MAFA-AS1, a long non-coding RNA, predicts for poor survival of hepatocellular carcinoma

Yuting Zhan¹, Xin-Yuan Guan¹², Yan Li¹

¹State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-Sen University Cancer Center, Guangzhou 510060, China; ²Department of Clinical Oncology, The University of Hong Kong, Hong Kong, China

Contributions: (I) Conception and design: Y Zhan, Y Li; (II) Administrative support: XY Guan; (III) Provision of study materials or patients: All authors; (IV) Collection and assembly of data: Y Zhan, Y Li; (V) Data analysis and interpretation: Y Zhan, Y Li; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Dr. Yan Li. State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-Sen University Cancer Center, Guangzhou 510060, China. Email: liy6@mail.sysu.edu.cn.

Background: Hepatocellular carcinoma (HCC) is a common cancer with high morbidity and mortality, especially in East Asia. Reliable biomarkers for HCC are indispensable given the absence of capable prediction of prognosis. Long non-coding RNAs (lncRNAs) are the most abundant products of transcription, which might serve as robust markers for diagnosis or treatment.

Methods: Multiple bioinformatics tools were utilized to screen indicative lncRNAs for HCC concurrently with the underlying mechanism of its role in tumorigenesis and development. We analyzed the genome-wide alterations and transcriptomics profiles of HCC and non-tumor samples from The Cancer Genome Atlas (TCGA). Survival analyses were also applied. Integrated bioinformatics analyses were applied to identify co-expressed genes along with chromosome loci, predictive RNA binding proteins (RBPs) and co-expressed miRNAs.

Results: Combining copy number variations (CNVs) with expressions of lncRNAs, 20 most aberrant lncRNAs were identified. MAFA-AS1 was identified as a prognostic indicator of HCC poor overall survival (OS) and disease-free survival (DFS). Multiple bioinformatics analyses suggest that MAFA-AS1 is involved in HCC progression via interacting with SRSF1/SRSF9 and coordinating with miR210.

Conclusions: MAFA-AS1 is a prognostic biomarker for poor OS and DFS of patients with HCC. It is involved in HCC progression.

Keywords: Hepatocellular carcinoma (HCC); long non-coding RNA (lncRNA); MAFA-AS1; prognosis

Submitted Nov 12, 2019. Accepted for publication Feb 21, 2020.
doi: 10.21037/tcr.2020.03.11

View this article at: http://dx.doi.org/10.21037/tcr.2020.03.11

Introduction

Hepatocellular carcinoma (HCC) is one of leading cancers worldwide (50% in China alone). It remains the second cause of death resulted from cancer (1). In China, HCC was the third most common cancer and the second most lethal tumor (2). Although there have been advancements in diagnosis and treatment recently, only a small group of patients receiving surgeries are completely relieved. Given the poor prognosis of patients with HCC, more precise and detailed work is indispensable. As we know, carcinogenesis is a multi-step and multi-factor driven process (3). In addition to some well-known factors [e.g. hepatitis B virus (HBV) infection or alcoholism], micro-RNAs (miRNAs), long non-coding RNAs (lncRNAs) and genetic alterations also contribute to HCC tumorigenesis and progression (4,5).

lncRNAs are a group of non-coding RNAs that are considered to be longer than 200 nucleotides (6). They make up most of our genome and was recognized as useless noises, since they are not able to produce proteins (7).
However, expression profiles of IncRNA bear tissue specificities, which means they are not functionally redundant (6,8). Actually, IncRNAs could act as transcription co-factors to regulate expression of adjacent protein-coding genes, interact with RNA binding proteins (RBPs) and regulate genes in cis, as well as affect epigenetics via chromatin modification (8,9). Other evidences implied that aberrant expression of IncRNAs was involved in pathological conditions like lung cancer, HCC and other cancers (10,11). Recent studies indicate that some exosomal IncRNAs are promising biomarkers for many cancer (12). Micro-RNAs belong to small non-coding RNA, which is about 21–22 nucleotides in length (13). Deregulation of miRNAs could become either engine or brake of oncogenesis (14). Previous literatures have shown that both IncRNAs and miRNAs could be key regulators in cancer stem cell differentiation and self-renew ability (15).

