Cancer Cell Research
35 (2022) 862-869
ISSN 2161-2609

Analysis of the expression and prognostic value of the KDM5C gene in hepatocellular carcinoma based on transcriptome sequencing

Zheng Wang
Shandong First Medical University, Jinan, Shandong 250117, China

Abstract: Hepatocellular carcinoma (HCC) is the sixth most common cancer in the world. Hepatocellular carcinoma (HCC) has high tumor heterogeneity. Although systemic treatments such as surgery and chemotherapy have made significant progress, the overall prognosis is poor. The study of liver cancer targeting genes has become a hot spot in the field of liver cancer research. In this paper, the transcriptome data of liver cancer patients enrolled in this research center were obtained, and verified by bioinformatics analysis and TCGA database, the KDM5C gene affecting the prognosis of liver cancer was screened and the clinical prognosis model was constructed.

Keywords: hepatocellular carcinoma; transcriptome sequencing; bioinformatics; prognostic analysis
Received 13 December 2022, Accepted 21 December 2022

1.Introduction

Hepatocellular carcinoma (hereinafter referred to as liver cancer) is the sixth most common cancer in the world and the third cause of cancer death [1,2]. China is also a big country of hepatitis B, about 50% of the new liver cancer and death cases occur in China every year, has become a major disease affecting the health of the Chinese people [3,4]. Surgical treatment and liver transplantation are still the first choice of treatment methods for early liver cancer. However, due to the disease characteristics of liver cancer, most patients are already in the middle and late stage of the disease at the first time, and lose the opportunity to undergo radical surgery, and the prognosis is very poor. Alpha-fetoprotein (AFP) is currently a blood test method for non-invasive diagnosis of liver cancer and a monitoring index of postoperative recurrence, but its clinical application is limited by sensitivity and specificity [5]. Despite the field of local therapy, hepatic artery perfusion chemotherapy based on mFOLFOX regimen (hepatic artery infusion chemotherapy, HAIC) has developed rapidly in recent years, the combination with targeted drugs and immunodrugs for advanced unresectable liver cancer can achieve an objective response rate of more than 30% (ORR), and the median survival period of patients once exceeded 20 months [6-8]. However, due to the heterogeneity of HCC, there are still a large number of HCC patients with poor treatment effects, and in-depth exploration of their drug resistance and metastasis mechanisms is expected to have important clinical significance to improving the prognosis of HCC patients. Transcriptomics technology is an important means to study cell phenotype and function, and bioinformatics has also become an indispensable and important part of tumor-related research. By screening abnormal differentially expressed genes, bioinformatics technology can be used to find the correlation with liver cancer occurrence, metastasis and recurrence, so as to provide support for early cancer diagnosis and patient prognosis analysis, and the reanalysis of clinical data can develop new diagnostic and treatment strategies for patients.

In this study selected our 20 routine radical resection of liver cancer patients with tumor samples and adjacent tissue samples, using transcriptome sequencing technology to obtain transcriptome data, combined with biological information analysis method, the screening of genes with obvious difference expression, in the TCGA data set to verify its expression and correlation with the prognosis of liver cancer, explore the feasibility as a target for diagnosis or treatment, so as to provide theoretical support for the individualized treatment of liver cancer.

Materials and Methods

1. 1 Transcriptome sequencing of the tissue samples

The tissue samples and adjacent tissue samples from 20 patients undergoing radical resection in our hospital were selected, and the high-throughput sequencing technology was applied, and the relevant differentially expressed genes in liver cancer tissues were selected. Informed consent was obtained from the patients and approved by the Ethics Committee of Qianfoshan Hospital in Shandong Province.

1.2 Download and collation of TCGA data set

Transcriptomic data and clinicopathological information for 374 HCC patients were downloaded from the The Cancer Genome Atlas database (TCGAhttps://portal.gdc.cancer.gov/). RNA sequencing data in FPKM format were converted to transcript per million reads (TPM) format.

35 (2022) 862-869

ISSN 2161-2609

1.3 Screening of the differentially expressed genes

Differentially expressed genes (DEGs) were filtered for transcriptomic data using the limma package in Rstduio. The selected DEGs with P <0.05 and | log2FC |> 2 (where p is the false discovery rate and the FC is the difference multiple) were used as the screening criteria, and the volcano plots DEGs were drawn using the ggplot 2 package.

