Am J Cancer Res 2024;14(4):1831-1849
www.ajcr.us /ISSN:2156-6976/ajcr0152922

Original Article

Construction of gastric cancer prognostic signature based on the E26 transcription factor and the identification of novel oncogene ELK3

Chenxi Liu[1], Liqiang Zhou[2,3], Zhiqing Chen[4]

[1]School of Optometry, Jiangxi Medical College, Nanchang University, Nanchang 330006, Jiangxi, P. R. China; [2]Department of General Surgery, The Second Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang 330006, Jiangxi, P. R. China; [3]Department of General Surgery, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang 330006, Jiangxi, P. R. China; [4]Jiangxi Provincial Key Laboratory of Molecular Medicine, The Second Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang 330006, Jiangxi, P. R. China

Received August 14, 2023; Accepted February 11, 2024; Epub April 15, 2024; Published April 30, 2024

Abstract: The aim of the present study was to investigate the function of 29 E26 (ETS) transcription factor families in gastric cancer (GC) and determine their association with prognosis. Our analysis of the expression of the ETS family revealed that 28 genes were dysregulated in GC, and that their expression was associated with multiple clinicopathological features (P<0.05). Based on the expression signature of the ETS family, consensus clustering was performed to generate two gastric cancer subtypes. These subtypes exhibited differences in overall survival (OS, P = 0.161), disease-free survival (DFS, P<0.05) and GC grade (P<0.01). Functional enrichment analysis of the target genes associated with the ETS family indicated that these genes primarily contribute to functions that facilitate tumor progression. A systematic statistical analysis was used to construct a prognostic model related to OS and DFS in association with the ETS family. This model demonstrated that the maximum area under the curve (AUC) values for predicting OS and DFS were 0.729 and 0.670, respectively, establishing ETS as an independent prognostic factor for GC Furthermore, a nomogram was created from the prognostic signature, and its predictive accuracy was confirmed by a calibration curve. Finally, the expression and prognostic significance of the six genes comprising the model were also examined. Among these, ELK3 was found to be significantly overexpressed in GC clinical samples. Subsequent *in vitro* and *in vivo* studies verified that ELK3 regulates GC proliferation and metastasis, highlighting its potential as a therapeutic target for gastric cancer.

Keywords: Gastric cancer, ETS transcription factors, overall survival, disease-free survival, prognostic model, ELK3

Introduction

Gastric cancer (GC) is a common malignant tumor within the digestive tract which originates from the epithelium of the gastric mucosa. Recent data indicate that annually, the global incidence of GC exceeds one million new cases. Despite a global decline in incidence and mortality rates over the past five years, GC continues to be the third leading cause of cancer-related deaths [1]. However, GC ranks second and third in morbidity and mortality in China, respectively [2]. While the widespread adoption of gastroscopy in China and other nations has somewhat enhanced the early diagnosis of GC, the disease often begins insidiously and lacks specific symptoms, leading patients to seek medical attention for dyspeptic symptoms like malignancy and vomiting, with a few experiencing melena due to upper gastrointestinal bleeding [1, 3]. Current treatment strategies are challenged by high recurrence rate and drug resistance, resulting in a 5-year survival rate of less than 30% [4]. Moreover, traditional tumor markers, such as carcinoembryonic antigen and CA199 offer limited sensitivity and specificity in the early stages of GC, underscoring the urgent need for novel markers capable of accurately predicting GC and its prognosis to improve patient quality of life.

https://doi.org/10.62347/RVBP7871

The E26 (ETS) transcription factor family is one of the largest families of transcription factors. A characteristic shared by all ETS transcription factors is the highly conserved ETS domain, comprising approximately 85 DNA binding sites that form a wing-like helix-turn-helix motif. This domain specifically recognizes the core consensus DNA sequence 5'-GGA(A/T)-3', known as the ETS binding site [5, 6]. To date, 29 ETS such factors have been discovered, with 28 in humans and 27 in mice. These transcription factors regulate various normal developmental and physiological processes, including cell cycle, differentiation, proliferation, apoptosis, tissue remodeling, and angiogenesis [7]. Furthermore, ETS transcription factors also play a crucial role in tumorigenesis, participating in tumor development through gene fusion or chromosomal rearrangement [8, 9], amplification or overexpression [10, 11], and point mutations [12]. In GC, specific ETS factors are associated with poor prognosis and contribute to tumor proliferation [13, 14].

RNA-seq data from The Cancer Genome Atlas (TCGA) and The Genotype-Tissue Expression (GTEx) databases to comprehensively analyze the expression patterns of 29 ETS transcription factors. Through consensus clustering analysis, two GC patient groups with distinct prognoses were identified. The association between the ETS transcription factors and with patient clinicopathological characteristics was analyzed, along with the investigation of their target genes and potential regulatory mechanisms in GC. A prognostic model comprising ETS transcription factors was established based on univariate Cox regression analysis and least absolute shrinkage and selection operator (LASSO) regression analysis to predict overall survival (OS) and disease-free survival (DFS) demonstrated good prognostic performance. Public databases and collected GC surgical specimens were employed to validate the expression and prognostic potential of the key genes constituting the OS and DFS models. Experimental research on ELK3 confirmed the role of ELK3 in promoting tumor cell proliferation and metastasis, highlighting its potential as a therapeutic target for GC.

Materials and methods

Data source and processing

Clinical information and fragments per kilobase million (FPKM) expression data of 32 normal stomachs and 375 stomach adenocarcinomas (STAD) were downloaded from TCGA (https://portal.gdc.cancer.gov/) and the FPKM data were converted into transcripts per million (TPM) [15, 16]. TPM data for normal tissues were obtained from the GTEx database (https://www.gtexportal.org/), from which the gastric normal tissue expression data for 359 normal gastric tissues were extracted. The 'sva' package in R was utilized to merge and standardize the expression data from TCGA and GTEx, thus mitigating the batch effect between the two databases [17]. The Ensembl ID and gene ID for 29 ETS transcription factor families were obtained from the Human Transcription Factor Database (Human TFDB, http://bioinfo.life.hust.edu.cn/HumanTFDB#!/) [18]. The expression levels of the 29 ETS transcription factors were extracted from the combined GC and normal gastric tissue expression datasets, and the Wilcoxon test of the 'limma' package in R was employed to examine the expression differences of these genes [19]. Additionally, the association of the 29 ETS transcription factors with clinicopathological features of gastric cancer was analyzed using the chi-square test.

Protein-protein interaction (PPI) network construction and correlation analysis

A total of 29 ETS transcription factors were uploaded to String (https://string-db.org/) [20], and the minimum correlation value was set to 0.150. The results were imported into Cytoscape 3.7.2 software for visualization [21], and network analysis tools were utilized to evaluate the connectivity degree of each gene, with genes ranked according to their degree of connectivity. R software version 3.6.1 (https://www.r-project.org/) was used to analyze the expression correlation between 29 tumor-related ETS transcription factors in 375 patients with STAD.

Consensus clustering of ETS transcription factors in GC

In order to examine the association between the expression of ETS transcription factors and clinical phenotypes, the consensus clustering method was employed to classify the TCGA-STAD samples into k (k = 2-9) clusters. This approach involved assessing similarities between consensus indices and gene expression

to determine the optimal k value. Furthermore, a survival analysis focusing on overall survival (OS) and disease-free survival (DFS) was conducted based on the clinical data.

Analysis of the mechanisms of ETS transcription factors regulating GC

The professional version of TRANSFAC database (http://gene-regulation.com/) [22] and Transcriptional Regulatory Relationships Unraveled by Sentence-base Text mining (TRRUST, https://www.grnpedia.org/) [23] were utilized to identify the target genes of ETS transcription factors as reported in the literature. Visualization was performed using Cytoscape 3.7.2 [21]. The 'clusterProfiler' R package (version 4.4.1; https://doi.org/10.1016/j.xinn.2021.100141) was employed to conduct Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) functional enrichment analyses on these transcription factors and their target genes [24], adopting $P<0.05$ and FDR <0.05 as thresholds to explore the potential mechanisms through which ETS transcription factors regulate the onset and progression of GC. The 'GOplot' R package was used for visualization [25].

Construction of ETS transcription factors-related OS and DFS prognostic models of GC

The relationship between the expression level of each ETS transcription factor and overall survival (OS) and disease-free survival (DFS) was investigated using univariate Cox regression analysis, with $P<0.05$ set as the selection criterion for hub genes associated with OS and DFS. LASSO regression analysis was then applied to further examine the hub ETS transcription factors related to OS and DFS, leading to the construction of corresponding prognostic models. Each patient with GC was assigned a risk score according to the formula: Risk score = $\beta1 \times EXP1 + \beta2 \times EXP2 + ... + \beta i \times EXPi$, where β represents the coefficient value and EXP denotes the gene expression level.

Based on these risk scores, patients with STAD were categorized into high- and low-risk groups. The 'survival' R package was utilized to generate Kaplan-Meier curves, and the log-rank test was applied to evaluate the survival differences between these two groups. The area under the receiver operating characteristic curve (ROC) was calculated to determine the predictive accuracy of the prognostic model. Additionally, the clinical information of each STAD patient was analyzed, and both univariate and multivariate Cox regression analyses were conducted to assess whether the prognostic model could independently predict patient outcomes, beyond other clinical variables.

Nomogram construction and evaluation

Based on the constructed risk model, the 'rms' package (Version 6.3-0, https://github.com/harrelfe/rms) was used to draw a nomogram, which allocates specific scores to different levels of gene expression. This allows for the visual determination of individual gene scores and the total score by drawing a vertical line, facilitating the evaluation of patient survival. Consequently, this tool can assist clinicians in making more precise patient assessments. Additionally, a calibration curve was generated to assess the accuracy of the nomogram.

Hub ETS transcription factors and prognosis, and expression verification

The Kaplan-Meier Plotter (http://kmplot.com/), which aggregates data from various GC expression microarrays [26], was utilized to verify the correlation between hub gene expression and prognosis. The Human Protein Atlas (HPA, https://www.proteinatlas.org/), containing expression data for approximately 24,000 proteins [27], was used to analyze the protein expression levels of hub genes.

Analysis of ELK3 expression in GC samples and cell lines using reverse transcription-quantitative PCR

This study highlights the significance of ELK3. After receiving written informed consent from the patients and approval from the Second Affiliated Ethics Committee of Nanchang University (ethical approval numbers: 2020093, 2022051), 30 pairs of GC surgical samples were collected. Clinical information relevant to these samples was also gathered for further analysis. Total RNA was extracted from these 50 mg tissue samples using TRIzol® reagent (Thermo Fisher Scientific, Inc.) and reverse-transcribed to cDNA using the PrimeScript RT

Reagent Kit with gDNA Eraser (RR047A, Takara Bio, Inc.), following the manufacturer's protocol. RT-qPCR analysis was conducted using the TB Green Premix Ex Taq II (RR820A, Takara Bio, Inc.) on the 7900HT Fast Real-Time PCR System (Applied Biosystems; Thermo Fisher Scientific, Inc.), with β-actin (ACTB) serving as the internal control. Primers, synthesized by Sangon Biotech (Shanghai) Co., Ltd., had the following sequences: ACTB forward, 5'-CACCATTGGCAATGAGCGGTTC-3' and reverse, 5'-AGGTCTTTGCGGATGTCCACGT-3'; ELK3 forward, 5'-GAGAGTGCA ATCACGCTGTG-3' and reverse, 5'-GTTCGAGGTCCAGCAGATCAA-3'. The $2^{-\Delta\Delta Cq}$ value for ELK3 was calculated using the formula: $\Delta\Delta Cq = \Delta Cq_{experimental\ group} - \Delta Cq_{control\ group}$, where $\Delta Cq = Cq_{target\ gene} - Cq_{internal\ reference}$ [28].

Cell culture and transfection

mRNA data for ELK3 in GC cell lines were downloaded from the Cancer Cell Line Encyclopedia (CCLE, https://sites.broadinstitute.org/ccle), and ELK3 expression was analyzed in each cell line. Based on the GC cell lines in CCLE, the Hs746t and AGS cell lines, which exhibited the highest ELK3 expression, were acquired from The Cell Bank of Type Culture Collection of the Chinese Academy of Sciences. The cells were cultured in DMEM or RPMI 1640 (Beijing Solarbio Science & Technology Co., Ltd.) supplemented with 10% fetal bovine serum (Gibco; Thermo Fisher Scientific, Inc.) and maintained in a 5% CO_2 atmosphere at 37°C. According to the manufacturer's instructions for Lipofectamine 3000® (Thermo Fisher Scientific, Inc.), shRNA#1 (5'-CCTGCGATACTATTATGAC-3'), shRNA#2 (5'-ATCAGGTTTGTGACCAATA-3'), shRNA#3 (5'-TGGATCAGAAACATGAGCA-3', mass/concentration: 7.5 μg, 960 ng/μL), and shNC (5'-TTCTCCGAACGTGTCACGT-3', mass/concentration: 7.5 μg, 1040 ng/μL) provided by Sangon Biotech (Shanghai) Co., Ltd., were transfected into cells in 6-well plates. Lentiviral packaging was carried out using psPAX2 and pMD2.G, and a GC cell line (Hs746t) with stable knockdown of ELK3 was established.

