Oncology and Translational Medicine
DOI 10.1007/s10330-020-0408-8

June 2020, Vol. 6, No. 3, P109–P115

ORIGINAL ARTICLE

Prognostic risk model construction and prognostic biomarkers identification in esophageal adenocarcinoma based on immune-related long noncoding RNA*

Kai Qin[1], Yi Cheng[1], Jing Zhang[1], Xianglin Yuan[1], Jianhua Wang[1], Jian Bai[2] (✉)

[1] Department of Oncology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China
[2] Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China

Abstract

Objective The aim of this study was to construct a prognostic model of esophageal adenocarcinoma (EAC) based on immune-related long noncoding RNAs (immune-related lncRNAs) and identify prognostic biomarkers using the Cancer Genome Atlas (TCGA) database.

Methods Whole genomic mRNA expression and clinical data of esophageal adenocarcinoma were obtained from the TCGA database. The software Strawberry Perl, R and R packets were used to identify the immune-related genes and lncRNAs of esophageal adenocarcinoma, and for data processing and analysis. The differentially expressed lncRNAs were detected while comparing esophageal adenocarcinoma and normal tissue samples. The key immune-related lncRNAs were screened using lasso regression analysis and univariate cox regression analysis, and used to construct the prognostic model using multivariate cox regression analysis.

To evaluate the accuracy of the risk prognostic model, all esophageal adenocarcinomas were divided into high-risk and low-risk groups according to the median risk score, after which Kaplan-Meier (K-M) survival curves, operating characteristic (ROC) curve and independent prognostic analysis of clinical traits were created. In addition, statistically significant immune-related lncRNAs and potential prognostic biomarkers were identified using the prognostic model and multifactor cox regression analysis for k-m survival analysis.

Results A total of 1322 differentially expressed immune-related lncRNAs were identified, 28 of which were associated with prognosis via univariate cox regression analysis. In addition, K-M survival analysis showed that the total survival time of the higher risk group was significantly shorter than that of the lower risk group (P = 1.063e−10). The area under the ROC curve of 5-year total survival rate was 0.90. The risk score showed independent prognostic risk for esophageal adenocarcinoma via single factor and multifactorial independent prognostic analyses. In addition, the HR and 95% CI of each key immune-related lncRNA were calculated using multivariate Cox regression. Using k-m survival analysis, we found that 5 out of 12 key significant immune-related lncRNAs had independent prognostic value [AL136115.1 (P = 0.006), AC079684.1 (P = 0.008), AC07916394.1 (P = 0.0386), AC087620.1 (P = 0.041) and MIRLET7BHG (P = 0.044)].

Conclusion The present study successfully constructed a prognostic model of esophageal adenocarcinoma based on the TCGA database, with moderate predictive accuracy. The model consisted of the expression level of 12 immune-related lncRNAs. Furthermore, the study identified one favorable prognostic biomarker, MIRLET7BHG, and four poor prognostic biomarkers (AL136115.1, AC079684.1, AC016394.1, and AC087620.1).

Received: 20 February 2020
Revised: 11 March 2020
Accepted: 25 March 2020

Key words: immune-related Cancer Genome Atlas (lncRNA); prognostic model; prognostic biomarker; esophageal adenocarcinoma (EAC); Cancer Genome Atlas (TCGA) database

✉ Correspondence to: Jian Bai. Email: 26062793@qq.com
* Supported by a grant from the Health Commission of Hubei Province Scientific Research Project (No. WJ2019M118).

Esophageal cancer is one of the most common malignant cancer of digestive tract [1]. In China, esophageal cancer, which is the sixth deadliest cancer worldwide, accounted for 6.25% of all new cases of malignant cancers in 2015, with an estimated 245,700 new cases [2]. The overall survival rate of esophageal cancer is still low and the burden of this malignancy is particularly high in China [2]. It accounts for 20% of the world's esophageal cancer [3]. About 50% of all global deaths from esophageal cancer occur in China [3–4]. Two major histological subtypes, esophageal adenocarcinoma (EAC) and squamous cell carcinoma (ESCC), show significantly different patterns, and the combination of these two types represent the vast majority of esophageal cancer [2]. EAC is the main type in western developed countries [5]. Melina Arnold *et al* predicted a dramatic increase of EAC in high-income countries, surpassing ESCC in the next few years [6]. Cowie A *et al* found that countries such as Japan are beginning to observe rising rates of EAC, likely due to the 'westernization' of their lifestyle [5]. Although there are many risk assessment models of esophageal carcinoma worldwide, most of them are prognostic models of ESCC with some limitations [7]. Therefore, constructing predictive models and prognostic markers of esophageal adenocarcinoma may prove helpful for prognosis judgment and treatment selection.

