Research Article

A Novel Four-Gene Signature as a Potential Prognostic Biomarker for Hepatocellular Carcinoma

Guangfeng Wang, Hao Zou, Yujie Feng, Wei Zhao, Kun Li, Kui Liu, Bingyuan Zhang ⓘ, and Chengzhan Zhu ⓘ

Department of Hepatobiliary and Pancreatic Surgery, The Affiliated Hospital of Qingdao University, No. 16 Jiangsu Road, Qingdao 266003, China

Correspondence should be addressed to Chengzhan Zhu; zhuchengz@qduhospital.cn

Hepatocellular carcinoma (HCC) is a common malignant tumor with high incidence and mortality rates. However, a reliable prognostic signature has not yet been confirmed. Thus, the purpose of the present study was to develop a biomarker with high specificity and sensitivity for the diagnosis and prognosis of patients with HCC. The mRNA expression profiles of HCC were obtained from the GSE19665, GSE41804, and TCGA databases. Subsequently, 193 differentially expressed genes (DEGs) were identified from the intersection of the data from the three datasets. Bioinformatics analysis showed that the identified DEGs are related to the cell cycle, oocyte meiosis, and p53 signaling pathway, among other factors, in cancers. A protein-protein interaction (PPI) and a functional analysis were performed to investigate the biological function of the DEGs and obtain the candidate genes using the MCODE of Cytoscape. The candidate genes were introduced into the TCGA database for survival analysis, and the four candidate genes that were hub genes and meaningful for survival were retained for further verification. We validated the gene and protein expression and determined the prognosis of our patient cohort. In addition, we evaluated the biological functions regulating tumor cell proliferation and metastasis *in vitro*. According to the ROC curve analysis of gene expression in clinical samples, it was found that the four genes can be used to predict the diagnosis. A survival analysis based on data from the TCGA database and clinical samples showed that the four genes may be used as biomarkers for providing prognoses for patients. The cell functional experiments revealed that these four genes were related to tumor proliferation, migration, and invasion. In conclusion, the genes identified in the present study could be used as markers to diagnose and predict the prognosis of patients with HCC and guide targeted therapy.

1. Introduction

Hepatocellular carcinoma (HCC) is the fourth leading cause of cancer mortality worldwide and is one of the most common malignant cancers because of limited treatment options and poor prognosis [1]. The main treatment strategies include hepatectomy, liver transplantation, and targeted therapy [2, 3]. Because of microvascular invasion and heterogenicity [4, 5], early recurrence and metastasis after the surgery and poor responses to the targeted therapy are the main causes of short long-term survival [6]. Therefore, significant targets that could predict the prognosis of HCC and be the probable targets of therapy are urgently required.

Bioinformatics is widely used to comprehensively analyze the datasets with large numbers of cases to assess the genes related to the prognosis of liver cancer and/or to identify the genes that can be used as therapeutic targets. At present, most gene biomarkers are used to predict the prognosis and survival of cancer patients [7, 8] and provide guidance for further treatment decisions. For instance, Li et al. used bioinformatics to identify several key biomarkers that provide a candidate the diagnostic target and treatment for HCC [9]. It is different from the genes we screened for in the present study. Similarly, the previous research has only used the TCGA database, however, these results are different from the results presented in the present study [10].

Furthermore, in the previous bioinformatics analyses, there were few functional experiments to verify the results, and we have included this in the present study.

In the present study, the datasets of the expression profiles were downloaded from the GEO and TCGA databases to obtain the DEGs. Bioinformatic functional analyses were conducted to identify the prognosis-related genes and cancer-related molecular mechanisms. A new signature has been identified as a prognostic biomarker for HCC. The biological functions of the hub genes were experimentally confirmed.

2. Materials and Methods

2.1. Datasets and DEGs Identification. Two datasets (GSE41804 and GSE19665) of mRNA gene expression were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). The gene expression profiles were downloaded from the TCGA database (https://cancergenome.nih.gov/). The GSE41804 dataset contains the paired samples of 20 HCC tissues and 20 adjacent tissues from 20 patients. The GSE19665 database contains 10 HCC and 10 non-HCC samples from 10 patients. We also obtained 371 tumor and 50 nontumor samples from the TCGA database for validation purposes.

In the GEO database, GEO2R is a convenient online tool for users to compare the datasets in a GEO series to distinguish the DEGs between the HCC and noncancerous samples. The p-values and the Benjamini–Hochberg test were used to coordinate the significance of the DEGs obtained and reduce the number of false positives. Subsequently, the DEGs were screened against the corresponding datasets based on a p-value < 0.05, and $|logFC|$ (fold change) ≥ 2 was used as a threshold to improve the credibility of the results. Then, the lncRNAs and miRNAs obtained from the TCGA database were eliminated. We acquired three groups of mRNA expression profiles after processing the data. The applet (http://bioinformatics.psb.ugent.be/webtools/Venn/) was used to determine which data in the three groups intersect.

2.2. PPI Network Construction. The PPI network was predicted using the Search Tool for the Retrieval of Interacting Genes (STRING; http://string-db.org) online database [11]. Research on the functional interactions between the proteins can provide a better understanding of the potential mechanisms underlying the occurrence or development of cancers. In the present study, the PPI interactions with a combined score >0.4 were regarded as statistically significant in the STRING database and were extracted to build a network of DEGs. We then visualized the molecular interaction network using Cytoscape (Version 3.6.1), which is an open-source bioinformatics software [12]. The Molecular Complex Detection (MCODE) [13] is a plug-in packaged with the Cytoscape software. The main function of MCODE is to cluster and construct the functional modules by means of topology in a huge gene (protein) network to find dense rendezvous spots. MCODE scores > 5, node density

cutoff $= 0.1$, degree cutoff and K-core $= 2$, node score cutoff $= 0.2$, and a maximum depth of 100 were used as the benchmarks for the gene module selection.

2.3. GO and KEGG Pathway Enrichment Analyses. The cluster profiler package [14] obtained from Bioconductor (http://bioconductor.org/) is a free online bioinformatics package in R. It contains biological data and analysis tools that provide a systematic and comprehensive biological functional annotation information of the large-scale genes or proteins that help the users extract biological information from them. Gene Ontology (GO) enrichment analysis is widely used for gene annotation and the analysis of the biological processes of DEGs [15]. Statistical significance was set at $p < 0.05$. A KEGG pathway enrichment analysis (http://www.genome.jp/kegg/pathway.html) provides an understanding of the advanced functions of the biological systems at the molecular level. It is widely used for large-scale molecular datasets produced by high-throughput experimental technologies [16].

2.4. Survival Analysis and Expression Levels of the Hub Genes. The survival analysis of the hub genes was performed using Kaplan–Meier analysis. Using GEPIA (http://gepia2.cancer-pku.cn), a TCGA visualization website, all of the expression information of the patients with HCC in the TCGA database were divided into high- and low-expression groups according to the median of each gene expression level. In addition, the gene expression of patients in our hospital was obtained using real-time PCR, and the corresponding survival analysis was performed according to the aforementioned method of analysis. Furthermore, the box plots of GEPIA were plotted to reflect the expression levels of each gene.

2.5. Establishment and Validation of the Prediction of the Signature. The signature was applied to a cohort of patients with HCC in our hospital to verify its ability to predict HCC. The expression of the genes in patients with HCC was measured, and the ROC curve was obtained using GraphPad Prism 7.

