

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/>

scNavigator: beta

Documentation

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

All scRNA-seq datasets

Curated datasets

Gene signature search

Name	Description	Organism	# of cells	Ext...
GSE93421/SRS1915190	Single-cell RNA-seq of 1.3 million brain cells from E18 mice	Mus Musculus	11911	
GSE123813/SRS4618072	Clonal replacement of tumor-specific T cells following PD-1 blockade [single cells]	Homo Sapiens	650	
GSE135985/SRS5293457	Acquisition of a unique mesenchymal precursor-like blastema state underlies successful adult mammalian digit tip regeneration	Mus Musculus	4253	
GSE121862/SRS3984446	A single-nucleus RNA-sequencing pipeline to decipher the molecular anatomy and pathophysiology of human kidneys	Homo Sapiens	541	
GSE121862/SRS3984436	A single-nucleus RNA-sequencing pipeline to decipher the molecular anatomy and pathophysiology of human kidneys	Homo Sapiens	1206	
GSE114461/SRS3305840	Genetic and transcriptional variation alters cancer cell line drug response [MCF7 strain AA]	Homo Sapiens	1463	
GSE123813/SRS4141611	Clonal replacement of tumor-specific T cells following PD-1 blockade [single cells]	Homo Sapiens	638	
GSE107185/SRS3601149	SEUSS: A scalable screening platform to assess transcriptomic and fitness effects of transcription factor overexpression	Homo Sapiens	641	
SAMEA5846745	Context-specific regulation of monocyte surface IL7R expression and soluble receptor secretion by a common autoimmune risk allele	Homo Sapiens	10538	
GSE103272/SRS2477906	Single cell RNA-seq of ten-thousand epithelium cells from the mammary glands of mice in puberty, estrus and diestrus phases	Mus Musculus	3776	

Previous

Page 1 of 205

10 rows

Next

Enter a secret token below:

Secret token

Go!

