




OPEN ACCESS

Original research

# Comprehensive single-cell analysis deciphered microenvironmental dynamics and immune regulator olfactomedin 4 in pathogenesis of gallbladder cancer

Huisi He,<sup>1,2</sup> Shuzhen Chen,<sup>1,2</sup> Yong Yu,<sup>3</sup> Zhecai Fan,<sup>1,2</sup> Youwen Qian,<sup>4</sup> Yaping Dong,<sup>1,5</sup> Yuting Song,<sup>1,2</sup> Caiming Zhong,<sup>1,6</sup> Xiaojuan Sun,<sup>6</sup> Qiqi Cao,<sup>1,2</sup> Shiyao Li,<sup>1,2</sup> Weihang Huang,<sup>1,6</sup> Wenxin Li,<sup>1,6</sup> Mingzhu Zhuang,<sup>6</sup> Jinxian Yang,<sup>1,2</sup> Xianming Wang,<sup>1,5</sup> Jiaqian Wang,<sup>7</sup> Dongfang Wu,<sup>7,8</sup> Hongyang Wang ,<sup>1,2,5</sup> Wen Wen <sup>1,6</sup>

► Additional supplemental material is published online only. To view, please visit the journal online (<https://doi.org/10.1136/gutjnl-2023-331773>).

For numbered affiliations see end of article.

## Correspondence to

Wen Wen, Hongyang Wang and Shuzhen Chen, Third Affiliated Hospital of Naval Medical University (Second Military Medical University), Shanghai, China; [wenwen\\_smmu@163.com](mailto:wenwen_smmu@163.com), [hywangk@vip.sina.com](mailto:hywangk@vip.sina.com), [superelechen@163.com](mailto:superelechen@163.com)

HH, SC and YY contributed equally.

Received 20 December 2023  
Accepted 20 April 2024



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

**To cite:** He H, Chen S, Yu Y, et al. *Gut* Epub ahead of print: [please include Day Month Year]. doi:10.1136/gutjnl-2023-331773

## ABSTRACT

**Objective** Elucidating complex ecosystems and molecular features of gallbladder cancer (GBC) and benign gallbladder diseases is pivotal to proactive cancer prevention and optimal therapeutic intervention.

**Design** We performed single-cell transcriptome analysis on 230 737 cells from 15 GBCs, 4 cholecystitis samples, 3 gallbladder polyps, 5 gallbladder adenomas and 16 adjacent normal tissues. Findings were validated through large-scale histological assays, digital spatial profiler multiplexed immunofluorescence (GeoMx), etc. Further molecular mechanism was demonstrated with *in vitro* and *in vivo* studies.

**Results** The cell atlas unveiled an altered immune landscape across different pathological states of gallbladder diseases. GBC featured a more suppressive immune microenvironment with distinct T-cell proliferation patterns and macrophage attributions in different GBC subtypes. Notably, mutual exclusivity between stromal and immune cells was identified and remarkable stromal ecosystem (SC) heterogeneity during GBC progression was unveiled. Specifically, SC1 demonstrated active interaction between Fibro-iCAF and Endo-Tip cells, correlating with poor prognosis. Moreover, epithelium genetic variations within adenocarcinoma (AC) indicated an evolutionary similarity between adenoma and AC. Importantly, our study identified elevated olfactomedin 4 (OLFM4) in epithelial cells as a central player in GBC progression. OLFM4 was related to T-cell malfunction and tumour-associated macrophage infiltration, leading to a worse prognosis in GBC. Further investigations revealed that OLFM4 upregulated programmed death-ligand 1 (PD-L1) expression through the MAPK-AP1 axis, facilitating tumour cell immune evasion.

**Conclusion** These findings offer a valuable resource for understanding the pathogenesis of gallbladder diseases and indicate OLFM4 as a potential biomarker and therapeutic target for GBC.

## INTRODUCTION

Gallbladder cancer (GBC) is a highly lethal disease, predominantly identified at an advanced stage owing to its atypical symptoms.<sup>1,2</sup> Early-stage GBC is often diagnosed incidentally after cholecystectomy for a

## WHAT IS ALREADY KNOWN ON THIS TOPIC

- ⇒ Gallbladder cancer (GBC) is one of the most common biliary tract cancers, with a particularly poor prognosis and limited therapeutic options when diagnosed at advanced stages.
- ⇒ Gallbladder diseases, ranging from benign to malignant conditions, involve persistent inflammation, irrespective of lithogenic or non-lithogenic origins.
- ⇒ The development of GBC primarily occurs through metaplasia-dysplasia-carcinoma or adenoma-carcinoma sequences, with the latter involving 5–10% of cases.

## WHAT THIS STUDY ADDS

- ⇒ We delineated a comprehensive landscape of the microenvironment within different pathological states of gallbladder diseases and highlighted the features of different GBC subtypes.
- ⇒ Stromal-immune cell exclusivity unveiled a complex stromal ecosystem (SC) in GBC progression, indicating SC1, characterised by active interaction between Fibro-iCAF and Endo-Tip cells was associated with poor prognosis.
- ⇒ Molecular profiling identified intrinsic transcriptional variations within adenocarcinoma, establishing an evolutionary similarity between adenoma and adenocarcinoma.
- ⇒ Olfactomedin 4 (OLFM4) was elevated in GBC, which played a crucial role in facilitating a tumour-evading microenvironment, particularly by enhancing programmed death-ligand 1 expression via MAPK-AP1 axis and recruiting tumour-associated macrophages.

presumed benign disease by pathologists.<sup>3</sup> Surgical resection is currently a recognised treatment for GBC, yet a limited number of cases qualify for such resection and most adjuvant therapy has shown an undesirable response rate.<sup>4</sup> Therefore, to improve GBC management, several critical challenges have to be addressed including the identification

**HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY**

- ⇒ The comprehensive transcriptome atlas offers insights into GBC heterogeneity and precancerous transitions, laying a groundwork for devising targeted therapies through understanding the complex microenvironment and genetic variations.
- ⇒ This study identified OLFM4 as a key modulator in GBC progression, highlighting its potential as a diagnostic marker and therapeutic target.

of reliable biomarkers for screening, clinical management and prognosis prediction.<sup>4,5</sup>

Covering a spectrum from benign to malignant conditions, gallbladder diseases pose a significant medical challenge.<sup>6</sup> Approximately 95% of GBCs originate from epithelial cells and adenocarcinoma (AC) is the predominant histological type, accounting for over 90% of cases.<sup>7</sup> The pathogenesis of GBC often follows metaplasia-dysplasia-carcinoma or adenoma-carcinoma sequence.<sup>8,9</sup> Persistent inflammation exists in both pathogenic ways and plays a pivotal role in driving the tumourigenesis of GBC, regardless of its lithogenic<sup>10</sup> or non-lithogenic origin.<sup>11,12</sup> Although 90% of patients with GBC exhibit gallstones,<sup>13–15</sup> only 1% of gallstone patients progress to GBC.<sup>16</sup> Chronic inflammation<sup>17</sup> also intertwines with risk factors such as *Salmonella enterica* serovar Typhi infection,<sup>18</sup> pancreaticobiliary duct anomalies,<sup>19</sup> metabolic syndrome<sup>20</sup> and gallbladder polypoid lesions to complicate the process. Thus, the transition of GBCs from precancerous lesions could be intervened with a better understanding of inflammation-driven tumour initiation and progression.<sup>19</sup>

The advances in omics technologies have provided novel insights into the mutational landscape, molecular characteristics and dysregulation of signalling pathways, leading to the emergence of several therapeutic targets, such as HER2, VEGFR and PD-(L)1.<sup>21–23</sup> However, trials targeting these molecules have shown minimal impact on overall survival.<sup>24,25</sup> Since the thriving application of single-cell RNA sequencing (scRNA-seq) in deciphering the tumour microenvironment, much emphasis was put on the heterogeneity of the tumour ecosystem to provide novel clues for cancer therapy, especially the complex interaction within different cell subtypes. Understanding the intricate crosstalk between cellular components and the microenvironment sheds light on the malignant transformation and what triggers GBC progression.<sup>26,27</sup>

In this study, we obtained whole-genome sequencing (WGS) and scRNA-seq data of gallbladder disease clinical samples, including cholecystitis, polyps, adenomas and GBCs to unveil the molecular signatures and heterogeneity of the microenvironment. Olfactomedin 4 (OLFM4), a modulator originating from the malignant epithelium of GBCs was identified and its role in regulating programmed death-ligand 1 (PD-L1) and the tumour microenvironment was validated. This investigation aims to illuminate the underlying mechanisms driving gallbladder disease progression and provide insights into potential therapeutic targets for GBCs.

**METHODS****Patient cohorts**

Specimens of patients who underwent surgical resection with gallbladder diseases at Eastern Hepatobiliary Surgery Hospital were collected and clinical data was sourced from electronic medical records under informed consent. The pathology of each patient was confirmed by at least three pathologists. Additional

samples for digital spatial profiler (DSP) GeoMx were from Shanghai Outdo Biotech (Refer to online supplemental material for more methodologies).

**RESULTS****Diverse tissue-lineage states and cell populations identified via scRNA-seq of 230 737 cells**

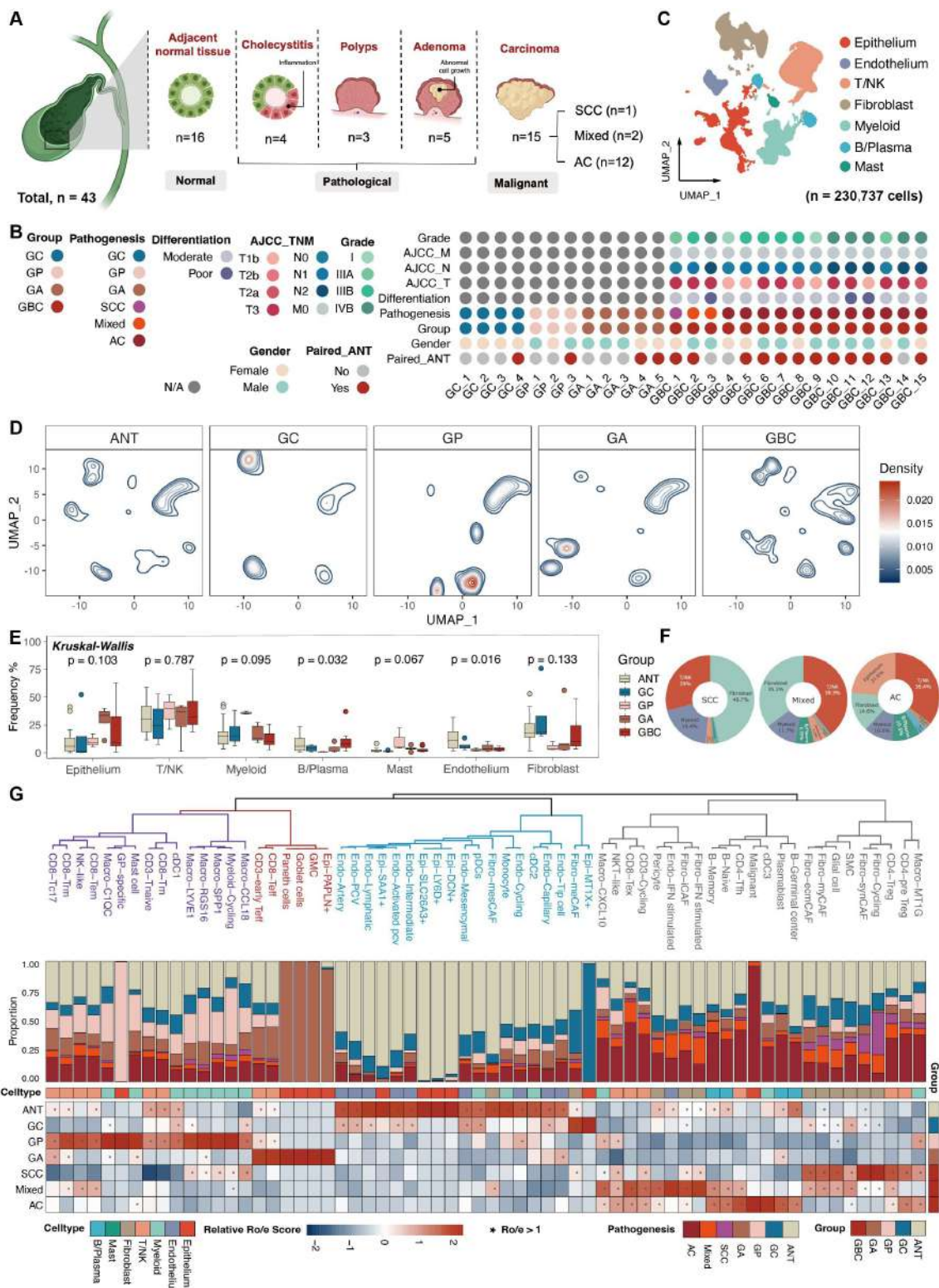
This study performed scRNA-seq analysis of 43 clinical samples from 27 patients, including 4 gallbladder cholecystitis (GC), 3 gallbladder polyps (GP), 5 gallbladder adenomas (GA) and 15 GBC (1 squamous cell carcinoma, SCC; 2 adenocarcinomas, Mixed; and 12 AC), along with 16 matched adjacent normal tissues (ANT, 1 from GC, 1 from GP, 2 from GA, 1 from SCC, 1 from mixed and 10 from AC) (figure 1A). The clinicopathological information of each patient was presented in figure 1B, online supplemental table S1 and figure S1A.

To get an overview of the mutational landscape, we first evaluated mutational patterns in samples from the WGS cohort with somatic mutation calling (online supplemental figure S1B). Notably, 58% of GBC samples exhibited *TP53* and *RTK-RAS* mutations. GA displayed the highest mutational burden with an *APC* gene mutation, aligning with previous reports.<sup>28</sup> While relevant genetic mutations were not detected in GP and GC.

For scRNA-seq, after rigorous quality control, 230 737 cells were classified into seven major cell types, encompassing epithelial, stromal (fibroblasts and endothelium), lymphoid (T/natural killer (NK), B/plasma and mast cells) and myeloid cells (monocytes, macrophages and dendritic cell) (figure 1C and online supplemental figure S1C). Traditionally, rapid cell proliferation is closely associated with pluripotency and stemness.<sup>29</sup> Consistently, the fewest lymphoid cells were in the G1 phase, while epithelial and stromal cells exhibited prolonged G1 phases. The relatively lower proliferative activity of epithelium may be attributed to the inclusion of normal epithelium. GBCs showed a higher proportion of cells in the G2 and S phases when compared with other groups, indicating increased proliferation (online supplemental figure S1D).

Further cellular composition analysis suggested a slightly different cell distribution pattern in each group. Myeloid cells and epithelium might be the predominant subpopulations in GP and GA, respectively. GBC and ANT showed a similar cell composition (figure 1D,E). While lymphoid cells and fibroblast proportion varied in different GBC subtypes (figure 1F). This observation underscores the necessity for unravelling the cellular heterogeneity across various pathological stages of gallbladder diseases.

After removing batch effects and clustering, we established a cell atlas with 69 subpopulations (online supplemental figure S1E and table S2). Comparative analysis highlighted the distinguished pattern of each group (figure 1G). Endothelium and epithelium clusters were enriched in ANT and GC, while GP and GA were enriched in immunocompetent cells, such as CD8<sup>+</sup> effector T (CD8-Teff) and CD8<sup>+</sup> effector memory T (CD8-Tem), and specific epithelial subtypes (ie, Goblet and Paneth cells). As expected, GBC exhibited abundant immunosuppressive immune cells (such as regulatory T (CD4-Treg) and CD8<sup>+</sup> exhausted T (CD8-Tex)) and stromal cells (Fibro-iCAF), aligning with their role in tumourigenesis. Despite the similarity of subpopulation distribution, differences among the three GBC pathological types were not to be neglected.



**Figure 1** Landscape of precancerous lesions of gallbladder and gallbladder cancer by scRNA-seq of 43 samples. (A) Overview of the study design and sample composition. (B) Clinicopathological profiles of samples enrolled in scRNA-seq cohort. (C) Uniform Manifold Approximation and Projection unveiling seven major cell lineages. (D) Visualisation of alterations in major lineages composition among groups through cell density mapping. (E) Boxplots revealing the frequency of major cell compartments across different groups. Statistical significance was evaluated via the Kruskal-Wallis' test. (F) Pie charts illustrating the predominant lineage composition among diverse GBC subtypes. (G) Phenotypic relationships and population abundance of 64 cell subsets excluding five patient-specific clusters. Unsupervised hierarchical clustering of cell subsets (top panel). Bar plot showing the distribution of cell subsets across seven tissue subtypes (middle panel). Heatmap at the bottom showing tissue prevalence estimated by Ro/e score for each cell subset. AC, adenocarcinomas; AJCC, American Joint Committee on Cancer; ANT, adjacent normal tissues; GA, gallbladder adenomas; GBC, gallbladder cancer; GC, gallbladder cholecystitis; GP, gallbladder polyps; NK, natural killer; SCC, squamous cell carcinoma; sc-RNA-seq, single-cell RNA sequencing; TNM, tumour, node, metastasis.

### Altered subpopulation patterns of immune cells in the pathogenesis of gallbladder diseases

Next, unsupervised clustering analysis revealed 14 subpopulations of lymphoid cells, including CD3<sup>+</sup>T cells (naive, early effector, cycling), CD4<sup>+</sup>T cells (pre-Treg, Treg, follicular helper T (Tfh)), CD8<sup>+</sup>T cells (effector, effector memory, memory, resident memory, Tc17, exhausted) and NK cell subsets (figure 2A and online supplemental figure S2A).

Detailed examination of T-cell receptor (TCR) clonal expansion highlighted CD8<sup>+</sup>T cells abundant in memory and exhausted subsets (figure 2B). An average of approximately 80% of CD8<sup>+</sup>T cells harboured TCRs, with minimal variations across tissue types. Notably, GP showed elevated TCR rates and increased identical CDR3 lengths (online supplemental figure S2B).

According to the subpopulation distribution presented in figure 1G, CD4-Tfh, CD4-Treg, CD8-*Tex* and CD3-Cycling were enriched in GBC (figure 2C and S2C). CD8-*Tex* exhibited extensive TCR expansion, notably in mixed GBC and GC (figure 2D and online supplemental figure S2C). Cells sharing identical TCR clonotypes with CD8-*Tex* were enriched within CD8-*Teff*, CD8-*Trm*, CD8-*Tem* and CD3-Cycling (figure 2E), suggesting a potential origin, as supported by RNA velocity analysis (figure 2F). Given the significant increase in CD3-Cycling in GBC, a subpopulation crucial for exhausted CD8<sup>+</sup>T cells, we then analysed its cellular composition illustrating the categorisation of proliferating T-cell subpopulations. In precancerous lesions, predominant proliferating T cells were functionally active, including CD8-*Teff* and CD8-*Tem*. Conversely, GBC showed proliferation of Treg, especially in SCC and exhausted CD8<sup>+</sup>T cells, particularly in AC. The Treg to CD8-*Tex* ratio in mixed GBC fell between SCC and AC, which was further validated through multiple immunohistochemistry (IHC) (figure 2G–2I and D).

