Discover Oncology

ANALYSIS Open Access

Helicobacter pylori–linked gene CFAP73 rewires epithelial programs and shapes the gastric cancer microenvironment

Haiwen Li[1†], Ran Wei[2†], Yuhan Liu[2†], Jiaju Wang[3†], Xinyi Wang[4], Li Chen[2], Bangjie Chen[3*] and Yong Yao[5*]

[†]Haiwen Li, Ran Wei, Yuhan Liu and Jiaju Wang have contributed equally equally to this work.

*Correspondence:
Bangjie Chen
bangjiechen@smail.nju.edu.cn
Yong Yao
yaoyong@ahmu.edu.cn
[1]Department of Gastroenterology, the Second Affiliated Hospital of Anhui Medical University, Hefei, China
[2]First Clinical Medical College, Anhui Medical University, Hefei, China
[3]Department of Oncology, Affiliated Drum Tower Hospital, Medical School of Nanjing University, Nanjing, China
[4]Department of Radiation Oncology, The First Affiliated Hospital of Anhui Medical University, Hefei, China
[5]College of Life Science, Anhui Medical University, Hefei, China

Abstract

Background Helicobacter pylori (HP) infection is the strongest environmental driver of gastric cancer, yet the epithelial programs that are progressively disrupted during infection and are associated with malignant transformation remain unclear.

Methods Trend-associated genes across HP-infection datasets (GSE60662, GSE60427) were identified using the Jonckheere–Terpstra test and intersected with survival-associated genes in TCGA-STAD. Random-forest modeling, multiple independent validation cohorts, functional analyses, single-cell RNA-seq (GSE249874), ligand–receptor inference, and spatial transcriptomics were integrated to define the biological role and microenvironmental impact of CFAP73.

Results CFAP73 emerged as the top tumor-protective gene progressively downregulated during HP infection and strongly predictive of favorable survival and cisplatin benefit. CFAP73 expression marked a tumor-suppressive epithelial state characterized by reduced proliferation, EMT inhibition, and activation of p53 and apoptotic pathways. Single-cell analysis showed CFAP73 predominantly in non-malignant epithelial cells, with HP infection driving its loss. CFAP73 + epithelial cells displayed increased LCN2 expression and attenuated oncogenic signaling. Microenvironmentally, CFAP73_high tumors were enriched for effector and Th17 T cells and showed reduced exhausted T cells, Tregs, and pro-tumorigenic CAF states (iCAF, apCAF). Ligand–receptor modeling revealed that CFAP73 + epithelial cells received weaker WNT, TGFβ, and PDGF signals from CAFs but stronger cytotoxic interactions from T cells. Spatial transcriptomics confirmed spatial segregation of CFAP73 + epithelial regions from proliferative, hypoxic, immune-checkpoint–active niches.

Conclusions CFAP73 is a previously unrecognized epithelial tumor suppressor suppressed early during HP infection. Loss of CFAP73 may contribute to epithelial malignant reprogramming and reshapes fibroblast and T-cell states toward an immunosuppressive, pro-tumor microenvironment. CFAP73 represents a promising biomarker linking HP-driven mucosal injury to gastric cancer initiation, progression, and therapeutic response.

Keywords Helicobacter pylori, Gastric cancer, CFAP73, Single-cell RNA-seq

1 Introduction

Gastric cancer (GC) remains one of the leading causes of cancer-related mortality worldwide and is frequently diagnosed at advanced stages when therapeutic options are limited and prognosis is poor [1, 2]. Despite recent advances in molecular stratification and immunotherapy, the overall survival of patients with GC has improved only modestly [3]. A major barrier to early interception lies in the incomplete understanding of the molecular events that underlie the transition from chronic mucosal injury to malignant transformation.

Among environmental factors, persistent infection with Helicobacter pylori (HP) is the most significant etiological driver of GC [4, 5]. HP colonization induces chronic inflammation, epithelial injury, and remodeling of the gastric niche, forming a multistep cascade that progresses from non-atrophic gastritis to intestinal metaplasia, dysplasia, and carcinoma. Although numerous studies have identified HP-associated alterations in inflammatory mediators, DNA damage responses, epigenetic regulation, and oncogenic signaling pathways [6–9], the specific epithelial genes whose progressive dysregulation during infection predisposes to tumor development remain poorly defined [10]. Most research has focused on canonical regulators—such as IL1B, CXCL8, CDH1, and TP53—while early epithelial determinants that integrate infection-induced stress with microenvironmental remodeling have been largely overlooked [11].

CFAP73 (Cilia and Flagella Associated Protein 73) is a poorly characterized gene with emerging links to ciliary biology and epithelial integrity, but its role in gastric mucosal homeostasis and carcinogenesis has not been investigated. Given that HP infection disrupts epithelial polarity, ciliary signaling, and barrier function—processes increasingly implicated in cancer susceptibility—genes such as CFAP73 may serve as missing molecular intermediaries that couple chronic infection to tumor initiation. However, no study to date has evaluated CFAP73 expression dynamics during HP infection, its prognostic relevance in gastric cancer, or its impact on the cellular ecosystem of the tumor microenvironment (TME).

The GC TME plays a decisive role in tumor evolution, therapeutic resistance, and patient survival. Recent single-cell and spatial transcriptomic studies have revealed complex interactions among epithelial cells, T-cell subsets, and cancer-associated fibroblasts (CAFs), including inflammatory CAFs (iCAFs) and antigen-presenting CAFs (apCAFs), which promote immunosuppression, extracellular matrix remodeling, and tumor progression. However, it remains unclear whether HP-driven epithelial alterations initiate specific microenvironmental changes that favor CAF activation, T-cell exhaustion, and immune evasion.

In this study, we integrated HP-infection transcriptomic datasets, TCGA survival analyses, multiple external cohorts, single-cell RNA sequencing, ligand–receptor modeling, and spatial transcriptomics to identify epithelial genes linking HP infection to malignant transformation. We discovered that CFAP73 is progressively downregulated during HP infection and defines a tumor-suppressive epithelial program associated with favorable clinical outcomes, reduced proliferation and EMT signaling, and activation of p53-mediated pathways. Moreover, CFAP73 expression stratifies distinct T-cell and CAF states, revealing a microenvironmental reprogramming axis through which HP-driven loss of CFAP73 fosters immunosuppressive and pro-tumor niches. Our findings identify CFAP73 as a previously unrecognized regulator of epithelial homeostasis and TME

architecture, bridging chronic infection to gastric carcinogenesis and offering a potential biomarker for early interception and therapeutic response prediction.

2 Results

2.1 Identification of HP-associated trend genes reveals CFAP73 as a key tumor-protective candidate

To identify genes progressively altered during Helicobacter pylori (HP) infection and associated with gastric cancer prognosis, we first applied the Jonckheere–Terpstra trend test to two independent HP-infection transcriptomic datasets (GSE60662 and GSE60427). This analysis yielded 197 genes exhibiting consistent downregulation and 5 genes showing progressive upregulation during the course of infection. Intersecting these trend-associated genes with survival-related genes identified by Cox regression and Kaplan–Meier analyses in TCGA-STAD revealed nine HP-suppressed genes associated with favorable overall survival, highlighting their potential tumor-protective roles (Fig. 1A). Random-forest modeling performed on this nine-gene signature further identified CFAP73 as the most informative predictor of patient outcome, ranking highest in variable importance (Fig. 1B). Kaplan–Meier analyses across TCGA-STAD and two external validation cohorts (GSE62254 and GSE15459) consistently demonstrated that high CFAP73 expression was strongly associated with improved overall survival ($P < 0.05$ for TCGA-STAD and GSE15459, with a similar trend in GSE62254), supporting the robustness of its prognostic value (Fig. 1C). In addition, our analysis within the GEPIA3 framework demonstrated that high CFAP73 expression was significantly associated with improved prognosis specifically in patients receiving cisplatin therapy, whereas this protective effect was less pronounced in the untreated group, suggesting a potential predictive value for chemotherapy benefit (Fig. 1D), suggesting that CFAP73 may influence chemosensitivity.

We next evaluated the broader oncologic significance of CFAP73 across cancer types. Pan-cancer survival analysis revealed that CFAP73 tended to correlate with favorable prognosis in multiple malignancies, including stomach adenocarcinoma, breast cancers, and others, while showing limited associations with adverse outcomes in any cancer subtype (Figs. 1E and Supplementary Figure S1A). Furthermore, disease-free survival analysis in GSE62254 showed a similar favorable trend for CFAP73-high tumors (Supplementary Figure S1B), indicating that CFAP73 may also be associated with reduced recurrence risk. To explore the clinical and molecular correlates of CFAP73 expression, we examined its levels across established gastric cancer classification systems. CFAP73 was significantly lower in diffuse-type tumors, high-grade or late-stage disease, and genomically unstable or proliferative molecular subtypes, whereas higher expression was observed in intestinal-type tumors and less aggressive subgroups (Fig. 1F).

Together, these findings identify CFAP73 as a progressively HP-suppressed, clinically favorable biomarker linked to less aggressive gastric cancer phenotypes and enhanced therapeutic response.

