ITMO UNIVERSITY

CT Lab
ITMO UNIVERSITY

Introduction into Single-Cell Explorer (SCE)

Konstantin “Kostya” Zaitsev, ITMO University

Systems biology workshop, Nice, Sep 22th

Visualizing scRNA-seq data

Main goals:

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

Extra goals:

- Fast
- Responsive

Visualizing scRNA-seq data

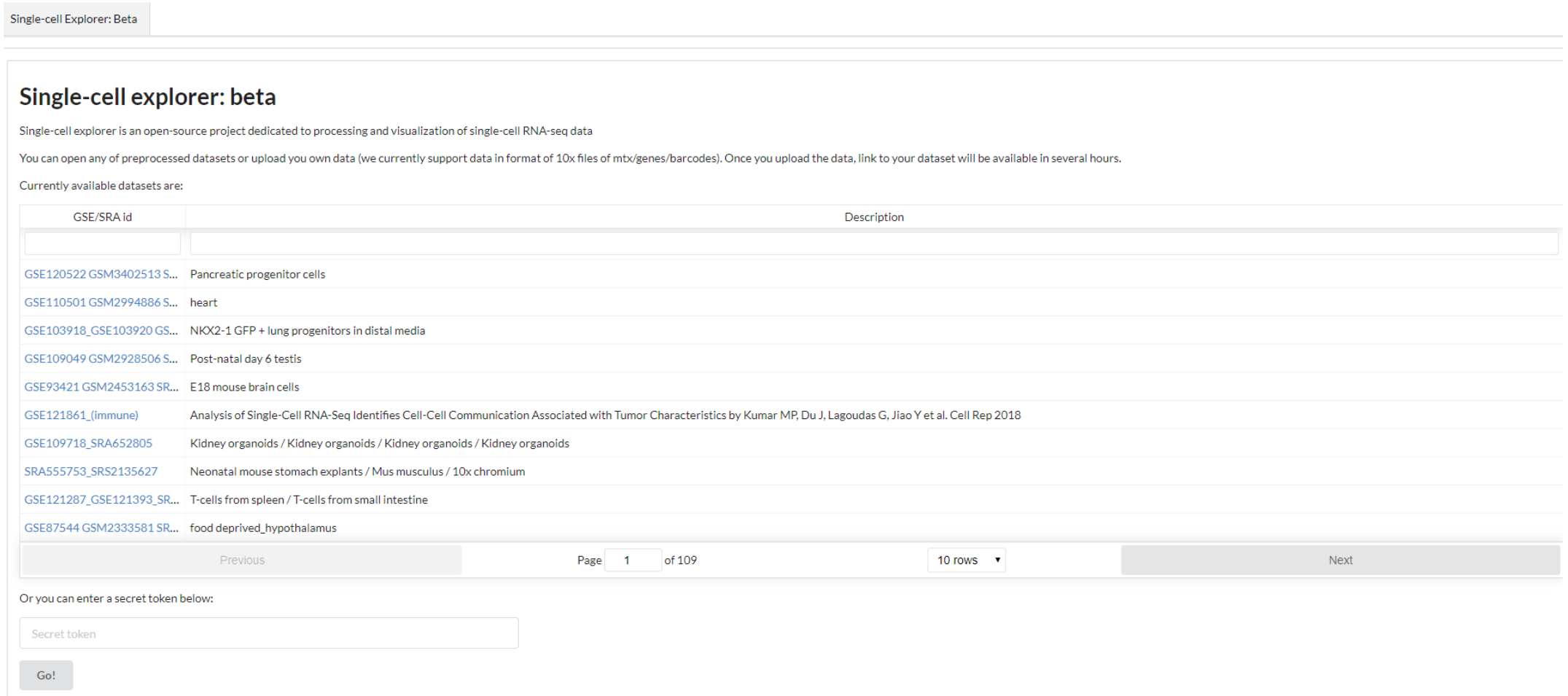
https://artyomovlab.wustl.edu/sce/

(still in production, so feedback is very welcome)

