MEDICAL SCIENCE MONITOR

DATABASE ANALYSIS

e-ISSN 1643-3750
© Med Sci Monit, 2020; 26: e920355
DOI: 10.12659/MSM.920355

Received: 2019.09.27
Accepted: 2020.01.13
Available online: 2020.02.01
Published: 2020.03.24

Secreted Phosphoprotein 1 (SPP1) and Fibronectin 1 (FN1) Are Associated with Progression and Prognosis of Esophageal Cancer as Identified by Integrated Expression Profiles Analysis

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

ABCDEF 1 Menglu Li
ABCDEF 2 Kaige Wang
ABCDEF 3 Yanhua Pang
BCDEF 2 Hongpan Zhang
BC 1 Hao Peng
DE 1 Qi Shi
F 2 Zhiyu Zhang
AG 1 Xiaobin Cui
AG 1,2 Feng Li

1 Department of Pathology and Key Laboratory for Xinjiang Endemic and Ethnic Diseases, The First Affiliated Hospital, Shihezi University School of Medicine, Shihezi, Xinjiang, P.R. China
2 Department of Pathology and Medical Research Center, Beijing Chaoyang Hospital, Capital Medical University, Beijing, P.R. China
3 Department of Gastroenterology, Beijing Chaoyang Hospital, Beijing, P.R. China

Corresponding Authors: Feng Li, e-mail: lifeng7855@126.com, Xiaobin Cui, e-mail: cuixiaobin4363@foxmail.com
Source of support: This study was supported by grants from the National Natural Science Foundation of China (No. 8177111318, 81460362), the Major Science and Technology Projects of Shihezi University (No. gxjs2014-zdgg06), the Applied Basic Research Projects of Xinjiang Production and Construction Corps (No. 2016AG020), and the High-Level Talent Project of Shihezi University (No. RCZX201533)

Background: Esophageal cancer is a malignant tumor with a complex pathogenesis and a poor 5-year survival rate, which encourages researchers to explore its molecular mechanisms deeper to improve the prognosis.

Material/Methods: DEGs were from 4 Gene Expression Omnibus (GEO) databases (GSE92396, GSE20347, GSE23400, and GSE45168) including 87 esophageal tumor samples and 84 normal samples. We performed Gene Ontology (GO) analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis, Protein-Protein interaction (PPI) analysis, and GeneMANIA to identify the DEGs. Gene set enrichment analysis (GSEA) and Kaplan-Meier survival analyses were performed.

Results: There was an overlapping subset consisting of 120 DEGs that was present in all esophageal tumor samples. The DEGs were enriched in extracellular matrix (ECM)-receptor interaction, as well as focal adhesion and transcriptional mis-regulation in cancer. The 2 most crucial regulatory pathways in esophageal cancer were the amebiasis pathway and the PI3K-Akt signaling pathway. Secreted phosphoprotein 1 (SPP1) and fibronectin 1 (FN1) were selected and verified in an independent cohort and samples using the TCGA and GTEx projects. Gene set enrichment analysis (GSEA) showed that proteasome and nucleotide excision repair were 2 most differentially enriched pathways in the SPP1 high-expression phenotype, and ECM-receptor interaction and focal adhesion in FN1 high-expression phenotype. Kaplan-Meier survival analysis showed that SPP1 and FN1 were significantly positively related to overall survival and had the potential to predict patient relapse.

Conclusions: Our analysis is the first to show that SPP1 and FN1 might work as biological markers of progression and prognosis in esophageal carcinoma (ESCA).

MeSH Keywords: Biological Markers • Esophageal Neoplasms • Fibronectins • Osteopontin

Full-text PDF: https://www.medscimonit.com/abstract/index/idArt/920355

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

Background

Esophageal cancer is the eighth most common cancer worldwide [1] and is the fifth leading cause of death in China. The 5-year survival rate of EC patients is <10% due to the lack of valid biomarkers for early diagnosis and efficacious drugs [2], and there are no reliable molecular targets in esophageal cancer. Therefore, it is vital to find novel biomarkers for the prognosis of esophageal cancer and to develop new targeted therapies.

Due to the lack of specific early symptoms and the lack of biomarkers, the prognosis of ESCA is poor. Recently, it was reported that several genes are related to the diagnosis of ESCA, including Bmi-1 [3], Ezrin [4], and LOC285194 [5]. However, there are currently no specific molecular markers of early-stage esophageal cancer. To screen biomarkers to enhance the diagnosis of ESCA, we searched for differences in gene expression between ESCA and normal samples.

To identify pathogenic genes, microarray analyses was performed to generate massive datasets in accessible cancer databases, which is an established effective method [6]. Cancer-related public databases such as the GEO and The Cancer Genome Atlas (TCGA) describe the matched normal tissues of 33 cancers and over 11 000 patients and have been widely used in the identification of prognostic gene properties of cancer [7]. To identify new potential therapeutic and prognostic targets in esophageal cancer, we integrated 4 separate esophageal cancer datasets in the GEO database, and we also demonstrated that the ECM-receptor interaction was the most important regulatory pathway in ESCA. We screened 2 genes, SPP1 and FN1, and found that the expression of SPP1 and FN1 in esophageal cancer samples was elevated in the GEO data and in data from the TCGA and GTEx projects. Gene set enrichment analysis (GSEA) was performed to study biologic pathways shared by different SPP1 or FN1 expression levels. Kaplan-Meier analysis demonstrated that patients with high expression levels of SPP1 and FN1 had worse overall survival. These results indicated that SPP1 and FN1 had the potential to be biomarkers for esophageal cancer, and dysregulation of SPP1 and FN1 was associated with the development and prognosis of this malignancy.

Material and Methods

Screening gene expression profile data

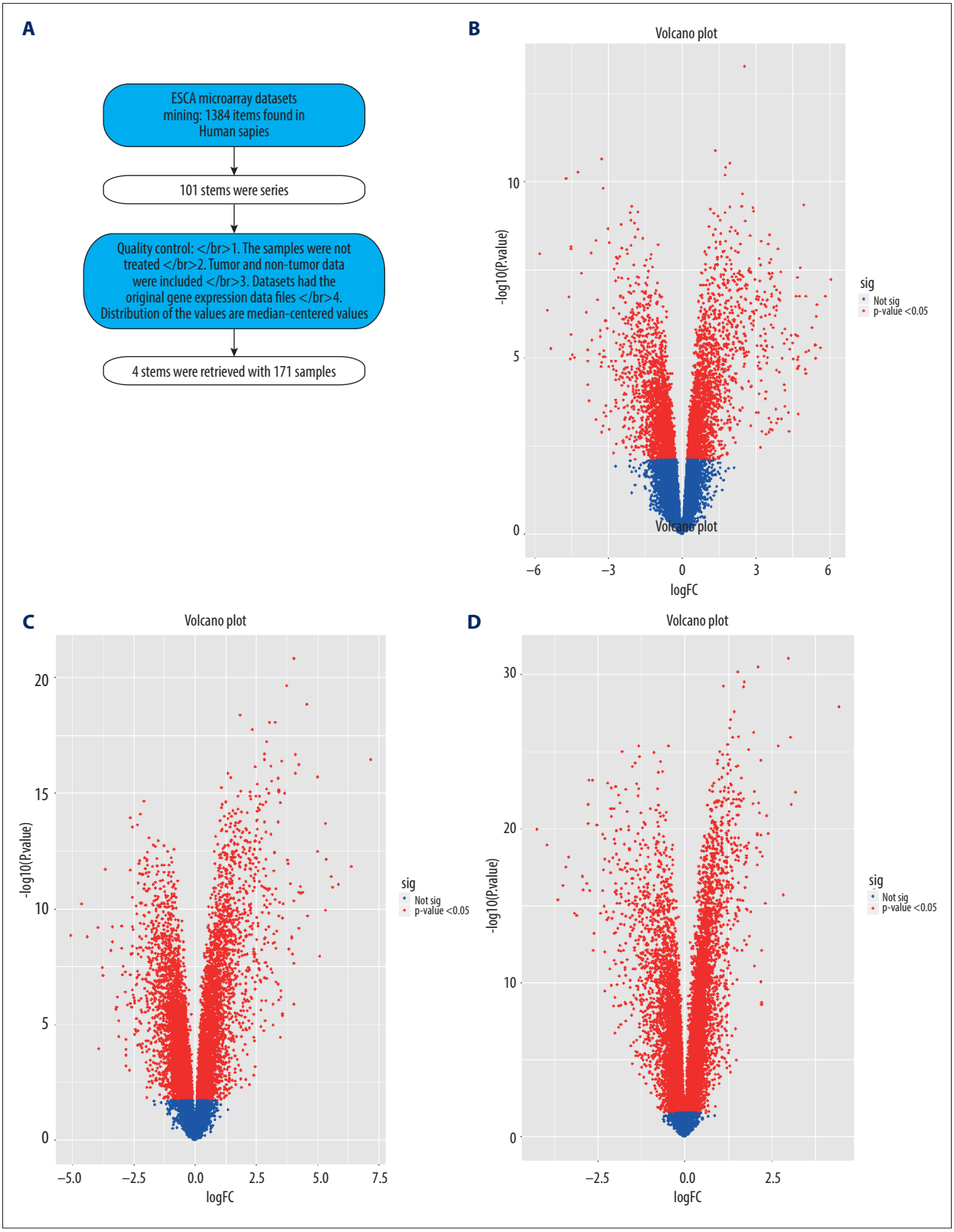
The National Center for Biotechnology Information (NCBI) GEO database (*https://www.ncbi.nlm.nih.gov/geo/*) is a databank with many gene expression datasets and online resources and is used to retrieve gene expression data from any species or man-made source. We used the keywords 'esophageal cancer' and the organism 'Homo sapiens'. The inclusion criteria were: (1) No samples received any chemical or physical treatment; (2) There were tumor and non-tumor data in the same study; (3) Datasets had the original gene expression data files; and (4) The distribution of the values for the samples we selected were median-centered values, which showed normalized and cross-comparable data. This was the basis for the qualified microarray.

