Development of a novel signature of long noncoding RNAs as a prognostic biomarker for esophageal cancer

Yan Miao¹, Jing Sui¹, Ying Zhang¹, Lihong Yin^{1,*}

¹ Key Laboratory of Environmental Medicine Engineering, Ministry of Education, School of Public Health, Southeast University, Nanjing, Jiangsu 210009, P.R. China

*Correspondence Author:

Lihong Yin

Email address: lh.yin62@gmail.com

ABSTRACT

Objectives. This study aims to develop a lncRNA signature based on RNA-Seq data to predict overall survival in esophageal cancer patients.

Methods. The lncRNA expression profiles and clinical data were downloaded from The Cancer Genome Atlas (TCGA) database on August 30, 2017. Differentially expressed lncRNAs were screened out between tumor tissues and adjacent normal tissues. The univariate and multivariate Cox regression models were used to develop a prognostic signature for all esophageal cancer patients. The receiver operating curve (ROC) was used to test the sensitivity and specificity of lncRNA signature. Survivals were compared via log-rank test. GO and KEGG enrichment analyses were used to explore the potential functions of prognostic lncRNAs.

Results. We identified two lncRNAs (RPL34-AS1 and GK3P) were significantly associated with the overall survival of the total 150 esophageal cancer patients. A novel two-lncRNA signature was constructed by Cox regression models. Signature low-risk cases showed better overall survival (median 625.560 days vs. 478.000 days, $p = 0.002$) than high-risk cases. Further analysis suggested that this two-lncRNA signature was independent of clinical characteristics. GO functional and KEGG pathway enrichment analyses revealed potential functional roles of the two prognostic lncRNAs in tumorigenesis.

Conclusions. Our findings suggest that the two-lncRNA signature may be a useful prognostic biomarker for predicting overall survival in esophageal cancer patients.

INTRODUCTION

Esophageal cancer (EC) ranks the eighth most common cancer worldwide and the sixth leading cause of cancer-related mortality (Ferlay et al. 2015). Unfortunately, EC is often associated with a poor prognosis that having an 18% overall five-year survival (Malhotra et al. 2017). As with most solid tumors, pathological tumor-node-metastasis (TNM) staging is still the primary indicator of survival in patients with EC (Malhotra et al. 2017; Rice et al. 2017). The molecular heterogeneity and complexity of EC make it difficult to predict its clinical outcomes (Hao et al. 2016; Hardie et al. 2005). Therefore, it is urgent to find specific prognostic factors of EC which would improve treatment efficacies for patients.

In the age of genomics, researchers have discovered tens of thousands of long noncoding RNAs (lncRNAs) by genome-wide techniques, which are defined as RNA transcripts exceeding 200 nucleotides in length without protein-coding capacity (Schmitt & Chang 2016; Young & Ponting 2013). Emerging evidence revealed that lncRNAs play a critical role in various aspects of cellular processes, including proliferation, migration, invasion, and genomic stability (Iyer et al. 2015; Ling et al. 2015). Based on the expression specificity among different tumor types, lncRNAs could serve as the basis for many clinical applications in oncology (Li & Chen 2013; Shi et al. 2013; Wahlestedt 2013). Till now, lncRNA biomarkers for diagnosis of EC have been reported in many studies (Tong et al. 2015; Wang et al. 2017; Wu et al. 2015). However, the understanding of the prognostic value of lncRNAs in patients with EC from previous studies was limited.

The Cancer Genome Atlas (TCGA) is a database, which provides a collection of DNA copy-number variation, DNA methylation, RNA sequence, and corresponding clinical data for the top 25 tumor types. In this study, we used the RNA-Seq dataset from the TCGA database to identify candidate prognostic lncRNAs for EC. After integrative analysis, we developed a prognostic lncRNA signature for the overall survival (OS) prediction in EC patients.

MATERIALS & METHODS

TCGA data source

185 EC patients' data were downloaded from the TCGA database on August 30, 2017. The exclusion criteria were listed in the following: i) histologic diagnosis ruled out EC; ii) another malignancy besides EC. After that, 150 EC patients were included in this study. As the data were downloaded from the public database, ethical approval was not applicable here. Data processing procedures followed the guidelines of TCGA data access (<http://cancergenome.nih.gov/publications/publicationguidelines>).

