



ITMO UNIVERSITY

CT Lab
ITMO UNIVERSITY

Visual and exploratory analysis in gene expression studies

Konstantin Zaitsev

November 30th, 2019

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 explorer
- ✓ JBR genome browser

Conflict of interest alert

- ✓ Phantasus (ITMO University)
- ✓ Single-cell explorer (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

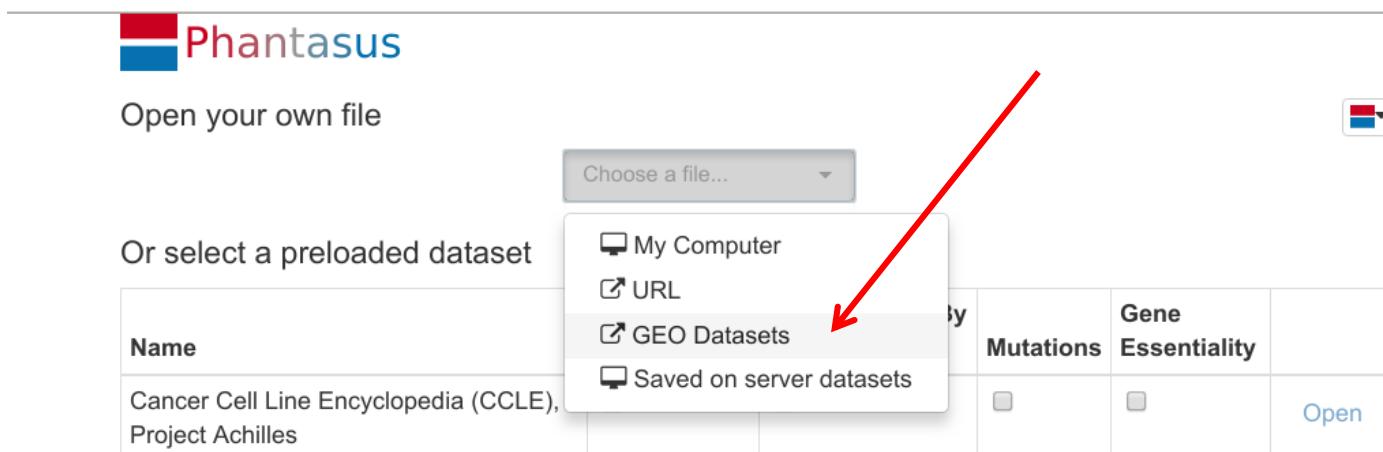
Samples (16)

[☰ 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

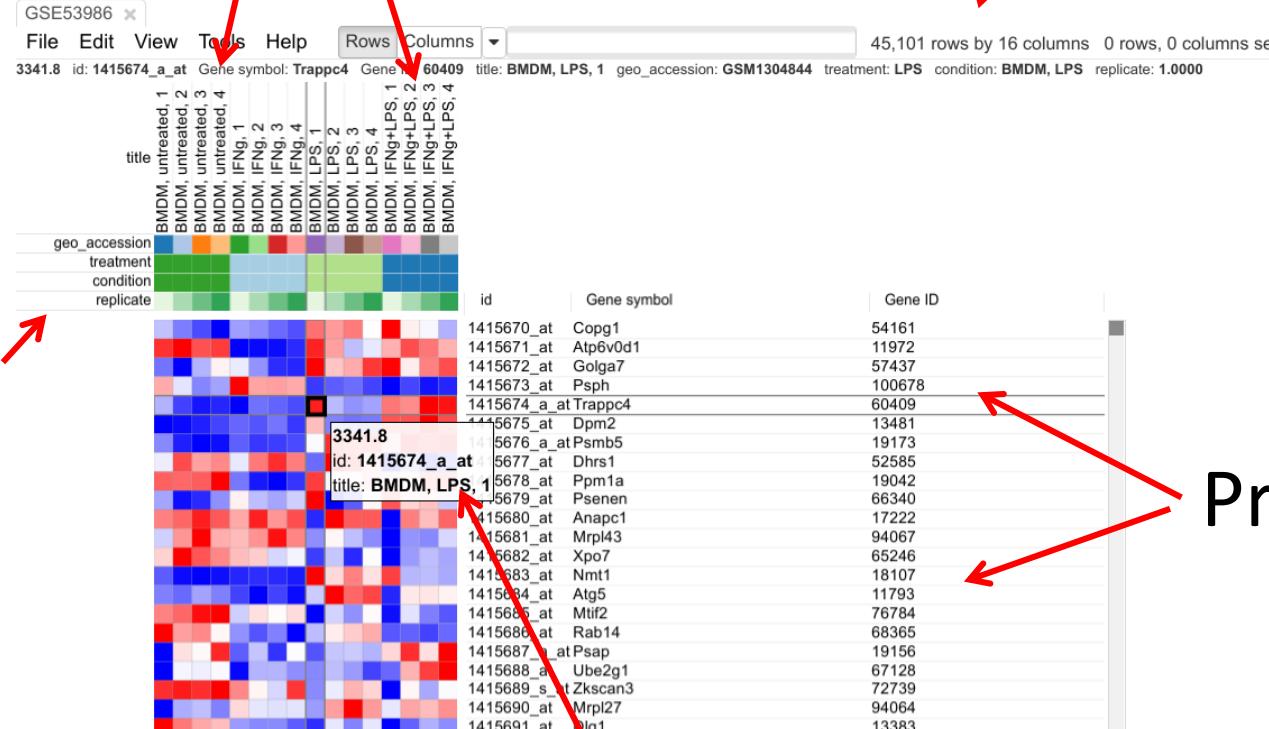


Interface overview

Samples (pData)

Dataset dimension

Sample
annotations
(right click
for context
menu)



Expression value (color scheme is relative)

Exploring individual genes

GSE53986 X

File Edit View Tools Help Rows Columns ▾ Acod1 1 match ⌂ ⌄ ⌁ 45,101 rows by 16 columns 1

title
geo_accession
treatment
condition
replicate

	id	Gene symbol	Gene ID
1427381_at	Acod1	16365	
1427382_a_at	Suv39h1	20937	
1427383_at	Irx6	64379	
1427384_at	Chd6	71389	
1427385_s_at	Actn1	109711	
1427386_at	Arhgef16	230972	
1427387_a_at	Itgb4	192897	
1427388_at	Lrrc2	74249	
1427389_at	Mfsd4b5	215928	
1427390_at	Bloc1s3	232946	
1427391_a_at	Col12a1	12816	
1427392_at	Dscaml1	114873	
1427393_at	F9	14071	
1427394_at	Igf2os	111975	
1427395_a_at	Aldh1a3	56847	
1427396_a_at	Csde1	229663	
1427397_at	Proser1	212127	
1427398_at	Muc4	140474	
1427399_a_at	Nxf7	170722	
1427400_at	Lbx1	16814	
1427401_at	Chrna5	110835	
1427402_at	Polr2f	69833	

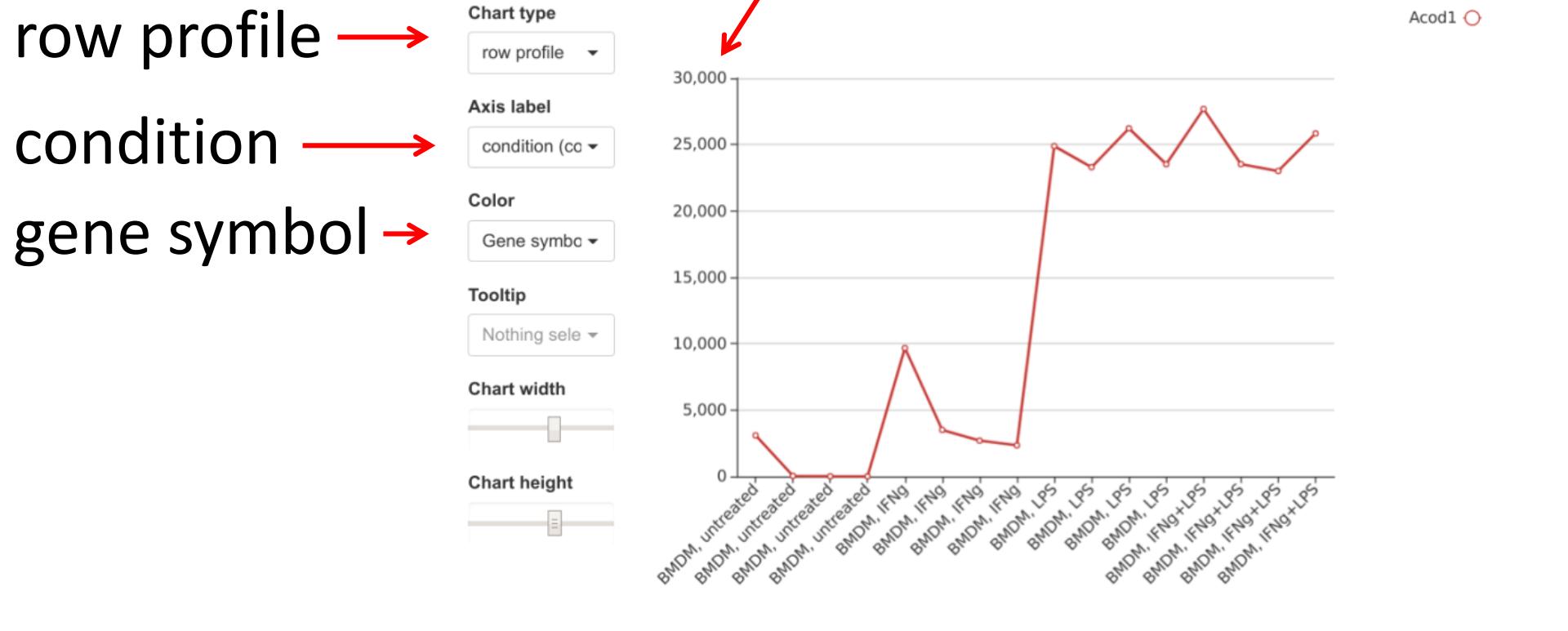
1. Enter “Acod1”

2. Click to scroll to the next hit

Row profile chart

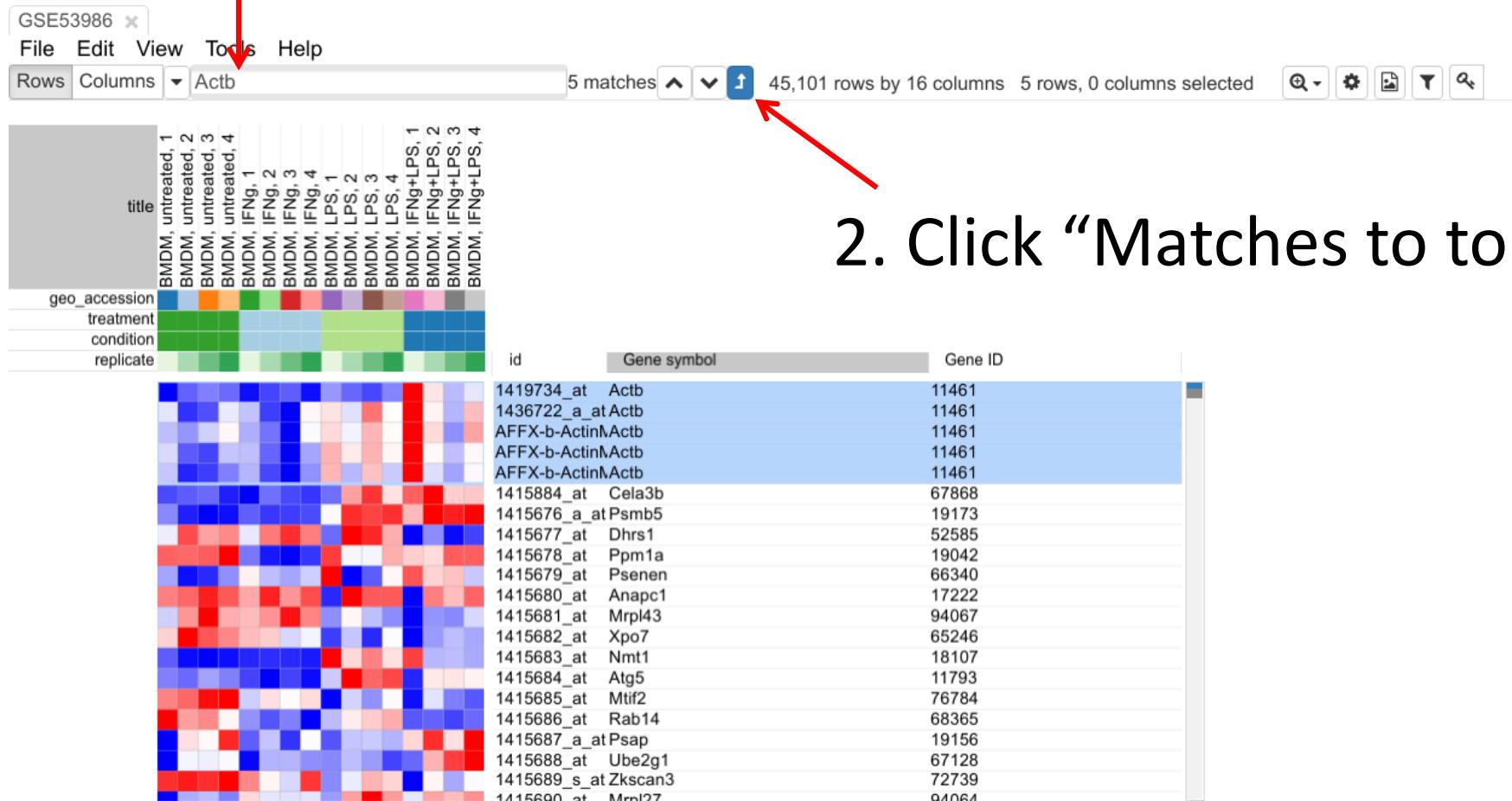
- ✓ Select all columns and Acod1 row
- ✓ Tools/Plots/Chart

Data is in linear scale!

