Spatial Analysis of breast cancer tumor microenvironment

Aymen Maqsood

2025-01-23

# Spatial Data Analysis Report: Xenium Human Breast Cancer Dataset

Independent Project of analysis workflow performed to study the spatial transcriptomics of a human breast cancer dataset generated using the Xenium platform. The analysis follows standard steps like:

* **Data Preprocessing**
* **Normalization**
* **Dimensionality Reduction**
* **Clustering**
* **Cell Type Annotation**
* **Spatial Visualization**

## Key Findings

The analysis found distinct cell populations, with invasive carcinoma cells concentrated in specific areas, potentially indicating tumor boundaries. FAS expression was notably high in these invasive regions, surprising given its role in tumor metabolism, while CEACAM6 marked ductal carcinoma areas. Immune cells were scattered, suggesting infiltration into the tumor.

## Introduction

The Xenium platform is a spatial transcriptomics technology that allows for the simultaneous measurement of gene expression and spatial location of cells in a tissue section. The dataset contains gene expression data from thousands of cells, as well as spatial information about the location of each cell in the tissue section. In this analysis, we will conduct a comprehensive analysis of the Dataset to identify cell types, spatial patterns, and marker genes associated with breast cancer.

**Data Source:** [10x Genomics Xenium In-Situ Preview Dataset](https://www.10xgenomics.com/products/xenium-in-situ/preview-dataset-human-breast)

### Loading Packages

library(Seurat)  
library(tidyverse)  
library(SeuratObject)

### 

### Loading Data

xenium.obj <- LoadXenium(  
 "~/Desktop/Projects/V1.0/Spatial-scRNA-Seq/Xenium Human Breast Cancer Dataset/data/outs/",  
 fov = "fov",   
 assay = "Xenium"  
)

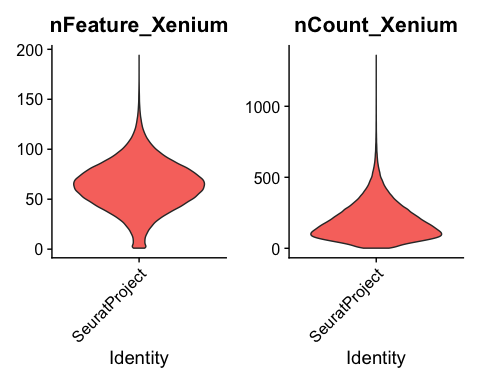
### 

### Data Preprocessing / Quality Control

Initial preprocessing removes low-quality cells and visualizes key metrics:

* **Remove cells with zero counts:** Ensures only cells with detectable transcripts are analyzed.
* **Visualize distributions:** Violin plots display genes per cell (nFeature\_Xenium) and transcript counts per cell (nCount\_Xenium).
* **Filtering:** Cells are subsetted to retain those with 5–200 features and 10–1000 counts, reducing noise from low-quality cells or outliers.

# remove cells with 0 counts  
xenium.obj <- subset(xenium.obj, subset = nCount\_Xenium > 0)  
  
# Genes/cell (nFeature\_Xenium) and transcript counts/cell (nCount\_Xenium)  
VlnPlot(xenium.obj, features = c("nFeature\_Xenium", "nCount\_Xenium"),   
 ncol = 2, pt.size = 0)



# Save the plot  
ggsave("VlnPlot\_QC.png", width = 8, height = 4)  
  
# 166363 cells X 313 genes , initially reduced to quality cells  
xenium.obj <- subset(xenium.obj,   
 subset = nFeature\_Xenium > 5 & nFeature\_Xenium < 200 &   
 nCount\_Xenium > 10 & nCount\_Xenium < 1000)

### 

### Normalization and Scaling

To correct for technical variations (e.g., sequencing depth), we apply SCTransform, a variance-stabilizing normalization method that accounts for gene expression dependencies on sequencing depth. note: The method also scales the data to account for differences in gene expression magnitude.

xenium.obj <- SCTransform(xenium.obj, assay = "Xenium")

### Dimensionality Reduction

Dimensionality reduction simplifies the dataset while preserving biologically relevant variation:

* **PCA:** Computes the top 30 principal components.
* **UMAP:** Projects the data into a 2D space for visualization.

