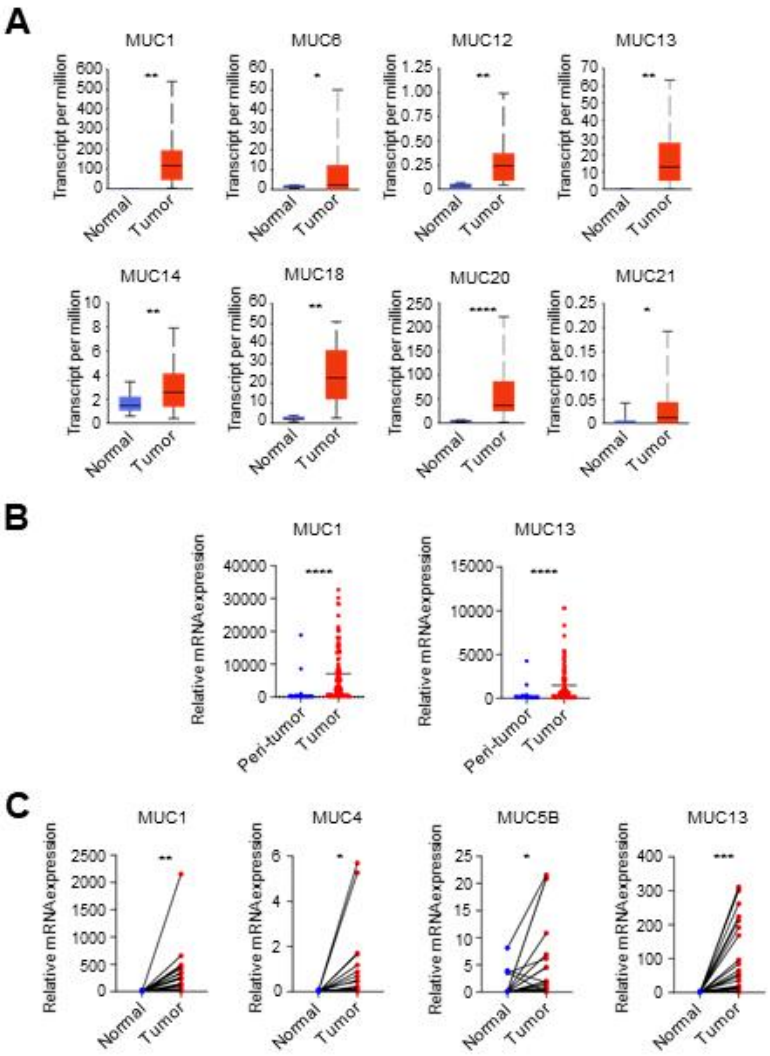
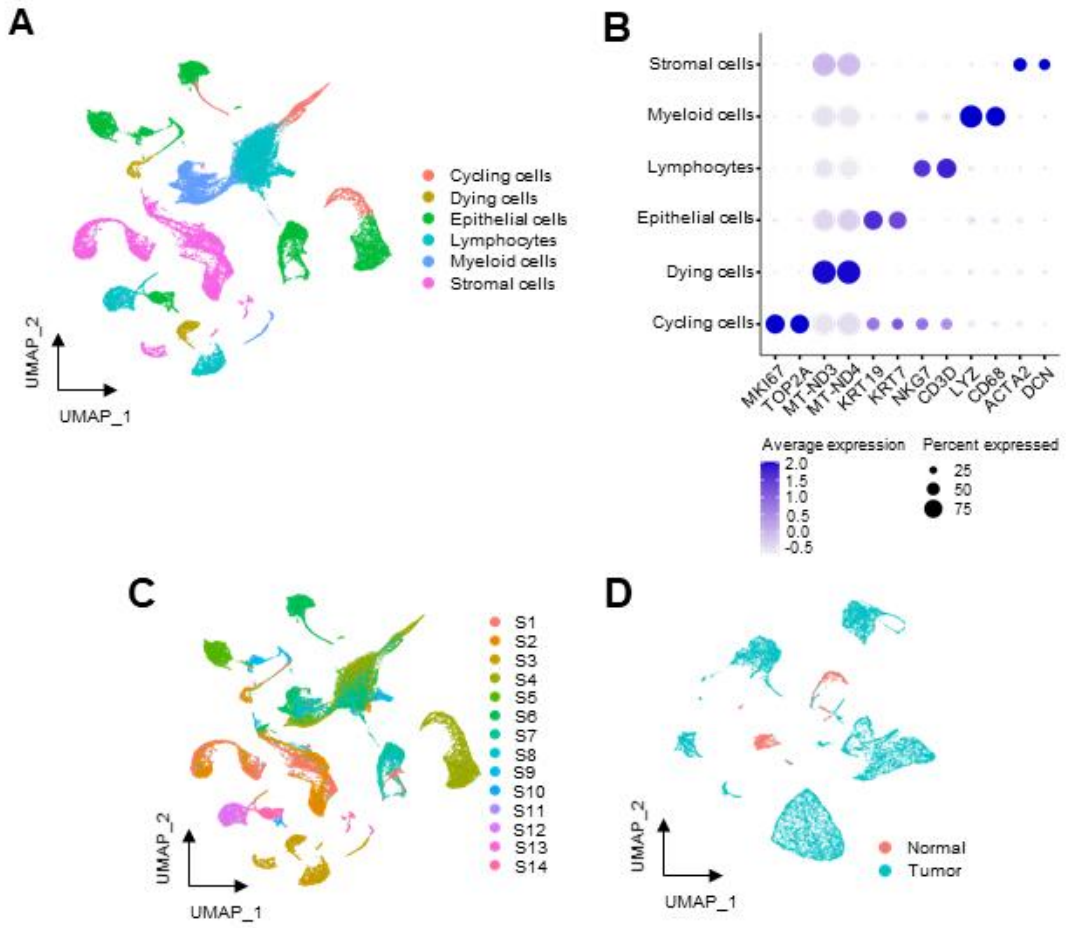
