A six-microRNA signature can better predict overall survival of patients with esophagus adenocarcinoma

Tian Lan, Yunyan Lu, Zunqiang Xiao, Haibin Xu, Junling He, Zujian Hu, Weimin Mao

1 The second clinical medical college, Zhejiang Chinese Medical University, Hangzhou, Zhejiang, People's Republic of China
2 Department of Breast Surgery, Hangzhou Hospital Of Traditional Chinese Medicine, Hangzhou, Zhejiang, People's Republic of China
3 Department of Cardiology, Hangzhou Xiaoshan First People's Hospital, Hangzhou, Zhejiang, People's Republic of China
4 Zhejiang Key laboratory of Diagnosis and Treatment Technology on Thoracic Oncology (Lung and Esophagus), Zhejiang Cancer Hospital, Hangzhou, Zhejiang, People's Republic of China
5 Department of Thoracic Surgery, Zhejiang Cancer Hospital, Hangzhou, Zhejiang, People's Republic of China

Corresponding Author: Weimin Mao
Email address: maowm1218@163.com

Background The MicroRNAs (miRNAs) have been validated as prognostic markers in many cancers. The aim of this study was to construct a miRNA-based signature for predicting the prognosis of esophagus adenocarcinoma (EAC).

Methods The RNA sequencing data set of EAC was downloaded from The Cancer Genome Atlas (TCGA). 84 patients with EAC were randomly divided into a train and a test set. Using univariate Cox regression analysis and The least absolute shrinkage and selection operator (LASSO), we identified prognostic factors and construct a prognostic miRNA signature. Receiver operating curve (ROC) analysis was applied to validate the accuracy of the signature.

Result In general, 6 miRNAs (has-let-7b, has-mir-23a, has-mir-3074, has-mir-424, has-mir-425, has-mir-505) were demonstrated to be predictive biomarkers of overall survival for EAC patients in train set. Patients assigned to the high-risk group based on the risk score of this miRNA model had significantly shorter overall survival than those in the low-risk group. This 6-miRNA model was validated in test and entire set. The Aera under curve (AUC) for ROC at 3 years was 0.868 in the entire set. Molecular functional analysis and pathway enrichment analysis indicated that the target mRNAs associated with 6-miRNA signature were closely related to multiple signaling pathways linked to carcinogenesis, especially cell cycle.

Conclusion In summary, we identified and validated a novel 6-miRNA-expression-based prognostic signature based on EAC data of TCGA.
A six-microRNA signature can better predict overall survival of patients with esophagus adenocarcinoma

Tian Lan¹,², Yunyan Lu³, Zunqiang Xiao¹, Haibin Xu², Junling He², Zujian Hu², WeiMin Mao⁴,⁵

¹ The second clinical medical college, Zhejiang Chinese Medical University, Hangzhou, Zhejiang, People's Republic of China
² Department of Breast Surgery, Hangzhou Hospital of Traditional Chinese Medicine, Hangzhou, Zhejiang, People's Republic of China
³ Department of Cardiology, Hangzhou Xiaoshan First People’s Hospital, Hangzhou, Zhejiang, People's Republic of China
⁴ Zhejiang Key laboratory of Diagnosis and Treatment Technology on Thoracic Oncology (Lung and Esophagus), Zhejiang Cancer Hospital, Hangzhou, Zhejiang, People's Republic of China
⁵ Department of Thoracic Surgery, Zhejiang Cancer Hospital, Hangzhou, Zhejiang, People's Republic of China

Corresponding Author:
WeiMin Mao⁴,⁵

Email address: maowm1218@163.com
ABSTRACT
Background
The MicroRNAs (miRNAs) have been validated as prognostic markers in many cancers. The aim of this study was to construct a miRNA-based signature for predicting the prognosis of esophagus adenocarcinoma (EAC).

Methods
The RNA sequencing data set of EAC was downloaded from The Cancer Genome Atlas (TCGA). 84 patients with EAC were randomly divided into a train and a test set. Using univariate Cox regression analysis and the least absolute shrinkage and selection operator (LASSO), we identified prognostic factors and constructed a prognostic miRNA signature. Receiver operating curve (ROC) analysis was applied to validate the accuracy of the signature.

