

Identification and validation of seven prognostic long non-coding RNAs in oral squamous cell carcinoma

TINGTING MIAO*, QINGZONG SI*, YUAN WEI, RUIHONG FAN, JUNJIE WANG and XIAOLI AN

School of Stomatology, Lanzhou University, Gansu, Lanzhou 730000, P.R. China

Received June 2, 2019; Accepted November 12, 2019

DOI: 10.3892/ol.2020.11603

Abstract. Oral squamous cell carcinoma (OSCC) is one of the most common malignancies worldwide, due to poor diagnosis and treatment. There is increasing evidence that demonstrates the involvement of long non-coding RNAs (lncRNAs) in carcinogenesis and cancer progression. Therefore, the aim of the present study was to explore potential lncRNA-associated features of patients with OSCC as a valuable and independent prognostic biomarker. A total of 268 lncRNA expression profiles and clinical patient information on OSCC were downloaded from The Cancer Genome Atlas database. The clinical information was exploited for prescreening, using Cox regression analysis, and differentially expressed lncRNAs (DELncRNAs) were identified using edgeR software. Using the 'caret' package, the datasets were categorized into test datasets and training datasets, respectively. Through bioinformatics, seven prognostic DELncRNAs were selected. Using the regression coefficients, a risk score based on the seven-DELncRNA signature was developed to assess the prognostic function of key DELncRNAs. According to the median risk score, patients were classified into high-risk and low-risk groups in the training and test datasets. Additionally, receiver operating characteristic (ROC) curve analysis was conducted to evaluate the sensitivity and specificity of the prognostic DELncRNAs, and the optimal cut-off point was obtained from ROC analysis. Based on the optimal cut-off point, the patients were also categorized into high-risk and low-risk groups. Notably, the optimal cut-off point was more sensitive than the median risk score, particularly in the test dataset. The Kaplan-Meier survival and log rank test analysis results indicated that the P-value, based on the optimal cut-off, was less than the median

risk cut-off. Additionally, stratified analysis results revealed that the seven-DELncRNAs signature was also independent of OSCC age. Furthermore, the findings of the present study suggested that the seven-DELncRNA signature can be used as a potential prognostic indicator and may have important clinical significance in OSCC.

Introduction

Oral squamous cell carcinoma (OSCC) is the sixth most common cancer worldwide, with high recurrence rate and metastasis, and remains the leading cause of cancer-associated mortality in the world (1,2). Despite advances in diagnosis and treatment, the 5-year survival rate has not improved significantly, and the prognosis of patients with metastatic disease has remained poor over the past decade (3,4). Currently, surgery and adjuvant therapy remain the most popular means of treatment for patients with OSCC (5,6). Hence, more exact molecular biomarkers should be identified for prognosis prediction and for the development of novel therapeutic strategies for OSCC.

Long non-coding RNAs (lncRNAs) are broadly defined as non-protein coding transcripts >200 nt in length that lack protein-coding ability (7-9). An increasing number of studies have reported that lncRNA is a critical regulator that is involved in multiple cellular biological processes and is associated with tumorigenesis, progression and metastasis (10), such as lncRNA P21-associated non-coding RNA (ncRNA) DNA damage-activated in hepatocellular carcinoma and lncRNA HOX transcript antisense intergenic RNA in 26 human tumor types (11,12). However, limited studies have indicated lncRNA signatures as independent biomarkers can predict OSCC prognosis with high efficiency (13-15).

The aim of the present study was to detect a lncRNA signature for the identification of OSCC prognostic biomarkers, by mining lncRNA expression profiles and clinical data in a large cohort of patients with OSCC in The Cancer Genome Atlas (TCGA) database. Multiple bioinformatics analyses were performed, such as Kaplan-Meier (K-M) curve survival analysis, univariate and multivariate Cox analyses and robust likelihood-based survival analysis. Seven lncRNAs (LINC01629, AC083967.1, AC067863.1, AC022092.1, AC005532.1, BX323046.1 and PRR29-AS1) were identified as potential novel independent prognostic biomarkers for the prediction of survival in patients with OSCC.

Correspondence to: Professor Xiaoli An, School of Stomatology, Lanzhou University, 199 Donggang West Road, Gansu, Lanzhou 730000, P.R. China
E-mail: anxl@lzu.edu.cn

*Contributed equally

Key words: long non-coding RNAs, oral squamous cell carcinoma, prognostic biomarkers, robust likelihood-based survival analysis, optimal cutoff

Materials and methods

Clinical information and lncRNA expression profile collection. OSCC lncRNA data and corresponding clinical information were downloaded from the TCGA data portal (<https://tcga-data.nci.nih.gov/tcga/>) using TCGAbiolinks (R version 3.6) (16). The OSCC datasets samples from the oral cavity were selected, including buccal mucosa, tongue, lip, hard palate, alveolar ridge, floor of the mouth and oral cavity. The latest genome annotation files were downloaded from GENCODE database (<https://www.gencodegenes.org/>). Moreover, patients were excluded from the present study for the following reasons: i) OSCC plus other malignancies; ii) tissue samples without complete RNA sequencing data; iii) patients receiving radiotherapy and chemotherapy prior to surgery; and iv) missing clinical information values. As a result, a total of 268 OSCC patients and 44 controls were enrolled in the study.

Differential expression analysis. Differentially expressed lncRNAs (DElncRNAs) screened from TCGA were analyzed using the 'edgeR' package (17). In order to improve the screening accuracy and simplify the screening process, the DElncRNAs were selected with the false discovery rate set at 0.05 and the lfold change value was >2-fold.