ITMO UNIVERSITY

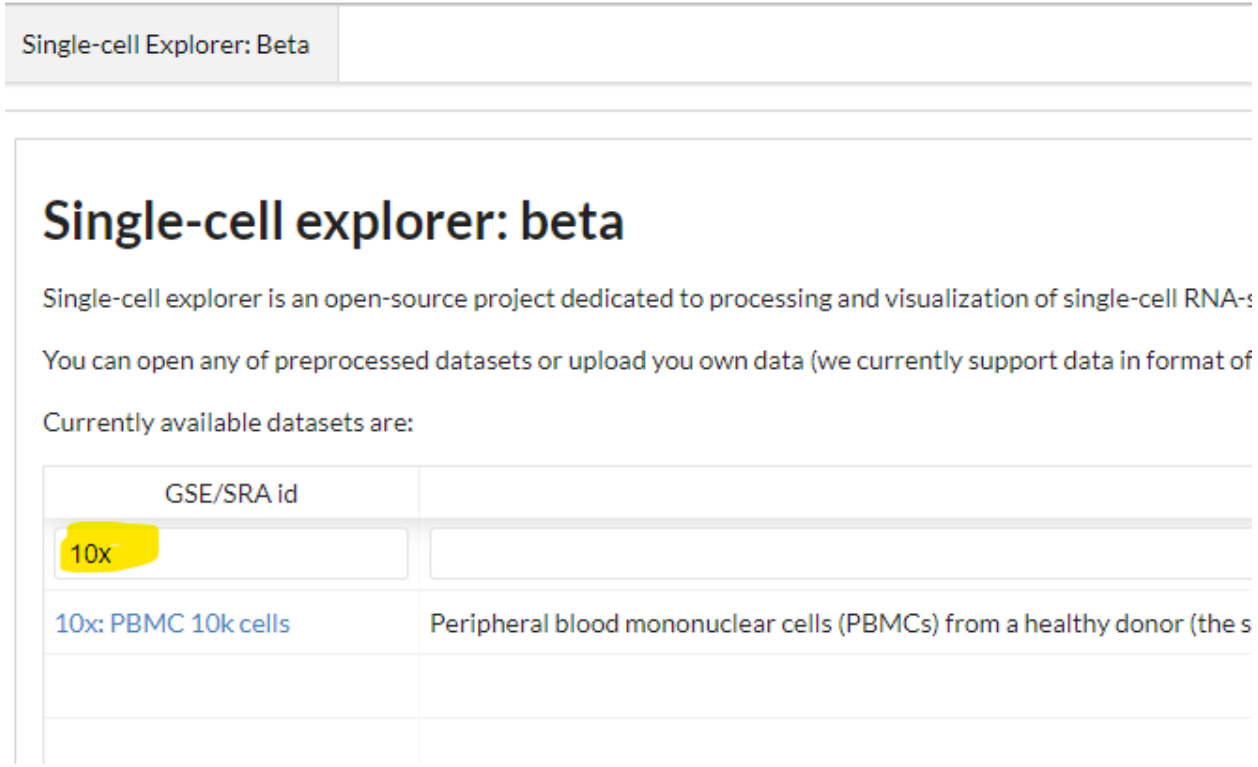
Let’s open the dataset

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



Let’s open the dataset

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

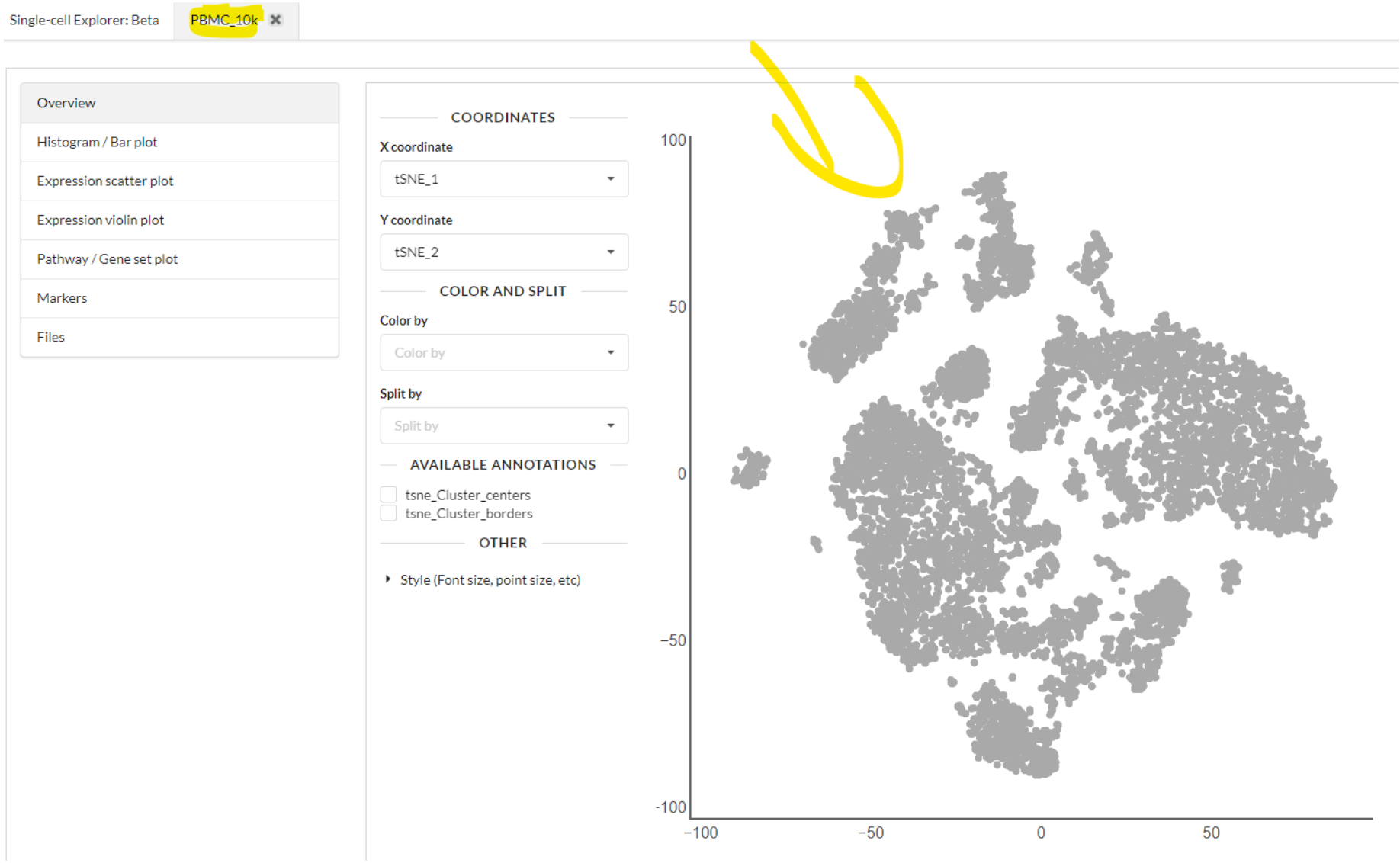


If you have any problem finding dataset

- Just go to https://artyomovlab.wustl.edu/sce/?token=PBMC_10k

ITMO UNIVERSITY

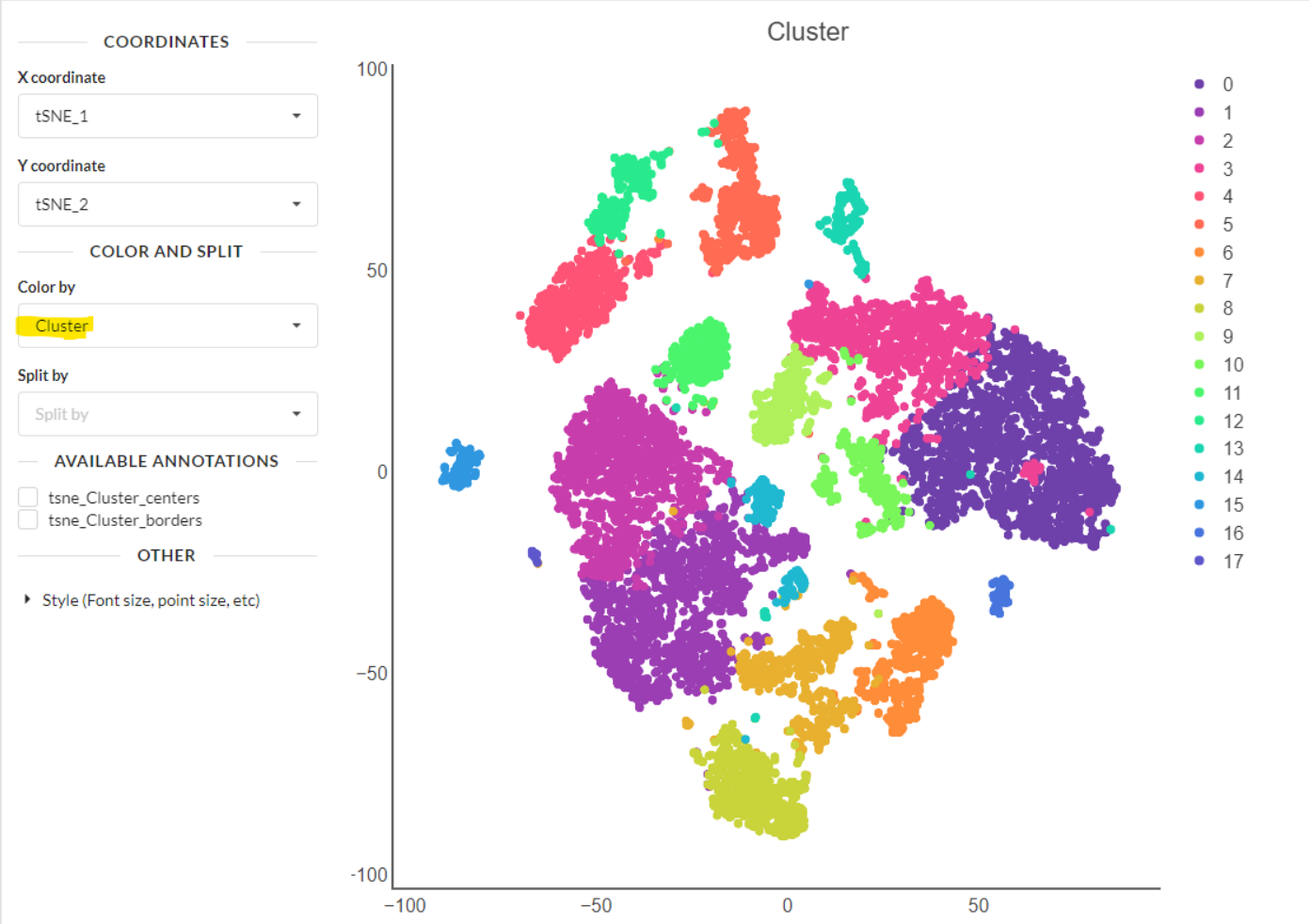
Result should look like that



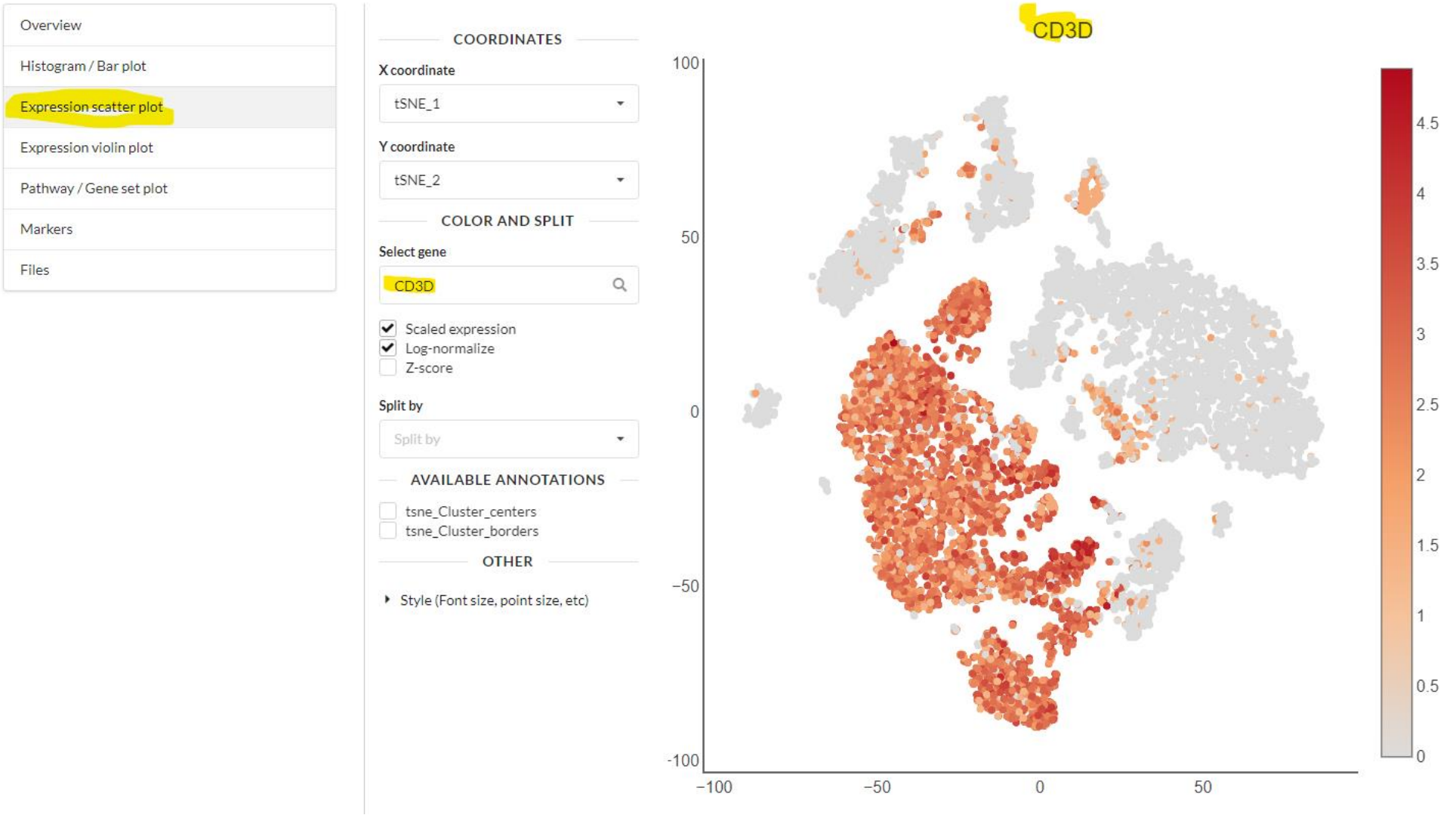
We can color the cells

- Cluster
- Number of UMIs
- Number of genes detected

- tsne_Cluster_centers

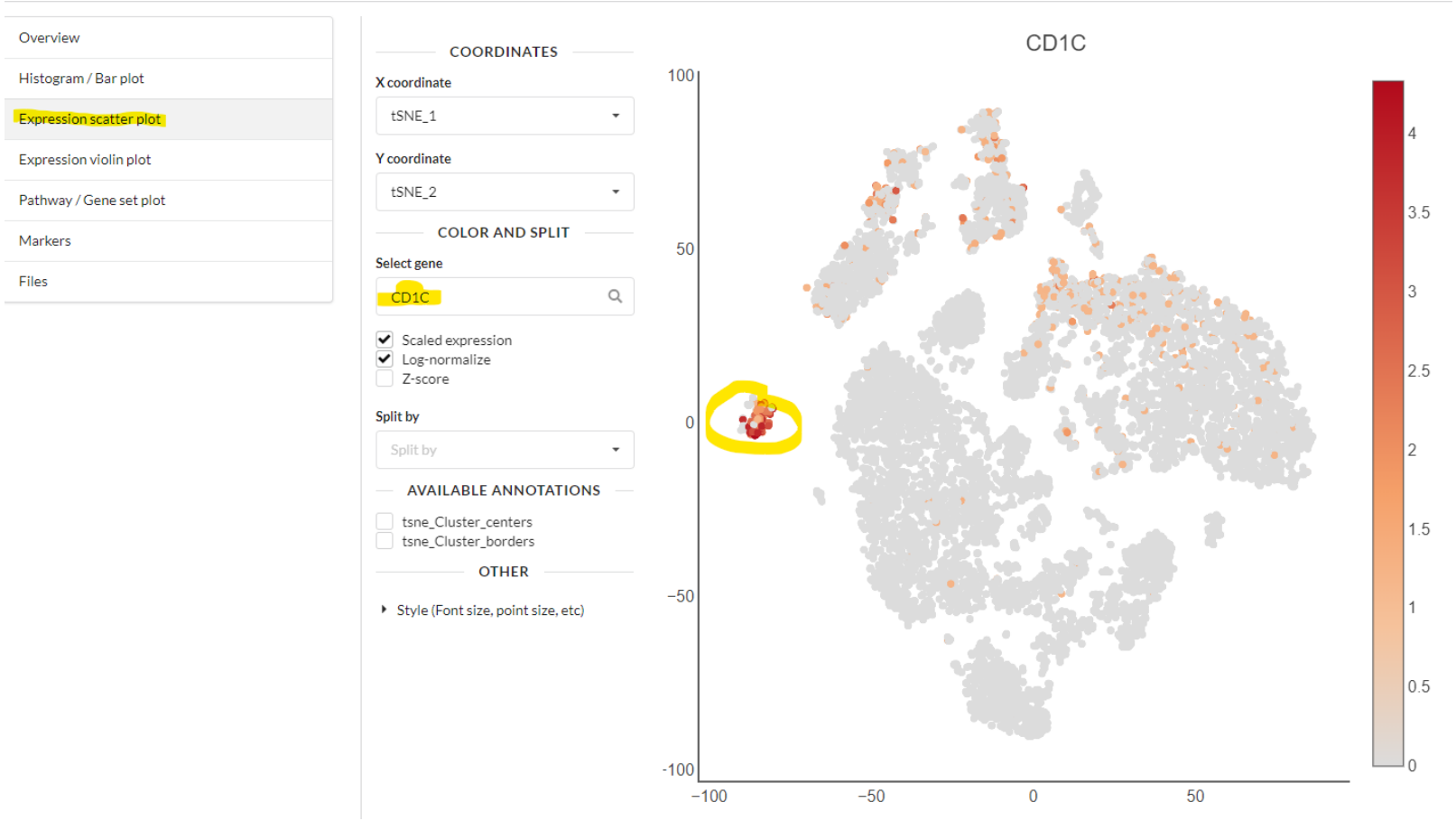


Expression of CD3d



ITMO UNIVERSITY

Or you can go for any of your favorite genes

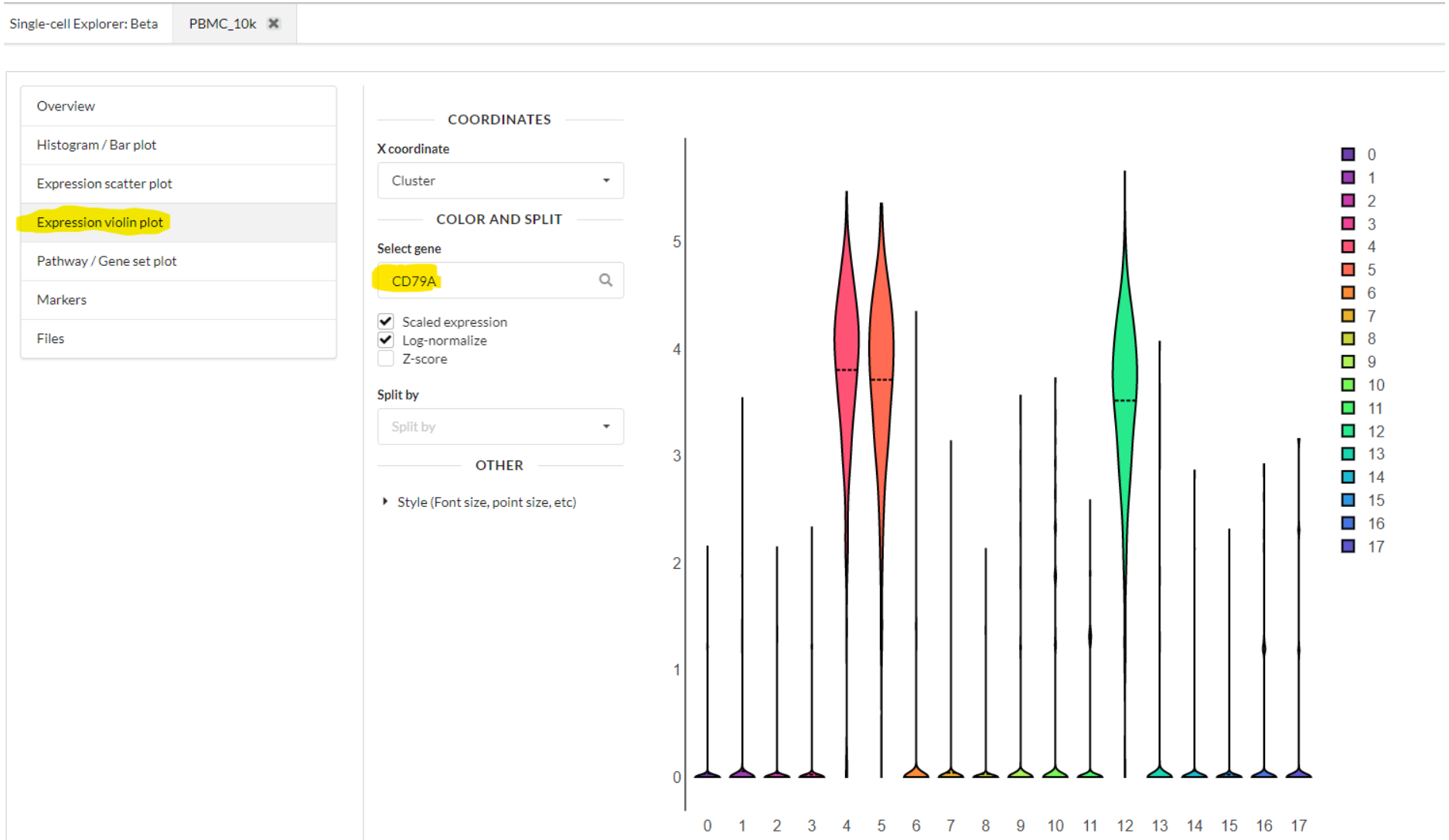


Expression scatter plot

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

ITMO UNIVERSITY

