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

**Tian Lan** [1, 2], **Yunyan Lu** [3], **Zunqiang Xiao** [1], **Haibin Xu** [2], **Junling He** [2], **Zujian Hu** [2], **Weimin Mao** [Corresp. 4, 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'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. Here, we aimed at developing 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). Eighty-four patients with EAC were classified into a training set and a test set randomly. Using univariate Cox regression analysis and the least absolute shrinkage and selection operator (LASSO), we identified prognostic factors and constructed a prognostic miRNA signature. The accuracy of the signature was evaluated by the receiver operating curve (ROC).

**Result** In general, in the training set, 6 miRNAs (hsa-mir-425, hsa-let-7b, hsa-mir-23a, hsa-mir-3074, hsa-mir-424, hsa-mir-505) displayed good prognostic power as markers of overall survival for EAC patients. Relative to patients in the low-risk group, those assigned to the high-risk group according to their risk scores of the designed miRNA model displayed reduced overall survival. 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 training, 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 several pathways involved in carcinogenesis, especially cell cycle.

**Conclusion** In summary, a novel 6-miRNA expression-based prognostic signature derived from the EAC data of TCGA was constructed and validated for predicting the prognosis of EAC.

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

Tian Lan[1,2], Yunyan Lu[3], Zunqiang Xiao[1], Haibin Xu[2], Junling He[2], Zujian Hu [2], Weimin Mao[4,5]

[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'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[4,5]

Email address: maowm1218@163.com

## ABSTRACT

**Background**

The MicroRNAs (miRNAs) have been validated as prognostic markers in many cancers. Here, we aimed at developing 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). Eighty-four patients with EAC were classified into a training set and a test set randomly. Using univariate Cox regression analysis and the least absolute shrinkage and selection operator (LASSO), we identified prognostic factors and constructed a prognostic miRNA signature. The accuracy of the signature was evaluated by the receiver operating curve (ROC).

**Result**

In general, in the training set, 6 miRNAs (hsa-mir-425, hsa-let-7b, hsa-mir-23a, hsa-mir-3074, hsa-mir-424, hsa-mir-505) displayed good prognostic power as markers of overall survival for EAC patients. Relative to patients in the low-risk group, those assigned to the high-risk group according to their risk scores of the designed miRNA model displayed reduced overall survival. The 6-miRNA model identified in the training set 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 training, 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 several pathways involved in carcinogenesis, especially cell cycle.

**Conclusion**

In summary, a novel 6-miRNA expression-based prognostic signature derived from the EAC data of TCGA was constructed and validated for predicting the prognosis of EAC.

**Keywords**: Esophagus adenocarcinoma, TCGA, Prognosis, Bioinformatics

## INTRODUCTION

Globally, esophagus cancer was ranked seventh among the leading types of cancers and sixth among the leading causes of cancer mortality in 2018 according to the Global Cancer Observatory (GCO) (Fitzmaurice 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). Generally, two types of malignancies are diagnosed: adenocarcinoma (10%) and squamous cell carcinoma (90% of cases). The prevalence of esophagus adenocarcinoma (EAC) has rapidly increased over the past few decades (Thrift & Whiteman 2012). The prognosis of EAC is poor and its 5-year overall survival rate is 30% (Hirst et al. 2011). Due to the poor outcomes of EAC, it is important to reveal the mechanisms leading to 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 small noncoding RNA transcripts that are made of estimated 22 nucleotides (Lujambio & Lowe 2012). The predominant function of miRNAs is to regulate protein translation by binding to target messenger RNAs (mRNAs), and inhibit mRNA translation (Krol et al. 2010). They have recently been validated and applied 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).

Many studies focused 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. Yet, the miRNA expression landscape in EAC is not clearly understood. Over the past few years, some studies reported the significant role of miRNAs in the molecular diagnosis and prognosis of EAC. A 4-miRNA expression profile score can provide a validated approach for predicting pathological complete response rates (pCR) to neoadjuvant treatment in EAC (Skinner et al. 2014). In addition, a 3-miRNA (miR-99b and miR-199a_3p and _5p) signature ~~is~~ correlated with patient survival and occurrence of lymph node metastasis (Feber et al. 2011). However, these findings were based on a small number of patients.

The Cancer Genome Atlas (TCGA), a landmark cancer genomics program, is a reservoir of large-scale miRNA-sequencing datasets spanning 33 cancer types. In the present investigation, we constructed a prognostic risk score system on the basis of miRNA datasets from TCGA to predict the prognosis of EAC. Furthermore, pathway enrichment and gene oncology annotation analyses were performed to understand the probable cellular functions of mRNAs associated with this signature.

## MATERIALS AND METHODS

### RNA-Seq and clinicopathological data of EAC patients

From TCGA data portal (https://portal.gdc.cancer.gov/), RNA-seq data and associated clinical information were downloaded in January 2019. The annotation information was provided by GENCODE datasets (www.gencodegenes.org). Given that some miRNAs and mRNAs display little or no expression in some tissues or do not vary sufficiently, only those with raw count value >20 in more than 80% of samples were retained for further analysis. Once normalization by edgeR was completed, this was followed by conversion of the expression patterns of miRNAs and mRNAs to log2 (normalized value +1) in preparation for the subsequent processing. Samples with a < 1-month censor time are removed, because they cannot 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, and 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 came from the TCGA database, no further approval was required from the Ethics Committee.

## Construction and validation of the miRNA risk score

Eighty-four patients were stratified to 2 categories in a random manner: training set = 42, test set = 42. Training set was analyzed to build a miRNA model that was later confirmed in test and entire sets. In the training 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 miRNA. 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). We determined the lambda by using the cross-validation routine cv.glmnet with an n-fold equal to 10. LASSO was performed to reduce above selected prognostic miRNAs further and to construct the risk score system.

For determination of survival risks, a prognostic model was created on the basis of miRNA data as follows:

$$\text{Risk score} = \sum_{i=1}^{n} \beta i * gene\ i$$

β stands for the coefficient of the miRNA, and gene refers to miRNA expression value.

Using the median score in training set as the cutoff, we stratified the subjects to low-risk and high-risk groups. The Kaplan–Meier (KM) and log-rank methods were applied to compare the survival rate between the groups by using the R “survival” package. The time-dependent receiver-operating characteristic (ROC) curve was plotted by using the R “timeROC” package to evaluate specificity and sensitivity of the miRNA expression-based prognostic signature. Thereafter, this signature was validated in test set and entire set. ROC and KM curves were also carried out to ~~the~~ validate accuracy and feasibility 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 $P < 0.05$ represented statistical significance.

## Gene set enrichment analysis

Subjects were stratified to two groups (high and low) based on the risk score of the 6-miRNA signature. 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) served 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

Using 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 platform having a large-scale set of functional annotation tools to understand biological mechanisms behind a large pool of genes (http://metascape.org/gp/index.html#/main/step1). We used Metascape to analyze functional enrichment of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and Gene Ontology (GO) based on the prognostic target genes of miRNAs and visualized by R "gglot2" package.

