Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Multi-omics analysis identifies CLDN1 as a key regulator of gallbladder cancer progression and prognosis

Zhicheng Pan
Department of general surgery,Nanjing Red Cross Hospital,Nanjing 210002, China
Guodong Liu
Department of general surgery,Nanjing Red Cross Hospital,Nanjing 210002, China
Fei Tong
Department of general surgery,Nanjing Red Cross Hospital,Nanjing 210002, China
Zhenggen Hu
Department of general surgery,Nanjing Red Cross Hospital,Nanjing 210002, China
Cheng Qian
Department of general surgery,Nanjing Red Cross Hospital,Nanjing 210002, China
Shijing Xing
Department of general surgery,Nanjing Red Cross Hospital,Nanjing 210002, China
Xinyi Shen
Department of general surgery,Nanjing Red Cross Hospital,Nanjing 210002, China
Dayun Lu
School of Medicine, Nanjing University of Chinese Medicine, Nanjing 210023, China
Yun Xiao
Department of general surgery,Nanjing Red Cross Hospital,Nanjing 210002, China
Jianbo Han
hanjianbo1982@163.com

Department of general surgery,Nanjing Red Cross Hospital,Nanjing 210002, China

Research Article

Keywords: gallbladder cancer, transcriptome, scRNA sequencing, CLDN1

Posted Date: April 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9088571/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Gallbladder cancer (GBC) is a highly aggressive biliary tract malignancy characterized by poor prognosis and a lack of reliable biomarkers for early diagnosis. Despite its clinical aggressiveness, the molecular drivers and cellular regulatory networks underlying GBC progression remain incompletely characterized. In this study, we identified *CLDN1* as a key epithelial-associated gene in GBC and revealed an epithelial-centered immune regulatory network potentially involved in tumor progression. Integrated bulk transcriptome and single-cell RNA sequencing (scRNA-seq) analyses revealed that *CLDN1* is markedly upregulated in GBC, with expression localized predominantly in epithelial cells. At the single-cell level, epithelial cells emerged as a central hub of cell–cell communication, exhibiting extensive interactions with neutrophils, natural killer (NK) cells, and monocytes, thereby associating with immunosuppressive or inflammatory landscape of the tumor microenvironment (TME). Functional enrichment and intercellular interaction analyses suggested that *CLDN1*-high epithelial cells promote immune modulation via cytokine secretion and facilitate TME remodeling and metastasis through crosstalk with mesenchymal and endothelial cells. Notably, GBC epithelial cells exhibited significantly higher *CLDN1* expression compared to those in chronic cholecystitis, reinforcing its disease-specific relevance. By integrating differential expression analysis, weighted gene co-expression network analysis (WGCNA), and machine learning approaches, *CLDN1* was consistently identified as a key molecular feature of GBC. Collectively, our findings suggest that *CLDN1* is closely associated with GBC initiation and progression through modulation of epithelial stability and epithelial–immune crosstalk, highlighting its potential as a biomarker for tumor classification, prognosis, and therapeutic targeting.

Introduction

Gallbladder cancer (GBC) is the most common malignancy of the biliary system and ranks among the top six gastrointestinal neoplasms worldwide [1]. Globally, GBC accounts for more than 150,000 cancer-related deaths annually, representing approximately 1.5% of all cancer fatalities [2]. Its incidence exhibits marked geographic and demographic disparities, with high prevalence reported in Chile, India, parts of Asia, Eastern Europe, and South America [3–6]. GBC is more frequently diagnosed in elderly individuals, with a mean age of approximately 65 years, and occurs 2–6 times more often in women than in men [7]. These variations are thought to result from complex interactions between genetic susceptibility and environmental risk factors, including gallstones, liver fluke infection, chronic inflammation, abnormal pancreaticobiliary duct junctions, microbial infections, and exposure to toxic substances or carcinogens [8–11].

Clinically, GBC remains one of the most lethal biliary tract malignancies due to its insidious onset and lack of specific early symptoms. As a result, most patients are diagnosed at advanced stages, and the rate of potentially curative radical resection is limited to approximately 20% [12]. Although current diagnostic strategies rely on blood-based biomarkers combined with imaging modalities, their sensitivity and specificity are insufficient for detecting early-stage or small lesions [13]. Even with aggressive treatments such as radical cholecystectomy and adjuvant chemotherapy, postoperative recurrence and

distant metastasis are common. The absence of reliable biomarkers for early diagnosis and disease progression prediction underscores an urgent need to identify novel molecular markers and regulatory mechanisms underlying GBC.

Accumulating evidence suggests that genetic, epigenetic, and transcriptomic alterations contribute to GBC initiation and progression. Variations in genes involved in immune and inflammatory signaling, such as TLR4, have been associated with increased GBC risk, particularly in female patients [14]. Epigenetic dysregulation also plays an important role; for example, elevated expression of the histone acetyltransferase KAT5 promotes GBC cell proliferation, while its knockdown induces caspase-9–mediated apoptosis and suppresses tumor growth [15]. Transcriptome-wide analyses have revealed widespread dysregulation of mRNAs, lncRNAs, and microRNAs in GBC, implicating key oncogenic pathways including PI3K-Akt, Hedgehog, Wnt, KRAS, androgen receptor, and interferon-γ signaling [16, 17]. However, most existing studies focus on individual molecules or pathways, and a systematic understanding of how these alterations integrate at the cellular and microenvironmental levels remains lacking.

To address these limitations, we performed an integrative analysis combining bulk transcriptome and single-cell RNA sequencing (scRNA-seq) data from public databases[18–20]. Through transcriptome-based screening, we identified *CLDN1* as a core characteristic gene significantly upregulated in GBC tissues. Subsequent single-cell analysis revealed that epithelial cells act as central regulators within the GBC tumor microenvironment, orchestrating immune cell interactions and signaling networks. Together, our study highlights a potential *CLDN1*-driven epithelial–immune regulatory axis in GBC and provides new insights into disease mechanisms and candidate targets for diagnosis and therapy.

Materials and methodsData acquisition The bulk RNA-seq dataset GSE276931 and GSE139682 used in this study were retrieved from Gene Expression Omnibus (GEO) database. Tissue specimens of gallbladder cancer of GSE276931 dataset including 7 tumors and 5 adjacent non-tumors. Tissue specimens of gallbladder cancer of GSE139682 dataset including 10 tumors and 10 adjacent non-tumors. The single-cell RNA dataset used in this study was obtained from a previously published multi-omic analysis of gallbladder cancer [46], which contains 4 gallbladder cancer tissue (GBC) and 4 cholecystitis tissue (CC) single-cell RNA sequencing samples. Chronic cholecystitis tissue has a high degree of similarity in cellular composition with normal gallbladder tissue[21].

Differential gene expression analysis in bulk-RNAseq Differentially expressed genes (DEGs) between the tumors and adjacent non-tumors in the GSE276931 dataset and GSE139682 dataset were visualized using the limma package and met the criteria of P values < 0.05 and Fold change log2FC ≥ 1 or log2FC ≤ − 1.The results were visualized through heatmaps and volcano plots to highlight the significantly upregulated and downregulated genes.

Weighted gene co-expression network analysis (WGCNA) in bulk-RNAseq To further screen the key genes closely related to gallbladder atrophy, the weighted gene coexpression network analysis (WGCNA) [22] was used to construct the gene coexpression module for the dataset GSE276931. Quality control

was conducted by removing genes and samples with excessive missing values or low expression levels. The soft-thresholding power was chosen the scale free topology fit index reached 0.8 for the first time, ensuring that the constructed network approximates a scale-free topology. This thresholding was used to calculate the adjacency matrix representing gene co-expression similarity. Genes were clustered into modules using average linkage hierarchical clustering based on topological overlap, and each module was assigned a unique color. Finally, cluster dendrogram and module-trait relationships were visualized to interpret the results.

