Journal of Advanced Research 74 (2025) 471–491

Contents lists available at ScienceDirect

Journal of Advanced Research

journal homepage: www.elsevier.com/locate/jare

Single-cell transcriptomic profiling uncovers cellular complexity and microenvironment in gastric tumorigenesis associated with *Helicobacter pylori*

Nianshuang Li [a,1], Sihai Chen [a,b,1], Xinbo Xu [a,1], Huan Wang [a,b], Pan Zheng [a], Xiao Fei [a], Huajing Ke [a], Yuting Lei [a], Yanan Zhou [a], Xiaoyu Yang [a], Yaobin Ouyang [a], Chuan Xie [a], Cong He [a], Yi Hu [a], Yi Cao [c], Zhengrong Li [c], Yong Xie [a], Zhongming Ge [d], Xu Shu [a], Nonghua Lu [a,*], Jianping Liu [a,*], Yin Zhu [a,*]

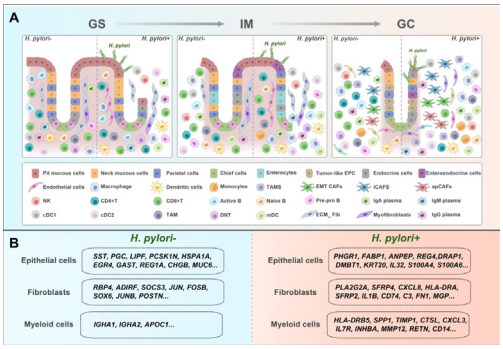
[a] *Department of Gastroenterology, Jiangxi Provincial Key Laboratory of Digestive Diseases, Jiangxi Clinical Research Center for Gastroenterology, Digestive Disease Hospital, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang, Jiangxi, China*
[b] *Postdoctoral Innovation Practice Base, The First Affiliated Hospital of Nanchang University, Nanchang, Jiangxi, China*
[c] *Department of General Surgery, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang, Jiangxi, China*
[d] *Division of Comparative Medicine, Massachusetts Institute of Technology, Cambridge, MA, USA*

HIGHLIGHTS

- ScRNA-seq data revealed cellular heterogeneity in gastric tumorigenesis.
- Enterocytes, was overwhelmingly abundant in gastric intestinal metaplasia lesion. HNF4G was predicted as the specific transcription factor.
- ScRNA-seq unveiled the transcriptional features of distinct cell subtypes in GC.
- The differentia expressed genes (DEGs) in epithelial, fibroblasts and myeolid cells associated with *H. pylori* infection have been identified.
- *H. pylori*-positive specimens exhibited enriched cell-cell communication, with significant active TNF signaling network.

GRAPHICAL ABSTRACT

Graphical summary of cellular heterogeneity and microenvironment along the consecutive gastric carcinogenesis associated with *H. pylori*. (A) Cellular heterogeneity along the histological cascade of gastric carcinogenesis with or without *H. pylori* infection. (B) The DEGs between *H. pylori* −positive and −negative cell types. GS, gastritis; IM, intestinal metaplasia; GC, gastric cancer.



ARTICLE INFO

Article history:
Received 29 June 2024
Revised 11 October 2024
Accepted 11 October 2024
Available online 15 October 2024

Keywords:
scRNA-seq
Cellular heterogeneity

ABSTRACT

Introduction: *Helicobacter pylori* (*H. pylori*) infection is the main risk for gastric cancer (GC). However, the cellular heterogeneity and underlying molecular mechanisms in *H. pylori*-driven gastric tumorigenesis are poorly understood.
Objective: Here, we generated a single-cell atlas of gastric tumorigenesis comprising 18 specimens of gastritis, gastric intestinal metaplasia (IM) and GC with or without *H. pylori* infection.
Methods: Single-cell RNA sequencing (scRNA-seq) was performed. Immunofluorescence, immunohistochemistry and qRT-PCR analysis were applied in a second human gastric tissues cohort for validation. Bioinformatics analyses of public TCGA and GEO datasets were applied.

* Corresponding authors at: Department of Gastroenterology, Jiangxi Provincial Key Laboratory of Digestive Diseases, Jiangxi Clinical Research Center for Gastroenterology, Digestive Disease Hospital, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang, Jiangxi, China.
E-mail addresses: lunonghua@ncu.edu.cn (N. Lu), ndyfy10307@ncu.edu.cn (J. Liu), ndyfy01977@ncu.edu.cn (Y. Zhu).
[1] These authors contributed equally to the study.

https://doi.org/10.1016/j.jare.2024.10.012

Microenvironment
H. pylori
Gastric carcinogenesis

Results: Single-cell RNA profile highlights cellular heterogeneity and alterations in tissue ecology throughout the progression of gastric carcinoma. Various cell lineages exhibited unique cancer-associated expression profiles, such as tumor-like epithelial cell subset (EPC), inflammatory cancer-associated fibroblasts (iCAFs) and Tumor-associated macrophage (TAM). Notably, we revealed that the specific epithelial subset enterocytes from the precancerous lesion GIM, exhibited elevated expression of genes related to lipid metabolism, and HNF4G was predicted as its specific transcription factor. Furthermore, we identified differentially expressed genes in *H. pylori*-positive and negative epithelial cells, fibroblasts and myeloid cells were identified. Futhermore, *H. pylori*-positive specimens exhibited enriched cell–cell communication, characterized by significantly active TNF, SPP1, and THY1 signaling networks.

Conclusions: Our study provides a comprehensive landscape of the gastric carcinogenesis ecosystem and novel insights into the molecular mechanisms of different cell types in *H. pylori*-induced GC.

Introduction

Gastric cancer (GC) is the fifth most common malignant tumor and the third leading cause of cancer-related deaths worldwide, particularly in East Asian countries [1]. The majority of GC cases are diagnosed at an advance stage [2]. *Helicobacter pylori* (*H. pylori*), a gram-negative bacterium that persistently colonize the gastric mucosa, is the widely-recognized known risk factor for GC. In fact, the World Health Organization (WHO) has classified *H. pylori* as a Group I carcinogen leading to gastric adenocarcinoma [3]. The classical Correa cascade of multistep gastric carcinogenesis generally involves a series of precancerous lesions, where chronic gastritis develops through metaplasia, dysplasia to ultimately intestinal type GC. *H. pylori* infection plays an initiating causative role in this process [4]. Although the benefit of *H. pylori* eradication in reducing GC risk has been reported by extensively clinical evidence [5], the issues become complicated if precancerous lesions have appeared. A thorough exploration of gastric tumorigenesis associated with *H. pylori* would help in understanding the molecular mechanisms and discover more effective strategies for GC diagnosis and treatment.

Multiple omics studies on genomics, transcriptomics and proteomics elaborated various molecular landscapes underlying the progression of GC [6,7]. However, GC is a complex ecosystem, defined by the heterogeneity and interactions between diverse cell populations, including epithelial cells, stromal cells and immune cells. Single-cell RNA sequencing (ScRNA-seq), a powerful tool to assess gene expression at the individual cell level, has been recently used to explore gastric carcinoma microenvironment with particular focus on the transcriptional patterns in epithelial and fibroblast cells [8–10].

However, cellular diversity and cell–cell communications in the context of gastric tissue microenvironments during the Correa cascade of multistep gastric carcinogenesis remains poorly understood. Given the causal determinant role of *H. pylori* in gastric inflammation and disease progression, extensive evidence focuses on the molecular biological and immune regulation of gastric epithelial cells and myeloid cells. *H. pylori* infection has been indicated to endow epithelial cells with multiple phenotypes, such as sustained inflammation, hyperproliferation, and invasiveness [11]. Additionally, adhesion of *H. pylori* triggers double-strand breaks (DSBs) and elicits genome instability in epithelial cells [12]. The transcriptional profiles of other cell types such as fibroblasts and intercellular communication are still lacking.

In this work, we dissected cellular diversity and microenvironment heterogeneity by analyzing 91,394 single-cell transcriptomes from 18 samples of non-atrophic gastritis (GS), intestinal metaplasia (IM), and GC that were either *H. pylori* positive or negative. Through our analysis, we identified diverse populations of epithelial cells, stromal cells and immune cells within the microenvironment as the disease progress from GS to IM, culminating in GC development. We unveiled the transcriptomic signatures of distinct cell subtypes, shedding light on the cellular heterogeneity caused by *H. pylori* infection. This included cellular composition, differentially expressed genes, and cell interactions. Our data provides valuable information on the molecular characteristics and potential intercellular signals controlling cell lineages states, offering a comprehensive understanding of the pathogenic mechanisms orchestrated by *H. pylori*. These findings pave the way for the development of new strategies for precise prevention and therapeutics of gastric carcinoma.

Materials and methods

Ethics statement

This work was approved by the Ethics Committee of the First Affiliated Hospital of Nanchang University under protocol number of (2023) CDYFYYLK (01–009). All tissue samples were obtained under informed consent.

Human tissue samples

A total of 18 patients pathologically diagnosed with gastric lesions at The First Affiliated Hospital of Nanchang University, including 6 of GS, 6 of IM, and 6 of GC tissues, were enrolled in this study (Table S1). The specimens of GS and IM were obtained during endoscopy, and the GC specimens were obtained from surgical samples. The patients with gastric mucosal lesions were then divided into two groups according to *H. pylori* status. The sections with H&E staining were reviewed and confirm disease diagnosis by senior pathologists. Status of *H. pylori* infection for these clinical specimens was confirmed by ^{13}C-urea breath test (UBT) and immunohistochemistry staining. *H. pylori* negative patients who have not undergone eradication therapy, were diagnosed using UBT and serological tests. The primary GC specimens were collected from patients who required surgery, without adjuvant therapy. Patients with diabetes, chronic cholesterol disease, liver and kidney damage, gastrointestinal surgery history, systemic infections, or other malignancies were excluded from the study.

In validation cohort 1 (n = 90), the formalin-fixed paraffin-embedded (FFPE) specimens including 30 cases each of GS, IM and 30 cases of GC were obtained from the Department of Pathology at the First Affiliated Hospital of Nanchang University. Validation cohort 2 (n = 84) of fresh gastric biopsy tissues, with 30 cases of GS, 30 cases of IM, and 24 cases of GC, were collected during endoscopic procedures and surgical resections, respectively.

Tissue dissociation and single-cell suspension preparation

The tissue samples were removed from patients and quickly immersed in Dulbecco's modified eagle medium (DMEM; Cat#11995065, Gibco) containing 10 % fetal bovine serum (FBS; Cat#10091148, Gibco). Samples were minced into about 0.5 mm^3 pieces and enzymatically digested in DMEM supplemented with trypsin (Cat#15090046, Gibco), collagenase I (Cat# 17018029, Gibco), collagenase II (Cat# 17101015, Gibco) and collagenase IV (Cat# 17104019, Gibco). The samples were incubated and digested at 37 °C for approximately 60 min and then filtered using a 100 μm strainer (Corning). Following centrifugation (300 g) at 4 °C for 5 min, the cells were resuspended with medium containing red cell lysate (Cat#130094183, MACS). After standing 10 min at 4 °C, the single-cell suspension was centrifuged at 300 g for 5 min and resuspended in 100 μl medium. Cell concentration and viability were detected under Luna Cell Counter.

Single cell droplet-based RNA sequencing

The single-cell suspensions were adjusted to 700–1200 cell/μl. The suspensions were then converted to separate barcoded scRNA-seq libraries using the Chromium Next GEM Single Cell 3′ Reagent Kits v3.1 (Cat#1000268, 10 × genomics) according to the manufacturer's instruction. Libraries were sequenced on an Illumina Nova 6000 PE150.

Single-cell RNA-seq data processing

The raw scRNA-seq reads were processed, including demultiplex cellular barcodes, read alignment, and generation of gene expression matrices using Cell Ranger (10 × genomics). The Seurat software package was used for further quality control. To remove low quality cells and likely multiple captures, the following quality control steps: the cells with gene numbers less than 200, UMI less than 1000 and log10Genes Per UMI less than 0.7 were filtered out. Furthermore, the cells where > 20 % of the counts belonged to mitochondrial genes and 5 % of the counts belonged to hemoglobin genes were removed. Additionally, the DoubletFinder package (version 2.0.2) was used to identify potential doublet. To achieve higher quality clustering, we removed likely cell doublets and multiplets. Finally, 124,528 cells (Median: 5319 cells/patient) were obtained for the downstream analysis. We obtained 1717 genes and 5763 UMIs per cell on average. The detailed sequencing data of quality control were shown in Table S1.

Cell-type clustering and marker identification

The FindVariableGenes function from Seurat R package was used to screen highly variable genes (HVGs). Then the principal component analysis (PCA) was performed based on HVGs, followed by visualization using Uniform Manifold Approximation and Projection (UMAP).

To detect cluster-specific expressed genes, the FindAllMarkers from Seurat R package was used to compare the differential genes of a given cluster with all other clusters. The identified marker genes were visualized through VlnPlot and FeaturePlot functions.

Cell type identification

The SingleR package was utilized to assign cell types based on a public reference dataset. This involved correlating the expression profiles of unidentified cells with those in the reference dataset, assigning the cell type from the refence dataset that exhibited the highest correlation to the unidentified cell type. For the identification of gastric epithelial cell subtypes, the specific embryonic development patterns of gastric epithelial cells were also considered.

CNV analysis

To investigate large-scale chromosomal CNV, the ScRNA-seq data obtained from these tissues were processed using InferCNV, a robust computation tool specifically sesigned to infer CNV numbers within individual cells based on intercellular variability in gene expression patterns. In this study, we employed InferCNV to explore CNV numbers in epithelial cells subsets with *H. pylori* infected or uninfected gastric tissues.

Pathway enrichment analysis

To compare the differences in biological states and pathways between cell subtypes, GO, KEGG, Gene set Enrichment analysis (GSEA) or Gene set variation analysis (GSVA) were performed. GO analysis was used to provide insights into molecular function, biological process, and cellular component of the gene sets in specific cell subpopulations. KEGG analysis was performed to identify the gene sets-enriched biological pathways. By utilizing predefined gene sets, such as metabolic pathways and signaling pathways, GSEA analysis was performed to identify whether the differential gene expression between two different cell subtypes was significantly enriched in these gene sets. GSVA analysis was performed to reveal functional differences between different cell lineages.

Trajectory analysis

To map differentiation within the gastric carcinogenesis microenvironment, pseudotime analysis was performed with Monocle 2 or 3 to determine the dynamic differentiation trajectories of cell subpopulations and gene expression changes at different gastric lesions. Utilizing the Monocle R package, genes exhibiting significant expression variation across cells were firstly identified, and their expression profiles were used for dimensionality reduction. Subsequently, a minimum spanning tree (MST) was constructed, representing the differentiation trajectories. Compared to Monocle 2, the Monocle 3 software package incorporates novel data procession methods, such as Destiny, and trajectory analysis algorithm DDRTree.

SCENIC analysis

Gene regulatory networks were constructed by SCENIC analysis. The analysis of transcriptional factor binding motifs was performed using RcisTarget R package. Grnboost was used to infer the co-expression network. Finally, the co-expressed modules between transcriptional factors and potential target genes, as well as the regulon activity score (RAS) of single cell. The regulon specificity score (RSS) inferred the specific correspondence of regulon to each cell. Connection specificity index (CSI) represented the association between different regulons.

Intercellular interaction analysis

To analyze cell–cell interaction, we applied the CellChatDB R package to identify significant ligand-receptor pairs between different cell types within all gastric tissue samples. We also compared the cell–cell interactions between *H. pylori* −positive and *H. pylori* −negative samples.

Establishment of human gastric organoids

To establish human gastric organoids, fresh tissue samples were acquired from gastric adjacent normal tissues in patients with

gastric adenocarcinoma who underwent curative gastrectomy at the First Affiliated Hospital of Nanchang University (Institutional Review Board No. (2023) CDYFYYLK (01–009)). The protocol for construction gastric organoid were provided from the manufacturer (BioGenous, China). In brief, the glands were dissociated from approximately 2 cm^2 strips of human gastric adjacent noncancerous tissues using EDTA buffer and digestion solution at 37 °C for 30 min. After cell counting, the cells were mixed with ice-cold Matrigel and plated in 48-well plate, followed by adding Human gastric epithelial organoid basal medium with growth factors (BioGenous, China). Organoids formed significant three-dimensional spherical structures after about a week of culture.

