Observational Study

Medicine®

OPEN

A five-gene signature for predicting overall survival of esophagus adenocarcinoma

Tian Lan, MD[a,b], Weiguo Liu, MS[c], Yunyan Lu, MS[d], Hua Luo, MS[a,*]

Abstract

Esophageal adenocarcinoma (EAC) is common and aggressive with increasing trend of incidence. Urgent need for an effective signature to assess EAC prognosis and facilitate tailored treatment is required.

Differentially expressed mRNAs (DEMs) were identified by analyzing EAC tissues and adjacent normal samples from The Cancer Genome Atlas (TCGA). Then univariate regression analyses were performed to confirm prognostic DEMs. We used least absolute shrinkage and selection operator (LASSO) to build a prognostic mRNA signature whose performance was assessed by Kaplan–Meier curve, receiver operating characteristic (ROC). GSE72874 were used as an external test set. The performances of the signature were also validated in internal TCGA and external test sets. Gene set enrichment analysis (GSEA) and tumor immunity analysis were performed to decipher the biological mechanisms of the signature.

A 5-mRNA signature consisted of SLC26A9, SINHCAF, MICB, KRT19, and MT1X was developed to predict prognosis of EAC. The 5-mRNA signature was promising as a biomarker for predicting 3-year survival rate of EAC in the internal test set, the entire TCGA set, and the external test set with areas under the curve (AUC) = 0.849, 0.924, and 0.747, respectively. Patients were divided into low- and high-risk groups based on risk scores of the signature. The high-risk group was mainly associated with cancer-related pathways and low levels of B cell infiltration.

The 5-mRNA prognostic signature we identified can reliably predict prognosis and facilitate individualized treatment decisions for EAC patients.

Abbreviations: AUC = areas under the curve, DEMs = differentially expressed mRNAs, EAC = esophageal adenocarcinoma, ESCA = esophagus cancer, ESCC = esophageal squamous cell carcinoma, GO = gene ontology, GSEA = gene set enrichment analysis, KEGG = Kyoto Encyclopedia of Genes and Genomes, LASSO = least absolute shrinkage and selection operator, ROC = receiver operating characteristic, TCGA = The Cancer Genome Atlas, TIMER = the tumor immune estimation resource, TNM = tumor–node–metastasis, TPM = transcripts per kilobase million.

Keywords: esophageal neoplasms, survival analysis, transcriptome

1. Introduction

Esophagus cancer (ESCA) is the 6th most lethal cancer and the 8th most prevalent malignancy globally.[1] The 2 major subtypes of ESCA include esophageal adenocarcinoma (EAC) and esophageal squamous cell carcinoma (ESCC). In recent decades, ESCC has decreased in prevalence whilst EAC has gradually increased.[2] EAC is a particularly aggressive form of ESCA with overall 5-year survival rates as low as ~18%.[3] The lack of early stage diagnostics coupled to high rates of metastasis and drug resistance contribute to the poor outcomes of EAC.[2,4] Despite the improvements in EAC therapeutics, their benefits must be balanced with their side effects. Precision therapy based on the molecular basis for EAC development offers the most hope for effective therapeutic interventions. Although tumor–node–metastasis (TNM) staging has been applied for prognostic prediction and individual treatment,[5] the patients with the same

Editor: Parth Mehta.

TL and WL contributed equally to this work.

Funding/Grant: The Natural Science Foundation of Zhejiang Province of China (LGD20H190002), The Science and Technology Project of Hangzhou City Health Bureau (2018A65), and The Science and Technology Project of The Xiaoshan Science and Technology Bureau (2018204).

The authors have no conflicts of interest to disclose.

Supplemental Digital Content is available for this article.

The datasets generated during and/or analyzed during the current study are publicly available.

[a] Department of Breast Surgery, Hangzhou TCM Hospital Affiliated to Zhejiang Chinese Medicine University, Hangzhou Hospital of Traditional Chinese Medicine, [b] The Second Clinical Medical College, Zhejiang Chinese Medical University, Hangzhou, [c] Department of Oncology, The People's Hospital of Jiangshan, Quzhou, [d] Department of Cardiology, The First People's Hospital of Xiaoshan District, Hangzhou, Zhejiang, People's Republic of China.

[*] Correspondence: Hua Luo, Department of Breast Surgery, Hangzhou TCM Hospital Affiliated to Zhejiang Chinese Medicine University, Hangzhou Hospital of Traditional Chinese Medicine, Hangzhou, Zhejiang, People's Republic of China (e-mail: luohua0313@163.com).

How to cite this article: Lan T, Liu W, Lu Y, Luo H. A five-gene signature for predicting overall survival of esophagus adenocarcinoma. Medicine 2021;100:14(e25305).

Received: 25 September 2020 / Received in final form: 24 February 2021 / Accepted: 3 March 2021

http://dx.doi.org/10.1097/MD.0000000000025305

stage can have significantly different outcomes in clinical practice. As such, urgent need for effective models to assess EAC prognosis and facilitate tailored treatment are required.

In recent years, with development of microarray and high-throughput sequencing, studies have focused on mRNAs for the prediction of EAC prognosis. Kim et al identified a 2-gene signature (SPARC and SPP1) associated with prognosis in EAC.[6] In addition, dysregulated DCK, PAPSS2, SIRT2, and TRIM44 have been linked to overall survival. As such, a 4-gene signature based on these genes was independently developed for EAC prognostics.[7] By evaluating the gene expression profiles in 64 patients with EAC, Pennathur et al constructed an internally cross-validated 59-mRNA prognostic signature.[8] There were however limitations in these studies, including a lack of the cross-platform validation datasets, small sample sizes, and the absence of external testing.

The Cancer Genome Atlas (TCGA) is a cancer genome project that provides large-scale genomic and clinical information spanning 33 cancer types. The Gene Expression Omnibus archives Array- and sequence-based data.[9] Here, we explored the available EAC data from the TCGA and GEO, and developed a 5-mRNA signature for EAC. This signature was validated in independent external test set, highlighting its promise for EAC prognostics.

2. Material and methods

2.1. Data sources and processing

RNAseq data and clinical data were downloaded from TCGA database (https://portal.gdc.cancer.gov/). Gene annotations were obtained from GENCODE datasets (www.gencodegenes.org). We converted the number of fragments per kilobase of non-overlapped exons per million fragments mapped into the value of transcripts per kilobase million (TPM) to mimic those of the microarrays.[10] Only mRNAs with TPM values >0.1 in at least 50% of the ESCA samples were enrolled for analysis. Expression was defined as log2 $(x+1)$, whilst x represented the TPM value. To predict the prognosis of EAC, we excluded samples with a <30-day censor time. We obtained 86 samples from the TCGA including 75 EAC cases and 11 normal tissue samples. Differentially expressed mRNAs (DEMs) were identified using the R package "limma" in EAC versus normal tissue. Selection criteria for significant DEMs were $P<.05$ and |log2 fold change| >1.5.