Macrophages, diverse across GBC and precancerous lesions, exhibited elevated subtypes in GBC (Macro-CCL18, Macro-CXCL10, Macro-MT1G, Macro-SPP1) and precancerous lesions (Macro-C1QC, Macro-LYVE1, Macro-RGS16) (figure 2J and online supplemental figure S2E,F), while cDC3 expressing LAMP3 upregulated in GBC (figure 2K–2L). Enriched macrophage subtypes in GBC manifest the highest functional scores in M2-like, anti-inflammatory, angiogenesis and immune checkpoint blockade (ICB) (figure 2M). Comparing macrophages from GBC with other precancerous samples identified a consistent expression pattern across the three pathological types (online supplemental figure S2G), encompassing upregulation of six key genes (figure 2N), including metalloproteinase family (eg, MMP12), osteopontin (ie, SPP1) and various chemokines (eg, CXCL10), which held promise as potential targets in cancer treatment. Proliferative myeloid cells were abundant in GP, probably due to the role of macrophages in inflammation which was also validated in clinical samples (online supplemental figure S2H,I).

Overall, lymphoid cell patterns and the core genes of tumour-associated macrophage (TAM) provided implications for the GBC pathogenesis.

### Enhanced plasticity and remodelling of stromal cells during GBC progression

Stromal compartment, encompassing fibroblasts, endothelium and other cell types like pericytes and smooth muscle cells, revealed remarkable heterogeneity in gallbladder diseases (figure 3A,B). A subgroup with glial cell characteristics resembling Schwann cells emerged, suggesting a potential similarity

between GBC and pancreatic cancer, given their shared developmental lineage with the biliary system,<sup>30</sup> particularly in terms of peripheral nerve infiltration (online supplemental figure S3A–C).

Gallbladder, characterised by substantial collagen content, exhibits a pathological landscape with a notable stromal presence.<sup>31</sup> A correlation analysis highlighted mutual exclusivity between stromal and immune cells (green box) and positive correlations among fibroblasts and endothelial cells (yellow box). A favourable correlation between lymphoid cells and myeloid cells, as well as pericytes and endothelium, suggested cell–cell interactions in microenvironment remodelling (figure 3C).

To understand phenotypic diversity within stromal cell subsets, we quantified cellular compositions across all samples, identifying three distinct stromal ecosystems (SC). Remarkably, SC1, mainly composed of Fibro-iCAF, Endo-Tip cells and pericytes, and SC3 enriched with fibroblast subtypes were predominantly observed in tumour samples. While SC2, dominant with endothelial subtypes prevailed in the non-tumour samples (figure 3D). Fibro-iCAF and Endo-Tip cells, known to spearhead angiogenesis in tumorigenesis,<sup>32</sup> were accumulated in samples with SC1 features (online supplemental figure S3D,E). Crosstalk analysis revealed that interactions among different cell types were significantly intensified in GBC compared with other groups (online supplemental figure S3F,G). Notably, among ligands interacting with Endo-Tip cells, TGFB1 and COL18A1 were tremendously upregulated in GBCs, triggering downstream signalling pathways (online supplemental figure S3H–J).

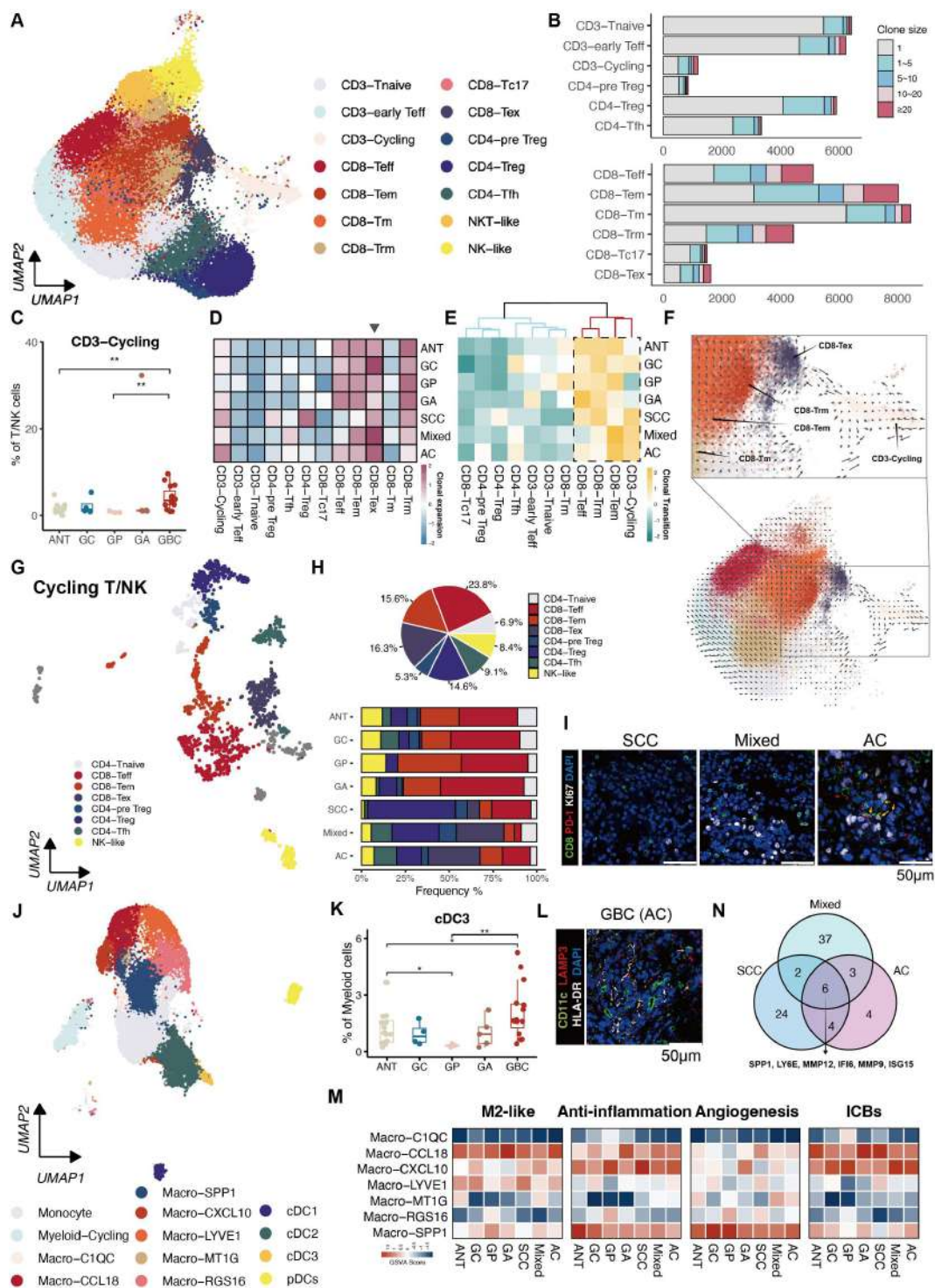
Accordingly, SC1 and SC3 signatures correlated with less favourable prognosis across cancers in The Cancer Genome Atlas (TCGA) cohort (n=10 535), particularly stroma-rich tumours like pancreatic and breast cancer (figure 3E; online supplemental table S3). Furthermore, unsupervised hierarchical clustering of bulk RNA sequencing (RNA-seq) data from GBC samples by Pandey *et al*,<sup>26</sup> also revealed SC1-like and SC3-like subgroups (figure 3F), with SC1-like patients exhibiting abnormally activated pro-tumour pathways, such as the PI3K-AKT and extracellular matrix pathways. While SC3-like patients retained partial gallbladder functions (figure 3G). Likewise, within the scRNA-seq data set, GBC samples with poor differentiation displayed elevated scores for the SC1 signature other than the SC3 signature (figure 3H), suggesting SC1 is a potential signature for worse patient prognosis.

These results emphasise a stromal-immune cell exclusivity in gallbladder diseases and enhanced interaction, notably between Fibro-iCAF and Endo-Tip cells in SC1, indicating worse patient prognosis.

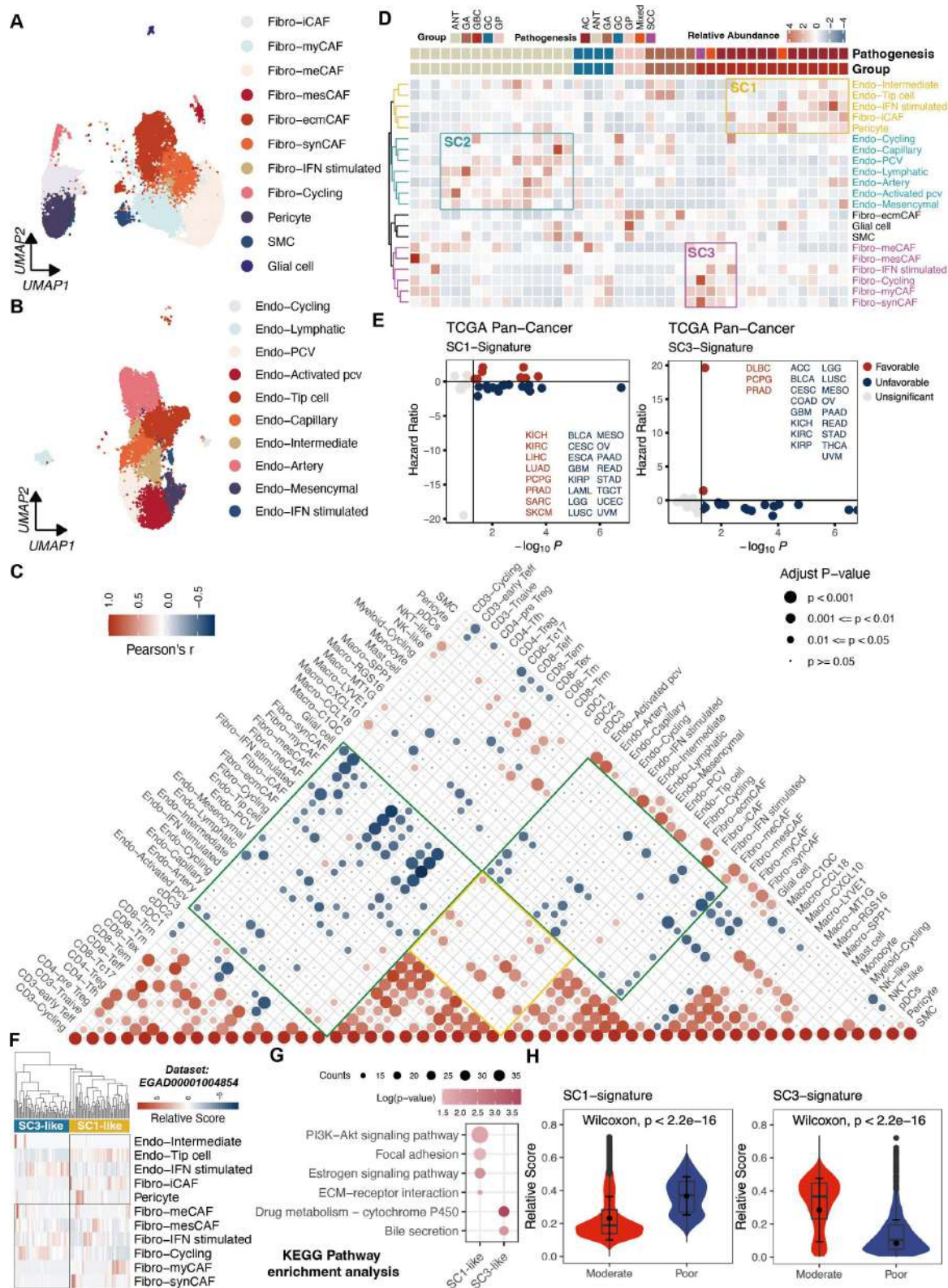
### Intricate molecular features of epithelium and evolutionary similarity between GA and AC

Next, copy number variation inference within epithelium distinguished malignant cells (online supplemental figure S4A). Shannon entropy revealed lower heterogeneity within the epithelium of GA and GBC compared with overall fibroblasts. Additionally, stromal cells displayed higher heterogeneity than immune cells (figure 4A). Precancerous tissues exhibited distinct epithelial cells (figure 4B and online supplemental figure S4B), including Paneth (expressing DEFA5 and DKK4) and Goblet cells (expressing PGC and MSMB) (figure 1G).

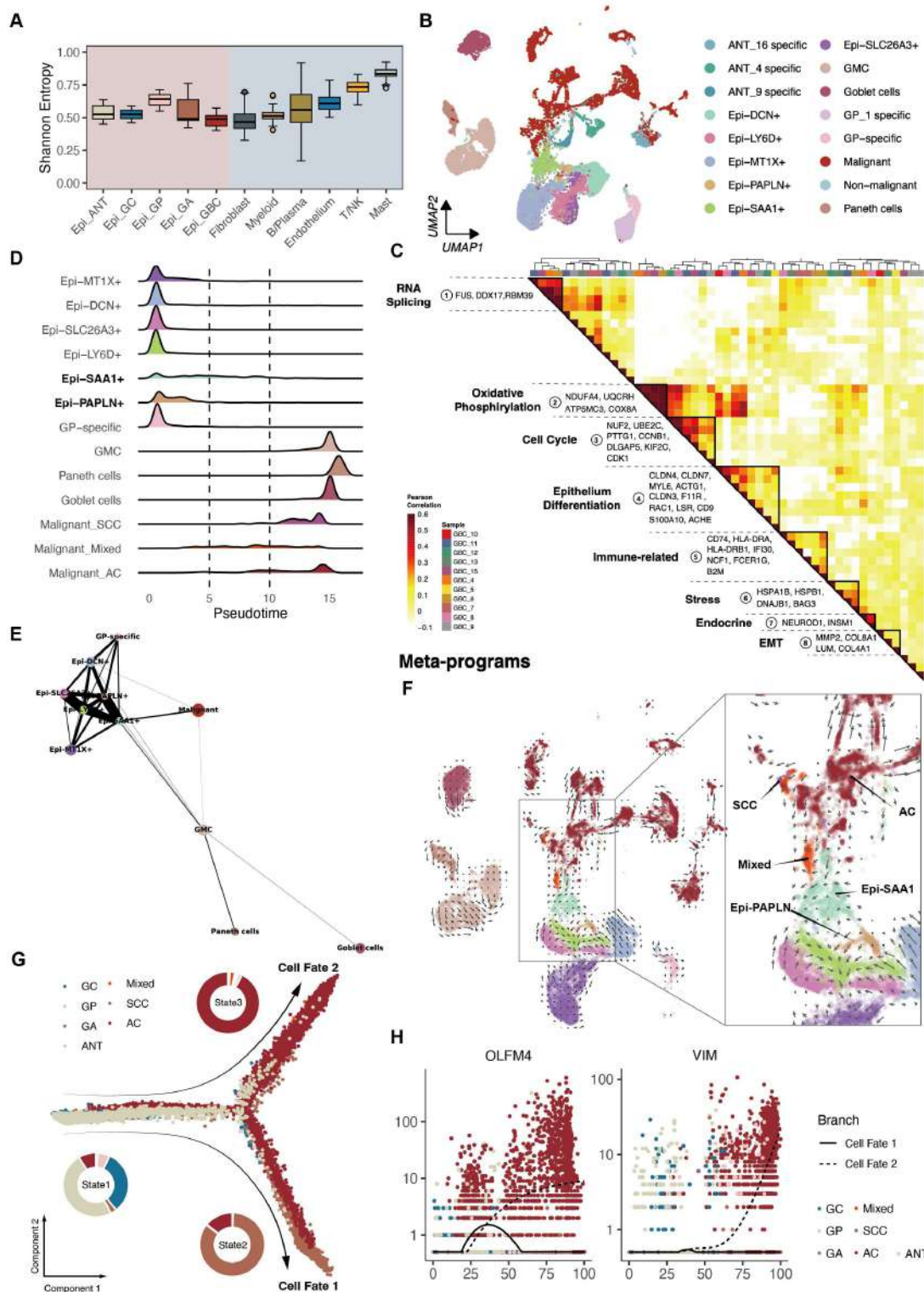
Non-negative matrix factorisation<sup>33</sup> was employed to uncover coherent genes preferentially co-expressed by malignant cells from AC. This analysis identified 8 meta-programmes with distinct biological activity among 11 AC samples (excluding GBC\_14 due to limited cells; figure 4C and online supplemental



**Figure 2** Characterisation of immune cell states throughout disease progression. (A) UMAP showing illustrating distinct subsets of T/NK cells. (B) Clonal expansion status of T cells shown as cell counts. (C) Proportion of the CD3-Cycling among different groups. (D) Heatmap displaying the clonal expansion of each T subset stratified by tissue subtypes. (E) Heatmap revealing clonal transitions between CD8-Tex and other clusters, stratified by tissue subtypes. (F) RNA velocity overlaid on UMAP of T cells, demonstrating potential transitional paths to CD8-Tex. (G) UMAP exhibiting subpopulations of Cycling T/NK cells. (H) Proportions of subpopulations of Cycling T/NK cells in each tissue subtype represented by Pie chart (top panel) and bar plot (bottom panel). (I) Immunofluorescence staining illustrating the dominance of cycling CD8-Tex in AC compared with the other two GBC subtypes. (J) UMAP showing subsets of myeloid cells identified in monocytes, macrophages and DCs. (K) Proportion of cDC3 among different groups. (L) Immunofluorescence staining confirms the presence of cDC3 in GBC. (M) Venn diagram illustrates six overlapping genes representing TAMs in GBC. (N) Heatmaps showing distinct expression patterns of function-associated signature genes among seven macrophage subsets in tissue subtypes. AC, adenocarcinomas; ANT, adjacent normal tissues; cDC, classical DC; DC, dendritic cell; GA, gallbladder adenomas; GBC, gallbladder cancer; GC, gallbladder cholecystitis; GP, gallbladder polyps; NK, natural killer; NKT, natural killer T; pDC, plasmacytoid DC; SCC, squamous cell carcinoma; TAM, tumour-associated macrophage; Teff, effector T; Tem, effector memory T; Tex, exhausted T; Tfh, follicular helper T; Tm, memory T; Treg, regulatory T; Trm, tissue-resident memory T; UMAP, Uniform Manifold Approximation and Projection.



**Figure 3** Phenotypic abundance and interactions of stromal cells. (A) UMAP of subpopulations of fibroblasts. (B) UMAP of subpopulations of endothelial cells. (C) Correlation analysis among TME subsets in all samples based on corresponding relative abundance. P values were calculated using the Spearman correlation test with Benjamini-Hochberg correction for multiple comparisons. (D) Inference of three ecotypes based on stromal cell compositions in the 43 samples. (E) Survival analysis of SC1 and SC3 gene signature across all cancer types from TCGA. Optimal cut points were determined using the survminer package. (F) Deconvolution of SC1 and SC3 cellular ecotypes in cohort by Pandey *et al.* (G) KEGG enrichment analysis of SC1 and SC3 ecotypes. (H) Violin plots showing significant differences in signature scores of SC1 and SC3 between patients with well-differentiated and poor-differentiated GBC. AC, adenocarcinomas; ANT, adjacent normal tissues; ECM, extracellular matrix; GA, gallbladder adenomas; GBC, gallbladder cancer; GC, gallbladder cholecystitis; GP, gallbladder polyps; KEGG, Kyoto Encyclopedia of Genes and Genomes; SCC, squamous cell carcinoma; TCGA, The Cancer Genome Atlas; TME, tumour microenvironment; UMAP, Uniform Manifold Approximation and Projection.



