

SCIENTIFIC REPORTS



OPEN

Seven lncRNA-mRNA based risk score predicts the survival of head and neck squamous cell carcinoma

Zhi-Li Zhang¹, Li-jing Zhao², Liang Chai¹, Shui-Hong Zhou¹, Feng Wang¹, Yan Wei¹, Ya-Ping Xu¹ & Peng Zhao³

Received: 2 December 2016

Accepted: 15 February 2017

Published online: 22 March 2017

Dysregulation of mRNAs and long non-coding RNAs (lncRNAs) is one of the most important features of carcinogenesis and cancer development. However, studies integrating the expression of mRNAs and lncRNAs to predict the survival of head and neck squamous cell carcinoma (HNSC) are still limited, hitherto. In current work, we identified survival related mRNAs and lncRNAs in three datasets (TCGA dataset, E-TABM-302, GSE41613). By random forest, seven gene signatures (six mRNAs and lncRNA) were further selected to develop the risk score model. The risk score was significantly associated with survival in both training and testing datasets (E-TABM-302, GSE41613, and E-MTAB-1324). Furthermore, correlation analyses showed that the risk score is independent from clinicopathological features. According to Cox multivariable hazard model and nomogram, the risk score contributes the most to survival than the other clinical information, including gender, age, histologic grade, and alcohol taking. The Gene Set Enrichment Analysis (GSEA) indicates that the risk score is associated with cancer related pathways. In summary, the lncRNA-mRNA based risk score model we developed successfully predicts the survival of 755 HNSC samples in five datasets and two platforms. It is independent from clinical information and performs better than clinical information for prognosis.

Head and neck carcinoma (HNSC) is one of the most causes of death in the world¹, and 600,000 new cases occurred every year, with 50% increase². Among those patients, more than 90% were identified as squamous cell carcinoma. Smoking, alcohol taking and human papilloma virus (HPV) are all important causes of HNSC carcinogenesis³. The prognosis of HNSC currently is unfavorable. The five-year survival rate of HNSC is about 50–60%⁴, and the survival rate of HNSC did not improved in the past 30 years⁵.

Long non-coding RNAs (lncRNAs) are newly discovered non-coding RNAs, with more than 200 nucleotides and no protein coding ability⁶. The gene expression abundance of lncRNAs is much lower than mRNA. Despite of this, lncRNAs play important roles in cell multiple biological processes⁷, including development⁸, diseases⁹, epigenetic modulation¹⁰, and scaffolding¹¹. Recently, studies have emphasized the role of lncRNAs on carcinogenesis and cancer progression^{12–16} among cancer types, including head and neck squamous cell carcinoma^{17, 18}. MALAT1 was identified as an oncogene, and the high expression of MALAT1 was associated with metastasis and poor survival across cancers^{19–21}. Another lncRNA, DANCR was reported to block the repression effect of miR-214, miR-320a, and miR-199a on CTNBN1 in hepatocellular carcinoma²². MEG3 was reported to involved in regulation of epithelial-mesenchymal transition in lung cancer cells²³. According to recent study on lncRNA GAS5, the allele deletion of rs145204276 was associated with decreased CRC risk and less possibility of lymph-node metastasis²⁴. On the other side, mRNAs encode proteins and thus control almost every cell progress. Thus, integrating abundance of lncRNAs and mRNAs enhances the prognostic effect²⁵.

In this article, we identified lncRNAs and mRNAs from three independent datasets with Cox univariate regression model, and selected seven lncRNA-mRNAs to develop the risk score with random survival forests variable hunting algorithm. Patients with high risk score have relatively poor prognoses than those with low risk score, in both training and test datasets. The risk score is independent with clinical observations including sex,

¹Department of otorhinolaryngology, the first affiliated hospital, Zhejiang University School of medicine, 310003, Qingchun Road 79, Hangzhou city, Zhejiang province, China. ²Department of otorhinolaryngology, the second affiliated hospital, Zhejiang University school of medicine, 310003, Qingchun Road 79, Hangzhou city, Zhejiang province, China. ³Department of Oncology, The first affiliated hospital, Zhejiang University School of medicine, 310003, Qingchun Road 79, Hangzhou city, Zhejiang province, China. Correspondence and requests for materials should be addressed to Z.-L.Z. (email: zhangzhili_zju@126.com)

age, histologic grade, alcohol taking, and primary tumor size. The prognostic performance of the risk score on survival is more favorable than aforementioned clinical information. In summary, the seven lncRNA-mRNA risk score is robust in predicting the survival of HNSC patients, and performs better than clinicopathological information.

Method and Material

Expression data processing. Processed TCGA expression data was downloaded from TCGA official website (<http://cancergenome.nih.gov/>), and the upper quantile normalized FPKM (fragments per kilobases per million) values were used. After removing normal samples, genes expressed in over 80% cancer samples were retained. The zero values in the expression matrix were replaced with (minimum non-zero FPKM value)/2 of corresponding gene. Then the expression data were log 2 transformed. The lncRNA expression data was downloaded from TANRIC (<http://bioinformatics.mdanderson.org/main/TANRIC:Overview>), and processed as mRNAs. GSE41613 data set was downloaded from GEO (<https://www.ncbi.nlm.nih.gov/geo/>), the raw data were normalized with “cyclicloess”²⁶.

Probes were matched to the gene names according to the manufacture provided annotation file. If a single gene matches multiple probes, probes were integrated by using the arithmetic mean to represent the expression level of single lncRNA. The workflow of this work was shown in Figure S1.

lncRNA and mRNA matching between platforms. Probe target sequences (HG U133 plus 2) was downloaded from Affymetrix website (<http://www.affymetrix.com/>), and these sequences were mapped to the lncRNA sequences from ENSEMBL (<http://ensembl.org/index.html>) with bowtie1 software (<http://bowtie-bio.sourceforge.net/index.shtml>)²⁷. This step allows no mismatch and only probes matched to the forward sequences were retained (strand-specific). Probes matched more than one lncRNAs were discarded. For single lncRNAs matching multiple probes, probes were integrated by using the arithmetic mean to represent the expression level of single lncRNA. Totally, 3505 lncRNAs were used for further analysis. Afterwards, the mRNAs and lncRNAs detected in both platforms (NGS and Affymetrix HG U133 plus 2) were retained for further analysis. Z-score was calculated for each gene among samples in each dataset.

