Research Article

Screening of Prognostic Markers for Hepatocellular Carcinoma Patients Based on Multichip Combined Analysis

Yong Dong[1], **Qian Miao**[2], **and Da Li**[1]

[1]*Department of Medical Oncology, Sir Run Run Shaw Hospital, College of Medicine, Zhejiang University, No. 3 Qingchun East Road, Jianggan District, Hangzhou, Zhejiang 310016, China*
[2]*Department of Medical Oncology, Quzhou People's Hospital, No. 2, Zhongloudi, Kecheng District, Quzhou, Zhejiang 324000, China*

Correspondence should be addressed to Da Li; lidaonconew@zju.edu.cn

Object. Due to challenges in HCC prognosis evaluation, novel prognosis-related markers are urgently needed. This study attempted to screen reliable prognostic HCC prognostic markers by bioinformatics methods. *Methods*. GSE (14520, 36376, 57957, 76427) datasets were accessed from GEO database. 55 differential mRNAs (DEGs) were obtained by differential analysis based on the datasets. GO and KEGG analysis results indicated that the DEGs were enriched in xenobiotic metabolic process and other pathways. Expression profiles and clinical data of TCGA-LIHC mRNAs were from TCGA database. We established a prognostic model of HCC through univariate and multivariate Cox risk regression analyses. ROC curve analysis was used to examine the prognostic model performance. GSEA analysis was performed between the high- and low-risk score sample groups. *Results*. A 4-gene HCC prognostic model was constructed, in which the gene expressions correlated to HCC patients' survival. The AUC value presented 0.734 in the ROC analysis for the prognostic model. *Conclusion*. The four-gene model could be introduced as an independent prognostic factors to assess HCC patients' survival status.

1. Introduction

Liver cancer is the sixth most commonly diagnosed malignancy, with increasing incidence and mortality year by year [1]. Primary liver cancer consists of 75%-85% hepatocellular carcinoma (HCC) and 10%-15% intrahepatic cholangiocarcinoma along with other rare types [2]. Clinical staging, age, and vascular invasion as important clinical factors may contribute to improvement of survival prediction [3]. Due to the complex molecular mechanism of tumor regulation, traditional clinical information prediction ability is limited. Though a lot of strategies such as hepatectomy, radiofrequency ablation, transcatheter arterial chemoembolization, liver transplantation, and chemotherapy are used in HCC treatment, HCC patient's prognosis is still poor [4]. About 70% of HCC patients have recurrence or metastasis within 5 years after surgery [5]. With the continuous improvement of tumor heterogeneity and molecular mechanism research, new tools like molecular markers are in urgent need for accurate prediction of patient's prognosis.

Genome sequencing technology and data emerged in this genomic era [6]. Microarray and bioinformatics analysis are applied to screen differentially expressed genes (DEGs) at the genomic level, which may aid in the identification of DEGs and pathways correlated with occurrence and development of HCC. Gene chips can rapidly detect DEGs, generate slicing data, and store them in a public database, which is a reliable technology [7]. Great contributions have been made to tumor diagnosis and prognosis prediction. On the basis of these data, valuable evidence can be found for new research. For example, through public databases, more and more potential biomarkers are being mined [8–10]. Feng et al. used lung adenocarcinoma (LUAD) dataset from TCGA and found that LUAD patients expressing high HMGB1 level had dismal overall survival (OS) [11]. Wang et al. take advantage of RNA sequencing (RNA-seq) data from TCGA-KIRC to confirm DEGs [12]. For HCC, Kong et al. selected the top 25% DEGs from the GSE62232 dataset and screened out prognostic-related modules to build protein-protein interaction (PPI) networks, who obtained 5 candidate genes (PCNA, RFC4, PTTG1, H2AFZ,

and RRM1) finally [13]. It is indicated that these fresh next-generation sequencing methods and data can screen biomarkers.

Despite the prognostic ability of signature genes being explored by bioinformatics analysis, these prognostic prediction effects are limited due to the heterogeneity of HCC [14]. These prognostic prediction effects are subject to certain challenges, and thus, it is necessary to develop reliable novel signature genes to design a more personalized diagnostic and therapeutic plan. In this study, we downloaded 4 sets of HCC mRNA chip data from GEO database and analyzed them to obtain the differentially expressed mRNAs (DEmRNAs) between HCC tissues and noncancer tissues. Subsequently, biological significance and key pathways of the DEGs were explored through GO and KEGG enrichment analyses. Cox regression analysis, risk assessment, and clinical information analysis helped us find markers that could predict HCC patients' prognoses. Here, a total of 55 DEmRNAs and 4 independent prognostic factors for HCC were identified, which may be candidate biomarkers for HCC, being utilized for prediction of HCC progression and prognosis. A new method for adjuvant therapy of clinical HCC patients was put forward in this study.

2. Materials and Methods

2.1. Data Downloading and Processing.
HCC-related matrix was downloaded from GEO database (https://www.ncbi.nlm.nih.gov/geo/). Criteria for obtaining matrix were (1) human HCC tissue samples, (2) cancer and noncancer samples, (3) the total number of samples ≥ 30, and (4) based on the same sequencing platform. Four groups of HCC mRNA chip data including GSE14520 (normal: 220, tumor: 225), GSE36376 (normal: 193, tumor: 240), GSE57957 (normal: 193, tumor: 240), and GSE76427 (normal: 52, tumor: 115) were selected for this research. The sequencing platform for GSE36376, GSE57957, and GSE76427 was GPL10558 Illumina HumanHT-12 V4.0 chip, and the sequencing platform for GSE14520 was GPL3921 HT_HG-U133A. Then, the combined differential analysis (logFC > |1.0|, padj < 0.05) was done with "limma" (PMID: 25605792) and "RobustRankAggreg" (PMID: 31638362) in R package to identify DEmRNAs in HCC tissues.

The mRNA expression data and clinical data of TCGA-LIHC were downloaded from TCGA database (https://portal.gdc.cancer.gov/). DEmRNA expression and clinical information screened by differential expression analyses were extracted for subsequent validation.

2.2. Functional Enrichment Analysis.
GO and KEGG enrichment analyses of DEGs were conducted using R package "clusterProfiler" (PMID: 22455463). GO enrichment analysis investigates the biological significance of DEmRNAs, including long-term enrichment analysis of biological processes, molecular functions, and cellular components. KEGG pathway enrichment analysis seeks key pathways closely related to DEmRNAs. P value < 0.05 and false discovery rate (FDR) < 0.05 were set as thresholds. P value was calculated based on hypergeometric separation.

2.3. Cox Regression Analysis and Model Assessment.
Univariate and multivariate Cox risk regression analyses were done on DEmRNAs using the R package "survival" [15]. At first, the clinical information provided by TCGA was used to conduct univariate Cox regression analysis on DEmRNAs, and the DEmRNAs with P value < 0.01 were selected (omnibus test). Then, these DEmRNAs were further analyzed by multivariate Cox regression analysis to establish a multigene risk model. Samples were divided into high- and low-risk groups by median risk score, and OS of two groups was compared by R package "survival". To assess the accuracy and predictive value of risk assessment model, we used R package "survivalROC" (PMID: 33790572) to draw receiver operating characteristic (ROC) curves and conduct independent ROC tests for genes in risk assessment model, respectively, to obtain area under curve (AUC) value.

2.4. GSEA Analysis for Hub Genes.
To investigate pathway changes in the high- and low-risk groups, the GSEA software (V 4.1.0) was utilized to perform KEGG enrichment pathway analysis to explore potential mechanism of action (PMID: 16199517). Normalized enrichment score (NES) and FDR were used to quantify enrichment and statistical significance, respectively [16].

2.5. Correlation Analysis of Risk Scores and Clinical Characteristics.
The expression levels of feature genes in the high- and low-risk groups were analyzed by using heat map, and their differences in different clinical information and risk score coefficients were analyzed. The performance of risk scores in different tumor stages was analyzed by box plots, and the reliability of the prognostic model in liver cancer was further validated. Kaplan-Meier survival curve was used to assess the effect of the combination of high- or low-risk score and tumor stage on survival rate (log-rank test was introduced for statistical test).