Identification and selection of prognostic DElncRNAs. The DElncRNA expression profile and clinical features were incorporated into the complete dataset and further randomly divided into training datasets and test datasets, using the 'caret' R package (version 6.0-84) (<http://caret.r-forge.r-project.org/>). The association between DElncRNAs and patients' OS was analyzed in the training dataset. Univariate Cox regression analysis was utilized to identify significant DElncRNAs with a $P < 0.05$ in the R environment using 'survival' packages (version 3.1-8; <https://cran.r-project.org/web/packages/survival/>). In order to ensure the reliability of these lncRNAs, a robust likelihood-based survival analysis was conducted using the R packages 'Rbsurv' (<http://bioconductor.org/packages/release/bioc/html/rbsurv.html>), following the procedures described previously by Mao *et al* (18). After selecting the DElncRNAs with the lowest Akaike information criterion (AICs) value, multivariate Cox regression analysis was performed on these DElncRNAs, with a $P < 0.05$ to screen for prognosis-associated DElncRNAs.

Construction of a risk score system for the key DElncRNAs. According to the results of multivariate Cox regression analysis, the regression coefficients of prognostic DElncRNAs were calculated and a risk scoring system was constructed. The risk score of each patient was determined based on the following risk formula: Risk score = $\sum \text{Coef}_{\text{DElncRNAs}} \times \text{Exp}_{\text{DElncRNAs}}$ (19). In the formula, $\text{Coef}_{\text{DElncRNAs}}$ represents the coefficient of each DElncRNA and $\text{Exp}_{\text{DElncRNAs}}$ is the expression of each DElncRNA.

Validation of the risk score formula. Based on this formula, the risk scores were calculated in the training and test datasets, and the patients were categorized into high-risk and low-risk groups using the median risk score as the cut-off point. To further

evaluate the specificity and sensitivity of the seven-lncRNA signature, receiver operating characteristic (ROC) curve analysis was performed and the ROC curve was obtained using the 'survivalROC' R package (<https://cran.r-project.org/web/packages/survivalROC>). Once the sensitivity and specificity reached a maximum, an optimal cut-off point was selected. Based on the cut-off point, the risk levels were used to classify patients into high-risk and low-risk groups in the training datasets, the test dataset and the complete dataset. The differences in survival were further assessed by K-M curve analyses and log-rank test analyses, using the 'survminer' R package (version, 0.4.6; <https://cran.r-project.org/web/packages/survminer/index.html>).

Correlation analysis and function enrichment. In order to explore the potential function for the prognostic lncRNA, correlation analysis was performed between gene expression and prognostic lncRNA expression. The correlated genes were screened based on the following criteria: $P < 0.05$ and $|\text{Pearson coefficient}| > 0.3$. The selected genes were further applied to conduct pathway enrichment analysis using the online tool 'metascape' (<http://metascape.org/gp/index.html>). The significant pathways were screened according to the criterion: q value < 0.05 .

Results

Preprocessing clinical features and screening for DElncRNAs. The RNA sequencing (RNAseq) expression profile data and detailed clinical information on patients with head and neck cell squamous carcinoma were downloaded from the TCGA database. To obtain the OSCC datasets, samples from the oral cavity were selected, including buccal mucosa, tongue, lip, hard palate, alveolar ridge, floor of the mouth and oral cavity. After excluding unclear clinical information, a final sample of 268 patients was considered (Table I). The RNAseq expression datasets consisted of coding and non-coding genes. To obtain the lncRNA datasets, the latest genome annotation files (<https://www.gencodegenes.org/>) were downloaded. Based on the annotation file, expression data of a total of 15,183 lncRNAs were extracted. In addition, 2,157 significant DElncRNAs (including 1,454 upregulated and 703 downregulated DElncRNAs) were identified between OSCC tumor and normal tissues (fold change > 2 ; $P < 0.05$; Fig. 1A and B). The expression data and clinical features of 2,157 lncRNAs were further integrated into a complete dataset and patient samples were randomly divided into training ($n=134$) and test ($n=134$) datasets.

Identification of prognostic lncRNAs associated with OSCC. In total, 2,157 DElncRNAs in the training dataset were first used for univariate Cox regression analysis, with $P < 0.05$, and 81 significant DElncRNAs were identified and further fitted to a robust likelihood-based survival analysis. As Table shown in Table II, 12 feasible and reliable DElncRNAs were selected (Table II). Furthermore, multivariate Cox regression analysis yielded seven prognosis-associated lncRNAs, including LINC01629, AC083967.1, AC067863.1, AC022092.1, AC005532.1, BX323046.1 and PRR29-AS1 (Fig. 2). To comprehensively understand the association between the seven-DElncRNA

Table I. Clinical information of patients and pre-screening of the clinical factors associated with the survival of OSCC based on the Cox regression analysis.

| TCGA OSCC set (n=268) | Cox regression analysis | |
|--------------------------|-------------------------|--------------------|
| | Hazard ratio (95% CI) | P-value |
| Age | 1.364 (1.064-1.748) | 0.014 ^a |
| Clinical_M | 1.923 (0.476-7.771) | 0.359 |
| Clinical_N | 1.128 (0.879-1.447) | 0.346 |
| Tumor_stage | 1.178 (0.888-1.563) | 0.257 |
| Sex | 0.957 (0.736-1.243) | 0.742 |

^aP<0.05. OSCC, oral squamous cell carcinoma; TCGA, The Cancer Genome Atlas; M, metastasis; N, Node.

signature and the prognosis of OSCC, a seven-DElncRNA risk scoring system was constructed, based on its Cox regression coefficient. Risk score = (0.1448) x Exp_{LINC01629} + (0.4073) x Exp_{AC083967.1} + (0.5187) x Exp_{AC067863.1} + (0.3046) x Exp_{AC022092.1} + (0.2533) x Exp_{AC005532.1} + (-0.4013) x Exp_{BX323046.1} + (0.6586) x Exp_{PRR29-AS1}. The risk score of the seven-DElncRNA signature was calculated for each sample in the training and test datasets separately. Patients were categorized into the high-risk group (N=67) and low-risk group (N=67), based on the median risk score. As shown in Fig. 3, the K-M curves revealed a significant difference between the high-risk and low-risk groups in the training dataset, which were not significantly associated with OSCC in the test dataset (P=0.092).

