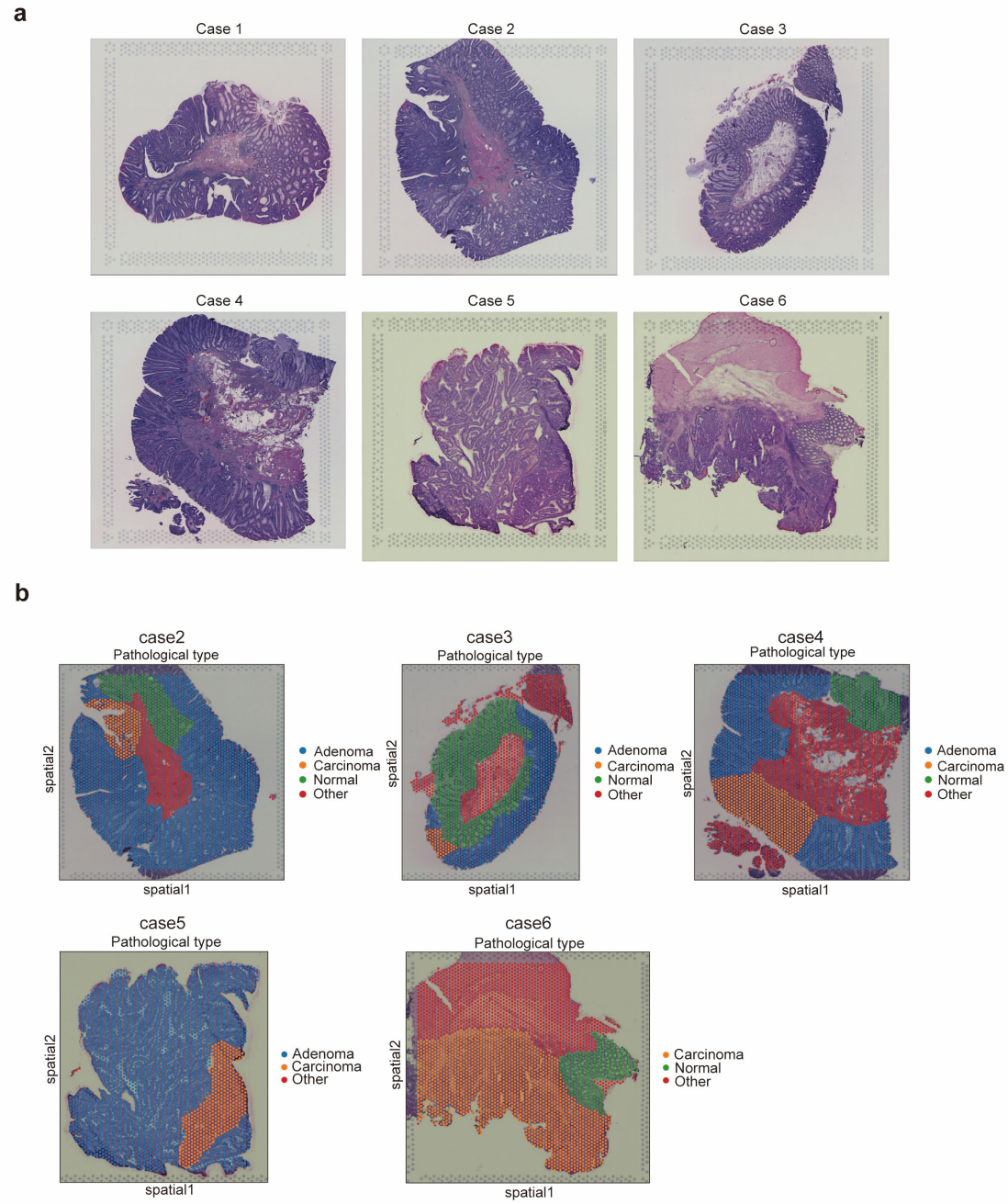
Supplementary Figures

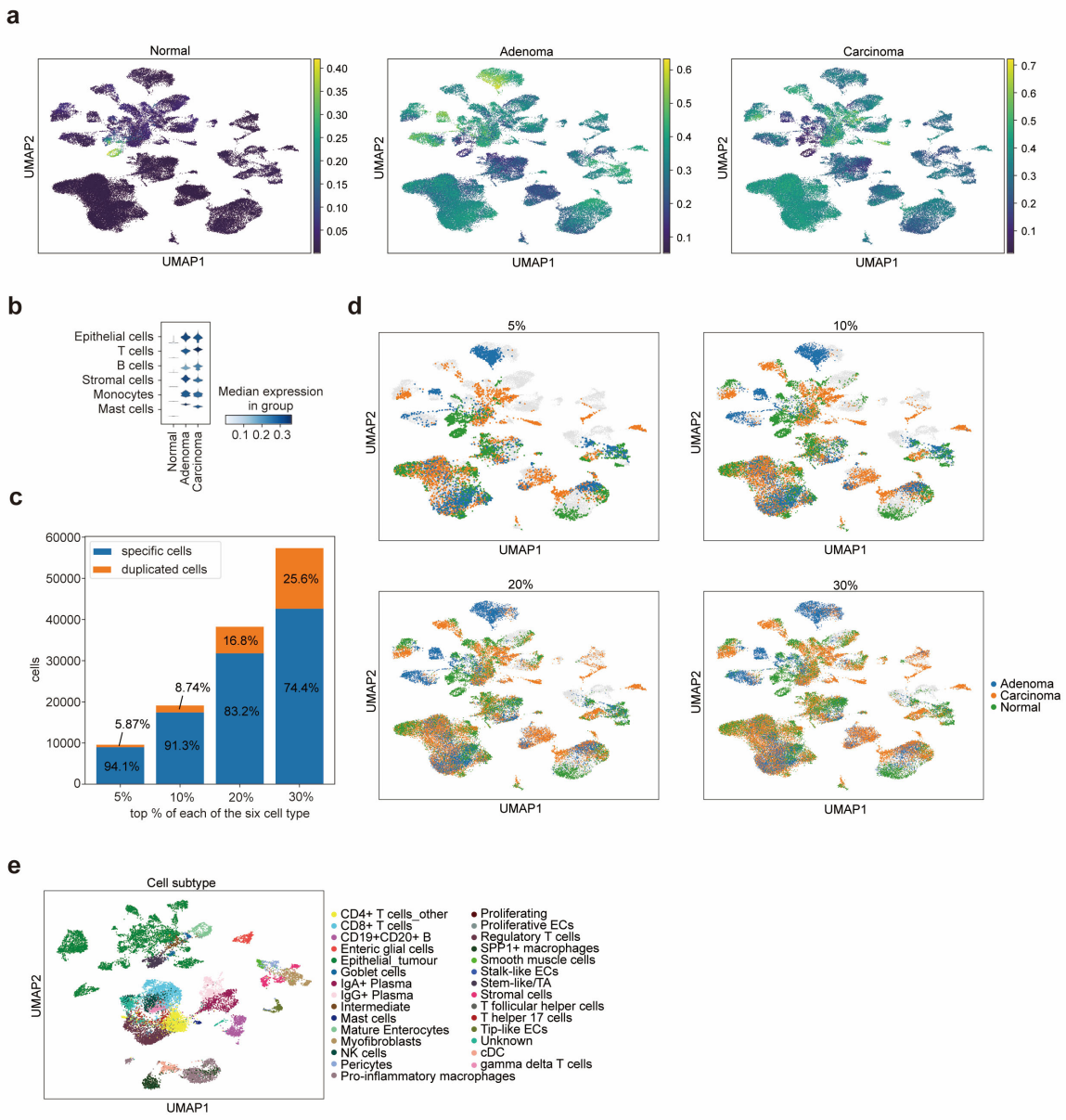
Supplementary Figure 1



Supplementary Figure 1. Slides and pathological diagnosis of all cases.

