

SUPPLEMENTARY INFORMATION

Spatially organized tumor-stroma boundary determines the efficacy of immunotherapy in colorectal cancer patients

Yu Feng^{1,2,3,4*}, Wenjuan Ma^{5,6*}, Yupeng Zang⁷, Yanying Guo², Young Li^{1,2}, Yixuan Zhang⁸, Xuan Dong², Yi Liu^{2,7}, Xiaojuan Zhan^{2,7}, Zhizhong Pan^{5,9}, Mei Luo^{2,7}, Miaoqing Wu^{5,9}, Ao Chen^{1,2}, Da Kang^{5,9}, Gong Chen^{5,9#}, Longqi Liu^{1,2,4#}, Jingying Zhou^{8#}, Rongxin Zhang^{5,9,10#}

¹BGI Research, Shenzhen 518083, China.

²BGI Research, Hangzhou 310030, China.

³Guangdong Provincial Key Laboratory of Human Disease Genomics, Shenzhen Key Laboratory of Genomics, BGI Research, Shenzhen 519083, China.

⁴Shanxi Medical University - BGI Collaborative Center for Future Medicine, Shanxi Medical University, Taiyuan 030605, China

⁵State Key Laboratory of Oncology in South China, Guangzhou 510060, Guangdong, People's Republic of China.

⁶Department of Intensive Care Unit, Sun Yat-sen University Cancer Centre, Guangzhou 510060, Guangdong, People's Republic of China.

⁷College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China.

⁸School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong SAR, 999077, China.

⁹Department of Colorectal Surgery, Sun Yat-sen University Cancer Centre, Guangzhou 510060, Guangdong, People's Republic of China

¹⁰Lead Contact

*These authors contributed equally to this work

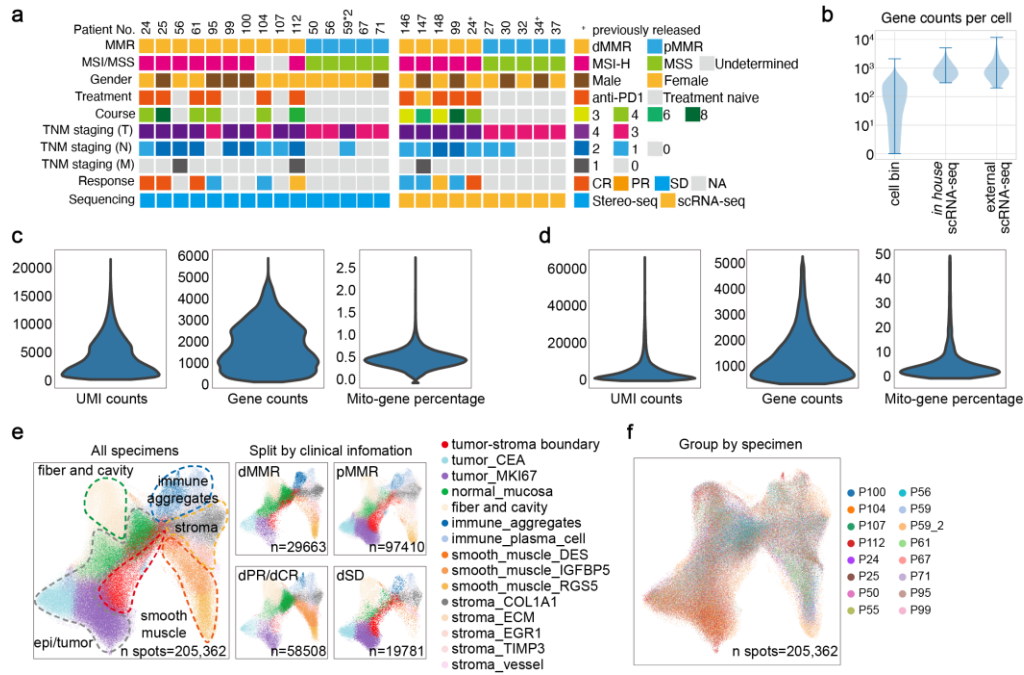
Correspondence:

Gong Chen, Department of Colorectal Surgery, Sun Yat-sen University Cancer Centre, State Key Laboratory of Oncology in South China, Guangzhou, Guangdong, 510060, P. R. China. TEL: +86 20 87343584; FAX: +86 20 87343584; EMAIL: chengong@sysucc.org.cn

Longqi Liu, BGI Research Hangzhou, Hangzhou, 310030, P. R. China. TEL: +86 0571 88226065; EMAIL: liulongqi@genomics.cn