Evaluation and validation of the seven-DElncRNAs signature. In order to evaluate the sensitivity and specificity of the seven-DElncRNA signature, ROC analysis was performed on the training dataset. Fig. 4A shows a value of the area under the curve of 0.827. It has been demonstrated that the cut-off points with the maximal sensitivity and specificity could achieve a good classification (15). The optimal cut-off point reached 1.444, and the optimal cut-off value was used to further divide patient samples into the high-risk group (N=43) and the low-risk group (N=91) in the training dataset (Fig. 4B). K-M curves and log-rank test results suggested that a significant difference existed in survival time between the high-risk and the low-risk group (P<0.0001). In addition, K-M curves and log-rank tests were also performed in the test and complete datasets. The patients were separated into the high-risk and the low-risk group in the test dataset (P=0.0044) and in the complete dataset (P<0.0001), using the same risk formula (Fig. 4C and D). Additionally, compared to the expression level of BX323046.1, the other lncRNAs were at a high expression level in the high-risk group. Most cases of mortality were observed in the high-risk group, and patients with extended survival time were observed in the low-risk group in the complete dataset (Fig. 5). Based on the aforementioned results, age factors were found to be associated with the survival time of OSCC. Therefore, the datasets were regrouped based on the median age (62 years) to investigate the applicability of the seven-DElncRNAs signature. Patients were divided into a younger group (N=141) and an older group (N=127). Using the

Table II. Prognostic value of lncRNAs was screened by performing forward selection in the training dataset (n=134).

| lncRNA | nloglik | AIC |
|-------------|---------|----------------------|
| LINC01629 | 606.51 | 1215.01 ^a |
| AP002989.1 | 603.15 | 1210.29 ^a |
| AC073578.1 | 597.29 | 1200.58 ^a |
| AC083967.1 | 592.62 | 1193.23 ^a |
| AC067863.1 | 588.19 | 1186.39 ^a |
| AC069503.1 | 586.91 | 1185.82 ^a |
| HCG14 | 586.51 | 1187.02 ^a |
| AC022092.1 | 585.00 | 1186.00 ^a |
| AC005532.1 | 579.36 | 1176.71 ^a |
| BX323046.1 | 577.73 | 1175.45 ^a |
| PRR29-AS1 | 573.64 | 1169.29 ^a |
| AC130456.2 | 572.02 | 1168.05 ^a |
| BLACE | 571.05 | 1168.10 |
| LINC00628 | 570.71 | 1169.43 |
| SMARCA5-AS1 | 569.99 | 1169.98 |
| KLHL30-AS1 | 569.56 | 1171.12 |
| AL391001.1 | 569.55 | 1173.11 |
| AC008011.2 | 569.10 | 1174.20 |
| AL035661.1 | 569.03 | 1176.06 |
| AC007786.1 | 567.94 | 1175.88 |
| LINC02453 | 567.09 | 1176.18 |
| AL021026.1 | 567.08 | 1178.15 |
| AC026471.6 | 566.29 | 1178.59 |
| AC093510.2 | 565.64 | 1179.29 |
| BDNF-AS | 565.43 | 1180.85 |
| MIR503HG | 562.80 | 1177.61 |
| AC097634.1 | 561.28 | 1176.56 |
| AC090337.1 | 559.59 | 1175.19 |
| CBR3-AS1 | 559.16 | 1176.32 |

^aP<0.05. lncRNA, long non-coding RNA; AIC, Akaike information criterion; nloglik, negative log-likelihoods.

same formula, the younger group was further categorized into the high-risk (N=31) and the low-risk group (N=110), based on the optimal cut-off (Fig. 6A). Similarly, the older group was divided into the high-risk (N=39) and the low risk-group (N=88; Fig. 6C). The K-M curves demonstrated that the patients in the high-risk group had shorter OS (P<0.0001), and the area under the curve (AUC) values were 0.760 and 0.742 for the younger and older group, respectively, indicating that the seven-DElncRNA signature was independent to age (Fig. 6B and D). Additionally, the association between the seven-DElncRNA signature and clinical phenotypes was investigated. The results demonstrated that the seven-DElncRNA signature can serve as an independent predictor among the clinical phenotypes (Fig. S1). In order to explore the potential functions of the seven DElncRNAs, a correlation analysis between genes and the seven lncRNAs was performed. As a result, a total of 287 genes were screened from the correlation analysis by setting the following criteria: P<0.001 and absolute

