Identification of potential biomarkers associated with immune infiltration in the esophageal carcinoma tumor microenvironment

Keywords
biomarker, immune infiltration, esophageal carcinoma, differentially expressed genes, COL1A2

Zhicheng Wang\textsuperscript{1,2}, Meilin Chen\textsuperscript{1,2}, Yanbing Qiu\textsuperscript{1,2}, Yuqin Yang\textsuperscript{1,2}, Yumei Huang\textsuperscript{1,2}, Xiaoxu Li\textsuperscript{1,2} and Wenling Zhang\textsuperscript{1,2}

1 Department of Laboratory Medicine, The Third Xiangya Hospital, Central South University, Changsha, Hunan, People’s Republic of China
2 Department of Laboratory Medicine, Xiangya School of Medicine, Central South University, Changsha, Hunan, People’s Republic of China

Corresponding Author:
Wenling Zhang\textsuperscript{1,2}
Department of Laboratory Medicine, The Third Xiangya Hospital and Xiangya School of Medicine Hospital, Central South University, Changsha, Hunan, People’s Republic of China
Email address: zhangwenling73@126.com

Abstract
Tumor immune cell infiltration was significantly correlated with the progression and the effect of immunotherapy in cancers including esophageal carcinoma (ESCA). However, no biomarkers were identified which were associated with immune infiltration in ESCA. In the present study, a total of 128 common differentially expressed genes (DEGs) were identified between esophageal squamous cell carcinomas (ESCC) and esophageal adenocarcinomas (EAC). The results of gene ontology (GO) enrichment and Reactome pathway analysis displayed that the upregulated DEGs were mainly involved in the regulation of extracellular matrix (ECM), while the downregulated DEGs were mainly involved in the regulation of cornification and keratinocyte differentiation. The most significant module of upregulated DEGs was selected by Molecular Complex Detection (MCODE). Top 10 similar genes of COL1A2 were explored, then validation and the prognostic analysis of these genes displayed that COL1A2, COL1A1, COL3A1, ZNF469 and POSTN had the prognostic value which were upregulated in ESCA. The expressions of COL1A2 and its 4 similar genes were mainly correlated with infiltrating levels of macrophages and dendritic cells (DCs) and showed strong correlations with diverse immune marker sets in ESCA. To summarize, COL1A2 and its 4 similar genes were identified as the potential biomarkers associated with immune infiltration in ESCA. These genes might be applied to immunotherapy for ESCA.

Introduction
As a kind of gastrointestinal cancers, esophageal carcinoma (ESCA) ranks seventh according to incidence (572,000 new cases) and sixth in mortality totally (509,000 deaths) in 2018 [1]. The progression and chemoresistance of ESCA are important factors that impact patient survival. For
those patients with locally advanced ESCA, the five-year survival rate is about 20%. The 
standard treatment includes chemotherapy or chemoradiotherapy with surgery [2]. In recent 
years, tumor immunologic therapies have also shown promising early results in ESCA. And 
tumor immune cell infiltration significantly correlates with the progression and the effect of 
immunotherapy in cancers [3, 4]. However, the specific molecular mechanism about 
immunotherapy for ESCA yet remains unclear and some genes associated with immune 
infiltration are required to be found.

With the development of gene chip technology and sequencing technology, bioinformatics is 
becoming more and more important [5-7]. Gene Expression Omnibus (GEO) and The Cancer 
Genome Atlas (TCGA) are two open databases, which provide gene chip data and sequencing 
data about the cancers. A lot of studies were published for analyzing the data from GEO and 
TCGA in ESCA, including the profiles of long noncoding RNA (lncRNA), microRNA 
(miRNA), and messenger RNA (mRNA). Interleukin 8 (IL8) and C-X-C chemokine receptor 
type 7 (CXCR-7) were predicted to differentially express in ESCA tissues compared to normal 
esophageal tissues and significantly correlate to the prognosis of patients with ESCA [8]. Liu et 
al constructed the network of lncRNA, miRNA, and mRNA and found 8 lncRNAs including 
WDFY3-AS2, CASC8, UGDH-AS1, RAP2C-AS1, AC007128.1, AC016205.1, AC092803.2, 
and AC079949.2 were correlated with overall survival of the patients with ESCA [9]. Zeng et al 
built the regulatory networks of miRNA-TF and miRNA-gene and identified the top five miRNA 
(miRNA-93, miRNA-21, miRNA-4746, miRNA-196a-1, and miRNA-196a-2) with diagnostic 
value [10]. However, the immunotherapy related genes were certainly not identified in these 
studies in ESCA.

In the present study, the common differentially expressed genes (DEGs) were screened 
between ESCC and EAC by analyzing the datasets from GEO. Then the gene ontology (GO) 
analysis, Reactome pathway analysis, and protein-protein interaction (PPI) analysis of common 
DEGs were performed. Subsequently, COL1A2 with its similar genes were selected in GEPIA 
and validated in UALCAN, meanwhile the prognostic significance analysis was performed. 
Finally, the immune infiltration analysis was carried out to explain the immunological 
significance of COL1A2 with its similar genes. The findings in this report revealed the important 
role of COL1A2 with its similar genes in regulating immune infiltration and suggested these 
genes might be some immunotherapy related genes in ESCA.