We searched and obtained 1384 items, and gene expression profiling of patients with esophageal cancer GSE92396, GSE20347, GSE23400, and GSE45168 was finally obtained. Dataset GSE92396, analyzed with the GPL6244 platform ([HuGene-10-st] Affymetrix Human Gene 1.0 ST Array), consisted of 21 samples including 12 esophageal tumor tissues and 9 normal esophageal tissues. Data from the pooled normal sample GSM2428987 in GSE92396 were not included in this study. The dataset GSE20347, analyzed with the GPL571 platform ([HG-U133A 2] Affymetrix Human Genome U133A 2.0 Array), included 17 matched micro-dissected tumors and normal tissues. For dataset GSE23400, analyzed with the GPL96 platform ([HG-U133A] Affymetrix Human Genome U133A Array), 106 samples were included with 53 esophageal tumor tissues and 53 normal esophageal tissues. Data of samples in platform GPL97 were not used in our study. The GSE45168 dataset, analyzed with the GPL13497 platform (Agilent-026652 Whole Human Genome Microarray 4x44K v2), consisted of 10 samples derived from ESCA patients (5 cancerous tissues and 5 normal tissues). The dataset GSE13898 was applied to validation cohorts, which was analyzed with the platform GPL6102 (Illumina human-6 v2.0 expression BeadChip), and included 75 cancer and 28 normal samples. The main characteristics of these 4 datasets that met our criteria are shown in Table 1.

Table 1. Main features of 4 selected studies of gene expression microarray data.

GEO datasets	Platform	Samples in total	Regions	Submission date	Citation(s) on
GSE92396	GPL6244	21	USA	Dec 14, 2016	Sci Rep, 2017; 7: 40729
GSE20347	GPL571	34	USA	Feb 16, 2010	BMC Genomics, 2010; 11: 576
GSE23400	GPL96	106	USA	Aug 03, 2010	Clin Cancer Res, 2011; 17(9): 2955–66
GSE45168	GPL13497	10	China	Mar 14, 2013	Int J Clin Exp Pathol, 2014; 7(6): 3132–40

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



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

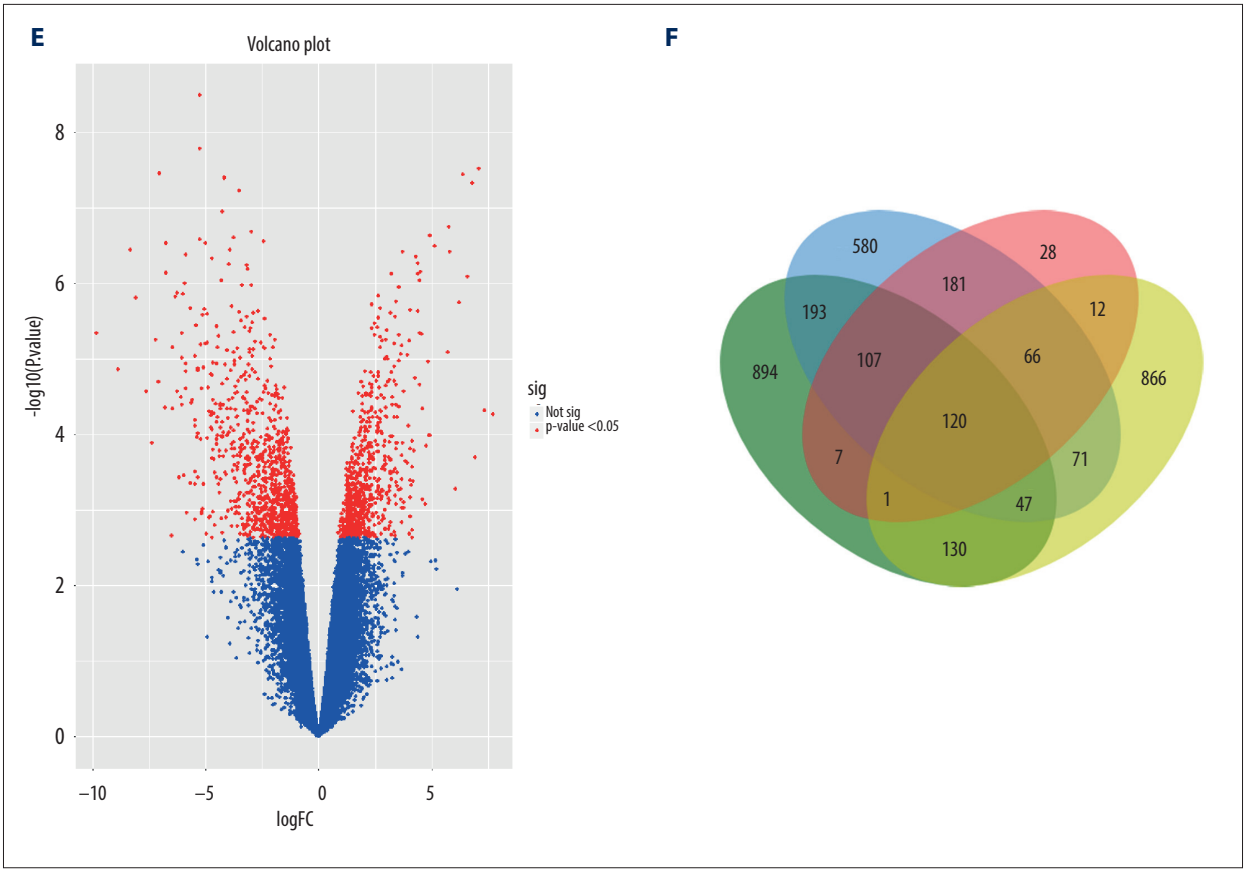


Figure 1. Identification of DEGs between esophageal carcinomas and non-malignant tissues. (**A**) Flow chart for study selection. **B–E**) Volcano plots of differentially expressed genes. X-axis: log fold-change; Y-axis: - log10 (P value) for each probe. There were 796 genes identified as being upregulated and 781 genes downregulated in GSE92396 (**B**), 964 genes were upregulated and 768 genes downregulated in GSE20347 (**C**), 316 genes were upregulated and 356 genes were downregulated in GSE23400 (**D**), 644 genes were upregulated and 879 genes were downregulated in GSE45168 (**E**). (**F**) Venn diagram of the overlapping genes of the 2 sets of DEGs. There were120 DEGs that were significantly differentially expressed in all DEGs sets.

DEG identification and clustering

GEO2R (*https://www.ncbi.nlm.nih.gov/geo/geo2r/*) was utilized to explore DEGs [8] by comparing the tumor and normal samples to identify DEGs across experimental conditions. The fold-change (FC) of gene expression was determined under the limitation of |logFC| ≥1 and P value <0.05 for DEG selection. A volcano plot was used to visualize the identified DEGs using ggplot2 of R [9]. Jvenn (*http://bioinfo.genotoul.fr/jvenn*), an open-source component to analyze data, was used to construct Venn diagrams [10].

Pathway analysis

GO and KEGG were utilized to confirm characteristic biological attributes and pathways, diseases, drugs, and chemical substances [11] by using online tools of the Database for Annotation Visualization and Integrated Discovery (DAVID, *https://david.ncifcrf.gov/*), a website that integrated and demonstrated the detailed function of a mass of proteins or genes using bioinformatics [12]. Enriched terms with P<0.05 were selected. After that, the R package ggplot2 was utilized to visualize the main biological processes, molecular functions, and cellular components [9].

Additionally, many functional relationships and interactions occurring among proteins whose systematic features revealed the context of molecular systems biology were at the core of cellular processing. The Search Tool for the Retrieval of Interacting Genes (STRING) database (*http://string-db.org*) was used in our study to evaluate and visualize the interactions of genes within a network [13]. Discounted nodes in the network were hidden. To visualize modules of the PPI network, the Molecular Complex Detection (MCODE) in the Cytoscape app was used, with the degree limitation of 2, the node score limitation of 0.2, the k-core limitation of 2, and the max.depth limitation of 100. For each module, DAVID was used to analyze the pathway of genes. To find and categorize related

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

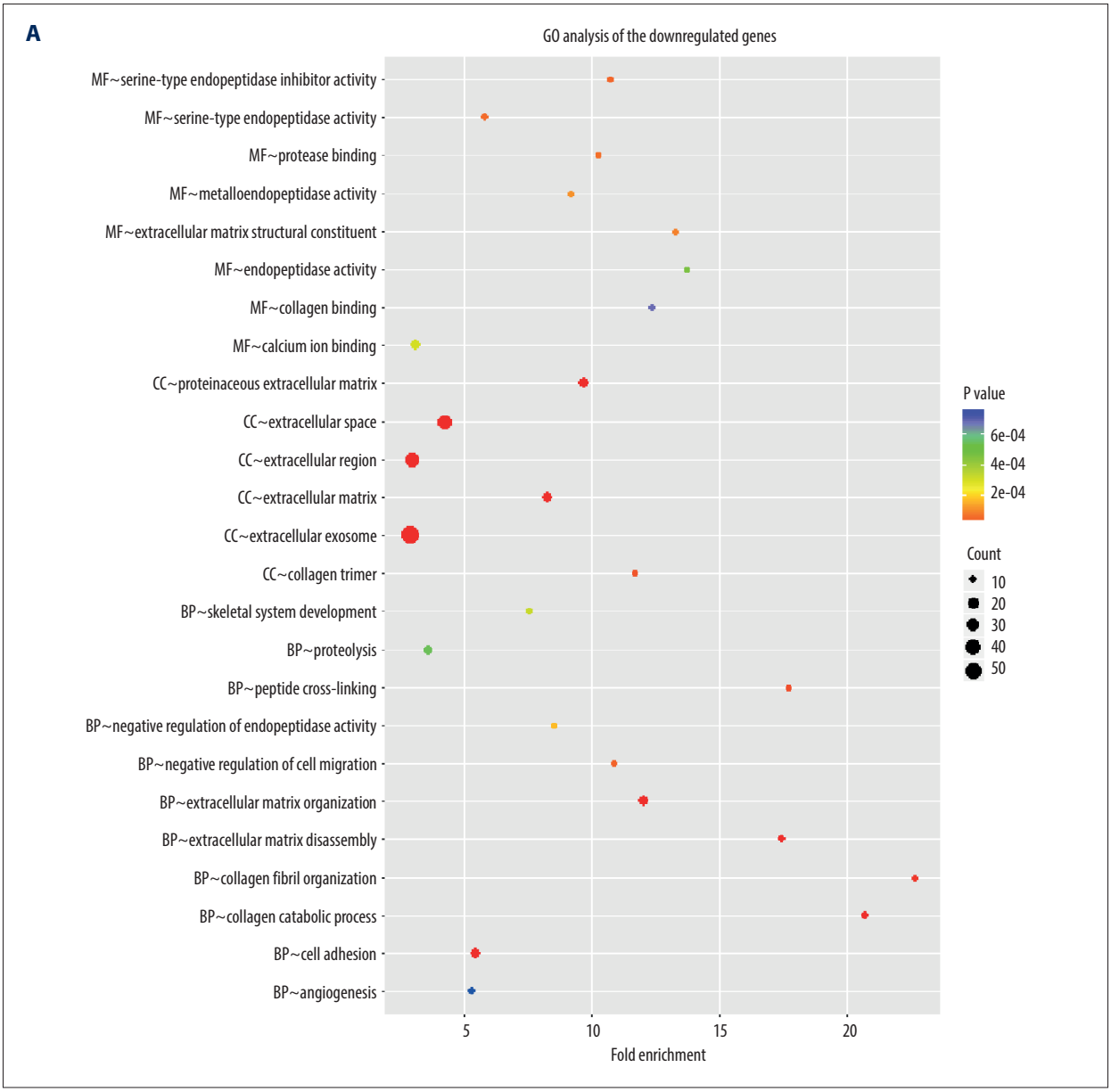
genes, including co-expression and co-localization, we used GeneMANIA (*http://genemania.org/*), which searched many biological datasets, including GEO, PFAM (*http://pfam.sanger.ac.uk*), BioGRID (*http://thebiogrid.org/*), and Pathway Commons (*http://www.pathwaycommons.org/*) and other databases [14].

