A novel long non-coding RNA, AC012456.4, as a valuable and independent prognostic biomarker of survival in oral squamous cell carcinoma

Xuegang Hu [1, 2] , **Zailing Qiu** [1, 2] , **Jianchai Zeng** [1, 2] , **Tingting Xiao** [1, 2] , **Zhihong Ke** [1, 2] , **Hongbing Lyu** Corresp. [1]

[1] Department of Endodontics and Operative Dentistry, School and Hospital of Stomatology, Fujian Medical University, Fuzhou, China

[2] Key laboratory of Stomatology, Fujian Province University, Fuzhou, China

Corresponding Author: Hongbing Lyu
Email address: hongbinglu@fjmu.edu.cn

Oral squamous cell carcinoma (OSCC) is a major malignant cancer of the head and neck. Long non-coding RNAs (lncRNAs) have emerged as critical regulators during the development and progression of cancers. This study aimed to identify a lncRNA-related signature with prognostic value for evaluating survival outcomes and to explore the underlying molecular mechanisms of OSCC. Associations between overall survival (OS), disease-free survival (DFS) and candidate lncRNAs were evaluated by Kaplan-Meier survival analysis and univariate and multivariate Cox proportional hazards regression analyses. The robustness of the prognostic significance was shown via the Gene Expression Omnibus (GEO) database. A total of 2493 lncRNAs were differentially expressed between OSCC and control samples (fold change>2, p <0.05). We used Kaplan-Meier survival analysis to identify 21 lncRNAs for which the expression levels were associated with OS and DFS of OSCC patients (p <0.05) and found that down-expression of lncRNA AC012456.4 especially contributed to poor DFS (p =0.00828) and OS (p =0.00987). Furthermore, decreased expression of AC012456.4 was identified as an independent prognostic risk factor through multivariate Cox proportional hazards regression analyses [DFS: p =0.004, hazard ratio (HR)=0.600, 95%confidence interval(CI)=0.423-0.851; OS: p =0.002, HR=0.672, 95%CI=0.523-0.863]. Gene Set Enrichment Analysis (GSEA) indicated that lncRNA AC012456.4 were significantly enriched in critical biological functions and pathways and was correlated with tumorigenesis, such as regulation of cell activation, and the JAK-STAT and MAPK signal pathway. Overall, these findings were the first to evidence that AC012456.4 may be an important novel molecular target with great clinical value as a diagnostic, therapeutic and prognostic biomarker for OSCC patients.

A novel long non-coding RNA, AC012456.4, as a valuable and independent prognostic biomarker of survival in oral squamous cell carcinoma

Xuegang Hu[1,2], Zailing Qiu[1,2], Jianchai Zeng [1,2], Tingting Xiao [1,2], Zhihong Ke[1,2] , Hongbing Lyu[1]

[1] Department of Endodontics and Operative Dentistry, School and Hospital of Stomatology, Fujian Medical University, Fuzhou, Fujian, China

[2] Key laboratory of Stomatology, Fujian Province University, Fuzhou, Fujian, China

Corresponding Author:

Hongbing Lyu[1]

246 Yangqiao Zhong Road, Fuzhou, Fujian, 350002, China

Email address: hongbinglu@fjmu.edu.cn

Abstract:

Oral squamous cell carcinoma (OSCC) is a major malignant cancer of the head and neck. Long non-coding RNAs (lncRNAs) have emerged as critical regulators during the development and progression of cancers. This study aimed to identify a lncRNA-related signature with prognostic value for evaluating survival outcomes and to explore the underlying molecular mechanisms of OSCC. Associations between overall survival (OS), disease-free survival (DFS) and candidate lncRNAs were evaluated by Kaplan-Meier survival analysis and univariate and multivariate Cox proportional hazards regression analyses. The robustness of the prognostic significance was shown via the Gene Expression Omnibus (GEO) database. A total of 2493 lncRNAs were differentially expressed between OSCC and control samples (fold change>2, p<0.05). We used Kaplan-Meier survival analysis to identify 21 lncRNAs for which the expression levels were associated with OS and DFS of OSCC patients (p<0.05) and found that down-expression of lncRNA AC012456.4 especially contributed to poor DFS (p=0.00828) and OS (p=0.00987). Furthermore, decreased expression of AC012456.4 was identified as an independent prognostic risk factor through multivariate Cox proportional hazards regression analyses [DFS: p=0.004, hazard ratio (HR)=0.600, 95%confidence interval(CI)=0.423-0.851; OS: p=0.002, HR=0.672, 95%CI=0.523-0.863]. Gene Set Enrichment Analysis (GSEA) indicated that lncRNA AC012456.4 were significantly enriched in critical biological functions and pathways and was correlated with tumorigenesis, such as regulation of cell activation, and the JAK-STAT and MAPK signal pathway. Overall, these findings were the first to evidence that AC012456.4 may be an important novel molecular target with great clinical value as a diagnostic, therapeutic and prognostic biomarker for OSCC patients.

1. Introduction

The 5-year survival rate is approximately 50% for oral squamous cell carcinoma (OSCC), which one of the most common malignancies of the head and neck region (Bozec et al. 2009; Ferlay et al. 2015; Kamangar et al. 2006; Kim et al. 2017; Verusingam et al. 2017).The predisposition of OSCC to distant metastases and metastases in the lymph nodes, its highly invasive nature, and its tendency towards local recurrence are important factors that contribute to the poor prognosis of OSCC patients(Massano et al. 2006; Singh & Schenberg 2013). Hence, more effective novel tumor diagnostic and prognostic biomarkers(Mehrotra & Gupta 2011),

which can improve the survival rate and can be used to assess treatment outcomes are urgently needed.

The Cancer Genome Atlas (TCGA) (http://cancergenome.nih.gov) database, which is primarily used to collate specimens from cancer patients and adjacent normal tissue specimens, contains large data sets collected with high-throughput methods at multiple genomic and proteomic levels(Chin et al. 2011; Wang et al. 2009). The Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) is the largest and most comprehensive public gene expression repository for high-throughput data at NCBI (Barrett & Edgar 2006; Clough & Barrett 2016). Both the GEO and TCGA collect macroscopic clinical information, such as stage and grade of tumor, survival time, age, sex, and race. Therefore, the TCGA and GEO databases can be analyzed systematically and comprehensively to explore important potential value and information.

In this study, we first sought to use the existing GEO microarrays and TCGA RNA-seq data to identify differential expression of lncRNAs between OSCC and control tissue samples. Then, the differentially expressed lncRNAs were evaluated by Kaplan-Meier survival analysis and univariate, multivariate Cox proportional hazards regression analyses and Gene Set Enrichment Analysis (GSEA). Ultimately, through systematic and objective analysis, we first discovered that lncRNA AC012456.4 is significantly associated with survival outcomes of OSCC patients based on TCGA data. Then, AC012456.4 was further successfully confirmed as a potential prognostic biomarker for the prediction of overall survival (OS) in the GEO database. We hope that the lncRNA AC012456.4 revealed in our study may serve as a novel biomarkers and potential targets for the diagnosis, treatment, and prognosis of OSCC.

2. Materials and Methods

2.1. Data source

The RNA-seq data and corresponding patient information data of head and neck cell carcinoma (HNSC) were downloaded from the TCGA database. Clinical samples from the oral cavity (buccal mucosa, tongue, lip, hard palate, alveolar ridge, floor of the mouth and oral cavity) were chosen, while some samples from other parts (hypopharynx, larynx, oropharynx and tonsil, for example) were excluded. The original microarray data between OSCC and adjacent normal tissue samples were downloaded from the NCBI GEO databases. The accession numbers were GSE36820 and GSE41613, respectively. The microarray data of GSE36820 and GSE41613 were based on GPL570 (Affymetrix Human Genome U133 Plus 2.0 Array).

2.2. Data pre-processing and differential expression analysis

The edgeR package was downloaded from the Stanford University website. The original microarray data from the GEO were converted into expression measures using the affy R package. Then, the differentially expressed lncRNAs were identified by Limma R package (Ritchie et al. 2015; Teufel et al. 2016). The differentially expressed lncRNAs that were screened from the TCGA were analyzed by the edgeR package (Robinson et al. 2010). To improve screen accuracy and simplify the screening process, the cut-off criteria, which was in accordance with the procedure of Benjamini & Hochberg (BH), was as follows: 1. the false discovery rate was

controlled at 0.01; 2. the fold change should be more than 2. The differentially expressed lncRNAs among GSE36820, GSE41613 and the TCGA were identified by the intersect function in R package. Tumor and normal tissue data were recorded and were statistically analyzed.

