Screening and validation of plasma long non-coding RNAs as biomarkers for the early diagnosis and staging of oral squamous cell carcinoma

HONGCHENG JIA1*, XUAN WANG2* and ZHENG SUN3

1Department of Stomatology, Beijing Ditan Hospital, Capital Medical University, Beijing 100015; 2Department of Stomatology, Beijing Tongren Hospital, Capital Medical University, Beijing 100730; 3Department of Oral Medicine, Beijing Stomatological Hospital, Capital Medical University, Beijing 100050, P.R. China

Received July 31, 2020; Accepted December 9, 2020

DOI: 10.3892/ol.2021.12433

Abstract. Oral squamous cell carcinoma (OSCC), characterized by a high recurrence rate, a poor prognosis and high morbidity, is the most prevalent malignancy of the oral cavity. The aberrant expression of long non-coding RNAs (lncRNAs) may lead to the development of various diseases, including cancer. Delayed diagnosis is the main reason for the poor prognosis. Therefore, the present study aimed to investigate the differential expression profiles of plasma lncRNAs in OSCC in order to screen target lncRNAs as biomarkers for the early diagnosis and staging of OSCC. The expression profiles of lncRNAs and mRNAs in OSCC were analyzed by microarray analysis. A total of 14 candidate lncRNAs were selected and analyzed using reverse transcription-quantitative polymerase chain reaction (RT-qPCR) using the array homologous samples. Subsequently, 4 target lncRNAs were measured by RT-qPCR in a large cohort, including 28 cases with TNM I/II [early-stage squamous cell carcinoma (ESCC) group], 36 cases with TNM III/IV [advanced-stage squamous cell carcinoma (ASCC) group], 16 cases with dysplasia [oral premalignant lesion (OPL) group] and 16 healthy controls (H group). Receiver operating characteristic (ROC) curves and logistic regression analyses revealed that the diagnostic efficacy of the combined lncRNAs was more prominent than that of a single lncRNA, particularly in the ESCC and ASCC groups. In conclusion, the present study identified the differential expression profiles of plasma lncRNAs in OSCC and demonstrated that ENST00000412740, NR_131012, ENST00000588803 and NR_038323 may be promising biomarkers for the early diagnosis and staging of OSCC. These findings may provide novel targets for the early diagnosis and staging of OSCC, which may provide an objective basis for clinical decision-making.

Introduction

Oral squamous cell carcinoma (OSCC) is the most prevalent malignancy of the oral cavity, accounting for >90% of oral cancer cases, with 354,864 estimated new cases and 177,384 mortalities worldwide in 2018 (1,2). OSCC, characterized by a high recurrence, a poor prognosis and high morbidity, severely affects the quality of life of patients. Therefore, OSCC poses a burden to global health. OSCC is preceded in 67% of cases by oral premalignant lesions (OPLs), of which 1-18% undergo malignant transformation into OSCC (3). Patients with early-stage squamous cell carcinoma (ESCC; TNM I and II) survive longer than those with advanced-stage squamous cell carcinoma (ASCC; TNM III and IV), with survival rates of 64.2 and 30.1% for early and late stages, respectively (4). Despite improvements in treatment modalities, the 5-year overall survival rate has improved only marginally, with 33% of cases surviving between 1973 and 2014, compared with 41% between 2006-2011 (5,6). Delayed diagnosis and the lack of accurate and timely treatment, derived from the bias of the standards of clinical decision-making based on the clinical experience and subjective judgment of doctors, are considered to be the major reasons for the poor prognosis.

Correspondence to: Professor Zheng Sun, Department of Oral Medicine, Beijing Stomatological Hospital, Capital Medical University, 4 Tiantanxili, Dongcheng, Beijing 100050, P.R. China E-mail: sunzheng12@vip.126.com

Dr Hongcheng Jia, Department of Stomatology, Beijing Ditan Hospital, Capital Medical University, 8 Jingshundongjie, Chaoyang, Beijing 100015, P.R. China E-mail: jiahongcheng@ccmu.edu.cn

*Contributed equally

Key words: oral squamous cell carcinoma, oral premalignant lesions, long non-coding RNA, biomarker, microarray
invasive, reliable and sensitive marker is urgently required to provide an objective basis for clinical decision-making.