1.4 Enrichment analysis of the functional genes

To clarify the potential biological function of differentially expressed genes, the differentially expressed genes for enrichment analysis, using the hypergeometric test statistical differentially expressed genes are enriched in an ontology (gene ontology) GO entry or Kyoto encyclopedia of genes (KEGG) included pathway, thus in the overall analysis of gene mutations may lead to which molecular function, biological processes or metabolic pathways change. The enrichment analysis of function and pathway was performed using KOBAS software (v3.0), using the metabolic pathway and functional databases were: KEGG and GO. The GO annotation includes three categories: biological processes (biological process, BP), cellular components (cellular components, CC), and molecular functions (molecular function, MF).

1.5 PPI network construction and core gene screening

The filtered DEGs were input into the STRING database (http://string-db.org/) to build the protein interaction network (protein-protein interaction network, PPIN), and the plug-in cytoNCA in the public bioinformatics software platform cytoscape software was used to visualized the degree of close molecular correlation, and select the top 20 most correlated differentially expressed genes, namely the core genes.

1.6 TCGA data set analysis of differentially expressed genes and prognosis correlation

Get the RNAseq data of RNASeq-TPM in the TCGA database (https://portal.gdc.cancer.gov/) LIHC (hepatocellular carcinoma) project, Divided into high and low expression groups according to the median expression level, The previously obtained 20 differentially expressed genes were used separately for the prognostic analysis, The software used for the analysis is: R (version 3.6.3) (statistical analysis and visualization), R package: survminer package [version 0.4.9] (for visualization), survival Package [version 3.2-10] (for statistical analysis of survival data), P <0.05 was considered to be statistically significant, To plot the Kaplan-Meier curve, Differentially expressed genes with a clear correlation with prognosis were obtained.

1.7 Expression of core differential genes in pan-cancer and correlation with immune infiltration

The selected genes were subjected to pan-cancer analysis and the UCSC XENA was collected (https: / / xenabrowser. The net / datapages /) database information, and through the Toil process (Vivian J et al., 2017) The TPM of TCGA, the RNAseq data in a unified format. We analyzed and compared the expression difference between tumor tissues and adjacent tissues in different organs. Secondly, the immune information of the patients in the liver cancer project in TCGA was obtained, and a set of immune cell surface marker genes was used in this study according to the results of Bindea et al. TCGA hepatocellular carcinoma transcriptomic data were grouped for line correlation analysis using 24 immune datasets and ssGSEA methods combined with the R software “GSVA” package.

1.8 ROC curve analysis, regression analysis, and nomogram construction

The differentially expressed genes were used as HCC variables, and the time-dependent ROC curves analysis were plotted to assess their potential as prognostic markers. Univariate and multivariate COX proportional hazards regression analyses were performed to determine the prognostic value of the differential genes. P <0.05 was considered to be statistically significant. All independent prognostic factors were used to establish the nomogram and to analyze the nomogram visually to assess the 1-, 3-, and 5-year survival probability of patients with HCC.

2.Comclusion

2.1 Screening of differentially expressed genes of HCC tissue samples

Transcriptome sequencing of HCC tissue samples and adjacent tissues was performed as described previously, and 2,145 differentially expressed genes were obtained. Using P <0.05 and | log2FC |> 2 as the screening criteria, 108 eligible DEGs were selected, and the volcano map was drawn using the ggplot 2 package in Rstudio (Figure 1). Downregulated genes are shown in blue left and upregulated genes are shown in red right.

