Using phantasus application

Daria Zenkova

2017-08-01

This is an extended version of the project Morpheus. We are integrating widely used gene epression analysis methods from Bioconductor. You can use it in multiple ways, either locally, using `servePhantasus`-function from this package, this way would be described in this tutorial, either on web-site.

Loading required libraries

That package needs GEOquery as a dependency. Although, we recommend to install this package from it's forked version due to its better cache support:

```
devtools::install_github('assaron/GEOquery')
```

Running

To run this package use its exported function:

```
library(phantasus)
servePhantasus('0.0.0.0', 8000, cacheDir = file.path(getwd(), 'cache'))
```

Then, open (http://localhost:8000) at your browser.

Loading a dataset for analysis

There are two ways to upload a dataset into application:

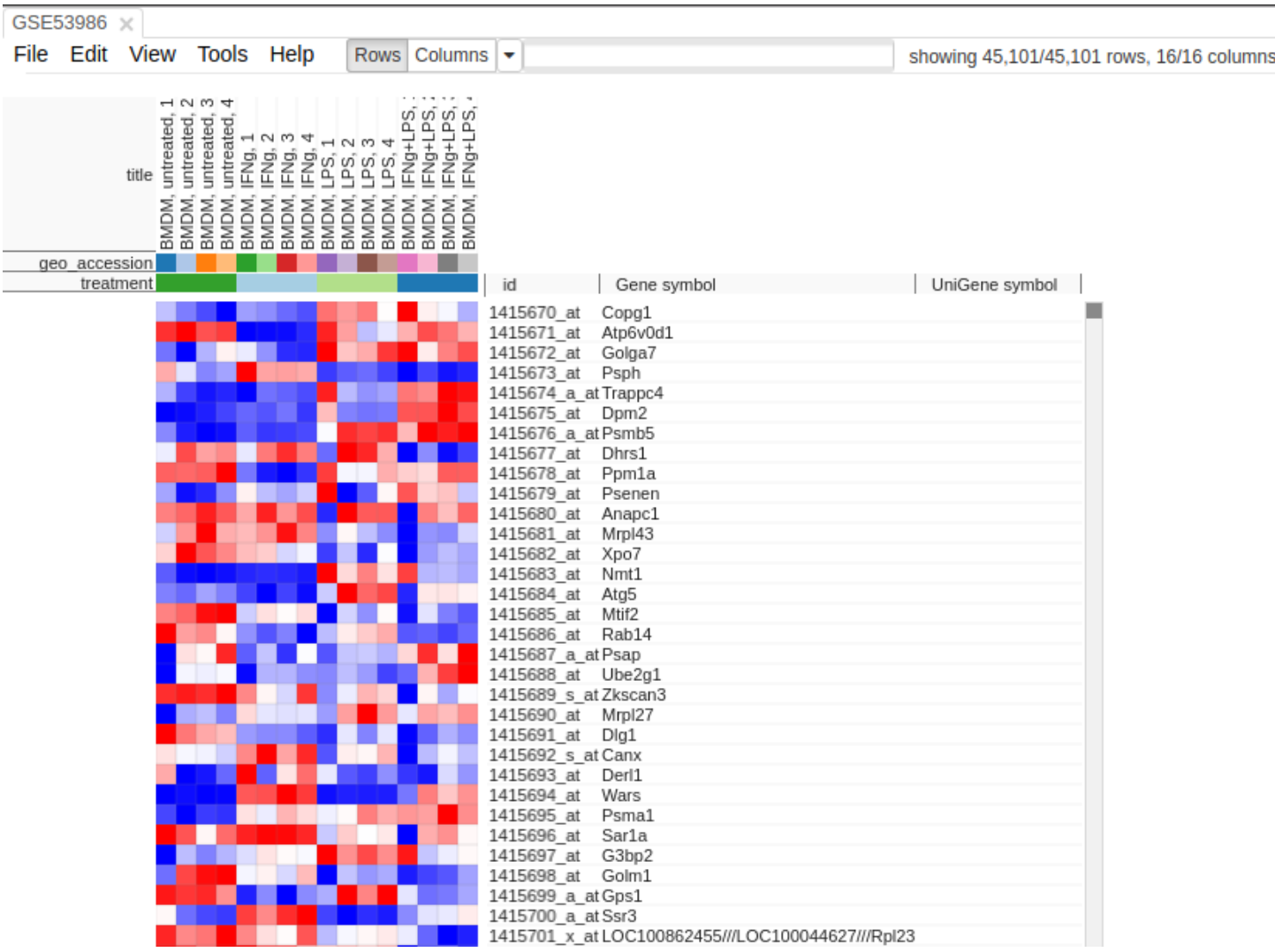
- As a file from
  - computer;
  - URL;
  - Dropbox;
- By GEO identifier.

Workflow example with GSE53986

Prepare the dataset for analysis

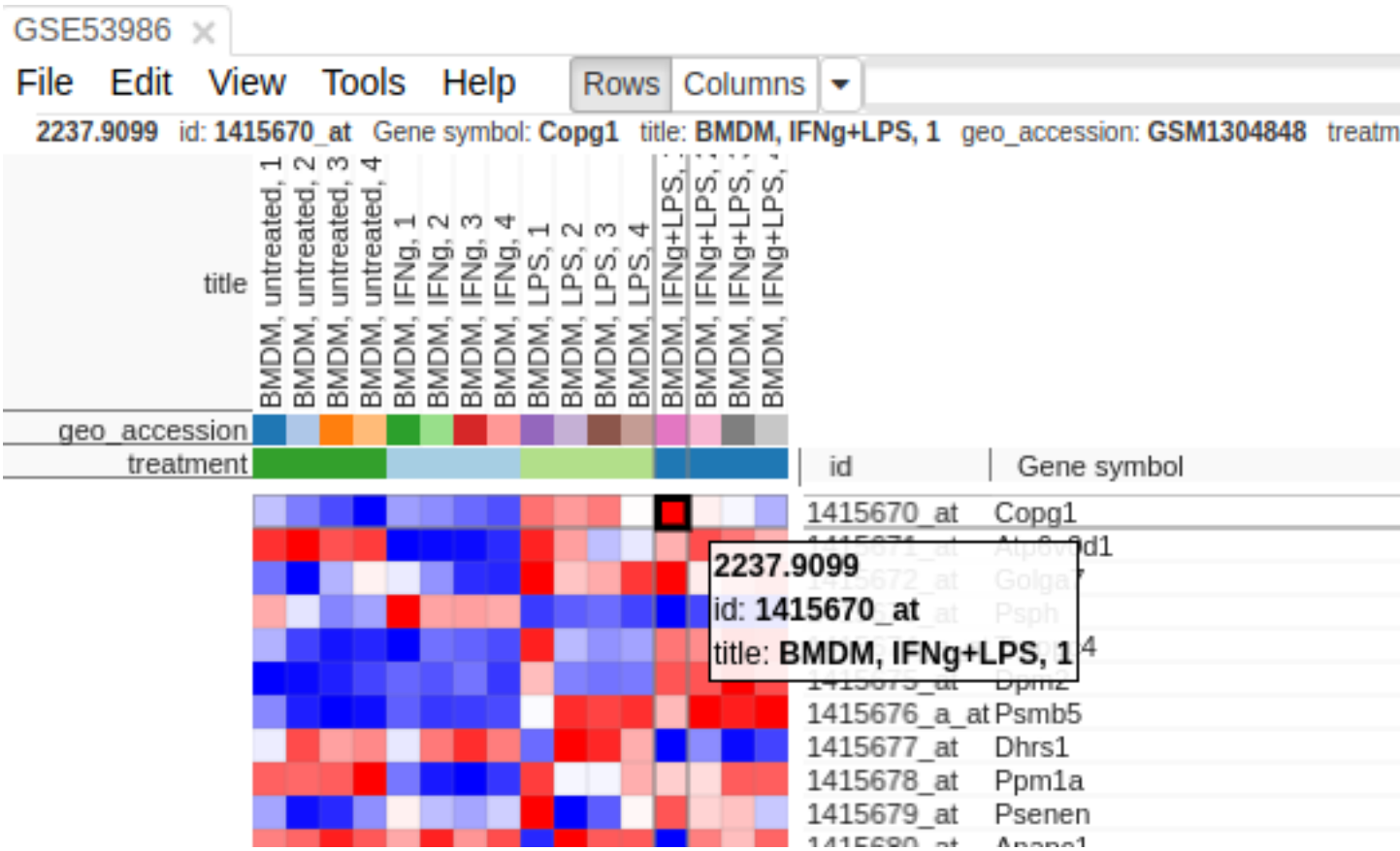
Open the dataset

Choose a loading option "GEO Datasets" and put "GSE53986" in the input field. After a few seconds, corresponding heatmap will be loaded.