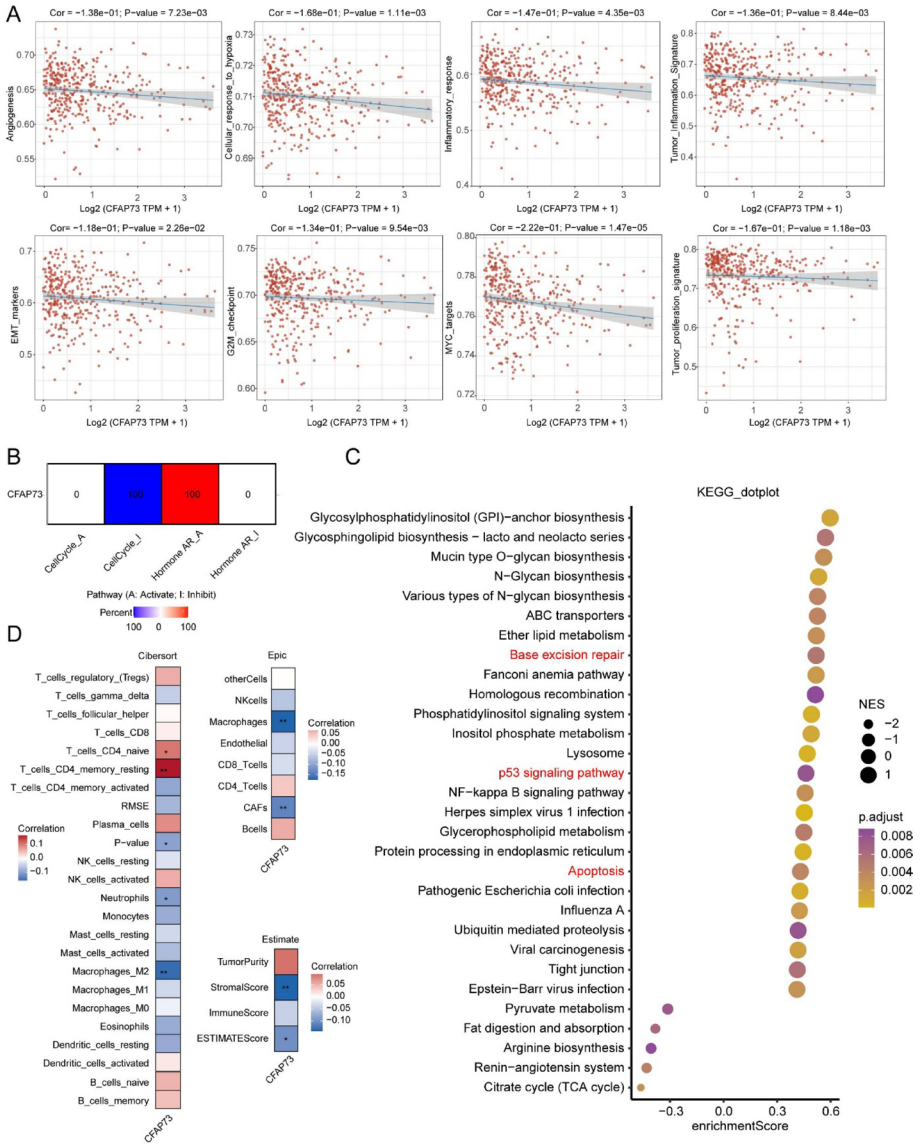

Fig. 1 Identification of CFAP73 as a progressively HP-suppressed and prognostically favorable epithelial gene. **A** Overlap between genes progressively altered during Helicobacter pylori infection (identified by the Jonckheere–Terpstra trend test in GSE60662 and GSE60427; control (uninfected) to mild/severe gastritis and intestinal metaplasia) and TCGA-STAD survival-associated genes (univariate Cox + Kaplan–Meier). Nine HP-decreased genes were associated with favorable overall survival. **B** Random-forest variable-importance plot ranking the nine HP-suppressed favorable-prognosis genes; CFAP73 showed the highest contribution to survival prediction. **C** Kaplan–Meier survival curves for CFAP73 expression in TCGA-STAD, GSE62254, and GSE15459 cohorts; high CFAP73 expression consistently predicted improved overall survival. Log-rank test was used. **D** Interaction between CFAP73 expression and cisplatin treatment. In cisplatin-treated TCGA tumors, CFAP73-low cases showed significantly poorer survival, while CFAP73-high cases benefited from cisplatin. NS = not significant. **E** Pan-cancer Cox regression analysis showing hazard ratios across TCGA cancer types. CFAP73 predominantly behaved as a protective gene across malignancies. **F** Associations between CFAP73 expression and gastric cancer molecular classifications. CFAP73 was lower in diffuse-type, high-grade, late-stage, and proliferative/unstable molecular subtypes, but higher in intestinal-type and less aggressive subgroups. Wilcoxon or Kruskal–Wallis tests were used

3 CFAP73 suppresses oncogenic epithelial programs and associates with favorable immune activity

To investigate the functional relevance of CFAP73 in gastric cancer biology, we first examined its relationship with hallmark oncogenic processes. CFAP73 expression

showed significant negative correlations with angiogenesis, inflammatory response, tumor-associated inflammation, EMT, cell adhesion signaling, and proliferation signatures (Fig. 2A), suggesting broad suppression of tumor-promoting epithelial programs. Conversely, pathway activity analysis demonstrated that CFAP73 predominantly activated tumor-suppressive modules while inhibiting cancer-promoting modules (Fig. 2B).

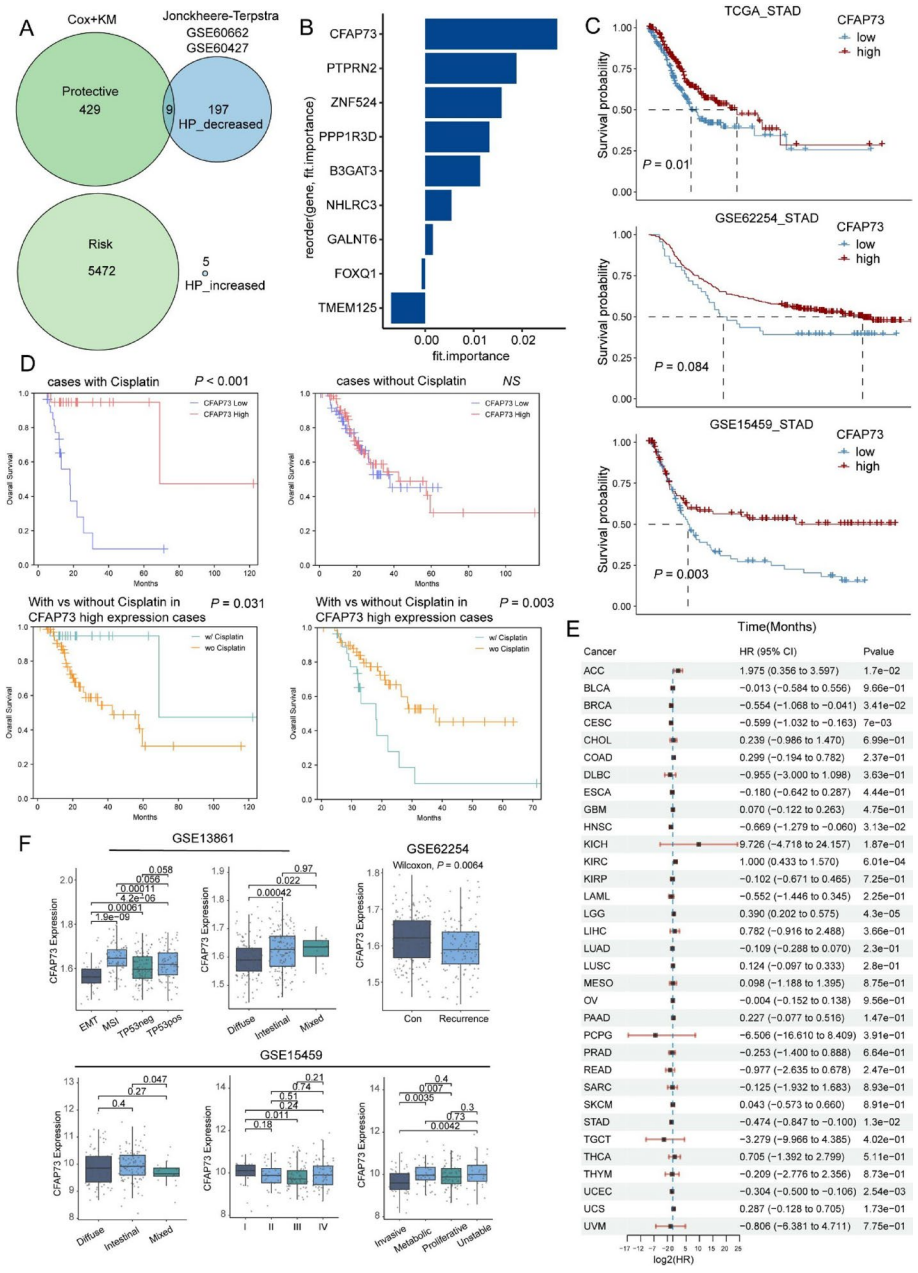

Fig. 2 CFAP73 suppresses oncogenic epithelial programs and aligns with favorable immune infiltration. **A** Scatter plots showing negative correlations between CFAP73 expression and angiogenesis, inflammatory response, EMT, cell adhesion, proliferation, and Myc-related signatures in TCGA-STAD (Spearman correlation). **B** Activation/inhibition heatmap showing CFAP73-associated pathway activity; CFAP73 predominantly activated tumor-suppressive pathways and inhibited oncogenic programs. **C** KEGG GSEA showing enrichment of apoptosis, p53 signaling, and DNA repair pathways in CFAP73-high tumors, versus metabolic/glycosylation pathways in CFAP73-low tumors. **D** Immune deconvolution across multiple algorithms (CIBERSORT, quanTIseq, EPIC, MCP-counter, ESTIMATE). CFAP73 correlated positively with effector T-cell subsets (CD8+, CD4 memory) and inversely with Tregs, M2 macrophages, NK resting cells, and CAF abundance. All correlations shown are Spearman coefficients

Gene-set enrichment analysis further confirmed that high CFAP73 expression was associated with enrichment of p53 signaling, apoptosis, base-excision repair, and antiviral pathways, whereas tumors with low CFAP73 expression showed activation of metabolic and glycosylation pathways linked to cancer aggressiveness (Fig. 2C). Given the strong functional signatures, we next assessed CFAP73 in the context of immune infiltration. Immune deconvolution revealed that CFAP73 expression correlated positively with multiple effector T-cell subsets—including CD8 T cells and CD4 memory T cells—while showing negative associations with Tregs, NK-cell resting states, tumor-promoting macrophage populations, and cancer-associated fibroblasts (CAFs) (Fig. 2D). CFAP73 was also associated with lower stromal content and higher immune activation, suggesting a more immunologically favorable tumor microenvironment.

To further validate these associations using external functional datasets, we interrogated the CITE CRISPR screening resource. Notably, CFAP73 sgRNAs were markedly enriched in T-cell–tumor co-culture screens (Supplementary Figure S2A), indicating that loss of CFAP73 conferred a selective disadvantage under T-cell–mediated cytotoxic pressure. This is consistent with our observation that CFAP73-high tumors are infiltrated by more active immune subsets. Across independent immunotherapy, TCGA, CPTAC, ICGC, TARGET, PRECOG, and clinical datasets, CFAP73 expression showed a consistent negative association with cancer risk and T-cell dysfunction modules (Supplementary Figure S2B), suggesting that CFAP73 may protect epithelial cells from acquiring immune-evasive phenotypes.

Collectively, these analyses reveal that CFAP73 functions as a suppressor of proliferation and EMT programs, activates p53-driven tumor-protective pathways, and aligns with an immunologically active tumor microenvironment, further supporting its role as a tumor-suppressive epithelial regulator.