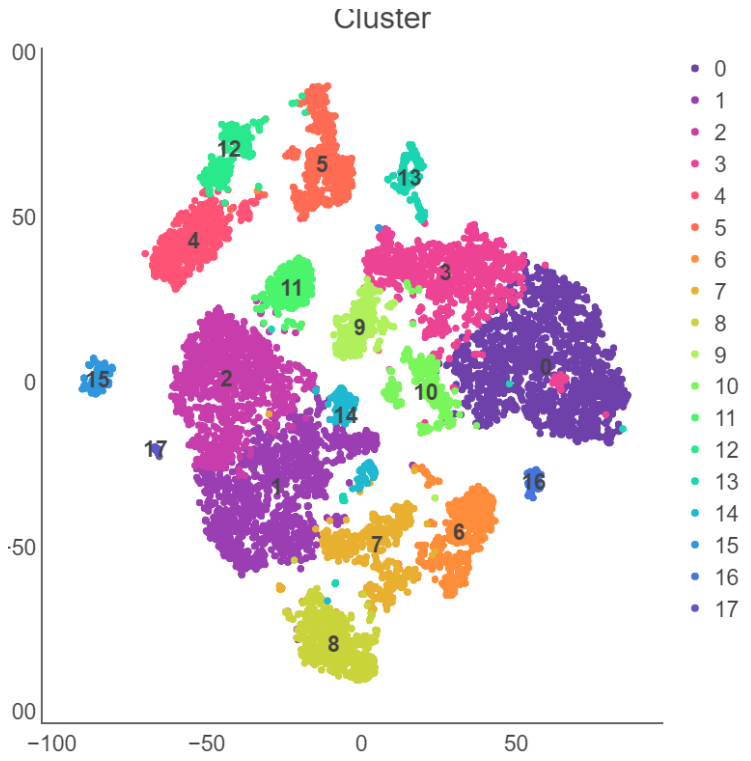
Violin plot

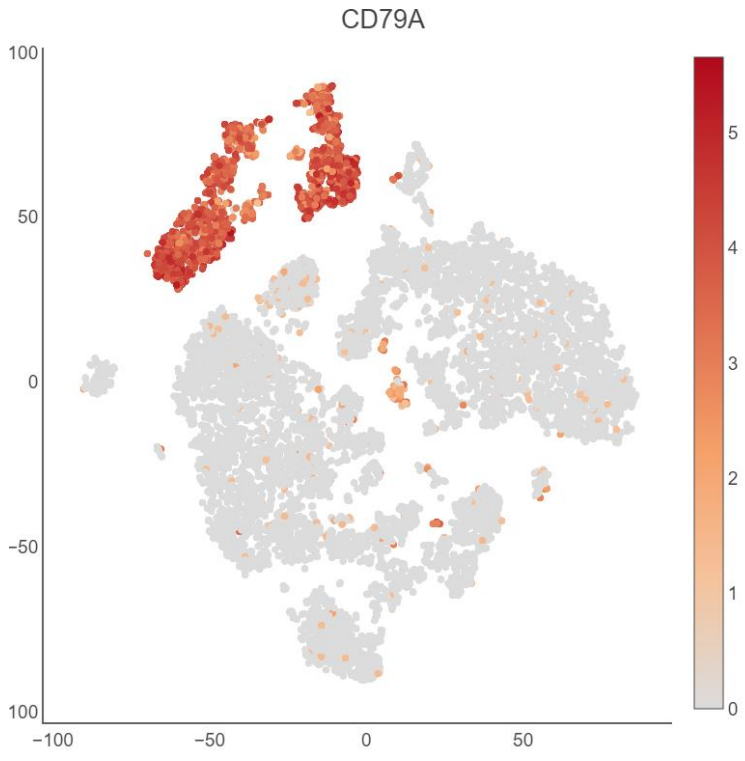


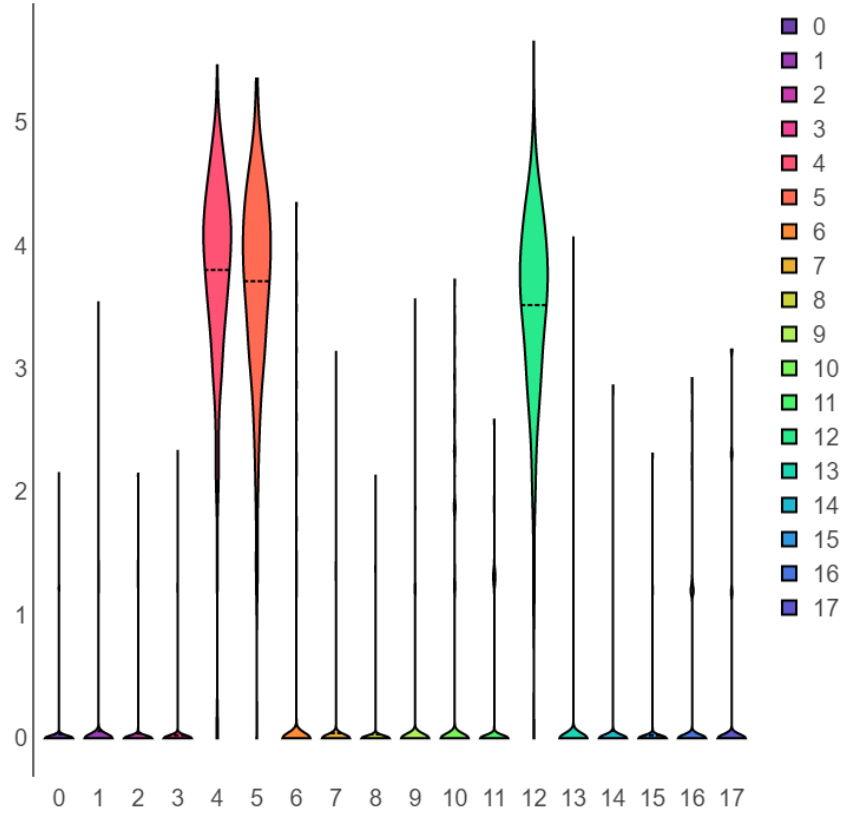
Violin plot

- 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

ITMO UNIVERSITY

Markers tab

Single-cell Explorer: Beta | PBMC_10k

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

Choose the table

Cluster

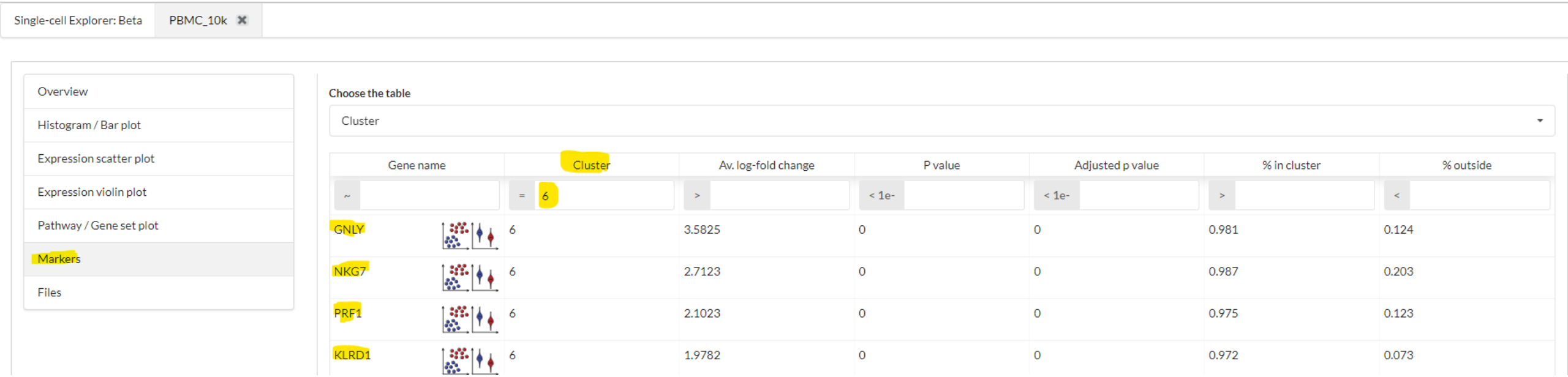
Gene name	Cluster	Av. log-fold change	P value	Adjusted p value	% in cluster	% outside
~	=	>	< 1e-	< 1e-	>	<
S100A8	0	2.6693	0	0	1	0.594
S100A9	0	2.4105	0	0	1	0.702
S100A12	0	2.2626	0	0	1	0.275
LYZ	0	1.8552	0	0	1	0.749
VCAN	0	1.8376	0	0	0.998	0.277
MNDA	0	1.6095	0	0	1	0.315
FCN1	0	1.53	0	0	1	0.332
FOS	0	1.3692	0	0	1	0.965
CTSS	0	1.3573	0	0	1	0.713
CD14	0	1.3368	0	0	0.968	0.201

Previous | Page 1 of 678 | 10 rows | Next

Download current table

ITMO UNIVERSITY

Markers tab: what's the cluster 6?

