Role of COL3A1 and POSTN on Pathologic Stages of Esophageal Cancer

Shao-wei Zhang, MD[1], Nan Zhang, MD[1], and Na Wang, MD[2] (ID)

Abstract

Background: Esophageal cancer (EC) is a primary malignant tumor originating from the esophageal of the epithelium. Surgical resection is a potential treatment for EC, but this is only appropriate for patients who have locally resectable lesions suitable for surgery. However, most patients with EC are at a late stage when diagnosed. Therefore, there is an urgent need to further explore the pathogenesis of EC to enable early diagnosis and treatment. **Methods:** Our study downloaded 2 expression spectrum datasets (GSE92396 and GSE100942) in the Gene Expression Omnibus (GEO) database. GEO2 R was used to identify the Differentially expressed genes (DEGs) between the samples of EC and control. Using the DAVID tool to make the Functional enrichment analysis. Constructing A protein–protein interaction (PPI) network. Identifying the Hub genes. The impact of hub gene expression on overall survival and their expression based on immunohistochemistry were analyzed. Associated microRNAs were also predicted. **Results:** There were 36 common DEGs identified. The analysis of GO and KEGG results shown that the variations were predominantly concentrated in the extracellular matrix (ECM), ECM organization, DNA binding, platelet activation, and ECM-receptor interactions. COL3A1 and POSTN had high expression in EC tissues which was compared with their expression in healthy tissues. Analysis of pathologic stages showed that when COL3A1 and POSTN were highly expressed, the stage of the pathologic of EC patients was relatively high (P < 0.005). **Conclusions:** COL3A1 and POSTN may play an important role in the advancement and occurrence of EC. These genes could provide some novel ideas and basis for the diagnosis and targeted treatment of EC.

Keywords

esophageal, cancer, bioinformatics, differentially expressed genes, hub gene

Received: April 09, 2020; Revised: October 31, 2020; Accepted: November 04, 2020.

Introduction

Esophageal cancer (EC) ranks seventh among malignant tumors in incidence and mortality. Its epidemiology has the following characteristics: it is regionally distributed, there is a higher incidence among men than women (men cases account for about 70%), and middle-aged and elderly people are most susceptible.[1] The incidence of EC among men in East Asia is the highest in the world, with incidence in Mongolia and China ranking among the top 5 countries around the world.[2] The occurrence of EC is closely associated with living conditions, dietary habits, carcinogens, and the genetic susceptibility of patients. Symptoms in patients with early-stage EC are often atypical and can easily be overlooked.[3] Patients with advanced-stage EC may experience symptoms such as progressive dysphagia, gastroesophageal reflux, and other symptoms. Insufficient food intake in patients with EC can lead to significant chronic dehydration, malnutrition, wasting, and

cachexia.[4] By the time the EC patients were diagnosed, they are often reach the advanced stage. The 5-year survival rate can't account about 10%. Surgical resection rates of EC can reach 80%–90%, and EC can be cured by early resection. Even if the surgical resection rate were to be improved, the long-term outlook remains unsatisfactory. Therefore, it is of great clinical and market application value to explore the pathogenesis of EC

[1] Department of Thoracic Surgery, The Fourth Hospital of Hebei Medical University, Shijiazhuang, Hebei, People's Republic of China
[2] Digestive Department, The Fourth Hospital of Hebei Medical University, Shijiazhuang, Hebei, People's Republic of China

Corresponding Author:
Na Wang, Digestive Department, The Fourth Hospital of Hebei Medical University, 12 Jiankang Road, Shijiazhuang, Hebei 050000, People's Republic of China.
Email: wangnahb532@163.com

through finding the potential diagnostic and therapeutic targets, making early detection and diagnosis.

Bioinformatics, a new and interdisciplinary field, combines life sciences with computer science. It mainly involves in the collection, storage, processing, dissemination, analysis, and interpretation of biological information. The combination of biological and informatics technologies enables a good deal of complex data of biology to be processed for analysis. Microarray data analysis was used widely to explore genetic relevance[5,6] in studies of tumors and other diseases. Microarray data analysis can simultaneously capture huge quantities of the information of gene expression and then explore genomic changes related to the generation and advancement of diseases. Numerous studies have been performed and many scholars[7,8] have used bioinformatics to analyze differentially expressed genes (DEGs) in tumor progression. The research of DEGs has also provided a theoretical basis for early diagnosis and treatment.

In our study, using bioinformatics techniques to find gene sequencing data from patients with EC and a comparison set of data from healthy people stored in Gene Expression Omnibus (GEO). Two high-quality gene datasets were analyzed to mine DEGs associated with EC and corresponding microRNAs. Using Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) to analyze the DEGs. Then, constructing a protein–protein interaction (PPI) network and the important modules were screened. Clinical samples from patients with EC were used to verify the screened genes and microRNAs. These genes and microRNAs will improve our comprehension of the underlying mechanisms of EC. They also can provide a basis for developing targeted therapies.

Methods

Differential Expression Analysis of the Whole Chromosome

Differential expression analysis in the Gene Expression Profiling Interactive Analysis tool (GEPIA, http://gepia.cancer-pku.cn/)[9] enables users to apply customized methods and thresholds of statistics to a given dataset and to obtain DEGs dynamically and the distribution of chromosomal. Here, ANOVA was selected as the method for differential analysis. The statistical thresholds were |log2FC| >1, and q-value <0.01.

Access to Public Data

An open-source, high-throughput genomic database named GEO (http://www.ncbi.nlm.nih.gov/geo)[10] that contains microarrays, gene expression data, and chips. We obtained 2 expression spectrum datasets (GSE92396 and GSE100942) in the GEO database. The GSE92396 dataset comprises 12 esophageal cancer tissues and 10 healthy esophageal tissues. The GSE100942 dataset comprises 4 esophageal carcinoma tissues and 4 healthy esophageal tissues. In the basis of the information

of the annotation on the platform, converting the whole probe numbers to gene symbols.

Repeatability Tests for the Data of Intra-Group

Pearson's correlation test was made to prove the repeatability of data. R,[11] as an open-source language and environment, was used to made statistical calculations and mapping. Using R language to make Pearson's correlation test and heat-mapping.

Screening of DEGs via GEO2 R

GEO2 R (http://www.ncbi.nlm.nih.gov/geo/geo2r)[12] is an online analysis system of data in the GEO. The advantages of GEO2 R are that it is an online tool, which is easy and efficient to operate. GEO2 R can perform a command to compare gene expression profiles between groups. Identification of DEGs between EC and control groups was carried out by it. In general, when a probe set has a corresponding gene symbol, the probe is considered valuable and will be retained. Statistically significant measures are a P-value <0.01 and a fold-change (FC) >1. There has another open-source tool for drawing volcano maps named SangerBox (https://shengxin.ren). Draw the Venn diagrams with an online Venn diagram tool (http://bioin formatics.psb.ugent.be/webtools/Venn/). The tool could visualize common DEGs shared between GSE92396 and GSE-100942.

Protein–Protein Interaction (PPI) Network

There has a network named the Search Tool for the Retrieval of Interacting Genes (STRING; http://string-db.org) (version 10.5)[13] that can be used to predict and track PPIs. Introducing DEGs into the tool enables intermolecular network analysis. The analysis of interactions in the different proteins can provide a novel idea for us of the mechanisms of the generation or advancement of tumors, such as EC. In this research, using STRING to obtain a PPI network with DEGs. The minimum requirement for interaction score is that medium confidence > 0.4. Cytoscape (version 3.6.1) is open-source software for visualization that can be used to visualize PPI networks.[14] Therefore, using STRING to construct a PPI network, which was then visualized through Cytoscape (version 2.8).[15]

Functional Annotation of DEGs via GO and KEGG Analysis

There has one bioinformatics database which is named DAVID (https://david.ncifcrf.gov/home.jsp) (version 6.8) integrating biological data and analytical tools. It can help users obtain biological information through providing systematic and comprehensive annotation information of the functions biology for massive gene or protein lists.[16] The Kyoto Encyclopedia of Genes and Genomes (KEGG) (https://www.kegg.jp/) is used to understand systems of biology and advanced functions, integrating a large amount of utility database resources. Gene

ontology (GO) contains 3 aspects: cell composition, molecular function and biological process,[17] that is used in bioinformatics widely. Using DAVID online tool to make the analysis of the GO and information of pathway enrichment of DEGs. P-value <0.05 was considered statistically significant.

Metascape (http://metascape.org/gp/index.html#/main/step1)[18] is an excellent tool in the annotation of analysis for gene functions. It can give a hand for researchers to make the batch analysis for genes and proteins, which help us understand the functions of genes or proteins. In the current study, we finished the function and pathway enrichment analyses with Metascape.

Identification of Hub Genes

Based on topological principles, the Molecular Complex Detection (MCODE) (version 1.5.1) can mine regions that were coupled closely from PPI networks. First, Cytoscape software is used to plot the PPI network. Second, the most essential modules in the PPI network graph were identified by MCODE. The criteria of the MCODE analysis were that the node score cut-off = 0.2, degree cut-off = 2, maximum depth = 100, MCODE score >5, and k-score = 2.