2.6. Expressions of Hub Genes and Survival Analysis.
GEPIA database (http://gepia.cancer-pku.cn/index.html) is an integration of RNA expression profiles of tumor and normal samples from TCGA and GTEx projects [17], which can be used for individual analysis. Kaplan-Meier plotter is a popular web tool for assessing the impact of a large number of genes on survival based on EGA, TCGA, and GEO databases [18]. In this study, GEPIA was utilized for expression confirmation and survival analysis of hub genes, and log-rank tests were conducted to measure statistical significance.

3. Results

3.1. Identification of DEmRNAs in HCC.
Four datasets GSE14520, GSE36376, GSE57957, and GSE76427 were obtained from GEO database. Subsequently, a differential gene expression analysis was performed among the 4 datasets and then integrated to remove batch effects; finally, 6 significantly upregulated mRNAs and 49 prominently downregulated mRNAs were screened (Figure 1). These DEGs were selected for further analysis.

3.2. GO and KEGG Enrichment Analyses for the DEmRNAs.
Then, functional enrichment analysis was conducted on

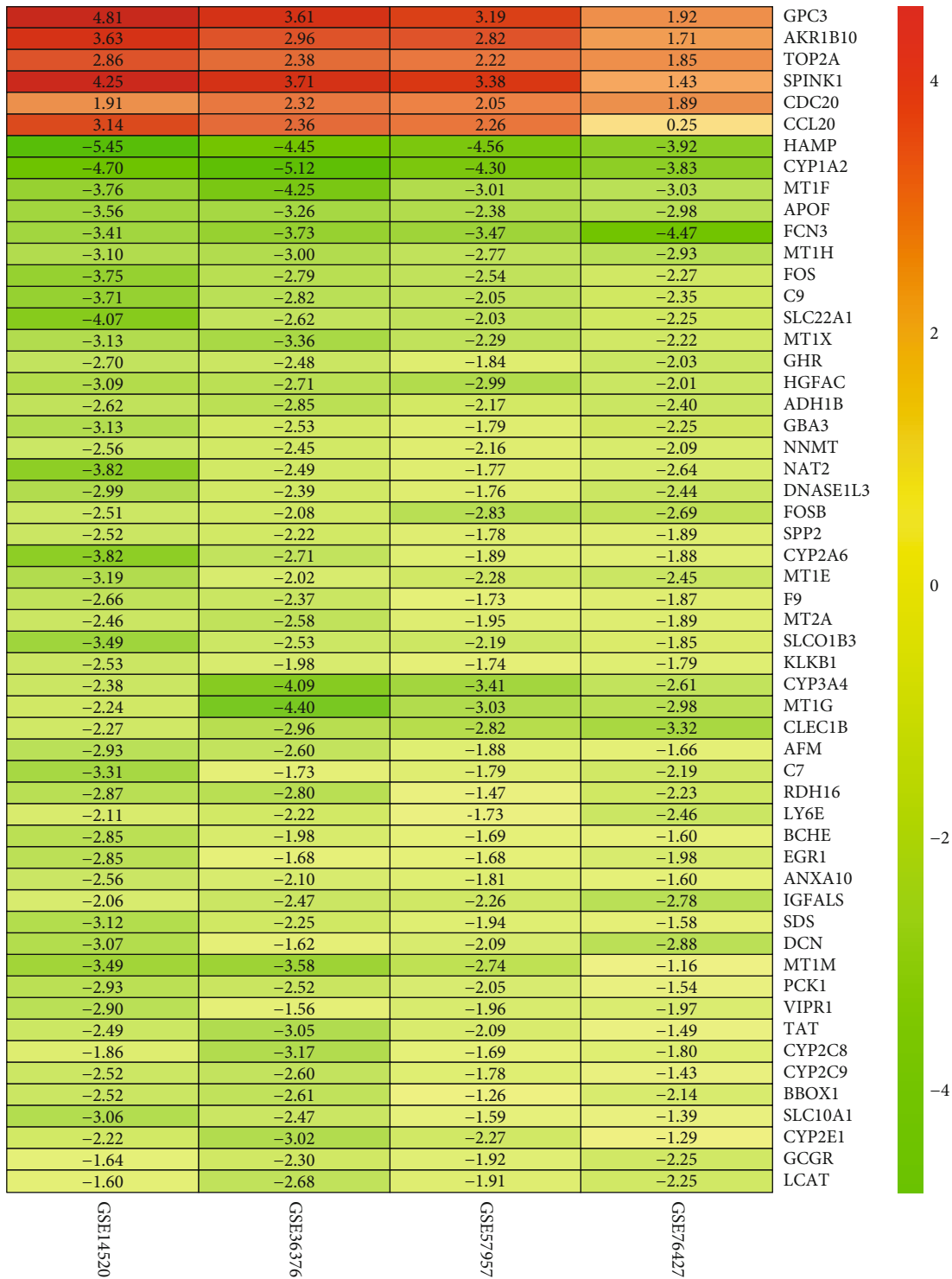| GSE14520	GSE36376	GSE57957	GSE76427	
4.81	3.61	3.19	1.92	GPC3
3.63	2.96	2.82	1.71	AKR1B10
2.86	2.38	2.22	1.85	TOP2A
4.25	3.71	3.38	1.43	SPINK1
1.91	2.32	2.05	1.89	CDC20
3.14	2.36	2.26	0.25	CCL20
−5.45	−4.45	-4.56	−3.92	HAMP
−4.70	−5.12	−4.30	−3.83	CYP1A2
−3.76	−4.25	−3.01	−3.03	MT1F
−3.56	−3.26	−2.38	−2.98	APOF
−3.41	−3.73	−3.47	−4.47	FCN3
−3.10	−3.00	−2.77	−2.93	MT1H
−3.75	−2.79	−2.54	−2.27	FOS
−3.71	−2.82	−2.05	−2.35	C9
−4.07	−2.62	−2.03	−2.25	SLC22A1
−3.13	−3.36	−2.29	−2.22	MT1X
−2.70	−2.48	−1.84	−2.03	GHR
−3.09	−2.71	−2.99	−2.01	HGFAC
−2.62	−2.85	−2.17	−2.40	ADH1B
−3.13	−2.53	−1.79	−2.25	GBA3
−2.56	−2.45	−2.16	−2.09	NNMT
−3.82	−2.49	−1.77	−2.64	NAT2
−2.99	−2.39	−1.76	−2.44	DNASE1L3
−2.51	−2.08	−2.83	−2.69	FOSB
−2.52	−2.22	−1.78	−1.89	SPP2
−3.82	−2.71	−1.89	−1.88	CYP2A6
−3.19	−2.02	−2.28	−2.45	MT1E
−2.66	−2.37	−1.73	−1.87	F9
−2.46	−2.58	−1.95	−1.89	MT2A
−3.49	−2.53	−2.19	−1.85	SLCO1B3
−2.53	−1.98	−1.74	−1.79	KLKB1
−2.38	−4.09	−3.41	−2.61	CYP3A4
−2.24	−4.40	−3.03	−2.98	MT1G
−2.27	−2.96	−2.82	−3.32	CLEC1B
−2.93	−2.60	−1.88	−1.66	AFM
−3.31	−1.73	−1.79	−2.19	C7
−2.87	−2.80	−1.47	−2.23	RDH16
−2.11	−2.22	-1.73	−2.46	LY6E
−2.85	−1.98	−1.69	−1.60	BCHE
−2.85	−1.68	−1.68	−1.98	EGR1
−2.56	−2.10	−1.81	−1.60	ANXA10
−2.06	−2.47	−2.26	−2.78	IGFALS
−3.12	−2.25	−1.94	−1.58	SDS
−3.07	−1.62	−2.09	−2.88	DCN
−3.49	−3.58	−2.74	−1.16	MT1M
−2.93	−2.52	−2.05	−1.54	PCK1
−2.90	−1.56	−1.96	−1.97	VIPR1
−2.49	−3.05	−2.09	−1.49	TAT
−1.86	−3.17	−1.69	−1.80	CYP2C8
−2.52	−2.60	−1.78	−1.43	CYP2C9
−2.52	−2.61	−1.26	−2.14	BBOX1
−3.06	−2.47	−1.59	−1.39	SLC10A1
−2.22	−3.02	−2.27	−1.29	CYP2E1
−1.64	−2.30	−1.92	−2.25	GCGR
−1.60	−2.68	−1.91	−2.25	LCAT

