

Research Article

# A three-lncRNA expression signature predicts survival in head and neck squamous cell carcinoma (HNSCC)

Peng Wang, Meng Jin, Chuan-hui Sun, Like Yang, Yu-shan Li, Xin Wang,  Ya-nan Sun, Lin-li Tian and Ming Liu

Department of Otorhinolaryngology, Head and Neck Surgery, the Second Affiliated Hospital of Harbin Medical University, Harbin, 150081, P.R. China

Correspondence: Ya-nan Sun (syn2767@126.com)



Increasing evidence has shown that long non-coding RNAs (lncRNAs) have important biological functions and can be used as a prognostic biomarker in human cancers. However, investigation of the prognostic value of lncRNAs in head and neck squamous cell carcinoma (HNSCC) is in infancy. In the present study, we analyzed the lncRNA expression data in a large number of HNSCC patients ( $n=425$ ) derived from The Cancer Genome Atlas (TCGA) to identify an lncRNA expression signature for improving the prognosis of HNSCC. Three lncRNAs are identified to be significantly associated with survival in the training dataset using Cox regression analysis. Three lncRNAs were integrated to construct an lncRNA expression signature that could stratify patients of training dataset into the high-risk group and low-risk group with significantly different survival time (median survival 1.85 years vs. 5.48 years;  $P=0.0018$ , log-rank test). The prognostic value of this three-lncRNA signature was confirmed in the testing and entire datasets, respectively. Further analysis revealed that the prognostic power of three-lncRNA signature was independent of clinical features by multivariate Cox regression and stratified analysis. These three lncRNAs were significantly associated with known genetic and epigenetic events by means of functional enrichment analysis. Therefore, our results indicated that the three-lncRNA expression signature can predict HNSCC patients' survival.

## Introduction

It is widely considered that head and neck squamous cell carcinoma (HNSCC) is one of the most prevalent and fatal cancers. Despite more efforts have been taken in the prevention and treatment of HNSCC, people who get HNSCC is still increasing in recent years and the outcome remains unsatisfactory with 5-year survival rates are less than 50% [1,2]. Smoking, alcohol and human papillomavirus (HPV) infections have been associated with HNSCC dramatically [3,4]. Due to the molecular heterogeneity and diverse etiology of head and neck tumors, it is of significance to identify novel molecular biomarkers to improve the outcome of HNSCC patients.

RNAs can be separated into coding RNAs and non-coding RNAs (ncRNAs) based on the protein-coding capacity. Long non-coding RNAs (lncRNAs) (>200 nt in length) is an important section of ncRNAs [5]. lncRNAs were thought to be transcriptional noise initially. Recent studies revealed that the lncRNAs can also serve as transcriptional, post-transcriptional and epigenetic levels [6–9]. Lots of evidence illustrated that lncRNAs have complex and wide functions in the development and progression of cancer [10–12]. For instance, *H19* is an estrogen-inducible lncRNA and it plays a key role in estrogen-induced cell proliferation in breast cancer cells [13]. The experimental study demonstrated that

Received: 03 September 2018  
Revised: 08 October 2018  
Accepted: 19 October 2018

Accepted Manuscript Online:  
24 October 2018  
Version of Record published:  
21 November 2018

*H19* served as a biomarker for breast cancer diagnosis and progression. LncRNA *MEG3* as a tumor suppressor is down-regulated in cervical cancer, and *MEG3* affects cell proliferation and apoptosis by regulating *miR-21* [14]. Several expression-based lncRNA signatures have been established in a variety of tumors [15–23]. For HNSCC, biological roles of microRNA in regulating the development of head and neck cancers have been summarized in recent some reviews [24–26]. Recently, some lncRNAs have been reported to be differentially expressed in HNSCC. For example, Cao et al. [27] identified a prognostic lncRNA signature by the orthogonal partial least square discrimination analysis. However, they did not verify this prognostic lncRNA signature model in the testing dataset. Zhang et al. [28] predicted the survival of HNSCC by seven lncRNA–mRNA based risk score. Because they analyzed the dataset based on two platforms (NGS and Affymetrix HG U133 plus 2), a relatively small number of lncRNAs were analyzed in their study. Therefore, the prognostic value of lncRNAs still needs to be investigated in HNSCC.

In the present study, we conducted a comprehensive study of lncRNA expression profiles across 425 HNSCC patients with clinical information to investigate the prognostic value of lncRNAs in HNSCC. Finally, we identified three lncRNAs associated with survival and constructed lncRNA expression signature based on the expression profiles of these three lncRNAs in the training dataset that was further confirmed in the testing and entire datasets.

## Materials and methods

### HNSCC datasets and patient information

The lncRNA expression profiles of HNSCC patients were downloaded from TANRIC (The Atlas of ncRNA in Cancer, <http://bioinformatics.mdanderson.org/>) [29]. Clinical information and features of HNSCC patients were obtained from The Cancer Genome Atlas (TCGA, <https://cancergenome.nih.gov/>). After removing patients without available survival information, a total of 425 HNSCC patients were used for further analysis. We annotated each of the samples according to patient barcode ID based on the available clinical information, including status, age, gender, history of neoadjuvant treatment, pathologic stage, alcohol history documented, margin status, primary lymph node presentation assessment (pN), lymph-vascular invasion present and tumor grade. Table 1 included more detailed clinical information of all 425 NSCC patients in the present study.

### Identification of potential prognostic lncRNA biomarkers associated with overall survival in patients with HNSCC

The 425 HNSCC patients were randomly divided into the training dataset composing of 213 patients and the testing dataset composing of 212 patients. A univariable Cox regression analysis was performed to evaluate the relationship between the continuous expression level of each lncRNA and patients' overall survival in the training dataset. The raw *P* values were adjusted by Benjamini and Hochberg multiple comparison methods to control the false discovery rate (FDR). If the FDR was less than 0.1, the corresponding lncRNAs were statistically significant and were considered as candidate prognostic lncRNAs. Subsequently, a multivariate Cox regression was performed to calculate the contribution of each lncRNA in survival prediction. The lncRNA-based risk score model was defined as the linear combination of the expression levels of the significant lncRNAs and the multivariable Cox regression coefficient as the weight. According to the median risk score in the training dataset, the patients with HNSCC in each dataset were classified into the high-risk group and low-risk group.

### Statistical analysis

The Kaplan–Meier method was used to estimate OS time for the two groups, and the statistical significance was obtained using the two-sided log-rank test [30]. Univariate and multivariate Cox proportional hazards regression analyses were carried out with OS as the dependent variable and other individual clinical features as explanatory variables in each dataset. Hazard ratios (HR) and 95% confidence intervals (CI) were obtained [31]. The time-dependent receiver operating characteristic (ROC) curve was performed to evaluate the prognostic performance for survival prediction of the lncRNA risk score and calculate the area under the ROC curves (AUC) value [32]. All analysis was performed under the environment of the R/Bio-Conductor (version 3.4.0).

### Functional enrichment analysis

In order to investigate potential biological roles of lncRNA, the co-expressed relationship between the prognostic lncRNAs and mRNA was calculated by Pearson correlation coefficients. DAVID Bioinformatics Tool (version 6.8) was used to further functional enrichment analysis. Only GO categories of 'Biological Process' were considered. Functional annotation with *P*-value of <0.05 was considered be statistically significant.

