84784_Auto_Edited.docx

WORD COUNT	2986	TIME SUBMITTED	31-MAY-2023 09:19AM
		PAPER ID	100109085

Name of Journal: *World Journal of Gastrointestinal Oncology*

Manuscript NO: 84784

Manuscript Type: ORIGINAL ARTICLE

Clinical and Translational Research

Expression patterns of CD147 impact the prognosis of liver hepatocellular carcinoma

CD147 impact the prognosis of liver hepatocellular carcinoma

Yun Ji Xu, Hong jie He, Peng Wu, Wen bing Li

Abstract

BACKGROUND

Hepatocellular carcinoma (HCC) is very low of overall survival. According to global cancer statistics, ~ 905,677 new cases were reported in 2020, with at least 830,180 of them being fatal. CD147 (cluster of differentiation 147), is a novel, transmembrane glycoprotein which is expressed in a wide variety of tumor cells and plays an important role in various stages of tumor development. Based on the reports described previously, we theorize that CD147 may be used as a novel biological indicator to predict the prognosis of HCC. To study this possibility, expression profiles of CD147 and corresponding clinical data from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases were analyzed, and a hazard ratio (HR) was also established.

AIM

To explore the pattern of CD147 expression and its applicability in the prognosis of hepatocellular carcinoma (HCC).

To establish hazard ratio and probability points for predicting the prognosis of HCC by correlating CD147 expression with clinical characteristics.

To determine if CD147 can be a reliable biomarker in HCC prognosis.

METHODS

CD147 expression profile in HCC and corresponding clinical data were obtained from The Cancer Genome Atlas (TCGA) database. The expression patterns of CD147 were then validated by analyzing data from the Gene Expression Omnibus (GEO) database. In addition, CD147 Immunohistochemistry (IHC) in HCC was obtained from the Human Protein Atlas (HPA). CD147 expression patterns and clinical characteristics in the prognosis of HCC were analyzed by accessing the UALCAN web resource. Accuracy, sensitivity, and specificity of the CD147 expression profile in predictive prognosis were determined by the Time-dependent Receiver Operating Characteristic (ROC) curves. Kaplan-Meier curves were plotted to estimate the hazard ratio (HR) of survival in HCC. Univariate and multivariate Cox regression proportional hazards analyses of CD147 expression levels and clinical characteristics as prognostic factors of HCC were performed. Nomograms were used to establish probability points and predict prognosis.

RESULTS

Data from TCGA and GEO databases revealed that CD147 was significantly overexpressed in HCC (p =1.624×10^{-12} and p =1.2×10^{-5}), respectively. The expression of CD147 and prognosis of HCC were significantly correlated with the clinical characteristics of HCC as per the data from the UALCAN web resource (p < 0.05). Kaplan-Meier analysis of CD147 expression in HCC revealed that the high expression groups showed poor prognosis and an HR of survival > 1 [(Log-rank test, p = 0.000542, HR (in high expression groups) = 1.856, 95%CI (1.308, 2.636)]. ROC curves were plotted to analyze the 1-year, 3-year, and 5-year survival rates, the Area under the ROC curve (AUC) values were 0.675, 95%CI (0.611–0.74); 0.623, 95%CI (0.555–0.692); and 0.664, 95%CI (0.582–9.745) respectively. Univariate Cox analysis of CD147 expression and clinical characteristics of HCC, and multivariate Cox analysis of CD147 patterns and

pTNM-stage showed significant differences [(Uni-Cox, p = 0.00013, HR = 1.42437, 95%CI (1.8838, 1.70723) and p = 0.00066, HR = 1.37612, 95%CI (1.14521, 1.65359); Multi-Cox, p = 0.00578, HR = 1.50746, 95%CI (1.12637, 2.0175) and p = 0.00336, HR = 1.44319, 95%CI (1.12941, 1.84415)]. Nomograms were plotted to establish the probability points and predict prognosis, the total points ranged from 0 to 180, and the C-index value was 0.673, 95%CI (0.6–1.0), p < 0.01.

