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. 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 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. 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 ~~A~~rea 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, ~~T~~here 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 ~~studys~~ 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 algorism capable of variable selection and regularization simultaneously (Gao et al. 2010). LASSO was performed to reduce aboved 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 = $\beta1 \times$ gene 1 + $\beta2 \times$gene 2 + … + $\beta n \times$ gene n, where β 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 “gglot2” 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 did 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 clinicalpathological factors (HR=2.95, CI 1.43-6.07, p= 0.00338, Table2) . 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 "cellcycle 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 cacer (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 paly 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 of 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, ~~D~~ue 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

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

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

Global Burden of Disease Cancer C, 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

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

Table1. Clinical characteristics of EAC patients

Variables	Case, n (%)
Sample number	84
Age	
＜60	29 (34.5)
≥60	55 (65.5)
Gender	
male	72 (85.7)
femal	12 (14.3)
Height	
＜175	46 (54.8)
≥175	38 (45.2)
NA	4 (4.8)
Weight	
＜85	46 (54.8)
≥85	37 (44.0)
NA	1 (1.2)
Race	
asian	1 (1.2)
white	66 (78.6)
NA	17 (20.2)
Event	
alive	46 (54.8)
dead	38 (45.2)
alcohol history	
no	27 (32.1)
yes	56 (66.7)
NA	1 (1.2)
Barrett's disease	
no	52 (61.9)
yes	26 (31.0)
NA	6 (7.1)
Tumor size	
1	21 (25.0)
2	14 (16.7)
3	45 (53.6)
4	1 (1.2)
TX	3 (3.6)
Lymph node status	
0	21 (25.0)
1	47 (56.0)

2	6 (7.1)
3	5 (6.0)
NX	5 (6.0)
metastsis (%)	
0	57 (67.9)
1	11 (13.1)
MX	12 (14.3)
NA	4 (4.8)
stage (%)	
I	12 (14.3)
II	24 (28.6)
III	33 (39.3)
IV	11 (13.1)
NA	4 (4.8)

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

variables	**Univariate analysis**			**Multivariate analysis**		
	HR	**95%CI**	**Pvalue**	**HR**	**95%CI**	**Pvalue**
Riskscore	3.41	1.70-6.84	0.00053	2.95	1.43-6.07	0.00338
Age (≥60 vs <60)	0.89	0.44-1.81	0.75169			
Gender (male vs female)	0.68	0.20-2.31	0.53893			
Height (≥175 vs <175cm)	0.80	0.39-1.62	0.53522			
Weight (≥85 vs <85kg)	1.07	0.53-2.15	0.84364			
Alcohol consumption (yes vs no)	0.46	0.23-0.92	0.02859	0.67	0.32-1.40	0.28696
Barrett's disease (yes vs no)	1.16	0.56-2.37	0.69146			
Stage (III+IV vs I+II)	2.30	1.08-4.91	0.03086	1.95	0.88-4.29	0.09782

HR, Hazard Ratio.

Figure 1 (on next page)