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

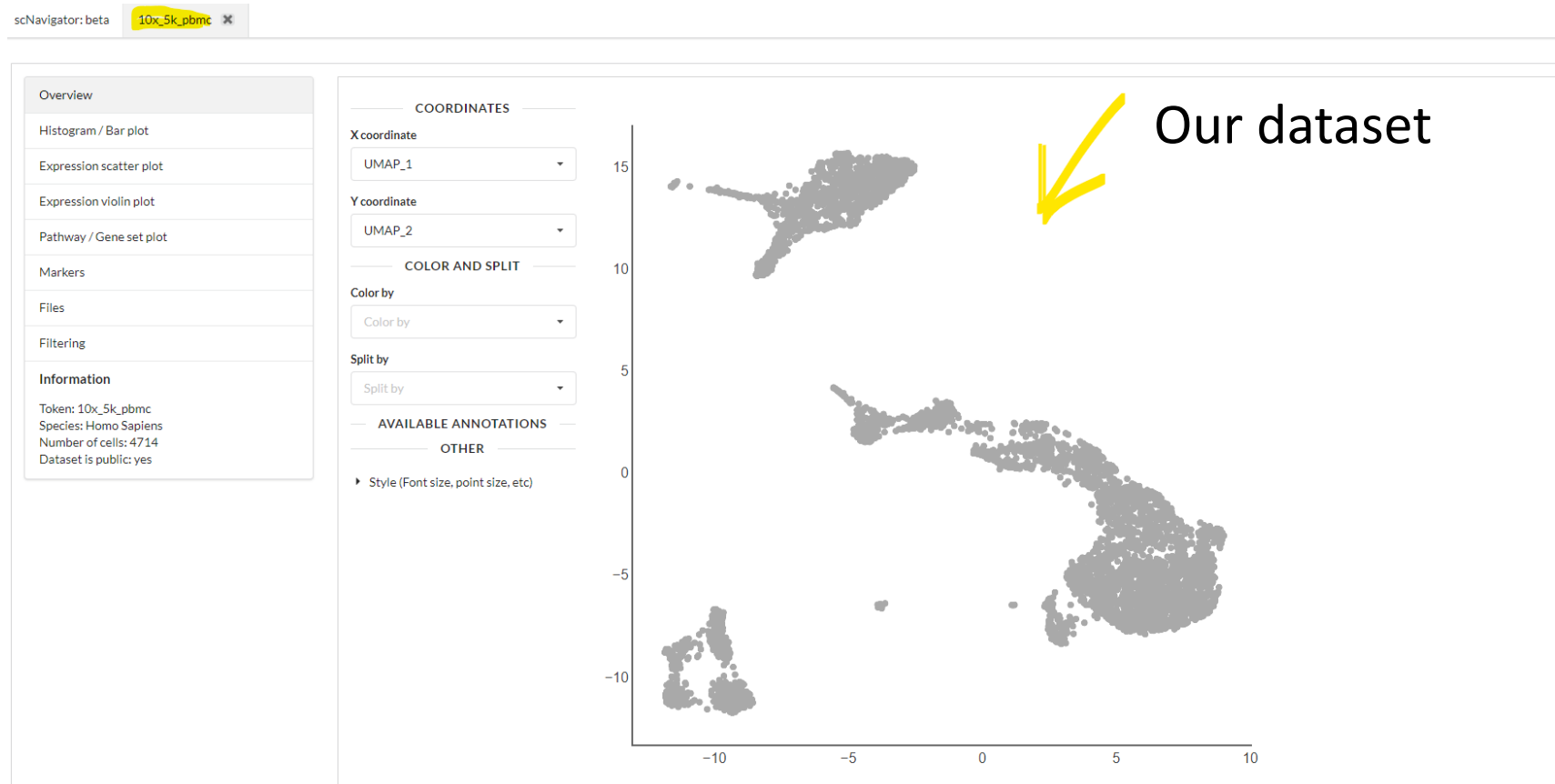
Below we have a large collection of datasets and tools to play with:

- 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



We can color the cells

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

COORDINATES

X coordinate

Y coordinate

COLOR AND SPLIT

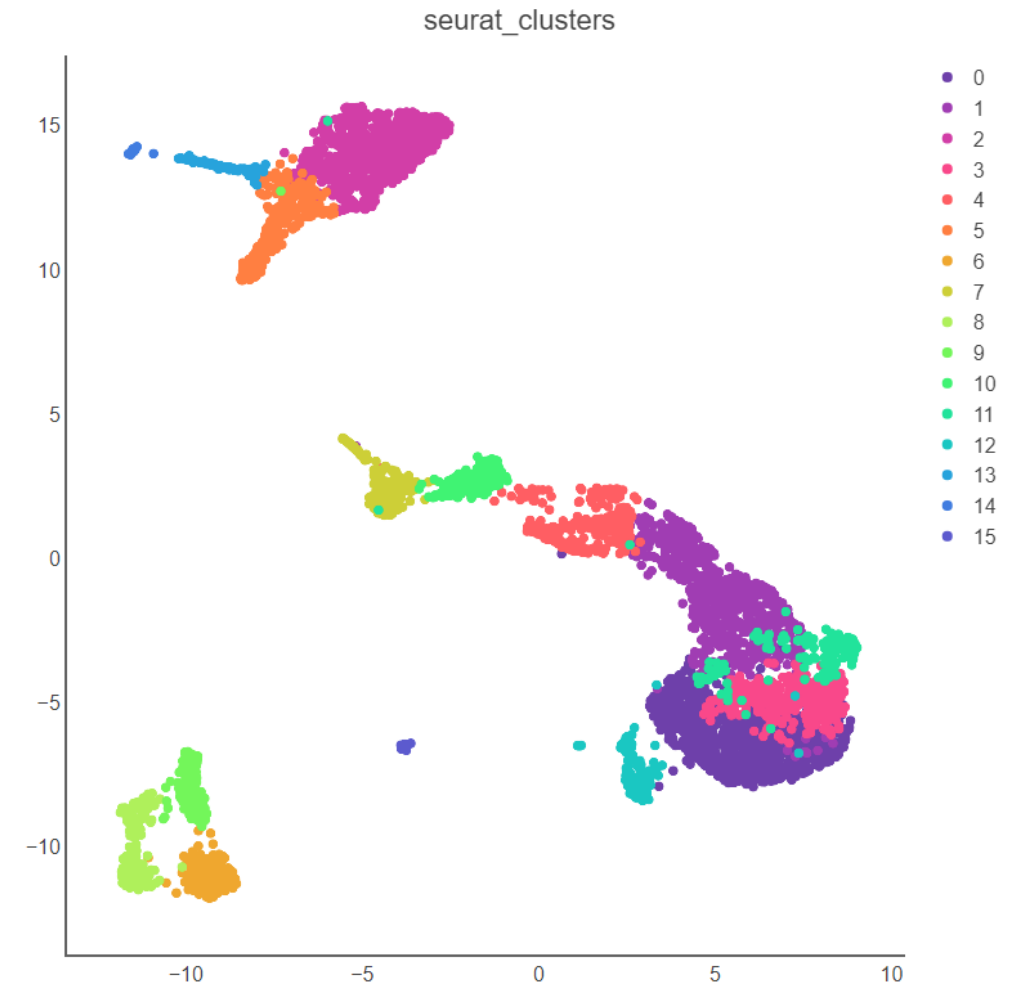
Color by