Cell proliferation assay

To evaluate the impact of ELK3 knockdown on GC cell proliferation, cells were stained using the 5-ethynyl-2'-deoxyuridine (EdU) staining kit (US Everbright Inc.) according to the manufacturer's protocol. Cells were incubated with EdU at 37°C for 4 hours, followed by sequential procedures at room temperature: fixation with 4% paraformaldehyde for 15 minutes, treatment with 100 μl YF®594 azide (Everbright Inc., USA) for 30 minutes, and DNA staining with 100 μl DAPI for 30 minutes. A fluorescence microscope (magnification, ×50; Olympus Corporation) was used to capture images from three randomly selected fields of view for analysis. Additionally, cell proliferation was assessed by measuring the optical density (OD) at 450 nm at 0, 24, 48, and 72 hours using a multi-functional reader, as per the CCK-8 kit (Hanbio Inc.) instructions and growth curves were plotted to assess cell proliferation capability.

Cell invasion and migration assay

To investigate the effects of ELK3 on tumor invasion and migration, a wound healing assay was conducted to assess cell migration. When cell confluency reached approximately 95%, a 200 μl pipette tip was used to create a vertical scratch in the cell monolayer. Subsequently, cell debris was removed with phosphate-buffered saline (PBS), and an inverted microscope (magnification, ×50; Olympus Corporation) was used to capture images of the wound gap at 0, 24, and 48 hours post-injury to evaluate migratory capability. For assessing vertical migration, cells were resuspended in serum-free medium, and 2×10^4 cells were plated in the upper chamber of an 8.0 μm Transwell® plate (Corning, Inc.). In the invasion assay, the upper chamber was pre-coated with Matrigel (BD Biosciences) for 2 hours, with subsequent steps mirroring those of the migration assay. The lower chamber was filled with 800 μl of serum-enriched medium and incubated for 72 hours. Cells that invaded the lower chamber were fixed with methanol for 15 minutes at room temperature and examined using an inverted microscope (magnification, ×50; Olympus Corporation).

Western blot analysis

Proteins were extracted from transfected cells using RIPA lysis buffer (Beijing Solarbio Science & Technology Co., Ltd.). Protein concentration was determined using a BCA kit (Beyotime Institute of Biotechnology). Proteins were then separated by 10% SDS-PAGE gel electrophoresis, followed by transfer onto PVDF membranes (MilliporeSigma) The membranes were blocked

with 5% non-fat dry milk for 1 hour at room temperature. Subsequently, membranes were incubated with primary antibodies against ELK3 (1:2500 dilution, cat. no. NBP1-83960; Novus Biologicals), proliferating cell nuclear antigen (PCNA; 1:1000 dilution, cat. no. A0264; ABclonal Biotech Co., Ltd.), cyclin D1 (1:1000 dilution, cat. no. A19038; ABclonal Biotech Co., Ltd.), N-cadherin (1:1000 dilution, cat. no. ab76057; Abcam), Vimentin (1:5000 dilution, cat. no. ab137321; Abcam), and GAPDH (1:10000 dilution, cat. no. 20536-1-AP; ProteinTech Group, Inc.) at 4°C for 12 hours. After aspirating primary antibody solutions and membrane washing, HRP-labeled goat anti-rabbit IgG (cat no. BA1070) and goat anti-mouse IgG (cat no. BM2002I) secondary antibodies (CWBio) were applied at a 1:10000 dilution and incubated at room temperature for 1 hour. Blots were washed and chemiluminescent detection was performed using an ultra-sensitive ECL kit (Everbright Inc., USA). Images were analyzed with Image Lab software 5.2.1 (Bio-Rad Laboratories, Inc.).

Nude mouse xenograft model and Immunohistochemistry analysis

Six female BALB/c nude mice (5 weeks old) were acquired from Hangzhou Ziyuan Laboratory Animal Technology Co., Ltd., China, and maintained in specific pathogen-free, normal conditions. Each nude mouse was injected with 5×10^6 stably transformed Hs746t GC cells. The length and width of the tumors were measured every 5 days following tumor formation, and the tumor volume = $(\text{width})^2 \times \text{length}/2$ was calculated. Mice were euthanized using carbon dioxide when tumor length reached 15 mm. Tumor tissues were then paraffin-embedded and sectioned for immunohistochemistry to assess the expression of Ki67 (dilution: 1:500, cat. no. A23722; ABclonal Biotech Co., Ltd.) and E-cadherin (dilution: 1:500, cat. no. ab40772; Abcam), aiming to determine the impact of ELK3 on gastric cancer proliferation and metastasis. The experiment was conducted at Nanchang Royo Biotech Co., Ltd. and received approval from The Institutional Animal Care and Use Committee of Nanchang Royo Biotech Co., Ltd. (approval number: RYE2022-011001).

Statistical analysis

R software version 3.6.1 was utilized for bioinformatics analysis. The analysis included the use of an unpaired t-test for assessing differences in ETS transcription factor expression, Pearson correlation for analyzing correlations among ETS transcription factors, the Log Rank test for evaluating prognostic differences, and the Chi-square test for investigating the association between ETS transcription factors and clinicopathological features. GraphPad Prism version 8.0 (GraphPad Software, Inc.) was employed for statistical analysis of both in vivo and in vitro experiments, with data presented as mean ± standard deviation. Experiments were conducted with greater than three replicates. Student's unpaired t-test was applied for comparisons between two groups. A *P*-value of <0.05 was used for determination of statistical significance.

Results

Expression of ETS transcription factors in GC

The present study first analyzed the differences in expression of ETS transcription factors in by integrating data from the TCGA and GTEx databases, and visualized the findings using heatmaps and violin plots. The analysis showed that (**Figure 1A** and **1B**), among the 29 ETS transcription factors, EHF, ELF1, ELF3, ELF4, ELK1, ELK3, ELK4, ERF, ERFL, ETS1, ETS2, ETV1, ETV3, ETV3L, ETV4, ETV5, ETV6, ETV7, GABPA, SPI1, SPIB, and SPIC were significantly overexpressed in GC ($P<0.001$); conversely, ELF2, ELF5, ERG, FEV, FLI1, and SPDEF were highly expressed in normal gastric tissues ($P<0.05$). Among these, ETV2 exhibited low expression in GC, though this difference did not reach statistical significance ($P=0.179$). Network analysis of the constructed PPI network revealed that the degree values of EHF, ELF3, ELF4, ETV6, ETS1, and ETS2 were all greater than 20, signifying a robust interaction capability with other ETS transcription factors (**Figure 1C**). Additionally, the study assessed the correlations among ETS transcription factors, finding varying degrees of association, predominantly positive. Specifically, ETV6, ELF1, EHF, ETS1, ELK4, ELK3, SPI1, and ELK1 showed positive correlations with other genes, whereas ERG was nega-

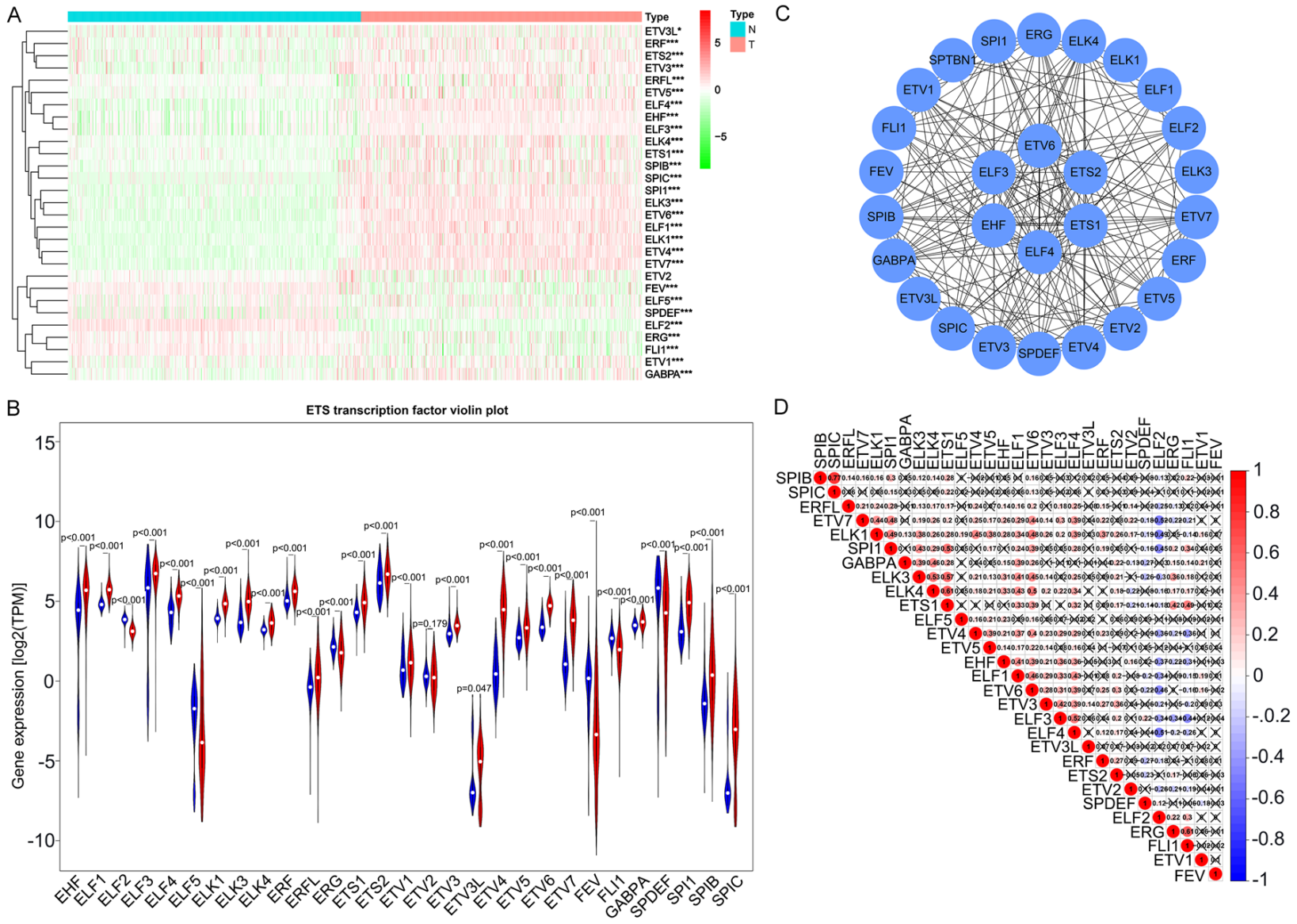


Figure 1. Expression pattern of ETS factors in gastric cancer. A. Heatmap illustrating the expression levels of ETS factors in gastric cancer and normal tissues. B. Violin plot illustrating the expression differences of ETS factors in gastric cancer. C. Protein-protein interaction network of 29 ETS factors. D. Pearson's correlation analysis of ETS factors in the TCGA-STAD cohort. *P<0.05, **P<0.01, ***P<0.001. TCGA, The Cancer Genome Atlas; STAD, stomach adenocarcinoma.

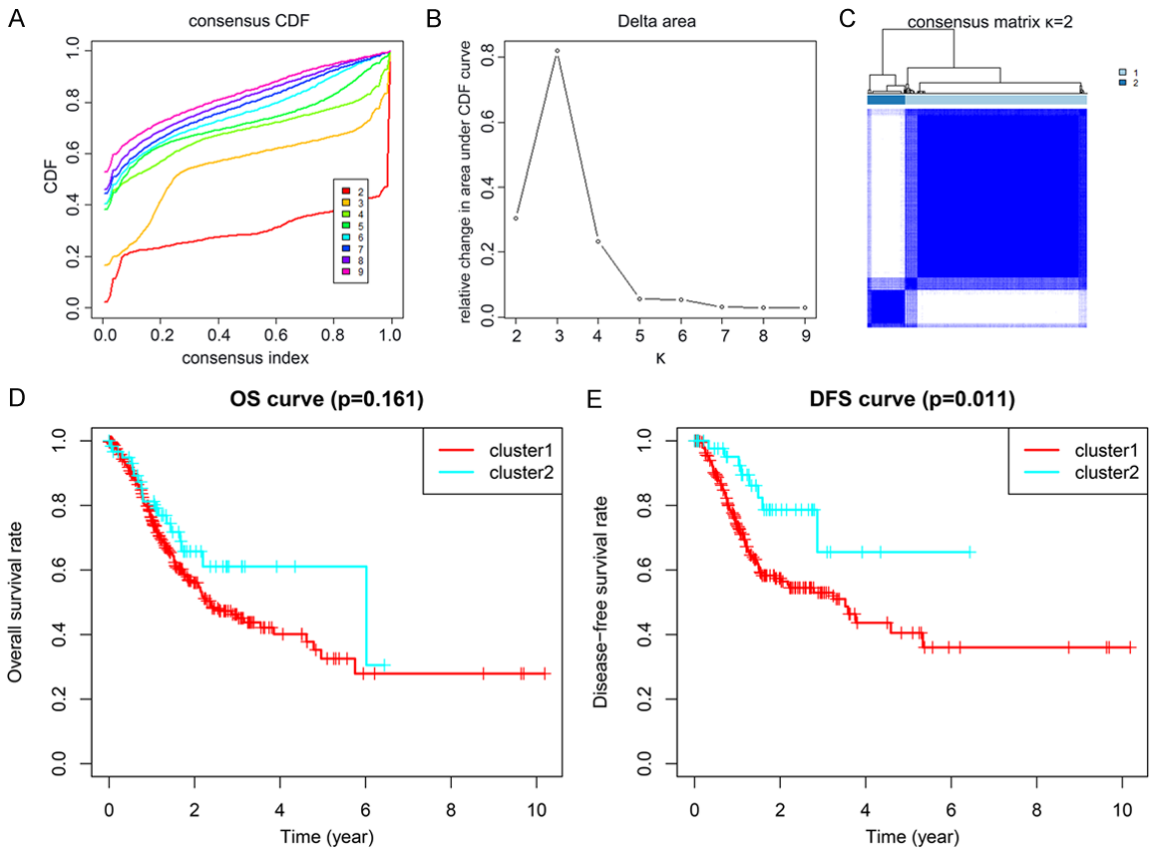


Figure 2. Differential expression patterns of TCGA-STAD patients in the two clusters. A. *k* = 2-9 Consensus cluster CDF. B. *k* = 2-9, the relative change of the area under the CDF curve. C. When *k* = 2, the TCGA-STAD is divided into two clusters. D. Kaplan-Meier analysis of the overall survival rate of the two clusters. E. Kaplan-Meier analysis of the disease-free survival rate of the two clusters. TCGA, The Cancer Genome Atlas; STAD, stomach adenocarcinoma; CDF, cumulative distribution function.

tively correlated with most ETS transcription factors (**Figure 1D**). These findings highlight ETV6, ETS1, and EHF as key genes within the ETS transcription factor PPI network.