**Table 1 Clinical features of all 425 HNSCC patients included in the present study**

Features	Training dataset (n=213)	Testing dataset (n=212)	Entire dataset (n=425)
<b>Status</b>			
Alive	82(38.5)	101(47.6)	183(43.1)
Dead	131(61.5)	111(52.4)	242(56.9)
<b>Age</b>			
Young (<60)	137(64.3)	142(67.0)	279(65.6)
Old (≥60)	76(35.7)	70(33.0)	146(34.4)
<b>Gender</b>			
Female	46(21.6)	69(32.5)	115(27.0)
Male	167(78.4)	143(67.5)	310(73.0)
<b>History of neoadjuvant treatment</b>			
Neoadjuvant treatment	5(2.3)	4(1.9)	9(2.1)
Without neoadjuvant	208(97.7)	208(98.1)	416(97.9)
<b>Pathologic stage</b>			
I/II	43(20.2)	43(20.3)	86(20.2)
III/IV	142(66.7)	131(61.8)	273(64.2)
Unknown	28(13.1)	38(17.9)	66(15.6)
<b>Alcohol history documented</b>			
Alcohol	148(69.5)	141(66.5)	289(68.0)
Without alcohol	60(28.2)	69(32.5)	129(30.4)
Unknown	5(2.3)	2(1.0)	7(1.6)
<b>Margin status</b>			
Ms positive	20(9.40)	22(10.4)	42(9.9)
Ms negative	149(70.0)	142(67.0)	291(68.5)
Unknown	44(20.6)	48(22.6)	92(21.6)
<b>pN</b>			
pN positive	178(83.6)	161(76.0)	339(79.8)
pN negative	18(8.5)	27(12.7)	45(10.6)
Unknown	17(7.9)	24(11.3)	41(9.6)
<b>Lymphovascular invasion present</b>			
Lymphovascular_invasion	47(22.1)	43(20.3)	90(21.2)
Without lymphovascular invasion	98(46.0)	87(41.0)	185(43.5)
Unknown	68(31.9)	82(38.7)	150(35.3)
<b>Tumor grade</b>			
G1/G2	150(70.4)	151(71.2)	301(70.8)
G3/G4	57(26.8)	54(25.5)	111(26.1)
Unknown	6(2.8)	7(3.3)	13(3.1)

## Results

### Identification of prognostic lncRNAs from training dataset

A total of 425 TCGA HNSCC patients were divided randomly into the training dataset ( $n=213$ ) and the testing dataset ( $n=212$ ). In order to evaluate whether the lncRNA expression was associated with patients survival, the training dataset was analyzed using univariate Cox proportional hazards regression analysis method. The experimental results were shown in Table 2. A total of three lncRNAs as prognostic lncRNAs were found to be significantly associated with patients' overall survival (FDR adjusted  $P<0.1$ ). These three lncRNAs were entered into the candidate pool for further selection. All of three lncRNAs (*AC002066.1*, *AC013652.1* and *AC016629.3*) reveal positive coefficient in univariate Cox proportional hazards regression analysis indicating that a higher level expression of these three lncRNAs was associated with shorter survival. Result demonstrated that these three lncRNAs were able to independently predict patients' OS at a statistically significant level of 0.01.

**Table 2 Three lncRNAs significantly associated with overall survival in HNSCC patients training set (n=213)**

Ensembl ID	Gene symbol	Chromosomal position (GRCh38)	P value <sup>1</sup>	Hazard ratio <sup>1</sup>	Coefficient <sup>1</sup>	$\beta^2$
ENSG00000237813	AC002066.1	Chromosome 7: 116,238,260-116,499,465 (-)	8.43E-05	2.207	0.792	1.504
ENSG00000259345	AC013652.1	Chromosome 15: 38,865,322-39,427,195 (-)	4.89E-05	3.748	1.321	1.337
ENSG00000269600	AC016629.3	Chromosome 19: 58,593,896-58,599,355 (-)	1.5E-04	39.778	3.683	3.714

<sup>1</sup>Obtained from the univariable Cox's proportional-hazards regression analysis.

<sup>2</sup>Obtained from the multivariate Cox's proportional-hazards regression analysis.

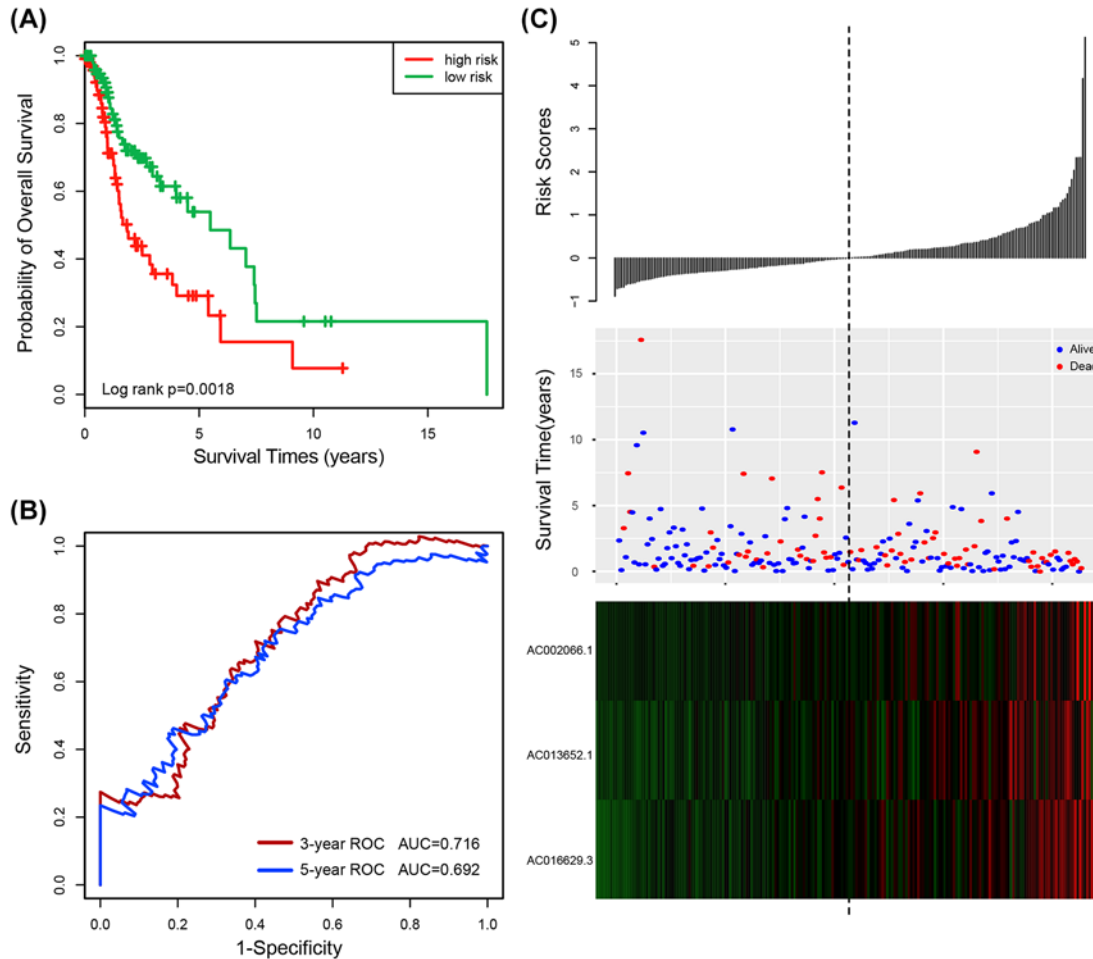
## The three-lncRNA signature predicts patients' survival in the training dataset

