Am J Transl Res 2022;14(11):8064-8084
www.ajtr.org /ISSN:1943-8141/AJTR0145040

Original Article

ER stress-related mRNA-lncRNA co-expression gene signature predicts the prognosis and immune implications of esophageal cancer

Feng Li*, Jiahao Ma*, Cheng Yan, Yonghua Qi

*School of Pharmacy, Key Laboratory of Nano-carbon Modified Film Technology of Henan Province, Diagnostic Laboratory of Animal Diseases, Xinxiang University, Xinxiang, Henan, China. *Equal contributors.*

Received June 29, 2022; Accepted October 27, 2022; Epub November 15, 2022; Published November 30, 2022

Abstract: Background: Esophageal cancer (EC) is one of the most common malignant cancers in the world. Endoplasmic reticulum (ER) stress is an adaptive response to various stress conditions and has been implicated in the development of various types of cancer. Long noncoding RNAs (lncRNAs) refer to a group of noncoding RNAs (ncRNAs), which regulate gene expression by interacting with DNA, RNA and proteins. Accumulating evidence suggests that lncRNAs are critical regulators of gene expression in development, differentiation, and human diseases, such as cancers and heart diseases. However, the prognostic model of EC based on ER stress-related mRNA and lncRNA has not been reported. Methods: Firstly, we downloaded RNA expression profiles from The Cancer Genome Atlas (TCGA) and obtained ER stress-related genes from the Molecular Signature Database (MSigDB). Next, Weighted Correlation Network Analysis (WGCNA) co-expression analysis was used to identify survival-related ER stress-related modules. Prognostic models were developed using univariate and Least absolute shrinkage and selection operator (LASSO) regression analyses on the training set and validated on the test set. Afterwards, The Receiver Operating Characteristic (ROC) curve and nomogram were used to evaluate the performance of risk prediction models. Differentially expressed gene (DEG) and enrichment analysis were performed between different groups in order to identify the biological processes correlated with the risk score. Finally, the fraction of immune cell infiltration and the difference of tumor microenvironment were identified in high-risk and low-risk groups. Results: The WGCNA co-expression analysis identified 49 ER genes that are highly associated with EC prognosis. Using univariate Cox regression and LASSO regression analysis, we developed prognostic risk models based on nine signature genes (four mRNAs and five lncRNAs). Both in the training and in the test sets, the overall survival (OS) of EC patients in the high-risk group was significantly lower than that in the low-risk group. The Kaplan-Meier curve and the ROC curve demonstrate the prognostic model we built can precisely predict the survival with more than 70% accuracy. The correlation analysis between the risk score and the infiltration of immune cells showed that the model can indicate the state of the immune microenvironment in EC. Conclusion: In this study, we developed a novel prognostic model for esophageal cancer based on ER stress-related mRNA-lncRNA co-expression profiles that could predict the prognosis, immune cell infiltration, and immunotherapy response in patients with EC. Our results also may provide clinicians with a quantitative tool to predict the survival time of patients and help them individualize treatment strategies for the patients with EC.

Keywords: Esophageal cancer, endoplasmic reticulum stress, mRNA, lncRNA, prognostic model

Introduction

Esophageal cancer (EC) is one of the most common malignancies worldwide, ranking as the eighth most common type of cancer globally [1]. Esophageal cancer has two main histological types: squamous cell carcinoma and adenocarcinoma. In China, esophageal squamous cell carcinoma is the main subtype of esophageal cancer, accounting for 90% of the total number of esophageal cancer [2]. Accordingly, esophageal adenocarcinomas count for the vast majority esophageal cancer in western countries. The five-year survival rate of patients with EC is less than 20% [3]. Most cases of EC are diagnosed at a late stage, missing the best opportunity for treatment and leading to a dismal prognosis [4]. Therefore, it is crucial to iden-

tify biomarkers that act as both prognosis-predictive markers and therapy targets for EC.

The Endoplasmic reticulum (ER) is a cellular organelle that responsible for protein synthesis and trafficking, protein folding, lipid and steroid synthesis, carbohydrate metabolism, and calcium storage [5]. Numerous abnormal cellular states, such as intracellular calcium abnormalities, disrupted glycosylation modifications, and redox disorders, cause ER dysfunction and induce endoplasmic reticulum stress (ER stress) [6]. In response to these abnormal factors, cells will initiate adaptive responses to decrease ER stress and restore cellular homeostasis, known as the unfolded protein response (UPR) [7]. Increasing evidence suggests that ER stress is a double-edged sword during EC development and progression [8-11]. During ER stress, EC cells either survive by inducing adaptive mechanisms or undergo cell death by apoptosis. Activation of the IRE1a pathway by AURKA can promote esophageal cancer cell survival in FLO-1 and OE33 [12]. Up-regulation of ATF4 promotes esophageal cancer cell metastasis by regulating matrix metalloproteinases [10]. Conversely, ER stress can induce autophagy and apoptosis in human esophageal cancer EC9706 cells by mediating the PI3K/Akt/mTOR signaling pathway [11]. These findings indicate that ER stress may play an important role in the proliferation and migration of EC, and ER stress related genes may serve as prognostic markers for EC. However, the specific functions and immune implications in EC have been less studied.

Long non-coding RNAs are a class of RNAs that are more than 200 nucleotides in length and the do not encode proteins, but tightly regulate gene expression [13]. LncRNAs have important roles in cancer, including epigenetic regulation, DNA damage and cell cycle regulation, regulation of microRNAs, involvement in signal transduction pathways, and hormone-induced cancer mediation [14-17]. Indeed, increasing studies have demonstrated that lncRNAs play an important role in esophageal cancer [18-21]. LncRNA SNHG7 can promote esophageal cancer cell proliferation and inhibit apoptosis by regulating P15 and P16 expression [18]. LncRNA UCA1 enhances cell proliferation by acting as ceRNA to inhibit miR-498 expression and thereby increasing ZEB2 expression in esophageal cancer [19]. Furthermore, lncRNAs also play a suppressive role in esophageal cancer. RPLS1 inhibits proliferation, migration, and invasion of esophageal cancer cells by down-regulating RPL34 expression [21]. Increasing evidence indicates that lncRNA-mRNA co-expression networks could provide new insights into tumorigenesis and tumor development [22, 23]. However, to the best of our knowledge, a prognostic model for esophageal cancer based on ER stress-related mRNA-lncRNA co-expression network profiles has not been reported.

In this study, we identified ER stress-related mRNAs that were associated with prognosis of EC by using WGCNA. Pearson correlation analysis was conducted to identify lncRNAs highly correlated with these ER stress-related mRNA. By performing univariate Cox regression and LASSO regression analysis, we developed a prognostic risk model based on nine signature genes (four mRNAs and five lncRNAs). Both in the training and the test sets, the overall survival of EC patients in the high-risk group was significantly lower than that in the low-risk group. The Kaplan-Meier curve and the ROC curve were performed to estimate the sensitivity and specificity of the prognostic signature. According to the pRRophetic algorithm, we identified four potential drugs (OSI.906, BI-2536, Lenalidomide, Bicalutamide) that might have potential therapeutic effect of EC. In conclusion, we developed a novel prognostic model for esophageal cancer based on ER stress-related mRNA-lncRNA co-expression profiles. Our results will not only provide clinicians with a quantitative tool to predict the survival time of patients and help them to individualize treatment strategies for the patients with EC, but also provide nine ER-stress related biomarkers for EC.

Methods

Data acquisition

The Cancer Genome Atlas (TCGA) database was created by the National Cancer Institute and contains genomic, transcriptomic, proteomic, and methylation data for 20,000 primary cancers (http://cancergenome.nih.gov/). The TCGA platform is publicly accessible, free to use and search, and download, in order to use the original data for integrated analysis [24]. From the TGCA, we collected transcrip-

tomic (lncRNA and mRNA) and clinicopathological data on 185 patients with EC (Supplementary Table 1). By searching the keyword "ER stress" in the Molecular Signature Database (MSigDB) database, 419 genes associated with ER stress were extracted (http://www.gsea-msigdb.org/gsea/msigdb/index.jsp) (Supplementary Table 2).

The weighted correlation network analysis (WGCNA)

In order to identify the new prognostic markers and potential therapeutic targets, we performed the WGCNA approach (WGCNA R package) to recognize gene modules highly correlated with the clinical phenotypes of patients with EC [25]. The analysis process includes: first, downloading expression data from the TCGA; second, creating gene co-expression networks by calculating the connection strength between genes; third, using hierarchical clustering and dynamic tree cutting to identify gene modules; fourth, construction of the relationships between gene module and clinical traits. Finally, a scale-free topology model was built. Power value = 16 and scale-free R^2 = 0.90 were chosen as soft threshold parameters to obtain a good scale-free topology model. Genes with high correlations were clustered into the same co-expression module for FlashClust analysis to generate the cluster dendrogram. Then, we transform the cluster dendrogram into a topology matrix to form the network heatmap plot. The relationships between modules and nine EC prognostic factors (survival time, survival state, age at initial pathologic diagnosis, gender, grade, stage, T, M, N) were analyzed with Pearson correlation coefficient and visualized by heatmap.

Identification of the mRNA-lncRNA co-expression pairs

We performed correlation analysis using R software to identify the mRNA-lncRNA co-expression pairs with the criteria of Pearson correlation coefficient > 0.6 and *p*-value < 0.001.

Identification of ER stress-related hub genes

In order to explore the relationship among the genes highly correlated with clinical phenotypes of patients with EC, we constructed the protein-protein interaction (PPI network using the Search Tool for the Retrieval of Interacting Genes (STRING) online tool (http://string-db.org) with the parameter of settings the confidence = 0.400 [26]. To identify the hub genes, we used the CytoHubba plugin to screen out key hub mRNAs and significant modules [27]. Maximal Clique Centrality (MCC) and Density of Maximum Neighborhood Component (DMNC) were developed for exploring and identifying key nodes from interactome networks [27, 28]. The intersection of hub genes was obtained by employing MCC and DMNC which employed the above two algorithms.

Construction and validation of the prognostic model

The entire set was randomly divided into train set and test set with a ratio of 5:5. According to the results of univariate Cox regression based on the ER stress-related mRNA and lncRNA, LASSO Cox regression was performed to construct a prognostic signature. Risk scores for EC patients were calculated as follows:

$$risk\ score = \beta i * Xj$$

Xj represents the relative expression levels of each selected gene, and βi represents the coefficient. Taking the median of risk score as the threshold, we classified the samples into high-risk and low-risk groups. The test set was used to validate the prognostic risk score model constructed according the train set. R "survival" package is a tool for statistical analysis and visualization of survival data and is widely used in scientific research [29, 30]. Using the "survival" package in R, we calculated the overall survival analysis and plotted the Kaplan-Meier survival curves. Chi-square tests were applied to the calculation of *p* values [31]. ROC curves were drawn using the R package "survivalROC" to verify the accuracy of the predictive model.

Construction of the nomogram for patients with EC

The nomogram containing the clinical characteristics was established to predict individual survival probability by the "rms" package of R software [32]. To assess the consistency between actual survival time and predicted prognosis in the nomogram, calibration curves

for predicting 1-, 2-, and 3-year survival rate were plotted.

Analysis of immune cell infiltration

The Cell-type Identification by Estimating Relative Subsets of RNA Transcripts (CIBERSORT) method is a general way to measure cell fractions based on the gene expression profiles (GEPs), which can accurately estimate the immune component of tumor biopsies [33]. We calculated the abundance of 22 immune cell infiltrates using the CIBERSORT deconvolution method in high- and low-risk groups, and then used the Wilcoxon test to compare the differential immune density in the two groups [34]. The level of statistical significance was set at $P < 0.05$.

Permutation test

The permutation test is a non-parametric method for analyzing the statistical significance, which is a very useful method to estimate the sampling distribution for the test statistic in situations where a null distribution of test statistics is unknown [35]. We also evaluated the immune infiltration results with a permutation test in patients, in order to verify that the statistically significant results have not been produced due to chance, but because of a systematic difference in the infiltration of these immune cell types. The sample distribution was created by permuting (randomly rearranging) the data one thousand times and calculating the *p*-value (the proportion of samples that have a test statistic larger than that of our observed data) on each permuted version of our data.

Correlation analysis between prognostic model and tumor microenvironment

We used "Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data" (ESTIMATE) package in R to predict the purity of tumors according to the gene expression of each tumor sample. These calculating results indicated that the microenvironment scores of tumors containing stromal, immune, and estimate scores. We divided patients into high- and low-risk groups based on the mean value of the risk scores and conducted the Wilcoxon test to compare the differences among the stromal, immune, and estimate scores in these two groups.

Differentially expressed gene (DEG) analysis between high and low risk groups

In order to explore the potential molecular mechanism underlying the tumorigenesis of EC, we obtained differentially expressed genes between the high-risk and the low-risk groups. We used the "limma" package in R software to select differentially expressed genes with $|\log \text{FoldChange}| \geq 1$ and false discovery rate (FDR) < 0.05.

Enrichment analysis

The gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of these differentially expressed genes between high and low risk groups were performed to find the enriched biological pathways and functions related to the ER stress-Related genes by clusterprofiler R package [36, 37]. The enriched results for GO and KEGG analysis were visualized by the "ggplot2" package. Through gene set enrichment analysis (GSEA), we identified the signaling pathways and biological processes (BPs) in which DEGs between high-risk and low-risk subgroups were enriched (software.broadinstitute.org/gsea).

Drug screening analysis

PRRophetic is an R package that uses tumor gene expression levels to predict clinical chemotherapy responses [38]. The half-maximal inhibitory concentration (IC50) of compounds obtained from the Genomics of Drug Sensitivity in Cancer (GDSC) website. Using the pRRophetic package in R software, we predicted the drug sensitive score of each sample from patients in the high-risk group and low-risk group. The statistical analysis was performed by Wilcox test with a *P* value less than 0.05 as the threshold. To visualize the conformation of drugs in 2D, PubChem online tool (https://pubchem.ncbi.nlm.nih.gov/) was used.