Expression and Correlation Analysis

According to the GSE92396 and GSE100942 datasets, heatmaps was used to make the Clustering analysis of the expression levels of the hub genes. Using Pearson's correlation test in order to make the correlation analysis between the hub genes. Heat maps to show associations among the hub genes were plotted by R language.

Overall Survival and Pathologic Stage Analysis

Analyzing the influence of hub gene expression on the overall survival and pathologic stage through the GEPIA.

Identification of Chemicals Related to Hub Genes

The Comparative Toxicogenomics Database (CTD)[19] is a public database. It plays an important role in predicting how environmental factors may influence human health. This database help us identify chemicals which may influence these hub genes and potentially lead to new discoveries.

Expression of Hub Genes in the Basis of Immunohistochemistry

The Human Protein Atlas can be used to analyze the all proteins about their distribution and relative abundance in healthy human cells and tissues and the subcellular localization of each protein can be determined. The Human Protein Atlas was employed to validate the change of the patterns of gene expression at the level of the protein in healthy esophageal tissue.

Prediction of microRNAs Associated With Hub Genes

Predicting microRNAs relative to hub genes via Targetscan, which shows potential correlation among genes and microRNAs.

Multivariate Cox Regression Analysis Based on the TCGA Dataset

Hazard ratios (HR) of clinicopathologic factors for overall survival was calculated via Multivariate Cox regression analysis. Conducting the statistical analyses through SPSS software, version 21.0 (IBM Corp., Armonk, NY, USA). A p-value < 0.05 was considered to be statistically significant.

Survival Analysis of COL3A1 for EC Patients in the PrognoScan Database

The PrognoScan database (http://dna00.bio.kyutech.ac.jp/PrognoScan/index.html) is an online analysis tool for determining overall survival rates, which will give a clear of esophageal cancer. We performed a survival analysis about COL3A1 in EC patients with the PrognoScan database.

Pan-Cancer Analysis of COL3A1 and POSTN Based on ESCA Datasets

The GEPIA was used to perform pan-cancer analyses for COL3A1 and POSTN, based on ESCA datasets. A total of 33 types of cancer were selected. Using GEPIA to calculate overall survival according to the expression of COL3A1 and POSTN. The Cox proportional hazard ratio and 95% confidence interval information were contained in overall survival plot.

Results

### Differential Expression Analysis of Whole Chromosomes

There were many DEGs located on the whole chromosomes between EC and control samples (Figure 1A).

High Repeatability of Data

Strong correlations existed between the samples in EC group. At the same time, strong correlations existed between the samples in control group in the GSE92396 dataset, as confirmed by Pearson's correlation test (Figure 1B). Similarly, the reproducibility of data in the GSE100942 was good (Figure 1C).

Identification of DEGs in EC and the PPI Network

After the results of microarray were standardized, 1,540 DEGs in GSE92396 (Figure 2A) and 345 DEGs in GSE100942 (Figure 2B) were identified. In the 2 datasets, the Venn diagram showed 36 genes that were overlapped (Figure 2C). The PPI network is also presented to show correlations among the common DEGs (Figure 2D).
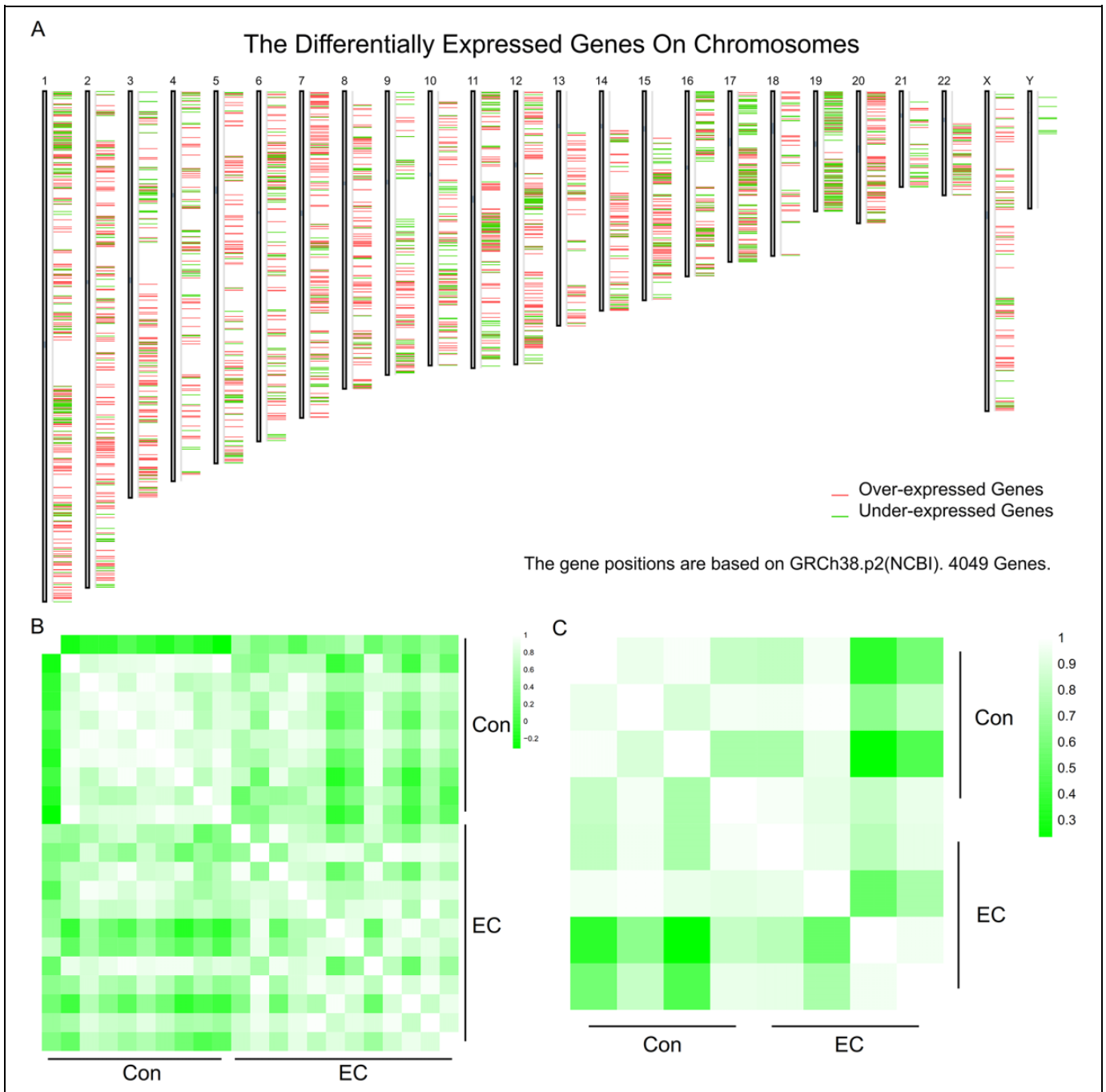
Figure 1. A, Differentially expressed genes (DEGs) located on chromosomes among esophageal cancer (EC) and control esophageal tissue. B, High repeatability of data in the GSE92396 dataset via Pearson's correlation test. C, High repeatability of data in the GSE100942 dataset via Pearson's correlation test.