## RESULTS

### The predictive 6-miRNA signature for the training set

The overall design and workflow of this study was presented in Fig. 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). The lambda value was set by using the lambda.min, which is the value of lambda giving minimum mean cross-validated error, and then 6 miRNAs with nonzero coefficients were defined (Fig. S1). Based on the LASSO Cox regression models, a risk score was determined for each subject according to 6-miRNA status: Risk score = (-0.6089×hsa-let-7b) + (-0.1974×hsa-mir-23a) + (0.3369×hsa-mir-3074) + (0.0294×hsa-mir-424) + (0.2421×hsa-mir-425) + (0.2435×hsa-mir-505).

In the training set, the patients with EAC were divided into a high-risk group and a low-risk group. The risk scores of patients were ranked, and the dotplot was developed for the survival status of each patient. Compared with the mortality of patients in the high-risk group, those in the low-risk group was much lower (figs. 2A, 2C). Moreover, based on a heatmap of the 6-miRNA profile, the levels of hsa-mir-3074, hsa-mir-424, hsa-mir-425 and hsa-mir-505 were lower in the low-risk group than those of the high-risk group. The level of hsa-let-7b and hsa-mir-23a were higher in the low-risk group than those of the high-risk group (fig. 2D). The KM curve indicated that the survival time of patients in the high-risk group was shorter than those in the low-risk group (fig. 2E). 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 (fig. 2B).

### The predictive power of 6-miRNA signature in test set and entire set

The 6-miRNA signature was applied to the test set and the entire set for evaluation of its prognostic value. The distribution of risk scores, the expression values of 6 miRNAs and the survival status of patients ranked according to the risk scores were presented in test set (Figs. 3A, 3C, and 3E) and entire set (Figs. 3B, 3D, and 3F). In test set and entire set, patients with the low-risk scores exhibited better overall survival than those with the high-risk scores based on the KM curve (Figs. 3G and 3H). 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. 3I and 3J).

To assess the independent prognostic value of the 6-miRNA signature, various clinicopathological factors were subjected to univariate Cox regression and multivariate Cox

~~regression~~. The result indicated that the 6-miRNA signature was an independent prognostic factor 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, caucasian, height, weight, alcohol consumption history, Barrett's disease, TNM stage), a nearly universal result was obtained for all subgroups (Figs. 4), showing that high risk score was strongly 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. Therefore, the present results suggest that the 6-miRNA signature can predict the clinical prognosis of EAC.

### Functional analysis of the 6-miRNA signature

BioCarta pathway enrichment was conducted through GSEA in high-risk group in the 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, 5B, 5C, and 5D). Through a miRNA prediction tool, starBase, 179 target mRNAs for hsa-let-7b, 147 for hsa-mir-23a, 382 for hsa-mir-424, 37 for hsa-mir-425 and 11 for hsa-mir-505 were obtained. Unfortunately, no target gene for hsa-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. 5E, 5F, and 5G). The top 20 KEGG pathways of these target genes were plotted (Figs. 5H). 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, microRNAs 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 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 a key role in the diagnosis of tumors, the prediction of chemotherapy efficacy, and ~~the biomarker 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 research. However, very few studies have assessed the prognostic value of miRNA signature for patients with EAC on the basis of TCGA data. 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 suggested that hsa-let-7b and hsa-mir-23a may enhance the survival rate of EAC patients, while hsa-mir-3074, hsa-mir-424, hsa-mir-425 and hsa-mir-505 may reduce the survival rate of EAC patients. Previous research has identified hsa-let-7b as a prognostic marker in NSCLC (Hosseini et al. 2018). Importantly, hsa-let-7b has been reported to inhibit 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 hsa-mir-23a played various roles in the initiation, progression, diagnosis, prognosis, and treatment of tumors (Wang et al. 2018). Meanwhile, hsa-mir-23a was associated with differentiation and carcinogenic process of esophageal squamous cell cancer (Zhu et al. 2013).

Few studies have been published on the function of hsa-mir-3074 in carcinogenesis~~,~~ it deserves further investigation. Hsa-mir-424 was recognized to play a dual role in various cancers. In colorectal cancer, hsa-mir-424 was identified as a tumor suppressor ~~by suppressing~~ cancer cell growth and ~~enhancing~~ apoptosis (Fang et al. 2018). In addition, hsa-mir-424 was upregulated and correlated with poor survival in esophageal squamous cell carcinoma~~,~~ it can promote cell proliferation by multilayered regulation of cell cycle (Wen et al. 2018).

The impact of hsa-mir-425 and hsa-mir-505 on other cancers seems to differ from its effect on EAC based on our bioinformatics analysis. A recent study indicated that hsa-mir-425 inhibited lung adenocarcinoma cell and promoted cell apoptosis (Liu et al. 2018). Hsa-mir-425 can also inhibit cell proliferation of renal cell carcinoma by targeting E2F6 (Cai et al. 2018). Meanwhile, several articles have reported that hsa-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 hsa-mir-425 and hsa-mir-505 was a poor prognostic factor in this study, 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 play 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 may be 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 after elimination of very low expressed 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~~ for prognosis may have been missed. Second, due to the ~~patient~~ limitation of TCGA, there are only 84 EAC patients, and fewer number of patients were included in subgroup analyses. Third, there were no external validation cohorts in this study 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 constructed a novel 6-miRNA-expression-based risk model based on TCGA dataset which displayed the potential to be 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

Blum AE, Venkitachalam S, Ravillah D, Chelluboyina AK, Kieber-Emmons AM, Ravi L, Kresak A, Chandar AK, Markowitz SD, Canto MI, Wang JS, Shaheen NJ, Guo Y, Shyr Y, Willis JE, Chak A, Varadan V, and Guda K. 2019. Systems Biology Analyses Reveal Hyperactivation of Transforming Growth Factor beta and JNK Signaling pathways in Esophageal Cancer. *Gastroenterology*. 10.1053/j.gastro.2019.01.263

Bochman ML, and Schwacha A. 2008. The Mcm2-7 complex has in vitro helicase activity. *Mol Cell* 31:287-293. 10.1016/j.molcel.2008.05.020

Cai Q, Zhao A, Ren LG, Chen J, Liao KS, Wang ZS, and Zhang W. 2018. MiR-425 involves in the development and progression of renal cell carcinoma by inhibiting E2F6. *Eur Rev Med Pharmacol Sci* 22:6300-6307. 10.26355/eurrev_201810_16040

Chen S, de Vries MA, and Bell SP. 2007. Orc6 is required for dynamic recruitment of Cdt1 during repeated Mcm2-7 loading. *Genes Dev* 21:2897-2907. 10.1101/gad.1596807

Chen S, Sun KX, Liu BL, Zong ZH, and Zhao Y. 2016. MicroRNA-505 functions as a tumor suppressor in endometrial cancer by targeting TGF-alpha. *Mol Cancer* 15:11. 10.1186/s12943-016-0496-4

Chou CH, Shrestha S, Yang CD, Chang NW, Lin YL, Liao KW, Huang WC, Sun TH, Tu SJ, Lee WH, Chiew MY, Tai CS, Wei TY, Tsai TR, Huang HT, Wang CY, Wu HY, Ho SY, Chen PR, Chuang CH, Hsieh PJ, Wu YS, Chen WL, Li MJ, Wu YC, Huang XY, Ng FL, Buddhakosai

W, Huang PC, Lan KC, Huang CY, Weng SL, Cheng YN, Liang C, Hsu WL, and Huang HD. 2018. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res* 46:D296-D302. 10.1093/nar/gkx1067

Choy B, LaLonde A, Que J, Wu T, and Zhou Z. 2016. MCM4 and MCM7, potential novel proliferation markers, significantly correlated with Ki-67, Bmi1, and cyclin E expression in esophageal adenocarcinoma, squamous cell carcinoma, and precancerous lesions. *Hum Pathol* 57:126-135. 10.1016/j.humpath.2016.07.013

Diffley JF, Cocker JH, Dowell SJ, and Rowley A. 1994. Two steps in the assembly of complexes at yeast replication origins in vivo. *Cell* 78:303-316.