GO and KEGG pathway enrichment analysis in bulk-RNAseq After identifying the overlaping genes by intersecting WGCNA results and DEGs from GSE276931 dataset and GSE139682 dataset, respectively, we conducted KEGG and GO enrichment analyses. Functional enrichment analyses, including Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG), were conducted using the R package clusterProfiler (version 4.14.6). Gene annotations and mappings were performed based on the org.Hs.eg.db (version 3.20.0). Specifically, the enrichGO and enrichKEGG functions were employed to identify significantly enriched pathways (P-value < 0.05). The KEGG pathway data were sourced from the KEGG API (https://www.genome.jp/kegg/pathway.html), and GO terms were retrieved via the Gene Ontology resource (http://geneontology.org/).Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses[23] were employed to detect the molecular functions and potential pathways of differentially expressed genes (DEGs). GO analysis primarily identifies key features of genes concerning molecular functions(MF), cellular components(CC), and biological processes(BP). KEGG analysis mainly reveals gene enrichment in specific signaling pathways, elucidating the potential roles and mechanisms of genes in disease development. KEGG and GO enrichment analyses were presented using bubble Charts.

Machine learning in bulk-RNAseq We used Least Absolute Shrinkage and Selection Operator(LASSO), Support Vector Machine - Recursive Feature Elimination (SVM-RFE), and Random Forest (RF) [24]to the overlaping genes to futher identify key regulatory genes for GBC. Three models were constructed using the R package. LASSO uses L1 regularized linear regression method to achieve feature sparsity through the absolute value of penalty coefficient and automatically screen important variables. The SVM-RFE method selects the genes with the strongest discrimination ability based on the principle of minimum classification error and highest accuracy. RF selected the top 20 key genes according to the importance score of genes.

Single-cell sequencing data analysis The scRNA-seq data was processed by the Seurat package. Canonical Correlation Analysis (CCA) was used to find Mutual Nearest Neighbours (MNNs)[25]. Cells with more than 2,500 or fewer than 300 gene counts was filtered out. Subsequently, cells exceeding a mitochondrial gene percentage of 5% was filtered out to eliminate low-quality cells. To identify cell clusters, PCA analysis was first performed on the list of highly variable genes. Clustering was then performed using the findclusters function. Finally, Umap was used for visualization. For normalized gene expression data, we listed markers for each cell cluster using the findallmarkers function. A total of 12 cell types were finally obtained.**Cell–cell interaction analysis**

To visualize and analyze intercellular communications, we conducted CellChat analysis[26]. CellChat identified differentially overexpressed ligands and receptorsfor each cell group and associated each interaction with a probability value to quantify communications between the two cell groups mediated by these signaling genes. Significant interactions were identified on the basis of a statistical test that randomly permuted the group labels of cells and then recalculated the interaction probability. Cell interaction with a P value < 0.05 was defined as significance.The results were visualized using a Circle plot. **Differential gene expression analysis in scRNA-seq**

To identify the differential expression genes (DEGs) of epithelial cells, we used model-based analysis of single-cell transcriptomics test with genes detected in a minimum of 10% of all cells, met the criteria of P values < 0.05 and |Fold change log2FC| ≥ 1 in the Findmarker function by Seurat.

Enrichment analysis in scRNA-seq After identifying the DEGs from scRNA-seq, we conducted KEGG and GO enrichment analyses.Gene enrichment analysis was conducted using the R package clusterProfiler. KEGG and GO enrichment analyses were presented using bar Charts. Subsequently, we used R software GSVA package to conduct ssGSEA to calculate the inflammation-related pathways score[27]. The result was visualized into heatmaps.

Results

Differential expression analysis of genes and functional analysis of DEGs We identified a total of 1387 differential expression genes (DEGs) from GSE276931 between the GBC and CC using the limma package with multiple testing correction. Applying a P value of < 0.05 and |Fold change log2FC| ≥ 1, 776 genes were significantly upregulated and 611 genes were significantly downregulated. The expression patterns of these DEGs are shown in the volcano plot (Fig. 1A). The Top50 gene of these DEGs are shown in the heatmap (Fig. 1B).In GSE139682, a total of 791 DEGs were identified, among them, 195 genes were significantly upregulated and 596 genes were significantly downregulated. The expression patterns of these DEGs are shown in the volcano plot (Fig. 1C). The Top50 gene of these DEGs are shown in the heatmap (Fig. 1D).

Identification of GBC-associated co-expression modules and functional signatures via WGCNA

To identify critical genes associated with GBC, weighted gene co-expression network analysis (WGCNA) was performed on 1,387 DEGs from the GSE276931 dataset to construct co-expression modules. Using a soft-threshold power of 30 (scale-free $R^2 = 0.77$), a total of 16 co-expression modules were identified (Fig. 2A). Among these, the black, yellow, and lavenderblush3 modules exhibited significant positive correlations with GBC (P < 0.05), with the black module showing the highest correlation coefficient and the most significant P value (Fig. 2B–C). By intersecting genes within the black module with DEGs from both GSE276931 and GSE139682, we identified 32 overlapping genes (Fig. 2D).Functional enrichment analyses indicated that these genes were primarily involved in GO terms related to cell junction

organization and epithelial–mesenchymal transition (EMT). Specifically, "apical junction complex" and "bicellular tight junction" were significantly enriched in the cellular component category, while "substrate adhesion-dependent cell spreading," "mitotic metaphase plate congression," and "positive regulation of protein kinase activity" dominated the biological process category (Fig. 2E). Furthermore, KEGG pathway analysis revealed significant enrichment in the *IL-17* signaling pathway, tight junctions, and cell adhesion molecules, suggesting that GBC progression is driven by inflammatory-mediated immune responses and junctional remodeling (Fig. 2F). Additionally, the enrichment of amino acid and drug metabolism pathways points toward systemic metabolic reprogramming and altered detoxification functions during GBC development.

Machine learning and screening of key regulatory genes To further screen the key regulatory genes in the process of gallbladder cancer, applied three machine learning methods-LASSO, SVM-RFE, and RF, to the 32 overlapping genes. LASSO identified 3 characteristic genes- *CLDN1*, *SHANK2* and *EFNA1* significantly associated with GBC from top 10 genes through lasso coefficient path analysis and cross validation error (Fig. 3A). Random forest analysis selected the top 20 key genes according to the importance score of genes (Fig. 3B-C). The SVM-RFE method selected the gene set with the strongest discrimination ability based on the principle of minimum classification error and highest accuracy (Fig. 3D). we focused on the intersection of genes consistently identified across all 3 methods, resulting in one common gene: *CLDN1* (Fig. 3E). The expression level of the *CLDN1* gene in the GBC group is significantly higher than that in the adjacent cancer group (Fig. 3F).

Integration and clustering of scRNA-seq data

Single-cell analysis was performed on openly available datasets. Cells were classified into 12 different distinct clusters following umap visual analysis after quality control and preprocessing of the data (Fig. 4A). Subsequently, we provided detailed annotations on the cell subpopulations of the GBC group and CC group. By analyzing the expression of established molecular markers, the specific types of cells we have identified were mainly immune cells and structural cells, such as: T cells, monocytes, NK cells, B cells, dendritic cells (DCS), Epithelial cells, Endothelial cells, Mesenchymal cells, Neutrophils, Mast cells and plasma cells, etc. in GBC group and CC group (Fig. 4B-C). Although the cell types of GBC group and CC group were similar, there were some differences in the proportion and distribution pattern (Fig. 4E).

In the CC group, T cells dominate, followed by Mononuclear cells and mesenchymal cells, which exhibited stronger cellular immune response characteristics and retains more mesenchymal matrix components in the cellular microenvironment. In the GBC group, mononuclear cells were the predominant population, followed by mesenchymal cells, reflecting enhanced innate immune activity and matrix remodeling. The accuracy of cell type annotation was further confirmed by typical marker gene expression heatmaps (Fig. 4D).