Materials & Methods

Selection of microarray data from GEO

Two microarray datasets (GSE20347 and GSE26886) were downloaded from Gene Expression 
Omnibus (GEO). GSE20347 was based on the GPL571 platform (Affymetrix Human Genome 
U133A 2.0 Array) which contained 34 samples, including 17 pairs of tumors and their matching 
normal adjacent tissues from esophageal squamous cell carcinomas (ESCC) patients. GSE26886 
was based on the GPL570 platform (Affymetrix Human Genome U133 Plus 2.0 Array) which 
contained 40 samples, including 21 esophageal adenocarcinomas (EAC) samples and 19 normal 
esophageal squamous epithelium samples.
Screening of common DEGs between ESCC and EAC

We downloaded the raw CEL data and standardized them by the Affy package of R language [11]. Then the differential gene analysis was carried out using the limma R package and genes with an adjusted P-value < 0.01 and |log2 fold change (FC)| > 2 were considered DEGs [12]. The heatmaps and volcano plots of DEGs were displayed. The common DEGs of two datasets were integrated using an online web tool, Venny 2.1, and the common DEGs including upregulated and downregulated genes were saved for subsequent analysis.

GO and Reactome pathway analysis of common DEGs

The GO PANTHER version 14.0 was used to perform GO functional and Reactome pathway analysis of the common DEGs [13]. And the GO functional analysis of integrated DEGs refers to three parts: biological process (BP), cell component (CC), and molecular function (MF). FDR < 0.05 was considered to have a statistically significant difference.

PPI network and module analysis

We constructed a protein-protein interaction (PPI) network of the common DEGs using the Search Tool for the Retrieval of Interacting Genes (STRING), an online web tool for predicting the correlation of protein function [14]. And the PPI network was visualized by Cytoscape software [15]. Then, the most significant module of upregulated DEGs was selected by Molecular Complex Detection (MCODE), which was a plug-in application of Cytoscape [16].

Validation of COL1A2 and its similar genes in UALCAN

UALCAN can analyze cancer OMICS data which is a comprehensive, user-friendly, and interactive web tool [17]. The expressions of COL1A2 and its similar genes were validated in primary esophageal carcinoma samples compared with normal esophageal squamous epithelium samples. P < 0.05 was considered to indicate a statistically significant difference.

Similar genes and prognostic significance analysis in GEPIA

Gene Expression Profiling Interactive Analysis (GEPIA) was a web-based tool to deliver fast and customizable functionalities based on TCGA and GTEx data. Its primary functions included differential expression analysis, patient survival analysis, similar gene detection, and so on [18]. Similar genes and prognostic significance analysis were executed here to study our target genes and P < 0.05 was considered to have a statistically significant difference.

Immune infiltration analysis in TIMER

Tumor Immune Estimation Resource (TIMER) was an interactive web tool to comprehensively investigate the molecular characterization of tumor-immune interactions. 6 major functional modules allow users to interactively explore the correlations between immune infiltrates and a wide range of factors, including gene expression, somatic mutations, and somatic copy number.
alterations, and so on [19]. The correlations between the expression levels of COL1A2 with its similar genes and the infiltration levels of various immune cells were analyzed in ESCA, including B cells, CD8+ T cells, CD4+ T cells, neutrophils, macrophages, and dendritic cells. And we also analyzed the relationships between the expression levels of COL1A2 with its similar genes and the immune marker sets including CD8+ T cell, T cell (general), B cell, monocytes, tumor-associated macrophages (TAMs), M1 macrophages, M2 macrophages, neutrophils, natural killer (NK) cells, dendritic cells (DCs), T-helper (Th) cells, T-helper 17 (Th17) cells, follicular helper T (Tfh) cells, Tregs and exhausted T cells based on previous studies [20-22].

**Results**

**Identification of common DEGs between ESCC and EAC**

To find the Common DEGs from the two subtypes of ESCA, we analyzed the ESCC chip expression dataset GSE20347 and EAC chip expression dataset GSE26886 (adjusted \( P < 0.01 \) and \(|\text{log2fold change (FC)}| > 2\)). The GSE20347 dataset contained 225 differentially expressed genes, including 61 upregulated genes and 164 downregulated genes (Fig. 1A,1C). The GSE26886 dataset contained 691 differentially expressed genes, including 296 upregulated genes and 395 downregulated genes (Fig. 1B,1D). The common upregulated differentially expressed genes and downregulated differentially expressed genes from the two kinds of datasets were identified respectively, which were displayed by the Venn diagram, including 19 upregulated genes and 109 downregulated genes (Fig. 1E,1F).

**Figure 1.** Identification of differentially expressed genes (DEGs) in GSE20347 and GSE26886. Heatmaps of DEGs in (A) GSE20347 and (B) GSE26886 were displayed. Volcano plots of DEGs in (C) GSE20347 and (D) GSE26886 were displayed. Orange: upregulated DEGs; Blue: downregulated DEGs; N: normal; T: tumor. (E) Upregulated DEGs and (F) downregulated DEGs were integrated between GSE20347 and GSE26886 by using the Venn diagram.

**GO and Reactome pathway analysis of common DEGs**

DEGs gene ontology (GO) and Reactome pathway analysis were performed with PANTHER. The data shown in the figure needed to meet the following criteria: 1) the FDR < 0.05; 2) the top five significant enrichment results. We can find the biological process (BP) of upregulated genes which were mainly enriched in the extracellular structure organization, extracellular matrix organization and collagen fibril organization. The downregulated genes were mainly enriched in the cornification, epidermis development, keratinocyte differentiation, epidermal cell differentiation and keratinization (Fig. 2A). And the cellular component (CC) of upregulated genes were mainly enriched in the complex of collagen trimers, extracellular matrix, collagen trimer, banded collagen fibril and fibrillar collagen trimer. The downregulated genes were mainly enriched in the cornified envelope, extracellular region, vesicle, extracellular organelle, and extracellular space (Fig. 2B). The molecular function (MF) of upregulated genes were mainly enriched in the extracellular matrix structural constituent, extracellular matrix...
structural constituent conferring tensile strength and structural molecule activity, but the results of downregulated genes were not displayed because the FDR > 0.05 (Fig. 2C). As for the pathway analysis, the upregulated DEGs were significantly enriched in the ECM proteoglycans, collagen degradation, extracellular matrix organization, degradation of the extracellular matrix, and integrin cell surface interactions, while the downregulated DEGs were significantly enriched in the formation of the cornified envelope, keratinization and developmental biology (Fig. 2D).