CONCLUSION

Overexpression of CD147 was correlated with poor prognosis in HCC. CD147 expression profile combined with clinical characteristics can reliably predict the prognosis of HCC. CD147 can serve as a biomarker to predict the prognosis of HCC.

Key Words: Hepatocellular carcinoma; CD147; Prognosis; Clinical characteristics; Hazard ratio

Xu YJ, He HJ, Wu P, Li WB. Expression patterns of CD147 impact the prognosis of liver hepatocellular carcinoma. *World J Gastrointest Oncol* 2023; In press

Core Tip: Hepatocellular carcinoma (HCC)
The Cancer Genome Atlas (TCGA)
Gene Expression Omnibus (GEO)
Immunohistochemistry (IHC)
Human Protein Atlas (HPA)
Receiver Operating Characteristic (ROC)
hazard ratio (HR)
Cholangiocarcinoma (CHOL)
Surveillance Epidemiology End Results (SEER)
Non-alcoholic Steatohepatitis (NASH)
hepatitis B virus (HBV)

hepatitis C virus (HCV)

Barcelona Clinic Liver Cancer (BCLC)

Transarterial Chemoembolization (TACE)

Basigin (*BSG*)

matrix metalloproteinase (MMP)

Pathological Tumor-node-metastasis (pTNM)

overall survival (OS)

Confidence Interval (CI)

Vascular Endothelial Growth Factor (VEGF)

esophageal squamous cell carcinoma (ESCC)

Human Immunodeficiency Virus type 1 (HIV-1)

Matuzumab (anti-CD147)

severe acute respiratory syndrome associated-coronavirus (SARS-CoV)

INTRODUCTION

Liver cancer is a highly malignant tumor, with a 5-year survival rate being only 10% [1]. Therefore, its treatment is challenging the worldwide [1,2]. HCC and Cholangiocarcinoma (CHOL) account for 80–90% and 10–15% of all primary liver cancers, respectively [3]. According to global cancer statistics, ~ 905,677 new cases were reported in 2020, with at least 830,180 of them being fatal [4]. It is estimated that by 2025, more than one million individuals would be affected by liver cancer annually [5]. Surveillance Epidemiology End Results (SEER) reported HCC to be the fastest-growing cause of cancer-related deaths in the US since the early 2000s and if this trend continues, is predicted to become the third leading cause of cancer-related mortality by 2030 [6]. The major risk factors for HCC include chronic alcohol consumption, diabetes or obesity-related Non-alcoholic Steatohepatitis (NASH), and infection by hepatitis B virus (HBV) or hepatitis C virus (HCV). The high-risk factors for HCC include chronic alcohol consumption, diabetes, and infection by HBV or HCV [7, 8]. According to Barcelona Clinic Liver Cancer (BCLC) staging system, several approaches are available for the

treatment of HCC, such as surgery (early stage), liver transplantation, Transarterial Chemoembolization (TACE), and targeted therapy and local ablation [9].

CD147 (cluster of differentiation 147), is a novel, transmembrane glycoprotein which is expressed in a wide variety of tumor cells and plays an important role in various stages of tumor development [10–12]. CD147 is encoded by the *Basigin* (*BSG*) gene located on chromosome 19 at the p13.3 Locus [11–12]. Epithelial and fetal tissues have low expression levels of CD147 [13]. CD147 promotes tumor proliferation, invasion, and metastasis [14, 15] probably by triggering a matrix metalloproteinase (MMP) on the tumor surface [16, 17]. The expression of CD147 is upregulated in various tumors including breast cancer, bladder cancer, colorectal cancer, ovarian cancer, melanoma, and osteosarcoma [18, 19] and also in HCC [20]. CD147 is highly expressed in ovarian cancer and in combination with Human Epididymis protein 4 (HE4) may act as a novel indicator for the diagnosis and treatment of early ovarian cancer [21].

Based on the reports described previously, we theorize that CD147 may be used as a novel biological indicator to predict the prognosis of HCC. To study this possibility, expression profiles of CD147 and corresponding clinical data from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases were analyzed, and a hazard ratio (HR) was also established.

