Cancer Immunology, Immunotherapy (2025) 74:218
https://doi.org/10.1007/s00262-025-04048-6

RESEARCH

Analyses of single-cell RNA sequencing uncover the role of intratumoral *Helicobacter pylori* in shaping tumor progression and immunity in gastric cancer

Jiao Xu[1] · Jin Yang[1,2,3,4] · Qi Sun[5] · Jingbo Chang[5] · Fan Wang[2]

Received: 4 December 2024 / Accepted: 8 April 2025

Abstract

The intratumoral microbiota is closely associated with tumor initiation and progression in multiple solid tumors, including gastric cancer (GC). Single-cell analysis of host–microbiome interactions (SAHMI) is a pipeline used to systematically recover and denoise microbial signals in human clinical tissues and examine tumor–microbiome interactions at the single-cell transcriptome level. In a large GC cohort, we used SAHMI to detect 12 bacteria, among which *Helicobacter pylori* (*H. pylori*) was widely present in multiple tumor and normal samples. Meanwhile, we verified the presence of *H. pylori* in GC tissues via fluorescence in situ hybridization and immunohistochemistry. We performed single-cell RNA sequencing to analyze 11 cell populations, including B cells, T cells, and epithelial cells, and these cell types contained large numbers of *H. pylori*. We detected obvious enrichment of *H. pylori* in cancer cells and identified 13 upregulated differentially expressed genes exhibiting significantly negative correlations with patient survival in the *H. pylori*-positive tumor group compared with the findings in the other groups, indicating that these genes could represent prognostic biomarkers or therapeutic targets for *H. pylori*-infected patients with GC. Moreover, *H. pylori*-enriched immune cells, including T cells, B cells, and macrophages, were associated with cell-type-specific gene expression and pathway activities, including cell fate and immune signaling. In summary, tumor–microbiome interactions might reflect or influence tumorigenesis in GC, which has implications for clinical practice.

Keywords Gastric cancer · SAHMI · Intratumoral microbiota · *Helicobacter pylori* · Single-cell RNA sequencing

Abbreviations

CagA	Cytotoxin-associated gene A
CCL17	C-C chemokine ligand 17
DEG	Differentially expressed gene
FISH	Fluorescence in situ hybridization
GC	Gastric cancer
GEO	Gene Expression Omnibus
GO	Gene ontology
H&E	Hematoxylin and eosin
H. pylori	*Helicobacter pylori*
IHC	Immunohistochemistry
KEGG	Kyoto Encyclopedia of Genes and Genomes
OS	Overall survival
PCA	Principal component analysis
SAHMI	Single-cell analysis of host–microbiome interaction
scRNA-seq	Single-cell RNA sequencing
STAD	Stomach adenocarcinoma
TCGA	The Cancer Genome Atlas
TME	Tumor microenvironment

✉ Fan Wang
wangfan230@outlook.com

1 Precision Medicine Center, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, Shaanxi, People's Republic of China

2 Phase I Clinical Trial Research Center, The First Affiliated Hospital of Xi'an Jiaotong University, No. 277 Yanta West Road, Xi'an 710061, Shaanxi, People's Republic of China

3 Cancer Center, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, Shaanxi, People's Republic of China

4 Department of Medical Oncology, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, Shaanxi, People's Republic of China

5 Department of General Surgery, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, Shaanxi, People's Republic of China

Published online: 24 May 2025

UMAP	Uniform Manifold Approximation and Projection
UMI	Unique molecular identifier
VacA	Vacuolation cytotoxin A

Introduction

Gastric cancer (GC) is the fifth most common malignancy and fifth leading cause of cancer-related death globally [1]. Because of the lack of effective clinical markers, delayed diagnosis of GC is common, resulting in patients frequently being diagnosed with advanced disease, and the 5-year survival rate is lower than 5% for patients with metastatic GC [2]. Intratumoral microbes have been found in various cancers, such as breast, pancreatic, lung, gastric, and liver cancers [3, 4]. Intratumoral microorganisms affect both the host immune system and the effectiveness of drug treatment. Increasing evidence illustrates that microbiota communities within GC tissues significantly differ from those in nonmalignant tissues [5–8]. In 1994, the International Agency for Research on Cancer classified *Helicobacter pylori* (*H. pylori*) as a class I carcinogen and reconfirmed this classification in 2009 [9]. *H. pylori* infection is a well-established risk factor for GC, being present in approximately 89% of distal GC cases globally [10]. *H. pylori* colonizes the human stomach mucosa, and its infection can lead to chronic gastritis, atrophic gastritis, intestinal metaplasia, dysplasia, and ultimately gastric carcinoma [11]. Hence, it is crucial to investigate the mechanisms underlying the adhesion and invasion of gastric mucosa epithelial cells by *H. pylori*. Various studies demonstrated that *H. pylori* can invade and proliferate in epithelial cells, indicating that this process might significantly contribute to chronic infection, disease induction, and immune escape [12–15]. *H. pylori* infection is associated with GC via the induction of aberrant DNA methylation, gene mutations, and disrupted intracellular signaling pathways. Even in the normal mucosa, the accumulation of point mutations and aberrant DNA methylation can occur, resulting in field cancerization [16]. *H. pylori* infection initiates inflammatory responses, including the activation of Toll-like receptor, interleukin, and NF-κB pathways. Meanwhile, chronic exposure to these inflammatory processes combined with oxidative stress induced by reactive oxygen and nitrogen species can impair immune defense mechanisms, leading to tissue injury and an elevated risk of GC [17].

Existing genomic sequencing data from bulk tissue samples can be used to study host–microbiome interactions. However, these traditional methods might miss microbe–somatic cell enrichment, associations with cell-type-specific activities, and microbial contributions to intercellular communication networks. Using single-cell RNA sequencing (scRNA-seq) and spatial transcriptomics data, it is possible to precisely detect somatic cells and microbes and analyze host–microbiome ecosystems. Single-cell analysis of host–microbiome interactions (SAHMI) is a computational pipeline utilized to systematically recover and denoise microbial signals in human clinical tissue and examine tumor–microbiome interactions at single-cell resolution [18, 19]. SAHMI provides a platform to investigate patterns of human–microbiome interactions using single-cell sequencing data, eliminating the necessity for additional experimental modifications, and facilitates the generation of testable hypotheses regarding host–microbiome relationships across multiple levels. Therefore, we employed SAHMI to investigate the distribution of intratumoral bacteria in GC using single-cell data, elucidated the role of *H. pylori* in promoting GC, and highlighted its effect on the tumor microenvironment (TME).

Materials and methods

Public data collection

For this study, we primarily utilized the GSE150290 dataset, which provided comprehensive scRNA-seq data for GC samples, encompassing 23 tumoral and 29 normal tissues. To broaden the context of our analysis and validate findings, we also incorporated several additional datasets, including GSE112302, GSE163558, and GSE134520 which provided scRNA-seq data for various gastric conditions. Furthermore, GSE191139 provided bulk RNA sequencing data related to GC. Clinical information about these datasets is presented in Table S1.

Microbial identification using the SAHMI pipeline

To clarify the intratumoral microbiota, we employed the SAHMI pipeline. This pipeline encompasses several steps, including the alignment of sequencing reads to a comprehensive database of microbial genomes, followed by classification using Kraken2Uniq. Our analysis concentrated on reads that were not classified as human. Subsequently, we filtered these reads to identify microbial species at both the genera and species levels. Additionally, the pipeline incorporates contamination control measures by comparing detected microbial signals against a large reference dataset of sterile RNA-seq runs, thereby ensuring the robustness of our microbial identifications. To investigate the true taxa, we performed Spearman's correlation analysis of total and unique k-mer counts for each microbial genus or specie identified across sample or barcode levels. Specifically, we calculated the total k-mer and unique k-mer counts for each microbial genera and species using

samples or barcodes and then used Spearman's rank correlation coefficient (ρ) to assess the relationships between these counts. Statistical significance was indicated by $p < 0.05$.

Hematoxylin and eosin (H&E) staining and immunohistochemistry (IHC)

Tumoral and matched normal tissues from patients with GC were obtained from the Department of General Surgery at the First Affiliated Hospital of Xi'an Jiaotong University (Xi'an, China). This study was approved by the Ethics Committee of the First Affiliated Hospital of Xi'an Jiaotong University (No. XJTU1AF2023LSK-417), and written informed consent was obtained from all participants. The clinical information of patients with GC is presented in Table S2. Briefly, gastric tissue samples obtained during surgery were cut into 3–5-mm pieces and washed five times with 1 × phosphate-buffer saline. Subsequently, the tissues were fixed in 4% paraformaldehyde, dehydrated through a graded ethanol series, and embedded in paraffin, and then, tissue sections were deparaffinized. For H&E staining, the sections were stained stepwise with hematoxylin solution and eosin solution. For IHC, antigen retrieval was performed using citrate buffer (pH 6.0), and endogenous peroxidase blocking was performed using 3% hydrogen peroxide solution. After blocking sections with 3% BSA for 30 min at room temperature, they were incubated with an antibody against cytotoxin-associated gene A (CagA; Santa Cruz, sc-28368, 1:100) overnight at 4 °C. The sections were then incubated with the corresponding secondary antibody, and 3,3′-diaminobenzidine tetrahydrochloride served as the chromogenic reagent. Sections were counterstained with hematoxylin and dehydrated. The resulting signals were visualized using a light microscope (Nikon Eclipse E100, Nikon, Tokyo, Japan).

