



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)

Let's open the dataset

✓ Go to <https://artyomovlab.wustl.edu/sce/>

Single-cell Explorer: Beta

Single-cell explorer: beta

Single-cell explorer is an open-source project dedicated to processing and visualization of single-cell RNA-seq data

You can open any of preprocessed datasets or upload you own data (we currently support data in format of 10x files of mtx/genes/barcodes). Once you upload the data, link to your dataset will be available in several hours.

Currently available datasets are:

GSE/SRA id	Description
GSE120522 GSM3402513 S...	Pancreatic progenitor cells
GSE110501 GSM2994886 S...	heart
GSE103918_GSE103920 GS...	NKX2-1 GFP + lung progenitors in distal media
GSE109049 GSM2928506 S...	Post-natal day 6 testis
GSE93421 GSM2453163 SR...	E18 mouse brain cells
GSE121861_(immune)	Analysis of Single-Cell RNA-Seq Identifies Cell-Cell Communication Associated with Tumor Characteristics by Kumar MP, Du J, Lagoudas G, Jiao Y et al. Cell Rep 2018
GSE109718_SRA652805	Kidney organoids / Kidney organoids / Kidney organoids / Kidney organoids
SRA555753_SRS2135627	Neonatal mouse stomach explants / Mus musculus / 10x chromium
GSE121287_GSE121393_SR...	T-cells from spleen / T-cells from small intestine
GSE87544 GSM2333581 SR...	food deprived_hypothalamus

Previous Page 1 of 109 10 rows Next

Or you can enter a secret token below:

Go!

Let's open the dataset

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

Single-cell Explorer: Beta

Single-cell explorer: beta

Single-cell explorer is an open-source project dedicated to processing and visualization of single-cell RNA-seq data. You can open any of preprocessed datasets or upload your own data (we currently support data in format of 10x Genomics). Currently available datasets are:

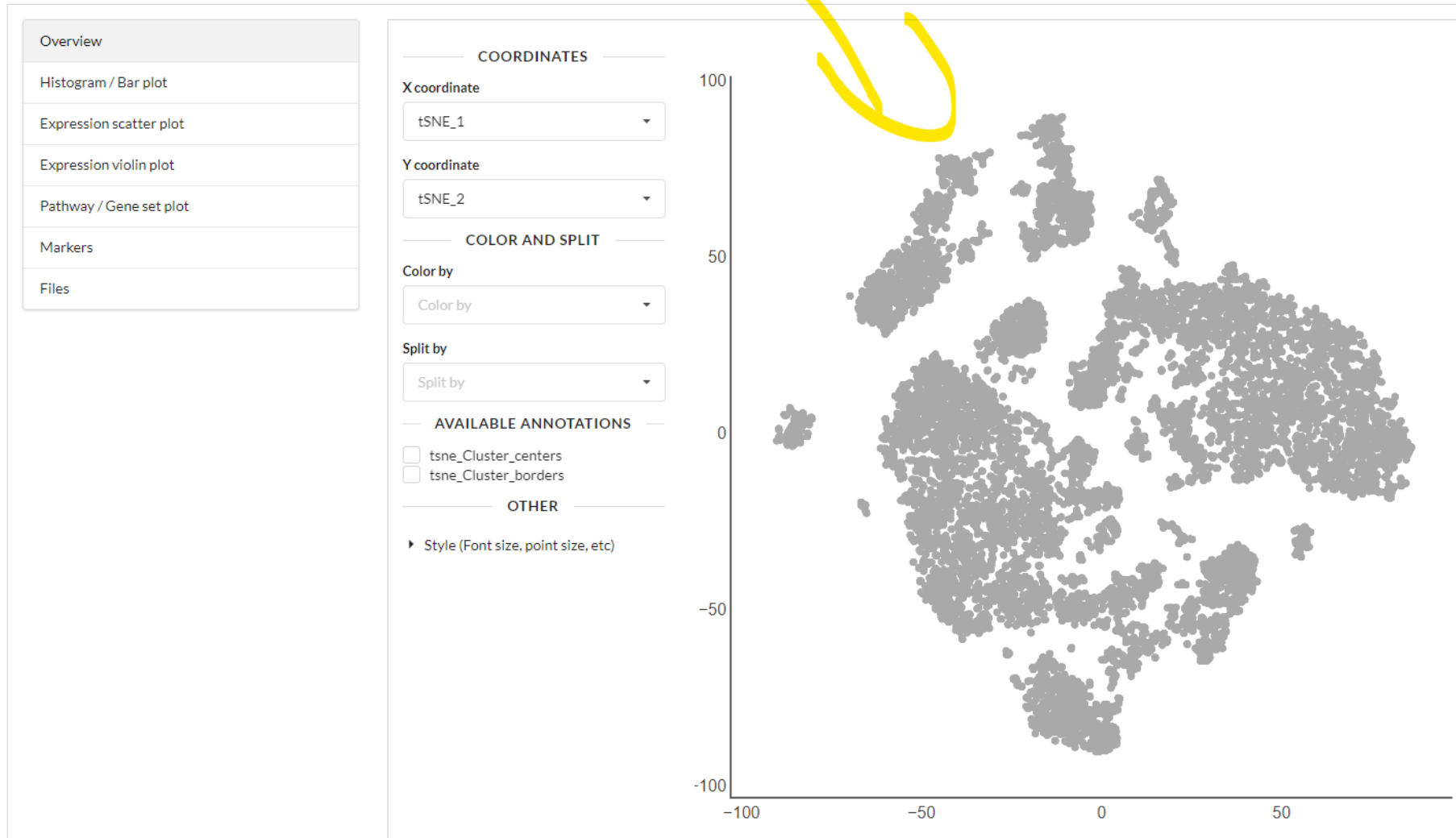
GSE/SRA id	Description
10x	
10x: PBMC 10k cells	Peripheral blood mononuclear cells (PBMCs) from a healthy donor (the s