Result
In general, 6 miRNAs (has-let-7b, has-mir-23a, has-mir-3074, has-mir-424, has-mir-425, has-mir-505) were demonstrated to be predictive biomarkers of overall survival for EAC patients in train set. Patients assigned to the high-risk group based on the risk score of this miRNA model had significantly shorter overall survival than those in the low-risk group. This 6-miRNA model was validated in test and entire set. The Area under curve (AUC) for ROC at 3 years was 0.959, 0.840, and 0.868 in train, test, and entire set respectively. Molecular functional analysis and pathway enrichment analysis indicated that the target mRNAs associated with 6-miRNA signature were closely related to multiple signaling pathways linked to carcinogenesis, especially cell cycle.

Conclusion
In summary, we identified and validated a novel 6-miRNA-expression-based prognostic signature based on EAC data of TCGA.

Keywords: Esophagus adenocarcinoma, TCGA, Prognosis, Bioinformatics

INTRODUCTION
Esophagus cancer is the seventh common cancer worldwide and the sixth most common cause of cancer death in 2018 according to the Global Cancer Observatory (GCO) (Global Burden of Disease Cancer et al. 2018). Although the diagnosis and treatment strategies have been developed, this cancer remains a major problem due to insufficient information on its etiology, and the overall five-year survival rate for patients with esophageal cancer is 15% to 25% worldwide (Pennathur et al. 2013). More often, there are two types of malignancies: squamous cell carcinoma (90% of cases) and adenocarcinoma (10%). The prevalence of esophagus adenocarcinoma (EAC) has rapidly increased over the past few decades (Thrift & Whiteman 2012). EAC carries a poor prognosis, with an overall 5-year survival rate of 30% (Hirst et al. 2011). Due to the poor outcomes of EAC, it is important to explore the molecular mechanisms involved in the occurrence and development of EAC. More biomarkers that can effectively
predict the genesis, progress and prognosis of EAC need to be found urgently. MicroRNAs (miRNAs) are a group of small noncoding RNA transcripts that are consisting of approximate 22 nucleotides (Lujambio & Lowe 2012). The predominant function of miRNAs is to regulate protein translation by binding to target messenger RNAs (mRNAs), and thereby regulate mRNA translation negatively (Krol et al. 2010). They have recently been validated and aided in diagnosis and prognosis of a variety of tumors, including hepatocellular carcinoma (Parizadeh et al. 2019), prostate cancer (Moya et al. 2019), and breast cancer (Yerukala Sathipati & Ho 2018) et al. Many studies focus on miRNAs in patients with Barrett’s Esophagus (Leidner et al. 2012; Li et al. 2018; Revilla-Nuin et al. 2013), a precursor lesion of EAC, comprehensive analysis of miRNA associated with prognosis of EAC remains poorly understood. Over the past few years, some studies reported the significant role of miRNAs in the molecular diagnosis and prognosis of EAC. A 4-miRNAs expression profile score can provide a validated approach of predicting pathological complete response rates (pCR) to neoadjuvant treatment in EAC (Skinner et al. 2014). In addition, three miRNAs (miR-99b and miR-199a_3p and _5p) signature is associated with patient survival and the presence of lymph node metastasis (Feber et al. 2011). However, the number of patients enrolled in these studies is small. The Cancer Genome Atlas (TCGA), a landmark cancer genomics program, provides open access to many comprehensive miRNA-sequencing datasets spanning 33 cancer types. In this study, we constructed a prognostic risk score system based on miRNAs dataset from TCGA to predict the prognosis of EAC. Furthermore, we conducted gene oncology annotation and pathway enrichment analyses to determine the potential biological functions of mRNAs associated with this signature.