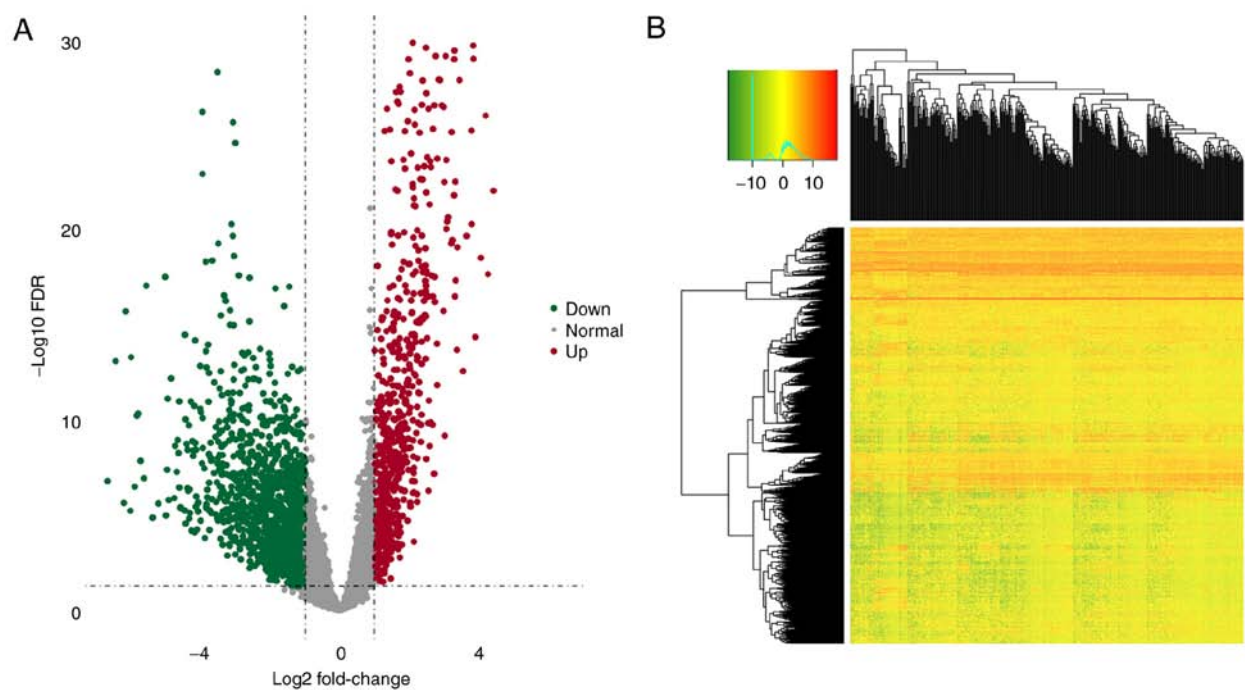


Figure 1. Landscape of DElncRNAs in oral squamous cell carcinoma samples. (A) Volcano plot for the DElncRNAs. The red points represent upregulation, the green points represent downregulation and the grey points represent statistical insignificance. (B) The normalized expression of DElncRNAs was plotted on a heatmap with the scale of -10 to 10. DElncRNA, differentially expressed long non-coding RNA; FDR, false discovery rate

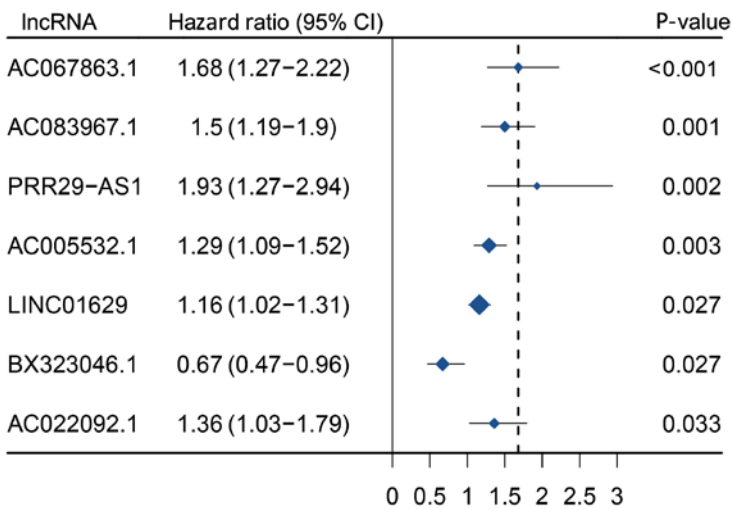


Figure 2. Prognostic value of the seven differentially expressed lncRNAs. Hazard ratios <1.0 suggests lncRNA as a favorable prognostic biomarker. lncRNA, long non-coding RNA.

value of Pearson correlation coefficient >0.3. Further pathway enrichment analysis for the corresponding genes revealed that these lncRNA may be involved in the ‘PPAR signaling pathway’ and ‘cell cycle’ (P<0.05; Fig. S2).

Discussion

OSCC is an aggressive malignancy of the head and neck, with a 5-year survival rate of 40-50% worldwide (20,21). Although great efforts have been made over the past few decades to develop signatures for prognostic predictions, the lack of specificity and sensitivity to predict survival remains due to the complex molecular and cellular heterogeneity of OSCC.

Therefore, the identification of an effective and independent molecular biomarker of OSCC is required.

The development of high-throughput technologies has promoted the discovery and study of ncRNAs, including lncRNAs, which account for only a small proportion (2%) of transcribed genes in eukaryotic species (22). A number of studies have reported that lncRNAs serve a pivotal role in complex biological processes (23-25), including tumor promotion and tumor suppression. For example, Liu *et al* (26) reported that DiGeorge syndrome critical region gene 5 is involved in cervical cancer progression by modulating the Wnt pathway. Guo *et al* (27) found that the lncRNA CEBPA-AS1 is associated with poor prognosis and promotes tumorigenesis

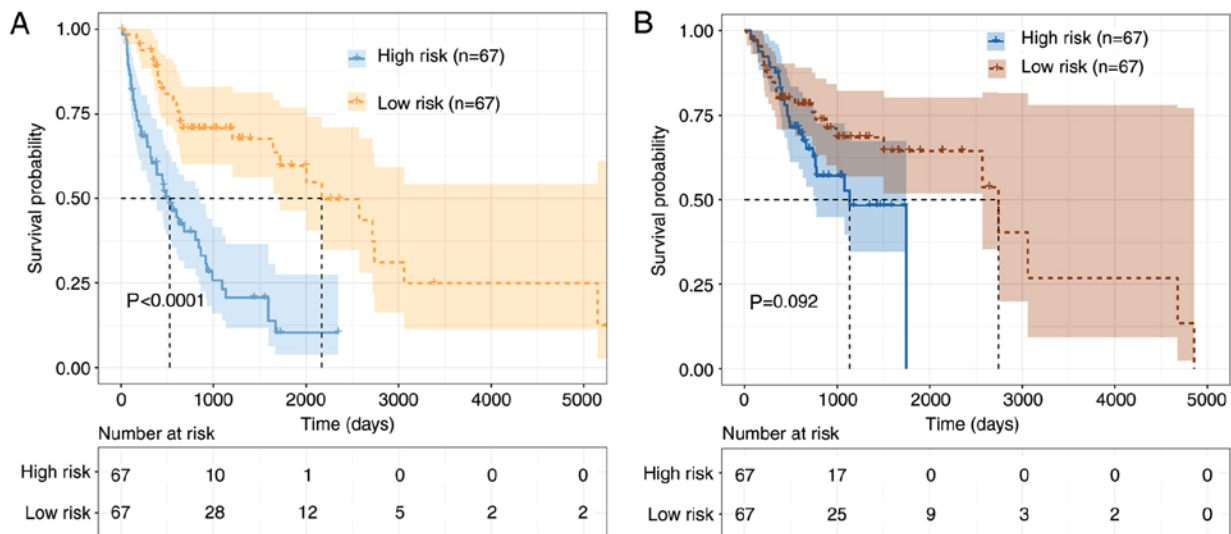


Figure 3. K-M survival curves of overall survival time in patients with oral squamous cell carcinoma, based on the median cut-off value. (A) K-M curve indicating the significance between the high-risk and low-risk groups, based on the median risk score, in the training dataset. (B) K-M curve demonstrating the significance between the high-risk and low-risk groups, based on the median risk score, in the test dataset. K-M, Kaplan-Meier.