Long noncoding RNAs (lncRNAs) belong to a class of endogenous RNAs with a length of more than 200 nucleotides, which regulate a variety of malignant tumor phenotypes through epigenetic modification, RNA decay and transcriptional regulation participate in biological processes such as proliferation, differentiation, and invasion of tumor cells [8]. Numerous lncRNAs are abnormally expressed in ESCC (*e.g.* actin filament-associated protein 1-antisense RNA 1 (AFAP1-AS1), metastatic lung adenocarcinoma transcript 1 (MALAT1), hox contraindic intergenic RNA (HOTAIR), taurine upregulated gene 1 (TUG1), etc. These aberrantly expressed lncRNAs play an important role in the occurrence and development of ESCC, possibly promoting or inhibiting cancer development, and therefore, could be potential biomarkers for ESCC prognostic assessment [9–11]. Esophageal adenocarcinoma is the main risk prediction model for esophageal cancer in Europe and Australia. Mostly, the research designs have focused on case-control and cohort studies, and non-genetic prediction variables are used to construct a risk scoring model. To date, there is no risk scoring model constructed with lncRNAs as predictive variables in esophageal adenocarcinoma [7].

Using whole genomic mRNA expression and clinical data from the TCGA database, the present study constructed an immune-related lncRNA prognostic risk model and identified prognostic biomarkers in esophageal adenocarcinoma.

Materials and methods

Acquisition of esophageal adenocarcinoma expression data

Whole genomic mRNA expression and clinical data for esophageal adenocarcinoma were downloaded from the TCGA database (https://portal.gdc.cancer.gov/). The screening conditions were as follows: (1) primary tumor site: esophageal cancer; (2) project: TCGA-ESCA; (3) disease type: adenocarcinoma or adenoma; (4) data category: transcriptome profiling; (5) data type: gene expression quantification; (6) experimental strategy: RNA-Seq; (7) workflow type: HTSeq-fpkm; (8) remaining screening criteria: default or unselected. The dataset downloaded in January 14, 2020 was integrated into a matrix using the Strawberry-Perl software (version 5.30.1.1–64 bits). We obtained the entire messenger RNA (mRNA) spectrum data and the clinical information of the samples (number of patients, survival time, survival status, sex, clinical stages and TNM stage).

Immune-related genes and lncRNA for esophageal adenocarcinoma

The mRNA and lncRNA matrices of the encoded proteins were obtained using Perl software. The GSEA website was used to search for gene sets: extracted protein-encoded immune-related genes applying to IMMUNE_RESPONSE (M19817) and IMMUNE_SYSTEM_PROCESS (M13664). The R (version 3.6.1) and Bioconductor (https://www.bioconductor.org/) packets were used to obtain immune-related lncRNAs and for the corresponding data processing and analysis.

Analysis of the differential expression of lncRNAs

Differentially expressed lncRNAs in EAC were screened using the R-package "edgeR." The screening criteria were a log2 (foldchange) > 2 and a false discovery rate (FDR) < 0.05. Then, a volcanic map was plotted using the R-package "gplot2."

Single factor Multivariate Cox regression analysis, Lasso regression analysis and prognostic model construction

To screen differentially expressed immune-related lncRNAs, single factor Cox regression analysis was performed using the R-package "survival " with $P < 0.001$. To avoid overfitting, the statistically significant differentially expressed lncRNAs were also analyzed by lasso regression analysis. Finally, multivariate cox regression analysis was used to identify immune-related lncRNAs and construct a prognostic model. To evaluate the accuracy of the risk prognostic model, Kaplan-Meier (K-M) survival curves, operating characteristic (ROC)

curve, and independent prognostic analysis of clinical traits were performed. Towards that, all esophageal adenocarcinomas were divided into high-risk and low-risk groups according to their median risk score. At the same time, we calculated HR and 95% CI of key lncRNAs in the prognostic model, A *P*-value < 0.05 was considered statistically significant.