Gene selection, Cox multivariate regression, and resampling. Cox univariate regression²⁸ was carried out on TCGA, E-TABM-302, and GSE41613 datasets. lncRNAs and mRNAs significantly associated survival ($p < 0.05$) in all of these three datasets were retained for further analyses, and 33 mRNAs and lncRNAs identified. Random forest variable hunting algorithm was used for select biomarker combination for prognosis (50 replications and 50 iterations)²⁹. Cox multivariate regression model were used to develop the risk score model:

$$\text{risk score} = \sum_{i=1}^n \beta_i * x_i$$

where β_i indicates the coefficient for each gene and x_i indicates the z-score transformed relative expression value of each gene.

Resampling was carried out, 80% of all samples randomly selected. After dividing these samples into high/low risk group, survival difference (p values) was calculated. Repeat this step for 10,000 times, distribution of p values (survival of high-risk vs low-risk) was plotted (Figure S2).

Statistical analysis. All statistical analysis were performed with R (<https://www.r-project.org/>, v3.0.1) and R packages. Normalization of Affymetrix microarray data were implemented with R package “limma”²⁶. The survival analysis, cox univariate regression, cox multivariate regression and cox probability hazard model were carried out on R package “survival”³⁰. Random forest variable hunting was performed with package “randomForestSRC”. The ROC curve was plotted with R package “pROC”³¹, and the nomogram were plotted with R package “rms”³². The Gene Set Enrichment Analysis was implemented with java software GSEA (<http://software.broadinstitute.org/gsea/index.jsp>)³³.

Results

Identification of survival related mRNAs and lncRNAs. In order to avoid result-data overfit, cox univariate regression was performed in three independent datasets, including TCGA (N = 407, overall survival), GSE41613 (N = 97, overall survival), and E-TABM-302 (N = 81, overall survival). Significant survival related mRNAs and lncRNAs ($p < 0.05$) in all of these datasets were identified, which yields 33 genes, including 2 lncRNAs and 31 mRNAs. To narrow down the panel, random forest variable hunting algorithm was implemented and 7 genes, including 6 mRNAs (LCLAT1, WDT1, MINK1, TOM1L2, AMPD3, CCDC43) and 1 lncRNA (ENSG00000269386), were selected. These genes are differentially expressed between normal and tumor tissues (Fig. 1A). The hazard ratio of each gene was evaluated (Fig. 1B). The hazard ratios of LCLAT1 and CCDC43 are positive, indicating the anti-survival function of these genes, while the coefficients of other 5 genes are negative, including ENSG00000269386 (RAB11B-AS1), TOM1L2, AMPD3, MINK1 and WDT1, indicating that they may be tumor suppressor genes.

Performance of risk score in training dataset. To integrate all these seven genes selected in our previous step, a cox multivariable regression model was developed on the largest cohort, TCGA dataset. The risk score of each patient were identified as the following: risk score = (1.6136310285262 × LCLAT1) + (−0.58247013428054 × WDT1) + (−0.351319390355873 × MINK1) + (−0.538752453473946 × TOM1L2) + (−0.303600743722099 × AMPD3) + (−0.200539523626066 × ENSG00000269386) + (0.386646135470189 × CCDC43). The corresponding p-values were shown in Table 1. The risk score for each patient in the TCGA dataset was calculated, and

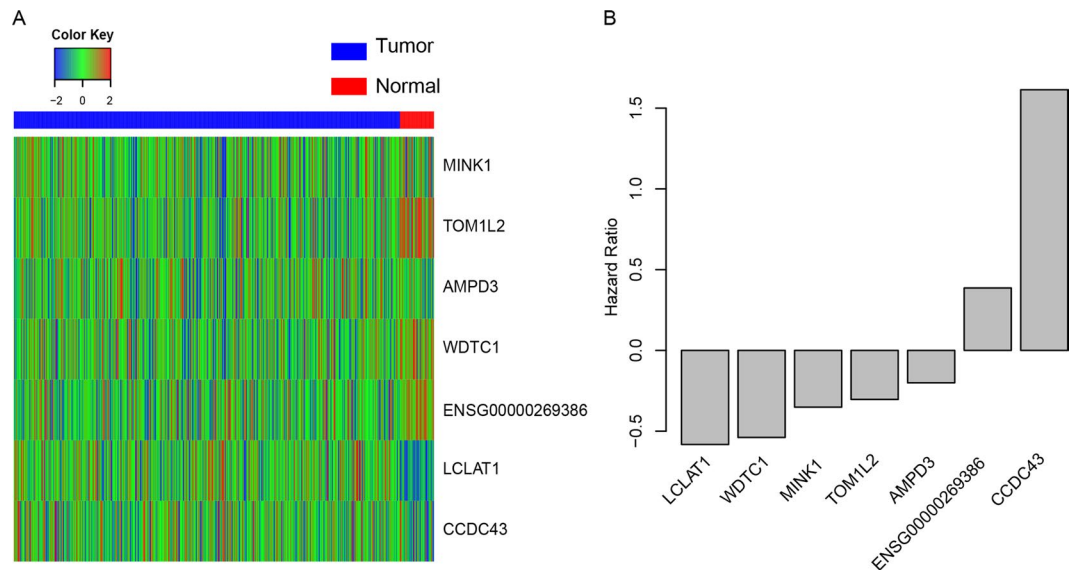


Figure 1. Seven mRNAs and lncRNA identified. **(A)** The expression pattern of the normal tissues (red) and tumor tissues (blue) is different. **(B)** The hazard ratio of seven RNAs.