Tissue microarray (TMA) and immunohistochemistry (IHC)

Human Tissue Microarrays (TMA) slides (HEsoSqu180Sur-02), included 90 ESCC tissues and 90 paired adjacent non-tumor tissues, which

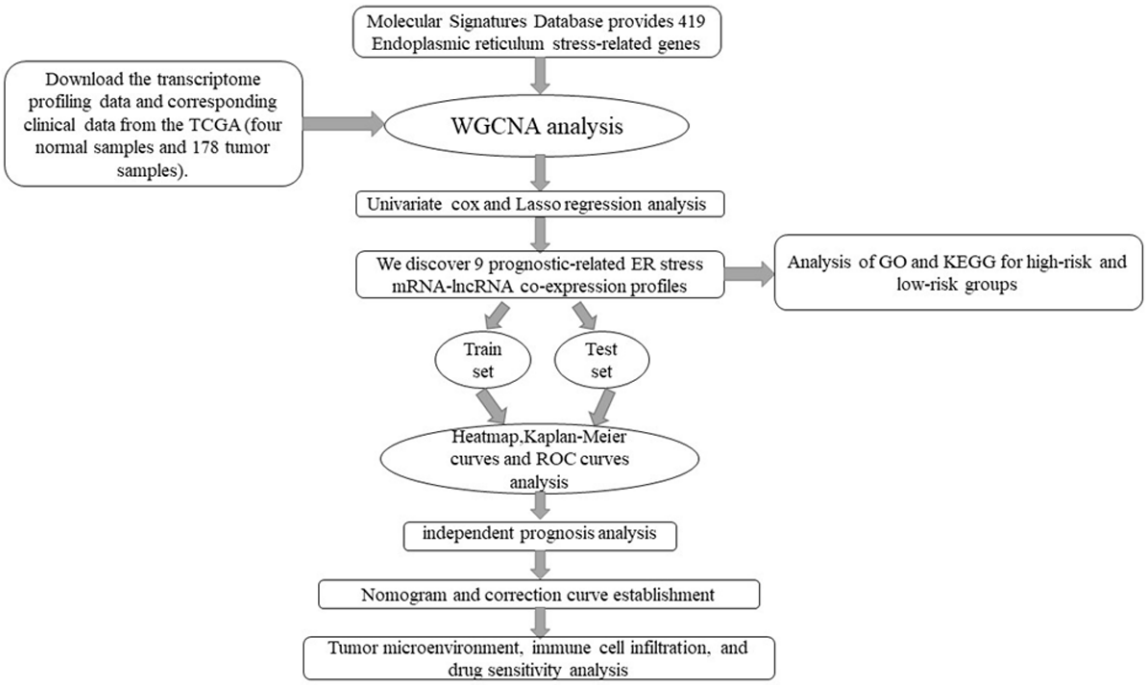


Figure 1. A flowchart of the major steps in this study. WGCNA: Weighted Correlation Network Analysis.

were purchased from Shanghai Outdo Biotech Co., Ltd. TMA slides were dewaxed, dehydrated, endogenous peroxidase blocked, and were antigens retrieved according to standard procedures. The sections were incubated with diluted SERP1 primary antibody (Proteintech, #17807-1-AP) at 4°C overnight and continuously with the anti-rabbit secondary antibody at 37°C for 1 h. Immunohistochemical staining was evaluated based on the scoring system as previously described [39]. High or low expression was defined based on the score of < 6 and ≥ 6, respectively. Use of the tissue microarrays complied with relevant regulations and was approved by the Ethics Committee of Xinxiang University.

Results

Construction of ER stress-related mRNA and lncRNA co-expression modules by WGCNA

The flow chart shows the overall experimental design of this study (**Figure 1**). In order to identify the ER Stress-Related mRNA, we performed WGNCA analysis based on the expression profile of 415 ER stress-related mRNAs from 87 EC samples with the complete clinical data (**Figure 2A**). We selected the power of β = 16 (scale-free R^2 = 0.90) as the soft thresholding to construct a scale-free network (**Figure 2B** and **2C**). Ultimately, there were 4 gene modules generated, and each one was represented in different color, including blue, brown, turquoise and grey (**Figure 2D**). As the brown module was the most relevant to the survival status, age and gender, we selected this module as candidate gene set for further research.

Identification of ER stress-related hub genes

In order to obtain key genes used as a tumor marker for prognosis and potential target for the therapy of EC, we performed the hub gene analysis using cytohubba plug-in Cytoscape software. Firstly, we uploaded the gene list in brown module to STRING database and download the analysis result of protein-protein interaction relationships (**Figure 2E**). Next, the top 10 genes with the highest score of MCC and DMNC were identified, respectively (**Figure 2F** and **2G**). Finally, after intersecting genes from these two methods, we screened out 6 genes as the hub genes including ERN1, ERN2, OS9, CLGN, ERLEC1 and DNAJB9 (**Figure 2H**). These hub genes might be potential biomarkers and new drug targets for EC.

Identification of mRNA-lncRNA co-expression profiles

In order to obtain the lncRNA highly correlated with ER stress-related mRNA, we first per-

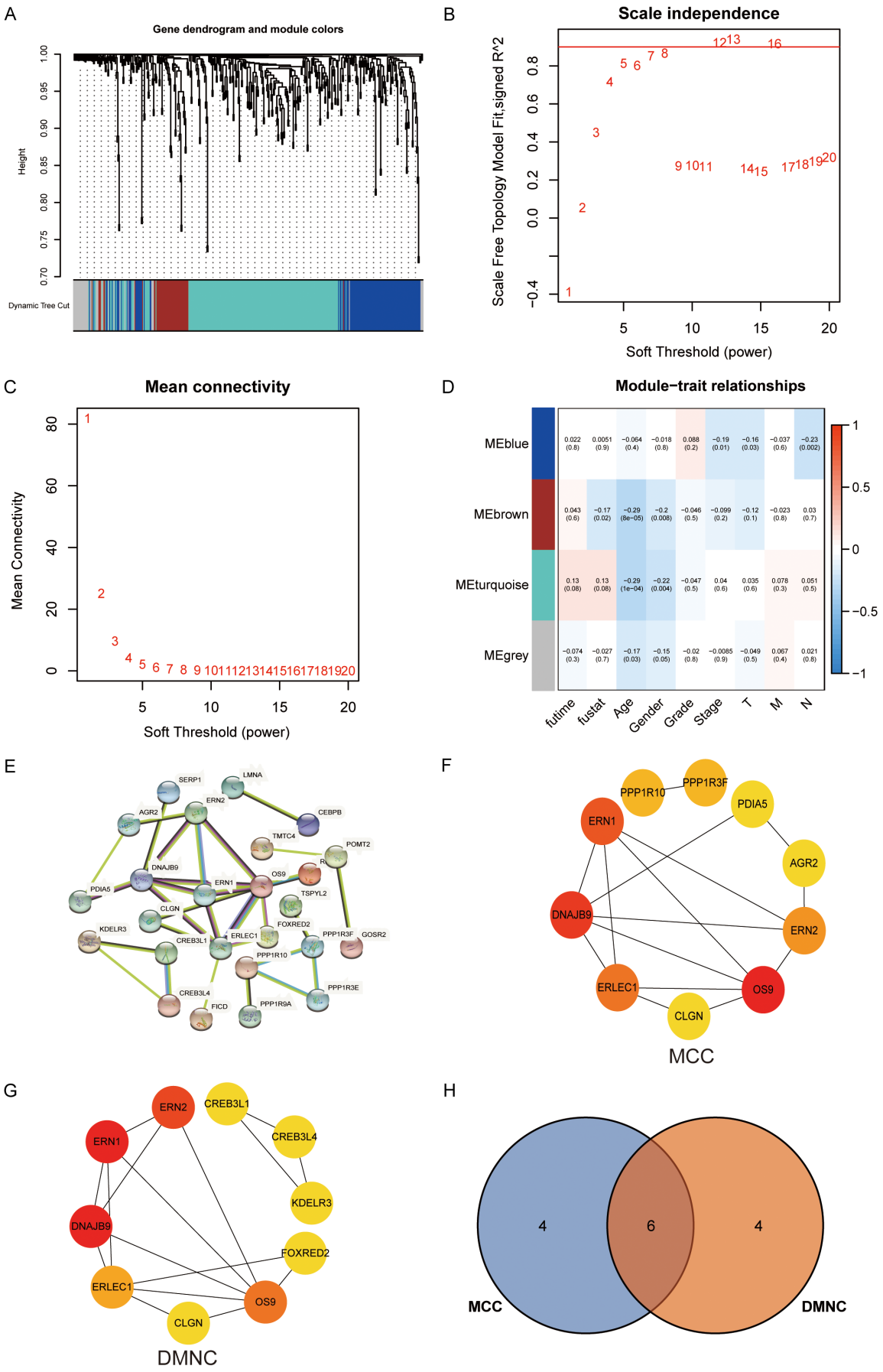

A
Gene dendrogram and module colors
Height
Dynamic Tree Cut
B
Scale independence
Scale Free Topology Model Fit,signed R^2
Soft Threshold (power)
C
Mean connectivity
Mean Connectivity
Soft Threshold (power)
D
Module−trait relationships
MEblue
MEbrown
MEturquoise
MEgrey
futime
fustat
Age
Gender
Grade
Stage
T
M
N
E
SERP1
LMNA
ERN2
AGR2
CEBPB
TMTC4
POMT2
DNAJB9
OS9
ERN1
PDIA5
TSPYL2
CLGN
KDELR3
FOXRED2
CREB3L1
ERLEC1
PPP1R3F
GOSR2
PPP1R10
CREB3L4
PPP1R3E
FICD
PPP1R9A
F
PPP1R10
PPP1R3F
PDIA5
ERN1
AGR2
DNAJB9
ERN2
ERLEC1
OS9
CLGN
MCC
G
ERN2
CREB3L1
ERN1
CREB3L4
KDELR3
DNAJB9
FOXRED2
ERLEC1
OS9
CLGN
DMNC
H
4
6
4
MCC
DMNC

Figure 2. Network analysis of Weighted Correlation Network Analysis (WGCNA) and identification of hub gene. A. The dendrogram was produced by clustering dissimilarity from the colored blocks based on topological overlap. Modules are represented by the colored horizontal bars below the dendrogram. B. Scale-free fit indexes for various soft-threshold powers (β). C. Various soft-thresholding powers and their mean connectivity. D. Heatmap of module-trait correlation. Abscissa represents clinical features and ordinate represents modules of different colors. Each cell displays the correlation coefficient and *p*-value associated with it. The blue, brown, turquoise and grey module were identified as clinical related modules. E. Protein-protein interactions (PPI) network of ER stress-related genes. F. Top 10 genes associated with ER stress as determined by the MCC (Maximal Clique Centrality) algorithm. G. Top 10 genes associated with ER stress as determined by the DMNC (Density of Maximum Neighborhood Component) algorithm. H. Venn diagram showing the intersection of top 10 hub genes by MCC and DMNC algorithms.

formed univariate Cox regression analysis ($P < 0.05$) with the 49 ER stress related genes belonging to the brown module and screened out 11 genes that were significantly associated with the prognosis of EC patients in TCGA dataset (**Figure 3A**). Then, we carried out the lncRNA-mRNA correlation analysis with these 11 mRNAs from the brown module and the total lncRNAs. With the threshold of correlation > 0.60 and $P < 0.001$, we obtained a network consisting 263 nodes and 821 connections, including 255 lncRNAs and 8 mRNAs (Supplementary Table 3). Subsequently, univariate Cox regression analysis was performed again with the lncRNAs in the network and 142 lncRNAs were obtained (Supplementary Table 4). We selected the top 11 significant lncRNA as representative terms shown in **Figure 3B**.

Construction of the prognostic model based on ER stress-related mRNA-lncRNA co-expression profiles

In order to construct the prognostic model and evaluate its performance, we divided the entire set randomly into training set and test set with a ratio of 1:1. Based on univariate Cox regression analysis with mRNA and lncRNA, we conducted the LASSO regression to build the ER stress-related mRNA and lncRNA prognostic model for EC (**Figure 3C**, **3D**). Then, we developed a prognostic model based on four mRNAs and five long noncoding RNAs. The risk score for each patient was calculated by the following formula: risk score = (0.75428 × SERP1) + (-0.13435 × LRRC8D-DT) + (-0.00630 × AL133520.1) + (-0.09910 × T20-AS1) + (-0.16841 × U47924.2) + (-0.10655 × YTHDF3-AS1) + (-0.20987 × TSPYL2) + (-0.03666 × PPP1R10) + (-0.06753 × ATP2A3). Based on a median risk score threshold, we classified EC patients into high- (n = 44) and low-risk (n = 45) groups. Next, we divided the patients in the training set into high-risk (n = 44) and low-risk groups (n = 45) according to the median of risk score. Patients were assigned to high-risk (n = 50) and low-risk groups (n = 51) according to the threshold of the median risk score (**Figure 4A**). EC Patients in the high-risk group were more likely to have a poor prognosis than those in the low-risk group (**Figure 4C**). The heatmap was used to visualize the expression levels of the 9 ER stress-related mRNA and lncRNA in the high- and low-risk groups (**Figure 4E**). Survival curves indicated that patients with EC in the high-risk group had a significantly lower survival probability compared to the patients in low-risk group ($P < 0.05$) (**Figure 4G**). ROC analysis revealed that the area under the curves (AUCs) for 1-, 3-, and 5-year OS were 0.711, 0.724, and 0.882, respectively (**Figure 4I**). The density plots was plotted to appropriately depict the distribution of low and high risk patients according to their survival time (Supplementary Figure 1A, 1B). These data suggested that our model has an acceptable capacity to predict the prognosis of EC patients.

Validation of the prognostic model in test set

In order to validate the prognostic model in the test set, we calculated the risk score of each patient in the test set according to the same risk score formula we constructed. The patients in the test set were divided into the high-risk group (n = 50) and low-risk group (n = 38) based on the cutoff value of the training set (**Figure 4B**). The survival status and the heatmap of these 9 prognostic genes in the test set are shown in **Figure 4D**, **4F**. Consistent with the results of the training set, patients from the high-risk group in the test set showed a worse prognosis compared to the patients from the low-risk group (**Figure 4H**). Furthermore, ROC analysis revealed that the area under the curves (AUCs) for 1-, 3-, and 5-year OS were 0.690, 0.756, and 0.959, respectively (**Figure 4J**). These data suggested that our prognostic

