

Single-cell Navigator: visualizing scRNA-seq data

Konstantin Zaitsev, ITMO University

Masters Program, Feb 12th, 2020

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/>

scNavigator: beta

scNavigator: beta

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

Below we have a large collection of datasets and tools to play with:

- Large collection of automatically processed datasets. We processed almost every scRNA-seq dataset from GEO Omnibus database. We make it available for you in our browser.
- Collection of curated datasets. Curated dataset are those that we process by hand. These will include datasets from Human Cell Atlas (HCA), Tabula Muris and some of the datasets that we generated in our lab.
- You can search for cell type specific gene signatures! When we processed all the public scRNA-seq datasets we also calculated all the markers of all the clusters in all these datasets. Just put a list of genes and we will tell you which cluster in which dataset it looks like.
- If you were provided with secret dataset token, you can use it at the very right of this page

Enter a secret token below:

Go!

All scRNA-seq datasetsCurated datasetsGene signature search

Name	Description	Organism	# of cells	Ext...
GSE101901/SRS2384613	Single cell sequencing of hippocampus tissues in traumatic brain injury	Mus Musculus	8878	↗
GSE103976/SRS2523512	Detecting Activated Cell Populations Using Single-Cell RNA-Seq	Mus Musculus	6488	↗
GSE129730/SRS4617144	Single cell RNA-seq shows cellular heterogeneity and lineage expansion in a mouse model of SHH-driven medulloblastoma support resistance to SHH inhibitor therapy	Mus Musculus	4552	↗
GSE103983/SRS2523775	Single-cell RNA-seq (Drop-seq) of MGE, CGE and LGE of E13.5 (MGE) and E14.5 (CGE, LGE) mouse embryos	Mus Musculus	11704	↗
GSE93374/SRS1913127	A Molecular Census of Arcuate Hypothalamus and Median Eminence Cell Types	Mus Musculus	61225	↗
GSE103983/SRS2523784	Single-cell RNA-seq (Drop-seq) of MGE, CGE and LGE of E13.5 (MGE) and E14.5 (CGE, LGE) mouse embryos	Mus Musculus	709	↗
GSE137007/SRS5355828	Proliferation-competent Tcf1+ CD8 T-cells in dysfunctional populations are CD4 T-cell help independent	Mus Musculus	434	↗
GSE106960/SRS2690039	The single cell RNA seq of pulmonary alveolar epithelial cells	Mus Musculus	2683	↗
GSE113111/SRS3165512	sc-RNA sequencing of skeletal muscle macrophages during T. gondii infection and injury	Mus Musculus	6625	↗
GSE129730/SRS4617149	Single cell RNA-seq shows cellular heterogeneity and lineage expansion in a mouse model of SHH-driven medulloblastoma support resistance to SHH inhibitor therapy	Mus Musculus	5110	↗

Previous

Page 1 of 35

10 rows

Next

Let's open the dataset

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

scNavigator: beta

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

Below we have a large collection of datasets and tools to play with:

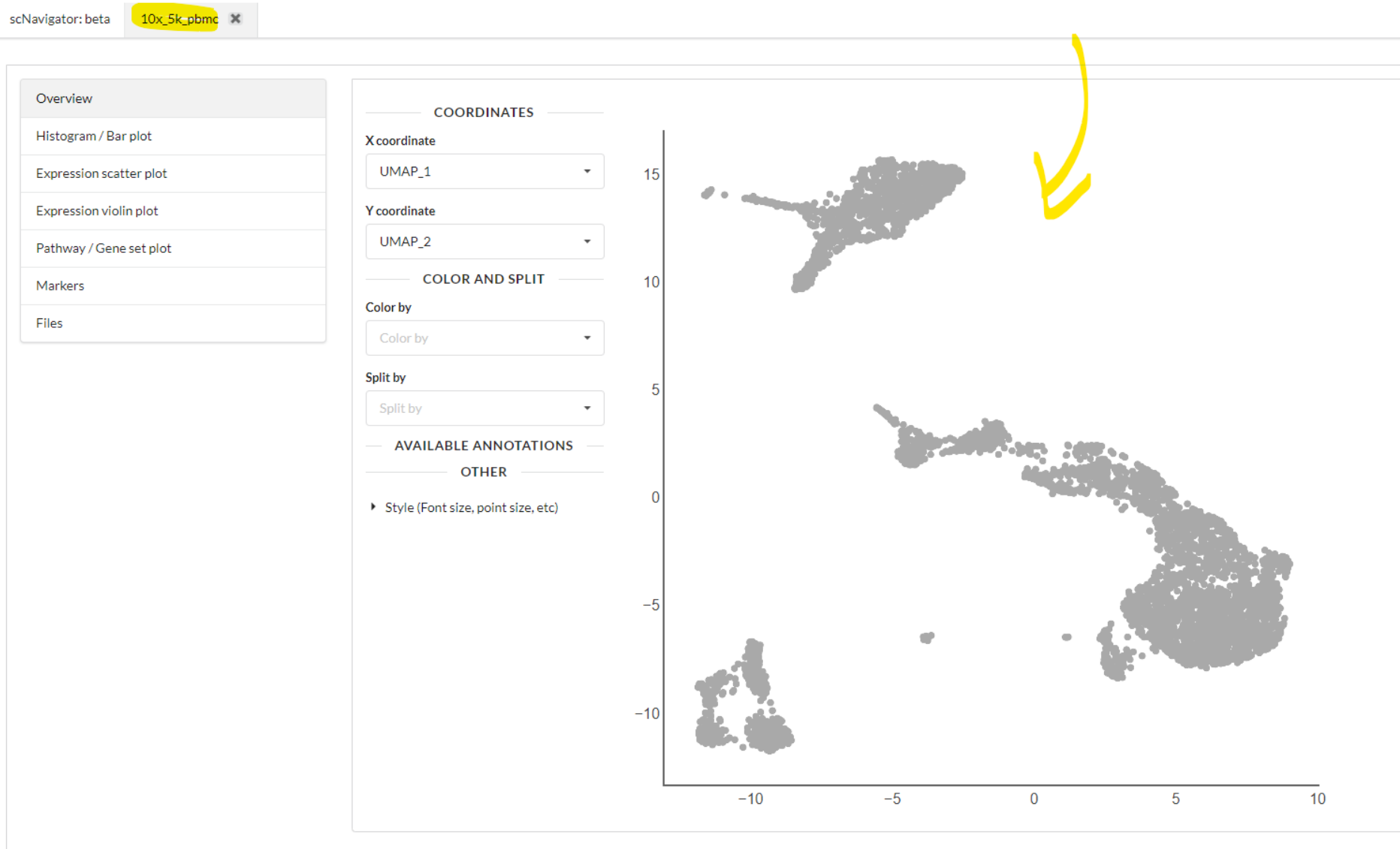
- Large collection of automatically processed datasets. We processed almost every scRNA-seq dataset from GEO Omnibus database. We make it available for you in our browser.
- Collection of curated datasets. Curated dataset are those that we process by hand. These will include datasets from Human Cell Atlas (HCA), Tabula Muris and some of the datasets that
- You can search for cell type specific gene signatures! When we processed all the public scRNA-seq datasets we also calculated all the markers of all the clusters in all these datasets. Just you which cluster in which dataset it looks like.
- If you were provided with secret dataset token, you can use it at the very right of this page

All scRNA-seq datasets		Curated datasets	Gene signature search
Name	Description		
10x			
10x_5k_pbmc	10x dataset of 5k peripheral blood mononuclear cells		