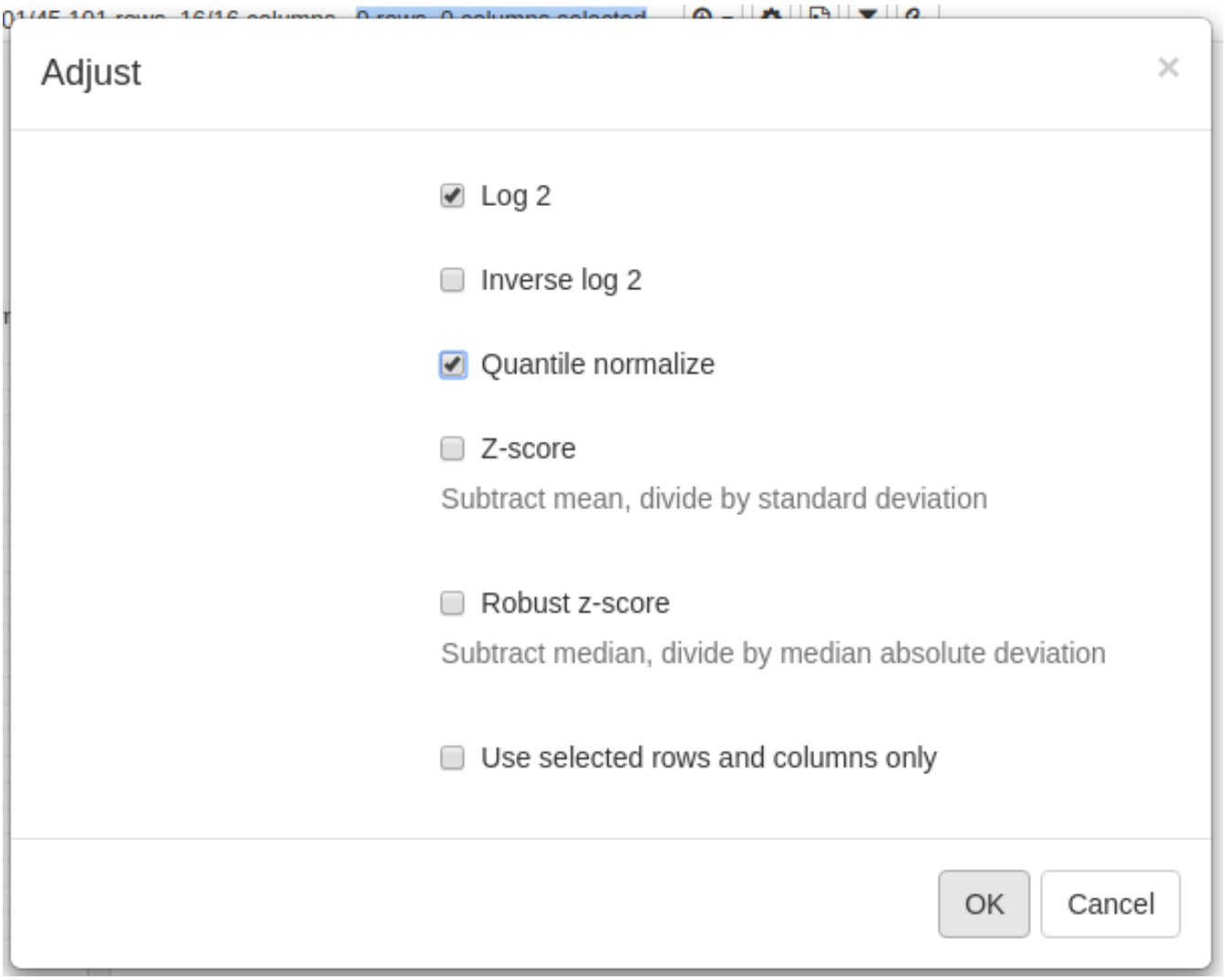


Adjust values

As you can see on the image, values are not scaled. So for the proper further analysis it is recommended to rescale the series matrix.

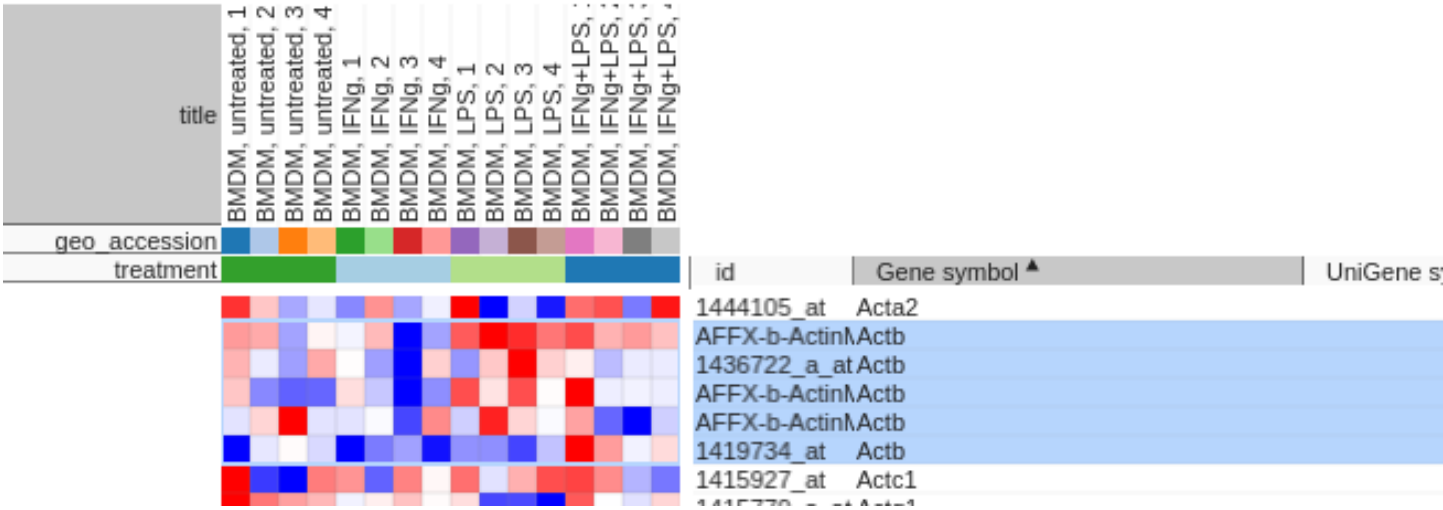


To adjust values go to Tools/Adjust and use Log2 and Quantile Normalization.

