

Single-cell Navigator: визуализируем данные scRNA-seq

Константин Зайцев, Университет ИТМО

ИНЦ РАН, 15 октября, 2020

Визуализация данных scRNA-seq

Основные цели

- ✓ Сделать генерацию гипотез проще
- ✓ Убрать «человека по середине»

Дополнительные цели:

- ✓ Быстро работающий
- ✓ Быстро реагирующий

Визуализация данных scRNA-seq

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

(всё ещё в разработке, шлите ваш фидбэк мне на почту)

Откроем датасет

✓ 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

Откроем датасет

- ✓ Переходим по <https://artyomovlab.wustl.edu/scn/>
- ✓ Ищем 10x
- ✓ Кликаем на датасет

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		

Если есть проблемы

- ✓ Просто переходите по ссылке
https://artyomovlab.wustl.edu/scn/?token=10x_5k_pbmc

Результат должен выглядеть как-то так



Можем покрасить клетки по

- ✓ Кластер
- ✓ Количество UMIs
- ✓ Количество генов
задетектировано

COORDINATES

X coordinate
UMAP_1

Y coordinate
UMAP_2

COLOR AND SPLIT

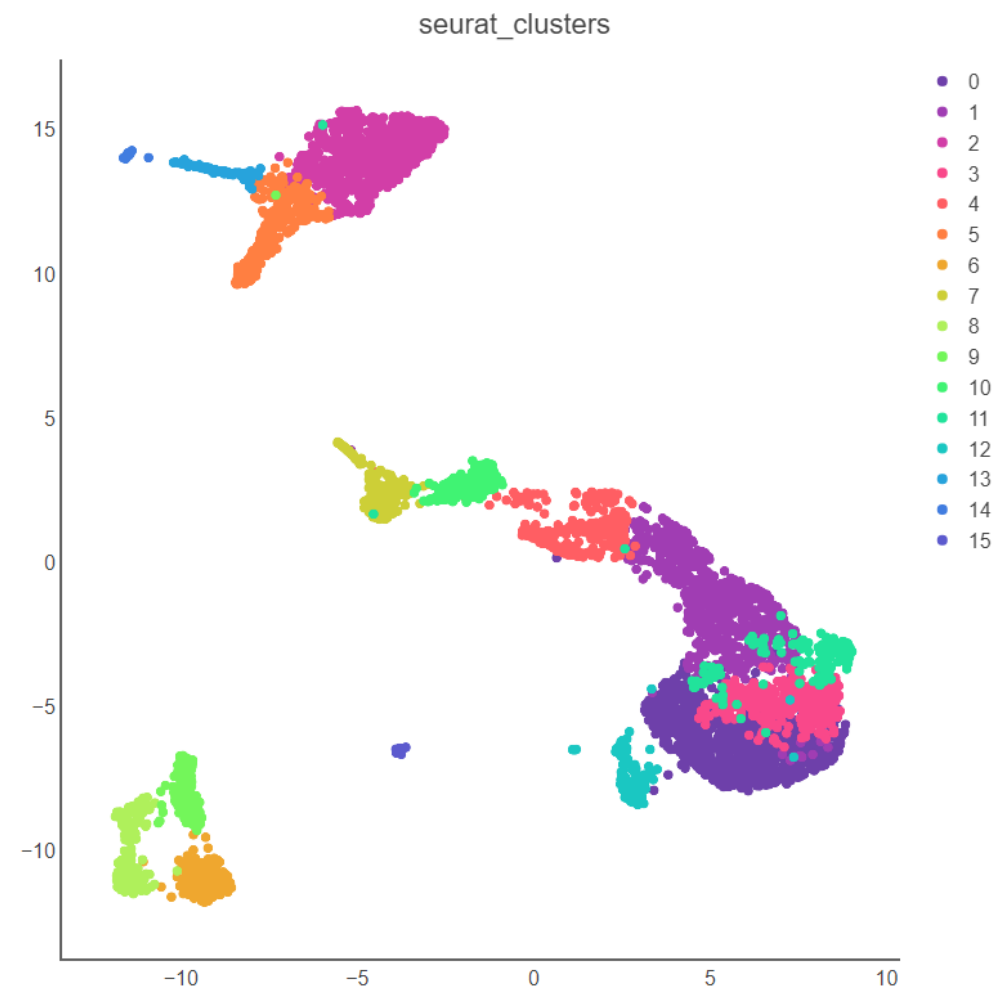
Color by
seurat_clusters

Split by
Split by

AVAILABLE ANNOTATIONS

OTHER

Style (Font size, point size, etc)