If you have any problem finding dataset

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

Result should look like that



We can color the cells

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

COORDINATES

X coordinate

Y coordinate

COLOR AND SPLIT

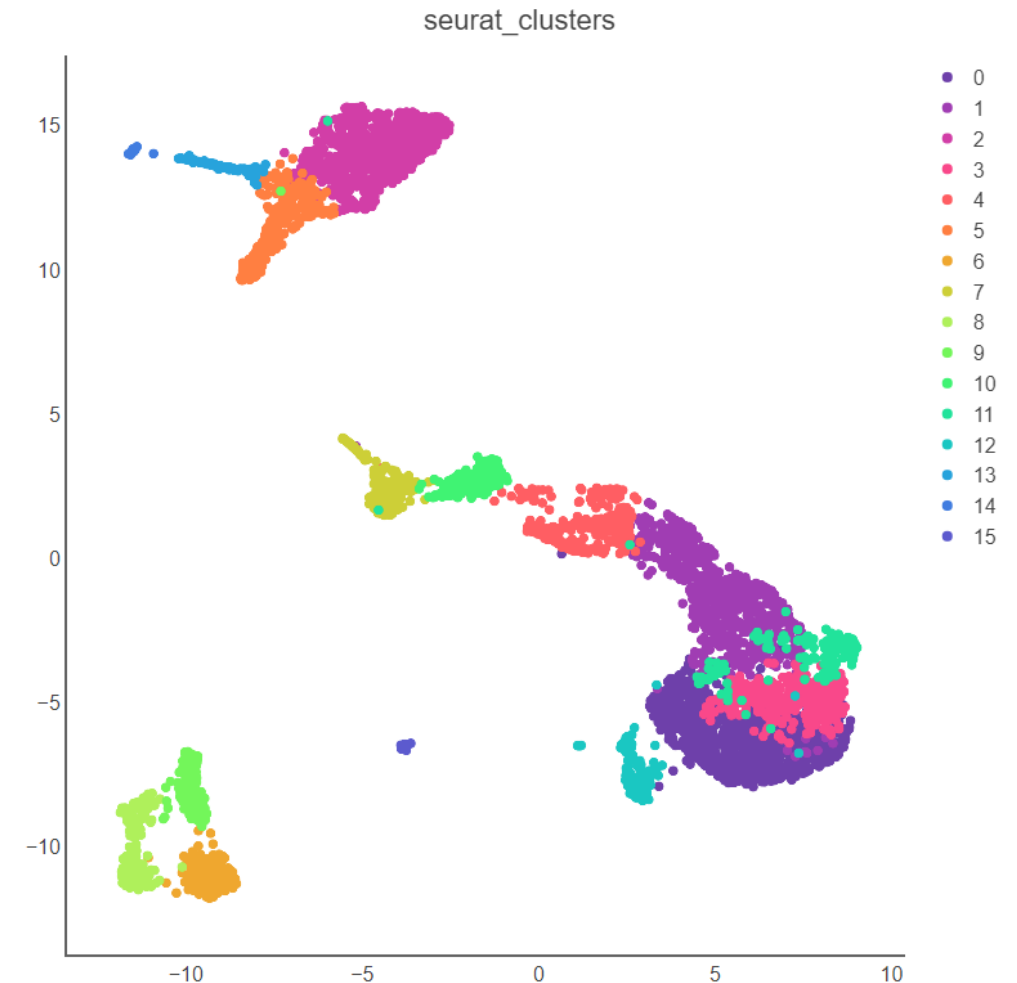
Color by

Split by

AVAILABLE ANNOTATIONS

OTHER

▸ Style (Font size, point size, etc)