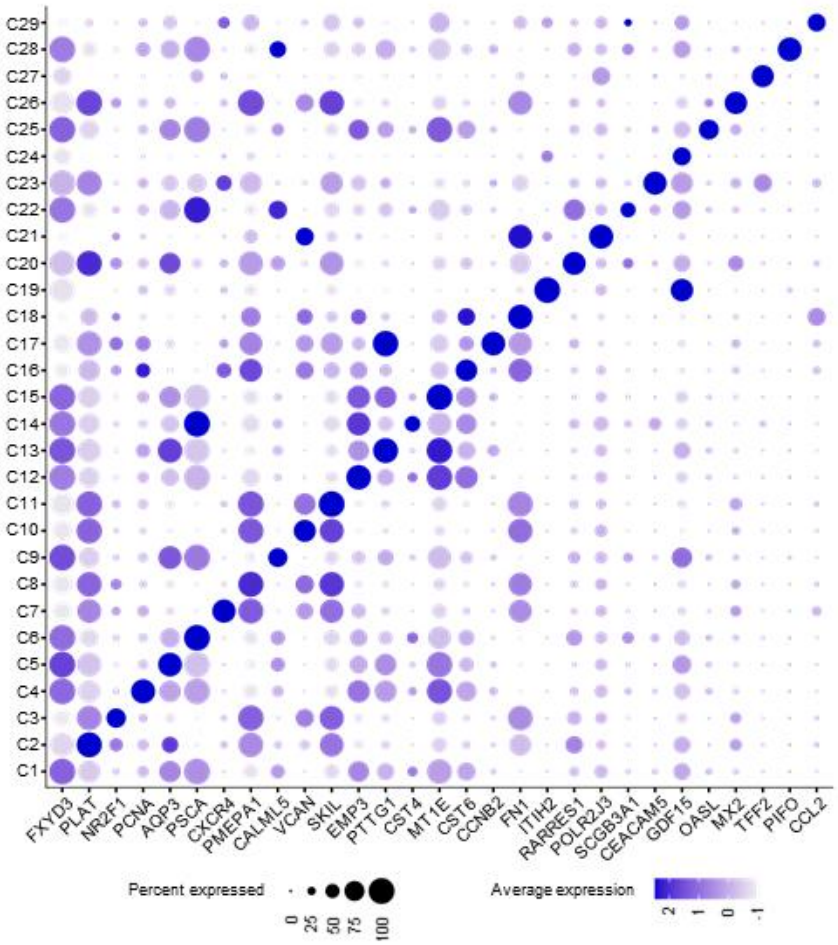


