



ITMO UNIVERSITY

CT Lab
ITMO UNIVERSITY

Visual and exploratory analysis in gene expression studies

Konstantin Zaitsev

November 28th, 2020

Visual exploratory analysis

- ✓ Sometimes visual data exploration is the only way to know if something went wrong
- ✓ Sometimes we can perform such analysis without even opening R

Tools

- ✓ Phantasus
- ✓ Single-cell Navigator
- ✓ JBR genome browser

Conflict of interest alert

- ✓ Phantasus (ITMO University)
- ✓ Single-cell Navigator (ITMO University)
- ✓ JBR genome browser (JetBrains Research)

We tell people about these tools because we believe these tools are awesome

Exploratory gene expression analysis

- ✓ GENE-E (Joshua Gould in Broad Institute) pioneered the way
<https://software.broadinstitute.org/GENE-E/>
- ✓ No longer supported :(
- ✓ They developed Morpheus that runs in web
<https://software.broadinstitute.org/morpheus/>
- ✓ Web-based, everything is done client-side
- ✓ Limited functionality

Phantasus – Morpheus integrated with R

- ✓ An extension developed by Daria Zenkova & Vlad Kamenev at ITMO University (and the grey eminence Alexey Sergushichev)
- ✓ Server-side application -> requires internet access (unless installed locally)
- ✓ Can be easily extended to support different R/Bioconductor packages
- ✓ Free and open-source



Phantasus can be accessed in multiple ways

Online:

- ✓ <https://ctlab.itmo.ru/phantasus/>
- ✓ <https://artyomovlab.wustl.edu/phantasus/>

It can be installed locally from Bioconductor

- ✓ <http://bioconductor.org/packages/phantasus>

As a docker image:

- ✓ <https://hub.docker.com/r/dzenkova/phantasus>

Phantasus perks

- ✓ We can perform some basic steps quickly
- ✓ We can access a lot of public datasets
(microarrays and RNA-seq data that is in ARCHS4)
- ✓ PCA plots / Basic DE / Gene Set enrichment

Let's look at GSE53986

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE53986>

Series GSE53986		Query DataSets for GSE53986
Status	Public on Mar 31, 2014	
Title	NRROS negatively regulates ROS in phagocytes during host defense and autoimmunity	
Organism	Mus musculus	
Experiment type	Expression profiling by array	
Summary	<p>Production of reactive oxygen species (ROS) is one of the important antimicrobial mechanisms of phagocytic cells. Enhanced oxidative burst requires these cells to be primed with agents such as IFNg and LPS with a synergistic effect of these agents on the level of the burst. However, excessive ROS generation will lead to tissue damage and has been implicated in a variety of inflammatory and autoimmune disease. Therefore, this process needs to be tightly regulated. In order to understand the genes regulating this process, we will treat bone marrow derived macrophages with above mentioned priming agents and study the gene expression.</p> <p>We used microarrays to determine the changes in gene expression that occur in bone marrow derived macrophages after treatment with IFNg, LPS, or a combination of IFNg and LPS</p>	
Overall design	Four condition experiment; Biological replicates: four replicates per condition	
Contributor(s)	Noubade R , Wong K , Ota N , Rutz S , Eidenschenk C , Ding J , Valdez PA , Peng I , Sebrell A , Caplazi P , DeVoss J , Soriano RH , Modrusan Z , Hackney JA , Sai T , Ouyang W	
Citation(s)	Noubade R, Wong K, Ota N, Rutz S et al. NRROS negatively regulates reactive oxygen species during host defence and autoimmunity. <i>Nature</i> 2014 May 8;509(7499):235-9. PMID: 24739962	

Samples from GSE53986

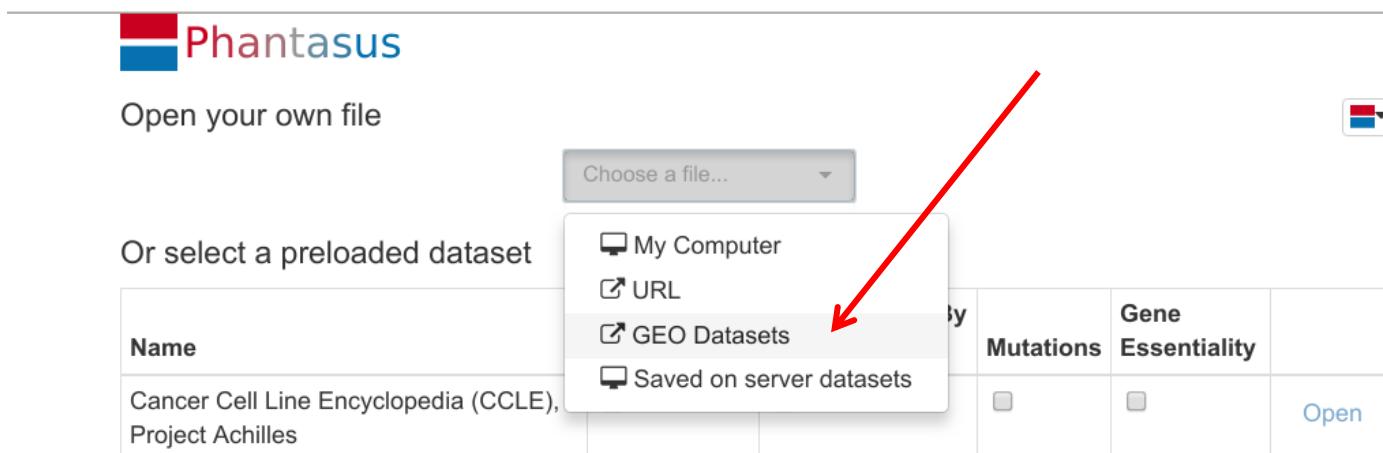
Samples (16)

[☰ Less...](#)

- [GSM1304836](#) BMDM, untreated, 1
- [GSM1304837](#) BMDM, untreated, 2
- [GSM1304838](#) BMDM, untreated, 3
- [GSM1304839](#) BMDM, untreated, 4
- [GSM1304840](#) BMDM, IFNg, 1
- [GSM1304841](#) BMDM, IFNg, 2
- [GSM1304842](#) BMDM, IFNg, 3
- [GSM1304843](#) BMDM, IFNg, 4
- [GSM1304844](#) BMDM, LPS, 1
- [GSM1304845](#) BMDM, LPS, 2
- [GSM1304846](#) BMDM, LPS, 3
- [GSM1304847](#) BMDM, LPS, 4
- [GSM1304848](#) BMDM, IFNg+LPS, 1
- [GSM1304849](#) BMDM, IFNg+LPS, 2
- [GSM1304850](#) BMDM, IFNg+LPS, 3
- [GSM1304851](#) BMDM, IFNg+LPS, 4

Let's explore this dataset

- ✓ Open <https://ctlab.itmo.ru/phantasus/> or
- ✓ Open <https://artyomovlab.wustl.edu/phantasus/>
- ✓ Load dataset into phantasus:
 - Choose a file/GEO Datasets/GSE53986

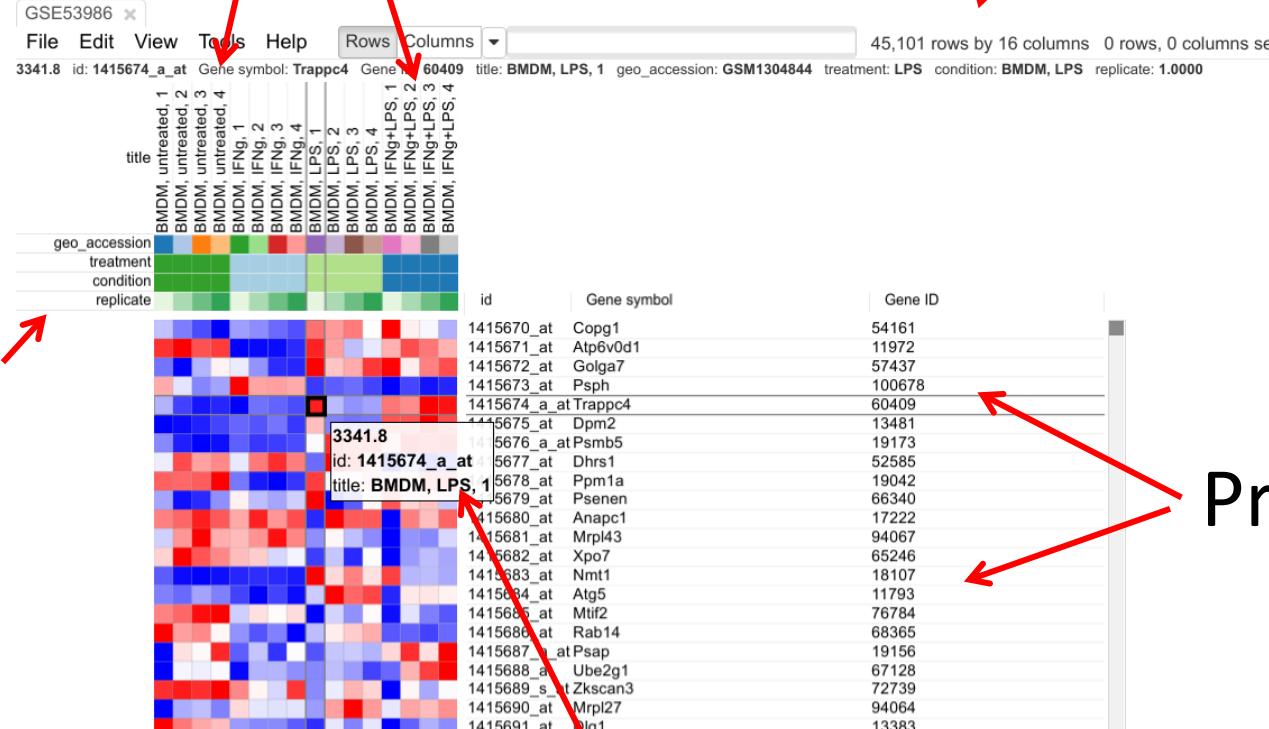


Interface overview

Samples (pData)

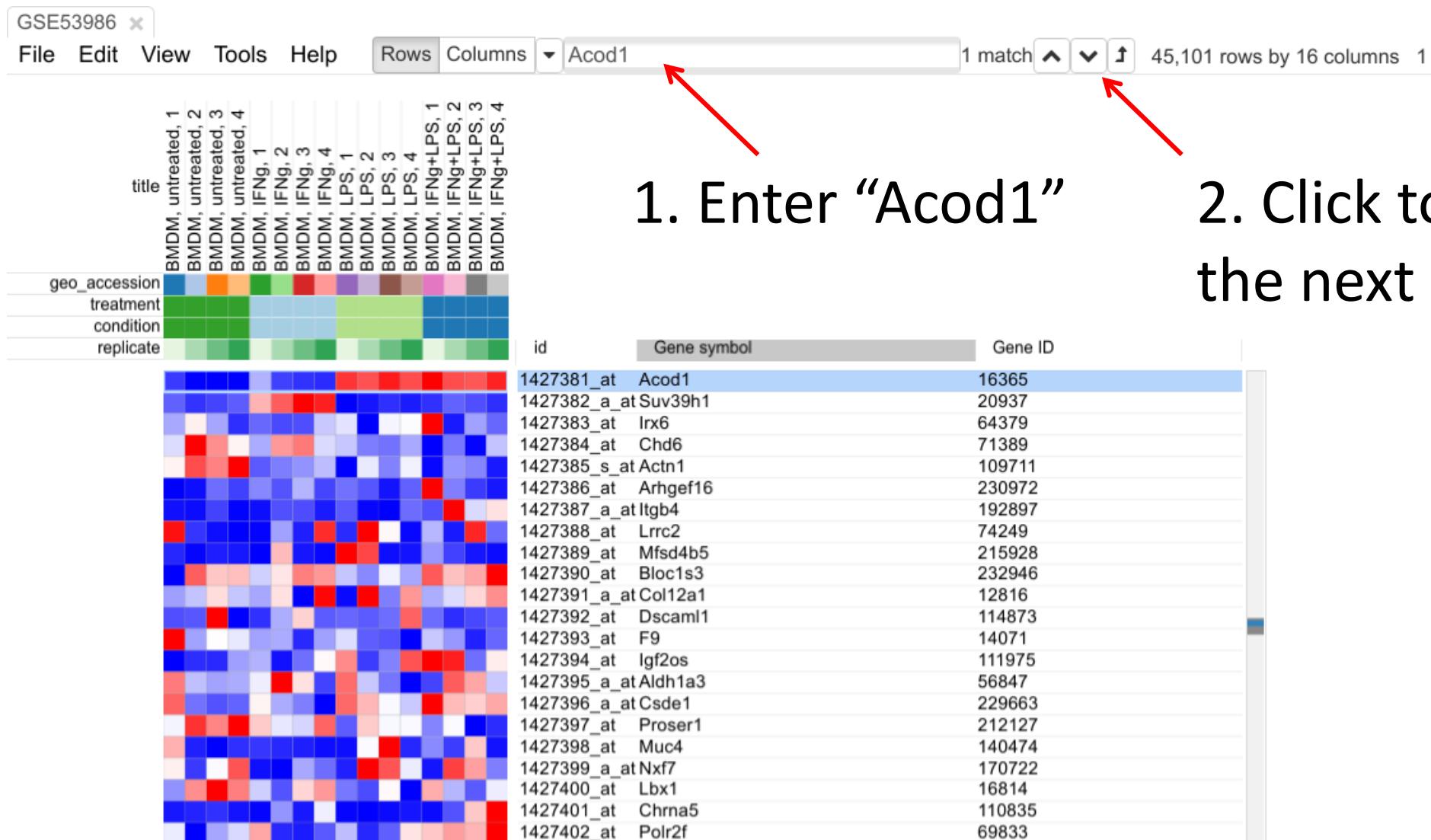
Dataset dimension

Sample
annotations
(right click
for context
menu)