Evaluation of prognostic models, K-M survival analysis, and prognostic analysis of clinical traits

To assess the predictive ability of the prognostic model, K-M survival analysis was performed on the patient risk score. Next, the roc curve of the 5-year overall survival rate was plotted, and the area under the curve (AUC) as well as other clinical traits (gender, staging, TNM, and risk score) were calculated using the R package "survivalROC." An AUC value of 0.7 to 0.9 accuracy rate is generally considered medium accuracy, while an AUC value > 0.9 is considered high accuracy. In addition, to explore the value and significance of lncRNAs in predicting the prognosis of EAC, we performed K-M survival analysis of individual lncRNAs that were statistically significant as observed by the results of multivariate cox regression analysis. The clinical characters of EAC were analyzed by single factor and multiple factor cox regression analyses.

Results

Differentially expressed lncRNAs in EAC

A total of 56754 genes from 78 EAC tissues samples and 9 paracancerous or normal esophageal tissues were obtained from the TCGA database. Furthermore, clinical data were downloaded for 87 patients with EAC (Table 1). These data were processed and analyzed using the R language and the corresponding R packets to obtain 14131 immune-related lncRNAs and 19659 protein-encoded immune-related genes. Using the edgeR package, 1322 differentially expressed immune-related lncRNAs were screened with a threshold of |log2FC| > 1 and FDR < 0.05, including 28 up-regulated and 66 down-regulated lncRNAs (Fig. 1a). A total of 28 immune-related lncRNAs identified as prognostic risk factors were preliminarily screened using single factor Cox regression analysis combined with the corresponding clinical data (Table 2). In addition, 12 key immune-related lncRNAs were further identified by Lasso regression analysis (Fig. 1b and 1c).

Construction and evaluation of the prognostic model of EAC

All esophageal adenocarcinomas were divided into high- and low-risk groups according to the median risk score. Then, we constructed a prognostic model based on the patient's prognostic risk scores. A total of 12 lncRNAs were used for the risk scoring model with a risk score $=(1.84)*Exp_{LINC01612}+(-4.40)*Exp_{MIRLET7BHG}+(1.86)*Exp_{AL136115.1}+(1.86)*Exp_{AL121992.3}+(2.00)*Exp_{AC087620.1}+(1.16)*Exp_{AC005841.1}+(1.37)*Exp_{AC016394.1}+(1.21)*Exp_{AC079684.1}+(-1.15)*Exp_{AC078778.1}+(0.81)*Expression_{AC010168.1}+(1.26)*Expression_{AC048344.4}+(-0.89)*Exp_{AL163051.2}$. The K-M survival analysis ($P = 1.063e{-}10$, Fig. 5f) indicated that the overall survival time of the higher risk group was significantly shorter than that of the lower risk group. The area under the ROC curve of 5-year total survival rate was 0.90 (Fig. 2). Prognostic risk score predicted the independent prognostic risk for esophageal adenocarcinoma via single factor and multifactorial independent prognostic analysis,