MATERIALS AND METHODS Microarray datasets preparation

From TCGA data portal, RNA-seq data and associated clinical information were downloaded in January 2019. On the basis of annotation information provided by GENCODE datasets (www.gencodegenes.org), Some miRNAs and mRNA are not expressed in certain tissue or show little variation, thus only miRNAs and mRNAs with raw count value >20 in more than 80% of samples were retained for further analysis. After normalization by edgeR, the expression profiles of miRNAs and mRNAs were further converted to log2 (normalized value +1) transformation to be used for the next operation. Samples with a < 1-month censor time are removed, because they can not be representative samples for analyzing prognostic factors. A total of 84 EAC subjects with the corresponding clinical data including age, gender, height, weight, race, alcohol history, barrett’s disease history, tumor size, lymph node status, metastasis status, TNM stage were collected in this study (Table 1). The EAC patients dataset contained 96 samples (84 EAC and 12 normal tissues) and 272 miRNAs. Since the data comes from the TCGA database, no further
approval was required from the Ethics Committee.

**Construction and validation of the miRNA risk score**

84 patients were randomly divided into 2 groups: train set = 42, test set = 42. Train set was analyzed to build a miRNA model that further validated in the test and entire set.

In the train set, we screen out miRNAs with a significant p value less than 0.1 by using univariate survival analysis based on Cox proportional hazards of each miRNAs. The least absolute shrinkage and selection operator (LASSO) is a generalized linear regression algorithm capable of variable selection and regularization simultaneously (Gao et al. 2010). LASSO was performed to reduce selected prognostic miRNAs further and to construct the risk score system.

To evaluate the survival risk, a miRNA-based prognostic model was established as the following formula: Risk score = $\beta_1 \times \text{gene } 1 + \beta_2 \times \text{gene } 2 + \ldots + \beta_n \times \text{gene } n$, where $\beta$ indicates the coefficient of the miRNA, and gene indicates the expression value of the miRNA.

Using the median score in train set as the cutoff, patients were divided into the high-risk and the low-risk groups. We employed Kaplan–Meier (KM) survival analysis by using the R “survival” package to compare the survival rate between the high- and low-risk group. The time-dependent receiver-operating characteristic (ROC) curve was plotted using the R “timeROC” package to evaluate the specificity and sensitivity of the miRNA expression-based prognostic signature.

Next, this signature were validated in test set and entire set. ROC and KM curves were carried out to validate the feasibility and accuracy of the miRNA model. Then stratified analysis based on clinical parameters was performed in the entire set.

All ROC and KM curves were plotted with R (version 3.5.2), and P value less than 0.05 was considered statistically significant.

**Gene set enrichment analysis**

All patients were divided into two groups (high and low) based on the risk score of the 6-miRNA signature, the median score in train set was set as the cutoff. We used gene set enrichment analysis (GSEA, http://software.broadinstitute.org/gsea) (Subramanian et al. 2005) to figure out potential functional annotations in the two groups. The BioCarta dataset (c2.cp.biocarta.v6.2.symbols.gmt) was chosen as the reference gene set. False discovery rate (FDR) < 0.05, enrichment score (ES) > 0.5 were set as the significance threshold.

**Functional enrichment analysis**

Utilizing the miRNA target prediction tool starBase (http://starbase.sysu.edu.cn/index.php), the target genes of the 6-miRNA signature were predicted based on 5 datasets, including TargetScan, PITA, miRmap, microT, and miRanda. Metascape is a free online program that provides a comprehensive set of functional annotation tools for researchers to understand biological
functions and characteristic behind large list of genes (http://metascape.org/gp/index.html#/main/step1). We used Metascape to analyzed functional enrichment of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway based on the prognostic target genes of miRNAs and visualized by R “ggplot2” package.

RESULT

The predictive 6-miRNA signature in the train set

The overall design and workflow of this study is presented in Figs. 1. According to the results of the univariate Cox regression analyses, 64 miRNAs associated with survival data were selected for patients with EAC (Table S1). Using the LASSO Cox regression models, we calculated a risk score for each patient based on the 6-miRNA status: Risk score = (-0.6089×has-let-7b) + (-0.1974×has-mir-23a) + (0.3369×has-mir-3074) + (0.0294×has-mir-424) + (0.2421×has-mir-425) + (0.2435×has-mir-505).

