Proceedings of Anticancer Research, 2025, Volume 9, Issue 5
http://ojs.bbwpublisher.com/index.php/PAR
ISSN Online: 2208-3553
ISSN Print: 2208-3545

Exploring Gastric Cancer-Related Genes and Clinical Significance Analysis Based on Bioinformatics

Liansi Ye[1], Chuanxin Zou[2]*

[1]Department of Gastroenterology, Jingzhou Hospital Affiliated to Yangtze University, Jingzhou, China
[2]Department of Gastroenterology, Jingzhou Hospital Affiliated to Yangtze University, Jingzhou, China

**Author to whom correspondence should be addressed.*

Abstract: *Objective*: Employing bioinformatics methodologies to identify core genes intricately associated with the pathogenesis and progression of gastric cancer, and to evaluate their clinical significance. *Method*: Gene expression datasets GSE19826 and GSE13911 were acquired from the Gene Expression Omnibus (GEO). Differential gene expression analysis was conducted using GEO2R. Common differentially expressed genes (DEGs) were discerned via Venn diagram analysis on a bioinformatics platform. Functional enrichment analyses, including Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG), were performed on these overlapping DEGs. A protein-protein interaction (PPI) network was constructed with the STRING database, and central hub genes were identified using Cytoscape software. The expression profiles, prognostic value, and immune infiltration correlations of these key genes were further examined utilizing the GEPIA, Kaplan-Meier plotter, Human Protein Atlas (HPA), and TIMER databases. *Results*: Analysis revealed 120 commonly differentially expressed genes. These genes were significantly enriched in biological pathways concerning muscle cell cytoskeleton regulation, nutrient absorption, and extracellular matrix receptor interactions. PPI network analysis highlighted 10 core genes, including COL1A1, COL1A2, BGN, THBS2, COL5A2, and TIMP1. These genes exhibited marked upregulation in GC tissues. Statistical evaluation confirmed a significant link between their elevated expression and unfavorable patient outcomes ($P < 0.01$). Furthermore, immune infiltration assessment indicated a positive correlation between the expression of these genes and macrophage infiltration within the tumor microenvironment, implying their involvement in modulating the immune response in GC, which could affect tumor behavior and clinical progression. *Conclusion*: The six genes identified may function as diagnostic biomarkers and represent promising therapeutic targets for gastric cancer.

Keywords: Gastric cancer; Biomarker; Bioinformatics; Differentially expressed genes; Immune microenvironment

Online publication: October 13, 2025

1. Introduction

Gastric cancer (GC) is a prevalent malignancy within the digestive tract, ranking among the most frequent cancers in China and representing the third primary cause of cancer-associated mortality globally [1]. Thanks to better understanding of the causes of GC and advancements in early screening and clinical diagnosis and treatment techniques, the overall incidence and mortality of GC have decreased significantly. However, in 2020, there were more than one million new cases and over 768,000 deaths worldwide. These figures are expected to increase by 2040, with 1.77 million new cases and 1.27 million deaths projected worldwide [2]. Probably because of the occult nature of early-stage gastric cancer, most patients have metastases at the time of initial diagnosis, reaching an advanced stage of diagnosis. This has a significant impact on treatment and prognosis. Therefore, adopting effective detection methods to improve early diagnosis, optimize treatment, reduce recurrence, and improve prognosis is the main challenge and prospect of GC patient management. With the rapid development of research on GC metabolic biomarkers, new GC biomarkers have been discovered, such as expression levels of various proteins and genes in GC samples, creating new opportunities for the diagnosis and monitoring of GC patients [3]. These findings may provide valuable targets for the early diagnosis and individualized treatment of GC. Therefore, the discovery of efficient tumor markers is of great significance for the diagnosis and treatment of GC and for improving patient survival rates. In this study, we screened and downloaded GC-related datasets from the GEO database, used bioinformatics methods to mine and analyze DEGs in GC tissues and adjacent normal tissues, selected the relevant key genes, analyzed their expression and survival curves, and the key was to identify possible GC diagnostic markers and potential therapeutic targets.

1.1 Data acquisition

The NCBI GEO database (https://www.ncbi.nlm.nih.gov/geo) was accessed to obtain the gene expression datasets GSE19826, comprising 12 samples (tumor/adjacent normal), and GSE13911, which includes 38 tumor and 31 normal samples.

1.2. Differential expression analysis

DEGs between GC tissues and matched normal tissues were identified using GEO2R. Thresholds were set at $|\log FC| > 1$ and an adjusted P-value < 0.05. Common DEGs across datasets were visualized and selected through Venn diagram analysis on a bioinformatics platform.

1.3. Functional enrichment analysis

GO/KEGG analysis of overlapping DEGs was performed using the microbioinformatics platform ($P < 0.05$), with GO focusing on biological functions and KEGG focusing on signaling pathways.

1.4. PPI network construction

Construct the PPI network in the STRING database (with a confidence level of 0.4) and screen the Top 10 hub genes using the degree plugin of Cytoscape 3.9.0.