**Figure 2.** Gene ontology (GO) and Reactome pathway analysis of differentially expressed genes (DEGs). (A) BP, (B) CC, (C) MF and (D) Reactome pathway enrichment analysis were displayed by the Bar Chart. The bars of upregulated DEGs enrichment results were marked in orange and the bars of downregulated DEGs enrichment results were marked in blue.

**Construction of PPI network and module analysis**

The PPI network of DEGs was constructed by using the Cytoscape, which contained 81 nodes and 174 edges (Fig. 3A). The most significant module of upregulated DEGs was identified by using MCODE, which contained COL1A2, COL5A2, COL4A2, LUM and COL4A1 (Fig. 3B).

**Figure 3.** Construction of protein-protein interaction (PPI) network of differentially expressed genes (DEGs) and identification of the most significant module of upregulated DEGs. (A) PPI network of DEGs. The upregulated genes were marked in orange and the downregulated genes were marked in blue. (B) The most significant module of upregulated DEGs. The upregulated genes were marked in orange.

**Prognostic significance analysis and validation of COL1A2 and its similar genes**

COL1A2 was a subtype of Type I collagen, the pieces of evidence suggest its messenger RNA expression in ESCC tumors was higher than that in normal tissues. And it was associated with the survival of ESCC patients [23]. In the following study, we selected COL1A2 for our important target gene. In the first place, we explored the top 10 similar genes of COL1A2 in ESCA by using GEPIA (Table 1). In order to validate our results, we analyzed the differential expression of COL1A2 and its similar genes between 11 normal esophageal squamous epithelium samples and 184 primary esophageal carcinoma samples by using UALCAN (Fig. 4A-4K). except for CTSK, SPARC, and CLEC11A, all of the genes expressed higher in primary esophageal carcinoma samples compared with normal esophageal squamous epithelium samples. Meanwhile, we investigated whether these genes contributed to the prognosis of ESCA patients, GEPIA was used to analyze the disease-free survival (DFS) of COL1A2 and its similar genes in patients. As shown in Fig. 5A-5K, the higher expression of COL1A2 and its similar genes showed worse disease-free survival in ESCA patients, except for COL5A1, LRRRC15, ADAM12
and CLEC11A. These genes that have been validated and had prognostic value will be used for
later study.

Table 1. The similar genes of COL1A2 explored by GEPIA and ranked by Pearson correlation
coefficient (PCC) in ESCA tumors.

| Genes symbol | Gene ID                       | PCC  |
|--------------|------------------------------|------|
| COL1A1       | ENSG00000108821.13           | 0.98 |
| COL3A1       | ENSG00000168542.12           | 0.97 |
| SPARC        | ENSG00000113140.10           | 0.93 |
| ZNF469       | ENSG00000225614.2            | 0.92 |
| POSTN        | ENSG00000133110.14           | 0.92 |
| CTSK         | ENSG00000143387.12           | 0.92 |
| COL5A1       | ENSG00000130635.15           | 0.91 |
| LRRC15       | ENSG00000172061.8            | 0.91 |
| ADAM12       | ENSG00000148848.14           | 0.90 |

Figure 4. Boxplots of mRNA expression levels of COL1A2 and its similar genes in normal
esophageal squamous epithelium samples and primary esophageal carcinoma samples.
(A) COL1A2 (B) COL5A1 (C) POSTN (D) LRRC15 (E) CTSK (F) SPARC (G) COL3A1 (H) ZNF469 (I) COL1A1 (J) CLEC11A (K) ADAM12. Normal, normal esophageal squamous
epithelium samples; Primary tumor, primary esophageal carcinoma samples. *Normal vs
Primary tumor and P < 0.05. P < 0.05 was considered statistically significant.

Figure 5. Disease-free survival curves of COL1A2 and its similar genes in esophageal
carcinoma from the TCGA database.
(A) COL1A2 (B) COL5A1 (C) POSTN (D) LRRC15 (E) CTSK (F) SPARC
(G) COL3A1 (H) ZNF469 (I) COL1A1 (J) CLEC11A (K) ADAM12. P < 0.05 was considered
statistically significant.

The expression levels of COL1A2 and its similar genes correlate with the infiltration levels
of immune cells in esophageal carcinoma
The infiltration of immune cells such as TAMs and Tregs in the microenvironment can promote
esophageal carcinogenesis [24]. Therefore, we explored the relationship between the expression
of COL1A2 with its similar genes and the infiltrating immune cells in ESCA using the TIMER
database. Tumor purity plays an important role in the analysis of immune cell infiltration about
clinical tumor samples by genomic approaches [25]. After adjusting the analysis by tumor purity,
we found that the expression levels of COL1A2, COL1A1, COL3A1, ZNF469 and POSTN were significantly correlated with tumor purity in ESCA. And the expression levels of these genes were positively correlated with the infiltration levels of macrophages and dendritic cells, as well as the expression levels of ZNF469 was also positively correlated with the infiltration levels of CD4+ T cells (Fig. 6).

**Figure 6.** Correlation between the expression levels of COL1A2 with its similar genes and immune infiltration level in esophageal carcinoma (ESCA).

(A) COL1A1 (B) ZNF469 (C) POSTN (D) COL3A1 (E) COL1A2. P < 0.05 was considered statistically significant.

