs forming a pathway web constitute tumor suppressor signature with high diagnostic power [40]. Importantly, Jelena et al. revealed that p53 might play multifunctional roles in different stages in HCC [41]. Moreover, the metastatic characteristics of liver cancer are the key factors affecting the survival and prognosis of tumor patients, and the process of cell migration is involved in tumor metastasis [42]. The COP9 signalosome is a highly conserved protein complex implicated in diverse biological functions that involve ubiquitin-mediated proteolysis [43]. Also, research demonstrates that COP9 signalosome is an important regulator of cell cycle and cell survival mediating the proliferation of HCC cells and highlight that COP9 signalosome might be a promising strategy for anti-HCC therapy [44]. The unfolded protein response

is a signal transduction cascade, which acts as a quality control mechanism for protein synthesis and consequently can act to protect cells from an adverse external microenvironment. The study declared that the unfolded protein response is activated in the majority of HCC and may play a role in chemoresistance to the most widely used chemotherapy agent, doxorubicin, and have effects on newer antiestrogen and multikinase inhibitor therapies [45]. The present results indicated that the three lncRNAs might have an important role in LIHC via their involvement in these known cancer-associated biological functions.

Besides, previous studies have demonstrated that RP11-150O12.3 was significantly and independently associated with survival of HCC patients [46]. Moreover, RP11-150O12.3 also has been recognized as an independent

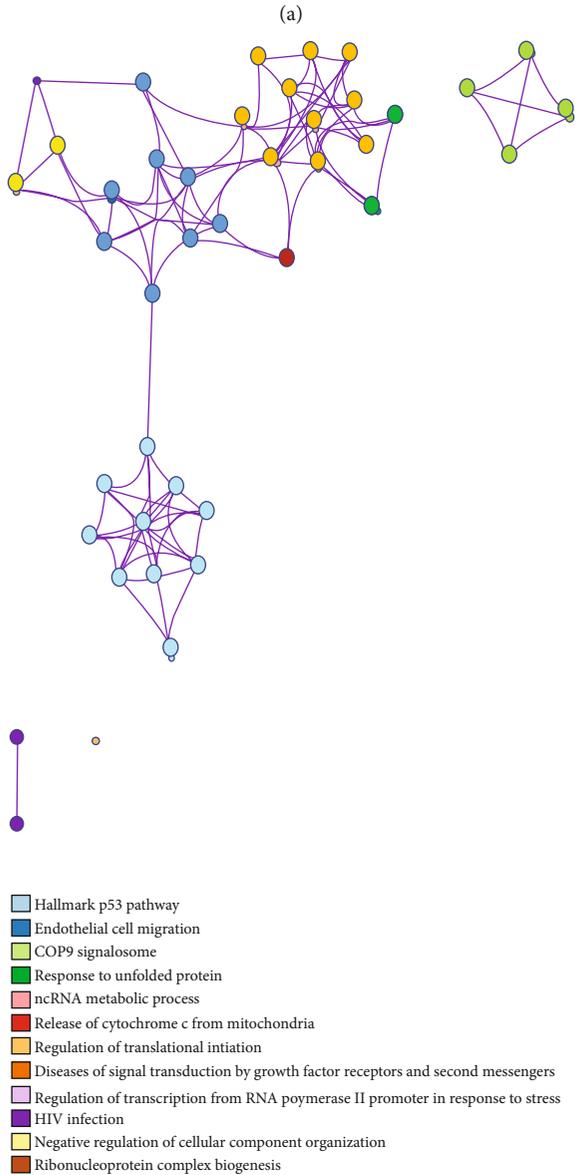
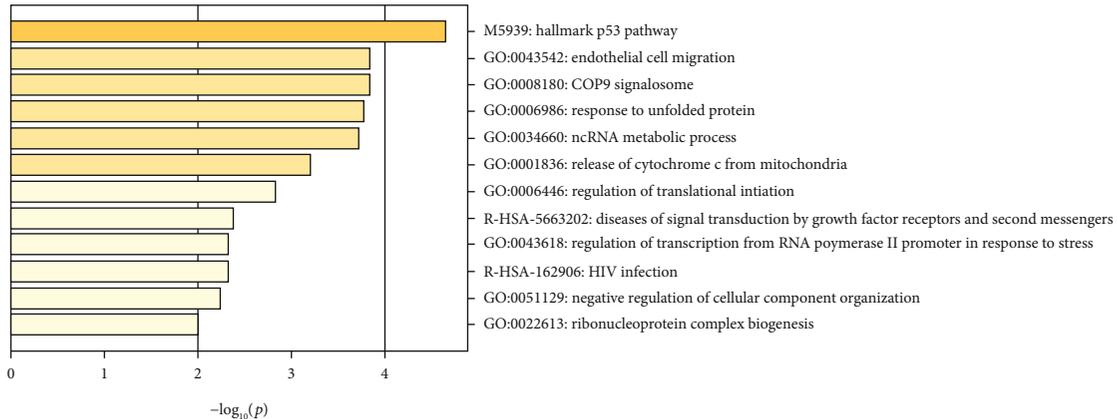