2.3. Identification of lncRNAs with prognostic value in OSCC

The differences between expressed lncRNAs (fold change > 2, $p < 0.05$) are involved in the prognostic value for OSCC. The OSCC patients were divided into two parts, depending on the average expression level of candidate lncRNAs: a high expression group and a low expression group. Survival differences and p-values were compared between the two groups and were evaluated using a Kaplan-Meier survival analysis and a log-rank test. After this, a univariate Cox proportional hazards regression analysis(Bair & Tibshirani 2004) was conducted to assess the correlation between candidate lncRNAs and patient overall survival (OS) and disease-free survival (DFS) ($p < 0.05$). Statistically significant lncRNAs and clinical candidate predictors were further evaluated by multivariate Cox proportional hazards regression analyses to identify independent prognostic lncRNAs. Candidate predictors included age, gender, grade, and stage. We then performed subgroup analyses. The hazard ratio (HR) and 95% confidence interval (CI) were also assessed.

2.4. Gene set enrichment analysis (GSEA)

GSEA 2-2.2.3 (JAVA version) was downloaded from the Gene Set Enrichment Analysis website (http://software.broadinstitute.org/gsea/index.jsp). Then,the downloaded dataset was imported using the GSEA software. Gene sets identified as related to biological signal conduction on the MSigDB (Molecular Signatures Database) (http://software.broadinstitute.org/gsea/msigdb), which may be found on the GSEA website, served as reference gene sets. This process was repeated 1000 times for each analysis according to the default weighted enrichment statistical method. Gene sets with a false discovery rate (FDR) < 0.25 and a family-wise error rate (FWR) < 0.05. The GSEA analysis includes four key statistics: Enrichment Score (ES), Normalized Enrichment Score (NES), False Discovery Rate (FDR) and P-value.

2.5. Statistical analysis

In this study, all analyses, including the t-test, heat map, and survival analyses, were performed with the R, GraphPad and SPSS software packages. p values less than 0.05 was considered significant. All statistical tests were two-sided.

3. Results

3.1. Characteristics of OSCC patients according to the TCGA

In this study, the datasets of 350 OSCC patient and 44 controls were acquired and downloaded from the TCGA (http://cancergenome.nih.gov) database; these datasets contained expression data and clinical information related to 14448 lncRNAs. The clinicopathological

features of all patients are shown in **Table 1**. The mean ± standard deviation (STDEV) for all patient ages is 61.590 ± 12.886.

3.2. Significant differentially expressed lncRNAs in OSCC

In all, 2493 differentially expressed lncRNAs were identified through analysis of 14448 lncRNAs using the edgeR packages (fold change > 2, $p < 0.05$) (**Fig 1**). Moreover, 855 lncRNAs were down-regulated and 1638 lncRNAs were up-regulated in the OSCC samples compared to normal tissue. Down-regulated and up-regulated lncRNAs account for 34.2% and 65.6% of the differentially expressed lncRNAs, respectively.

3.3. Identification of survival differences lncRNAs in OSCC

We used a Kaplan-Meier survival analysis with the log-rank test to identify relationships between the above 2493 lncRNA signatures and the survival of OSCC patients. Then, we determined the levels of 21 lncRNA signatures that were significantly related to OS and DFS. Among these 21 lncRNAs, a significant positive correlation was observed between the signatures of 13 lncRNAs (TTC39A-AS1, RP11-93B14.9, AC012456.4, RP11-87C12.5, RP11-464F9.21, LINC01549, RP11-897M7.1, AP003900.6, LINC01343, RP11-181E10.3, CTD-2545H1.2, RP11-796E2.4 and LINC01108) and OS/DFS. In contrast, the signatures of the remaining 8 lncRNAs (AC007879.2, BOK-AS1, CTB-161M19.4, CTD-2033A16.3, FAM95B1, RP11-1C8.7, RP11-285G1.14 and RP11-286E11.1) were significantly negatively correlated with OS and DFS. That is, low expression of the 13 lncRNAs described above correlated with a poor prognosis of OSCC patients, while the up-regulation of the latter 8 lncRNAs correlated with a shorter survival time (**Fig 2**) (**Table 2**)

Through the above Kaplan-Meier survival analysis, the variables of age, gender, grade, tumor stage, and TNM stage were identified as statistically significant factors that are related to the above 21 lncRNAs and patient prognosis. We also applied univariate and multivariate Cox regression analyses to evaluate the ability of 21 candidate lncRNA signatures to serve as independent prognostic variables. The univariate analysis indicated that decreased AC012456.4 expression (HR = 0.706, 95% CI: 0.551-0.903, $p = 0.006$), age, tumor stage, and TNM stage were all significantly related to worse OS in OSCC patients (**Table 3**). Decreased AC012456.4 expression (HR = 0.601, 95% CI: 0.423-0.853, $p = 0.004$) was the only variable that could predict poorer DFS for OSCC. Finally, multivariate Cox regression analysis revealed that low expression of AC012456.4 was the only independent prognostic variable for both OS (HR = 0.672, 95% CI: 0.523-0.863, $p = 0.002$) and DFS (HR = 0.600, 95% CI: 0.423-0.851, $p = 0.004$) in OSCC patients (**Table 4**). In addition, age and N stage were highly significantly correlated with shorter OS or DFS.

3.4. lncRNA AC012456.4 was low expressed in OSCC tissues and associated with clinicopathological parameters

OSCC patients were further classified into high or low expression groups based on the

median value of the relative lncRNA expression. The expression of lncRNA AC012456.4 was significantly weaker in OSCC tissue samples (1.360 ± 0.05569) relative to normal tissue samples (3.062 ± 0.2304) in the TCGA (p <0.0001) (**Fig 3**). The correlation between lncRNA AC012456.4 expression and clinicopathologic parameters of OSCC patients was also further analyzed. As shown in Table 5, lncRNA AC012456.4 expression was significantly correlated with alcohol history consumption (p = 0.033). Additionally, decreased expression of lncRNA AC012456.4 expression nearly significantly associated with T stage (p = 0.075). However, no significant association was found between other clinicopathological factors and lncRNA AC012456.4 expression.

3.5. Evaluation of the prognostic value of lncRNA AC012456.4 via the GEO

For the purpose of evaluating the robustness of lncRNA AC012456.4 expression in th prediction of OS of OSCC patients, we acquired another independent datasets from the GEO with accession numbers of GSE36820 and GSE41613, which contained OSCC samples, but samples with incomplete clinical information were excluded. The prognostic signatures and the Kaplan-Meier analysis were calculated and performed for each OSCC sample. In agreement with the result of the TCGA datasets, low expression levels of lncRNA AC012456.4 were associated with lower OS (**Fig 4**). The lncRNA AC012456.4 was also expressed at low levels in OSCC tissues (p < 0.0001).

3.6. Relationship between lncRNA AC012456.4 and biological pathways and functions

Biological pathways and functions of lncRNA AC012456.4 were identified by GSEA. This analysis revealed that lncRNA AC012456.4 was involved in many critical pathways and correlated with tumorigenesis. A total of 150 pathways listed in the high-risk group were enriched, including KEGG MAPK SIGNALING PATHWAY, KEGG JAK-STAT SIGNALING PATHWAY, KEGG CALCIUM SIGNALING PATHWAY and KEGG PATHWAYS IN CANCER. Twenty-seven pathways in the low-risk group were also identified, including the KEGG OXIDATIVE PHOSPHORYLATION, KEGG PROTEASOME and KEGG SPLICEOSOME (**Fig 5**). Similarly, 3073 GO annotations in the high-risk group and 516 GO annotations in the low-risk group were enriched (**Fig 6**). Relevant partial results for KEGG pathways and GO analysis are listed in **Table 6** and **Table 7**.

4. Discussion

OSCC is a common, highly invasive type of oral cancer prone to early recurrence and metastasis (Massano et al. 2006; Singh & Schenberg 2013). Therefore, early diagnosis and treatment of OSCC is essential (Bozec et al. 2009). While cytology- and pathology-based methods have been applied to the clinical differential diagnosis of OSCC, limitations in the detection methods and poor prognoses have limited the 5-year survival rat (Omar 2013). Hence, more reliable, accurate and sensitive prognosis biomarkers and tools for early diagnosis are urgently needed(Mehrotra & Gupta 2011). In recent years, many studies have revealed a close

association between aberrant expression of lncRNAs and tumorigenesis (A & I 2014; Batista & Chang 2013; JM 2017; Rinn & Chang 2012; Slaby et al. 2017), which may aid in cancer diagnosis and prognosis.

Fewer than 2% of genes in the human genome are transcribed, and up to 98% of these transcripts are non-coding RNAs (Jandura & Krause 2017; JT 2013; Quinn & Chang 2016). lncRNAs are a class of non-coding transcripts ≥ 200 nucleotides in length that are actively involved in many biological processes, such as epigenetic regulation, cell cycle regulation, chromatin modulation and regulation of multiple gene expression (Rinn & Chang 2012; Wang et al. 2017). These non-coding transcripts also play key roles in the occurrence, development and progression of malignant tumors (JM 2017; Kopp & Mendell 2018; Spizzo et al. 2012). An increasing number of studies have reported that lncRNAs can play essential roles as oncogenes or tumor suppressor genes involved in the development and progression of various cancers (Batista & Chang 2013; JM 2017; Kopp & Mendell 2018; Reik 2009; Rinn & Chang 2012; Slaby et al. 2017; Spizzo et al. 2012), including OSCC (Fang et al. 2017; Gomes et al. 2017; Guo et al. 2017b; Li et al. 2017). For example, the down-regulation of HOTAIR is associated with cancer progression in 26 human tumor types (Bhan & Mandal 2015).