Fang Y, Liang X, Xu J, and Cai X. 2018. miR-424 targets AKT3 and PSAT1 and has a tumor-suppressive role in human colorectal cancer. *Cancer Manag Res* 10:6537-6547. 10.2147/CMAR.S185789

Feber A, Xi L, Pennathur A, Gooding WE, Bandla S, Wu M, Luketich JD, Godfrey TE, and Litle VR. 2011. MicroRNA prognostic signature for nodal metastases and survival in esophageal adenocarcinoma. *Ann Thorac Surg* 91:1523-1530. 10.1016/j.athoracsur.2011.01.056

Fitzmaurice C, Akinyemiju TF, Al Lami FH, Alam T, Alizadeh-Navaei R, Allen C, Alsharif U, Alvis-Guzman N, Amini E, Anderson BO, Aremu O, Artaman A, Asgedom SW, Assadi R, Atey TM, Avila-Burgos L, Awasthi A, Ba Saleem HO, Barac A, Bennett JR, Bensenor IM, Bhakta N, Brenner H, Cahuana-Hurtado L, Castaneda-Orjuela CA, Catala-Lopez F, Choi JJ, Christopher DJ, Chung SC, Curado MP, Dandona L, Dandona R, das Neves J, Dey S, Dharmaratne SD, Doku DT, Driscoll TR, Dubey M, Ebrahimi H, Edessa D, El-Khatib Z, Endries AY, Fischer F, Force LM, Foreman KJ, Gebrehiwot SW, Gopalani SV, Grosso G, Gupta R, Gyawali B, Hamadeh RR, Hamidi S, Harvey J, Hassen HY, Hay RJ, Hay SI, Heibati B, Hiluf MK, Horita N, Hosgood HD, Ilesanmi OS, Innos K, Islami F, Jakovljevic MB, Johnson SC, Jonas JB, Kasaeian A, Kassa TD, Khader YS, Khan EA, Khan G, Khang YH, Khosravi MH, Khubchandani J, Kopec JA, Kumar GA, Kutz M, Lad DP, Lafranconi A, Lan Q, Legesse Y, Leigh J, Linn S, Lunevicius R, Majeed A, Malekzadeh R, Malta DC, Mantovani LG, McMahon BJ, Meier T, Melaku YA, Melku M, Memiah P, Mendoza W, Meretoja TJ, Mezgebe HB, Miller TR, Mohammed S, Mokdad AH, Moosazadeh M, Moraga P, Mousavi SM, Nangia V, Nguyen CT, Nong VM, Ogbo FA, Olagunju AT, Pa M, Park EK, Patel T, Pereira DM, Pishgar F, Postma MJ, Pourmalek F, Qorbani M, Rafay A, Rawaf S, Rawaf DL, Roshandel G, Safiri S, Salimzadeh H, Sanabria JR, Santric Milicevic MM, Sartorius B, Satpathy M, Sepanlou SG, Shackelford KA, Shaikh MA, Sharif-Alhoseini M, She J, Shin MJ, Shiue I, Shrime MG, Sinke AH, Sisay M, Sligar A, Sufiyan MB, Sykes BL, Tabares-Seisdedos R, Tessema GA, Topor-Madry R, Tran TT, Tran BX, Ukwaja KN, Vlassov VV, Vollset SE, Weiderpass E, Williams HC, Yimer NB, Yonemoto N, Younis MZ, Murray CJL, and Naghavi M. 2018. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2016: A Systematic Analysis for the Global Burden of Disease Study. *JAMA Oncol* 4:1553-1568. 10.1001/jamaoncol.2018.2706

Gao J, Kwan PW, and Shi D. 2010. Sparse kernel learning with LASSO and Bayesian inference algorithm. *Neural Netw* 23:257-264. 10.1016/j.neunet.2009.07.001

He Z, Deng W, Jiang B, Liu S, Tang M, Liu Y, and Zhang J. 2018. Hsa-let-7b inhibits cell proliferation by targeting PLK1 in HCC. *Gene* 673:46-55. 10.1016/j.gene.2018.06.047

Hirst J, Smithers BM, Gotley DC, Thomas J, and Barbour A. 2011. Defining cure for esophageal cancer: analysis of actual 5-year survivors following esophagectomy. *Ann Surg Oncol* 18:1766-1774. 10.1245/s10434-010-1508-z

Hosseini SM, Soltani BM, Tavallaei M, Mowla SJ, Tafsiri E, Bagheri A, and Khorshid HRK. 2018. Clinically Significant Dysregulation of hsa-miR-30d-5p and hsa-let-7b Expression in Patients with Surgically Resected Non-Small Cell Lung Cancer. *Avicenna J Med Biotechnol* 10:98-104.

Janse van Rensburg HJ, and Yang X. 2016. The roles of the Hippo pathway in cancer metastasis. *Cell Signal* 28:1761-1772. 10.1016/j.cellsig.2016.08.004

Kim CG, Lee H, Gupta N, Ramachandran S, Kaushik I, Srivastava S, Kim SH, and Srivastava SK. 2018. Role of Forkhead Box Class O proteins in cancer progression and metastasis. *Semin Cancer Biol* 50:142-151. 10.1016/j.semcancer.2017.07.007

Krol J, Loedige I, and Filipowicz W. 2010. The widespread regulation of microRNA biogenesis, function and decay. *Nat Rev Genet* 11:597-610. 10.1038/nrg2843

Labib K, Tercero JA, and Diffley JF. 2000. Uninterrupted MCM2-7 function required for DNA replication fork progression. *Science* 288:1643-1647.