(a) Hematoxylin and eosin-stained images of all cases. (b) Pathological diagnosis distribution of cases 2-6. Cases 2-5 were classified as “Normal,” “Adenoma,” and “Carcinoma” by mucosa layer, while the others were classified as “Other.” Case 6, which was an advanced colorectal cancer, was classified via normal and cancer regions.

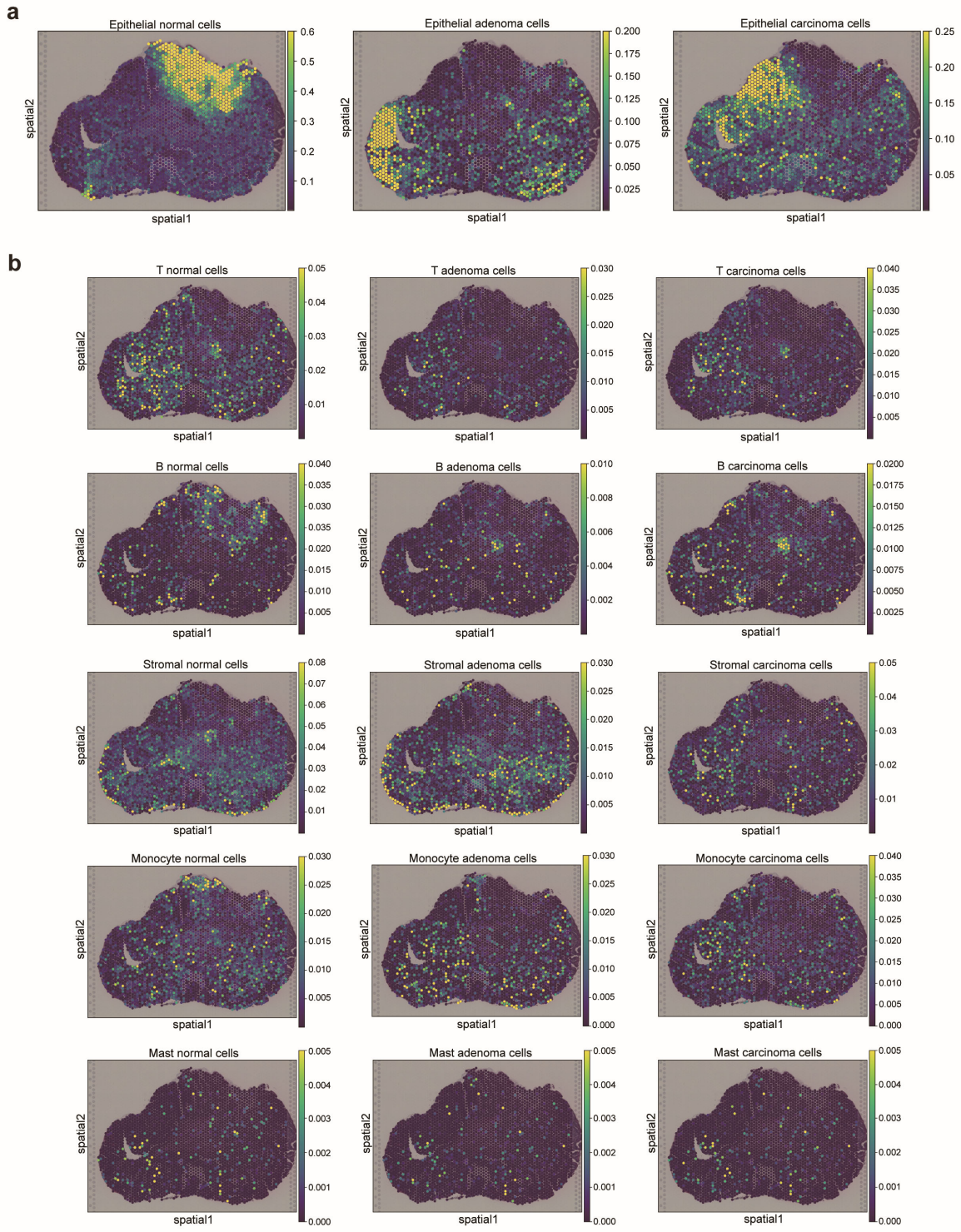
Supplementary Figure 2



Supplementary Figure 2. Spatial cell distribution after processing with integrative analysis.

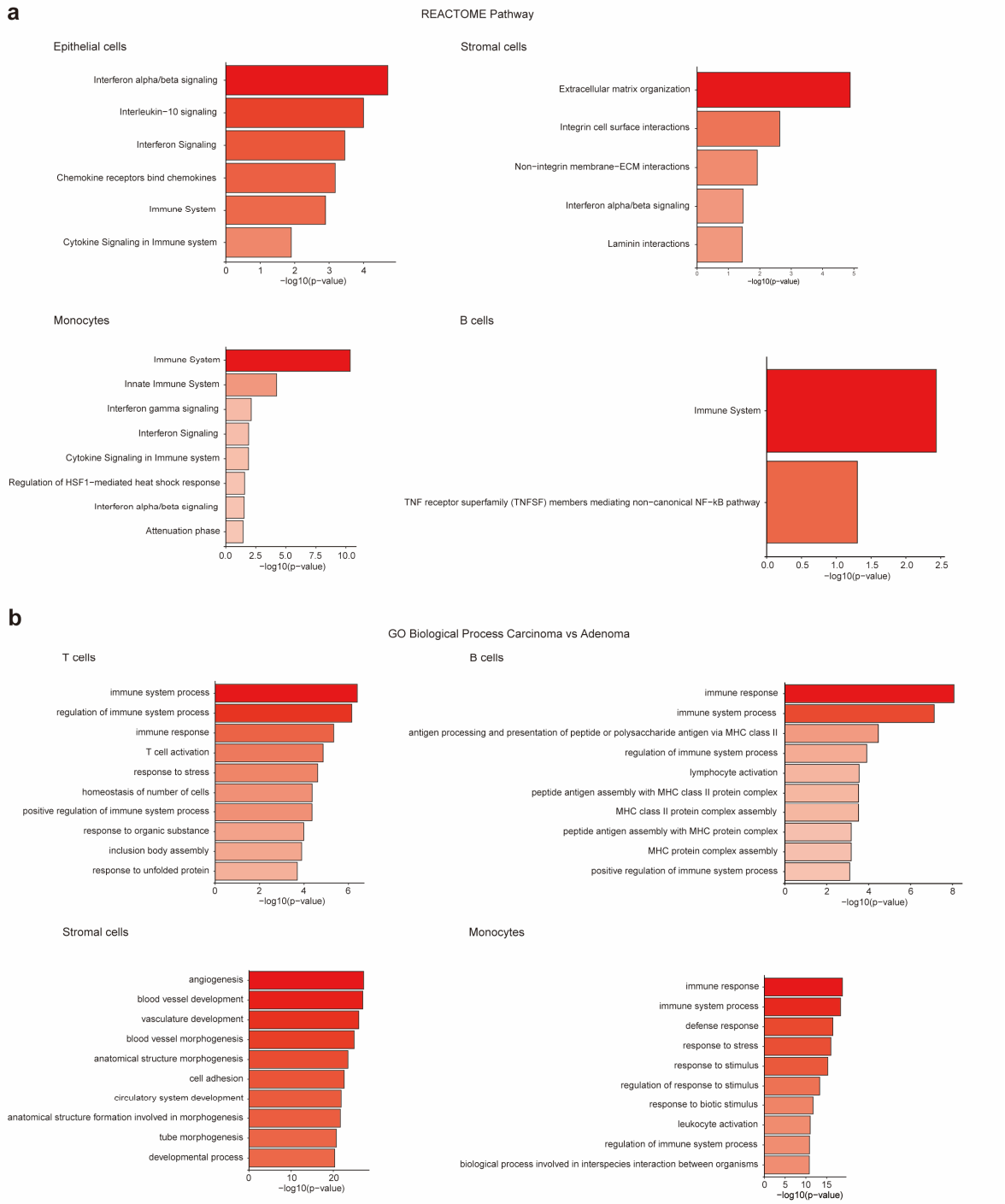
(a) Abundance at the single cell level in spatial distribution by each spatial pathological diagnosis. (b) Stacked violin plots of spatial assignment to each pathological diagnosis for the six cell types. (c) Bar plot presenting the number of cells with top-5, -10, -20, -30% defined as distinctive tissue origin cells and the fraction of specific cells (blue) and duplicated cells (orange). (d) Uniform manifold approximation and projection (UMAP) of the specific cells after definition of cell origin filtering of the top-5, -10, -20, -30% in all cells. (e) UMAP of cell subtypes based on the definition of cell origins filtering of the top 10% for each pathological diagnosis.

