

DEEP LEARNING IN LIFE SCIENCE PROJECT



Team Arkusz5



EXPLORATORY DATA ANALYSIS

Part I. General data overview

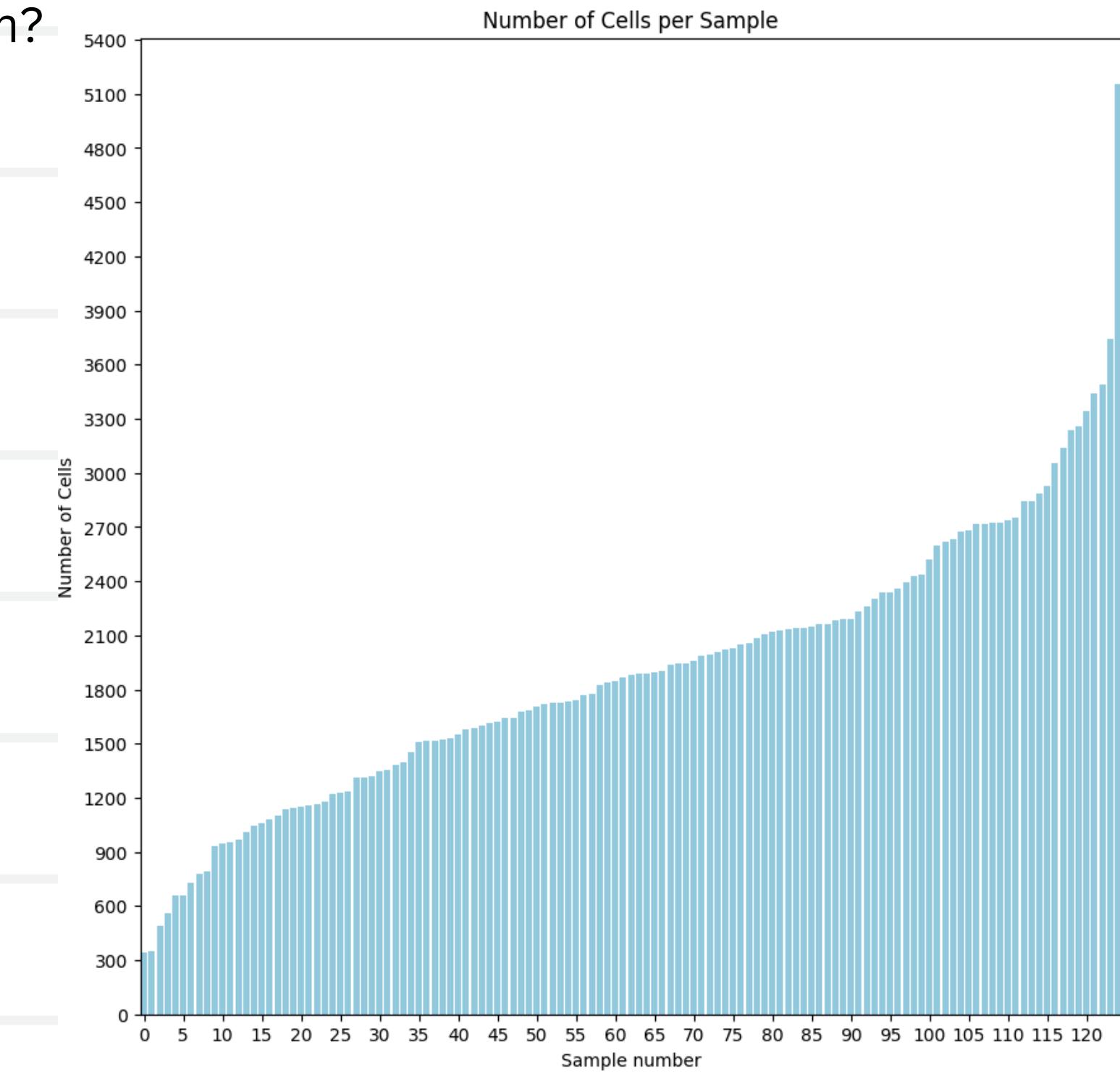
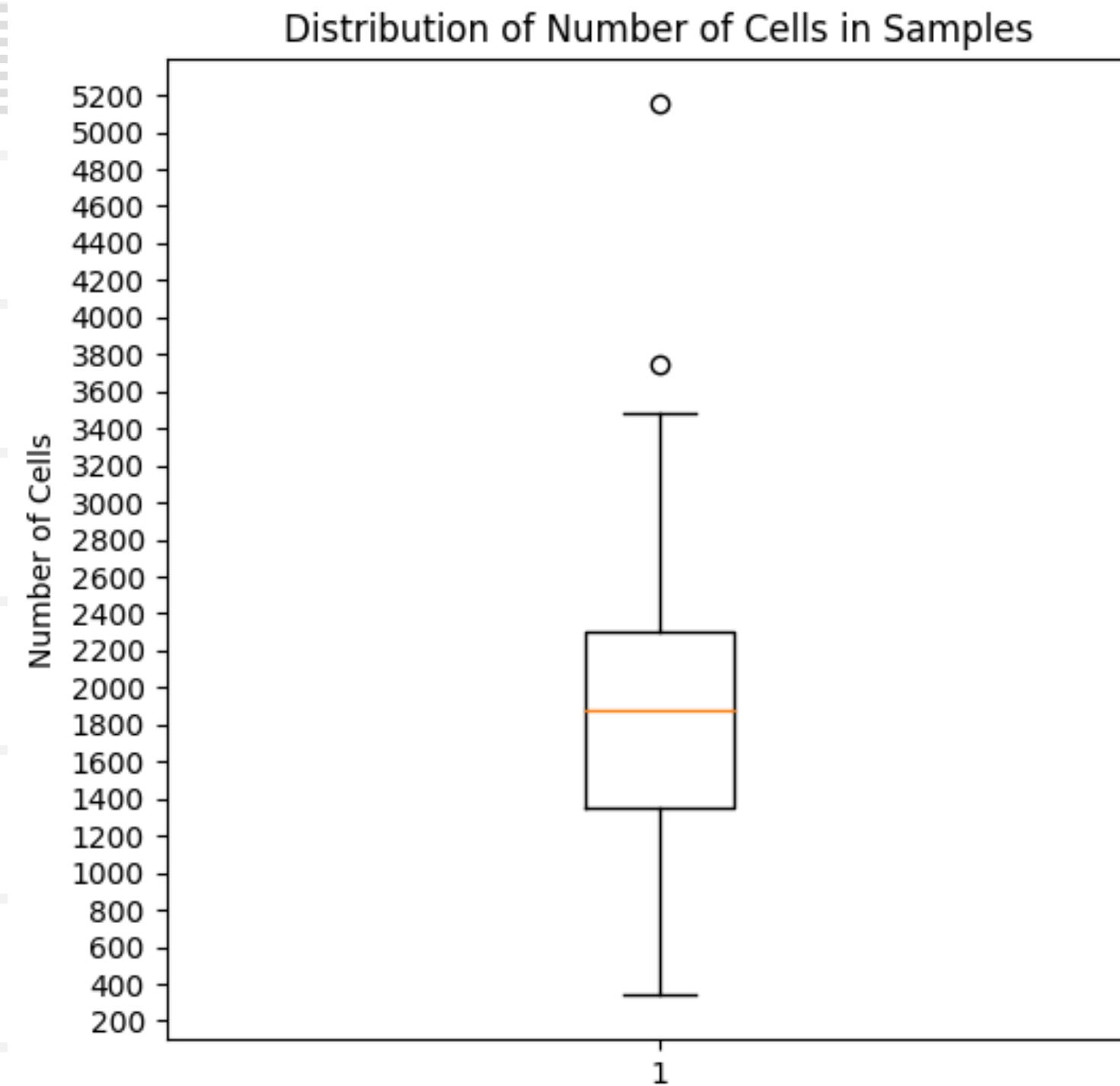
We have **236791** records of cell data with predefined labels

For each cell we have information about expression of **40** markers

There are **14** cell types: various immune cells and tumor

SAMPLES

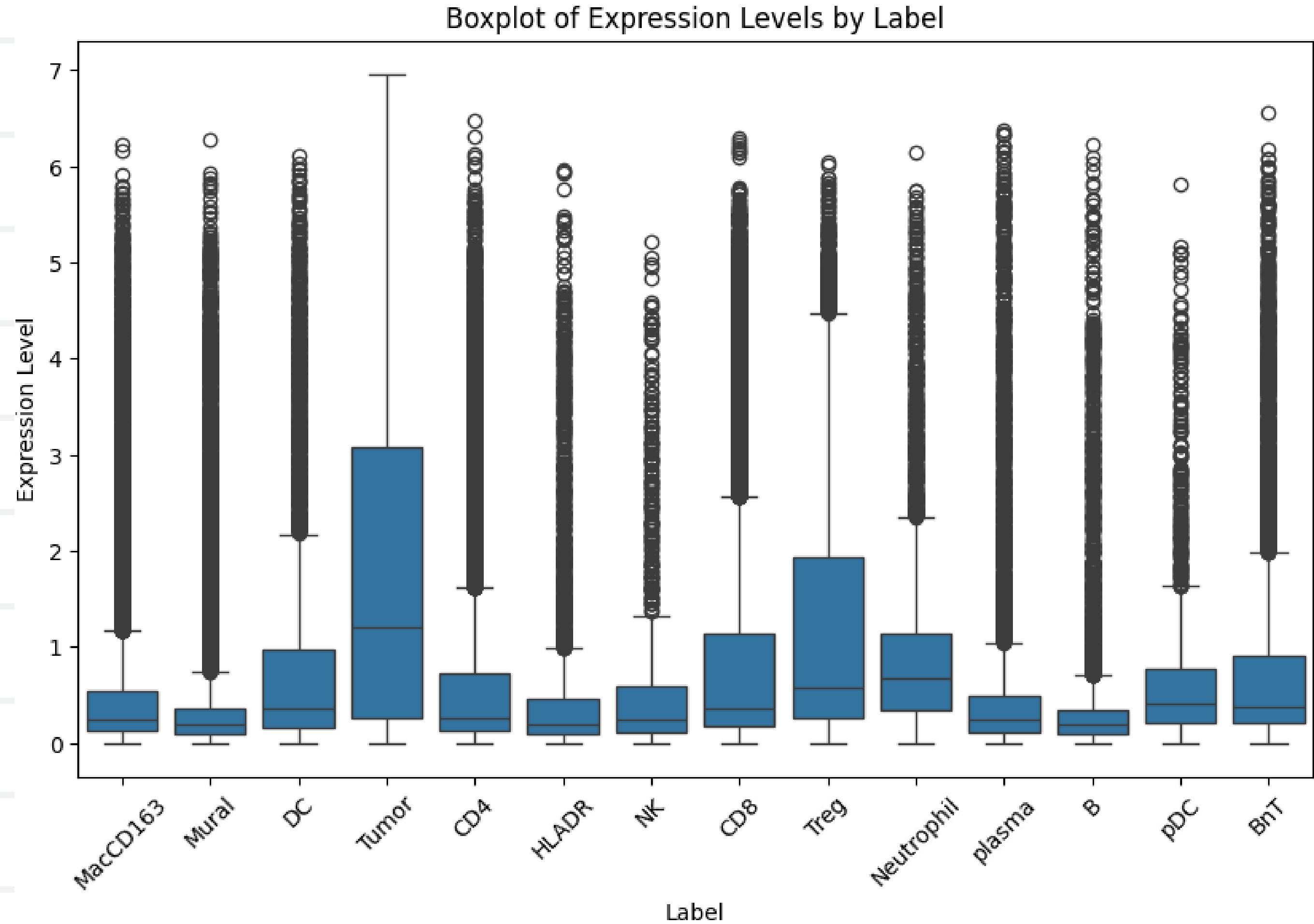
We have at our disposal 125 samples. How many cells are within them?



Sample with max number of cells (5152): IMMUcan_Batch20210921_LUNG_10041543-LUNG-
VAR-TIS-01-IMC-01_003

We wrote a script that takes as an input name of the marker and returns the profile of its expression in every cell type

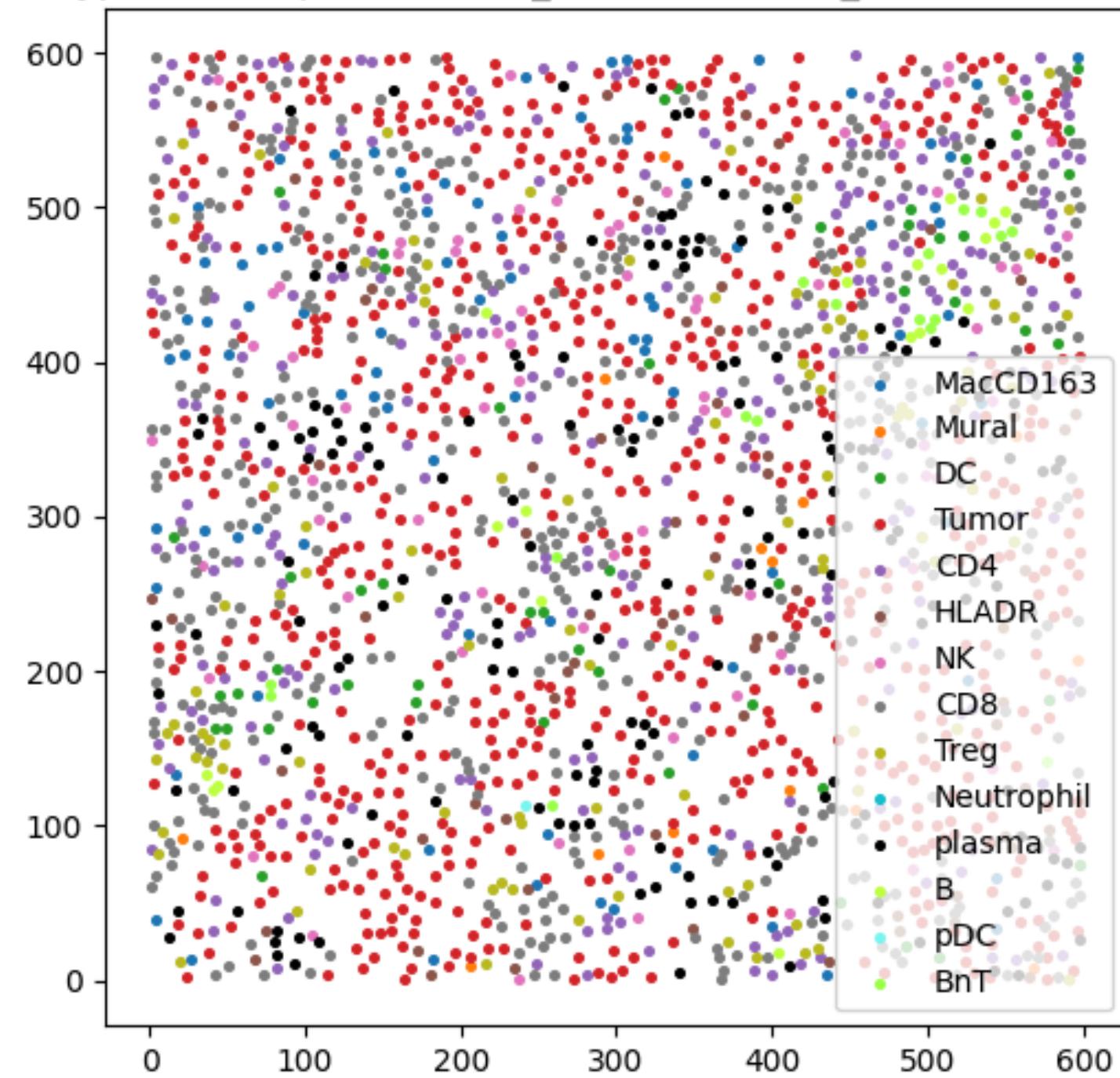
Example of its performance, with Ki67 marker as an input:



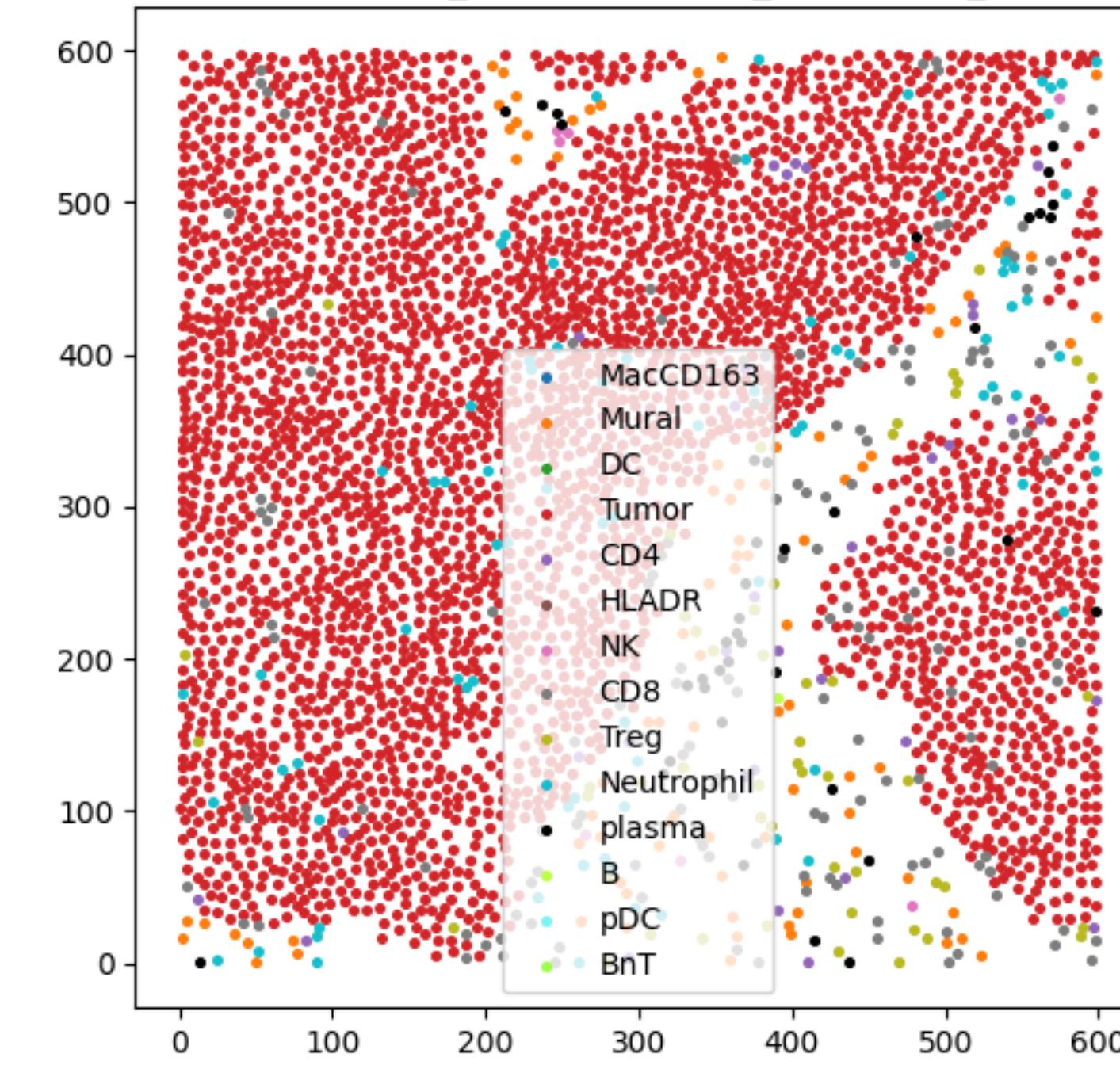
Cell's types distribution over samples

We grouped cells by samples and created scatterplots, showing all the cell types present within the sample, each cell type with different colour. Here are some interesting ones:

Cell Types of sample IMMUcan_Batch20220309_S-220114-00026_005

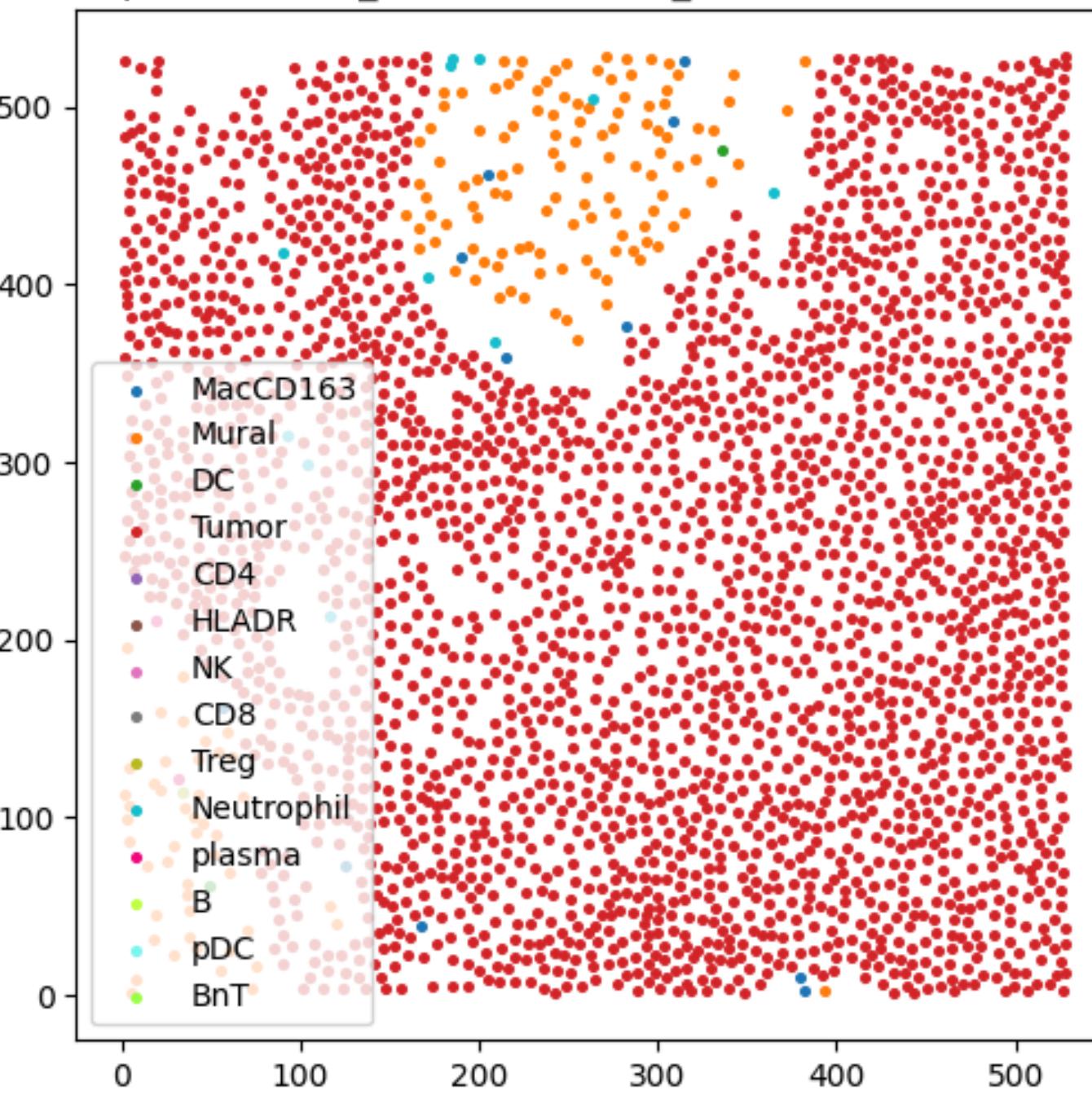


Cell Types of sample IMMUcan_Batch20210921_UPSTREAM_S-210622-00014_004

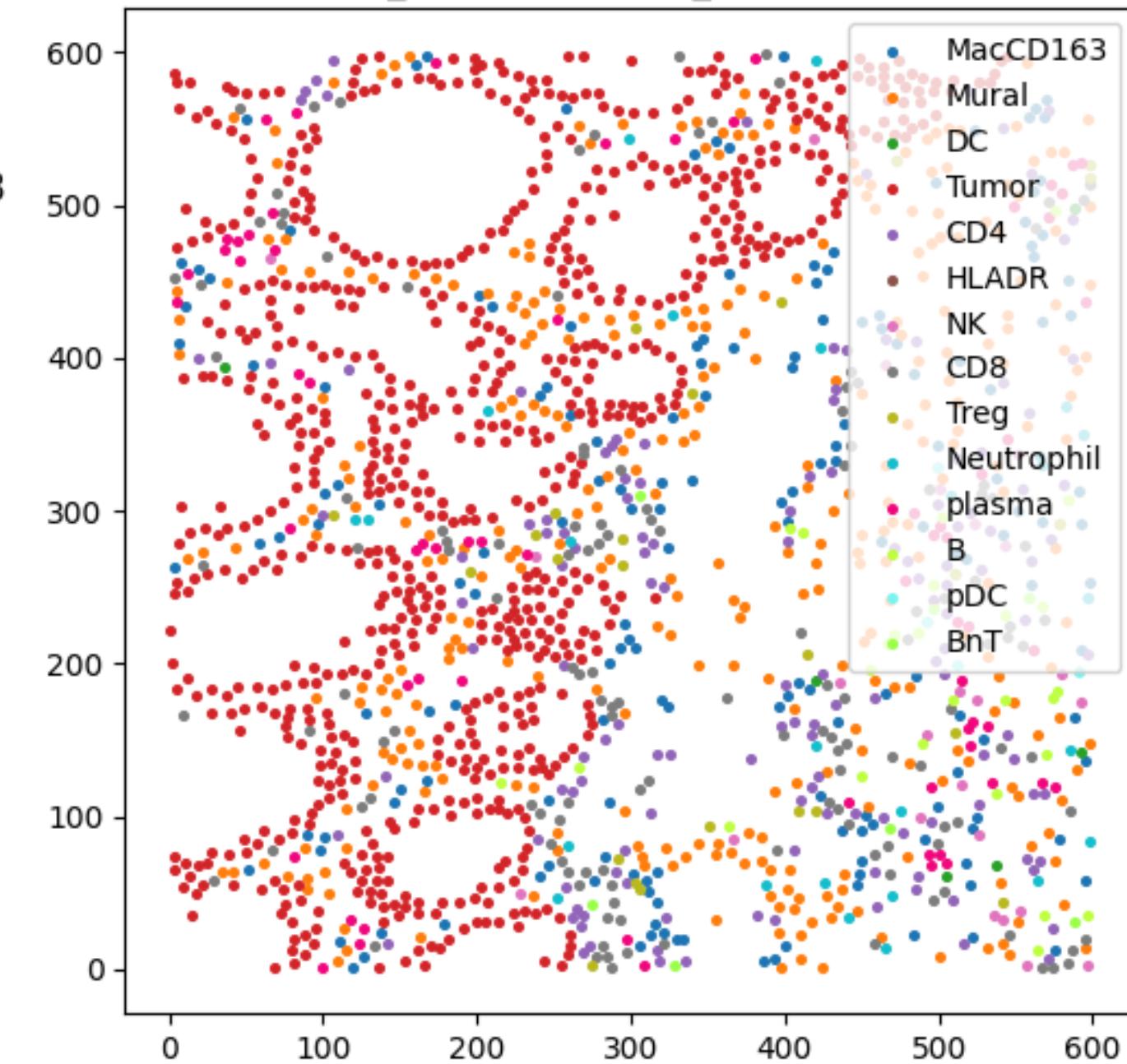


Here are the samples where tumor cells are not mixed with immune cells:

Cell Types of sample IMMUcan_Batch20191023_10032401-HN-VAR-TIS-01-IMC-01_003

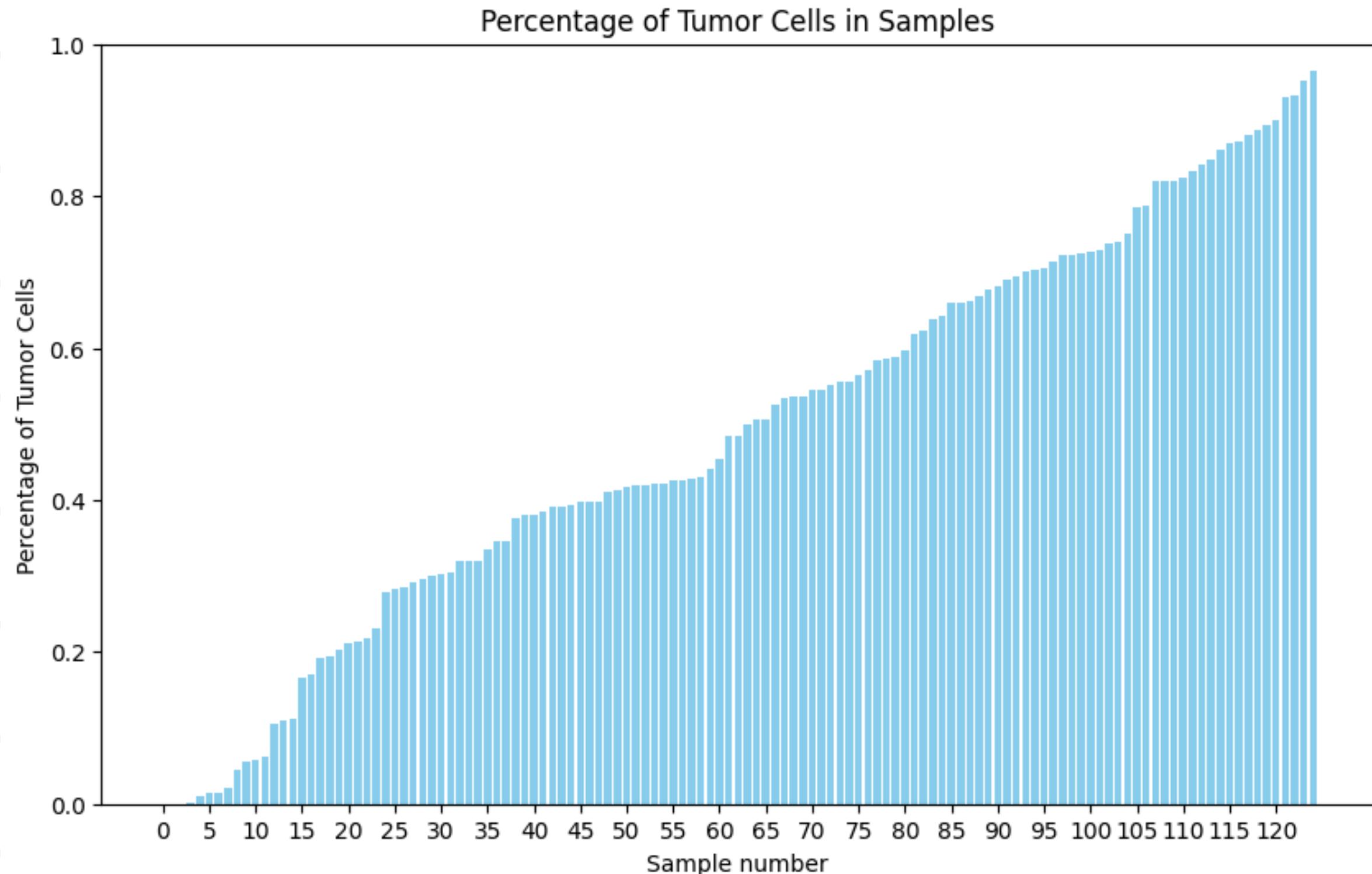


Cell Types of sample IMMUcan_Batch20210506_10074065-SPECT-VAR-TIS-UNST-03_001



Tumor cells

It turns out that 3 samples contain **no tumor** cells at all - two of them have the **smallest** number of observed cells, as well, which indicates the sample could have been too small to register the presence of the tumor cells.