Cancer Cell Research
35 (2022) 862-869
ISSN 2161-2609

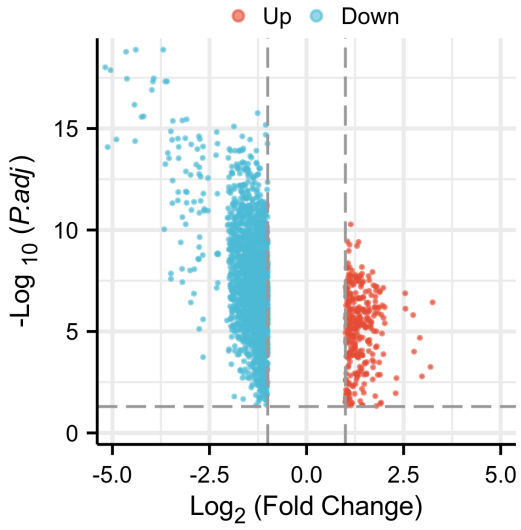
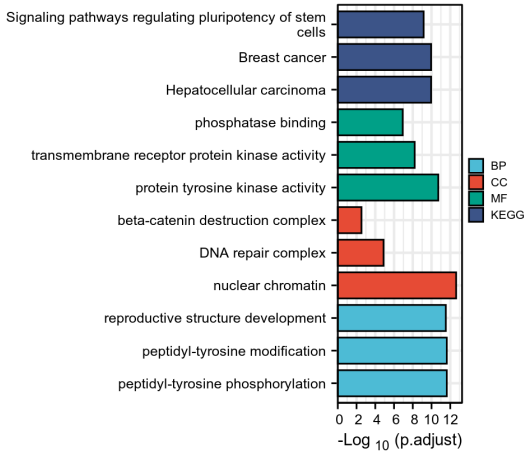


Figure1 Volcano plot of differentially expressed genes in HCC and adjacent tissues

2.2 Functional enrichment and pathway analysis



Enrichment analysis of gene function and pathways can provide clues to study the effects of gene variation on molecular function, metabolic pathways, signal transduction pathways, etc. The selected differentially expressed genes were used for enrichment analysis, and the enrichment results and figure examples are shown below. In terms of BP, the differential genes were significantly enriched in peptidyl tyrosine phosphorylation, peptidyl tyrosine modification, and regeneration and development. For CC, differential genes were enriched in nuclear chromatin, DNA repair complexes and catenin destruction complexes. For MF, the differential genes were significantly enriched in protein tyrosine kinase activity, transmembrane receptor protein kinase activity, phosphatase binding, etc. Moreover, KEGG analysis indicated that most differential gene-related pathways were significantly associated with immune response, and found a high enrichment rate in stem cells of liver cancer, breast cancer, and pluripotency regulatory signaling pathways. The results are shown in Figure 2.Biological processes (biological process, BP), cellular components (cellular components, CC), and molecular functions (molecular function, MF), and Figure2 Biological processes (biological process, BP), cellular components (cellular components, CC), and molecular functions (molecular function, MF), and KEGG analysis in the GO analysis. The abscissa represents the fold difference of expression, and the larger the absolute value of the abscissa is, the greater the fold difference of expression between the two samples. The ordinate indicates the pathway name

2.3 PPI network construction and core genes

The selected 108 differentially expressed genes input into STRING database, analyze the interaction relationship of gene expression, the interaction analysis data application cytoscope, software depth analysis, 105 are located in the PPIN complex, using the cytoNCA plug-in visualization to analyze the close correlation of differential genes, according to the correlation score BETWEENNESS (BC), and select the core genes in the top 20 BC value. .

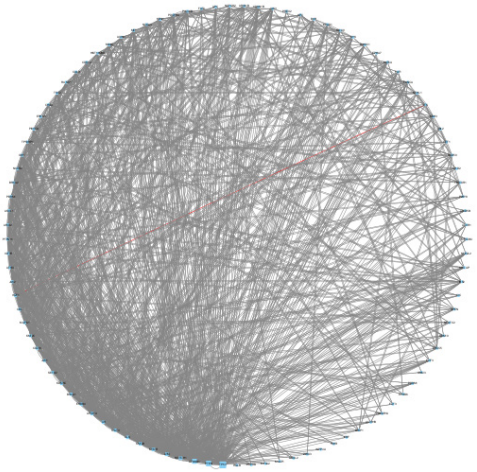


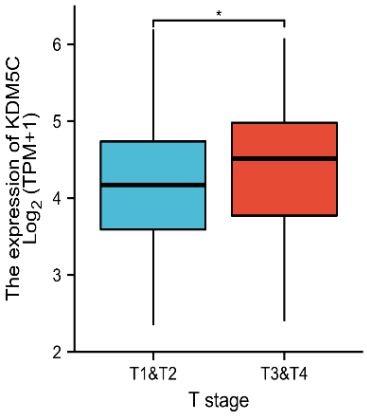
Figure 3 Figure left Cytoscape CytoNCA analysis (visualizing the closeness of molecules, the more connections, the more correlation. Figure right shows the results after screening for differential genes of the top 20 samples during cytoscape mapping.