1.5. Validation of key genes

mRNA expression differences were analyzed by GEPIA, survival associations were evaluated by Kaplan-Meier plotter, protein expression was verified by HPA, and immune infiltration was analyzed by TIMER.

2. Results

2.1. Screening of differentially expressed genes

A total of 120 overlapping DEGs were consistently identified across the two analyzed datasets (**Figure 1** and **Figure 2**).

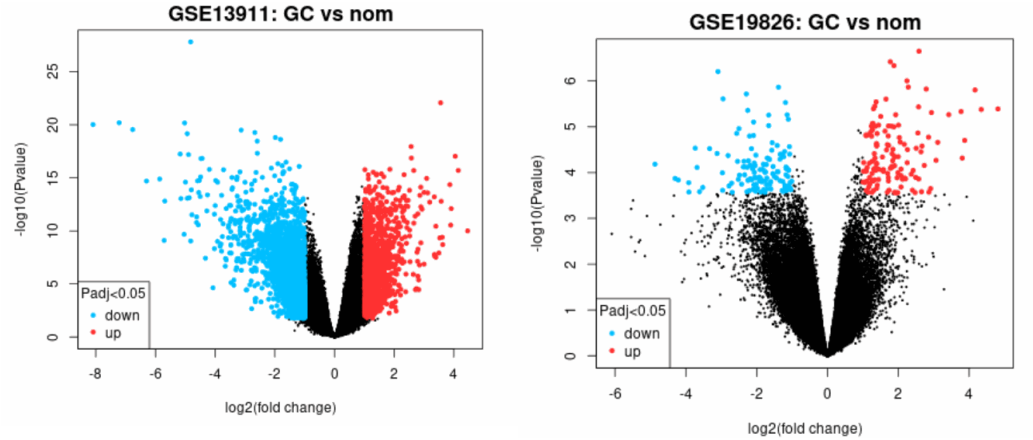

Figure 1. Volcano maps of the GSE19826 and GSE13911 datasets.

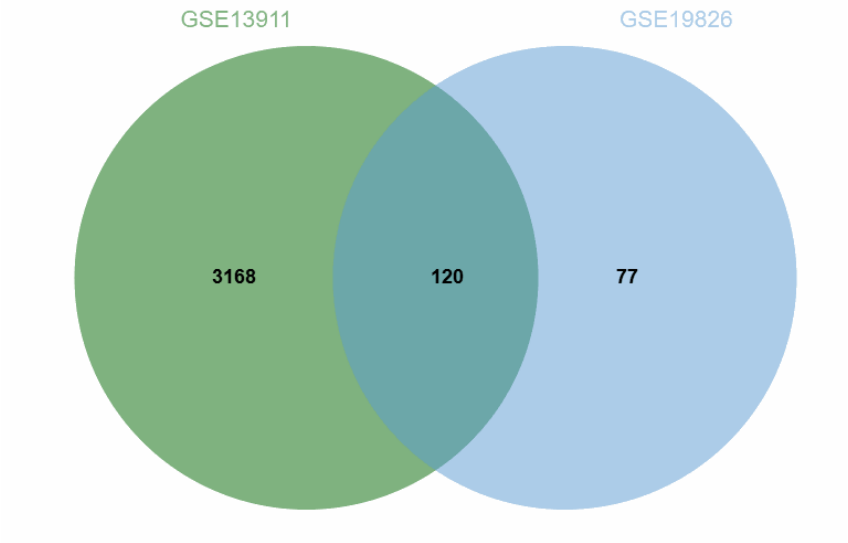

Figure 2. Venn diagrams of the GSE19826 and GSE13911 datasets DEGs.

2.2. Functional enrichment analysis of DEGs

DEGs are primarily involved in extracellular matrix formation (BP), collagen trimer formation (CC), and glycosaminoglycan binding (MF). The KEGG pathway is enriched in myocytoskeletal regulation and extracellular matrix receptor interaction (**Figure 3**).