Gene	Coefficients	p value
LCLAT1	1.613631029	0.024
WDTC1	-0.582470134	0.036
MINK1	-0.35131939	0.05
TOM1L2	-0.538752453	0.006
AMPD3	-0.303600744	0.019
ENSG00000269386	-0.200539524	0.018
CCDC43	0.386646135	0.0054

Table 1. The Coefficients and p values of each gene. P values were generated from univariate cox regression in TCGA dataset.

high risk group/low risk group were divided using the median risk score as cutoff. We evaluated the survival of the high risk group and low risk group, and the survival time of high risk group is significantly shorter than the low risk group (Fig. 2A). In addition, the tumor-free survival (TFS) difference between high/low risk group was evaluated, and the result shows high risk group has a significantly less tumor-free survival rate, in consistent with the overall survival (Fig. 2B). To avoid overfit due to cancer heterogeneity³⁴, resampling (80% samples, 10,000 repeats) was used to evaluate the performance of risks score, and the results indicates that 97.8% resampling was significantly associated with risk score ($p < 0.05$), as shown in Figure S1. And according to the three-year survival receiving operating characteristic (ROC) curve, the area under curve (AUC) of risk score reached 0.66 (Fig. 2C), which is significantly higher than other clinical information, indicating that the seven lncRNA-mRNA based risk score is a good indicator for prognosis. The detailed risk score, seven lncRNA-mRNA expression and survival information was shown in Fig. 2D.

Performance of risk score in test datasets. To further validate the robustness of lncRNA-mRNA risk score model developed in the TCGA dataset (RNA-seq), the performance of the risk score was also evaluated in three independent datasets (microarray), E-TABM-302 (N = 81) and GSE41613 (N = 97) using coefficients in training datasets. The high risk and low risk groups were also divided according to the median risk score. The overall survival time in high risk group was significantly shorter than that in low risk group, in both datasets (Fig. 3A,B). Considering that the expression of candidates lncRNAs and mRNAs were generated from those two datasets, over fit may exists. We utilized another totally independent datasets, E-MTAB-1328 (N = 89) for further validation. The tumor-free survival profile and gene expression pattern resemble our previously conclusion (Fig. 3C).

We also randomly selected seven genes from 33 previously identified genes, and developed a risk score model with these genes. However, these models are not robust in the other validation datasets (Table S1), indicating that our method performs better. In summary, our results indicate that the seven lncRNA-mRNA based risk score is robust across platforms and datasets.

Risk score and clinicopathological information for prognosis. Correlation analyses were implemented between risk score and clinical indicator, which include gender, age, alcohol taking, histologic grade,

