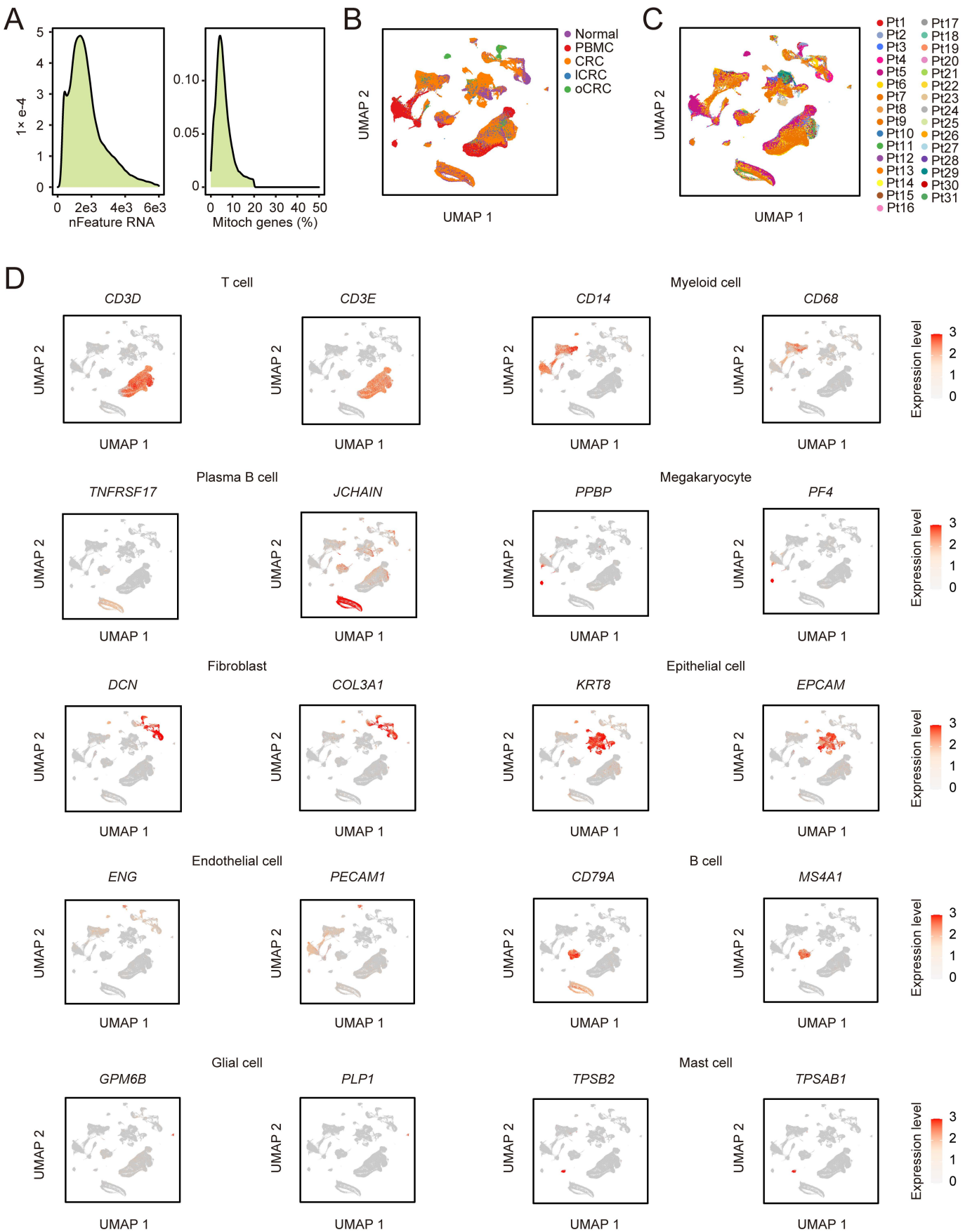
supplemental figure 1

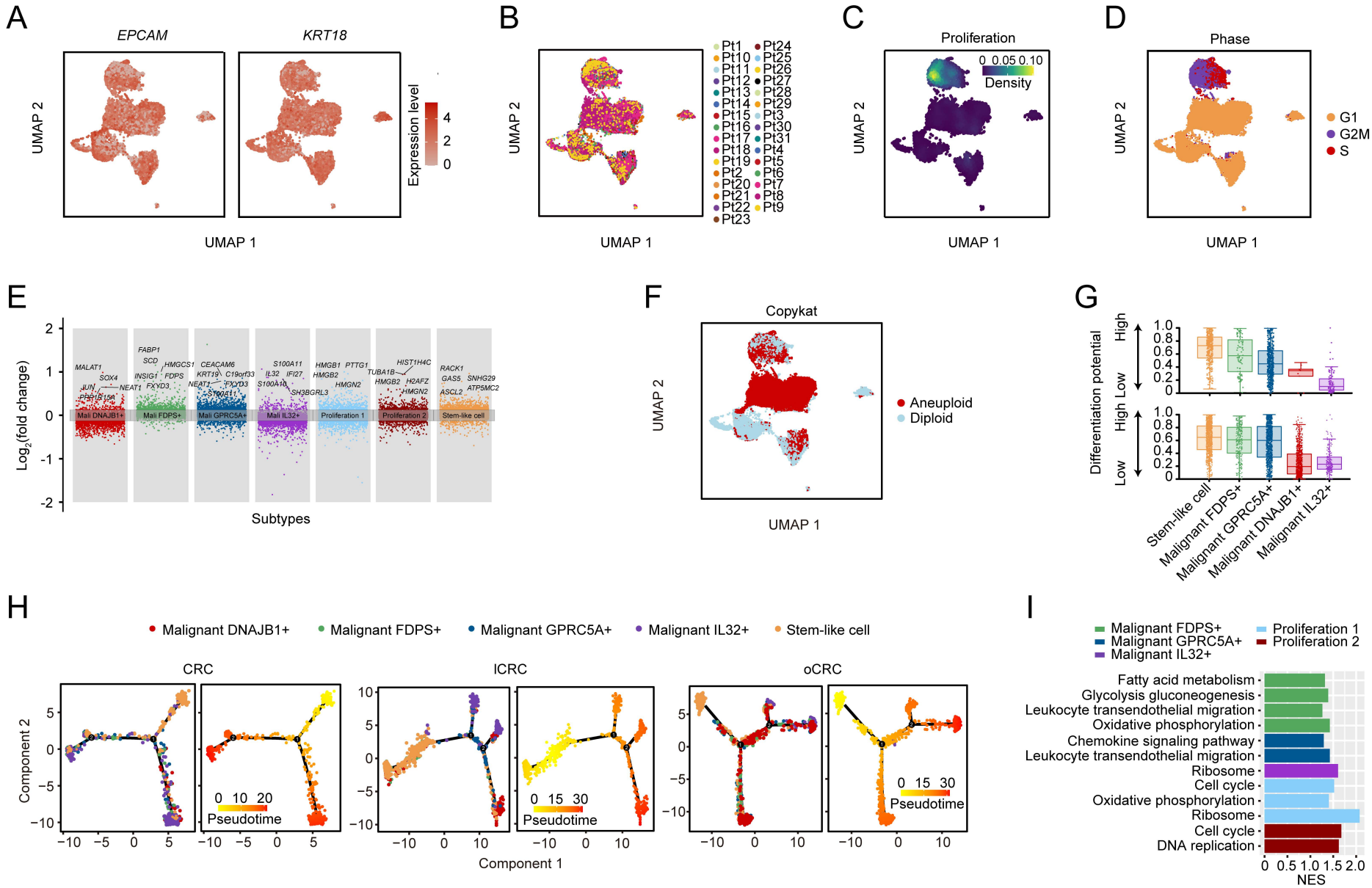


Supplemental figure 1 scRNA-seq profiling of all cells in CRC. Related to Figure 1.