Leidner RS, Ravi L, Leahy P, Chen Y, Bednarchik B, Streppel M, Canto M, Wang JS, Maitra A, Willis J, Markowitz SD, Barnholtz-Sloan J, Adams MD, Chak A, and Guda K. 2012. The microRNAs, MiR-31 and MiR-375, as candidate markers in Barrett's esophageal carcinogenesis. *Genes Chromosomes Cancer* 51:473-479. 10.1002/gcc.21934

Li X, Kleeman S, Coburn SB, Fumagalli C, Perner J, Jammula S, Pfeiffer RM, Orzolek L, Hao H, Taylor PR, Miremadi A, Galeano-Dalmau N, Lao-Sirieix P, Tennyson M, MacRae S, Cook MB, and Fitzgerald RC. 2018. Selection and Application of Tissue microRNAs for Nonendoscopic Diagnosis of Barrett's Esophagus. *Gastroenterology* 155:771-783 e773. 10.1053/j.gastro.2018.05.050

Liu R, Wang F, Guo Y, Yang J, Chen S, Gao X, and Wang X. 2018. MicroRNA-425 promotes the development of lung adenocarcinoma via targeting A disintegrin and metalloproteinases 9 (ADAM9). *Onco Targets Ther* 11:4065-4073. 10.2147/OTT.S160871

Lujambio A, and Lowe SW. 2012. The microcosmos of cancer. *Nature* 482:347-355. 10.1038/nature10888

Mari L, Hoefnagel SJM, Zito D, van de Meent M, van Endert P, Calpe S, Sancho Serra MDC, Heemskerk MHM, van Laarhoven HWM, Hulshof M, Gisbertz SS, Medema JP, van Berge Henegouwen MI, Meijer SL, Bergman J, Milano F, and Krishnadath KK. 2018. microRNA 125a Regulates MHC-I Expression on Esophageal Adenocarcinoma Cells, Associated With Suppression of Antitumor Immune Response and Poor Outcomes of Patients. *Gastroenterology* 155:784-798. 10.1053/j.gastro.2018.06.030

Maru DM, Singh RR, Hannah C, Albarracin CT, Li YX, Abraham R, Romans AM, Yao H, Luthra MG, Anandasabapathy S, Swisher SG, Hofstetter WL, Rashid A, and Luthra R. 2009.

MicroRNA-196a is a potential marker of progression during Barrett's metaplasia-dysplasia-invasive adenocarcinoma sequence in esophagus. *Am J Pathol* 174:1940-1948. 10.2353/ajpath.2009.080718

Moya L, Meijer J, Schubert S, Matin F, and Batra J. 2019. Assessment of miR-98-5p, miR-152-3p, miR-326 and miR-4289 Expression as Biomarker for Prostate Cancer Diagnosis. *Int J Mol Sci* 20. 10.3390/ijms20051154

Nguyen GH, Schetter AJ, Chou DB, Bowman ED, Zhao R, Hawkes JE, Mathe EA, Kumamoto K, Zhao Y, Budhu A, Hagiwara N, Wang XW, Miyashita M, Casson AG, and Harris CC. 2010. Inflammatory and microRNA gene expression as prognostic classifier of Barrett's-associated esophageal adenocarcinoma. *Clin Cancer Res* 16:5824-5834. 10.1158/1078-0432.CCR-10-1110

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

Parizadeh SM, Jafarzadeh-Esfehani R, Ghandehari M, Goldani F, Parizadeh SMR, Hassanian SM, Ghayour-Mobarhan M, Ferns GA, and Avan A. 2019. MicroRNAs as potential diagnostic and prognostic biomarkers in hepatocellular carcinoma. *Curr Drug Targets*. 10.2174/1389450120666190307095720

Pennathur A, Gibson MK, Jobe BA, and Luketich JD. 2013. Oesophageal carcinoma. *Lancet* 381:400-412. 10.1016/S0140-6736(12)60643-6

Revilla-Nuin B, Parrilla P, Lozano JJ, de Haro LF, Ortiz A, Martinez C, Munitiz V, de Angulo DR, Bermejo J, Molina J, Cayuela ML, and Yelamos J. 2013. Predictive value of MicroRNAs in the progression of barrett esophagus to adenocarcinoma in a long-term follow-up study. *Ann Surg* 257:886-893. 10.1097/SLA.0b013e31826ddba6

Skinner HD, Lee JH, Bhutani MS, Weston B, Hofstetter W, Komaki R, Shiozaki H, Wadhwa R, Sudo K, Elimova E, Song S, Ye Y, Huang M, Ajani J, and Wu X. 2014. A validated miRNA profile predicts response to therapy in esophageal adenocarcinoma. *Cancer* 120:3635-3641. 10.1002/cncr.28911

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, and Mesirov JP. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102:15545-15550. 10.1073/pnas.0506580102

Sun W, Ping W, Tian Y, Zou W, Liu J, and Zu Y. 2018. miR-202 Enhances the Anti-Tumor Effect of Cisplatin on Non-Small Cell Lung Cancer by Targeting the Ras/MAPK Pathway. *Cell Physiol Biochem* 51:2160-2171. 10.1159/000495835

Thrift AP, and Whiteman DC. 2012. The incidence of esophageal adenocarcinoma continues to rise: analysis of period and birth cohort effects on recent trends. *Ann Oncol* 23:3155-3162. 10.1093/annonc/mds181

Tian L, Wang ZY, Hao J, and Zhang XY. 2018. miR-505 acts as a tumor suppressor in gastric cancer progression through targeting HMGB1. *J Cell Biochem*. 10.1002/jcb.28082

Wang N, Tan HY, Feng YG, Zhang C, Chen F, and Feng Y. 2018. microRNA-23a in Human Cancer: Its Roles, Mechanisms and Therapeutic Relevance. *Cancers (Basel)* 11. 10.3390/cancers11010007

Wang Y, Zhao Y, Herbst A, Kalinski T, Qin J, Wang X, Jiang Z, Benedix F, Franke S, Wartman T, Camaj P, Halangk W, Kolligs FT, Jauch KW, Nelson PJ, and Bruns CJ. 2016. miR-221 Mediates Chemoresistance of Esophageal Adenocarcinoma by Direct Targeting of DKK2 Expression. *Ann Surg* 264:804-814. 10.1097/SLA.0000000000001928

Wen J, Hu Y, Liu Q, Ling Y, Zhang S, Luo K, Xie X, Fu J, and Yang H. 2018. miR-424 coordinates multilayered regulation of cell cycle progression to promote esophageal squamous cell carcinoma cell proliferation. *EBioMedicine* 37:110-124. 10.1016/j.ebiom.2018.10.043

Xu H, Liu C, Zhang Y, Guo X, Liu Z, Luo Z, Chang Y, Liu S, Sun Z, and Wang X. 2014. Let-7b-5p regulates proliferation and apoptosis in multiple myeloma by targeting IGF1R. *Acta Biochim Biophys Sin (Shanghai)* 46:965-972. 10.1093/abbs/gmu089

Yerukala Sathipati S, and Ho SY. 2018. Identifying a miRNA signature for predicting the stage of breast cancer. *Sci Rep* 8:16138. 10.1038/s41598-018-34604-3

Yu J, Feng J, Zhi X, Tang J, Li Z, Xu Y, Yang L, Hu Z, and Xu Z. 2015. Let-7b inhibits cell proliferation, migration, and invasion through targeting Cthrc1 in gastric cancer. *Tumour Biol* 36:3221-3229. 10.1007/s13277-014-2950-5

Zhang YF, Zhang AR, Zhang BC, Rao ZG, Gao JF, Lv MH, Wu YY, Wang SM, Wang RQ, and Fang DC. 2013. MiR-26a regulates cell cycle and anoikis of human esophageal adenocarcinoma cells through Rb1-E2F1 signaling pathway. *Mol Biol Rep* 40:1711-1720. 10.1007/s11033-012-2222-7

Zhu L, Jin L, Jiang R, Wang Q, Jiang J, Mao C, and Chen D. 2013. [Correlations between miRNAs and TGF-beta1 in tumor microenvironment of esophageal squamous cell cancer]. *Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi* 29:524-528.

