

From Healthy to Tumor: Spatial Shifts in the Human Intestine

A Comparative Study of Visium V2 Datasets in Seurat (R)

Alberto Lupatin, University of Trento - CiBio

29-05-2025

Load Libraries

```
library(Matrix)
library(Seurat)
library(tidyverse)
library(ggplot2)
library(ggemb1)
library(patchwork)
library(corrplot)

colors <- c("#6E100B", "#9EF94B", "#2B1333")
```

Load Data

Intestinal Cancer Dataset

Dataset taken from 10xGenomics (source).

```
cnc_path <- './Data/Cancer'

# Create Seurat Object
counts <- ReadMtx(
  mtx = file.path(cnc_path, "matrix.mtx.gz"),
  features = file.path(cnc_path, "features.tsv.gz"),
  cells = file.path(cnc_path, "barcodes.tsv.gz")
)

coordinates <- Read10X_Coordinates(file.path(cnc_path, "spatial/tissue_positions_list.csv"),
                                     filter.matrix = T)

cnc <- CreateSeuratObject(counts, assay = "RNA")

# Add image
image <- Read10X_Image(file.path(cnc_path, "spatial"))

image_fixed <- new(
  Class = "VisiumV1",
  assay = "RNA",
  key = "slice1_",
  image = image@image,
  scale.factors = image@scale.factors,
```

```

coordinates = coordinates
)
image_fixed@spot.radius <- Radius(image_fixed)

cnc@images$slice1 <- image_fixed

cnc <- subset(cnc, cells = intersect(colnames(cnc), rownames(coordinates)))
cnc <- AddMetaData(cnc, metadata = coordinates)

```

Healthy Intestine Dataset

Dataset taken from Human Atlas database (source). It was composed on different tissues and areas. Loading only the epithelium and fibroblast one to make comparisons with the tumor sample.

```

ctr_path <- "./Data/Control"

# Loading Seurat objects
epithelium <- readRDS(file.path(ctr_path, "epithelium.RDS"))
fibroblasts <- readRDS(file.path(ctr_path, "fibroblasts.RDS"))

# Merging the two objects into 1
ctr <- merge(epithelium, fibroblasts)
rm(epithelium, fibroblasts)

# Subset cells
set.seed(123) # Reproducibility purpose
cells_to_keep <- sample(seq_len(ncol(ctr)), size = min(20000, ncol(ctr)))
ctr <- ctr[, cells_to_keep]

```

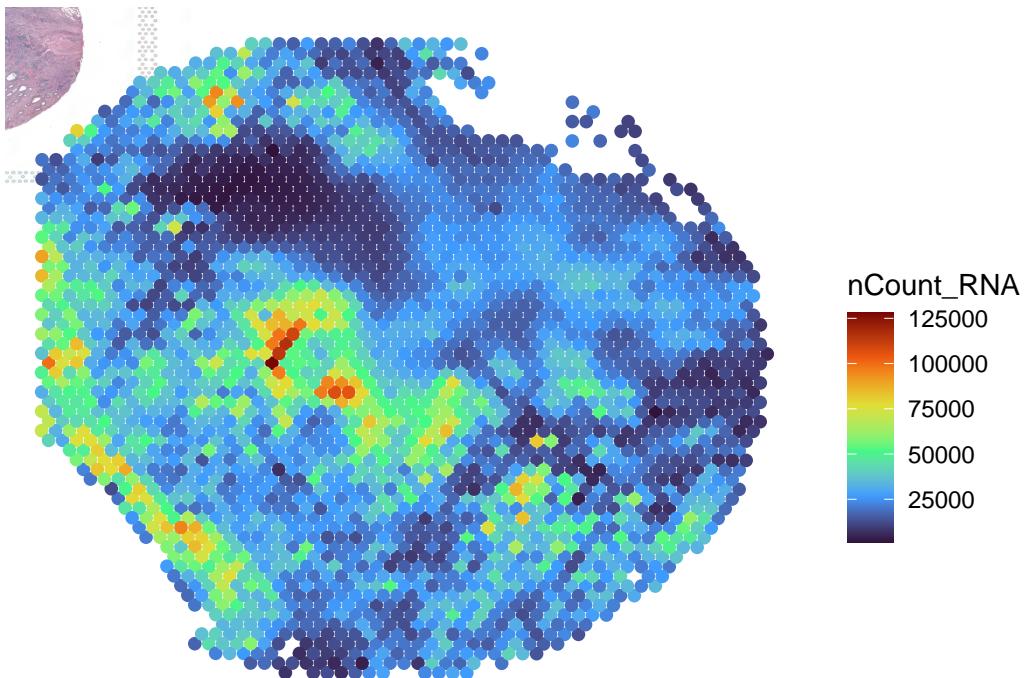
Preprocessing

Pre-Normalization Expression

```

SpatialFeaturePlot(cnc, features = "nCount_RNA",
                    image.scale = "hires", pt.size.factor = 1) +
  theme(legend.position = "right") +
  scale_fill_viridis_c(option = "H")

```



```
ggsave("./Images/Pre-Normalization RNA_Count.pdf")
```

Normalization, feature selection and scaling

```
cnc <- SCTtransform(cnc,
                      assay = "RNA",
                      verbose = FALSE)

ctr <- SCTtransform(ctr,
                     assay = "RNA",
                     verbose = FALSE)
```