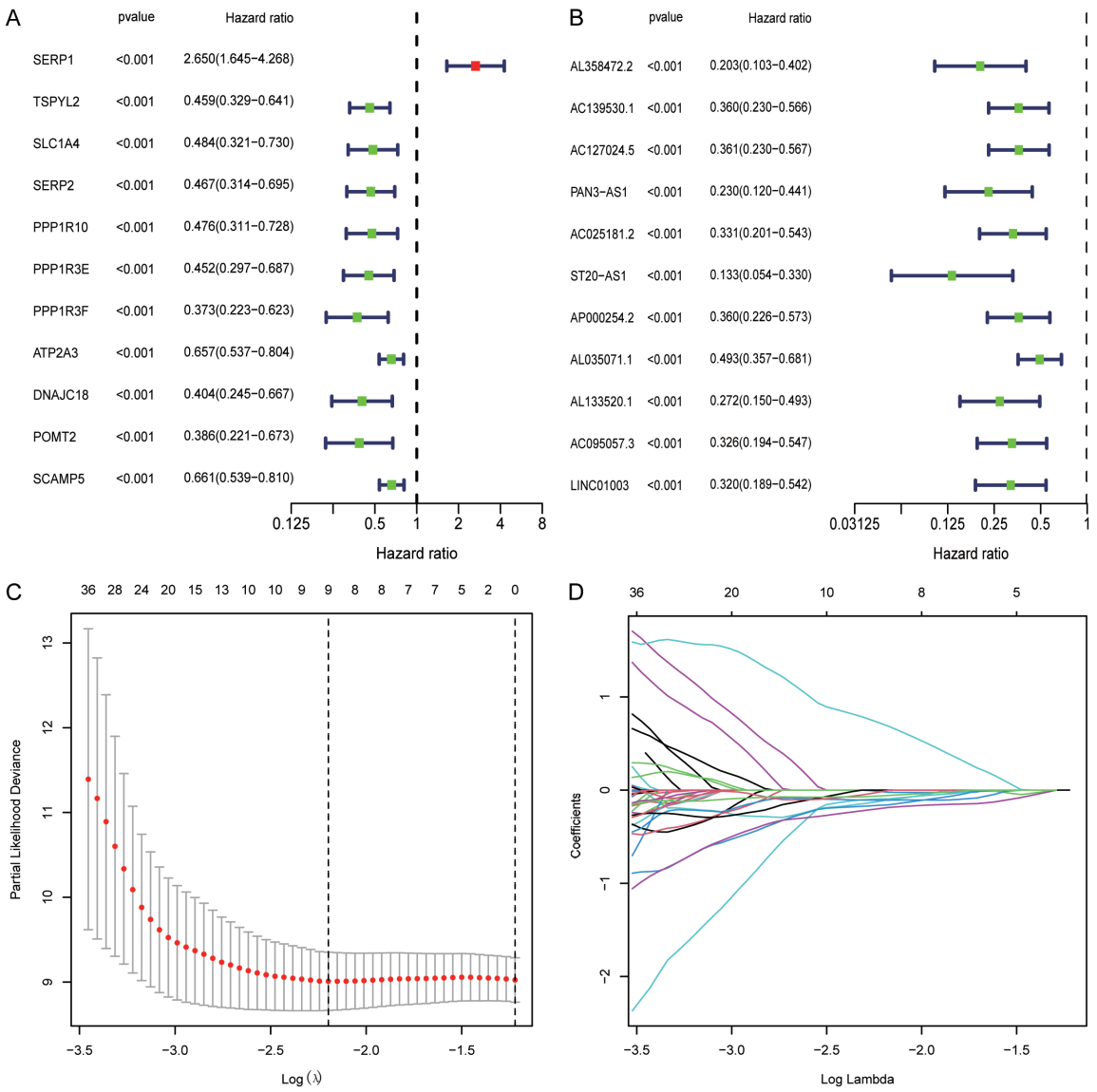


Figure 3. Identification of the endoplasmic reticulum (ER) stress-related mRNAs and lncRNAs by univariate Cox regression and Lasso regression analysis in TCGA esophageal cancer cohort. (A) Univariate Cox regression analysis of ER stress-related mRNAs. (B) Univariate Cox regression analysis of ER stress-related lncRNAs. We selected the top 11 significant lncRNA as representative terms shown in (B). (C) Partial likelihood deviance with changing of log (λ) plotted through LASSO Cox regression in 10-fold cross-validations. (D) Coefficients with changing of log (λ) plotted through LASSO Cox regression in 10-fold cross-validations. TCGA: The Cancer Genome Atlas.

model could also accurately predict prognosis of EC patients from test set.

Functional analysis of differentially expressed genes between high- and low-risk groups

To indicate the biological functions and pathways that were associated with the risk score, we obtained the differentially expressed genes between the high- and low-risk groups with a cutoff value of $|\log_2 \text{fold change}| > 1$ and false discovery rate (FDR) < 0.05, including 34 up-regulated genes and 517 down-regulated genes (Supplementary Table 5). Then, we performed GO and KEGG enrichment analysis of these DEGs to identify the biological processes correlated with the risk score (Supplementary Table 6). GO enrichment analysis involving the biological process (BP) category indicated that these DEGs are predominantly associated with signal release, hormone transport and insulin secretion (**Figure 5A**, **5B**). For the cellular com-

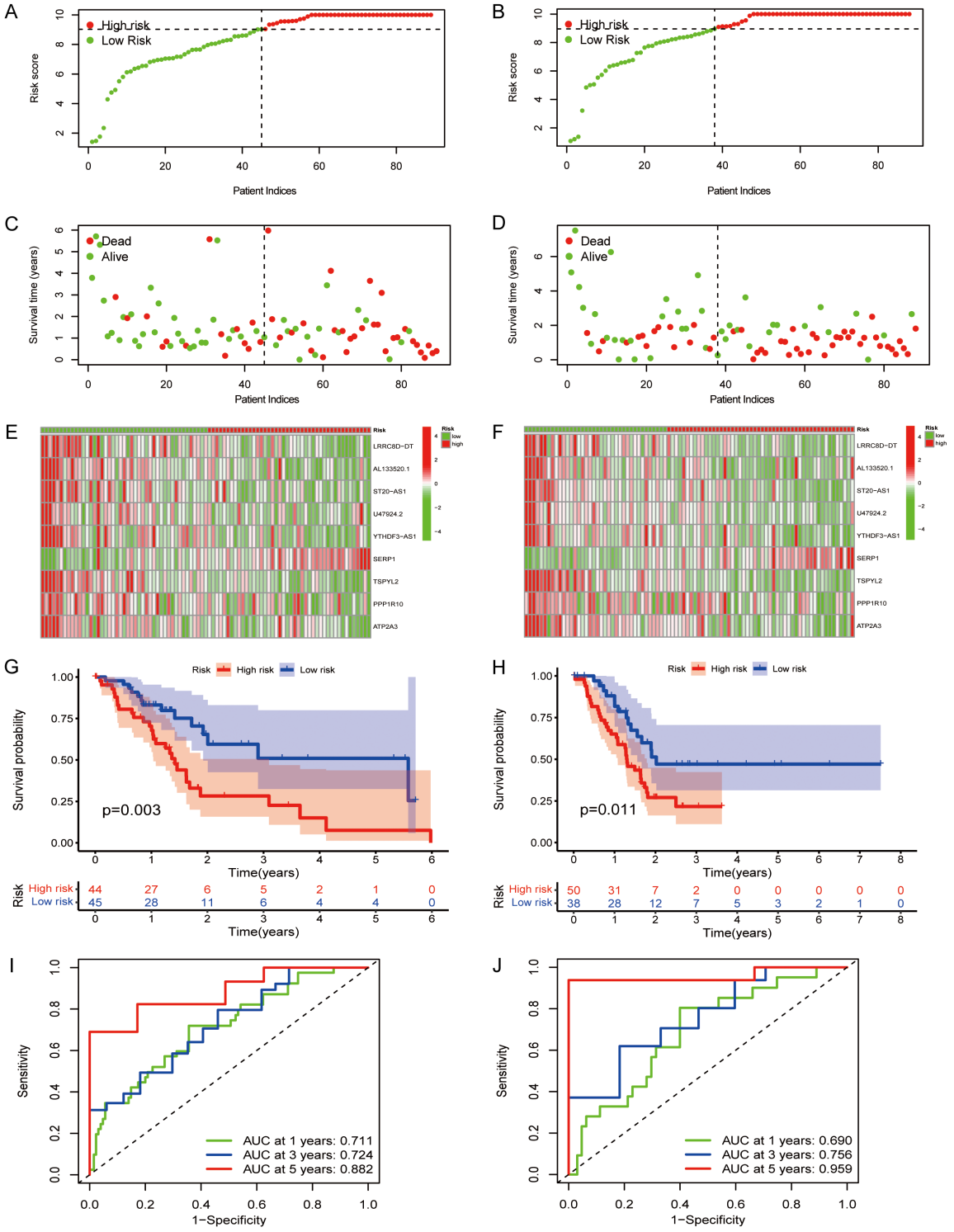


Figure 4. Correlation between the risk score and overall survival of EC patients in the training and test set. A. The distribution of risk scores in the training set. B. The distribution of risk scores in the validation set. C. The survival status of patients in the training set. D. The survival status of patients in the test set. E. Heat map of 9 genes expression in the training set. F. Heat map of 9 genes expression in the test set. G. Kaplan-Meier curves of survival in training set. H. Kaplan-Meier curves of survival in test set. I. Time-dependent receiver operating characteristic (ROC) curve of the risk score model for predicting 1, 3, and 5 years in training set. J. Time-dependent ROC curve of the risk score model for predicting 1, 3, and 5 years in test set. EC: Esophageal Cancer.

ponent (CC) category, enriched DEGs were mainly related to transport vesicle, synaptic vesicles and exocytosis vesicles. For the molecular function (MF) category, enriched DEGs were largely related to G protein-coupled receptor binding, hormone activity and neuropeptide receptor binding. KEGG pathway analysis showed that Insulin secretion, cAMP signaling pathway and dopamine synapses were significantly enriched with the DEGs (**Figure 5C**, **5D**).

In the TCGA cohort, GSEA was used to analyze pathways significantly enriched in high-risk and low-risk groups. We found that high-risk groups were significantly enriched in proteasomes, starch and sucrose metabolism and pentose phosphate pathways. We found that low-risk populations were significantly enriched in neuroactive ligand receptor interactions, calcium signaling pathways, vascular smooth muscle contraction and mTOR signaling pathways (**Figure 5E**). GSEA results suggest that these pathways may contribute to EC involvement in tumorigenesis and the progression of these ER stress-related mRNAs and lncRNAs.

Identification of independent prognostic indicators

To verify whether our prognostic model risk score could be an independent prognostic factor to predictor the prognosis of patients with EC, we performed univariate and multivariate Cox regression analyses in training set and test set. In the training set, the univariate and multivariate Cox regression analysis showed that Grade and risk score were independent prognostic factors (Supplementary Figure 2A, 2B). In the test set, independent univariate regression analysis showed Age and risk Score were independent prognostic factors (Supplementary Figure 2C, 2D). ASNS, DNAJB9 and HERPUD1 are well known transcriptional ER stress markers. We also detected the possibility that these genes could be independent prognostic factors. Interestingly, we found that these three genes were not associated with OS of EC patients from both training and test set (Supplementary Figure 2A-D). These data indicated that the risk score calculated based our prognostic model was an independent prognostic indicator in EC.

Construction of nomogram and calibration curves

In order to accurately estimate survival for individual patients with EC, we established a nomogram to evaluate the survival probability at 1, 3, and 5 years based on risk scores and other clinicopathological characteristics (**Figure 6A**). Our results demonstrated that nomograms could serve as an effective tool for the prognostic evaluation of patients with EC. Moreover, calibration curves for OS indicated that the predicted prognosis was in good agreement with the actual mortality at 1, 3, and 5 years (**Figure 6B-D**). These findings revealed that the nomogram we built could accurately assess the OS of patients with EC.

Correlation between tumor immune cell infiltration and risk score

The CIBERSORT algorithm was performed to indicate the infiltration proportion of 22 types of immune cells in the high-risk and low-risk groups. As shown in **Figure 7A** and **7B**, Barplot and heatmap showed the composition of 22 subpopulations of immune cells in the high-risk and low-risk groups. Furthermore, the proportion of TME cells in EC samples varies significantly between high-risk and low-risk groups (**Figure 7C**). The fraction of B cells naïve was higher in the low-risk group than high-risk group. On the contrary, the proportion of macrophage M1 in the low-risk group was higher than that in high-risk group (**Figure 7C**). In addition, to determine whether these significant results are caused by the separation of samples in low and high-risk groups, and not produced by random groups of patients, we performed permutation tests. The results of the permutation test showed that only B cells naive had a P value < 0.05, while macrophage M1 had a P value > 0.05 (Supplementary Table 7). The histogram of random permutation tests of mixed change proportions among naive B cells and M1 macrophages are shown in Supplementary Figure 3. These results confirmed that the proportion of B cells naive was higher in the low-risk group than high-risk group.

Correlation between tumor microenvironments and risk score

To understand the immunological evaluation value of the model, we performed quantitative analysis of TME of tumors belonging to the high- and low-risk groups, respectively. We found that the immune score, stromal score, and ESTIMATE score of the low-risk group were significantly higher than those of the high-risk group (Supplementary Figure 4A, 4C, 4E).

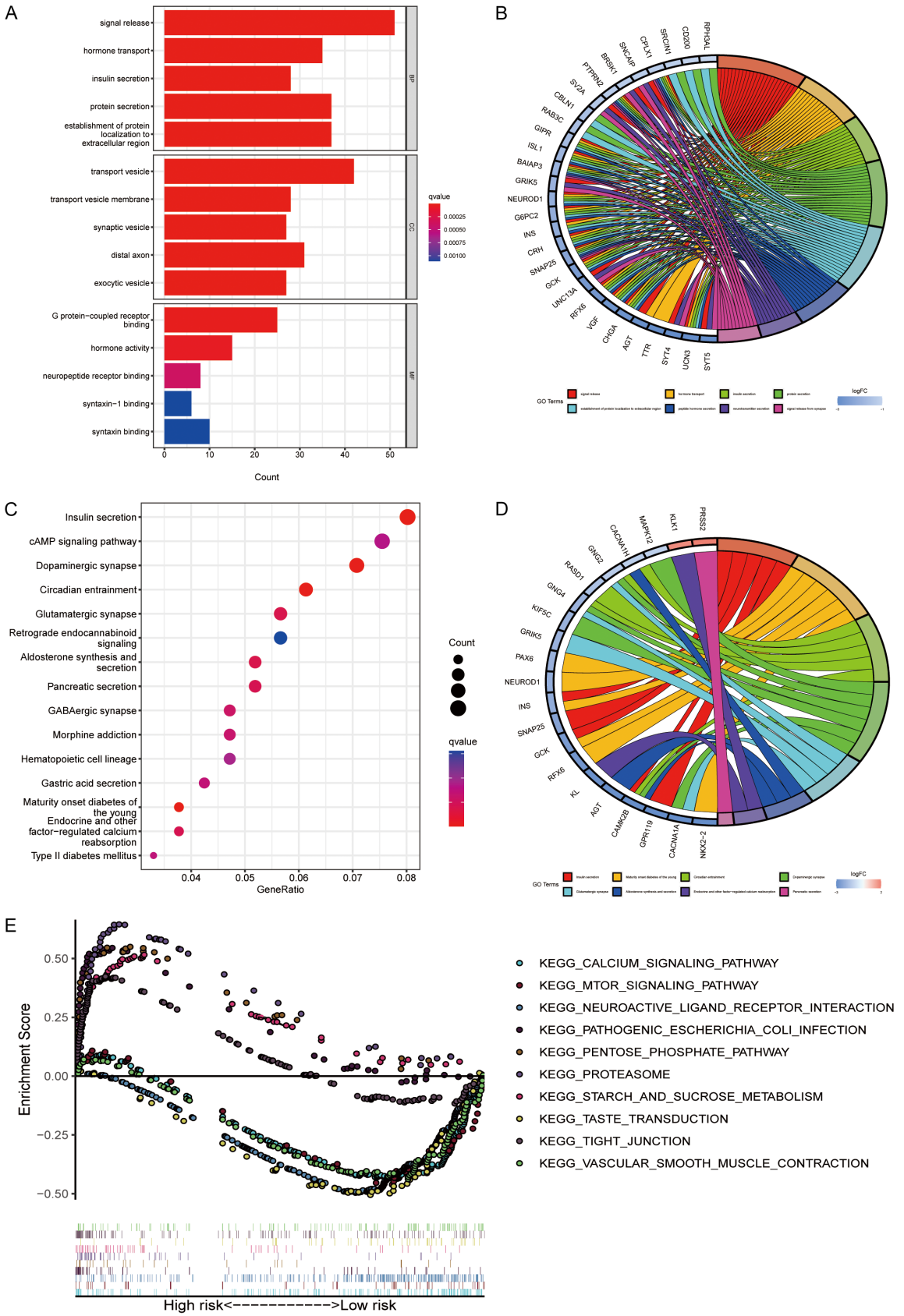


Figure 5. Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Set Enrichment Analysis (GSEA) analysis of differentially expressed genes between the high- and low-risk groups. A. The bar plot of GO

enrichment analysis. The top 5 terms were significantly enriched in GO categories for BP, CC, and MF, respectively. B. The circos plot of interconnection between GO terms and genes. C. The bubble plot for KEGG enrichment analysis. D. The circos plot of interconnection between KEGG terms and genes. E. GSEA pathway enrichment of DEGs between high and low risk groups. DEG: Differentially Expressed Gene.

