



CXCR4 is a Novel Biomarker Correlated With Malignant Transformation and Immune Infiltrates in Gastric Precancerous Lesions

Xiaotao Jiang^{1,2†}, Junhui Zheng^{1,2†}, Lanxing Liu^{1,2}, Kailin Jiang^{1,2}, Yi Wen¹, Yanhua Yan^{1,2}, Yufeng Liu³, Limei Zhong⁴, Yuancheng Huang^{1,2}, Zhengyang Yao², Kechao Nie^{1,2}, Zhihua Zheng^{1,2}, Jinglin Pan⁵, Peng Liu^{1,2}, Kunhai Zhuang^{1,6*}, Fengbin Liu^{1,6,3*}, Shijie Xu^{7*} and Peiwu Li^{1*}

OPEN ACCESS

Edited by:

Peng Li,
National Heart, Lung, and Blood
Institute (NHLBI), United States

Reviewed by:

Tae Jin Lee,
Augusta University, United States
Sayan Chakraborty,
Institute of Molecular and Cell Biology
(A*STAR), Singapore

*Correspondence:

Kunhai Zhuang
sea-zhuang007@163.com
Fengbin Liu
liufengbin163@163.com
Shijie Xu
xsj@gzucm.edu.cn
Peiwu Li
doctorlipw@gzucm.edu.cn

[†]These authors share first authorship

Specialty section:

This article was submitted to
Molecular Diagnostics and
Therapeutics,
a section of the journal
Frontiers in Molecular Biosciences

Received: 20 April 2021

Accepted: 23 August 2021

Published: 05 October 2021

Citation:

Jiang X, Zheng J, Liu L, Jiang K,
Wen Y, Yan Y, Liu Y, Zhong L, Huang Y,
Yao Z, Nie K, Zheng Z, Pan J, Liu P,
Zhuang K, Liu F, Xu S and Li P (2021)
CXCR4 is a Novel Biomarker
Correlated With Malignant
Transformation and Immune Infiltrates
in Gastric Precancerous Lesions.
Front. Mol. Biosci. 8:697993.
doi: 10.3389/fmolb.2021.697993

¹Department of Gastroenterology, The First Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou, China, ²First Clinical Medical College, Guangzhou University of Chinese Medicine, Guangzhou, China, ³Lingnan Medical Research Center, Guangzhou University of Chinese Medicine, Guangzhou, China, ⁴Department of Laboratory Medicine, Guangdong Second Provincial General Hospital, Guangzhou, China, ⁵Department of Gastroenterology, Hainan Provincial Hospital of Traditional Chinese Medicine, Haikou, China, ⁶Baiyun Hospital of The First Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou, China, ⁷Guangzhou University of Chinese Medicine, Guangzhou, China

Background: As early gastric cancer (EGC) has a far better prognosis than advanced gastric cancer (GC), early diagnosis and treatment are essential. However, understanding the mechanism of the process from gastric precancerous lesion (GPL) becoming EGC has made little advances. Besides, biomarkers that can monitor the progression of GPL-to-GC are still much insufficient.

Methods: Key gene modules associated with GPL progression to EGC were identified by integrating two GPL-related data sets, GSE55696 and GSE130823, using the WGCNA method. Combining with the TCGA-STAD cohort, hub genes were identified. Immunofluorescence was conducted to validate the expression. To explore the implication of hub genes in GPL malignant transformation, a correlation test was conducted to identify their co-expression genes, co-expression cytokines, and co-expression immune cells. Least absolute shrinkage and selection operator (LASSO) Cox regression was applied to shrink CXCR4-related predictors and construct a prognostic model. Functional enrichment was applied for exploring the potential mechanism.

Results: The green module in GSE55696 and the yellow module in GSE130823 were regarded as key gene modules associated with GPL progression to EGC, and 219 intersection genes from them were mainly enriched in critical immune biological processes. Combining with the TCGA-STAD cohort, CXCR4 was identified as a novel biomarker correlated with the malignant transformation of GPL, the positive rate of which was increased with GPL progression according to immunofluorescence. CXCR4 co-expression genes were found mainly involved in regulation of actin. CXCR4 co-expression cytokines were enriched in regulation of chemotaxis, cell chemotaxis, mononuclear cell migration, leukocyte chemotaxis, etc. As for co-expression immune

cells, the expression level of CXCR4 was positively correlated with the abundance of macrophages but negatively correlated with that of effector memory T cells and NKT cells during GPL malignant transformation. In addition, the CXCR4-related prognostic model was able to predict the prognosis of GC and serve as an independent predictor for overall survival (OS).

Conclusions: CXCR4 was a novel biomarker correlated with malignant transformation of GPL and played a vital role in the control of tumor immunity. CXCR4 is possible to serve as a therapeutic target for malignant transformation of GPL.

Keywords: CXCR4, gastric precancerous lesions, gastric cancer, malignant transformation, immune infiltrates

INTRODUCTION

Gastric cancer (GC) is the fifth most common malignancy and the third leading cause of cancer-related mortality (Freddie et al., 2018). Histologically, Lauren categorized GC into two subtypes: intestinal and diffuse types (Laurén, 1965). Intestinal-type GC (IGC), with a clear and multistep histological evolution starting from chronic inflammation and progressing to atrophy, intestinal metaplasia, gastric precancerous lesion [GPL, including low-grade intraepithelial neoplasia (LGIN) and high-grade intraepithelial neoplasia (HGIN)], and frank malignancy (Correa, 1988), and diffuse-type GC (DGC), without having to go through the Correa pathway, directly become cancerous through highly active inflammation (Nardone et al., 2004; Watanabe et al., 2012). As early gastric cancer (EGC) has a far better prognosis than advanced GC, early diagnosis and treatment are essential (Gotoda et al., 2000). However, biomarkers that can monitor the progression of GPL-to-GC are still much insufficient. The relatively fixed evolutionary paradigm of IGC makes it possible to develop biomarkers to monitor the early onsets.

With the development of public databases such as the Gene Expression Omnibus (GEO) database, huge amounts of genomic, epigenomic, transcriptional, and proteomic data become easier to access (Barrett et al., 2013). Correspondingly, various algorithms designed for analyzing omics data have also been developed over the past few years (Lin and Lane, 2017), which provide more opportunities to explore the molecular mechanism and biomarkers of carcinogenesis.

Recently, cancer immunotherapy has been a research “hot spot”. CXC chemical receptors (CXCRs), as a key component of the immune system, have received increasing attention. Belonging to a subfamily of G-protein coupled receptors, CXCR family members (comprising CXCR1-7) serve as crucial regulators of cancer progression through binding to the corresponding ligands (Ben-Baruch, 2006; Lee et al., 2014). CXCRs and their ligands could facilitate tumor cell activation, proliferation, invasion, and migration (Zhu et al., 2012). In addition, CXCR-associated angiogenesis and tumor infiltrating immune cells also attracted more and more attention.

Herein, we carried out a series of bioinformatics analysis on GPL and GC-related data sets and identified CXCR4 as a novel biomarker correlated with malignant transformation, which was

also validated by immunofluorescence. We also explored the immune implication of CXCR4 in GPL malignant transformation and constructed a CXCR4-related prognostic model.

METHODS

Data Acquisition and Preprocessing

Raw data of two GPL-related gene expression profiles [GSE55696 (Xu et al., 2014) and GSE130823 (Zhang et al., 2020)] were downloaded from the GEO database on NCBI (www.ncbi.nlm.nih.gov/geo/). Then, they were processed and normalized using tools from the limma package (Smyth et al., 2010) (the R code was provided in the supplementary file). The annotation function in Sangerbox (<http://soft.sangerbox.com/>) software was used to annotate the probe. GSE55696 performed on an Agilent-014850 Whole Human Genome Microarray 4 × 44K G4112F includes 19 LGIN, 20 HGIN, 19 EGC, and 19 chronic gastritis (CG) samples. GSE130823 performed on an Agilent-039494 SurePrint G3 Human GE v2 8 × 60K Microarray 039381 includes 17 LGIN, 14 HGIN, 15 EGC, and 47 paired inflammation controls. In order to investigate the difference between individuals from different pathological stages, data of 47 paired inflammation controls in GSE130823 were excluded. As stomach adenocarcinoma (STAD) is the most common form of GC and regarded as the final stage of the Correa cascade (Bockerstett and Dipaolo, 2017), RNA sequencing data of STAD samples with clinical information in TCGA were adopted and downloaded from the UCSC Xena browser (<https://xenabrowser.net/>) (Goldman et al., 2015). After removing samples with 0-day follow-up duration and incomplete clinical information, 350 STAD samples and 32 adjacent samples were obtained. Among 350 STAD samples, 46 were in Stage I, 110 were in Stage II, 145 were in Stage III, 35 were in Stage IV, and 14 were unknown.

