

A multi-lncRNA signature for survival prediction in esophageal squamous cell carcinoma (ESCC) patients

Lan zhang

Department of Pathology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou University

Pan Li

Department of Pathology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou University

Chenju Xing

Department of Pathology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou University

Di Zhu

Department of Pathology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou University

Jianying Zhang

Institute of Medical and Pharmaceutical Sciences, Zhengzhou University

Weiwei Wang

Department of Pathology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou University

Guozhong Jiang (✉ guozhongjiang@zzu.edu.cn)

Zhengzhou University First Affiliated Hospital <https://orcid.org/0000-0001-6306-726X>

Research

Keywords: lncRNA, esophageal squamous cell carcinoma, overall survival, signature, prognosis

Posted Date: May 12th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-26795/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: The aim of this study was to identify prognostic long non-coding RNAs (lncRNAs) and develop a multi-lncRNA signature for survival prediction in esophageal squamous cell carcinoma (ESCC) patients.

Methods: The clinical and gene expression data from 301 ESCC patients were downloaded, including a cohort used as training set from Gene Expression Omnibus database (GSE53624, $n=119$), another cohort of 98 paired ESCC tumor and normal tissues as test set and an independent validation cohort including 84 ESCC tissues. Survival analyses, Cox regression and Kaplan–Meier analysis were used.

Results: we screened a prognostic marker of ESCC from the GSE53624 dataset and named it as the five-lncRNA signature including AC007179.1, MORF4L2-AS1, RP11-488I20.9, RP13-30A9.2, RP4-735C1.6, which could classify patients into high- and low-risk groups with significantly different survival (median survival: 1.75 years vs. 4.01 years, log rank $P<0.05$). Then test dataset and validation dataset confirmed that the five-lncRNA signature can determine the prognosis of ESCC patients. Predictive independence of the prognostic marker was proved by multivariable Cox regression analyses in the three datasets ($P<0.05$). In addition, the signature was found to be better than TNM stage in terms of prognosis.

Conclusion: The five-lncRNA signature could be a good prognostic biomarker for ESCC patients and has important clinical value.

Background

Esophageal carcinoma is a malignant tumor of esophageal mucosa epithelium or gland. Globally, esophageal cancer ranks seventh in morbidity and sixth mortality among all cancers^[1]. China is one of the countries with the highest incidence of esophageal cancer, accounting for approximately half of all esophageal cancer cases^[2]. There were 477,900 new cases diagnosed in China in 2015^[3]. Among these Chinese esophageal cancer patients, esophageal squamous cell carcinoma (ESCC) patients account for nearly 90%^[2]. In terms of mortality, esophageal cancer deaths in China account for half of global esophageal cancer deaths^[4]. Therefore, ESCC is a major public health challenge in the world, especially in China. Using Chinese ESCC patients as research samples is of great significance in revealing the pathogenesis and prognostic factors of ESCC.

ESCC is an extremely aggressive malignancy. Along with the clinical symptoms of early esophageal squamous cell carcinoma are not obvious and can be easily ignored by patients, most ESCC patients are diagnosed in advanced stages. As a result, ESCC patients have a extremely poor survival. The 5-year survival rate of ESCC patients in some less developed areas is less than 10%^[5]. In the United States, where medical conditions are more advanced, the 5-year survival rate is 18%^[6]. In China, the 5-year relative survival of postoperative ESCC patients is 8%-41%^[7,8]. Despite recent advances in diagnostic and therapeutic approaches, the life expectancy of ESCC patients have not improved significantly^[9-11].

Therefore, identification of a sensitive and effective prognostic marker of ESCC is in urgent need to evaluate disease progression and patient's overall outcome.

Next generation sequencing (NGS) and bioinformatics analysis tools provides great help for humans to understand tumors^[12-15]. In the process of exploring the pathogenesis and prognostic factors of tumors by NGS and bioinformatics analysis, long non-coding RNAs (lncRNAs), defined as transcripts longer than 200 bp, have been discovered to play an important role. For example, a HOX transcript antisense RNA, lncRNA HOTAIR, was found to correlate with poor prognosis of lung cancer and promote tumor progression^[16]. LncRNA REG1CP was shown to promote the development and progression of tumor^[17]. For ESCC, several lncRNAs have been detected and associated with shorter survival, such as lncRNA H19^[18], lncRNA CASC9^[19] and lncRNA LUCAT1^[20]. Subsequently, prognostic multi-lncRNA signatures with high potential clinical application significance were constructed based on lncRNA expression profile data from public database such as Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) database. For instance, a three-lncRNA signature (ENST00000435885.1, XLOC_013014, ENST00000547963.1) was found that it can classify the ESCC patients into two groups with significantly different overall survival.^[21] A seven-lncRNA (RP5-1172N10.2, RP11-579D7.4, RP11-89N17.4, LA16c-325D7.2, RP1-251M9.2, RP11-259O2.2, LINC00173) signature can predict overall survival of ESCC patients^[22]. Although prognostic lncRNA signatures have been identified for ESCC, there are few studies that can verify the effectiveness of the predictive model in independent experimental data set and confirm the presence of the prognostic lncRNA in ESCC tissues.

Here, we used our lncRNA expression data tested by RNA sequencing and followed up the clinical 5-year-survival information of 98 ESCC patients, then, combined it with 119 ESCC public expression profiles from GEO and another independent 84 ESCC tissues for qPCR validation to construct a clinically valuable lncRNA signature that can accurately predict survival in ESCC patients.

Materials And Methods

Sample collection and preparation

lncRNA expression profile and corresponding clinical data of 119 ESCC cases (GSE53624) were obtained from the publicly available GEO database (<https://www.ncbi.nlm.nih.gov/geo/>). As Anyang is a high-risk area of ESCC in China, we collected 98 postoperative ESCC tissues and paired non-tumor tissues from Anyang Tumor Hospital during 2014-2019, and organized relevant patient clinical information. Then we examined the PCG and lncRNA expression profile of the 98 pair ESCC tissues by next-generation sequencing (NGS, Hereinafter referred to as RNA-seq dataset). In addition, we collected an independent validation cohort including 84 postoperative ESCC patients from the same hospital and detected the lncRNA expression level using the qRT-PCR. The patients were coded to protect anonymity and all pathological information of the ESCC samples in this study was shown in Table 1. Tumor-node-metast

asis (TNM) classification of the International Union against Cancer, 7th edition was used to categorize. Documentation of informed consent was obtained through the institutional review board. The study was approved by the Anyang Tumor Hospital Ethical Committee.

