Original Article

Plasma IncRNAs acting as fingerprints in predicting hepatocellular carcinoma from HBV positive chronic hepatitis

Li Gong*, Xi Zhu*, Wei Zhang, Huihui Ni, Weihua Yin, Maoying Fu

Department of Infectious Disease, The First People’s Hospital of Kunshan Affiliated with Jiangsu University, Suzhou, China. *Equal contributors.

Received February 6, 2018; Accepted July 14, 2018; Epub November 15, 2018; Published November 30, 2018

Abstract: Current studies have identified that circulating (either from plasma or serum) long non-coding RNAs (IncRNAs) present a potential ability for predicting occurrence or prognosis of multiple human malignant tumors. The present study aimed to screen potential IncRNAs as fingerprints for occurrence of hepatocellular carcinoma in patients with chronic hepatitis (CH). The IncRNA microarray was applied to screen potential biomarkers for HCC from CH patients. Each group contained three individual plasma samples. Multi-stage validation and risk score formula detection was used for validation. Eight dysregulated IncRNAs were obtained after Venny analysis. Further validation in a larger cohort, including 200 HCC patients, 100 CH patients, and 200 healthy controls, confirmed that increased nc-HOXC8-143, XLOC_000667, and AK123675 might be potential biomarkers for predicting early progress of HCC with an area under curve (AUC) of 0.821, 0.660, and 0.728, respectively. The merged AUC of the three factors was 0.907. This study also identified that circulating levels of three IncRNAs were associated with poor post-surgery prognosis of HCC patients.

Keywords: Hcc, long non-coding RNAs, AUC, chronic hepatitis, HBV

Introduction

Hepatocellular carcinoma (HCC) is one of the most common malignancies, closely correlated with hepatitis B virus (HBV) infections [1, 2]. HCC is defined as a primary tumorigenesis in the liver, mainly in patients suffering from chronic liver cirrhosis or hepatitis B or C [3, 4]. The tumor gradually spreads to hepatocytes and, in advanced stages, metastasizes to other organs, such as the lungs and brain [5, 6]. HCC has become very common, causing death and affecting more than 500,000 people in the world. Poor outcomes of HCC are, in large part, due to the lack of timely diagnosis and frequent intrahepatic metastasis [7]. Indeed, vascular invasion, as a representative of intrahepatic metastasis, is a major cause of tumor recurrence within 2 years after resection [8]. To date, alpha-fetoprotein (AFP) measurements have been widely used in clinical practice. However, a recent study has shown that this assessment lacked adequate sensitivity and specificity for effective surveillance and diagnosis of HCC [9]. Therefore, seeking effective early biomarkers for diagnosis and targets for therapy is essential for diagnosis and treatment of HCC.

The human genome contains about 20,000 protein-coding genes, accounting for less than 2% of genome sequence [10]. Long non-coding RNAs (IncRNAs) are an important kind of non-coding RNA transcripts that range from 200 nt to 100 kb without protein-coding capacity [11]. Recently, many investigators have reported that IncRNAs play critically important roles in biological regulation, occurrence, and development of disease [12, 13]. It has been widely recognized that IncRNAs play crucial roles in the regulation of multiple biological processes, including proliferation, differentiation, apoptosis, tumorigenesis, and metastasis [14, 15]. An increasing number of circulating IncRNAs have been demonstrated to be dysregulated in plasma or serum, demonstrating their high potential as powerful and noninvasive tumor markers.
Circulating lncRNA in HCC from CH

For example, Tang et al. reported that three lncRNAs (RP11-160H22.5, XLOC_014172, and LOC149086) were upregulated in HCC relative to cancer-free controls. Furthermore, XLOC_014172 and LOC149086 have been confirmed to be markedly increased in metastatic HCC [18, 19]. However, little evidence has been found for predicting HCC patients from HBV positive chronic hepatitis.

The present study aimed to characterize the genome-wide lncRNAs expression profile, in plasma from HCC patients and CH patients. Comparing with cancer-free controls, this study used lncRNAs microarrays to identify a panel of plasma lncRNAs that might serve as a novel biomarker for diagnosis of HCC.