Expression value (color scheme is relative)

Exploring individual genes



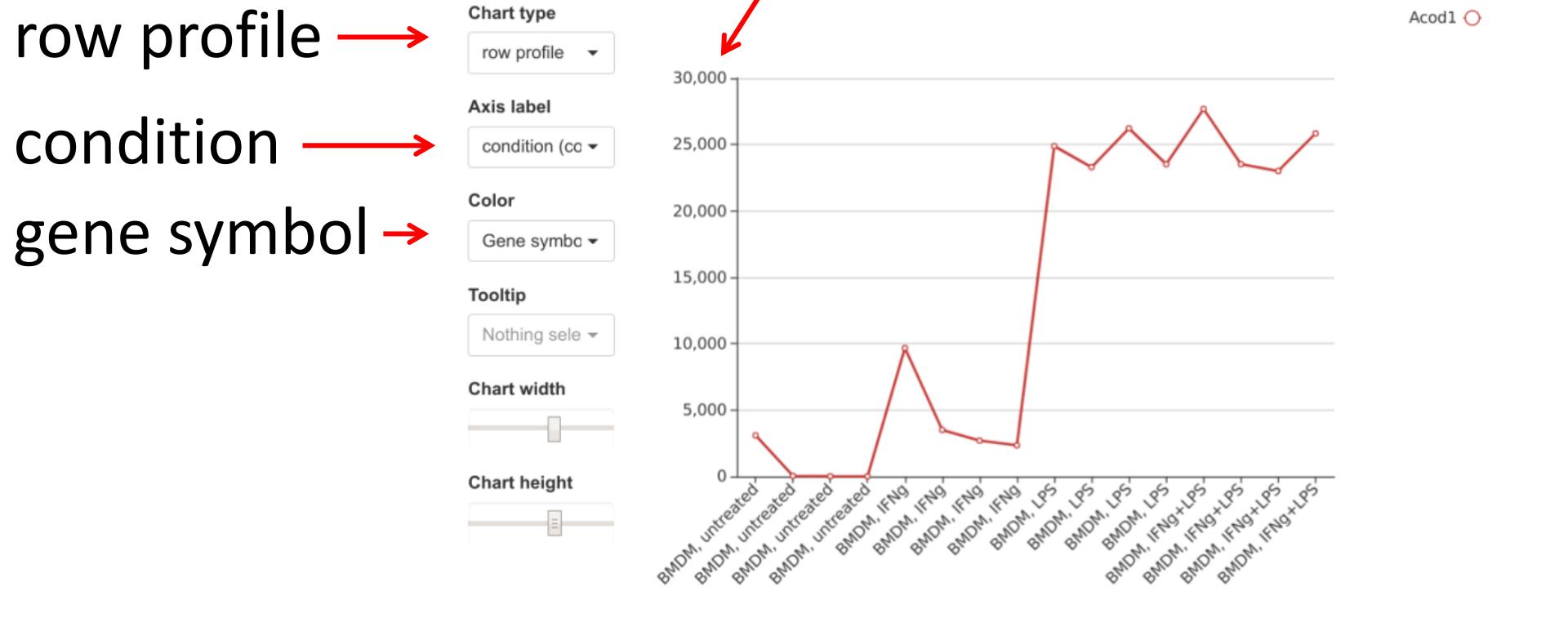
1. Enter “Acod1”

2. Click to scroll to
the next hit

Row profile chart

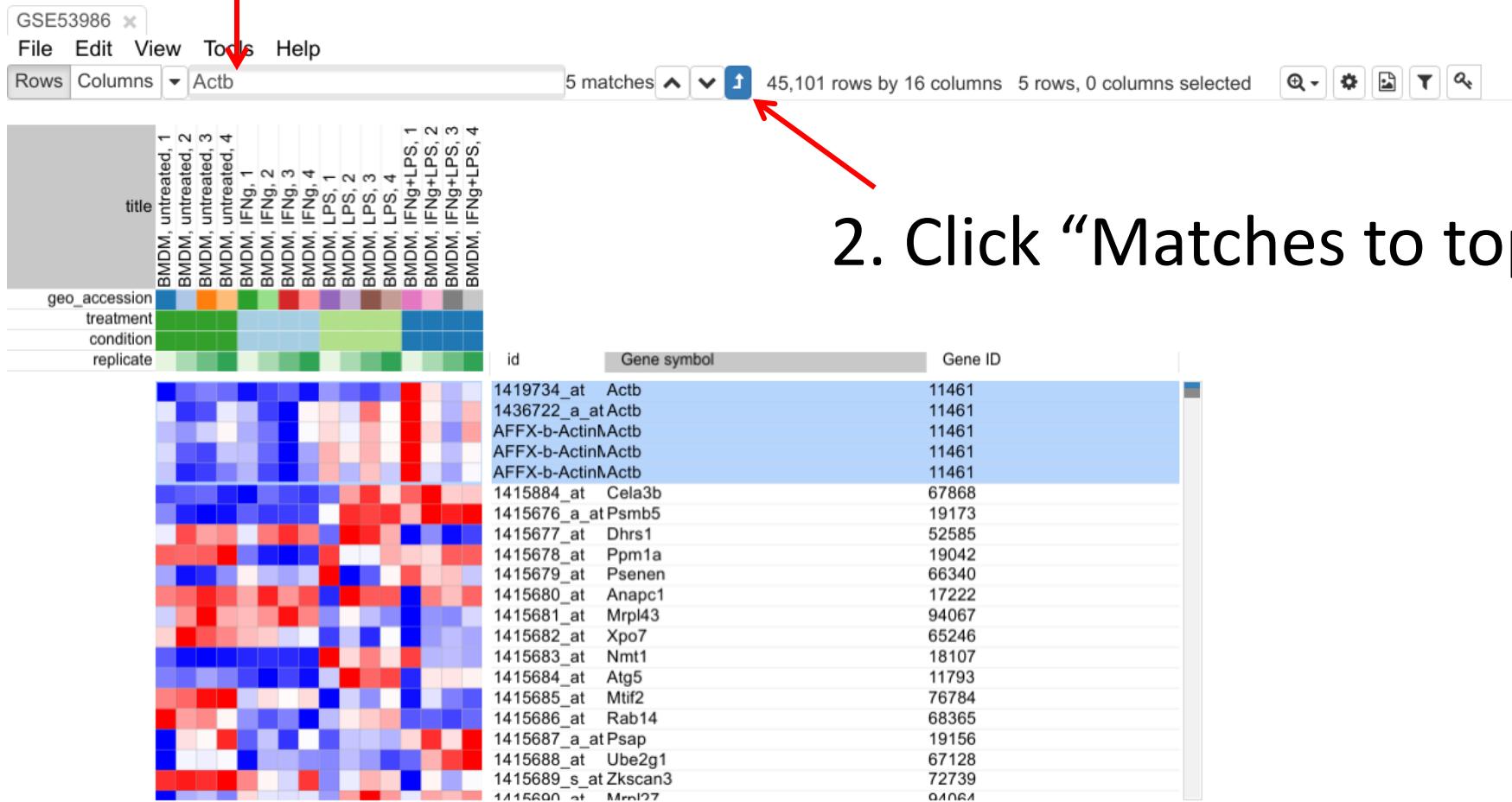
- ✓ Select all columns and Acod1 row
- ✓ Tools/Plots/Chart

Data is in linear scale!



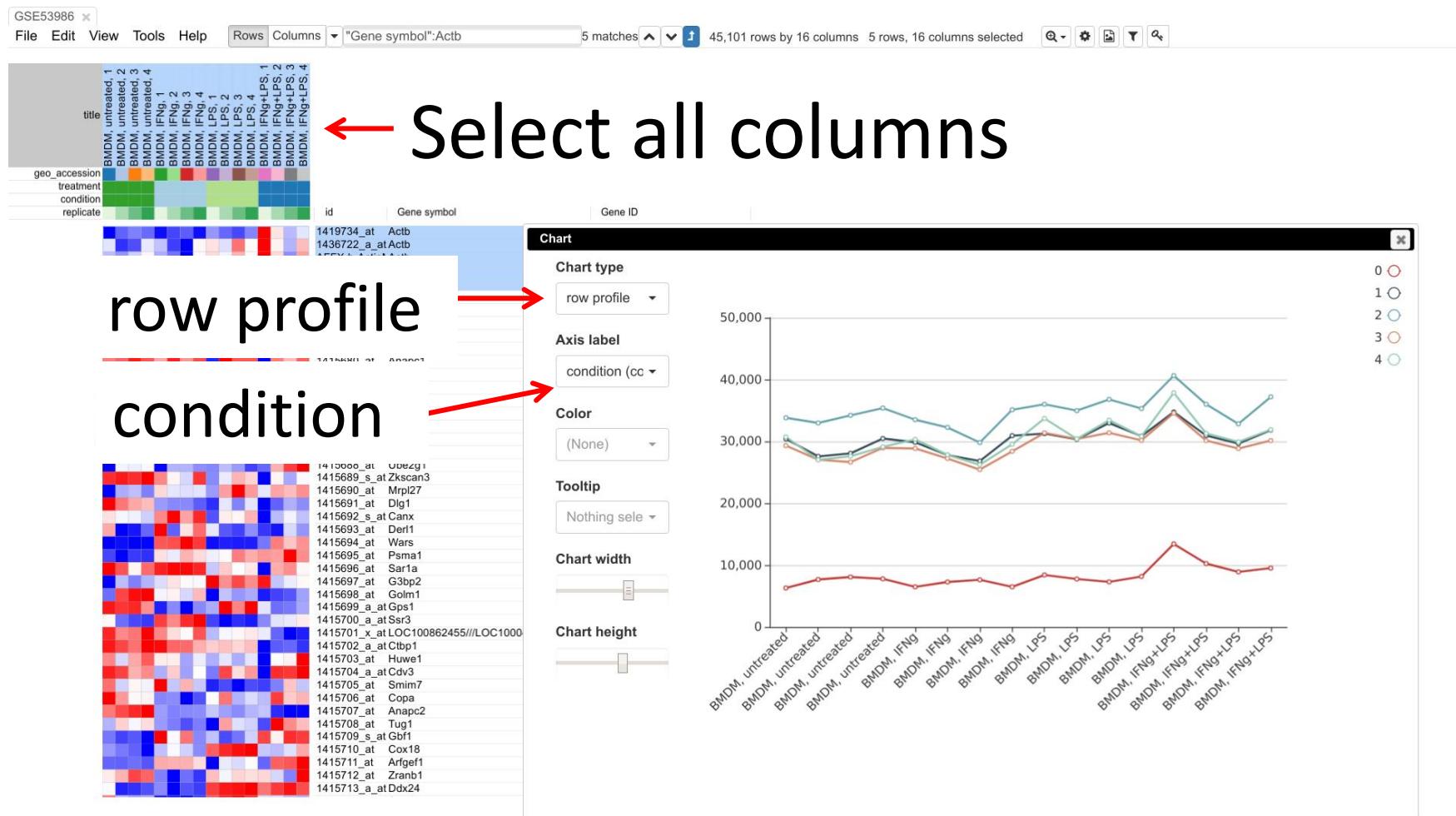
Let's look at Actb as a control

1. Enter “Actb”



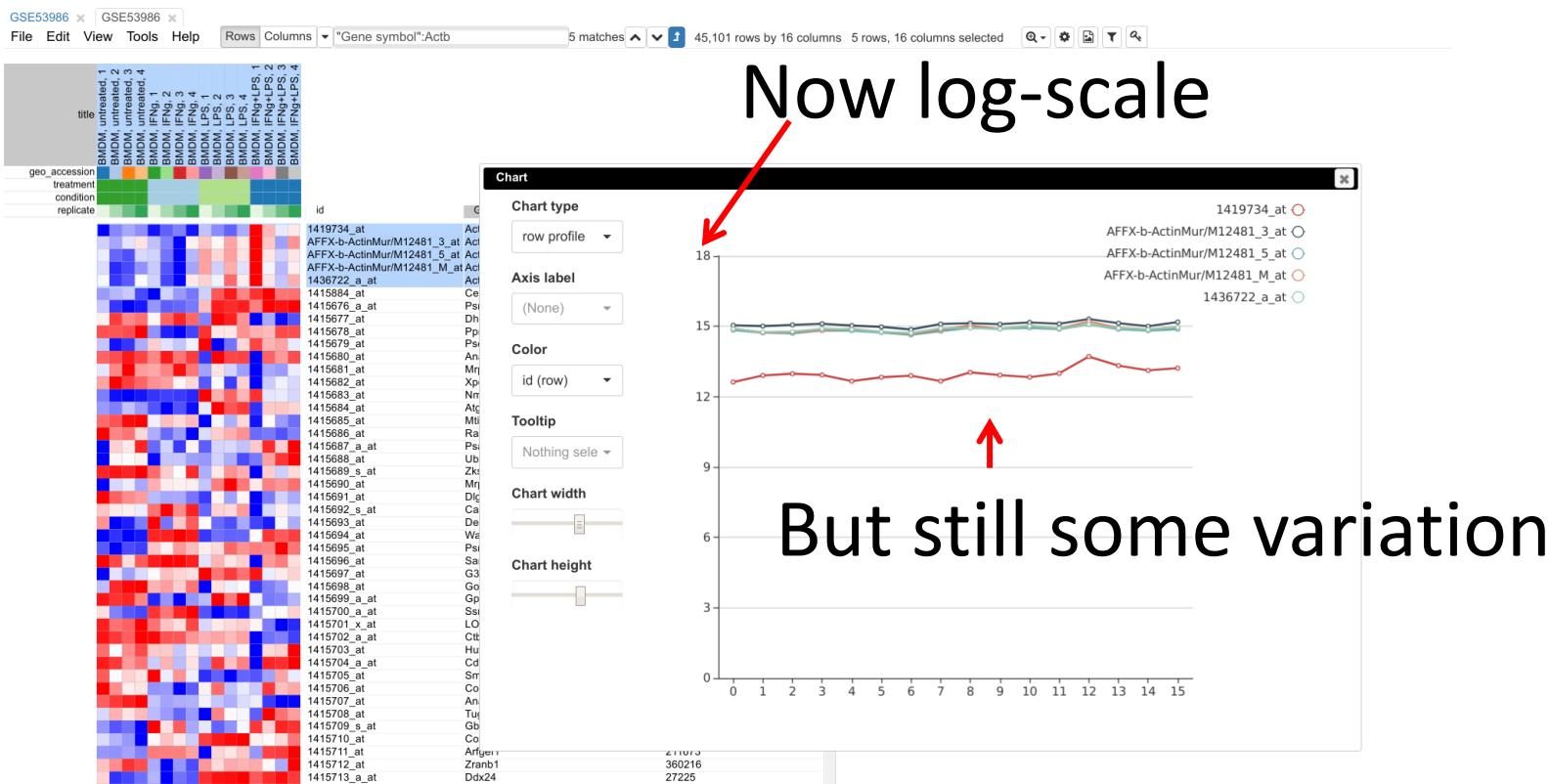
2. Click “Matches to top”