(A) Histogram indicating the total number of detected genes (left panel) and mitochondrial genes (right panel). (B-C) UMAP plot showing different tissue origin (B) and patient origin (C) by color. (D) UMAP plot showing the expression levels of marker genes, defined for all cell types.

Li R, *et al*. *Gut* 2024; 73:470–484. doi: 10.1136/gutjnl-2023-330243

supplemental figure 2

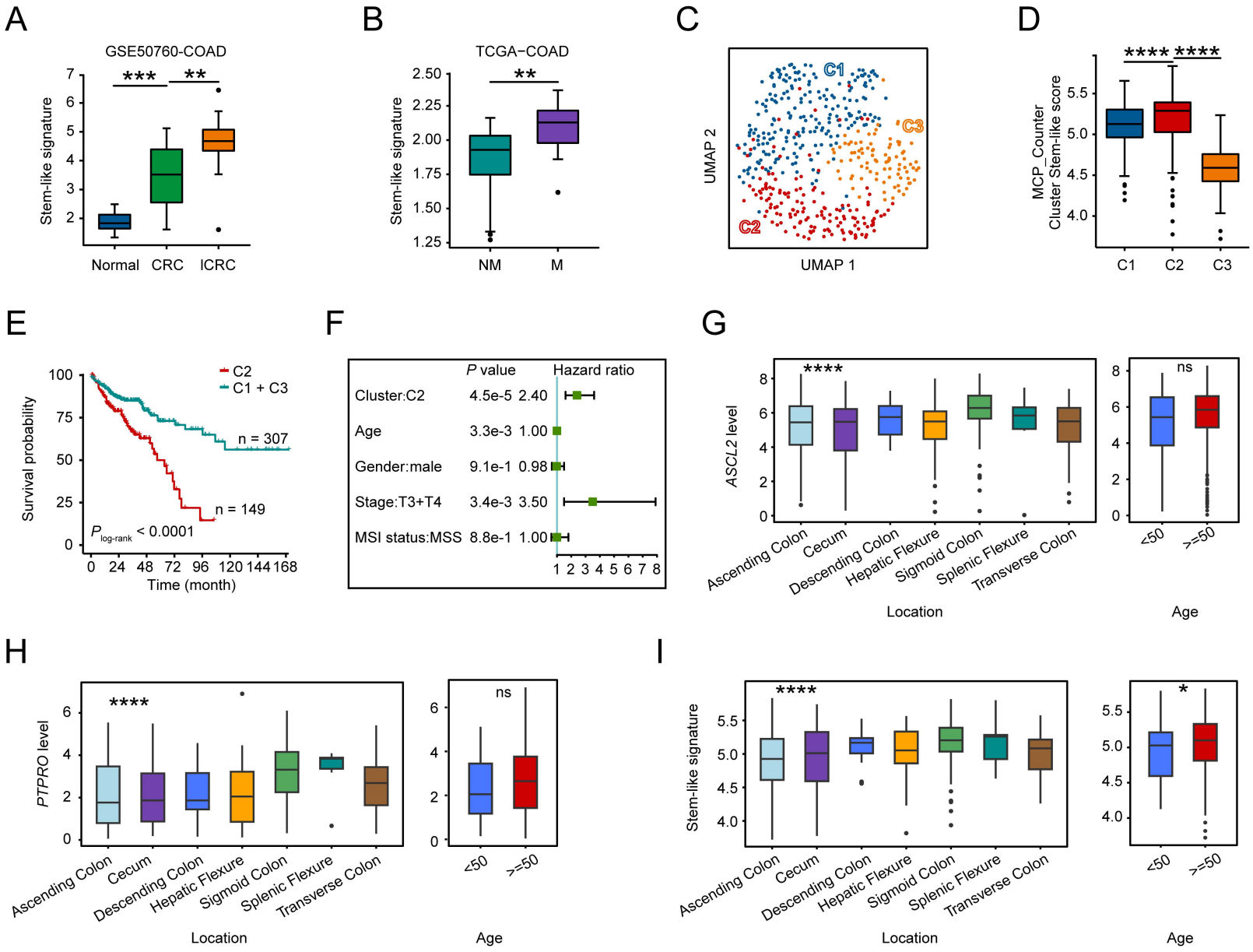


Supplemental figure 2 Intrinsic epithelial cell subtypes underlying tumor metastasis. Related to Figure 2.

(A) UMAP plot showing the expression levels of *EPCAM* and *KRT18*, defined for epithelial cells. (B) UMAP plot of epithelial cells showing different patient origins. (C) UMAP plot showing the proliferation score of all epithelial cells. (D) UMAP plot showing the cell cycle phase at the single-cell level. (E) Representative upregulated genes of each cell subtype based on DEG analysis. (F) Distribution of aneuploid and diploid cells. (G) Distribution of CytoTRACE scores among 5 malignant cell subtypes in oCRC (upper panel) or lCRC (lower panel). (H) Pseudotime-ordered analysis of 5 malignant cell subtypes in primary CRC, oCRC and lCRC inferred by monocle2. (I) Enrichment of Hallmark pathways in epithelial cell subtypes. The x-axis represents the NES value of each pathway (FDR < 0.05 and NES > 1).