By applying the median risk score as the cutoff, EAC patients were classified into a high-risk group and a low-risk group. The risk scores of patients were ranked in the train set, and survival status were also plotted for each patient on a dotplot. The mortality of patients in the high-risk group was much higher than those in the low-risk group (figs. 2A). In addition, a heatmap of the 6-miRNA expression profiles, which were ranked by the risk score of each patient in the train set, showed that the levels of has-mir-3074, has-mir-424, has-mir-425 and has-mir-505 were higher in the high-risk group than those of the low-risk group. The level of has-let-7b and has-mir-23a were lower in the high-risk group than those of the low-risk group (figs. 2C). The Kaplan–Meier curve indicated that patients with lower risk scores generally had higher survival than patients with higher risk scores (figs. 2D). We described the predictive value of the 6-miRNA signature by using a time-dependent ROC curve. The AUC at 1, 2, and 3 years of the signature was 0.860, 0.962, 0.959 respectively (figs. 2B).

Predictive value of 6-miRNA signature in test set and entire set

To predict the prognostic value, we applied the 6-miRNA signature to the test set and entire set. With the median risk score contained from train set as cutoff threshold, patients in test set and entire set were classified into high-risk and low-risk groups. The distribution of risk scores, the expression values of the 6 miRNAs and the survival status of patients ranked according to the risk scores are shown in test set and entire set (Figs. 3A and 3B). In both test set and entire set, patients with the high-risk scores exhibited poorer overall survival significantly than those with the low-risk scores did according to the Kaplan–Meier curve (Figs. 3C and 3D). The 3-year AUC of the 6-miRNA based signature was 0.840 and 0.868 respectively for the test set and the entire set (Figs. 3E and 3F).

To evaluate the independent prognostic value of the 6-miRNA signature, various
Clinicopathological factors were subjected to univariate Cox regression and multivariate Cox regression. The result demonstrated that the risk score of 6-miRNA signature was an independent correlation with OS after adjustment for other clinicopathological factors (HR=2.95, CI 1.43-6.07, p= 0.00338, Table 2). When stratified by clinical factors (age, gender, race, height, weight, alcohol consumption history, Barrett’s disease, TNM stage), a nearly universal result was achieved among the subgroups (Figs. 4), demonstrating that high risk score was highly associated with poor prognosis and vice versa. Regardless of height, weight, TNM stage, alcohol consumption history and Barrett’s disease, the 6-miRNA signature is significantly effective. Moreover, this signature seemed more applicable to male Caucasian patients over 60. Therefore, our findings suggest that the six-miRNA signature can provide predictive value that complements clinical prognostic features.

**Function analysis of the 6-miRNA signature**

BioCarta pathway enrichment was conducted through GSEA in high-risk group in entire set. It revealed that high-risk patients were associated with some pathways, including “proteasome pathway”, “MCM pathway”, “G2 pathway” and “cell cycle pathway” (Figs. 5A). Through a miRNA prediction tool, starBase, it yielded 179 target mRNAs of has-let-7b, 147 target mRNAs of has-mir-23a, 382 target mRNAs of has-mir-424, 37 target mRNAs of has-mir-425 and 11 target mRNAs of has-mir-505. Unfortunately, no target gene for has-mir-3074 was predicted. We conducted functional enrichment of these target genes by GO and KEGG categories. Cellular component, molecular function and biological process of these target genes based on p-values were showed (Figs. 5B, 5C, and 5D). The top 20 KEGG pathways of these target genes were plotted (Figs. 5E). Among these pathways, MAPK signaling pathway, hippo signaling pathway, foxo signaling pathway and TGF-beta signaling pathway were reported to be related to metastasis of cancer (Blum et al. 2019; Janse van Rensburg & Yang 2016; Kim et al. 2018; Sun et al. 2018). Some other pathways are also known to be associated with cancers, such as pathways in cancer, mircoRNAs in cancer, cell cycle, autophagy.