However, most early studies focused on a single gene or the results obtained from a single cohort study of lncRNAs and OSCC. Sun et al.(Sun et al. 2017) used qRT-PCR to analyze the expression levels of lncRNA PDIA3P in 58 OSCC and paired noncancerous tissue samples. This study found that the overexpression of lncRNA PDIA3P correlated with lower survival rates for OSCC patients. One study by Wu et al.(Wu et al. 2015) suggested that high expression of lncRNA HOTAIR in OSCC patients would contribute to the development and progression of cancer, leading to a poor prognosis. Similarly, LINC00668 expression is increased in both 50 OSCC tissues and cells, and over-expression is significantly correlated with poorer survival for OSCC patients; Therefore, this might be a negative predictive factor for the prognosis of OSCC patients(Zhang 2017). In the era of big data, the development of TCGA and GEO technology has allowed researchers to predict and identify new biomarkers, which has enhanced the reliability and accuracy of current research. Cui et al.(Cui et al. 2017) used TCGA and GEO data to determine that the expression levels of several lncRNAs, including RP1-228H13.5, TMCC1-AS1, LINC00205, and RP11-307C12.11, were associated with OS and recurrence-free survival of hepatocellular carcinoma patients. Three lncRNAs (LINC01140, TGFB2-OT1, and RP11-347C12.10) were significantly correlated with prognoses of hepatocellular carcinoma patients, independent of some clinical characteristics. Using the database, three lncRNAs, which may play key roles in the development, progression, and recurrence in gastric cancer, were identified (Song et al. 2017). However, the functions, roles, and molecular mechanisms of lncRNAs associated with OSCC remain unclear.

In this study, we identified lncRNAs that are dysregulated in OSCC and evaluated the relationships between the TCGA database and the clinicopathological features of these OSCC patients. Based on the above analysis, a total of 21 lncRNAs were correlated with patient prognoses, of which thirteen lncRNAs (TTC39A-AS1, RP11-93B14.9, AC012456.4, RP11-87C12.5, RP11-464F9.21, LINC01549, RP11-897M7.1, AP003900.6, LINC01343, RP11-181E10.3, CTD-2545H1.2, RP11-796E2.4 and LINC01108) were significantly positively

associated with OS and DFS, while the up-regulation of the latter 8 lncRNAs (AC007879.2, BOK-AS1, CTB-161M19.4, CTD-2033A16.3, FAM95B1, RP11-1C8.7, RP11-285G1.14 and RP11-286E11.1) were correlated with poorer prognoses. Lan et al. (Lan et al. 2017) have also reported that RP11-1C8.7 predicted the progression and outcome of patients with kidney renal papillary cell carcinoma and was regarded as an independent prognostication factor for kidney renal papillary cell carcinoma. Thus far in the published literature, no report has evaluated the biological function and molecular mechanisms of other lncRNAs associated with human cancers.

To our knowledge, this study is pioneering research and identified the lncRNA AC012456.4, which exhibited significantly lower expression in OSCC tissues than in adjacent normal tissues. Additionally, a Kaplan-Meier survival analysis (Gyorffy et al. 2012) as well as univariate and multivariate Cox regression analyses revealed that lncRNA AC012456.4 was an independent prognostic factor and was significantly correlated with shorter OS and DFS. Further validation via the GEO database was consistent with the TCGA database analysis results. Moreover, we further evaluated the relationship between AC012456.4 expression and the clinicopathological features of OSCC patients. Low levels of AC012456.4 were found to be significantly associated with the history of alcohol consumption in OSCC patients. Interestingly, according to previous studies, we found that alcohol consumption can increase the probability of G:C to A:T transitions and that alcohol drinkers exhibited a significantly higher incidence of p53 mutations in OSCC (Hsieh et al. 2001), which suggested that alcohol may play a critical role in the progression of OSCC.

Since lncRNAs perform their biological function by specifically binding to target genes, we further explored the possible biological functions and molecular pathways of AC012456.4. Through GSEA, AC012456.4 was found to be significantly involved with tumor-related signaling pathways and crucial biological functions in tumorigenesis. Key pathways and functions for tumor initiation and progression were identified, such as GO biological function annotation and KEGG pathways, including the adaptive immune response, RRNA metabolic processes, CALCIUM, MAPK, and the JAK/STAT signaling pathway. Additionally, mutation, aberrant expression and modification of these GO annotations and signaling pathways have been frequently reported in OSCC and other cancers. We found that the MAPK pathway could be activated by the low expression of the tumor suppressor QKI-5, which can promote the proliferation of OSCC cells (Fu & Feng 2015). We also revealed the strong relationships between HOXC10 and gastric cancer cell proliferation and metastasis, which occur through the MAPK pathway (Guo et al. 2017a). Other pathways and biological functions have also been reported in pancreatic ductal adenocarcinoma (Huang et al. 2017a), hepatocellular carcinoma (Huang et al. 2017b; Wonganan et al. 2017), and human papillomavirus-transformed tumors (Skeate et al. 2018).

Dysregulated expression of lncRNA signatures has tremendous potential value, but this research has limitations. Above all, we have explored the correlation between AC012456.4 expression and OSCC prognosis based on the TCGA and GEO databases, which signifies that the exploration was performed using a bioinformatics approach. Then, further research, such as quantitative real-time PCR, as well as in vivo and in vitro experiments, will require collaborative

efforts to explore the potential molecular functions and related mechanisms of these lncRNAs in OSCC.

5. Conclusions

In summary, this study was the first to discover that lncRNA AC012456.4 was poorly expressed in OSCC, with decreased survival rates for OSCC patients. This may be a potential novel, independent biomarker and therapeutic target for the early diagnosis, pathological classification, clinical treatment and outcome prediction for OSCC. Nevertheless, these assumptions require validation and confirmation by larger, multicenter studies.

Acknowledgments: We thank the patients and investigators who participated in TCGA Research Network (http://cancergenome.nih.gov/), which provides a Web resource for exploring, visualizing, and analyzing multidimensional cancer genomics data.

Abbreviations

LncRNAs	long non-coding RNAs
OSCC	Oral squamous cell carcinoma
HR	hazard ratio
CI	confidence interval
DFS	disease-free survival
OS	overall survival
TCGA	The Cancer Genome Atlas
GEO	Gene Expression Omnibus
GSEA	Gene Set Enrichment Analysis
KEGG	the Kyoto Encyclopedia of Genes and Genomes
GO	Gene Ontology
STDEV	standard deviation

References

A F, and I B. 2014. Long non-coding RNAs: new players in cell differentiation and development. Nat Rev Genet 15:7-21. DOI: 10.1038/nrg3606.

Bair E, and Tibshirani R. 2004. Semi-supervised methods to predict patient survival from gene expression data. PLoS Biol 2:E108. DOI: 10.1371/journal.pbio.0020108.

Barrett T, and Edgar R. 2006. Mining microarray data at NCBI's Gene Expression Omnibus (GEO)*. Methods Mol Biol 338:175-190. DOI: 10.1385/1-59745-097-9:175.

Batista P, and Chang H. 2013. Long noncoding RNAs: cellular address codes in development and disease. Cell 152:1298-1307. DOI: 10.1016/j.cell.2013.02.012.

Bhan A, and Mandal S. 2015. LncRNA HOTAIR: A master regulator of chromatin dynamics and cancer. Biochim Biophys Acta 1856:151-164. DOI: 10.1016/j.bbcan.2015.07.001.

Bozec A, Peyrade F, Fischel JL, and Milano G. 2009. Emerging molecular targeted therapies in the treatment of head and neck cancer. Expert Opin Emerg Drugs 14:299. DOI: 10.1517/14728210902997947.

Chin L, Andersen JN, and Futreal PA. 2011. Cancer genomics: from discovery science to personalized medicine. Nature Medicine 17:297-303. DOI: 10.1038/nm.2323.

Clough E, and Barrett T. 2016. The Gene Expression Omnibus Database. Methods Mol Biol 1418:93-110.

Cui H, Zhang Y, Zhang Q, Chen W, Zhao H, and Liang J. 2017. A comprehensive genome‐wide analysis of long noncoding RNA expression profile in hepatocellular carcinoma. Cancer Medicine 6:2932-2941. DOI: 10.1002/cam4.1180.

Fang Z, Zhao J, Xie W, Sun Q, Wang H, and Qiao B. 2017. LncRNA UCA1 promotes proliferation and cisplatin resistance of oral squamous cell carcinoma by sunppressing miR-184 expression. Cancer Med 6:2897-2908. DOI: 10.1002/cam4.1253.

Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin D, Forman D, and Bray F. 2015. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 136:E359-386. DOI: 10.1002/ijc.29210.