Epithelial cells were located in the core regulatory position in GBC

To quantify intercellular communications, we employed CellChat to construct signaling networks within the GBC microenvironment. Epithelial cells emerged as the central hub of the interaction network, exhibiting significant crosstalk with neutrophils, monocytes, NK cells, and stromal cells (Fig. 5A). Compared to the balanced interaction pattern in chronic cholecystitis (CC) (Fig. 5C), GBC tissues showed distinct remodeling of the tumor microenvironment (TME) (Fig. 5B). Notably, interactions between epithelial cells and neutrophils or mesenchymal cells were markedly enhanced in GBC, with the latter showing the most significant intensity, potentially driving TME formation (Fig. 5D).Further investigation using pseudo-temporal trajectory analysis revealed that epithelial cells in GBC exhibited a more extensive distribution compared to CC (Fig. 5E). Specifically, a subset of cells in GBC shifted toward advanced evolutionary stages, suggesting they may have undergone epithelial–mesenchymal transition (EMT) to promote tumor progression. Collectively, these findings indicate that epithelial cells act as a primary orchestrator of immune regulation and structural remodeling in GBC.

Differential expression analysis and functional enrichment of epithelial cells

To elucidate the molecular characteristics of epithelial cells in GBC, we performed differential expression and functional enrichment analyses. We identified 3,302 significantly upregulated and 2,976 downregulated genes in GBC epithelial cells compared to the CC group (Fig. 6A). GO enrichment analysis (Fig. 6B) revealed that these genes were primarily associated with neutrophil-mediated immunity, antigen processing and presentation (MHC class I), and cell–substrate junctions. These findings suggest that GBC epithelial cells actively orchestrate immune responses and facilitate intercellular signaling through specialized membrane structures. KEGG pathway analysis further highlighted the enrichment of phagosome, cell cycle, and endocytosis pathways (Fig. 6C), suggesting their contribution to the malignant phenotype of GBC. Notably, inflammatory pathways were consistently enriched across both bulk and single-cell datasets. Subsequent ssGSEA scoring confirmed that GBC epithelial cells possessed significantly higher inflammation scores than those in the CC group (Fig. 6D), reflecting a pro-inflammatory state within the tumor microenvironment that may drive disease progression.

Expression of CLDN1 in epithelial cells

We expressed the *CLDN1* gene overlapped by machine learning in the U-Map plot and found that it is mainly expressed in epithelial cells (Fig. 7A-B).Comparing the expression levels of *CLDN1* in the epithelial cells of CC group and GBC group, it was found that the expression level in GBC group was significantly higher than that of CC group with $P < 0.05$ (Fig. 7C).

Discussion

Gallbladder cancer (GBC) is typically diagnosed at advanced stages, rendering early detection a formidable clinical challenge. While early-stage diagnosis significantly improves survival, the overall prognosis remains suboptimal, with only a fraction of patients achieving favorable long-term outcomes

[29, 30]. These therapeutic limitations underscore an urgent imperative to identify robust biomarkers and to elucidate the fundamental biological mechanisms driving GBC progression.

In this study, we identified *CLDN1* as a core molecular feature of GBC and, for the first time, systematically linked its expression to an epithelial-centered immune regulatory network at single-cell resolution. Moving beyond the traditional focus on isolated genes, our findings illuminate the pivotal role of epithelial cells as primary coordinators of tumor progression and immune landscape modulation in GBC.

CLDN1 is a scaffold component of tight junction complexes, which, alongside occludin and junctional adhesion molecules, maintains epithelial barrier integrity and intercellular cohesion [31]. Dysregulated *CLDN1* has been implicated in the initiation and invasion of various malignancies, driven by oncogenic signaling, genetic aberrations, and microenvironmental cues. For instance, BRAF V600E mutations upregulate *CLDN1* via NF-κB activation in thyroid cancer [32], with parallel associations observed in colorectal polyps [33]. In head and neck squamous cell carcinoma and hepatocellular carcinoma (HCC), *CLDN1* overexpression facilitates tumor advancement through Wnt/β-catenin signaling [34], contributing to epithelial–mesenchymal transition (EMT), metabolic reprogramming, and tumor immune microenvironment (TIME) remodeling [35].

While the specific role of *CLDN1* in GBC has remained largely elusive, our functional enrichment analyses suggest it exerts analogous oncogenic effects. GO and KEGG pathways were significantly enriched in substrate adhesion, cadherin binding, and tight junction organization—processes integral to cell migration and metastatic potential [36–38]. The disruption of these adhesive architectures often triggers EMT, thereby enhancing tumor aggressiveness [37, 38]. Crucially, these functional patterns were predominantly localized within epithelial cells, reinforcing their role as the primary engine of GBC progression.

Beyond structural maintenance, we discovered that epithelial cells possess potent immunomodulatory properties. Our cell–cell communication analysis identified epithelial cells as a central hub within the interaction network, orchestrating crosstalk between neutrophils, natural killer (NK) cells, and monocytes. Epithelial cells are known to secrete a spectrum of cytokines, such as *IL-1, IL-8, IL-18, TNF-α*, and *TGF-β*, which act as critical mediators of inflammatory responses and immune polarization [38–40]. Furthermore, epithelial-derived metabolites like retinoic acid are essential for maintaining immune homeostasis [41]. Our findings support a paradigm shift: viewing GBC epithelial cells not merely as structural units but as active immunometabolic regulators of the TME.

The interplay between epithelial-mesenchymal plasticity and the TME is a hallmark of metastasis. The intensified interactions between epithelial, mesenchymal, and endothelial cells observed in our study likely facilitate microenvironmental remodeling and systemic dissemination. This aligns with evidence that microRNA clusters, such as miR-214/miR-3120, can modulate EMT and autophagy to suppress GBC progression [42]. Consequently, this epithelial-centered interaction axis represents a critical vulnerability in GBC pathogenesis that could be exploited for therapeutic intervention.

Finally, our single-cell validation confirmed that *CLDN1* is both cell-type specific and significantly upregulated in GBC compared to chronic cholecystitis, emphasizing its diagnostic relevance. Given that *CLDN1* correlates with prognosis in colorectal cancer [43], and considering the clinical success of *CLDN1*8.2-targeted therapies in gastric cancers [44, 45], *CLDN1* emerges as a compelling target for GBC. Our data suggest that *CLDN1* could serve as a versatile tool for early diagnosis, molecular classification, and the development of precision therapies.

In conclusion, our study identifies a previously unrecognized *CLDN1*–epithelial–immune regulatory axis in GBC. These insights not only deepen our understanding of GBC biology but also provide a conceptual framework for developing epithelial-targeted diagnostic and therapeutic strategies.

Despite the insights gained from our multi-omics analysis, this study has several limitations. First, the sample sizes of the primary datasets (GSE276931 and GSE139682) are relatively small, which may affect the generalizability of the findings. Second, while CLDN1 was identified as a key regulator through bioinformatics, further *in vitro* and *in vivo* experimental validations are required to confirm its functional roles and underlying mechanisms in gallbladder cancer. Future studies with larger clinical cohorts and functional assays are warranted to strengthen these conclusions.

Conclusion

This study integrates bulk and single-cell transcriptomic analyses to elucidate key regulatory mechanisms in gallbladder cancer. We demonstrate that epithelial cells act as the central hub of the cell–cell interaction network, coordinating immune regulation through interactions with neutrophils, natural killer cells, and monocytes, while crosstalk with mesenchymal and endothelial cells may promote tumor microenvironment remodeling and metastasis. Notably, *CLDN1* was identified as a core gene predominantly expressed in epithelial cells and may contribute to GBC progression by regulating epithelial integrity, inflammatory responses, and the tumor microenvironment. These findings suggest that *CLDN1* and epithelial-centered regulatory networks represent promising candidates for tumor classification, targeted therapy, and prognosis assessment, warranting further functional and clinical validation.

Declarations

Authors contribution

Jianbo Han, Yun Xiao, and Dayun Lu conceptualized the study, developed the methodology, and supervised the project and manuscript revision. Zhicheng Pan, Guodong Liu, and Fei Tong executed data collection, performed machine learning-based formal analysis, and drafted the original manuscript. Zhenggen Hu, Cheng Qian, Shijing Xing, and Xinyi Shen contributed to data curation, methodological implementation, and the visualization of research results.