MATERIALS AND METHODS

2.1 Data collection

The transcriptome profile of HCC was obtained from TCGA by using the TCGAbiolinks package. The dataset included 50 normal and 371 tumor samples (TCGA-LIHC). The GSE112790 dataset (based on the GPL570 platform) which included 15 normal and 183 tumor tissues was acquired from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) for external validation of the prognostic gene signature. The workflow was illustrated in Figure 1.

2.2 Analysis of patterns in CD147 expression

To determine the expression patterns of CD147, gene expression analysis was performed on the TCGA-LIHC dataset by using the R package cluster profiler. The samples with no significant expression value and insufficient survival information were excluded. The clinical characteristics including age, gender, Pathological Tumor-node-metastasis (pTNM), tumor grade, metastatic status, overall survival (OS) time, and survival status were obtained from the patient's data. The expression profile of CD147 was analyzed and its prognostic value was validated in the GSE112790 dataset.

2.3 Analysis of mRNA expression profile of CD147 by using UALCAN

UALCAN (http://ualcan.path.uab.edu) is an interactive web resource designed to analyze the relative mRNA expression patterns of potential genes (TCGA and MET500 transcriptome sequencing) and their relationship with various tumor subtypes. UALCAN was utilized to obtain the mRNA expression profile of CD147 in HCC tissues and ascertain its association with clinical characteristics.

2.4 Analysis of survival parameters

The correlation between the expression of CD147 and OS of patients with liver HCC was analyzed by univariate and multivariate Cox proportional hazards regression analysis. The time-dependent Receiver Operating Characteristic (ROC) curve and Kaplan-Meier curve were generated to assess the prognostic ability of CD147. A prognostic nomogram was also constructed based on the results obtained from the multivariate Cox regression analysis to predict the 1, 3, and 5-year survival rates and overall recurrence. The log-rank test was used for calculating HR with a Confidence Interval (CI) of 95%.

2.5 Statistical analysis

All analyses were performed using R software (version 4.0.3, foundation for statistical computing, 2020) and its related packages unless stated otherwise. The Student's *t*-test was used to determine if the differences between the two groups were statistically significant. A value of $p < 0.05$ was considered to be statistically significant.

RESULTS

The expression of CD147 and Immunohistochemistry (IHC) in HCC

CD147 expression profile was obtained from the TCGA database and GSE112790 dataset from the GEO database. Analysis of these revealed that CD147 was significantly overexpressed in tumors than in normal tissues (p =1.624×10^{-12} and p =1.2×10^{-5}, Figures 2 A and B). In addition, the IHC profile of CD147 in HCC was obtained from the Human Protein Atlas (HPA) [https://www.proteinatlas.org] (Figure 3).

The expression of CD147 in the clinical classification of HCC

The expression of CD147 in clinical characteristics of HCC obtained from the TCGA database was analyzed using UALCAN. The expression of CD147 in age, gender, tumor grade, pTNM, weight, histology, metastatic status, and mutation in the *TP53* gene, respectively were compared. The expression of CD147 showed significant differences in each criterion ($p < 0.05$, Figures 4 A, B, C, D, E, F, G, H).

CD147 is correlated with prognosis and HR in HCC

The correlation between CD147 expression and prognosis of HCC concerning age, tumor grade, weight, gender, and OS time data obtained from the TCGA database showed statistically significant differences (p = 0.065, p = 0.065, p = 0.0016, p = 0.0078, p = 0.0016, respectively; Figures 5, A, B, C, D, E, F). Kaplan-Meier curve based survival analysis in low- and high-expression groups of CD147 was done by the log-rank test(Figure 6). The high-expression groups demonstrated poor prognosis and HR > 1 [Log-rank p = 0.000542, HR (in high expression groups) = 1.856, 95%CI (1.308, 2.636) (Figure 6, B)]. Based on the AUC values obtained from ROC curves, the 1-year, 3-year, and 5-year survival rates were 0.675, 95%CI (0.611–0.74); 0.623, 95%CI (0.555–0.692); and 0.664, 95%CI (0.582–9.745) (Figure 6, C).