**Correlation analysis between mRNA levels of COL1A2 with its similar genes and immune marker sets**

To investigate the relationship between COL1A2 with its similar genes and the different immune infiltrating cells, the correlations between COL1A2 with its similar genes and immune marker sets of various immune cells of ESCA were explored in the TIMER database. We analyzed the correlations between the expression of COL1A2 with its similar genes and immune marker genes of diverse immune cells, including CD8+ T cells, T cells (general), B cells, monocytes, TAMs, M1 and M2 macrophages, neutrophils, NK cells and DCs in ESCA. Besides, the diverse T cells, for instance, Th1 cells, Th2 cells, Thf cells, Th17 cells, Tregs, and exhausted T cells were also analyzed. Before the correlation analysis, we adjusted it for tumor purity because the tumor purity of clinical samples can impact the immune infiltration analysis. As shown in Table 2, the expression of COL1A2 was significantly correlated with the expression of marker genes from tumor-infiltrating B cells (CD19 and CD79A), monocytes (CD86 and CD115), TAMs (CCL2 and IL10), M1 macrophages (COX2), neutrophils (CD66b, CD11b, and CCR7), DCs (BDCA-4), T-helper (STAT4, GATA3, STAT6, BCL6, and IL17A), Treg (FOXP3, CCR8, and TGFβ) and exhausted T cells (CTLA4 and TIM-3). The expression of COL1A1 was significantly correlated with the expression of marker genes from tumor-infiltrating B cells (CD8B), B cells (CD19 and CD79A), monocytes (CD86 and CD115), TAMs (CCL2 and IL10), M1 macrophages (COX2), neutrophils (CD66b, CD11b, and CCR7), DCs (HLA-DPA1), T-helper (STAT4, STAT5A, STAT6, BCL6, and IL17A), Treg (FOXP3, CCR8, and TGFβ) and exhausted T cells (CTLA4 and TIM-3). The expression of COL3A1 was significantly correlated with the expression of marker genes from tumor-infiltrating B cells (CD19 and CD79A), monocytes (CD86 and CD115), M1 macrophages (COX2), neutrophils (CD66b and CD11b), DCs (HLA-DQb1 and HLA-DPA1), T-helper (STAT4, GATA3, STAT6, BCL6, and IL17A), Treg (FOXP3, CCR8, and TGFβ) and exhausted T cells (CTLA4 and TIM-3). The expression of ZNF469 was significantly correlated with the expression of marker genes from tumor-infiltrating B cells (CD19 and CD79A), monocytes (CD86 and CD115), neutrophils (CD66b and CD11b), T-helper (STAT4, STAT6, STAT5A, BCL6, and IL17A), Treg (FOXP3, CCR8, and TGFβ) and exhausted T cells (CTLA4 and TIM-3). The expression of POSTN was significantly correlated with the expression of
marker genes from tumor-infiltrating B cells (CD19 and CD79A), monocytes (CD86 and CD115), TAMs (CD68), M1 macrophages (COX2), neutrophils (CD66b, CD11b, and CCR7), T-helper (STAT4, GATA3, STAT6, STAT5A, BCL6, and IL17A), Treg (FOXP3, CCR8, and TGFβ) and exhausted T cells (CTLA4 and TIM-3). All of the results demonstrated that COL1A2 with its similar genes were correlated with the infiltration of the immune cells in esophageal carcinoma.

Table 2. Correlation analysis between COL1A2 with its similar genes and related genes and markers of immune cells in TIMER.