Dimensionality reduction, and clustering

```
# Cancer
cnc <- RunPCA(cnc, assay = "SCT", verbose = FALSE)

cnc <- FindNeighbors(cnc, reduction = "pca", dims = 1:30, verbose = FALSE)

cnc <- FindClusters(cnc, verbose = FALSE)

cnc <- RunUMAP(cnc, reduction = "pca", dims = 1:30, verbose = FALSE)

# Control
ctr <- RunPCA(ctr, assay = "SCT", verbose = FALSE)
```

```

ctr <- FindNeighbors(ctr, reduction = "pca", dims = 1:30, verbose = FALSE)
ctr <- FindClusters(ctr, verbose = FALSE)
ctr <- RunUMAP(ctr, reduction = "pca", dims = 1:30, verbose = FALSE)

```

Spatially Variable Features

Cancer Marker Selection

Choosing the following gene markers:

- **Cytokeratine 20 (KRT20)**: cytokeratin protein expressed in differentiated epithelial cells of the gastrointestinal tract, often overexpressed in CRC as a marker of tumor cells(Spisak et al. 2024).
- **alpha-Smooth Muscle Actin (ACTA2)**: normally expressed by fibroblasts, but its overexpression marks cancer-associated fibroblasts (CAFs) in the tumor stroma. CAFs contribute to tumor progression by secreting growth factors, remodeling the extracellular matrix, and promoting invasion (Gu et al. 2019).

And the following immune markers:

- **CD3D, CD4, FOXP3, CD8A**: T-cell markers (pan-T, helper T, regulatory T, cytotoxic T respectively); their presence indicates immune infiltration in the tumor microenvironment (Kuwahara et al. 2019a) (Kuwahara et al. 2019b).
- **CD68, CD163**: macrophage markers (pan-macrophage and M2-like macrophage, respectively), indicating macrophage infiltration and potential immunosuppression (*CD163*) in the tumor microenvironment (Hou et al. 2024)

```

gene_markers = c("KRT20", "ACTA2")
immune_markers <- c("CD3D", "CD4", "FOXP3", "CD8A", "CD68", "CD163")

```

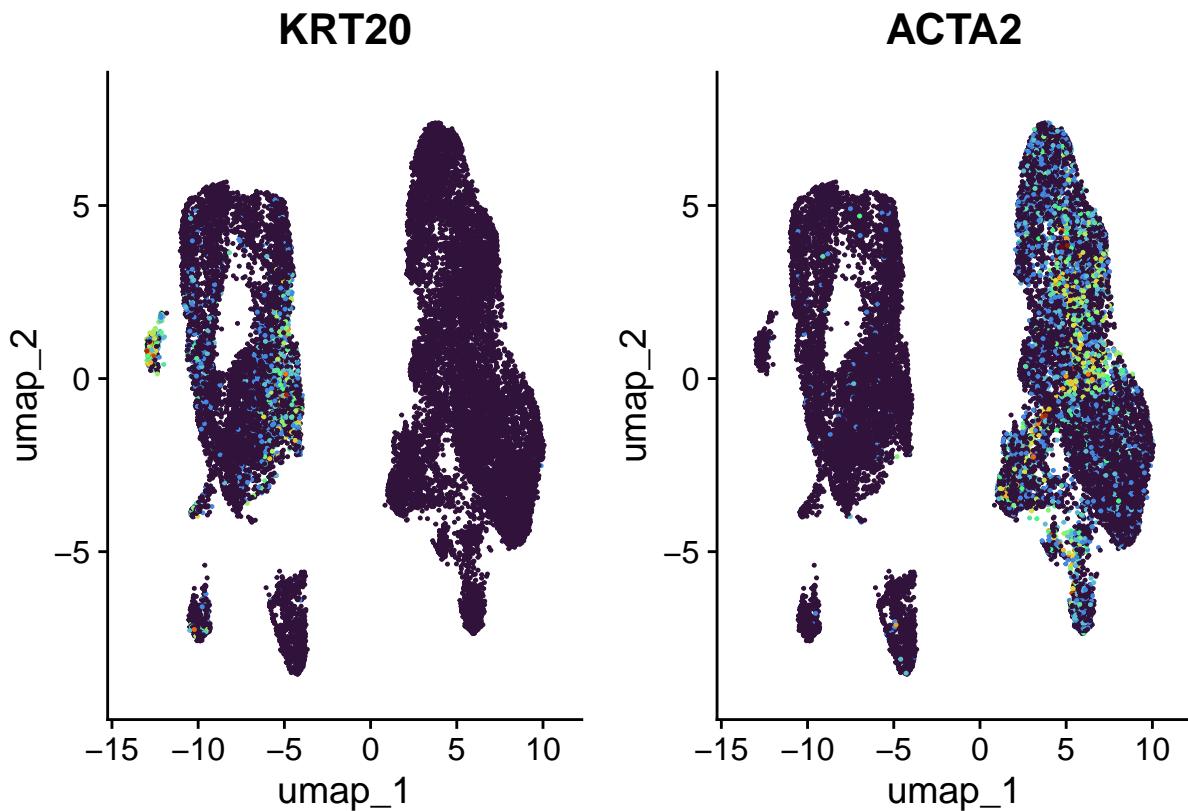
Identifying the possible tissue sample composition

Based on the marker expression, we can infer the possible cell composition of the tissues.

