Construction and Verification of a Hypoxia-Stemness-Based Gene Signature for Risk Stratification in Esophageal Cancer

Authors' Contribution:
Study Design A
Data Collection B
Statistical Analysis C
Data Interpretation D
Manuscript Preparation E
Literature Search F
Funds Collection G

BCDE 1 **Kang Tang**
AF 2 **Yong Cheng**
AF 3 **Qian Li** [iD]

1 Department of Frontier Surgery, Chiba University Graduate School of Medicine, Chiba, Japan
2 Department of Gastrointestinal Surgery, The First Affiliated Hospital of Chongqing Medical University, Chongqing, PR China
3 Department of Thoracic Surgery, The Fourth Hospital of Hebei Medical University, Shijiazhuang, Hebei, PR China

Corresponding Author: Qian Li, e-mail: liqian@hebmu.edu.cn
Financial support: None declared
Conflict of interest: None declared

Background: Numerous studies have shown that esophageal cancer (ESCA) contains areas of intertumoral hypoxia. It is widely accepted that the association of hypoxia with cancer stemness in the tumor microenvironment of ESCA is of profound clinical significance. However, reliable prognostic signatures based on hypoxia and cancer stemness are still lacking in ESCA.

Material/Methods: The t-SNE algorithm was used to estimate the hypoxia status based on the transcriptome profiles of the discovery cohort in the TCGA database. Median values of the stemness index were used to group and identify stemness-associated differentially expressed genes (DEGs). The LASSO method and Cox regression model were combined to screen for prognostic genes and to establish a genetic signature based on hypoxia-stemness. The robustness of the prognostic model was then tested in an external independent validation cohort of the GEO database.

Results: A total of 8 genes – FBLN2, IL17RB, CYP2W1, AMTN, FABP1, FOXA2, GAS1, and CTSF – were identified to construct a gene signature for ESCA risk stratification. Overall survival was significantly lower in the high-risk group than in the low-risk group in both the internal discovery set and the external validation set. The risk score was found to be an independent prognostic factor for ESCA patients. In addition, a higher risk score was significantly associated with the sensitivity of ESCA patients to gefitinib, bexarotene, dasatinib, and imatinib.

Conclusions: The hypoxia-stemness-based genetic signature established for the first time in our study could be a promising tool for ESCA cancer risk stratification.

Keywords: **Esophageal Squamous Cell Carcinoma • Neoplastic Stem Cells • Tumor Hypoxia**

Full-text PDF: https://www.medscimonit.com/abstract/index/idArt/934359

📄 4614   📊 3   📈 8   📚 46

Background

Esophageal cancer (ESCA) is one of the most commonly diagnosed cancer types and is the sixth most lethal cancer worldwide. With an increasing annual incidence and an overall 5-year survival rate of approximately 15-20%, the prognosis for patients with esophageal cancer remains poor [1]. To improve the prognosis of patients with esophageal cancer, multimodal combination therapy has been incorporated into clinical applications in recent decades. However, surgery remains the mainstay of treatment for patients with esophageal cancer. The burgeoning of adjuvant treatment has played a central role in improving survival outcomes, especially in low-stage esophageal cancer patients. Neoadjuvant chemotherapy has been shown to be superior to surgery alone, with a yield of 12-15%, thus becoming an essential component of current treatment for patients with curable esophageal cancer [2]. Unfortunately, despite advances in early screening, radical surgical resection, and recent advances in immunotherapy, esophageal cancer remains a silent disease usually diagnosed in the middle to late stages, leading to its high mortality rate. Therefore, there is an urgent need to explore the characteristics of ESCA to develop new therapeutic approaches.

As a hallmark of cancer, hypoxia is also associated with a higher capacity for metastatic progression, tumor recurrence, immune evasion, and resistance to radio-chemotherapy [3]. Since tumor hypoxia cannot be predicted based on tumor size, TNM stage, or pathological grade, there is an urgent need for validated molecular biomarkers capable of assessing ESCA hypoxia status. To the best of our knowledge, only HIF-1α and carbonic anhydrase IX (CAIX) are valid molecular markers capable of evaluating the hypoxic response in ESCA [4,5]. An early study demonstrated a significant relationship between HIF-1α expression and VEGF expression, venous infiltration, and, ultimately, clinical prognosis in ESCA [6]. During cancer development and progression, tumor cells, including CSCs, are often in a hypoxic state, and this hypoxic microenvironment maintains CSCs to acquire a more aggressive phenotype [7].

As the name implies, cancer stem cells (CSCs) are a subpopulation of cancer cells with characteristics similar to those of normal stem cells. The current widely accepted Cancer Stem Cell theory claims that a single stem cell is sufficient to initiate and generate the remaining tumor cells. The heterogeneity of specific tumor cell subpopulations with different potential for self-renewal and differentiation has led to a significant focus on the concept of cancer stem cells (CSCs) [8]. It is known that hypoxia regulates CSCs stemness by targeting the Wnt/β-catenin, Notch, Hedgehog, PI3K/AKT, and STAT3 pathways [9]. Studies have shown that esophageal cancer cells exposed to hypoxia are arrested in the G1/S phase. Since cells in the G2/M phase are most sensitive to irradiation, hypoxia-induced arrest in the

G0-G1 phase could be considered as a potential mechanism leading to radiation resistance in esophageal cancer cells [10].

The present work hypothesized that the interaction between hypoxia and stemness has some prognostic significance in ESCA. Through a series of systematic bioinformatics analyses, we developed a novel genetic signature by integrating stemness and hypoxic states followed by incorporation into the current clinicopathological signature and staging system, aimed at improving the prognosis of patients with esophageal cancer and supporting personalized treatment.

Material and Methods

Datasets

As the largest publicly available database of cancer genetic information, TCGA (https://portal.gdc.cancer.gov/) covers data including gene expression data, miRNA expression data, copy number variants (CNVs), DNA methylation, and single-nucleotide polymorphisms (SNPs). We downloaded processed esophageal cancer mRNA expression data (level 3 data) from the TCGA esophageal cancer RNA-seq dataset, including a normal group (n=11) and a tumor group (n=161). Similarly, we also downloaded the series matrix file data file of GSE53624 (n=119) from the NCBI GEO public database (http://www.ncbi.nlm.nih.gov/geo/), annotated with the platform GPL18109, and extracted the complete gene expression profiles and clinical survival information of 119 ESCA patients.

Esophageal Cancer Subtype Classification and Differentially Expressed Genes (DEGs) Analysis

A total of 480 hallmark hypoxia-related gene sets were obtained through the GeneCards database (https://www.genecards.org/) (Relevance score>5). Unsupervised nonnegative matrix factorization (NMF) [11] clustering was performed on the expression profile of hypoxia-related genes through the NMF package. We next assessed the correlation of all candidate genes with overall survival (OS) by using the "survival" package in R packages. The best cluster value is defined as the value of k at which the correlation coefficient starts to decrease. Then, based on the t-distributed Stochastic Neighbor Embedding (t-SNE) method [12], we further used the mRNA expression data of the aforementioned hypoxia genes to verify subtype assignment. Finally, we analyzed the differentially expressed genes (DEGs) between esophageal cancer subtypes through the Limma package [13]. The threshold for differential gene screening was: |logFC| >0.58 and P<0.05. In addition, TCGA esophageal cancer samples were further grouped according to the median of stemness index, and stemness-related differential genes were obtained by differential expression

analysis between the 2 groups through the Limma package. The intersection of differential genes related to hypoxia and differential genes related to stemness index yielded the differential gene of the hub (candidate DEGs).

Functional Annotation and PPI Network Construction of Candidate DEGs

To fully explore the functional relevance of candidate DEGs, the R package cluster profile (R4.0) [14] was used for functional annotation of candidate DEGs. Gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment analysis was further performed to evaluate related functional categories. The GO and KEGG enrichment terms with p and FDR q values less than 0.05 were considered significant. The STRING online database 11.0 (https://string-db.org) was used to explore the protein-protein interaction (PPI) network of candidate DEGs, and a confidence score >0.4 was used as the cut-off criterion. The generated PPI network is further visualized by Cytoscape software to explore the regulatory correlation between candidate DEGs.

Constitution of Hypoxia-Stemness Index Risk Model Based on Hub DEGs

We collect clinical information of ESCA patients based on the candidate DEGs, Cox single factor regression, and LASSO regression algorithms [15] were used in combination to screen out essential genes (Hub DEGs) in esophageal cancer to further construct prognosis-related models. Specifically, risk scores were obtained for each patient with esophageal cancer after incorporating values for each hub gene at the transcriptional level, the estimated regression coefficients derived from LASSO regression analysis were used for weighting. According to the risk score formula, the median risk score value was defined as the cut-off point, and esophageal cancer patients were further stratified into low- and high-risk subgroups. The probability of survival curves was estimated by the Kaplan-Meier estimator and any statistical difference in survival between groups was assessed by the log-rank test. The predictive performance of the models was finally evaluated by the ROC curve.

Drug Susceptibility Tests

Based on the largest publicly available pharmacogenomics database, the Genomics of Drug Sensitivity in Cancer (GDSC, www.cancerrxgene.org), we initially utilized the R software package "pRRophetic" to predicted the chemotherapy response of each esophageal cancer patient. The regression method was used to obtain the IC_{50} estimated value for each specific chemotherapy drug treatment, and we performed the tenfold cross-validation approach on the GDSC training set. All parameters were run at default values, and batch effects were corrected using ComBat.

Tumor Immune Infiltration Analysis

We combine the xCell algorithm [16] and the CIBERSORT algorithm [17] to evaluate the abundance of tumor-infiltrating immune cells. The correlation between risk score and immune cell abundance was analyzed by the Spearman correlation test. $P<0.05$ was considered as the threshold value of statistical difference.

Gene Set Variation Analysis (GSVA)

GSVA was used to transform the gene expression data from expression matrices with individual genes as features to expression matrices with specific sets of genes as features and quantified the gene enrichment results [18]. The multi-category gene set was downloaded from the molecular signatures database (MSigDB v7.0) (http://software.broadinstitute.org/gsea/msigdb/), and the GSVA algorithm converted gene-level quantification into pathway-level quantification, thereby identifying differentially expressed pathways between samples.

Statistical Analysis

The Kaplan-Meier method and log-rank tests were used for survival analysis. Multivariate analysis was performed using the Cox proportional hazard model. All statistical analyses were performed in R (version 4.0). All statistical tests were two-sided, and P values <0.05 were regarded as statistically significant.

Results

Hypoxia-Stemness-Related DEGs in ESCA Cohort

The baseline clinicopathological characteristics of the TCGA-ESCA cohort and GSE53624 dataset from GEO are shown in **Table 1**. We download the processed ESCA original mRNA expression data (FPKM) in the TCGA database. A total of 480 hypoxia-related genes were obtained through the GeneCards database (https://www.genecards.org/) (Relevance score >5), and we extracted a total of 469 hypoxia-related regulatory factor gene sets from the TCGA data. The NMF consensus clustering method was used to cluster the TCGA data set containing esophageal cancer samples based on the expression profiles of the above 469 candidate genes, and k=2 was selected as the best clustering value (**Figure 1A**). When k=2, the dimensionality of the feature was reduced by the t-SNE method, and it was found that the subtypes were mainly consistent with the two-dimensional t-SNE distribution pattern (**Figure 1B**). In addition, we observed a significant prognostic difference in the TCGA-ESCA data set. Compared with the C2 cluster, cluster C1 shown a longer median survival time ($P=0.036$) (**Figure 1C**). Finally, a total of 1308 hypoxia-related DEGs were obtained through the difference analysis between the 2 subtypes through the Limma package in R (**Figure 1D, 1F**).