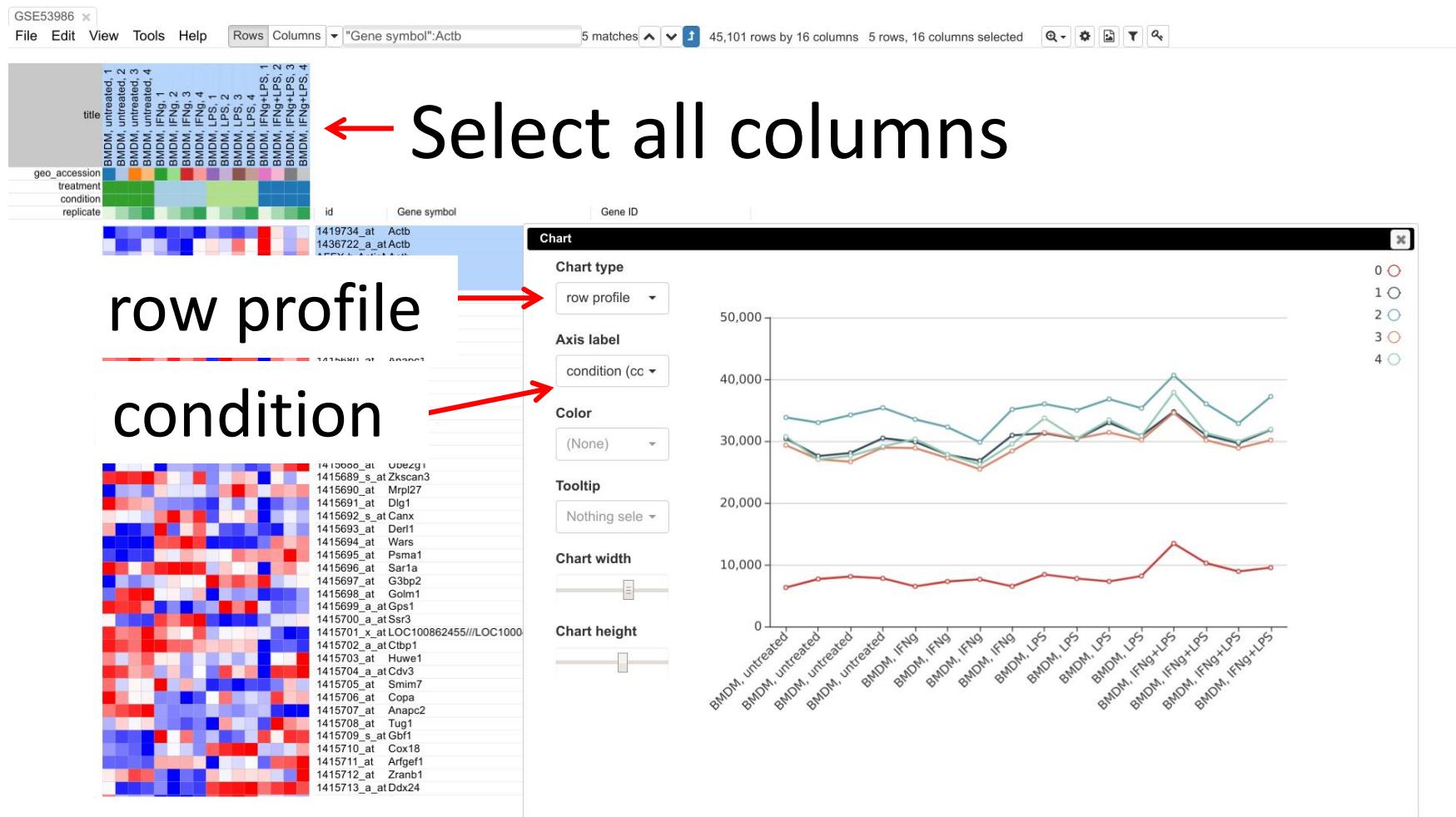


Let's look at Actb as a control

1. Enter “Actb”

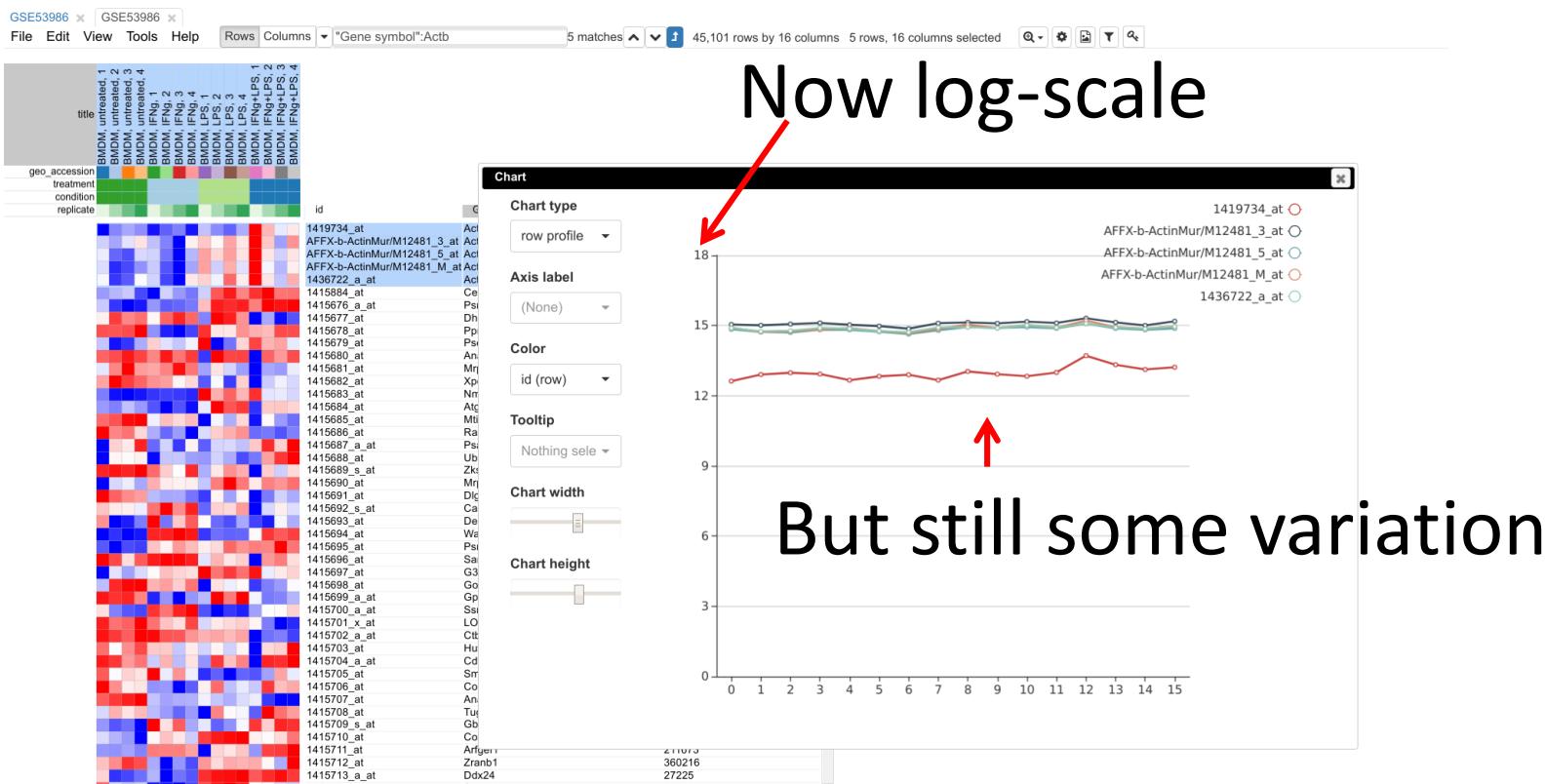


Actb expression chart: high variation (but in a linear scale)