**Figure 4** Epithelium cell subtypes recapitulated dynamics of cell differentiation and specific gene expression. (A) Inter-epithelium heterogeneity across groups, measured by Shannon entropy, alongside other major cell types. (B) UMAP of subpopulations of epithelium. (C) Pearson correlation clustering of intratumour expression programmes generating eight meta-programmes. The colour was proportional to the absolute value of the correlation. (D) Distribution of epithelium subtypes during the transition, illustrated alongside pseudotime. (E) PAGA analysis of epithelium, where each dot represented an epithelium cluster. (F) RNA velocities overlaid on UMAP depicting potential transition paths from AC to mixed and SCC. Arrows on a grid show the RNA velocity field, and dots are coloured by meta-clusters. (G) Trajectory reconstruction of all epitheliums revealed three branches: pre-branch (before bifurcation), cell fate one branch, and cell fate two branch (after bifurcation). Pie charts indicated tissue subtype proportions of the cell fate 1/2 branches, respectively. (H) Two-dimensional plots showing the dynamic expression of OLFM4 and VIM in a pseudotime manner. AC, adenocarcinomas; ANT, adjacent normal tissues; GA, gallbladder adenomas; GBC, gallbladder cancer; GC, gallbladder cholecystitis; GP, gallbladder polyps; NK, natural killer; OLFM4, olfactomedin 4; PAGA, partition-based graph abstraction; SCC, squamous cell carcinoma; UMAP, Uniform Manifold Approximation and Projection; VIM, vimentin.

figure S4C and table S4). The eight programmes featured RNA splicing, oxidative phosphorylation, cell cycle, epithelial differentiation, immune-related genes such as Major Histocompatibility Complex, Class II, DR Alpha (HLA-DRA), stress response, endocrine-related genes and epithelial-mesenchymal transition (EMT), respectively. This classification unveiled the heterogeneity within AC, shedding new light on molecular classification-based therapy.

Monocle 2<sup>34</sup> analysis revealed that epithelium in GA displayed similarities to malignant epithelium, suggesting their proximity to terminal stages of differentiation. Epithelium in the mixed subtype appeared to be undergoing a differentiation process actively. Two clusters of non-malignant epitheliums in an intermediate state, Epi-SAA1 and Epi-PAPLN, were identified (figure 4D). Connectivity structures between these subtypes were determined as the previous report indicated,<sup>35</sup> which revealed a robust connection between Epi-SAA1 and both malignant and non-malignant epithelium, supporting Monocle 2 results (figure 4E). RNA velocity analysis suggested that Epi-SAA1 may play a pivotal role in facilitating the transition from non-malignant to malignant epithelium (figure 4F).

Epithelium pseudotime analysis unveiled three stages and two directions of cell fate. State 1 included precancerous samples. State 2 was predominantly made up of GA, with an emerging presence of GBC, while State 3 was primarily characterised by GBC. (figure 4G). In Cell Fate 1, the enrichment of pathways like the Wnt pathway and glandular development was commonly found elevated in adenoma cells. In Cell Fate 2, enriched signals were predominantly associated with cell adhesion and negative immune regulation pathways, which were associated with GBC differentiation. Robust activation of kinase signalling, oxidative phosphorylation, stress response and hypoxia signalling occurred throughout the ‘branching’ process (online supplemental figure S4D).

In summary, these findings revealed intricate gene expression programmes in AC and suggested a potential evolutionary similarity of GA and AC.

### OLFM4 elevated in malignant epithelium of GBC

Within epithelium, a chronological augmentation of specific genes, such as vimentin (VIM) and OLFM4 over the pseudo-temporal progression was observed, particularly within cells undergone Cell Fate 2, a trajectory directed towards GBC (figure 4H).

Remarkably, OLFM4, a glycoprotein belonging to the olfactomedin family, prominently featured in the malignant but normal epithelium (online supplemental figure S4E). OLFM4 exhibited the highest expression within malignant cells of AC, followed by Epi-SAA1 (figure 5A,B). Multiplex IHC confirmed the expression of OLFM4 in AC epithelium (figure 5C). An appreciable escalation of OLFM4 in tumour bulks was discovered compared with ANT (figure 5D). IHC staining on 136 GBC samples indicated a significantly worse prognosis in the OLFM4-High group (figure 5E,F). Subsequently, OLFM4 levels in plasma specimens from patients, encompassing GC (n=13), GP (n=12), GA (n=12) and AC (n=47), were examined via ELISA. The findings showed a substantial elevation of OLFM4 levels in GA and GBC, surpassing those in GC and GP (figure 5G), providing further evidence of tumour evolution from GA to GBC. This insight was further validated through IHC analysis of patient tissues (figure 5H). Given that GA, a recognised precancerous lesion for AC also serves as an indication for surgical intervention, OLFM4 may play a role as a marker for patient screening.

### Amplified TAM interaction and enhanced PD-(L)-1 signalling in OLFM4-positive GBC

To assess the spatial characteristics of GBC associated with OLFM4 expression, GeoMx DSP technology was employed on tissue microarrays (figure 6A). The DSP panel included 6 internal references and 43 proteins, which were categorised into 4 modules (online supplemental table S5). Based on the quantification of PanCK<sup>+</sup>OLFM4<sup>+</sup> cells, 90 samples were divided into two groups: OLFM4-negative (n=58) and OLFM4-positive (n=32) groups (online supplemental figure S5A). The protein expression level in one region of interest of each sample was quantified for further analysis. Upregulation of PD-L1, CD163 and fibronectin was observed in the OLFM4-positive group, while STING declined (figure 6C). Furthermore, tumour-killing effector molecules granzyme B and OX40L, along with CD45RO, were downregulated in the OLFM4-positive group (figure 6D and online supplemental figure S5B). Notably, the proportion of OLFM4<sup>+</sup> cells positively correlated with signals related to TAMs and PD-(L)1 signalling (figure 6E), further demonstrated in 15 independent GBC samples (online supplemental figure S5C,D).

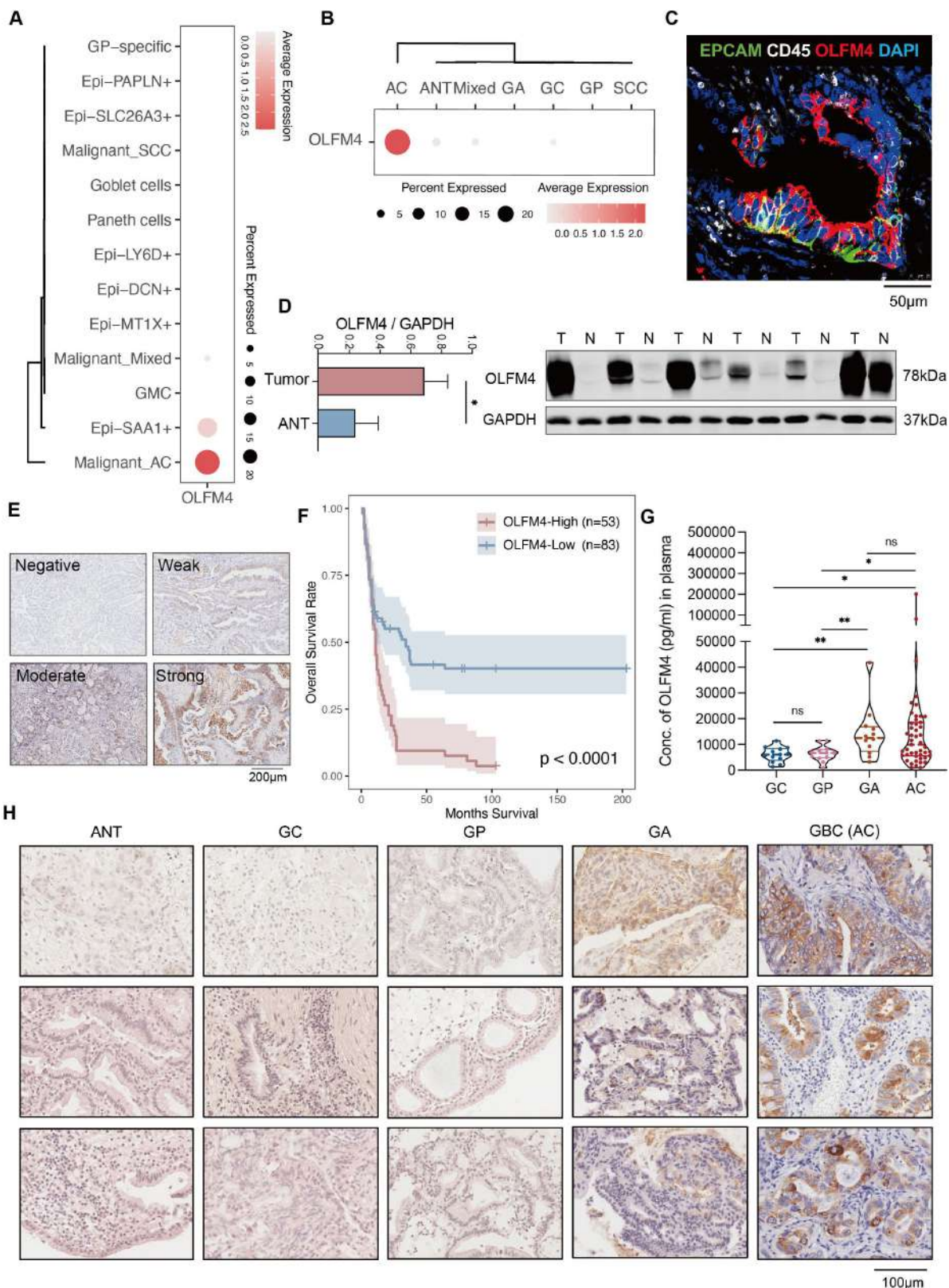
Furthermore, scRNA-seq data revealed that OLFM4<sup>+</sup> epithelium and TAMs interact significantly via ligand-receptor pairs such as ICAM1-integrin, CCL15-CCR1 and AREG-MMP9 (online supplemental figure S5E). The top predicted ligands were more abundant in OLFM4<sup>+</sup> epithelium than in OLFM4<sup>-</sup> epithelium, with TAMs expressing corresponding ligands like transforming growth factor beta 1 (TGFB1). Receptors in OLFM4<sup>+</sup> epithelium for tumour necrosis factor ligands from TAMs are linked to enhanced EMT and tumour invasion (online supplemental figure S5F). Additionally, programmed cell death protein 1 (PD-1) and PD-L1 levels were higher in the OLFM4-positive group, indicating a potential for increased tumour immunity interaction. Besides, PD-1 on lymphoid cells and PD-L1 on malignant epithelium were higher in the OLFM4-positive group, stratified by OLFM4 level in AC (online supplemental figure S6A,B).

To elucidate the potential role of OLFM4 in the immune microenvironment of GBC, loss-of-function studies using the GBC cell lines were conducted (online supplemental figure S6C). Cell Counting Kit-8 assays revealed that the proliferation of GBC was unaffected on OLFM4 knockdown (online supplemental figure S6D). Subcutaneous xenografts in immunodeficient NCG mice implanted with GBC-SD OLFM4 knockdown cells (sh-OLFM4), or control cells (CTRL) showed no significant difference in tumour growth (online supplemental figure S6E,F).

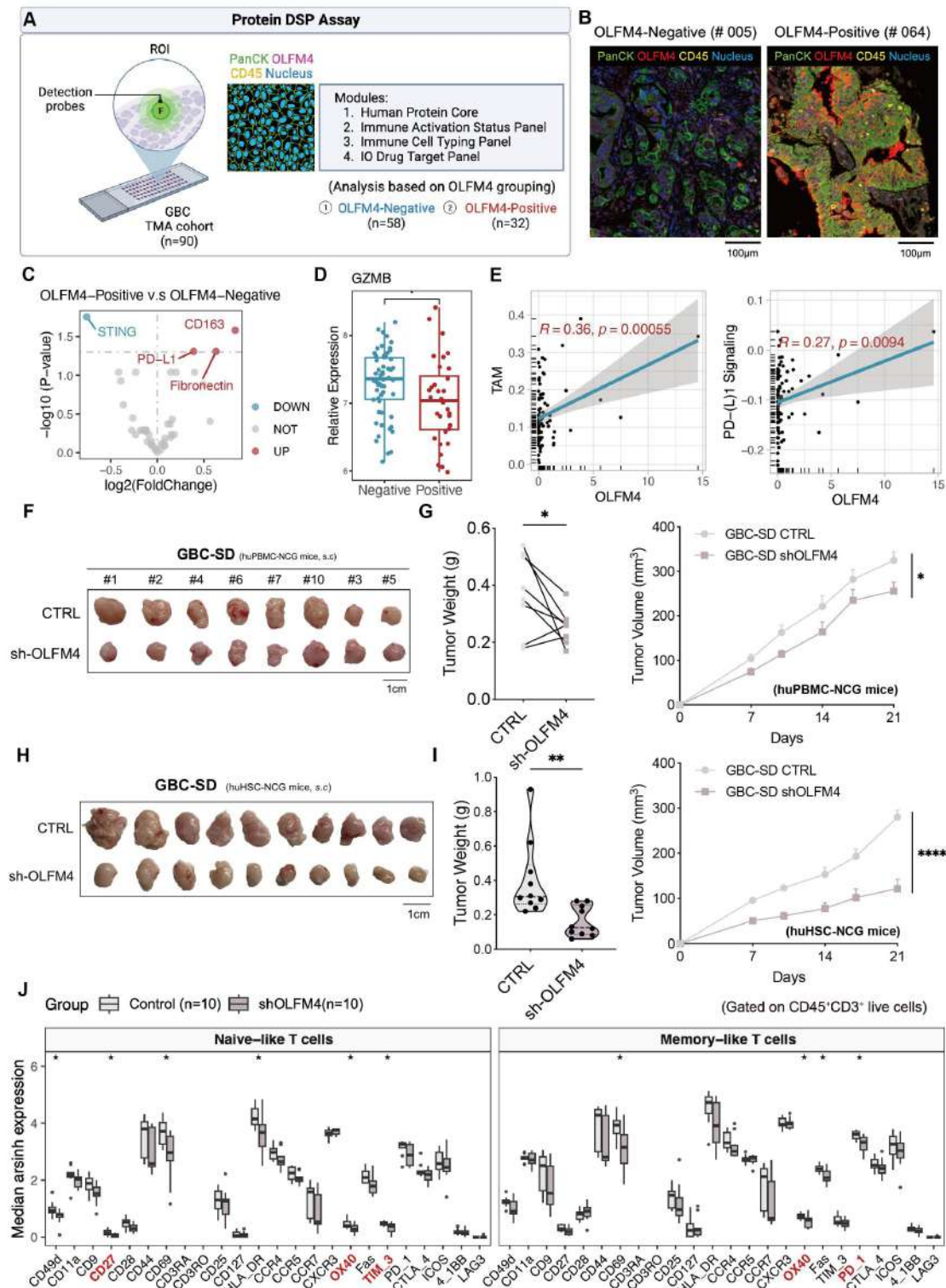
In a subcutaneous xenograft model with pre-activated human peripheral blood mononuclear cells (huPBMC-NCG model) to reconstruct immunity, tumour sizes of GBC-SD sh-OLFM4 were significantly smaller than that of CTRL group, suggesting that OLFM4 knockdown had a noticeable impact on tumour growth when the immune system of NCG mice was partially reconstructed (figure 6F,G). Human CD3<sup>+</sup>T cells were more abundant in GBC-SD sh-OLFM4 than in the CTRL group (online supplemental figure S6G). The disparities between the two models signify that OLFM4's role in tumour growth *in vivo* is imperative to the active involvement of the immune microenvironment.

Next, we used CD34<sup>+</sup> haematopoietic stem cell humanised mice (huHSC-NCG) to further confirm the role of OLFM4 in GBC progression. In this model, GBC-SD sh-OLFM4 xenografts were significantly smaller in tumour size and weight compared with CTRL, displaying greater disparity between groups than observed in the huPBMC-NCG model (figure 6H and I). Immune cell subset analysis using cytometry by time of