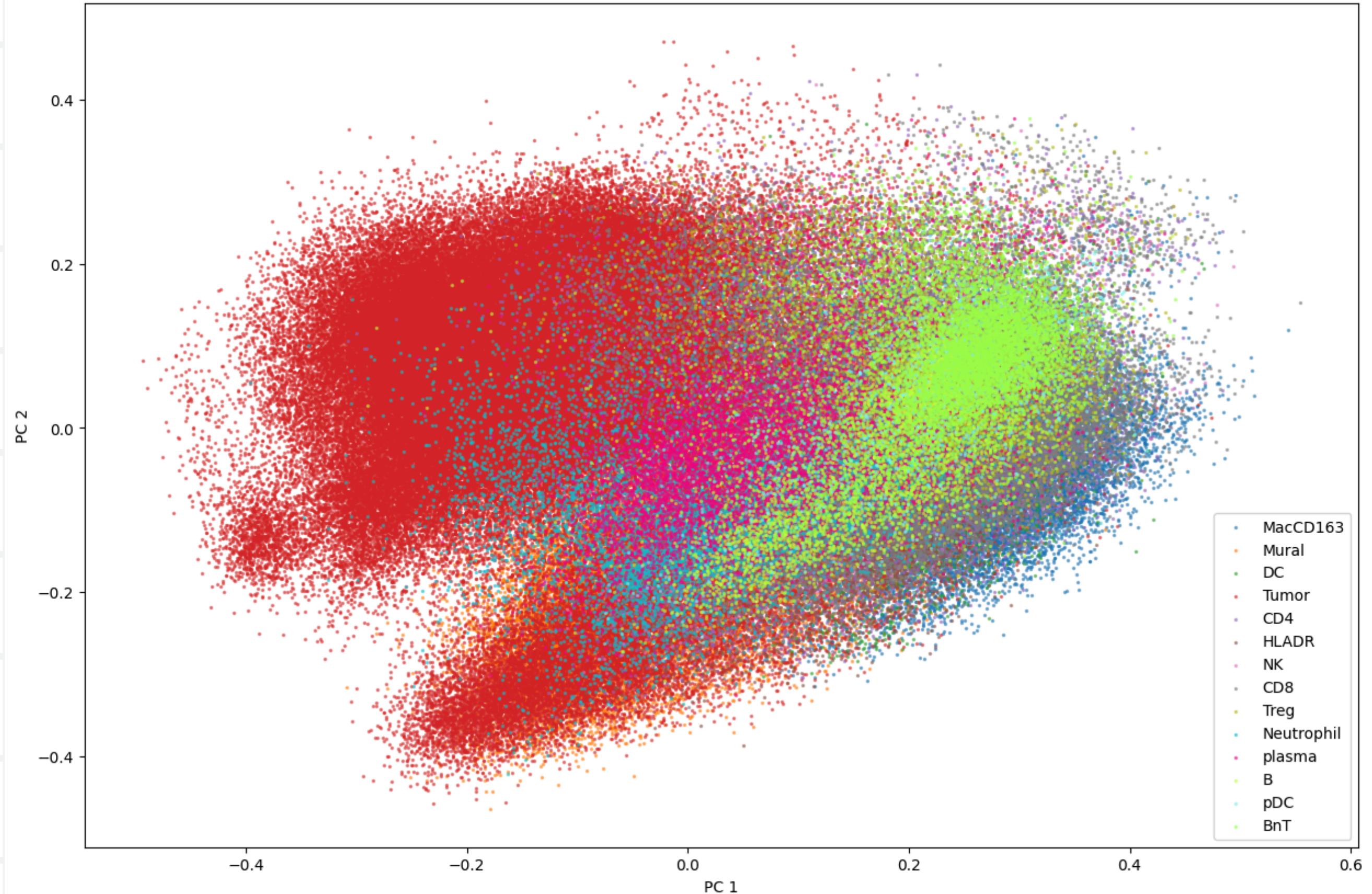
The **biggest percentage of tumor** cells in sample is 96.5062%, and this sample is also the **biggest** (in terms of cells' number) one. This results point out the need to consider total number of cells in sample when analysing their content

Principal Component Analysis (PCA)

PCA from 40-dimentional space of markers to 2 dimentions

Cells are mostly clustering by their cell type - we expect hight model accuracy

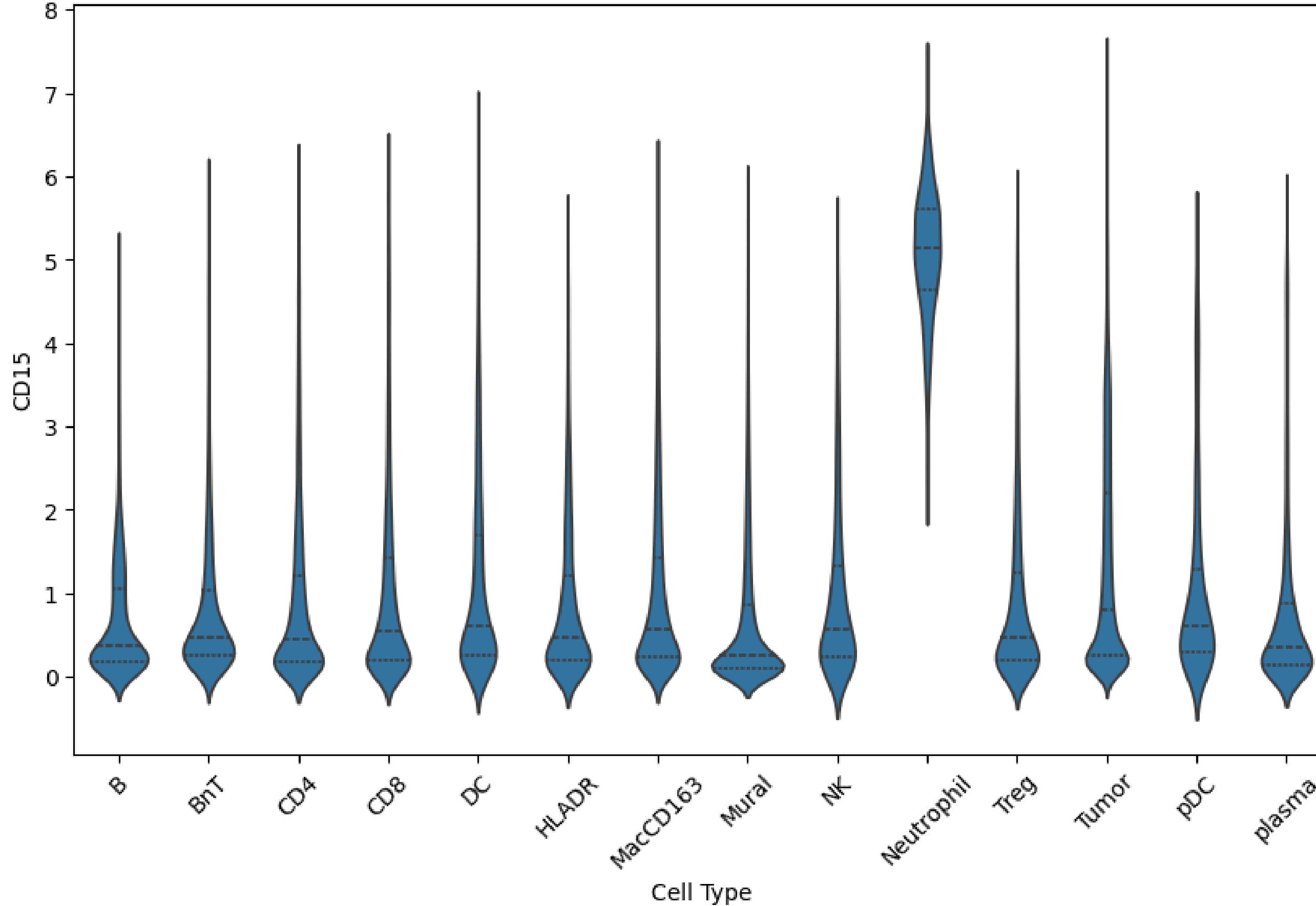
PCA Visualization



EXPLORATORY DATA ANALYSIS

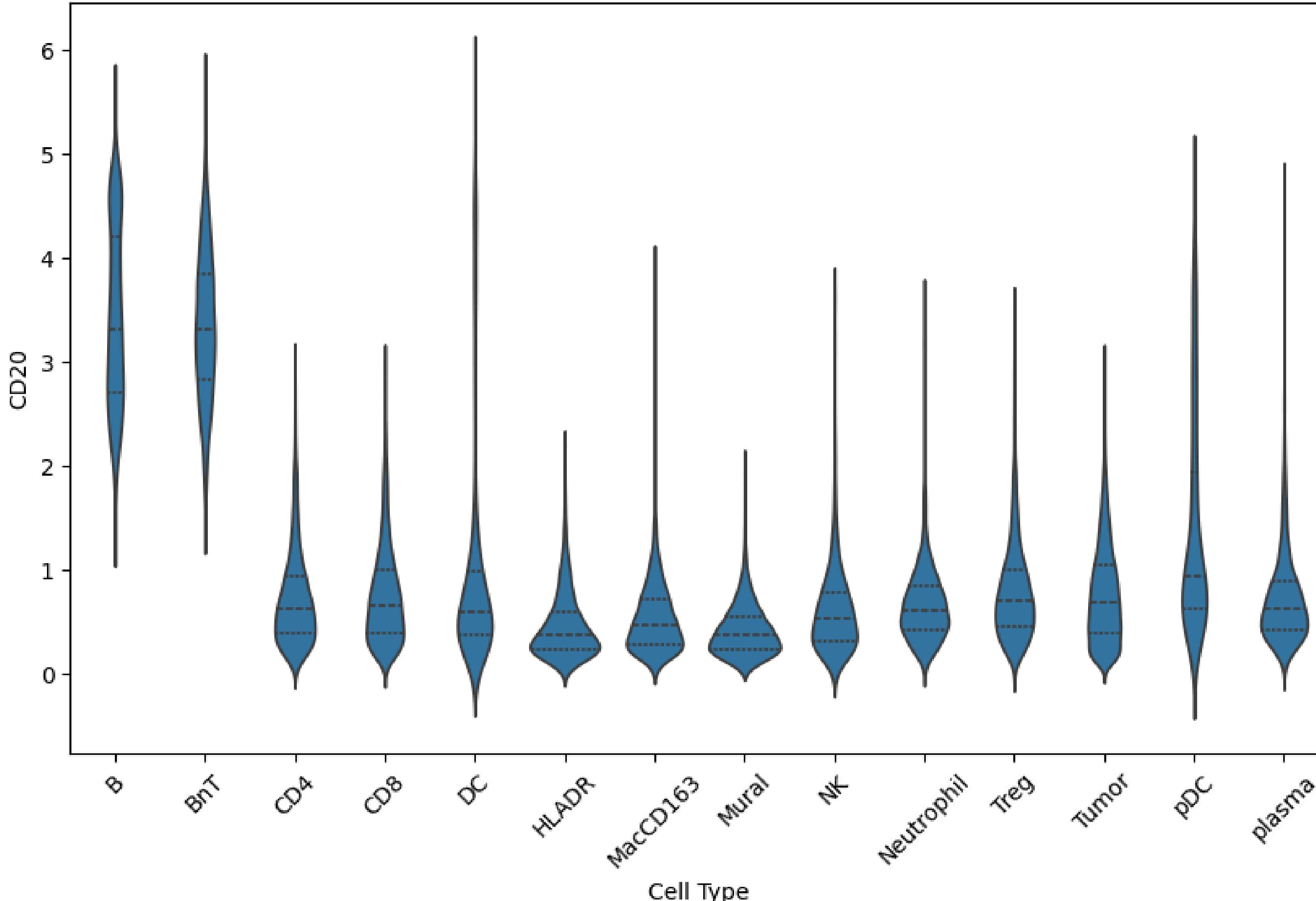
Part II.
Correlation patterns between markers and cell types

Violin Plot of CD15 across Cell Types



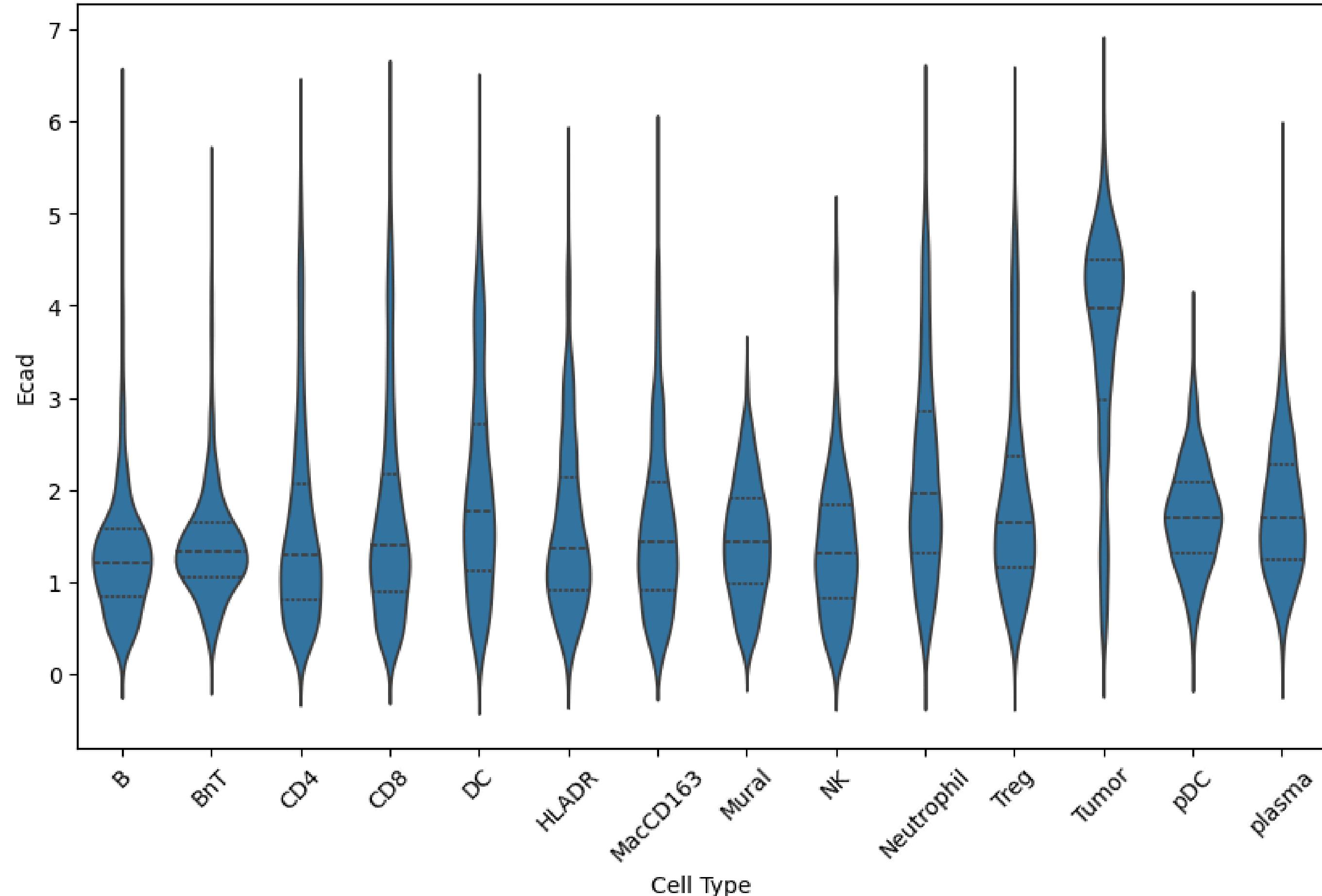
CD15, an antigen
expressed on the
surface of immature
neutrophils -
neutrophil marker

Violin Plot of CD20 across Cell Types



**CD20 - marker for
B cells**

Violin Plot of Ecad across Cell Types



Ecad - marker for tumor

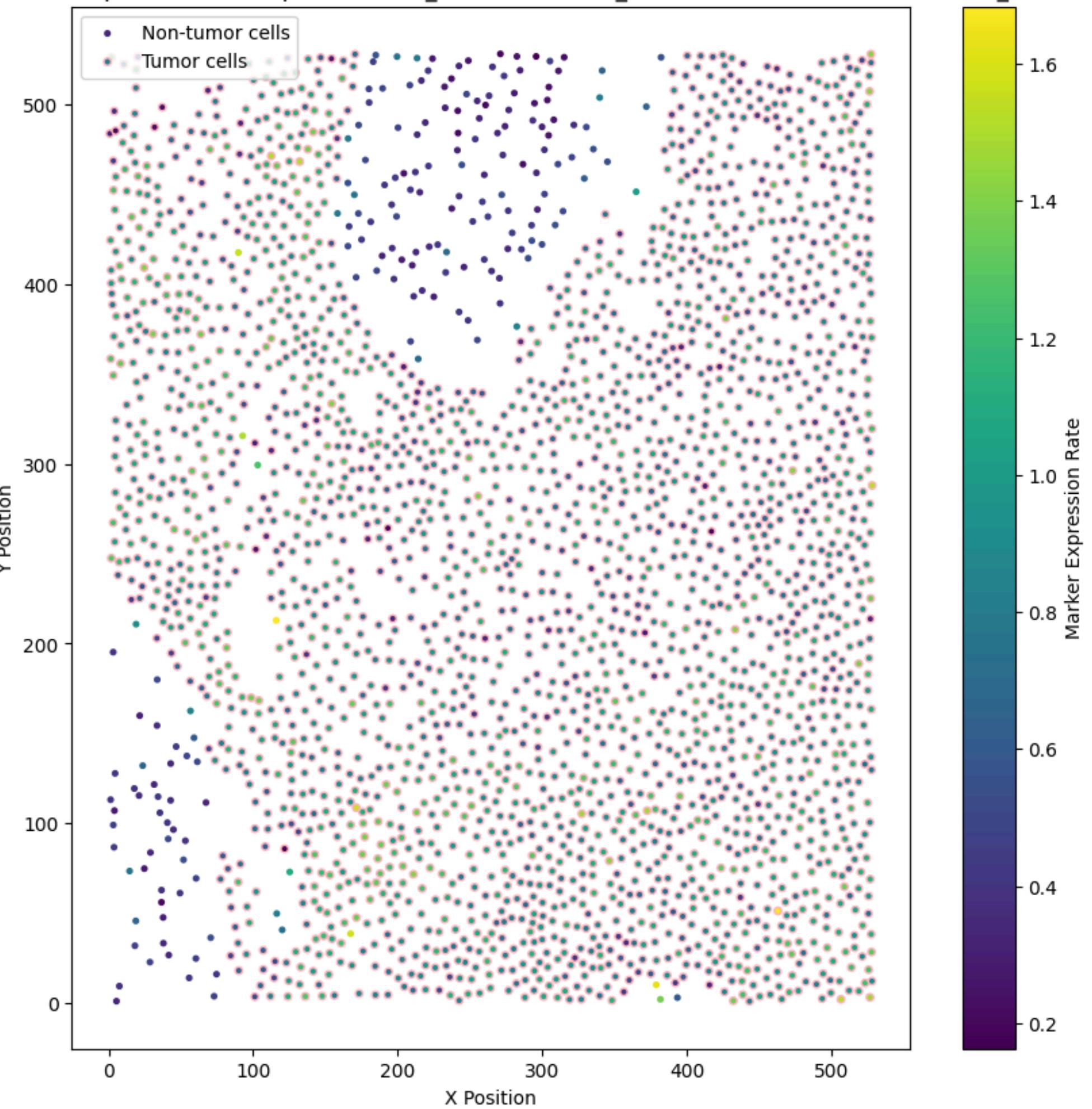
Corso G, Figueiredo J, De Angelis SP, Corso F, Girardi A, Pereira J, Seruca R, Bonanni B, Carneiro P, Pravettoni G, Guerini Rocco E, Veronesi P, Montagna G, Sacchini V, Gandini S. E-cadherin deregulation in breast cancer. *J Cell Mol Med*. 2020 Jun;24(11):5930-5936. doi: 10.1111/jcmm.15140. Epub 2020 Apr 16. PMID: 32301282; PMCID: PMC7294130.