KEGG and GO Enrichment Analyses of DEGs

Using DAVID to carry out functional and pathway enrichment analyses to make the analysis of biological classification of DEGs. Our study has some finding from the results through GO analysis that variation in biological process (BP) DEGs were prominently concentrated in ECM organization, cell–substrate adhesion regulation, intracellular signaling cascades, collagen biosynthetic processes, and cell cycle processes (Figure 3A). There have some variations in cell component (CC) of DEGs that were mainly concentrated in ECM, actin cytoskeleton, proteinaceous ECM, extracellular region parts, and fibrillar collagen (Figure 3B). In addition, there are some changes in molecular function (MF) of DEGs were mainly concentrated in single-stranded DNA binding, growth factor binding, platelet-derived growth factor binding, cytoskeletal protein binding, actin
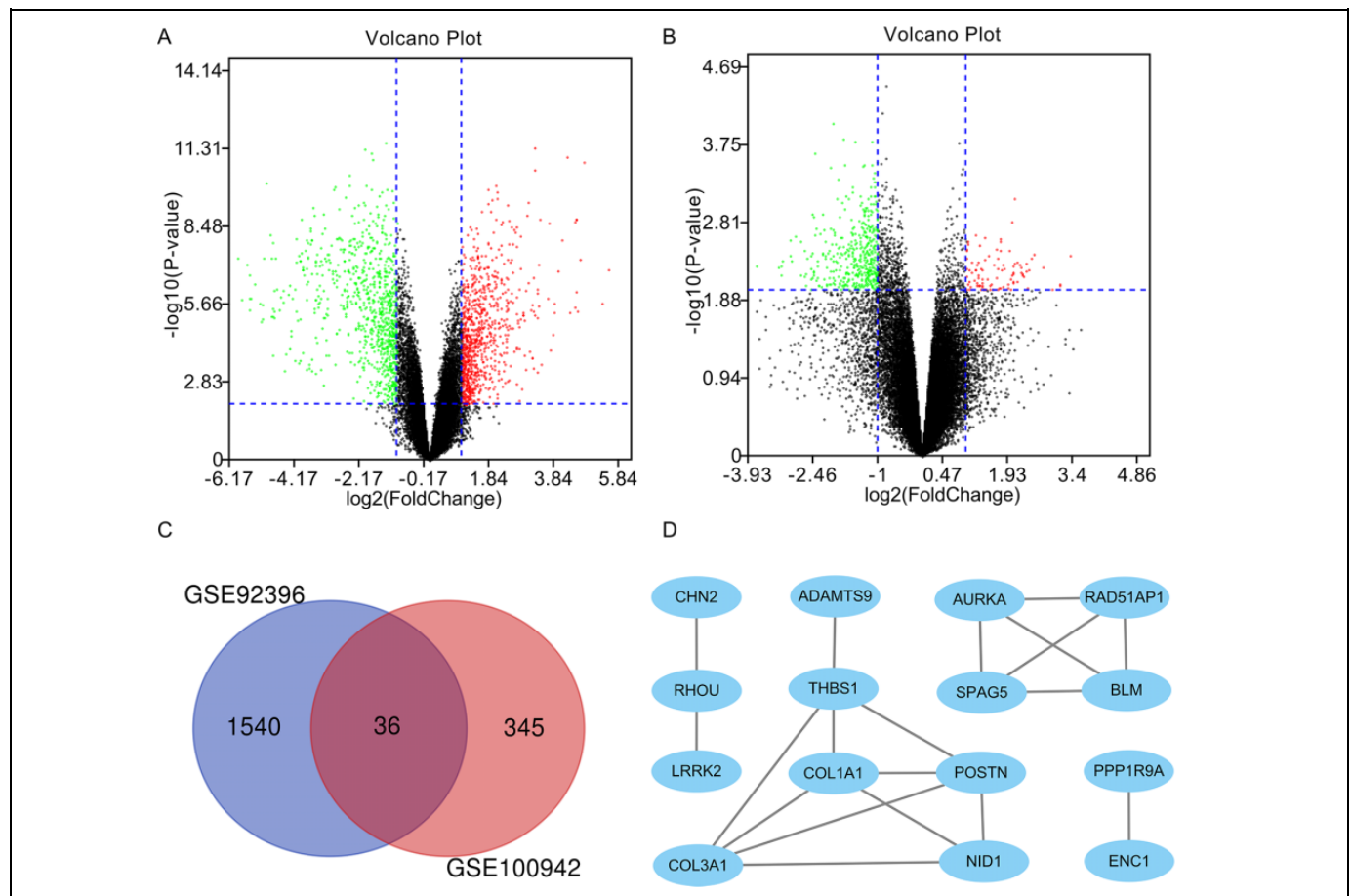
Figure 2. Identification of DEGs in esophageal cancer and the protein–protein interaction (PPI) network. A, Volcano plot showing the DEGs in GSE92396 dataset. B, Volcano plot showing the DEGs in GSE100942 dataset. C, The Venn diagram shows that 36 DEGs were simultaneously present. D, The protein–protein interaction (PPI) network of the common DEGs.

binding, and structure-specific DNA binding (Figure 3C). The whole DEGs were mainly concentrated in focal adhesion, platelet activation, and ECM–receptor interactions shown by the KEGG pathway analysis (Figure 3D).

Metascape Enrichment Analysis of DEGs

We found that the DEGs in EC and control tissues were mainly concentrated in cell–matrix adhesion, supramolecular fiber organization, cell–matrix adhesion regulation, positive regulation of cytokine secretion, organelle fission, developmental maturation, cellular responses to heat, regulation of neurotransmitter transport, SUMO E3 ligases SUMOylate target proteins, the Notch signaling pathway, regulation of organelle assembly, and renal system development after the functional enrichment analysis that made by Metascape (P < 0.05, Figure 3A, B, and C).

Selection and Analysis of Hub Gene

Finally, the hub genes contain 5 genes (THBS1, NID1, COL3A1, POSTN, and COL1A1). The names, abbreviations, and functions of these hub genes are shown in Table 1.

Expression Level and Correlation of the Hub Genes

Making a comparison between the samples in the EC group and control samples in the GSE92396 dataset, all hub genes were expressed higher in the former what was shown by one heat map (Figure 4B). The expression of COL1A1, THBS1, and NID1 was lower in EC samples than in the control samples in the GSE100942 dataset also shown by another heat map. However, COL3A1, and POSTN were expressed higher in samples of EC (Figure 4C).

Using Pearson's correlation test, the hub genes in the GSE92396 (Figure 4D) and GSE100942 (Figure 4E) datasets were correlated strongly shown by these heat maps.

Association Between Hub Gene Expression, Overall Survival, and Pathologic Stage

Our study carried out the overall survival analysis of hub genes via a Kaplan–Meier curve. EC Patients that were expressed lowly of THBS1 showed worse overall survival times (P < 0.05, Figure 5A). The times of overall survival of EC patients who were expressed highly levels of NID1 was poorer compared with those who had low NID1 expression levels
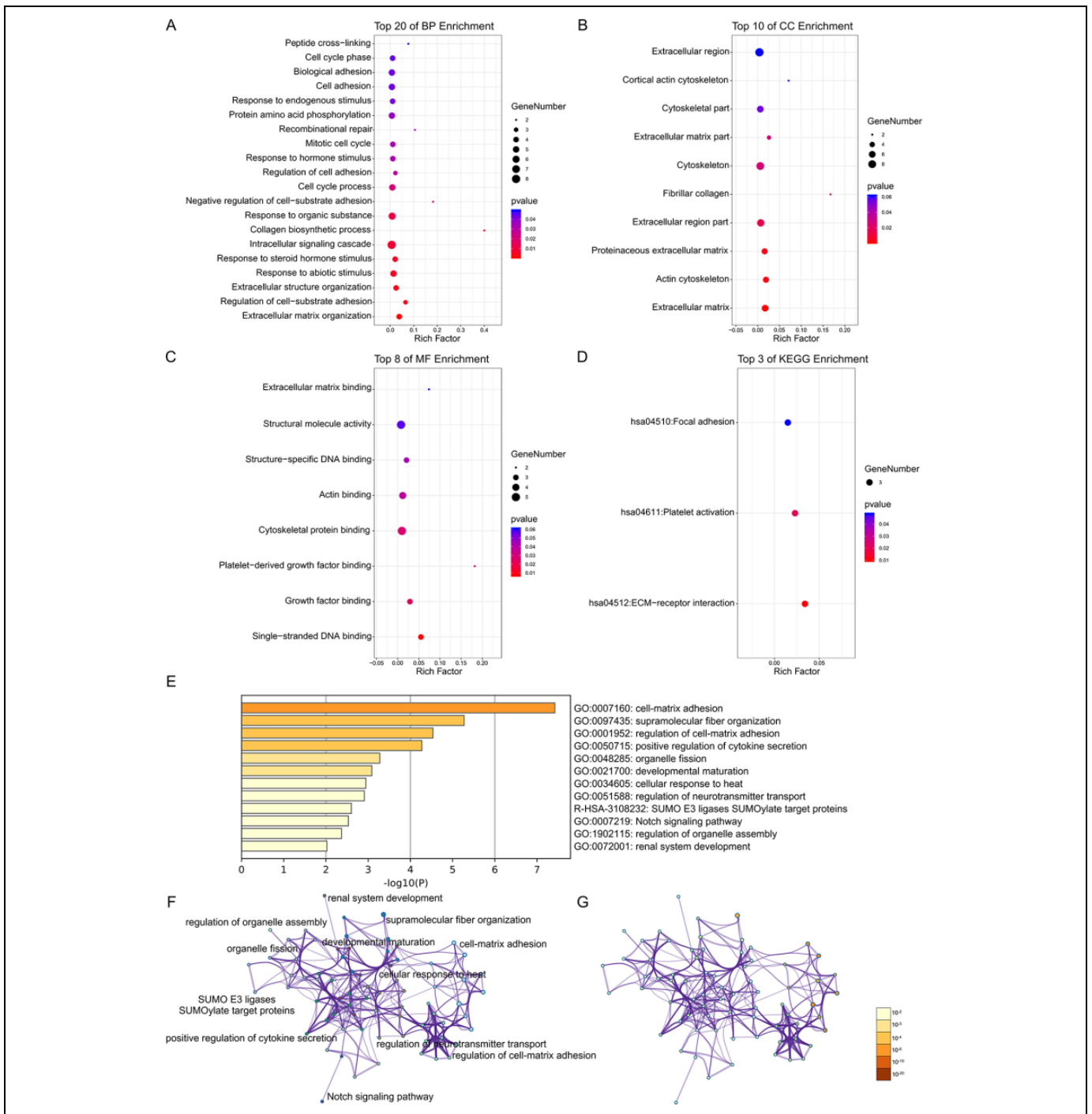
Figure 3. Common DEGs enrichment analysis. (A) Biological processes (BP) analysis, (B) cellular components (CC) analysis, (C) molecular function (MF) analysis, (D) KEGG analysis. (E) Heatmap of enriched terms across inputted lists of DEG, colored with p-values, through Metascape. (F) An enriched terms network colored with cluster identity. The nodes that have the uniform cluster identity meaning they are typically approximate to mutual. (G) An enriched terms network colored with p-values. More genes contained by the terms; more significant p-value will have.