Li R, *et al*. *Gut* 2024; 73:470–484. doi: 10.1136/gutjnl-2023-330243

supplemental figure 3

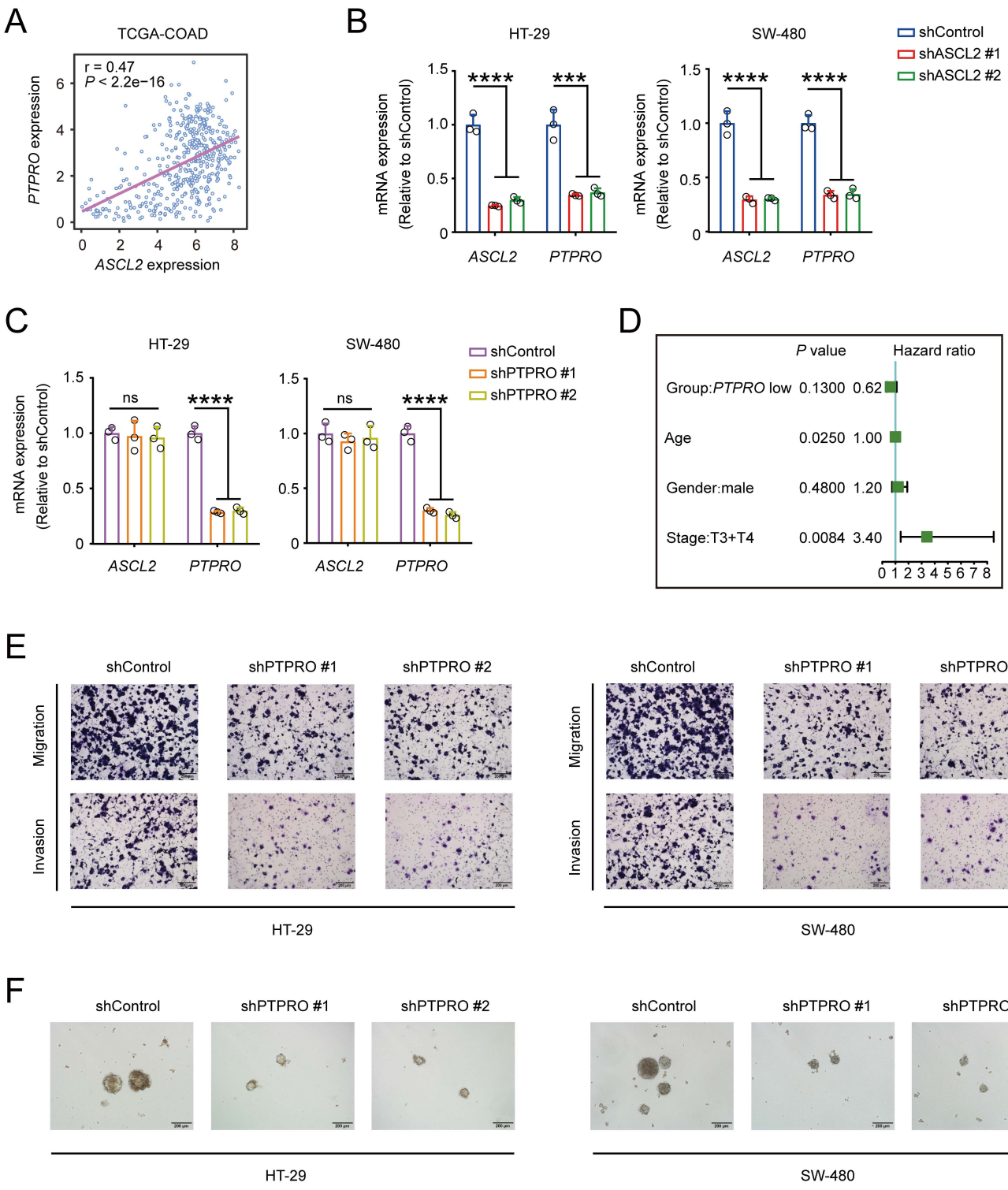


Supplemental figure 3 The correlation between stem-like gene signature scores and survival time in TCGA-COAD patient cohort. Related to Figure 2.

(A) Comparison of stem-like cell signature scores in normal, primary CRC, and liver CRC metastases samples (GSE50760-COAD cohort). (B) Comparison of stem-like cell signature scores in non-metastatic CRC and metastatic CRC (TCGA-COAD cohort). (C) UMAP plot showing unsupervised clustering of TCGA-COAD patients. (D) Comparison of stem-like cell signature scores among C1, C2 and C3 cell populations shown in (C). (E) Kaplan–Meier estimation of overall survival time in patients by different clusters. (F) Multivariate Cox regression analysis of TCGA-COAD data shows C2 cluster and tumor stage were risk factors for death of CRC. Bars represent 95% confidence interval of hazard ratios. (G-I) Comparison of *ASCL2* (G), *PTPRO* (H) expression levels or stem-like signature scores (I) in different groups of CRC anatomical location and patients' age in TCGA-COAD cohort. *P* value in (A) was determined by two-sided paired Wilcoxon signed-rank test, *P* values in (B), (D) and (G-I) were calculated by Wilcoxon rank-sum tests, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$; ns, not significant.

Li R, *et al*. *Gut* 2024; 73:470–484. doi: 10.1136/gutjnl-2023-330243

supplemental figure 4

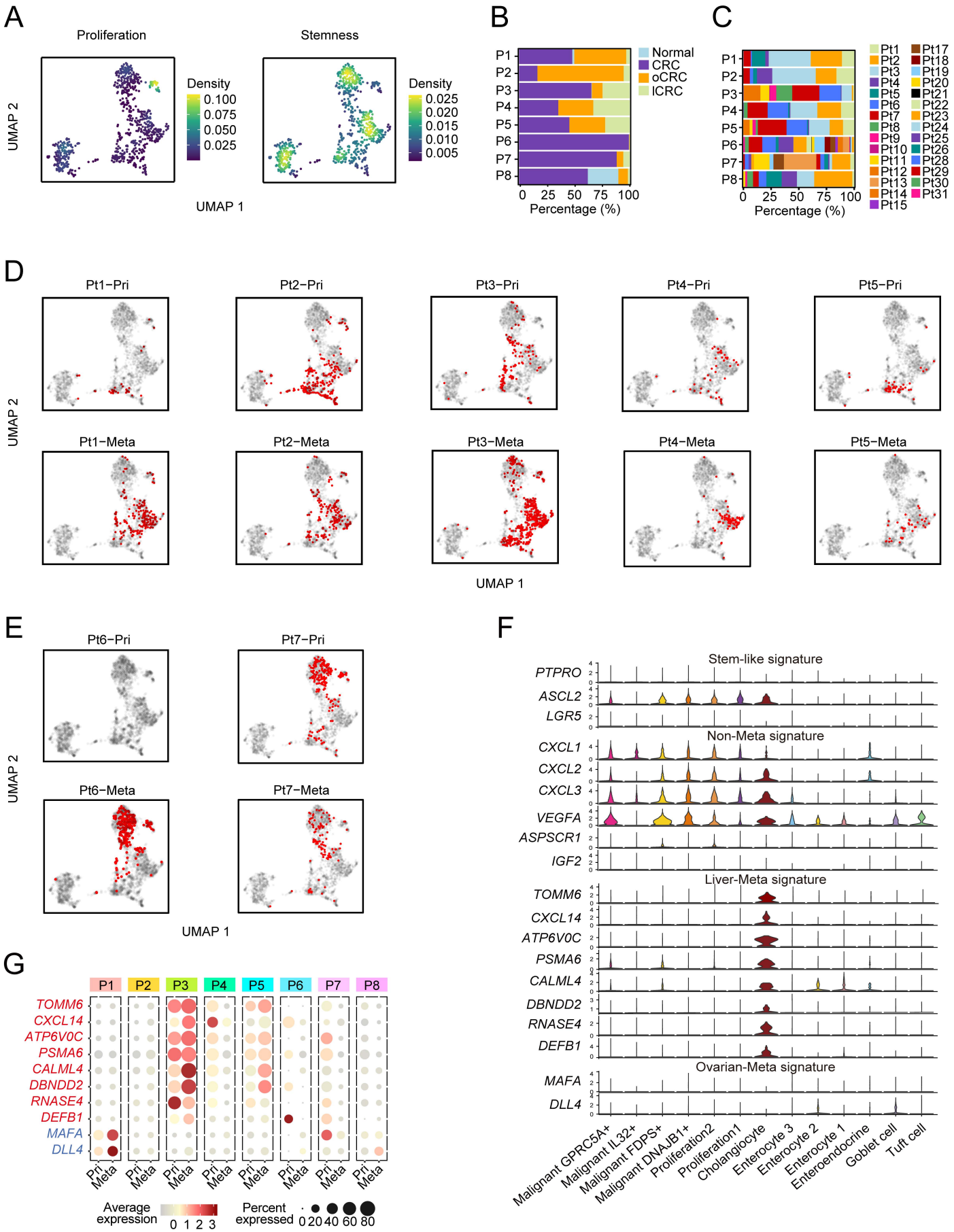


Supplemental figure 4 Effects of *PTPRO* depletion on CRC cell metastasis and stemness. Related to Figure 3.