16S rRNA fluorescence in situ hybridization (FISH)

Slides were treated with 20 μg/mL proteinase K at 40 °C for 10 min. Then, 50 μL prehybridization solution without probe was applied for prehybridization at 37 °C for 1 h. A Cy3-labeled probe (5′-CACACCTGACTGACTATCCCG-3′) targeting *H. pylori* 16S rRNA (Wuhan Servicebio Technology Co., Ltd.) was synthesized and used. The slides were incubated with hybridization solution (1.5:100) overnight at 40 °C. Subsequently, the slides were washed with saline sodium citrate. Finally, nuclei were counterstained using DAPI, and sections were observed and imaged under a Nikon positive fluorescence microscope.

Processing of scRNA-seq data

The initial Unique Molecular Identifier (UMI) count matrix was subjected to rigorous filtering using the R package Seurat (v4.9.9.9039) [20]. Cells that did not meet specific quality criteria—namely, those with UMI numbers < 500, detected gene counts < 200 or > 8000, or the presence of more than 15% mitochondrial-derived UMI counts—were classified as low-quality and subsequently excluded from the analysis. Additionally, genes detected in fewer than three cells were removed prior to downstream analyses. Following stringent quality control, the count values were normalized using the 'SCTransform' function in Seurat by employing the 'glmGamPoi' method. Then, the 'SelectIntegrationFeatures' function of Seurat was utilized to identify the top 3000 variable genes. To reduce dimensionality and mitigate potential batching effects, principal component analysis (PCA) and Harmony (v0.1.1) [21] were executed in sequence. Using the 'Elbowplot' function of Seurat, the top 50 principal components were identified for downstream analysis. Following this, the primary cell clusters were delineated using the 'FindNeighbors' (dims = 1:50) and 'FindClusters' functions in Seurat (resolution = 0.6), resulting in the categorization of cells into 33 major cell clusters. These cell clusters were visualized using Uniform Manifold Approximation and Projection (UMAP) with the 'RunUMAP' function (dims = 1:50). To determine the cell type for each cluster, we utilized a compilation of previously published marker genes representing various cell types in the stomach. Initially, we employed the ScType automatic cell-type annotation package in R [22] to annotate the cell types. Subsequently, a meticulous manual review and corrections were undertaken to refine the annotations. Ultimately, we successfully annotated 11 distinct cell types. Similarly, epithelial cells, T cells, B cells, and macrophages subtypes were annotated using a comparable strategy, albeit with reduced dimensions (dims = 1:25) and a different resolution parameter (resolution = 0.3).

Pseudotime analysis of epithelial cells

In the pseudotime analysis of epithelial cells, we isolated epithelial cells and utilized Monocle2 (version 2.26.0) for trajectory construction and pseudotime-dependent gene expression calculations [23]. We specifically selected highly variable genes with a mean expression level of at least 0.1 for the unsupervised ordering of the cells. The DDRTree algorithm within Monocle2 was employed for trajectory reconstruction. To identify genes exhibiting significant expression differences along the developmental trajectory, we applied the differentialGeneTest function, retaining genes with FDR lower than 0.001. Visualization was conducted using

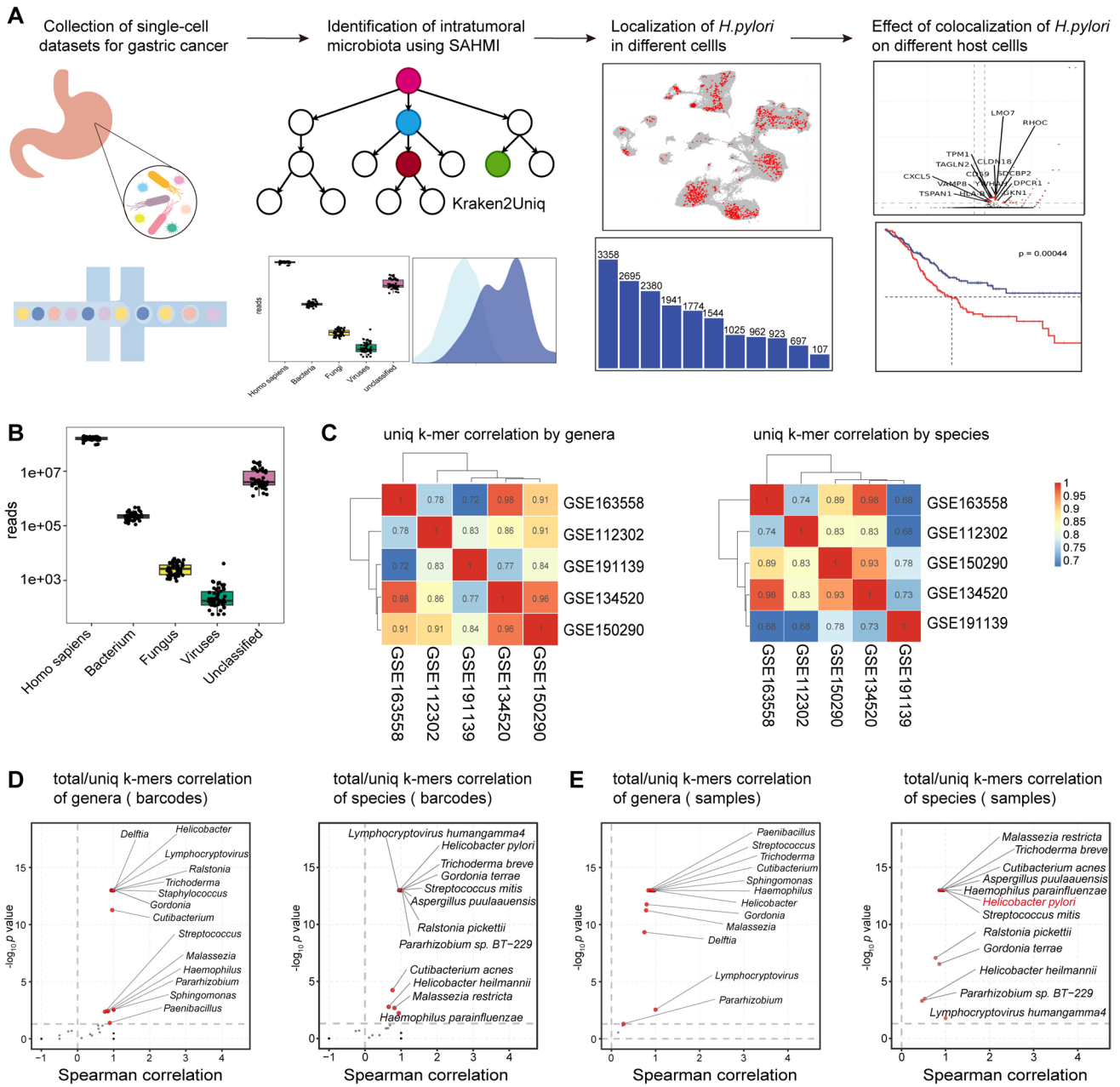


Fig. 1 Detection and validation of the GC microbiome using scRNA-seq data. **A** Schematic illustration of the study design and analytical workflow. Study design included GC data collection, identification of intratumoral microbiota, localization of *H. pylori* within different cell types, and assessment of its effect on host cells. **B** Box plot showing the classification of reads from the GSE150290 dataset by SAHMI into various categories including humans, bacterium, fungus, viruses, and unclassified species. **C** Heatmap depicting the Spearman correlation of unique k-mer counts for microbial genera/species identified across GC samples from multiple studies, illustrating the reproducibility of the microbial profiles across different sequencing datasets. **D** Scatterplot showing the Spearman correlation between the number of reads and the number of the total/unique k-mers assigned to each genus/specie across scRNA-seq barcodes. Each point represents a genus/specie, with the x-axis representing the correlation coefficient and the y-axis representing the -$\log_{10}$ p value. Each red point indicates a genus/specie with significant correlation ($p < 0.05$). **E** Scatterplot similar to (D), but across different samples rather than barcodes, each point represents a genus/specie, with the x-axis representing the correlation coefficient and the y-axis representing the -$\log_{10}$ p value. Each red point indicates a genus/specie with significant correlation ($p < 0.05$). GC, gastric cancer; scRNA-seq, single-cell RNA sequencing; and SAHMI, single-cell analysis of host–microbiome interactions

the 'plot_cell_trajectory' and 'plot_pseudotime_heatmap' functions of Monocle2.