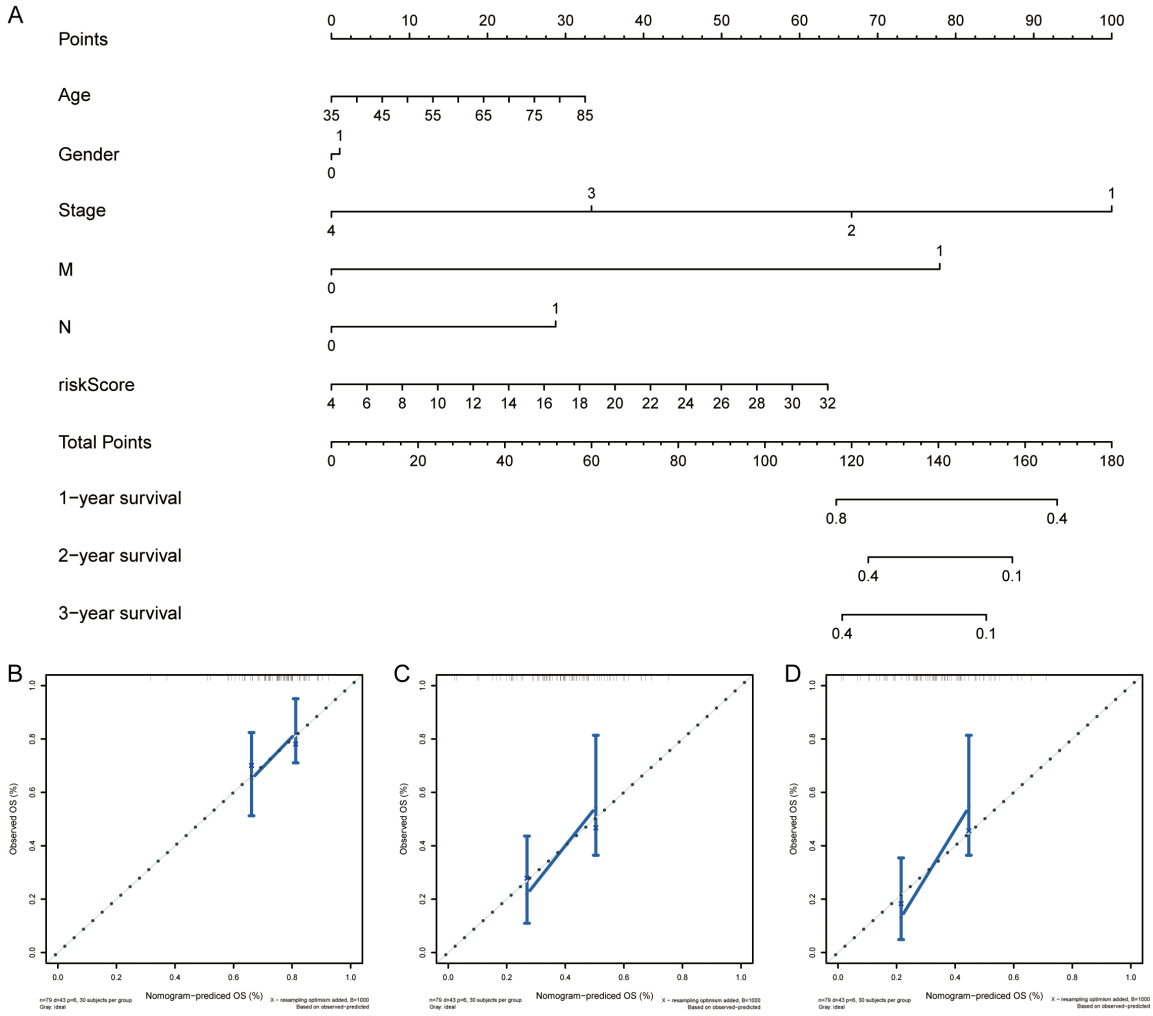


Figure 6. Establishment of the nomogram to predict overall survival of esophageal cancer patients based on TCGA cohort. A. The nomogram for predicting survival proportion of patients in 1-, 2-, and 3-year. B-D. The calibration plots for predicting patient survival at 1-, 2- and 3-years. TCGA: The Cancer Genome Atlas.

Moreover, Kaplan-Meier survival curves demonstrated that patients with a high score of immune score, stromal score or ESTIMATES cores exhibited a worse prognosis in comparison to those patients with a low score (Supplementary Figure 4B, 4D, 4F). These results reveal that the immune proportion in the tumor microenvironment is a key factor in the prognosis of patients with EC.

Assessment of response of high-risk and low-risk patients with EC to candidate drugs

To further evaluate the response of EC patients to candidate drugs in the high-risk and low-risk groups, we assessed the sensitivity score for each compound for each patient in the high-risk and low-risk groups. We identified 23 compounds with IC50 values that differed between the high-risk and low-risk groups (Supplementary Figure 5). We chose the top four (OSI.906, BI-2536, Lenalidomide, Bicalutamide) of the most divergent compounds for analysis and permutation testing of them. The results of the permutation test showed that the *P* values of the four drugs were less than 0.001 (Supplementary Table 8), demonstrating that our findings that came out of these four drugs are significant. Histograms of the permuta-

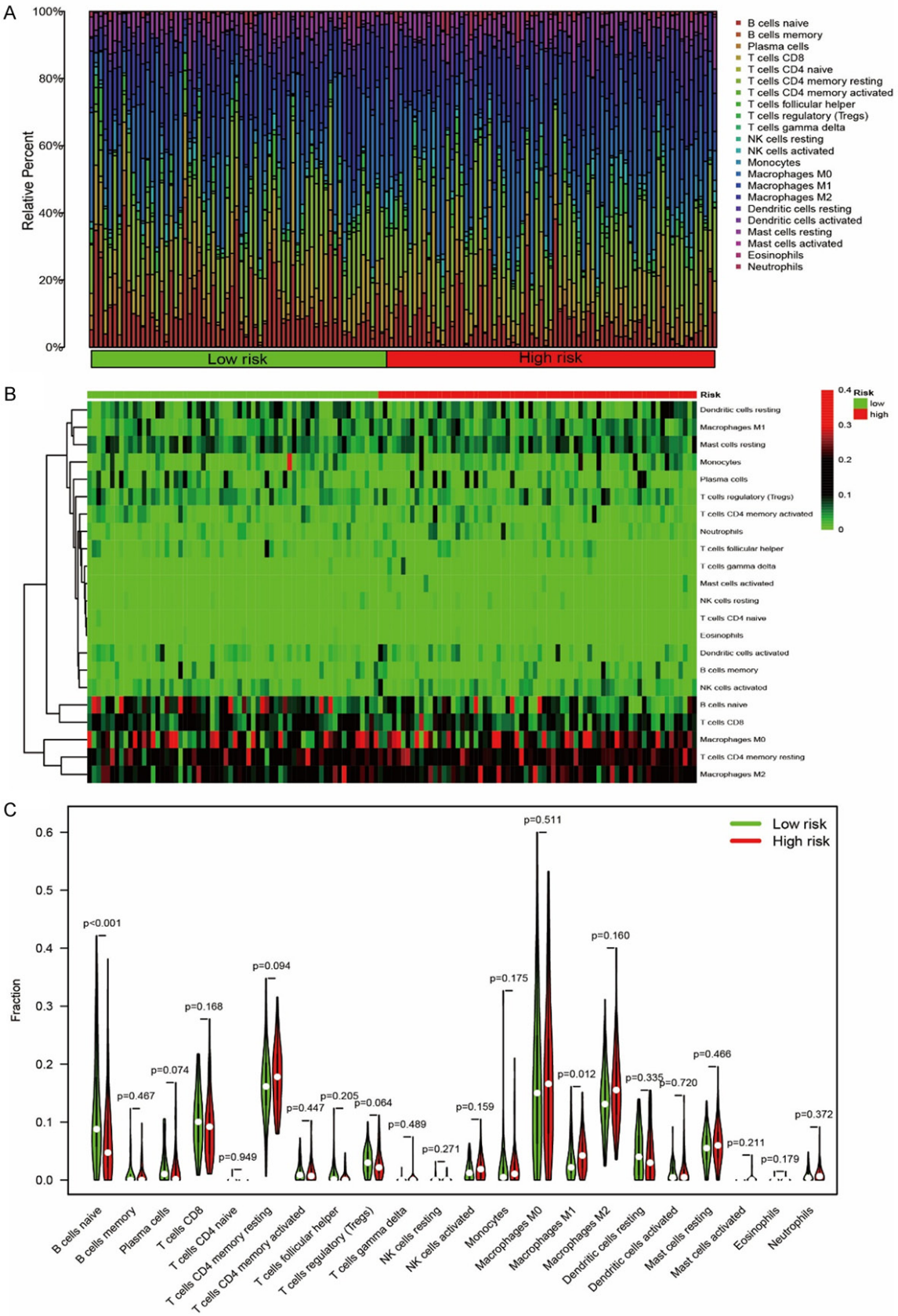

A
Relative Percent
100%
80%
60%
40%
20%
0%
Low risk
High risk
B cells naive
B cells memory
Plasma cells
T cells CD8
T cells CD4 naive
T cells CD4 memory resting
T cells CD4 memory activated
T cells follicular helper
T cells regulatory (Tregs)
T cells gamma delta
NK cells resting
NK cells activated
Monocytes
Macrophages M0
Macrophages M1
Macrophages M2
Dendritic cells resting
Dendritic cells activated
Mast cells resting
Mast cells activated
Eosinophils
Neutrophils
B
Risk
low
high
C
Low risk
High risk
Fraction
p<0.001
p=0.467
p=0.074
p=0.168
p=0.949
p=0.094
p=0.447
p=0.205
p=0.064
p=0.489
p=0.271
p=0.159
p=0.175
p=0.511
p=0.012
p=0.160
p=0.335
p=0.720
p=0.466
p=0.211
p=0.179
p=0.372

Figure 7. The immune infiltration of 22 immune cell types in high and low risk patients with esophageal cancer. A. The comparison of the proportion of immune cells infiltrating in high- and low-risk patients. B. The heatmaps plot of immune cell infiltrating in high- and low-risk groups. C. The violin plot of immune cell infiltrating in high- and low-risk patients.

tion tests for the four drugs we placed in Supplementary Figure 6. In addition, using the PubChem website, 2D conformations of the four compounds with the most significant differences in susceptibility between the high- and low-risk groups were visualized, including OSI.906 (**Figure 8A**), BI-2536 (**Figure 8B**), Lenalidomide (**Figure 8C**) and Bicalutamide (**Figure 8D**). Among them, IC50 values of BI-2536 (**Figure 8B**) and Bicalutamide (**Figure 8D**) were higher in the low-risk group than in the high-risk group, while IC50 values of OSI.906 (**Figure 8A**) and Lenalidomide (**Figure 8C**) were higher in the high-risk group than in the low-risk group. These data suggest that high-risk patients are more sensitive to BI-2536 and Bicalutamide, whereas low-risk patients are more sensitive to OSI.906 and Lenalidomide.

The expression level of SERP1 was significantly higher in EC tissues

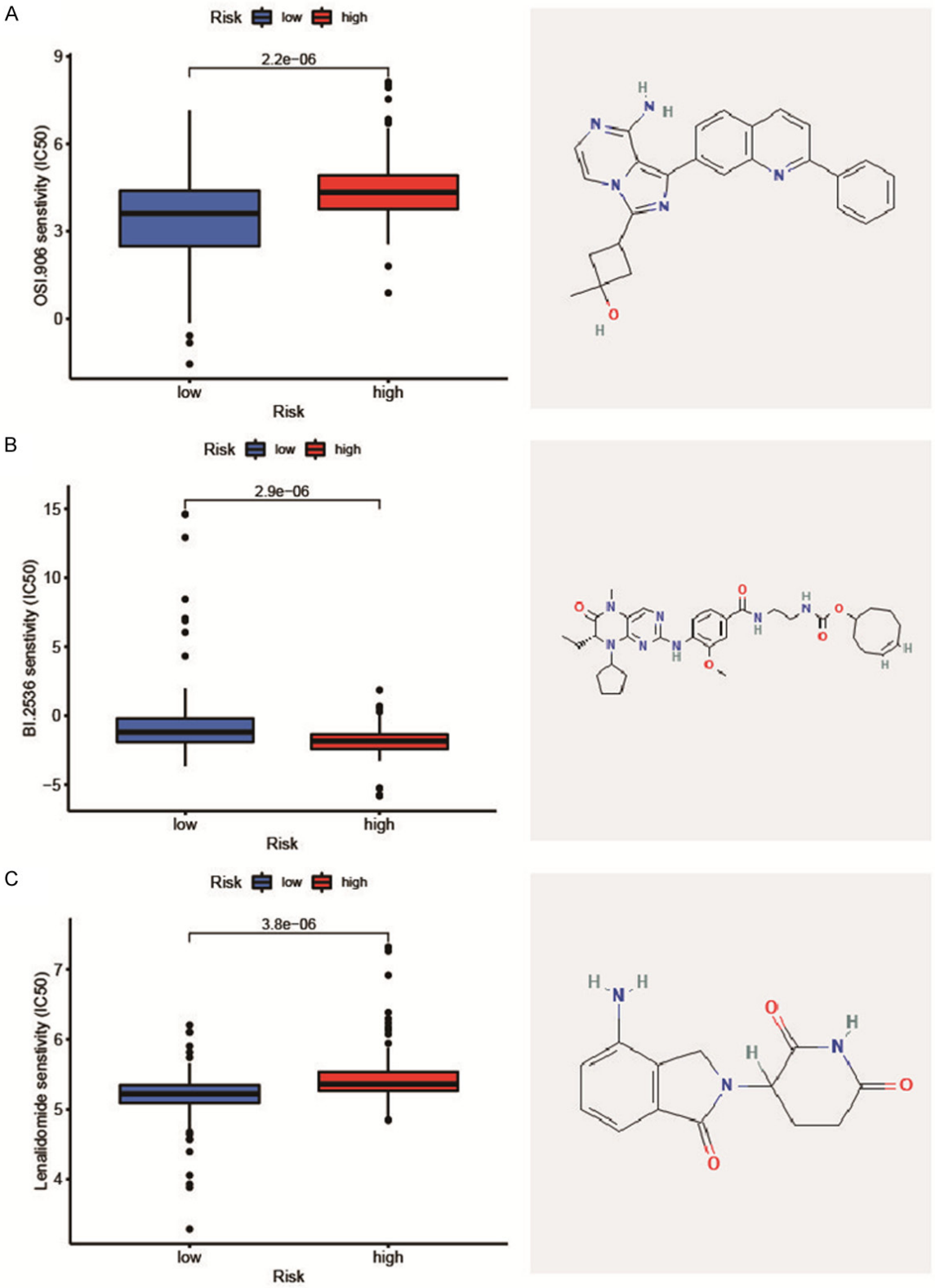
To verify the expression level of these genes used to construct a prognosis model, we evaluated SERP1 protein expression in a tumor tissue microarray (TMA) with 90 pairs of tumor and para-tumor tissues from patients with EC. High SERP1 protein expression was detected in 58.89% (53/90) of EC tissues, significantly higher than 21.11% (19/90) detected in adjacent noncancerous tissues ($\chi^2 = 26.76$, $P < 0.001$). Consistent with the mRNA transcription levels, SERP1 protein expression level was significantly elevated in tumor tissues relative to their paired non-tumor tissues (**Figure 9**).