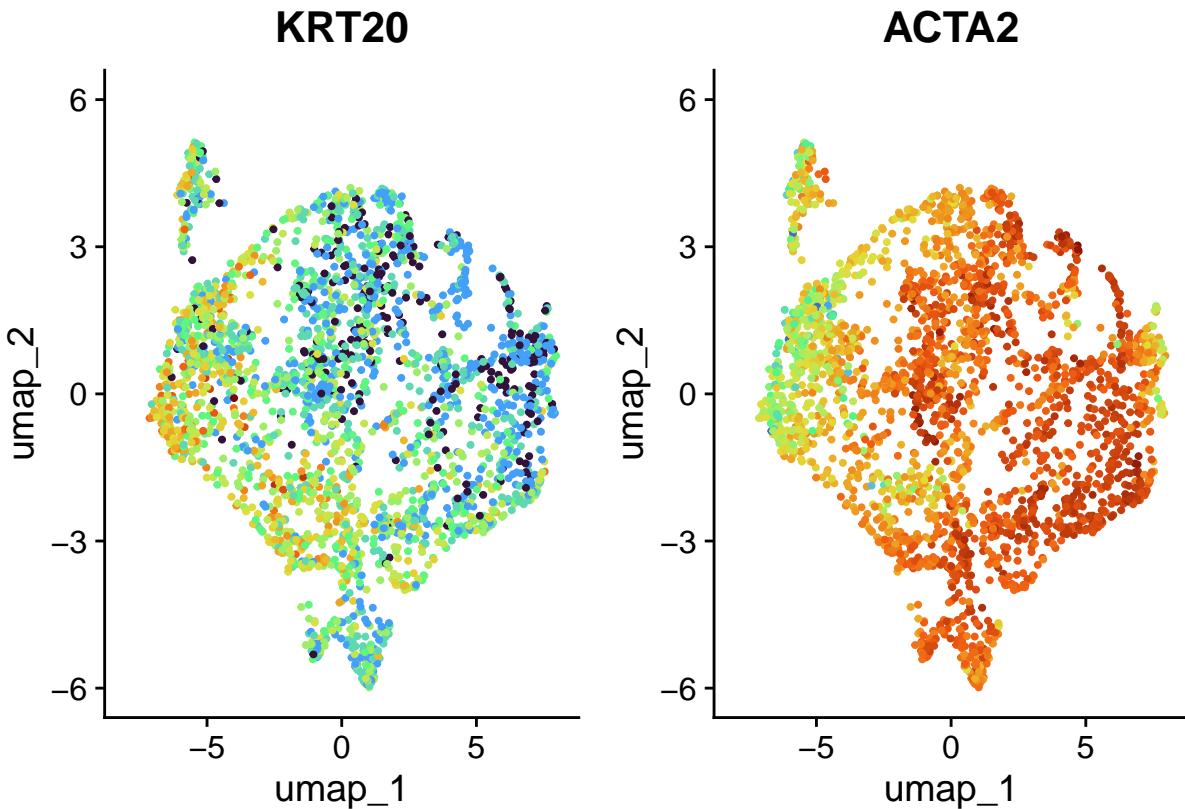
```

FeaturePlot(ctr, features = gene_markers, ncol = 2) &
  scale_color_viridis_c(option = "H") &
  NoLegend()

```



```
ggsave("./Images/Tissue Identification Normal.pdf")  
  
FeaturePlot(cnc, features = gene_markers, ncol = 2) &  
  scale_color_viridis_c(option = "H") &  
  NoLegend()
```



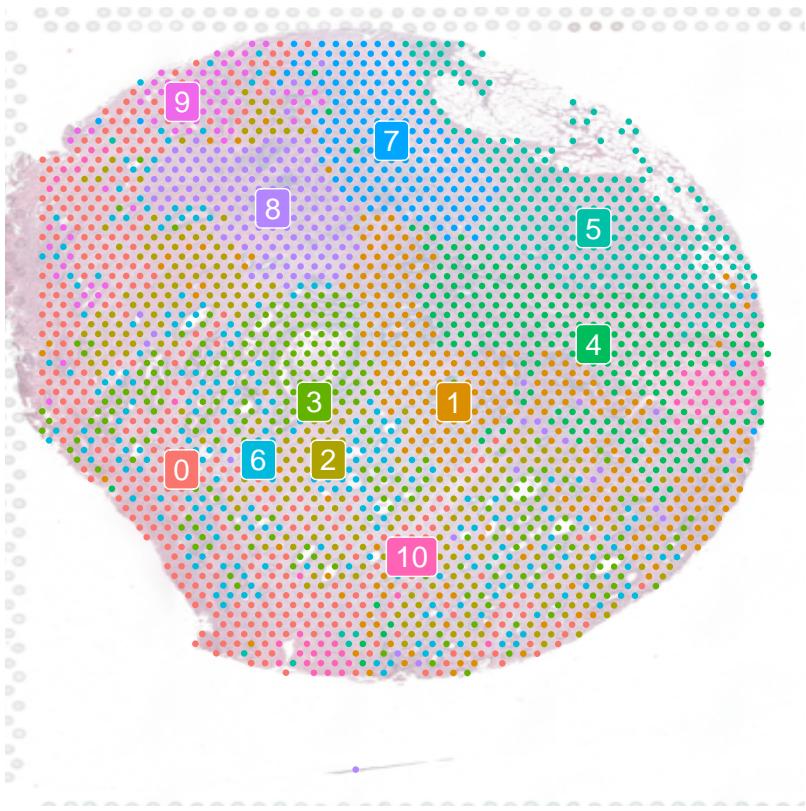
```
ggsave("./Images/Tissue Identification Cancer.pdf")
```

- **Control Tissue:**
 - Left part: moderate *KRT20* expression and almost no *ACTA2* expression, compatible with **epithelial** tissue;
 - Right part: moderate *ACTA2* expression and almost no *KRT20* expression, compatible with **connective** tissue.
- **Tumor Tissue:** *KRT20* is expressed in some areas with low *ACTA2* and viceversa, indicating a possible mutual exclusivity of over-expression in some clusters