Materials and methods

Samples and screening phase

The present study enrolled a total of 200 patients diagnosed with HCC. They were pathologically diagnosed in The First People’s Hospital of Kunshan, Affiliated with Jiangsu University, between 2012 and 2016. This study also enrolled 100 patients suffering from HBV induced CH. All HCC patients and CH patients were confirmed with HBV infection. Patients with other malignant disorders, previous history of cancers, radiotherapy, and chemotherapy were excluded. The 200 cancer-free control subjects were confirmed without any malignant disorders or congenital diseases. This study was approved by the Institutional Ethics Committee of Jiangsu University. All research was performed in compliance with government policies and the Helsinki Declaration. Experiments were undertaken with the understanding and written consent of each subject.

Clinicopathological relevance analysis of all 500 individuals is summarized in Table 1. All 200 HCC patients enrolled in this study were clinically and pathologically diagnosed. There were no significant differences in distribution of age and sex between cancer/CH patients and cancer-free controls. The screening phase was divided into a training set and validation set. Twenty pairs of samples, including 20 patients with HCC, 20 patients with CH, and 20 healthy controls were enrolled, while the validation set contained two cohorts.

Training set

All candidates were tested in an independent cohort of 20 plasma samples obtained from patients. Expression levels of these candidates were analyzed in all samples and comparative $2^{\Delta\Delta Ct}$ method algorithms were used to analyze differences between patients and health controls.

Validation set

A case-control study was designed to validate obvious differences of relative expression levels of selected potential biomarker candidates in another two independent cohorts, including 200 HCC patients, 100 CH patients, and 200 healthy controls. These were named as cohort I and II, respectively.

### Table 1. Clinic pathological analysis of HCC patients, CH patients, and cancer-free control samples

<table>
<thead>
<tr>
<th></th>
<th>HCC</th>
<th>CH</th>
<th>Control</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>200</td>
<td>100</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>Age Mean (SE) year</td>
<td>56.11 (10.2)</td>
<td>57.09 (9.3)</td>
<td>57.29 (10.3)</td>
<td>0.52a</td>
</tr>
<tr>
<td>Sex (male/female)</td>
<td>125/75</td>
<td>68/32</td>
<td>126/74</td>
<td>0.22b</td>
</tr>
<tr>
<td>Tumor Number</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solitary</td>
<td>160</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Differentiation grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Well</td>
<td>66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poorly</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tumor Size (cm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤3 cm</td>
<td>118</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;3 cm</td>
<td>82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNM stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-II</td>
<td>128</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>72</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metastasis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>122</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>78</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*a* Student’s t-test. *b* Chi-squared test.
Circulating lncRNA in HCC from CH

Quantitative real-time PCR (qRT-PCR)

Total RNA was obtained from plasma samples using TRIzol Reagent, as described by the manufacturer (Invitrogen Life Technologies Co, Carlsbad, CA, USA). For mRNA detection, total RNAs (500 ng) were reverse transcribed using a reverse transcription kit (Takara, Tokyo, Japan). The quality of RNA samples was assessed by a UV spectrophotometer (Bio-Rad, USA).
Circulating lncRNA in HCC from CH


Hercules, CA, USA) and the 260/280 nm absorbance ratio of samples was limited to 1.8-2.0. QRT-PCR was performed using ABI Prism 7900 HT (Applied Biosystems, CA, USA).

Risk score analysis

Risk score analysis was performed to evaluate association between the concentrations of plasma lncRNA expression levels. The upper 95% reference interval of each lncRNA value in controls or the non-metastasis group was set as the threshold to code the expression level of corresponding lncRNAs for each sample as 0 and 1, in the training set. A risk score function (RSF) to predict TNBC group was defined according to a linear combination of the expression level for each lncRNA. For example, the RSF for sample $i$ using information from three lncRNAs was: $rsfi = \sum_{j=1}^{3} W_j \cdot sij$. In the above equation, $sij$ is the risk score for lncRNA $j$ on sample $i$ and $W_j$ is the weight of the risk score of lncRNA $j$. The risk score of three lncRNAs was calculated using the weight by the regression coefficient derived from univariate logistic regression analysis of each lncRNAs. Samples were ranked according to their RSF. They were then divided into a high-risk group, representing TNBC patients, and a low-risk group, representing predicted control individuals or non-metastasis patients. Frequency tables and ROC curves were then used to evaluate the diagnostic effects of profiling, to find the appropriate cut-off point, and to validate the procedure and cutoffs in the next validation sample set.