All authors read and approved the final manuscript.

Conflict of interests

The authors declare no competing interests.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data Availability Statement

The datasets analyzed in this study are publicly available in the Gene Expression Omnibus (GEO) repository. The bulk RNA-seq data can be accessed via the following links: **GSE276931** (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE276931) and **GSE139682** (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE139682). The single-cell RNA-seq data of gallbladder cancer (Reference [46]) is available at [https://zenodo.org/records/15400138]. Functional enrichment resources can be accessed via **Gene Ontology** (http://geneontology.org/) and **KEGG** (https://www.genome.jp/kegg/pathway.html).

Ethics declaration

Ethics declaration: Not applicable.

Clinical trial number

Clinical trial number: not applicable.

Consent to Participate declaration

Consent to Participate declaration: not applicable.

Consent to Publish declaration

Consent to Publish declaration: not applicable.

References

1. Song, X., *et al*., Overview of current targeted therapy in gallbladder cancer. Signal Transduct Target Ther, 2020. **5**(1): p. 230.
2. Rawla, P., *et al*., Epidemiology of gallbladder cancer. Clin Exp Hepatol, 2019. **5**(2): p. 93-102.
3. Krell, R.W. and A.C. Wei, Gallbladder cancer: surgical management. Chin Clin Oncol, 2019. **8**(4): p. 36.
4. Thomas, T.S., *et al*., Advancing age and the risk of bleomycin pulmonary toxicity in a largely older cohort of patients with newly diagnosed Hodgkin Lymphoma. J Geriatr Oncol, 2020. **11**(1): p. 69-74.

5. Kumar, A., *et al*., Gallbladder cancer: Progress in the Indian subcontinent. World J Clin Oncol, 2024. **15**(6): p. 695-716.
6. Noel, R., *et al*., Does the Frequency of Cholecystectomy Affect the Ensuing Incidence of Gallbladder Cancer in Sweden? A Population-Based Study with a 16-Year Coverage. World J Surg, 2016. **40**(1): p. 66-72.
7. Van Dyke, A.L., *et al*., Biliary tract cancer incidence and trends in the United States by demographic group, 1999-2013. Cancer, 2019. **125**(9): p. 1489-1498.
8. Shin, H.R., *et al*., Epidemiology of cholangiocarcinoma: an update focusing on risk factors. Cancer Sci, 2010. **101**(3): p. 579-85.
9. Rawla, P., T. Sunkara, and A. Barsouk, Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. Prz Gastroenterol, 2019. **14**(2): p. 89-103.
10. Dixit, R., *et al*., Comparative Analysis of Mutational Profile of Sonic hedgehog Gene in Gallbladder Cancer. Dig Dis Sci, 2017. **62**(3): p. 708-714.
11. Lee, W. and V.S. Chandan, Gallbladder carcinomas: review and updates on morphology, immunohistochemistry, and staging. Hum Pathol, 2023. **132**: p. 149-157.
12. Ren, T., *et al*., [Analysis of treatment modalities and prognosis of patients with gallbladder cancer in China from 2010 to 2017]. Zhonghua Wai Ke Za Zhi, 2020. **58**(9): p. 697-706.
13. Deo, K.B., *et al*., Misdiagnosis of carcinoma gall bladder in endemic regions. BMC Surg, 2022. **22**(1): p. 343.
14. Seretis, C., *et al*., Metaplastic changes in chronic cholecystitis: implications for early diagnosis and surgical intervention to prevent the gallbladder metaplasia-dysplasia-carcinoma sequence. J Clin Med Res, 2014. **6**(1): p. 26-9.
15. Feinberg, A.P. and B. Tycko, The history of cancer epigenetics. Nat Rev Cancer, 2004. **4**(2): p. 143-53.
16. Dixit, R., *et al*., Unravelling of the comparative Transcriptomic Profile of Gallbladder Cancer using mRNA sequencing. Mol Biol Rep, 2022. **49**(7): p. 6395-6403.
17. Saxena, R., *et al*., Next generation sequencing uncovers multiple miRNAs associated molecular targets in gallbladder cancer patients. Sci Rep, 2023. **13**(1): p. 19101.
18. Ozsolak, F. and P.M. Milos, RNA sequencing: advances, challenges and opportunities. Nat Rev Genet, 2011. **12**(2): p. 87-98.
19. Zhang, X., F. Zhang, and X. Xu, Single-cell RNA sequencing in exploring the pathogenesis of diabetic retinopathy. Clin Transl Med, 2024. **14**(7): p. e1751.
20. Van Hove, I., *et al*., Single-cell transcriptome analysis of the Akimba mouse retina reveals cell-type-specific insights into the pathobiology of diabetic retinopathy. Diabetologia, 2020. **63**(10): p. 2235-2248.
21. Zhou, T., *et al*., Multi-omic analysis of gallbladder cancer identifies distinct tumor microenvironments associated with disease progression. Nat Genet, 2025. **57**(8): p. 1935-1949.

22. Langfelder, P. and S. Horvath, WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics, 2008. **9**: p. 559.
23. Kanehisa, M., *et al*., KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res, 2023. **51**(D1): p. D587-d592.
24. Alazaidah, R., *et al*., Potential of Machine Learning for Predicting Sleep Disorders: A Comprehensive Analysis of Regression and Classification Models. Diagnostics (Basel), 2023. **14**(1).
25. Slovin, S., *et al*., Single-Cell RNA Sequencing Analysis: A Step-by-Step Overview. Methods Mol Biol, 2021. **2284**: p. 343-365.
26. Jin, S., *et al*., Inference and analysis of cell-cell communication using CellChat. Nat Commun, 2021. **12**(1): p. 1088.
27. Hänzelmann, S., R. Castelo, and J. Guinney, GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics, 2013. **14**: p. 7.
28. Liu, L., *et al*., SIRT3 inhibits gallbladder cancer by induction of AKT-dependent ferroptosis and blockade of epithelial-mesenchymal transition. Cancer Lett, 2021. **510**: p. 93-104.
29. Feo, C.F., *et al*., Current management of incidental gallbladder cancer: A review. Int J Surg, 2022. **98**: p. 106234.
30. Roa, J.C., *et al*., Gallbladder cancer. Nat Rev Dis Primers, 2022. **8**(1): p. 69.
31. Ding, L., *et al*., The claudin family of proteins in human malignancy: a clinical perspective. Cancer Manag Res, 2013. **5**: p. 367-75.
32. Ross-Orozco, A., *et al*., Association between BRAFV600E mutation and gene expression of CLDN1, TIMP1, and KRT19 in papillary thyroid cancer. Mol Cell Endocrinol, 2025. **607**: p. 112614.
33. Caruso, M., *et al*., Claudin-1 Expression Is Elevated in Colorectal Cancer Precursor Lesions Harboring the BRAF V600E Mutation. Transl Oncol, 2014. **7**(4): p. 456-63.
34. Kayamori, K., *et al*., Role of Stromal Fibroblast-Induced WNT7A Associated with Cancer Cell Migration Through the AKT/CLDN1 Signaling Axis in Oral Squamous Cell Carcinoma. Lab Invest, 2023. **103**(10): p. 100228.
35. Wang, W., *et al*., Claudins and hepatocellular carcinoma. Biomed Pharmacother, 2024. **171**: p. 116109.
36. Al Azim, M. and J.S. Di Martino, ECM, integrins, and DDRs: A nexus of cancer progression, therapy, and future directions. Matrix Biol, 2025. **138**: p. 27-43.
37. Danilova, N.V. and N.A. Oleynikova, [E-cadherin: structure and functions, role in gastric cancer carcinogenesis]. Arkh Patol, 2023. **85**(4): p. 70-77.
38. Andrews, C., M.H. McLean, and S.K. Durum, Cytokine Tuning of Intestinal Epithelial Function. Front Immunol, 2018. **9**: p. 1270.
39. Scarpa, M., *et al*., The epithelial danger signal IL-1α is a potent activator of fibroblasts and reactivator of intestinal inflammation. Am J Pathol, 2015. **185**(6): p. 1624-37.

