Mathematical Biosciences and Engineering

MBE, 20(8): 13798–13823.
DOI: 10.3934/mbe.2023614
Received: 08 April 2023
Revised: 08 May 2023
Accepted: 23 May 2023
Published: 16 June 2023

http://www.aimspress.com/journal/MBE

Research article

Identification and validation of biomarkers for epithelial-mesenchymal transition-related cells to estimate the prognosis and immune microenvironment in primary gastric cancer by the integrated analysis of single-cell and bulk RNA sequencing data

Kaiyu Shen[1], Shuaiyi Ke[2], Binyu Chen[1], Tiantian Zhang[1], Hongtai Wang[3], Jianhui Lv[3] and Wencang Gao[4,*]

[1] The Second Clinical Medical College, Zhejiang Chinese Medical University, Hangzhou 310053, China
[2] Department of Internal Medicine, XianJu People's Hospital, XianJu 317399, China
[3] Department of General Surgery, XianJu People' Hospital, XianJu 317399, China
[4] Department of Oncology, Zhejiang Chinese Medical University, Hangzhou 310005, China

* **Correspondence:** Email: 20164580@zcmu.edu.cn.

Abstract: Background: The epithelial-mesenchymal transition (EMT) is associated with gastric cancer (GC) progression and immune microenvironment. To better understand the heterogeneity underlying EMT, we integrated single-cell RNA-sequencing (scRNA-seq) data and bulk sequencing data from GC patients to evaluate the prognostic utility of biomarkers for EMT-related cells (ERCs), namely, cancer-associated fibroblasts (CAFs) and epithelial cells (ECs). Methods: scRNA-seq data from primary GC tumor samples were obtained from the Gene Expression Omnibus (GEO) database to identify ERC marker genes. Bulk GC datasets from the Cancer Genome Atlas (TCGA) and GEO were used as training and validation sets, respectively. Differentially expressed markers were identified from the TCGA database. Univariate Cox, least-absolute shrinkage, and selection operator regression analyses were performed to identify EMT-related cell-prognostic genes (ERCPGs). Kaplan-Meier, Cox regression, and receiver-operating characteristic (ROC) curve analyses were adopted to evaluate the prognostic utility of the ERCPG signature. An ERCPG-based nomogram was constructed by integrating independent prognostic factors. Finally, we evaluated the correlations between the ERCPG signature and immune-cell infiltration and verified the expression of ERCPG prognostic signature genes by in vitro cellular assays. Results: The ERCPG signature was comprised of seven genes

(COL4A1, F2R, MMP11, CAV1, VCAN, FKBP10, and APOD). Patients were divided into high- and low-risk groups based on the ERCPG risk scores. Patients in the high-risk group showed a poor prognosis. ROC and calibration curves suggested that the ERCPG signature and nomogram had a good prognostic utility. An immune cell-infiltration analysis suggested that the abnormal expression of ERCPGs induced the formation of an unfavorable tumor immune microenvironment. In vitro cellular assays showed that ERCPGs were more abundantly expressed in GC cell lines compared to normal gastric tissue cell lines. Conclusions: We constructed and validated an ERCPG signature using scRNA-seq and bulk sequencing data from ERCs of GC patients. Our findings support the estimation of patient prognosis and tumor treatment in future clinical practice.

Keywords: gastric cancer; biomarker; epithelial-mesenchymal transition; prognosis; bioinformatics

1. Introduction

Gastric cancer (GC) is among the most common gastrointestinal malignancies worldwide and originates from the epithelial cells (ECs) of the gastric mucosa. GC accounts for 12.4% of cancer-related deaths in China and ranks fifth in incidence among all malignant tumors [1]. Surgical resection remains the primary treatment for early-stage GC [2], and the widespread use of chemotherapy and radiotherapy improves the prognosis and quality of life of postoperative patients and patients with advanced-stage GC [3,4]. However, the high postoperative recurrence rate among patients with GC and the low survival rate of patients with advanced-stage GC remain as major challenges for GC treatment [5,6]. Hence, the effective prediction of prognosis is required to design treatment strategies with greater clinical significance for patients.

Recent advancements in genome-sequencing technologies have ushered in a new era for tumor diagnosis and prognosis evaluation, and biomarkers have now become an integral part of auxiliary cancer diagnosis in clinical practice [7]. Further investigation into the mechanisms of action of GC-related biomarkers is of great significance for determining the prognosis of patients with GC and identifying effective therapeutic targets.

The epithelial-mesenchymal transition (EMT) is a reversible conversion of ECs into cells with a mesenchymal phenotype that are involved in gastrulation, wound healing, and the development of malignant tumors [8]. Extensive evidence suggests that the EMT process can accelerate tumor progression and drug resistance [9] and is associated with the malignant biological behavior of GC cells [10,11].

As two types of key cells involved in EMT, cancer-associated fibroblasts (CAFs) and ECs play important roles in the occurrence, proliferation, and invasion of most epithelium-derived malignant tumors. The tumor microenvironment (TME) is a complex meshwork consisting of the extracellular matrix (ECM) and cells, including CAFs, inflammatory cells, cellular components of blood vessels, and mesenchymal stem cells [12]. The stromal component of tumors can affect tumor development [13]. CAFs are important cellular components of the tumor stroma and can affect the tumor-stroma ratio (TSR). Patients with GC and a high TSR had a more advanced tumor stage and experienced a shorter survival period [14]. In addition, CAFs can promote angiogenesis and reshape the ECM by synthesizing soluble molecules, such as vascular endothelial growth factor (VEGF), transforming growth factor-β (TGF-β), matrix metalloproteinase (MMP), and ECM components, to induce EMT and create a microenvironment conducive for tumor invasion [15–18]. In ECs, cell-cell junctions are disrupted by

the loss of E-cadherin expression, with an upregulation of mesenchymal markers such as S100A4, enabling ECs to acquire invasive and metastatic potential during EMT [8,19]. In contrast, therapies targeting CAFs and EC-related biomarkers have shown positive antitumor efficacy. For example, the TGF-β receptor kinase inhibitor galunisertib inhibited EMT in hepatocellular carcinoma cell lines [20]. Moreover, catumaxomab can target EpCAM, while ADH-1 can target N-cadherin [21].

The drawback of conventional RNA-sequencing technologies lies in the fact that they only reveal the average gene-expression profile in studied tumor samples and cannot reflect the heterogeneity within tumor cell populations. At present, the rapid development of single-cell RNA sequencing (scRNA-seq) technologies has fueled research breakthroughs in biomedicine have previously not been achieved using conventional sequencing techniques [22]. scRNA-seq technologies have been applied in research on circulating tumor cells, the TME, and tumor drug resistance [23], enabling more accurate analysis of genetic data in precision medicine. Sathe et al. [24] used scRNA-seq technology to compare normal gastric tissues and GC cell subpopulations differences and revealed multiple cancer-promoting cell subpopulations reprogramming process. The scRNA-seq technique has not only been applied to explore the transcriptional heterogeneity of GCs with different degrees of differentiation [25], but also enriched the understanding of the interaction between immune cells and stromal cells in the GC immunosuppressive microenvironment [26]. In addition, scRNA-seq efficiently identifies novel GC biomarkers. Wang et al. [27] identified GC lymph node metastasis-related biomarkers using scRNA-seq, which provided new insights into the development of GC therapeutic targets.

In this study, we identified marker genes of EMT-related CAFs and ECs based on scRNA-seq data from primary GC tumor samples, constructed a prognostic model by integrating the scRNA-seq and bulk sequencing data from patients with GC, and used external data from the Gene Expression Omnibus (GEO) database to validate the model. Finally, we verified the expression of ERCPGs by via quantitative real-time PCR (qRT-PCR) in vitro cells.

2. Materials and methods

2.1. Data acquisition

The scRNA-seq data of three primary GC tumor samples were obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) (accession number: GSE163558). The dataset contained three primary tumor samples, one paracancer tissue sample, and six metastatic tumor samples from different tissues and organs. Our study aimed to explore the biological mechanisms of cell subpopulations associated with primary GC; therefore, normal paracancer samples were not included. Considering that there may be characteristics of cells and the TME different metastases [28,29] and to avoid interference due to sample heterogeneity, we only included scRNA-seq data from three primary tumor samples; the other six metastasis samples were excluded from our analysis. The transcriptome data and clinical data of the TCGA-Stomach adenocarcinoma (STAD) cohort were downloaded from the TCGA database (https://portal.gdc.cancer.gov/) and used as the bulk sequencing data. In this study, we included 32 normal samples and 375 GC samples, among which 317 GC samples contained transcriptome data and complete paired clinical data were used as the training set for subsequent prognosis-related studies. The clinical data for the GC samples included age, sex, grade, and tumor stage. The GSE15459 dataset (GEO database) contained 192 GC samples and was used as the external validation set. The paired clinical data included sex, age, and tumor stage.

2.2. scRNA-seq data analysis

The scRNA-seq data for the GC samples were converted into analyzable Seurat objects using the CreatSeuratObject function in the "Seurat" package of the R software (v4.1.0). We used the "harmony" package to remove the batch effect from the three primary GC samples. The proportion of mitochondrial genes in each cell was determined using the "PercentageFeatureSet" function. During the quality-control screening, high-quality genes and cells were selected using the "subset" function based on the following criteria: 1) containing < 10% mitochondrial genes (parameter min.cells = 10), 2) genes expressed in at least 10 cells (parameter percent.mt < 10), and 3) cells with genes expression > 200 (parameter nFeature_RNA > 200). Subsequently, the "FindVariableFeatures" function was used to identify highly variable genes. After the data were scaled using the "ScaleData" function, a principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) were performed for dimensionality reduction, clustering, and visualization of the highly variable genes. The parameters of the "FindAllMarkers" function were set up (min.pct = 0.3, logfc.threshold = 0.25) to identify the marker genes of cell clusters. A heatmap was generated to identify the top five discriminative genes per cluster. Based on the CellMarker database [30] and a review of previous studies [31–36], the cellular clusters were annotated based on cell-specific marker genes. Finally, we identified the eligible marker genes of ERCs based on |log2 (fold-change)| ≥ 1 and adjusted P (adjPval) < 0.01 parameters.

2.3. Identifying Differentially Expressed Marker Genes (DEMGs) based on bulk data

The "limma" package of R was used to screen for differentially expressed genes (DEGs) in 32 normal samples and 375 GC samples with sequencing data deposited in the TCGA database (screening criteria: |log2 (fold-change) ≥ 1| and adjPval < 0.05). We performed set-intersection analysis using the DEGs obtained from the bulk sequencing data of patients with GC and the ERC-marker genes to identify the DEMGs. The "VennDiagram" package was used to plot the differentially expressed marker genes (DEMGs).

2.4. Functional enrichment analysis

To explore the common characteristics of the DEMGs, we used the "org.Hs.eg.db" and "clusterProfiler" packages of R to perform a gene ontology (GO) analysis with the "enrichGO" function. The "enrichKEGG" function was used to perform a Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway-enrichment analyses (parameter organization = "hsa", pvalueCutoff = 1, qvalueCutoff = 1). A P value of < 0.05 indicated significant enrichment.

2.5. Constructing the ERCPG risk-scoring system and prognostic analysis

For the TCGA database, we used the "survival," "survminer," and "glmnet" packages of R. A univariate Cox regression analysis was performed to preliminarily select DEMGs with a prognostic significance. A least absolute shrinkage and selection operator (LASSO) analysis was performed based on the optimal value of λ to construct the ERCPG signature ("glmnet" function parameters were family = "cox", maxit = 1000). The ERCPG risk score of each GC sample was calculated using Eq (1) as follows:

$$ERCPG\ Risk\ Score = \sum_{i=0}^{n} ExG * LASSOcoef\ G \quad (1)$$

Here, ExG is the gene-expression level and LASSOcoef G is the regression coefficient. TCGA-STAD training set and the GSE15459 validation set were divided into two subgroups, namely the high-risk and low-risk groups, based on the median values of the ERCPG risk scores. We used the R software and the Gene Expression Profiling Interactive Analysis (GEPIA) database (http://gepia.cancer-pku.cn/) to perform a Kaplan-Meier (K-M) survival analysis of the ERCPG-based high- and low-risk groups and the ERCPGs to investigate their correlations with the overall survival (OS) rate. Differences with P values < 0.05 were considered statistically significant. Additionally, we performed multiple testing corrections and calculated false discovery rate (FDR) values based on the "fdrtool" package. We also explored the value of the ERCPG risk scores in predicting the survival and determining the prognosis of patient subgroups with different clinical and prognostic characteristics. The "timeROC" package was used to generate receiver operating characteristic (ROC) curves for the ERCPG risk scores using TCGA-STAD training set and GSE15459 validation set. Univariate and multivariate Cox regression analyses were performed to evaluate the ERCPG risk scores and clinical prognostic characteristics and verify the independent prognostic significance of the ERCPG signature.