Supplementary Figure 3



Supplementary Figure 3. Spatial distribution of cell types after definition of the specific tissue origins.
(a) Spatial distribution of epithelial cells based on the definition of specific tissue origins. (b) Spatial distribution of other five cell types (T, B, Stromal, Monocyte, Mast cells) based on the definition of specific tissue origins.

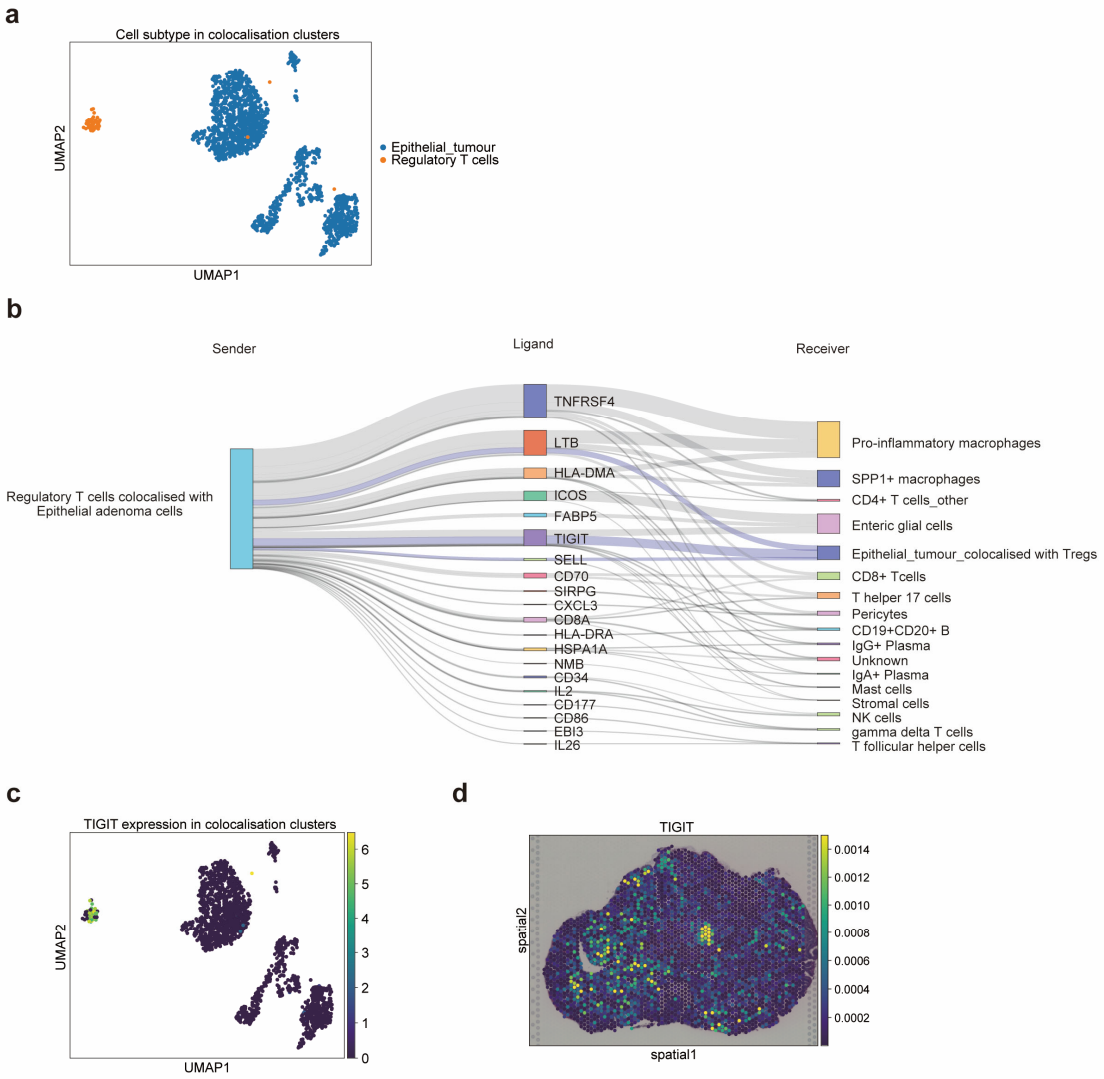
Supplementary Figure 4



Supplementary Figure 4. Enrichment analysis comparing carcinoma and adenoma.