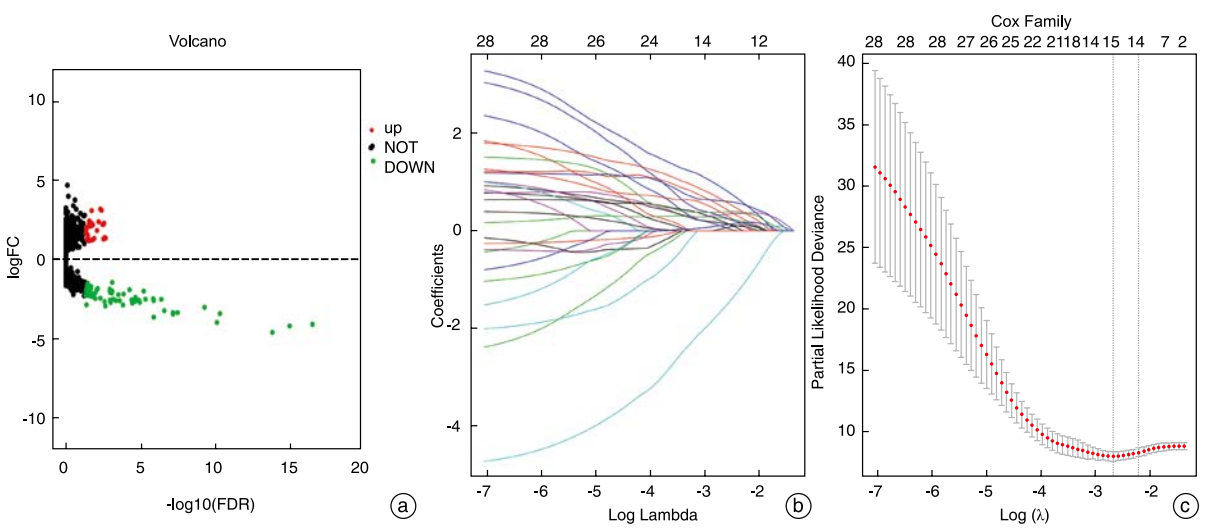

Fig. 1 Identification of the differentially expressed lncRNA by Lasso regression analysis in EAC. (a) Volcano plot of the differentially expressed lncRNAs; (b) tuning parameter (lambda) selection in the Lasso regression using 10 fold cross-validation via minimum criteria; (c) Lasso coefficient profiles of the features against the log2 (lambda)

Table 1 Clinical characteristics of esophageal adenocarcinoma

Clinical characteristics	*n* (%)
Gender	
Male	75 (86.2)
Female	12 (13.8)
Stage	
I	11 (12.6)
II	22 (25.3)
III	29 (33.3)
IV	5 (5.8)
Unknow	20 (23.0)
Tumor	
T1	24 (27.6)
T2	11 (12.6)
T3	37 (42.6)
T4	1 (1.1)
Unknow	14 (16.1)
Fustat	
Alive	43 (49.4)
Dead	44 (50.6)
Node	
N0	21 (24.1)
N1	40 (46.0)
N2	6 (6.9)
N3	5 (5.7)
Unknow	15 (17.3)
M	
M0	51 (58.6)
M1	5 (5.7)
Unknow	31 (35.7)

$P < 0.001$ (Fig. 3). The heatmap of 12 lncRNAs involved in constructing the risk scoring model is shown in Fig. 4. In addition, we also calculated the HR and 95% CI for these lncRNAs. Our results showed that among these 12 lncRNAs, 8 were independent prognostic risk factors for EAC ($P < 0.05$; Table 3).

Prognostic biomarkers of esophageal adenocarcinoma

Using a K-M survival analysis of 8 lncRNAs with independent prognostic risk factors, we further short-listed 5 immune-related lncRNAs with independent prognostic value: AL136115.1 ($P = 0.006$), AC079684.1 ($P = 0.008$), AC07916394.1 ($P = 0.0386$), AC087620.1 ($P = 0.041$) and MIRLET7BHG ($P = 0.044$). Based on our results, we identified a favorable prognostic biomarker, MIRLET7BHG, and four poor prognostic biomarkers (AL136115.1, AC079684.1, AC016394.1 and AC087620.1) (Fig. 5a–5e).

Table 2 Prognostic immune-related lncRNA in esophageal adenocarcinoma identified preliminary by univariate Cox regression analysis

Immune-related lncRNA name	HR	*P*
AC005841.1	2.640 (1.286–5.417)	0.008
AL136115.1	2.753 (1.361–5.570)	0.005
JPX	2.973 (1.393–6.346)	0.005
AL121992.3	3.275 (1.534–6.990)	0.002
MIRLET7BHG	0.285 (0.112–0.724)	0.008
AL080317.1	4.001 (2.019–7.926)	< 0.001
AC010168.1	2.534 (1.310–4.899)	0.006
AC048344.4	3.310 (1.337–8.197)	0.01
AL163051.2	3.234 (1.415–7.390)	0.005
AC079684.1	4.566 (1.932–10.787)	< 0.001
KCNQ1OT1	2.191 (1.290–3.720)	0.004
AC087752.4	2.728 (1.312–5.672)	0.007
AL133243.2	2.837 (1.438–5.598)	0.003
AL024508.2	1.785 (1.160–2.748)	0.008
AF117829.1	4.296 (1.645–11.221)	0.003
AC016394.1	2.597 (1.416–4.762)	0.002
AC078778.1	2.734 (1.290–5.793)	0.009
AC127024.3	2.311 (1.254–4.261)	0.007
AL590822.2	1.855 (1.215–2.831)	0.004
LINC01612	1.783 (1.166–2.726)	0.008
AC139100.2	2.741 (1.433–5.244)	0.002
AC087620.1	3.910 (1.642–9.309)	0.002
AL031673.1	2.056 (1.340–3.153)	< 0.001
ZNF337-AS1	3.150 (1.484–6.688)	0.003
AL596247.1	2.686 (1.394–5.174)	0.003
AC009318.2	3.699 (1.537–8.903)	0.004
AC012467.2	2.604 (1.522–4.455)	< 0.001
AC097468.3	3.344 (1.513–7.393)	0.003