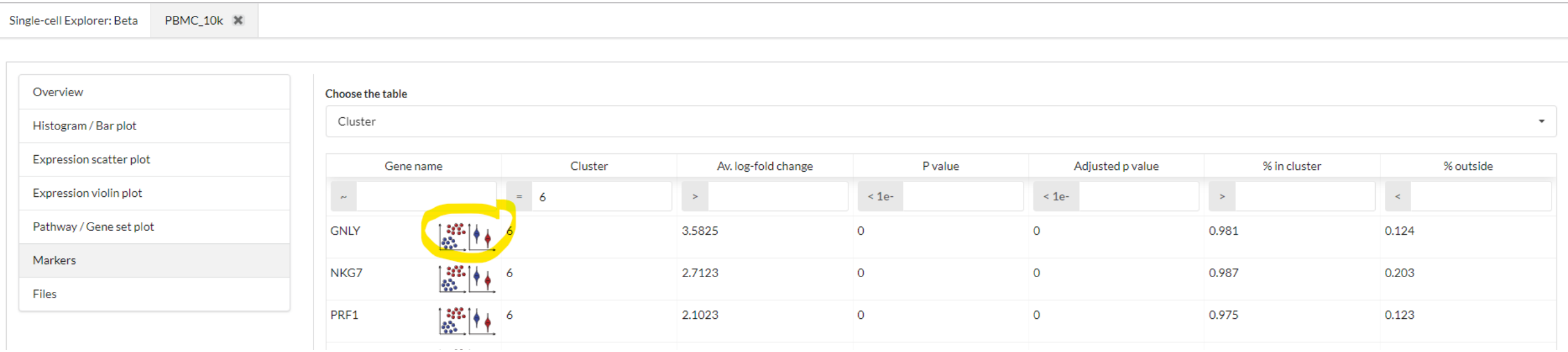


- GNLY – gene name
- Cluster 6 – we are checking results for cluster 6 vs other clusters
- Average log-fold change: average difference between expression of GNLY in cluster 6 and in other clusters
- P value is p value :D
- P adjusted – adjusted p value for multiple hypothesis
- % in and outside of the cluster – in how many cells GNLY is detected in cluster 6 and in other clusters

Markers tab: what's the cluster 6?

- 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

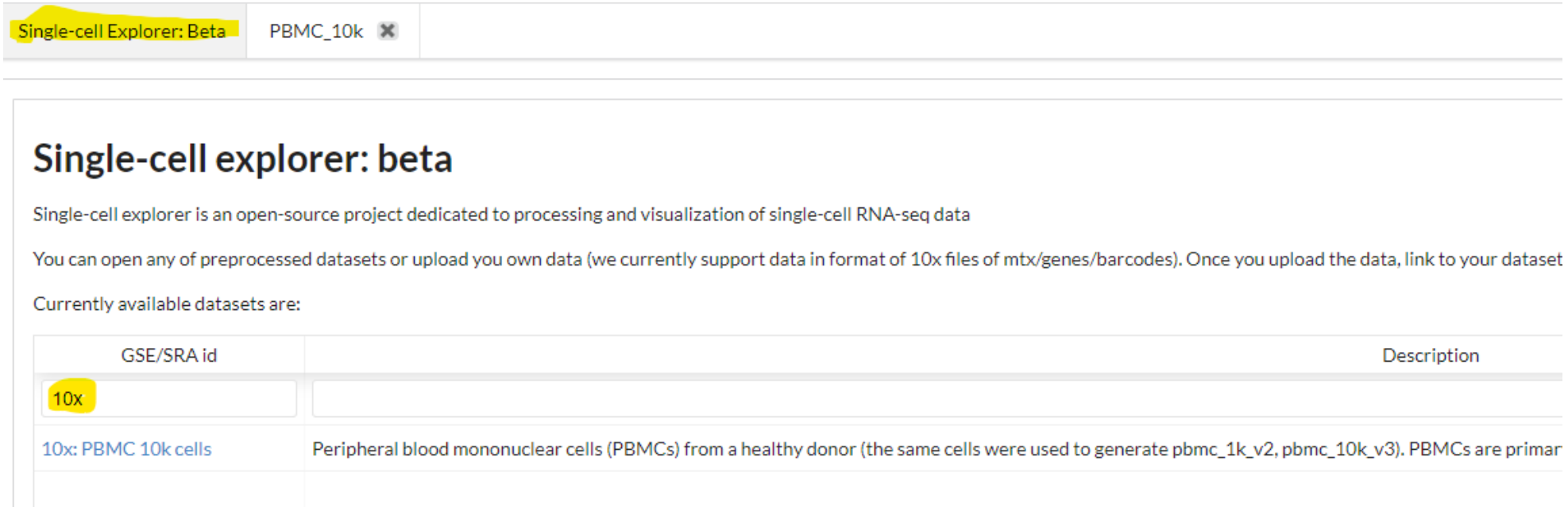


Now let's play with it

- I want you to check out any other genes

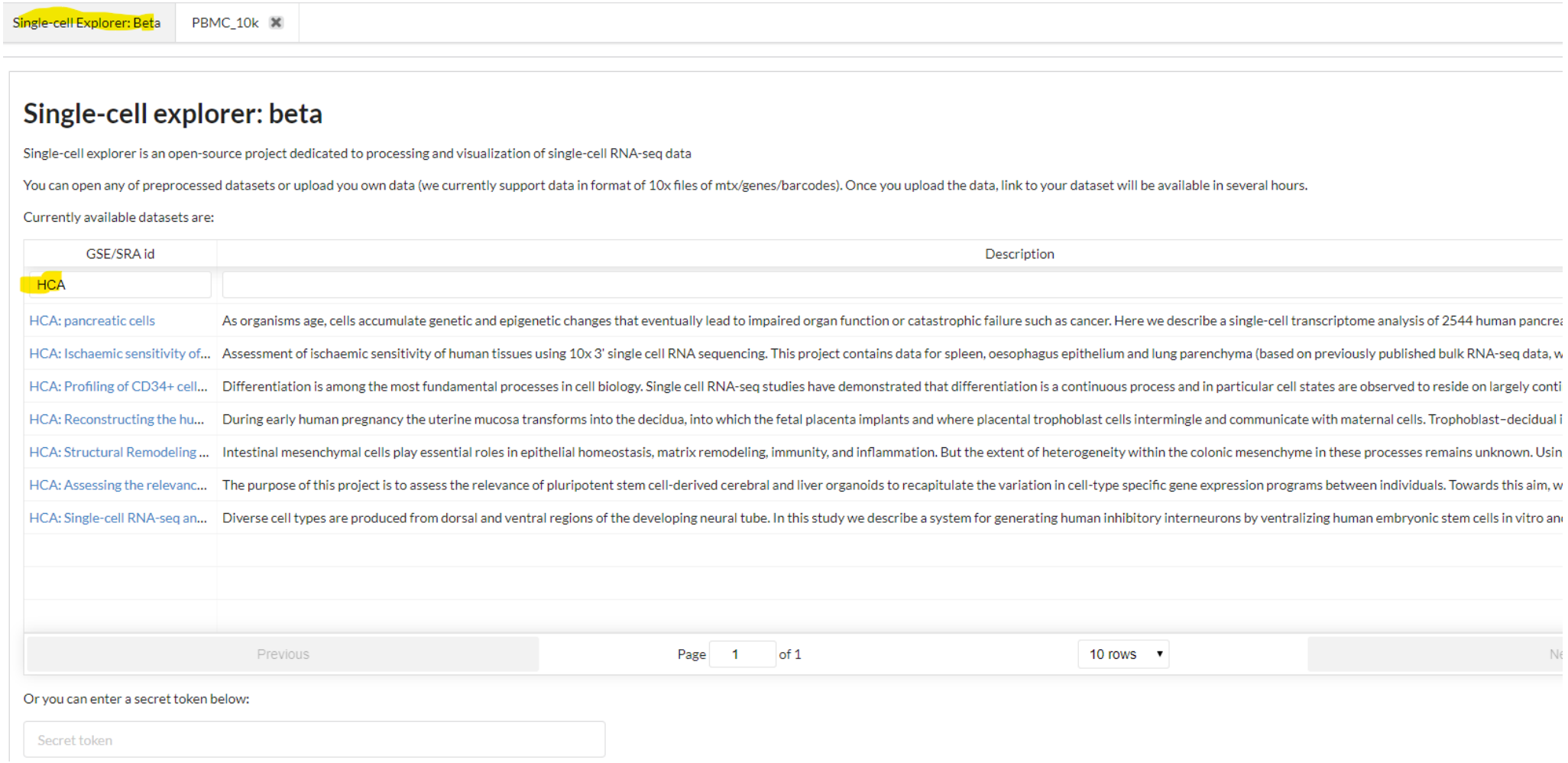
Public datasets

- We try to process many other public datasets trying to make them available to scientific community
- Right now we processed around 1100 of different scRNA-seq datasets
- You can always go back to the main tab (top left corner)



ITMO UNIVERSITY

Public datasets
including datasets from Human Cell Atlas



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

PanglaoDB

- https://panglaodb.se/

Pros:

- They provide count tables for a lot of datasets

Cons:

- Their analysis sometimes has different issues
- Their website is not responsive at al
- A lot of datasets are not present

Datasets at SCE

- Everything from Panglao DB
- We also try to process GEO datasets that are not present in Panglao
- We want to process “milestone” datasets: HCA, Tabula Muris, Mouse Cell Atlas, million mouse brain cells …