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

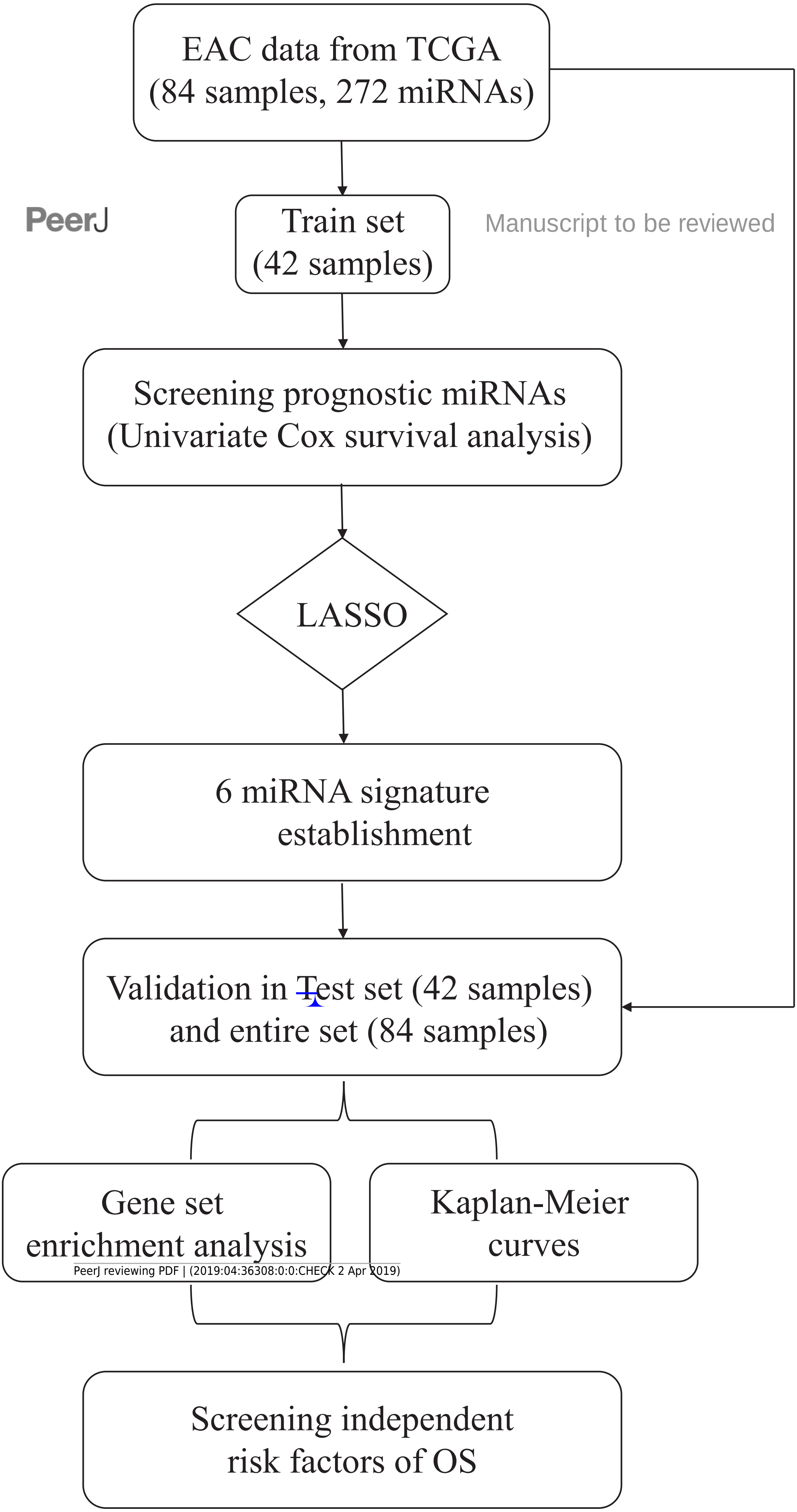

EAC data from TCGA
(84 samples, 272 miRNAs)
Train set
(42 samples)
Screening prognostic miRNAs
(Univariate Cox survival analysis)
LASSO
6 miRNA signature
establishment
Validation in Test set (42 samples)
and entire set (84 samples)
Gene set
enrichment analysis
Kaplan-Meier
curves
Screening independent
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.