We downloaded GSE72874 from the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) database as an external validation set using R package "GEOquery." Corresponding clinical data were achieved from Supplemental data submitted by Krause et al.[11] Since the data came from the TCGA database and the Gene Expression Omnibus database, no ethical approval was required.

The tumor immune estimation resource (TIMER) algorithm (https://cistrome.shinyapps.io/timer/) was used to measure the abundance of B cells, CD4+ T cells, CD8+ T cells, neutrophils, macrophages, and dendritic cells in EAC.[12] This provided details on the immune infiltration in EAC and its relevance to the identified gene panel signature.

2.2. Construction and validation of the mRNA signature

To further screen DEM correlating with prognosis, univariate COX proportional hazards regression analysis was performed. Only prognostic DEMs with $P<.05$ were considered as statistically significant mRNAs.

We divided the TCGA datasets into training (n = 37) and tests (n = 38). Least absolute shrinkage and selection operator (LASSO) regression was selected as the optimal dimension reduction technique,[13] previous reported studies have applied it to filter significant genes.[14] LASSO was performed to screen and confirm the selected DEMs in the training set. Risk scores for each patient were calculated with the linear combinational of the selected gene expression weighted by their regression coefficients. The formula was set as follow:

$$\text{Risk score} = \beta 1 \times \text{gene } 1 + \beta 2 \times \text{gene } 2 + \ \ldots \ + \beta n \times \text{gene n},$$

where β is the coefficient of each gene, and gene indicates relative expression of gene.

Datasets were divided into low- and high-risk groups using the median score. Survival curves between these groups were assessed via the Kaplan–Meier method, whilst *P*-values were estimated using log-rank tests. Receiver operating characteristic (ROC) and areas under the curve (AUC) were calculated to determine the accuracy of the mRNA signature by using the R "timeROC" package. Further, validations in the internal and external tests were performed using the risk score formula.

2.3. Biological functions of the mRNA signature

Gene set enrichment analysis (GSEA) was performed using the R package "clusterProfiler"[15,16] to study the biological mechanisms between low- and high-risk EAC groups. Kyoto Encyclopedia of Genes and Genomes (KEGG, c2.cp.kegg.v7.0.entrez.gmt) and Reactome (c2.cp.reactome.v7.0.entrez.gmt) datasets were selected from the molecular signature database. Gene Ontology (GO)[17] of the Biological Process (GO_BP), Cellular Components (GO_CC), and Molecular Functions (GO_MF) was performed to examine the biological functions of the signature. Cut-off values for significant enriched terms were $P<.05$.

2.4. Statistical analyses

The R statistical package (R version 3.5.2; http://www.Rproject.org) was used to perform all statistical analyses and for plot construction. We used "survival" package of R to carry out univariate and multivariate Cox proportional hazard assessments. Hazard ratios (HR), 95% confidence intervals, and *P*-values were calculated. A Student *t* test was performed to compare continuous variables. A 1-way analysis of variance was used for multiple group comparisons. The association between tumor immunity infiltration and gene expression level was estimated using the Spearman correlation test. *P*-values $<.05$ were considered significant.

3. Results

3.1. Study datasets

A flowchart of the study is shown in Figure 1. We obtained data from 75 patients with EAC from the TCGA dataset, divided into training (n = 37) and internal test sets (n = 38). In total, 43 patients with EAC from the GSE72874 dataset were enrolled for further analysis. Basic characteristics in these 2 datasets are shown in Table 1.

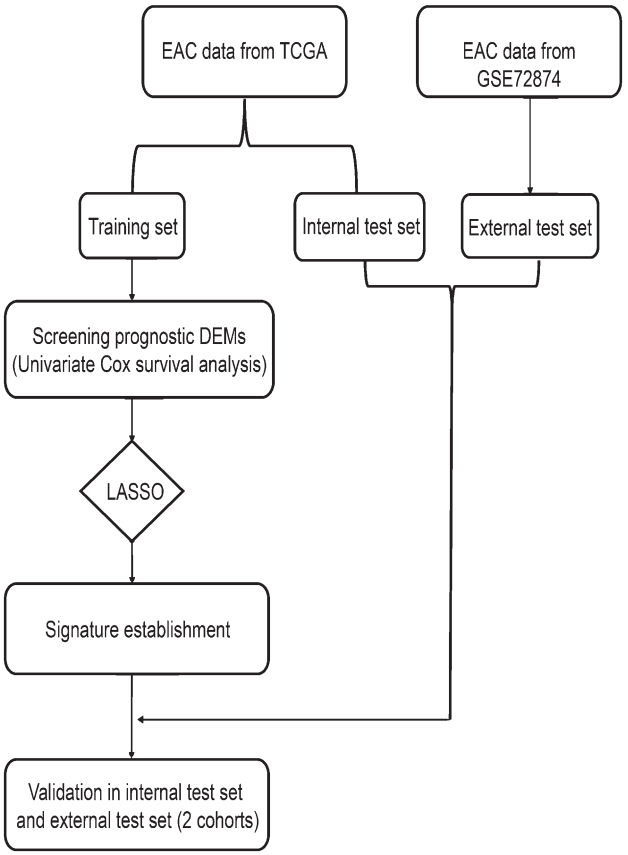

Figure 1. A flow diagram of establishment and validation of the prognostic mRNA signature. DEMs = differentially expressed mRNAs, EAC = esophageal adenocarcinoma, LASSO = least absolute shrinkage and selection operator, TCGA = The Cancer Genome Atlas.

3.2. DEMs associated with EAC prognosis

Through the analysis of EAC and normal tissues in the TCGA dataset, we extracted 852 upregulated and 160 downregulated genes (Supplemental Fig. S1, http://links.lww.com/MD2/A24). By subjecting DEMs to univariable COX regression analysis, 138 differentially expressed genes (DEGs) related to overall survival were identified for the prognostic signature (Supplemental Table S1, http://links.lww.com/MD2/A30).