ITMO UNIVERSITY

What are the issues

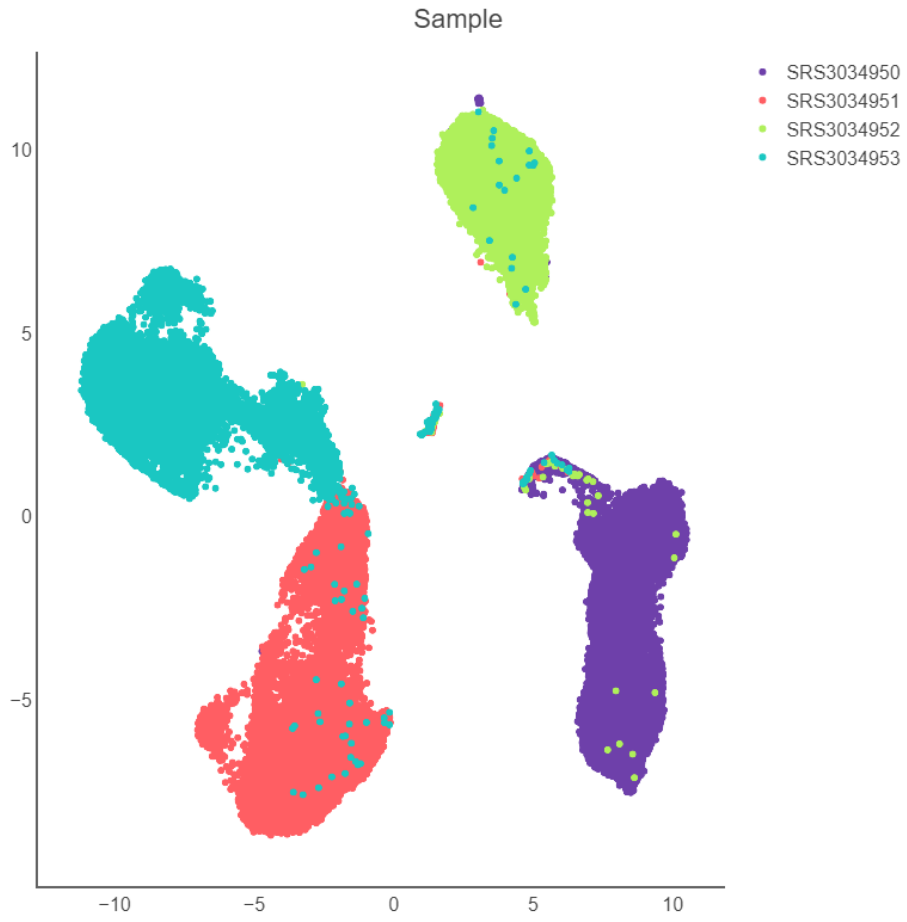
When we first analyzed 1000 dataset two main issues were identified:

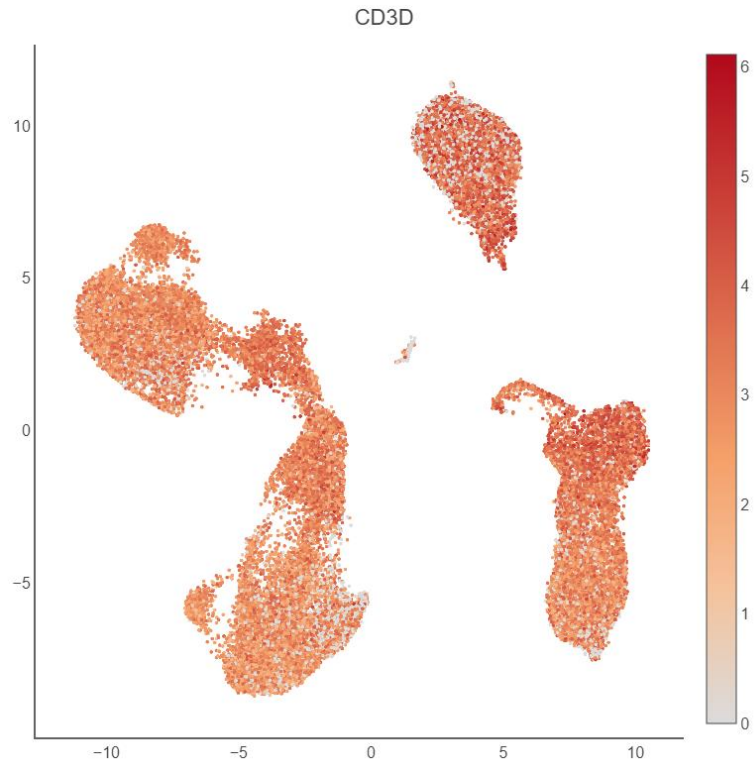
1) Donor effect in human data
2) UMI distribution affects the analysis

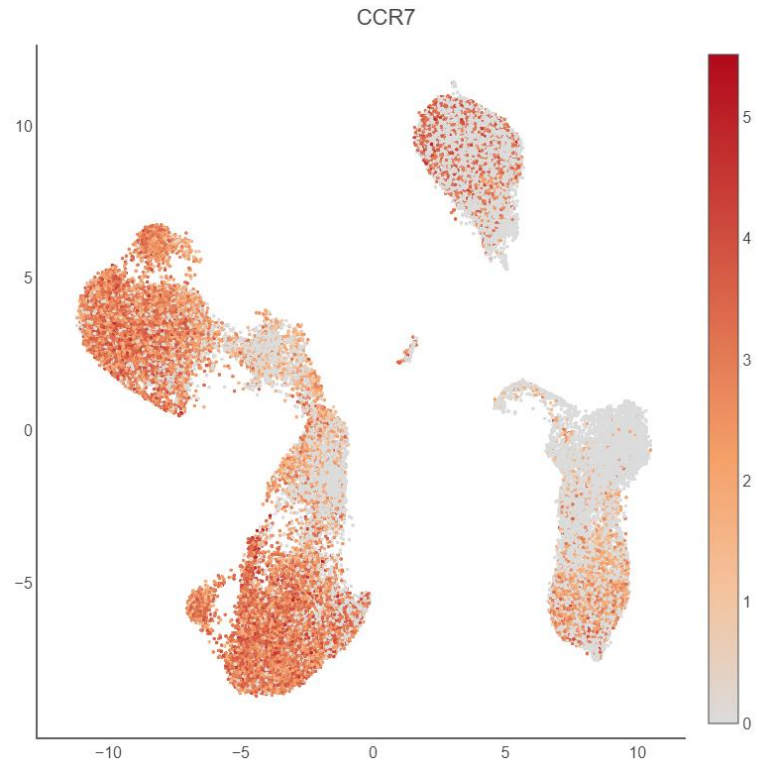
Most of the dataset processing was done by Maria Firuleva

