Transcription Elongation Factor A-like 7 is a Stemness-Related Prognostic Factor in Gastric Cancer

Haisheng Qian (✉ qianhschn@foxmail.com)

Jiangsu Province Hospital and Nanjing Medical University First Affiliated Hospital

<https://orcid.org/0000-0002-5329-0182>

Xin Gao

Jiangsu Province Hospital and Nanjing Medical University First Affiliated Hospital

Rui Ma

Jiangsu Health Vocational College

Wenjie Li

Jiangsu Province Hospital and Nanjing Medical University First Affiliated Hospital

Zhen Yang

Jiangsu Province Hospital and Nanjing Medical University First Affiliated Hospital

Guoxin Zhang

Jiangsu Province Hospital and Nanjing Medical University First Affiliated Hospital

Xuan Li

Jiangsu Province Hospital and Nanjing Medical University First Affiliated Hospital

<https://orcid.org/0000-0003-2703-1950>

Research Article

Keywords: TCEAL7, stemness, OCLR, mRANsi, gastric cancer

Posted Date: April 16th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-415488/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Stemness is described as the potential for self-renewal and differentiation from the cell-of-origin. A previous study calculated the mRNA expression-based stemness index (mRNAsi) based on a one-class logistic regression machine learning algorithm for describing stemness features of cancer. We aim to identify stemness-related prognostic genes in gastric cancer (GC) based on mRNAsi using bioinformatics analysis.

Methods: The WGCNA analysis was performed to find the relevant gene modules to mRNAsi. Gene ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) signaling pathways annotation analysis were performed on genes in blue module. The overall survival analysis, univariate Cox regression analysis and the least absolute shrinkage and selection operator (LASSO) regression model were used to identify prognostic genes highly associated with survival. The multivariate Cox regression analysis was performed to analysis prognostic factors. The nomogram was constructed according to the result of multivariate analysis. qPCR, Western Blot and IHC staining were applied

Results: The mRNAsi of tumors is higher than normal tissues, and there was a significant difference in overall survival (OS) between the high and low mRNAsi GC groups. TCEAL7 was selected to be the key gene associated with mRNAsi and prognosis according to the result of least absolute shrinkage and selection operator (LASSO) Cox regression. The expression level of TCEAL7 was lower in tumors than in normal tissues, but high TCEAL7 level group showed a worse OS than low TCEAL7 level group in GC. Based on the result of multivariable Cox regression analysis which including TCEAL7 and clinical characteristics, a nomogram for predicting GC 1-, 3-, and 5-year survival was established. The C-index and the AUC (Area Under Curve) of the model indicated that the model has a good discrimination ability. Additionally, the calibration curves of 3- and 5-year OS rates showed the model fits well. The experimental validation of the expression of TCEAL7 in GC and normal tissues were consistent with the above.

Conclusions: In summary, we verified mRNAsi was associated with the prognosis of GC patients. And TCEAL7 was finally identified as the key gene correlated with stemness features and prognosis in GC.

Introduction

As an important disease worldwide, gastric cancer (GC) is the 5th most common cancer and 3rd most common cause of cancer-related death, with about 784,000 deaths globally in 2018[1]. Though surgery and chemotherapy improve survival and quality of life for patients with advanced gastric cancer, overall survival (OS) in GC patients is still poor due to recurrence, metastasis, and chemoresistance[2]. Stemness is described as the potential for self-renewal and differentiation from the cell-of-origin. The progression of cancer involves gradual loss of differentiated phenotype and the acquisition of stemness features[3]. Previous studies have shown that tumors with more stemness features are more apt to result in metastasis and chemoresistance, leading to disease progression and poor prognosis[3–5]. Therefore, it

would be helpful to identify key genes associated with cancer stemness features to predict and improve prognosis.

In a previous study, Malta et al. applied a one-class logistic regression machine learning algorithm (OCLR) to extract transcriptomic and epigenetic feature sets derived from non-transformed pluripotent stem cells and their differentiated progeny[6]. Then, they calculated the mRNA expression-based stemness index (mRNAsi), DNA methylation-based stemness index (mDNAsi), and mRNAsi and the epigenetic regulation based-index (EREG-mRNAsi) based on the OCLR-derived feature sets. This provided a new way for describing stemness features of cancer. Many scholars have identified stemness features-associated genes and possible signal pathways in bladder cancer (BLCA), breast cancer (BRCA), and lung adenocarcinoma (LUAD) based on the OCLR-based stemness index[7–9]. Preceding studies also suggested the importance of mRNAsi-related genes in risk stratification and survival prediction in lower-grade glioma (LGG) and gastric cancer[10, 11]. However, the predictive power of the nine-mRNAsi-related-gene risk model in the study of gastric cancer was not very satisfactory compared to that of LGG, and this gene model also lacked external validation.

The transcription elongation factor A-like 7 (TCEAL7) is a member of the transcription elongation factor A (SII)-like gene family, which was first cloned as a proapoptotic nuclear protein and shown to function as a tumor suppressor gene contributing to the activation of oncogenic pathways in ovarian cancer[12–15]. Besides, decreased expression of TCEAL7 has been reported in different types of cancer cell lines including endometrial cancer, glioblastoma, non-small cell lung cancer and gastric cancer[16–20], but the role of TCEAL7 in these cancers is still not very clear.