Discussion

EC usually has a poor prognosis with a five-year survival rate of less than 20%. Therefore, construction of a robust prognostic model for predicting prognosis of patients is urgently needed. lncRNAs play an important role in many biological processes, including tumorigenesis, cellular differentiation and metabolism [40]. In recent years, these novel non-coding transcripts have gained considerable attention because of their extensive and complicated roles in cancer migration and progression [41]. It has become possible to identify tens of thousands of lncRNAs across the mammalian genome using RNA sequencing technologies [42]. The prognostic models integrating lncRNA expression profiles with coding gene expression profiles have drawn increasing attention [43-47].

ER is a highly dynamic organelle where secreted and transmembrane proteins are synthesized, folded and modified [48]. ER stress is a cellular process that occurs when several stress circumstances occur in cancer cells or tumor microenvironment, such as the accumulation of unfolded or misfolded proteins, energy or nutrient depletion and disturbances in calcium homeostasis. In order to relieve ER stress in tumors or TME, cells in the tumor and stroma will activate an adaptive response, called UPR, that aims to restore ER homeostasis [7, 48]. ER stress not only modulates tumor growth and anti-tumor immunity, but also plays an important role in cancer immunotherapy [49]. Although an EC prognostic model based on ER stress-related mRNA was established, the model integrated ER stress-related lncRNA and mRNA expression has not been reported. Therefore, it is necessary to establish a higher quality model based on the co-expression profile of ER stress-associated mRNA-lncRNA in esophageal cancer to help clinical decision-making in pursuit of individual patient care.

In this study, we developed a reliable esophageal cancer prognostic model based on ER stress-related mRNA-LncRNA co-expression profiles through univariate and LASSO regression analyses. These nine mRNA and lncRNAs were used to develop an ER stress-related prognostic model that plays an important role in progression and metastasis of many types of cancer. It has been reported that lncRNA AL133520.1 is defined as a protective effector in pancreatic cancer [50]. Additionally, another study reported lncRNA ST20-AS1 is identified as a protective gene in glioma [51]. These results are consistent with our results that AL133520.1 and ST20-AS1 have a higher


A
Risk low high
2.2e−06
OSI.906 senstivity (IC50)
9
6
3
0
low
high
Risk
B
Risk low high
2.9e−06
BI.2536 senstivity (IC50)
15
10
5
0
−5
low
high
Risk
C
Risk low high
3.8e−06
Lenalidomide senstivity (IC50)
7
6
5
4
low
high
Risk

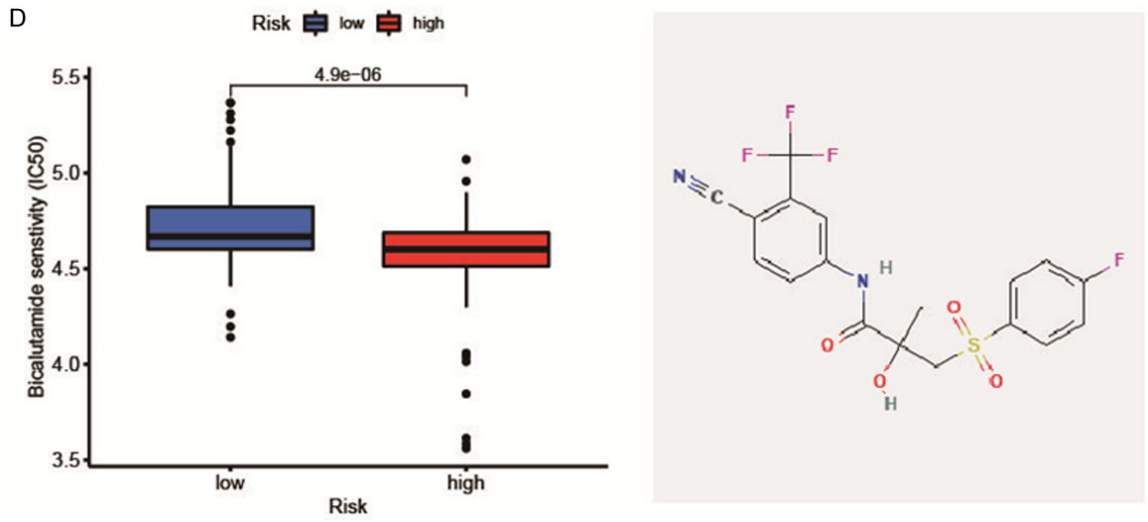


Figure 8. Drug sensitivity correlated with high- and low-risk patients in esophageal cancer. A. IC50 value of OSI.906 in high- and low-risk patients with EC. B. IC50 value of BI-2536 in high- and low-risk patients with EC. C. IC50 value of Lenalidomide in high- and low-risk patients with EC. D. IC50 value of Bicalutamide in high- and low-risk patients with EC. IC50: Half Maximal Inhibitory Concentration; EC: Esophageal Cancer.

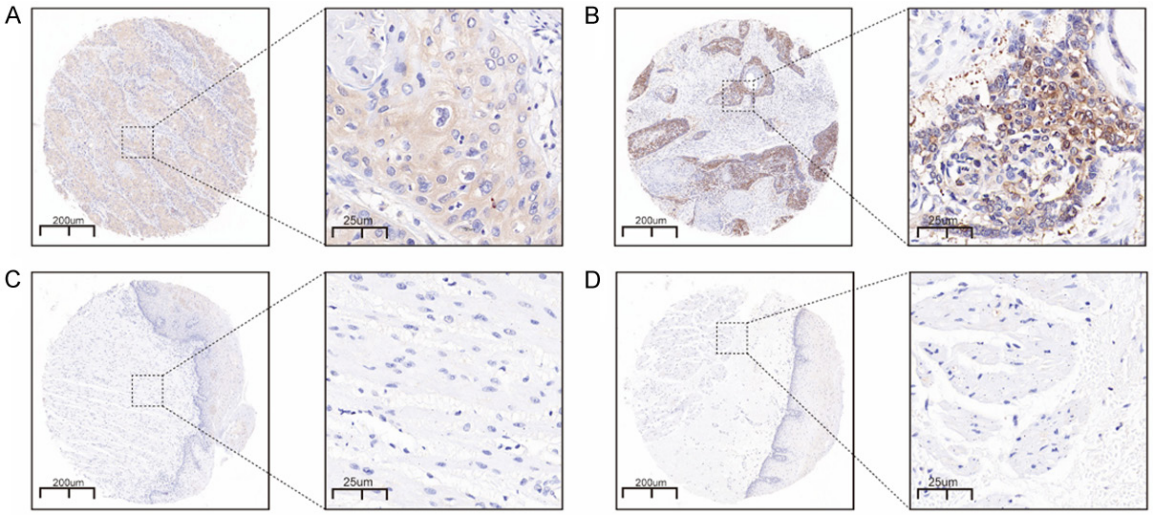


Figure 9. Representative immunohistochemistry (IHC) staining of SERP1 in EC tissue microarrays (TMA). A, B. Representative IHC staining of SERP1 in EC tissue. C, D. Representative IHC staining of SERP1 in normal esophageal tissue.

expression level in the low-risk group compared with the how-risk group, respectively. YTHDF3-AS1 was expressed highly in the high-risk group according to a model based on Ferroptosis-Related lncRNAs in breast cancer, suggesting that this lncRNA might be a high-risk oncogene [52]. However, our study showed that YTHDF3-AS1 exhibits the opposite effect in EC, suggesting that the same gene may have a completely opposite role in different cancers. The underlying mechanism needs to be studied in the future. SERP1 is a protein coding gene involved in preventing unfolded target proteins from degradation during ER stress [53]. It is suggested that SERP1 is a novel marker of poor prognosis in pancreatic ductal adenocarcinoma and glioblastoma patients [54, 55], which is consistent with our study.

Next, we screened and identified six ER stress-related hub genes: ERN1, ERN2, OS9, CLGN, ERLEC1 and DNAJB9. ERN1 (endoplasmic retic-

ulum requiring to nucleus signaling 1), also known as IRE1α (inositol-enzyme-1 α) is a transmembrane protein localized in the endoplasmic reticulum that has both protein kinase and ribonuclease activities [56]. ERN1 acts as an ER stress sensor, and it has been demonstrated that blocking the expression of ERN1 leads to reduced tumor growth by inhibiting angiogenic and pro-proliferative processes [57]. ERN2 (endoplasmic reticulum requiring to nucleus signaling 2), also known as IRE1β (inositol-requiring enzyme-1 β), is a paralog of the ER sensor ERN1/IRE1α [58]. It has been shown that inhibition of ERN2 expression promotes colorectal carcinogenesis [59]. Whereas DNAJB9 can suppress metastasis in triple-negative breast cancer by activating FBXO45 and decreasing ZEB1 levels [60]. These hub genes are likely to be new potential biomarkers and may provide new therapeutic targets for EC patients.

The tumor microenvironment, comprising immune cells and secreted chemokines, has important roles in the biological behavior of cancer [61]. Previous studies have shown that tumor cells may attract naive B cells into the tumor microenvironment, promoting tumor metastasis in breast cancer [62]. Our results showed that the fraction of naive B cells was significantly higher in the low-risk group compared with low-risk group. The function of naive B cells in EC needs to be elucidated in the future.

KEGG pathway analysis showed that differential genes between high and low risk groups were significantly enriched for cAMP signaling pathways. cAMP plays a critical role in cell signaling and regulates many physiological and pathological processes. Many studies have shown that ER stress is often accompanied by activation of the cAMP pathway [63-65]. cAMP regulates transcription of a variety of target genes primarily via protein kinase A (PKA), and aberrant cAMP-PKA signaling has been implicated in various types of human tumors [66]. It has been shown that cAMP-PKA signaling can regulate the growth, migration, invasion and metabolism of cancer cells [66, 67]. However, in esophageal cancer, the cAMP signaling pathway has not been fully investigated.

ER stress also plays an key role in promoting chemotherapeutic efficiency in various cancers [68]. There were 23 drugs with significant differences between high-risk and low-risk groups according to estimated IC50s. OSI-906 has been found to inhibit colorectal cancer growth by increasing apoptosis *in vivo* and *in vitro* [69]. BI2536 inhibits tumor growth *in vitro* by blocking polo-like kinase 1 [70]. Bicalutamide is an effective drug that has been proved by FDA for the treatment of prostate cancer [71]. However, further studies are still needed to evaluate the effectiveness of these drugs in the treatment of EC. Our results may provide new insights into the treatment of patients with EC.

Validation of the prognostic model in external test set is an important part of evaluating its reproducibility and accuracy. In this study, we used ER stress-related mRNAs and lncRNAs to construct the prognostic models, which is different from the former prognostic model in EC. Furthermore, the prognostic model based on mRNA and lncRNA may not only provide new insight into survival-related biomarkers and therapeutic targets, but also offer a more accurate prognostic prediction of EC patients and improving clinical decision-making for individualized treatment. In order to validate our model in external test sets, we have to get an extra transcriptional set containing lncRNA profiling, mRNA profiling and survival data. However, to the best of our knowledge, we haven't found a dataset meeting the demand for this to be analyzed. These transcriptional data sets (GSE179267, GSE137867, GSE103373, GSE173475, GSE130078, GSE43732, GSE53624, GSE54995, GSE54994, GSE13937, GSE21293) are all excluded either because they lack mRNA or lncRNA expression data. So we divided the TCGA set into internal train set and test set. Although the results of internal validation are not as solid as external validation, internal validation remains an effective method to verify the reproducibility of prognostic models as previous literature has mentioned [72-74]. Our model still has certain significance in offering biomarkers and therapeutic targets of EC and helping clinical decision-making.

In this study, we used univariate and LASSO regression analyses to construct the prognostic models [75-77]. Lasso regression has been widely used in the construction of prognostic model in various cancers in order to shrink multicollinearity and improve the accuracy of linear regression [78-84]. Although some mRNA or

lncRNA in our model may have certain correlations, the prognostic model we constructed still has the positive value for predicting prognosis of EC patients.

This study needs to be improved in the future for the following reasons: first, external validation is still needed to verify the reproducibility of the prognostic model. Second, more research is needed to unravel the potential role of ER stress-related genes in EC.

Conclusion

In our study, we developed a prognostic prediction model based on nine ER stress mRNA-lncRNA co-expression signatures (LRRC8D-DT, AL133520.1, ST20-AS1, U47924.2, YTHDF3-AS1, SERP1, TSPYL2, PPP1R10 and ATP2A3). Our study suggests that these nine ER stress-related mRNAs and lncRNAs are potential biomarkers for EC. Furthermore, 4 drugs closely related to the treatment of EC patients were also screened out. Our results may provide a reliable tool for clinicians to assess prognoses and make treatment decisions.

Acknowledgements

Thanks to Xinxiang University participants and staff. The Key Scientific Research Projects of Henan Colleges and Universities (Grant No. 21A310008), Natural Science Foundation for Young Scientists of Henan Province (Grant No. 222300420261) and the Natural Science Foundation of Henan Province (222300420510).

Disclosure of conflict of interest

None.

Abbreviations

EC, Esophageal cancer; ER, Endoplasmic reticulum; LncRNAs, Long noncoding RNAs; NcRNAs, Noncoding RNAs; TCGA, The Cancer Genome Atlas; MSigDB, Molecular Signature Database; WGCNA, Weighted Correlation Network Analysis; LASSO, Least Absolute Shrinkage and Selection Operator; ROC, Receiver Operating Characteristic; AUCs, Area Under the Curves; DEG, Differentially Expressed Gene; OS, Overall Survival; UPR, Unfolded Protein Response; STRING, Search Tool for the Retrieval of Interacting Genes; PPI, Protein-Protein Interaction; MCC, Maximal Clique Centrality; DMNC, Density of Maximum Neighborhood Component; CIBERSORT, Cell-type Identification by Estimating Relative Subsets of RNA Transcripts; GEPs, Gene Expression Profiles; ESTIMATE, Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; BP, Biological Process; CC, Molecular Function; MF, Cellular Component; GSEA, Gene Set Enrichment Analysis; IC50, Half Maximal Inhibitory Concentration; GDSC, Genomics of Drug Sensitivity in Cancer; TMA, Tissue Microarrays; IHC, Immunohistochemistry.