Clustering of ETS transcription factors in GC

The study utilized 375 clinical samples from the TCGA-STAD dataset for consensus clustering. The stability of clustering varied from k = 2 to 9, with analyses of gene expression similarity and cumulative distribution function (CDF) suggesting k = 2 as the optimal cluster solution. Cluster 1 comprised 310 samples, whereas cluster 2 included 65 samples (**Figure 2A-C**). Kaplan-Meier survival analysis was performed to compare OS and DFS between these two subgroups. The analysis indicated that cluster 1 exhibited poorer OS and DFS compared to cluster 2 (P = 0.011); however, it is important to note that the Kaplan-Meier analysis for OS yielded a *P*-value of 0.161 (**Figure 2D** and **2E**). Furthermore, the study examined the relationship between cluster categorization and various clinical parameters including microsatellite instability (MSI), AJCC stage, T stage, N stage, M stage, grade, sex, age, and survival time, using the Chi-squared test. The results revealed (Supplementary Figure 1A) that the majority of GC samples in cluster 1 were classified as grade 3, while samples in cluster 2 were predominantly grades 2 and 3 (P<0.01). The association between ETS transcription factors and these clinical characteristics was further analyzed and visualized through heatmaps. This analysis revealed (Supplementary Figure 1B) that the expression of ETS transcription factors was most strongly associated with grade, T stage, AJCC stage, and MSI, and to a lesser extent with M stage and N stage. Specifically, ETS1, FLI1, ETV2, ETV3, and ETV6 were all sig-

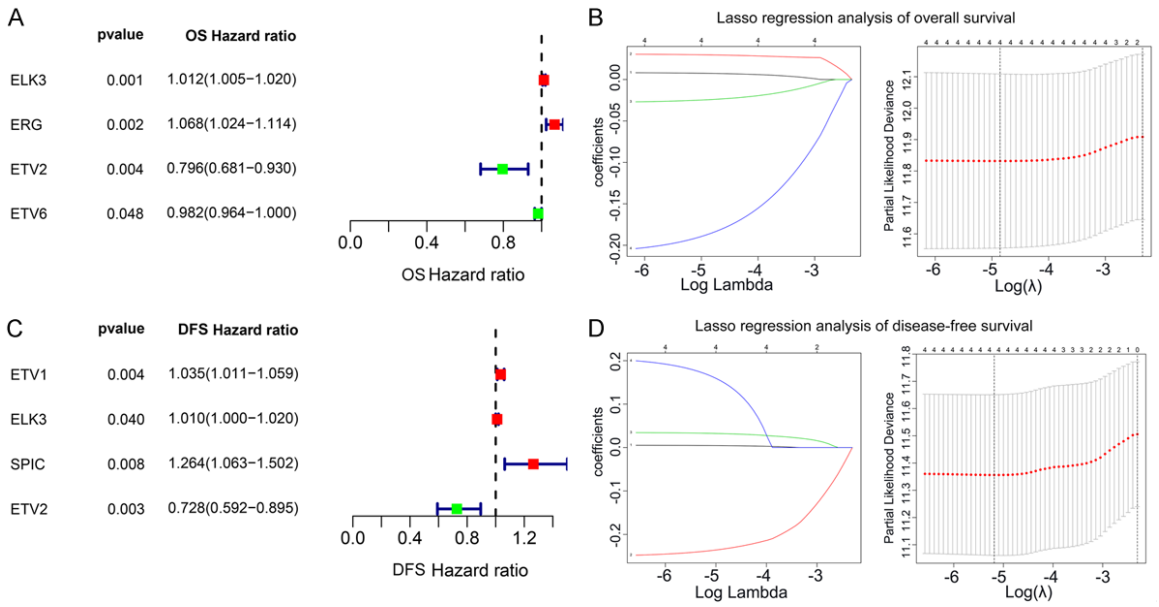


Figure 3. Screening of ETS factors that can predict the prognosis of gastric cancer. A. Univariate Cox regression analysis identified four ETS factors associated with OS. B. LASSO regression analysis further identified four hub ETS factors of OS. C. Univariate Cox regression analysis identified four ETS factors associated with DFS. D. LASSO regression analysis further identified four hub ETS factors of DFS. OS, overall survival; DFS, disease-free survival.

nificantly associated with the clinical characteristics of the four cases (P<0.05).

Regulatory mechanisms of ETS transcription factors in GC

In the TRANSFAC database and TRRUST, 26 ETS transcription factor target genes were identified (Supplementary Figure 2A). Notably, ETS1, ETS2, SPI1, and ELK1 had a greater number of target genes, whereas no target genes were identified for ERFL, ETV2, and ETV3L. GO and KEGG enrichment analyses were conducted to elucidate the potential regulatory mechanisms of the ETS transcription factors in GC. The GO enrichment results revealed (Supplementary Figure 2B) that the genes were enriched in biological processes, such as the regulation of hemopoiesis, leukocyte migration, and response to molecules of bacterial origin; cellular components like the external side of the plasma membrane, secretory granule lumen, and membrane raft; and molecular functions including cytokine receptor binding, growth factor receptor binding, and growth factor binding. The KEGG enrichment analysis results revealed (Supplementary Figure 2C) that these genes were mainly enriched in PI3K/Akt signaling, microRNAs in cancer, JAK-STAT signaling, Epstein-Barr virus infection, NF-κB signaling, IL-17 signaling, TNF signaling, HIF-1 signaling, MAPK signaling, PD-1/PD-L1 immune checkpoint and other key pathways, suggesting that ETS transcription factors play a critical role in these signaling pathways in tumors.

Construction of OS and DFS prognostic signatures based on ETS transcription factors

Univariate Cox regression analysis was utilized to assess the prognostic value of the 29 ETS transcription factors for OS and DFS of patients with GC. The analysis revealed that four factors were significantly associated with OS and DFS (P<0.05). Among the OS-related genes, ELK3 (HR = 1.012, 95% CI = 1.005-1.020) and ERG (HR = 1.068, 95% CI = 1.024-1.114) were considered to be predictive OS risk genes, and ETV2 (HR = 0.796, 95% CI = 0.681-0.930) and ETV6 (HR = 0.982, 95% CI = 0.964-1.000) were considered to be protective genes for predicting OS (**Figure 3A**). For DFS, ETV1 (HR = 1.035, 95% CI = 1.011-1.059), ELK3 (HR = 1.010, 95% CI = 1.000-1.020), and SPIC (HR = 1.264, 95% CI = 1.063-1.502) were identified as risk genes, and ETV2 (HR = 0.782, 95% CI = 0.592-0.895) as a protective gene (**Figure 3C**). LASSO regression analysis was subsequently applied to fur-

ther refine the ETS transcription factors related to OS and DFS, resulting in the construction of risk proportion prognostic models. The results of LASSO regression analysis suggested that the prognostic signatures for OS included ELK3, ERG, ETV2, and ETV6 (**Figure 3B**), and the risk score calculation formula related to the OS of each patient was as follows: Risk score = (0.007 × EXP_{ELK3}) + (0.030 × EXP_{ERG}) + (-0.024 × EXP_{ETV6}) + (-0.186 × EXP_{ETV2}). The prognostic signatures for constructing DFS included ETV1, ELK3, SPIC, and ETV2 (**Figure 3D**). The risk score calculation formula for DFS of each patient was as follows: Risk score = (0.033 × EXP_{ETV1}) + (0.005 × EXP_{ELK3}) + (0.169 × EXP_{SPIC}) + (-0.238 × EXP_{ETV2}).

Validation of OS and DFS prognostic models

Based on the median value of the OS-related risk score and DFS-related risk score, all patients in the OS prognostic model and DFS prognostic model were divided into the high- and low-risk groups. First, a scatter plot of the survival status against risk score was drawn. The plot demonstrates that with an increase in risk score, the mortality rate of GC patients gradually increased (Supplementary Figure 3A and 3B). Kaplan-Meier analyses indicated that patients with high risk scores had significantly shorter OS (P<0.0001, Supplementary Figure 3C and 3E) and DFS (P<0.05, Supplementary Figure 3H) compared to those with low risk scores. Furthermore, the predictive accuracy of the OS and DFS prognostic models was assessed using the area under the ROC curve (AUC). For the OS prognostic model (Supplementary Figure 3D), AUC values were 0.660 at 1 year, 0.612 at 3 years, 0.646 at 5 years, and 0.729 at 7 years, suggesting moderate prognostic capability. The DFS prognostic model (Supplementary Figure 3F) exhibited AUC values of 0.590 at 1 year, 0.664 at 3 years, 0.670 at 5 years, and 0.639 at 7 years, indicating median predictive performance. Additionally, the integration of clinicopathological features from the TCGA-STAD dataset was performed to evaluate whether the risk score functioned as an independent prognostic factor. Univariate Cox regression analysis (Supplementary Figure 3G) identified N stage, T stage, AJCC stage, and OS risk score as risk factors for GC patient survival. Multivariate Cox regression analysis (Supplementary Figure 3H) further determined that age (HR = 1.039, 95% CI = 1.019-1.059, P<0.001) and the OS risk score (HR = 2.899, 95% CI = 1.817-4.628, P<0.001) were independent risk factors for OS. For DFS, univariate analysis (Supplementary Figure 3J) identified sex, N stage, AJCC stage, and DFS risk score as risk factors, while multivariate analysis (Supplementary Figure 3K) showed sex (HR = 1.873, 95% CI = 1.140-3.079, P = 0.013) and DFS risk score (HR = 1.458, 95% CI = 1.230-1.729, P<0.001) as independent prognostic factors for DFS. The 5-year AUC values were calculated to compare the prognostic abilities of risk scores and various clinicopathological characteristics. The results demonstrated that in the OS prognostic model, the risk score (0.654) surpassed that of age (0.592) (Supplementary Figure 3I), while in the DFS model, risk score AUC was marginally lower than that of sex (Supplementary Figure 3L).

Nomogram construction and verification

To accurately predict the prognosis of GC, nomograms were developed by integrating the four signatures from both the OS and DFS models, enabling the prediction of OS and DFS probabilities from 1 to 5 years (**Figure 4A** and **4C**). The calibration chart (**Figure 4B** and **4D**) demonstrated that the nomogram closely predicted the actual survival outcomes with a slope approximating 1 for the 5-year OS and DFS, indicating their excellent predictive accuracy. These nomograms can thus provide valuable insights for patients with GC and assist clinicians in making more precise clinical decisions.

Verification of the prognostic value and expression of Hub ETS transcription factors

To further investigate the prognostic significance of key ETS transcription factors in developing OS and DFS prognostic models, Kaplan-Meier plotter was used to analyze the association between these hub genes and OS and DFS. The Kaplan-Meier plotter determined that the expression levels of ELK3 and ERG were associated with an overall poor prognosis of patients with GC, while the expression levels of ETV2 and ETV6 were associated with an improved OS (Supplementary Figure 4A). For