Actb expression chart: high variation (but in a linear scale)



Log 2 normalization

- ✓ Close the chart window
- ✓ Tools/Adjust, check
“Log 2”
- ✓ Redo the plot

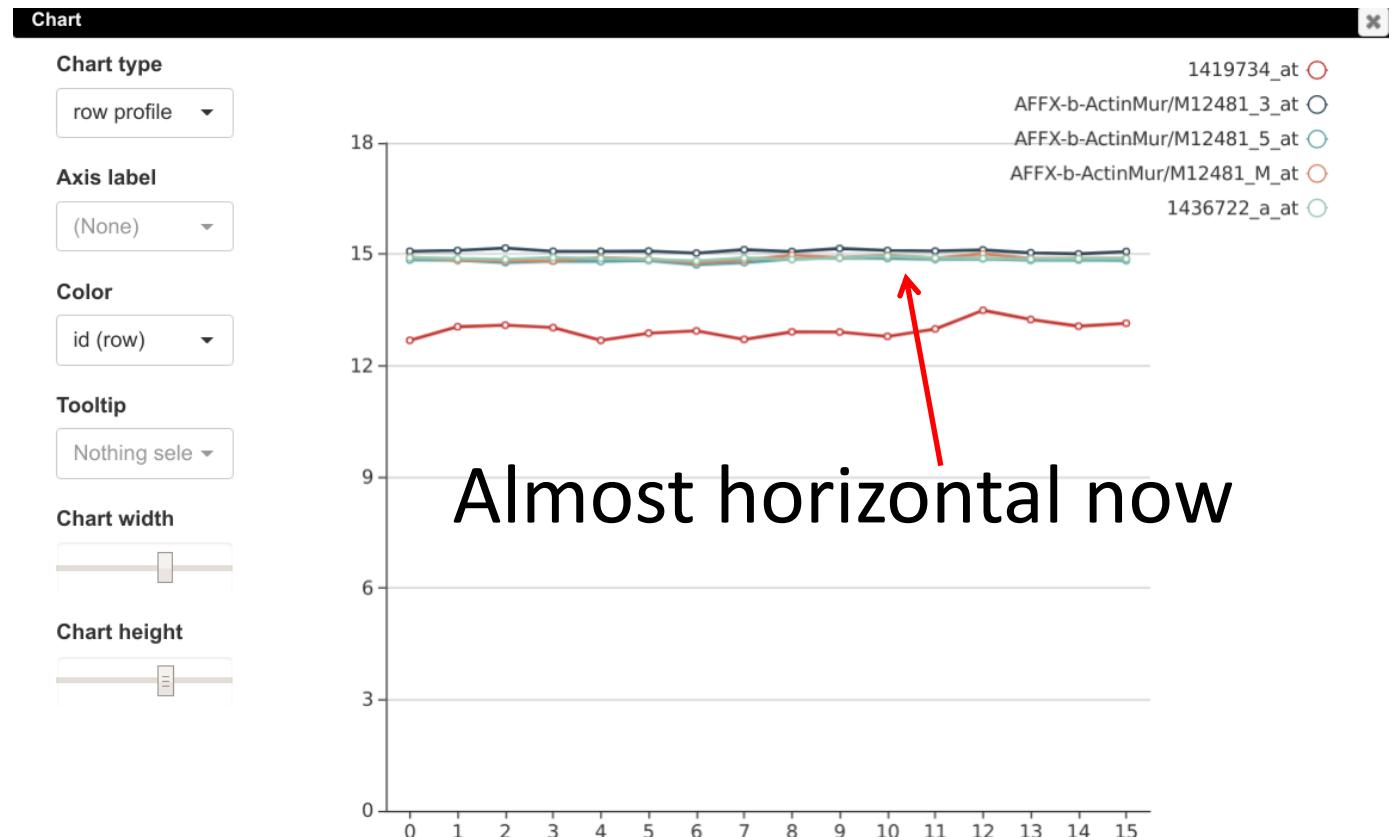


But still some variation

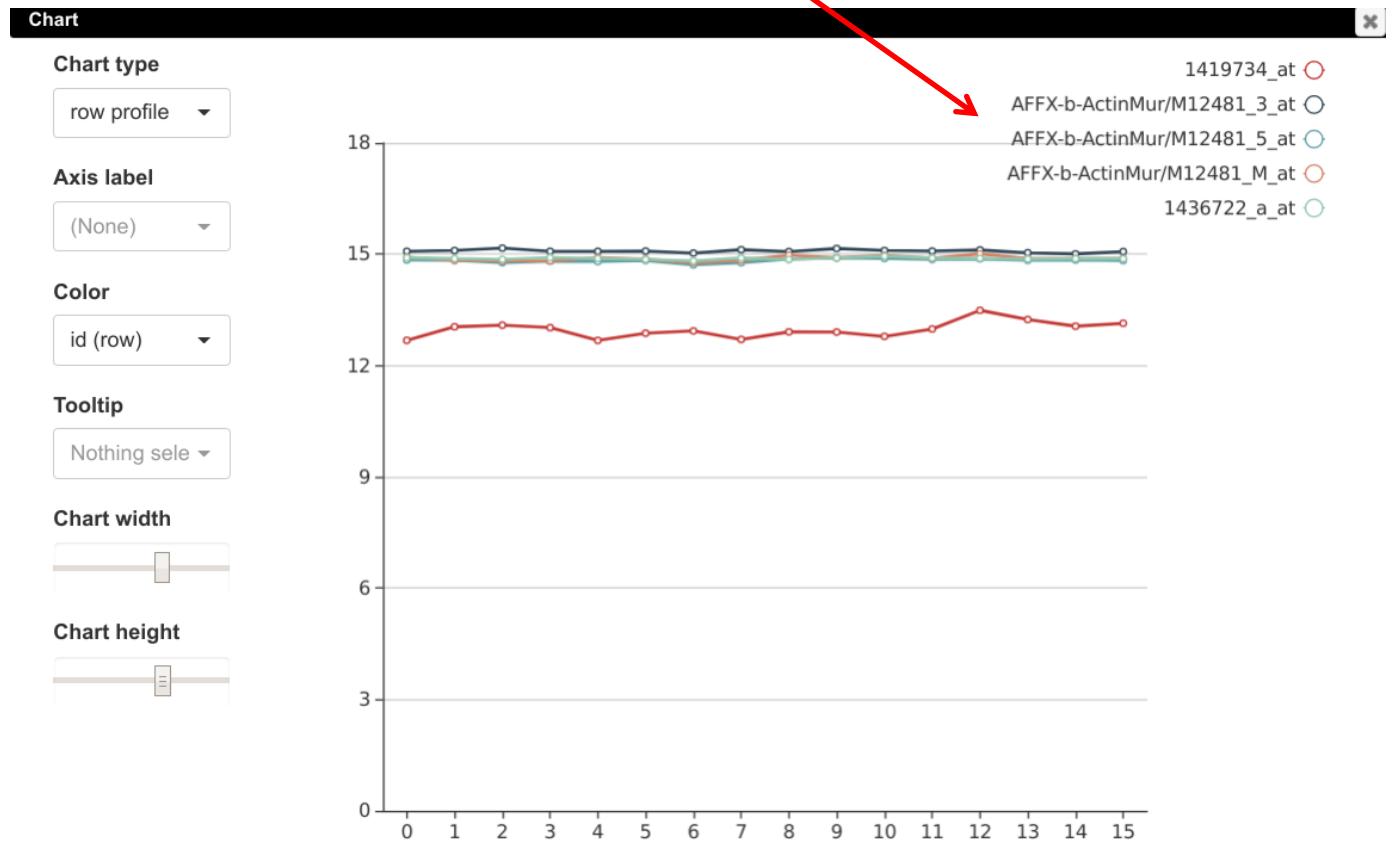
Quantile normalization

- ✓ Close the chart window
- ✓ Tools/Adjust, check “quantile”
- ✓ Redo the plot

- ✓ Log2 and quantile can be done in one step
- ✓ Don’t do Log2 twice, twice quantile is OK