If you have any problem finding dataset

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

Result should look like that

Single-cell Explorer: Beta

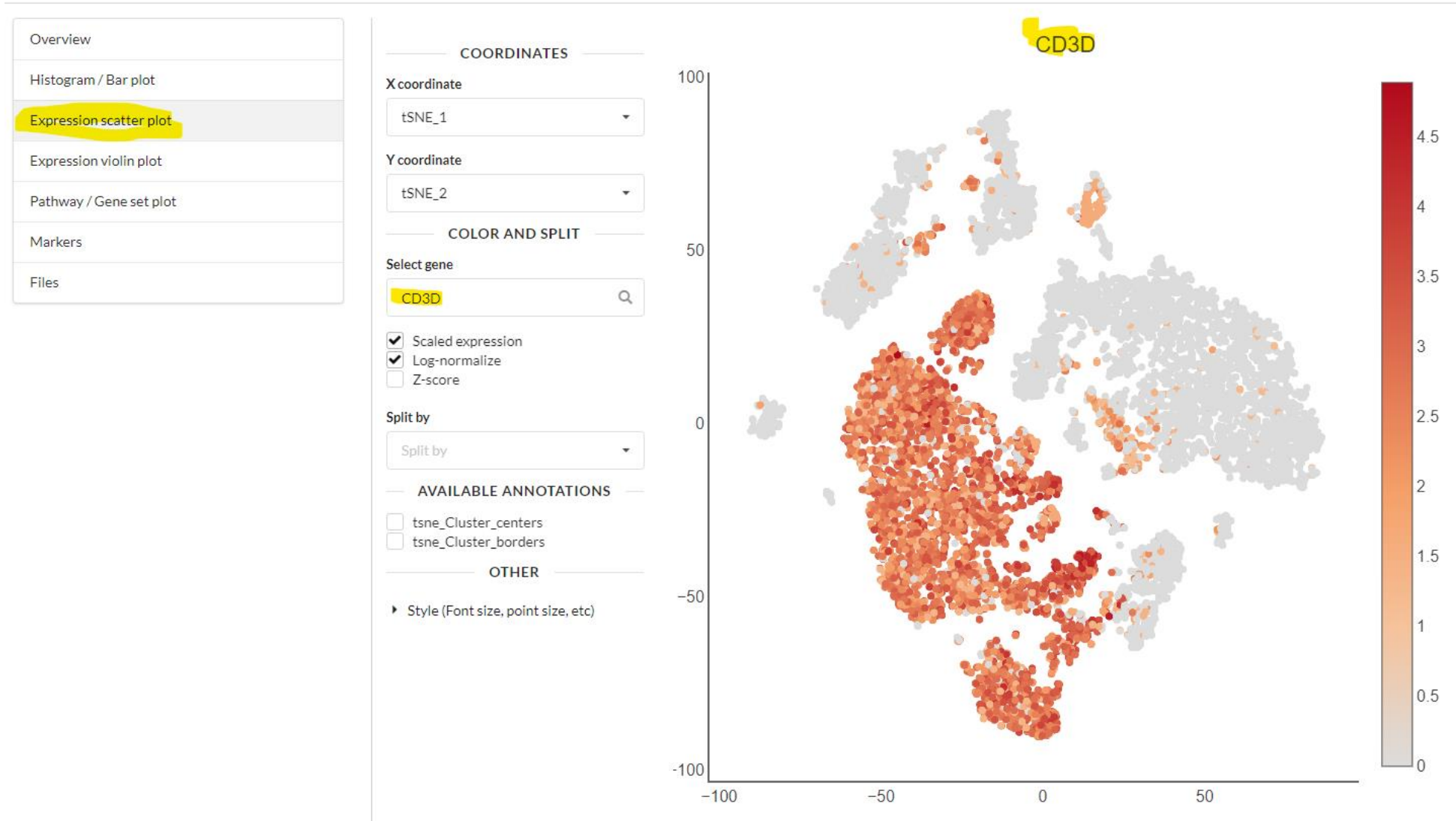
 PBMC_10k ✕


We can color the cells

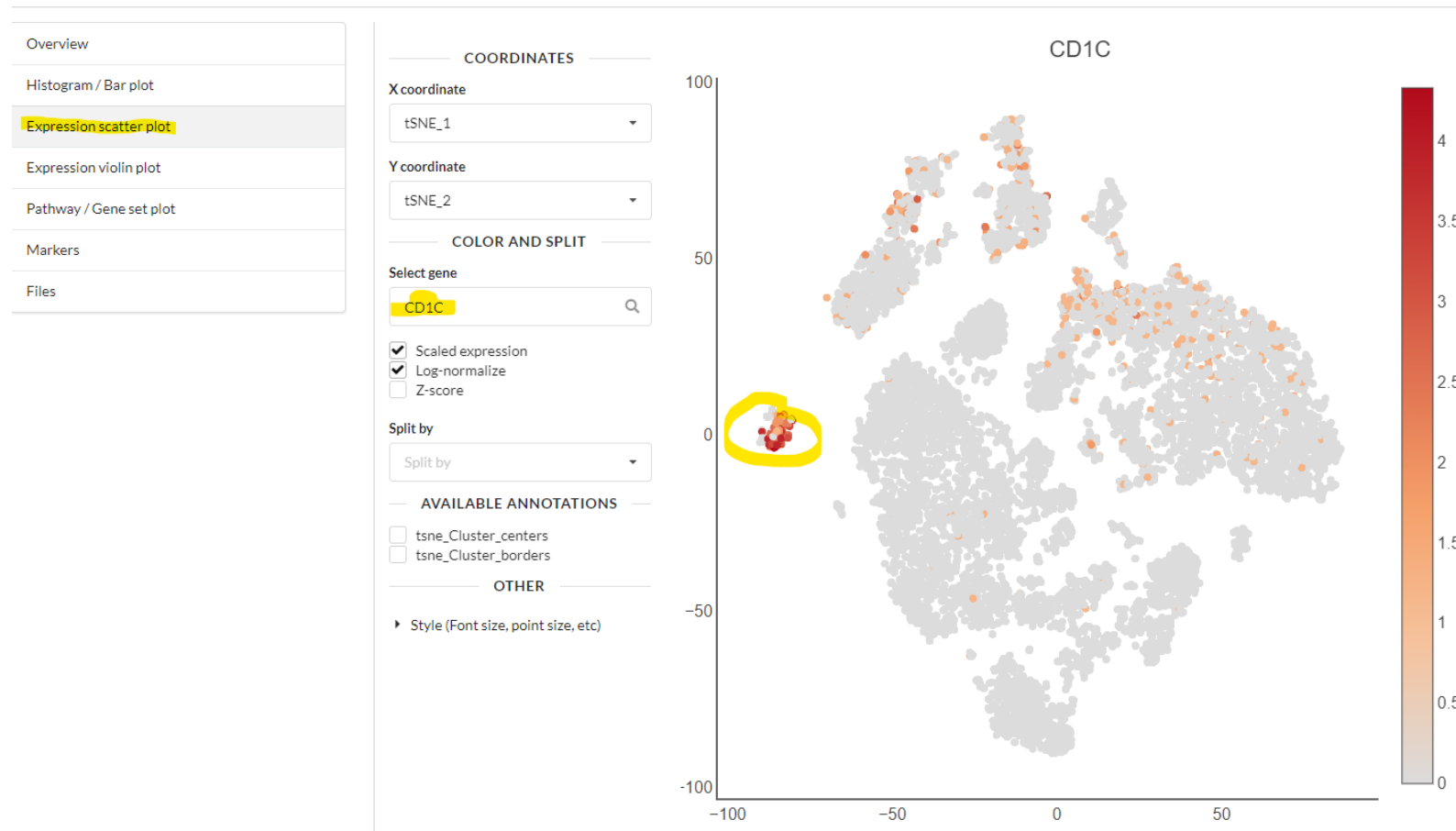
- ✓ Cluster
- ✓ Number of UMIs
- ✓ Number of genes detected
- ✓ tsne_Cluster_centers