**DISCUSSION**

Although great progress has been made in the field of the pathogenesis and clinical treatment of EAC, the overall morbidity and mortality for EAC have not been improved significantly, which can be attributed to the lack of reliable biomarkers and genetic signatures for proper individualized treatment. Therefore, it is urgent to build the molecular signature of EAC to improve the survival rate and tailor effective personalized treatment. A large number of studies reported that miRNAs can play an important role in the diagnosis of tumors, the prediction of chemotherapy efficacy, and the genetic marker of cancer risk (Mari et al. 2018). The miRNAs have been reported to predict Barrett’s disease development to EAC, Diagnosis, prognosis, and treatment effect in EAC (Maru et al. 2009; Nguyen et al. 2010; Wang et al. 2016; Zhang et al. 2013). Data mining of TCGA is an effective way to identify genetic alterations related to clinical outcomes and screen novel therapeutic targets. In the last decade, miRNAs have attracted...
increasing attention in cancer researches. However, the miRNA prognostic signature of EAC based on TCGA data has been rarely investigated. In this study we used univariate Cox regression analyses to identify 64 miRNAs, among which 6 miRNAs are selected to construct the risk score system for EAC prognosis through LASSO.

Through our analysis, we proposed that has-let-7b and has-mir-23a may increase the survival rate of EAC patients, while has-mir-3074, has-mir-424, has-mir-425 and has-mir-505 may reduce the survival rate of EAC patients. Previous research has identified has-let-7b as a prognostic marker in NSCLC (Hosseini et al. 2018). Importantly, has-let-7b has been reported to inhibits cell proliferation, migration, and invasion in various malignant tumor by targeting different proteins (He et al. 2018; Xu et al. 2014; Yu et al. 2015). It was reported that has-mir-23a played various roles in the initiation, progression, diagnosis, prognosis, and treatment of tumors (Wang et al. 2018). Meanwhile, has-mir-23a was associated with differentiation and carcinogenic process of esophageal squamous cell cancer (Zhu et al. 2013).

There is a few published literature about The function of has-mir-3074 in carcinogenesis, it deserves further investigating. Has-mir-424 has been reported to play a dual role in various cancers. In colorectal cancer, has-mir-424 was identified as a tumor suppressor by suppressing cancer cell growth and enhancing apoptosis (Fang et al. 2018). On the other hand, has-mir-424 was upregulated and correlated with poor survival in ESCC, it can promote cell proliferation by multilayered regulation of cell cycle (Wen et al. 2018).

The impact of has-mir-425 and has-mir-505 on the other cancers seems to be different from its effect on EAC in this bioinformatics analysis. Recent study indicated that has-mir-425 inhibited lung adenocarcinoma cell and promoted cell apoptosis (Liu et al. 2018). Has-mir-425 can also inhibit cell proliferation of renal cell carcinoma by targeting E2F6 (Cai et al. 2018). Meanwhile, a wide range of articles have reported that has-mir-505 suppresses cell proliferation and invasion by targeting certain mRNAs in endometrial carcinoma and gastric cancer (Chen et al. 2016; Tian et al. 2018). However, overexpression of has-mir-425 and has-mir-505 was a poor prognostic factor in our bioinformatics analysis, and they may play a role as oncogenes of EAC.

Functional annotations in high-risk patients with EAC revealed that MCM pathway, G2 pathway and cell cycle pathway was enriched significantly. There are 10 proteins in the family of Minichromosome maintenance complex (MCM), named MCM 1-10 (Nowinska & Dziegiel 2010). It has been reported that MCM2-7 paly an important role as the eukaryotic replicative helicase due to its unwinding DNA and traveling with the fork (Bochman & Schwacha 2008; Labib et al. 2000), along with the cyclin dependent kinases (CDKs) as master regulators of the cell cycle and the initiator proteins of DNA replication, such as the Origin Recognition Complex (ORC), Cdc6/18 (Chen et al. 2007; Diffley et al. 1994). There is evidence that high expression of MCM4 and MCM7 were associated with lymph node metastasis and shorter survival in EAC (Choy et al. 2016). Based on the result of GSEA, molecular function of GO, and KEGG, 6-
miRNA signature maybe is involved in regulation of cell cycle and DNA replication.