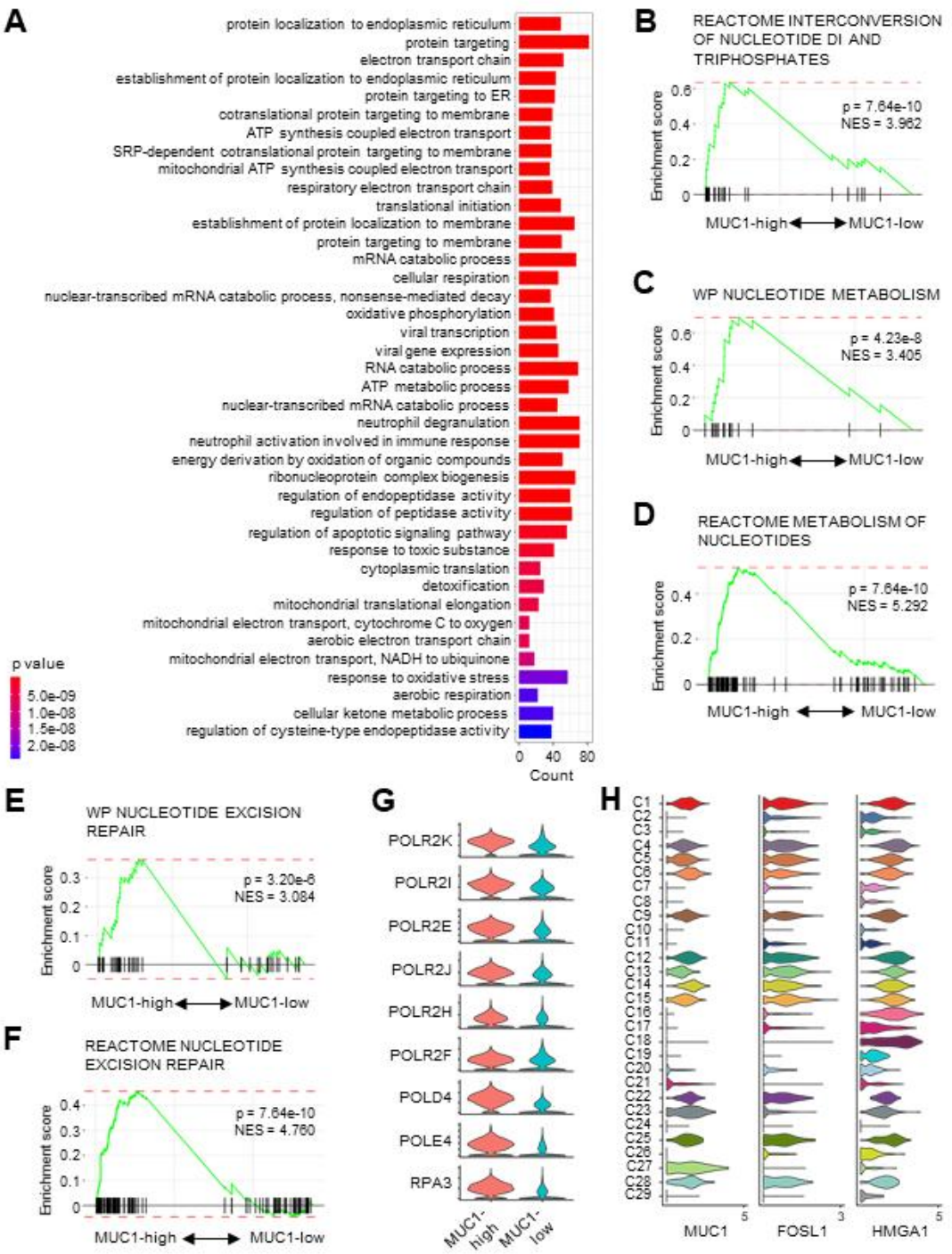
Supplementary Figure 1 Overexpression of mucins in cholangiocarcinoma tumor tissues. A: Box plots showing the different expression levels of MUC genes. The expression levels of MUC1, MUC6, MUC12, MUC13, MUC14, MUC18, MUC20, and MUC21 were detected in tumor versus normal tissues from cohort 1, $p < 0.05$ *, $p < 0.01$ **, $p < 0.0001$ ****; B: Dot plots showing the different expression level of MUC genes in peritumoral (n = 59) and tumor (n = 104) tissues of cholangiocarcinoma (CCA) from cohort 4, $p < 0.0001$ ****; C: Dot plots showing the different expression level of MUC genes in paired (n = 27) CCA tumor and peritumoral normal regions from cohort 3, $p < 0.05$ *, $p < 0.01$ **, $p < 0.001$ ***.