Long non-coding RNAs (lncRNAs) are a class of RNA molecules with transcripts of >200 nucleotides in lengths, which were first discovered as 'transcriptional noise' in 1989. However, increasing evidence has suggested that lncRNAs are involved in gene expression regulations at the epigenetic, transcriptional and post-transcriptional levels, and are essential in physiological events (7-9). The aberrant expression of lncRNAs may directly or indirectly lead to the development of various diseases, including cancer (10,11). lncRNAs may be promising biomarkers in cancer diagnosis and prognosis (12,13). For example, MALAT1 may be used as a marker for the early diagnosis of prostate cancer (14), the upregulation of HOTAIR expression is indicative of a poor prognosis in colon and breast cancer (15), and the downregulation of GAS5 expression is indicative of a poor prognosis in gastric cancer (16). lncRNAs have also been revealed to be differentially expressed in tissues and salivary samples of the normal oral mucosa, in OPLs and OSCC (17-20). However, to the best of our knowledge, the differential expression profiles of lncRNAs in the plasma of patients with OSCC has not yet been reported.

In the present study, the differential expression profiles of plasma lncRNAs in OSCC were first investigated using microarray analysis to screen candidate lncRNAs, followed by reverse transcription-quantitative polymerase chain reaction (RT-qPCR) analysis. Subsequently, the target lncRNAs were further validated by RT-qPCR to estimate the diagnostic efficacy of plasma lncRNAs from patients with OPL, ESCC and ASCC. The results of the present study may provide novel targets for the early diagnosis and staging of OSCC, which may also provide an objective basis for clinical decision-making for the early diagnosis, reasonable implementation of the treatment plan and prognosis evaluation of OSCC.

Materials and methods

Samples. A total of 67 patients with OSCC (39 men and 28 women; age range, 47-75 years; mean age, 63.5 years), 16 patients with OPL (3 cases of mild dysplasia, 7 cases of moderate dysplasia and 6 cases of severe dysplasia) and 19 healthy control individuals (H group) were recruited between December 2013 and May 2015 from Capital Medical University Beijing Stomatological Hospital. A total of three patients with TNM III/IV OSCC and 3 healthy controls (8 males and 8 females; age range, 37-61 years; mean age, 58.1 years; all with epithelial dysplasia) and 16 patients with OPL (9 males and 7 females; age range, 43-72 years; mean age, 63.3 years; TNM staging is presented in Table SII), 16 patients were used for PCR validation, including 64 patients with OSCC microarray analysis (Table SI), and the remaining samples were |log FC|≥10 and co-existing in the general list and the differentially expressed lncRNA list in OPL based on SAGE (21), namely GAS5-AS1, LOC100507156, RP11-539G18.2 and ARHGEF26-AS1; ii) lncRNAs with |FC|≥5 and original expression ≥200 in the sub-category analysis list with tissue-specific set to metastasis, namely CTD-2008L17.1 and LINCO1539; iii) lncRNAs with |FC|≥5 and co-existing in the general list and the differentially expressed lncRNA list in OPL based on SAGE (21), namely LINCO00665 and NEAT1; iv) lncRNAs with |FC|≥10 and co-existing in the general list and the differentially expressed lncRNA list in OSCC based on the microarray (20), namely RP11-250B2.3 and AP001347; and v) the top 3 lncRNAs with the largest FC coexisting in the general list and the tumor-related lncRNA list downloaded from the Lnc2Cancer database (http://www.bio-bigdata.com/Lnc2cancer/down.jsp), namely HOTAIR_4, BCAR4 and MNXI-ASI.
The primers were designed using primer premier 5.0 software (Premier Biosoft International) (Table SIII). The reverse transcription of total RNA was performed in a 20 µl volume containing 500 ng total RNA, 1 µl 10 µM primers, 1.6 µl of 2.5 mM dNTPs mixture, 4 µl 5X First-Strand Buffer, 1 µl 0.1 M DTT, 0.3 µl RNase inhibitor, 0.2 µl SuperScript III RT (Invitrogen; Thermo Fisher Scientific, Inc.) and 14.5 µl water. The program was as follows: 50°C for 60 min, 70°C for 5 min, and 4°C hold. The ViiA 7 Real-time PCR System (Applied Biosystems; Thermo Fisher Scientific, Inc.) was used for the RT-qPCR assay. A total of 2 µl of the cDNA product was used as a template in 10 µl reaction on a 384-well plate containing 5 µl of 2X PCR master mix (Arraystar), 1 µl of 10 µM specific primer, 2 µl of RNase-free water. The conditions were as follows: A denaturation step for 10 min at 95°C, followed by 40 cycles of 10 sec at 95°C and 60 sec at 60°C. Following amplification, the operation of the instrument was performed according to the procedure (95°C, 10 sec; 60°C, 60 sec; 95°C, 15 sec) and slowly heated from 60°C to 99°C (Ramp Rate was 0.05°C/sec). Each experiment was repeated in triplicate. The housekeeping gene used was β-actin. The 2^ΔΔCt method was used to measure relative expression levels (REls) (22).

