ITMO UNIVERSITY

CT Lab
ITMO UNIVERSITY

Single-cell Navigator: visualizing scRNA-seq data

Konstantin Zaitsev, ITMO University

Systems Biology Program, March 18th, 2021

Visualizing scRNA-seq data

Main goals:

- To make hypothesis generations easier
- Remove “man-in-the-middle”

Extra goals:

- Fast
- Responsive

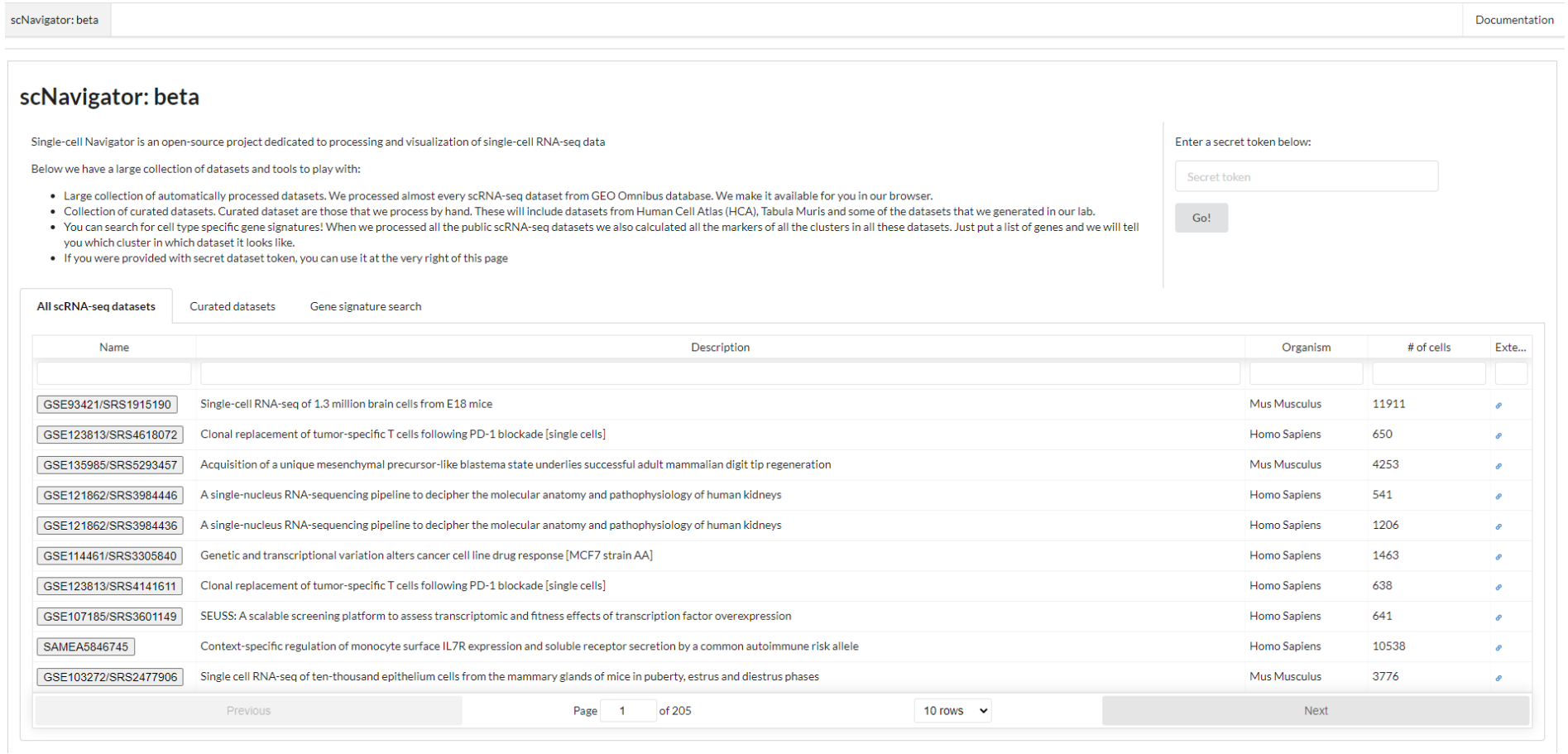
Visualizing scRNA-seq data

https://artyomovlab.wustl.edu/scn/

(still in production, so feedback is very welcome)

Let’s open the dataset

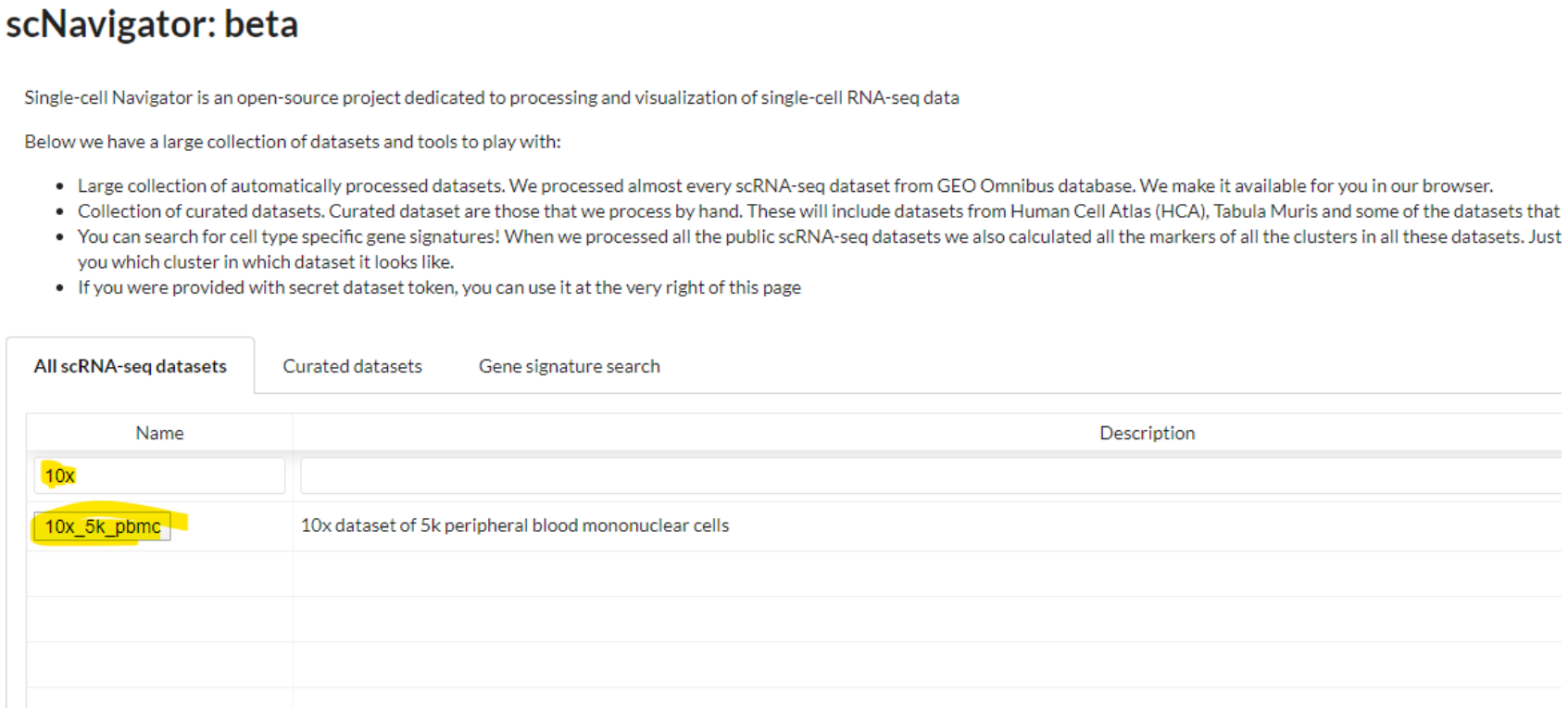
- Go to https://artyomovlab.wustl.edu/scn/