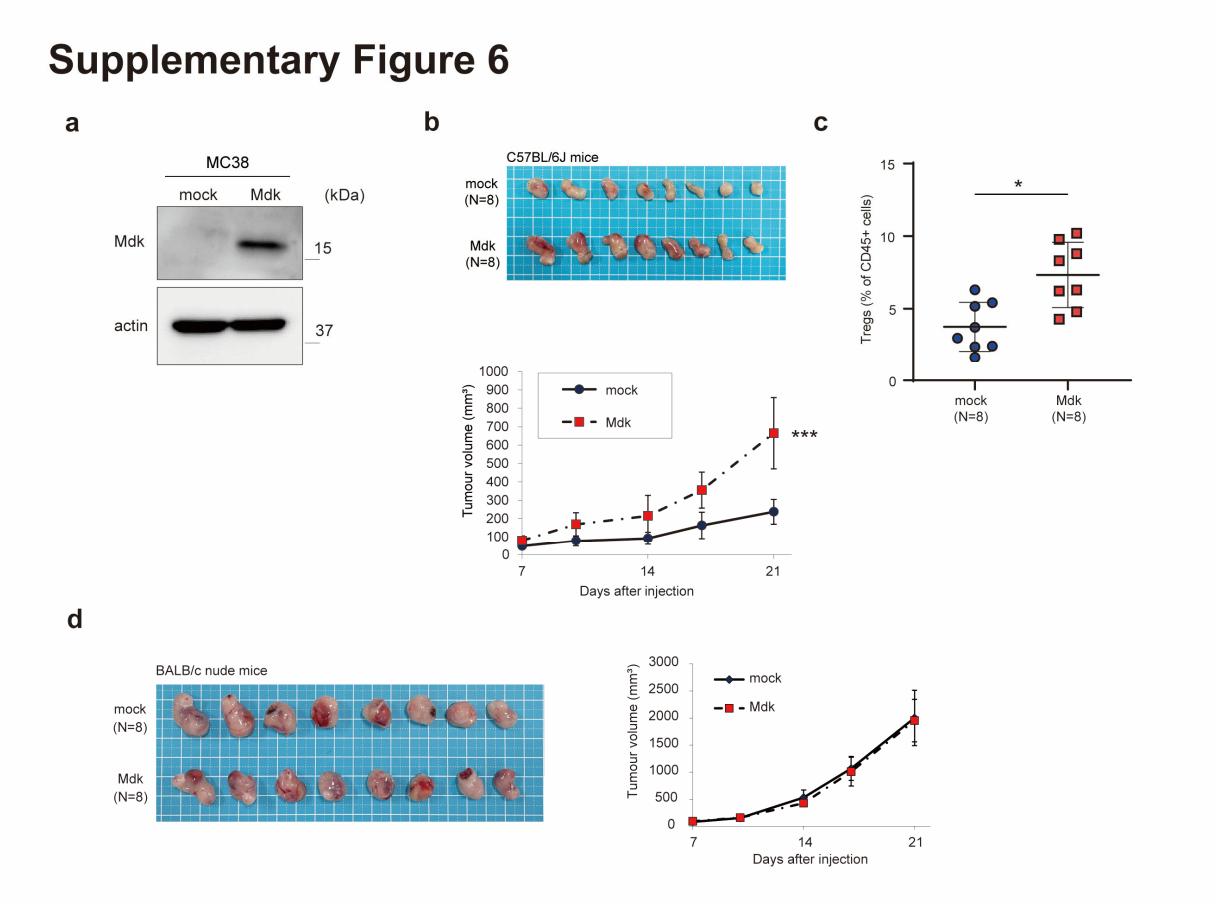
(a) Reactome pathway analysis in epithelial cells, B cells, stromal cells, and monocytes. (b) Gene ontology (GO) analysis of biological process in T cells, B cells, stromal cells, and monocytes.

Supplementary Figure 5



Supplementary Figure 5. Cell–cell interaction from regulatory T cells colocalised with epithelial adenoma cells to other cells.

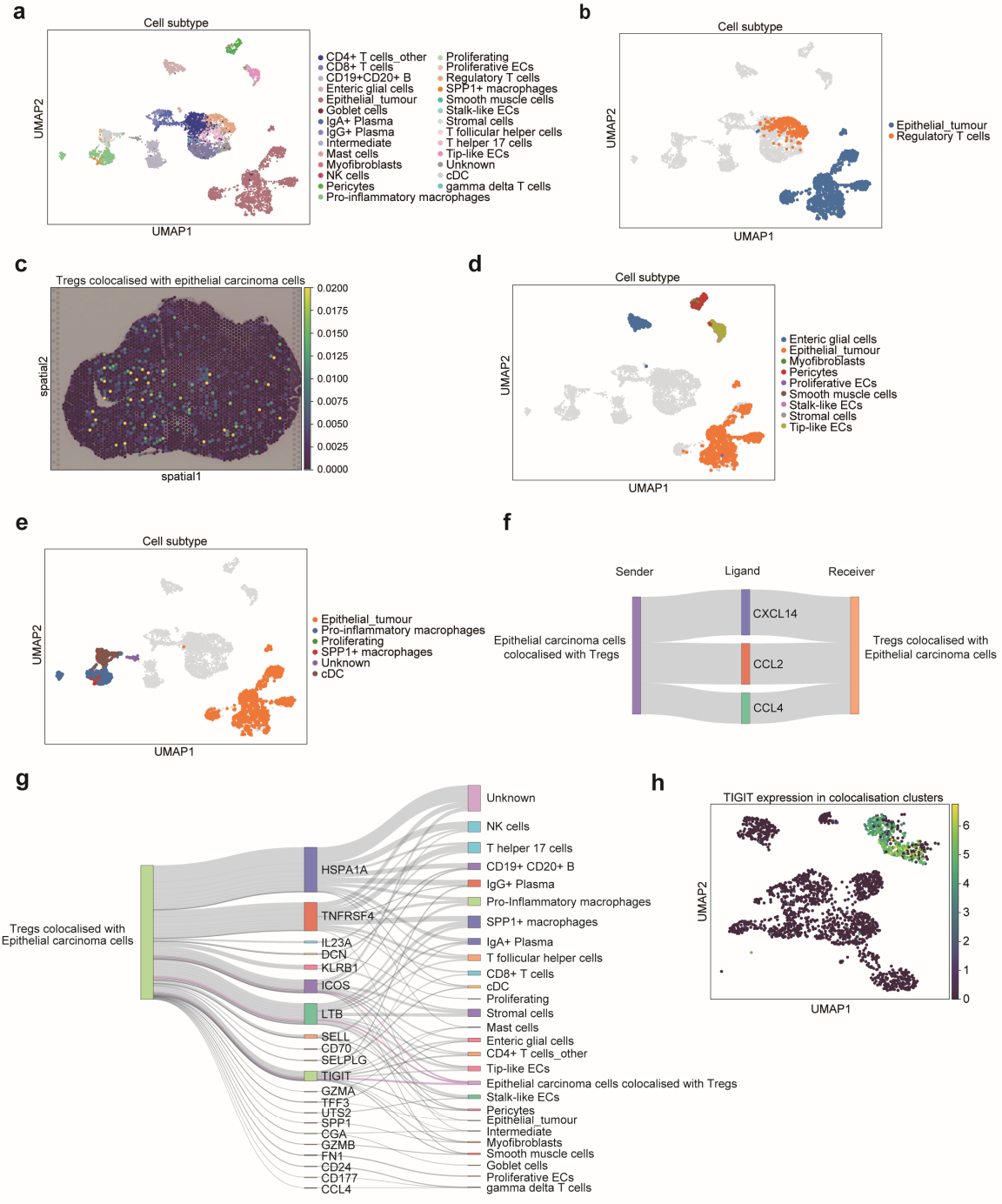
(a) Colocalisation clusters between adenoma epithelial cells and regulatory T cells (Tregs) in UMAP representation. (b) Comparison of ligand activity between colocalised single cells from epithelial regulatory T cells colocalised with epithelial adenoma cells and other cells. The widths of lines correspond to the ligand activity scores. (c) TIGIT expression in colocalised cell populations in uniform manifold approximation and projection (UMAP) representation. (d) Imputed TIGIT expression in spatial distribution.



Supplementary Figure 6. MDK is involved in proliferative potential via the induction of regulatory T cells in mice.

