Supplemental information

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

He H, *et al. Gut* 2024;0:1–14. doi: 10.1136/gutjnl-2023-331773

Figure S1

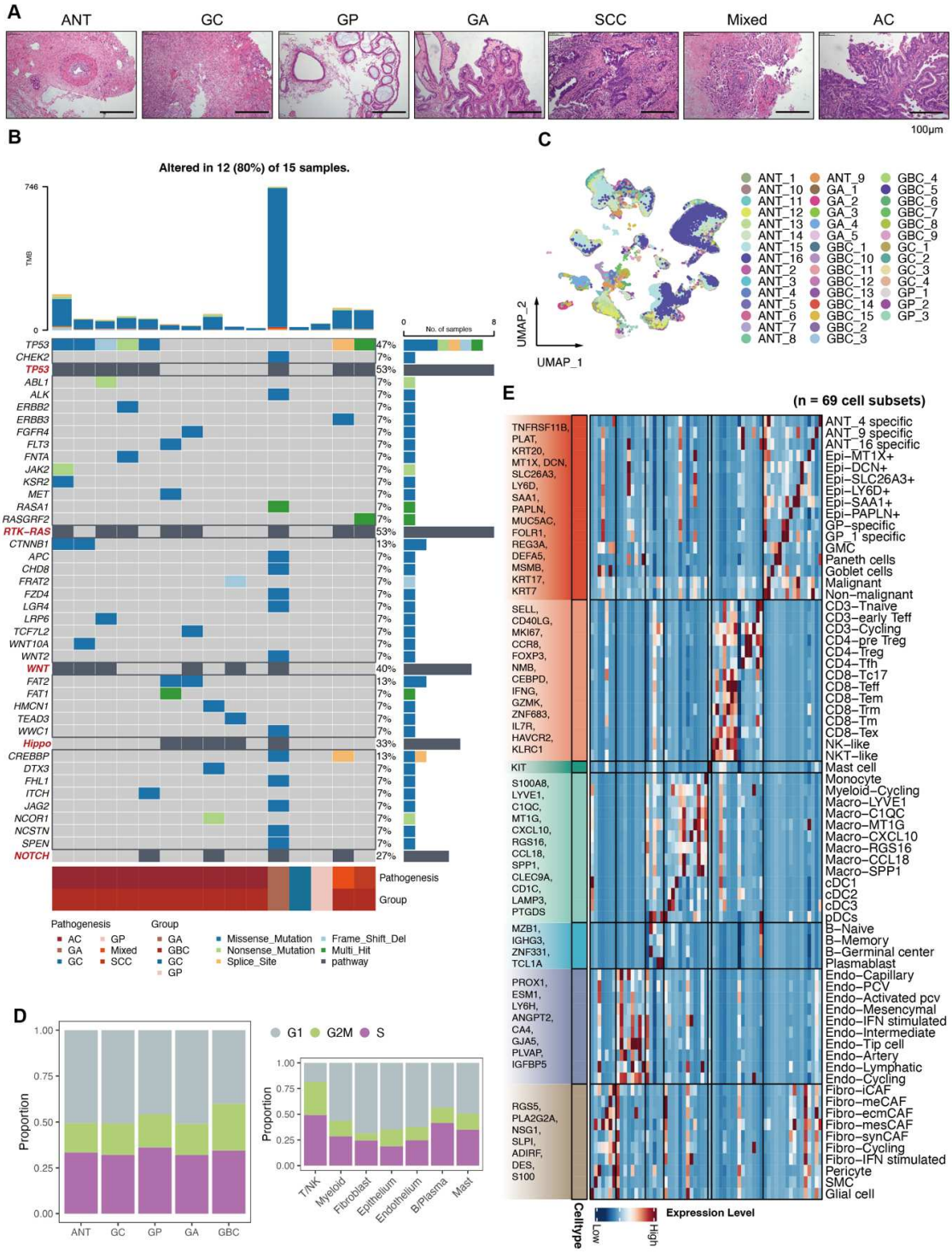

A
ANT
GC
GP
GA
SCC
Mixed
AC
100μm
B
Altered in 12 (80%) of 15 samples.
TMB
No. of samples
TP53 47%
CHEK2 7%
TP53 53%
ABL1 7%
ALK 7%
ERBB2 7%
ERBB3 7%
FGFR4 7%
FLT3 7%
FNTA 7%
JAK2 7%
KSR2 7%
MET 7%
RASA1 7%
RASGRF2 7%
RTK–RAS 53%
CTNNB1 13%
APC 7%
CHD8 7%
FRAT2 7%
FZD4 7%
LGR4 7%
LRP6 7%
TCF7L2 7%
WNT10A 7%
WNT2 7%
WNT 40%
FAT2 13%
FAT1 7%
HMCN1 7%
TEAD3 7%
WWC1 7%
Hippo 33%
CREBBP 13%
DTX3 7%
FHL1 7%
ITCH 7%
JAG2 7%
NCOR1 7%
NCSTN 7%
SPEN 7%
NOTCH 27%
Pathogenesis
Group
Pathogenesis
AC
GA
GC
GP
Mixed
SCC
Group
GA
GBC
GC
GP
Missense_Mutation
Nonsense_Mutation
Splice_Site
Frame_Shift_Del
Multi_Hit
pathway
C
UMAP_2
UMAP_1
ANT_1
ANT_10
ANT_11
ANT_12
ANT_13
ANT_14
ANT_15
ANT_16
ANT_2
ANT_3
ANT_4
ANT_5
ANT_6
ANT_7
ANT_8
ANT_9
GA_1
GA_2
GA_3
GA_4
GA_5
GBC_1
GBC_10
GBC_11
GBC_12
GBC_13
GBC_14
GBC_15
GBC_2
GBC_3
GBC_4
GBC_5
GBC_6
GBC_7
GBC_8
GBC_9
GC_1
GC_2
GC_3
GC_4
GP_1
GP_2
GP_3
D
Proportion
ANT
GC
GP
GA
GBC
G1
G2M
S
Proportion
T/NK
Myeloid
Fibroblast
Epithelium
Endothelium
B/Plasma
Mast
E
(n = 69 cell subsets)
TNFRSF11B, PLAT, KRT20, MT1X, DCN, SLC26A3, LY6D, SAA1, PAPLN, MUC5AC, FOLR1, REG3A, DEFA5, MSMB, KRT17, KRT7
SELL, CD40LG, MKI67, CCR8, FOXP3, NMB, CEBPD, IFNG, GZMK, ZNF683, IL7R, HAVCR2, KLRC1
KIT
S100A8, LYVE1, C1QC, MT1G, CXCL10, RGS16, CCL18, SPP1, CLEC9A, CD1C, LAMP3, PTGDS
MZB1, IGHG3, ZNF331, TCL1A
PROX1, ESM1, LY6H, ANGPT2, CA4, GJA5, PLVAP, IGFBP5
RGS5, PLA2G2A, NSG1, SLPI, ADIRF, DES, S100
ANT_4 specific
ANT_9 specific
ANT_16 specific
Epi–MT1X+
Epi–DCN+
Epi–SLC26A3+
Epi–LY6D+
Epi–SAA1+
Epi–PAPLN+
GP–specific
GP_1 specific
GMC
Paneth cells
Goblet cells
Malignant
Non–malignant
CD3–Tnaive
CD3–early Teff
CD3–Cycling
CD4–pre Treg
CD4–Treg
CD4–Tfh
CD8–Tc17
CD8–Teff
CD8–Tem
CD8–Trm
CD8–Tm
CD8–Tex
NK–like
NKT–like
Mast cell
Monocyte
Myeloid–Cycling
Macro–LYVE1
Macro–C1QC
Macro–MT1G
Macro–CXCL10
Macro–RGS16
Macro–CCL18
Macro–SPP1
cDC1
cDC2
cDC3
pDCs
B–Naive
B–Memory
B–Germinal center
Plasmablast
Endo–Capillary
Endo–PCV
Endo–Activated pcv
Endo–Mesencymal
Endo–IFN stimulated
Endo–Intermediate
Endo–Tip cell
Endo–Artery
Endo–Lymphatic
Endo–Cycling
Fibro–iCAF
Fibro–meCAF
Fibro–ecmCAF
Fibro–mesCAF
Fibro–synCAF
Fibro–Cycling
Fibro–IFN stimulated
Pericyte
SMC
Glial cell
Celltype
Expression Level
Low
High