3.3. Construction of the mRNA signature and evaluation of its prognostic ability in the training set

Lambda value was set using the lambda.min. Five mRNAs with non-zero coefficients were identified (Supplemental Fig. S2, http://links.lww.com/MD2/A25). We created a risk score formula as follows:

Risk score =
(−0.0755 × relative expression of SLC26A9) +
(0.6688 × relative expression of SINHCAF) +
(0.3071 × relative expression of MICB) +
(−0.5899 × relative expression of KRT19) +
(−0.4242 × relative expression of MT1X)

In the training set, we classified patients into high (n = 19) and low risk groups (n = 19) using a median risk score of −1.60 as the cut-off. Details on the survival status according to risk scores are shown in Figure 2A. Upon analysis of the heatmaps of the mRNAs (Fig. 2D), 2 mRNAs had positive coefficients including SINHCAF and MICB which indicated that the higher expression levels of these mRNAs were linked with poor survival. The other 3 mRNAs with negative coefficients were SLC26A9, KRT19, and MT1X which demonstrated that it was a positive correlation between their expression levels and clinical outcome. Survival curves indicated that patients in the lower-risk group had higher

Table 1

Clinical characteristics of patients with EAC in TCGA and GSE72874 datasets.

	TCGA Training dataset	Testing dataset	GSE72874 External validation dataset
Sample number	37	38	43
Survival time (yr [SD])	1.29 (1.24%)	1.57 (1.44%)	1.71 (1.36%)
Status			
Alive	23 (62.2%)	19 (50.0%)	26 (60.5%)
Dead	14 (37.8%)	19 (50.0%)	17 (39.5%)
Age			
<60	10 (27.0%)	18 (47.4%)	11 (25.6%)
>60	27 (73.0%)	20 (52.6%)	32 (74.4%)
Sex			
Female	5 (13.5%)	6 (15.8%)	3 (7.0%)
Male	32 (86.5%)	32 (84.2%)	40 (93.0%)
Height (cm)			
<175	18 (48.6%)	17 (44.7%)	
>175	16 (43.2%)	19 (50.0%)	
NA	3 (8.1%)	2 (5.3%)	
Weight (kg)			
<85	19 (51.4%)	24 (63.2%)	40 (93.0%)
>85	17 (45.9%)	14 (36.8%)	3 (7.0%)
NA	1 (2.7%)	0 (0.0%)	0 (0.0%)
Race			
Non-white	0 (0.0%)	1 (2.6%)	
White	28 (75.7%)	30 (78.9%)	
NA	9 (24.3%)	7 (18.4%)	
Alcohol history			
No	14 (37.8%)	12 (31.6%)	
Yes	22 (59.5%)	26 (68.4%)	
NA	1 (2.7%)	0 (0.0%)	
Barrett disease			
No	21 (56.8%)	25 (65.8%)	
Yes	14 (37.8%)	10 (26.3%)	
NA	2 (5.4%)	3 (7.9%)	
Tumor size			
I+II	14 (37.8%)	15 (39.5%)	
III+IV	22 (59.5%)	22 (57.9%)	
NA	1 (2.7%)	1 (2.6%)	
Node status			
Negative	11 (29.7%)	8 (21.1%)	
Positive	24 (64.9%)	28 (73.7%)	
NA	2 (5.4%)	2 (5.3%)	
Metastasis			
0	29 (78.4%)	26 (68.4%)	
1	2 (5.4%)	8 (21.1%)	
NA	6 (16.2%)	4 (10.5%)	
Stage			
I+II	16 (43.2%)	16 (42.1%)	
III+IV	21 (56.8%)	20 (52.6%)	
NA	0 (0.0%)	2 (5.3%)	

EAC = esophageal adenocarcinoma, NA = not available, SD = standard deviation, TCGA = The Cancer Genome Atlas.