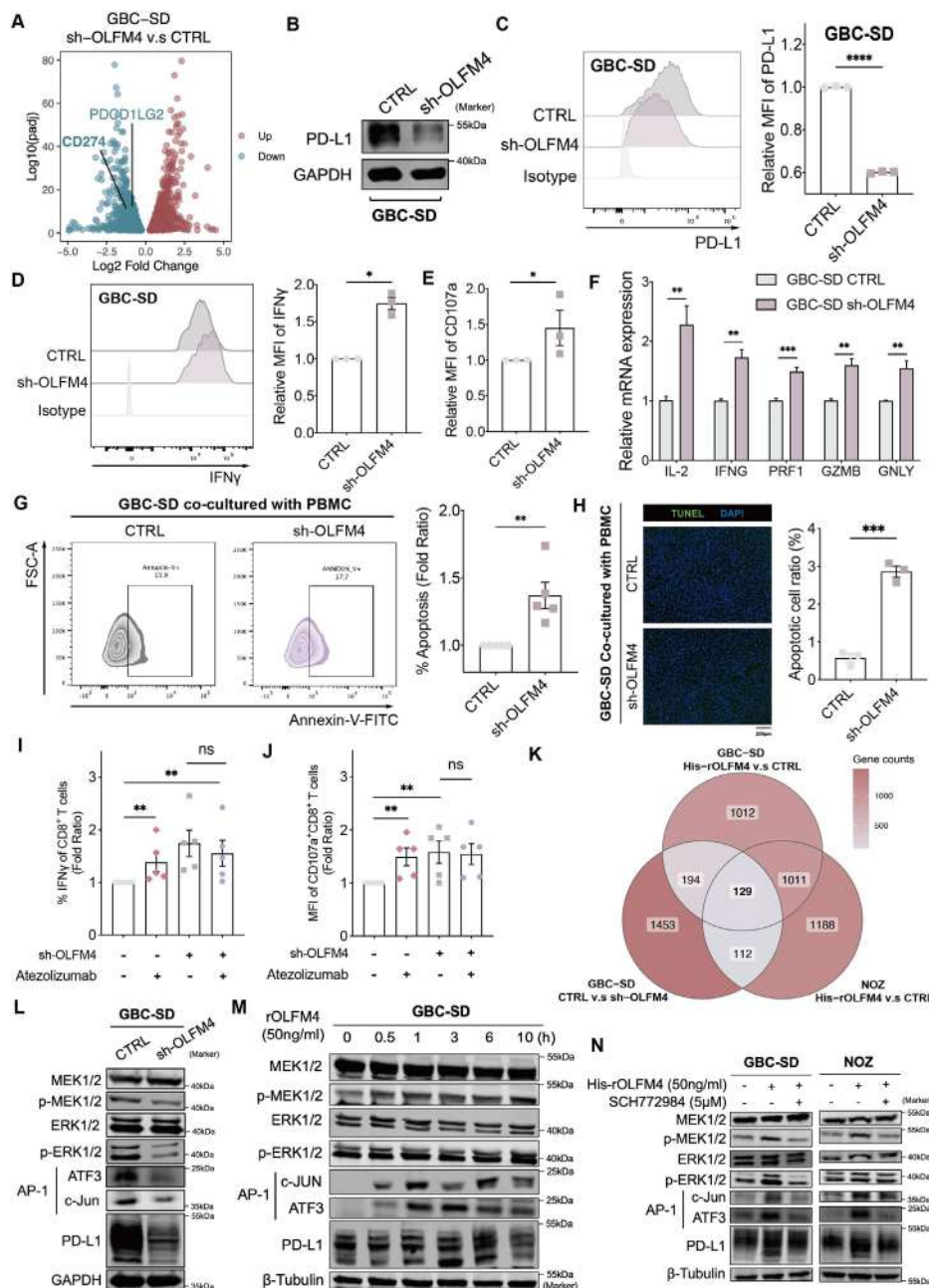




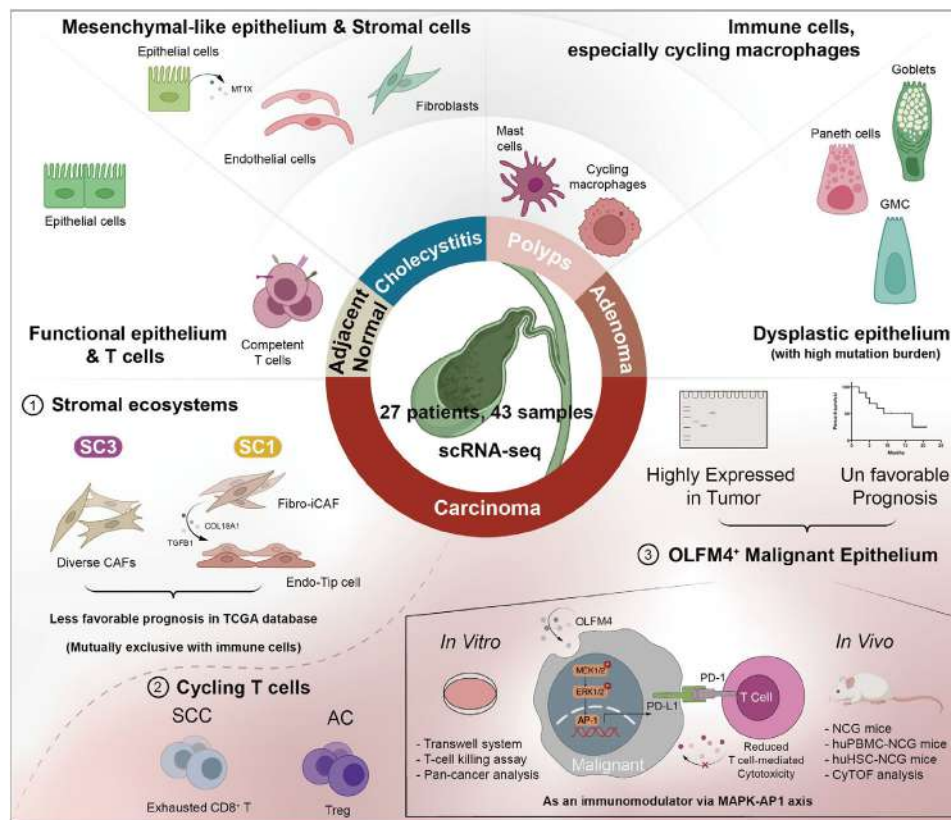
**Figure 5** OLFM4 was upregulated along with GBC progression and could be detected in peripheral blood. (A) Dot plot showing the expression of OLFM4 across epithelial subsets. (B) Dot plot depicting the expression of OLFM4 across tissue subtypes. (C) Multiplex IHC staining confirmed the specific expression of OLFM4 in gallbladder epithelium. (D) Western blot analysis of OLFM4 expression in GBC tissues compared with corresponding adjacent non-cancerous tissues. Quantification of results was displayed in the bar chart on the right. (E) Representative images of OLFM4 IHC staining, related to figure 5F. (F) Overall survival analysis of patients with GBC stratified by the OLFM4 expression level in 136 IHC samples. (G) Plasma levels of OLFM4 across the progression of gallbladder disease. (H) Representative IHC images showing the distribution pattern of OLFM4 (each group, n=3). AC, adenocarcinomas; ANT, adjacent normal tissues; GA, gallbladder adenomas; GBC, gallbladder cancer; GC, gallbladder cholecystitis; GP, gallbladder polyps; IHC, immunohistochemistry; OLFM4, olfactomedin 4; SCC, squamous cell carcinoma.



**Figure 6** OLFM4 positively correlated with TAM infiltration and suppressed antitumour T-cell immunity *in vivo*. (A) Overview of DSP assay design to investigate spatial heterogeneity of GBC. (B) Representative images of the OLFM4-negative sample and OLFM4-positive sample. (C) Volcanic plot of differentially expressed proteins among groups stratified by OLFM4 expression. (D) Box plot comparing GZMB expression between groups. (E) Pearson correlation of TAM and PD-(L)1 signalling with OLFM4 expression. (F) GBC-SD CTRL/sh-OLFM4 cells were injected subcutaneously into the right flank of NCG mice and transferred with activated PBMC to build the xenograft model. Gross morphology of tumours in the huPBMC-NCG model (CTRL, n=8; sh-OLFM4, n=8). (G) Tumour volume (left) and tumour weight (right) of F were calculated. (H) Subcutaneous injection of GBC-SD CTRL/sh-OLFM4 cells into CD34<sup>+</sup> humanised mice to obtain tumour xenografts. Gross morphology of tumours in the huHSC-NCG model (CTRL, n=10; sh-OLFM4, n=10). (I) Tumour volume (left) and tumour weight (right) of H were calculated. (J) Relative expression levels of functional markers of naive-like<sup>+</sup> T cells and memory-like<sup>+</sup> T cells with tumours in H were determined by cytometry. (J) Median expression levels of functional markers of naive-like<sup>+</sup> T cells and memory-like<sup>+</sup> T cells with tumours in H were determined by cytometry by time of flight; DSP, digital spatial profiler; GBC, gallbladder cancer; GZMB, granzyme B; OLFM4, olfactomedin 4; PBMC, peripheral blood mononuclear cells; PD-L1, programmed cell death ligand 1; ROI, region of interest; TAM, tumour-associated macrophage.



**Figure 7** OLFM4 impaired immune response via the MAPK-AP1-PD-L1 axis. (A) Volcano plot for the differentially expressed genes after OLFM4-knockdown in GBC-SD (adjusted  $p < 0.05$ ). (B) Expression levels of PD-L1 determined by western blot in OLFM4 knockdown GBC cells. (C) PD-L1 expression in OLFM4-knockdown GBC-SD cells analysed by flow cytometry (left). Bar plot comparing the relative mean fluorescence intensity (MFI) of PD-L1 (right). (D) Intracellular IFN- $\gamma$  of CD8 $^+$ T cells detected by flow cytometry after co-culturing with GBC-SD CTRL/sh-OLFM4 cells for 72 hours. Bar plot compared the relative MFI of IFN- $\gamma$  ( $n=3$ , \* $p < 0.05$ ). (E) Bar plot comparing the relative MFI of CD107a from CD8 $^+$ T cells co-cultured with GBC-SD CTRL/sh-OLFM4 cells ( $n=3$ , \* $p < 0.05$ ). (F) Quantitative RT-PCR performed to detect IL-2 (interleukin-2), IFN-G (IFN- $\gamma$ ), PRF1 (perforin-1), GZMB (granzyme) and GNLV (granulysin) in activated PBMCs co-cultured with GBC-SD CTRL/sh-OLFM4 cells ( $n=3$ , \* $p < 0.05$ ). (G) Activated PBMCs were co-cultured with GBC-SD CTRL/sh-OLFM4 cells for 72 hours at the ratio of 4:1. The PBMCs were collected and stained with FITC-annexin V, then subjected to flow cytometry analysis. The percentage of apoptotic cells was analysed (right panel). (H) Cell apoptosis in PBMC co-cultured GBC cells evaluated by transferase dUTP nick end labelling assay (left panel). The apoptotic cell ratios were shown (right panel). (I) Relative proportion of IFN- $\gamma^+$ CD8 $^+$ T cells following a 72-hour co-culture with GBC-SD CTRL/sh-OLFM4 cells, in the presence or absence of atezolizumab (10  $\mu$ g/mL) treatment. (J) Relative MFI of CD107a $^+$ CD8 $^+$ T cells following a 72-hour co-culture with GBC-SD CTRL/sh-OLFM4 cells, in the presence or absence of atezolizumab (10  $\mu$ g/mL) treatment. (K) Venn diagram illustrating overlapping genes in GBC cell lines with OLFM4 stimulation or knockdown. (L) Western blot analyses of total MEK1/2, p-MEK1/2, total ERK1/2, p-ERK1/2, AP-1 complex and PD-L1 levels in GBC-SD CTRL and GBC-SD sh-OLFM4. (M) Western blot analyses of total MEK1/2, p-MEK1/2, total ERK1/2, p-ERK1/2, AP-1 complex and PD-L1 levels in GBC-SD treated with OLFM4 (50 ng/mL). (N) Western blot analyses of total MEK1/2, p-MEK1/2, total ERK1/2, p-ERK1/2, AP-1 complex and PD-L1 levels in GBC-SD and NOZ treated with OLFM4 (50 ng/mL) combined with or without ERK1/2 pathway inhibitor (SCH772984). GBC, gallbladder cancer; IFN, interferon; mRNA, messenger RNA; OLFM4, olfactomedin 4; PBMC, peripheral blood mononuclear cells; PD-L1, programmed cell death ligand 1; RT-PCR, reverse transcription polymerase chain reaction.



**Figure 8** Summary of immune features and dynamics of the ecosystem among gallbladder diseases within this study. AC, adenocarcinoma; CAF, cancer-associated fibroblast; CyTOF, cytometry by time of flight; GMC, glandular mucosal cell; OLFM4, olfactomedin 4; PD-1, programmed cell death protein 1; PD-L1, programmed cell death ligand 1; SCC, squamous cell carcinoma; scRNA-seq, single-cell RNA sequencing; TCGA, The Cancer Genome Atlas.

flight focused primarily on T-cell markers identified four T-cell subsets: naive-like, memory-like, natural killer T (NKT) and CD161<sup>+</sup>T cells (online supplemental figure S6H,I and table S6). In the OLFM4-CTRL group, functionally exhausted T cells marked by CD27 and OX40 belonging to the tumour necrosis factor receptor superfamily were elevated, and ICB genes like T-cell immunoglobulin and mucin domain 3 (TIM-3) and PD-1 were upregulated (figure 6J and online supplemental figure S6J). Furthermore, multiplex IHC staining highlighted increased TAM infiltration in OLFM4-CTRL tumours (online supplemental figure S6K,L). In summary, our data indicated that OLFM4-positive tumour cells regulated immune response in GBC by intensive crosstalk with TAM and modulating T-cell function.

### OLFM4 impaired immune responses by upregulating PD-L1 through MAPK-AP1 axis

To delineate the mechanism behind OLFM4-mediated immune response, we conducted bulk RNA-seq analysis on GBC-SD sh-OLFM4 and CTRL cells. Among immune checkpoint molecules, CD274, encoding PD-L1, emerged as the key candidate (figure 7A). OLFM4 knockdown led to a reduction in PD-L1 expression at both messenger RNA (mRNA) and protein levels, with or without exogenous interferon (IFN)- $\gamma$  stimulation (figure 7B,C, online supplemental figure S7A–C). Conversely, introducing recombinant OLFM4 protein (50 ng/mL) induced a time-dependent increase in PD-L1 mRNA and protein expression within 4 hours (online supplemental figure S7D,E). Notably, this time-dependent elevation of PD-L1 by OLFM4 was demonstrated in several cancer cell lines, including human colorectal

AC (SW620 and HT-29), human gastric cancer (MGC803 and AGS) and human pancreatic cancer (SW1990 and PANC-1) cell lines (online supplemental figure S7F). These data strongly suggest that OLFM4 may substantially impact the expression of PD-L1 in multiple cancer types beyond GBC.

T cell-mediated killing assay was conducted to assess whether OLFM4-regulated PD-L1 influenced T-cell function *in vitro*. Cytotoxic T-cell activities, as indicated by effectors (IFN- $\gamma$ , CD107a and granzyme), were upregulated when co-cultured with GBC shOLFM4 (figure 7D, E and F; online supplemental table S7). Simultaneously, apoptotic tumour cells increased in OLFM4-knockdown cells after co-culturing with PBMC, as detected by flow cytometry and terminal deoxynucleotidyl transferase dUTP nick end labelling assay (figure 7G,H). Treatment with atezolizumab (anti-PD-L1 antibody, 10  $\mu$ g/mL) reversed T-cell-mediated killing enhancement in a co-culture system. While administration of atezolizumab to the sh-OLFM4 group did not result in enhanced IFN- $\gamma$  secretion by CD8<sup>+</sup>T cells (figure 7I,J, online supplemental figure S7G,H).

Given the *in vitro* and *in vivo* induction of PD-L1 expression by OLFM4, we investigated the underlying mechanism. Tunica-mycin inhibition and cycloheximide chase assays were conducted to assess the role of OLFM4 in PD-L1 expression, revealing no significant influence at the protein level (online supplemental figure S7I,J). Then, transcriptomic profiling of GBC cell lines identified 129 downstream genes enriched in the MAPK cascade and AP-1 pathway (figure 7K and online supplemental figure S7K). In OLFM4-knockdown GBC-SD cells, a significant decrease in p-MEK1/2 and p-ERK1/2 expression within the

MAPK pathway was confirmed by western blot. Subsequently, protein levels of the downstream transcription factor AP-1 and target gene PD-L1 were also reduced (figure 7L). Likewise, the introduction of exogenous OLFM4 to GBC cell lines GBC-SD and NOZ resulted in elevated levels of p-MEK1/2, AP-1 and PD-L1 (figure 7M and online supplemental figure S7L). Conversely, ERK1/2 pathway inhibitor SCH72984 could block the activation effect of recombinant OLFM4 (figure 7N).

Collectively, these findings underscore the pivotal role of OLFM4 in activating the PD-1/PD-L1 axis and subsequently inhibiting T-cell function within but possibly not limited to GBC.

## DISCUSSION

GBC manifests as a highly lethal disease with wide geographical prevalence, which is closely linked to persistent inflammation. It is noteworthy that the onset of GBC does not always correlate with cholelithiasis, and polypoid lesions, particularly adenomas, are found associated with GBC progression. Despite the well-known pathological sequence of GBC, limited investigations have explored the association of their ecosystem with disease progression. Our study comprehensively investigated the cellular landscape and molecular features of gallbladder diseases of cholecystitis, polypoid lesions and GBC by multiomics including WGS, bulk RNA-seq and scRNA-seq to decipher the microenvironment heterogeneity in disease evolution (figure 8).

Exploration of the ecosystem offers insights into GBC pathogenesis and targeted molecular therapy. ScRNA-seq revealed a cell atlas of 69 subpopulations, showcasing diverse cell types within gallbladder diseases. Distribution patterns highlighted similarities between ANT and GC, as well as GP and GA. GBC displayed an enrichment of immune suppressive and stromal cells. Notably, the variation in TCR expansion patterns, cycling T cells, and myeloid cells significantly inform potential immunotherapeutic strategies for gallbladder diseases by targeting T-cell populations key to disease progression.<sup>36</sup> Stromal cell analysis unveiled significant heterogeneity, categorising stromal cells into SC1, SC2 and SC3, with SC1 and SC3 predominantly in GBC samples. SC1-like patients exhibited enriched pro-tumour pathways and poorer outcomes, which could serve as an accountable prognosis indicator. In addition, epithelium among AC samples could be classified into eight meta-programmes according to molecular features indicating distinct biological functions such as immune response, EMT, etc, unveiling the heterogeneity within AC and offering valuable insights into molecular classification-based therapy.

Notably, a significant elevation of OLFM4 expression was found in malignant GBC epithelium by scRNA-seq data. Subsequent investigations revealed that OLFM4-positive cells were intensively enriched in AC and moderately elevated in GA tissues. Given that OLFM4 could be secreted to extracellular space, our data also indicated a significant elevation of OLFM4 in the serum of GA and AC patients. Pseudotime analysis of epithelium evolution also revealed overlapped states of GA and AC, suggesting a potential evolutionary similarity between them. Considering GA is a recognised precancerous lesion of GBC, this similarity suggests a plausible evolution of GBC from GA.

Previous research has delineated an upregulation of OLFM4 across the continuum from GC, dysplasia to GBC and associated with patient survival, aligning with our findings.<sup>37</sup> While this study mainly highlighted the sensitisation of GBC to cisplatin on OLFM4 depletion, we focused on the role of OLFM4<sup>+</sup>GBC cells on the immune microenvironment, particularly the inhibitory effect of OLFM4 on modulating T cells via regulating PD-L1

expression. Spatial proteomics revealed that the OLFM4-positive group exhibits strong interactions with TAM and enhanced PD-(L)1 signalling, with further *in vitro* and *in vivo* experiments confirming OLFM4's role in modulating PD-L1 and recruiting TAM. Besides, cancer-associated fibroblast (CAF) is prominent components of the microenvironment of GBC. iCAF has been linked to tumour escape via chemoattraction, participating in promoting angiogenesis and induction of immunosuppressive cells.<sup>38,39</sup> ScRNA-seq data also indicated SC1, predominantly comprising iCAFs and Endo-Tip cells within GBC, was associated with a worse prognosis. Given the intricate interplay among malignant cells, immune constituents and CAFs, it is plausible that OLFM4 may modulate the functionality of CAFs, either directly or indirectly. Moreover, OLFM4 upregulated PD-L1 in other cell lines such as gastric cell lines and pancreatic cell lines, suggesting its potential role on TME in more tumours other than AC, possibly tumours of epithelial origin.

Our study sheds light on key molecular elements of GBC pathogenesis but needs further exploration due to limitations like small sample sizes in certain GBC types, affecting the generalisability of our findings. The specific functions of identified T cells and myeloid cells in gallbladder disease warrant further study. Furthermore the function of OLFM4 in the progression of GBC should be further elucidated to evaluate its potential as a prognosis marker or therapeutic target for gallbladder diseases. In summary, the intricate interaction between different cell subtypes constructs a complex ecosystem promoting GBC pathogenesis, further research was needed to make breakthroughs in understanding the diseases and ultimately improve clinical outcomes for patients.