Log 2 normalization

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



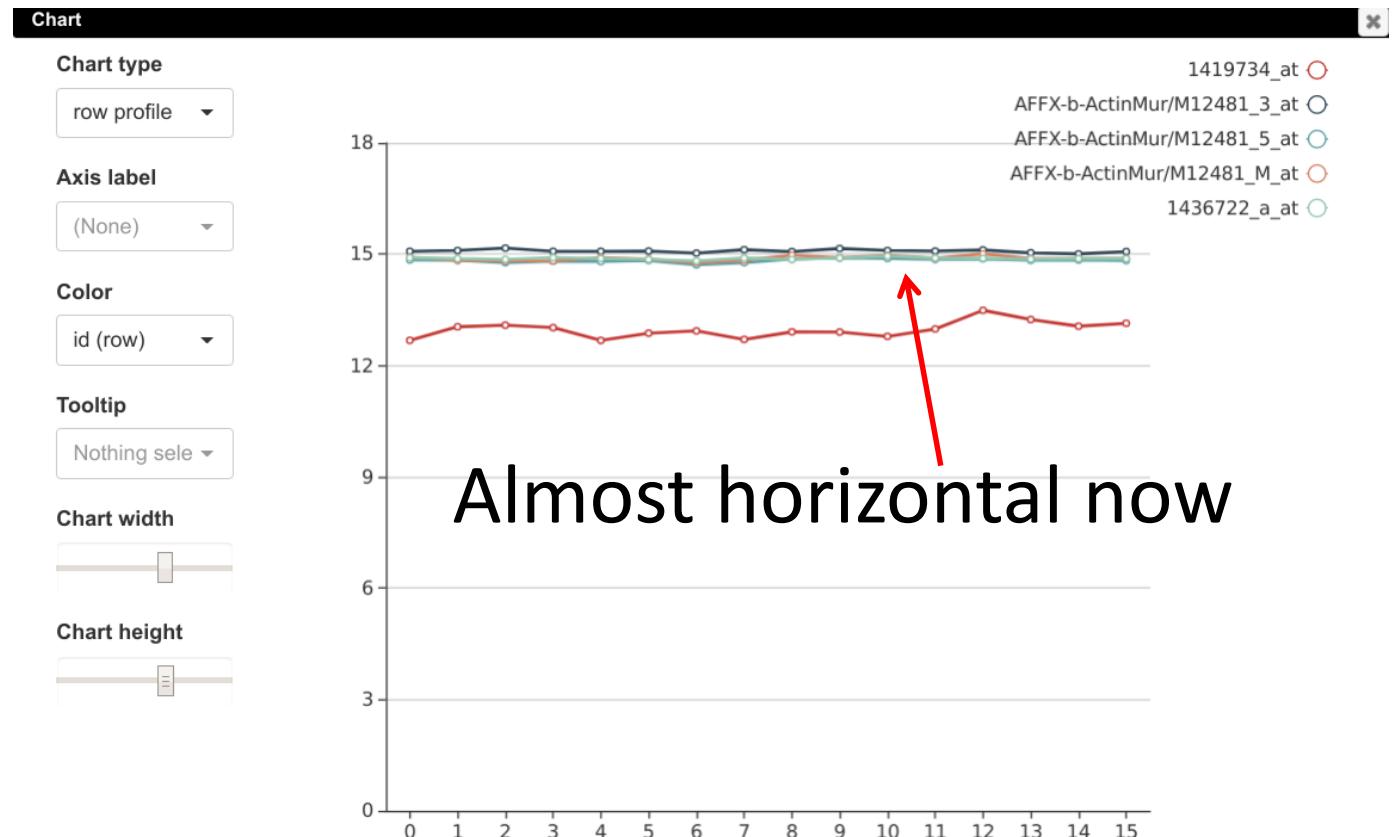
Now log-scale

But still some variation

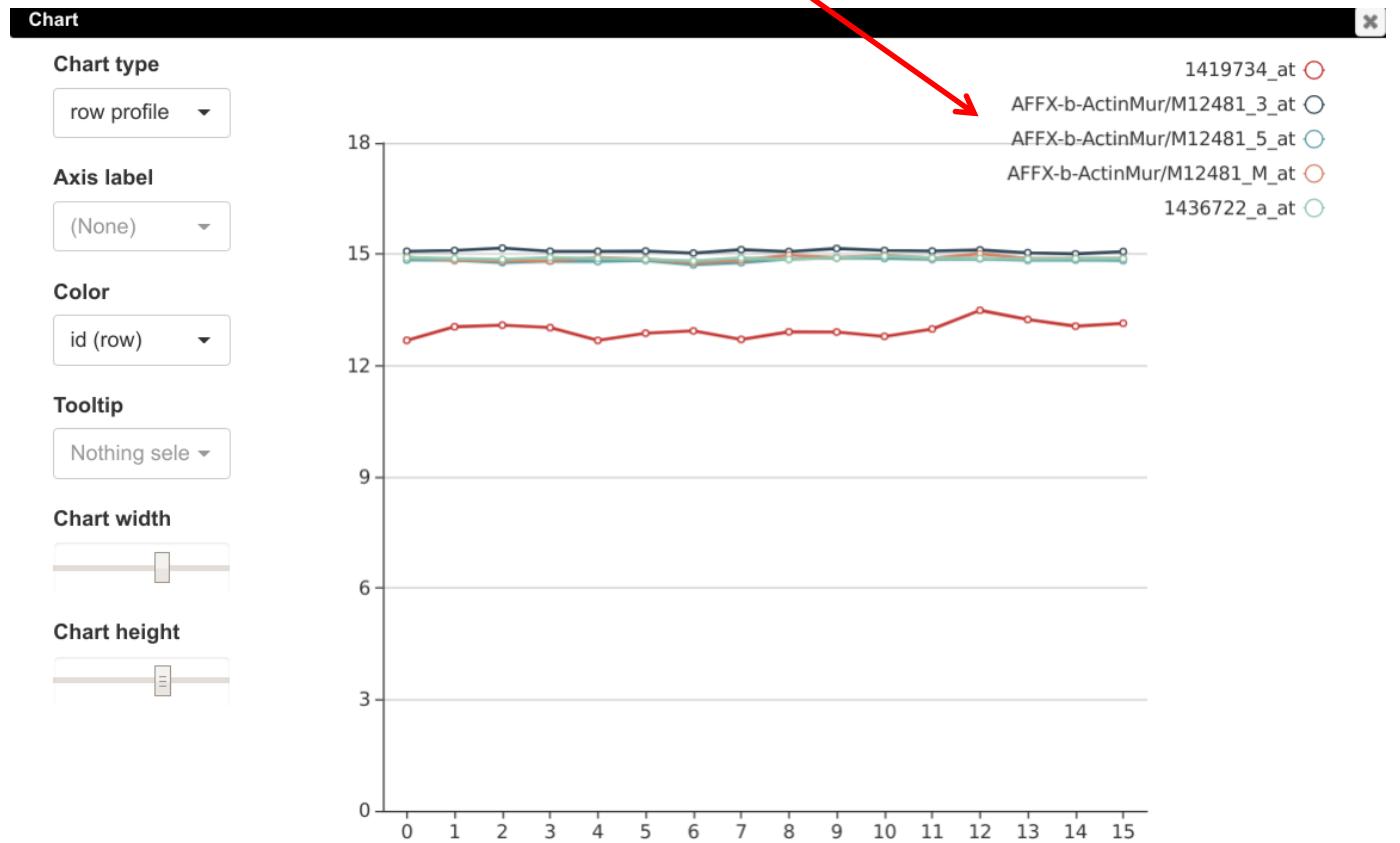
Quantile normalization

- ✓ 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

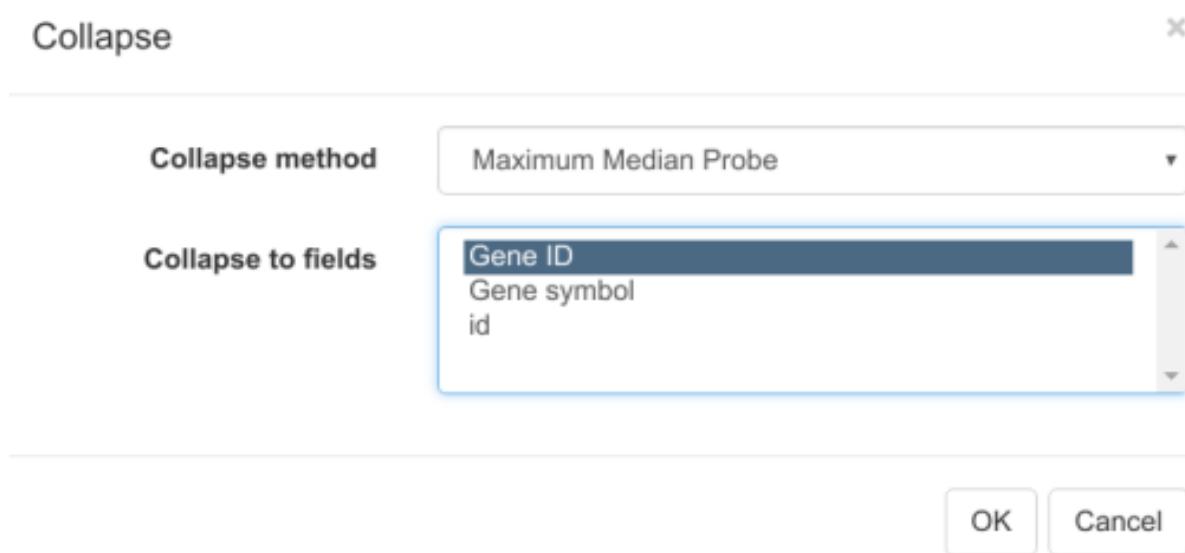


There are multiple probes per gene in microarrays



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

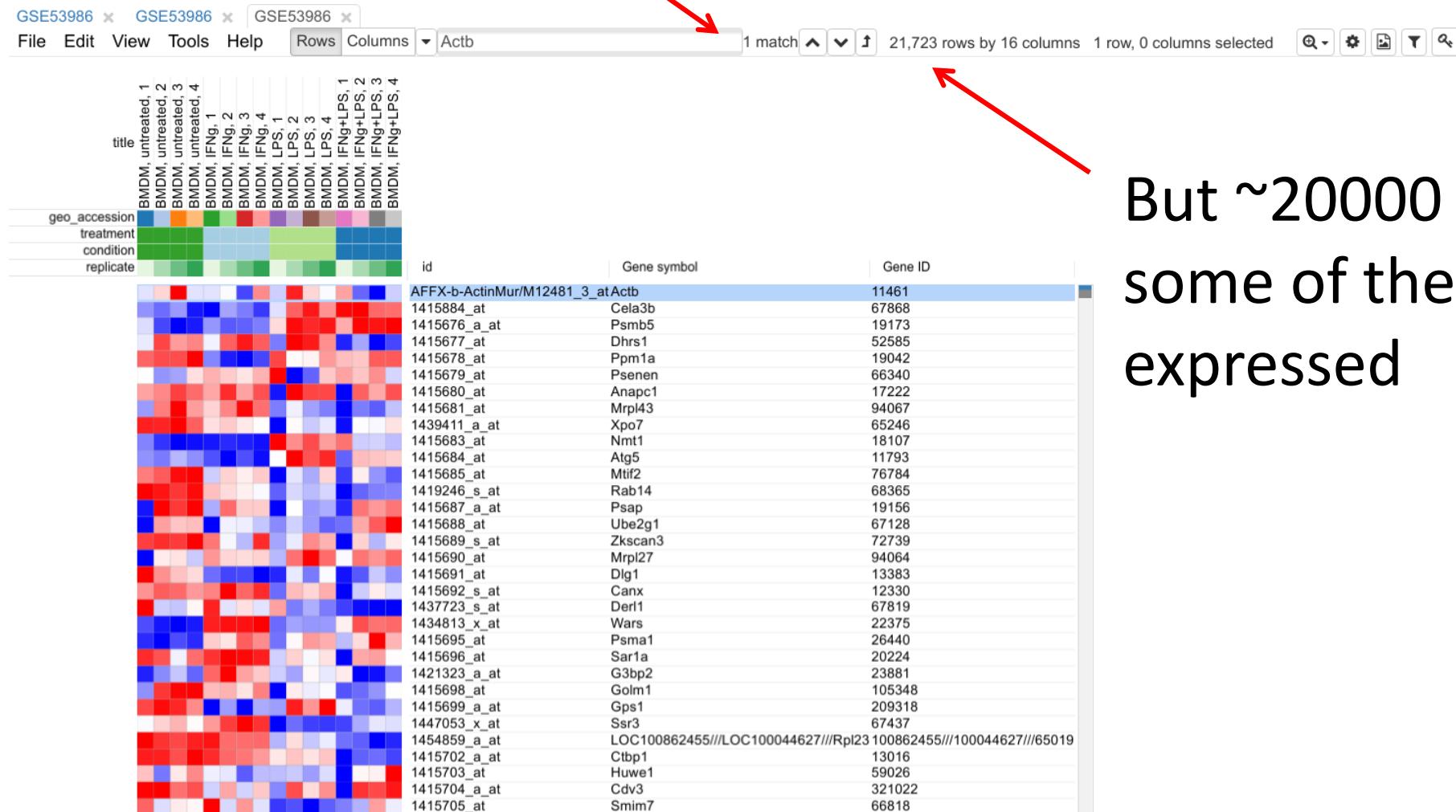
- ✓ Tools/Collapse



method = maximum
median probe

Grouping by
Gene ID

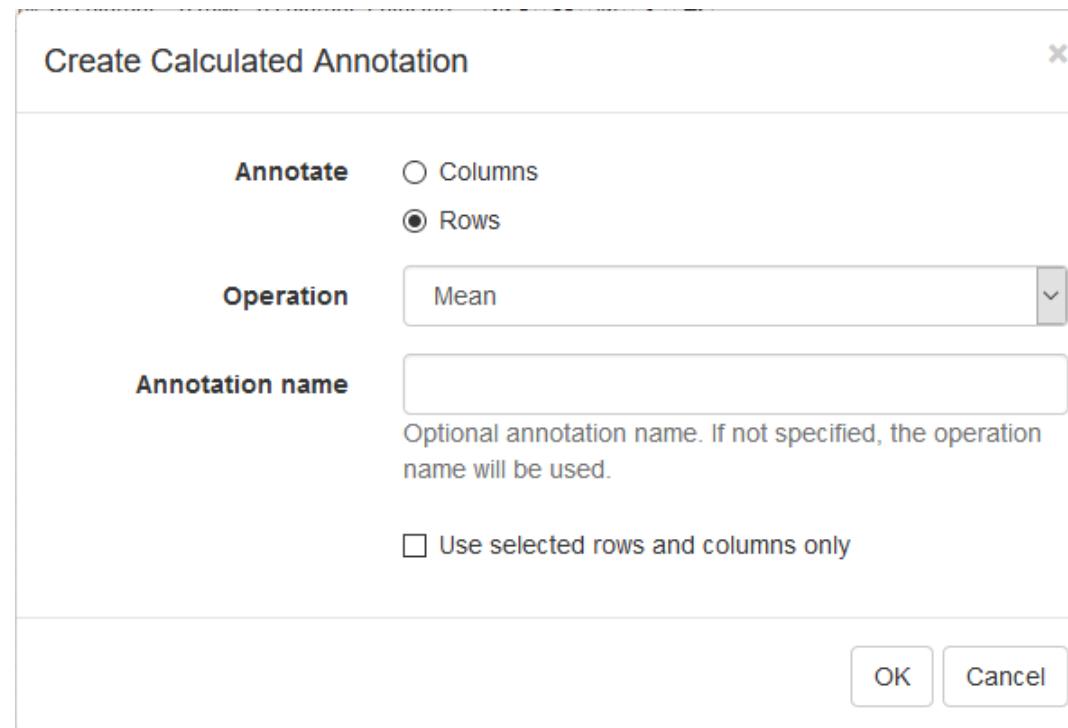
No more duplicates



But ~20000 genes,
some of them are not
expressed

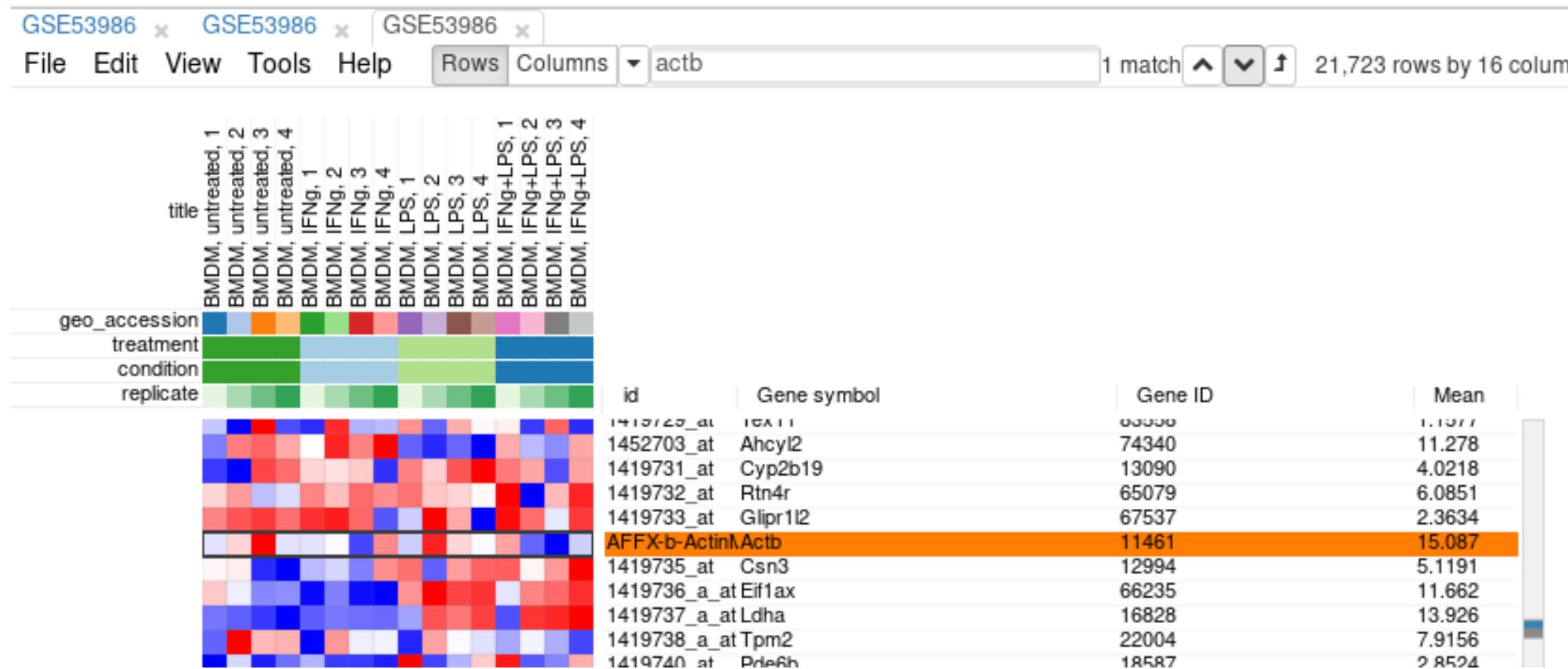
Filtering lowly expressed genes: calculating mean expression

- ✓ Tools/Create Calculated Annotation



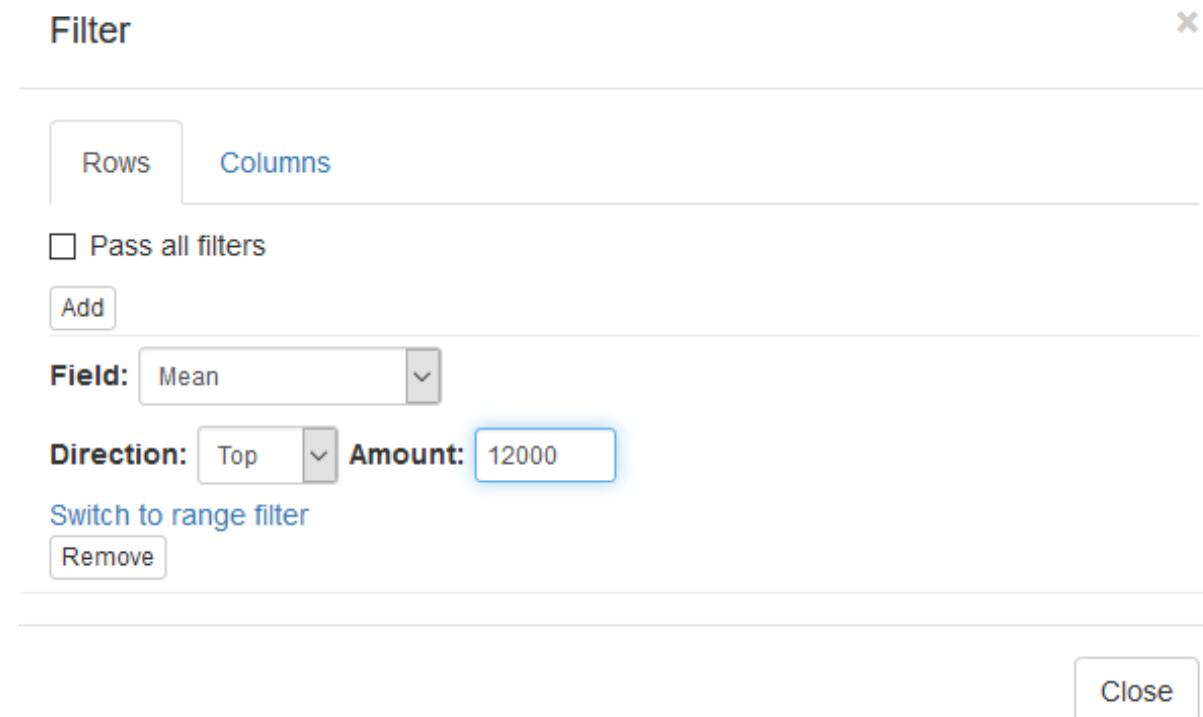
Operation: Mean
Optional name
(e.g. “mean_expression”)