Ecad expression over all the cells in the sample

Tumor cells have rather **high** expression of Ecad comparing to the other cell types

There are probably **subpopulations** of Tumor cells with **lower** expression

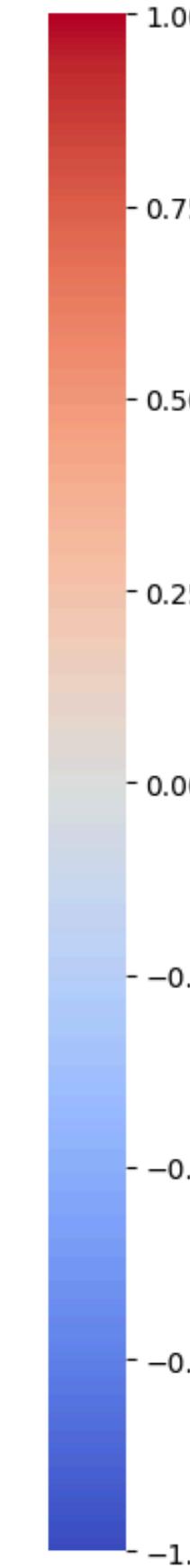
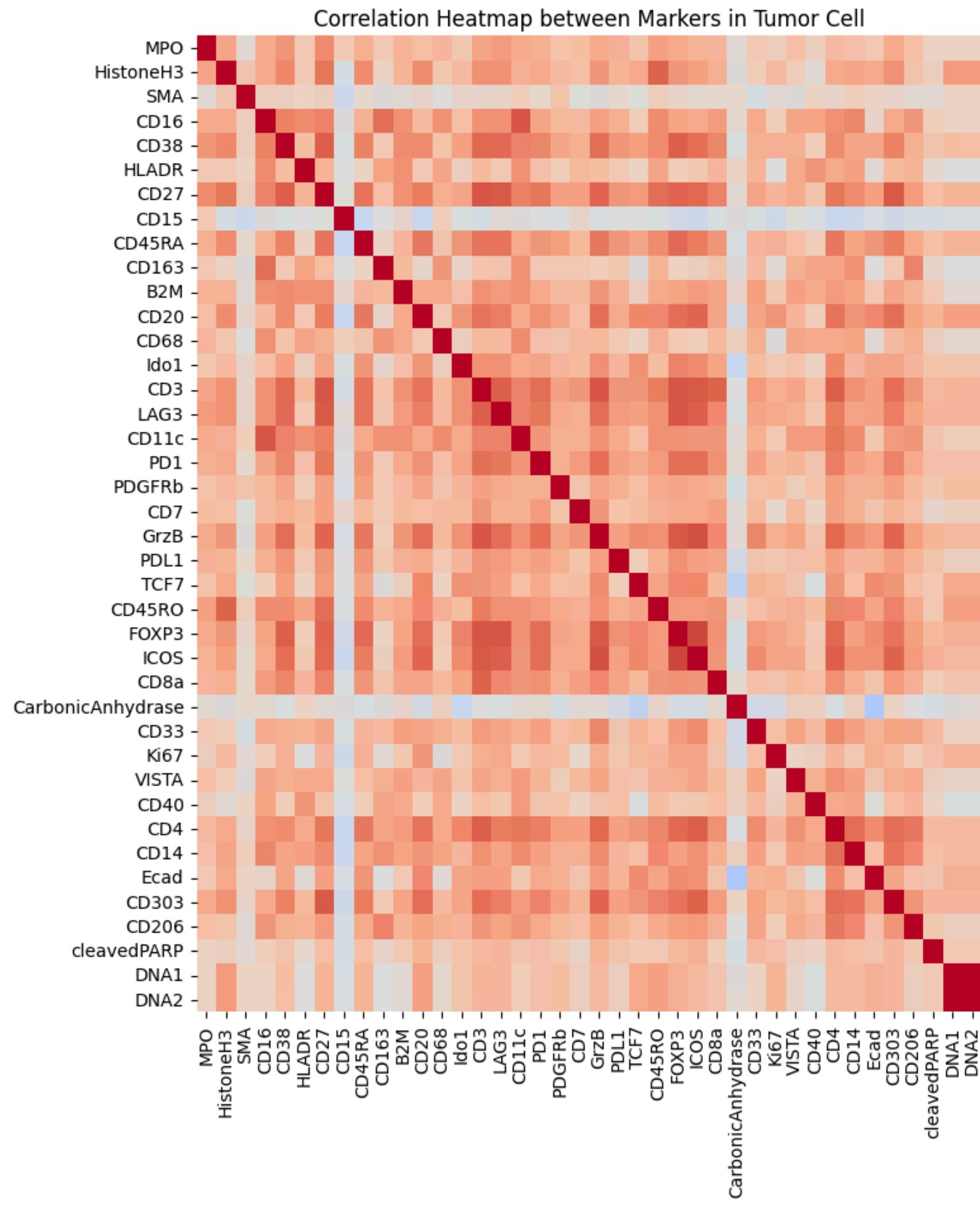
Marker Ecad expression of sample IMMUcan_Batch20191023_10032401-HN-VAR-TIS-01-IMC-01_003



Correlation heatmap

Tumor cells

pairwise relationships
between different markers
based on their expression
levels



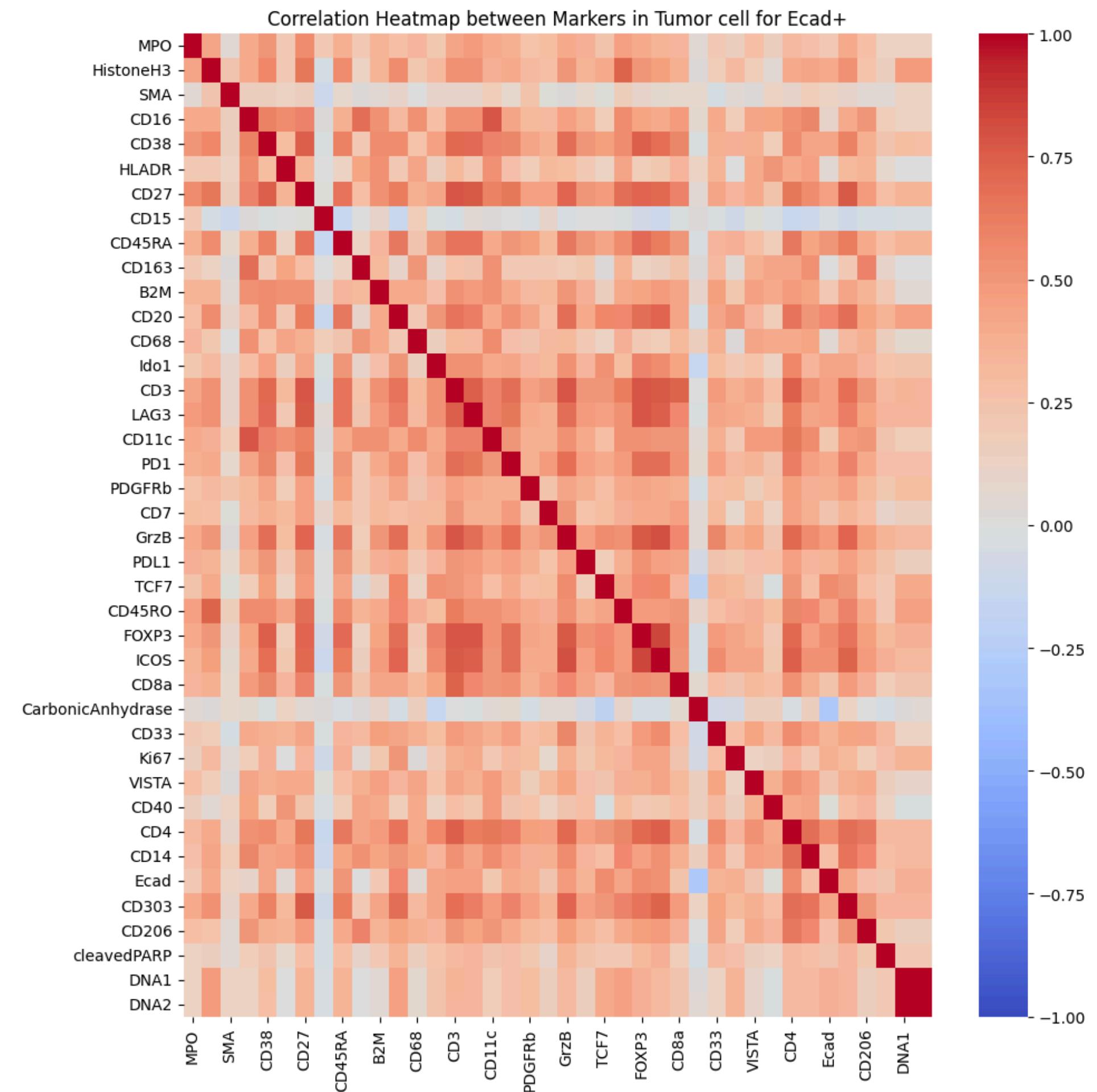
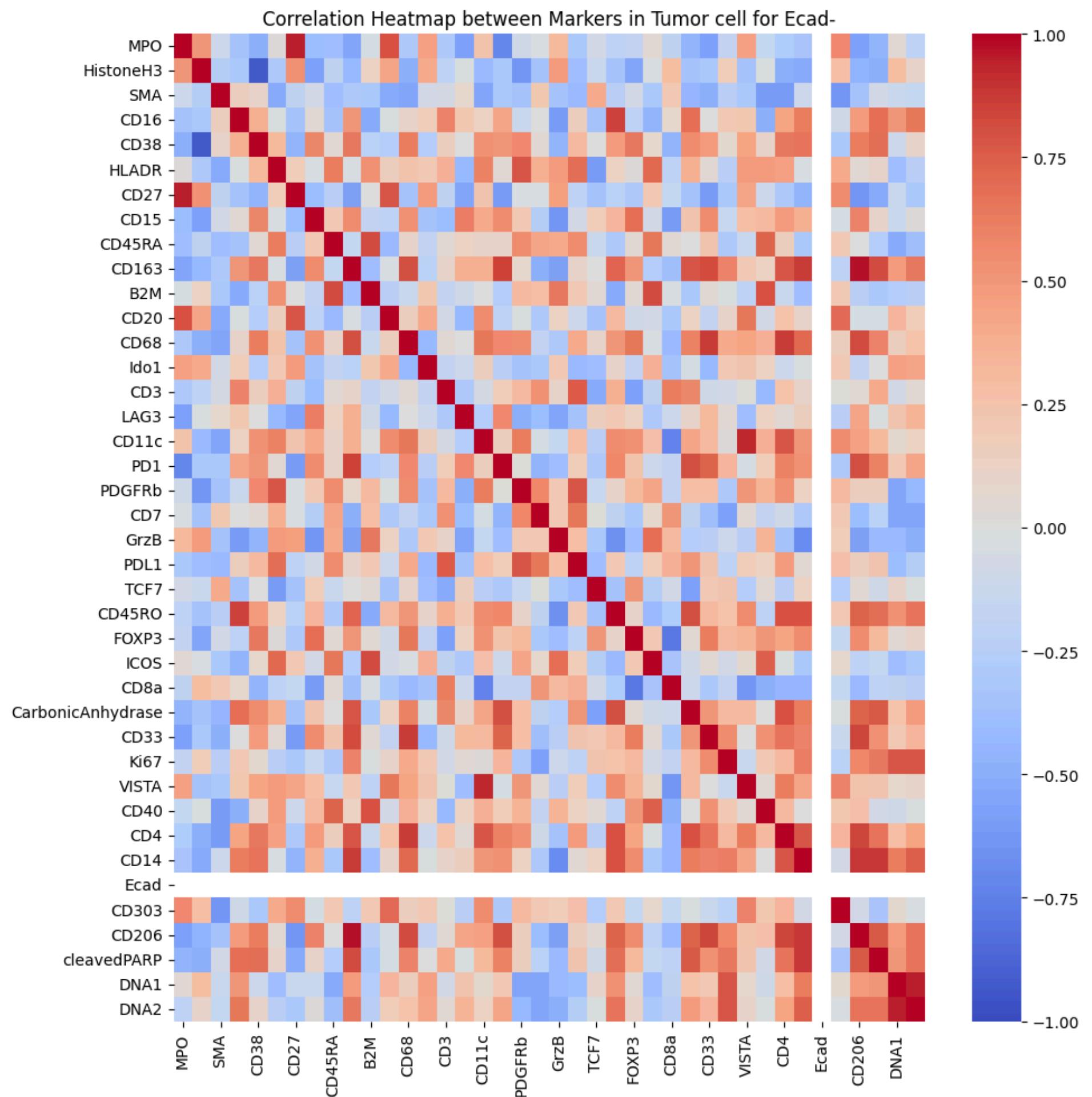
EXPLORATORY DATA ANALYSIS

Part III.