A variety of IncRNAs have been proved to be aberrantly expressed in tumor, for example, Yan et al. found the imbalanced up-regulation of H19 in gastric cancer tissues (16). Recently, some IncRNAs have been recognized as novel prognostic biomarkers of tumor, for example, IncRNA GAS5 expression is higher in non-small cell lung cancer (NSCLC) tumor tissues than in normal tissues, and high expression of GAS5 was correlated with advanced clinical stage and poorer outcome (17). Additionally, many studies showed that abnormal expression of IncRNAs such as H19, HOTAIR and MALAT1 might contribute to tumor development (16,18,19). Many IncRNAs promote or inhibit carcinogenesis via interacting with mRNAs (20). MiRNAs can conditionally bind IncRNA on specific site and reduce the decay of target genes (21). Since complicated interaction exist among IncRNA, miRNAs and associated proteins of HCC, a more detailed visualization of the network is necessary. In this study, some candidate oncogenic or tumor-suppressive IncRNAs in HCC were screened by integrative bioinformatics analyses. MAFA-AS1 was identified as a candidate IncRNA because of its prognostic value for overall and disease-free survival (DFS) of patients with HCC. It also has potential to interact with SRSF1 and/or SRSF9 and affect DNA replication and cell cycle. Additionally, MAFA-AS1 could interact with miR210, an important oncogenic miRNA in many cancers (22,23). The molecular interactions among IncRNAs, miRNAs and associated proteins were explored.

### Methods

#### Statement of ethics approval

The Statement of Ethics Approval is not required because this is a bioinformatics study.

#### Screening of aberrantly expressed and genomically altered IncRNAs in HCC samples

The Cancer Genome Atlas (TCGA) (http://cancergenome.nih.gov/) is a public cancer database of genomic and clinical information containing RNA-seq, single-nucleotide variations (SNVs) and copy number variations (CNVs), in 33 kinds of malignant tumors. Firstly, RNA-Seq-HT seq-count raw data of LIHC including 371 HCC tissues and 50 liver tissues was downloaded from TCGA (24). The raw data were inputted into R to analyze and visualize the aberrantly expressed IncRNAs. Fold change (FC) >2 ($\log_{2}^{\text{FC}}$>1) and P<0.05 between tumor and non-tumor tissues were considered as significant. As further narrow down the threshold to $\log_{2}^{\text{FC}}$>1.5, the eligible IncRNAs were screened by edgeR package and limma package, respectively. Two lists of altered IncRNAs were integrated, and finally 1384 IncRNAs were achieved. These IncRNAs were then imported into DAVID (www.DAVID.com) and HGNC (www.genenames.org) to get gene symbol of each IncRNA (25,26).

To acquire those officially acknowledged IncRNAs in transcriptome and genome simultaneously, the most differentially expressed officially approved IncRNAs were imported into the OncoPrint module of cBioportal (http://www.cbioportal.org) for genomic analysis (27,28). The IncRNAs changed in more than 5% of the cases were taken as remarkable ones. The overlapping of genomic altered IncRNAs and the abnormally expressed IncRNAs (log$_{2}$ FC>2, P<0.05) are implemented by VENN (http://bioinfogp.cnb.csic.es/tools/venny/index.html). Hierarchical cluster analyses are performed by R in ggplot2 package (29).

#### Clinical properties and survival analysis in HCC

To further investigate the association between IncRNAs and the Clinicopathologic characters (survival, stage and grade, etc.), GEPIA (http://gepia.cancer-pku.cn/) was applied to analyze the survival and stages of patients with IncRNAs.
based on TCGA RNA-Seq data (30). CBioportal, another online tool, was used for analyzing survival data of HCC cases with or without CNVs (27,28). Visualization of overall survival (OS), DFS and stages were analyzed by GEPIA and CBioportal. Correlations between tumor grades and lncRNAs were evaluated and visualized by TANRIC (http://ibl.mdanderson.org/tanric/_design/basic/index.html) (31). Kaplan-Meier method was used to generate survival curves. The association of lncRNAs with TNM stage or grade was evaluated by ANOVA or student t-test. P<0.05 was set as statistically significant.

Integration of co-expressed genes, RNA-protein interaction network, gene ontology/pathway enrichment analysis and associated miRNAs

In order to probe the underlying molecular mechanism of MAFA-AS1 in HCC, we took advantage of circlncRNA database (http://120.126.1.61/circlnc/circlncRNAet/lncRNA_TCGA/index.php) to calculate the co-expressed genes of MAFA-AS1 along with their chromosomal locations. Co-expressed genes with Pearson P value below 0.05 on the basis of TCGA LIHC samples were regarded as significant. Besides, we conducted GO/KEGG analysis and RBPs were predicted by circlncRNAet (32). In addition to looking for related lncRNAs in TANRIC, we utilized LinkedOmics website (http://www.linkedomics.org/login.php) and GEPIA to analyze the survival information of patients according to profiles of the RBPs and miRNAs.