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).

He H, *et al. Gut* 2024;0:1–14. doi: 10.1136/gutjnl-2023-331773

Figure S2

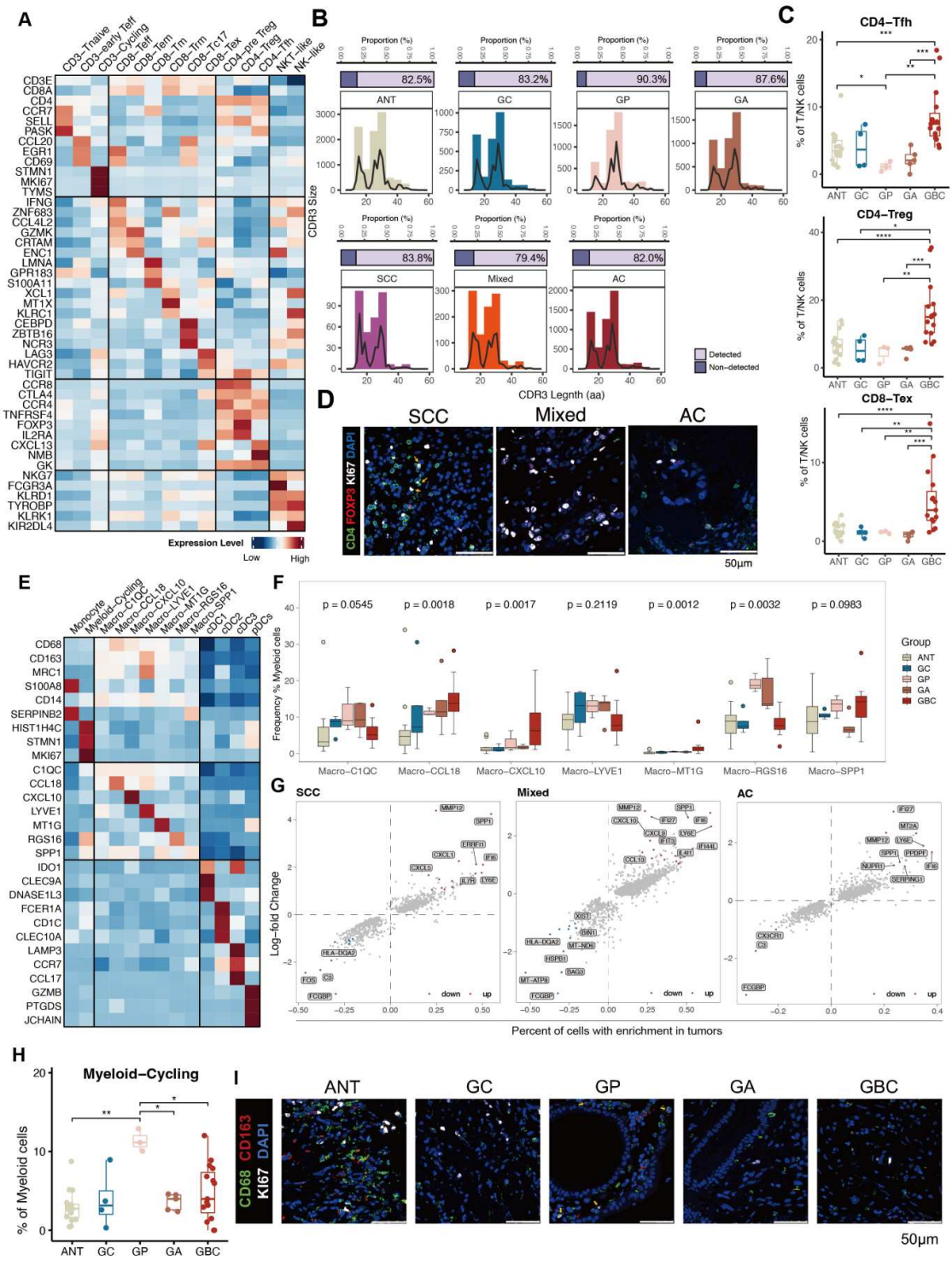

A
CD3_Tnaive
CD3_early Teff
CD3_Cycling
CD8_Teff
CD8_Tem
CD8_Trm
CD8_Tc17
CD8_Tex
CD4_pre Treg
CD4_Treg
CD4_Tfh
NKT_like
NK_like
CD3E
CD8A
CD4
CCR7
SELL
PASK
CCL20
EGR1
CD69
STMN1
MKI67
TYMS
IFNG
ZNF683
CCL4L2
GZMK
CRTAM
ENC1
LMNA
GPR183
S100A11
XCL1
MT1X
KLRC1
CEBPD
ZBTB16
NCR3
LAG3
HAVCR2
TIGIT
CCR8
CTLA4
CCR4
TNFRSF4
FOXP3
IL2RA
CXCL13
NMB
GK
NKG7
FCGR3A
KLRD1
TYROBP
KLRK1
KIR2DL4
Expression Level
Low
High
B
Proportion (%)
82.5%
ANT
83.2%
GC
90.3%
GP
87.6%
GA
83.8%
SCC
79.4%
Mixed
82.0%
AC
CDR3 Size
CDR3 Legnth (aa)
Detected
Non-detected
C
CD4−Tfh
CD4−Treg
CD8−Tex
% of T/NK cells
ANT
GC
GP
GA
GBC
D
SCC
Mixed
AC
CD4 FOXP3 KI67 DAPI
50μm
E
Monocyte
Myeloid−Cycling
Macro−C1QC
Macro−CCL18
Macro−CXCL10
Macro−LYVE1
Macro−MT1G
Macro−RGS16
Macro−SPP1
cDC1
cDC2
cDC3
pDCs
CD68
CD163
MRC1
S100A8
CD14
SERPINB2
HIST1H4C
STMN1
MKI67
C1QC
CCL18
CXCL10
LYVE1
MT1G
RGS16
SPP1
IDO1
CLEC9A
DNASE1L3
FCER1A