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

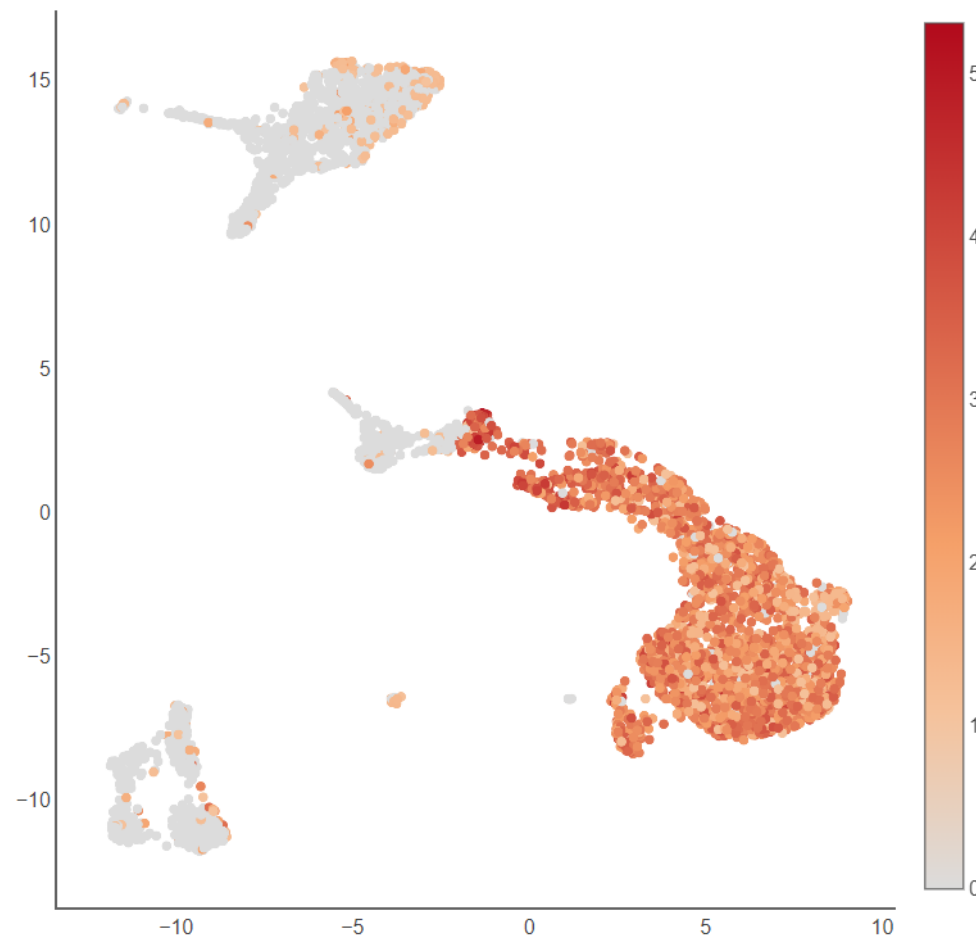
Split by

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

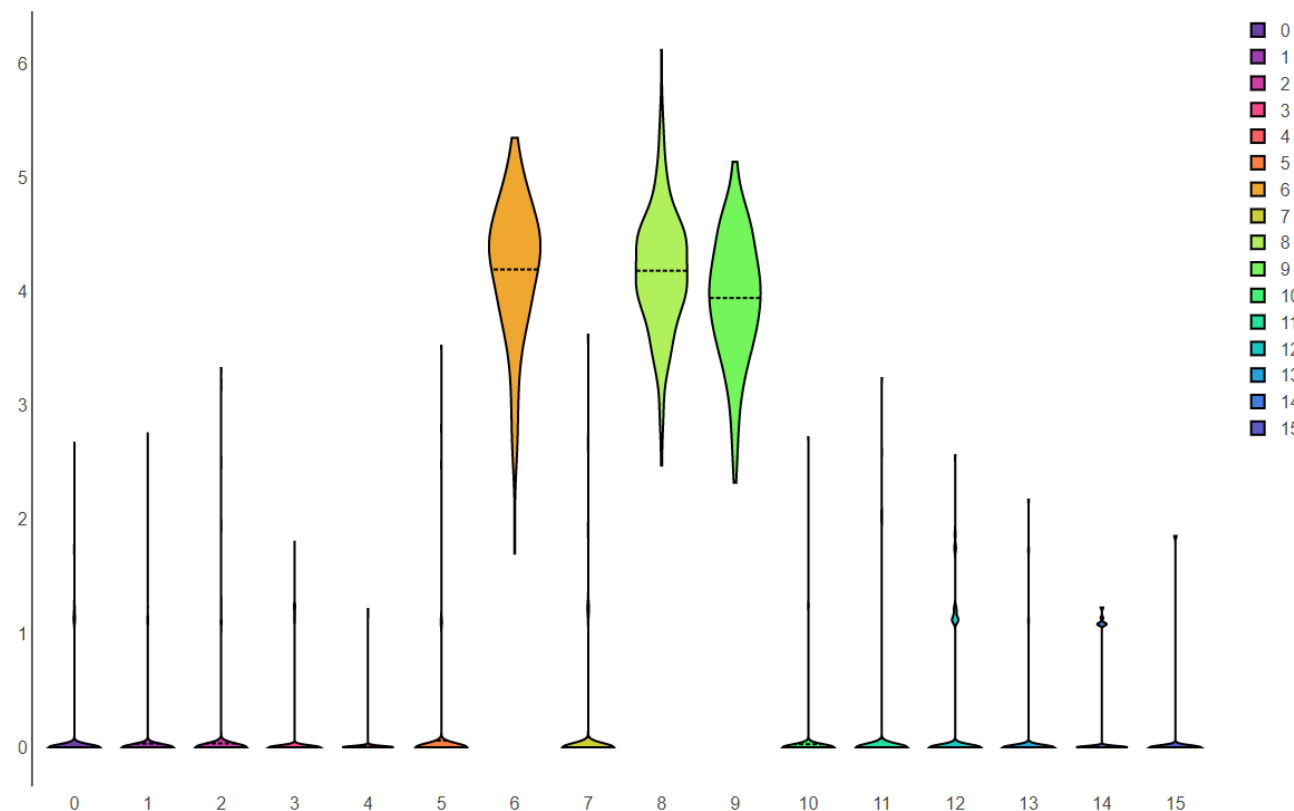
- ☒ Scaled expression
- ☒ Log-normalize
- ☐ Z-score