2.6. Cox Regression Analysis and Prognostic Validation of the Signature. The intersection of the DEGs among the three cohorts of mRNA expression profiles was selected to construct the predictive character for survival. The aforementioned hub genes in the TCGA cohort were incorporated into a multivariate Cox regression model using the online Kaplan–Meier plotter [17] to obtain the survival analysis and verification of the biomarkers. The prognosis risk score for predicting the overall survival (OS) of HCC patients was determined by multiplying the expression level of these genes (exp) by a regression coefficient (β) obtained from the multivariate Cox regression model. The algorithm used was Risk score $= \mathrm{EXP}_{\mathrm{gene}_1} * \beta_{\mathrm{gene}_1} + \mathrm{EXP}_{\mathrm{gene}_2} * \beta_{2\mathrm{gene}_2} + \cdots \mathrm{EXP}_{\mathrm{gene}n} * \beta_{\mathrm{gene}_n}$. A total of 364 HCC patients with accessible data were selected for the individual survival analyses. The

median risk score was treated as a threshold. Based on this, the HCC patients were divided into low-risk and high-risk groups. Similarly, the aforementioned algorithm was applied to a cohort of HCC patients from our hospital to verify the prognostic value of the hub genes.

2.7. Cell Culture and Plasmid Transfection.

The human HCC cell line M3 was obtained from the Shanghai Institute for Biological Sciences, Chinese Academy of Sciences. The cell line was cultured in HyClone Dulbecco's Modified Eagle's Medium (DMEM) mixed with 10% fetal bovine serum (FBS) and 1% penicillin and streptomycin in a humidified incubator (Thermo Scientific, USA) at 37°C and 5% CO_2. The control and gene-overexpressing plasmids were purchased from Genechem Co. Ltd (Shanghai, China). EXO1, CYP2C8, and CLEC1B share the GV141 vector, which contains a multiple cloning site followed by a 3FLAG tag downstream of the CMV promoter, as well as the neomycin gene downstream of the SV40 promoter. In addition, the vector of GYS2 is GV230, which contains a multiple cloning site followed by EGFP downstream of the CMV promoter, as well as the neomycin gene downstream of the SV40 promoter. Transient transfection was performed using Lipofectamine 2000 (Invitrogen, USA) according to the manufacturer's instructions.

2.8. Wound Healing Assay.

The cells were cultured in a six-well cell culture plate overnight at 37°C and 5% CO_2. The following day, transient transfection was performed, and the cells were cultured for 24 h. The cells were washed twice with phosphate-buffered saline (PBS). A $200 \mu L$ tip of a pipette was then used to make cross scratches at the bottom of the wells. The images were captured at 0 and 24 h using a microscope (Nikon, Japan).

2.9. Cell Migration and Invasion Assays.

The migration assay (Transwell assay) was conducted after transfection and culture. The cells of the control and experimental groups were counted and placed in the transwell chambers. The HCC cells (105 cells) were fully mixed with the serum-free DMEM and added to the interior of the chambers. Then, $600 \mu L$ of DMEM with 10% FBS was added to the bottom of the 24-well cell culture plate. The chambers were placed in the plate wells. Matrigel was added to the chambers for invasion assay. During culturing in the incubator for 10 h to 24 h, methanol was added to the chambers to fix the cells that had migrated or invaded. At the end of the treatment, the cells were stained with crystal violet, observed, and counted under a microscope.

2.10. Cell Growth and Cloning Assays.

The transfected cells were incubated in a 96-well cell culture plate overnight at 37°C and 5% CO_2 to attach CYP2C8: 5000 cells/well, EXO1, CLEC1B, and GYS2: 2000 cells/well. A Cell Counting Kit-8 (CCK8) was used to detect the viability of the cells at different points in time. The cells were cultivated for 2 h in an incubator after the addition of $100 \mu L$ DMEM with $10 \mu L$ CCK-8 reagent per well. Subsequently, the absorbance optical density (OD) was measured at 450 nm. For the cloning assay, 1000 transfected cells per well were implanted in a six-well culture plate and cultured for 14 days at 37°C and 5% CO_2. The cells were fixed with 10% formalin, stained with crystal violet, and counted.

2.11. HCC Tissue Samples.

We ethically collected a total of 40 pairs of tumor and paratumor tissues from the patients who underwent hepatectomy between 2016 and 2019 at the Affiliated Hospital of Qingdao University. All patients were confirmed to have HCC by postoperative pathology, and their clinical and follow-up data were integrated into the present study. The present study complied with the Declaration of Helsinki, and all sample collections were approved by the Ethics Committee of The Affiliated Hospital of Qingdao University. Informed consent was obtained from each patient. None of the tissue samples was exposed to any preoperative radiotherapy or chemotherapy. Paired cancer and paracancerous tissues were collected in RNA protective additive-filled microtubes and immediately frozen in a medical refrigerator at −80°C.

2.12. Real-Time PCR Analysis for Verification of the Signature.

Total RNA was extracted from the collected tumor and nontumor tissue samples using RNAiso Plus (total RNA extraction reagent, Takara Inc. Japan) according to the manufacturer's instructions. Subsequently, the fineness of the total RNA was estimated using the A260/A280 ratio, and the corresponding DNA was obtained using the PrimeScript™ RT reagent Kit (Takara Inc. Japan). Quantitative real-time PCR (qRT-PCR) was performed using a LightCycler 480 (Roche, Basel, Switzerland) and Takara TB Green™ Premix Ex Taq™ II (Tli RNaseH Plus, Takara Inc. Japan). The thermocycling procedure was as follows: preincubation at 95°C for 5 min, followed by 35 temperature cycles (each involving denaturation at 95°C for 20 s, annealing at 60°C for 30 s, and extension at 72°C for 20 s). The procedure was repeated 3 times. An analysis of the melting curve was performed on each sample to verify amplification specificity. The primer sequences of nine genes are shown in Table 1. The data were standardized using a control group of β-actin.

2.13. Statistical Analysis.

Data analyses were performed using GraphPad Prism software (version 7.0; San Diego, CA, USA). All experimental data are expressed as mean ± SD. The statistical significance of the results was analyzed by the one-way analysis of variance (ANOVA) for comparisons between three or more groups and the Student's t-tests for comparisons between the two groups. The differences were considered statistically significant at $p < 0.05$. All experiments were performed in triplicate.

3. Results

3.1. DEGs Identification and Identification of the Hub Genes.

The entire research process is illustrated in Figure 1. After processing the GSE19665, GSE41804, and TCGA data, the DEGs

TABLE 1: mRNA PCR primer.