Table 1. Baseline clinicopathological characteristics of patients in the TCGA-ESCA cohort and GSEA53624 cohort.

Characteristics	TCGA-ESCA cohort n (%)	GSEA53624 cohort n (%)
Gender		
Male	137 (85%)	98 (82%)
Female	24 (15%)	21 (18%)
Age		
Mean (SD)	62.29 (12.05)	59 (8.89)
T stage		
T1	27 (17%)	8 (7%)
T2	39 (24%)	20 (17%)
T3	88 (55%)	62 (52%)
T4	5 (3%)	29 (24%)
Tx	2 (1%)	–
N stage		
N0	68 (42%)	54 (45%)
N1	71 (44%)	42 (35%)
N2	10 (6%)	13 (11%)
N3	6 (4%)	10 (9%)
Nx	6 (4%)	–
M stage		–
M0	129 (80%)	–
M1	14 (9%)	–
Mx	15 (9%)	–
N/A	3 (2%)	
TNM stage		
I	17 (10%)	6 (5%)
II	74 (46%)	47 (40%)
III	56 (35%)	66 (55%)
IV	14 (9%)	–

SD – standard deviation; Tx – uncertain T stage; Nx – uncertain N stage; Mx – uncertain M stage; N/A – not applicable.

A total of 1071 DEGs related to stemness index were obtained by analyzing the difference between the 2 groups through the Limma package (**Figure 1E, 1G**). We used Venn diagrams to further screen the genes that are differentially expressed in both hypoxia subtypes and stemness index to obtain 282 genes as hypoxia-stemness-related DEGs for subsequent analysis (**Figure 1H**).

Functional Enrichment of Hypoxia-Stemness-Related DEGs and Protein–Protein Interaction Network Construction

GO and KEGG enrichment analysis was used to annotate the functions of these hypoxia-stemness-related DEGs. As shown in **Figure 2A**, GO enrichment showed significantly enriched GO biological processes (BP) terms such as extracellular matrix organization, extracellular structure organization, connective tissue development, and cartilage development. The significant GO-cellular component (CC) terms included collagen-containing extracellular matrix and cell-cell junction. For molecular function (MF) analysis, the significant terms included extracellular matrix structural constituent, growth factor binding, and cell adhesion molecule binding. KEGG pathway analysis also identified numerous signaling pathways that were significantly enriched in the candidate DEGs, for example, the AGE-RAGE signaling pathway and PI3K-Akt signaling pathway (**Figure 2B**). To analyze the role of candidate DEGs in the occurrence and development of esophageal cancer, we explored the regulatory correlation between candidate DEGs and established a PPI network based on the candidate DEGs. As shown in **Figure 2C**, according to the PPI network, the top 10 candidate DEGs with higher degrees were screened, including Fibronectin 1 (FN1), Collagen type I alpha 1 chain (COL1A1), Matrix metallopeptidase 2 (MMP2), Collagen type III alpha 1 chain (COL3A1), Collagen type I alpha 2 chain (COL1A2), Secreted protein acidic and cysteine rich (SPARC), Periostin (POSTN), Biglycan (BGN), Lysyl oxidase (LOX), and Collagen type V alpha 1 chain (COL5A1). Among these genes, FN1 had the highest node degree (82).

Construction of Hypoxia-Stemness Risk Signature

To further identify the essential genes (Hub DEGs) in the hypoxia-stemness-related DEGs sets, we extracted the clinical information of TCGA-ESCA patients, combining univariate Cox regression analysis and LASSO regression analysis to screen out essential genes (Hub DEGs) (**Figure 3A, 3B**). The results showed that a total of 14 prognosis-related genes were screened by Cox univariate regression (**Table 2**). We randomly divide TCGA-ESCA patients into a training set and a validation set at a ratio of 4: 1. As shown in **Figure 3C**, the optimal risk score value corresponding to each esophageal tumor sample was obtained by LASSO regression analysis for follow-up analyses (Risk score=FBLN2 x (-0.1731926)+IL17RB×0.01043405+CYP2W1×0.03672903+AMTN× 0.09540487+FABP1× 0.20776812+FOXA2×0.21556933+GAS1× 0.2433375+CTSF× 0.2849325). Two sets of TCGA-ESCA patients are further divided into high-risk groups and low-risk groups according to the risk score individually. Kaplan-Meier curves were used for survival analysis. As shown in **Figure 3D and 3E**, the overall survival of high-risk esophageal cancer patients in the training set (*P*=0.005) and validation set (*P*=0.041) was significantly lower than that of the low-risk group. In addition, the ROC curve analysis of the training
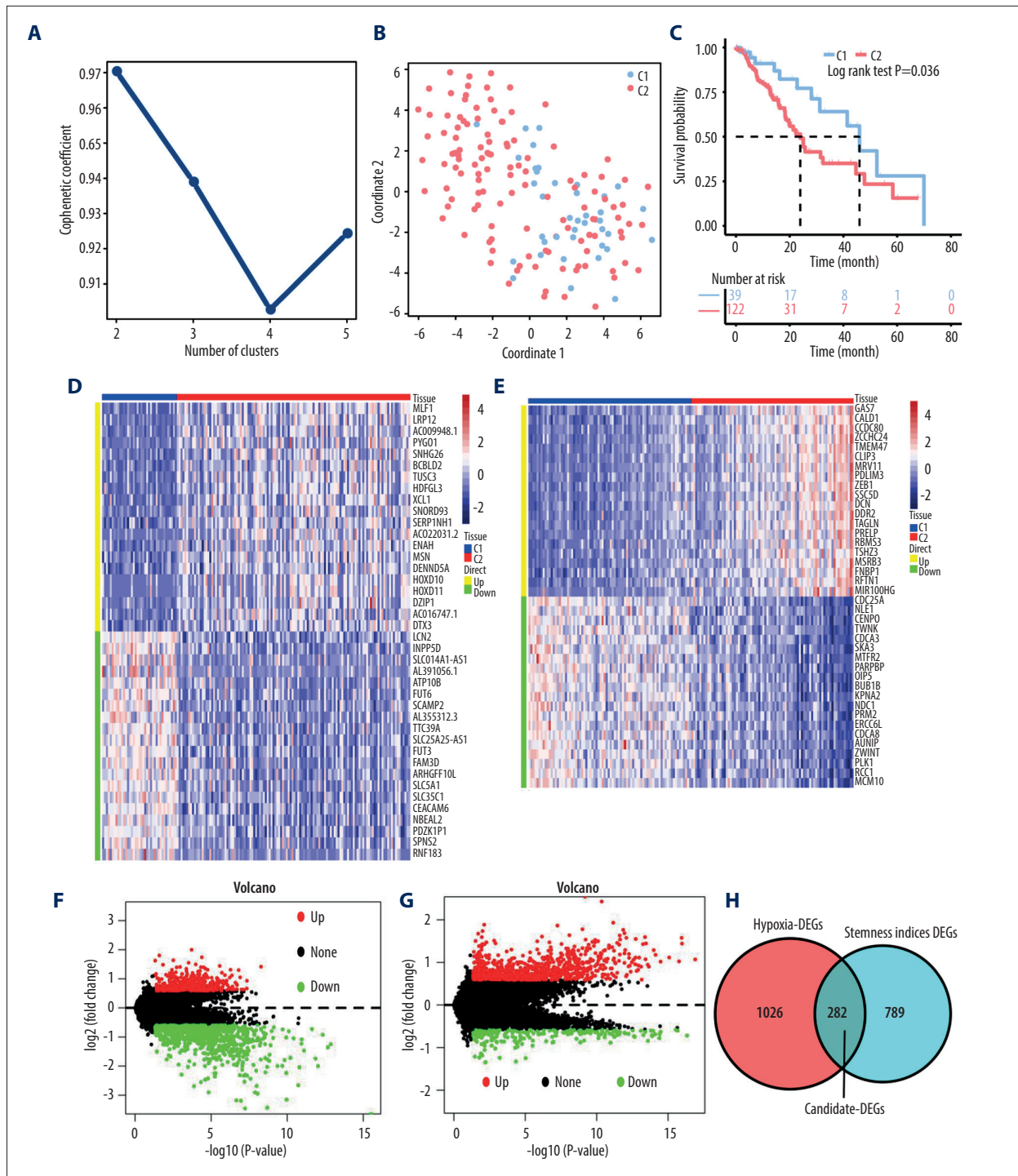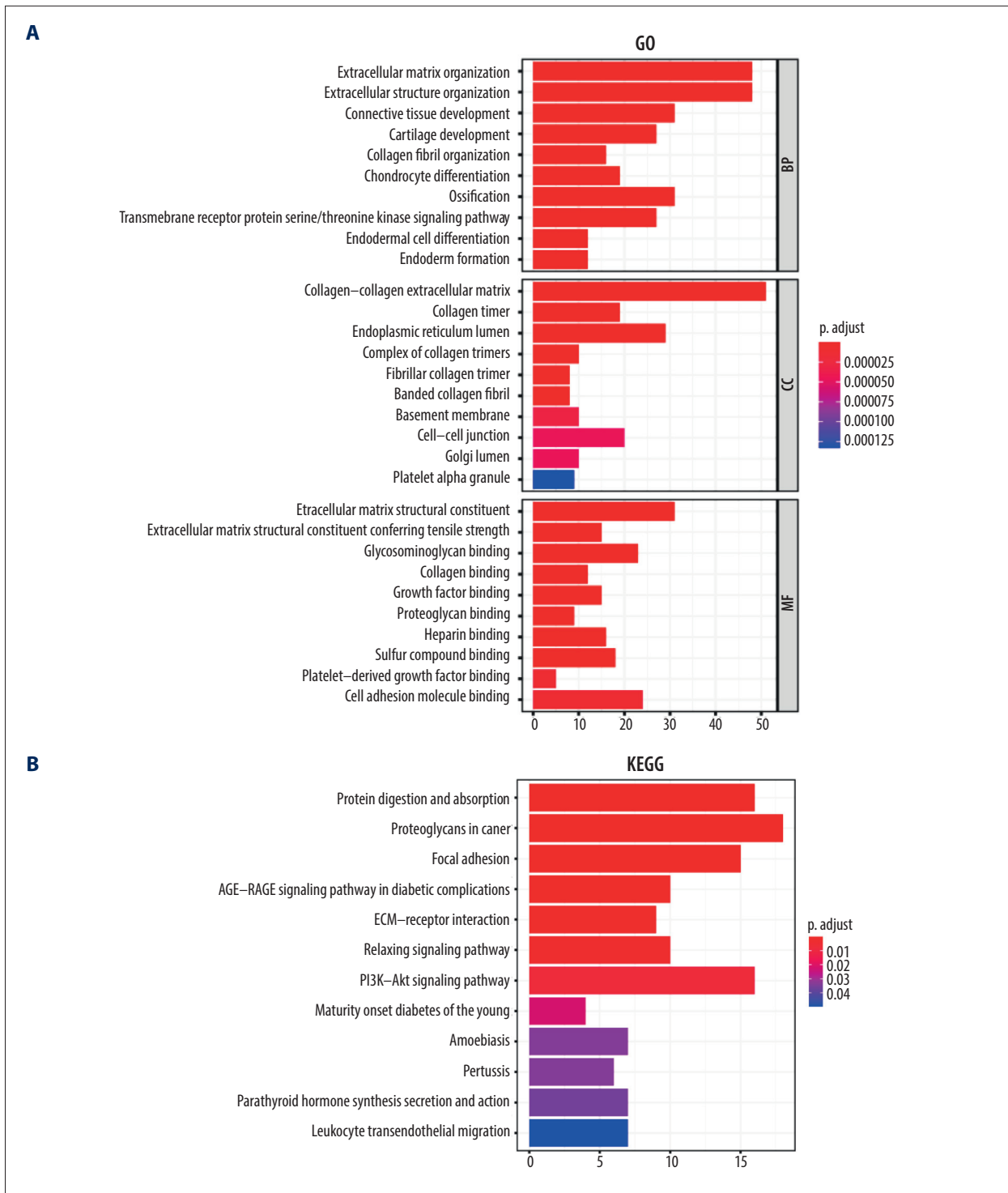
Figure 1. Identification of hypoxia- and stemness-related DEGs. (**A**) Cophenetic coefficient for clusters k=2 to 5. The figure shows that the most significant cointegration correlation coefficient occurs in cluster k=2. (**B**) Scatter plots of 2 different clusters identified by the t-SNE algorithm based on 469 hypoxic candidate genes. (**C**) Kaplan-Meier plots of overall survival for 2 hypoxia subgroups of ESCA patients. (**D**) Heat map showing the expression of hypoxia-associated DEGs and comparison between ESCA subgroups based on hypoxia. (**E**) Heatmap showing expression profiles for stemness-related DEGs with the comparison between ESCA subgroups according to the median of stemness index. (**F**) Volcano plot of hypoxia-associated DEGs gene expression. (**G**) Volcano plot of stemness-associated DEGs gene expression. (**H**) Venn diagrams show overlaps of hypoxia-related DEGs with stemness-related DEGs.
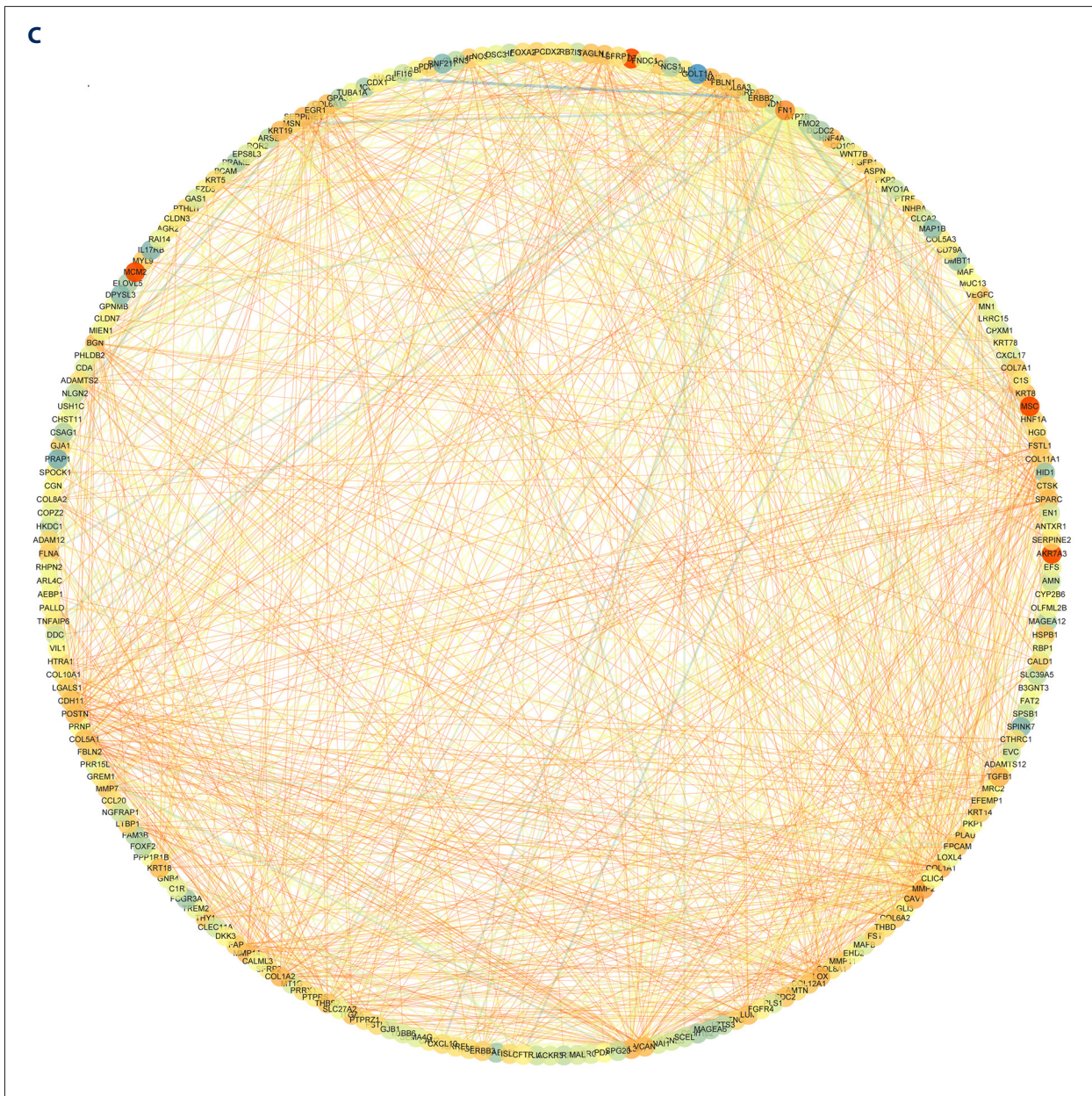
This work is licensed under Creative Common Attribution-
NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