LncRNA data mining and processing

The level 3 RNA-Seq data were obtained from the TCGA database. Raw data of lncRNA sequencing were processed and normalized by TCGA RNASeqv2 system (Zand et al. 2016). Here, only lncRNAs with the description in NCBI (<http://ncbi.nlm.nih.gov/gene>) or Ensembl (<http://ensembl.org/index.html>) were selected for further analysis. To identify differentially expressed lncRNAs, the expression of lncRNAs in tumor tissues were screened and compared with that in adjacent non-tumor tissues. The candidate lncRNAs were gathered for further analysis.

Construction of the prognostic signature

The expression profile of each lncRNA was normalized by log2 transformation for statistical analysis. The univariate Cox regression model was used to evaluate the EC-specific lncRNAs that were associated with OS. The multivariate Cox regression model was used to evaluate the prognostic value of these OS-related lncRNAs. The risk-score model was constructed based on the combination of the expression profiles of each prognostic lncRNA, weighted by the regression coefficients that were calculated by multivariate Cox regression analysis. The formula was listed as follows: $risk\ score = exp_{lncRNA_1} * \beta_{lncRNA_1} + exp_{lncRNA_2} * \beta_{lncRNA_2} + \dots + exp_{lncRNA_n} * \beta_{lncRNA_n}$ (exp : expression level; β : regression coefficient derived from the multivariate Cox regression model) (Zeng et al. 2017). The median risk score was set to the cutoff point, and EC patients were divided into high or low groups (Zhou et al. 2016). Further univariate and multivariate Cox regression models were used to investigate the effects

of risk score and clinical characteristics in EC patients. The hazard ratio (HR) and the 95% confidence interval (CI) were used for quantification. The time-dependent receiver operating curve (ROC) within five years was used to assess the predictive value of risk score for time-dependent outcomes (Heagerty et al. 2000). The Kaplan-Meier (K-M) curve and log-rank test were used to assess the differences in survival distribution among different groups. Here, the p value <0.05 was considered to be significant. IBM SPSS Statistics 22.0 (IBM Corp., Armonk, NY, USA) was used to perform the analyses mentioned above.

Functional enrichment analysis

To investigate the biological features of these two lncRNAs, we selected the genes that were strongly correlated with these two lncRNAs expressions (Pearson $|R| > 0.5$) from TCGA database. Pathways and biological processes were predicted by functional enrichment analysis of the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) in the DAVID database (<http://david.ncifcrf.gov/>). The p -value <0.05 was considered to be significant. Afterward, the protein-protein interaction (PPI) network was constructed based on the co-expressed genes via the STRING database (<http://string-db.org/>).

RESULTS

Patient characteristics

There were 150 EC patients included in this study. According to the cancer staging standards of the American Joint Committee on Cancer (AJCC) (Amin & Edge 2017), the EC patients were classified into four groups: stage I (16 cases), stage II (74 cases), stage III (52 cases), and stage IV (8 cases). The mean (and standard deviation) age of all EC patients was 61.233 (± 11.311) years. The mean OS was 551.780 (± 505.185) days. 61 of 150 (40.667%) patients died during follow-up.

Identification of differentially expressed lncRNAs

269 lncRNAs were identified as differentially expressed in EC tissues and adjacent non-tumor tissues, including 152 lncRNAs in stage I, 219 lncRNAs in stage II, 183 lncRNAs in stage III, and 109 lncRNAs in stage IV. We used “Fold change > 2 or < 0.5 , $p < 0.05$,

and $FDR < 0.05$ ” as the screening method to identify the significant differentially expressed lncRNAs. Then, 74 candidate lncRNAs were identified as differentially expressed in all tumor stages (Fig. 1 and Fig. 2).

Construction of a prognostic signature for EC patients

Based on the 74 lncRNAs and the clinical information extracted from TCGA database, two candidate lncRNAs were identified to be significantly associated with OS ($p < 0.05$) in EC patients by univariate Cox regression analysis (Table 1). The multivariate Cox regression analysis was used to verify the correlation between two candidate lncRNAs and OS (Table 1 and Fig 3A). These two lncRNAs, RPL34-AS1 and GK3P, also showed prognostic value for EC patients (Fig 3B).

A risk-score formula was established based on the combination of the expression profiles of each prognostic lncRNA, weighted by the regression coefficients that were calculated by multivariate Cox regression analysis. The formula was listed as follows: $Risk\ score = exp_{RPL34-AS1} * (-0.685) + exp_{GK3P} * (0.697)$. We calculated the risk score of the two-lncRNA expression signature for each patient.