ITMO UNIVERSITY

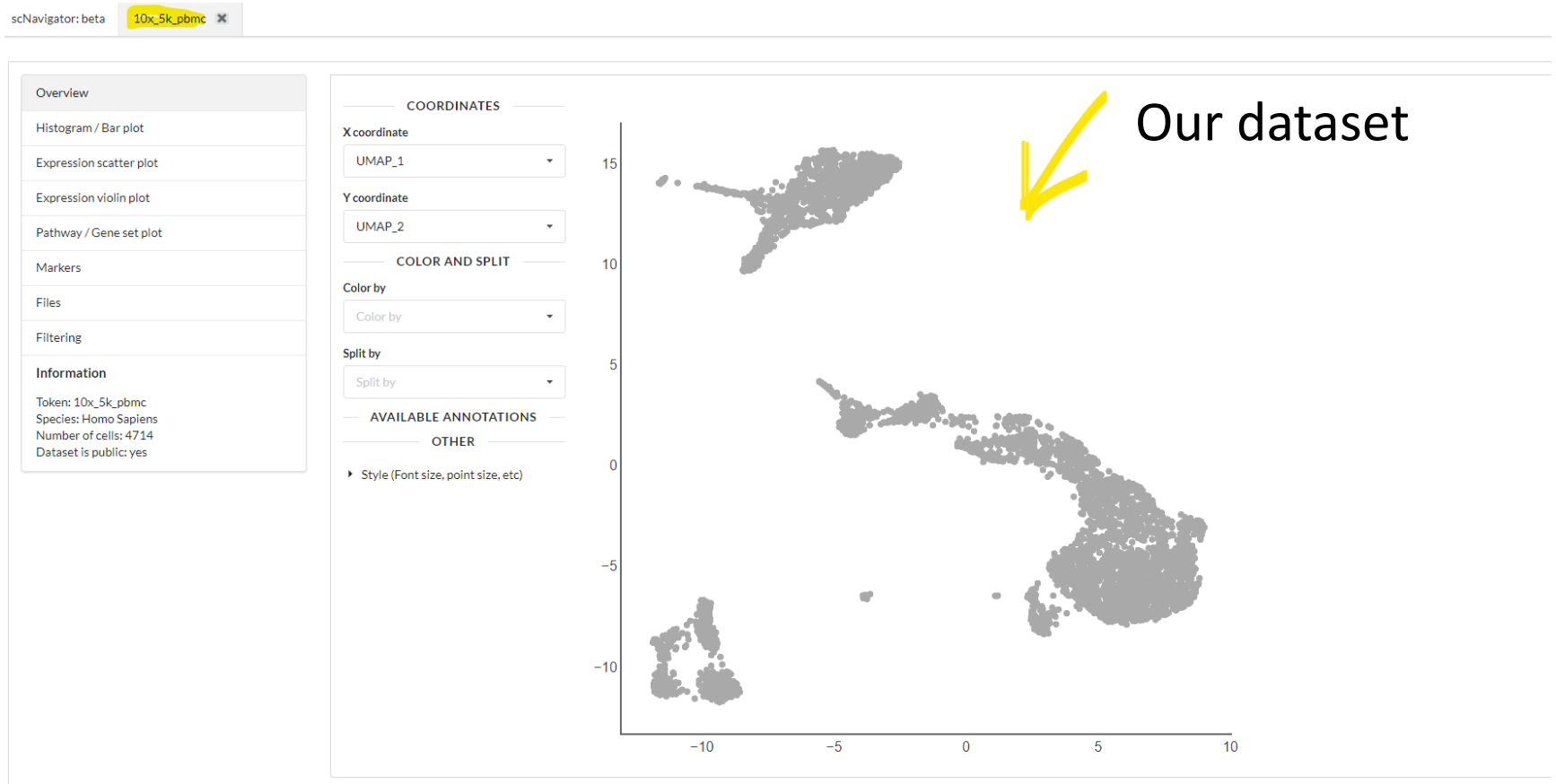
Let’s open the dataset

- Go to https://artyomovlab.wustl.edu/scn/
- Search for 10x
- And click on the dataset



If you have any problem finding dataset

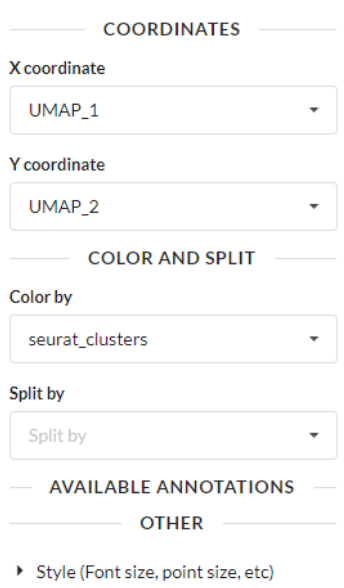
- Just go to https://artyomovlab.wustl.edu/scn/?token=10x_5k_pbmc