| Description     | Gene markers | COL1A2 | COL1A1 | COL3A1 | ZNF469 | POSTN |
|-----------------|--------------|--------|--------|--------|--------|-------|
|                  |              | Cor    | P      | P      | P      | P     |
| CD8+ T cell     | CD8A         | -0.070 | 0.353  | 0.127  | 0.090  | 0.055 |
|                 | CD8B         | -0.133 | 0.076  | -0.197 | **     | -0.112|
| T cell (general)| CD3D         | -0.063 | 0.399  | -0.116 | 0.122  | 0.045 |
|                 | CD3E         | -0.042 | 0.573  | -0.094 | 0.207  | 0.025 |
|                 | CD2          | -0.048 | 0.525  | -0.101 | 0.178  | 0.031 |
| B cell          | CD19         | -0.156 | *      | -0.231 | **     | -0.157|
|                 | CD79A        | -0.181 | *      | -0.259 | ***    | -0.153|
| Monocyte        | CD86         | 0.447  | ***    | 0.405  | ***    | 0.455 |
|                 | CD115        | 0.408  | ***    | 0.333  | ***    | 0.431 |
| TAM             | CCL2         | 0.482  | ***    | 0.419  | ***    | 0.202 |
|                 | CD68         | -0.003 | 0.972  | -0.018 | 0.812  | 0.203 |
|                 | IL10         | 0.387  | ***    | 0.351  | ***    | 0.215 |
| M1 Macrophage   | INOS         | 0.223  | 0.058  | 0.128  | 0.280  | 0.211 |
|                 | IRF5         | -0.027 | 0.822  | -0.093 | 0.433  | -0.147|
|                 | COX2         | 0.382  | ***    | 0.369  | **     | 0.413 |
| M2 Macrophage   | CD163        | 0.140  | 0.237  | 0.039  | 0.740  | 0.095 |
|                 | VSIG4        | 0.163  | 0.168  | 0.093  | 0.434  | 0.090 |
|                 | MS4A4A       | 0.111  | 0.348  | -0.005 | 0.967  | 0.055 |
| Neutrophils | CD66b | -0.195 | ** | -0.209 | ** | -0.185 | * | -0.202 | ** | -0.200 | ** |
| CD11b      | 0.324 | ***    | 0.244 | ***    | 0.335 | ***    | 0.302 | ***    | 0.305 | ***    |
| CCR7       | -0.154 | *     | -0.241 | **    | -0.134 | 0.072 | -0.124 | 0.098 | -0.161 | *     |
| Natural killer | KIR2DL1   | 0.016 | 0.828 | 0.021 | 0.777 | 0.019 | 0.805 | -0.006 | 0.939 | -0.035 | 0.640 |
|            | KIR2DL3   | -0.036 | 0.629 | -0.031 | 0.682 | -0.022 | 0.768 | -0.003 | 0.969 | 0.001 | 0.990 |
|            | KIR2DL4   | -0.115 | 0.125 | -0.121 | 0.106 | -0.105 | 0.160 | -0.079 | 0.293 | -0.105 | 0.162 |
|            | KIR3DL1   | -0.010 | 0.898 | -0.023 | 0.754 | -0.022 | 0.769 | 0.048 | 0.524 | -0.067 | 0.372 |
|            | KIR3DL2   | 0.006 | 0.938 | 0.017 | 0.817 | 0.002 | 0.978 | 0.005 | 0.949 | -0.005 | 0.946 |
|            | KIR3DL3   | -0.034 | 0.654 | -0.034 | 0.651 | -0.023 | 0.762 | -0.051 | 0.494 | -0.048 | 0.526 |
|            | KIR2DS4   | -0.003 | 0.971 | -0.012 | 0.869 | -0.013 | 0.868 | 0.022 | 0.765 | -0.049 | 0.516 |
| Dendritic cell | HLA-DPB1   | -0.045 | 0.708 | -0.188 | 0.111 | -0.165 | 0.163 | -0.081 | 0.495 | -0.041 | 0.731 |
|            | HLA-DQB1  | -0.152 | 0.199 | -0.222 | 0.059 | -0.254 | *     | -0.108 | 0.365 | 0.057 | 0.635 |
|            | HLA-DRA   | -0.028 | 0.815 | -0.138 | 0.244 | -0.104 | 0.380 | -0.093 | 0.436 | -0.057 | 0.633 |
|            | HLA-DPA1  | -0.115 | 0.332 | -0.246 | *     | -0.234 | *     | -0.161 | 0.173 | 0.012 | 0.918 |
|            | BDCA-1    | 0.058 | 0.625 | -0.155 | 0.191 | 0.031 | 0.792 | -0.084 | 0.480 | 0.209 | 0.077 |
|            | BDCA-4    | 0.245 | *     | 0.185 | 0.117 | 0.186 | 0.116 | 0.134 | 0.260 | 0.058 | 0.626 |
|            | CD11c     | 0.027 | 0.820 | -0.055 | 0.645 | -0.048 | 0.689 | 0.023 | 0.848 | -0.047 | 0.692 |
| Th1        | T-bet     | -0.028 | 0.708 | -0.083 | 0.269 | -0.013 | 0.862 | 0.023 | 0.755 | -0.005 | 0.952 |
|            | STAT4     | 0.238 | **    | 0.172 | *     | 0.276 | ***    | 0.238 | *     | 0.292 | ***    |
|            | STAT1     | -0.005 | 0.947 | -0.020 | 0.794 | 0.016 | 0.834 | 0.041 | 0.587 | 0.048 | 0.522 |
|            | IFN-γ     | 0.017 | 0.817 | -0.001 | 0.990 | 0.024 | 0.745 | 0.061 | 0.415 | 0.008 | 0.918 |
|            | TNF-α     | 0.041 | 0.581 | 0.063 | 0.400 | 0.036 | 0.636 | 0.112 | 0.133 | 0.018 | 0.811 |
| Th2        | GATA3     | 0.154 | *     | 0.140 | 0.061 | 0.158 | *     | 0.126 | 0.092 | 0.154 | *     |
|            | STAT6     | -0.311 | ***    | -0.298 | ***    | -0.298 | ***    | -0.270 | ***    | -0.282 | ***    |
|            | STAT5A    | -0.139 | 0.063 | -0.171 | *     | -0.114 | 0.127 | -0.160 | *     | -0.154 | *     |
|            | IL13      | 0.091 | 0.227 | 0.047 | 0.530 | 0.109 | 0.143 | 0.076 | 0.309 | 0.055 | 0.462 |
| Tfh        | BCL6      | 0.288 | ***    | 0.280 | ***    | 0.282 | ***    | 0.344 | ***    | 0.294 | ***    |
| Abbreviations: TAM, tumor-associated macrophage; Th, T helper cell; Tfh, Follicular helper T cell; Treg, regulatory T cell; Cor, R value of Spearman’s correlation; Purity, correlation adjusted by purity.

**Discussion**

Over the past decade, there are many studies about bioinformatics analysis that have been published to explore the potential pathological mechanisms in ESCA. However, the mechanism of immune cell infiltration and the genes associated with immune infiltration in ESCA yet remain unclear. We firstly analyzed two subtypes of ESCA respectively, which contained the datasets of ESCC and EAC from GEO to find the common DEGs from the different histologic subtypes. Altogether 128 DEGs including 19 upregulated genes and 109 downregulated genes were screened. To demonstrate the functional mechanism of common DEGs in ESCA, we performed GO enrichment and Reactome pathway analysis. The results of GO enrichment and Reactome pathway analysis displayed that the upregulated DEGs were mainly involved in the regulation of extracellular matrix (ECM) such as extracellular structure organization and extracellular matrix organization, while the downregulated DEGs were mainly involved in the regulation of cornification and keratinocyte differentiation. These results suggest the major variations in ESCA were involved in the remodeling of the tumor microenvironment (TME) and cell differentiation. The TME is an internal and complex environment in the tumor, which includes all kinds of ECM (for instance, collagen and laminin) and stromal cells including fibroblasts, immune cells, and vascular endothelial cells [26, 27]. During tumor progression, the TME interacts with tumor cells depending on different kinds of immune cells (for example,
dendritic cells, macrophages, and myeloid-derived suppressor cells) to mediate their immunologic tolerance, subsequently impacting the clinical effect of immunotherapy [28, 29].