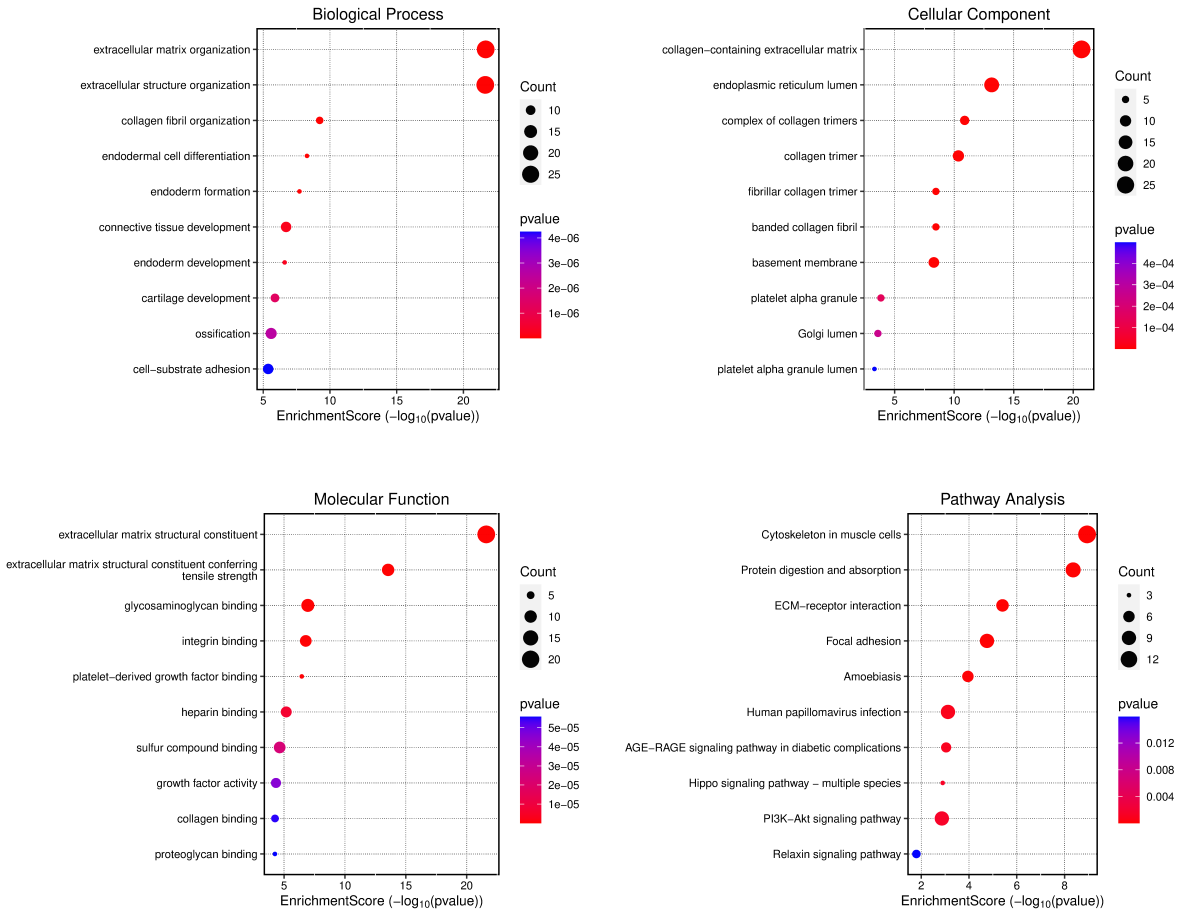

Figure 3. Enrichment analysis of overlapping differentially expressed genes based on GO and KEGG.

2.3. PPI network analysis

10 hub genes, such as COL3A1 and FN1 were screened out (Figure 4).

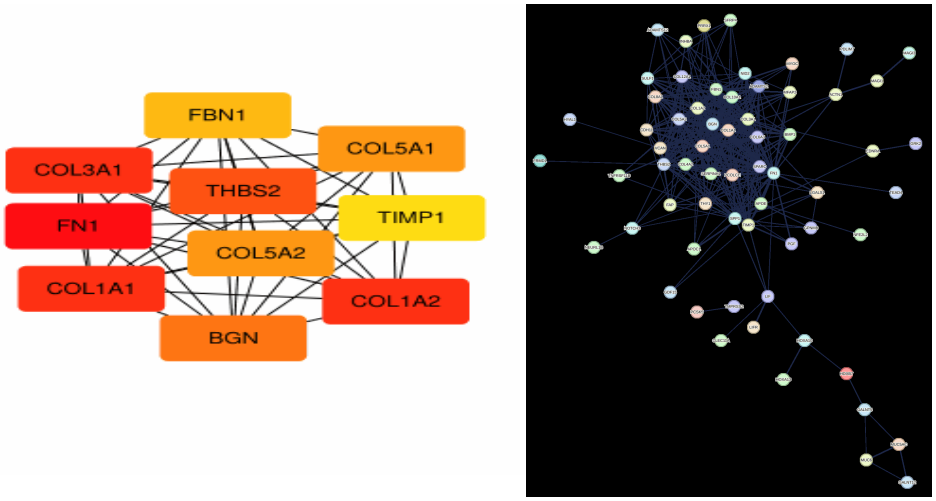

Figure 4. PPI network map of overlapping DEGs and key genes (A.PPI network B. Key genes - The redder the color, the more generation connection points).