FIGURE 6: Pathway and process enrichment analysis. (a) KEGG functional analysis of the coexpressed genes of the three lncRNAs. (b) Detailed net structure of the coexpressed genes of the three lncRNAs.

predictor of gastric cancer prognosis [47], which also related to survival time of patients with colorectal adenocarcinoma [48]. RP13-143G15.3 is abnormally expressed in liver cancer and may act as an important role of tumor suppressor in human HCC [49]. The results of these studies provided some grounding to validate our findings.

In our study, the single-sample networks were constructed for identifying the lncRNA biomarkers associated with LIHC patients. Actually, a single-sample network in this study is not a real molecular network for each sample but a perturbation network for a single sample against the reference network. It mainly reflects the variation between normal and disease samples in terms of interactions or a network. Similarly, a differential expression of a gene is not the real gene expression level for each sample but the variation of the gene expression between normal and disease samples. The advantage of the single-sample network is that it cannot only characterize the personalized features for all the single samples but also can be directly applied to the data analysis of single samples at the molecular network level, which is superior to traditional single molecular biomarker.

There are several limitations to the present study. First, the primary purpose of the present study was to identify lncRNA biomarker for LIHC at the molecular network level. The lncRNA biomarkers selected in our study should be further investigated incorporating more specific clinical characteristics to fully understand the associations involved. Second, the three lncRNA biomarkers were only validated in the datasets from TCGA database, which required additional confirmation in other large cohorts of LIHC patients in the future. Third, the lncRNA biomarkers identified in our study have no validation in fresh samples and experiment studies.

5. Conclusion

In this study, we constructed the single-sample networks for LIHC patients and identified three lncRNA biomarkers associated with the OS of LIHC patients. A 3-lncRNA-based risk score model with the ability to effectively predict the survival of LIHC patients was established, which was validated to be an independent predictor for the OS of LIHC patients. Functional enrichment analysis revealed that three lncRNAs may be associated with LIHC via their involvement in many known cancer-associated biological functions. This study provides a new perspective for identifying the lncRNA biomarkers at the molecular network level.

Abbreviations

LIHC:	Liver hepatocellular carcinoma
HCC:	Hepatocellular carcinoma
HBV:	Hepatitis B virus
HCV:	Hepatitis C virus
lncRNA:	Long noncoding RNA
OS:	Overall survival
PCCs:	Pearson correlation coefficients
SSN:	Single-sample network
Go:	Gene Ontology
KEGG:	Kyoto Encyclopedia of Genes and Genomes.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 11701379, No. 11801374, and No. 61602460) and Hainan Province Natural Science Foundation (No. 2019RC184).

Supplementary Materials

Supplementary 1. Table S1: list of 27 differentially expressed lncRNAs in LIHC.

Supplementary 2. Table S2: the results of GO and KEGG enriched terms.