40. Eckmann, L., M.F. Kagnoff, and J. Fierer, Epithelial cells secrete the chemokine interleukin-8 in response to bacterial entry. Infect Immun, 1993. **61**(11): p. 4569-74.
41. Cao, Y.G., *et al*., Faecalibaculum rodentium remodels retinoic acid signaling to govern eosinophil-dependent intestinal epithelial homeostasis. Cell Host Microbe, 2022. **30**(9): p. 1295-1310.e8.
42. Li, W., *et al*., The microRNA cluster miR-214/miR-3120 prevents tumor cell switching from an epithelial to a mesenchymal-like phenotype and inhibits autophagy in gallbladder cancer. Cell Signal, 2021. **80**: p. 109887.
43. Ouban, A., Claudin-1 role in colon cancer: An update and a review. Histol Histopathol, 2018. **33**(10): p. 1013-1019.
44. Nakayama, I., *et al*., Claudin 18.2 as a novel therapeutic target. Nat Rev Clin Oncol, 2024. **21**(5): p. 354-369.
45. Shah, M.A., *et al*., Zolbetuximab plus CAPOX in CLDN18.2-positive gastric or gastroesophageal junction adenocarcinoma: the randomized, phase 3 GLOW trial. Nat Med, 2023. **29**(8): p. 2133-2141.
46. Multi-omic analysis of gallbladder cancer identifies distinct tumor microenvironments associated with disease progression. https://zenodo.org/records/15400138. Accessed 03 Feb 2026.

Figures

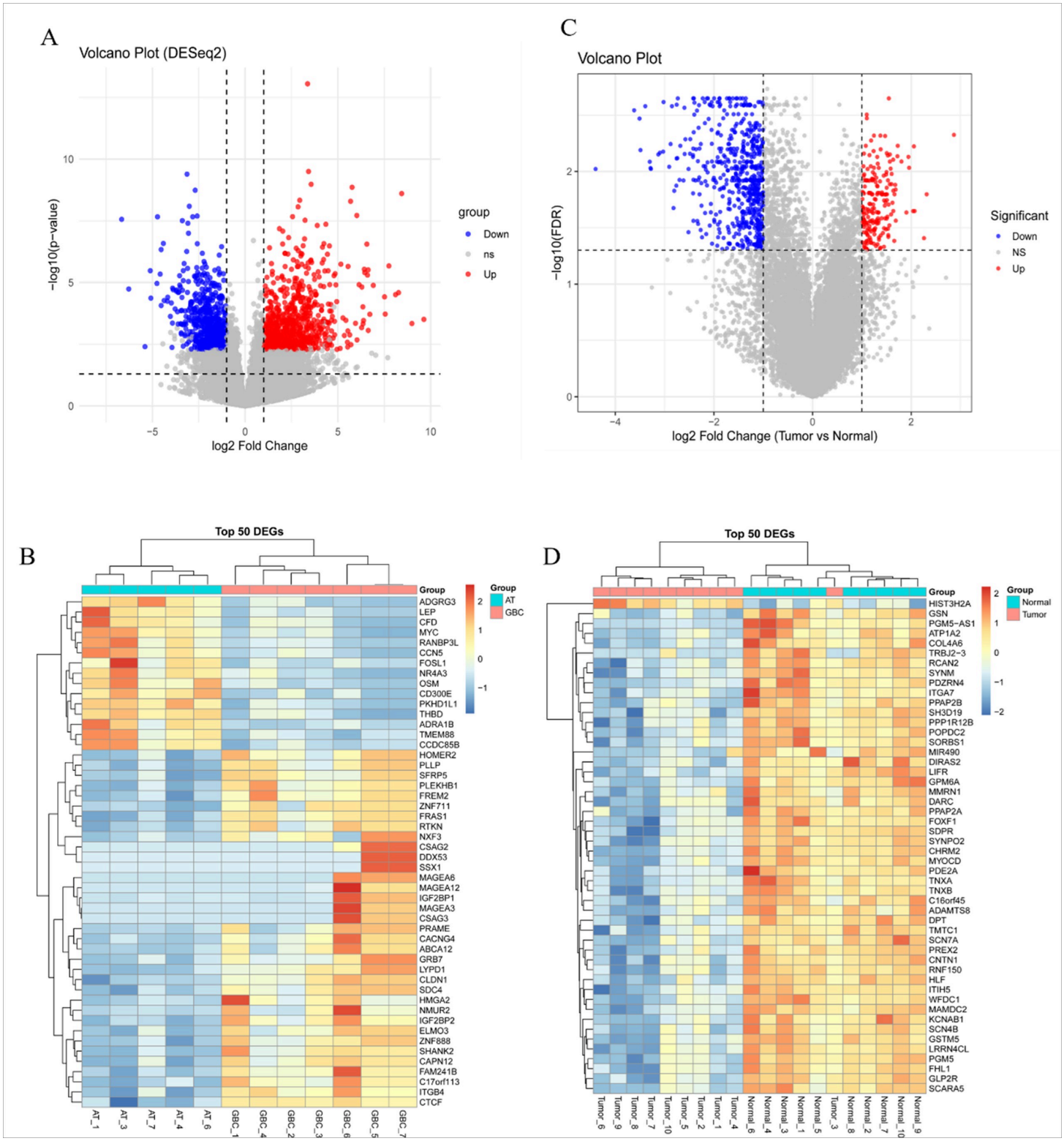


Figure 1

Different expression genes identified from GSE276931 and GSE139682 between tumor group and adjacent tumor group. (A) The red and blue dots refer to DEGs with upregulated and downregulated gene expression, respectively, from GSE276931. **(B)** Hierarchical clustering analysis of the DEGs from GSE276931. The color bar was employed to represent the expression levels of the DEGs, where red color was indicative of high expression and blue color was indicative of low expression. **(C)** The red and blue

dots refer to DEGs with upregulated and downregulated gene expression, respectively, from GSE139682. **(D)** Hierarchical clustering analysis of the DEGs from GSE139682. The color bar was employed to represent the expression levels of the DEGs, where red color was indicative of high expression and blue color was indicative of low expression.