e934359-6

Indexed in: [Current Contents/Clinical Medicine]   [SCI Expanded]   [ISI Alerting System]
[ISI Journals Master List]   [Index Medicus/MEDLINE]   [EMBASE/Excerpta Medica]
[Chemical Abstracts/CAS]

Figure 2. **Functional enrichment of hypoxia-stemness-related DEGs and construction of protein interaction networks.** (**A**) Top 10 representative gene ontology (GO) terms for the enrichment of DEGs associated with hypoxia-stemness-related DEGs. (**B**) Top 12 representative Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways for the enrichment of hypoxia-stemness-related DEGs. (**C**) Hypoxia-stemness-related DEGs of the protein-protein interaction (PPI) network. BP – biological process; CC – cellular components; MF – molecular function.

set and the validation set indicates that the model has effective verification performance (**Figure 3G, 3H**).

External Dataset Validated the Robustness of the Hypoxia-Stemness Risk Classifier

The hypoxia-stemness risk model was further validated by an external independent cohort (GSE53624). Using the risk score classifier, Kaplan-Meier analysis was used to evaluate the survival difference between groups to verify the stability of the hypoxia-stemness risk model. As shown in **Figure 3F**, similar to the findings in the discovery cohort, the external validation cohort also showed high-risk groups with poor survival and low-risk groups with better outcomes (log-rank test, P=0.021). We analyzed the ROC curve to further verify the accuracy of the risk score model (**Figure 3I**). The results showed that the risk score model had strong predictive power for the prognosis of patients with esophageal cancer.
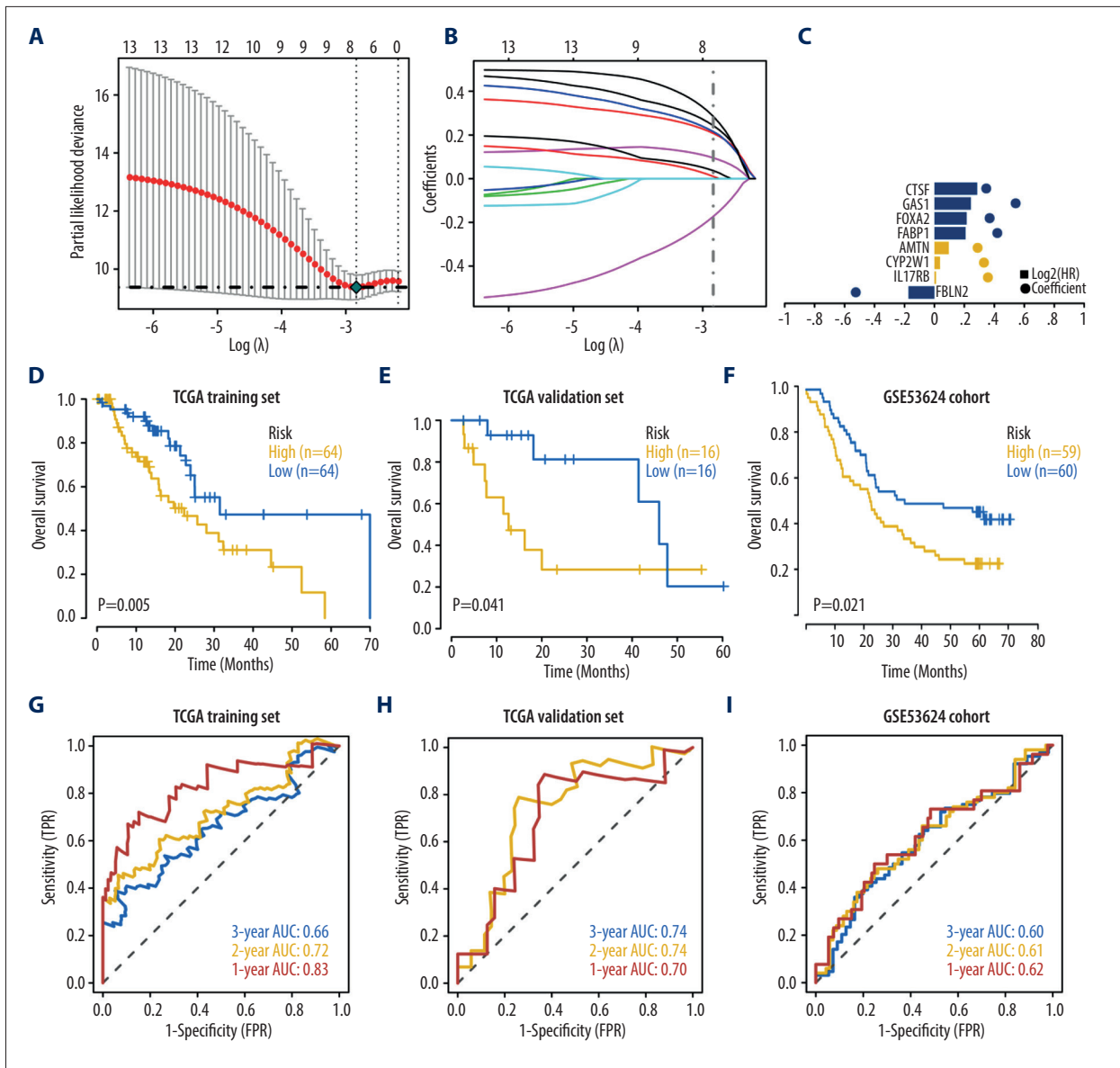
Figure 3. Construction and validation of a hypoxia-stemness-based prognostic classifier in the ESCA cohort. (**A**) Coefficients of the selected features are shown by the lambda parameter; partial likelihood deviations were derived using the LASSO Cox regression model versus log (λ). The partial likelihood deviation values are shown, and the error bars represent SE. (**B**) LASSO coefficient analysis of prognostic DEGs associated with hypoxic-stemness-related DEGs. The dotted lines in the graph represent the values selected after passing the 3-fold cross-validation. (**C**) Forest plot of hazard ratios for eight hypoxic stemness-related prognostic DEGs screened by LASSO Cox regression. (**D, E**) TCGA-ESCA patients were separated into training and validation sets in the ratio of 4: 1, Kaplan-Meier curve analysis of overall survival in high-risk and low-risk groups. (**F**) The Kaplan-Meier survival analysis was used to assess the overall survival differences between the high- and low-risk groups in the GSEA53624 cohort. (**G, H**) ROC curve analysis of TCGA-ESCA training and validation cohort. (**I**) ROC curve analysis of external independent verification cohort (GSEA53624).