Infection of cells with H. Pylori strains

The monocytes THP-1 cells and gastric organoids were used in this study. THP-1 cells were cultured in RPMI-1640 medium supplemented with 10 % FBS (Gibco, Australia) and 1 % P/S Before experimentation, the THP-1 cells were differentiated into macrophages by phorbol 12-myristate-13-acetate (PMA, HY-18739, MedChemExpress, Shanghai, China). This study employed $CagA^+$ *H. pylori* strain PMSS1 (pre-mouse Sydney strain 1), as described previously [13]. Briefly, the *H. pylori* strains were cultured on Brucella agar (BD Bioscience) supplemented with 5 % sheep blood under microaerophilic conditions for in vitro massage. *H. pylori* PMSS1 strain were used for in vitro co-culture with THP1 at a multiplicity of infection of 100:1. For infection of organoids, *H. pylori* PMSS1 strain was injected into Matrigel for 24 h or 48 h.

GEO and TCGA database analysis

To validate gene expression levels in bulk RNA-seq, TCGA-STAD (The Cancer Genome Atlas-Stomach carcinoma) dataset from the National Cancer Institute, and a series of RNA-seq datasets of GC and paired adjacent normal tissues (GSE122401, GSE65801 and GSE13195) from GEO database. Firstly, RNA-seq data was retrieved from public database and performed quality control and data cleaning to ensure the reliability of the results. The ggplot2 R package was then used to generate box plots illustrating the gene expression patterns in cancer and adjacent tissues.

Immunohistochemical staining

Immunohistochemistry analysis was performed to assess protein expression in human gastric tissues, as described previously [13]. All tissues were collected from the patients with 30 of GS, 30 of IM, and 30 of GC at The First Affiliated Hospital of Nanchang University. Each lesion stage was divided into half of the positive and negative specimens of *H. pylori*. Briefly, tissue sections were deparaffinized in a 70 °C oven for 1.5 h, followed by rehydration using a graded ethanol series. Antigen retrieval was performed by boiling the sections in citrate buffer in a microwave for 15 min. Endogenous peroxidase activity was quenched by the incubation in 3 % H_2O_2 for 10 min. The sections were then incubated with the primary antibodies overnight at 4 °C. For a full list of primary and secondary antibodies used for immunohistochemistry, see Table S2. Subsequently, tissue sections were incubated with the secondary antibody and visualized using a DAB staining kit according to the manufactures' instructions (Zhongshan Biotech, Beijing, China). The quantitative analysis of immunohistochemistry staining was performed as described previously [13]. In brief, the immunoreactivity of targets was evaluated and scored for intensity (scaled 0–3) and frequency (scaled 0–4) by two trained pathologists blinded to sample identity. For statistical analysis, the expression levels of these proteins were displayed by a score in range from 0 to 12 using formula intensity × frequency. All images were acquired under a microscope (Nikon Eclipse).

Flow cytometry

The fresh human tissue specimens were cut into 1–2 cm pieces, which were then processed into single cell suspension by digestive enzyme (2 mg/ml collagenase IV and 0.5 mg/ml DNase I included in 1 mM EDTA) at 37 °C for 30 min, followed by filtration through a 70 mm filter. The Zombie Aqua Fixable Viability Kit (Cat#423102, Biolegend) was used for detection and gating out of dead cells. Then cells were resuspended with staining media containing FITC-CD45 (Cat#11–0459-42, ThermoFisher) and PE-CD68 (Cat#333808, Biolegend) antibodies. After centrifugation and washing with incubation buffer, the cells were resuspended in PBS and analyzed in a flow cytometer. Multiparameter FACs data were acquired on the NovoCyte D3000 instrument (Agilent). After gating of lymphocytes according to FSC and SSC, $CD68^+$-macrophage (FITC/PE +) was identified. Data were analyzed with FlowJo software.

Immunofluorescence staining

The tissue sections underwent deparaffinization and rehydration steps, followed by antigen retrieval in citrate buffer (pH 6.0). Then tissue sections were permeabilized with 0.3 % Triton X-100 in PBS for 15 min, followed by blocking with 3 % BSA in PBS for 1 h. The primary antibodies were then applied overnight at 4 °C, see Table S2. For immunofluorescence staining, the tissue slides were washed in PBS and incubated with Alexa-flour secondary antibodies (Invitrogen) at room temperature for 60 min, then mounted with an appropriate medium containing 4′,6-diamidino-2-phenylindole (DAPI) nuclear counterstain. The mounted sections were then imaged using a confocal microscope (LEICA Stellaris 5).

qRT-PCR analysis

Gastric biopsy tissues containing 30 of GS, 30 of IM and 24 of GC were derived from the remaining specimens from the previous study [14]. Each gastric lesion stage was divided into infected and uninfected groups based on *H. pylori* infection status. The qRT-PCR analysis was performed to detect the mRNA levels of target genes, as described previously [13]. Briefly, total RNA was extracted from the tissue samples using Trizol reagent (TIANGEN Biotech, Beijing, China). Then RNA was converted into cDNA using a reverse transcription kit according to the manufacturer's instructions. The qPCR assay was performed with a QuantStudio 5 Real-time PCR system (Life Technologies). The primers for qRT-PCR analysis used in this study were listed in Table S3.

Statistical analysis

All statistical analyses and graph were performed using SPSS software (V 21.0) and graphpad prism 9. The Wilcoxon rank-sum test was used to compare continuous variables between two independent samples. The *t*-test was used to evaluate association with parametric variables between two independent samples. The results were considered statistically significant at $p < 0.05$ (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

Results

ScRNA-seq landscape along the consecutive gastric carcinogenesis associated with H. pylori

To explore the cellular heterogeneity and microenvironment in gastric tumorigenesis associated with *H. pylori* infection, we collected 18 gastric specimens representing the progressive histologic stages, i.e., 6 cases of GS, 6 cases of IM, and 6 cases of GC specimens (Fig. S1A). Each stage of the lesion was divided into two groups based on *H. pylori* infection status. Detailed patient information, including age, sex, pathological information, *H. pylori* infection status, was provided in Table S1. Tissue samples were dissociated into single-cell suspensions and processed using the 10 × Genomic Chromium workflow (Fig. 1A).

Following data quality control and filtering processes, a total of 91,394 cells were used for further analysis (2563–7677 cells; median 5077 cells/patient) (Table S1). The resulting 26 cell clusters

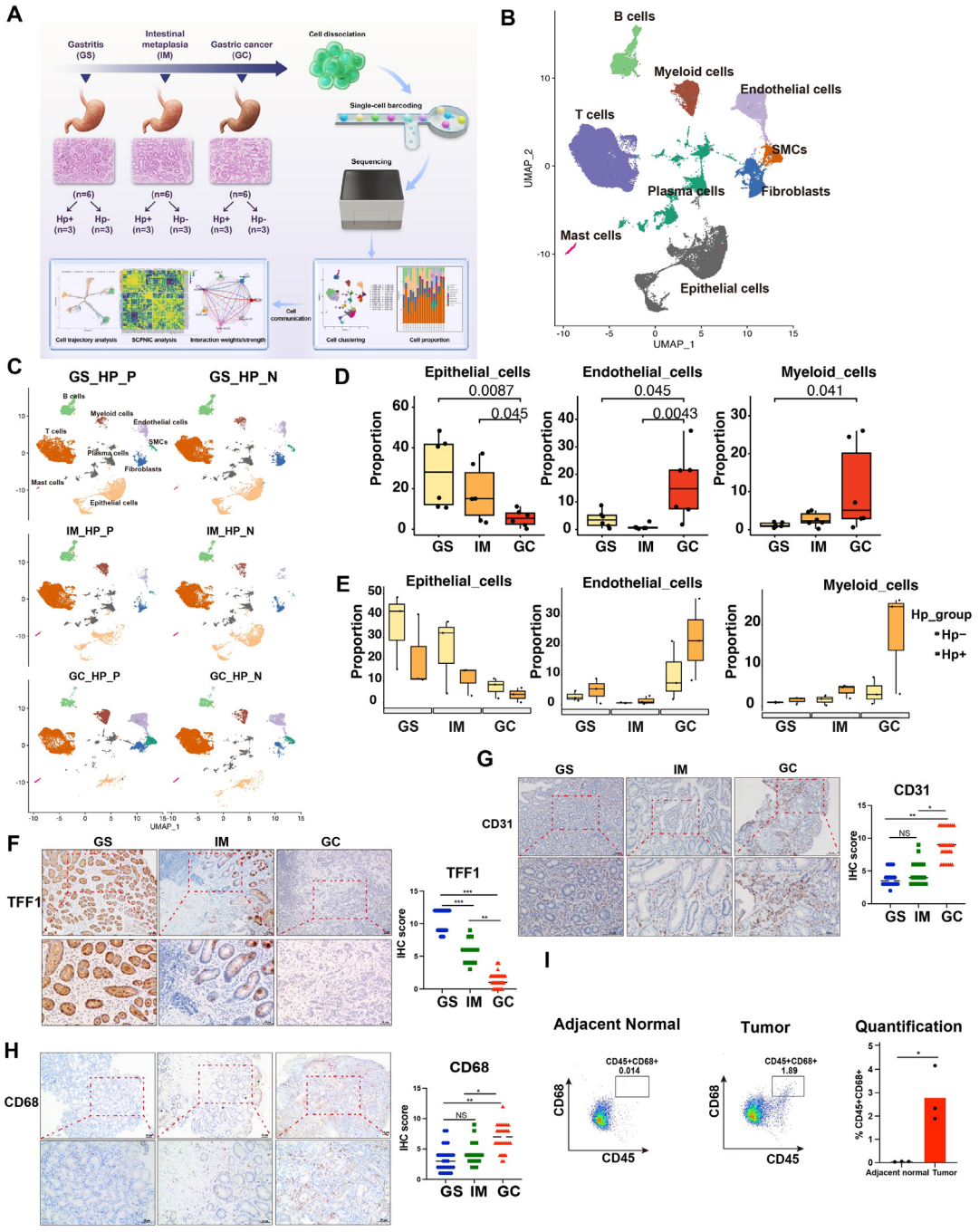


Fig. 1. ScRNA-seq profiles identified increased myeloid cell proportions, as well as decreased epithelial cell proportion in gastric carcinogenic cascade. (A) Schematic representation of the experimental design of this study. Human gastric tissues of 18 patients with GS, IM and GC were harvested for scRNA-seq using 10 × genomic platform. (B) UMAP plot showing the annotation and color codes for cell types during gastric carcinogenesis cascade. (C) The UMAP plot displaying the cell types in *H. pylori*-positive and −negative GS, IM and GC tissues. (D) The box plot indicating the proportions of epithelial cells, endothelial cells (ECs), and myeloid cells during the progression from GS, IM to GC. (E) The box plot showing the fraction of epithelial cells, endothelial cells (ECs), and myeloid cells in *H. pylori* −positive and −negative gastric tissues. (F-H) Immunohistochemical staining showing the expression of TFF1 (F), CD31 (G), CD68 (H) in human gastric tissues, respectively. Scale bar, 20 μm. (I) Flow cytometry analysis for $CD45^{+}CD68^{+}$ cells in human gastric carcinoma and adjacent normal tissues. *NS, no significance*; $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

were annotated as 9 cell types based on established marker genes [15,16] (Fig. 1B, Fig. S1B). Non-immune cells included epithelial cells (*PECAM*, *KRT19*), endothelial cells (ECs; *PECAM1*, *CDH5*), fibroblasts (*COL1A1*, *PDGFRA*), and smooth muscle cells (SMCs; *TAGLN*, *ACTA2*); immune cells included T cells (*CD3D*, *CD3G*), B cells (*CD79A*), myeloid cells (*CD68*, *CSF1R*), plasma cells (*MZB1*, *IGKC*), and mast cells (*TPSAB1*, *TPSB2*) (Fig. S1C, D; Table S4). Notably, epithelial cells expressed high levels of gastric mucosa-specific genes GAST and TFF1 (Fig. S1D), as expected, indicating the accuracy of our data.

Gradually decreased proportion of epithelial cells, accompanied by increased percentages of myeloid cells during gastric carcinogenic cascade

In the gastric specimens, all identified cell types were present regardless of *H. pylori* infection, with epithelial cells and T cells predominant (Fig. 1C). When comparing the cellular compositions across the three progressively pathological stages (GS, IM and GC), a gradual decrease in epithelial cell proportion was noted, accompanied by a corresponding increase in myeloid cells as the disease advanced towards cancer. The ECs were greatly expanded in cellular proportion and significantly increased in GC tissues, compared to GS and IM tissues (Fig. 1D). Furthermore, *H. pylori*-infected samples contained relatively lower proportion of epithelial cells compared to uninfected tissues, while the trend was opposite for ECs and myeloid cells (Fig. 1E). Subsequent immunohistochemical staining for cell type markers during the cascade of gastric carcinogenesis supported these findings. Specifically, downregulation of the gastric epithelial cell-specific marker TFF1 (Fig. 1F), and upregulation of marker genes CD31 for ECs and CD68 for myeloid cells, were observed in GC tissues, and, compared to GS and IM specimens (Fig. 1G, H). However, there seemed to be no trend in epithelial cell marker EPCAM expression in different gastric diseases (Fig. S1E). Flow cytometry analysis confirmed a notable increase in myeloid cells within GC tissues compared to adjacent normal tissues (Fig. 1I), with the gating strategy shown in Fig. S1F. In line with these findings, the upregulation of *CD68* and loss of *TFF1* in GC were corroborated by the data from the gastric carcinoma cohort from public TCGA (Fig. S1G), and GEO database (GSE122401, GSE65801 and GSE13195) (Fig. S1H, I). These findings from our local cohort and multiple datasets strengthen the evidence indicating the significant alterations in cellular compositions during gastric tumorigenesis, emphasizing the dynamic changes that occur in the tumor microenvironment.

Distinct cell lineage and transcriptional patterns of epithelial cells during multistep gastric carcinogenesis

To explore lineage states and transcriptional changes in gastric tumorigenesis, we next performed unsupervised dimensionality reduction of epithelial cells. A total of 19 clusters emerged, including pit mucous cells (*MUC5AC*, *TFF1*), neck mucous cells (*MUC6*), chief cells (*PGA3*, *PGA4*), parietal cells (*ATP4A*, *ATP4B*), Endocrine cells (*GAST*, *SST*), enteroendocrine cells (*CHGA* and *TPH1*), enterocytes (*FABP1* and *PHGR1)*, and tumor-like epithelial cells (tumor-like EPC, *CEACAM5*) (Fig. 2A, D; Fig S2A, B) [17,18]. Notably, endocrine and enteroendocrine cells exhibited clearly different states from other cell lineages (Fig. 2A). Of particular interest was cluster 10 defined as enterocytes, based on the well-known markers *FABP1* and *PHGR1* (Fig. S2C). Additionally, cluster 7 designated as tumor-like EPC, expressing cancer embryonic antigen *CEACAM5*, were enriched in GC tissues (Fig. 2C, D; Fig. S2C). Given that cancer cells often occur multiple mutations and chromosomal alterations, we further utilized copy number variants (CNV) analysis to classify epithelial cells as either non-malignant or malignant cells. As expected, tumor-like EPC displayed higher CNV levels compared to other cell types (Fig. S2D). When comparing the CNV scores of epithelial cells across all samples, we observed a significant elevation in GC tissues, particularly within *H. pylori* −positive subgroups, compared to GS or IM tissues (Fig. 2B).