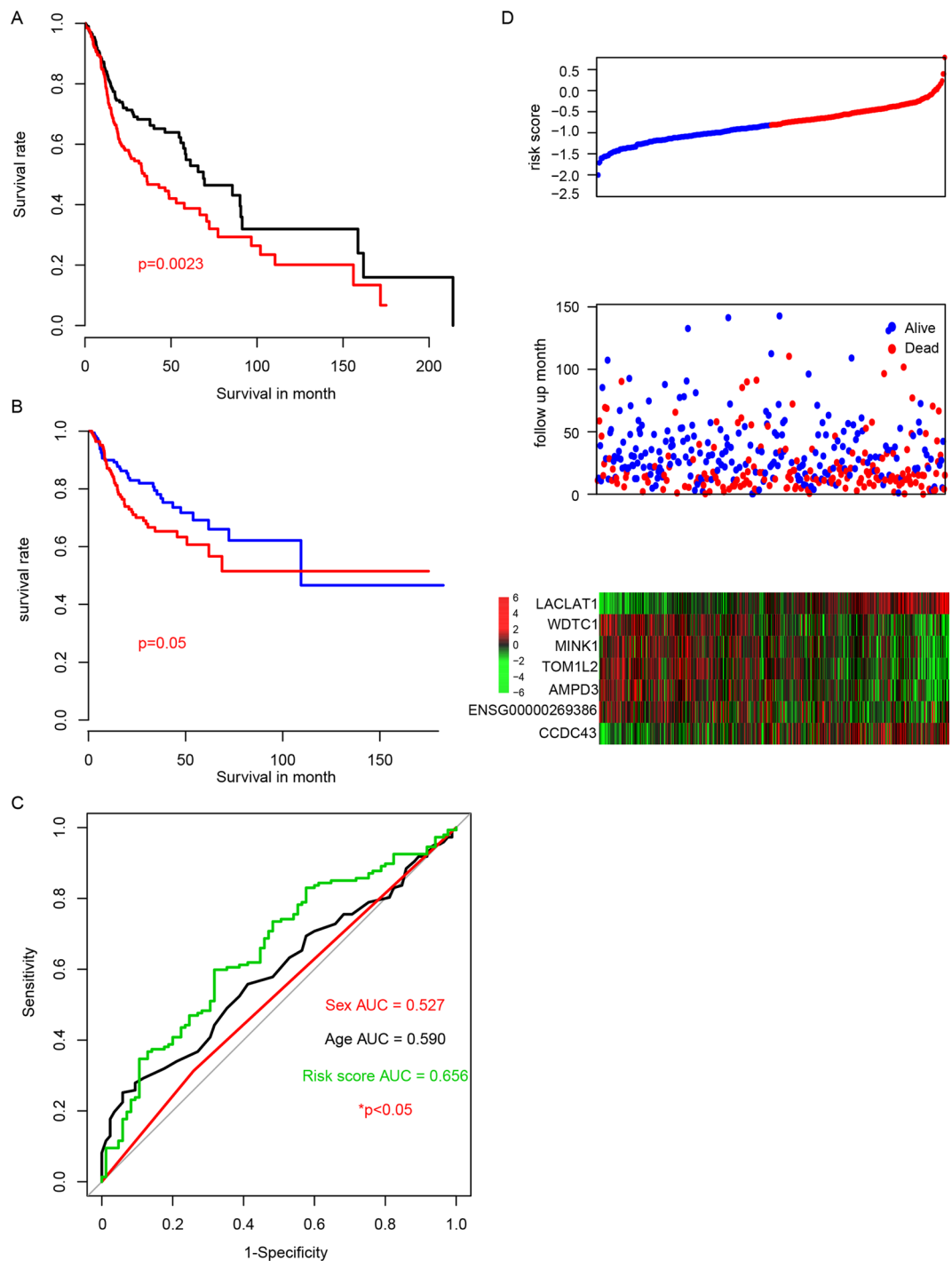


Figure 2. The risk score performance in the TCGA (training) datasets. The overall survival of high risk score group and low risk score group (A), and the recurrence-free survival between high/low risk score groups (B). The 3-year survival receiving operating characteristic curve (ROC) of age (black), gender (red) and risk score (green) (C). (D) The relationship between risk score (upper), survival information (middle) and z-score transformed expression value (bottom) was shown.

diameter of primary tumor, on the largest cohort, TCGA dataset. As show in Fig. 4A, none of these clinical information is associated with our seven lncRNA-mRNA based risk score, indicating that the risk score is independent from the clinical information.

Furthermore, to identify the importance of the clinicopathological information and the risk score, cox multi-variable probability hazard model was employed (Fig. 4B). The risk score is the most significant correlated with the survival information ($p=0.0013$) and highest median risk score (HR = 2, 95% CI = 1.2–3.1), indicating that the risk score performs better than the other clinical information. To facilitate the utilization of risk score, a 3-year survival nomogram were plotted (Fig. 4C) considering risk score and aforementioned clinicopathological

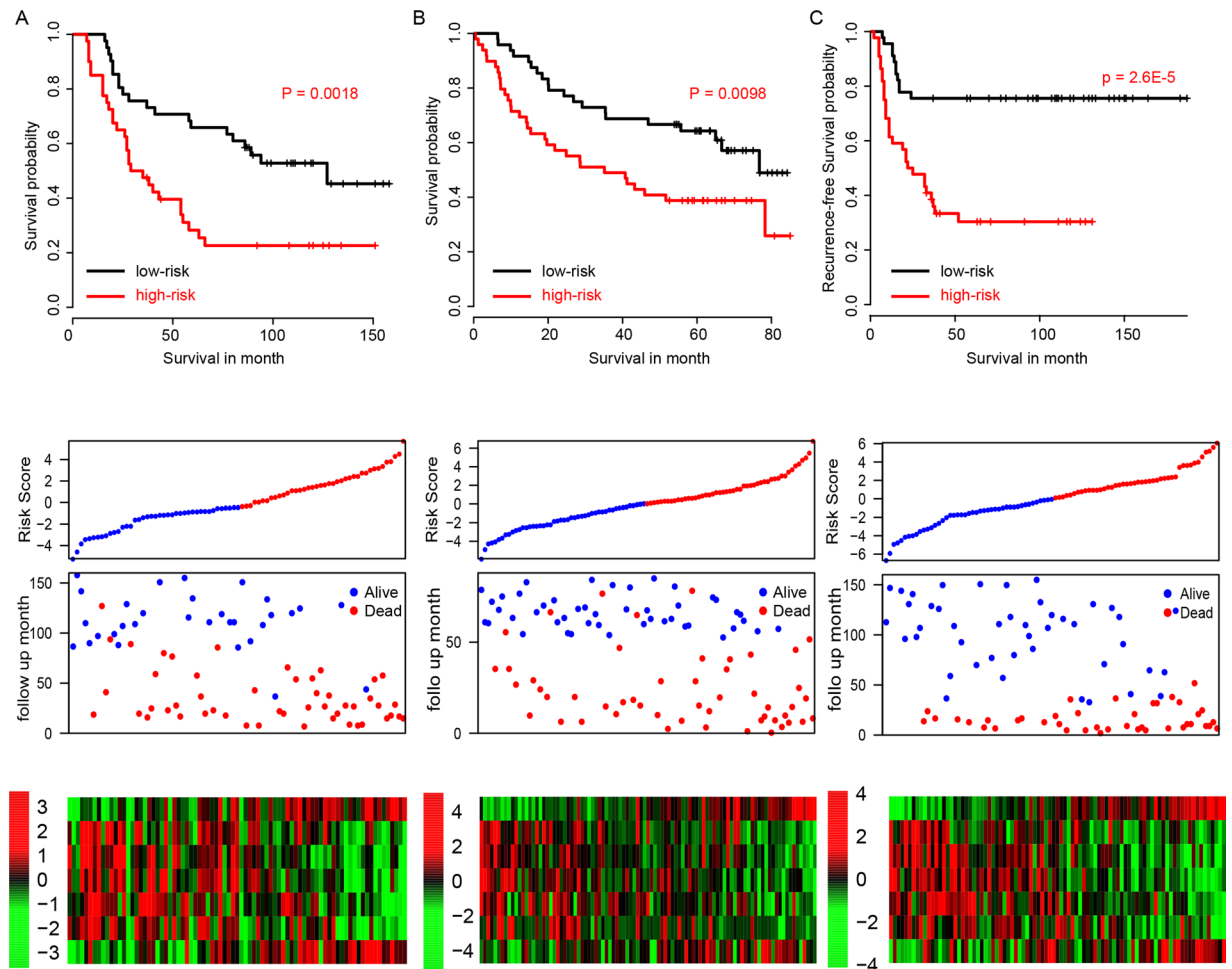


Figure 3. The survival information of three independent test datasets. The survival plot (upper panel), risk score (2nd panel), survival information (3rd panel) and z-score transformed expression value (top-down, LACLAT1, WDTC1, MINK1, TOM1L2, AMPD3, ENSG00000269386, CCDC43) were shown in E-TABM-302 (A), GSE41613 (B), and E-MTAB-1328 (C).

observations. In consistent with our cox multivariable regression results, the risk score contributes the most risk points (ranged 0–100), whereas the other clinical information contributes much less (diameter of primary tumor, ranged 0–65).