2.4. Expression and prognostic validation

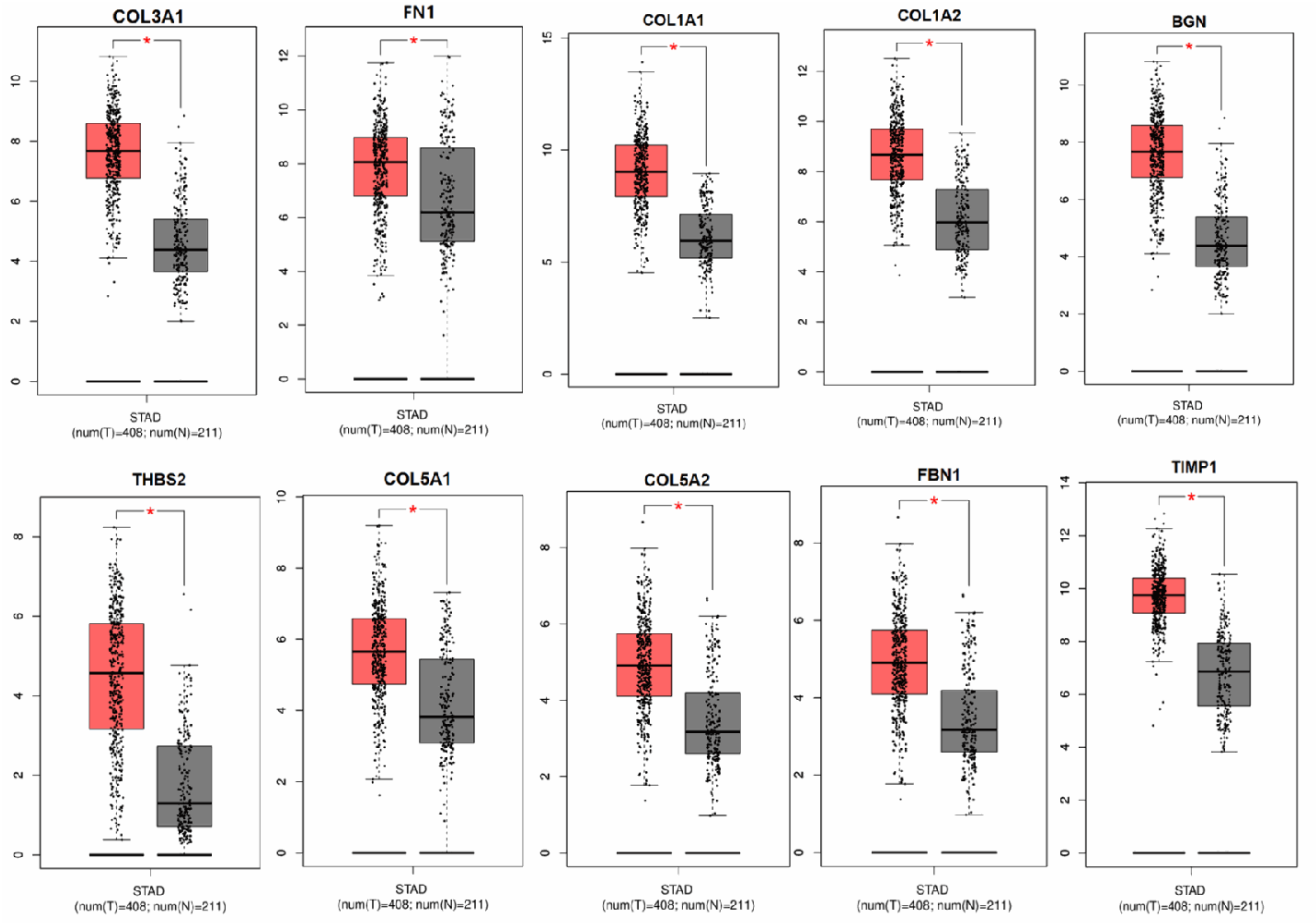

Figure 5. High expression of key genes in gastric cancer tissues (red - gastric cancer tissues, gray - normal gastric tissues).