Expression of CD3d



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

Violin plot

Single-cell Explorer: Beta

 PBMC_10k ✕

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

COORDINATES

X coordinate

 Cluster ▾

COLOR AND SPLIT

Select gene

 CD79A 🔍
 Scaled expression

 Log-normalize

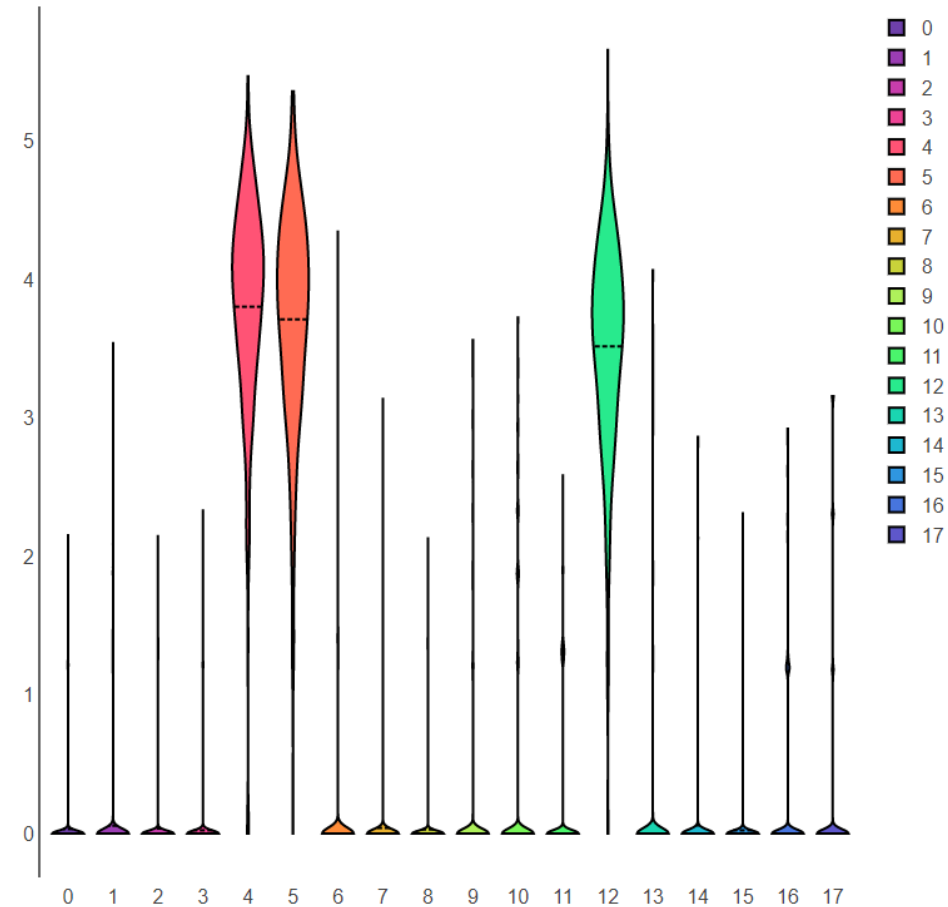
 Z-score

Split by

 Split by ▾

OTHER

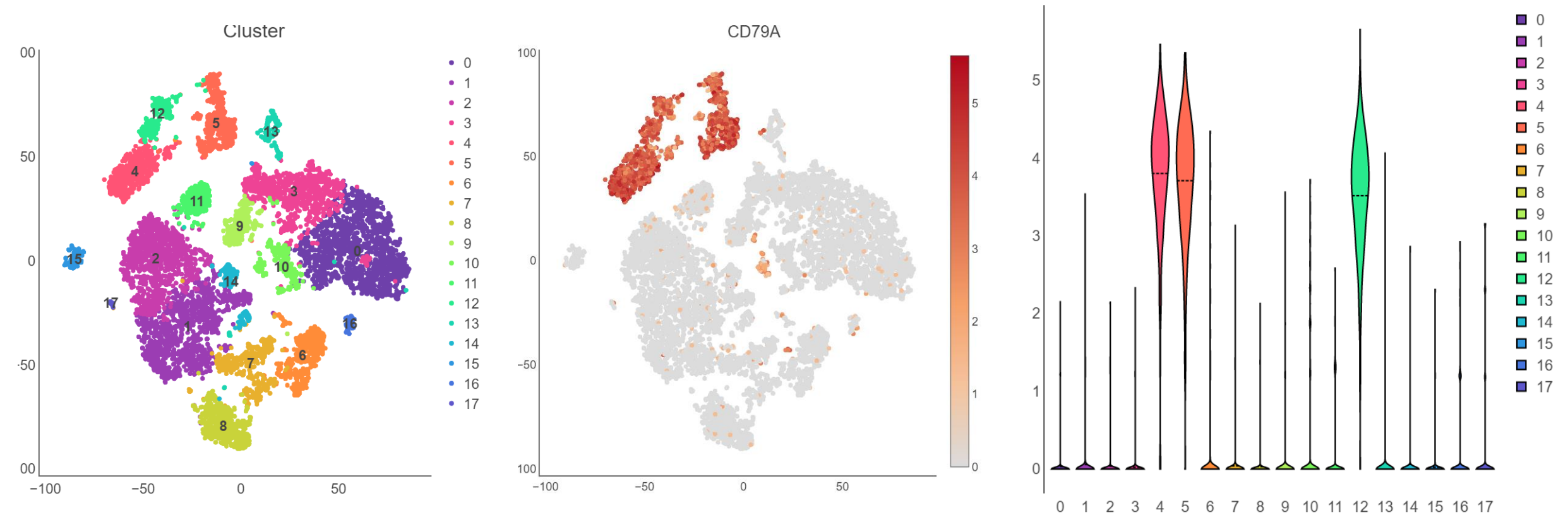
▶ Style (Font size, point size, etc)



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

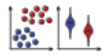
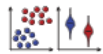
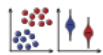
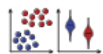
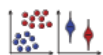





Markers tab

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

Download current table

Markers tab: what's the cluster 6?

Single-cell Explorer: Beta PBMC_10k

Overview

Histogram / Bar plot

Expression scatter plot

Expression violin plot

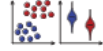
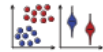


Pathway / Gene set plot

Markers

Files

Choose the table

Cluster

Gene name	Cluster	Av. log-fold change	P value	Adjusted p value	% in cluster	% outside
~	= 6	>	< 1e-	< 1e-	>	<
GNL1	 6	3.5825	0	0	0.981	0.124
NKG7	 6	2.7123	0	0	0.987	0.203
PRF1	 6	2.1023	0	0	0.975	0.123
KLRD1	 6	1.9782	0	0	0.972	0.073

- ✓ GNL1 – gene name
- ✓ Cluster 6 – we are checking results for cluster 6 vs other clusters
- ✓ Average log-fold change: average difference between expression of GNL1 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 GNL1 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

Single-cell Explorer: Beta PBMC_10k ✕

Overview

Histogram / Bar plot

Expression scatter plot

Expression violin plot

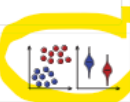


Pathway / Gene set plot

Markers

Files

Choose the table

Cluster

Gene name	Cluster	Av. log-fold change	P value	Adjusted p value	% in cluster	% outside
~	= 6	>	< 1e-	< 1e-	>	<
GNLY 	6	3.5825	0	0	0.981	0.124
NKG7 	6	2.7123	0	0	0.987	0.203
PRF1 	6	2.1023	0	0	0.975	0.123

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)

Single-cell Explorer: Beta
PBMC_10k ✕

Single-cell explorer: beta

Single-cell explorer is an open-source project dedicated to processing and visualization of single-cell RNA-seq data

You can open any of preprocessed datasets or upload you own data (we currently support data in format of 10x files of mtx/genes/barcodes). Once you upload the data, link to your dataset

Currently available datasets are:

GSE/SRA id	Description
10x	
10x: PBMC 10k cells	Peripheral blood mononuclear cells (PBMCs) from a healthy donor (the same cells were used to generate pbmc_1k_v2, pbmc_10k_v3). PBMCs are primar

Public datasets

including datasets from Human Cell Atlas

Single-cell Explorer: Beta PBMC_10k ✕

Single-cell explorer: beta

Single-cell explorer is an open-source project dedicated to processing and visualization of single-cell RNA-seq data

You can open any of preprocessed datasets or upload your own data (we currently support data in format of 10x files of mtx/genes/barcodes). Once you upload the data, link to your dataset will be available in several hours.