Expression of CD3d

scNavigator: beta

10x_5k_pbmc

Overview

Histogram / Bar plot

Expression scatter plot

Expression violin plot

Pathway / Gene set plot

Markers

Files

COORDINATES

X coordinate

UMAP_1

Y coordinate

UMAP_2

COLOR AND SPLIT

Select gene

CD3D

☒ Scaled expression

☒ Log-normalize

☐ Z-score

Split by

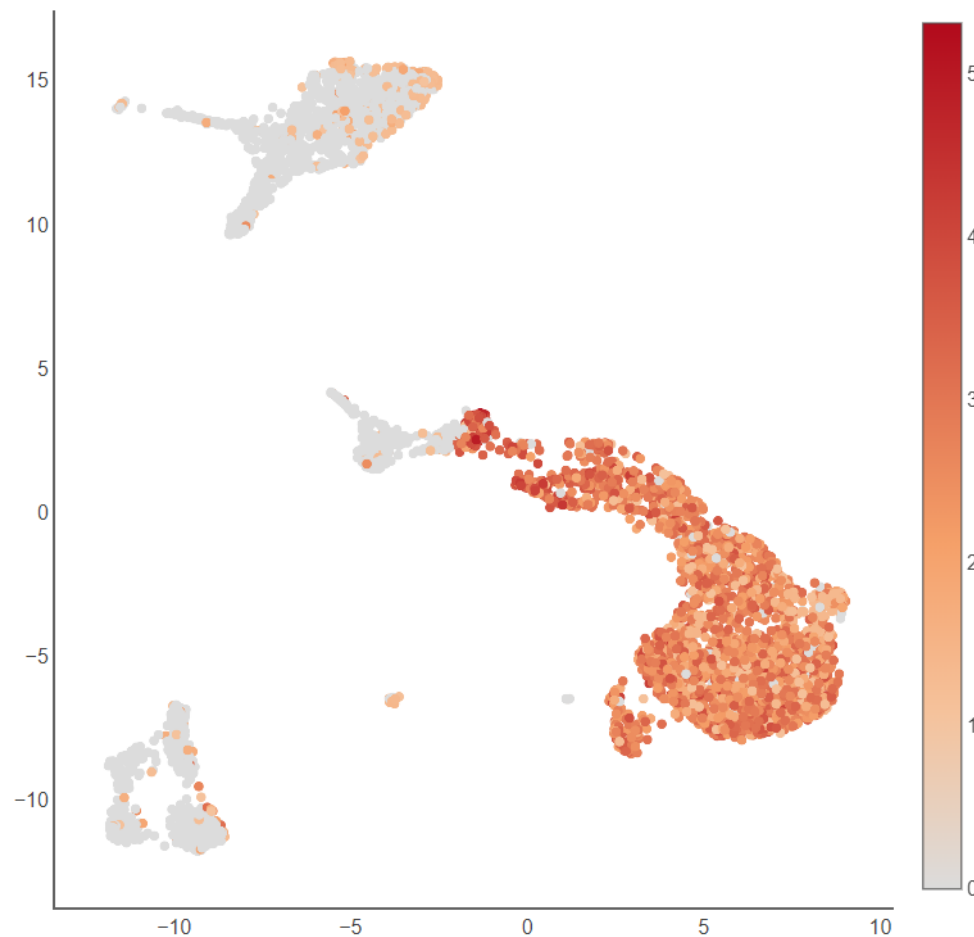
Split by

AVAILABLE ANNOTATIONS

OTHER

► Style (Font size, point size, etc)