Address correspondence to: Drs. Yonghua Qi and Cheng Yan, School of Pharmacy, Key Laboratory of Nano-carbon Modified Film Technology of Henan Province, Diagnostic Laboratory of Animal Diseases, Xinxiang University, No. 191 Jinsui Avenue, Xinxiang 453000, Henan, China. E-mail: qyh@xxu.edu.cn (YHQ); yanchengxx@163.com (CY)

References

[1] Huang FL and Yu SJ. Esophageal cancer: risk factors, genetic association, and treatment. Asian J Surg 2018; 41: 210-215.

[2] Liang H, Fan JH and Qiao YL. Epidemiology, etiology, and prevention of esophageal squamous cell carcinoma in China. Cancer Biol Med 2017; 14: 33-41.

[3] Jain S and Dhingra S. Pathology of esophageal cancer and Barrett's esophagus. Ann Cardiothorac Surg 2017; 6: 99-109.

[4] Ansa BE, Coughlin SS, Alema-Mensah E and Smith SA. Evaluation of colorectal cancer incidence trends in the United States (2000-2014). J Clin Med 2018; 7: 22.

[5] Schwarz DS and Blower MD. The endoplasmic reticulum: structure, function and response to cellular signaling. Cell Mol Life Sci 2016; 73: 79-94.

[6] Cao SS, Luo KL and Shi L. Endoplasmic reticulum stress interacts with inflammation in human diseases. J Cell Physiol 2016; 231: 288-294.

[7] Mohamed E, Cao Y and Rodriguez PC. Endoplasmic reticulum stress regulates tumor growth and anti-tumor immunity: a promising opportunity for cancer immunotherapy. Cancer Immunol Immunother 2017; 66: 1069-1078.

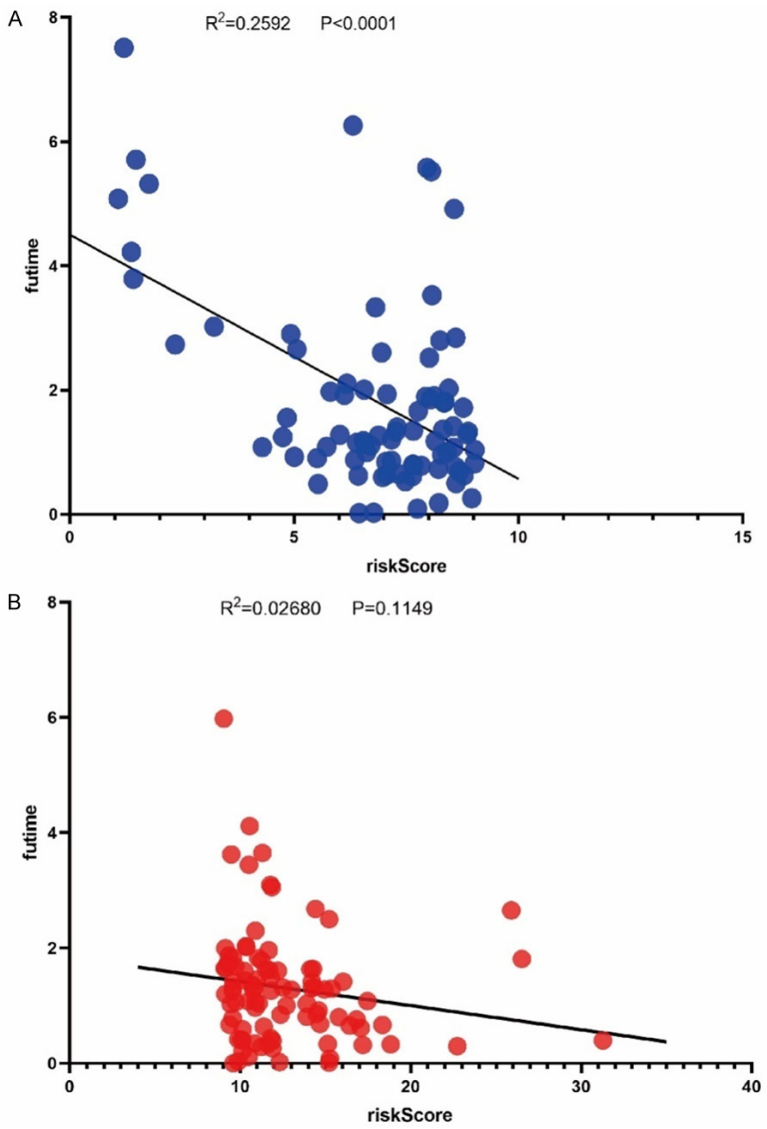
[8] Ma AG, Yu LM, Zhao H, Qin CW, Tian XY and Wang Q. PSMD4 regulates the malignancy of esophageal cancer cells by suppressing endoplasmic reticulum stress. Kaohsiung J Med Sci 2019; 35: 591-597.

[9] Huang ZL, Chen RP, Zhou XT, Zhan HL, Hu MM, Liu B, Wu GD and Wu LF. Long non-coding RNA MEG3 induces cell apoptosis in esophageal

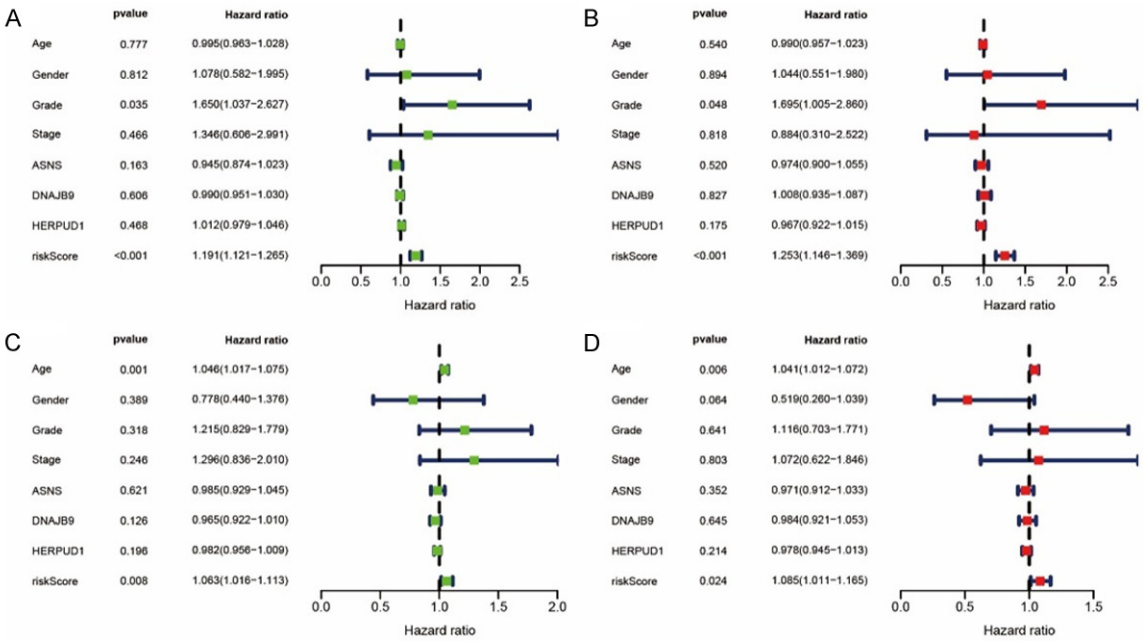
cancer through endoplasmic reticulum stress. Oncol Rep 2017; 37: 3093-3099.
[10] Zhu H, Chen X, Chen B, Chen B, Song W, Sun D and Zhao Y. Activating transcription factor 4 promotes esophageal squamous cell carcinoma invasion and metastasis in mice and is associated with poor prognosis in human patients. PLoS One 2014; 9: e103882.
[11] Zhou F, Li YH, Wang JJ, Pan J and Lu H. Endoplasmic reticulum stress could induce autophagy and apoptosis and enhance chemotherapy sensitivity in human esophageal cancer EC9706 cells by mediating PI3K/Akt/mTOR signaling pathway. Tumour Biol 2017; 39: 1010428317705748.
[12] Lu H, Gomaa A, Wang-Bishop L, Ballout F, Hu T, McDonald O, Washington MK, Livingstone AS, Wang TC, Peng D, El-Rifai W and Chen Z. Unfolded protein response is activated by aurora kinase A in esophageal adenocarcinoma. Cancers (Basel) 2022; 14: 1401.
[13] Statello L, Guo CJ, Chen LL and Huarte M. Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol 2021; 22: 96-118.
[14] Lin A, Li C, Xing Z, Hu Q, Liang K, Han L, Wang C, Hawke DH, Wang S, Zhang Y, Wei Y, Ma G, Park PK, Zhou J, Zhou Y, Hu Z, Zhou Y, Marks JR, Liang H, Hung MC, Lin C and Yang L. The LINK-A lncRNA activates normoxic HIF1α signalling in triple-negative breast cancer. Nat Cell Biol 2016; 18: 213-224.
[15] Su M, Wang H, Wang W, Wang Y, Ouyang L, Pan C, Xia L, Cao D and Liao Q. LncRNAs in DNA damage response and repair in cancer cells. Acta Biochim Biophys Sin (Shanghai) 2018; 50: 433-439.
[16] Jiang MC, Ni JJ, Cui WY, Wang BY and Zhuo W. Emerging roles of lncRNA in cancer and therapeutic opportunities. Am J Cancer Res 2019; 9: 1354-1366.
[17] Wang C, Wang L, Ding Y, Lu X, Zhang G, Yang J, Zheng H, Wang H, Jiang Y and Xu L. LncRNA structural characteristics in epigenetic regulation. Int J Mol Sci 2017; 18: 2659.
[18] Xu LJ, Yu XJ, Wei B, Hui HX, Sun Y, Dai J and Chen XF. LncRNA SNHG7 promotes the proliferation of esophageal cancer cells and inhibits its apoptosis. Eur Rev Med Pharmacol Sci 2018; 22: 2653-2661.
[19] Jiao C, Song Z, Chen J, Zhong J, Cai W, Tian S, Chen S, Yi Y and Xiao Y. lncRNA-UCA1 enhances cell proliferation through functioning as a ceRNA of Sox4 in esophageal cancer. Oncol Rep 2016; 36: 2960-2966.
[20] Zheng X, Hu H and Li S. High expression of lncRNA PVT1 promotes invasion by inducing epithelial-to-mesenchymal transition in esophageal cancer. Oncol Lett 2016; 12: 2357-2362.
[21] Gong Z, Li J, Cang P, Jiang H, Liang J and Hou Y. RPL34-AS1 functions as tumor suppressive lncRNA in esophageal cancer. Biomed Pharmacother 2019; 120: 109440.
[22] Zhang Y, Xu Y, Feng L, Li F, Sun Z, Wu T, Shi X, Li J and Li X. Comprehensive characterization of lncRNA-mRNA related ceRNA network across 12 major cancers. Oncotarget 2016; 7: 64148-64167.
[23] Zhang J, Le TD, Liu L and Li J. Inferring and analyzing module-specific lncRNA-mRNA causal regulatory networks in human cancer. Brief Bioinform 2019; 20: 1403-1419.
[24] Tomczak K, Czerwińska P and Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Pozn) 2015; 19: A68-77.
[25] Langfelder P and Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008; 9: 559.
[26] Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ and von Mering C. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015; 43: D447-D452.
[27] Chin CH, Chen SH, Wu HH, Ho CW, Ko MT and Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol 2014; 8 Suppl 4: S11.
[28] Lin CY, Chin CH, Wu HH, Chen SH, Ho CW and Ko MT. Hubba: hub objects analyzer-a framework of interactome hubs identification for network biology. Nucleic Acids Res 2008; 36: W438-W443.
[29] Therneau T and Lumley T. R survival package. R Core Team 2013.
[30] Moore DF. Applied survival analysis using R. Springer; 2016.
[31] Greenwood PE and Nikulin MS. A guide to chi-squared testing. John Wiley & Sons; 1996.
[32] Harrell FE Jr, Harrell MFE Jr and Hmisc D. Package 'rms'. Vanderbilt University 2017; 229.
[33] Chen B, Khodadoust MS, Liu CL, Newman AM and Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. In: editors. Cancer Systems Biology. Springer; 2018. pp. 243-259.
[34] Hsiung T, Olejnik S and Huberty CJ. Comment on a Wilcox test statistic for comparing means when variances are unequal. J Educ Stat 1994; 19: 111-118.
[35] Good P. Permutation tests: a practical guide to resampling methods for testing hypotheses. Springer Science & Business Media 2013.
[36] Kanehisa M and Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000; 28: 27-30.