## Author affiliations

<sup>1</sup>National Center for Liver Cancer, Third Affiliated Hospital of Naval Medical University (Second Military Medical University), Shanghai, China

<sup>2</sup>International Cooperation Laboratory on Signal Transduction, Third Affiliated Hospital of Naval Medical University (Second Military Medical University), Shanghai, China

<sup>3</sup>Department I of Biliary Tract Surgery, Third Affiliated Hospital of Naval Medical University (Second Military Medical University), Shanghai, China

<sup>4</sup>Department of Pathology, Third Affiliated Hospital of Naval Medical University (Second Military Medical University), Shanghai, China

<sup>5</sup>Fudan University Shanghai Cancer Center, Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China

<sup>6</sup>Department of Laboratory Diagnosis, Third Affiliated Hospital of Naval Medical University (Second Military Medical University), Shanghai, China

<sup>7</sup>Engineering Center for Translational Medicine of Precision Cancer Immunodiagnosis and Therapy, YuceBio Technology Co Ltd, Shenzhen, China

<sup>8</sup>Key Laboratory of Gene Engineering of the Ministry of Education, Institute of Healthy Aging Research, School of Life Sciences, Sun-Yat-sen University, Guangzhou, China

**Acknowledgements** We appreciate the support from the Shanghai Key Laboratory of Hepatobiliary Tumor Biology and the Ministry of Education Key Laboratory on Signaling Regulation and Targeting Therapy of Liver Cancer, Naval Medical University (Second Military Medical University), Shanghai, China. We also want to thank Shan-Hua Tang, Dan Cao, Ji Hu and Lu Chen for their technical assistance.

**Contributors** HH, SC, YY, HW and WW designed the study. HH, YY, YQ and WW developed the methodology. HH, SL, ZF, WH and WL acquired data. HH, SC, YD, ZF and WW analysed and interpreted data. HH and WW wrote the manuscript. HH, SC, HW and WW reviewed the manuscript. SL, XS, YS, JY, CQ, CZ, MZ and XW provided technical and material support. DW and JW performed the DSP experiment. HW and WW supervised the study. As the guarantor, WW is fully responsible for the overall work and the conduct of the study.

**Funding** This work was supported by the National Natural Science Foundation of China (82173146, 81830054, 81988101 and 82072600). The Major Program of the Shanghai Education Committee (2019010700007E00065), the plan of the Shanghai Municipal Health Commission (2022XD036), the National Science Foundation of Shanghai (21JC1406600, 20DZ2291900 and 23692123600). The Shanghai Science and Technology Committee Rising-Star Program (21QA1411700) also provided funding to conduct this project.

**Competing interests** None declared.

**Patient consent for publication** Consent obtained directly from patient(s).

**Ethics approval** The study adhered to the principles of the Declaration of Helsinki and was approved by the ethics committee of Eastern Hepatobiliary Surgery Hospital (EHBH) (No. EHBHKY2023-K049-P001). Participants gave informed consent to participate in the study before taking part.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available upon reasonable request.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iDs

Hongyang Wang <http://orcid.org/0000-0002-4709-3334>

Wen Wen <http://orcid.org/0000-0002-7835-922X>

#### REFERENCES

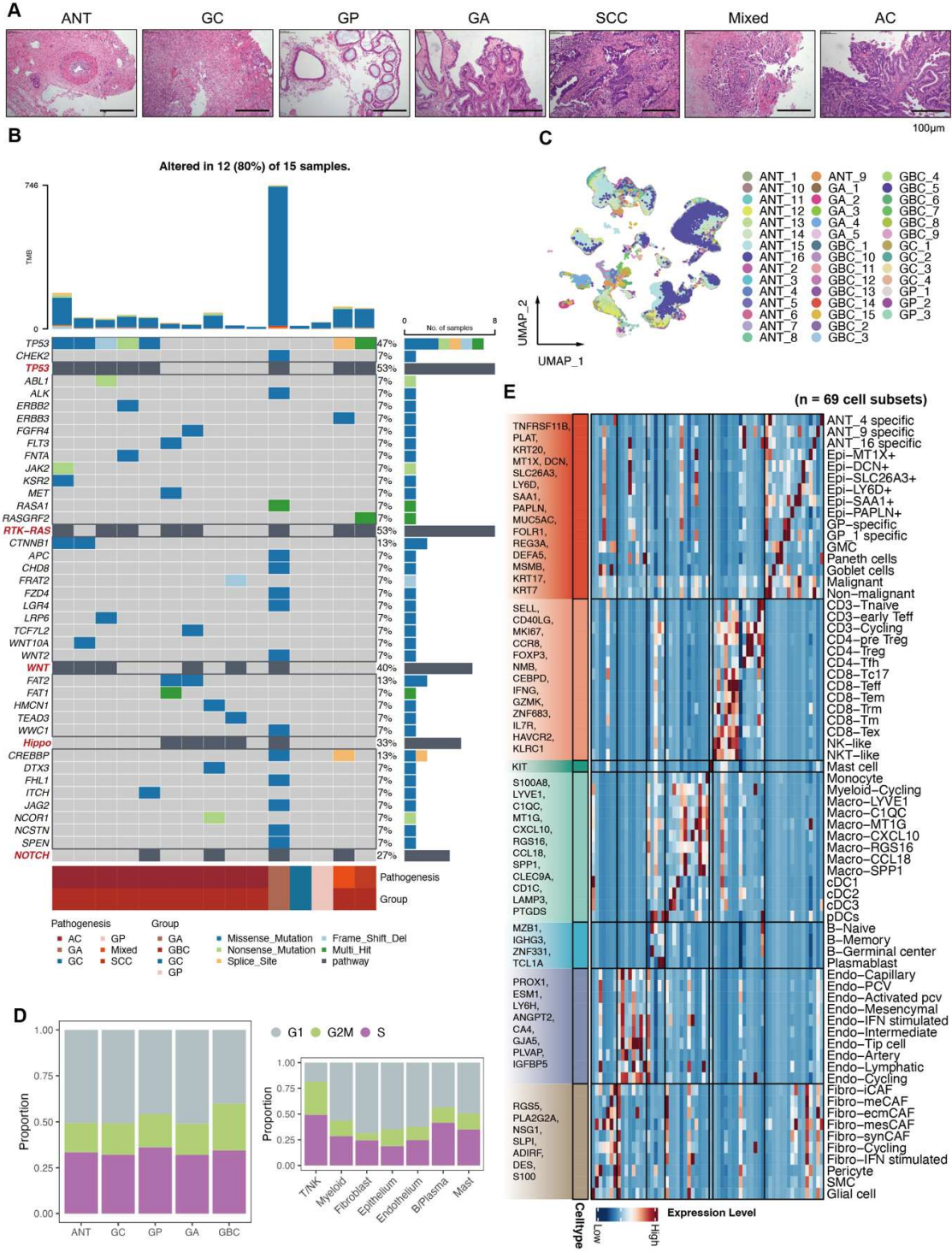
- Huang J, Patel HK, Boakye D, *et al.* Worldwide distribution, associated factors, and trends of gallbladder cancer: A global country-level analysis. *Cancer Letters* 2021;521:238–51.
- Sung H, Ferlay J, Siegel RL, *et al.* Global cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021;71:209–49.
- Roa JC, Garcia P, Kapoor VK, *et al.* Gallbladder cancer. *Nat Rev Dis Primers* 2022;8:69.
- Javle MM, Hainsworth JD, Swanton C, *et al.* Pertuzumab + Trastuzumab for Her2-positive metastatic biliary cancer: preliminary data from MyPathway. *JCO* 2017;35:402.
- Javle M, Zhao H, Abou-Alfa GK, *et al.* Systemic therapy for gallbladder cancer. *Chin Clin Oncol* 2019;8:44.
- Roa JC, Basturk O, Adsay V, *et al.* Dysplasia and carcinoma of the gallbladder: pathological evaluation, sampling, differential diagnosis and clinical implications. *Histopathology* 2021;79:2–19.
- Roa JC, Basturk O, Adsay V. Dysplasia and carcinoma of the gallbladder: pathological evaluation, sampling, differential diagnosis and clinical implications. *Histopathology* 2021;79:2–19.
- Jang K-T, Ahn S. Tumoral versus flat intraepithelial Neoplasia of Pancreatobiliary tract, gallbladder, and Ampulla of Vater. *Arch Pathol Lab Med* 2016;140:429–36.
- Wistuba II, Gazdar AF. Gallbladder cancer: lessons from a rare tumour. *Nat Rev Cancer* 2004;4:695–706.
- Hsing AW, Gao Y-T, Han T-Q, *et al.* Gallstones and the risk of biliary tract cancer: a population-based study in China. *Br J Cancer* 2007;97:1577–82.
- Hemminki K, Försti A, Hemminki O, *et al.* Long-Term incidence and survival trends in cancer of the gallbladder and extrahepatic bile ducts in Denmark, Finland, Norway and Sweden with Etiological implications related to Thorotrast. *Int J Cancer* 2022;151:200–8.
- Kapoor VK. Etiology and Pathogenesis of Gall Bladder Cancer. Springer Singapore, 2021:35–55.
- Shaffer EA. Epidemiology and risk factors for gallstone disease: has the paradigm changed in the 21st century *Curr Gastroenterol Rep* 2005;7:132–40.
- Lazcano-Ponce EC, Miquel JF, Muñoz N, *et al.* Epidemiology and molecular pathology of gallbladder cancer. *CA Cancer J Clin* 2001;51:349–64.
- Dutta U, Bush N, Kalsi D, *et al.* Epidemiology of gallbladder cancer in India. *Chin Clin Oncol* 2019;8:33.
- Espinoza JA, Bizama C, Garcia P, *et al.* The inflammatory inception of gallbladder cancer. *Biochim Biophys Acta* 2016;1865:245–54.
- Randi G, Franceschi S, La Vecchia C. Gallbladder cancer worldwide: geographical distribution and risk factors. *Int J Cancer* 2006;118:1591–602.
- Iyer P, Barreto SG, Sahoo B, *et al.* Non-Typhoidal salmonella DNA traces in gallbladder cancer. *Infect Agent Cancer* 2016;11:12.
- Muraki T, Pehlivanoglu B, Memis B, *et al.* Pancreatobiliary Maljunction-associated gallbladder cancer is as common in the West, shows distinct Clinicopathologic characteristics and offers an invaluable model for anatomy-induced reflux-associated Physio-chemical carcinogenesis. *Ann Surg* 2022;276:e32–9.
- Barahona Ponce C, Scherer D, Brinster R, *et al.* Gallstones, body mass index, C-Reactive protein, and gallbladder cancer: Mendelian randomization analysis of Chilean and European genotype data. *Hepatology* 2021;73:1783–96.
- Cao L, Bridle KR, Shrestha R, *et al.* Cd73 and PD-L1 as potential therapeutic targets in gallbladder cancer. *Int J Mol Sci* 2022;23:1565.
- Li M, Liu F, Zhang F, *et al.* Genomic Erbb2/Erbb3 mutations promote PD-L1-mediated immune escape in gallbladder cancer: a whole-Exome sequencing analysis. *Gut* 2019;68:1024–33.
- Song X, Hu Y, Li Y, *et al.* Overview of current targeted therapy in gallbladder cancer. *Signal Transduct Target Ther* 2020;5:230.
- Cai W, Yuan Y, Ge W, *et al.* EGFR target therapy combined with Gemox for advanced biliary tract cancers: a meta-analysis based on Rcts. *J Cancer* 2018;9:1476–85.
- Ramanathan RK, Belani CP, Singh DA, *et al.* A phase II study of Lapatinib in patients with advanced biliary tree and hepatocellular cancer. *Cancer Chemother Pharmacol* 2009;64:777–83.
- Pandey A, Stawiski EW, Durinck S, *et al.* Integrated Genomic analysis reveals Mutated Elf3 as a potential gallbladder cancer vaccine candidate. *Nat Commun* 2020;11:4225.
- Nepal C, Zhu B, O'Rourke CJ, *et al.* Integrative molecular Characterisation of gallbladder cancer reveals micro-environment-associated subtypes. *J Hepatol* 2021;74:1132–44.
- Li M, Zhang Z, Li X, *et al.* Whole-Exome and targeted gene sequencing of gallbladder carcinoma identifies recurrent mutations in the Erbb pathway. *Nat Genet* 2014;46:872–6.
- Liu L, Michowski W, Kolodziejczyk A, *et al.* The cell cycle in stem cell proliferation, Pluripotency and differentiation. *Nat Cell Biol* 2019;21:1060–7.
- Demir IE, Friess H, Ceyhan GO. Neural plasticity in Pancreatitis and Pancreatic cancer. *Nat Rev Gastroenterol Hepatol* 2015;12:649–59.
- Memis B, Saka B, Roa JC, *et al.* Eosinophilic Cholecystitis and Eosinophils in gallbladder injuries: A Clinicopathological analysis of 1050 Cholecystectomies. *Diagnostics (Basel)* 2023;13:2559.
- Goveia J, Rohlenova K, Taverna F, *et al.* An integrated gene expression landscape profiling approach to identify lung tumor endothelial cell heterogeneity and angiogenic candidates. *Cancer Cell* 2020;37:21–36.
- Kotliar D, Veres A, Nagy MA, *et al.* Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. *Life* 2019;8:e43803.
- Qiu X, Mao Q, Tang Y, *et al.* Reversed graph Embedding resolves complex single-cell Trajectories. *Nat Methods* 2017;14:979–82.
- Wolf FA, Hamey FK, Plass M, *et al.* PAGA: graph abstraction reconciles clustering with trajectory inference through a Topology preserving map of single cells. *Genome Biol* 2019;20:59.
- Chen P, Wang Y, Li J, *et al.* Diversity and Intratumoral heterogeneity in human gallbladder cancer progression revealed by Single-Cell RNA sequencing. *Clin Transl Med* 2021;11:e462.
- Lin Z, Yang S, Zhou Y, *et al.* Olfm4 depletion sensitizes gallbladder cancer cells to cisplatin through the Arl6ip1/Caspase-3 axis. *Transl Oncol* 2022;16:101331.
- Tsoumakidou M. The advent of immune stimulating Cafs in cancer. *Nat Rev Cancer* 2023;23:258–69.
- Monteran L, Erez N. The dark side of fibroblasts: cancer-associated fibroblasts as mediators of immunosuppression in the tumor Microenvironment. *Front Immunol* 2019;10:1835.

## Supplemental information

### Comprehensive single-cell analysis deciphered microenvironmental dynamics and immune regulator Olfactomedin 4 in pathogenesis of gallbladder cancer

Supplementary Figure S1	-----	P2-3
Supplementary Figure S2	-----	P4-5
Supplementary Figure S3	-----	P6-8
Supplementary Figure S4	-----	P9-10
Supplementary Figure S5	-----	P11-13
Supplementary Figure S6	-----	P14-16
Supplementary Figure S7	-----	P17-19
Supplementary materials and methods	-----	P20-34

Figure S1

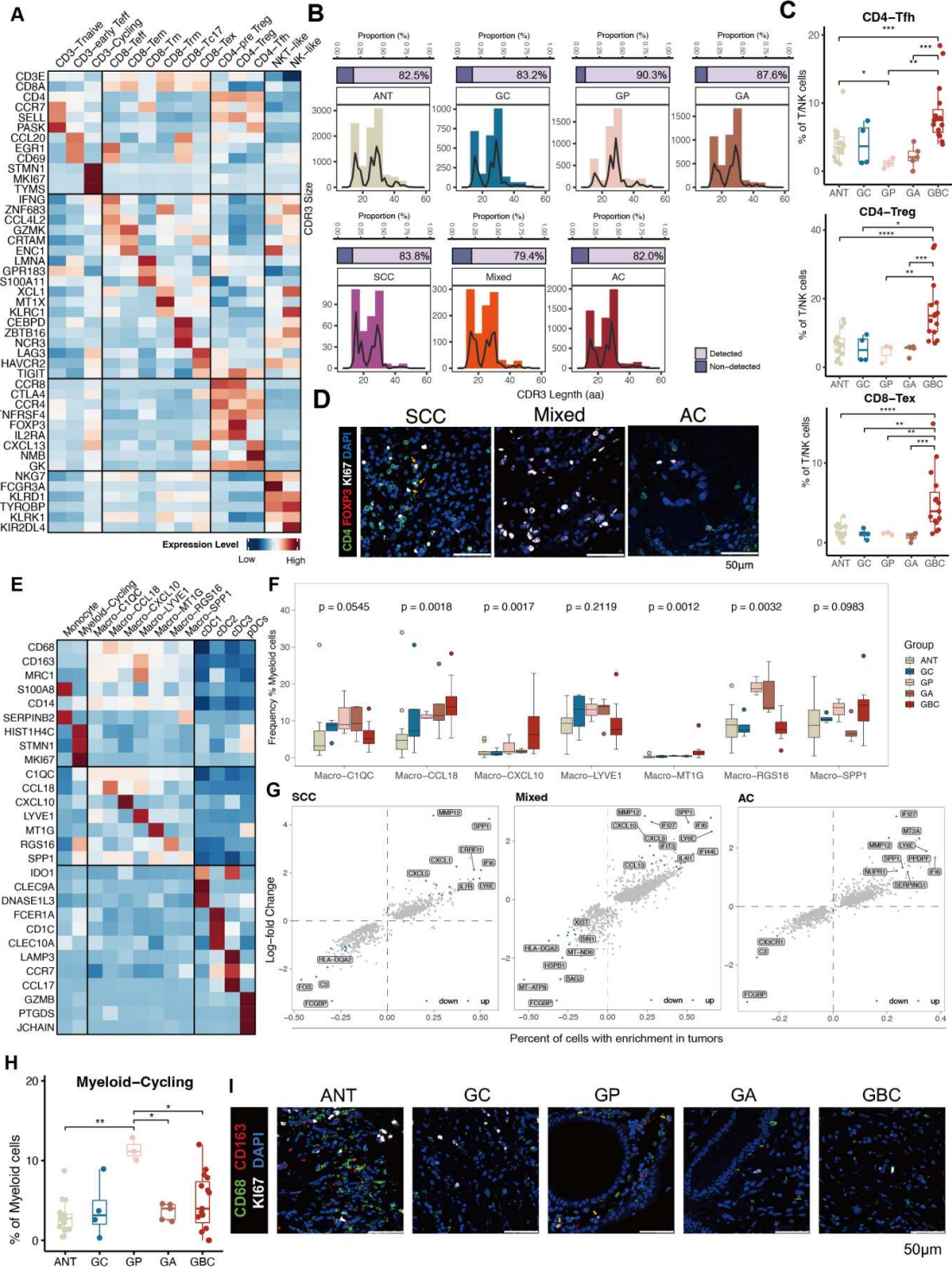




**Figure S1. Molecular landscape heterogeneity in human GBC and associated precancerous lesions.**

- A. Representative histological images representing 7 tissue subtypes.
- B. Oncoplot representation of the mutational landscape of 15 specimens detected through bulk whole-genome sequencing and mutational variant calling.
- C. UMAP visualization of individual samples.
- D. The phases of the cell cycle and their distribution. Proportions across groups (left) and proportions across major cell types (right).
- E. Heatmap displaying average expressions of marker genes for all cell subsets (n=69).