FIGURE 1: Heat map of DEGs screened by combined analysis of 4 datasets from GEP database. The numbers in the grids represent the fold changes of genes in different datasets between the tumor group and the normal group.

above 55 DEmRNAs to explore their potential biological functions. GO enrichment analysis revealed that these genes were mainly concentrated in cellular response to cadmium ion, detoxification of inorganic compound, and cellular response to zinc ion (Figure 2(a)). KEGG unveiled those genes were enriched mainly in mineral absorption, chemical carcinogenesis DNA adducts, and drug metabolism-cytochrome P450 (Figure 2(b)). Therefore, these DEmRNAs may influence the progression of HCC by influencing these biological functions and pathways.
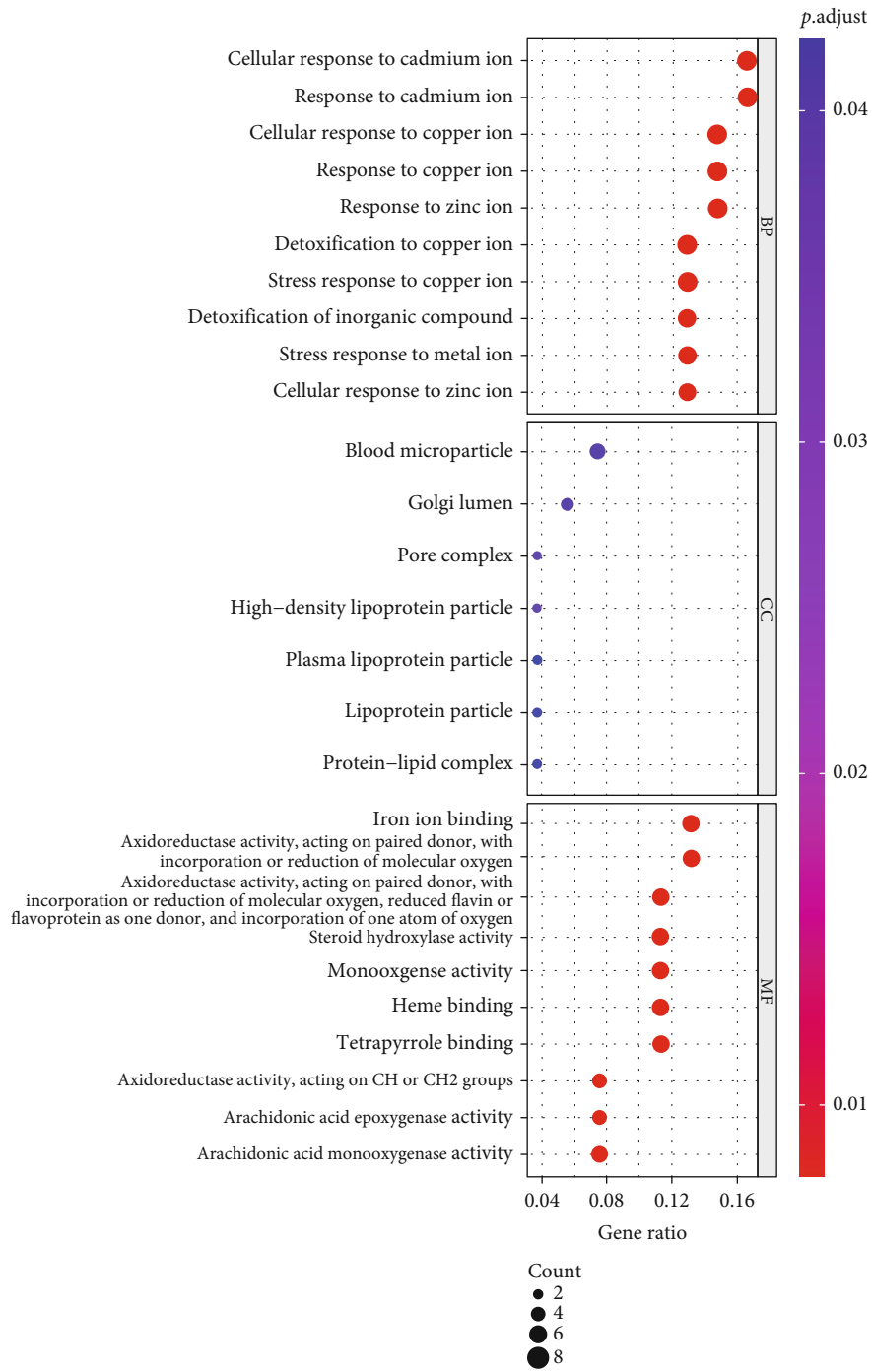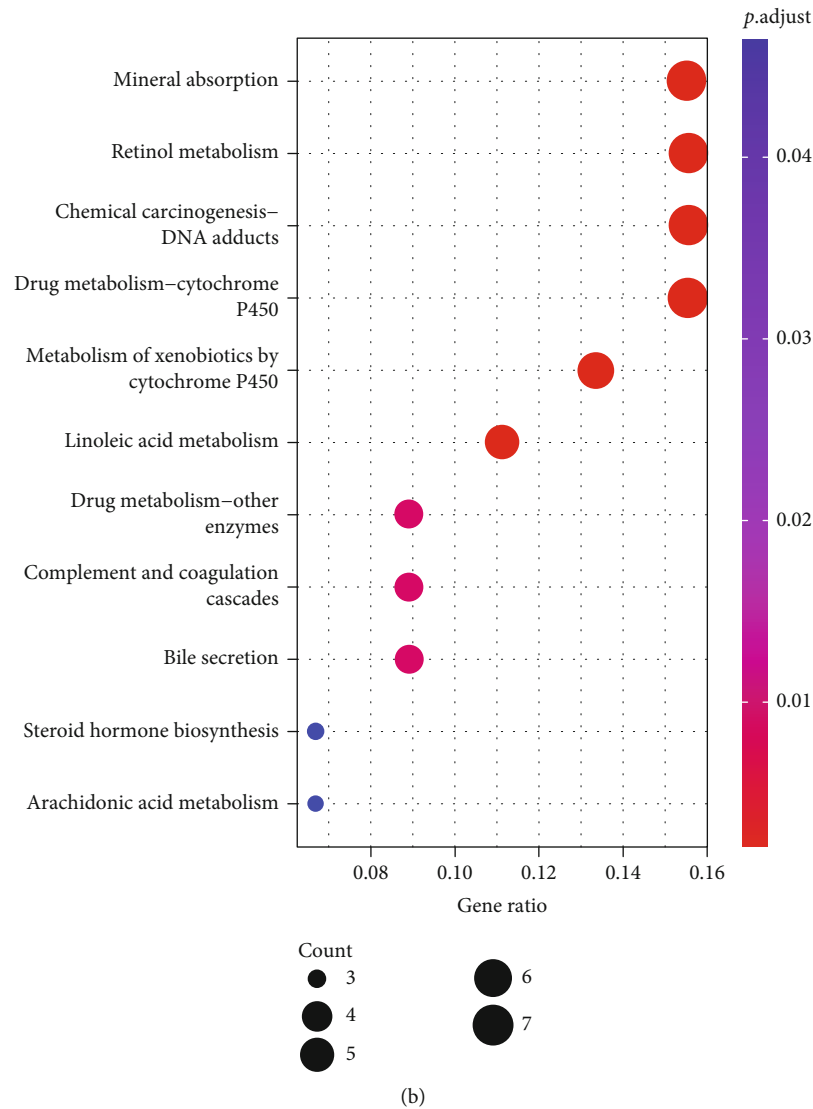
(a)

FIGURE 2: Continued.