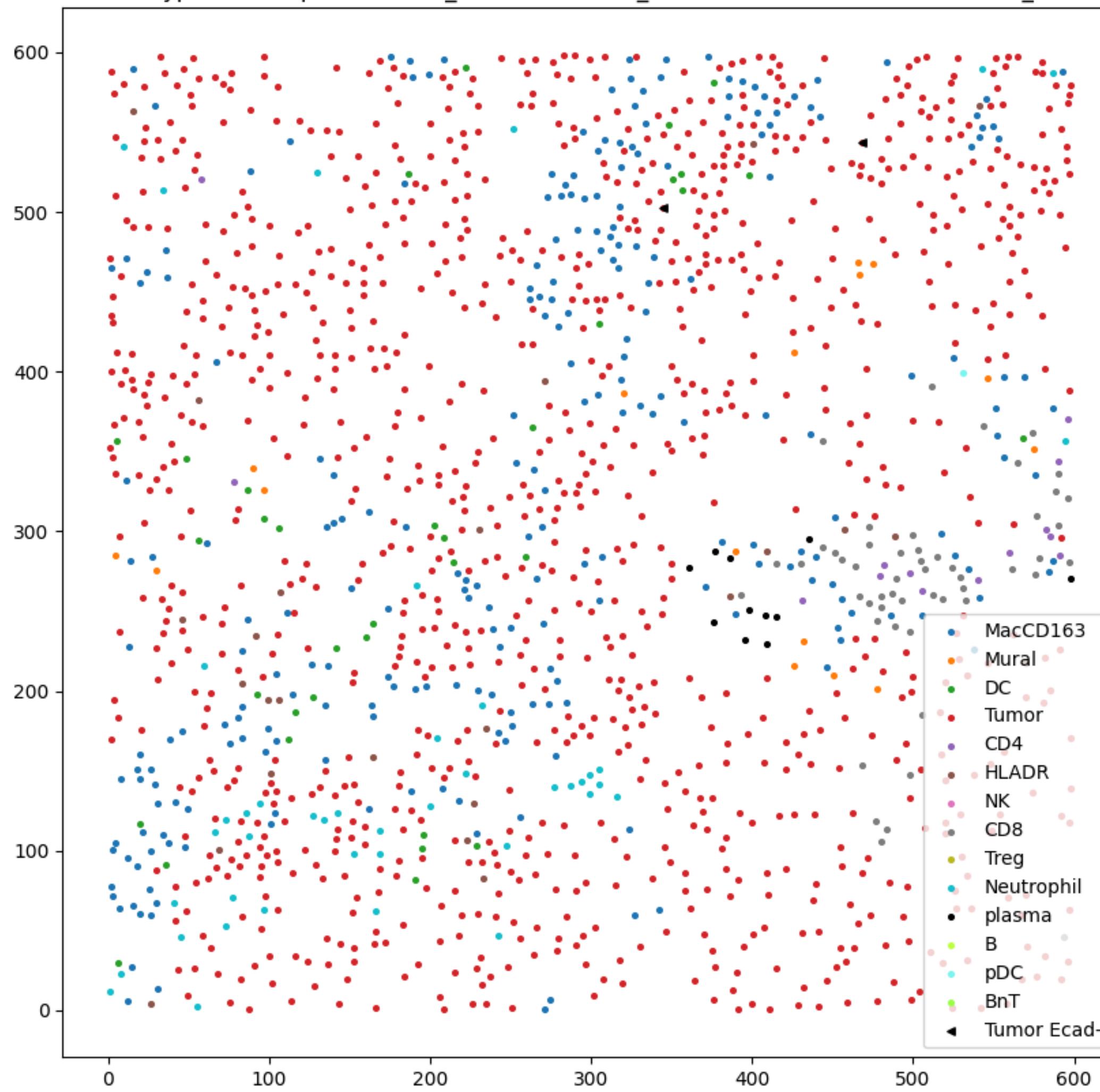
**Three biologically driven patterns of intertype marker
differentiation**



There are 2 phenotypes for Tumor cells Ecad+ and Ecad-:

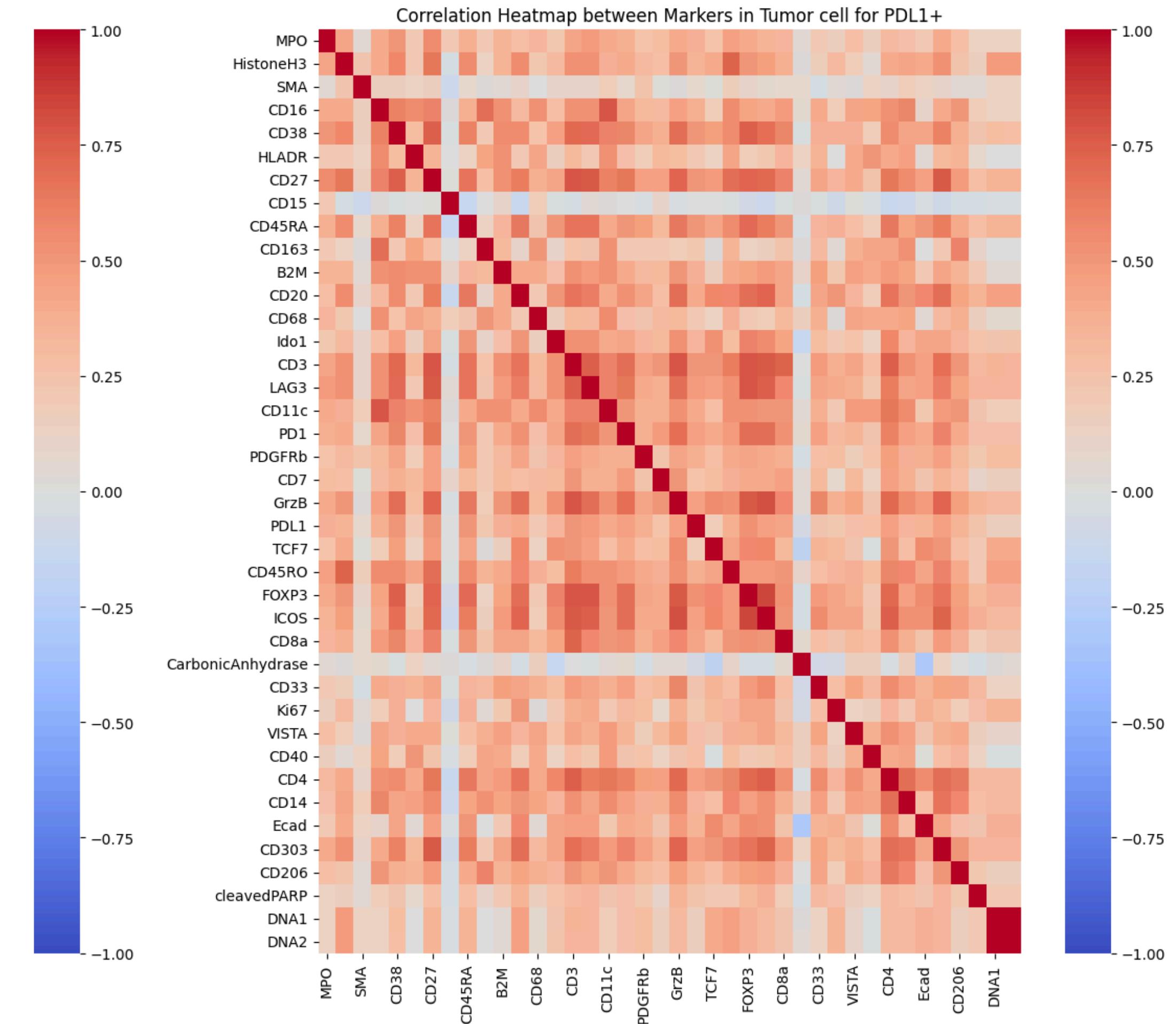
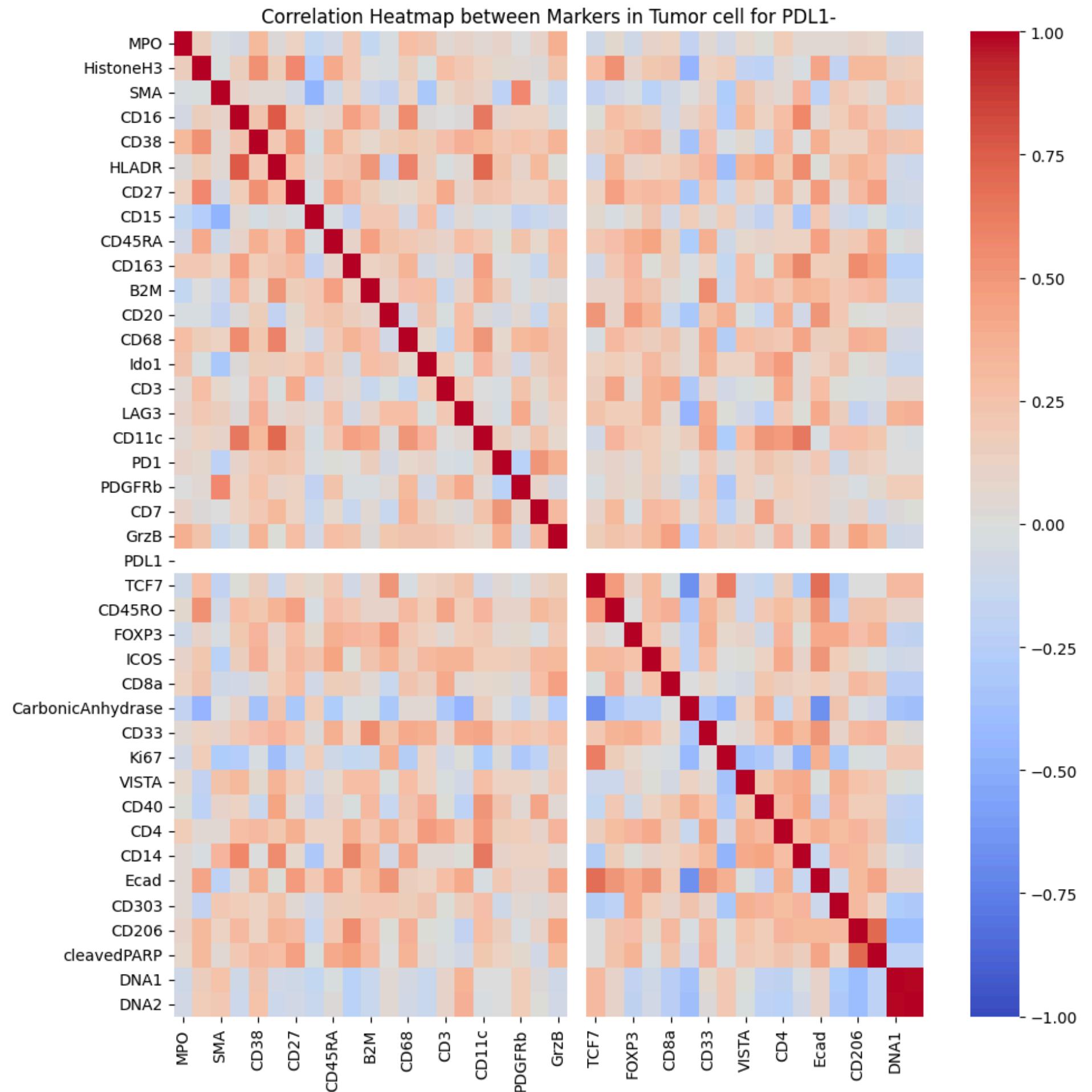


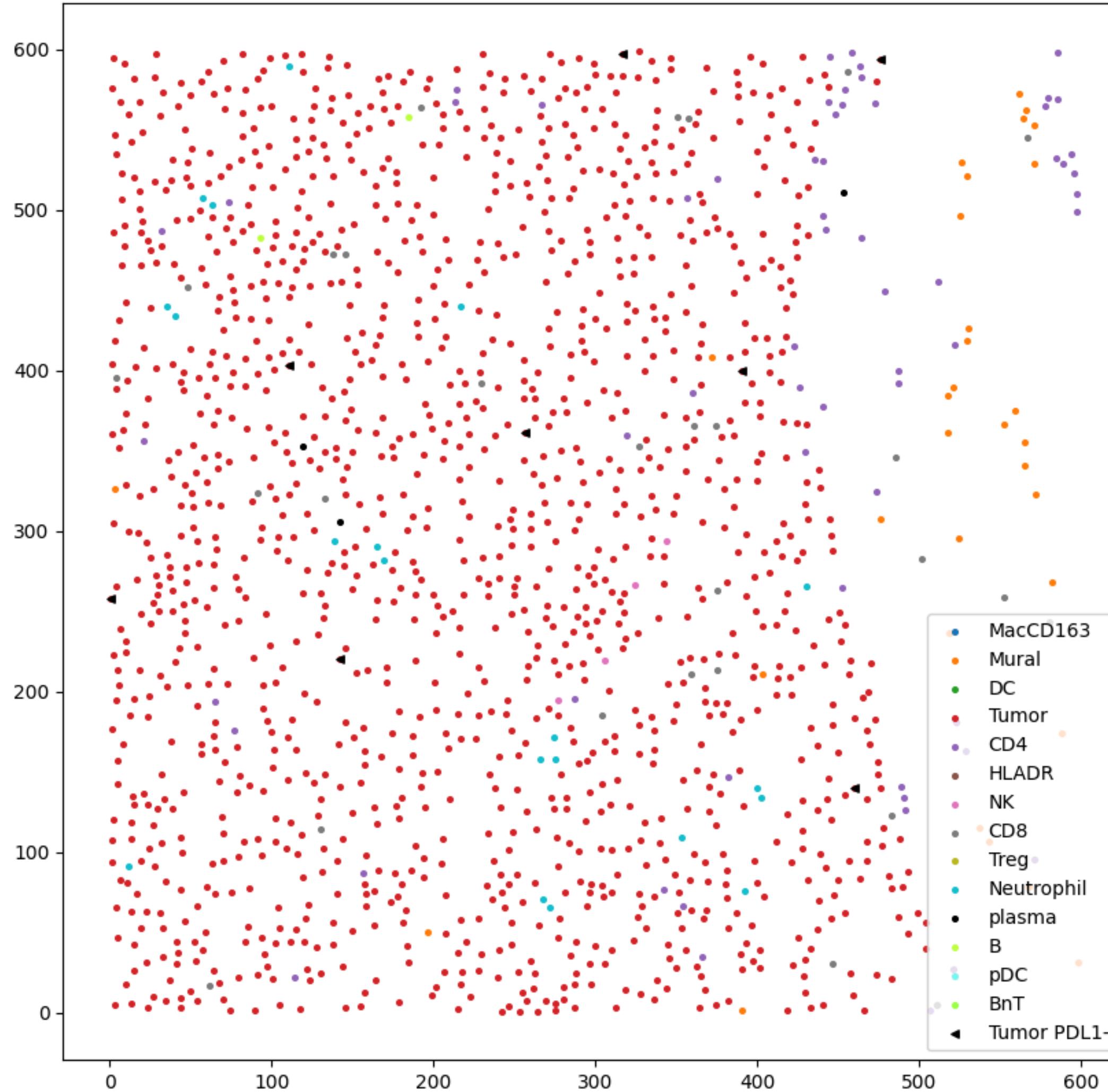
Cell Types of sample IMMUcan_Batch20201113_10042701-GU-VAR-TIS-01-IMC-01_001