Gene name	Primer sequence
EXO1	F: TCGGATCTCCTAGCTTTTGGCTG
	R: AGCTGTCTGCACATTCCTAGCC
CYP2C8	F: GAGACAACAAGCACCACTCTGAG
	R: CAGTGTAAGGCATGTGGCTCCT
CLEC1B	F: TGGTGGCGTGTGATGGCTTTGA
	R: CACCTTGTAGGTAATTGCGCTGC
GYS2	F: CCAGTGACCACGCACAACATGA
	R: GTAAGGGACTGGTGGAGGATAG
β-actin	F: CCTCTCCCAAGTCCACACAG
	R: GGGCACGAAGGCTCATCATT
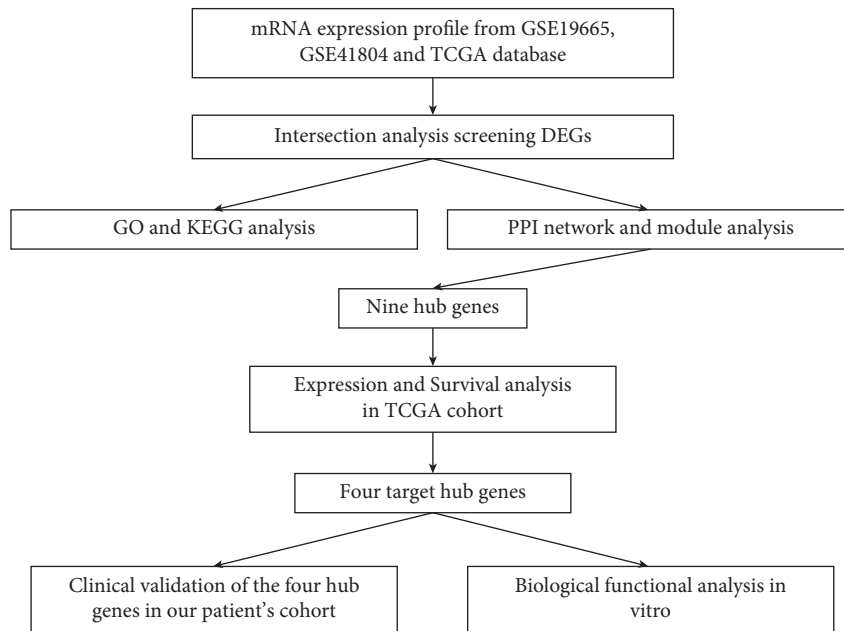
FIGURE 1: Workflow of this study to construct a four-gene signature in HCC.

were identified between HCC and nontumor tissue using $p < 0.05$ and $|logFC| \geq 2$ as the thresholds. There were 1341, 155, and 943 upregulated genes in GSE19665, GSE41804, and TCGA, respectively. There were 224, 389, and 362 downregulated genes in GSE19665, GSE41804, and TCGA, respectively. The up- and downregulated genes are shown with a volcano plot in Figure 2(a). One hundred and ninety-three DEGs were identified by the intersection of the genes between the three cohorts (Figure 2(b)). The detailed positions on the chromosomes of these 193 genes are shown in the Circos plot (Figure 2(c)).

A PPI dataset was obtained from STRING and used to construct a PPI network of the DEGs. Subsequently, an interaction analysis was performed to visualize the interaction network using Cytoscape (Figure 3(a)). The results showed that MT1M, CYP2C8, CFP, EXO1, CLEC1B, GRHL2, SLCO1B3, HAMP, and GYS2 were the nine most highly ranked genes (Table 2).

3.2. Functional Enrichment and Survival Analysis of the Hub Genes. The GO enrichment analysis was conducted to investigate the biological functions, which indicated that the cellular processes and biological regulation were significantly enriched in the biological processes (BP). The main enrichments included the binding of iron ions, activity of monooxygenase, heme binding, and oxidoreductase activity (Figure 3(b)). KEGG pathway analysis showed that these genes were significantly enriched in the pathways related to retinol metabolism, the cell cycle, oocyte meiosis, and the p53 signaling pathway in cancers (Figure 3(c)). These results suggest that these genes are important for the pathogenesis and progression of HCC.

A Kaplan–Meier analysis was performed to screen out the genes in the TCGA database that were related to overall survival (OS). Four of the nine genes were significantly correlated with the prognosis. The patients with a high expression level of CLEC1B ($p = 0.017$), GYS2 ($p = 0.00052$), and CYP2C8 ($p = 0.0066$) and a low expression level of EXO1 ($p = 0.00032$) had a favorable prognosis (Figure 4(a)). Then, we validated the function of predicting the prognosis of patients in our cohort using Kaplan–Meier analysis as shown in Figure 4(b). We then
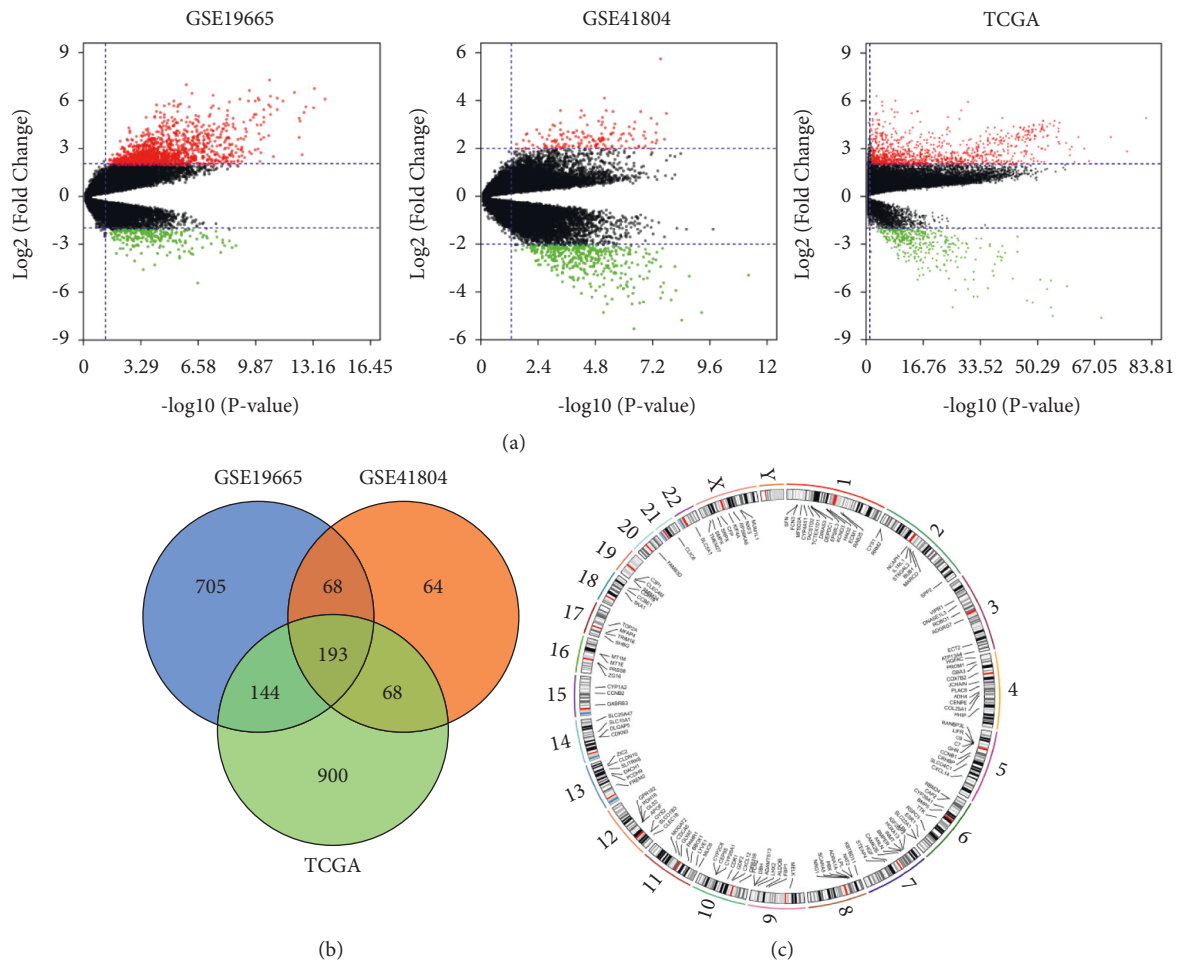
(a)

(b)

(c)

FIGURE 2: DEGs in HCC. (a) Volcano plot of all genes expression profiles in GSE16515, GSE28735, and TCGA. The red represents the mRNA with high expression level, while the green represents the mRNA with low expression level. (b) Venn diagram showing DEGs in GSE16515, GSE28735, and TCGA. (c) Circos plots showing the position of DEGs on the chromosome.

investigated the level of gene expression in the TCGA database (Figure 4(c)) and in our cohort ($n = 40$, $p < 0.05$, Figure 5(a)). Clinicopathological information is listed in Table 3. CYP2C8, CLEC1B, and GYS2 were downregulated, whereas EXO1 was upregulated in HCC ($p < 0.05$).