(b)

FIGURE 2: GO and KEGG pathway analyses of DEmRNAs in HCC: (a, b) GO and KEGG enrichment analyses of DEGs.

3.3. Screening and Verifying the Optimal Prognostic Hub Genes. To screen out hub genes with the best prognosis, we carried out univariate Cox regression analysis on 55 DEmR-NAs and selected 20 genes noticeably associated with HCC patients' prognoses (Table 1). Multivariate Cox regression analysis was done on these 20 related genes to establish a risk model of four genes, among which CDC20 was a high-risk factor while CYP2C9, CLEC1B, and LCAT were low-risk factors (Figure 3(a)). Next, patients were assorted into high- and low-risk groups by median risk score. R package "survival" was utilized to compare survival time of two groups, and survival curves were plotted. A significant difference in OS was seen between the two groups, which demonstrated that survival time of patients in the high-risk group was dramatically shorter than the other group, indicating that the risk score could be used for risk grading and prognosis assessment of HCC patients (Figure 3(b)). The ROC curves manifested that AUC value of the model was 0.734 (Figure 3(c)), indicating

that the model had a certain accuracy in diagnosing patients' prognoses. Thus, risk scores of these four hub genes were dramatically associated with HCC patients' prognoses.

3.4. Functional Analysis of 4 Selected Genes via GSEA. In order to analyze function changes of the high- and low-risk groups, we used the GSEA software to carry out KEGG pathway analysis on the two groups, and the results displayed that compared with the low-risk group, complement and blood coagulation cascade, fatty acid metabolism, peroxidase, and primary bile acid biosynthesis pathways are significantly activated in the high-risk group (Figure 4). These pathways are closely related to the liver and are the major metabolic processes in the liver. Hence, changes in these pathways can affect the metabolic function of the liver. Therefore, we speculated that hub genes may affect the biological function of cancer cells by regulating the metabolic process of the liver, so as to play a role in prognosis.

TABLE 1: Univariate Cox regression analysis screened genes that were significantly associated with the prognosis of HCC patients.

Gene	HR	z	P value
CDC20	1.303474527	4.659540595	3.17E-06
LCAT	0.800676874	-4.422698682	9.75E-06
DNASE1L3	0.834137148	-4.228589383	2.35E-05
GHR	0.813122352	-4.187942095	2.81E-05
CYP2C9	0.887041102	-4.126833672	3.68E-05
ANXA10	0.877021937	-3.969016876	7.22E-05
SPP2	0.907350691	-3.964707036	7.35E-05
SLC10A1	0.914033265	-3.67849838	0.000234611
SLC22A1	0.909870575	-3.598475946	0.000320087
AFM	0.902676346	-3.557458905	0.00037446
TAT	0.912717953	-3.472545816	0.000515547
ADH1B	0.901724063	-3.429051233	0.000605695
KLKB1	0.842899464	-3.320604451	0.000898227
VIPR1	0.851232022	-3.312650071	0.000924165
TOP2A	1.210739633	3.308652423	0.000937461
RDH16	0.903679968	-3.278095563	0.0010451
CLEC1B	0.860469147	-3.090652798	0.00199717
FCN3	0.8707386	-2.958448406	0.00309192
C7	0.919120279	-2.896960425	0.003767973
IGFALS	0.912348703	-2.779238461	0.005448651

3.5. Relationship between the Hub Genes and Clinical Features. To validate the reliability of the prognostic model, we explored the performance of the risk score on clinical features. Heat map results showed that CDC20 level increased gradually, while CYP2C9, CLEC1B, and LCAT levels decreased gradually, with increased risk score (Figure 5(a)). Significant differences were seen in different grades, pathological stages, and T stages between high- and low-risk group (Figure 5(a)). Survival time of patients with risk score combined with different clinical stages, pathological grades, and T stages was significantly different (Figure 5(b)), which may be reference factors predicting patient's prognosis. With progression of tumor, the risk value increased (Figure 5(c)). This further indicated that the four-gene model could predict the risk of tumors, and the risk scores and some clinical features established by us (stage, grade, T) could be utilized to comprehensively evaluate patient's prognosis.

3.6. Analysis of Hub Genes by the Kaplan-Meier Plotter and GEPIA. GEPIA dataset was used to conduct survival and ROC analysis on 4 hub genes in cancer and normal tissues to test significance of the 4 hub genes to patients' prognosis. We searched the Kaplan-Meier OS curves and progression-free survival (PFS) curves of the four genes in the risk model in the GEPIA database and manifested that these four-gene levels had a dramatic impact on OS of patients (Figure 6(a)), among which the HCC patients with high expressions of CLEC1B, CYP2C9, and LCAT presented poor prognosis. Moreover, CDC20, CLEC1B, and LCAT expression levels also

had significant effects on PFS (Figure 6(b)). Independent ROC tests were conducted for the four genes in the risk model, and it was noted that the AUC value of each gene reached more than 0.6 (Figure 6(c)). Therefore, the above results suggested that the expressions of these four genes in HCC were related to the progression of cancer and correlated with prognosis, and these four genes could be used as independent prognostic factors for HCC.

4. Discussion

Liver cancer is a malignant tumor with different histological characteristics, which is mainly caused by chronic hepatitis virus infection, gene mutation, cell damage, alcoholic liver disease, and aflatoxin poisoning. Despite advances in cancer treatment, prognosis is unfavorable with increasing morbidity and mortality [19]. Most HCC patients without early detection are not suitable for radical treatment, which may be a reason for dismal prognosis. Many studies have demonstrated that biomarkers can determine the progression of cancer. For example, CCND1, c-myc, and RAS mutations, and hypermethylation of CCND2 promoter have been associated with HCC [20, 21]. Splicing changes of NT5E, Sulf1, and SLC39A14 were also associated with HCC [22–24]. All of these factors can be used as markers to predict the progression of HCC. However, the molecular mechanism of HCC is very complex, including many genetic and epigenetic changes, chromosomal aberrations, gene mutations, and alternative molecular pathways [25]. Therefore, potential and efficient diagnostic and therapeutic markers are in great need. Microarray technology enables the exploration of genetic changes in HCC and has been proved to be an effective strategy for mining novel biomarkers in other diseases.

We analyzed four microarray datasets by multichip combined analysis to obtain DEmRNAs between HCC and noncancer tissues. 55 DEmRNAs were identified, containing 6 upregulated genes and 49 downregulated mRNAs. It is well known that main metabolic pathways in the liver include bile acid synthesis, fatty acid metabolism, complement and coagulation cascade, etc., which play a vital role in disease and human homeostasis [26–29]. Therefore, we used GO and KEGG enrichment analyses to explore biological functions and pathway enrichment of DEmRNAs. The results exhibited that DEmRNAs were mainly concentrated in biological functions such as response to exotic biological stimulation, drug catabolism, complement activation, cAMP response, and signaling pathways such as cytochrome P450, drug metabolism, complement and coagulation cascade, fatty acid metabolism, peroxisome, and primary bile acid biosynthesis. Another study has suggested that the activation of complement can promote the tumor [30]. Oxidoreductase activity may lead to antioxidant defense and can encode tumor repressors that are often altered in tumors [31, 32]. The accumulation of bile acids in the liver inhibits cell growth and improves survival [33]. These conclusions are all consistent with our predictions.