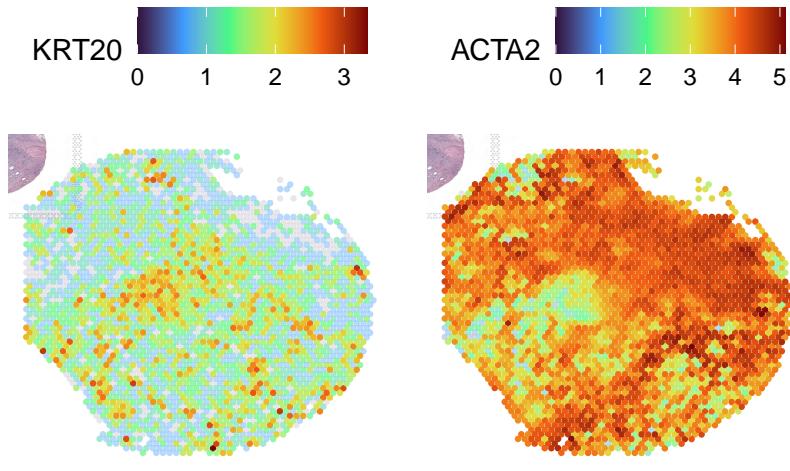
Cancer Dataset Analysis

Cluster Marker Expression

```
SpatialDimPlot(cnc, label = TRUE, label.size = 4, alpha = 8, image.alpha = .5) +
  NoLegend()
```

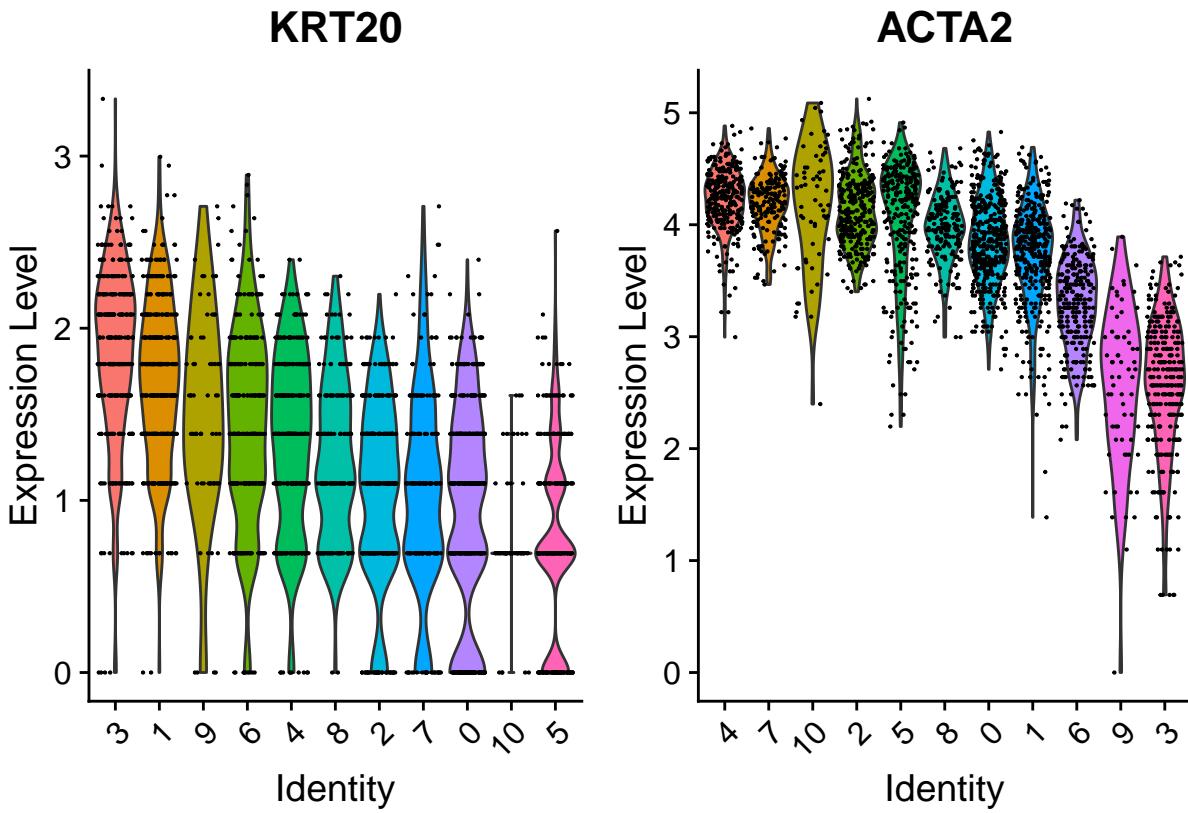


```
ggsave("./Images/Clusters.pdf")
SpatialFeaturePlot(cnc, features = gene_markers,
                  pt.size.factor = 1, alpha = c(0.1, 1.5), ncol = 3, image.scale = "hires") &
  scale_fill_viridis_c(option = "H")
```



```
ggsave("./Images/Gene Markers Cancer.pdf")

VlnPlot(cnc, features = gene_markers,
        group.by = "seurat_clusters", sort = "increasing")
```



```
ggsave("./Images/Gene Marker Cancer Violin.pdf")
```