3.3. Predictive and Prognostic Indication of the Four-Gene Signature.
To evaluate the four-gene signature for predicting HCC, an analysis of the ROC curve of each gene was performed according to the expression of each gene. As presented in Figure 5(b), the AUCs of EXO1, CLEC1B, GYS2, and CYP2C8 were 0.7339, 0.8866, 0.7963, and 0.8455, respectively.

To validate the function as a prognostic indicator, we divided the 364 patients in the TCGA database into low-risk (182 patients) and high-risk cohorts (182 patients) according to their median risk score. The OS of the high-risk cohort was superior to that of the low-risk (log-rank $p = 0.0003$, Figure 5(c)). A similar result was obtained for our cohort (log-rank $p = 0.026$) with further validation, which indicated that the signature could be a favorable predictive biomarker of HCC (Figure 5(d)).

3.4. Biological Functions of the Four Hub Genes.
To investigate the biological functions of the genes, we overexpressed the genes by plasmid transfection in the M3 HCC cell line. The CCK8 and cloning assays indicated that the overexpression of CLEC1B, GYS2, and CYP2C8 inhibited cell proliferation, while EXO1 significantly promoted M3 cell proliferation (Figures 6(a) and 6(b)). The wound healing assay and Transwell assay demonstrated that the upregulation of CLEC1B, GYS2, and CYP2C8 restrained migration, while EXO1 accelerated the migration of M3 cells (Figures 6(c) and 6(d)). In addition, the Transwell assay with Matrigel showed that the overexpression of CLEC1B, GYS2, and CYP2C8 inhibited cell invasion, while the upregulation of EXO1 promoted the invasion of M3 cells (Figure 6(e)).
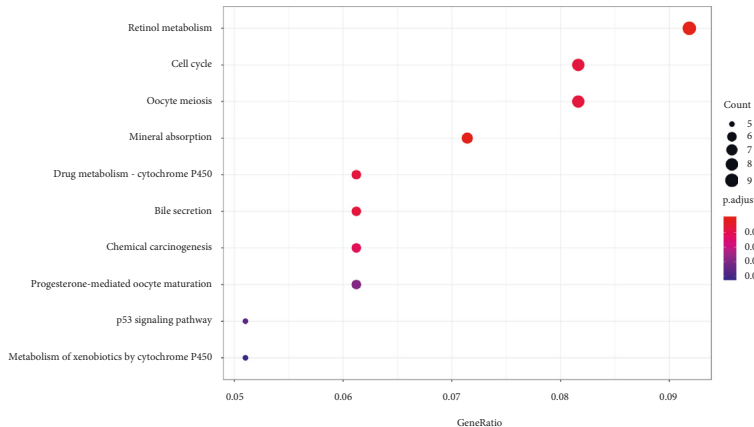
4. Discussion

Bioinformatics analysis is a useful tool for exploring the prognostic indicators of cancer and possible therapeutic targets. We performed an integrated analysis of two GEO datasets and the TCGA database and identified four hub genes based on a biofunctional analysis. A signature was
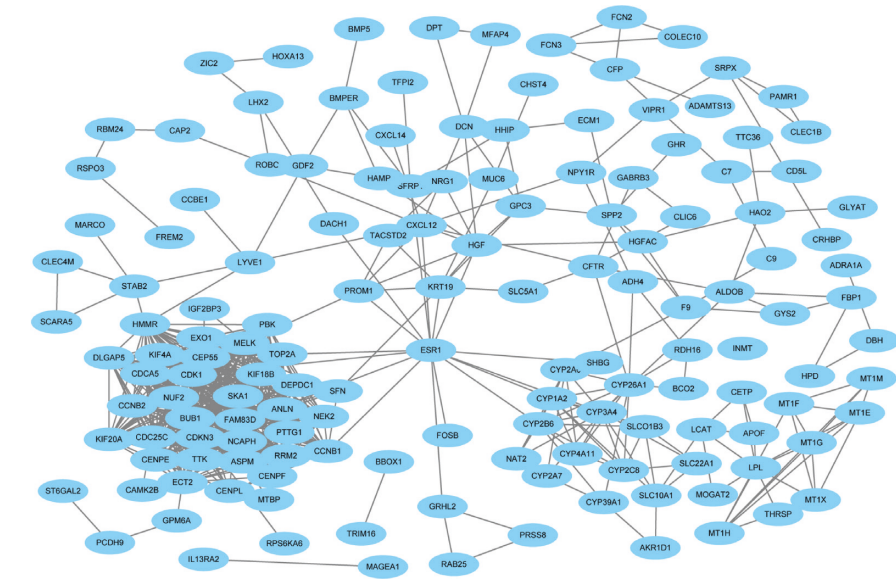
(a)
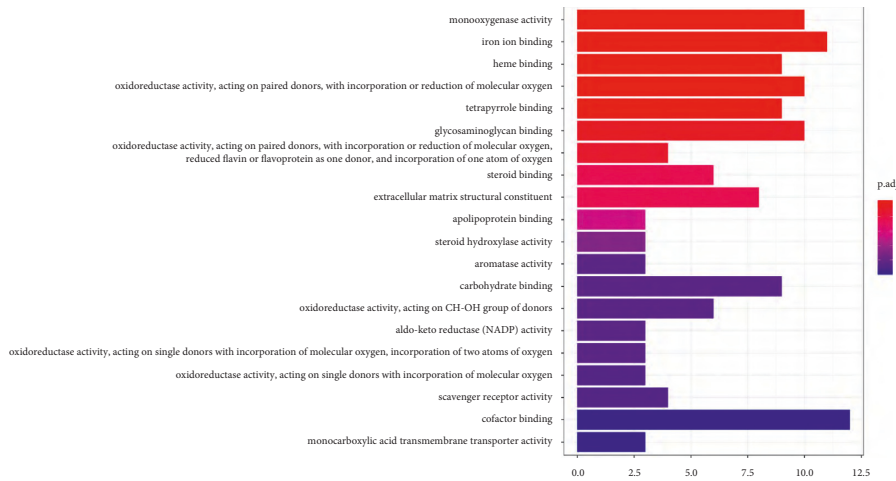
(b)

(c)

FIGURE 3: Function analysis of the DEGs. (a) PPI network of DEGs. (b) GO function analysis of the DEGs. (c) KEGG pathway enrichment of DEGs.