CD1C
CLEC10A
LAMP3
CCR7
CCL17
GZMB
PTGDS
JCHAIN
F
p = 0.0545
p = 0.0018
p = 0.0017
p = 0.2119
p = 0.0012
p = 0.0032
p = 0.0983
Frequency % Myeloid cells
Group
ANT
GC
GP
GA
GBC
Macro−C1QC
Macro−CCL18
Macro−CXCL10
Macro−LYVE1
Macro−MT1G
Macro−RGS16
Macro−SPP1
G
SCC
Mixed
AC
Log−fold Change
down
up
Percent of cells with enrichment in tumors
H
Myeloid−Cycling
% of Myeloid cells
I
ANT
GC
GP
GA
GBC
CD68 CD163
KI67 DAPI
50μm

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 $CD8^+$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.

He H, *et al. Gut* 2024;0:1–14. doi: 10.1136/gutjnl-2023-331773

Figure S3-1

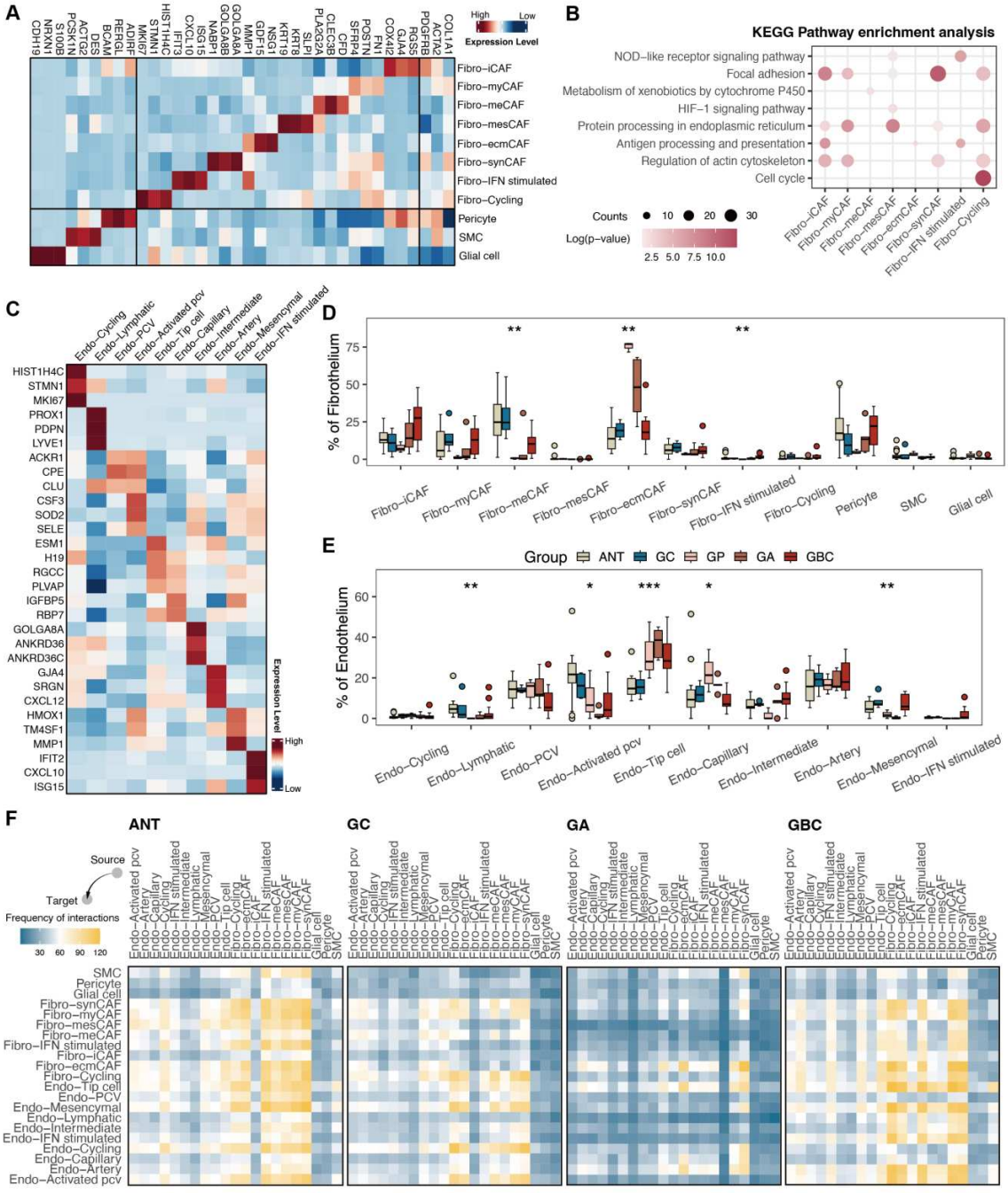

A
B
KEGG Pathway enrichment analysis
NOD−like receptor signaling pathway
Focal adhesion
Metabolism of xenobiotics by cytochrome P450
HIF−1 signaling pathway
Protein processing in endoplasmic reticulum
Antigen processing and presentation
Regulation of actin cytoskeleton
Cell cycle
Counts
Log(p−value)
C
D
% of Fibrothelium
E
Group ANT GC GP GA GBC
% of Endothelium
F
ANT
GC
GA
GBC
Source
Target
Frequency of interactions