Table 2. Univariate analysis of expression of 14 hypoxia-stemness base prognostic genes with overall survival.

Gene symbol	Univariate z score	HR	95% CI	*P* value
GAS1	3.024	1.449	(1.139, 1.842)	0.002
FABP1	2.973	1.331	(1.102, 1.607)	0.003
EPCAM	2.457	1.364	(1.065, 1.748)	0.014
ARSE	2.456	1.326	(1.059, 1.661)	0.014
CLDN3	2.438	1.334	(1.058, 1.682)	0.015
AMTN	2.304	1.215	(1.029, 1.433)	0.021
CYP2W1	2.294	1.251	(1.033, 1.514)	0.022
IL17RB	2.259	1.274	(1.033, 1.572)	0.024
CLRN3	2.193	1.330	(1.031, 1.717)	0.028
CLDN7	2.079	1.287	(1.015, 1.633)	0.038
FOXA2	2.017	1.284	(1.007, 1.638)	0.044
HHLA2	2.000	1.238	(1.004, 1.527)	0.046
FBLN2	-1.980	0.691	(0.480, 0.996)	0.048
CTSF	1.980	1.264	(1.002, 1.594)	0.048

HR – hazard ratio risk; CI – confidence interval.

Table 3. Univariate and multivariate Cox regression analysis of clinicopathological characteristics associated with overall survival in the TCGA-ESCA cohort.

Variables	Univariate analysis		Multivariate analysis	
	HR (95% CI)	P-value	HR (95% CI)	P-value
Gender	2.354 (0.728, 7.605)	0.153	–	–
Stage	2.671 (1.800, 3.964)	<0.001*	1.582 (0.593, 4.220)	0.359
T stage	1.257 (0.866, 1.824)	0.228	–	–
N stage	1.819 (1.294, 2.556)	<0.001*	1.580 (0.887, 2.815)	0.120
M stage	4.847 (2.211, 10.626)	<0.001*	2.551 (0.413, 15.742)	0.313
Risk score	3.346 (2.054, 5.452)	<0.001*	4.172 (2.429, 7.165)	<0.001*

HR – hazard ratio; CI – confidence interval. * Statistically significant.

Hypoxia-Stemness Risk Classifier and Clinicopathologic Characteristics in Esophageal Cancer

Univariate Cox regression analysis showed that the clinical features of tumor stage (HR=2.671, *P*<0.001), N stage (HR=1.819, P<0.001), M stage (HR=4.847, P<0.001) and risk score (HR=3.346, *P*<0.001) were significantly related to survival rate in esophageal cancer patients, while multivariate Cox regression analysis indicated risk score (HR=4.172, *P*<0.001) to be an independent risk factor of overall survival in patients with esophageal cancer (**Table 3**). Next, we divided esophageal cancer patients into different groups according to the clinicopathologic characteristics, and the corresponding risk score values were included in different groups of each sample. The results of each clinical feature subgroup are displayed in the form of box plots. The result of the Kruskal-Wallis rank sum test indicates that the distribution of risk score was significant between the 2 clinical characteristics of N stage (*P*=1.939e-02) (**Figure 4C**) and fustat (*P*=3.3939e-03) (**Figure 4E**). There were no significant relations between risk score and stage (**Figure 4A**), T stage (**Figure 4B**) or M stage (**Figure 4D**). The comprehensive results show that the risk score derived from the hypoxia-stemness model has great applicability and stability for the clinicopathologic classification of esophageal cancer patients.
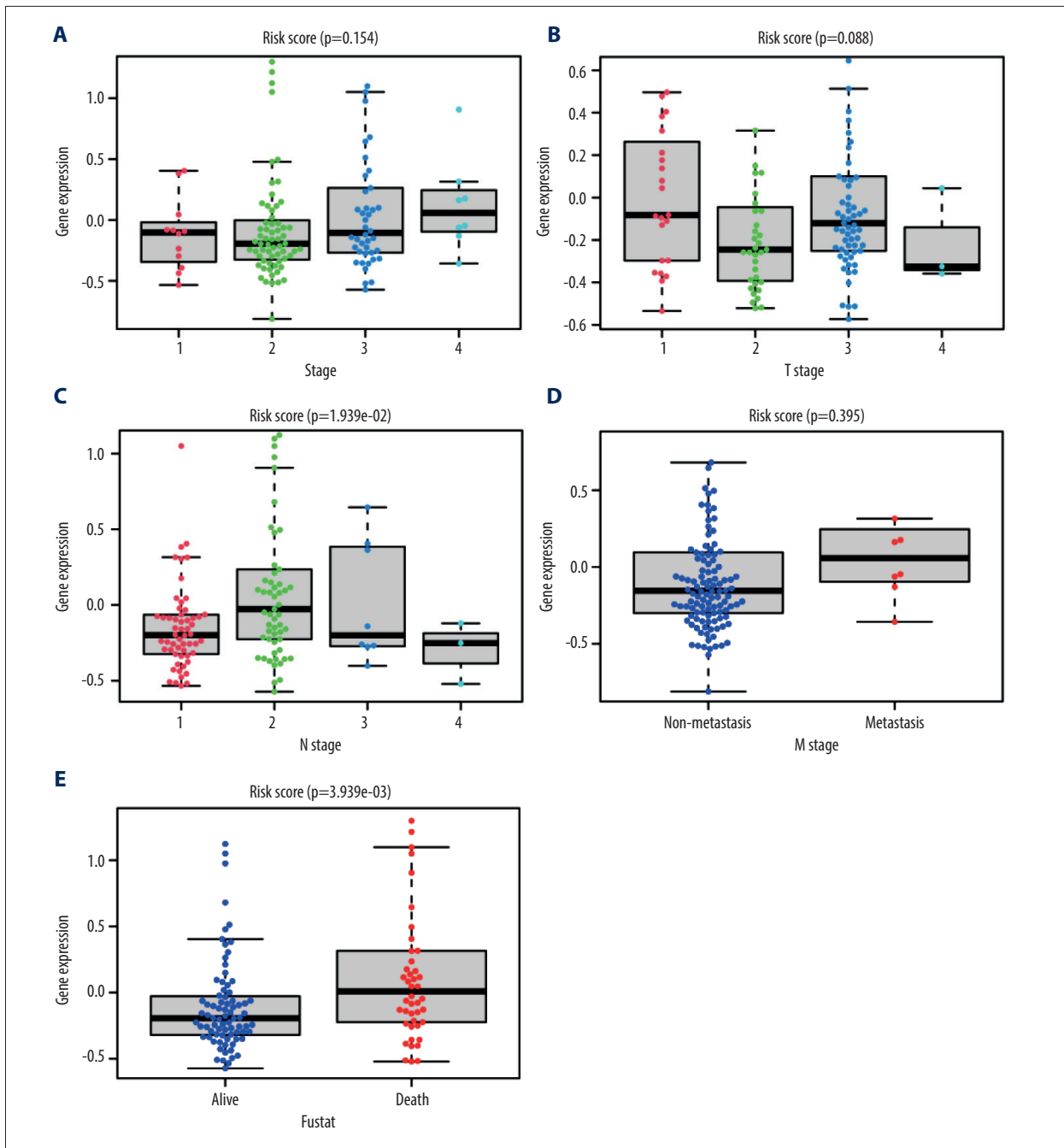
Figure 4. Correlation of hypoxia-stemness risk classifier with clinicopathological features based on the TCGA-ESCA cohort. The relationship between risk score and stage (**A**), T stage (**B**), N stage (**C**), M stage (**D**), fustat (**E**). Fustat – censoring status.

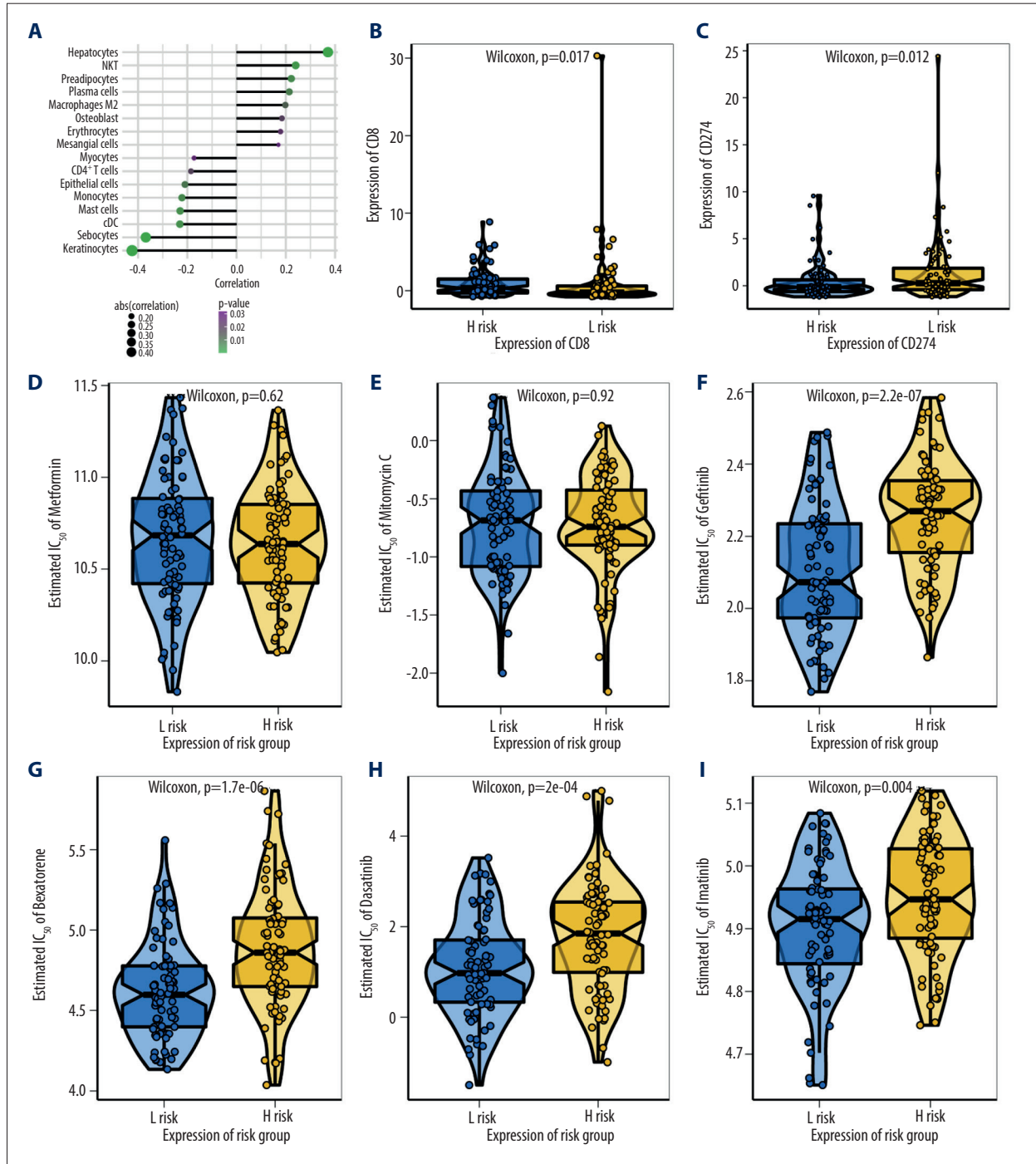## Multi-Omics Analysis Validated the Clinical Predictive Value of the Hypoxia-Stemness Risk Classifier