With the progression of gastric neoplasm, the fractions of endocrine cells gradually decreased (Fig. 2C), accompanied by a reduction in the expression of *GAST* and *SST* (Fig. 2E). Additionally, a loss of *MUC5AC* and *TFF1*, typically expressed in pit mucous cells, was observed in GC tissues (Fig. 2E). Of note, the enterocytes were almost predominantly originated from IM tissues, while the tumor-like EPC showed a significant increase in GC tissues (Fig. 2C). Further validation through multicolor immunostaining for EPCAM and CEACAM5 revealed the abundance of tumor-like EPC population in GC tissues (Fig. 2F). The trajectory analysis suggested that tumor-like EPCs may stem from pit mucous cells (Fig. 2G, Fig. S2E). This evidence was further supported by the co-localization of MUC5AC (pit mucous cell marker) with CEACAM5 (tumor-like EPC marker) in GC samples (Fig. 2H). The GSEA analysis revealed that epithelial mesenchymal transition (EMT), inflammatory NF-κB signaling, interferon γ response, and hypoxia were the highly active pathways in the tumor-like EPC populations (Fig. 2I). Specifically, within the tumor-like EPC cluster, there was notable overexpression of genes involved in the EMT pathway, which links to the intracellular cytoskeleton and extracellular matrix (Fig. 2J, Fig. S2F). Notably, we noticed that *KRT17* gene was extremely enriched in tumor-like EPC, suggesting its potential cancer-promoting role. This observation was substantiated by the confirmation of *KRT17* overexpression in GC tissues compared with matched normal specimens using GEO datasets and TCGA cohort analyses (Fig. 2K). This was further supported by immunostaining showing elevated protein levels of KRT17 in GC tissues, compared to non-cancerous tissues (Fig. 2M, N; Fig. S2G). Taken together, these data revealed the diversity and distinct cell states of epithelial cells in the cascade of gastric carcinogenesis and underscore the significance of the oncogenic transcriptomic features (including KRT17), functional characteristics, and cell origin of tumor-like EPC subpopulation that were abundant in GC tissues.

A distinctive epithelial cell population enterocyte from precancerous lesion gastrointestinal metaplasia

Given that the enterocyte cell populations were overwhelmingly enriched in gastrointestinal tissues (Fig. 2C), we focused on characterizing the transcriptional patterns and enriched pathways of this distinct cell cluster. The unique presence of *CDX1* and *CDX2*, members of the CDX family known to confer gastrointestinal metaplasia phenotypes [19], was detected in enterocyte cells (Fig. S3A), reinforcing the specificity of enterocyte subclusters in gastrointestinal metaplasia, a crucial precursor lesion of GC. Spasmolytic polypeptide-expressing metaplasia (SPEM) is a regenerative pathological state of gastric corpus mucosa in response to injury, associated with loss of chief and parietal cells, and hyperplasia of neck mucous cells [20]. We observed a gradual decline in SPEM cells, defined by the co-expression of *MUC6* and *TFF2*, from GS to IM, and ultimately GC stages (Fig. S3B), aligning with the notion that SPEM represents as a precursor lesion to metaplasia [17,18]. Functional analysis of enterocyte cells through KEGG enrichment based on the DEGs of this subcluster highlighted the relation to glycolysis and lipid metabolism, such as fat digestion and absorption, PPAR signaling pathway, and cholesterol metabolism (Fig. 3A). The dot plot showed that the markers regulating lipid metabolism (*FABP1*, *APOC3*, *ANPEP*, *APOA1*, *APOA4*, *APOC3* and *ALDOB*) and those regulating glycolysis (*PCK1*, and *ADH4*) were specifically overexpressed in enterocytes (Fig. 3B), with significantly higher expression levels in IM tissues compared to GS and GC tissues (Fig. 3C). Immunofluorescence and immunohistochemical staining validated the

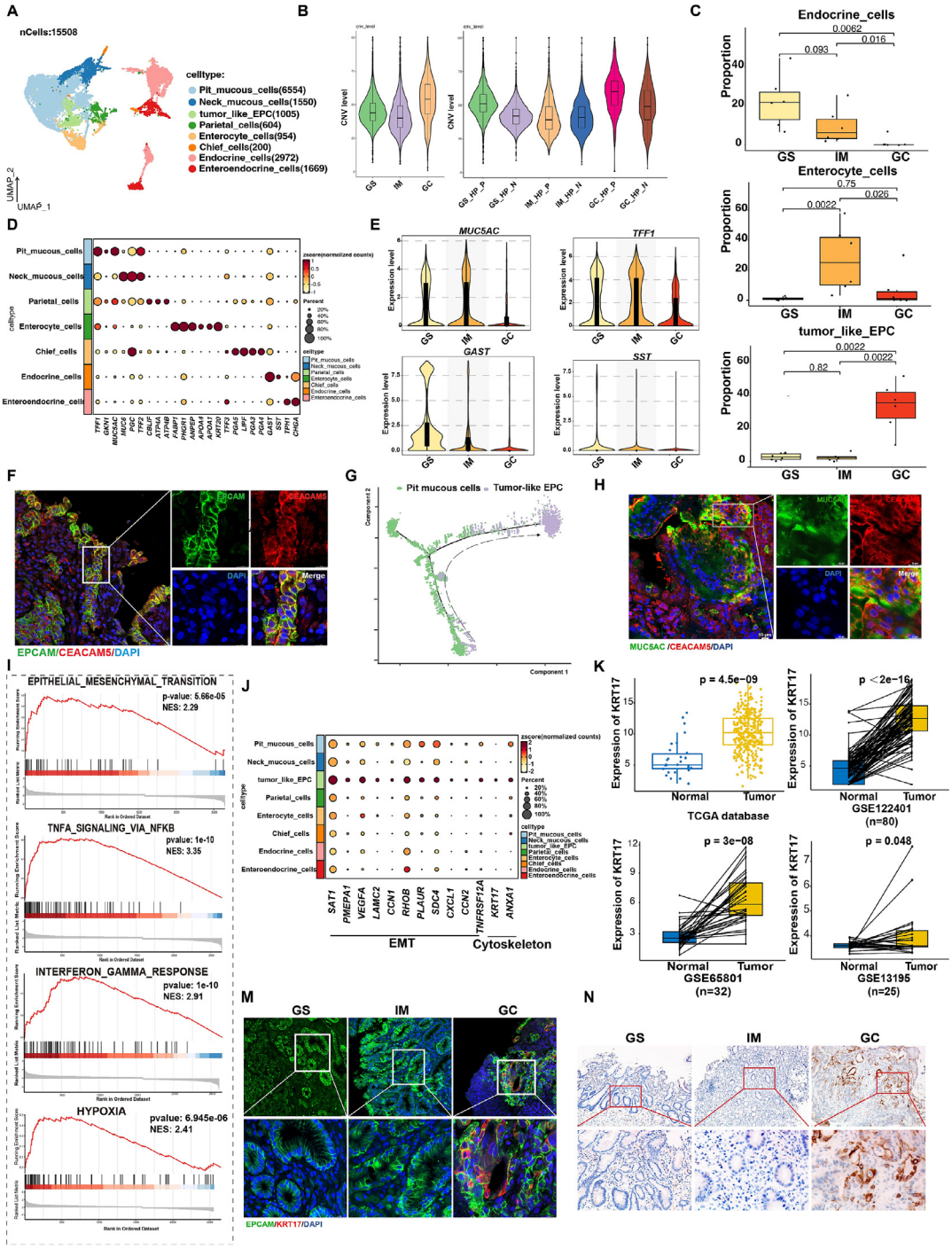


Fig. 2. Distinct cell subtypes and transcriptional patterns of epithelial cells during multistep gastric carcinogenesis. (A) UMAP plot showing the major clusters of epithelial cells. All epithelial cells were color-coded by condition. (B) InferCNV assessing the genomic variation and malignancy of epithelial cells during the progress of gastric neoplasm. (C) The box plot indicating the proportions of epithelial subtypes, including endocrine cells, enterocytes, and tumor-like EPC in the progression from GS, IM to GC lesions. (D) Dot plot illustrating the expression of markers in gastric epithelial cell subpopulations. (E) Violin plot showing the expression of *MUC5AC*, *TFF1*, *GAST*, and *SST* in GS, IM and GC tissues. (F) Combined immunofluorescence staining for EPCAM (green) and CEACAM5 (red) validating the presence of epithelial-derived tumor cells in GC tissues. Scale bar, 10 μm. (G) Trajectories analysis of pit mucous cells and tumor-like EPC using Monocle 2. (H) Representative images of immunostaining indicating the MUC5AC^{+}CEACAM5^{+} cells in GC tissues. Scale bar, 10 μm. (I) GSVA analysis indicating the enriched pathways in tumor-like EPC populations. (J) Dot plot showing the expression of EMT and intracellular cytoskeleton-related genes in gastric epithelial subclusters. (K) The expression of *KRT17* gene in bulk RNA-seq from TCGA-STAD cohort and GEO database (GSE122401, GSE65801, GSE13195), gastric carcinoma and normal specimens. (M, N) Representative images of immunofluorescence and immunohistochemical staining for detecting the cellular localization and expression of KRT17 in GS, IM and GC tissues. Scale bar, 20 μm. NS, no significance; $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

predominant expression of FABP1 in epithelial cells, particularly in gastric tissues with preneoplastic IM lesions by (Fig. 3D, E; Fig. S3C). Notably, the enterocytes exhibited elevated expression of *KRT20*, a well-known marker of intestinal differentiation (Fig. 3F, Fig S3D), which was confirmed by immunostaining showing the overexpression of KRT20 in IM tissues (Fig. 3G, H).

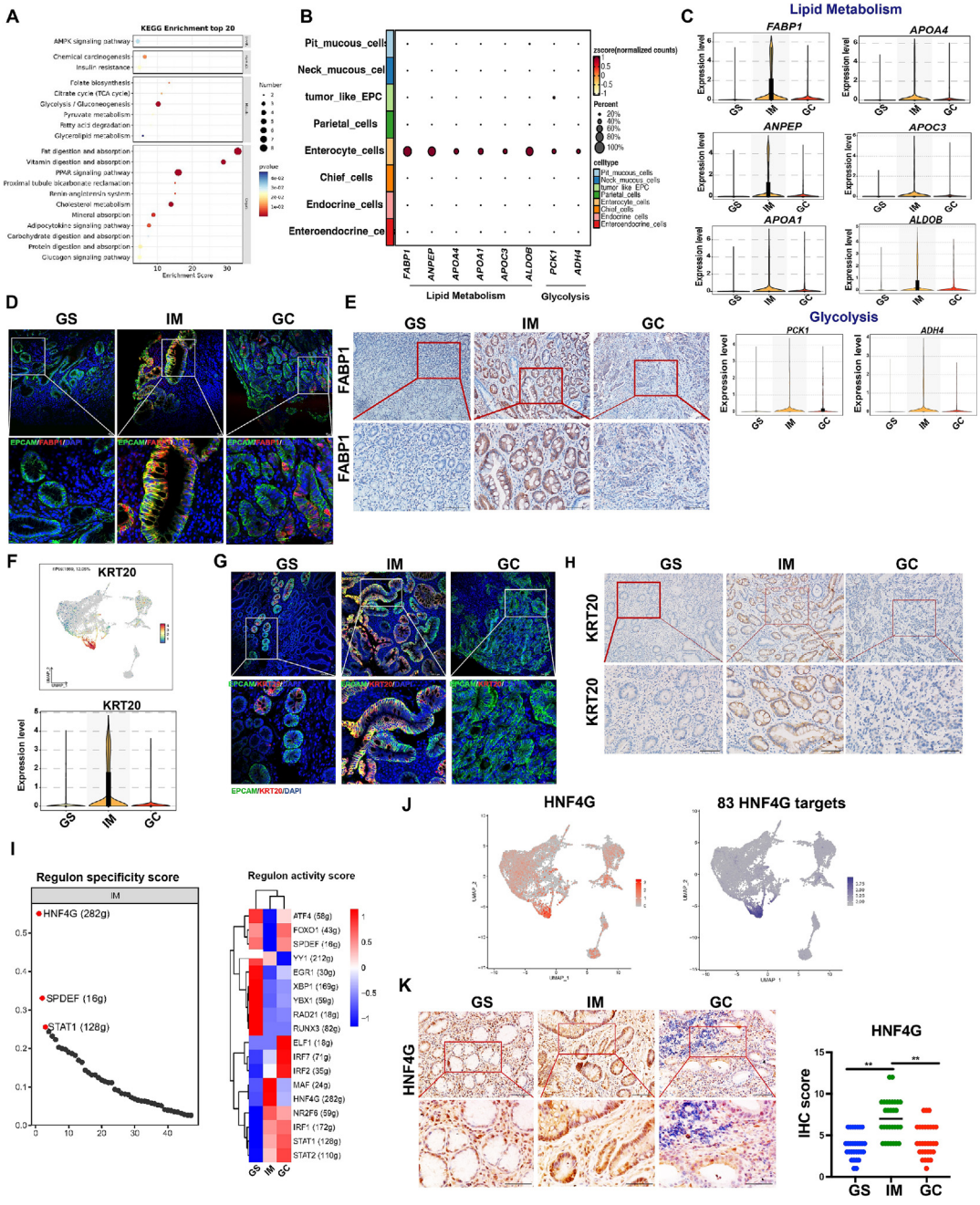


Fig. 3. The molecular features of enterocyte, a distinct epithelial subset in gastric IM. (A) KEGG analysis investigating enriched functional processes and pathways of enterocyte cells. (B) Dot plot showing the expression of DEGs involved in lipid metabolism and glycolysis in gastric epithelial cell subsets. (C) Violin plot showing their expression in GS, IM and GC specimens. (D) Immunostaining verifying the colocalization of EPCAM and FABP1 in GS, IM and GC tissue sections. Scale bar, 20 μM. (E) Immunohistochemical staining for FABP1 expression in human gastric tissues. Scale bar, 20 μM. (F) UMAP plot (upper panel) and violin plot showing the expression of KRT20 from gastric epithelial cells in gastric carcinogenesis. (G-H) Immunofluorescence (G) and immunohistochemistry staining (H) detecting the expression of KRT20 in epithelial cells in GS, IM and GC specimens. (I) SCENIC analysis predicting the potential transcriptional factors targeting enterocyte cells. (J) UMAP plot showing the expression of transcriptional factor HNF4G (left panel) and its target genes (right panel) in epithelial cells. (K) Immunohistochemistry analysis of HNF4G in GS, IM and GC tissues. Scale bar, 20 μM. Representative images (left panel), Quantification analysis (right panel). *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

Afterwards, we applied SCENIC to infer TF-target regulatory network (regulons) of enterocytes. Notably, the transcription factor *HNF4G* exhibited specific and heightened transcriptional prosperities within enterocytes (Fig. 3I), aligning with previous studies highlighting its crucial role of *HNF4G* in regulation of fatty acid oxidation and intestinal stem cells renewal [21]. Of note, *HNF4G*, together with its 83 downstream target genes, such as *FABP1*, *FABP2*, *APOB*, and *VIL*, demonstrated specific expression patterns within enterocytes (Fig. 3J; Table S5). Immunohistochemistry analysis verified the increased levels of HNF4G in IM samples when compared with GS or GC samples (Fig. 3K). In summary, these findings uncovered enterocytes as the distinct epithelial sublineages in premalignant intestinal metaplastic lesion and HNF4G as its specific transcriptional regulator.

Differentially expressed genes associated with H. pylori infection in epithelial cells

H. pylori exhibited a predilection for colonizing the pit cells within gastric epithelial mucous layer, as evidenced by

immunostaining showing co-localization of MUC5AC and *H. pylori* (Fig. 4A), consistent with previous studies [22]. Given that chronic infection with *H. pylori* causes gastric mucosal injury and initiates neoplastic transformation, we aimed to identify the DEGs associated with *H. pylori* in epithelial cells. By comparing the cell number of epithelial subsets, endocrine cells expressing *GAST* and *SST* were reduced in *H. pylori*−positive samples. Conversely, as previously observed enrichment in GC, enterocyte and tumor-like EPC subclusters were increased in *H. pylori*-positive tissues. These findings suggested a significant impact of *H. pylori* infection on the regulation of various cell types (Fig. S3E).