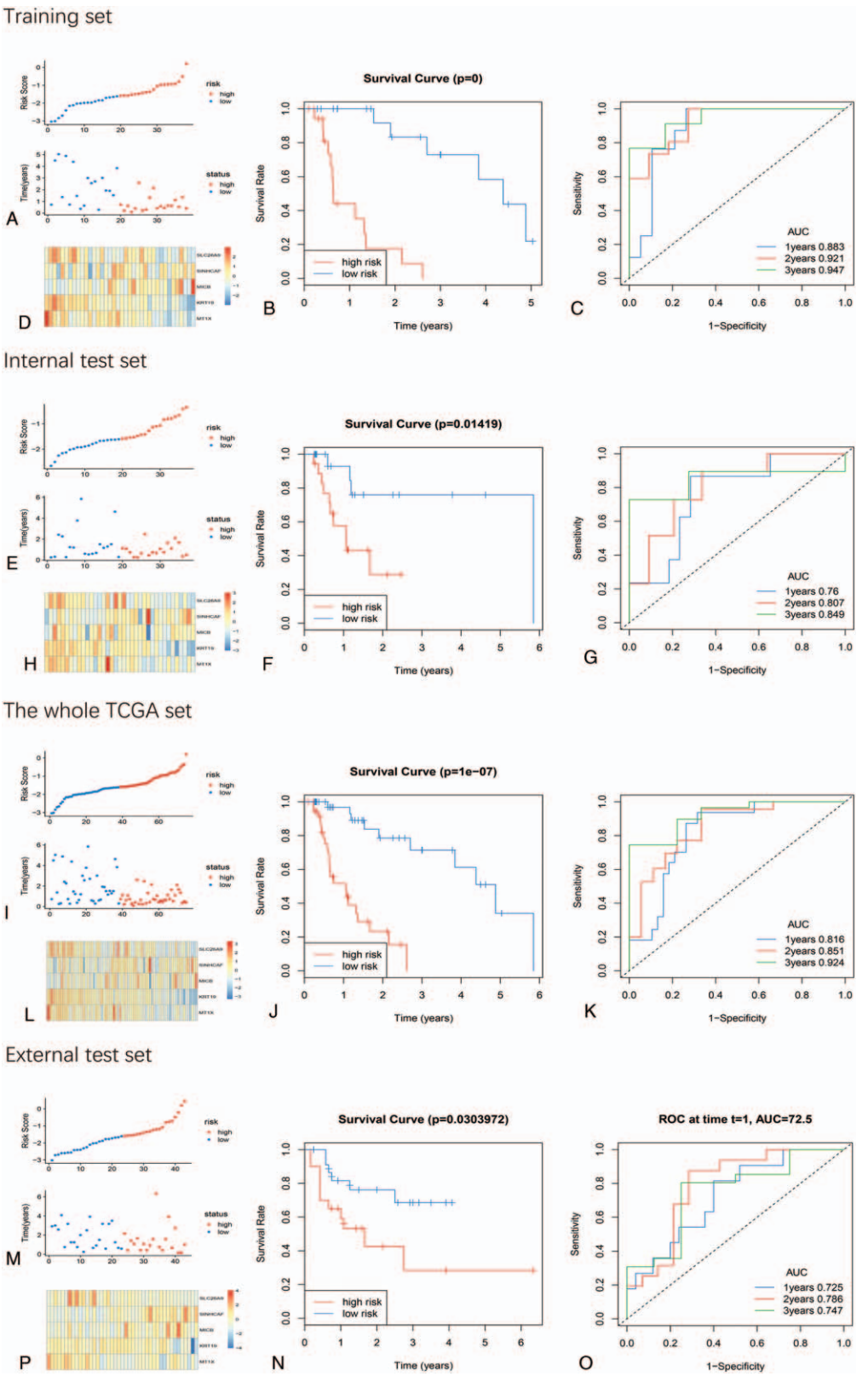

Figure 2. Risk score based on 5-mRNA signature significantly associates with prognosis. Distribution of each patient risk score (A, E, I, M) and individual mRNA expression profiles (D, H, L, P); the Kaplan–Meier survival analysis between high- and low-risk groups (B, F, J, N); time dependent ROC curves at 1, 2, and 3 years (C, G, K, O) in training set, internal test set, the entire TCGA set, and external test set. ROC = receiver operating characteristic, TCGA = The Cancer Genome Atlas.

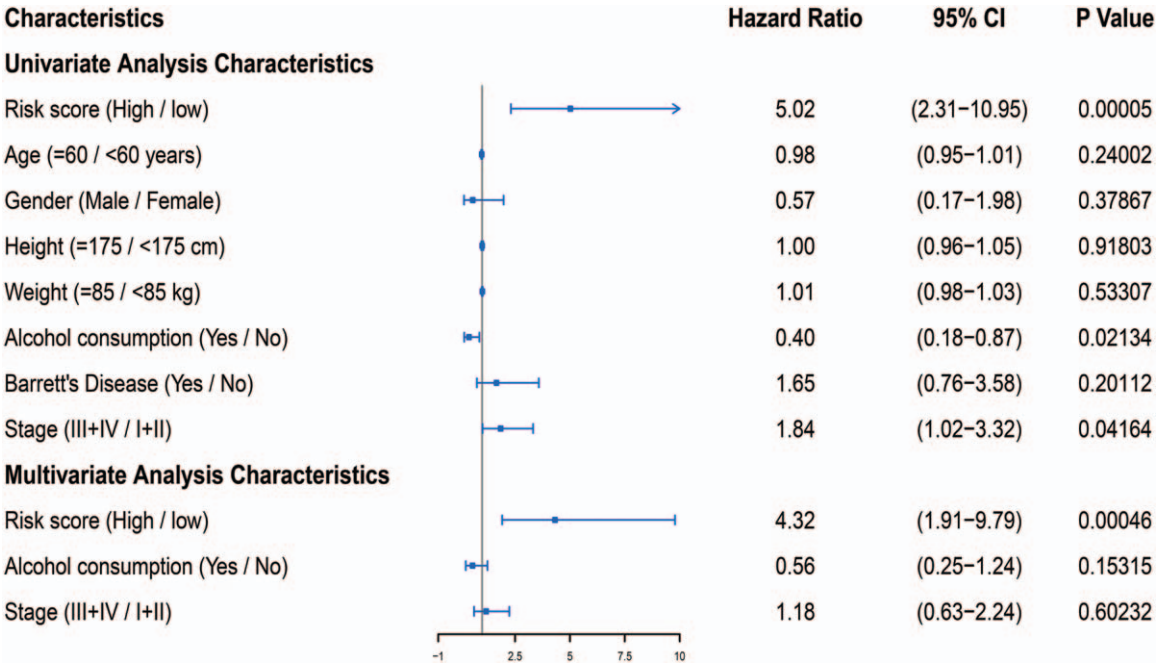

Figure 3. Forest plot summary of univariate and multivariate Cox regression analysis of overall survival in the whole TCGA set. CI = confidence interval, TCGA = The Cancer Genome Atlas.

survival rates (Fig. 2B). Time-dependent ROC analyses demonstrated that AUCs of the mRNA-based signature were 0.760, 0.807, and 0.849 at 1-, 2-, and 3-year survival times, respectively (Fig. 2C).

3.4. Validation of the signature

To further assess the prognostic power of the identified mRNA signature, we determined its prognostic ability in internal TCGA and external datasets. Risk scores were evaluated using the training set formula. The distribution of the 5-mRNA risk score, patient survival, and mRNA expression were obtained in the internal test set (Fig. 2E and H), the entire TCGA set (Fig. 2I and L), and external test set (Fig. 2M and P). Kaplan–Meier curves demonstrated that patients in the low-risk group had improved outcomes compared to high-risk patients (Fig. 2F, J, and N). Comparable data were achieved through ROC analysis. The AUCs of the signature at 3-year survival time were 0.849, 0.924, and 0.747 in the internal test set (Fig. 2G), the entire TCGA set (Fig. 2K), and the external test set (Fig. 2O), respectively. These data were consistent with the training set data and validated the mRNA signature as a reliable predictor for overall survival in patients with EAC.

To investigate if the prognostic signature can be used independently of other risk factors, we performed univariable and multivariable Cox regression analyses across the TCGA dataset considering age, sex, height, weight, alcohol consumption, Barrett esophagitis, and TNM stage (Supplemental Table S2, http://links.lww.com/MD2/A31). As shown in the Figure 3, the mRNA signature was most significantly related to overall survival (hazard ratio, 4.32; 95% confidence interval, 1.91–9.79; $P = .00046$). In addition, the sensitivity and specificity of the mRNA signature performed numerically better than the TNM stage for 1-, 2-, and 3-year prognostic evaluation of EAC (Supplemental Fig. S3, http://links.lww.com/MD2/A27). When patients were stratified by clinicopathological factors, the mRNA signature represented a statistically significant prognostic model (Fig. 4). These findings demonstrate that high-risk scores possessed a strong association with a poor prognosis.

3.5. Clinical and immune relevance of the mRNA signature

We next evaluated the correlation between the mRNA signature and clinical parameters (tumor size, node status, metastasis, and TNM stage) of EAC in TCGA set. Those with a positive lymph node status possessed higher risk scores (Fig. 5A). Stage III+IV patients also showed higher risk scores upon comparison to stage I+II patients (Fig. 5B).

Using the TIMER analysis, the low-risk group was significantly related to higher levels of B cell infiltration (Fig. 5C). The expression of SLC25A45 negatively correlated with the infiltration of B cell ($P < .0001$; Fig. 5D). High levels of B cell infiltration prolonged survival in the EAC patients ($P = .078$; Fig. 5E).