Gene set enrichment analysis (GSEA)

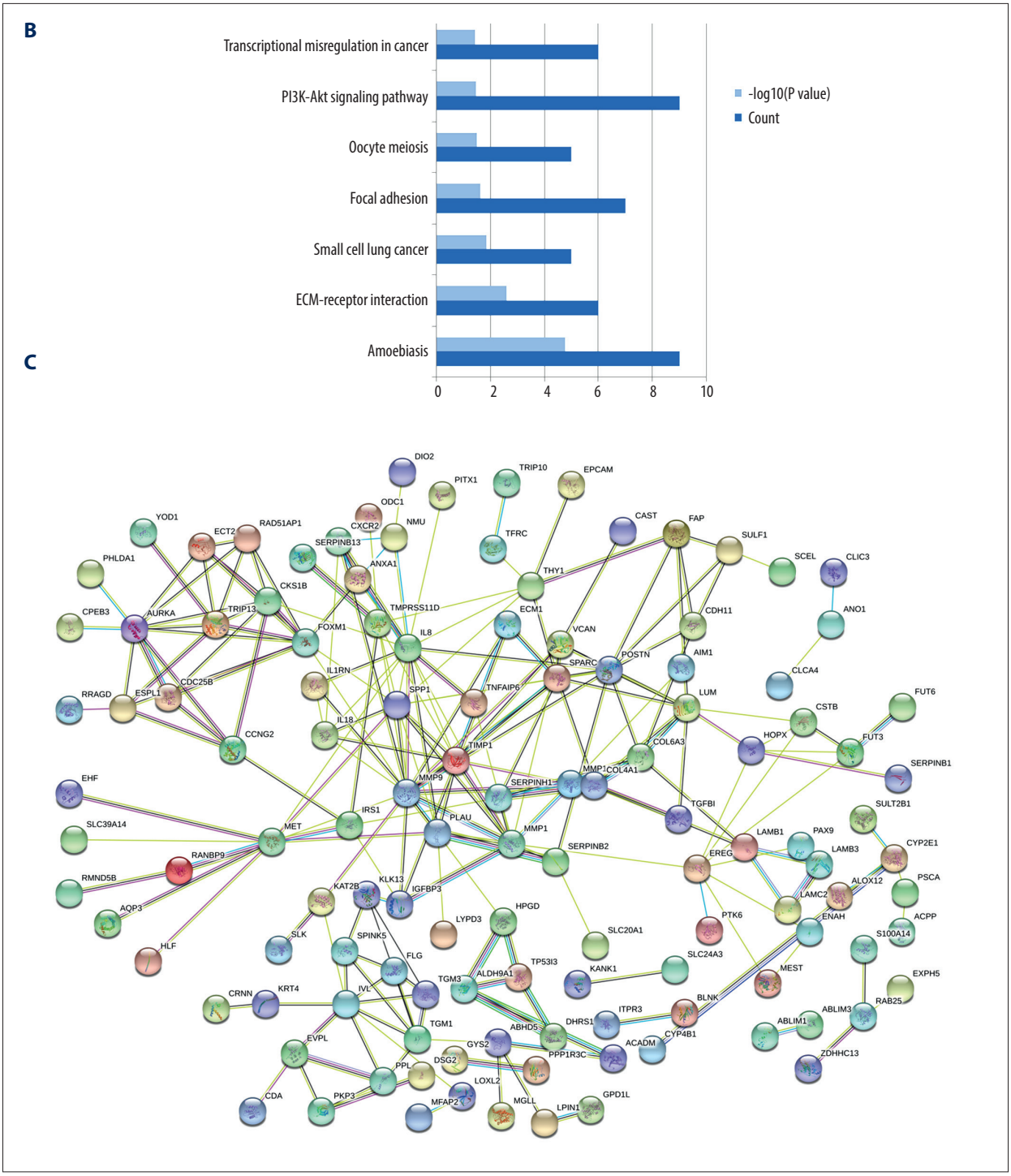
GSEA is a computational method (*https://software.broadinstitute.org/gsea/index.jsp*) used to assess correlations between pathways and gene expression [15]. To identify the correlations between common gene expression and related pathways in esophageal cancer, GSEA was used to look for significant differences between high expression and low expression. Terms enriched with nominal p value <0.05 and normalized enrichment score |NES| ≥1.8 were selected.

Expression level and survival analysis of the key genes

The Gene Expression Profiling Interactive Analysis (GEPIA) (*http://gepia.cancer-pku.cn/index.html*) is a mature network using a standard processing pipeline, which is used to analyze expression data of RNA sequence. The web server provides differential expression analysis and survival analysis between tumor and normal samples with 9736 tumor samples and 8587 normal samples in TCGA and the GTEx projects [16]. Gene expression in individual stages of cancer (11 normal samples and 155 ESCA samples based on TCGA) was characterized in



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



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

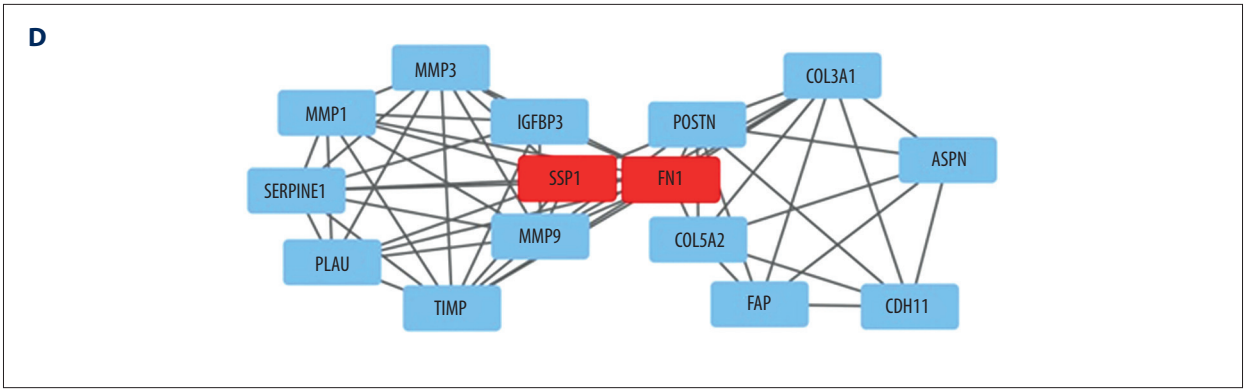


Figure 2. GO and KEGG pathway enrichment analysis of the 120 DEGs. (**A**) The significantly enriched GO terms, with P<0.001. (**B**) Gene networks identified through KEGG analysis of the differentially expressed genes, with P<0.05. (**C**) The protein-protein interaction network of 120 DEGs. (**D**) The top module from the protein-protein interaction network.

Table 2. KEGG pathway analysis of differentially expressed genes associated with esophageal cancer.

Category	Term	Count	%	P value	Genes	Fold Enrichment	FDR
KEGG_ PATHWAY	hsa05146: Amebiasis	8	6.67	7.19E-06	COL3A1, SERPINB2, SERPINB1, LAMC2, COL1A1, SERPINB13, COL5A2, FN1	10.64	8.11E-03
KEGG_ PATHWAY	hsa04512: ECM-receptor interaction	7	5.83	2.69E-05	COL3A1, COL6A3, LAMC2, COL1A1, COL5A2, SPP1, FN1	11.35	3.03E-02
KEGG_ PATHWAY	hsa04510: Focal adhesion	7	5.83	2.82E-03	COL3A1, COL6A3, LAMC2, COL1A1, COL5A2, SPP1, FN1	4.79	3.13E+00
KEGG_ PATHWAY	hsa04151: PI3K-Akt signaling pathway	8	6.67	9.14E-03	COL3A1, COL6A3, GYS2, LAMC2, COL1A1, COL5A2, SPP1, FN1	3.27	9.84E+00
KEGG_ PATHWAY	hsa04974: Protein digestion and absorption	4	3.33	2.29E-02	COL3A1, COL6A3, COL1A1, COL5A2	6.41	2.30E+01
KEGG_ PATHWAY	hsa05202: Transcriptional mis-regulation in cancer	5	4.17	2.86E-02	MMP9, MMP3, IGFBP3, HPGD, PLAU	4.20	2.79E+01

UALCAN (*http://ualcan.path.uab.edu/index.html*) [17]. To visualize the expression level, prognosis, and relationship, boxplots and disease-free survival curves were constructed.