Supplementary Figure 2 Single cell sequencing of cholangiocarcinoma samples. A: Uniform manifold approximation and projection (UMAP) plot of the integrated scRNA-seq data divided by cell partition; B: Dot plot showing the marker gene expression levels and proportions in CCA sub-clusters; C: UMAP plot of the integrated scRNA-seq data divided by sample origin. Cells from different patients mixed well, especially immune cells and stromal cells; D: UMAP plot of the epithelial partition in integrated scRNA-seq data divided by tissue origin. Cells from normal tissues are shown in pink, and cells from tumor tissues are shown in blue.

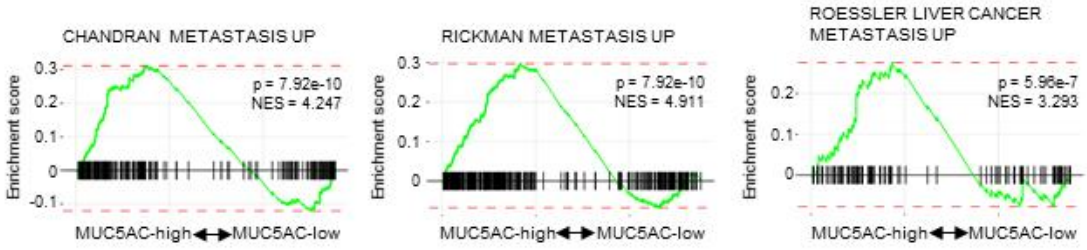


Supplementary Figure 3 Dot plot showing the marker gene expression levels and proportions in 29 cholangiocarcinoma tumor cell sub-clusters.

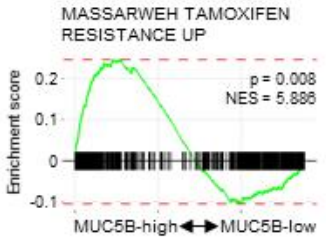


Supplementary Figure 4 Metabolic characteristics of *MUC1*-high tumor cells of cholangiocarcinoma. A: Top 40 enriched signaling pathways of the differentially expressed genes (DEGs) in *MUC1*-high cells; B – F: Gene set enrichment analysis (GSEA) plots showing the enrichment of genes from the *MUC1*-high cells (left) versus *MUC1*-low cells (right) in nucleotide metabolic signaling pathways, p values and normalized enrichment score (NES) are indicated on the plots; G: Violin plots showing the relative expression levels