3.6. Identification of biological processes associated with the mRNA signature

To infer the potential mechanisms of the mRNA signature, GSEA was applied to the TCGA and GSE72874 datasets. Gene sets based on KEGG and REACTOME pathways were associated with high-risk patients and identified in the TCGA datasets (Supplemental Table S3, http://links.lww.com/MD2/A32 and S4, http://links.lww.com/MD2/A34). The top 5 KEGG and REACTOME pathways in the TCGA are shown in Figure 6A and B. Meanwhile, GSEA was also performed in the GSE72874 dataset, and the results associated with KEGG and REACTOME pathways were presented in Supplemental Tables S5, http://links.lww.com/MD2/A35 and S6, http://links.lww.com/MD2/

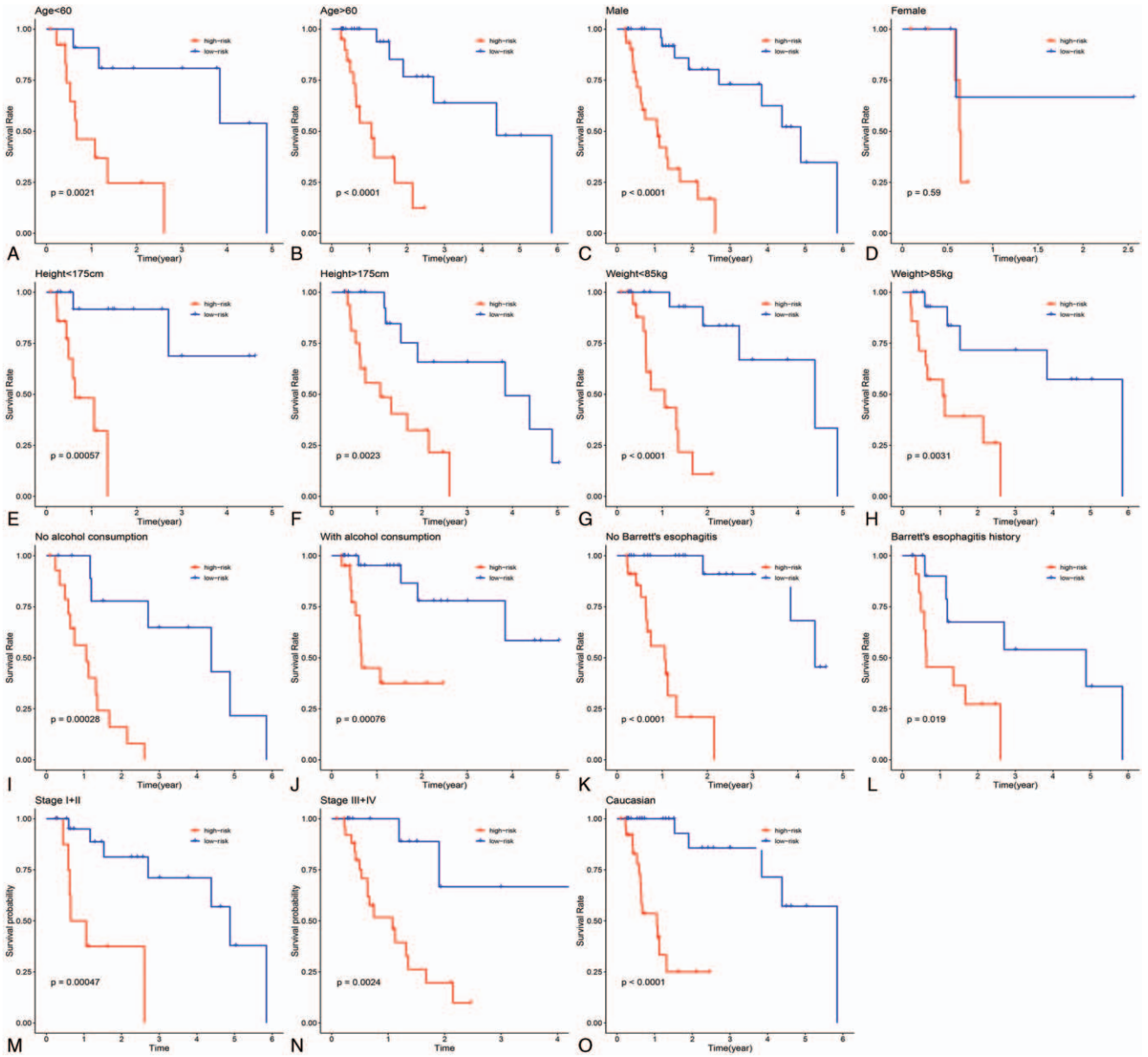

Figure 4. Survival curves for EAC patients in different subgroup in the entire TCGA set. Kaplan–Meier analysis (low risk vs high risk) in subgroups stratified by age (A, B), sex (C, D), height (E, F), weight (G, H), alcohol consumption (I, J), Barrett esophagitis (K, L), TNM stage (M, N), and race (O). EAC = esophageal adenocarcinoma, TCGA = The Cancer Genome Atlas, TNM = tumor–node–metastasis.

A37. Two commonly enriched KEGG pathways were screened including cell cycle and PI3K–Akt signaling pathway. The enriched REACTOME pathways were involved in nucleotide metabolism (DNA replication, DNA synthesis, and repair) and cell cycle progression (G1/S phase, S phase, and G2/M phase). We constructed enrichment maps and organized the enriched terms into networks with overlapping gene sets (Supplemental Fig. S4, http://links.lww.com/MD2/A29).

As shown in Figure 6C to E, the mRNA signature correlated with chromosomes in the CC, mediated MFs such as DNA-dependent ATPase activity and catalytic activity, and regulated DNA replication, chromosome segregation, and RNA splicing of the BP. These corresponding biological functions may contribute to poor prognosis of EAC patients with high-risk score.

4. Discussion

As a highly malignant neoplasm, the incidence of EAC is increasing and has surpassed ESCC in some areas of North America.[2] TNM staging is a conventional and effective tool based on anatomical information. It helps us to improve current empirical treatment decisions and predict prognosis in patients with EAC. It is however unable to achieve adequate assessments of disease outcome in EAC. Molecular prognostic biomarkers such as mRNAs, microRNAs, and lncRNAs can supplement or substitute TNM staging in EAC.[6,14] Gene expression and relevant clinical data of EAC available in TCGA and GEO facilitate the establishment of novel molecular signatures related to prognosis. Here, we employed TCGA and GEO datasets to build prognostic mRNA signatures for EAC.