Altogether, these results indicate that the seven lncRNA-mRNA based risk score is independent from the clinical information, and performs better in the survival prediction than that clinical information.

Risk score and radiation therapy. Radiation therapy is the most important adjuvant HNSC treatment method. Thus, the correlation between risk score and the radiation outcome were evaluated. We used two independent datasets providing the radiation therapy information to test whether our risk score is also available for survival prediction of patients underwent radiation therapy. As we expected, the patients with high risk score who received radiation had a worse prognosis in both TCGA (Fig. 5A) and E-TABM-302 (Fig. 5B) dataset, compared to the low risk group. These results indicate that the risk score is also available for the prognosis of HNSC patients with radiation therapy.

Altered pathways in high and low risk score group. To investigate the potential altered pathways in the high risk group, Gene Set Enrichment Analysis (GSEA) was implemented between high/low risk groups. According to the results, we noticed that KEGG pathways including “homologous recombination”, “RNA polymerase”, “RNA degradation”, “pentose and glucuronate interconversions” and “DNA replication” were significantly enriched (p value < 0.01, Fig. 6A). Of these pathways, DNA replication and pentose pathways were noted for its role in carcinogenesis and cancer maintenance (Fig. 6B,C), suggesting that the seven lncRNA-mRNA based risk score may reflect the status of these cell processes, and thus predict the survival of HNSC patients.

Discussion

The heterogeneity of head and neck squamous carcinoma is high according to a lot of studies^{35, 36}, which may due to the heterogeneity of causes of HNSC³⁷, including alcohol abusing, HPV infection and diet. It makes the

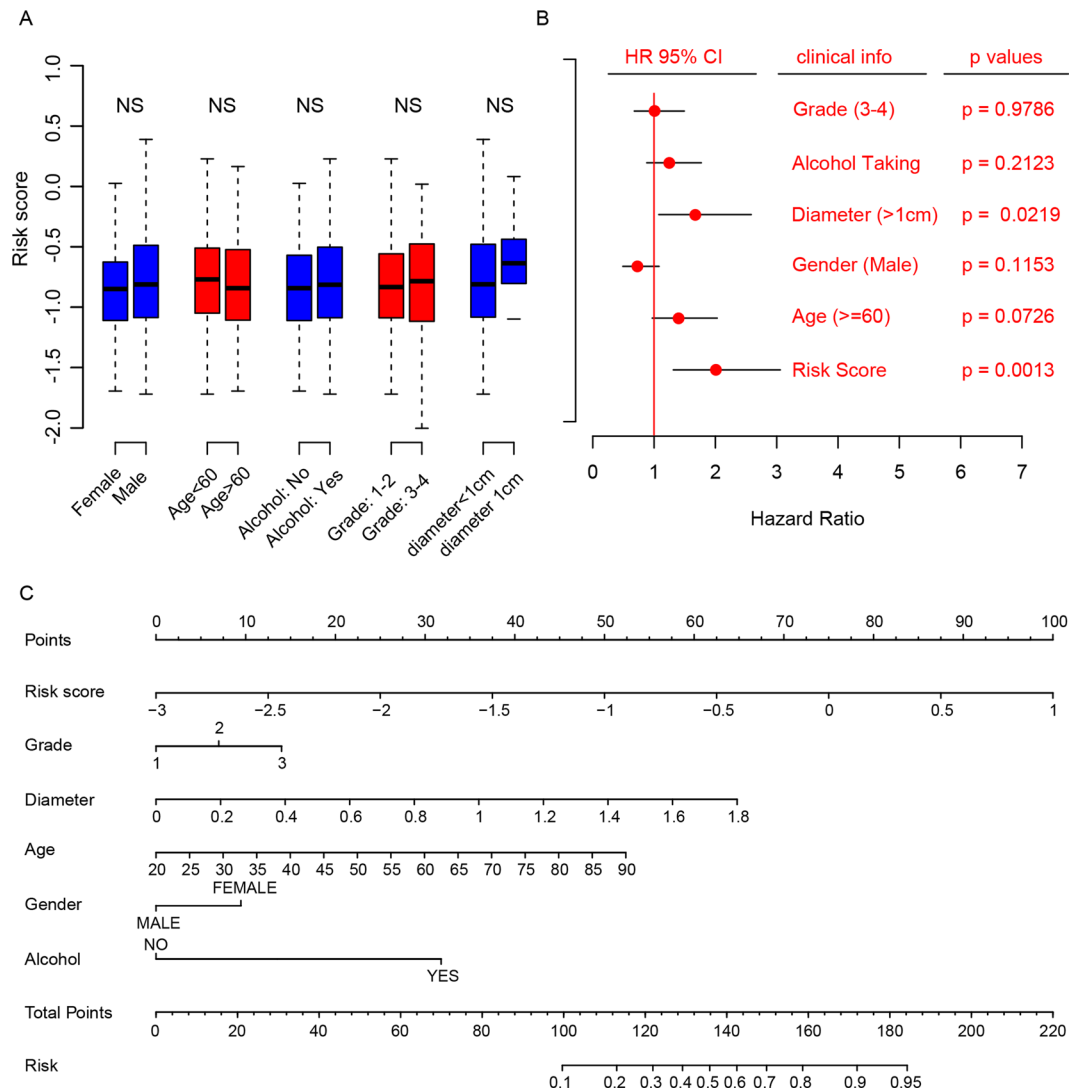


Figure 4. The relationship between risk score and other clinical information. **(A)** The correlation between risk score and clinical information is not significantly different ($p > 0.05$, NS). **(B)** Cox multivariate regression with clinical information and risk score for survival. **(C)** Nomogram for predicting the 3-year event (death) with risk score and clinical information.