($P < 0.05$, Figure 5B). Patients with EC who had low expression level of COL3A1 showed worse overall survival times ($P < 0.05$, Figure 5C). The overall survival times of EC patients who were expressed POSTN highly was poorer than those who had low expression levels ($P < 0.05$, Figure 5D). EC patients

who had low expression of COL1A1 showed worse overall survival times ($P < 0.05$, Figure 5E).

Subsequently, we found that the THBS1 (Figure 6A) and NID1 (Figure 6B) expression were not associated with the stage of the pathologic of EC, while the expression of COL3A1

Table 1. Summary of 5 Hub Genes Functions.

No.	Gene symbol	Full name	Function
1	THBS1	Thrombospondin 1	It could mediate the interactions between cell-to-cell and cell-to-matrix.
2	POSTN	Periostin	Play a role in cell attachment, and cell adhesion.
3	COL3A1	Collagen Type III Alpha 1 Chain	Regulate cortical development.
4	COL1A1	Collagen Type I Alpha 1 Chain	Type I collagen in the group I collagens.
5	NID1	Nidogen 1	Regulate cell-extracellular matrix interactions.
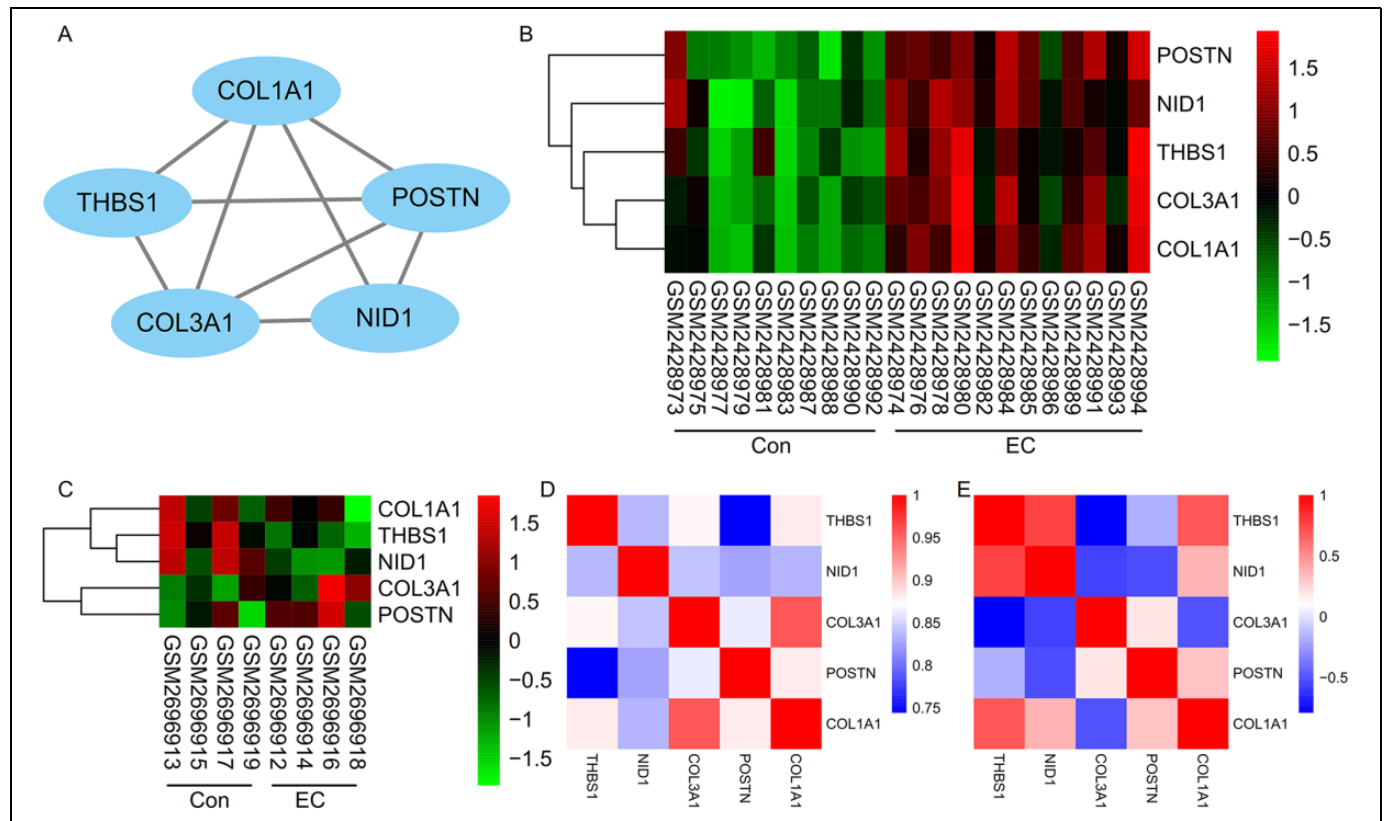
Figure 4. Hub gene network, analysis of expression and correlation. A, Five hub genes were identified (THBS1, NID1, COL3A1, POSTN, and COL1A1). B, A heat map showing that the whole hub genes were expressed higher in EC samples by comparing with the control samples in GSE92396 dataset. C, A heat map showing the whole hub genes expression in GSE100942 dataset. D, Associations in hub genes of the GSE92396 dataset showing by Heat maps. E, Associations in hub genes of the GSE100942 dataset showing by heat maps.

(Figure 6C), POSTN (Figure 6D), and COL1A1 (Figure 6E) were significantly positively related with pathologic stage in patients with EC ($P < 0.05$).

Identification of Chemicals Associated With Hub Genes

Esophageal neoplasms was targeted by the hub genes shown by the CTD database; the results are shown in Figure 7A-E.

Immunohistochemistry Validation of Genes Using the Human Protein Atlas Database

These tissue samples are from healthy esophageal tissues in the Human Protein Atlas. The expression level of THBS1, NID1,

COL3A1, POSTN, and COL1A1 were investigated via immunohistochemical staining (Figure 8A-E).

Prediction of microRNAs Related to Hub Genes

In order to figure out the mechanism and regulatory network of the hub genes, we hypothesized that the microRNAs related to hub genes for miRNAs. Table 2 shows the microRNAs connected with the hub genes.

Clinicopathologic Correlation Analysis Based on TCGA Datasets

Based on the multivariate Cox regression analysis, age had no effect on overall survival times (HR = 1.187, P = 0.453).
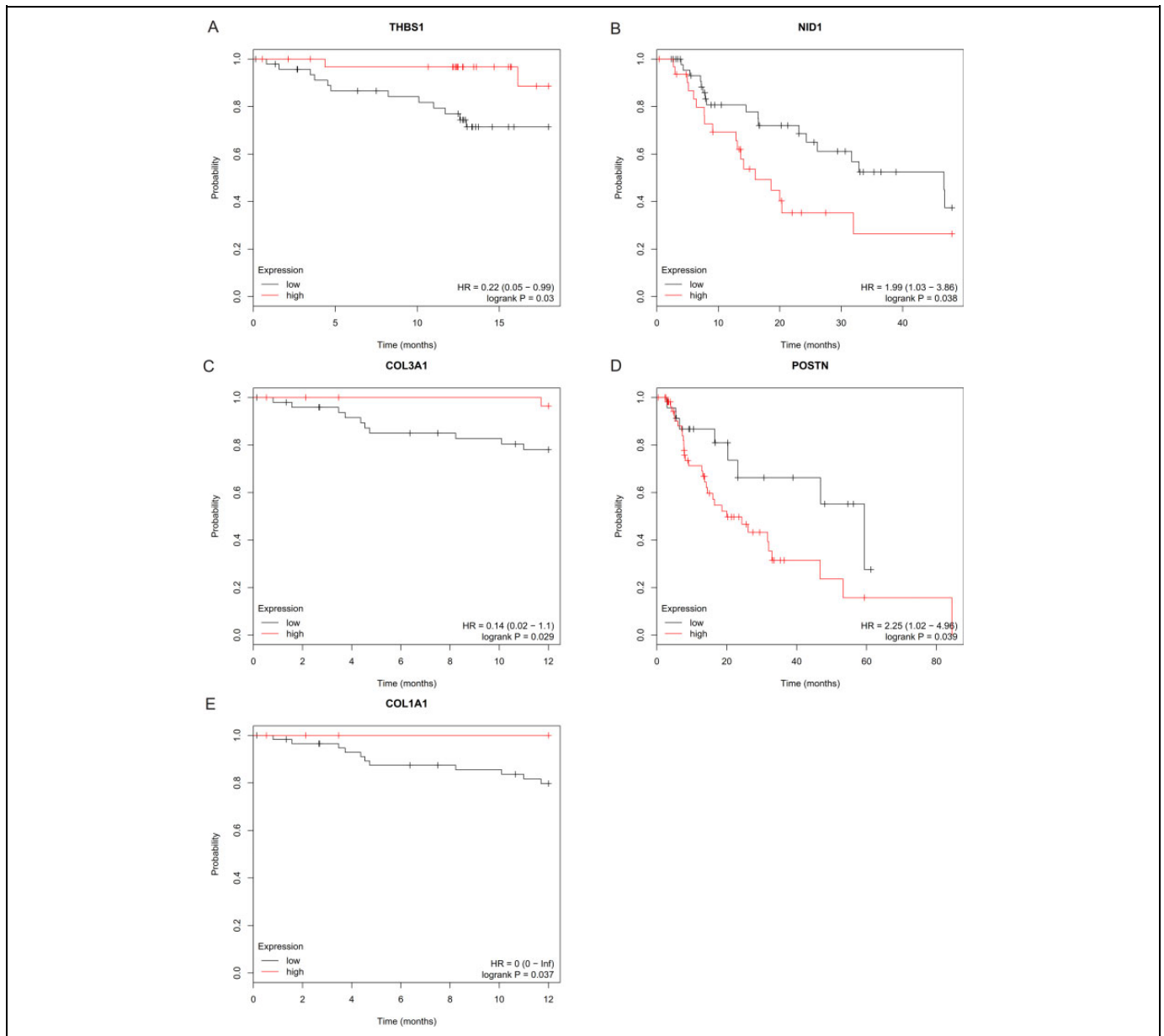
Figure 5. The relationship between the overall survival and the hub genes. (A) THBS1, (B) NID1, (C) COL3A1, (D) POSTN, (E) COL1A1.