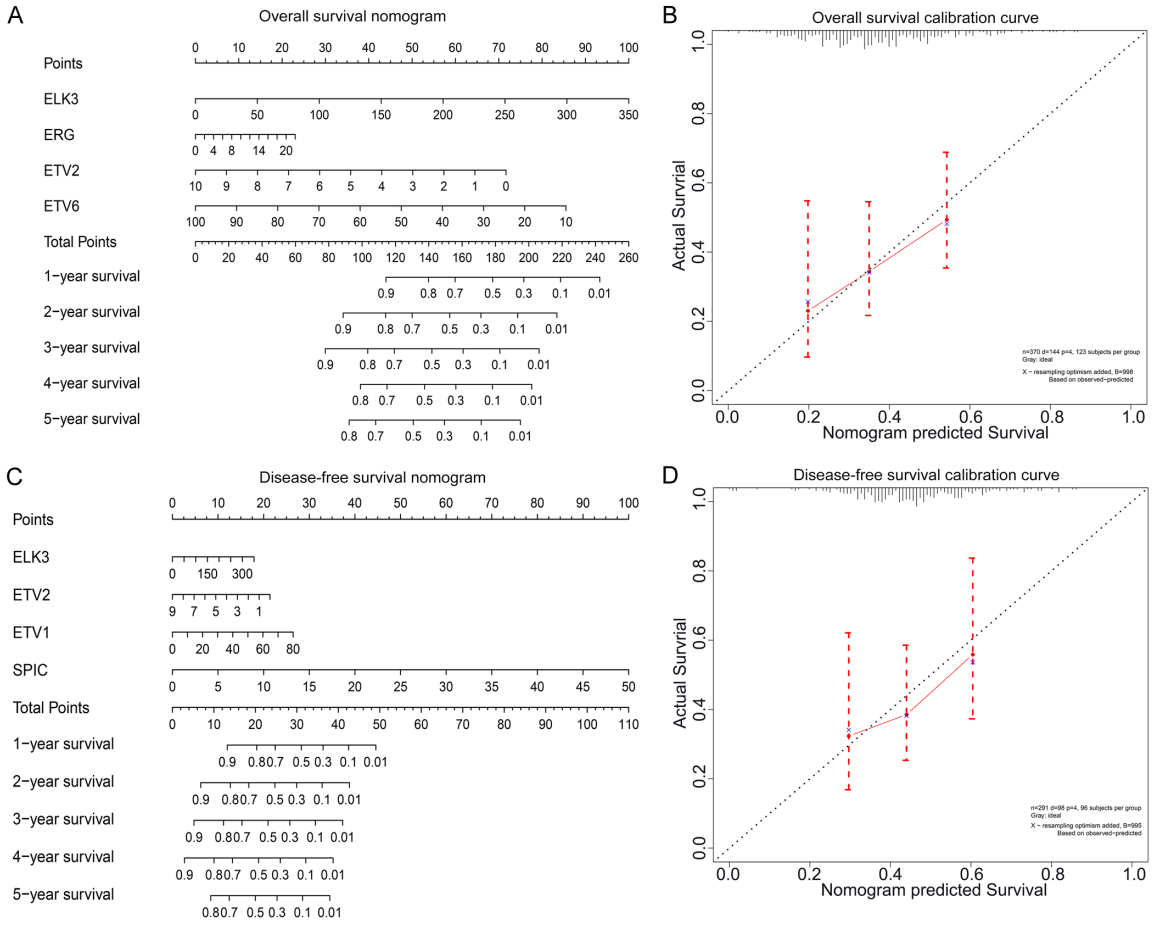


Figure 4. Construction and verification of OS/DFS related nomograms. A. Multivariate Cox regression was used to construct OS-related nomograms and predict the 1-5-year survival rates. B. Calibration detected the difference between the 5-year predicted survival rate and the actual survival rate of the OS nomogram. C. Multivariate Cox regression was used to construct DFS-related nomograms and predict the 1-5-year survival rates. D. Calibration detected the difference between the 5-year predicted survival rate and the actual survival rate of the DFS nomogram. OS, overall survival; DFS, disease-free survival.

patients with GC, the expression levels of ELK3, ETV1, and SPIC were associated with poor DFS; however, the log-rank test of survival in association with SPIC did not yield significant results (P>0.05), and the expression of ETV2 was associated with an improved DFS (Supplementary Figure 4A). The HPA database was also used to verify the protein expression of the six hub genes that constitute the prognostic model of OS and DFS. In this database, only three genes were found. The results indicated that the protein expression levels of these three, ELK3, ERG and ETV6, were consistent with their mRNA levels (Supplementary Figure 4B). Among these ELK3 was implicated in influencing both OS and DFS in GC patients. ELK3 expression was further examined in GC clinical samples (**Figure 5A**) revealing elevated ELK3 levels in 30 pairs of gastric cancer samples, with 18 samples showing high ELK3 expression and 12 exhibiting low expression. Analysis of the clinical and pathological characteristics of ELK3 indicated a significant association with T stage (**Table 1**).

ELK3 acts as an oncogene in GC

In vitro and *in vivo* experiments were performed to assist in elucidating the potential role of ELK3 in GC. The CCLE database identified that ELK3 was most highly expressed in Hs746t and AGS cell lines (**Figure 5B**). Three shRNA plasmids were utilized to suppress ELK3 expression in Hs746t cells, with Western blot results showing that all three shRNAs effectively inhibited ELK3 protein expression, while RT-PCR

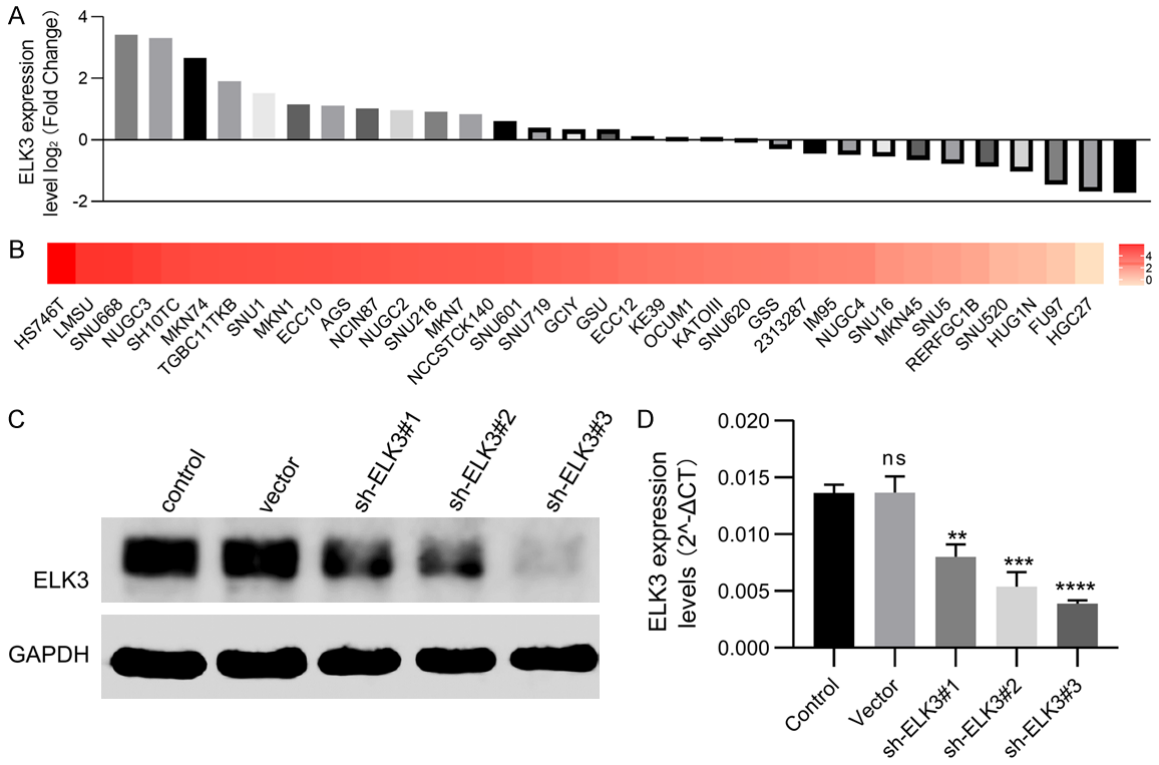


Figure 5. A. The high expression of ELK3 was confirmed in 30 pairs of samples. B. The Cancer Cell Line Encyclopedia database demonstrated that the Hs746t cell line had the highest expression of ELK3. C. Western-blot verifies the efficiency of ELK3 inhibition. D. rt-PCR verified the inhibition efficiency of ELK3. **P<0.01, ***P<0.001, ****P<0.0001.

confirmed significant suppression of ELK3 mRNA levels (**Figure 5C** and **5D**, P<0.01). Subsequent experiments were conducted using the shELK3#2 and shELK3#3 plasmids. The present study first analyzed the effects of ELK3 on proliferation. EdU staining indicated reduced proliferation in cells with ELK3 knockdown (P<0.05, **Figure 6A** and **6B**), accompanied by decreased levels of proliferation-associated proteins, PCNA and cyclin D1 (**Figure 6C**). Furthermore, ELK3 knockdown resulted in diminished horizontal migration (P<0.05, **Figure 7A**), reduced vertical migration and invasion (P<0.01, **Figure 7C**), and reduced expression of metastasis-related proteins, N-cadherin, and vimentin (**Figure 7B**). These findings suggest the involvement of ELK3 in GC cell metastasis. Notably, *in vitro* experiments demonstrated significantly reduced tumorigenic potential in Hs746t cells following ELK3 knockdown compared to the control group (P<0.01, **Figure 8A**). Immunohistochemical analysis of tumor tissues indicated that reduced ELK3 expression was associated with decreased Ki67 and increased E-cadherin expression (**Figure 8B**). *In vivo* experiments showed that ELK3 can promote gastric cancer proliferation and metastasis, collectively indicating the ability of ELK3 to function as an oncogene in GC.

Discussion

The ETS transcription factor family has been discovered for more than 30 years. Over the past 10 years, the understanding of the function of this family in solid tumors has increased exponentially Studies have highlighted that the abnormal expression of ETS transcription factors plays a critical role in the occurrence and development of tumors [7]. This study leverages data from the GTEx and TCGA to perform a comprehensively analysis of the expression of 29 ETS transcription factors in GC, finding that 28 of these are abnormally expressed. The majority of these factors are upregulated and are closely associated with tumor progression. For example, the study by Zhang *et al.* [29] demonstrated that PDEF was highly expressed in GC and was closely related to GC cell prolifera-

Table 1. Relationship between ELK3 and clinicopathological features of gastric cancer

Clinicopathologic	High expression	Low expression	*P* value
Gender			0.88
Male	11	7	
Female	7	5	
Age (years)			0.87
≤60	13	9	
>60	5	3	
CEA (ng/ml)			0.77
≤5	8	6	
>5	10	6	
CA199 (U/ml)			0.18
≤37	12	5	
>37	6	7	
Histologic grade			0.15
Poor	4	8	
Moderate-Well	14	4	
T stage			0.02
Tris-T2	7	11	
T3-T4	11	3	
N stage			0.37
N0	9	8	
N1-3	9	4	

tion, migration, and invasion, potentially through a positive feedback loop with FOXM1 that stimulates cell proliferation [14]. Notably, this study found that ETS2 is highly expressed in GC and positively correlates with the grade of GC. This conclusion differs from the findings reported in the study by Liao *et al.* [30]. They confirmed that ETS2 was expressed at low levels in GC and functioned as a tumor suppressor gene [30]. However, Das *et al.* demonstrated that ETS2 is significantly upregulated in *Helicobacter pylori*-infected GC, promoting the expression of Siah1 and enhancing the degradation of membrane-bound β-catenin, thereby significantly promoted the invasiveness of GC [31, 32]. These findings suggest the unique role of ETS2 in GC. In addition, based on the expression patterns of ETS transcription factors and using consensus clustering, the present study identified two unique subgroups of GC. Systematic analysis revealed significant differences between these two subgroups OS, DFS and grade. This observation further suggests that the expression patterns of ETS transcription factors are closely associated with the progression and prognosis of gastric cancer (GC). Several of these factors, including ETS1 [33-36], EHF [13], ETV1 [37], and ETV4 [38, 39], have been extensively reported in GC. These factors are highly expressed in GC and promote the progression of GC, thereby serving as effective biomarkers.

Previous research has established that ETS transcription factors are involved in various critical processes related to tumor development, including tumor cell self-renewal and survival, DNA repair and genome stabilization, the regulation of chromatin dynamics and epigenetics, the regulation of metabolism, and tumor microenvironment [40]. The present study analyzed the role of the ETS transcription factor family in GC by identifying and conducting functional enrichment analyses on the target genes of each factor, in conjunction with examining their association with the clinicopathological characteristics of GC. Analysis revealed a positive association between ETS transcription factors and clinicopathological parameters of grade, M stage, N stage, T stage, and AJCC stage, suggesting their potential role in promoting GC progression. KEGG analysis of all target genes suggested that the ETS family might influence GC cell development and progression through various pathways, including PI3K-Akt signaling, JAK-STAT signaling, NF-κB signaling, microRNAs in cancer, HIF-1 signaling, MAPK signaling, platinum resistance, TGF-β signaling, and other pathways regulating the occurrence and development of GC cells. Although some of these conclusions have not yet been confirmed in GC, they have been reported in other tumors. For instance, Xu *et al.* [41] demonstrated that ETV4 promotes clear cell carcinoma metastasis via a PI3K-Akt-dependent activation of FOSL1. ELK3 was found to breast cancer sensitivity to adriamycin through the same pathway [42]. In addition, ELK3 has been implicated in enhancing liver cancer migration and invasion by regulating HIF-1α [43]. ETS1 has been shown to promote cisplatin resistance in triple-negative breast cancer by activating the NF-κB pathway through direct binding to the IKKα promoter [44]. NF-κB can also form a positive feedback loop with ELF3, which constitutively activates NF-κB and drives prostate cancer progression [45].