Currently available datasets are:

GSE/SRA id	Description
HCA	
HCA: pancreatic cells	As organisms age, cells accumulate genetic and epigenetic changes that eventually lead to impaired organ function or catastrophic failure such as cancer. Here we describe a single-cell transcriptome analysis of 2544 human pancrea
HCA: Ischaemic sensitivity of...	Assessment of ischaemic sensitivity of human tissues using 10x 3' single cell RNA sequencing. This project contains data for spleen, oesophagus epithelium and lung parenchyma (based on previously published bulk RNA-seq data, w
HCA: Profiling of CD34+ cell...	Differentiation is among the most fundamental processes in cell biology. Single cell RNA-seq studies have demonstrated that differentiation is a continuous process and in particular cell states are observed to reside on largely conti
HCA: Reconstructing the hu...	During early human pregnancy the uterine mucosa transforms into the decidua, into which the fetal placenta implants and where placental trophoblast cells intermingle and communicate with maternal cells. Trophoblast-decidual i
HCA: Structural Remodeling...	Intestinal mesenchymal cells play essential roles in epithelial homeostasis, matrix remodeling, immunity, and inflammation. But the extent of heterogeneity within the colonic mesenchyme in these processes remains unknown. Usin
HCA: Assessing the relevanc...	The purpose of this project is to assess the relevance of pluripotent stem cell-derived cerebral and liver organoids to recapitulate the variation in cell-type specific gene expression programs between individuals. Towards this aim, w
HCA: Single-cell RNA-seq an...	Diverse cell types are produced from dorsal and ventral regions of the developing neural tube. In this study we describe a system for generating human inhibitory interneurons by ventralizing human embryonic stem cells in vitro an

Previous Page 1 of 1 10 rows ▾ Ne

Or you can enter a secret token below:

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 all

✓ 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 ...

What are the issues

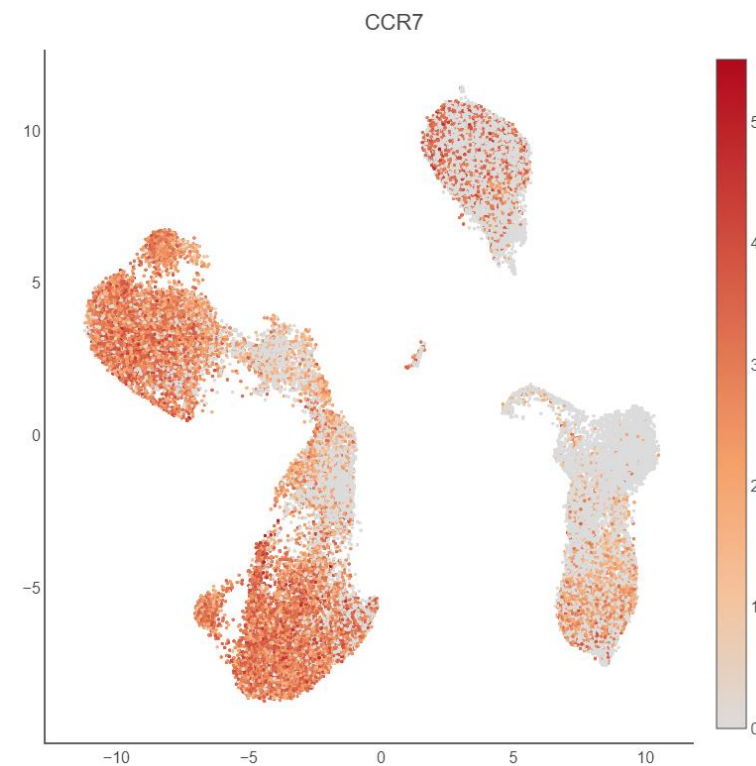
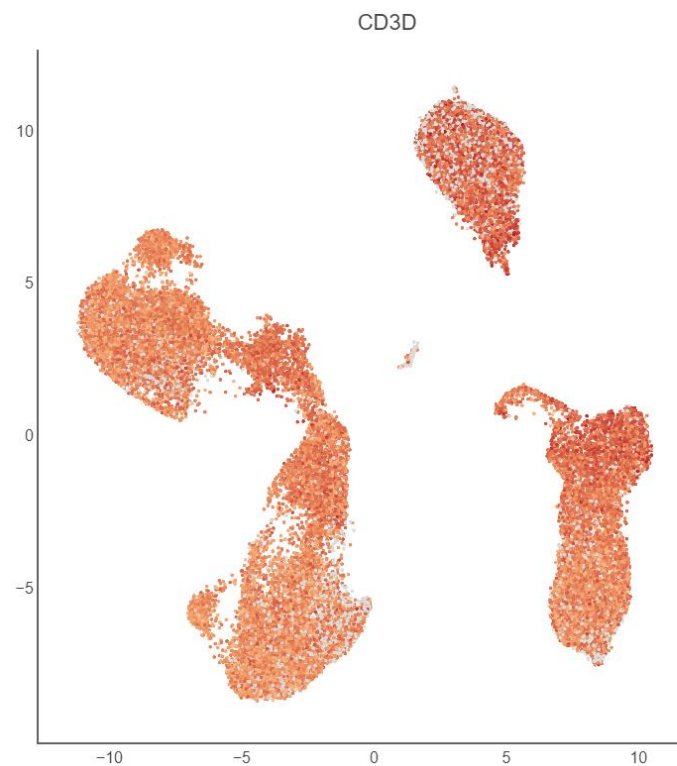
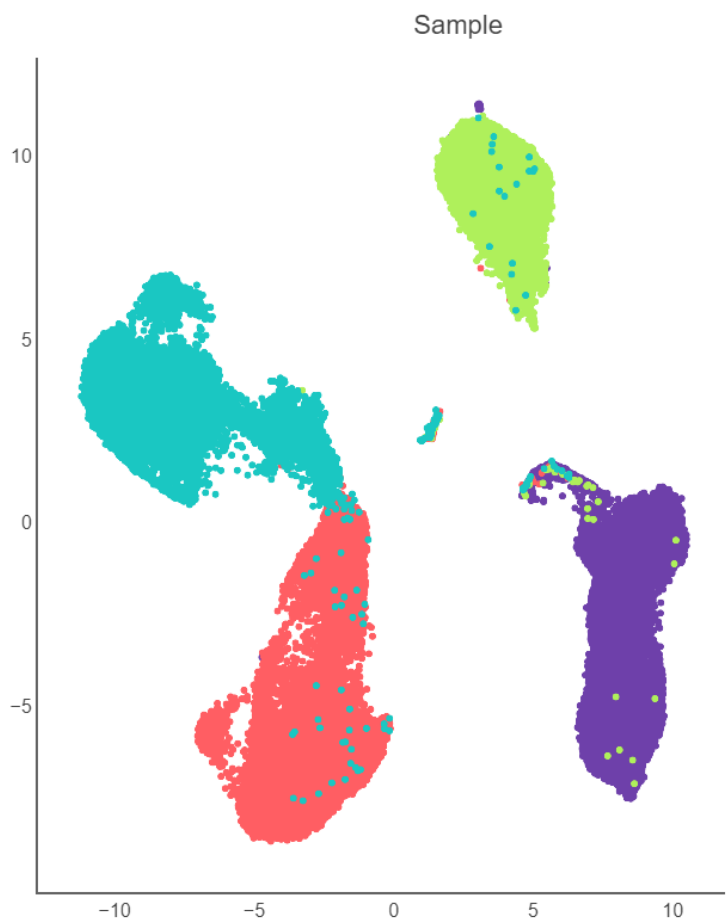
When we first analyzed 1000 dataset two main issues were identified:

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

Most of the dataset processing was done
by Maria Firuleva

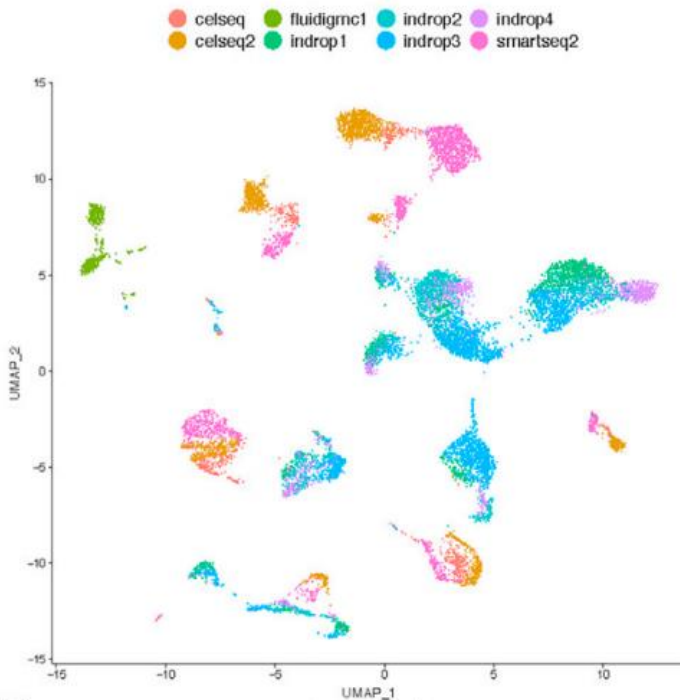


Issues: donor effect

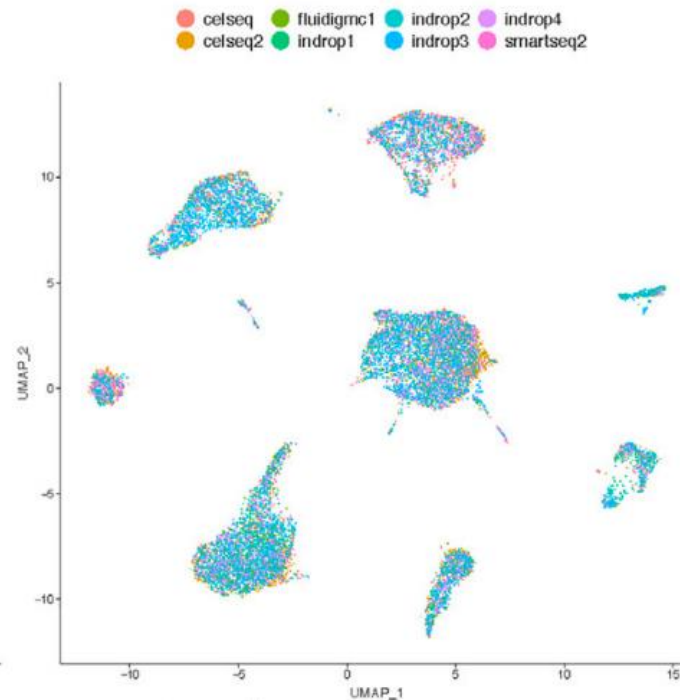


Recent developments of methods

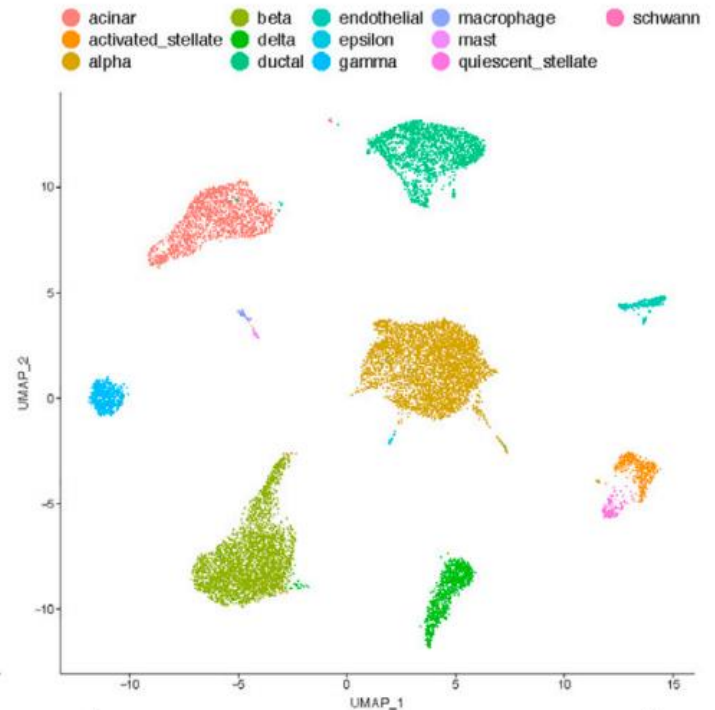
A Unintegrated Datasets



B Integrated Datasets



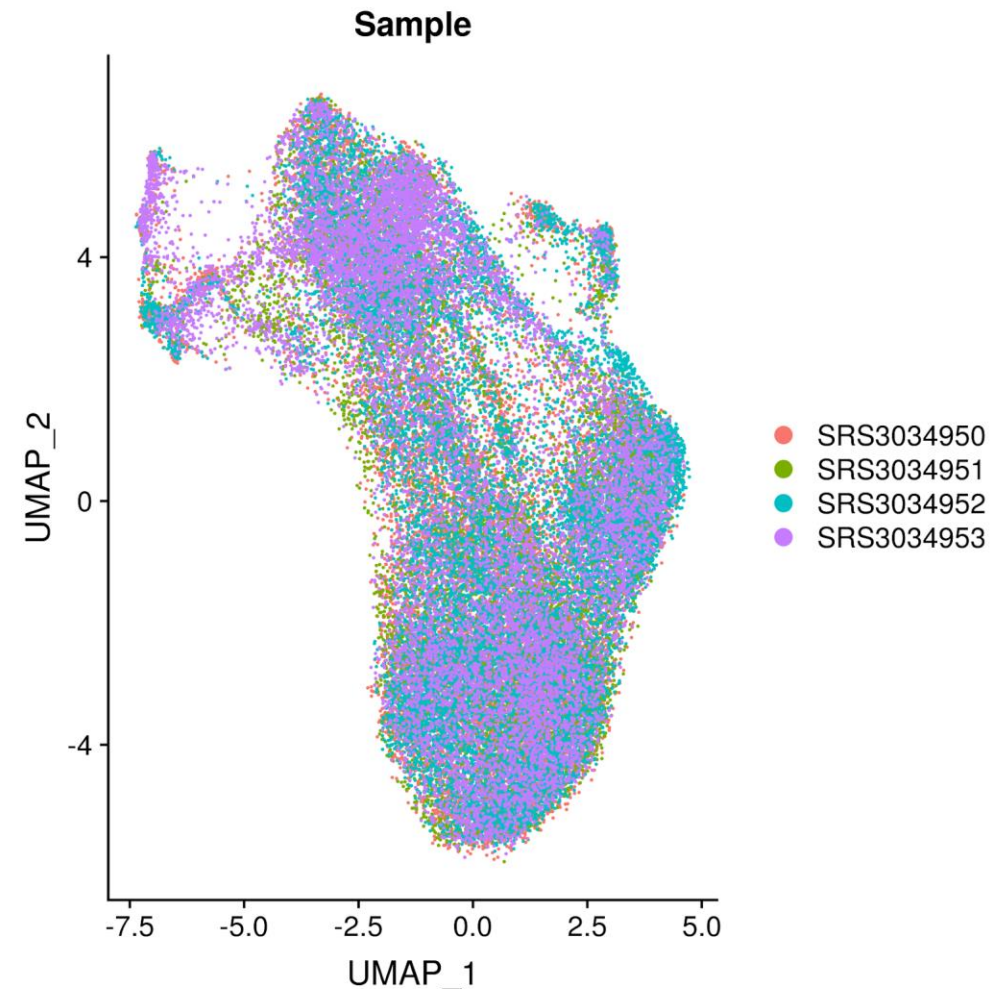
C



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

Issues: donor effect

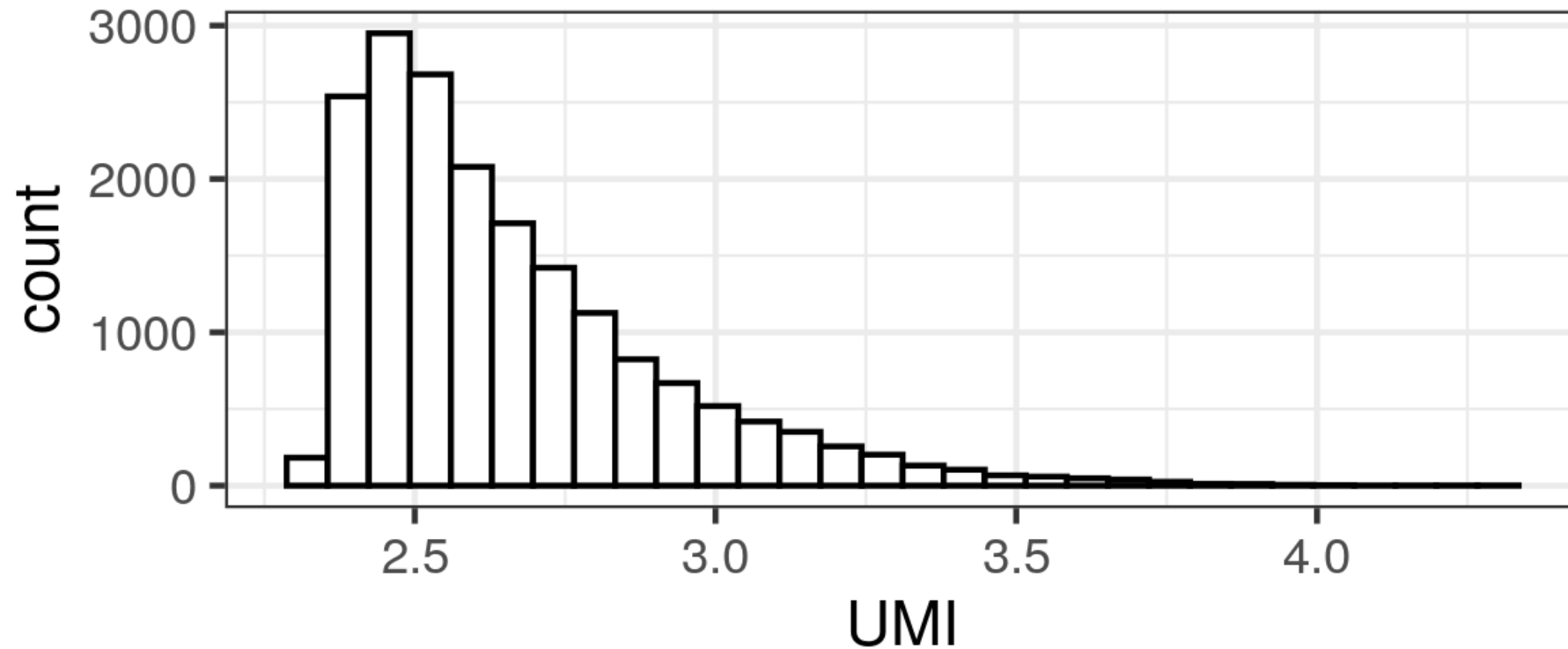
- ✓ Integration methods remove batch/donor effects
- ✓ Integration methods can be run automatically



Issues: UMI distribution