(A) Spearman's correlation between the expression levels of *PTPRO* and *ASCL2* in TCGA-COAD dataset. (B) Effects of *ASCL2* depletion on the mRNA level of each other gene. (C) Effect of *PTPRO* knockdown on the mRNA levels of *PTPRO* or *ASCL2*. (D) Multivariable Cox regression analysis shows low *PTPRO* level is a protective factor for death of MSS CRC patients in TCGA-COAD cohort. Bars represent 95% confidence intervals of hazard rations. (E) Representative transwell assay pictures of *PTPRO* depletion on CRC cell migration and invasion. (F) Representative pictures of *PTPRO* depletion on CRC cell sphere-propagating capacity. Data in (B) and (C) are mean ± SD from 3 independent experiments, ***$P < 0.001$ and ****$P < 0.0001$ were determined by one-way ANOVA test with Dunnett's T3 multiple-comparison. ns, not significant.

Li R, *et al*. *Gut* 2024; 73:470–484. doi: 10.1136/gutjnl-2023-330243

supplemental figure 5

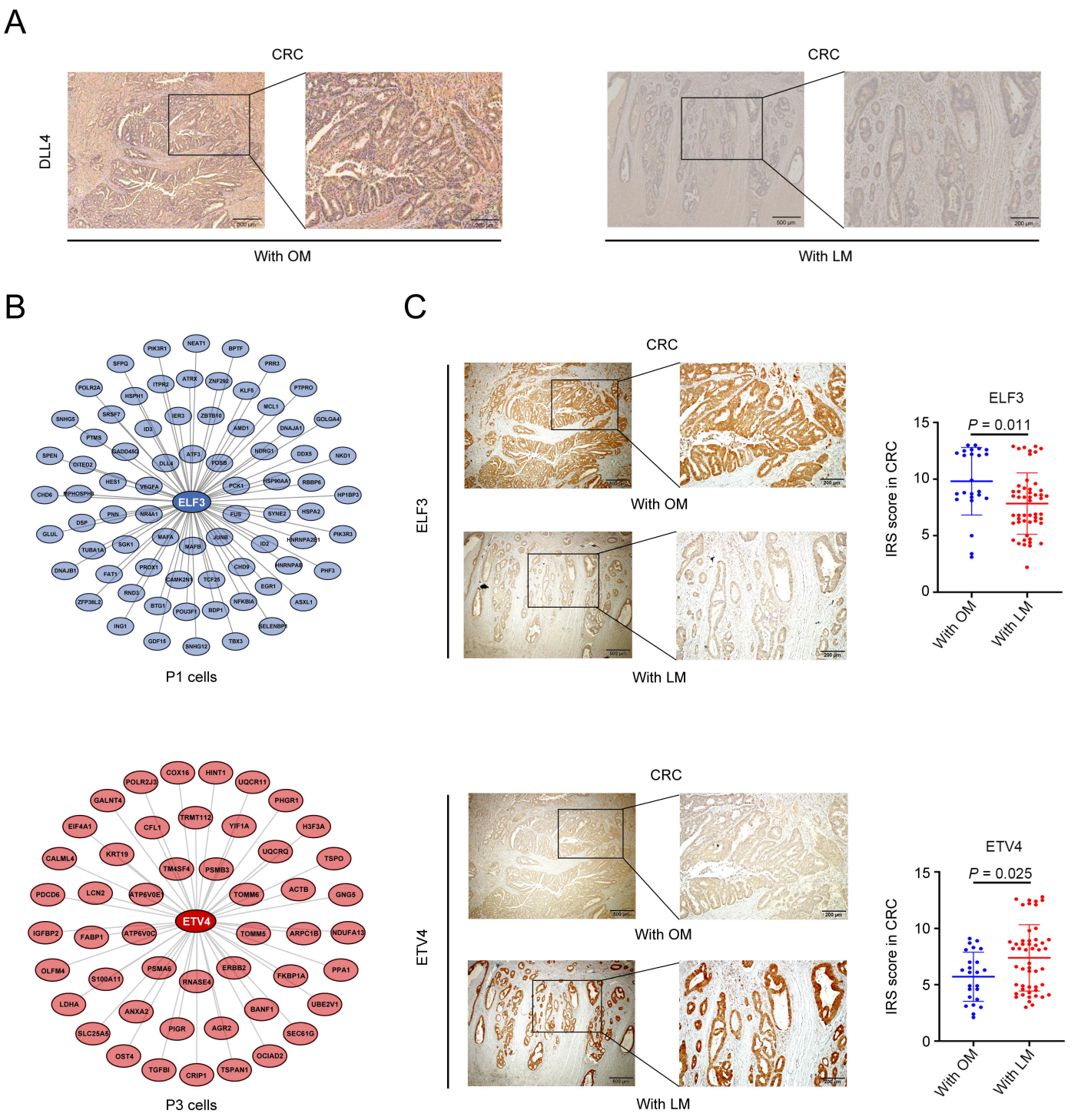


Supplemental figure 5 Transcription patterns and master TFs of stem-like cell subtypes. Related to Figure 4.

(A) UMAP plot showing the stemness and proliferation scores of stem-like cell subtypes. (B) Proportions of stem-like cell subtypes across different tissue types. (C) Proportions of stem-like cell subtypes across each patient. (D) UMAP plot showing the cells derived from primary CRC and ovarian metastases of patients 1 to 5 (Pt1−Pt5). (E) UMAP plot showing the cells derived from primary CRC and liver metastases of patients 6 and 7 (Pt6−Pt7). (F) Violin plots showing the signature gene expression in other epithelial cell subtypes. (G) Signature gene expression levels across stem-like cell subtypes in primary CRC and metastatic CRC. Dot size indicates the fraction of expressing cells and color represents normalized expression levels.