Table 1 Detailed GEO database information used for the analysis of GC microbiome

GEO number	SRP number	Sample type	Description	Link
GSE134520	SRP215370	scRNA-seq	a scRNA-seq survey of 56,440 cells derived from 13 gastric antral mucosa biopsies from 9 patients with NAG, CAG, IM or EGC, using 10 × genomics	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE134520
GSE163558	SRP300226	scRNA-seq	scRNA-seq analysis of 42,968 cells derived from 10 fresh human tissue samples from 6 patients. 3 primary tumor and 1 adjacent non-tumoral samples and 6 metastases from different organs or tissues (liver, peritoneum, ovary, lymph node) were evaluated	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE163558
GSE112302	SRP136407	scRNA-seq	scRNA-seq analysis of 6 tumor tissues and 4 normal samples	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE112302
GSE150290	SRP261119	scRNA-seq	scRNA-seq analysis derived from 23 tumor samples and 29 adjacent normal tissues from 29 patients with stomach problems	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE150290
GSE191139	SRP351365	bulk RNA-seq	RNA profiling of primary gastric cancer, normal mucosae and metastasis by deep sequencing using Illumina NovaSeq	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE191139

NAG, non-atrophic gastritis; CAG, chronic atrophic gastritis; IM, intestinal metaplasia; and EGC, early gastric cancer

Association between *H. pylori* and host cells

To assess the association between *H. pylori* and host cells, we analyzed the distribution of *H. pylori*-positive cells across various cell types identified in scRNA-seq data from the GSE150290 dataset. The presence of *H. pylori* was inferred by detecting microbial reads linked to specific barcodes corresponding to individual cells. Meanwhile, we conducted a differential gene expression analysis to compare *H. pylori*-positive and *H. pylori*-negative cells within both tumor and adjacent normal samples. The parameters for identifying differentially expressed genes (DEGs) included average $\log_2$ fold change ≥ 0.5 and adjusted $p \leq 0.05$. These thresholds ensured that the identified DEGs were both biologically significant and statistically robust.

Functional enrichment analysis

To identify functional categories of genes, we employed the clusterProfiler package (v4.6.2) [24], which enabled us to determine Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways.

Stomach adenocarcinoma (STAD) data in The Cancer Genome Atlas (TCGA)

To validate the prognostic significance of the *H. pylori*-associated gene signature, we utilized RNA sequencing and clinical data from TCGA STAD cohort. Data were obtained from the GDC Data Portal (https://portal.gdc.cancer.gov/). We downloaded FPKM-normalized gene expression data and the corresponding clinical information. The FPKM values were converted to TPM values. The samples incorporated complete overall survival (OS) and vital status information. A gene signature score was calculated for each patient based on the average TPM value of the 13 genes identified in our scRNA-seq analysis. Patients were then stratified into high- and low-risk groups based on the median signature score for subsequent Kaplan–Meier survival analysis using the 'survival' and 'survminer' packages in R.

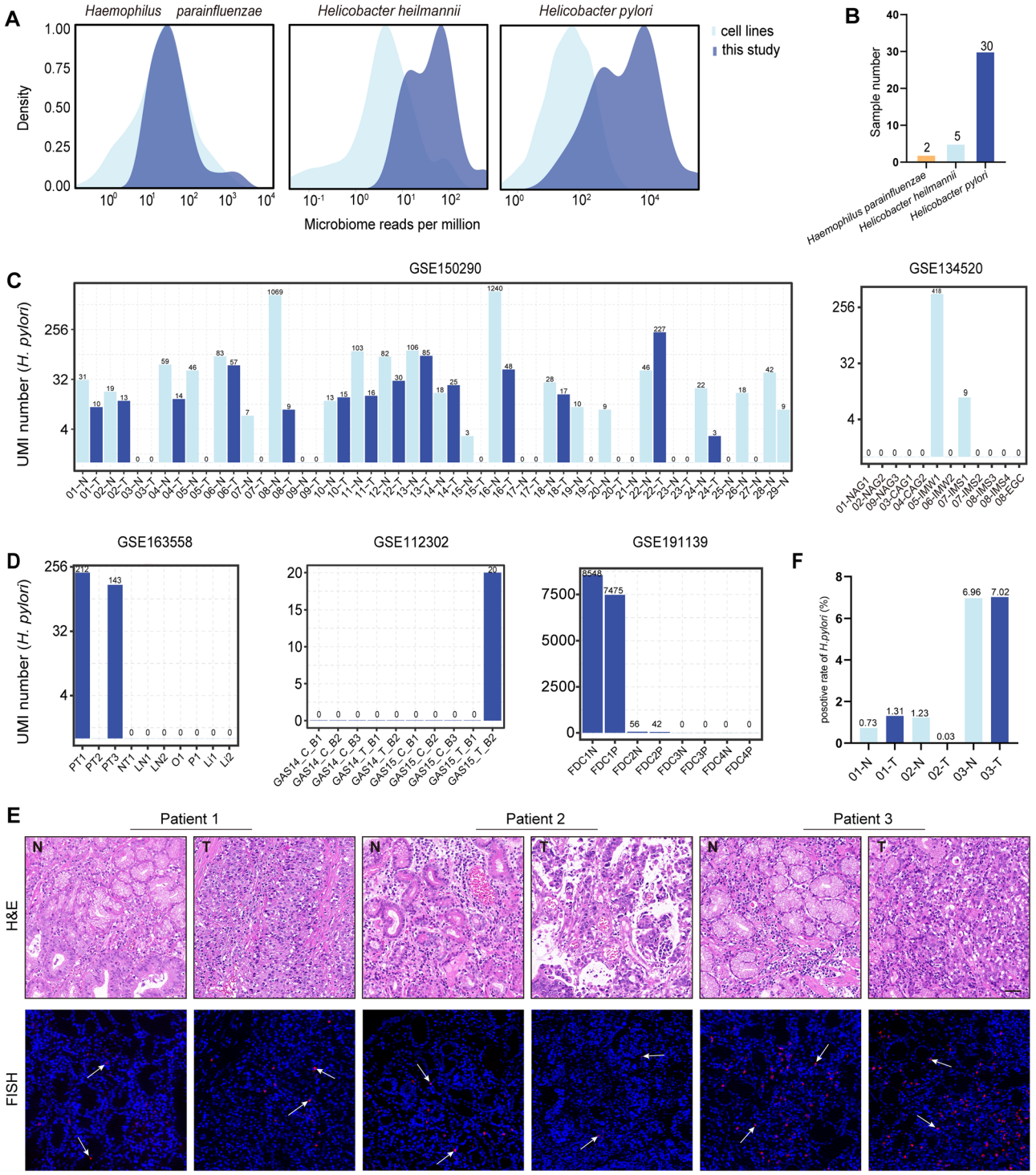
Statistical analysis

The data were presented as the mean ± SD. Statistical analyses were conducted using GraphPad Prism 8 software, R software, and Python, employing packages such as scanpy (v1.9.3) [25] and ggplot2 for data visualization. The pheatmap package (v1.0.12, https://cran.r-project.org/web/packages/pheatmap/index.html) was used to conduct clustering based on Euclidean distances. Student's *t* test was used to assess statistical significance, which was defined as $p < 0.05$.

Results

Identification and prevalence of intratumoral microbes in GC using scRNA-seq data

SAHMI [19], a computational pipeline, can systematically recover and denoise microbial signals in human clinical