Identification of Key Gene Modules Associated With GPL Progression to EGC

To identify key gene modules associated with GPL progression, weighted gene co-expression analysis (WGCNA) was used to construct a co-expressed gene network and explore its correlation

with clinical traits. First, through utilizing the “WGCNA” package in R (Langfelder and Horvath, 2008), top 25% genes with the largest variance differences were used to construct weight gene co-expression networks in GSE55696 and GSE130823 data sets. Then, a similarity matrix was established based on the Pearson’s correlation value between the paired genes. Next, an adjacency matrix was created using the formula (Zhang and Horvath, 2005) $a_{mn} = |c_{mn}|^\beta$ (a_{mn} : adjacency matrix between gene m and gene n , c_{mn} = Pearson’s correlation between paired genes, β : soft power value) and converted into a topological overlap matrix. The value of soft threshold power was chosen with the scale-free topology scale R^2 exceeding 0.85. Genes with similar expression patterns were categorized into modules by average linkage hierarchical clustering with the module minimum size set as 100, and colors were used to label modules. Those genes that cannot be assigned to any of the modules were placed in a gray module. That is, the genes in the gray module were not co-expressed. To explore the correlation between gene modules and clinical traits, we defined “Stage” representing the severity of pathology, in which CG was represented by “1,” LGIN by “2,” HGIN by “3,” and EGC by “4.” Then, the correlation between module eigengenes (MEs) and the clinical traits including “Stage” were calculated using Pearson’s correlation test in GSE55696 and GSE130823 data sets, with a p -value < 0.05 as the cutoff. Those gene modules harboring the highest correlation coefficient with “Stage” were considered as key gene modules associated with GPL progression to EGC.

Identification of Novel Genes Associated With GPL Progression

Among the intersecting genes of key gene modules from GSE55696 and GSE130823 data sets, we regarded that novel genes should be differentially expressed between the tumor and adjacent normal tissue and significantly correlated with prognosis. Therefore, novel genes were identified with the following three criteria met meanwhile in the TCGA-STAD cohort: 1) significantly associated with survival (filtered by Kaplan–Meier (K-M) survival curves utilizing the “survival” package with a Log-rank p -value < 0.05 as the cutoff), 2) harboring a predictive ability for overall survival (OS) with the area under the curve (AUC) ≥ 0.60 (utilizing the “timeROC” package), 3) differently expressed between the tumor and adjacent normal tissue (utilizing the linear model and the empirical Bayes method in the “limma” package with $|\log_{2}FC| \geq 0.585$ and a false discovery rate (FDR) < 0.05 as the cutoff).

Patients and Stomach Tissues

A total of 13 stomach tissue samples including 4 CG, 4 LGIN, 3 HGIN, and 3 GC were obtained from the patients treated at the First Affiliated Hospital of Guangzhou University of Chinese Medicine. All these samples were stored at -80°C . All patients signed informed consents before operation. The study was approved by the Ethics Committee of the First Affiliated Hospital of Guangzhou University of Chinese Medicine.

Validation of CXCR4 Expression Based on Immunofluorescence

Stomach tissues were fixed in 4% paraformaldehyde, embedded in paraffin, and sectioned into five μm slices for the immunofluorescence assay. The slice was then put into dimethylbenzene xylene I for 15 min, dimethylbenzene xylene II for 15 min, ethanol I for 5 min, anhydrous ethanol II for 5 min, 75% alcohol for 5 min, 85% alcohol for 5 min, and distilled water in turn. Then, the slice was put into a repair box which was filled with the ethylenediaminetetraacetic acid antigen repair buffer (pH 8.0). The antigen repair was performed in a microwave with the medium boiling for 5 min. After natural cooling, the slice was placed into phosphate-buffered saline (PBS) (pH 7.4) and washed on a decoloring table three times, 5 min each time. After spin-drying, a circle was drawn on the slice with a Pap Pen to prevent the antibody from running away. Dropping 3% bovine serum albumin in the circle, the tissue was covered and closed for 30 min at room temperature. The sealing solution was gently shaken off, PBS was dropped on the slice to prepare the first antibody in a certain proportion, and the slice was laid flat in the wet box at 4°C for incubation overnight. Then, the slice was put into PBS (pH 7.4) for decoloration (washing three times, 5 min each time). After slightly drying, the secondary antibody with the corresponding species of the primary antibody was added dropwise in the circle to cover the tissue and incubated for 50 min at room temperature in the dark. The slide was put in PBS (pH 7.4), and it was shaken on the bleaching shaker three times, each time for 5 min. After spin-drying, the slice was added with 4,6-diamidino-2-phenylindole and incubated for 10 min at room temperature in the dark. After being washed three times and sealed, the slice was inverted on a fluorescence microscope to collect images. The stained sections were scanned with a panoramic scanner (panoramic MIDI II, 3DHISTH). Quantification of the immunostained images generated was carried out using HistoQuant software (Quant Center 2.0 software, 3DHISTH). The percentage of the positively stained area was obtained by dividing the stained area by the total tissue area, multiplied by 100.

Identification of CXCR4 Co-expression Genes

The Spearman correlation test was conducted to identify genes significantly correlated with CXCR4 expression in GPL-related cohorts (GSE55696 and GSE130823) and the GC cohort (TCGA-STAD), with a p -value < 0.05 as the cutoff. Intersecting CXCR4-related genes from the above three cohorts were considered as crucial CXCR4 co-expression genes.

Identification of CXCR4 Co-expression Cytokines

Annotation 456 cytokines were downloaded from ImmPort Portal (<https://www.immport.org/>). Gene expression data of cytokines in GSE55696, GSE130823, and TCGA-STAD cohorts

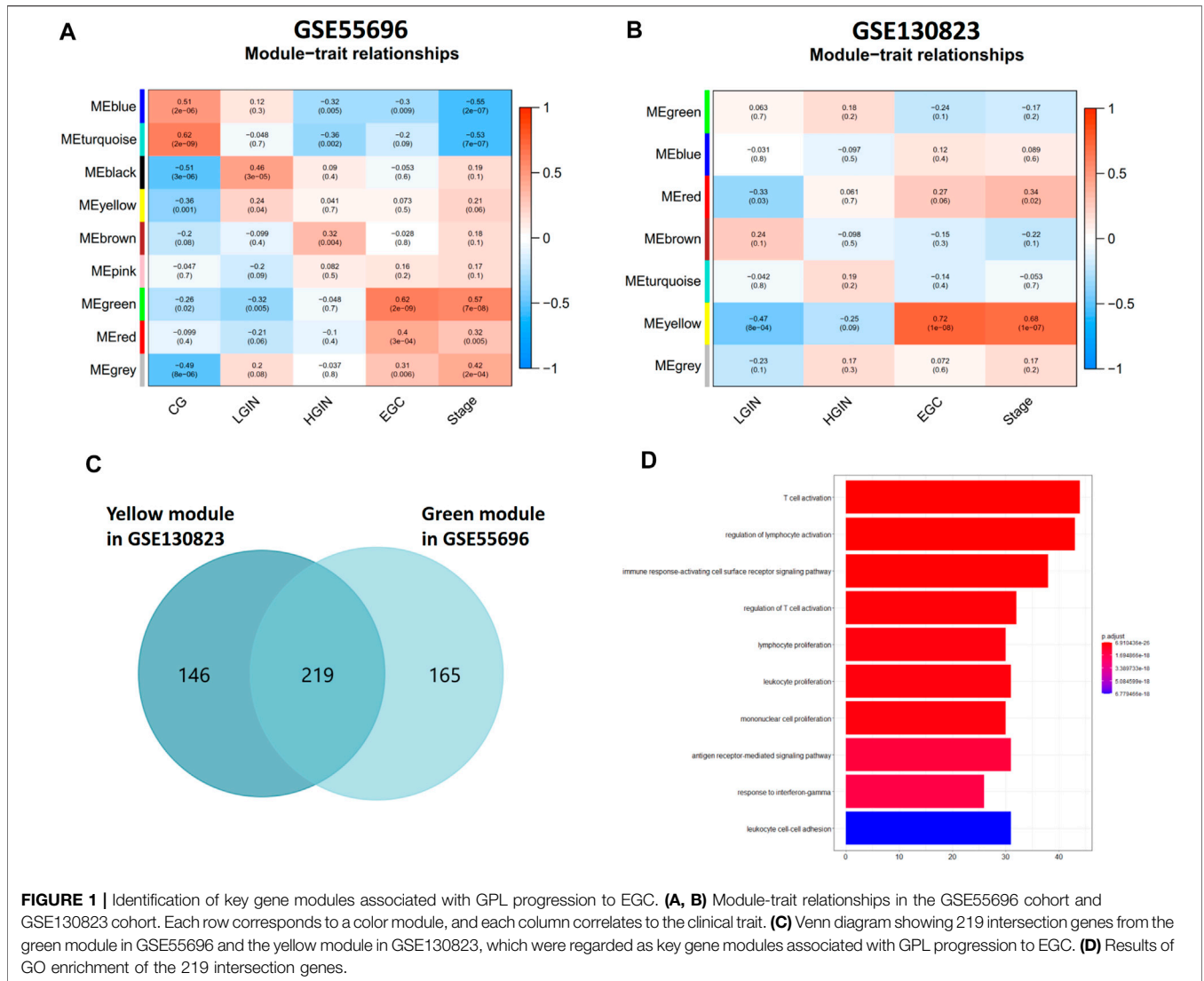


FIGURE 1 | Identification of key gene modules associated with GPL progression to EGC. **(A, B)** Module-trait relationships in the GSE55696 cohort and GSE130823 cohort. Each row corresponds to a color module, and each column correlates to the clinical trait. **(C)** Venn diagram showing 219 intersection genes from the green module in GSE55696 and the yellow module in GSE130823, which were regarded as key gene modules associated with GPL progression to EGC. **(D)** Results of GO enrichment of the 219 intersection genes.

were obtained, and the correlations with the CXCR4 expression level were calculated based on the Spearman correlation test with a *p*-value < 0.05 as the cutoff. Similarly, intersecting co-expressed cytokines from the above three cohorts were regarded as crucial CXCR4 co-expression cytokines.