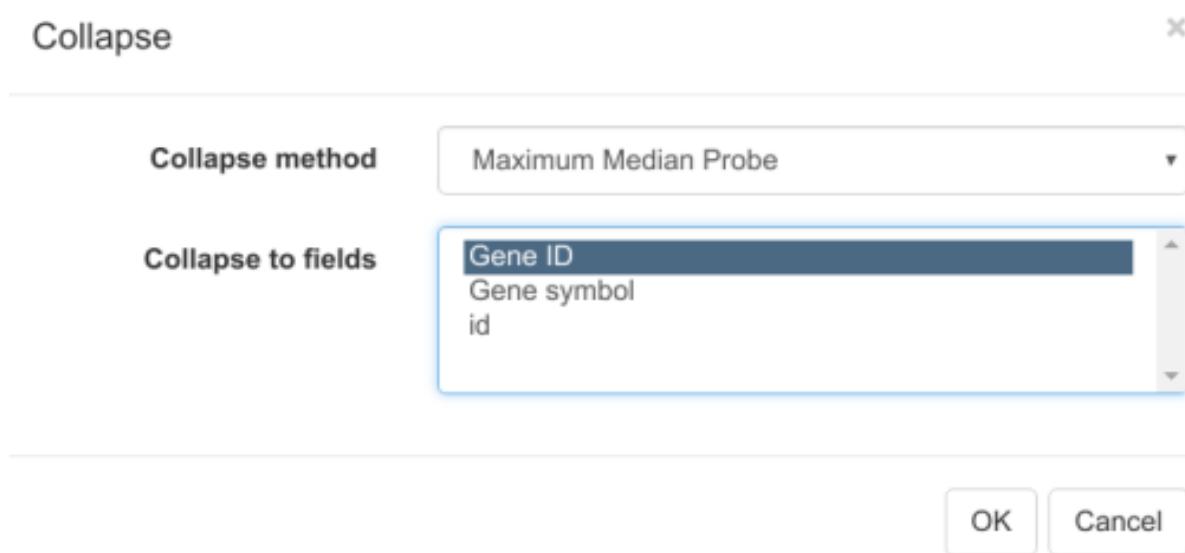


There are multiple probes per gene in microarrays



Collapsing duplicated probes to genes: keeping only one probe per gene

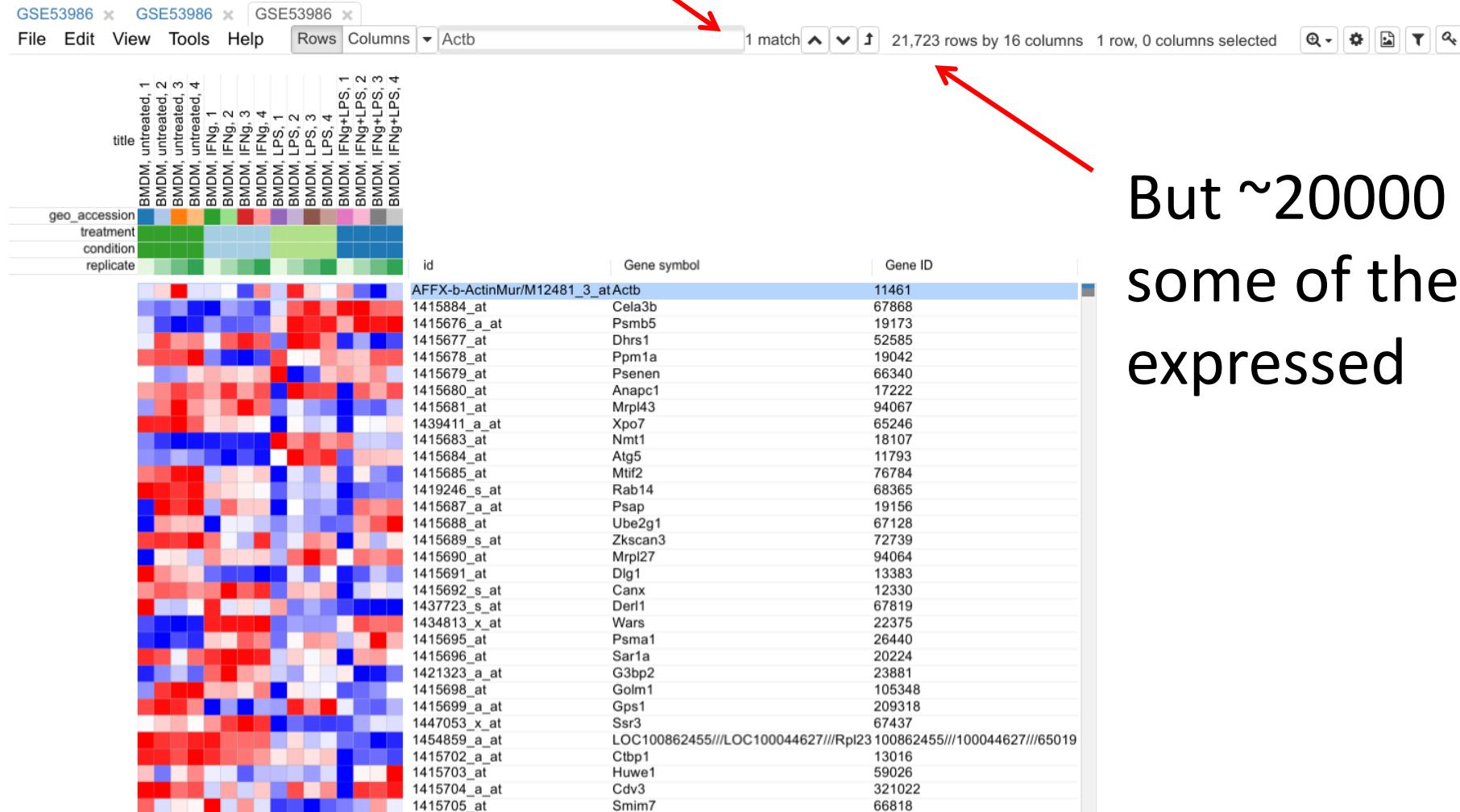
- ✓ Tools/Collapse



method = maximum
median probe

Grouping by
Gene ID

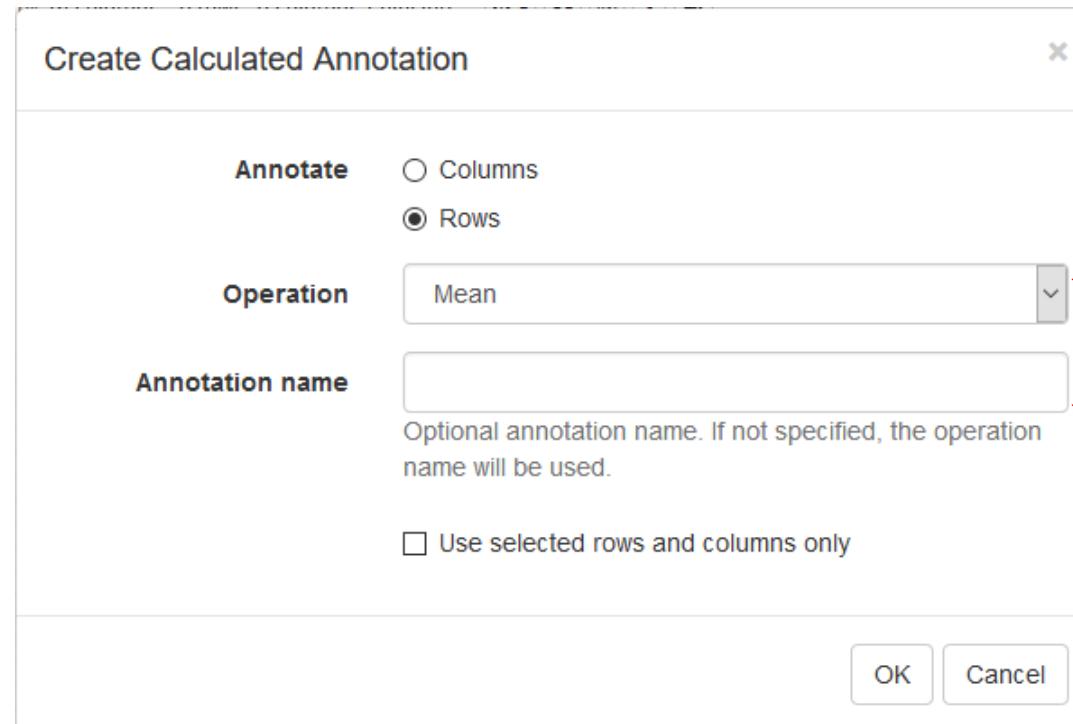
No more duplicates



But ~20000 genes,
some of them are not
expressed

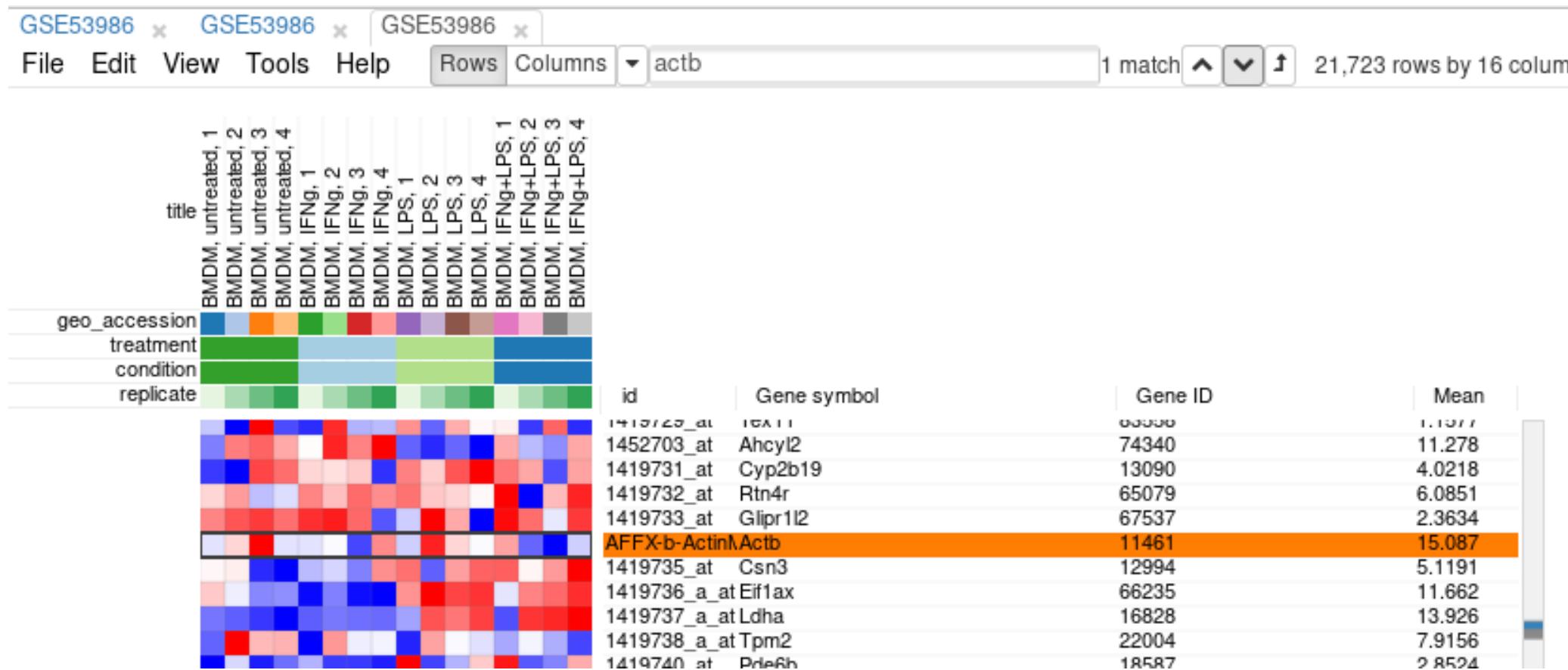
Filtering lowly expressed genes: calculating mean expression

- ✓ Tools/Create Calculated Annotation



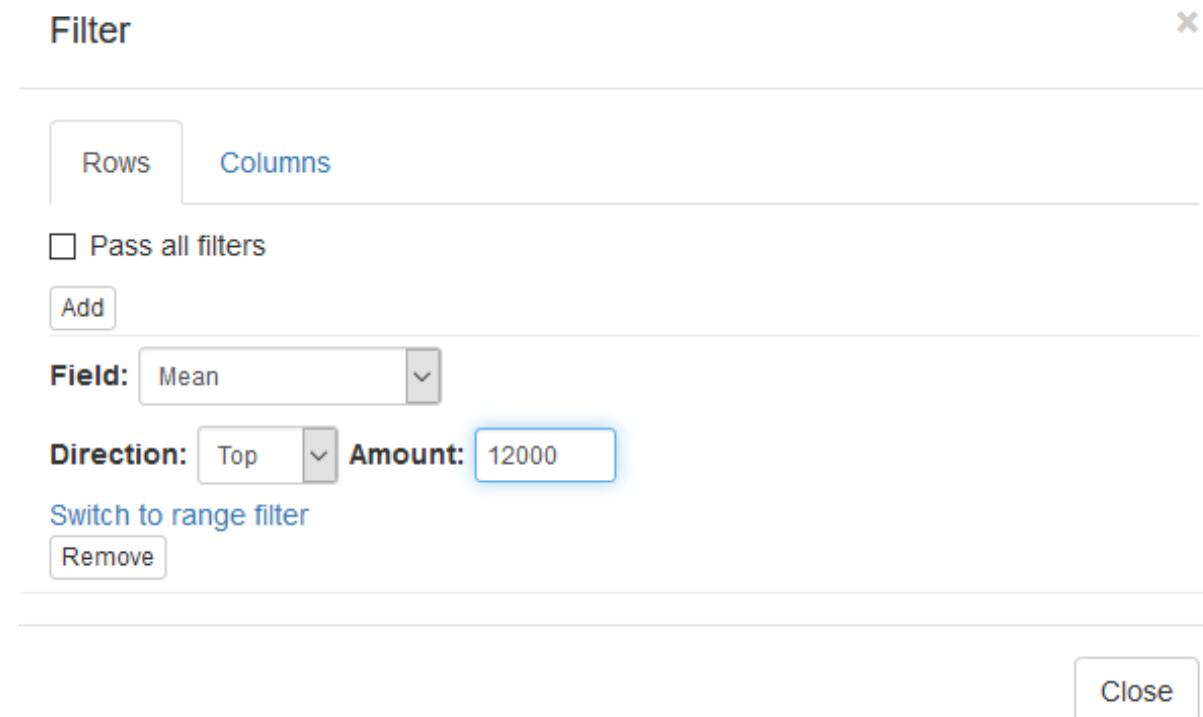
Operation: Mean
Optional name
(e.g. “mean_expression”)

Filtering lowly expressed genes: calculating mean expression result



Filtering lowly expressed genes: keeping only top 12000 genes

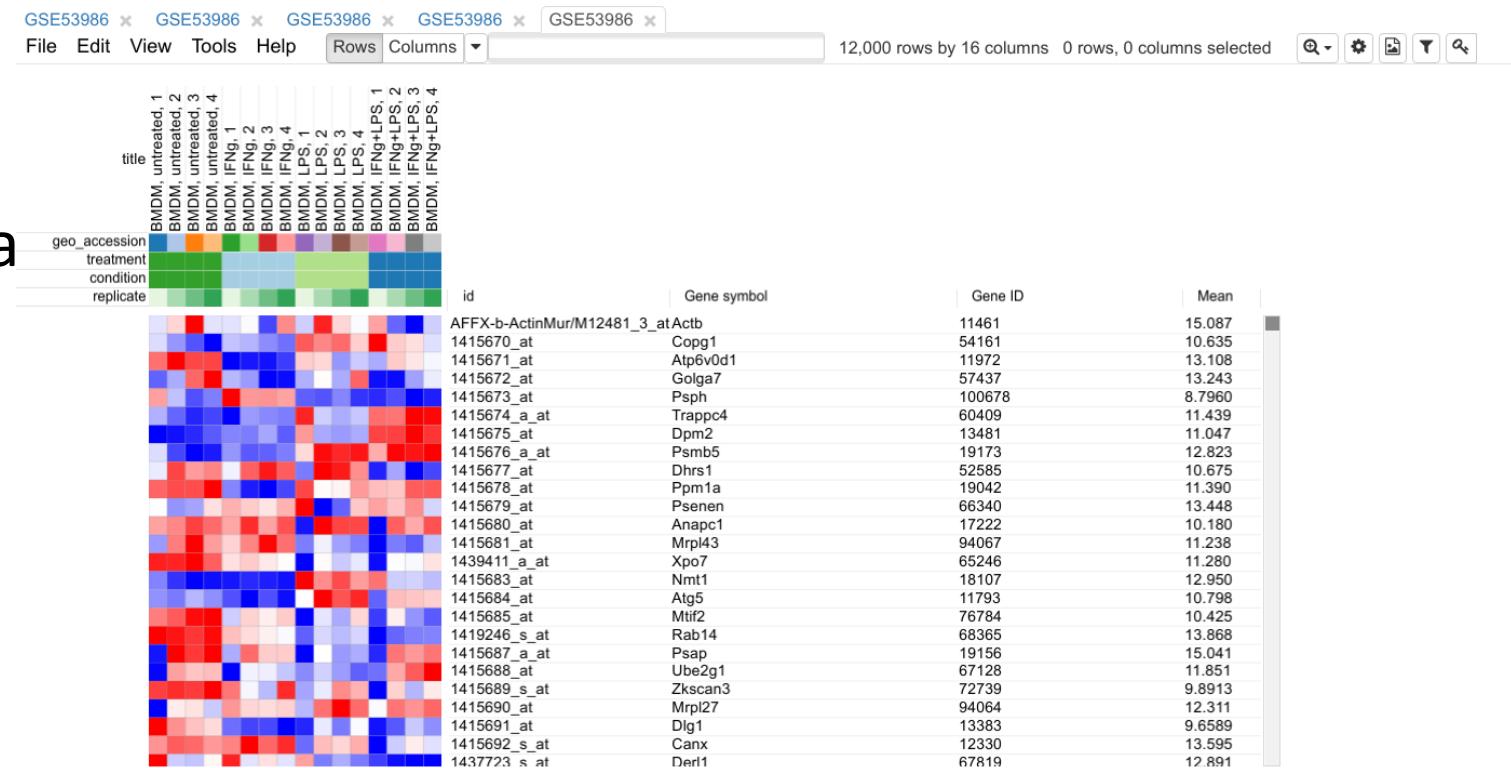
- ✓ Tools/Filter
- ✓ Add
- ✓ Field <- Mean
- ✓ Switch to top filter
- ✓ N <- 12000



