Original Article

Stemness-related gene signature for predicting therapeutic response in patients with esophageal cancer

Shaojin Zhu[1]^, Gengxin Zhang[1], Qi You[1], Fei Li[1], Boying Ding[1], Feng Liu[2], De Bi[3], Lan Jiang[1,4,5,6]^

[1]Department of Thoracic Surgery, The First Affiliated Hospital of Wannan Medical College (Yijishan Hospital of Wannan Medical College), Wuhu, China; [2]Department of Thoracic Surgery, Lishui Branch of Zhongda Hospital Affiliated to Southeast University, Nanjing, China; [3]Suzhou Polytechnic Institute of Agriculture, Suzhou, China; [4]Key Laboratory of Non-Coding RNA Transformation Research of Anhui Higher Education Institution, Yijishan Hospital of Wannan Medical College, Wuhu, China; [5]Central Laboratory, Yijishan Hospital of Wannan Medical College, Wuhu, China; [6]Clinical Research Center for Critical Respiratory Medicine of Anhui Province, Wuhu, China

Contributions: (I) Conception and design: L Jiang, S Zhu; (II) Administrative support: L Jiang, S Zhu; (III) Provision of study materials or patients: L Jiang, S Zhu; (IV) Collection and assembly of data: All authors; (V) Data analysis and interpretation: L Jiang, S Zhu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Lan Jiang. Key Laboratory of Non-Coding RNA Transformation Research of Anhui Higher Education Institution, The First Affiliated Hospital of Wannan Medical College (Yijishan Hospital of Wannan Medical College), Wuhu, China. Email: jianglanhi@163.com; De Bi. Suzhou Polytechnic Institute of Agriculture, Suzhou, China. Email: bide1125@vip.163.com.

Background: Extensive research has indicated that tumor stemness promotes tumor progression. However, the underlying role of stemness-related genes (SRGs) in esophageal cancer (ESCA) remains unclear.

Methods: This study identified differentially expressed stemness-related (DESR) messenger RNAs (mRNAs), microRNAs (miRNAs), and long non-coding RNAs (lncRNAs) in ESCA, and correlated them with the clinical features of patients with ESCA to develop a prognostic risk assessment model. Functional analysis, protein-protein interaction (PPI) analysis, competing endogenous RNA (ceRNA) networks, and tumor-infiltrating immune cell analyses were performed to corroborate the results obtained from the model.

Results: Correlation analysis of the stemness enrichment scores revealed 1,106 DESR genes (DESRGs), 84 DESRmiRNAs, and 320 DESRlncRNAs were identified from The Cancer Genome Atlas Esophageal Carcinoma (TCGA-ESCA) dataset. Network clustering was performed and the top 20 connection points were identified, including *CDC20* that connects to 136 adjacent nodes. A ceRNA network was constructed, including 17 DESRmiRNAs, 44 DESRlncRNAs, and 55 DESRGs.

Conclusions: *NCAPG* [$\log_2$fold change (FC) =1.81; q value =2.68×10^{-11}] was significantly upregulated in ESCA and positively correlated with resting natural killer (NK) cells, suggesting that human NK cells rest via the overexpression of *NCAPG* in ESCA. *hsa-miR-1269a* is significantly upregulated in ESCA patients with poor prognostic features. $CD4^+$ resting memory T cells ($P<0.01$) were significantly negatively correlated with *hsa-miR-1269a*. The insights presented in this study will contribute to the development of innovative therapeutics for the treatment of patients with ESCA.

Keywords: Stemness; biomarker; esophageal cancer (ESCA); prognostic; survival

Submitted May 30, 2022. Accepted for publication Jul 10, 2022.
doi: 10.21037/tcr-22-1723
View this article at: https://dx.doi.org/10.21037/tcr-22-1723

^ ORCID: Shaojin Zhu, 0000-0002-0633-3987; Lan Jiang, 0000-0003-2778-1081.

Introduction

Esophageal cancer (ESCA) is the second deadliest gastrointestinal cancer. Patients generally have poor prognosis, with a 5-year survival rate of 40% for early-stage patients (1). There are two pathological types of ESCA that show regional specificities because of differences in living habits and genetic backgrounds. Esophageal squamous cell carcinoma is more prevalent in Southeast Asia, while esophageal adenocarcinoma is more frequent in Europe and America (2). Patients with ESCA can be divided into resectable patients and those who are inoperable (3). Traditional treatments such as surgery, radiotherapy, chemotherapy, immunotherapy, and targeted therapy, have limited efficacy in both patients with resectable and those with inoperable ESCA, leading to poor outcomes (4). In patients with advanced metastatic ESCA, nanoparticles Abraxane™ and Doxil™ can be used as treatment options (3). Carboplatin and paclitaxel are the most common preoperative treatment agents for ESCA (5), while capecitabine and fluorouracil are administered postoperatively (6). Minimally invasive esophagectomy is a common surgical method for patients with operable ESCA to reduce postoperative morbidity. For patients with inoperable ESCA, cebox and cisplatin are used in combination to prolong survival (3). Pembrolizumab™, which targets specific genetic features, has been approved by the Food and Drug Administration (FDA) for the treatment of advanced ESCA (7). Therefore, current research on ESCA should be continuously adjusted to develop more effective treatment strategies.

Cancer stem cells (CSCs) are highly relevant for tumorigenesis, progression, recurrence, and metastasis of various cancers (8). The messenger RNA (mRNA) expression-based stemness index (mRNAsi) is used to evaluate the unique characteristics of CSCs and to assess tumor development (9). There is mounting evidence to support the existence of CSCs in lung adenocarcinoma (10), breast cancer (11), colorectal cancer (12), and prostate cancer (13), as well as their roles in metastasis, drug resistance, and cancer adaptation to changing microenvironments. For instance, single-cell transcriptome analysis revealed the landscape of intra-tumoral heterogeneity and stemness-related subpopulations of liver cancer cells (14).

Stemness-related genes (SRGs) have been identified in numerous cancers, such as cervical squamous cell carcinoma (15), colorectal cancer (16), renal clear cell carcinoma (17), lung cancer (18), gastric cancer (19), and ESCA (20). Moreover, SRGs have been shown to promote the formation, progression, and metastasis of different cancers (21). The *AGR3* gene has been shown to promote the stemness of CSCs by upregulating SRGs via Frizzled 4 (*FZD4*), which is involved in the Wnt/β-catenin signaling pathway (22). Study examining ESCA stemness and clinical characteristics have indicated that higher-stage and metastatic tumors feature great phenotypic dedifferentiation (20). This study examined how SRGs promote ESCA formation, progression, and metastasis. The differentially expressed stemness-related (DESR) genes (DESRGs), DESR microRNAs (DESRmiRNAs), and DESR long non-coding RNAs (DESRlncRNAs) were firstly identified and functional analyses were performed. CSCs could change tumor microenvironment into immunosuppressive (23). Tumors with significant positive DESRGs, DESRlncRNAs, and DESRmiRNAs expression may exist higher immune infiltration in the tumor microenvironment (21). The DESRGs, DESRmiRNAs, and DESRlncRNAs were used to develop a prognostic risk assessment model and to construct a protein-protein interaction (PPI) network and a competing endogenous RNA (ceRNA) network, which together with the tumor-infiltrating immune cell analyses, corroborated the model results (*Figure 1*). Potential stemness-related prognostic markers for ESCA were identified. We present the following article in accordance with the TRIPOD reporting checklist (available at https://tcr.amegroups.com/article/view/10.21037/tcr-22-1723/rc).

Methods

Data collection and analysis of differential gene expression

The ESCA gene expression profile was obtained from The Cancer Genome Atlas-ESCA (TCGA-ESCA) database containing 11 normal tissues and 152 tumor tissues, and analyzed using the R package TCGA biolinks (24). Among these, 145 tumor tissues had complete clinical information, including age, sex, survival status, and tumor, node, metastasis (TNM) stage. The read counts were converted to transcripts per kilobase million (TPM) to compare the relative gene expression level between samples, while the miRNA data were converted to reads per kilobase million (RPKM) (25).

Differential gene expression between tumor and normal tissue samples was analysis using the R package DESeq2 (26). The significant threshold for the mRNA and lncRNA differential expression was defined as the

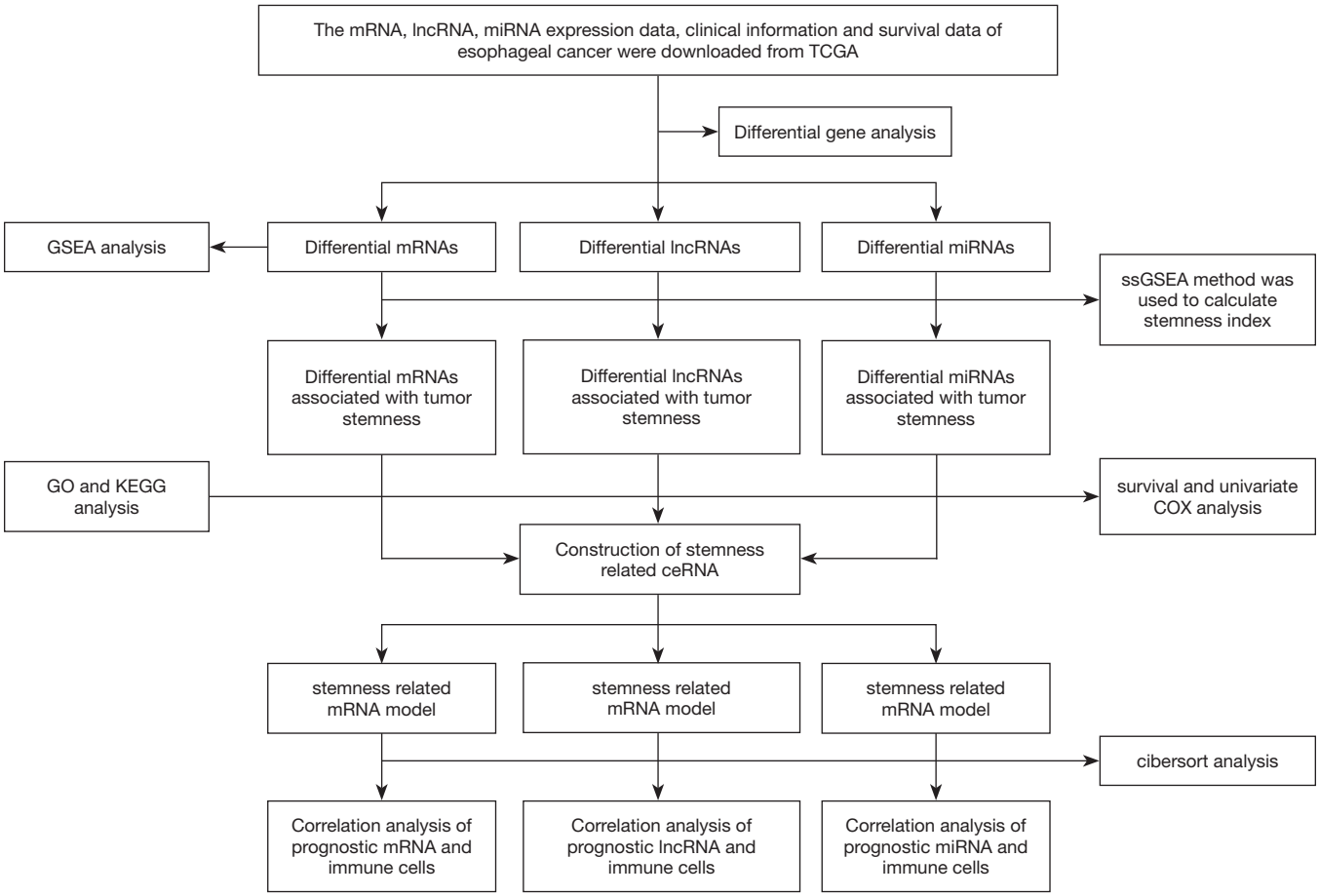

Figure 1 Flowchart showing the process involved in this study. mRNA, messenger RNA; lncRNA long non-coding RNA; miRNA, microRNA; TCGA, The Cancer Genome Atlas; GSEA, gene set enrichment analysis; ssGSEA, single-sample GSEA; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; ceRNA, competing endogenous RNA.