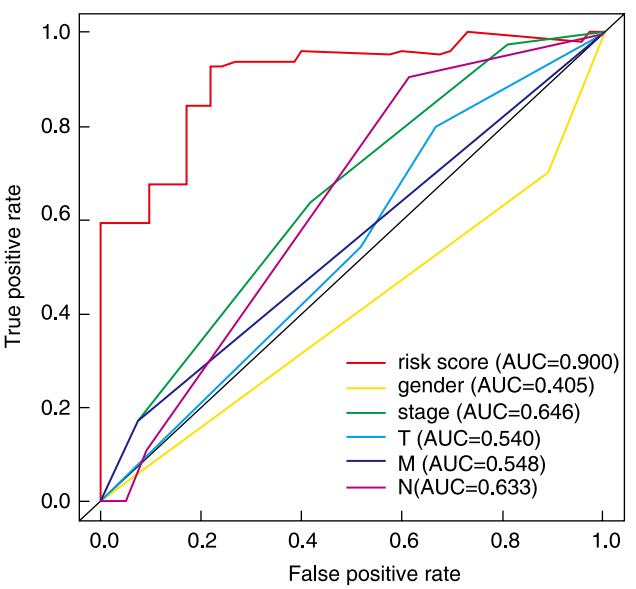

Fig. 2 ROC curve of multivariate Cox analysis model

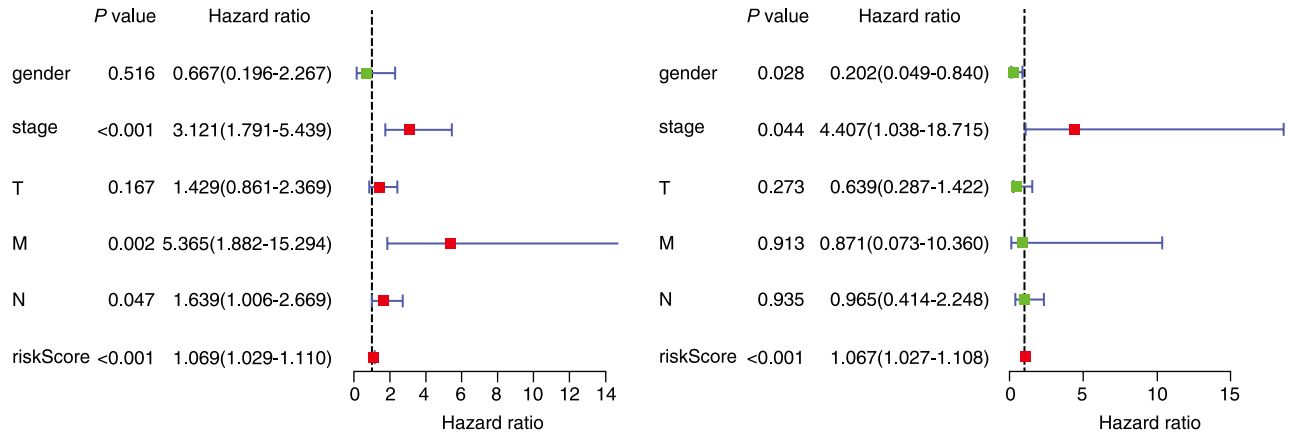

Fig. 3 Univariate and multivariate independent prognostic analysis of clinical features of esophageal adenocarcinoma

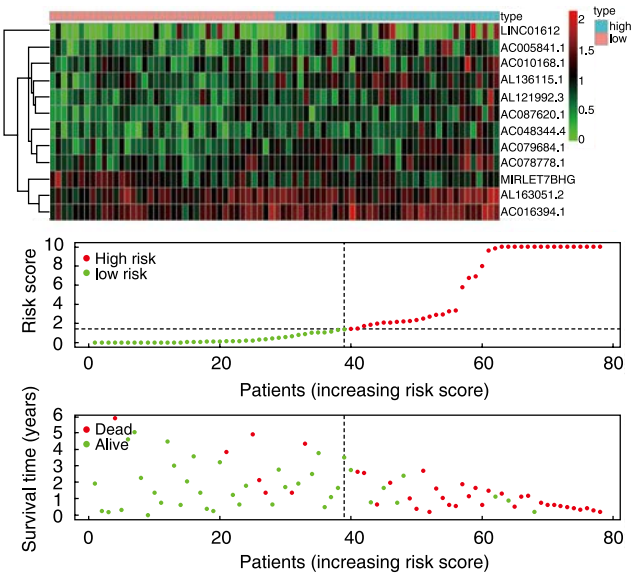

Fig. 4 Risk score and heat map of Immune-related lncRNA expression and scattered plots of survival time