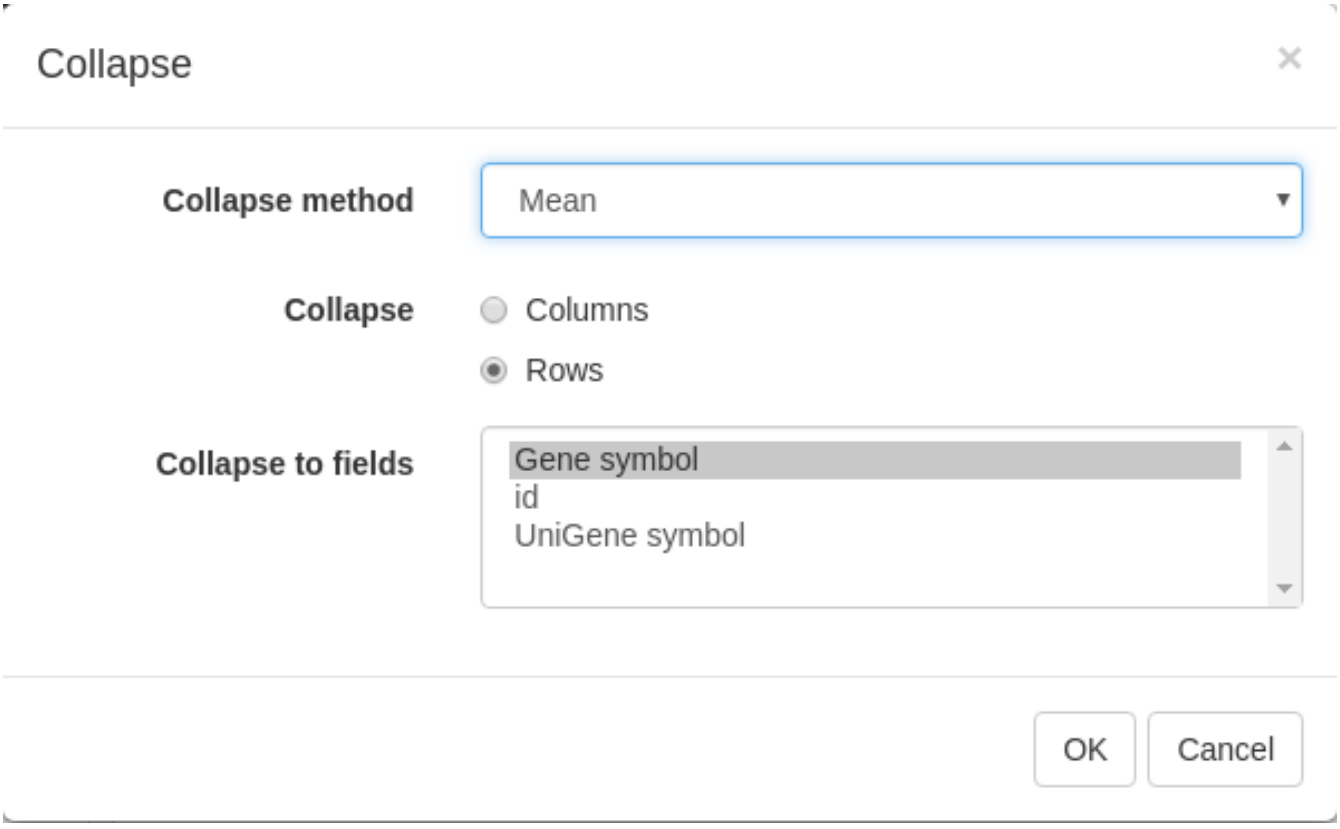


Remove duplicates

There can be duplicated genes, it is important to collapse their values.



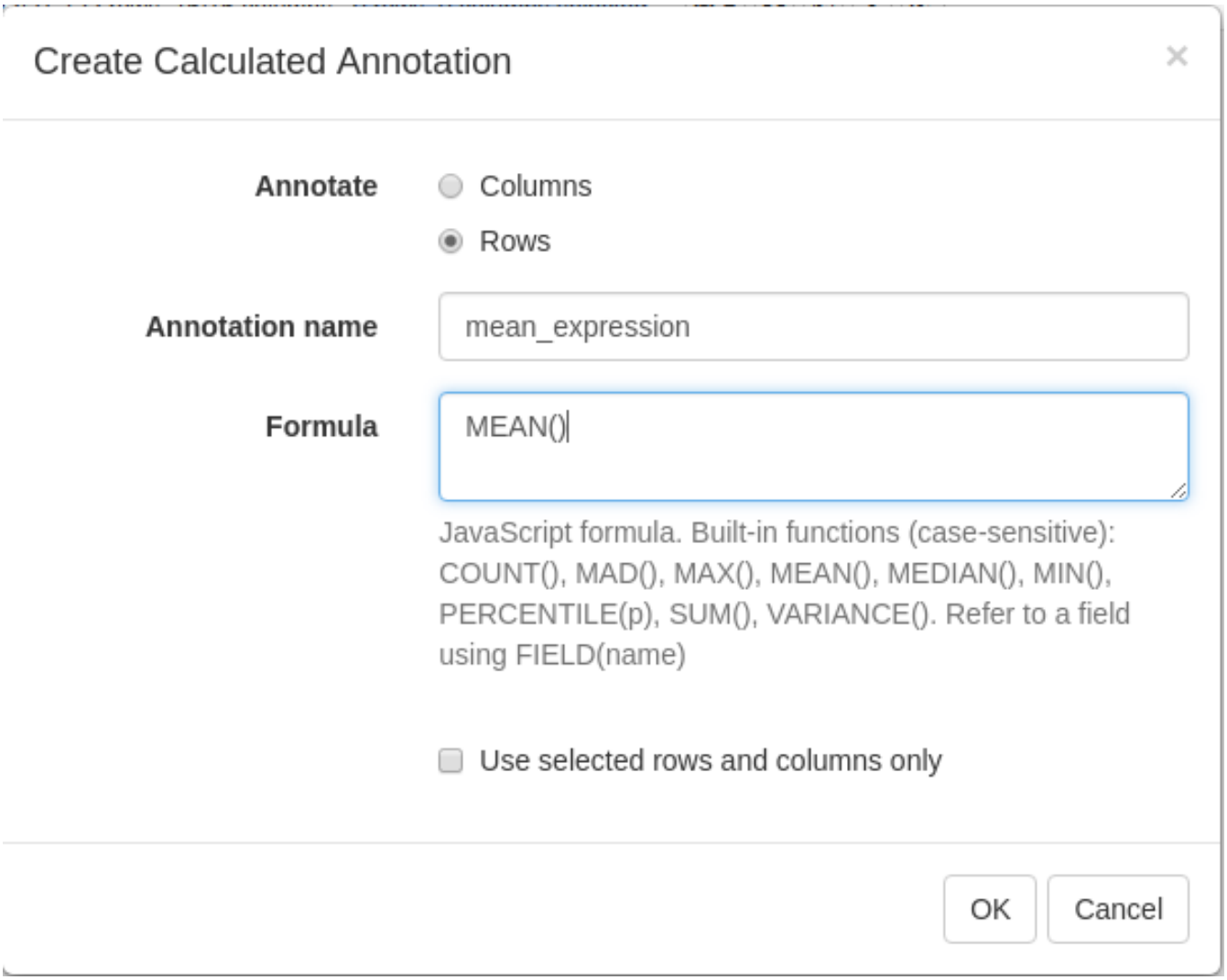
For that go to Tools/Collapse and choose "Mean" as the method and "Gene Symbol" as the collapse field.



