SOX4 as biomarker in hepatitis B virus-associated hepatocellular carcinoma

Jian-Lv Huang1, Xiang-Kun Wang1, Xi-Wen Liao1, Chuang-Ye Han1, Ting-Dong Yu1, Ke-Tuan Huang1, Cheng-Kun Yang1, Xiao-Guang Liu1,2, Long Yu1,3, Guang-Zhi Zhu1, Hao Su1, Wei Qin1, Quan-Fa Han1, Zheng-Qian Liu1, Xin Zhou1, Jun-Qi Liu1, Xin-Ping Ye1, Tao Peng1

2. Department of Hepatobiliary Surgery, Affiliated Hospital of Guangdong Medical University, Zhanjiang, 524001, Guangdong Province, China.

Corresponding author: Prof. Tao Peng, Department of Hepatobiliary Surgery, The First Affiliated Hospital of Guangxi Medical University, Shuang Yong Road 6, Nanning, 530021, Guangxi Zhuang Autonomous Region, People’s Republic of China. Tel: (+86)-771-5356528. Fax: (+86)-771-5350031. E-mail: pengtao@gxmu.edu.cn. ORCID ID: http://orcid.org/0000-0001-6133-7078.

Abstract

Background: Hepatitis B virus infection is associated with liver disease, including cancers. In this study, we assessed the power of sex-determining region Y (SRY)-related high-mobility group (HMG)-box 4 (SOX4) gene to predict the clinical course of hepatocellular carcinoma (HCC).

Methods: To evaluate the differential expression of SOX4 and its diagnostic and prognostic potential in HCC, we analyzed the GSE14520 dataset. Stratified analysis and joint-effect analysis were done using SOX4 and clinical factor. We then designed a nomogram for predicting the clinical course of HCC. Differential SOX4 expression and its correlation with tumor stage as well as its diagnostic and prognostic value were analyzed on the oncomine and GEPIA websites. Gene set enrichment analysis was explored as well as candidate gene ontology and metabolic pathways modulated by SOX4 in HCC.

Results: Our analysis revealed that the level of SOX4 was significantly upregulated in tumor issue (P < 0.001). This observation was validated through oncomine dataset and MERAV analysis (all P < 0.05). Diagnostic receiver operating characteristic (ROC) analysis of SOX4 suggested it has diagnostic potential in HCC (GSE14520 dataset: P < 0.001, area under curve (AUC) = 0.782; Oncomine: Wurmbach dataset P = 0.002, AUC = 0.831 and Mas dataset P < 0.001, AUC = 0.947). In addition, SOX4 exhibited high correlation with overall survival of HBV-associated HCC (adjusted P = 0.004, hazard ratio (HR) (95% confidence interval (CI)) = 2.055 (1.261-3.349) and recurrence-free survival (adjusted P = 0.008, HR (95% CI) = 1.721 (1.151-2.574). These observations which were verified by GEPIA analysis for overall survival (P = 0.007) and recurrence-free survival (P = 0.096). Gene enrichment analysis revealed that affected processes included lymphocyte differentiation, pancreatic endocrine pathways, and insulin signaling pathway. SOX4 prognostic value was evaluated using nomogram analysis for HCC 1, 3, and 5-year, survival.

Conclusion: Differential SOX4 expression presents an avenue of diagnosing and predicting clinical course of HCC. In HCC, SOX4 may affect TP53 metabolic processes, lymphocyte differentiation and the insulin signaling pathway.

Key words: SOX4, hepatocellular carcinoma, hepatitis B virus, mRNA expression

Introduction

Cancers affecting liver tissues have been on the rise, making liver cancer the fourth cause of deaths and sixth most prevalent cancer globally [1]. Specifically, hepatocellular carcinoma is the most common type of primary liver cancers. Liver cancer is estimated to be the fourth commonest cancer among Chinese males [2]. Majority of liver cancers have been associated hepatitis B virus (HBV) infection. While advances in diagnostic and treatment strategies have improved HCC clinical outcomes, its 5-year survival remains low (<15%) [3]. Early detection and more effective management of liver cancer is therefore...
necessary. While some HCC prognostic biomarkers have been recommended, such as α-fetoprotein (AFP) [4] and PIVKA-II [5], HCC survival is still poor. Better understanding of the mechanisms of HCC development and progression, as well as the identification of novel prognostic biomarkers is needed.