Subsequently, Cox regression analysis and the construction of risk model obtained the 4-gene risk model consisting of CDC20, CLEC1B, CYP2C9, and LCAT. The risk values of these four-gene sets were evaluated and verified to be
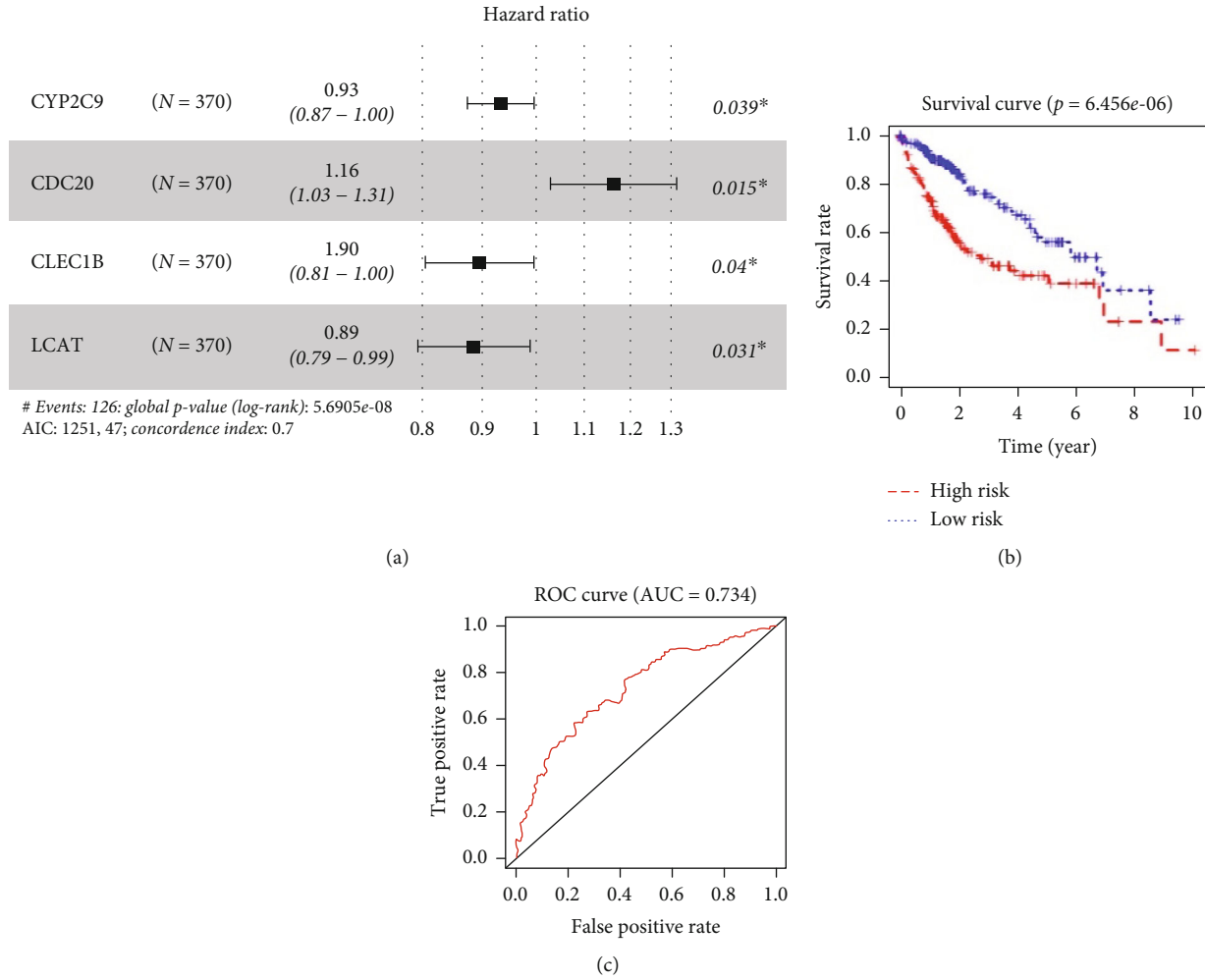
(a)

(b)

(c)

FIGURE 3: Constructing and verification the prognostic model. (a) Forest map of the 4-gene prognostic model established by multivariate Cox risk regression analysis; (b) Kaplan-Meier OS curves for the high- and low-risk groups are plotted based on risk score. Horizontal axis: survival time; vertical axis: overall rate; red line: high-risk samples; blue line: low-risk samples. (c) ROC curves based on the risk model.

prognostic factors for HCC. We also found from the extensive literature that these hub genes played an essential role in many cancers including HCC. Based on the gene expression and clinical information of liver cancer patients obtained from the TCGA database, Long et al. used univariate, LASSO, and multivariate Cox regression analyses to establish the prognostic model of liver cancer and used GEO database to verify the feasibility of the model. Finally, a prognostic model based on four marker genes is obtained, and the model can predict the overall survival of liver cancer patients [34]. Similarly, in our study, based on 4 GSE datasets in the GEO database, the liver cancer prognostic model was established by univariate and multivariate Cox regression models and finally verified by the GEPIA database. Since we built the model based on multiple gene sets, the results of this study are more reliable to a certain extent as compared with the earlier studies. Dysregulation of cell cycle processes is vital in the development of tumors [20, 35, 36]. CDC20 is an important regulatory factor in the cell cycle process, which forms ubiquitin-proteolytic enzyme complex

by binding with APC and participates in the degradation of various proteins to modulate cell cycle process [37]. It has been reported that upregulation of CDC20 may predict the decrease of OS and DFS in HCC patients [38]. Increased CDC20 in HCC was related to gender, differentiation, and TNM stage [39], which is similar to our results. Through risk model, we manifested that CDC20 was a high-risk factor and was also associated with clinical features. CYP2C9 is a drug metabolism enzyme and is a decreased low-risk factor in HCC [40]. Shuaichen and Guangyi have discovered that CYP2C9 may promote the development of HCC, especially can be a diagnostic biomarker in drug metabolism [41]. CLEC2 is expressed in platelets and some hematopoietic cells. Wang et al. have put forward that CLEC2 inhibits gastric cancer metastasis, prevents the activation of AKT and glycogen synthase signals, and affects the invasion and expression of EMT markers, which can be a potential biomarker in gastric cancer [42]. However, fewer studies on CLEC2 in HCC have been reported. So, our study provided a data source for the role of this gene in HCC. Finally, LCAT
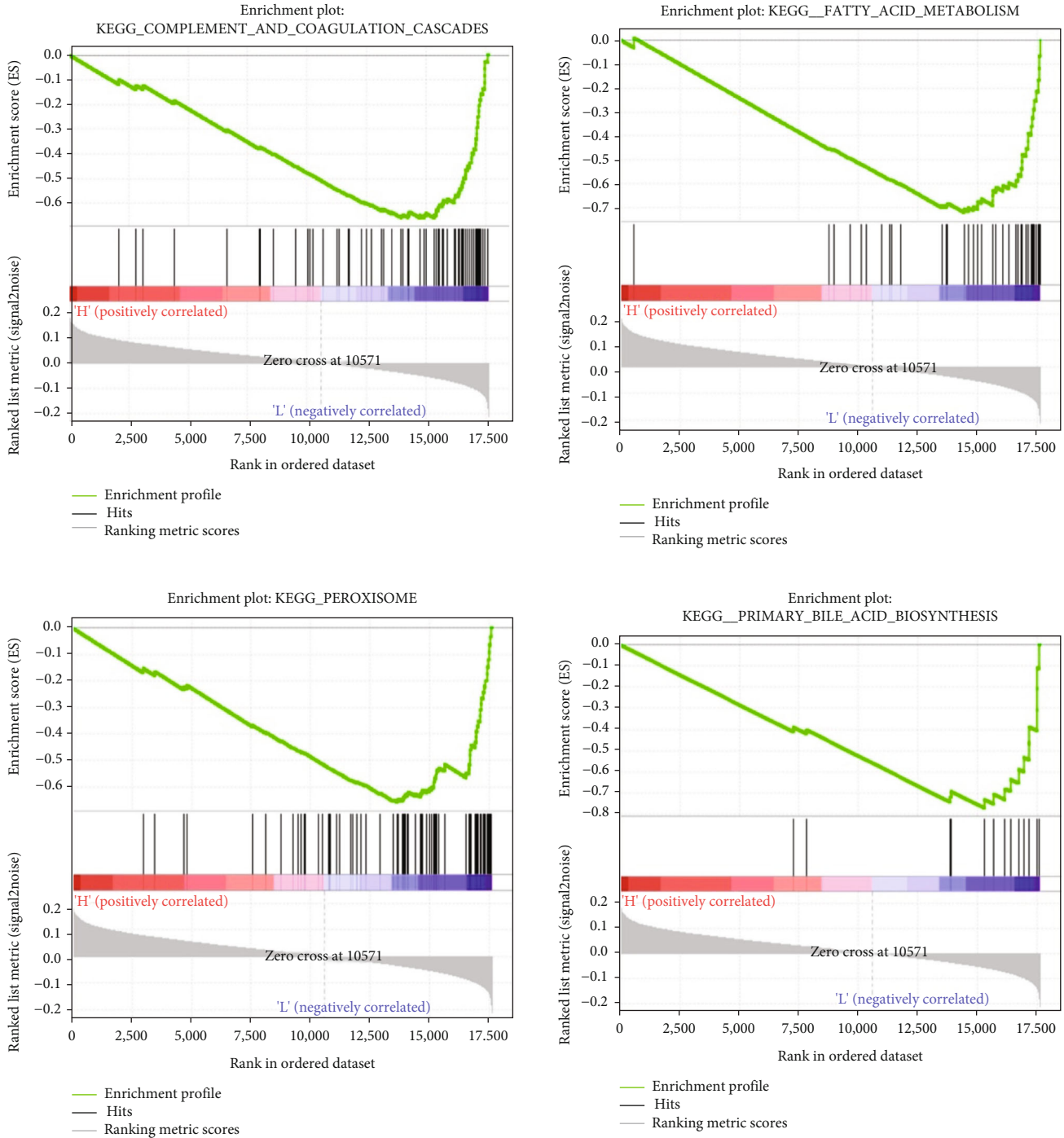
Figure 4: GSEA pathway enrichment results based on median risk score grouping.