File Edit View Tools Help Rows Columns ▾ actb 1 match ↕ ↖ ↘ ↙ 12,000/21,723 rows by 16 columns

Filtering lowly expressed genes: creating new dataset

- ✓ Select all genes (click on any gene and Ctrl+A)
- ✓ Hit Ctrl-X to create new dataset (or Tools/New Heat Map)

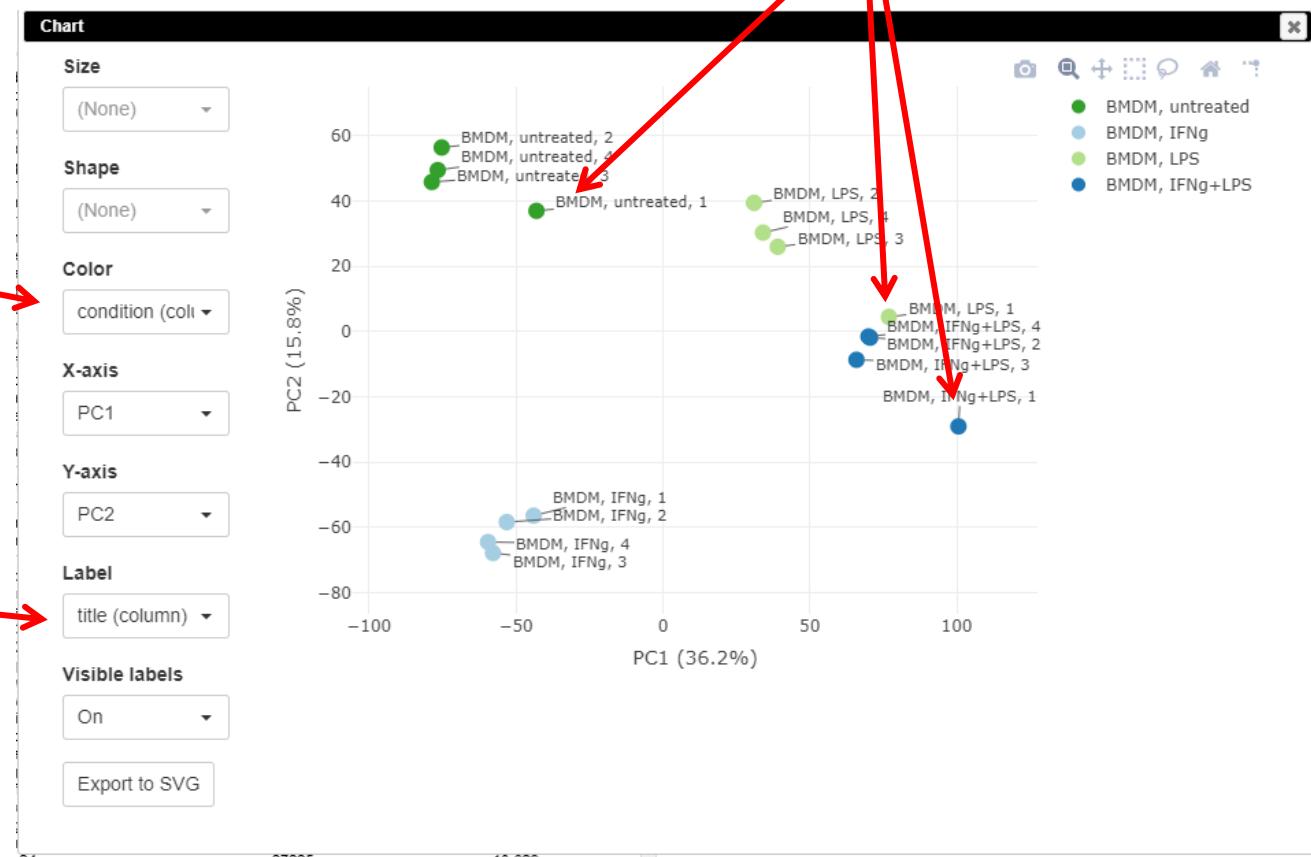


PCA

✓ Tools/Plots/PCA plot

color <-
treatment

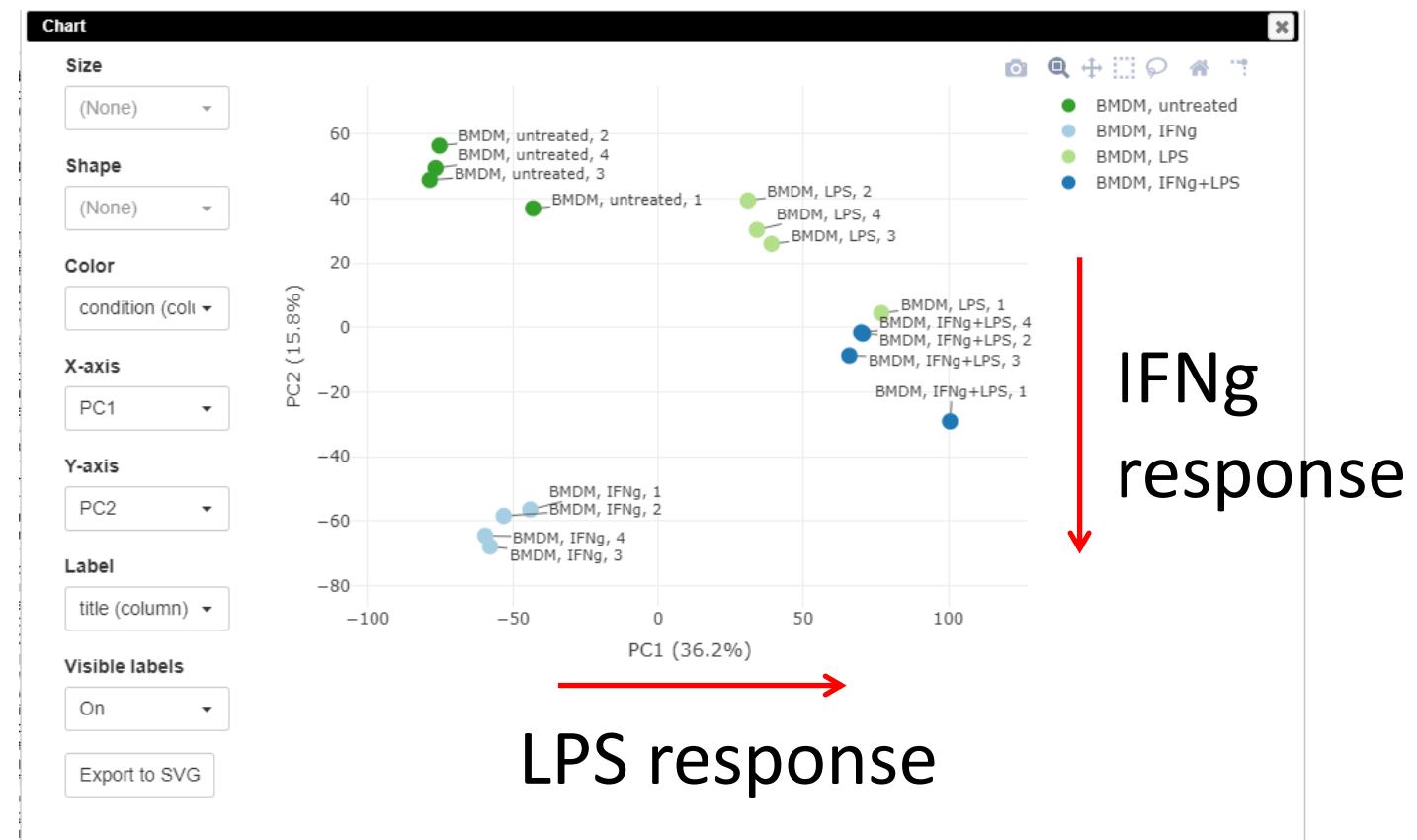
label <- title



Scale should be ~10-100, not 1000000

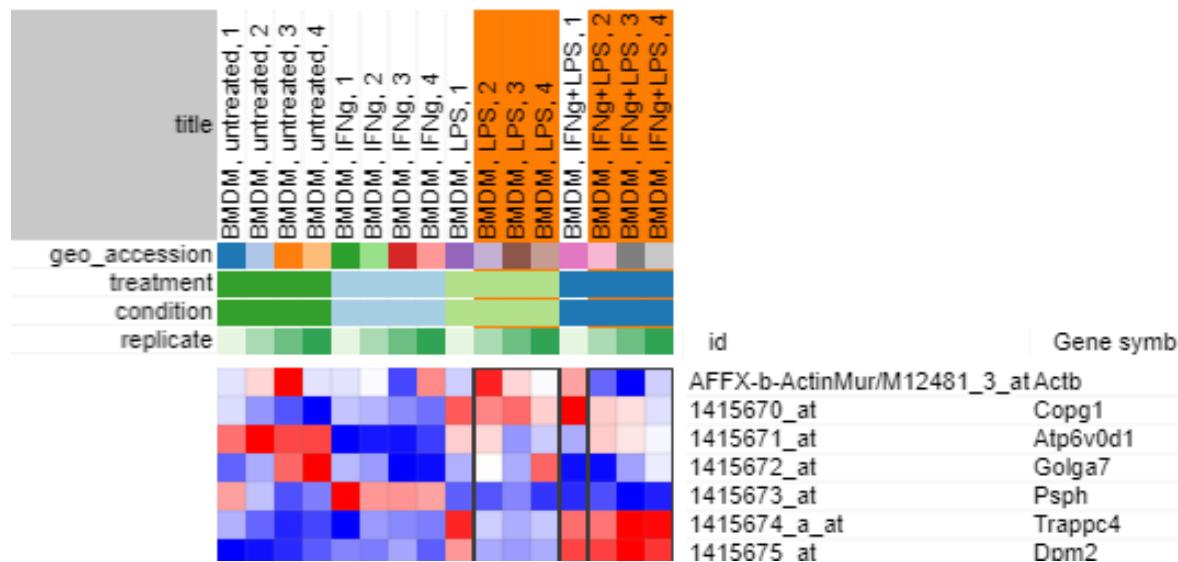
PCA

- ✓ In some cases PCA plots might be very meaningful



DE

- ✓ What's different between LPS vs LPS + IFNg?
- ✓ Select all interesting samples without an outlier and create new dataset (ctrl + x)



Tools > Diff expression > limma

Comparison	id	Gene symbol	Gene ID	mean_expression	logFC	AveExpr	t	P.Value	adj.P.Val	B
1435477_s_at	Fcgr2b	14130	13.352	2.0677	13.482	34.375	7.3114e-8	0.00015972	9.0015	
1417936_at	Ccl9	20308	12.943	2.8340	13.064	31.182	1.2756e-7	0.00015972	8.5498	
1448475_at	Olfml3	99543	8.0623	3.5999	7.5726	28.070	2.3231e-7	0.00015972	8.0343	
1440666_at	4933430I17Rik	214106	6.2900	2.4749	7.8986	27.434	2.6468e-7	0.00015972	7.9185	
1419561_at	Ccl3	20302	11.941	2.0743	12.223	27.059	2.8629e-7	0.00015972	7.8481	
1418069_at	Apoc2	11813	10.565	2.3018	10.210	27.051	2.8675e-7	0.00015972	7.8467	
1419132_at	Tlr2	24088	11.646	2.4945	12.043	25.908	3.6663e-7	0.00015972	7.6236	
1442434_at	D8Ertd82e	244418	7.8873	2.3431	7.5717	25.666	3.8676e-7	0.00015972	7.5745	
1422010_at	Tlr7	170743	11.394	2.3556	10.962	25.246	4.2489e-7	0.00015972	7.4877	
1438306_at	Rnf180	71816	9.0500	3.3620	7.5052	24.640	4.8782e-7	0.00017616	7.3591	
1417926_at	Ncapg2	76044	11.434	2.4804	10.238	24.105	5.5272e-7	0.00018304	7.2417	
1434572_at	Hdac9	79221	8.1260	2.5108	7.4893	23.766	5.9903e-7	0.00018304	7.1656	
1416882_at	Rgs10	67865	11.888	1.5579	10.775	23.630	6.1884e-7	0.00018304	7.1347	
1428484_at	Osbpl3	71720	9.9793	1.6935	11.452	23.095	7.0493e-7	0.00018389	7.0103	
1417611_at	Tmem37	170706	8.8936	2.6647	8.0207	22.749	7.6797e-7	0.00019191	6.9279	
1418172_at	Hebp1	15199	9.1728	2.5260	8.9768	22.668	7.8363e-7	0.00019191	6.9085	
1420819_at	Sla	20491	10.572	3.1577	10.078	22.372	8.4449e-7	0.00019885	6.8361	

Tools > pathway analysis > fgsea

Perform FGSEA ×

Pathway database

GO Biological Processes - Mus Musculus (Entrez) ▾

Rank by

t ▾

Column with gene ID

Gene ID ▾

Omit ambiguous genes

Submit Cancel

Tools > pathway analysis > fgsea

FGSEA:

pathway	pval	padj	log2err	ES	NES	size	leadingEdge
Translation	2.86e-10	3.01e-7	0.814	0.500	2.08	220	19896 19935 20084 19988 27207 ...
RRNA processing	0.00000710	0.00374	0.611	0.493	1.89	122	57444 59028 19942 20115 20085 ...
Antigen processing and presentation	0.0000397	0.0139	0.557	-0.732	-2.13	27	14998 14969 16149 14960 14961 ...
Defense response to protozoan	0.000122	0.0214	0.538	-0.724	-2.10	26	74481 15900 21939 14468 60440 ...
Positive regulation of transcription, DNA-templated	0.000119	0.0214	0.538	-0.340	-1.53	407	16600 16149 16362 327987 15900 ...
Phosphatidylinositol 3-kinase signaling	0.000105	0.0214	0.538	0.712	1.98	23	15559 104709 320207 16000 12192 ...
Cellular response to interferon-beta	0.000231	0.0346	0.519	-0.661	-1.99	32	16362 14468 620913 60440 240327 ...
Release of cytochrome c from mitochondria	0.000401	0.0404	0.498	-0.759	-2.02	18	12122 110175 11977 58801 12018 ...
Ubiquitin-dependent protein catabolic process	0.000461	0.0404	0.498	-0.389	-1.61	212	24108 268291 71745 242960 75234 ...