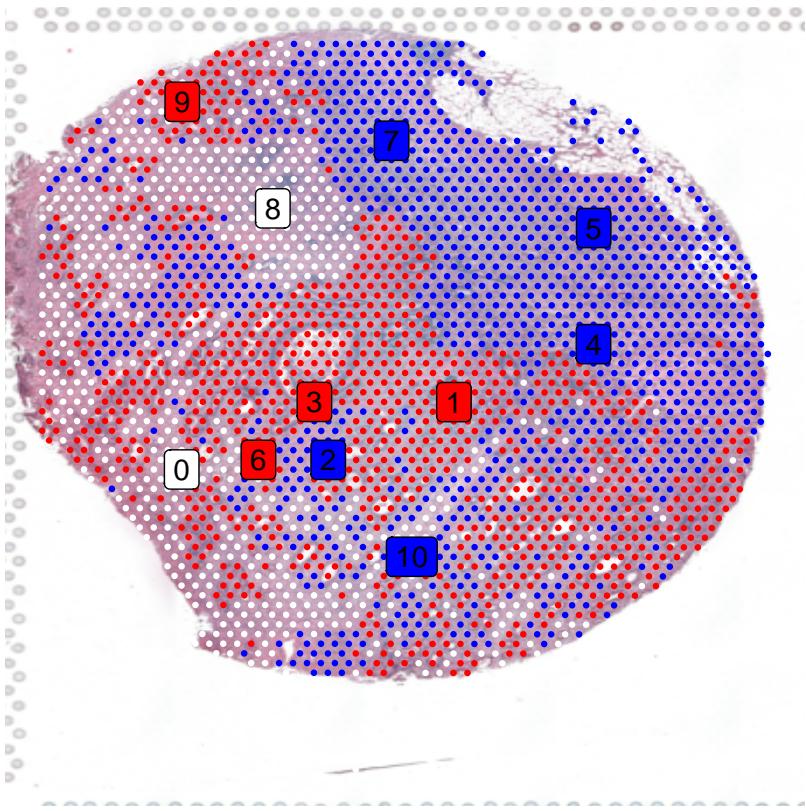
The markers expression gives us important hints on the cluster composition information. High *KRT20* expression in Clusters 1, 3, 6, and 9 indicates these are primarily CRC cells, with Cluster 3 as the main tumor mass. Low *ACTA2* expression in these clusters confirms their epithelial composition, with fewer fibroblasts.

On the other hand, high expression of *ACTA2* in Clusters 2, 4, 5, 7, and 10 indicates denote the presence of fibroblasts, typical of connective tissue.

Cancer associated fibroblasts are expected to be found in areas surrounding the tumor, forming a stromal barrier or support network, rather than within the tumor mass itself.

Positional Correlations between Markers' Expressions

```
SpatialDimPlot(cnc,
  label = TRUE, label.color = "black",
  label.size = 4) +
scale_fill_manual(values = c(
  "0" = "white",
  "1" = "red",
  "2" = "blue",
  "3" = "red",
  "4" = "blue",
  "5" = "blue",
  "6" = "red",
  "7" = "blue",
  "8" = "white",
  "9" = "red",
  "10" = "blue")) +
NoLegend()
```



```

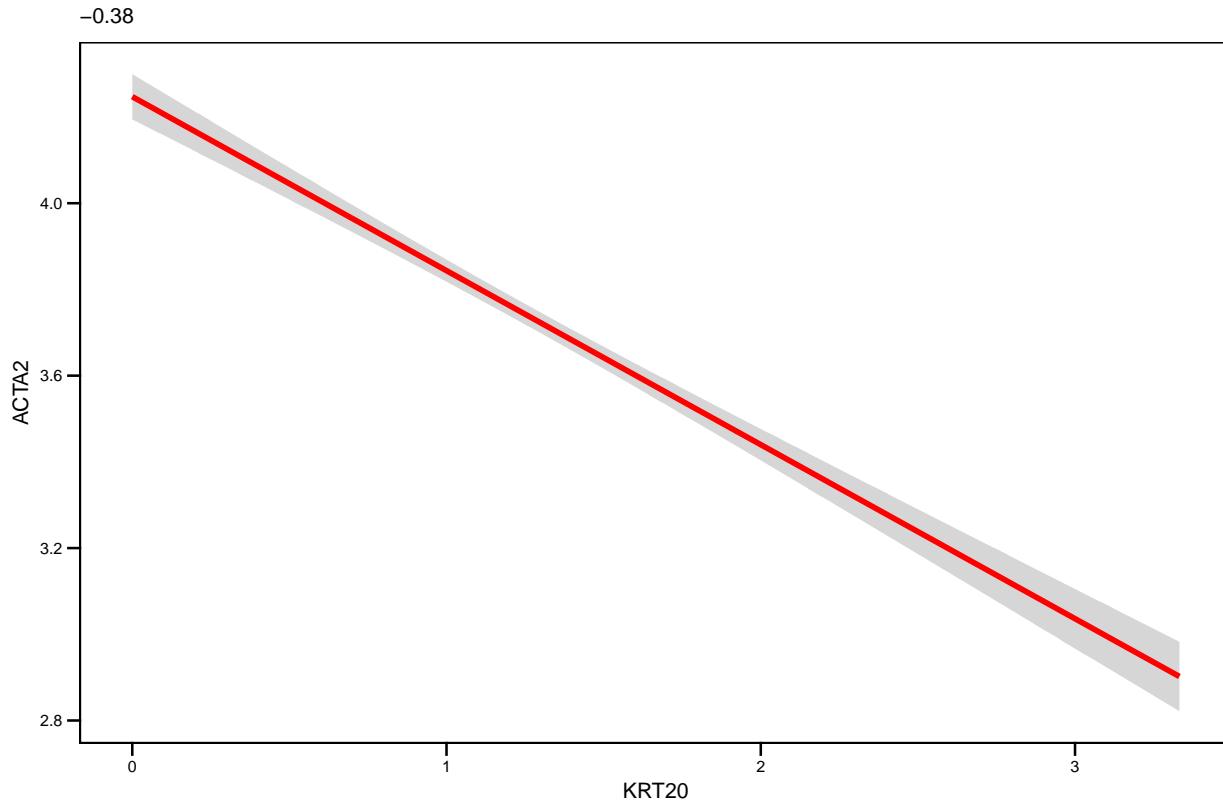
ggsave("./Images/Cancer Cluster Positioning.pdf")

# Test the correlation
corr_expr <- GetAssayData(cnc) %>%
  as.data.frame() %>%
  filter(rownames(.) %in% gene_markers) %>%
  t(.) %>%
  as.data.frame()

cor_krt_acta <- cor(corr_expr$KRT20, corr_expr$ACTA2, method = "pearson")
ggplot(corr_expr, aes(x = KRT20, y = ACTA2)) +
  geom_smooth(method = "lm", color = "red", se = TRUE) +
  labs(title = "Person Correlation KRT20 - ACTA2",
       subtitle = round(cor_krt_acta, 2)) +
  theme_publication()