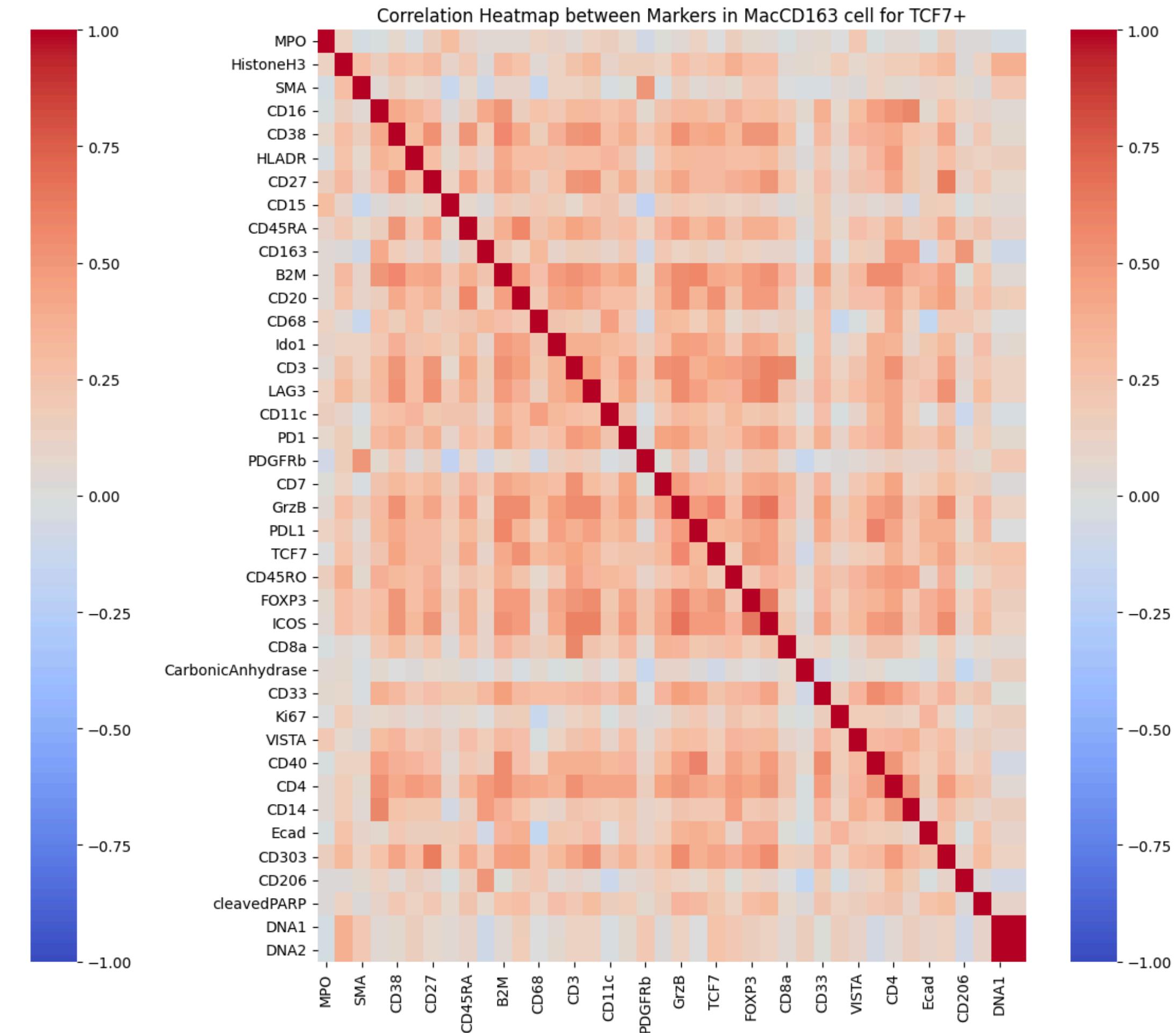
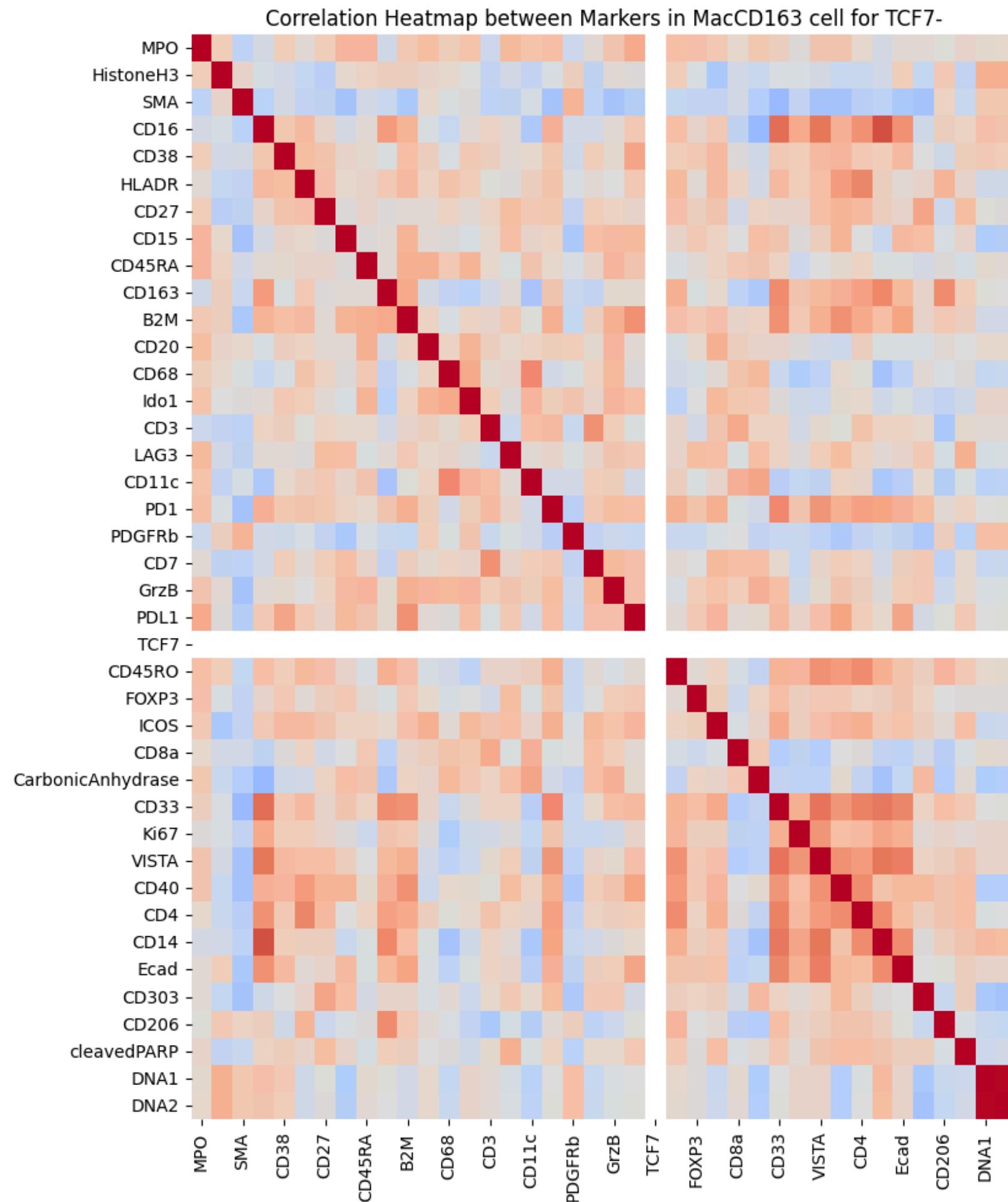
Sample with 2 phenotypes of tumor cells: Ecad+ and Ecad-

There are 2 phenotypes for Tumor cells PDL1+ and PDL1-:

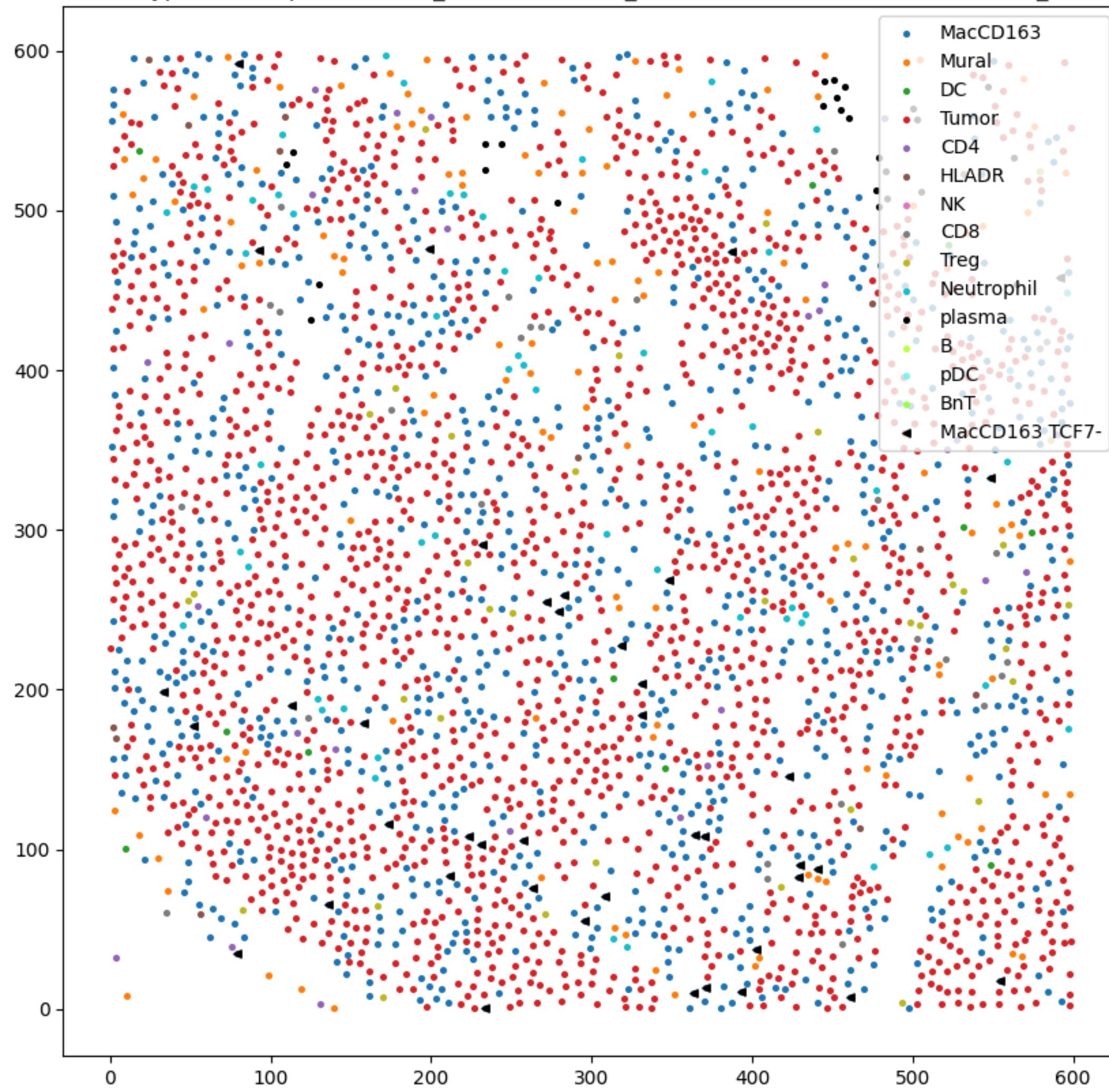




There are 2 phenotypes for MacCD163 cells TCF7+ and TCF7-:



Cell Types of sample IMMUcan_Batch20201113_10034034-THOR-VAR-TIS-01-IMC-01_002



2 phenotypes for MacCD163 cells
TCF7+ and TCF7-

There is a sample with 41 cells of
phenotype MacCD163 TCF7-

EXPLORATORY DATA ANALYSIS

Biologically driven patterns of tumor intertype marker differentiation:

- MPO - Myeloperoxidase
- α -SMA - alpha-smooth muscle actin
- CA - Carbonic anhydrase

MPO

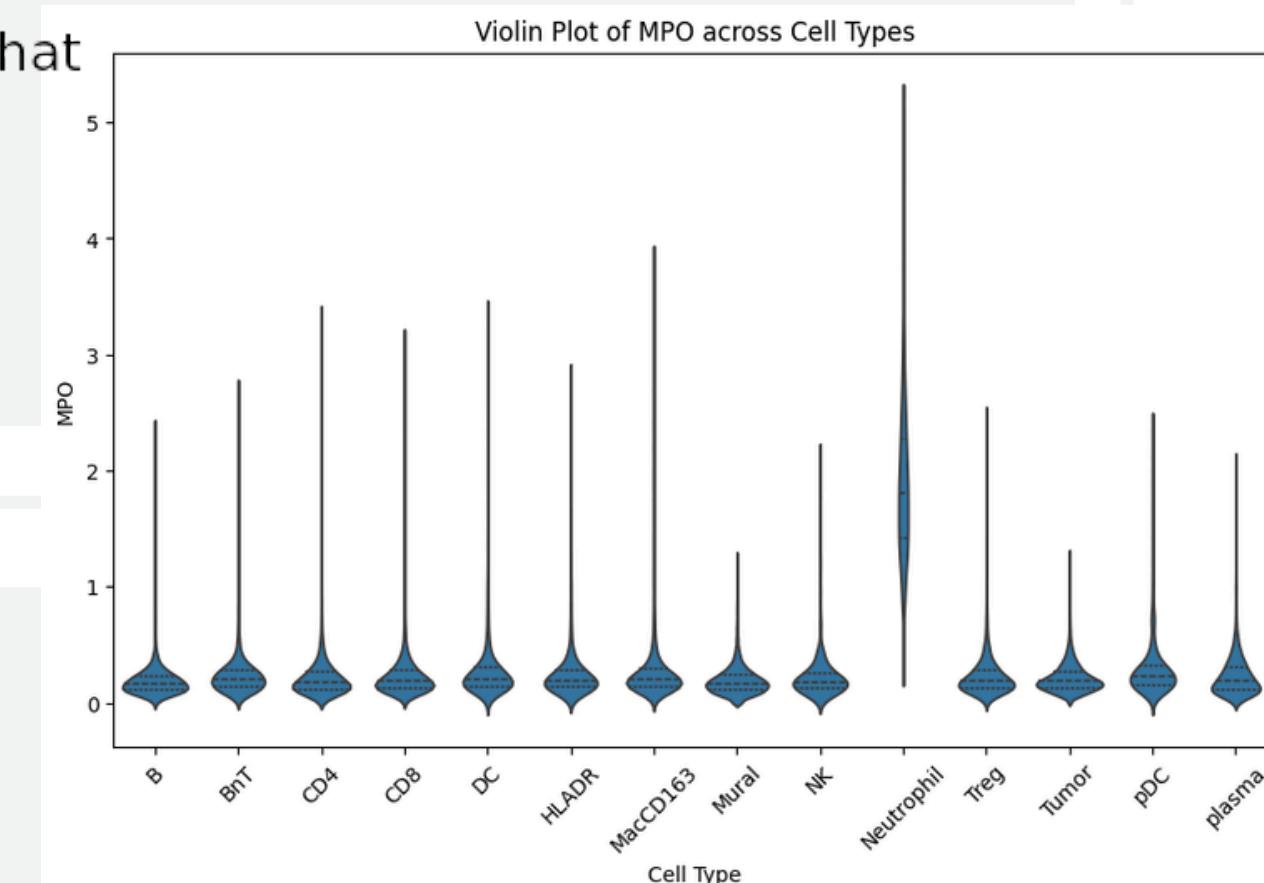
Myeloperoxidase (MPO) is a peroxidase enzyme and produces hypochlorous acid (HOCl) from hydrogen peroxide (H_2O_2) and chloride anion (Cl^-) during the neutrophil's respiratory burst.

Respiratory burst (or oxidative burst) is the rapid release of the reactive oxygen species (ROS) from different cell types.

Reactive oxygen species (ROS) are highly reactive chemicals.

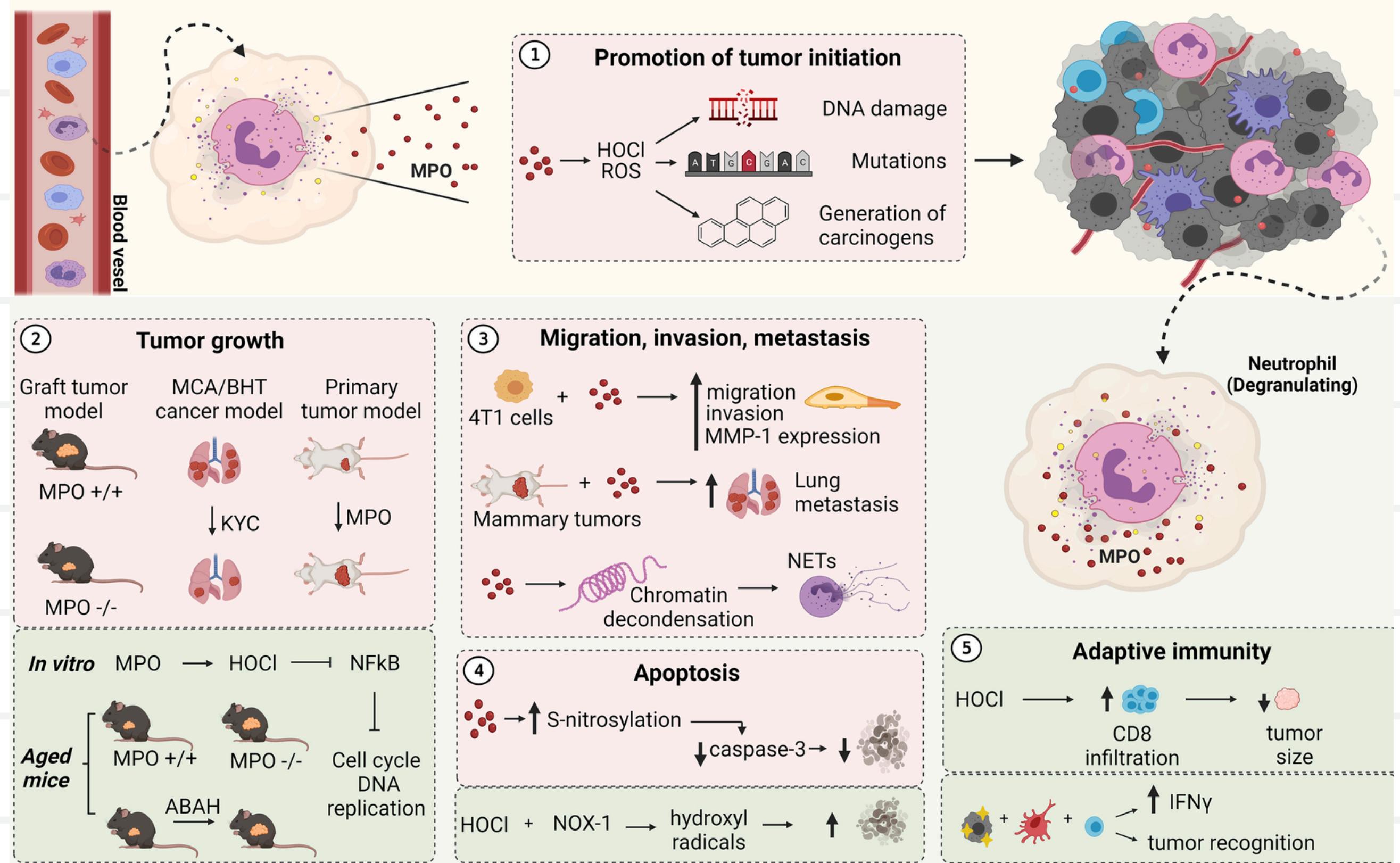
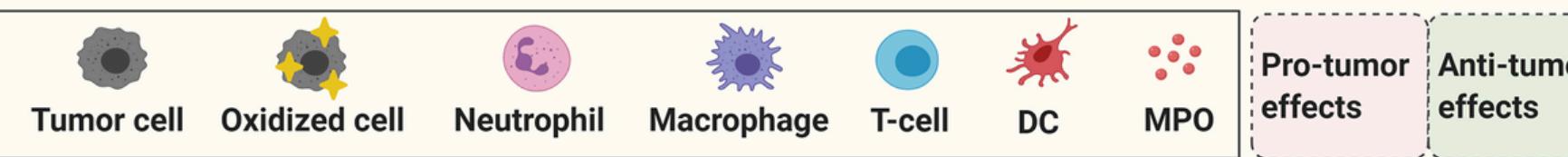
ROS are constantly generated and eliminated in the biological system and are required to drive regulatory pathways.