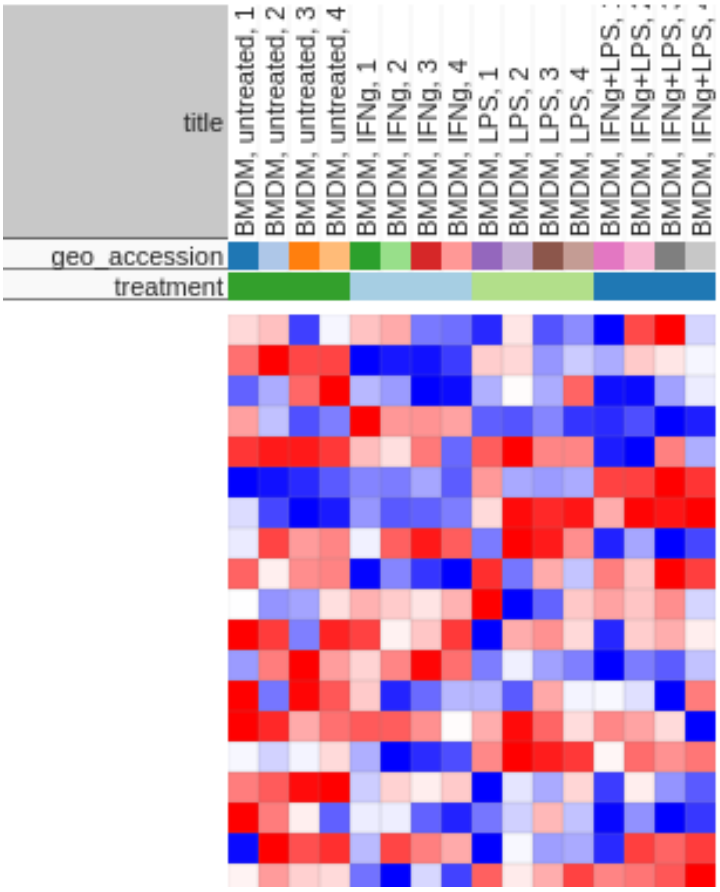
Filter lowly expressed genes

To calculate mean expression of each gene go to Tools/Create Calculated Annotation.

Put there annotation name and formula for calculation.

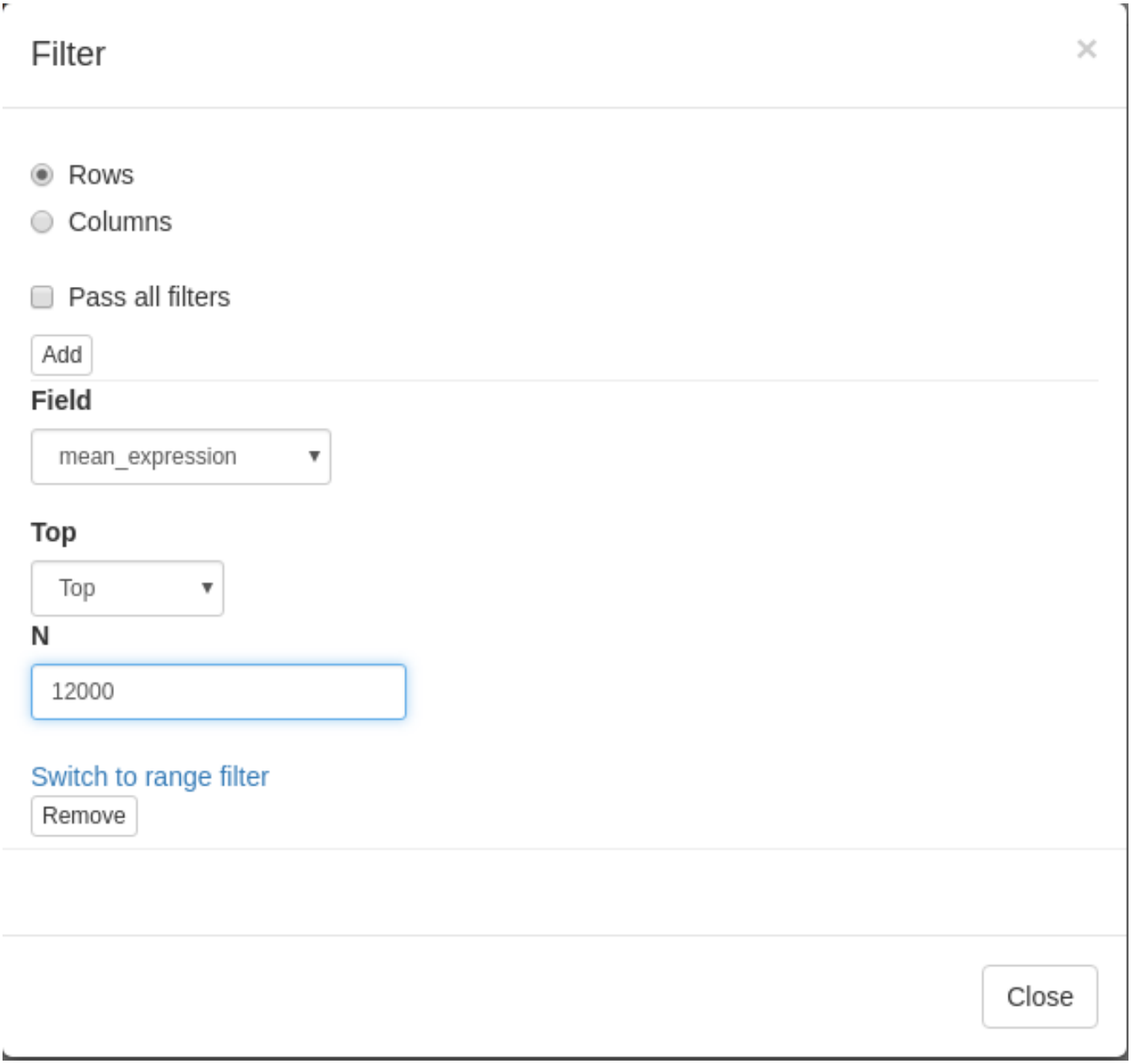


The result should look like this, now you can use this annotation to sort genes by:



Gene symbol	mean_expression
Copg1	7.5543
Atp6v0d1	13.1083
Golga7	13.2429
Psph	8.796
Trappc4	8.3935
Dpm2	11.0475
Psmb5	12.823
Dhrs1	10.675
Ppm1a	7.6373
Psenen	13.4482
Anapc1	7.8757
Mrpl43	11.238
Xpo7	5.1941
Nmt1	9.3109
Atg5	10.1685
Mtif2	10.4253
Rab14	9.7619
Psap	14.7023
Ube2g1	10.3002

To filter genes go to Tools/Filter, press “Add”, choose “mean_expression” as the field, then press to “Switch to top filter”, and input the number of genes you want to proceed with.