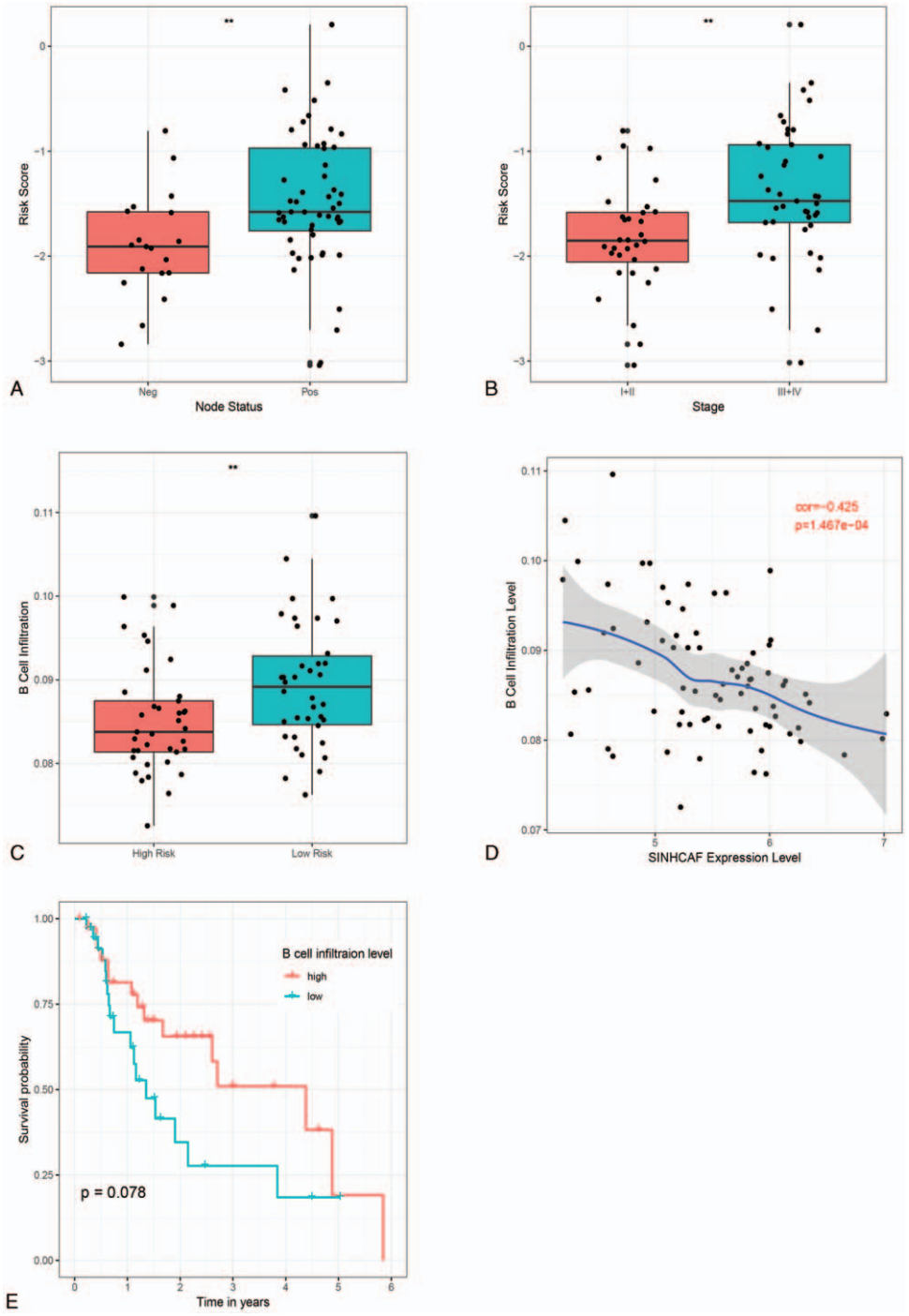

Figure 5. Clinical- and tumor immunity relevance of the mRNA signature. The distribution of the risk score in different status of node and TNM stage in the whole TCGA set (A, B). The B cell infiltration level in high risk and low risk group in the entire TCGA set (C). The abundances of B cells were estimated using the TIMER algorithm (https://cistrome.shinyapps.io/timer/). The correlation of SINHCAF expression with B cell infiltration level in EAC patients (D). The Kaplan–Meier survival curves of different B cell infiltration level (E). $^{*}P < .05$, $^{**}P < .01$, $^{***}P < .001$, $^{****}P < .0001$. EAC = esophageal adenocarcinoma, TCGA = The Cancer Genome Atlas, TIMER = the tumor immune estimation resource, TNM = tumor–node–metastasis.

We screened and selected significant DEMs associated with survival in the training set. Based on LASSO, we constructed a 5-mRNA signature in the training set and further validated its accuracy in internal and external test sets. As a supplement of TNM staging, the 5-mRNA signature could guide the stratification of patients with EAC and aid in decision making for individualized treatments.

SLC26A9, SINHCAF, MICB, KRT19, and MT1X formed the 5-mRNA signature that could predict the prognosis of EAC in the independent datasets. SINHCAF, also known as FAM60A, is a component of SIN3/HDAC deacetylase complex.[18] In esophageal carcinoma, FAM60A was reported as a driver gene with a significant correlation with prognosis. FAM60A silencing could inhibit the proliferation, migration, and invasion of cells, increase