4 CFAP73 defines a tumor-suppressive epithelial state that is diminished by HP infection

To elucidate the cellular context of CFAP73 expression, we analyzed gastric single-cell RNA-seq data (GSE249874). Uniform manifold approximation and projection (UMAP) revealed major cellular compartments, including epithelial cells, fibroblasts, T cells, endothelial cells, and myeloid subsets (Fig. 3A). Stratification by HP status demonstrated marked alterations in epithelial composition during HP-induced disease progression (Fig. 3B). Marker-based cell annotation confirmed that CFAP73 expression was highly specific to epithelial cells, with minimal expression in fibroblasts, endothelial cells, or immune populations (Fig. 3C–D). Copy-number variation analysis identified malignant vs. non-malignant epithelial subsets, showing that CFAP73 expression was enriched in non-malignant epithelial cells and progressively lost in malignant epithelial populations (Fig. 3E–F).

Analysis of epithelial subgroups across disease states demonstrated that CFAP73 was most highly expressed in non-atrophic and intestinal metaplasia epithelium, followed by HP-negative normal mucosa, and sharply reduced in HP-positive gastric cancer (Fig. 3G). Differential expression analysis between CFAP73 + and CFAP73 – epithelial cells identified LCN2 as the most significantly upregulated gene in CFAP73 + tumor-adjacent epithelial cells [12] (Fig. 3H), consistent with the known tumor-suppressive and anti-metastatic roles of LCN2 in gastric tissue biology.

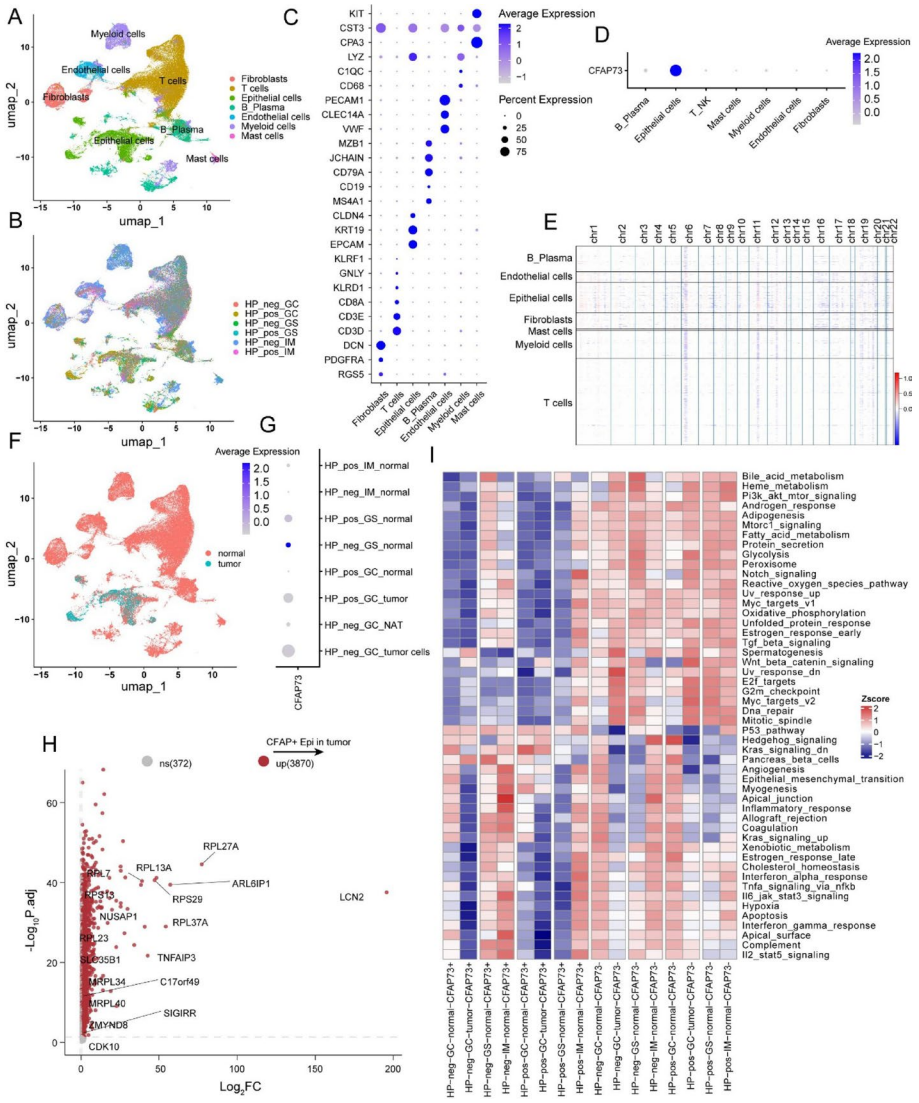

Fig. 3 CFAP73 defines a tumor-suppressive epithelial state diminished by HP infection. **A** UMAP of gastric scRNA-seq (GSE249874) showing major cell types (94542 cells). **B** Distribution of cell types across HP-negative and HP-positive conditions. **C**–**D** Dot plots showing high epithelial specificity of CFAP73; minimal expression in stromal or immune compartments. **E** inferCNV analysis identifying malignant vs. non-malignant epithelial cells.**F** UMAP showing segregation of malignant and non-malignant epithelial subsets; CFAP73 enriched in non-malignant cells. **G** CFAP73 expression across HP-negative normal, IM, NAG, and HP-positive GC epithelium. **H** Volcano plot of differentially expressed genes between CFAP73 + and CFAP73 − epithelial cells; LCN2 is most upregulated in CFAP73 + epithelium. **I** Heatmap of pathway activity showing that CFAP73 + epithelial cells are enriched for oxidative phosphorylation, apoptosis, and p53 signaling, and depleted of EMT, hypoxia, and cell-cycle programs

To characterize the functional identity of CFAP73 + epithelial populations, we performed pathway enrichment analysis. CFAP73 + epithelial cells showed significant enrichment of oxidative phosphorylation, p53 signaling, apoptosis, Myc targets downregulation, and reduced inflammatory and stress-response signatures (Fig. 3I). GO and KEGG analyses of CFAP73 + enriched genes highlighted strong enrichment in mitochondrial translation, ribosomal biogenesis, RNA splicing, proteostasis, and nucleocytoplasmic transport (Supplementary Figure S3A–B), suggesting a metabolically stable, differentiated epithelial state. Consistent with this, gene-set enrichment analysis revealed that CFAP73 + epithelial cells were positively associated with extrinsic apoptotic

signaling, responses to bacterial stimuli, and IL-6–related pathways—signatures linked to epithelial integrity and anti-tumor immunity—while suppressing pathways that promote EMT and proliferation (Supplementary Figure S3C). Direct quantification confirmed that CFAP73 + epithelial cells displayed significantly lower EMT scores and reduced expression of proliferation markers such as MKI67 and PCNA (Supplementary Figure S3D–E).

Together, these findings indicate that CFAP73 marks a non-malignant, differentiation-associated epithelial program that restrains proliferative and EMT-driven malignant transformation. Its progressive loss in HP-associated gastric pathology suggests that HP infection may co-occurs with carcinogenesis by eroding this tumor-suppressive epithelial state.

5 Loss of CFAP73 reshapes epithelial–immune–fibroblast interactions toward a pro-tumor microenvironment

Because CFAP73 expression was confined to non-malignant epithelial cells and associated with anti-proliferative epithelial states, we next investigated how epithelial CFAP73 levels influence immune and stromal compositions. Stratification of gastric cancer epithelial cells into CFAP73_high and CFAP73_low groups revealed that CFAP73 expression decreased markedly in HP-positive tumor epithelium across all patients (Fig. 4A).

High-resolution clustering of tumor-infiltrating T cells identified multiple functional subsets, including cytotoxic CD8 T cells, Th17 cells, tissue-resident memory T cells (Trm), exhausted T cells (Tex), follicular helper T cells (Tfh), and regulatory T cells (Tregs) (Fig. 4B). Comparison between CFAP73_high and CFAP73_low tumors showed striking compositional differences. CFAP73_high tumors were enriched for cytotoxic effector T cells, Th17 cells, and Trm subsets, whereas CFAP73_low tumors accumulated Tregs, Tex cells, and pre-exhausted T cell states (Fig. 4B–D). These results suggest that epithelial CFAP73 supports a more active anti-tumor T-cell landscape, while its loss is associated with T-cell dysfunction.

We next examined cancer-associated fibroblast (CAF) heterogeneity, identifying inflammatory CAFs (iCAF), myofibroblastic CAFs (mCAF), and antigen-presenting CAFs (apCAF) (Fig. 4E–F). CFAP73 expression strongly influenced CAF composition: CFAP73_low tumors showed substantial expansion of iCAF and apCAF populations, whereas CFAP73_high tumors contained predominantly mCAFs (Fig. 4G). Pathway analysis revealed that iCAFs were enriched in extracellular matrix remodeling, PI3K–Akt signaling, complement/coagulation cascades, and cytokine signaling (Supplementary Figure S4A–B), while apCAFs were enriched for mitochondrial metabolism, oxidative phosphorylation, and antigen-processing pathways (Supplementary Figure S4C–D). Heatmap analyses showed that iCAFs exhibited high levels of inflammatory programs, KRAS signaling, hypoxia, and TGF-β signaling—hallmarks of pro-tumor fibroblast phenotypes (Supplementary Figure S4E). Importantly, Kaplan–Meier analysis confirmed that high iCAF abundance predicted significantly poorer overall survival (Fig. 4H), consistent with their established tumor-promoting functions.

Together, these results demonstrate that epithelial CFAP73 expression shapes both immune and fibroblast landscapes. High CFAP73 aligns with active cytotoxic immunity and repressed CAF activity, whereas CFAP73 loss—driven by HP infection co-occurs with T-cell exhaustion and expansion of tumor-supportive iCAF and apCAF states.