Aiming to find the most significant module involved in immune infiltration, we constructed the PPI network of DEGs and identified the most significant module of upregulated DEGs by using MCODE. The most significant module of upregulated DEGs contained COL1A2, COL5A2, COL4A2, LUM and COL4A1. One of five upregulated genes, COL1A2, is a subtype of Type I collagen, which can be produced by stromal fibroblasts and cancer cells [30-32]. And COL1A2 plays an important role in the invasion and metastasis of ovarian cancer cells since it promotes the migration of cancer cells by interacting with α2β1-integrin [33]. However, what roles it plays in immune cell infiltration and immunotherapy about ESCA are still indistinct.

Subsequently, we explored the similar genes of COL1A2 in ESCA by using GEPIA. The top 10 similar genes were screened including COL1A1, COL3A1, SPARC, ZNF469, POSTN, CTSK, COL5A1, LRRC15, ADAM12, and CLEC11A. We validated the expressions of COL1A2 and its similar genes in primary esophageal carcinoma samples compared with normal esophageal squamous epithelium samples, while CTSK, SPARC, and CLEC11A were not significantly upregulated. Among these genes, the high expression of COL1A2, COL1A1, COL3A1, SPARC, ZNF469, POSTN, and CTSK were negatively correlated with the disease-free survival (DFS) of ESCA patients. Then these genes (COL1A2, COL1A1, COL3A1, ZNF469 and POSTN) that have been validated and had prognostic value were used for immune-related research. The expression levels of COL1A2, COL1A1, COL3A1, ZNF469 and POSTN were positively correlated with the infiltration levels of macrophages and dendritic cells, and the expression levels of ZNF469 was also positively correlated with the infiltration levels of CD4+ T cells. These results indicated these genes might be the candidate genes for assessing the immune infiltration levels in ESCA.

Finally, to further explore the relationship between COL1A2 with its similar genes and the infiltration levels of various immune cells comprehensively, we analyzed the correlation between these genes and immune marker sets. The expression of COL1A2 was significantly correlated with the expression of marker genes from tumor-infiltrating B cells (CD19 and CD79A), monocytes (CD86 and CD115), TAMs (CCL2 and IL10), M1 macrophages (COX2), neutrophils (CD66b, CD11b, and CCR7), DCs (BDCA-4), T-helper (STAT4, GATA3, STAT6, BCL6, and IL17A), Treg (FOXP3, CCR8, and TGFβ) and exhausted T cells (CTLA4 and TIM-3). The expression of COL1A1 was significantly correlated with the expression of marker genes from tumor-infiltrating CD8+ T cell (CD8B), B cells (CD19 and CD79A), monocytes (CD86 and CD115), TAMs (CCL2 and IL10), M1 macrophages (COX2), neutrophils (CD66b, CD11b, and CCR7), DCs (HLA-DPA1), T-helper (STAT4, STAT5A, STAT6, BCL6, and IL17A), Treg (FOXP3, CCR8, and TGFβ) and exhausted T cells (TIM-3). Type I collagen (COL1A1 and COL1A2) was correlated with the resembling of TAMs in non-small cell lung carcinoma (NSCLC) while it was not correlated with the paucity of T-cell accumulation in pancreatic ductal adenocarcinoma (PDAC), which were consistent with our results [34, 35]. The expression of COL3A1 was significantly correlated with the expression of marker genes from tumor-
infiltrating B cells (CD19 and CD79A), monocytes (CD86 and CD115), M1 macrophages (COX2), neutrophils (CD66b and CD11b), DCs (HLA-DQB1 and HLA-DPA1), T-helper (STAT4, GATA3, STAT6, BCL6, and IL17A), Treg (FOXP3, CCR8, and TGFβ) and exhausted T cells (CTLA4 and TIM-3). COL3A1, a subtype of Type III collagen, which is similar to Type I collagen, and primarily refers to the growth and metastasis of tumors such as osteosarcoma and nasopharyngeal tumors [36, 37]. While our results suggested that COL3A1 also can be a potential biomarker for assessing the infiltration of the immune cells in ESCA. The high expression of COL3A1 may promote the growth and metastasis of esophageal carcinoma by recruiting various immune cells. The expression of ZNF469 was significantly correlated with the expression of marker genes from tumor-infiltrating B cells (CD19 and CD79A), monocytes (CD86 and CD115), neutrophils (CD66b and CD11b), T-helper (STAT4, STAT6, STAT5A, BCL6, and IL17A), Treg (FOXP3, CCR8, and TGFβ) and exhausted T cells (CTLA4 and TIM-3). Zinc finger protein (ZNF469) gene mainly refers to the pathological process of corneal related disease and it is similar to three types of collagen including COL1A1, COL1A2, and COL4A1 [38-40]. Interestingly, we firstly found the important roles of ZNF469 in the assessment of prognosis and immune cell infiltration in ESCA. This gene may be an important biomarker associated with immune infiltration in the esophageal carcinoma tumor microenvironment. The expression of POSTN was significantly correlated with the expression of marker genes from tumor-infiltrating B cells (CD19 and CD79A), monocytes (CD86 and CD115), TAMs (CD68), M1 macrophages (COX2), neutrophils (CD66b, CD11b, and CCR7), T-helper (STAT4, GATA3, STAT6, STAT5A, BCL6, and IL17A), Treg (FOXP3, CCR8, and TGFβ) and exhausted T cells (CTLA4 and TIM-3). Periostin (POSTN) is a soluble ECM protein and it is upregulated in many cancers such as esophageal carcinoma, head and neck squamous cell carcinoma, and non-small cell lung cancer [41-43]. Recent studies demonstrated that it was an important biomarker associated with the pathogenesis and prognosis of esophageal carcinoma, and our results further underlined the important role of POSTN in immune cells infiltration [44, 45].