Statistical analysis

GraphPad Prism 6.0 software was used to analyze data. Comparisons of the 2 groups were made using the *t* test, or with non-parametric test when the overall distribution was unclear. The difference was considered statistically significant when P values were less than 0.05.

Results

Identification of differentially expressed genes

To confirm possible biomarkers of esophageal cancer, we first screened the literature at first as shown in Figure 1A. We found 101 items of microarray-based esophageal cancer study. Among these items, 106 items did not meet our inclusion criteria and were eliminated. We selected 171 samples from 4 items (Table 1) with 87 tumor and 84 non-tumor samples. Finally, GSE92396, GSE20347, GSE23400, and GSE45168 were applied to discover differentially expressed genes in esophageal cancer. To study gene expression alteration related to esophageal tumor progression, the DEGs of these 4 datasets

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

Table 3. Gene ontology analysis of differentially expressed genes associated with esophageal cancer.

Category	Term	Count	%	P value	Genes	Fold enrichment	FDR
GOTERM_BP_DIRECT	GO: 0030198~ extracellular matrix organization	16	13.33	3.95E-12	PXDN, COL3A1, POSTN, SPINK5, COL5A2, BGN, ERO1A, SERPINE1, TGFBI, COL6A3, LAMC2, MFAP2, VCAN, COL1A1, SPP1, FN1	12.02	5.98E-09
GOTERM_BP_DIRECT	GO: 0030574~ collagen catabolic process	9	7.50	1.06E-08	MMP9, COL3A1, COL6A3, COL1A1, MMP3, COL5A2, MMP12, MMP1, MMP11	20.71	1.60E-05
GOTERM_BP_DIRECT	GO: 0022617~ extracellular matrix disassembly	9	7.50	4.22E-08	MMP9, LAMC2, MMP3, MMP12, MMP1, SPP1, MMP11, FN1, TIMP1	17.44	6.39E-05
GOTERM_BP_DIRECT	GO: 0007155~ cell adhesion	17	14.17	7.28E-08	PTPRK, SLURP1, POSTN, THY1, TNFAIP6, SORBS2, FAP, FAT1, COL6A3, TGFBI, LAMC2, VCAN, COL1A1, LOXL2, SPP1, FN1, CDH11	5.46	1.10E-04
GOTERM_BP_DIRECT	GO: 0030199~ collagen fibril organization	6	5.00	6.06E-06	COL3A1, COL1A1, LOXL2, SERPINH1, COL5A2, MMP11	22.66	9.18E-03
GOTERM_BP_DIRECT	GO: 0018149~ peptide cross-linking	6	5.00	2.10E-05	CRCT1, SPRR1A, COL3A1, TGM1, TGM3, FN1	17.68	3.18E-02
GOTERM_BP_DIRECT	GO: 0030336~ negative regulation of cell migration	7	5.83	4.34E-05	PTPRK, SLURP1, SULF1, SERPINE1, PTN, KANK1, THY1	10.85	6.57E-02
GOTERM_BP_DIRECT	GO: 0010951~ negative regulation of endopeptidase activity	7	5.83	1.67E-04	COL6A3, SERPINE1, SERPINB2, SERPINB1, SERPINB13, SERPINH1, TIMP1	8.52	2.52E-01
GOTERM_BP_DIRECT	GO: 0001501~ skeletal system development	7	5.83	3.27E-04	MMP9, COL3A1, POSTN, VCAN, COL1A1, COL5A2, CDH11	7.53	4.94E-01
GOTERM_BP_DIRECT	GO: 0006508~ proteolysis	12	10.00	5.51E-04	CLCA4, FAP, MMP9, TMPRSS11E, ENDOU, KLK12, MMP3, MMP12, PLAU, MMP1, MMP11, KLK13	3.54	8.31E-01
GOTERM_BP_DIRECT	GO: 0001525~ angiogenesis	8	6.67	7.78E-04	IL18, FAP, TGFBI, SERPINE1, RORA, ECM1, FN1, THY1	5.28	1.17E+00
GOTERM_CC_DIRECT	GO: 0005615~ extracellular space	37	30.83	8.29E-14	PXDN, SLURP1, MMP9, IL18, COL3A1, ENDOU, POSTN, MMP3, SERPINH1, TIMP1, FAP, COL6A3, TGFBI, SERPINE1, KLK12, PTN, SERPINB13, LOXL2, SPP1, KLK13, FN1, CRISP3, IL1RN, CST1, ECM1, ACPP, TNFAIP6, SULF1, NUCB2, SERPINB2, SERPINB1, LAMC2, VCAN, COL1A1, IGFBP3, PLAU, IL36A	4.21	9.93E-11
GOTERM_CC_DIRECT	GO: 0070062~ extracellular exosome	53	44.17	1.05E-13	PXDN, SLURP1, MMP9, IL18, CRABP2, SULT2B1, SPINK5, SLK, SERPINE1, TGFBI, KLK12, AHNAK, KLK13, CRISP3, ACADM, CLCA4, CRYAB, KRT13, GLTP, THY1, BGN, CLIC3, NUCB2, SERPINB1, PSCA, EPS8L1, ALDH9A1, ALOX12, CRNN, SERPINH1, TIMP1, KRT24, GPD1L, ADIRF, PPL, FAT1, TGM1, COL6A3, TGM3, SERPINB13, NDRG2, SCNN1B, TRIP10, FN1, SPP1, IL1RN, ECM1, UBL3, ACPP, IGFBP3, HPGD, PLAU, CDH11	2.89	1.25E-10

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

Table 3 continued. Gene ontology analysis of differentially expressed genes associated with esophageal cancer.

Category	Term	Count	%	P value	Genes	Fold enrichment	FDR
GOTERM_CC_DIRECT	GO: 0005578~ proteinaceous extracellular matrix	17	14.17	1.86E-11	ASPN, CRISP3, PXDN, MMP9, POSTN, MMP3, ECM1, COL5A2, MMP12, MMP1, TIMP1, MMP11, BGN, COL6A3, TGFBI, VCAN, FN1	9.71	2.23E-08
GOTERM_CC_DIRECT	GO: 0031012~ extracellular matrix	16	13.33	8.17E-10	ASPN, PXDN, COL3A1, POSTN, ECM1, COL5A2, MMP1, MMP11, BGN, SERPINE1, TGFBI, COL6A3, VCAN, COL1A1, LOXL2, FN1	8.28	9.78E-07
GOTERM_CC_DIRECT	GO: 0005576~ extracellular region	31	25.83	8.14E-08	SLURP1, MMP9, IL18, COL3A1, ENDOU, MMP3, SPINK5, MMP1, TIMP1, COL6A3, TGFBI, SERPINE1, SPP1, KLK13, FN1, CRISP3, EPHX3, COL5A2, ECM1, MMP12, MMP11, BGN, TMPRSS11E, SERPINB2, LAMC2, MFAP2, VCAN, COL1A1, IGFBP3, PLAU, IL36A	2.95	9.75E-05
GOTERM_CC_DIRECT	GO: 0005581~ collagen trimer	7	5.83	2.94E-05	COL3A1, COL6A3, COL1A1, SERPINH1, COL5A2, MMP1, TIMP1	11.65	3.52E-02
GOTERM_MF_DIRECT	GO: 0004867~ serine-type endopeptidase inhibitor activity	7	5.83	4.74E-05	COL6A3, SERPINE1, SERPINB2, SERPINB1, SERPINB13, SERPINH1, SPINK5	10.69	6.13E-02
GOTERM_MF_DIRECT	GO: 0004252~ serine-type endopeptidase activity	10	8.33	5.41E-05	FAP, MMP9, TMPRSS11E, KLK12, MMP3, MMP12, PLAU, MMP1, MMP11, KLK13	5.81	6.99E-02
GOTERM_MF_DIRECT	GO: 0002020~ protease binding	7	5.83	5.95E-05	FAP, SERPINE1, CST1, SERPINB13, ECM1, FN1, TIMP1	10.26	7.69E-02
GOTERM_MF_DIRECT	GO: 0005201~ extracellular matrix structural constituent	6	5.00	8.53E-05	PXDN, BGN, COL3A1, VCAN, COL1A1, COL5A2	13.26	1.10E-01
GOTERM_MF_DIRECT	GO: 0004222~ metalloendopeptidase activity	7	5.83	1.11E-04	CLCA4, FAP, MMP9, MMP3, MMP12, MMP1, MMP11	9.17	1.44E-01
GOTERM_MF_DIRECT	GO: 0005509~calcium ion binding	15	12.50	3.06E-04	ASPN, CRNN, MMP3, ITPR3, MMP1, MMP12, MMP11, FLG, FAT1, SULF1, NUCB2, TGM3, VCAN, EHD3, CDH11	3.10	3.95E-01
GOTERM_MF_DIRECT	GO: 0004175~ endopeptidase activity	5	4.17	4.65E-04	FAP, MMP9, MMP3, MMP12, MMP1	13.71	6.00E-01
GOTERM_MF_DIRECT	GO: 0005518~ collagen binding	5	4.17	6.95E-04	ASPN, MMP9, TGFBI, SERPINH1, FN1	12.34	8.95E-01

were then explored. The GEO2R was used to explore the DEGs with the cut-off criteria whose adjusted P value was less than 0.05 and |log fold-change| was not less than 1 after being normalized with the RMA algorithm. A total of 1577 genes, 1732 genes, 672 genes, and 1523 genes were identified as DEGs between tumor samples and normal samples after the analysis GSE92396, GSE20347, GSE23400, and GSE45168, respectively (Figure 1B–1E). Additionally, the common DEGs identified in the databases were visualized in Venn diagrams made by use of Jvenn (Figure 1F). There were 120 common DEGs in the esophageal tumor samples, which were considered in the development and progression of esophageal cancer (Figure 1F).