RNA isolation and next generation RNA sequencing analysis

After TRIZOL lysis and purification, total RNA was isolated by the miRNeasy Mini Kit (QIAGEN) with DNase digestion step. A total amount of 5 ug RNA per sample was used as input material for the RNA sample preparations. Firstly, ribosomal RNA was removed by Epicentre Ribozero™ rRNA Removal Kit (Epicentre, USA), and rRNA free residue was cleaned up by ethanol precipitation. The sequencing libraries were generated by NEBNext Ultra™ Directional RNA Library Prep Kit for Illumina (NEB, USA) following manufacturer's recommendations and sequenced on an Illumina Hiseq platform. The 150 bp paired-end reads were generated. Normalized fragments per kilobase per million mapped reads (FPKM) by cufflinks was used to estimate the gene expression values.

Validation of lncRNA expression using RT-PCR

lncRNAs reverse transcription were amplified with TIANScript[®] RT Kit (KR107, TIANGEN, Beijing, China). We used Real-time quantitative PCR (qRT-PCR) to measure the lncRNAs expression with TB Green[®] Premix Ex Taq™ (Tli RNaseH Plus, TaKaRa, Dalian, China). Relative quantification of lncRNA expression was normalized by the $-\Delta\Delta C_t$ method, and GAPDH was used for normalization with the corresponding primers (Supplementary Table S3). All reactions were carried out in triplicate by StepOnePlus™ Real-Time PCR System (Applied Biosystems) [23-25].

Construction of multi-lncRNA prognostic signature

GSE53624 and RNA-seq dataset were analyzed respectively by Cox analysis and Kaplan–Meier analysis to identify lncRNA that associated with OS of ESCC patients. We selected the prognostic lncRNA using Cox p & log rank P < 0.05 in two datasets. Model was estimated as follows [23,24,26]: **Risk Score (RS) = $\sum_{i=1}^N (ExpLncRNA_i * CoefCox_i)$** , where N is prognostic lncRNA number, *ExpLncRNA_i* represents the lncRNA expression value, and *CoefCox_i* is the Cox regression coefficient of lncRNA. We plotted ROC curves^[14] and calculated their area under the curve (AUC) values to screen out the prognostic signature with largest AUC value in the GSE53624 set [26]. R program (Version 3.5.1) was used to perform the above analyses, including packages called pROC, TimeROC and survival from Bio-conductor (<http://www.bioconductor.org/>). Gene set enrichment analysis (GSEA) were used to find the dysregulated pathways in the low- and high-risk groups. Normalized enrichment score (NES) and P value were calculated to verify the statistical difference for GSEA analysis^[27].

Results

Screening prognostic lncRNAs in ESCC

We used 119 ESCC data from public GSE53624 as training group, 98 RNA-seq data as test group and 84 qPCR data as validation group. Detailed clinicopathological characteristics of all these ESCC patients were showed in Table 1. There were a total of 217 samples with lncRNA expression data including GSE53624 dataset and test dataset. From the two datasets, we found a total of 6253 lncRNAs were expressed in ESCC tissue. We then performed univariate cox analysis and Kaplan–Meier analysis to analyze the relationship between the lncRNA expression and ESCC OS. Based on the two ESCC profiling datasets and the corresponding clinical follow-up information, univariate cox analyses identified 368 lncRNAs in the GSE53624 and 290 lncRNAs in the RNA-seq dataset which were significantly associated with ESCC OS (Cox $P < 0.05$) and Kaplan–Meier analyses identified 386 survival associated lncRNAs in the GSE53624 and 312 survival associated lncRNAs in the RNA-seq dataset Kaplan–Meier (log rank $P < 0.05$). Through comparing the above four groups of survival related lncRNA data, we found 19 identical lncRNAs significantly correlated with OS in two datasets (COX $P < 0.05$, log rank $P < 0.05$) (Figure 1A). Subsequently, we found that 10 of the 19 prognostic-related lncRNAs showed a consistent risk trend in the four groups data, displaying a risk or protective role in ESCC (Supplementary table S1).

Constructing prognostic multi-lncRNA signature

Using 10 prognostic lncRNAs, we constructed $2^{10}-1=1023$ signatures. In order to obtain a signature with the strongest prognostic ability, we performed multiple ROC analyses in the training dataset (Supplementary table S2). We screened the 5-lncRNA combination with the largest AUC value, namely the prognostic signature containing AC007179.1, MORF4L2-AS1, RP11-488I20.9, RP13-30A9.2, RP4-735C1.6 (Figure 1B, Table 2). The lncRNA risk score was calculated as follows: $(0.47 \times \text{expression value of AC007179.1}) + (0.58 \times \text{expression value of MORF4L2-AS1}) + (-0.47 \times \text{expression value of RP11-488I20.9}) + (-0.64 \times \text{expression value of RP13-30A9.2}) + (0.62 \times \text{expression value of RP4-735C1.6})$. As shown in Figure 1C, the AUC of the screened lncRNA signature was 0.71.

Predictive power of the lncRNA signature for patients with ESCC

After calculated risk scores of ESCC patients in the GSE53624 dataset, the median risk classified ESCC patients into high- and low- risk groups ($n = 59/60$). Kaplan–Meier analysis found the survival of ESCC patients in the low-risk group were significantly longer than those in the high-risk group (median survival: 1.75 years vs. 4.01 years, log-rank test $P < 0.001$; Figure 2A). The 3-year survival rate for patients in the high-risk group was only 23.73%, while that of patients in the low-risk group reached as high as 66.67%.