Sex-determining region Y (SRY)-related high-mobility group (HMG)-box(SOX) genes are evolutionarily conserved and are thought to be regulate in cell fate determination during development [6]. During embryogenesis, this family of genes participates in the development of neuronal tissue, nervous system and as well as skeletal tissue [6]. SOX4 comprises three domains - a serine-rich region, a glycine-rich region and an HMG box [7]. This gene participates in tumorigenesis and progression. It has also been verified that SOX4 regulates lymphocyte differentiation and development, and drive endocardial ridge development [8]. It is emerging that SOX4 is markedly upregulated in various human cancers, including breast cancer [9, 10], colorectal cancer [11], gastric cancer [12] and HCC [13, 14]. SOX4 expression has also been associated with prognosis of some cancer types [7, 15]. Lack of SOX4 expression in normal adult liver does not affect normal liver function [16]. Nevertheless, the exact role of SOX4 gene along the clinical course of HCC is yet to be fully uncovered. This study evaluated the prognostic and diagnostic value of SOX4 in HBV-associated HCC.

Materials and methods

Date source


Bioinformatic analysis and SOX4 diagnostic potential

In order to investigate the biological function and pathways associated with SOX4, we performed a gene ontology (GO) term analysis of SOX4 using the bioware gene ontology tool (BinGO) in Cytoscape_ version3.4.0. GeneMANIA (http://www.genemania.org/, accessed December 17, 2017) [20, 21] and STRING (https://string-db.org/, accessed December 17, 2017) [22, 23] were used to investigate SOX4 gene-gene and protein-protein interactions, respectively. In order to explore the diagnostic value of the SOX4, we used T-test analysis to compare its expression in tumor vs non-tumor tissues in the GSE14520 dataset and then conducted a receiver operating characteristic (ROC) analysis. Diagnostic value was considered statistically significant when P<0.05 and area under curve (AUC) >0.7.

Survival analysis, joint-effect and stratified analysis

For survival analysis, patient data were divided into 2 categories on the basis of median SOX4 mRNA expression. RFS (recurrence-free survival) and OS (overall survival) were estimated using Cox proportional hazards regression and Kaplan-Meier models. Clinical factors found to be statistically significant were adjusted for survival analysis and joint effects survival analysis for SOX4. SOX4 gene was then combined with AFP for survival analysis. Furthermore, SOX4 gene expression was subclassified to effectively perform analyses for clinical factors. Next, factors found to be significant were included in multivariate analysis.

Gene set enrichment analysis (GSEA)

To investigate the prognostic value of SOX4 in HCC, the difference in biological functions and pathways between high and low SOX4 expression groups were analyzed using GSEA (http://software.broadinstitute.org/gsea/index.jsp, accessed December 19, 2017) [24, 25]. This analysis evaluated the molecular signatures database (MSigDB) of c2 (c2.cp.kegg.v6.1.symbols) and c5 (c5.all.v6.1.symbols). Differences were considered statistically significant is the P value <0.05 and false discovery rate <0.25.

Nomogram construction of SOX4 and prognosis-related clinical factors

Nomogram analysis was used to 1, 3-, and 5-year OS and RFS. The nomogram was constructed using prognosis-related clinical factors and SOX4 expression. Different factors and genes had different expression scores.

Validation cohorts

To validate the diagnostic value of SOX4, we analyzed SOX4 expression in tumor tissue vs tumor-adjacent normal tissue using the Mas [26] and Wurmbach [27] liver datasets on Oncomine (http://www.oncomine.org/, accessed December 21, 2017). Differential SOX4 expression analysis was also performed using the aforementioned datasets. SOX4 expression scatter plots, staging verification and prognosis verification were conducted using gene

Statistical analysis

Data were analyzed using SPSS version 24.0 (IBM corporation, Armonk, NY, USA) and R 3.6.0. The log-rank $P$ and median survival time (MST) were determined using Kaplan-Meier method. The hazard ratio (HR) and 95% confidence interval (CI) were estimated using univariate and multivariate Cox proportional hazards regression models. Differential SOX4 expression between tumor and non-tumor tissue was analyzed by T-test. $P < 0.05$ was considered significant.