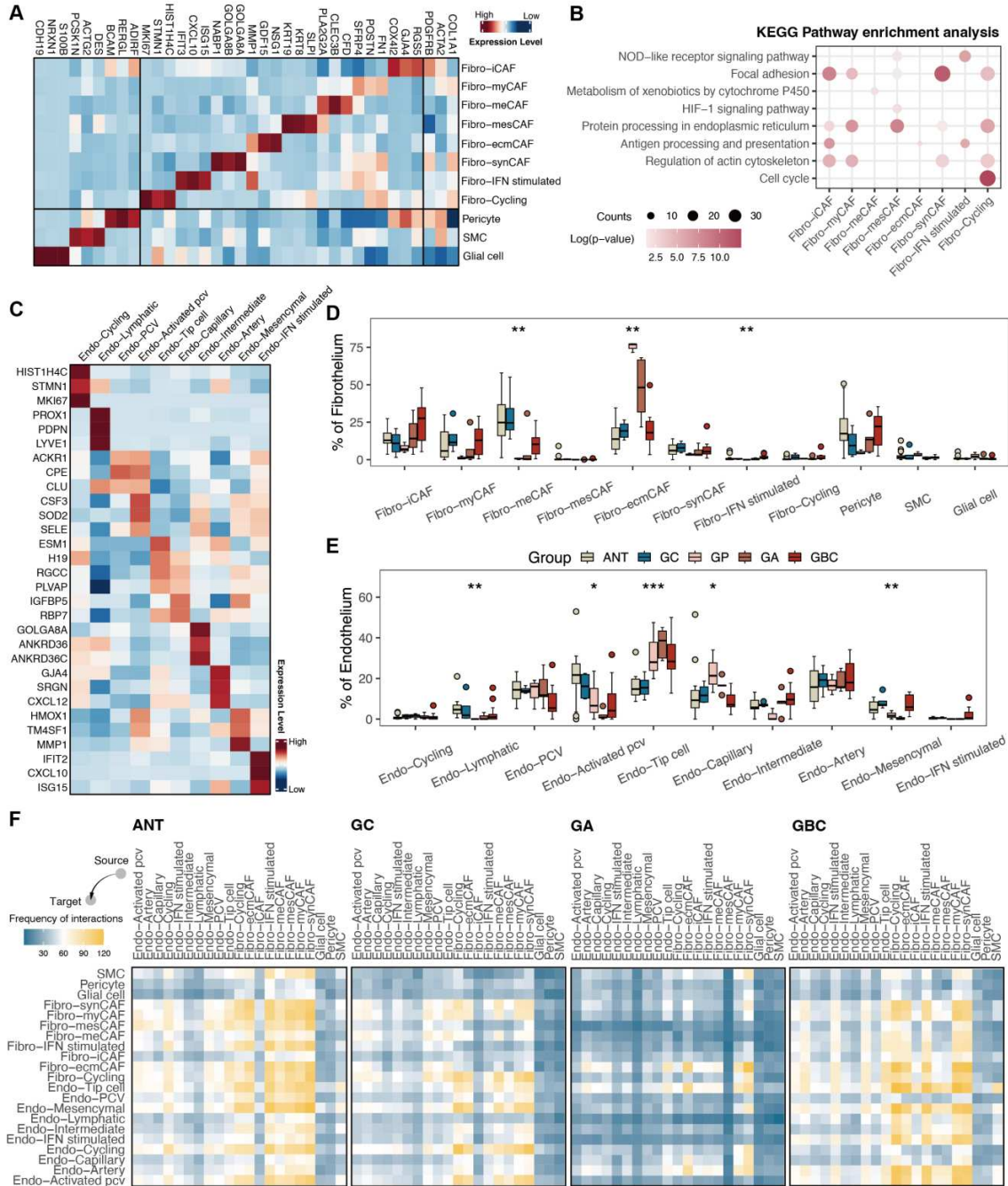
Figure S2



**Figure S2. Immunological profiles of immune cells.**

- A. Gene expression heatmap within each cell cluster of T/NK cells.
- B. Proportion of detected TCR and CDR3 length distribution in CD8<sup>+</sup>T cells.
- C. Boxplots showing the distribution of CD4-Tfh, CD4-Treg, and CD8-Tex were dominant in GBC.
- D. Multiplex IHC staining confirmed cycling Treg in SCC, indicated by yellow solid arrows.
- E. Gene expression heatmap in each cell cluster of Myeloid cells.
- F. Boxplots showing the distribution of macrophage subsets among groups.
- G. Scatter plots showing significant expressed genes in each subtype of GBC compared with other tissue types.
- H. Proportion of Myeloid-Cycling among different groups.
- I. Multiplex IHC staining confirmed the presence of cycling macrophage in GP, indicated by yellow solid arrows.

Figure S3-1

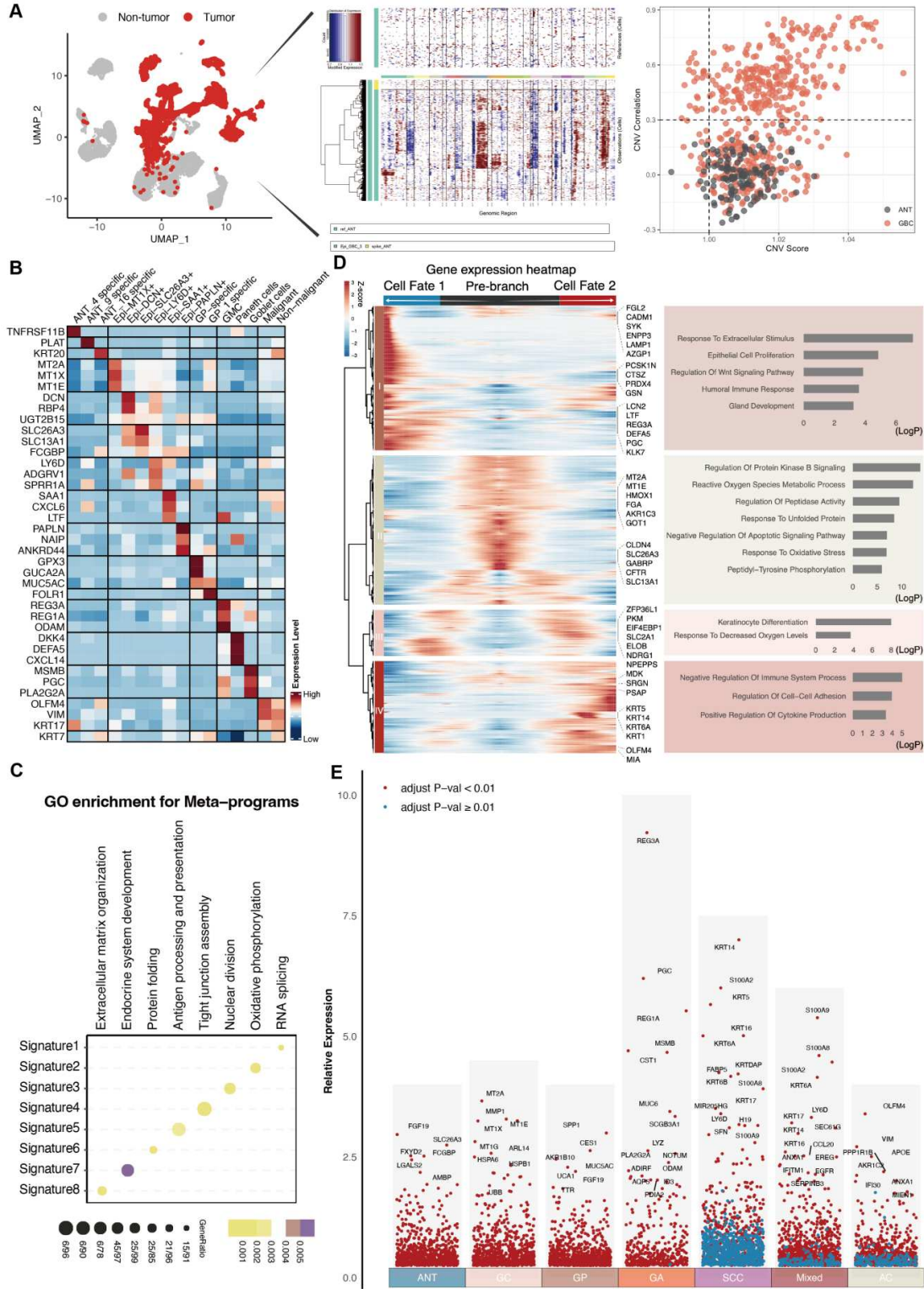




**Figure S3. Analysis of stromal cell subsets.**

- A. Heatmap depicting the expression of representative genes within the fibroblast subpopulations.
- B. KEGG enrichment analysis of fibroblast subsets.
- C. Heatmap illustrating the expression of representative genes of endothelium.
- D. Boxplots showing the distribution of fibroblast subsets across groups.
- E. Boxplots showing the distribution of endothelium subsets among groups.
- F. Heatmap illustrating patterns of cell-cell interactions in ANT, GC, GA, and GBC.
- G. Heatmap illustrating patterns of cell-cell interactions in GP.
- H. Heatmap displaying the potential ligands from Fibro-iCAF and their corresponding targeted gene in Endo-Tip cell.
- I. Heatmap depicting relative expression across groups of the top predicted ligands expressed by Fibro-iCAF using scRNA-seq (left). Heatmap highlighting significant ligand-receptor pairs between Fibro-iCAF and Endo-Tip cells in scRNA-seq (middle). Top predicted ligands color-coded by activity (right).
- J. Dot plot demonstrating the average expression of three candidate ligands associated with endothelium remodeling across different fibroblast clusters.

**Figure S4**



**Figure S4. Deciphering subtype-specific regulatory programs in epithelium cells of gallbladder diseases.**

- A. CNV inference analysis of malignant cells of GBC in an individual sample. Representative UMAP plot highlighting malignant cells (left). Representative heatmap of inferred CNV (middle). The CNV score and correlation for each cell (right).
- B. Heatmap showing the expression of representative genes of epithelium.
- C. Dotplot illustrating the GO enrichment results of each meta-program.
- D. Gene expression heatmap of DEGs (categorized in four clusters) in a pseudo-temporal order (left panel). GO analysis of upregulated genes in each cluster (right panel).
- E. Plot displaying the DEGs in each tissue subtype. Representative genes were indicated, and significant expressed genes were colored red.



Figure S5-1

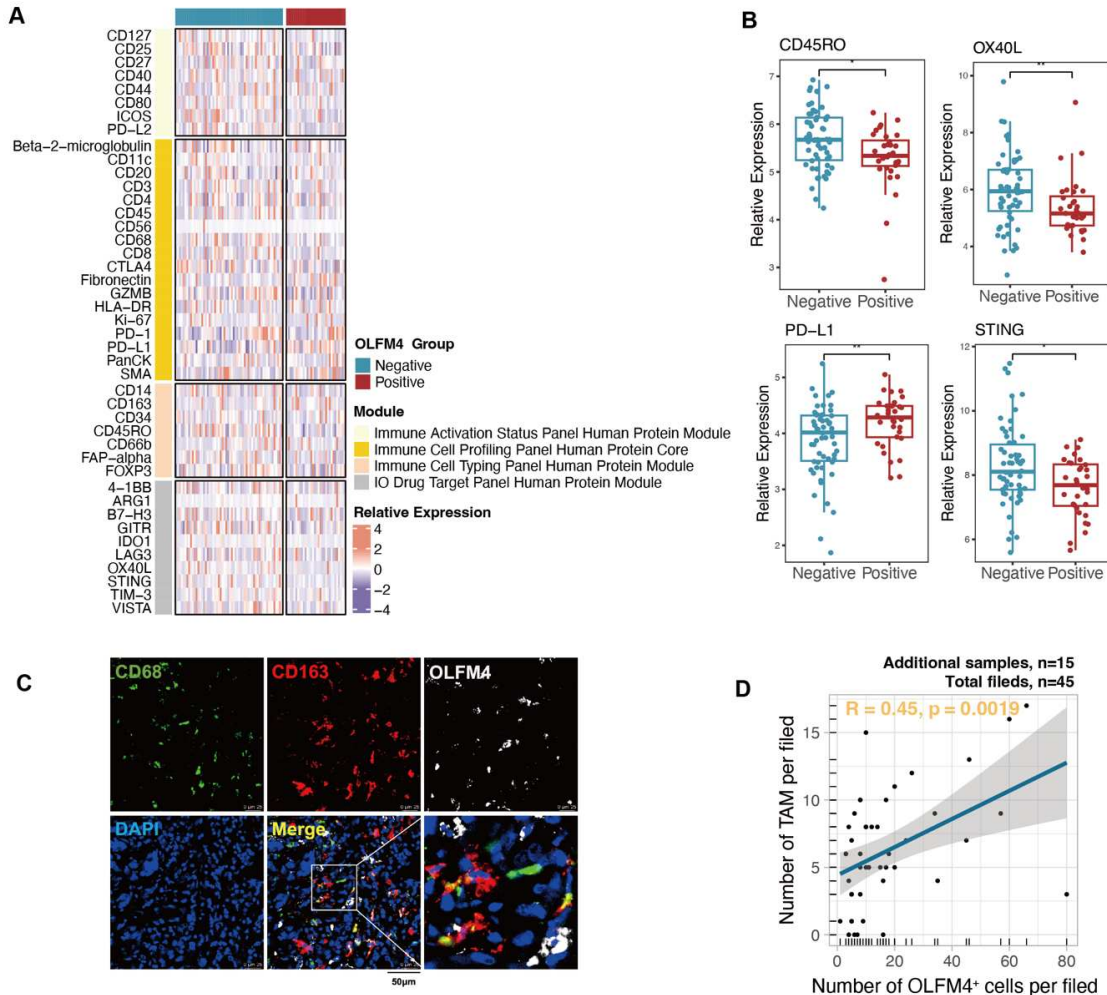
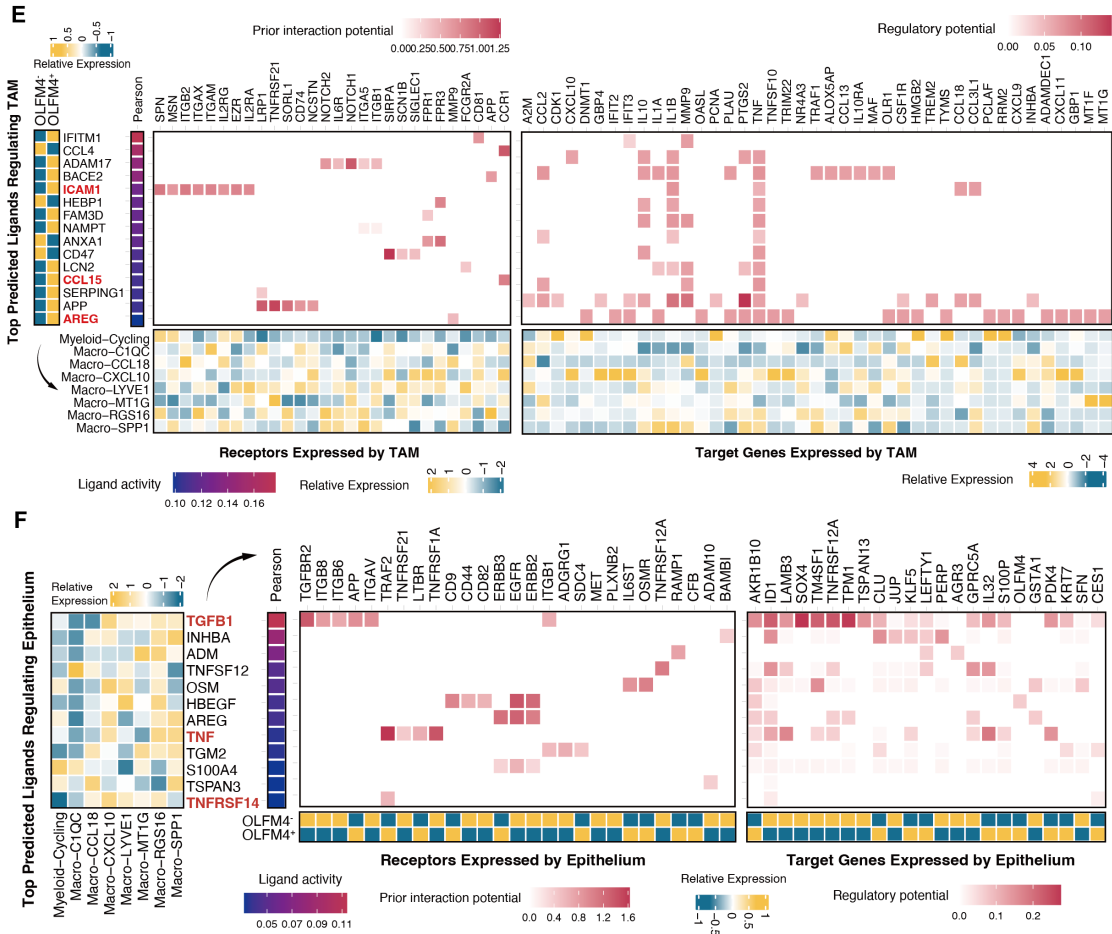


Figure S5-2



**Figure S5. Cell-cell interactome related to OLFM4 grouping.**

- A. Heatmap of DSP assay proteins stratified by OLFM4 expression level.
- B. Box plots comparing CD45RO, OX40L, PD-L1, and STING expression between groups.
- C. Immunofluorescence staining representing the presence of TAM (CD68<sup>+</sup>CD1163<sup>+</sup>) and OLFM4 expression in an additional clinical sample cohort (Sample n=15, 3 randomly selected fields per sample, total fields n=45).
- D. Pearson correlation between the number of TAM and OLFM4<sup>+</sup> cells in all fields.
- E. Putative signal sensed from malignant epithelium in GBC to TAM. Relative expression of top-ranked ligands (left panel). Top predicted ligand colored by activity (middle panel [left]). Heatmap of ligand-receptor pairs (middle panel [right]). Genes activated by top predicted ligands (right panel).
- F. Putative signaling pathways mediating communication between TAMs and malignant epithelium in GBC. Relative expression of top-ranked ligands (left panel). Top predicted ligand colored by activity (middle panel [left]). Heatmap of ligand-receptor pairs (middle panel [right]). Genes activated by top predicted ligands (right panel).