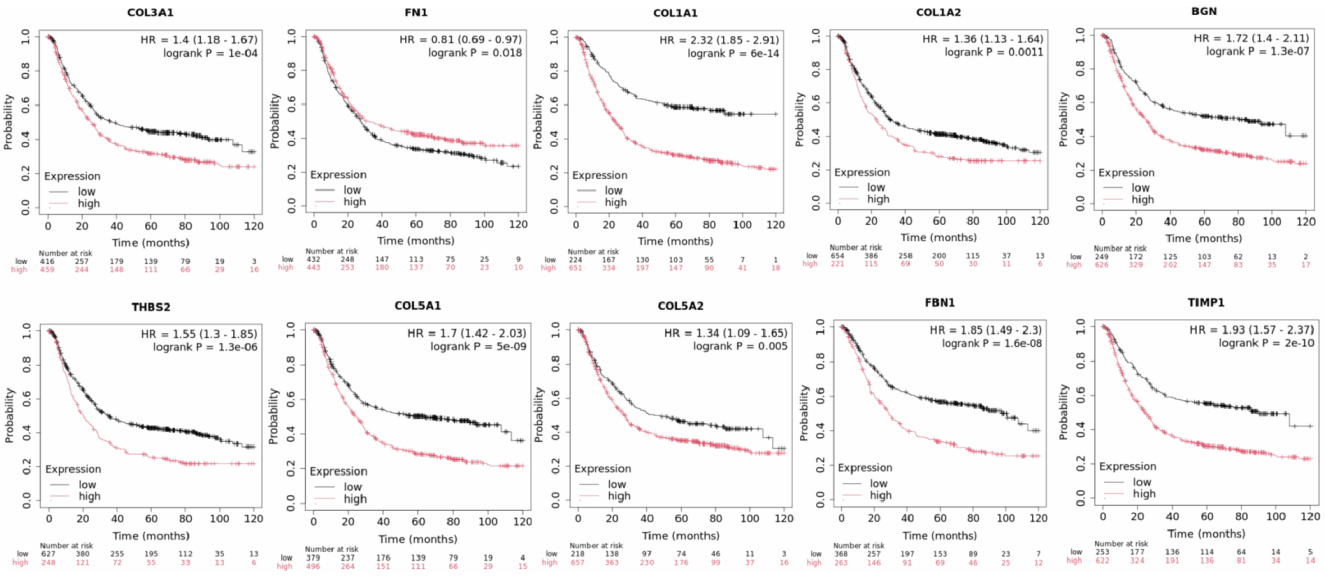

Figure 6. Kaplan-Meier survival curves of key genes (red - high expression, black - low expression).

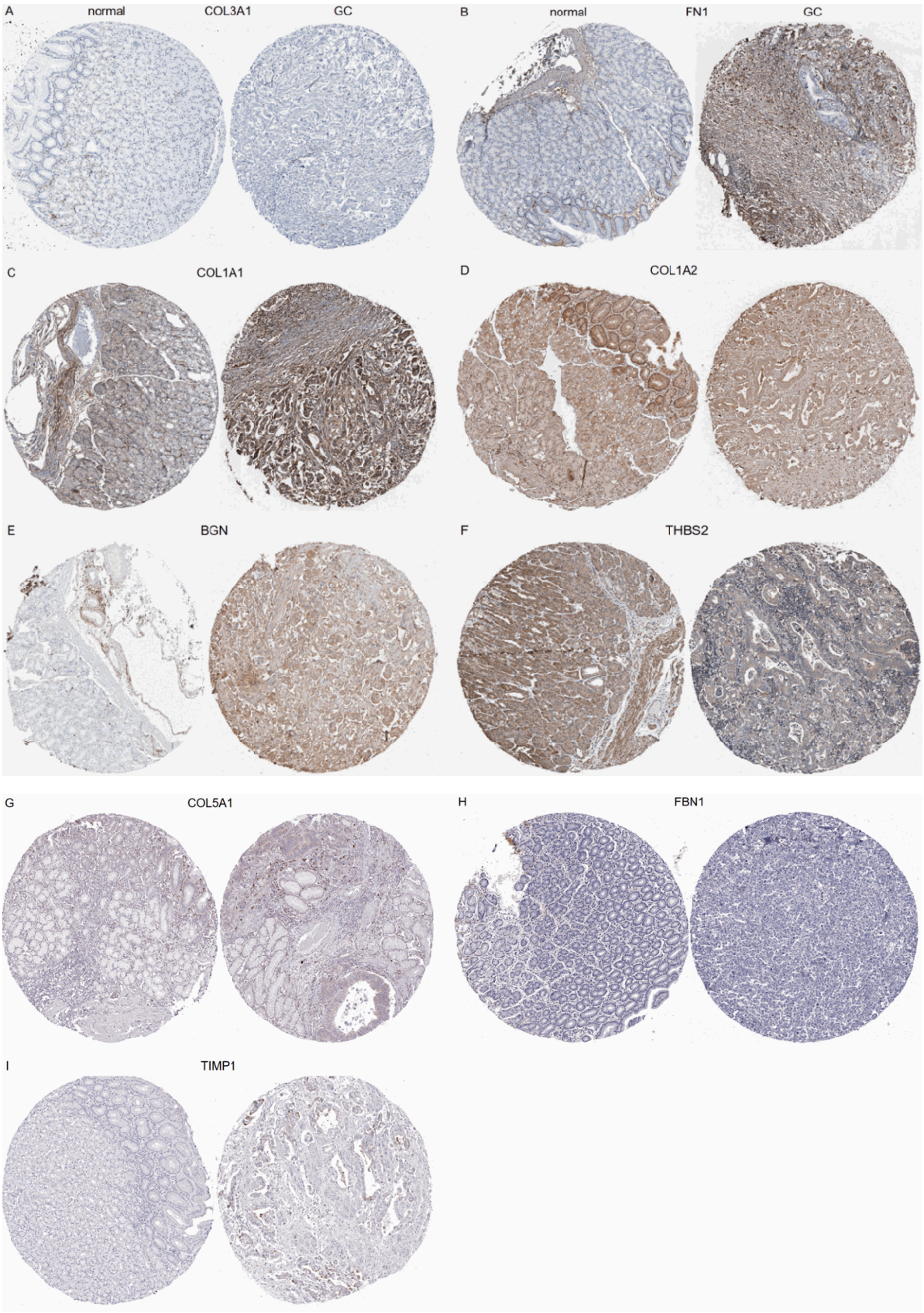

Figure 7. Immunohistochemical images of key genes in gastric cancer tissues and normal gastric tissues in the HPA database; (A-I) represent COL3A1, FN1, COL1A1, COL1A2, BGN, THBS2, COL5A1, FBN1, TIMP1, respectively.