Good case scenario

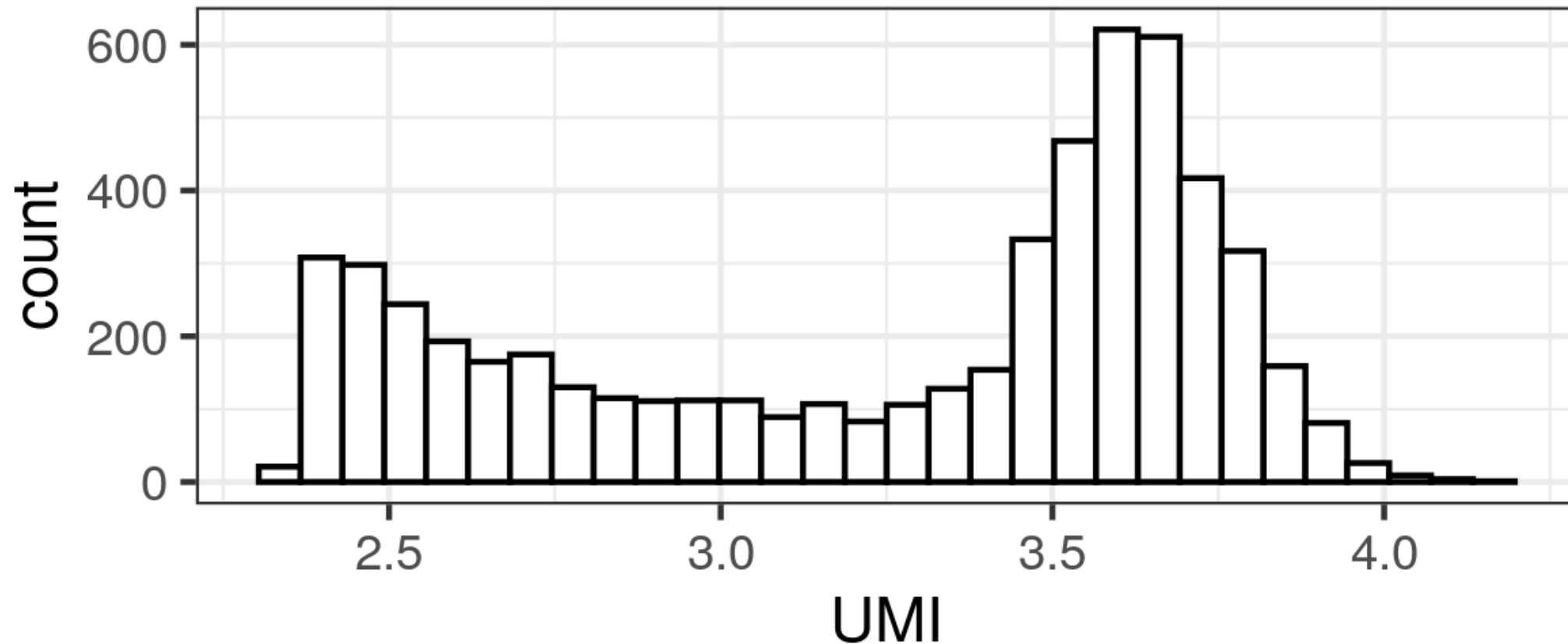
Unimodal nUMI distribution



Issues: UMI distribution

Bad case scenario

Bimodal nUMI distribution



Let's open SRS3954313

Single-cell Explorer: Beta

Single-cell explorer: beta

Single-cell explorer is an open-source project dedicated to processing and visualization of single-cell RNA-seq data

You can open any of preprocessed datasets or upload you own data (we currently support data in format of 10x files of mtx/genes/

Currently available datasets are:

GSE/SRA id	
<input type="text" value="SRS3954313"/>	
GSE121540 GSM3439020 SRS3954313	C57/Bl6J SVZ-derived neural stem cells