Under normal physiological conditions, cells control ROS levels by balancing the generation of ROS with their elimination by scavenging systems. But under oxidative stress conditions, excessive ROS can damage cellular proteins, lipids and DNA, leading to fatal lesions in the cell that contribute to carcinogenesis.



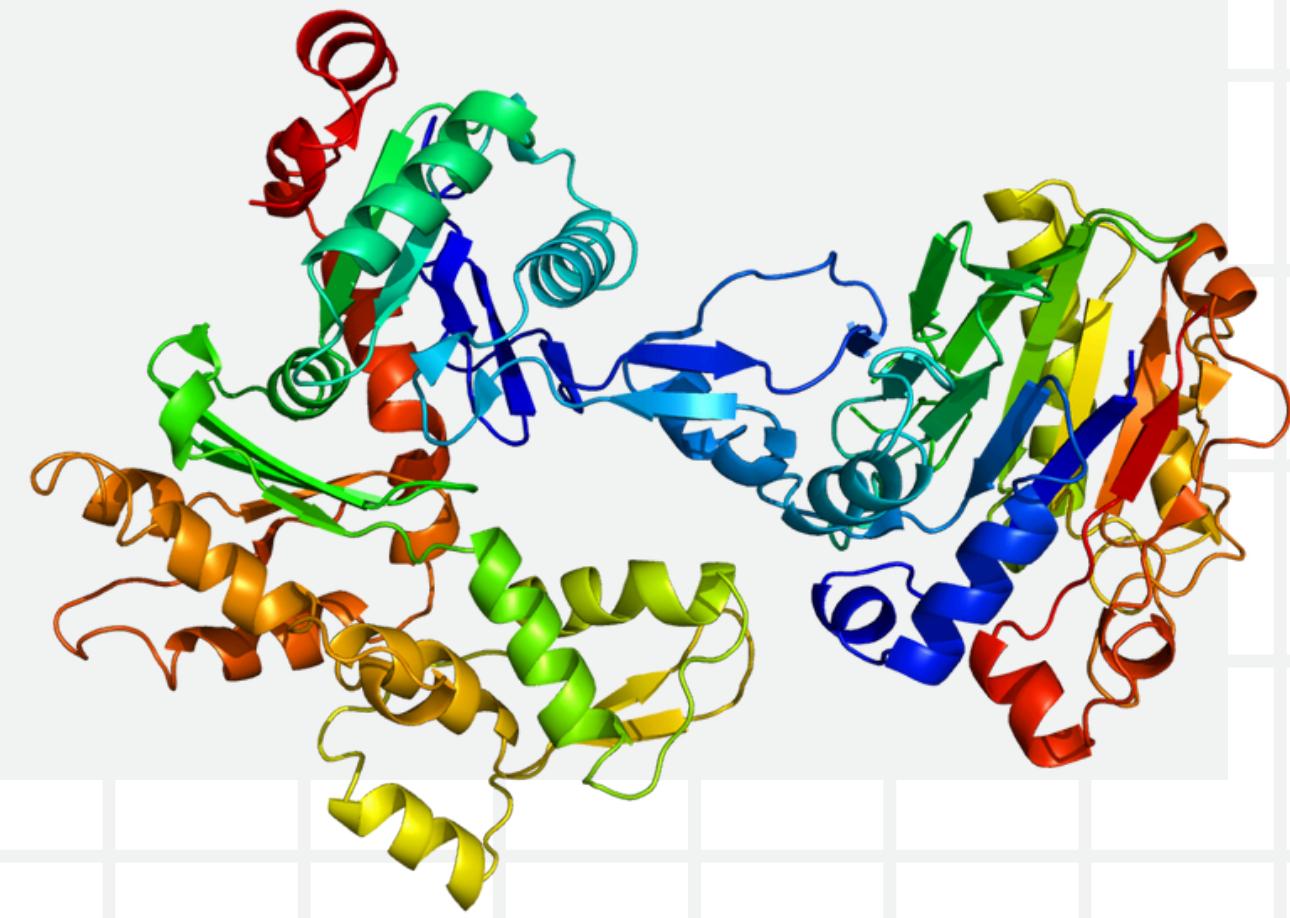
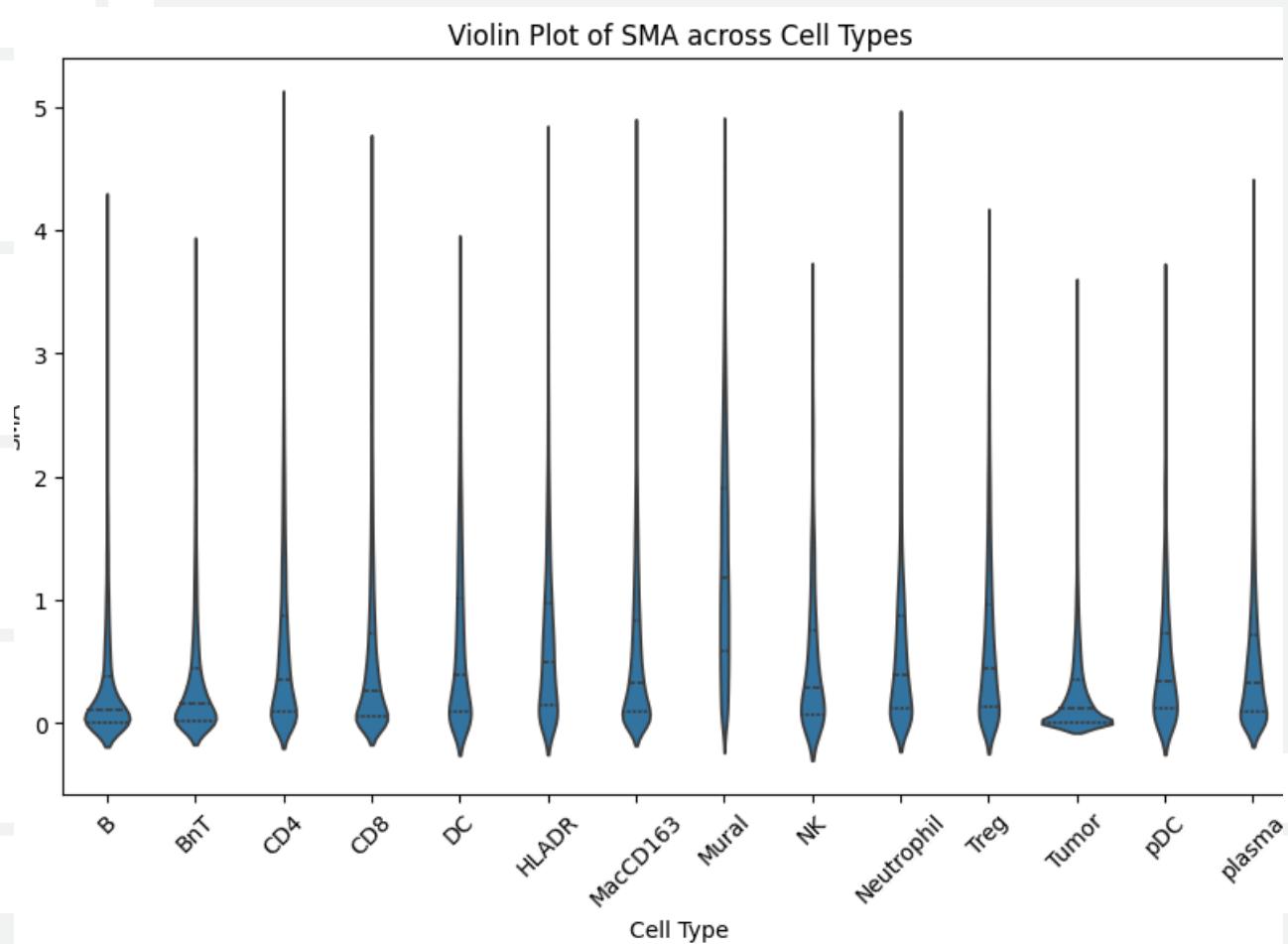
MPO

- the impact of MPO in the setting of cancer has gained attention over the past years and several reports describe the participation of this enzyme in the regulation of cancer.
- the bulk of evidence highlights MPO as a molecule that favors tumor initiation and progression. MPO has been implicated in tumor initiation through the support of a hypermutagenic environment due to the action of MPO-derived oxidants that are able to oxidize and modify DNA.
- cancer progression is also influenced by the presence of MPO which has been involved in the regulation of tumor growth, apoptosis, cell migration and metastasis.



SMA - ALPHA-SMOOTH MUSCLE ACTIN

Actins - family of globular multi-functional proteins that form microfilaments
α-SMA is one of 6 different actin isoforms and is involved in the contractile apparatus of smooth muscle

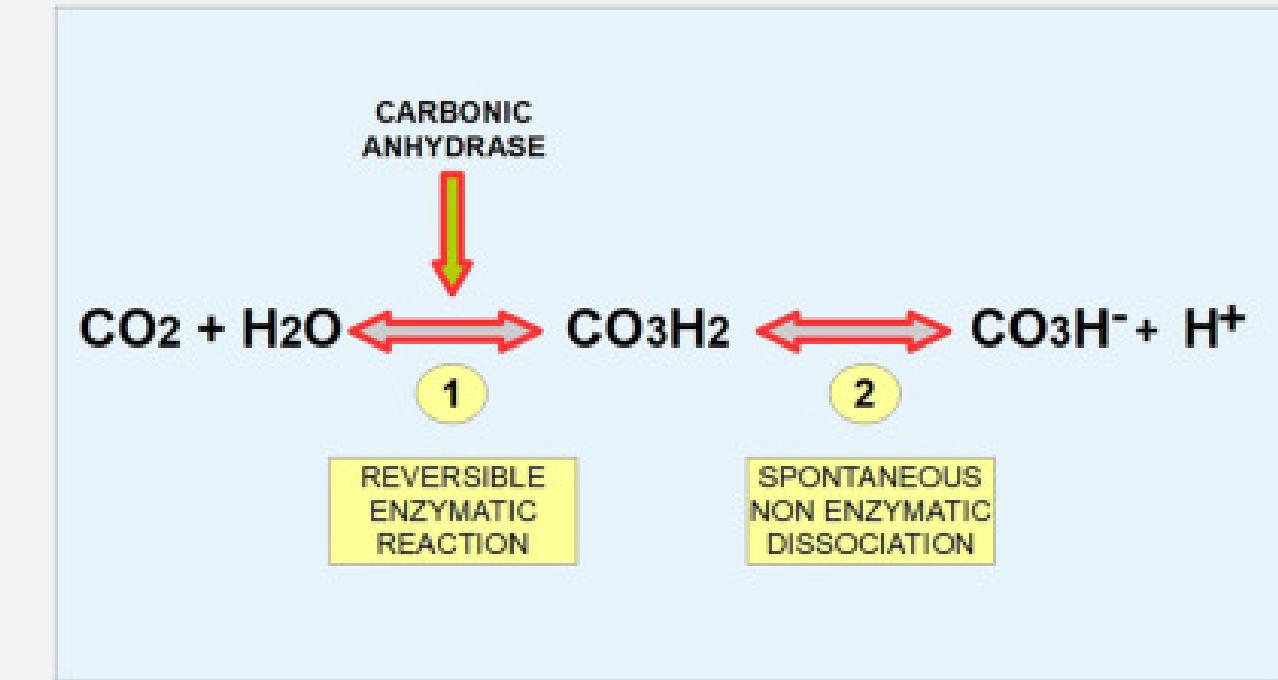
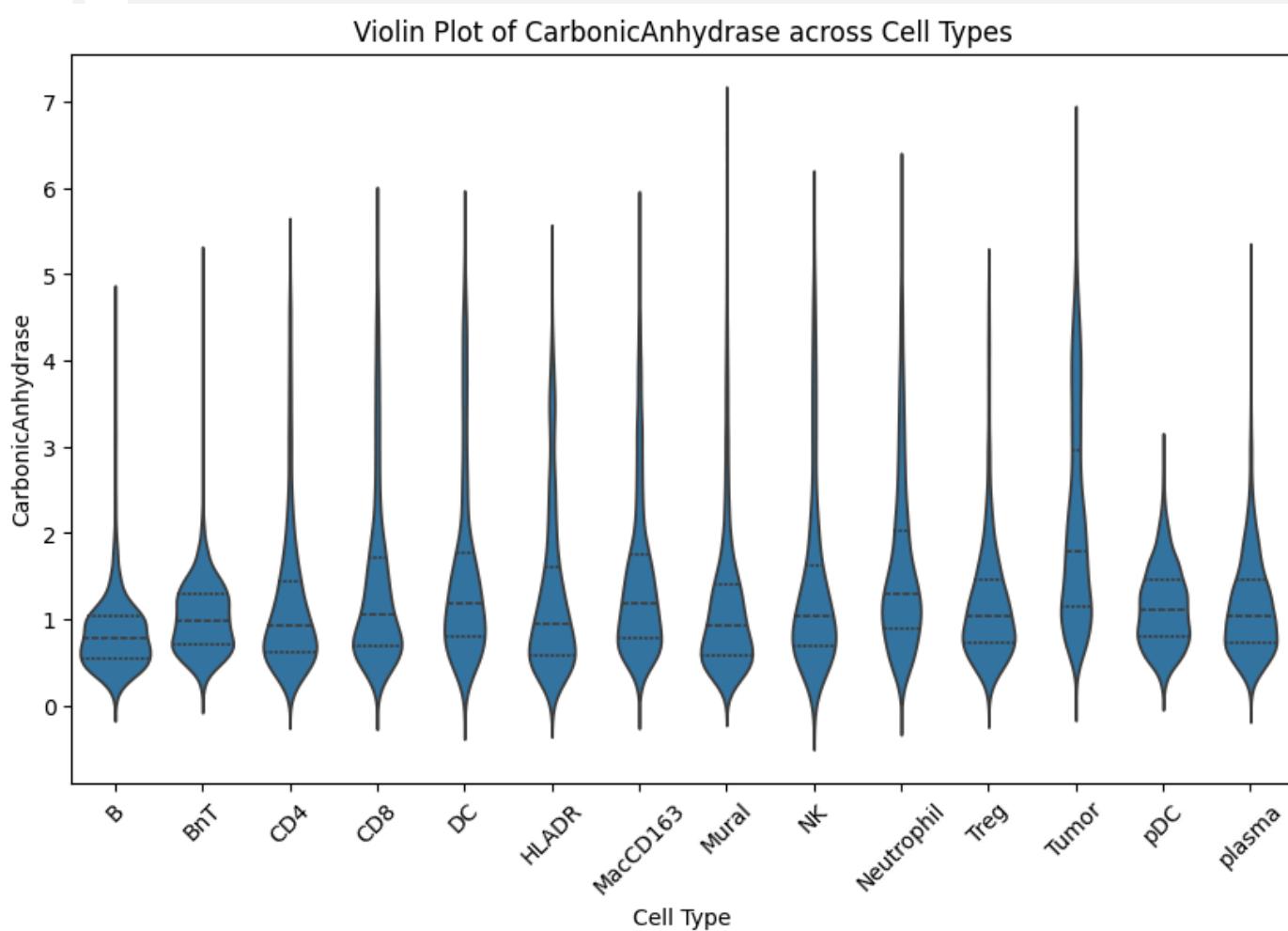


SMA - ALPHA-SMOOTH MUSCLE ACTIN

- α -SMA is positive in myofibroblasts and is an epithelial to mesenchymal transition (EMT) marker.
- α -SMA correlates with activation of fibroblast to myofibroblast. Myofibroblasts differ from fibroblasts because of its contractile ability.
- EMT is a process by which tumor cells develop to be more hostile and able to metastasize.
- Progression of tumor cells is always followed by cell composition and extracellular matrix component alteration.