$\log_2$ of the absolute fold change (FC) value greater than 1 ($\log_2$FC >1) and adjusted P value less than 0.05 (q value <0.05). The threshold to determine differentially expressed miRNA was set as $\log_2$FC >0.5 and q value <0.05. The differential analysis results are displayed as heat maps using the R package pheatmap (https://CRAN.R-project.org/package=pheatmap) and as volcano maps using the ggplot2 R package (27). The DESRGs were obtained using the gene set enrichment analysis (GSEA) (https://www.gsea-msigdb.org/gsea/index.jsp) tool of the GSEA software (28). The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Construction of the PPI network and the ceRNA network

The single-sample GSEA (ssGSEA) method (29) was used to obtain the enrichment score for cell stemness of each sample based on the ESCA expression data and the DESRG set. To identify the DESRmiRNAs and DESRlncRNAs, correlation analysis was performed between the differentially expressed mRNA and the lncRNA or miRNA levels. Significant correlation coefficient greater than 0.2 (ρ>0.2; P<0.05) indicated strong positive correlation.

To construct the PPI network, the STRING database (https://string-db.org/), containing 2,031 species, 9.6 million proteins, and 1.38 million PPIs was used. Our PPI analysis included experimental results, text-mined results from PubMed abstracts, and synthesized and predicted results from bioinformatics analyses (30). A correlation coefficient above 0.7 (ρ>0.7; P<0.05) indicated a strong association between the differentially expressed genes and cell stemness. The PPI results were exported

from the STRING database and visualized using the Cytoscape StringApp (31). In addition, the molecular complex detection (MCODE) plugin from Cytoscape was used to analyze hub genes in the PPI network.

MiRcode (https://bio.tools/miRcode) provides transcriptome-wide human miRNA target predictions based on comprehensive GENCODE gene annotations (https://www.gencodegenes.org/), including 10,000 lncRNA genes (32). Therefore, MiRcode was used to identify the miRNAs and lncRNAs in our dataset. The TargetScan database was used to predict the biological targets of the identified miRNAs by searching for the presence of conserved 6mer-8mer sites on the mRNA sequences matching the seed region of each miRNA (33). The miRDB database (http://mirdb.org/) was used to confirm the predicted miRNA targets and functional annotations by implementing its MirTarget tool, which uses thousands of miRNA-target interactions (MTIs) from high-throughput sequencing experiments (34). miRTarBase (https://mirtarbase.cuhk.edu.cn) is an experimentally validated MTI database that has accumulated more than 360,000 MTIs obtained from natural language processing after a manual survey of relevant literature (35). For the analysis using miRTarBase, research articles were systematically screened to collect those related to miRNA functional studies. The MTIs from TargetScan, miRDB, and miRTarBase analysis were combined with the identified DESRGs, DESRmiRNAs, and DESRlncRNAs to construct a ceRNA network, which was visualized using Cytoscape.

Functional analysis

Gene Ontology (GO) analysis is a standard method for large-scale functional enrichment studies that includes terms related to biological processes (BPs), molecular functions (MFs), and cellular components (CCs) (36). The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a widely used database for storing information on genomes, biological pathways, diseases, and drugs (37). GO and KEGG pathway enrichment analyses of DESRGs were performed using the clusterProfiler (38) R package, with a cutoff value of false discovery rate (FDR) <0.05 for statistical significance.

To investigate the differences in enriched BPs between normal and tumor tissues, GSEA was performed based on the TCGA gene expression profiling dataset of patients with ESCA (39). For GSEA analysis, the control "c2.cp.kegg.v7.1.entrez" dataset used was downloaded from the MsigDB database and compared with the gene expression profiling dataset of patients with ESCA (FDR <0.25).

Development and validation of a prognostic risk assessment model

Univariate Cox (uni-Cox) regression analysis (P<0.05) was performed for the DESRGs, DESRmiRNAs, and DESRlncRNAs to estimate ESCA prognosis. The uni-Cox results were visualized using forest plot. Survival analysis was performed on the DESRGs, DESRmiRNAs, and DESRlncRNAs and the results were displayed using the R package survminer (40).

Least absolute shrinkage and selection operator (LASSO) regression analysis was performed to filter the DESRGs, DESRmiRNAs, and DESRlncRNAs with a 10-fold cross-validation. Furthermore, the DESRGs, DESRmiRNAs, and DESRlncRNAs screened by the LASSO method were used for multivariate Cox (multi-Cox) proportional hazards regression and risk model construction. The LASSO glmnet package (41) was used to divide the TCGA samples with complete clinical information into a training set and a validation set at a ratio of 1:1. The training dataset was composed of the DESRGs, DESRmiRNAs, and DESRlncRNAs (single factor P<0.05). The accuracy of the final model was verified through the validation set. The R package ggrisk (https://CRAN.R-project.org/package=ggrisk) was used to display the risk factor and perform the survival analysis on the risk scores of the training and validation sets. The R package pROC was used to obtain the receiver operating characteristic (ROC) curve, nomogram, and standard curve, showing the accuracy of the model. Finally, multi-Cox analysis with clinical traits was used to determine whether the risk score is independent of other traits.

Tumor-infiltrating immune cells analyses

CIBERSORT (https://cibersort.stanford.edu/index.php) was used to analysis 22 tumor-infiltrating immune cell types in each tissue, using the LM22 signature matrix with 1,000 permutations (42). The gene expression matrix data (TPM) was uploaded to CIBERSORT combined with the LM22 eigengene matrix, and the immune cell infiltration matrix was derived. A histogram was drawn showing the distribution of the 22 infiltrating immune cells in each sample. The R package corrplot (https://github.com/taiyun/corrplot) was used to calculate and display the correlation between the prognosis-related mRNA, lncRNA, miRNA, and the 22 tumor-infiltrating immune cell types.

Table 1 Baseline data of TCGA patients with ESCA

Type	Alive (n=98)	Deceased (n=46)	P value
Tumor grade, n (%)			0.238
G1	16 (16.3)	2 (4.3)	
G2	38 (38.8)	19 (41.3)	
G3	21 (21.4)	12 (26.1)	
GX	23 (23.5)	13 (28.3)	
Gender, n (%)			0.448
Female	17 (17.3)	5 (10.9)	
Male	81 (82.7)	41 (89.1)	
Age (years), n (%)			0.912
<60	43 (43.9)	19 (41.3)	
≥60	55 (56.1)	27 (58.7)	
Stage, n (%)			0.0011
I	17 (17.3)	3 (6.5)	
II	48 (49.0)	14 (30.4)	
III	31 (31.6)	22 (47.8)	
IV	2 (2.0)	7 (15.2)	
T staging, n (%)			0.191
T1	18 (18.4)	9 (19.6)	
T2	25 (25.5)	9 (19.6)	
T3	55 (56.1)	26 (56.5)	
T4	0 (0.0)	2 (4.3)	
M staging, n (%)			0.00238
M0	86 (87.8)	34 (73.9)	
M1	1 (1.0)	7 (15.2)	
MX	11 (11.2)	5 (10.9)	
N staging, n (%)			0.00909
N0	48 (49.0)	11 (23.9)	
N1	33 (33.7)	28 (60.9)	
N2	6 (6.1)	4 (8.7)	
N3	6 (6.1)	0 (0.0)	
NX	5 (5.1)	3 (6.5)	

TCGA, The Cancer Genome Atlas; ESCA, esophageal cancer.

Statistical analyses

All statistical analyses were performed using R software version 4.1.3 (https://www.r-project.org/). The ROC curves assessed the accuracy of the risk score in estimating the prognosis by calculating the area under the ROC curve (AUC). Spearman's correlation (ρ) was used to determine all correlations between two samples. All statistical analyses were two-sided, with a significance of $P<0.05$.

Results

Differential expression analyses

DESeq2 was used to analyze the differential expression between TCGA-ESCA tumors [152] and normal tissues [11]. Baseline data of TCGA patients with ESCA was detected in *Table 1*. A total of 3,504 differentially expressed mRNAs (including 1,900 upregulated and 1,604 downregulated mRNAs; *Figure 2A*), 1,564 differentially expressed lncRNAs (including 993 upregulated and 571 downregulated lncRNAs; *Figure 2B*), and 209 differentially expressed miRNAs (including 130 upregulated and 79 downregulated miRNAs; *Figure 2C*) were identified. Heatmaps were used to identify the most significant differentially expressed mRNAs, lncRNAs, and miRNAs (*Figure 2D-2F*).

Stemness-related ceRNA network

Correlation analysis of the stemness enrichment scores revealed 1,106 DESRGs, 84 DESRmiRNAs, and 320 DESRlncRNAs (*Figure 3A*). Network clustering was performed (*Figure 3B-3D*) and the top 20 connection points were identified, including CDC20 that connects to 136 adjacent nodes (*Figure 3E*). A ceRNA network was constructed, including 17 DESRmiRNAs, 44 DESRlncRNAs, and 55 DESRGs (*Figure 3F*).

Functional analysis of the DESRGs

GO enrichment analysis for the DESRGs (*Figure 4A*,*4B*) identified nuclear division, organelle fission, mitotic nuclear division, DNA-dependent DNA replication, and DNA replication as enriched initiation terms. KEGG pathway enrichment analysis revealed that the DESRGs were mainly enriched in biological pathways such as cell cycle, DNA replication, human papillomavirus infection, Cushing syndrome, and protein digestion and absorption (*Figure 4C*). GSEA functional analysis revealed that the DESRGs were enriched in the following KEGG pathways: cytokine-cytokine receptor interaction, cell cycle, calcium signaling pathway, neuroactive ligand-receptor interaction, fatty acid metabolism,