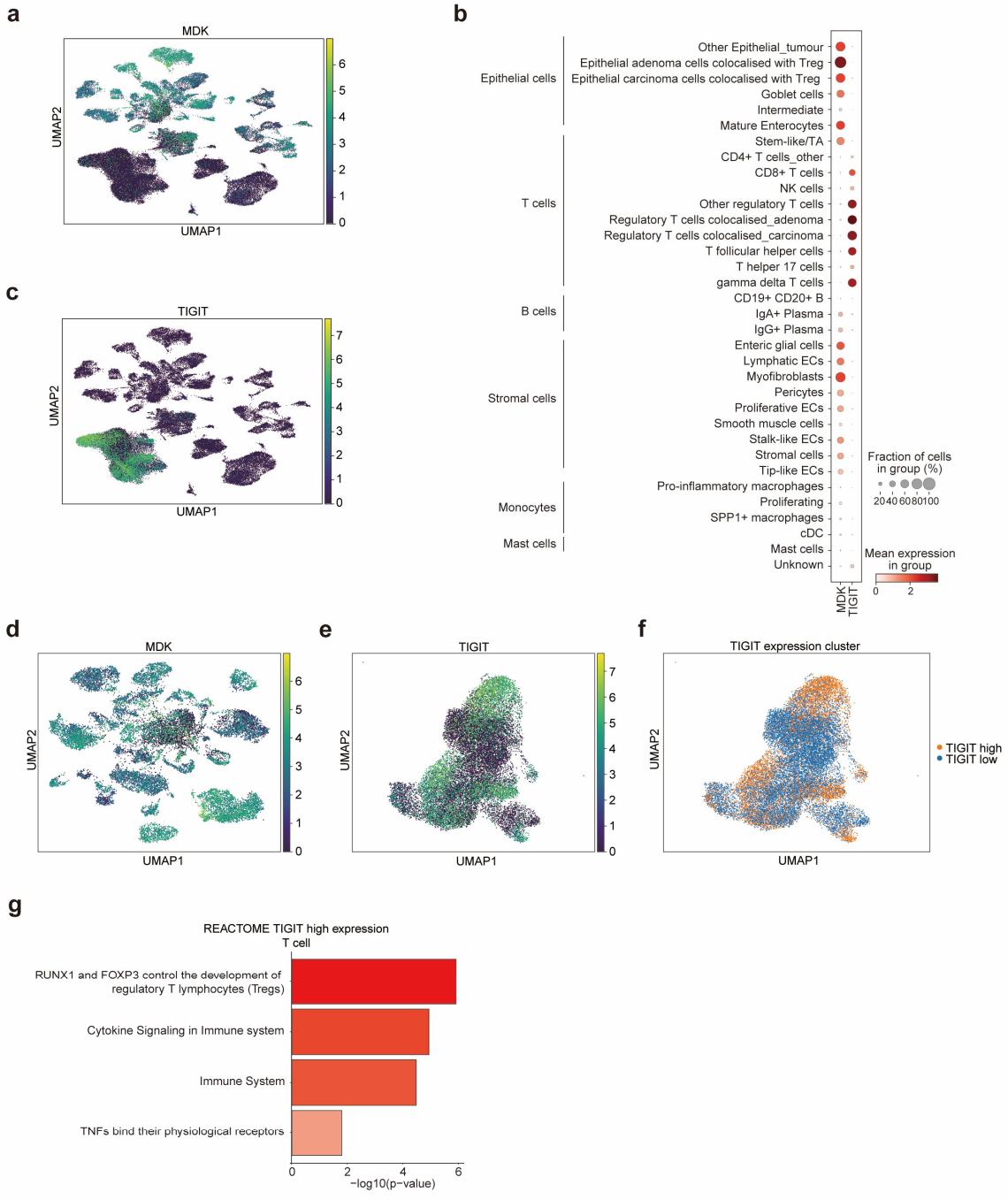
(a) Immunoblotting for Mdk in mock and Mdk-overexpressing cells (MC38). (b) Mock and Mdk-overexpressing MC-38 cells were injected subcutaneously into C57BL/6J mice. Photos of the tumours on day 21; Bar=1 cm; Growth curves of tumour volume on the indicated day. N=8 per group. Error bars the indicate standard error of the mean (SEM). ***, $P<0.001$. P values were determined using the Wilcoxon rank sum test. (c) Comparison of the percentage of Tregs (CD4+CD25+) in CD45+ cells in tumour tissues from Mock and Mdk-overexpressing MC-38 cells, as assessed using flow cytometry. (d) Mock and Mdk-overexpressing MC-38 cells were injected subcutaneously into nude mice.

Supplementary Figure 7



Supplementary Figure 7. Colocalisation analysis and cell–cell interaction in carcinoma cells.
(a) Uniform manifold approximation and projection (UMAP) of cell subtypes in carcinoma cluster. (b) Colocalisation clusters between carcinoma epithelial cells and regulatory T cells (Tregs) in UMAP representation across all carcinoma cells. (c) Spatial distribution of regulatory T cells (Tregs) colocalised with epithelial carcinoma cells. (d) Colocalisation clusters between carcinoma epithelial cells colocalised with Tregs and stromal cells in UMAP representation across all carcinoma cells. (e) Colocalisation clusters between carcinoma epithelial cells colocalised with Tregs and monocytes in UMAP representation across all carcinoma cells. (f) Comparison of ligand activity initiating from epithelial carcinoma cells colocalised with Tregs to Tregs colocalised with epithelial carcinoma cells. (g) Comparison of ligand activity initiating from Tregs to Tregs colocalised with epithelial carcinoma cells and other cells. The widths of lines correspond to the ligand activity scores. (h) TIGIT expression in colocalised cell populations in UMAP.