# Run PCA  
xenium.obj <- RunPCA(xenium.obj, npcs = 30, features = rownames(xenium.obj))  
  
# Run UMAP  
xenium.obj <- RunUMAP(xenium.obj, dims = 1:30)

### Clustering

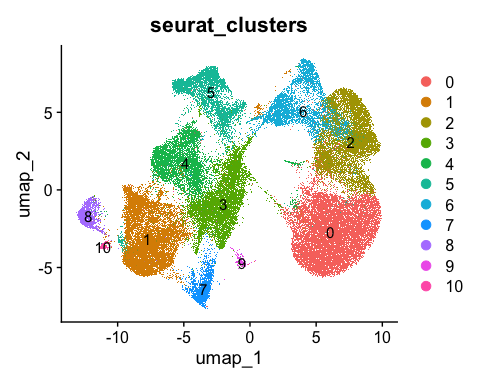
groupeing cells based on shared expression profiles:

* **Find Neighbors:** Uses PCA-reduced data to identify cell neighbors.
* **Find Clusters:** Applies a resolution of 0.2 to define clusters, visualized with a UMAP plot.

# Find neighbors  
xenium.obj <- FindNeighbors(xenium.obj, reduction = "pca", dims = 1:30)  
  
# Find clusters  
xenium.obj <- FindClusters(xenium.obj, resolution = 0.2)

## Modularity Optimizer version 1.3.0 by Ludo Waltman and Nees Jan van Eck  
##   
## Number of nodes: 163779  
## Number of edges: 5297889  
##   
## Running Louvain algorithm...  
## Maximum modularity in 10 random starts: 0.9488  
## Number of communities: 11  
## Elapsed time: 100 seconds

# Visualize clusters  
DimPlot(xenium.obj, group.by = "seurat\_clusters", label = TRUE)



# Save the plot  
ggsave("Cluster\_UMAP.png", width = 6, height = 6)

### 

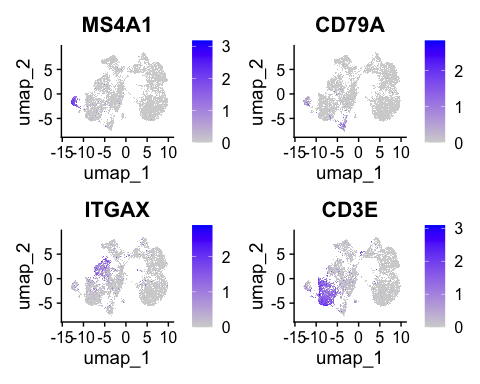
### Cell Type Annotation

The above Clusters are annotated using differential expression analysis and known marker genes:

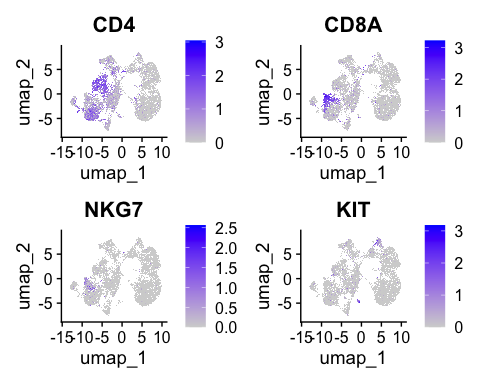
* **Find Markers:** Identifies genes differentially expressed in each cluster.
* **Visualize Markers:** Feature plots display expression of cell-type-specific genes.
* **Assign Identities:** Clusters are renamed based on marker expression.