Based on the median risk score as a cutoff value, all EC patients were divided into two groups: the low-risk score group ($n = 75$) and the high-risk score group ($n = 75$) (Fig 4). Meanwhile, K-M curves confirmed that the mean survival time (and standard deviation) of patients in the low-risk score group was 625.560 (± 507.079) days, significantly better than patients in the high-risk score group (478.000 $\pm$ 427.816 days, $p = 0.002$) (Fig 5A). Furthermore, this risk score model could be used to predict 5-year survival of EC patients as the area under ROC curve was 0.647 (Fig 5B).

The significant expression patterns of these two candidate lncRNAs in the tumor tissues/adjacent normal tissues group (Fig 6A), and in the low-risk score/high-risk score group were presented in Figure 6B.

Correlation between the two-lncRNA signature and clinical characteristics

To verify the correlation between the two-lncRNA signature and clinical characteristics in EC patients, we used univariate Cox regression analysis and multivariate Cox regression analysis.

The univariate Cox regression analysis indicated that gender, TNM stage, N stage,

M stage, additional treatment completion success outcome, residual tumor, and neoplasm tumor status could significantly predict poorer survival of EC patients ($p < 0.05$, Table 2). The multivariate Cox regression analysis showed that neoplasm tumor status ($p = 0.036$) and risk score ($p = 0.001$) could serve as independent prognostic factors in EC patients (Table 2).

The K-M curves of these clinical characteristics indicated that TNM stage ($p < 0.001$), N stage ($p = 0.001$), M stage ($p < 0.001$), and residual tumor ($p = 0.005$), and neoplasm tumor status ($p = 0.001$) were significantly associated with OS of EC patients (Fig 7A). Moreover, we assessed the association between the risk score and clinical characteristics and found that the risk score might have the prognostic significance in predicting some clinical characteristics, including TNM stage (AUC = 0.643, $p = 0.040$) and N stage (AUC = 0.579, $p = 0.032$) (Fig 7B).

Functional assessment of the two-lncRNA signature

629 genes (extracted from TCGA database) were co-expressed with these two lncRNAs (Pearson $|R| > 0.5$), including 622 genes co-expressed with RPL34-AS1 and seven genes co-expressed with GK3P (Supplemental Table 1). We identified 89 enriched GO terms and seven pathways ($p < 0.05$, enrichment score > 1.5 , Supplemental Table 2) by gene enrichment and functional annotation analysis. The top GO biological process and KEGG pathway (Table 3 and Fig 8) were detection of chemical stimulus involved in sensory perception of smell (GO: 0050911) and olfactory transduction (hsa: 04740), respectively. Finally, we constructed a PPI network of 283 genes that were considered as hub genes (Fig 9).

DISCUSSION

Esophageal cancer (EC) is among the deadliest malignancies in the world, with high morbidity and mortality (Griffin & Berrisford 2013; Torre et al. 2015). The onset of dysphagia is associated symptom of EC, and a small proportion of patients with EC survive for five years (di Pietro et al. 2017). The novel biomarker for early diagnosis, process monitoring, and prognostic evaluation might increase the survival rate for EC patients. In the past decade, researchers have reported that about 70% of the genome is

transcribed in various cell types and contexts (Djebali et al. 2012; Okazaki et al. 2002), nearly 80% of the genome is biochemically active (2012), and many DNAs code for RNA as end products instead of proteins (Pennisi 2012). LncRNAs are engaged in a wide range of cellular mechanisms, such as DNA methylation and histone modifications (O'Leary et al. 2015; Plath et al. 2003). In addition, increasing evidence has proved that lncRNAs function as a critical player in gene regulation and tend to be expressed in a more tissue- and cell-specific manner, which demonstrated the crucial advantages of lncRNAs as diagnostic or prognostic biomarkers in cancers (Bolha et al. 2017; Evans et al. 2016; Yang et al. 2014). Therefore, development of novel lncRNA-related biomarkers in prognostic prediction of EC patients may improve treatment decisions regarding the aggressiveness or recurrence of the disease. In the present study, we investigated the impact of lncRNAs on EC patients' survival, which were based on RNA sequencing technology.

EC usually presents late and carries a poor prognosis (Bird-Lieberman & Fitzgerald 2009; Meves et al. 2015). Therefore, we aimed to find a novel lncRNA signature and to investigate whether it affects the prognosis of all stages of EC patients. In this study, we first screened out the differentially expressed lncRNAs in tumor tissues, by comparing with the expression in non-tumor tissues. Among candidate lncRNAs, we identified two lncRNAs as having significant prognostic value for survival in all EC patients by using univariate and multivariate Cox regression models. Then we constructed a risk-score model by combining the expressions of these two lncRNAs and found that the two-lncRNA expression signature could predict OS for all EC patients.