Экспрессия 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

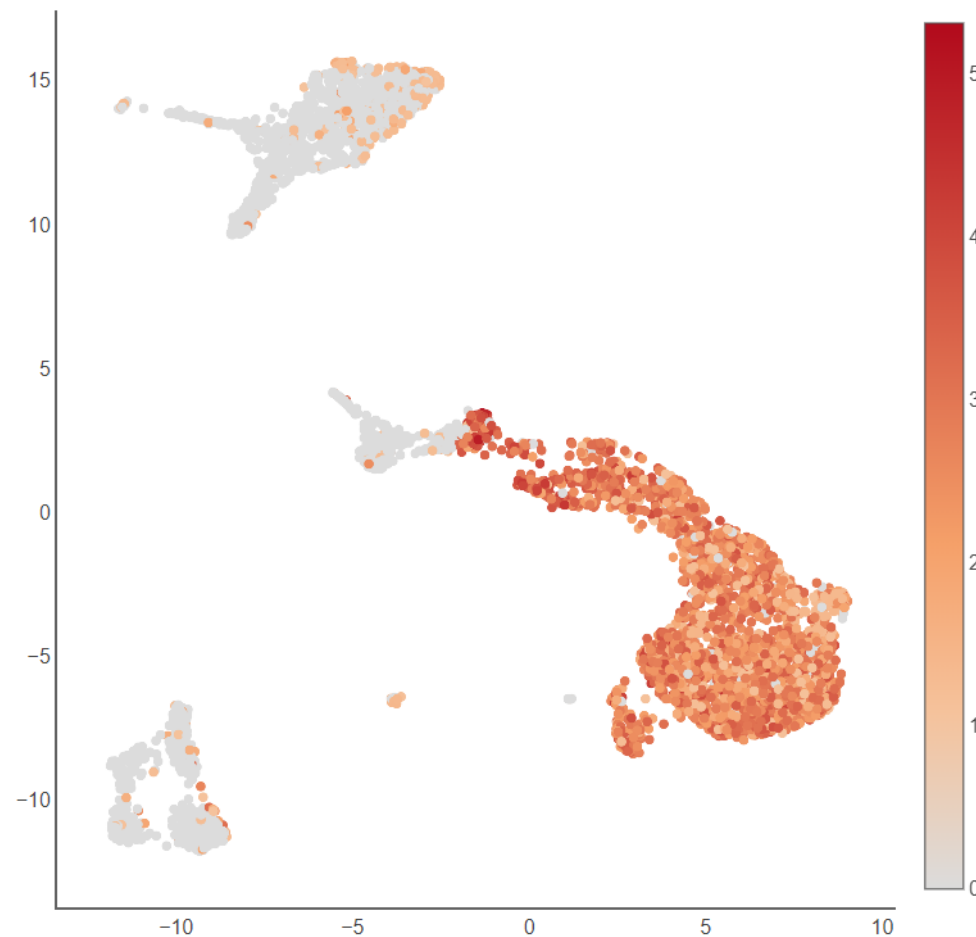
Split by

AVAILABLE ANNOTATIONS

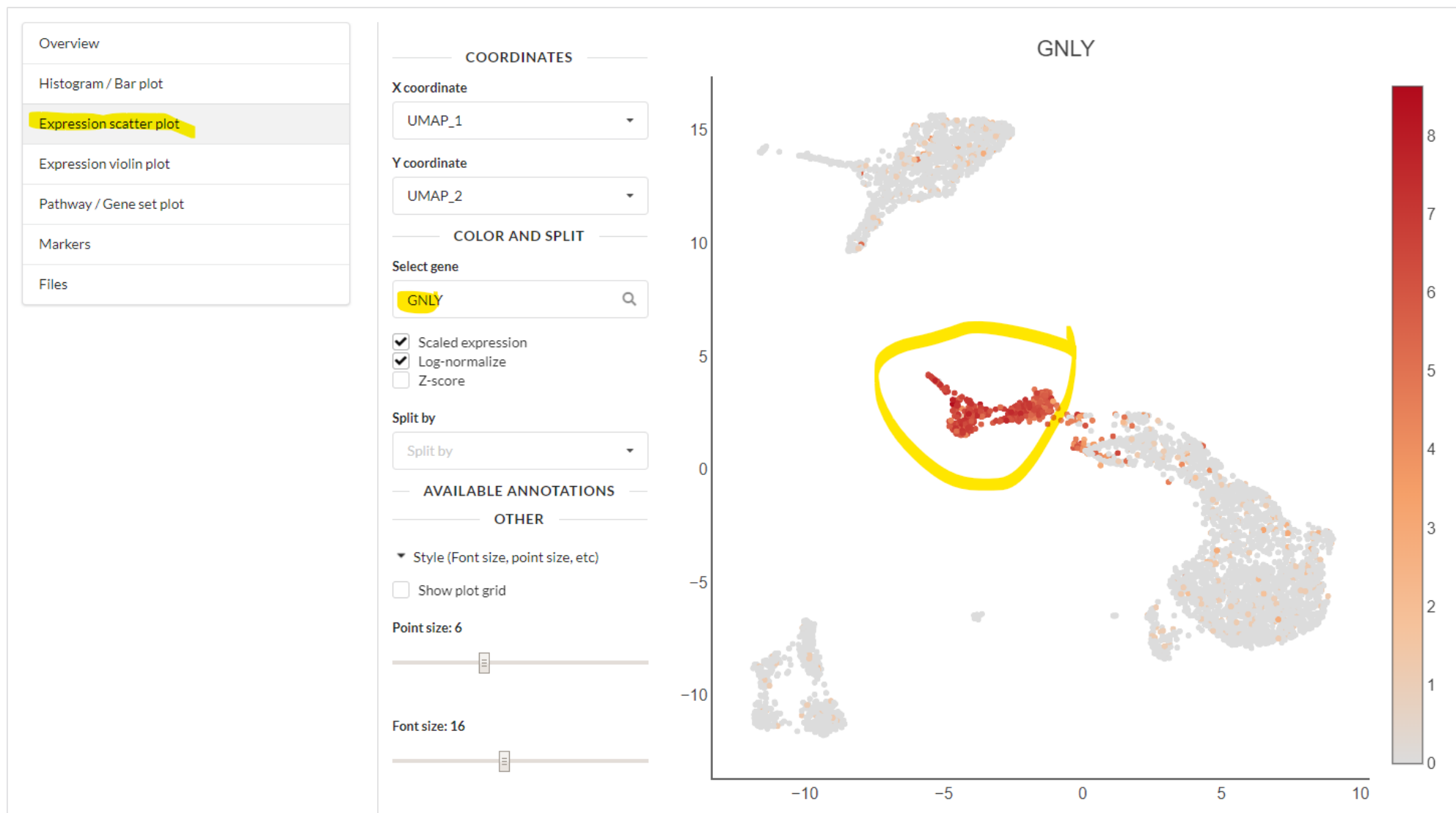
OTHER

- Style (Font size, point size, etc)

CD3D



Можете выбрать ваш любимый ген



Expression scatter plot

- ✓ Expression scatter plot показывает экспрессию генов в каждой клетке поверх графика понижения размерности
- ✓ Можно видеть как экспрессия некоторых генов локализуется вместе с кластерами