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

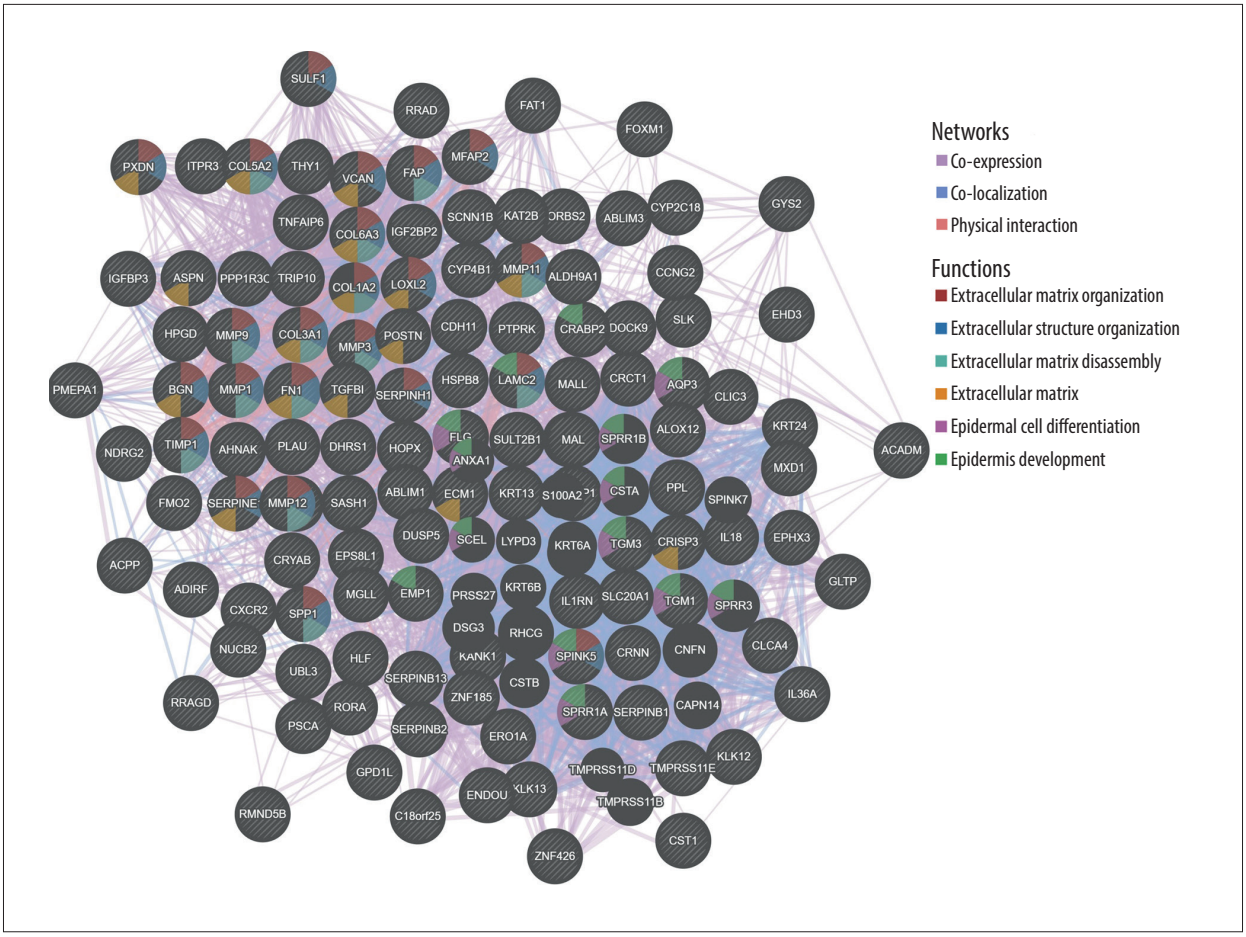


Figure 3. Functional enrichment of genes participates in pathways of ECM-receptor interaction and focal adhesion.

Functional enrichment of differentially expressed genes

To gain a deeper understanding of the biological roles of the 120 DEGs, DAVID was applied to analyze the GO and KEGG pathways. For 3 domains of GO, approximately half were related to the biological process category, and extracellular matrix organization and collagen catabolic process were the top 2 relevant biological processes. For cellular component, extracellular terms were mainly enriched. Regarding molecular function (MF), most of them were concentrated mainly in enzymatic activity, including serine-type endopeptidase inhibitor, serine-type endopeptidase, protease binding, and extracellular matrix structural constituent (Figure 2A, Table 2), and KEGG pathway analysis demonstrated that amebiasis and ECM-receptor interaction were 2 of the most enriched pathways (Figure 2B, Table 3), among which amebiasis was reported to show a strong association with chronic infection in esophagus tumors [18].

Furthermore, to identify the important genes related to esophageal cancer, an integrated PPI (protein-protein interaction) network was used. One hundred and twenty DEGs involved in interactions between each other and their KEGG pathway were demonstrated in a PPI network with STRING 10 (Figure 2C). The top module (score =8) that used the MCODE plug-in was selected to explore crucial modules in this PPI network (Figure 2D). ECM-receptor interaction and transcriptional mis-regulation in cancer were the main KEGG pathways (Figure 2D). We found that the common enriched KEGG pathways were amebiasis, ECM-receptor interaction, focal adhesion, PI3K-Akt signaling pathway, protein digestion and absorption, and transcriptional mis-regulation in cancer. These 120 genes enriched in skin development and extracellular matrix were categorized based on their functions using GeneMANIA (Figure 3). We found that the most commonly enriched function was ECM-receptor interaction.

SPP1 and FN1 were overexpressed in independent esophageal carcinomas

To further assess the altered expressions of DEGs, GSE13898 was used as a validation cohort including 75 esophageal tumor samples and 28 normal esophageal samples. Because both SPP1 and FN1 are related to the ECM-receptor interaction (Figure 3),

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

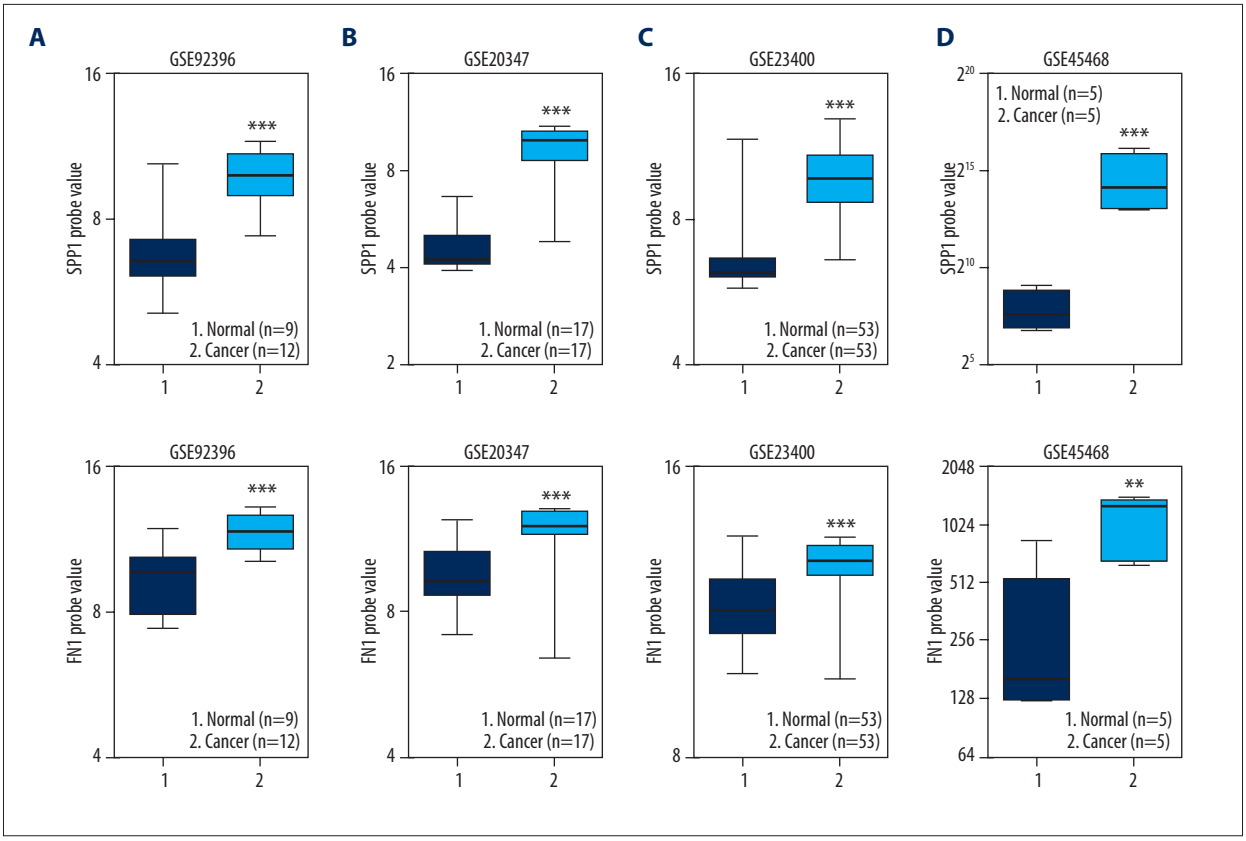


Figure 4. Differential expression of SPP1 and FN1 in the discovery datasets. (**A**) SPP1 and FN1 expression were significantly increased in ESCA tissues in GSE92396. (**B**) SPP1 and FN1 expression were remarkably increased in ESCA tissues in GSE20347. (**C**) SPP1 and FN1 expression were notably increased in ESCA tissues in GSE23400. (**D**) SPP1 and FN1 expression were significantly higher in ESCA than in normal tissues in GSE45168.

we selected these 2 genes for further study. SPP1 and FN1 expression levels were markedly increased in esophageal carcinoma tissues compared with normal samples in each of the 4 discovery datasets (Figure 4A–4D). Common to these studies, increased expression of SPP1 and FN1 was validated in another cohort with malignant samples (Figure 5A). Both revealed that the dysregulation of SPP1 and FN1 was associated with tumorigenesis of esophageal carcinoma. Additionally, we validated the overexpression of SPP1 and FN1 in 182 ESCA tissues and 286 normal tissue samples from TCGA and GTEx using the GEPIA web. The expression of SPP1 and FN1, which was in accordance with the discovery datasets, was significantly different between the tumor and normal samples (Figure 5B). These results revealed that both SPP1 and FN1 were overexpressed in ESCA. We further demonstrated that patients with enhanced expression of SPP1 and FN1 had a higher pathological stage (Figure 5C).