Different from previous studies, we analyzed the common DEGs between the ESCC and EAC to find the most important variations in ESCA and identified the important roles of COL1A2 with its similar genes in influencing the immune infiltration in ESCA. However, some limitations still exist in the present study. Firstly, some technical defects of gene chip and sequencing technology existed here. Secondly, we only analyzed the potential biomarker in ESCA and the specific mechanism of immune cell infiltration induced by these genes are needed to be demonstrated in vivo and in vitro experiments.

In summary, we identified COL1A2 and its similar genes as the potential biomarkers for assessing the infiltration levels of immune cells using a comprehensive analysis of bioinformatics, which could help us better understand the mechanism of tumor immune cell infiltration in ESCA. However, more studies are needed to verify these results.

**Competing Interests**

The authors declare that they have no conflict of interests.
Funding
This work was supported by grants from the National Natural Science Foundation of China (no. 81672688) and the Natural Science Foundation of Hunan Province, China (Grant No. 2019J50359) and College Students' Innovative Entrepreneurial Training Plan Program of China (no. 2020105330254).

Author Contribution
ZW, MC and YQ designed this study. ZW, MC, YQ, YY, YH and XL collected the raw data and prepared the analysis. ZW, MC and WZ were involved in writing and revising the paper. All authors read and approved the final manuscript.

Data Availability Statement
The gene expression data are available in public databases including GEO (https://www.ncbi.nlm.nih.gov/geo/) and TCGA (https://portal.gdc.cancer.gov).

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. 10.3322/caac.21492.
[2] Lagergren J, Smyth E, Cunningham D, Lagergren P. Oesophageal cancer. Lancet 2017;390(10110):2383-96. 10.1016/S0140-6736(17)31462-9.
[3] Smyth EC, Lagergren J, Fitzgerald RC, Lordick F, Shah MA, Lagergren P, et al. Oesophageal cancer. Nat Rev Dis Primers 2017;3:17048. 10.1038/nrdp.2017.48.
[4] Burugu S, Dancsok AR, Nielsen TO. Emerging targets in cancer immunotherapy. Semin Cancer Biol 2018;52(Pt 2):39-52. 10.1016/j.semcancer.2017.10.001.
[5] Sarwat M, Yamdagni MM. DNA barcoding, microarrays and next generation sequencing: Recent tools for genetic diversity estimation and authentication of medicinal plants. Crit Rev Biotechnol 2016;36(2):191-203. 10.3109/07388551.2014.947563.
[6] Bai G, Bai R, Saeidian AH, Bai M. Microarray to deep sequencing: Transcriptome and miRNA profiling to elucidate molecular pathways in systemic lupus erythematosus. Immunol Res 2016;64(1):14-24. 10.1007/s12026-015-8672-y.
[7] Mantione KJ, Kream RM, Kuzelova H, Ptacek R, Raboch J, Samuel JM, et al. Comparing bioinformatic gene expression profiling methods: Microarray and RNA-Seq. Med Sci Monit Basic Res 2014;20:138-42. 10.12659/MSMBR.892101.
[8] Yue Y, Song M, Qiao Y, Li P, Yuan Y, Lian J, et al. Gene function analysis and underlying mechanism of esophagus cancer based on microarray gene expression profiling. Oncotarget 2017;8(62):35681-99. 10.18632/oncotarget.16051.
[9] Liu H, Zhang Q, Lou Q, Zhang X, Cui Y, Wang P, et al. Differential analysis of lncRNA, miRNA and mRNA expression profiles and the prognostic value of lncRNA in esophageal cancer. Pathol Oncol Res 2019. 10.1007/s12253-019-00655-8.
[10] Zeng JH, Xiong DD, Pang YY, Zhang Y, Tang RX, Luo DZ, et al. Identification of molecular targets for esophageal carcinoma diagnosis using miRNA-seq and RNA-seq data from the Cancer Genome Atlas: A study of 187 cases. Oncotarget 2017;8(22):35681-99. 10.18632/oncotarget.16051.
[11] Gautier L, Cope L, Bolstad BM, Irizarry RA. Affy--analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 2004;20(3):307-15. 10.1093/bioinformatics/btg405.
[12] Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 2004;3:e3. 10.2202/1544-6115.1027.
[13] Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res 2019;47(D1):D419-26. 10.1093/nar/gky1038.
[14] Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 2019;47(D1):D607-13. 10.1093/nar/gky1131.
[15] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13(11):2498-504.
10.1101/gr.1239303.

[16] Bandettini WP, Kellman P, Mancini C, Booker OJ, Vasu S, Leung SW, et al. MultiContrast Delayed Enhancement (MCODE) improves detection of subendocardial myocardial infarction by late gadolinium enhancement cardiovascular magnetic resonance: A clinical validation study. J Cardiovasc Magn Reson 2012;14:83. 10.1186/1532-429X-14-83.

[17] Chandrashekar DS, Bashel B, Balasubramanyan S, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, et al. UALCAN: A portal for facilitating tumor subgroup gene expression and survival analyses. Neoplasia 2017;19(8):649-58. 10.1016/j.neo.2017.05.002.

[18] 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(W1):W98-102. 10.1093/nar/gkx247.

[19] Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. TIMER: A web server for comprehensive analysis of Tumor-Infiltrating immune cells. Cancer Res 2017;77(21):e108-10. 10.1158/0008-5472.CAN-17-0307.

[20] Danaher P, Warren S, Dennis L, D’Amico L, White A, Disis ML, et al. Gene expression markers of Tumor Infiltrating Leukocytes. J Immunother Cancer 2017;5:18. 10.1186/s40425-017-0215-8.

[21] Siemers NO, Holloway JL, Chang H, Chasalow SD, Ross-Macdonald PB, Voliva CF, et al. Genome-wide association analysis identifies genetic correlates of immune infiltrates in solid tumors. Plos One 2017;12(7):e179726. 10.1371/journal.pone.0179726.