Then we evaluated the predictive ability of the five-lncRNA signature in ESCC RNA-seq dataset ($n=98$). The median risk score of patients in the test dataset was calculated and used to divide patients into high- and low-risk groups. Figure 2B showed the Kaplan–Meier analysis result. The survival rate of ESCC patients in the low-risk group was significantly higher than patients in the high-risk group. (Median survival: 2.44 years vs. 3.97 years, log-rank test $P=0.014$; Figure 2B). The 3-year survival rate for the patients of high-risk group was only 40.82%, while that of the low-risk group was 61.22%.

In addition, we performed qPCR experiment with the primers (Supplementary table S3) in 84 ESCC tissues and obtained the five lncRNAs expression level. Through calculated risk scores and used the median risk score as the cutoff value, we obtained the high risk score group and low risk score group. Kaplan–Meier analysis discovered the survival difference between two risk groups. The 3-year survival rate of ESCC patients with low-risk scores was still significantly higher than patients with high-risk scores. (Median survival: 2.82 years vs. 4.08 years, log-rank test $P=0.011$; Figure 2C).

Figure 3 displayed the survival time, risk score and lncRNA expression level of each patients in training, test and qPCR validation datasets. As shown, the higher expression of AC007179.1, MORF4L2-AS1 and RP4-735C1.6, the higher the risk and the more deaths. On the contrary, patients with lower-risk scores tended to have higher expression level of protective lncRNAs (RP13-30A9.2, RP11-488I20.9) and have a lower mortality rate .

The five-lncRNA signature can predict the survival of ESCC patients independently

The Chi-square test found lncRNA signature was not associated with clinicopathological factors, including age, sex and pTNM stage in the training, test and qPCR validation groups (Table 3). Univariate cox analysis of the training, test and independent validation datasets showed TNM staging was associated with prognosis. The results of multivariable Cox regression analyses showed only the lncRNA signature was an independent prognostic factor for ESCC patients (High-risk group vs. Low-risk group, HR training = 3.50, 95% CI 2.13–5.76, $P < 0.001$, $n=119$; HR test = 1.03, 95% CI 1.00–1.06, $P = 0.014$, $n=98$; HR independent = 2.95, 95% CI 1.21–7.17, $P = 0.017$, $n=84$, Table 4).

The prediction performance of the five-lncRNA signature and TNM stage

TNM stage is currently used as the main indicator for prognosis evaluation of ESCC. Our study confirmed that TNM staging and the five-lncRNA signature were related to the prognosis of patients. Therefore, we performed stratification analysis to explore the predictive power of the five-lncRNA signature in patients with known TNM stage. We first divided all the ESCC patients involved in this study ($n=301$) into two groups, one group was patients with TNM low (I+II) stage and the other group was patients with high TNM stage (III+IV). Then we used the lncRNA signature to conduct stratification analyses of above two groups of patients. Kaplan–Meier results showed that the lncRNA signature could further separated the patients with TNM low (I+II) stage (Figure 4A) or patients with high TNM stage (III+IV)(Figure 4B) into two subgroups with significantly different survival(log-rank test $P<0.001$).

To analyze the five-lncRNA signature's predictive performance advantages, TimeROC analyses were performed in the three datasets ($n=301$). The AUCs of the signature were 0.690/0.648/0.639 at 5/3/1 years (Figure 4C), while the AUCs of TNM were 0.643/0.654/0.637 at 5/3/1 years (Figure 4D), suggesting the five-lncRNA signature was superior to TNM stage in terms of ESCC prognosis evaluation, especially at 5 years survival. Moreover, we found AUC values were maximized when signature and TNM staging were combined to assess patient prognosis (0.742/0.723/0.702 at 5/3/1 years, Figure 4E), indicating the lncRNA signature improve the accuracy of TNM in evaluating prognosis.

Functional prediction of five prognostic lncRNAs

In order to predict function of the five prognostic lncRNAs, we performed GSEA analyses based on high-risk and low-risk groups in the GSE53624, and RNAseq datasets, respectively. The four-lncRNA risk signature was significantly enriched in 83 different GO categories and KEGG pathways in the GSE53624, and enriched in 385 GO categories and KEGG pathways in RNA seq dataset, simultaneously ($P < 0.05$, Supplementary Table S4), especially involved in RNA metabolism and immune response processes such as cell division, mRNA processing, ncRNA metabolic process, RNA splicing, activation of immune response, immune response regulating cell surface receptor signaling pathway, innate immune response, immune effector process (Figure 5).

Discussion

Esophageal squamous cell carcinoma is a rapidly progressive disease faced with many difficulties, such as difficulty in early diagnosis, low five-year survival rate, and lack of effective prognostic markers. Recently, lncRNA has been reported to be involved in the occurrence and progression of tumors^[28,29]. However, the expression characteristics and roles of lncRNAs in ESCC are still fairly elusive. Thus, this paper revealed the survival related lncRNAs expression level in ESCC and constructed a prognostic five-lncRNA signature by collected public ESCC lncRNA expression data. Then, we measured the lncRNA expression of 98 ESCC patients by RNA seq and test the lncRNA expression of 84 ESCC tissues for validating the predictive performance of the five-lncRNA signature. Meanwhile, the 98 ESCC lncRNA expression profiles provide scientists with reference data to study the association of ESCC and lncRNAs.

Despite numerous research have reported some ESCC lncRNA models for prognostic markers^[22], PCR validation was rarely performed in the independent validation group and some lncRNAs in the predicting model might not yet exist in ESCC tissue. In this study, the predictive model was constructed through the analyzing two sets of high-throughput sequencing data, and lncRNAs of the five-lncRNA signature were shown to be present in collected ESCC samples by qPCR.

In the five-lncRNA survival predicting model, the high expression level of lncRNA MORF4L2-AS1, AC007179.1 and RP4-735C1.6 were associated with short survival (Cox coefficient >0), indicating these three lncRNAs were risk lncRNAs for ESCC patients. The high expression of RP13-30A9.2 and RP11-488I20.9 were associated with long survival (Cox coefficient <0), suggesting the two lncRNAs were protective lncRNAs for ESCC. The significant correlation of the five lncRNAs with ESCC prognosis was found in two data sets, emphasizing the important clinical research value of the five prognostic lncRNAs in ESCC. However, the function of prognostic lncRNAs in ESCC have not been reported so far and biological roles of the five lncRNAs in ESCC progression should be investigated in further experimental studies. The functional annotation found the signature may play a role by participating in RNA metabolism and immune regulation.