Li R, *et al*. *Gut* 2024; 73:470–484. doi: 10.1136/gutjnl-2023-330243

supplemental figure 6

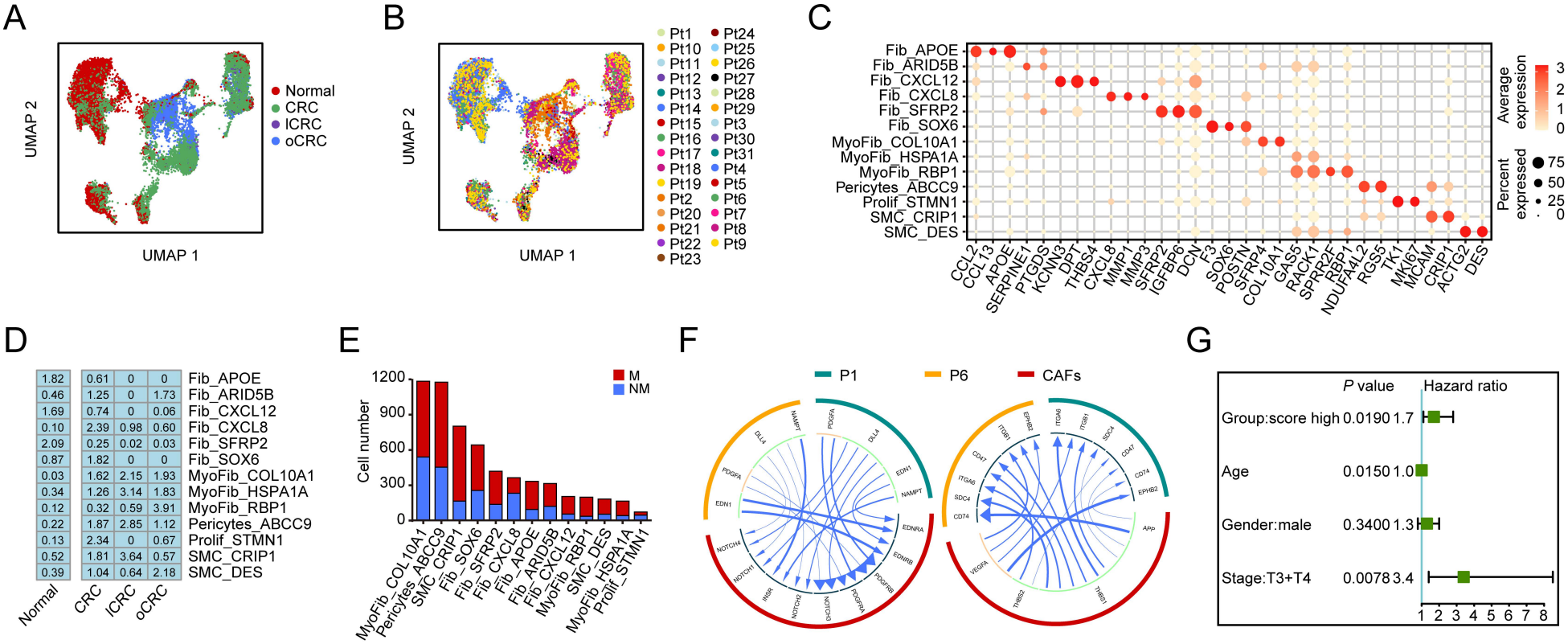


Supplemental figure 6 DLL4 and master TFs of stem-like cell subtypes. Related to Figure 4.

(A) Representative IHC images of DLL4 in primary CRC from patients with liver metastasis (LM) or ovarian metastasis (OM). Scale bars, 500 and 200 μm. (B) Regulatory network showing master TFs and their target genes in P1 cells and P3 cells. (C) Representative IHC images of ELF3 and ETV4 in primary CRC from patients with liver (LM) or ovarian metastasis (OM). Scale bars, 500 and 200 μm. Comparison of the ELF3 and ETV4 levels in primary CRC from patients with liver (LM) or ovarian metastases (OM). *P* values were calculated by two-sided Wilcoxon rank-sum test.

Li R, *et al*. *Gut* 2024; 73:470–484. doi: 10.1136/gutjnl-2023-330243

supplemental figure 7

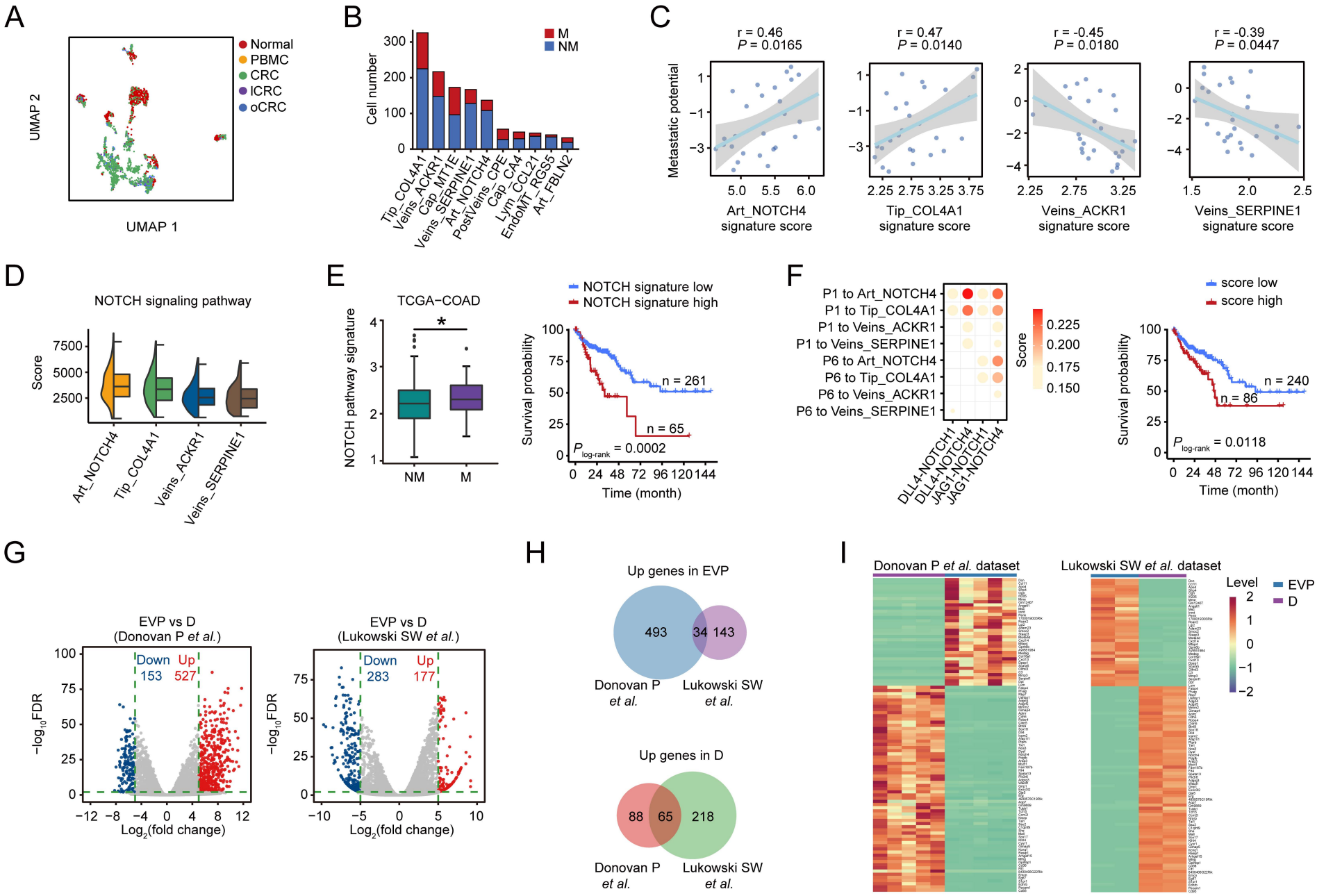


Supplemental figure 7 Fibroblast subtypes in primary CRC and their metastases. Related to Figure 6.