The differential gene expression analysis revealed that epithelial cells from *H. pylori* −positive specimens exhibited decreased expression of genes involved associated with gastric acid and

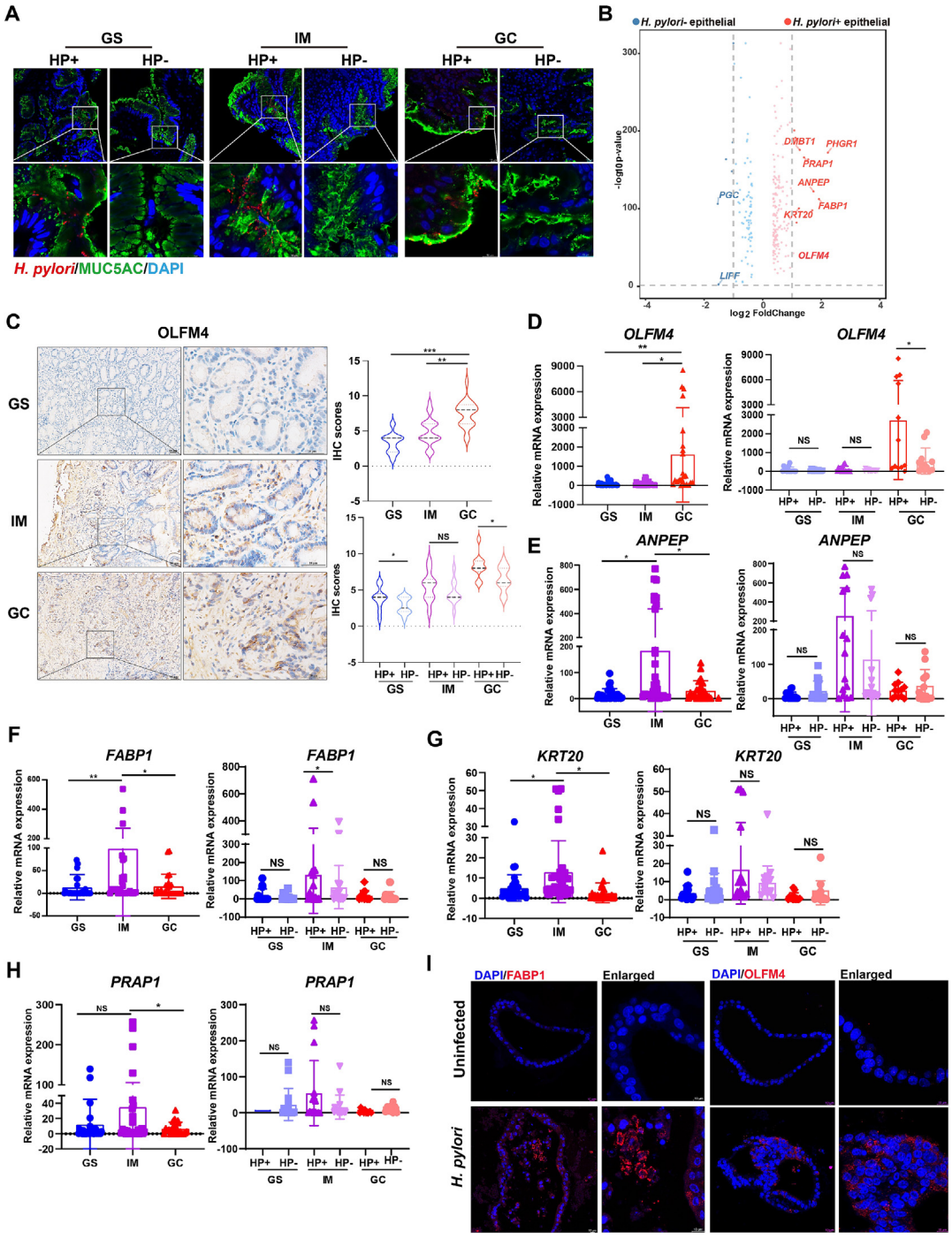


Fig. 4. Transcriptome heterogeneity of epithelial cells depending on *H. pylori* infection status. (A) Representative images of immunostaining for *H. pylori* (red) and MUC5AC (green) in human GS, IM and GC samples. Scale bar, 10 μm. (B) Volcano plot showing differential gene expression analysis of *H. pylori* −positive versus negative epithelial cells. (C) Immunostaining for OLFM4 expression in GS, IM and GC specimens (left panel, representative images; right panel, quantification analysis. Scale bar, 10 μm. (D)Violin plot, presenting the expression score of OLFM4 in different gastric pathological tissues with or without *H. pylori* infection, based on immunohistochemistry staining results. (E-H) qRT-PCR analysis detecting the RNA levels of *ANPEP* (E), *PARP1* (F), *FABP1* (G) and *KRT20* (H) in the human gastric tissues used in this study. (I) Immunofluorescence staining for FABP1 or OLFM4 expression in human gastric organoid following *H. pylori* PMSS1 strain infection. *NS, no significance*; $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

digestive enzymes secretion (e.g., *PGC*, *LIPF*, and *GAST*). Conversely, there was an elevation in the expression levels of genes typically overexpressed in metaplasia lesions (*PHGR1*, *FABP1*, *ANPEP*, *KRT20*, *PRAP1* and *DMBT1*) and extracellular matrix *OLFM4* oncogene (Fig. 4B, Table S6). This observation was further supported by the abnormal increase in both mRNA and protein levels of OLFM4 in GC tissues (Fig. 4C, D), in consistent with previous reports [23]. Notably, *H. pylori* −positive GC specimens exhibited higher levels of OLFM4 than *H. pylori*-negative subjects (Fig. 4D). Public datasets analysis also confirmed the upregulation of *OLFM4* in human stomach carcinoma (Fig. S3F). In accordance with scRNA-seq, the mRNA levels of *ANPEP*, *PRAP1*, *FABP1* and *KRT20* were significantly augmented in IM tissues, particularly in *H. pylori*-positive IM individuals, although without statistically significant differences (Fig. 4E–H). Given that organoids can faithfully recapitulate many aspects of tissues such as differentiation ability into tissue-specific lineages, we have established organoids from human gastric normal adjacent tissues. As a result, *H. pylori* infection significantly upregulated FABP1 and OLFM4 expression in gastric organoids (Fig. 4I), indicating a potential involvement of these genes in the pathogenesis of *H. pylori* infection within the gastric environment, highlighting their significance in gastric pathology.

The heterogeneity of ECs and ECs-tumor-like EPC subtypes interactome predictions

Endothelial cells (ECs) play a crucial role in tumorigenesis by governing angiogenesis to provide nutrients and oxygen for cell proliferation [24]. Considering the progressive rise in ECs proportion throughout gastric carcinogenesis cascade, our subsequent focus was to conduct a comprehensive exploration of the transcriptomic heterogeneity of ECs and to identify tissue-specific subpopulations. We successfully obtained 6266 high-quality ECs in which unsupervised clustering revealed 7 distinct subclusters (Fig. S4A). Among these subpopulations, we identified 4 cell subtypes corresponding to traditional vascular beds, i.e., venous ECs (*VCAM1*, *SELP*, *ACKR1*), lymphatic ECs (*LYVE1*, PROX1), capillary ECs (*PLPP3*, *RGCC*, *CA4*) and arteries ECs (*SEMA3G*, *GJA4*, *GJA5*) (Fig. 5A, Fig. S4B). Remarkably, all EC subsets were largely enriched in GC tissues, compared with those in GS and IM tissues (Fig. 5B). A specific cluster designated as tumor-like ECs due to remarkably increase in GC tissues (Fig. 5C) and highly expressed marker genes known to be enriched in angiogenesis and extracellular matrix (ECM) remodeling (*LGALS1*, *COL4A1*, *COL4A2* and *ESM1*) (Fig. 5D). The Endo_3 cluster exhibited significantly lower UMI counts compared to others, with highly expressed lncRNAs (e.g., *HES1*, *MALAT1*, *AC020916.1*), was not be further analyzed in depth (Fig. S4C). Both venous and capillary populations accounted for higher percentage of all ECs, whereas lymphatic population were rare (Fig. 5C). However, these populations manifested no significance between *H. pylori* −positive and −negative patients, except for tumor-like ECs (Fig. S4D, E). Enriched pathways identified in tumor-like ECs were related to cell adhesion and PI3K/Akt signaling pathway (Fig. 5E, F), particularly in *H. pylori*-positive subgroup, suggesting the potential role of *H. pylori* infection in promoting angiogenesis (Fig. 5G).

Since GC originated from the abnormal changes in epithelial cells [25], and tumor-like EPC was the specific epithelial-derived subset in GC, we speculated that ECs and tumor-like EPC cluster communicate closely. Using CellChat analysis [26], we observed that ECs subtypes predominantly drove outgoing signaling, while tumor-like EPC was the target for incoming signaling (Fig. 5H). The interactions between these cell types involved activation of COLLAGEN and VEGF signaling pathways (Fig. S4F, G), with COL4A1 and COL4A2 as major ligands secreted by ECs, while CD44 and SDC44 serving as their receptors on the tumor-like EPC. Moreover, the pro-angiogenic VEGF signaling pathway was mainly driven by VEGFA ligand from tumor-like EPC to VEGFR receptors (FLT1, KDR) from ECs (Fig. 5I, J). By independently re-analyzing the bulk RNA-seq data from gastric carcinoma GEO datasets (GSE122401, GSE65801, GSE13195) and the TCGA database, we found that COL4A1 and COL4A2, the secreted factors of COLLAGEN signaling pathways, were expressed at high levels in gastric malignancies compared with paired normal tissues by reanalyzing bulk RNA-seq gastric carcinoma GEO datasets (GSE122401, GSE65801, GSE13195) and TCGA database (Fig. S5A, B). However, only some RNA-seq data from public databases showed that the receptors *SDC4* and *CD44* were upregulated in gastric carcinomas (Figure S5C, D). Altogether, these findings highlighted the molecular signatures of GC tissue-specific EC subsets, and provided further insights into potential interactions between ECs and tumor-like EPC that may contribute to gastric carcinogenesis microenvironment.

Distinct fibroblast cell subtypes in cancerous and non-cancerous samples

Fibroblasts are highly organized stromal cells embedded in the ECM. Accumulating evidence has highlighted their remarkable diversity, with distinct phenotypes and functions associated with specific-cell lineage [27,28]. In our study, we employed unsupervised clustering method and identified 11 subsets of fibroblasts, with cluster 11 being a mixed cell population excluded from further analysis (Fig. S6A). Intriguingly, when comparing fibroblasts in non-cancerous tissue to those in gastric carcinoma tissues, we observed distinct lineage states in the latter, leading to their classification as cancer-associated fibroblasts (CAFs) (Fig. 6A). In detail, ECM fibroblasts (ECM_Fib; *COL5A1* and *POSTN*) and myofibroblasts (*ACTG2*) were characterized, mainly derived from GS and IM tissues. Swann cells was named due to specifically overexpression of S100B and PLP1 [29]. Additionally, three CAFs subpopulations were identified: antigen-presenting CAFs (apCAFs; *HLA-DRA* and *HLA-DRB1*), epithelial-mesenchymal transition CAFs (EMT_CAFs; *KRT8* and *KRT19*), and inflammatory CAFs (iCAFs; *FBLN1*, *IGFBP6* and *CXCL1*) (Fig. 6B; Fig. S6B, C) [30]. To further explore the gene set signature of CAFs, we performed a gene GSEA and differential gene expression analysis comparing all CAFs (iCAFs, apCAFs and EMT_CAFs) against other fibroblast subsets (ECM_Fib and Myofibroblasts) (Fig. 6C, D). Of note, CAFs were enriched for gene sets associated with immune and inflammation-related pathways (*IGFBP4*, *CXCL8*, *CXCL2*), complement cascades (*C3*, *C7*), and Wnt signaling pathway (*SFRP2*, *SFRP4*), while demonstrating lower expression of gene sets related to metabolism (Fig. 6C, D; Table S7). These findings underscore the crucial role of CAFs in shaping tumor immune microenvironment.

Given the higher fraction of iCAFs in GC tissues (Fig. 6E), we explored the characteristics of this subpopulation. We observed that iCAFs displayed elevated expression levels of complement components (*C3* and *C7*), extracellular regulators of Wnt signaling (*SFRP2* and *SFRP4*) [31]. Validation through dual-color immunofluorescence analysis confirmed the abundance of iCAFs ($COL1A1^{+}$-$FBLN1^{+}$ or $COL1A1^{+}C3^{+}$) within GC tissues (Fig. 6F). We next investigated the dynamic transition of fibroblast cells using pseudotime analysis and excluding schwann cells due to their distinct characteristics. This analysis revealed that ECM_Fib and myofibroblasts appeared at the initial stages of the trajectory path, while all types of CAFs were situated at terminally differentiated states (Fig. S6D). Of note, ECM_Fib subset was obsereved to differentiated into iCAF population (Fig. 6G; Fig. S6E). Moreover, according to the regulatory network using SCENIC analysis, NFATC1-driven regulons and transcriptional factor expression were inferred to be specific and enriched in iCAF subsets (Fig. S6F).

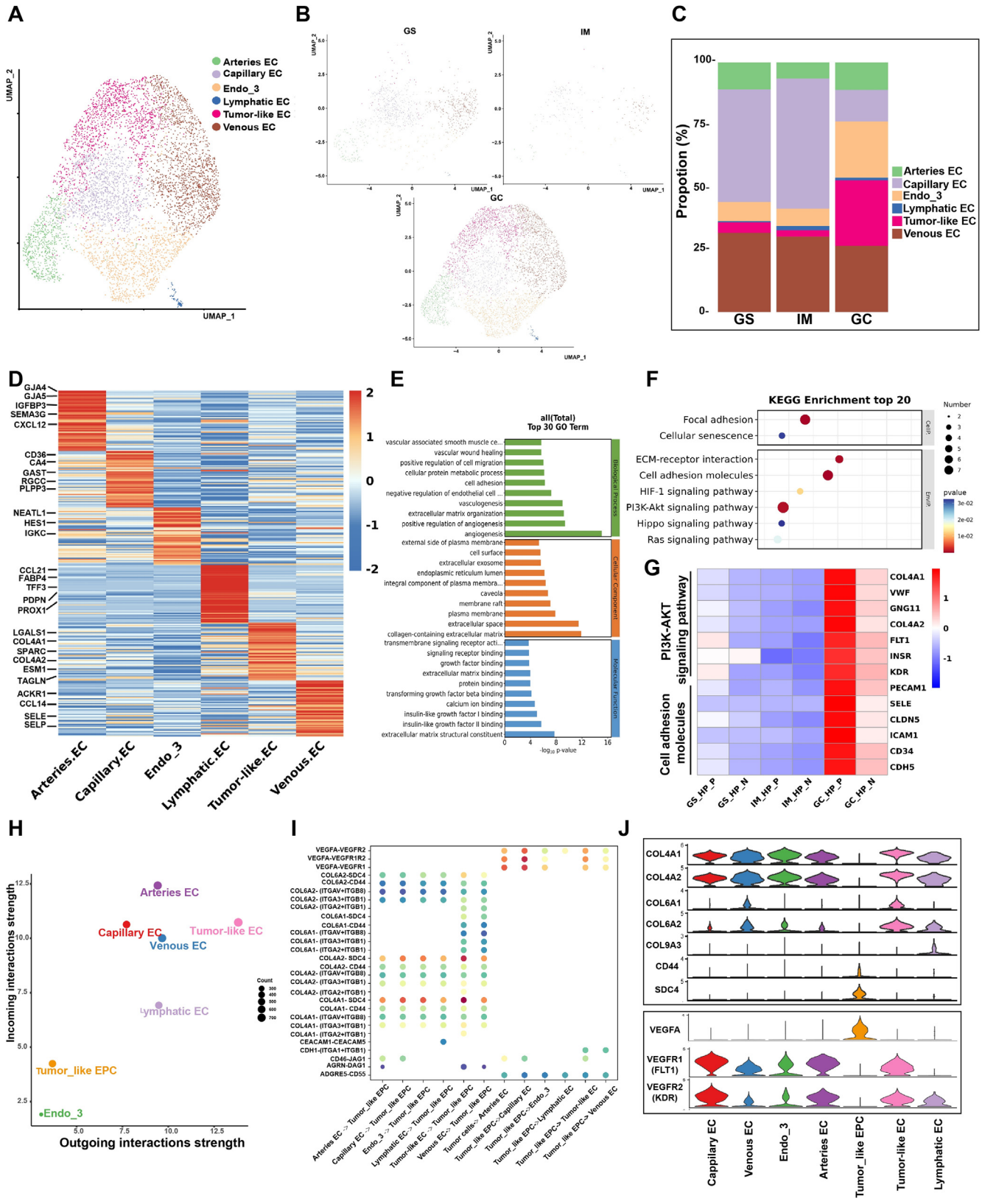
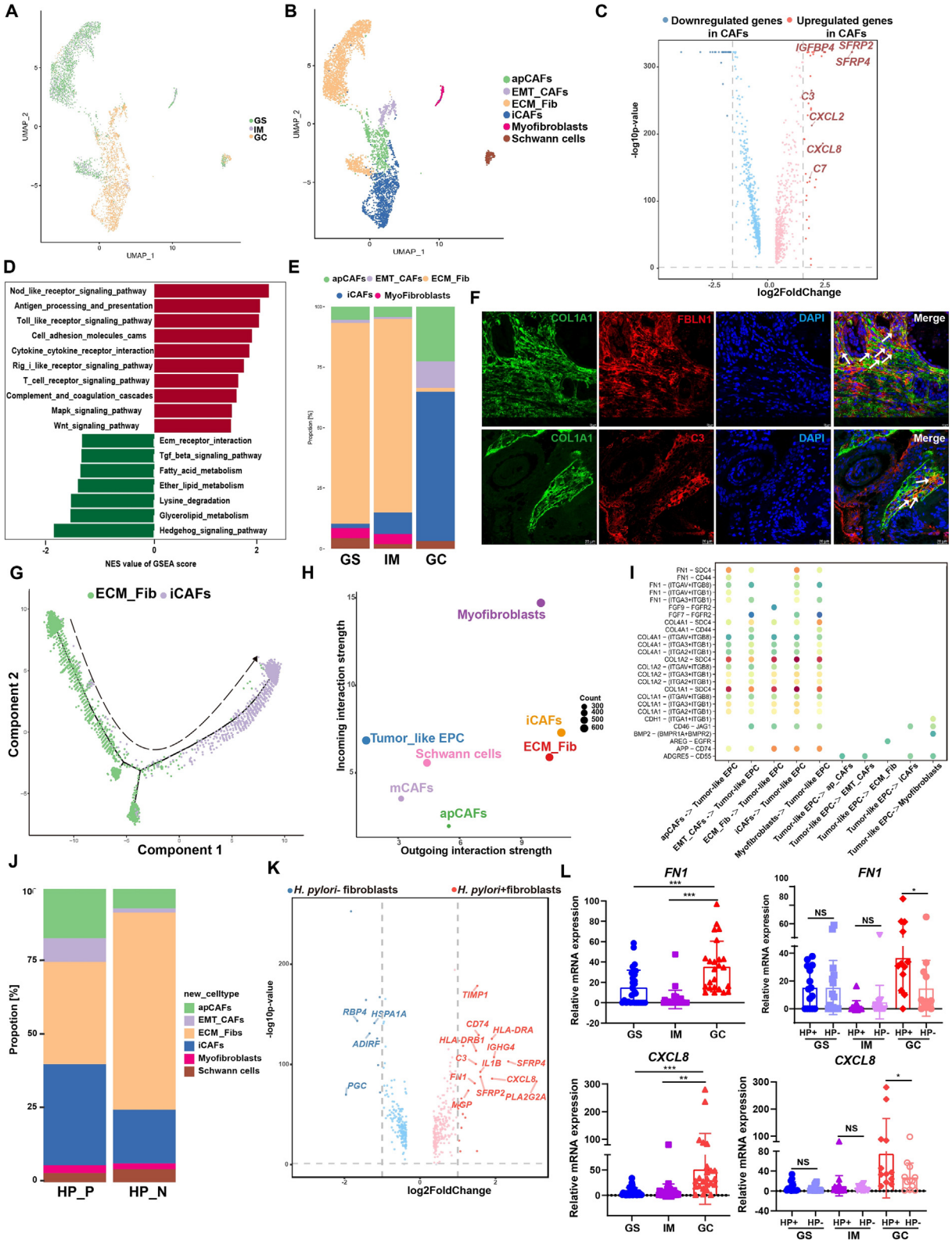