Gene set enrichment analysis for SPP1 and FN1 expression

GSEA was conducted to investigate the potential function and their correlations between SPP1 or FN1 and esophageal cancer. Proteasome and nucleotide excision repair were the 2 most differentially enriched in the SPP1 high-expression phenotype (Figure 6A). Figure 6B shows that ECM-receptor interaction, focal adhesion, cell adhesion molecules cams, TGF-beta signaling pathway, regulation of actin cytoskeleton, cytokine-cytokine receptor interaction, and glycosaminoglycan biosynthesis chondroitin sulfate were differentially enriched in the FN1 high-expression phenotype.

SPP1 and FN1 act as prognostic markers for patients with ESCA

We next assessed whether the overexpression of SPP1 and FN1 was connected with esophageal cancer patient survival. The Cancer Genome Atlas (TCGA) and the GTEx projects, which provided not only 182 ESCA patients' gene expression information but also clinical follow-up research, were utilized to evaluate disease-free survival (DFS). The GEPIA, a web server that analyzed gene expression data and clinical information from the TCGA and GTEx projects, was used to investigate the prognostic significance in esophageal carcinoma. We compared the survival curves with the P values count using the

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

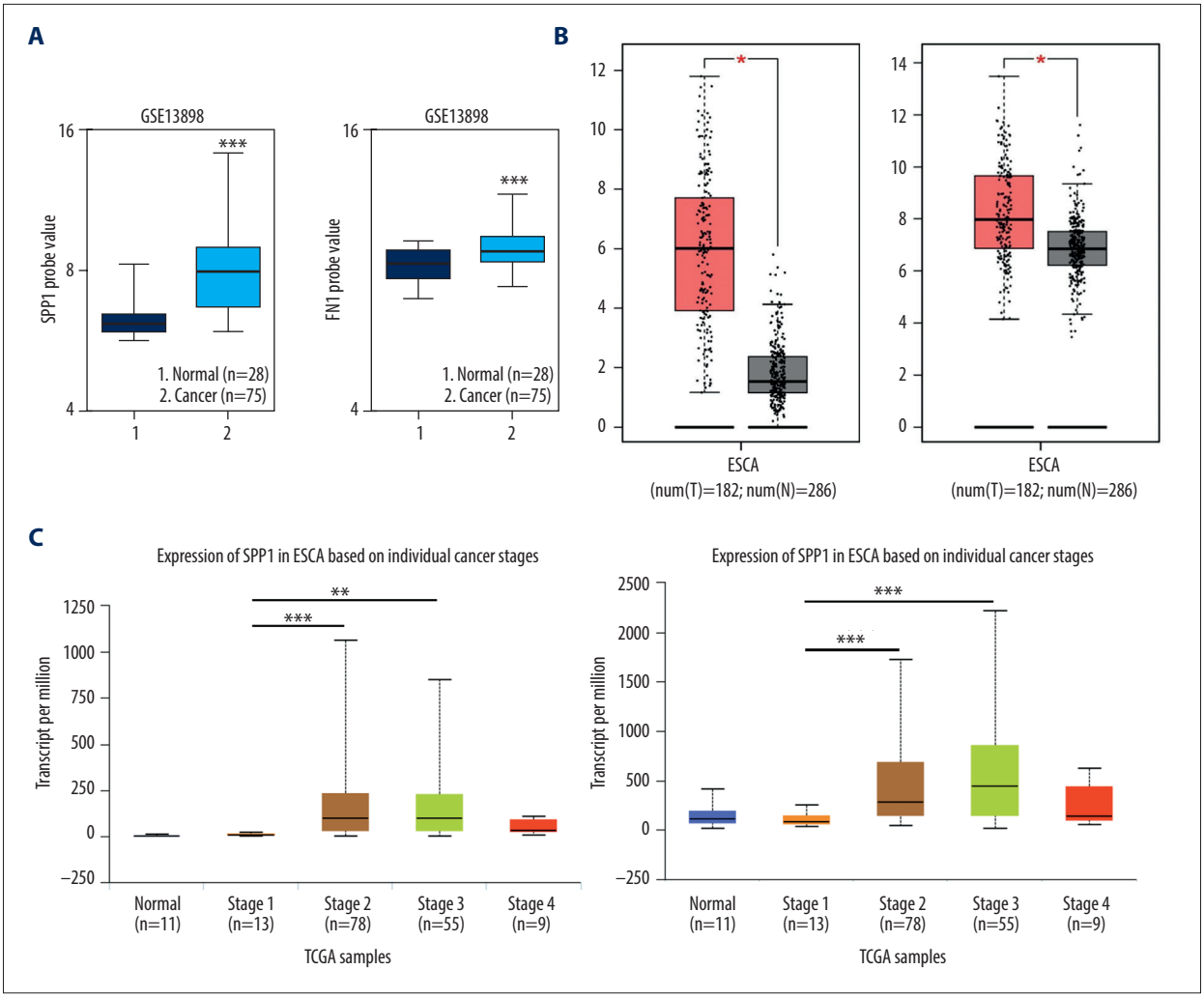


Figure 5. Validation of differential expression of SPP1 and FN1. (**A**) Elevated expression of SPP1 and FN1 in esophageal carcinoma. The differences between the 2 groups were determined by *t* tests (**the left**) or non-parametric test (**the right**). *** P<0.001. (**B**) Validation of expression of SPP1 and FN1 in 182 ESCA tissues and 286 normal tissue samples from the TCGA and the GTEx projects. (**C**) Gene expression of SPP1 and FN1 in ESCA based on individual cancer stages.

log-rank (Mantel-Cox) test. Plots of the Kaplan-Meier estimator showed that the overexpression of SPP1 and FN1 does not affect the overall survival of patients, while dysregulation of SPP1 and FN1 was remarkably connected with shorter disease-free survival (DFS) and earlier recurrence (P=0.001, HR: 2.2, Figure 7A; P=0.0064, HR: 1.9, Figure 7B). In short, these results suggest that overexpressed SPP1 and FN1 are prognostic markers for the prediction of ESCA relapse and survival.

Discussion

Our study determined signatures of gene expression in esophageal carcinoma with genome-wide profiles by integrating 4 esophageal cancer microarray datasets of 87 esophageal cancer samples from GEO. Our results suggest that alterations at the molecular level can identify esophageal cancer with high accuracy, in contrast to studies using only a single dataset. We analyzed and identified 112 common DEGs in all esophageal tumor samples. Functional enrichment analysis of the DEGs using KEGG demonstrated that these DEGs had carcinogenic features in ESCA, including ECM-receptor interaction, focal adhesion, and transcriptional mis-regulation in cancer. Two genes – SPP1 and FN1 – were selected and later corroborated with the valid dataset and 468 samples from TCGA and the GTEx projects. In our study, we found that pathways related to proteasome and nucleotide excision repair were the 2 most differentially enriched in SPP1 high-expression phenotype, and ECM-receptor interaction and focal adhesion were the 2 most differentially enriched in FN1 high-expression phenotype. We also found that SPP1 and FN1 are strongly associated with short DFS and early recurrence in ESCA.

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

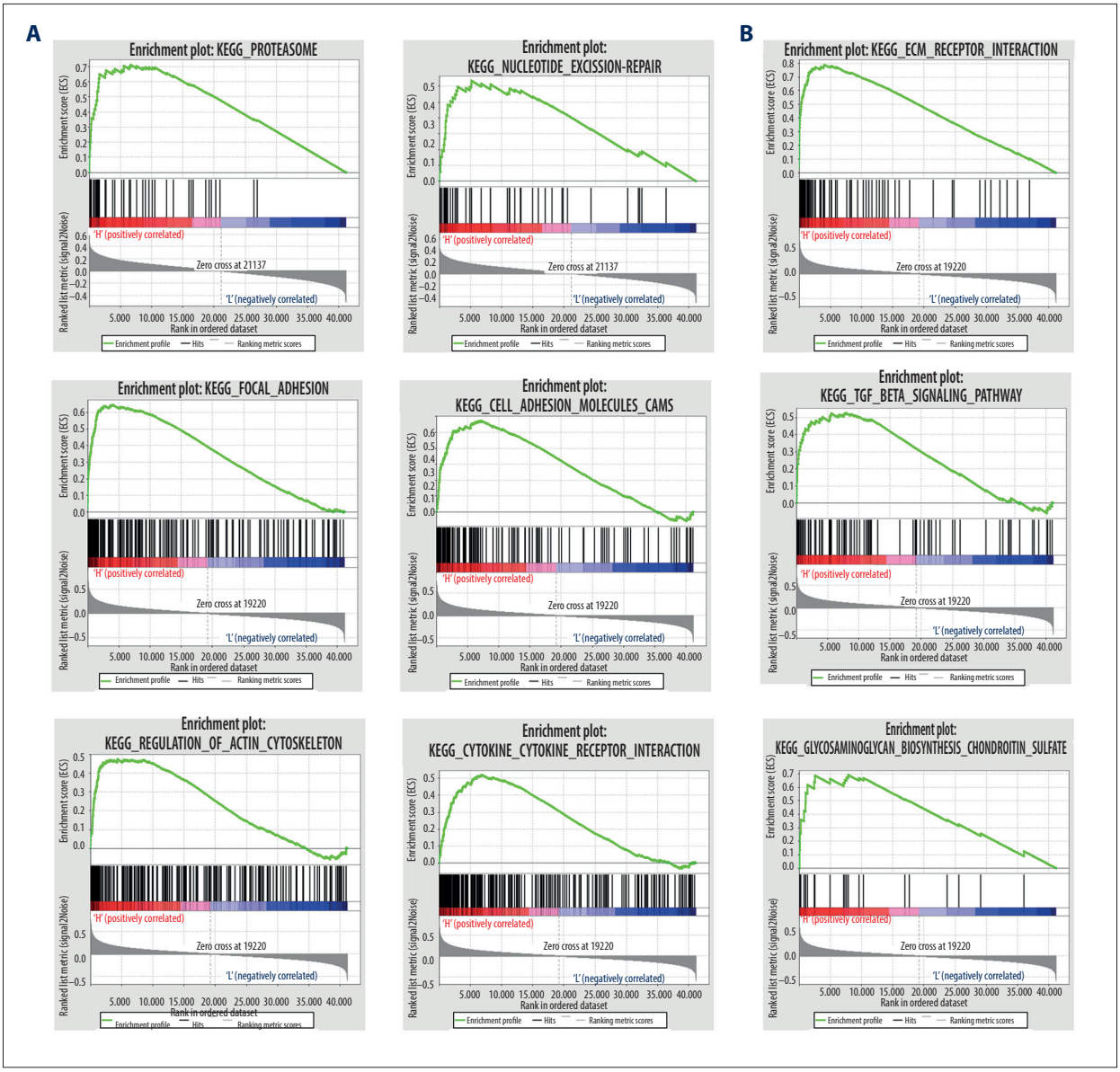


Figure 6. Gene set enrichment analysis for SPP1 and FN1 expression. (**A**) Gene set enrichment analysis for SPP1 expression; (**B**) Gene set enrichment analysis for FN1 expression.