**Target IncRNA screening and RT-qPCR validation.** A total of 4 IncRNAs were selected to be measured and validated by RT-qPCR in a large cohort. The screening strategies were as follows: i) The two IncRNAs with the top FC among the aforementioned 14 IncRNAs, namely ENST00000412740 and ENST00000588803; ii) the IncRNA with the largest FC in the general list, namely NR_038323; and iii) the key IncRNA in OPL, namely NR_131012. The RT-qPCR procedure was the same as that described earlier, and the primers used are listed in Table SIII.

### Results

**Differentially expressed profiles of IncRNAs and mRNAs in the plasma of patients with OSCC.** Following image acquisition and data analysis, the expression matrices of IncRNAs and mRNAs were obtained. The volcano plot indicated that several IncRNAs and mRNAs were differentially expressed between the OSCC and normal samples (Fig. 1A and B). According to the screening standard, a total of 6,606 IncRNAs and 4,196 mRNAs were differentially expressed in the plasma of patients with OSCC. Furthermore, 3,511 IncRNAs and 1,766 mRNAs were upregulated, and 3,095 IncRNAs and 2,430 mRNAs were downregulated. The top 20 dysregulated IncRNAs and mRNAs are presented in Tables SIV and SV, respectively. Hierarchical clustering analysis revealed that the expression profiles exhibited a good clustering effect on OSCC and normal plasma (Fig. 1C and D). The results of GO and KEGG analyses are presented in Tables SVI-SIX. The data have been deposited in NCBI’s Gene Expression Omnibus and are accessible through GEO series accession no. GSE97251 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE97251).

**Validation of candidate IncRNAs by RT-qPCR.** Compared with the H group, 14 candidate IncRNAs were all differentially expressed in OSCC (Table I). Apart from NR_024050, which was downregulated (no statistical significance), the remaining 13 IncRNAs were all upregulated, and the FC value of 9 IncRNAs exhibited a statistically significant difference. These results were consistent with those of the microarray.

**Validation of target IncRNAs by RT-qPCR.** ENST00000412740, NR_131012, ENST00000588803 and NR_038323 were differentially expressed among the H, OPL, ESCC and ASCC groups (P<0.05). Furthermore, the differential expression of 4 target IncRNAs was compared between groups (Fig. 2). Compared with the H group, ENST00000412740, ENST00000588803 and NR_038323 were downregulated in the OPL group, and notably, they were upregulated in the ASCC group compared with the OPL group (P<0.05). Compared with the H group, NR_038323 was downregulated in the ESCC and ASGC groups (P<0.05). Compared with the OPL group, NR_131012 was upregulated in the ESCC group (P<0.05). Compared with the ESCC group, ENST00000412740 and ENST00000588803 were upregulated in the ASCC group (P<0.05).