Fig. 5. The heterogeneity of ECs and ECs-epithelial tumor-like EPC subtypes interactome predictions. (A) UMAP plot of subtypes of endothelial cells (ECs) with color code. Cell type annotations were provided in the figure. (B) UMAP plot showing changes in cell number of ECs in GS, IM and GC tissues. (C) Histogram presenting the fraction of EC subtypes in GS, IM and GC, respectively. (D) Heatmap showing the top marker genes of all EC subtypes. Several marker genes were indicated. (E, F) Bar plot showing enriched functions and pathways of upregulated genes in tumor-like ECs, according to GO (E) and KEGG (F) enrichment analysis. (G) Heatmap showing the overall RNA levels of genes enriched in PI3K/AKT signaling pathway and cell adhesion in gastric tissues, according to pathological stages or *H. pylori* infection. (H) CellChat analysis predicting the communication between ECs and epithelial-derived tumor cells. Cell subclusters were projected onto a two-dimensional manifold based on their function. Each color-coded dot representing the communication network of a cell population. Dot size suggests the overall communication probability. (I) Bubble plots showing significant ligand-receptor pairs identified from ECs-tumor cells interaction. (J) Violin plot showing the expression of major ligands and receptors identified from ECs-tumor cells interaction.


A
UMAP_2
UMAP_1
GS
IM
GC
B
apCAFs
EMT_CAFs
ECM_Fib
iCAFs
Myofibroblasts
Schwann cells
C
Downregulated genes in CAFs
Upregulated genes in CAFs
IGFBP4
SFRP2
SFRP4
C3
CXCL2
CXCL8
C7
-log10p-value
log2FoldChange
D
Nod_like_receptor_signaling_pathway
Antigen_processing_and_presentation
Toll_like_receptor_signaling_pathway
Cell_adhesion_molecules_cams
Cytokine_cytokine_receptor_interaction
Rig_i_like_receptor_signaling_pathway
T_cell_receptor_signaling_pathway
Complement_and_coagulation_cascades
Mapk_signaling_pathway
Wnt_signaling_pathway
Ecm_receptor_interaction
Tgf_beta_signaling_pathway
Fatty_acid_metabolism
Ether_lipid_metabolism
Lysine_degradation
Glycerolipid_metabolism
Hedgehog_signaling_pathway
NES value of GSEA score
E
apCAFs EMT_CAFs ECM_Fib
iCAFs MyoFibroblasts
Propotion [%]
GS
IM
GC
F
COL1A1
FBLN1
DAPI
Merge
COL1A1
C3
DAPI
Merge
G
ECM_Fib iCAFs
Component 2
Component 1
H
Myofibroblasts
iCAFs
Tumor_like EPC
Schwann cells
ECM_Fib
mCAFs
apCAFs
Count
Incoming interaction strength
Outgoing interaction strength
I
FN1 – SDC4
FN1 – CD44
FN1 – (ITGAV+ITGB8)
FN1 – (ITGAV+ITGB1)
FN1 – (ITGA3+ITGB1)
FGF9 – FGFR2
FGF7 – FGFR2
COL4A1 – SDC4
COL4A1 – CD44
COL4A1 – (ITGAV+ITGB8)
COL4A1 – (ITGA3+ITGB1)
COL4A1 – (ITGA2+ITGB1)
COL1A2 – SDC4
COL1A2 – (ITGAV+ITGB8)
COL1A2 – (ITGA3+ITGB1)
COL1A2 – (ITGA2+ITGB1)
COL1A1 – SDC4
COL1A1 – (ITGAV+ITGB8)
COL1A1 – (ITGA3+ITGB1)
COL1A1 – (ITGA2+ITGB1)
CDH1 – (ITGA1+ITGB1)
CD46 – JAG1
BMP2 – (BMPR1A+BMPR2)
AREG – EGFR
APP – CD74
ADGRE5 – CD55
apCAFs -> Tumor-like EPC
EMT_CAFs -> Tumor-like EPC
ECM_Fib -> Tumor-like EPC
iCAFs-> Tumor-like EPC
Myofibroblasts-> Tumor-like EPC
Tumor-like EPC-> apCAFs
Tumor-like EPC-> EMT_CAFs
Tumor-like EPC-> ECM_Fib
Tumor-like EPC-> iCAFs
Tumor-like EPC-> Myofibroblasts
J
Propotion [%]
new_celltype
apCAFs
EMT_CAFs
ECM_Fibs
iCAFs
Myofibroblasts
Schwann cells
HP_P
HP_N
K
H. pylori- fibroblasts
H. pylori+fibroblasts
-log10p-value
log2FoldChange
RBP4
HSPA1A
ADIRF
PGC
TIMP1
CD74
HLA-DRA
HLA-DRB1
IGHG4
C3
IL1B
SFRP4
FN1
CXCL8
SFRP2
PLA2G2A
MGP
L
FN1
CXCL8
Relative mRNA expression
GS
IM
GC
HP+
HP-
NS

Prediction of the interactome between fibroblast-tumor-like EPC populations and identification of DEGs in fibroblast cells associated with H. pylori

Given the increased attention of fibroblasts in multicellular interaction in tumor microenvironment [32], we next predicted the interaction between fibroblasts and tumor-like EPC, to explore the mechanism through which microenvironment favors gastric tumorigenesis. Interestingly, CellChat analysis revealed that fibroblast subsets, particularly iCAFs, predominantly engaged in outgoing signaling by secreting a greater number of ligands. The corresponding incoming receptors were expressed in tumor-like EPC (Fig. 6H). In agreement with our previous analysis of ECs-tumor-like EPC interaction, the COLLAGEN signaling members, such as *COL1A1* and *COL1A2*, along with Fibronectin 1(FN1) signaling, were among the top-ranked ligands produced by fibroblasts. Their corresponding receptors in tumor-like EPC were similar, such as *CD44* and *SDC4* (Fig. 6I).

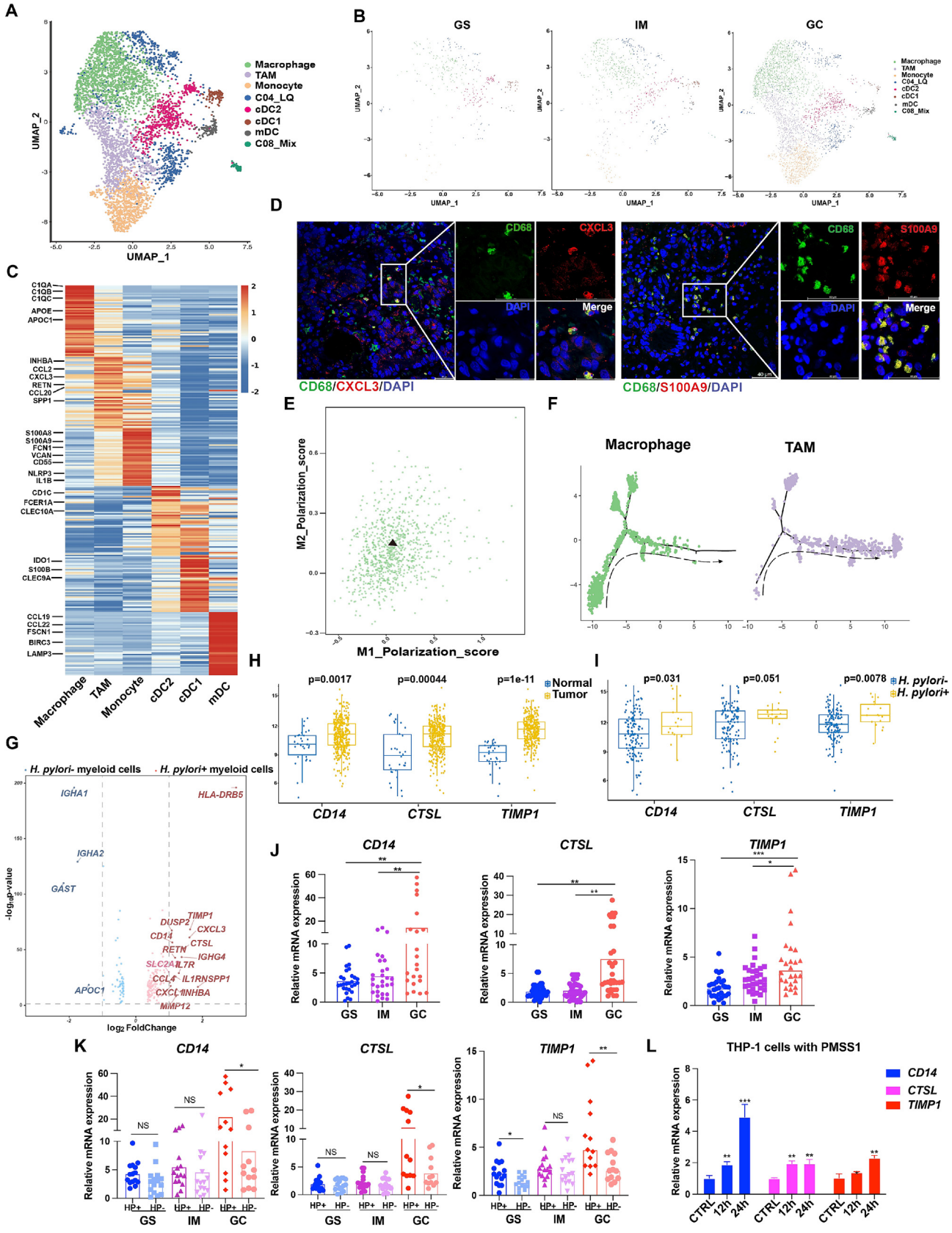
Existing evidence indicates the critical impact of *H. pylori* on gastric epithelial cells or macrophages, while the responses of stromal cells to *H. pylori* remains less understood. Considering the high heterogeneity of fibroblast cells in gastric malignant transformation, we investigated the effect of *H. pylori* on fibroblast cell composition. Notably, *H. pylori*-positive gastric specimens showed increased percentage of CAFs, as indicated by apCAFs, EMT_CAFs and iCAFs, whereas relatively reduced ECM_Fibs (Fig. 6J). Identification of DEGs between *H. pylori*-positive and −negative groups showed that several ligands and receptors for cellular signaling pathways were the top DEGs, such as *FN1*, *C3*, *IL1B*, *CXCL8* and *CD74* (Fig. 6K; Table S8). They were also significantly overexpressed at bulk RNA levels in GC tissues, particularly in *H. pylori* −positive GC tissues (Fig. S7A). Querying bulk RNA-seq datasets from the TCGA and GEO databases further supported the higher expression of *FN1*, *CXCL8*, *C3*, but not *IL1B*, in gastric cancerous tissues than that in paired adjacent normal tissues (Fig. S7B–E). To validate these findings, RT-PCR analysis conducted on an expanded sample size to confirm the elevated mRNA levels of *FN1* and *CXCL8* in GC tissues, particularly those associated *H. pylori* infection (Fig. 6L). Altogether, these findings suggest that *H. pylori* infection may drive the interaction between fibroblasts and epithelial cells in GC microenvironment via activation of FN1 signaling network.

Tissue-specific patterns of myeloid cells in H. pylori−related gastric carcinogenesis

To characterize the subsets of myeloid cells that exhibited a marked increase in GC, we identified five common cell subsets based on canonical cell markers: macrophage (*CD68*, *CD163* and *MRC1*), Monocyte (*VCAN* and *LYZ*), type 1 conventional dendritic cells (cDC1; *XCR1* and *CLEC9A*), type 2 cDC (cDC2; *CD1C* and *CLEC9A*), and mature dendritic cells (mDC; *LAMP3*) (Fig. 7A; Fig. S8A). Two clusters, cluster 4 and cluster 8, were excluded from further analyses due to low quality and mixed cells, respectively (Fig. S8B). A specific cluster of TAMs was designated, based on highly expressed marker genes of macrophages and enriched in tumor tissues (Fig. 7A, B; Fig. S8D). Of note, total myeloid cells number were much higher in GC tissues, particularly in *H. pylori*-infected GC tissues, compared to GS and IM tissues (Fig. 7B, Fig. S8C). Apart from TAM, monocytes also showed a relatively high proportion in GC, whereas the cDC2 subtype tended to gradually decrease in gastric cancerous cascade (Fig. S8D). The complement components (*C1QA*, *C1QB* and *C1QC*), and cholesterol metabolism genes (*APOE* and *APOC1*) were enriched in macrophages, as described previously [33]. Monocytes highly expressed a set of neutrophil-associated genes (*S100A8* and *S100A9*), pro-inflammatory cytokine *IL1B* and inflammasome *NLRP3* (Fig. 7C). Dual-color staining further validated the presence of TAM subsets, as indicated by co-expression of CD68 and CXCL3. Likewise, monocytes were verified by combined immunostaining of CD68 and S100A9 (Fig. 7D). GSVA analysis revealed enrichment of monocyte gene sets involved in metabolism pathways, such as fatty acid biosynthesis, bile secretion. The gene sets of TAMs were involved in IL-17, TNF pathway and cytokines interactions (Fig. S8E). To clarify M1 and M2 polarization of TAMs, we calculated pro-and anti-inflammatory scores using related gene sets (Table S9). In agreement with previous studies [15], TAMs exhibited dominant M2-like phenotype, suggesting its immunosuppressive role in tumor microenvironment (Fig. 7E). Pseudotime analysis using monocle 2 suggested that both TAM in neoplastic tissues were likely derived from macrophages in GS and IM tissues (Fig. 7F).