GEPIA showed significantly high expression of 10 genes in gastric cancer tissues ($P < 0.01$, **Figure 5**). Survival analysis indicated that high expression of nine genes, including COL1A1, was associated with poor prognosis (**Figure 6**). HPA confirmed that six genes, including COL1A1, were specifically highly expressed at the protein level (**Figure 7**).

2.5. Immune infiltration analysis

TIMER showed that the key genes were significantly positively correlated with macrophage infiltration ($r > 0.3$, $P < 0.05$) and with infiltration of CD4+T cells, etc. (**Figure 8**)

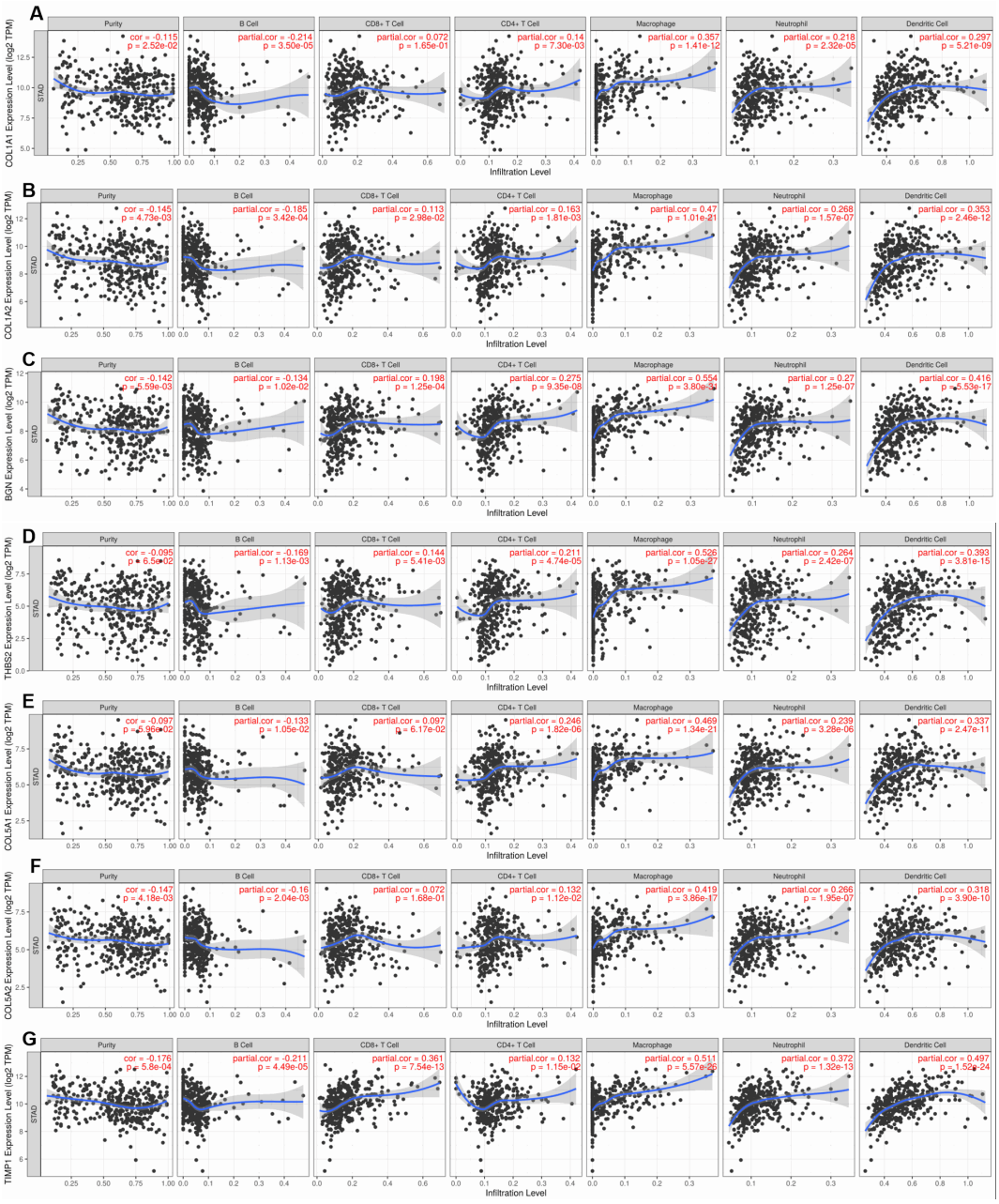

Figure 8. Correlation between the expressions of (A-G) COL1A1, COL1A2, BGN, THBS2, COL5A1, COL5A2, and TIMP1 in gastric cancer tissues and immune infiltrating cell populations (B cells, CD8+T cells, CD4+T cells, macrophages, neutrophils, dendritic cells). $P < 0.05$ was statistically significant.