Statistical analysis

Venny online tool (http://bioinfogp.cnb.csic.es/tools/venny/index.html) was employed in screening increased lncRNAs in the HCC group and CH group. All data are expressed as mean ± SEM. Statistical analysis was performed with Student’s t-test for comparison of two groups and analysis of variance (ANOVA) for multiple comparisons. The statistical significance of microarray results was analyzed in terms of fold change using Student’s t-test. False discovery rate (FDR) was calculated to correct the P-value. Fold changes ≥4 or ≤0.25 (P<001) were used as threshold values used to screen differentially expressed lncRNAs. Risk score analysis was applied in calculating the potential ability for predicting the ESCC receiver operating characteristic curve, created by STATA 10. Kaplan-Meier survival curves were plotted and log rank test was conducted. The significance of various variables for survival was analyzed by the Cox proportional hazards model in multivariate analysis. Statistical analysis was performed using STATA 10 and presented with GraphPad Prism 5.0 software. Results are considered statistically significant at P<0.05.

Results

Aberrant lncRNA expression profile in ESCC and dysplasia patients

Human LncRNA Array v3.0 was used to detect lncRNAs derived from plasma of 3 patients with HCC, 3 patients with chronic hepatitis, and 3 cancer-free controls. Clustering analysis revealed an aberrant different expression lncRNA profile in these three groups (Figure 1A). A total of 233 lncRNA transcripts were specifically dys-regulated in HCC patients and 318 lncRNAs in dysplasia patients, comparing to the control group. Of these, only 8 lncRNAs were selected, based on aberrant increased levels in both HCC patients and dysplasia patients with a cutoff of 4/0.25 (Figure 1B).

Multiple phase validation

These 8 selected lncRNAs were further amplified in a larger sample, including 200 HCC patients, 100 CH patients, and 200 healthy controls. Only three lncRNAs (nc-FOX0-8-143, XLOC_000667, and AK123675) were validated as significantly upregulated in both HCC patients and CH patients, comparing to the control group, while the remaining 5 lncRNAs presented a disqualification. Results are presented in Figure 1C.

To assess the diagnostic value of the three lncRNAs profiling system, a risk score formula

<table>
<thead>
<tr>
<th>Score</th>
<th>0-4.883</th>
<th>4.883-10.612</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training set</td>
<td>0.95</td>
<td>0.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HCC</td>
<td>1</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CH</td>
<td>19</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Validation set</td>
<td>0.89</td>
<td>0.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HCC</td>
<td>21</td>
<td>179</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CH</td>
<td>175</td>
<td>35</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*PPV, positive predictive value. *NPV, negative predictive value.

Table 2. Risk score analysis

Student’s t-test for comparison of two groups and analysis of variance (ANOVA) for multiple comparisons. The statistical significance of microarray results was analyzed in terms of fold change using Student’s t-test. False discovery rate (FDR) was calculated to correct the P-value. Fold changes ≥4 or ≤0.25 (P<001) were used as threshold values used to screen differentially expressed lncRNAs. Risk score analysis was applied in calculating the potential ability for predicting the ESCC receiver operating characteristic curve, created by STATA 10. Kaplan-Meier survival curves were plotted and log rank test was conducted. The significance of various variables for survival was analyzed by the Cox proportional hazards model in multivariate analysis. Statistical analysis was performed using STATA 10 and presented with GraphPad Prism 5.0 software. Results are considered statistically significant at P<0.05.
Circulating lncRNA in HCC from CH

Figure 2. ROC analysis. Receiver operating characteristics (ROC) curves were drawn with the data of plasma IncRNAs separately, merged from 200 patients with HCC, 100 CH patients, and 200 healthy controls. Merged indicates the combination of three IncRNAs.
was used to calculate the risk score function for cases and control samples. Taking the three lncRNAs as candidates, 60 individuals were chosen in the three groups mentioned above (each group contained 20 individuals), randomly, as the training set. Expression of the three lncRNAs was calculated through risk score analysis. Based on risk score analysis, the optimal cutoff value (Value=4.883) was obtained with the value of sensitivity + specificity considered to be maximal. Plasma samples were then divided into a high-risk group, representing the possible HCC group, and a low-risk group, representing predicted controls. Positive predictive and negative predictive values were 95% and 95%, in the training set, respectively. Similarly, when the same cutoff value was applied to calculate the risk score of samples in the larger validation sets (the rest samples in each group), positive predictive and negative predictive values were 89% and 87% (Table 2).