Smoking and alcohol also had no effect on times of overall survival (HR = 1.104, P = 0.675 and HR = 1.133, P = 0.613, respectively). However, there was a prominent association between overall survival times and pathologic stage (HR = 2.307, P < 0.001), the expression of COL3A1 (HR = 0.159, P < 0.001), and the expression of POSTN (HR = 2.902, P = 0.002) (Table 3).

Effect of COL3A1 on the Survival Analysis for Patients With EC in the PrognoScan Database

Based on the PrognoScan database, the overall survival times of EC patients who expressed COL3A1 lowly is worse (P < 0.02, HR = 1.3, Figure 9).

Pan-Cancer Analysis for COL3A1 and POSTN

In the pan-cancer analysis based on the esophageal squamous cell carcinoma datasets, EC patients who were expressed COL3A1 and POSTN highly showed worse overall survival times (P < 0.05, Figure 10).

Discussion

EC is a primary malignancy originating from the esophageal epithelium. According to the latest data from China,[20] the incidence of EC in the Chinese population is 11.28/100,000 that is 1.79-times higher than the global average. Squamous cell carcinoma originating from squamous epithelium is the main
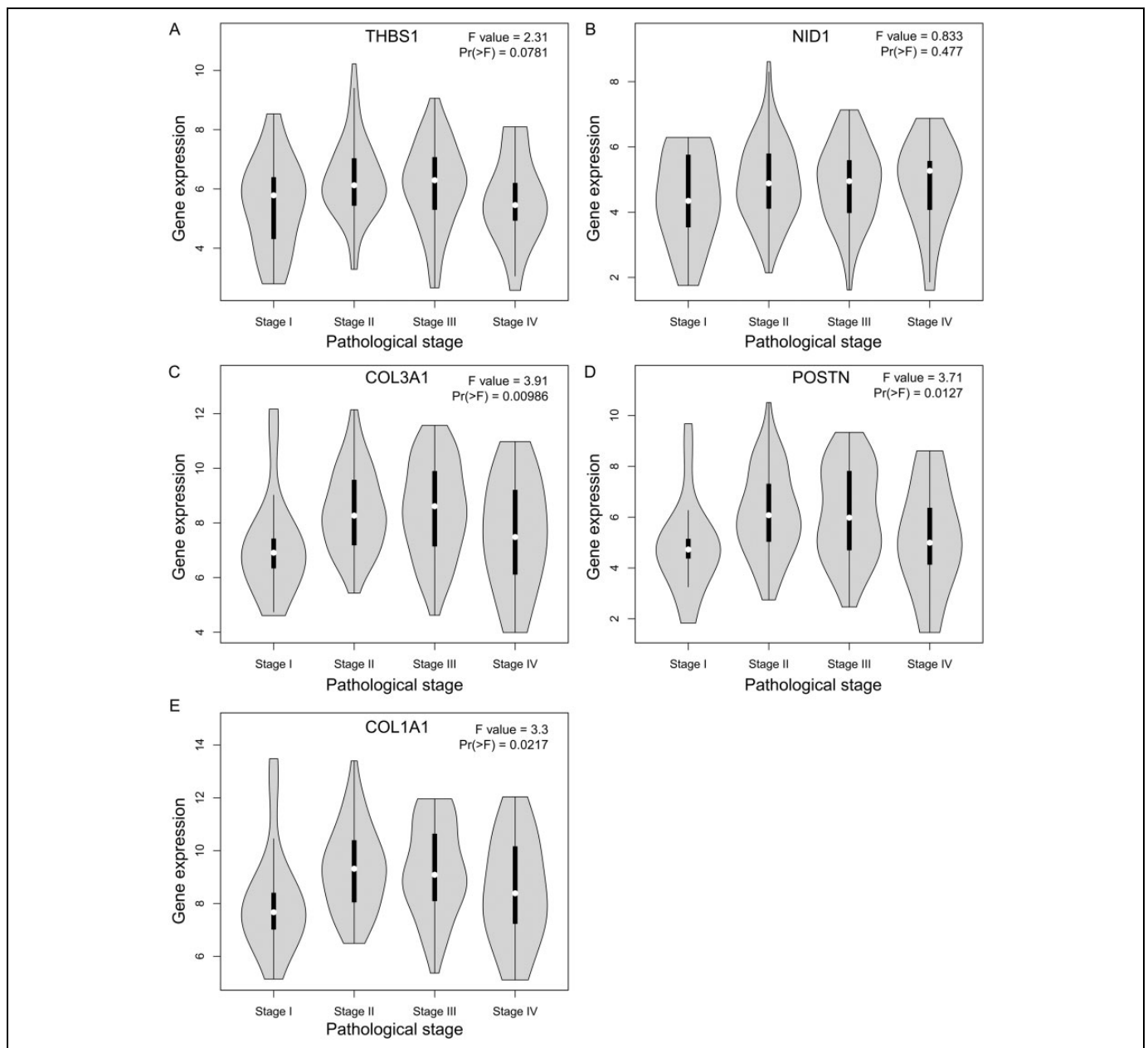
Figure 6. Effect of the hub genes expression on the pathologic stage of patients with EC. (A) THBS1, (B) NID1, (C) COL3A1, (D) POSTN, (E) COL1A1.

histological type, about 90% in all cases.[21] Adenocarcinoma develops from degenerate Barrett's mucosa, located in the lower esophagus.[22] Clinically, progressive dysphagia is a typical symptom. EC is more sensitive to chemotherapy and radiotherapy than other gastrointestinal tumors.[23] Surgical resection is a potential treatment for EC, but it is only appropriate for patients who are able to undergo surgery and who have locally resectable lesions with no signs of distant metastasis. Unfortunately, most people experience tumor recurrence in the first few years following surgery. Palliative treatment is the only option for metastasis, and the 5-year survival rate for patients