Split by

Split by

OTHER

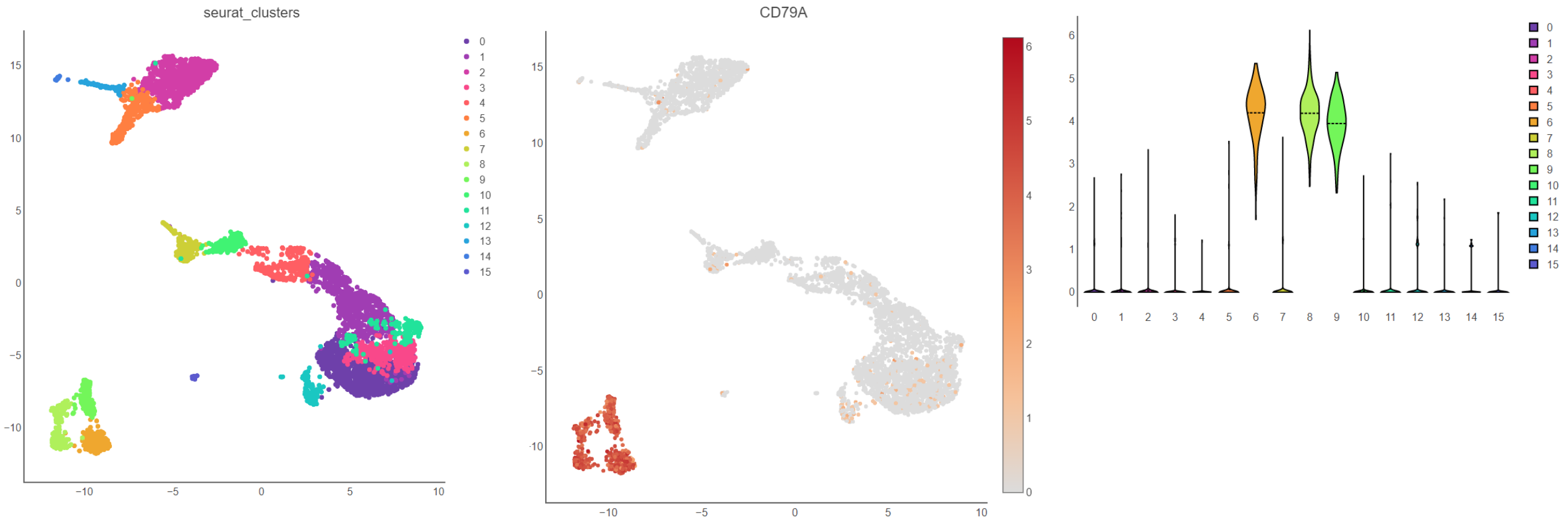
► Style (Font size, point size, etc)