Figure S3-2

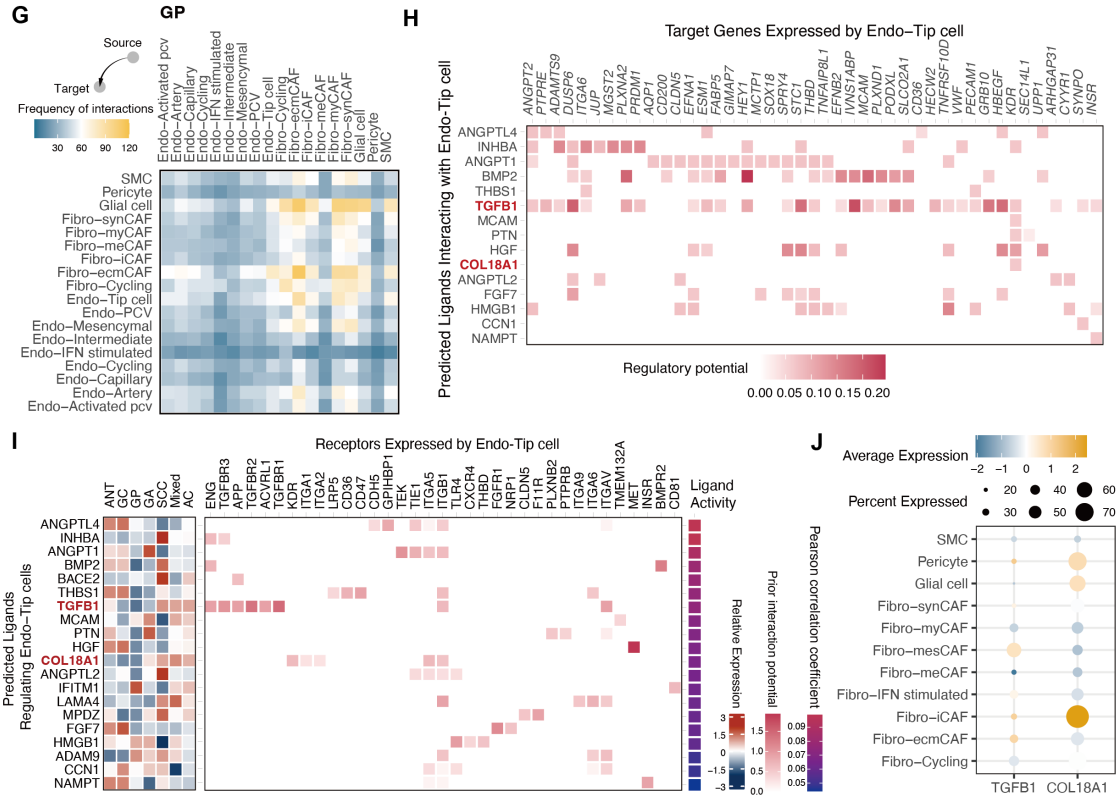

G
GP
Source
Target
Frequency of interactions
30 60 90 120
Endo–Activated pcv
Endo–Artery
Endo–Capillary
Endo–Cycling
Endo–IFN stimulated
Endo–Intermediate
Endo–Mesencymal
Endo–PCV
Endo–Tip cell
Fibro–Cycling
Fibro–ecmCAF
Fibro–iCAF
Fibro–meCAF
Fibro–myCAF
Fibro–synCAF
Glial cell
Pericyte
SMC
H
Target Genes Expressed by Endo–Tip cell
Predicted Ligands Interacting with Endo–Tip cell
ANGPTL4
INHBA
ANGPT1
BMP2
THBS1
TGFB1
MCAM
PTN
HGF
COL18A1
ANGPTL2
FGF7
HMGB1
CCN1
NAMPT
Regulatory potential
0.00 0.05 0.10 0.15 0.20
I
Receptors Expressed by Endo–Tip cell
Predicted Ligands Regulating Endo–Tip cells
Ligand Activity
Relative Expression
Prior interaction potential
Pearson correlation coefficient
J
Average Expression
Percent Expressed
TGFB1
COL18A1