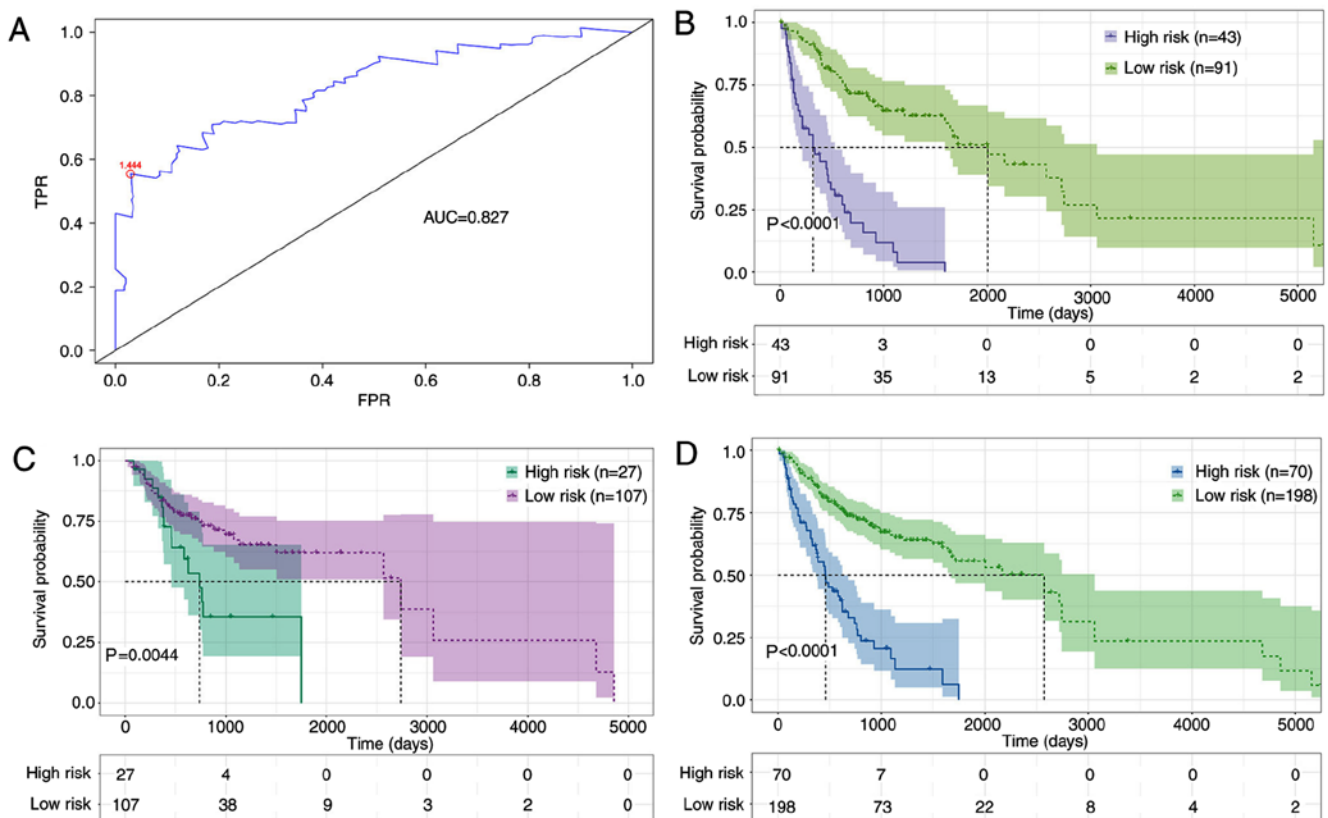


Figure 4. Evaluation of the seven-DElncRNAs signature based on the risk score and classification of different risk level groups and an optimal cut-off point. (A) Evaluation of the accuracy of seven-DElncRNA signature by conducting the receiver operating characteristic analysis, and the red point represents the optimal cut-off. (B) K-M curve showing the significance between the high-risk and the low-risk groups based on the optimal cut-off in the training dataset. (C) K-M curve demonstrating the significance between the high-risk and the low-risk groups based on the optimal cut-off in the test dataset. (D) K-M curve demonstrating the significance between the high-risk and low-risk groups, based on the optimal cut-off of the complete dataset. AUC, area under the curve; K-M, Kaplan-Meier; DElncRNA, differentially expressed long non-coding RNA; TPR, true positive rate; FPR, false positive rate.

via CEBPA/Bcl2 in OSCC, which may contribute to improving the effects of clinical treatment in OSCC. However, to the best of our knowledge, the prognostic values of lncRNAs in OSCC have not been comprehensively examined. Therefore,

it is of great significance to explore the lncRNAs associated with prognosis, which will provide a potential mechanism and help to identify effective therapeutic targets for patients with OSCC.