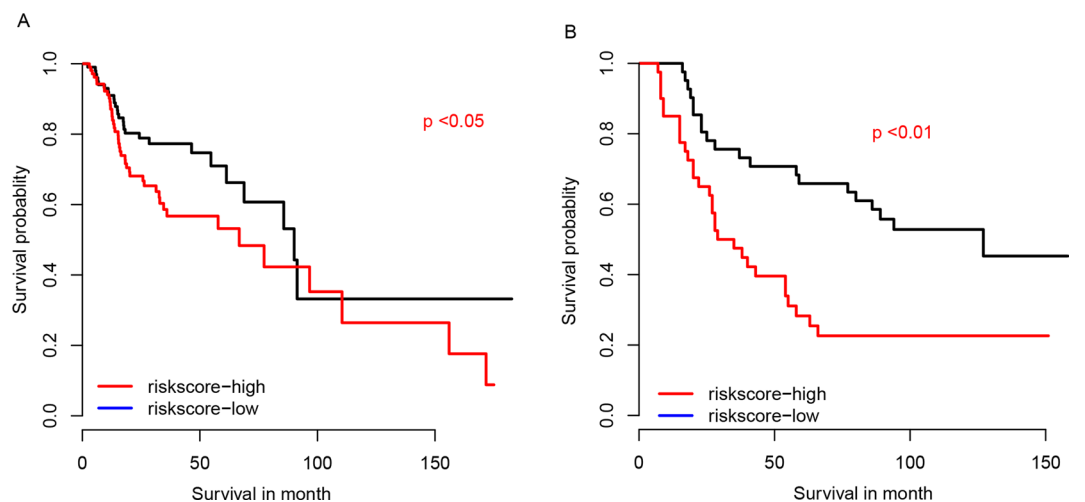


Figure 5. The survival rate of patients with radiation therapy. The survival of patients received radiation in TCGA **(A)** and E-TABM-302 datasets **(B)**.

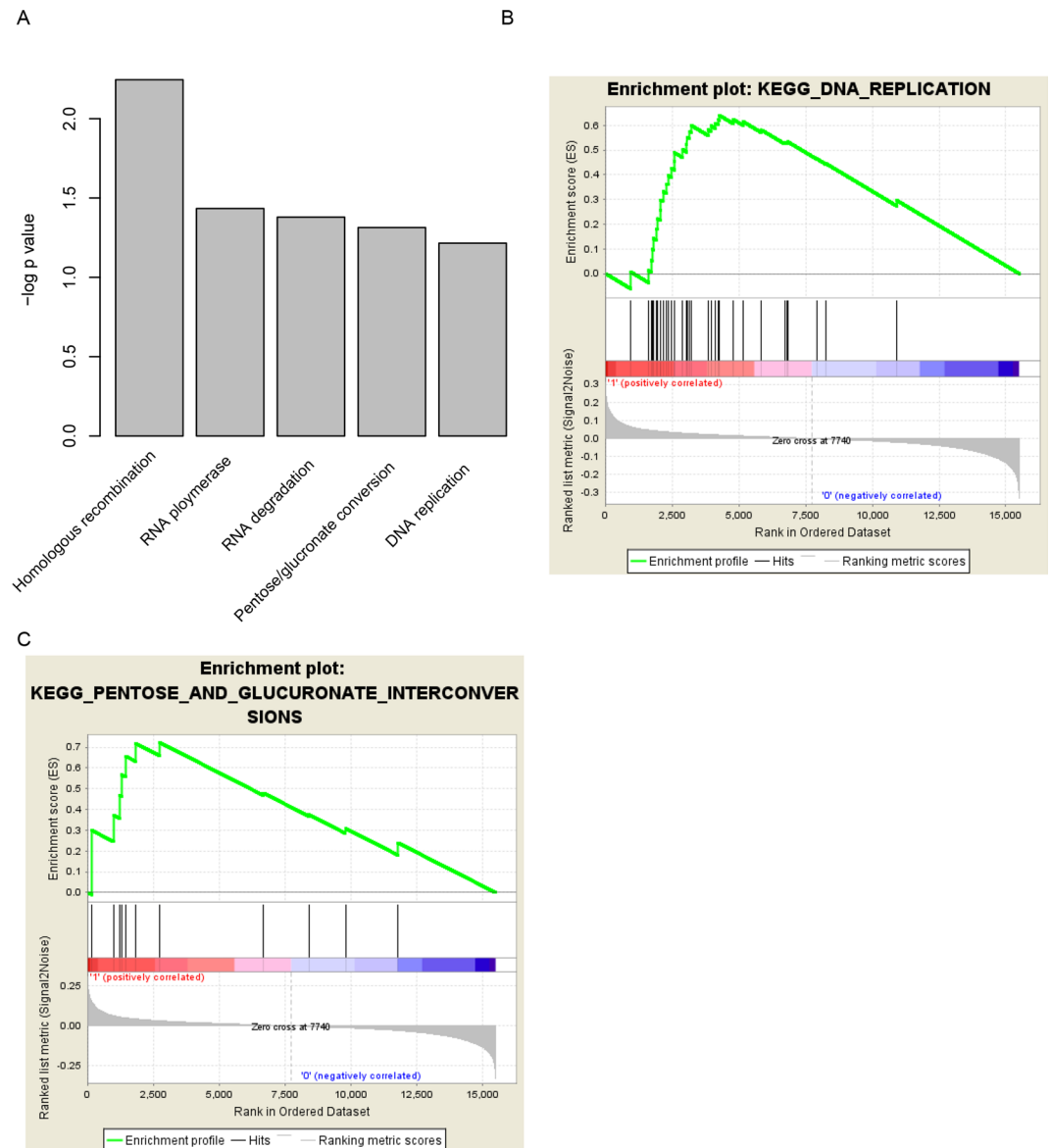


Figure 6. KEGG pathways associated with risk score. The pathways associated with risk score (A), include DNA replication (B) and pentose/glucuronate conversion (C).

prognosis and therapy of head and neck squamous cell carcinoma difficult. In the past, the prognosis of HNSC is largely depend on the TNM staging system and histologic grade³⁸. However, the effect is unfavorable.