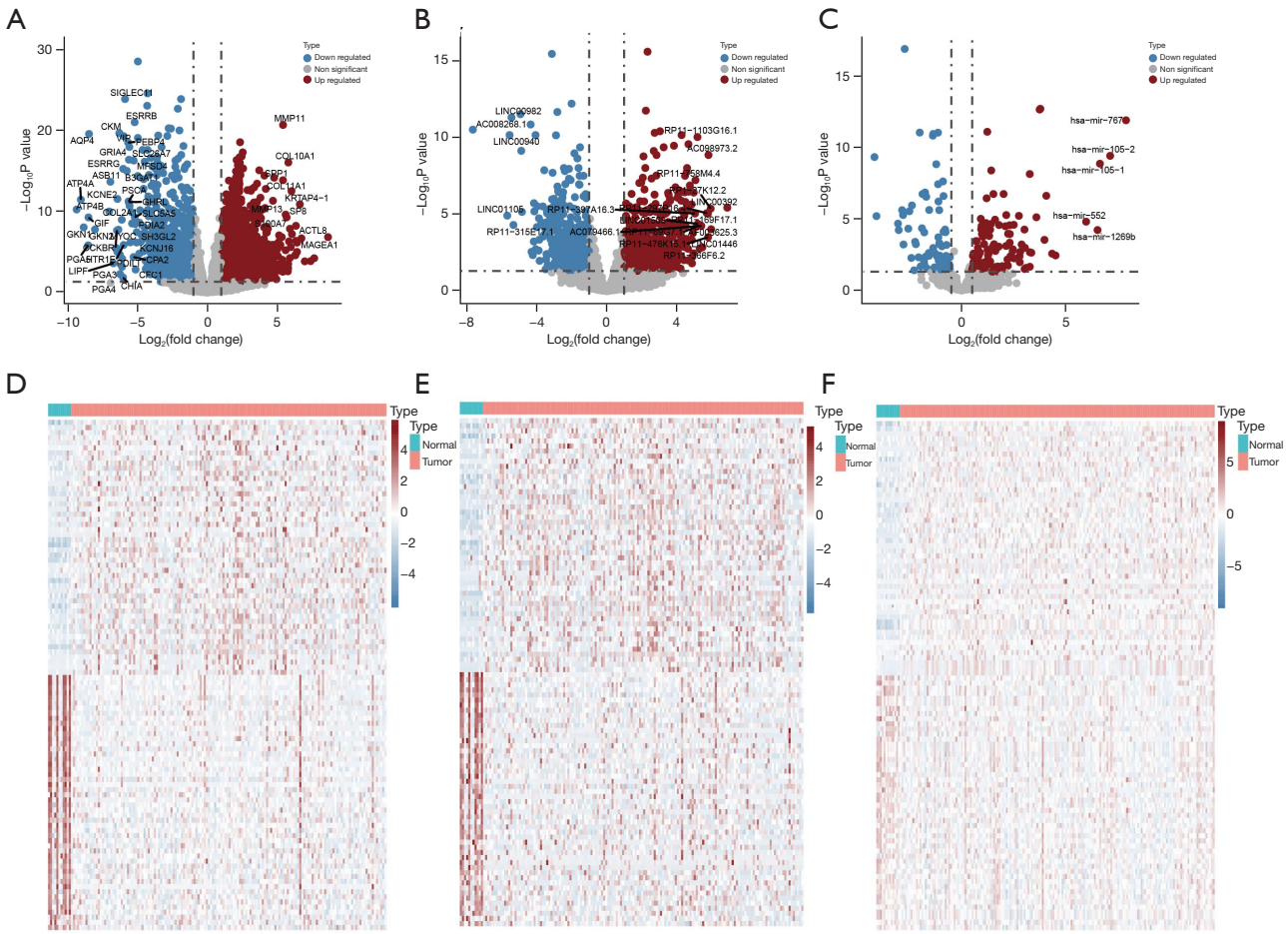

Figure 2 The DEGs. (A-C) The volcano plots show the DEmRNAs, DElncRNAs, and DEmiRNAs, respectively. The red nodes represent upregulated DEGs, the blue nodes represent downregulated DEGs, and the gray nodes represent DEGs that were not significantly different. (D-F) The heatmaps show the gene expression in TCGA-ESCA tumors samples and normal tissues. DEGs, differentially expressed genes; DEmRNAs, differentially expressed messenger RNAs; DElncRNAs, differentially expressed long non-coding RNAs; DEmiRNAs, differentially expressed microRNAs.

DNA replication, spliceosome, NOD-like receptor signaling pathway, vascular smooth muscle contraction pathway, and graft-versus-host disease (*Figure 5A*). The most enhanced pathways were cytokine-cytokine receptor interaction, cell cycle, and DNA replication (*Figure 5B*), whereas the most inhibited pathways were related to calcium signaling pathway, neuroactive ligand-receptor interaction, and fatty acid metabolism (*Figure 5C*).

Survival analysis of the DESRGs, DESRmiRNAs, and DESRlncRNAs

Uni-Cox and survival analysis of the DESRGs, DESRmiRNAs, and DESRlncRNAs was performed to identify the correlation on the survival of patients with ESCA. The uni-Cox model (P<0.05) used 73 DESRGs, 7 DESRmiRNAs, and 23 DESRlncRNAs (Table S1). The pseudogene *MTRNR2L13* and the chromosome 15 open reading frame 54 (*C15orf54*) were identified as high-risk factors associated with patient mortality. Notably, the centromere protein I (*CENPI*), PRICKLE2 antisense RNA 1 (*PRICKLE2-AS1*), and human miRNAs *hsa-miR-101-1* and *hsa-miR-101-2* were found to negatively correlate with patient mortality. Survival analysis revealed six highly significant DESRGs, DESRmiRNAs, and DESRlncRNAs (Figure S1).

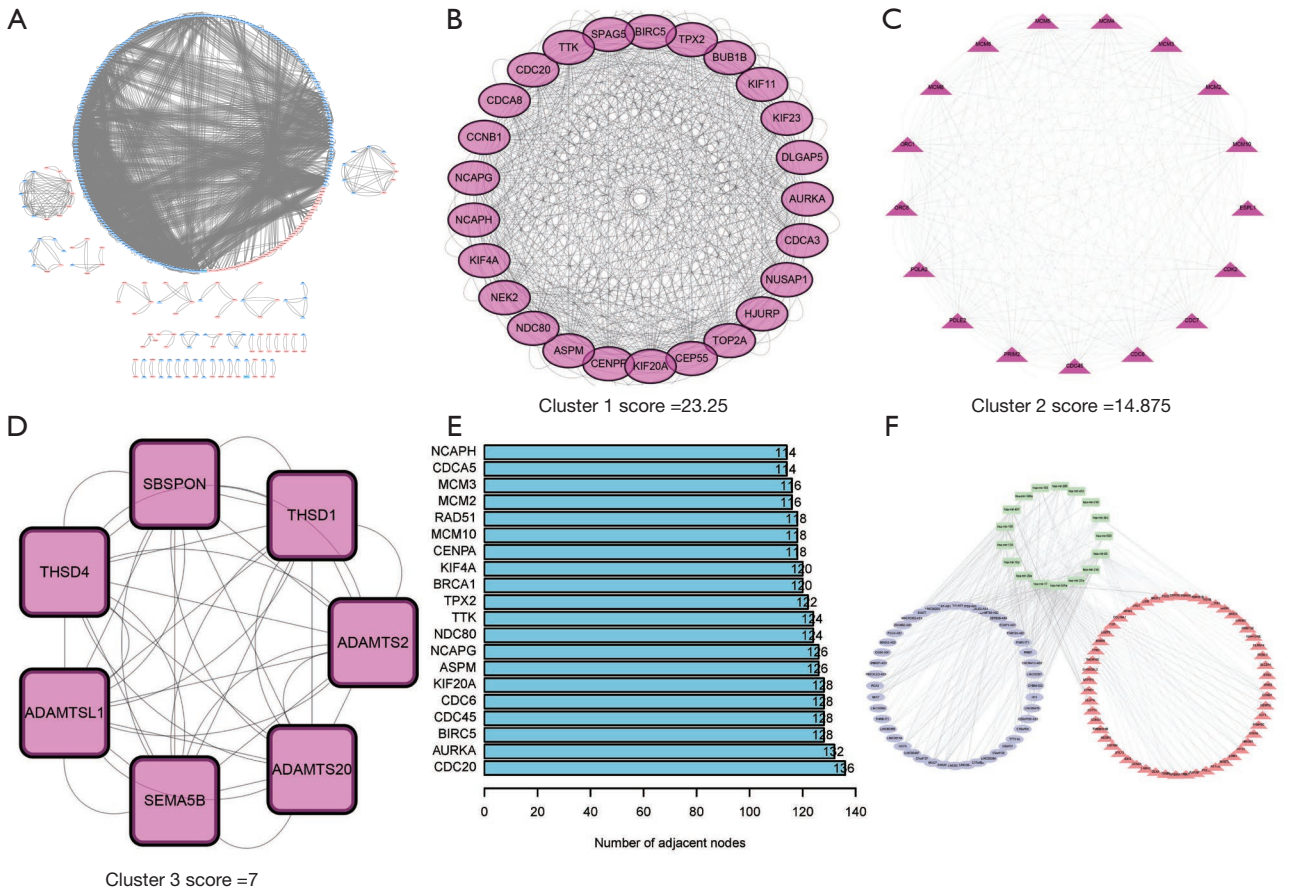

Figure 3 The stemness-related ceRNAs. (A) A PPI network diagram of the DESRGs. Blue indicates high expression in tumor tissue, while red indicates low expression in tumor tissue. (B-D) Gene network maps of the hubs screened with the cytoscape plugin MCODE. (E) Histogram depicting the PPI network nodes for the DESRGs. (F) A diagram of the stemness-related ceRNA network. Green presents miRNA, purple presents lncRNA, and orange presents mRNA. ceRNA, competing endogenous RNA; PPI, protein-protein interaction; DESRGs, differentially expressed stemness-related genes; MCODE, molecular complex detection; miRNA, microRNA; lncRNA, long non-coding RNA; mRNA, messenger RNA.

Construction and validation of the prognostic model

The TCGA-ESCA datasets were randomly divided into training and validation sets, and univariate analysis was performed to detect prognostic-related DESRGs. LASSO regression was used to identify meaningful genes in the uni-Cox analysis, including six DESRGs, namely, *NCAPG*, *HIST1H4J*, *KIAA1456*, *SCUBE2*, *PCCA*, and *ARID3A*, that were selected to build the prediction model (*Figure 6A*). The expression profiles of these selected genes were used to calculate the risk scores in the training and validation sets separately. The total risk score was calculated as follows: expression level of (*NCAPG* × 0.0245 + *HIST1H4J* ×1.2293 + *KIAA1456* × 0.0358 + *SCUBE2* × 0.1375 + *PCCA* × 0.0738 + *ARID3A* × 0.0134). The distribution of risk scores in patients with ESCA and the correlation between survival time and risk score in the training set revealed that patients with high scores were associated with a higher mortality (*Figure 6B*), indicating that high-risk patients generally had poor overall survival (OS). A similar trend was observed in the validation set (*Figure 6C*). Based on the median value of the risk scores, the patients were divided into a high-risk group and a low-risk group. The survival rate of patients in the high-risk group was significantly lower than that in the low-risk group in both the training (P<0.0001; *Figure 6D*) and validation sets (P=0.01; *Figure 6E*). The AUC for 1-, 2-, and 3-year survival rates in the training group were 0.995, 0.937, and 0.818, respectively (*Figure 6F*), while