2.6. Constructing and evaluating the ERCPG nomogram

We used the "rms", "regplot", and "survival" packages to construct a prediction model for the ERCPG signature. The "regplot" function was used to generate an ERCPG nomogram based on factors with an independent prognostic significance in the TCGA-STAD training set to predict the 1, 3, and 5-year OS rates. Calibration curves drawn using the "calibrate" function were used to evaluate the prognostic utility of the ERCPG nomogram in TCGA-STAD training set and GSE15459 validation set, using the following parameters: cmethod = "KM", method = "boot."

2.7. GSEA of ERCPG signature

To investigate the biological mechanisms related to ERCPG risk scores, we performed a Gene Set Enrichment Analysis (GSEA) using the GSEA software (v4.1.0), with a permutation test parameter of 1000, and gene set parameters of "h.all.v2022.1.Hs.symbols.gmt" and "c2.cp.kegg.v2022.1.Hs.symbols.gmt." P < 0.05 indicated significant enrichment. The "AUCell", "clusterProfiler" and "ggplot2" packages were used to present the GSEA results for different cell subpopulations.

2.8. Immune-cell infiltration analysis of ERCPG signature

Based on the "cibersort" and "limma" packages, we obtained the abundance of 22 immune cell subtypes in GC. We applied the CIBERSORT algorithm to analyze the TCGA-STAD gene-expression data and evaluate the infiltration levels of 22 types of immune cells in the ERCPG-based low- and high-risk groups. We also evaluated the correlation between the seven ERCPGs and tumor immune cells using the Tumor Immune Estimation Resource (TIMER) online tool (https://cistrome.shinyapps.io/timer/).

2.9. Cell culture

Normal gastric tissue cells (GES-1) and four GC cell lines (MKN-45, HGC-27, BGC-823, and Hs-746T) were purchased from Hangzhou Freden Biotechnology Co. (Hangzhou, China). MKN-45,

BGC-823, and Hs-746T cells were grown in RPMI-1640 medium, whereas GES-1 and HGC-27 cells were grown in DMEM medium supplemented with 10% fetal bovine serum (FBS). Cells were incubated at 37°C with 5% CO_2.

2.10. Quantitative real-time PCR (qRT-PCR)

To quantify the expression of the seven ERCPG signatures, Trizol reagent was used to extract the total RNA from GES-1 and the remaining four GC cell lines according to the manufacturer's protocol. qRT-PCR was performed using the SYBR Premix Ex Taq according to the manufacturer's protocol. Each set of experiments was performed in triplicate. β-actin was used as an internal control. The expression level of the ERCPG signature was quantified using the $2^{-\Delta\Delta CT}$ method. The primers are as follows: β-actin, F 5'-TGGCACCCAGCACAATGAA-3', R 5'-CTAAGTCATAGTCCGCCTAGAAGCA-3'; COL4A1, F 5'-ACCCCCGGGAGAAATAGGT-3', R 5'-GGATTTGAAAAAGCAATGGCACT-3'; F2R, F 5'-TGCCTACCTCCTCTGTGTCTGTG-3', R 5'-ACTGCTGGGATCGGAACTTTCTTTG-3'; MMP11, F 5'-GATCGACTTCGCCAGGTACT-3', R 5'-CCCCGATAGTCCAGGTCTCA-3'; CAV1, F 5'-TCTTCCAACACGTAGCTGCC-3', R 5'-CGGTGTAGAGATGTCCCTGCG-3'; VCAN, F 5'-GCAAGTGATGCGGGTCTTTAC-3', R 5'-TTGCCGCCCTGTAGTGAAAC-3'; FKBP10 F 5'-CATGGGCATGTGTGTCAACG-3', R 5'-GAATGAGCCCCGCCAGG-3'; and APOD, F 5'-AATCGAAGGTGAAGCCACCC-3', R 5'-GTGCCGATGGCATAAACCAG-3'.

2.11. Statistical analysis

All statistical analyses and graphing were performed using the R software. Univariate Cox and LASSO regression analyses were performed to construct the ERCPG signature. The K-M method and the log-rank test were adopted to calculate survival differences. An independent prognostic analysis was performed based on the results of multivariate Cox analysis. Data were compared between multiple groups using an analysis of variance (ANOVA). A Spearman's correlation analysis was performed to identify the correlations between ERCPGs and different types of immune cells. Statistical significance was set at $P < 0.05$.

3. Results

3.1. scRNA-seq Analysis of GC samples

Data for three primary GC tumor samples containing 15,729 cells were obtained by downloading the GSE163558 dataset. After a quality-control screening, 8,381 cells and 21,635 genes were identified (Figure 1A). The correlation coefficient between the number of genes and depth of sequencing was 0.86, suggesting a considerable positive correlation trend (Figure 1B). PCA analysis of the obtained 2,000 highly variable genes (Figure 1C) was performed (Figure 1D) to obtain 40 principal components (PC). The t-SNE analysis was subsequently used to identify the top eight PC ($p < 0.05$) as 16 cell clusters (Figure 1E), and the cell-cluster distribution of the three GC samples is shown in Figure 1F. Figure 1G shows the proportion of cell clusters in the GC samples. The heatmap in Figure 1H shows the top five cluster-discriminative genes for each of the 16 cell clusters. We calculated the specific marker genes in each cell cluster (Table S1) and annotated the cell types with previous literature [31–36]. Cell cluster 0 was

annotated as natural killer cells; cell clusters 1 and 11 were annotated as neutrophils; cell clusters 2 and 8 were annotated as regulatory T (Treg) cells; cell cluster 3 was annotated as CD4+ T cells; cell clusters 4 and 13 were annotated as ECs; cell clusters 5 and 14 were annotated as B cells; cell clusters 6 and 10 were annotated as macrophages, cell cluster 12 was annotated as endothelial cells; cell cluster 7 was annotated as T helper 17 (Th17) cells; cell cluster 9 was annotated as CAFs; and cell cluster 15 was annotated as CD8+ T cells. The cell clusters annotation results are shown in Figure 1I. We identified 379 CAFs and 685 ECs. Finally, we excluded the genes that did not satisfy the criteria of |log2 (fold-change) ≥ 1| and adjPval < 0.01 and obtained 740 ERC marker genes (Table S2).

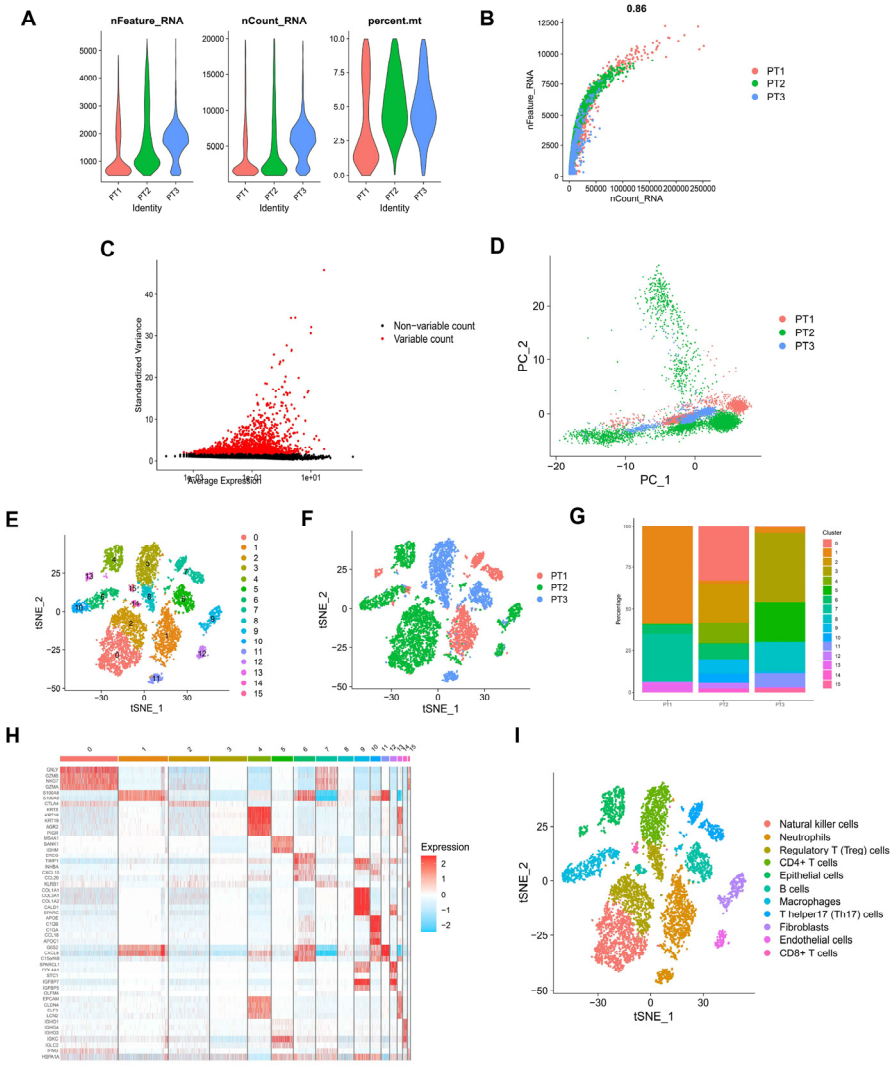

Figure 1. Characterization of the scRNA-seq data from 15,729 cells. (A) Quality-control analysis of the scRNA-seq data from GC cell samples. (B) Correlation between the number of expressed genes and sequencing depth. (C) Variogram displaying 2,000 genes with highly variable expression. (D) Preliminary classification of cells via PCA and identification of significantly usable dimensions. (E) Identification of 16 clusters based on the t-SNE algorithm. (F) Distribution of cell clusters from three primary GC tumor samples. (G) Proportional distribution of cell clusters in GC samples. (H) Heatmap of the top five significantly DEMGs for each cell cluster. (I) Annotated phenotype of each cell cluster.

3.2. Identification of DEMGs

We performed a differential analysis of normal and tumor samples in the TCGA-STAD dataset and obtained 718 DEGs, including 552 downregulated genes and 166 upregulated genes (Figure 2A) (Table S3). We performed a set-intersection analysis of the 740 ERC-marker genes and 718 DEGs to obtain 115 DEMGs for subsequent analysis (Figure 2B) (Table S4).