Results

Differential expressions and diagnostic analysis

Our analysis of the GSE14520 and MERAV dataset revealed elevated SOX4 expression in HCC tissues relative to normal tissue ($P < 0.001$, Figure 1A-B). The ROC analysis of SOX4 in the GSE14520 dataset, HBV-related HCC cohort indicated that SOX4 had a high accuracy of distinguishing tumor tissues from adjacent non-tumor liver tissues ($P < 0.001$, AUC of the ROC curves = 0.782; Figure 1C).

Survival analysis of SOX4 in OS and RFS

In order to avoid the batch effect of microarray data, only the dataset of Affymetrix HT Human Genome U133A Array of GSE14520 was included in the current study. Because most of the patients in GSE14520 were HBV-related HCC, we excluded those patients without HBV infection reports and survival information. As a result, there were 212 HBV-related HCC patients were included in the current study, and all of the 212 HBV-related HCC patients had prognosis information. Our analysis of suggested that the cirrhosis and Barcelona Clinic Liver Cancer (BCLC) stage significantly correlate with OS and RFS (OS: $P = 0.041$, $< 0.001$; RFS: $0.036$, $< 0.001$ respectively; Supplementary Table 1) [29]. Tumor size and AFP significantly associate with OS ($P = 0.002$, 0.049, respectively; Supplementary Table 1). This analysis further revealed that gender significantly correlates with RFS ($P = 0.021$; Supplementary Table 1).

Univariate OS analysis, revealed that SOX4 expression significantly correlates with survival (crude $P < 0.001$, HR = 2.397, 95% CI = 1.522-3.775; Table 1, Figure 2). Similar results were obtained from multivariate OS analysis (adjusted $P = 0.004$, HR = 2.055, 95% CI = 1.261-3.349; Table 1, Figure 2). Univariate analysis of RFS revealed that SOX4 expression significantly correlates with survival (crude $P = 0.001$, HR = 1.896, 95% CI = 1.307-2.750; Table 1, Figure 2) and similar results were obtained by multivariate RFS analysis (adjusted $P = 0.008$, HR = 1.721, 95% CI = 1.151-2.574; Table 1, Figure 2).

Stratified analysis and joint-effect analysis

Stratified analysis of how SOX4 influences OS and RFS indicated age ($\geq$60), being male and single nodular significantly correlate with HCC OS ($P = 0.024$, 0.005 and 0.013, respectively; Figure 3; Table 2). An age of $\geq$60 years, tumor size $> 5$ cm, single nodular and AFP $> 300$ ng/mL were associated with a longer RFS relative to others ($P = 0.025$, 0.019, 0.012 and 0.007 respectively; Figure 4; Table 2).

Analysis of survival on the GSE14520 cohort revealed that SOX4 expression is significantly associated with HCC OS and RFS. Previous studies have reported that AFP is associated with the HCC diagnosis and prognosis. We therefore investigated the combined role of SOX4 and AFP expression on HCC OS and RFS. Analysis of the GSE14520 cohort indicated that the risk of death and recurrence was significantly higher in patients exhibiting high AFP and SOX4 expression when compared to those with low (Figure 2; Table 3).

Prognostic nomogram for survival prediction

Next, we constructed a nomogram for OS based on the following clinical features: BCLC stage, cirrhosis, serum AFP level, tumor size and SOX4 expression. The following clinical features were used to construct a nomogram for RFS: BCLC stage, cirrhosis, gender and SOX4 expression. The nomograms may enable individualized prognosis prediction. Nomogram analysis was performed for the probabilities of 1-, 3- and 5-year OS (Figure 7) and RFS (Figure 8). These analyses revealed that SOX4 expression levels were correlated with the patients’ clinical prognosis.