Identification of CXCR4 Co-expression Immune Cells

The Immune Cell Abundance Identifier (ImmuCellAI, <http://bioinfo.life.hust.edu.cn/web/ImmuCellAI/>) is a gene set signature-based method for precisely estimating the abundance of 24 immune cell types, especially T-cell properties (18 T-cell subsets) (Miao et al., 2020). It is reported that it has the best performance in immune cell abundance estimation (Miao et al., 2020), compared with other five methods [xCell (Aran et al., 2017), CIBERSORT (Newman et al., 2015), EPIC (Racle et al., 2017), MCP-counter (Becht et al., 2016), and TIMER (Li et al., 2017)]. Gene expression data of GSE55696, GSE130823, and

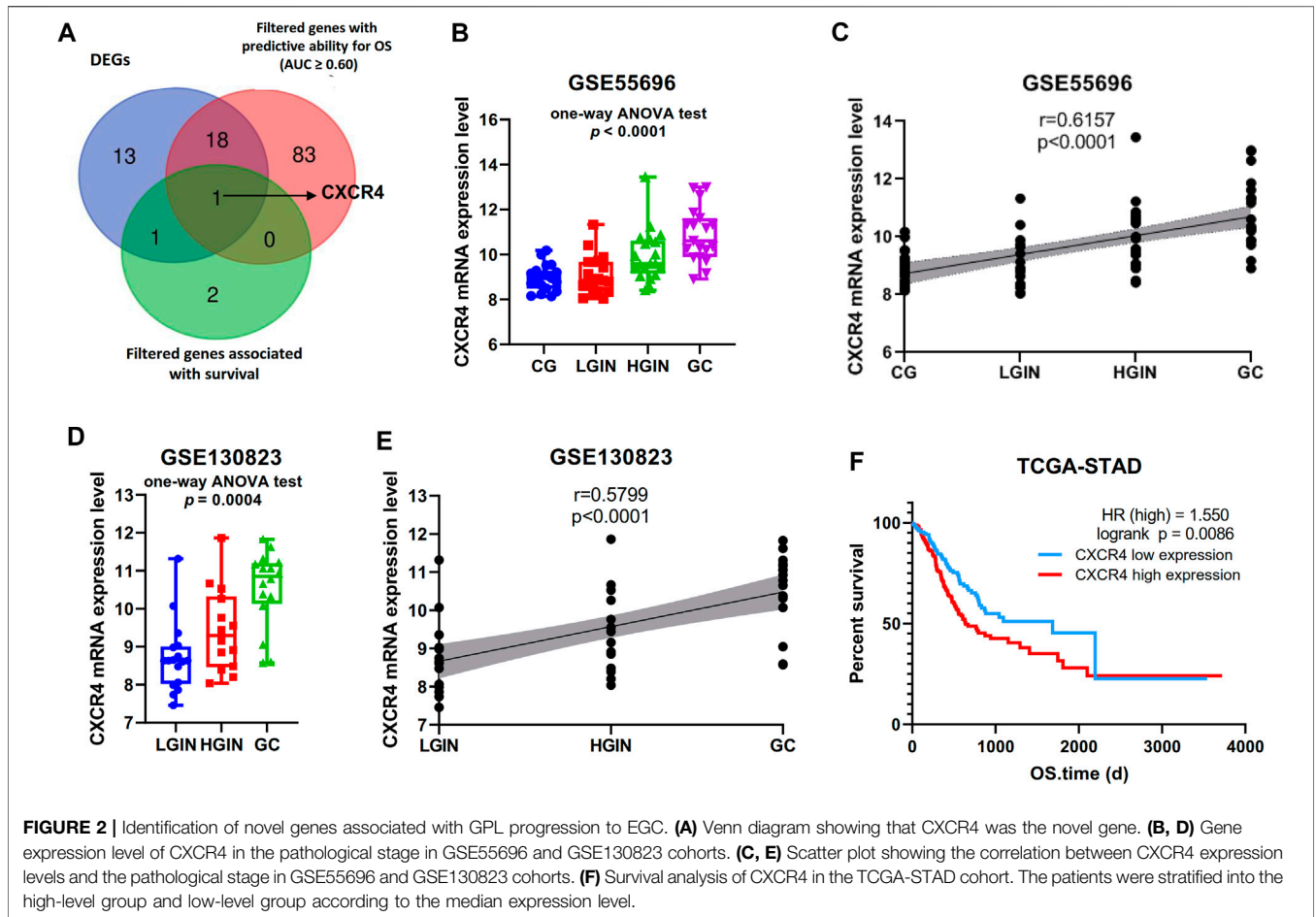
TCGA-STAD cohorts were submitted to ImmuCellAI to acquire the estimation of immune cell abundance. Likewise, the Spearman correlation test was also conducted to identify immune cells co-expressed with CXCR4 based on a *p*-value < 0.05 as the cutoff. Intersecting co-expressed immune cells of CXCR4 from the above three cohorts were regarded as crucial CXCR4 co-expression immune cells.

Construction of the Protein-Protein Interaction (PPI) Network

Target genes were uploaded to STRING (<http://string-db.org>) to acquire the information about interactions, with a limitation of the species to “*Homo sapiens*” and a confidence score greater than 0.4.

Functional Enrichment Analysis

The enrichment analysis of Gene Ontology (GO) was carried out according to target genes to explore potential molecular



mechanisms by utilizing the “clusterProfiler” R package. The *p*-values are adjusted using the BH method to control the FDR.

Construction of a CXCR4-Related Prognostic Model

In the TCGA-STAD cohort, the LASSO Cox regression analysis by the R package “glmnet” was performed to select the valuable prognostic markers from the co-expression genes of CXCR4 and the optimal penalty parameter (λ) was estimated by tenfold cross-validation to prevent overfitting. The risk score of each sample was determined by the following formula: risk score $\propto e^{\sum(\text{each gene's expression} \times \text{corresponding coefficient})}$. The patients were separated into high- and low-risk groups based on the optimal cutoff values. A K-M survival curve and time-dependent ROC analysis were conducted to compare the survival and evaluate the model’s predictive ability, respectively. Survival analysis was implemented within the “survivalROC” R package. Univariate and multivariate Cox regression analyses were used to determine the independent predictors for OS.

Cell Culture

Purchased from the American Type Culture Collection (ATCC, Manassas, VA, USA), GC cell line AGS and normal human

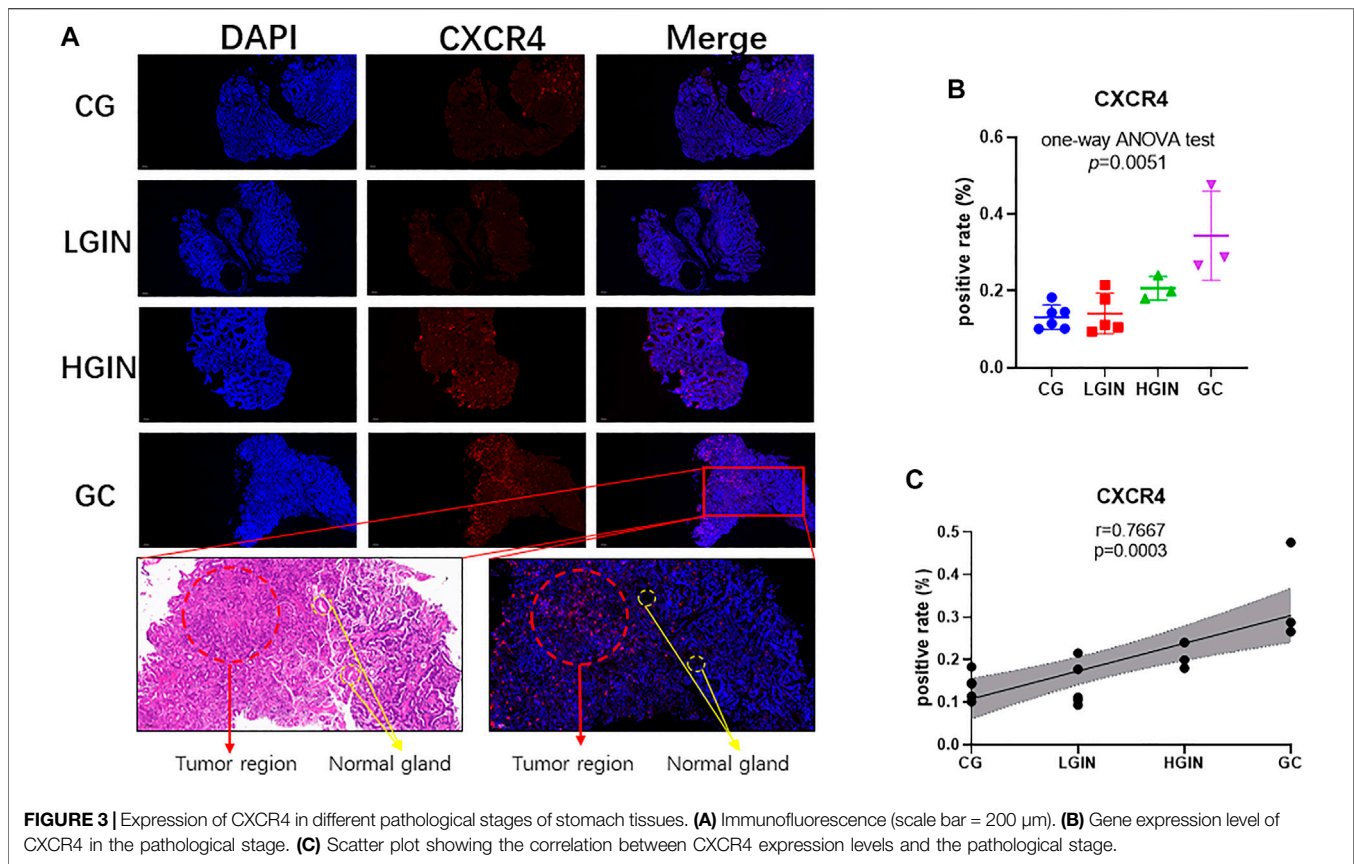
gastric epithelial cell line GES-1 were cultured in the RPMI-1640 Medium (Life Technologies, Grand Island, NY, USA) supplemented with 10% fetal bovine serum (Life Technologies) at 37°C in a humidified atmosphere with 5% CO₂.