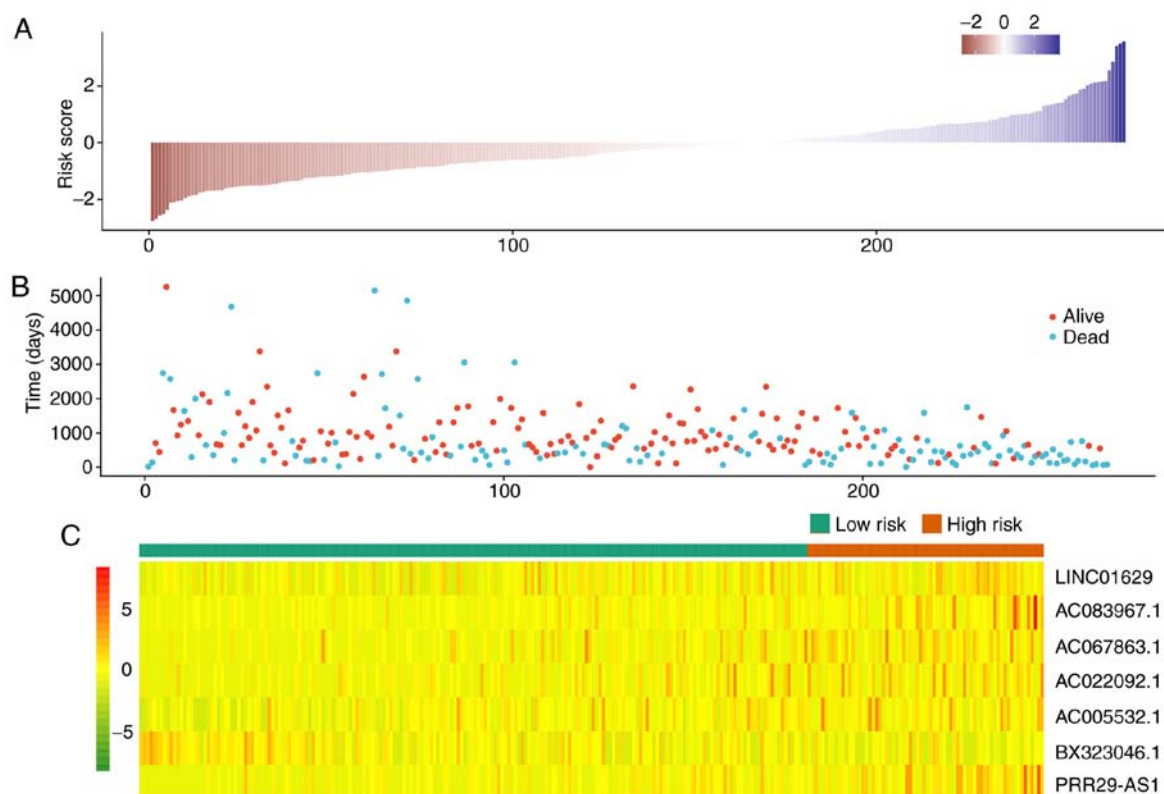


Figure 5. Risk score distribution of the seven-DElncRNAs signature and the DElncRNAs expression profile. Each panel consists of three rows: (A) Top rows showed a risk score distribution for the high risk score group and low risk score group; (B) middle rows represent the distribution and survival status of patients with OSCC; and (C) the bottom rows shows the heatmap of the expression of 7 prognostic DElncRNA.

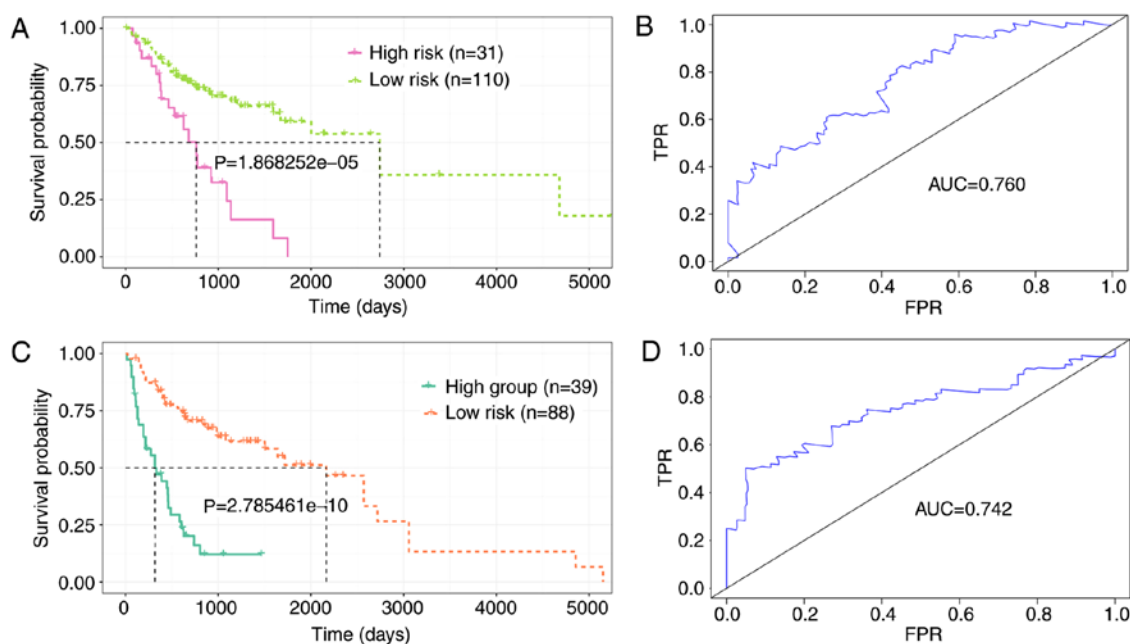


Figure 6. K-M and ROC analyses of patients with oral squamous cell carcinoma in different age cohorts (young vs. old group). (A) K-M analysis with log-rank test was conducted for the young group to evaluate the survival differences between the low-risk and the high-risk patients, based on the optimal cut-off point. (B) ROC curves were applied to estimate the sensitivity and specificity of the seven DElncRNAs signature in the younger group. (C) K-M analysis with log-rank test was conducted for the older group to evaluate the survival differences between the low-risk and high-risk patients, based on the optimal cut-off point. (D) ROC curves were applied to estimate the sensitivity and specificity of the five-DNA methylation signature in the older group. K-M, Kaplan-Meier; ROC, receiver operating characteristic; AUC, area under the curve; TPR, true positive rate; FPR, false positive rate.