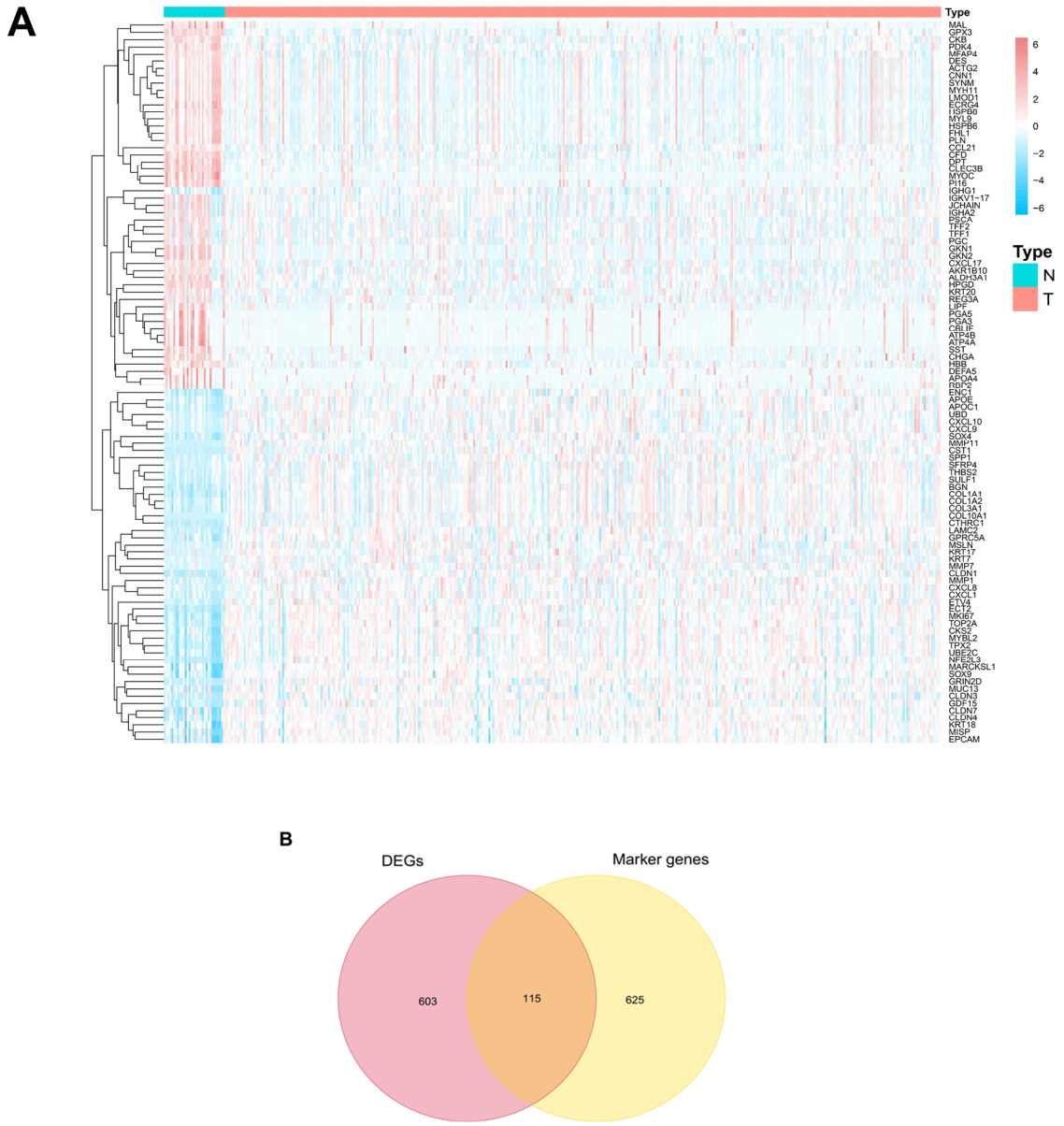

Figure 2. Identification of DEMGs in TCGA-STAD cohort. (A) A heatmap of the DEGs(Red represents up-regulated genes, blue represents down-regulated genes). (B) DEMGs were obtained as the intersection of DEGs and ERC marker genes.

3.3. GO- and KEGG-based functional-enrichment analyses of DEMGs

A GO analysis indicated that the DEMGs were enriched for several terms associated with

biological processes, including extracellular structure organization, focal adhesion, and a cellular response to transforming the growth factor beta stimulus (Figure 3A). The KEGG pathway analysis suggested that DEMGs were mainly enriched for signaling-pathway terms, including proteoglycans in cancer, the PI3K-Akt signaling pathway, the TGF-β signaling pathway, the ECM-receptor interaction, and signaling pathways regulating the pluripotency of stem cells (Figure 3B). Our results suggest that DEMGs may promote the progression of GC through these pathways.

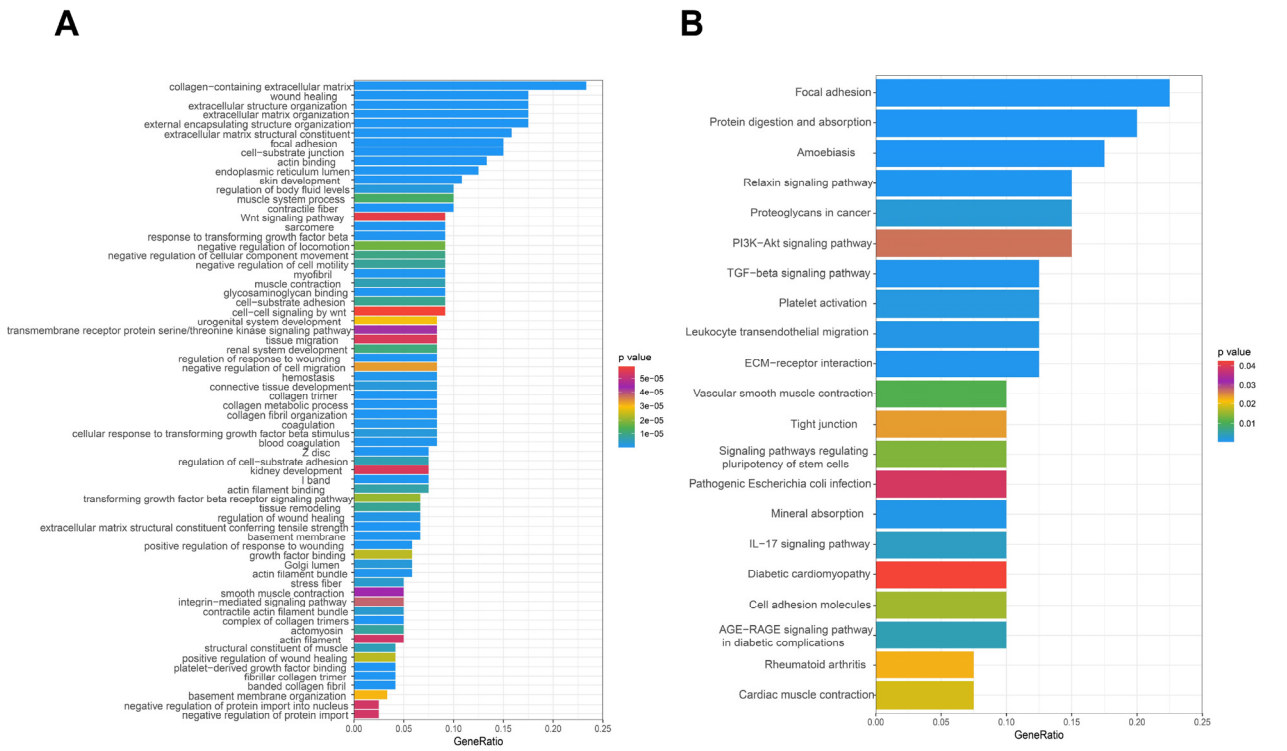

Figure 3. Functional enrichment analyses of DEMGs. (A) GO analysis. (B) KEGG analysis.

3.4. Construction of ERCPG signature

A univariate Cox regression analysis was performed using the TCGA-STAD training set to identify 30 cell-specific marker genes ($P < 0.05$) (Figure 4A). A subsequent LASSO regression analysis revealed the final ERCPG signature out of these 30 genes (Figure 4B). Seven ERCPGs were included and the ERCPG-based risk score was calculated as follows: 0.0046959 Ex (COL4A1) + 0.0952825 Ex (F2R) + 0.0715635 Ex (MMP11) + 0.0124528 Ex (CAV1) + 0.0619419 Ex (VCAN) + 0.0481557 Ex (FKBP10) + 0.0524272 Ex (APOD).

All seven ERCPGs were identified as risk factors (hazard ratio [HR] > 1). The log-rank test and K-M survival analysis were performed using the GEPIA database, and the results suggested that the high expression of six ERCPGs (COL4A1, F2R, CAV1, VCAN, FKBP10, and APOD) significantly correlated with a poor prognosis in patients with GC ($P < 0.05$) (Figure 4C). As can be seen from the distribution of ERCPGs in cell clusters, COL4A1, F2R, MMP11, CAV1, VCAN, and FKBP10 were distributed in CAF-related subpopulations (cell cluster 9), whereas APOD was mainly distributed in EC-related subpopulations (cell cluster 13) (Figure 4D).

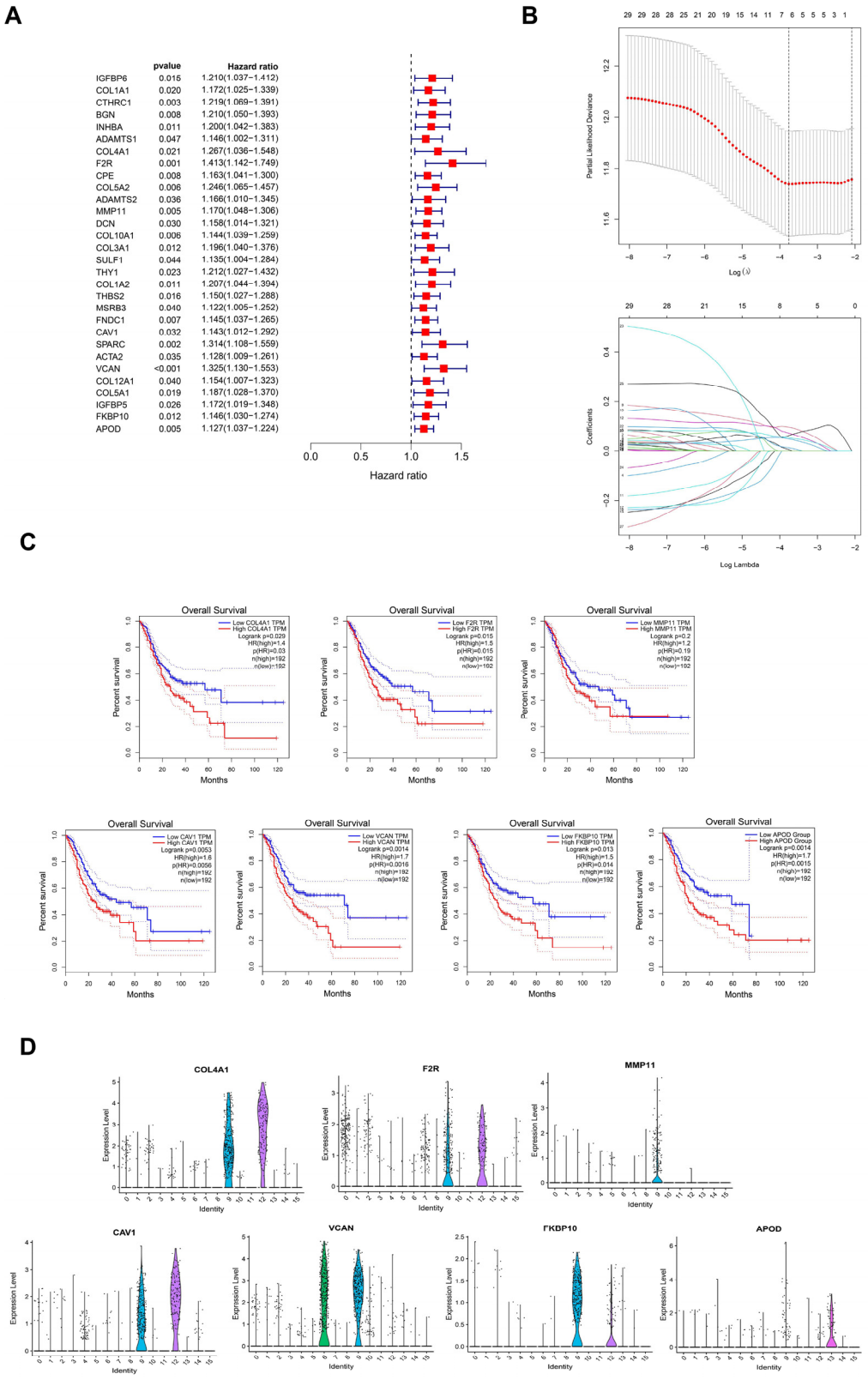

Figure 4. Screening of marker genes, *COL4A1*, *F2R*, *CAV1*, *VCAN*, *FKBP10*, and *APOD*, in the ERC subpopulation with prognostic significance. ERCPGs with prognostic significance were identified by univariate Cox (A) and LASSO regression analyses (B). (C) K-M analysis of ERCPGs. (D) Distribution of ERCPGs in cell clusters.