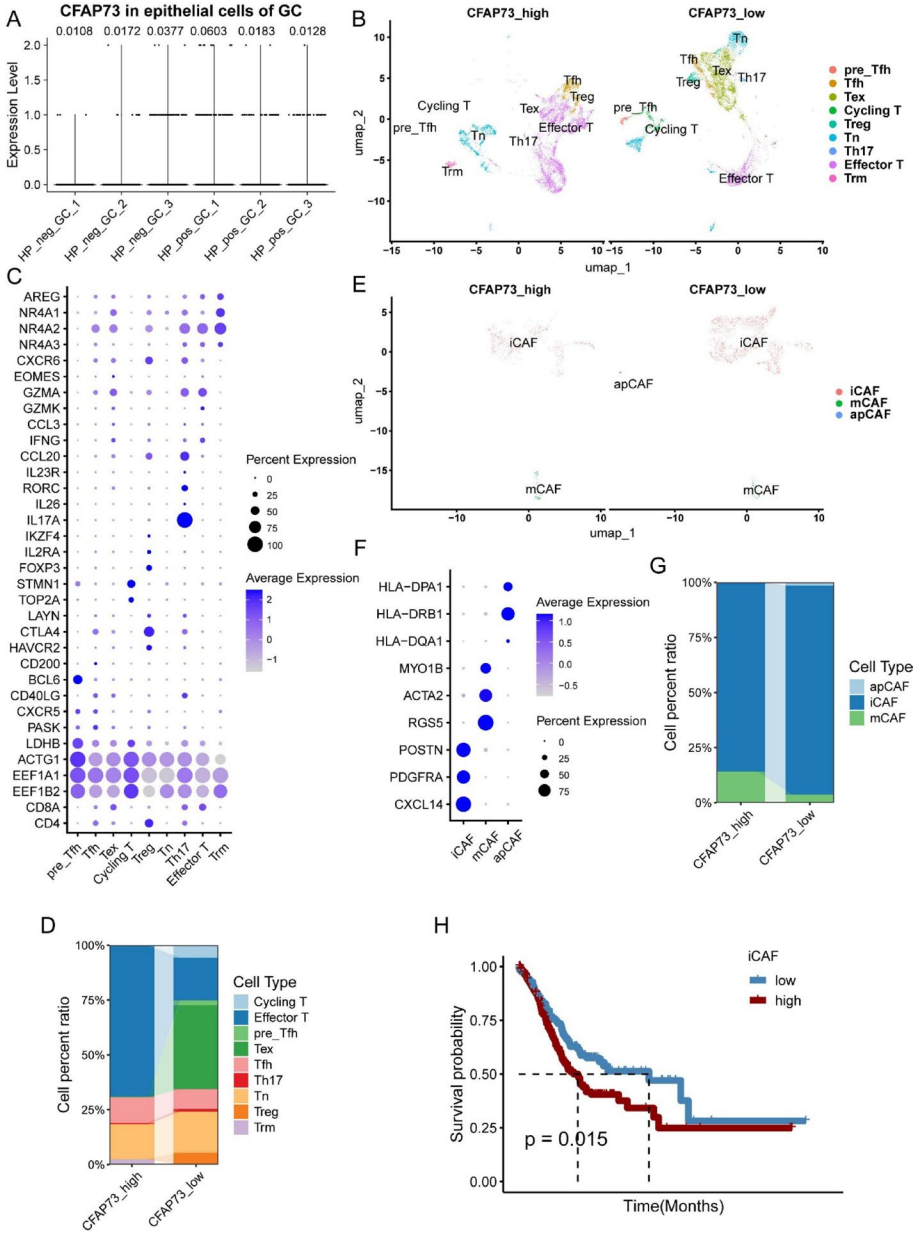

Fig. 4 Loss of CFAP73 reshapes T-cell and CAF landscapes toward immunosuppressive, pro-tumor states. **A** CFAP73 expression across epithelial cells of HP-positive and HP-negative gastric cancers. **B**–**D** UMAP and proportion plots of tumor-infiltrating T-cell subsets. CFAP73-high tumors show enrichment of cytotoxic CD8 + T cells, Th17, and Trm subsets; CFAP73-low tumors accumulate Tregs, Tex, and pre-exhausted T cells. **E**–**G** UMAP and proportion plots of CAF subtypes. CFAP73-low tumors display expanded iCAF and apCAF populations, whereas CFAP73-high tumors contain predominantly mCAFs. **H** Kaplan–Meier curve demonstrating that high iCAF abundance predicts significantly shorter overall survival

These findings indicate that CFAP73 serves as a critical epithelial regulator of microenvironmental homeostasis, and its loss initiates a coordinated shift toward an immunosuppressive, pro-tumor TME.

6 CFAP73 + epithelial cells receive stronger cytotoxic T-cell engagement and weaker CAF-derived oncogenic signaling

To determine how epithelial CFAP73 status influences microenvironmental communication, we performed ligand–receptor interaction analysis across epithelial, fibroblast, and T-cell compartments. Global interaction heatmaps revealed that CFAP73 + epithelial cells engaged in markedly fewer interactions with iCAFs and apCAFs, whereas CFAP73 – epithelial cells displayed the strongest epithelial–CAF crosstalk (Fig. 5A). Detailed mapping of ligand–receptor pairs further demonstrated that fibroblast-derived oncogenic cues—including multiple WNT5A–FZD interactions and TGFβ1–TGFBR1/2 signals—were preferentially directed toward CFAP73 – epithelial cells (Fig. 5B). In contrast, these interactions were substantially attenuated toward CFAP73 + epithelial cells, consistent with their association with reduced EMT potential and a more differentiated epithelial state.

In striking contrast to the weakened CAF interactions, CFAP73 + epithelial cells exhibited significantly enhanced communication with cytotoxic and inflammatory T-cell subsets (Fig. 5C). Ligand–receptor analysis revealed prominent engagement of pro-apoptotic and immune-activating pairs, including TNFSF10–TNFRSF10B, IFNG–IFNGR, TNFSF14–TNFRSF14, and multiple CCL5–CCR interactions (Fig. 5D). These interactions are known to potentiate epithelial apoptosis, promote T-cell cytotoxicity, and enhance anti-tumor immune surveillance. Meanwhile, CFAP73 – epithelial cells showed increased interactions with immune checkpoint and inhibitory receptors—including CD274–PDCD1 (PD-1/PD-L1 axis), HLA-E–NKG2A, and HLA-F/LI–LILRB families—suggesting that epithelial loss of CFAP73 is associated with acquisition of immune-evasive signaling (Fig. 5E).

Taken together, these findings highlight a coordinated shift in epithelial communication associated with CFAP73 loss: fibroblast-derived WNT and TGFβ signals preferentially target CFAP73 – epithelial cells, while cytotoxic T-cell engagement is directed toward CFAP73 + cells. This shift provides a mechanistic link between HP-driven reduction of CFAP73, enhanced CAF–epithelial oncogenic signaling, and increased immune-checkpoint–mediated evasion, collectively promoting the emergence of an immunosuppressive, tumor-permissive niche.

6.1 Spatial transcriptomics reveals the physical segregation of CFAP73 + epithelium from proliferative, hypoxic, and immunosuppressive niches

To determine how CFAP73 influences the spatial organization of the gastric tumor microenvironment, we performed spatial transcriptomic profiling across multiple gastric cancer sections. CFAP73 + and CFAP73 – epithelial regions exhibited strikingly distinct spatial distributions. CFAP73 + epithelial cells were concentrated in well-differentiated, structurally preserved peripheral epithelial areas, whereas CFAP73 – epithelial regions were located deeper within tumor cores marked by architectural disruption (Fig. 6A and Supplementary Figure S5A–B). Spatial maps of T-cell subsets revealed that cytotoxic effector T cells and Th17 cells localized preferentially near CFAP73 + epithelial