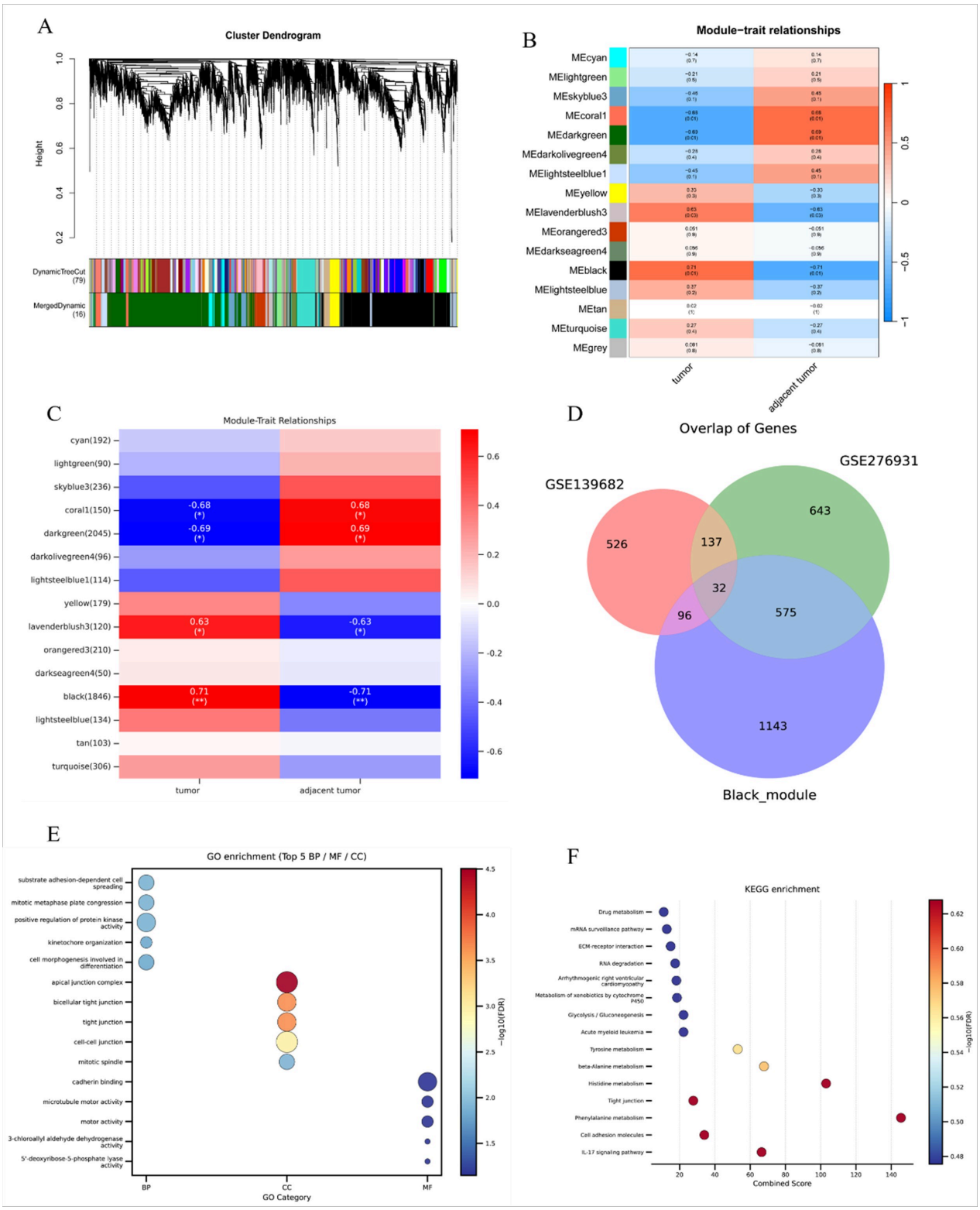


Figure 2

Co-expression network of DEGs from GSE276931 and Screening and functional enrichment analysis of key genes associated with GBC. (A) Cluster tree dendrogram of the co-expression modules. Different colors represent distinct co-expression modules. **(B)** Module-Trait Relationships Heatmap, correlation analysis between module eigengenes and clinical status. Each row represents a module; each column represents a clinical status. **(C)** Module-Trait Relationships barplot, correlation coefficients between modules and traits are expressed as bar lengths, * $p < 0.05$, * * $p < 0.01$, Numbers in parentheses indicate the number of genes included in the module. **(D)** The Venn plot shows that 32 overlapping genes were identified by intersecting black module genes from WGCNA and DEGs from GSE276931 and GSE139682, respectively. **(E)**Go functional enrichment of overlapping genes, including biological processes, cellular components, and molecular functions. **(F)** KEGG analysis of overlapping genes.

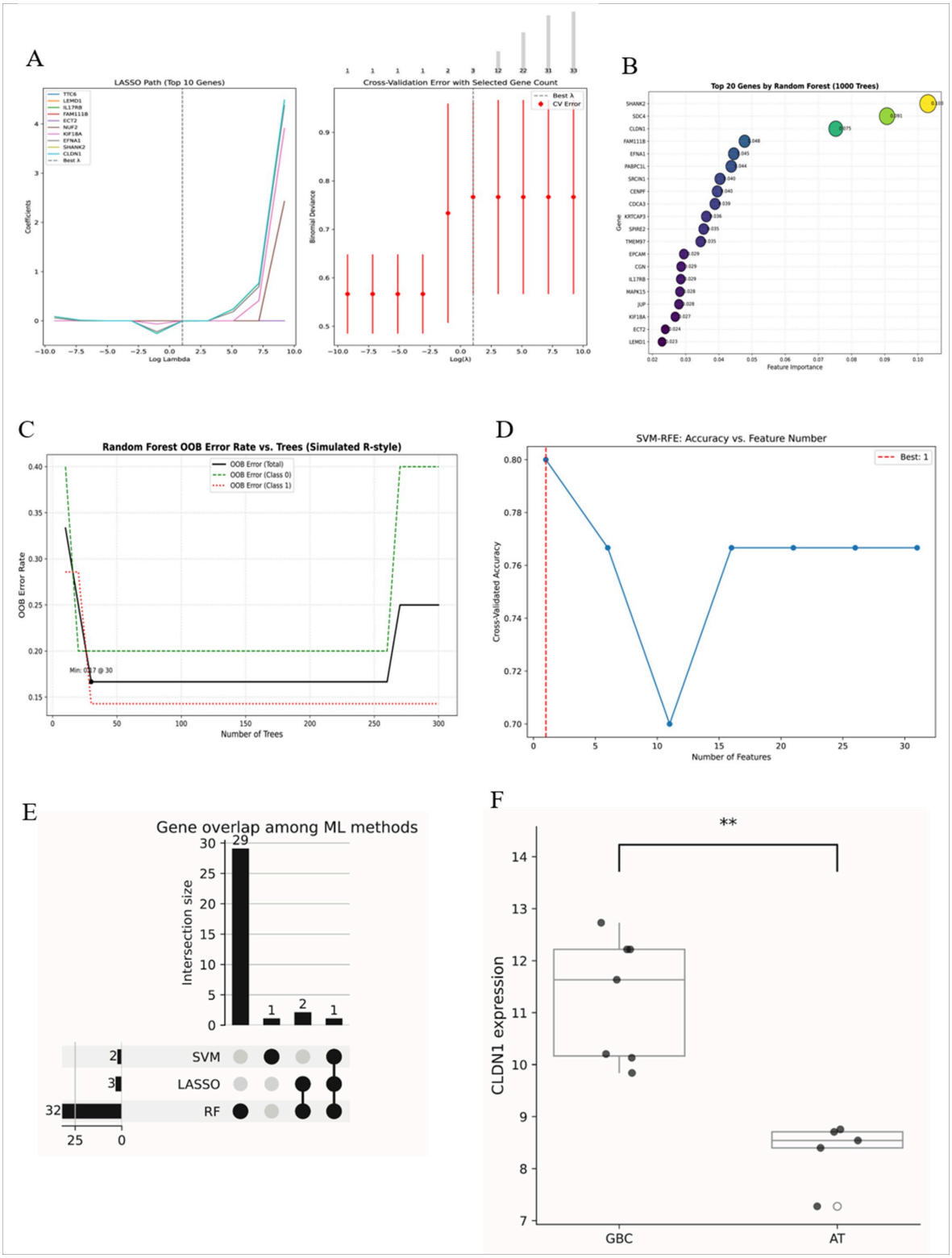


Figure 3

Three machine learning methods for identifying the core characteristic genes. (A) Lasso regression analysis to screen GBC key genes, Left: the coefficient of top 10 gene changes with log lambda; Right: cross validation error curve. **(B)** Top 20 gallbladder cancer related genes screened by random forest model and their feature importance scores. **(C)** Variation curve of OOB error rate of random forest with the number of trees. **(D)** Variation curve of cross validation accuracy of SVM-RFE model with the number

of features. **(E)** Wayne diagram of core characteristic genes screened by SVM-RFE, Lasso, and random forest. **(F)** Expression of *CLDN1* gene in tumor group and adjacent tumor group.