In this study, we explored the mRNAsi level in different clinical features and verified its relationship with prognosis. By applying weighted gene co-expression network analysis (WGCNA) and least absolute shrinkage and selection operator (LASSO) Cox regression analysis based on The Cancer Genome Atlas (TCGA) database, TCEAL7 was identified as the most relevant gene to mRNAsi and prognosis, which showed a good prognostic value in GC. The microarray GSE66229 (composed of GSE62254 and GSE66222) dataset from the Gene Expression Omnibus (GEO) database was utilized to further confirm the expression and prognostic value of TCEAL7.

Materials And Methods

Bioinformatics analysis

Data Processing

The RNA-seq data of 32 normal and 375 stomach adenocarcinoma (STAD) samples from The Cancer Genome Atlas (TCGA) database and 174 normal gastric tissue samples from the Genotype-Tissue Expression (GTEx) project were downloaded from the University of California Santa Cruz (UCSC) Xena website. The clinical data of tumor samples were also obtained, which included: gender, age, gender, TNM stages, survival time, and vital status. The RNA-seq expression profiles were quantified by transcripts per

kilobase of exon model per million mapped reads (TPM) normalized estimation and log2-based transformation. Next, R package “sva” was used to remove the batch effects by the methods of ComBat. Then, differentially expressed genes (DEGs) were identified by “limma” package of R software under the selection criteria of absolute value of the log2-transformed fold change (FC) > 1, the adjusted P-value < 0.05, and false discovery rate (FDR) < 0.05.

The Expression Level and Survival Analysis of mRNAsi

In a previous study[6], the gene expression-based stemness index of the STAD samples from TCGA database was calculated by a one class logistic regression machine learning algorithm (OCLR), which ranges from zero (no gene expression) to one (complete gene expression). The mRNAsi was obtained from this research. Boxplots were drawn to observe mRNAsi expression in different clinical features. The “surv_cutpoint” function of “survminer” package was used to find the optimal cutoff value for the dichotomy of mRNAsi related to overall survival. The Kaplan Meier analysis was performed for samples with high and low mRNAsi.

Weighted Gene Co-expression Network Analysis (WGCNA)

The WGCNA analysis was performed by the “WGCNA” package[21] to build a co-expression network of DEGs. First, the co-expression similarity matrix and Pearson correlation matrix were constructed. Next, the power function $a_{mn} = |c_{mn}|^\beta$ (c_{mn} = Pearson correlation between gene m and gene n; a_{mn} = adjacency between gene m and gene n) was used to construct a weighted adjacency matrix. The parameter β was set to achieve a scale-free co-expression network. Then, a topological overlap matrix (TOM) was built to measure the network connectivity of genes defined as the sum of adjacent genes generated by all other networks. Finally, a hierarchical clustering tree and gene modules were performed based on TOM-based dissimilarity measurements. The minimum module size was 30 and cut height was 0.2.

The mRNAsi and epigenetically regulated mRNAsi (EREG-mRNAsi) were selected as the external traits to identify the mRNAsi-related modules and genes. The relationship between the modules and mRNAsi was calculated, and the most relevant module was selected for subsequent analysis to explore key genes. Gene significance (GS, the correlation between genes and sample traits) and module membership (MM, correlation between the module’s genes and gene expression profiles) for each gene in the module were calculated. The thresholds for screening key genes in the module were as follows: cor. gene MM > 0.8 and cor. gene GS > 0.7.

Functional Enrichment Annotation and Pathway Analysis of Key Genes

Gene ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) signaling pathways annotation analysis were performed by clusterProfiler package[22] to explore the biological functions of the key genes. The statistical significance values were set as follows: P-value < 0.05, and false discovery rate (FDR) < 0.05.

LASSO Cox Regression Analysis and Construction of TCEAL7 Expression-based Model

The univariate Cox regression analysis was used to identify prognostic genes highly associated with survival. Next, the least absolute shrinkage and selection operator (LASSO) regression model was performed to further optimize the prognostic genes, and the TCEAL7 was finally selected as the prognostic gene. As mentioned above in the mRNAsi analysis, boxplots were drawn to observe the expression of TCEAL7 in different clinical features, and the Kaplan Meier analysis was performed for samples with high and low expression level. Multivariate Cox regression analysis was performed to analysis prognostic factors in the model, including age, gender, TNM stages, TCEAL7. The nomogram was constructed according to the result of multivariate analysis using “rms” package.

The Cox model validation was performed by AUC (Area Under Curve) and C-index for discrimination ability, calibration curves for calibration ability. Bootstraps with 1000 resamples were adopted to decrease overfit bias.

To compare the accuracy and discrimination of different models, the net reclassification improvement (NRI), the integrated discrimination improvement (IDI), and the median improvement in risk score (MIRS) were applied by using “survIDINRI” package[23].

Data Validation

To further verify the expression and prognostic value of TCEAL-7, the microarray GSE66229 (composed of GSE62254 and GSE66222) dataset was downloaded from the Gene Expression Omnibus (GEO) database, which had 100 normal and 300 tumor samples. Boxplots for expression of TCEAL-7 in different clinical features and the Kaplan Meier analysis of high and low expression group were performed. We also built Cox regression models to manifest the prognostic value of TCEAL-7.

Clinical specimens

5 pairs of GC tissues and paired adjacent normal tissues were obtained from patients who underwent surgery at Nanjing Medical University's First Affiliated Hospital in China. All samples were confirmed by histopathology. The samples were frozen in liquid nitrogen and then stored at -80 °C until use. Written consent was obtained from all patients. The study were approved by First Affiliated Hospital of Nanjing Medical University Ethics Committee.