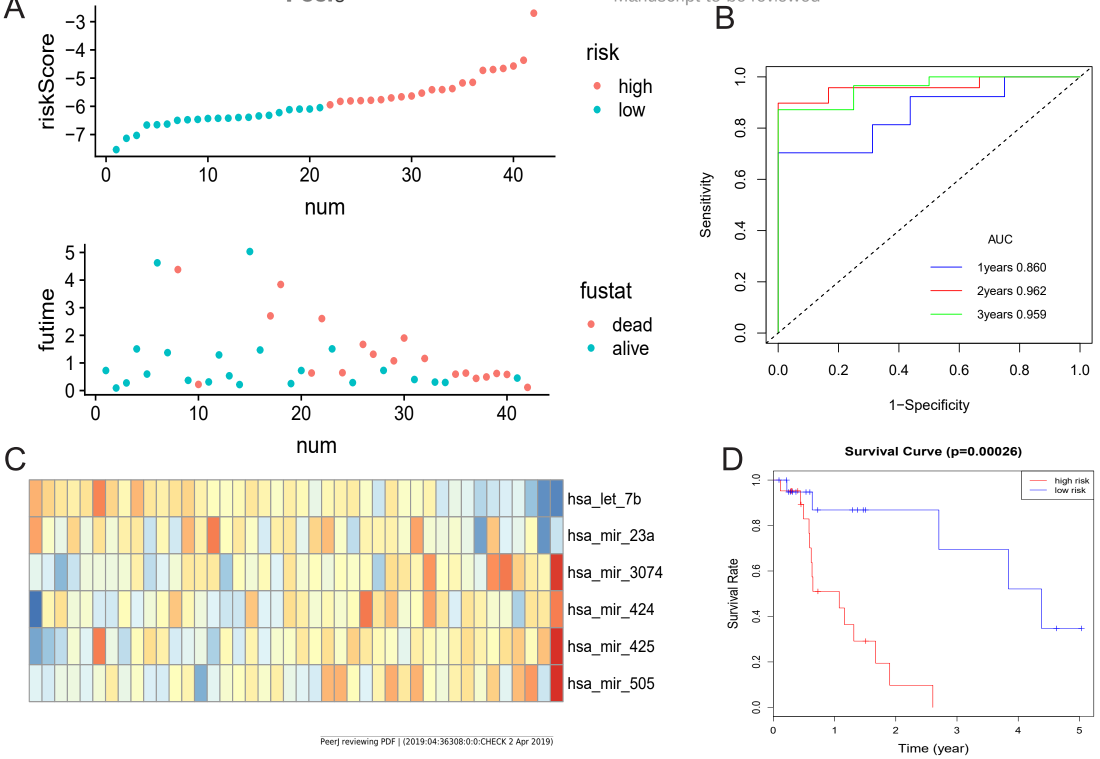

A
riskScore
−3
−4
−5
−6
−7
0 10 20 30 40
num
risk
high
low
futime
5 4 3 2 1 0
0 10 20 30 40
num
fustat
dead
alive
B
Sensitivity
1.0 0.8 0.6 0.4 0.2 0.0
0.0 0.2 0.4 0.6 0.8 1.0
1−Specificity
AUC
1years 0.860
2years 0.962
3years 0.959
C
hsa_let_7b
hsa_mir_23a
hsa_mir_3074
hsa_mir_424
hsa_mir_425
hsa_mir_505
D
Survival Curve (p=0.00026)
high risk
low risk
Survival Rate
1.0 0.8 0.6 0.4 0.2 0.0
0 1 2 3 4 5
Time (year)

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) .