TABLE 2: List of 9 hub genes identified in the PPI network.

| Gene symbol	Description
MT1M	Metallothionein 1M
CYP2C8	Cytochrome P450 family 2 subfamily C member 8
CFP	Complement factor properdin
EXO1	Exonuclease 1
CLEC1B	C-Type lectin domain family 1 member B
GRHL2	Grainyhead like transcription factor 2
SLCO1B3	Solute carrier organic anion transporter family member 1B3
HAMP	Hepcidin antimicrobial peptide
GYS2	Glycogen synthase 2

PPI, protein–protein interaction.
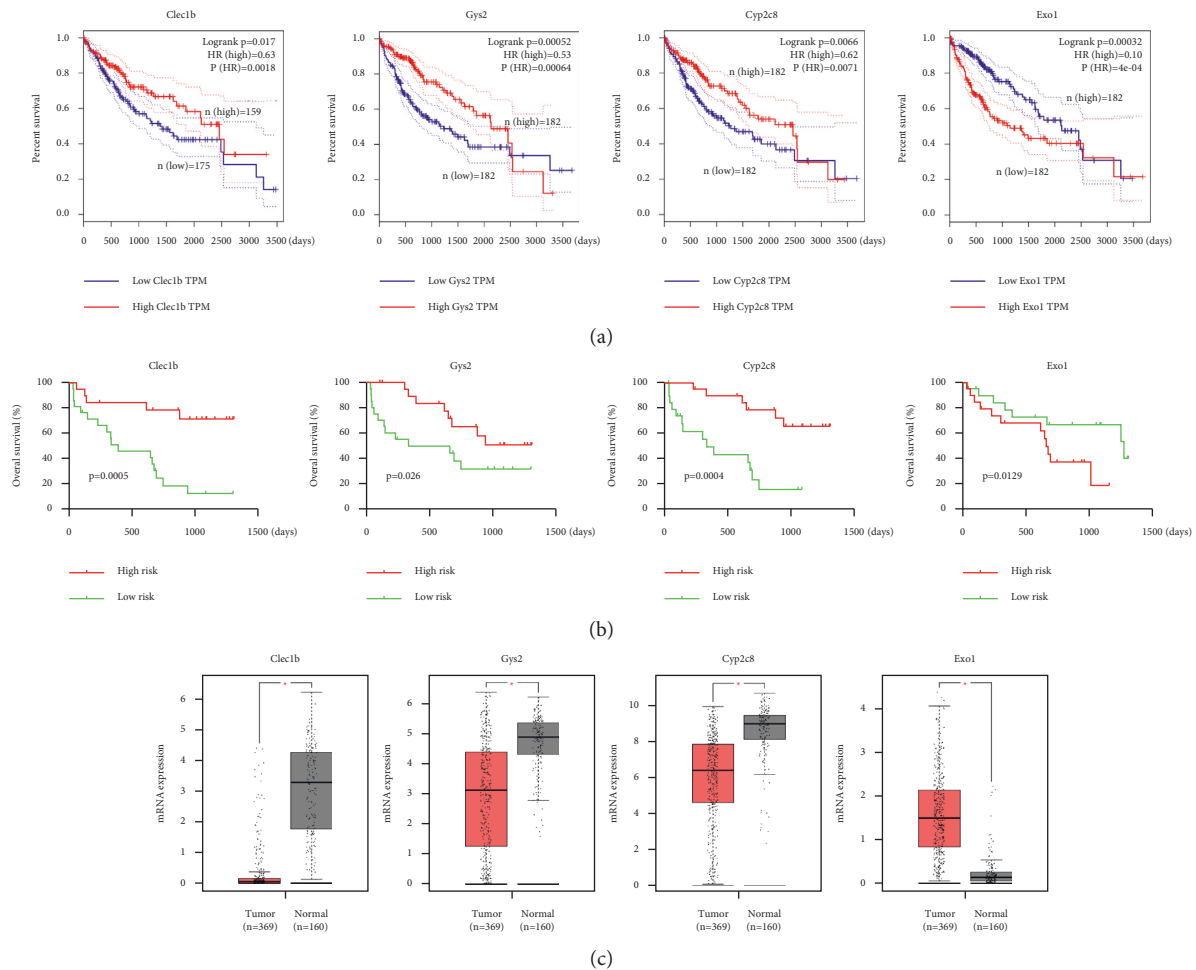
FIGURE 4: The situation of signatures in the TCGA database. (a) Kaplan–Meier survival plots of four genes in TCGA cohort. (b) Kaplan–Meier survival plots of four genes in our patient cohort. (c) Box plots showing the expression of four genes in TCGA cohort.

established and validated to predict the prognosis of patients with HCC. In addition, the biological functions of these four genes were investigated experimentally. In the present study, the genes expressed in the nontumor liver tissues and tumor liver tissues from the two GEO and TCGA databases were analyzed and filtered to identify a novel set of four-gene signatures (CLEC1B, GYS2, CYP2C8, and EXO1) to diagnose and determine the prognosis of patients with HCC. CLEC1B, GYS2, and CYP2C8 were identified as tumor suppressors for the prognosis, however, EXO1 was

determined to be an oncogene. In spite of GYS2 [18] and CYP2C8 [19], which are regarded as one hub gene with two signatures, their screening criterion was |logFC| > 1, while | logFC| ≥ 2 was adopted in our study. CLEC1B was identified as a new biomarker [20], however, because of a lack of verification of the survival model, its reliability remains low. Here, we used more stringent screening criteria and analyzed these four genes together to establish a survival model. The performance of the Kaplan–Meier analysis obtained from the TCGA information on the HCC patients
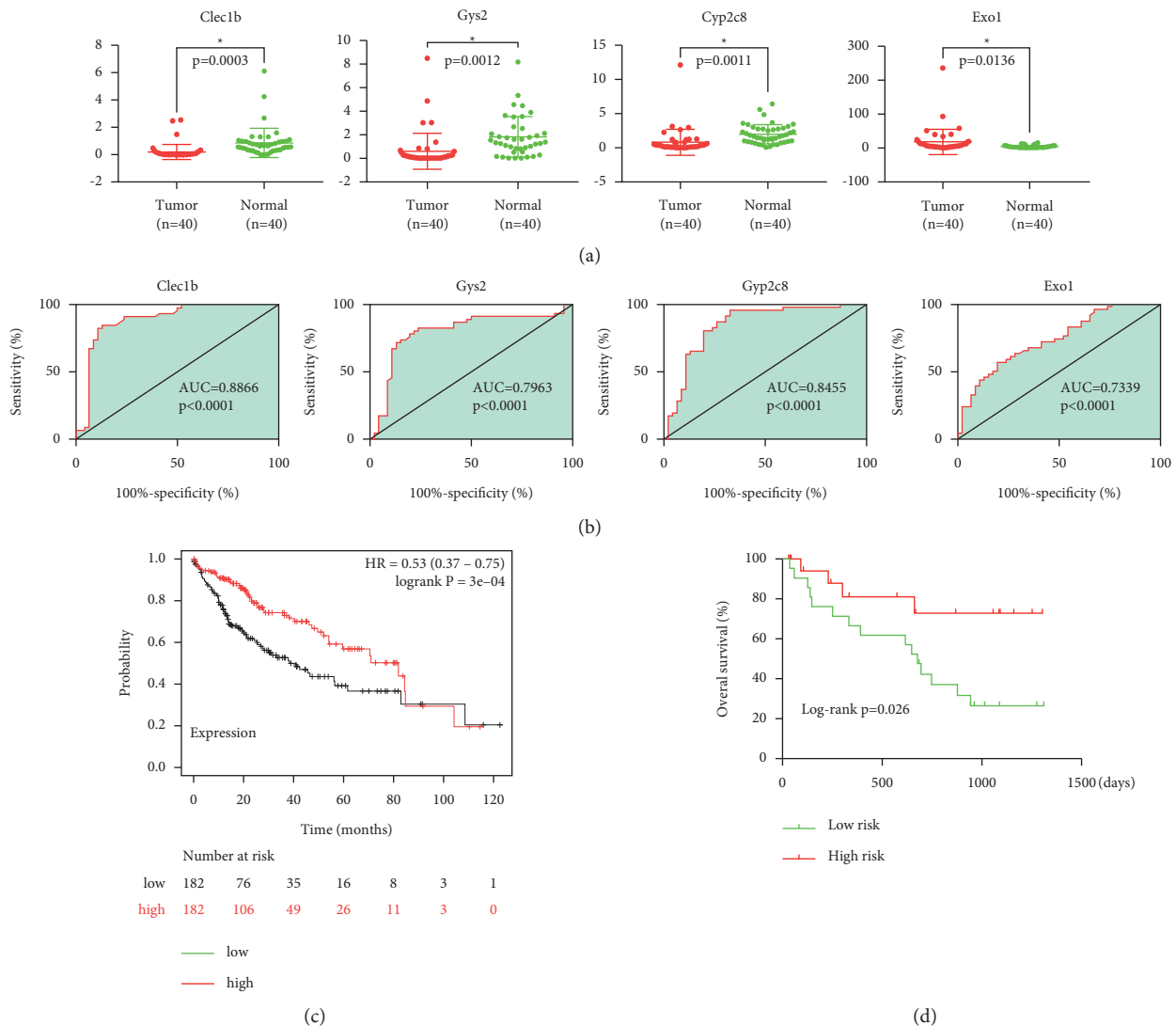
Figure 5: The validation of signatures in our patient cohort. (a) qRT-PCR analysis: mRNA expression levels of four genes in HCC and paired nontumorous liver tissues. (b) ROC analysis of four genes. (c) Kaplan–Meier survival plots of the four-gene signature in TCGA cohort. (d) Kaplan–Meier survival plots of the four-gene signature in our patient cohort.