First, based on the transcription data, we used CIBERSORT and xCell algorithms to evaluate the distribution of tumor-infiltrating immune cells. Utilizing the xCell algorithm, the risk score was significantly positively correlated with hepatocytes, NKT, preadipocytes, plasma cells, and macrophages M2. In contrast,

a significantly negative correlation was found between the risk score and keratinocytes, sebocytes, cDC, mast cells, monocytes, epithelial cells, CD4+ T cells, and myocytes (**Figure 5A**). Meanwhile, the result of CIBERSORT was employed to generate a boxplot of immune cells in all tumor samples based on the risk score group. The risk score was significantly positively correlated with T cells regulatory (Tregs), B cells naïve, and T cells CD4 memory resting. On the other hand, there was

a significantly negative correlation between risk score and mast cells resting and dendritic cells resting (**Supplementary Figure 1A**). In addition, we also observed increased CD8 expression (**Figure 5B**) and decreased CD274(PD-L1) (**Figure 5C**) expression in high-risk groups compared to low-risk groups. However, we failed to find any significant difference between the high- and low-risk groups for tumor mutation burden (TMB) (P=0.083) (**Supplementary Figure 1B**), microsatellite instable

(MSI) (P=0.18) (**Supplementary Figure 1C**) or T cell dysfunction and exclusion (TIDE) (P=0.65) (**Supplementary Figure 1D**).

In this study, drug response data were obtained from the Genomics of Drug Sensitivity in Cancer (GDSC) dataset, and we used the R package "pRRophetic" to predict the chemotherapy sensitivity of each tumor sample. The results show that the level of risk score is significantly related to the sensitivity
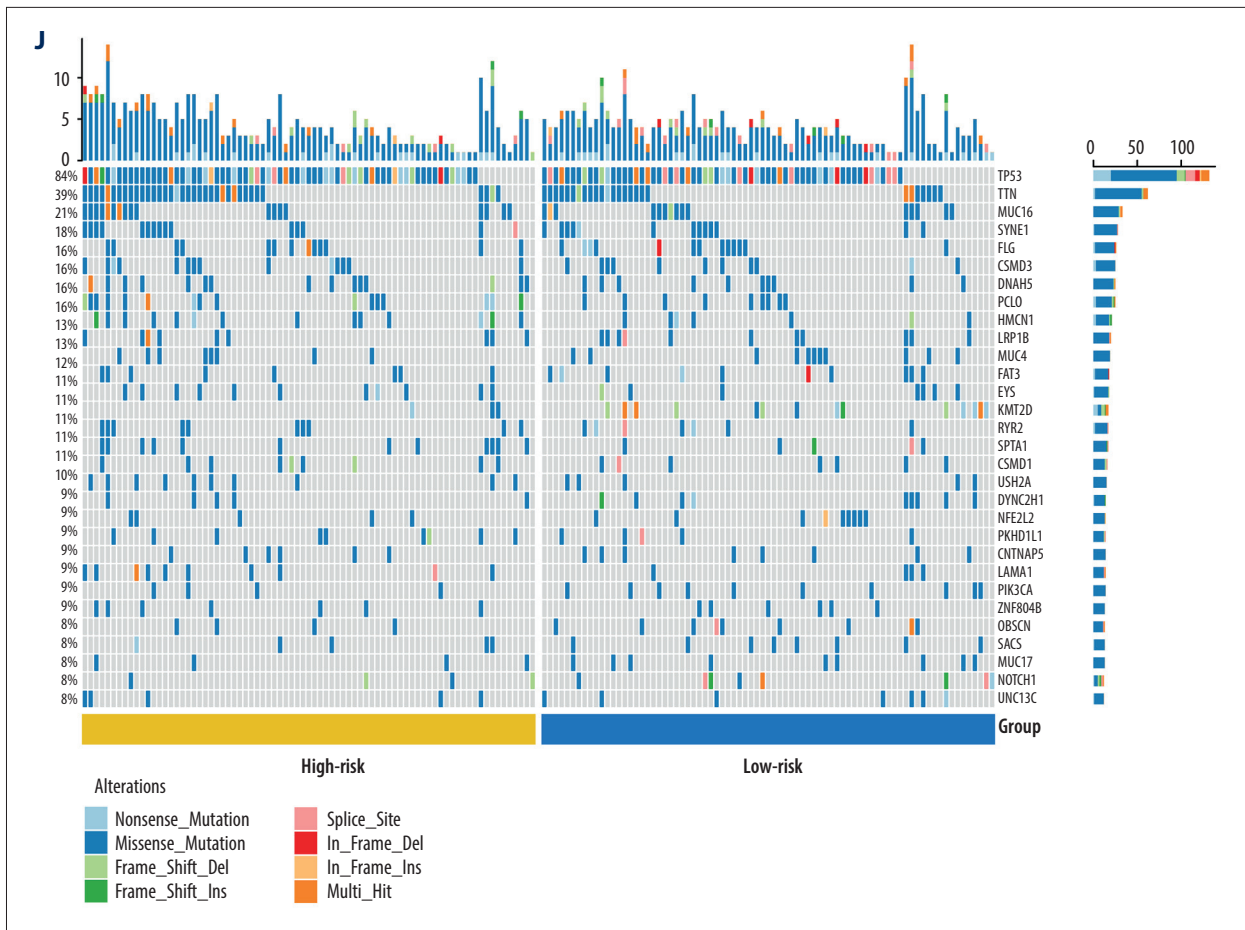
This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

e934359-11

Indexed in: [Current Contents/Clinical Medicine]   [SCI Expanded]   [ISI Alerting System]
[ISI Journals Master List]   [Index Medicus/MEDLINE]   [EMBASE/Excerpta Medica]
[Chemical Abstracts/CAS]

Figure 5. **Multi-omics study to investigate the clinical value of hypoxia-stemness risk classifier.** (**A**) Correlation of patient risk score with immune infiltration based on xCell algorithm. (**B**) CD8 expression in high- and low-risk groups. (**C**) CD274 (PD-L1) expression in high-risk and low-risk patients. (**D-I**) Comparison of common chemotherapeutic agents' sensitivity in high- and low-risk groups of ESCA patients, including Metformin (**D**), Mitomycin (**E**), Gefitinib (**F**), Bexarotene (**G**), Dasatinib. (**H**), and Imatinib (**I**). The significance was assessed by the Wilcoxon test. (**J**) Mutation mapping of ESCA patients between high and low-risk esophageal cancer patients, genes with the top 30 mutation frequencies were represented.

of patients to Gefitinib (*P*=2.2e-07) (**Figure 5F**), Bexarotene (*P*=1.7e-06) (**Figure 5G**), Dasatinib (*P*=2e-04) (**Figure 5H**), and Imatinib (*P*=0.004) (**Figure 5I**). While there was no significant relationship between risk score and Metformin (**Figure 5D**) or Mitomycin C (**Figure 5E**).

We further explored the mutation profiles of patients in the high-risk and low-risk groups. The results showed that the proportion of TP53 and other gene mutations in the high-risk group was significantly higher than that in the low-risk group (**Figure 5J**).

GSVA Analysis Revealed a Close Relationship Between the Hypoxia-Stemness Risk Classifier and Tumor Proliferation

Since the constructed hypoxia-stemness risk model is significantly related to the prognosis of patients with esophageal cancer, it is valuable to explore its potential molecular

mechanisms. GSVA is an improved GSE method, which is essentially a non-parametric and unsupervised method. Its value lies in integrating traditional analysis methods such as clustering, survival analysis, and correlation analysis into a pathway-centric analysis [18]. As shown in **Figure 6**, different pathways of esophageal cancer patients in high-risk and low-risk groups are mainly enriched in angiogenesis, G2M checkpoint, glycolysis androgen response, protein secretion, xenogamy metabolism, and P53 pathway. It is suggested that the disturbance of these signaling pathways in the high-risk and low-risk groups can activate the proliferative process and ultimately affects the prognosis of patients with esophageal cancer.
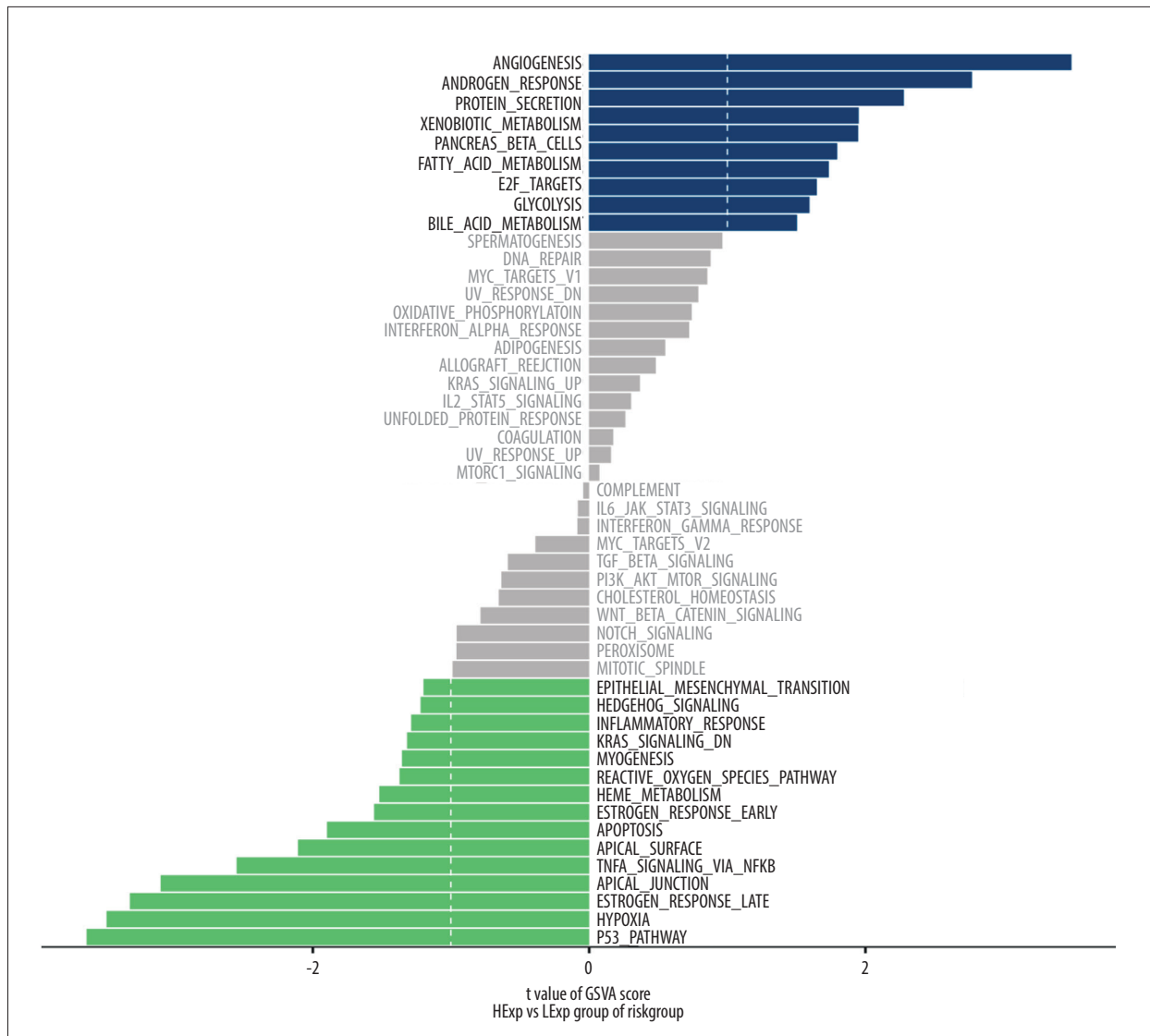
Figure 6. Heatmap of GSVA enrichment scores for the changes in the pathways in the high-risk and low-risk patients. The vertical axis represents the individual gene sets (signaling pathways in the figure), and the horizontal axis represents the degree of expression differences of each gene set in different subgroups. GSVA – gene set variation analysis.