**Marker Genes Used:** - **B cells:** MS4A1, CD79A - **Macrophages:** ITGAX - **T cells:** CD3E, CD4, CD8A - **NK cells:** NKG7 - **Mast cells:** KIT - **Endothelial cells:** PECAM1 - **Myoepithelial cells:** KRT15 - **Fibroblasts:** LUM - **Proliferating cells:** MKI67 - **Ductal carcinoma:** CEACAM6 - **Invasive tumor:** FASN

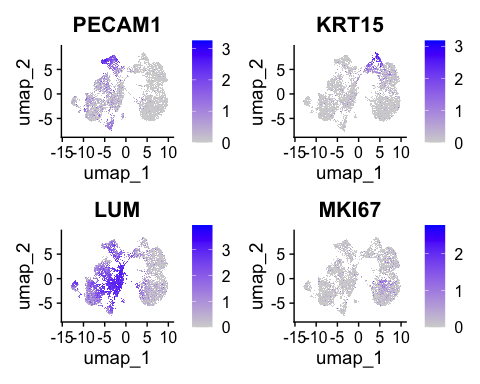
# Find marker genes for all clusters  
AllMarkers <- FindAllMarkers(xenium.obj)  
  
# Feature plots for key marker genes  
FeaturePlot(xenium.obj, features = c("MS4A1", "CD79A", "ITGAX", "CD3E")) # B cells, macrophages, T cells



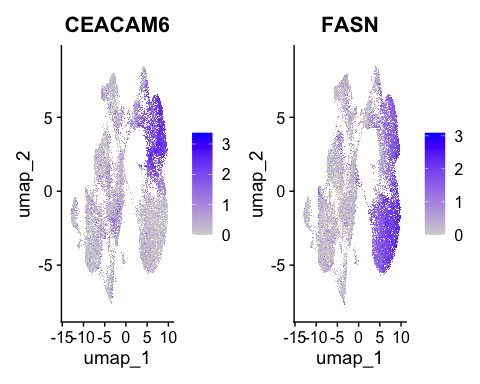
ggsave("FeaturePlot\_Immune.png", width = 8, height = 8)  
  
FeaturePlot(xenium.obj, features = c("CD4", "CD8A", "NKG7", "KIT")) # T cells, NK cells, Mast cells



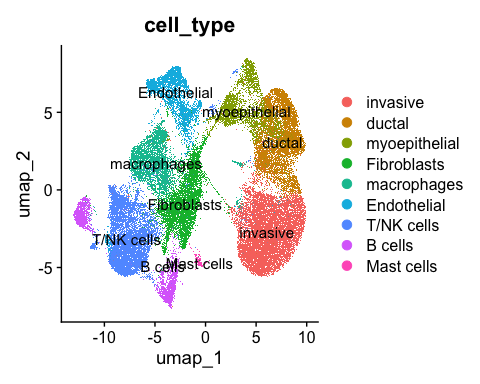
ggsave("FeaturePlot\_Immune2.png", width = 8, height = 8)  
  
FeaturePlot(xenium.obj, features = c("PECAM1", "KRT15", "LUM", "MKI67")) # Endothelial, myoepithelial, fibroblasts, proliferating



ggsave("FeaturePlot\_Stroma.png", width = 8, height = 8)  
  
FeaturePlot(xenium.obj, features = c("CEACAM6", "FASN")) # Ductal carcinoma, invasive tumor



ggsave("FeaturePlot\_Tumor.png", width = 8, height = 8)  
  
# Rename clusters to cell types  
xenium.obj <- RenameIdents(xenium.obj,   
 `0` = "invasive", `2` = "ductal", `11` = "ductal", `6` = "myoepithelial",   
 `3` = "Fibroblasts", `4` = "macrophages", `5` = "Endothelial",   
 `1` = "T/NK cells", `10` = "T/NK cells", `7` = "B cells", `8` = "B cells",   
 `9` = "Mast cells")  
  
# Assign cell types to metadata  
xenium.obj$cell\_type <- Idents(xenium.obj)  
  
# Visualize cell types  
DimPlot(xenium.obj, group.by = "cell\_type", label = TRUE)



ggsave("CellType\_UMAP.png", width = 6, height = 6)