Fu X, and Feng Y. 2015. QKI-5 suppresses cyclin D1 expression and proliferation of oral squamous cell carcinoma cells via MAPK signalling pathway. Int J Oral Maxillofac Surg 44:562-567. DOI: 10.1016/j.ijom.2014.10.001

Gomes C, de Sousa S, Calin G, and Gomez R. 2017. The emerging role of long noncoding RNAs in oral cancer. Oral Surg Oral Med Oral Pathol Oral Radiol 123:235-241. DOI: 10.1016/j.oooo.2016.10.006.

Guo C, Hou J, Ao S, Deng X, and Lyu G. 2017. HOXC10 up-regulation promotes gastric

351 cancer cell proliferation and metastasis through MAPK pathway. Chin J Cancer Res 29:572-
352 580. DOI: 10.21147/j.issn.1000-9604

353 **Guo Y, Ma Y, Hu X, Song R, Zhu L, and Zhong M. 2018.** Long non-coding RNA CEBPA-
354 AS1 correlates with poor prognosis and promotes tumorigenesis via CEBPA/Bcl2 in oral
355 squamous cell carcinoma. Cancer Biol Ther 19:205-213. DOI: 10.1080/15384047.

356 **Gyorffy B, Lánczky A, and Szállási Z. 2012.** Implementing an online tool for genome-wide
357 validation of survival-associated biomarkers in ovarian-cancer using microarray data from
358 1287 patients. Endocr Relat Cancer 19:197-208. DOI: 10.1530/ERC-11-0329.

359 **Hsieh L, Wang P, Chen I, Liao C, Wang H, Chen M, Chang J, and Cheng A. 2001.**
360 Characteristics of mutations in the p53 gene in oral squamous cell carcinoma associated with
361 betel quid chewing and cigarette smoking in Taiwanese. Carcinogenesis 22:1497-1503. DOI:
362 10.1093/carcin/22.9.1497

363 **Huang L, Ma B, Ma J, and Wang F. 2017.** Fractalkine/CX3CR1 axis modulated the
364 development of pancreatic ductal adenocarcinoma via JAK/STAT signaling pathway.
365 Biochem Biophys Res Commun 493:1510-1517. DOI:10.1016/j.bbrc.2017.10.006.

366 **Huang Q, Cao H, Zhan L, Sun X, Wang G, Li J, Guo X, Ren T, Wang Z, Lyu Y, Liu B, An**
367 **J, and Xing J. 2017.** Mitochondrial fission forms a positive feedback loop with cytosolic
368 calcium signaling pathway to promote autophagy in hepatocellular carcinoma cells. Cancer
369 Lett 403:108-118. DOI：10.1016/j.canlet.2017.05.034.

370 **Jandura A, and Krause HM. 2017.** The New RNA World: Growing Evidence for Long
371 Noncoding RNA Functionality. Trends in Genetics Tig 33:665-676. DOI：
372 10.1016/j.tig.2017.08.002.

373 **JM E. 2017.** On the Origin of lncRNAs: Missing Link Found. Trends Genet 33:660. DOI:
374 10.1016/j.tig.2017.07.005.

375 **JT L. 2013.** Epigenetic regulation by long noncoding RNAs. Chromosome Research 21:685-693.
376 DOI: 10.1126/science.1231776.

377 **Kamangar F, Dores GM, and Anderson WF. 2006.** Patterns of cancer incidence, mortality,
378 and prevalence across five continents: defining priorities to reduce cancer disparities in
379 different geographic regions of the world. J Clin Oncol 24:2137-2150. DOI:
380 10.1200/JCO.2005.05.2308.

Kim H, Kang Y, Byun J, Jang S, Rho G, Lee J, and Park B. 2017. Midkine and NANOG Have Similar Immunohistochemical Expression Patterns and Contribute Equally to an Adverse Prognosis of Oral Squamous Cell Carcinoma. Int J Mol Sci 18:2339. DOI: 10.3390/ijms18112339.

Kopp F, and Mendell J. 2018. Functional Classification and Experimental Dissection of Long Noncoding RNAs. Cell 172:393-407. DOI: 10.1016/j.cell.2018.01.011.

Lan H, Zeng J, Chen G, and Huang H. 2017. Survival prediction of kidney renal papillary cell carcinoma by comprehensive LncRNA characterization. Oncotarget 8: 110811 – 110829. DOI: 10.18632/oncotarget.22732.

Li X, Ma C, Zhang L, Li N, Zhang X, He J, He R, Shao M, Wang J, Kang L, and Han C. 2017. LncRNAAC132217.4, a KLF8-regulated long non-coding RNA, facilitates oral squamous cell carcinoma metastasis by upregulating IGF2 expression. Cancer Lett 407:45-56. DOI: 10.1016/j.canlet.2017.08.007.

Maciejewski H. 2014. Gene set analysis methods: statistical models and methodological differences. Brief Bioinformatics 15:504-518. DOI: 10.1093/bib/bbt002

Massano J, Regateiro FS, Januário G, and Ferreira A. 2006. Oral squamous cell carcinoma: review of prognostic and predictive factors. Oral Surgery Oral Medicine Oral Pathology Oral Radiology & Endodontology 102:67-76. DOI: 10.1016/j.tripleo.2005.07.038.

Mehrotra R, and Gupta DK. 2011. Exciting new advances in oral cancer diagnosis: avenues to early detection. Head & Neck Oncology 3:1-9. DOI: 10.1186/1758-3284-3-33.

Omar EA. 2013. The Outline of Prognosis and New Advances in Diagnosis of Oral Squamous Cell Carcinoma (OSCC): Review of the Literature. Journal of Oral Oncology 2013:13. DOI: 10.1155/2013/519312.

Quinn J, and Chang H. 2016. Unique features of long non-coding RNA biogenesis and function. Nat Rev Genet 17:47-62. DOI: 10.1038/nrg.2015.10.

Reik W. 2009. Evolution and functions of long noncoding RNAs. Cell 136:629. DOI: 10.1016/j.cell.2009.02.006.

Rinn J, and Chang H. 2012. Genome regulation by long noncoding RNAs. Annu Rev Biochem 81:145-166. DOI: 10.1146/annurev-biochem-051410-092902.

Ritchie M, Phipson B, Wu D, Hu Y, Law C, Shi W, and Smyth G. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43:e47. DOI：10.1093/nar/gkv007.

Robinson M, McCarthy D, and Smyth G. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139-140. DOI: 10.1093/bioinformatics/btp616

Singh T, and Schenberg M. 2013. Delayed diagnosis of oral squamous cell carcinoma following dental treatment. Annals of the Royal College of Surgeons of England 215:369. DOI: 10.1308/003588413X13629960045599.

Skeate J, Da Silva D, Chavez-Juan E, Anand S, Nuccitelli R, and Kast W. 2018. Nano-Pulse Stimulation induces immunogenic cell death in human papillomavirus-transformed tumors and initiates an adaptive immune response. PLoS ONE 13:e0191311. DOI: 10.1371/journal.pone.0191311.

Slaby O, Laga R, and Sedlacek O. 2017. Therapeutic targeting of non-coding RNAs in cancer. Biochemical Journal 474:4219. DOI: 10.1042/BCJ20170079.

Song P, Jiang B, Liu Z, Ding J, Liu S, and Guan W. 2017. A three‐lncRNA expression signature associated with the prognosis of gastric cancer patients. Cancer Medicine 6:1154-1164. DOI: 10.1002/cam4.1047.

Spizzo R, Almeida M, Colombatti A, and Calin G. 2012. Long non-coding RNAs and cancer: a new frontier of translational research? Oncogene 31:4577. DOI: 10.1038/onc.2011.621.

Sun CC, Zhang L, Li G, Li SJ, Chen ZL, Fu YF, Gong FY, Bai T, Zhang DY, and Wu QM. 2017. The lncRNA PDIA3P Interacts with miR-185-5p to Modulate Oral Squamous Cell Carcinoma Progression by Targeting Cyclin D2. Molecular Therapy Nucleic Acids 9:100. DOI: 10.1016/j.omtn.2017.08.015.

Teufel A, Itzel T, Erhart W, Brosch M, Wang X, Kim Y, von Schönfels W, Herrmann A, Brückner S, Stickel F, Dufour J, Chavakis T, Hellerbrand C, Spang R, Maass T, Becker T, Schreiber S, Schafmayer C, Schuppan D, and Hampe J. 2016. Comparison of Gene Expression Patterns Between Mouse Models of Nonalcoholic Fatty Liver Disease and Liver Tissues From Patients. Gastroenterology 151:513-525.e510. DOI: 10.1053/j.gastro.2016.05.051

Verusingam N, Yeap S, Ky H, Paterson I, Khoo S, Cheong S, Ong A, and Kamarul T. 2017. Susceptibility of Human Oral Squamous Cell Carcinoma (OSCC) H103 and H376 cell lines to Retroviral OSKM mediated reprogramming. PeerJ 5:e3174. DOI:10.7717/peerj.3174.