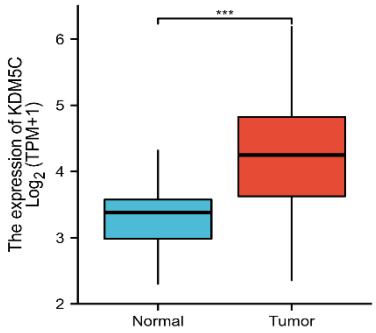
Cancer Cell Research
35 (2022) 862-869
ISSN 2161-2609

1	TP53	Betweenness: 1907.4716
2	PTEN	Betweenness: 1273.6497
3	CTNNB1	Betweenness: 1058.175
4	NOTCH1	Betweenness: 517.3907
5	ATM	Betweenness: 500.83072
6	MTOR	Betweenness: 467.32202
7	KMT2D	Betweenness: 418.87253
8	ALK	Betweenness: 356.82462
9	STAT3	Betweenness: 343.75644
10	AR	Betweenness: 333.95105
11	ARID1A	Betweenness: 320.81564
12	ERBB4	Betweenness: 307.56598
13	BRCA1	Betweenness: 249.90355
14	KDM5C	Betweenness: 225.0475
15	SETD2	Betweenness: 202.54855
16	JAK1	Betweenness: 185.40361
17	CD274	Betweenness: 181.76392
18	KMT2C	Betweenness: 173.91037
19	MAPK1	Betweenness: 170.13805
20	CBL	Betweenness: 163.82245

2.4 Acquisition gene closely related to HCC prognosis—KDM5C

Applied RNAseq data from TCGA database provided in Xiantao Academic Database (https://www.xiantao.love/products), Divided into high and low expression groups according to the median expression level, The previously obtained 20 differentially expressed genes were used separately for the prognostic analysis, The software used for analysis is: survminer package (for visualization) and survival package (for statistical analysis), P <0.05 was considered to be statistically significant, To plot the Kaplan-Meier curve, KDM5C was selected to be closely related with the prognosis of HCC. After the analysis of clinically relevant information of this gene in HCC samples, it was found that KDM5C was significantly differential expressed in normal tissues and HCC tissues, and KDM5C expression in HCC tissues was significantly increased, and its high expression predicted a poor prognosis (4A). In the tumor TNM grade, KDM5C expression was closely related to T stage (4B), and KDM5C expression was significantlydifferent between T3, T4 and T1, and T2 stages (4C).





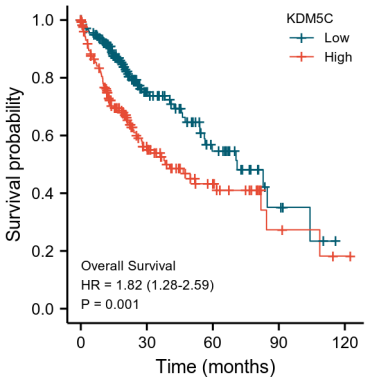


Figure4A Survival curves of the different expression of KDM5C and the prognosis of HCC
Figure4B KDM5C was differentially expressed in HCC tumor
Figure4C Correlation between different T stages and KDM5C expression

2.5 Pan-cancer analysis and correlation analysis with immune cells

Using the data source of UCSC XENA in Xiantao Academic database (https://www.xiantao.love/products), the RNAseq data in TPM format combined with KDM5C gene for pan-cancer analysis was used, and the expression differences of tumor tissues and adjacent tissues in different organs were analyzed and compared. As shown in Figure 5A, the KDM5C expression varied significantly between tumor tissues and normal tissues in many different cancer species, especially in liver cancer. The correlation between KDM5C gene and 24 immune cells was analyzed, (as shown in Figure 5B), pDC, cytotooxic, DC, and T helper cell, TF 2, and TFH. As shown in Figure 5C, the strongly correlated immune cells pDC, cytotooxic, DC negatively correlated with

Cancer Cell Research
35 (2022) 862-869
ISSN 2161-2609

KDM5C gene (p <0.05), and T helper cell, TF 2, TFH positively correlated with KDM5C gene (p <0.05)