```

Person Correlation KRT20 – ACTA2



```
ggsave("./Images/Gene Markers Correlations.pdf")
```

In the plot above the clusters have colored based on their tissue composition:

- Red cells -> high *KRT20* - low *ACTA2* expression
- Blue cells -> low *KRT20* - high *ACTA2* expression
- White cells -> both *KRT20* and *ACTA2* highly expressed

The negative correlation between *KRT20* and *ACTA2* (e.g., low *ACTA2* in Clusters 3 and 9) reflects the distinct cellular compartments in the tumor microenvironment. *KRT20* is expressed by epithelial tumor cells, while *ACTA2* is expressed by stromal fibroblasts. In CRC, the tumor core (high *KRT20*) is typically devoid of CAFs, which are more abundant in the surrounding stroma (e.g., Clusters 2, 4, 10). This segregation is due to the epithelial origin of CRC cells versus the mesenchymal origin of fibroblasts. The lack of *ACTA2* in *KRT20*-high regions suggests that CAFs are excluded from the tumor mass, instead forming a desmoplastic stroma that supports tumor growth indirectly. This separation is a hallmark of CRC, where stromal fibroblasts enhance tumor invasiveness but do not co-localize with cancer cells. (Jia et al. 2025)

Immune Markers Analysis

Expression Comparison between Control and Cancer Tissues

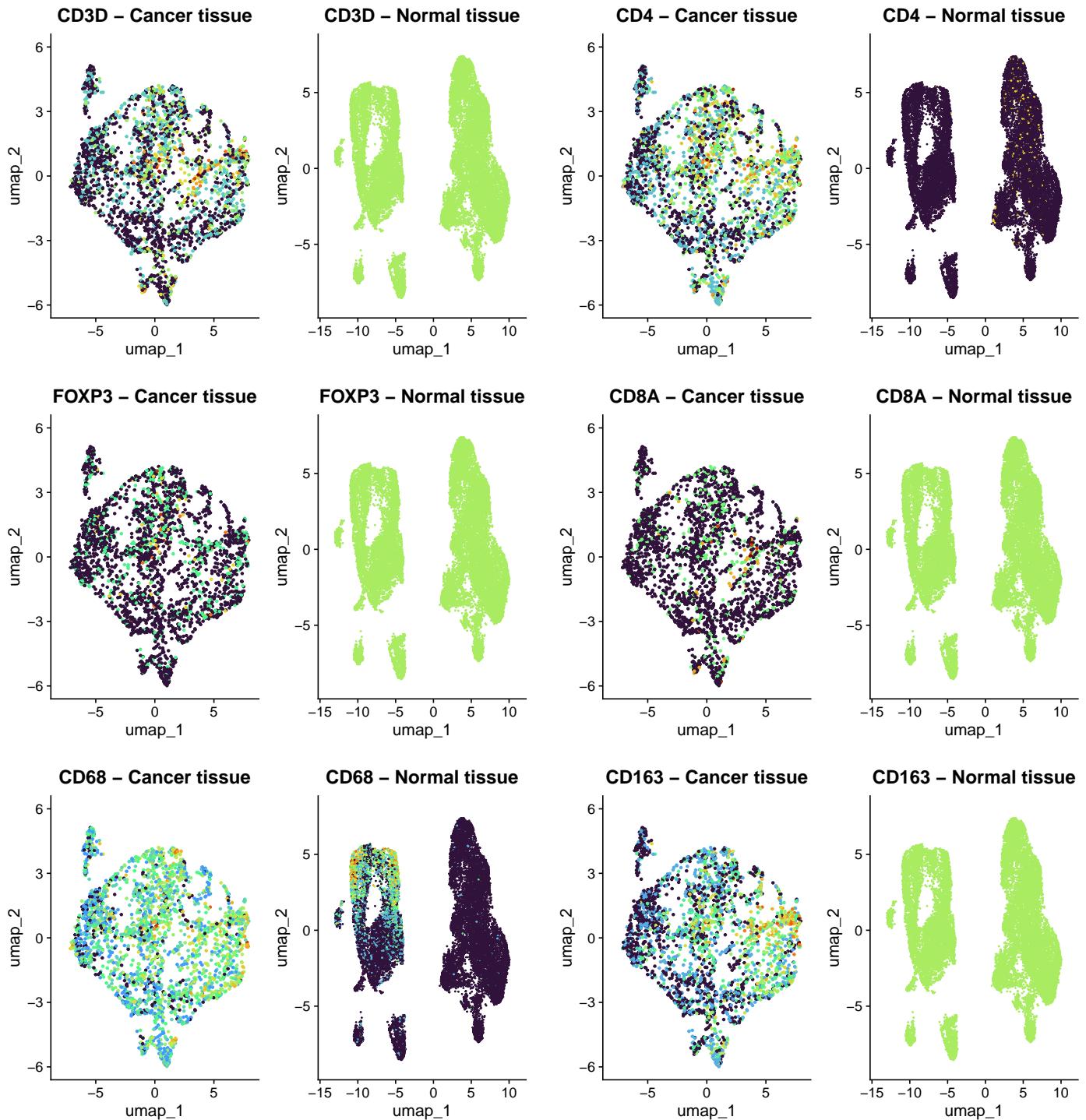
```
for (mrk in immune_markers) {  
  cnc_plot_tmp <- FeaturePlot(cnc, features = mrk) &  
  scale_color_viridis_c(option = "H") &  
  labs(title = paste0(mrk, " - Cancer tissue")) &  
  NoLegend()
```

```

crt_plot_tmp <- FeaturePlot(ctr, features = mrk) &
  scale_color_viridis_c(option = "H") &
  labs(title = paste0(mrk, " - Normal tissue")) &
  NoLegend()

plot(cnc_plot_tmp | crt_plot_tmp)
ggsave(filename = paste0("./Images/Expressions/", mrk, ".pdf"))
}