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

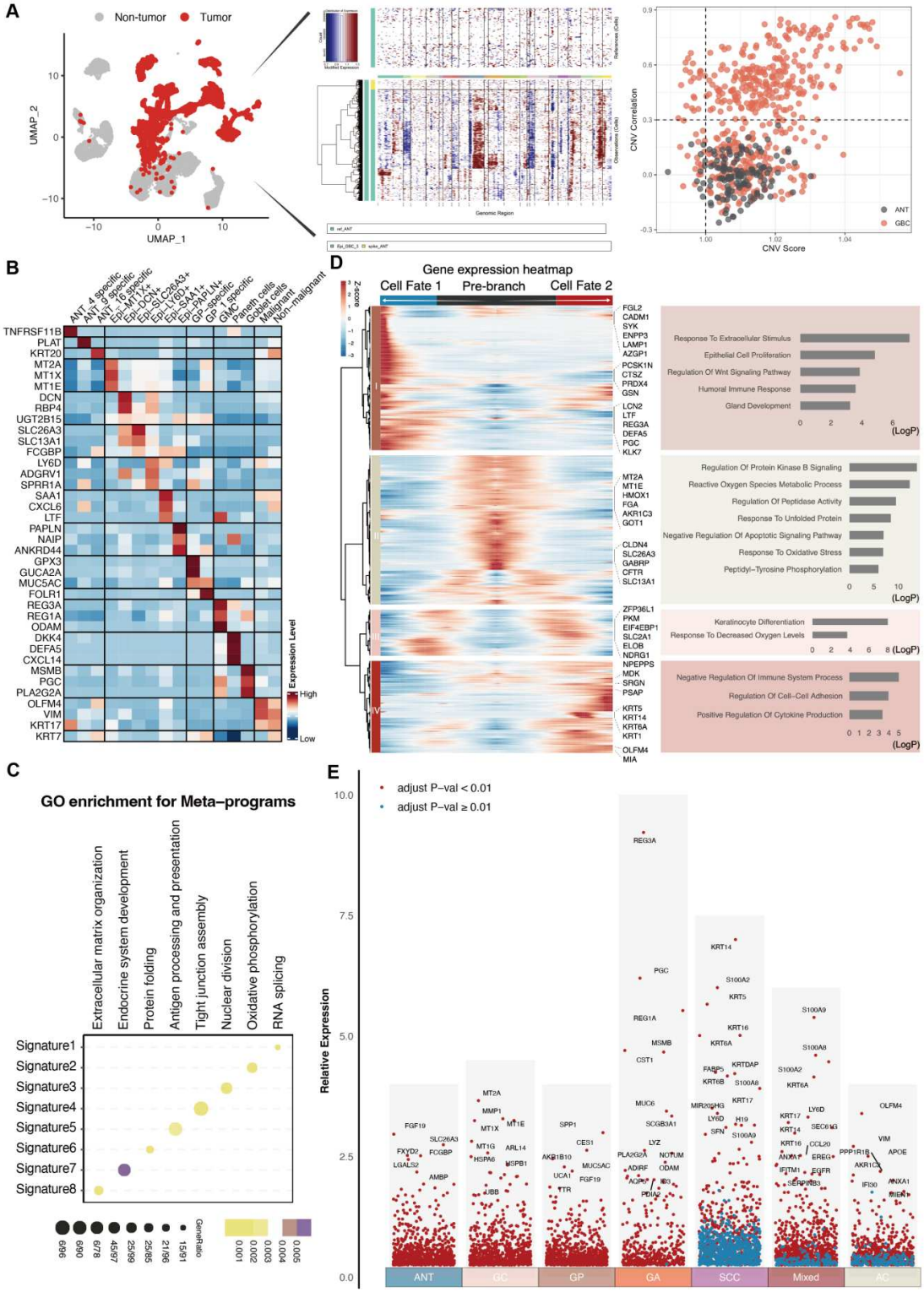

A
Non-tumor
Tumor
UMAP_2
UMAP_1
CNV Correlation
CNV Score
ANT
GBC
B
Expression Level
High
Low
C
GO enrichment for Meta−programs
Extracellular matrix organization
Endocrine system development
Protein folding
Antigen processing and presentation
Tight junction assembly
Nuclear division
Oxidative phosphorylation
RNA splicing
Signature1
Signature2
Signature3
Signature4
Signature5
Signature6
Signature7
Signature8
GeneRatio
D
Gene expression heatmap
Cell Fate 1
Pre-branch
Cell Fate 2
Z-score
Response To Extracellular Stimulus
Epithelial Cell Proliferation
Regulation Of Wnt Signaling Pathway
Humoral Immune Response
Gland Development
Regulation Of Protein Kinase B Signaling
Reactive Oxygen Species Metabolic Process
Regulation Of Peptidase Activity
Response To Unfolded Protein
Negative Regulation Of Apoptotic Signaling Pathway
Response To Oxidative Stress
Peptidyl−Tyrosine Phosphorylation
Keratinocyte Differentiation
Response To Decreased Oxygen Levels
Negative Regulation Of Immune System Process
Regulation Of Cell−Cell Adhesion
Positive Regulation Of Cytokine Production
(LogP)
E
adjust P−val < 0.01
adjust P−val ≥ 0.01
Relative Expression
ANT
GC
GP
GA
SCC
Mixed
AC

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).

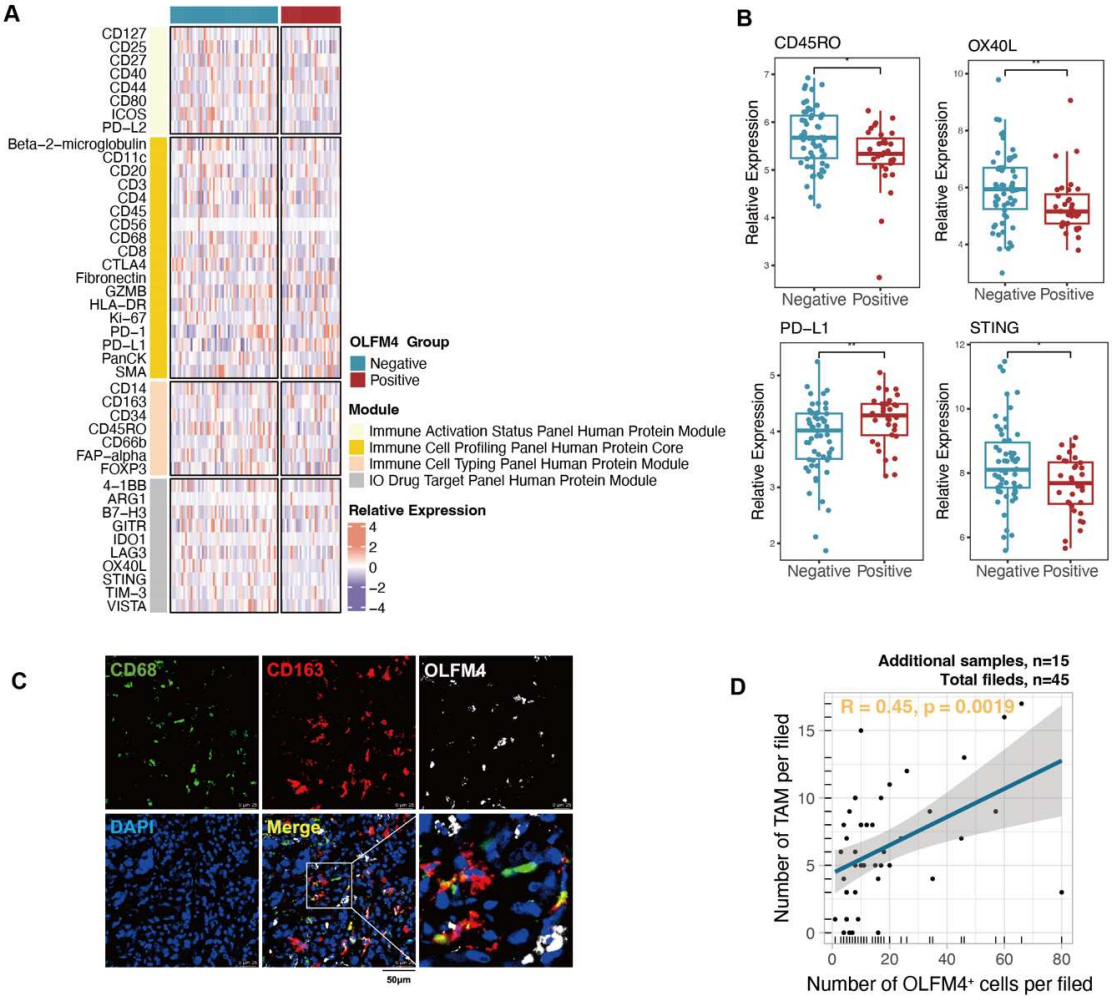
B. Heatmap showing the expression of representative genes of epithelium.

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.

Figrue S5-1


A
CD127
CD25
CD27
CD40
CD44
CD80
ICOS
PD-L2
Beta-2-microglobulin
CD11c
CD20
CD3
CD4
CD45
CD56
CD68
CD8
CTLA4
Fibronectin
GZMB
HLA-DR
Ki-67
PD-1
PD-L1
PanCK
SMA
CD14
CD163
CD34
CD45RO
CD66b
FAP-alpha
FOXP3
4-1BB
ARG1
B7-H3
GITR
IDO1
LAG3
OX40L
STING
TIM-3
VISTA
OLFM4 Group
Negative
Positive
Module
Immune Activation Status Panel Human Protein Module
Immune Cell Profiling Panel Human Protein Core
Immune Cell Typing Panel Human Protein Module
IO Drug Target Panel Human Protein Module
Relative Expression
4
2
0
−2
−4
B
CD45RO
OX40L
PD-L1
STING
Relative Expression
Negative Positive
C
CD68
CD163
OLFM4
DAPI
Merge
50μm
D
Additional samples, n=15
Total fileds, n=45
R = 0.45, p = 0.0019
Number of TAM per filed
Number of OLFM4+ cells per filed