[22] Sousa S, Maatta J. The role of tumour-associated macrophages in bone metastasis. J Bone Oncol 2016;5(3):135-38. 10.1016/j.jbo.2016.03.004.

[23] Fang S, Dai Y, Mei Y, Yang M, Hu L, Yang H, et al. Clinical significance and biological role of cancer-derived Type I collagen in lung and esophageal cancers. Thorac Cancer 2019;10(2):277-88. 10.1111/1759-7714.12947.

[24] Miyashita T, Tajima H, Shah FA, Oshima M, Makino I, Nakagawara H, et al. Impact of inflammation-metaplasia-adenocarcinoma sequence and inflammatory microenvironment in esophageal carcinogenesis using surgical rat models. Ann Surg Oncol 2014;21(6):2012-19. 10.1245/s10434-014-3537-5.

[25] Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inhibiting tumour purity and stromal and immune cell admixture from expression data. Nat Commun 2013;4:2612. 10.1038/ncomms3612.

[26] Wu T, Dai Y. Tumor microenvironment and therapeutic response. Cancer Lett 2017;387:61-68. 10.1016/j.canlet.2016.01.043.

[27] Karlou M, Tzeli V, Efstathiou E. Therapeutic targeting of the prostate cancer microenvironment. Nat Rev Urol 2010;7(9):494-509. 10.1038/nrevurol.2010.134.

[28] Gao S, Yang D, Fang Y, Lin X, Jin X, Wang Q, et al. Engineering nanoparticles for targeted remodeling of the tumor microenvironment to improve cancer immunotherapy. Theranostics 2019;9(1):126-51. 10.7150/thno.29431.

[29] Zou W. Immunosuppressive networks in the tumour environment and their therapeutic relevance. Nat Rev Cancer 2005;5(4):263-74. 10.1038/nrc1586.

[30] Steinberg J. The turnover of collagen in fibroblast cultures. J Cell Sci 1973;12(1):217-34.

[31] Steinberg J. Collagen turnover and the growth state in 3T6 fibroblast cultures. Lab Invest 1978;39(5):491-97.

[32] Fishman DA, Kearsn A, Chilukuri K, Bafetti LM, O'Toole EA, Georgacopoulos J, et al. Metastatic dissemination of human ovarian epithelial carcinoma is promoted by alpha2beta1-mediated interaction with type I collagen. Invasion Metastasis 1998;18(1):15-26. 10.1159/000024495.

[33] Rebello SP, Pinto C, Martins TR, Harrer N, Estrada MF, Loza-Alvarez P, et al. 3D-3-culture: A tool to unveil macrophage plasticity in the tumour microenvironment. Biomaterials 2018;163:185-97. 10.1016/j.biomaterials.2018.02.030.

[34] Carstens JL, Correa DSP, Yang D, Barua S, Wang H, Rao A, et al. Spatial computation of intratumoral T cells correlates with survival of patients with pancreatic cancer. Nat Commun 2017;8:15095. 10.1038/ncomms15095.

[35] Xu W, Li Z, Zhu X, Xu R, Xu Y. MiR-29 family inhibits resistance to methotrexate and promotes cell apoptosis by targeting COL3A1 and MCL1 in osteosarcoma. Med Sci Monit 2018;24:8812-21. 10.12659/MSM.911972.

[36] Qiu F, Sun R, Deng N, Guo T, Cao Y, Yu Y, et al. MiR-29a/b enhances cell migration and invasion in
[38] Yu X, Chen B, Zhang X, Shentu X. Identification of seven novel ZNF469 mutations in keratoconus patients in a Han Chinese population. Mol Vis 2017;23:296-305.

[39] Micheal S, Siddiqui SN, Zafar SN, Gabriela NI, Khan MI, Bergen A. Identification of a novel ZNF469 mutation in a pakistani family with brittle cornea syndrome. Cornea 2019;38(6):718-22. 10.1097/ICO.0000000000001828.

[40] Abu A, Frydman M, Marek D, Pras E, Nir U, Reznik-Wolf H, et al. Deleterious mutations in the Zinc-Finger 469 gene cause brittle cornea syndrome. Am J Hum Genet 2008;82(5):1217-22. 10.1016/j.ajhg.2008.04.001.

[41] Michaylira CZ, Wong GS, Miller CG, Gutierrez CM, Nakagawa H, Hammond R, et al. Periostin, a cell adhesion molecule, facilitates invasion in the tumor microenvironment and annotates a novel tumor-invasive signature in esophageal cancer. Cancer Res 2010;70(13):5281-92. 10.1158/0008-5472.CAN-10-0704.

[42] Sasaki H, Lo KM, Chen LB, Auclair D, Nakashima Y, Moriyama S, et al. Expression of Periostin, homologous with an insect cell adhesion molecule, as a prognostic marker in non-small cell lung cancers. Jpn J Cancer Res 2001;92(8):869-73. 10.1111/j.1349-7006.2001.tb01174.x.

[43] Kudo Y, Ogawa I, Kitajima S, Kitagawa M, Kawai H, Gaffney PM, et al. Periostin promotes invasion and anchorage-independent growth in the metastatic process of head and neck cancer. Cancer Res 2006;66(14):6928-35. 10.1158/0008-5472.CAN-05-4540.

[44] Zhang H, Zhong J, Tu Y, Liu B, Chen Z, Luo Y, et al. Integrated bioinformatics analysis identifies hub genes associated with the pathogenesis and prognosis of esophageal squamous cell carcinoma. Biomed Res Int 2019;2019:2615921. 10.1155/2019/2615921.

[45] Heidari P, Esfahani SA, Turker NS, Wong G, Wang TC, Rustgi AK, et al. Imaging of secreted extracellular periostin, an important marker of invasion in the tumor microenvironment in esophageal cancer. J Nucl Med 2015;56(8):1246-51. 10.2967/jnumed.115.156216.