```

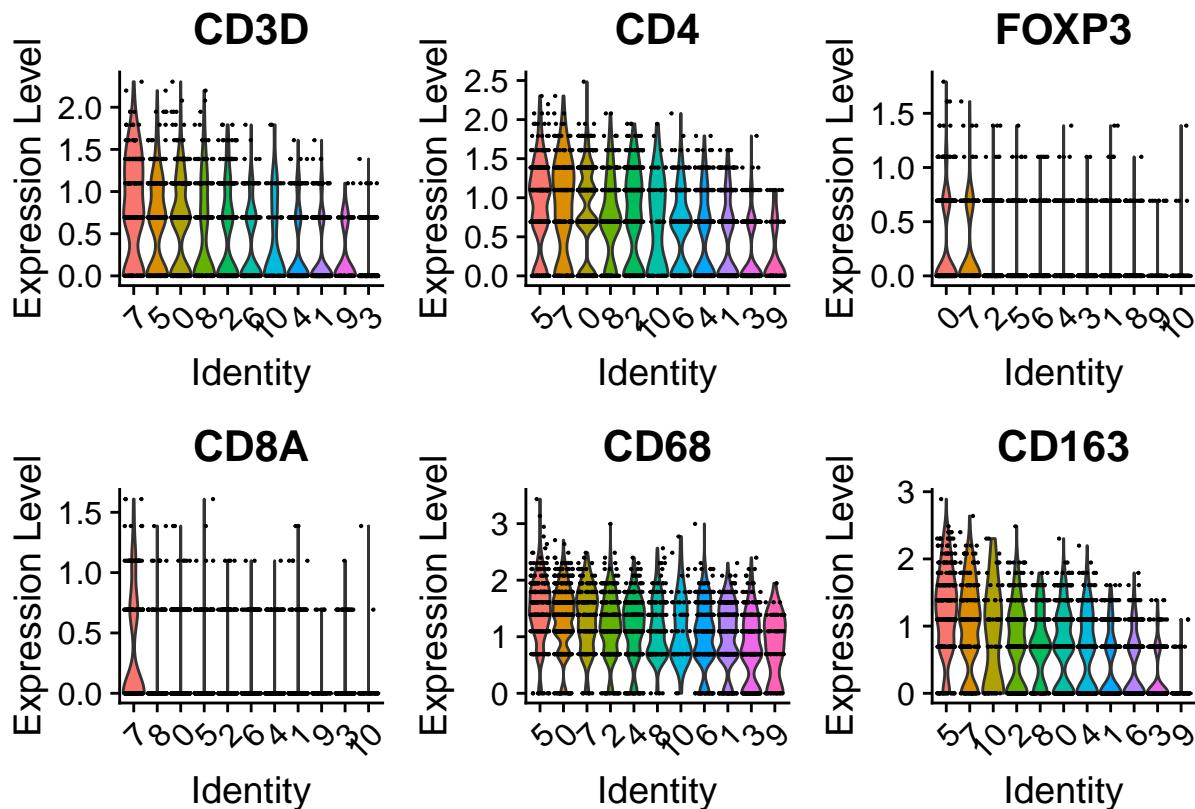


Most immune markers (*CD3D*, *CD4*, *FOXP3*, *CD20*, *CD279*, *CD163*, *CD8A*) are absent in healthy tissue but present in cancer tissue, reflecting an altered immune environment. In particular:

- *CD3D*, *CD4*, *FOXP3*, *CD8A*: T-cell markers indicate some immune infiltration, but low *CD8A* expression suggests a possibly immune-cold tumor, i.e. tumors that do not respond to immunotherapy, lacking of neo-antigens for T-cells recognition;
- *CD68*: presence in some areas of the epithelial fibroblast tissue, likely a “surveillance” area. Much higher expression in all the cancer tissue’s clusters;
- *CD163*: presence of M2-like macrophages, indicating a tumor-driven immuno suppression, while healthy epithelium doesn’t recruit *M2* macrophages

Clusters for Immune-Specific Markers in Cancer Tissue

```
VlnPlot(cnc, features = immune_markers, sort = "increasing")
```



```
ggsave("./Images/Immune Markers Violin.pdf")
```