TNM staging is a commonly used tumor classification standard in clinical practice and a recognized prognostic marker^[30,31]. However, TNM staging remains flawed in prognostic assessment. We found that the prognostic predictive power of signature is better than TNM staging, suggesting that the strong prognostic ability of the five-lncRNA signature. Consistent with some scholars' results that combined TNM classification with molecular marker can predict outcome of ESCC patients more accurately^[32], we found the prognostic prediction of the combination of signature with TNM staging was the best, indicating the signature combined with TNM staging is useful for prognosis evaluation.

Conclusion

We identified a prognostic five-lncRNA signature from a large cohort of ESCC patients. The signature could predict the survival of patients with ESCC based on lncRNA expression profile and have strong clinical value.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The data generated during this study are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by grant from National Natural Science Foundation of China (No. 81272371, U1904148), Outstanding talents in science and technology innovation in Henan Province (No. 184200510013), the National Science and Technology Major Project of China (No. 2018ZX10302205), Zhengzhou Major Project for Collaborative Innovation (Zhengzhou University, No. 18XTZX12007), the Major Project of Science and Technology in Henan Province (No.161100311400), the Key Scientific and Technological Research Projects in Henan Province (No.192102310050), and the Key Research Projects of Henan Higher Education (No.16A320053) & (No.18A310033), and the Youth Innovation Fund of The First Affiliated Hospital of Zhengzhou University to Wei-Wei Wang & Li Wang.

Authors' Contributions

The authors contributed in the following way: Lan zhang: data collection, data analysis, interpretation, and drafting; Guozhong Jiang, Weiwei Wang: study design, study supervision, and final approval of the manuscript; Pan Li, Chenju Xing, Di Zhu, Jianying Zhang: technical support, experiment advice and data collection. All authors read and approved the final manuscript.

Acknowledgements

The authors gratefully thank the academic editor and the anonymous reviewers for their insightful comments and suggestions to improve this manuscript.

Data availability statement

The data that support the findings of this study are available in Gene Expression Omnibus database (GEO). These data were derived from the following resources available in the public domain: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE53624>.

Disclosure of Potential Conflicts of Interest

None

Abbreviations

ESCC: esophageal squamous cell carcinoma; qPCR: Real-time quantitative PCR; ROC: receiver operating characteristic; AUC: area under the ROC curve; CI: confidence interval; OS: overall survival.

References

- [1] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. *CA: a cancer journal for clinicians*, 2018,68(6):394-424.
- [2] Malhotra GK, Yanala U, Ravipati A, et al. Global trends in esophageal cancer [J]. *Journal of surgical oncology*, 2017,115(5):564-579.
- [3] Chen W, Zheng R, Baade PD, et al. Cancer statistics in china, 2015 [J]. *CA: a cancer journal for clinicians*, 2016,66(2):115-132.
- [4] Kou K, Guo X, Baade P, et al. Spatial analysis of esophageal cancer mortality in a high-risk population in china: Consistent clustering pattern in 1970-74 and 2011-13 [J]. *Asian Pacific journal of cancer prevention : APJCP*, 2018,19(11):3161-3166.

- [5] Aghcheli K, Marjani HA, Nasrollahzadeh D, et al. Prognostic factors for esophageal squamous cell carcinoma—a population-based study in golestan province, iran, a high incidence area [J]. PloS one, 2011,6(7):e22152.
- [6] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015 [J]. CA: a cancer journal for clinicians, 2015,65(1):5-29.
- [7] LI Xuemin TS, PENG Xiuqing, HUANG Jia, WANG Weipeng, ZHANG Peng,ZHAO Xueke, SONG Xin, WANG Ran, FAN Zongmin, LEI Xiaoyu,WANG Lidong. Changes of five-year survival rates in patients with esophageal squamous cell carcinoma from 1973 to 2005 [J]. Cancer Research on Prevention and Treatment, 2014,41(03):237-242.
- [8] Liu J, Xie X, Zhou C, et al. Which factors are associated with actual 5-year survival of oesophageal squamous cell carcinoma? [J]. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery, 2012,41(3):e7-11.
- [9] Zeng H, Zheng R, Zhang S, et al. Esophageal cancer statistics in china, 2011: Estimates based on 177 cancer registries [J]. Thoracic cancer, 2016,7(2):232-237.
- [10] Chen W, Zheng R, Baade PD, et al. Cancer statistics in china, 2015 [J]. CA: a cancer journal for clinicians, 2016.
- [11] Patel N, Benipal B. Incidence of esophageal cancer in the united states from 2001-2015: A united states cancer statistics analysis of 50 states [J]. Cureus, 2018,10(12):e3709.
- [12] Tripathy D, Harnden K, Blackwell K, et al. Next generation sequencing and tumor mutation profiling: Are we ready for routine use in the oncology clinic? [J]. BMC medicine, 2014,12(140).
- [13] Biesecker LG, Burke W, Kohane I, et al. Next-generation sequencing in the clinic: Are we ready? [J]. Nature reviews Genetics, 2012,13(11):818-824.
- [14] Guo JC, Fang SS, Wu Y, et al. Cnit: A fast and accurate web tool for identifying protein-coding and long non-coding transcripts based on intrinsic sequence composition [J]. Nucleic acids research, 2019,47(W1):W516-W522.
- [15] Wu Y, Zhang F, Yang K, et al. Symmap: An integrative database of traditional chinese medicine enhanced by symptom mapping [J]. Nucleic acids research, 2019,47(D1):D1110-D1117.
- [16] Loewen G, Jayawickramarajah J, Zhuo Y, et al. Functions of lncrna hotair in lung cancer [J]. Journal of hematology & oncology, 2014,7(90).
- [17] Yari H, Jin L, Teng L, et al. Lncrna reg1cp promotes tumorigenesis through an enhancer complex to recruit fancj helicase for reg3a transcription [J]. Nature communications, 2019,10(1):5334.