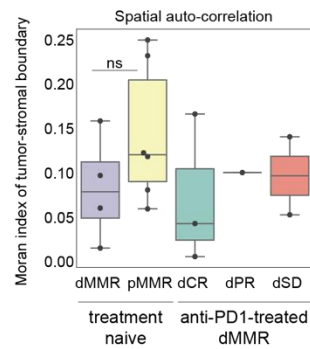
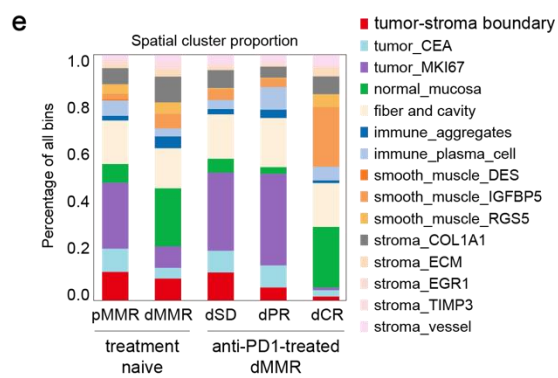
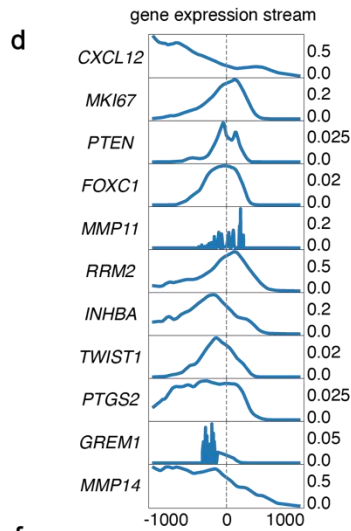
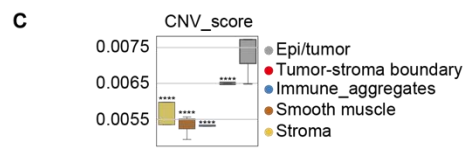
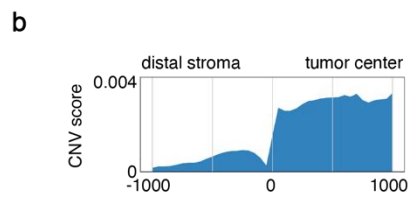
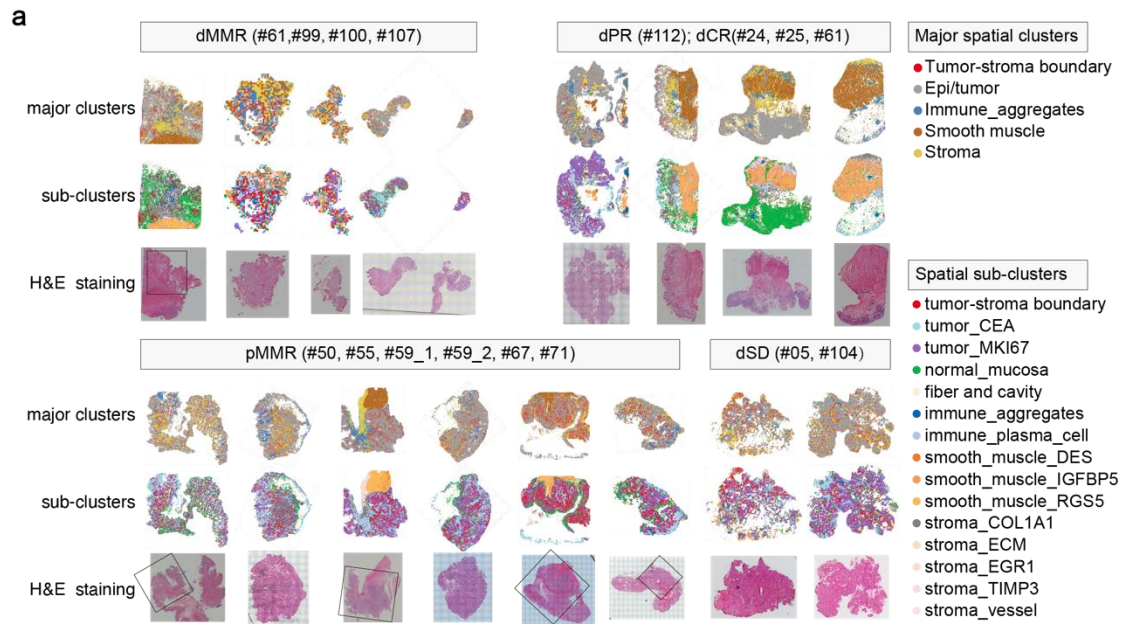
Jingying Zhou, School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong SAR, 999077, P.R. China. TEL: +852 39431350; FAX: +852 26035139; EMAIL: zhoujy@cuhk.edu.hk

Rongxin Zhang, Department of Colorectal Surgery, Sun Yat-sen University Cancer Centre, State Key Laboratory of Oncology in South China, Guangzhou, Guangdong, 510060, P. R. China. TEL: +86 20 87343584; FAX: +86 20 87343584; EMAIL: zhangrx@sysucc.org.cn



Supplementary Figure 1. Basic information of the Stereo-seq data.

a Matrix plots of the clinical informations of our CRC patient cohort are shown. **b** Violin plots of the gene counts per cell in cell bin segmented of Stereo-seq data from a representative patient #59, our in house scRNA-seq data and a representative public scRNA-seq data (GSE178341) are shown. **c** Violin plots of unique molecular identifier (UMI) counts, gene counts and mitochondria-related (Mito-gene) genes percentage in the Stereo-seq, or **d** scRNAseq data are shown. **e** UMAP of the spatial transcriptome data from 205,362 bin spots from 15 CRC patients are shown. Bins are colored by spatial cell clusters, split by clinical information or **f** colored by patient ID.