(A-B) UMAP plots of fibroblasts from different tissues (A) and patients (B). (C) Marker genes across 13 fibroblast subtypes. Dot size indicates the fraction of expressing cells and color represents normalized expression levels. (D) Tissue prevalence of fibroblasts estimated by Ro/e score. (E) The number of different fibroblast cell subtypes in primary CRC from patients with or without distant metastases. (F) Circos plots of CAFs, P1 cells and P6 cells, showing the ligand-receptor interactions. (G) Multivariate Cox regression analysis of TCGA-COAD data shows risk factors for death of MSS CRC patients (related to figure 6C). Bars represent 95% confidence intervals of hazard ratios.

Li R, *et al*. *Gut* 2024; 73:470–484. doi: 10.1136/gutjnl-2023-330243

supplemental figure 8

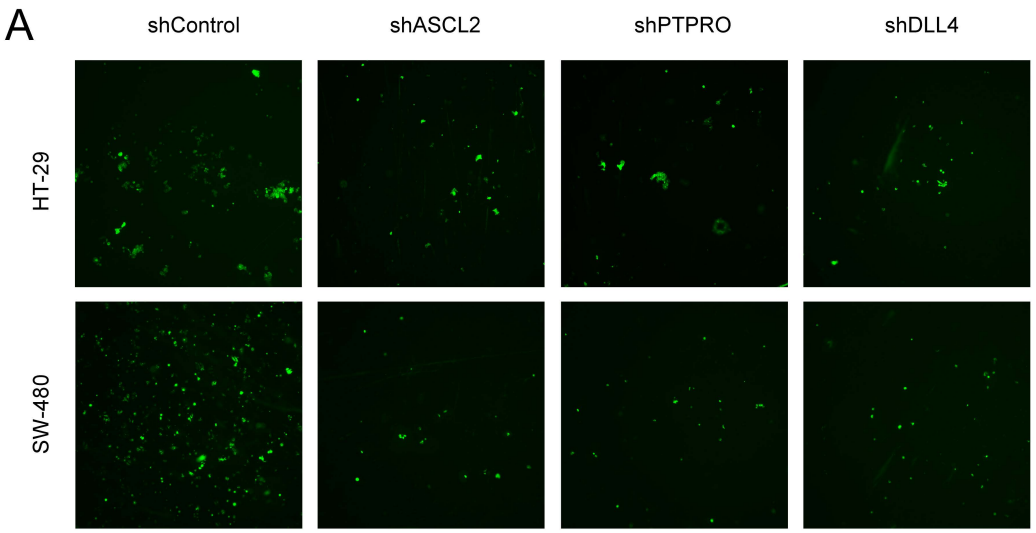


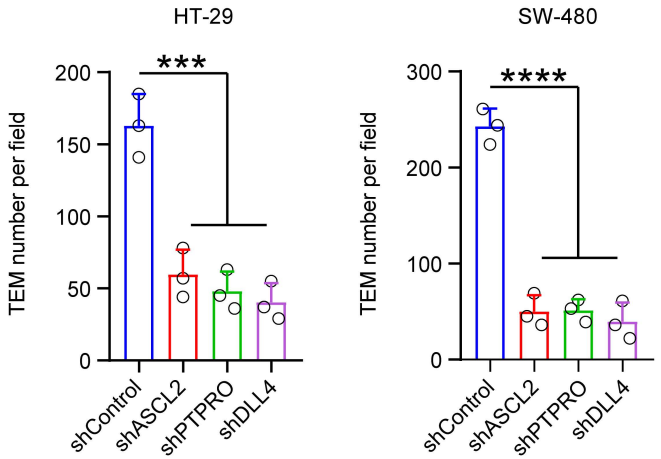
Supplemental figure 8 Endothelial cell subtypes in primary CRC and their metastases. Related to Figure 6.

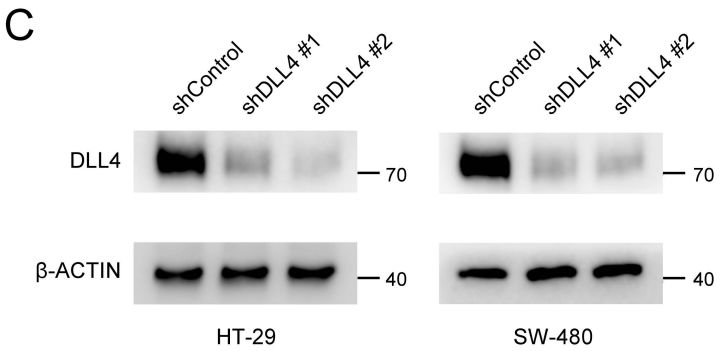
(A) UMAP plot of endothelial cells from different tissues. (B) The number of endothelial cell subtypes in primary CRC from patients with or without metastases. (C) Spearman's correlations between the signature scores of four endothelial cell subtypes and metastatic potential score in the CCLE database. (D) The levels of NOTCH signaling pathway in four endothelial cell subtypes. (E) Comparison of NOTCH pathway signature scores in patients with or without metastases (left panel), *P* values were determined by Wilcoxon rank-sum tests, $*P < 0.05$. Kaplan–Meier estimation of overall survival time of MSS CRC patients by the NOTCH pathway signature score (right panel, TCGA-COAD cohort). (F) The ligand-receptor pairs involved in the NOTCH pathway between P1/P6 cells and four endothelial cell subtypes in primary CRC (left panel) and Kaplan–Meier estimation of survival time of MSS CRC patients by the score of P1-endothelial cell interaction network (right panel, TCGA-COAD cohort). (G) Differentially expressed genes between EVP cells and D cells. (H) The overlap of upregulated genes in two independent bulk RNA-seq datasets shown in (G). Upper panel shows upregulated genes in EVP cells and lower panel shows upregulated genes in D cells. (I) Heatmaps showing the expression levels of overlap genes shown in (H) in EVP cells and D cells.