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

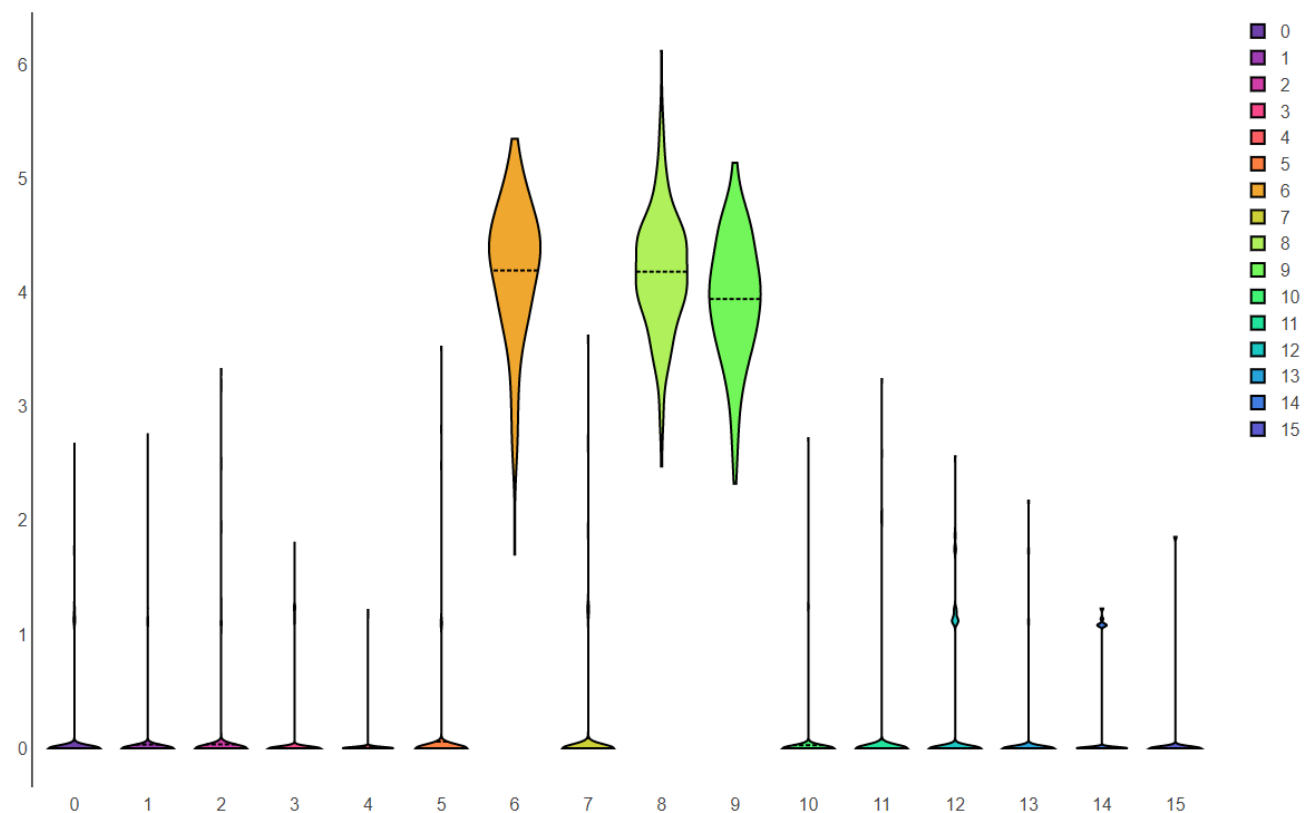
Select gene
CD79A

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

Split by
Split by

OTHER

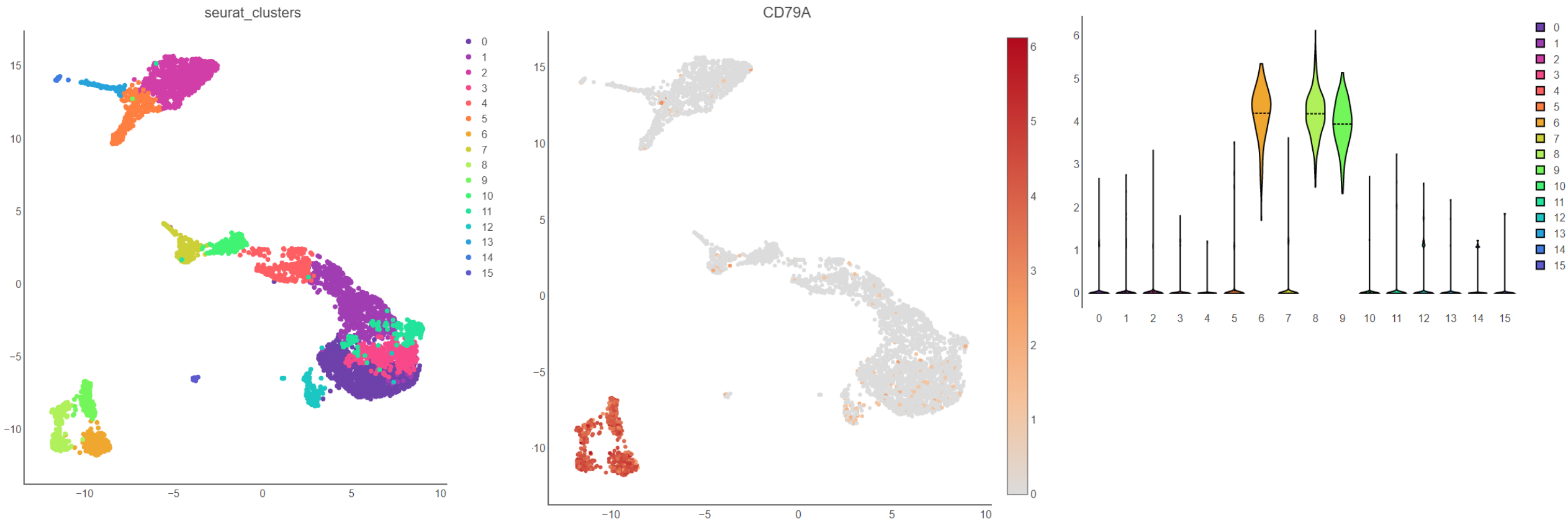
► Style (Font size, point size, etc)



Violin plot

- ✓ Violin plot показывает **распределение** значений экспрессии генов в нескольких группах клеток (например в кластерах)
- ✓ Выше распределение – выше экспрессия в группе