Table 3 Result of the multivariate Cox regression analysis based on the 12 key immune-related lncRNA

immune-related lncRNA name	Coef	HR (95%CI)	*P*
LINC01612	1.842304116	6.311(3.150–12.643)	2.03E-07
MIRLET7BHG	-4.395979721	0.012(0.002–0.069)	5.95E-07
AL136115.1	1.862898963	6.442(2.414–17.193)	0.000199466
AL121992.3	1.857241154	6.406(2.088–19.652)	0.001165136
AC087620.1	2.001926502	7.403(1.902–28.814)	0.003885086
AC005841.1	1.157845653	3.183(1.356–7.471)	0.00782035
AC016394.1	1.369822912	3.934(1.338–11.575)	0.012837805
AC079684.1	1.210612392	3.355(1.033–10.896)	0.043959335
AC078778.1	-1.153082814	0.315(0.098–1.016)	0.053285834
AC010168.1	0.814394289	2.257(0.981–5.195)	0.055437015
AC048344.4	1.260896544	3.528(0.845–14.735)	0.083815226
AL163051.2	-0.891848958	0.410(0.128–1.310)	0.132527493

HR: hazard ratio; coef: coefficient

Discussion

Numerous studies have confirmed that lncRNAs play a wide range of regulatory roles in tumor occurrence, immune response, and tumor progression [12–14]. LncRNAs play a significant role in different stages of tumor immunity, such as antigen recognition, immune activation, immune cell infiltration, and tumor clearance [15]. Lv *et al* found that lncRNAs are involved in regulatory pathways of the immune system, such as T cell differentiation, immune deficiency, and cytotoxicity of natural killer cells, affecting patient's prognosis in hepatocellular carcinoma [16].

In this study, we obtained transcript data and clinical data from patients with esophageal adenocarcinoma from the TCGA database. Using correlation analysis, we found 14131 immune-related lncRNAs and 19659 protein-encoded immune-related genes in esophageal adenocarcinoma. We used univariate cox regression analysis and identified 28 immune-related lncRNAs that were associated with prognosis. In addition, K-M survival analysis (P = 1.063e–10) showed that the total survival time of the higher risk group was significantly shorter than that of the lower risk group. The area under the ROC curve of 5-year total survival rate was 0.90, indicating that the prediction accuracy of this prognostic model was relatively high. Prognostic risk score showed independent prognostic risk for esophageal adenocarcinoma via single factor and multifactorial independent prognostic analysis. In addition, the HR and 95% CI of each key immune-related lncRNA were calculated using multivariate Cox regression, and 12 key significant immune-related lncRNAs were identified. Moreover, K-M survival analysis identified five immune-related lncRNAs with independent prognostic value: a favorable prognostic biomarker, MIRLET7BHG, and four poor prognostic biomarkers (AL136115.1, AC079684.1,