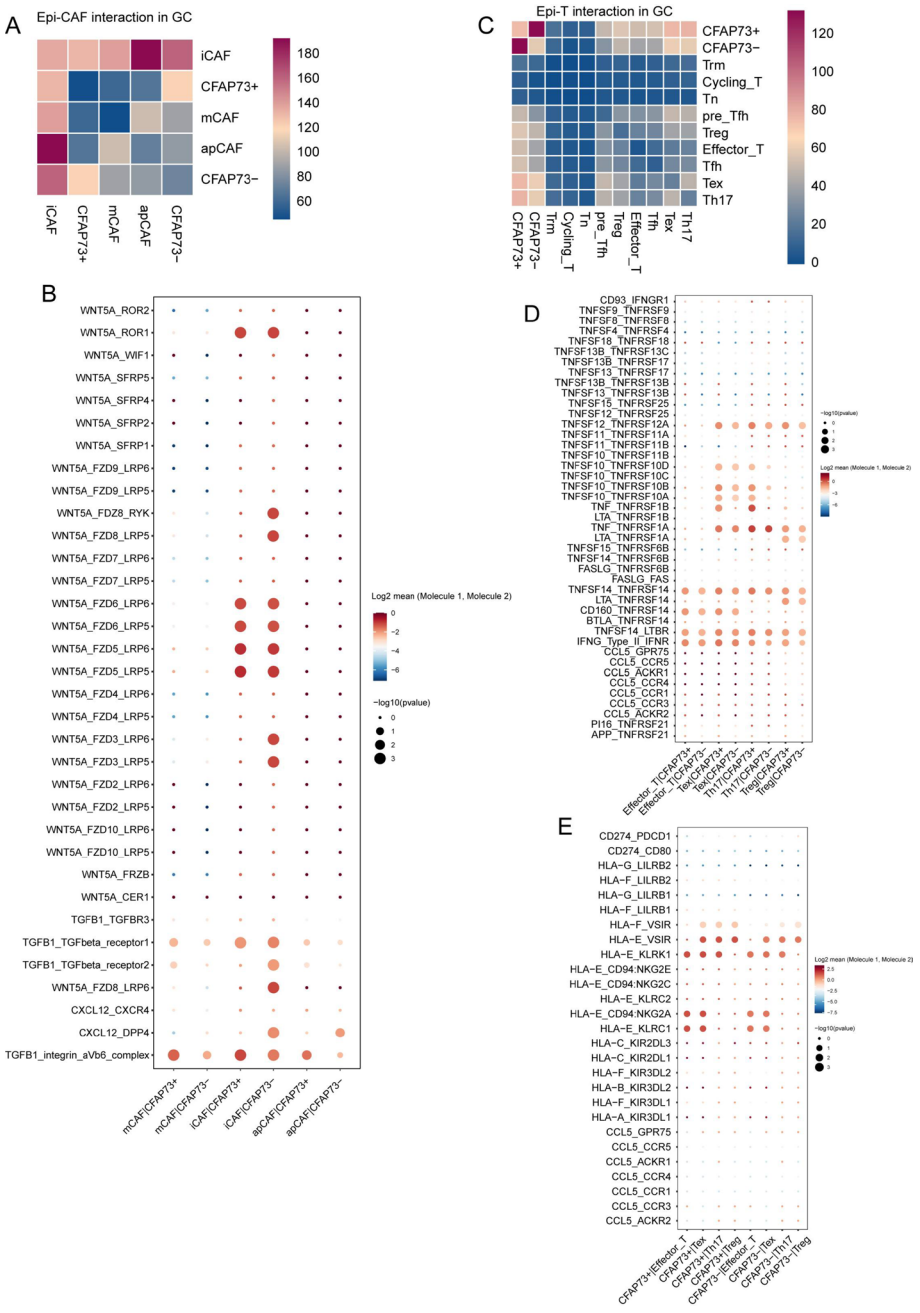

Fig. 5 CFAP73 + epithelial cells receive stronger cytotoxic T-cell signaling and reduced CAF oncogenic input. **A** Global epithelial–CAF interaction heatmap showing reduced CAF interactions toward CFAP73 + epithelium. **B** WNT and TGFβ ligand–receptor interactions preferentially targeting CFAP73 − epithelial cells, including multiple WNT5A–FZD and TGFB1–TGFBR pairs. **C** Global epithelial–T-cell interaction heatmap. CFAP73 + epithelial cells show stronger engagement with cytotoxic and inflammatory T-cell subsets. **D** Dot plot of pro-apoptotic and immune-activating interactions (TNFSF10–TNFRSF10B, IFNG–IFNGR, TNFSF14–TNFRSF14, and CCL5–CCR). **E** Immune checkpoint ligand–receptor interactions preferentially enriched toward CFAP73 − epithelial cells, including PD-1/PD-L1, HLA-E–NKG2A, LILRB, and CCL5–CCR5 inhibitory axes

zones, while exhausted T cells and Tregs were enriched in CFAP73 − regions, reflecting the immunological polarity observed in single-cell data.

Overlaying CAF signatures showed a similar pattern: mCAFs were positioned predominantly near CFAP73 + epithelium, whereas iCAFs and apCAFs clustered tightly

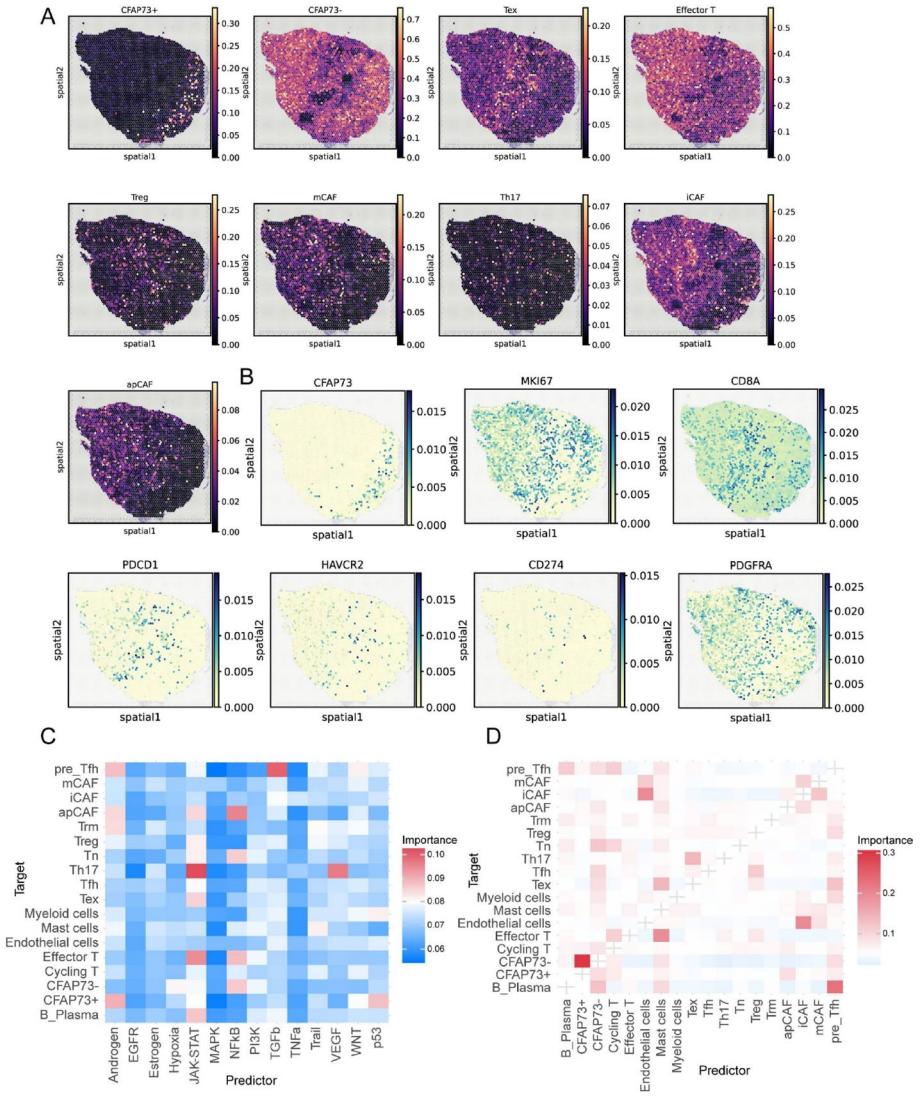

Fig. 6 Spatial transcriptomics reveals physical segregation between CFAP73 + and oncogenic/hypoxic/immunosuppressive niches. **A** Tangram-inferred spatial maps showing CFAP73 + and CFAP73 – epithelial regions, CAF subtypes, Tex, Th17, Tregs, and effector T cells across representative gastric cancer tissue sections ($n = 10$; 32291 spots). CFAP73 + epithelium localizes to well-differentiated peripheries; CFAP73 – regions cluster in hypoxic tumor cores with abundant iCAFs/apCAFs and dysfunctional T cells. **B** Spatial distribution of CFAP73, proliferation marker MKI67, CD8A, PDCD1, HAVCR2, CD274, and PDGFRA. CFAP73 inversely correlates with proliferation, immune checkpoints, and fibroblast markers. Total sample sizes and counts are detailed in the corresponding Methods section. **C** mistyR pathway–cell spatial dependency analysis showing spatial coupling of CFAP73 – epithelium with hypoxia, MAPK, EGFR, and JAK–STAT pathways. **D** mistyR cell–cell spatial dependency matrix showing strong co-localization of CFAP73 – epithelium with iCAFs/apCAFs, and large spatial distance between CFAP73 + epithelium and CAF hubs

around CFAP73 – epithelial regions (Fig. 6A). This spatial co-segregation supports a model in which CFAP73 – epithelial cells preferentially interface with pro-tumor fibroblast niches, consistent with CAF-to-epithelium ligand–receptor interactions identified earlier. To further validate these spatial relationships, we mapped the distribution of key pathway markers. CFAP73 expression inversely correlated with MKI67, PDCD1, HAVCR2, CD274, and PDGFRA, all of which were spatially enriched in regions with low CFAP73 expression (Fig. 6B). These findings highlight that CFAP73 – epithelial zones coincide with proliferative, immune-checkpoint–active, and fibroblast-dense

microenvironments, while CFAP73 + epithelial regions align with lower proliferation and reduced immunosuppression.

To further characterize how CFAP73 shapes the spatial architecture of the tumor microenvironment, we quantified spatial dependency patterns across pathways and cell populations. Pathway–cell distance modeling revealed that CFAP73 – epithelial regions were preferentially located near hypoxia, EGFR, JAK–STAT, and MAPK pathway activation, while CFAP73 + epithelial areas were spatially associated with lower levels of these oncogenic signaling programs (Fig. 6C). Notably, Th17 cells and iCAFs showed strong spatial proximity to hypoxic and JAK–STAT–active regions, whereas cytotoxic effector T cells localized preferentially to regions with lower pathway activation, consistent with the previously observed immune and fibroblast polarization. We next computed cell–cell spatial dependencies to determine how different cell populations co-localize within the tumor tissue. CFAP73 – epithelial cells showed the strongest spatial coupling with iCAFs and apCAFs, while CFAP73 + epithelial cells were positioned farther away from these fibroblast populations (Fig. 6D). Conversely, CFAP73 + epithelial regions exhibited closer spatial proximity to effector T cells and cycling T cells, whereas CFAP73 – regions were enriched near Tregs, exhausted T cells, and fibroblast-dense zones. These spatial relationships reinforce the model that CFAP73 loss shifts epithelial cells into immunosuppressive and fibroblast-dominated niches, while CFAP73 + epithelial cells remain embedded within more immunologically active microenvironments.

Taken together, spatial transcriptomics establishes that CFAP73 + epithelial cells are physically segregated from oncogenic, hypoxic, and immunosuppressive niches, whereas CFAP73 loss corresponds to direct spatial coupling with proliferative tumor cores, immune-evasive signaling centers, and pro-tumor fibroblast hubs. This spatial organization provides in situ evidence that CFAP73 acts as a structural and molecular barrier against the establishment of malignant epithelial–stromal ecosystems.

7 Materials and methods

7.1 Bulk RNA-seq preprocessing and differential expression analysis

TCGA-STAD RNA-seq raw count matrices and corresponding clinical annotations were downloaded from the UCSC Xena data portal (https://xenabrowser.net/) [13], which provides harmonized data aligned to the GDC pipeline. GEO datasets used in this study (including HP-infection cohorts and external validation cohorts) were obtained from the NCBI Gene Expression Omnibus. All datasets were inspected for sample quality, annotation consistency, and format compatibility prior to downstream processing. Raw count matrices were normalized and variance-stabilized using the DESeq2 package [14]. Size factors were estimated by the median-of-ratios method, followed by dispersion estimation using empirical Bayes shrinkage. Differential expression analysis between predefined groups was performed using the DESeq2 Wald test. Differentially expressed genes (DEGs) were defined using $|\log2 \text{ fold change}| > 1$ and $P < 0.05$. Raw count matrices were utilized for differential expression analysis via DESeq2 to ensure statistical robustness, whereas log2(TPM + 1) normalized values were employed for survival modeling and visualization to ensure cross-sample comparability. To identify genes exhibiting monotonic expression changes across HP-infection progression, we applied the Jonckheere–Terpstra trend test to infection-stage–ordered samples from GSE60662 and GSE60427. The stages of H. pylori-associated progression in GSE60662 and GSE60427

were defined based on histological diagnosis, ranging from control (uninfected) to mild/severe gastritis and intestinal metaplasia (IM). The ordering was consistently maintained across both cohorts to ensure biological comparability during the Jonckheere–Terpstra trend test. Pairwise comparisons between HP-infected and uninfected gastric mucosa were evaluated using the Wilcoxon rank-sum test, unless otherwise specified. Batch effects in GEO datasets were assessed but not detected at levels requiring correction. Functional and pathway enrichment analyses for DEGs were performed using the clusterProfiler package [15], incorporating Gene Ontology biological process terms, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, and MSigDB Hallmark gene sets. Multiple testing correction was performed using the Benjamini–Hochberg procedure, and enrichment significance was set at adjusted $P < 0.05$.