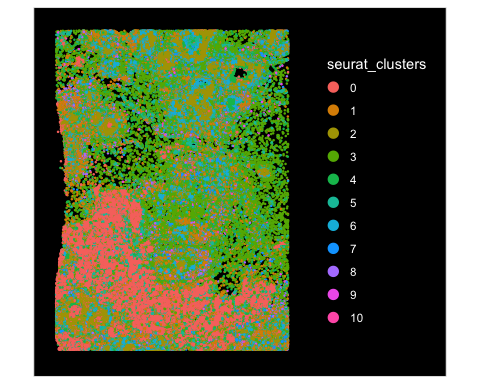
### 

### Spatial Visualization

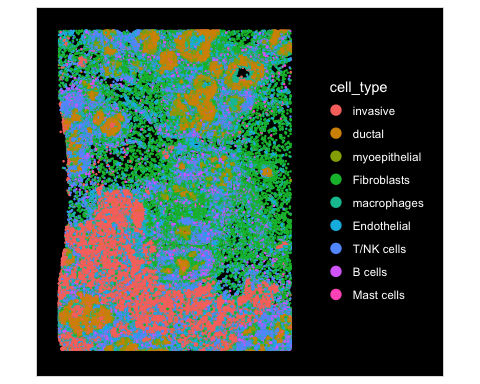
Spatial plots integrate gene expression with tissue coordinates:

* **Cluster/Cell Type Maps:** Spatial plots of clusters and cell types using ImageDimPlot.
* **Gene Expression Maps:** Gene expression maps for FAS and CEACAM6 using ImageFeaturePlot, with color scales (e.g., white to red for FAS, blue to red for CEACAM6).
* **Tumor Markers:** Tumor-specific marker genes overlaid on spatial coordinates.

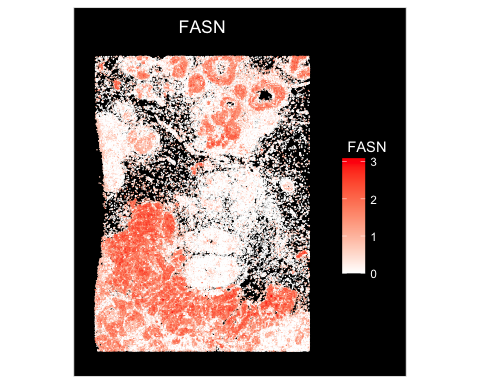
# Spatial plot of clusters  
ImageDimPlot(xenium.obj, group.by = "seurat\_clusters", size = 0.9)



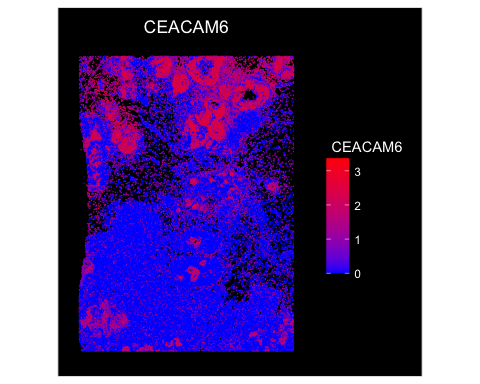
ggsave("Spatial\_Clusters.png", width = 6, height = 6)  
  
# Spatial plot of cell types  
ImageDimPlot(xenium.obj, group.by = "cell\_type", size = 0.9)



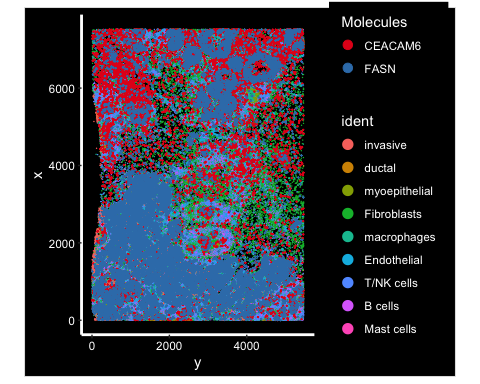
ggsave("Spatial\_CellTypes.png", width = 6, height = 6)  
  