## Table 1(on next page)

Clinical characteristics of EAC patients

Table 1:

Clinical characteristics of EAC patients

| Variables | Case, n (%) |
|---|---|
| Sample number | 84 |
| Age | |
| <60 | 29(34.52) |
| ≥60 | 55(65.48) |
| Gender | |
| Male | 72(85.71) |
| Female | 12(14.29) |
| Height | |
| <175 | 41(48.81) |
| ≥175 | 38(45.24) |
| NA | 5(5.95) |
| Weight | |
| <85 | 46(54.76) |
| ≥85 | 37(44.05) |
| NA | 1(1.19) |
| Race | |
| Asian | 1(1.19) |
| White | 66(78.57) |
| NA | 17(20.24) |
| Event | |
| Alive | 46(54.76) |
| Dead | 38(45.24) |
| Alcohol history | |
| No | 27(32.14) |
| Yes | 56(66.67) |
| NA | 1(1.19) |
| Barrett's disease | |
| No | 52(61.91) |
| Yes | 26(30.95) |
| NA | 6(7.14) |
| Tumor size | |
| 1 | 21(25.00) |
| 2 | 14(16.67) |
| 3 | 45(53.57) |
| 4 | 1(1.19) |
| NA | 3(3.57) |
| Lymph node status | |
| 0 | 21(25.00) |

| | |
|---|---|
| 1 | 47(55.95) |
| 2 | 6(7.15) |
| 3 | 5(5.95) |
| NA | 5(5.95) |
| Metastasis | |
| 0 | 57(67.86) |
| 1 | 11(13.09) |
| NA | 16(19.05) |
| Stage | |
| I | 12(14.29) |
| II | 24(28.57) |
| III | 33(39.29) |
| IV | 11(13.09) |
| NA | 4(4.76) |

3 NA, non available.

## Table 2(on next page)

Univariate and multivariate COX regression analyses of the six-microRNA signature and clinicopathologic factors in the entire set.

Table 2:

Univariate and multivariate COX regression analyses of the six-microRNA signature and clinicopathologic factors in the entire set.

| Variables | Univariate analysis | | | Multivariate analysis | | |
|---|---|---|---|---|---|---|
| | HR | 95% CI | Pvalue | HR | 95% CI | Pvalue |
| miRNA risk score | 3.41 | 1.70-6.84 | 0.001* | 2.95 | 1.43-6.07 | 0.003* |
| Age (≥60 vs <60) | 0.89 | 0.44-1.81 | 0.752 | | | |
| Gender (male vs female) | 0.68 | 0.20-2.31 | 0.539 | | | |
| Height (≥175 vs <175cm) | 0.80 | 0.39-1.62 | 0.535 | | | |
| Weight (≥85 vs <85kg) | 1.07 | 0.53-2.15 | 0.844 | | | |
| Alcohol consumption (yes vs no) | 0.46 | 0.23-0.92 | 0.029* | 0.67 | 0.32-1.40 | 0.287 |
| Barrett's disease (yes vs no) | 1.16 | 0.56-2.37 | 0.691 | | | |
| Stage (III+IV vs I+II) | 2.30 | 1.08-4.91 | 0.031* | 1.95 | 0.88-4.29 | 0.098 |

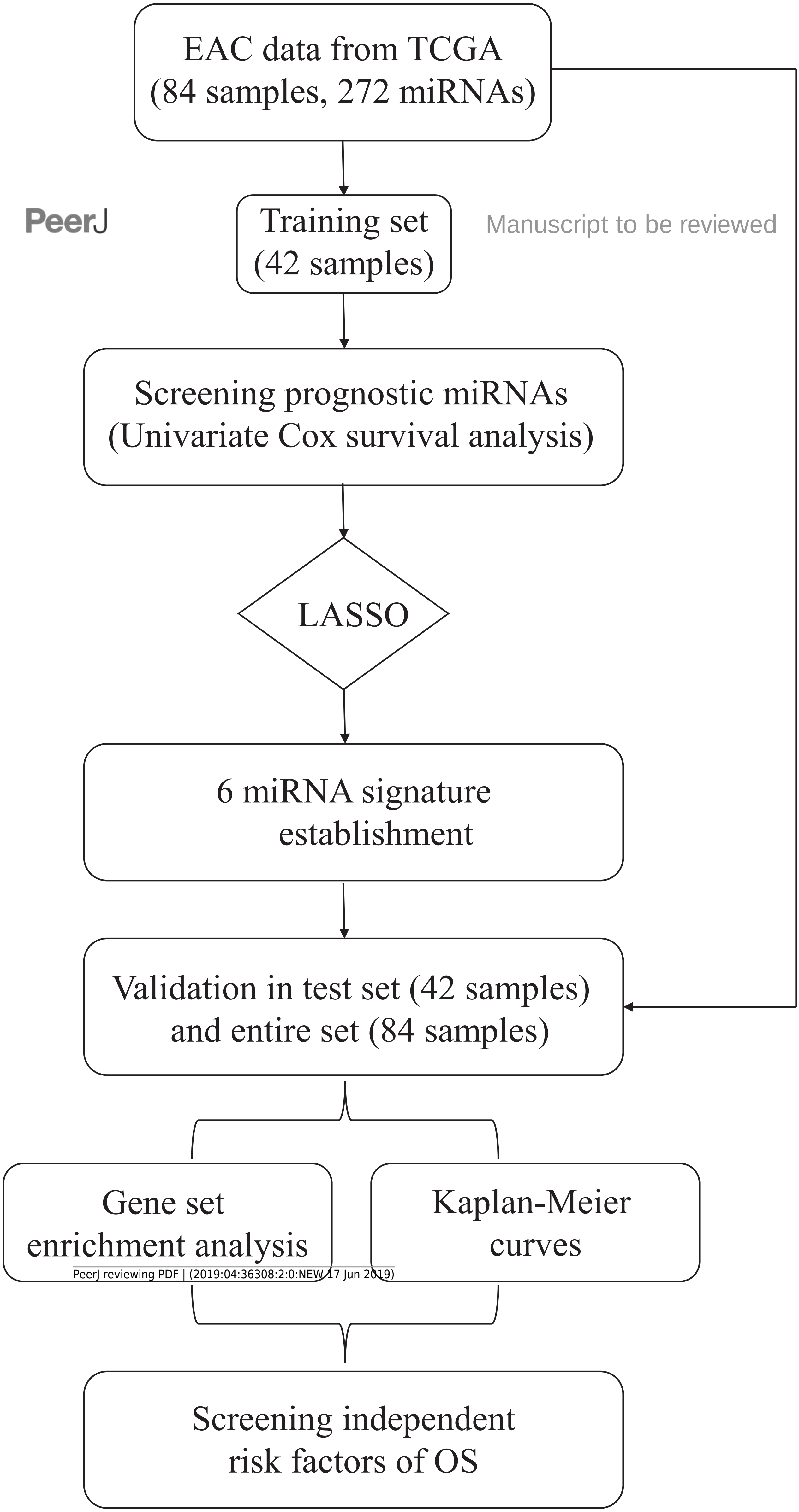
HR, Hazard Ratio.

95% CI, 95%confidence interval.

*p<0.05, statistically significant.