Quantitative Real-Time Polymerase Chain Reaction

Total RNA was extracted from cells with the TRIzol reagent (Invitrogen, China) according to the manufacturer’s protocol. Reverse transcription was performed according to the manufacturer’s instructions using a PrimeScript RT Reagent Kit (Takara, China). A SYBR PrimeScript RT-PCR Kit (Takara) was applied for the analysis of quantitative reverse transcription-polymerase chain reaction (qRT-PCR). The $2^{-\Delta\Delta C_t}$ statistic was used to calculate the expression levels of genes.

Statistical Analysis

Student’s *t* test or the Mann–Whitney test was used to compare gene expression levels of different groups depending on the distribution of data. Pearson’s or Spearman’s correlation tests were performed in order to evaluate statistical correlation depending on the distribution of data. One-way analysis of



variance (ANOVA) was used to perform a comparison for continuous variables among groups ≥ 3 . The OS between groups was compared by using the Kaplan–Meier analysis with the log-rank test. Also, the identification of independent predictors of OS was conducted by the analysis of univariate and multivariate Cox regression. All statistical analyses were performed with R software (Version 3.6.3) or GraphPad Prism software (Version 8.0). All p -values are two-tailed, and a p -value less than 0.05 was considered statistically significant.

RESULTS

Key Gene Modules Associated With GPL Progression to EGC

A total of 4,359 genes in GSE55696 and 8,020 genes in GSE130823 with the highest expression variance (top 25%) were selected for subsequent WGCNA. $\beta = 6$ (scale-free $R^2 = 0.87$) and $\beta = 16$ (scale-free $R^2 = 0.85$) were the lowest power fit scale-free indices over 0.85 and determined as the soft-thresholding power parameters to ensure a scale-free network in GSE55696 and GSE130823, respectively (**Supplementary Figure S1**). Eventually, genes with similar expression patterns were grouped into 9 and 7 co-expression modules, respectively, in GSE55696 and GSE130823 (**Figures 1A,B**). The gray module is a set of genes that cannot be clustered to any module. We found that the green module ($R = 0.57$, $p = 3e-07$) in GSE55696 and the

yellow module ($R = 0.68$, $p = 1e-07$) in GSE130823 showed the highest correlation with Stage. Regarding as key gene modules associated with GPL progression to EGC, the green module in GSE55696 possessed 365 genes, while the yellow module in GSE130823 harbored 384 genes. Venn diagrams obtained 219 intersection genes from the above two key modules (**Figure 1C**), and they were mainly enriched in critical immune biological processes, including T-cell activation, regulation of lymphocyte activation, the immune response-regulating cell surface receptor signaling pathway, and so on (**Figure 1D**).

Identification and Validation of Novel Genes Associated With GPL Progression to EGC

Among the 219 intersection genes, 33 genes were differentially expressed between the tumor and adjacent normal tissue, 102 genes harbored predictive ability for OS with $AUC \geq 0.60$, and 4 genes were significantly associated with survival. After an intersection, CXCR4 were found to meet the above three criteria meanwhile (**Figure 2A**). Expression profile data showed that CXCR4 expression was a significant difference among pathological stages (GSE55696: p in the one-way ANOVA test < 0.0001 , **Figure 2B**; GSE130823, p in the one-way ANOVA test = 0.0004, **Figure 2D**) and significantly positively correlated with tumorigenesis (GSE55696: $r = 0.6157$, $p < 0.0001$, **Figure 2C**; GSE130823, $r = 0.5799$, $p = 0.0001$, **Figure 2E**). Survival analysis indicated that the increased

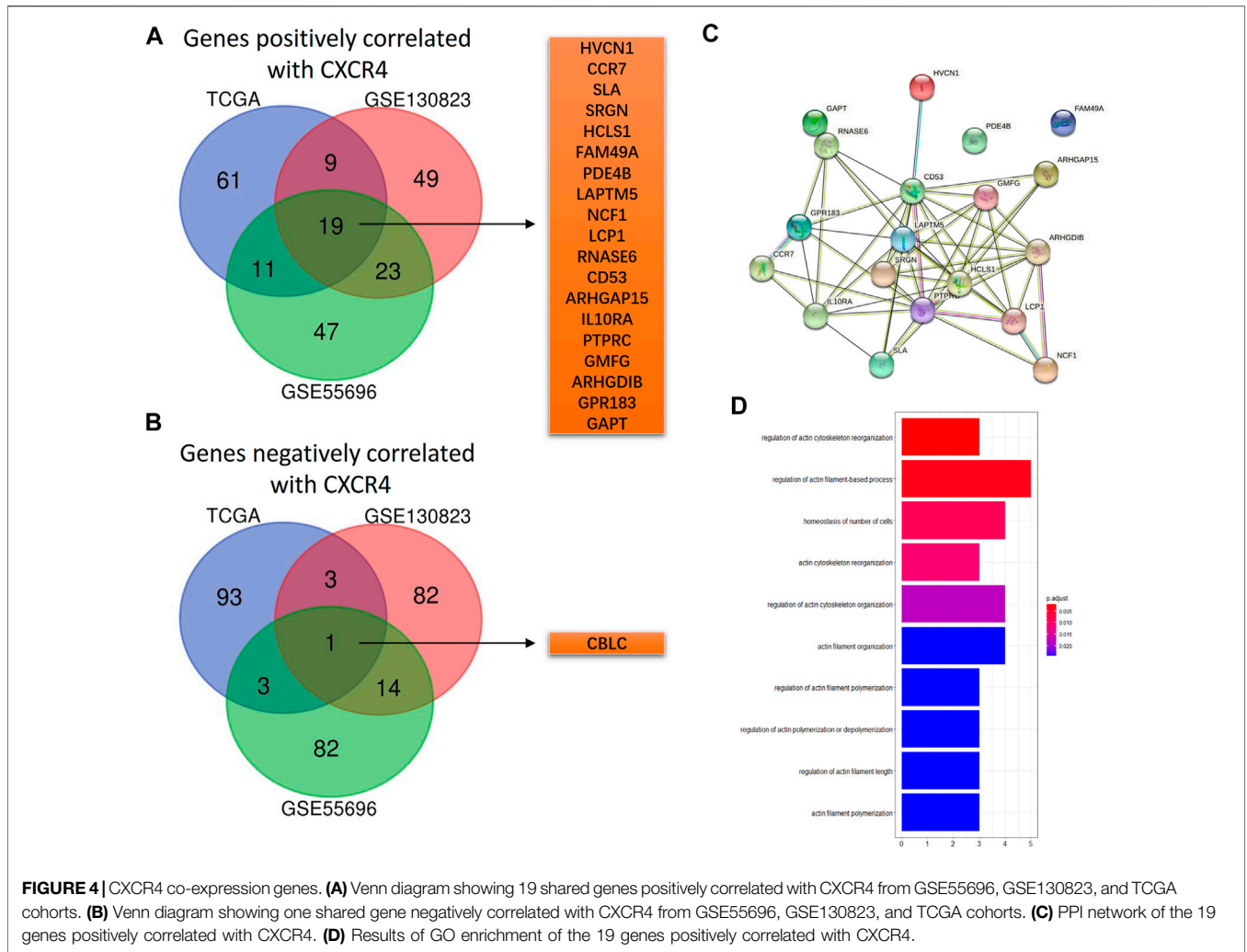


FIGURE 4 | CXCR4 co-expression genes. **(A)** Venn diagram showing 19 shared genes positively correlated with CXCR4 from GSE55696, GSE130823, and TCGA cohorts. **(B)** Venn diagram showing one shared gene negatively correlated with CXCR4 from GSE55696, GSE130823, and TCGA cohorts. **(C)** PPI network of the 19 genes positively correlated with CXCR4. **(D)** Results of GO enrichment of the 19 genes positively correlated with CXCR4.

expression of CXCR4 was significantly associated with poorer prognosis (Figure 2F, $p = 0.0086$). We validated CXCR4 expression based on immunofluorescence in 13 stomach tissue samples including 4 CG, 4 LGIN, 3 HGIN, and 3 GC. The positive rate of CXCR4 was also a significant difference among pathological stages (p in the one-way ANOVA test = 0.0051, Figures 3A,B) and significantly positively correlated with tumorigenesis ($r = 0.7667$, $p = 0.0003$, Figure 3C). We also observed that CXCR4 mainly expressed inside the tumor region but rarely in the peri-tumor (Figure 3A).