3. Discussion

GC remains a leading cause of cancer mortality worldwide, posing a substantial public health challenge. The discovery of reliable biomarkers is thus critical for enhancing early detection rates and improving survival outcomes [3]. In this study, two microarray datasets (GSE19826 and GSE13911) from GEO were analyzed bioinformatically, yielding 120 common DEGs. Functional enrichment analysis indicated these genes were predominantly associated with extracellular matrix (ECM) organization and glycosaminoglycan binding. In healthy tissue, the ECM provides structural support and facilitates cellular communication, function, and morphology. Within tumors, the ECM additionally contributes to the tumor microenvironment (TME), fostering tumor progression, chemoresistance, and metastasis [4]. Glycosaminoglycans are common structural and functional components in the extracellular matrix, which can alter the physical properties and are actively involved in TME dynamics, regulating cancer cell proliferation, angiogenesis, invasion, and metastasis through interactions with growth factors and signaling pathways [5,6]. Pathway analysis further indicated enrichment in pathways such as regulation of the actin cytoskeleton, protein digestion and absorption, and ECM-receptor interaction. SERPINE1, a known robust prognostic marker in several cancers, has been closely linked to ECM-receptor interaction pathways in GC. Its high expression correlates with enrichment in these pathways, and in vitro studies show that SERPINE1 knockdown attenuates malignant behaviors in GC cells, potentially predicting prognosis and immunotherapy response [7,8]. From the common DEGs, 10 hub genes (COL3A1, FN1, COL1A1, COL1A2, BGN, THBS2, COL5A1, COL5A2, FBN1, TIMP1) were identified using STRING and Cytoscape. GEPIA database queries confirmed significant upregulation of their mRNA expression in GC tissues compared to normal stomach tissues from 408 patients. Survival analysis indicated that high expression of all except FN1 predicted poorer prognosis, underscoring their prognostic value. Protein-level validation via HPA confirmed high expression in GC for all genes except COL3A1, FBN1, and COL5A1, leading to the selection of COL1A1, COL1A2, BGN, THBS2, COL5A2, and TIMP1 as final candidate genes. Investigation into immune infiltration using TIMER revealed that these six genes correlated with infiltration levels of at least four types of immune cells, suggesting significant roles in GC pathogenesis.

COL1A1, COL1A2, and COL5A2 belong to the collagen family, crucial ECM structural components. In cancer, collagens contribute to tumorigenesis, progression, metastasis, and tissue fibrosis [9]. Upregulated COL1A1 and COL1A2 expression in colon cancer epithelium implicates them in angiogenesis and stromal formation during cancer progression [10]. Li *et al.* [11] reported elevated COL1A1 mRNA in premalignant and malignant tissues versus normal, and higher COL1A2 in malignant tissues, identifying them as prognostic factors in GC, consistent with our findings. COL1A1 can modulate proliferation and migration in GC cell lines, suggesting a role in early pathogenesis and potential as an early detection marker [12]. COL5A2 is implicated in focal adhesions and the PI3K-Akt pathway, essential for cell migration and angiogenesis in GC [13,14]. Reported dysregulation in various cancers and bioinformatics analyses suggest COL5A2 is a significant hub gene influencing GC prognosis [15]. While COL5A1 showed weak expression and was linked to poorer survival in our analysis, its role in GC requires further investigation due to the limited existing reports.

BGN, an extracellular matrix proteoglycan, is overexpressed in multiple cancers and predicts adverse outcomes. It associates with epithelial-mesenchymal transition (EMT) via integrating TGFβ/Snail and TNFα/NF-κB pathways within the TME [16]. EMT promotes the migration and invasion of tumor cells, thereby supporting the process of metastasis [17]. Bioinformatics techniques were utilized to detect and evaluate genes linked to gastric cancer, along with assessing their clinical relevance [18].

THBS2, which belongs to the family of matricellular proteins, plays a regulatory role in angiogenesis [19]. While some studies report lower THBS2 in GC correlating with better prognosis [20], others, including Zhang *et al.* [21], associate high THBS2 expression with poor prognosis and demonstrate that its knockdown inhibits GC cell proliferation and metastasis. It appears highly expressed during EMT, ECM remodeling, and invasion, indicating a complex role as a tumor regulator. TIMP1, an inhibitor of matrix metalloproteinases, can facilitate tumor invasion and metastasis when imbalanced [22]. In rectal cancer, TIMP1 levels distinguish patients from healthy controls and indicate disease progression [23]. In GC, elevated TIMP1 is associated with recurrence and poorer overall survival [24].

4. Conclusion

In summary, the six key genes COL1A1, COL1A2, BGN, THBS2, COL5A2, and TIMP1, which were screened and validated using bioinformatics methods in this study, may serve as biomarkers and potential therapeutic targets for GC diagnosis. However, the specific mechanisms remain unclear, and many studies have limitations, all of which require further validation of theoretical predictions in vivo and in vitro experiments.

Disclosure statement

The authors declare no conflict of interest.

References

[1] Smyth E, Nilsson M, Grabsch H, et al., 2020, Gastric Cancer. Lancet, 396(10251): 635–648.

[2] Conti C, Agnesi S, Scaravaglio M, et al., 2023, Early Gastric Cancer: Update on Prevention, Diagnosis and Treatment. Int J Environ Res Public Health, 20(3): 2149.