## Figure 1(on next page)

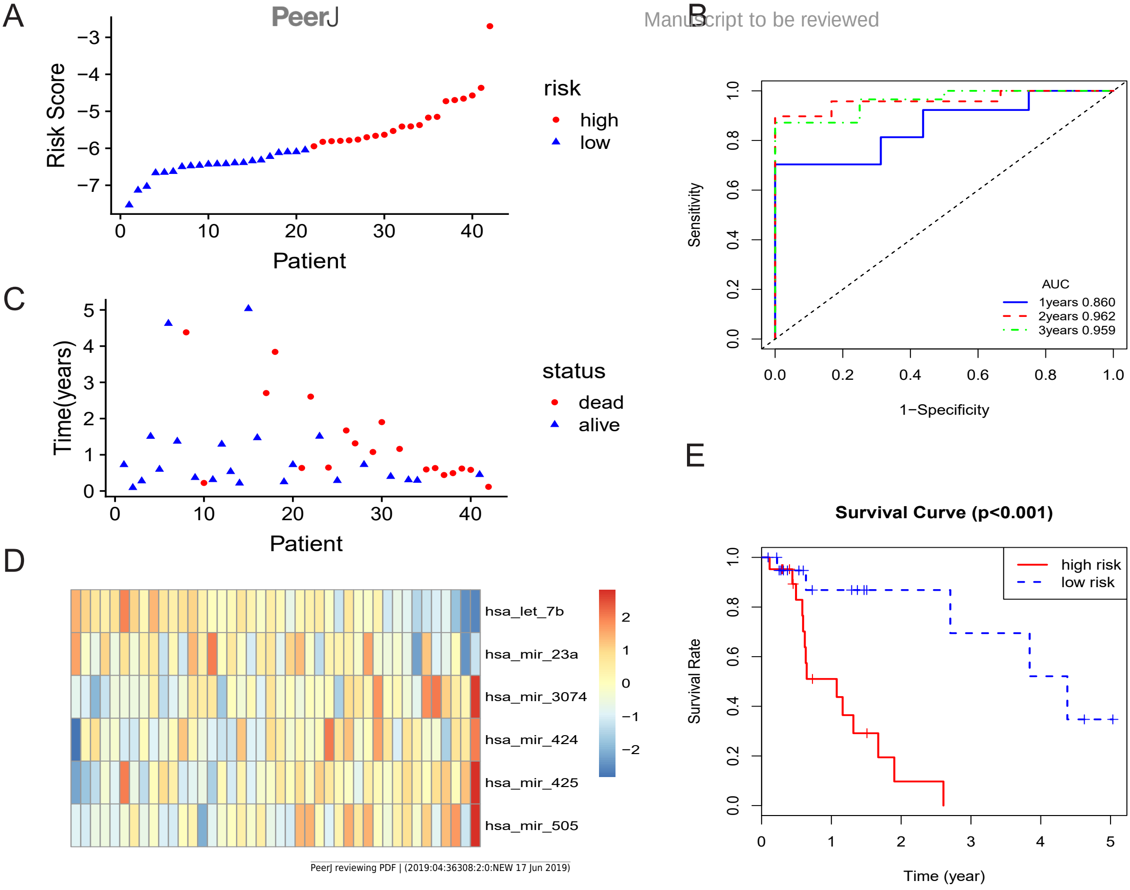
Flow chart of data preparation, processing, analysis and validation in this study


EAC data from TCGA
(84 samples, 272 miRNAs)
Training 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 training set.

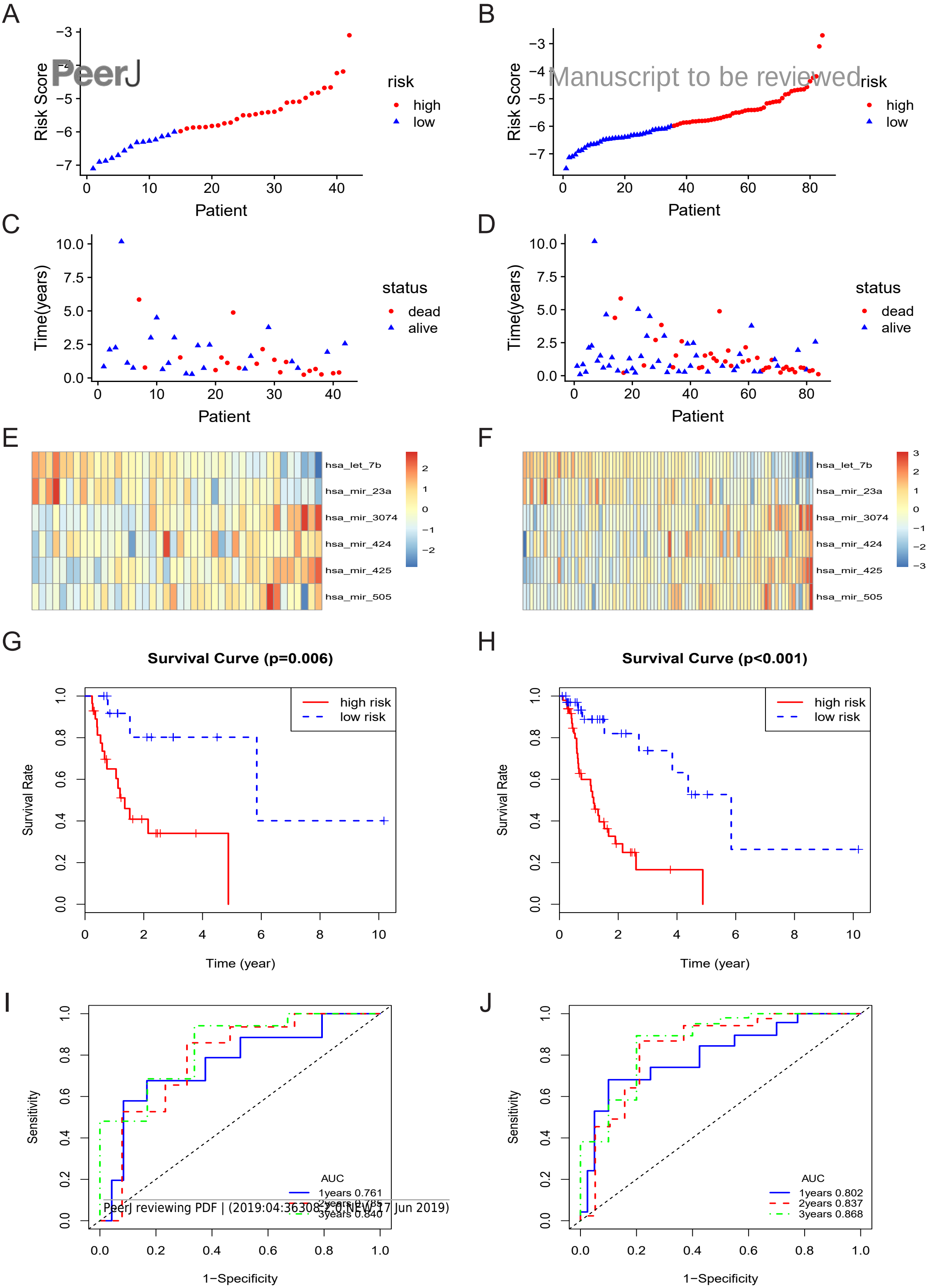
(A, C) The 6-miRNA based risk score and survival status of EAC patients. (B) Receiver operating characteristic (ROC) analyzes the sensitivity and specificity of the survival time by risk score based on the 6-miRNA signature. (D) Expression heatmap of the 6 miRNAs corresponding to each sample which ranks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