Figrue S5-2

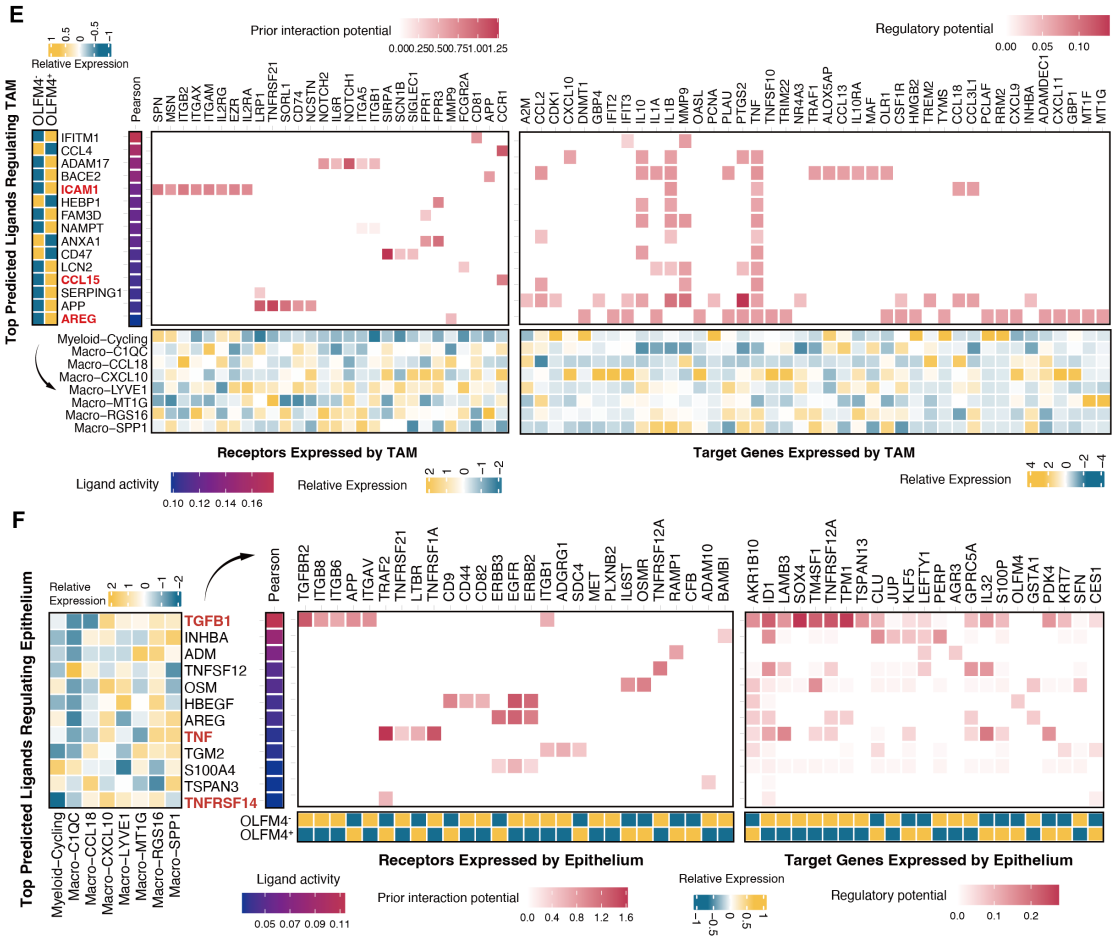

E
Prior interaction potential
Regulatory potential
Relative Expression
Top Predicted Ligands Regulating TAM
OLFM4⁺
OLFM4⁻
Pearson
IFITM1
CCL4
ADAM17
BACE2
ICAM1
HEBP1
FAM3D
NAMPT
ANXA1
CD47
LCN2
CCL15
SERPING1
APP
AREG
Myeloid–Cycling
Macro–C1QC
Macro–CCL18
Macro–CXCL10
Macro–LYVE1
Macro–MT1G
Macro–RGS16
Macro–SPP1
Receptors Expressed by TAM
Target Genes Expressed by TAM
Ligand activity
F
Top Predicted Ligands Regulating Epithelium
TGFB1
INHBA
ADM
TNFSF12
OSM
HBEGF
AREG
TNF
TGM2
S100A4
TSPAN3
TNFRSF14
Receptors Expressed by Epithelium
Target Genes Expressed by Epithelium

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 ($CD68^{+}CD1163^{+}$) 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 $OLFM4^{+}$ 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