Cell culture

Human GC cell lines (SGC-7901, HGC-27, and AGS) and the normal human gastric epithelial cell line (GES-1) were purchased from the American Type Culture Collection (ATCC, USA). All these cells were cultured at 37 °C in a humidified atmosphere with 5% CO₂, and cultured in RPMI 1640 medium supplemented with 1% penicillin/streptomycin, 10% fetal bovine serum (FBS). All culture medium reagents were obtained from Gibco, Brazil.

RNA extraction and quantitative real-time PCR (qRT-PCR)

Total RNA extraction was isolated from cell lines with TRIzol reagent (Invitrogen, USA). The cDNA of the RNAs were created with the protocol of the PrimeScript RT Reagent Kit (TaKaRa, Japan). TB Green® Premix Ex Taq™ (TaKaRa, Japan,) was used to conduct the quantitative real-time PCR (qRT -PCR). qRT -PCR was used to detect the expression levels of TCEAL7. The mRNA levels were normalized to that of GAPDH. The PCR primer sequences were synthesized by Tsingke (Nanjing, China) (human-TCEAL7-F: 5-GAAAAACGCCCGTATGGAGAA-3, human-TCEAL7-R: 5-GCAGCCTCTGTCTAAAATTCCCT-3; human-GAPDH-F: 5-GGAGTCCACTGGCGTCTTCA-3, human-GAPDH-R: 5-GTCATGAGTCCTTCCACGATACC-3). RNA expression fold changes was determined with the $2^{-\Delta\Delta C_t}$ formula.

Protein extraction and western blot analysis (WB)

Radioimmunoprecipitation assay (RIPA) buffer was used for protein extraction from cells. Bicinchoninic acid (BCA) kit (Beyotime, China) was used to determine protein concentration. The equal amounts of protein samples were run on 15% acrylamide gels by SDS-PAGE and then moved onto to a polyvinylidene difluoride (PVDF) membrane (Millipore, USA). Antibodies against TCEAL7 (1:500 dilution, #11218-1-AP, proteintech) GAPDH (1:10000 dilution, #10494-1-AP, proteintech) and an HRP-conjugated secondary antibody (1:50000 dilution, #111-035-003, Jackson ImmunoResearch) were used for western blotting. Chemiluminescence western blotting detection system (proteintech) was used for protein detection.

Immunohistochemistry (IHC)

Clinical samples were fixed in 4% paraformaldehyde and embedded in paraffin, and were cut into 4- μ m-thick sections, followed by incubated with an anti-TCEAL7 (1:20, #orb357063, Biorbyt) antibody at 4 °C overnight. Finally, images were acquired for analysis.

Statistical Analysis

The statistical analysis was carried out using the R software (version 3.6.2). Wilcox test was used to assess the difference in mRNAsi and TCEAL7 between tumor samples and normal samples. Kruskal-Wallis test was applied to explore the correlation in mRNAsi and TCEAL7 versus clinical characteristics. Pearson regression was applied to illustrate the correlation between TCEAL7 and mRNAsi. Survival was analyzed using the Kaplan-Meier method and log-rank test. Statistical significance was set at $P < 0.05$.

Results

Data Processing for DEGs

The workflow of our study was briefly summarized in **Figure 1**. The RNA-seq data of 174 normal gastric tissue samples from GTEx project, 32 normal and 375 stomach adenocarcinoma (STAD) tissue samples from the TCGA database were included in the current study. We confirmed the necessity of normalization and batch effect removal before identification of DEGs (**Supplementary Fig. S1a-b**). And it was not

difficult to find out the TCGA and GTEx samples performed better after normalization and batch effect removal (**Supplementary Fig. S1c-d**). Then 5416 DEGs were identified between normal and tumor tissue samples (**Supplementary Fig. S1e**), of which 3744 of them were upregulated and 1672 were downregulated, using the criteria of the absolute value of the log2-transformed fold change (FC) > 1 and the adjusted P-value < 0.05 (**Fig. 2a**).

The Expression Level and Survival Analysis of mRNAsi

The mRNAsi of normal tissues and tumors were respectively illustrated in **Figure 2b**, with the expression level of tumors obviously higher than that of normal tissues. According to the calculated mRNAsi, all the GC samples were further assigned to either a high mRNAsi score group or a low mRNAsi score group by a cutoff value of 0.42. As presented in **Figure 2c**, the OS showed notable difference between the two groups ($P < 0.001$) by applying Kaplan-Meier survival analysis. Furthermore, the expression level of mRNAsi was compared among T stages, lymphatic metastasis, M stages, and AJCC stages. However, as shown in **Figures 2d-g**, there was no statistically significant difference among these subgroups.