CXCR4 Co-expression Genes

According to the Spearman correlation test, there were 19 genes positively correlated with CXCR4 (HVCN1, CCR7, SLA, SRGN, HCLS1, FAM49A, PDE4B, LAPTMS, NCF1, LCP1, RNASE6, CD53, ARHGAP15, IL10RA, PTPRC, GMFG, ARHGDIB, GPR183, and GAOT, Figure 4A), and the interaction network is shown in Figure 4C. However, only one gene was negatively correlated with CXCR4 (Figure 4B). According to functional enrichment, the 19 genes positively correlated with CXCR4 were mainly involved in regulation of actin (Figure 4D).

CXCR4 Co-expression Cytokines

In order to explore the immune implication of CXCR4, CXCR4 co-expression cytokines were identified. The number of cytokines positively correlated with CXCR4 was up to 59, and their interaction was significant (Figures 5A,C). However, there are only seven cytokines negatively correlated with CXCR4, and no significant interaction was observed (Figures 5B,D). Figure 5E shows the predominate biological processes of CXCR4 positively related cytokines such as regulation of chemotaxis, cell chemotaxis, mononuclear cell migration, leukocyte chemotaxis, etc.

CXCR4 Co-expression Immune Cells

In a similar manner, we identified CXCR4 co-expression immune cells. In GPL and GC cohorts, the expression level of CXCR4 was positively correlated with the abundance of macrophages (GSE55696: $r = 0.6$, $p = 6.8e-09$, GSE130823: $r = 0.57$, $p = 3.1e-05$, TCGA: $r = 0.38$, $p = 1.1e-13$, Figure 6A,C) but negatively correlated with that of effector memory T cells (GSE55696: $r = -0.75$, $p = 5.1e-15$, GSE130823: $r = -0.77$, $p = 2.7e-10$, TCGA: $r = -0.45$, $p < 2.2e-16$, Figures 6B,C) and

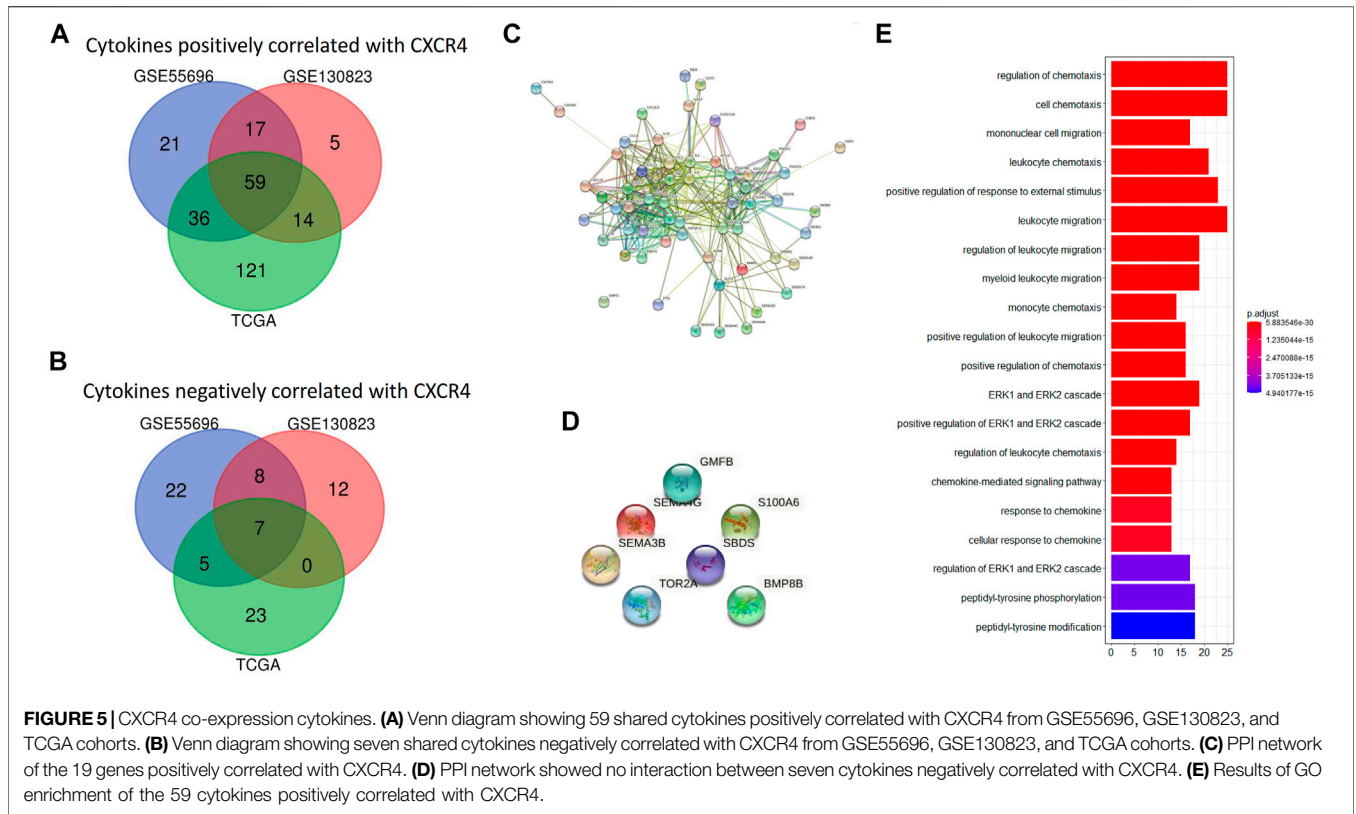


FIGURE 5 | CXCR4 co-expression cytokines. **(A)** Venn diagram showing 59 shared cytokines positively correlated with CXCR4 from GSE55696, GSE130823, and TCGA cohorts. **(B)** Venn diagram showing seven shared cytokines negatively correlated with CXCR4 from GSE55696, GSE130823, and TCGA cohorts. **(C)** PPI network of the 19 genes positively correlated with CXCR4. **(D)** PPI network showed no interaction between seven cytokines negatively correlated with CXCR4. **(E)** Results of GO enrichment of the 59 cytokines positively correlated with CXCR4.