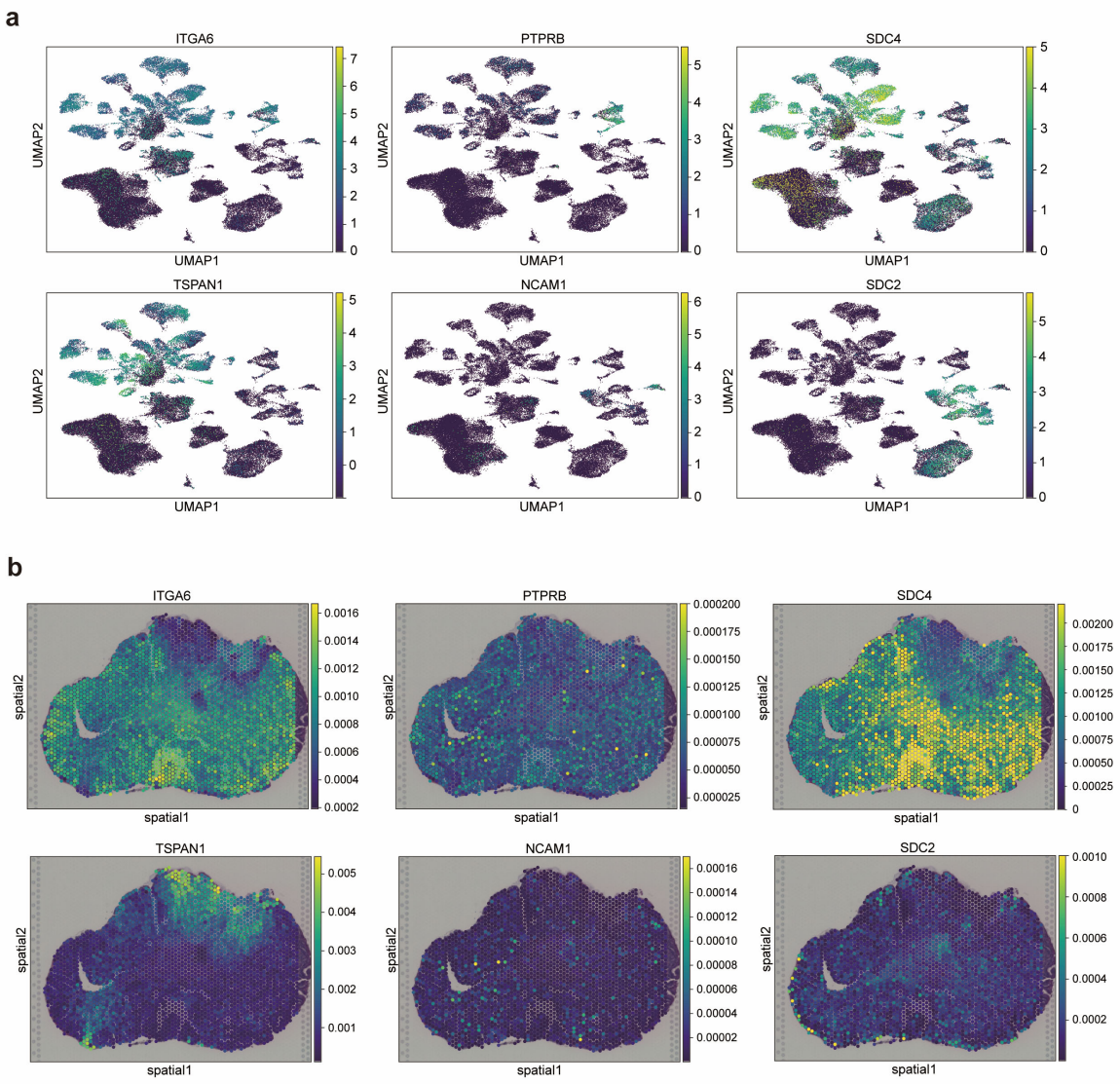
Supplementary Figure 8



Supplementary Figure 8. MDK and TIGIT expression in single-cell RNA sequencing (scRNA-seq) data and pathway analysis.

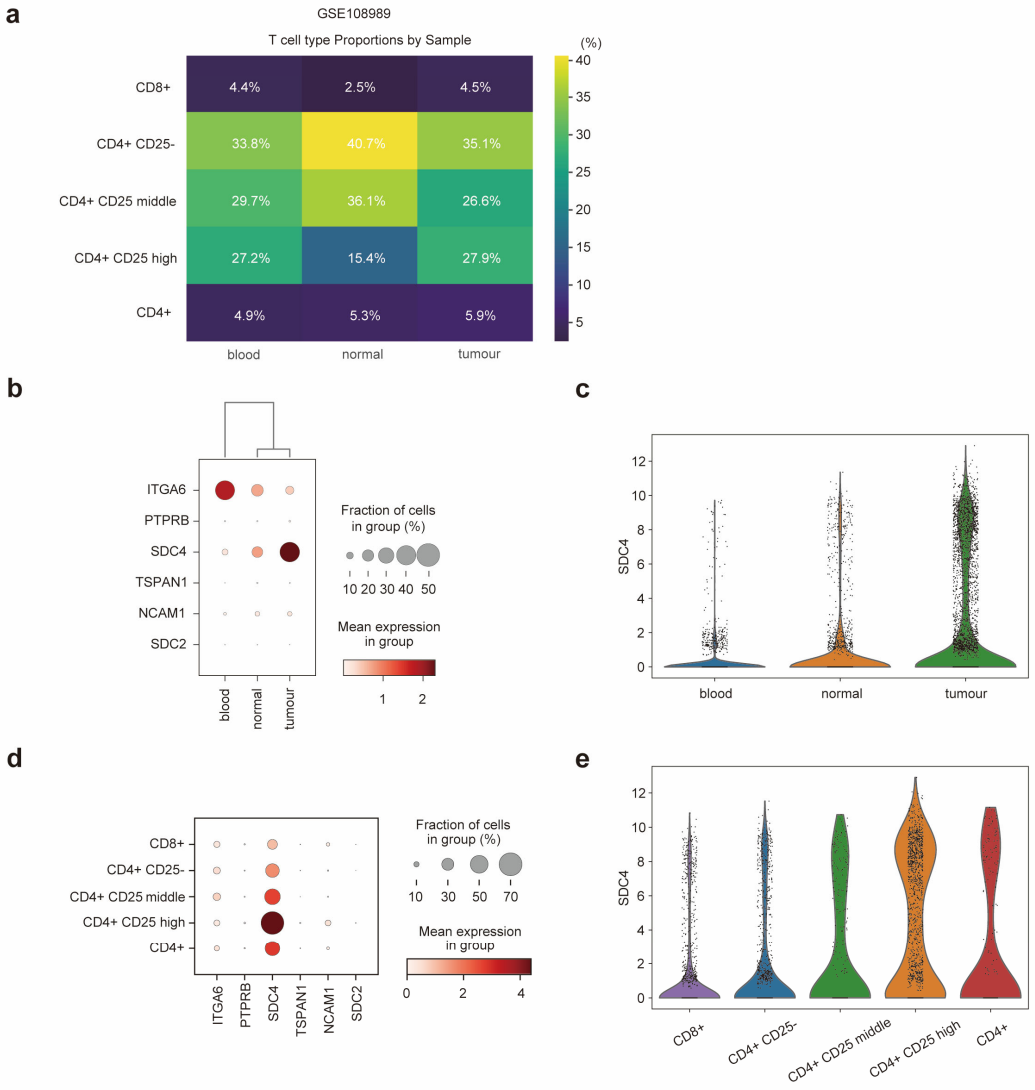
(a) Uniform manifold approximation and projection (UMAP) distribution of MDK expression in scRNA-seq data. (b) Dot plot of the expression and proportion of MDK and TIGIT per cell subtype; the circle size represents the cell proportion. (c) UMAP distribution of TIGIT expression in scRNA-seq data. (d) UMAP distribution of MDK expression in epithelial cells. (e) UMAP distribution of TIGIT expression in epithelial cells. (f) UMAP distribution in the comparison of T cells with high and low TIGIT expression divided by median TIGIT expression levels. (g) Reactome pathway analysis comparing T cells with high and low TIGIT expression.