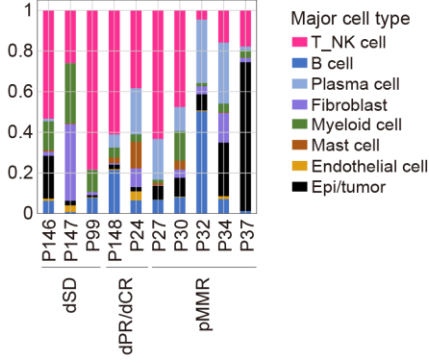


Supplementary Figure 2. *Global landscape of spatial cell clusters*

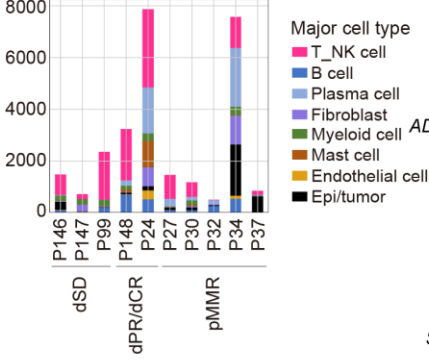
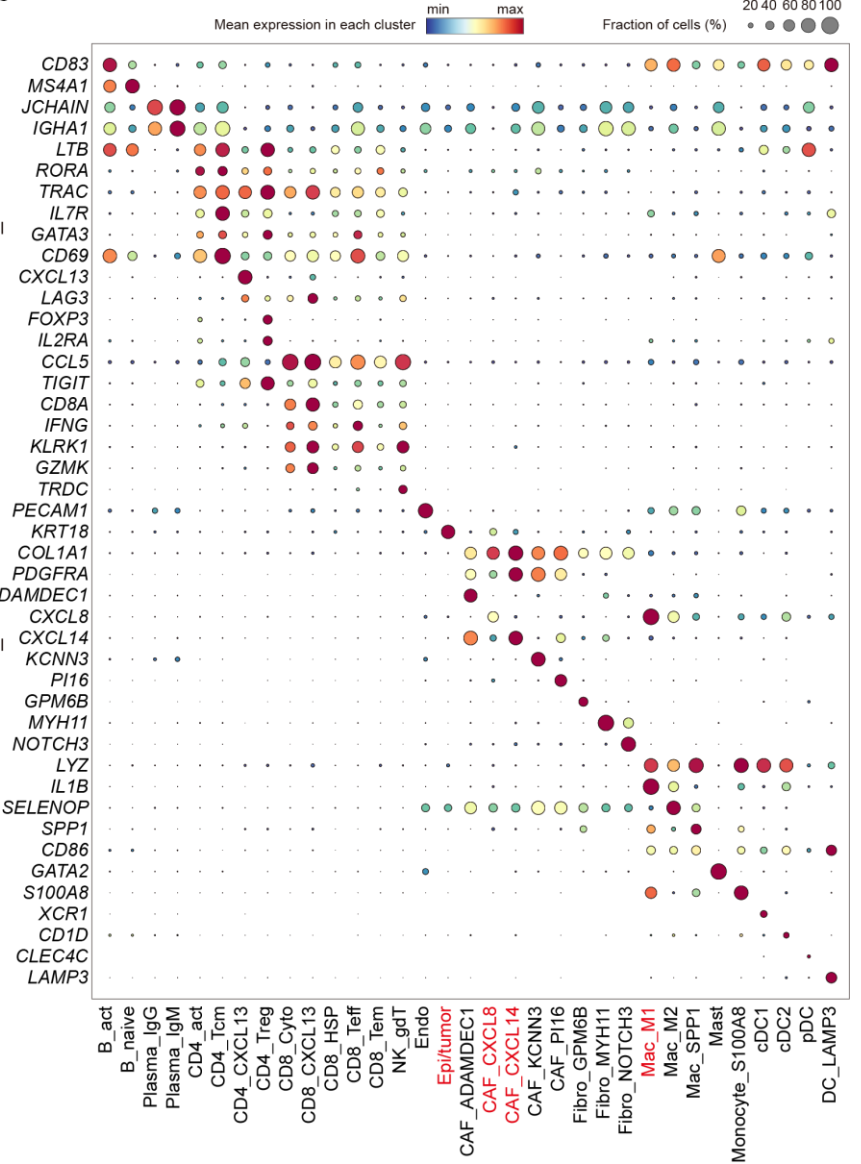
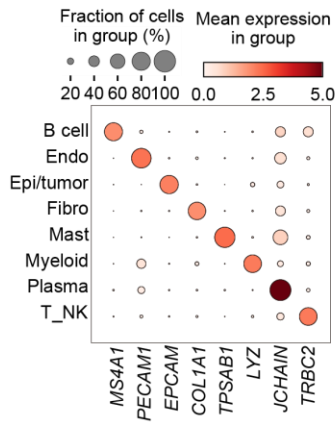
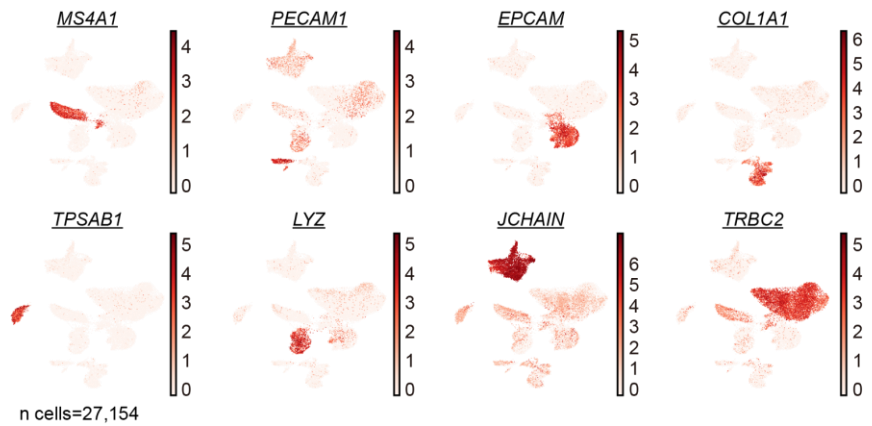
a The spatial plots of the major spatial clusters, sub-clusters and H&E image in each specimen are shown. **b** The stacked stream plot of the CNV scores from the distal stroma (-1000 μ m, left) to the tumor center (1000 μ m, right) is shown. The mean CNV score in each 1mm interval is smoothed using slinger model. The distance of boundary was set to 0 μ m. **c** The box-whisker plots of the CNV scores in each major spatial cluster are shown (n spots: Epi/tumor=82,655, tumor-stroma boundary=17,639, immune aggregates=16,304, smooth muscle=24,172, stroma=28,012). The asterisk represents the comparison of the epi/tumor clusters towards other spatial clusters, analyzed by two-tailed unpaired Student-t test with Bonferroni correction. ****, $p < 0.0001$. Data are represented mean \pm IQR (25%~75%) with minima (lower bar) and maxima (upper bar). **d** The stacked stream plots of the indicated gene expressions from the distal stroma (-1000 μ m, left) to the tumor center (1000 μ m, right) are shown. The mean expression level in each 1mm interval is smoothed using slinger model. The distance of boundary was set to 0 μ m. **e** Stacked box plots show the proportions of spatial clusters in the indicated patient groups. **f** Spatial auto-correlation of tumor-stroma boundary is shown by Moran index. Data are represented mean \pm IQR (25%~75%) with minima (lower bar) and maxima (upper bar). Statistics are analyzed by two-tailed unpaired Student-t test with Bonferroni correction. ns, not significant. N number of the patients: pMMR=6; dMMR=4; dSD=2; dPR=1; dCR=3.