- [18] Li X, Yang H, Wang J, et al. High level of lncrna h19 expression is associated with shorter survival in esophageal squamous cell cancer patients [J]. *Pathology, research and practice*, 2019,215(11):152638.
- [19] Liang Y, Chen X, Wu Y, et al. Lncrna casc9 promotes esophageal squamous cell carcinoma metastasis through upregulating lamc2 expression by interacting with the creb-binding protein [J]. *Cell death and differentiation*, 2018,25(11):1980-1995.
- [20] Yoon JH, You BH, Park CH, et al. The long noncoding rna lucat1 promotes tumorigenesis by controlling ubiquitination and stability of DNA methyltransferase 1 in esophageal squamous cell carcinoma [J]. *Cancer letters*, 2018,417(47-57).
- [21] Li J, Chen Z, Tian L, et al. Lncrna profile study reveals a three-lncrna signature associated with the survival of patients with oesophageal squamous cell carcinoma [J]. *Gut*, 2014,63(11):1700-1710.
- [22] Mao Y, Fu Z, Zhang Y, et al. A seven-lncrna signature predicts overall survival in esophageal squamous cell carcinoma [J]. *Scientific reports*, 2018,8(1):8823.
- [23] Guo JC, Xie YM, Ran LQ, et al. L1cam drives oncogenicity in esophageal squamous cell carcinoma by stimulation of ezrin transcription [J]. *Journal of molecular medicine*, 2017,95(12):1355-1368.
- [24] Guo JC, Li CQ, Wang QY, et al. Protein-coding genes combined with long non-coding rnas predict prognosis in esophageal squamous cell carcinoma patients as a novel clinical multi-dimensional signature [J]. *Molecular bioSystems*, 2016,12(11):3467-3477.
- [25] Zhang XD, Huang GW, Xie YH, et al. The interaction of lncrna ezr-as1 with smyd3 maintains overexpression of ezr in escc cells [J]. *Nucleic acids research*, 2018,46(4):1793-1809.
- [26] Guo JC, Wu Y, Chen Y, et al. Protein-coding genes combined with long noncoding rna as a novel transcriptome molecular staging model to predict the survival of patients with esophageal squamous cell carcinoma [J]. *Cancer communications*, 2018,38(1):4.
- [27] Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2005,102(43):15545-15550.
- [28] Bhan A, Soleimani M, Mandal SS. Long noncoding rna and cancer: A new paradigm [J]. *Cancer research*, 2017,77(15):3965-3981.
- [29] Slack FJ, Chinnaiyan AM. The role of non-coding rnas in oncology [J]. *Cell*, 2019,179(5):1033-1055.
- [30] Sobin LH, Fleming ID. Tnm classification of malignant tumors, fifth edition (1997). *Union internationale contre le cancer and the american joint committee on cancer* [J]. *Cancer*, 1997,80(9):1803-1804.

[31] Sobin LH, Hermanek P, Hutter RV. Tnm classification of malignant tumors. A comparison between the new (1987) and the old editions [J]. Cancer, 1988,61(11):2310-2314.

[32] Takeno S, Noguchi T, Takahashi Y, et al. Assessment of clinical outcome in patients with esophageal squamous cell carcinoma using tnm classification score and molecular biological classification [J]. Annals of surgical oncology, 2007,14(4):1431-1438.

Tables

Table 1 Summary of ESCC patients and clinical characteristics in this study

Characteristic	GSE53624	RNA seq	qPCR validation
group	training	test	validation
Age (years)			
> 61	39	49	52
≤ 61	80	49	32
Sex			
female	21	32	36
Male	98	66	48
Vital status			
Living	46	46	59
Dead	13	52	25
Pathological stage			
Stage I	6	3	4
Stage II	47	51	51
Stage III	66	40	26
Stage IV	0	4	3

Table 2. lncRNAs in the prognostic signature of ESCC.

Gene symbol	Coefficient ^a	P value ^a	KM P value	Gene expression level association with prognosis
RP13-30A9.2	-0.635	0.008	0.007	Low
RP11-488I20.9	-0.468	0.048	0.046	Low
MORF4L2-AS1	0.582	0.014	0.013	High
AC007179.1	0.470	0.047	0.045	High
RP4-735C1.6	0.616	0.010	0.008	High

^aDerived from the univariable Cox regression analysis in the GSE53624 set.

Table 3. Association of the signature with clinicopathological characteristics in ESCC patients

Variables	GSE53624		P	RNA-seq		P	qPCR		P
	Low *	High *		Low *	High *		Low *	High*	
Age			0.217			0.106			0.261
≤ 62	44	36		29	20		13	19	
> 62	16	23		20	29		29	23	
Sex			1.000			0.132			0.825
Female	11	10		20	12		17	19	
Male	49	49		29	37		25	23	
TNM stage			0.654			0.588			0.681
Stage I	2	4		1	2		3	1	
Stage II	25	22		28	23		25	26	
Stage III	33	33		19	21		12	14	
Stage IV				1	3		2	1	

* Low risk ≤Median of risk score, High risk >Median of risk score; The Chi-squared test P value <0.05 was considered significant.

Table 4. Cox regression analysis of the signature for the ESCC patients from the three datasets.