Supplementary Figure 9



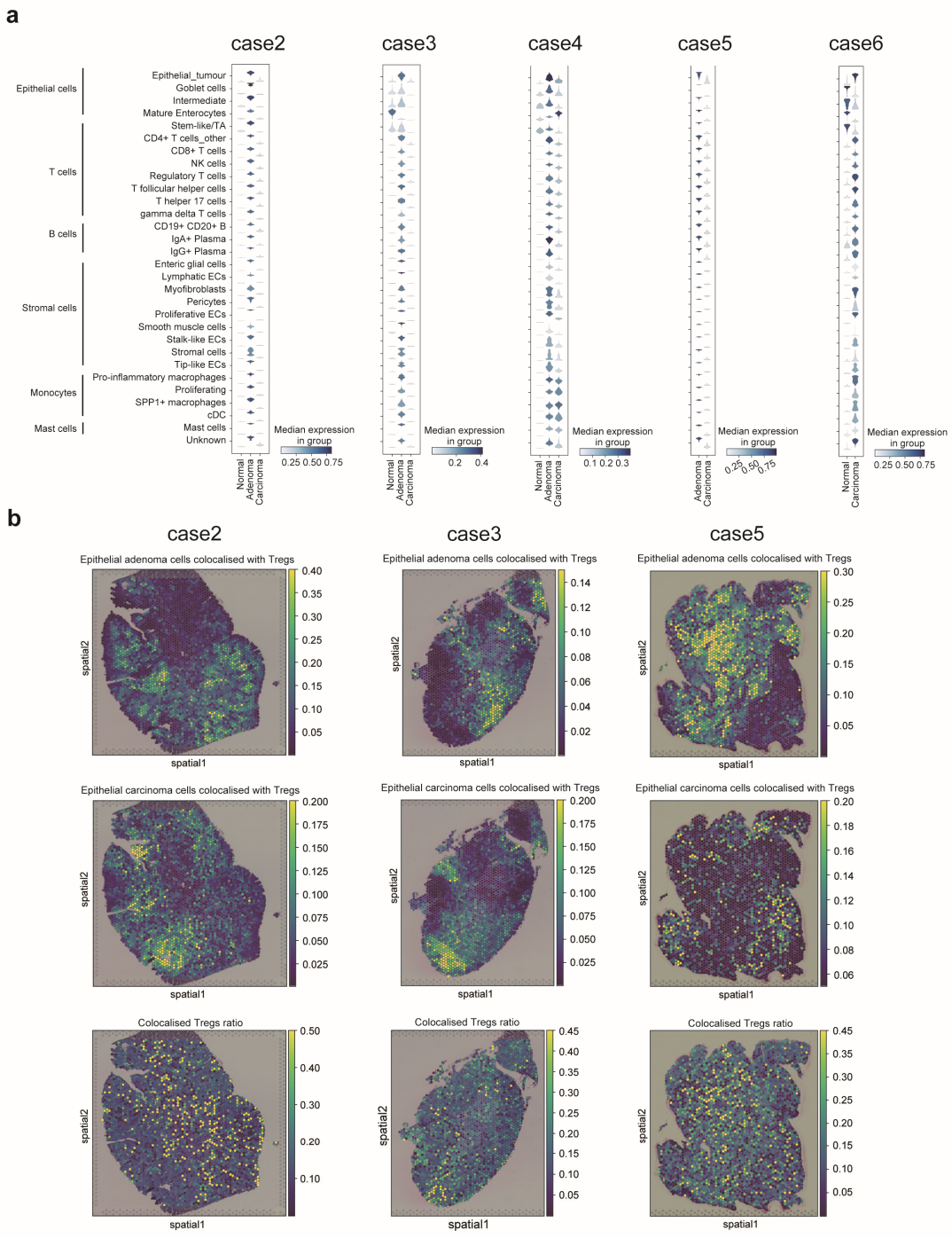
Supplementary Figure 9. MDK receptor gene expression at single-cell level and spatial distribution.
(a) Uniform manifold approximation and projection (UMAP) distribution of the expression levels of genes encoding six MDK receptors in single-cell RNA sequencing data. (b) spatial distribution of imputed expression levels of genes encoding MDK receptors.

Supplementary Figure 10



Supplementary Figure 10. MDK receptor gene analysis in single T cell sequencing data of colorectal cancer. (a) Heatmap of the cell proportion values corresponding to different tissues of each T cell subtype. (b) Dot plot of the expression and proportion of MDK receptor genes per tissue origin type; circle size represents cell proportion. (c) Violin plot of SDC4 expression by tissue origin type. (d) Dot plot of the expression and proportion of MDK receptor genes per T cell subtype. (e) Violin plot of SDC4 expression by T cell subtype.

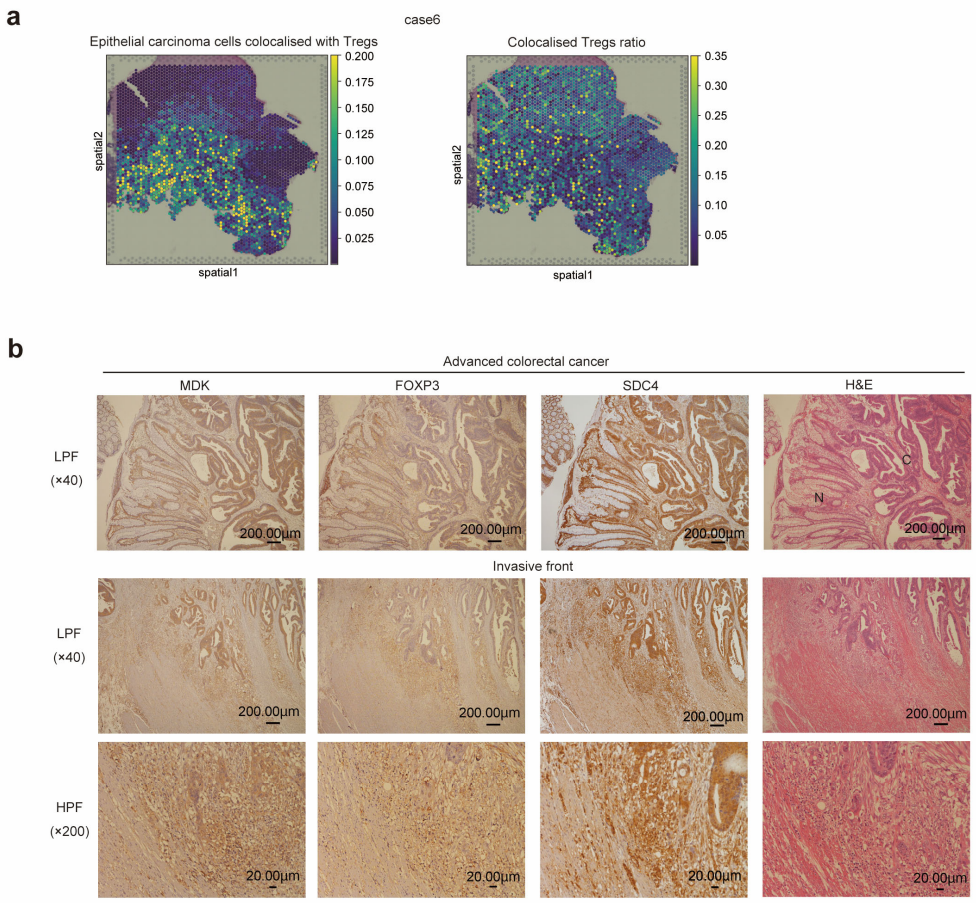
Supplementary Figure 11



Supplementary Figure 11. Spatial cell distribution after processing with integrative analysis using DeepCOLOR in other cases.