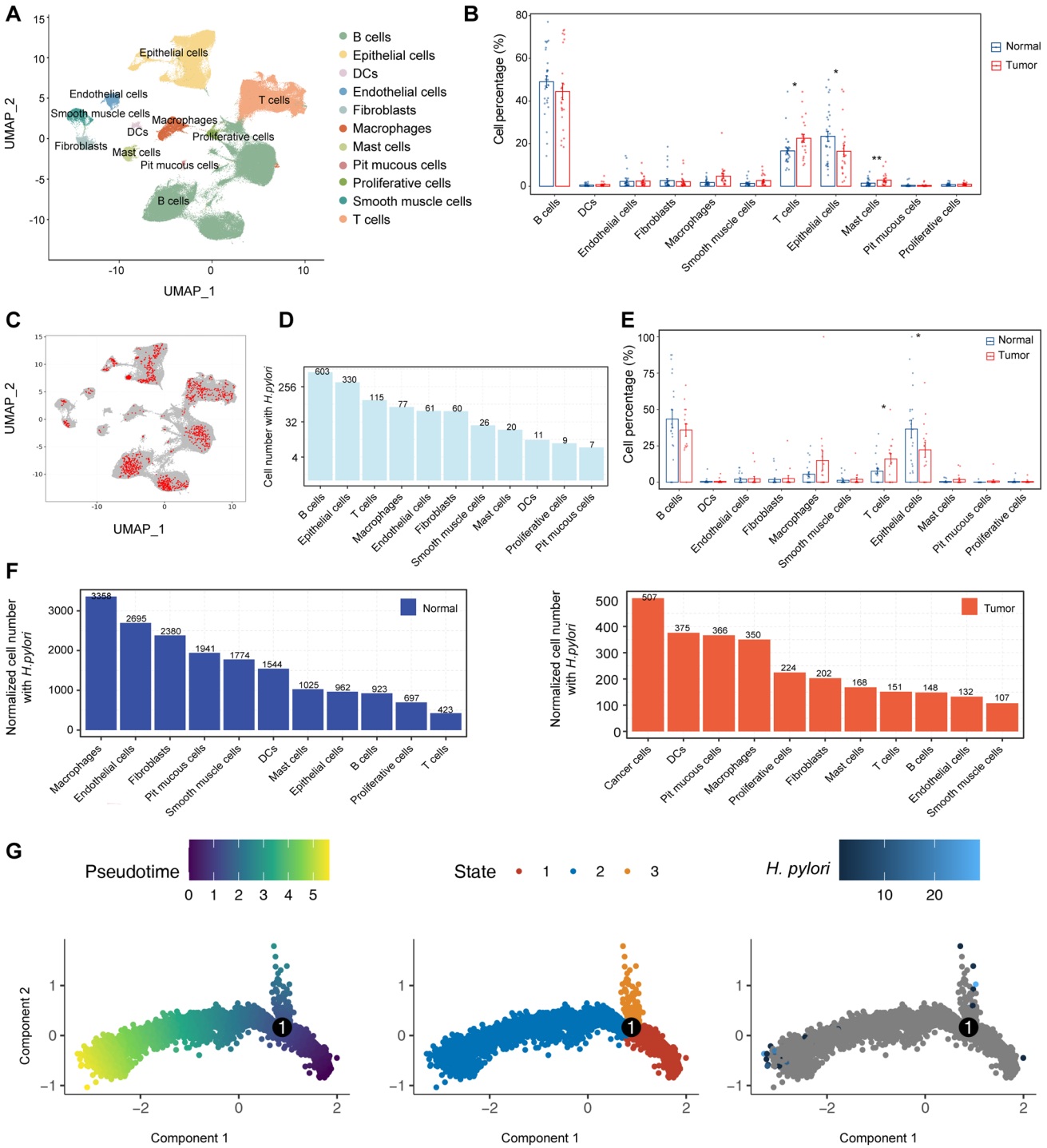
tissue and evaluate host–microbiome interactions at the single-cell level (Fig. 1A). By mining GC-related data from the Gene Expression Omnibus (GEO) database, we selected four GC scRNA-seq datasets (GSE134520, GSE163558, GSE112302, GSE150290) and one bulk RNA sequencing dataset (GSE191139) to analyze intratumoral bacteria at single-cell resolution. The scRNA-seq data included samples from human normal tissues, gastritis, early-stage cancer, primary GC, and metastatic GC to ensure that datasets were diverse and comprehensive. Detailed information about the datasets is provided in Table 1. We applied SAHMI to a large independent GC scRNA-seq cohort (GSE150290). Our analysis included 13,022 cells derived from 23 tumoral samples and 29 normal samples collected

◂**Fig. 2** Identification and prevalence of intratumoral *H. pylori* in GC samples. **A** Normalized count density plots comparing the reads per million for selected species detected in GC with the same species detected in thousands of cell line experiments, serving as a negative control. The three plots show species detected above the contamination and noise threshold. **B** Bar plot showing the sample number of the three detected bacterium passing the SAHMI denoising criteria in the GSE150290 dataset. **C**, **D** Bar plot showing the UMI number for *H. pylori* co-localized with host cells in each sample from five datasets, plots highlight the prevalence of *H. pylori* across different gastric samples. **E** Representative images of H&E and FISH for *H. pylori* (white arrows) in GC tumoral tissues and matched normal tissues. Scale bars, 50 μm. **F** Bar plot presenting the positive rate of *H. pylori* in tumoral and normal tissues from patients with GC via FISH. *H. pylori*, *Helicobacter pylori*; UMI, unique molecular identifier; N, normal sample; T, tumoral sample; NAG, non-atrophic gastritis; CAG, chronic atrophic gastritis; IMW, intestinal metaplasia with wild level; IMS, intestinal metaplasia with severe level; EGC, early gastric cancer; PT, primary GC sample; NT, non-tumor sample; LN, metastatic lymph nodes sample; O, ovary tumor sample; P, peritoneal tumor sample; Li, liver tumor sample; GAS*_C, gastric cancer; GAS*_T, gastric tissue; FDC*N, normal gastric mucosa; FDC*P, gastric adenocarcinoma tissue; H&E, hematoxylin and eosin staining; and FISH, fluorescence in situ hybridization

from 29 patients with stomach problems [26]. In total, 8599 million sequencing reads were generated from these gastric tissues. On average, Kraken2Uniq classified 95.62% ± 3.2% of the reads as human, 3.96% ± 3.2% as unclassified, and resolved 0.14% ± 0.04% of the reads were resolved to the bacterial level (Fig. 1B). In addition, we investigated the Spearman correlations of unique k-mer counts for microbial genera and species identified across GC samples from five datasets, demonstrating the reproducibility of the microbial profiles across different sequencing data (Fig. 1C). First, we screened 41 genera and 45 species of microbes with host-matched barcodes. We then applied the SAHMI k-mer correlation test, which required a significant correlation ($p < 0.05$) between the number of reads and the numbers of total and unique k-mers across barcodes, and we identified 14 genera and 12 species (Fig. 1D). Additionally, a significant correlation ($p < 0.05$) was also required between the number of reads and the numbers of total and unique k-mers across samples, and 12 genera and 12 species were finally screened (Fig. 1E).

We employed the SAHMI cell line quantile test to identify potential false taxonomic assignments and contaminant species. This test compared taxon frequencies with their distribution observed across thousands of sterile RNA-seq runs globally, encompassing more than 1000 human cell lines derived from both normal and diseased tissues. Four species, namely *Trichoderma breve*, *Aspergillus puulaauensis*, *Pararhizobium sp.* BT-229, and *Lymphocryptovirus humangamma* 4, were excluded because of the lack of reference datasets. Meanwhile, five taxa, namely *Cutibacterium acnes*, *Gordonia terrae*, *Malassezia restricta*, *Ralstonia pickettii*, and *Streptococcus mitis*, were detected at levels consistent with the cell line data (Fig. S1A). The remaining three significant taxa, *Haemophilus parainfluenzae*, *Helicobacter heilmannii*, and *H. pylori*, had microbial frequencies greater than the 95th percentile in the reference datasets (Fig. 2A). Hence, these three species passing the SAHMI denoising criteria were retained for subsequent analysis. However, *H. parainfluenzae* and *H. heilmannii* were detected in two and five samples in the GSE150290 dataset, respectively, whereas *H. pylori* was detected in 30 samples (Fig. 2B). Therefore, we finally focused on *H. pylori* for further analysis. Moreover, we calculated the UMI number of *H. pylori* detections in different samples using five datasets (Fig. 2C–D). Meanwhile, we also conducted additional examinations of the UMI number of *H. heilmannii* and *H. parainfluenzae* across five datasets (Fig. S1B–C), reinforcing the reliability of SAHMI in identifying these microorganisms. Among the five datasets examined, only GSE150290 and GSE134520 provided clinical information on *H. pylori* infection (Table S1). Notably, the *H. pylori* infection outcomes in the two analyzed datasets and clinical reports were remarkably similar.

In a subset of three patients with GC who had available tumoral tissues and matched non-tumoral tissues, we performed H&E staining to determine the histological morphology of their lesions (Fig. 2E). All three patients with *H. pylori*-positive exhibited features of gastric adenocarcinoma, and Patient 2 additionally demonstrated certain pathological characteristics of signet ring cell carcinoma. Then, we further verified the prevalence of *H. pylori* using FISH with specific labeled probes for hybridization with 16S rRNA of *H. pylori* in the gastric tissues (Fig. 2E). The tumors arose in the gastric body and antrum in Patient 1 and Patient 3, respectively, but in the cardia in Patient 2. The results demonstrated the presence of *H. pylori* in the cardia (upper), body (middle) and antrum (lower) of the stomach in both tumoral and matched normal tissues. Compared with the findings in Patient 1 and Patient 3, Patient 2 exhibited greater abundance of *H. pylori* in the normal tissues of the upper sites than in the tumoral tissues (Fig. 2F), and a similar distribution of *H. pylori* in different sites of the stomach was also reported in another paper [27]. *H. pylori* secretes several virulence factors, such as CagA and vacuolation cytotoxin A (VacA) protein, which enable persistence of the bacterium in the acidic environment of the human stomach. These proteins play key roles in the occurrence and development of GC [28]. Through IHC, we verified the presence of the CagA protein in human gastric tissues (Fig. S1D), demonstrating the colonization of epithelial cells by *H. pylori*. Thus, FISH and IHC verified the presence of *H. pylori* in GC tissues, and the results confirmed the accuracy and reliability of SAHMI in detecting tumor microbes.