Variables		Univariable analysis			Multivariable analysis				
		HR	95% CI of HR		HR	95% CI of HR		P	
			lower	upper		lower	upper		
GSE53624 dataset(n=119)									
Age	>61 vs.≤61	1.643	1.024	2.636	0.040	1.441	0.894	2.322	0.134
Sex	Male vs. Female	0.827	0.468	1.461	0.513	0.782	0.440	1.391	0.403
TNM stage	III, IV vs. I, II	1.901	1.226	2.948	0.004	2.017	1.288	3.159	0.002
lncRNA signature	High risk vs. low risk	3.264	1.998	5.331	<0.001	3.502	2.131	5.756	<0.001
RNA seq set (n=98)									
Age	>61 vs.≤61	1.287	0.744	2.225	0.367	1.314	0.711	2.430	0.383
Sex	Male vs. Female	1.534	0.818	2.875	0.182	1.116	0.561	2.221	0.754
TNM stage	III, IV vs. I, II	1.795	1.152	2.796	0.010	1.629	0.887	2.991	0.115
lncRNA signature	High risk vs. low risk	1.994	1.139	3.491	0.016	1.032	1.002	1.062	0.035
independent qPCR validation set (n=84)									
Age	>61 vs.≤61	0.949	0.892	1.011	0.105	0.961	0.893	1.034	0.289
Sex	Male vs. Female	0.797	0.352	1.804	0.585	0.872	0.381	1.997	0.747
TNM stage	III, IV vs. I, II	2.699	1.607	4.534	0.000	4.910	2.133	11.302	<0.001
lncRNA signature	High risk vs. low risk	2.981	1.235	7.192	0.015	2.945	1.210	7.168	0.017

Figures

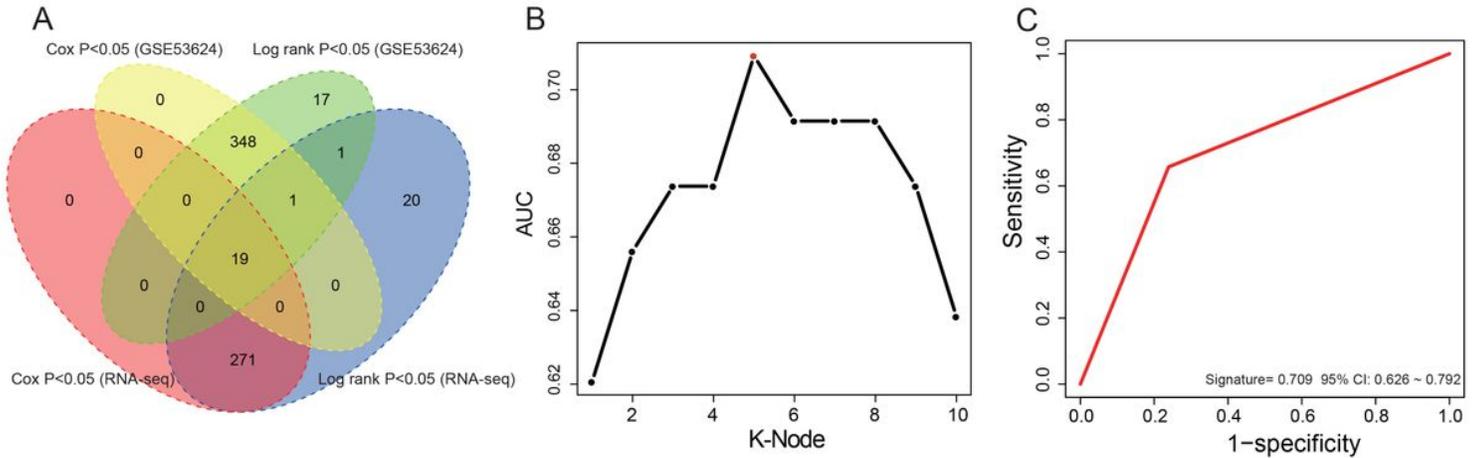


Figure 1

Constructing the prognostic lncRNA model in the high throughput sequencing datasets. (A) Venn diagram for analyzing the prognostic lncRNAs. (B) Screening out the lncRNA signature with largest AUC from all 1023 signatures which were calculated by ROC for k=1, 2,..... 10. (C) The AUC of the screened lncRNA signature.

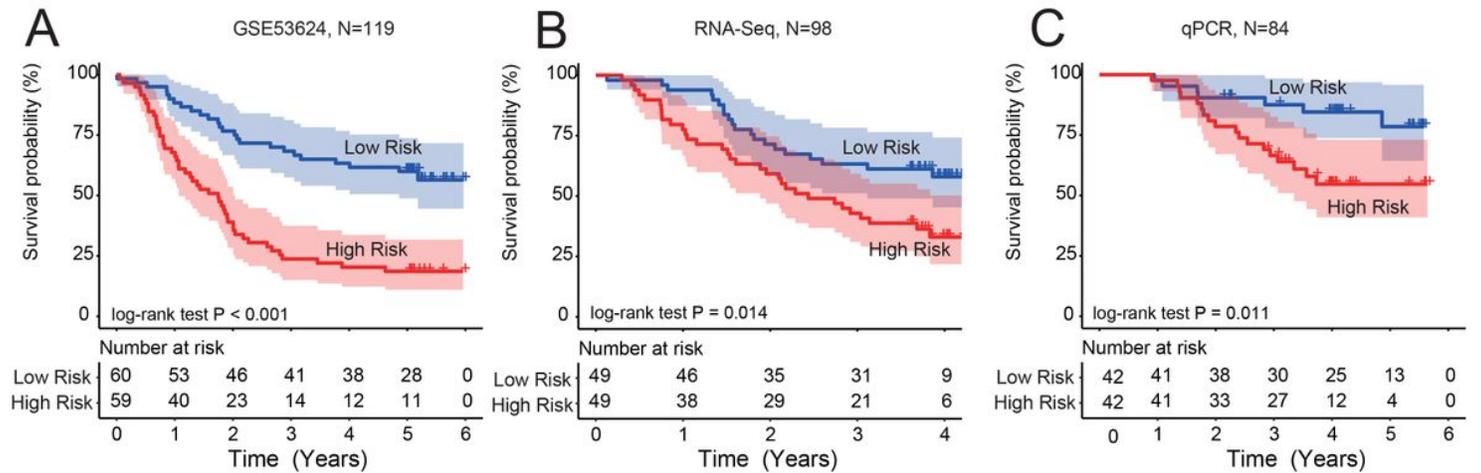


Figure 2

The lncRNA signature classification power for ESCC prognosis. Kaplan–Meier curves found ESCC patients were classified into two different risk groups based on the risk score of the signature in the GSE53624 (A), RNA-seq (B) and qPCR validation datasets(C).