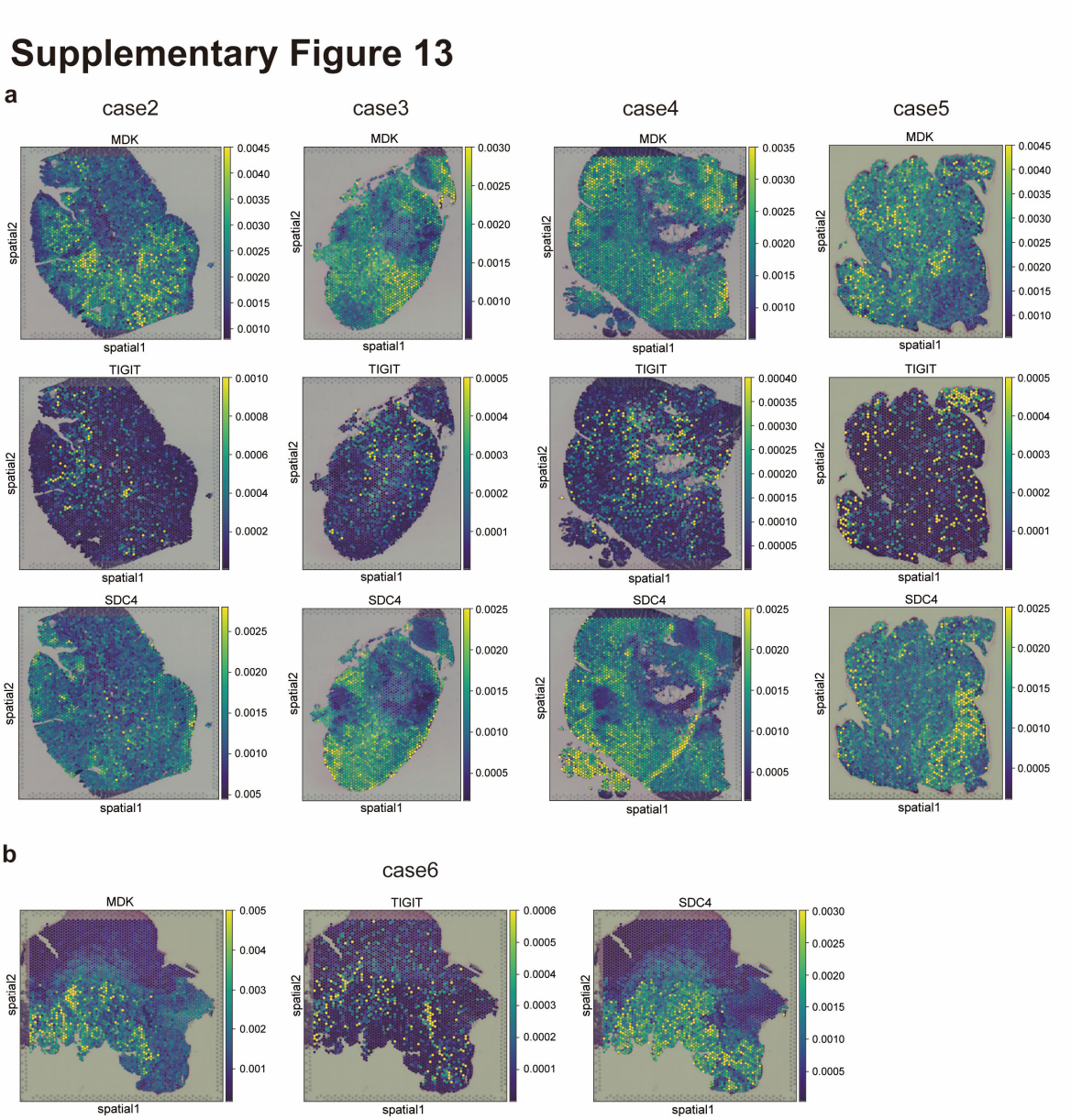
(a) Stacked violin plots of expression levels corresponding to each spatial pathological diagnosis for 30 cell subtypes in cases 2–6. (b) Spatial distribution of epithelial cells colocalised with regulatory T cells (Tregs) and colocalised Tregs ratio in other carcinoma in adenomatous polyps, cases 2, 3, and 5. Colocalised Tregs ratio is calculated as the proportion of Tregs colocalised with epithelial tumour cells among all Tregs.

Supplementary Figure 12



Supplementary Figure 12. Spatial distribution and immunohistochemical analysis in advanced colorectal cancer (CRC).

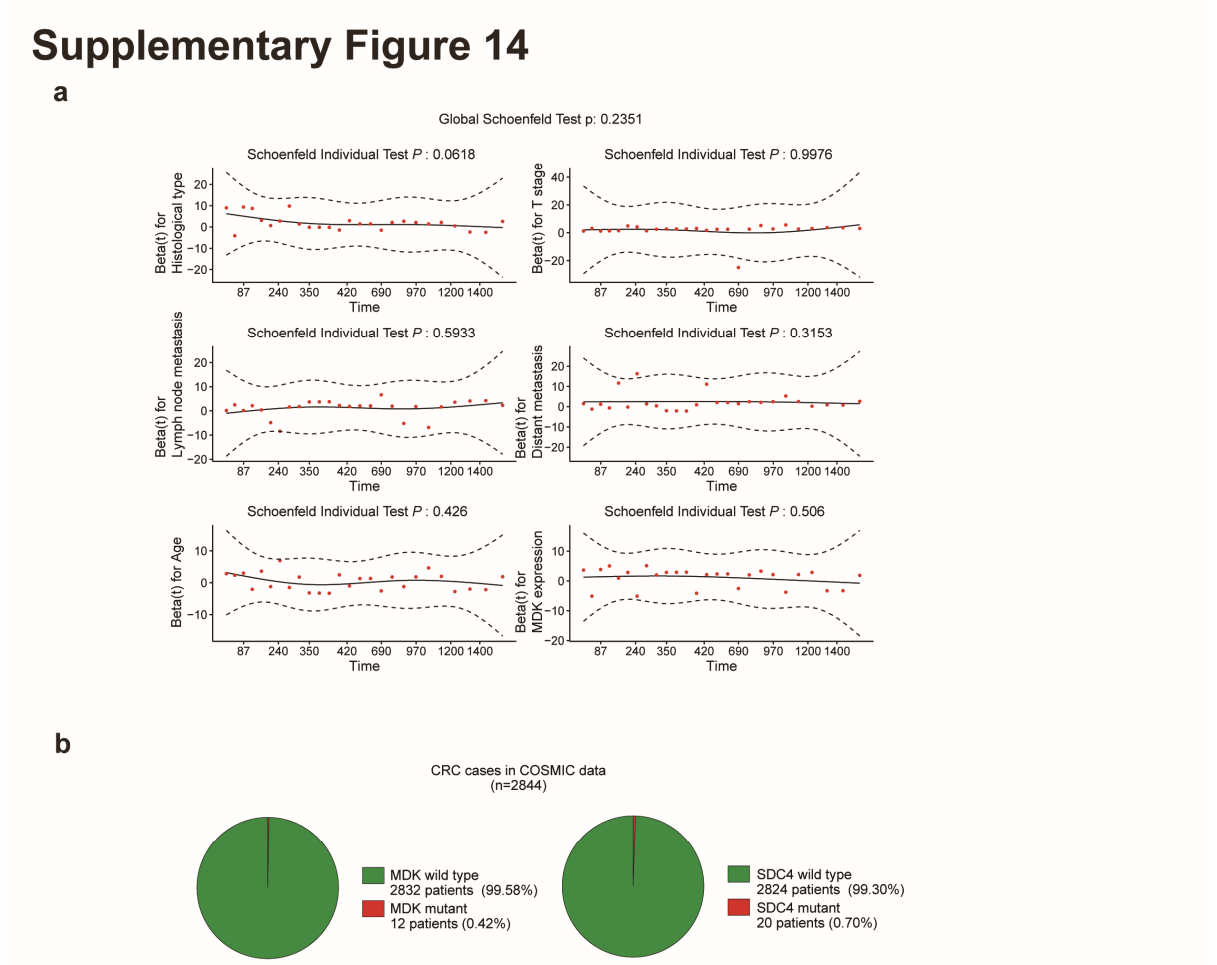
(a) Spatial distribution of epithelial cells colocalised with regulatory T cells (Tregs) and colocalised Tregs ratio in advanced colorectal cancer, case 6. (b) Immunostaining of MDK, FOXP3, and SDC4 in advanced CRC. Pathological diagnosis by H&E staining is as follows; N: normal tissue, C: carcinoma tissue.



Supplementary Figure 13. Spatial distribution of imputed MDK, TIGIT, and SDC4 expression.

Spatial distribution of MDK, TIGIT, and SDC4 expression in cases 2–6.

Supplementary Figure 14



Supplementary Figure 14. Further clinical information of MDK and SDC4.

(a) Schoenfeld residual diagram. The curve of each diagram represents the trend of risk factor change as time goes by. $P > 0.05$ indicates that the variable meets the conditional proportional hazard assumption.

(b) The frequency of mutations in MDK and SDC4 among CRC cases in the COSMIC dataset.