3.5. Prognostic-utility analysis of ERCPG signature

We tested the prognostic utility of the ERCPG signature. The GSE15459 validation set and the TCGA-STAD training set were divided into high- and low-risk groups based on the median values of the ERCPG risk scores. The patients in the high-risk subgroup had a shorter OS time (Figure 5A, B). The risk distribution and survival status of patients with GC are shown in Figure 5C and D. The area under the ROC curve (AUC) values in patients in the TCGA-STAD training set were 0.636, 0.635, and 0.787, after 1, 3, and 5-years, respectively (Figure 5E). The AUC values in the GSE15459 validation set were 0.688, 0.698, and 0.730, after 1, 3, and 5-years, respectively (Figure 5F). Univariate Cox and multivariate Cox regression analyses were performed to evaluate the prognostic value based on the TCGA-STAD training set, the ERCPG signature, and other clinical data of patients with GC. Age (HR: 1.024 and 1.032; 95% confidence interval [CI]: 1.006–1.042 and 1.012–1.051; and $P = 0.010$ and $P = 0.001$, respectively), staging (HR: 1.549 and 1.533; 95% CI: 1.244–1.929 and 1.066–2.164; and $P < 0.001$ and $P = 0.015$, respectively), and risk scores (HR: 3.741 and 3.768; 95% CI: 2.048–6.835 and 1.999–7.105; and $P < 0.001$ and $P < 0.001$, respectively) were independent risk factors for GC (Figure 6A, B). Cox analyses of the GSE15459 validation set showed that staging (HR: 2.789 and 2.625; 95% CI: 2.140–3.635 and 1.992–3.459; and $P < 0.001$ and $P < 0.001$, respectively) and risk scores (HR: 3.351 and 3.800; 95% CI: 2.122–6.212 and 2.065–5.160; and $P < 0.001$ and $P < 0.001$, respectively) had independent prognostic values (Figure 6 C, D).

We further analyzed the prognostic utility of the ERCPG risk scores in patient subgroups with different clinical and prognostic characteristics. As shown in Figure 7, patients aged ≤ 65 years ($P = 0.012$, FDR = 0.046), male patients ($P = 0.010$, FDR = 0.007), female patients ($P = 0.028$, FDR = 0.046), patients with grade 3 GC ($P = 0.040$, FDR = 0.042), patients with stage III-IV GC ($P = 0.007$, FDR = 0.006), patients with T3-4 GC ($P = 0.004$, FDR = 0.005), patients with N1-3 GC ($P = 0.003$, FDR = 0.008), patients with M0 GC ($P = 0.009$, FDR = 0.007), and in the low-risk group survived longer and had a better prognosis than those in the high-risk group. These results indicate that the ERCPG signature can be used as an independent predictor of the prognosis of patients with GC.

3.6. Construction and validation of ERCPG nomogram

To improve the clinical utility of the ERCPG signature, we constructed an ERCPG nomogram with the independent prognostic factors obtained from the multivariate Cox regression analysis of the TCGA-STAD training set (Figure 8A). The generated 1, 3, and 5-year calibration curves confirmed the accuracy of the ERCPG nomogram using both the TCGA-STAD training set and the GSE15459 validation set (Figure 8B, C). Hence, our results suggest that the ERCPG nomogram has a good clinical value.

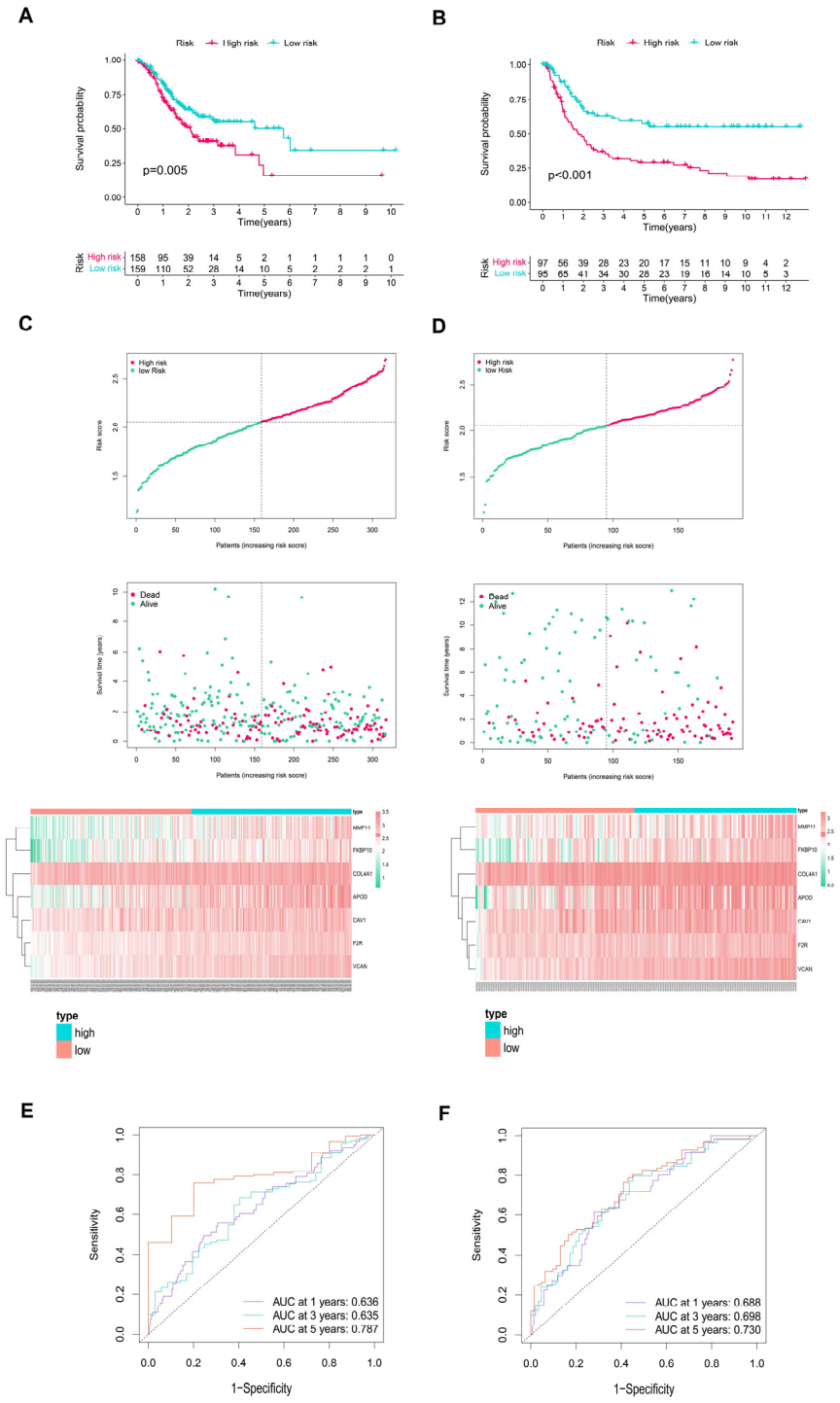

Figure 5. Correlation of the ERCPG signature with the survival and prognosis of patients with GC. K-M curves for the ERCPG-based high- and low-risk groups, as determined using the TCGA-STAD training set (A) and the GSE15459 validation set (B). Distribution of the ERCPG risk score and patient survival status, and heatmaps for ERCPG expression levels, as determined using the TCGA-STAD training cohort (C) and GSE15459 validation cohort (D). The ROC curves depict the accuracy of the ERCPG signature in predicting the 1, 3, and 5-year survival rates of patients with GC in the TCGA-STAD training cohort (E) and the GSE15459 validation cohort (F).

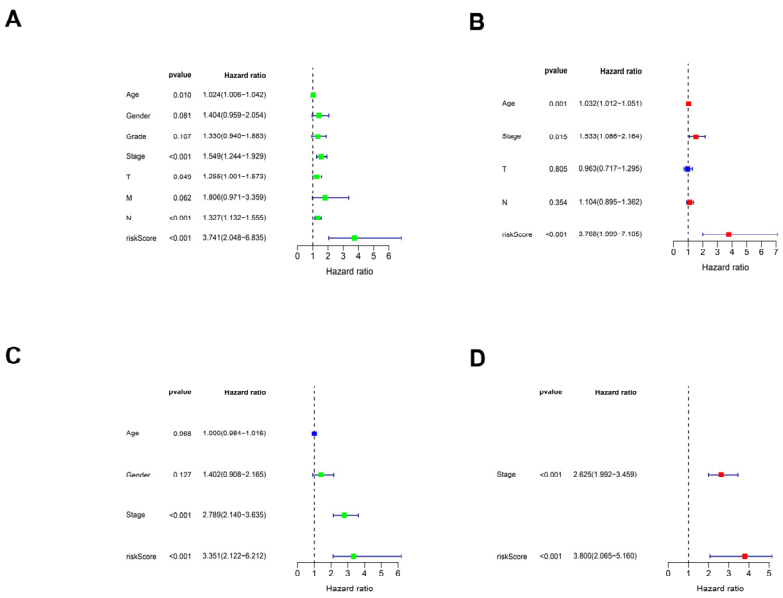

Figure 6. Independent prognostic analysis of the ERCPG signature in different clinical subgroups. Univariate Cox analysis (A) and multivariate Cox analysis (B) using the TCGA-STAD training set. Univariate Cox analysis (C) and the multivariate Cox analysis (D) using the GSE15459 validation set.

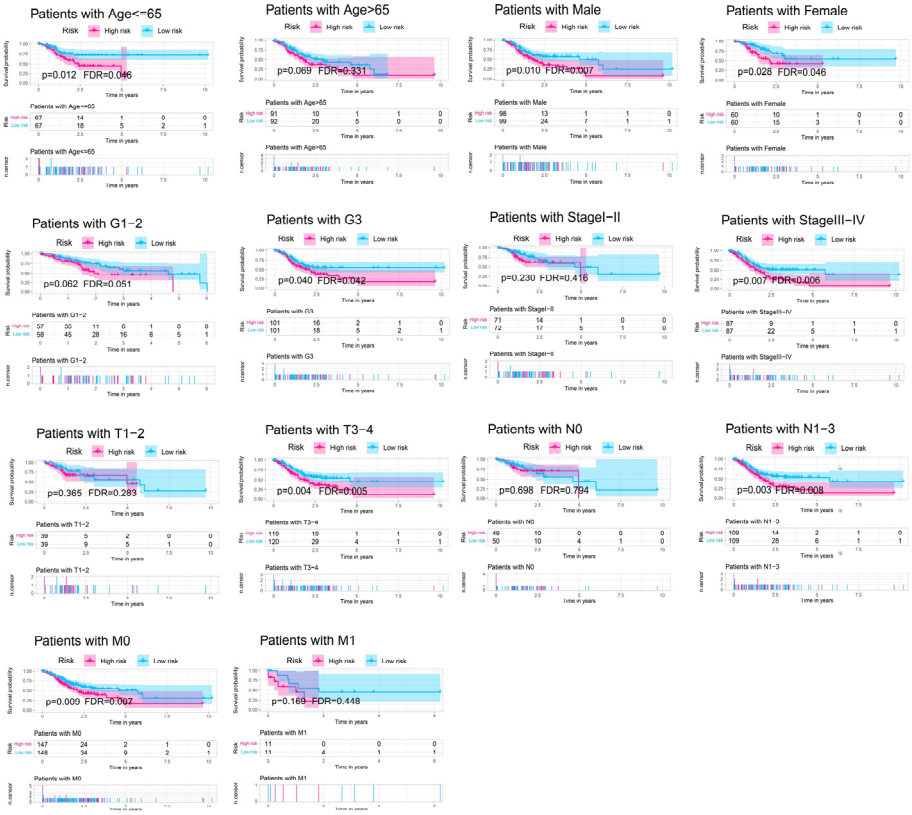

Figure 7. K-M analyses of patients with different clinical characteristics, as determined using the TCGA-STAD cohort.