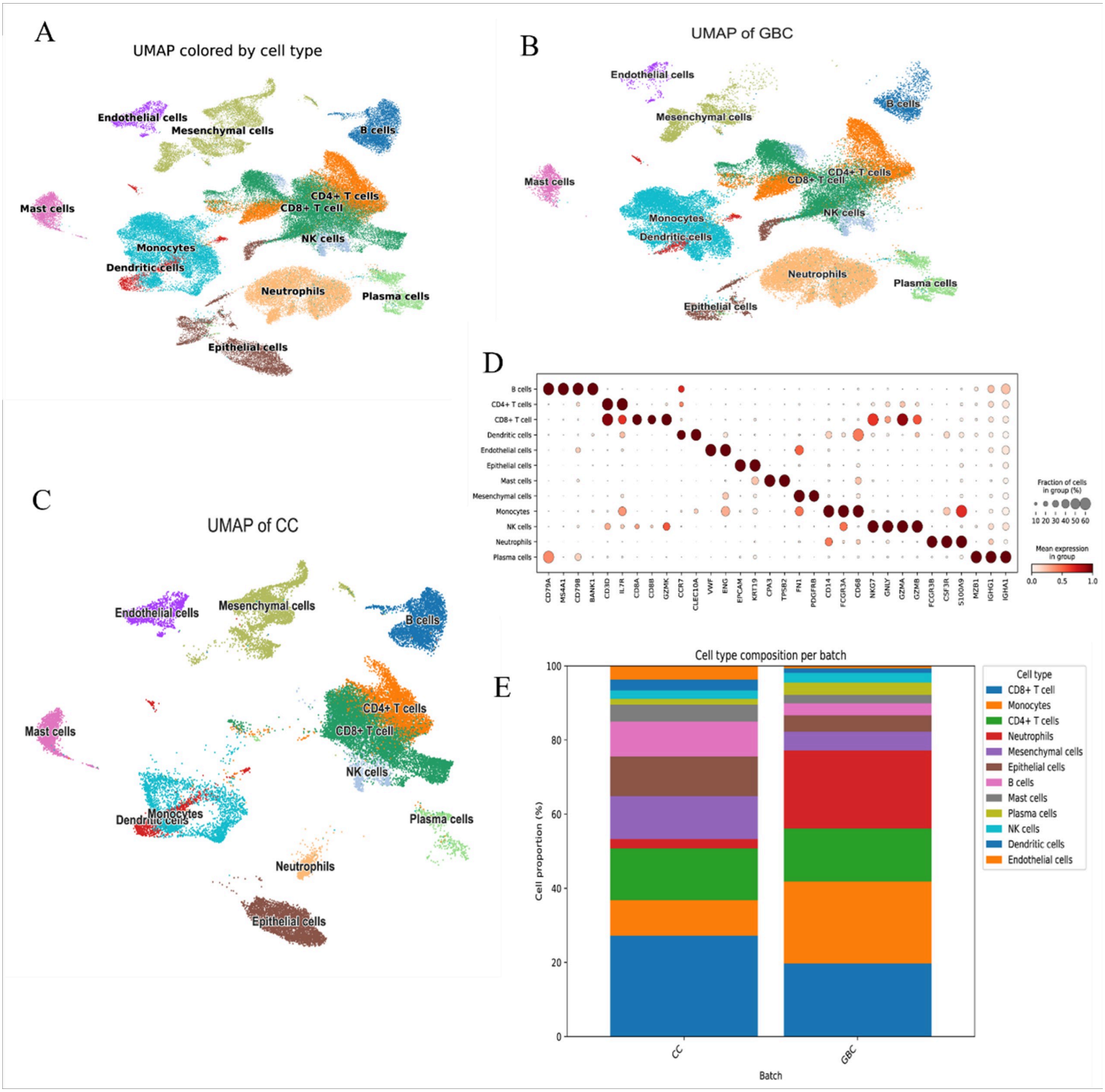


Figure 4

scRNA-seq analysis of gallbladder cancer tissue (GBC) and cholecystitis tissue (CC). (A) 12 different distinct clusters analysis by PCA and UMAP. **(B)** U-MAP plot showing clusters of cells in GBC. **(C)** U-MAP plot showing clusters of cells in CC. **(D)** Heatmap showing the expression of marker genes in each cell type. **(E)** Proportion of cell type clusters present in each GBC and CC groups.

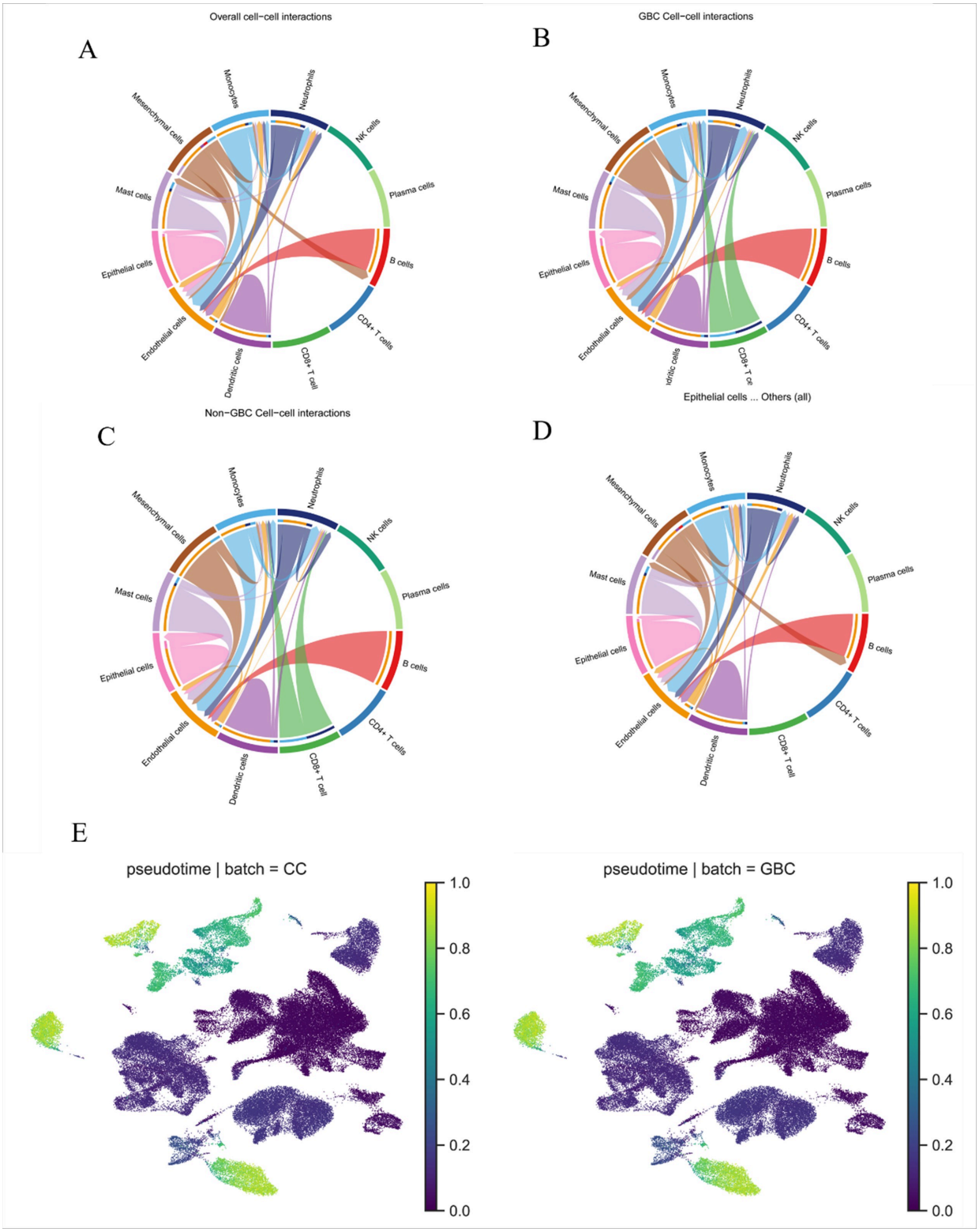


Figure 5

Cell–cell interaction analysis and pseudotime analysis. (A)overall cell–cell interaction. **(B)** GBC cell–cell interaction. **(C)** Non-GBC cell–cell interaction. **(D)** Epithelial cells-other interaction. **(E)** Pseudotime analysis.