demonstrated that the four-gene signature was a sensitive potential biomarker for predicting the prognosis of patients with HCC not only for each individual gene but also for the four-gene association. Synchronously, it is different from most other simple bioinformatics studies that use only one dataset [21, 22]. Furthermore, the clinical tissue data were analyzed using ROC to diagnose HCC. The ROC analysis consequently proved the accuracy and specificity of the four-gene signature in the diagnosis of HCC. The functional verification of these genes has seldom been conducted in other studies. Other studies remained theoretical. We also investigated the functions of the four genes corresponding to the signature at the cellular level and the level of expression of the corresponding proteins in the cancer and para-cancerous tissues. In summary, a multidimensional analysis of these four genes firmly demonstrated that the combination of these four genes could effectively predict the prognosis of HCC patients.

Glycogen synthase 2 (GYS2) is a key enzyme in glycogen biosynthesis. GYS2 was significantly downregulated in HCC with glycogen loss, resulting in a poor prognosis. GYS2 inhibited tumor growth in HBV-related HCC by negative feedback in the p53 signaling pathway [23]. By a series of *in vitro* experiments, we confirmed that the overexpression of GYS2 can lead to the proliferation, metastasis, and invasion of HCC cells.

Exonuclease 1 (EXO1) is an exonuclease from the $5'$ to $3'$ end that participates in the regulation of the cell cycle checkpoint, the maintenance of replication forks, and the postreplication repair of DNA [24]. A deficiency in restarting the DNA replication pathway may lead to double-strand breaks, cell cycle arrest, cell death, or transformation, which may lead to cancer [25], and EXO1 is involved in this process. Therefore, the variations in EXO1 have been linked to various types of cancers [26]. Furthermore, several lines of previous research have reported a negative relationship

TABLE 3: The correlation of HCC clinic pathological variables with gene expression level in tissue samples.

Clinic pathological features	Low risk ($n = 20$)	High risk ($n = 20$)	p-value
Age (years)			
<60	13	8	0.113
≥60	7	12	
Gender			
Male	15	13	0.49
Female	5	7	
Smoking			
Yes	14	7	0.027
No	6	13	
Alcohol			
Yes	16	8	0.01
No	4	12	
AFP level (ng/L)			
≤400	8	16	0.01
>400	12	4	
Microvascular invasion			
Yes	13	6	0.027
No	7	14	
TNM stage			
I-II	8	6	0.507
III-IV	12	14	

HCC, hepatocellular carcinoma; TNM, tumor, lymph node and metastasis.

(a)

(b)

(c)

(d)

FIGURE 6: Continued.

Matrigel-transwell assay

(e)
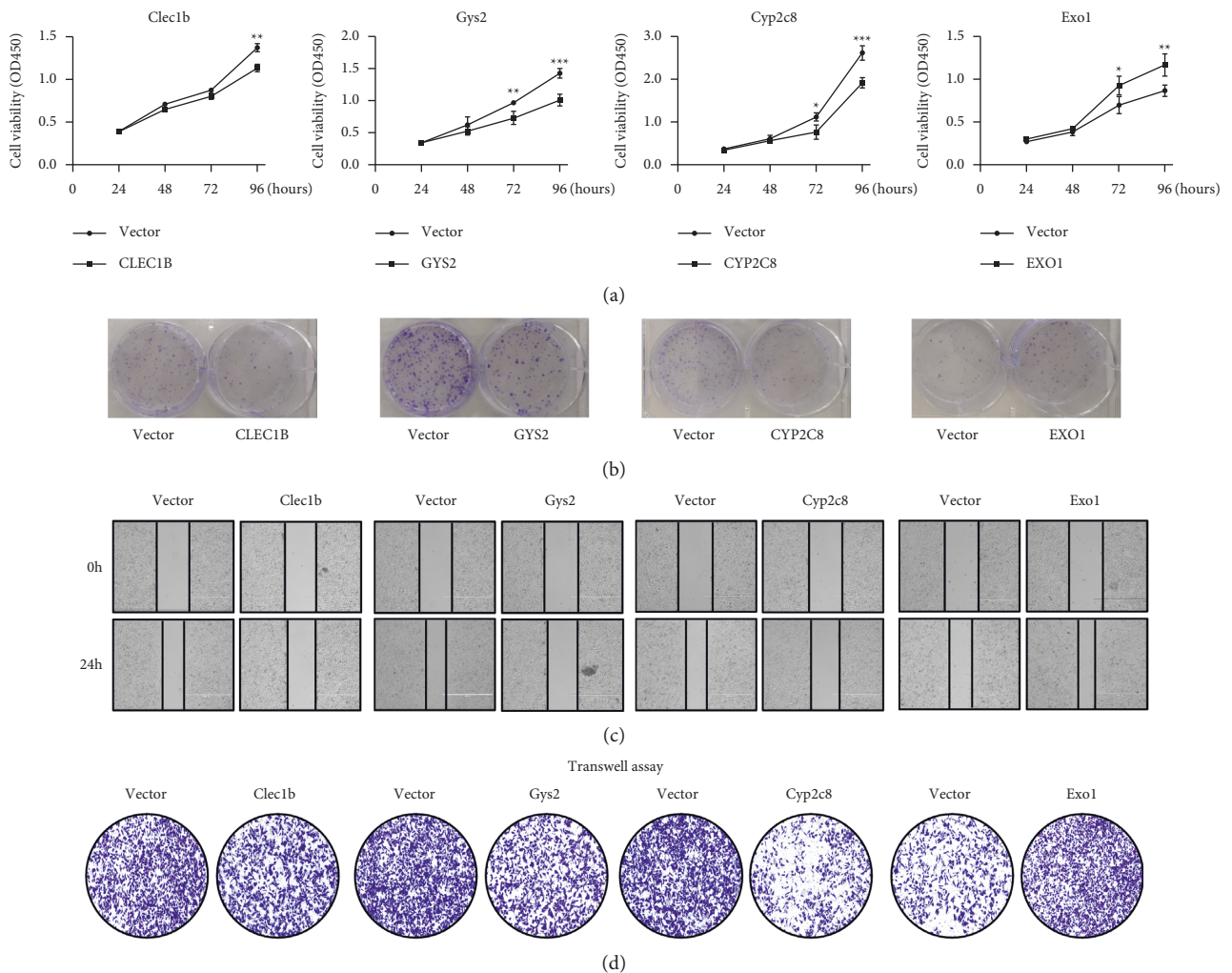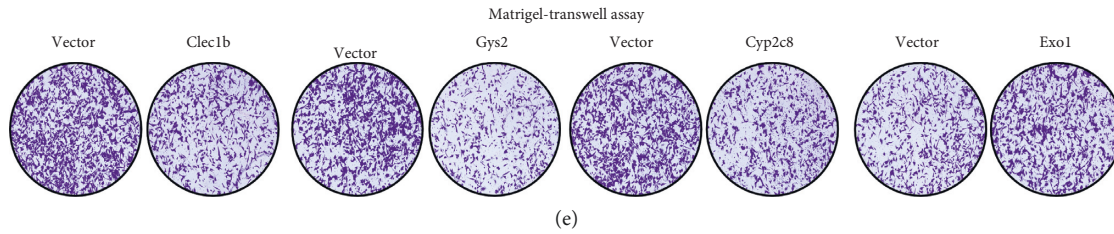
Figure 6: Biological function of gene overexpression of M3 cells. (a) Cell proliferation curves of the vector group and gene-overexpression group. (b) Colony assays of the vector group and gene-overexpression group. (c) Wound healing assays verifying the migration ability of M3 cells in vector group and transfected group. (d) Transwell assays validating the motility of vector group and gene-overexpression group. (e) Matrigel-transwell assays contrasting the invasiveness between the vector group and gene-overexpression group.