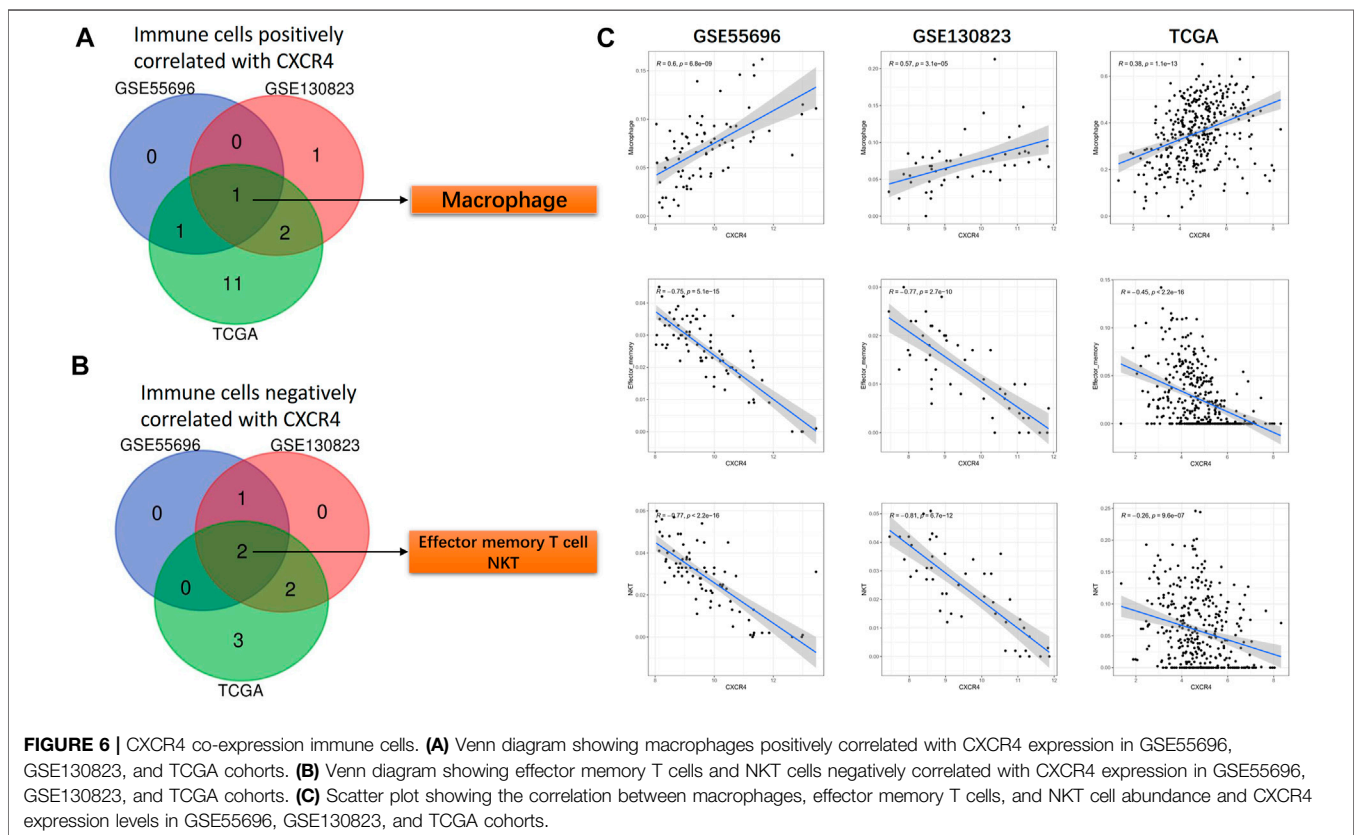
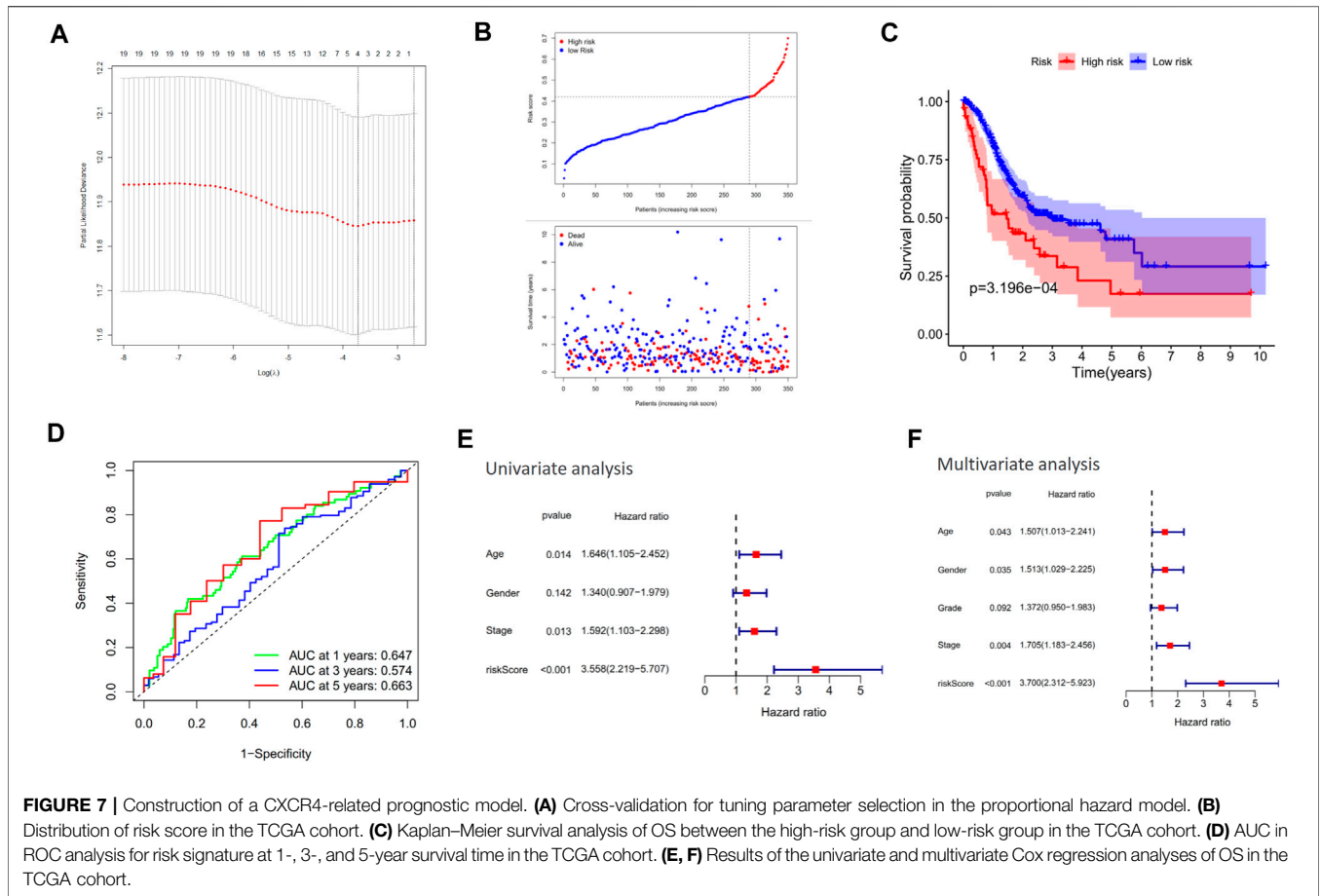


FIGURE 6 | CXCR4 co-expression immune cells. **(A)** Venn diagram showing macrophages positively correlated with CXCR4 expression in GSE55696, GSE130823, and TCGA cohorts. **(B)** Venn diagram showing effector memory T cells and NKT cells negatively correlated with CXCR4 expression in GSE55696, GSE130823, and TCGA cohorts. **(C)** Scatter plot showing the correlation between macrophages, effector memory T cells, and NKT cell abundance and CXCR4 expression levels in GSE55696, GSE130823, and TCGA cohorts.



NKT cells (GSE55696: $r = -0.77$, $p < 2.2e-16$, GSE130823: $r = -0.81$, $p = 6.7e-12$, TCGA: $r = -0.26$, $p = 9.6e-07$, **Figures 6C**).

Construction of a CXCR4-Related Prognostic Model

Among the 20 co-expression genes of CXCR4, the LASSO Cox regression analysis identified 4 predictors (i.e., FAM49A, GPR183, CCR7, and GAPT) with the greatest impact on the OS of GC patients (**Figure 7A**). Based on the optimal cutoff of risk score, the patients were divided into high-risk ($n = 60$) and low-risk ($n = 290$) groups and the patients with a higher risk score tended to possess a poorer survival (**Figure 7B**). The K-M curve revealed that patients in the high-risk group had worse survival outcomes compared to those in the low-risk group (**Figure 7C**, $p = 3.196e-04$) and the AUC of ROC for 1-, 3-, and 5-year OS reached 0.647, 0.574, and 0.663, respectively (**Figure 7D**). Univariate and multivariate analyses also confirmed the independent prognostic value of this model (**Figures 7E,F**).

Validation of Novel Gene Expression Levels and Correlationship in Cell Lines

Among the 19 genes positively correlated with CXCR4, 4 genes (FAM49A, GPR183, CCR7, and GAPT) were identified with

the greatest impact on the OS of GC patients according to LASSO Cox regression analysis. FAM49A and GPR183 were selected for further validation. CBLC, the only one gene negatively correlated with CXCR4 in this research, was also sent to validate. The expression levels of CXCR4, FAM49A, GPR183, and CBLC were evaluated in GES-1 and AGS by qRT-PCR. Compared with GES-1, CXCR4 was significantly upregulated in AGS (**Figure 8A₂**, $p < 0.01$). For CXCR4 positively correlated genes, FAM49A was significantly upregulated (**Figure 8B₂**, $p < 0.01$), while GPR183 was downregulated in AGS (**Figure 8C₂**, $p < 0.01$). CBLC, the only one gene negatively correlated with CXCR4, was remarkably downregulated in AGS (**Figure 8D₂**, $p < 0.01$). Consistently, no matter in GES-1 or AGS, FAM49A was positively correlated with CXCR4 (GES-1: $r = -0.8545$, $p = 0.0302$, **Figure 8E**; AGS: $r = 0.8247$, $p = 0.0434$, **Figure 8F**) and CBLC was remarkably negatively correlated (GES-1: $r = -0.8543$, $p = 0.0303$, **Figure 8G**; AGS: $r = -0.8545$, $p = 0.0302$, **Figure 8H**).

DISCUSSION

Over the past few decades, although the research on gastric tumors itself has received much attention, understanding the

Validation of novel genes expression levels and relationship in cell lines.