ROC curve analysis was used to evaluate the predicting diagnosis value of lncRNAs for HCC from CH patients. Areas under the ROC curves (AUC) of the validated three lncRNAs signature were 0.767, 0.874, and 0.855, respectively, in the training set. Combination of the three factors possessed a moderate ability for discrimination between HCC patients and controls, with an area under the ROC curve of 0.942. Further validation set analysis revealed AUC of 0.821, 0.660, 0.728, and 0.907, respectively (Figure 2).

Circulating levels of lncRNAs associated with poor prognosis of HCC patients

Five-year survival analysis, reflecting the diagnosis value of these three indices on prognosis, was carried out using a Kaplan-Meier curve according to the follow up data. All patients were sub-divided into high and low expression groups, according to 95% CI in the normal control group. All three lncRNAs indices were bad prognosis indicators for post-surgery HCC patients. Of these, significant survival differences were found in three lncRNAs: nc-HOXC8-143 (P=0.00882, HR=1.667), XLOC_000667 (P=0.0121, HR=1.778), and AK123675 (P=0.0231, HR=1.771) (Figure 3).

Stability detection of lncRNAs in human plasma

Next, this study amplified the three lncRNAs in five healthy controls and detected the product of amplification by agarose electrophoresis. The bands presented in Figure 4 indicate that all three lncRNAs were detectable in human plasma. Human plasma obtained from three healthy controls was incubated at room temperature for 0 hours, 12 hours, 24 hours, and 48 hours, then treated with frozen-thawing for 3 cycles. It was found that neither the expression level of the three lncRNAs was alternated, indicating that nc-HOXC8-143, XLOC_000667, and AK123675 were stably expressed and detectable in human plasma.
Discussion

The current study comprehensively analyzed IncRNA expression in HBV-related HCC, identifying three IncRNAs that might be fingerprints for HCC. Compared with previous reports, the present study had several advantages. First, HBV positive CH patients and primary HCC were treated as control groups to identify IncRNAs that were specifically differentially expressed in HBV-related HCC. Second, based on the development of HCC in a Chinese Han population, the biomarker for HCC from CH patients might be more effective. Third, compared with other reports, the present sample size was larger, perhaps providing better statistical efficacy.

Biomarkers are a vitally important part of clinical diagnosis and treatment. Many studies have focused on identifying effective biomarkers (especially blood biomarkers) because they are easy to obtain, do not require invasive methods, and have a lower economic burden. To date, the serum concentration of AFP is the most frequently used marker for diagnosis of HCC. However, increasing data has revealed that the sensitivity and specificity of AFP are not sufficient for effective diagnosis [20]. Therefore, there is an urgent need for more effective biomarkers for HCC. Researchers have identified multiple factors that might be potential fingerprints. However, most of these were conducted by comparing expression of these IncRNAs in a case-control study. The control group was healthy volunteers. This kind of biomarker might be useful for distinguishing HCC patients from healthy humans. However, the HCC patients were developed from HBV positive patients and most of them were suffering with CH or liver cirrhosis. Biomarkers for predicting HCC from patients with CH or liver cirrhosis might be useful for early diagnosis of HCC.

Generally, IncRNAs have been considered junk RNA and transcription noise. Recent evidence, however, has proven that IncRNAs are involved in various biological and pathological processes [14, 21, 22]. The present study found that 3 of the 8 candidate IncRNAs could be detected in HCC patients, CH patients, and healthy individuals, indicating that some IncRNAs might have extreme tissue specificity. Using qRT-PCR, it was demonstrated that expression levels of 3 cancer-associated IncRNAs varied significantly between the plasma of HCC patients, CH patients, and healthy controls. Additionally, risk score analysis revealed that three IncRNAs demonstrated great ability to discriminate between HCC patients and CH patients, while also predicting poor prognosis.

In conclusion, this study identified three IncRNAs, nc-HOXC8-143, XLOC_000667, and AK123675, as potential biomarkers for occurrence of HCC. These factors may serve as both diagnosis and prognosis biomarkers for HCC, especially for patients with an HBV infection background.

Disclosure of conflict of interest

None.

Address correspondence to: Maoying Fu, Department of Infectious Disease, The First People's Hospital of Kunshan Affiliated with Jiangsu University, Suzhou, China. Tel: +86-512-57559009; Fax: +86-512-57559009; E-mail: fumaoying66@yeah.net
References


Circulating IncRNA in HCC from CH