ECM proteins, including collagen, proteoglycan, laminin, elastin, and fibronectin, are pre-conditions for tumor invasion and metastasis [19]. The reduced expressions of matrix metalloproteinase 2 (MMP2) promotes the invasion and migration in anoikis-resistant prostate cancer cells [20]. In breast cancer, low expression levels of a2V, which regulates breast tissue ECM stiffness, contributes to high risk of developing metastases in breast cancer patients [21]. Furthermore, as a key component that stabilizes the extracellular matrix (ECM) network, Nidogen-2 (NID2) is crucial to suppression of metastasis of esophageal squamous cell carcinoma [22]. The upregulation of LAMC2 contributes to the prognosis and metastasis of ESCC [23]. Amebiasis was recently reported in a patient diagnosed with colonic amebiasis and signet-ring cell carcinoma of the ileocecal valve [24]. In the African community, it was reported that 4 patients developed esophageal cancer from the chronic tract, and there was a strong association between chronic infection with gastrointestinal schistosomiasis and amebiasis and the likelihood of developing tumors [18]. Amebiasis can promote the development of esophageal cancer, and the interaction between amebiasis and esophageal cancer could be a new research focus. Our study demonstrated that amebiasis and the ECM-receptor interaction were the most significantly regulated pathways in ESCA.

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

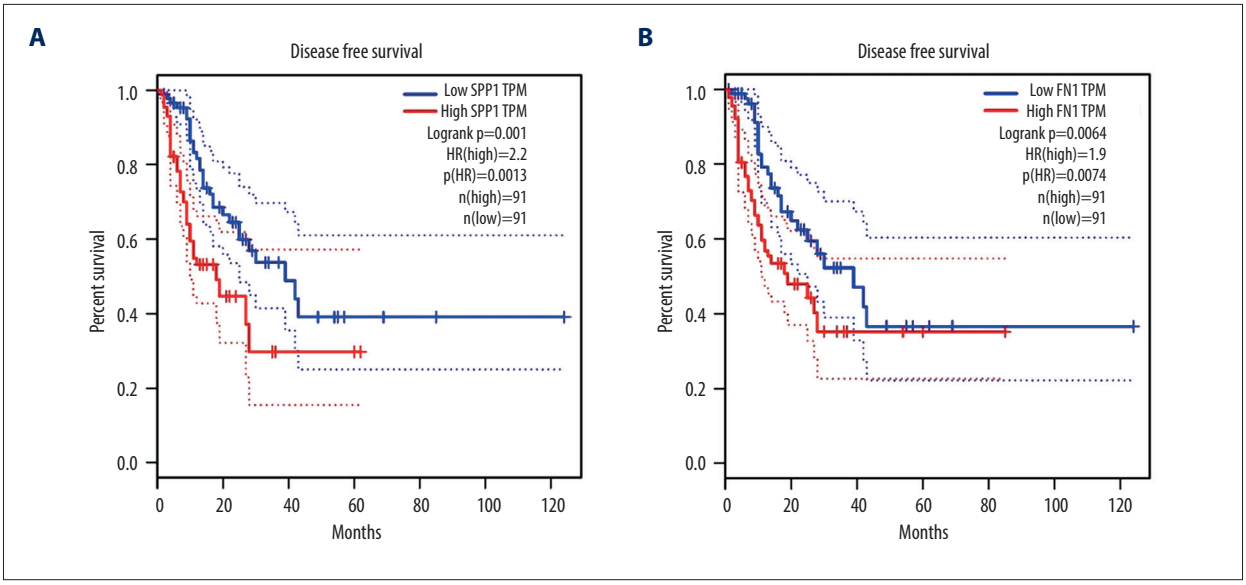


Figure 7. Kaplan-Meier survival curves by different levels of SPP1 and FN1 expression in 182 ESCA patients, and the correlation of their expression with immune infiltration level. (**A**) Disease-free survival (DFS) by low and high SPP1 expression; (**B**) Disease-free survival (DFS) by low and high FN1 expression.

SPP1 is rich in chemokine-like matrix phosphoglycoprotein; it is located on chromosome 4 (4q13) and contains 7 exons. In addition, SPP1 is a secreted non-collagen protein that affects cell proliferation, invasion, stem-like behavior, survival, and drug resistance. SPP1 is involved in the progression of invasion and metastasis of many cancers and has been considered a prognostic and diagnostic indicator for several cancers [25]. Additionally, SPP1 recruits potential meiotic double-strand break sites on the chromosomal axis [26]. SPP1 mediates the radioresistance of lung cancer with KRAS mutations [27], and the loss of SPP1 produces a microenvironment that promotes glioblastoma [28]. Furthermore, SPP1 is a potential novel molecular therapeutic biomarker in patients with rectal cancer [29]. Studies showed that isoforms of SPP1 promote metastasis of esophageal adenocarcinoma cells [30]. Although SPP1 overexpression can contribute to the development of esophageal squamous cell carcinoma [31], there is currently no data showing that SPP1 is related to proteasome and nucleotide excision repair. The development and progression mechanism(s) driven by SPP1 in ESCA need to be studied. PLCE1 proved to be a tumor promoter in our previous research [32]. Interestingly, we also found that there was a negative relationship between SPP1 and PLCE1 (Figure 8A). However, the mechanism by which SPP1 and PLCE1 are negatively related requires further exploration in ESCA, and we will further explore the relationship between them.

FN1, which plays vital roles in embryogenesis, differentiation, carcinogenesis, and metastasis [33], is reported to promote metastasis in ovarian cancer by activating MMP9 via the PI3K/Akt pathway [34]. It was reported that FN1 is associated with advanced stage in patients with renal cancer [35]. In breast cancer, the FN1 expression level is associated with prognosis [36]. In colon cancer cells, fibronectin is related to invasion ability [37]. It was found that a single-nucleotide polymorphism in FN1 affects tumor migration and metastasis of colon cancer [38]. In esophageal squamous cell carcinoma, FN1 promotes activation of the Erk pathway in the advanced stage [39], and overexpression of FN1 increases proliferation and survival [40]. We found that there was no obvious relationship between FN1 and PLCE1, but there was a marked positive relationship between SPP1 and FN1 (Figure 8B, 8C), but the molecular mechanism underlying their relationship has not yet been studied.

Although we identified 2 diagnostic biomarkers and their pathways in ESCA and our research might be the first to combine GEO, TCGA, ROC analysis, experimental validation, and bioinformatics, the Western blot and immunohistochemistry experiments are needed to provide stronger support for the conclusion. In fact, we are considering studying the specific mechanisms of 2 key molecules in further research. Studies have shown that loss of SPP1 produces a microenvironment that promotes glioblastoma [28], which stimulates our interests in studying the relationship between SPP1 and immune microenvironment in esophageal cancer. FN1 is related to the invasion of colon cancer cells [37], and FN1 knockdown experiments in ESCA cells would provide stronger support for the relationship between FN1 and invasion in ESCA. We believe that further research will provide a stronger theoretical basis for targeted therapy of esophageal cancer.