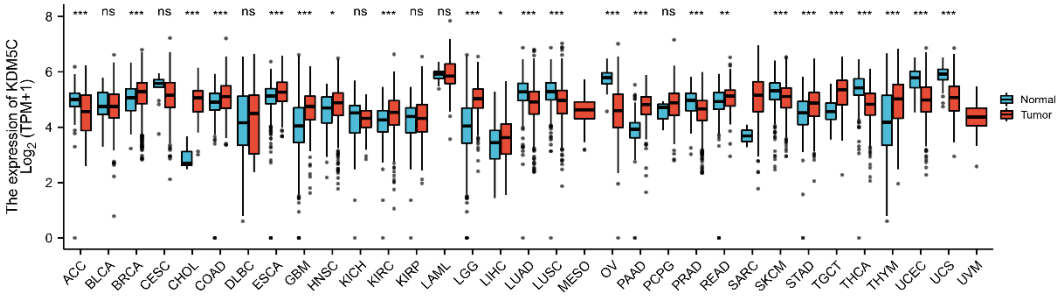


Figure 5A KDM5C Gene expression in the pan-cancer

2.6 Construction of time-dependent ROC curves and prognostic correlation analysis

Time-dependent ROC curves were used to analyze the potential of KDM5C as a prognostic marker for HCC. The area under the curve (AUC) values for 1-year, 3-year overall survival were 0.673,0.614, and 0.574, respectively (Figure 6A), indicating that the KDM5C gene has a higher warning value for the 1-year survival rate of HCC patients after surgery compared with the other two time points.

Univariate analysis of KDM5C gene with clinical indicators including degree of tumor invasion, age, degree of vascular invasion, AFP, Child grade, tumor resection margin, and tumor clearance status. The degree of tumor invasion, KDM5C, and tumor clearance status were the clinical factors with a clear correlation. Multivariate regression analysis included KDM5C, degree of tumor invasion and tumor clearance status, and the three were still independent prognostic factors, as shown in Figure 6B

2.7 nomogram construction and visual analysis

As shown in Figure 7, a nomogram was constructed based on the next three independent factors of OS: the degree of tumor invasion, KDM5C, and tumor clearance status. The different independent factors were matched to the corresponding score, and the total score was obtained by adding it. On the nomogram, an increase in total points was associated with a worse prognosis.1,3 and 5 year survival probabilities were analyzed for each patient. The results are shown in Figure 7, performing calibration analysis and visualizing it, 45° line represents the best prediction situation, and patients have a good fit at 3 years, that is, the accuracy of survival risk