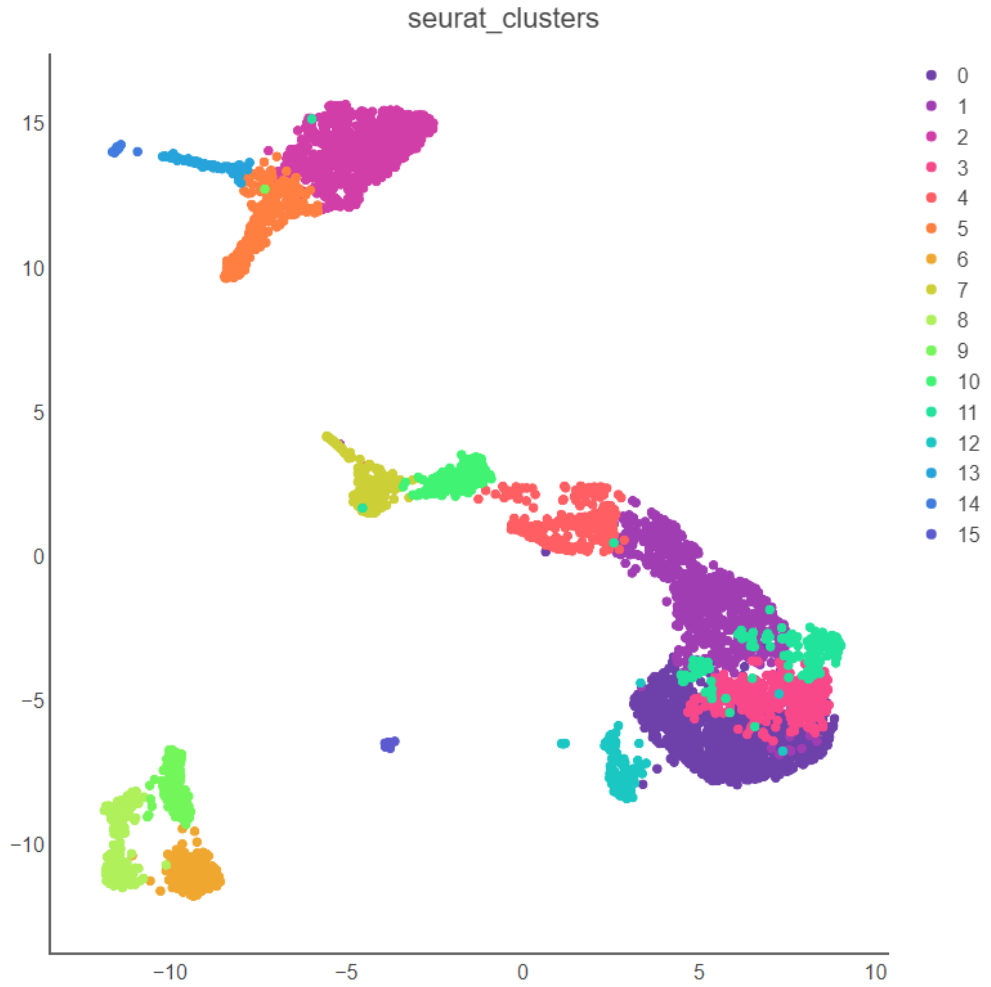


We can color the cells

- Cluster
- Number of UMIs
- Number of genes detected

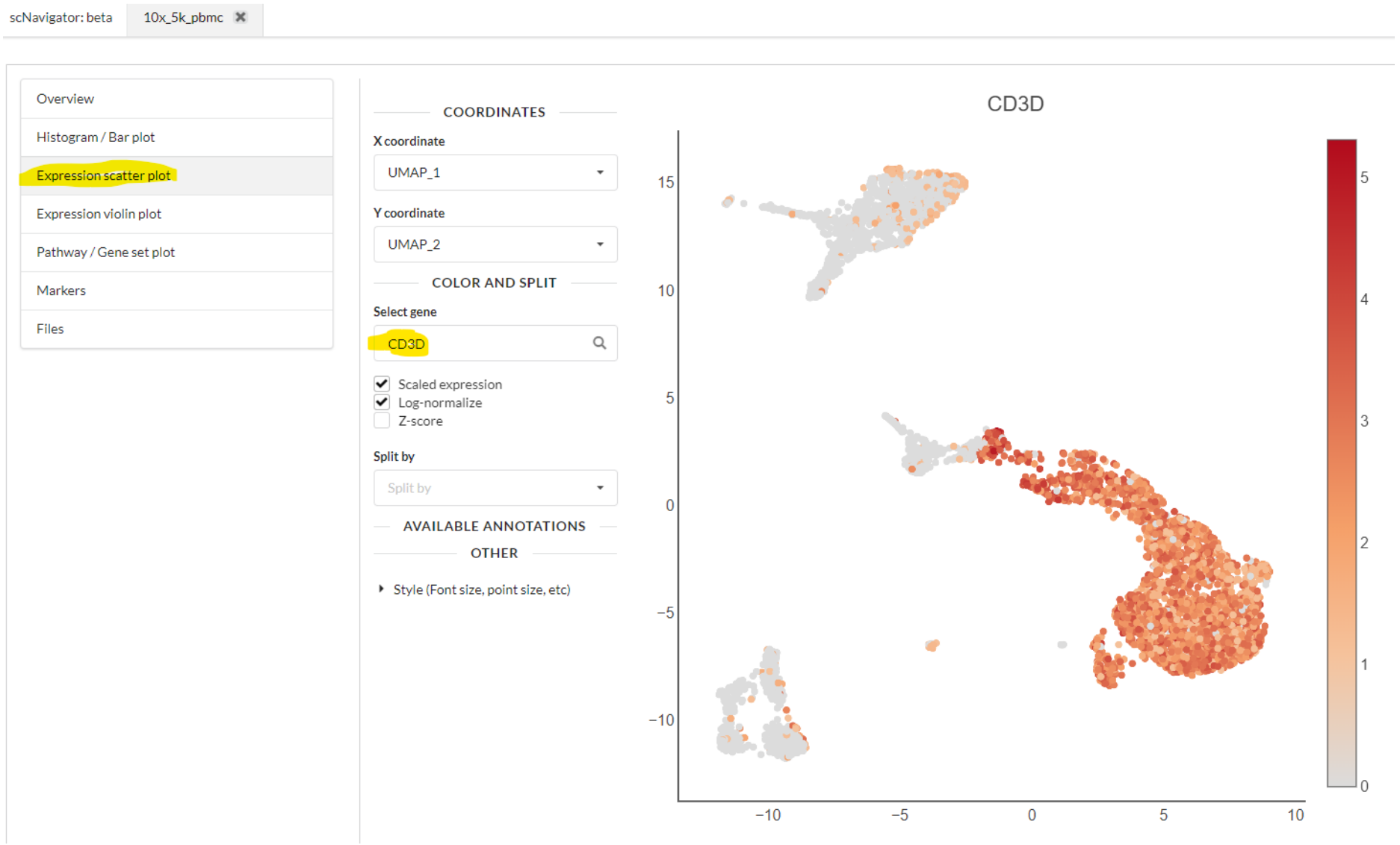
- tsne_Cluster_centers





ITMO UNIVERSITY

Expression of CD3d



ITMO UNIVERSITY