a

Major cell type composition in scRNA-seq

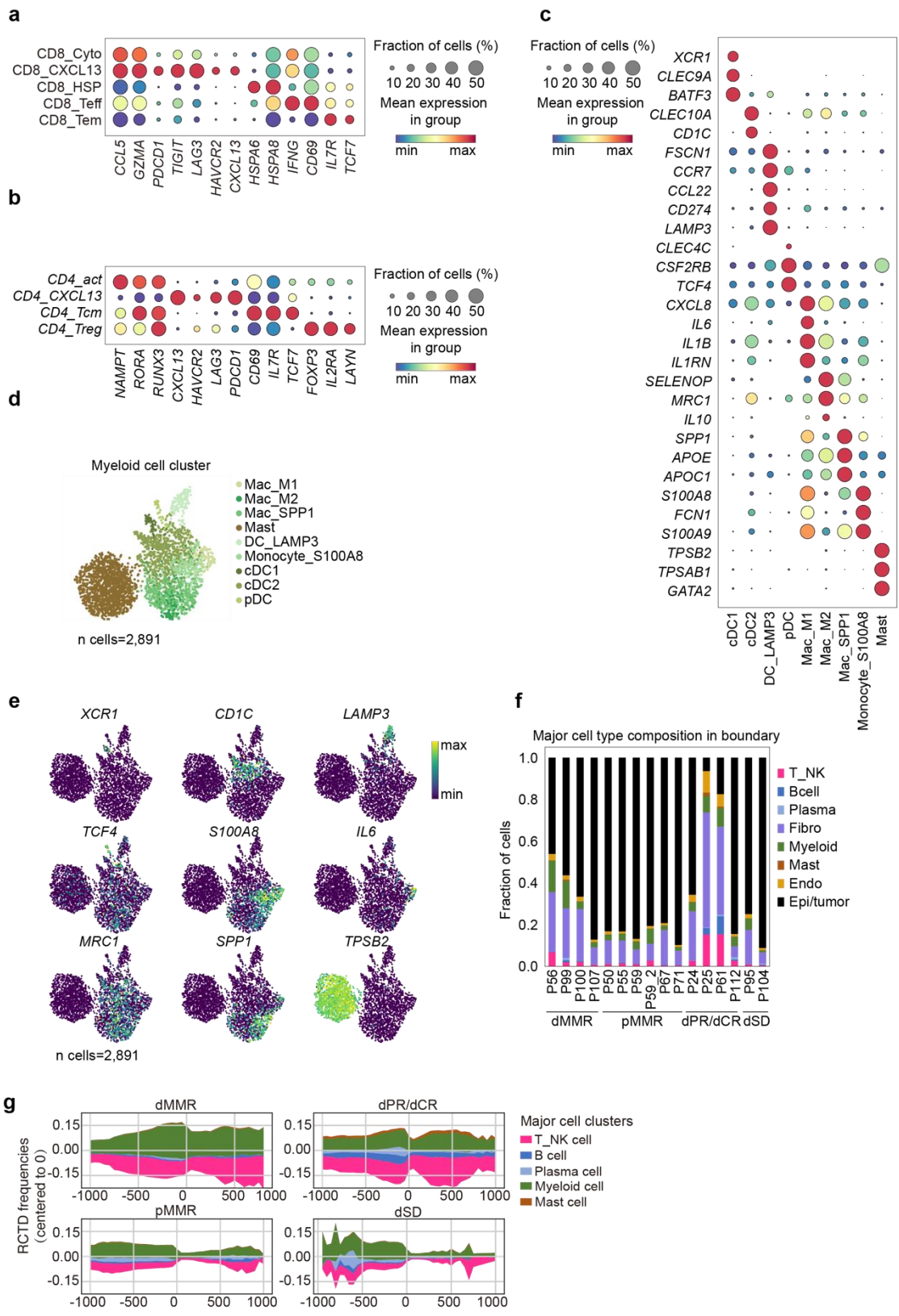


Major cell counts in scRNA-seq

**b****c****d**

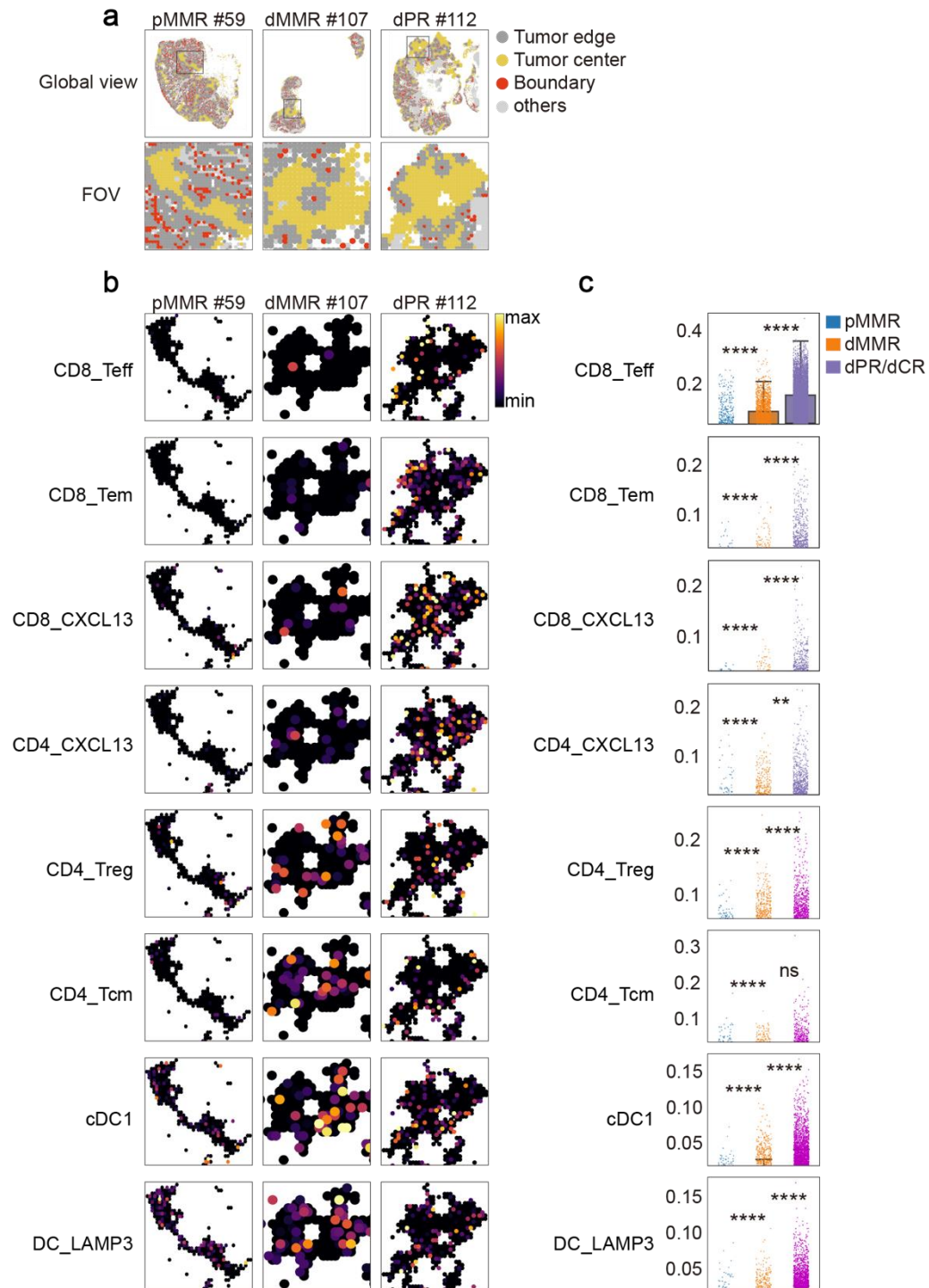
Supplementary Figure 3. *Basic information of the scRNA-seq data.*

a The stacked bar plots of proportions or absolute cell counts of the major cell clusters in the scRNA-seq data of each patients are shown. **b** Bubble plots of marker gene expressions in all cell clusters or **c** major cell clusters from the scRNA-seq dataset are shown. The plots are sized by the fraction of cells with positive gene expression, while the color represents the gene expression level. **d** Expression level of canonical marker genes for each major cell cluster is shown in UMAP (n cells=27,154; n samples: 5 pMMR, 2 dPR/CR, 3 dSD).



Supplementary Figure 4. Characterization of T cell and myeloid cell clusters.