[37] Alexa A and Rahnenführer J. Gene set enrichment analysis with topGO. Bioconductor Improv 2009; 27: 1-26.
[38] Geeleher P, Cox N and Huang RS. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. PLoS One 2014; 9: e107468.
[39] Zhou X, Li Y, Wang W, Wang S, Hou J, Zhang A, Lv B, Gao C, Yan Z, Pang D, Lu K, Ahmad NH, Wang L, Zhu J, Zhang L, Zhuang T and Li X. Regulation of Hippo/YAP signaling and esophageal squamous carcinoma progression by an E3 ubiquitin ligase PARK2. Theranostics 2020; 10: 9443-9457.
[40] Goyal N, Kesharwani D and Datta M. Lnc-ing non-coding RNAs with metabolism and diabetes: roles of lncRNAs. Cell Mol Life Sci 2018; 75: 1827-1837.
[41] Kopp F. Molecular functions and biological roles of long non-coding RNAs in human physiology and disease. J Gene Med 2019; 21: e3104.
[42] Jathar S, Kumar V, Srivastava J and Tripathi V. Technological developments in lncRNA biology. Adv Exp Med Biol 2017; 1008: 283-323.
[43] Zhang J, Liu L, Li J and Le TD. LncmiRSRN: identification and analysis of long non-coding RNA related miRNA sponge regulatory network in human cancer. Bioinformatics 2018; 34: 4232-4240.
[44] Shi YM, Li YY, Lin JY, Zheng L, Zhu YM and Huang J. The discovery of a novel eight-mRNA-lncRNA signature predicting survival of hepatocellular carcinoma patients. J Cell Biochem 2019; 120: 7539-7550.
[45] Dai W, Feng Y, Mo S, Xiang W, Li Q, Wang R, Xu Y and Cai G. Transcriptome profiling reveals an integrated mRNA-lncRNA signature with predictive value of early relapse in colon cancer. Carcinogenesis 2018; 39: 1235-1244.
[46] Zhang Y, Li Y, Wang Q, Zhang X, Wang D, Tang HC, Meng X and Ding X. Identification of an lncRNA-miRNA-mRNA interaction mechanism in breast cancer based on bioinformatic analysis. Mol Med Rep 2017; 16: 5113-5120.
[47] Mao Y, Chen R, Xia M, Guo P, Zeng F, Huang J and He M. Identification of an immune-based mRNA-lncRNA signature for overall survival in cervical squamous cell carcinoma. Future Oncol 2021; 17: 2365-2380.
[48] Chen X and Cubillos-Ruiz JR. Endoplasmic reticulum stress signals in the tumour and its microenvironment. Nat Rev Cancer 2021; 21: 71-88.
[49] Cubillos-Ruiz JR, Mohamed E and Rodriguez PC. Unfolding anti-tumor immunity: ER stress responses sculpt tolerogenic myeloid cells in cancer. J Immunother Cancer 2017; 5: 5.
[50] Wei C, Liang Q, Li X, Li H, Liu Y, Huang X, Chen X, Guo Y and Li J. Bioinformatics profiling utilized a nine immune-related long noncoding RNA signature as a prognostic target for pancreatic cancer. J Cell Biochem 2019; 120: 14916-14927.
[51] Wang W, Zhao Z, Yang F, Wang H, Wu F, Liang T, Yan X, Li J, Lan Q, Wang J and Zhao J. An immune-related lncRNA signature for patients with anaplastic gliomas. J Neurooncol 2018; 136: 263-271.
[52] Jia CL, Yang F and Li R. Prognostic model construction and immune microenvironment analysis of breast cancer based on ferroptosis-related lncRNAs. Int J Gen Med 2021; 14: 9817-9831.
[53] Yan C, Liu Q, Nie M, Hu W and Jia R. Comprehensive analysis of the immune and prognostic implication of TRIM8 in breast cancer. Front Genet 2022; 13: 835540.
[54] Ma Q, Wu X, Wu J, Liang Z and Liu T. SERP1 is a novel marker of poor prognosis in pancreatic ductal adenocarcinoma patients via anti-apoptosis and regulating SRPRB/NF-κB axis. Int J Oncol 2017; 51: 1104-1114.
[55] Mucaj V, Lee SS, Skuli N, Giannoukos DN, Qiu B, Eisinger-Mathason TK, Nakazawa MS, Shay JE, Gopal PP and Venneti S. MicroRNA-124 expression counteracts pro-survival stress responses in glioblastoma. Oncogene 2015; 34: 2204-2214.
[56] Chaurasia M, Gupta S, Das A, Dwarakanath B, Simonsen A and Sharma K. Radiation induces EIF2AK3/PERK and ERN1/IRE1 mediated pro-survival autophagy. Autophagy 2019; 15: 1391-1406.
[57] Minchenko OH, Kubaichuk KI, Minchenko DO, Kovalevska OV, Kulinich AO and Lypova NM. Molecular mechanisms of ERN1-mediated angiogenesis. International Journal of Physiology and Pathophysiology 2014; 5.
[58] Grey MJ, De Luca H, Ward DV, Kreulen IA, Gwilt KB, Foley SE, Thiagarajah JR, McCormick BA, Turner JR and Lencer WI. The epithelial-specific ER stress sensor ERN2/IRE1β enables host-microbiota crosstalk to affect colon goblet cell development. J Clin Invest 2022; 132: e153519.
[59] Jiang Y, Zhou Y, Zheng Y, Guo H, Gao L, Chen P, Feng D, Qi R, Li X, Chang Y, Chu FF and Gao Q. Expression of inositol-requiring enzyme 1β is downregulated in colorectal cancer. Oncol Lett 2017; 13: 1109-1118.
[60] Kim HY, Kim YM and Hong S. DNAJB9 suppresses the metastasis of triple-negative breast cancer by promoting FBXO45-mediated degradation of ZEB1. Cell Death Dis 2021; 12: 461.

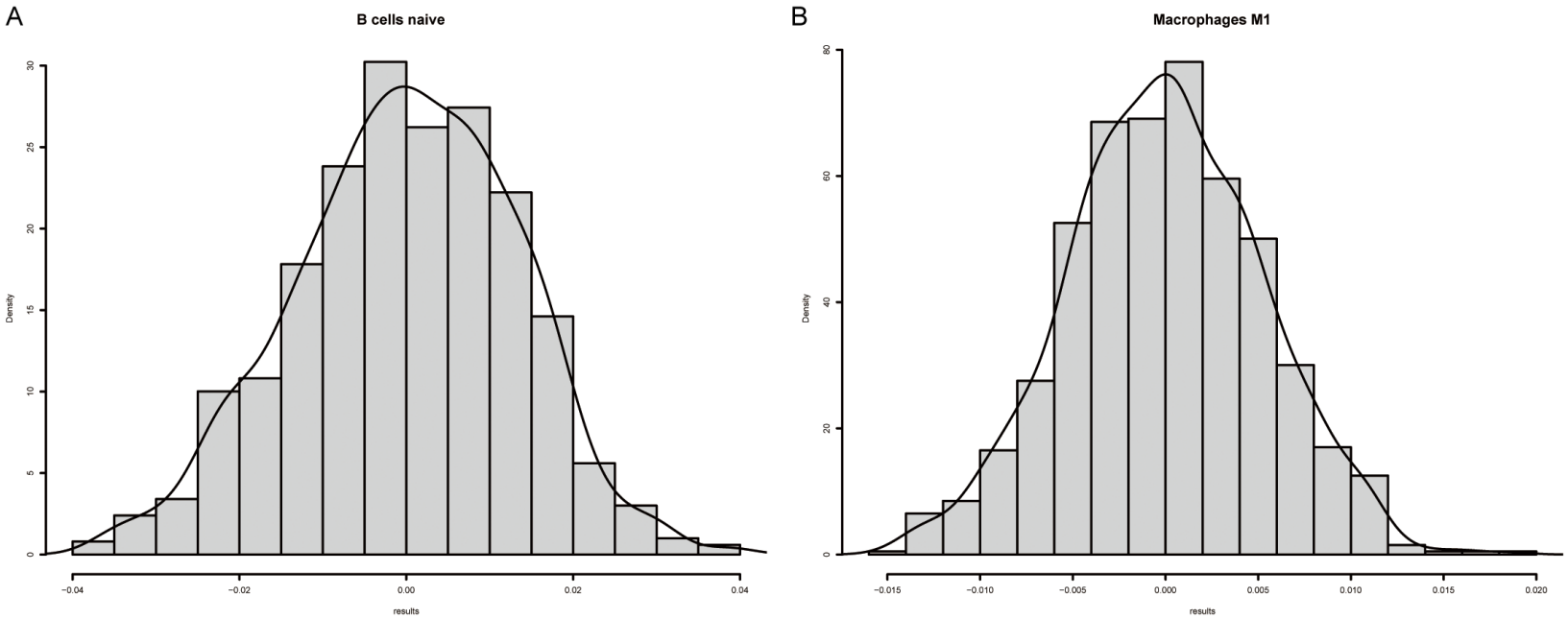
[61] Ino Y, Yamazaki-Itoh R, Shimada K, Iwasaki M, Kosuge T, Kanai Y and Hiraoka N. Immune cell infiltration as an indicator of the immune microenvironment of pancreatic cancer. Br J Cancer 2013; 108: 914-923.
[62] Olkhanud PB, Damdinsuren B, Bodogai M, Gress RE, Sen R, Wejksza K, Malchinkhuu E, Wersto RP and Biragyn A. Tumor-evoked regulatory b cells promote breast cancer metastasis by converting resting CD4+ T cells to T-regulatory cellscancer-promoting role of B cellscancer-promoting role of B cells. Cancer Res 2011; 71: 3505-3515.
[63] Song YF, Hogstrand C, Wei CC, Wu K, Pan YX and Luo Z. Endoplasmic reticulum (ER) stress and cAMP/PKA pathway mediated Zn-induced hepatic lipolysis. Environ Pollut 2017; 228: 256-264.
[64] Li J, Dou X, Li S, Zhang X, Zeng Y and Song Z. Nicotinamide ameliorates palmitate-induced ER stress in hepatocytes via cAMP/PKA/CREB pathway-dependent Sirt1 upregulation. Biochim Biophys Acta 2015; 1853: 2929-2936.
[65] Park K, Elias PM, Oda Y, Mackenzie D, Mauro T, Holleran WM and Uchida Y. Regulation of cathelicidin antimicrobial peptide expression by an endoplasmic reticulum (ER) stress signaling, vitamin D receptor-independent pathway. J Biol Chem 2011; 286: 34121-34130.
[66] Zhang H, Kong Q, Wang J, Jiang Y and Hua H. Complex roles of cAMP-PKA-CREB signaling in cancer. Exp Hematol Oncol 2020; 9: 32.
[67] Cho-Chung YS. CAMP signaling in cancer genesis and treatment. Cancer Treat Res 2003; 115: 123-143.
[68] Xu Y, Wang C and Li Z. A new strategy of promoting cisplatin chemotherapeutic efficiency by targeting endoplasmic reticulum stress. Mol Clin Oncol 2014; 2: 3-7.
[69] Leiphrakpam PD, Agarwal E, Mathiesen M, Haferbier KL, Brattain MG and Chowdhury S. In vivo analysis of insulin-like growth factor type 1 receptor humanized monoclonal antibody MK-0646 and small molecule kinase inhibitor OSI-906 in colorectal cancer. Oncol Rep 2014; 31: 87-94.
[70] Steegmaier M, Hoffmann M, Baum A, Lénárt P, Petronczki M, Krššák M, Gürtler U, Garin-Chesa P, Lieb S and Quant J. BI 2536, a potent and selective inhibitor of polo-like kinase 1, inhibits tumor growth in vivo. Curr Biol 2007; 17: 316-322.
[71] Fradet Y. Bicalutamide (Casodex®) in the treatment of prostate cancer. Expert Rev Anticancer Ther 2004; 4: 37-48.
[72] Altman DG, Vergouwe Y, Royston P and Moons KG. Prognosis and prognostic research: validating a prognostic model. BMJ 2009; 338: b605.
[73] Song M, Yang Y, He J, Yang Z, Yu S, Xie Q, Xia X, Dang Y, Zhang Q, Wu X, Cui Y, Hou B, Yu R, Xu R and Jiang T. Prognostication of chronic disorders of consciousness using brain functional networks and clinical characteristics. Elife 2018; 7: e36173.
[74] Merli M, Visco C, Spina M, Luminari S, Ferretti VV, Gotti M, Rattotti S, Fiaccadori V, Rusconi C and Targhetta C. Outcome prediction of diffuse large B-cell lymphomas associated with hepatitis C virus infection: a study on behalf of the Fondazione Italiana Linfomi. Haematologica 2014; 99: 489-96.
[75] Januaviani TMA, Gusriani N, Joebaedi K, Supian S and Subiyanto S. The best model of LASSO with the LARS (least angle regression and shrinkage) algorithm using Mallow's Cp. World Scientific News 2019; 116: 245-252.
[76] Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, Marquéz JRG, Gruber B, Lafourcade B and Leitão PJ. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 2013; 36: 27-46.
[77] Algamal ZY and Lee MH. Adjusted adaptive lasso in high-dimensional poisson regression model. Modern Applied Science 2015; 9: 170.
[78] Yan C, Liu Q and Jia R. Construction and validation of a prognostic risk model for triple-negative breast cancer based on autophagy-related genes. Front Oncol 2022; 12: 829045.
[79] Yang T, Hao L, Cui R, Liu H, Chen J, An J, Qi S and Li Z. Identification of an immune prognostic 11-gene signature for lung adenocarcinoma. PeerJ 2021; 9: e10749.
[80] Zhou X, Chi Y, Dong Z, Tao T, Zhang X, Pan W and Wang Y. A nomogram combining PPARγ expression profiles and clinical factors predicts survival in patients with hepatocellular carcinoma. Oncol Lett 2021; 21: 319.
[81] Yang F, Liu C, Zhao G, Ge L, Song Y, Chen Z, Liu Z, Hong K and Ma L. Long non-coding RNA LINC01234 regulates proliferation, migration and invasion via HIF-2α pathways in clear cell renal cell carcinoma cells. PeerJ 2020; 8: e10149.
[82] Yang F, Song Y, Ge L, Zhao G, Liu C and Ma L. Long non-coding RNAs as prognostic biomarkers in papillary renal cell carcinoma. Oncol Lett 2019; 18: 3691-3697.
[83] Huo J, Wu L and Zang Y. Development and validation of a novel metabolic-related signature predicting overall survival for pancreatic cancer. Front Genet 2021; 12: 561254.
[84] Jiang J, Bi Y, Liu XP, Yu D, Yan X, Yao J, Liu T and Li S. To construct a ceRNA regulatory network as prognostic biomarkers for bladder cancer. J Cell Mol Med 2020; 24: 5375-5386.



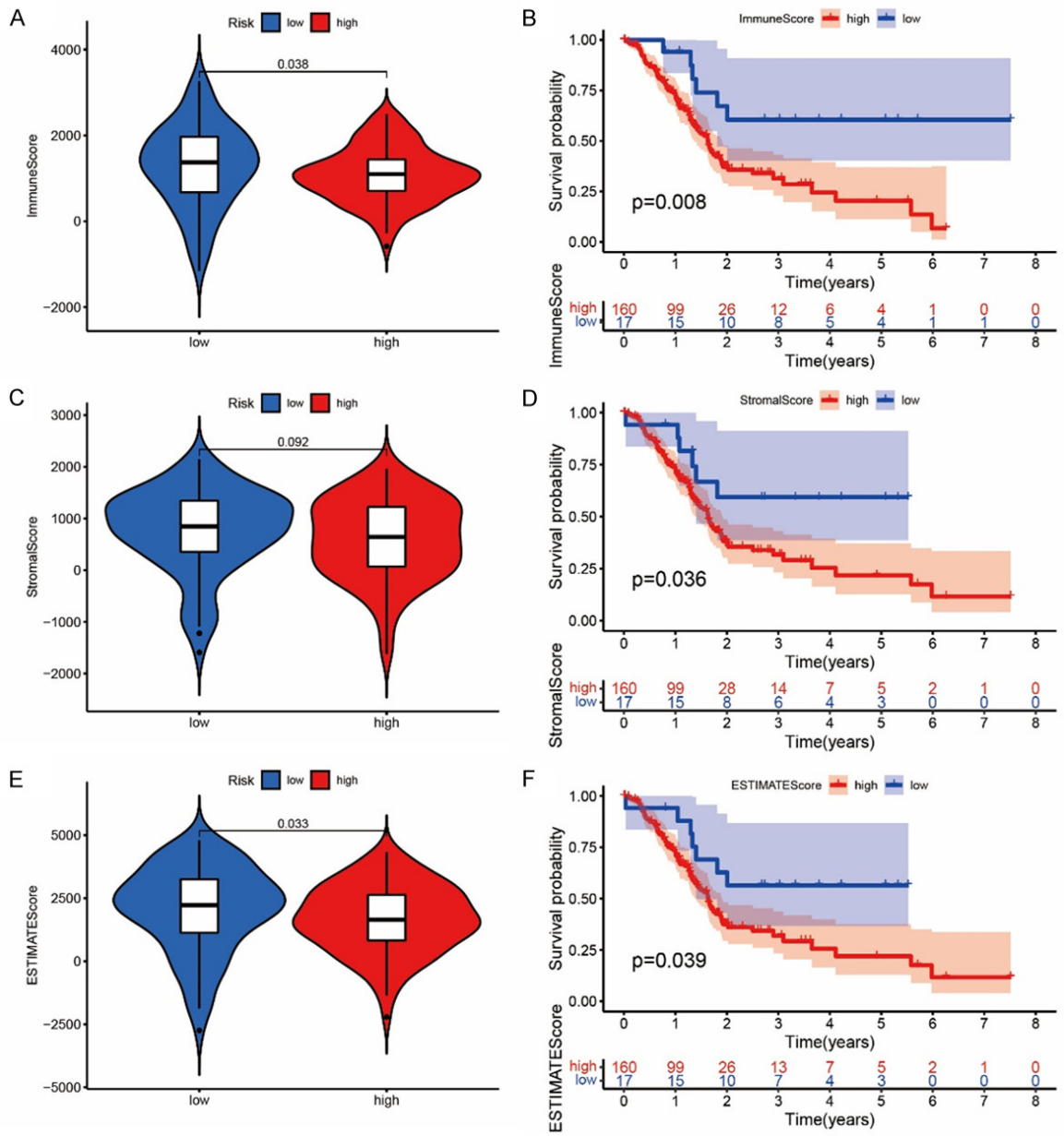
Supplementary Figure 1. Risk score and survival time distribution for low and high risk patients. Blue represents patients in low risk group (A), red represents patients in high risk group (B).