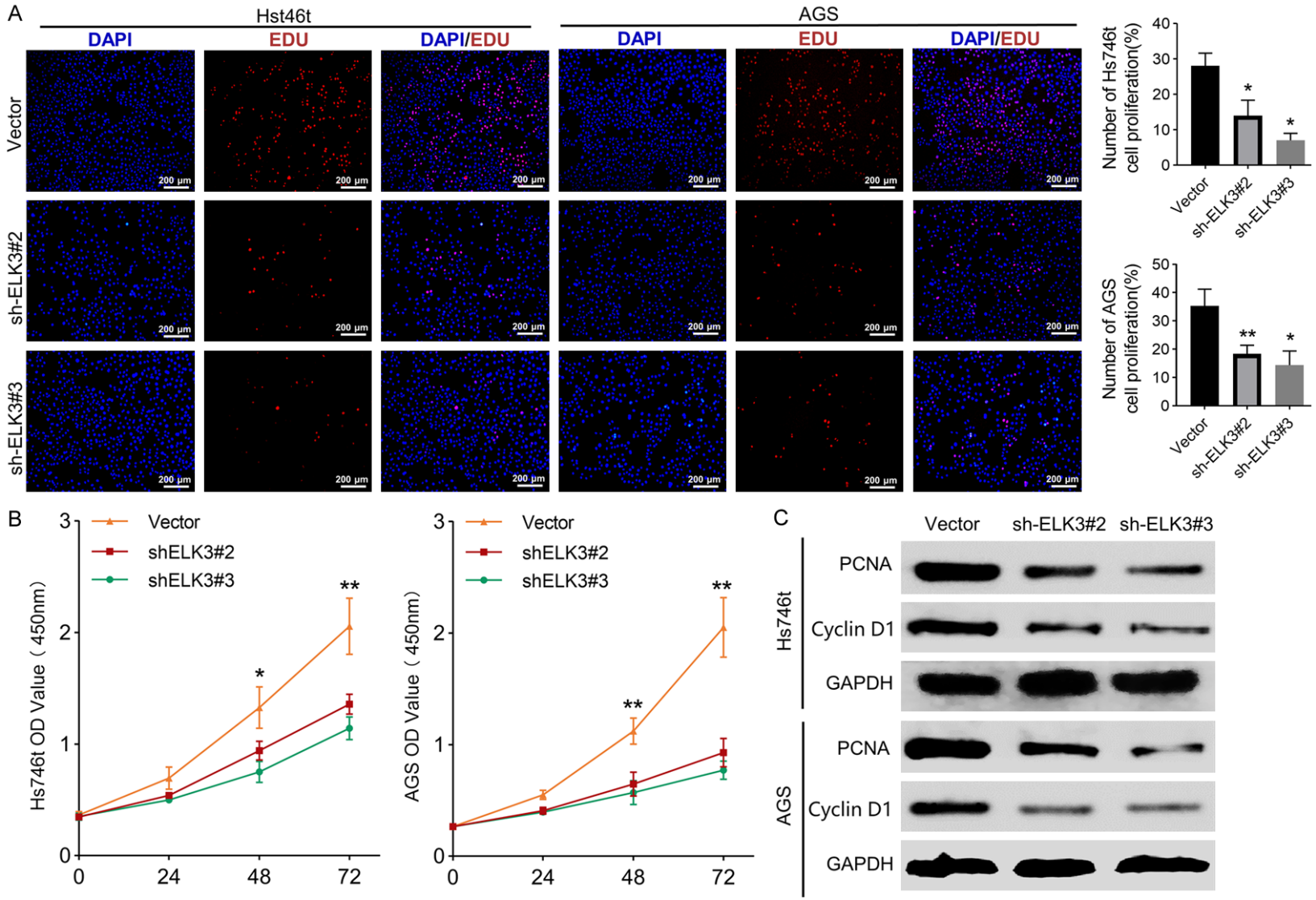


Figure 6. The cell proliferation ability decreased after ELK3 inhibition. A. EdU staining revealed that the proliferation of gastric cancer cells was inhibited by ELK3 knockdown. B. CCK-8 detection of cell proliferation ability showed that inhibition of ELK3 inhibited the growth of gastric cancer cells. C. Western-blot detecting revealed that downregulation of ELK3 resulted in a decrease in PCNA and Cyclin D1 expression. *P<0.05, **P<0.01.

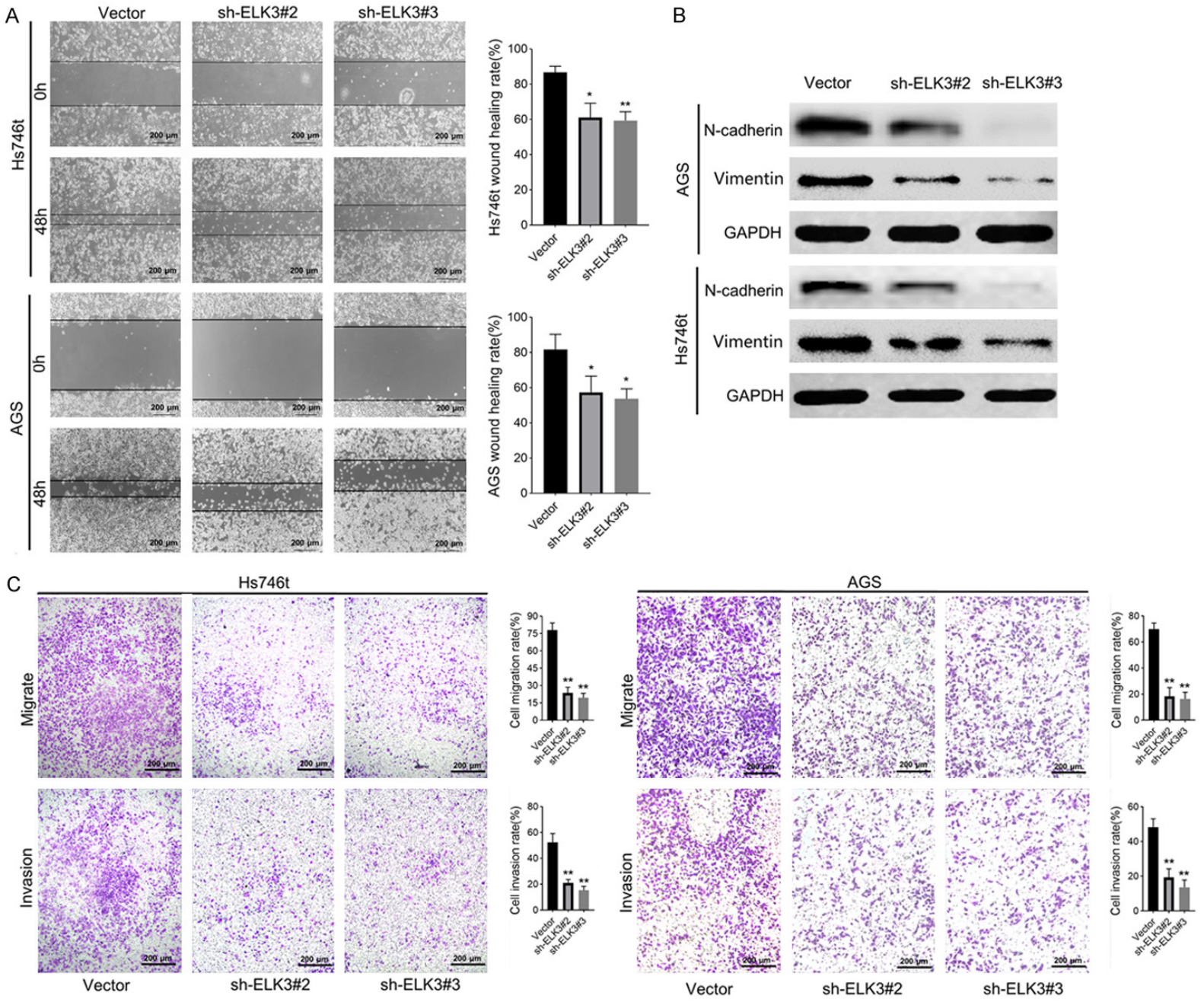

A
Vector
sh-ELK3#2
sh-ELK3#3
Hs746t
AGS
0h
48h
200 μm
Hs746t wound healing rate(%)
AGS wound healing rate(%)
Vector
sh-ELK3#2
sh-ELK3#3
B
Vector
sh-ELK3#2
sh-ELK3#3
AGS
Hs746t
N-cadherin
Vimentin
GAPDH
C
Hs746t
AGS
Migrate
Invasion
Cell migration rate(%)
Cell invasion rate(%)
Vector
sh-ELK3#2
sh-ELK3#3

Figure 7. A. Knockdown of ELK3 reduced the horizontal migratory ability of gastric cancer cell. B. Knockdown of ELK3 led to a decreased expression of N-cadherin and vimentin in gastric cancer cells. C. Knockdown of ELK3 reduced the ability of gastric cancer cells to invade and migrate vertically. *P<0.05, **P<0.01.

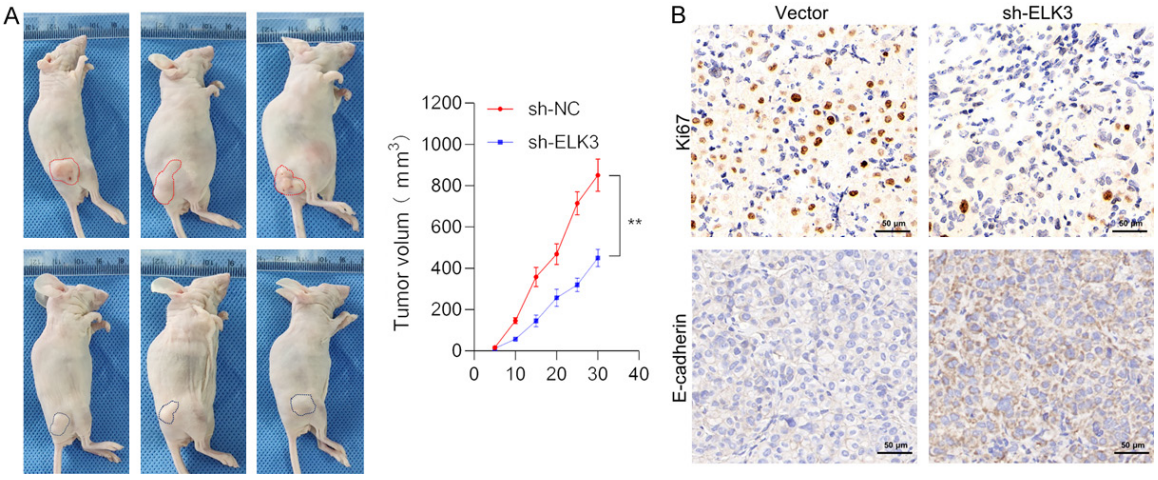


Figure 8. A. *In vivo* experiments revealed that the tumorigenic ability of Hs746t cells in which ELK3 was stably knocked down decreased. B. Immunohistochemistry showed that knocking down ELK3 resulted in downregulation of Ki67 and upregulation of E-cadherin expression in the tumorigenesis of nude mice. **P<0.01.

Furthermore, the ERG fusion gene can activate NF-κB via Toll-like receptor 4 [46]. It has also been demonstrated that this family is involved in other inflammatory pathways. EHF can directly inhibit IL-6, thereby preventing the activation of STAT3 and inhibiting the growth of prostate cancer stem cells [47]. ELF3 can also participate in other pathways. In a previous study on thyroid cancer, ELF3 was found to not only act as a prognostic marker, but also to regulate the transcription of the human epidermal growth factor receptor 2 family and regulate the activity of the MAPK/Erk pathway, forming a positive feedback loop. It was also found to play a role in promoting cancer [48]. The TGF-β signaling pathway, crucial for cell growth, differentiation, and apoptosis, is differentially regulated by the ETS transcription factor family. Yao *et al.* [49] confirmed that ELF5 inhibits TGF-β-driven epithelial-mesenchymal transition in prostate cancer by blocking SMAD3 phosphorylation. Conversely, evidence suggests that the ETS family plays a significant role in promoting certain pathways. Wang *et al.* [50] suggested that EHF directly upregulated the expression of TGF-β1 at the transcriptional level and activated the signal transduction. Similarly, the ERG fusion gene is known to induce this signaling pathway in prostate cancer, fulfilling a similar function [51].

Another key finding of this study involves utilizing univariate Cox regression analysis and LASSO regression analysis to construct prognostic signatures for predicting OS and DFS. The OS prognostic signatures comprise ELK3, ERG, ETV2, and ETV6, while the DFS prognostic signatures include ETV1, ELK3, SPIC, and ETV2. Among these, ELK3, ERG, ETV1 and SPIC are posited to act as oncogenes, a theory supported by multiple reports. ELK3, in particular, has been identified with high expression in various cancers including breast cancer [52-54], liver cancer [43], non-small cell lung carcinoma, and prostate cancer, playing roles in tumor development and progression [55, 56]. Studies have shown that ELK3 can activate the PI3K-Akt pathway to regulate autophagy-mediated sensitivity of breast cancer to adriamycin chemotherapy [42]. Yoo et al. demonstrated that ELK3 can also activate the transcription of c-fos and enhance breast cancer cell growth and transformation [57]. Moreover, ELK3 has been reported to promote liver cancer cell migration and invasion by regulating HIF-1α [43]. This study confirms the overexpression of ELK3 in GC and demonstrates that inhibiting its expression both *in vitro* and *in vivo* can suppress GC cell progression. ERG is critically involved in cell proliferation, differentiation, angiogenesis, inflammation, and apoptosis; it

can influence the Warburg effect by binding to the HK2 promoter and affecting cervical cancer progression [58]. Research on prostate cancer has indicated that ERG directly targets and activates FZD8 of the Wnt pathway by binding to its promoter, promoting tumor metastasis [59]. ETV1, known to be upregulated in GC, acts as a favorable prognostic factor by inducing epithelial-mesenchymal transition through Snail transcription regulation [37]. In pancreatic cancer, it can regulate the expansion and metastasis of the stroma [60]. In aggressive prostate cancer, overexpressed ETV1 promotes TGF-β signal transduction and tumor progression [61]. Research in lung cancer suggests SPIC activates SNHG6 transcriptionally, inducing the miR-485-3p/VPS45 axis to promote non-small cell lung cancer progression [62], and upregulates PARP9 in cervical cancer to enhance tumorigenicity [63].