Bioinformatics analysis of SOX4 gene

Go term analysis indicated that SOX4 gene is involved in the modulation primary alcohol metabolic processes, fatty acid beta oxidation, lipid oxidation, cellular respiration, alpha amino acid metabolic process, small molecule biosynthetic process, organelle inner membrane, mitochondrial matrix and microbody (Figure 5). KEGG functional analysis indicated that the SOX4 gene is involved in various signaling pathways, including insulin and adipocytokine signaling etc. (Figure 6). Detailed representations of the GSEA results are shown in Figure 5 and 6. The visualized interactions of GO terms were constructed using BinGO (Figure 11). This analysis revealed that SOX4 may be involved WNT signaling, lymphocyte differentiation and pancreatic endocrine development. Analysis of gene-gene
interaction found that SOX4 is associated with TP53 etc. (Figure 1D) while analysis of protein-protein interaction found that SOX4 is associated with CTNNB1 and TP53 etc. (Figure 1E).

**Figure 1.** Scatter plots, ROC analysis, and interaction network of SOX4 gene. (A) Scatter plots of SOX4 expression by MERAV website. (B) Scatter plots of SOX4 expression in GSE14520 cohort. (C) Diagnostic receiver operating characteristic curve of SOX4 in GSE14520 cohort. (D) Gene-gene interaction network of SOX4 gene. (E) Protein-protein interaction network of SOX4 protein.

**Table 1.** Prognostic analysis of SOX-4 gene in HBV-related HCC of GSE14250 cohort

<table>
<thead>
<tr>
<th>Type</th>
<th>Gene expression</th>
<th>Patients (n=212)</th>
<th>OS</th>
<th>NO. of event</th>
<th>MRT (months)</th>
<th>Crude HR</th>
<th>Crude P</th>
<th>Adjusted HR (95% CI)</th>
<th>Adjusted P*</th>
</tr>
</thead>
<tbody>
<tr>
<td>OS</td>
<td>Low</td>
<td>106</td>
<td>29</td>
<td>NA</td>
<td>7.9</td>
<td>Ref.</td>
<td>Ref.</td>
<td>Ref.</td>
<td>Ref.</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>106</td>
<td>47</td>
<td>57.9</td>
<td>2.397 (1.522-3.775)</td>
<td>0.000</td>
<td>2.055 (1.261-3.349)</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td>RFS</td>
<td>Low</td>
<td>106</td>
<td>53</td>
<td>46.1</td>
<td>2.397 (1.522-3.775)</td>
<td>0.000</td>
<td>2.055 (1.261-3.349)</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>106</td>
<td>69</td>
<td>26.4</td>
<td>1.896 (1.307-2.750)</td>
<td>0.001</td>
<td>1.721 (1.151-2.574)</td>
<td>0.008</td>
<td></td>
</tr>
</tbody>
</table>

Note: #: P values were adjusted for age, gender, tumor size, multinodular, cirrhosis, AFP and BCLC stage; Bold indicates significant P values.
Figure 2. Univariate prognosis analysis and joint-effect analysis of SOX4. (A) OS analysis plot of SOX4. (B) Joint-effect OS analysis plot of SOX4 and AFP. (C) RFS analysis plot of SOX4. (D) Joint-effect RFS analysis plot of SOX4 and AFP. OS = overall survival, RFS = recurrence-free survival.

Figure 3. Statistical OS analysis plot of SOX4 by clinical factors. (A-B): Statistical OS analysis plot of SOX4 by, (A) age ≤60, (B) >60, (C) male, (D) female, (E) multinodular, and (F) single nodular.
Figure 4. Statistical RFS analysis plot of SOX4 by clinical factors. (A-B) Statistical OS analysis plot of SOX4 by (A) age ≤60, (B) >60, (C) tumor ≤5 cm, (D) tumor >5 cm, (E) multinodular, (F) single nodular, (G) AFP ≤300 ng/ml and (I) AFP >300 ng/ml.