The interactome network between myeloid and epithelial cells was further studied. Interestingly, the outgoing and incoming signaling patterns were obviously observed between myeloid cell subpopulations in tumor microenvironment, compared to myeloid-epithelial cells. Macrophages, TAMs, monocytes and cDC2 cells not only secreted cytokines, but also expressing corresponding receptors (Fig. S8F). To identify myeloid cell-derived genes regulated by *H. pylori*, we compared the DEGs from myeloid cells between *H. pylori*-positive and *H. pylori*-negative tissues. Notably, myeloid cells from *H. pylori*-positive tissues displayed high expression of *CXCL3*, *TIMP1*, *CD14* and *CTSL* genes (Fig. 7G; Table S10). Further analysis of gastric carcinoma bulk RNA-seq datasets from GEO and TCGA database showed that CD14, *CTSL* and *TIMP1* were significantly upregulated in GC tissues compared to the paired adjacent non-cancerous tissues (Fig. 7H). Of note, *H. pylori*-positive patients showed even higher expression (Fig. 7I; Fig. S8G-I). Likewise, these results were further validated by qRT-PCR analysis with additional samples (Fig. 7J–K), reinforcing the association of *CD14*, *CTSL*, and *TIMP1* with *H. pylori* infection in GC. Moreover, our results demonstrated that *H. pylori* infection significantly increased the mRNA levels of *CD14*, *CTSL* and *TIMP1* in PMA-differentiated THP-1 cells (Fig. 7L), further supporting the regulatory impact of *H. pylori* on these specific genes in myeloid cells.

Fig. 6. Characterization of fibroblast cell subclusters in *H. pylori*-associated gastric precancerous and cancerous samples. (A) UMAP plot of fibroblasts in GS, IM and GC tissues which were color coded. (B) UMAP plot showing the subclusters of fibroblast cells, with color code. Each cluster was a color-coded according to cell types. (C) Volcano plot showing differential gene expression analysis between CAFs (including apCAFs, EMT_CAFs and iCAFs) and normal fibroblasts (including ECM_Fibs and myofibroblasts). (D) Waterfall plot indicating the enriched pathways in CAFs (red color) compared to normal fibroblasts (green color). (E) The histogram plot showing the fraction of fibroblast subtypes in GS, IM and GC tissues. (F) Immunostaining confirming the presence of iCAFs in GC samples by identification of $COL1A1^{+}FBLN1^{+}$ and $COL1A1^{+}C3^{+}$ cells. Scale bar, 20 μm. (G) Pseudotime analysis of ECM_Fib (green color) and iCAFs (purple color) from gastric tissues. (H) CellChat analysis predicting the communication network between fibroblast subtypes and epithelial-derived tumor cells. Each cell cluster was distributed in the plot according to the outgoing or incoming interaction strength. (I) Bubble plots showing major ligand-receptor pairs between fibroblast subtypes and epithelial-derived tumor cells. (J) The proportion of fibroblast cell subsets in *H. pylori* −positive and negative specimens, shown by histogram plot. (K) The volcano plot showing DEGs in fibroblasts between *H. pylori* positive and negative gastric samples. (L) The qRT-PCR analysis illustrating the expression of *FN1*, and *CXCL8* in GS, IM and GC specimens with or without *H. pylori* infection, respectively. CAFs, cancer-associated fibroblasts; ap_CAFs, antigen-presenting CAFs; EMT_CAFs, epithelial-mesenchymal transition CAFs; iCAFs, inflammatory CAFs; ECM_Fibs, extracellular matrix fibroblasts. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)


A
Macrophage
TAM
Monocyte
C04_LQ
cDC2
cDC1
mDC
C08_Mix
UMAP_1
UMAP_2
B
GS
IM
GC
C
C1QA
C1QB
C1QC
APOE
APOC1
INHBA
CCL2
CXCL3
RETN
CCL20
SPP1
S100A8
S100A9
FCN1
VCAN
CD55
NLRP3
IL1B
CD1C
FCER1A
CLEC10A
IDO1
S100B
CLEC9A
CCL19
CCL22
FSCN1
BIRC3
LAMP3
Macrophage
TAM
Monocyte
cDC2
cDC1
mDC
D
CD68
CXCL3
DAPI
Merge
S100A9
CD68/CXCL3/DAPI
CD68/S100A9/DAPI
E
M2_Polarization_score
M1_Polarization_score
F
Macrophage
TAM
H
p=0.0017
p=0.00044
p=1e-11
Normal
Tumor
CD14
CTSL
TIMP1
I
p=0.031
p=0.051
p=0.0078
H. pylori-
H. pylori+
CD14
CTSL
TIMP1
G
H. pylori- myeloid cells
H. pylori+ myeloid cells
IGHA1
HLA-DRB5
IGHA2
GAST
APOC1
TIMP1
DUSP2
CXCL3
CD14
CTSL
RETN
IGHG4
SLC2A3
IL7R
CCL4
IL1RNSPP1
CXCL1
INHBA
MMP12
-log10p-value
log2 FoldChange
J
CD14
CTSL
TIMP1
Relative mRNA expression
GS
IM
GC
K
CD14
CTSL
TIMP1
Relative mRNA expression
NS
HP+
HP-
GS
IM
GC
L
THP-1 cells with PMSS1
CD14
CTSL
TIMP1
Relative mRNA expression
CTRL
12h
24h

T, B and plasma cell clustering and subtype analysis

We next applied unsupervised clustering of immune T and B cells to reveal their intrinsic substructures. A total of 14 subclusters of T cells were classified based on well-known markers CD4 and CD8 (Fig. 8A). The first CD4 cluster CD4_IL7R highly expressed Naïve marker IL17R that is critical for T cells activation and development [34]. The second cluster CD4_PDCD1, showing gradually reduced proportion in gastric carcinogenesis cascade, specifically expressed T follicular helper (Tfh) markers, such as *TOX2* and *PDCD1* [35,36]. The final cluster CD4_FOXP3 was well-known Treg cells with immunosuppressive effects based on the dominant *CTLA4*, *FOXP3* expression [37] (Fig. 8A–C). According to signature genes expression and top-ranked genes, $CD8^+$T cells comprised 7 subtypes: CD8_GZMK and CD8_GNLY cells highly expressed cytotoxicity-associated genes (*GZMK* and *GZMB*). CD8_IFNG showed activated resident memory-related markers (*CCL4*, *IFNG* and *CCL3*) [38,39]; CD8_CAPG exhibited elevated expression of *CAPG* and novel immune checkpoint *KLRC1* [40]. CD8_IL17A (*IL17A* and *KLRB1*) cells showed similar activation mode and cytokines secretion with TH17 cells [41]. CD8_XCL1 cells showed enriched expression of *TRDC*, *XCL1*, *XCL2*, and *CD160* genes. The CD8_MKI67 cluster was designated based on uniquely expression of *MKI67* [36,42] (Fig. 8A, B and D). In addition, three clusters were characterized as DNT (*HSPA1A* and *HSPA1B*), C14_KRT81 (*KRT81*) [43]and natural killer (NK) cells (*NKG7*).

We found that Treg cells represent higher proportion in GC samples than that in GS or IM sample (Fig. 8E, F). When assessing the role of *H. pylori* on T cell heterogeneity, we observed a trend of increased CD8_IL17A subset in *H. pylori*−positive (versus *H. pylori*-negative) GS tissues, consistent with previous reported viewpoints regarding the role of IL17 cytokine in driving protective immunity against pathogen and promoting inflammatory response [41] (Fig. 8F). Combined immunofluorescence staining labeled CD4 and FOXP3 also revealed Treg cells to be enriched for GC, as previously reports where Treg cells are increased in several solid malignancies and associated with poor prognosis [44] (Fig. 8G). The trajectory analysis using monocle 3 indicated that $CD4^+$T cells originated from CD4_IL7R cells and then differentiated into CD4_PDCD1 and CD4_Treg cells (Fig. 8H), as previously described that the sustained expression of IL7R in the naïve states is critical for the maintenance and development of all T subpopulations [34]. As for $CD8^+$T cells, we observed a notable enrichment of the CD8_MKI67 subpopulation in GC tissues (Fig. 8F). Multi-color immunohistochemical staining for CD3, CD8 and Ki67 further confirmed the presence of CD8_MKI67 cells in gastric carcinoma tissues (Fig. 8I), supporting the previous findings indicating the proliferative function of T cells in tumors [38]. Taken together, these findings suggested the potential functional role of Treg cells, and IL17A-expressing and cytotoxic characteristics $CD8^+$T cells in *H. pylori* −associated gastric neoplasm.

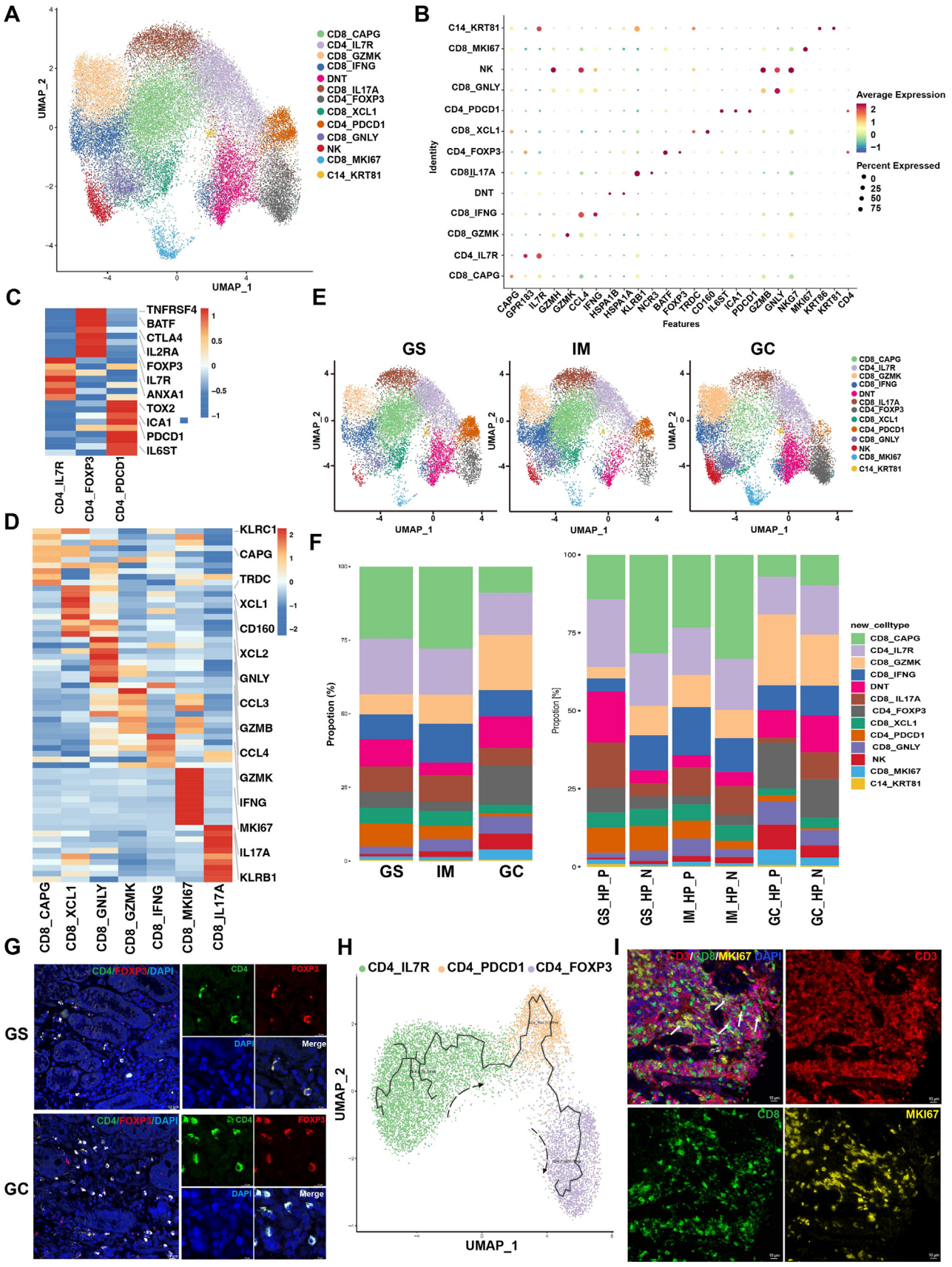
The re-clustering of B and plasma cells revealed 6 populations, including active B (*CD69*), Naïve B (*TCL1A*, *IL4R*), Pre-pro B (*CD7*, *IL7R*), IgA plasma (*IGHA1*, *IGHA2*), IgG plasma (*IGHG1*, *IGHG2*), and IgM plasma cells (*IGHM*) (Figure S9A, B). Overall, B and plasma cells defined clearly separate clusters, with the lower abundance in GC samples (Fig. S9C). The relative cellular proportion of plasma cells various across tissues. IgA^+ plasma cells were highly prevalent in GS/IM, while IgG^+ plasma subsets were enriched in GC samples (Fig. S9D), in agreement with previous work [10]. Only slight difference was existed between *H. pylori*-positive and negative gastric tissues (Fig. S9E). Nevertheless, IgG^+ plasma population appears to be more abundant in *H. pylori*-infected individuals (Fig. S9F).

Multicellular communication network shapes H. pylori−driven carcinogenesis microenvironment

To understand the cellular signaling pathways driving gastric carcinogenesis cascade, we used CellChat to infer autocrine and paracrine signaling network among all cell types, including epithelial cells, stromal cells and immune cells. The analysis suggested that stromal cells were the dominant senders of signaling, while immune T and myeloid cells were the receivers (Fig. S10A). ECs, as well as fibroblasts, also showed relatively strong interaction with epithelial cell subpopulations, pit mucous cells and tumor-like EPC (Fig. S10B, C). The COLLAGEN signaling component, which included the classic marker *COL1A1*, was mainly driven by fibroblasts and almost implicated in communication with all other cell types (Fig. S10D). Another fibroblast-derived C3 signaling pathway uniquely recruited to myeloid cells through C3-C3AR1 ligand-receptor pair (Fig. S10E, H). In line with the previous results, the VEGF signaling pathway was dominant in ECs and tumor-like EPC interaction, as well as ECs-myeloid cells interaction (Fig. S10F). Additionally, we found that epithelial-derived pit mucous cells secreted IL1B and that recruited ILR2-expressing myeloid cells (Fig. S10G–H). Importantly, the communication network of all cell subtypes inferred that iCAFs subsets with high expression in GC dominantly drive outgoing signaling (Fig. 9A). Due to the significant increase of myeloid cells in GC and the immunomodulatory role of iCAFs, we further focused on exploring the interaction between iCAFs and myeloid cell subsets. The analysis suggested that iCAFs primarily secreted cytokines and were recruited to myeloid cells, particularly the TAM subpopulations (Fig. 9B). The COL1A1, MIF, C3, and APP signaling regulatory networks were mainly responsible for the communication between iCAFs and myeloid cells (Fig. 9C). Based on the exclusive enrichment of iCAFs and myeloid cells, these findings suggest their interaction as an important decisive factor for poor prognosis in GC.