A
Risk Score
−3
−4
−5
−6
−7
0
10
20
30
40
Patient
risk
high
low
B
Sensitivity
1−Specificity
0.0
0.2
0.4
0.6
0.8
1.0
AUC
1years 0.860
2years 0.962
3years 0.959
C
Time(years)
5
4
3
2
1
0
Patient
status
dead
alive
D
hsa_let_7b
hsa_mir_23a
hsa_mir_3074
hsa_mir_424
hsa_mir_425
hsa_mir_505
2
1
0
−1
−2
E
Survival Curve (p<0.001)
Survival Rate
Time (year)
high risk
low risk
0
1
2
3
4
5

## Figure 3(on next page)

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

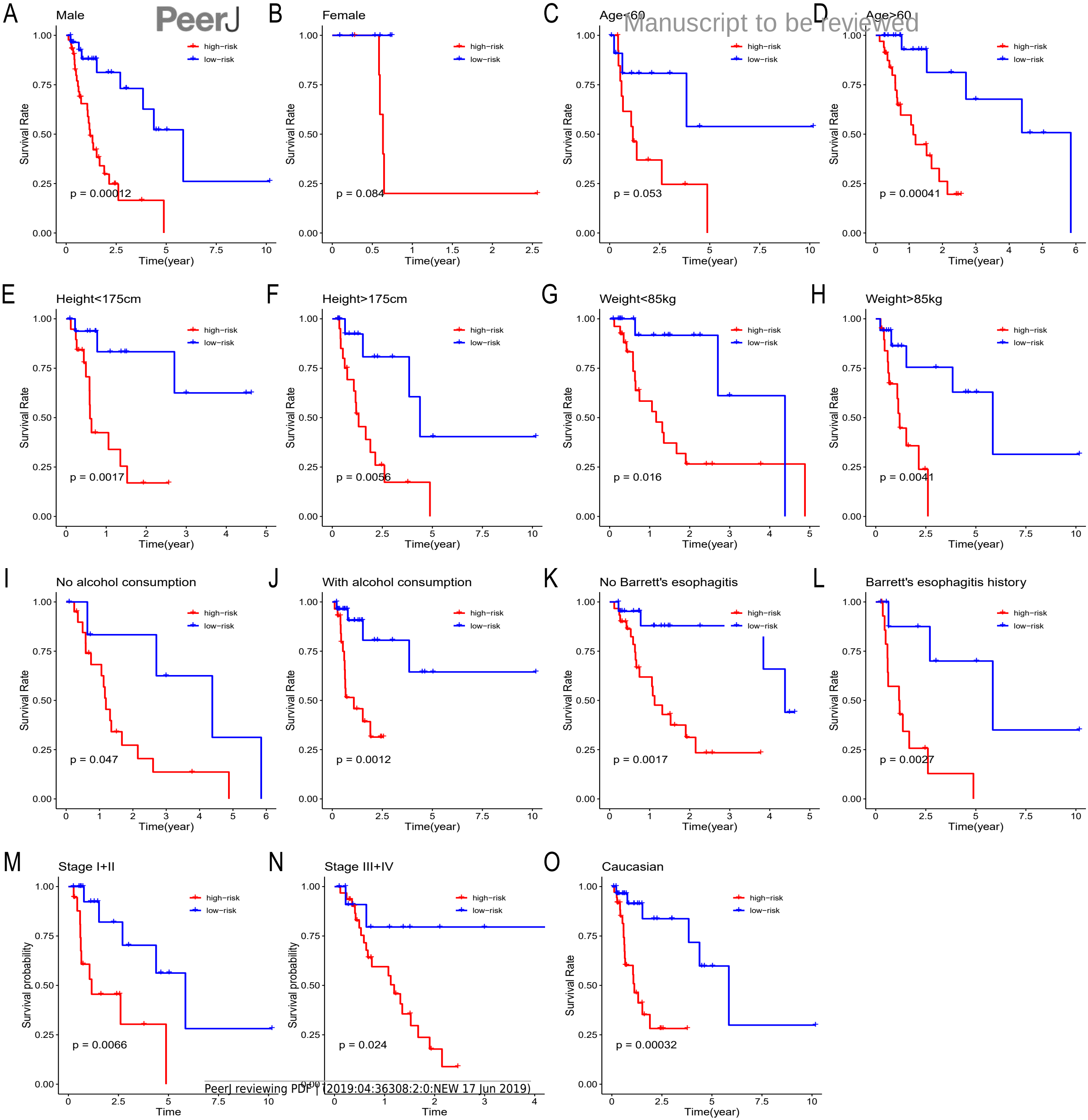
The miRNA signature risk score distribution and heatmap of the miRNA expression profiles in test set (A, C, and E) and entire set (B, D, and F). survival curves of high- and low- risk samples in test set (G) and entire set (H). Time dependent ROC curve for accuracy of the predicting risk score system in test set (I) and entire set (J) .


A
Risk Score
−3
−4
−5
−6
−7
0
10
20
30
40
Patient
risk
high
low
B
Risk Score
−3
−4
−5
−6
−7
0
20
40
60
80
Patient
risk
high
low
C
Time(years)
10.0
7.5
5.0
2.5
0.0
0
10
20
30
40
Patient
status
dead
alive
D
Time(years)
10.0
7.5
5.0
2.5
0.0
0
20
40
60
80
Patient
status
dead
alive
E
hsa_let_7b
hsa_mir_23a
hsa_mir_3074
hsa_mir_424
hsa_mir_425
hsa_mir_505
2
1
0
−1
−2
F
hsa_let_7b
hsa_mir_23a
hsa_mir_3074
hsa_mir_424
hsa_mir_425
hsa_mir_505
3
2
1
0
−1
−2
−3
G
Survival Curve (p=0.006)
Survival Rate
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10
Time (year)
high risk
low risk
H
Survival Curve (p<0.001)
Survival Rate
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10
Time (year)
high risk
low risk
I
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
1−Specificity
AUC
1years 0.761
2years 0.785
3years 0.840
J
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
1−Specificity
AUC
1years 0.802
2years 0.837
3years 0.868