Construction of Hypoxia-Stemness Risk Classifier-Related ceRNA Network

To construct the ceRNA network related to the prognostic model, this study analyzed the DEGs between the high- and low-risk groups and screened a total of 57 differential lncRNAs and 985 differential mRNAs. We obtained possible target genes related to 57 differential lncRNAs through multi-database predictive analysis. We extracted 354 pairs of interactions (including 7 lncRNA and 171 miRNA) between lncRNA and miRNA from the miRcode database. Then, we predicted the target mRNA based on these miRNAs and predicted a total of 1790 pairs of miRNA-mRNA interactions (including 35 miRNAs and 1179 mRNAs). After intersecting with 958 differential

mRNAs, a ceRNA network was constructed through Cytoscape, including 158 pairs of lncRNA-miRNA-mRNA interactions (6 lncRNA, 23 miRNAs, and 47 mRNA) (**Figure 7A**). In addition, 47 target mRNAs were used in the Metascape online tool for functional enrichment analysis. As shown in **Figure 7B-7D**, we identified that these genes were significantly enriched in developmental growth, modified amino acid transport, morphogenesis of epithelium, negative regulation of cell population proliferation, negative regulation of growth, wound healing, actin filament-based process, regulation of insulin secretion, mononuclear cell differentiation, gland development, glucose homeostasis, and sensory organ morphogenesis (GO Biological Processes); SUMO E3 ligases SUMOylate target proteins (Reactome Gene Sets).
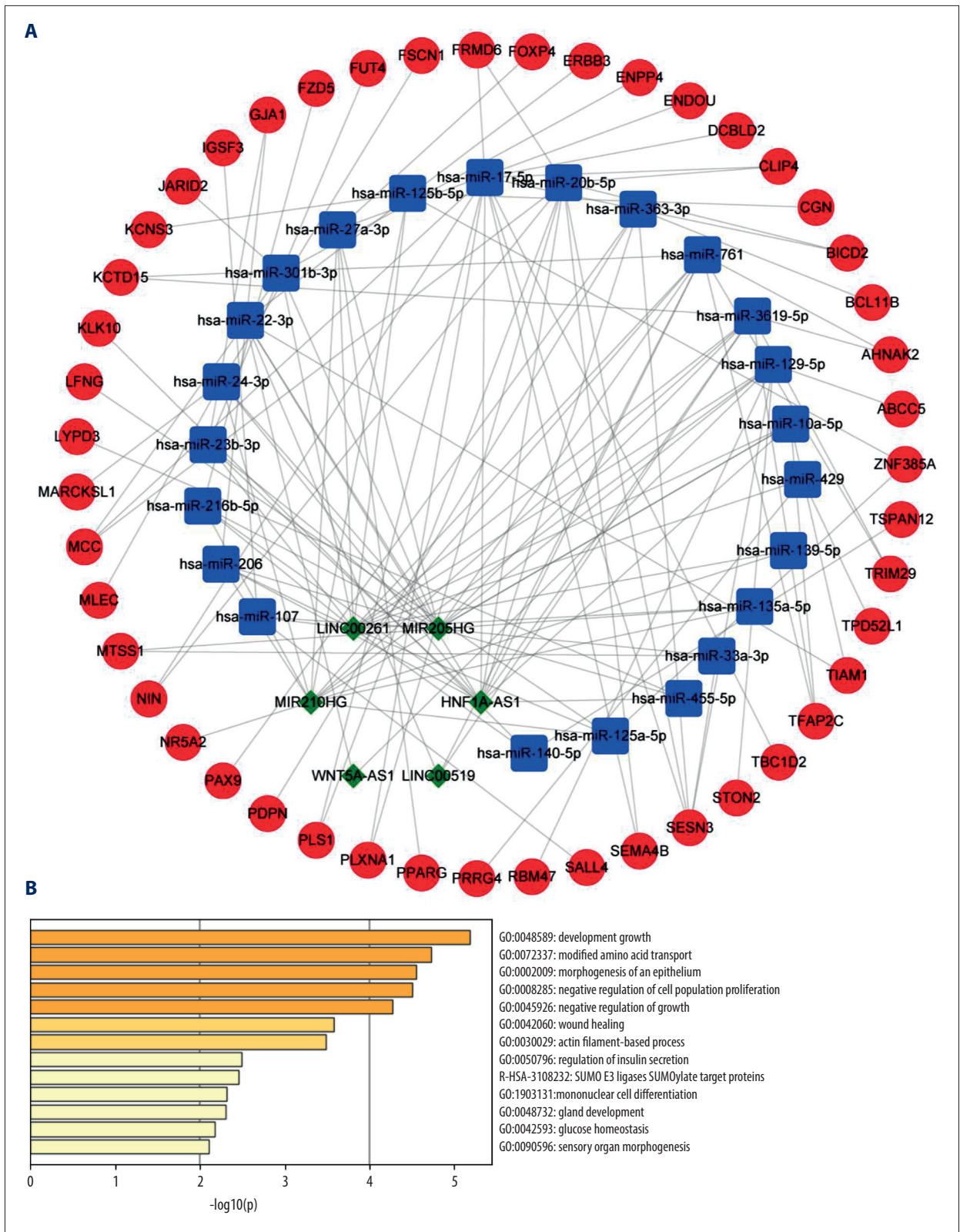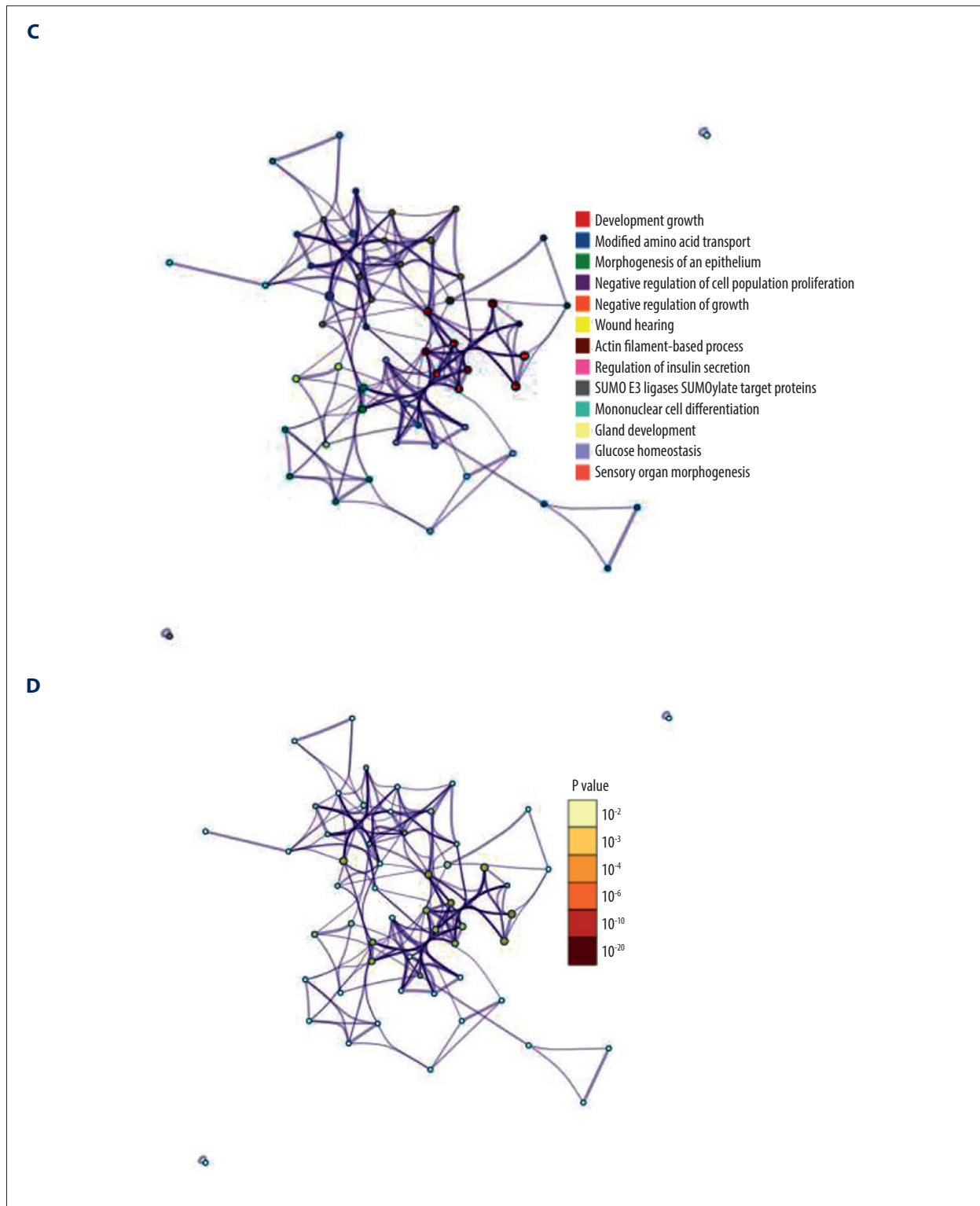
Figure 7. Construction of prognostic model-related ceRNA networks. (**A**) The ceRNA network of differential mRNAs (red) and their target miRNAs (blue) and lncRNAs (green) between high- and low-risk groups. (**B**) Heat map of enriched terms for 47 mRNAs, the top 13 clusters, and their representative enriched terms (1 for each cluster). Color scale represents the significance of the *P* value. (**C, D**) Network of enriched terms. Colored by cluster ID (**C**), where nodes with the same cluster ID are usually close to each other; Colored by *P* value (**D**), where terms containing more genes tend to have more significant *P* values.