Wang C, Wang L, Ding Y, Lu X, Zhang G, Yang J, Zheng H, Wang H, Jiang Y, and Xu L. 2017. LncRNA Structural Characteristics in Epigenetic Regulation. Int J Mol Sci 18. DOI: 10.3390/ijms18122659.

Wang Z, Gerstein M, and Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57-63. DOI:10.1038/nrg2484.

Wonganan O, He Y, Shen X, Wongkrajang K, Suksamrarn A, Zhang G, and Wang F. 2017. 6-Hydroxy-3-O-methyl-kaempferol 6-O-glucopyranoside potentiates the anti-proliferative effect of interferon α/β by promoting activation of the JAK/STAT signaling by inhibiting SOCS3 in hepatocellular carcinoma cells. Toxicol Appl Pharmacol 336:31-39. DOI: 10.1016/j.taap.2017.10.004

Wu Y, Zhang L, Zhang L, Wang Y, Li H, Ren X, Wei F, Yu W, Liu T, and Wang X. 2015. Long non-coding RNA HOTAIR promotes tumor cell invasion and metastasis by recruiting EZH2 and repressing E-cadherin in oral squamous cell carcinoma. International Journal of Oncology 46:2586. DOI: 10.3892/ijo.2015.2976.

Zhang CZ. 2017. Long intergenic non-coding RNA 668 regulates VEGFA signaling through inhibition of miR-297 in oral squamous cell carcinoma. Biochem Biophys Res Commun. DOI: 10.1016/j.bbrc.2017.05.155.

Figure 1(on next page)

A heat map drawn to show differential lncRNA expression in OSCC and normal tissue samples from the TCGA datasets, which were analyzed with R software

Representative genes of each cluster were selected and represented as a heat map. Genes shown in red are upregulated and genes in blue are downregulated. The magnitude of the regulation is illustrated by the intensity of the color.

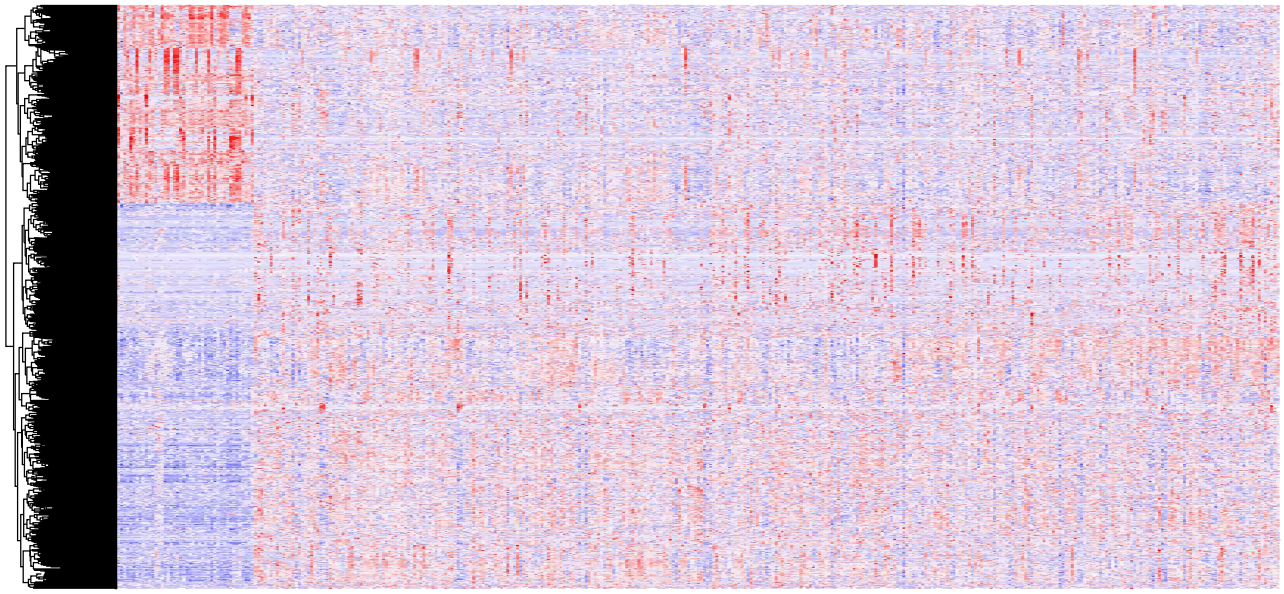

Figure 2(on next page)

Kaplan–Meier survival analyses and log-rank tests for OS and DFS in OSCC

(A) OS and (B) DFS rates of all patients according to AC012456.4 expression. (C) OS and (D) DFS rates of all patients according to AP003900.6 expression. (E) OS and (F) DFS rates of all patients according to BOK-AS1 expression. (G) OS and (H) DFS rates of all patients according to LINC01108 expression. (I) OS and (J) DFS rates of all patients according to RP11-1C8.7 expression. (K) OS and (L) DFS rates of all patients according to RP11-87C12.5 expression

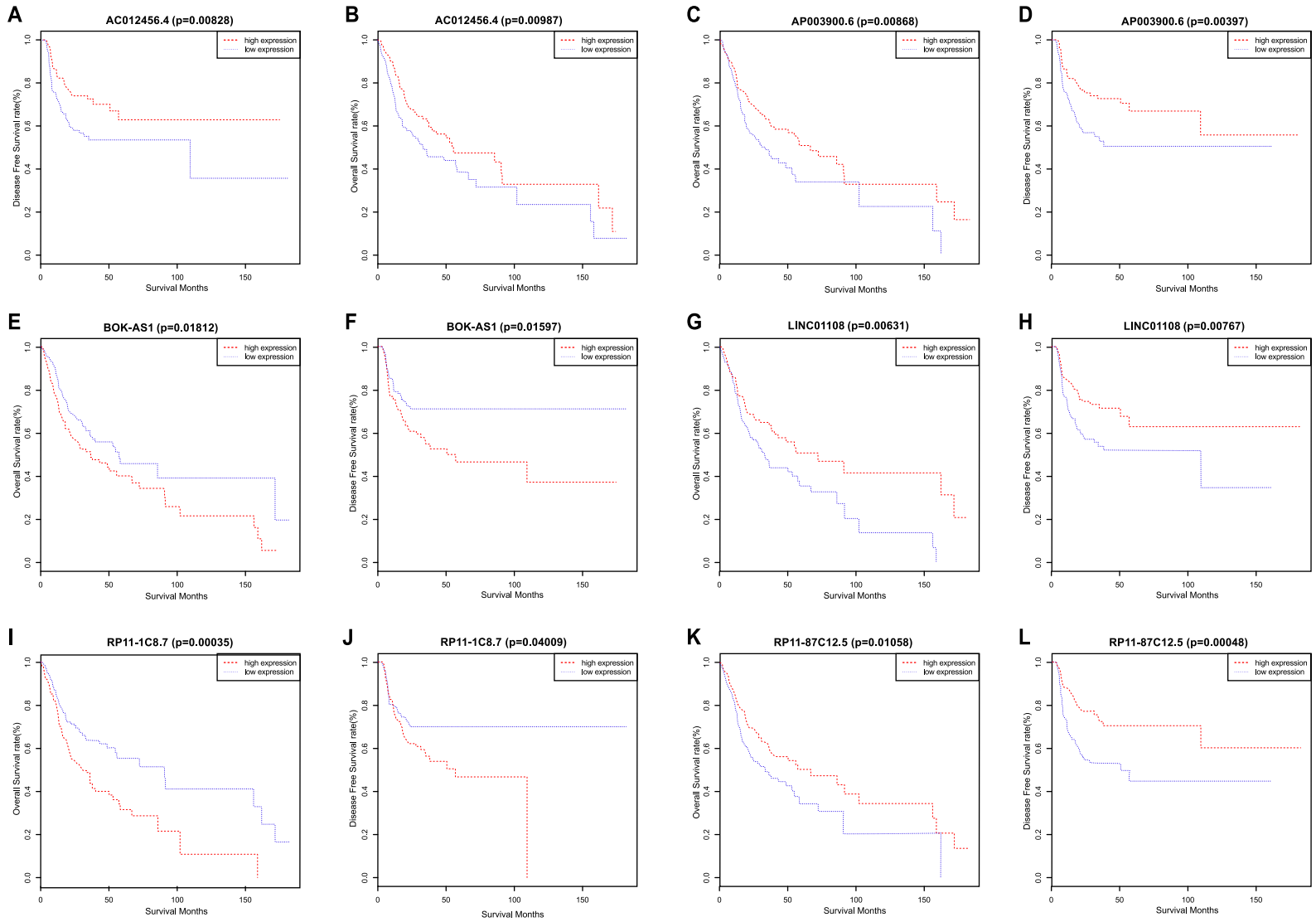
A
AC012456.4 (p=0.00828)
B
AC012456.4 (p=0.00987)
C
AP003900.6 (p=0.00868)
D
AP003900.6 (p=0.00397)
E
BOK-AS1 (p=0.01812)
F
BOK-AS1 (p=0.01597)
G
LINC01108 (p=0.00631)
H
LINC01108 (p=0.00767)
I
RP11-1C8.7 (p=0.00035)
J
RP11-1C8.7 (p=0.04009)
K
RP11-87C12.5 (p=0.01058)
L
RP11-87C12.5 (p=0.00048)
high expression
low expression
Survival Months
Disease Free Survival rate(%)
Overall Survival rate(%)

Figure 3(on next page)

Expression of AC012456.4 in normal tissues and OSCC tissues.