H. pylori is significantly enriched in gastric tumor cells

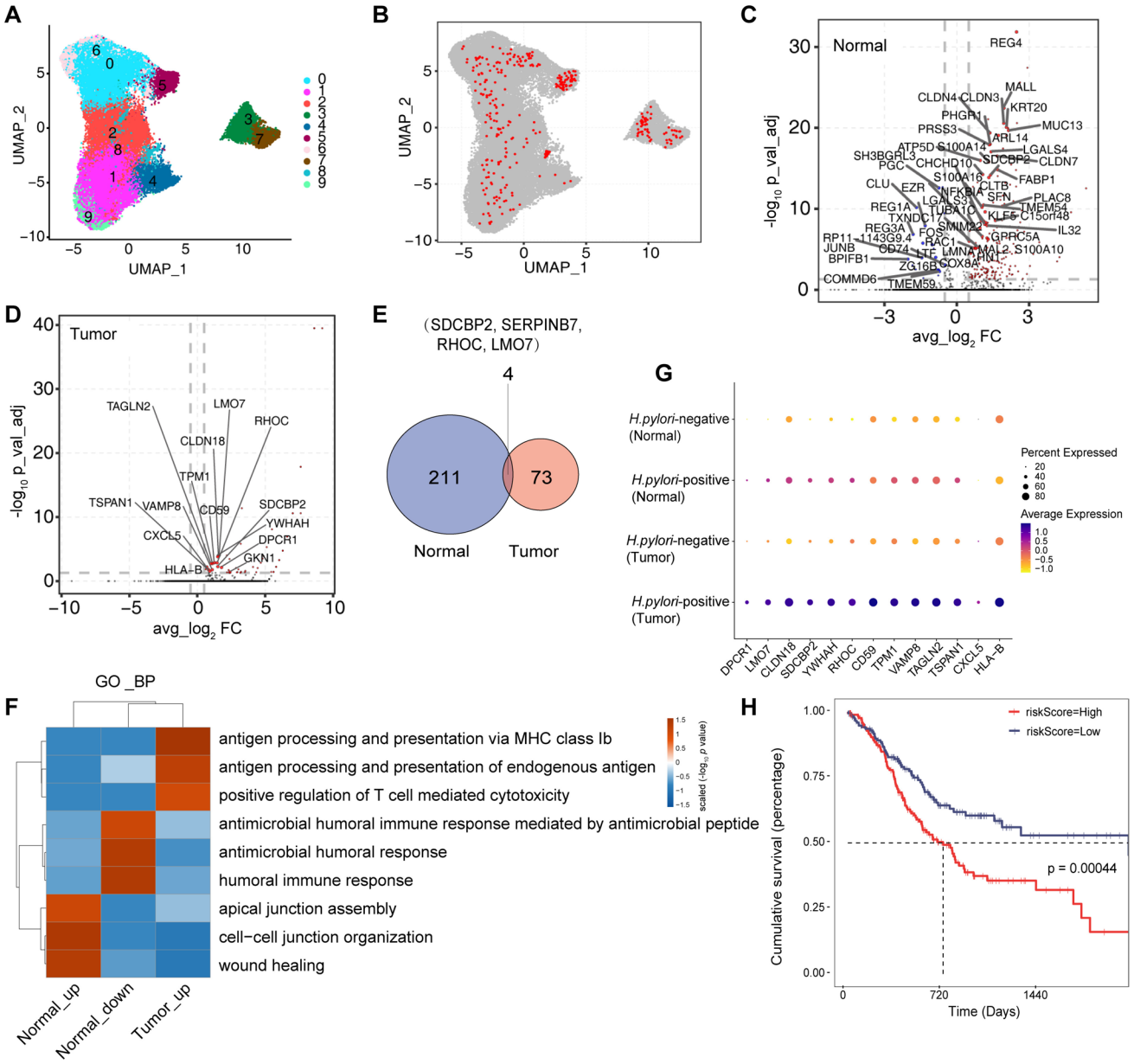
We performed scRNA-seq to determine distinct cell populations in the GSE150290 dataset. To reduce dimensionality and address batching effects, we performed PCA followed by Harmony integration (Fig. S2A-B). When we visualized the cell data using UMAP, 33 cell clusters were finally identified (Fig. S2C). Based on the expression of known markers, the atlas mainly comprised B cells, epithelial cells, dendritic cells, endothelial cells, fibroblasts, macrophages, mast cells, pit mucous cells, proliferative cells, smooth muscle cells, and T cells (Fig. S2D; Fig. 3A). In addition, B cells, T cells, and epithelial cells represented the three most abundant cell types in both tumoral and normal tissues (Fig. 3B). Upon visualizing the

◂**Fig. 3** Cell-associated *H. pylori* distributions in tumor and adjacent tissues. **A** UMAP plot displaying the composite single-cell transcriptomic profiles from tumor and normal samples. Different colors represent various annotated cell types, including epithelial cells, T cells, B cells, and others. **B** Bar plot comparing the relative proportions of each cell type between normal and tumor groups, highlighting the distribution of different cell types in both conditions. $^{*}p < 0.05$; $^{**}p < 0.01$. **C** UMAP plot focusing on cells containing *H. pylori*, highlighted in red, illustrating the distribution of *H. pylori*-positive cells within the broader cellular landscape. **D** Bar plot showing the total number of cells containing *H. pylori* across different cell types. **E** Bar plot comparing the relative proportions of each cell type infected by *H. pylori* between normal and tumor groups. $^{*}p < 0.05$. **F** Bar plot presenting the normalized cell number containing *H. pylori* in relation to each cell type in normal/tumor samples. This normalization allows for a clearer comparison of *H. pylori* prevalence among different cell types. **G** Pseudotime trajectory analysis of tumor-derived epithelial cells by Monocle2. Each dot represents a single epithelial cell, positioned according to two principal components (Component 1 and Component 2). The left panel is colored by pseudotime (from 0 to 5), the middle panel by inferred cellular 'state' (1, 2, or 3), and the right panel by the relative abundance of *H. pylori*. (Darker blue indicates higher abundance.) UMAP, Uniform Manifold Approximation and Projection; DC, dendritic cell

cell data using UMAP, the distribution of bacteria was not uniform across different cell types (Fig. 3C). Bacteria were discovered in all cell types, consistent with the imaging of tumors displaying intracellular bacteria in malignant and immune cells [29]. Among them, B cells, epithelial cells, and T cells had the highest *H. pylori* counts (Fig. 3D). Meanwhile, a significant difference emerged in the proportions of *H. pylori*-associated T and epithelial cells between adjacent normal and tumor tissues, suggesting that these cells have distinct roles in *H. pylori* colonization with the TME (Fig. 3E). Bacterium-associated cells generally clustered together within their respective cell types (Fig. 3C), demonstrating shared and broad gene expression changes compared with the findings in unassociated cells. By normalizing the number of cells infected by *H. pylori* by the total number of cells of each type, we found significant differences in the distribution of *H. pylori*-enriched cells between normal and tumor tissues. In normal tissues, *H. pylori* was significantly enriched in macrophages, whereas in tumor tissues, it was obviously enriched in cancer cells, indicating that co-clustering of bacterium-associated host cells enhanced the possibility of transcriptional changes in host cells in response to the presence of bacteria (Fig. 3F). Specifically, pseudotime analyses of tumor-derived epithelial cells revealed distinct cellular differentiation states and clear branching trajectories throughout tumor progression (Fig. 3G). Cells at the pseudotime origin (state 1) represented an initial differentiated state, whereas state 2 and state 3 diverged into two distinct evolutionary branches. Notably, *H. pylori* demonstrated preferential enrichment in cells belonging to state 2, suggesting a stronger association of this bacterium with a specific differentiation pathway within the tumor epithelium (Fig. 3G). Further analysis of gene expression dynamics confirmed distinct transcriptional profiles across these cellular states, reinforcing the potential role of *H. pylori* in influencing epithelial cell differentiation and consequently shaping tumor heterogeneity and progression (Fig. S2E). These data indicated that the presence of *H. pylori* could both reflect tumor progression and potentially contribute to its evolution. These findings provided important clues for further investigation of the relationship between *H. pylori* and GC.

Intratumoral *H. pylori* is associated with poor GC prognoses

Using UMAP plots visualizing clusters of epithelial cells derived from scRNA-seq data, we identified 10 cell cluster subgroups and demonstrated the distribution of epithelial cells with *H. pylori* (Fig. 4A-B). Initially, we investigated whether the presence of *H. pylori* was associated with epithelial cell-specific gene expression. We divided the samples into *H. pylori*-positive and *H. pylori*-negative groups for both normal and tumor tissues. Comparing *H. pylori*-positive with *H. pylori*-negative cells, 236 DEGs were identified in the normal group, including 215 upregulated and 21 downregulated genes (Fig. 4C). In the tumor group, 77 DEGs were identified, all of which were upregulated (Fig. 4D). Moreover, our analysis revealed that among upregulated genes in the tumor and normal groups, only four genes, namely *SDCBP2*, *SERPINB7*, *RHOC* and *LMO7*, were common to both groups, demonstrating that the regulation of host cell-specific genes by *H. pylori* differed significantly between the tumor and normal groups (Fig. 4E). Then, we performed GO enrichment analysis using the cell-type- and bacterium-specific DEGs to examine the biological processes associated with cell-associated bacteria (Fig. 4F). Specifically, upregulated DEGs were significantly enriched in the antigen processing and presentation-related pathways in the tumor group. However, in the normal group, upregulated and downregulated DEGs mainly involved the cell junction and humoral immune response, respectively. Meanwhile, we identified the 13 most significant DEGs (*DPCR1*, *LMO7*, *CLDN18*, *SDCBP2*, *YWHAH*, *RHOC*, *CD59*, *TPM1*, *VAMP8*, *TAGLN2*, *TSPAN1*, *CXCL5*, and *HLA-B*) between *H. pylori*-positive and *H. pylori*-negative tumor tissues. Notably, these genes exhibited significantly higher expression in the *H. pylori*-positive tumor group, whereas their expression in the other groups was comparatively lower (Fig. 4G). Finally, we investigated whether intratumoral *H. pylori* was correlated with OS. We used the 13 most significant DEGs to draw the Kaplan–Meier curves exhibiting differences in OS among patients with STAD in TCGA database, and the results indicated that cell-associated *H. pylori* was associated with significantly decreased OS in the tumor group ($p = 0.00044$; Fig. 4H). Thus, these host cell-specific genes regulated by *H.*



pylori might represent prognostic biomarkers or therapeutic targets for *H. pylori*-infected patients with GC. These findings suggested that intratumoral *H. pylori* has clinical relevance in a subset of individuals with GC.