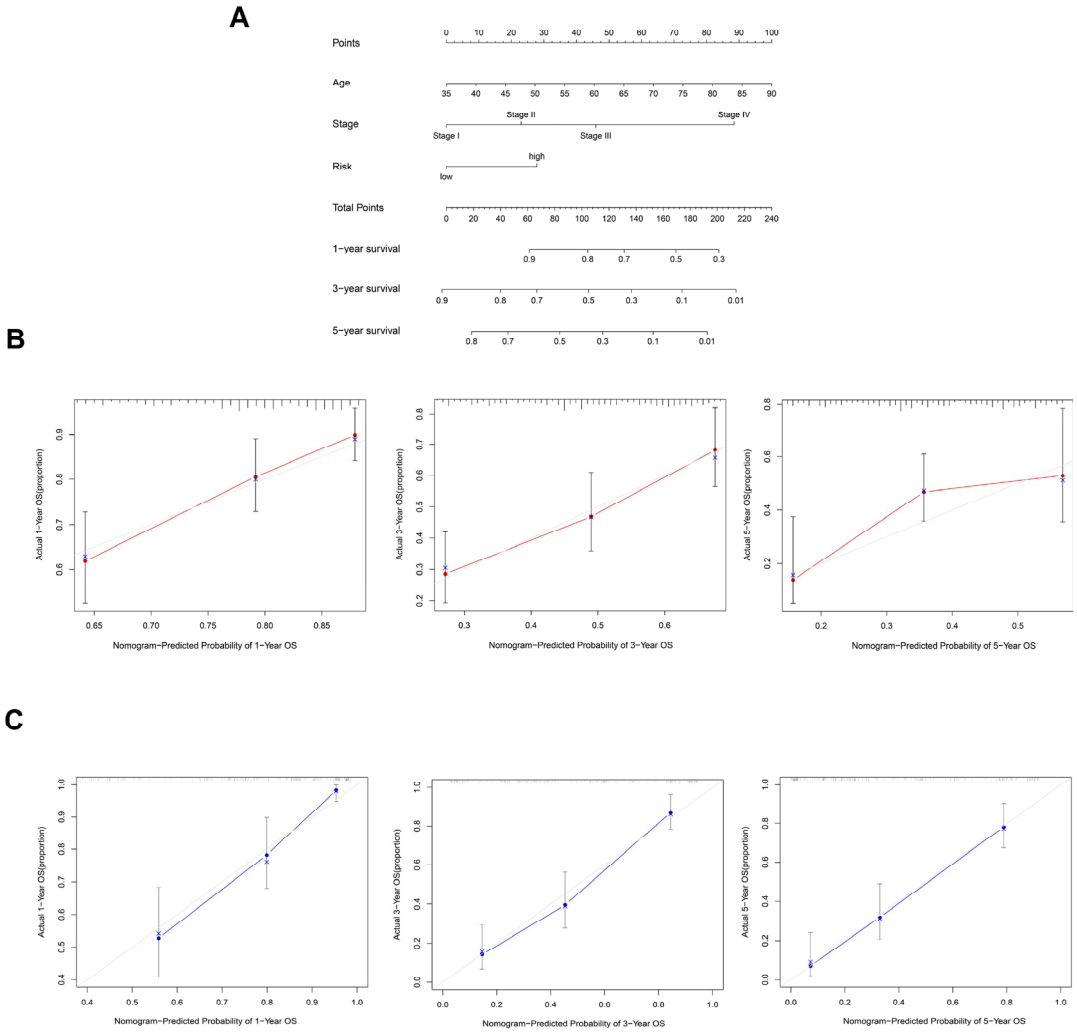

Figure 8. Construction and validation of the ERCPG nomogram. (A) Nomogram was constructed based on age, staging, and grouping by using the ERCPG risk score. (B) Calibration curves for the TCGA-STAD training set. (C) Calibration curves for the GSE15459 validation set.

3.7. Correlations between ERCPG risk scores and associated signaling pathways

We analyzed the signaling pathways related to ERCPG risk scores using the TCGA-STAD training set. Patients in the high-risk subgroup had a higher expression of genes that are closely associated with several pathway terms, including cell adhesion molecules (CAMs), ECM receptor interaction, focal adhesion, the Wnt signaling pathway, apical junction, the epithelial mesenchymal transition, and the TGF-β signaling pathway (Figure 9A and B). ECM receptor interaction, focal adhesion, and EMT were considerably enriched terms in the ERC subpopulation (cell clusters 4, 9 and 13) (Figure 9C). These results indicate that the ERCPG signature could play a role in promoting GC invasion through CAFs and ECs mediating the abovementioned pathways. Notably, the EMT signaling pathway was the most enriched in the CAF subpopulation among the other cell subpopulations, suggesting that CAFs may be a key cell subpopulation mediating the EMT signaling pathway in GC.

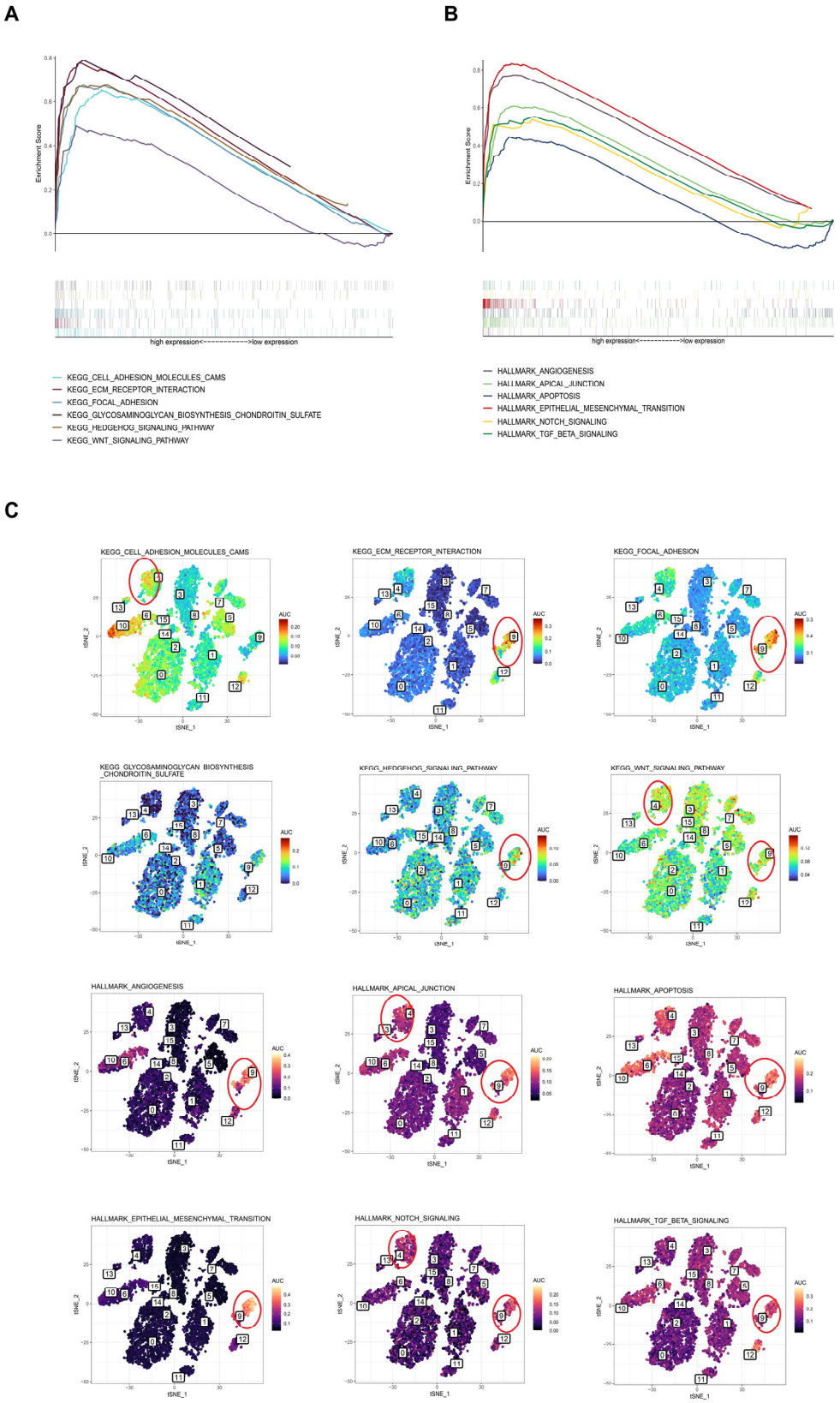

Figure 9. Functional enrichment analysis of the ERCPG signature. (A) Functional enrichment analysis based on the KEGG gene set. (B) Functional enrichment analysis based on the HALLMARK gene set. (C) Distribution of enriched pathways in the cell clusters.

3.8. Relationships Between ERCPGs and Immune-Cell Infiltration

Tumoral immune tolerance usually results from interactions between tumor-associated immune cells and cancer cells [37]. Regulation of the TME by immune cells can greatly affect the efficacy of cancer immunotherapy. Hence, we evaluated the relationship between ERCPG risk scores and immune-cell infiltration. As shown in Figure 10, significantly higher M2 macrophage infiltration ($P < 0.001$) was found in patients with GC with a high ERCPG risk score than in those with a low ERCPG risk score. The infiltration levels of plasma cells ($P = 0.013$), activated memory CD4 T cells ($P = 0.006$), and M1 macrophages ($P = 0.022$) were higher in low-risk patients with GC than in high-risk patients. Analysis of the seven ERCPGs and six types of immune cells in the TIMER database suggested that the expression of ERCPGs was closely correlated with immune cells in GC (Figure 11). The seven ERCPGs correlated significantly and positively with macrophages ($P < 0.01$).

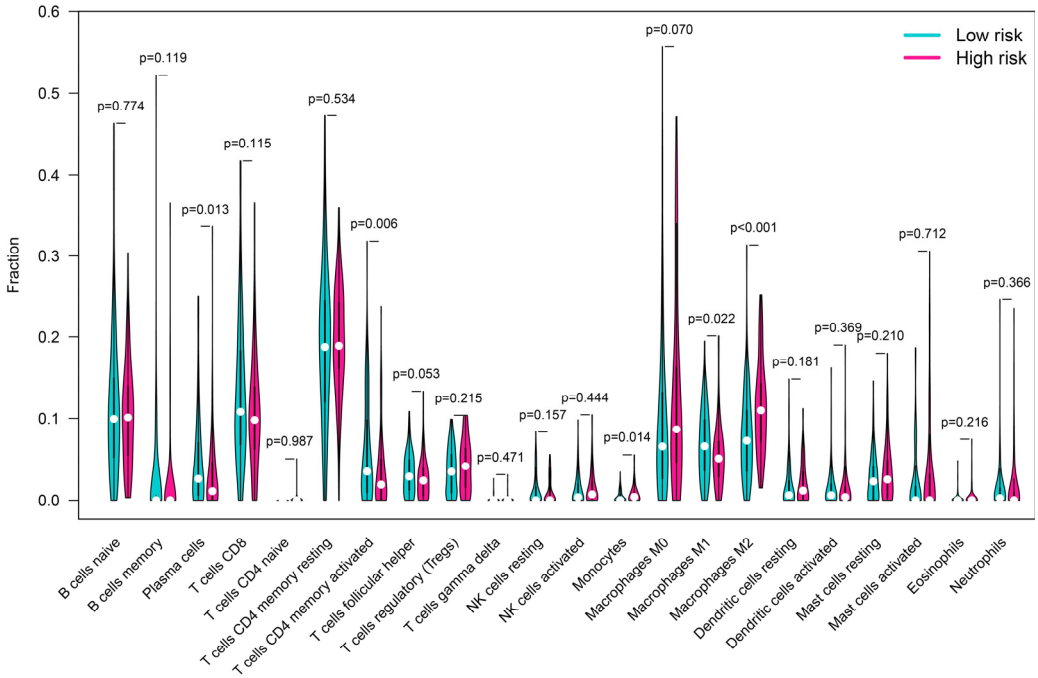

Figure 10. CIBERSORT algorithm-based investigation of the distribution of immune cells in the low-ERCPG risk and high-ERCPG risk subgroups.

3.9. Expression of ERCPGs in GC cell lines

Based on normal gastric cell lines and GC cell lines, we verified the expression of the ERCPG signature through in vitro experiments. A qRT-PCR analysis confirmed that COL4A1 was highly expressed in HGC-27 and BGC-823, F2R, MMP11, CAV1, and VCAN were highly expressed in MKN-45 and Hs-746T compared with GES-1. FKBP10 was expressed to a greater extent in all four GC cell lines than in normal gastric mucosal cells. APOD was highly expressed in MKN-45, HGC-27, and BGC-823 (Figure 12).