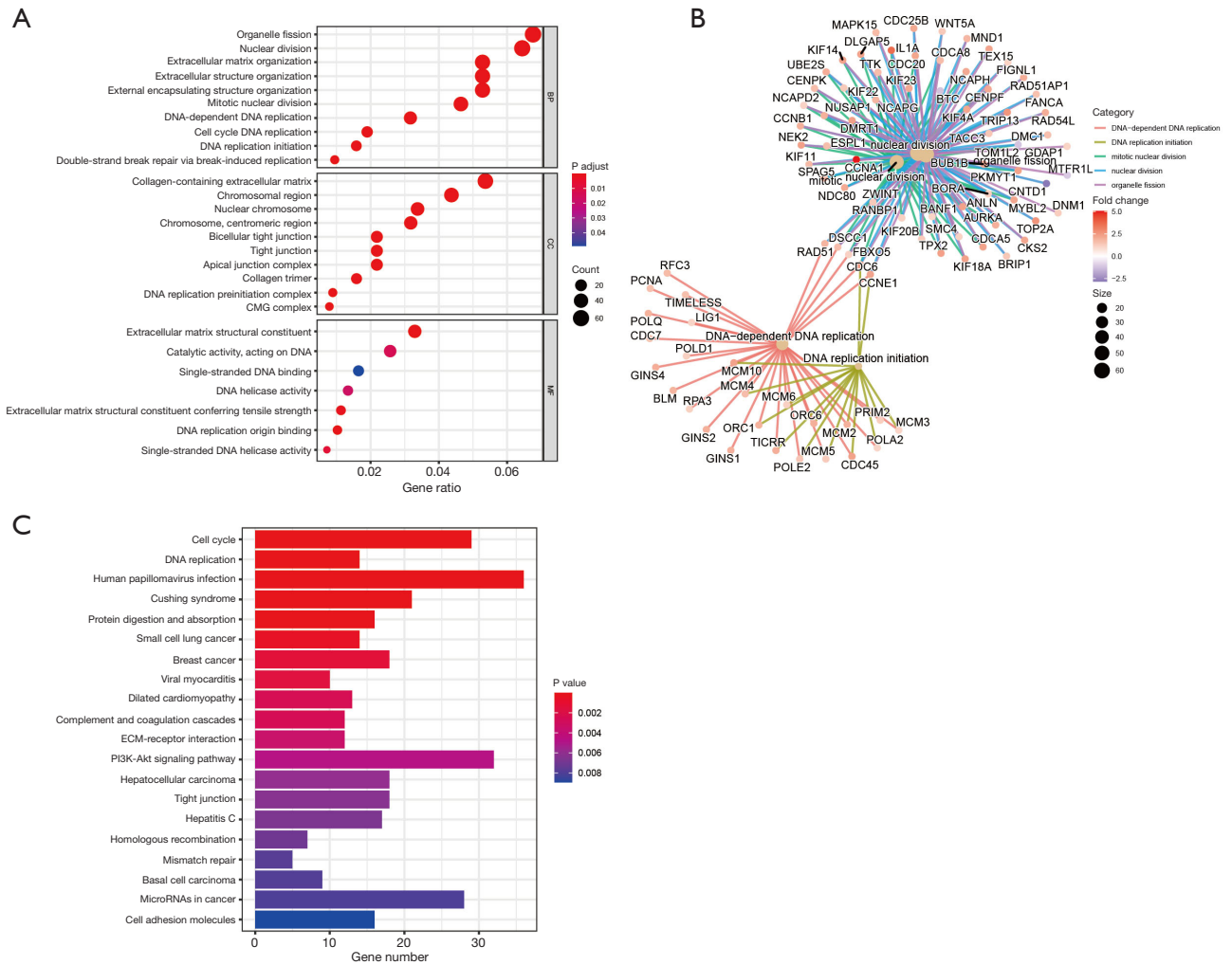

Figure 4 Enrichment functional analysis of the DESRGs. (A,B) GO analysis showing enrichment of the DESRGs in the BP, CC, and MF. The network graph shows the overall enrichment. (C) KEGG analysis of the DESRGs. BP, biological process; CC, cellular component; MF, molecular function; DESRGs, differentially expressed stemness-related genes; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

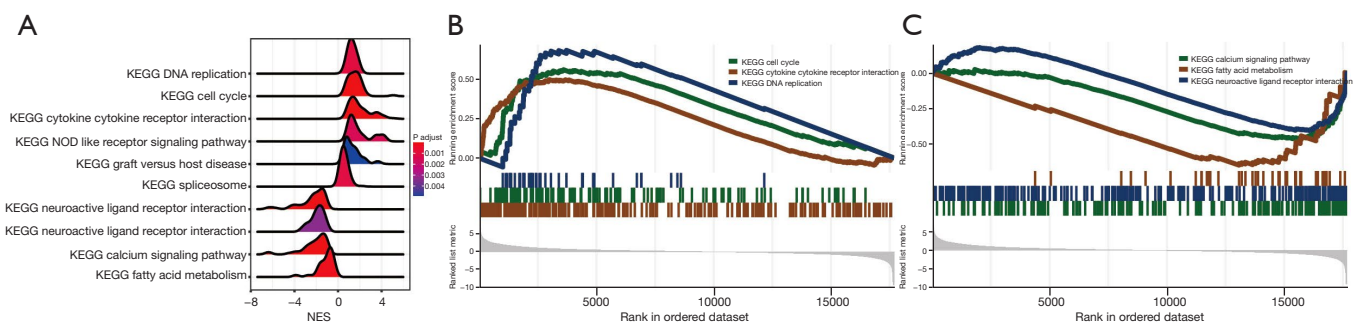

Figure 5 GSEA analysis. (A) The general enrichment pathway of the Mountain Map Exhibition. (B) The most significant top three pathways activated in ESCA. (C) The top three pathways most significantly inhibited in ESCA. NES, normalize enrichment score; GSEA, gene set enrichment analysis; ESCA, esophageal cancer.

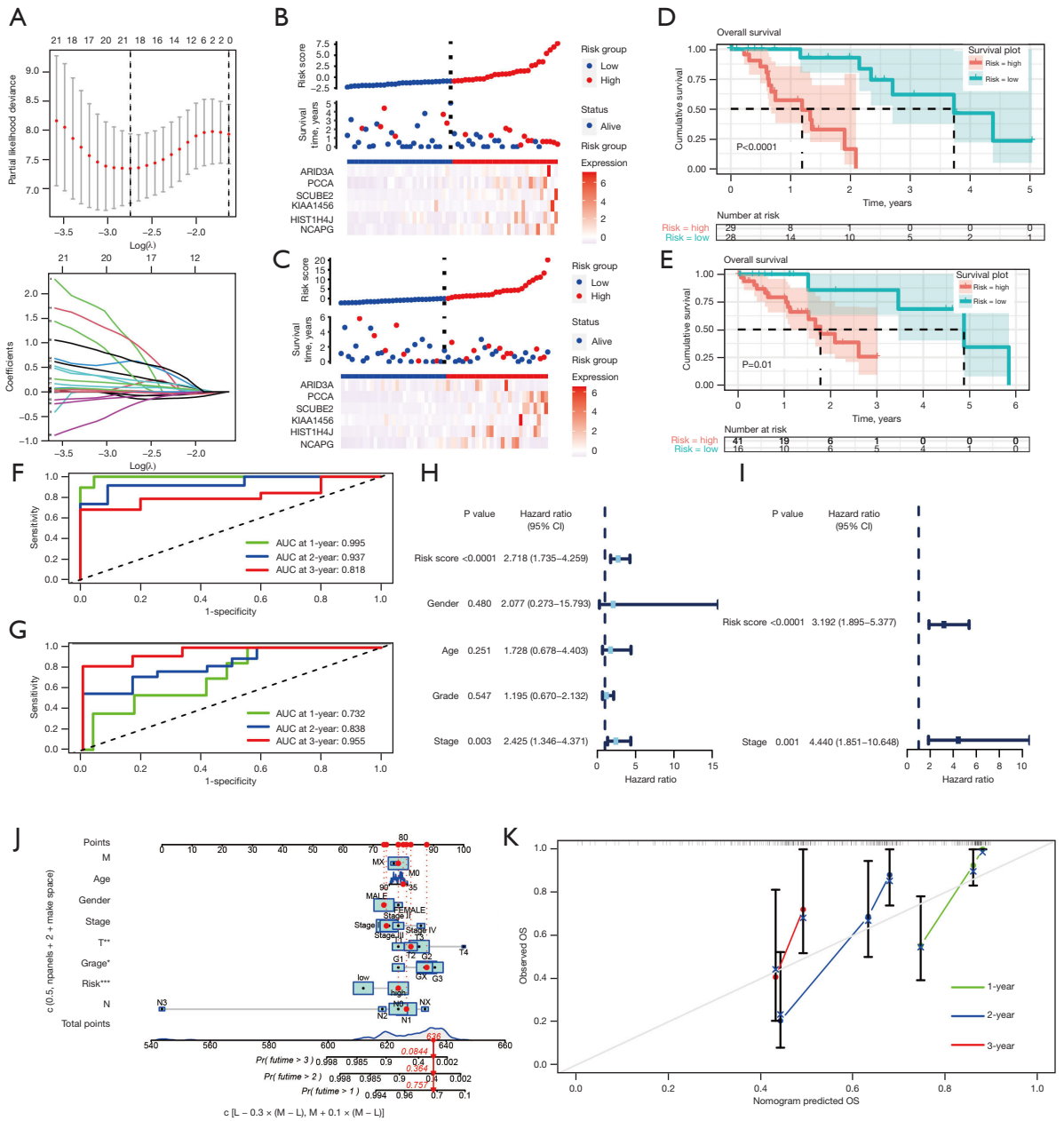

Figure 6 Construction and validation of the DESRG prognostic model. (A) LASSO regression analysis plot of meaningful genes in uni-Cox analysis. (B) A risk factor plot for the training set. (C) A risk factor plot for the validation set. The Kaplan-Meier survival curve analysis revealed an OS in the high-risk group from the training set (D) and the validation set (E). The red line represents the high-risk group. The green line represents the low-risk group. Survival-dependent ROC curves validated the prognostic significance of DESRGs as prognostic indicators in the training set (F) and the validation set (G). The green line represents the 1-year AUC; the blue line represents the 2-year AUC; and the red line represents the 3-year AUC. A forest plot showing the uni-Cox (H) and multi-Cox (I) regression analyses. (J) A nomogram for predicting the survival probability of ESCA patients for predicting 1-, 2-, and 3-year survival outcomes of ESCA. (K) Calibration curves of the OS-nomogram. *, P<0.05; **, P<0.01; ***, P<0.001. AUC, area under the ROC curve; ROC, receiver operating characteristic; CI, confidence interval; OS, overall survival; DESRG, differentially expressed stemness-related gene; LASSO, least absolute shrinkage and selection operator; uni-Cox, univariate Cox; multi-Cox, multivariate Cox; ESCA, esophageal cancer.

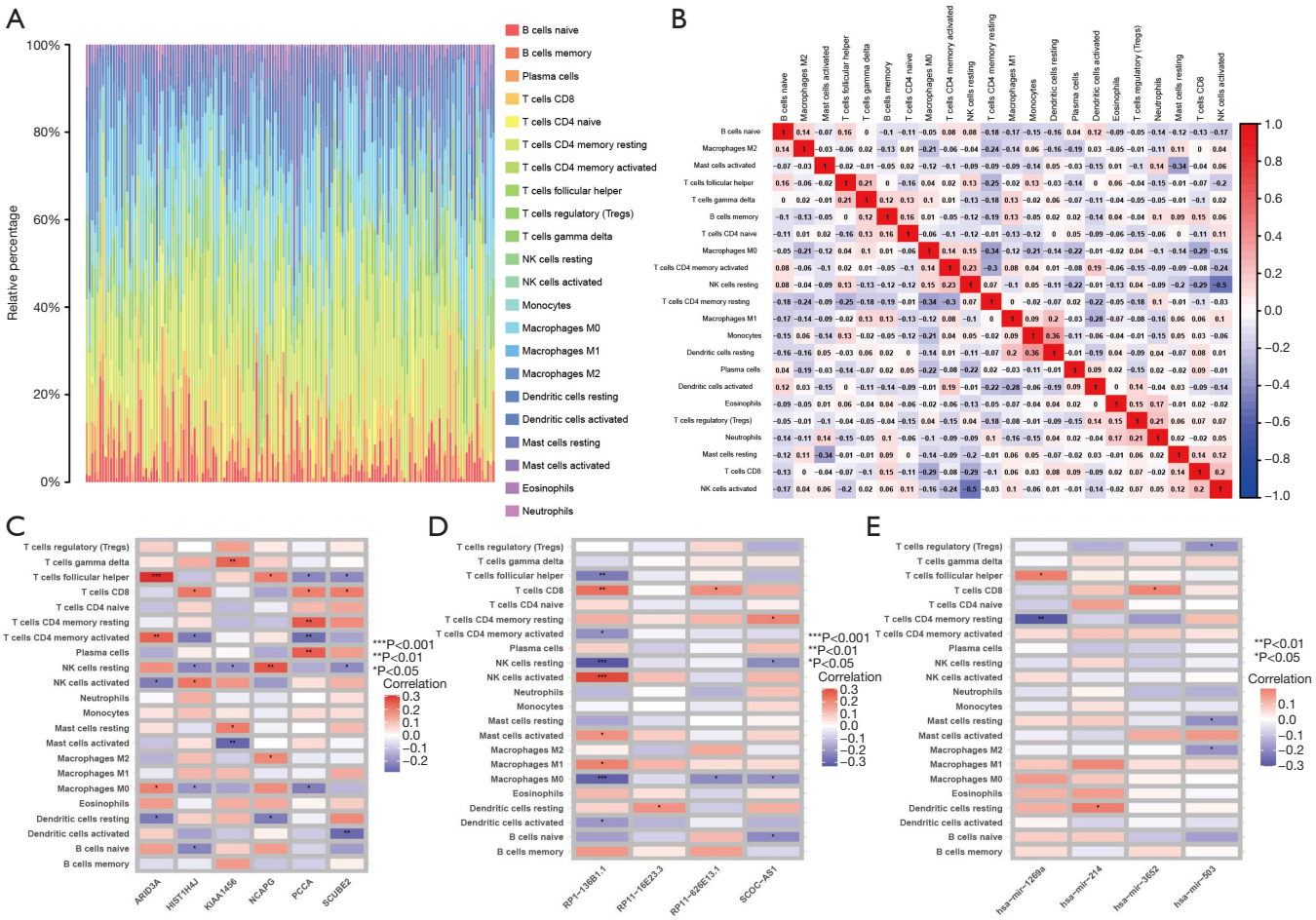

Figure 7 The landscape of immune infiltration in ESCA. (A) The difference in immune infiltration. (B) A correlation heat map of 22 types of immune cells. Red represents a positive correlation, blue represents a negative correlation. Candidate DESRGs (C), DESRlncRNAs (D), and DESRmiRNAs (E) that show significant positive correlations with immune infiltration cells are shown in red, whereas that show negative correlations with immune infiltration cells are shown in blue. NK, natural killer; ESCA, esophageal cancer; DESRG, differentially expressed stemness-related genes; DESRlncRNA, differentially expressed stemness-related long non-coding RNA; DESRmiRNA, differentially expressed stemness-related microRNA.

those for the validation group were 0.732, 0.838, and 0.955, respectively (*Figure 6G*). These results indicated that the DESRGs selected as prognostic indicators have robust potential for survival prediction in patients with ESCA.