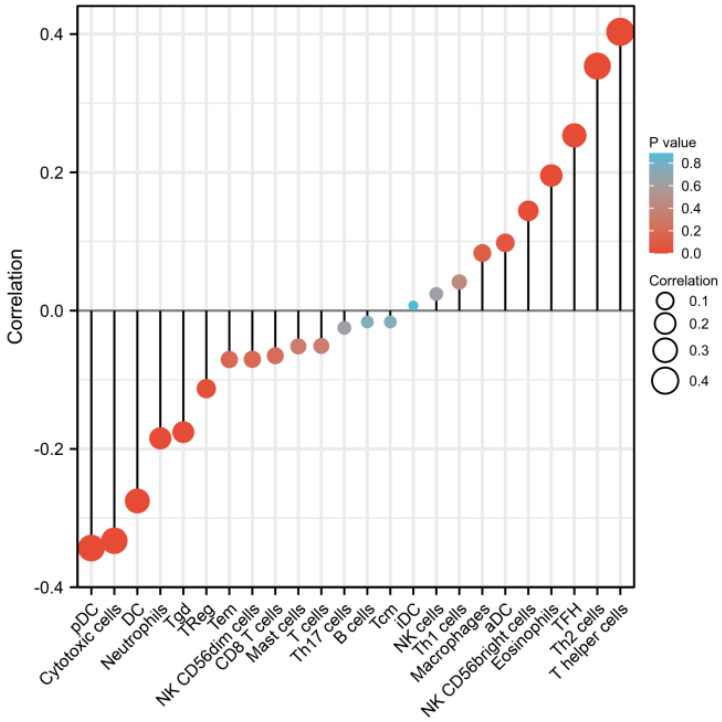


Figure5B Correlation between KDM5C gene and immune cells in liver cancer

Cancer Cell Research
35 (2022) 862-869
ISSN 2161-2609

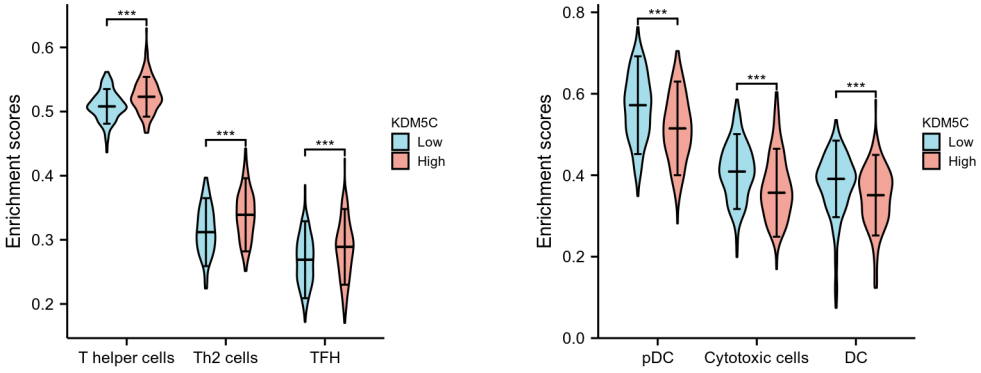
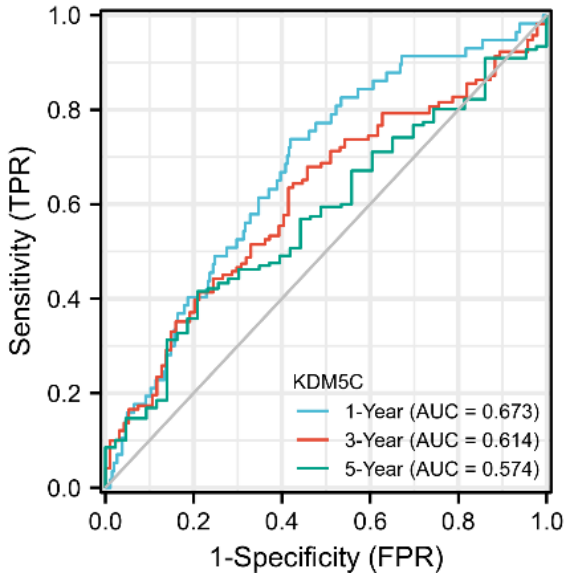


Figure5C Immune cell correlation of KDM 5 C gene



T1&T2	277	Reference				
T4&T3	93	2.598 (1.826-3.697)	<0.001		2.192 (1.509-3.184)	<0.001
Age	373					
<=60	177	Reference				
>60	196	1.205 (0.850-1.708)	0.295			
Vascular invasion	317					
No	208	Reference				
Yes	109	1.344 (0.887-2.035)	0.163			
Gender	373					
Male	252	Reference				
Female	121	1.261 (0.885-1.796)	0.200			
KDM5C	373					
Low	187	Reference				
High	186	1.824 (1.284-2.592)	<0.001		1.672 (1.155-2.421)	0.006
AFP	373					
Low	187	Reference				
High	186	1.089 (0.770-1.539)	0.630			
Child-Pugh grade	240					
A	218	Reference				
B&C	22	1.643 (0.811-3.330)	0.168			
Residual tumor	344					
R0	326	Reference				

Figure6A ROC curve shows the prognostic value of KDM 5 C in patients with hepatocellular carcinoma (HCC)

Figure6B Univariate and multivariate analysis in the correlation analysis of liver cancer prognosis