Nomograms establish the probability points of CD147 expression in HCC

Through Uni- and Multi-Cox regression analysis CD147 expression, age, gender, pTNM-stage, tumor grade, and HR of a new tumor in HCC. The CD147 (BSG) and pTNM-stage showed significant differences [(Uni-Cox, p = 0.00013, HR = 1.42437, 95%CI (1.8838, 1.70723) and p = 0.00066, HR = 1.37612, 95%CI (1.14521, 1.65359),

respectively (Figure 7, A); Multi-Cox, $p = 0.00578$, HR = 1.50746, 95%CI (1.12637, 2.0175) and $p = 0.00336$, HR = 1.44319, 95%CI (1.12941, 1.84415), respectively (Figure 7, B)]. Accordingly, the factors of CD147 expression and pTNM-stage ($p < 0.05$) were selected to establish a prognostic nomogram for HCC. The total points ranged from 0 to 180, and the C-index value was 0.673, 95%CI (0.6–1.0), $p < 0.01$ (Figure 7, C).

DISCUSSION

The expression profile of CD147 in HCC was obtained from the TCGA database and was verified in the GSE112790 dataset obtained from the GEO database. The expression of CD147 was significantly higher in HCC than in normal tissue. In addition, analysis using UALCAN revealed that the expression of CD147 was closely associated with clinical characteristics of HCC, including age, gender, pTNM, tumor grader, metastasic status, weight, histological, and mutation in the *TP53* gene. This suggests that CD147 is overexpression and closely related to clinical characteristics in HCC.

Even though CD147 is closely associated with tumor proliferation, invasion, and metastasis [14, 15], the underlying mechanisms are still unclear. CD147 overexpression may be associated with tumor cell migration and activation of the extracellular-signal-regulated kinase signaling pathway [22]. In addition, CD147 regulated MMPs and Vascular Endothelial Growth Factor (VEGF) implicated in tumor and stromal cells [22–23]. CD147 induced MEK-mediated intracellular signaling pathway and MMP-9 activity which promoted tumor proliferation, invasion, and metastasis in hypopharyngeal carcinoma [24]. CD147 was overexpressed in HCC cells and the knockdown of CD147 significantly inhibited the proliferation, migration, and invasion of HCC cells [25]. Hypo-phosphorylation of CD147 promotes the invasion and metastasis of HCC and CD147 may be utilized as a novel biomarker in the prognosis of HCC [26]. CD147 expression has been associated with lymph node metastasis in cervical cancer and laryngeal squamous cell carcinoma [27]. CD147 was overexpressed in oral cancer, knockout of CD147 significantly reduced the proliferation and invasion of cal27 cells, and CD147 may be a potential therapeutic target in oral cancer [28].

CD147 has been proposed as a novel, prognostic biomarker in HCC. We analyzed the role of the expression of CD147 in the prognosis of HCC. High expression of CD147 significantly shortened the prognosis of HCC and related clinical characteristics. In our study, we found CD147 had an HR > 1 implying that is a poorly reliable factor in prognosis. The CI of Uni-Cox and Multi-Cox analyses of the expression of CD147 and clinical characteristics was 0.673, significantly greater than 0.5, indicating that it has a high predictive value. Since CD147 plays an important role in the prognosis of HCC, it may be considered a potential biomarker in the prediction of tumor prognosis.

Several studies have shown that CD147 has a good prognostic value in many tumors. The methylation levels of CD147 were ascertained using cfDNA in non-small cell lung cancer (NSCLC) tissues and were inversely related to tumor size, lymph node metastasis, and TNM stage [29]. Targeted methylation of CD147 could inhibit NSCLC invasion and metastasis. CD147 and MMP-9 were closely correlated with the pathological stage, metastasis, and differentiation of tumors in breast cancer cases ($p < 0.05$) but were poor independent risk factors for prognosis in Triple-Negative Breast Cancer (TNBC) (PCD147 = 0.023, PMMP-9 = 0.015) [30]. In cancer patients older than 50 years, CD147 was an independent prognostic indicator [30]. Expression patterns of CD147 in different stages of esophageal squamous cell carcinoma (ESCC) were able to reliably predict prognosis in patients [31].