Phantasus

- ✓ Quick/brief analysis
- ✓ Only simple differential expression designs are supported
- ✓ Heatmaps can be saved as publication-ready SVGs

Personal experience

- ✓ RNA-seq data quantification (done on cluster)
- ✓ RNA-seq analysis (DESeq2 and fgsea done in R)
- ✓ rlog matrix from DESeq2 goes into Phantasus
- ✓ Most of the results I validate in Phantasus by eye
- ✓ I usually do heatmaps in Phantasus

Visualizing scRNA-seq data

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

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:

All scRNA-seq datasets	Curated datasets	Gene signature search		
Name	Description	Organism	# of cells	Exte...
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/SRS53555828	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 <i>T. gondii</i> 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 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:

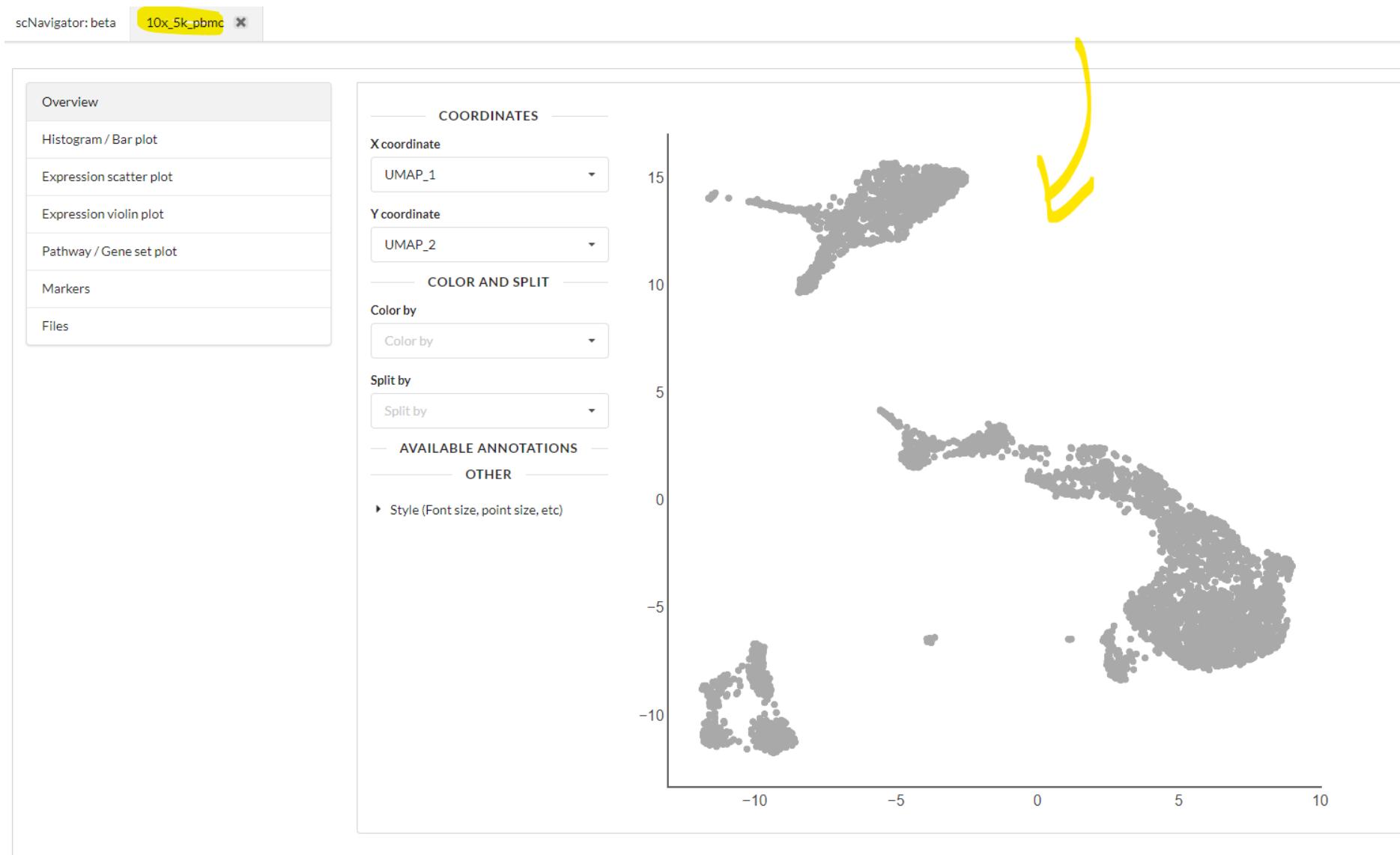
- Large collection of automatically processed datasets. We processed almost every scRNA-seq dataset from GEO Omnibus database. We make it available for you in our browser.
- Collection of curated datasets. Curated dataset are those that we process by hand. These will include datasets from Human Cell 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