Filtering lowly expressed genes: calculating mean expression result



Filtering lowly expressed genes: keeping only top 12000 genes

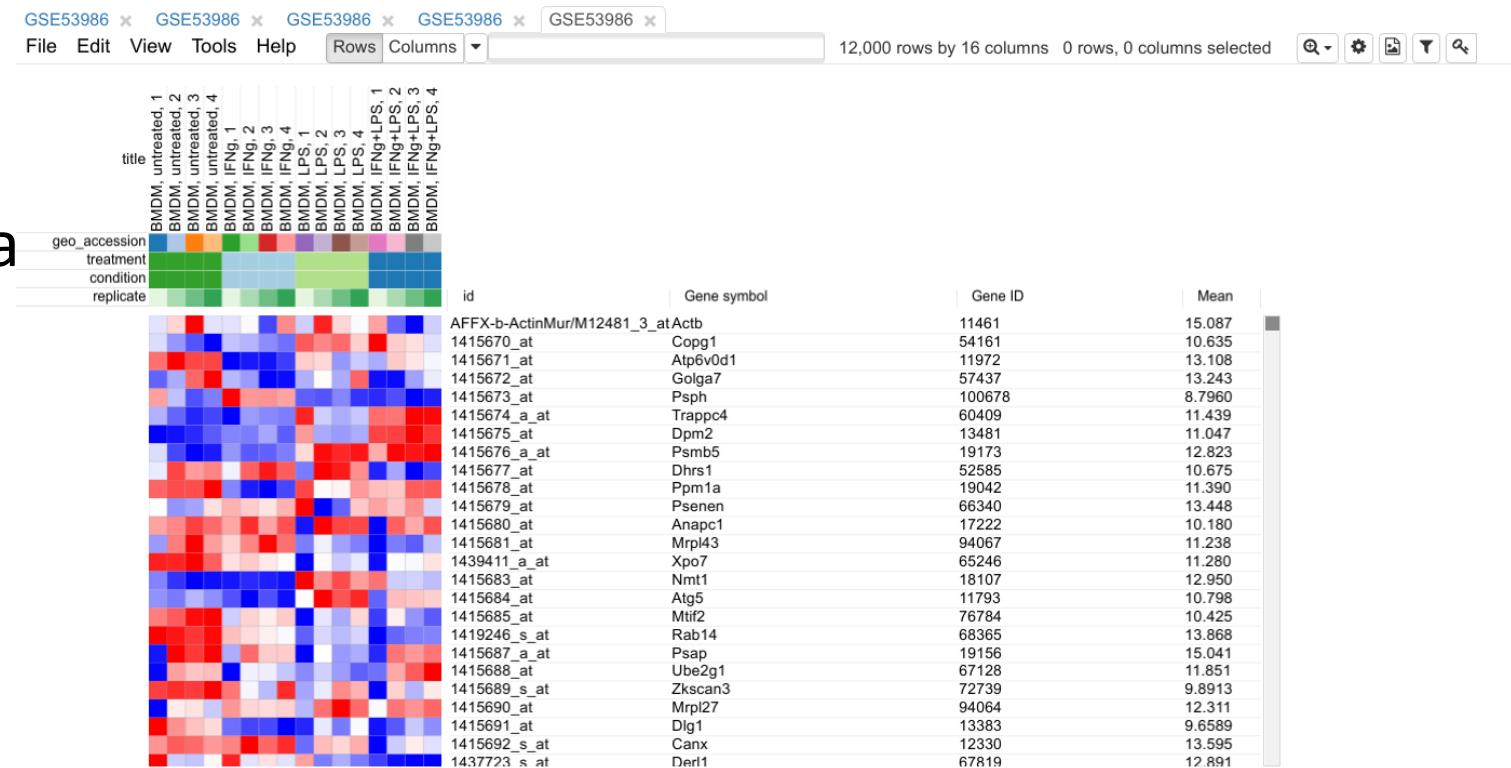
- ✓ Tools/Filter
- ✓ Add
- ✓ Field <- Mean
- ✓ Switch to top filter
- ✓ N <- 12000



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)

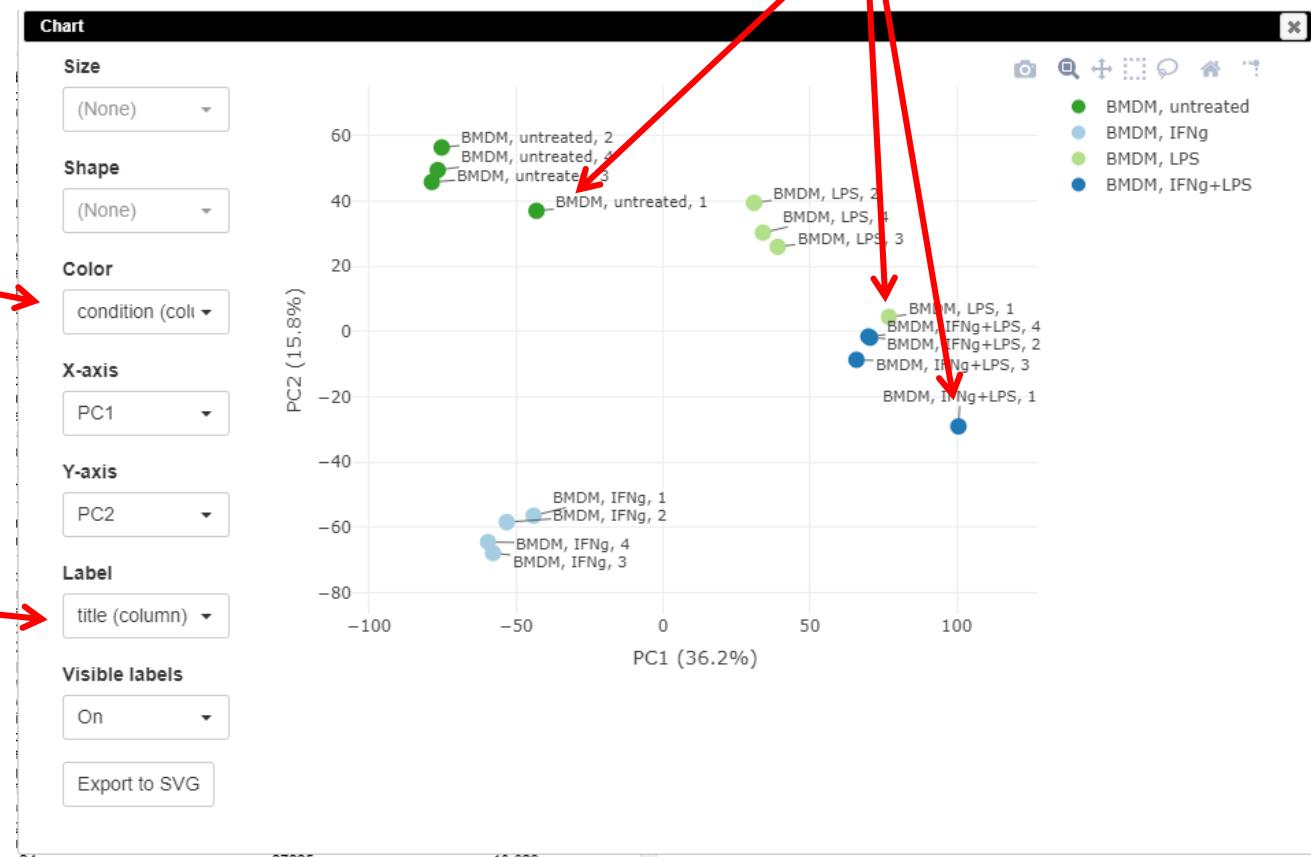


PCA

✓ Tools/Plots/PCA plot

color <-
treatment

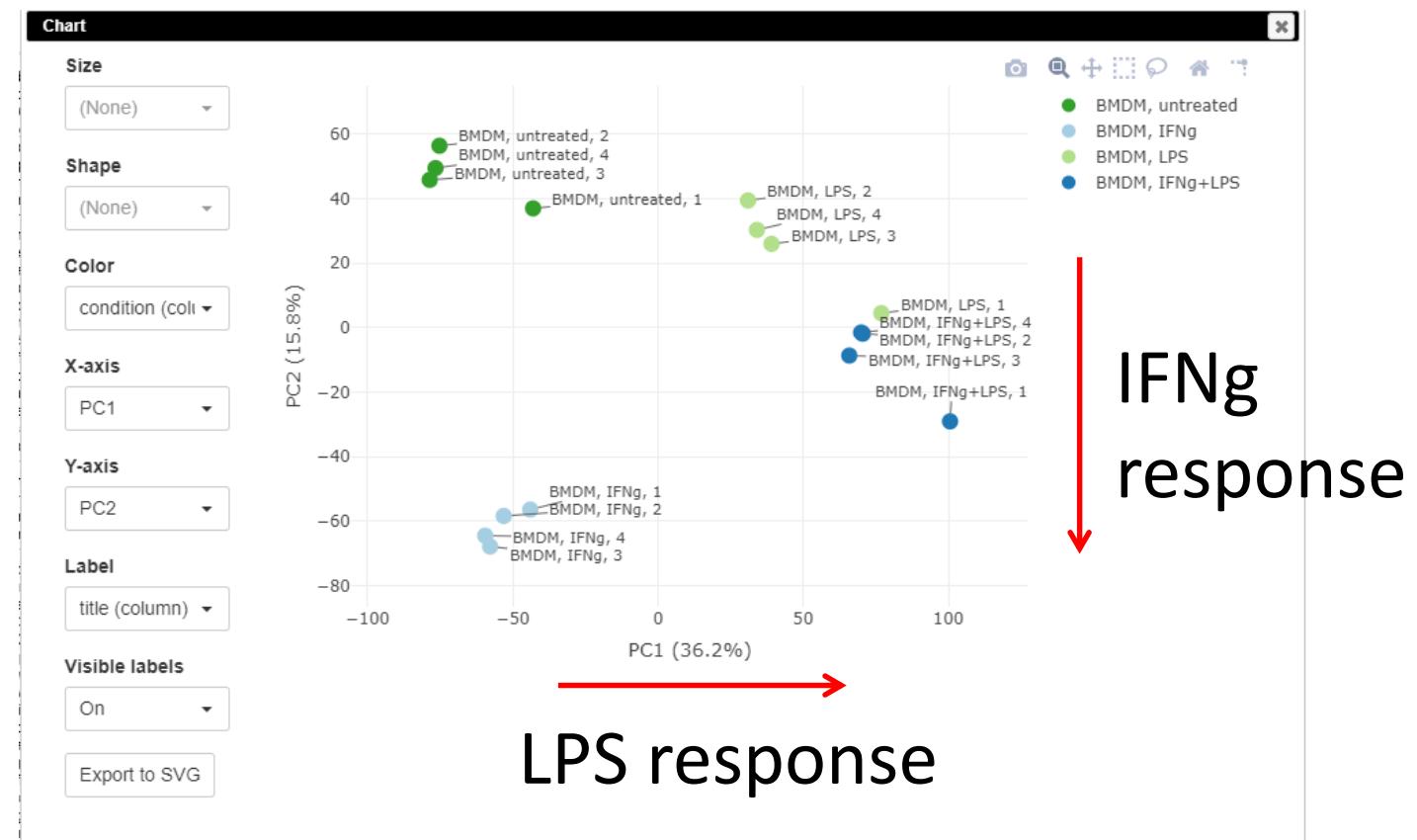
label <- title



Scale should be ~10-100, not 1000000

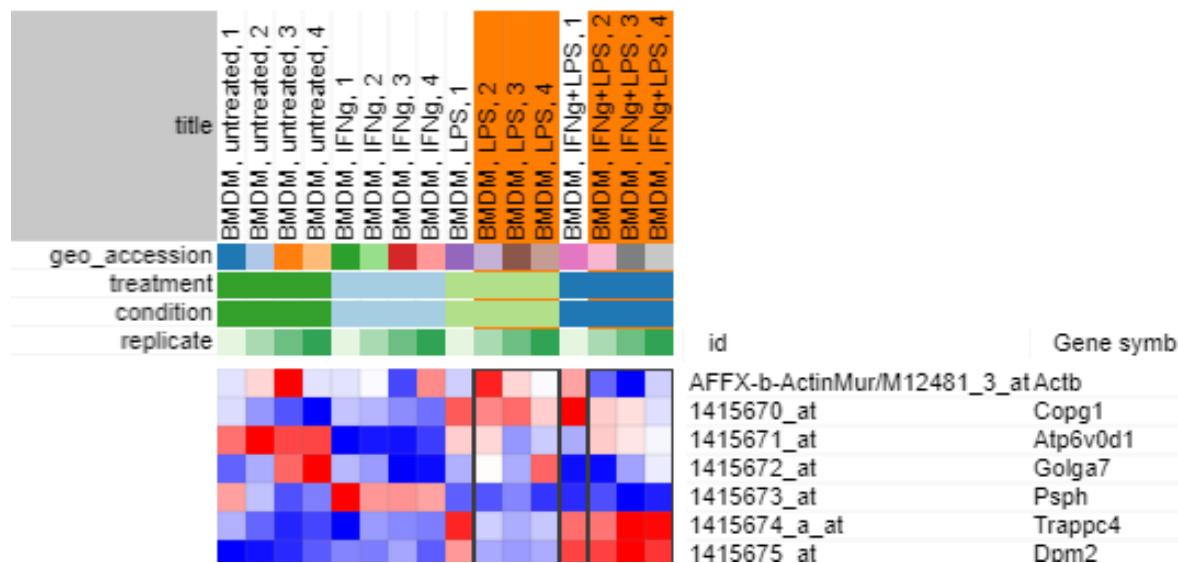
PCA

- ✓ 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/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 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
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).