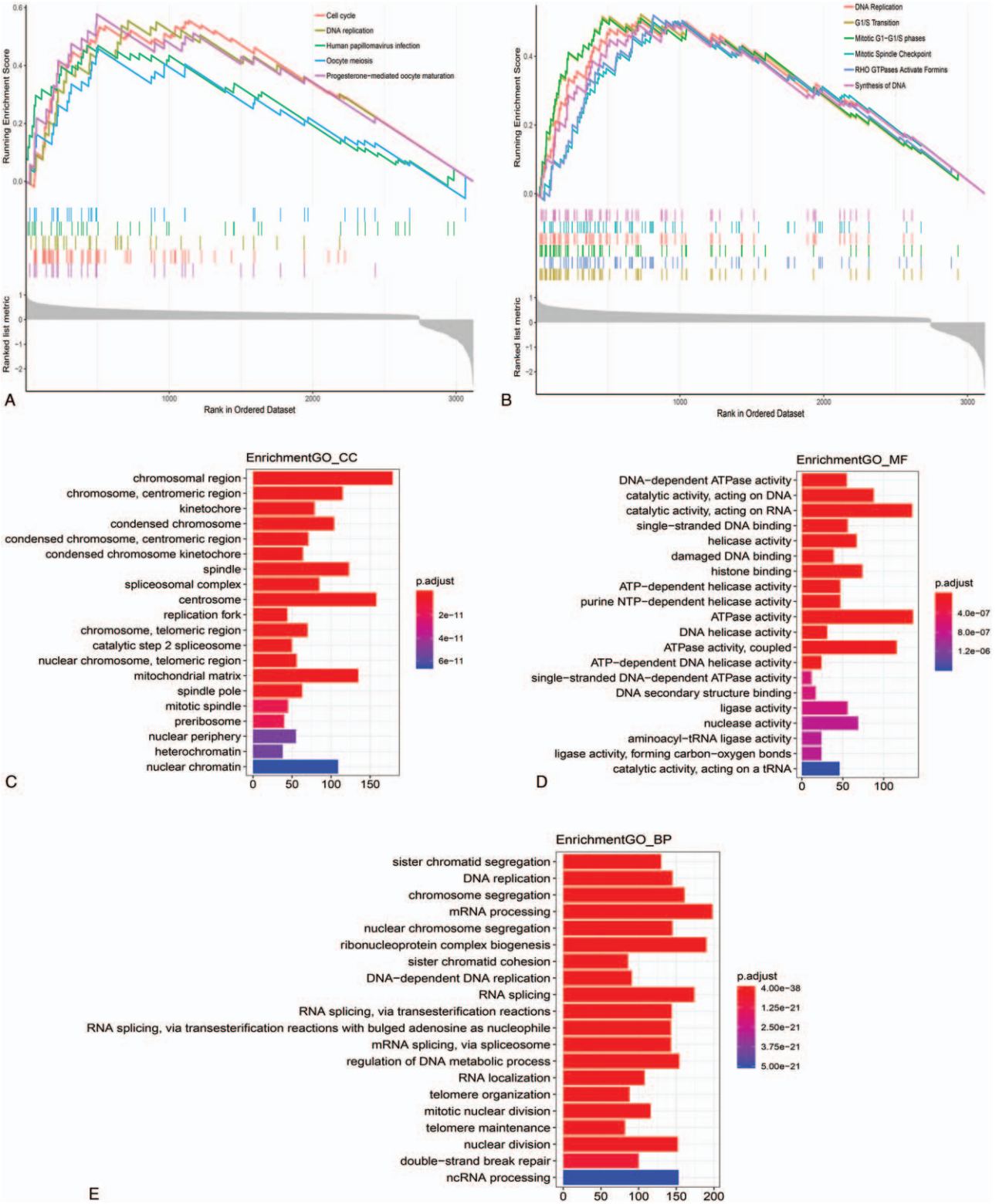

Figure 6. Functional analysis depicted the potential biological mechanism associated with the mRNA signature. The top 5 enriched KEGG pathways in patients with high risk (A). The top 5 enriched REACTOME pathways in patients with high risk (B). Significant GO terms enriched by the mRNA signature (C, D, E). GO = Gene Ontology, GSEA = Gene Set Enrichment Analysis, KEGG = Kyoto Encyclopedia of Genes and Genomes.

apoptosis and arrest cells in the G2/M phase.[19] FAM60A is overexpressed in gastric cancer tissues compared to adjacent tissue, and its overexpression enhances the gastric cancer cell proliferation through the PI3K/AKT pathway.[20] In lung cancer and liver cancer cells, FAM60A acts as a tumor suppressor,[18] which is not consistent with present study.

Major histocompatibility complex (MHC) class I chain-related protein B (MICB) is a ligand of NKG2D receptors on NK cells

and T cells.[21] When MICB is expressed on the cell surface, engagement between MICB and NKG2D actives the tumor killing effects of NK cells and T cells.[22] In hepatocellular carcinoma and cervical cancer, patients with high expression of MICB showed an improved outcome.[23,24] However, when MICB is retained in the cytoplasm, higher rates of immune evasion of cancer cells occurs, decreasing patient survival.[25] The function of MICB markedly differs depending on the cancer type. Li et al[26] reported that high levels of MICB are linked to poor prognosis in ovarian cancer. In ESCA, the expression of MICB was upregulated and associated with the histological grade of ESCA.[27]

The function of KRT19 in a range of cancer types differs to its role in EAC based on our findings. KRT19 promoted cancer cell proliferation and invasion and was identified as a marker of poor prognosis for hepatocellular carcinoma, pancreatic ductal adenocarcinoma, and breast cancer.[28–31] Liu et al[32] found that MT1X could serve as a favorable prognostic marker and inhibit the progression and metastasis of hepatocellular carcinoma. In oral squamous cell carcinoma, the upregulation of MT1X was related with a lack of metastasis which indicated that MT1X may aid prediction of prognosis.[33] However, few cancer-related studies have been performed regarding the function of SLC26A9, which warrants further investigation.

In this study, we found that B cell infiltration was significantly higher in low-risk groups and suggestive of an improved outcome. T cells are therapeutic targets with the immune checkpoint inhibitors in ESCA, but the efficacy of this therapy is variable.[34,35] B cells are major effector cells of humoral immunity and mediate immune responses. They prevent cancer progression through immunoglobulins secretion, T cell responses enhancement and cancer cell killing.[36] To balance immunotherapy and adverse events, further stratification based on B cell infiltration or risk scores are necessary to make informed treatment decisions. Further experimental studies are similarly required to decipher the function of B cells during cancer progression.

To our knowledge, this mRNA signature associated with prognosis built using the TCGA and GEO datasets have not been reported previously. Inevitably, there are some limitations in the current study. First, although 2 datasets were used to construct and validate the signature, the sample size was small. This signature requires further validated in large prospective clinical trials. In addition, Dong et al[19] demonstrated the function of SINHCAF in ESCA, but no experimental studies on the other genes of the signature in ESCA have been reported. To better understand the potential mechanisms behind the signature, in vitro and in vivo experiments should be performed.

In conclusion, despite the limitations described, the 5-mRNA signature that we constructed through a comprehensive analysis of the mRNA profiles of EAC may serve as a novel and reliable tool to aid prognosis predictions and tailored treatments for EAC. This study also provides significant information for molecular research and clinical treatment in EAC.

Acknowledgments

The authors sincerely thank the TCGA and GEO datasets for offering access to their resources.

Author contributions

Conceptualization: Tian Lan, Hua Luo.
Supervision: Hua Luo.
Validation: Yunyan Lu.
Visualization: Yunyan Lu.
Writing – original draft: Tian Lan.
Writing – review & editing: Tian Lan, Weiguo Liu, Hua Luo.