is the only enzyme that can esterify cholesterol in plasma, which determines maturation of high-density lipoprotein as a key enzyme for reverse cholesterol transport with reports of its role in atherosclerosis, cholesterol deposition, and kidney [43–45]. Besides, Russell et al. have pointed out that LCAT plays a role as a diagnostic marker in epithelial ovarian cancer [46]. It can also predict OS of HCC [47]. In conclusion, the four hub genes found in this study played a crucial role in HCC. These genes also function in pathways such as complement and blood coagulation cascade, fatty acid metabolism, peroxisome, and primary bile acid biosynthesis. Their expressions in HCC will indirectly affect the changes of these pathways, thus affecting the incidence of HCC, which fully illustrates the importance of these hub genes of HCC progression.

Taken together, we attempted to screen DEmRNAs that may be related to HCC occurrence and development. 55 DEmRNAs and 4 hub genes had been identified as diagnostic
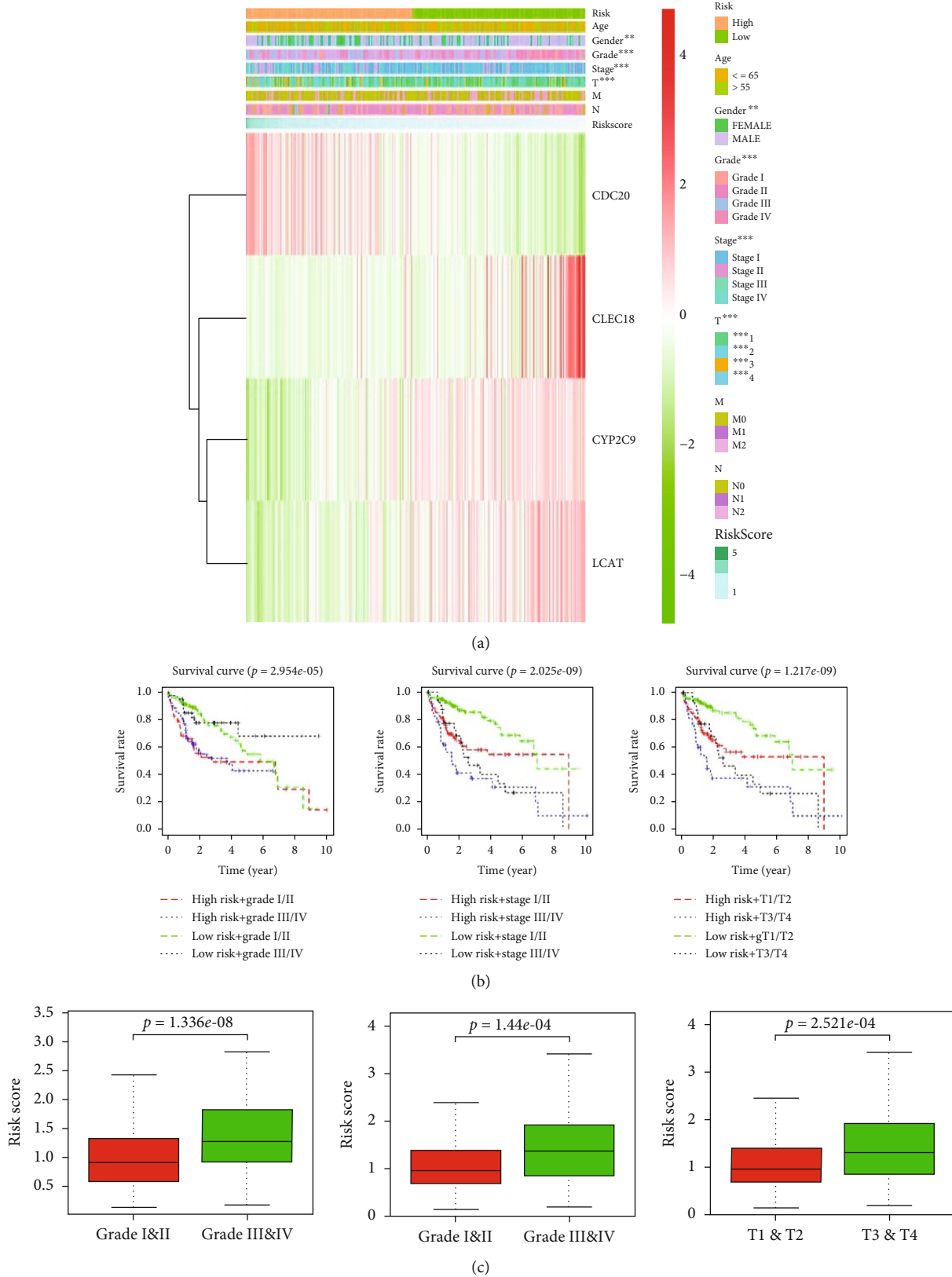
FIGURE 5: Analysis of DEmRNAs and clinical relationship and patients' prognosis. (a) Heat map of 4 gene expressions in the high- and low-risk groups of risk model and clinicopathological differences between the two groups; (b) OS curves combined risk score and the patients' different tumor stages, with the horizontal axis representing survival time and the vertical axis representing OS rate. Different color curves represent different combinations of risk and tumor stage. (c) Boxplots of differences in risk scores of different clinical stages, grades, and T stages.
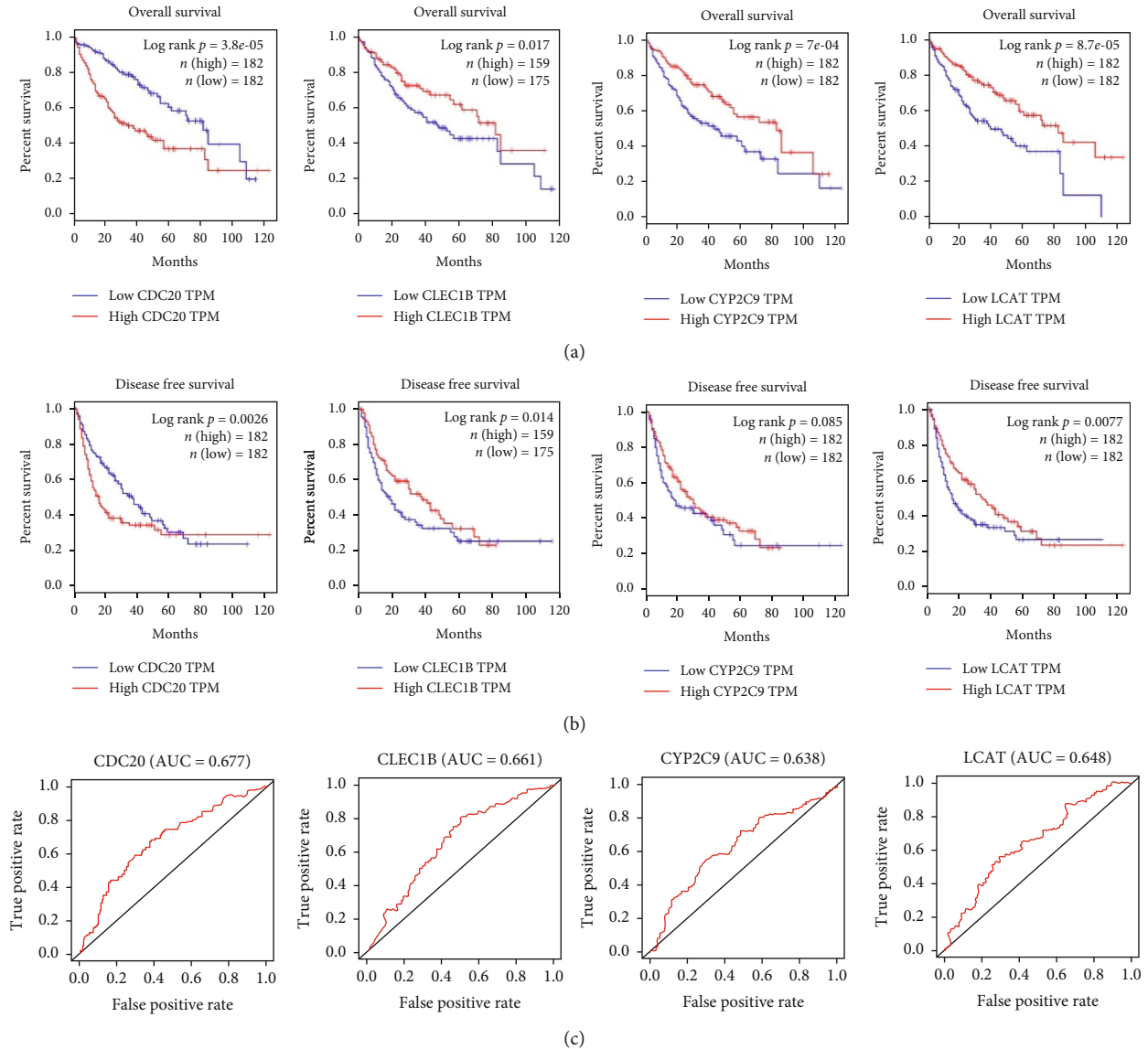
FIGURE 6: Analysis of hub genes by the Kaplan-Meier plotter and GEPIA. (a, b) Kaplan-Meier OS curves and PFS curves of four genes in the risk model are retrieved in the GEPIA database. Horizontal axis: survival time; vertical axis: survival rate; red line: high-expression group; blue line: low-expression group. (c) Independent ROC curves of four genes in the risk model for HCC patient's prognosis.