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

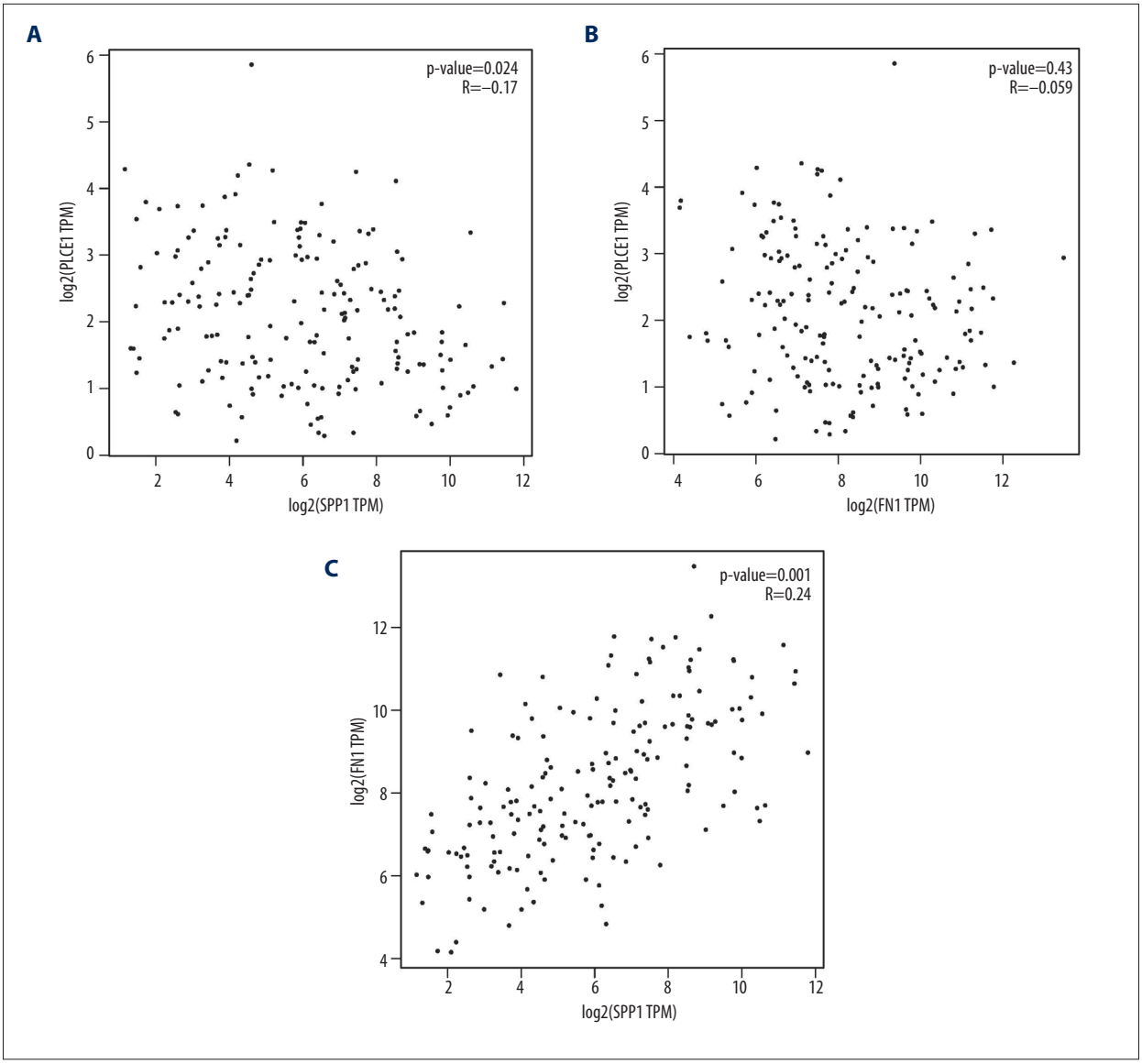


Figure 8. The correlation analysis of SPP1, FN1, and PLCE1. (**A**) SPP1 and PLCE1 are significantly negatively correlated. (**B**) FN1 and PLCE1 are significantly negatively correlated. (**C**) SPP1 and FN1 are significantly positively correlated.

Conclusions

We showed that SPP1 and FN1 are overexpressed in esophageal cancer tissues, enriched in proteasome, nucleotide excision repair, ECM-receptor interaction, and focal adhesion, and they indicate poorer survival rates and a higher risk of tumor relapse and are potential markers of prognosis in patients with ESCA. These findings showed the need for further investigation into the molecular mechanisms of SPP1 and FN1 and their functions as early targets in ESCA, as well as the relationship between SPP1 and FN1.

Conflicts of interest

None.

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

References:

1. Torre LA, Bray F, Siegel RL et al: Global cancer statistics, 2012. Cancer J Clin, 2015; 65(2): 87–108
2. Pennathur A, Gibson MK, Jobe BA, Luketich JD: Oesophageal carcinoma. Lancet, 2013; 381(9864): 400–12
3. Xu YW, Peng YH, Chen B et al: Autoantibodies as potential biomarkers for the early detection of esophageal squamous cell carcinoma. Am J Gastroenterol, 2014; 109(1): 36–45
4. Xie JJ, Xu LY, Xie YM et al: Roles of ezrin in the growth and invasiveness of esophageal squamous carcinoma cells. Int J Cancer, 2009; 124(11): 2549–58
5. Tong YS, Zhou XL, Wang XW et al: Association of decreased expression of long non-coding RNA LOC285194 with chemoradiotherapy resistance and poor prognosis in esophageal squamous cell carcinoma. J Transl Med, 2014; 12: 233
6. Zhu GZ, Yang YL, Zhang YJ et al: High expression of AHSP, EPB42, GYPC and HEMGN predicts favorable prognosis in FLT3-ITD-negative acute myeloid leukemia. Cell Physiol Biochem, 2017; 42(5): 1973–84
7. Kim YW, Koul D, Kim SH et al: Identification of prognostic gene signatures of glioblastoma: A study based on TCGA data analysis. Neurooncology, 2013; 15(7): 829–39
8. Davis S, Meltzer PS: GEOquery: A bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics, 2007; 23(14): 1846–47
9. Ito K, Murphy D: Application of ggplot2 to Pharmacometric Graphics. CPT Pharmacometrics Syst Pharmacol, 2013; 2: e79
10. Bardou P, Mariette J, Escudie F et al: jvenn: An interactive Venn diagram viewer. BMC Bioinformatics, 2014; 15: 293
11. Ashburner M, Ball CA, Blake JA et al: Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. Nat Genet, 2000; 25(1): 25–29
12. Huang da W, Sherman BT, Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc, 2009; 4(1): 44–57
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(Database issue): D447–52
14. Warde-Farley D, Donaldson SL, Comes O et al: The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res, 2010; 38(Web Server issue): W214–20
15. Wan Q, Tang J: Exploration of potential key pathways and genes in multiple ocular cancers through bioinformatics analysis. Graefes Arch Clin Exp Ophthalmol, 2019;2 57(10): 2329–41
16. Tang Z, Li C, Kang B et al: GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res, 2017; 45(W1): W98–102
17. Chandrashekar DS, Bashel B, Balasubramanya SAH et al: UALCAN: A portal for facilitating tumor subgroup gene expression and survival analyses. Neoplasia, 2017; 19(8): 649–58
18. Waku M, Napolitano L, Clementini E et al: Risk of cancer onset in sub-Saharan Africans affected with chronic gastrointestinal parasitic diseases. Int J Immunopathol Pharmacol, 2005; 18(3): 503–11
19. Verma S, Kesh K, Ganguly N et al: Matrix metalloproteinases and gastrointestinal cancers: Impacts of dietary antioxidants. World J Biol Chem, 2014; 5(3): 355–76
20. Zhang P, Song Y, Sun Y et al: AMPK/GSK3beta/beta-catenin cascade-triggered overexpression of CEMIP promotes migration and invasion in anoikis-resistant prostate cancer cells by enhancing metabolic reprogramming. FASEB J, 2018; 32(7): 3924–35
21. Katara GK, Kulshrestha A, Mao L et al: Mammary epithelium-specific inactivation of V-ATPase reduces stiffness of extracellular matrix and enhances metastasis of breast cancer. Mol Oncol, 2018; 12(2): 208–23
22. Chai AW, Cheung AK, Dai W et al: Metastasis-suppressing NID2, an epigenetically-silenced gene, in the pathogenesis of nasopharyngeal carcinoma and esophageal squamous cell carcinoma. Oncotarget, 2016; 7(48): 78859–71
23. Liang Y, Chen X, Wu Y et al: LncRNA CASC9 promotes esophageal squamous cell carcinoma metastasis through upregulating LAMC2 expression by interacting with the CREB-binding protein. Cell Death Differ, 2018; 25(11): 1980–95
24. Grosse A: Diagnosis of colonic amebiasis and coexisting signet-ring cell carcinoma in intestinal biopsy. World J Gastroenterol, 2016; 22(36): 8234–41
25. Shevde LA, Samant RS: Role of osteopontin in the pathophysiology of cancer. Matrix Biol, 2014, 37: 131–41
26. Acquaviva L, Szekvolgyi L, Dichtl B et al: The COMPASS subunit Spp1 links histone methylation to initiation of meiotic recombination. Science, 2013; 339(6116): 215–18
27. Wang M, Han J, Marcar L et al: Radiation resistance in KRAS-mutated lung cancer is enabled by stem-like properties mediated by an osteopontin-EGFR pathway. Cancer Res, 2017; 77(8): 2018–28
28. Szulzewsky F, Schwendinger N, Guneykaya D et al: Loss of host-derived osteopontin creates a glioblastoma-promoting microenvironment. Neurooncology, 2018; 20(3): 355–66
29. Huang RH, Quan YJ, Chen JH et al: Osteopontin promotes cell migration and invasion, and inhibits apoptosis and autophagy in colorectal cancer by activating the p38 MAPK signaling pathway. Cel Physiol Biochem, 2017; 41(5): 1851–64
30. Lin J, Myers AL, Wang Z et al: Osteopontin (OPN/SPP1) isoforms collectively enhance tumor cell invasion and dissemination in esophageal adenocarcinoma. Oncotarget, 2015; 6(26): 22239–57
31. Ito T, Hashimoto Y, Tanaka E et al: An inducible short-hairpin RNA vector against osteopontin reduces metastatic potential of human esophageal squamous cell carcinoma *in vitro* and *in vivo*. Clin Cancer Res, 2006; 12(4): 1308–16
32. Cui XB, Peng H, Li RR et al: MicroRNA-34a functions as a tumor suppressor by directly targeting oncogenic PLCE1 in Kazakh esophageal squamous cell carcinoma. Oncotarget, 2017; 8(54): 92454–69
33. Jerhammar F, Ceder R, Garvin S et al: Fibronectin 1 is a potential biomarker for radioresistance in head and neck squamous cell carcinoma. Cancer Biol Ther, 2010; 10(12): 1244–51
34. Balanis N, Wendt MK, Schiemann BJ et al: Epithelial to mesenchymal transition promotes breast cancer progression via a fibronectin-dependent STAT3 signaling pathway. J Biol Chem, 2013; 288(25): 17954–67
35. Waalkes S, Atschekzei F, Kramer MW et al: Fibronectin 1 mRNA expression correlates with advanced disease in renal cancer. BMC Cancer, 2010; 10: 503
36. Helleman J, Jansen MP, Ruigrok-Ritstier K et al: Association of an extracellular matrix gene cluster with breast cancer prognosis and endocrine therapy response. Clin Cancer Res, 2008; 14(17): 5555–64
37. Ding J, Li D, Wang X et al: Fibronectin promotes invasiveness and focal adhesion kinase tyrosine phosphorylation of human colon cancer cell. Hepatogastroenterology, 2008; 55(88): 2072–76
38. Kida H, Takano Y, Yamamoto K et al: A single nucleotide polymorphism in fibronectin 1 determines tumor shape in colorectal cancer. Oncol Rep, 2014; 32(2): 548–52
39. Zhang J, Zhi H, Zhou C et al: Up-regulation of fibronectin in oesophageal squamous cell carcinoma is associated with activation of the Erk pathway. J Pathol, 2005; 207(4): 402–9
40. Song G, Liu K, Yang X et al: SATB1 plays an oncogenic role in esophageal cancer by up-regulation of FN1 and PDGFRB. Oncotarget, 2017; 8(11): 17771–84

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