References

- [1] A. P. Venook, C. Papandreou, J. Furuse, and L. Ladrón de Guevara, "The incidence and epidemiology of hepatocellular carcinoma: a global and regional perspective," *Oncologist*, vol. 15, Supplement 4, pp. 5–13, 2010.
- [2] Y. R. Liu, R. X. Tang, W. T. Huang et al., "Long noncoding RNAs in hepatocellular carcinoma: novel insights into their mechanism," *World Journal of Hepatology*, vol. 7, no. 28, pp. 2781–2791, 2015.
- [3] K.-H. Bai, S. Y. He, L. L. Shu et al., "Identification of cancer stem cell characteristics in liver hepatocellular carcinoma by WGCNA analysis of transcriptome stemness index," *Cancer Medicine*, vol. 9, no. 12, pp. 4290–4298, 2020.
- [4] H. B. Elserag, "Epidemiology of viral hepatitis and hepatocellular carcinoma," *Gastroenterology*, vol. 142, no. 6, pp. 1264–1273.e1, 2012.
- [5] C. H. Su, Y. Lin, and L. Cai, "Genetic factors, viral infection, other factors and liver cancer: an update on current progress," *Asian Pacific Journal of Cancer Prevention Apjcp*, vol. 14, no. 9, pp. 4953–4960, 2013.
- [6] K. A. Mcglynn, J. L. Petrick, and W. T. London, "Global epidemiology of hepatocellular carcinoma: an emphasis on demographic and regional variability," *Clinics in Liver Disease*, vol. 19, no. 2, pp. 223–238, 2015.
- [7] K. Zhu, Z. Dai, and J. Zhou, "Biomarkers for hepatocellular carcinoma: progression in early diagnosis, prognosis, and personalized therapy," *Biomarker Research*, vol. 1, no. 1, p. 10, 2013.
- [8] C. H. Wang, K. C. Wey, L. R. Mo, K. K. Chang, R. C. Lin, and J. J. Kuo, "Current trends and recent advances in diagnosis, therapy, and prevention of hepatocellular carcinoma," *Asian Pacific Journal of Cancer Prevention Apjcp*, vol. 16, no. 9, pp. 3595–3604, 2015.
- [9] F. H. C. Crick, L. Barnett, S. Brenner, and R. J. Watts-Tobin, "General nature of the genetic code for proteins," *Nature*, vol. 192, no. 4809, pp. 1227–1232, 1961.