Discussion

Hypoxia occurs in tumors when oxygen consumption exceeds oxygen supply in the tumor tissue. Intra-tumor hypoxia can promote the malignancy and aggressiveness of tumors and further adversely affect the clinical prognosis of cancer patients, especially in solid tumors [19]. However, uncontrolled proliferation of tumors can stimulate the growth of microvasculature. However, unlike the precise distribution of blood vessels in normal tissues, neovascularization in tumor tissues is characterized by disorderly distribution, and excessive misdistribution of tumor capillaries induced by hypoxic microenvironment can increase the distance between capillaries, thus exceeding the upper limit of oxygen exchange and diffusion capacity between capillaries [20]. Hypoxic areas within the tumor have been identified as one of the independent prognostic factors for cancer. Differences in the severity and duration of hypoxia result in oxygen gradients within the tumor, leading to heterogeneity and plasticity and promoting a more metastatic and aggressive phenotype [21]. More importantly, hypoxia mediates resistance to conventional chemotherapy and radiotherapy, which is ultimately associated with poor prognosis in cancer patients [22]. A growing number of studies have shown that hypoxia-mediated drug delivery contributes significantly to the development of chemoresistance in many carcinomas, including esophageal cancer [23]. As a critical mediator of the hypoxic response in the tumor microenvironment, the crucial role of hypoxia-inducible factors (HIFs) in resistance to conventional anti-cancer therapy for esophageal cancer has been demonstrated [24]. A recent study showed that high expression levels of the long noncoding RNA (lncRNA) E2F1 messenger RNA (mRNA) stabilizing factor (EMS), Wilms' tumor 1-associating protein (WTAP), and low expression levels of miR-758-3p were associated with poorer survival in patients with esophageal cancer. In addition, they demonstrated that the interaction between lncRNA-EMS/miR-758-3p/WTAP might be a potential mechanism for hypoxia-mediated cisplatin resistance in patients with esophageal cancer[25].

Cancer stem cells (CSCs) are defined as a small subpopulation of malignant cells at the top of the cell pyramid, characterized by the ability to self-renew and to give rise to differentiated cells [26]. Several clinical trials have shown that current surgical procedures alone or in combination with neoadjuvant therapy do not significantly improve the prognosis of patients with esophageal cancer [27-29]. The presence of CSCs is one of the leading causes of cancer recurrence and is mainly responsible for treatment resistance, which explains the inadequate response to conventional clinical treatment for esophageal cancer [30]. CSCs in esophageal cancer are directly involved in the regulation of carcinogenesis, tumor proliferation, distant metastasis, immune evasion, and drug resistance [31]. Notch, Wnt/β-catenin, Hedgehog, Hippo pathways, and immune system interactions have been shown to be central mechanisms involved in regulating the stemness of CSCs [32]. Intratumoral hypoxic regions are thought to be niches where treatment-resistant and more aggressive cell subtype, cancer stem cells (CSCs), localize and accumulate [33]. CSCs-associated genes, including Lgr5, ALDH1A1, NANOG, SOX2, and p75NTR, were reported to be upregulated in ESCC under Cocl2 mimic hypoxic conditions. In addition, they demonstrated that targeting hypoxia by specific knockdown of HIF-1α inhibits cell proliferation and spheroid formation, suppresses the expression of CSC-related genes, and inhibits the activity of the Wnt/β-catenin pathway [34].

Considering the widely varying prognosis of patients with esophageal cancer, it would be valuable to establish an effective classification system to stratify patients with different levels of risk. Currently, due to the complexity of the tumor microenvironment (TME), few transcriptional expression profile-based biomarkers have been developed and assessed for their prognostic value in ESCC. Solid tumors are highly disorganized normal organs with many cell types, including fibroblasts, endothelial cells, immune cells, and malignant tumor cells. Complex metabolite exchange mechanisms within the tumor synergistically manage the metabolism of the tumor [35]. Therefore, it is difficult for a single biomarker to effectively assess the state of hypoxia within a tumor. Since hypoxia and stemness in the tumor microenvironment are symbiotic, there are no published biomarkers based on transcriptional expression patterns to jointly assess their status in esophageal cancer. In our study, mRNA expression data of hypoxic genes were used to verify esophageal cancer subtype assignment based on the T-SNE approach, and differentially expressed genes (DEGs) among esophageal cancer subtypes were obtained by the Limma method. Regarding stemness status, the idea of a stemness index was adopted for grouping patients with esophageal cancer and further to obtain differentially expressed genes (DEGs) associated with stemness index. In the present study, a total of 242 hypoxia-stemness index overlap DEGs were selected. Prognostic prediction models were further constructed by integrating differential genes associated with hypoxia-stemness index using LASSO regression. Finally, 8 prognosis-related genes (FBLN2, IL17RB, CYP2W1, AMTN, FABP1, FOXA2, GAS1, and CTSF) were selected to construct a risk scoring system. A risk score formula was built for each esophageal cancer patient, and the patients were divided into low- and high-risk groups. The stability of the prediction model was finally validated by assessing the survival differences between the 2 subgroups in both internal (TCGA-ESCC) and external datasets (GSEA53624), and univariate and multivariate Cox analyses showed that only the risk score was an independent prognostic variable in the TCGA-ESCC cohort.

The vital role of risk signature genes identified in this study has been previously reported in several types of cancer. The

human FBLN2 gene encodes a protein that is a member of the fibronectin family. Fibulin-2 was initially found to be expressed at the site of the conversion of polarized epithelial cells into mesenchyme during endocardial cushion tissue development. In addition, fibulin-2 has been reported to be involved in cell migration and adhesion activities [36]. Notably, an isomer of FBLN2, FBLN2S, has previously shown tumor-suppressive and anti-angiogenic effects in nasopharyngeal carcinoma [37]. These results are consistent with the results of the present study, in which FBLN2 was identified as a gene associated with good prognosis of esophageal cancer. In contrast, overexpression of FBLN2 is related to unfavorable pathological features and poor prognosis in urethral epithelial cancer patients [36]. Another risk gene, IL17RB, has been reported to enhance cancer stemness and resistance to the chemotherapeutic drug gemcitabine in pancreatic cancer [38]. Moreover, FoxA2, a forkhead box transcription factor, is a member of the FoxA subfamily, and the expression level of FoxA2 was significantly increased under hypoxia environment in prostate adenocarcinomas. More importantly, the interaction between FpxA2 and HIF-1α stimulates a specific subset of HIF target genes, including Sox9, Jmjd1a, and Plod2, in the presence of hypoxia [39]. However, the above-mentioned signature genes have rarely been studied in the context of combined hypoxia and cancer stemness. The risk signature genes identified in this study may provide targets for elucidating the molecular mechanisms of esophageal cancer.

In this study, we also conducted a multi-omics study to investigate the clinical assessment validity of the risk model. We observed a positive correlation between high-risk scores and M2 macrophages, Tregs. As a drug delivery vehicle, macrophages play an essential role in hypoxia-mediated immune evasion due to their targeting tropism to the intra-tumor hypoxic microenvironment. M2 macrophages, also known as reparative macrophages, are involved in constructive tissue processes, including wound healing and tissue repair. In addition, M2 macrophages evade the immune surveillance system in the organism by secreting various immunosuppressive cytokines [40]. We further explored the mutation profiles of patients in the high- and low-risk groups, and the results showed that the proportion of TP53 mutations was significantly higher in the high-risk group than in the low-risk group. Early studies have shown that tumors with TP53 mutations have higher hypoxia scores in various tumor types, including breast and lung adenocarcinoma, thus supporting the idea that TP53 mutations are a genomic consequence of tumor hypoxia [41]. In addition, multiple enriched biological processes and signaling pathways, including EMT, significantly associated with risk scores were identified. Esophageal cancer cells undergo EMT through activation

of the Wnt/beta-catenin, Notch, and Hedgehog pathways in the tumor hypoxic microenvironment, acquiring characteristics of CSCs such as enhanced treatment resistance and malignancy. This makes EMT a vital bridge between hypoxia and cancer stem cells [42]. It is known that intratumorally hypoxia is associated with treatment resistance. The presence of cancer stem cells (CSCs) is also critical for tumor progression and chemotherapy resistance. More importantly, hypoxia affects patient prognosis not only by inducing tumor regeneration after treatment, but also by promoting the dissemination of stem cells [43]. In our study, patients with esophageal cancer in the high-risk group were significantly less sensitive to gefitinib, bexarotene, dasatinib, and imatinib, compared with the low-risk group. A randomized trial of second-line treatment for esophageal cancer reported that patients with epidermal growth factor receptor (EGFR) amplification had a better prognosis after treatment with gefitinib [44]. The relationship between EGFR and hypoxia has been demonstrated in various tumors [45,46].

Conclusions

In conclusion, 282 overlapping DEGs based on the hypoxia-stemness index were screened for the first time in ESCA, and 8 prognosis-related genes (FBLN2, IL17RB, CYP2W1, AMTN, FABP1, FOXA2, GAS1, and CTSF) were selected to construct a risk scoring system. Furthermore, this novel risk scoring system effectively predicted the clinical prognosis of ESCA patients, including TCGA-ESCA and GSEA53624. However, a limitation of this study is that the robustness of our risk scoring model still needs to be further validated in an external independent ESCA cohort.