This study has certain limitations. First, the initial screening univariate Cox regression analyses included only 272 miRNAs due to eliminating very low expression of miRNAs, whereas more than 4000 human miRNAs have been discovered at present (Chou et al. 2018). Although the 6-miRNA signature can predict prognosis of EAC well, other miRNAs which have good predictive ability of prognosis may have been missed. Second, Due to the patients limitation of TCGA, there are only 87 EACs, and the stratified sample size in the subgroup analysis become very small. Third, in this study, we lack the external validation cohorts, which can convincingly validate the miRNA signature. Therefore, further studies will be needed to validate these findings using larger numbers of patients, and to explore potential molecular functions of the six separate miRNAs in EAC.

**CONCLUSIONS**

In summary, we construct a novel 6-miRNA-expression-based risk model based on TCGA dataset which could be used as an independent prognostic factor for patients with EAC. In addition, the miRNA signature can help improve our understanding of clinical decision-making as potential biomarkers and targets for patients with EAC.

**ACKNOWLEDGEMENTS**

This study is based on data from the Cancer Genome Atlas (TCGA) database.

**REFERENCES**


Nowinska K, and Dziegiel P. 2010. The role of MCM proteins in cell proliferation and tumorigenesis. *Postepy Hig Med Dosw (Online)* 64:627-635.


**Table 1** (on next page)

Clinical characteristics of EAC patients
<table>
<thead>
<tr>
<th>Variables</th>
<th>Case, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample number</td>
<td>84</td>
</tr>
<tr>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>&lt;60</td>
<td>29 (34.5)</td>
</tr>
<tr>
<td>≥60</td>
<td>55 (65.5)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>72 (85.7)</td>
</tr>
<tr>
<td>female</td>
<td>12 (14.3)</td>
</tr>
<tr>
<td>Height</td>
<td></td>
</tr>
<tr>
<td>&lt;175</td>
<td>46 (54.8)</td>
</tr>
<tr>
<td>≥175</td>
<td>38 (45.2)</td>
</tr>
<tr>
<td>NA</td>
<td>4 (4.8)</td>
</tr>
<tr>
<td>Weight</td>
<td></td>
</tr>
<tr>
<td>&lt;85</td>
<td>46 (54.8)</td>
</tr>
<tr>
<td>≥85</td>
<td>37 (44.0)</td>
</tr>
<tr>
<td>NA</td>
<td>1 (1.2)</td>
</tr>
<tr>
<td>Race</td>
<td></td>
</tr>
<tr>
<td>asian</td>
<td>1 (1.2)</td>
</tr>
<tr>
<td>white</td>
<td>66 (78.6)</td>
</tr>
<tr>
<td>NA</td>
<td>17 (20.2)</td>
</tr>
<tr>
<td>Event</td>
<td></td>
</tr>
<tr>
<td>alive</td>
<td>46 (54.8)</td>
</tr>
<tr>
<td>dead</td>
<td>38 (45.2)</td>
</tr>
<tr>
<td>alcohol history</td>
<td></td>
</tr>
<tr>
<td>no</td>
<td>27 (32.1)</td>
</tr>
<tr>
<td>yes</td>
<td>56 (66.7)</td>
</tr>
<tr>
<td>NA</td>
<td>1 (1.2)</td>
</tr>
<tr>
<td>Barrett's disease</td>
<td></td>
</tr>
<tr>
<td>no</td>
<td>52 (61.9)</td>
</tr>
<tr>
<td>yes</td>
<td>26 (31.0)</td>
</tr>
<tr>
<td>NA</td>
<td>6 (7.1)</td>
</tr>
<tr>
<td>Tumor size</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>21 (25.0)</td>
</tr>
<tr>
<td>2</td>
<td>14 (16.7)</td>
</tr>
<tr>
<td>3</td>
<td>45 (53.6)</td>
</tr>
<tr>
<td>4</td>
<td>1 (1.2)</td>
</tr>
<tr>
<td>TX</td>
<td>3 (3.6)</td>
</tr>
<tr>
<td>Lymph node status</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>21 (25.0)</td>
</tr>
<tr>
<td>1</td>
<td>47 (56.0)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>2</td>
<td>6 (7.1)</td>
</tr>
<tr>
<td>3</td>
<td>5 (6.0)</td>
</tr>
<tr>
<td>NX</td>
<td>5 (6.0)</td>
</tr>
<tr>
<td>metastasis (%)</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>57 (67.9)</td>
</tr>
<tr>
<td>1</td>
<td>11 (13.1)</td>
</tr>
<tr>
<td>MX</td>
<td>12 (14.3)</td>
</tr>
<tr>
<td>NA</td>
<td>4 (4.8)</td>
</tr>
<tr>
<td>stage (%)</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>12 (14.3)</td>
</tr>
<tr>
<td>II</td>
<td>24 (28.6)</td>
</tr>
<tr>
<td>III</td>
<td>33 (39.3)</td>
</tr>
<tr>
<td>IV</td>
<td>11 (13.1)</td>
</tr>
<tr>
<td>NA</td>
<td>4 (4.8)</td>
</tr>
</tbody>
</table>