between the overexpression of EXO1 and the prognosis of various cancers [27–30]. It has been reported that the overexpression of EXO1 leads to a poor prognosis in patients with HCC [31]. In addition, it has been shown that the overexpression of EXO1 is associated with a poor prognosis in breast cancer [32].

CYP2C8 is located in the cytochrome P450 gene cluster on chromosome 10q24 and can metabolize approximately 30% of the body's clinical drugs and various chemicals from the environment [33]. Moreover, KEGG analysis showed that CYP2C8 is related to retinol and chemical metabolism. A previous study pointed out that the OS of patients with HCC with low CYP2C8 was worse than that for those with high CYP2C8 [34, 35]. In addition, the low expression level of CYP2C8 was related to advanced clinicopathological features, including tumor stage and intrahepatic metastasis. According to the database, CYP2C8 is well-expressed in normal human livers, and CYP2C8 metabolizes paclitaxel [36]. Paclitaxel is prescribed in combination with the cytochrome P450 inhibitors to enhance its anticancer effects against various malignant tumors [37]. Therefore, this finding may explain why paclitaxel has effective antitumor activity *in vitro* but has no significant clinical effect on patients with HCC. Our research has further promoted the use of paclitaxel in patients with clinical liver cancer *in vitro*.

CLEC1B, a member of the C-type lectin domain family 1, is mainly related to the thromboses caused by platelet aggregation, platelet-mediated tumor proliferation, and metastasis [38, 39]. Furthermore, it has been previously reported that CLEC2 is significantly downregulated in the HCC tissues [40], which agrees with our results. A recent study also revealed that the downregulation of CLEC2 is related to the depth to which the tumor has invaded, lymph node metastasis, and the 5-year survival rate [41]. In the present study, we confirmed the role of CLEC1B, as reported by previous studies, that the overexpression of CLEC1B distinctly suppressed the proliferation, metastasis, and invasion of the HCC cells. We also confirmed that CLEC1B is a marker gene highly related to the progression of HCC and the low expression level of CLEC1B may be a significant prognostic factor, suggesting a poor clinical outcome. Furthermore, it can be used as a target for immunotherapy, which is consistent with the views of Hu et al. [42].

We believe that the signature of the four genes combined is a promising prognostic indicator for patients with HCC.

However, there are some limitations to the present study. Firstly, the mechanism of gene regulation in HCC progression needs further investigation. Secondly, because of the shortage of clinical specimens and short-term follow-up, the data needs further validation with more patients [43].

5. Conclusion

Based on the results of the present study, which involved an analysis of the data in the databases and the experimental verification of the functions of genes, we identified a reliable signature to predict the prognosis of patients with HCC. In addition, the signature can be used as a guide for targeted cancer therapy. However, further studies are needed to develop the prognostic and diagnostic signatures for HCC.

Data Availability

The datasets used in this study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Authors' Contributions

Guangfeng Wang and Hao Zou contributed equally to this work.

Acknowledgments

References

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[2] A. Takeda, N. Sanuki, Y. Tsurugai et al., "Phase 2 study of stereotactic body radiotherapy and optional transarterial chemoembolization for solitary hepatocellular carcinoma not amenable to resection and radiofrequency ablation," *Cancer*, vol. 122, no. 13, pp. 2041–2049, 2016.

[3] A. A. Mokdad, H. Zhu, J. A. Marrero, J. C. Mansour, A. G. Singal, and A. C. Yopp, "Hospital volume and survival after hepatocellular carcinoma diagnosis," *American Journal of Gastroenterology*, vol. 111, no. 7, pp. 967–975, 2016.

[4] Z. Peng, S. Chen, H. Xiao et al., "Microvascular invasion as a predictor of response to treatment with sorafenib and transarterial chemoembolization for recurrent intermediate-stage hepatocellular carcinoma," *Radiology*, vol. 292, no. 1, pp. 237–247, 2019.

[5] K.-C. Lim, P. K.-H. Chow, J. C. Allen et al., "Microvascular invasion is a better predictor of tumor recurrence and overall survival following surgical resection for hepatocellular carcinoma compared to the Milan criteria," *Annals of Surgery*, vol. 254, no. 1, pp. 108–113, 2011.

[6] C. L. Chaffer and R. A. Weinberg, "A perspective on cancer cell metastasis," *Science*, vol. 331, no. 6024, pp. 1559–1564, 2011.

[7] S. Klahan, W.-C. Huang, C.-M. Chang et al., "Gene expression profiling combined with functional analysis identify integrin beta1 (ITGB1) as a potential prognosis biomarker in triple negative breast cancer," *Pharmacological Research*, vol. 104, pp. 31–37, 2016.

[8] A. Suzuki, T. Horie, and Y. Numabe, "Investigation of molecular biomarker candidates for diagnosis and prognosis of chronic periodontitis by bioinformatics analysis of pooled microarray gene expression datasets in Gene Expression Omnibus (GEO)," *BMC Oral Health*, vol. 19, no. 1, p. 52, 2019.

[9] 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.

[10] B. Li, W. Feng, O. Luo et al., "Development and validation of a three-gene prognostic signature for patients with hepatocellular carcinoma," *Scientific Reports*, vol. 7, no. 1, p. 5517, 2017.

[11] A. Franceschini, D. Szklarczyk, S. Frankild et al., "STRING v9.1: protein-protein interaction networks, with increased coverage and integration," *Nucleic Acids Research*, vol. 41, pp. D808–D815, 2013.

[12] B. Demchak, T. Hull, M. Reich et al., "Cytoscape: the network visualization tool for GenomeSpace workflows," *F1000Research*, vol. 3, p. 151, 2014.

[13] W. P. Bandettini, P. Kellman, C. Mancini et al., "MultiContrast Delayed Enhancement (MCODE) improves detection of subendocardial myocardial infarction by late gadolinium enhancement cardiovascular magnetic resonance: a clinical validation study," *Journal of Cardiovascular Magnetic Resonance*, vol. 14, no. 1, p. 83, 2012.