WGCNA of DEGs

After all the data process was completed, a correlation analysis of 5416 DEGs was conducted. To assure a scale-free topology model, the soft threshold power of β was set at 5 (**Supplementary Fig. S2b**). Next, eleven various modules (magenta, red, turquoise, brown, yellow, green, blue, purple, black, pink, and gray) were validated by a clustering analysis (module size>30 and cut height>0.2), with each color module presenting the same gene expression pattern (**Fig. 3a**). And the correlation relationships between every two modules were displayed in the figures (**Supplementary Fig. S2a, c**). Under the condition of fold enrichment > 1 and P-value < 0.05, a total of eight modules among all the eleven modules were validated significantly associated with mRNAsi. The green, blue, purple, black and pink modules were correlated negatively with mRNAsi ($ME_{\text{green}}: r = -0.19, P = 2E^{-4}$, $ME_{\text{blue}}: r = -0.79, P = 6E^{-79}$, $ME_{\text{purple}}: r = 0.52, P = 1E^{-26}$, $ME_{\text{black}}: r = -0.39, P = 2E^{-14}$, $ME_{\text{pink}}: r = -0.64, P = 4E^{-43}$). The turquoise, brown and yellow modules were correlated positively to mRNAsi ($ME_{\text{turquoise}}: r = 0.15, P = 0.004$, $ME_{\text{brown}}: r = 0.7, P = 2E^{-55}$, $ME_{\text{yellow}}: r = 0.38, P = 6E^{-14}$) (**Fig. 3b**). According to the module-trait relationships, the blue module was identified as the most significantly relevant module to mRNAsi, with the highest correlation value ($r = -0.79, P = 6E^{-79}$). The blue module was therefore selected for subsequent analyses to explore key genes. And based on threshold limits (cor. gene GS > 0.7 and cor. Gene MM > 0.8), 55 out of 504 hub genes were identified after selection in the blue module (**Fig. 3c**).

Function Annotation of the Key Genes in the Blue Module

To elaborate the definite functions of key genes in blue module, the KEGG signaling pathway annotation analysis was applied. As a result, the key genes in blue module were confirmed closely related to 18 pathways which were associated with "Signal transduction", "Endocrine system", "Circulatory system", "Digestive system", "Cardiovascular disease", "Cell growth and death" and "Excretory system"

(**Supplementary Fig. S3a**). Then 9 biological processes (BP) terms, 9 cellular components (CC) terms and 6 molecular functions (MF) terms were enriched as a total 24 GO processes based on the Gene ontology (GO) enrichment analysis, with GO terms consisting of “regulation of cardiac conduction,” “potassium ion transport,” “Z disc” and “n ATPase binding” ranked at the top of the list (**Supplementary Fig. S3b**). These results could preliminarily explain the negative correlation between mRNAsi and the key genes in blue module.

LASSO Cox Regression Analysis

By applying univariable Cox regression analysis, 23 out of 55 genes were identified as the prognostic genes highly associated with survival, which were subjected to LASSO Cox regression. Then, TCEAL7 was selected to be the key gene associated with mRNAsi and survival according to the result of LASSO Cox regression (**Supplementary Fig. S4a, b**).

The Expression Level and Survival Analysis of TCEAL7

As shown in the **Fig. 4a**, the expression level of TCEAL7 was obviously lower in tumors than in normal tissues ($P < 0.001$). Moreover, TCEAL7 was validated at the lowest expression level in T1 among all T stages of GC ($P < 0.001$, P-value for T1 vs T2, T1 vs T3, and T1 vs T4 were all < 0.001) and so was in AJCC classification ($P = 0.003$, P-value for stage I vs stage II, stage I vs stage III, and stage I vs stage IV were all < 0.05) (**Fig. 4d, g**). But in lymphatic metastasis ($P = 0.203$) and M stage groups ($P = 0.301$), the expression level showed no statistically significant difference (**Fig. 4e, f**). Simultaneously, Pearson regression was applied to illustrated TCEAL7 had a negative correlation with mRNAsi (Pearson $r = -0.8404$, $P < 0.001$) (**Fig. 4c**). Kaplan-Meier survival analysis found a significant difference in OS between the high TCEAL7 level group and the low TCEAL7 level group (**Fig. 4b**, cutoff value = 1.82, $P = 0.003$).

To further investigate this gene, we also analyzed the expression level of TCEAL7 in other tumors from TCGA database. It was shown that the expression level of TCEAL7 was lower in tumors than in normal tissues in most types of tumors. (**Supplementary Fig. S5a**), which was consistent with the current results. However, univariable Cox regression analysis showed that TCEAL7 was considered as a risk factor in BLCA (HR = 1.17, $P = 0.0150$), KIRP (Kidney renal papillary cell carcinoma) (HR = 1.29, $P = 0.0170$), STAD (HR = 1.26, $P = 0.0023$), KICH (Kidney chromophobe) (HR = 2.93, $P = 0.0320$), and ACC (Adrenocortical carcinoma) (HR = 2.14, $P = 0.0036$) (**Supplementary Fig. S5b**), which confirmed the necessity of our current study.

All these results confirmed the prognostic value of TCEAL7. Nevertheless, it was truly worth thinking that the expression level of TCEAL7 was lower in tumors while high expression level of TCEAL7 was associated with the worse survival outcome in GC.

Construction and Evaluation of TCEAL7 Expression-Based Model

As shown in the univariable Cox regression analysis of 6 clinicopathologic features, age (HR = 1.02, $P = 0.011$), gender (HR = 1.58, $P = 0.027$), T stages (HR = 1.66, $P = 0.030$), lymphatic metastasis (HR = 1.60,

P = 0.032), and TCEAL7 (HR = 1.25, P = 0.001) were significantly associated with OS (**Fig. 5a**). Furthermore, the multivariable Cox regression analysis validated that age (HR = 1.03, P = 0.001), gender (HR = 1.58, P = 0.026), T stages (HR = 1.66, P = 0.035), M stage (HR = 2.43, P = 0.010), TCEAL7 (HR = 1.54, P = 0.001) were statistically significant (**Fig. 5b**). Based on the result of multivariable Cox regression analysis, a nomogram for predicting GC 1-, 3-, and 5-year survival was established (**Fig. 5c**). The C-index for nomogram was 0.677 (95% CI: 0.627-0.727) indicating that the model has a good discrimination ability. And the AUC of the model were 0.704 for 1-year OS, 0.689 for 3-year OS, and 0.755 for 5-year OS, which also proved this conclusion (**Fig. 5d**, Cox model 2). Additionally, the calibration curves of 3- and 5-year OS rates showed the model fits well (**Fig. 5e, f**).