Results

Aberrantly expressed lncRNAs in TCGA LIHC samples

Raw data of TCGA LIHC was inputted into R and differentially expressed lncRNAs were analyzed. Following the criteria (log_{2FC}>1, P<0.05), there are 1,932 lncRNAs considered to be dysregulated between tumor and nontumor tissues, with 1722 upregulated and 210 downregulated lncRNAs (Figure 1A). With log_{2FC} >1.5 (P<0.05) threshold, 1,384 lncRNAs were enrolled into our study. These lncRNAs were integrated into DAVID (www.DAVID.com) and HGNC (www.genenames.org) to blast with official gene symbols, only 454 symbols were acknowledged by HGNC. Indeed, 230 acknowledged lncRNAs in total conformed to a more stringent standard (log_{2FC}>2).

LncRNAs with highest frequency in CNVs

Genomic changes tend to play a key role in many tumors including HCC. We looked for meaningful mutations in genome, but no valuable lncRNA mutation was observed. However, the CNVs, including amplification and deep deletion, were found in those lncRNAs. CNV information of 453 lncRNAs was obtained, except AC005150.1, whose data was not available in CBioportal. There are 45 of 453 lncRNAs were altered in more than 5% of the HCC patients according to CBioportal. Among them, PVT1 had the highest frequency (24%) and most lncRNAs amplified rather than deleted (Figure 1B).

LncRNAs are systematically dysregulated in genomics and transcriptomics in HCC

It is believed that lncRNAs with concurrent changes in transcriptome and genome might be critical in HCC. We extracted 230 differentially expressed lncRNAs mentioned above and 45 lncRNAs with obvious CNVs. By integrating bioinformatics analysis in VENNY, a total of 11 lncRNAs including CDKN2A-AS, BPESC1, ELF9N2, CASC9, C17ORF82, RMST, TSPER-AS2, PVT1, LINC00200, C2ORF48, GUSBP11, were identified from the datasets (Figure 1C). Ten most differentially expressed lncRNAs and 10 most frequently CNV-altered lncRNAs were also included for gene expression analysis (Figure 1D).

LncRNAs correlate to poor survival rates of patients with HCC

Twenty lncRNAs were retrieved in GEPIA and CBioportal to validate the correlation between lncRNAs and the clinical outcomes based on TCGA database. Among them, the high expression of LINC00200, MAFA-AS1, CASC8, CASC9, LINC01667, CDKN2A-AS1 indicates poor five-year OS rate in HCC (Figure 2A). Kaplan-Meier analysis revealed that high expression of MAFA-AS1, CASC9, LINC01667, CDKN2A-AS1 were significantly correlated with poorer DFS rates of HCC patients (Figure 2B). With the help from CBioportal, we also analyzed the survival data of those lncRNAs to determine whether HCC patients with genomic alterations in those lncRNAs had poor survival outcome. However, no association between the CNVs of the lncRNAs and survival has been found.
MAFA-AS1 is a candidate biomarker for poor prognosis of HCC patients

The clinical-pathological properties also reflect for prognosis in HCC (33). We used GEPIA and TANRIC to determine the correlations between the lncRNAs and clinical characters of HCC. The profile of MAFA-AS1 was able to distinguish early-stage HCC patients from advanced stage patients (Figure 3A). The other three lncRNAs (LINC01667, CASC9, CDKN2A-AS1) showed positive correlation with stages except stage IV (Figure 3A). Besides, high expression of MAFA-AS1 was observed in higher tumor grade, which means loss of differentiation, a sign of extremely malignancy (Figure 3B). Taken together, these results indicate that MAFA-AS1 might be a novel prognostic marker for HCC.

MAFA-AS1 interacts with SRSF1/SRSF9 and miR210

Because MAFA-AS1 high expression was significantly associated with worse outcome in HCC, the role of MAFA-AS1 in HCC progression was further explored. Recent studies have revealed that one of the major
Figure 2 lncRNAs associated with HCC survival. (A) Overall survival curves based on six different lncRNAs (red, high expression; blue, low expression; Kaplan-Meier method) (log-rank test). (B) Kaplan-Meier analysis indicates the correlation of four lncRNAs with disease-free survival rates of HCC patients. HCC, hepatocellular carcinoma; lncRNA, long non-coding RNA.
functions of IncRNA was to mediate the expression of nearby genes (8). We were wondering whether some oncogenes or tumor suppressors were regulated by MAFA-AS1. We evaluated certain genes correlating with MAFA-AS1 in HCC by TCGA co-occurrence analyses (circlncRNA.net). Fifty most correlated co-expressing genes based on TCGA LIHC samples were presented by heatmap (Figure 4A). The chromosomal locations of these genes are presented by Circos plot. As shown in Figure 4B, the co-expressing genes are randomly located on almost all chromosomes except chromosome X and Y. The locations of those genes are far from the MAFA-AS1 locus, suggesting that MAFA-AS1 might not affect those genes via space interactions.