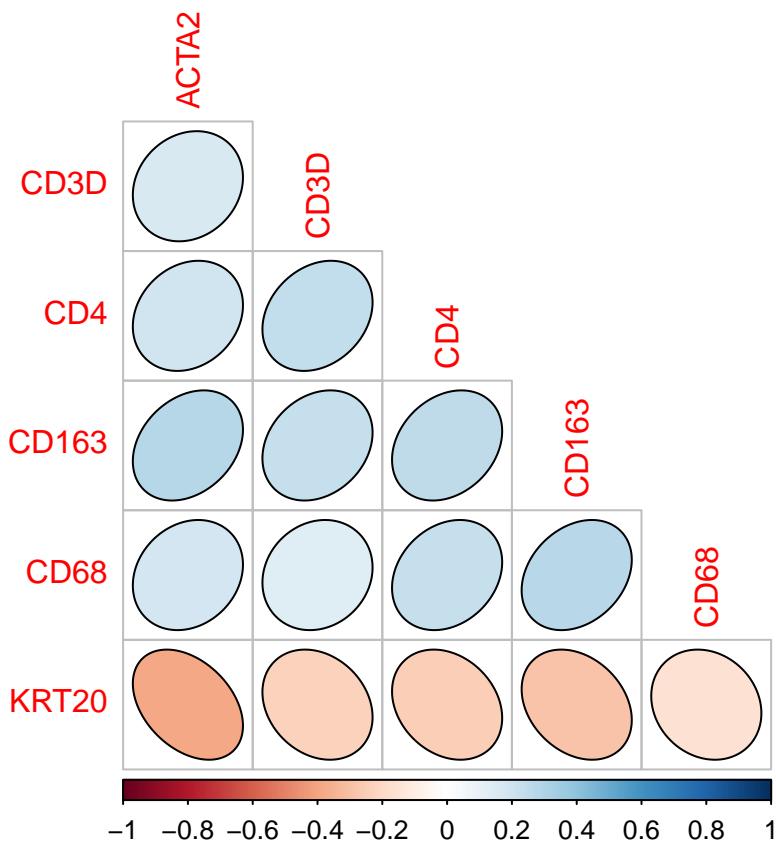
Contrary to the previous tissue-specific markers, immunological ones don't have a strong expression in some specific clusters but they exhibit a moderate expression in clusters 7, 5, and 0. These results could indicate that there is a small infiltration of pan- and helper-T cells in non-tumor regions, which is concordant with the previous results (clusters with the lowest tumor content). Additionally, since these clusters are located in the external areas of the sample, it could be that these regions are formed by mixed cells (tumor and healthy).

Not finding almost any immune markers in the tumor area could possibly mean that this is a cold-tumor or a relatively newly formed tumor, that has yet to activate a strong immune response. The latter hypothesis seems less likely, as the tumor mass appears large. Longitudinal data would help in this sense to compare immune markers expression over time.

Correlation between low immune expression and tumor (*KRT20*) / stromal (*ACTA2*) dominance.

```
corr_imm_expr <- GetAssayData(cnc) %>%
  as.data.frame() %>%
  filter(rownames(.) %in% c("KRT20", "ACTA2", "CD3D", "CD4", "CD68", "CD163")) %>%
  t(.) %>%
  as.data.frame()

cor_imm_matrix <- cor(corr_imm_expr, method = "pearson")
corrplot(cor_imm_matrix, method = "ellipse",
         type = "lower", diag = F, is.corr = T, outline = T)
```



```
pdf("./Images/Immune Markers Correlations.pdf")
corrplot(cor_imm_matrix, method = "ellipse",
         type = "lower", diag = F, is.corr = T, outline = T)
dev.off()
```

```
## pdf
## 2
```

A weak negative correlation between *KRT20* and immune markers (*CD3D*, *CD4*, *FOXP3*, *CD8A*, *CD68*, *CD163*) indicates immune markers are less expressed in high tumor regions (Clusters 3, 9). While a weak positive correlation between *ACTA2* and every immune marker suggests some immune presence in stromal regions (Clusters 2, 4).

This suggests a layered tumor microenvironment: the tumor epithelial core (Clusters 3, 9) is highly immune-excluded, possibly due to chemical or physical barriers. The surrounding stroma (Clusters 2, 4, 5, 10) acts as an immune buffer zone, where some immune cells, especially pro-tumorigenic M2 macrophages, are present [@CD163].

Alternatively, stromal tissue may interact with T cells, but this requires further verification.

Bibliography

- Gu, Jun Jie, Zhao Sun, Lin Zhao, and Chun Mei Bai. 2019. “Expression of α -Smooth Muscle Actin in Advanced Colorectal Cancer Tissue and Its Significance in Chemotherapy Response Prediction and Prognosis.” *Zhongguo Yi Xue Ke Xue Yuan Xue Bao* 41 (1): 63–67.
- Hou, Siyu, Yuanchun Zhao, Jiajia Chen, Yuxin Lin, and Xin Qi. 2024. “Tumor-Associated Macrophages in Colorectal Cancer Metastasis: Molecular Insights and Translational Perspectives.” *J. Transl. Med.* 22 (1): 62.
- Jia, Hongyuan, Xingmin Chen, Linling Zhang, and Meihua Chen. 2025. “Cancer Associated Fibroblasts in Cancer Development and Therapy.” *Journal of Hematology & Oncology* 18 (1): 36. <https://doi.org/10.1186/s13045-025-01688-0>.
- Kuwahara, Taichi, Shoichi Hazama, Nobuaki Suzuki, Shin Yoshida, Shinobu Tomochika, Yuki Nakagami, Hiroto Matsui, et al. 2019a. “Intratumoural-Infiltrating CD4 + and FOXP3 + T Cells as Strong Positive Predictive Markers for the Prognosis of Resectable Colorectal Cancer.” *Br. J. Cancer* 121 (8): 659–65.
- , et al. 2019b. “Intratumoural-Infiltrating CD4 + and FOXP3 + T Cells as Strong Positive Predictive Markers for the Prognosis of Resectable Colorectal Cancer.” *Br. J. Cancer* 121 (8): 659–65.
- Spisak, Sandor, David Chen, Pornlada Likasitwatanakul, Paul Doan, Zhixin Li, Pratyusha Bala, Laura Vizkeleti, et al. 2024. “Identifying Regulators of Aberrant Stem Cell and Differentiation Activity in Colorectal Cancer Using a Dual Endogenous Reporter System.” *Nat. Commun.* 15 (1): 2230.