In order to obtain the relative contribution of three prognostic lncRNAs to survival prediction of HNSCC, expression data of three lncRNAs were fitted by the multivariate Cox regression model in the training dataset. The contribution values ( $\beta$ ) are shown in Table 2. Subsequently, a risk score model was constructed based on the expression levels of these three lncRNAs to predict the survival of HNSCC. The risk score model as follows: Risk score = (1.504  $\times$  expression value of *AC002066.1*) + (1.337  $\times$  expression value of *AC013652.1*) + (3.714  $\times$  expression value of *AC016629.3*). According to the risk score, the three-lncRNA signature risk score was computed for each patient in the training dataset. Using the median risk score as the cutoff point (cutoff = 1.010079), all patients (n=213) were classified into a high-risk group (n=107) and a low-risk group (n=106). The Kaplan–Meier curves illustrated that patients in the high-risk group tend to have shorter survival than those in the low-risk group (median survival 1.85 years vs. 5.48 years;  $P=0.0018$ , log-rank test) (Figure 1A). To determine the performance of the risk score model, the time-dependent ROC curve was analyzed. The AUC for the three-lncRNA signature risk score model at 3 and 5 years for overall survival (OS) were 0.716 and 0.692, respectively (Figure 1B). The results demonstrated the three-lncRNA signature risk score model had better performance in the training dataset. Meanwhile, the risk score was evaluated by the univariable Cox regression model analysis. The experimental results depicted that the risk score was significantly associated with patients' OS when the risk score was analyzed as a continuous variable (Hazard ratios (HR) = 2.718, 95% confidence intervals (CI) 2.016–3.666,  $P = 5.54E-11$ ) (Table 3). The distribution of the three-lncRNA signature risk score, the survival status and expression pattern in the training dataset were shown in Figure 1C. The expression of these lncRNAs tended to be up-regulated in patients with high-risk score (Figure 1C).

## Validation of the three-lncRNA signature risk score model for survival prediction in the testing and entire datasets

In order to further validate the prognostic power of the three-lncRNA signature risk score model for patients' OS prediction, the three-lncRNA signature risk score model was analyzed in the testing dataset (n=212). Using the median risk score in training dataset as the cutoff value, the patients of the testing dataset were classified into low-risk group (n=110) and high-risk group (n=102). Consistent with our findings in the training dataset, patients in the high-risk group exhibited poorer OS than those in the low-risk group (median survival 3.11 years vs. 13.04 years;  $P=0.00712$ , log-rank test) (Figure 2A). The AUC for the three-lncRNA signature risk score model at 3 and 5 years for OS were 0.6 and 0.67, respectively (Figure 2B). Figure 2C showed the distribution of the risk score, the survival status and the three-lncRNA expression in the testing dataset. Patients in the high-risk group tended to express risky lncRNAs at a higher level than those in the low-risk group. In the testing dataset, the significant association between risk score and OS has also been observed when the risk score was as a continuous variable in the univariate Cox regression analysis (HR = 1.457, 95% CI 1.093–1.942,  $P=0.010$ ) (Table 3).

The three-lncRNA signature risk score model was then tested in the entire dataset of 425 HNSCC patients. Kaplan–Meier curves showed that patients with the high-risk scores (n=209) had significantly shorter survival than those with the low-risk scores (n=216) (median survival 2.36 years vs. 7.04 years;  $P = 5.63E-05$ , log-rank test) (Figure 3A). Similar results were obtained by the univariable Cox regression model analysis of the three-lncRNA signature (HR = 1.770, 95% CI 1.481–2.115,  $P = 3.49E-10$ ) (Table 3). The AUC at 3 and 5 years for overall survival was 0.654



**Figure 1. The three-lncRNA signature in the prognosis of overall survival of HNSCC patients in the training dataset**  
(A) Kaplan–Meier survival estimates overall survival of HNSCC patients according to the three-lncRNA signature in the training dataset. (B) ROC analysis for overall survival prediction by the three-lncRNA signature within 3 and 5 years as the defining point in the training dataset. (C) The risk score distribution, patients' survival status and heatmap of the three-lncRNA expression profiles in the training dataset.

and 0.669, respectively (Figure 3B). The distribution of risk score, the survival status and lncRNA expression of HNSCC in the entire datasets also yielded similar results (Figure 3C).

## Independence of the three-lncRNA signature risk score model for survival prediction from clinical features

To determine whether the predictive ability of the three-lncRNA signature risk score model was independent of other clinical features of HNSCC patients, the multivariable Cox regression analysis was performed using lncRNA risk score and other clinical features (age, gender, pathologic stage, alcohol history documented and primary lymph node presentation assessment) in each dataset. Results demonstrated that the three-lncRNA signature was significantly associated with the survival of HNSCC patients in the training dataset (HR = 2.528, 95% CI 1.735–3.683,  $P = 1.36E-06$ ), the testing dataset (HR = 2.224, 95% CI 1.194–4.139,  $P = 0.012$ ) and the entire dataset (HR = 2.096, 95% CI 1.574–2.790,  $P = 4.02E-07$ ) (Table 3). Because the primary lymph node presentation assessment (pN) and pathologic stage were significant in the multivariate Cox analysis, stratification analysis was performed based on pN and pathologic stage subsequently. All patients were stratified into two subgroups with positive or negative pN. Patients with positive pN were divided into a high-risk and low-risk group based on their risk score. The Kaplan–Meier analyses demonstrated that high-risk group ( $n=171$ ) has shorter survival than those in the low-risk group ( $n=168$ ) (median survival 2.51 years vs. 7.41 years;  $P = 2.76E-05$ , log-rank test) (Figure 4A). For the patient with negative pN, low-risk group ( $n=21$ )



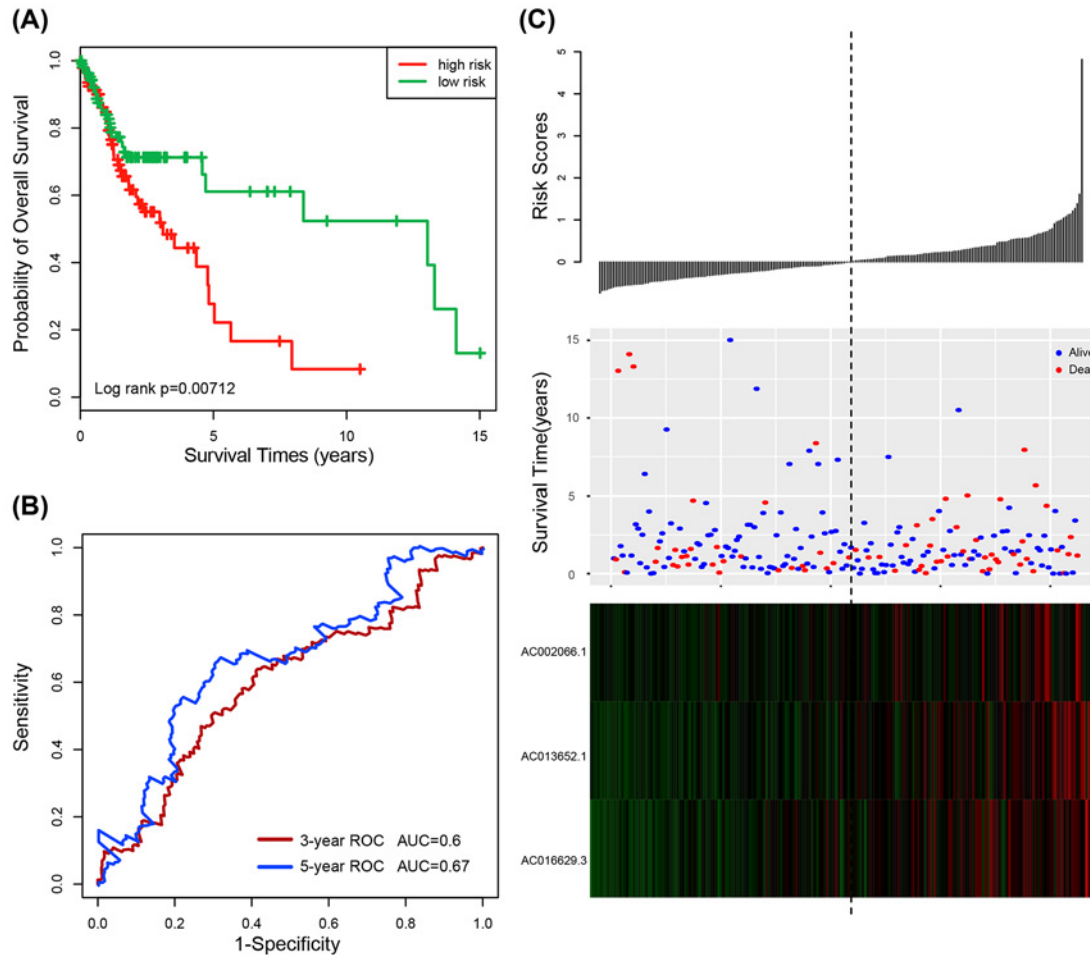
**Table 3 Univariable and multivariable Cox regression analysis of the three-lncRNA signature and overall survival in each dataset**