Table 2. Stratified analysis of SOX4 for overall survival and recurrence-free survival

<table>
<thead>
<tr>
<th>Variables</th>
<th>Overall survival</th>
<th>Recurrence-free survival</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤60</td>
<td>80</td>
<td>92</td>
</tr>
<tr>
<td>&gt;60</td>
<td>23</td>
<td>13</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>Female</td>
<td>13</td>
<td>15</td>
</tr>
<tr>
<td>Tumor size (cm)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤5</td>
<td>73</td>
<td>63</td>
</tr>
<tr>
<td>&gt;5</td>
<td>30</td>
<td>42</td>
</tr>
<tr>
<td>Cirrhosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>93</td>
<td>98</td>
</tr>
<tr>
<td>No</td>
<td>10</td>
<td>7</td>
</tr>
</tbody>
</table>
Variables | Overall survival | Recurrence-free survival
--- | --- | ---
| | Low | High | Adjusted HR (95%CI) | Adjusted P value | Low | High | Adjusted HR (95%CI) | Adjusted P value
Multinodular<br>Yes | 19 | 26 | 1.489 (0.534-4.153) | 0.447 | 19 | 26 | 1.322 (0.535-2.628) | 0.546
No | 84 | 79 | 2.044 (1.163-3.592) | 0.013 | 87 | 80 | 1.725 (1.130-2.636) | 0.012
AFP (ng/ml)<br≤300 | 74 | 41 | 1.539 (0.791-2.992) | 0.204 | 74 | 41 | 1.448 (0.856-2.451) | 0.168
>300 | 29 | 64 | 3.553 (1.542-18.186) | 0.003 | 30 | 64 | 2.557 (1.297-5.031) | 0.007
BCLC stage<br>0 | 11 | 9 | 1.871 (0.117-29.956) | 0.658 | 11 | 9 | 4.624 (0.686-31.164) | 0.116
A | 78 | 61 | 1.773 (0.969-3.245) | 0.063 | 81 | 62 | 1.399 (0.884-2.126) | 0.152
B | 8 | 14 | 1.206 (0.247-5.898) | 1.206 | 8 | 14 | 1.026 (0.310-3.393) | 0.996
C | 6 | 21 | 3.801 (0.838-17.234) | 3.801 | 6 | 21 | 2.170 (0.630-7.471) | 0.220

Figure 5. Gene ontology analysis results of SOX4 gene.

Analysis of correlation between SOX4 expression and tumor stage
Analysis of the GSE14520 dataset for SOX4 expression at various BCLC stages revealed significantly elevated expression in each BCLC stage \( P < 0.001 \), Figure 9A, but least expressed in the BCLC stage C. Next, we combined BCLC stage 0 and A to
Constitute the early-stage cancer category and BCLC stage B and C to constitute the advanced-stage cancer category. Interestingly, there was significance lower in former one. Similar results were obtained by GEPIA analysis ($P = 0.00373$; Figure 10C).

**Differential expression, diagnostic and prognostic validation analysis**

Next, we analyzed SOX4 expression in the Wurmbach and Mas liver datasets and found markedly elevated SOX4 mRNA levels in tumor tissue in relative to normal tissue ($P = 0.003, <0.001$, respectively; Figure 9D, F). Moreover, the potential diagnostic value of SOX4 expression was revealed by ROC analysis of these two databases (AUC = 0.831, 0.947 respectively; $P = 0.002, <0.001$, respectively; Figure 9C, E). SOX4 expression was also found to be significantly upregulated in tumor tissue following GEPIA analysis (Figure 10A and B). Analysis of the possible impact of SOX4 expression on survival indicated that patients with low SOX4 expression levels in the GEPIA analysis, exhibit longer OS relative to those with high expression ($P = 0.007$, Figure 10D). Similar results were obtained for RFS ($P = 0.096$, Figure 10E), although this was not statistically significant. In addition, differences in SOX4 gene at various stages of HCC were statistically significant ($P = 0.004$; Figure 10C).

![Figure 6. KEGG pathway analysis results of SOX4 gene.](http://www.jcancer.org)
Figure 7. Nomograms constructed using OS-related clinical factors and SOX4.

Discussion

Here, we assessed the relationship between SOX4 levels and various parameters of HBV-related HCC. Results reveal that SOX4 gene possesses significant value for HCC diagnosis, a finding that is in agreement with previous reports (Wurmbach E et al. and Mas VR et al.) [26, 27]. In addition, we find that low SOX4 expression correlates with better HCC prognosis. Next, we carried out joint-effect and stratified analyses of the value of SOX4 as a prognostic indicator in HCC. GSEA analysis indicated that SOX4 positively modulates primary alcohol metabolic process, fatty acid beta oxidation, lipid oxidation, cellular respiration, small molecule biosynthetic process, alpha amino acid metabolic process, organelle inner membrane, mitochondrial matrix and microbody.