who experience recurrence is less than 3%.[3,24] Combining the advantages of each treatment approach and working closely with the treatment team, including paramedic staff, can improve the prognosis for patients with EC. Sato et al[25] showed that the examination of preoperative dental and dental care in patients with EC are critical to reducing the risk of severe postoperative pneumonia. In recent years, multimodality treatment has shown good prospects in improving patients' prognosis and survival while reducing morbidity. However, these measures are still not enough. Whether diagnosis and treatment are timely affected the patient prognosis. Therefore, it is worth
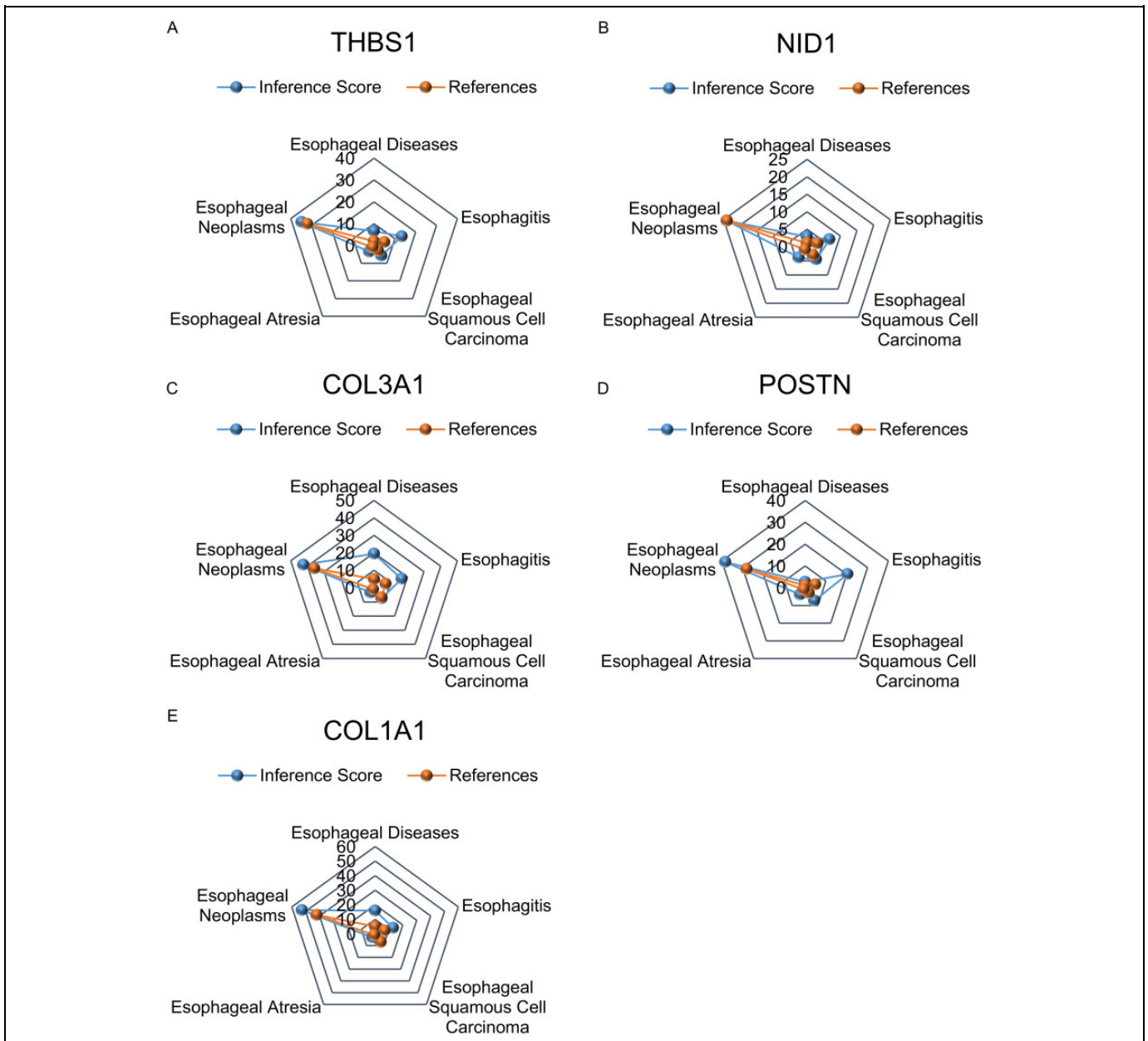
Figure 7. Relationships of hub genes to tumors or metastasis on the basis of the CTD database. (A) THBS1, (B) NID1, (C) COL3A1, (D) POSTN, (E) COL1A1.

exploring the pathogenesis of EC more, find early target therapeutic genes, and molecules, and then develop methods for early diagnosis and timely, individualized treatment.

Bioinformatics have been as a widely used methods to discover genetic molecules related to tumorigenesis and tumor development and then find genes and molecules that can be used as therapeutic targets. Wang et al. used bioinformatics methods and discovered that the molecule MMP1 could discriminate between esophageal adenocarcinoma and Barrett's esophagus, suggesting that it might serve as a potential biomarker.[26] Our study analyzed 2 microarray datasets to identify DEGs. 36 DEGs were identified in the 2 datasets. By means

of bioinformatics analysis, it was found that collagen type III alpha 1 chain (COL3A1) and periostin (POSTN) proteins had high expression in EC tissues than their expression in healthy tissues. In the overall survival analysis, it was concluded that the prognosis for patients with EC was poor if POSTN was highly expressed. However, if COL3A1 was expressed at low levels, patients with EC had a poor prognosis. Analysis of pathologic stages showed that when both COL3A1 and POSTN were highly expressed, the pathologic stage of patients with EC was relatively high.

COL3A1 plays a part in cell adhesion, migration, proliferation, and differentiation via interactions with cell-surface
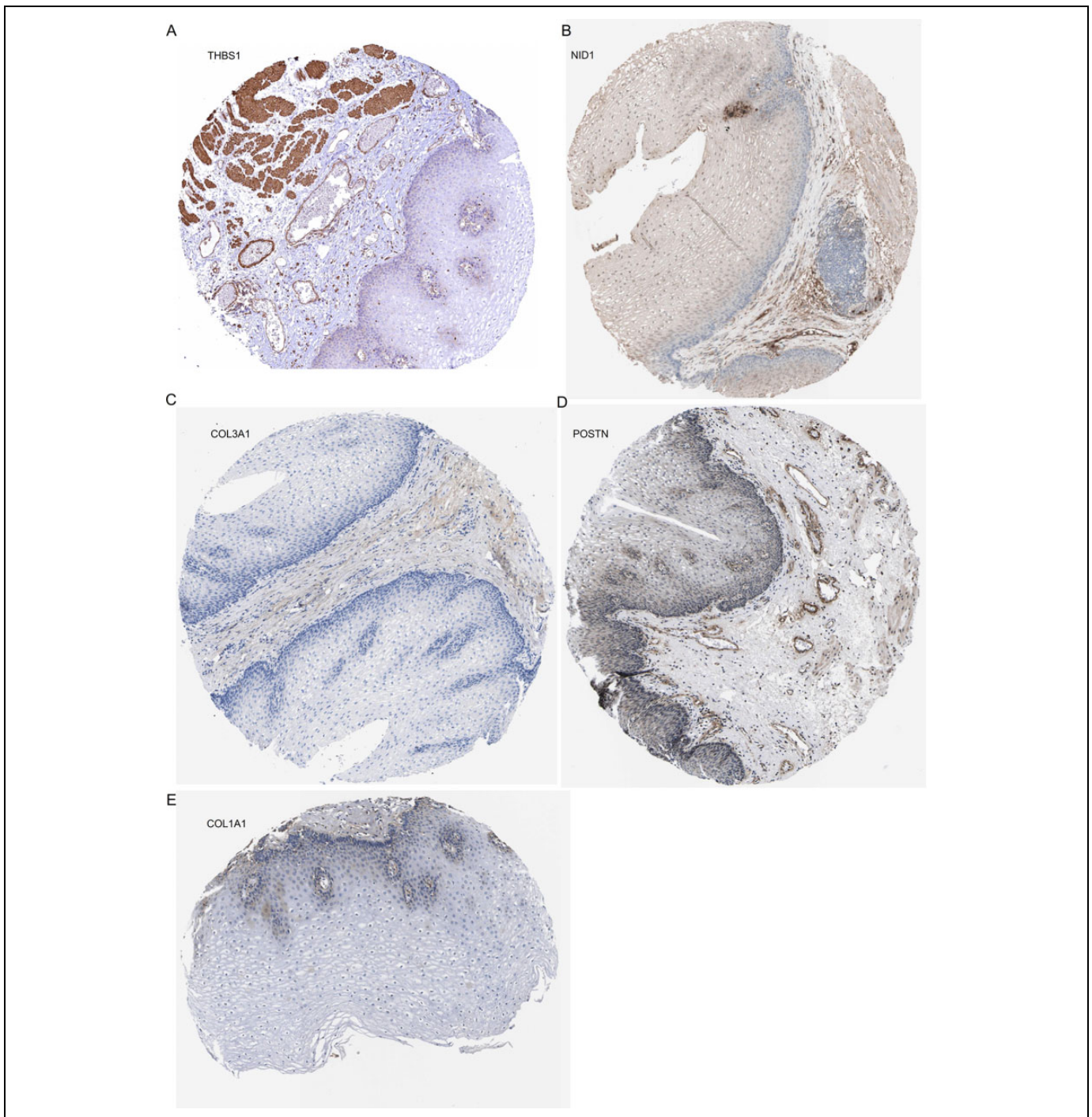
Figure 8. Verification of the hub genes on a level of translation in healthy esophageal tissues through the Human Protein Atlas database. (A) THBS1, (B) NID1, (C) COL3A1, (D) POSTN, (E) COL1A1.

receptor integrins.[27] COL3A1 is situated on the long arm of chromosome 2 and is mainly expressed in elastic connective tissue. COL3A1 plays a part in cell adhesion, migration, proliferation and differentiation by interacting with cell surface receptor integrins.[28] A mutation in COL3A1 causes vascular Ehlers–Danlos syndrome (vEDS), a rare, life-threatening genetic disease.[29] The abnormal expression of COL3A1 in

various cancer is significantly connected with bad overall survival rates, e.g. in bladder cancer, breast cancer, and smoking-related lung cancer[30-32]; its expression may affect the tumor microenvironment and regulate the ECM through collagen degradation and deposition to promote tumor progression.[33,34] Engqvist et al. suggested that COL3A1 is related to a bad prognosis of breast cancer. It maybe influences the

Table 2. Summary of miRNAs That Regulate the Hub Genes.

|	Gene	Predicted MiR
1	THBS1	hsa-miR-338-3p
2	NID1	hsa-miR-1297
		hsa-miR-4465
		hsa-miR-26a-5p
		hsa-miR-26b-5p
3	COL3A1	hsa-miR-29a-3p
		hsa-miR-29b-3p
		hsa-miR-29c-3p
4	POSTN	hsa-miR-19b-3p
		hsa-miR-19a-3p
5	COL1A1	hsa-miR-129-5p

Table 3. Correlative Clinicopathologic Factors and Their Impact on Overall Survival (OS) According to Multivariate Cox Proportional Regression Analysis.