To evaluate the prognostic signature against other clinicopathological features, ROC curves, univariate, and multivariate Cox regression analyses were employed. The constructed prognostic signature outperformed other clinicopathological features in predicting GC patient prognosis. A nomogram was developed to predict 1-5 years OS and DFS probabilities for GC patients. Calibration plot verification indicated that both nomograms accurately forecast GC patient survival, offering valuable insights into effective treatment options for clinicians. Furthermore, hub genes within the prognostic signature were validated using an external database, confirming their association with survival and protein-level expression. Specifically, ELK3, pivotal in both OS and DFS prognostic signatures, was found highly expressed in GC compared to matched normal samples in 30 clinical specimens. High ELK3 expression correlates with significant clinical and pathological characteristics. *In vitro* and *in vivo* experiments showed that ELK3 knockdown in GC cells diminished tumor proliferation, invasion, and migration capabilities, suggesting ELK3 as a potential therapeutic target for GC treatment.

Acknowledgements

We sincerely thank TCGA and GTEx database for the massive shared resources.

All gastric cancer clinical samples involved were approved by the patient's informed consent.

Disclosure of conflict of interest

None.

Abbreviations

ETS, E26; GC, Gastric cancer; OS, overall survival; DFS, disease-free survival; TCGA, The Cancer Genome Atlas; GTEx, Genotype-Tissue Expression; LASSO, least absolute shrinkage and selection operator; FPKM, fragments per kilobase million; STAD, stomach adenocarcinomas; TPM, transcripts per million; PPI, Protein-protein interaction; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; HPA, Human Protein Atlas; CCLE, Cancer Cell Line Encyclopedia.

Address correspondence to: Liqiang Zhou, Department of General Surgery, The Second Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang 330006, Jiangxi, P. R. China. E-mail: doczhoulq@163.com; Zhiqing Chen, Jiangxi Provincial Key Laboratory of Molecular Medicine, The Second Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang 330006, Jiangxi, P. R. China. E-mail: czq033021@163.com

References

[1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018; 68: 394-424.

[2] Chen W, Zheng R, Zhang S, Zhao P, Zeng H and Zou X. Report of cancer incidence and mortality in China, 2010. Ann Transl Med 2014; 2: 61.

[3] Van Cutsem E, Sagaert X, Topal B, Haustermans K and Prenen H. Gastric cancer. Lancet 2016; 388: 2654-2664.

[4] Sun P, Xiang J and Chen Z. Meta-analysis of adjuvant chemotherapy after radical surgery for advanced gastric cancer. Br J Surg 2009; 96: 26-33.

[5] Sementchenko VI and Watson DK. Ets target genes: past, present and future. Oncogene 2000; 19: 6533-6548.

[6] Sharrocks AD. The ETS-domain transcription factor family. Nat Rev Mol Cell Biol 2001; 2: 827-837.

[7] Findlay VJ, Larue AC, Turner DP, Watson PM and Watson DK. Understanding the role of ETS-mediated gene regulation in complex biological processes. Adv Cancer Res 2013; 119: 1-61.

[8] Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, Pienta KJ, Rubin MA and Chinnaiyan AM. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 2005; 310: 644-648.

[9] Kumar-Sinha C, Tomlins SA and Chinnaiyan AM. Recurrent gene fusions in prostate cancer. Nat Rev Cancer 2008; 8: 497-511.

[10] Baldus CD, Liyanarachchi S, Mrózek K, Auer H, Tanner SM, Guimond M, Ruppert AS, Mohamed N, Davuluri RV, Caligiuri MA, Bloomfield CD and de la Chapelle A. Acute myeloid leukemia with complex karyotypes and abnormal chromosome 21: amplification discloses overexpression of APP, ETS2, and ERG genes. Proc Natl Acad Sci U S A 2004; 101: 3915-3920.

[11] Ghadersohi A and Sood AK. Prostate epithelium-derived Ets transcription factor mRNA is overexpressed in human breast tumors and is a candidate breast tumor marker and a breast tumor antigen. Clin Cancer Res 2001; 7: 2731-2738.

[12] Mueller BU, Pabst T, Osato M, Asou N, Johansen LM, Minden MD, Behre G, Hiddemann W, Ito Y and Tenen DG. Heterozygous PU.1 mutations are associated with acute myeloid leukemia. Blood 2003; 101: 2074.

[13] Shi J, Qu Y, Li X, Sui F, Yao D, Yang Q, Shi B, Ji M and Hou P. Increased expression of EHF via gene amplification contributes to the activation of HER family signaling and associates with poor survival in gastric cancer. Cell Death Dis 2016; 7: e2442.

[14] Wu J, Qin W, Wang Y, Sadik A, Liu J, Wang Y, Song P, Wang X, Sun K, Zeng J and Wang L. SPDEF is overexpressed in gastric cancer and triggers cell proliferation by forming a positive regulation loop with FoxM1. J Cell Biochem 2018; 119: 9042-9054.

[15] Cancer Genome Atlas Research Network; Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C and Stuart JM. The cancer genome atlas pan-cancer analysis project. Nat Genet 2013; 45: 1113-1120.

[16] Li B, Ruotti V, Stewart RM, Thomson JA and Dewey CN. RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics 2010; 26: 493-500.

[17] Leek JT, Johnson WE, Parker HS, Jaffe AE and Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 2012; 28: 882-883.

[18] Hu H, Miao YR, Jia LH, Yu QY, Zhang Q and Guo AY. AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. Nucleic Acids Res 2019; 47: D33-D38.

[19] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W and Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015; 43: e47.

[20] Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ and von Mering C. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res 2011; 39: D561-568.

[21] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003; 13: 2498-2504.

[22] Wingender E. The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. Brief Bioinform 2008; 9: 326-332.

[23] Han H, Cho JW, Lee S, Yun A, Kim H, Bae D, Yang S, Kim CY, Lee M, Kim E, Lee S, Kang B, Jeong D, Kim Y, Jeon HN, Jung H, Nam S, Chung M, Kim JH and Lee I. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. Nucleic Acids Res 2018; 46: D380-D386.

[24] Yu G, Wang LG, Han Y and He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics 2012; 16: 284-287.

[25] Walter W, Sánchez-Cabo F and Ricote M. GOplot: an R package for visually combining expression data with functional analysis. Bioinformatics 2015; 31: 2912-2914.

[26] Szász AM, Lánczky A, Nagy Á, Förster S, Hark K, Green JE, Boussioutas A, Busuttil R, Szabó A and Győrffy B. Cross-validation of survival associated biomarkers in gastric cancer using transcriptomic data of 1,065 patients. Oncotarget 2016; 7: 49322-49333.

[27] Luck K, Kim DK, Lambourne L, Spirohn K, Begg BE, Bian W, Brignall R, Cafarelli T, Campos-Laborie FJ, Charloteaux B, Choi D, Coté AG, Daley M, Deimling S, Desbuleux A, Dricot A, Gebbia M, Hardy MF, Kishore N, Knapp JJ, Kovács IA, Lemmens I, Mee MW, Mellor JC, Pollis C, Pons C, Richardson AD, Schlabach S, Teeking B, Yadav A, Babor M, Balcha D, Basha O, Bowman-Colin C, Chin SF, Choi SG, Colabella C, Coppin G, D'Amata C, De Ridder D, De Rouck S, Duran-Frigola M, Ennajdaoui H, Goebels F, Goehring L, Gopal A, Haddad G, Hatchi E, Helmy M, Jacob Y, Kassa Y, Landini S, Li R, van Lieshout N, MacWilliams A, Markey D,

Paulson JN, Rangarajan S, Rasla J, Rayhan A, Rolland T, San-Miguel A, Shen Y, Sheykhkarimli D, Sheynkman GM, Simonovsky E, Taşan M, Tejeda A, Tropepe V, Twizere JC, Wang Y, Weatheritt RJ, Weile J, Xia Y, Yang X, Yeger-Lotem E, Zhong Q, Aloy P, Bader GD, De Las Rivas J, Gaudet S, Hao T, Rak J, Tavernier J, Hill DE, Vidal M, Roth FP and Calderwood MA. A reference map of the human binary protein interactome. Nature 2020; 580: 402-408.

[28] Livak KJ and Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods 2001; 25: 402-408.

[29] Zhang YQ, Pei JH, Shi SS, Guo XS, Cui GY, Li YF, Zhang HP and Hu WQ. CRISPR/Cas9-mediated knockout of the PDEF gene inhibits migration and invasion of human gastric cancer AGS cells. Biomed Pharmacother 2019; 111: 76-85.

[30] Liao YL, Hu LY, Tsai KW, Wu CW, Chan WC, Li SC, Lai CH, Ho MR, Fang WL, Huang KH and Lin WC. Transcriptional regulation of miR-196b by ETS2 in gastric cancer cells. Carcinogenesis 2012; 33: 760-769.

[31] Das L, Kokate SB, Rath S, Rout N, Singh SP, Crowe SE, Mukhopadhyay AK and Bhattacharyya A. ETS2 and Twist1 promote invasiveness of Helicobacter pylori-infected gastric cancer cells by inducing Siah2. Biochem J 2016; 473: 1629-1640.

[32] Das L, Kokate SB, Dixit P, Rath S, Rout N, Singh SP, Crowe SE and Bhattacharyya A. Membrane-bound β-catenin degradation is enhanced by ETS2-mediated Siah1 induction in Helicobacter pylori-infected gastric cancer cells. Oncogenesis 2017; 6: e327.

[33] Yu Y, Zhang YC, Zhang WZ, Shen LS, Hertzog P, Wilson TJ and Xu DK. Ets1 as a marker of malignant potential in gastric carcinoma. World J Gastroenterol 2003; 9: 2154-2159.

[34] Nakayama T, Ito M, Ohtsuru A, Naito S, Nakashima M, Fagin JA, Yamashita S and Sekine I. Expression of the Ets-1 proto-oncogene in human gastric carcinoma: correlation with tumor invasion. Am J Pathol 1996; 149: 1931-1939.

[35] Tsutsumi S, Kuwano H, Nagashima N, Shimura T, Mochiki E and Asao T. Ets-1 expression in gastric cancer. Hepatogastroenterology 2005; 52: 654-656.

[36] Tsutsumi S, Kuwano H, Asao T, Nagashima K, Shimura T and Mochiki E. Expression of Ets-1 angiogenesis-related protein in gastric cancer. Cancer Lett 2000; 160: 45-50.

[37] Li Z, Zhang L, Ma Z, Yang M, Tang J, Fu Y, Mao Y, Hong X and Zhang Y. ETV1 induces epithelial to mesenchymal transition in human gastric cancer cells through the upregulation of Snail expression. Oncol Rep 2013; 30: 2859-2863.

[38] Yamamoto H, Horiuchi S, Adachi Y, Taniguchi H, Nosho K, Min Y and Imai K. Expression of ets-related transcriptional factor E1AF is associated with tumor progression and over-expression of matrilysin in human gastric cancer. Carcinogenesis 2004; 25: 325-332.

[39] Keld R, Guo B, Downey P, Cummins R, Gulmann C, Ang YS and Sharrocks AD. PEA3/ETV4-related transcription factors coupled with active ERK signalling are associated with poor prognosis in gastric adenocarcinoma. Br J Cancer 2011; 105: 124-130.

[40] Sizemore GM, Pitarresi JR, Balakrishnan S and Ostrowski MC. The ETS family of oncogenic transcription factors in solid tumours. Nat Rev Cancer 2017; 17: 337-351.

[41] Xu L, Hu H, Zheng LS, Wang MY, Mei Y, Peng LX, Qiang YY, Li CZ, Meng DF, Wang MD, Liu ZJ, Li XJ, Huang BJ and Qian CN. ETV4 is a theranostic target in clear cell renal cell carcinoma that promotes metastasis by activating the pro-metastatic gene FOSL1 in a PI3K-AKT dependent manner. Cancer Lett 2020; 482: 74-89.

[42] Park JH, Kim KP, Ko JJ and Park KS. PI3K/Akt/mTOR activation by suppression of ELK3 mediates chemosensitivity of MDA-MB-231 cells to doxorubicin by inhibiting autophagy. Biochem Biophys Res Commun 2016; 477: 277-282.

[43] Lee JH, Hur W, Hong SW, Kim JH, Kim SM, Lee EB and Yoon SK. ELK3 promotes the migration and invasion of liver cancer stem cells by targeting HIF-1α. Oncol Rep 2017; 37: 813-822.

[44] Zhang Y, Wu J, Ye M, Wang B, Sheng J, Shi B and Chen H. ETS1 is associated with cisplatin resistance through IKKα/NF-κB pathway in cell line MDA-MB-231. Cancer Cell Int 2018; 18: 86.