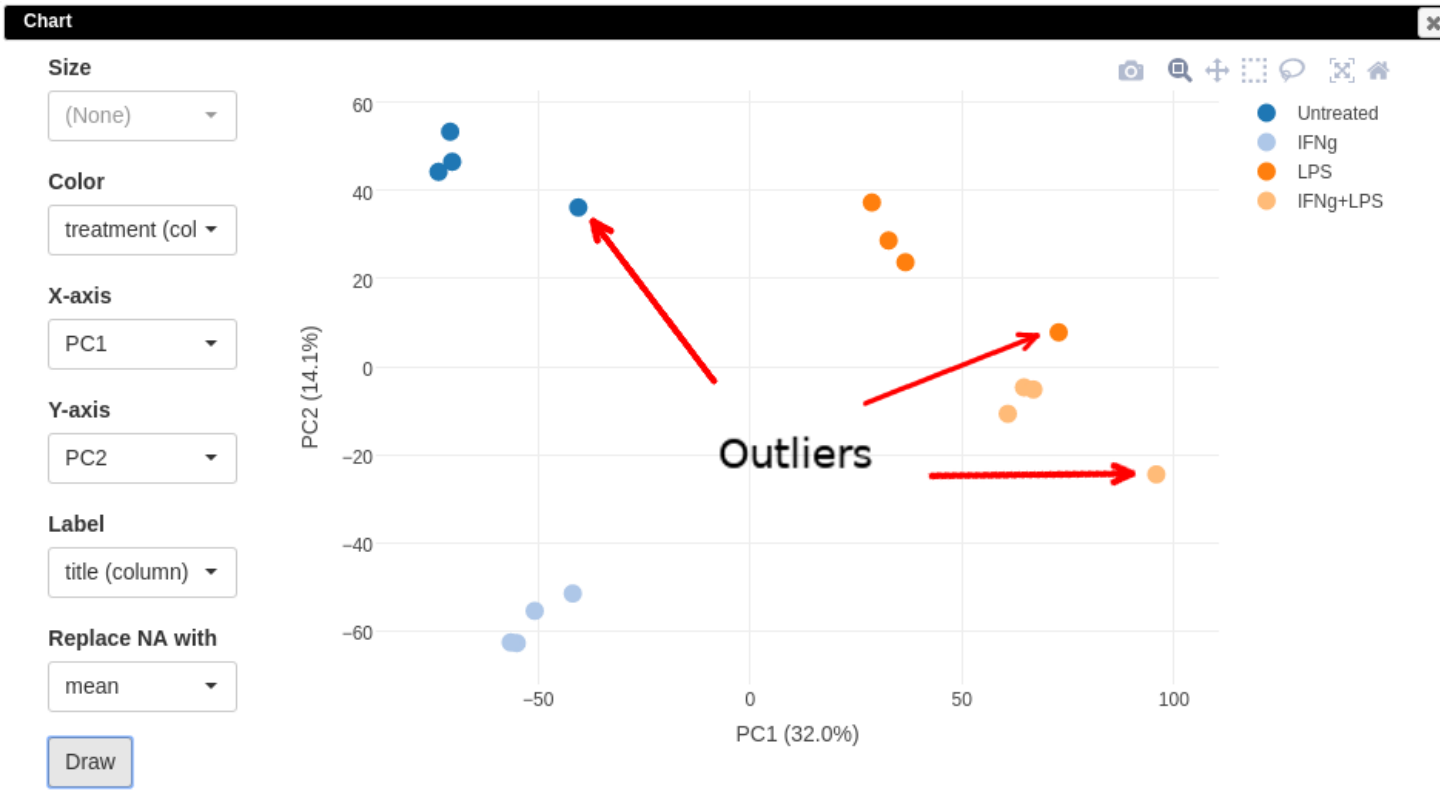
Then you can close tool, the added filter will be still active. Although it is more convenient to extract this genes to a new dataset, so select all genes (click on any gene and press Ctrl+A) and use Tools/New Heat Map (Ctrl+X).

Explore the dataset

PCA Plot

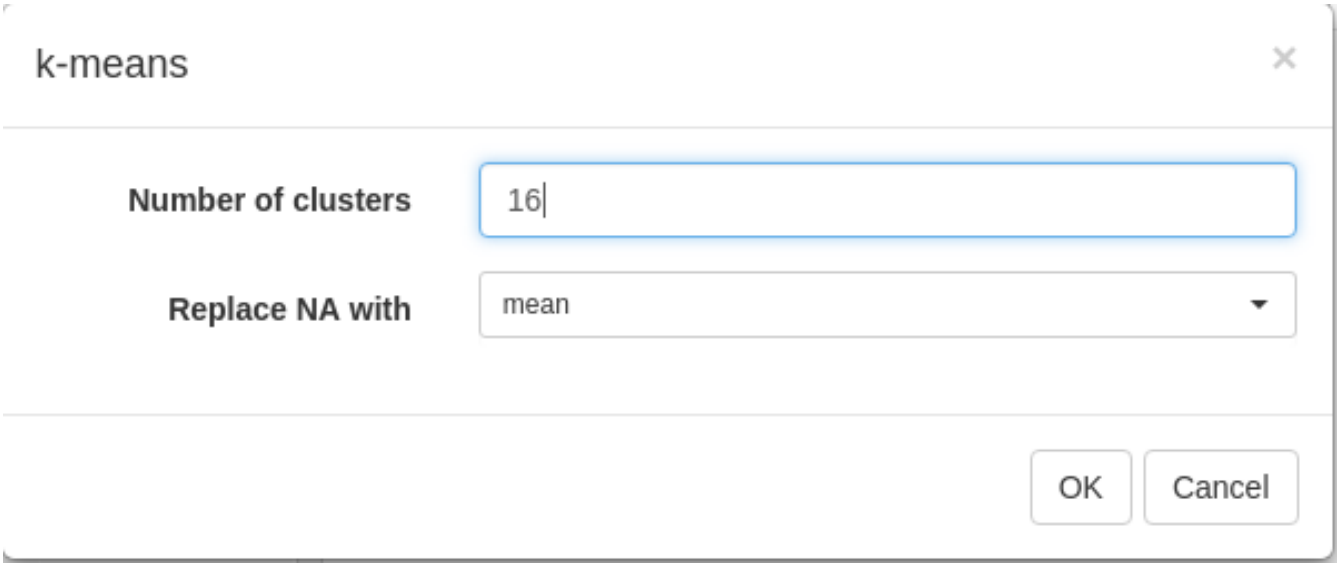
Use Tools/PCA Plot to analyse the data with Principal Component Analysis method.

You can customize color, size and label of points on the chart by values in annotation. By analysing the plot you may find some outliers that need to be analysed further to see if they can be eliminated.