Characteristic	OS		
	HR	95% CI	P-value
Age	1.218	0.781-1.900	0.385
Smoking	1.201	0.752-1.919	0.443
Alcohol	0.983	0.591-1.635	0.948
Pathologic stage	1.853	1.397-2.458	<0.001*
COL3A1	0.159	0.077-0.326	<0.001*
POSTN	2.902	1.497-5.623	0.002*

HR, hazard ratio; 95% CI, 95% confidence interval. * $P < 0.05$.
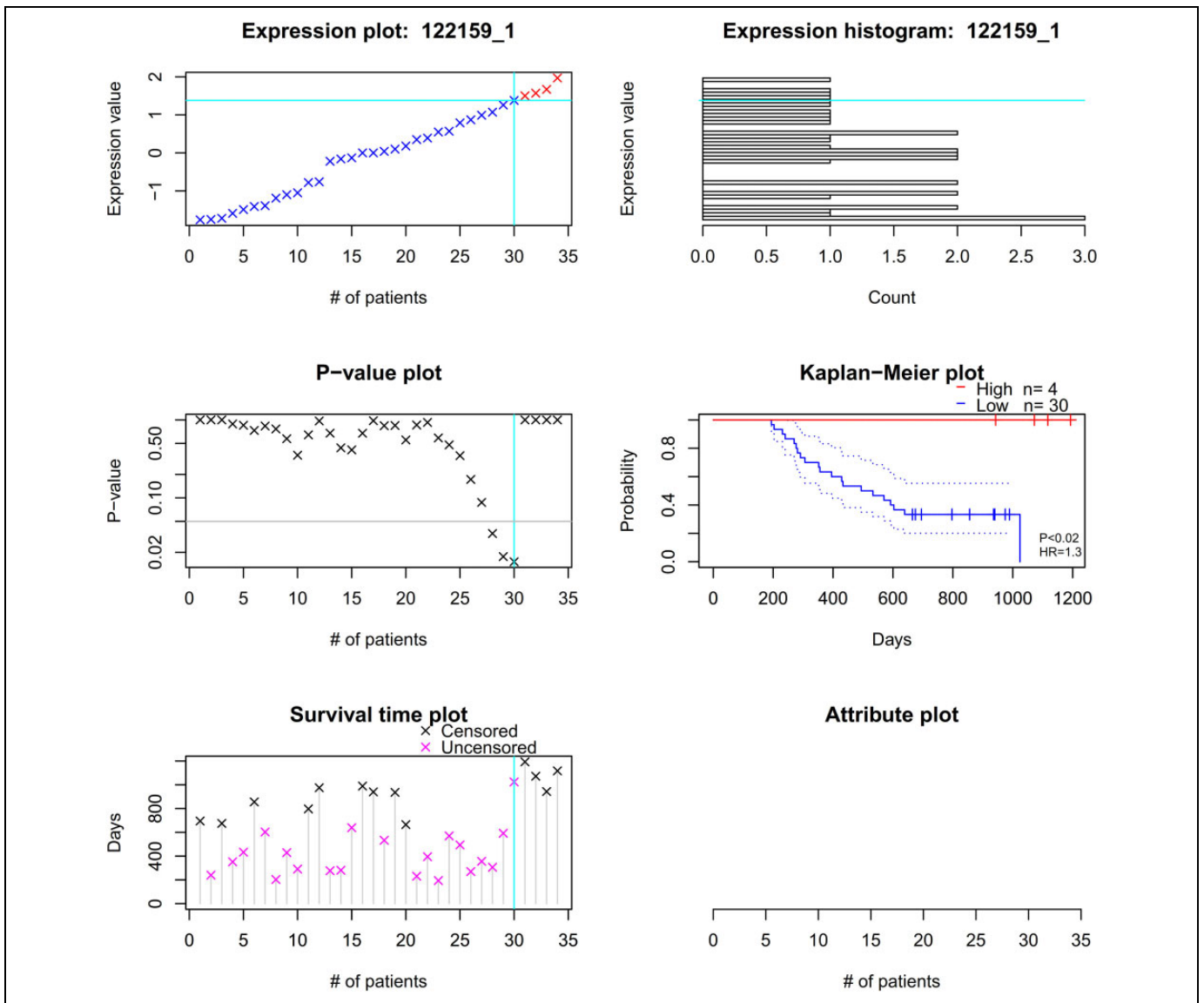
Figure 9. Effect of COL3A1 on the survival analysis for patients with EC in the PrognoScan database.
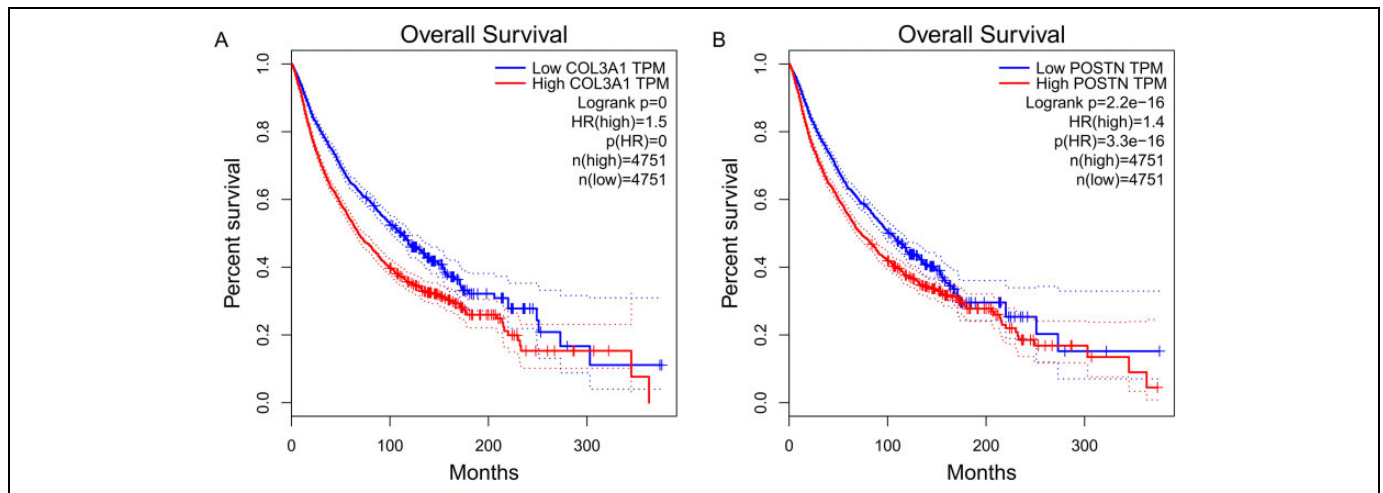
Figure 10. Pan-cancer analysis for COL3A1 and POSTN. A, Effect of COL3A1 on overall survival. B, Effect of POSTN on overall survival.

development of early breast cancer.[35] Some studies have shown that COL3A1 can promote proliferation and migration of colorectal cancer cells through stimulating PI3K-AKT signaling pathway or the WNT/PCP signaling pathway.[35,36] Our bioinformatics analysis showed that there has high COL3A1 expression in EC tissues. When COL3A1 was highly expressed, patients with EC were at a higher pathologic stage, which may be closely connected with the 3 main collagen functions. First, collagen provides scaffolds and sites for cell adhesion. Second, collagen acts as a reservoir of ECM proteins, proteoglycans, and growth factors. Third, the growth, differentiation, and invasion of the tumor were affected by ligands for the signal transduction networks.