The SOX4 gene belongs to group C SOX transcription factors [30]. The products of these genes consist of three domains: a serine-rich region (SRR, aa 333-397), which encodes a protein of 474 amino acids (aa), a glycine-rich region (aa 152-227) and an HMG box (aa 57-135) [30, 31]. The HMG box possesses DNA binding, which has been take part in various developmental processes through its transcriptional activity, while SRR domain acts as a deactivation domain. Glycine-rich central region (CD), located between the SRR region and the HMG box is a recently identified functional region that promotes apoptosis [15, 31].
The SOX4 gene modulates tumor development and growth, epithelial-mesenchymal transition and metastasis [14, 32-34]. Furthermore, SOX4 drives several components of the RNAi machinery, transcriptional regulators, and cellular proteins [35-37]. Thus, SOX4 is a momentous transcription factor that regulates various cellular functions.

Multiple studies have reported the action of SOX4 as an oncogene in solid tumors [7, 38, 39]. It has been reported that SOX4 is upregulated in various malignancies, including HCC, pancreatic cancer, bladder carcinoma, prostate cancer, breast cancer, colorectal cancer, gastric cancer and melanoma [11, 12, 14, 32, 40-44], raising the potential of this gene as a...
diagnostic marker. The Mas and Wurmbch liver cancer datasets have reported that SOX4 is highly expressed in hepatitis C virus-associated HCC [26, 27], which is consistent with our results. Moreover, various reports suggest that SOX4 can aid in predicting marker in some cancer types. High SOX4 expression has been associated with poor prognosis in prostate cancer, gastric cancer, colorectal cancer, breast cancer and HCC [11, 12, 14, 32, 45]. On the contrary, low expression has been associated with better prognosis in bladder carcinoma and melanoma [41, 43].

Majority of HCCs are attributable to HBV infection [46, 47]. Shang et al. reported that HBV increases expression of SOX4 gene by upregulating transcription factor YY1 via the mitogen-activated protein kinase pathway, epigenetically suppressing miR-203, miR-335, and miR-129-2 by protecting SOX4 from HBsAg mediated degradation [48]. On the other hand, SOX4 has been shown to promote HBV replication by stimulating viral DNA replication and protein expression in liver cancer cells [49]. As a consequence, SOX4 interacts with HBV and synergistically promotes the occurrence and development of HCC.

It was initially found that SOX4 acted as a transcription factor that drive B and T lymphocyte differentiation [8, 50]. Wilson et al. have reported that SOX4 is involved in pancreatic endocrine development [49]. It has been reported that WNT pathway promotes SOX4 expression in colorectal cancer (Van der Flier LG et al. and Reichling T et al.) [51, 52]. SOX4 in turn enhances WNT pathway by stabilizing beta-catenin and directly promoting transcription factor 4 expression [35, 52-55]. Multiple studies have demonstrated that SOX4 interacts with the tumor suppressor p53 during DNA-damage and apoptosis in HCC [15, 56, 57]. A study by Yang Jiao et al. reported that tribbles homolog3 is a SOX4 target [58]. Tribbles homolog3 is a pseudo-kinase that disrupts the insulin signaling pathway in the liver by binding to Protein Kinase B and blocking its activation [59, 60]. However, none has verified the link between TP53, insulin signaling and SOX4. Based on our findings, we hypothesis that SOX4 may modulate TP53 activity and insulin signaling pathway. However, the mechanism still needs further investigated. Consistent with the aforementioned studies, our data show that SOX4 might influence WNT signaling, lymphocyte differentiation and TP53 activity.