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

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...
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
GSE93421/SRS1915190	Single-cell RNA-seq of 1.3 million brain cells from E18 mice	Mus Musculus	11911	↗
GSE123813/SRS4618072	Clonal replacement of tumor-specific T cells following PD-1 blockade [single cells]	Homo Sapiens	650	↗
GSE135985/SRS5293457	Acquisition of a unique mesenchymal precursor-like blastema state underlies successful adult mammalian digit tip regeneration	Mus Musculus	4253	↗
GSE121862/SRS3984446	A single-nucleus RNA-sequencing pipeline to decipher the molecular anatomy and pathophysiology of human kidneys	Homo Sapiens	541	↗
GSE121862/SRS3984436	A single-nucleus RNA-sequencing pipeline to decipher the molecular anatomy and pathophysiology of human kidneys	Homo Sapiens	1206	↗
GSE114461/SRS3305840	Genetic and transcriptional variation alters cancer cell line drug response [MCF7 strain AA]	Homo Sapiens	1463	↗
GSE123813/SRS4141611	Clonal replacement of tumor-specific T cells following PD-1 blockade [single cells]	Homo Sapiens	638	↗
GSE107185/SRS3601149	SEUSS: A scalable screening platform to assess transcriptomic and fitness effects of transcription factor overexpression	Homo Sapiens	641	↗
SAMEA5846745	Context-specific regulation of monocyte surface IL7R expression and soluble receptor secretion by a common autoimmune risk allele	Homo Sapiens	10538	↗
GSE103272/SRS2477906	Single cell RNA-seq of ten-thousand epithelium cells from the mammary glands of mice in puberty, estrus and diestrus phases	Mus Musculus	3776	↗
Previous		Page 1 of 205	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

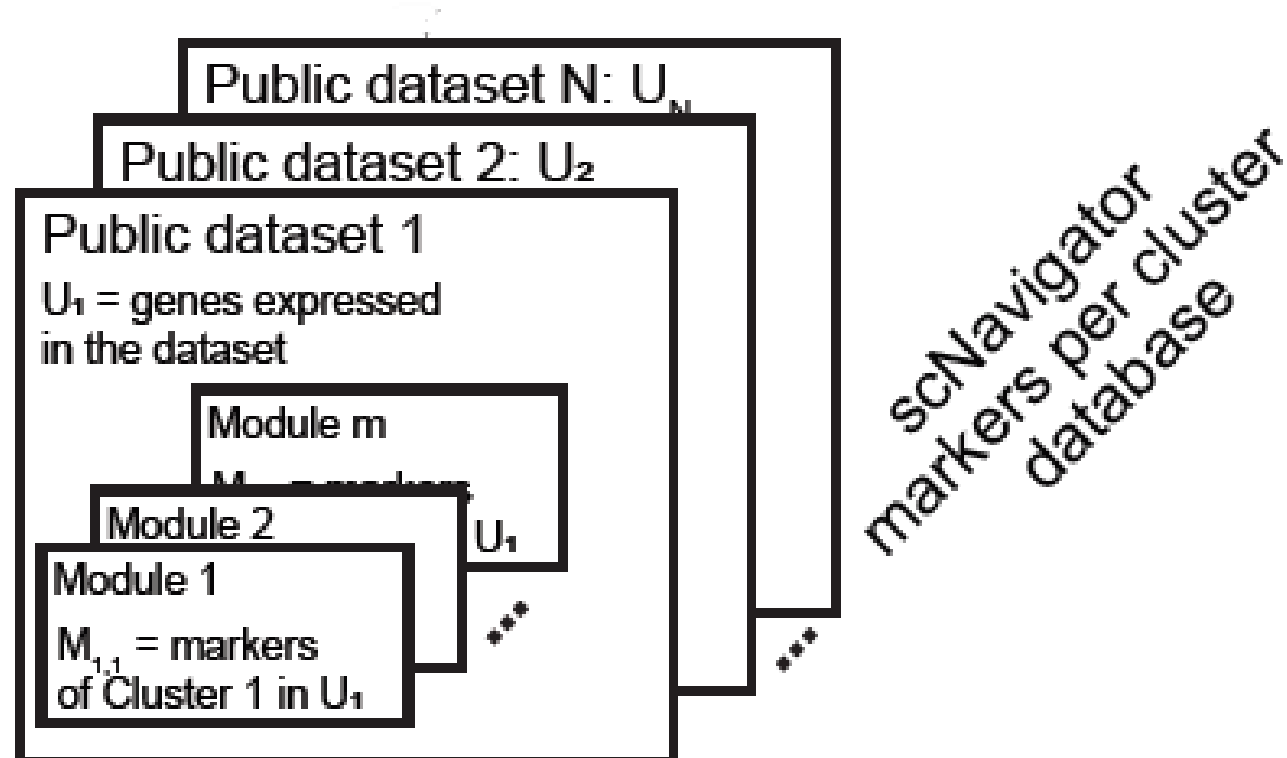
Most of the dataset processing was done
by Maria Firuleva