The POSTN gene is situated on the long arm of chromosome 13(13q13.3) and is about 36 kb in length. It can induce cell adhesion and diffusion and plays a crucial part in in the maintenance, regulation, and amplification of tumor stem cells during metastasis.[37,38] Many researches had found that POSTN expressed various tumor tissues abnormally.[39] Kikuchi et al. found that POSTN promotes the proliferation of gastric cancer cells while activating the ERK pathway.[40] Another study showed that the migration and invasion of renal cancer cells were promoted by POSTN via the integrin/focal adhesion kinase/c-Jun N-terminal kinase pathway.[41] Wang et al. proposed that POSTN levels expressed highly related to poor prognosis and short overall survival times in esophageal squamous cell carcinoma patients. A multivariate analysis found that the POSTN expression was an independent prognostic factor in the aspect of the tumor differentiation, venous invasion, and TNM (tumor, node, metastasis) staging.[42] POSTN regulates epithelial-mesenchymal transition (EMT) by activating the ERK and p38 pathways and down-regulating the expression of miR-381, an miRNA that targets Twist and Snail mRNAs thought to be related to tumor cell invasion and metastasis.[43] Our bioinformatics analysis showed that the higher the POSTN expression, the worse the prognosis. Our analysis of pathologic stages showed that patients with EC were at a relatively higher

pathologic stage when the expression level of POSTN was higher.

Although the bioinformatics analysis methods used in the study was rigorous, some deficiencies remain. Larger clinical sample datasets and animal experiments are necessary to make a comprehensive verification of the results and to obtain a better understanding of the pathogenesis of EC.

Conclusion

COL3A1 and POSTN might play an important role in the worse pathologic stages of EC, and they may provide some novel ideas and basis for the diagnosis and targeted treatment of EC.

Authors' Note

Acknowledgments

Declaration of Conflicting Interests

ORCID iD

Na Wang ⬤ https://orcid.org/0000-0002-0121-8816

References

1. He Y, Li D, Shan B, et al. Incidence and mortality of esophagus cancer in China, 2008-2012. *Chin J Cancer Res*. 2019;31(3):426-434.

2. Bray F, Ferlay J, Soerjomataram I, et al. 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.

3. Hur C, Miller M, Kong CY, et al. Trends in esophageal adenocarcinoma incidence and mortality. *Cancer*. 2013;119(6):1149-1158.

4. Rubenstein JH, Shaheen NJ. Epidemiology, diagnosis, and management of esophageal adenocarcinoma. *Gastroenterology*. 2015;149(2):302-317.E1.

5. Zhou Y, Liepe J, Sheng X, Stumpf MPH, Barnes C. GPU accelerated biochemical network simulation. *Bioinformatics*. 2011;27(6):874-876.

6. Nobile MS, Cazzaniga P, Tangherloni A, Besozzi D. Graphics processing units in bioinformatics, computational biology and systems biology. *Brief Bioinformatics*. 2017;18(5):870-885.

7. Li L, Lei Q, Zhang S, Kong L, Qin B. Screening and identification of key biomarkers in hepatocellular carcinoma: evidence from bioinformatic analysis. *Oncol Rep*. 2017;38(5):2607-2618.

8. Milan T, Wilhelm BT. Mining cancer transcriptomes: bioinformatic tools and the remaining challenges. *Mol Diagn Ther*. 2017;21:249-258.

9. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. 2017;45(3):W98-W102.

10. Wang XQ, Tang ZX, Yu D, et al. Epithelial but not stromal expression of collagen alpha-1(III) is a diagnostic and prognostic indicator of colorectal carcinoma. *Oncotarget*. 2016;7:8823-8838.

11. Lin Z, Jaberi-Douraki M, He C, et al. Performance assessment and translation of physiologically based pharmacokinetic models from acslX to Berkeley Madonna, MATLAB, and R language: oxytetracycline and gold nanoparticles as case examples. *Toxicol Sci*. 2017;158(1):23-35.

12. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res*. 2013;41(D1):D991-D995.

13. Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015;43(D1):D447-D452.

14. Su G, Morris JH, Demchak B, Bader GD. Biological network exploration with Cytoscape 3. *Current Protocol Bioinformatics*. 2014;47(1):8.13.1-8.13.24.

15. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*. 2011;27(3):431-432.

16. Huang DW, Sherman BT, Tan Q, et al. The DAVID gene functional classification tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol*. 2007;8:R183.

17. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25:25-29.

18. Zhou Y, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. 2019;10:1523.

19. Davis AP, Grondin CJ, Johnson RJ, et al. The comparative toxicogenomics database: update 2019. *Nucleic Acids Res*. 2019;47(D1):D948-D954.

20. Zheng RS, Sun KX, Zhang SW, et al. Report of cancer epidemiology in China, 2015 [in Chinese]. *Zhonghua Zhong Liu Za Zhi*. 2019;41(1):19-28.

21. Kitagawa Y, Uno T, Oyama T, et al. Esophageal cancer practice guidelines 2017 edited by the Japan Esophageal Society: part 1. *Esophagus*. 2019;16:1-24.

22. Enzinger PC, Mayer RJ. Esophageal cancer. *N Engl J Med*. 2003;349:2241-2252.

23. Sohda M, Kuwano H. Current status and future prospects for esophageal cancer treatment. *Ann Thorac Cardiovasc Surg*. 2017;23(1):1-11.

24. Janmaat VT, Steyerberg EW, van der Gaast A, et al. Palliative chemotherapy and targeted therapies for esophageal and gastroesophageal junction cancer. *Cochrane Database Syst Rev*. 2017;11:CD004063.

25. Sato Y, Motoyama S, Takano H, et al. Esophageal cancer patients have a high incidence of severe periodontitis and preoperative dental care reduces the likelihood of severe pneumonia after esophagectomy. *Dig Surg*. 2016;33:495-502.

26. Yue Y, Song M, Qiao Y, et al. Gene function analysis and underlying mechanism of esophagus cancer based on microarray gene expression profiling. *Oncotarget*. 2017;8(7):105222-105237.

27. Chen Z, Soutto M, Rahman B, et al. Integrated expression analysis identifies transcription networks in mouse and human gastric neoplasia. *Genes Chromosomes Cancer*. 2017;56:535-547.

28. Kim JK, Xu Y, Xu X, et al. A novel binding site in collagen type III for integrins alpha1beta1 and alpha2beta1. *J Biol Chem*. 2005;280:32512-32520.

29. Kuivaniemi H, Tromp G. Type III collagen (COL3A1): Gene and protein structure, tissue distribution, and associated diseases. *Gene*. 2019;707:151-171.

30. Chen Y, Pan Y, Ji Y, Sheng L, Du X. Network analysis of differentially expressed smoking-associated mRNAs, lncRNAs and miRNAs reveals key regulators in smoking-associated lung cancer. *Exp Ther Med*. 2018;16(6):4991-5002.

31. Shi S, Tian B. Identification of biomarkers associated with progression and prognosis in bladder cancer via co-expression analysis. *Cancer Biomark*. 2019;24(2):183-193.

32. Srour MK, Gao B, Dadmanesh F, et al. Gene expression comparison between primary triple-negative breast cancer and paired

axillary and sentinel lymph node metastasis. *Breast J*. 2019;26(5): 904-910.

33. Fang M, Yuan J, Peng C, Li Y. Collagen as a double-edged sword in tumor progression. *Tumour Biol*. 2014;35:2871-2882.

34. Wang M, Zhao J, Zhang L, et al. Role of tumor microenvironment in tumorigenesis. *J Cancer*. 2017;8(5):761-773.

35. Engqvist H, Parris TZ, Kovács A, et al. Immunohistochemical validation of COL3A1, GPR158 and PITHD1 as prognostic biomarkers in early-stage ovarian carcinomas. *BMC Cancer*. 2019;19:928.

36. Wang Z, Monteiro CD, Jagodnik KM, et al. Extraction and analysis of signatures from the gene expression omnibus by the crowd. *Nat Commun*. 2016;7:12846.

37. Litvin J, Selim AH, Montgomery MO, et al. Expression and function of periostin-isoforms in bone. *J Cell Biochem*. 2004;92(5): 1044-1061.

38. Park SY, Piao Y, Jeong KJ, Dong J, de Groot JF. Periostin (POSTN) regulates tumor resistance to antiangiogenic therapy in glioma models. *Mol Cancer Ther*. 2016;15(9):2187-2197.

39. Gonzalez-Gonzalez L, Alonso J. Periostin: a matricellular protein with multiple functions in cancer development and progression. *Front Oncol*. 2018;8:225.

40. Kikuchi Y, Kunita A, Iwata C, et al. The niche component periostin is produced by cancer-associated fibroblasts, supporting growth of gastric cancer through ERK activation. *Am J Pathol*. 2014;184(3):859-870.

41. Chuanyu S, Yuqing Z, Chong X, Guowei X, Xiaojun Z. Periostin promotes migration and invasion of renal cell carcinoma through the integrin/focal adhesion kinase/c-Jun N-terminal kinase pathway. *Tumour Biol*. 2017;39(4):1010428317694549.

42. Wang W, Sun QK, He YF, et al. Overexpression of periostin is significantly correlated to the tumor angiogenesis and poor prognosis in patients with esophageal squamous cell carcinoma. *Int J Clin Exp Pathol*. 2014;7(2):593-601.

43. Hu WW, Chen PC, Chen JM, et al. Periostin promotes epithelial-mesenchymal transition via the MAPK/miR-381 axis in lung cancer. *Oncotarget*. 2017;8:62248-62260.