Intratumoral *H. pylori* in immune cells is involved in GC progression

The extensive colocalization of bacterium-immune cells and their association with immune-related signaling indicated that intratumoral *H. pylori* can influence the immune response to GC. We investigated the existence of a functional difference in T cell subtypes with or without cell-associated

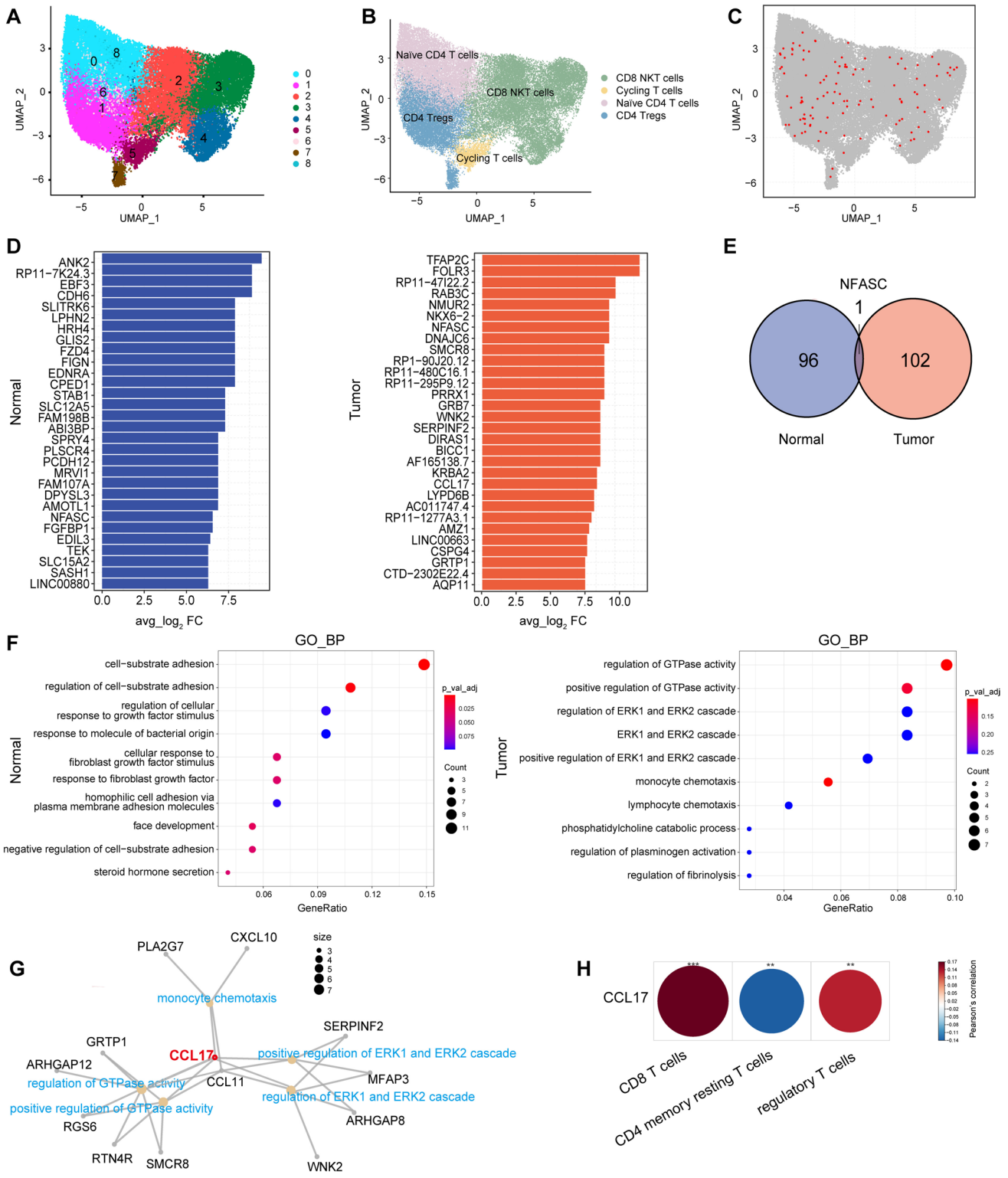
◂**Fig. 4** Comparative genomic analysis of tumor cells with and without Intratumoral *H. pylori*. **A** UMAP plot displaying the distribution of 10 epithelial cell clusters identified using the Seurat package. The colors represent different cell clusters within the epithelial cells. **B** UMAP plot highlighting epithelial cells containing *H. pylori* in red, showing the distribution of *H. pylori*-positive cells within these clusters. **C**, **D** Scatter plot depicting DEGs between *H. pylori*-positive and *H. pylori*-negative cells in normal/tumor samples. The x-axis represents the average $\log_2$ fold change (avg_$\log_2$FC), and the y-axis represents the -$\log_{10}$ adjusted *p* value (p_val_adj). Each red point indicates an upregulated gene (avg_$\log_2$FC ≥ 0.5 and p_val_adj ≤ 0.05), and each blue point indicates a downregulated gene (avg_$\log_2$FC ≤ -0.5 and p_val_adj ≤ 0.05). **E** Venn diagram illustrating the overlap of upregulated DEGs in *H. pylori*-positive cells from both normal and tumor samples, highlighting common genes such as *SDCBP2*, *SERPINB7*, *RHOC*, and *LMO7*. **F** GO enrichment analysis of BP associated with DEGs in *H. pylori*-positive cells. The heatmap shows the top 3 enriched GO terms for each group, with color intensity indicating the enrichment -$\log_{10}$ *p* value. **G** Dot plot displaying the relative average expression levels (z-score, column scaled) of 13 significantly upregulated genes in *H. pylori*-positive cells from tumor samples compared with other groups (diff pct > 0.15). **H** Kaplan–Meier survival curves showing overall survival differences in STAD patients from TCGA database based on the signature score of the 13 genes highlighted in (G). The *p* value indicates the statistical significance of survival differences between the high and low gene expression groups. DEG, differentially expressed gene; GO, gene ontology; BP, biological process; STAD, stomach adenocarcinoma; TCGA, The Cancer Genome Atlas; Normal_up, upregulated genes of normal samples; Normal_down, downregulated genes of normal samples; and Tumor_up, upregulated genes of tumor samples

H. pylori infection in tumor and normal tissues. By visualizing cell data using UMAP, nine T cell clusters were identified (Fig. 5A). Based on the expression of known markers, the atlas mainly comprised $CD8^+$ natural killer T cells, cycling T cells, naïve $CD4^+$ T cells, and $CD4^+$ regulatory T cells (Fig. 5B). *H. pylori* was relatively dispersed among T cells, suggesting that *H. pylori* infection alone is not sufficient to significantly influence the transcriptional levels of T cells on a broad scale (Fig. 5C). Compared with the findings in *H. pylori*-negative tissues, 97 or 103 DEGs were identified in *H. pylori*-infected T cells from normal and tumor tissues, respectively, and these DEGs were all upregulated (Fig. 5D). Moreover, two *H. pylori*-positive groups had only one common gene among the DEGs, suggesting that the effect of *H. pylori* on T cells differed between normal and tumor tissues (Fig. 5E). Pathways upregulated in *H. pylori*-positive T cells from the normal group were significantly enriched in cell-substrate adhesion signaling. However, in the tumor group, pathways upregulated in *H. pylori*-positive T cells included monocyte chemotaxis, GTPase activation, and ERK1 and ERK2 cascade regulation, which played important roles in anti-*H. pylori* infection responses in T cells (Fig. 5F). One study reported that C–C chemokine ligand 17 (CCL17) within GC was related to the increased abundance of $Foxp3^+$ Tregs, which could create an immunosuppressive microenvironment [30]. In our study, we found that CCL17 was an extremely critical hub gene among the pathways in the tumor group (Fig. 5G). We then used STAD data in TCGA to characterize the correlation between CCL17 and the proportion of immune cells, finding that CCL17 was positively correlated with the proportion of Tregs (Fig. 5H). These results implied that *H. pylori*-positive T cell may recruit Tregs to impair anti-tumor immunity.