Herein, we report that SOX4 is elevated in HCC with BCLC stage B+C than with BCLC stage 0+A. The OS and RFS nomograms indicated that SOX4 is associated with HCC prognosis. Previous studies have shown that SOX4 expression is upregulated in breast cancer [9] and promotes HCC metastases [14], suggesting it might lead to poor

---

**Figure 9.** Scatter plot of SOX4 gene in BCLC stage of GSE14520 cohort and validation using Oncomine database. (A) Scatter plot of SOX4 gene by BCLC stage (0, A, B, C) of GSE14520 cohort. (B) Scatter plot of SOX4 gene by BCLC stage (0+A (early stage), B+C (advance stage) of GSE14520 cohort. (C) ROC curve of SOX4 in Wurmbach liver dataset. (D) ROC curve of SOX4 in Mas liver dataset. (E) ROC curve of SOX4 in tumor tissue and non-tumor tissue in Wurmbach liver dataset. (F) ROC curve of SOX4 in Mas liver dataset.
metastasis-free survival. It has been reported that SOX4 contributes to hepatocarcinogenesis and its expression can reflect the clinical course of HCC after surgical resection [15].

This study is limited by the small sample size, consisting of 212 HBV-associated liver cancer cases. Future studies should utilize a larger sample size. Our analysis was limited to HBV-associated HCC. It is necessary to explore the diagnostic and prognostic value of SOX4 all the HCCs, irrespective of HBV status. Since the data of the two cohorts in this study are from public databases, there is no additional validation cohort. This study still needs to be independently verified in an additional cohort. Relative to past studies, this study only assessed the relationship between SOX4 RNA levels and HCC clinical course. Thus, further investigation is advocated to provide better understanding.

**Figure 10.** Validation of SOX4 gene in GEPIA website. (A) The SOX4 expression profile across all tumor and normal tissues. (B) Differential expressions of SOX4 in GEPIA website. (C) Violin plot of SOX4 by tumor stage in GEPIA website. (D) OS analysis plot of SOX4 in GEPIA website. (E) RFS analysis plot of SOX4 in GEPIA website.
Figure 11. BiNGO analysis results for SOX4 gene.

Conclusions

This study found that SOX4 expression is significantly upregulated in HCC tumor tissues. Our data indicate that this gene has potential value in HCC diagnosis. Further survival analysis of SOX4 gene in two cohorts suggests that it significantly correlates with HCC OS and RFS. Bioinformatics analysis suggested that SOX4 may affect HCC prognosis by modulating TP53 activity, lymphocyte differentiation, pancreatic endocrine development and insulin signaling.

Supplementary Material

Supplementary table S1.
http://www.jcancer.org/v12p3486s1.pdf

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (No.: 81560535), Natural Science Foundation of Guangxi Province of China (Grant No. 2017JBJ140189y, 2018GXNSFAA050119), 2009 Program for New Century Excellent Talents in University (NCET), Guangxi Natural Sciences Foundation (No.: GuiKeGong 1104003A-7), and Guangxi Health Ministry Medicine Grant (Key-Scientific Research-Grant Z201018). The present study is also partly supported by Scientific Research Fund of the Health and Family Planning Commission of Guangxi Zhuang Autonomous Region (Z2016318), the Guangxi Key R & D Program (GKEAB18221019), Key laboratory of High-Incidence-Tumor Prevention & Treatment (Guangxi Medical University), Ministry of Education (GKE2018-01, GKE2019-11), Guangxi Key Laboratory for the Prevention and Control of Viral Hepatitis (No. GXCCKL201902). As well as, the present study is also partly supported by Research Institute of Innovative Think-tank in Guangxi Medical University (The gene-environment interaction in hepatocarcinogenesis in Guangxi HCCs and its translational applications in the HCC prevention). We would also acknowledge the supported by the Key laboratory of High-Incidence-Tumor Prevention & Treatment (Guangxi Medical University), Ministry of Education.

Author contribution

Jian-Lv Huang and Tao Peng designed this manuscript; Xiang-Kun Wang, Xi-Wen Liao, Chuang-Ye Han, Ting-Dong Yu, Ke-Tuan Huang, Cheng-Kun Yang, Xiao-Guang Liu, Long Yu, Guang-Zhi Zhu, Hao Su, Wei Qin, Quan-Fa Han, Zheng-Qian Liu, Xin Zhou, Jun-Qi Liu, Xin-Ping Ye and Tao Peng conducted the study and analyzed the data. Jian-Lv Huang wrote the manuscript, and Tao Peng guided the writing.

http://www.jcancer.org
Availability of data and materials

The datasets analyzed during the current study are available from the corresponding author on reasonable request.

Competing Interests

The authors have declared that no competing interest exists.

References