7.2 Random forest analysis

The 9-gene set was prioritized using a significance threshold of $P < 0.05$ in both the Jonckheere–Terpstra trend test and univariate Cox regression in TCGA-STAD. For the Random Forest model, the 'randomForest' R package was used with 500 trees (ntree = 500), and genes were ranked based on the Mean Decrease Gini index.

7.3 Survival and pan-cancer analysis

Survival associations between gene expression and patient outcomes were assessed using both univariate Cox proportional hazards regression and Kaplan–Meier estimators. Overall survival (OS) time and event status for TCGA STAD were obtained from the UCSC Xena clinical archive. Gene expression values were log2-transformed (TPM + 1) prior to analysis. For Cox regression, hazard ratios (HRs), 95% confidence intervals, and Wald test P values were calculated using the survival package. For Kaplan–Meier analysis, patients were stratified into high and low expression groups based on median expression thresholds, and statistical significance was determined using the log rank test. Pan cancer prognostic associations were computed independently across all available TCGA cancer types. For each cancer type, univariate Cox models were fitted using gene expression as a continuous variable to derive hazard ratios and P values. Cancer types with fewer than 30 survival events were excluded to ensure model stability. Resulting hazard ratios were aggregated to visualize cross cancer prognostic patterns and determine whether gene expression was broadly protective, neutral, or risk associated across tumor contexts. Pathway activation scores used for correlation analyses were calculated by single sample gene set enrichment analysis (ssGSEA) implemented in the GSVA package [16]. Hallmark, KEGG, and curated oncogenic pathway gene sets were obtained from MSigDB. ssGSEA scores were Z scaled within each cohort, and associations between pathway activity and genes of interest were quantified using Spearman correlation, with significance corrected by the Benjamini–Hochberg method. For all survival analyses, including overall survival (OS) and disease-free survival (DFS) analysis across all independent cohorts, the median expression value of CFAP73 was strictly employed as the universal cutoff to stratify patients into CFAP73-high and CFAP73-low groups. This approach was chosen to maintain consistency and minimize the risk of over-optimizing results through cohort-specific cutoffs. The association between CFAP73 expression and cisplatin response was evaluated using the GEPIA3 database, based on the TCGA-STAD cohort. Patients were stratified into 'Cisplatin-treated' and

'Untreated' subgroups. Survival differences were assessed using the Log-rank test, and Hazard Ratios (HR) were calculated via the Cox proportional hazards model.

7.4 Immune infiltration and deconvolution

Tumor immune cell infiltration in TCGA-STAD bulk RNA-seq samples was quantified using the IOBR Immune Deconvolution and Signature Repository (IOBR) package (https://github.com/IOBR/IOBR) [17], which provides a unified framework integrating multiple widely used deconvolution algorithms. Five independent immune-inference methods were applied: TIMER, quanTIseq, CIBERSORT, MCP-counter, and EPIC. For each method, normalized expression matrices were supplied as input, and default signature matrices curated in IOBR were used without modification. Immune infiltration estimates were computed for major lymphoid and myeloid lineages, including CD8 T cells, effector T cells, regulatory T cells, NK cells, macrophage subsets, dendritic cells, and B cells. To ensure robustness, cell-type abundance estimates across algorithms were standardized and compared to assess consistency. Correlations between gene expression and inferred immune infiltration were computed using Spearman coefficients and adjusted for multiple testing using the Benjamini–Hochberg method. To integrate bulk RNA-seq with single-cell data and obtain higher-resolution cell-type composition estimates, we performed Bayesian deconvolution using the BayesPrism [18] package. Single-cell RNA-seq data from our gastric tissue atlas were first curated to define high-confidence cell-type reference profiles. Quality control included filtering for mitochondrial content, minimum UMI counts, and removal of doublets. Annotated cell types—including epithelial subclusters, CAF subtypes, endothelial cells, myeloid subsets, and T-cell lineages—were used as priors for BayesPrism. Reference expression matrices were generated by averaging normalized gene counts (log-normalized counts) within each annotated cell type. To improve deconvolution accuracy, genes with low expression across all single-cell types (< 1 TPM in > 90% of samples) were excluded, and highly variable genes were retained. BayesPrism was run using the default Gibbs sampling scheme, with 10,000 iterations and a burn-in period of 2,000 iterations. The algorithm estimates posterior distributions of cell-type proportions and gene expression contributions per cell type. For each bulk sample, final cell-type fraction estimates were computed using posterior means. Convergence was assessed by monitoring Markov chain traceplots and ensuring stable posterior distributions.

7.5 Single-cell RNA-seq processing, integration, and annotation

Raw single-cell RNA-seq data were processed using Seurat. Gene–barcode matrices for each sample were first filtered to remove low-quality cells and potential doublets based on global distribution characteristics of gene counts, UMI counts, and mitochondrial transcript percentage. Instead of fixed cutoffs, sample-specific QC thresholds were determined using median absolute deviation (MAD)–based statistical outlier detection, whereby cells falling outside the typical distribution range for any QC metric were excluded. Genes expressed in fewer than three cells were removed from downstream analyses. For each dataset, expression values were log-normalized using a scale factor of 10,000, and highly variable genes were identified using Seurat's vst method. All samples were merged and scaled, while regressing out technical covariates including sequencing depth and mitochondrial transcript percentage. To correct for batch effects across

HP-negative, HP-positive, non-malignant, and tumor tissues, we applied Harmony on the top principal components. Harmony-adjusted embeddings were used for subsequent graph-based clustering using the Louvain algorithm and for visualization using UMAP. Malignant and non-malignant epithelial populations were distinguished using infercnvpy. Stromal and immune cells were used as reference "normal" profiles. infercnvpy computes gene-expression intensity along chromosomes, applies window smoothing, and identifies large-scale amplifications and deletions characteristic of malignant genomes. Epithelial cells with clear CNV signatures were classified as malignant, while those lacking broad chromosomal alterations were designated non-malignant. In malignance, a cell was defined as CFAP73-positive if its log-normalized expression value was >0. Analysis confirmed that CFAP73 is predominantly detected in the epithelial compartment.

7.6 Pathway and functional scoring

Gene-set enrichment analyses were performed using clusterProfiler. Differentially expressed genes were ranked by log2 fold change, and pre-ranked GSEA was conducted against the following resources: Molecular Signatures Database (MSigDB) Hallmark gene sets (https://www.gsea-msigdb.org), Gene Ontology (GO) (https://geneontology.org), Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database (https://www.kegg.jp), Reactome Pathway Knowledgebase (Reactome) pathway database (https://reactome.org), and curated oncogenic signature collections from MSigDB (https://www.gsea-msigdb.org). Enrichment significance was determined using 1,000 permutations, and adjusted P-values were computed using the Benjamini–Hochberg procedure. Over-representation analyses were performed as complementary evidence for up- or down-regulated pathways. To quantify pathway activation at the single-cell level, we computed single-sample enrichment scores using Seurat's AddModuleScore. Gene sets such as EMT, proliferation, apoptosis, inflammatory response and hypoxia were derived from MSigDB (https://www.gsea-msigdb.org) and Progeny (https://github.com/saezlab/progeny). Module scores were calculated after regressing out technical covariates and were used to compare pathway activity across HP-negative, HP-positive, malignant, and non-malignant epithelial subsets.

7.7 Cell–cell communication analysis

Cell–cell communication was inferred using CellPhoneDB (v5.0) (https://www.cellphonedb.org/). Significant interactions were filtered based on a p-value <0.05, with 1000 permutations and a minimum cell expression threshold of 10% within each cluster. For each cell-type pair, significant ligand–receptor interactions were calculated using 1,000 permutations. Interaction strengths and pathway-specific ligand–receptor pairs (e.g., WNT, TGFβ, TNFSF, CXCL families) were visualized using kplots for systematic heatmap and dot-plot rendering.

7.8 Spatial transcriptomic deconvolution and mapping

To map single-cell–defined transcriptional states onto spatial transcriptomic sections, we employed Tangram. Single-cell RNA-seq data were first preprocessed to obtain normalized expression matrices and cell-type labels for major lineages (epithelial subsets, CAF subtypes, endothelial cells, T-cell states, myeloid populations, and B/plasma cells).

Tangram was run in gene-mode, aligning the single-cell expression matrix to the spatial gene-expression matrix while constraining the optimization with cell-type abundance priors derived from the single-cell data. Gene sets used for alignment were restricted to highly variable and consistently detected genes shared between single-cell and spatial datasets. The Tangram optimization was performed using stochastic gradient descent until convergence of the alignment objective. For each spatial spot, Tangram generated cell-type probability distributions and continuous abundance scores, enabling spatial reconstruction of epithelial malignant states, differentiation gradients, CAF subtype localization, and T-cell compartmentalization. Tangram alignment was performed using a set of 2,000 highly variable genes (HVGs) shared between the single-cell and spatial datasets. Major lineages (epithelial, T cells, CAFs) were mapped to spatial spots, where the output 'abundance/probability' represents the inferred relative density of each cell state within a given spot. Spatial abundance maps were generated using the native Tangram visualization functions and custom plotting scripts and were used to evaluate co-localization patterns between CFAP73 + epithelium, CFAP73 – malignant epithelium, fibroblast niches, and immune infiltrates.

7.9 Spatial neighborhood and dependency modeling

To quantify spatial dependencies between cellular populations and pathway activities, we used mistyR [19], which implements a multi-view random-forest framework to model how spatial predictors influence the spatial abundance of each target cell type or pathway. For each spatial transcriptomic section, we constructed three complementary "views":

- The intrinsic view (spot-level expression or cell-type abundance).
- The juxtaview (immediate spatial neighbors based on spot adjacency).
- The paraview (distance-weighted neighborhood within a predefined radius).