Recently, Fan and Liu identified an eight-lncRNA signature (GS1-600G8.5, CTD-2357A8.3, LINC00365, RP11-705O24.1, RP1-90J4.1, RP11-327J17.1, LINC01554, and LINC00176) which might be able to predict survival of EC patients (Fan & Liu 2016). Compared to their work, our study used “Fold change > 2 or < 0.5, $p < 0.05$, and FDR < 0.05” as the screening criteria, which was more rigorous for bioinformatics analysis. Therefore, the number of candidate prognostic lncRNAs in both studies were different. Moreover, we combined RNA sequencing data with clinical information and

found that this two-lncRNA signature could be served as an indicator of specific clinical characteristics (TNM stage and N stage) for EC patients.

A series of evidence-based papers have proved that lncRNA could serve as a carcinogen or tumor suppressor of EC. Despite this, the role of most lncRNAs in EC remains unknown. Lin et al. reported that HOXA13 might promote metastasis and tumorigenesis in EC cells (Lin et al. 2017). Yao et al. found that RP11-766N7.4 was associated with carcinogenesis of EC (Yao et al. 2017). Furthermore, Wu et al. suggested that NORAD was related to poor prognosis in EC patients (Wu et al. 2017).

Regarding these two lncRNAs, Zhao et al. reported that the decreased expression of RPL34-AS1 was correlated with larger gastric tumors (Zhao et al. 2015). For GK3P, it has not been reported yet. Besides, the biological functions of these two lncRNAs have not been investigated in EC. Here, we assessed the functional relevance of these two lncRNAs by using DAVID bioinformatics resources. 629 genes were identified to have significant co-expression with these two lncRNAs. The relevant genes were mainly enriched in olfactory transduction, neuroactive ligand-receptor interaction, detection of chemical stimulus involved in sensory perception of smell, and defense response to bacterium. Additionally, 283 genes were identified as hub genes regulated by these two lncRNAs via the PPI network.

The findings of our study may have essential clinical significance. However, some limitations should be taken into consideration as well. Firstly, the patient cohort of TCGA was obtained from multi-institutional sites, and the results still need to be validated in other cohorts in the future study. Secondly, only lncRNAs were involved in our work, and this study could not represent the whole transcription alteration associated with EC. Thirdly, the functional role of RPL34-AS1 and GK3P are still unknown. Thus, we need well-designed experiments to validate the biological functions of these two lncRNAs in EC.

CONCLUSIONS

We identified two lncRNAs associated with the survival of EC patients in a large cohort

from the TCGA database and developed a lncRNA signature. Further analysis indicated that the two-lncRNA signature could be a prognostic indicator independent of clinical characteristics. It can serve as a novel biomarker of prognostic prediction for EC patients. Further functional validations are needed to explore the regulatory mechanisms of these lncRNAs in EC.

ACKNOWLEDGMENTS

We thank Mr. Dong-Ling Chen for his help in technical assistance.

REFERENCES

- ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489:57-74. 10.1038/nature11247
- Amin MB, and Edge SB. 2017. *AJCC cancer staging manual*: Springer.
- Bird-Lieberman EL, and Fitzgerald RC. 2009. Early diagnosis of oesophageal cancer. *British Journal of Cancer* 101:1-6. 10.1038/sj.bjc.6605126
- Bolha L, Ravnik-Glavač M, and Glavač D. 2017. Long Noncoding RNAs as Biomarkers in Cancer. *Disease Markers* 2017:7243968. 10.1155/2017/7243968
- di Pietro M, Canto MI, and Fitzgerald RC. 2017. Endoscopic Management of Early Adenocarcinoma and Squamous Cell Carcinoma of the Esophagus-Screening, Diagnosis, and Therapy. *Gastroenterology*. 10.1053/j.gastro.2017.07.041
- Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Roder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Bar NS, Batut P, Bell K, Bell I, Chakraborty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Falconnet E, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena H, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Luo OJ, Park E, Persaud K, Preall JB, Ribeca P, Risk B, Robyr D, Sammeth M, Schaffer L, See LH, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Wrobel J, Yu Y, Ruan X, Hayashizaki Y, Harrow J, Gerstein M, Hubbard T, Reymond A, Antonarakis SE, Hannon G, Giddings MC, Ruan Y, Wold B, Carninci P, Guigo R, and Gingeras TR. 2012. Landscape of transcription in human cells. *Nature* 489:101-108. 10.1038/nature11233
- Evans JR, Feng FY, and Chinnaiyan AM. 2016. The bright side of dark matter: lncRNAs in cancer. *J Clin Invest* 126:2775-2782. 10.1172/jci84421
- Fan Q, and Liu B. 2016. Identification of a RNA-Seq Based 8-Long Non-Coding RNA Signature Predicting Survival in Esophageal Cancer. *Med Sci Monit* 22:5163-5172. 10.12659/MSM.902615
- Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, 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. 10.1002/ijc.29210
- Griffin SM, and Berrisford RG. 2013. Surgery for cancer of the oesophagus. *Oesophagogastric Surg Saunders*:81-106.
- Hao JJ, Lin DC, Dinh HQ, Mayakonda A, Jiang YY, Chang C, Jiang Y, Lu CC, Shi ZZ, Xu X, Zhang Y, Cai Y, Wang JW, Zhan QM, Wei WQ, Berman BP, Wang MR, and Koeffler HP. 2016. Spatial intratumoral heterogeneity and temporal clonal evolution in esophageal squamous cell carcinoma. *Nat Genet* 48:1500-1507. 10.1038/ng.3683
- Hardie LJ, Darnton SJ, Wallis YL, Chauhan A, Hainaut P, Wild CP, and Casson AG. 2005. p16 expression in Barrett's esophagus and esophageal adenocarcinoma: association with genetic and epigenetic alterations. *Cancer Lett* 217:221-230. 10.1016/j.canlet.2004.06.025
- Heagerty PJ, Lumley T, and Pepe MS. 2000. Time-Dependent ROC Curves for Censored Survival Data and a Diagnostic Marker. *Biometrics* 56:337-344. 10.1111/j.0006-341X.2000.00337.x
- Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, Barrette TR, Prensner JR, Evans JR, Zhao