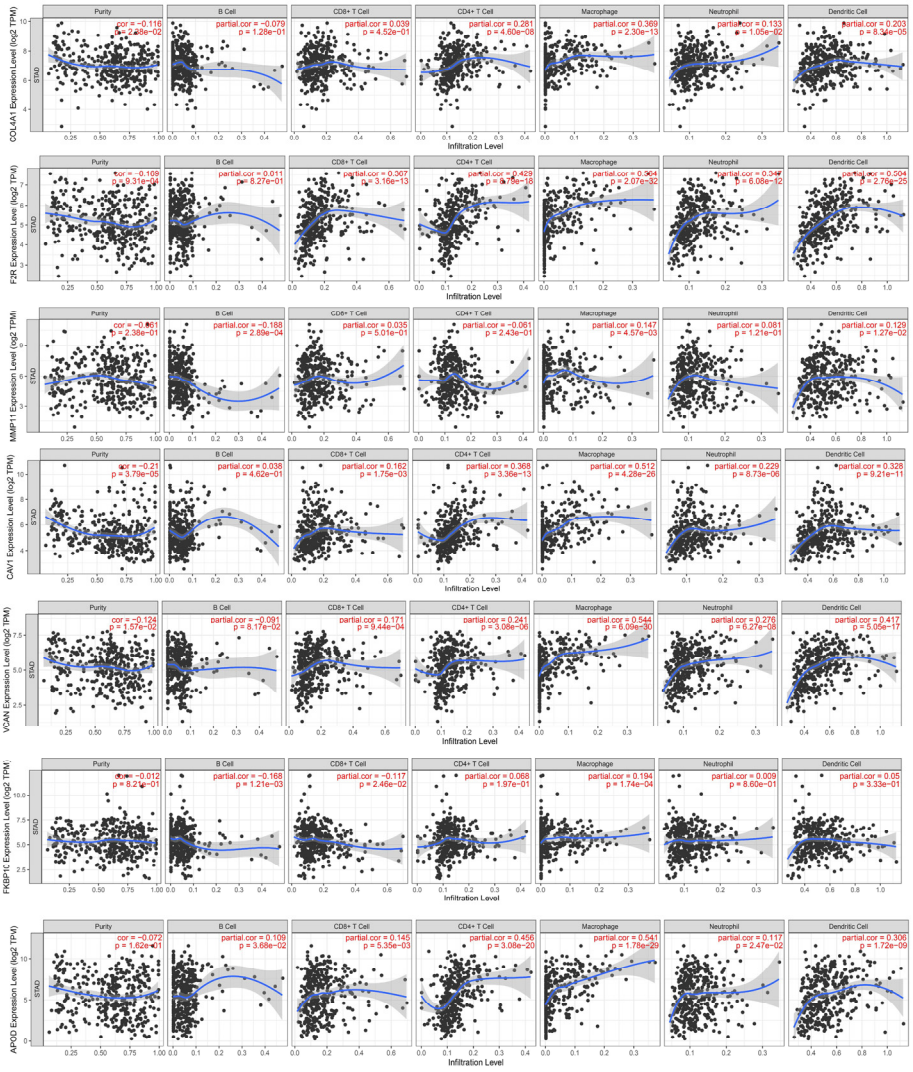

Figure 11. Correlation of the seven ERCPGs with immune cells, as determined using the TIMER database.

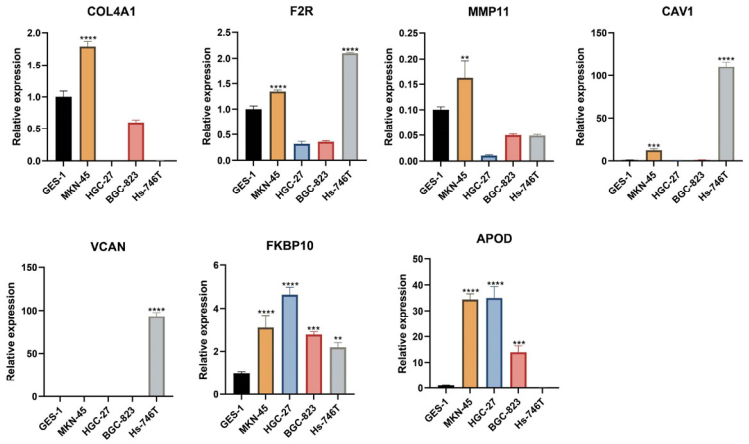

Figure 12. Differential expression of ERCPG signature in normal gastric cell lines and GC cell lines. (****:P < 0.0001,***:P < 0.001,**:P < 0.01,*:P < 0.05)

4. Discussion

Recently, the carcinogenic mechanism of EMT has been extensively investigated, and the roles of CAFs and ECs in the development of human malignant tumors have attracted increasing attention. As an important member of the stromal cell subpopulation in the TME, CAFs have inspired novel approaches related to the current development of targeted drugs for cancer immunotherapy [38]. ECs are particularly associated with the progression of epithelial malignancies induced by activation of the EMT program owing to their ability to transform into cells with different phenotypes under different cellular states [39]. Liu and Chu [19] reported that the E-cadherin expression of ECs may offer new opportunities for GC diagnosis and treatment. However, the prognostic utility of CAF- and EC-related biomarkers for patients with tumor remains not fully understood.

With the development of high-throughput sequencing technology, studies on malignant tumor biomarkers using omics technologies have also emerged. In particular, the rise of scRNA-seq technology compensates for the limitations of conventional high-throughput sequencing, providing an effective tool for analyzing the biological behaviors of tumors [40]. Prognostic models for predicting the OS of patients with cancer have been proposed based on bioinformatics investigations of EMT-related biomarkers. Zhang et al. [41] used 11 EMT-related genes to construct a prognostic model of colorectal cancer. EMT-related signals have also been applied to determine the prognosis of patients with hepatic cancer [42]. Dai et al. [43] developed five EMT-related prognostic signals for GC. Unlike these studies (which focused on conventional bulk sequencing data), this study performed an in-depth analysis of two types of ERCs (CAFs and ECs) at the cell-subpopulation level based on scRNA-seq data to identify prognostic biomarkers for GC.

To the best of our knowledge, this study was the first to analyze the characteristics of marker genes for CAFs and ECs in GC and construct an ERCPG signature for predicting the OS of patients with GC. First, we identified 740 marker genes for CAFs and ECs based on the scRNA-seq data from three primary GC tumor samples and performed a set-intersection analysis of the marker genes and 718 DEGs from the bulk data to obtain 115 DEMGs. Then, we performed univariate Cox regression and LASSO regression analyses to construct an ERCPG signature comprised of seven genes (COL4A1, F2R, MMP11, CAV1, VCAN, FKBP10, and APOD). Patients with GC were divided into high- and low-risk groups based on the ERCPG risk scores obtained using the TCGA-STAD training set and the GSE15459 validation set. A K-M analysis suggested that patients in the high-risk group had a shorter OS. In addition, the ERCPG risk scores remarkably distinguished between patients in the high- and low-risk groups based on the following clinical characteristics: age < 65 years; sex; grade 3, T3-4, N1-3, or M0 GC; and stage III-IV GC. Univariate and multivariate Cox analyses results suggested that the ERCPG risk scores were independent prognostic indicators. The ROC curves showed a good prognostic value for the ERCPG signature in predicting the 1, 3, and 5-year survival rates of patients with GC. Furthermore, to improve the efficiency of the ERCPG signature in predicting the OS of patients with GC, we constructed an ERCPG nomogram using independent prognostic factors. The calibration curves demonstrated good consistency between the predicted and actual values. Finally, cellular assays confirmed that ERCPGs were expressed in higher abundance in GC cell lines than in normal gastric cells.

We reviewed the literature regarding the seven ERCPGs. COL4A1 upregulation by lncRNA SND1-IT1 and miR-124 in GC promoted the occurrence of EMT [44]. In hepatic cancer, COL4A1 upregulation by RUNX1 facilitated the growth and invasion of tumor cells through the activation of

FAK-Src signaling. COL4A1 overexpression contributed to a poor prognosis in patients with hepatic cancer [45]. Moreover, COL4A1 was considerably upregulated in bladder cancer tissues, and COL4A1 produced by bladder cancer cells facilitated tumor budding and was suggestive of a poor prognosis [46,47], which is consistent with our results.

CAV1 expression was regulated by miR-103/107 (miR-103 and miR-107) and closely associated with multidrug resistance in GC [48]. In addition, CAV1 expression was upregulated in GC metastatic lymph nodes, which were indicative of a poor prognosis in patients with GC [49]. CAV1 was overexpressed in pulmonary cancer tissues, and its knockout in a non-small cell lung cancer cell line (A549) impaired the proliferative and invasive capacities of tumor cells [50]. These findings on CAV1 are consistent with our results.

The activation of F2R (also known as PAR1) by agonists enhanced the proliferative and invasive capacities of GC cells and induced the activation of other oncogenic factors, such as TN-C and NF-κB [51]. F2R can be overexpressed in triple-negative breast cancer due to the absence of ARRDC3 and can promote tumor progression by regulating WWTR1 and activating Hippo signaling [52]. Moreover, an elevated expression of F2R mediated by the TGFβ signaling pathway was observed in A549 lung cancer cells and might contribute to the EMT process [53]. Our results confirmed the malignant biological behavior of F2R in malignant tumors.

Furthermore, the proliferative capacity of GC cells was considerably impaired by MMP11 silencing in the GC cell line BGC823 [54]. Another GC study on MMP11 showed that its negative regulation by miR-139 suppressed GC progression [55]. Let-7c inhibited the growth and metastasis of colon cancer by downregulating MMP11 [56]. Moreover, MMP11 was highly expressed in breast cancer cells and clinical specimens, and its expression negatively correlated with the OS of patients with breast cancer; MMP11 can enhance the structural stability of Smad2 in the TGFβ signaling pathway, hence promoting breast cancer progression [57]. These findings are consistent with our findings on MMP11, indicating MMP11 as a potential therapeutic target for treating malignant tumors.

VCAN knockout in two GC cell lines (AGS and NCI-N87) remarkably inhibited the proliferation and migration of tumor cells [58]. Platelet-derived growth factors can regulate VCAN expression and thereby stimulate the proliferation of arterial smooth muscle cells; VCAN expression can promote ECM expansion [59], which might be associated with tumor invasion and metastasis. Snail protein in breast cancer cells mediated the overexpression of PAPSS2 and VCAN, subsequently disrupting ECM homeostasis in the tumor stroma, triggering EMT, and ultimately facilitating the invasion and metastasis of tumor cells [60]. Similarly, our results suggest that a high VCAN expression in patients with GC is correlated with a shorter OS.

FKBP10 was upregulated in GC cell lines and may promote GC progression by mediating the PI3K/AKT signaling pathway [61]. In contrast, its downregulation inhibited the progression of KRAS-driven lung cancer, and its overexpression was dependent on PPIase activity during oncogenesis [62]. Moreover, FKBP10 was upregulated in gliomas, stimulating the growth of tumor cells by mediating the AKT-CREB-PCNA signaling pathway [63]. Our results also suggest that low FKBP10 expression is associated with a good OS in patients with GC.

APOD upregulation was observed in drug-resistant breast cancer cells [64]. Ashida et al. [65] found that APOD was upregulated during the progression of invasive prostate cancer and was highly expressed in prostate cancer cells. Additionally, Vazquez et al. [66] detected higher APOD expression in ovarian cancer tissues than in normal ovaries. Interestingly, however, patients with ovarian cancer with high APOD expression had a better prognosis than those with low APOD expression. Our results

showed that the expression of APOD (a marker gene for ECs) in patients with GC was associated with a poor prognosis. Other bioinformatics analyses of APOD in GC also showed that APOD expression was a poor prognostic factor [67,68]; however, further relevant studies are still required to clarify its role in GC.

This is the first study to investigate biomarkers of ERCs using scRNA-seq data. We identified marker genes for CAFs (COL4A1, F2R, MMP11, CAV1, VCAN, and FKBP10) and a marker gene for ECs (APOD). Notably, the relationship between APOD and GC ECs is first reported in this study; further fundamental studies are urgently required to confirm this correlation. Results from enrichment analyses showed that 115 DEMGs were active in multiple pathways, including those related to focal adhesion, the PI3K-Akt pathway, and ECM-receptor interactions. The high-risk subgroup was enriched in several terms, including cell adhesion molecules (CAMs), TGF-β and Wnt signaling pathways, and cell-cluster graphs, suggesting that these pathways were strongly activated in the CAF or EC subpopulations. The correlation of the EMT with the PI3K-Akt and TGF-β signaling pathways in GC has been elucidated [69]. The ERCPG signature may mediate tumor invasion and metastasis via these pathways.