ITMO UNIVERSITY

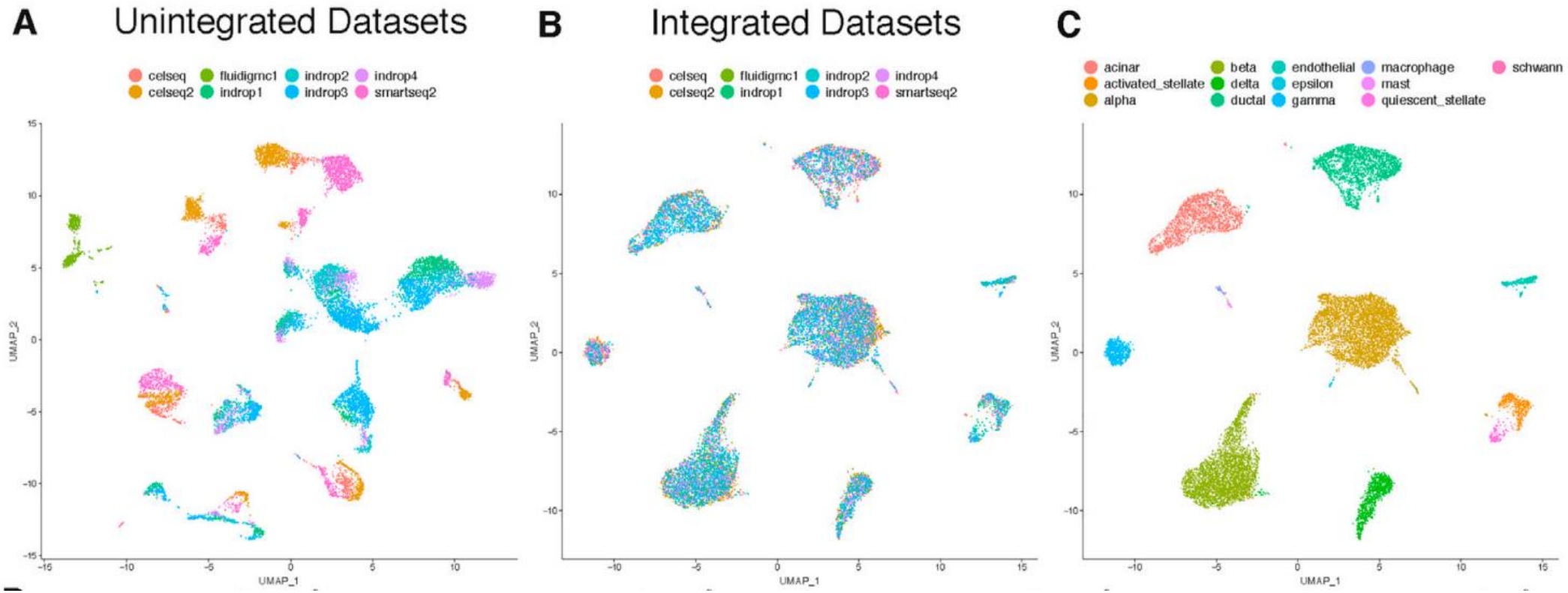
Issues: donor effect







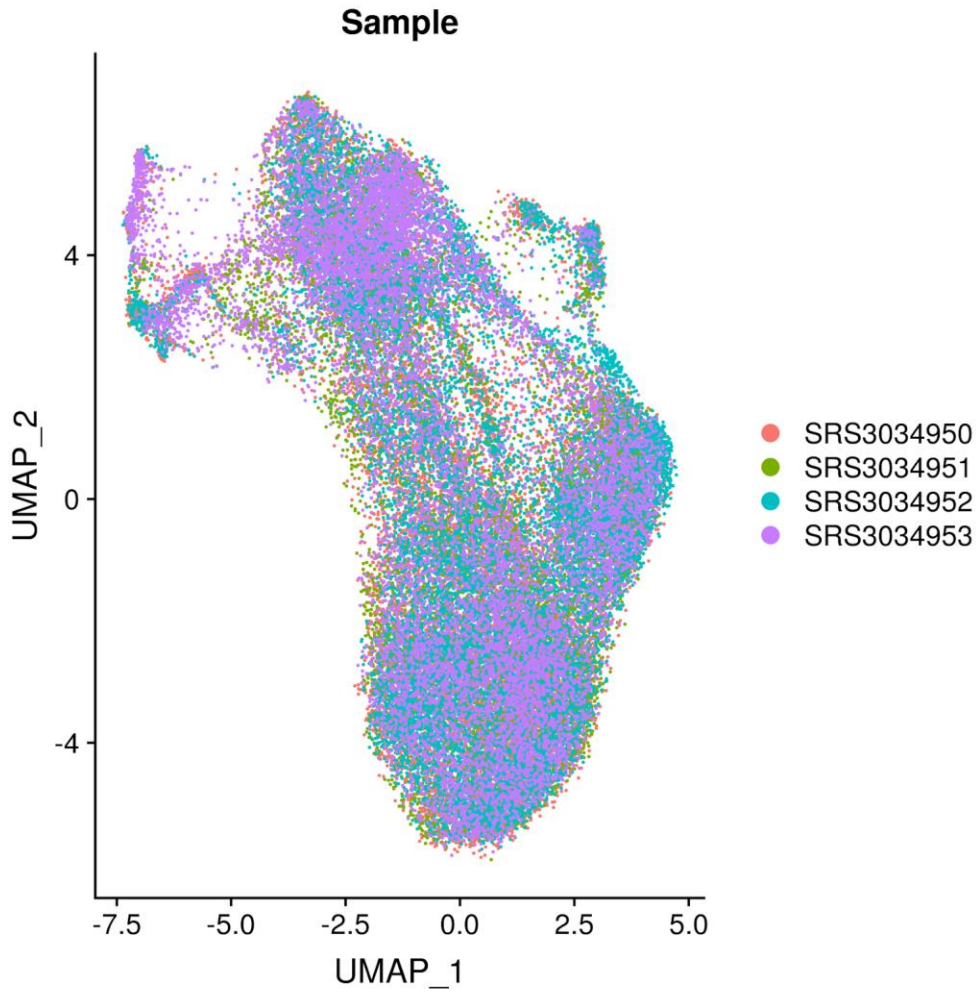
Recent developments of methods



- Taken from https://www.cell.com/cell/pdf/S0092-8674(19)30559-8.pdf

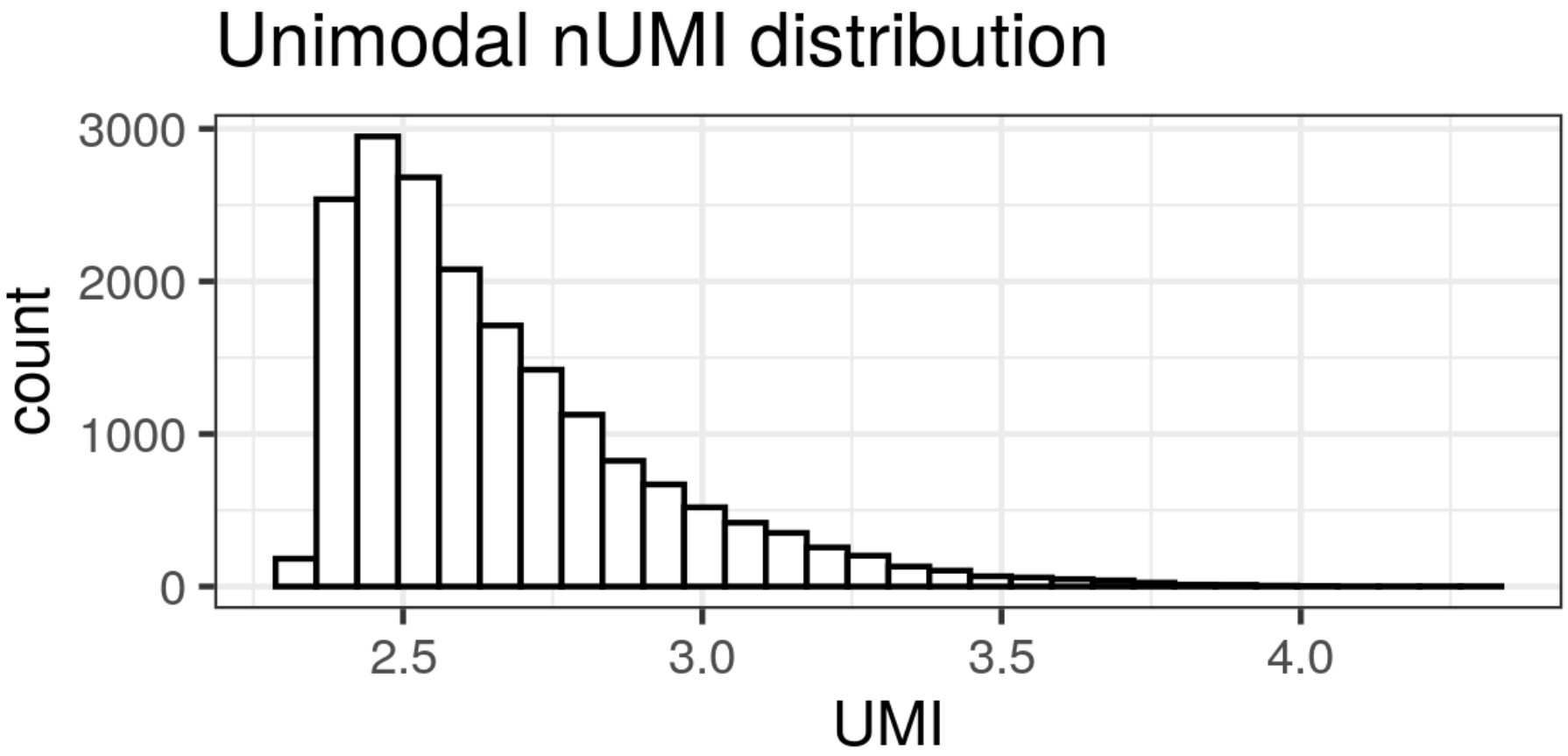
Issues: donor effect

- Integration methods remove batch/donor effects
- Integration methods can be run automatically