We then compared the overall communication probability in *H. pylori*-positive and negative samples. Notably, *H. pylori*-infected gastric tissues exhibited a significant increase in the number and strength of inferred intercellular communications (Fig. 9D). Moreover, our ligand-receptor pairs analysis predicted several enriched pathways in *H. pylori* −positive specimens, such as SPP1, TNF, THY1 and complement pathways, while GAS and IL10 signaling networks were more abundant in *H. pylori*-negative specimens (Fig. 9E). We subsequently focused on TNF pathways that has been implicated in inflammatory diseases [45]. The analysis predicted TNF-TNFRSF1A (TNFR1) was found to be highly active in *H. pylori* −positive samples, suggesting the important role of TNF-TNFR1

Fig. 7. Tissue-specific patterns of myeloid cells in *H. pylori*-related gastric carcinogenesis. (A) UMAP plot of subtypes of myeloid cells. (B) UMAP plot showing the differences of myeloid cell subsets in GS, IM and GC tissues. (C) Heatmap of DEGs among different myeloid cell subtypes. (D) Representative images of immunostaining indicating the presence of TAMs ($CD68^+CXCL3+$) and monocytes ($CD68^+S100A9^+$) in GC tissues. Scale bar, 40 μm. (E) The polarization phenotype scores of TAM subset. (F) Trajectory analysis predicting the cell origin of TAM using Monocle 2. (G) Volcano analysis showing the differential gene expression analysis between *H. pylori* −positive and negative myeloid cells. (H-I) The expression levels of *CD14*, *CTSL* and *TIMP1* in normal and tumor tissues (the upper panel) or in *H. pylori*-positive and negative tumor tissues (the lower panel), from TCGA-STAD cohort. (J, K) The qRT-PCR analysis detecting the expression of *CD14*, *CTSL* and *TIMP1* in GS, IM and GC specimens with or without *H. pylori* infection. (L) The qRT-PCR analysis for *CD14*, *CTSL* and *TIMP1* mRNA expression in PMA-stimulated THP1 cells following *H. pylori* infection. *NS, no significance*; $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$. TAM, Tumor-associated macrophage; LQ, low quality; cDC1, type 1 conventional dendritic cells; cDC2, type 2 conventional dendritic cells; mDC, and mature dendritic cells.


A
UMAP_2
UMAP_1
CD8_CAPG
CD4_IL7R
CD8_GZMK
CD8_IFNG
DNT
CD8_IL17A
CD4_FOXP3
CD8_XCL1
CD4_PDCD1
CD8_GNLY
NK
CD8_MKI67
C14_KRT81
B
Identity
Features
Average Expression
Percent Expressed
C14_KRT81
CD8_MKI67
NK
CD8_GNLY
CD4_PDCD1
CD8_XCL1
CD4_FOXP3
CD8IL17A
DNT
CD8_IFNG
CD8_GZMK
CD4_IL7R
CD8_CAPG
CAPG
GPR183
IL7R
GZMH
GZMK
CCL4
IFNG
HSPA1B
HSPA1A
KLRB1
NCR3
BATF
FOXP3
TRDC
CD160
IL6ST
ICA1
PDCD1
GZMB
GNLY
NKG7
MKI67
KRT86
KRT81
CD4
C
TNFRSF4
BATF
CTLA4
IL2RA
FOXP3
IL7R
ANXA1
TOX2
ICA1
PDCD1
IL6ST
CD4_IL7R
CD4_FOXP3
CD4_PDCD1
E
GS
IM
GC
UMAP_2
UMAP_1
D
KLRC1
CAPG
TRDC
XCL1
CD160
XCL2
GNLY
CCL3
GZMB
CCL4
GZMK
IFNG
MKI67
IL17A
KLRB1
CD8_CAPG
CD8_XCL1
CD8_GNLY
CD8_GZMK
CD8_IFNG
CD8_MKI67
CD8_IL17A
F
Propotion (%)
Propotion [%]
new_celltype
GS
IM
GC
GS_HP_P
GS_HP_N
IM_HP_P
IM_HP_N
GC_HP_P
GC_HP_N
G
GS
GC
CD4/FOXP3/DAPI
CD4
FOXP3
DAPI
Merge
H
CD4_IL7R
CD4_PDCD1
CD4_FOXP3
UMAP_2
UMAP_1
I
CD3/CD8/MKI67/DAPI
CD3
CD8
MKI67

pathway in the pathogenesis of *H. pylori* infection (Fig. 9F and G). These results highlighted the putative importance of intercellular crosstalk in *H. pylori* −induced gastric neoplastic transformation.

Discussion

Gastric adenocarcinoma has the high incidence and death rates, especially in Eastern Asia countries. Non-cardia cancer, the most common GC and arises in distal regions of the stomach, is almost attributable to *H. pylori* infection [46]. Non-cardia GC is most commonly of the intestinal-type gastric adenocarcinoma, which develops through a sequential progression of gastric lesions known as the Correa cascade [3].

In this study, we employed scRNA-seq to comprehensively analyze the cellular transcriptomic heterogeneity and communication among all cell types including epithelial, stromal and immune cells in gastric carcinogenesis microenvironment. We identified specific cell populations linked to gastrointestinal metaplasia and gastric carcinoma (Fig. 9H). Importantly, our notable finding delineated the impact of *H. pylori* infection in different cell subtypes heterogeneity across GC progression and identified potential biomarkers for *H. pylori* −associated GC (Fig. 9H). Our findings were orthogonally validated by expanding sample size and utilizing bulk RNA-seq data from public GEO database and TCGA-STAD cohort, and mechanically confirmed in gastric epithelial cells and myeloid cells for studying the effect of *H. pylori* infection. Building upon these insights, we have proposed a working model to provide a novel and comprehensive landscape of *H. pylori*-associated-gastric carcinogenesis ecosystem.

Previous studies have explored the diversity of cell types in gastric precancerous and cancerous lesions [47,48], but this analysis was often limited to specific cell populations within small sample size. For instance, Kumar et al. compared the tumor microenvironment of gastric cancer tissues with different Lauren pathological subtypes, with specific focus on investigating the states and transcriptomic patterns of plasma cells and fibroblasts lineages [8]. Similarly, Zhang and colleagues performed scRNA-seq in human gastric tissues (56440 cells) to study the molecular features of epithelial cells in gastric precancerous lesions and early GC [49]. Our study represents an innovative approach by performing single-cell RNA sequencing analyses on uniform samples of progressively gastric pathological lesions, including six each of GS, IM, and GC. With a substantial increase in the number of cell analyzed ($>$ 90000 cells), we aimed to create a cellular heterogeneity landscape encompassing all cell subtypes across gastric carcinogenic cascade. A recent single-cell study also comprehensively investigated the transcriptomic heterogeneity of stromal and immune cells. There were several concordant findings with our study, such as high enrichment of IgA^+ plasma cells in precancerous lesions, stromal cells remodeling and high abundance of M2-like TAM cluster in GC [10]. Significantly, our study goes beyond previous studies by emphasizing the interaction among all cell subtypes and elucidated the impact of *H. pylori* infection in the heterogeneity and expression profiles of various cell types throughout different stages of gastric pathology. By providing a holistic view of cellular dynamics and intercellular crosstalk within the gastric microenvironment, our study contributes valuable insights into the complex landscape of gastric carcinogenesis and the role of *H. pylori* infection in shaping the cellular ecosystem during disease progression.

Gastric adenocarcinoma often preceded from precancerous lesions, such as metaplasia and dysplasia, which originate from alteration in epithelial cells [50]. The primary innovation of our findings may have relevance for the management of precancerous lesions IM-induced by *H. pylori* infection. Notably, we observed the enrichment of epithelial cell subpopulations enterocytes in IM tissues, especially in *H. pylori* −infected samples. Enterocytes exhibited specific high expression of intestinal mucosa markers *FABP1*, *DMBT1* and *KRT20*, which were enriched in metabolic pathways. Our observations align with previous spatiotemporal genomic profiling of IM tissues, which revealed the intestinal-type lineages in IM, including enterocytes (*FABP1*, *KRT20*) and transit amplifying cells (*DMBT1*). Importantly, we further determined that the transcription factor HNF4G that has been previously shown to regulate intestinal differentiation [21,51], serves as a potential driver for enterocytes identity. The signature genes of enterocytes (*ANPEP*, *PRAP1*, *FABP1*, *KRT20*), could be regulated following *H. pylori* infection, which was validated by RT-PCR analysis of our further expanded samples. These results provide potential molecular interventions for inhibiting the progression of *H. pylori*-induced IM.

One notable finding from our study was the identification of an extremely abundant subset of epithelial cells in GC tissues, defined as tumor-like EPC with the following features: 1. Higher CNV numbers; 2. Significant high expression of KRT17 and CEACAM5; 3. Their origin from pit mucous cells. Notably, the gene set of tumor cells was involved in cancer-related signaling pathways such as EMT and NF-κB signaling pathway. Moreover, we found that the potential oncogene *OLFM4* was positively related to *H. pylori* infection. Earlier report showed two specific subpopulations in IM, one marked by OLFM4 responsible for intestinal stem cell phenotype and the other HES6-marked globet cells [52]. Our findings revealing the high expression of OLFM4 in GC cells contrast somewhat with previous observations in IM, possibly due to individual variations in human IM and GC sampling. To delve deeper into the role of gastric microenvironment during the progression from gastritis to GC, we analyzed the crosstalk between tumor-like EPC and stromal cell populations. We uncovered that the COLLAGEN signaling network mediated the communications between stromal cells and tumor-like EPC. Additionally, the VEGF signaling was involved in the tumor-like EPC communication with ECs, facilitated by ligand-receptor pairs *VEGFA* and *FLT1*, *KDR*. These insights shed light on the importance of cell–cell interactions in gastric carcinogenesis.

Fibroblasts were highly organized constituents of stromal cells. Functionally, fibroblasts are fundamental regulators of ECM synthesis and supporting nearby epithelial cells growth and differentiation by paracrine signaling [53]. Recently, CAFs, widely considered as the distinct cell population in tumors, have been shown to play cardinal roles in shaping tumor environment. Accumulating evidence indicates that CAFs largely influence tumor initiation, growth and dissemination through ECM remodeling, tumor immunity and angiogenesis [54]. In this study, compared to GS and IM, gastric carcinoma presented a completely different fibroblast cell states, named CAFs. Immunomodulatory pathways were significantly upregulated in CAFs, such as antigen presentation, Nod-like receptor and Toll-like receptor pathways. In detail, we observed that CAFs in GC exhibited high heterogeneity, including

Fig. 8. Characterization of T cell subtypes. (A) UMAP plot showing T cell subtypes. (B) Dotplot showing the expression levels of gene signatures across T cell clusters. (C, D) Heatmap showing the expression levels of top10 gene signatures across $CD4^+$T (C) and $CD8^+$T (D), respectively. (E) UMAP plot showing tissue prevalence of each T cell subset in GS, IM and GC tissues. (F) Histogram plot showing the cellular proportions of T cell populations across gastric neoplastic stages, with or without *H. pylori* infection. (G) Representative images of immunofluorescence indicating the presence of Treg cells ($CD4^+FOXP3^+$) in GS and GC tissues. Scale bar, 10 μm. (H) Monocle trajectory inference of $CD4^+$T and $CD8^+$T cells, colored by cell clusters. (I) Multi-color staining of CD3/CD8/MKI67 showing the presence of $CD8^+MKI67^+$ cluster in GC tissues. Scale bar, 20 μm.

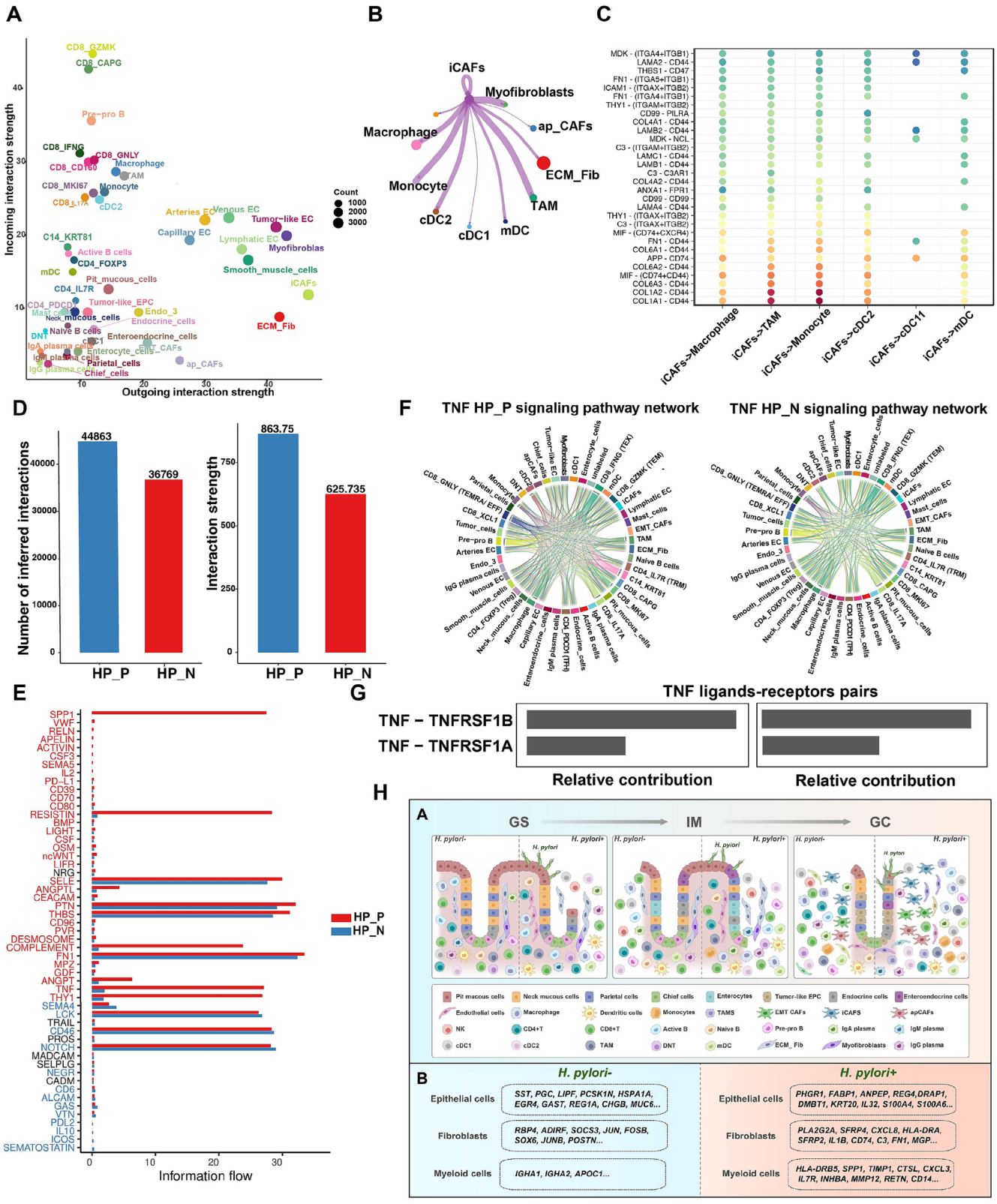


Fig. 9. Multicellular communication network shapes *H. pylori* –driven carcinogenesis microenvironment. (A) CellChat analysis predicting the communication network between all cell subtypes. (B) The communication intensity network between fibroblast and myeloid cell subsets. iCAFs primarily secreted ligands. (C) Bubble plots showing significant ligand-receptor pairs between iCAFs and myeloid cell subtypes. (D) Histogram showing the number and strength of inferred interactions of *H. pylori* positive and negative gastric tissues, respectively. (E) The significant signaling pathways ranked based on the differences of overall information flow in the inferred networks between *H. pylori* positive and negative gastric tissues. (F) The chord diagram showing inferred intercellular communication network of TNF signaling between *H. pylori* positive and negative samples. (G) The histogram showing differences of relative contribution of ligand-receptor pairs in TNF signaling between *H. pylori* positive and negative specimens. (H) Graphical summary of cellular heterogeneity and microenvironment along the consecutive gastric carcinogenesis associated with *H. pylori*. Cellular heterogeneity along the histological cascade of gastric carcinogenesis with or without *H. pylori* infection. The DEGs between *H. pylori* –positive and –negative cell types. GS, gastritis; IM, intestinal metaplasia; GC, gastric cancer.

apCAFs, iCAFs and EMT_CAFs, aligning with the earlier studies that have recognized these subpopulations and their roles in regulating tumor immunity, potentially contributing to tumor immune evasion [55,56]. In addition, given that *H. pylori* favors epithelial cell colonization and causes inflammatory response, previous studies have mostly focused on epithelial cells or macrophages, instead of exploring the effects of *H. pylori* on stromal cells. Here, we compared the cellular heterogeneity and expression programs in fibroblast subpopulations between *H. pylori* infected and uninfected tissues. Bulk-RNA seq data from GEO and TCGA cohort and experiments with expanded gastric samples, have shown that fibroblast-secreted ligands *FN1* and *CXCL8* may be involved in the pathogenesis of *H. pylori*. However, further research is needed to elucidate the molecular mechanisms through which genes expressed in fibroblasts contribute to *H. pylori* infection-induced gastric tumorigenesis.