- S, Poliakov A, Cao X, Dhanasekaran SM, Wu YM, Robinson DR, Beer DG, Feng FY, Iyer HK, and Chinnaiyan AM. 2015. The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet* 47:199-208. 10.1038/ng.3192
- Li CH, and Chen Y. 2013. Targeting long non-coding RNAs in cancers: progress and prospects. *Int J Biochem Cell Biol* 45:1895-1910. 10.1016/j.biocel.2013.05.030
- Lin C, Wang Y, Wang Y, Zhang S, Yu L, Guo C, and Xu H. 2017. Transcriptional and posttranscriptional regulation of HOXA13 by lncRNA HOTTIP facilitates tumorigenesis and metastasis in esophageal squamous carcinoma cells. *Oncogene* 36:5392-5406. 10.1038/onc.2017.133
- Ling H, Vincent K, Pichler M, Fodde R, Berindan-Neagoe I, Slack FJ, and Calin GA. 2015. Junk DNA and the long non-coding RNA twist in cancer genetics. *Oncogene* 34:5003-5011. 10.1038/onc.2014.456
- Malhotra GK, Yanala U, Ravipati A, Follet M, Vijayakumar M, and Are C. 2017. Global trends in esophageal cancer. *J Surg Oncol* 115:564-579. 10.1002/jso.24592
- Meves V, Behrens A, and Pohl J. 2015. Diagnostics and Early Diagnosis of Esophageal Cancer. *Viszeralmedizin* 31:315-318. 10.1159/000439473
- O'Leary VB, Ovsepiyan SV, Carrascosa LG, Buske FA, Radulovic V, Niyazi M, Moertl S, Trau M, Atkinson MJ, and Anastasov N. 2015. PARTICLE, a Triplex-Forming Long ncRNA, Regulates Locus-Specific Methylation in Response to Low-Dose Irradiation. *Cell Rep* 11:474-485. 10.1016/j.celrep.2015.03.043
- Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S, Nikaido I, Osato N, Saito R, Suzuki H, Yamanaka I, Kiyosawa H, Yagi K, Tomaru Y, Hasegawa Y, Nogami A, Schonbach C, Gojobori T, Baldarelli R, Hill DP, Bult C, Hume DA, Quackenbush J, Schriml LM, Kanapin A, Matsuda H, Batalov S, Beisel KW, Blake JA, Bradt D, Brusica V, Chothia C, Corbani LE, Cousins S, Dalla E, Dragani TA, Fletcher CF, Forrest A, Frazer KS, Gaasterland T, Gariboldi M, Gissi C, Godzik A, Gough J, Grimmond S, Gustincich S, Hirokawa N, Jackson IJ, Jarvis ED, Kanai A, Kawaji H, Kawasawa Y, Kedzierski RM, King BL, Konagaya A, Kurochkin IV, Lee Y, Lenhard B, Lyons PA, Maglott DR, Maltais L, Marchionni L, McKenzie L, Miki H, Nagashima T, Numata K, Okido T, Pavan WJ, Pertea G, Pesole G, Petrovsky N, Pillai R, Pontius JU, Qi D, Ramachandran S, Ravasi T, Reed JC, Reed DJ, Reid J, Ring BZ, Ringwald M, Sandelin A, Schneider C, Semple CA, Setou M, Shimada K, Sultana R, Takenaka Y, Taylor MS, Teasdale RD, Tomita M, Verardo R, Wagner L, Wahlestedt C, Wang Y, Watanabe Y, Wells C, Wilming LG, Wynshaw-Boris A, Yanagisawa M, Yang I, Yang L, Yuan Z, Zavolan M, Zhu Y, Zimmer A, Carninci P, Hayatsu N, Hirozane-Kishikawa T, Konno H, Nakamura M, Sakazume N, Sato K, Shiraki T, Waki K, Kawai J, Aizawa K, Arakawa T, Fukuda S, Hara A, Hashizume W, Imotani K, Ishii Y, Itoh M, Kagawa I, Miyazaki A, Sakai K, Sasaki D, Shibata K, Shinagawa A, Yasunishi A, Yoshino M, Waterston R, Lander ES, Rogers J, Birney E, and Hayashizaki Y. 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* 420:563-573. 10.1038/nature01266
- Pennisi E. 2012. Genomics. ENCODE project writes eulogy for junk DNA. *Science* 337:1159, 1161. 10.1126/science.337.6099.1159
- Plath K, Fang J, Mlynarczyk-Evans SK, Cao R, Worringer KA, Wang H, de la Cruz CC, Otte AP, Panning B, and Zhang Y. 2003. Role of histone H3 lysine 27 methylation in X inactivation. *Science* 300:131-135. 10.1126/science.1084274
- Rice TW, Patil DT, and Blackstone EH. 2017. 8th edition AJCC/UICC staging of cancers of the