CARBONIC ANHYDRASE (CA)

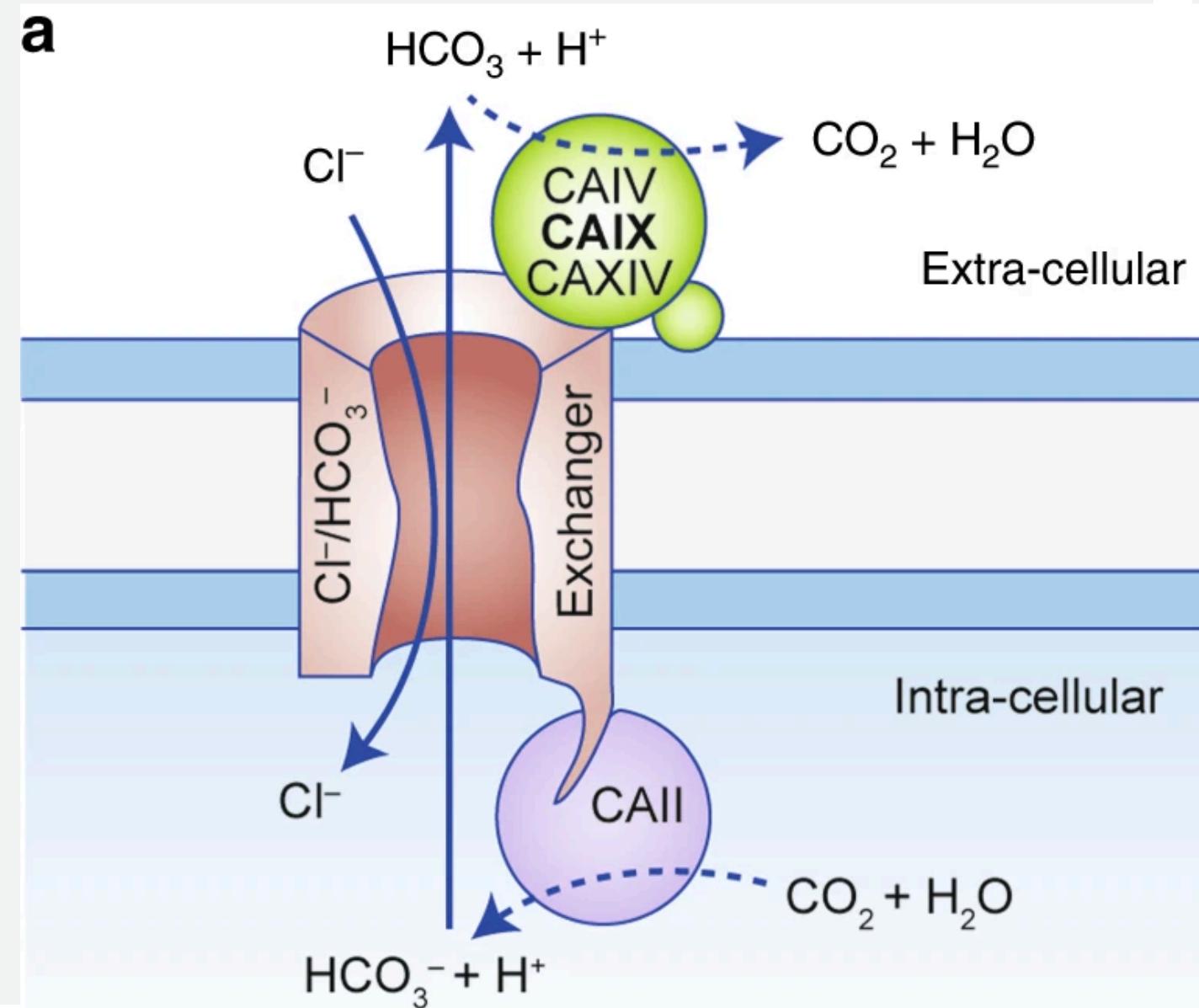
The family of carbonic anhydrases (CAs), also known as carbonic dehydratases, are a group of metalloenzymes that catalyze in a reversible form the conversion of carbon dioxide and water to bicarbonate and a proton



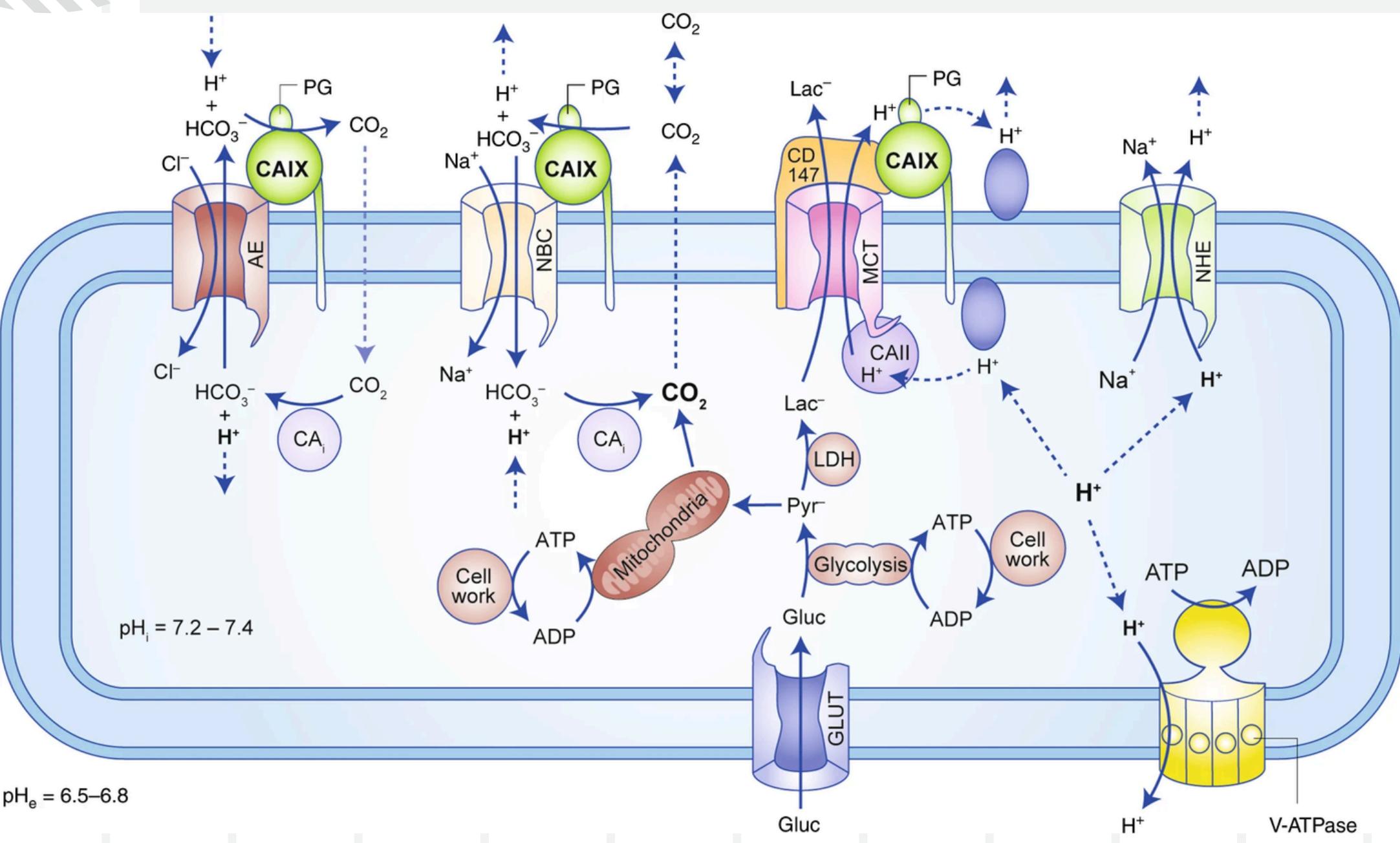
CARBONIC ANHYDRASE (CA)

Malignant tissues modify their own microenvironmental parameters. And they do so with a very high flexibility and adaptation to permanently changing conditions. pH, pO₂, and interstitial fluid pressure have additive influence on tumor growth, proliferation, migration, and invasion, and enhance cell survival of the more fit in a Darwinian selection fashion.

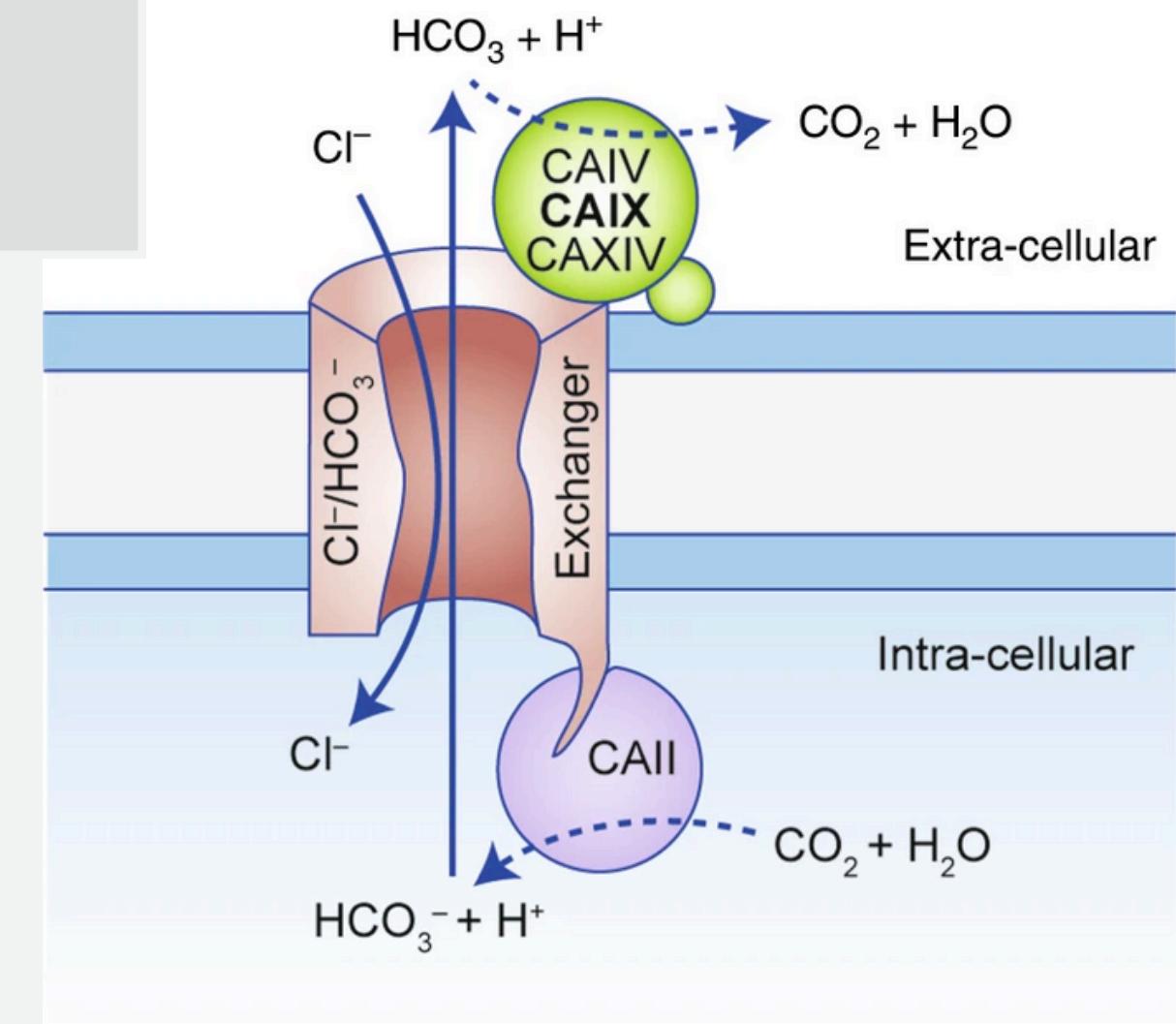
Carbonic anhydrases are a group of enzymes with the ability to modify intra and extracellular pH and therefore influence proliferation, migration and invasion, and contribute significantly to the pH gradient inversion.



CARBONIC ANHYDRASE (CA)

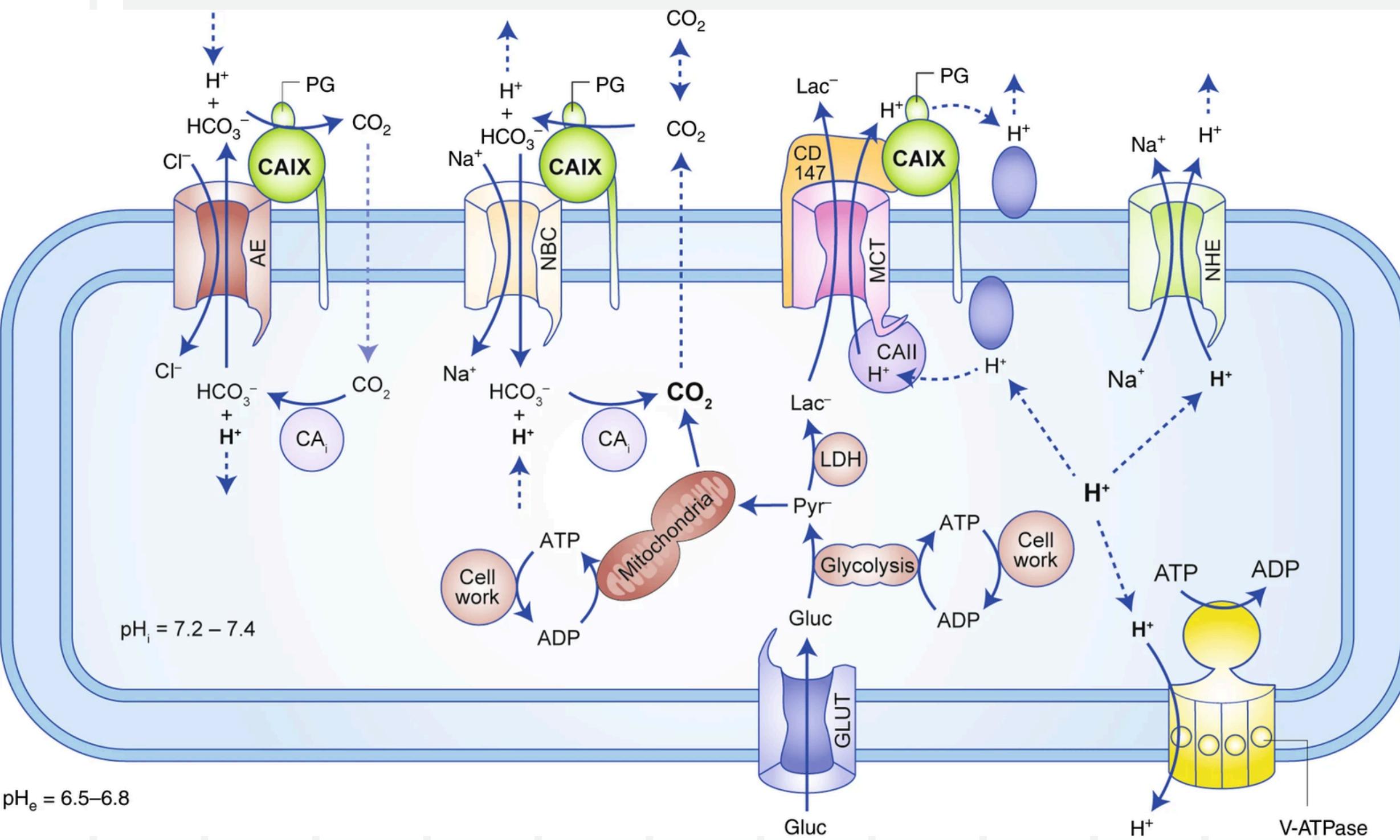


Holger M. Becker Carbonic anhydrase IX and acid transport in cancer, *British Journal of Cancer* volume 122, pages 157-167 (2020)



Extracellular acidification promotes tumour progression via various mechanisms, including pH-dependent modulation of integrin-mediated cell-matrix adhesion, degradation of the extracellular matrix via activation of cathepsins and various matrix metalloproteases and by killing of adjacent host cells.

CARBONIC ANHYDRASE (CA)



Cancer cells primarily express the plasma-membrane-associated CA isoforms CAIX and CAXII, as well as intracellular CAs such as CAI and CAII. Amongst the cancer-related CAs, CAIX has gained most attention, since expression of this isoform in healthy tissue is restricted to epithelial cells in the stomach and gut, but is strongly upregulated in many tumour tissues.