In this study, we performed immune-cell infiltration analysis based on the ERCPG signature and found that the infiltration level of M2 macrophages was elevated in high-risk patients with GC, whereas that of plasma cells, activated memory CD4 T cells, and M1 macrophages were elevated in low-risk patients with GC. The expression of all seven ERCPGs considerably and positively correlated with the extent of macrophage infiltration. M2 macrophages sensitized by mesenchymal stromal cells induced EMT in the GC TME [70]. Su et al. [71] suggested that breast cancer cells can secrete granulocyte macrophage colony-stimulating factor during EMT to induce macrophage infiltration and promote tumor metastasis. In addition, CCL18 derived from activated macrophages can induce EMT progression. IL-8 produced by hepatic cancer cells promoted the proliferation of M2 macrophages and facilitated the EMT program and tumor invasion [72].

Our results showed that all seven ERCPGs considerably and positively correlated with macrophage levels and the abnormal expression of ERCPGs in the high-risk subgroup, which may have promoted EMT by affecting the proliferation of M2 macrophages in GC, ultimately leading to a poor patient prognosisa. Wouters et al. [73] reported that infiltration of plasma cells into tumor tissues was indicative of a good prognosis, and the anti-tumor activity of activated memory T cells and M1 macrophages have been confirmed [74,75]. Hence, we deduced that the proliferation of plasma cells, activated memory T cells, and M1 macrophages might exhibit anti-tumor activity in patients with GC in the low-risk subgroup.

This study had some limitations. First, the scRNA-seq data and bulk data were obtained from public databases. Thus, more clinicopathological data is needed to analyze the prognostic value of ERCPGs. Future prospective studies should be conducted to confirm our results. In addition, more basic studies are still needed to identify the relationship between ERCPGs and cell subpopulations and their effects on the GC immune microenvironment.

5. Conclusions

We integrated scRNA-seq data from patients with GC with bulk sequencing data to construct an EMT-related prognostic signature comprised of seven genes (COL4A1, F2R, MMP11, CAV1, VCAN, FKBP10, and APOD) in TCGA-STAD training set. All seven ERCPGs were found to be closely related with GC survival and were effectively validated in the GSE15459 validation set and in vitro cell experiments. A functional enrichment analysis revealed that seven ERCPGs may promote GC

progression through signaling pathways such as CAMs, ECM receptor interaction, focal adhesion, and the Wnt signaling pathway. In addition, our findings lay a foundation for future research on the immune microenvironment of GC.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This study was supported by the Zhejiang Chinese Medicine "13th Five-Year" Key Discipline Construction Program [grant number 2017-XK-A12] and the General Project of Zhejiang Province Traditional Chinese Medicine Science and Technology Program [grant number 2021ZB128].

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, et al., Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA: Cancer J. Clin*., **71** (2021), 209–249. https://doi.org/10.3322/caac.21660
2. M. Das, Neoadjuvant chemotherapy: survival benefit in gastric cancer, *Lancet Oncol*., **18** (2017), 307. https://doi.org/10.1016/s1470-2045(17)30321-2
3. National Health Commission of the People's Republic of China, Chinese guidelines for diagnosis and treatment of gastric cancer 2018 (English version), *Chin. J. Cancer Res.*, **31** (2019), 707–737. https://doi.org/10.21147/j.issn.1000-9604.2019.05.01
4. M. Orditura, G. Galizia, V. Sforza, V. Gambardella, A. Fabozzi, M. M. Laterza, et al., Treatment of gastric cancer, *World J. Gastroenterol*., **20** (2014), 1635–1649. https://doi.org/10.3748/wjg.v20.i7.1635
5. M. P. Lutz, J. R. Zalcberg, M. Ducreux, J. A. Ajani, W. Allum, D. Aust, et al., Highlights of the EORTC St. Gallen International Expert Consensus on the primary therapy of gastric, gastroesophageal and oesophageal cancer-Differential treatment strategies for subtypes of early gastroesophageal cancer, *Eur. J. Cancer*, **48** (2012), 2941–2953. https://doi.org/10.1016/j.ejca.2012.07.029
6. I. Thomassen, Y. R. van Gestel, B. van Ramshorst, M. D. Luyer, K. Bosscha, S. W. Nienhuijs, et al., Peritoneal carcinomatosis of gastric origin: A population-based study on incidence, survival and risk factors, *Int. J. Cancer*, **134** (2014), 622–628. https://doi.org/10.1002/ijc.28373
7. H. Nakagawa, M. Fujita, Whole genome sequencing analysis for cancer genomics and precision medicine, *Cancer Sci*., **109** (2018), 513–522. https://doi.org/10.1111/cas.13505

8. A. Dongre, R. A. Weinberg, New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer, *Nat. Rev. Mol. Cell. Biol*., **20** (2019), 69–84. https://doi.org/10.1038/s41580-018-0080-4
9. M. Izumiya, A. Kabashima, H. Higuchi, T. Igarashi, G. Sakai, H. Iizuka, et al., Chemoresistance is associated with cancer stem cell-like properties and epithelial-to-mesenchymal transition in pancreatic cancer cells, *Anticancer Res*., **32** (2012), 3847–3853.
10. R. B. Hazan, R. Qiao, R. Keren, I. Badano, K. Suyama, Cadherin switch in tumor progression, *Ann. N. Y. Acad. Sci*., **1014** (2004), 155–163. https://doi.org/10.1196/annals.1294.016
11. C. Cai, J. Yu, J. Wu, R. Lu, X. Ni, S. Wang, et al., CD133 promotes the invasion and metastasis of gastric cancer via epithelial-mesenchymal transition, *Chin. J. Gastrointest. Surg.*, **16** (2013), 662–667.
12. C. Zeltz, I. Primac, P. Erusappan, J. Alam, A. Noel, D. Gullberg, Cancer-associated fibroblasts in desmoplastic tumors: emerging role of integrins, *Semin. Cancer Biol*., **62** (2020), 166–181. https://doi.org/10.1016/j.semcancer.2019.08.004
13. D. F. Quail, J. A. Joyce, Microenvironmental regulation of tumor progression and metastasis, *Nat. Med*., **19** (2013), 1423–1437. https://doi.org/10.1038/nm.3394
14. N. Kemi, M. Eskuri, A. Herva, J. Leppanen, H. Huhta, O. Helminen, et al., Tumour-stroma ratio and prognosis in gastric adenocarcinoma, *Br. J. Cancer*, **119** (2018), 435–439. https://doi.org/10.1038/s41416-018-0202-y
15. L. Huang, A. M. Xu, S. Liu, W. Liu, T. J. Li, Cancer-associated fibroblasts in digestive tumors, *World. J. Gastroenterol*., **20** (2014), 17804–17818. https://doi.org/10.3748/wjg.v20.i47.17804
16. A. C. Johansson, A. Ansell, F. Jerhammar, M. B. Lindh, R. Grenman, E. Munck-Wikland, et al., Cancer-associated fibroblasts induce matrix metalloproteinase–mediated cetuximab resistance in head and neck squamous cell carcinoma cells, *Mol. Cancer Res*., **10** (2012), 1158–1168. https://doi.org/10.1158/1541-7786.Mcr-12-0030
17. R. A. Saito, P. Micke, J. Paulsson, M. Augsten, C. Pena, P. Jonsson, et al., Forkhead box F1 regulates tumor-promoting properties of cancer-associated fibroblasts in lung cancer, *Cancer Res*., **70** (2010), 2644–2654. https://doi.org/10.1158/0008-5472.Can-09-3644
18. K. Pietras, K. Rubin, T. Sjoblom, E. Buchdunger, M. Sjoquist, C. Heldin, et al., Inhibition of PDGF receptor signaling in tumor stroma enhances antitumor effect of chemotherapy, *Cancer Res*., **62** (2002), 5476–5484.
19. X. Liu, K. M. Chu, E-Cadherin and gastric cancer: Cause, consequence, and applications, *BioMed Res. Int*., **2014** (2014), 637308. https://doi.org/10.1155/2014/637308
20. S. Herbertz, J. S. Sawyer, A. J. Stauber, I. Gueorguieva, K. E. Driscoll, S. T. Estrem, et al., Clinical development of galunisertib (LY2157299 monohydrate), a small molecule inhibitor of transforming growth factor-beta signaling pathway, *Drug Des*., *Dev. Ther.*, **9** (2015), 4479–4499. https://doi.org/10.2147/dddt.S86621
21. M. Singh, N. Yelle, C. Venugopal, S. K. Singh, EMT: Mechanisms and therapeutic implications, *Pharmacol. Ther*., **182** (2018), 80–94. https://doi.org/10.1016/j.pharmthera.2017.08.009
22. X. Zhang, S. L. Marjani, Z. Hu, S. M. Weissman, X. Pan, S. Wu, Single-cell sequencing for precise cancer research: Progress and prospects, *Cancer Res*., **76** (2016), 1305–1312. https://doi.org/10.1158/0008-5472.Can-15-1907

23. G. Sun, Z. Li, D. Rong, H. Zhang, X. Shi, W. Yang, et al., Single-cell RNA sequencing in cancer: Applications, advances, and emerging challenges, *Mol. Ther Oncolytics*., **21** (2021), 183–206. https://doi.org/10.1016/j.omto.2021.04.001
24. A. Sathe, S. M. Grimes, B. T. Lau, J. Chen, C. Suarez, R. J. Huang, et al., Single-cell genomic characterization reveals the cellular reprogramming of the gastric tumor microenvironment, *Clin. Cancer Res*., **26** (2020), 2640–2653. https://doi.org/10.1158/1078-0432.Ccr-19-3231
25. M. Zhang, S. Hu, M. Min, Y. Ni, Z. Lu, X. Sun, et al., Dissecting transcriptional heterogeneity in primary gastric adenocarcinoma by single cell RNA sequencing, *Gut*, **70** (2021), 464–475. https://doi.org/10.1136/gutjnl-2019-320368
26. Y. Li, X. Hu, R. Lin, G. Zhou, L. Zhao, D. Zhao, et al., Single-cell landscape reveals active cell subtypes and their interaction in the tumor microenvironment of gastric cancer, *Theranostics*, **12** (2022), 3818–3833. https://doi.org/10.7150/thno.71833
27. B. Wang, Y. Zhang, T. Qing, K. Xing, J. Li, T. Zhen, et al., Comprehensive analysis of metastatic gastric cancer tumour cells using single-cell RNA-seq, *Sci. Rep*., **11** (2021), 10. https://doi.org/10.1038/s41598-020-80881-2
28. A. C. Obenauf, J. Massague, Surviving at a distance: Organ-specific metastasis, *Trends Cancer*, **1** (2015), 76–91. https://doi.org/10.1016/j.trecan.2015.07.009
29. D. X. Nguyen, P. D. Bos, J. Massague, Metastasis: from dissemination to organ-specific colonization, *Nat. Rev. Cancer*, **9** (2009), 274–284. https://doi.org/10.1038/nrc2622
30. X. Zhang, Y. Lan, J. Xu, F. Quan, E. Zhao, C. Deng, et al., CellMarker: a manually curated resource of cell markers in human and mouse, *Nucleic. Acids. Res*., **47** (2019), 721–728. https://doi.org/10.1093/nar/gky900
31. T. Yan, W. Qiu, J. Song, Y. Fan, Z. Yang, ARHGAP36 regulates proliferation and migration in papillary thyroid carcinoma cells, *J. Mol. Endocrinol*., **66** (2021), 1–10. https://doi.org/10.1530/jme-20-0230
32. C. Han, T. Liu, R. Yin, Biomarkers for cancer-associated fibroblasts, *Biomark. Res*., **8** (2020). https://doi.org/10.1186/s40364-020-00245-w
33. S. Togo, U. M. Polanska, Y. Horimoto, A. Orimo, Carcinoma-associated fibroblasts are a promising therapeutic target, *Cancers*, **5** (2013), 149–169. https://doi.org/10.3390/cancers5010149
34. G. Corso, J. Figueiredo, S. P. De Angelis, F. Corso, A. Girardi, J. Pereira, et al., E-cadherin deregulation in breast cancer, *J. Cell. Mol. Med*., **24** (2020), 5930–5936. https://doi.org/10.1111/jcmm.15140
35. Y. A. Lyons, S. Y. Wu, W. W. Overwijk, K. A. Baggerly, A. K. Sood, Immune cell profiling in cancer: molecular approaches to cell-specific identification, *npj Precision Onc*., **1** (2017). https://doi.org/10.1038/s41698-017-0031-0
36. Z. Chen, Z. Han, H. Nan, J. Fan, J. Zhan, Y. Zhang, et al., A novel pyroptosis-related gene signature for predicting the prognosis and the associated immune infiltration in colon adenocarcinoma, *Front. Oncol*., **12** (2022). https://doi.org/10.3389/fonc.2022.904464
37. S. Han, K. Huang, Z. Gu, J. Wu, Tumor immune microenvironment modulation-based drug delivery strategies for cancer immunotherapy, *Nanoscale*, **12** (2020), 413–436. https://doi.org/10.1039/c9nr08086d