Uni-Cox (*Figure 6H*) and multi-Cox (*Figure 6I*) regression analyses with clinical variables (risk score, sex, age, grade, and stage) revealed that the risk score and stage were independent prognostic indicators for patients with ESCA. A nomogram combining risk score and clinicopathological parameters was selected to predict ESCA patient prognosis (*Figure 6J*). Calibration curves were used to compare the actual probabilities of survival and the predicted survival rates. The results showed a significant correlation between the probability of OS and the actual survival rates for 1, 2, and 3 years (*Figure 6K*).

A prognostic model for the DESRlncRNAs and DESRmiRNAs was constructed and validated. The regression analysis identified *SCOC-AS1*, *RP11-626E13.1*, *RP1-136B1.1*, and *RP11-16E23.3* as meaningful lncRNAs for the prediction model. The risk score for DESRlncRNAs was calculated as follows: expression level of *SCOC-AS1* × 0.7209 + expression level of *RP11-626E13.1* × 2.2361 + expression level of *RP1-136B1.1* × 1.4229 + expression level of *RP11-16E23.3* × 6.7802. Furthermore,

the miRNAs selected for the risk analysis were *hsa-miR-269a*, *hsa-miR-214*, *hsa-miR-3652*, and *hsa-miR-503*. The DESRmiRNAs risk score was calculated as follows: expression level of *hsa-miR-1269a* × 0.003 + expression level of *hsa-miR-214* × 0.03181 + expression level of *hsa-miR-3652* × 0.1195 + expression level of *hsa-miR-503* × 0.04297. The risk score results for the DESRlncRNAs (Figure S2) and DESRmiRNAs (Figure S3) showed that the model provided high accuracy in predicting the prognosis of patients with ESCA.

Tumor-infiltrating immune cells

To analyze the effect of tumor-infiltrating immune cells on the model, the CIBERSORT algorithm was used to calculate the degree of infiltration of 22 types of infiltrating cells and display their overall distribution scores (*Figure 7A*) and correlations (*Figure 7B*). The correlation heatmap of the 22 types of infiltrating cells revealed that activated and resting natural killer (NK) cells, activated and resting mast cells, resting M0 macrophages, and $CD4^+$ memory T cells had significant negative correlations. Monocytes were positively correlated with resting dendritic cells.

The association between infiltrating levels of the immune cells and the expression of DESRGs (*Figure 7C*), DESRlncRNAs (*Figure 7D*), and DESRmiRNAs (*Figure 7E*) from patients with ESCA was analyzed. *ARID3A* was found to be positively correlated with infiltrating levels of immune cells, including follicular helper T cells (P<0.001) and activated $CD4^+$ memory T cells (P<0.01). *KIAA1456* expression was positively correlated with gamma delta (γδ) T cells (P<0.01) and negatively correlated with activated mast cells (P<0.01). The expression of *NCAPG* was positively correlated with resting NK cells (P<0.01). The expression of *PCCA* was positively correlated with resting $CD4^+$ memory T cells (P<0.01) and plasma cells (P<0.01), and negatively correlated with activated $CD4^+$ memory T cells (P<0.01). *SCUBE2* expression was negatively correlated with activated dendritic cells (P<0.01). *RP1-136B1.1* levels were positively correlated with activated NK cells (P<0.001) and $CD8^+$ T cells (P < 0.01), and negatively correlated with follicular helper T cells (P<0.01), resting NK cells (P<0.001), and M0 macrophage (P<0.001). Finally, *hsa-miR-1269a* expression was negatively correlated with resting $CD4^+$ memory T cells (P<0.01).

Discussion

To the best of our knowledge, this is the first study to perform a comprehensive characterization of ESCA stemness based on combined analyses of mRNA, miRNA, and lncRNA expression in ESCA tumors and normal tissues. A total of 1,106 DESRGs, 84 DESRmiRNAs, and 320 DESRlncRNAs were identified using the TCGA-ESCA dataset.

Non-coding RNAs, including miRNAs and lncRNAs, play important roles in transcriptional and translational regulation in several cancers (43). lncRNAs act as miRNA sponges to reduce their regulatory abilities (44). Interactions between lncRNAs, miRNAs, and mRNAs form ceRNA networks (45). Understanding the role of ceRNA networks in pan-cancer analysis is crucial (46). However, to date, no such study has been conducted on stemness-related ceRNA networks in ESCA. In this present investigation, we constructed a stemness-related ceRNA network based on 55 DESRGs, 17 DESRmiRNAs, and 44 DESRlncRNAs, and analyzed its correlation with the survival data of patients with ESCA. *PRICKLE2-AS1*, *FAM13A-AS1*, *C15orf54*, and *LINC00304* were identified as high-risk DESRlncRNAs that correlated with high ESCA severity. Previous study has demonstrated that *LINC00304* directly promotes cell proliferation and cell cycle progression in prostate cancer (47). *PRICKLE2-AS1* has been shown to be differentially methylated in vulvar squamous cell carcinoma (48) and *FAM13A-AS1* is involved in regulating RNA splicing and mRNA processing in thyroid cancer (49). Herein, *PRICKLE2-AS1*, *FAM13A-AS1*, and *LINC00304* were shown to be downregulated in ESCA tumors, whereas *C15orf54* was upregulated (log_2FC =1.51; q value =0.02). *C15orf54* is known to be positively associated with high-grade gastric cancer (50) and indeed, *C15orf54* may be a potential biomarker for ESCA prognosis.

NCAPG is one of the DESRGs selected to build the prediction model. Increased expression of *NCAPG* promotes oncogenesis in non-small cell lung cancer (51), glioma (52), lung adenocarcinoma (53), prostate cancer (54), and esophageal squamous cell carcinoma (55). *NCAPG* (log_2FC =1.81; q value =2.68×10−11) was significantly upregulated in ESCA and positively correlated with resting NK cells, suggesting that human NK cells rest via the overexpression of *NCAPG* in ESCA. Candidate DESRGs, DESRlncRNAs, and DESRmiRNAs that show significant positive correlations with immune infiltration cells, increased expression of SRGs and increased immune infiltration may be related to poor prognosis. Notably, the oncogene *NCAPG* in prostate cancer is directly targeted by the antitumor *hsa-miR-145-3p*, which is associated with cancer pathogenesis (54). In the present study, we found that hsa-miR-1269a was significantly upregulated

in ESCA tumors (log_2FC =3.98; q value =2.92×10^{-4}) with poor prognostic features. Additionally, the number of resting $CD4^+$ memory T cell as negatively corrected with *hsa-miR-1269a* (P<0.01). *hsa-miR-145-3p* is known as a diagnostic and oncogenic marker in hepatocellular carcinoma (56), colorectal cancer (57), and non-small cell lung cancer (58), and it may be involved in the molecular pathogenesis of various cancers, including ESCA (55).

ARID3A acts as an oncogene, and its overexpression affects the development of colorectal cancer (59), ovarian cancer (60), gastric cancer (61), hepatocellular carcinoma (62), and rectal cancer (63). This report revealed that *ARID3A* was upregulated and positively associated with infiltrating follicular helper T cells and activated $CD4^+$ memory T cells, and thus, may play an immunological role in ESCA. Follicular helper T cells are a subset of T-helper cells, which provide critical support to B cells and are essential for maintaining humoral immune responses (64), such as germinal center formation, affinity maturation, and the development of high-affinity antibodies (65). Therefore, *ARID3A* is a promising independent prognostic biomarker corresponding to follicular helper T cells and activated $CD4^+$ memory T cells.

A previous report showed that a *de novo* variant in the human *HIST1H4J* gene causes a syndrome analogous to *HIST1H4C*-associated neurodevelopmental disorder (66). Furthermore, *HIST1H4J* is the strongest negative predictor of ovarian cancer (67) and an important prognostic and chemoresistance predictor of ESCA (68). Herein, upregulation expression of *HIST1H4J* (log_2FC =1.39; q value =1.87×10^{-3}) was observed in patients with ESCA.

Hsa-miR-1269a serves as a diagnostic and oncogenic marker in hepatocellular carcinoma (56), colorectal cancer (57) and non - small cell lung cancer (58). Our investigations showed that *hsa-miR-1269a* is significantly upregulated in ESCA patients with poor prognostic features. $CD4^+$ resting memory T cells (P<0.01) were significantly negatively correlated with *hsa-miR-1269a*.

This is the first study to perform a comprehensive characterization of ESCA stemness based on a combined analysis of mRNA, miRNA, and lncRNA expression, and correlation with prognosis prediction. SRG network will facilitate improved treatment regimens for patients with ESCA and may promote the development of innovative therapeutics in the future.

Acknowledgments

We thank the editor and the four anonymous reviewers for their thoughtful comments.

Funding: This project was supported by the Talent Scientific Research Start-Up Foundation of Yijishan Hospital, Wannan Medical College (Grant Nos. YR202001 and YR201806), the Opening Foundation of the Key Laboratory of Non-Coding RNA Transformation Research of the Anhui Higher Education Institution (Grant No. RNA202004), and the Key Projects of Natural Science Research of Universities in Anhui Province (Grant No. KJ2020A0622).

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://tcr.amegroups.com/article/view/10.21037/tcr-22-1723/rc

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://tcr.amegroups.com/article/view/10.21037/tcr-22-1723/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

References

1. Lundberg E, Lagergren P, Mattsson F, et al. Life Expectancy in Survivors of Esophageal Cancer Compared with the Background Population. Ann Surg Oncol 2022;29:2805-11.
2. Li K, Chen J, Lou X, et al. HNRNPA2B1 Affects the Prognosis of Esophageal Cancer by Regulating the miR-17-92 Cluster. Front Cell Dev Biol 2021;9:658642.