	Univariate analysis			Multivariate analysis		
	HR	95% CI of HR	P-Value	HR	95% CI of HR	P-value
<b>Training dataset (n=213)</b>						
Train score	2.718	2.016–3.666	5.54E-11	2.528	1.735–3.683	1.36E-06
Age (old/young)	1.035	0.648–1.654	0.884	0.860	0.499–1.482	0.588
Gender (male/female)	0.835	0.489–1.424	0.508	0.892	0.456–1.744	0.739
Pathologic stage (Stage III.IV/I.II)	1.558	0.864–2.808	0.140	1.536	0.751–3.145	0.240
Alcohol history documented (Yes/No)	1.027	0.629–1.676	0.915	0.992	0.548–1.796	0.979
pN (positive/negative)	0.735	0.334–1.615	0.443	0.709	0.238–2.114	0.537
<b>Testing dataset (n=212)</b>						
Test score	1.457	1.093–1.942	0.010	2.224	1.194–4.139	0.012
Age (old/young)	1.540	0.947–2.505	0.082	1.210	0.655–2.236	0.543
Gender (male/female)	0.783	0.480–1.277	0.327	1.178	0.592–2.346	0.640
Pathologic stage (Stage III.IV/I.II)	1.476	0.773–2.817	0.238	2.169	0.870–5.407	0.097
Alcohol history documented (Yes/No)	0.795	0.489–1.293	0.355	0.806	0.404–1.609	0.541
pN (positive/negative)	0.542	0.292–1.007	0.053	0.215	0.089–0.516	0.001
<b>Entire dataset (n=425)</b>						
Entire score	1.770	1.481–2.115	3.49E-10	2.096	1.574–2.790	4.02E-07
Age (old/young)	1.286	0.918–1.802	0.144	1.073	0.722–1.595	0.727
Gender (male/female)	0.837	0.588–1.192	0.325	1.124	0.704–1.796	0.624
Pathologic stage (Stage III.IV/I.II)	1.494	0.969–2.305	0.069	1.787	1.022–3.122	0.0418
Alcohol history documented (Yes/No)	0.893	0.633–1.258	0.516	0.842	0.542–1.310	0.446
pN (positive/negative)	0.640	0.396–1.034	0.068	0.356	0.182–0.695	0.002

also has longer survival than the high-risk group ( $n=24$ ) (median survival 4.0 years vs. 1.82 years;  $P=0.189$ , log-rank test) (Figure 4B). Here, the  $P$  that is slightly above the 0.05 significance level may be caused by too little sample size. Another clinical feature was pathologic stage that also stratified all patients into two subgroups, I/II stage patient group and III/IV stage patient group. Using the three-lncRNA signature, patients of each group could be classified into either the high-risk or low-risk group. Figure 4C showed that patients with high-risk score ( $n=145$ ) have shorter survival than those in the low-risk group ( $n=128$ ) in the III/IV stage patient group (median survival 2.36 years vs. 4.71 years;  $P=0.0498$ , log-rank test). The same results were obtained in the I/II stage patient group. The OS was significantly different in the high-risk group ( $n=34$ ) and low-risk group ( $n=52$ ) (mean survival 3.83 vs. 8.38 years;  $P=0.0995$ , log-rank test) (Figure 4D). In summary, the results of multivariate Cox regression analyses and stratification analysis illustrated that the three-lncRNA signature was independent of other clinical factors for the survival prediction of HNSCC patients.

### Functional analysis of the three prognostic lncRNAs

To further investigate the potential biological functions of the three lncRNAs, the co-expression relationship between the expression level of the three-lncRNA and mRNA was measured. The top 5% mRNA was selected as co-expressed mRNAs with prognostic lncRNA biomarkers. The expression of 294 mRNA was correlated with the three prognostic lncRNAs. GO and KEGG pathway function enrichment analysis was employed for these co-expressed mRNA. The result revealed that 294 mRNA were significantly enriched in 11 GO terms ( $P<0.05$ ) (Figure 5A) and 10 KEGG pathways ( $P<0.05$ ) (Figure 5B). Some related GO terms were observed including angiogenesis, positive regulation of MAPK cascade, extracellular matrix organization, cell adhesion, extracellular matrix disassembly, negative regulation of anoikis, positive regulation of cell division, inositol phosphate metabolic process, positive regulation of cell migration, epidermis development, positive regulation of insulin secretion involved in cellular response to glucose stimulus.