Next, other two Cox models were constructed to further assess the accuracy and discrimination of Cox model 2, with age, gender, TNM stages enrolled in Cox model 1 and age, gender, TCEAL7 enrolled in Cox model 3. As shown in **Figure 5d**, the AUC for 1- to 6-year OS of Cox model 2 is higher than Cox 1 and Cox model 3. The C-index was 0.648 (95% CI: 0.593-0.704) for model 1 and 0.636 (95% CI: 0.585-0.687) for model 3. Moreover, when took model 1 as the reference, the IDI, continuous NRI, and MIRS for 3-year OS were significant higher in model 2, with 0.045 (P < 0.001) for IDI, 0.243 (P < 0.001) for continuous NRI, and 0.146 (P = 0.013) for MIRS respectively. Meanwhile, similar results were found in the continuous NRI, and MIRS for the 1-year OS, with 0.173 (P < 0.027) for continuous NRI and 0.068 (P < 0.04) for MIRS. Corresponding results were not found in 5-year OS between model 2 and model 1. In addition, there were no significant IDI, continuous NRI, and MIRS in model 3 compared to model 1 (**Table1**). These results showed that TCEAL7 could improve the accuracy and discrimination of model 2 compared with model 1.

Validation of the Prognostic Value of TCEAL7

To further verify the expression level and the prognostic value of TCEAL-7, the microarray GSE66229 dataset was obtained, which had 100 normal and 300 tumor samples. Consistent with the results above, the expression level of TCEAL7 was significantly lower in tumors than in normal tissues (P < 0.001, **Fig. 6a**), and the OS of high TCEAL7 level group was obviously worse than that of low TCEAL7 level group (**Fig. 6b**, cutoff value = 1.54, P < 0.001). It seemed that lower TCEAL7 expression level was associated with lower TNM stages and AJCC stages (**Supplementary Fig. S6a-d**). Correspondingly, the AUC of the three models were shown in **Figure 6e** to explain the accuracy and discrimination of models. And the C-index was 0.646 (95% CI: 0.6-0.692) for model 1, 0.663 (95% CI: 0.619-0.708) for model 2 and 0.596 (95% CI: 0.55-0.643) for model 3. These results verified the prognostic value of TCEAL7. Moreover, the calibration plots of model 2 were also established (**Fig. 6d, e**).

Validation of the TCEAL7 expression in GC tissues and cells

The expression of TCEAL7 was verified in GES-1 cells and in another three GC cell lines, including SGC-7901, HGC27 and AGS. The gastric cancer cell lines had a relatively low expression level of TCEAL7 compared with the GES-1 cells, both at the mRNA and protein levels (**Fig. 7a-b**). To further confirmed the low expression of TCEAL7 in GC tissues, IHC staining was employed and the results obtained were consistent (**Fig. 7c-d**).

Discussion

Since the pan-cancer cohorts study of Malta et al. provided a new way for describing stemness features of cancer, many studies have focused on stemness features-associated key genes and possible signal pathways in cancers based on the OCLR-based stemness index[7–11]. These studies verified that tumor samples had a higher stemness index when compared with normal samples in BLCA, BRCA, LUAD, LGG, and STAD. However, the high level of mRNAsi was not always related to a poor prognosis. In our study, we analyzed the mRNAsi expression in different clinical features and its relationship with survival in GC patients. It showed that the high level of mRNAsi might represent a good prognosis. In the pan-cancer cohorts study[6], a trend toward higher stemness index with better outcome was also found in COAD (Colon adenocarcinoma) and LGG. Zheng et al. used a relative expression orderings (REOs)-based way to describe the stemness features of cancers, and in this study, there was no significant relationship was found between REO-based stemness index and prognosis of STAD with Hazard Ratio (HR) value < 1 but $P\text{-value} > 0.05$ [24]. Possible explanations for these results might be that whether OCLR-based or REO-based stemness index can not accurately describe the stemness features in these cancers or the relationship between cancer stemness features and progression or prognosis of cancers is more complicated than we know, which may be associated with intratumoral heterogeneity. Further studies on stemness features of these cancers will need to be undertaken. In short, we preliminarily demonstrated the prognostic value of OCLR-based stemness index despite the negative correlation between mRNAsi level and prognosis.

WGCNA is a tool to analyze the gene expression pattern in multiple samples, and it can classify those genes with similar expression patterns into clusters and further analyse the correlations between different gene clusters and certain characteristics. By applying WGCNA, we identified more than one gene module related to mRNAsi. Different from the study of Chen et al. who selected the highest positive correlation module with mRNAsi[11], the blue module which had the highest correlation value ($r = -0.79$, $P = 6E^{-79}$) with mRNAsi was selected and 55 out of 504 hub genes in this module were further analyzed according to the threshold limits. The KEGG signaling pathway annotation analysis and GO enrichment analysis showed that these gene were more associated with some organismal system-specific functions such as secretion of some digestive juices, which preliminarily explained the negative correlation between the key genes in blue modul and stemness features.