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

scNavigator: beta 10x_5k_pbmc ✕

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

COORDINATES

X coordinate
SCT_snn_res.0.6

COLOR AND SPLIT

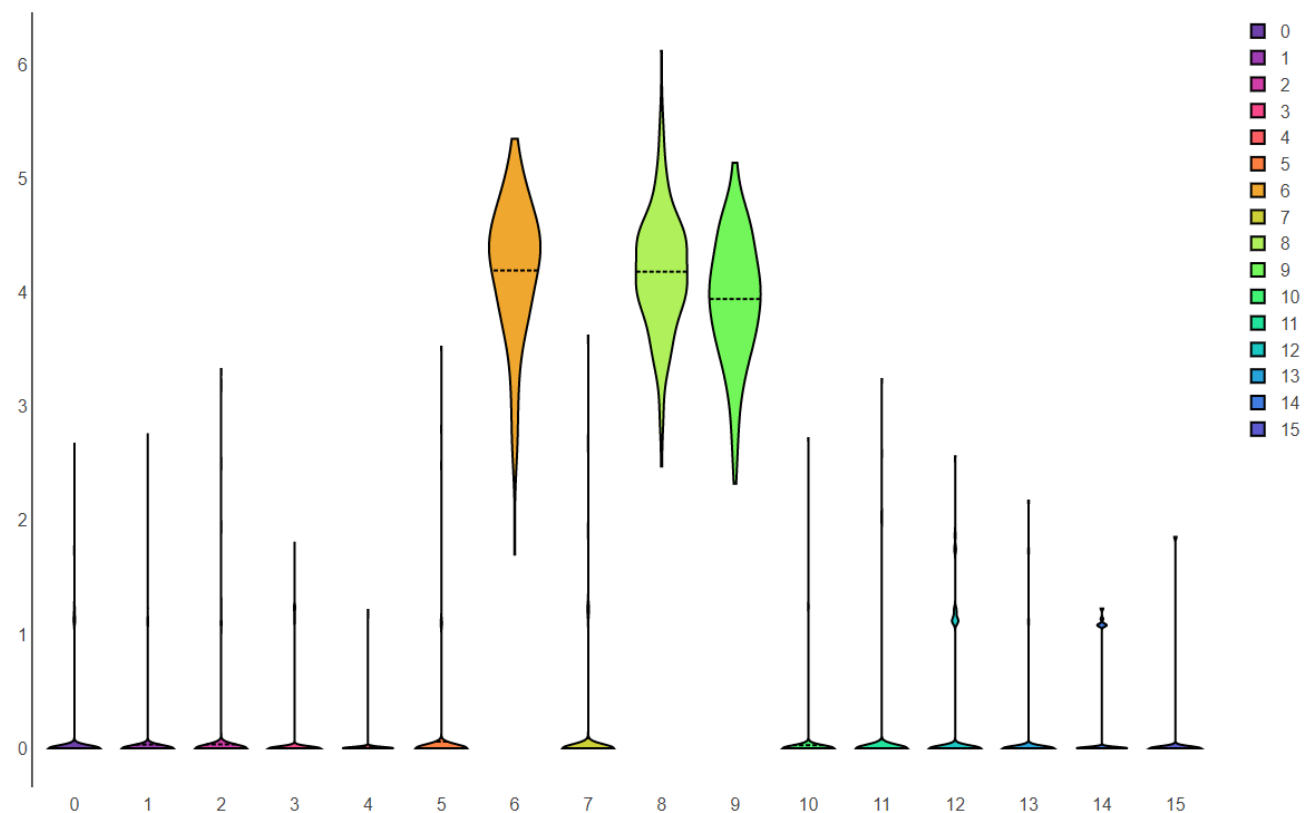
Select gene
CD79A

- ☒ Scaled expression
- ☒ Log-normalize
- ☐ Z-score

Split by
Split by

OTHER

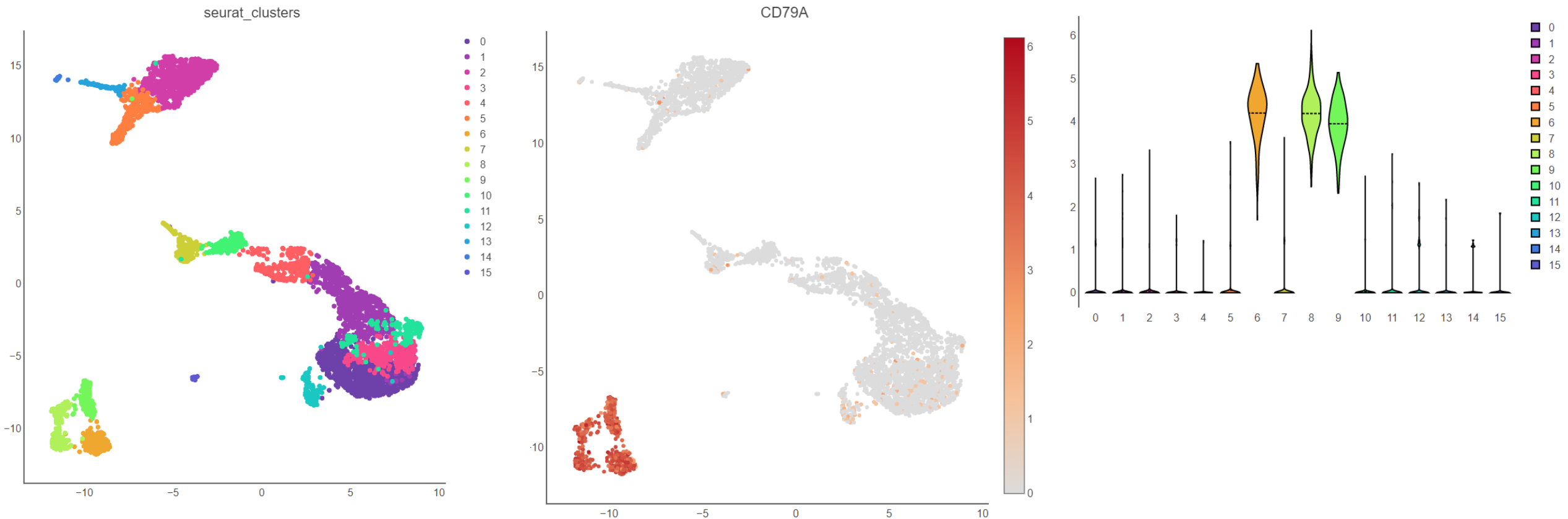
► Style (Font size, point size, etc)