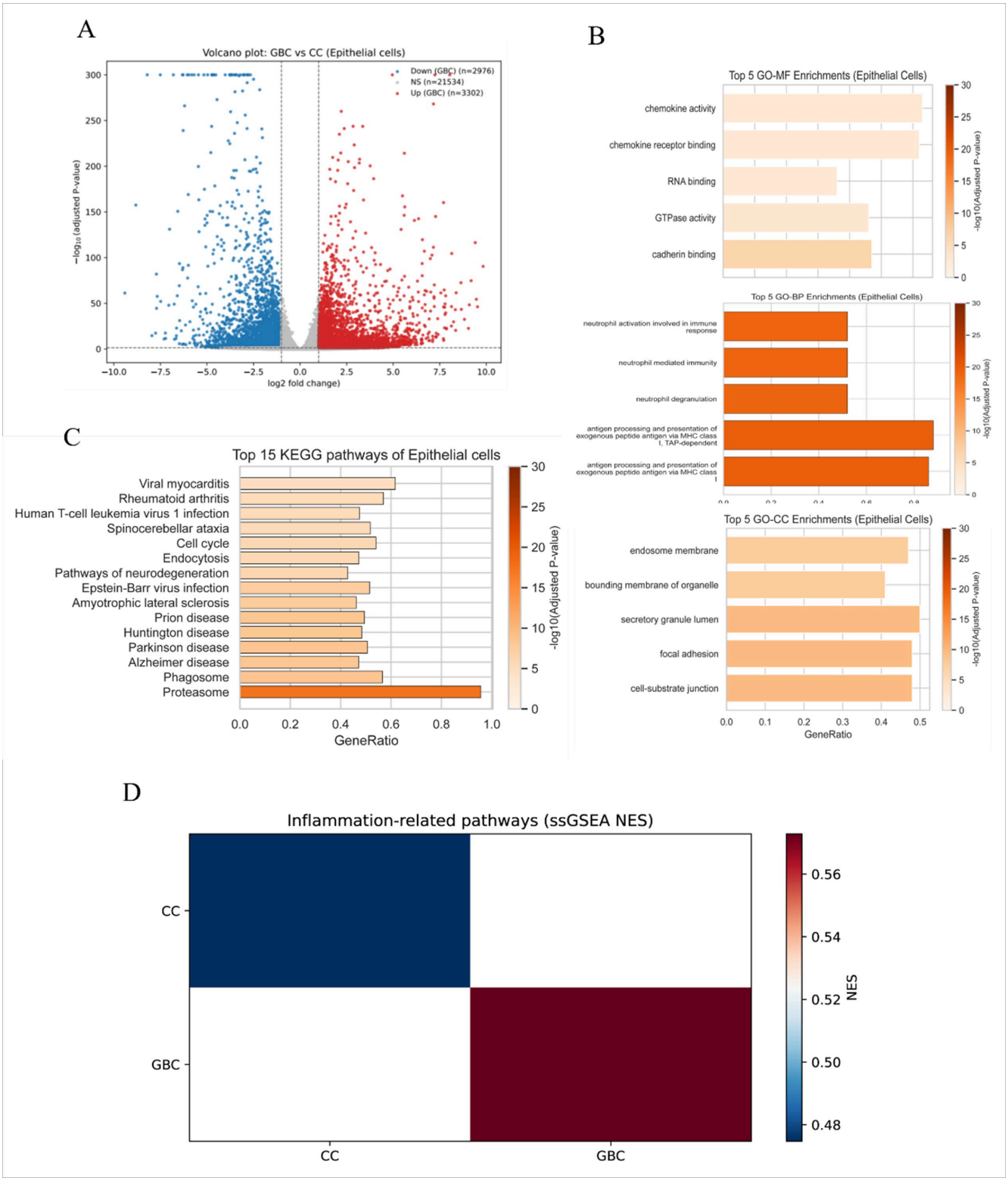


Figure 6

Functional enrichment of epithelial cells. (A) Differential expression analysis on epithelial cells. **(B)** Top 5 GO terms in BP, CC and MF of epithelial cells. **(C)** Top 15 KEGG pathways of epithelial cells. **(D)** The ssGSEA analysis of inflammation-related pathways.

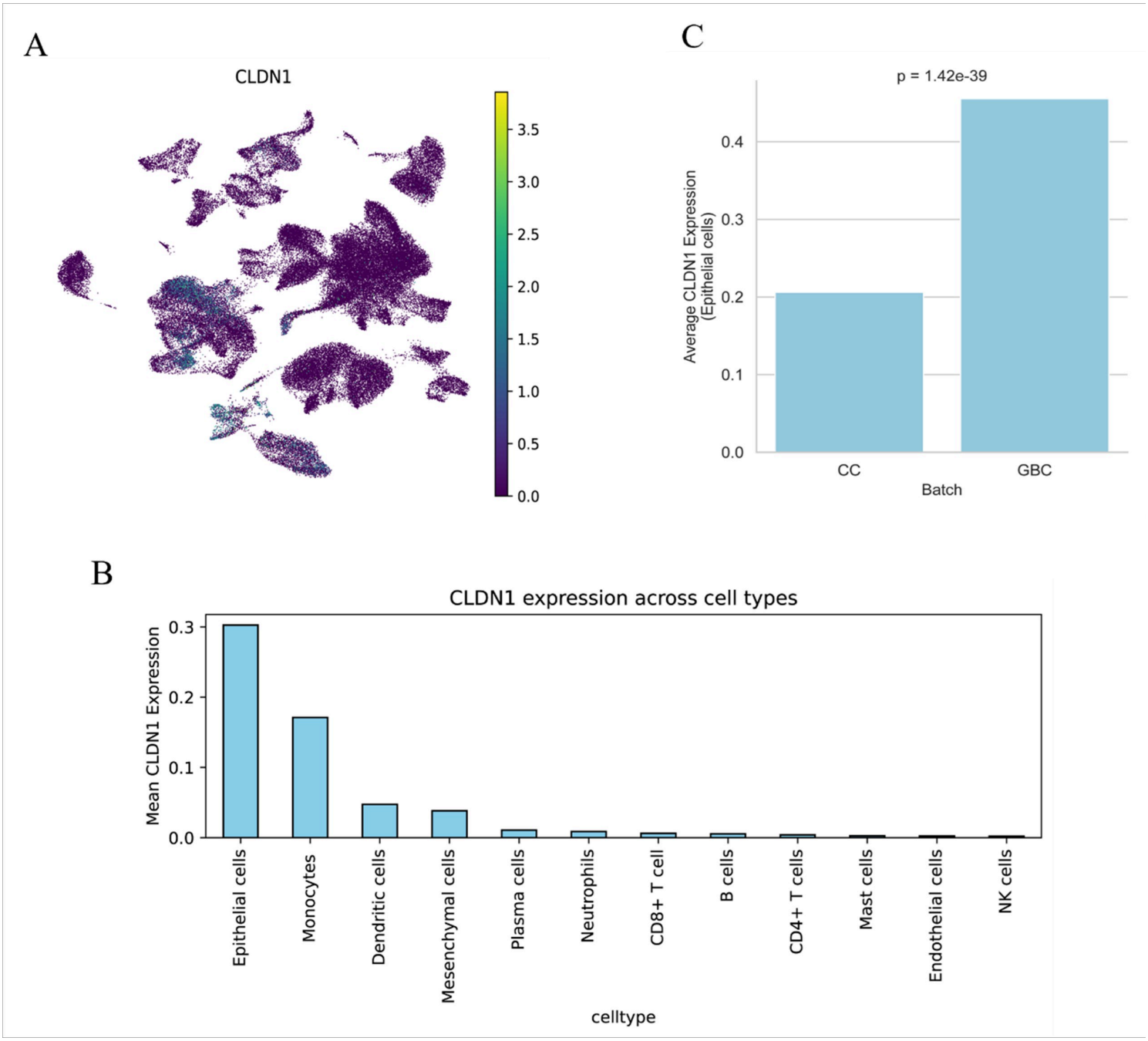


Figure 7

Expression validation of *CLDN1*. (A)The expression distribution of *CLDN1*. **(B)** *CLDN1* expression across cell types. **(C)** *CLDN1* expression in CC group and GBC group.