The clinical prognostic effect of mRNAs in HNSC have been widely reported in the past decades. Recently, the utilization of lncRNAs as prognostic marker and potential therapy targets has been reported³⁹, including head and neck squamous cell carcinoma^{40,41}. The usage of lncRNAs for disease association have been reported in the past years^{7,9,37,39}, and achievements have been obtained. However, the combination of lncRNAs and mRNAs is currently not reported for prognosis, to our knowledge. In current work, we developed a cox multivariate regression model with survival-related genes and random forest variable selection. Afterwards, the performance of the seven lncRNA-mRNA based risk score were assessed in the training dataset and validated in another two datasets, and further validated in a totally independent dataset. Patients in the high risk group significantly have a shorter survival than the low risk group. And more, the seven lncRNA-mRNA based risk score is independent from the clinicopathological observations and performs better than this information.

The pathways significantly associated with the risk score were “RNA polymerase” and “DNA replication”, suggesting that our risk score may reflects the basic status of HNSC cells, including DNA replication and basic metabolism.

References

1. Parkin, D. M., Bray, F., Ferlay, J. & Pisani, P. Global cancer statistics, 2002. *CA: a cancer journal for clinicians* **55**, 74–108 (2005).
2. Jemal, A., Siegel, R., Xu, J. & Ward, E. Cancer statistics, 2010. *CA: a cancer journal for clinicians* **60**, 277–300, doi:10.3322/caac.20073 (2010).