**Screening of diagnostic combination and evaluation of the diagnostic efficacy of the 4 target IncRNAs.** ROC curve analysis revealed that the 4 target IncRNAs exhibited excellent discriminative ability for OPL vs. H, OSCC vs. OPL and ASCC vs. ESCC, with an AUC >0.7, apart from NR_131012 and NR_038323, which were considered as having moderate discriminative ability only for ASCC vs. ESCC, with an AUC of 0.558 (95% CI, 0.418-0.698) and 0.590 (95% CI, 0.451-0.728), respectively (Fig. 3). However, they exhibited no discriminative ability for OSCC vs. H, apart from NR_038323 with an AUC >0.7 (Fig. S1). The logistic regression model was performed using the logistf package of R language, and variables were screened by the stepwise optimization method to determine the IncRNA combination with a high diagnostic efficiency. All tests were two-sided and P<0.05 was considered to indicate a statistically significant difference.
revealed that the combined IncRNAs provided a more prominent diagnostic efficacy than a single IncRNA, particularly for ASCC vs. ESCC (Table II). The sensitivity, specificity and cut-off value of each combination of IncRNAs are illustrated in Fig. 4.

**Discussion**

To the best of our knowledge, the present study was the first to identify the differential expression profiles of plasma IncRNAs in OSCC by microarray analysis. The reliability of microarray and quality of the array samples were verified to be credible by RT-qPCR using the array homologous samples. The results revealed that the profile of IncRNAs in plasma from patients with OSCC differed significantly from that of the healthy controls. The majority of the differentially expressed genes have been proven to be involved in the biological process of OSCC by GO and KEGG analyses (20,21). However, there are only limited studies available on the diagnostic role of circulating IncRNAs in OSCC (23).

In the present study cohort, patients with TNM I/II stage OSCC accounted for 39.58% of primary OSCC cases, which
ONCOLOGY LETTERS  21:  172,  2021

was 22% of those involving the posterior third of the tongue reported in the literature (24), indicating that the early diagnosis of OSCC remains relatively low. However, patients with TNM I/II stage OSCC accounted for 47.37% of recurrent OSCC cases, which was slightly higher than that of primary OSCC, which may be associated with regular follow-up after surgery. A specialist may improve the early diagnosis of OSCC; however, this remains insufficient. The early diagnosis and staging of OSCC may aid doctors in determining effective and appropriate treatment strategies, including the scope of surgery, radiotherapy, chemotherapy and other adjuvant therapy, which has an important impact on the quality of life and prognosis of patients. These decisions are largely dependent on the clinical experience and subjective judgment of doctors; however, the lack of objective indicators leads to a bias in the making of these decisions. Therefore, an objective, accurate and minimally invasive biomarker is urgently required.

To date, >1,000 lncRNAs have been proven to be involved in various biological processes, and an increasing number of studies have demonstrated that plasma lncRNAs have great potential for use in tumor diagnosis, prognosis and in the evaluation of the therapeutic effects (14,25‑29). Circulating lncRNAs are derived from apoptosis, necrotic tissue and the active secretion of cells and lysis of circulating cells. Endogenous circulating lncRNAs are bound with proteins, which may be stable at room temperature and may endure multiple cycles of freezing and thawing (30,31). According to Schlosser et al (32), the level of lncRNAs in plasma has a certain association with its level in tissues, and lncRNAs may partly be derived from tissues. In the present study, when target lncRNAs were screened, the profiles of plasma lncRNAs and tissue lncRNAs in OSCC were compared and it was identified that only some of the differentially expressed lncRNAs was the same, which also indicated that the differentially expressed lncRNAs in the plasma were derived partly from tumor tissues. The expression of lncRNAs is tissue‑specific (32‑34). Therefore, the analysis of plasma lncRNA expression levels may be used as a minimally invasive diagnostic method for diseases.