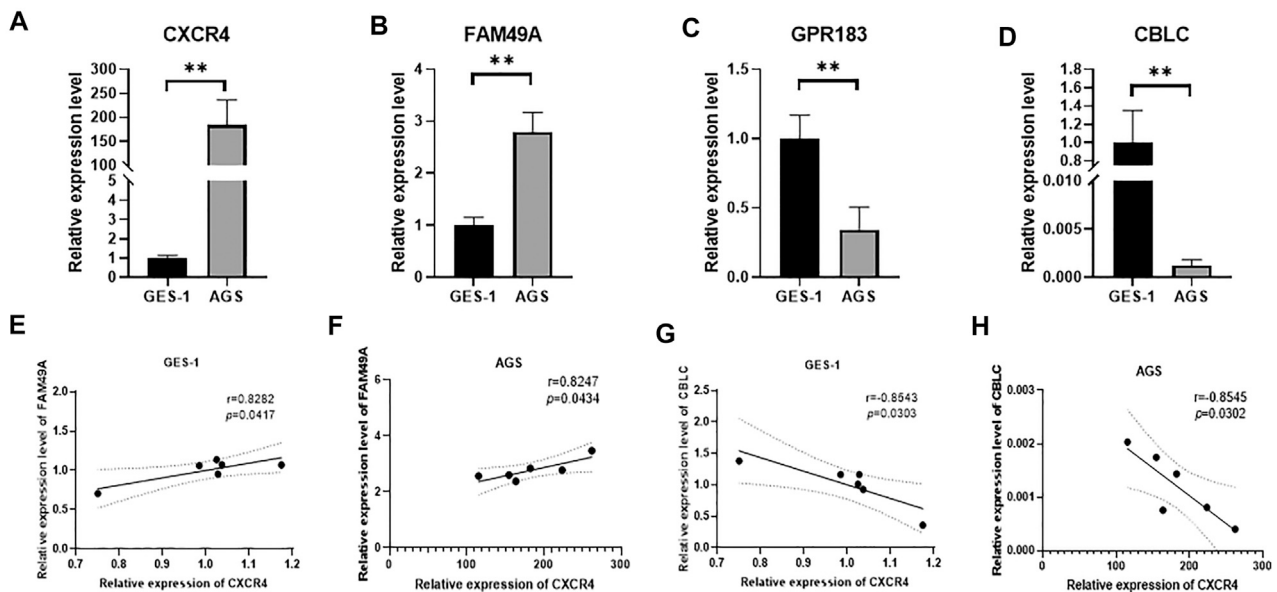


FIGURE 8 | Relative expression levels of (A) CXCR4, (B) FAM49A, (C) GPR183, and (D) CBLC between AGS and GES-1. ns: not significant; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$. Correlation between CXCR4 and FAM49A in GES-1 (E) and AGS (F). Correlation between CXCR4 and CBLC in GES-1 (G) and AGS (H).

mechanism of the process from GPL becoming cancer has made little advances. Besides, biomarkers that can monitor the progression of GPL-to-GC are still much insufficient.

In this study, the expression level and the immunofluorescence positive rate of CXCR4 were observed increasing with GPL progression. Besides, they were significantly different among pathological stages and remarkably positively correlated with tumorigenesis. Therefore, we identified CXCR4 as a novel biomarker correlated with the malignant transformation of GPL. CXCR4 is a seven-span transmembrane G-protein coupled receptor that is the primary receptor for CXCL12 (Loetscher et al., 1994). CXCR4 plays an important role in tumor biological behaviors, such as growth, metastasis, angiogenesis, and cancer cell-microenvironment interactions (Sleightholm et al., 2017). It is reported that CXCR4 is upregulated in GC and associated with poor prognosis (He et al., 2013; Xu et al., 2020).

We also explored the implication of CXCR4 in GPL malignant transformation. CXCR4 co-expression genes were identified, and they were found to be mainly involved in regulation of actin. The malignant transformation of GPL is accompanied by the enhancement of invasion and metastasis capabilities (Correa and Piazuolo, 2015), which is significantly associated with extracellular matrix (ECM) remodeling (Uloza et al., 2015). The reconstructed ECM forms a loose microenvironment for cancer cells, leading to high proliferation as well as invasion and metastasis of tumor

cells (Schedin and Keely, 2011). In ECM remodeling, it is the regulation of actin that serves as the key modulator (Sahai, 2005). Accordingly, from immunofluorescence results, CXCR4 mainly expressed inside the tumor region, which further indicated a crucial role of CXCR4 in the regulation of actin and ECM remodeling in the tumor microenvironment (TME).

CXCR4 co-expression cytokines and immune cells were also identified. CXCR4 co-expression cytokines were mainly enriched in regulation of chemotaxis, cell chemotaxis, mononuclear cell migration, leukocyte chemotaxis, etc. As for co-expression immune cells, we found that the expression level of CXCR4 was positively correlated with the abundance of macrophages but negatively correlated with that of effector memory T cells and NKT cells during GPL malignant transformation. In our research, CXCR4 was observed to be mainly expressed inside the tumor region but rarely in the peri-tumor. It is reported that the SDF-1/CXCR4 axis could attract CXCR4-expressing tumor-associated macrophage (TAM) and myeloid-derived suppressor cell (MDSC) accumulation in TME and suppresses T-cell immune responses (Obermajer et al., 2011; Vatner et al., 2014; Séhédic et al., 2017; Jiang et al., 2019). Therefore, the accumulation of CXCR4-expressing TAMs and MDSCs may be the reason why CXCR4 was mainly observed in tumor areas. Besides, the enhancement of protumor immunity and the impairment of antitumor immunity may result in the malignant transformation of GPL. Last, we constructed a

CXCR4-related prognostic model. Our results demonstrated that the risk model was able to predict the prognosis of GC and serve as an independent predictor for OS, which further illustrated that CXCR4 played a vital role in malignant transformation.

Some limitations should be addressed in this study. The mechanisms underpinning CXCR4-mediated malignant transformation of GPL and alteration of tumor immunity should be explored in the future. Second, our model was constructed and validated based on retrospective data. Prospective clinical validation is needed henceforth.

In conclusion, CXCR4 was a novel biomarker correlated with malignant transformation of GPL and played a vital role in the control of tumor immunity. Besides, the prognostic signatures derived from CXCR4-related genes could serve as an independent predictor for OS in GC. CXCR4 is possible to serve as a therapeutic target for malignant transformation of GPL.

DATA AVAILABILITY STATEMENT

Publicly available data sets were analyzed in this study. These data can be found here: TCGA and the GEO database.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the First Affiliated Hospital of Guangzhou University of Chinese Medicine. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

KZ, FL, SX, and PLi designed the study. XJ, KJ, and JZ carried out the analysis and wrote the manuscript. YW and YL performed the experiment. YY, YH, and KN participated in coordination of the study and interpretation of the results. ZY, ZZ, and JP participated in manuscript writing. PLiu collected data. All authors read and approved the final manuscript.

REFERENCES

- Aran, D., Hu, Z., and Butte, A. J. (2017). xCell: Digitally Portraying the Tissue Cellular Heterogeneity Landscape. *Genome Biol.* 18 (1), 220. doi:10.1186/s13059-017-1349-1
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2013). NCBI GEO: Archive for Functional Genomics Data Sets—Update. *Nucleic Acids Res.* 39, 1005–1010. doi:10.1093/nar/gks1193
- Becht, E., Giraldo, N. A., Lacroix, L., and Buttard, B. (2016). Estimating the Population Abundance of Tissue-Infiltrating Immune and Stromal Cell Populations Using Gene Expression. *Genome Biol.* 17 (1), 218. doi:10.1186/s13059-016-1070-5

FUNDING

This study was supported by the following funds: the First Affiliated Hospital of Guangzhou University of Chinese Medicine “Innovative Strong Hospital” clinical research project (Grant No. No.2019IIT19); the Natural Science Foundation of Guangdong Province (Grant No. 2019A1515011145); the National Natural Science Foundation of China (Grant No. 81973819); the Youth program of the National Natural Science Foundation of China (Grant No. 81904139); the Guangdong Medical Science and Technology Research Fund (Grant No. A2020186); the Innovation Development Project of the First Affiliated Hospital, Guangzhou University of Chinese Medicine (Grant No. 2019QN01); the Traditional Chinese Medicine Bureau of Guangdong Province (Grant No. 20211013); the Medical Scientific Research Foundation of Guangdong Province (Grant No. A2021373); the Major Subject Research Projects of Guangzhou University of Traditional Chinese Medicine (Grant No. A1-2606-19-110-007); FL, Guangdong famous traditional Chinese medicine inheritance studio (Grant No. Guangdong TCM office (2020) no. 1); the Guangdong Medical Science and Technology Research Fund (Grant No. B2021089); and the Innovation and Strengthening project of the First Affiliated Hospital of Guangzhou University of traditional Chinese Medicine (Grant No. 211010010101). Guangzhou Science and Technology Program Basic Research Program - City School (Hospital) Jointly Funded Project - City School (Hospital) Jointly Funded (Dengfeng Hospital) Basic Research Project (Grant No.2021438).