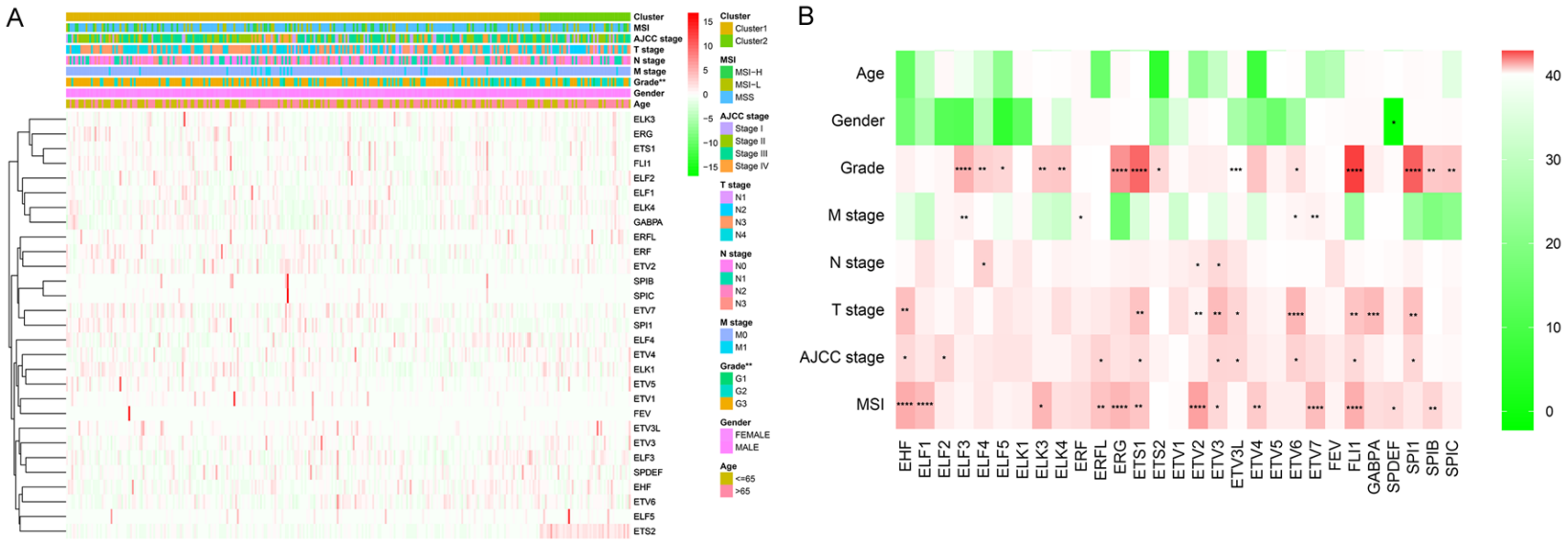
[45] Longoni N, Sarti M, Albino D, Civenni G, Malek A, Ortelli E, Pinton S, Mello-Grand M, Ostano P, D'Ambrosio G, Sessa F, Garcia-Escudero R, Thalmann GN, Chiorino G, Catapano CV and Carbone GM. ETS transcription factor ESE1/ELF3 orchestrates a positive feedback loop that constitutively activates NF-κB and drives prostate cancer progression. Cancer Res 2013; 73: 4533-4547.

[46] Wang J, Cai Y, Shao LJ, Siddiqui J, Palanisamy N, Li R, Ren C, Ayala G and Ittmann M. Activation of NF-{kappa}B by TMPRSS2/ERG fusion isoforms through toll-like receptor-4. Cancer Res 2011; 71: 1325-1333.

[47] Albino D, Civenni G, Rossi S, Mitra A, Catapano CV and Carbone GM. The ETS factor ESE3/EHF represses IL-6 preventing STAT3 activation and expansion of the prostate cancer stem-like compartment. Oncotarget 2016; 7: 76756-76768.

[48] Chen H, Chen W, Zhang X, Hu L, Tang G, Kong J and Wang Z. E26 transformation (ETS)-specif-

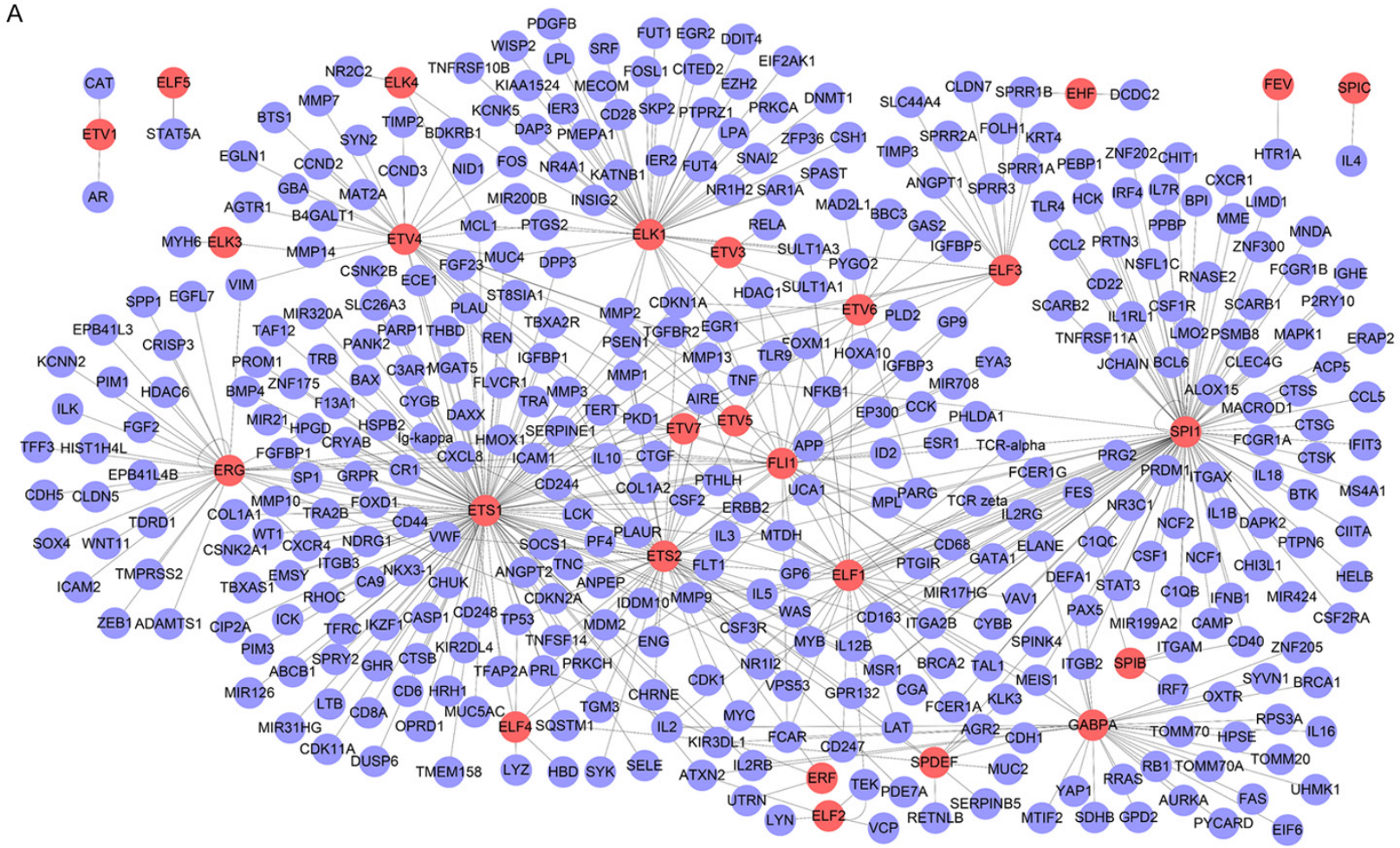
ic related transcription factor-3 (ELF3) orchestrates a positive feedback loop that constitutively activates the MAPK/Erk pathway to drive thyroid cancer. Oncol Rep 2019; 41: 570-578.
[49] Yao B, Zhao J, Li Y, Li H, Hu Z, Pan P, Zhang Y, Du E, Liu R and Xu Y. Elf5 inhibits TGF-β-driven epithelial-mesenchymal transition in prostate cancer by repressing SMAD3 activation. Prostate 2015; 75: 872-882.
[50] Wang L, Ai M, Nie M, Zhao L, Deng G, Hu S, Han Y, Zeng W, Wang Y, Yang M and Wang S. EHF promotes colorectal carcinoma progression by activating TGF-β1 transcription and canonical TGF-β signaling. Cancer Sci 2020; 111: 2310-2324.
[51] Ratz L, Laible M, Kacprzyk LA, Wittig-Blaich SM, Tolstov Y, Duensing S, Altevogt P, Klauck SM and Sültmann H. TMPRSS2:ERG gene fusion variants induce TGF-β signaling and epithelial to mesenchymal transition in human prostate cancer cells. Oncotarget 2017; 8: 25115-25130.
[52] Kong SY, Kim KS, Kim J, Kim MK, Lee KH, Lee JY, Oh N, Park JI, Park JH, Heo SH, Shim SH, Lee DR, Kim KP and Park KS. The ELK3-GATA3 axis orchestrates invasion and metastasis of breast cancer cells in vitro and in vivo. Oncotarget 2016; 7: 65137-65146.
[53] Oh N, Park JI, Park JH, Kim KS, Lee DR and Park KS. The role of ELK3 to regulate peritumoral lymphangiogenesis and VEGF-C production in triple negative breast cancer cells. Biochem Biophys Res Commun 2017; 484: 896-902.
[54] Heo SH, Lee JY, Yang KM and Park KS. ELK3 expression correlates with cell migration, invasion, and membrane type 1-matrix metalloproteinase expression in MDA-MB-231 breast cancer cells. Gene Expr 2015; 16: 197-203.
[55] Sloan KA, Marquez HA, Li J, Cao Y, Hinds A, O'Hara CJ, Kathuria S, Ramirez MI, Williams MC and Kathuria H. Increased PEA3/E1AF and decreased Net/Elk-3, both ETS proteins, characterize human NSCLC progression and regulate caveolin-1 transcription in Calu-1 and NCI-H23 NSCLC cell lines. Carcinogenesis 2009; 30: 1433-1442.
[56] Mao Y, Li W, Hua B, Gu X, Pan W, Chen Q, Xu B, Wang Z and Lu C. Silencing of ELK3 induces S-M phase arrest and apoptosis and upregulates SERPINE1 expression reducing migration in prostate cancer cells. Biomed Res Int 2020; 2020: 2406159.
[57] Yoo SM, Lee CJ, An HJ, Lee JY, Lee HS, Kang HC, Cho SJ, Kim SM, Park J, Kim DJ and Cho YY. RSK2-mediated ELK3 activation enhances cell transformation and breast cancer cell growth by regulation of c-fos promoter activity. Int J Mol Sci 2019; 20: 1994.
[58] Zhang Z, Chen F, Li S, Guo H, Xi H, Deng J, Han Q and Zhang W. ERG the modulates Warburg effect and tumor progression in cervical cancer. Biochem Biophys Res Commun 2020; 522: 191-197.
[59] Chakravarthi BVSK, Chandrashekar DS, Hodigere Balasubramanya SA, Robinson AD, Carskadon S, Rao U, Gordetsky J, Manne U, Netto GJ, Sudarshan S, Palanisamy N and Varambally S. Wnt receptor Frizzled 8 is a target of ERG in prostate cancer. Prostate 2018; 78: 1311-1320.
[60] Heeg S, Das KK, Reichert M, Bakir B, Takano S, Caspers J, Aiello NM, Wu K, Neesse A, Maitra A, Iacobuzio-Donahue CA, Hicks P and Rustgi AK. ETS-transcription factor ETV1 regulates stromal expansion and metastasis in pancreatic cancer. Gastroenterology 2016; 151: 540-553, e514.
[61] Oh S, Shin S, Song H, Grande JP and Janknecht R. Relationship between ETS transcription factor ETV1 and TGF-β-regulated SMAD proteins in prostate cancer. Sci Rep 2019; 9: 8186.
[62] Gao N and Ye B. SPI1-induced upregulation of lncRNA SNHG6 promotes non-small cell lung cancer via miR-485-3p/VPS45 axis. Biomed Pharmacother 2020; 129: 110239.
[63] Tao L, Wang X and Zhou Q. Long noncoding RNA SNHG16 promotes the tumorigenicity of cervical cancer cells by recruiting transcriptional factor SPI1 to upregulate PARP9. Cell Biol Int 2020; 44: 773-784.