In the present study, the four target lncRNAs were significantly downregulated in the plasma of patients with OPLs and gradually increased with the malignant transformation process. The differential expression of these four lncRNAs in different stages of OSCC indicated that they had the potential to be used as diagnostic markers for OPL and OSCC staging. The single lncRNAs ENST00000412740, NR_131012, ENST00000588803 or NR_038323 may distinguish OPL...
from the healthy controls, with an AUC of 0.901, 0.924, 0.839 and 0.849, respectively, but was not effective for the determination of OSCC stage. To further prove the efficacy of the four lncRNAs for the diagnosis of OPLs and OSCC, ROC curve and logistic regression analyses were performed with optimal combinations. The results revealed that the AUCs of the combined lncRNAs were generally larger than those of single lncRNAs in distinguishing OSCC and OPLs, with a high sensitivity (93.8%) and specificity (91.0%), particularly in distinguishing ESCC from ASCC more effectively than all single lncRNAs with a high sensitivity (94.9%) and specificity (78.6%). The sensitivity of all combinations was far greater

**Table II. The results of Delong test of receiver operating characteristic curve for the combination of lncRNAs.**

<table>
<thead>
<tr>
<th>Comparison</th>
<th>D</th>
<th>Group</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combind2 vs. NR_038323</td>
<td>2.515717</td>
<td>OSCC vs. OPL</td>
<td>0.011879</td>
</tr>
<tr>
<td>Combind3 vs. NR_131012</td>
<td>4.764378</td>
<td>ASCC vs. ESCC</td>
<td>0.000002</td>
</tr>
<tr>
<td>Combind3 vs. NR_038323</td>
<td>4.631511</td>
<td>ASCC vs. ESCC</td>
<td>0.000004</td>
</tr>
<tr>
<td>Combind3 vs. ENST00000588803</td>
<td>2.027116</td>
<td>ASCC vs. ESCC</td>
<td>0.042651</td>
</tr>
<tr>
<td>Combind3 vs. ENST00000412740</td>
<td>2.677019</td>
<td>ASCC vs. ESCC</td>
<td>0.007428</td>
</tr>
</tbody>
</table>

lncRNAs, long non-coding RNAs; Combind2, combination of ENST00000588803 and NR_131012; Combind3, combination of ENST00000588803, ENST00000412740 and NR_131012; D, Delong test statistic; ASCC, advanced oral squamous cell carcinoma (TNM III/IV); ESCC, early oral squamous cell carcinoma (TNM I/II); OSCC, oral squamous cell carcinoma; OPL, oral premalignant lesions.

**Figure 3. ROC curve analysis of single lncRNAs.** ROC curve for (A) OPL vs. H, (B) OSCC vs. OPL and (C) ASCC vs. ESCC. AUC values are presented in the graphs. ROC, receiver operating characteristic; lncRNAs, long non-coding RNAs; OSCC, oral squamous cell carcinoma; OPL, oral premalignant lesion; ESCC, early-stage squamous cell carcinoma; ASCC, advanced-stage squamous cell carcinoma; AUC, area under the ROC curve.

**Figure 4. ROC curve analysis of combinations of lncRNAs.** ROC curve of (A) ENST00000588803 combined with NR_038323 for OPL vs. H, (B) ENST00000588803 combined with NR_131012 for OSCC vs. OPL and (C) ENST00000588803 combined with ENST00000412740 and NR_131012 for ASCC vs. ESCC. AUC values are presented on the graphs. ROC, receiver operating characteristic; lncRNAs, long non-coding RNAs; OSCC, oral squamous cell carcinoma; OPL, oral premalignant lesion; ESCC, early-stage squamous cell carcinoma; ASCC, advanced-stage squamous cell carcinoma; AUC, area under the curve.
than that of the most well-known available biomarker, SCCA, with a sensitivity of 38.1% (35). Therefore, they may be very promising biomarkers for the early diagnosis and staging of OSCC. However, the expression levels of four IncRNAs in the ESCC group were similar to those of the H group; therefore, the dynamic monitoring of IncRNAs needs to be combined with clinical examinations to distinguish the difference between the H group and ESCC group.