NA, non available.
Table 2 (on next page)

Univariate and multivariate COX regression analyses of clinicopathologic factors
Table 2. Univariate and multivariate COX regression analyses of clinicopathologic factors

<table>
<thead>
<tr>
<th>variables</th>
<th>Univariate analysis</th>
<th></th>
<th></th>
<th>Multivariate analysis</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR</td>
<td>95%CI</td>
<td>Pvalue</td>
<td>HR</td>
<td>95%CI</td>
</tr>
<tr>
<td>Riskscore</td>
<td>3.41</td>
<td>1.70-6.84</td>
<td>0.00053</td>
<td>2.95</td>
<td>1.43-6.07</td>
</tr>
<tr>
<td>Age (≥60 vs &lt;60)</td>
<td>0.89</td>
<td>0.44-1.81</td>
<td>0.75169</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (male vs female)</td>
<td>0.68</td>
<td>0.20-2.31</td>
<td>0.53893</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Height (≥175 vs &lt;175cm)</td>
<td>0.80</td>
<td>0.39-1.62</td>
<td>0.53522</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight (≥85 vs &lt;85kg)</td>
<td>1.07</td>
<td>0.53-2.15</td>
<td>0.84364</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol consumption (yes vs no)</td>
<td>0.46</td>
<td>0.23-0.92</td>
<td>0.02859</td>
<td>0.67</td>
<td>0.32-1.40</td>
</tr>
<tr>
<td>Barrett's disease (yes vs no)</td>
<td>1.16</td>
<td>0.56-2.37</td>
<td>0.69146</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage (III+IV vs I+II)</td>
<td>2.30</td>
<td>1.08-4.91</td>
<td>0.03086</td>
<td>1.95</td>
<td>0.88-4.29</td>
</tr>
</tbody>
</table>

HR, Hazard Ratio.
Figure 1 (on next page)

Flow chart of data preparation, processing, analysis and validation in this study
EAC data from TCGA (84 samples, 272 miRNAs)

Train set (42 samples)

Screening prognostic miRNAs (Univariate Cox survival analysis)

LASSO

6 miRNA signature establishment

Validation in Test set (42 samples) and entire set (84 samples)

Gene set enrichment analysis

Kaplan-Meier curves

Screening independent risk factors of OS
Figure 2 (on next page)

The 6-miRNA signature predicted the OS of EAC patients in the train set.

(A) The 6-miRNA based risk score and survival status of EAC patients. (B) Receiver operating characteristic (ROC) analyzes the sensitivity and specificity of the survival time by risk score based on the 6-miRNA signature. (C) Expression heatmap of the 6 miRNAs corresponding to each sample which ranks in order of risk score. (D) Kaplan-Meier analysis for OS using the 6-miRNA signature.
Figure 3 (on next page)

The 6-miRNA signature predicted the OS of EAC patients in test and entire set.