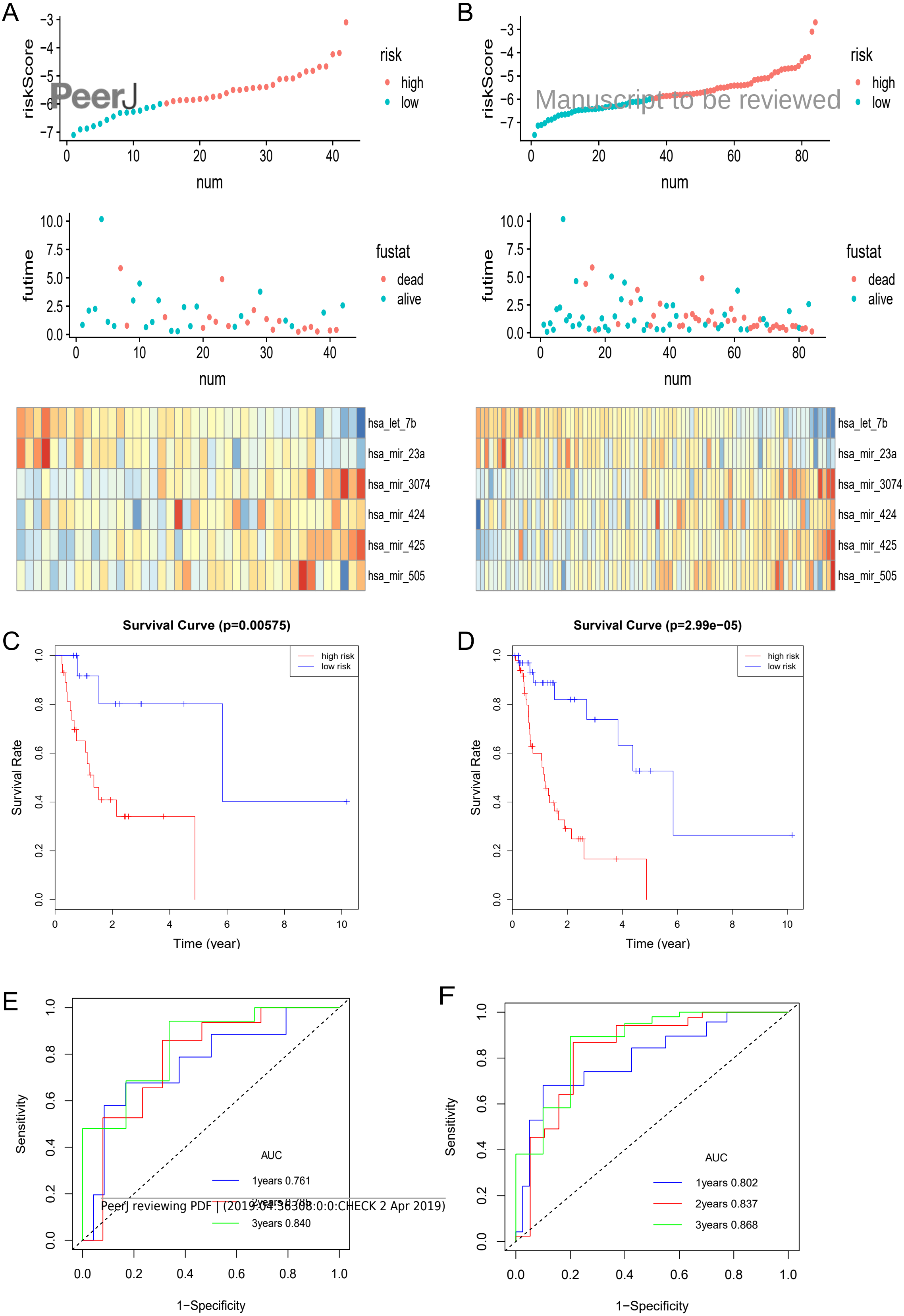

A
riskScore
num
risk
high
low
futime
fustat
dead
alive
hsa_let_7b
hsa_mir_23a
hsa_mir_3074
hsa_mir_424
hsa_mir_425
hsa_mir_505
B
C
Survival Curve (p=0.00575)
Survival Rate
Time (year)
high risk
low risk
D
Survival Curve (p=2.99e−05)
E
Sensitivity
1−Specificity
AUC
1years 0.761
3years 0.840
F
AUC
1years 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 age (A), gender (B), height (C), weight (D), Alcohol consumption (E), Barrett's esophagitis (F), Caucasian (G).