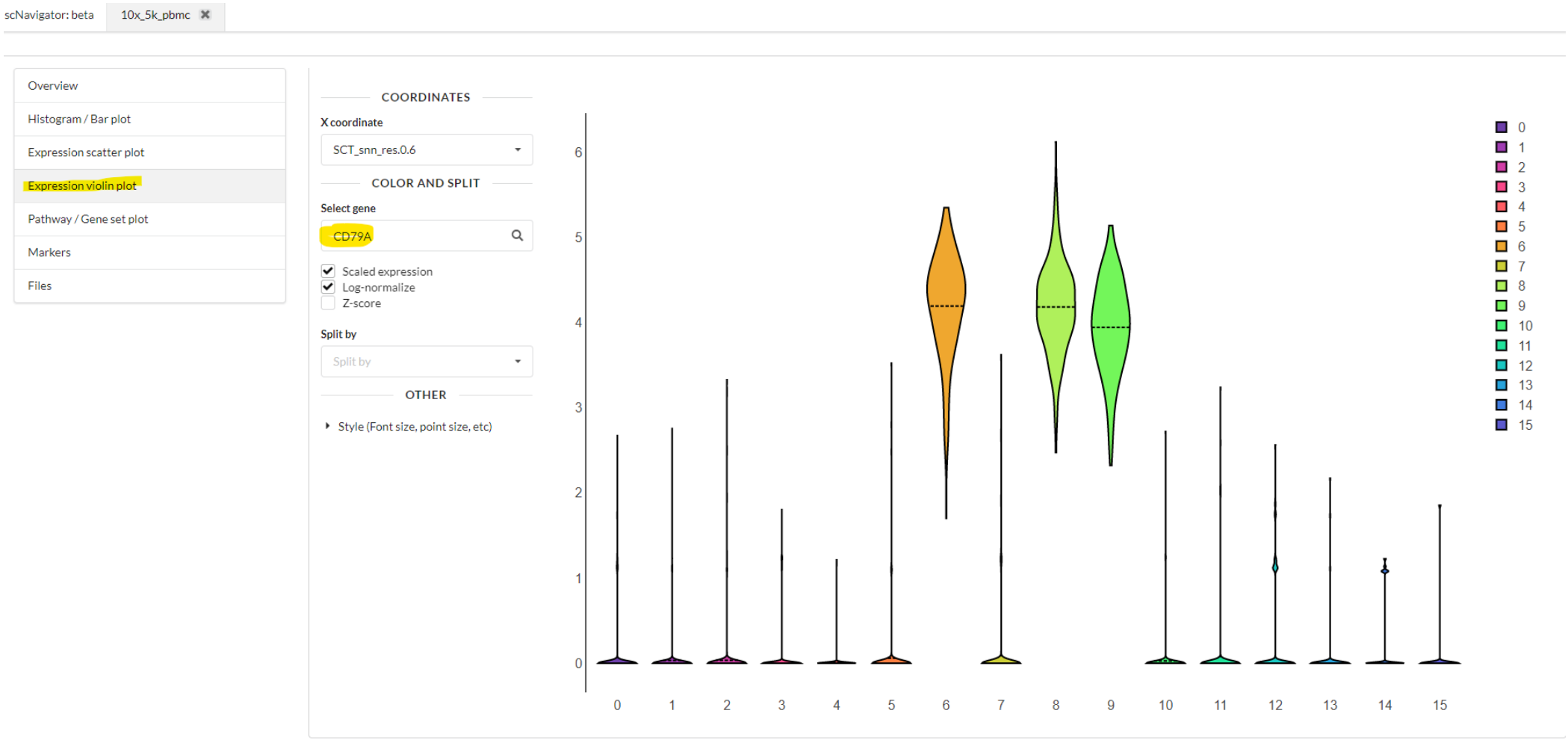
Or you can go for any of your favorite genes

Expression scatter plot

- Expression scatter plot shows gene expression **in each cell**
- We can see that expression of some genes is localized with clusters