LncRNAs were reported to bind to RBPs as partner and adjust stability of those proteins only on the condition that lncRNAs have a considerable firm structural conformation (34). We next characterized the secondary structure of MAFA-AS1 in RNAfold web server. As Figure 5A shown, MAFA-AS1 has a stable secondary structure. Through analysis of related RBPs in circlncRNA.net, SRSF1 and SRSF9 were predicted to interact with MAFA-AS1 (Figure 5B). Survival analyses suggested that high expression of SRSF1 or SRSF9 indicated poor OS and DFS (Figure 5C, D). SRSF1/SRSF9 was reported to play a key role in HCC progression by modulating DNA damage repair mechanism and cell
cycle (35). Additionally, MAFA-AS1 mainly affects DNA replication and cell cycle by GO and KEGG analyses (Figure 5E,F).

Besides, it was also predicted that MAFA-AS1 correlated with hsa-miR-210 via TANRIC (R=0.401, P<0.001). MiR210 is a hypoxia specific miRNA participating in many cancers (23). MiR210 high expression predicts for poor survival of patients with HCC (Figure 5G). It is significantly upregulated in HCC patients with lymphoid node metastasis (Figure 5H).

**Discussion**

Although lots of studies were conducted to investigate the mechanism of HCC tumorigenesis and progress, it remains perplexing and contradictory. One reason is that most
Figure 5 MAFA-AS1 potentially interacts with SRSF1/SRSF9 and coordinates with miR210. (A) Structure of MAFA-AS1 implies several stem loops and stable secondary structure. (B) Predictive RNA binding proteins of MAFA-AS1. (C) Kaplan-Meier plots show the OS and DFS of SRSF1 (red, high expression; blue, low expression). (D) Kaplan-Meier plots show the OS and DFS of SRSF9 (red, high expression; blue, low expression). (E,F) GO analysis (E) and KEGG analysis (F) results were listed. (G) Kaplan-Meier plot shows OS of miR210 (red, high expression; blue, low expression). (H) Box plot shows the relative expression of miR210 in the N1 and N0 groups (P<0.05, Wilcoxon Test).
results were generated from a single cohort and/or small sample study. Integrating the genomic and transcriptomic information from multiple databases and conducting bioinformatics analyses is a practical way to find potential biomarkers and therapeutic targets in HCC tumorigenesis and progression. Many lncRNAs were reported to be correlated with clinical outcomes of HCC patients (36). However, considering the complex factors and mechanisms accounting for HCC tumorigenesis, more detailed and accurate analyses were necessary. Here, we identified a cluster of upregulated lncRNAs (MAFA-AS1, CASC9, LINC01667, CDKN2A-AS1) as indicators for poor prognosis of HCC patients by data-mining LIHC (a TCGA HCC database). One of these lncRNAs, MAFA-AS1, can distinguish early-stage HCC patients from advanced stage patients. The high expression of MAFA-AS1 predicts for poor OS and disease-free survival of HCC patients. Instead of co-acting with the neighboring genes, MAFA-AS1 interacts with RBPs SRSF1 or SRSF9 and it might affect DNA replication and cell cycle. Both SRSF1 and SRSF9 are members of the SR (splicing regulators) protein family (37). SRSF1 is involved in some key parts of mRNA metabolism (for example, mRNA splicing, mRNA stability and mRNA translation) and other processes (for instance, nucleolar stress response, protein sumoylation and miRNA processing) (37). Moreover, SRSF1 and SRSF9 were reported previously to activate Wnt signaling pathways by increasing biosynthesis of β-catenin (38). MAFA-AS1 might exert oncogenic effect in HCC by interacting with SRSF1/ SRSF9 via being scaffold. MAFA-AS1 was also positively correlated with miR210, an onco-miRNA that predicts for poor OS in patients with HCC.

The combination of multiple bioportals with TCGA database provides an efficient way to seek out lncRNAs that can predict clinical outcome. It opens a window to elucidate the possible molecular mechanisms of oncogenic role of lncRNAs in HCC. Our study demonstrates for the first time that MAFA-AS1 is a candidate lncRNA for HCC carcinogenesis via binding splicing factors SRSF1/SRSF9 and co-acting with miR210. Unlike other studies that lncRNA affects tumor progression mainly via the mediation of connected genes (39), our research displays a more comprehensive network of lncRNA-RBPs and miRNAs modulation.

**Acknowledgments**

We thank the website (https://www.shengxin.ren/) and its engineer Rangfei Zhu for the guidance in bioinformatics. We also thank Dr. Lei Li for assistance on the manuscript.

**Footnote**

*Conflicts of Interest:* The authors have no conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The Statement of Ethics Approval is not required because this is a bioinformatics study.

**References**


Cite this article as: Zhan Y, Guan XY, Li Y. MAFA-AS1, a long non-coding RNA, predicts for poor survival of hepatocellular carcinoma. Transl Cancer Res 2020. doi: 10.21037/tcr.2020.03.11