Let's open SRS3954313

Single-cell Explorer: Beta **GSE121540 GSM3439020 SRS3954313** ✕

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

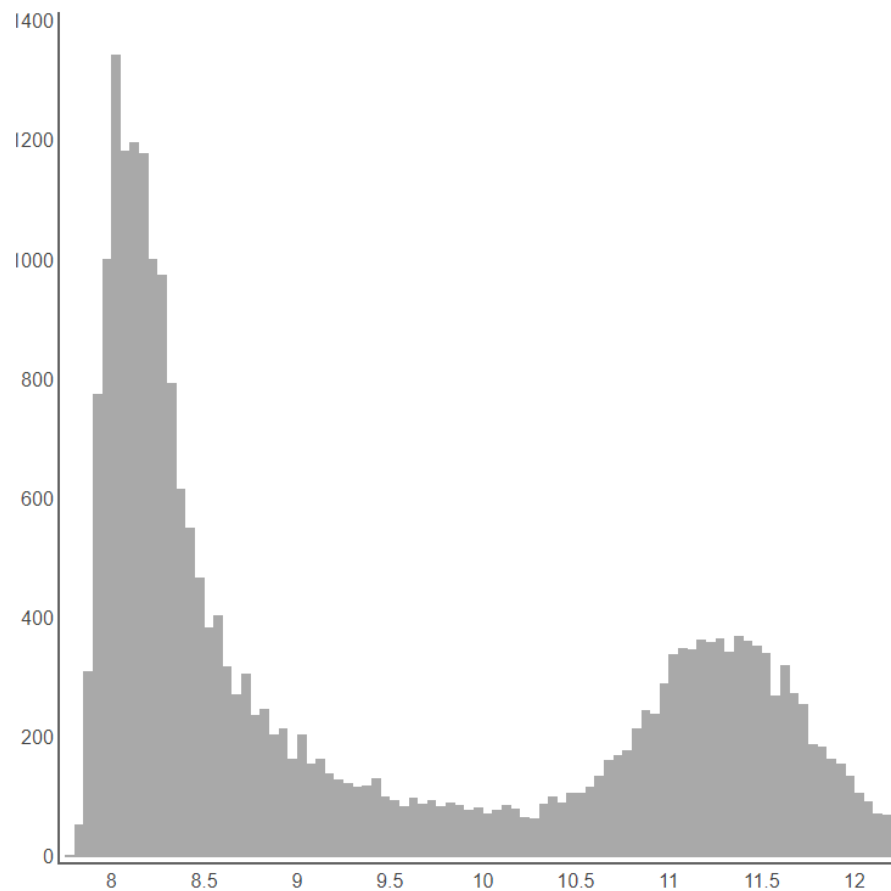
BAR PLOT INPUTS

X coordinate

Split by

OTHER

▸ Style (Font size, point size, etc)

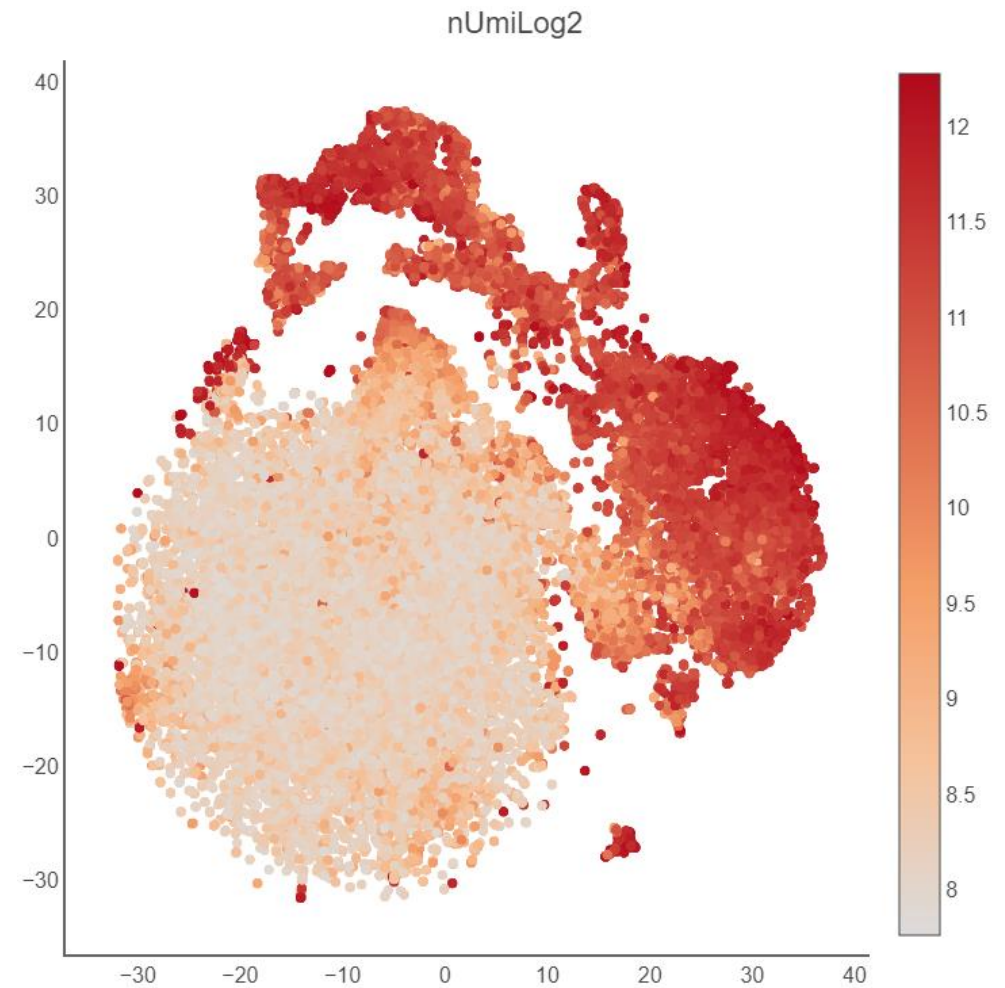
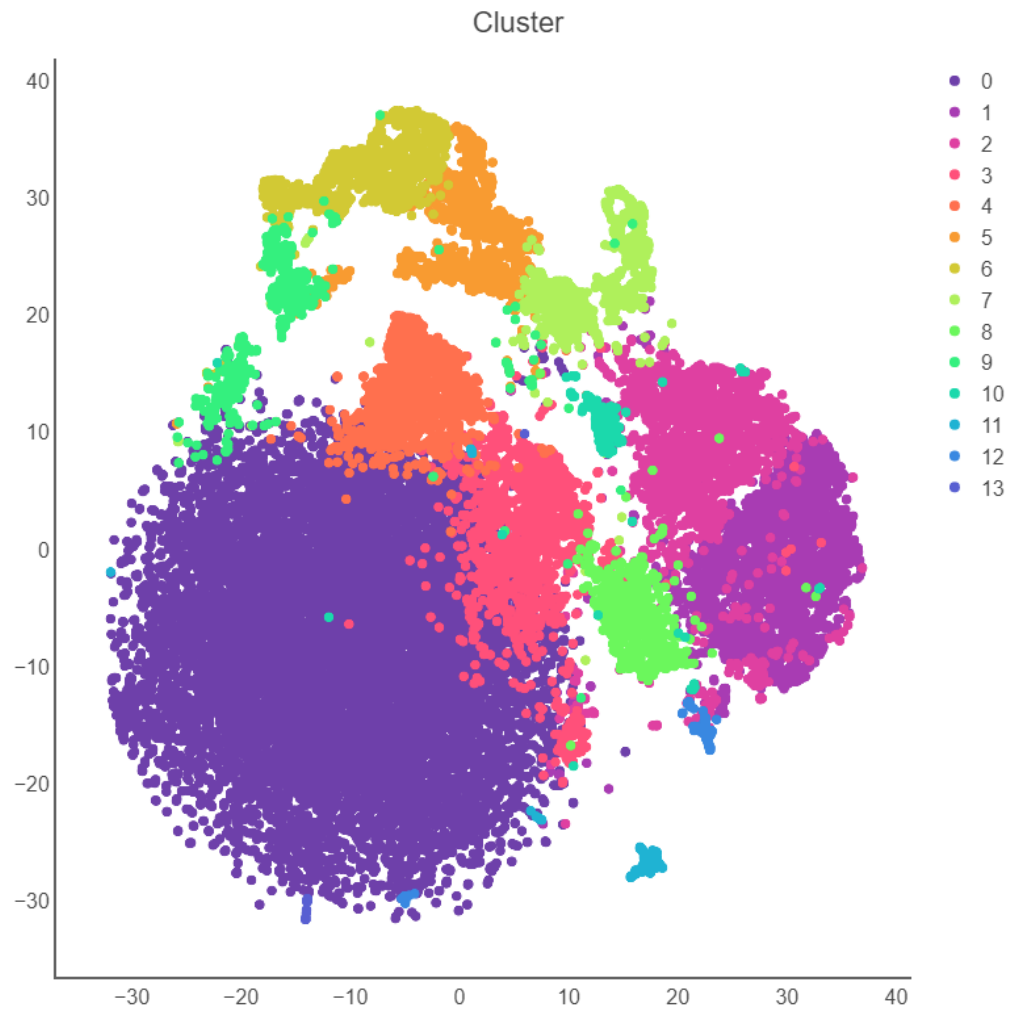