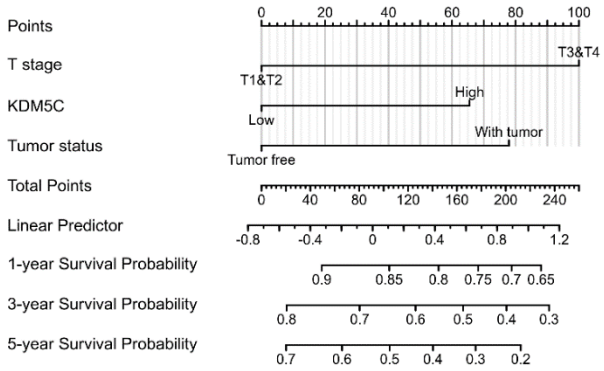


Figure7A T stage, KDM5C, Tumor status predicts 1,3,5-year survival risk

Cancer Cell Research
35 (2022) 862-869
ISSN 2161-2609

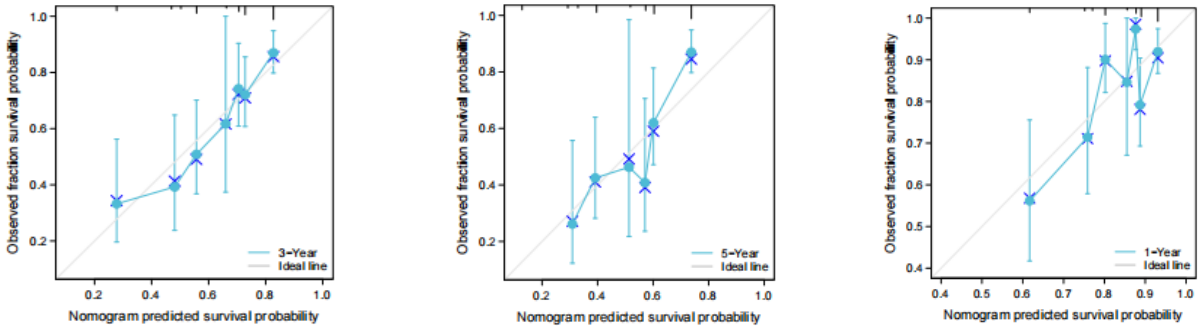


Figure7B calibration Visual analysis evaluates the nomograms

assessment at 3 years is higher.

3.Discussion

Liver cancer has a high tumor heterogeneity, although systematic treatment has been significantly improved, but the overall prognosis is poor [9, 10]. Sorafenib, lenvatinib, Donafenib and other targeted drugs are the first-line treatment of advanced liver cancer at present, but they all have the problems of primary drug resistance and acquired drug resistance. Therefore, in-depth analysis of differential genes in liver cancer can provide a basis for targeted therapy or screening reliable biomarkers of liver cancer.

Lysine specific demethylase 5C (KDM5C) is a member of the JmjC domain protein family, which specifically removes methyl residues from trimethyl, dimethyl and monomethylated lysine 4 on histone H3 lysine 4 (H3K4) associated with active genes. KDM5C is a transcriptional inhibitor with inherent histone demethylase activity [12]. The trimethylation of H3K4 is an important histone marker related to active transcriptional genes. KDM5C specifically demethylates H3K4me3 into a state of transcriptional inactivity. The expression of KDM5C is up-regulated in prostate cancer, breast cancer and other tumor cell lines [15-17]. It can also promote the proliferation of cancer cells. Down-regulation of its expression will lead to a significant delay in the transition of G1UniS [17]. KDM5C gene promotes tumorigenesis by specifically suppressing anti-proliferation genes. According to the study of XueningJi, down-regulation of KDM5C expression can inhibit the migration, invasion and epithelial-mesenchymal transformation of HCC cells in vitro, and significantly reduce the metastatic ability of invasive HCC cells in liver and lung. Secondly, KDM5C induces protein inactivation by regulating BMP7 in bone regulatory protein BMP (Bonemorphogeneticproteins), which promotes cell migration, invasion and epithelial-mesenchymal transformation. Knockout of BMP7 significantly promoted the inhibition of cell migration induced by KDM5C gene [20]. In this study, through the biological information analysis of differentially expressed genes in the whole genome sequencing data of 20 liver cancer tissue samples, it was found that KDM5C gene was significantly differentially expressed in hepatocellular carcinoma. The correlation between KDM5C gene and prognosis, differential expression in liver cancer tissue, immune infiltration and clinical related indexes were verified in TCGA database. The degree of tumor invasion and KDM5C were found. Three indicators, including tumor clearance status, are closely related to clinical prognosis, which enrich the prognostic indicators of liver cancer and play a certain guiding value in the prognosis of liver cancer.