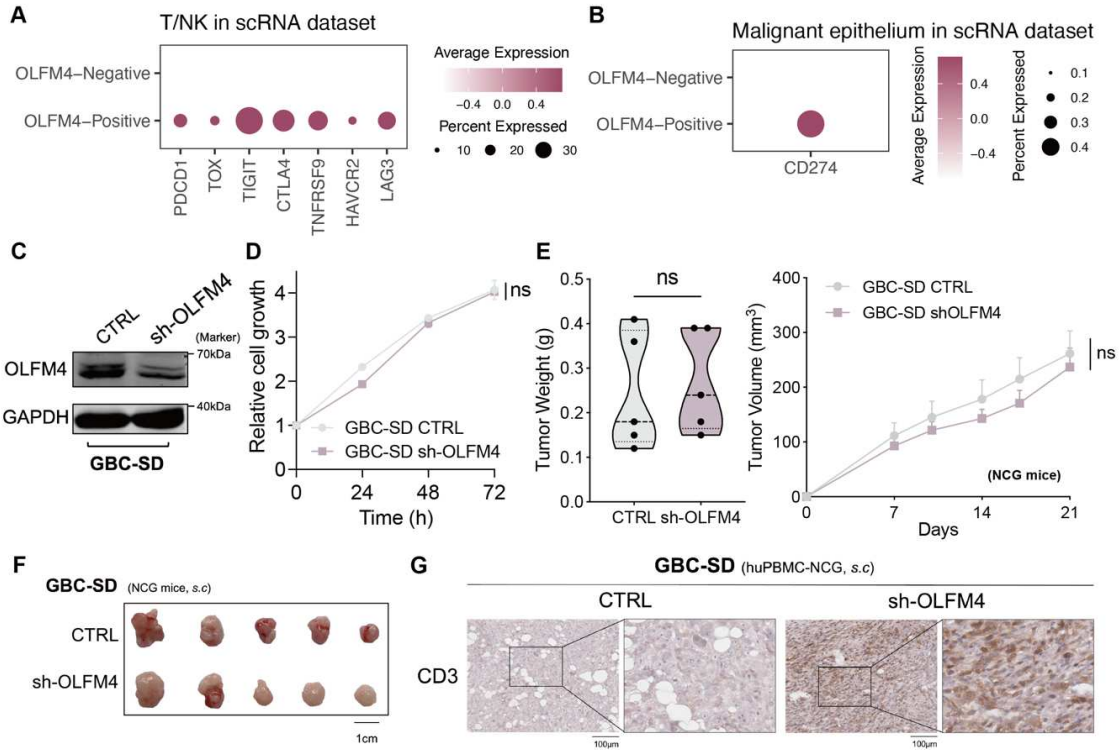

A
T/NK in scRNA dataset
OLFM4–Negative
OLFM4–Positive
PDCD1
TOX
TIGIT
CTLA4
TNFRSF9
HAVCR2
LAG3
Average Expression
Percent Expressed
B
Malignant epithelium in scRNA dataset
CD274
C
CTRL
sh-OLFM4
OLFM4
GAPDH
GBC-SD
D
Relative cell growth
GBC-SD CTRL
GBC-SD sh-OLFM4
Time (h)
ns
E
Tumor Weight (g)
CTRL sh-OLFM4
Tumor Volume (mm3)
GBC-SD CTRL
GBC-SD shOLFM4
(NCG mice)
Days
F
GBC-SD (NCG mice, s.c)
CTRL
sh-OLFM4
1cm
G
GBC-SD (huPBMC-NCG, s.c)
CTRL
sh-OLFM4
CD3

Figure S6-2

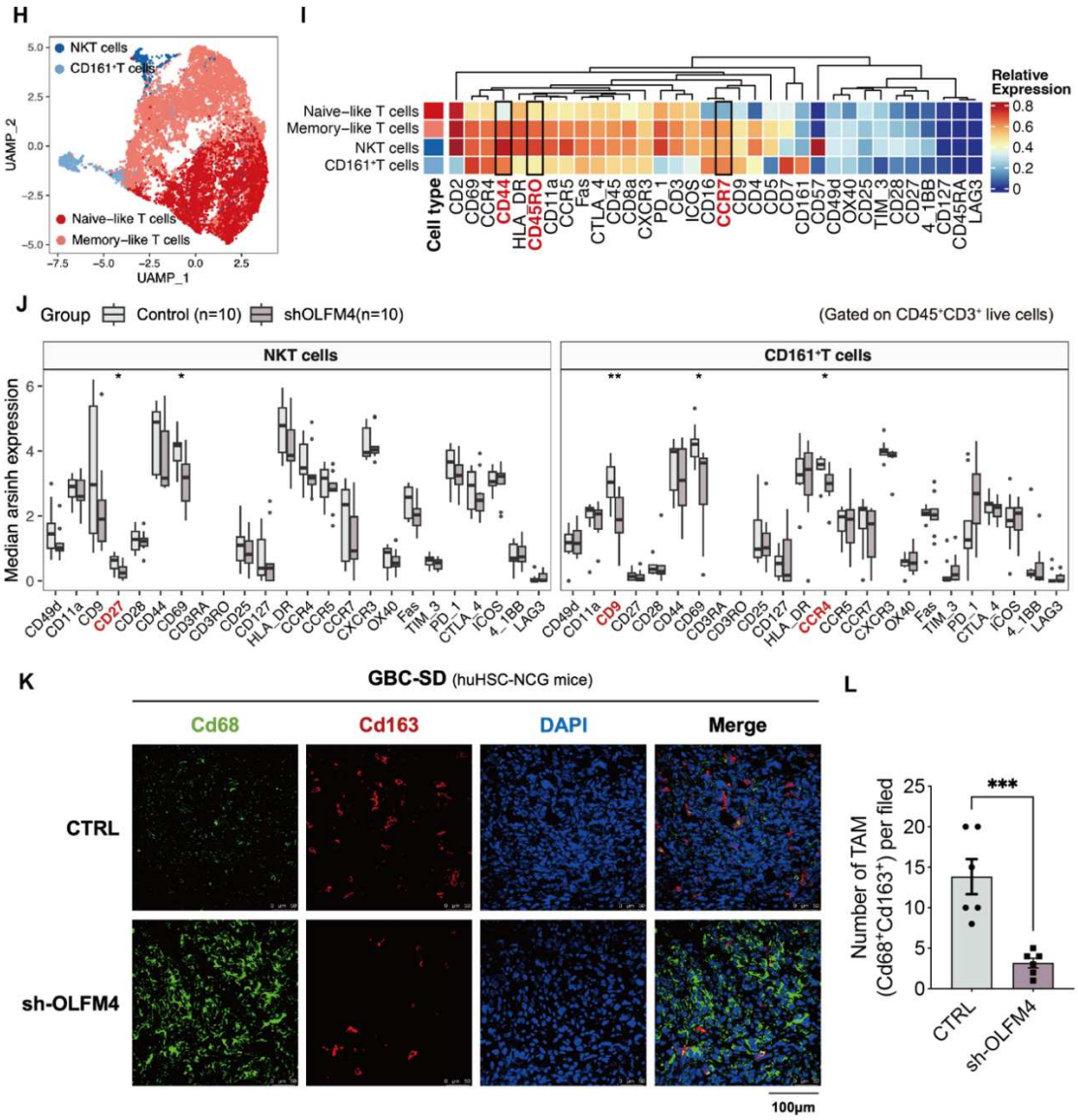

H
NKT cells
CD161⁺T cells
Naive−like T cells
Memory−like T cells
UMAP_1
UMAP_2
I
Relative Expression
Cell type
J
Group Control (n=10) shOLFM4(n=10)
(Gated on CD45⁺CD3⁺ live cells)
NKT cells
CD161⁺T cells
Median arsinh expression
K
GBC-SD (huHSC-NCG mice)
Cd68
Cd163
DAPI
Merge
CTRL
sh-OLFM4
100μm
L
Number of TAM (Cd68⁺Cd163⁺) per filed
***
CTRL
sh-OLFM4

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 $CD161^+$ 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 $CD34^+$ humanized mice across groups.