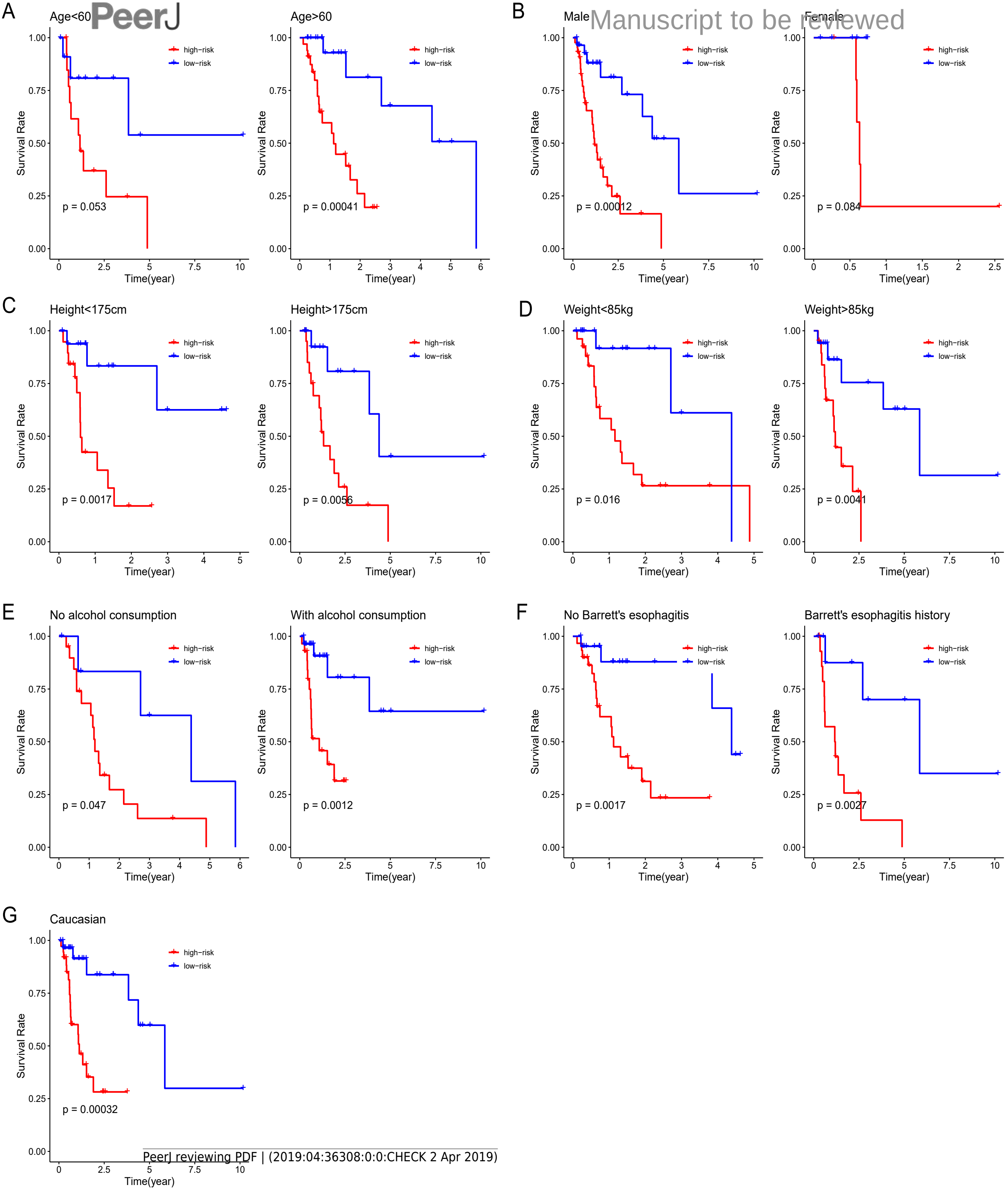
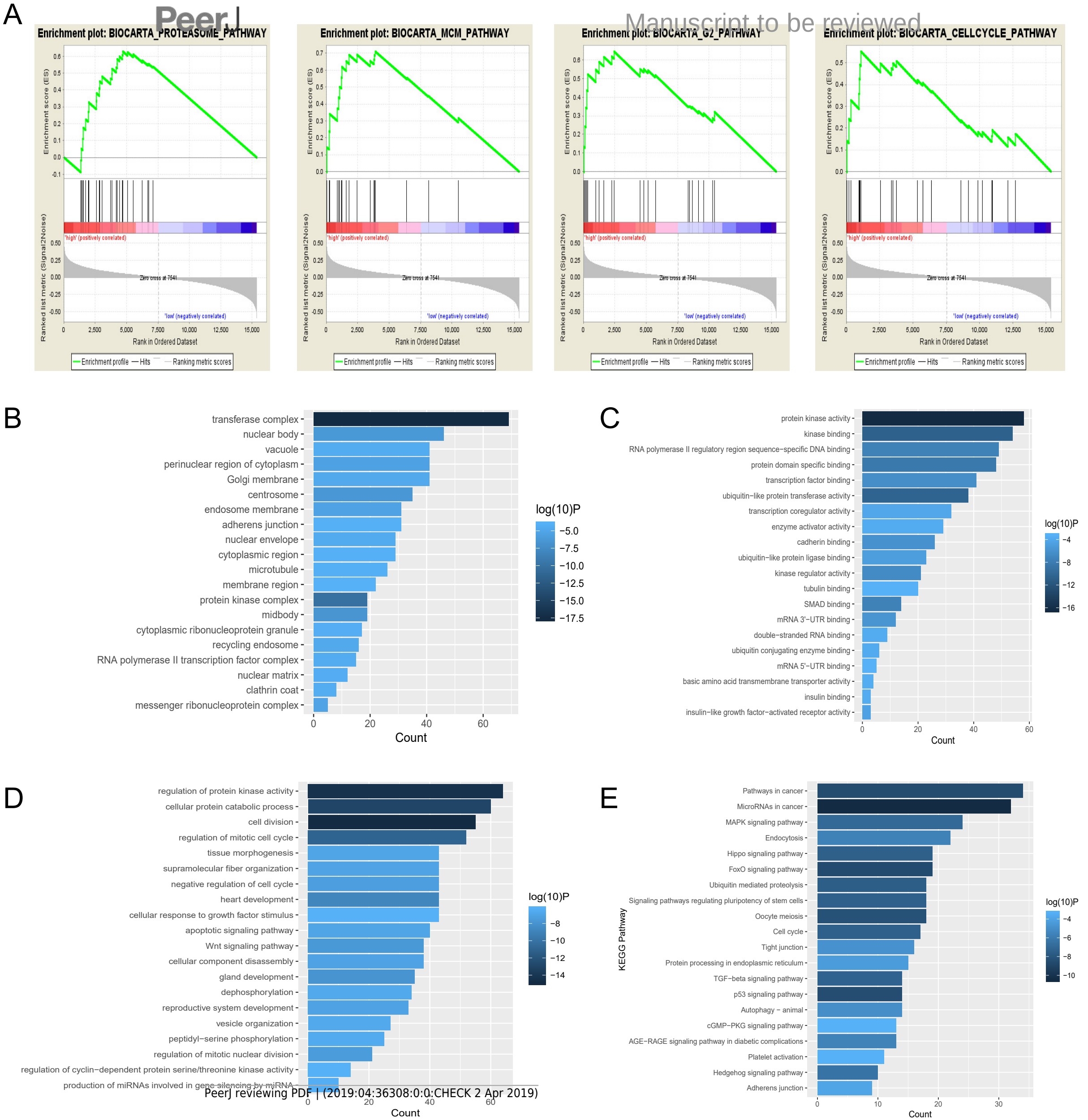

A
Age<60
Age>60
B
Male
Female
C
Height<175cm
Height>175cm
D
Weight<85kg
Weight>85kg
E
No alcohol consumption
With alcohol consumption
F
No Barrett's esophagitis
Barrett's esophagitis history
G
Caucasian
high-risk
low-risk
Survival Rate
Time(year)
p = 0.053
p = 0.00041
p = 0.00012
p = 0.084
p = 0.0017
p = 0.0056
p = 0.016
p = 0.0041
p = 0.047
p = 0.0012
p = 0.0017
p = 0.0027
p = 0.00032

Figure 5(on next page)

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.


A
Enrichment plot: BIOCARTA_PROTEASOME_PATHWAY
Enrichment plot: BIOCARTA_MCM_PATHWAY
Enrichment plot: BIOCARTA_G2_PATHWAY
Enrichment plot: BIOCARTA_CELLCYCLE_PATHWAY
Enrichment score (ES)
Ranked list metric (Signal2Noise)
'high' (positively correlated)
'low' (negatively correlated)
Zero cross at 7641
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores
B
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
C
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
D
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
E
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