3. Li X, Chen L, Luan S, et al. The development and progress of nanomedicine for esophageal cancer diagnosis and treatment. Semin Cancer Biol 2022. [Epub ahead of print]. doi: 10.1016/j.semcancer.2022.01.007.
4. Zayac A, Almhanna K. Esophageal, gastric cancer and immunotherapy: small steps in the right direction? Transl Gastroenterol Hepatol 2020;5:9.
5. Ai D, Ye J, Wei S, et al. Comparison of 3 Paclitaxel-Based Chemoradiotherapy Regimens for Patients With Locally Advanced Esophageal Squamous Cell Cancer: A Randomized Clinical Trial. JAMA Netw Open 2022;5:e220120.
6. Shim YM, Yun J, Im YH, et al. The efficacy of adjuvant chemotherapy with capecitabine and cisplatin after surgery in locally advanced esophageal squamous cell carcinoma: a multicenter randomized phase III trial. Dis Esophagus 2022;35:doab040.
7. Harada K, Yamamoto S, Kato K. Pembrolizumab for the treatment of advanced esophageal cancer. Future Oncol 2022;18:2311-9.
8. Najafi M, Farhood B, Mortezaee K. Cancer stem cells (CSCs) in cancer progression and therapy. J Cell Physiol 2019;234:8381-95.
9. Chen D, Liu J, Zang L, et al. Integrated Machine Learning and Bioinformatic Analyses Constructed a Novel Stemness-Related Classifier to Predict Prognosis and Immunotherapy Responses for Hepatocellular Carcinoma Patients. Int J Biol Sci 2022;18:360-73.
10. Zeng H, Ji J, Song X, et al. Stemness Related Genes Revealed by Network Analysis Associated With Tumor Immune Microenvironment and the Clinical Outcome in Lung Adenocarcinoma. Front Genet 2020;11:549213.
11. Yin P, Bai Y, Wang Z, et al. Non-canonical Fzd7 signaling contributes to breast cancer mesenchymal-like stemness involving Col6a1. Cell Commun Signal 2020;18:143.
12. Wei R, Quan J, Li S, et al. Development and Validation of a Unique Signature of Cancer Stemness-Related Genes for Predicting the Overall Survival of Colorectal Cancer Patients. SSRN 2020. doi: 10.2139/ssrn.3772767.
13. Zhuang J, Li M, Zhang X, et al. Construction of Bone Metastasis-Specific Regulation Network Based on Prognostic Stemness-Related Signatures in Prostate Cancer. Dis Markers 2022;2022:8495923.
14. Ho DW, Tsui YM, Sze KM, et al. Single-cell transcriptomics reveals the landscape of intra-tumoral heterogeneity and stemness-related subpopulations in liver cancer. Cancer Lett 2019;459:176-85.
15. Guo H, Wang S, Ju M, et al. Identification of Stemness-Related Genes for Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma by Integrated Bioinformatics Analysis. Front Cell Dev Biol 2021;9:642724.
16. Wang W, Xu C, Ren Y, et al. A Novel Cancer Stemness-Related Signature for Predicting Prognosis in Patients with Colon Adenocarcinoma. Stem Cells Int 2021;2021:7036059.
17. Liu Y, Wang J, Li L, et al. AC010973.2 promotes cell proliferation and is one of six stemness-related genes that predict overall survival of renal clear cell carcinoma. Sci Rep 2022;12:4272.
18. Jiang W, Xie N, Xu C. Characterization of a prognostic model for lung squamous cell carcinoma based on eight stemness index-related genes. BMC Pulm Med 2022;22:224.
19. Lu YJ, Lian L, Shen XM, et al. A bioinformatics analysis to evaluate the prognostic value of stemness-related genes in gastric cancer. Transl Cancer Res 2021;10:174-83.
20. Yi L, Huang P, Zou X, et al. Integrative stemness characteristics associated with prognosis and the immune microenvironment in esophageal cancer. Pharmacol Res 2020;161:105144.
21. Hong L, Zhou Y, Xie X, et al. A stemness-based eleven-gene signature correlates with the clinical outcome of hepatocellular carcinoma. BMC Cancer 2021;21:716.
22. Chi J, Zhang H, Hu J, et al. AGR3 promotes the stemness of colorectal cancer via modulating Wnt/ -catenin signalling. Cell Signal 2020;65:109419.
23. Müller L, Tunger A, Plesca I, et al. Bidirectional Crosstalk Between Cancer Stem Cells and Immune Cell Subsets. Front Immunol 2020;11:140.
24. Colaprico A, Silva TC, Olsen C, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res 2016;44:e71.
25. Jiang L, Zhong M, Chen T, et al. Gene regulation network analysis reveals core genes associated with survival in glioblastoma multiforme. J Cell Mol Med 2020;24:10075-87.
26. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.
27. Wickham H, Chang W, Wickham MH. Package 'ggplot2'. 2014. Available online: https://cran.microsoft.com/snapshot/2015-01-06/web/packages/ggplot2/ggplot2.pdf
28. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl

Acad Sci U S A 2005;102:15545-50.
29. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics 2013;14:7.
30. von Mering C, Huynen M, Jaeggi D, et al. STRING: a database of predicted functional associations between proteins. Nucleic Acids Res 2003;31:258-61.
31. Doncheva NT, Morris JH, Gorodkin J, et al. Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. J Proteome Res 2019;18:623-32.
32. Jeggari A, Marks DS, Larsson E. miRcode: a map of putative microRNA target sites in the long non-coding transcriptome. Bioinformatics 2012;28:2062-3.
33. McGeary SE, Lin KS, Shi CY, et al. The biochemical basis of microRNA targeting efficacy. Science 2019;366:eaav1741.
34. Wong N, Wang X. miRDB: an online resource for microRNA target prediction and functional annotations. Nucleic Acids Res 2015;43:D146-52.
35. Hsu SD, Lin FM, Wu WY, et al. miRTarBase: a database curates experimentally validated microRNA-target interactions. Nucleic Acids Res 2011;39:D163-9.
36. Harris MA, Clark J, Ireland A, et al. The Gene Ontology (GO) database and informatics resource. Nucleic Acids yoto2004;32:D258-61.
37. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28:27-30.
38. Wu T, Hu E, Xu S, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb) 2021;2:100141.
39. Subramanian A, Kuehn H, Gould J, et al. GSEA-P: a desktop application for Gene Set Enrichment Analysis. Bioinformatics 2007;23:3251-3.
40. Kassambara A, Kosinski M, Biecek P, et al. Package 'survminer'. 2017. Available online: https://cran.microsoft.com/snapshot/2017-04-21/web/packages/survminer/survminer.pdf
41. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw 2010;33:1-22.
42. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015;12:453-7.
43. Wang M, Mao C, Ouyang L, et al. Long noncoding RNA LINC00336 inhibits ferroptosis in lung cancer by functioning as a competing endogenous RNA. Cell Death Differ 2019;26:2329-43.
44. Sun P, Hamblin MH, Yin KJ. Non-coding RNAs in the regulation of blood-brain barrier functions in central nervous system disorders. Fluids Barriers CNS 2022;19:27.
45. Lin P, Wen DY, Li Q, et al. Genome-Wide Analysis of Prognostic lncRNAs, miRNAs, and mRNAs Forming a Competing Endogenous RNA Network in Hepatocellular Carcinoma. Cell Physiol Biochem 2018;48:1953-67.
46. Zhang Y, Han P, Guo Q, et al. Oncogenic Landscape of Somatic Mutations Perturbing Pan-Cancer lncRNA-ceRNA Regulation. Front Cell Dev Biol 2021;9:658346.
47. Kumar S, Prajapati KS, Singh AK, et al. Long non-coding RNA regulating androgen receptor signaling in breast and prostate cancer. Cancer Lett 2021;504:15-22.
48. Dasgupta S, Ewing-Graham PC, Swagemakers SMA, et al. Exploring Differentially Methylated Genes in Vulvar Squamous Cell Carcinoma. Cancers (Basel) 2021;13:3580.
49. Rao Y, Liu H, Yan X, et al. In Silico Analysis Identifies Differently Expressed lncRNAs as Novel Biomarkers for the Prognosis of Thyroid Cancer. Comput Math Methods Med 2020;2020:3651051.
50. Liu M, Li J, Huang Z, et al. Gastric cancer risk-scoring system based on analysis of a competing endogenous RNA network. Transl Cancer Res 2020;9:3889-902.
51. Sun H, Zhang H, Yan Y, et al. NCAPG promotes the oncogenesis and progression of non-small cell lung cancer cells through upregulating LGALS1 expression. Mol Cancer 2022;21:55.
52. Zheng G, Han T, Hu X, et al. NCAPG Promotes Tumor Progression and Modulates Immune Cell Infiltration in Glioma. Front Oncol 2022;12:770628.
53. Wang X, Tian X, Sui X, et al. Increased expression of NCAPG (Non-SMC condensing I complex subunit G) is associated with progression and poor prognosis of lung adenocarcinoma. Bioengineered 2022;13:6113-25.
54. Shimonosono M, Idichi T, Seki N, et al. Molecular pathogenesis of esophageal squamous cell carcinoma: Identification of the antitumor effects of miR-145-3p on gene regulation. Int J Oncol 2019;54:673-88.
55. Chen N, Zhang G, Fu J, et al. Identification of Key Modules and Hub Genes Involved in Esophageal Squamous Cell Carcinoma Tumorigenesis Using WCGNA. Cancer Control 2020;27:1073274820978817.
56. Hu Y, Dingerdissen H, Gupta S, et al. Identification of key differentially expressed MicroRNAs in cancer patients through pan-cancer analysis. Comput Biol Med 2018;103:183-97.
57. Qin S, Shi X, Wang C, et al. Transcription Factor and miRNA Interplays Can Manifest the Survival of ccRCC Patients. Cancers (Basel) 2019;11:1668.

58. Wang X, Jiang X, Li J, et al. Serum exosomal miR-1269a serves as a diagnostic marker and plays an oncogenic role in non-small cell lung cancer. Thorac Cancer 2020;11:3436-47.
59. Tang J, Yang L, Li Y, et al. ARID3A promotes the development of colorectal cancer by upregulating AURKA. Carcinogenesis 2021;42:578-86.
60. Dausinas P, Pulakanti K, Rao S, et al. ARID3A and ARID3B induce stem promoting pathways in ovarian cancer cells. Gene 2020;738:144458.
61. Xu G, Li K, Zhang N, et al. Screening Driving Transcription Factors in the Processing of Gastric Cancer. Gastroenterol Res Pract 2016;2016:8431480.
62. Wang X, Zhou Y, Dong K, et al. Exosomal lncRNA HMMR-AS1 mediates macrophage polarization through miR-147a/ARID3A axis under hypoxia and affects the progression of hepatocellular carcinoma. Environ Toxicol 2022;37:1357-72.
63. Zhou D, Ji G, Wei G, et al. MiR-361-3p promotes tumorigenesis of osteosarcoma cells via targeting ARID3A. Tissue Cell 2022;76:101759.
64. Zhang X, Ing S, Fraser A, et al. Follicular helper T cells: new insights into mechanisms of autoimmune diseases. Ochsner J 2013;13:131-9.
65. Ise W. Development and function of follicular helper T cells. Biosci Biotechnol Biochem 2016;80:1-6.
66. Tessadori F, Rehman AU, Giltay JC, et al. A de novo variant in the human HIST1H4J gene causes a syndrome analogous to the HIST1H4C-associated neurodevelopmental disorder. Eur J Hum Genet 2020;28:674-8.
67. Bachmayr-Heyda A, Aust S, Auer K, et al. Integrative Systemic and Local Metabolomics with Impact on Survival in High-Grade Serous Ovarian Cancer. Clin Cancer Res 2017;23:2081-92.
68. Zhang G, Zhang Y, Yang H, et al. Genome-Wide Profiling of a Prognostic RNA-Binding Protein Signature in Esophageal Cancer. Research Square 2022. doi: 10.21203/rs.3.rs-1439005/v1.