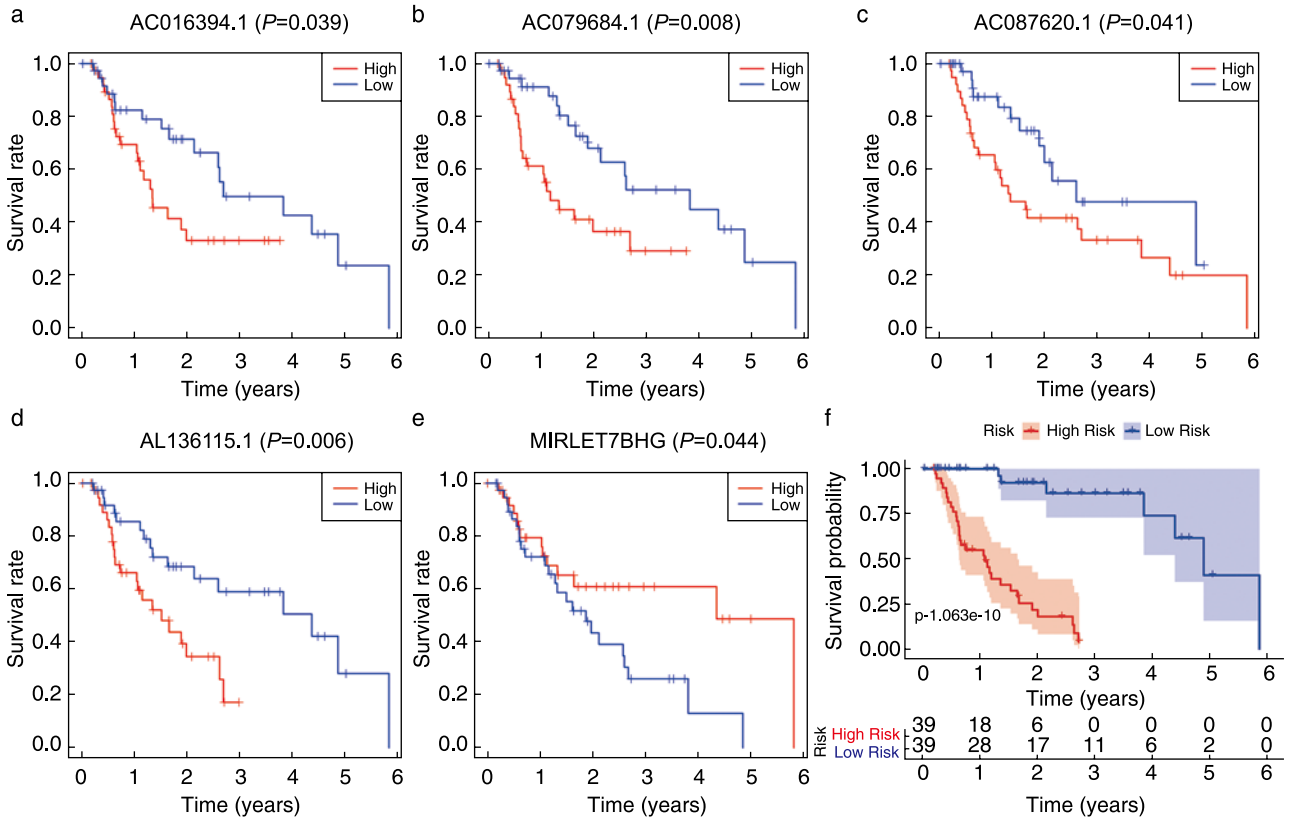

Fig. 5 K-M survival curves of the 8 significant key immune-related lncRNA identified using multivariate Cox regression and evaluation of the prognostic model of esophageal adenocarcinoma using the K-M survival curve

AC016394.1, and AC087620.1). Since there is a dearth of studies on these 4 immune-related lncRNAs [17–18], further research is needed to validate their role in the progression and prognosis of EAC disease. Liu et al. found that the MIRLET7BHG gene polymorphism may be an important predictor of asbestos exposure-related lung cancer [18]. Studies on lncRNA (MIRLET7BHG) have not been reported yet. Although immune-related lncRNAs are dysregulated during cell carcinogenesis, they are rarely reported in EAC [15, 19–20]. Wu *et al* reported that EAC exhibited reduced non-coding region methylation. Methylation of the long noncoding RNA AFAP1-AS1 is reduced in EAC, and its expression inhibits the cancer-related biological functions of EAC cells [19]. However, this study had some limitations. The conclusions were obtained from the TCGA database, and the prediction model and predictive markers of EAC lack domestic data for confirmation. Further, the role and mechanism of immune-related lncRNAs in EAC need to be further validated.

Conclusions

In summary, we screened lncRNAs from TCGA data base analyzed by various statistical tools. Based on our results, we identified 12 key immune-related lncRNAs in the present work and constructed a prognostic model of EAC with moderate predictive accuracy. In addition, we identified five immune-related lncRNAs as independent prognostic factors for EAC: AL136115.1, AC079684.1, AC016394.1, AC087620.1, and MIRLET7BHG. While MIRLET7BHG was identified as a favorable prognostic biomarker, the remaining four (AL136115.1, AC079684.1, AC016394.1, and AC087620.1) were identified as poor prognostic biomarkers.

Conflicts of interest

The authors indicated no potential conflicts of interest.

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. CA Cancer J Clin, 2018, 68: 394–424.