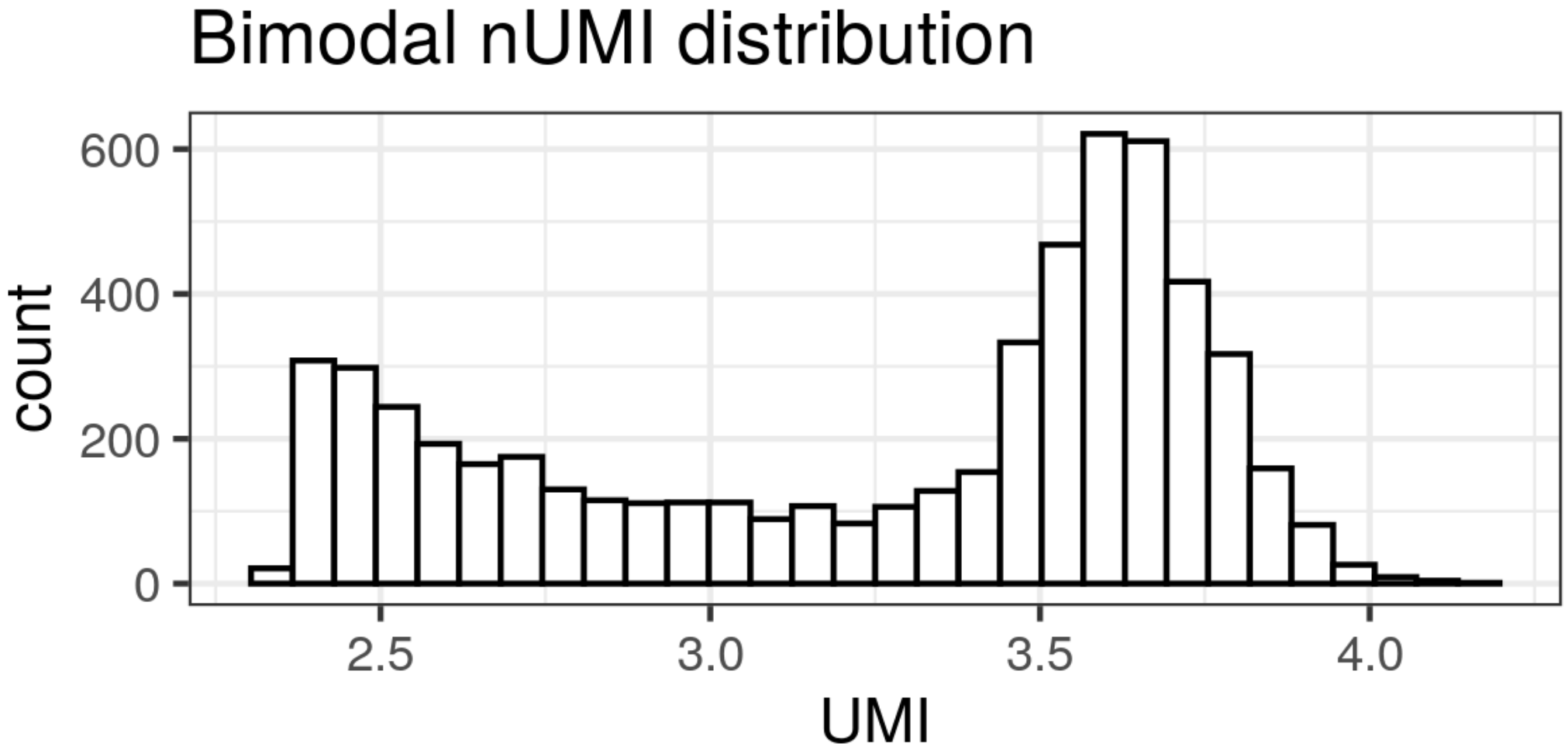
Issues: UMI distribution

Good case scenario



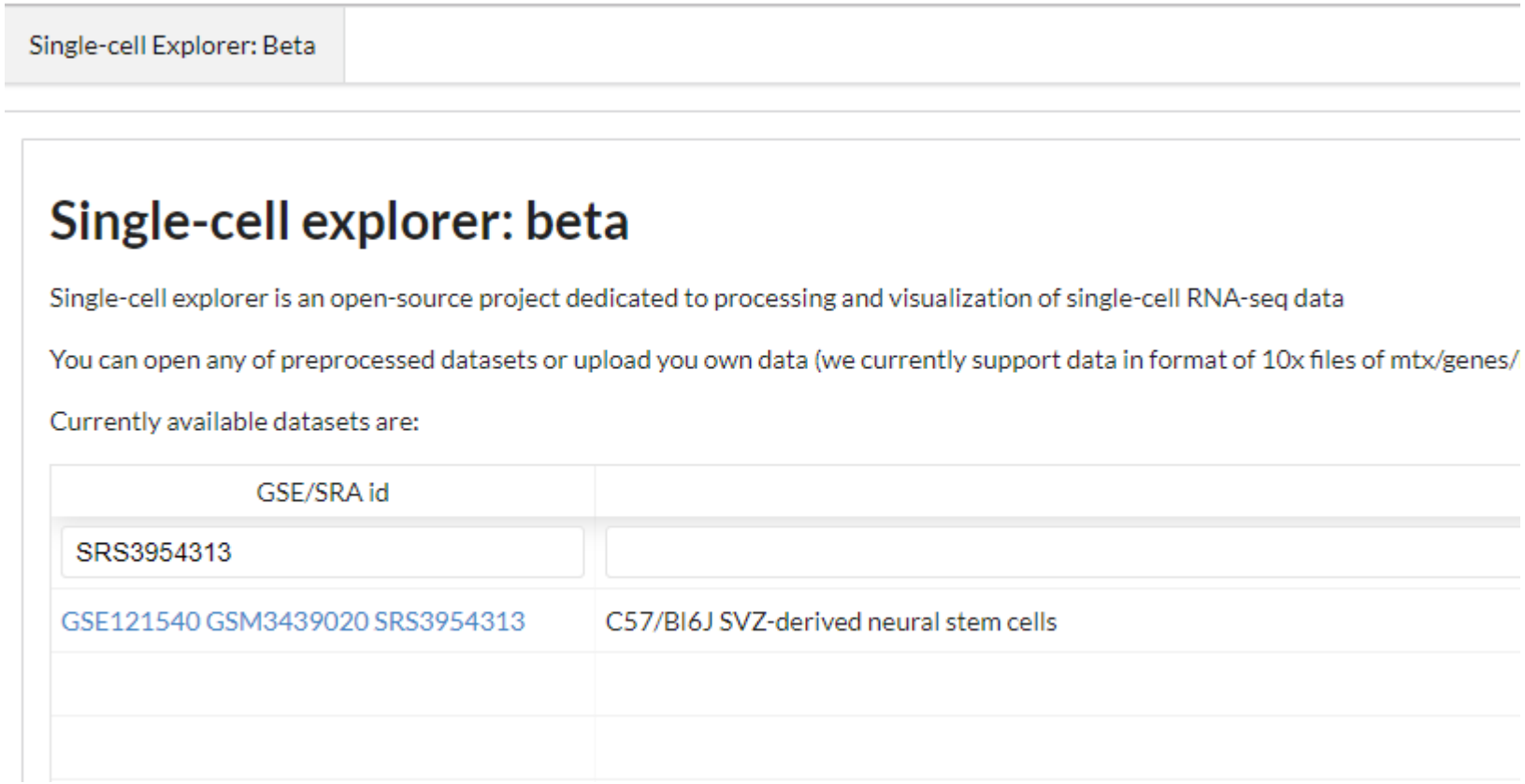
Issues: UMI distribution

Bad case scenario



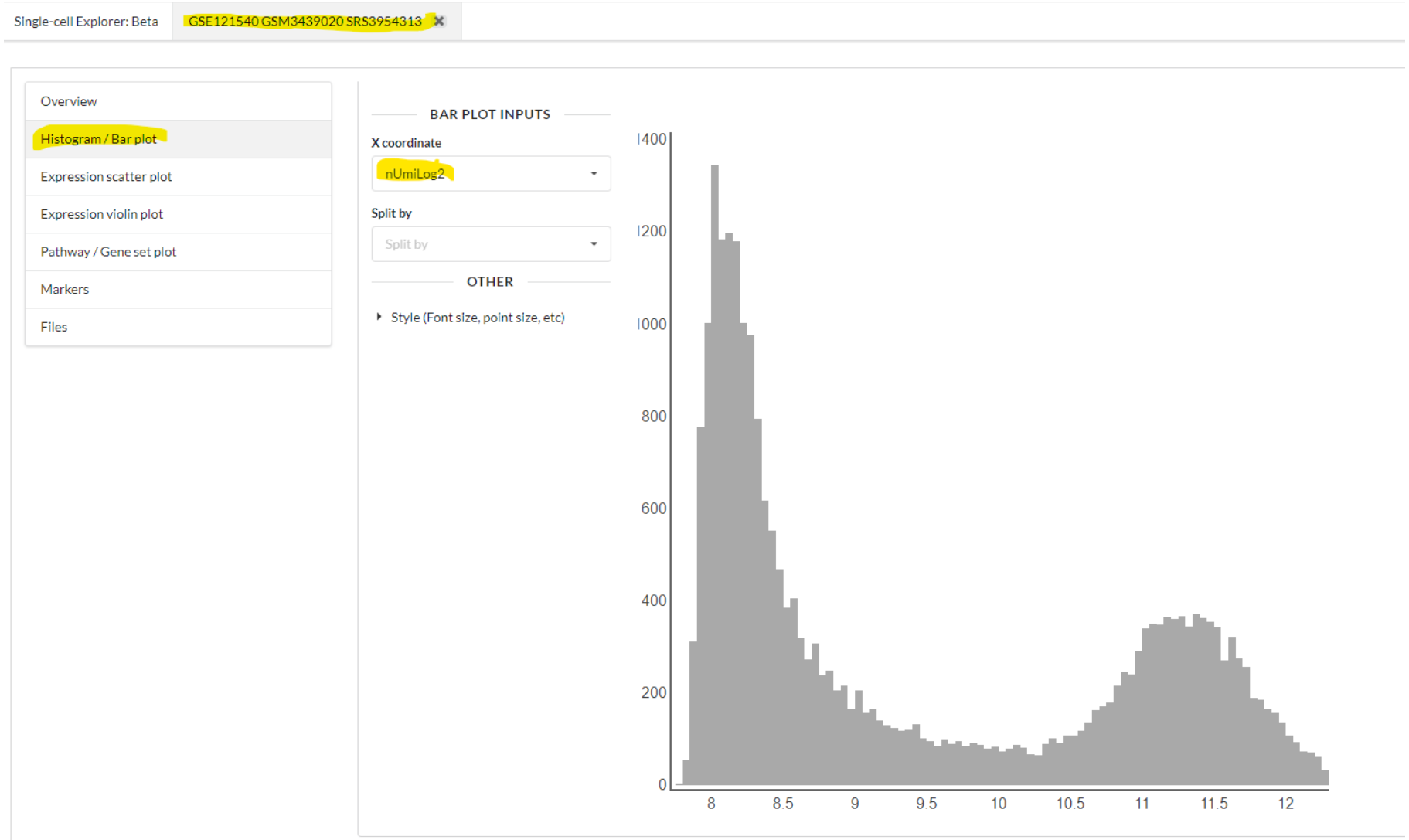
ITMO UNIVERSITY

Let’s open SRS3954313



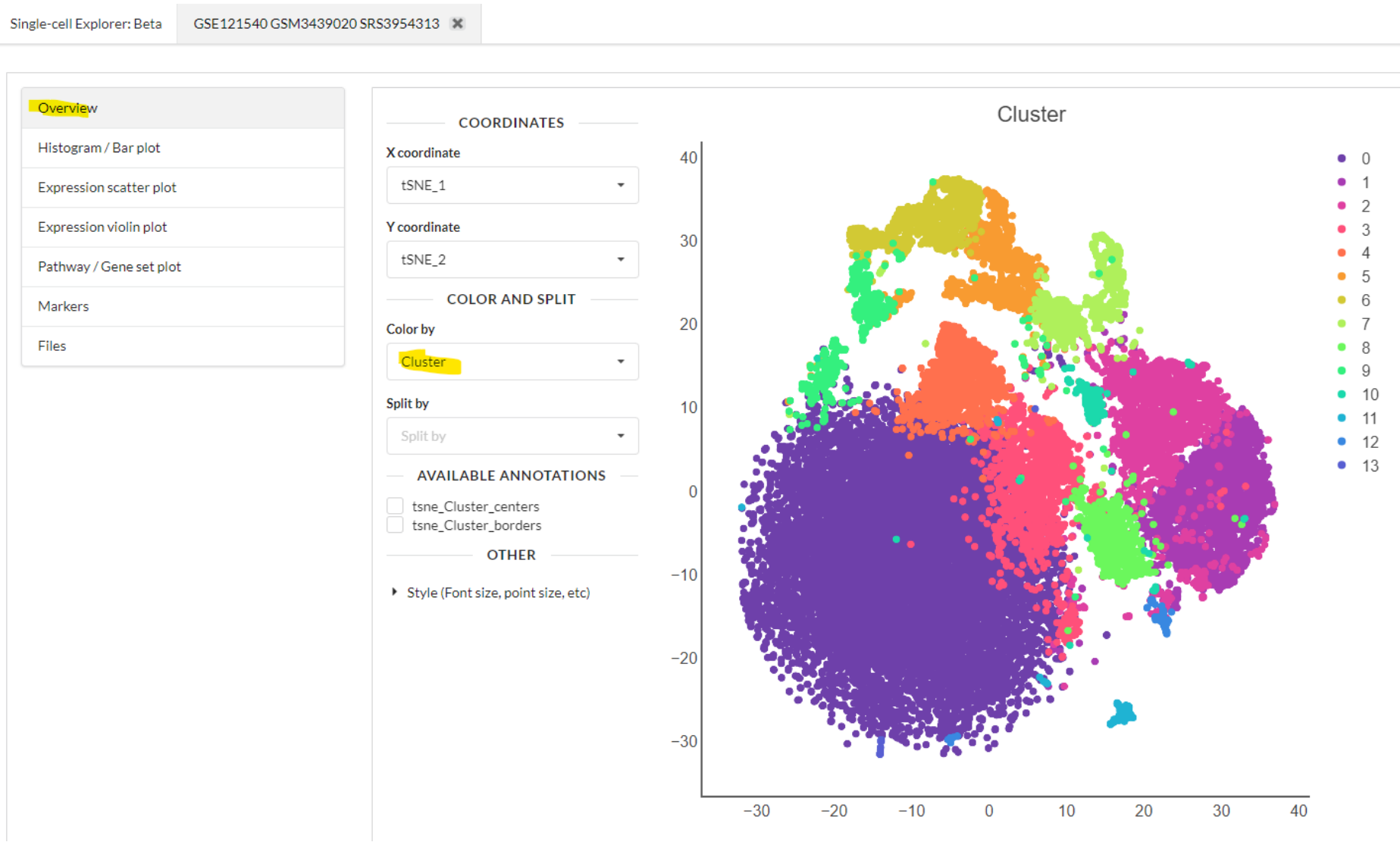
ITMO UNIVERSITY

Let’s open SRS3954313



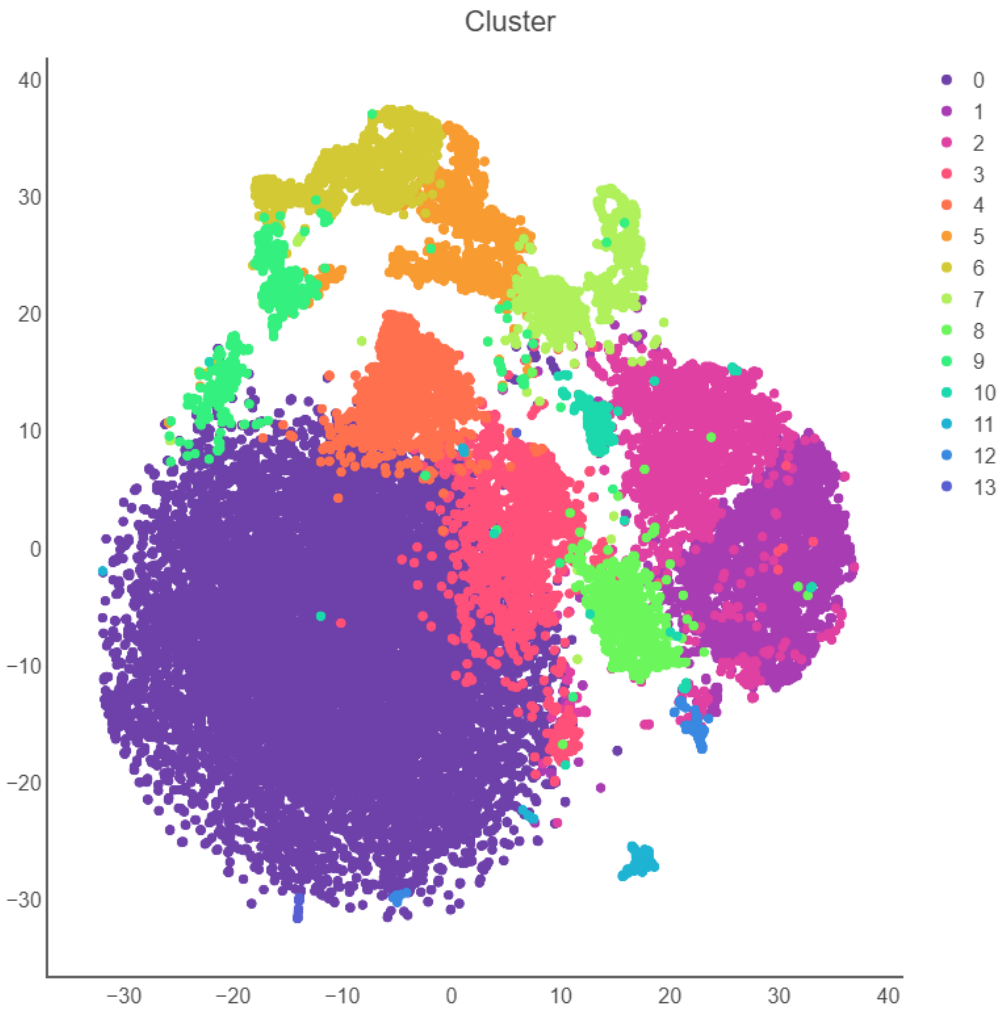
ITMO UNIVERSITY

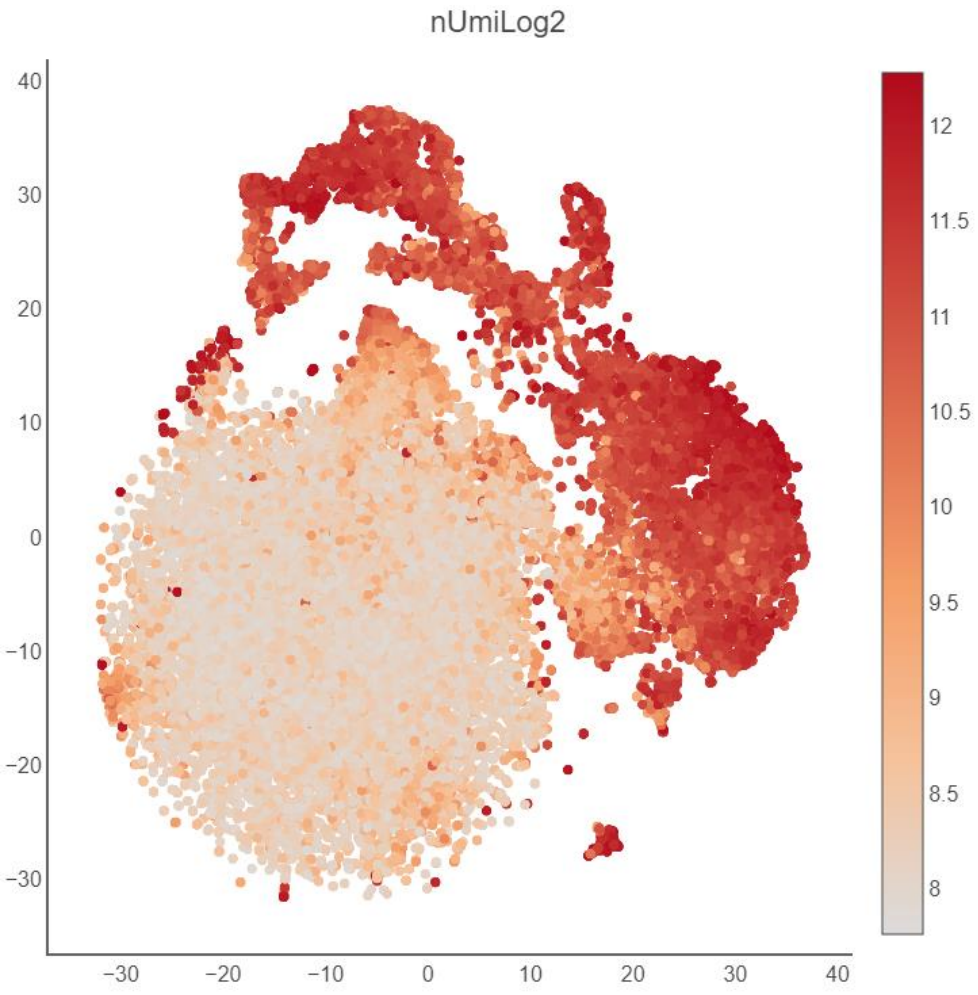
Let’s open SRS3954313



ITMO UNIVERSITY

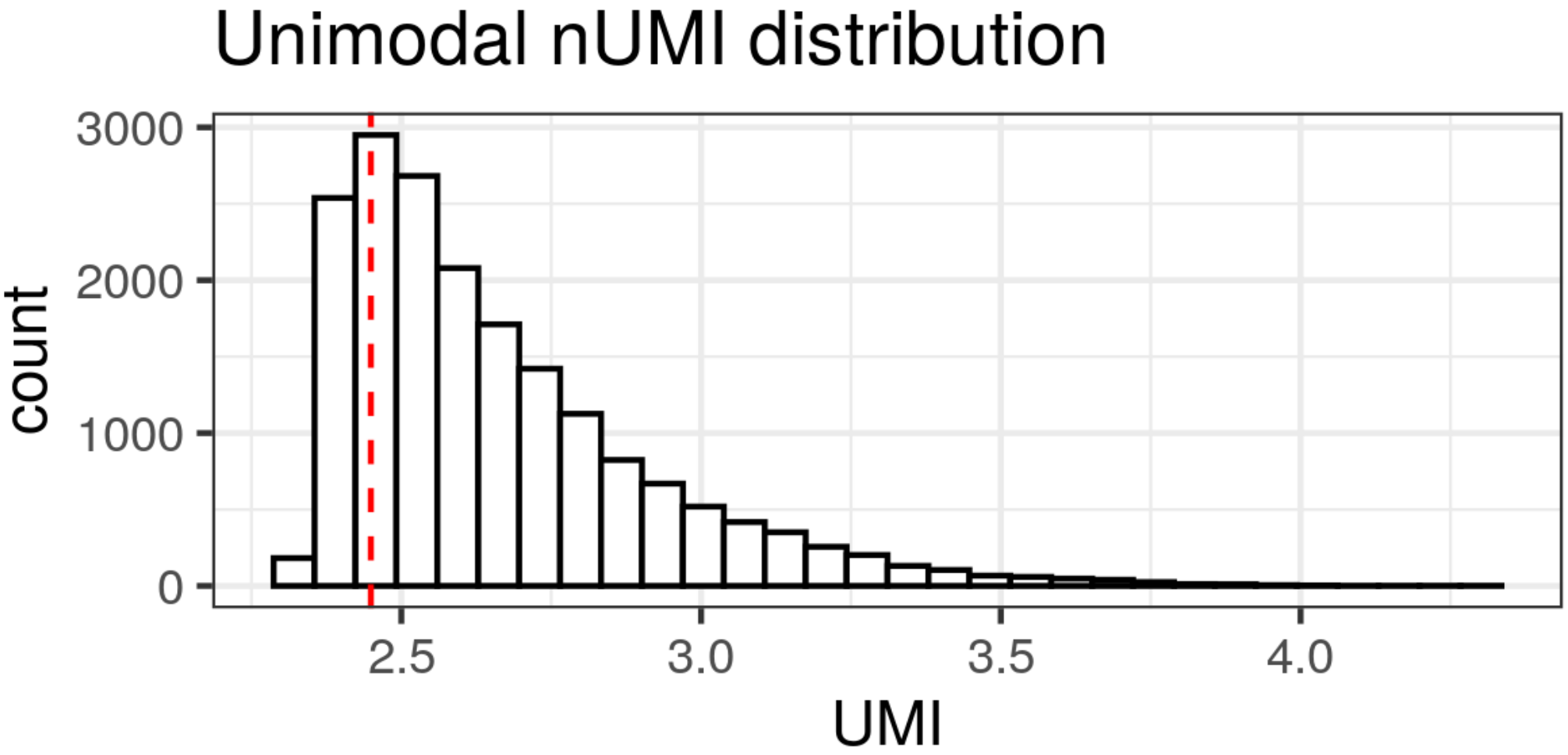
Let’s open SRS3954313