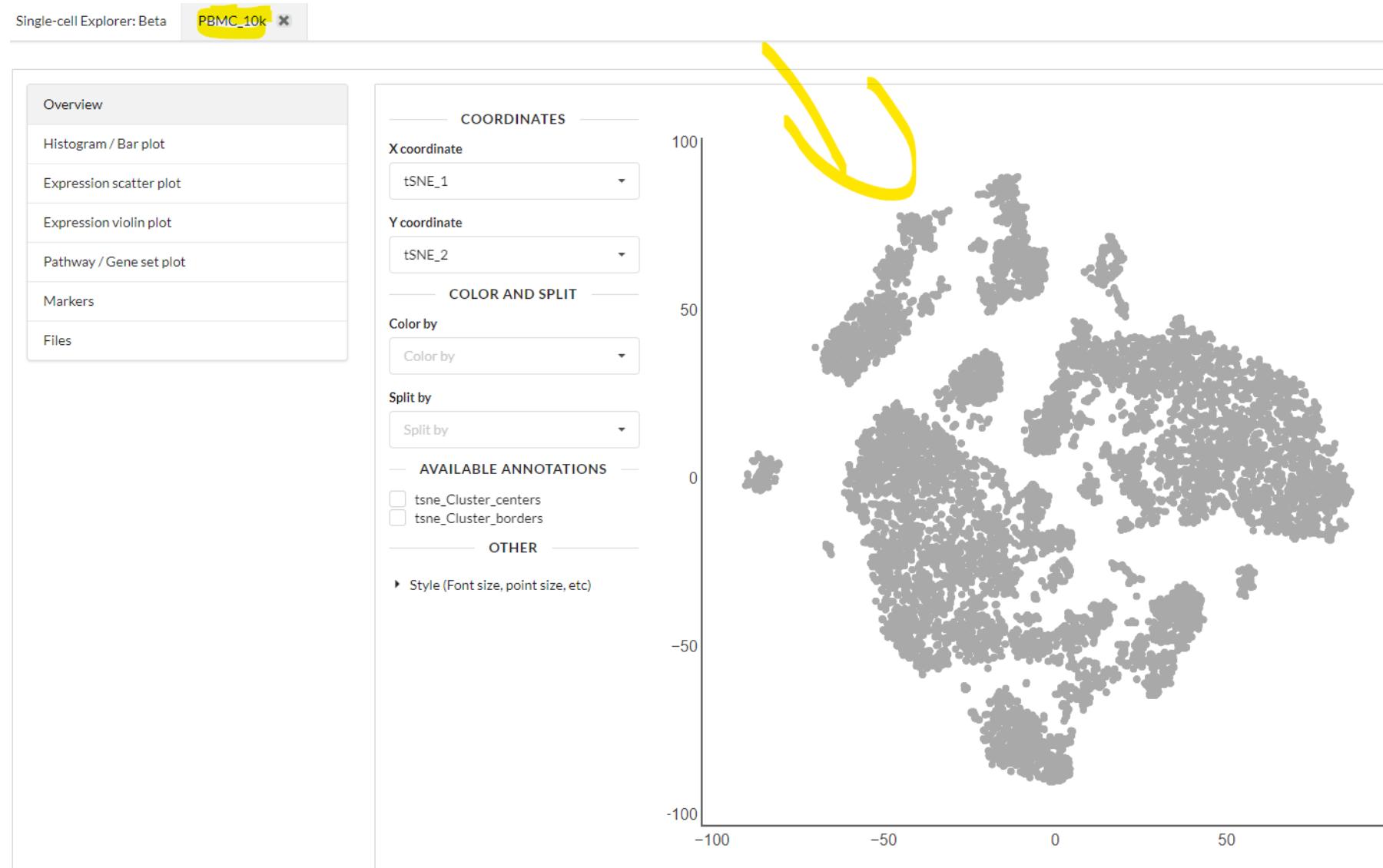
Currently available datasets are:

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

If you have any problem finding dataset

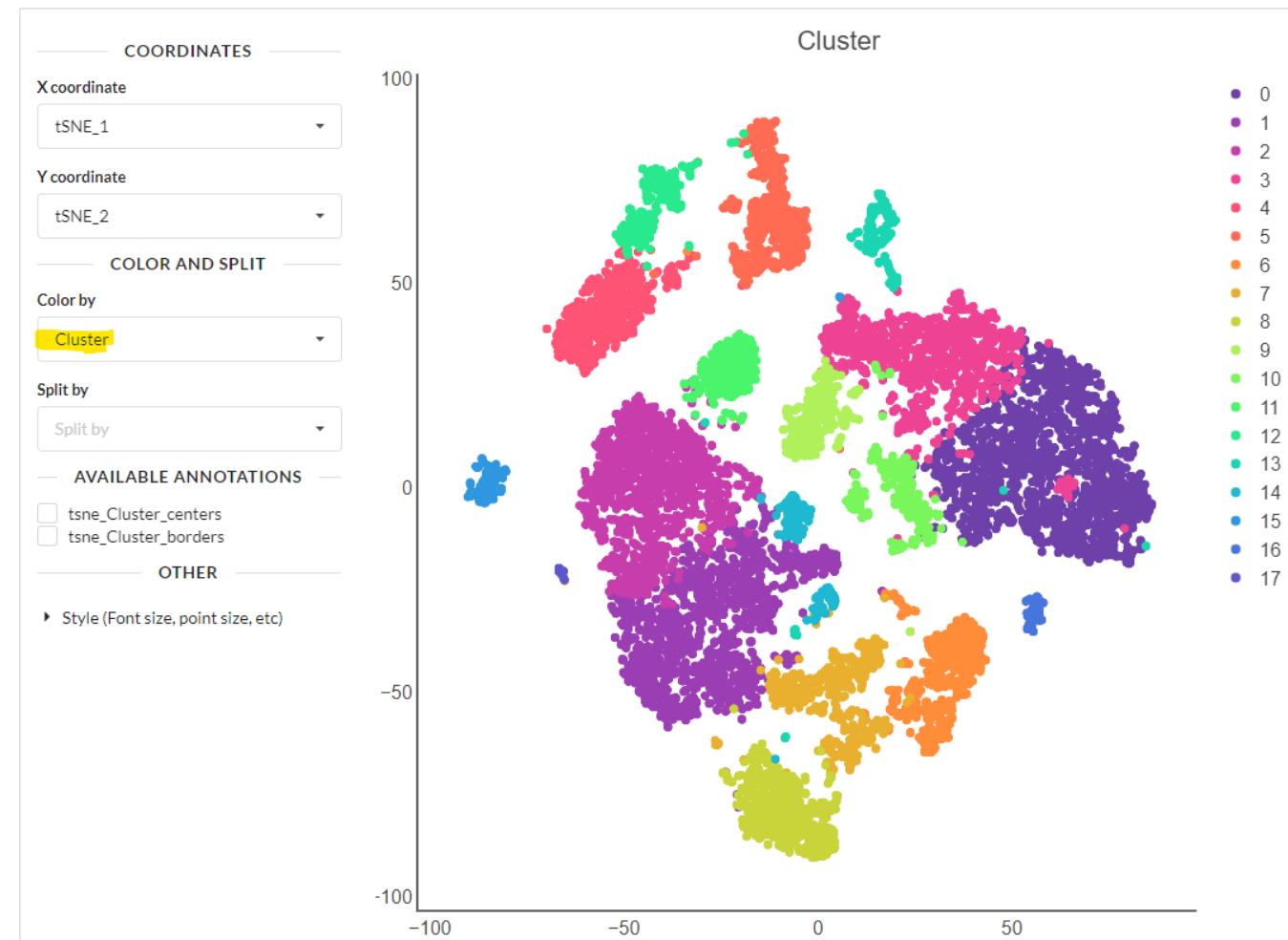
- ✓ Just go to https://artyomovlab.wustl.edu/sce/?token=PBMC_10k

Result should look like that

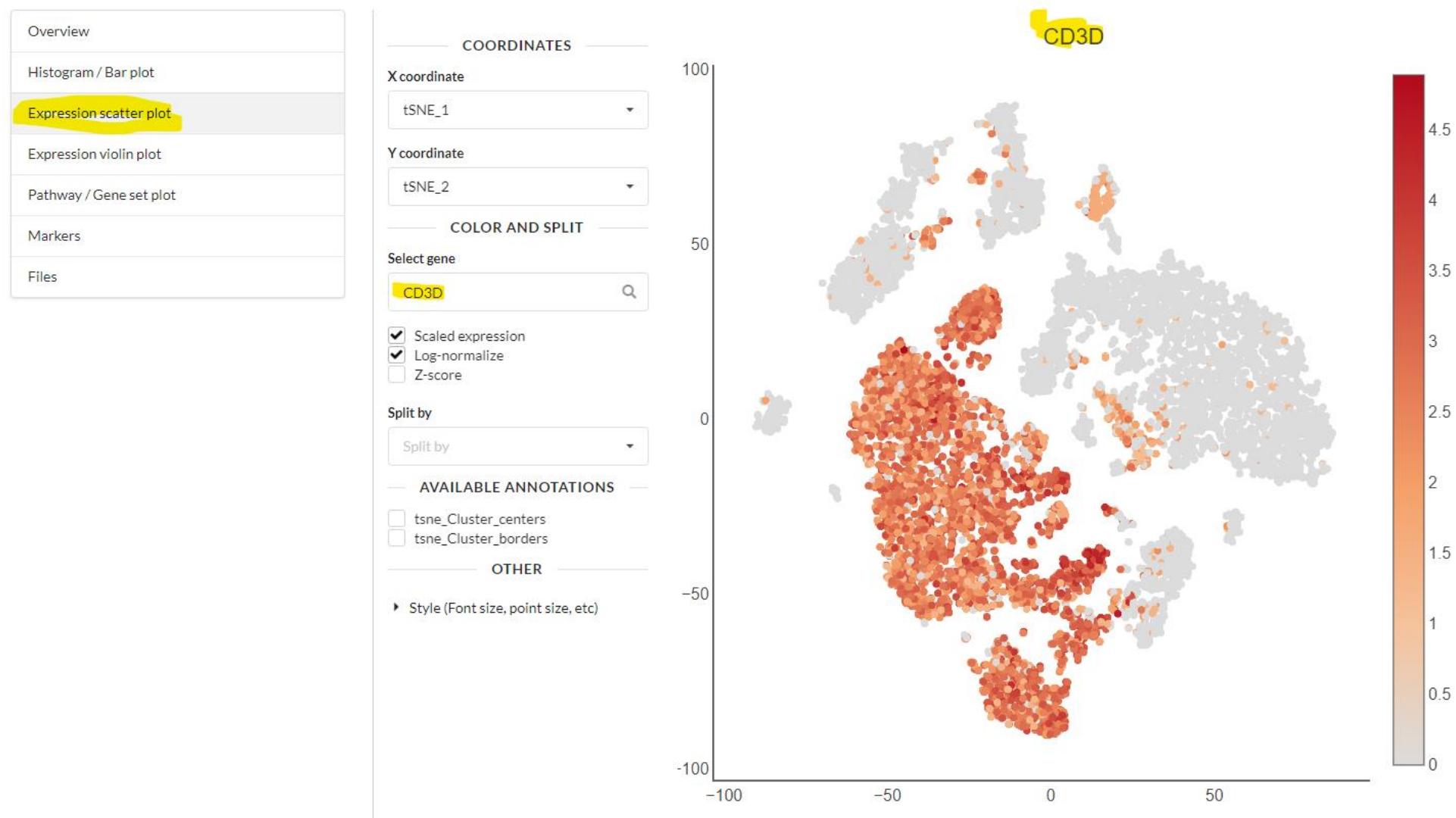


We can color the cells

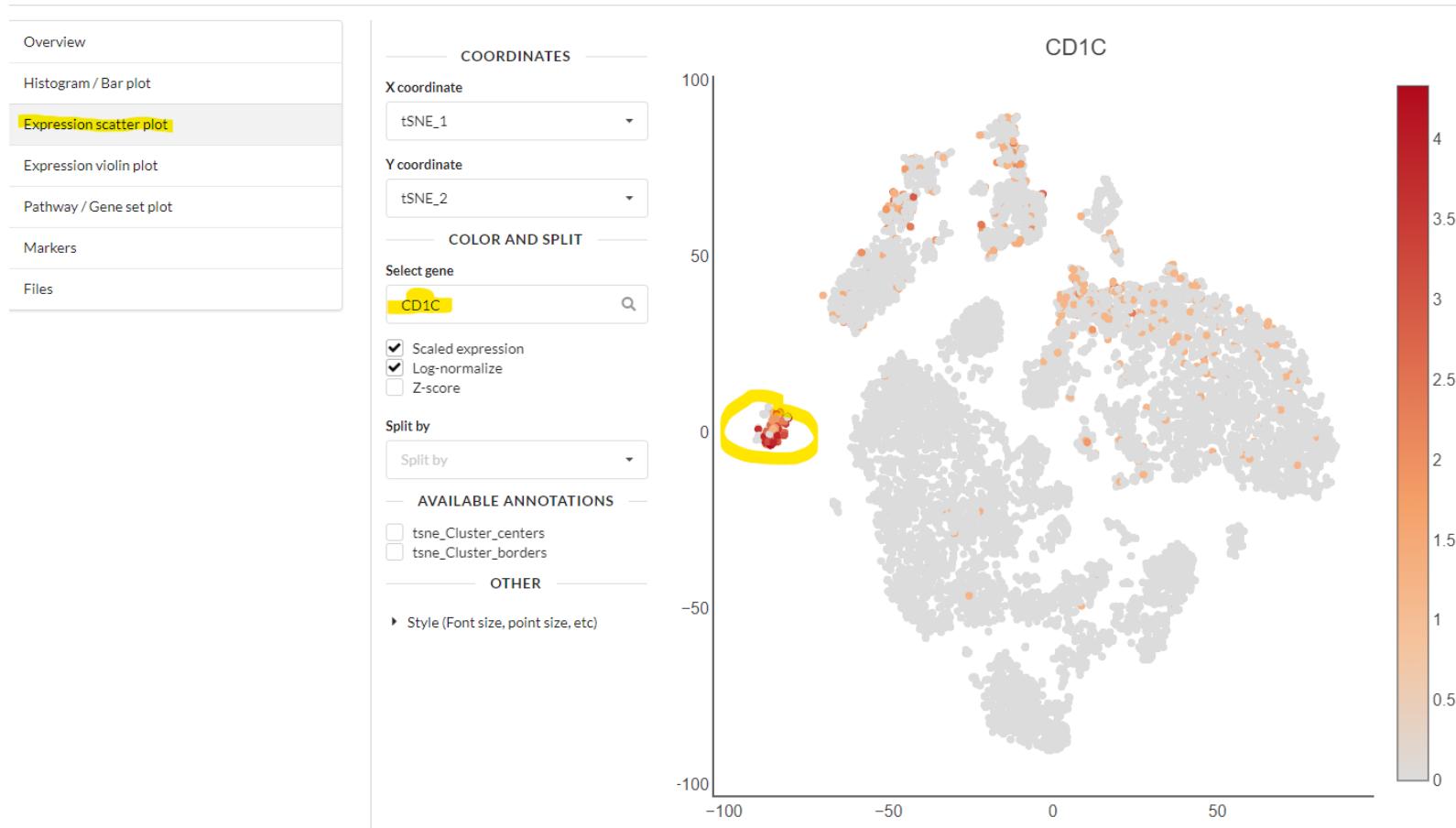
- ✓ Cluster
- ✓ Number of UMIs
- ✓ Number of genes detected
- ✓ tsne_Cluster_centers