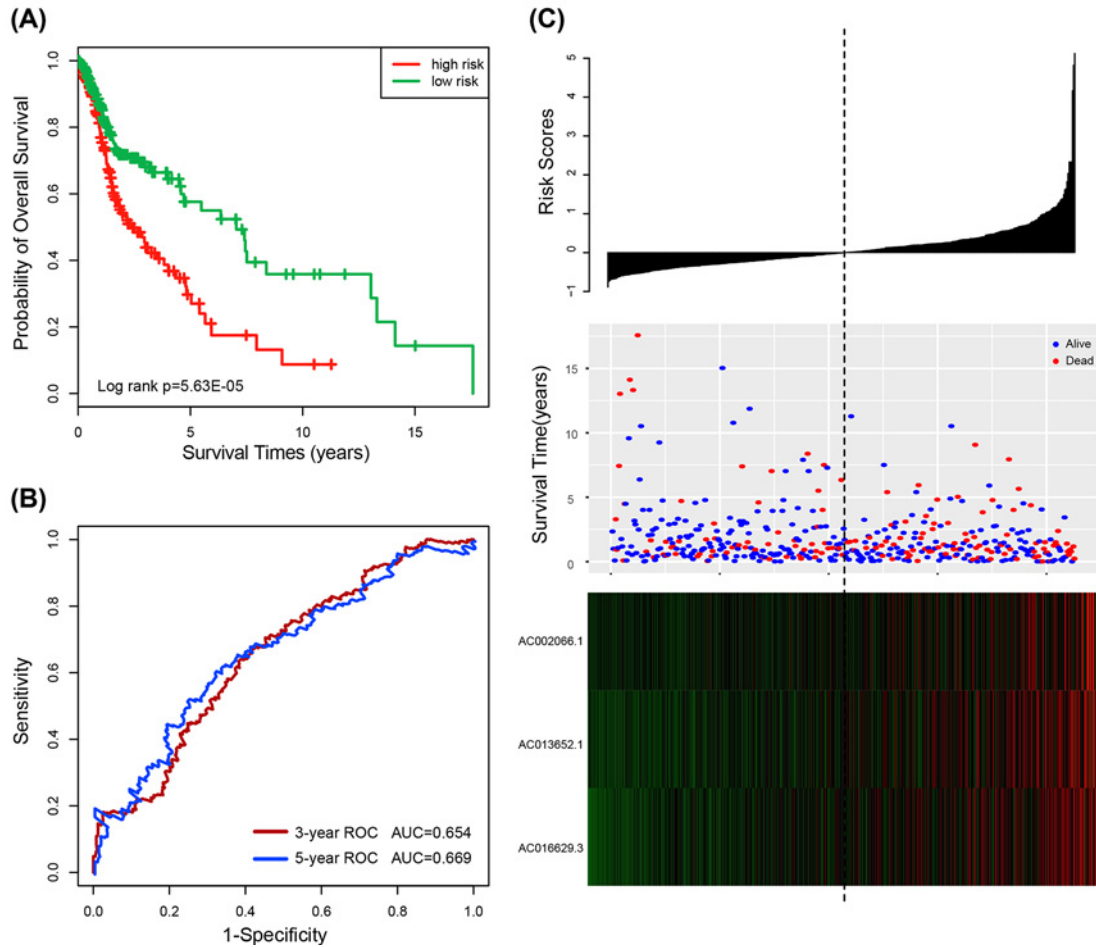


**Figure 2. The three-lncRNA signature in the prognosis of overall survival of HNSCC patients in the testing dataset**  
(A) Kaplan–Meier survival estimates overall survival of HNSCC patients according to the three-lncRNA signature in the testing dataset. (B) ROC analysis for overall survival prediction by the three-lncRNA signature within 3 and 5 years as the defining point in the testing dataset. (C) The risk score distribution, patients' survival status and heatmap of the three-lncRNA expression profiles in the testing dataset.

## Discussion

HNSCC is one of the most commonly occurring cancers [33]. Despite considerable advances in clinical research and new therapies, the overall survival rate of HNSCC remains low. Even if the patient has the same clinical and pathological stages, their prognosis is also different. Studies have found that molecular markers can effectively distinguish between these conditions. Therefore, it is an urgent need to identify novel biomarkers to predict the patient's prognosis. Previous some studies have illustrated that mRNA and miRNA could act as molecular markers to predict the outcome of HNSCC [34–37]. Recent studies have found that lncRNAs can also act as molecular markers in HNSCC. Through comparing 60 pairs of HNSCC tissues/non-tumor tissues samples and 7 cohorts of HNSCC cell lines, lncRNA HNSCC glycolysis-associated 1 was up-regulated in tumor tissues [38]. The results indicated that the differential expression of lncRNA was associated with the occurrence and development of HNSCC. Therefore, lncRNA can be used as a molecular marker to open a new vision for further investigation. Recently, researchers have shown that lncRNA played an important role in different cancers by analyzing expression profiles, and several lncRNA signatures in HNSCC have been identified [27,39,40]. However, the prognostic power of expression-based lncRNA signature for predicting survival in HNSCC patients still needs further investigation.

In the present study, lncRNA expression profiles of 425 HNSCC patients were obtained from TCGA except for the patient whose survival time data were missing. We identified the three-lncRNA signature by Cox regression analysis. Moreover, the regression coefficients using multivariable Cox regression analysis was obtained. Based on the risk

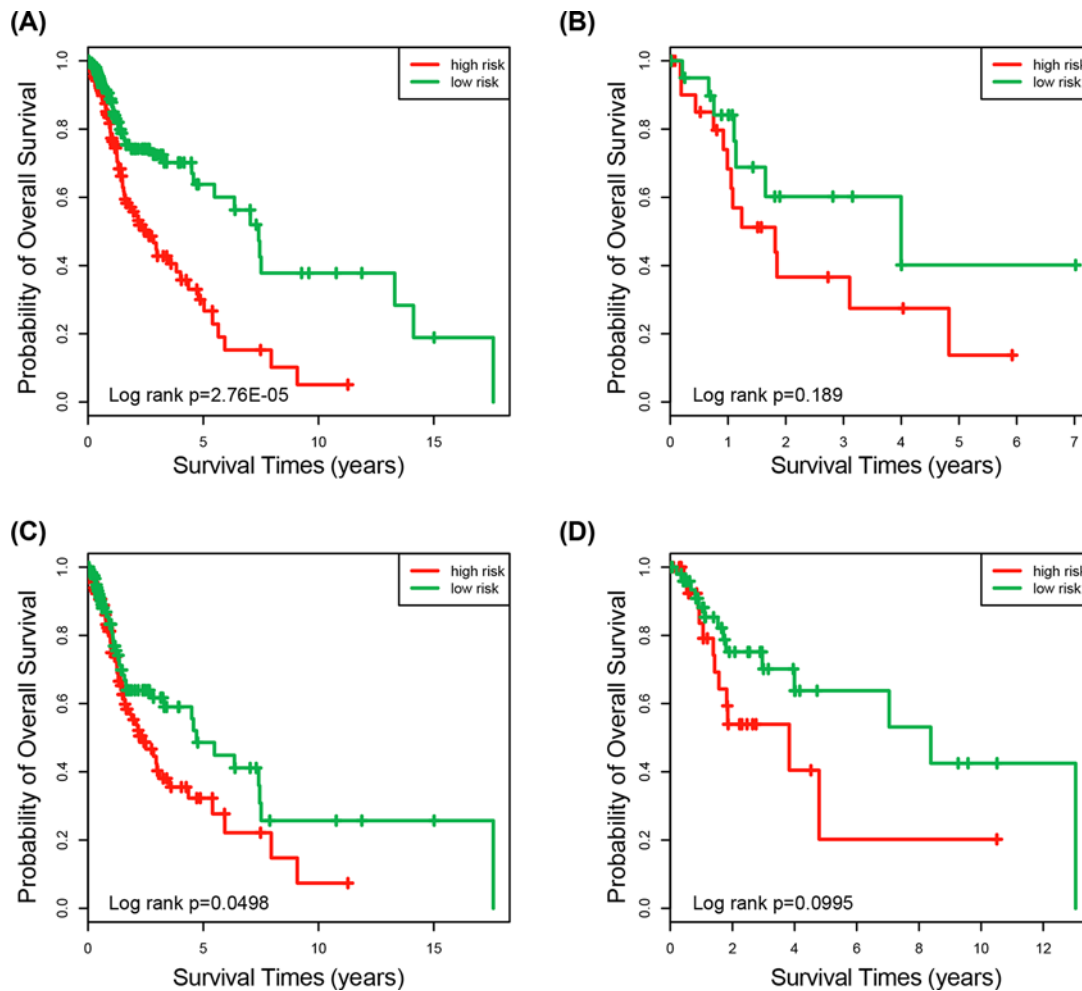


**Figure 3. The three-lncRNA signature in the prognosis of overall survival of HNSCC patients in the entire dataset**

(A) Kaplan–Meier survival estimates overall survival of HNSCC patients according to the three-lncRNA signature in the entire dataset. (B) ROC analysis for overall survival prediction by the three-lncRNA signature within 3 and 5 years as the defining point in the entire dataset. (C) The risk score distribution, patients’ survival status and heatmap of the three-lncRNA expression profiles in the entire dataset.