By virtue of Cox regression analysis and LASSO regression, TCEAL7 was finally identified as the key gene correlated with mRNAsi and prognosis in GC. Consistent with previous study[19, 20], the expression level of TCEAL7 was significantly lower in tumors than in normal tissues in two GC cohorts including samples from TCGA database and GSE66229 dataset. However, contrary to the results of miRNAsi mentioned above, OS of high TCEAL7 exprssion level group was worse than low TCEAL7 expression level group in both two GC cohorts. Moreover, a trend that lower TCEAL7 expression level was associated with lower TNM stage and AJCC stage was discovered in the cohorts of GSE66229 dataset. These findings are unexpected and are contrary to the study of Huang et al. who found low expression of TCEAL7 was associated with worse prognosis of gastric adenocarcinoma[20]. The inconsistency may be explained by differences in races and areas between the cohorts in the two studies. Several reports have shown that

TCEAL7 functions as a tumor suppressor gene in glioblastoma, ovarian and gastric cancer[13, 16, 19]. As a cell death-regulatory protein, TCEAL7 has been reported to be associated with a reduced risk of invasive serous ovarian cancers[12]. Further studies proved that TCEAL7 negatively modulated Myc activity by binding to E-boxes of gene promoters in ovarian cancer[15]. Moreover, TCEAL7 could negatively regulate nuclear factor κ B (NF- κ B) signaling by modulating the transcriptional activity of NF- κ B on its target gene promoters in ovarian cancer cells[13]. In glioblastoma, exo-miR-301a was manifested to activate wnt/ β -catenin signaling and promote radiation resistance by targeting TCEAL7[16]. What's more, Heart and Neural Crest Derivatives Expressed 2 antisense RNA 1 (HAND2-AS1) was found to regulate the migration, invasion, and apoptosis of GC cells by suppressing TCEAL7 expression via targeting miR-769–5p[19]. In this study, we confirmed that the expression level of TCEAL7 was lower in tumors than in normal tissues in most types of tumors. Nevertheless, univariable Cox regression analysis showed that TCEAL7 was considered as a risk factor in BLCA, KIRP, STAD, KICH, and ACC. This result may be explained by the fact that TCEAL7 might play different roles in different tumors and the prognosis of tumors is affected by various factors including TCEAL7. Therefore, further research should be undertaken to investigate the function of TCEAL7 and its correlation with stemness features in GC.

To explore the prognostic value of TCEAL7, three models were constructed to predict the prognosis of GC, with age, gender, TNM stages enrolled in Cox model 1, age, gender, TNM stages, TCEAL7 enrolled in Cox model 2 and age, gender, TCEAL7 expression level enrolled in Cox model 3. Cox model 2 showed better accuracy and discrimination than Cox model 1 and Cox model 3 in both two GC cohorts. Moreover, the accuracy and discrimination of this one gene-based model are comparable with a nine-mRNAsi-related-gene risk model of a previous study[11]. These results verified the prognostic value of TCEAL7, and the TCEAL7 expression based model could be expected to serve as an efficient tool to predict the prognosis of GC patients for clinicians.

There are several limitations in the current study. First of all, as mentioned above, the OCLR-based stemness index was used to describe the stemness features of tumors, which might not accurately describe the stemness features in GC. Next, though this study was based on two cohorts of GC from TCGA database and GSE66229 dataset, validation in cellular experiments, and animal and tissue models are needed. Finally, considering inadequate treatment information in both TCGA database and GSE66229 dataset, the role of TCEAL7 in chemoresistance of GC patients needs to be further explored.

Conclusion

In conclusion, we verified mRNAsi was associated with the prognosis of GC patients. And TCEAL7 was finally identified as the key gene correlated with stemness features and prognosis in GC, which was demonstrated as a risk factor. Nomogram including TCEAL7 and clinical features showed good accuracy and discrimination for predicting the prognosis of GC patients. Further study should be conducted to confirm our conclusion and investigate the role of TCEAL7 in GC and its correlation with stemness features.

Abbreviations

TCEAL7

The transcription elongation factor A-like 7; GC:Gastric cancer; OS:Overall survival; OCLR:One-class logistic regression machine learning algorithm; mRNAsi:mRNA expression-based stemness index; mDNAsi:DNA methylation-based stemness index; EREG-mRNAsi:mRNAsi and the epigenetic regulation based-index; BLCA:Bladder cancer; BRCA:Breast cancer; LUAD:lung adenocarcinoma; LGG:lower-grade glioma; WGCNA:Weighted Gene Co-expression Network Analysis; LASSO:The least absolute shrinkage and selection operator; AUC:Area Under Curve; TCGA:The Cancer Genome Atlas; GEO:Gene Expression Omnibus; GTEx:Genotype-Tissue Expression; DEGs:differentially expressed genes; STAD:stomach adenocarcinoma; FDR:false discovery rate; GO:Gene ontology; KEGG:Kyoto Encyclopedia of Genes and Genomes; NRI:net reclassification improvement; IDI:integrated discrimination improvement; MIRS:median improvement in risk score;

Declarations

Ethics approval and consent to participate

The study were approved by First Affiliated Hospital of Nanjing Medical University Ethics Committee. Written consent was obtained from all participants.

Consent for publication

Not applicable.