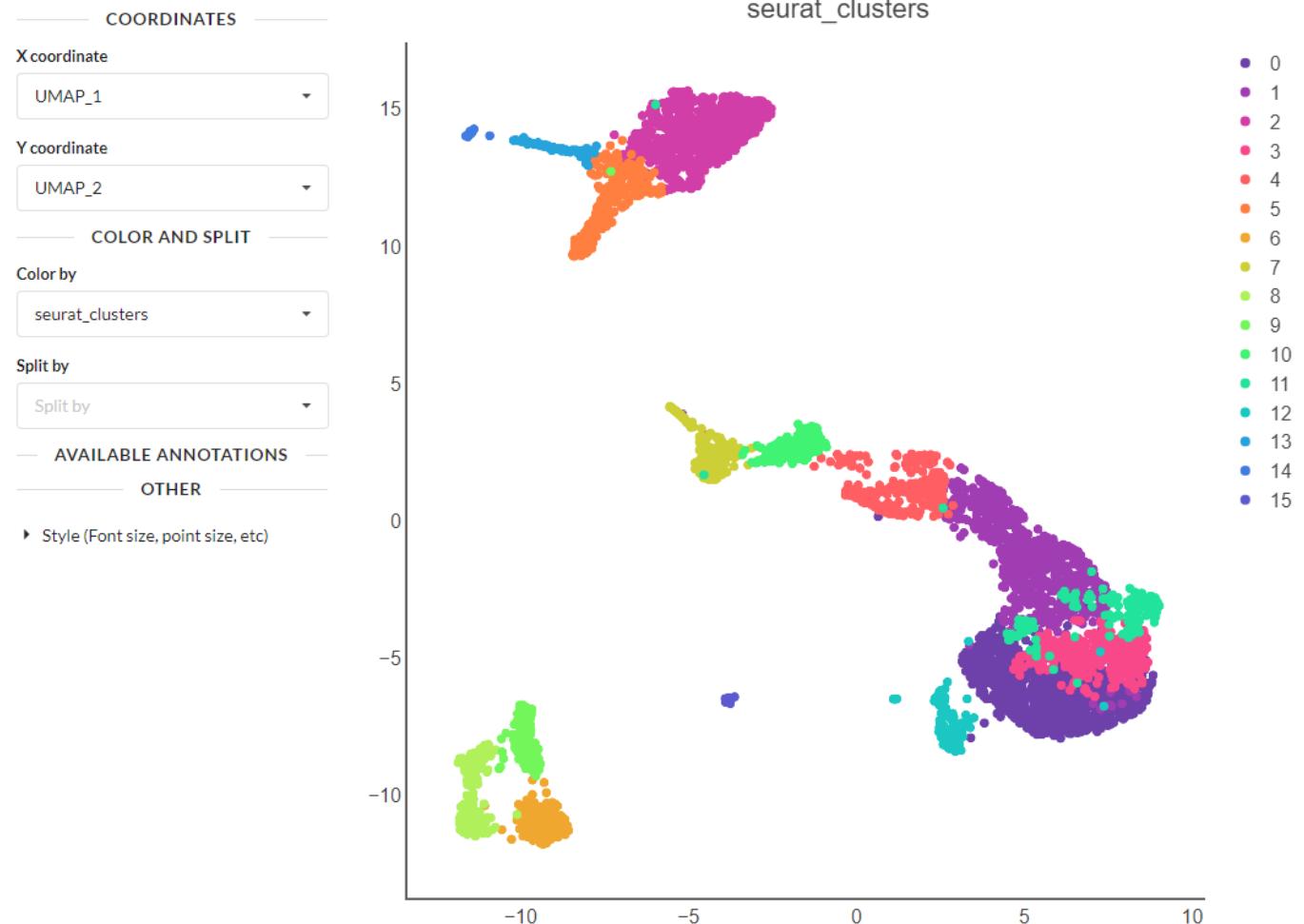
Result should look like that



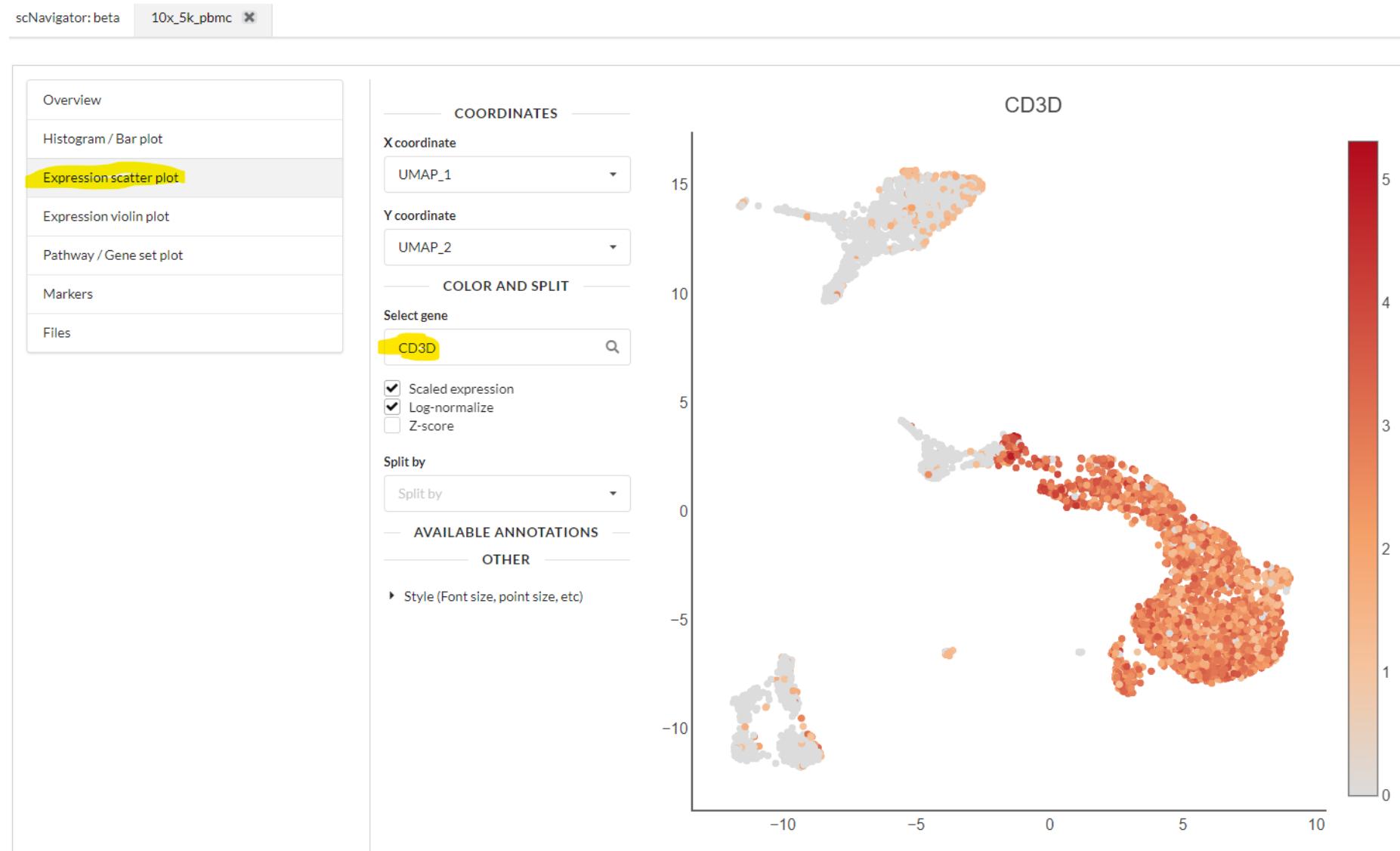
We can color the cells

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

- ✓ tsne_Cluster_centers



Expression of CD3d



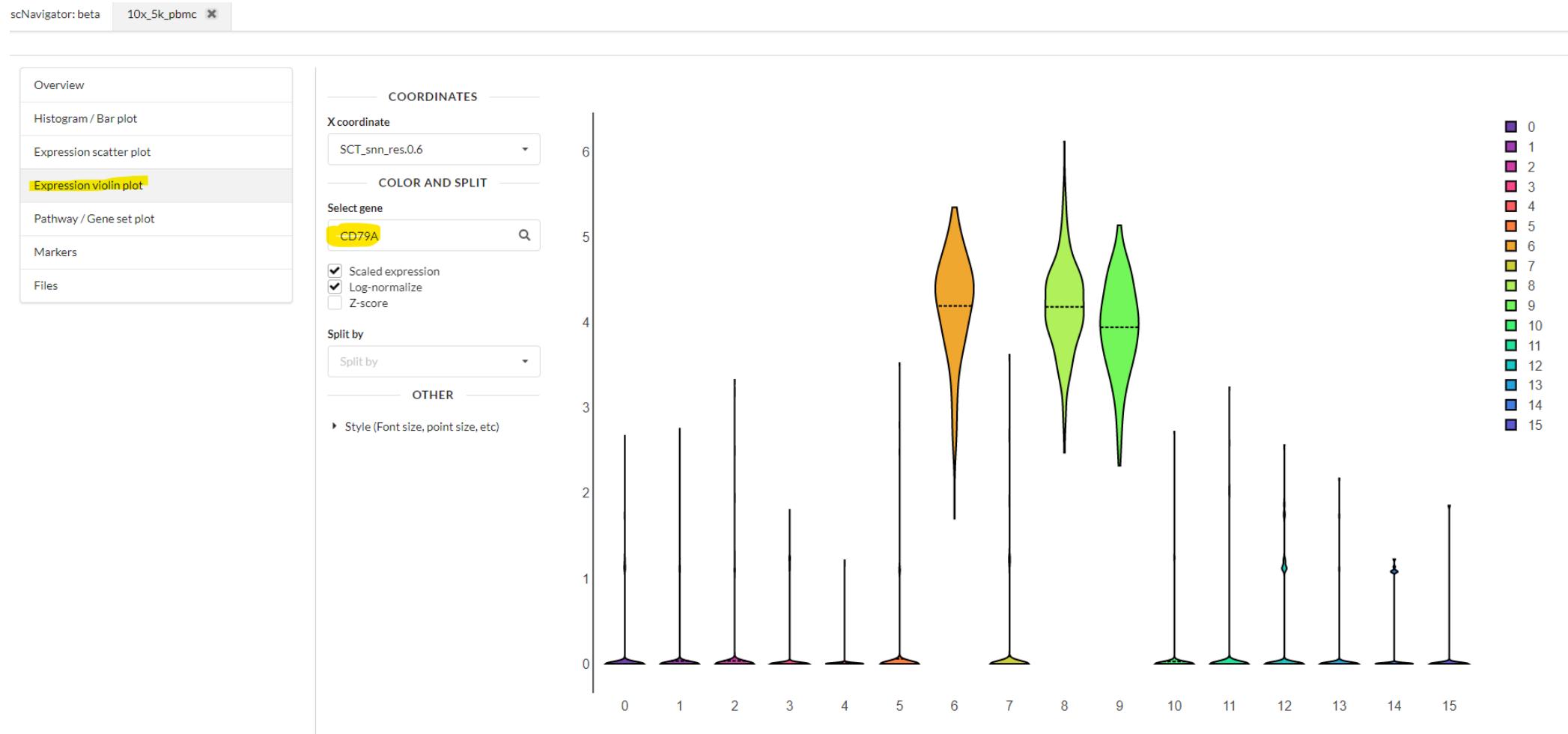
Or you can go for any of your favorite genes



Expression scatter plot

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

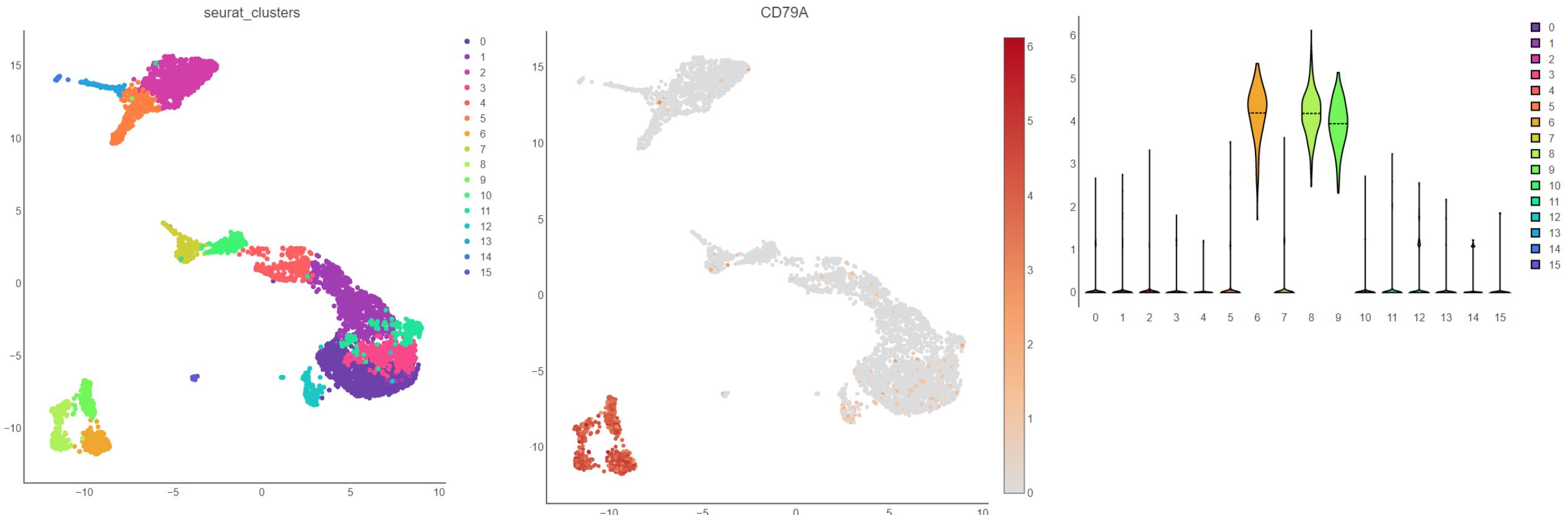
Violin plot



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

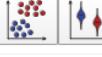
Markers tab

scNavigator: beta 10x_5k_pbmc X

- 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
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 X

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	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
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 X

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	Exte...
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 <i>T. gondii</i> 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:

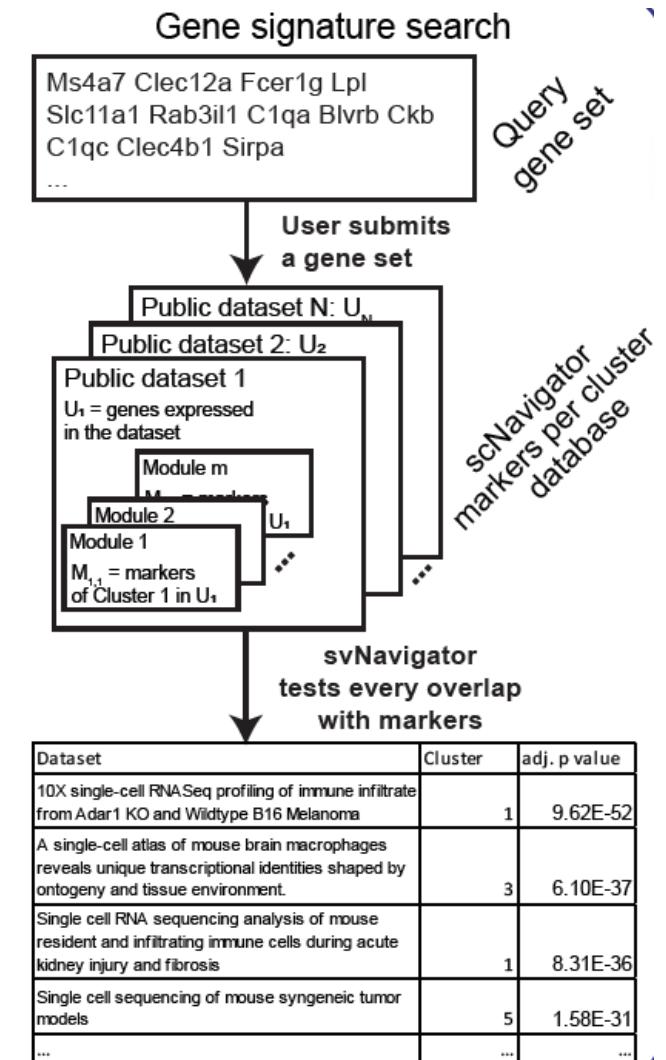
- ✓ Different technologies used to perform experiment (10x, DropSeq, SmartSeq2, C1 Fluidigm etc)
- ✓ Different pipelines were used to analyze
- ✓ Different formats in which data is kept

Most of the dataset processing was done
by Maria Firuleva



Gene Signature search

- ✓ User queries a gene set
- ✓ We then look for a cluster with similar markers
- ✓ Return significant hits with adjusted p.values



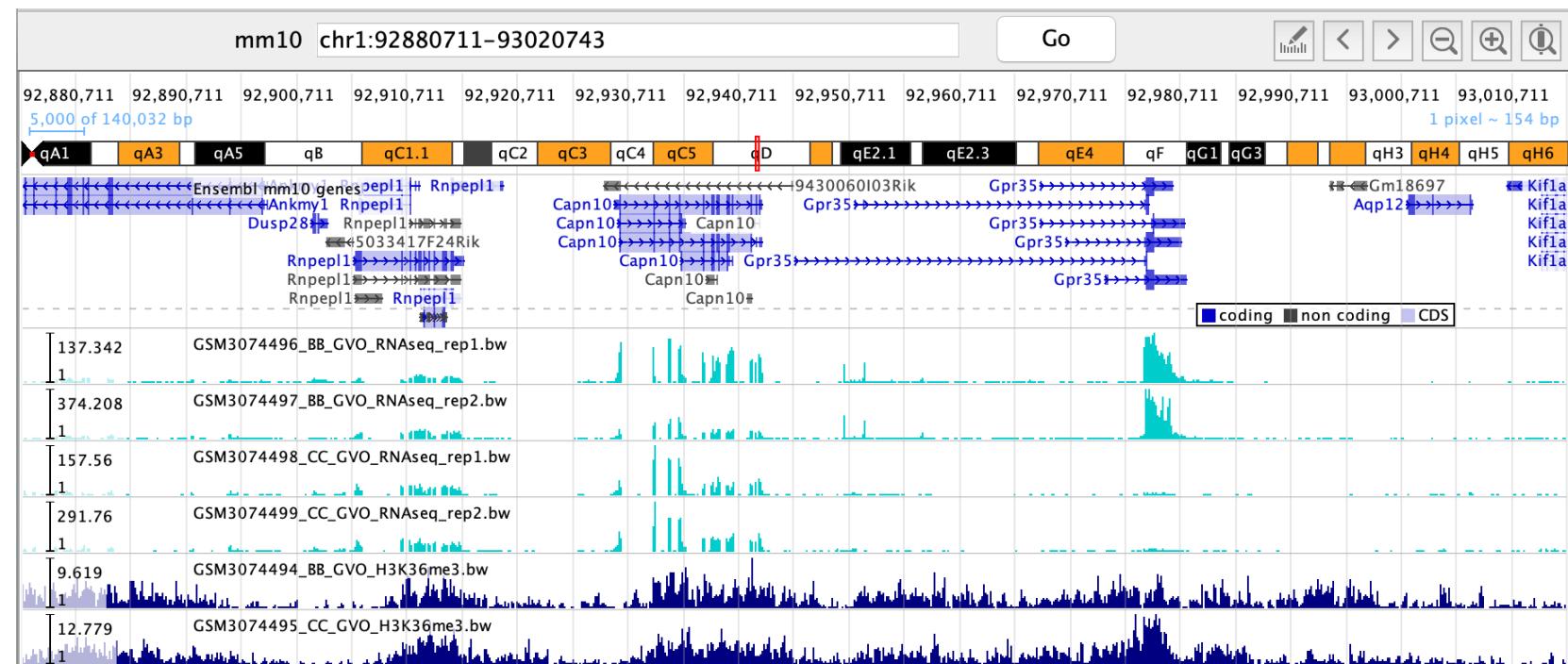
Genome browsers

- ✓ Genome browsers are amazing when you try to figure out how did your RNA-seq experiment go
- ✓ Gene coverage
- ✓ Visualizing differential splicing

JBR Genome Browser

Desktop Browser

- ✓ Supports semi-supervised peak calling
 - ✓ Optimized for large sessions (lot's of opened tracks)



Prerequisites: Download JBR

<https://research.jetbrains.org/groups/biolabs/tools/jbr-genome-browser>

Downloads

JBR Genome Browser (build 1.0.beta.5244), released on Aug 12, 2020

Windows
Mac and Linux

Download	Description
jbr-1.0.beta.5244.zip	Windows archive (includes bundled 64-bit Java Runtime)
jbr-1.0.beta.5244.dmg	Mac installer (includes bundled 64-bit Java Runtime)
jbr-1.0.beta.5244.tar.gz	Linux archive (includes bundled 64-bit Java Runtime)

Install JBR

<https://research.jetbrains.org/groups/biolabs/tools/jbr-genome-browser>

Installation

Download suitable build for your OS from [Downloads](#) section.

Windows:

- Unpack the browser 1.0.beta.5244.zip file:
- Run `jbr.exe`.

MacOS:

- Download the 1.0.beta.5244.dmg macOS Disk Image file
- Mount it as another disk in your system
- Copy JBR Genome Browser to your Applications folder

If you wan't to open multiple JBR instances launch second, third, etc. instances using command:

```
open -n "/Applications/JBR 1.0.app"
```

Linux:

- Unpack the browser 1.0.beta.5244.tar.gz file using the following command:
`tar -xzf 1.0.beta.5244.tar.gz`
- Run `jbr.sh` from the bin subfolder.