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

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

Markers tab: what's the cluster 7?

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
GNLY	= 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
KLRD1	7	1.872	1.528e-48	2.448e-44	0.99	0.109
PRF1	7	2.076	3.824e-48	6.126e-44	0.995	0.169

- ✓ 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

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)

Public datasets

scNavigator: beta 10x_5k_pbmc

scNavigator: beta

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

Below we have a large collection of datasets and tools to play with:

- Large collection of automatically processed datasets. We processed almost every scRNA-seq dataset from GEO Omnibus database. We make it available for you in our browser.
- Collection of curated datasets. Curated dataset are those that we process by hand. These will include datasets from Human Cell Atlas (HCA), Tabula Muris and some of the datasets that we generated in our lab.
- You can search for cell type specific gene signatures! When we processed all the public scRNA-seq datasets we also calculated all the markers of all the clusters in all these datasets. Just put a list of genes and we will tell you which cluster in which dataset it looks like.
- If you were provided with secret dataset token, you can use it at the very right of this page

Enter a secret token below:

Go!

All scRNA-seq datasets

Curated datasets

Gene signature search

Name	Description	Organism	# of cells	Ext...
GSE101901/SRS2384613	Single cell sequencing of hippocampus tissues in traumatic brain injury	Mus Musculus	8878	🔗
GSE103976/SRS2523512	Detecting Activated Cell Populations Using Single-Cell RNA-Seq	Mus Musculus	6488	🔗
GSE129730/SRS4617144	Single cell RNA-seq shows cellular heterogeneity and lineage expansion in a mouse model of SHH-driven medulloblastoma support resistance to SHH inhibitor therapy	Mus Musculus	4552	🔗
GSE103983/SRS2523775	Single-cell RNA-seq (Drop-seq) of MGE, CGE and LGE of E13.5 (MGE) and E14.5 (CGE, LGE) mouse embryos	Mus Musculus	11704	🔗
GSE93374/SRS1913127	A Molecular Census of Arcuate Hypothalamus and Median Eminence Cell Types	Mus Musculus	61225	🔗
GSE103983/SRS2523784	Single-cell RNA-seq (Drop-seq) of MGE, CGE and LGE of E13.5 (MGE) and E14.5 (CGE, LGE) mouse embryos	Mus Musculus	709	🔗
GSE137007/SRS5355828	Proliferation-competent Tcf1+ CD8 T-cells in dysfunctional populations are CD4 T-cell help independent	Mus Musculus	434	🔗
GSE106960/SRS2690039	The single cell RNA seq of pulmonary alveolar epithelial cells	Mus Musculus	2683	🔗
GSE113111/SRS3165512	sc-RNA sequencing of skeletal muscle macrophages during T. gondii infection and injury	Mus Musculus	6625	🔗
GSE129730/SRS4617149	Single cell RNA-seq shows cellular heterogeneity and lineage expansion in a mouse model of SHH-driven medulloblastoma support resistance to SHH inhibitor therapy	Mus Musculus	5110	🔗

Previous

Page 1 of 35

10 rows

Next

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



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