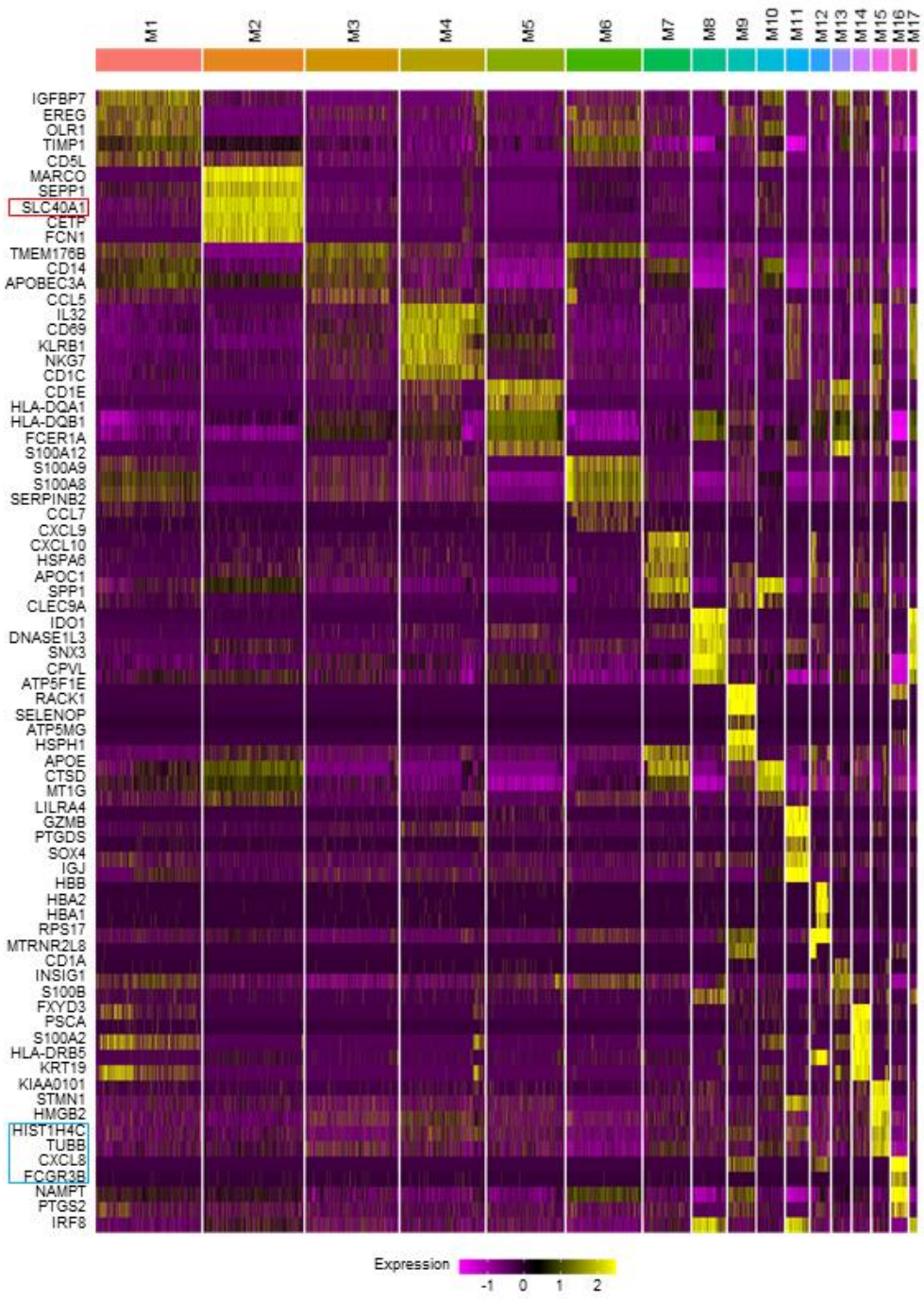
of RNA polymerase components and DNA repair genes in *MUC1*-high cells versus *MUC1*-low cells; H: Violin plots showing relative expression levels of *MUC1*, *FOSL1*, and *HMGA1* in CCA tumor sub-clusters.