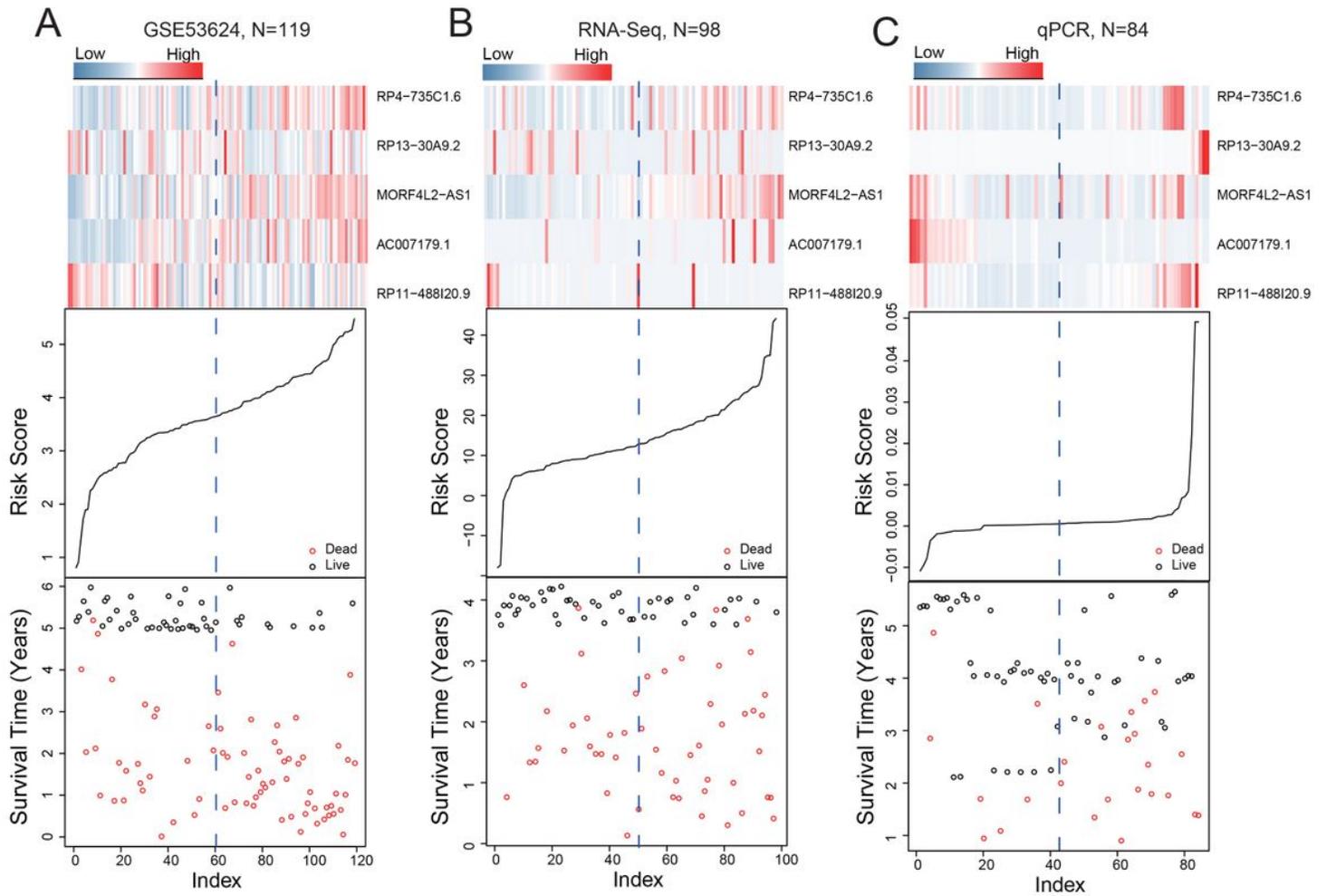


Figure 3

Risk score distribution, survival status and gene expression of ESCC patients in high- and low-risk groups classified by the four-lncRNA signature in the GSE53624 (A), RNA-seq (B) dataset and qPCR validation set (C).