In addition, we used the same approach to analyze the effects of cell-associated *H. pylori* on the biological processes of B cells (Fig. S3A–F) and macrophages (Fig. S4A–F). MHC class II protein in B cells is mainly responsible for presenting exogenous antigens to $CD4^+$ T cells, and *H. pylori* infection in normal stomach tissues can induce antibody production and attract immune cells to the infected area [31]. In our study, downregulated DEGs of B cells in the normal group were significantly enriched in MHC class II protein-related pathways, demonstrating that the downregulation of normal genes might activate MHC II molecules in B cells in response to *H. pylori* infection (Fig. S3F). Macrophages also play a vital role in GC progression induced by pathogenic *H. pylori* infection. In our study, upregulated DEGs were significantly enriched in the Notch signaling pathway in tumor tissues (Fig. S4F). Previous research reported that the Notch signaling ligand Jagged1, which is upregulated in activated macrophages in response to *H. pylori*, can enhance macrophage-mediated bactericidal activities [32]. Thus, Notch signaling pathway-related molecules might represent new therapeutic targets for *H. pylori* infection. In summary, immune cells in the TME played an important role in the response to *H. pylori*, and further studies are warranted to elucidate the immunopathogenesis of *H. pylori*-induced GC and identify potential therapeutic targets to combat this health burden globally.

Discussion

Recently, conceptual development in cancer biology has focused on the identification of microbes residing within cancer tissues, classified as the 'intratumor microbiota.' Compared with the rapid understanding of the role of the gut microbiota in cancer progression, our knowledge of the intratumor microbiota remains limited. The intratumor microbiota is an integral tumor component and live inhabitant of tumor tissues in various cancer types, including breast, bone, and pancreatic cancers. In addition, each tumor type has a distinct microbiome composition, and intratumoral bacteria are mostly located intracellularly and present in cancer and immune cells [29]. Intratumoral bacteria have a significantly lower abundance in cancer tissues than in the gastrointestinal tract, and the diversity of the microbial community is generally reduced in breast cancer tissues relative to that in normal tissue, indicating that tumors can create a



unique environment that selectively promotes the expansion of certain bacterial species [33]. Similarly, GC has an obviously lower microbial diversity than nonmalignant tissues, and *H. pylori* was significantly more enriched in GC tissues in both Memorial Sloan Kettering Cancer Center and TCGA cohorts in a retrospective analysis [5]. One of the most notable characteristics of intratumor bacteria is their ability to be recognized by the immune system, triggering specific immune responses. Intratumoral *Methylobacterium* has been significantly associated with poor prognoses in patients with

◂**Fig. 5** Immune microenvironment modulation by intratumoral *H. pylori* in T Cells. **A** UMAP plot showing the distribution of 9 T cell subclusters identified using the Seurat package. The colors indicate different T cell subclusters. **B** UMAP plot displaying the composite single-cell transcriptomic profiles of T cells from tumor and normal samples. Different colors represent various annotated T cell subtypes. **C** UMAP plot highlighting *H. pylori*-infected T cells, with red points representing *H. pylori*-positive T cells, indicating their distribution within the T cell population. **D** Bar plot displaying the top 30 upregulated genes in *H. pylori*-positive cells compared to *H. pylori*-negative cells in both normal and tumor samples. The genes are ranked by average $\log_2$ fold change (avg_$\log_2$FC). **E** Venn diagram showing the overlap of upregulated genes in *H. pylori*-positive T cells from normal and tumor samples, indicating shared and unique genes affected by *H. pylori* presence in immune microenvironment. **F** Dot plot representing the top 10 enriched GO analysis of BP terms associated with upregulated genes in T cells containing *H. pylori* versus those without in normal/tumor group. The size and color of the dots reflect the enrichment score and adjusted *p* value (p_val_adj), respectively. **G** Network diagram illustrating the top 5 enriched GO biological processes related to upregulated genes in *H. pylori*-positive T cells in tumor samples. This diagram highlights key processes such as monocyte chemotaxis, regulation of GTPase activity and ERK1/ERK2 signaling cascades. CCL17 is a hub gene involved in all these pathways. **H** Using CIBERSORT, the proportions of various T cell subtypes were estimated from bulk RNA-seq data in the TCGA stomach adenocarcinoma (STAD) cohort. The bubbles show the correlation between CCL17 expression and the fractions of each T cell subtype: color encodes the direction and strength of the correlation (red for positive; blue for negative), and bubble size reflects the Pearson correlation coefficient. Asterisks indicate statistically significant correlations. $^{**}p<0.01$; $^{***}p<0.001$. CCL17, C-C chemokine ligand 17

GC, and it has displayed an inverse correlation with the frequency of $CD8^+$ tissue-resident memory T cells within the TME, suggesting that intratumor bacteria play a crucial role in gastric carcinogenesis [6].

H. pylori secretes several virulence factors in the form of proteins that enable persistence in the acidic environment of the human stomach, which also leads to immune evasion. CagA and VacA proteins, as the most important virulence factors, play key roles in the incidence of GC [28]. *H. pylori* typically adheres to the gastric mucosa epithelial cells through its outer membrane proteins and injects CagA protein into host cells through a type IV secretion system, resulting in alterations in cytokine signaling and cell cycle regulation [13, 34]. In this study, we confirmed the presence of CagA in gastric tissues through IHC, indicating that *H. pylori* successfully colonized the epithelial cells of GC. CagA contributes to the onset and progression of GC through several physical and chemical mechanisms. Consequently, eradicating *H. pylori* might decrease the risk of GC, particularly in regions with a high prevalence of both GC and precancerous lesions [35]. Additionally, VacA protein affects the internal environment of gastric epithelial cells. It specifically establishes conditions that are conducive to the colonization and survival of *H. pylori* by modulating the intracellular energy state and ion balance [36, 37]. VacA has also been found to impair the function of host cells [38], thereby suppressing the therapeutic responses and creating a favorable ecological niche for *H. pylori* infection [17]. Most GC tumors are immunologically 'cold' including intestinal-type GC initiated by *H. pylori* infection. *H. pylori*-induced GC is unresponsive to chemotherapies or immune checkpoint inhibitors. One possible explanation is that *H. pylori* initiates immune escape processes for several years before GC develops. Briefly, when the human stomach is infected by *H. pylori*, both bone marrow-derived and tissue-resident immune cells are activated, and these cells express specific cell signatures that orchestrate neovascularization, establish an immune suppressive microenvironment, and contribute to the development of intestinal metaplasia [39]. In short, these evasion strategies led to immune system dysfunction, chronic inflammation, and prolonged exposure to carcinogenic stimuli, creating a microenvironment conducive to GC [17]. In this study, we used SAHMI to analyze tumor–microbiome interactions at single-cell resolution in a GC scRNA-seq cohort. Although a large number of microbial reads were initially detected, only three species, namely *H. parainfluenzae*, *H. heilmannii*, and *H. pylori*, passed all SAHMI denoising criteria. *H. pylori* was detected in most samples in this cohort, and it was chosen for further analysis. A previous study revealed the presence of *H. pylori* in GC tissues [27], and we also confirmed its presence by FISH and IHC.

In our study, *H. pylori* was predominantly enriched in macrophages in normal tissues, indicating that these cells play a significant role in the initial defense against *H. pylori* infection. However, in tumor tissues, *H. pylori* was primarily found in epithelial cells, which are the principal components of cancer cells. The regulation of epithelial cell-specific genes by *H. pylori* differed significantly between tumor and normal tissues, indicating different patterns of gene expression and transcriptional levels under the influence of cell-associated *H. pylori*. Meanwhile, we also identified one gene set including 13 genes with elevated expression in the tumor group, and this gene set was negatively correlated with patient prognosis. In subsequent research about cell-associated *H. pylori*, we plan to elucidate its impact on the progression of GC. T cells also play a crucial role in anti-*H. pylori* infection responses. In our research, pathways upregulated in *H. pylori*-infected T cells in tumors mainly included monocyte chemotaxis, GTPase activation, and ERK1 and ERK2 cascade regulation. It has been reported that GTPases activation is an important event in the coordination of immune responses, particularly in the activation of T cells [40]. The extracellular ERK1/2 signaling pathway plays a key role in various cell behaviors and significantly influences many aspects of the inflammatory response and the fate of

immune cells. In addition, ERK1/2 activation occurs in Toll-like receptor pathways, providing mechanistic insights into its recruitment, compartmentalization, and activation in cells of the innate immune system [41]. Meanwhile, downregulated DEGs in B cells in normal tissues were significantly enriched in MHC class II protein-related pathways, indicating these genes can activate MHC II molecules in B cells to product antibodies in response to *H. pylori* infection. The activity of TME components is regulated by several immunomodulatory factors including cytokines and chemokines. CCL17 is produced by DCs, endothelial cells, keratinocytes, and fibroblasts. CCL17 is highly expressed in the thymus, in which it plays an important role in T cell development and trafficking and the activation of mature T cells [42]. CCL17 can recruit immunosuppressive $CCR4^+$ Tregs in the TME, resulting in the functional suppression of $CD8^+$ effector T cells and immune escape by tumor cells [43].