2. Chen R, Zheng RS, Zhang SW, *et al*. Analysis of incidence and mortality of esophageal cancer in China, 2015. Chin J Prev Med (Chinese), 2019, 53: 1094–1097.
3. Gao X, Wang Z, Kong C, *et al*. Trends of esophageal cancer mortality in Rural China from 1989 to 2013: An age-period-cohort analysis. Int J Environ Res Public Health, 2017, 14: 218.
4. Zhang HY, Pei XF, Mo XQ, *et al*. Incidence and survival changes in patients with esophageal adenocarcinoma during 1984–2013. Biomed Res Int, 2019, 2019: 7431850.
5. Cowie A, Noble F, Underwood T. Strategies to improve outcomes in esophageal adenocarcinoma. Expert Rev Anticancer Ther, 2014, 14: 677–687.
6. Arnold M, Laversanne M, Brown LM, *et al*. Predicting the future burden of esophageal cancer by histological subtype: international trends in incidence up to 2030. Am J Gastroenterol, 2017, 112: 1247–1255.
7. Li H, Chen WQ. Progress in research of high risk assessment models for esophageal cancer. Chin Cancer (Chinese), 2019, 28: 883–890.
8. Yang FL, Hong K, Zhao GJ, *et al*. Construction of prognostic model and identification of prognostic biomarkers based on the expression of long non-coding RNA in bladder cancer via bioinformatics. J Peking Uni (Health Sci) (Chinese), 2019, 51: 615–622.
9. Li XM, Xing Y, ZA WJ, *et al*. Research progress of long non-coding RNA in esophageal cancer. Shandong Med J (Chinese), 2019, 59: 94–97.
10. Talebi A, Masoodi M, Mirzaei A, *et al*. Biological and clinical relevance of metastasis-associated long noncoding RNAs in esophageal squamous cell carcinoma: A systematic review. J Cell Physiol, 2020, 235: 848–868.
11. Yan YS, Li SB, Wang SH, *et al*. Long noncoding RNA HAND2-AS1 inhibits cancer cell proliferation, migration, and invasion in esophagus squamous cell carcinoma by regulating microRNA-21. J Cell Biochem, 2019, 120: 9564–9571.
12. Leucci E, Coe EA, Marine JC, *et al*. The emerging role of long non-coding RNAs in cutaneous melanoma. Pigment Cell Melanoma Res, 2016, 29: 619–626.
13. Li X, Meng Y. Survival analysis of immune-related lncRNA in low-grade glioma. BMC Cancer, 2019, 19: 813.
14. Wang B, Liang T, Li J. Long noncoding RNA LINC01296 is associated with poor prognosis in ESCC and promotes ESCC cell proliferation, migration and invasion. Eur Rev Med Pharmacol Sci, 2018, 22: 4524–4531.
15. Yu WD, Wang H, He QF, *et al*. Long noncoding RNAs in cancer-immunity cycle. J Cell Physiol, 2018, 233: 6518–6523.
16. Lv Y, Wei W, Huang Z, *et al*. Long non-coding RNA expression profile can predict early recurrence in hepatocellular carcinoma after curative resection. Hepatol Res, 2018, 48: 1140–1148.
17. Feng Q, Zhang H, Yao D, *et al*. Emerging role of non-coding RNAs in esophageal squamous cell carcinoma. Int J Mol Sci, 2019, 21: 258.
18. Liu CY, Stucker I, Chen C, *et al*. Genome-wide gene-asbestos exposure interaction association study identifies a common susceptibility variant on 22q13.31 associated with lung cancer risk. Cancer Epidemiol Biomark Prev, 2015, 24: 1564–1573.
19. Wu W, Bhagat TD, Yang X, *et al*. Hypomethylation of noncoding DNA regions and overexpression of the long noncoding RNA, AFAP1-AS1, in Barrett's esophagus and esophageal adenocarcinoma. Gastroenterology, 2013, 144: 956–966.
20. Li S, Xu Y, Sun Z, *et al*. Identification of a lncRNA involved functional module for esophageal cancer subtypes. Mol Biosyst, 2016, 12: 3312–3323.

DOI 10.1007/s10330-020-0408-8
Cite this article as: Qin K, Cheng Y, Zhang J, *et al*. Prognostic risk model construction and prognostic biomarkers identification in esophageal adenocarcinoma based on immune-related long noncoding RNA. Oncol Transl Med, 2020, 6: 109–115.