NEAT1 (NR_131012) is essential for the assembly and structural integrity of nuclear subunit paraspeckles (36) and is regulated by TP53. p53 and pRb pathway disruptions are an important step in the early stage of oral carcinogenesis (37), which may lead to the immortalization of oral epithelial cells (38). Among these, p53 is the ‘guardian’ of genome integrity, which has been found to upregulate NEAT1 expression. In oral premalignant lesions, TP53 mutation damages the p53 signaling pathway (39) and the expression of NEAT1 is down-regulated. With the malignant transformation process of cells, p53 becomes activated under the effects of replication stress to upregulate NEAT1 expression, which promotes the formation of nuclear paraspeckles and the growth of highly divided cancer cells. Furthermore, NEAT1 promotes ATR signaling in response to replication stress to inhibit replication-related DNA damage and p53 activation, thereby forming a negative feedback loop that attenuates the activation of p53 in cells with DNA damage (40). This indicates that NEAT1 is downregulated in OPL and upregulated in ESCC and is expressed in ASCC. NEAT1 is highly expressed in various types of cancer, and its expression is associated with tumor size, TNM stage and distant metastasis; it is also a risk factor for a shorter overall survival (41). However, to the best of our knowledge, there are no studies available to date on the molecular mechanisms of the other 3 IncRNAs.

However, the exact mechanisms of these IncRNAs in the occurrence and development of OSCC remain unclear and cytological experiments are required to verify their functions. In addition, the sample size of the present study was small. For cross-sectional analysis, the validation sample needs to be further expanded, and the prognosis of patients requires follow-up, in order to make a comprehensive and accurate assessment of the clinical value of IncRNAs as diagnostic markers.

In conclusion, the present study demonstrated that the expression profiles of plasma IncRNAs are altered in OSCC compared with normal controls. ENST00000412740, NR_131012, ENST00000588803 and NR_038323 were differentially expressed in different stages of OSCC and their expression became altered with the malignant progression of OSCC. This suggests that these four IncRNAs may be promising biomarkers for the early diagnosis and staging of OSCC. Furthermore, the diagnostic efficacy of the combined IncRNAs was more prominent than that of a single IncRNA.

Acknowledgements

The authors would like to thank Professor Xiaofei Tang and Professor Xinyan Zhang (Institute of Stomatology, Capital Medical University) for their suggestions and comments, Professor Zhengxue Han, Dr Yao Liu and Dr Meihua Zhang (Beijing Stomatological Hospital, Capital Medical University) for their help in sample collection, and Mr Zhigang Wang (Medical Data Processing Center, Peking Union Medical College Hospital) for data processing.

Funding

The present study was supported by a grant from the National Natural Science Foundation of China (grant no. 81372897).

Availability of data and materials

The datasets used and/or analyzed during the present study are available from the corresponding author upon reasonable request.

Authors’ contributions

HJ and XW acquired the data, performed the experiments and drafted the initial manuscript. HJ, XW and ZS designed the experiments, interpreted the data and analyzed the results. SZ and HJ revised and approved the final version of the manuscript. All authors have read and approved the final manuscript, and agreed to be accountable for all aspects of the research in ensuring that the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Ethics approval and consent to participate

The present study was approved by the Ethics Committee of the Capital Medical University Beijing Stomatological Hospital (Beijing, China; approval no. 201314) and written informed consent was provided by all participants prior to the study start.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References