Availability of data and materials

The data that support the findings of this study are available in TCGA database (<https://portal.gdc.cancer.gov/>) and GEO database (<https://www.ncbi.nlm.nih.gov/geo/>). Other datasets are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

The present study was supported by the National Natural Science Foundation of China (No. 81770561), National Science and Technology Major Project (AD18), and Jiangsu medical leading talent and innovation team (CXTDA2017033).

Authors' Contributions

G.Z. and X.L.: designing the project; H.Q., X.G., R.M. and W. L.: data analysis and writing the article; Y.Z.: revising the syntax.

Acknowledgements

The authors would like to acknowledge all the members in Dr Zhang's laboratory for help with this study.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, 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(6):394–424.
2. Smyth EC, Nilsson M, Grabsch HI, van Grieken NC, Lordick F. Gastric cancer. *Lancet*. 2020;396(10251):635–48.
3. Friedmann-Morvinski D, Verma IM. Dedifferentiation and reprogramming: origins of cancer stem cells. *EMBO Rep*. 2014;15(3):244–53.
4. Shibue T, Weinberg RA. EMT, CSCs, and drug resistance: the mechanistic link and clinical implications. *Nat Rev Clin Oncol*. 2017;14(10):611–29.
5. Ge Y, Gomez NC, Adam RC, Nikolova M, Yang H, Verma A, Lu CP, Polak L, Yuan S, Elemento O, et al. Stem Cell Lineage Infidelity Drives Wound Repair and Cancer. *Cell*. 2017;169(4):636–50 e614.
6. Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, Kaminska B, Huelsken J, Omberg L, Gevaert O, et al. Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell*. 2018;173(2):338–54 e315.
7. Pan S, Zhan Y, Chen X, Wu B, Liu B. Identification of Biomarkers for Controlling Cancer Stem Cell Characteristics in Bladder Cancer by Network Analysis of Transcriptome Data Stemness Indices. *Front Oncol*. 2019;9:613.
8. Zhao M, Chen Z, Zheng Y, Liang J, Hu Z, Bian Y, Jiang T, Li M, Zhan C, Feng M, et al. Identification of cancer stem cell-related biomarkers in lung adenocarcinoma by stemness index and weighted correlation network analysis. *J Cancer Res Clin Oncol*. 2020;146(6):1463–72.
9. Pei J, Wang Y, Li Y. Identification of key genes controlling breast cancer stem cell characteristics via stemness indices analysis. *J Transl Med*. 2020;18(1):74.
10. Zhang M, Wang X, Chen X, Guo F, Hong J. Prognostic Value of a Stemness Index-Associated Signature in Primary Lower-Grade Glioma. *Front Genet*. 2020;11:441.
11. Chen X, Zhang D, Jiang F, Shen Y, Li X, Hu X, Wei P, Shen X. Prognostic Prediction Using a Stemness Index-Related Signature in a Cohort of Gastric Cancer. *Front Mol Biosci*. 2020;7:570702.
12. Chien J, Staub J, Avula R, Zhang H, Liu W, Hartmann LC, Kaufmann SH, Smith DI, Shridhar V. Epigenetic silencing of TCEAL7 (Bex4) in ovarian cancer. *Oncogene*. 2005;24(32):5089–100.
13. Rattan R, Narita K, Chien J, Maguire JL, Shridhar R, Giri S, Shridhar V. TCEAL7, a putative tumor suppressor gene, negatively regulates NF-kappaB pathway. *Oncogene*. 2010;29(9):1362–73.

14. Peedicayil A, Vierkant RA, Shridhar V, Schildkraut JM, Armasu S, Hartmann LC, Fridley BL, Cunningham JM, Phelan CM, Sellers TA, et al. Polymorphisms in TCEAL7 and risk of epithelial ovarian cancer. *Gynecol Oncol.* 2009;114(2):260–4.
15. Chien J, Narita K, Rattan R, Giri S, Shridhar R, Staub J, Belefond D, Lai J, Roberts LR, Molina J, et al. A role for candidate tumor-suppressor gene TCEAL7 in the regulation of c-Myc activity, cyclin D1 levels and cellular transformation. *Oncogene.* 2008;27(58):7223–34.
16. Yue X, Lan F, Xia T. Hypoxic Glioma Cell-Secreted Exosomal miR-301a Activates Wnt/beta-catenin Signaling and Promotes Radiation Resistance by Targeting TCEAL7. *Mol Ther.* 2019;27(11):1939–49.
17. Orhan C, Bulut P, Dalay N, Ersen E, Buyru N. Downregulation of TCEAL7 expression induces CCND1 expression in non-small cell lung cancer. *Mol Biol Rep.* 2019;46(5):5251–6.
18. Guo Y, Liao Y, Jia C, Ren J, Wang J, Li T. MicroRNA-182 promotes tumor cell growth by targeting transcription elongation factor A-like 7 in endometrial carcinoma. *Cell Physiol Biochem.* 2013;32(3):581–90.
19. Yu L, Luan W, Feng Z, Jia J, Wu Z, Wang M, Li F, Li Z. Long non-coding RNA HAND2-AS1 inhibits gastric cancer progression by suppressing TCEAL7 expression via targeting miR-769-5p. *Dig Liver Dis.* 2021;53(2):238–44.
20. Huang CY, Chen YM, Zhao JJ, Chen YB, Jiang SS, Yan SM, Zhao BW, Pan K, Wang DD, Lv L, et al. Decreased expression of transcription elongation factor A-like 7 is associated with gastric adenocarcinoma prognosis. *PLoS One.* 2013;8(1):e54671.
21. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559.
22. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16(5):284–7.
23. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med.* 2008;27(2):157–72. discussion 207 – 112.
24. Zheng H, Song K, Fu Y, You T, Yang J, Guo W, Wang K, Jin L, Gu Y, Qi L, et al: **An absolute human stemness index associated with oncogenic dedifferentiation.** *Brief Bioinform* 2020.