Myeloid cells, constituting a key cellular component of immune cells, plays an essential role in the pathogenesis of bacterial infection. Macrophage is the major type of myeloid cells that could efficient phagocytosis and kill *H. pylori* [57]. However, the persistent infection with *H. pylori* exerts inflammation and subsequent carcinogenesis through influencing macrophage function, implying the substantial heterogeneity of macrophages in *H. pylori* −associated gastric lesions. In agreement with the recent studies that revealed a remarkable enrichment of myeloid cells in several solid tumors [58,59], we also confirmed the significant role of myeloid cells in gastric carcinoma. Additionally, we identified distinct subsets of TAMs and monocytes in GC tissues, and revealed their enriched immunometabolic functions and inflammatory pathways, suggesting the potential mechanisms of tumor immunosuppression. Emerging evidence suggests that TAMs, acting as the driver of immunosuppression in the tumor microenvironment, contribute to cancer progression and metastasis. TAMs promote the expansion of immunosuppressive Treg cells by the secretion of IL-10 and TGFβ, while concurrently suppressing the activity of cytotoxic T and NK cells that have the potential to kill cancers [60,61]. Moreover, our data provides evidence supporting that the involvement of *CD14*, *CTSL* and *TIMP1* genes expressed in myeloid cells in *H. pylori* −induced carcinogenesis. Further studies are warranted to reveal their functions and the underlying molecular mechanisms.

The complex interactions between epithelial cells, stromal cells and immune cells shape the tumor ecosystem and profoundly modulate tumor biology. Accumulating evidence indicates the crucial role of interactions among immune cells feature immunosuppressive and pro-tumorigenic functions [62]. The intercellular communication network between all cell types in gastric carcinogenesis. This study highlighted the pivotal role of stromal cells, particularly iCAFs and tumor-like subsets, in serving as key secretory ligands, while immune cells express receptors. Notably, our findings revealed that the intercellular communications were more active in *H. pylori* −positive gastric tissues. Several significant signaling pathways, such as SPP1, TNF and THY1 signaling networks, were indicated to regulate the cellular microenvironment in response to *H. pylori* infection. Therefore, to prevent *H. pylori* −related gastric mucosal lesions, it is crucial to consider not only the factors related to individual cell types but also the complex communications between various cell types. Documented evidence shows that the interaction between cancer-derived fibroblasts and metaplastic epithelial cells promotes transition of SPEM-expressing metaplasia into dysplasia lesions [63]. Future studies should consider how intercellular interactions contribute to GC initiation and progression, which would provide novel insights for the diagnosis and treatment of GC. In-depth mechanistic studies of interaction between ligands and receptors in the tumor microenvironment would pave the way to identify novel therapeutic targets in GC treatment.

In summary, as one of the largest single-cell analysis of Correa's cascade of gastric carcinogenesis to date, we comprehensively unveiled both common and specific characteristics of nearly all cell types from pre-cancerous lesions to primary cancers related to *H. pylori* infection. Notably, for the first time, we provide insights into the influence of *H. pylori* infection on cellular heterogeneity in the gastric microenvironment, which may contribute to a better understanding of the pathogenesis of *H. pylori*, and further offer valuable resources for the prevention and early diagnosis of GC.

As this is a preliminary study, there are several limitations in this study. Firstly, the acquisition of human gastric tissues, especially gastritis and IM tissues obtained via endoscopy, is inherently challenging. This resulted in a relatively small sample size for scRNA-seq. Moreover, in compliance with medical ethics guidelines, we were unable to obtain normal human stomach tissues. Secondly, scRNA-seq has limitations of low capture efficiency, high dropouts, and amplification biases [64,65]. Finally, we have made preliminary investigations into the DEGs and potential regulatory networks that involved in *H. pylori*-associated gastric tumorigenesis. Further in-depth research is needed to elucidate, for example, how intercellular interactions regulated by signaling network such as TNF and THY1 modulate cell–cell communication following *H. pylori* infection through in vivo and in vitro experiments. And the functional implications of the identified cell subsets CAFs, tumor-like ECs and TAMs within gastric carcinoma microenvironment were further investigated.

Declarations

Data sharing statement

The raw single-cell sequencing data of 18 human gastric samples from this study have been deposited in the gene expression omnibus database (GEO), under accession code GSE249874 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi). The publicly available gastric cancer datasets GSE122401, GSE65801, and GSE13195 used in this study were available in GEO databases. The TCGA pan-cancer dataset from the National Cancer Institute was also used.

Summary: ScRNA-seq profiling reveals cellular heterogeneity and microenvironment in *H. pylori*-related gastric carcinogenesis.

Ethics approval and consent to participate.

This work was approved by the Ethics Committee of the First Affiliated Hospital of Nanchang University under protocol number of (2023) CDYFYYLK (01–009). All tissue samples were obtained under informed consent.

CRediT authorship contribution statement

Nianshuang Li: Conceptualization, Methodology, Software, Validation, Writing – original draft, Writing – review & editing, Visualization, Funding acquisition, Supervision. **Sihai Chen:** Software. **Xinbo Xu:** Methodology, Validation. **Huan Wang:** . **Pan Zheng:** Software, Validation. **Xiao Fei:** Validation. **Huajing Ke:** . **Yuting Lei:** Validation. **Yanan Zhou:** . **Xiaoyu Yang:** . **Yaobin Ouyang:** . **Chuan Xie:** Software. **Cong He:** Methodology. **Yi Hu:** . **Yi Cao:** . **Zhengrong Li:** . **Yong Xie:** . **Zhongming Ge:** Writing – review & editing. **Xu Shu:** . **Nonghua Lu:** Conceptualization, Methodology, Writing – review & editing, Funding acquisition, Supervision. **Jianping Liu:** Software, Writing – review & editing, Visualization. **Yin Zhu:** Conceptualization, Methodology, Writing – review & editing, Funding acquisition, Supervision.

Funding

This work was supported by the National Natural Science Foundation of China (82260119, 82170580, and 82260118), Academic and Technical Leader of major disciplines in Jiangxi Province

(20225BCJ23021), the Natural Science Foundation of Jiangxi Province (20224ACB216004), Chunhui Program of the Ministry of Education (HZKY20220394) and Jiangxi Provincial Health Commission project (202410013).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Key Laboratory Project of Digestive Diseases in Jiangxi Province (2024SSY06101), and Jiangxi Clinical Research Center for Gastroenterology (20223BCG74011). We thank all the patients for their generous donation of tissue samples in this study. We also thank Dr. Yongbing Ba and Wenyang Ding from OE biotech company (Shanghai, China) for the support of bioinformatic analysis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jare.2024.10.012.

References

[1] Arnold M et al. Global burden of oesophageal and gastric cancer by histology and subsite in 2018. Gut 2020;69(9):1564–71.
[2] Bray F et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68(6):394–424.
[3] Crowe SE. Helicobacter pylori Infection. N Engl J Med 2019;380(12):1158–65.
[4] Malfertheiner P et al. Helicobacter pylori infection. Nat Rev Dis Primers 2023;9(1):19.
[5] Yan L et al. Effect of helicobacter pylori eradication on gastric cancer prevention: updated report from a randomized controlled trial with 26.5 years of follow-up. Gastroenterology 2022.
[6] Cancer Genome Atlas Research N. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 2014;513(7517):202-9.
[7] Mun DG et al. Proteogenomic characterization of human early-onset gastric cancer. Cancer Cell 2019;35(1):111–24.
[8] Kumar V et al. Single-cell atlas of lineage states, tumor microenvironment, and subtype-specific expression programs in gastric cancer. Cancer Discov 2022;12(3):670–91.
[9] Zhang M et al. Dissecting transcriptional heterogeneity in primary gastric adenocarcinoma by single cell RNA sequencing. Gut 2021;70(3):464–75.
[10] Wang R et al. Evolution of immune and stromal cell states and ecotypes during gastric adenocarcinoma progression. Cancer Cell 2023;41(8):1407–26.
[11] Fei X et al. Macrophage biology in the pathogenesis of Helicobacter pylori infection. Crit Rev Microbiol 2024;1–18.
[12] Xie C et al. Inhibition of autophagy aggravates DNA damage response and gastric tumorigenesis via Rad51 ubiquitination in response to H. pylori infection. Gut Microbes 2020;11(6):1567–89.
[13] Li N et al. YAP and beta-catenin cooperate to drive H. pylori-induced gastric tumorigenesis. Gut Microbes 2023;15(1):2192501.
[14] He C et al. Convergent dysbiosis of gastric mucosa and fluid microbiome during stomach carcinogenesis. Gastric Cancer 2022;25(5):837–49.
[15] Sun Y et al. Single-cell landscape of the ecosystem in early-relapse hepatocellular carcinoma. Cell 2021;184(2):404–21.
[16] Zhang M et al. Dissecting transcriptional heterogeneity in primary gastric adenocarcinoma by single cell RNA sequencing. Gut 2020.
[17] Bockerstett KA et al. Single-cell transcriptional analyses identify lineage-specific epithelial responses to inflammation and metaplastic development in the gastric corpus. Gastroenterology 2020.
[18] Bockerstett KA et al. Single-cell transcriptional analyses of spasmolytic polypeptide-expressing metaplasia arising from acute drug injury and chronic inflammation in the stomach. Gut 2020;69(6):1027–38.
[19] Chen HY et al. Caudal type homeoboxes as a driving force in Helicobacter pylori infection-induced gastric intestinal metaplasia. Gut Microbes 2020;12(1):1–12.
[20] Miao ZF et al. DDIT4 licenses only healthy cells to proliferate during injury-induced metaplasia. Gastroenterology 2021;160(1):260–71 e10.
[21] Chen L et al. HNF4 regulates fatty acid oxidation and is required for renewal of intestinal stem cells in mice. Gastroenterology 2020;158(4):985–99 e9.
[22] Aguilar C et al. Helicobacter pylori shows tropism to gastric differentiated pit cells dependent on urea chemotaxis. Nat Commun 2022;13(1):5878.
[23] Oh HK et al. Genomic loss of miR-486 regulates tumor progression and the OLFM4 antiapoptotic factor in gastric cancer. Clin Cancer Res 2011;17(9):2657–67.
[24] Palikuqi B et al. Adaptable haemodynamic endothelial cells for organogenesis and tumorigenesis. Nature 2020;585(7825):426–32.
[25] Hayakawa Y et al. Stem cells and origins of cancer in the upper gastrointestinal tract. Cell Stem Cell 2021;28(8):1343–61.
[26] Jin S et al. Inference and analysis of cell-cell communication using Cell Chat. Nat Commun 2021;12(1):1088.
[27] Kobayashi H et al. Cancer-associated fibroblasts in gastrointestinal cancer. Nat Rev Gastroenterol Hepatol 2019;16(5):282–95.
[28] Buechler MB et al. Cross-tissue organization of the fibroblast lineage. Nature 2021;593(7860):575–9.
[29] Lovatt D et al. scRNA-seq generates a molecular map of emerging cell subtypes after sciatic nerve injury in rats. Commun Biol 2022;5(1):1105.
[30] Zhang M et al. Single-cell transcriptomic architecture and intercellular crosstalk of human intrahepatic cholangiocarcinoma. J Hepatol 2020;73(5):1118–30.
[31] Liu J et al. Wnt/beta-catenin signalling: function, biological mechanisms, and therapeutic opportunities. Signal Transduct Target Ther 2022;7(1):3.
[32] Mao X et al. Crosstalk between cancer-associated fibroblasts and immune cells in the tumor microenvironment: new findings and future perspectives. Mol Cancer 2021;20(1):131.
[33] Zilionis R et al. Single-cell transcriptomics of human and mouse lung cancers reveals conserved myeloid populations across individuals and species. Immunity 2019;50(5):1317–34 e10.
[34] Barata JT et al. Flip the coin: IL-7 and IL-7R in health and disease. Nat Immunol 2019;20(12):1584–93.
[35] Xu W et al. The transcription factor Tox2 drives T follicular helper cell development via regulating chromatin accessibility. Immunity 2019;51(5):826–39 e5.
[36] Guo X et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. Nat Med 2018;24(7):978–85.
[37] Zheng C et al. Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing. Cell 2017;169(7):1342–56 e16.
[38] Szabo PA et al. Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. Nat Commun 2019;10(1):4706.
[39] Sasson SC et al. Interferon-gamma-producing CD8(+) tissue resident memory T cells are a targetable hallmark of immune checkpoint inhibitor-colitis. Gastroenterology 2021;161(4):1229–44 e9.
[40] Salome B et al. NKG2A and HLA-E define an alternative immune checkpoint axis in bladder cancer. Cancer Cell 2022;40(9):1027–43.
[41] Mills KHG. IL-17 and IL-17-producing cells in protection versus pathology. Nat Rev Immunol 2023;23(1):38–54.
[42] Liu Y et al. Immune phenotypic linkage between colorectal cancer and liver metastasis. Cancer Cell 2022;40(4):424–37 e5.
[43] Miheecheva N et al. Multiregional single-cell proteogenomic analysis of ccRCC reveals cytokine drivers of intratumor spatial heterogeneity. Cell Rep 2022;40(7):111180.
[44] Tay C et al. Tumor-infiltrating regulatory T cells as targets of cancer immunotherapy. Cancer Cell 2023;41(3):450–65.
[45] Kalliolias GD et al. TNF biology, pathogenic mechanisms and emerging therapeutic strategies. Nat Rev Rheumatol 2016;12(1):49–62.
[46] Morgan E et al. The current and future incidence and mortality of gastric cancer in 185 countries, 2020–40: a population-based modelling study. EClinicalMedicine 2022;47:101404.
[47] Wang R et al. Single-cell dissection of intratumoral heterogeneity and lineage diversity in metastatic gastric adenocarcinoma. Nat Med 2021;27(1):141–51.
[48] Huang XZ et al. Single-cell sequencing of ascites fluid illustrates heterogeneity and therapy-induced evolution during gastric cancer peritoneal metastasis. Nat Commun 2023;14(1):822.
[49] Zhang P et al. Dissecting the single-cell transcriptome network underlying gastric premalignant lesions and early gastric cancer. Cell Rep 2019;27(6):1934–47 e5.
[50] Dissecting the Single-Cell Transcriptome Network Underlying Gastric Premalignant Lesions and Early Gastric Cancer - PubMed.
[51] Chen L et al. A reinforcing HNF4-SMAD4 feed-forward module stabilizes enterocyte identity. Nat Genet 2019;51(5):777–85.
[52] Zhang P et al. Dissecting the single-cell transcriptome network underlying gastric premalignant lesions and early gastric cancer. Cell Rep 2020;30(12):4317.
[53] Younesi FS et al. Fibroblast and myofibroblast activation in normal tissue repair and fibrosis. Nat Rev Mol Cell Biol 2024;25(8):617–38.
[54] Yang D et al. Cancer-associated fibroblasts: from basic science to anticancer therapy. Exp Mol Med 2023;55(7):1322–32.
[55] Elyada E et al. Cross-species single-cell analysis of pancreatic ductal adenocarcinoma reveals antigen-presenting cancer-associated fibroblasts. Cancer Discov 2019;9(8):1102–23.
[56] Huang H et al. Mesothelial cell-derived antigen-presenting cancer-associated fibroblasts induce expansion of regulatory T cells in pancreatic cancer. Cancer Cell 2022;40(6):656–73.
[57] Chaturvedi R et al. Polyamines impair immunity to helicobacter pylori by inhibiting L-arginine uptake required for nitric oxide production. Gastroenterology 2010;139(5):1686–98.
[58] Zhang Q et al. Landscape and Dynamics of Single Immune Cells in Hepatocellular Carcinoma. Cell 2019;179(4):829–45.

[59] Moncada R et al. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. Nat Biotechnol 2020;38(3):333–42.
[60] Mantovani A et al. Macrophages as tools and targets in cancer therapy. Nat Rev Drug Discov 2022;21(11):799–820.
[61] Cassetta L et al. A timeline of tumour-associated macrophage biology. Nat Rev Cancer 2023;23(4):238–57.
[62] Ozga AJ et al. Chemokines and the immune response to cancer. Immunity 2021;54(5):859–74.
[63] Lee SH et al. Apposition of fibroblasts with metaplastic gastric cells promotes dysplastic transition. Gastroenterology 2023;165(2):374–90.
[64] Jovic D et al. Single-cell RNA sequencing technologies and applications: a brief overview. Clin Transl Med 2022;12(3):e694.
[65] Chen G et al. Single-cell RNA-Seq technologies and related computational data analysis. Front Genet 2019;10:317.