Lessons from GeneQuery

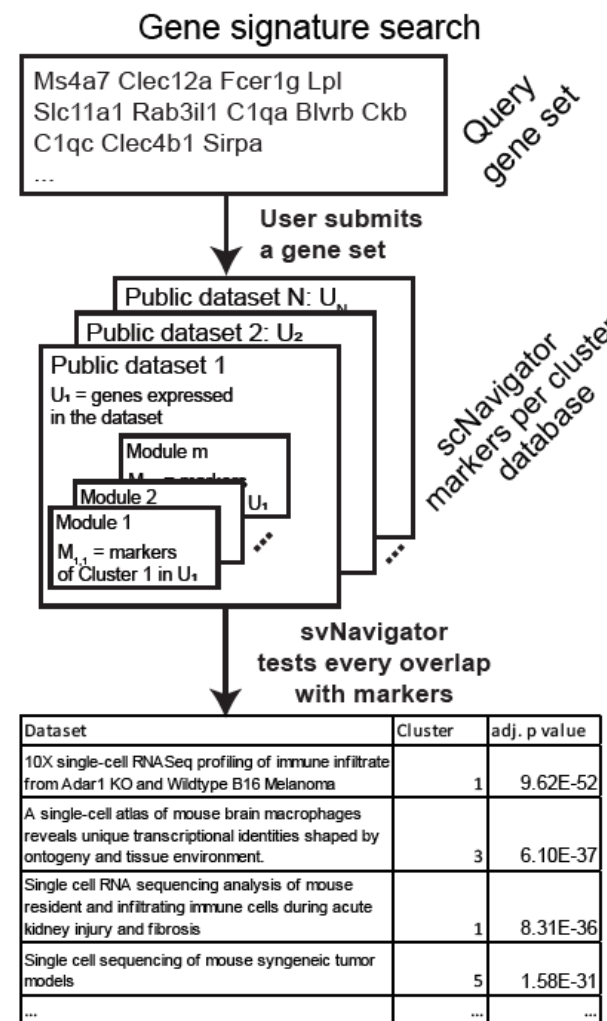
- ✓ We can test gene signatures against other public data
- ✓ We can use public scRNA-seq data: we can calculate markers for all populations found in the data

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



Let's test it:

Human monocyte markers:

CTSS

S100A9

S100A8

LYZ

S100A6

MNDA

VCAN

FTL

Let's test it:

scNavigator: beta

[Documentation](#)

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

Selected species of gene set

☐ Mus Musculus ☒ Homo Sapiens ☐ Rattus Norvegicus

Selected species of dataset (if it's different from gene set we will use orthology to convert genes)