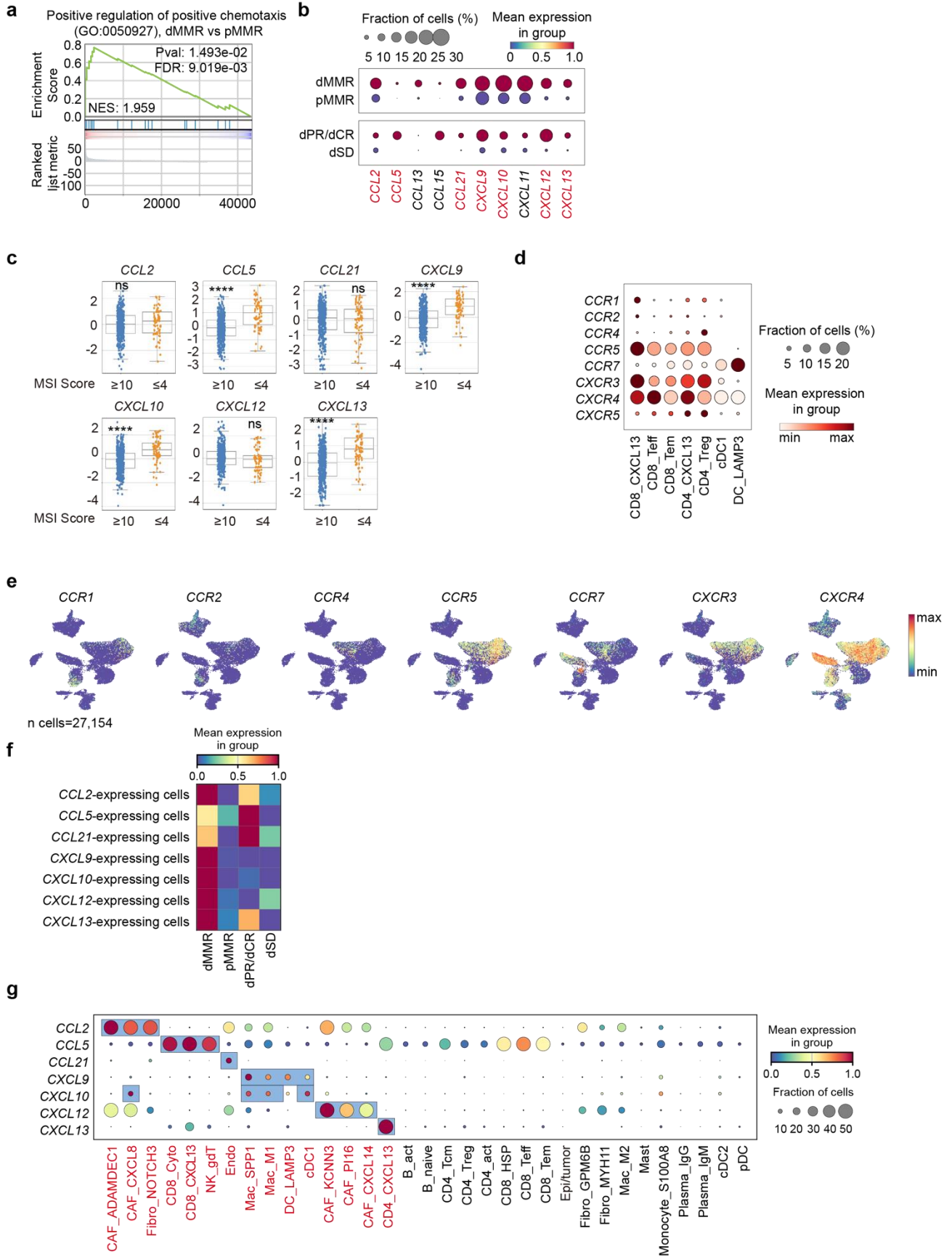
a Bubble plots of marker gene expressions in CD8⁺T cells, **b** CD4⁺T cells, and **c** myeloid cells from the scRNA-seq dataset are shown. The plot are sized by the fraction of cells with positive gene expression, while the color represents the gene expression level. **d** UMAP of myeloid cell clusters (n cells=2,891; n samples: 5 pMMR, 2 dPR/CR, 3 dSD) and **e** expression level of canonical marker genes for each cluster are shown. **f** Stacked box plots of the proportions of major cell clusters in the tumor-stromal boundary of each patient from Stereo-seq data are shown. **g** The stacked stream plots of major immune cell distribution patterns from distal stroma (-1000 μ m, left) to tumor center (1000 μ m, right) in indicated patient groups are shown. The mean RCTD frequencies of each immune cell in each 1mm interval was smoothed using slinger model and colored by cell clusters in accordance with **Fig. 2b**.



Supplementary Figure 5. The distributions of immune cell clusters in the spatial map.

a Representative images of overviewed FOVs and **b** indicated immune cell clusters in the spatial map of pMMR patient #59, dMMR patient #107 and dPR patient #112 are shown. **c** The RCTD frequencies of indicated immune cell clusters are shown in dot

plot graphs in indicated patient groups. Data are represented as box-whisker plot (mean±IQR: 25%~75%) and analyzed by two-tailed unpaired Student-t test with Bonferroni correction. ns, not significant; **, $p<0.01$; ****, $p<0.0001$. N spots: pMMR=7,694; dMMR=2,595; dPR/dCR=10.283.