- esophagus and esophagogastric junction: application to clinical practice. *Annals of Cardiothoracic Surgery* 6:119-130. 10.21037/acs.2017.03.14
- Schmitt AM, and Chang HY. 2016. Long Noncoding RNAs in Cancer Pathways. *Cancer Cell* 29:452-463. 10.1016/j.ccell.2016.03.010
- Shi X, Sun M, Liu H, Yao Y, and Song Y. 2013. Long non-coding RNAs: a new frontier in the study of human diseases. *Cancer Lett* 339:159-166. 10.1016/j.canlet.2013.06.013
- Tong YS, Wang XW, Zhou XL, Liu ZH, Yang TX, Shi WH, Xie HW, Lv J, Wu QQ, and Cao XF. 2015. Identification of the long non-coding RNA POU3F3 in plasma as a novel biomarker for diagnosis of esophageal squamous cell carcinoma. *Mol Cancer* 14:3. 10.1186/1476-4598-14-3
- Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, and Jemal A. 2015. Global cancer statistics, 2012. *CA Cancer J Clin* 65:87-108. 10.3322/caac.21262
- Wahlestedt C. 2013. Targeting long non-coding RNA to therapeutically upregulate gene expression. *Nat Rev Drug Discov* 12:433-446. 10.1038/nrd4018
- Wang W, He X, Zheng Z, Ma X, Hu X, Wu D, and Wang M. 2017. Serum HOTAIR as a novel diagnostic biomarker for esophageal squamous cell carcinoma. *Mol Cancer* 16:75. 10.1186/s12943-017-0643-6
- Wu X, Lim ZF, Li Z, Gu L, Ma W, Zhou Q, Su H, Wang X, Yang X, and Zhang Z. 2017. NORAD Expression Is Associated with Adverse Prognosis in Esophageal Squamous Cell Carcinoma. *Oncol Res Treat* 40:370-374. 10.1159/000464465
- Wu Y, Liu X, Hu L, Tao H, Guan X, Zhang K, Bai Y, and Yang K. 2015. Copy number loss of variation_91720 in PIK3CA predicts risk of esophageal squamous cell carcinoma. *Int J Clin Exp Pathol* 8:14479-14485.
- Yang G, Lu X, and Yuan L. 2014. LncRNA: a link between RNA and cancer. *Biochim Biophys Acta* 1839:1097-1109. 10.1016/j.bbarm.2014.08.012
- Yao GL, Pan CF, Xu H, Wei K, Liu B, Zhai R, and Chen YJ. 2017. Long noncoding RNA RP11-766N7.4 functions as a tumor suppressor by regulating epithelial-mesenchymal transition in esophageal squamous cell carcinoma. *Biomed Pharmacother* 88:778-785. 10.1016/j.biopha.2017.01.124
- Young RS, and Ponting CP. 2013. Identification and function of long non-coding RNAs. *Essays Biochem* 54:113-126. 10.1042/bse0540113
- Zand B, Previs RA, Zacharias NM, Rupaimoole R, Mitamura T, Nagaraja AS, Guindani M, Dalton HJ, Yang L, Baddour J, Achreja A, Hu W, Pecot CV, Ivan C, Wu SY, McCullough CR, Gharpure KM, Shoshan E, Pradeep S, Mangala LS, Rodriguez-Aguayo C, Wang Y, Nick AM, Davies MA, Armaiz-Pena G, Liu J, Lutgendorf SK, Baggerly KA, Eli MB, Lopez-Berestein G, Negrath D, Bhattacharya PK, and Sood AK. 2016. Role of Increased n-acetylaspartate Levels in Cancer. *J Natl Cancer Inst* 108:djv426. 10.1093/jnci/djv426
- Zeng JH, Liang L, He RQ, Tang RX, Cai XY, Chen JQ, Luo DZ, and Chen G. 2017. Comprehensive investigation of a novel differentially expressed lncRNA expression profile signature to assess the survival of patients with colorectal adenocarcinoma. *Oncotarget* 8:16811-16828. 10.18632/oncotarget.15161
- Zhao J, Liu Y, Zhang W, Zhou Z, Wu J, Cui P, Zhang Y, and Huang G. 2015. Long non-coding RNA Linc00152 is involved in cell cycle arrest, apoptosis, epithelial to mesenchymal transition, cell migration and invasion in gastric cancer. *Cell Cycle* 14:3112-3123. 10.1080/15384101.2015.1078034
- Zhou X, Huang Z, Xu L, Zhu M, Zhang L, Zhang H, Wang X, Li H, Zhu W, Shu Y, and Liu P. 2016. A