38. X. Geng, H. Chen, L. Zhao, J. Hu, W. Yang, G. Li, et al., Cancer-Associated Fibroblast (CAF) heterogeneity and targeting therapy of CAFs in pancreatic cancer, *Front. Cell Dev. Biol*., **9** (2021). https://doi.org/10.3389/fcell.2021.655152
39. L. A. Aparicio, M. Blanco, R. Castosa, A. Concha, M. Valladares, L. Calvo, et al., Clinical implications of epithelial cell plasticity in cancer progression, *Cancer Lett*., **366** (2015), 1–10. https://doi.org/10.1016/j.canlet.2015.06.007
40. T. Baslan, J. Hicks, Unravelling biology and shifting paradigms in cancer with single-cell sequencing, *Nat. Rev. Cancer*, **17** (2017), 557–569. https://doi.org/10.1038/nrc.2017.58
41. Z. Zhang, S. Zheng, Y. Lin, J. Sun, N. Ding, J. Chen, et al., Genomics and prognosis analysis of epithelial-mesenchymal transition in colorectal cancer patients, *BMC Cancer*, **20** (2020). https://doi.org/10.1186/s12885-020-07615-5
42. C. Xiong, G. Wang, D. Bai, A novel prognostic models for identifying the risk of hepatocellular carcinoma based on epithelial-mesenchymal transition-associated genes, *Bioengineered*, **11** (2020), 1034–1046. https://doi.org/10.1080/21655979.2020.1822715
43. W. Dai, Y. Xiao, W. Tang, J. Li, L. Hong, J. Zhang, et al., Identification of an EMT-related gene signature for predicting overall survival in gastric cancer, *Front. Genet*., **12** (2021). https://doi.org/10.3389/fgene.2021.661306
44. Y. Hu, Z. Hu, T. Liao, Y. Li, Y. Pan, LncRNA SND1-IT1 facilitates TGF-beta 1-induced epithelialto-mesenchymal transition via miR-124/COL4A1 axis in gastric cancer, *Cell Death Discov*., **8** (2022). https://doi.org/10.1038/s41420-021-00793-6
45. T. Wang, H. Jin, J. Hu, X. Li, H. Ruan, H. Xu, et al., COL4A1 promotes the growth and metastasis of hepatocellular carcinoma cells by activating FAK-Src signaling, *J. Exp. Clin. Cancer Res*., **39** (2020). https://doi.org/10.1186/s13046-020-01650-7
46. X. Xie, H. He, N. Zhang, X. Wang, W. Rui, D. Xu, et al., Overexpression of DDR1 promotes migration, invasion, though EMT-Related molecule expression and COL4A1/DDR1/MMP-2 signaling axis, *Technol. Cancer Res. Treat*., **19** (2020). https://doi.org/10.1177/1533033820973277
47. M. Miyake, Y. Morizawa, S. Hori, Y. Tatsumi, S. Onishi, T. Owari, et al., Diagnostic and prognostic role of urinary collagens in primary human bladder cancer, *Cancer Sci*., **108** (2017), 2221–2228. https://doi.org/10.1111/cas.13384
48. Y. Zhang, X. Qu, C. Li, Y. Fan, X. Che, X. Wang, et al., miR-103/107 modulates multidrug resistance in human gastric carcinoma by downregulating Cav-1, *Tumor Biol.*, **36** (2015), 2277–2285. https://doi.org/10.1007/s13277-014-2835-7
49. D. S. Sun, S. A. Hong, H. S. Won, S. H. Yoo, H. H. Lee, O. Kim, et al., Prognostic value of metastatic tumoral caveolin-1 expression in patients with resected gastric cancer, *Gastroenterol. Res. Pract.*, **2017** (2017). https://doi.org/10.1155/2017/5905173
50. W. Liu, N. Yin, H. Liu, K. Nan, Cav-1 promote lung cancer cell proliferation and invasion through lncRNA HOTAIR, *Gene*, **641** (2018), 335–340. https://doi.org/10.1016/j.gene.2017.10.070
51. D. Fujimoto, Y. Hirono, T. Goi, K. Katayama, S. Matsukawa, A. Yamaguchi, The activation of Proteinase-Activated Receptor-1 (PAR1) mediates gastric cancer cell proliferation and invasion, *BMC Cancer*, **10** (2010). https://doi.org/10.1186/1471-2407-10-443
52. A. K. S. Arakaki, W. Pan, H. Wedegaertner, I. Roca-Mercado, L. Chinn, T. S. Gujral, et al., α-Arrestin ARRDC3 tumor suppressor function is linked to GPCR-induced TAZ activation and breast cancer metastasis, *J. Cell Sci*., **134** (2021). https://doi.org/10.1242/jcs.254888

53. N. Smoktunowicz, M. Plate, A. O. Stern, V. D. Antongiovanni, E. Robinson, V. Chudasama, et al., TGF beta upregulates PAR-1 expression and signalling responses in A549 lung adenocarcinoma cells, *Oncotarget*, **7** (2016), 65471–65484. https://doi.org/10.18632/oncotarget.11472
54. H. Deng, R. Guo, W. Li, M. Zhao, Y. Lu, Matrix metalloproteinase 11 depletion inhibits cell proliferation in gastric cancer cells, *Biochem. Biophys. Res. Commun*., **326** (2005), 274–281. https://doi.org/10.1016/j.bbrc.2004.11.027
55. G. Xu, B. Zhang, J. Ye, S. Cao, J. Shi, Y. Zhao, et al., Exosomal miRNA-139 in cancer-associated fibroblasts inhibits gastric cancer progression by repressing MMP11 expression, *Int. J. Biol. Sci*., **15** (2019), 2320–2329. https://doi.org/10.7150/ijbs.33750
56. H. B. Han, J. Gu, H. Zuo, Z. Chen, W. Zhao, M. Li, et al., Let-7c functions as a metastasis suppressor by targeting MMP11 and PBX3 in colorectal cancer, *J. Pathol*., **226** (2012), 544–555. https://doi.org/10.1002/path.3014
57. Y. Zhuang, X. Li, P. Zhan, G. Pi, G. Wen, MMP11 promotes the proliferation and progression of breast cancer through stabilizing Smad2 protein, *Oncol. Rep*., **45** (2021). https://doi.org/10.3892/or.2021.7967
58. L. Zhai, W. Chen, B. Cui, B. Yu, Y. Wang, H. Liu, Overexpressed versican promoted cell multiplication, migration and invasion in gastric cancer, *Tissue Cell*, **73** (2021). https://doi.org/10.1016/j.tice.2021.101611
59. S. P. Evanko, P. Y. Johnson, K. R. Braun, C. B. Underhill, J. Dudhia, T. N. Wight, Platelet-derived growth factor stimulates the formation of versican-hyaluronan aggregates and pericellular matrix expansion in arterial smooth muscle cells, *Arch. Biochem. Biophys.*, **394** (2001), 29–38. https://doi.org/10.1006/abbi.2001.2507
60. Y. Zhang, X. Zou, W. Qian, X. Weng, L. Zhang, L. Zhang, et al., Enhanced PAPSS2/VCAN sulfation axis is essential for Snail-mediated breast cancer cell migration and metastasis, *Cell Death Differ*., **26** (2019), 565–579. https://doi.org/10.1038/s41418-018-0147-y
61. R. Wang, D. Zhang, C. Zhao, Q. Wang, H. Qu, Q. He, FKBP10 functioned as a cancer-promoting factor mediates cell proliferation, invasion, and migration via regulating PI3K signaling pathway in stomach adenocarcinoma, *Kaohsiung J. Med. Sci*., **36** (2020), 311–317. https://doi.org/10.1002/kjm2.12174
62. G. Ramadori, R. M. Ioris, Z. Villanyi, R. Firnkes, O. O. Panasenko, G. Allen, et al., FKBP10 regulates protein translation to sustain lung cancer growth, *Cell Rep*., **30** (2020), 3851–3863. https://doi.org/10.1016/j.celrep.2020.02.082
63. H. Cai, M. Zhang, Z. Cheng, J. Yu, Q. Yuan, J. Zhang, et al., FKBP10 promotes proliferation of glioma cells via activating AKT-CREB-PCNA axis, *J. Biomed. Sci*., **28** (2021). https://doi.org/10.1186/s12929-020-00705-3
64. R. Murad, A. Avanes, X. Ma, S. Geng, A. Mortazavi, J. Momand, Transcriptome and chromatin landscape changes associated with trastuzumab resistance in HER2+breast cancer cells, *Gene*, **799** (2021), 145808. https://doi.org/10.1016/j.gene.2021.145808
65. S. Ashida, H. Nakagawa, T. Katagiri, M. Furihata, M. Iiizumi, Y. Anazawa, et al., Molecular features of the transition from prostatic intraepithelial neoplasia (PIN) to prostate cancer: genome-wide gene-expression profiles of prostate cancers and PINs, *Cancer Res*., **64** (2004), 5963–5972. https://doi.org/10.1158/0008-5472.Can-04-0020

66. J. Vazquez, L. Gonzalez, A. Merino, F. Vizoso, Expression and clinical significance of apolipoprotein D in epithelial ovarian carcinomas, *Gynecol. Oncol.*, **76** (2000), 340–347. https://doi.org/10.1006/gyno.1999.5678
67. J. Huo, L. Wu, Y. Zang, Construction and validation of a universal applicable prognostic signature for gastric cancer based on seven immune-related gene correlated with tumor associated macrophages, *Front. Oncol*., **11** (2021). https://doi.org/10.3389/fonc.2021.635324
68. X. Guo, X. Liang, Y. Wang, A. Cheng, H. Zhang, C. Qin, et al., Significance of tumor mutation burden combined with immune infiltrates in the progression and prognosis of advanced gastric cancer, *Front. Genet*., **12** (2021). https://doi.org/10.3389/fgene.2021.642608
69. Z. Peng, C. Wang, E. Fang, G. Wang, Q. Tong, Role of epithelial-mesenchymal transition in gastric cancer initiation and progression, *World J. Gastroenterol*., **20** (2014), 5403–5410. https://doi.org/10.3748/wjg.v20.i18.5403
70. W. Li, X. Zhang, F. Wu, Y. Zhou, Z. Bao, H. Li, et al., Gastric cancer-derived mesenchymal stromal cells trigger M2 macrophage polarization that promotes metastasis and EMT in gastric cancer, *Cell Death Dis*., **10** (2019). https://doi.org/10.1038/s41419-019-2131-y
71. S. Su, Q. Liu, J. Chen, J. Chen, F. Chen, C. He, et al., A positive feedback loop between mesenchymal-like cancer cells and macrophages is essential to breast cancer metastasis, *Cancer Cell*, **25** (2014), 605–620. https://doi.org/10.1016/j.ccr.2014.03.021
72. P. Xiao, X. Long, L. Zhang, Y. Ye, J. Guo, P. Liu, et al., Neurotensin/IL-8 pathway orchestrates local inflammatory response and tumor invasion by inducing M2 polarization of Tumor-Associated macrophages and epithelial-mesenchymal transition of hepatocellular carcinoma cells, *OncoImmunology*, **7** (2018). https://doi.org/10.1080/2162402x.2018.1440166
73. M. C. A. Wouters, B. H. Nelson, Prognostic significance of tumor-infiltrating B cells and plasma cells in human cancer, *Clin. Cancer Res*., **24** (2018), 6125–6135. https://doi.org/10.1158/1078-0432.Ccr-18-1481
74. C. Gu-Trantien, S. Loi, S. Garaud, C. Equeter, M. Libin, A. de Wind, et al., CD4(+) follicular helper T cell infiltration predicts breast cancer survival, *J. Clin. Investigation*, **123** (2013), 2873–2892. https://doi.org/10.1172/jci67428
75. G. R. Gunassekaran, S. M. P. Vadevoo, M. Baek, B. Lee, M1 macrophage exosomes engineered to foster M1 polarization and target the IL-4 receptor inhibit tumor growth by reprogramming tumor-associated macrophages into M1-like macrophages, *Biomaterials*, **278** (2021), 121137. https://doi.org/10.1016/j.biomaterials.2021.121137

AIMS Press