AC012456.4 expression is significantly down-regulated in OSCC samples (1.360 ± 0.05569) in comparison to adjacent non-cancerous tissues (3.062 ± 0.2304) in the TCGA dataset.

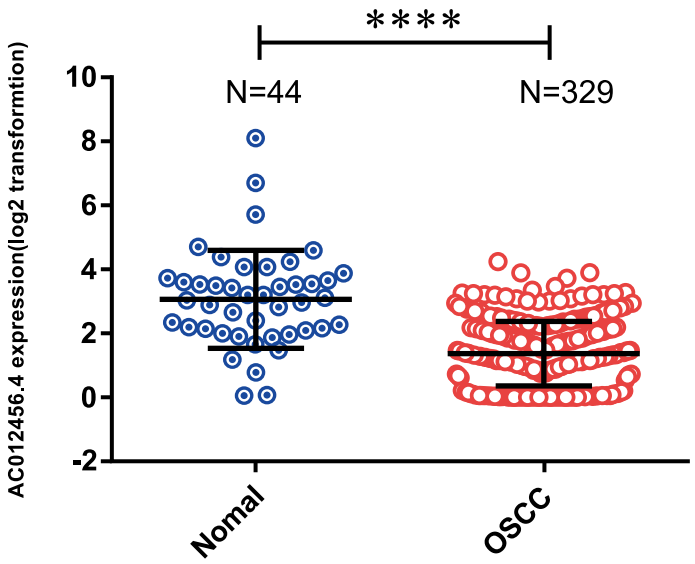
****
N=44
N=329
AC012456.4 expression(log2 transformtion)
10
8
6
4
2
0
-2
Nomal
OSCC

Figure 4(on next page)

Evaluation of the prognostic value of lncRNA AC012456.4 via the GEO

(A) Heatmap of lncRNA AC012456.4 expression in GEO. (B) lncRNA AC012456.4 expression was significantly low in OSCC. (C) OSCC patients were divided into the high expression group and the low expression group according to the median lncRNA AC012456.4 expression. (D) The low expression of lncRNA AC012456.4 was significantly associated with poor prognosis in patients with OSCC (p ˂0.0001).

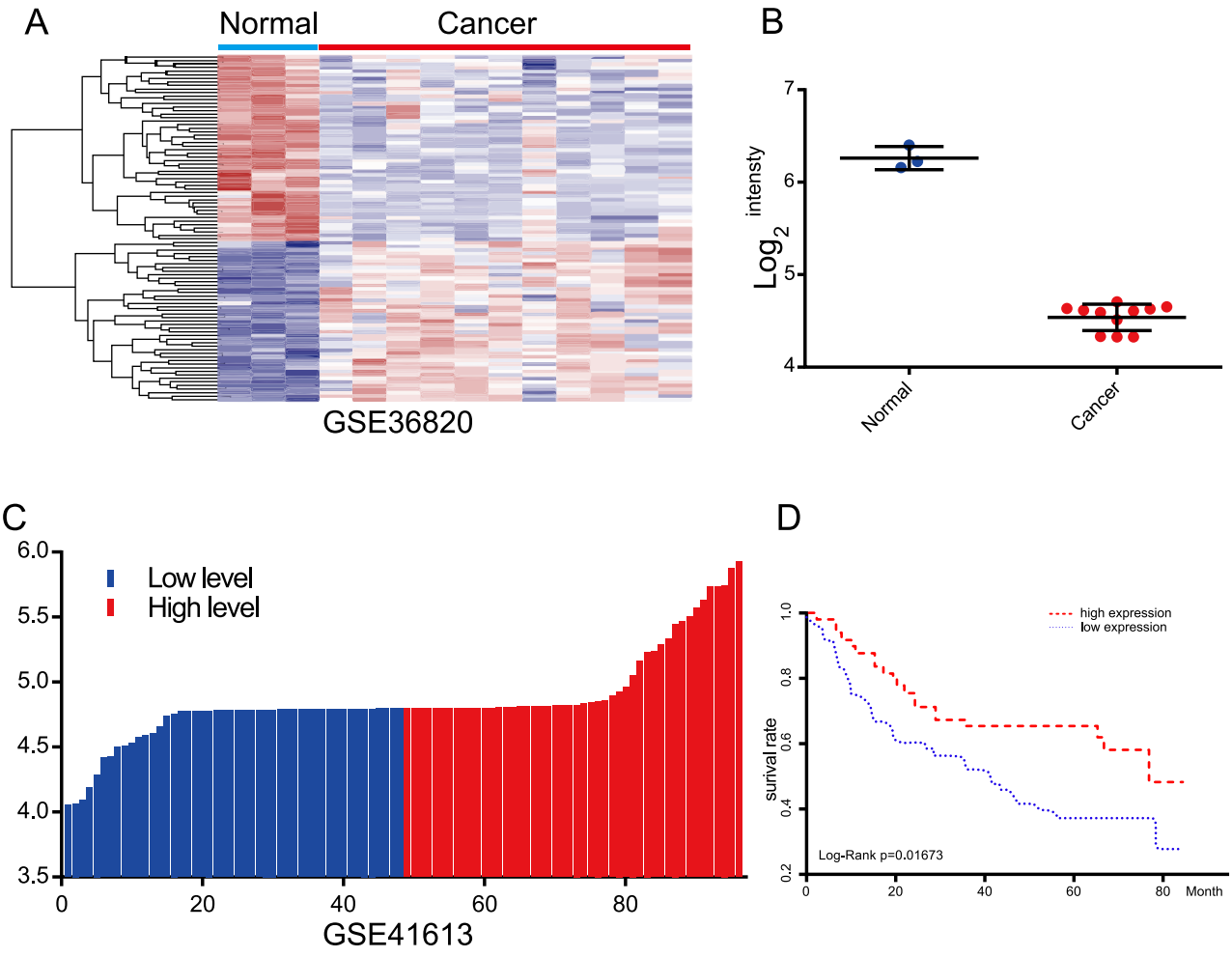
A
Normal
Cancer
GSE36820
B
intensty
Log2
7
6
5
4
Normal
Cancer
C
Low level
High level
6.0
5.5
5.0
4.5
4.0
3.5
0
20
40
60
80
GSE41613
D
high expression
low expression
surival rate
1.0
0.8
0.6
0.4
0.2
Log-Rank p=0.01673
0
20
40
60
80
Month

Figure 5(on next page)

KEGG pathway enrichment analysis of lncRNA AC012456.4

(A) Enrichment of genes in the KEGG MAPK SIGNALING PATHWAY by GSEA. (B) Heat map of core enrichment genes in the gene set KEGG MAPK SIGNALING PATHWAY. (C) Enrichment of genes in KEGG PATHWAYS IN CANCER by GSEA. (D) Heat map of core enrichment genes from the gene set KEGG PATHWAYS IN CANCER. (E) Enrichment of genes in KEGG OXIDATIVE PHOSPHORYLATION by GSEA. (F) Heat map of core enrichment genes from the gene set KEGG OXIDATIVE PHOSPHORYLATION. (G) Enrichment of genes in KEGG SPLICEOSOME by GSEA. (H) Heat map of core enrichment genes from the gene set KEGG SPLICEOSOME. The GSEA software was used to calculate enrichment levels.