panel of 13-miRNA signature as a potential biomarker for predicting survival in pancreatic cancer. *Oncotarget* 7:69616-69624. 10.18632/oncotarget.11903

Figure 1

Venn diagram analysis of differentially expressed lncRNA in esophageal cancer.

Figure 2

Heatmap of 74 lncRNAs differentially expressed in esophageal cancer.

Figure 3

Two differentially expressed lncRNAs in esophageal cancer patients.

(A) Kaplan-Meier curves show the correlation between two lncRNAs and overall survival; (B) ROC curves compare the sensitivity, specificity, and prognostic value of two lncRNAs and assess the efficacy of the signature.

Figure 4

Expression of two lncRNAs, risk score distribution and survival in esophageal cancer patients. The risk scores for all patients are shown in the top; following up and survival of each patient are plotted in the middle; expression distribution of two lncRNAs by low-risk and high-risk score groups are shown in the bottom.

Figure 5

Risk score model for the outcome of esophageal cancer patients.

(A) Kaplan-Meier curve tests the risk score model in overall survival; (B) Time-dependent ROC curve evaluates the efficacy of the risk score model in predicting 5-year survival.

Figure 6

Comparison of the expression level of two lncRNAs.

(A) The expression level of lncRNAs between tumor tissues and adjacent normal tissues; (B) The expression level of lncRNAs between low-risk and high-risk score groups.

Figure 7

Prognostic value of the clinical characteristics in overall survival and predictive value of the risk score model for specific clinical characteristics.

(A) Kaplan-Meier curves reveal five independent prognostic indicators; (B) ROC curves assess the risk score model in predicting clinical characteristics.

Figure 8

Top 20 enrichment of GO terms and KEGG pathways of co-expressed mRNAs.

Figure 9

Map presented the PPI network of co-expressed genes.

Table 1

Prognostic value of the differentially expressed lncRNAs by univariate and multivariate

Cox regression analyses.

Variables	Estimate	Standard error	Chi-square	<i>p</i> value	HR (95%CI)
Univariate analysis					
RPL34-AS1	-0.747	0.266	7.905	0.005*	0.474(0.281-0.798)
GK3P	0.757	0.277	7.489	0.006*	2.132(1.240-3.667)
Multivariate analysis					
RPL34-AS1	-0.685	0.266	6.622	0.010*	0.504(0.299-0.849)
GK3P	0.697	0.278	6.272	0.012*	2.008(1.164-3.464)

Table 2

Predictive values of the related clinical characteristics and risk score.