3. Katiyar, S. K. Emerging Phytochemicals for the Prevention and Treatment of Head and Neck Cancer. *Molecules (Basel, Switzerland)* **21**, doi:10.3390/molecules21121610 (2016).
4. de Andrade, D. A. & Machiels, J. P. Treatment options for patients with recurrent or metastatic squamous cell carcinoma of the head and neck, who progress after platinum-based chemotherapy. *Current opinion in oncology* **24**, 211–217, doi:10.1097/CCO.0b013e3283510773 (2012).
5. Pfister, D. G. *et al.* Head and neck cancers. *Journal of the National Comprehensive Cancer Network: JNCCN* **9**, 596–650 (2011).
6. Jiang, C., Li, X., Zhao, H. & Liu, H. Long non-coding RNAs: potential new biomarkers for predicting tumor invasion and metastasis. *Molecular cancer* **15**, 62, doi:10.1186/s12943-016-0545-z (2016).
7. Chen, X., You, Z. H., Yan, G. Y. & Gong, D. W. IRWRLDA: improved random walk with restart for lncRNA-disease association prediction. *Oncotarget* **7**, 57919–57931, doi:10.18632/oncotarget.11141 (2016).
8. Frank, S., Aguirre, A., Hescheler, J. & Kurian, L. A lncRNA Perspective into (Re)Building the Heart. *Frontiers in cell and developmental biology* **4**, 128, doi:10.3389/fcell.2016.00128 (2016).
9. Chen, X. & Yan, G. Y. Novel human lncRNA-disease association inference based on lncRNA expression profiles. *Bioinformatics (Oxford, England)* **29**, 2617–2624, doi:10.1093/bioinformatics/btt426 (2013).
10. Meller, V. H., Joshi, S. S. & Deshpande, N. Modulation of Chromatin by Noncoding RNA. *Annual review of genetics* **49**, 673–695, doi:10.1146/annurev-genet-112414-055205 (2015).
11. Deniz, E. & Erman, B. Long noncoding RNA (lincRNA), a new paradigm in gene expression control. *Functional & integrative genomics*, doi:10.1007/s10142-016-0524-x (2016).
12. Song, W. & Zou, S. B. Prognostic role of lncRNA HOTAIR in esophageal squamous cell carcinoma. *Clinica chimica acta; international journal of clinical chemistry* **463**, 169–173, doi:10.1016/j.cca.2016.10.035 (2016).
13. Chen, Z., He, A., Wang, D., Liu, Y. & Huang, W. -Long noncoding RNA HOTTIP as a novel predictor of lymph node metastasis and survival in human cancer: a systematic review and meta-analysis. *Oncotarget*, doi:10.18632/oncotarget.12981 (2016).
14. He, A. *et al.* Role of long noncoding RNA UCA1 as a common molecular marker for lymph node metastasis and prognosis in various cancers: a meta-analysis. *Oncotarget*, doi:10.18632/oncotarget.12463 (2016).
15. Chen, J. *et al.* Long Non-coding RNAs in Urologic Malignancies: Functional Roles and Clinical Translation. *Journal of Cancer* **7**, 1842–1855, doi:10.7150/jca.15876 (2016).
16. Chen, X., Yan, C. C., Zhang, X. & You, Z. H. Long non-coding RNAs and complex diseases: from experimental results to computational models. *Briefings in bioinformatics*, doi:10.1093/bib/bbw060 (2016).
17. Nohata, N., Abba, M. C. & Gutkind, J. S. Unraveling the oral cancer lncRNAome: Identification of novel lncRNAs associated with malignant progression and HPV infection. *Oral oncology* **59**, 58–66, doi:10.1016/j.oraloncology.2016.05.014 (2016).
18. Denaro, N., Merlano, M. C., Russi, E. G. & Lo Nigro, C. Non coding RNAs in head and neck squamous cell carcinoma (HNSCC): a clinical perspective. *Anticancer research* **34**, 6887–6896 (2014).
19. Tian, X. & Xu, G. Clinical value of lncRNA MALAT1 as a prognostic marker in human cancer: systematic review and meta-analysis. *BMJ open* **5**, e008653, doi:10.1136/bmjopen-2015-008653 (2015).
20. Jadalaha, M. *et al.* Functional and prognostic significance of long non-coding RNA MALAT1 as a metastasis driver in ER negative lymph node negative breast cancer. *Oncotarget* **7**, 40418–40436, doi:10.18632/oncotarget.9622 (2016).
21. Huang, C., Yu, Z., Yang, H. & Lin, Y. Increased MALAT1 expression predicts poor prognosis in esophageal cancer patients. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie* **83**, 8–13, doi:10.1016/j.biopha.2016.05.044 (2016).
22. Yuan, S. X. *et al.* Long noncoding RNA DANCR increases stemness features of hepatocellular carcinoma by derepression of CTNBN1. *Hepatology (Baltimore, Md.)* **63**, 499–511, doi:10.1002/hep.27893 (2016).
23. Terashima, M., Tange, S., Ishimura, A. & Suzuki, T. MEG3 long noncoding RNA contributes to the epigenetic regulation of epithelial-mesenchymal transition in lung cancer cell lines. *The Journal of biological chemistry*, doi:10.1074/jbc.M116.750950 (2016).
24. Tao, R. *et al.* Association between indel polymorphism in the promoter region of lncRNA GAS5 and the risk of hepatocellular carcinoma. *Carcinogenesis* **36**, 1136–1143, doi:10.1093/carcin/bgv099 (2015).
25. Guo, Q. *et al.* Comprehensive analysis of lncRNA-mRNA co-expression patterns identifies immune-associated lncRNA biomarkers in ovarian cancer malignant progression. *Scientific reports* **5**, 17683, doi:10.1038/srep17683 (2015).
26. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* **43**, e47, doi:10.1093/nar/gkv007 (2015).
27. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**, R25, doi:10.1186/gb-2009-10-3-r25 (2009).
28. Cox, D. R. Regression Models and Life-Tables. *Journal of the Royal Statistical Society* **34**, 527–541 (1972).
29. Ishwaran, H. *et al.* Random survival forests for competing risks. *Biostatistics (Oxford, England)* **15**, 757–773, doi:10.1093/biostatistics/kxu010 (2014).
30. Moreno-Betancur, M., Sadaoui, H., Piffaretti, C. & Rey, G. Survival analysis with multiple causes of death: Extending the competing risks model. *Epidemiology (Cambridge, Mass.)*, doi:10.1097/ede.0000000000000531 (2016).
31. Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC bioinformatics* **12**, 77, doi:10.1186/1471-2105-12-77 (2011).
32. Miller, M. E., Hui, S. L. & Tierney, W. M. Validation techniques for logistic regression models. *Statistics in medicine* **10**, 1213–1226 (1991).
33. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15545–15550, doi:10.1073/pnas.0506580102 (2005).
34. Li, J. *et al.* Identification of high-quality cancer prognostic markers and metastasis network modules. *Nature communications* **1**, 34, doi:10.1038/ncomms1033 (2010).
35. Sawabe, M. *et al.* Heterogeneous impact of drinking according to treatment method on survival in head and neck cancer: a prospective study. *Cancer science*, doi:10.1111/cas.13115 (2016).
36. Ausoni, S. *et al.* Targeting cellular and molecular drivers of head and neck squamous cell carcinoma: current options and emerging perspectives. *Cancer metastasis reviews* **35**, 413–426, doi:10.1007/s10555-016-9625-1 (2016).
37. Chen, X. *et al.* HGIMDA: Heterogeneous graph inference for miRNA-disease association prediction. *Oncotarget* **7**, 65257–65269, doi:10.18632/oncotarget.11251 (2016).
38. Alsaffar, H. A. *et al.* Correlation between clinical and MRI assessment of depth of invasion in oral tongue squamous cell carcinoma. *Journal of otolaryngology - head & neck surgery = Le Journal d'oto-rhino-laryngologie et de chirurgie cervico-faciale* **45**, 61, doi:10.1186/s40463-016-0172-0 (2016).
39. Chen, X., Huang, Y. A., Wang, X. S., You, Z. H. & Chan, K. C. FMLNCSIM: fuzzy measure-based lncRNA functional similarity calculation model. *Oncotarget* **7**, 45948–45958, doi:10.18632/oncotarget.10008 (2016).
40. Schmitt, A. M. & Chang, H. Y. Long Noncoding RNAs in Cancer Pathways. *Cancer cell* **29**, 452–463, doi:10.1016/j.ccell.2016.03.010 (2016).
41. Weng, M. *et al.* Noncoding RNAs in the development, diagnosis, and prognosis of colorectal cancer. *Translational research: the journal of laboratory and clinical medicine*, doi:10.1016/j.trsl.2016.10.001 (2016).

Acknowledgements

Zhejiang provincial health and Family Planning Commission, Grant No. 2015115433 to Zhi-Li Zhang; Zhejiang province administration of traditional Chinese medicine, Grant No. 2015ZA051 to Zhi-Li Zhang; Zhejiang provincial health and Family Planning Commission, Grant No. 201493689 to Zhi-Li Zhang.

Author Contributions

Experiment design: Z.Z., L.Z., L.C., S.Z. Writing: Z.Z., F.W., Y.W., Y.X., P.Z. Data interpretation: Z.Z., S.Z., F.W., P.Z.

Additional Information

Supplementary information accompanies this paper at doi:[10.1038/s41598-017-00252-2](https://doi.org/10.1038/s41598-017-00252-2)

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017