ITMO UNIVERSITY

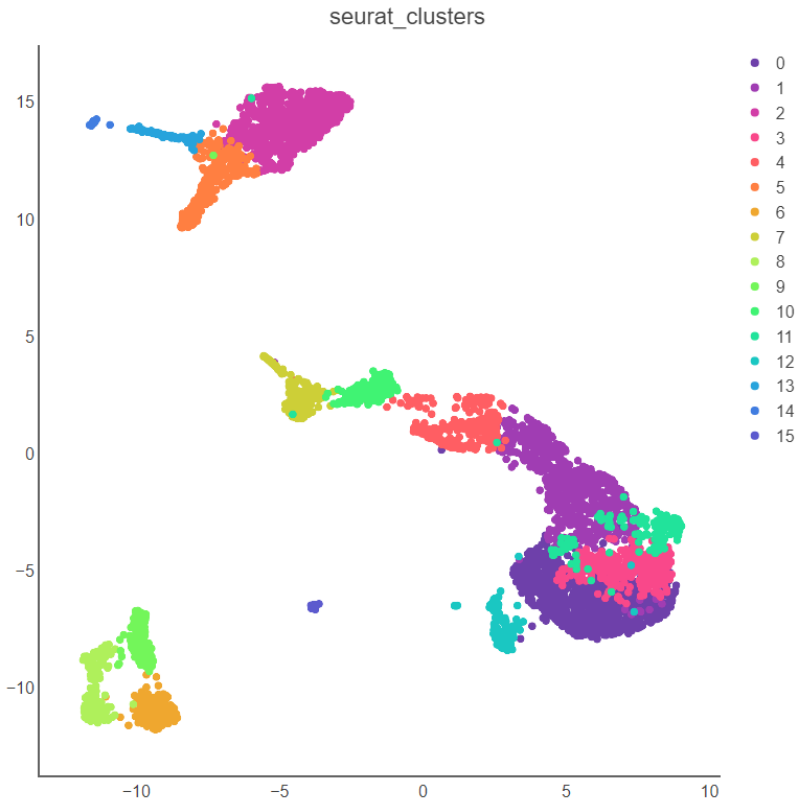
Violin plot

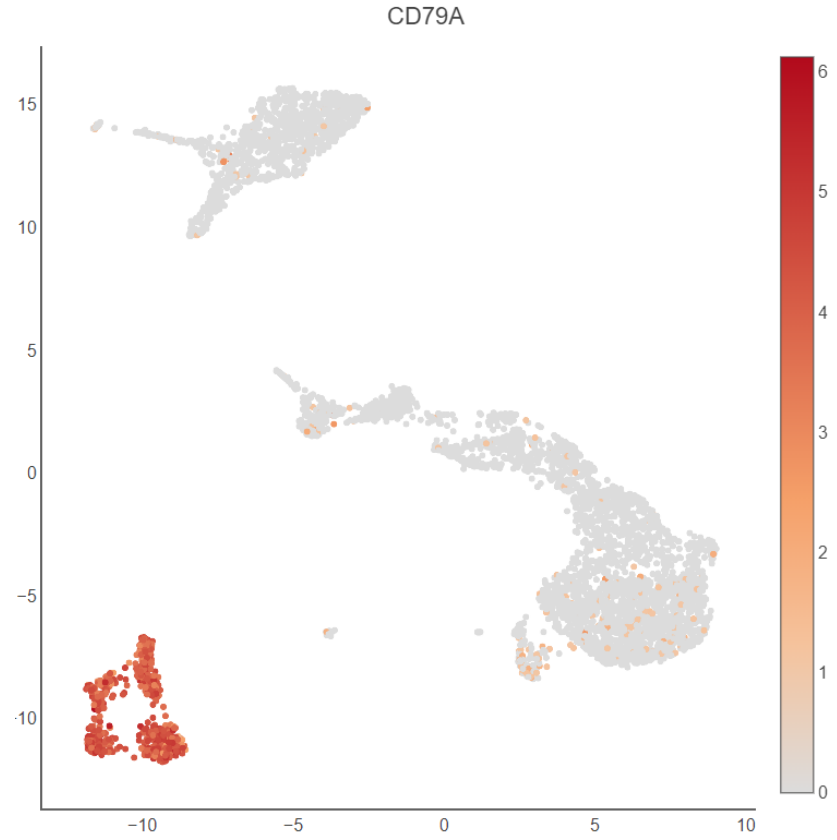


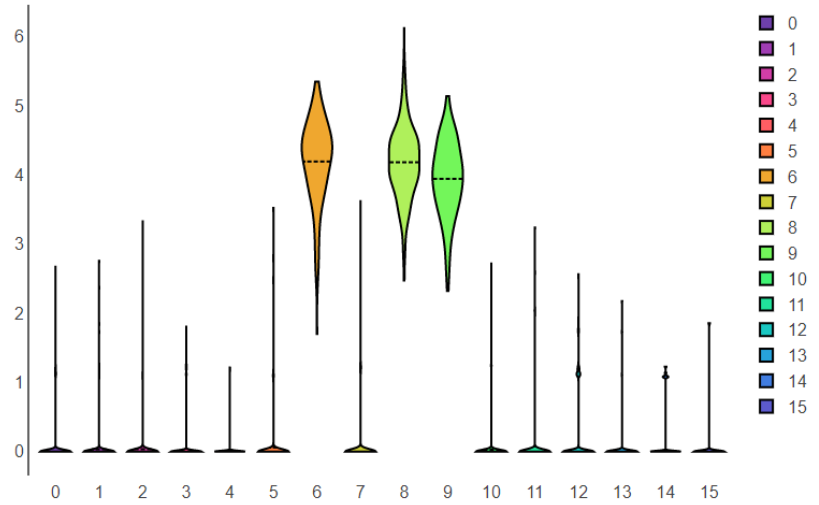
Violin plot

- Violin plot shows **distribution** of gene expression within several groups of cells (in our case groups are clusters)
- Higher the violin – higher the expression in the group

Cd79a:
expression scatter and expression violin







Markers

- Usually, we run differential expression to identify cluster markers
- You can compare a cluster against all the other clusters and identify genes that have higher expression than in the other clusters

Markers tab

scNavigator: beta | 10x_5k_pbmc

- Overview
- Histogram / Bar plot
- Expression scatter plot
- Expression violin plot
- Pathway / Gene set plot
- Markers
- Files

Choose the table

markers

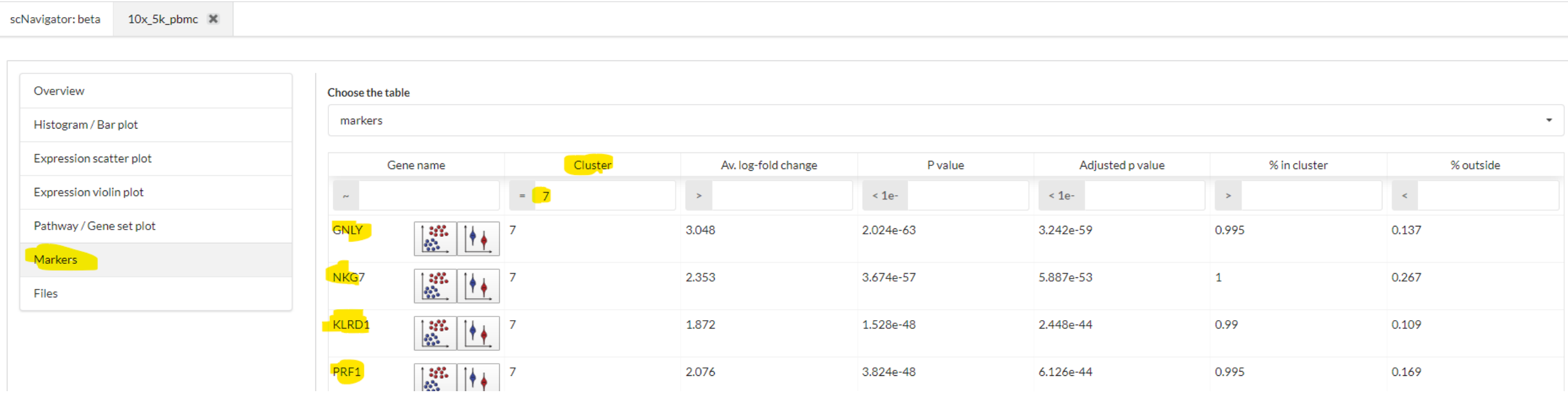
Gene name	Cluster	Av. log-fold change	P value	Adjusted p value	% in cluster	% outside
~	=	>	< 1e-	< 1e-	>	<
RPS3A	0	0.5289	9.674e-25	1.55e-20	1	0.999
SARAF	0	0.6701	5.816e-24	9.318e-20	0.999	0.914
RPL30	0	0.489	8.246e-23	1.321e-18	1	0.999
RPL35A	0	0.4283	1.383e-22	2.216e-18	1	1
RPS16	0	0.4549	2.21e-22	3.54e-18	1	0.999
RPL9	0	0.4875	5.119e-22	8.202e-18	1	1
RPL21	0	0.4461	1.053e-21	1.687e-17	1	1
LDHB	0	0.7275	1.466e-21	2.348e-17	0.998	0.772
RPS15A	0	0.464	4.221e-21	6.763e-17	1	1
TRABD2A	0	0.7406	6.847e-20	1.097e-15	0.822	0.233

Previous | Page 1 of 646 | 10 rows | Next

Download current table

ITMO UNIVERSITY

Markers tab: what's the cluster 7?