Launch JBR



Genome Browser

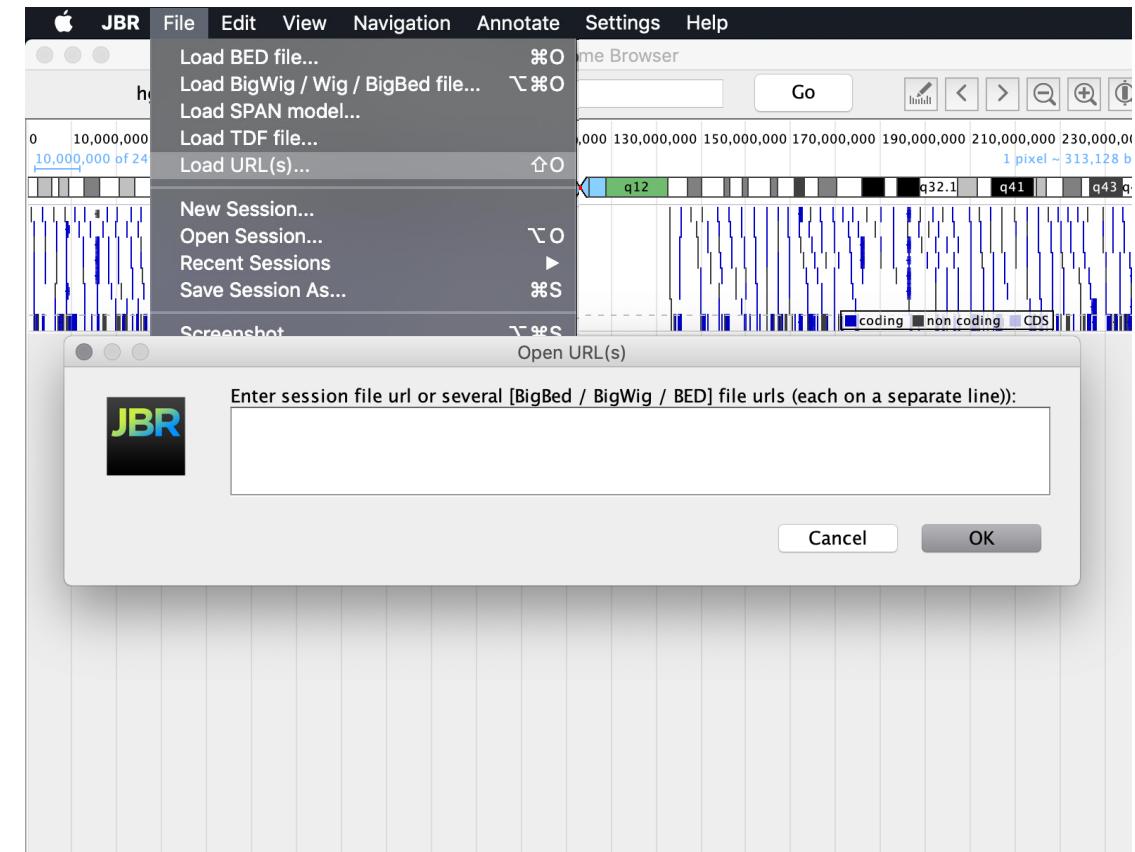
[May 30, 2018 22:55:14] Loading genes /Users/romeo/.jbr_browser/genomes/hg19/Homo_sapiens.GRCh37.87.gtf.gz: done in 802.5 ms

JBR Is Ready To Use



JBR: Supported Files

- BED-like peaks files:
 - *.bed, *.bigBed, *.bb
 - MACS2, SICER, SPAN peak files
- Signal Profile: *.bigWig, *.bw, *.tdf
- SPAN model file *.span
- Sessions file:
 - IGV session *.xml file
 - JBR *.yaml file



Let's Open JBR demo session to explore epigenetics data

Recent nature aging paper

- ✓ Describing healthy aging of human monocytes



Enhanced epigenetic profiling of classical human monocytes reveals a specific signature of healthy aging in the DNA methylome

Irina Shchukina^{1,13}, Juhi Bagaitkar^{2,13}, Oleg Shpynov^{1,3,13}, Ekaterina Loginicheva¹, Sofia Porter¹, Denis A. Mogilenko¹, Erica Wolin⁴, Patrick Collins¹, German Demidov^{1,5,6}, Mykyta Artomov^{1,7,8}, Konstantin Zaitsev^{1,12}, Sviatoslav Sidorov⁹, Christina Camell⁹, Monika Bambouskova¹, Laura Arthur¹, Amanda Swain¹, Alexandra Panteleeva¹, Aleksei Dievskii³, Evgeny Kurbatsky³, Petr Tsurinov^{1,3}, Roman Chernyatchik^{1,3}, Vishwa Deep Dixit⁹, Marko Jovanovic⁴, Sheila A. Stewart^{1,10}, Mark J. Daly^{7,8,11}, Sergey Dmitriev³, Eugene M. Oltz¹ and Maxim N. Artyomov¹

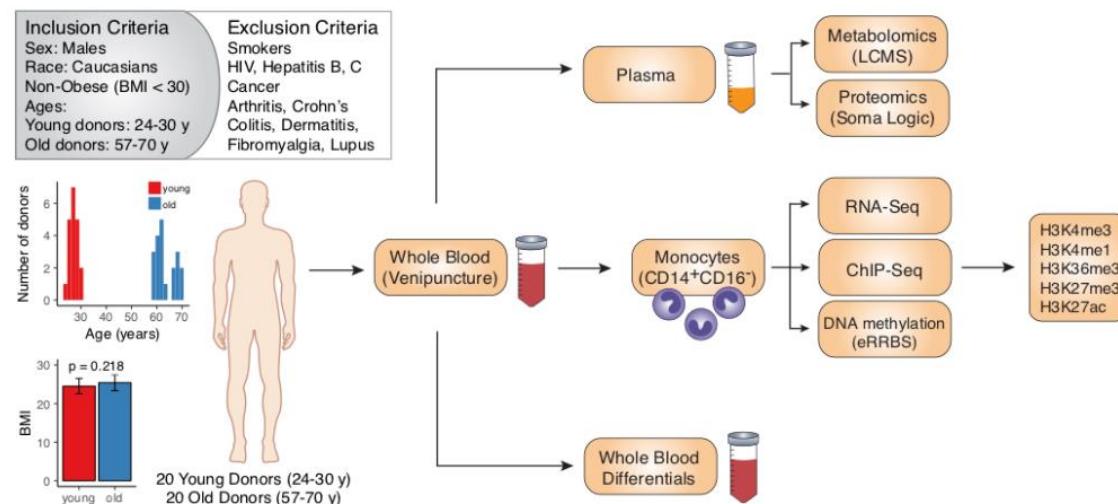
The impact of healthy aging on molecular programming of immune cells is poorly understood. Here we report comprehensive characterization of healthy aging in human classical monocytes, with a focus on epigenomic, transcriptomic and proteomic alterations, as well as the corresponding proteomic and metabolomic data for plasma, using healthy cohorts of 20 young and 20 older males (~27 and ~64 years old on average). For each individual, we performed enhanced reduced representation bisulfite sequencing-based DNA methylation profiling, which allowed us to identify a set of age-associated differentially methylated regions (DMRs)—a novel, cell-type-specific signature of aging in the DNA methylome. Hypermethylation events were associated with H3K27me3 in the CpG Islands near promoters of lowly expressed genes, while hypomethylated DMRs were enriched in H3K4me1-marked regions and associated with age-related increase of expression of the corresponding genes, providing a link between DNA methylation and age-associated transcriptional changes in primary human cells.

Website with all the data

- ✓ <https://artyomovlab.wustl.edu/aging/index.html>

Experiment

The impact of healthy aging on molecular programming of immune cells is poorly understood. Here, we report comprehensive characterization of healthy aging in human monocytes, with a focus on transcriptomic and epigenomic alterations, as well as the corresponding proteomic and metabolomic data for plasma, using healthy cohorts of 20 young and 20 older individuals (~27 and ~64 years old on average). For each individual, we have performed RRBS-based DNA methylation profiling, which revealed a set of age-related differentially methylated regions (DMRs) that either gain or lose methylation with age. Optimized ultra-low-input ChIP-seq (ULI-ChIP-seq) data acquisition and analysis pipelines allowed us to link hypo- and hypermethylated DMRs to distinct chromatin modification patterns. Specifically, hypermethylation events were associated with H3K27me3 marked CpG islands, while hypomethylated DMRs were enriched in H3K4me1 marked regions. Further integrative analysis of the transcriptomic data established a connection between identified age-associated alterations in the DNA methylome and corresponding age-associated transcriptional changes.



Website with all the data

- ✓ https://artyomovlab.wustl.edu/aging/explore_data.html

Sessions

For those who prefer desktop genome browser applications, we prepared dedicated session files for the following genome browsers:

[IGV Genome Browser](#) and [JBR Genome Browser](#).

Click on the corresponding link to download the application.

Preconfigured session file for [IGV](#) or [JBR](#) genome browsers: [DNA methylation session](#).

ChIP-Seq

- [View](#) data in the JBR and UCSC genome browsers.
- Download preconfigured [session files](#).
- Download [data](#) used in sessions.

View

Explore preconfigured online session in web [JBR Genome Browser](#) in one click.

Modification	Session
H3K27ac	2018_h3k27ac_aging
H3K27me3	2018_h3k27me3_aging
H3K36me3	2018_h3k36me3_aging
H3K4me1	2018_h3k4me1_aging
H3K4me3	2018_h3k4me3_aging



Tracks, I love tracks

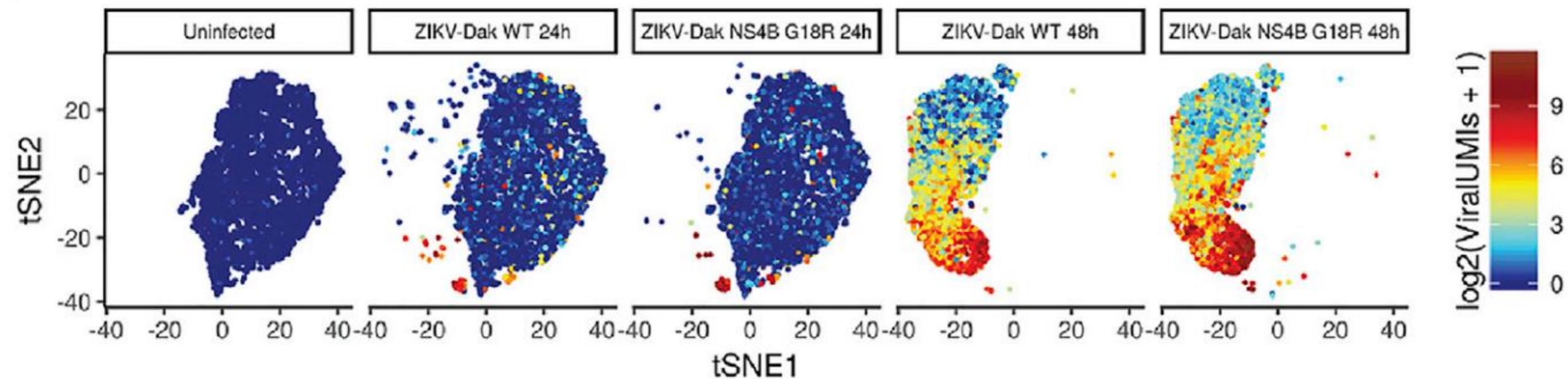


Why would you use genome browser for RNA-seq

- ✓ Checking that experiment works: reads aligning to exons, 3' vs 5' vs full-length sequencing, strand-specific experiments etc
- ✓ Investigating why something didn't work

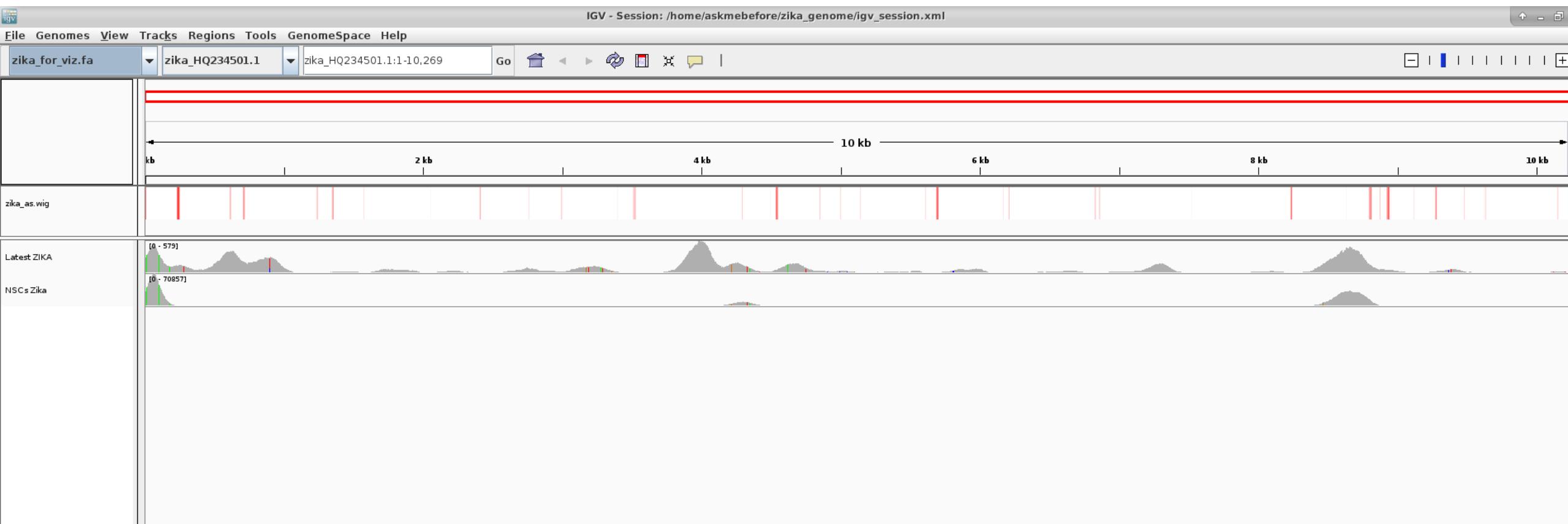
Zika story

- ✓ 5 samples: uninfected aNSCs and infected with two Zika strains at 24 and 48 hours
- ✓ We were able to get Zika reads and count them
- ✓ Zika is not polyadenylated



Zika story

- ✓ We went to IGV to see the coverage



Conclusions

- ✓ Visual data exploration can help to understand some processes
- ✓ Visual data exploration can help to identify possible errors in your analysis