The miRNA signature risk score distribution and heatmap of the miRNA expression profiles in test set (A) and entire set (B). survival curves of high- and low- risk samples in test set (C) and entire set (D). Time dependent ROC curve for accuracy of the predicting risk score system in test set (E) and entire set (F).
ROC at time t=1, AUC=76.1

Survival Curve (p=0.00575)

Survival Curve (p=2.99e−05)

AUC at 1 year: 0.761
AUC at 2 years: 0.785
AUC at 3 years: 0.840

AUC at 1 year: 0.802
AUC at 2 years: 0.837
AUC at 3 years: 0.868

Survival Rate

Time (year)

Risk Score

Risk

High

Low

num

futime

fustat

dead

alive

hsa_let_7b
hsa_mir_23a
hsa_mir_3074
hsa_mir_424
hsa_mir_425
hsa_mir_505

Risk Score

Risk

High

Low

num

futime

fustat

dead

alive

hsa_let_7b
hsa_mir_23a
hsa_mir_3074
hsa_mir_424
hsa_mir_425
hsa_mir_505

AUC

1−Specificity

Sensitivity

AUC

1−Specificity

Sensitivity

AUC
Figure 4 (on next page)

Stratified analysis of overall survival in the entire set

Kaplan-Meier analysis for OS in subgroups stratified by age (A), gender (B), height (C), weight (D), Alcohol consumption (E), Barrett's esophagitis (F), Caucasian (G).
A) Survival rates over time for different age groups:
- Age ≤ 60:
  - High-risk group: Red line
  - Low-risk group: Blue line
  - Survival rate at 10 years: 0.25
  - P-value: 0.053
- Age > 60:
  - High-risk group: Red line
  - Low-risk group: Blue line
  - Survival rate at 10 years: 0.25
  - P-value: 0.00041

B) Survival rates over time for different genders:
- Male:
  - High-risk group: Red line
  - Low-risk group: Blue line
  - Survival rate at 10 years: 0.25
  - P-value: 0.0017
- Female:
  - High-risk group: Red line
  - Low-risk group: Blue line
  - Survival rate at 10 years: 0.25
  - P-value: 0.0027

C) Survival rates over time for different height groups:
- Height < 175cm:
  - High-risk group: Red line
  - Low-risk group: Blue line
  - Survival rate at 10 years: 0.25
  - P-value: 0.0017
- Height > 175cm:
  - High-risk group: Red line
  - Low-risk group: Blue line
  - Survival rate at 10 years: 0.25
  - P-value: 0.0056

D) Survival rates over time for different weight groups:
- Weight < 85kg:
  - High-risk group: Red line
  - Low-risk group: Blue line
  - Survival rate at 10 years: 0.25
  - P-value: 0.0041
- Weight > 85kg:
  - High-risk group: Red line
  - Low-risk group: Blue line
  - Survival rate at 10 years: 0.25
  - P-value: 0.053

E) Survival rates over time for different alcohol consumption habits:
- No alcohol consumption:
  - High-risk group: Red line
  - Low-risk group: Blue line
  - Survival rate at 10 years: 0.25
  - P-value: 0.047
- With alcohol consumption:
  - High-risk group: Red line
  - Low-risk group: Blue line
  - Survival rate at 10 years: 0.25
  - P-value: 0.0012

F) Survival rates over time for Barrett's esophagitis:
- No Barrett's esophagitis:
  - High-risk group: Red line
  - Low-risk group: Blue line
  - Survival rate at 10 years: 0.25
  - P-value: 0.0017
- Barrett's esophagitis history:
  - High-risk group: Red line
  - Low-risk group: Blue line
  - Survival rate at 10 years: 0.25
  - P-value: 0.0027

G) Survival rates over time for different ethnic groups:
- Caucasian:
  - High-risk group: Red line
  - Low-risk group: Blue line
  - Survival rate at 10 years: 0.25
  - P-value: 0.0032
Gene enrichment analysis, GO, and KEGG pathways of mRNA associated with the 6-miRNA signature.

(A) Gene enrichment analysis in high-risk patients. The cellular component (B), molecular function (C) and biological process (D) of GO of the target genes. (F) The bar chart of significantly KEGG pathways of the target genes.