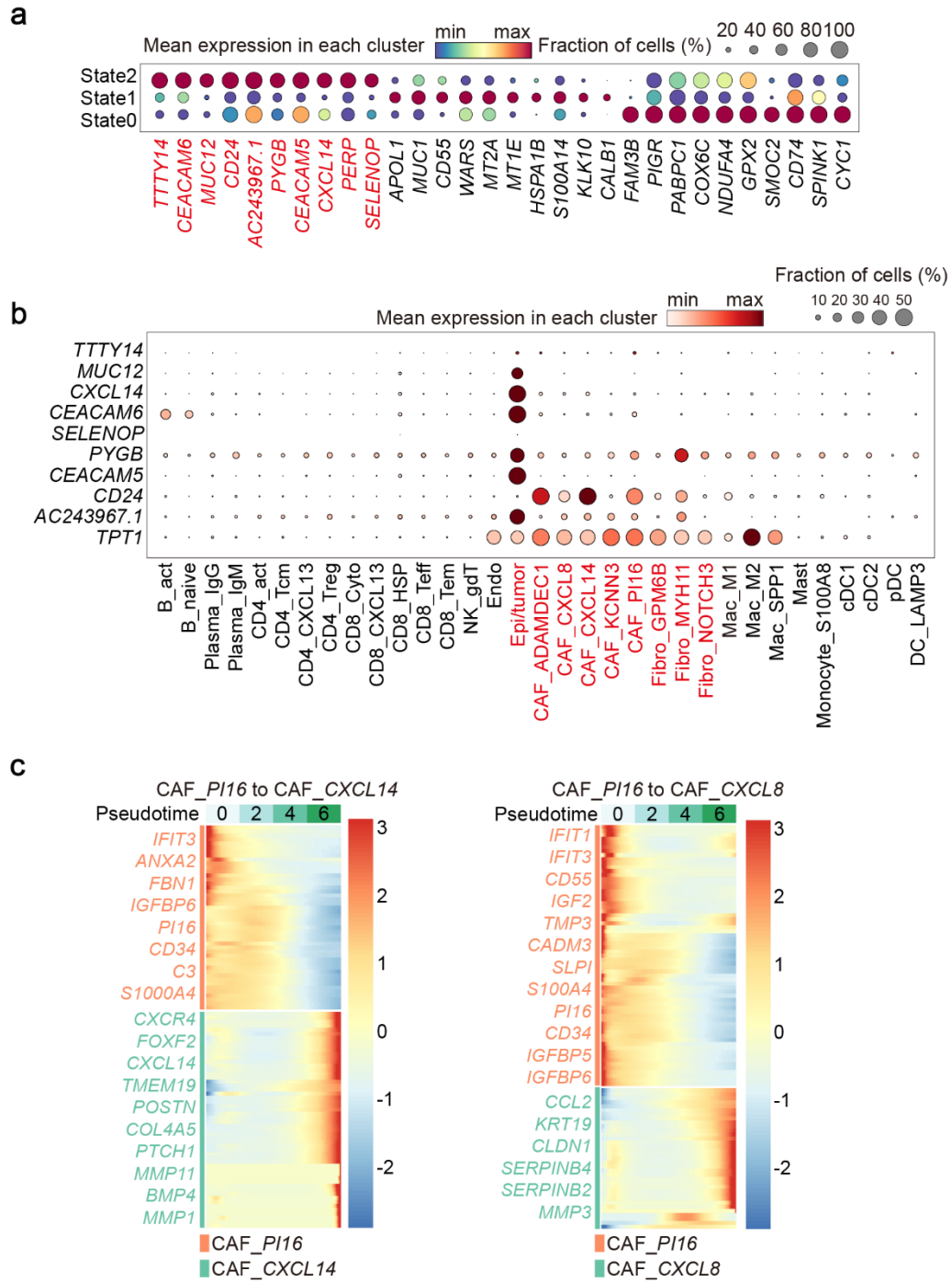


Supplementary Figure 6. Key chemotaxis pathways in T cells and DCs from the tumor-stroma boundary of treatment naïve dMMR and anti-PD1-treated dPR/dCR patients.

a GSEA plot of the upregulated genes related to positive regulation of positive chemotaxis (GO:0050927) in the tumor-stroma boundary of treatment naïve dMMR compared to pMMR is shown. GSEA plot showing the up-regulation of the genes related to in dMMR boundaries (in comparison to pMMR boundaries). **b** Bubble plots of indicated chemokine expressions in the tumor-stroma boundary of indicated patient groups. The plots are sized by the fraction of cells with positive gene expression, while the color represents the gene expression level. The chemokines with higher expressions in treatment naïve dMMR compared to pMMR, as well as in anti-PD1-treated dPR/dCR compared to dSD are highlighted in red. **c** The expressions of indicated chemokines from MSI-hi (MSI score \geq 10, n = 78 patients) and MSI-lo (MSI score \geq 10, n = 494 patients) groups in COAD TCGA data are shown as box-whisker plot. Data are represented mean \pm IQR (25%~75%) with minima (lower bar) and maxima (upper bar). Statistics are analyzed by two-tailed unpaired Student-t test with Bonferroni correction. ns, not significant; ****, p<0.0001. **d** Bubble plots of chemokine receptors of chemokines highlighted in **b** from indicated immune cell clusters are shown. The plots are sized by the fraction of cells with positive gene expression, while the color represents the gene expression level. **e** UMAP of indicated chemokine receptor expressions from the scRNA-seq dataset (n cells=27,154; n samples: 5 pMMR, 2 dPR/CR, 3 dSD). **f** Heatmap of chemokine-expressing cells in the tumor-stroma boundary of indicated patient groups. **g** Bubble plots of the chemokine expressions profiling in the cell clusters from the scRNA-seq dataset. The cells of interests are highlighted in red and framed.