Expression of CD3d



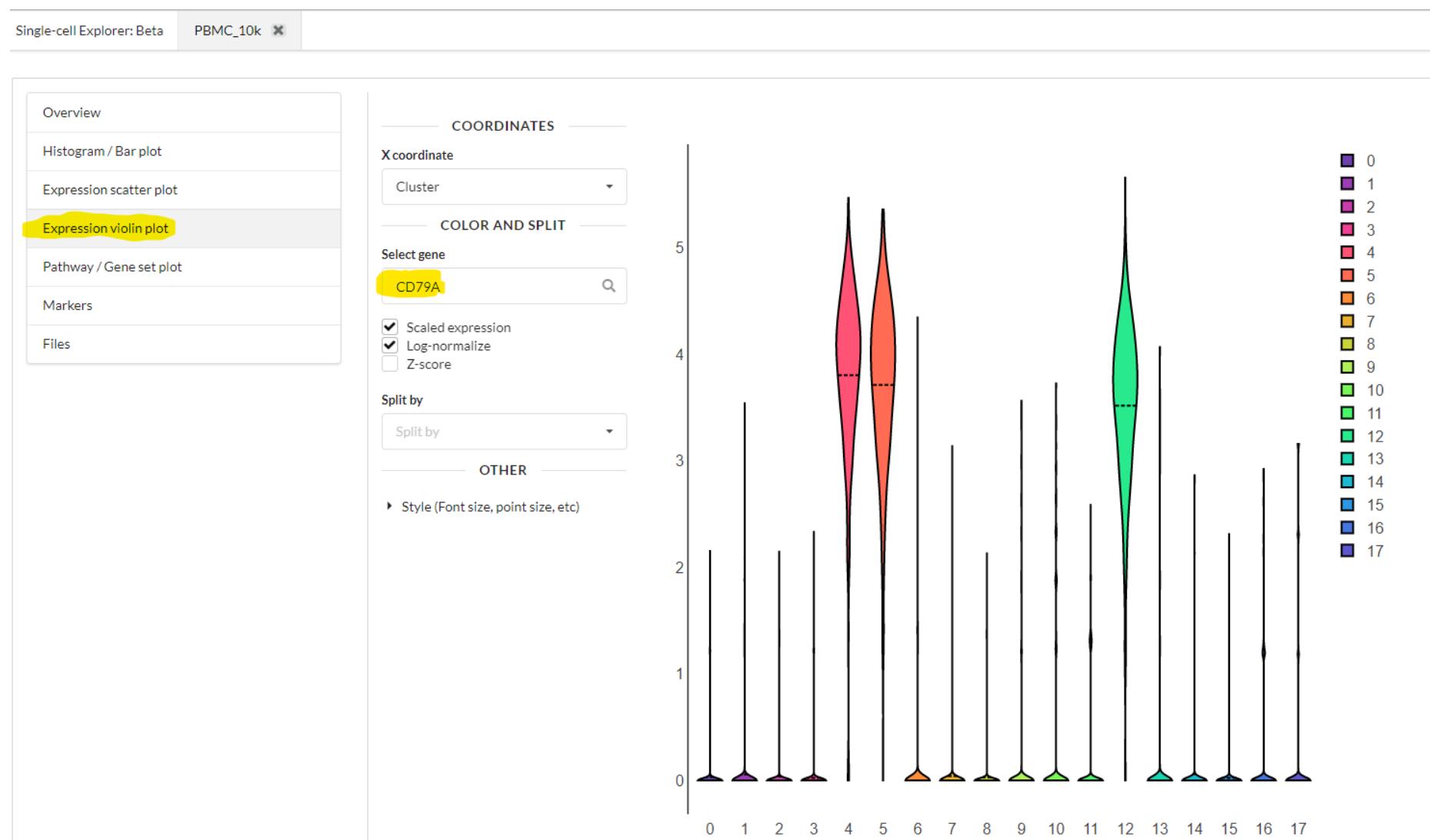
Or you can go for any of your favorite genes



Expression scatter plot

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

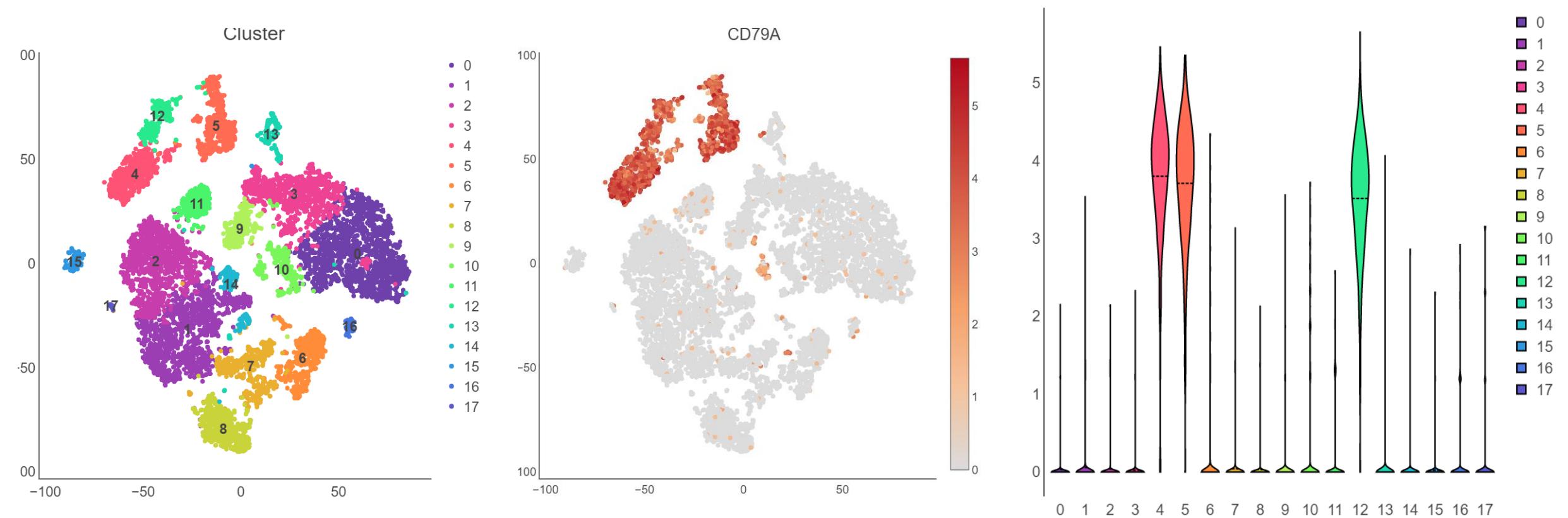
Violin plot



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

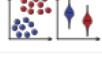
Markers tab

Single-cell Explorer: Beta | PBMC_10k X

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
S100A8	~	=	>	< 1e-	< 1e-	>
S100A9		0	2.4105	0	0	1
S100A12		0	2.2626	0	0	1
LYZ		0	1.8552	0	0	1
VCAN		0	1.8376	0	0	0.998
MNDA		0	1.6095	0	0	1
FCN1		0	1.53	0	0	1
FOS		0	1.3692	0	0	1
CTSS		0	1.3573	0	0	1
CD14		0	1.3368	0	0	0.968

Previous
Page **1** of 678
10 rows ▾
Next

[Download current table](#)

Markers tab: what's the cluster 6?

Single-cell Explorer: Beta PBMC_10k X

- 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
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
KLRD1	6	1.9782	0	0	0.972	0.073

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

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
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)

The screenshot shows a web browser window with a yellow highlight on the title bar 'Single-cell Explorer: Beta'. The main content area displays the 'Single-cell explorer: beta' page. The page includes a brief description of the project, instructions for opening datasets or uploading own data, and a list of currently available datasets. A table lists datasets by GSE/SRA id and description, with one entry highlighted.

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

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 X

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	<p>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 pancreatic cells. This study shows that the transcriptional landscape of these cells is highly heterogeneous, with many distinct cell types and subtypes. We find that some cell types are highly enriched for specific genes, such as <i>CCKAR</i>, <i>PPARγ</i>, and <i>ATM</i>. We also identify several novel cell types, including a population of cells that express both <i>ATM</i> and <i>PPARγ</i>. These findings provide new insights into the biology of the pancreas and may help to identify new therapeutic targets for cancer.</p> <p>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, with additional samples from the same tissues obtained by single-cell RNA-seq). The data show that ischaemic sensitivity varies significantly between different cell types and tissues, with some cells showing high sensitivity to oxygen deprivation and others being more resistant. This information could be used to predict which cells are most likely to be affected by ischaemia and to develop strategies for protecting them.</p> <p>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 continuous trajectories. In this study, we performed a comprehensive analysis of the transcriptional landscape of CD34+ cells, which are multipotent stem cells that can differentiate into various cell types. We found that these cells exhibit a wide range of gene expression patterns, with some cells showing high levels of differentiation and others remaining relatively undifferentiated. This information could be used to better understand the biology of these cells and to develop new strategies for their use in medical applications.</p> <p>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-decidua interactions are critical for successful implantation and development of the embryo. In this study, we used single-cell RNA-seq to analyze the transcriptional landscape of these cells, with the aim of reconstructing the complex interactions between them. We found that trophoblast cells express a unique set of genes, including <i>PLXNC1</i>, <i>PLXNC2</i>, and <i>PLXNC3</i>, which are involved in adhesion and migration. We also identified several novel cell types, including a population of cells that express both <i>PLXNC1</i> and <i>PLXNC2</i>.</p> <p>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. Using single-cell RNA-seq, we analyzed the transcriptional landscape of these cells and found that they exhibit a wide range of gene expression patterns, with some cells showing high levels of differentiation and others remaining relatively undifferentiated. This information could be used to better understand the biology of these cells and to develop new strategies for their use in medical applications.</p> <p>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, we performed a comprehensive analysis of the transcriptional landscape of these organoids, with the aim of identifying the key genes and pathways that drive their development and differentiation. We found that these organoids exhibit a wide range of gene expression patterns, with some cells showing high levels of differentiation and others remaining relatively undifferentiated. This information could be used to better understand the biology of these cells and to develop new strategies for their use in medical applications.</p> <p>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 and then differentiating them into interneurons. We found that this approach yields a high proportion of interneurons, with a low rate of differentiation into other cell types. This information could be used to better understand the biology of these cells and to develop new strategies for their use in medical applications.</p>

Previous Page 1 of 1 10 rows ▾ Next

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

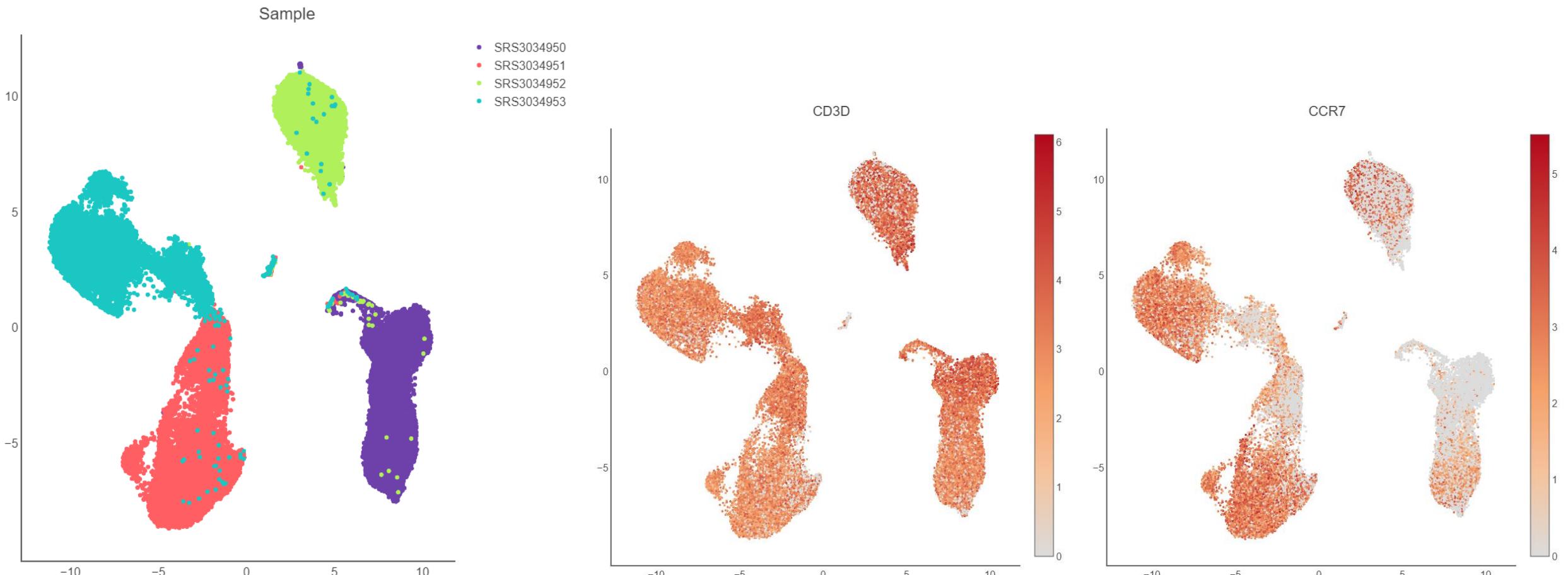
When we first analyzed 1000 dataset two main issues were identified:

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

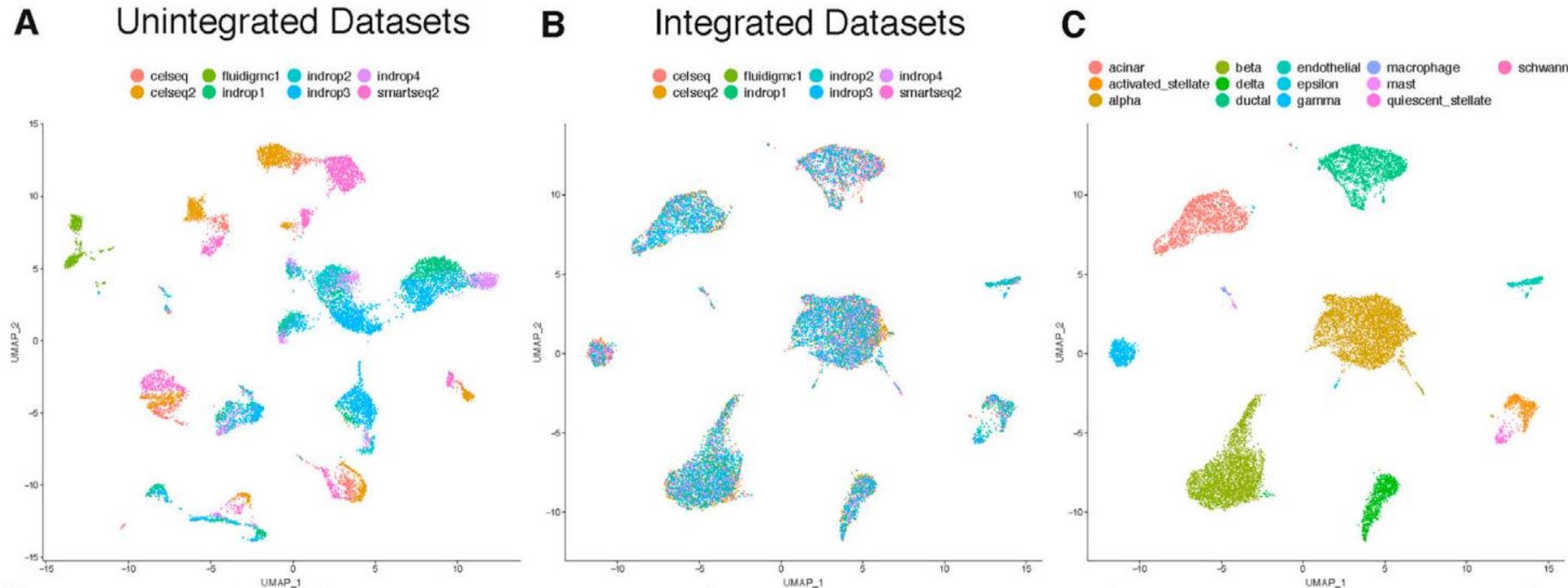
Most of the dataset processing was done
by Maria Firuleva



Issues: donor effect



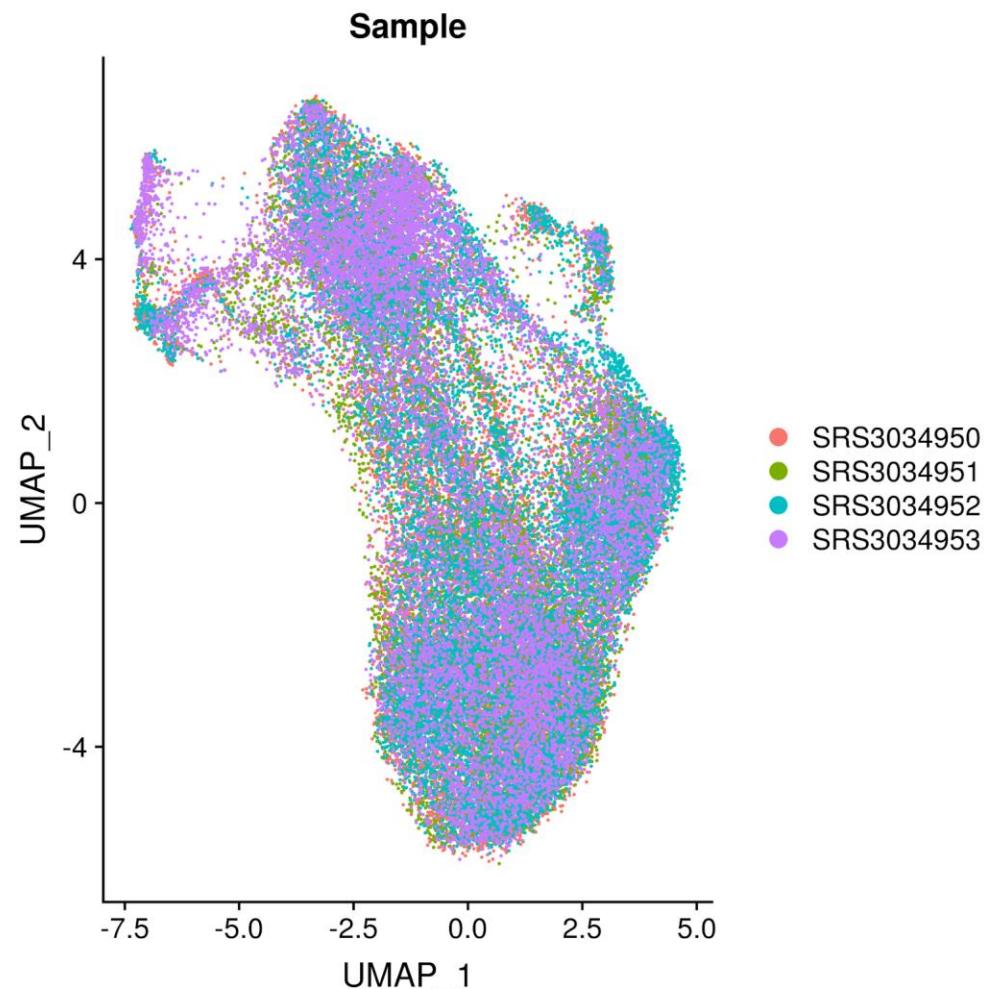
Recent developments of methods



- ✓ 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



Single-cell Explorer

- ✓ Easy, quick and friendly way to look at public single-cell RNA-seq data
- ✓ We try to get as many datasets as possible
- ✓ We will implement more features soon (like searching for clusters with specific signatures)

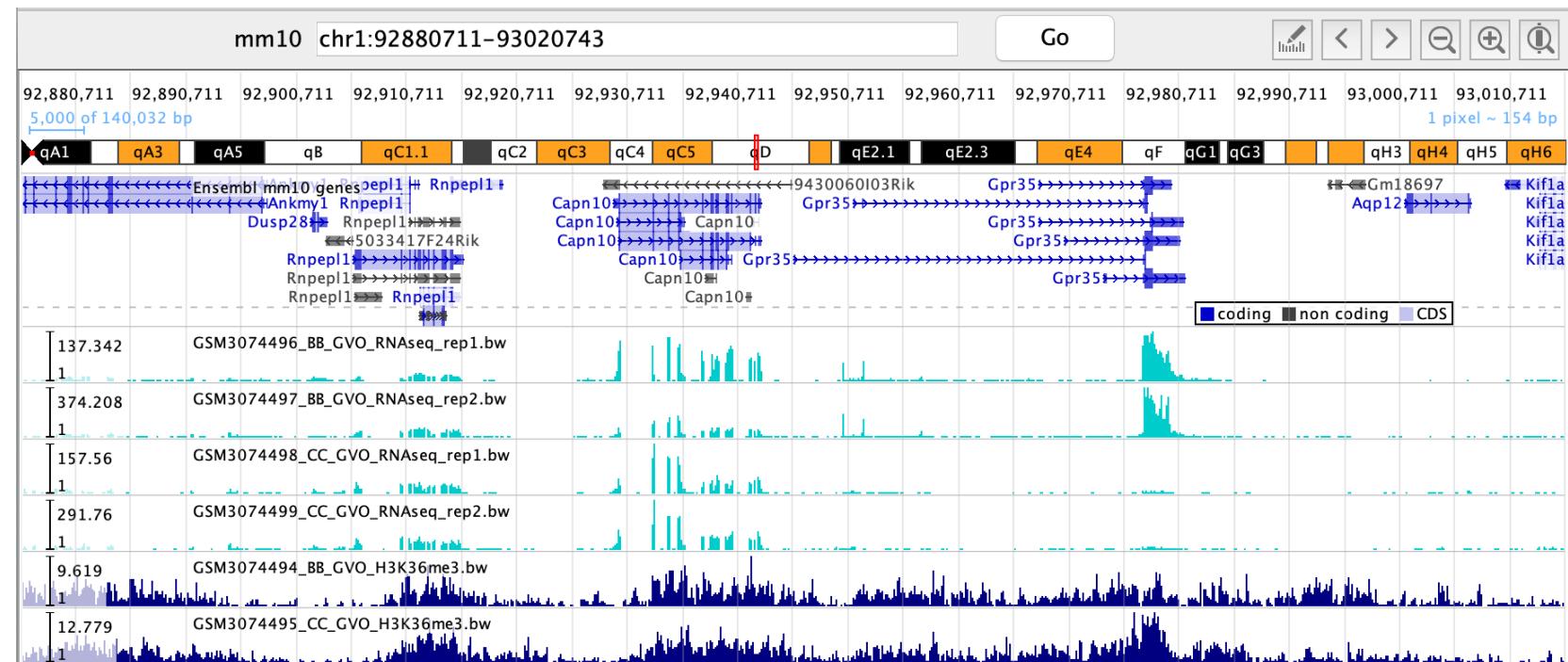
Genome browsers

- ✓ 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.4882), released on May 17, 2019

Windows
Mac and Linux

Download	Description
jbr-1.0.beta.4882_x64.zip	Windows 64-bit ZIP archive (includes bundled 64-bit Java Runtime)
jbr-1.0.beta.4882_x86.zip	Windows 32-bit ZIP archive (includes bundled 32-bit Java Runtime)
jbr-1.0.beta.4882.dmg	Mac installer (includes bundled 64-bit Java Runtime)
jbr-1.0.beta.4882.tar.gz	Linux archive (includes bundled 64-bit Java Runtime)

Prerequisites: Download JBR

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

Downloads

JBR Genome Browser (build 1.0.beta.4882), released on May 17, 2019

Windows
Mac and Linux

Download	Description
jbr-1.0.beta.4882_x64.zip	Windows 64-bit ZIP archive (includes bundled 64-bit Java Runtime)
jbr-1.0.beta.4882_x86.zip	Windows 32-bit ZIP archive (includes bundled 32-bit Java Runtime)
jbr-1.0.beta.4882.dmg	Mac installer (includes bundled 64-bit Java Runtime)
jbr-1.0.beta.4882.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. We recommend choosing 64-bit version if your OS is 64-bit.

Windows:

- Unpack the browser 1.0.beta.4882_x64.zip file:
- Run `jbr.exe`.

MacOS:

- Download the 1.0.beta.4882.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 commandline: `open -n "/Applications/JBR 1.0.app"`

Linux:

- Unpack the browser 1.0.beta.4882.tar.gz file using the following command:
`tar -xzf 1.0.beta.4882.tar.gz`
- Run `jbr.sh` from the bin subdirectory.

Launch JBR



Genome Browser

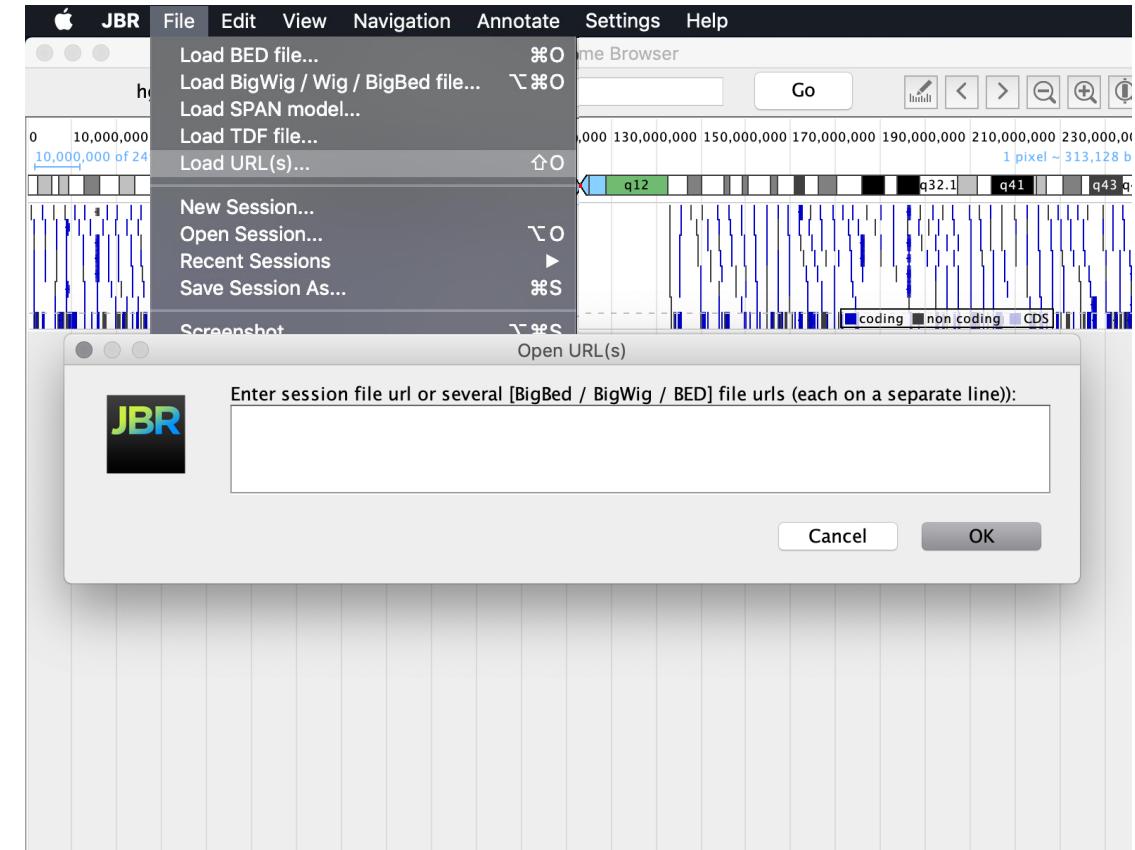
[May 30, 2018 22:55:14] Loading genes /Users/romeo/.jbr_browser/genomes/hg19/Homo_sapiens.GRCh37.87.gtf.gz: done in 802.5 ms