Variables		Univariate analysis		Multivariate analysis	
		HR (95% CI)	p value	HR (95% CI)	p value
Race	White	1(reference)			
	Black	0.931(0.126-6.869)	0.944		
	Asian	0.744(0.347-1.597)	0.448		
Gender	Female	1(reference)			
	Male	2.961(1.069-8.207)	0.037*		
Age	<=65	1(reference)			
	>65	1.326(0.791-2.223)	0.284		
TNM stage	I	1(reference)		1(references)	
	II	1.600(0.603-4.246)	0.345	1.434(0.315-6.530)	0.642
	III	4.128(1.510-11.282)	0.006*	5.852(1.295-26.442)	0.022*
	IV	8.087(2.441-26.794)	0.001*	3.055(0.473-19.733)	0.241
T stage	T1	1(reference)			
	T2	0.916(0.439-1.908)	0.814		
	T3	1.212(0.621-2.365)	0.573		
	T4	2.510(0.697-9.043)	0.159		
N stage	N0	1(reference)			
	N1	2.272(1.274-4.051)	0.005*	1.718(0.589-5.008)	0.258
	N2	3.748(1.371-10.246)	0.010*	0.866(0.173-4.342)	0.230
	N3	3.042(1.019-9.086)	0.046*	2.287(0.351-14.914)	0.820
M stage	M0	1(reference)			
	M1	3.760(1.740-8.124)	0.001*		
Neoplasm histologic grade	G1	1(reference)		1(references)	
	G2	2.341(0.800-6.852)	0.121	1.262(0.300-5.309)	0.453
	G3	2.150(0.720-6.425)	0.170	0.903(0.212-3.849)	0.481
Additional treatment completion success outcome	CR	1(reference)			
	SD	1.079(0.137-8.490)	0.943		
	PD	2.452(1.013-5.940)	0.047*		
Radiotherapy	No	1(reference)			
	Yes	0.834(0.411-1.689)	0.614		
Columnar mucosa dysplasia	No	1(reference)			
	Low	3.558(0.952-13.300)	0.059		
	High	1.229(0.554-2.728)	0.613		
Lymph node metastasis	No	1(reference)		1(reference)	
	Yes	1.674(0.951-2.945)	0.074	0.663(0.279-1.575)	0.274
Residual tumor	R0	1(reference)		1(reference)	
	R1+R2	2.523(1.290-4.933)	0.007*	2.152(0.829-5.581)	0.107
Neoplasm tumor status	Tumor free	1(reference)		1(reference)	
	With tumor	2.483(1.451-4.248)	0.001*	2.132(1.331-5.766)	0.036*
Risk score	Low	1(reference)		1(reference)	
	High	2.210(1.312-3.720)	0.003*	3.500(1.685-7.274)	0.001*

Table 3

Top 20 KEGG and GO terms enriched by the coding genes.

ID	Biological process and pathway	Count	Enrichment	-lg <i>P</i>	-lg FDR
GO					
GO:0050911	detection of chemical stimulus involved in sensory perception of smell	12	30.048	13.860	11.040
GO:0042742	defense response to bacterium	16	14.926	13.219	10.700
GO:0048387	negative regulation of retinoic acid receptor signaling pathway	11	31.717	13.002	10.660
GO:0045596	negative regulation of cell differentiation	11	26.167	11.954	9.737
GO:0007608	sensory perception of smell	12	19.030	11.238	9.117
GO:0007601	visual perception	17	9.037	10.400	8.359
GO:0007186	G-protein coupled receptor signaling pathway	21	6.384	9.925	7.950
GO:0007283	spermatogenesis	18	5.143	7.047	5.131
GO:0050912	detection of chemical stimulus involved in sensory perception of taste	5	47.576	6.914	5.049
GO:0030154	cell differentiation	16	3.985	4.815	2.995
GO:0050909	sensory perception of taste	5	18.298	4.558	2.780
GO:0018298	protein-chromophore linkage	4	29.277	4.494	2.754
GO:0008284	positive regulation of cell proliferation	16	3.704	4.417	2.712
KEGG					
hsa: 04740	Olfactory transduction	201	46.085	289.793	287.644
hsa: 04080	Neuroactive ligand-receptor interaction	22	7.557	11.850	10.002
hsa: 04742	Taste transduction	11	20.128	10.586	8.914
hsa: 05320	Autoimmune thyroid disease	5	9.149	3.059	1.512
hsa: 05033	Nicotine addiction	4	9.515	2.485	1.035
hsa: 04744	Phototransduction	3	10.195	1.905	0.534
hsa: 04140	Regulation of autophagy	3	7.319	1.496	0.192