[14] G. Yu, L.-G. Wang, Y. Han, and Q.-Y. He, "ClusterProfiler: an R package for comparing biological themes among gene clusters," *OMICS: A Journal of Integrative Biology*, vol. 16, no. 5, pp. 284–287, 2012.

[15] M. Ashburner, C. A. Ball, J. A. Blake et al., "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium," *Nature Genetics*, vol. 25, no. 1, pp. 25–29, 2000.

[16] M. Kanehisa, "The KEGG database," *Novartis Foundation Symposium*, vol. 247, pp. 91–52, 2002.

[17] B. Győrffy, A. Lanczky, A. C. Eklund 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 Research and Treatment*, vol. 123, no. 3, pp. 725–731, 2010.

[18] C. Li, D. Zhou, X. Jiang, M. Liu, H. Tang, and Z. Mei, "Identifying hepatocellular carcinoma-related hub genes by bioinformatics analysis and CYP2C8 is a potential prognostic biomarker," *Gene*, vol. 698, pp. 9–18, 2019.

[19] G.-M. Liu, H.-D. Zeng, C.-Y. Zhang, and J.-W. Xu, "Identification of a six-gene signature predicting overall survival for hepatocellular carcinoma," *Cancer Cell International*, vol. 19, no. 1, p. 138, 2019.

[20] D. W.-H. Ho, A. K.-L. Kai, and I. O.-L. Ng, "TCGA whole-transcriptome sequencing data reveals significantly dysregulated genes and signaling pathways in hepatocellular carcinoma," *Frontiers of Medicine*, vol. 9, no. 3, pp. 322–330, 2015.

[21] J. Kong, T. Wang, Z. Zhang, X. Yang, S. Shen, and W. Wang, "Five core genes related to the progression and prognosis of hepatocellular carcinoma identified by analysis of a coexpression network," *DNA and Cell Biology*, vol. 38, no. 12, pp. 1564–1576, 2019.

[22] L. Zhang, Y. Guo, B. Li et al., "Identification of biomarkers for hepatocellular carcinoma using network-based bioinformatics methods," *European Journal of Medical Research*, vol. 18, no. 1, p. 35, 2013.

[23] S.-L. Chen, C. Z. Zhang, L.-L. Liu et al., "A GYS2/p53 negative feedback loop restricts tumor growth in HBV-related hepatocellular carcinoma," *Cancer Research*, vol. 79, no. 3, pp. 534–545, 2019.

[24] C. Cotta-Ramusino, D. Fachinetti, C. Lucca et al., "Exo1 processes stalled replication forks and counteracts fork reversal in checkpoint-defective cells," *Molecular Cell*, vol. 17, no. 1, pp. 153–159, 2005.

[25] P. Szankasi and G. R. Smith, "A role for exonuclease I from S. pombe in mutation avoidance and mismatch correction," *Science*, vol. 267, no. 5201, pp. 1166–1169, 1995.

[26] G. Keijzers, D. Bakula, M. A. Petr et al., "Human exonuclease 1 (EXO1) regulatory functions in DNA replication with putative roles in cancer," *International Journal of Molecular Sciences*, vol. 20, 2018.

[27] F. Song, A. A. Qureshi, J. Zhan et al., "Exonuclease 1 (EXO1) gene variation and melanoma risk," *DNA Repair*, vol. 11, no. 3, pp. 304–309, 2012.

[28] F. Luo, Y. Z. Wang, D. Lin, J. Li, and K. Yang, "Exonuclease 1 expression is associated with clinical progression, metastasis, and survival prognosis of prostate cancer," *Journal of Cellular Biochemistry*, vol. 120, no. 7, pp. 11383–11389, 2019.

[29] C. Kretschmer, A. Sterner-Kock, F. Siedentopf, W. Schoenegg, P. M. Schlag, and W. Kemmner, "Identification of early molecular markers for breast cancer," *Molecular Cancer*, vol. 10, no. 1, p. 15, 2011.

[30] T. Shi, R. Jiang, P. Wang et al., "Significant association of the EXO1 rs851797 polymorphism with clinical outcome of ovarian cancer," *OncoTargets and Therapy*, vol. 10, pp. 4841–4851, 2017.

[31] Y. Dai, Z. Tang, Z. Yang et al., "EXO1 overexpression is associated with poor prognosis of hepatocellular carcinoma patients," *Cell Cycle*, vol. 17, no. 19-20, pp. 2386–2397, 2018.

[32] J. Liu and J. Zhang, "Elevated EXO1 expression is associated with breast carcinogenesis and poor prognosis," *Annals of Translational Medicine*, vol. 9, no. 2, p. 135, 2021.

[33] N. L. Makia and J. A. Goldstein, "CYP2C8 is a novel target of peroxisome proliferator-activated receptor alpha in human liver," *Molecular Pharmacology*, vol. 89, no. 1, pp. 154–164, 2016.

[34] X. Ren, Y. Ji, X. Jiang, and X. Qi, "Downregulation of CYP2A6 and CYP2C8 in tumor tissues is linked to worse overall survival and recurrence-free survival from hepatocellular carcinoma," *BioMed Research International*, vol. 2018, Article ID 5859415, 2018.

[35] K. Li and Y. Chen, "CYP2C8 regulated by GAS5/miR-382-3p exerts anti-cancerous properties in liver cancer," *Cancer Biology & Therapy*, vol. 21, no. 12, pp. 1145–1153, 2020.

[36] J. T. Backman, A. M. Filppula, M. Niemi, and P. J. Neuvonen, "Role of cytochrome P450 2C8 in drug metabolism and interactions," *Pharmacological Reviews*, vol. 68, no. 1, pp. 168–241, 2016.

[37] Y. Wang, M. Wang, H. Qi et al., "Pathway-dependent inhibition of paclitaxel hydroxylation by kinase inhibitors and assessment of drug-drug interaction potentials," *Drug Metabolism and Disposition*, vol. 42, no. 4, pp. 782–795, 2014.

[38] L. Navarro-Nunez, S. A. Langan, G. B. Nash, and S. P. Watson, "The physiological and pathophysiological roles of platelet CLEC-2," *Thrombosis & Haemostasis*, vol. 109, pp. 991–998, 2013.

[39] K. Suzuki-Inoue, O. Inoue, and Y. Ozaki, "The novel platelet activation receptor CLEC-2," *Platelets*, vol. 22, no. 5, pp. 380–384, 2011.

[40] R. Critelli, F. Milosa, F. Faillaci et al., "Microenvironment inflammatory infiltrate drives growth speed and outcome of hepatocellular carcinoma: a prospective clinical study," *Cell Death & Disease*, vol. 8, no. 8, Article ID e3017, 2017.

[41] L. Wang, J. Yin, X. Wang et al., "C-type lectin-like receptor 2 suppresses AKT signaling and invasive activities of gastric cancer cells by blocking expression of phosphoinositide 3-kinase subunits," *Gastroenterology*, vol. 150, no. 5, pp. 1183–1195, 2016.

[42] K. Hu, Z.-M. Wang, J.-N. Li, S. Zhang, Z.-F. Xiao, and Y.-M. Tao, "CLEC1B expression and PD-L1 expression predict clinical outcome in hepatocellular carcinoma with tumor hemorrhage," *Translational Oncology*, vol. 11, no. 2, pp. 552–558, 2018.

[43] W. Xue, B. Dong, Y. Zhao et al., "Upregulation of TTYH3 promotes epithelial-to-mesenchymal transition through Wnt/beta-catenin signaling and inhibits apoptosis in cholangiocarcinoma," *Cellular Oncology*, vol. 44, no. 6, pp. 1351–1361, 2021.