Let's open SRS3954313

Single-cell Explorer: Beta GSE121540 GSM3439020 SRS3954313 ✕



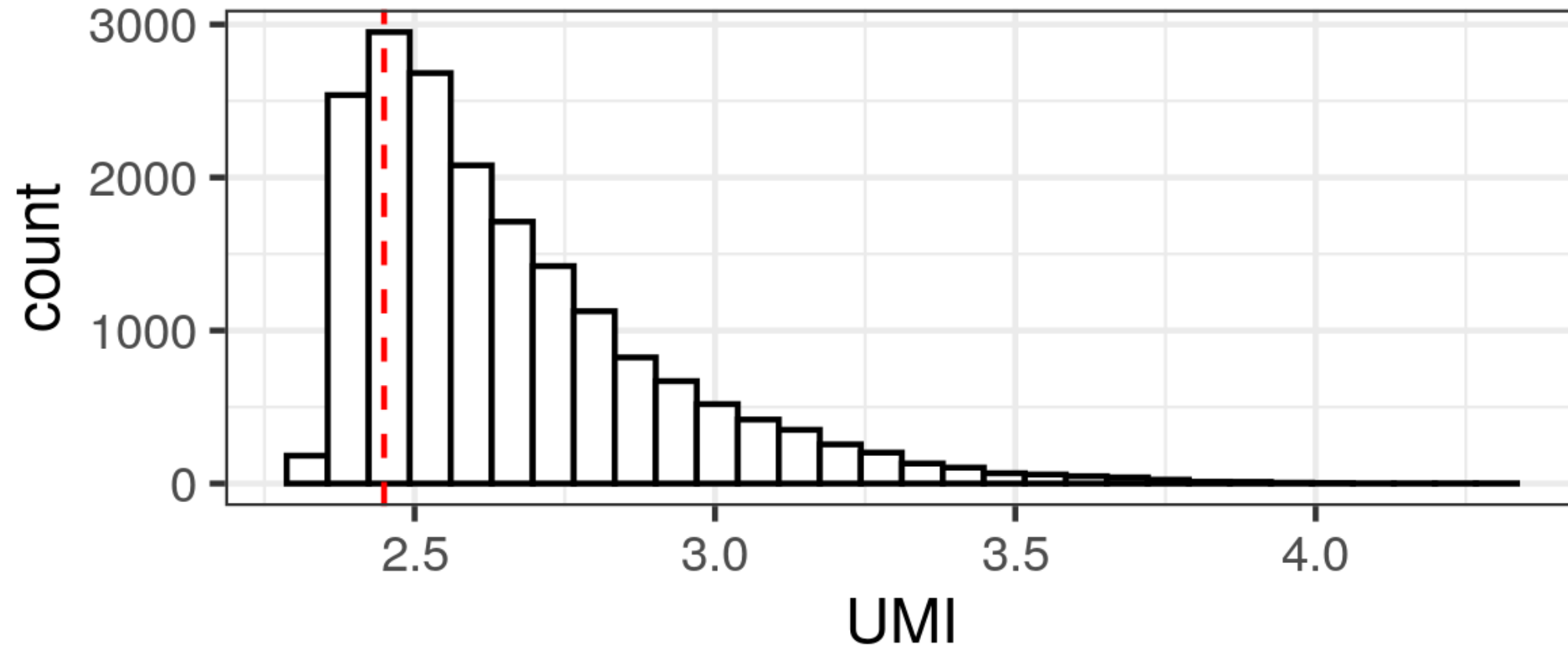
Let's open SRS3954313



Issues: UMI distribution

Good case scenario

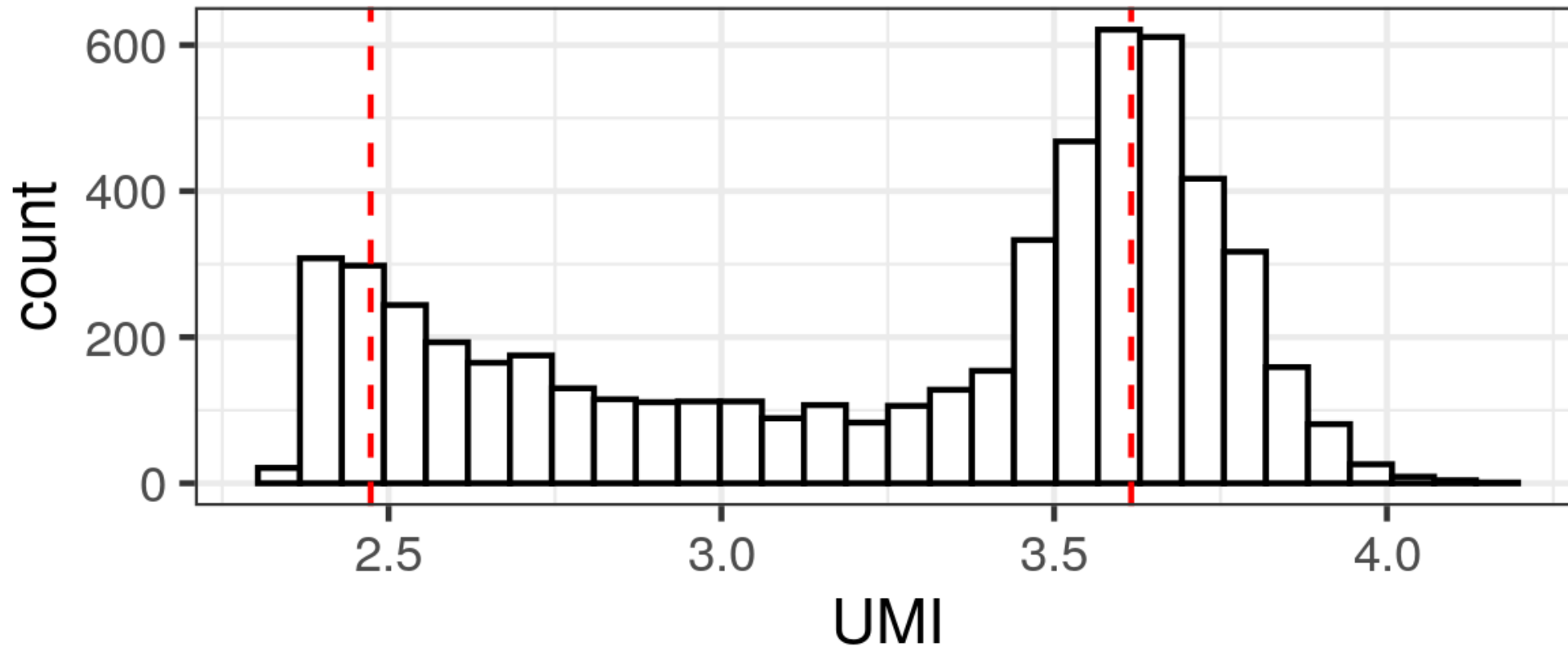
Unimodal nUMI distribution



Issues: UMI distribution

Bad case scenario

Bimodal nUMI distribution



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