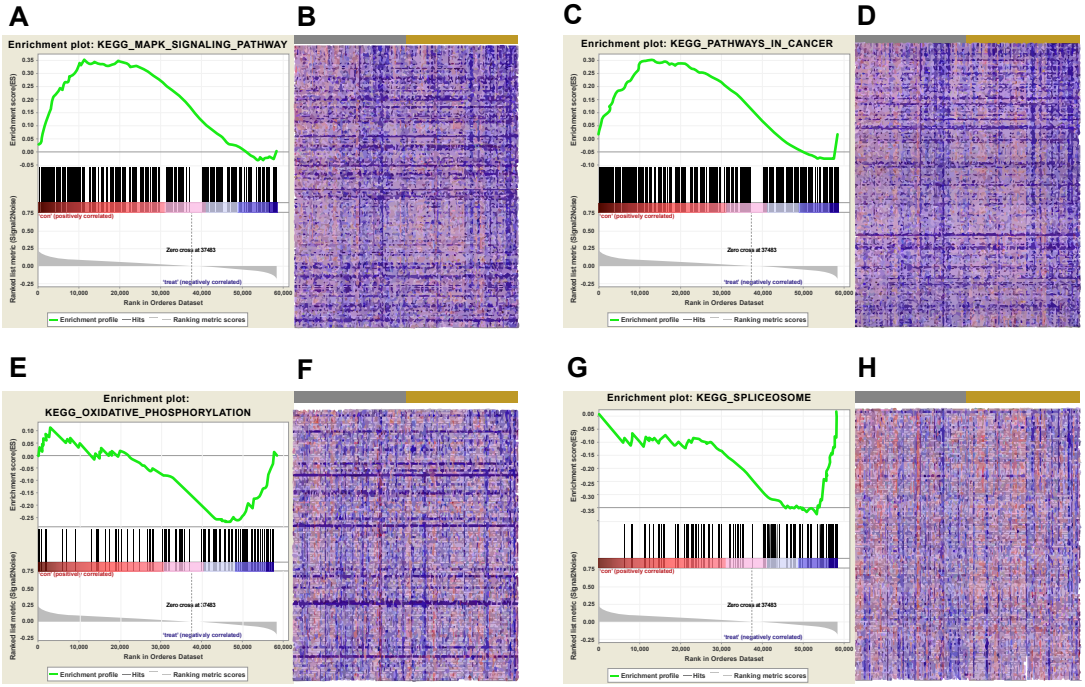
A
B
Enrichment plot: KEGG_MAPK_SIGNALING_PATHWAY
C
D
Enrichment plot: KEGG_PATHWAYS_IN_CANCER
E
F
Enrichment plot:
KEGG_OXIDATIVE_PHOSPHORYLATION
G
H
Enrichment plot: KEGG_SPLICEOSOME
Enrichment score(ES)
Ranked list metric (Signal2Noise)
'con' (positively correlated)
'treat' (negatively correlated)
Rank in Orderes Dataset
Enrichment profile
Hits
Ranking metric scores

Figure 6(on next page)

GSEA were carried out to identify upregulated or downregulated GO.

(A) Enrichment of genes in GO ADAPTIVE IMMUNE RESPONSE by GSEA. (B) Heat map of core enrichment genes in the gene set GO ADAPTIVE IMMUNE RESPONSE. (C) Enrichment of genes in GO POSITIVE REGULATION OF CELL ACTIVATION by GSEA. (D) Heat map of core enrichment genes in the gene set GO POSITIVE REGULATION OF CELL ACTIVATION. (E) Enrichment of genes in GO RRNA METABOLIC PROCESS by GSEA. (F) Heat map of core enrichment genes in the gene set GO RRNA METABOLIC PROCESS. (G) Enrichment of genes in GO RIBOSOME BIOGENESIS by GSEA. (H) Heat map of core enrichment genes in the gene set GO RIBOSOME BIOGENESIS. The GSEA software was used to calculate the enrichment levels.

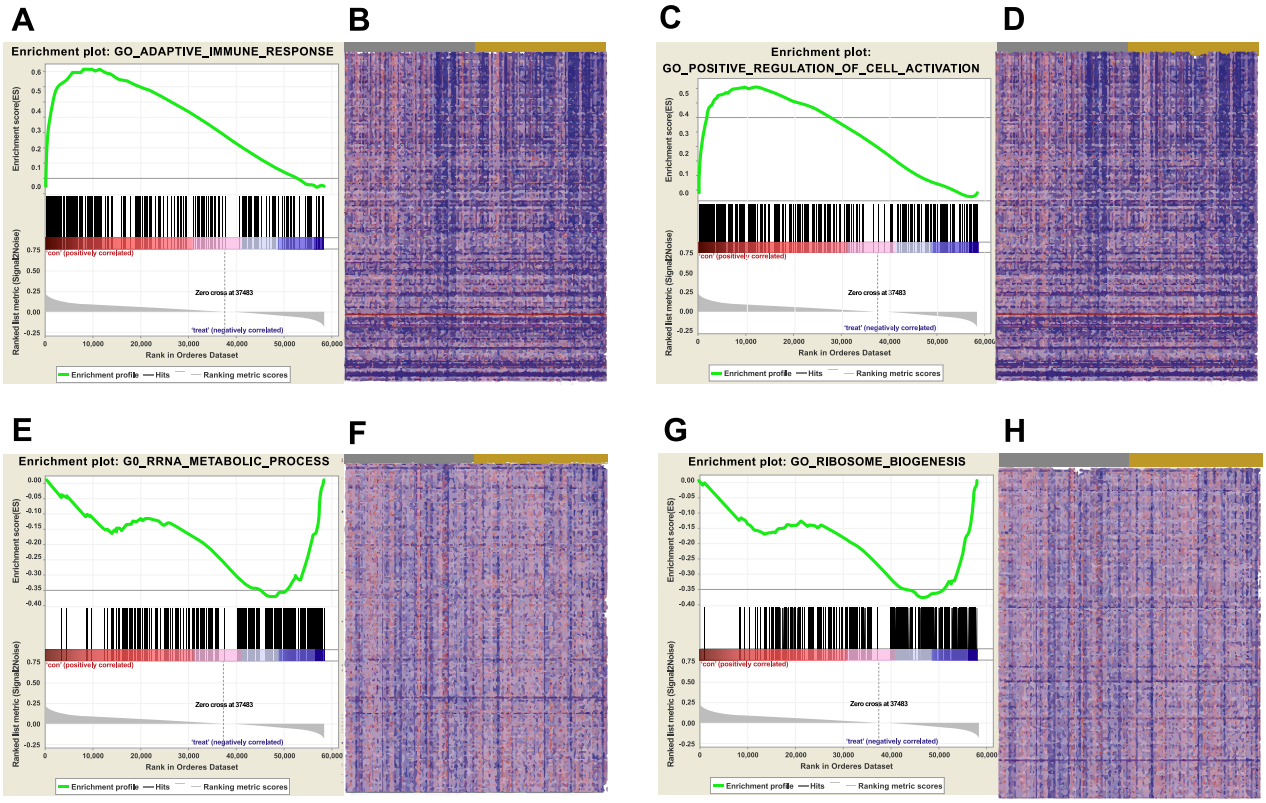
A
B
C
D
Enrichment plot: GO_ADAPTIVE_IMMUNE_RESPONSE
Enrichment plot: GO_POSITIVE_REGULATION_OF_CELL_ACTIVATION
E
F
G
H
Enrichment plot: G0_RRNA_METABOLIC_PROCESS
Enrichment plot: GO_RIBOSOME_BIOGENESIS
Enrichment score(ES)
Ranked list metric (Signal2Noise)
'con' (positively correlated)
'treat' (negatively correlated)
Zero cross at 37483
Rank in Orderes Dataset
Enrichment profile
Hits
Ranking metric scores

Table 1(on next page)

The clinicopathological characteristics of patients from the TCGA database.

Table 1. The clinicopathological characteristics of patients from the TCGA database.

Characteristics	Number of case	No. of Patients (%)
Age (years)	346	
≦60		152(41.33%)
≧60		194(58.67%)
Median (range)		61.590(19-90)
Gender	347	
Male		236(68.01%)
Female		111(31.99%)
Alcohol history	339	
No		111(32.74%)
Yes		228(67.26%)
Perineural invasion present	263	
No		123(46.77%)
Yes		140(53.23%)
Margin status	324	
Close		39(12.04%)
Negative		244(75.31%)
Positive		41(12.65%)
Lymphovascular invasion present	250	
Yes		76(30.40%)
No		174(69.60%)
Tumor stage	314	
Stage I		21(6.69%)
Stage II		56(17.83%)
Stage III		64(20.38%)
Stage IV		173(55.10%)
T stage	335	
T1		34(10.15%)
T2		103(%)
T3		70(%)
T4		128(%)
N stage	334	
NO		126(37.72%)
N1		52(15.57%)
N2		110(32.93%)
N3		46(13.77%)
M stage	170	
M0		125(73.53%)
M1		45(26.47%)
Histologic grade	344	
G1		53(15.41%)
G2		210(61.05%)
G3		71(20.64%)
G4		10(2.91%)

vital status	347	
Alive		227(65.42%)
Dead		120(34.58%)

1

Table 2(on next page)

21 lncRNA levels significantly correlated to OS and DFS.

Table 2. 21 lncRNA levels significantly correlated to OS and DFS.

LncRNA	Gene ID	Chromosome	OS(P value)	DFS(P value)
AC012456.4	ENSG00000230790	chr2	0.00987	0.00828
AP003900.6	ENSG00000271308	chr21	0.00868	0.00397
BOK-AS1	ENSG00000234235	chr2	0.01812	0.01597
LINC01108	ENSG00000226673	chr6	0.00631	0.00767
RP11-1C8.7	ENSG00000271830	chr8	0.00035	0.04009
RP11-87C12.5	ENSG00000255856	chr12	0.01058	0.00048
TTC39A-AS1	ENSG00000261664	chr1	0.04276	0.00371
RP11-93B14.9	ENSG00000277496	chr20	0.01279	0.00352
AC007879.2	ENSG00000234902	chr2	0.00811	0.03607
RP11-464F9.21	ENSG00000234606	chr10	0.01486	0.03221
LINC01549	LINC01549	chr21	0.00021	0.0165
CTB-161M19.4	ENSG00000249494	chr5	0.04807	0.01152
RP11-286E11.1	ENSG00000245293	chr4	0.03618	0.0041
RP11-897M7.1	ENSG00000256209	chr12	0.03129	0.02265
LINC01343	ENSG00000237290	chr1	0.01115	0.03191
FAM95B1	ENSG00000223839	chr9	0.04778	0.01648
RP11-181E10.3	ENSG00000271590	chr2	0.00597	0.00934
CTD-2545H1.2	ENSG00000262445	chr17	0.02892	0.02929
RP11-796E2.4	ENSG00000245904	chr12	0.04276	0.00371
CTD-2033A16.3	ENSG00000262136	chr16	0.04586	0.02714
RP11-285G1.14	ENSG00000273363	chr10	0.01276	0.00503

1

Table 3(on next page)

Univariate and multivariate Cox regression analysis for OS in patients with OSCC.