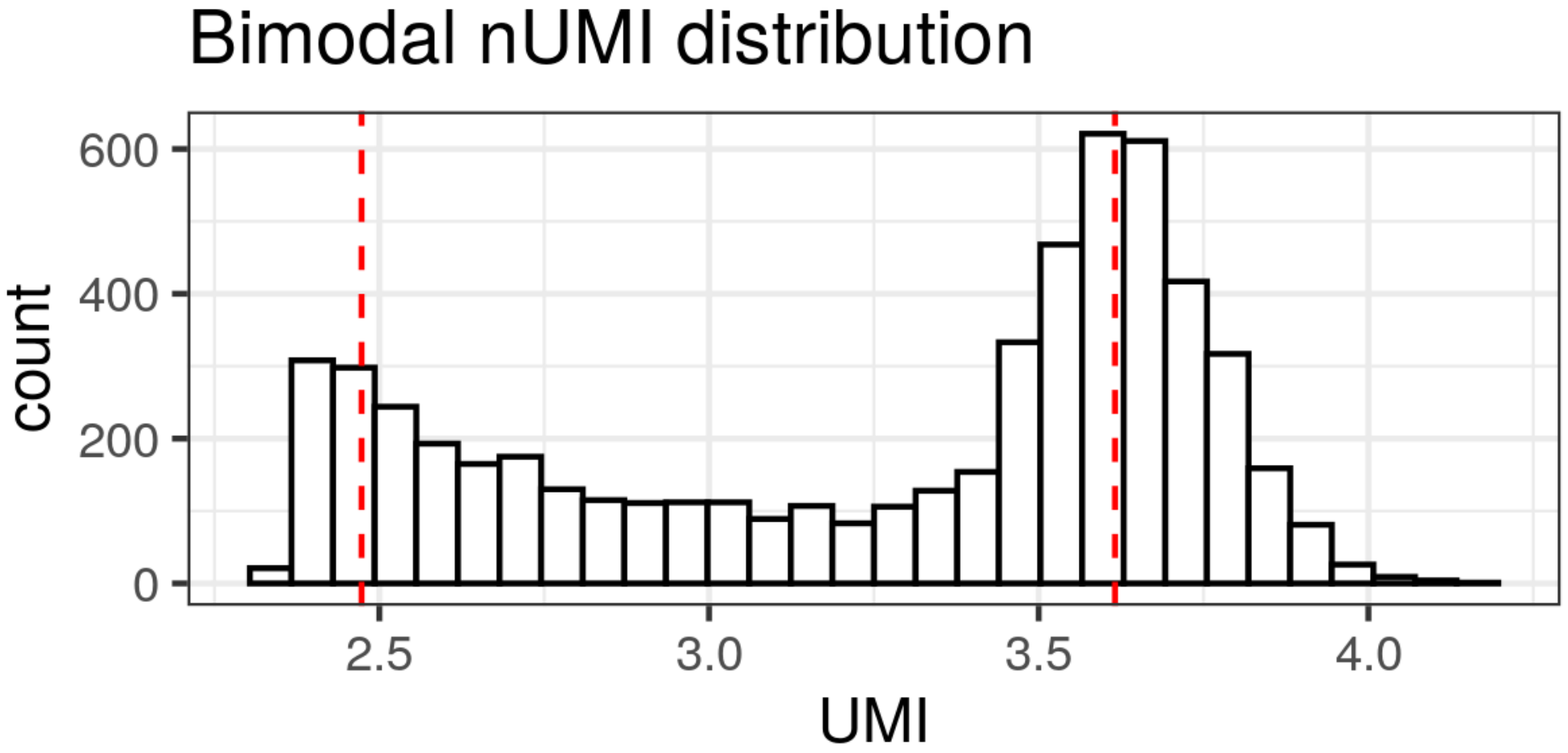
Issues: UMI distribution

Good case scenario



Issues: UMI distribution

Bad case scenario



Conclusion

- We hope that single-cell explorer will make interpretation of scRNA-seq data easier
- https://artyomovlab.wustl.edu/sce/

- We try to get there as much datasets as we can
- If you want to use SCE for your private data:
  - You can just e-mail me zayats1812@gmail.com, and I will give you a private link to your data
  - Wait until it gets published (ETA?), you will be able to host SCE locally, or for your department