- GNLY – gene name
- Cluster 7 – we are checking results for cluster 7 vs other clusters
- Average log-fold change: average difference between expression of GNLY in cluster 7 and in other clusters
- P value (we test difference between average expression of this gene inside and outside cluster 7)
- P adjusted – adjusted p value for multiple hypothesis
- % in and outside of the cluster – in how many cells GNLY is detected in cluster 7 and in other clusters

ITMO UNIVERSITY

Markers tab: what's the cluster 7?

- You have two buttons next to the gene name

1) First will open gene expression on scatter plot
2) Second will open gene expression on violin plot

Choose the table

markers

Gene name	Cluster	Av. log-fold change	P value	Adjusted p value	% in cluster	% outside
~	= 7	>	< 1e-	< 1e-	>	<
GNLY	7	3.048	2.024e-63	3.242e-59	0.995	0.137
NKG7	7	2.353	3.674e-57	5.887e-53	1	0.267

Now let's play with it

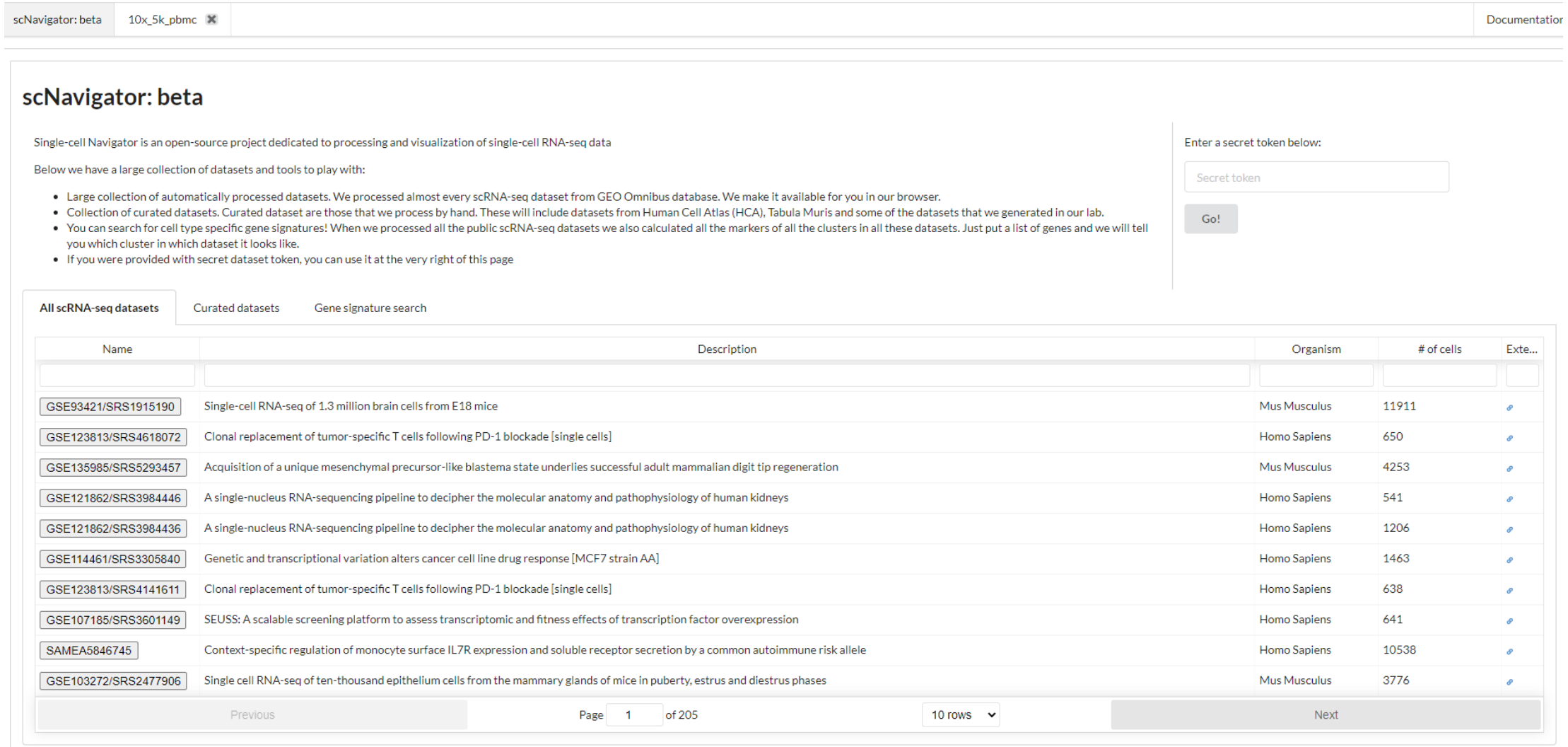
- I want you to check out any other genes

Public datasets

- We try to process many other public datasets trying to make them available to scientific community
- You can always go back to the main tab (top left corner)

ITMO UNIVERSITY

Public datasets



Public scRNA-seq datasets

Most of the scRNA-seq datasets are available at NCBI GEO (or SRA)

Problems are:

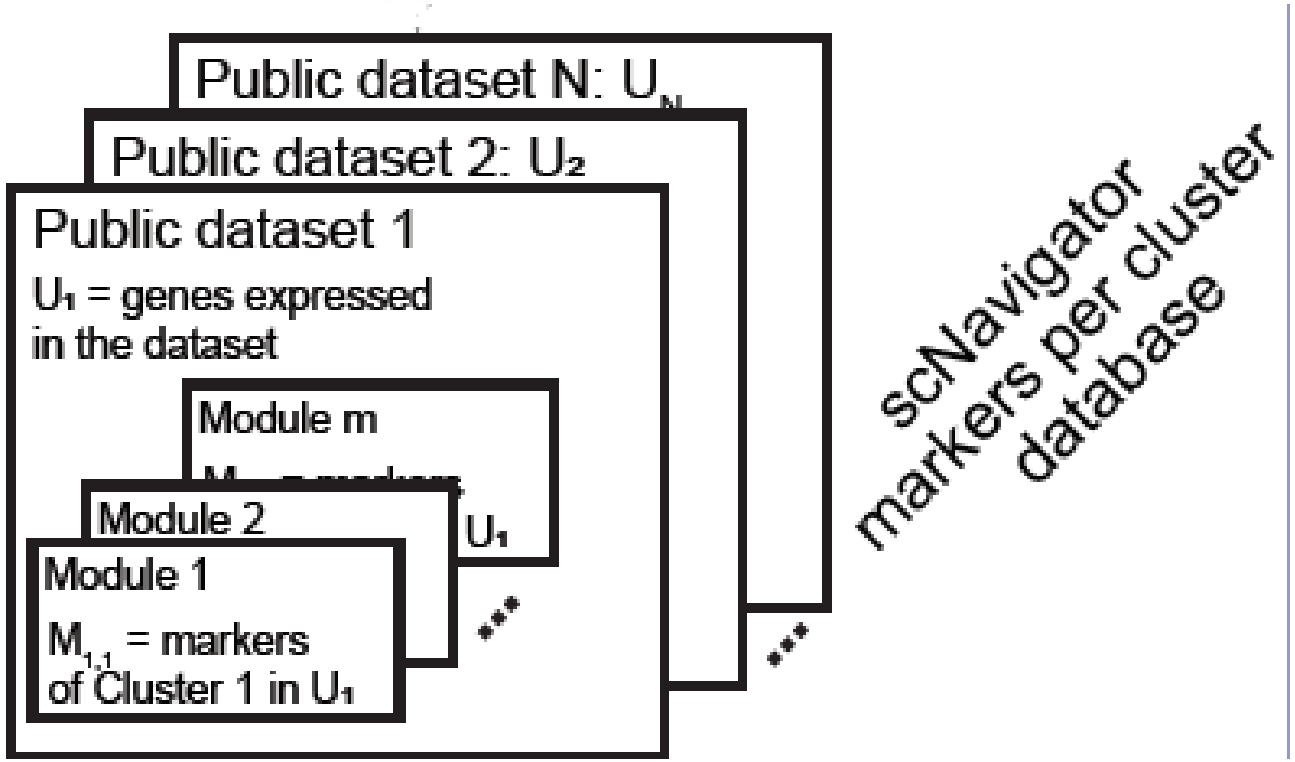
- Different technologies used to perform experiment (10x, DropSeq, SmartSeq2, C1 Fluidigm etc)
- Different pipelines were used to analyze
- Different formats in which data is kept

Most of the dataset processing was done by Maria Firuleva

Lessons from GeneQuery

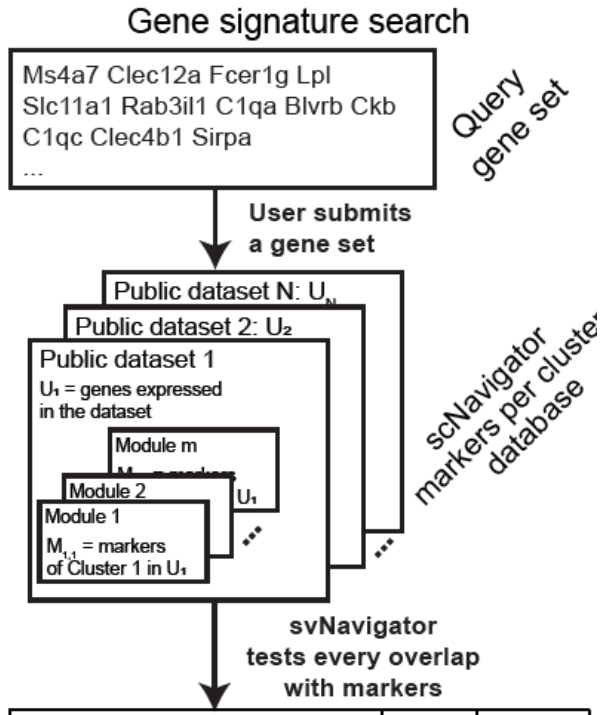
- We can test gene signatures against other public data
- We can use public scRNA-seq data: we can calculate markers for all populations found in the data

GeneQuery (but for single-cell)



GeneQuery (but for single-cell)

- Query Gene Set:
  - List of your DE genes
  - List of markers from your single cell
  - Gene related to pathways / biological processes
  - Any other gene set
- Calculating overlaps with markers from all public scRNA-seq dataset
- Report p-values and adjusted p-value for all the dataset with the overlap



Dataset	Cluster	adj. p value
10X single-cell RNASeq profiling of immune infiltrate from Adar1 KO and Wildtype B16 Melanoma	1	9.62E-52
A single-cell atlas of mouse brain macrophages reveals unique transcriptional identities shaped by ontogeny and tissue environment.	3	6.10E-37
Single cell RNA sequencing analysis of mouse resident and infiltrating immune cells during acute kidney injury and fibrosis	1	8.31E-36
Single cell sequencing of mouse syngeneic tumor models	5	1.58E-31
...	...	...

Let’s test it:

Human monocyte markers:

CTSS

S100A9

S100A8

LYZ

S100A6

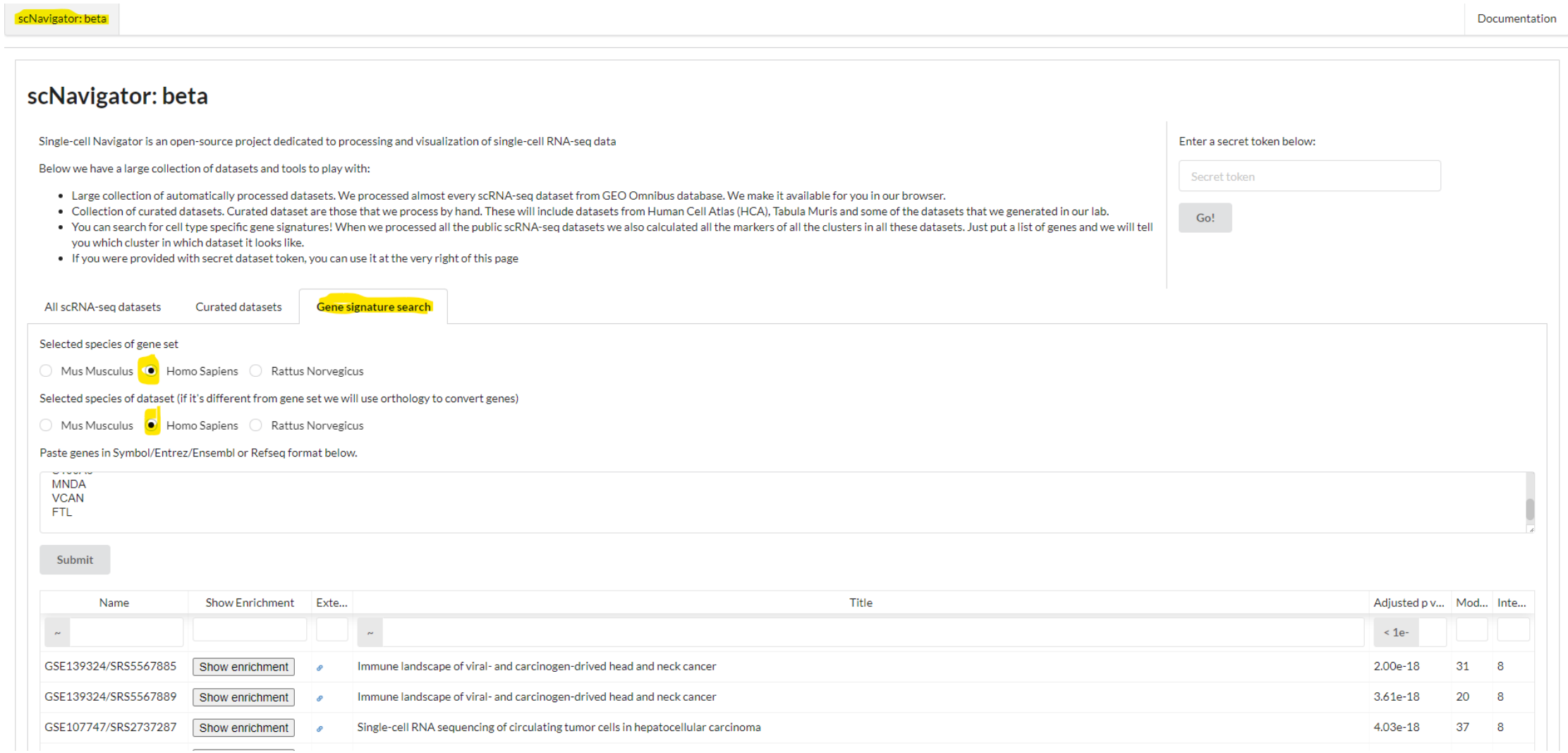
MNDA

VCAN

FTL

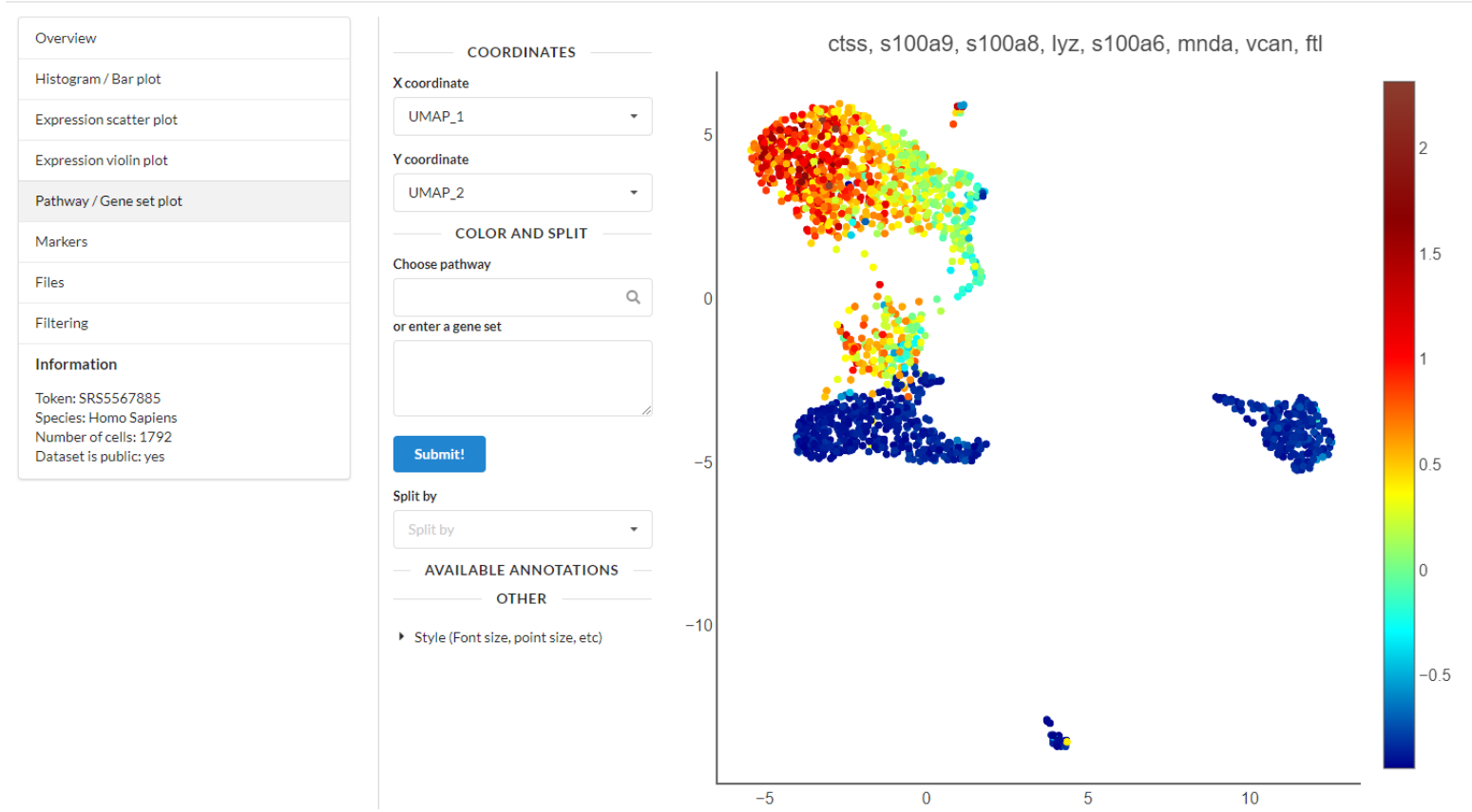
ITMO UNIVERSITY

Let’s test it:



Let’s test it, “Show enrichment”

- Averaged expression across all the cells
- We can clearly see upregulated expression of these genes in a population
- We can guess these cells are monocytes



Conclusion

- We hope that scNavigator will make interpretation of scRNA-seq data easier
- https://artyomovlab.wustl.edu/scn/

- We try to get there as much datasets as we can
- If you want to use scNavigator for your private data:
  - Check out the docs on the file format https://artyomovlab.wustl.edu/scn/docs/, and send converted data to me
  - Email me if you have any questions: kzaitsev@itmo.ru