model of the three-lncRNA signature in training dataset, the patients were classified into two groups. Patients in the high-risk group tended to have lower OS than patients in the low-risk group. Subsequently, risk score was further confirmed in the testing dataset and entire dataset. The results showed that the prognostic value of the three-lncRNA signature risk model is robust and reliable for survival prediction in HNSCC. Afterward, multivariate Cox regression analysis was performed to determine whether the predictive ability of the three-lncRNA signature was independent of other clinical features of HNSCC patients. The estimations of HR for OS were 2.528, 2.224 and 2.096 in the training dataset, the testing dataset and the entire dataset, respectively. Because pN and pathologic stage were significantly correlated with patients’ OS, the independence of the three-lncRNA signature risk model for survival prediction needs further evaluation. Stratification analysis for pN and pathologic stage was carried out. The results of stratification analysis revealed that two subgroups’ OS were significantly different in two clinical prognostic variables, individually. Studies have confirmed that *AC002066.1* has been proved to up-regulate in H2 vs. LM3 cell lines. H2 is high potential for metastasis to lymph nodes (HCCLYM-H2), and LM3 is high potential for metastasis to the lung [41]. The experimental results illustrated lncRNA expression profiles related to organ-specific metastasis in hepatocellular carcinoma. Bryzghalov et al. found that *AC016629.3* showed high expression in K562 derived from a female patient with chronic myelogenous leukemia than other non-cancer cell lines. *AC016629.3* masked miRNA target sites in seven splice forms of RPL23A [42]. In conclusion, the three-lncRNA signature risk model was an independent prognostic factor for survival prediction in HNSCC.





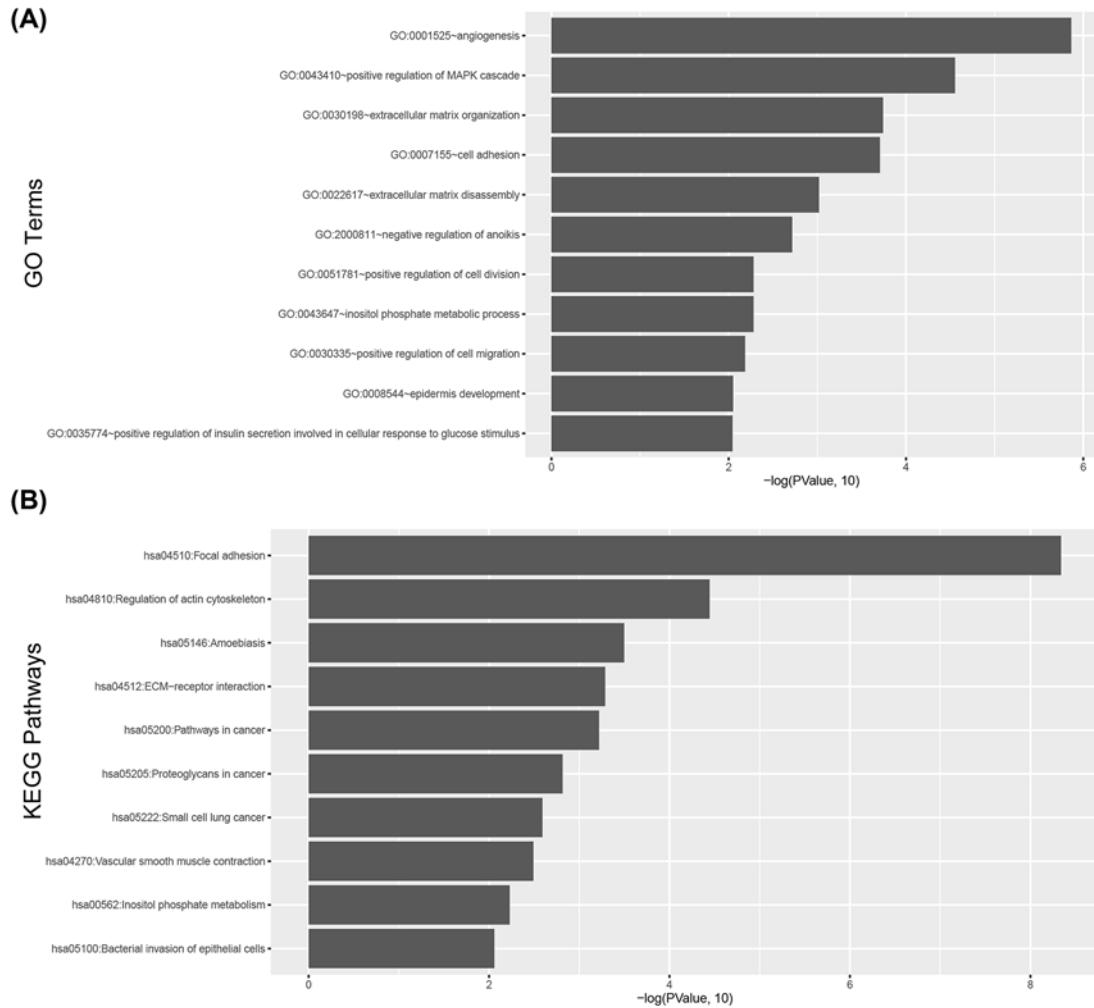
**Figure 4. Survival analysis of all patients with pN and pathologic stage information**

(A) Kaplan–Meier curves for patients with positive of pN ( $n=339$ ). (B) Kaplan–Meier curves for patients with negative pN ( $n=45$ ). (C) Kaplan–Meier curves for patients with stage III/IV ( $n=273$ ). (D) Kaplan–Meier curves for patients with stage I/II ( $n=86$ ).

Thousands of lncRNAs have been discovered in humans during the past decades with the development of ncRNA prediction algorithms and software [43–46], many functions of lncRNAs are still unknown. Therefore, functional enrichment analysis was used to predict their function. The analysis revealed that the three lncRNAs mainly involved in 11 GO terms and 10 KEGG pathways. Some of the GO terms were approved to be related to cancers. For example, experiments have confirmed that angiogenesis was involved in the occurrence and development of HNSCC [47], and cell adhesion was also confirmed to be related to HNSCC [48], Nohata et al. [49] found that *CAV1* mRNA mediated tumor cell migration and invasion in HNSCC.

Three major limitations existed in the present study. The first one is the three-lncRNA signature only identified and validated in TCGA dataset. Therefore, the signature required further confirmation in large cohorts in the future studies. The other one we only used is bioinformatics method to predict the three-lncRNA signature in HNSCC, thus further experiments need to be conducted. Last, the two characteristics of smoking and radiotherapy were not considered in the present study. Our team will continue to investigate this field in the future.

Through the study of lncRNA expression profiles in HNSCC patients, a three-lncRNA signature associated with OS in HNSCC patients was identified. A three-lncRNA signature was constructed to predict the survival of HNSCC patients in the training dataset, and then, the signature was validated in the testing dataset and the entire dataset. Subsequently, multivariable Cox regression analysis of the three-lncRNA signature was investigated. The results demonstrated that the prognostic power of the three-lncRNA signature was independent of other clinical features. Therefore, the three-lncRNA signature can provide novel insights to predict HNSCC patient survival.



**Figure 5. Functional enrichment results of the co-expressed protein-coding genes with prognostic lncRNAs**  
(A) Significantly enriched GO terms. (B) Significantly enriched KEGG pathway.

## Funding

This work was financially supported by grants from the National Natural Science Foundation of China [81772874 and 81272965].

## Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

## Author Contribution

P.W. and Y.S. conceived and designed the experiments. P.W., M.J., C.S., L.Y., Y.L., X.W., L.T. and M.L. analyzed data. P.W. and Y.S. wrote this manuscript. All authors read and approved the final manuscript.

## Abbreviations

AUC, area under the ROC curves; FDR, false discovery rate; HNSCC, head and neck squamous cell carcinoma; HPV, human papillomavirus; ncRNA, non-coding RNA; ROC, receiver operating characteristic; TCGA, The Cancer Genome Atlas.

## References

- 1 Leemans, C.R., Braakhuis, B.J. and Brakenhoff, R.H. (2011) The molecular biology of head and neck cancer. *Nat. Rev. Cancer* **11**, 9–22, <https://doi.org/10.1038/nrc2982>
- 2 Rothenberg, S.M. and Ellisen, L.W. (2012) The molecular pathogenesis of head and neck squamous cell carcinoma. *J. Clin. Invest.* **122**, 1951–1957, <https://doi.org/10.1172/JCI59889>

- 3 Lewin, F., Norell, S.E., Johansson, H., Gustavsson, P., Wennerberg, J., Biorklund, A. et al. (1998) Smoking tobacco, oral snuff, and alcohol in the etiology of squamous cell carcinoma of the head and neck: a population-based case-referent study in Sweden. *Cancer* **82**, 1367–1375, [https://doi.org/10.1002/\(SICI\)1097-0142\(19980401\)82:7%3c1367::AID-CNCR21%3e3.0.CO;2-3](https://doi.org/10.1002/(SICI)1097-0142(19980401)82:7%3c1367::AID-CNCR21%3e3.0.CO;2-3)
- 4 Li, G. and Sturgis, E.M. (2006) The role of human papillomavirus in squamous carcinoma of the head and neck. *Curr. Oncol. Rep.* **8**, 130–139, <https://doi.org/10.1007/s11912-006-0048-y>
- 5 Wahlestedt, C. (2013) Targeting long non-coding RNA to therapeutically upregulate gene expression. *Nat. Rev. Drug Discov.* **12**, 433–446, <https://doi.org/10.1038/nrd4018>
- 6 Dhamija, S. and Diederichs, S. (2016) From junk to master regulators of invasion: lncRNA functions in migration, EMT and metastasis. *Int. J. Cancer* **139**, 269–280, <https://doi.org/10.1002/ijc.30039>
- 7 Dykes, I.M. and Emanuelli, C. (2017) Transcriptional and post-transcriptional gene regulation by long non-coding RNA. *Genomics Proteomics Bioinformatics* **15**, 177–186, <https://doi.org/10.1016/j.gpb.2016.12.005>
- 8 Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J. et al. (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071–1076, <https://doi.org/10.1038/nature08975>
- 9 Mercer, T.R., Dinger, M.E. and Mattick, J.S. (2009) Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* **10**, 155–159, <https://doi.org/10.1038/nrg2521>
- 10 Li, C.H. and Chen, Y. (2013) Targeting long non-coding RNAs in cancers: progress and prospects. *Int. J. Biochem. Cell Biol.* **45**, 1895–1910, <https://doi.org/10.1016/j.biocel.2013.05.030>
- 11 Matouk, I.J., DeGroot, N., Mezan, S., Ayesb, S., Abu-lail, R., Hochberg, A. et al. (2007) The H19 non-coding RNA is essential for human tumor growth. *PLoS One* **2**, e845, <https://doi.org/10.1371/journal.pone.0000845>
- 12 Zhou, Y., Zhang, X. and Klibanski, A. (2012) MEG3 noncoding RNA: a tumor suppressor. *J. Mol. Endocrinol.* **48**, R45–R53, <https://doi.org/10.1530/JME-12-0008>
- 13 Sun, H., Wang, G., Peng, Y., Zeng, Y., Zhu, Q.N., Li, T.L. et al. (2015) H19 lncRNA mediates 17beta-estradiol-induced cell proliferation in MCF-7 breast cancer cells. *Oncol. Rep.* **33**, 3045–3052, <https://doi.org/10.3892/or.2015.3899>
- 14 Zhang, J., Yao, T., Wang, Y., Yu, J., Liu, Y. and Lin, Z. (2016) Long noncoding RNA MEG3 is downregulated in cervical cancer and affects cell proliferation and apoptosis by regulating miR-21. *Cancer Biol. Ther.* **17**, 104–113, <https://doi.org/10.1080/15384047.2015.1108496>
- 15 Chen, M., Liu, B., Xiao, J., Yang, Y. and Zhang, Y. (2017) A novel seven-long non-coding RNA signature predicts survival in early stage lung adenocarcinoma. *Oncotarget* **8**, 14876–14886
- 16 Li, J., Chen, Z., Tian, L., Zhou, C., He, M.Y., Gao, Y. et al. (2014) lncRNA profile study reveals a three-lncRNA signature associated with the survival of patients with oesophageal squamous cell carcinoma. *Gut* **63**, 1700–1710, <https://doi.org/10.1136/gutjnl-2013-305806>
- 17 Hu, Y., Chen, H.Y., Yu, C.Y., Xu, J., Wang, J.L., Qian, J. et al. (2014) A long non-coding RNA signature to improve prognosis prediction of colorectal cancer. *Oncotarget* **5**, 2230–2242, <https://doi.org/10.18632/oncotarget.1895>
- 18 Zhou, M., Zhao, H., Xu, W., Bao, S., Cheng, L. and Sun, J. (2017) Discovery and validation of immune-associated long non-coding RNA biomarkers associated with clinically molecular subtype and prognosis in diffuse large B cell lymphoma. *Mol. Cancer* **16**, 16, <https://doi.org/10.1186/s12943-017-0580-4>
- 19 Zhou, M., Zhang, Z., Zhao, H., Bao, S., Cheng, L. and Sun, J. (2018) An immune-related six-lncRNA signature to improve prognosis prediction of glioblastoma multiforme. *Mol. Neurobiol.* **55**, 3684–3697, <https://doi.org/10.1007/s12035-017-0572-9>
- 20 Zhou, M., Hu, L., Zhang, Z., Wu, N., Sun, J. and Su, J. (2018) Recurrence-associated long non-coding RNA signature for determining the risk of recurrence in patients with colon cancer. *Mol. Ther. Nucleic Acids* **12**, 518–529, <https://doi.org/10.1016/j.omtn.2018.06.007>
- 21 Sun, J., Cheng, L., Shi, H., Zhang, Z., Zhao, H., Wang, Z. et al. (2016) A potential panel of six-long non-coding RNA signature to improve survival prediction of diffuse large-B-cell lymphoma. *Sci. Rep.* **6**, 27842, <https://doi.org/10.1038/srep27842>
- 22 Sun, J., Chen, X., Wang, Z., Guo, M., Shi, H., Wang, X. et al. (2015) A potential prognostic long non-coding RNA signature to predict metastasis-free survival of breast cancer patients. *Sci. Rep.* **5**, 16553, <https://doi.org/10.1038/srep16553>
- 23 Zhou, M., Guo, M., He, D., Wang, X., Cui, Y., Yang, H. et al. (2015) A potential signature of eight long non-coding RNAs predicts survival in patients with non-small cell lung cancer. *J. Transl. Med.* **13**, 231, <https://doi.org/10.1186/s12967-015-0556-3>
- 24 Nagadia, R., Pandit, P., Coman, W.B., Cooper-White, J. and Punyadeera, C. (2013) miRNAs in head and neck cancer revisited. *Cell Oncol.* **36**, 1–7, <https://doi.org/10.1007/s13402-012-0122-4>
- 25 Satapathy, S., Batra, J., Jeet, V., Thompson, E.W. and Punyadeera, C. (2017) MicroRNAs in HPV associated cancers: small players with big consequences. *Expert Rev. Mol. Diagn.* **17**, 711–722, <https://doi.org/10.1080/14737159.2017.1339603>
- 26 Wan, Y., Vagenas, D., Salazar, C., Kenny, L., Perry, C., Calvopina, D. et al. (2017) Salivary miRNA panel to detect HPV-positive and HPV-negative head and neck cancer patients. *Oncotarget* **8**, 99990–100001, <https://doi.org/10.18632/oncotarget.21725>
- 27 Cao, W., Liu, J.N., Liu, Z., Wang, X., Han, Z.G., Ji, T. et al. (2017) A three-lncRNA signature derived from the Atlas of ncRNA in cancer (TANRIC) database predicts the survival of patients with head and neck squamous cell carcinoma. *Oral. Oncol.* **65**, 94–101, <https://doi.org/10.1016/j.oraloncology.2016.12.017>
- 28 Zhang, Z.L., Zhao, L.J., Chai, L., Zhou, S.H., Wang, F., Wei, Y. et al. (2017) Seven lncRNA-mRNA based risk score predicts the survival of head and neck squamous cell carcinoma. *Sci. Rep.* **7**, 309, <https://doi.org/10.1038/s41598-017-00252-2>
- 29 Li, J., Han, L., Roebuck, P., Diao, L., Liu, L., Yuan, Y. et al. (2015) TANRIC: An Interactive Open Platform to Explore the Function of lncRNAs in Cancer. *Cancer Res.* **75**, 3728–3737, <https://doi.org/10.1158/0008-5472.CAN-15-0273>
- 30 Therneau T. A package for survival analysis in S. R package version 2.37-4. URL <http://CRAN.R-project.org/package=survival>. **2013**; 980032: 23298-0032