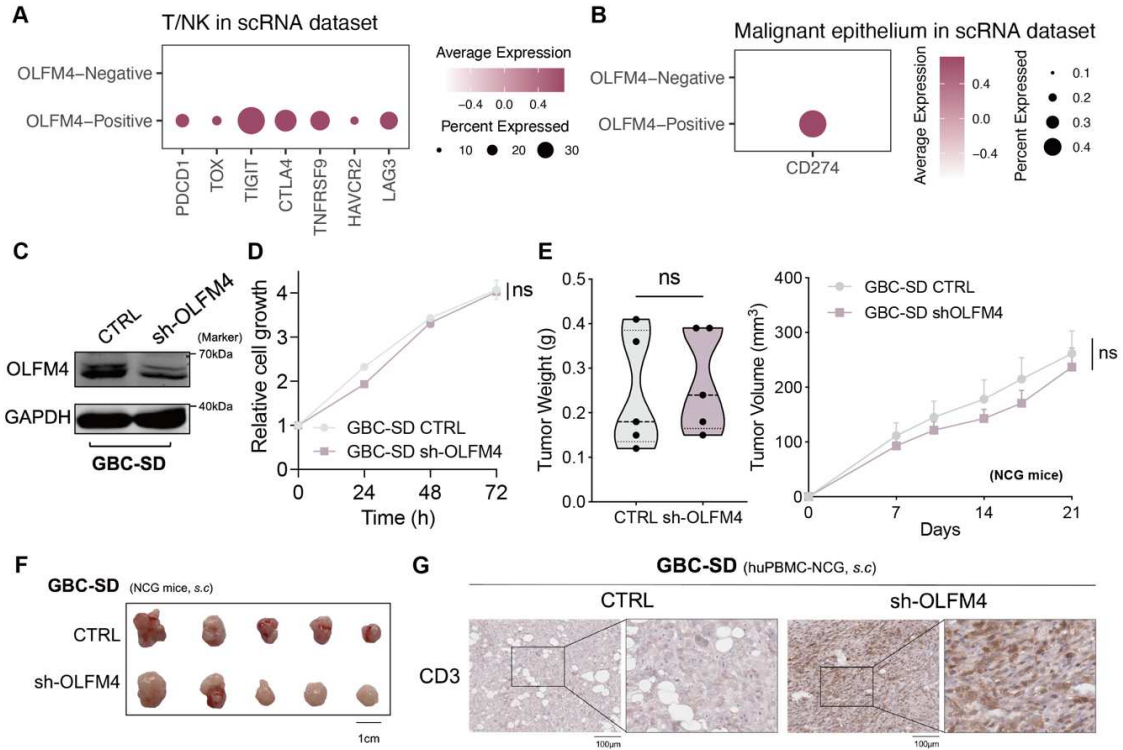
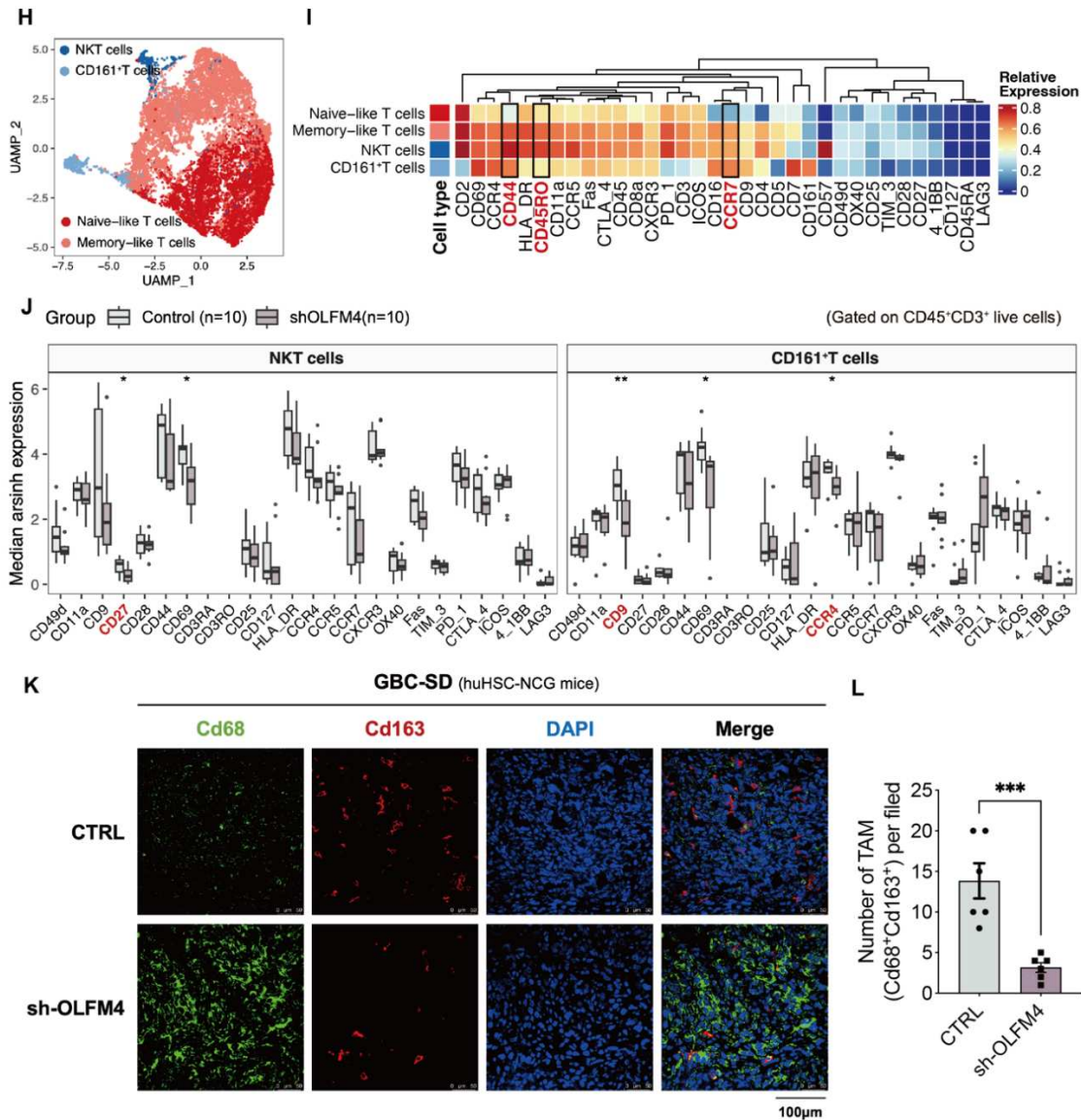
**Figure S6-1**

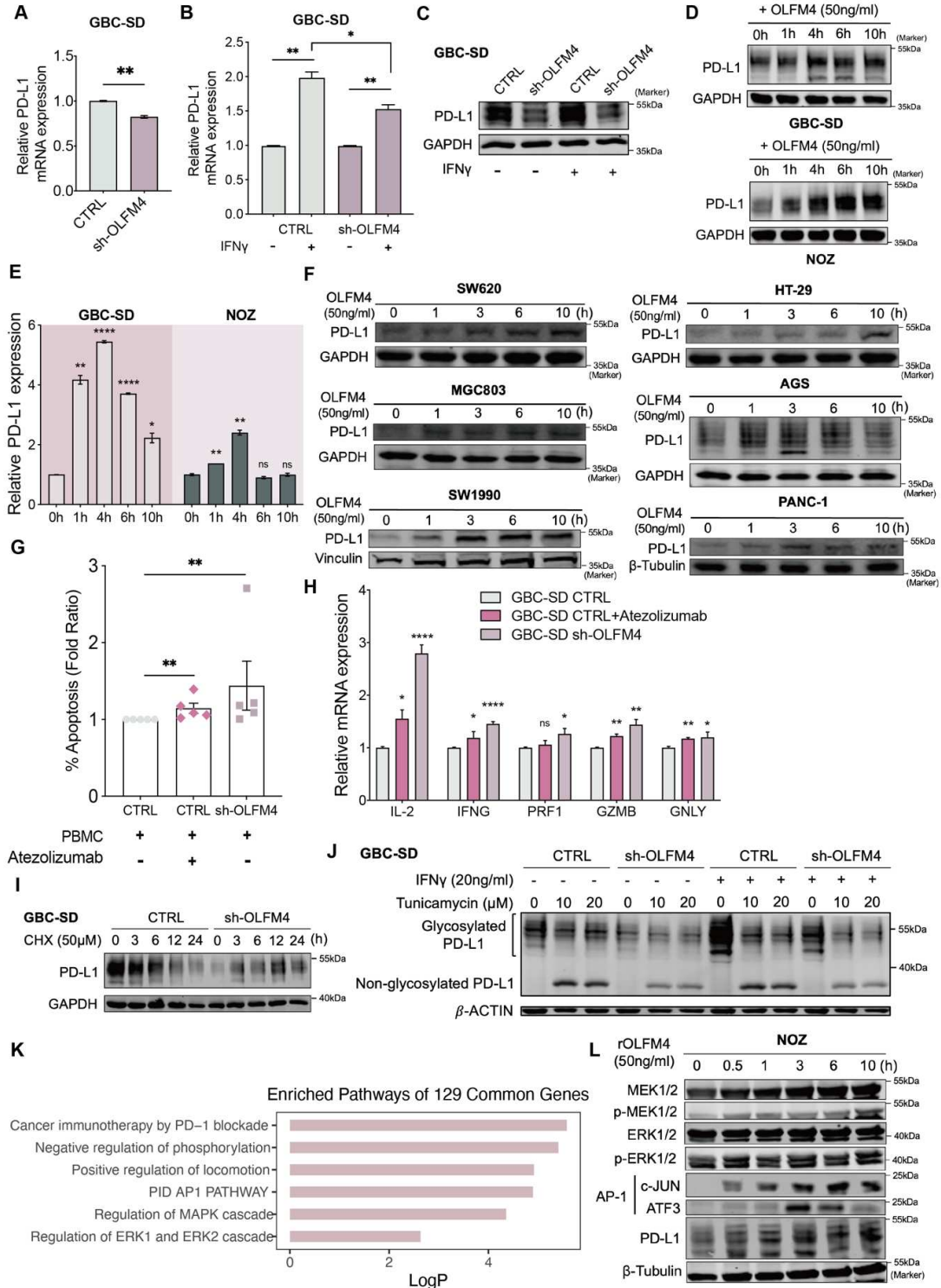
Figure S6-2



**Figure S6. OLFM4 regulated the tumor microenvironment *in vivo*.**

- A. Dotplot depicting the expression of T-cell exhaustion markers across groups stratified by OLFM4 expression.
- B. Dotplot depicting the expression of PD-L1 across groups stratified by OLFM4 expression.
- C. Confirmation of OLFM4 expression level by western blot in OLFM4 knockdown GBC cells.
- D. Assessment of the effect of OLFM4 knockdown on GBC cell proliferation using a CCK8 Assay.
- E. Subcutaneous injection of GBC-SD CTRL/sh-OLFM4 cells into NCG mice to obtain tumor xenografts. Tumor volume (right) and tumor weight (left).
- F. Gross morphology of tumors in the NCG model (CTRL, n=5; sh-OLFM4, n=5).
- G. Representative CD3 expression in CTRL/sh-OLFM4 groups detected by IHC.
- H. UMAP plots identifying 4 T-cell subsets in tumor-infiltrating lymphocytes of humanized mice.
- I. Heatmap depicting the protein expression from CyTOF analysis.
- J. Relative expression levels of a functional marker of NKT cells and CD161<sup>+</sup> T cells across recruited CyTOF cohort. \* $P < 0.05$  using a Wilcoxon test.
- K. Representative images of TAM in each group (\* $P < 0.05$ ).
- L. The proportion of TAM in CD34<sup>+</sup> humanized mice across groups.

**Figure S7**



**Figure S7. The MAPK-AP1 axis was involved in OLFM4-mediated regulation of PD-L1**

- A. Confirmation of PD-L1 mRNA Levels by quantitative RT-PCR in OLFM4-knockdown GBC cells.
- B. PD-L1 mRNA levels in GBC-SD CTRL/sh-OLFM4 under IFN $\gamma$  stimulation or unstimulated conditions.
- C. PD-L1 protein levels in GBC-SD CTRL/sh-OLFM4 under IFN $\gamma$  stimulation or unstimulated conditions.
- D. Elevated PD-L1 expression in GBC cell lines (GBC-SD [Top], NOZ [Bottom]) after treatment with OLFM4 (50 ng/mL).
- E. Time-dependent changes in PD-L1 mRNA levels in response to exogenous stimulation of OLFM4 (50 ng/mL).
- F. Elevated PD-L1 expression in multiple cancer cell lines after treatment with OLFM4 (50 ng/mL). Colorectal cancer, SW620, and HT-29; gastric cancer, MGC803, and AGS; pancreatic cancer, SW1990, and PANC-1.
- G. Proportion of apoptotic GBC-SD cells following a 72-hour co-culture with activated PBMC, with or without Atezolizumab (10  $\mu$ g/mL) treatment.
- H. Quantitative RT-PCR was performed to detect IL-2 (Interleukin-2), IFNG (IFN $\gamma$ ), PRF1 (perforin-1), GZMB (granzyme), and GNLY (granulysin) in activated PBMCs cocultured with GBC-SD CTRL/sh-OLFM4 cells, in the presence or absence of Atezolizumab (10  $\mu$ g/mL) treatment.



- I. PD-L1 protein levels in GBC-SD CTRL/sh-OLFM4 in the presence or absence of CHX (50  $\mu$ M).
- J. PD-L1 protein levels in GBC-SD CTRL/sh-OLFM4 under IFN $\gamma$  stimulation or unstimulated conditions, in the presence or absence of tunicamycin (10 or 20  $\mu$ M, 24 h).
- K. Enriched pathways of 129 common genes, related to Figure 7L.
- L. Western blot analyses of the levels of total MEK1/2, p-MEK1/2, total ERK1/2, p-ERK1/2, AP-1, and PD-L1 in NOZ treated with OLFM4 (50 ng/mL).

## Supplementary materials and methods

### Cell lines

GBC-SD, NOZ, SW620, HT-29, MGC803, AGS, SW1990, and PANC-1 were purchased from ATCC, authenticated through the STR characterization method, and regularly tested for Mycoplasma. Specifically, GBC-SD, MGC803, SW620, and AGS cell lines were cultured in RPMI 1640 medium, while NOZ, SW1990, and PANC-1 were cultured in DMEM. HT-29 cells were cultivated in McCoy's 5A medium. In all cases, the culture media were supplemented with 10% fetal bovine serum and 1% penicillin and streptomycin.

### Sample processing

Upon arrival at the laboratory, tissue samples were subjected to mechanical and enzymatic dissociation using the tumor dissociation kit (Miltenyi) and the GentleMACS Octo Dissociator with Heaters (Miltenyi). Resection samples were finely chopped and introduced into a GentleMACS tube containing 7.5 mL of enzyme mix. In comparison, core needle biopsies and fine needle aspiration samples were combined with 2.5 mL of enzyme mix in the same type of tube. After an incubation period of 15 to 30 minutes, which varied based on sample size and consistency, larger specimens were filtered through MACS SmartStrainers (70  $\mu$ m) (Miltenyi) into 50 mL tubes. Subsequently, dead cells and cellular debris were effectively removed using the Debris Removal Solution from Milenyi Biotec. The samples were then subjected to centrifugation at 800g for 1 minute, and the resulting supernatant was carefully discarded. Following this, the cells were subjected to two wash cycles and resuspended in PBS containing 0.5% bovine serum albumin, in preparation for library construction and sequencing.

### **OLFM4 knockdown**

Lentivirus for OLFM4 knockdown was produced and procured from Obio Technology in Shanghai, China. Cells, which were at a confluence level of 60-70%, were incubated in a growth medium containing appropriately diluted lentivirus along with polybrene. After 48 hours of transfection, the cells underwent puromycin selection at a concentration of 5 mg/mL to isolate and establish stable transfected cell lines.

### **Real-time quantitative PCR**

Total cellular RNA was isolated using Trizol reagent from Invitrogen. This RNA was then reverse transcribed into cDNA utilizing Superscript III reverse transcriptase (Invitrogen) and random primers, following the manufacturer's instructions. The resulting cDNA served as a template for amplifying target gene transcripts through real-time PCR, employing SYBR Green PCR Master Mix from Applied Biosystems, and an ABI PRISM 7300HT Sequence Detection System (also from Applied Biosystems). GAPDH was employed as a control for normalization. For a comprehensive list of primers, please refer to Table S7.

### **Western blot**

The Western blotting analysis was carried out following established procedures. In brief, cells were lysed using IP lysis buffer from Beyotime Biotechnology in Shanghai, China, with the addition of 1 mM PMSF, and kept on ice for 30 minutes. Protein concentrations were quantified using the Pierce™ BCA Protein Assay Kit from ThermoFisher Scientific in MA, USA. Equal quantities of protein were loaded onto SDS-PAGE gels and subsequently transferred onto 0.22 µm nitrocellulose membranes from Millipore in Cork, Ireland. The membranes were then

incubated with the respective primary antibodies overnight at 4°C, followed by incubation with IRDye 800 goat anti-rabbit antibody (LI-COR Biosciences, Lincoln, USA) for 1 hour at room temperature. Following the removal of unbound antibodies through washing, the labeled bands were scanned using the Odyssey® CLx Infrared Imaging System from LI-COR Biosciences in MA, USA.

### **Multiplex immunofluorescence tissue staining**

For fluorescent multiplex immunohistochemistry analysis, a four-color fluorescence kit based on tyramine signal amplification (TSA) was employed according to the manufacturer's protocol. In a nutshell, slides underwent deparaffinization and rehydration. Antigen retrieval was performed, followed by treatment with 3% H<sub>2</sub>O<sub>2</sub> for 20 minutes. After washing, the slides were blocked using 1% BSA. Primary antibodies were applied, followed by the TSA solution. Following the final TSA cycle, DAPI was used for counterstaining at a dilution of 1:1000 for 10 minutes. Photomicrographs of the stained sections were captured using the Leica TCS SP8 system from Leica Biosystems in MA, USA.

### **Assessment of T Cell Cytotoxicity**

T cell cytotoxicity was assessed based on the expression levels of IFN $\gamma$ , CD107a, IL-2, perforin, granulysin, and Granzyme B. Peripheral blood mononuclear cells (PBMCs) from healthy donors were pre-activated using anti-CD3 (5 $\mu$ g/mL, Biolegend) and anti-CD28 (5 $\mu$ g/mL, Biolegend) for 2-3 days. Meanwhile, GBC-SD CTRL/sh-OLFM4 cells were seeded into a 12-well plate and allowed to culture overnight.

The pre-activated PBMCs were introduced into the same well for co-culture with the tumor cells at a 4:1 ratio, and this co-culture was maintained for 72 hours. Following incubation, the suspended cells (primarily PBMCs) were collected, washed twice, and subsequently subjected to RNA extraction or analyzed using flow cytometry. The residual cells in the cell plate were washed twice with PBS and subjected to the TUNEL assay as per the manufacturer's provided protocol. Atezolizumab was added at a concentration of 10µg/mL to inhibit PD-L1 function.

### ***In vitro* tumorigenic surrogate analyses**

In the context of growth curves, numerous 96-well plates were seeded with 3,000 cells per well and cell density was assessed using a luminescent assay. Cell proliferation was determined by normalization against the cell density measurement on day 0. To evaluate chemoresistance to gemcitabine, GBC-SD cells were exposed to specified concentrations of gemcitabine for 72 hours. Regarding the migration assay, GBC-SD CTRL/sh-OLFM4 cells were positioned in the upper chamber. In contrast, for the invasion assay, Matrigel-coated membranes were employed to replicate the extracellular matrix environment. Following a 24-hour incubation period, non-invading or non-migrating cells were removed, and the remaining cells on the lower side of the membrane were stained and quantified.