- [10] J. E. Wilusz, H. Sunwoo, and D. L. Spector, "Long noncoding RNAs: functional surprises from the RNA world," *Genes & Development*, vol. 23, no. 13, pp. 1494–1504, 2009.
- [11] A. F. Palazzo and E. S. Lee, "Non-coding RNA: what is functional and what is junk?," *Frontiers in Genetics*, vol. 6, p. 2, 2015.
- [12] International Human Genome Sequencing Consortium, "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, no. 6822, pp. 860–921, 2001.
- [13] T. R. Mercer, M. E. Dinger, and J. S. Mattick, "Long non-coding RNAs: insights into functions," *Nature Reviews Genetics*, vol. 10, no. 3, pp. 155–159, 2009.
- [14] V. Mohanty, Y. Gokmen-Polar, S. Badve, and S. C. Janga, "Role of lncRNAs in health and disease—size and shape matter," *Briefings in Functional Genomics*, vol. 14, no. 2, pp. 115–129, 2015.
- [15] X. Chen, C. C. Yan, X. Zhang, and Z. H. You, "Long non-coding RNAs and complex diseases: from experimental results to computational models," *Briefings in Bioinformatics*, vol. 18, no. 4, article bbw060, 2016.
- [16] M. Klingenberg, A. Matsuda, S. Diederichs, and T. Patel, "Non-coding RNA in hepatocellular carcinoma: mechanisms, biomarkers and therapeutic targets," *Journal of Hepatology*, vol. 67, no. 3, pp. 603–618, 2017.
- [17] D. Shi, Q. Qu, Q. Chang, Y. Wang, Y. Gui, and D. Dong, "A five-long non-coding RNA signature to improve prognosis prediction of clear cell renal cell carcinoma," *Oncotarget*, vol. 8, no. 35, pp. 58699–58708, 2017.
- [18] M. Zhou, Z. Zhang, H. Zhao, S. Bao, L. Cheng, and J. Sun, "An immune-related six-lncRNA signature to improve prognosis prediction of glioblastoma multiforme," *Molecular Neurobiology*, vol. 55, no. 5, pp. 3684–3697, 2018.
- [19] Z. Chen, H. Lan, Y. Wu, W. Zhai, P. Zhu, and Y. Gao, "LncSox4 promotes the self-renewal of liver tumour-initiating cells through Stat3-mediated Sox4 expression," *Nature Communications*, vol. 7, no. 1, article 12598, 2016.
- [20] T. Li, J. Xie, C. Shen et al., "Upregulation of long noncoding RNA ZEB1-AS1 promotes tumor metastasis and predicts poor prognosis in hepatocellular carcinoma," *Oncogene*, vol. 35, no. 12, pp. 1575–1584, 2016.
- [21] J.-Y. Zhang, M. Z. Weng, F. B. Song et al., "Long noncoding RNA AFAP1-AS1 indicates a poor prognosis of hepatocellular carcinoma and promotes cell proliferation and invasion via upregulation of the RhoA/Rac2 signaling," *International Journal of Oncology*, vol. 48, no. 4, pp. 1590–1598, 2016.
- [22] L. Li, Q. Lei, S. Zhang, L. Kong, and B. Qin, "Screening and identification of key biomarkers in hepatocellular carcinoma: evidence from bioinformatic analysis," *Oncology Reports*, vol. 38, no. 5, pp. 2607–2618, 2017.
- [23] C.-L. Liu and S.-T. Fan, "Nonresectional therapies for hepatocellular carcinoma," *American Journal of Surgery*, vol. 173, no. 4, pp. 358–365, 1997.
- [24] J. M. Llovet, A. Villanueva, A. Lachenmayer, and R. S. Finn, "Advances in targeted therapies for hepatocellular carcinoma in the genomic era," *Nature Reviews Clinical Oncology*, vol. 12, no. 7, pp. 408–424, 2015.
- [25] L. N. Chen, R. S. Wang, and X. S. Zhang, *Biomolecular Network: Methods and Applications in Systems Biology*, John Wiley & Sons, Inc., Hoboken, 2009.
- [26] A. L. Barabási, N. Gulbahce, and J. Loscalzo, "Network medicine: a network-based approach to human disease," *Nature Reviews Genetics*, vol. 12, no. 1, pp. 56–68, 2011.
- [27] T. Ideker and N. J. Krogan, "Differential network biology," *Molecular Systems Biology*, vol. 8, no. 1, p. 565, 2012.
- [28] X. P. Liu, Z. P. Liu, X.-M. Zhao, and L. Chen, "Identifying disease genes and module biomarkers by differential interactions," *Journal of the American Medical Informatics Association Jamia*, vol. 19, no. 2, pp. 241–248, 2012.
- [29] X. Liu and X. Chang, "Identifying module biomarkers from gastric cancer by differential correlation network," *Oncotargets & Therapy*, vol. 9, pp. 5701–5711, 2016.
- [30] W. Zhang, T. Zeng, and L. Chen, "EdgeMarker: identifying differentially correlated molecule pairs as edge-biomarkers," *Journal of Theoretical Biology*, vol. 362, pp. 35–43, 2014.
- [31] Z. Fang, W. Tian, and H. Ji, "A network-based gene-weighting approach for pathway analysis," *Cell Research*, vol. 22, no. 3, pp. 565–580, 2012.
- [32] W. Zhang, T. Zeng, X. Liu, and L. Chen, "Diagnosing phenotypes of single-sample individuals by edge biomarkers," *Journal of Molecular Cell Biology*, vol. 7, no. 3, pp. 231–241, 2015.
- [33] R. Liu, X. Wang, K. Aihara, and L. Chen, "Early diagnosis of complex diseases by molecular biomarkers, network biomarkers, and dynamical network biomarkers," *Medicinal Research Reviews*, vol. 34, no. 3, pp. 455–478, 2014.
- [34] X. Liu, Y. Wang, H. Ji, K. Aihara, and L. Chen, "Personalized characterization of diseases using sample-specific networks," *Nucleic Acids Research*, vol. 44, no. 22, pp. e164–e164, 2016.
- [35] Y. Zhou, B. Zhou, L. Pache et al., "Metascape provides a biologist-oriented resource for the analysis of systems-level datasets," *Nature Communications*, vol. 10, no. 1, article 1523, 2019.
- [36] Y. Ren, S. Zhao, D. Jiang et al., "Proteomic biomarkers for lung cancer progression," *Biomarkers in Medicine*, vol. 12, no. 3, pp. 205–215, 2018.
- [37] M. J. Hangauer, I. W. Vaughn, and M. T. Mcmanus, "Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs," *PLoS Genetics*, vol. 9, no. 6, article e1003569, 2013.
- [38] Y. Sasaki, H. Negishi, M. Idogawa et al., "p53 negatively regulates the hepatoma growth factor HDGF," *Cancer Research*, vol. 71, no. 22, pp. 7038–7047, 2011.
- [39] D. J. Voce, G. M. Bernal, L. Wu et al., "Temozolomide treatment induces lncRNA MALAT1 in an NF- κ B and p53 codependent manner in Glioblastoma," *Cancer Research*, vol. 79, no. 10, pp. 2536–2548, 2019.
- [40] Y. Sánchez, V. Segura, O. Marín-Béjar et al., "Genome-wide analysis of the human p53. transcriptional network unveils a lncRNA tumour suppressor signature," *Nature Communications*, vol. 5, no. 1, article 5812, 2014.
- [41] J. Krstic, M. Gallhuber, T. Schulz, M. Schupp, and A. Prokesch, "p53 as a dichotomous regulator of liver disease: the dose makes the medicine," *International Journal of Molecular Sciences*, vol. 19, no. 3, p. 921, 2018.
- [42] H. B. El-Serag, "Hepatocellular carcinoma: an epidemiologic view," *Journal of Clinical Gastroenterology*, vol. 35, no. 5, Supplement 2, pp. S72–S78, 2002.
- [43] D. A. Wolf, C. Zhou, and S. Wee, "The COP9 signalosome: an assembly and maintenance platform for cullin ubiquitin ligases?," *Nature Cell Biology*, vol. 5, no. 12, pp. 1029–1033, 2003.
- [44] Y. S. Yu, Z. H. Tang, Q. C. Pan, X. H. Chen, X. N. Liu, and G. Q. Zang, "Inhibition of Csn3 expression induces growth arrest and apoptosis of hepatocellular carcinoma cells," *Cancer*