Our findings provided evidence that intratumoral *H. pylori* can influence GC progression and dynamic alternations of the TME. The greatest limitation of our study was the lack of our own gastric tissues for single-cell sequencing to confirm the impact of *H. pylori* on GC. In future research, we plan to collect gastric samples from patients with GC for scRNA-seq to verify our findings and further study the effects of intratumoral bacteria on GC progression and the immune microenvironment at single-cell resolution. Although SAHMI can detect microbial nucleic acids captured in single-cell experiments, it also has some disadvantages [19]. First, it cannot replace gold-standard microbiology practices, because data processing filters can eliminate certain taxa, and non-specific RNA might also influence the rate of certain taxa and microbial profiles. Second, scRNA-seq data primarily capture host gene expression, and bacterial transcripts are often underrepresented or not captured effectively. Lastly, further research is necessary to determine their cellular localization, including whether they are freely expressed within the tissue microenvironment, within the intracellular region, or on the cell surface.

In general, SAHMI represents a significant advancement in the study of host–microbiome interactions and facilitates the examination of patterns derived from single-cell sequencing data, eliminating the need for additional experimental modifications. This approach generates testable hypotheses regarding host–microbiome relationships across multiple levels. We believe that in-depth research on the intratumoral microbiome will provide new strategies for GC treatment.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00262-025-04048-6.

Acknowledgements We are very grateful to Dr. Chao Cheng for his support and help in data analysis.

Author contributions JX and FW designed the research and supervised the research procedures. JX, QS, and JC participated in data organization and analysis. JX drafted the manuscript. JX, JY, and FW revised the manuscript and were responsible for project administration and funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding This study was supported by Institutional Foundation of The First Affiliated Hospital of Xi'an Jiaotong University (2024-QN-17), National Natural Science Foundation of China (81902680), and Cancer Precision Medical Science System and Service Platform Building-National Major Disease Multidisciplinary Collaborative Diagnosis and Treatment Capacity Building Program (QT264).

Data availability No datasets were generated or analyzed during the current study. Data will be made available on request.

Declarations

Conflict of interest The authors declare no competing interests. The authors have no relevant financial or non-financial interests to disclose.

Ethical approval This study about gastric cancer patient samples (XJTU1AF2023LSK-417) has received approval from the Ethics Committee of the First Affiliated Hospital of Xi'an Jiaotong University.

References

1. Bray F et al (2024) Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 74(3):229–263
2. Thrift AP, El-Serag HB (2020) Burden of Gastric Cancer. Clin Gastroenterol Hepatol 18(3):534–542
3. Xue C et al (2023) Current understanding of the intratumoral microbiome in various tumors. Cell Rep Med 4(1):100884
4. Fu A et al (2023) Emerging roles of intratumor microbiota in cancer metastasis. Trends Cell Biol 33(7):583–593
5. Abate M et al (2022) A novel microbiome signature in gastric cancer: a two independent cohort retrospective analysis. Ann Surg 276(4):605–615
6. Peng R et al (2022) Gastric microbiome alterations are associated with decreased CD8+ tissue-resident memory T cells in the tumor microenvironment of gastric cancer. Cancer Immunol Res 10(10):1224–1240

7. Yu G et al (2017) Molecular characterization of the human stomach microbiota in gastric cancer patients. Front Cell Infect Microbiol 7:302
8. Shao D et al (2019) Microbial characterization of esophageal squamous cell carcinoma and gastric cardia adenocarcinoma from a high-risk region of China. Cancer 125(22):3993–4002
9. Iarc L (1994) Schistosomes, liver flukes and *Helicobacter pylori*. IARC Monogr Eval Carcinog Risks Hum 61:1–241
10. Plummer M et al (2015) Global burden of gastric cancer attributable to *Helicobacter pylori*. Int J Cancer 136(2):487–490
11. Wroblewski LE, Peek RM Jr, Wilson KT (2010) *Helicobacter pylori* and gastric cancer: factors that modulate disease risk. Clin Microbiol Rev 23(4):713–739
12. Zhang X et al (2015) Analysis of the relationship between invasive capability of *Helicobacter pylori* and gastroduodenal diseases. J Med Microbiol 64(Pt 5):498–506
13. Huang Y et al (2016) Adhesion and invasion of gastric mucosa epithelial cells by *Helicobacter pylori*. Front Cell Infect Microbiol 6:159
14. Sit WY et al (2020) Cellular evasion strategies of *Helicobacter pylori* in regulating its intracellular fate. Semin Cell Dev Biol 101:59–67
15. Necchi V et al (2007) Intracellular, intercellular, and stromal invasion of gastric mucosa, preneoplastic lesions, and cancer by *Helicobacter pylori*. Gastroenterology 132(3):1009–1023
16. Ohba R, Iijima K (2016) Pathogenesis and risk factors for gastric cancer after *Helicobacter pylori* eradication. World J Gastrointest Oncol 8(9):663–672
17. Shirani M et al (2024) The immunopathogenesis of *Helicobacter pylori*-induced gastric cancer: a narrative review. Front Microbiol 15:1395403
18. Ghaddar B, Blaser MJ, De S (2023) Denoising sparse microbial signals from single-cell sequencing of mammalian host tissues. Nat Comput Sci 3(9):741–747
19. Ghaddar B et al (2022) Tumor microbiome links cellular programs and immunity in pancreatic cancer. Cancer Cell 40(10):1240-1253.e5
20. Butler A et al (2018) Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 36(5):411–420
21. Korsunsky I et al (2019) Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods 16(12):1289–1296
22. Ianevski A, Giri AK, Aittokallio T (2022) Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. Nat Commun 13(1):1246
23. Trapnell C et al (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol 32(4):381–386
24. Wu T et al (2021) clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innovation (Camb) 2(3):100141
25. Wolf FA, Angerer P, Theis FJ (2018) SCANPY: large-scale single-cell gene expression data analysis. Genome Biol 19(1):15
26. Kim J et al (2022) Single-cell analysis of gastric pre-cancerous and cancer lesions reveals cell lineage diversity and intratumoral heterogeneity. NPJ Precis Oncol 6(1):9
27. Lei L et al (2024) Distinct oral-associated gastric microbiota and *Helicobacter pylori* communities for spatial microbial heterogeneity in gastric cancer. mSystems 9(7):e0008924
28. Con SA et al (2009) Role of bacterial and genetic factors in gastric cancer in Costa Rica. World J Gastroenterol 15(2):211–218
29. Nejman D et al (2020) The human tumor microbiome is composed of tumor type-specific intracellular bacteria. Science 368(6494):973–980
30. Mizukami Y et al (2008) CCL17 and CCL22 chemokines within tumor microenvironment are related to accumulation of Foxp3+ regulatory T cells in gastric cancer. Int J Cancer 122(10):2286–2293
31. Reyes VE, Peniche AG (2019) *Helicobacter pylori* Deregulates T and B Cell Signaling to Trigger Immune Evasion. Curr Top Microbiol Immunol 421:229–265
32. Wen J et al (2021) Notch signaling ligand Jagged1 enhances macrophage-mediated response to *Helicobacter pylori*. Front Microbiol 12:692832
33. Fu A et al (2022) Tumor-resident intracellular microbiota promotes metastatic colonization in breast cancer. Cell 185(8):1356-1372.e26
34. Tegtmeyer N, Wessler S, Backert S (2011) Role of the cag-pathogenicity island encoded type IV secretion system in *Helicobacter pylori* pathogenesis. Febs j 278(8):1190–1202
35. Obayashi Y et al (2021) Risk factors for gastric cancer after the eradication of *Helicobacter pylori* evaluated based on the background gastric mucosa: a propensity score-matched case-control study. Intern Med 60(7):969–976
36. Cover TL, Blanke SR (2005) *Helicobacter pylori* VacA, a paradigm for toxin multifunctionality. Nat Rev Microbiol 3(4):320–332
37. Boquet P, Ricci V (2012) Intoxication strategy of *Helicobacter pylori* VacA toxin. Trends Microbiol 20(4):165–174
38. Son YS et al (2025) *Helicobacter pylori* VacA-induced mitochondrial damage in the gastric pit cells of the antrum and therapeutic rescue. Biomaterials 314:122842
39. Zavros Y, Merchant JL (2022) The immune microenvironment in gastric adenocarcinoma. Nat Rev Gastroenterol Hepatol 19(7):451–467
40. Pernis AB (2009) Rho GTPase-mediated pathways in mature CD4+ T cells. Autoimmun Rev 8(3):199–203
41. Lucas RM, Luo L, Stow JL (2022) ERK1/2 in immune signalling. Biochem Soc Trans 50(5):1341–1352
42. Imai T et al (1997) The T cell-directed CC chemokine TARC is a highly specific biological ligand for CC chemokine receptor 4. J Biol Chem 272(23):15036–15042
43. Ketcham JM, Marshall LA, Talay O (2018) CCR4 antagonists inhibit T(reg) trafficking into the tumor microenvironment. ACS Med Chem Lett 9(10):953–955