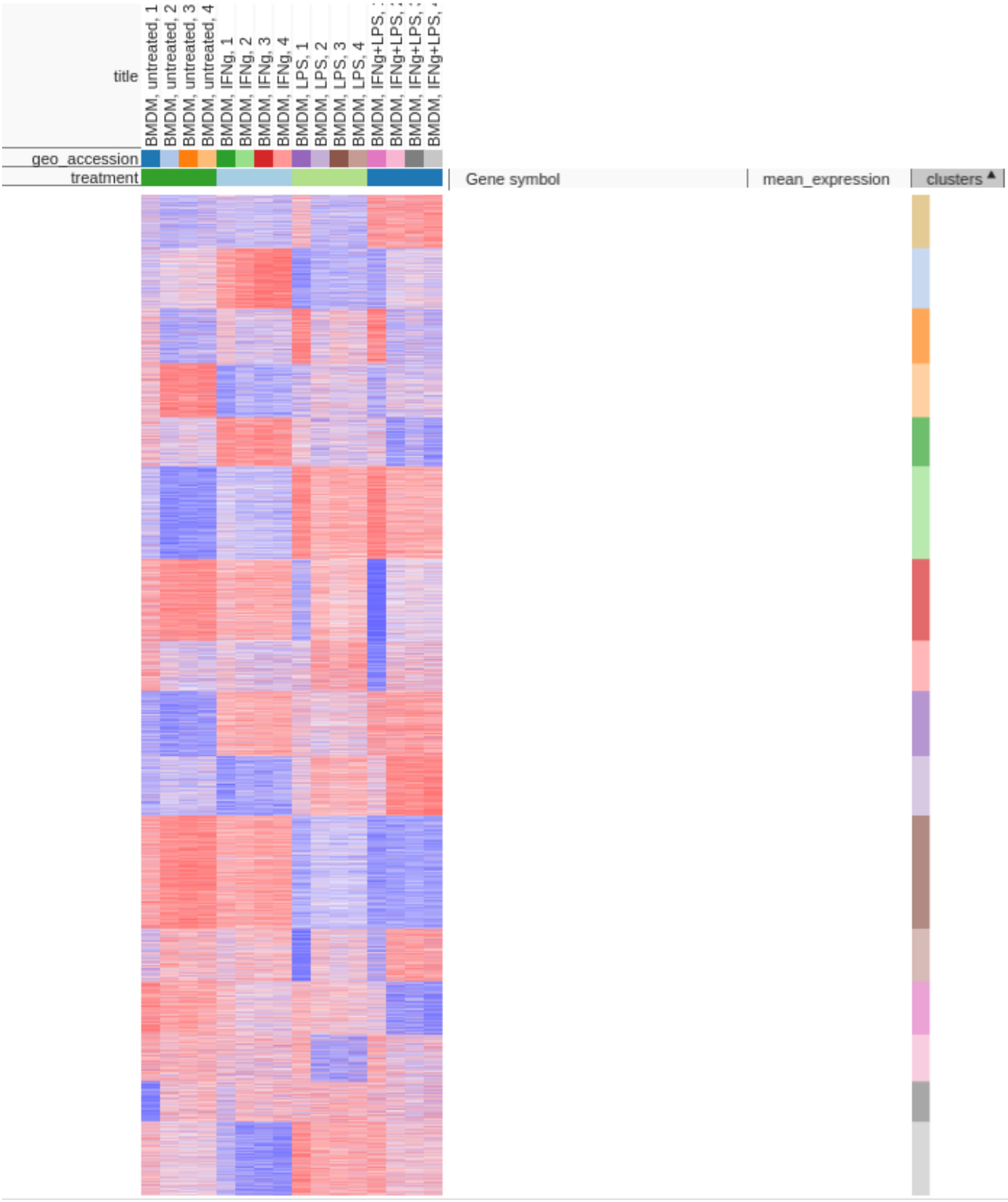


K-means clustering

Use Tools/k-means to cluster genes into predefined number of clusters.

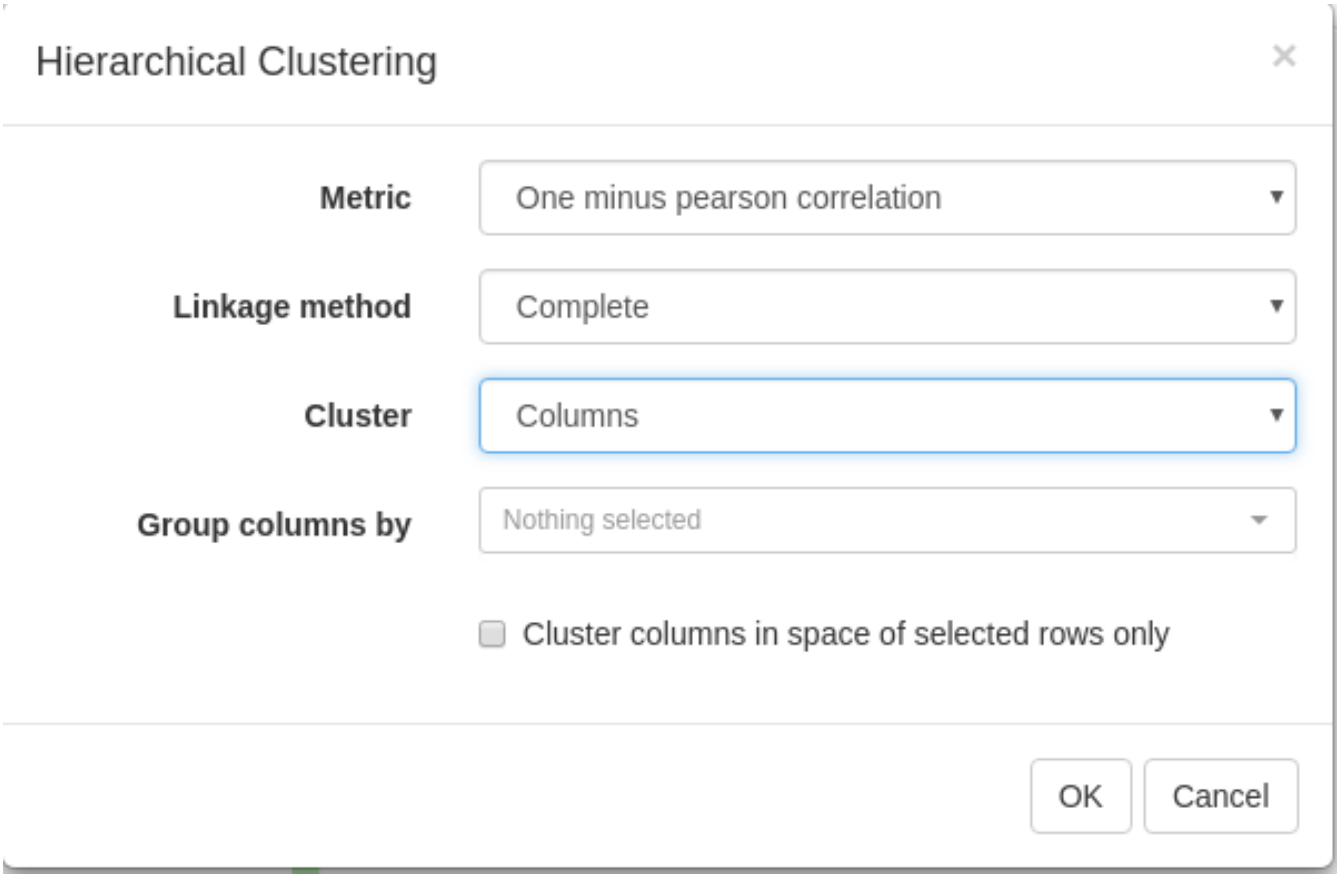


Then you can sort by annotation "clusters" and view the whole dataset by using View/Fit to window. Here you can also see outlier samples.


title
BMDM, untreated, 1
BMDM, untreated, 2
BMDM, untreated, 3
BMDM, untreated, 4
BMDM, IFNg, 1
BMDM, IFNg, 2
BMDM, IFNg, 3
BMDM, IFNg, 4
BMDM, LPS, 1
BMDM, LPS, 2
BMDM, LPS, 3
BMDM, LPS, 4
BMDM, IFNg+LPS,
BMDM, IFNg+LPS,
BMDM, IFNg+LPS,
BMDM, IFNg+LPS,
geo_accession
treatment
Gene symbol
mean_expression
clusters