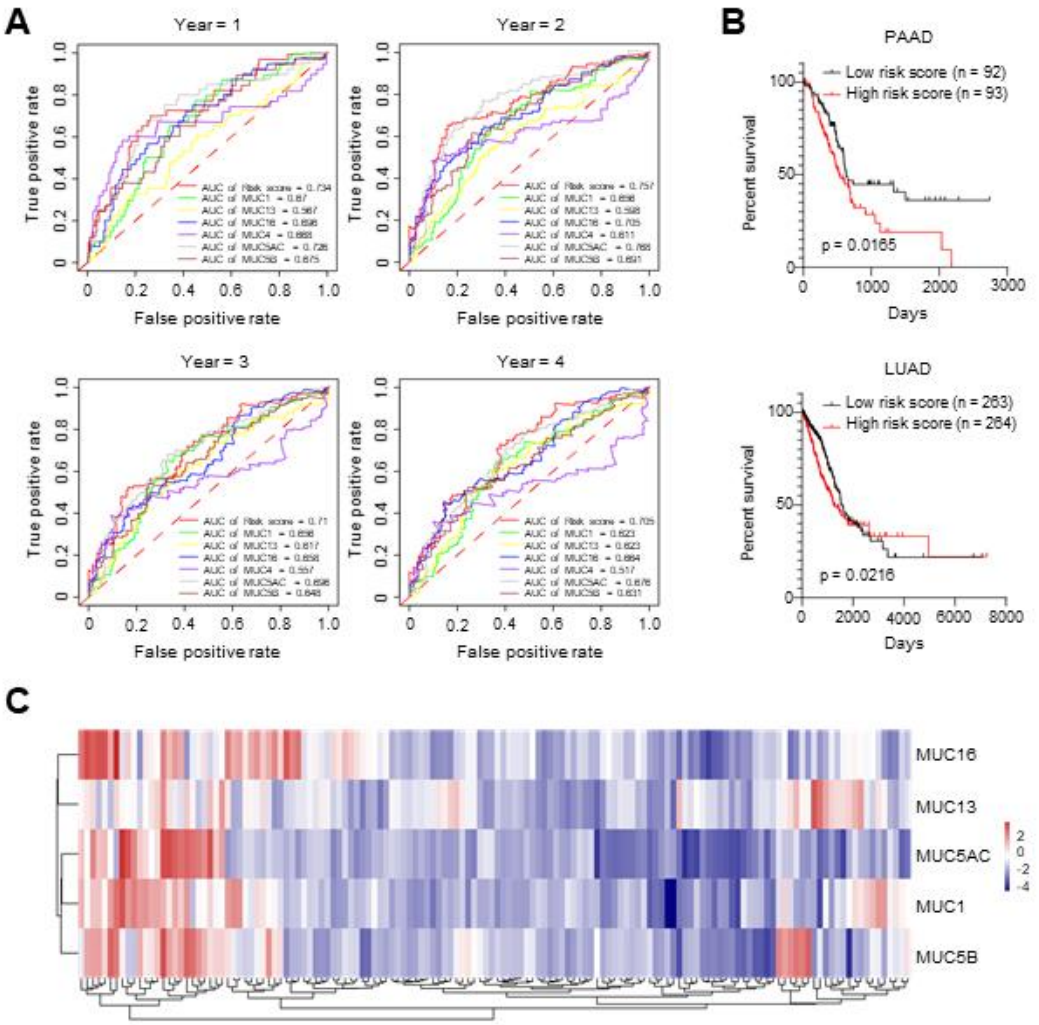
Supplementary Figure 5 Gene set enrichment analysis results showing the enrichment of genes from the *MUC5AC*-high cells (left) versus *MUC5AC*-low cells (right) in metastasis signaling pathways, p values and NES are indicated on the plots.