- 31 Heagerty, P.J. and Zheng, Y. (2005) Survival model predictive accuracy and ROC curves. *Biometrics* **61**, 92–105, <https://doi.org/10.1111/j.0006-341X.2005.030814.x>
- 32 Heagerty, P.J., Lumley, T. and Pepe, M.S. (2000) Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* **56**, 337–344, <https://doi.org/10.1111/j.0006-341X.2000.00337.x>
- 33 Siegel, R., Ma, J., Zou, Z. and Jemal, A. (2014) Cancer statistics, 2014. *CA Cancer J. Clin.* **64**, 9–29, <https://doi.org/10.3322/caac.21208>
- 34 Ginos, M.A., Page, G.P., Michalowicz, B.S., Patel, K.J., Volker, S.E., Pambuccian, S.E. et al. (2004) Identification of a gene expression signature associated with recurrent disease in squamous cell carcinoma of the head and neck. *Cancer Res.* **64**, 55–63, <https://doi.org/10.1158/0008-5472.CAN-03-2144>
- 35 Qiu, Z., Sun, W., Gao, S., Zhou, H., Tan, W., Cao, M. et al. (2017) A 16-gene signature predicting prognosis of patients with oral tongue squamous cell carcinoma. *Peer J.* **5**, e4062, <https://doi.org/10.7717/peerj.4062>
- 36 Guo, W., Chen, X., Zhu, L. and Wang, Q. (2017) A six-mRNA signature model for the prognosis of head and neck squamous cell carcinoma. *Oncotarget* **8**, 94528–94538
- 37 Shen, S., Bai, J., Wei, Y., Wang, G., Li, Q., Zhang, R. et al. (2017) A seven-gene prognostic signature for rapid determination of head and neck squamous cell carcinoma survival. *Oncol. Rep.* **33**, 3403–3411
- 38 Wang, Y. (2017) Abstract 2545: Upregulated lncRNA-HNGA1, a target of miR-378a, contributes to aerobic glycolysis of head and neck squamous cell carcinoma through increasing levels of the C-C chemokine receptor type 7 (CCR7).
- 39 Bao, Z., Zhang, W. and Dong, D. (2017) A potential prognostic lncRNA signature for predicting survival in patients with bladder urothelial carcinoma. *Oncotarget* **8**, 10485–10497, <https://doi.org/10.18632/oncotarget.14441>
- 40 Meng, J., Li, P., Zhang, Q., Yang, Z. and Fu, S. (2014) A four-long non-coding RNA signature in predicting breast cancer survival. *J. Exp. Clin. Cancer Res.* **33**, 84, <https://doi.org/10.1186/s13046-014-0084-7>
- 41 Fang, T.T., Sun, X.J., Chen, J., Zhao, Y., Sun, R.X., Ren, N. et al. (2014) Long non-coding RNAs are differentially expressed in hepatocellular carcinoma cell lines with differing metastatic potential. *Asian Pac. J. Cancer Prev.* **15**, 10513–10524, <https://doi.org/10.7314/APJCP.2014.15.23.10513>
- 42 Bryzghalov, O., Szczesniak, M.W. and Makalowska, I. (2016) Retroposition as a source of antisense long non-coding RNAs with possible regulatory functions. *Acta Biochim. Pol.* **63**, 825–833, <https://doi.org/10.18388/abp.20161354>
- 43 Liu, B., Yang, F. and Chou, K.C. (2017) 2L-piRNA: a two-layer ensemble classifier for identifying Piwi-interacting RNAs and their function. *Mol. Ther. Nucleic Acids* **7**, 267–277, <https://doi.org/10.1016/j.omtn.2017.04.008>
- 44 Liu, B., Fang, L., Liu, F., Wang, X. and Chou, K.C. (2016) iMiRNA-PseDPC: microRNA precursor identification with a pseudo distance-pair composition approach. *J. Biomol. Struct. Dyn.* **34**, 223–235, <https://doi.org/10.1080/07391102.2015.1014422>
- 45 Liu, B., Wang, S., Long, R. and Chou, K.C. (2017) iRSpot-EL: identify recombination spots with an ensemble learning approach. *Bioinformatics* **33**, 35–41, <https://doi.org/10.1093/bioinformatics/btw539>
- 46 Liu, B., Liu, F., Wang, X., Chen, J., Fang, L. and Chou, K.C. (2015) Pse-in-One: a web server for generating various modes of pseudo components of DNA, RNA, and protein sequences. *Nucleic Acids Res.* **43**, W65–W71, <https://doi.org/10.1093/nar/gkv458>
- 47 Leelahavanichkul, K., Amornphimoltham, P., Molinolo, A.A., Basile, J.R., Koontongkaew, S. and Gutkind, J.S. (2014) A role for p38 MAPK in head and neck cancer cell growth and tumor-induced angiogenesis and lymphangiogenesis. *Mol. Oncol.* **8**, 105–118, <https://doi.org/10.1016/j.molonc.2013.10.003>
- 48 Samara, G.J., Lawrence, D.M., Chiarelli, C.J., Valentino, M.D., Lyubsky, S., Zucker, S. et al. (2004) CXCR4-mediated adhesion and MMP-9 secretion in head and neck squamous cell carcinoma. *Cancer Lett.* **214**, 231–241, <https://doi.org/10.1016/j.canlet.2004.04.035>
- 49 Nohata, N., Hanazawa, T., Kikkawa, N., Mutallip, M., Fujimura, L., Yoshino, H. et al. (2011) Caveolin-1 mediates tumor cell migration and invasion and its regulation by miR-133a in head and neck squamous cell carcinoma. *Int. J. Oncol.* **38**, 209–217