# Gene expression spatial plots  
ImageFeaturePlot(xenium.obj, features = "FASN", cols = c("white", "red"))



ggsave("Spatial\_FASN.png", width = 6, height = 6)  
  
ImageFeaturePlot(xenium.obj, features = "CEACAM6", cols = c("blue", "red"))



ggsave("Spatial\_CEACAM6.png", width = 6, height = 6)  
  
# Tumor-specific marker genes on spatial coordinates  
ImageDimPlot(xenium.obj, fov = "fov", nmols = 20000, axes = TRUE,  
 molecules = c("CEACAM6", "FASN"))



ggsave("Spatial\_TumorMarkers.png", width = 6, height = 6)

## Results

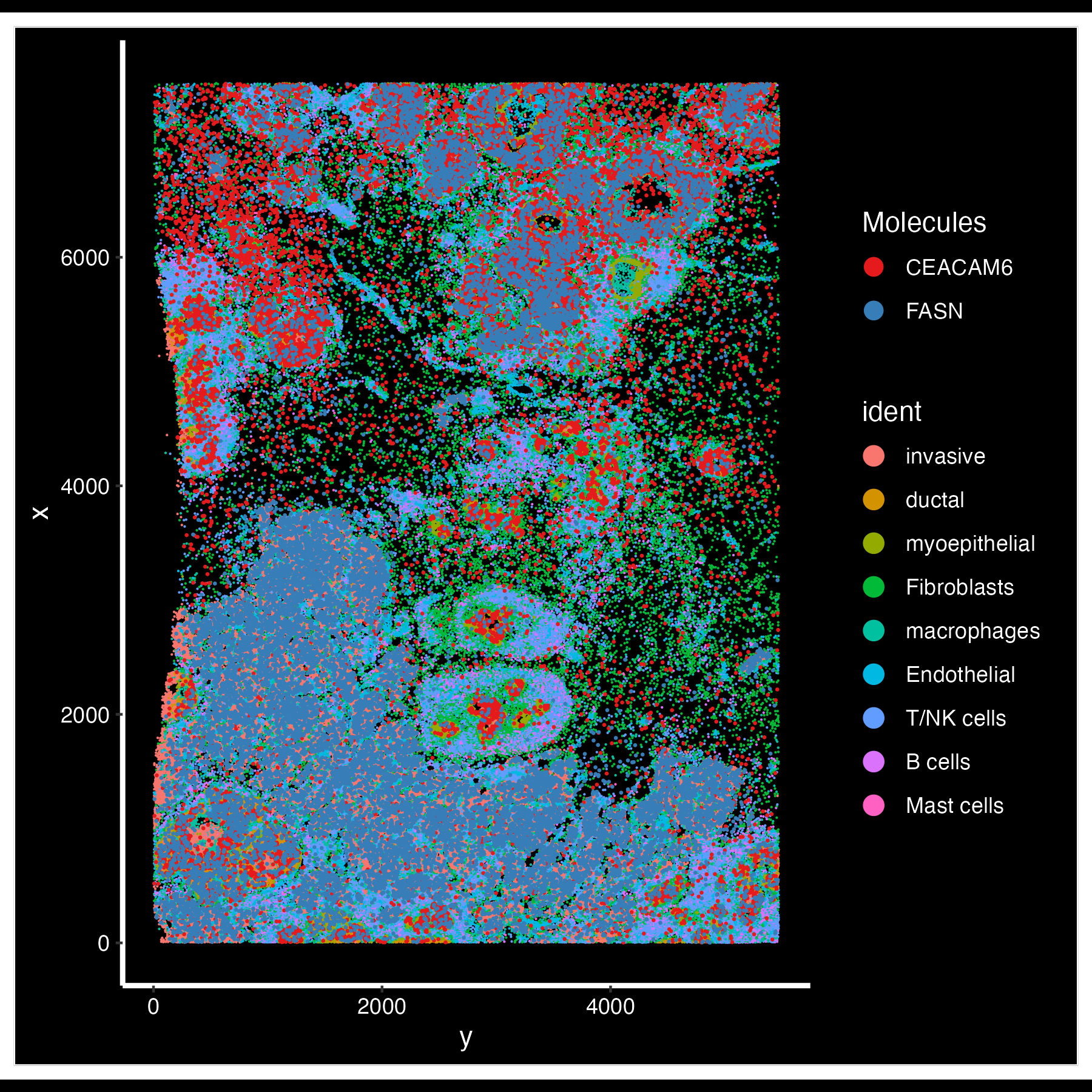
# Results

The analysis of the Xenium Human Breast Cancer Dataset provided key insights into cellular composition and spatial organization:

1. **Cell Type Identification**
   * Clustering identified distinct populations, annotated as invasive and ductal carcinoma cells, myoepithelial cells, fibroblasts, macrophages, endothelial cells, T/NK cells, B cells, and mast cells.
   * Marker gene expression validated these annotations (e.g., FASN for invasive tumors [1], CEACAM6 for ductal carcinoma [2]). See the table below for key markers:

| Cell Type | Marker Genes | Reference |
| --- | --- | --- |
| Invasive Tumor | FASN | Swinnen et al., 2006 [1] |
| Ductal Carcinoma | CEACAM6 | Blumenthal et al., 2007 [2] |
| B Cells | MS4A1, CD79A | Standard markers |
| Macrophages | ITGAX | Standard markers |
| T/NK Cells | CD3E, CD4, CD8A, NKG7 | Standard markers |
| Endothelial | PECAM1 | Standard markers |
| Myoepithelial | KRT15 | Standard markers |
| Fibroblasts | LUM | Standard markers |
| Mast Cells | KIT | Standard markers |
|  |  |  |

1. **Spatial Organization**
   * The spatial plot (Spatial\_TumorMarkers.png) revealed invasive carcinoma cells concentrated in peripheral regions, potentially marking tumor boundaries, while ductal carcinoma cells aligned with central ductal structures (see Figure 1).