Expression-derived pathway scores (e.g., hypoxia, EGFR, JAK–STAT, MAPK, WNT, TGFβ, TRAIL signaling) were computed by ssGSEA and included as predictors in the mistyR model. For each target population or pathway, mistyR fits a random-forest model to estimate the relative importance of each predictor—cell type, pathway score, or spatial view—toward explaining its spatial distribution. Importance scores were aggregated across iterations and normalized to yield interpretable dependence matrices. These matrices quantify:

- Pathway–cell spatial proximity (Fig. 6C).
- Cell–cell spatial dependencies (Fig. 6D).
- Spatial segregation or co-localization tendencies among epithelial, fibroblast, and immune compartments. Importance matrices were visualized as heatmaps, where higher importance values indicate strong spatial coupling (close proximity or coordinated localization), whereas low values indicate spatial independence or segregation.

8 Statistical analysis

Statistical analyses and data visualization were performed using R software. For continuous variables, the Wilcoxon rank-sum test or Mann-Whitney U test was employed for comparisons between two groups, while the Kruskal-Wallis test was used for

multi-group comparisons. To ensure the robustness of our single-cell and spatial transcriptomic analyses, P-values were adjusted for multiple testing using the Benjamini-Hochberg (BH) False Discovery Rate (FDR) method where applicable.

9 Discussion

Gastric cancer is a paradigmatic inflammation-driven malignancy in which persistent *Helicobacter pylori* infection reprograms epithelial identity, remodels the stromal environment, and promotes immune tolerance [9, 20]. Although many studies have described how *H. pylori* induce epithelial damage, EMT activation, and genomic instability, the early epithelial events that determine whether chronically injured mucosa progresses toward malignant transformation remain incompletely understood [10]. Our study identifies CFAP73 as a previously uncharacterized epithelial regulator whose loss integrates these pathological processes and helps explain the transition from chronic infection to a cancer-permissive state.

Multiple transcriptomic studies of *H. pylori*–infected gastric mucosa have shown broad suppression of epithelial differentiation genes and disruption of ciliogenesis-associated pathways [20]—features highly consistent with the progressive downregulation of CFAP73 we observed. Prior work has demonstrated that chronic *H. pylori*–induced inflammation downregulates genes responsible for epithelial polarity maintenance, mucociliary function, and structural stability, creating a vulnerable epithelium prone to metaplastic transformation [21]. Our finding that CFAP73 decreases monotonically with *H. pylori* infection severity aligns with these observations and suggests that its loss represents a core epithelial response to sustained microbial stress.

Importantly, our single-cell analyses demonstrate that CFAP73 low or absent epithelial cells rapidly acquire malignant-like transcriptional features, including increased MAPK, JAK–STAT, EGFR, and hypoxia signaling. These pathway activations mirror mechanisms previously shown to drive gastric epithelial transformation in the context of chronic inflammation. Hypoxia signaling, for example, has been shown to initiate gastric dysplastic programs and promote tumor-initiating cell emergence [22]; similarly, aberrant MAPK and JAK–STAT signaling represent well-established downstream consequences of *H. pylori* CagA-driven epithelial perturbation [23–26]. Thus, our data provide a unifying explanation whereby CFAP73 loss not only marks but may permit the activation of these canonical oncogenic programs.

A major conceptual advance of our study is the integration of epithelial transcriptional states with stromal and immune microenvironmental remodeling. Previous research has highlighted the pivotal role of cancer-associated fibroblasts (CAFs)—particularly inflammatory CAFs (iCAFs) and antigen-presenting CAFs (apCAFs)—in promoting gastric cancer progression, angiogenesis, and immunosuppression [27–31]. Our ligand–receptor analyses show that CFAP73 – epithelial cells engage strongly with these fibroblast subsets through WNT, TGFβ, TNFSF, and ECM-modifying pathways, consistent with known mechanisms of CAF-mediated epithelial dedifferentiation and tumor expansion [32–34]. Conversely, CFAP73 + epithelial regions demonstrated weak stromal interaction, reinforcing the idea that CFAP73 maintains an epithelial phenotype resistant to fibroblast-derived oncogenic cues. Immune ecosystem remodeling is another hallmark of gastric carcinogenesis. Many studies have shown that *H. pylori* infection induces chronic T-cell dysfunction, elevates PD-1/PD-L1 signaling, and fosters an immunosuppressive

tumor milieu. Our results echo these findings, demonstrating that CFAP73 – epithelial regions spatially co-localize with exhausted T cells and Tregs, and show strong engagement of checkpoint pathways such as PD-1/PD-L1, HAVCR2, and LILRB families. These molecules have been repeatedly implicated in gastric cancer immune escape and poor immunotherapy responsiveness. In contrast, CFAP73 + epithelial cells preferentially interact with cytotoxic and effector T-cell states, consistent with observations that well-differentiated or less-damaged epithelium supports more active immune surveillance. This polarity suggests that CFAP73 acts as a determinant of whether epithelial cells may contribute to immune engagement or contribute to immune tolerance.

Spatial transcriptomics provides an in situ perspective supporting this model. Several spatial studies of gastric cancer have reported that malignant epithelial clusters reside within hypoxic, fibroblast-dense, immune-excluded cores, while better-differentiated epithelium remains at peripheral regions [35, 36]. Our spatial maps reproduce this architectural pattern with striking clarity: CFAP73 – epithelium accumulates in proliferative, hypoxic niches enriched for CAFs and immune checkpoints, whereas CFAP73 + cells remain in differentiated peripheral layers associated with active immune niches. The spatial dependency modeling further specifies that CFAP73 – epithelium forms structured, predictable relationships with iCAFs and apCAFs, while CFAP73 + epithelium is more strongly coupled to cytotoxic T-cell infiltration. These findings align with emerging evidence that epithelial differentiation status determines microenvironmental architecture in gastrointestinal tumors.

Together, these findings position CFAP73 at a previously unrecognized convergence point between epithelial integrity, stromal activation, and immune regulation. While prior studies have emphasized inflammatory and microbial triggers of gastric carcinogenesis, few have identified epithelial gatekeepers capable of protecting tissue architecture and immune competence. Our data suggest that CFAP73 fulfills this role by maintaining epithelial stability and preventing malignant ecosystem formation. When CFAP73 is lost—first through *H. pylori*–induced repression and later through tumor evolution—the epithelial compartment becomes permissive to CAF engagement, oncogenic signaling, and immune evasion.

Clinically, these results have several implications. First, CFAP73 may serve as an early biomarker to identify *H. pylori*-infected individuals at high risk of progression to gastric cancer, addressing a major gap in current surveillance strategies. Second, the strong association between CFAP73 loss, CAF-driven pathways, and immune-checkpoint upregulation suggests predictive value for stromal-targeted therapies and immunotherapy responsiveness. Third, CFAP73-defined epithelial states may guide patient stratification in precision oncology, especially as spatial biology becomes increasingly integrated into diagnostic workflows.

9.1 Limitations

This study has several limitations. First, although we integrated multiple bulk and single-cell transcriptomic datasets, the sample sizes of some publicly available cohorts remain relatively modest and may not fully capture the heterogeneity of Helicobacter pylori–associated gastric cancer. Second, several conclusions rely on computational inferences, including pathway scoring, immune deconvolution, and cell–cell communication analyses, which may introduce method-specific biases. Third, while we validated

key findings in independent datasets and through limited experimental assays, further in-depth mechanistic and functional studies are required to fully establish causality and to characterize the biological roles of CFAP73 in gastric tumorigenesis. These limitations should be considered when interpreting our results.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s12672-026-04891-8.

Supplementary material 1. Figure S1. Pan-cancer and disease-free validation of CFAP73 as a favorable biomarker. (A) Pan-cancer Cox regression showing predominantly protective effect of CFAP73 across TCGA cancer types. (B) Disease-free survival (DFS) analysis in GSE62254 demonstrating better prognosis for CFAP73-high tumors.

Supplementary material 2. Figure S2. External functional validation of CFAP73 antitumor roles. (A) CITE CRISPR screen: CFAP73 sgRNAs enriched in T-cell–tumor co-culture assays, indicating selective disadvantage when CFAP73 is lost. (B) CFAP73 expression negatively associated with cancer-risk modules and T-cell dysfunction across immunotherapy, TCGA, CPTAC, ICGC, TARGET, PRECOG, and clinical cohorts.

Supplementary material 3. Figure S3. Functional characterization of CFAP73 + epithelial cells. (A–B) GO and KEGG enrichment: CFAP73 + epithelial cells enriched for mitochondrial translation, RNA processing, proteostasis, and ribosomal biogenesis programs. (C) GSEA of extrinsic apoptosis, bacterial response, and IL-6 signaling. (D–E) EMT scores and expression of MKI67 and PCNA demonstrating reduced proliferation and EMT in CFAP73 + epithelial cells.

Supplementary material 4. Figure S4. Functional characterization of iCAF, apCAF, and mCAF programs. (A–B) GO and KEGG enrichment of iCAFs highlighting ECM remodeling, PI3K–Akt, cytokine signaling, and complement cascades. (C–D) GO and KEGG enrichment of apCAFs showing mitochondrial and oxidative phosphorylation programs. (E) Heatmap comparing pathway activity across CAF subtypes.

Supplementary material 5. Figure S5. Expanded spatial transcriptomics across multiple tissue sections. (A) UMAP showing clustering and library IDs for all spatial sections. (B) Full Tangram-inferred maps of CFAP73+, CFAP73−, Tex, effector T, Th17, iCAF, mCAF, apCAF across all spatial samples.

Acknowledgements
Not applicable.

Author contributions
Haiwen Li and Ran Wei conducted analysis and wrote the manuscript, Yuhan Liu and Jiaju Wang created figures and tables and revised the manuscript, Xinyi Wang and Li Chen searched for literature and revised the manuscript, Bangjie Chen and Yong Yao designed the study, supervised the project, and revised the manuscript. Haiwen Li and Yong Yao provided financial support. All authors have reviewed and approved the manuscript.

Funding
This study is supported by Natural Science Research Major Project of Anhui Educational Committee (No.2023AH040077), Hefei Municipal Natural Science Foundation (No.2022033) and Basic and Clinical Cooperation Improvement Program of Anhui Medical University (No.2022xkjT038).