(English Language Editor: J. Teoh)

Cite this article as: Zhu S, Zhang G, You Q, Li F, Ding B, Liu F, Bi D, Jiang L. Stemness-related gene signature for predicting therapeutic response in patients with esophageal cancer. Transl Cancer Res 2022;11(7):2359-2373. doi: 10.21037/tcr-22-1723

Table S1 mRNA, lncRNA, miRNA for uni-Cox analysis

Type	Gene	HR	Lower	Upper	P value
mRNA	*PODN*	1.00324714	1.000381736	1.006120751	0.026317216
mRNA	*FCGR1B*	1.225197756	1.087937253	1.379775843	0.000807375
mRNA	*S100A6*	1.000132726	1.000020772	1.000244691	0.020144423
mRNA	*MFSD4*	1.003896826	1.0010547	1.006747021	0.007172789
mRNA	*RD3*	9.686973427	1.222493358	76.75907079	0.03154037
mRNA	*MAP1LC3C*	9.909903301	2.448197547	40.11366793	0.001304158
mRNA	*LBH*	1.002492379	1.000798014	1.004189613	0.003923848
mRNA	*PTH2R*	1.024213254	1.00595957	1.042798162	0.009118662
mRNA	*GBX2*	1.162964375	1.03895711	1.301772831	0.008683245
mRNA	*P2RY14*	1.034625196	1.002493988	1.06778625	0.034454507
mRNA	*ECT2*	1.000884699	1.000119407	1.001650577	0.023457019
mRNA	*NCAPG*	1.009635496	1.003840711	1.015463733	0.001093683
mRNA	*MTRNR2L13*	11.96603594	1.998156885	71.65904596	0.00656814
mRNA	*BRIX1*	1.00601046	1.00209252	1.009943718	0.002613369
mRNA	*PLCXD3*	1.006577701	1.001103146	1.012082193	0.018462868
mRNA	*ANXA2R*	1.041780733	1.013337135	1.071022721	0.003755365
mRNA	*F2R*	1.001167104	1.000163813	1.002171401	0.022597905
mRNA	*TMEM232*	1.316746374	1.094985569	1.583419054	0.003452075
mRNA	*SOWAHA*	1.033868334	1.003405156	1.065256367	0.029055336
mRNA	*ECI2*	1.009864713	1.002134859	1.01765419	0.012281938
mRNA	*HIST1H4J*	1.564509606	1.072090069	2.28310137	0.020290381
mRNA	*ATP6V1G2*	1.4873922	1.04772262	2.111566091	0.026369744
mRNA	*FBXO5*	1.013126265	1.002037096	1.024338153	0.020213059
mRNA	*STEAP1B*	1.041449037	1.000369087	1.084215926	0.047935932
mRNA	*STK31*	1.022053615	1.003115979	1.041348771	0.02225451
mRNA	*KIAA1456*	1.024296663	1.009681504	1.039123378	0.001060423
mRNA	*MBOAT4*	1.87459532	1.206266256	2.913210576	0.005211303
mRNA	*TEX15*	1.039375982	1.005666937	1.074214922	0.021681647
mRNA	*ABRA*	1.992266299	1.095058859	3.624576864	0.02398563
mRNA	*PSCA*	1.001112083	1.000148391	1.002076704	0.023701718
mRNA	*TPM2*	1.001297589	1.000140327	1.00245619	0.02796486
mRNA	*FNBP1*	1.003245661	1.000449606	1.006049531	0.022867729
mRNA	*IFIT1*	1.005054142	1.00065053	1.009477132	0.024434132
mRNA	*SVIP*	1.003810442	1.000194377	1.007439581	0.038873821
mRNA	*MS4A4E*	1.517375598	1.027378242	2.241072092	0.036107518
mRNA	*BANF1*	1.00250939	1.000946599	1.004074621	0.001640415
mRNA	*CST6*	1.023088654	1.004576641	1.041941801	0.014282239
mRNA	*CARNS1*	1.012654989	1.000820463	1.024629456	0.036019953
mRNA	*NOX4*	1.035195951	1.006147711	1.065082836	0.017218103
mRNA	*PCED1B*	1.020595086	1.000004332	1.041609817	0.049951328
mRNA	*ADCY6*	1.004574666	1.001432996	1.007726192	0.004290194
mRNA	*FAM19A2*	1.04124358	1.009978676	1.07347632	0.009368428
mRNA	*FGD6*	1.003877446	1.000994344	1.006768853	0.00835838
mRNA	*PXMP2*	1.032804175	1.008270549	1.057934763	0.008502061
mRNA	*PCCA*	1.009721838	1.001541774	1.017968713	0.019744437
mRNA	*RNASE4*	1.058263055	1.014720652	1.103673894	0.008250202
mRNA	*CHRNA7*	1.090171138	1.011816488	1.17459354	0.023289605
mRNA	*MYZAP*	1.045775929	1.015019956	1.077463835	0.003294647
mRNA	*CPEB1*	1.261233732	1.040651504	1.528571785	0.017971125
mRNA	*NETO2*	1.008593485	1.002321789	1.014904424	0.007174012
mRNA	*CES5A*	1.055968764	1.020003163	1.093202522	0.002068794
mRNA	*MT3*	1.871944415	1.010177739	3.468870632	0.046355814
mRNA	*FOXL1*	1.026298867	1.009175703	1.043712567	0.002494814
mRNA	*WSCD1*	1.04274854	1.006289644	1.08052838	0.021152414
mRNA	*FBXO39*	1.254608709	1.017668813	1.546714404	0.033672697
mRNA	*CLDN7*	1.000861121	1.000009453	1.001713515	0.047509668
mRNA	*GRIN3B*	1.040611102	1.008474328	1.07377197	0.012875002
mRNA	*GNG7*	1.02878942	1.004691234	1.053465618	0.018926503
mRNA	*ZNF69*	1.065589813	1.016451225	1.117103923	0.008354967
mRNA	*ZNF568*	1.103815426	1.037505847	1.174363015	0.001779274
mRNA	*ZNF677*	1.03702122	1.009900582	1.064870176	0.007175125
mRNA	*ZNF583*	1.055045041	1.012046614	1.099870326	0.011602025
mRNA	*ZFP28*	1.060267265	1.018434445	1.103818393	0.004380818
mRNA	*ZNF470*	1.069066834	1.009432604	1.132224074	0.022575092
mRNA	*PCNA*	1.00068121	1.000036564	1.001326271	0.038343032
mRNA	*FOXS1*	1.081767842	1.019362105	1.147994083	0.009527507
mRNA	*CPNE1*	1.000270287	1.0001006	1.000440003	0.001795694
mRNA	*OSM*	1.076369433	1.004919899	1.152899009	0.035727969
mRNA	*PNPLA5*	1.835240896	1.274827676	2.642011315	0.001090493
mRNA	*CSF2RA*	1.005326399	1.001445725	1.00922211	0.007100766
mRNA	*REPS2*	1.005841152	1.000026981	1.011689126	0.048943231
mRNA	*MTRNR2L10*	1.475029014	1.075887699	2.022246928	0.015764606
lncRNA	*CENPI*	0.956151986	0.91489512	0.999269315	0.046322293
lncRNA	*C15orf54*	13.70002718	2.445218728	76.75826404	0.002911662
lncRNA	*FAM13A-AS1*	1.164985633	1.049213514	1.293532256	0.004242355
lncRNA	*RP11-348N5.7*	1.146861831	1.039238092	1.265631109	0.006420873
lncRNA	*RP11-182J1.17*	3.783981949	1.414301714	10.1240911	0.008041888
lncRNA	*RP11-636O21.1*	2.195423447	1.203686861	4.004267447	0.01033054
lncRNA	*RP11-522B15.3*	1.030207733	1.006233497	1.054753172	0.013240969
lncRNA	*ZFHX4-AS1*	1.66609615	1.110198097	2.500343306	0.013713115
lncRNA	*RP11-510N19.5*	1.030398167	1.005990913	1.055397588	0.014352505
lncRNA	*RP11-626E13.1*	2.746290777	1.143831247	6.593728801	0.023779145
lncRNA	*RP11-462G2.1*	1.007757162	1.001001925	1.014557987	0.024335659
lncRNA	*PRICKLE2-AS1*	0.005377968	4.50E-05	0.642647288	0.032262757
lncRNA	*RP11-266N13.2*	7.51993236	1.166039388	48.49697469	0.03388058
lncRNA	*RP1-136B1.1*	1.398188357	1.024293439	1.908565073	0.034759448
lncRNA	*RP5-1159O4.1*	1.079643763	1.005336208	1.159443623	0.035183321
lncRNA	*MAMDC2-AS1*	1.221694334	1.009639103	1.478287679	0.039533938
lncRNA	*LINC00304*	3.892700743	1.060305649	14.29127449	0.040539274
lncRNA	*RP6-74O6.6*	1.081829563	1.002939965	1.166924487	0.041755029
lncRNA	*RP11-325K4.2*	1.118697393	1.003612719	1.246978874	0.042859978
lncRNA	*RP11-1399P15.1*	1.94707145	1.009623778	3.754950422	0.04675119
lncRNA	*RP11-79H23.3*	1.064936777	1.000740638	1.13325101	0.047333628
lncRNA	*RP3-426I6.6*	2.366105534	1.004367295	5.574111607	0.048846213
lncRNA	*RP11-139H15.7*	3.133567638	1.004772391	9.772607439	0.049048698
miRNA	*hsa-miR-101-1*	0.999662129	0.999343099	0.999981262	0.037984048
miRNA	*hsa-miR-101-2*	0.999675281	0.999363188	0.999987472	0.041489546
miRNA	*hsa-miR-3074*	1.058380837	1.010632503	1.108385089	0.01599622
miRNA	*hsa-miR-3651*	1.084241971	1.026215504	1.145549493	0.003950585
miRNA	*hsa-miR-3652*	1.08412749	1.006833747	1.167355005	0.032319846
miRNA	*hsa-miR-421*	1.044978925	1.00639907	1.085037721	0.02188809
miRNA	*hsa-miR-503*	1.030179924	1.007310327	1.053568743	0.009435251

mRNA, messenger RNA; lncRNA, long non-coding RNA; miRNA, microRNA; uni-Cox, univariate Cox; HR, hazard ratio.

https://dx.doi.org/10.21037/tcr-22-1723

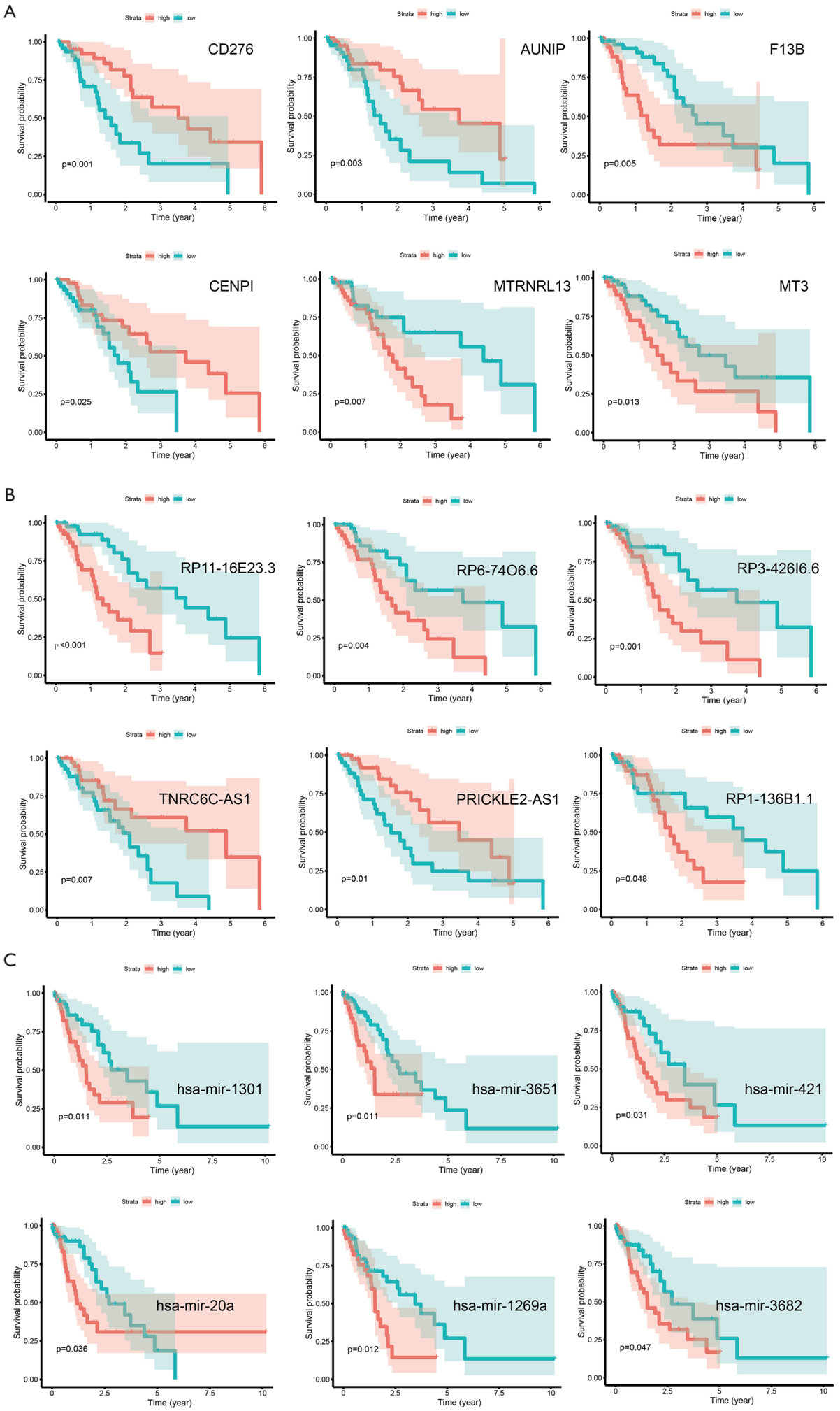

Figure S1 Survival analysis of the DESRGs (A), the DESRlncRNAs (B), and the DESRmiRNAs (C). Red indicates high expression group, light green indicates low expression group. DESRGs, differentially expressed stemness-related genes; DESRlncRNA, differentially expressed stemness-related long non-coding RNA; DESRmiRNA, differentially expressed stemness-related microRNA.