ACKNOWLEDGMENTS

We thank Lingnan Medical Research Center for experiment support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.697993/full#supplementary-material>

- Ben-Baruch, A. (2006). The Multifaceted Roles of Chemokines in Malignancy. *Cancer Metastasis Rev.* 25 (3), 357–371. doi:10.1007/s10555-006-9003-5
- Bockerstett, K. A., and Dipaolo, R. J. (2017). Regulation of Gastric Carcinogenesis by Inflammatory Cytokines. *Cell Mol. Gastroenterol. Hepatol.* 4 (1), 47–53. doi:10.1016/j.jcmgh.2017.03.005
- Correa, P. (1988). A Human Model of Gastric Carcinogenesis. *Cancer Res.* 48 (13), 3554–3560.
- Correa, P., and Piazuelo, M. B. (2015). The Gastric Precancerous cascade. *J. Dig. Dis.* 13, 2. doi:10.1111/j.1751-2980.2011.00550.x
- Freddie, B., Jacques, F., Isabelle, S., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer J. Clin.* 68, 394. doi:10.3322/caac.21492

- Goldman, M., Craft, B., Swatoski, T., Cline, M., Morozova, O., Diekhans, M., et al. (2015). The UCSC Cancer Genomics Browser: Update 2015. *Nucleic Acids Res.* 43, D812–D817. doi:10.1093/nar/gku1073
- Gotoda, T., Yanagisawa, A., Sasako, M., Ono, H., Nakanishi, Y., Shimoda, T., et al. (2000). Incidence of Lymph Node Metastasis from Early Gastric Cancer: Estimation with a Large Number of Cases at Two Large Centers. *Gastric Cancer* 3 (4), 219–225. doi:10.1007/pl00011720
- He, H., Wang, C., Shen, Z., Fang, Y., Wang, X., Chen, W., et al. (2013). Upregulated Expression of C-X-C Chemokine Receptor 4 Is an Independent Prognostic Predictor for Patients with Gastric Cancer. *PLoS one* 8 (8), e71864. doi:10.1371/journal.pone.0071864
- Jiang, K., Li, J., Zhang, J., Wang, L., Zhang, Q., Ge, J., et al. (2019). SDF-1/CXCR4 axis Facilitates Myeloid-Derived Suppressor Cells Accumulation in Osteosarcoma Microenvironment and Blunts the Response to Anti-PD-1 Therapy. *Int. Immunopharmacol.* 75, 105818. doi:10.1016/j.intimp.2019.105818
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R Package for Weighted Correlation Network Analysis. *BMC Bioinformatics* 9, 559. doi:10.1186/1471-2105-9-559
- Laurén, P. (1965). The Two Histological Main Types of Gastric Carcinoma: Diffuse and So-Called Intestinal-Type Carcinoma. An Attempt at A Histo-Clinical Classification. *Acta Pathologica Microbiologica Scand.* 64 (1), 31. doi:10.1111/apm.1965.64.1.31
- Lee, H. J., Song, I. C., Yun, H. J., Jo, D. Y., and Kim, S. (2014). CXC Chemokines and Chemokine Receptors in Gastric Cancer: From Basic Findings towards Therapeutic Targeting. *World J. Gastroenterol.* 20 (7), 1681–1693. doi:10.3748/wjg.v20.i7.1681
- Li, T., Fan, J., Wang, B., Traugh, N., Chen, Q., Liu, J. S., et al. (2017). TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res.* 77 (21), e108. doi:10.1158/0008-5472.can-17-0307
- Lin, E., and Lane, H. Y. (2017). Machine Learning and Systems Genomics Approaches for Multi-Omics Data. *Biomarker Res.* 5 (1), 2. doi:10.1186/s40364-017-0082-y
- Loetscher, M., Geiser, T., O'Reilly, T., Zwahlen, R., Baggiolini, M., and Moser, B. (1994). Cloning of a Human Seven-Transmembrane Domain Receptor, LESTR, that Is Highly Expressed in Leukocytes. *J. Biol. Chem.* 269 (1), 232–237. doi:10.1016/s0021-9258(17)42339-8
- Miao, Y. u., Zhang, Q., Lei, Q., Luo, M., and Guo, A. u. (2020). ImmuCellAI: A Unique Method for Comprehensive T X ell Subsets Abundance Prediction and its Application in Cancer Immunotherapy. *Adv. Sci.* 7 (7), 1902880. doi:10.1002/advs.201902880
- Nardone, G., Rocco, A., and Malfertheiner, P. (2004). *Helicobacter pylori* and Molecular Events in Precancerous Gastric Lesions. *Aliment. Pharmacol. Ther.* 20, 261. doi:10.1111/j.1365-2036.2004.02075.x
- Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust Enumeration of Cell Subsets from Tissue Expression Profiles. *Nat. Methods* 12 (5), 453–457. doi:10.1038/nmeth.3337
- Obermajer, N., Muthuswamy, R., Odunsi, K., Edwards, R., and Kalinski, P. (2011). PGE(2)-induced CXCL12 Production and CXCR4 Expression Controls the Accumulation of Human MDSCs in Ovarian Cancer Environment. *Cancer Res.* 71 (24), 7463–7470. doi:10.1158/0008-5472.can-11-2449
- Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D. E., and Gfeller, D. (2017). Simultaneous Enumeration of Cancer and Immune Cell Types from Bulk Tumor Gene Expression Data. *Elife* 6, e26476. doi:10.7554/eLife.26476
- Sahai, E. (2005). Mechanisms of Cancer Cell Invasion. *Curr. Opin. Genet. Dev.* 15 (1), 87–96. doi:10.1016/j.gde.2004.12.002
- Schedin, P., and Keely, P. (2011). Mammary Gland ECM Remodeling, Stiffness, and Mechanosignaling in normal Development and Tumor Progression. *Cold Spring Harbor Perspect. Biol.* 3 (1), a003228. doi:10.1101/cshperspect.a003228
- Séhédic, D., Chourpa, I., Tétaud, C., Griveau, A., Loussouarn, C., Avril, S., et al. (2017). Locoregional Confinement and Major Clinical Benefit of (188)Re-Loaded CXCR4-Targeted Nanocarriers in an Orthotopic Human to Mouse Model of Glioblastoma. *Theranostics* 7 (18), 4517–4536. doi:10.7150/thno.19403
- Sleightholm, R., Neilsen, B., Li, J., Steele, M., Singh, R., Hollingsworth, M., et al. (2017). Emerging Roles of the CXCL12/CXCR4 axis in Pancreatic Cancer Progression and Therapy. *Pharmacol. Ther.* 179, 158–170. doi:10.1016/j.pharmthera.2017.05.012
- Smyth, G. K., Ritchie, M., Thorne, N., Wettenhall, J., and Shi, W. (2010). “Limma: Linear Models for Microarray Data,” in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Melbourne, VIC, Australia: Walter and Eliza Hall Institute of Medical Research, 397. doi:10.1007/0-387-29362-0_23
- Uloza, V., Liutkevicius, V., Pangonyte, D., and Lesauskaite, V. (2015). Characteristics of Expression of Matrix Metalloproteinases (MMP-2 and MMP-9) in Glottic Squamous Cell Carcinoma and Benign Vocal Fold Lesions. *Clin. Exp. Otorhinolaryngol.* 8 (1), 57–64. doi:10.3342/ceo.2015.8.1.57
- Vatner, R. E., Cooper, B. T., Vanpouille-Box, C., Demaria, S., and Formenti, S. C. (2014). Combinations of Immunotherapy and Radiation in Cancer Therapy. *Front. Oncol.* 4, 325. doi:10.3389/fonc.2014.00325
- Watanabe, M., Kato, J., Inoue, I., Yoshimura, N., Yoshida, T., Mukoubayashi, C., et al. (2012). Development of Gastric Cancer in Nonatrophic Stomach with Highly Active Inflammation Identified by Serum Levels of Pepsinogen and *Helicobacter pylori* Antibody Together with Endoscopic Rugal Hyperplastic Gastritis. *Int. J. Cancer J. Int. Du Cancer* 131 (11), 2632–2642. doi:10.1002/ijc.27514
- Xu, G., Lu, K., Shen, M., Zhang, Q., Pan, W., and Tang, Z. (2020). Correlation between Chemokine CXCL-12 and Its Receptor CXCR4 Expression Is Associated with Clinical Prognosis of Gastric Cancer. *Clin. Lab.* 66 (4). doi:10.7754/Clin.Lab.2019.190217
- Xu, X., Feng, L., Liu, Y., Zhou, W. X., Ma, Y. C., Fei, G. J., et al. (2014). Differential Gene Expression Profiling of Gastric Intraepithelial Neoplasia and Early-Stage Adenocarcinoma. *World J. Gastroenterol.* 20 (47), 17883–17893. doi:10.3748/wjg.v20.i47.17883
- Zhang, B., and Horvath, S. (2005). A General Framework for Weighted Gene Co-expression Network Analysis. *Stat. Appl. Genet. Mol. Biol.* 4, 17. doi:10.2202/1544-6115.1128
- Zhang, Y., Wu, X., Zhang, C., Wang, J., Fei, G., Di, X., et al. (2020). Dissecting Expression Profiles of Gastric Precancerous Lesions and Early Gastric Cancer to Explore Crucial Molecules in Intestinal-type Gastric Cancer Tumorigenesis. *J. Pathol.* 251(2), 135–146. doi:10.1002/path.5434
- Zhu, Q., Han, X., Peng, J., Qin, H., and Wang, Y. (2012). The Role of CXC Chemokines and Their Receptors in the Progression and Treatment of Tumors. *J. Mol. Histology* 43 (6), 699–713. doi:10.1007/s10735-012-9435-x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article or claim that may be made by its manufacturer is not guaranteed or endorsed by the publisher.

Copyright © 2021 Jiang, Zheng, Liu, Jiang, Wen, Yan, Liu, Zhong, Huang, Yao, Nie, Zheng, Pan, Liu, Zhuang, Liu, Xu and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.