biomarkers for HCC, which offered an effective basis for the treatment of HCC. But this experiment still has certain limitations. This study only conducted a pure bioinformatics analysis through the GEO database, further verification in multicenter clinical trials, and prospective studies is required. Nevertheless, biological functions of these genes in HCC need further study.

Data Availability

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

Consent

All authors consent to submit the manuscript for publication.

Conflicts of Interest

The authors declare that they have no potential conflicts of interest.

Authors' Contributions

DY and LD both contribute to the work, including conception and design and article drafting and revising. MQ is the guarantor for the article who accept full responsibility for the work.

References

[1] H. B. El-Serag and K. L. Rudolph, "Hepatocellular carcinoma: epidemiology and molecular carcinogenesis," *Gastroenterology*, vol. 132, no. 7, pp. 2557–2576, 2007.

[2] 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.

[3] K. Ikeda, S. Saitoh, A. Tsubota et al., "Risk factors for tumor recurrence and prognosis after curative resection of hepatocellular carcinoma," *Cancer*, vol. 71, no. 1, pp. 19–25, 1993.

[4] A. Villanueva, Y. Hoshida, C. Battiston et al., "Combining clinical, pathology, and gene expression data to predict recurrence of hepatocellular carcinoma," *Gastroenterology*, vol. 140, no. 5, pp. 1501–1512.e2, 2011.

[5] F. Ji, Y. Liang, S. Fu et al., "Prognostic value of combined preoperative prognostic nutritional index and body mass index in HCC after hepatectomy," *HPB : the official journal of the International Hepato Pancreato Biliary Association*, vol. 19, no. 8, pp. 695–705, 2017.

[6] Z. Wang, M. Gerstein, and M. Snyder, "RNA-Seq: a revolutionary tool for transcriptomics," *Nature Reviews Genetics*, vol. 10, no. 1, pp. 57–63, 2009.

[7] B. Vogelstein, N. Papadopoulos, V. E. Velculescu, S. Zhou, L. A. Diaz, and K. W. Kinzler, "Cancer genome landscapes," *Science*, vol. 339, no. 6127, pp. 1546–1558, 2013.

[8] B. Li, K. Pu, and X. Wu, "Identifying novel biomarkers in hepatocellular carcinoma by weighted gene co-expression network analysis," *Journal of cellular biochemistry*, vol. 120, no. 7, pp. 11418–11431, 2019.

[9] E. Martinez-Ledesma, R. G. Verhaak, and V. Treviño, "Identification of a multi-cancer gene expression biomarker for cancer clinical outcomes using a network-based algorithm," *Scientific Reports*, vol. 5, no. 1, p. 11966, 2015.

[10] D. Huang, W. Sun, Y. Zhou et al., "Mutations of key driver genes in colorectal cancer progression and metastasis," *Cancer Metastasis Reviews*, vol. 37, no. 1, pp. 173–187, 2018.

[11] A. Feng, Z. Tu, and B. Yin, "The effect of HMGB1 on the clinicopathological and prognostic features of non-small cell lung cancer," *Oncotarget*, vol. 7, pp. 20507–20519, 2016.

[12] Y. Wang, L. Chen, G. Wang et al., "Fifteen hub genes associated with progression and prognosis of clear cell renal cell carcinoma identified by coexpression analysis," *Journal of Cellular Physiology*, vol. 234, no. 7, pp. 10225–10237, 2019.

[13] 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.

[14] Z. Wang, L. Pan, D. Guo et al., "A novel five-gene signature predicts overall survival of patients with hepatocellular carcinoma," *Cancer Medicine*, vol. 10, no. 11, pp. 3808–3821, 2021.

[15] T. M. Therneau and P. M. Grambsch, *Modeling Survival Data: Extending the Cox Model*, Springer, New York, 2000.

[16] A. Eraso-Pichot, M. Brasó-Vives, A. Golbano et al., "GSEA of mouse and human mitochondriomes reveals fatty acid oxidation in astrocytes," *Glia*, vol. 66, no. 8, pp. 1724–1735, 2018.