Cd79a: expression scatter and expression violin plots



Маркеры

- ✓ Обычно мы запускаем тесты дифференциальной экспрессии, чтобы найти маркеры кластеров
- ✓ Дифференциальная экспрессия – для каждого кластера, сравним кластер со всеми остальными

Маркеры

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

Page1of 646

10 rows

Next

Previous

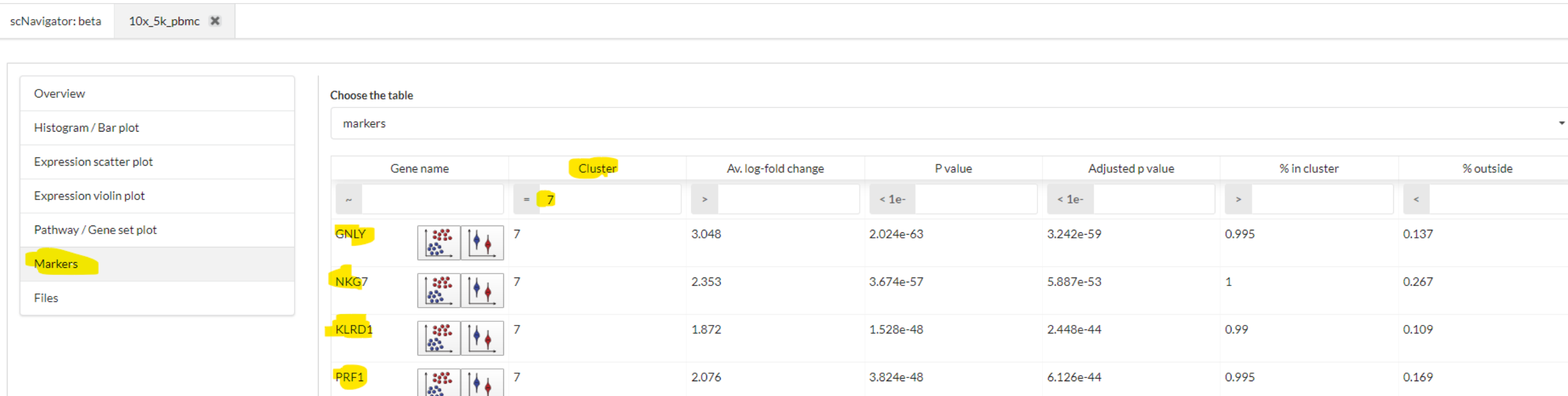
Page 1 of 646

10 rows

Next

Download current table

Маркеры : что находится в кластере 7 ?



- ✓ GNLY – имя гена
- ✓ Cluster 7 – мы тестировали кластер 7 против остальных
- ✓ Average log-fold change: разница в средней экспрессии гена GNLY между кластером 7 и другими кластерами
- ✓ P value (мы тестируем разницу средней экспрессии)
- ✓ P adjusted – p value поправленное на множественное сравнение
- ✓ % in and outside of the cluster – в сколько процентах клеток GNLY детектирован внутри и вне кластера 7

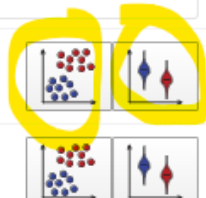
Маркеры : что находится в кластере 7 ?

- ✓ Две кнопки для удобства
- 1) Первая откроет expression scatter plot
- 2) Вторая откроет expression 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



Можно поиграть и позадавать мне вопросы

Публичные датасеты

- ✓ В данный момент мы пытаемся обработать как можно больше публичных датасетов и сделать их доступными в навигаторе
- ✓ Вы всегда можете проверить наличие того или иного датасета на главной странице

Публичные датасеты

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

Публичные датасеты scRNA-seq

Большинство публичных датасетов single-cell RNA-seq доступны через NCBI GEO (or SRA)

Проблемы:

- ✓ Разные технологии используются для проведения эксперимента (10x, DropSeq, SmartSeq2, C1 Fluidigm etc)
- ✓ Исследователи часто используют разные пайплайны
- ✓ Разные форматы для хранения данных

Большая часть работы по процессингу датасетов была проделана Марией Фирулевой



Выводы

- ✓ Мы надеемся, что навигатор сделает интерпретацию данных scRNA-seq проще и быстрее
- ✓ <https://artyomovlab.wustl.edu/scn/>
- ✓ Мы пытаемся обработать как можно больше датасетов
- ✓ Если вы хотите использовать SCN для своих данных:
 - Можете просто написать мне email kzaitsev@itmo.ru, и мы обсудим, как это делать
 - Нужно подождать пока мы это опубликуем (ETA?), и тогда вы сможете поднять scNavigator у себя сами