Supplementary Figure 6 Gene set enrichment analysis plots showing the enrichment of genes from the *MUC5B*-high cells (left) versus *MUC5B*-low cells (right) in tamoxifen-resistance signaling pathway, p value and NES are indicated on the plot.



Supplementary Figure 7 Heatmap showing the marker genes of each macrophage sub-clusters. The red frame shows specific high expression of *SLC40A1* in M2 sub-cluster. The blue frame shows specific high expression of *CXCL8*, *FCGR3B*, *NAMPT*, and *PTGS2* in M16 sub-cluster.



Supplementary Figure 8 Clinical significance of mucins. A: Area under curve of risk factor versus different mucins in evaluating the prognosis of CCA patients in different time points; B: Survival curves generated using the pancreatic carcinoma (PAAD) and lung adenocarcinoma (LUAD) data showing the relationship between patient survival and the risk factor; C: Protein levels of MUC1, MUC13, MUC16, MUC5AC and MUC5B in CCA samples from cohort 2.

Supplementary Table 1 Brief information of the clinical cohorts in our study

Cohort number	Identifier	Data type	Number of samples	Tumor site	Sample source
Cohort 1	TCGA database	Bulk RNA-seq data	T (n = 36) N (n = 9)	iCCA + Extrahepatic CCA	Global
Cohort 2	OEP001105, NODE database	Bulk RNA-seq data and protein mass spectrometry data	T (n = 262)	iCCA	China
Cohort 3	GSE26566, GEO database	Microarray	T (n = 104) N (n = 59)	iCCA	America, Belgium, Australia
Cohort 4	GSE107943, GEO database	Bulk RNA-seq data	T (n = 27) N (n = 27)	iCCA	Korea
Cohort 5	GSE154170, GEO database	scRNA-seq data	T (n = 1)	pCCA	America
Cohort 6	GSE138709, GEO database	scRNA-seq data	T (n = 6) N (n = 3)	iCCA	China
Cohort 7	GSE142784, GEO database	scRNA-seq data	T (n = 2)	iCCA	China
Cohort 8	GSE125449, GEO database	scRNA-seq data	T (n = 5)	iCCA	America
Cohort 9	HRA000437, GSA human	ST data	T (n = 1)	iCCA	China
Cohort 10	GSE89749, GEO database	Bulk RNA-seq data	T (n = 112)	iCCA + pCCA	Global

TCGA: The cancer genome atlas; NODE: National omics data encyclopedia; GEO: Gene expression omnibus; GSA: Genome sequence archive; scRNA-seq:

Single cell RNA-sequencing; ST: Spatial transcriptomics; T: Tumor; N: Normal; CCA: cholangiocarcinoma; iCCA: intrahepatic cholangiocarcinoma; pCCA: perihilar cholangiocarcinoma.

Supplementary Table 2 Brief information of samples involved in the single cell RNA-sequencing data

Patient number	Original identifier	Age	Gender	Stage	Tumor site	Virus infection	Sample source
S1	24S	NR	Male	II	iCCA	None	GSE142784
S2	32S	NR	Male	II	iCCA	None	GSE142784
S3	AJ064	NR	NR	NR	pCCA	NR	GSE154170
S4	ICC18	NR	Female	III	iCCA	HBV	GSE138709
S5	ICC20	NR	Female	II	iCCA	None	GSE138709
S6	ICC23	NR	Male	III	iCCA	HBV	GSE138709
S7	ICC24.1	NR	Male	II	iCCA	None	GSE138709
S8	ICC24.2	NR	Male	II	iCCA	None	GSE138709
S9	ICC25	NR	Male	III	iCCA	None	GSE138709
S10	C25	47	Female	IV	iCCA	None	GSE125449
S11	C39	61	Male	IV	iCCA	None	GSE125449
S12	C60	80	Female	III	iCCA	None	GSE125449
S13	C56	52	Female	IV	iCCA	None	GSE125449
S14	C66	71	Female	IV	iCCA	None	GSE125449

NR: Not reported.