☐ Mus Musculus ☒ Homo Sapiens ☐ Rattus Norvegicus

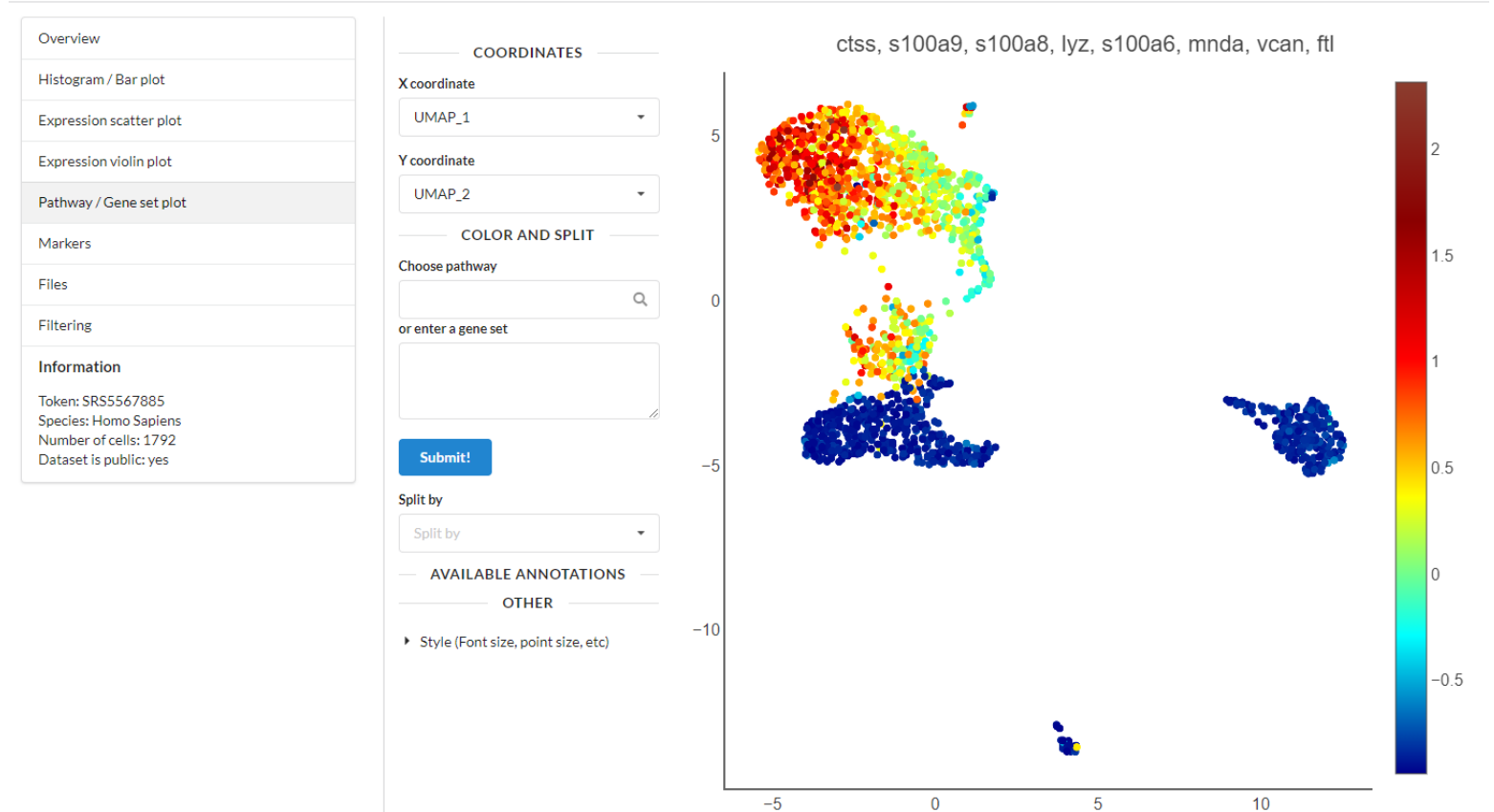
Paste genes in Symbol/Entrez/Ensembl or Refseq format below.

Submit

Name	Show Enrichment	Ext...	Title	Adjusted p v...	Mod...	Inte...
~			~	< 1e-		
GSE139324/SRS5567885	Show enrichment		Immune landscape of viral- and carcinogen-driven head and neck cancer	2.00e-18	31	8
GSE139324/SRS5567889	Show enrichment		Immune landscape of viral- and carcinogen-driven head and neck cancer	3.61e-18	20	8
GSE107747/SRS2737287	Show enrichment		Single-cell RNA sequencing of circulating tumor cells in hepatocellular carcinoma	4.03e-18	37	8

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