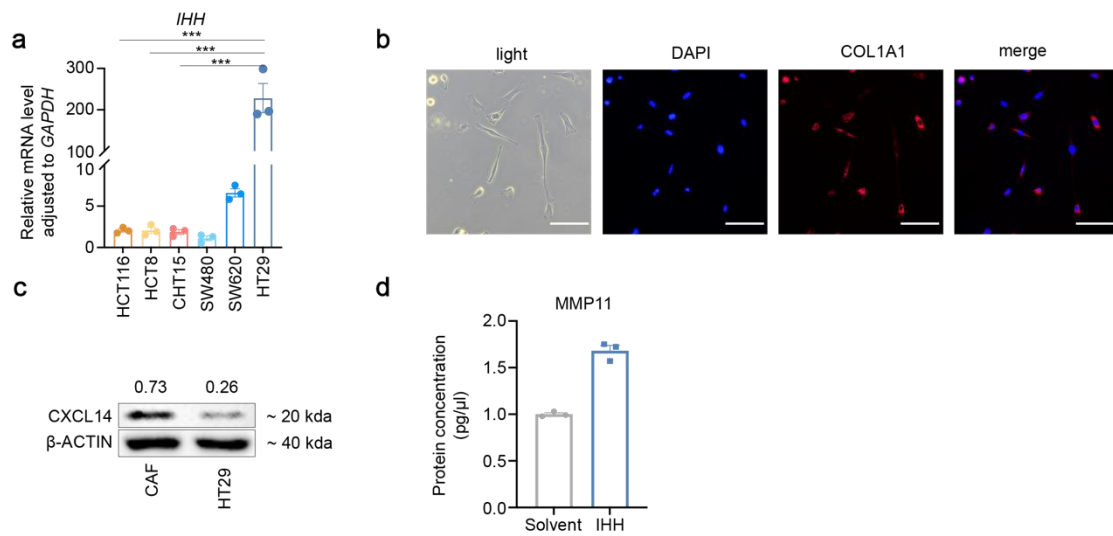
Supplementary Figure 7. Key pathways in the potential interactions between DC_LAMP3 and T cell clusters in the tumor-stroma boundary.

a Correlations of DC_LAMP3 and T cell clusters in the Stereo-seq data from treatment naïve dMMR patients. Each scattered dot represents a single spatial bin. **b** Pearson correlation of the signature scores of DC_LAMP3 and T cell clusters from MSI-hi (MSI score \geq 10, n = 78 patients) and MSI-lo (MSI score \geq 10, n = 494 patients) groups in COAD TCGA data are shown. The signature Z-scores are calculated using the top10 DEGs from each cell cluster in our scRNA-seq data. **c** The box and whisker plot shows the absolute distance of each T cell cluster to proximal DC_LAMP3. Data are represented as box-whisker plot (mean \pm IQR: 25%~75%) and analyzed by two-tailed unpaired Student-t test with Bonferroni correction. N cells: CD4_CXCL13=3,036, CD4_Treg=3,806, CD8_CXCL13=980, CD8_Tem=7,568. **d** The chord plots show the ligand-receptor interaction of CD80_CD28 (left) and CD80_CTLA4 (right) between DC_LAMP3 and CD4⁺T cell clusters. **e** Bubble plots of expression profiling of ligand-receptor pairs CD80_CTLA4, CD80_CD28 and CD274_PDCD1 in the tumor-stromal boundary of indicated patient groups are shown. The plots are sized by the fraction of cells with positive gene expression, while the color represents the gene expression level. **f** UMAP of DC_LAMP3 cluster and **g** its marker genes are shown. N cells=27,154; n samples: 5 pMMR, 2 dPR/CR, 3 dSD. **h** Bubble plots of the expressions of co-stimulatory and co-inhibitory ligands in the single cell clusters. The plot are sized by the fraction of cells with positive gene expression, while the color represents the gene expression level. **i** GSEA plot of the upregulated genes related to regulation of leukocyte activation (GO:0002694) in the tumor-stroma boundary of treatment naïve dMMR compared to pMMR patients.



Supplementary Figure 8. *The features of CAFs in the tumor-stroma boundary.*

a Bubble plots of the expressions of the top 10 genes (ranked by p-value) in each cell state or **b** the top 10 highlighted genes in state 0 in the single cell clusters. The plots are sized by the fraction of cells with positive gene expression, while the color represents the gene expression level. **c** Heatmaps of the pseudo-temporal expression patterns of the top 100 genes during the lineage 0 (CAF_ *P116* to CAF_ *CXCL14*) or lineage 1 (CAF_ *P116* to CAF_ *CXCL8*) differentiation. Genes of interests are labelled.



Supplementary Figure 9. *MMP11* expression is upregulated by *IHH* in *CXCL14*⁺*CAF* in vitro.

a The relative expression of *IHH* to *GAPDH* in the 6 human CRC cell lines. Data are represented as mean±SD and analyzed by two-tailed unpaired Student-t test. N=3 independent experiment. **b** The representative images of the morphology and IF staining of COL1A1 in the CAFs are shown. DAPI was used as a positive control for cell nuclei staining. Scale bars, 20µm. **c** Western blot analysis of CXCL14 in CAF and HT29 cell lines is shown. β-actin serves as loading control. Number indicates the relative expression towards β-actin. This result represents 3 independent experimental repetitions. **d** The relative expression of *MMP11* to *GAPDH* in HT29 cells treated with PBS or IHH recombinant protein. Data are represented as mean±SD and analyzed by unpaired Student-t test. ***, p<0.001. N=3 independent experiment.