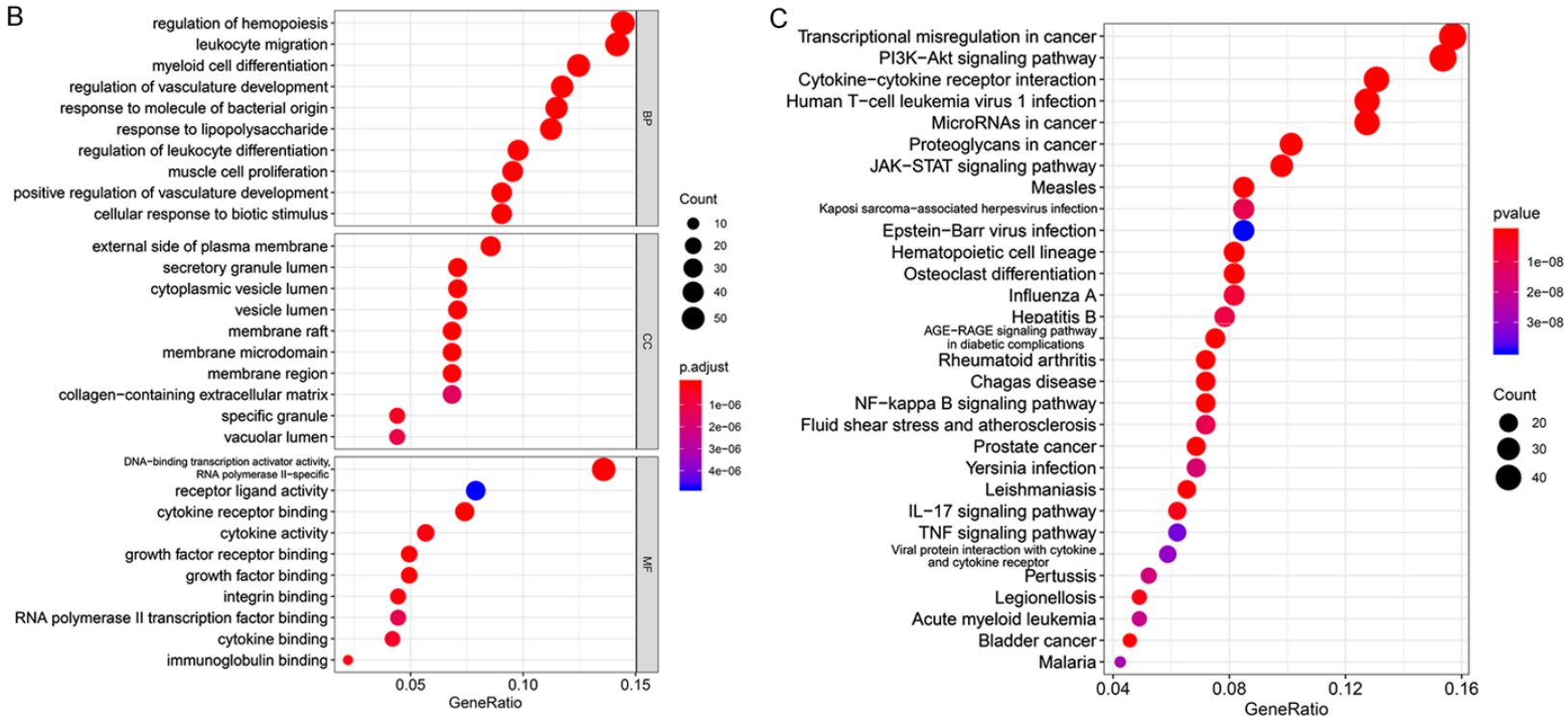


Supplementary Figure 1. Clinical outcomes of TCGA-STAD patients have in association with ETS factors and the two different clusters. A. Heatmap and clinicopathological characteristics of the two clusters defined by the shared expression of ETS factors. B. Heatmap illustrating the association between ETS factors and patient clinical characteristics. *P<0.05, **P<0.01, ***P<0.001, ****P<0.001. TCGA, The Cancer Genome Atlas; STAD, stomach adenocarcinoma.

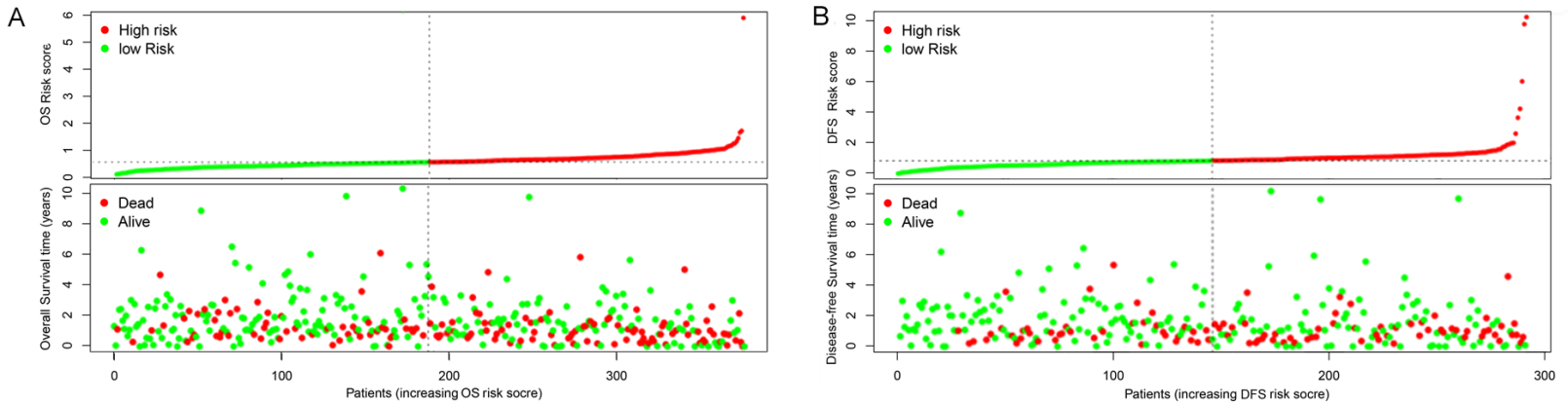
The oncogene ELK3 was identified from the prognostic model of E26 TFs in GC

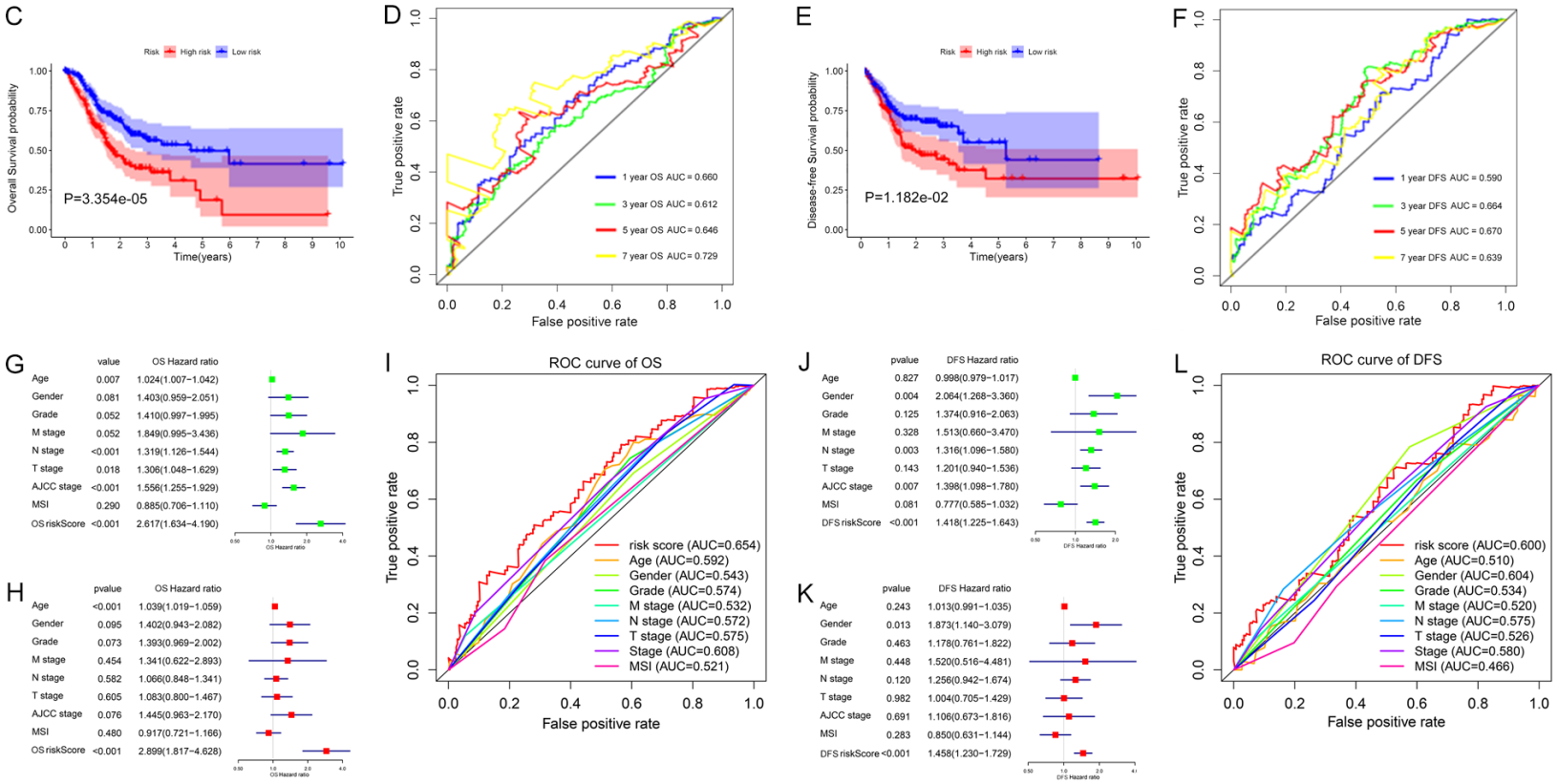
A



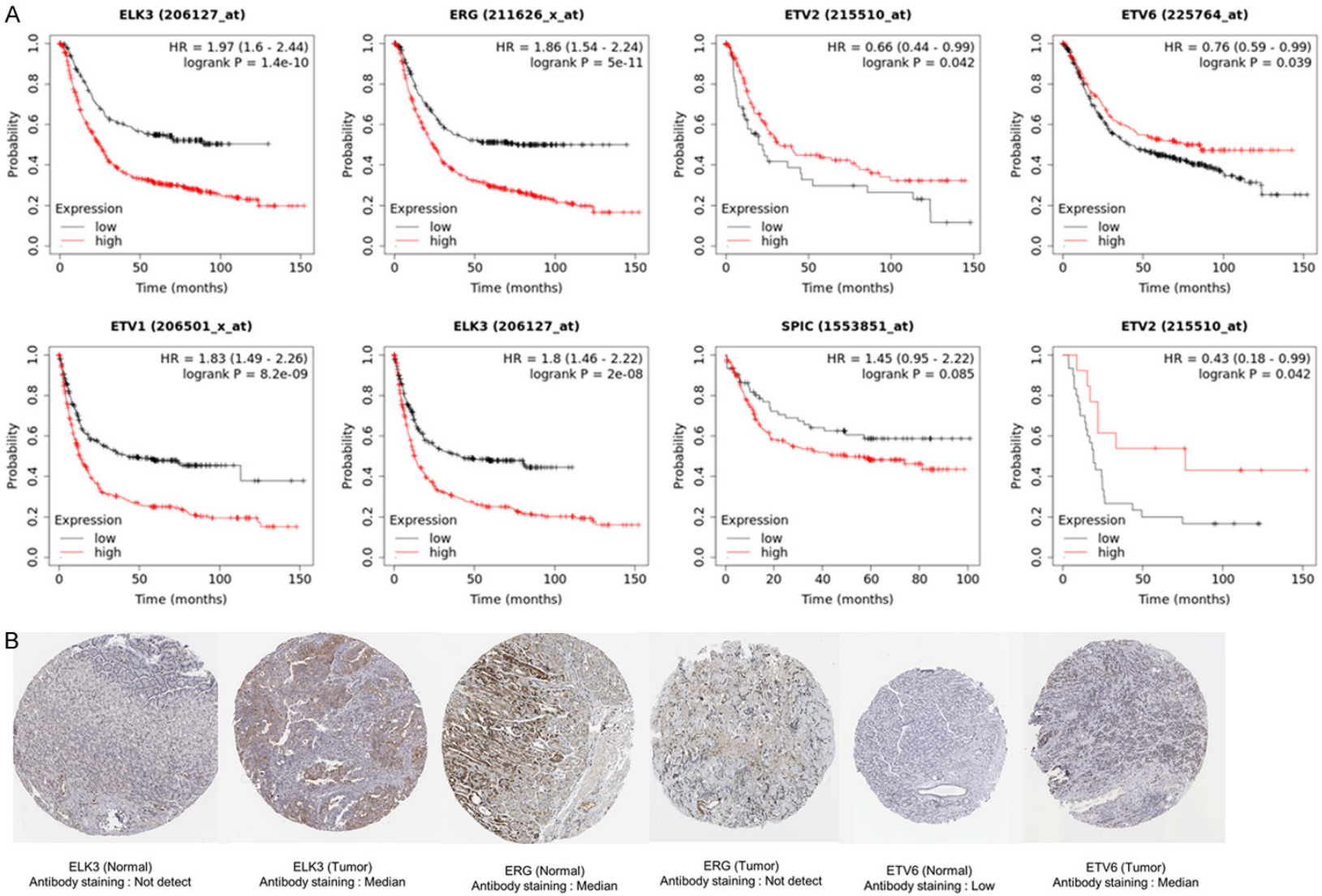


Supplementary Figure 2. ETS factor regulatory network and enrichment analysis. A. ETS transcription factor target genes were obtained from TRANSFAC and TRRUST and a transcriptional regulatory network was established. B. Top 10 results of each entry for GO enrichment analysis of ETS transcription factor target genes. C. Top 30 results of Kyoto Encyclopedia of Genes and Genomes enrichment analysis of ETS transcription factor target genes.





Supplementary Figure 3. Construction and verification of OS and DFS prognostic signatures. A and B. Distribution of survival status in the OS/DFS signature as the risk score increased. C and E. Survival analysis of the two subgroups stratified based on the median of OS/DFS risk scores calculated by multivariate Cox regression analysis. D and F. ROC curve for evaluating the predicting ability of the prognostic signature. G and J. Univariate Cox analyses of the OS/DFS signature based on the risk score and clinicopathological parameters in the TCGA-STAD cohort. H and K. Multivariate Cox analyses of the OS/DFS signature based on the risk score and clinicopathological parameters in the TCGA-STAD cohort. I and L. ROC curve evaluating the prognostic performance of the OS/DFS risk score and other clinicopathological parameters. OS, overall survival; DFS, disease-free survival; TCGA, The Cancer Genome Atlas; STAD, stomach adenocarcinoma.



Supplementary Figure 4. Validation of the prognosis and expression of hub ETS factors. A. Validation of OS for four OS-related hub ETS factors using the Kaplan-Meier plotter; Validation of DFS for four DFS-related hub ETS factors using the Kaplan-Meier plotter. B. The Human Protein Atlas database was used to verify the protein expression of ELK3, ERG and ETV6. OS, overall survival; DFS, disease-free survival.