CD147 also plays an important role in other diseases such as Human Immunodeficiency Virus type 1 (HIV-1), HCV, HBV, Kaposi's sarcoma-associated herpes virus, and severe acute respiratory syndrome associated-coronavirus (SARS-CoV) infections [32]. Interestingly, CD147 is involved in SARS-CoV-2 tropism and may be a potential therapeutic target for COVID-19 [32].

Matuzumab (anti-CD147) was confirmed to be a safe treatment method for NSCLC [33]. Recently, chimeric antigen receptor T-cell immunotherapy targeting CD147 had demonstrated antitumor efficacy in patients with HCC [20, 34]. Thus, CD147 expression in correlation with clinical characteristics may serve as a predictive biomarker for pathological type prognosis of tumor and also as a target for tumor treatment.

CONCLUSION

Overexpression of CD147 was correlated with poor prognosis in HCC. CD147 expression profile combined with clinical characteristics can reliably predict the prognosis of HCC. CD147 can serve as a biomarker to predict the prognosis of HCC.

ARTICLE HIGHLIGHTS

Research background

Hepatocellular carcinoma (HCC) is very low of overall survival. Searching for a new biomarker that related to prognosis is helpful to improve the prognosis of HCC. CD147 is a novel that transmembrane glycoprotein which is expressed in a wide variety of tumor cells and plays an important role in various stages of tumor development. In our study, we found CD147 correlated with clinical characteristics and prognosis of HCC. Therefore, we suggest CD147 can serve as a biomarker to predict the prognosis of HCC.

Research motivation

CD147 is highly expressed in various tumors and is associated with prognosis. To explore the correlation between clinical characteristics and prognosis of CD147 in HCC. And to predict whether CD147 can become a reliable biomarker in HCC prognosis.

Research objectives

To explore the pattern of CD147 expression and its applicability in the prognosis of hepatocellular carcinoma (HCC). And establish hazard ratio and probability points for predicting the prognosis of HCC by correlating CD147 expression with clinical characteristics. In addition, to determine if CD147 can be a reliable biomarker in HCC prognosis.

Research methods

Using TCGA and GEO databases, R language was used to analyze the expression of CD147 in HCC. The online website UALCAN analyzes the correlation between clinical characteristics and survival time of TCGA-LIHC and CD147 expression. Subsequently, Time dependent Receiver Operating Characteristic (ROC) curves were used to analyze the accuracy, sensitivity, and specificity of CD147 in HCC. Finally, univariate and multivariate Cox regression proportional hazards analyses of CD147 expression levels and clinical characteristics as prognostic factors of HCC were performed. Nomograms were used to establish probability points and predict prognosis.

Research results

CD147 is overexpressed and the prognosis of HCC were significantly correlated with the clinical characteristics of HCC. The overexpression of CD147 showed poor prognosis and HR of survival > 1 in HCC. Multivariate Cox analysis of CD147 patterns and pTNM-stage showed significant differences. Nomograms showed CD147 can predict prognosis

Research conclusions

CD147 is overexpressed and is associated with clinical characteristics in HCC. CD147 overexpressed in HCC has a poor prognosis. In addition, CD147 can predict the prognosis of HCC.

Research perspectives

CD147 is a novel, transmembrane glycoprotein which is expressed in a wide variety of tumor cells and plays an important role in various stages of tumor development. Based on our study and the reports described previously, we theorize that CD147 may be used as a novel biological indicator to predict the prognosis of HCC.

84784_Auto_Edited.docx

ORIGINALITY REPORT

54%
SIMILARITY INDEX

PRIMARY SOURCES

1	www.researchsquare.com Internet	1691 words — 54%

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	< 1%
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	< 10 WORDS