# Figure 4(on next page)

Stratified analysis of overall survival in the entire set

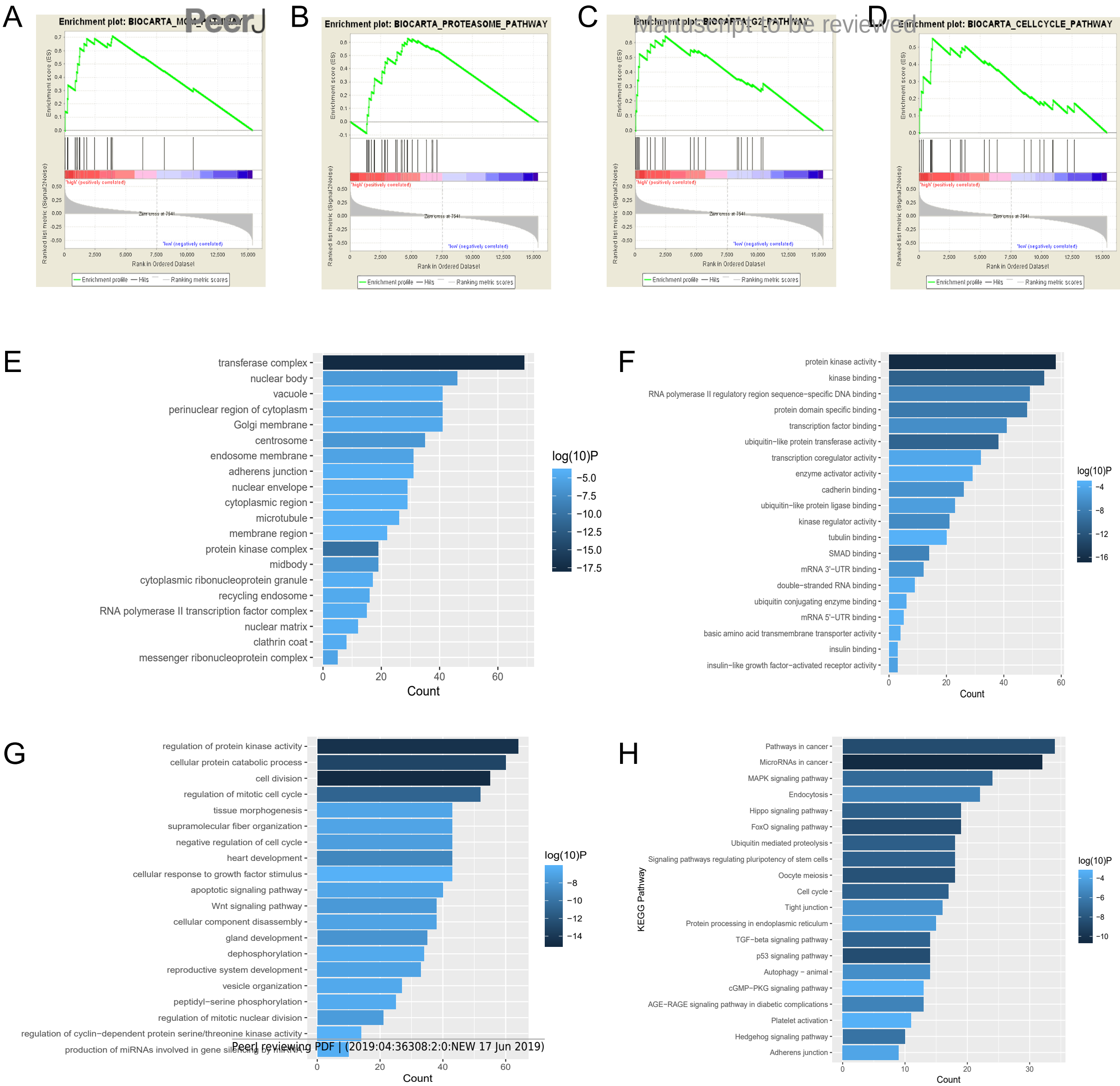
Kaplan-Meier analysis for OS in ~~subgroups~~ stratified by gender(A, B), age (C, D), height (E, F), weight (G, H), alcohol consumption (I, J), Barrett's esophagitis (K, L), TNM stage (M, N), caucasian (O).


A
Male
p = 0.00012
B
Female
p = 0.084
C
Age<60
p = 0.053
D
Age>60
p = 0.00041
E
Height<175cm
p = 0.0017
F
Height>175cm
p = 0.0056
G
Weight<85kg
p = 0.016
H
Weight>85kg
p = 0.0041
I
No alcohol consumption
p = 0.047
J
With alcohol consumption
p = 0.0012
K
No Barrett's esophagitis
p = 0.0017
L
Barrett's esophagitis history
p = 0.0027
M
Stage I+II
p = 0.0066
N
Stage III+IV
p = 0.024
O
Caucasian
p = 0.00032
high−risk
low−risk
Survival Rate
Survival probability
Time(year)
Time
0.00
0.25
0.50
0.75
1.00

## Figure 5(on next page)

Gene enrichment analysis, GO, and KEGG pathways of mRNA associated with the 6-miRNA signature.

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


A
Enrichment plot: BIOCARTA_MCM_PATHWAY
B
Enrichment plot: BIOCARTA_PROTEASOME_PATHWAY
C
Enrichment plot: BIOCARTA_G2_PATHWAY
D
Enrichment plot: BIOCARTA_CELLCYCLE_PATHWAY
Enrichment score (ES)
Ranked list metric (Signal2Noise)
'high' (positively correlated)
Zero cross at 7541
'low' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile — Hits — Ranking metric scores
E
transferase complex
nuclear body
vacuole
perinuclear region of cytoplasm
Golgi membrane
centrosome
endosome membrane
adherens junction
nuclear envelope
cytoplasmic region
microtubule
membrane region
protein kinase complex
midbody
cytoplasmic ribonucleoprotein granule
recycling endosome
RNA polymerase II transcription factor complex
nuclear matrix
clathrin coat
messenger ribonucleoprotein complex
Count
log(10)P
−5.0
−7.5
−10.0
−12.5
−15.0
−17.5
F
protein kinase activity
kinase binding
RNA polymerase II regulatory region sequence−specific DNA binding
protein domain specific binding
transcription factor binding
ubiquitin−like protein transferase activity
transcription coregulator activity
enzyme activator activity
cadherin binding
ubiquitin−like protein ligase binding
kinase regulator activity
tubulin binding
SMAD binding
mRNA 3'−UTR binding
double−stranded RNA binding
ubiquitin conjugating enzyme binding
mRNA 5'−UTR binding
basic amino acid transmembrane transporter activity
insulin binding
insulin−like growth factor−activated receptor activity
Count
log(10)P
−4
−8
−12
−16
G
regulation of protein kinase activity
cellular protein catabolic process
cell division
regulation of mitotic cell cycle
tissue morphogenesis
supramolecular fiber organization
negative regulation of cell cycle
heart development
cellular response to growth factor stimulus
apoptotic signaling pathway
Wnt signaling pathway
cellular component disassembly
gland development
dephosphorylation
reproductive system development
vesicle organization
peptidyl−serine phosphorylation
regulation of mitotic nuclear division
regulation of cyclin−dependent protein serine/threonine kinase activity
production of miRNAs involved in gene silencing by miRNA
Count
log(10)P
−8
−10
−12
−14
H
KEGG Pathway
Pathways in cancer
MicroRNAs in cancer
MAPK signaling pathway
Endocytosis
Hippo signaling pathway
FoxO signaling pathway
Ubiquitin mediated proteolysis
Signaling pathways regulating pluripotency of stem cells
Oocyte meiosis
Cell cycle
Tight junction
Protein processing in endoplasmic reticulum
TGF−beta signaling pathway
p53 signaling pathway
Autophagy − animal
cGMP−PKG signaling pathway
AGE−RAGE signaling pathway in diabetic complications
Platelet activation
Hedgehog signaling pathway
Adherens junction
Count
log(10)P
−4
−6
−8
−10