### **Mouse xenograft models**

All animal experiments adhered to NIH guidelines and were approved by the Ethics Committees of Eastern Hepatobiliary Surgery Hospital (EHBH) (No. DWLL-004). Adult female NCG mice (NOD-*Prkdc*<sup>em26Cd52</sup>*Il2rg*<sup>em26Cd22</sup>/NjuCrl; 6–8 weeks old) were procured from the

Nanjing Biomedical Research Institute of Nanjing University. They were randomly allocated into experimental groups. GBC-SD CTRL/sh-OLFM4 cells at a concentration of  $5 \times 10^6$  were injected into the right flank of NCG mice. Tumor size (calculated as length  $\times$  width<sup>2</sup>  $\times$  0.5) was assessed twice per week following the injection. PBMCs from healthy donors were activated and expanded as described previously<sup>1,2</sup>. On the day before tumor cell injection, PBMCs ( $1 \times 10^7$  cells) were adoptively transferred to NCG mice via the tail vein.

CD34<sup>+</sup> humanized NCG mice were also obtained from the Nanjing Biomedical Research Institute of Nanjing University and were generated as outlined in previous reports. After 21 days of cell injection, CD34<sup>+</sup> humanized NCG mice were humanly euthanized, and the tumor-infiltrating leukocytes were isolated for subsequent CyTOF analysis.

### **Mass CyTOF and data processing**

A set of pre-conjugated antibodies comprising 34 markers was procured from the supplier (cat no. 201321, 201307, and 201305 [Fludigm, USA]; Table S6). Tumor-infiltrating lymphocytes were isolated from freshly resected tumors of the huHSC-NCG model. These cells were stained for viability, using 5  $\mu$ M cisplatin, for 2 minutes, and then exposed to surface markers for 30 minutes at room temperature. Subsequently, the cells were fixed and subjected to analysis using a Helios mass cytometer from Fludigm, USA. The resulting files in .fcs format were uploaded to Cytobank (<https://community.cytobank.org>), where total T cells were manually gated, and events of interest were exported as .fcs files. The high-dimensional raw data underwent dimension reduction as part of the initial processing. A random sampling was conducted from each .fcs file using the cytofWorkflow package within the R software environment.

## GeoMx DSP

Formalin-fixed paraffin-embedded (FFPE) slides (4  $\mu\text{m}$ ) were baked at 60°C for 1.5 hours, and then deparaffinized and rehydrated as follows: 3 $\times$ 5 min in CitriSolv, 2 $\times$ 5 min in 100% ethanol, 2 $\times$ 5 min in 95% ethanol, and 2 $\times$ 5 min in double-distilled water. For antigen retrieval, slides were placed in a staining jar containing 1 $\times$  citrate buffer with pH 6 at 25°C. The staining jar containing the slides was placed in a preheated pressure cooker and run at high pressure and temperature for 15 min. After carefully releasing the pressure, transferring the staining jar to the lab bench, removing the lid, and letting it stand for 25 min, the slides were then washed with 1 $\times$  tris-buffered saline with Tween-20 (TBST) for 5 min. Blocking was performed by placing the slide in a humidity chamber in a horizontal position and covering it with sufficient Buffer W (NanoString). The slides were then incubated with Buffer W for 1 hour at 25°C in a humidity chamber. Ultraviolet (UV)-photocleavable oligo antibody sets (Immune Cell Profiling Core, Immuno-oncology (IO) Drug Target Module, Immune Cell Typing Module, and Immune Activation Status Module), containing 44 targets, were used for protein detection. A mixture of UV-photocleavable oligo antibody sets and morphological markers panCK, CD45, and OLFM4 was diluted in Buffer W. The slides were removed from the humidity chamber and Buffer W was discarded then placed back into the humidity chamber and covered with diluted antibody solution. The humidity chamber was then transferred to a 4°C freezer and incubated overnight. Postfix was performed by removing the slide from the humidity chamber and carefully aspirating the antibody solution from the slide. The slides were washed for 3 $\times$ 10 min in TBST. The samples were covered with 4% paraformaldehyde and incubated for 30 min at 25°C in a humidity chamber. After incubation, the slides were washed for 2 $\times$ 5 min in TBST. For nuclear staining,

the slides were incubated with SYTO 13 for 15 min at 25°C in a humidity chamber and rinsed with 1× TBST. Finally, the slides were loaded onto the GeoMx instrument.

Whole-slide image analysis employed HALO® image analysis software (version v3.3.2541.323, Indica Labs, Inc.). Quantification of PANCK<sup>+</sup>OLFM4<sup>+</sup> Epithelium was conducted utilizing the High-Plex FL module. The OLFM4-Positive group was identified when PANCK<sup>+</sup>OLFM4<sup>+</sup> cells constituted more than 0.2 of all cells in the region of interest (ROI) field of view, and conversely, deemed negative otherwise.

### **ScRNA-seq data pre-processing**

The 5'-expression sequencing data, obtained off the machine, underwent demultiplexing and alignment to the human transcriptome (GRCh38) using Cell Ranger v2.1.1 (10x Genomics). The outputs for the 16 samples were aggregated to create a combined raw expression matrix, accomplished through the 'cell ranger aggression' function.

The unique molecular identifier (UMI) count matrix was transformed into Seurat objects via the R package Seurat (version 4.1.1)<sup>3</sup>. Cells that met specific criteria, including detected gene numbers between 200 and 6,000, UMI numbers between 1,000 and 50,000, and a percentage of mitochondrial genes below 10%, were considered qualified and retained. Following quality control, a dataset consisting of 230,737 cells and 29,418 genes was prepared for downstream analysis. Raw gene expression values for each cell were normalized by dividing the total expression and subsequently scaled (multiplied by 10,000) and log-transformed using the 'NormalizeData' function within the Seurat toolkit (UMI-per-10,000+1).



To mitigate batch effects, we employed the harmony algorithm<sup>4</sup> to integrate samples based on patient samples. Essentially, we divided the combined Seurat object into a list of Seurat objects, with each dataset as an element, by executing the 'SplitObject' command. Each Seurat dataset within the list was normalized, and variable genes were identified using 'NormalizeData' and 'FindVariableFeatures' (SeuratObject, selection.method = "vst," features = 2,000). Subsequently, 'RunHarmony' was conducted, returning a Seurat object with an integrated expression matrix that had corrected batch effects. This object included a "harmony" assay with the integrated expression matrix, while the original uncorrected values were stored in the "RNA" assay, allowing flexibility in switching between them. The integrated expression matrix was employed for downstream analysis. Initially, we scaled the integrated data for principal component analysis (PCA) and UMAP visualization. Cells were subsequently clustered by cell type, rather than by batch effects.

### **The ratio of Observed to Expected Cell Numbers in Pathogenesis Analysis**

We calculated the ratio of observed to expected (Ro/e) cell numbers within each cluster to dissect significant variances in cell distribution among various pathogenic states based on methods previously reported in the literature<sup>5</sup>. This quantification is pivotal for revealing deviations from expected distributions, assuming no specific association between cell types and pathogenic conditions. The corrected formula, in alignment with your code's functionality, is articulated as  $Ro/e_{ij} = O_{ij}/E_{ij}$ . Here  $O_{ij}$  represents the observed number of cells of type  $i$  within pathogenesis  $j$ , while  $E_{ij}$  denotes the expected number of cells, determined by:  $E_{ij} = T_i \times P_j / T$ .  $T_i$  is the total number of cells of type  $i$  across all pathogenic states,  $P_j$  is the total count of cells in pathogenesis  $j$ , and  $T$  signifies the total of cells observed. This computation effectively highlights

areas of enrichment or depletion in cell types within specific pathogenic contexts, essential for understanding cellular dynamics and contributions to disease pathology.

### **Quantitative analysis of clonal expansion and transition**

The analysis utilized the scRepertoire<sup>6</sup> package in conjunction with Seurat to integrate TCR sequencing with scRNA-seq data, enabling the assessment of clonal expansion in T cell subpopulations across various samples. By aggregating TCR sequences with metadata annotations like sample identity and pathogenesis, the study categorized clones based on frequency, from Single (appearing once) to Hyperexpanded (more than 250 appearances). This categorization quantified clonal expansion using the frequency distribution of TCR sequences. Moreover, the investigation focused on clonal transitions across different cell types and pathogenic states, leveraging scRepertoire to track TCR sequence presence and frequency. This approach facilitated a detailed analysis of clonal dynamics, comparing shared sequences among cell types and conditions to quantify clonal overlap and assess the impact of pathogenic stimuli on clonal populations.

### **ROGUE analysis to assess cellular heterogeneity in scRNA-seq data**

To analyze cellular heterogeneity in scRNA-seq data using ROGUE<sup>7</sup>, begin by ensuring that the celltype\_ROGUE metadata accurately reflects cell types, particularly refined epithelial categories according to pathology groupings. For ROGUE analysis, convert Seurat's sparse expression matrix to a dense format using Rcpp for compatibility. Execute the rogue function on this dense matrix, providing cell type labels, sample identifiers, and specifying parameters like platform for UMI counts and span for data smoothing. For visualization, if starting from pre-

computed ROGUE scores, load them and use `rogue.boxplot` to create boxplots representing the variability of gene expression robustness across cell types. Incorporate statistical comparisons such as Kruskal-Wallis tests directly into the boxplots with `stat_compare_means`, highlighting significant expression robustness differences among cell types.

### **Analysis of malignant cell heterogeneity using cNMF**

The process began with the preprocessing of scRNA-seq datasets using the Seurat package to isolate malignant epithelial cells based on specific metadata annotations. This subset was further refined by excluding cells from designated samples to focus on the most relevant cellular populations for analysis. Subsequent steps included a rigorous filtering process to remove genes associated with mitochondrial processes, ribosomal proteins, immunoglobulin genes, and other non-epithelial markers to prepare the data for computational non-negative matrix factorization (cNMF) analysis<sup>8</sup>. Each sample's expression data was exported into separate text files, tailored for cNMF compatibility.

The core of our analysis involved executing a series of Python scripts to perform cNMF, a computational method designed to identify gene expression programs that underpin cellular heterogeneity and infer cellular states. This included data preparation, factorization, and results combination across varying component numbers to determine the optimal representation of cellular states. The integration of cNMF analysis aimed to reveal the underlying gene expression programs contributing to the observed cellular heterogeneity. Post-cNMF analysis, the results were analyzed within the R environment, focusing on quality control and the correlation between identified gene expression programs. We generated correlation heatmaps to evaluate the distinctiveness and consistency of these programs across the cellular landscape. Enrichment

analysis on genes associated with each program was conducted using tools like clusterProfiler against various databases, including gene ontology and KEGG pathways, to interpret the biological significance of the expression patterns.

### **Integrative analysis of cellular trajectories in epithelium**

Our study deployed a comprehensive analytical framework combining Monocle2, CytoTRACE, RNA velocity analysis, and Partition-based graph abstraction (PAGA) to dissect the cellular trajectories and underlying gene expression programs within epithelial tumor cells.

We began by collating data from various sources, including Monocle2 for cellular trajectories, CytoTRACE for estimating cellular states, and scRNA-seq data focusing on epithelial cells. The integration process entailed aligning datasets based on cell barcodes, ensuring a coherent foundation for subsequent analyses. Utilizing Monocle2, we visualized cellular trajectories, color-coded by pathogenesis, to delineate cellular progression pathways. Pie charts representing different cellular states further detailed the distribution of pathological conditions. CytoTRACE scores were incorporated to refine our understanding of cellular states, enhancing trajectory analysis by integrating a measure of cellular 'stemness' or differentiation potential. RNA velocity analysis was conducted to estimate the direction and speed of cellular transitions, adding a temporal dimension to our trajectory insights. This approach allowed us to predict future cellular states based on the current transcriptional dynamics. PAGA was employed to construct a graph abstraction of the data, providing a simplified yet informative representation of the complex cellular transitions and interactions within the dataset. This facilitated the identification of key branching points and transition pathways between cellular states.

Focused gene expression analysis, including BEAM to pinpoint genes associated with trajectory branch points and heatmap visualization of differentially expressed genes, highlighted distinct expression programs. These analyses were complemented by enrichment studies to elucidate biological functions and pathways characterizing each trajectory segment. Enrichment analyses leveraged GO and KEGG databases, alongside custom gene lists from unique and time-differentially expressed gene compilations, to annotate the functional implications of identified gene expression patterns.

### **Cell-cell interactome**

We utilized a method called CellPhoneDB<sup>9</sup>, tailored for single-cell transcriptome data, to investigate cell-cell communication. This method relies on a manually curated repository of interacting ligands and receptors. In essence, it infers potential cell-cell interactions by evaluating the expression of interacting ligand-receptor pairs between two clusters. For a gene encoding a receptor or ligand to be considered in downstream analysis, it should be expressed in more than 30% of cells within a specific cluster. To assess the significance of a ligand-receptor pair between two clusters, a permutation test was performed by randomly assigning cluster labels to each cell 1,000 times. An empirical P-value was determined by ranking the actual average expression of a given ligand and receptor pair in two clusters among the 1,000 permutations.

For NicheNet analysis<sup>10</sup>, we generated cell type signatures by selecting the top differentially expressed genes (with an average  $\text{Log}_2\text{FC} > 1$ ) in cells isolated from tumors, including epithelium and TAMs. These signatures were then input into NicheNet to derive a comprehensive set of predicted ligands that modulate TME cell-type signatures. For example, to predict ligands modulating Endo-Tip cells, we employed the top differentially expressed genes in

Endo-Tip cells. In each case, we presented the top 20% of predicted ligands, based on regulatory potential, that also demonstrated significance in our single-cell RNA-seq ligand-receptor interaction analysis, as described earlier. These findings are depicted in Figure S3 and Figure S7.

### **Quantitative correlation analysis of cellular composition**

In our analysis, we meticulously aggregated metadata that included sample identifiers and non-epithelial cell types to quantitatively evaluate the cellular composition within the microenvironment. A comprehensive table was constructed to count each cell type's occurrences across samples, facilitating the calculation of their percentage representations. This enriched dataset, augmented with additional metadata such as group and pathogenesis, served as the foundation for our correlation analysis. We employed Pearson's correlation tests to examine the relationships between the percentage representations of all cell types across samples. To ensure the reliability of our findings, we adjusted the p-values from Spearman's correlation tests using the Benjamini-Hochberg method, categorizing them into four significance levels. The resulting correlations were visualized on a heatmap, with Pearson's correlation coefficients depicted through a color gradient and the significance categories through point sizes.

### **Hierarchical clustering of sample similarities based on cellular composition**

To analyze the similarities in cellular composition across samples within the microenvironment, we first organized our data to include sample identifiers, cell types, and their respective percentages within each sample. This data was transformed into a matrix where columns represented samples and rows corresponded to cell types, with values indicating the percentage of each cell type per sample. Utilizing the vegan package, we computed Bray-Curtis

dissimilarities within pathogenic groups (ANT, GC, GP, GA, GBC) to capture the ecological distances that underscore compositional differences between samples. Hierarchical clustering was then applied to these dissimilarity matrices using the 'average' linkage method, allowing us to identify clusters of samples with similar cellular compositions. The clustering results informed the ordering of samples, integrating these insights across all pathogenic conditions.

### **Analysis of TCGA and bulk RNA-seq cohorts for stromal systems**

In the analysis of the TCGA cohort, we obtained preprocessed gene expression data (TOIL RSEM tpm) and clinical data for the TCGA Pan-Cancer (PANCAN) RNA-seq gene expression dataset from UCSC Xena (<http://xena.ucsc.edu>). Differential expression analysis was employed to determine specific markers for stromal cells, such as fibroblasts and endothelial cells, leading to the creation of signatures like SC1 and SC3 (Table S3). Subsequently, we conducted SC1/SC3-specific gene signature scoring, employing the GSVA package. Survival analysis, including Kaplan-Meier and Cox regression models, was conducted to evaluate the prognostic value of these stromal signatures in predicting patient outcomes. The pan-cancer approach allowed us to examine the generalizability and potential universal relevance of these stromal markers across different cancer contexts.

GSVA was also conducted on bulk RNA-seq data of GBC from Pandey., et al<sup>11</sup> to assess the activity levels of the identified stromal signatures. This analysis provided insights into the functional states of stromal cells within the tumor context. Enrichment of KEGG pathway analyses was performed on signature genes to uncover the biological processes and pathways enriched in each stromal cell category. Selected significant pathways were further visualized, emphasizing their relevance to stromal cell functions.

**Data and code availability**

The raw FASTQ files from this study can be made available for scientific research purposes upon request while ensuring compliance with relevant privacy laws due to human patient privacy concerns. Additionally, the code used for all data processing and analysis is also accessible upon request.



## Reference

1. Qiu, X., Yang, S., Wang, S., Wu, J., Zheng, B., Wang, K., Shen, S., Jeong, S., Li, Z., Zhu, Y., et al. (2021). M6A Demethylase ALKBH5 Regulates PD-L1 Expression and Tumor Immunoenvironment in Intrahepatic Cholangiocarcinoma. *Cancer Res.* *81*, 4778–4793. 10.1158/0008-5472.CAN-21-0468.
2. Huang, C., Chen, B., Wang, X., Xu, J., Sun, L., Wang, D., Zhao, Y., Zhou, C., Gao, Q., Wang, Q., et al. (2023). Gastric cancer mesenchymal stem cells via the CXCR2/HK2/PD-L1 pathway mediate immunosuppression. *Gastric Cancer* *26*, 691–707. 10.1007/s10120-023-01405-1.
3. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell* *184*, 3573–3587.e29. 10.1016/j.cell.2021.04.048.
4. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* *16*, 1289–1296. 10.1038/s41592-019-0619-0.
5. Zhang, Q., He, Y., Luo, N., Patel, S.J., Han, Y., Gao, R., Modak, M., Carotta, S., Haslinger, C., Kind, D., et al. (2019). Landscape and Dynamics of Single Immune Cells in Hepatocellular Carcinoma. *Cell* *179*, 829–845.e20. 10.1016/j.cell.2019.10.003.
6. Borcherding, N., Bormann, N.L., and Kraus, G. (2020). scRepertoire: An R-based toolkit for single-cell immune receptor analysis. *F1000Research* *9*, 47. 10.12688/f1000research.22139.2.
7. Liu, B., Li, C., Li, Z., Wang, D., Ren, X., and Zhang, Z. (2020). An entropy-based metric for assessing the purity of single cell populations. *Nat. Commun.* *11*, 3155. 10.1038/s41467-020-16904-3.
8. Kotliar, D., Veres, A., Nagy, M.A., Tabrizi, S., Hodis, E., Melton, D.A., and Sabeti, P.C. (2019). Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. *eLife* *8*, e43803. 10.7554/eLife.43803.
9. Efremova, M., Vento-Tormo, M., Teichmann, S.A., and Vento-Tormo, R. (2020). CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat. Protoc.* *15*, 1484–1506. 10.1038/s41596-020-0292-x.
10. Browaeys, R., Saelens, W., and Saeys, Y. (2020). NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* *17*, 159–162. 10.1038/s41592-019-0667-5.
11. Pandey, A., Stawiski, E.W., Durinck, S., Gowda, H., Goldstein, L.D., Barbhuiya, M.A., Schröder, M.S., Sreenivasamurthy, S.K., Kim, S.-W., Phalke, S., et al. (2020). Integrated

genomic analysis reveals mutated ELF3 as a potential gallbladder cancer vaccine candidate.  
Nat. Commun. 11, 4225. 10.1038/s41467-020-17880-4.