References

[1] Fitzmaurice C, Dicker D, Pain A, et al. Global Burden of Disease Cancer CollaborationThe Global Burden of Cancer 2013. JAMA Oncol 2015;1:505–27.
[2] Rustgi AK, El-Serag HB. Esophageal carcinoma. N Engl J Med 2014; 371:2499–509.
[3] Alsop BR, Sharma P. Esophageal cancer. Gastroenterol Clin North Am 2016;45:399–412.
[4] Saxena R, Klochkova A, Murray MG, et al. Roles for autophagy in esophageal carcinogenesis: implications for improving patient outcomes. Cancers 2019;11:1697.
[5] Rice TW, Ishwaran H, Blackstone EH, et al. Recommendations for clinical staging (cTNM) of cancer of the esophagus and esophagogastric junction for the 8th edition AJCC/UICC staging manuals. Dis Esophagus 2016;29:913–9.
[6] Kim SM, Park YY, Park ES, et al. Prognostic biomarkers for esophageal adenocarcinoma identified by analysis of tumor transcriptome. PLoS One 2010;5:e15074.
[7] Peters CJ, Rees JR, Hardwick RH, et al. A 4-gene signature predicts survival of patients with resected adenocarcinoma of the esophagus, junction, and gastric cardia. Gastroenterology 2010;139:1995.e15–2004.e15.
[8] Pennathur A, Xi L, Litle VR, et al. Gene expression profiles in esophageal adenocarcinoma predict survival after resection. J Thorac Cardiovasc Surg 2013;145:505–12.
[9] Clough E, Barrett T. The Gene Expression Omnibus Database. Methods Mol Biol 2016;1418:93–110.
[10] Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci 2012;131:281–5.
[11] Krause L, Nones K, Loffler KA, et al. Identification of the CIMP-like subtype and aberrant methylation of members of the chromosomal segregation and spindle assembly pathways in esophageal adenocarcinoma. Carcinogenesis 2016;37:356–65.
[12] Li T, Fan J, Wang B, et al. TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. Cancer Res 2017;77: e108–10.
[13] Gao J, Kwan PW, Shi D. Sparse kernel learning with LASSO and Bayesian inference algorithm. Neural Netw 2010;23:257–64.
[14] Lan T, Lu Y, Xiao Z, et al. A six-microRNA signature can better predict overall survival of patients with esophagus adenocarcinoma. PeerJ 2019;7:e7353.
[15] Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545–50.
[16] Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012; 16:284–7.
[17] Gene Ontology C, Blake JA, Dolan M, et al. Gene Ontology annotations and resources. Nucleic Acids Res 2013;41:D530–5.
[18] Smith KT, Sardiu ME, Martin-Brown SA, et al. Human family with sequence similarity 60 member A (FAM60A) protein: a new subunit of the Sin3 deacetylase complex. Mol Cell Proteomics 2012;11:1815–28.
[19] Dong G, Mao Q, Yu D, et al. Integrative analysis of copy number and transcriptional expression profiles in esophageal cancer to identify a novel driver gene for therapy. Sci Rep 2017;7:42060.
[20] Yao X, Liu D, Zhou L, et al. FAM60A, increased by Helicobacter pylori, promotes proliferation and suppresses apoptosis of gastric cancer cells by targeting the PI3K/AKT pathway. Biochem Biophys Res Commun 2020;521:1003–9.
[21] Schmiedel D, Mandelboim O. NKG2D ligands-critical targets for cancer immune escape and therapy. Front Immunol 2018;9:2040.
[22] Baginska J, Viry E, Paggetti J, et al. The critical role of the tumor microenvironment in shaping natural killer cell-mediated anti-tumor immunity. Front Immunol 2013;4:490.
[23] Fang L, Gong J, Wang Y, et al. MICA/B expression is inhibited by unfolded protein response and associated with poor prognosis in human hepatocellular carcinoma. J Exp Clin Cancer Res 2014;33:76.

[24] Cho H, Chung JY, Kim S, et al. MICA/B and ULBP1 NKG2D ligands are independent predictors of good prognosis in cervical cancer. BMC Cancer 2014;14:957.
[25] Zhao Y, Chen N, Yu Y, et al. Prognostic value of MICA/B in cancers: a systematic review and meta-analysis. Oncotarget 2017;8: 96384–95.
[26] Li K, Mandai M, Hamanishi J, et al. Clinical significance of the NKG2D ligands, MICA/B and ULBP2 in ovarian cancer: high expression of ULBP2 is an indicator of poor prognosis. Cancer Immunol Immunother 2009;58:641–52.
[27] Mei JZ, Zhao JZ, Yang GY, et al. Expression of MICA/B protein in esophageal cancer and its clinical significance in Chinese]. Zhonghua Zhong Liu Za Zhi [Chin J Oncol] 2012;34:745–7.
[28] Takano M, Shimada K, Fujii T, et al. Keratin 19 as a key molecule in progression of human hepatocellular carcinomas through invasion and angiogenesis. BMC Cancer 2016;16:903.
[29] Rhee H, Kim HY, Choi JH, et al. Keratin 19 expression in hepatocellular carcinoma is regulated by fibroblast-derived HGF via a MET-ERK1/2-AP1 and SP1 axis. Cancer Res 2018;78:1619–31.
[30] Yao H, Yang Z, Liu Z, et al. Glypican-3 and KRT19 are markers associating with metastasis and poor prognosis of pancreatic ductal adenocarcinoma. Cancer Biomark 2016;17:397–404.
[31] Kabir NN, Ronnstrand L, Kazi JU. Keratin 19 expression correlates with poor prognosis in breast cancer. Mol Biol Rep 2014;41:7729–35.
[32] Liu Z, Ye Q, Wu L, et al. Metallothionein 1 family profiling identifies MT1X as a tumor suppressor involved in the progression and metastastatic capacity of hepatocellular carcinoma. Mol Carcinog 2018;57:1435–44.
[33] Brazao-Silva MT, Rodrigues MF, Eisenberg AL, et al. Metallothionein gene expression is altered in oral cancer and may predict metastasis and patient outcomes. Histopathology 2015;67:358–67.
[34] Tokunaga R, Naseem M, Lo JH, et al. B cell and B cell-related pathways for novel cancer treatments. Cancer Treat Rev 2019;73:10–9.
[35] Vrana D, Matzenauer M, Neoral C, et al. From tumor immunology to immunotherapy in gastric and esophageal cancer. Int J Mol Sci 2018;20:
[36] Wang SS, Liu W, Ly D, et al. Tumor-infiltrating B cells: their role and application in anti-tumor immunity in lung cancer. Cell Mol Immunol 2019;16:6–18.