Figure S7

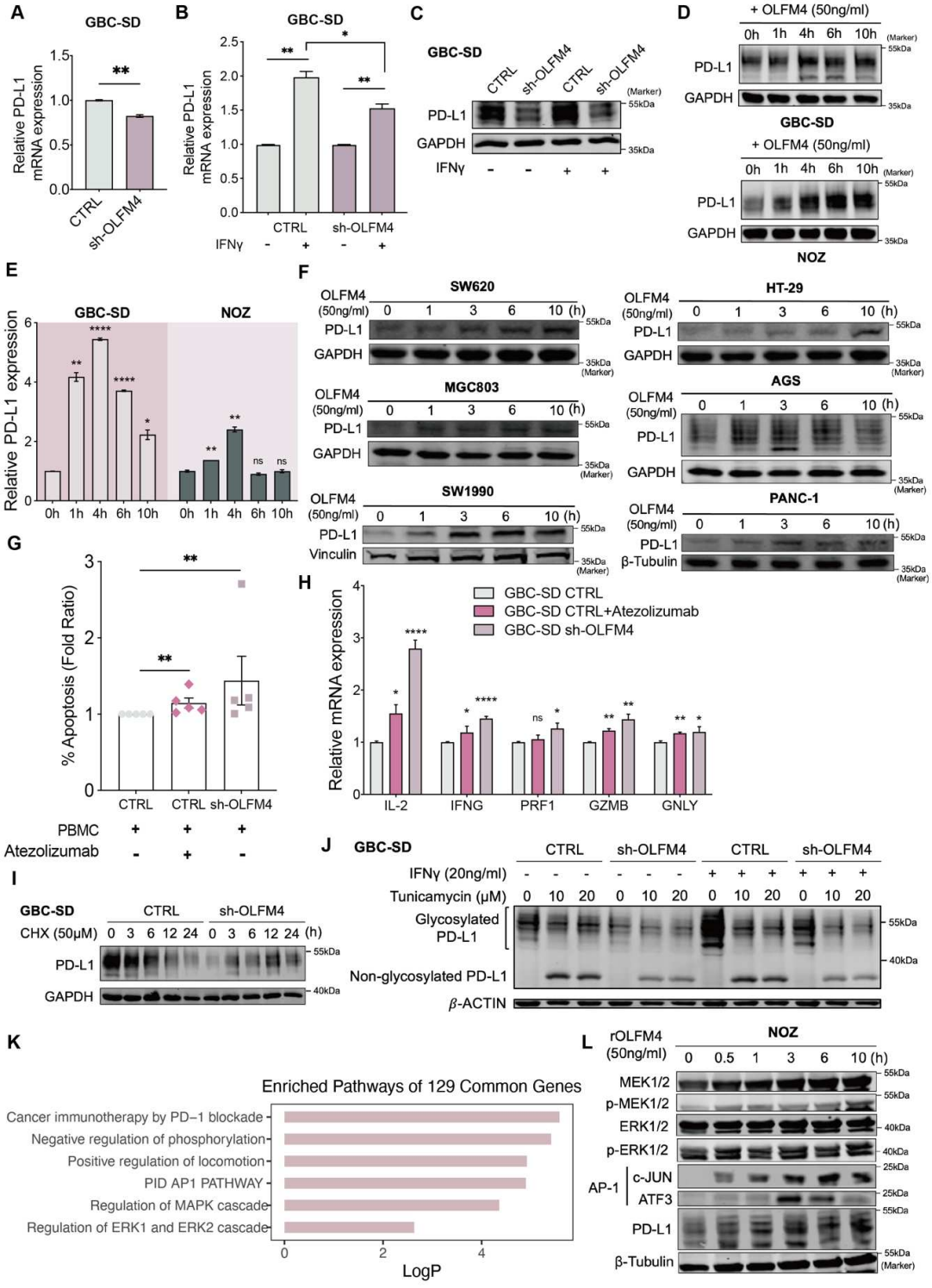

A
GBC-SD
Relative PD-L1 mRNA expression
CTRL
sh-OLFM4
B
GBC-SD
IFNγ
C
GBC-SD
PD-L1
GAPDH
D
+ OLFM4 (50ng/ml)
0h 1h 4h 6h 10h
NOZ
E
GBC-SD
NOZ
Relative PD-L1 expression
F
SW620
HT-29
MGC803
AGS
SW1990
PANC-1
Vinculin
β-Tubulin
G
% Apoptosis (Fold Ratio)
PBMC
Atezolizumab
H
GBC-SD CTRL
GBC-SD CTRL+Atezolizumab
GBC-SD sh-OLFM4
Relative mRNA expression
IL-2
IFNG
PRF1
GZMB
GNLY
I
CHX (50μM)
J
Tunicamycin (μM)
Glycosylated PD-L1
Non-glycosylated PD-L1
β-ACTIN
K
Enriched Pathways of 129 Common Genes
Cancer immunotherapy by PD−1 blockade
Negative regulation of phosphorylation
Positive regulation of locomotion
PID AP1 PATHWAY
Regulation of MAPK cascade
Regulation of ERK1 and ERK2 cascade
LogP
L
rOLFM4 (50ng/ml)
MEK1/2
p-MEK1/2
ERK1/2
p-ERK1/2
AP-1
c-JUN
ATF3

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 IFNγ stimulation or unstimulated conditions.

C. PD-L1 protein levels in GBC-SD CTRL/sh-OLFM4 under IFNγ 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 μg/mL) treatment.

H. Quantitative RT-PCR was performed to detect IL-2 (Interleukin-2), IFNG (IFNg), 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 μg/mL) treatment.

I. PD-L1 protein levels in GBC-SD CTRL/sh-OLFM4 in the presence or absence of CHX (50 μM).

J. PD-L1 protein levels in GBC-SD CTRL/sh-OLFM4 under IFNγ stimulation or unstimulated conditions, in the presence or absence of tunicamycin (10 or 20 μM, 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 μm) (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_2O_2 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 IFNγ, CD107a, IL-2, perforin, granulysin, and Granzyme B. Peripheral blood mononuclear cells (PBMCs) from healthy donors were pre-activated using anti-CD3 (5μg/mL, Biolegend) and anti-CD28 (5μg/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*^{em26Cd52}*Il2rg*^{em26Cd22}/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 5×10^6 were injected into the right flank of NCG mice. Tumor size (calculated as length × width^2 × 0.5) was assessed twice per week following the injection. PBMCs from healthy donors were activated and expanded as described previously[1,2]. On the day before tumor cell injection, PBMCs (1×10^7 cells) were adoptively transferred to NCG mice via the tail vein.

$CD34^+$ 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, $CD34^+$ 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 5μM 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 μm) were baked at 60°C for 1.5 hours, and then deparaffinized and rehydrated as follows: 3×5 min in CitriSolv, 2×5 min in 100% ethanol, 2×5 min in 95% ethanol, and 2×5 min in double-distilled water. For antigen retrieval, slides were placed in a staining jar containing 1× 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× 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 3×10 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 2×5 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 $PANCK^+OLFM4^+$ Epithelium was conducted utilizing the High-Plex FL module. The OLFM4-Positive group was identified when $PANCK^+OLFM4^+$ 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)[3]. 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[4] 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[5]. 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[6] 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[7], 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[8]. 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[9], 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[10], we generated cell type signatures by selecting the top differentially expressed genes (with an average $Log_2FC > 1$) 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[11] 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.

He H, *et al. Gut* 2024;0:1–14. doi: 10.1136/gutjnl-2023-331773

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.