Data availability
Bulk RNA-seq data for gastric adenocarcinoma were downloaded from the TCGA-STAD project through the UCSC Xena data portal (https://xenabrowser.net/). For Helicobacter pylori–associated transcriptomic profiling, gastric mucosa datasets were obtained from the NCBI Gene Expression Omnibus (GEO). Specifically, GSE60427 and GSE60662 were included as well-established cohorts profiling gastric epithelial responses across H. pylori infection stages ("Con", "Mild", "Severe" and "IM"). Raw or processed expression matrices and sample annotation files were downloaded directly from GEO and harmonized for downstream differential expression and trend analyses. To characterize single-cell transcriptional heterogeneity, publicly available gastric single-cell RNA-seq datasets were integrated. Single-cell gene-barcode matrices were obtained from GSE249874. Spatial transcriptomics dataset was retrieved under Accession Number GSE203612.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 18 December 2025 / Accepted: 17 March 2026
Published online: 04 April 2026

References

1. Thrift AP, Wenker TN, El-Serag HB. Global burden of gastric cancer: epidemiological trends, risk factors, screening and prevention. Nat Rev Clin Oncol. 2023;20(5):338–49. https://doi.org/10.1038/s41571-023-00747-0.
2. Sundar R, Nakayama I, Markar SR, Shitara K, van Laarhoven HWM, Janjigian YY, Smyth EC. Gastric cancer. Lancet. 2025;405(10494):2087–102. https://doi.org/10.1016/S0140-6736(25)00052-2.
3. Mamun TI, Younus S, Rahman MH. Gastric cancer-Epidemiology, modifiable and non-modifiable risk factors, challenges and opportunities: An updated review. Cancer Treat Res Commun. 2024;41:100845. https://doi.org/10.1016/j.ctarc.2024.100845.
4. Duan Y, Xu Y, Dou Y, Xu D. Helicobacter pylori and gastric cancer: mechanisms and new perspectives. J Hematol Oncol. 2025;18(1):10. https://doi.org/10.1186/s13045-024-01654-2.
5. Yan T, Sun H, Liao Y, Zhou J. Helicobacter pylori mediated niche environment aberrations promote the progression of gastric cancer. Genes Dis. 2024;11(5):101207. https://doi.org/10.1016/j.gendis.2024.101207.
6. Sharndama HC, Mba IE. Helicobacter pylori: an up-to-date overview on the virulence and pathogenesis mechanisms. Braz J Microbiol. 2022;53(1):33–50. https://doi.org/10.1007/s42770-021-00675-0.
7. Mommersteeg MC, Simovic I, Yu B, van Nieuwenburg SAV, Bruno IMJ, Doukas M, et al. Autophagy mediates ER stress and inflammation in Helicobacter pylori-related gastric cancer. Gut Microbes. 2022;14(1):2015238. https://doi.org/10.1080/19490976.2021.2015238.
8. Bakhti SZ, Latifi-Navid S. Interplay and cooperation of Helicobacter pylori and gut microbiota in gastric carcinogenesis. BMC Microbiol. 2021;21(1):258. https://doi.org/10.1186/s12866-021-02315-x.
9. Hu W, Chen ZM, Wang Y, Yang C, Wu ZY, You LJ, et al. Single-cell RNA sequencing dissects the immunosuppressive signatures in Helicobacter pylori-infected human gastric ecosystem. Nat Commun. 2025;16(1):3903. https://doi.org/10.1038/s41467-025-59339-4.
10. Alzahrani S, Lina TT, Gonzalez J, Pinchuk IV, Beswick EJ, Reyes VE. Effect of Helicobacter pylori on gastric epithelial cells. World J Gastroenterol. 2014;20(36):12767–80. https://doi.org/10.3748/wjg.v20.i36.12767.
11. Biagi E, Nylund L, Candela M, Ostan R, Bucci L, Pini E, et al. Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians. PLoS ONE. 2010;5(5):e10667. https://doi.org/10.1371/journal.pone.0010667.
12. Huang Z, Li Y, Qian Y, Ye L, Zhang T, Cheng Y, et al. An LCN2-Dependent Positive-Feedback Loop Between Gastric Cancer Cells and Tumor-Associated-Macrophages Mediates Lymphangiogenesis and Lymphatic Metastasis. Adv Sci (Weinh). 2025;12(44):e08352. https://doi.org/10.1002/advs.202508352.
13. Li S, Peng Y, Chen M, Zhao Y, Xiong Y, Li J, et al. Facilitating integrative and personalized oncology omics analysis with UCSCXenaShiny. Commun Biol. 2024;7(1):1200. https://doi.org/10.1038/s42003-024-06891-2.
14. Nearing JT, Douglas GM, Hayes MG, MacDonald J, Desai DK, Allward N, et al. Microbiome differential abundance methods produce different results across 38 datasets. Nat Commun. 2022;13(1):342. https://doi.org/10.1038/s41467-022-28034-z.
15. Xu S, Hu E, Cai Y, Xie Z, Luo X, Zhan L, et al. Using clusterProfiler to characterize multiomics data. Nat Protoc. 2024;19(11):3292–320. https://doi.org/10.1038/s41596-024-01020-z.
16. Zhang T, Hasegawa Y, Waldor MK. Enteric bacterial infection stimulates remodelling of bile metabolites to promote intestinal homeostasis. Nat Microbiol. 2024;9(12):3376–90. https://doi.org/10.1038/s41564-024-01862-z.
17. Zeng D, Fang Y, Qiu W, Luo P, Wang S, Shen R, et al. Enhancing immuno-oncology investigations through multidimensional decoding of tumor microenvironment with IOBR 2.0. Cell Rep Methods. 2024;4(12):100910. https://doi.org/10.1016/j.crmeth.2024.100910.
18. Chu T, Wang Z, Pe'er D, Danko CG. Cell type and gene expression deconvolution with BayesPrism enables Bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology. Nat Cancer. 2022;3(4):505–17. https://doi.org/10.1038/s43018-022-00356-3.
19. Tanevski J, Flores ROR, Gabor A, Schapiro D, Saez-Rodriguez J. Explainable multiview framework for dissecting spatial relationships from highly multiplexed data. Genome Biol. 2022;23(1):97. https://doi.org/10.1186/s13059-022-02663-5.
20. Li N, Chen S, Xu X, Wang H, Zheng P, Fei X, et al. Single-cell transcriptomic profiling uncovers cellular complexity and microenvironment in gastric tumorigenesis associated with Helicobacter pylori. J Adv Res. 2024;74:471–91. https://doi.org/10.1016/j.jare.2024.10.012.
21. Li D, Zhou Z, Li X, Guo Q, Yuan L, Zhang X, et al. Hypermethylation-mediated HNF4A silencing by Helicobacter pylori infection drives gastric cancer by disrupting epithelial cell polarity and activating EMT signaling. Cell Death Dis. 2025;16(1):688. https://doi.org/10.1038/s41419-025-08029-6.
22. Schito L, Rey-Keim S. Hypoxia signaling and metastatic progression. Semin Cancer Biol. 2023;97:42–9. https://doi.org/10.1016/j.semcancer.2023.11.001.
23. Xi Y, Zhang X-L, Luo Q-X, Gan H-N, Liu Y-S, Shao S-H, Mao X-H. Helicobacter pylori regulates stomach diseases by activating cell pathways and DNA methylation of host cells. Front Cell Dev Biol. 2023;11:1187638. https://doi.org/10.3389/fcell.2023.1187638.
24. Koch KN, Müller A. Helicobacter pylori activates the TLR2/NLRP3/caspase-1/IL-18 axis to induce regulatory T-cells, establish persistent infection and promote tolerance to allergens. Gut Microbes. 2015;6(6):382–7. https://doi.org/10.1080/19490976.2015.1105427.
25. Ding S-Z, Smith MF, Goldberg JB. Helicobacter pylori and mitogen-activated protein kinases regulate the cell cycle, proliferation and apoptosis in gastric epithelial cells. J Gastroenterol Hepatol. 2008;23(7 Pt 2):e67–78. https://doi.org/10.1111/j.1440-1746.2007.04912.x.
26. Tavares R, Pathak SK. Helicobacter pylori protein JHP0290 exhibits proliferative and anti-apoptotic effects in gastric epithelial cells. PLoS ONE. 2015;10(4):e0124407. https://doi.org/10.1371/journal.pone.0124407.
27. Tsoumakidou M. The advent of immune stimulating CAFs in cancer. Nat Rev Cancer. 2023;23(4):258–69. https://doi.org/10.1038/s41568-023-00549-7.
28. Papait A, Romoli J, Stefani FR, Chiodelli P, Montresor MC, Agoni L et al. Fight the Cancer, Hit the CAF! Cancers (Basel). 2022;14(15). https://doi.org/10.3390/cancers14153570
29. Caligiuri G, Tuveson DA. Activated fibroblasts in cancer: Perspectives and challenges. Cancer Cell. 2023;41(3):434–49. https://doi.org/10.1016/j.ccell.2023.02.015.
30. Chen X, Song E. Turning foes to friends: targeting cancer-associated fibroblasts. Nat Rev Drug Discov. 2019;18(2). https://doi.org/10.1038/s41573-018-0004-1.

31. Kennel KB, Bozlar M, De Valk AF, Greten FR. Cancer-Associated Fibroblasts in Inflammation and Antitumor Immunity. Clin Cancer Res. 2023;29(6):1009–16. https://doi.org/10.1158/1078-0432.CCR-22-1031.
32. Chang J, Lu J, Liu Q, Xiang T, Zhang S, Yi Y, et al. Single-cell multi-stage spatial evolutional map of esophageal carcinogenesis. Cancer Cell. 2025;43(3). https://doi.org/10.1016/j.ccell.2025.02.009.
33. Zhang H, Qian Y, Zhang Y, Zhou X, Shen S, Li J, et al. Multi-omics analysis deciphers intercellular communication regulating oxidative stress to promote oral squamous cell carcinoma progression. NPJ Precis Oncol. 2024;8(1):272. https://doi.org/10.1038/s41698-024-00764-x.
34. Chhabra Y, Weeraratna AT. Fibroblasts in cancer: Unity in heterogeneity. Cell. 2023;186(8):1580–609. https://doi.org/10.1016/j.cell.2023.03.016.
35. Zhang G, Zhang X, Pan W, Chen X, Wan L, Liu C, et al. Dissecting the Spatial and Single-Cell Transcriptomic Architecture of Cancer Stem Cell Niche Driving Tumor Progression in Gastric Cancer. Adv Sci (Weinh). 2025;12(18):e2413019. https://doi.org/10.1002/advs.202413019.
36. Zhang X, Ren B, Liu B, Wang R, Li S, Zhao Y, Zhou W. Single-cell RNA sequencing and spatial transcriptomics reveal the heterogeneity and intercellular communication of cancer-associated fibroblasts in gastric cancer. J Transl Med. 2025;23(1):344. https://doi.org/10.1186/s12967-025-06376-8.

Publisher's Note