Li R, *et al*. *Gut* 2024; 73:470–484. doi: 10.1136/gutjnl-2023-330243

supplemental figure 9





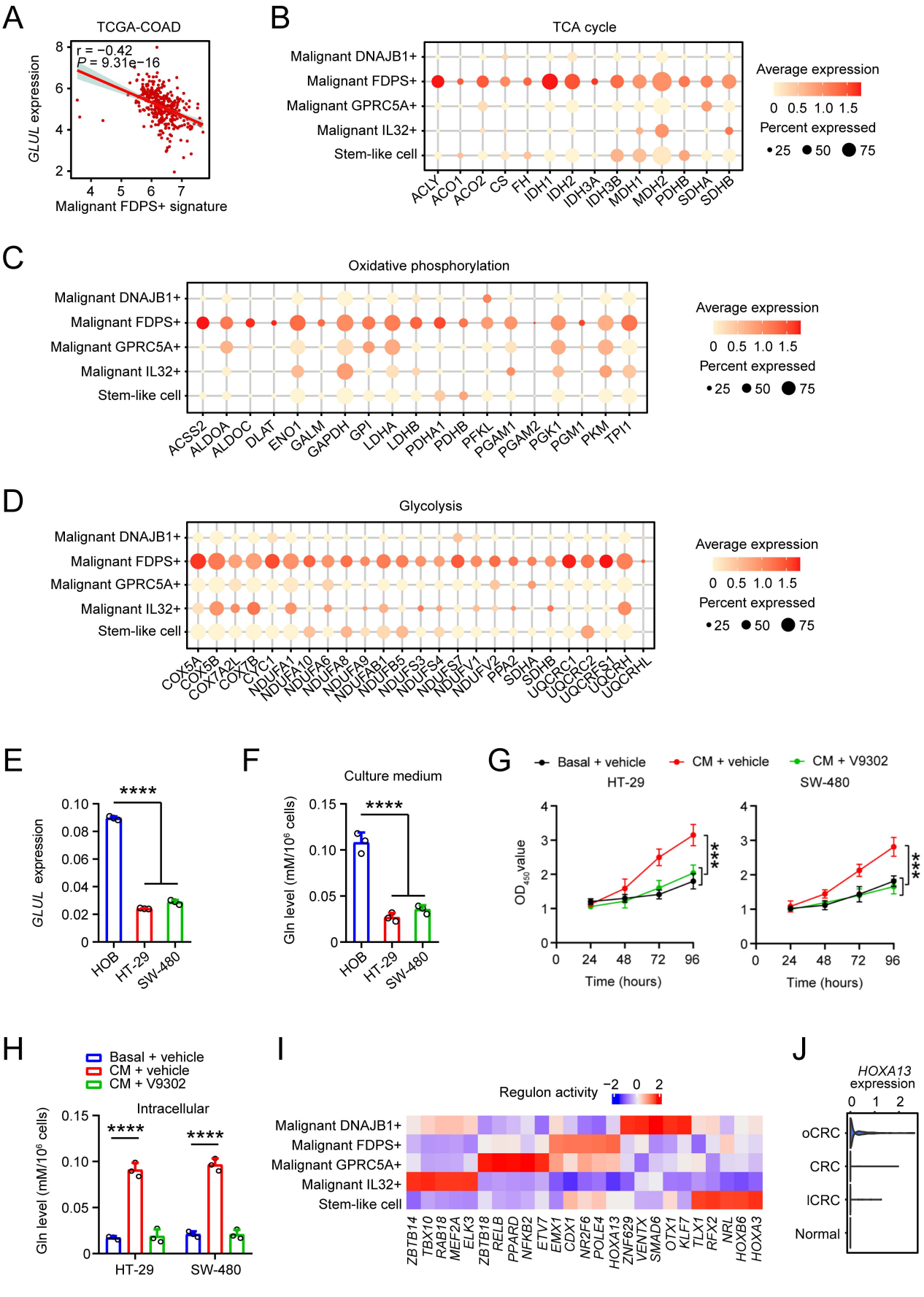


Supplemental figure 9 *DLL4*, *ASCL2* or *PTPRO* knockdown in CRC cells suppresses tumor cells migrated through endothelial cell. Related to Figure 6

(A, B) Representative images (A) and quantitative analysis (B) of DiO labeled *DLL4*, *ASCL2* or *PTPRO* knockdown CRC cells transmigrated to the lower face of the filter. CRC cells were labeled with green fluorescence. TEM, transendothelial migration. Data are mean ± SD from 3 independent experiments, ***$P < 0.001$ and ****$P < 0.0001$ for one-way ANOVA test with Dunnett's T3 multiple-comparison. (C) The effect of *DLL4* knockdown on DLL4 protein level in CRC cells.