- Chemotherapy and Pharmacology*, vol. 69, no. 5, pp. 1173–1180, 2012.
- [45] F. Y. al-Rawashdeh, P. Scriven, I. C. Cameron, P. V. Vergani, and L. Wyld, “Unfolded protein response activation contributes to chemoresistance in hepatocellular carcinoma,” *European Journal of Gastroenterology & Hepatology*, vol. 22, no. 9, pp. 1099–1105, 2010.
- [46] Z. Wang, Q. Wu, S. Feng, Y. Zhao, and C. Tao, “Identification of four prognostic lncRNAs for survival prediction of patients with hepatocellular carcinoma,” *PeerJ*, vol. 5, no. 7, article e3575, 2017.
- [47] W. Ren, J. Zhang, W. Li et al., “A tumor-specific prognostic long non-coding RNA signature in gastric cancer,” *Medical Science Monitor*, vol. 22, pp. 3647–3657, 2016.
- [48] J. H. Zeng, L. Liang, R. Q. He et al., “Comprehensive investigation of a novel differentially expressed lncRNA expression profile signature to assess the survival of patients with colorectal adenocarcinoma,” *Oncotarget*, vol. 8, no. 10, pp. 16811–16828, 2017.
- [49] Q. Liao, X. Y. Yang, and X. Q. Qiu, “Expression of long-coding RNA RP13-143G15.3 in human hepatocellular carcinoma and its clinical significance,” *Chinese Journal of Cancer Prevention and Treatment*, vol. 24, no. 12, pp. 829–834, 2017.