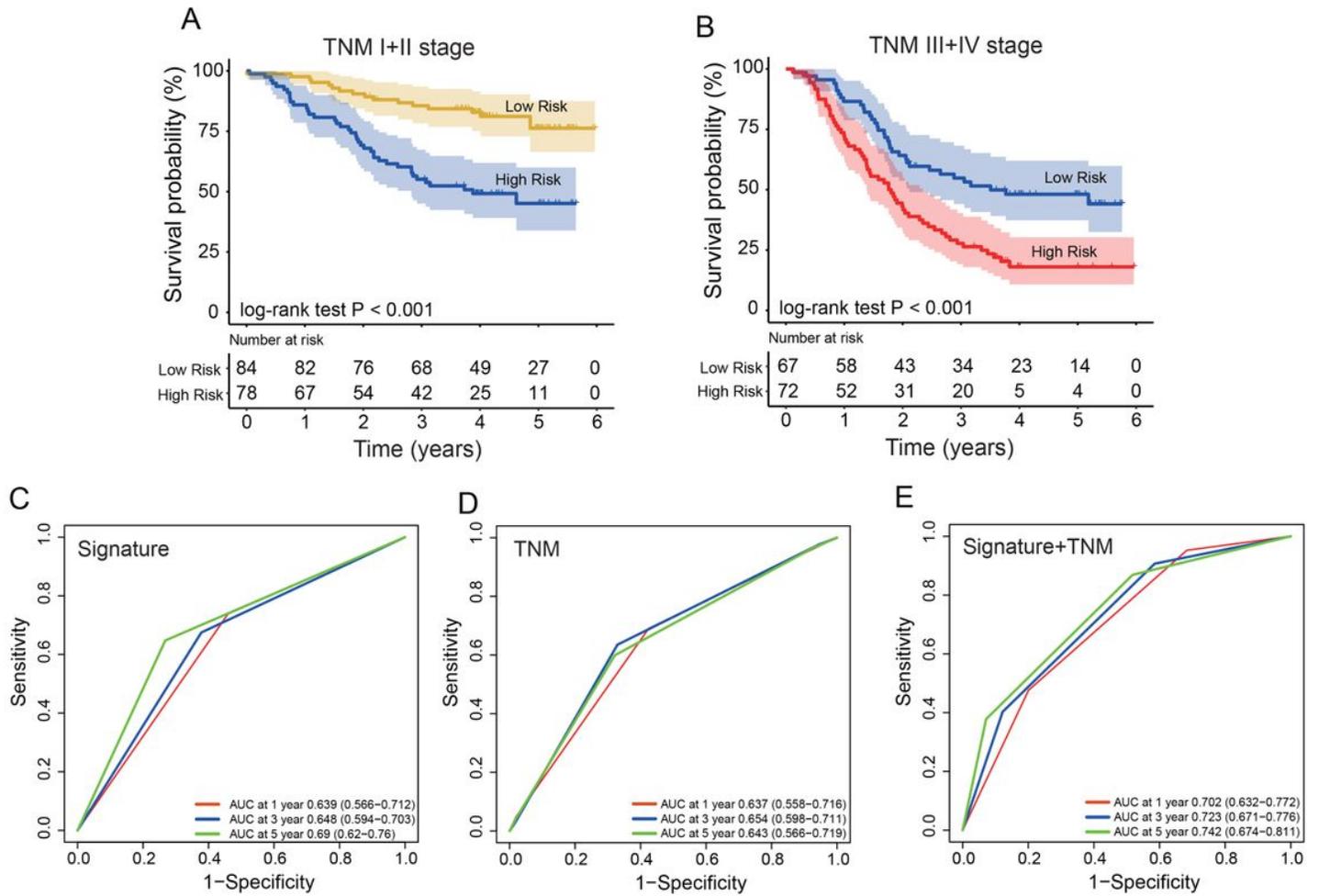


Figure 4

Stratification analysis of TNM stage using the five-lncRNA signature. (A, B) Stratification analysis of TNM low/high stage using the signature by Kaplan–Meier curves. (C,D) Comparing the survival prediction power of the signature with that of TNM stage by ROC in the entire datasets. (E) TimeROC analysis to explore the survival prediction power of combination of the signature and TNM stage in the entire dataset.

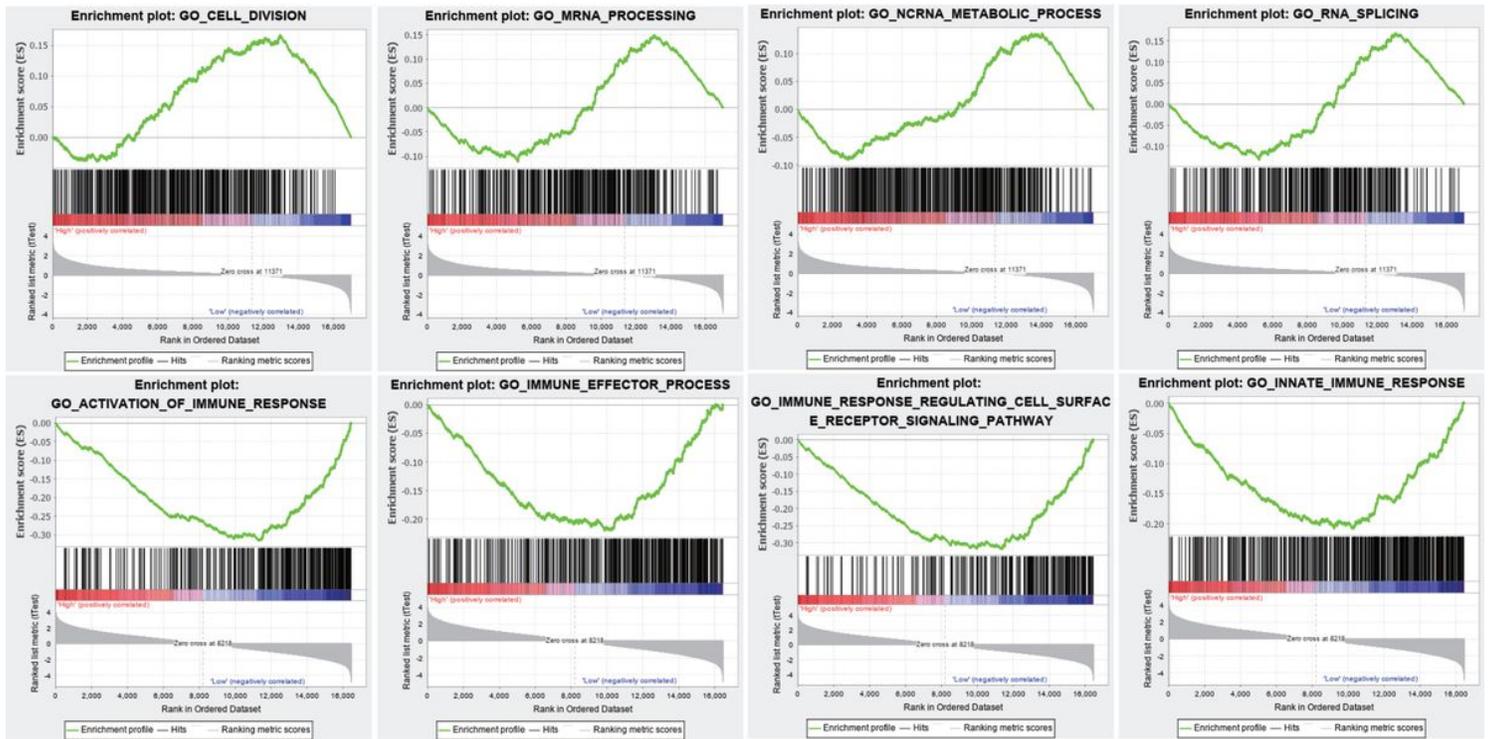


Figure 5

GSEA functional analysis of the six-lncRNA signature.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementarytableS4GSEA.xlsx](#)
- [SupplementarytableS1N.xlsx](#)
- [SupplementarytableS2.xlsx](#)
- [SupplementarytableS3.xlsx](#)