[3] Xu B, Shi Y, Yuan C, et al., 2024, Integrated Gene-Metabolite Association Network Analysis Reveals Key Metabolic Pathways in Gastric Adenocarcinoma. Heliyon, 10(17): e37156.

[4] Karamanos N, Piperigkou Z, Passi A, et al., 2021, Extracellular Matrix-Based Cancer Targeting. Trends Mol Med, 27(10): 1000–1013.

[5] Wieboldt R, Laubli H, 2022, Glycosaminoglycans in Cancer Therapy. Am J Physiol Cell Physiol, 322(6): C1187–C1200.

[6] Wei J, Hu M, Huang K, et al., 2020, Roles of Proteoglycans and Glycosaminoglycans in Cancer Development and Progression. Int J Mol Sci, 21(17): 5983.

[7] Chen S, Li Y, Zhu Y, et al., 2022, SERPINE1 Overexpression Promotes Malignant Progression and Poor Prognosis of Gastric Cancer. J Oncol, 2022: 2647825.

[8] Xu X, Zhang L, Qian Y, et al., 2022, A SERPINE1-Based Immune Gene Signature Predicts Prognosis and Immunotherapy Response in Gastric Cancer. Pharmaceuticals (Basel), 15(11): 1401.

[9] Cikach F, Koch C, Mead T, et al., 2018, Massive Aggrecan and Versican Accumulation in Thoracic Aortic Aneurysm and Dissection. JCI Insight, 3(5): e97167.

[10] Croix B, Rago C, Velculescu V, et al., 2000, Genes Expressed in Human Tumor Endothelium. Science, 289(5482): 1197–1202.

[11] Li J, Ding Y, Li A, 2016, Identification of COL1A1 and COL1A2 as Candidate Prognostic Factors in Gastric Cancer. World J Surg Oncol, 14(1): 297.

[12] Li A, Si J, Shang Y, et al., 2010, Construction of COL1A1 Short Hairpin RNA Vector and Its Effect on Cell Proliferation and Migration of Gastric Cancer Cells. Journal of Zhejiang University, 39(3): 257–263.

[13] Paluch E, Aspalter I, Sixt M, 2016, Focal Adhesion-Independent Cell Migration. Annu Rev Cell Dev Biol, 32: 469–490.

[14] Hu M, Zhu S, Xiong S, et al., 2019, MicroRNAs and the PTEN/PI3K/Akt Pathway in Gastric Cancer. Oncol Rep, 41(3): 1439–1454.

[15] Jin Y, Song X, Sun X, et al., 2023, Up-Regulation of Collagen Type V Alpha 2 (COL5A2) Promotes Malignant Phenotypes in Gastric Cancer Cell via Inducing Epithelial-Mesenchymal Transition (EMT). Open Med (Wars), 18(1): 20220593.

[16] Li H, Zhong A, Li S, et al., 2017, The Integrated Pathway of TGFβ/Snail with TNFα/NFκB May Facilitate the Tumor-Stroma Interaction in the EMT Process and Colorectal Cancer Prognosis. Sci Rep, 7(1): 4915.

[17] Schulz G, Grimm T, Sers C, et al., 2019, Prognostic Value and Association with Epithelial-Mesenchymal Transition and Molecular Subtypes of the Proteoglycan Biglycan in Advanced Bladder Cancer. Urol Oncol, 37(8): 530.e9–530.e18.

[18] Wu H, Xiang Z, Huang G, et al., 2023, BGN/FAP/STAT3 Positive Feedback Loop Mediated Mutual Interaction Between Tumor Cells and Mesothelial Cells Contributes to Peritoneal Metastasis of Gastric Cancer. Int J Biol Sci, 19(2): 465–483.

[19] Bornstein P, 2009, Thrombospondins Function as Regulators of Angiogenesis. J Cell Commun Signal, 3(3–4): 189–200.

[20] Sun R, Wu J, Chen Y, et al., 2014, Down Regulation of Thrombospondin2 Predicts Poor Prognosis in Patients with Gastric Cancer. Mol Cancer, 13(1): 225.

[21] Chang Z, Gao Y, Chen P, et al., 2024, THBS2 Promotes Gastric Cancer Progression and Stemness via the Notch Signaling Pathway. Am J Cancer Res, 14(7): 3433–3450.

[22] Thaysen-Andersen M, Thøgersen I, Lademann U, et al., 2008, Investigating the Biomarker Potential of Glycoproteins Using Comparative Glycoprofiling—Application to Tissue Inhibitor of Metalloproteinases-1. Biochim Biophys Acta, 1784(3): 455–463.

[23] Fuksiewicz M, Kotowicz B, Rutkowski A, et al., 2015, The Matrix Metalloproteinase-7 and Pro-Enzyme of Metalloproteinase-1 as a Potential Marker for Patients with Rectal Cancer Without Distant Metastasis. Tumour Biol, 36(5): 3629–3635.

[24] Grunnet M, Mau-Sørensen M, Brunner N, 2013, Tissue Inhibitor of Metalloproteinase 1 (TIMP-1) as a Biomarker in Gastric Cancer: A Review. Scand J Gastroenterol, 48(8): 899–905.