https://dx.doi.org/10.21037/tcr-22-1723

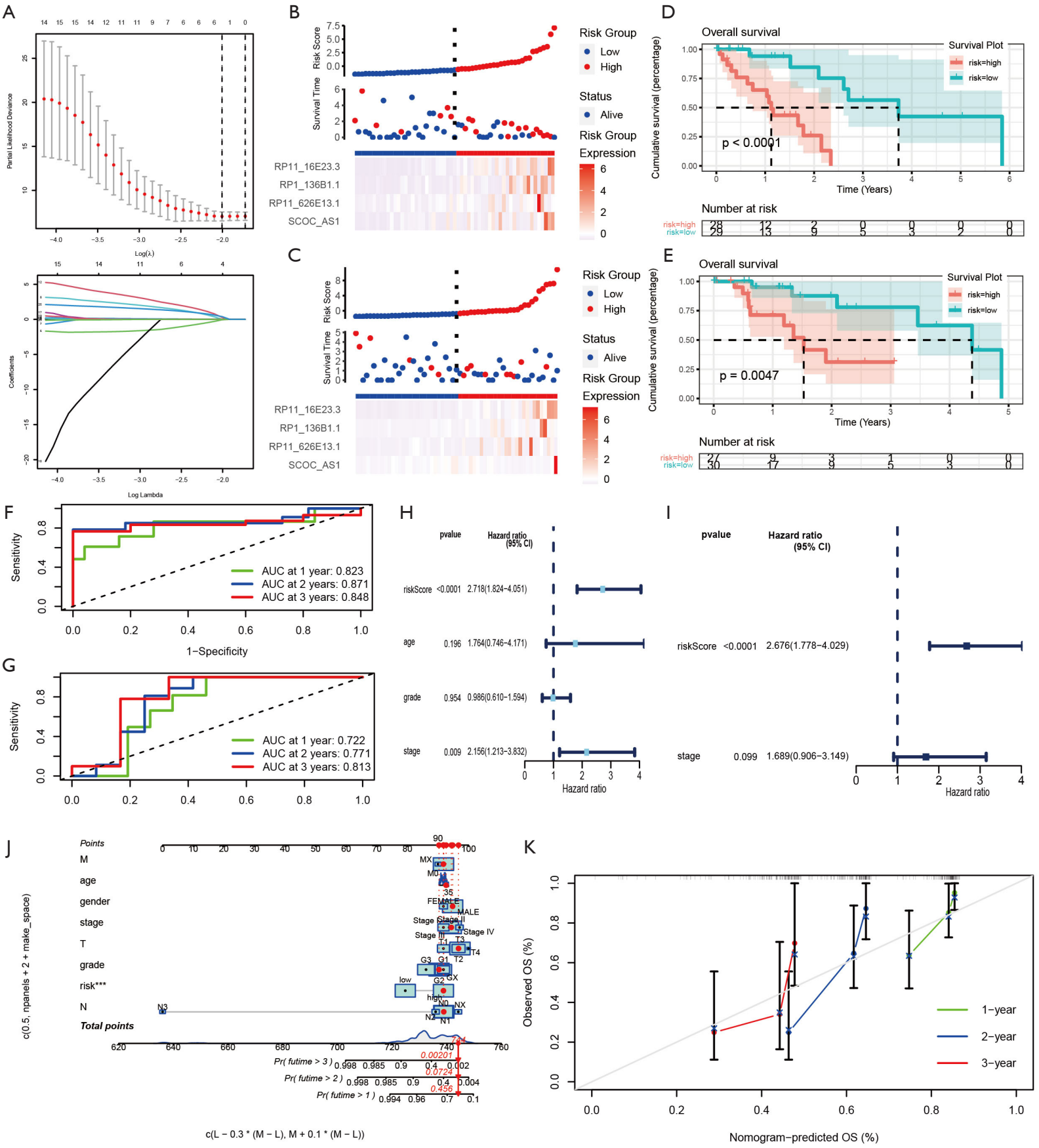

Figure S2 Construction and validation of the DESRlncRNA prognostic model. (A) LASSO regression analysis plot of the meaningful genes in the uni-Cox analysis. (B) A risk factor plot for the training set. (C) A risk factor plot for the validation set. The Kaplan-Meier survival curve analysis revealed an OS in the high-risk group from the training set (D) and the validation set (E). The red line represents the high-risk group. The green line represents the low-risk group. Survival-dependent ROC curves validated the prognostic significance of DESRlncRNAs as prognostic indicators in the training set (F) and the validation set (G). The green line represents the 1-year AUC; the blue line represents the 2-year AUC; and the red line represents the 3-year AUC. A forest plot of the uni-Cox (H) and multi-Cox (I) regression analyses. (J) A nomogram for predicting the survival probability of ESCA patients for predicting 1-, 2-, and 3-year survival outcomes of ESCA. (K) Calibration curves of the OS-nomogram. ***, P<0.001. AUC, area under the ROC curve; ROC, receiver operating characteristic; CI, confidence interval; OS, overall survival; DESRlncRNA, differentially expressed stemness-related long non-coding RNA; LASSO, least absolute shrinkage and selection operator; uni-Cox, univariate Cox; multi-Cox, multivariate Cox; ESCA, esophageal cancer.

https://dx.doi.org/10.21037/tcr-22-1723

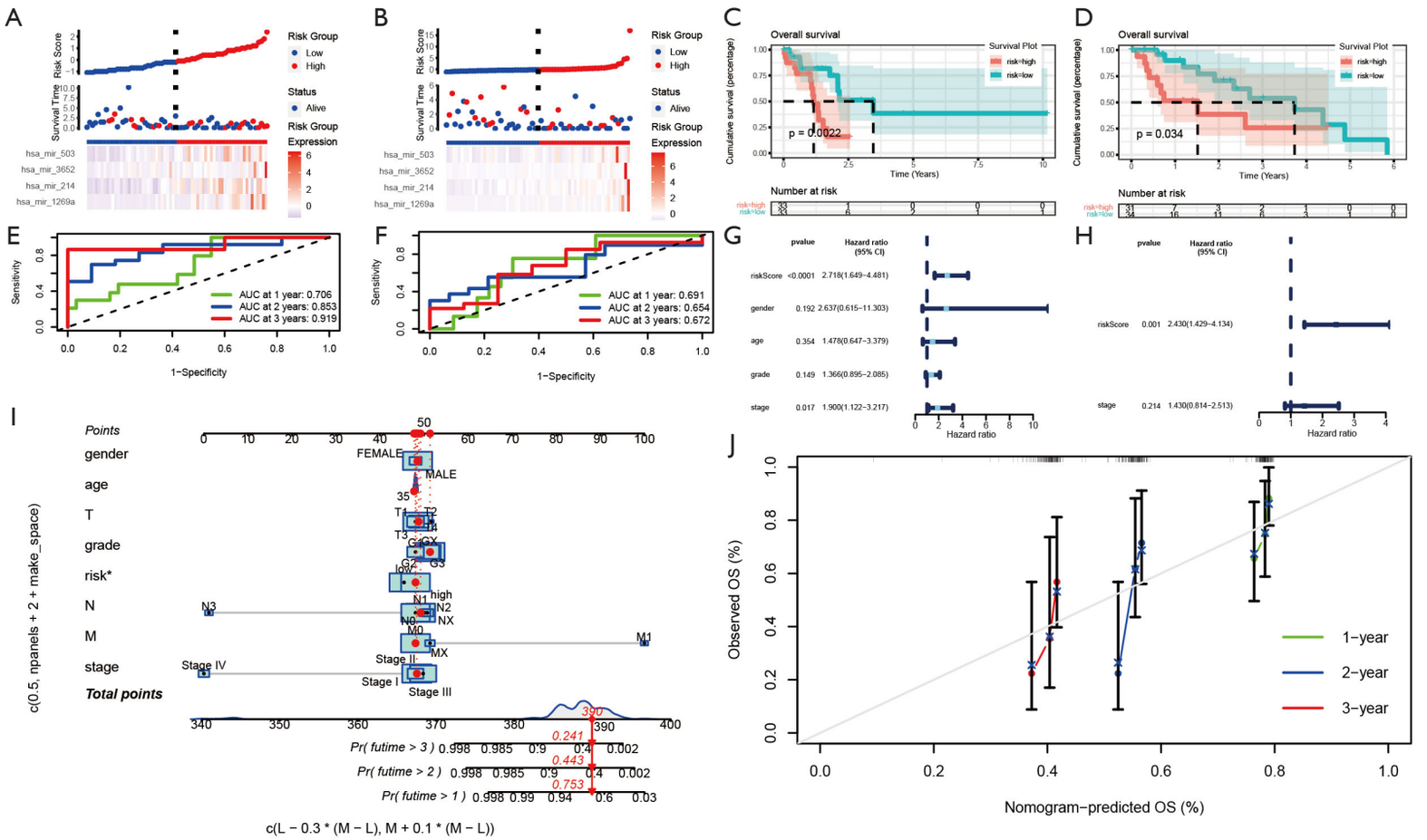

Figure S3 Construction and validation of the DESRmiRNA prognostic model. (A) A risk factor plot for the training set. (B) A risk factor plot for the validation set. The Kaplan-Meier survival curve analysis revealed an OS in the high-risk group from the training set (C) and the validation set (D). The red line represents the high-risk group. The green line represents the low-risk group. Survival-dependent ROC curves validated the prognostic significance of DESRmiRNAs as prognostic indicators in the training set (E) and the validation set (F). The green line represents the 1-year AUC; the blue line represents the 2-year AUC; and the red line represents the 3-year AUC. A forest plot of the uni-Cox (G) and multi-Cox (H) regression analyses. (I) A nomogram for predicting the survival probability of ESCA patients for predicting 1-, 2-, and 3-year survival outcomes of ESCA. (J) Calibration curves of the OS-nomogram. *, P<0.05. AUC, area under the ROC curve; ROC, receiver operating characteristic; CI, confidence interval; OS, overall survival; DESRmiRNA, differentially expressed stemness-related microRNA; LASSO, least absolute shrinkage and selection operator; uni-Cox, univariate Cox; multi-Cox, multivariate Cox; ESCA, esophageal cancer.

https://dx.doi.org/10.21037/tcr-22-1723