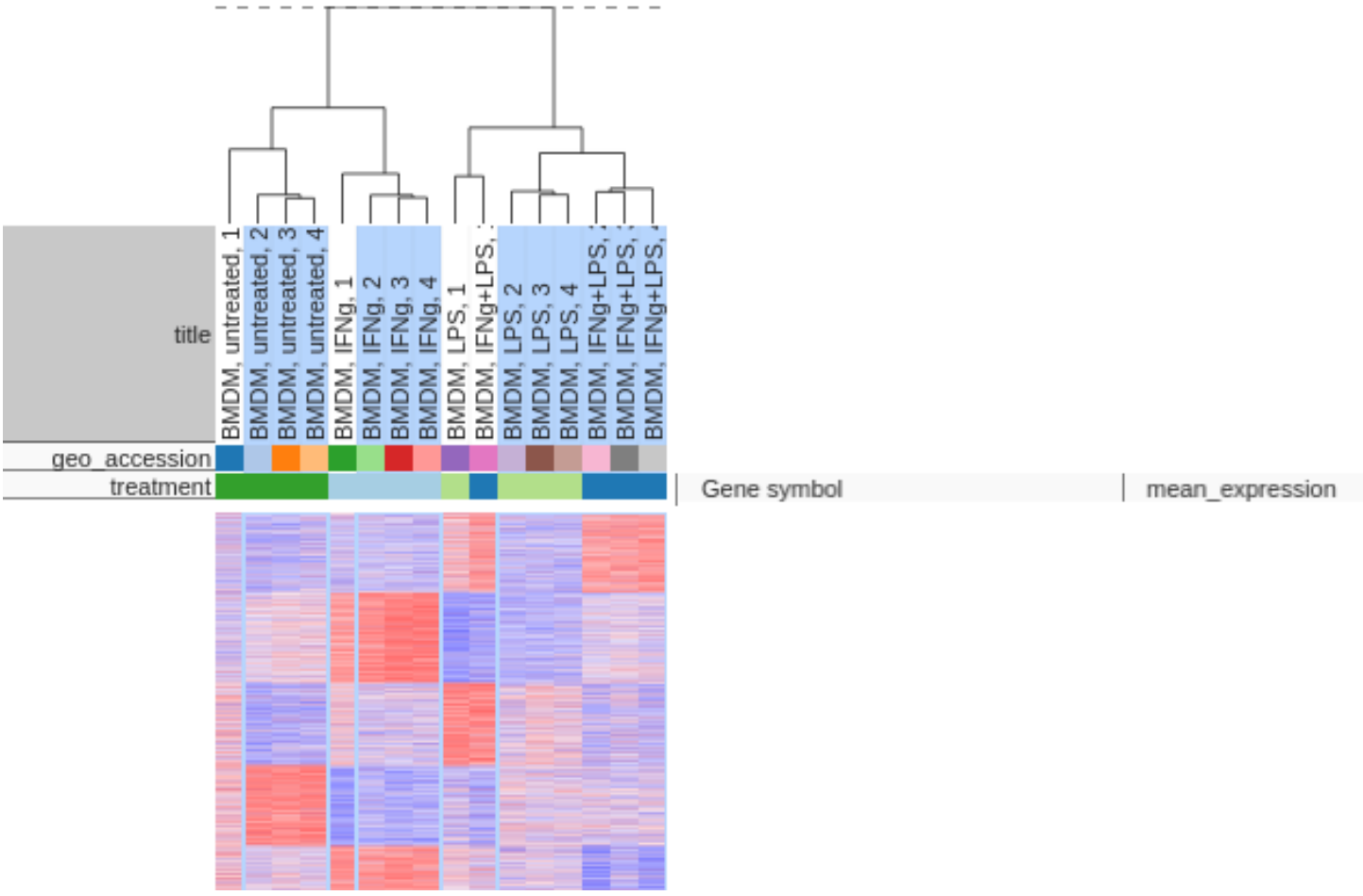
Hierarchical clustering

Use Tool/Hierarchical clustering to cluster samples.



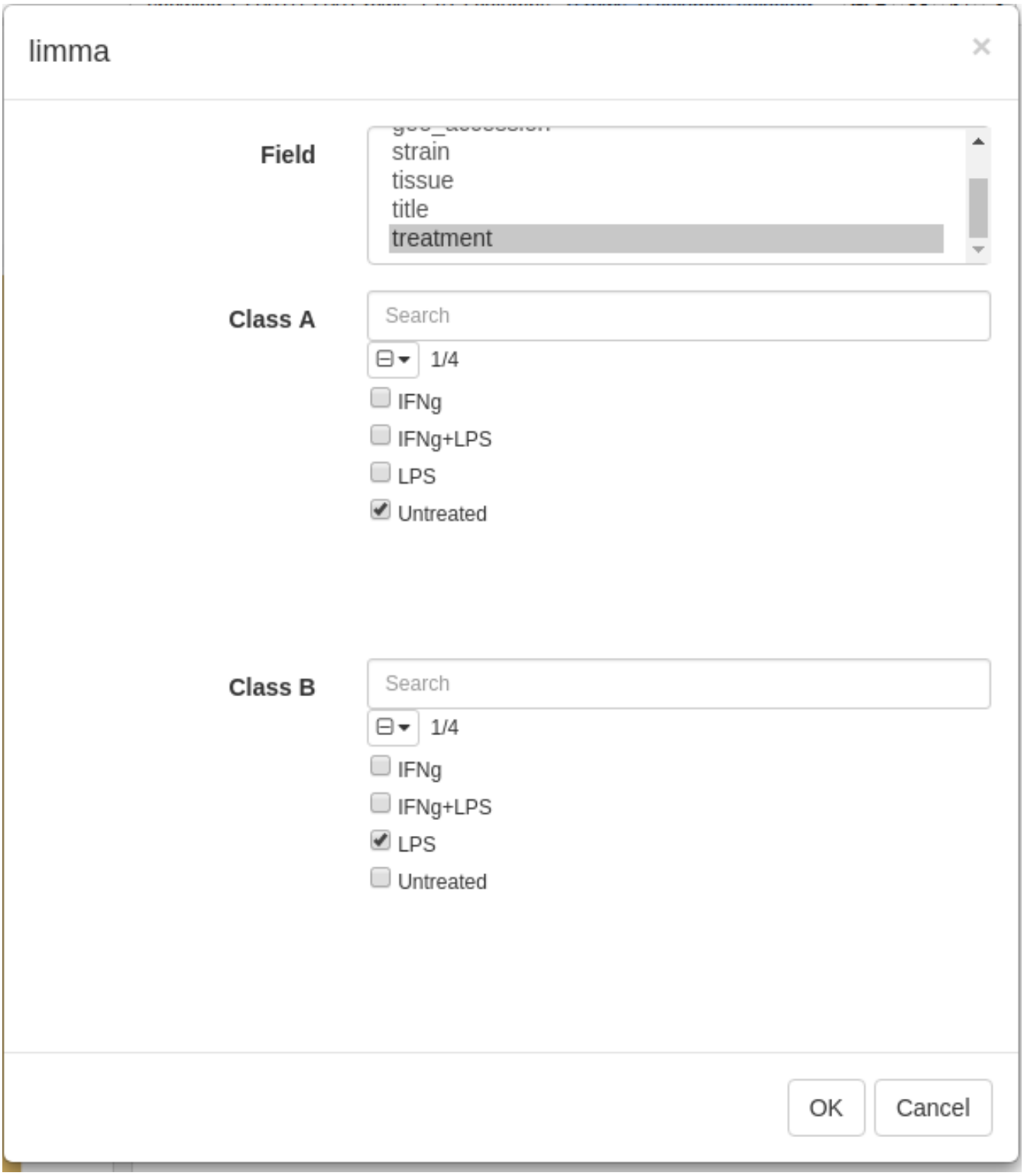
Filtering outliers

Now, when outliers are confirmed, you can choose good samples and extract them into another heatmap (Ctrl+X).