Li R, *et al*. *Gut* 2024; 73:470–484. doi: 10.1136/gutjnl-2023-330243

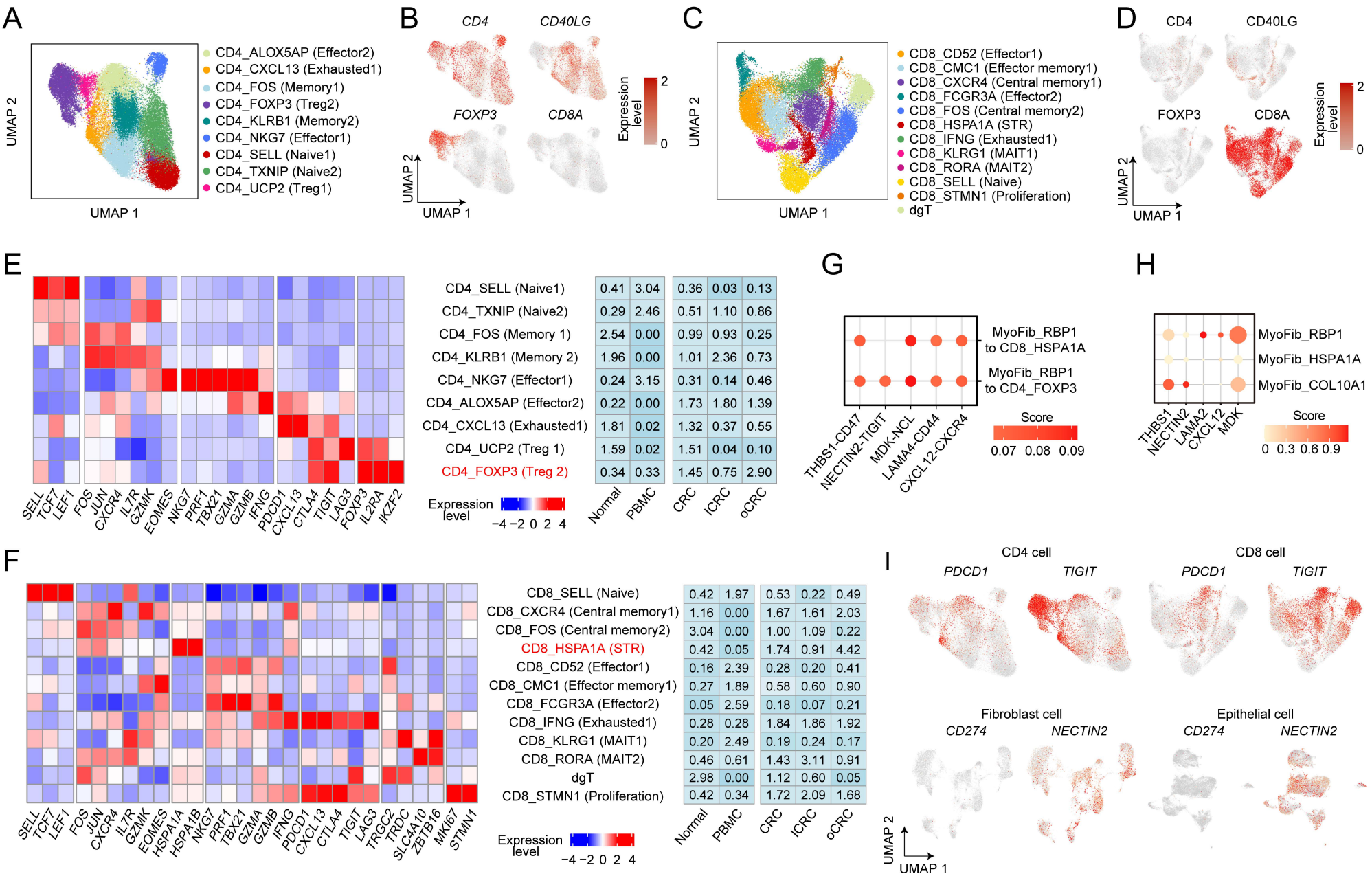
supplemental figure 10



Supplemental figure 10 Characterization of the metabolic features in oCRC. Related to Figure 7. (A) Spearman's correlation between *GLUL* mRNA levels and the signature scores of malignant FDPS^{+} cells in TCGA-COAD dataset. (B-D) Bubble heatmap showing the expression levels of genes in TCA cycle (B), oxidative phosphorylation (C), and glycolysis (D) pathways among 5 malignant cell subtypes. Dot size indicates the fraction of expressing cells and the color indicates the normalized expression levels. (E) Comparison of the *GLUL* RNA levels in human ovarian fibroblast (HOB) and two CRC cell lines. (F) Gln levels in culture medium before conditional medium collection. (G) Effect of V9302 and conditional medium on proliferation of CRC cells. (H) Intracellular gln levels of CRC cells after V9302 and conditional medium treatment. (I) Heatmap showing the activity of regulatory TFs across 5 malignant cell subtypes. (J) The expression levels of *HOXA13* in FDPS^{+} malignant cells from different tissues, *P* values were determined by Wilcoxon rank-sum tests, ****$P < 0.0001$. Data in (E), (F), (G) and (H) are mean ± SD from 3 independent experiments, ***$P < 0.001$ and ****$P < 0.0001$ in (E), (F) and (G) for one-way ANOVA test with Dunnett's T3 multiple-comparison, in (H) for two-way ANOVA test with Sidak's multiple comparisons test.

Li R, *et al*. *Gut* 2024; 73:470–484. doi: 10.1136/gutjnl-2023-330243

supplemental figure 11



Supplemental figure 11 CD4 and CD8 cell subtypes in primary CRC and their metastases.

(A) UMAP plot showing 9 CD4 cell subtypes. (B) UMAP plot showing the expression levels of *CD4*, *CD40LG*, *FOXP3* and *CD8A* in CD4 cells. (C) UMAP plot showing 12 CD8 cell subtypes. (D) UMAP plot showing the expression levels of *CD4*, *CD40LG*, *FOXP3* and *CD8A* in CD8 cells. (E) Heatmap showing the expression of representative genes across CD4 cell subtypes. Tissue distribution of different CD4 cells estimated by Ro/e score. (F) Heatmap showing the expression of representative genes across CD8 cell subtypes. Tissue distribution of different CD8 cells estimated by Ro/e score. (G) The ligand-receptor pairs between MyoFib_RBP1 cells and two oCRC enriched T cells. (H) Bubble heatmap showing expression levels of selected ligands in 3 myofibroblast cell subtypes. Dot size indicates the fraction of expressing cells and the color indicates the normalized expression level. (I) UMAP plot showing the expression levels of *PDCD1* and *TIGIT* in CD4 or CD8 cells, *CD274* and *NECTIN2* in fibroblast and epithelial cells.

Li R, *et al*. *Gut* 2024; 73:470–484. doi: 10.1136/gutjnl-2023-330243

Online supplemental table 1. Baseline and clinical information of CRC patients in the present study.

Patient ID	Gender	Age	Anatomic region	Pathological subtype	Differentiation	MSI[a] status	Tumor stage	Metastasis	Data source
Pt1	F	48	Transverse	Adenocarcinoma	Moderate	MSS[b]	IV	Ovary-M[c]	In-house
Pt2	F	52	Ascending	Mix/Mucinous	High	MSS	IV	Ovary-M	In-house
Pt3	F	54	Splenic flexure	Adenocarcinoma	Moderate	MSS	IV	Ovary-M	In-house
Pt4	F	56	Ileocecal junction	Mucinous	Undefined	MSS	IV	Ovary-M	In-house
Pt5	F	65	Descending	Adenocarcinoma	Moderate	MSS	IV	Ovary-M	In-house
Pt6	M	41	Rectal	Adenocarcinoma	Low-Moderate	MSS	IV	Liver-M	In-house
Pt7	F	34	Sigmoid	Adenocarcinoma	Moderate-High	MSS	IV	Liver-M	In-house
Pt8	F	50	Sigmoid	Adenocarcinoma	Moderate	MSS	IV	Unknown-M	KUL3
Pt9	F	86	Rectosigmoid	Adenocarcinoma	Moderate	MSS	III	NM[d]	KUL3
Pt10	M	52	Sigmoid	Adenocarcinoma	Moderate	MSS	II	NM	KUL3
Pt11	M	84	Ascending	Adenocarcinoma	Moderate	MSS	II	NM	KUL3
Pt12	M	85	Sigmoid	Adenocarcinoma	High	MSS	I	NM	KUL3
Pt13	F	64	Rectal	Adenocarcinoma	High	MSS	II	NM	SMC
Pt14	M	66	Rectal	Adenocarcinoma	High	MSS	III	NM	SMC
Pt15	M	69	Sigmoid	Adenocarcinoma	Moderate	MSS	III	NM	SMC
Pt16	F	58	Ascending	Adenocarcinoma	High	MSS	II	NM	SMC
Pt17	F	67	Ascending	Adenocarcinoma	High	MSS	I	NM	SMC
Pt18	M	68	Sigmoid	Adenocarcinoma	Moderate	MSS	III	NM	SMC
Pt19	M	75	Sigmoid	Adenocarcinoma	High	MSS	II	NM	SMC
Pt20	F	38	Sigmoid	Adenocarcinoma	High	MSS	III	NM	SMC

Pt21	M	77	Rectosigmoid	Adenocarcinoma	Moderate	MSS	III	NM	SMC
Pt22	M	56	Sigmoid	Adenocarcinoma	High	MSS	II	NM	SMC
Pt23	M	59	Ascending	Adenocarcinoma	Moderate	MSS	III	NM	SMC
Pt24	M	47	Hepatic flexure	Mucinous	Undefined	MSS	III	NM	SMC
Pt25	F	63	Sigmoid	Adenocarcinoma	Moderate	MSS	II	NM	SMC
Pt26	F	80	Ascending	Adenocarcinoma	Moderate	MSS	III	NM	SMC
Pt27	F	65	Ascending	Adenocarcinoma	Undefined	MSS	III	NM	SMC
Pt28	M	51	Rectal	Adenocarcinoma	Moderate	MSS	IV	Unknown-M	SMC
Pt29	M	76	Sigmoid	Adenocarcinoma	Moderate	MSS	III	NM	SMC
Pt30	F	67	Ascending	Adenocarcinoma	Moderate	MSS	III	NM	SMC
Pt31	F	57	Sigmoid	Adenocarcinoma	Moderate	MSS	IV	Unknown-M	SMC

[a]MSI, Microsatellite instability.

[b]MSS, Microsatellite stability.

[c]M, Metastasis.

[d]NM, Non-metastasis.

Online supplemental table 2. Baseline information on spatial cohort in the present study.

Sample ID	Patient ID	Tissue type	Data source
Pt1-CRC	Patient1 (Pt1)	CRC	In-house
Pt2-oCRC	Patient2 (Pt2)	oCRC	In-house
Pt7-CRC	Patient7 (Pt7)	CRC	In-house
Pt7-ICRC	Patient7 (Pt7)	lCRC	In-house