Table 3. Univariate and multivariate Cox regression analysis for OS in patients with OSCC.

Variables	Univariate analysis			Multivariate analysis		
	P value	HR	95% CI	P value	HR	95% CI
Age (years)	0.003	1.021	1.007,1.036	0.001	1.026	1.011,1.041
Gender	0.459	1.150	0.794,1.665	0.481	1.145	0.786,1.666
Grade	0.127	1.215	0.946,1.560	0.062	1.276	0.988,1.648
Stage (age≦60)	0.034	1.425	1.026,1.978	0.210	0.765	0.503,1.163
(age>60)	0.523	1.080	0.853,1.367			
N	0.015	1.263	1.046,1.524	0.011	1.279	1.059,1.546
T (age≦60)	0.003	1.551	1.160,2.075	0.293	1.101	0.921,1.316
(age>60)	0.873	0.982	0.783,1.230			
AC012456.4	0.006	0.706	0.551,0.903	0.002	0.672	0.523,0.863

N: REGIONAL LYMPH NODES

T: PRIMARY TUMOR

1

Table 4(on next page)

Univariate and multivariate Cox regression analysis for DFS in patients with OSCC.

Table 4. Univariate and multivariate Cox regression analysis for DFS in patients with OSCC.

Variables	Univariate analysis			Multivariate analysis		
	P value	HR	95% CI	P value	HR	95% CI
Age (years)	0.093	1.017	0.997,1.036	0.071	1.018	0.999,1.037
Gender	0.627	1.132	0.687,1.867	0.678	1.113	0.672,1.841
Grade	0.817	1.043	0.732,1.485	0.533	1.125	0.777,1.627
Stage	0.625	1.064	0.830,1.363	0.482	0.852	0.545,1.332
N	0.539	1.085	0.7837,1.407	0.167	1.286	0.900,1.836
T	0.191	1.167	0.926,1.470	0.295	1.134	0.896,1.434
AC012456.4	0.004	0.601	0.423,0.853	0.004	0.600	0.423,0.851

1

Table 5(on next page)

AC012456.4 expression and clinicopathological characteristics of patients with OSCC.

Table 5. AC012456.4 expression and clinicopathological characteristics of patients with OSCC.

Characteristics	Number of case	AC012456.4 expression		P value
		Decreased Number(%)	Non-decreased Number(%)	
Age (years)				0.082
≥60	186	96(51.61%)	90(48.39%)	
＜60	143	60(41.96%)	83(58.08%)	
Gender				0.745
Female	102	47(46.08%)	55(59.92%)	
Male	227	109(48.02%)	118(51.98%)	
Alcohol history				0.033
Yes	213	109(51.17%)	104(48.83%)	
No	104	40(38.46%)	64(61.54%)	
M stage				0,511
M0	119	56(47.06%)	63(52.94%)	
M1	39	16(41.03%)	23(58.97%)	
T stage				0.075
T1+T2	128	54(42.19%)	74(57.81%)	
T3+T4	189	99(52.38%)	90(47.62%)	
N stage				0.163
N0+N1	168	87(51.79%)	81(48.21%)	
N2+N3	148	65(43.92%)	83(56.08%)	

M0: No distant metastasis (no pathologic M0; use clinical M to complete stage group.

M1: Distant metastasis.

N0: No regional lymph node metastasis.

N1: Metastasis in a single ipsilateral lymph node, 3 cm or less in greatest dimension.

N2: Metastasis in a single ipsilateral lymph node, more than 3 cm but not more than 6 cm in greatest dimension; or in multiple ipsilateral lymph nodes, none more than 6 cm in greatest dimension; or in bilateral or contralateral lymph nodes, none more than 6 cm in greatest dimension.

N3: Metastasis in a lymph node more than 6 cm in greatest dimension.

T1: Tumor 2 cm or less in greatest dimension.

T2: Tumor more than 2 cm but not more than 4 cm in greatest dimension.

T3: Tumor more than 4 cm in greatest dimension

T4a: Moderately advanced local disease.

T4b: T4b Very advanced local disease.

Tumor invades masticator space, pterygoid plates, or skull base and/or encases internal carotid

artery

1

Table 6(on next page)

KEGG Pathways enriched in high-risk and low-risk groups by using GSEA.

1

Table 6. KEGG Pathways enriched in high-risk and low-risk groups by using GSEA.

NAME	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val	RANK AT MAX	LEADING EDGE
KEGG_PRIMARY_IMMUNODEFICIENCY	35	0.783950	2.003367	0.002036	0.080199	0.032	5022	tags=63%, list=9%, signal=69%
KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION	258	0.503302	1.751613	0.016227	0.351688	0.258	13393	tags=46%, list=23%, signal=60%
KEGG_JAK_STAT_SIGNALING_PATHWAY	151	0.462485	1.585162	0.051020	0.356088	0.496	11252	tags=35%, list=19%, signal=43%
KEGG_PATHWAYS_IN_CANCER	324	0.296756	1.015304	0.442386	0.570524	0.968	12772	tags=28%, list=22%, signal=36%
KEGG_MAPK_SIGNALING_PATHWAY	265	0.353983	1.239951	0.226804	0.527153	0.881	11268	tags=29%, list=19%, signal=35%
KEGG_PROTEASOME	46	-0.542264	-1.310828	0.249049	1	0.849	11204	tags=54%, list=19%, signal=67%
KEGG_CYTOSOLIC_DNA_SENSING_PATHWAY	55	-0.342477	-1.059409	0.361581	1	0.958	6866	tags=33%, list=12%, signal=37%
KEGG_SNARE_INTERACTIONS_IN_VESICULAR_TRANSPORT	38	-0.365953	-0.983674	0.481132	1	0.969	6863	tags=32%, list=12%, signal=36%
KEGG_OXIDATIVE_PHOSPHORYLATION	118	-0.269338	-0.724989	0.681050	1	0.992	11643	tags=38%, list=20%, signal=48%
KEGG_SPLICEOSOME	123	-0.362891	-0.936620	0.566473	1	0.978	5025	tags=24%, list=9%, signal=26%

Table 7 (on next page)

GO annotation enriched in high-risk and low-risk groups by using GSEA.

1

Table 7. GO annotation enriched in high-risk and low-risk groups by using GSEA.

NAME	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val	RANK AT MAX	LEADING EDGE
GO_B_CELL_RECEPTOR_SIGNALING_PATHWAY	54	0.749803	1.963207	0.003838	0.974954	0.161	6389	tags=67%, list=11%, signal=75%
GO_ADAPTIVE_IMMUNE_RESPONSE	279	0.614785	1.932954	0.007648	0.761521	0.202	7793	tags=46%, list=13%, signal=53%
GO_NEGATIVE_REGULATION_OF_INTERLEUKIN_6_PRODUCTION	33	0.711452	1.897834	0	0.660863	0.28	10264	tags=67%, list=18%, signal=81%
GO_REGULATION_OF_B_CELL_ACTIVATION	121	0.626420	1.886616	0.003883	0.617897	0.294	9579	tags=55%, list=16%, signal=65%
GO_POSITIVE_REGULATION_OF_CELL_ACTIVATION	305	0.540650	1.725436	0.031496	0.464684	0.631	11768	tags=47%, list=20%, signal=58%
GO_CELLULAR_RESPONSE_TO_ZINC_ION	16	-0.60868	-1.550511	0.056310	1	0.883	4440	tags=56%, list=8%, signal=61%
GO_RIBOSOMAL_LARGE_SUBUNIT_BIOGENESIS	48	-0.60318	-1.496404	0.109343	1	0.925	5330	tags=42%, list=9%, signal=46%
GO_POSITIVE_REGULATION_OF_PEPTIDYL_SERINE_PHOSPHORYLATION_OF_STAT_PROTEIN	21	-0.52630	-1.392874	0.115079	1	0.962	6411	tags=48%, list=11%, signal=53%
GO_RRNA_METABOLIC_PROCESS	249	-0.38387	-1.055244	0.457925	1	0.998	10606	tags=36%, list=18%, signal=44%
GO_RIBOSOME_BIOGENESIS	300	-0.38284	-1.050548	0.456692	1	0.998	11706	tags=38%, list=20%, signal=47%