The association between dysregulated lncRNAs and the prognosis of OSCC has been studied using a single biomarker

or in small-scale studies (28,29). However, compared with single clinical biomarkers, integrating multiple biomarkers

in a large clinical cohort can improve predictive accuracy (30). At present, the TCGA database has introduced a novel approach to this genomic analysis (31). TCGA is a comprehensive expression database of a variety of cancer types, which collects high-throughput methods at various genomic and proteomic levels as well as clinical information, including stage and grade of tumor, survival time, age, sex and ethnicity (32). Based on large datasets provided by the public TCGA database, a number of studies have investigated the diagnostic and prognostic value of lncRNAs in various types of cancer (33-35).

In the present study, RNAseq and relevant clinical data of the TCGA OSCC cohort were downloaded, resulting in the identification of seven DElncRNAs associated with OSCC survival in the training dataset, including LINC01629, AC083967.1, AC067863.1, AC022092.1, AC005532.1, BX323046.1 and PRR29-AS1. Based on regression coefficients from multivariate Cox regression analysis, a seven-DElncRNAs risk scoring system was built and used to calculate the risk score for each patient. According to the median risk score, patients were classified into high-risk and low-risk groups in the training and test datasets. The K-M curve analysis results demonstrated that the low-risk group had a longer survival time compared with the high-risk group. Furthermore, ROC analysis was performed to estimate the sensitivity and specificity of the seven-DElncRNAs signature. In addition, an optimal cut-off point was identified from the ROC analysis and the patients were further divided into the high-risk and the low-risk group. Additionally, the risk score formula and the optimal cut-off points were validated in the test and complete datasets. According to K-M curve and log-rank test analysis, the patients in the low-risk group had a significantly longer OS and fewer cases of mortality compared with patients in the high-risk group. Furthermore, the associations of the seven DElncRNAs were also identified, and pathway enrichment analysis for these genes revealed that they may be involved in 'PPAR signaling pathway' and 'cell cycle'.

To the best of our knowledge, the present study was the first to identify seven DElncRNAs associated with the prognosis of OSCC survival and their role in the prognosis of cancer, which will provide useful information for further studies on OSCC. Additionally, the optimal cut-off point for the risk level classification was more effective and accurate compared with the median risk cut-off point. For the median risk score, the high-risk group and the low-risk group were not significant in the test dataset ($P=0.092$), whereas a significant difference was observed based on the optimal cut-off ($P=0.0044$). To the best of our knowledge, this was the first attempt to determine an optimal cut-off in OSCC. Additionally, age is an important factor that is associated with survival in OSCC, and therefore, the seven-DElncRNA signature was employed to analyze the clinical effect in different age groups. The results indicated that the seven-DElncRNA signature was a good classification system and further categorized patients into the high-risk and low-risk groups with significant P-values, according to the median age (62 years), demonstrating that the seven-DElncRNA signature was an independent predictor for the prognosis of OSCC according to age.

However, several limitations of the present study should be taken into consideration. Firstly, only the seven-lncRNAs

signature was analyzed and validated in the TCGA datasets, and therefore it is necessary to gain more expression profiles of OSCC-associated lncRNAs for further validation. Additionally, *in vivo* and *in vitro* experiments should be conducted to verify the molecular function and mechanisms of the seven lncRNAs in OSCC. In addition, in the risk formula, the combination of differential multiples of the lncRNAs expression was not considered, since the risk formula was not only established on the lncRNA expression, and the regression coefficient for each lncRNA was also important.

In conclusion, the present study explored the aberrantly expressed lncRNAs in OSCC profiles from the large-scale TCGA database. Furthermore, a seven-DElncRNA signature was identified, which was associated with OS in OSCC and could act as a potential independent biomarker for the prediction of prognosis in patients with OSCC. Nevertheless, future studies are required to evaluate and validate the molecular mechanisms of these DElncRNAs in prospective clinical trials.

Acknowledgements

Not applicable.

Funding

No funding was received.

Availability of data and materials

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

Authors' contributions

XA designed the study. YW, RF and JW collected the clinical information, lncRNA expression data, and revised the figures and manuscript. TM and QS analyzed the data and wrote the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Chan GG, Tai BC, Liang S, Lim DT and Soo KC: Squamous cell carcinoma of the head and neck (HNSCC)-multi-modality treatment and impact on survival. *Asian J Surg* 25: 35-40, 2002.
2. Friemel J, Foraita R, Günther K, Heibeck H, Günther G, Pflueger M, Pohlabein H, Behrens T, Bullerdiek J, Nimzyk R and Ahrens W: Pretreatment oral hygiene habits and survival of head and neck squamous cell carcinoma (HNSCC) patients. *BMC Oral Health* 16: 33, 2016.