There are some shortcomings in this study. Due to the small number of patients and short time in the group, the prognosis of 20 patients was not directly observed but verified in the database. In the later stage, we can continue to increase the enrollment list, screen potential driving genes, and conduct a long-term follow-up survey of patients who have been enrolled in the group. In addition, molecular biology methods and experimental data are needed for further verification.

References

[1]. Bray, F., et al., Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin, 2018. 68(6): p. 394-424.

[2]. Forner, A., M. Reig and J. Bruix, Hepatocellular carcinoma. Lancet, 2018. 391(10127): p. 1301-1314.

[3]. Sperber, A.D., et al., Worldwide Prevalence and Burden of Functional Gastrointestinal Disorders, Results of Rome Foundation Global Study. Gastroenterology, 2021. 160(1): p. 99-114.e3.

[4]. Schweitzer, A., et al., Estimations of worldwide prevalence of chronic hepatitis B virus infection: a systematic review of data published between 1965 and 2013. Lancet, 2015. 386(10003): p. 1546-55.

[5]. Johnson, P.J., The role of serum alpha-fetoprotein estimation in the diagnosis and management of hepatocellular carcinoma. Clin Liver Dis, 2001. 5(1): p. 145-59.

[6]. Lee, B.H., et al., Role and limitation of

35 (2022) 862-869
ISSN 2161-2609

neoadjuvant hepatic arterial infusion chemotherapy in advanced hepatocelluar carcinoma patients with Child-Pugh class A. World J Surg Oncol, 2019. 17(1): p. 143.
[7]. Lyu, N., et al., Hepatic arterial infusion of oxaliplatin plus fluorouracil/leucovorin vs. sorafenib for advanced hepatocellular carcinoma. J Hepatol, 2018. 69(1): p. 60-69.
[8]. Long, G.B., et al., Effects of hepatic arterial infusion chemotherapy in the treatment of hepatocellular carcinoma: A meta-analysis. Medicine (Baltimore), 2020. 99(26): p. e20745.
[9]. El-Serag, H.B., Hepatocellular carcinoma. N Engl J Med, 2011. 365(12): p. 1118-27.
[10]. Wang, N., et al., Berberine induces autophagic cell death and mitochondrial apoptosis in liver cancer cells: the cellular mechanism. J Cell Biochem, 2010. 111(6): p. 1426-36.
[11]. Jensen, L.R., et al., Mutations in the JARID1C gene, which is involved in transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. Am J Hum Genet, 2005. 76(2): p. 227-36.
[12]. Abidi, F.E., et al., Mutations in JARID1C are associated with X-linked mental retardation, short stature and hyperreflexia. J Med Genet, 2008. 45(12): p. 787-93.
[13]. Wynder, C., L. Stalker and M.L. Doughty, Role of H3K4 demethylases in complex neurodevelopmental diseases. Epigenomics, 2010. 2(3): p. 407-18.
[14]. Stein, J., et al., KDM5C is overexpressed in prostate cancer and is a prognostic marker for prostate-specific antigen-relapse following radical prostatectomy. Am J Pathol, 2014. 184(9): p. 2430-7.
[15]. Patani, N., et al., Histone-modifier gene expression profiles are associated with pathological and clinical outcomes in human breast cancer. Anticancer Res, 2011. 31(12): p. 4115-25.
[16]. Kaiser, J., Genes link epigenetics and cancer. Science, 2010. 330(6004): p. 577.
[17]. Brookes, E., et al., Mutations in the intellectual disability gene KDM5C reduce protein stability and demethylase activity. Hum Mol Genet, 2015. 24(10): p. 2861-72.
[18]. Rondinelli, B., et al., H3K4me3 demethylation by the histone demethylase KDM5C/JARID1C promotes DNA replication origin firing. Nucleic Acids Res, 2015. 43(5): p. 2560-74.
[19]. Shao, G.B., et al., Dynamic patterns of histone H3 lysine 4 methyltransferases and demethylases during mouse preimplantation development. In Vitro Cell Dev Biol Anim, 2014. 50(7): p. 603-13.
[20]. Ji, X., et al., Lysine-specific demethylase 5C promotes hepatocellular carcinoma cell invasion through inhibition BMP7 expression. BMC Cancer, 2015. 15: p. 801.