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

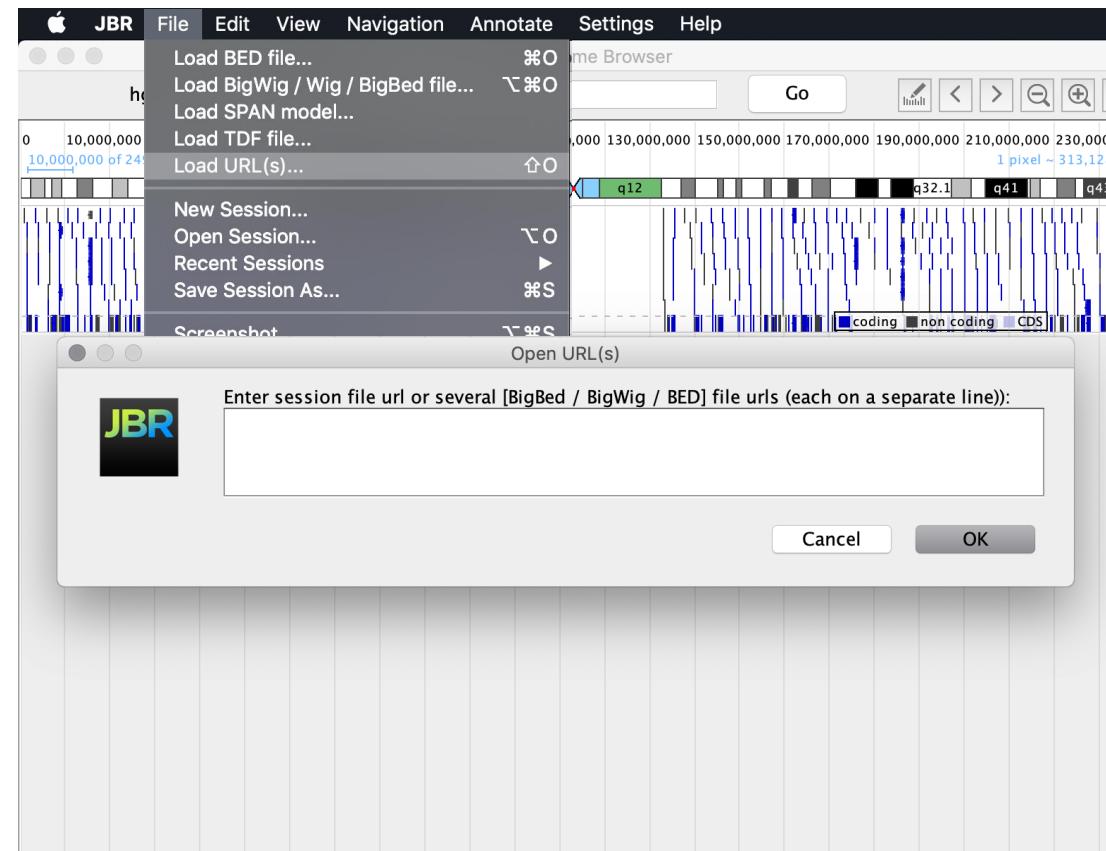


Let's Open JBR demo session to explore epigenetics data

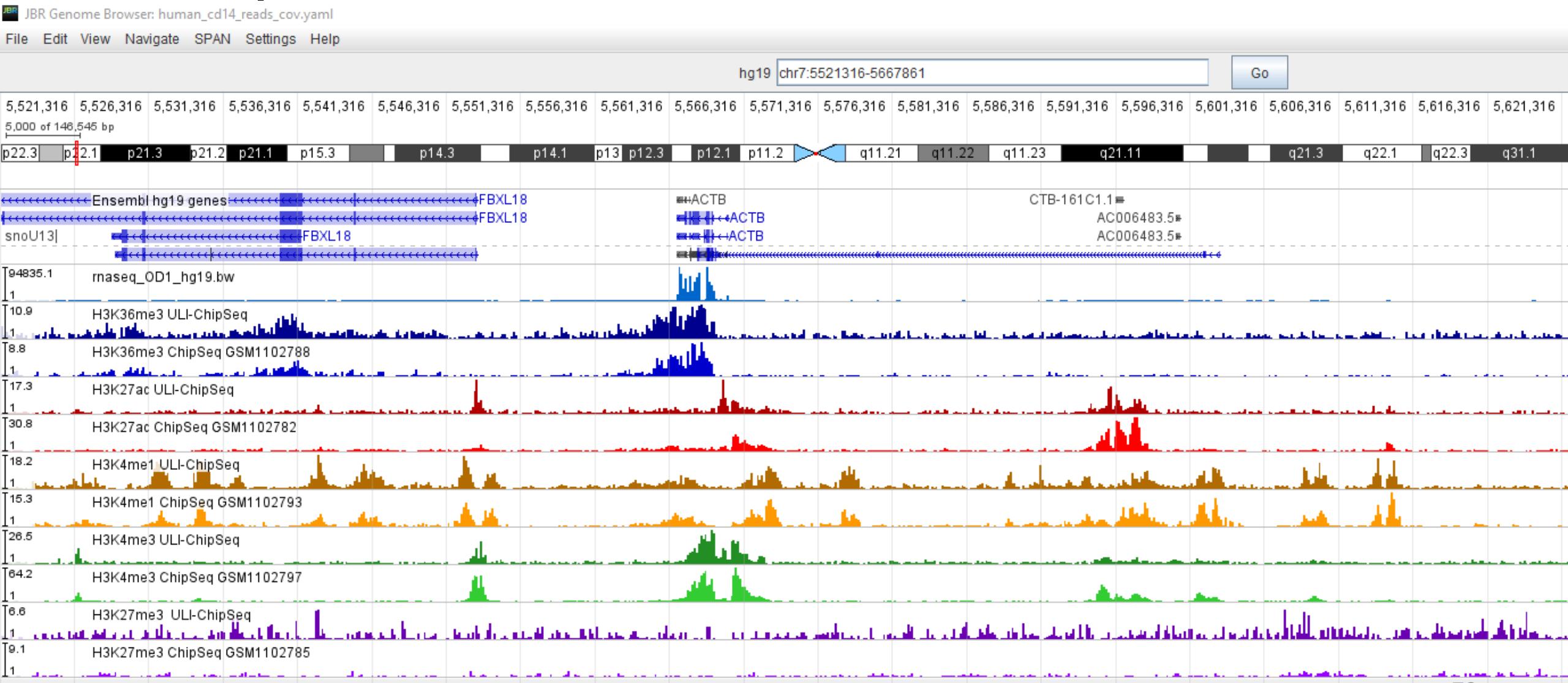
Explore Epigenetics Data Demo

Open session:

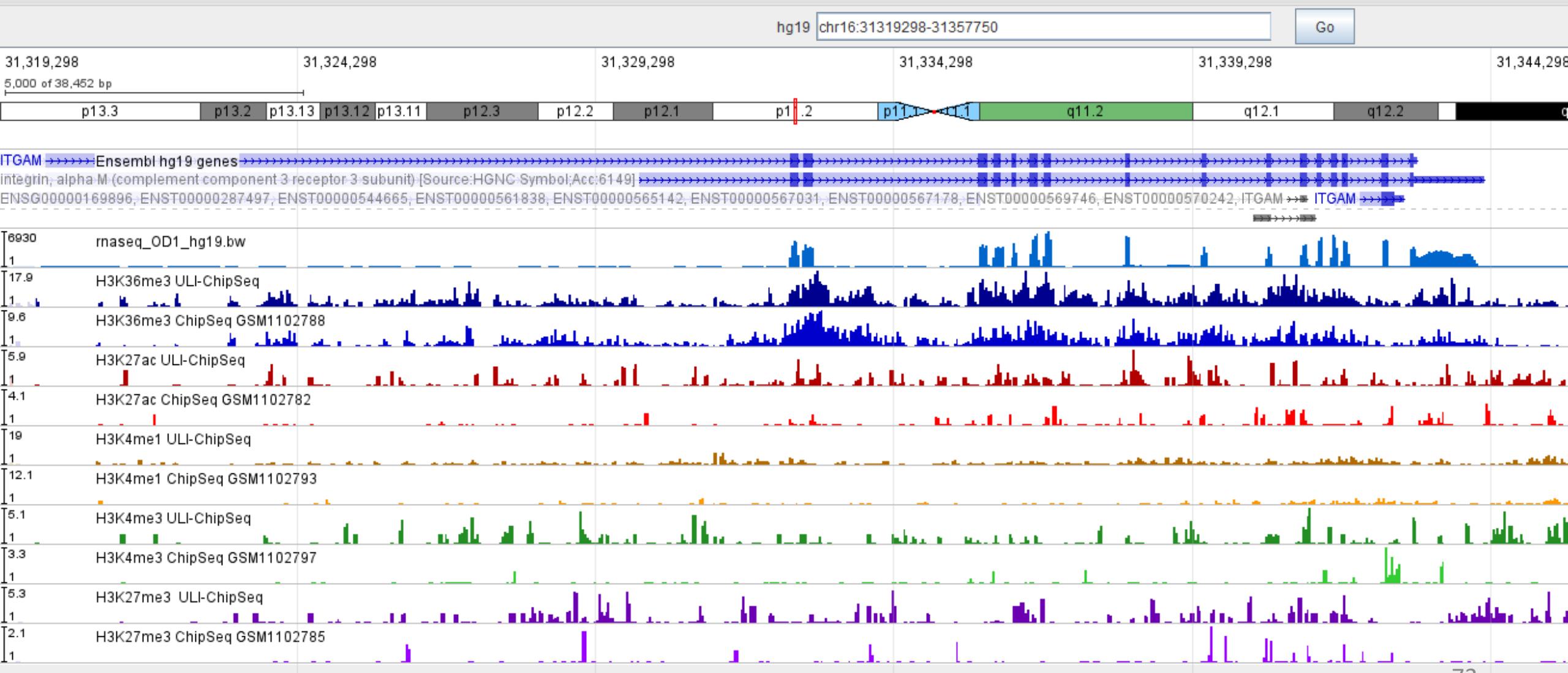
[Get human cd14 reads cov.yaml](#)



Tracks, I love tracks

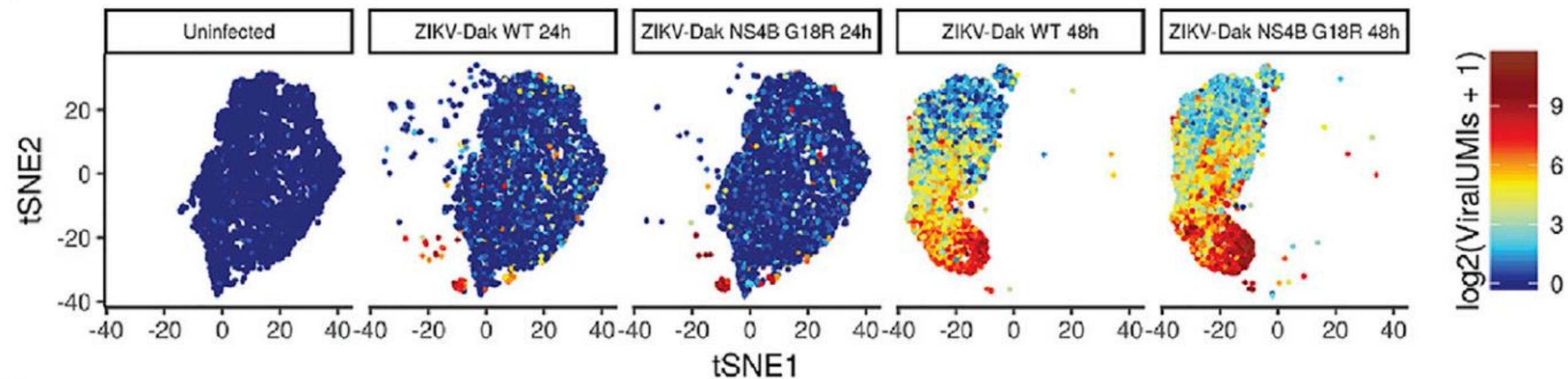


Let's have a look at ITGAM expression



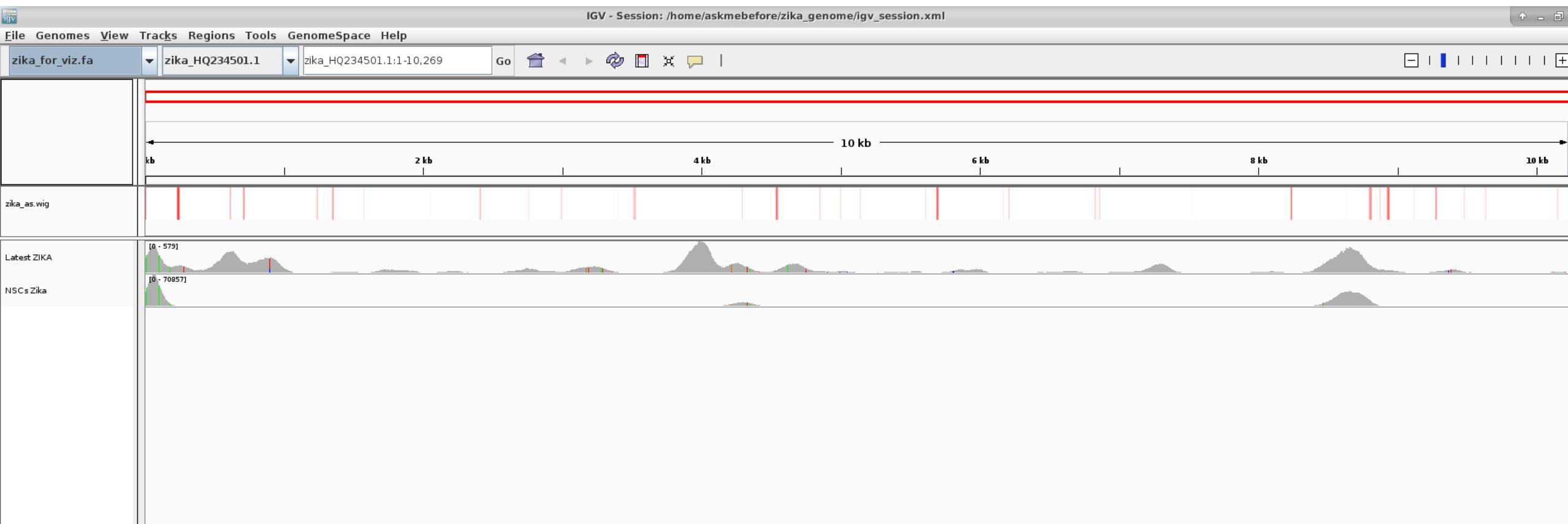
Zika story

- ✓ 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