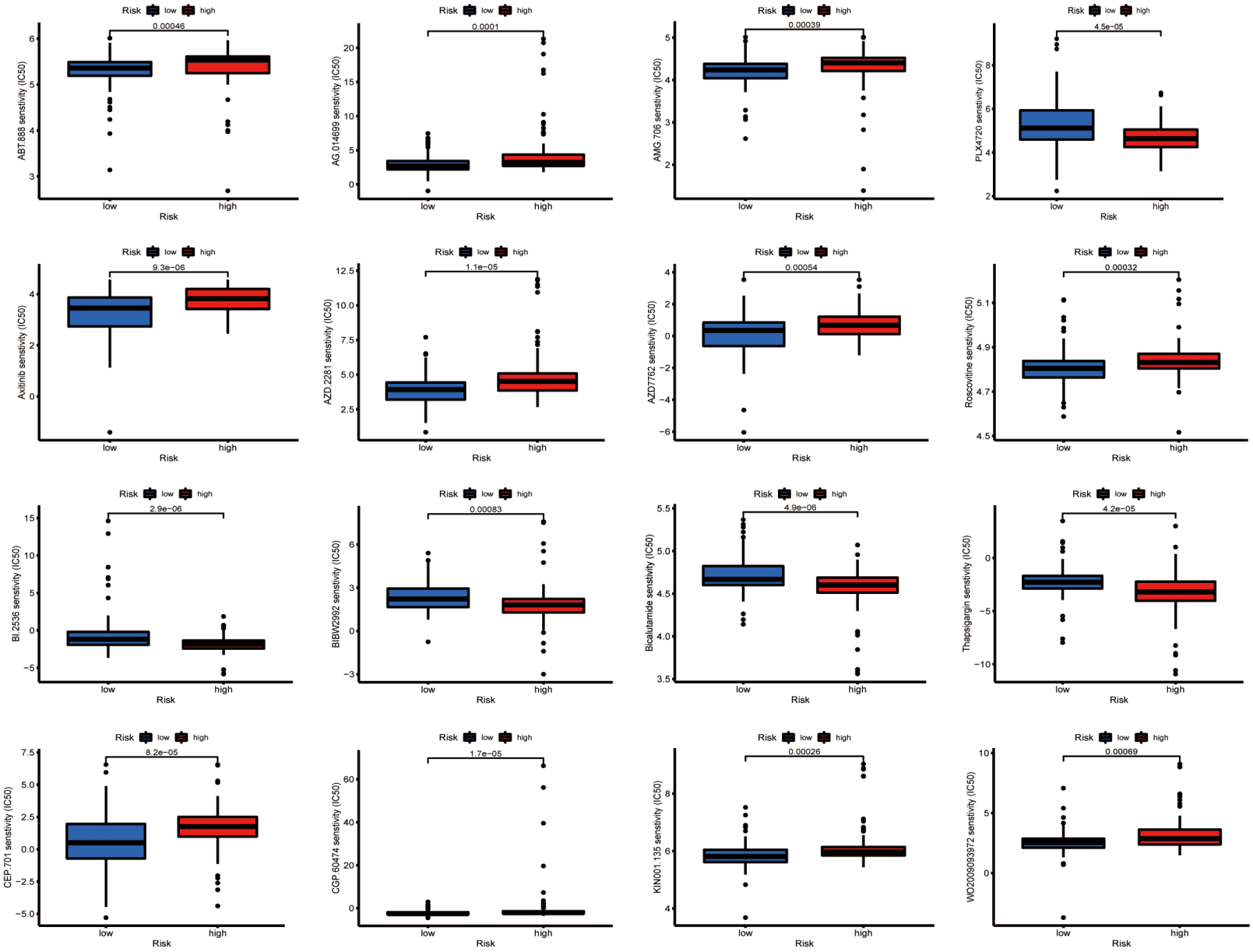
Supplementary Figure 2. An independent prognostic analysis of clinical parameters and risk scores. A. The univariate Cox regression analysis of the associations between the risk scores and clinical parameters and the OS of patients in training set. B. The multivariate Cox regression analysis of the associations between the risk scores and clinical parameters and the OS of patients in training set. C. The univariate Cox regression analysis of the associations between the risk scores and clinical parameters and the OS of patients in test set. D. The multivariate Cox regression analysis of the associations between the risk scores and clinical parameters and the OS of patients in test set. OS: Overall Survival.

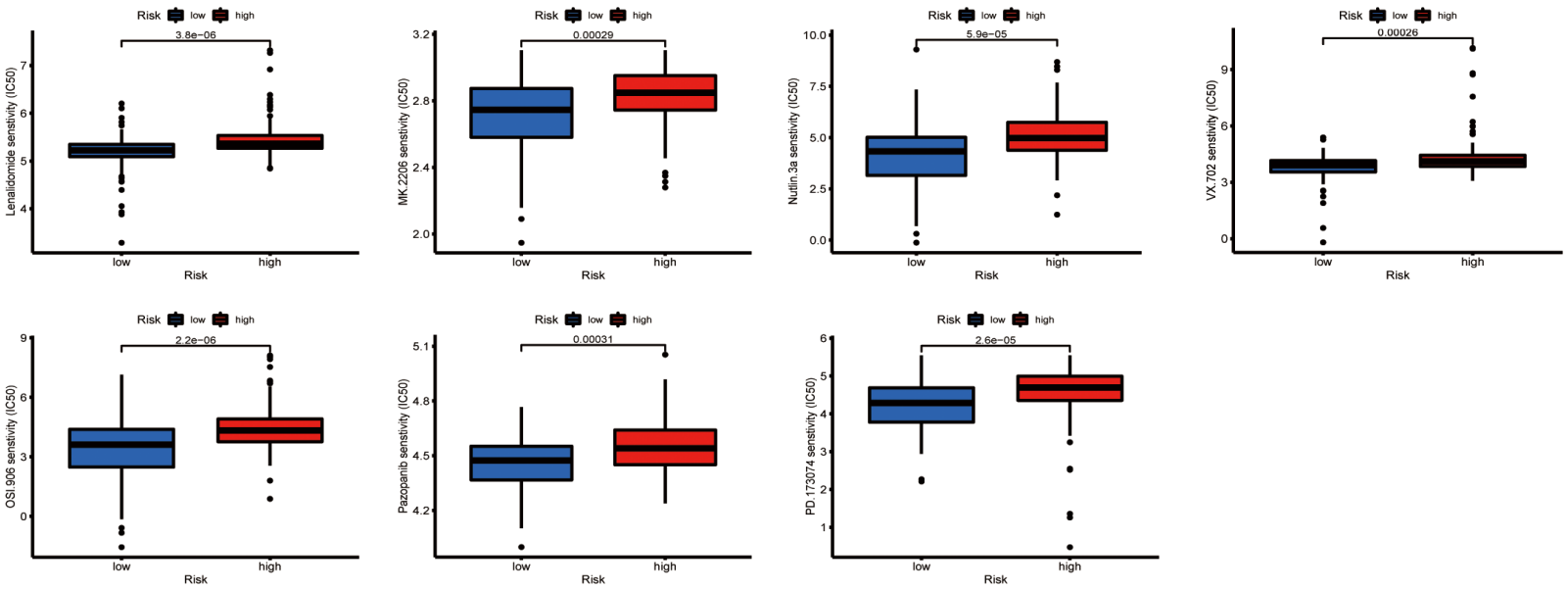


Supplementary Figure 3. The Histogram plots show the permutation test results of naive B cells (A) and macrophage M1 (B).

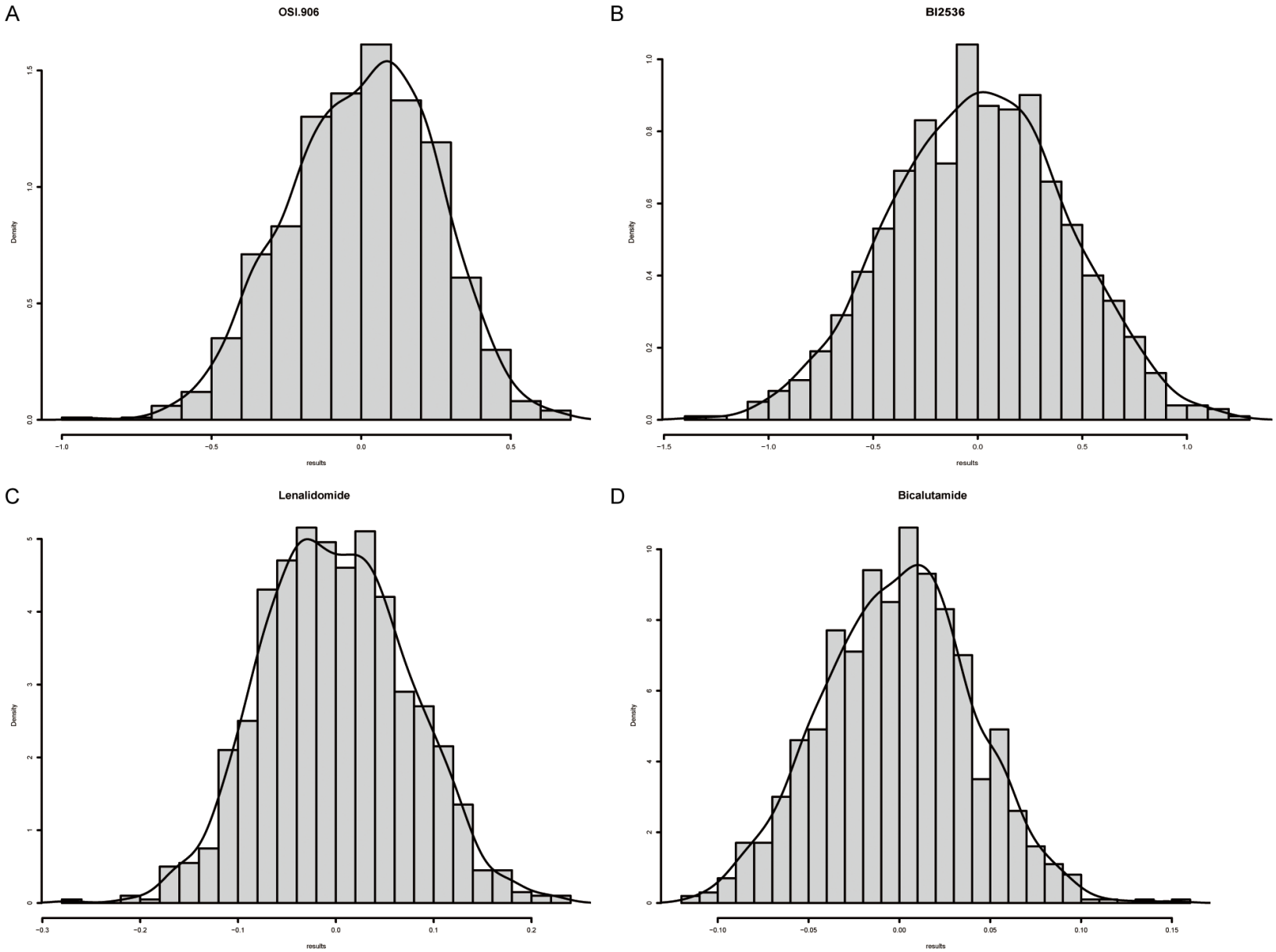


Supplementary Figure 4. The immune microenvironment analysis of the high- and low-risk groups. A. The box plot shows the difference of immune scores between the low- and high-risk group. B. Kaplan-Meier curves of OS for the high-risk and low-risk groups according to the immune scores. C. The box plot shows the difference of stromal scores between the low- and high-risk group. D. Kaplan-Meier curves of OS for the high-risk and low-risk groups according to the stromal scores. E. The box plot shows the difference of ESTIMATE scores between the low- and high-risk group. F. Kaplan-Meier curves of OS for the high-risk and low-risk groups according to the ESTIMATE scores. OS: Overall Survival.


Risk low high
0.00046
ABT.888 senstivity (IC50)
low high
Risk
Risk low high
0.0001
AG.014699 senstivity (IC50)
low high
Risk
Risk low high
0.00039
AMG.706 senstivity (IC50)
low high
Risk
Risk low high
4.5e−05
PLX4720 senstivity (IC50)
low high
Risk
Risk low high
9.3e−06
Axitinib senstivity (IC50)
low high
Risk
Risk low high
1.1e−05
AZD.2281 senstivity (IC50)
low high
Risk
Risk low high
0.00054
AZD7762 senstivity (IC50)
low high
Risk
Risk low high
0.00032
Roscovitine senstivity (IC50)
low high
Risk
Risk low high
2.9e−06
BI.2536 senstivity (IC50)
low high
Risk
Risk low high
0.00083
BIBW2992 senstivity (IC50)
low high
Risk
Risk low high
4.9e−06
Bicalutamide senstivity (IC50)
low high
Risk
Risk low high
4.2e−05
Thapsigargin senstivity (IC50)
low high
Risk
Risk low high
8.2e−05
CEP.701 senstivity (IC50)
low high
Risk
Risk low high
1.7e−05
CGP.60474 senstivity (IC50)
low high
Risk
Risk low high
0.00026
KIN001.135 senstivity (IC50)
low high
Risk
Risk low high
0.00069
WO2009093972 senstivity (IC50)
low high
Risk



Supplementary Figure 5. Drug sensitivity correlated with high- and low-risk patients in esophageal cancer.



Supplementary Figure 6. The Histogram plots show the permutation test results of OSI.906 (A), BI-2536 (B), Lenalidomide (C) and Bicalutamide (D).