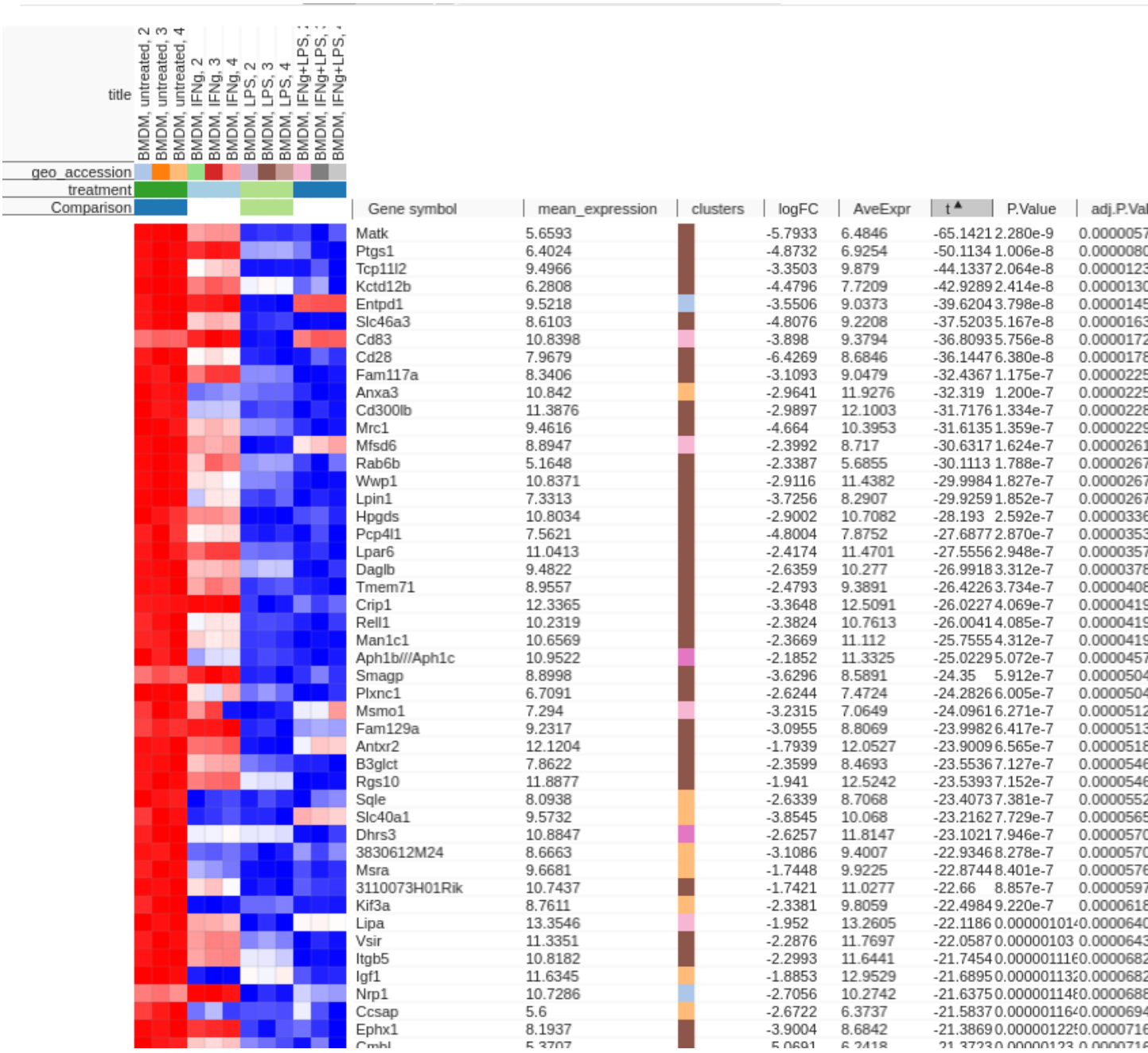


Differential expression

Use Tools/Limma to compare samples. Choose “treatment” as a field, “Untreated” and “LPS” as classes.



Now you can sort by t-statistic and see genes with the highest difference between classes.



Gene symbol	mean_expression	clusters	logFC	AveExpr	t	P.Value	adj.P.Val
Matk	5.6593		-5.7933	6.4846	-65.1421	2.280e-9	0.0000057
Ptgs1	6.4024		-4.8732	6.9254	-50.1134	1.006e-8	0.0000080
Tcp11l2	9.4966		-3.3503	9.879	-44.1337	2.064e-8	0.0000123
Kctd12b	6.2808		-4.4796	7.7209	-42.9289	2.414e-8	0.0000130
Entpd1	9.5218		-3.5506	9.0373	-39.6204	3.798e-8	0.0000145
Slc46a3	8.6103		-4.8076	9.2208	-37.5203	5.167e-8	0.0000163
Cd83	10.8398		-3.898	9.3794	-36.8093	5.756e-8	0.0000172
Cd28	7.9679		-6.4269	8.6846	-36.1447	6.380e-8	0.0000178
Fam117a	8.3406		-3.1093	9.0479	-32.4367	1.175e-7	0.0000225
Anxa3	10.842		-2.9641	11.9276	-32.319	1.200e-7	0.0000225
Cd300lb	11.3876		-2.9897	12.1003	-31.7176	1.334e-7	0.0000228
Mrc1	9.4616		-4.664	10.3953	-31.6135	1.359e-7	0.0000229
Mfsd6	8.8947		-2.3992	8.717	-30.6317	1.624e-7	0.0000261
Rab6b	5.1648		-2.3387	5.6855	-30.1113	1.788e-7	0.0000267
Wwp1	10.8371		-2.9116	11.4382	-29.9984	1.827e-7	0.0000267
Lpin1	7.3313		-3.7256	8.2907	-29.9259	1.852e-7	0.0000267
Hpgds	10.8034		-2.9002	10.7082	-28.193	2.592e-7	0.0000336
Pcp4l1	7.5621		-4.8004	7.8752	-27.6877	2.870e-7	0.0000353
Lpar6	11.0413		-2.4174	11.4701	-27.5556	2.948e-7	0.0000357
Daglb	9.4822		-2.6359	10.277	-26.9918	3.312e-7	0.0000378
Tmem71	8.9557		-2.4793	9.3891	-26.4226	3.734e-7	0.0000408
Crip1	12.3365		-3.3648	12.5091	-26.0227	4.069e-7	0.0000419
Rell1	10.2319		-2.3824	10.7613	-26.0041	4.085e-7	0.0000419
Man1c1	10.6569		-2.3669	11.112	-25.7555	4.312e-7	0.0000419
Aph1b///Aph1c	10.9522		-2.1852	11.3325	-25.0229	5.072e-7	0.0000457
Smagp	8.8998		-3.6296	8.5891	-24.35	5.912e-7	0.0000504
Plxnc1	6.7091		-2.6244	7.4724	-24.2826	6.005e-7	0.0000504
Msmo1	7.294		-3.2315	7.0649	-24.0961	6.271e-7	0.0000512
Fam129a	9.2317		-3.0955	8.8069	-23.9982	6.417e-7	0.0000513
Antxr2	12.1204		-1.7939	12.0527	-23.9009	6.565e-7	0.0000518
B3glct	7.8622		-2.3599	8.4693	-23.5536	7.127e-7	0.0000546
Rgs10	11.8877		-1.941	12.5242	-23.5393	7.152e-7	0.0000546
Sqle	8.0938		-2.6339	8.7068	-23.4073	7.381e-7	0.0000552
Slc40a1	9.5732		-3.8545	10.068	-23.2162	7.729e-7	0.0000565
Dhrs3	10.8847		-2.6257	11.8147	-23.1021	7.946e-7	0.0000570
3830612M24	8.6663		-3.1086	9.4007	-22.9346	8.278e-7	0.0000570
Msra	9.6681		-1.7448	9.9225	-22.8744	8.401e-7	0.0000576
3110073H01Rik	10.7437		-1.7421	11.0277	-22.66	8.857e-7	0.0000597
Kif3a	8.7611		-2.3381	9.8059	-22.4984	9.220e-7	0.0000618
Lipa	13.3546		-1.952	13.2605	-22.1186	0.00000101	0.0000640
Vsir	11.3351		-2.2876	11.7697	-22.0587	0.00000103	0.0000643
Itgb5	10.8182		-2.2993	11.6441	-21.7454	0.00000111	0.0000682
Igf1	11.6345		-1.8853	12.9529	-21.6895	0.00000113	0.0000682
Nrp1	10.7286		-2.7056	10.2742	-21.6375	0.00000114	0.0000688
Ccsap	5.6		-2.6722	6.3737	-21.5837	0.00000116	0.0000694
Ephx1	8.1937		-3.9004	8.6842	-21.3869	0.00000122	0.0000716
Cmbl	5.3707		5.0691	6.2418	21.3723	0.00000123	0.0000716

Feedback

You can see known issues and place yours here: (https://github.com/ctlab/phantasus/issues)