3. Koch FP, Kunkel M, Biesterfeld S and Wagner W: Diagnostic efficiency of differentiating small cancerous and precancerous lesions using mucosal brush smears of the oral cavity-a prospective and blinded study. *Clin Oral Investig* 15: 763-769, 2011.
4. Snow GB, Annyas AA, Slooten EA Van, Bartelink H and Hart AA: Prognostic factors of neck node metastasis. *Clin Otolaryngol Allied Sci* 7: 185-192, 2010.
5. McGuire S: Geneva, Switzerland: World Health Organization, International Agency for Research on cancer, WHO Press, 2015. *Adv Nutr* 7: 418-419, 2016.
6. Kokemueller H, Rana M, Rublack J, Eckardt A, Tavassol F, Schumann P, Lindhorst D, Ruecker M and Gellrich NC: The Hannover experience: Surgical treatment of tongue cancer-A clinical retrospective evaluation over a 30 years period. *Head Neck Oncol* 3: 27, 2011.
7. Stein LD: Human genome: End of the beginning. *Nature* 431: 915-916, 2004.
8. Brosnan CA and Voinnet O: The long and the short of noncoding RNAs. *Curr Opin Cell Biol* 21: 416-425, 2009.
9. Gibb EA, Brown CJ and Wan Lam WL: The functional role of long non-coding RNA in human carcinomas. *Mol Cancer* 10: 38, 2011.
10. Tano K and Akimitsu N: Long non-coding RNAs in cancer progression. *Front Genet* 3: 219, 2012.
11. Peng C, Hu W, Weng X, Tong R, Cheng S, Ding C, Xiao H, Lv Z, Xie H, Zhou L, *et al*: Over expression of long Non-coding RNA PANDA promotes hepatocellular carcinoma by inhibiting senescence Associated inflammatory factor IL8. *Sci Rep* 7: 4186, 2017.
12. Lu X, Liu Z, Ning X, Huang L and Jiang B: The long noncoding RNA HOTAIR promotes colorectal cancer progression by sponging miR-197. *Oncol Res* 26: 473-481, 2018.
13. Zhao C, Zou H, Wang J, Shen J and Liu H: A three long noncoding RNA-Based signature for oral squamous cell carcinoma prognosis prediction. *DNA Cell Biol* 37: 888-895, 2018.
14. Diao P, Song Y, Ge H, Wu Y, Li J, Zhang W, Wang Y and Cheng J: Identification of 4-lncRNA prognostic signature in head and neck squamous cell carcinoma. *J Cell Biochem* 120: 10010-10020, 2019.
15. Song Y, Pan Y and Liu J: Functional analysis of lncRNAs based on competitive endogenous RNA in tongue squamous cell carcinoma. *PeerJ* 7: e6991, 2019.
16. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, Sabedot TS, Malta TM, Pagnotta SM, Castiglioni I, *et al*: TCGAblinks: An R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res* 44: e71, 2016.
17. Robinson MD, McCarthy DJ and Smyth GK: edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139-140, 2010.
18. Mao X, Qin X, Li L, Zhou J, Zhou M, Li X, Xu Y, Yuan L, Liu QN and Xing H: A 15-long non-coding RNA signature to improve prognosis prediction of cervical squamous cell carcinoma. *Gynecol Oncol* 149: 181-187, 2018.
19. Dong J and Xu M: A 19-miRNA Support Vector Machine classifier and a 6-miRNA risk score system designed for ovarian cancer patients. *Oncol Rep* 41: 3233-3243, 2019.
20. Qiu Z, Sun W, Gao S, Zhou H, Tan W, Cao M and Huang W: A 16-gene signature predicting prognosis of patients with oral tongue squamous cell carcinoma. *PeerJ* 5: e4062, 2017.
21. Zhou Y, Kolokythas A, Schwartz JL, Epstein JB and Adami GR: microRNA from brush biopsy to characterize oral squamous cell carcinoma epithelium. *Cancer Med* 6: 67-78, 2017.
22. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, *et al*: Landscape of transcription in human cells. *Nature* 489: 101-108, 2012.
23. Li CH and Chen Y: Targeting long non-coding RNAs in cancers: Progress and prospects. *Int J Biochem Cell Biol* 45: 1895-1910, 2013.
24. Fatica A and Bozzoni I: Long non-coding RNAs: New players in cell differentiation and development. *Nat Rev Genet* 15: 7-21, 2014.
25. Rinn JL and Chang HY: Genome regulation by long noncoding RNAs. *Ann Rev Biochem* 81: 145-166, 2012.
26. Liu Y, Chang Y, Lu S and Xiang YY: Downregulation of long noncoding RNA DGCR5 contributes to the proliferation, migration, and invasion of cervical cancer by activating Wnt signaling pathway. *J Cell Physiol* 234: 11662-11669, 2019.
27. Guo Y, Ma Y, Hu X, Song R, Zhu L and Zhong M: Long non-coding RNA CEBPA-AS1 correlates with poor prognosis and promotes tumorigenesis via CEBPA/Bcl2 in oral squamous cell carcinoma. *Cancer Biol Therapy* 19: 205-213, 2017.
28. Fang Z, Zhao J, Xie W, Sun Q, Wang H and Qiao B: LncRNA UCA1 promotes proliferation and cisplatin resistance of oral squamous cell carcinoma by suppressing miR-184 expression. *Cancer Med* 6: 2897-2908, 2017.
29. Liang S, Zhang S, Wang P, Yang C, Shang C, Yang J and Wang J: LncRNA, TUG1 regulates the oral squamous cell carcinoma progression possibly via interacting with Wnt/beta-catenin signaling. *Gene* 608: 49-57, 2017.
30. Li S, Chen X, Liu X, Yu Y, Pan H, Haak R, Schmidt J, Ziebolz D and Schmalz G: Complex integrated analysis of lncRNAs-miRNAs-mRNAs in oral squamous cell carcinoma. *Oral Oncol* 73: 1-9, 2017.
31. Tomczak K, Czerwinska P and Wiznerowicz M: The cancer genome atlas (TCGA): An immeasurable source of knowledge. *Contemp Oncol (Pozn)* 19: A68-A77, 2015.
32. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C and Stuart JM: The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* 45: 1113-1120, 2013.
33. Yin H, Wang X, Zhang X, Wang Y, Zeng Y, Xiong Y, Li T, Lin R, Zhou Q, Ling H, *et al*: Integrated analysis of long noncoding RNA associated-competing endogenous RNA as prognostic biomarkers in clear cell renal carcinoma. *Cancer Sci* 109: 3336-3349, 2018.
34. Zhou M, Zhang Z, Zhao H, Bao S, Cheng L and Sun J: An immune-related six-lncrna signature to improve prognosis prediction of glioblastoma multiforme. *Mol Neurobiol* 55: 3684-3697, 2018.
35. Wang R, Du L, Yang X, Jiang X, Duan W, Yan S, Xie Y, Zhu Y, Wang Q, Wang L, *et al*: Identification of long noncoding RNAs as potential novel diagnosis and prognosis biomarkers in colorectal cancer. *J Cancer Res Clin Oncol* 142: 2291-2301, 2018.