Tables

Table 1 Comparison of the accuracy and discrimination in three models

Index	Cox model 2 vs Cox model 1			Cox model 3 vs Cox model 1		
	Estimated value	95%CI	P value	Estimated value	95%CI	P value
IDI (1 year)	0.018	-0.001-0.057	0.073	-0.024	-0.096-0.016	0.239
Continuous NRI (1 year)	0.173	0.022-0.322	0.027	0.015	-0.256-0.183	1.103
RSMI (1 year)	0.068	0.000-0.132	0.04	0.004	-0.085-0.076	1.07
IDI (3 year)	0.045	0.007-0.108	<0.001	0.025	-0.057-0.088	0.558
Continuous NRI (3 year)	0.243	0.095-0.384	<0.001	0.156	-0.147-0.331	0.352
RSMI (3 year)	0.146	0.003-0.234	0.013	0.064	-0.042-0.195	0.419
IDI (5 year)	0.032	-0.023-0.192	0.239	-0.047	-0.197-0.118	0.452
Continuous NRI (5 year)	0.186	-0.268-0.769	0.259	-0.106	-0.589-0.531	0.591
RSMI (5 year)	0.125	-0.053-0.227	0.286	-0.018	-0.266-0.172	0.631

Cox model 1: age, gender, and TNM stages were enrolled; Cox model 2: age, gender, TNM stages, and TCEAL7 expression level were enrolled; Cox model 3: age, gender, and TCEAL7 expression level were enrolled. NRI, net reclassification improvement; IDI, integrated discrimination improvement. MIRS, median improvement in risk score.

Figures

Figure 1

The complete process of stemness-related prognostic genes identification and verification.

Figure 2

Screening of DEGs, expression level and survival analysis of mRNAsi. (A) The volcano plot showed that 3744 DEGs were upregulated and 1672 were downregulated. (B) The difference in mRNAsi of normal tissues and tumors respectively. (C) The Kaplan-Meier survival analysis showed a notable difference in OS between high and low mRNAsi level groups. (D-G) Difference of mRNAsi level in TNM stages and AJCC stages. DEGs, differentially expressed genes; OS, overall survival; mRNAsi, mRNA expression-based stemness index.

Figure 3

Weighted gene co-expression network analysis (WGCNA) of DEGs. (A) The clustering analysis validated eleven various modules, with each color module presenting the same gene expression pattern. (B) The green, blue, purple, black, and pink modules were correlated negatively with mRNA_{si}, and turquoise, brown and yellow modules were positively correlated with mRNA_{si}. (C) Scatter plot analysis of blue modules. DEGs, differentially expressed genes; mRNA_{si}, mRNA expression-based stemness index. mRNA_{si}, mRNA expression-based stemness index.

Figure 4

The expression level and survival analysis of TCEAL7 based on TCGA database. (A) The plot illustrated the expression level of TCEAL7 was obviously lower in tumors than in normal tissues. (B) The Kaplan-Meier survival analysis proved a significant difference in OS between high TCEAL7 level group and low TCEAL7 level group. (C) Pearson correlation analysis illustrated TCEAL7 had a negative correlation with mRNAsi. (D-G) The difference of TCEAL7 expression level in TNM stages and AJCC stages. TCGA, The Cancer Genome Atlas; OS, overall survival.

Figure 5

Construction and evaluation of TCEAL-7 expression-based model. Univariable (A) and multivariable (B) Cox regression analysis of 6 clinicopathologic features. (C) A nomogram for predicting 1-, 3- and 5-year OS of GC. (D) AUC of three models for 1- to 6-year OS. Calibration plot of the nomogram for predicting probabilities of 3- (E) and 5-year (F) OS of GC. OS, overall survival; GC, gastric cancer; AUC, Area Under Curve.

Figure 6

The validation of the expression and prognostic value of TCEAL7 based on GSE66229 dataset. (A) The difference of TCEAL7 expression level in tumors and normal tissues. (B) Kaplan-Meier survival analysis for high TCEAL7 level group and low TCEAL7 level group. (C) AUC of three models for 1- to 6-years OS. Calibration plot of the nomogram for predicting probabilities of 3- (E) and 5-year (F) OS of GC. OS, overall survival; GC, gastric cancer; AUC, Area Under Curve.

Figure 7

The expression of TCEAL7 was verified in GES-1 cells and in another three GC cell lines, including SGC-7901, HGC27 and AGS. The gastric cancer cell lines had a relatively low expression level of TCEAL7 compared with the GES-1 cells, both at the mRNA and protein levels (Fig. 7a-b). To further confirmed the low expression of TCEAL7 in GC tissues, IHC staining was employed and the results obtained were consistent

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFigure1.pdf](#)
- [SupplementaryFigure2.pdf](#)
- [SupplementaryFigure3.pdf](#)

- [SupplementaryFigure4.pdf](#)
- [SupplementaryFigure5.pdf](#)
- [SupplementaryFigure6.pdf](#)