   * Immune cells (macrophages, T/NK cells) were dispersed throughout the tissue, suggesting infiltration into the tumor microenvironment, consistent with immune surveillance roles [3].

* 
* Figure 1: Spatial distribution of tumor markers CEACAM6 (ductal) and FASN (invasive)

1. **Gene Expression Patterns**
   * FASN expression was elevated in invasive regions (Spatial\_FASN.png), aligning with its role in lipid metabolism supporting tumor growth [1].
   * CEACAM6 marked ductal carcinoma areas (Spatial\_CEACAM6.png), consistent with its association with epithelial-derived cancers [2].

These findings underscore the cellular diversity and spatial architecture of the breast cancer microenvironment, with implications for tumor progression and immune interactions.

## Conclusion

This spatial transcriptomics analysis demonstrates the utility of the Xenium platform in dissecting the breast cancer tumor microenvironment. By combining gene expression with spatial data, I identified key cell types and their distributions, offering insights into tumor-immune interactions and potential therapeutic targets.The elevated FAS expression in invasive regions underscores its metabolic role, suggesting avenues for targeting lipid metabolism in cancer therapy. Future work could integrate additional datasets or functional assays to validate these findings and explore clinical implications.

## References

1. Swinnen, J. V., et al. (2006). “Fatty acid synthase drives the growth of prostate cancer cells.” *Cancer Research*, 66(8), 3814-3820.
2. Blumenthal, R. D., et al. (2007). “Carcinoembryonic antigen (CEA) and CEACAM6 in cancer progression.” *Cancer Biology & Therapy*, 6(6), 831-837.
3. Hanahan, D., & Weinberg, R. A. (2011). “Hallmarks of cancer: The next generation.” *Cell*, 144(5), 646-674.