[17] Z. Tang, C. Li, B. Kang, G. Gao, C. Li, and Z. Zhang, "GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses," *Nucleic Acids Research*, vol. 45, no. W1, pp. W98–w102, 2017.

[18] A. M. Szász, A. Lánczky, Á. Nagy et al., "Cross-validation of survival associated biomarkers in gastric cancer using transcriptomic data of 1,065 patients," *Oncotarget*, vol. 7, no. 31, pp. 49322–49333, 2016.

[19] D. Sia, A. Villanueva, S. L. Friedman, and J. M. Llovet, "Liver cancer cell of origin, molecular class, and effects on patient prognosis," *Gastroenterology*, vol. 152, no. 4, pp. 745–761, 2017.

[20] Y. Wang, J. Cheng, C. Xu et al., "Quantitative methylation analysis reveals gender and age differences in p16INK4a hypermethylation in hepatitis B virus-related hepatocellular carcinoma," *Liver international : official journal of the International Association for the Study of the Liver*, vol. 32, no. 3, pp. 420–428, 2011.

[21] M. Gorath, T. Stahnke, T. Mronga, O. Goldbaum, and C. Richter-Landsberg, "Developmental changes of tau protein and mRNA in cultured rat brain oligodendrocytes," *Glia*, vol. 36, no. 1, pp. 89–101, 2001.

[22] N. T. Snider, P. J. Altshuler, S. Wan, T. H. Welling, J. Cavalcoli, and M. B. Omary, "Alternative splicing of human NT5E in cirrhosis and hepatocellular carcinoma produces a negative regulator of ecto-5'-nucleotidase (CD73)," *Molecular Biology of the Cell*, vol. 25, no. 25, pp. 4024–4033, 2014.

[23] R. B. Gill, A. Day, A. Barstow, G. Zaman, C. Chenu, and G. K. Dhoot, "Mammalian Sulf1 RNA alternative splicing and its significance to tumour growth regulation," *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine*, vol. 33, no. 5, pp. 1669–1680, 2012.

[24] R. B. Franklin, B. A. Levy, J. Zou et al., "ZIP14 zinc transporter downregulation and zinc depletion in the development and progression of hepatocellular cancer," *Journal of Gastrointestinal Cancer*, vol. 43, no. 2, pp. 249–257, 2012.

[25] S. T. Awwad and J. P. McCulley, "Wavefront-guided LASIK: recent developments and results," *International Ophthalmology Clinics*, vol. 46, no. 3, pp. 27–38, 2006.

[26] C. M. Pusec, A. de Jesus, M. W. Khan et al., "Hepatic HKDC1 expression contributes to liver metabolism," *Endocrinology*, vol. 160, no. 2, pp. 313–330, 2019.

[27] J. Muciño-Bermejo, R. Carrillo-Esper, M. Uribe, and N. Méndez-Sánchez, "Coagulation abnormalities in the cirrhotic patient," *Annals of Hepatology*, vol. 12, no. 5, pp. 713–724, 2013.

[28] W. Li, W. Zhang, M. Chang et al., "Metabonomics reveals that triclocarban affects liver metabolism by affecting glucose metabolism, β-oxidation of fatty acids, and the TCA cycle in male mice," *Toxicology Letters*, vol. 299, pp. 76–85, 2018.

[29] M. Li, S. Y. Cai, and J. L. Boyer, "Mechanisms of bile acid mediated inflammation in the liver," *Molecular Aspects of Medicine*, vol. 56, pp. 45–53, 2017.

[30] M. M. Markiewski and J. D. Lambris, "Unwelcome complement," *Cancer Research*, vol. 69, no. 16, pp. 6367–6370, 2009.

[31] M. Abu-Remaileh and R. I. Aqeilan, "The tumor suppressor WW domain-containing oxidoreductase modulates cell metabolism," *Experimental biology and medicine*, vol. 240, no. 3, pp. 345–350, 2015.

[32] P. Roszczenko, K. A. Radomska, E. Wywial, J. F. Collet, and E. K. Jagusztyn-Krynicka, "A novel insight into the oxidoreductase activity of Helicobacter pylori HP0231 protein," *PLoS One*, vol. 7, no. 10, article e46563, 2012.

[33] W. Tang, V. Putluri, C. R. Ambati, T. H. Dorsey, N. Putluri, and S. Ambs, "Liver- and microbiome-derived bile acids accumulate in human breast tumors and inhibit growth and improve patient survival," *Clinical cancer research : an official*

journal of the American Association for Cancer Research, vol. 25, no. 19, pp. 5972–5983, 2019.

[34] J. Long, L. Zhang, X. Wan et al., "A four-gene-based prognostic model predicts overall survival in patients with hepatocellular carcinoma," *Journal of Cellular and Molecular Medicine*, vol. 22, no. 12, pp. 5928–5938, 2018.

[35] Y. L. Choi, S. H. Park, J. J. Jang, and C. K. Park, "Expression of the G1-S modulators in hepatitis B virus-related hepatocellular carcinoma and dysplastic nodule: association of cyclin D1 and p53 proteins with the progression of hepatocellular carcinoma," *Journal of Korean Medical Science*, vol. 16, no. 4, pp. 424–432, 2001.

[36] V. Tripathi, Z. Shen, A. Chakraborty et al., "Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB," *PLoS Genetics*, vol. 9, no. 3, article e1003368, 2013.

[37] L. Wang, J. Zhang, L. Wan, X. Zhou, Z. Wang, and W. Wei, "Targeting Cdc20 as a novel cancer therapeutic strategy," *Pharmacology & Therapeutics*, vol. 151, pp. 141–151, 2015.

[38] L. Zhuang, Z. Yang, and Z. Meng, "Upregulation of BUB1B, CCNB1, CDC7, CDC20, and MCM3 in tumor tissues predicted worse overall survival and disease-free survival in hepatocellular carcinoma patients," *BioMed Research International*, vol. 2018, Article ID 7897346, 8 pages, 2018.

[39] N. Rahman, "Mainstreaming genetic testing of cancer predisposition genes," *Clinical medicine*, vol. 14, pp. 436–439, 2014.

[40] D. Yu, B. Green, A. Marrone et al., "Suppression of CYP2C9 by microRNA hsa-miR-128-3p in human liver cells and association with hepatocellular carcinoma," *Scientific Reports*, vol. 5, no. 1, p. 8534, 2015.

[41] L. Shuaichen and W. Guangyi, "Bioinformatic analysis reveals CYP2C9 as a potential prognostic marker for HCC and liver cancer cell lines suitable for its mechanism study," *Cellular and molecular biology*, vol. 64, no. 7, pp. 70–74, 2018.

[42] 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.e16, 2016.

[43] A. Ossoli, S. Simonelli, C. Vitali, G. Franceschini, and L. Calabresi, "Role of LCAT in atherosclerosis," *Journal of Atherosclerosis and Thrombosis*, vol. 23, no. 2, pp. 119–127, 2016.

[44] R. Flores, X. Jin, J. Chang et al., "LCAT, ApoD, and ApoA1 expression and review of cholesterol deposition in the cornea," *Biomolecules*, vol. 9, no. 12, p. 785, 2019.

[45] A. Ossoli, E. B. Neufeld, S. G. Thacker et al., "Lipoprotein X causes renal disease in LCAT deficiency," *PLoS One*, vol. 11, no. 2, article e0150083, 2016.

[46] M. R. Russell, C. Graham, A. D'Amato et al., "Diagnosis of epithelial ovarian cancer using a combined protein biomarker panel," *British Journal of Cancer*, vol. 121, no. 6, pp. 483–489, 2019.

[47] Y. Zheng, Y. Liu, S. Zhao et al., "Large-scale analysis reveals a novel risk score to predict overall survival in hepatocellular carcinoma," *Cancer Management and Research*, vol. Volume 10, pp. 6079–6096, 2018.