Data Availability Statements

Some or all data, models, or code generated or used during this study are available from the corresponding author by reasonable request.
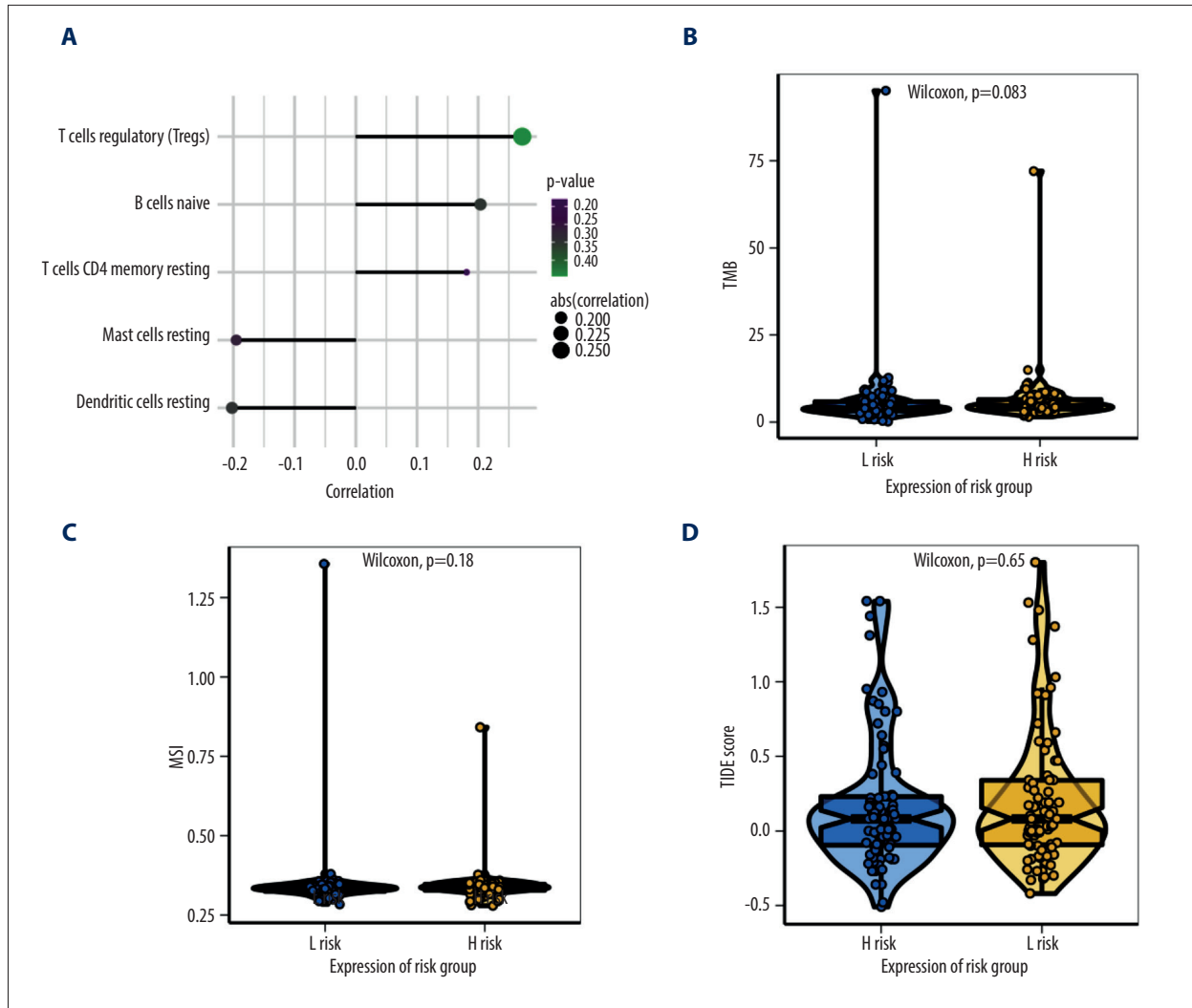
Declaration of Figures' Authenticity

All figures submitted have been created by the authors who confirm that the images are original with no duplication and have not been previously published in whole or in part.

Supplementary Material

Supplementary Figure 1. Multi-omics study to investigate the clinical value of hypoxia-stemness risk classifier. (**A**) Relationship between risk score and tumor immune infiltration based on CIBERSORT algorithm. (**B**) TMB level in high and low-risk groups. (**C**) MSI level in high and low-risk groups. (**D**) TIDE score in high and low-risk groups of ESCA patients. TMB – tumor mutational burden; MSI – microsatellite instability; TIDE – tumor immune dysfunction and exclusion.

References:

1. Pennathur A, Gibson MK, Jobe BA, Luketich JD. Oesophageal carcinoma. Lancet. 2013;381:400-12

2. Sjoquist KM, Burmeister BH, Smithers BM, et al. Survival after neoadjuvant chemotherapy or chemoradiotherapy for resectable oesophageal carcinoma: An updated meta-analysis. Lancet Oncol. 2011;12:681-92

3. Riera-Domingo C, Audigé A, Granja S, et al. Immunity, hypoxia, and metabolism-the Ménage à Trois of cancer: Implications for immunotherapy. Physiol Rev. 2020;100:1-102

4. Peerlings J, Van De Voorde L, Mitea C, et al. Hypoxia and hypoxia response-associated molecular markers in esophageal cancer: A systematic review. Methods. 2017;130:51-62

5. Driessen A, Landuyt W, Pastorekova S, et al. Expression of carbonic anhydrase IX (CA IX), a hypoxia-related protein, rather than vascular-endothelial growth factor (VEGF), a pro-angiogenic factor, correlates with an extremely poor prognosis in esophageal and gastric adenocarcinomas. Ann Surg. 2006;243:334-40

6. Kimura S, Kitadai Y, Tanaka S, et al. Expression of hypoxia-inducible factor (HIF)-1alpha is associated with vascular endothelial growth factor expression and tumour angiogenesis in human oesophageal squamous cell carcinoma. Eur J Cancer. 2004;40:1904-12

7. Lee KE, Simon MC. From stem cells to cancer stem cells: HIF takes the stage. Curr Opin Cell Biol. 2012;24:232-35

8. Bhaduri A, Di Lullo E, Jung D, et al. Outer radial glia-like cancer stem cells contribute to heterogeneity of glioblastoma. Cell Stem Cell. 2020;26:48-63.e6

9. Sun X, Lv X, Yan Y et al. Hypoxia-mediated cancer stem cell resistance and targeted therapy. Biomed Pharmacother. 2020;130:110623

10. Kato Y, Yashiro M, Fuyuhiro Y, et al. Effects of acute and chronic hypoxia on the radiosensitivity of gastric and esophageal cancer cells. Anticancer Res. 2011;31:3369-75

11. Wang D, Gao X, Wang X. Semi-supervised nonnegative matrix factorization via constraint propagation. IEEE Trans Cybern. 2016;46:233-44

12. Kobak D, Berens P. The art of using t-SNE for single-cell transcriptomics. Nat Commun. 2019;10:5416

13. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47

14. Yu G, Wang LG, Han Y, He QY. clusterProfiler: An R package for comparing biological themes among gene clusters. OMICS. 2012;16:284-87

15. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33:1-22

16. Aran D, Hu Z, Butte AJ. xCell: Digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. 2017;18:220

17. Yang S, Liu T, Cheng Y, et al. Immune cell infiltration as a biomarker for the diagnosis and prognosis of digestive system cancer. Cancer Sci. 2019;110:3639-49

18. Hänzelmann S, Castelo R, Guinney J. GSVA: Gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7

19. Rankin EB, Giaccia AJ. Hypoxic control of metastasis. Science. 2016;352:175-80

20. Jain RK. Normalization of tumor vasculature: An emerging concept in anti-angiogenic therapy. Science. 2005;307:58-62

21. Jing X, Yang F, Shao C, et al. Role of hypoxia in cancer therapy by regulating the tumor microenvironment. Mol Cancer. 2019;18:157

22. Wigerup C, Påhlman S, Bexell D. Therapeutic targeting of hypoxia and hypoxia-inducible factors in cancer. Pharmacol Ther. 2016;164:152-69

23. Sharma A, Arambula JF, Koo S, et al. Hypoxia-targeted drug delivery. Chem Soc Rev. 2019; 48:771-813

24. Rohwer N, Cramer T. Hypoxia-mediated drug resistance: Novel insights on the functional interaction of HIFs and cell death pathways. Drug Resist Updat. 2011;14:191-201

25. Zhu ZJ, Pang Y, Jin G, et al. Hypoxia induces chemoresistance of esophageal cancer cells to cisplatin through regulating the lncRNA-EMS/miR-758-3p/WTAP axis. Aging (Albany NY). 2021;13:17155-76

26. Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. Nature. 2001;414:105-11

27. Arnott SJ, Duncan W, Kerr GR, et al. Low dose preoperative radiotherapy for carcinoma of the oesophagus: Results of a randomized clinical trial. Radiother Oncol. 1992;24:108-13

28. Yano T, Muto M, Minashi K et al. Photodynamic therapy as salvage treatment for local failure after chemoradiotherapy in patients with esophageal squamous cell carcinoma: A phase II study. Int J Cancer. 2012;131:1228-34

29. Bao Y, Liu S, Zhou Q, et al. Three-dimensional conformal radiotherapy with concurrent chemotherapy for postoperative recurrence of esophageal squamous cell carcinoma: clinical efficacy and failure pattern. Radiat Oncol. 2013;8:241

30. Islam F, Gopalan V, Wahab R, et al. Cancer stem cells in oesophageal squamous cell carcinoma: Identification, prognostic and treatment perspectives. Crit Rev Oncol Hematol. 2015;96:9-19

31. Islam F, Gopalan V, Lam AK. Detention and identification of cancer stem cells in esophageal squamous cell carcinoma. Methods Mol Biol. 2020;2129:177-91

32. Clara JA, Monge C, Yang Y, Takebe N. Targeting signalling pathways and the immune microenvironment of cancer stem cells – a clinical update. Nat Rev Clin Oncol. 2020;17:204-32

33. Najafi M, Farhood B, Mortezaee K, et al. Hypoxia in solid tumors: A key promoter of cancer stem cell (CSC) resistance. J Cancer Res Clin Oncol. 2020;146:19-31

34. Lv Z, Liu RD, Chen XQ, et al. HIF-1α promotes the stemness of oesophageal squamous cell carcinoma by activating the Wnt/β-catenin pathway. Oncol Rep. 2019;42:726-34

35. Lyssiotis CA, Kimmelman AC. Metabolic interactions in the tumor microenvironment. Trends Cell Biol. 2017;27:863-75

36. de Vega S, Iwamoto T, Yamada Y. Fibulins: Multiple roles in matrix structures and tissue functions. Cell Mol Life Sci. 2009;66:1890-902

37. Law EW, Cheung AK, Kashuba VI, et al. Anti-angiogenic and tumor-suppressive roles of candidate tumor-suppressor gene, Fibulin-2, in nasopharyngeal carcinoma. Oncogene. 2012;31:728-38

38. Tsai LH, Hsu KW, Chiang CM, et al. Targeting interleukin-17 receptor B enhances gemcitabine sensitivity through downregulation of mucins in pancreatic cancer. Sci Rep. 2020;10:17817

39. Qi J, Nakayama K, Cardiff RD, et al. Siah2-dependent concerted activity of HIF and FoxA2 regulates formation of neuroendocrine phenotype and neuroendocrine prostate tumors. Cancer Cell. 2010;18:23-38

40. Li X, Liu R, Su X et al: Harnessing tumor-associated macrophages as aids for cancer immunotherapy. Mol Cancer, 2019;18:177

41. Bhandari V, Hoey C, Liu LY, et al. Molecular landmarks of tumor hypoxia across cancer types. Nat Genet. 2019;51:308-18

42. Wang D, Plukker J, Coppes RP. Cancer stem cells with increased metastatic potential as a therapeutic target for esophageal cancer. Semin Cancer Biol, 2017;44:60-66

43. van den Beucken T, Koch E, Chu K, et al. Hypoxia promotes stem cell phenotypes and poor prognosis through epigenetic regulation of DICER. Nat Commun. 2014;5:5203

44. Petty RD, Dahle-Smith A, Stevenson D, et al. Gefitinib and EGFR gene copy number aberrations in esophageal cancer. J Clin Oncol. 2017;35:2279-87

45. Swinson DE, O'Byrne KJ. Interactions between hypoxia and epidermal growth factor receptor in non-small-cell lung cancer. Clin Lung Cancer. 2006;7:250-56

46. Zhong H, Chiles K, Feldser D, et al. Modulation of hypoxia-inducible factor 1alpha expression by the epidermal growth factor/phosphatidylinositol 3-kinase/PTEN/AKT/FRAP pathway in human prostate cancer cells: Implications for tumor angiogenesis and therapeutics. Cancer Res. 2000;60:1541-45