

Using phantasus application

Daria Zenkova

2017-09-13

This is an extended version of the project Morpheus. We are integrating widely used gene expression 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
 - with “Open file” interface;
 - with adding parameter “geo” to the link (e.g. <http://localhost:8000/?geo=GSE27112>).

Workflow example with GSE53986

Prepare the dataset for analysis

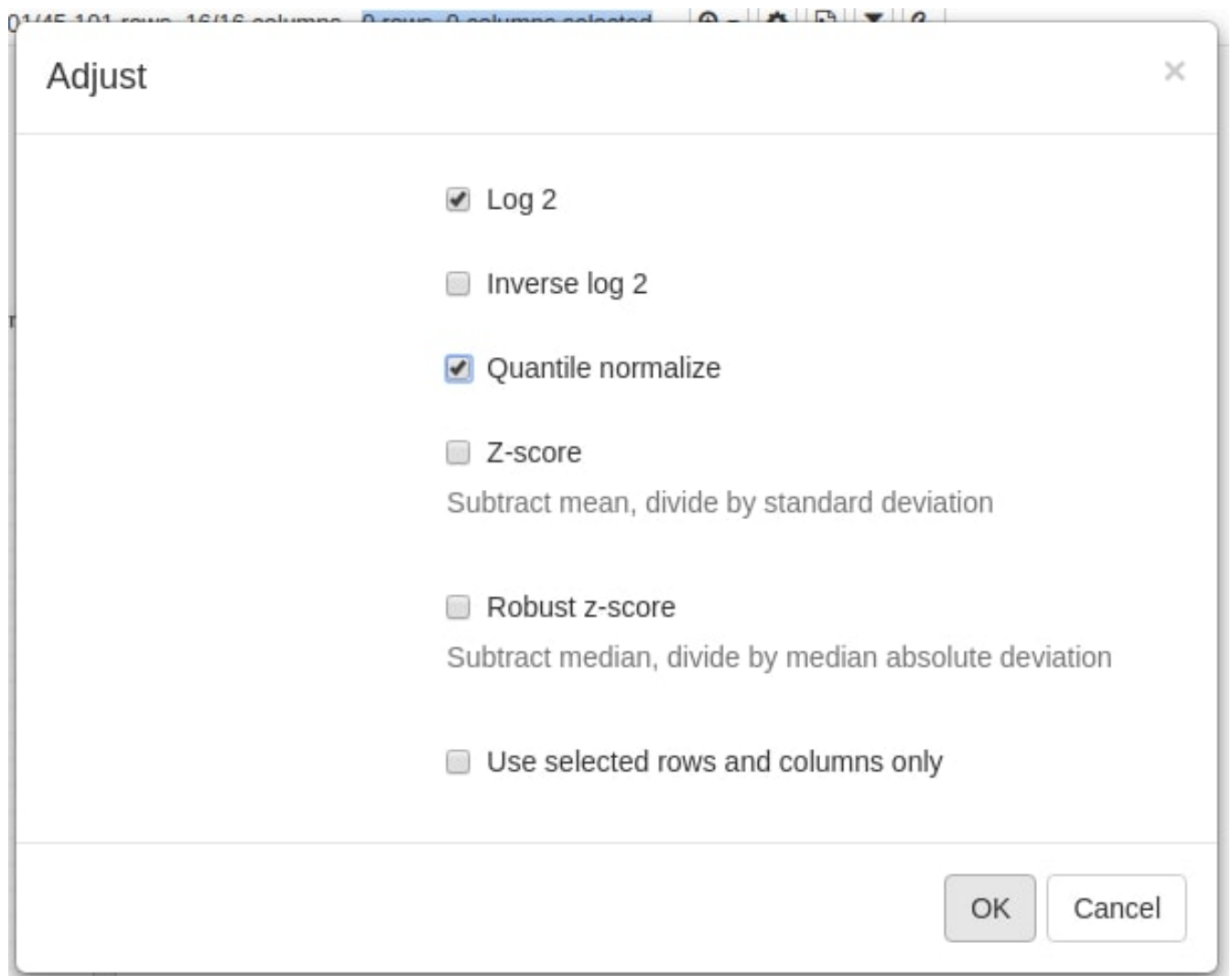
Open the dataset

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



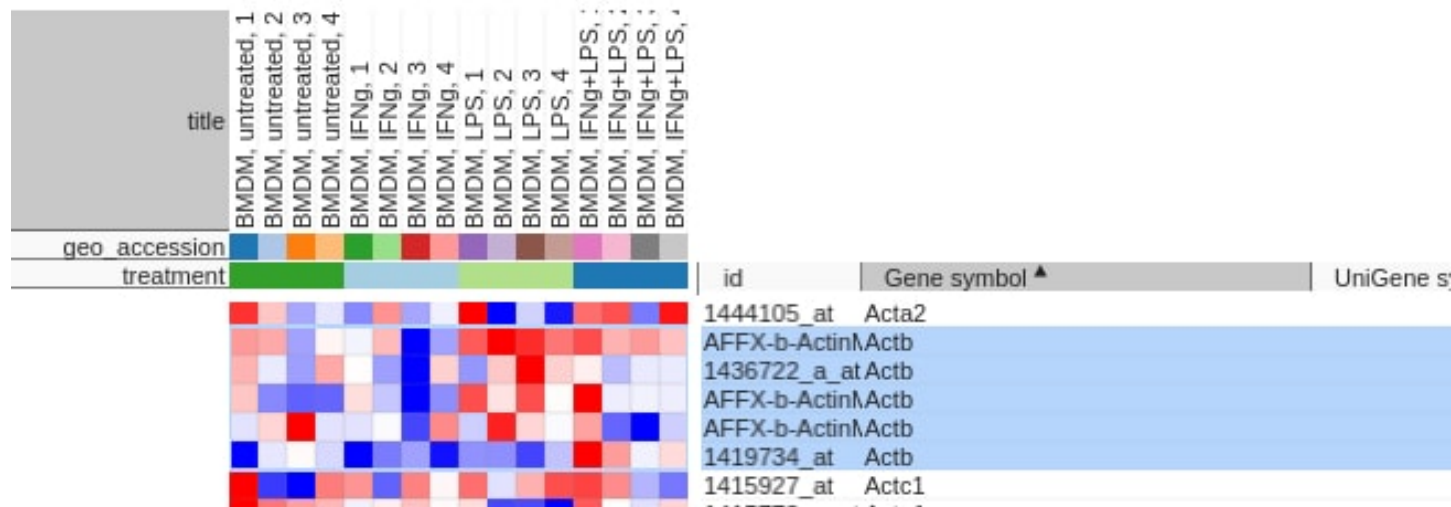
Adjust values

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.



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.

Collapse

Collapse method

Mean

Collapse

☐ Columns
 ☒ Rows

Collapse to fields

Gene symbol
 id
 UniGene symbol

OK

Cancel

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.

Create Calculated Annotation

Annotate

☐ Columns

☒ Rows

Annotation name

mean_expression

Formula

MEAN()

JavaScript formula. Built-in functions (case-sensitive):
COUNT(), MAD(), MAX(), MEAN(), MEDIAN(), MIN(),
PERCENTILE(p), SUM(), VARIANCE(). Refer to a field
using FIELD(name)

☐ Use selected rows and columns only

OK

Cancel

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

Filter

Rows

Columns

Pass all filters

Add

Field

mean_expression

Top

Top

N

12000

Switch to range filter

Remove

Close

Chart

Size

(None)

Color

(None)

X-axis

PC1

Y-axis

PC2

Label

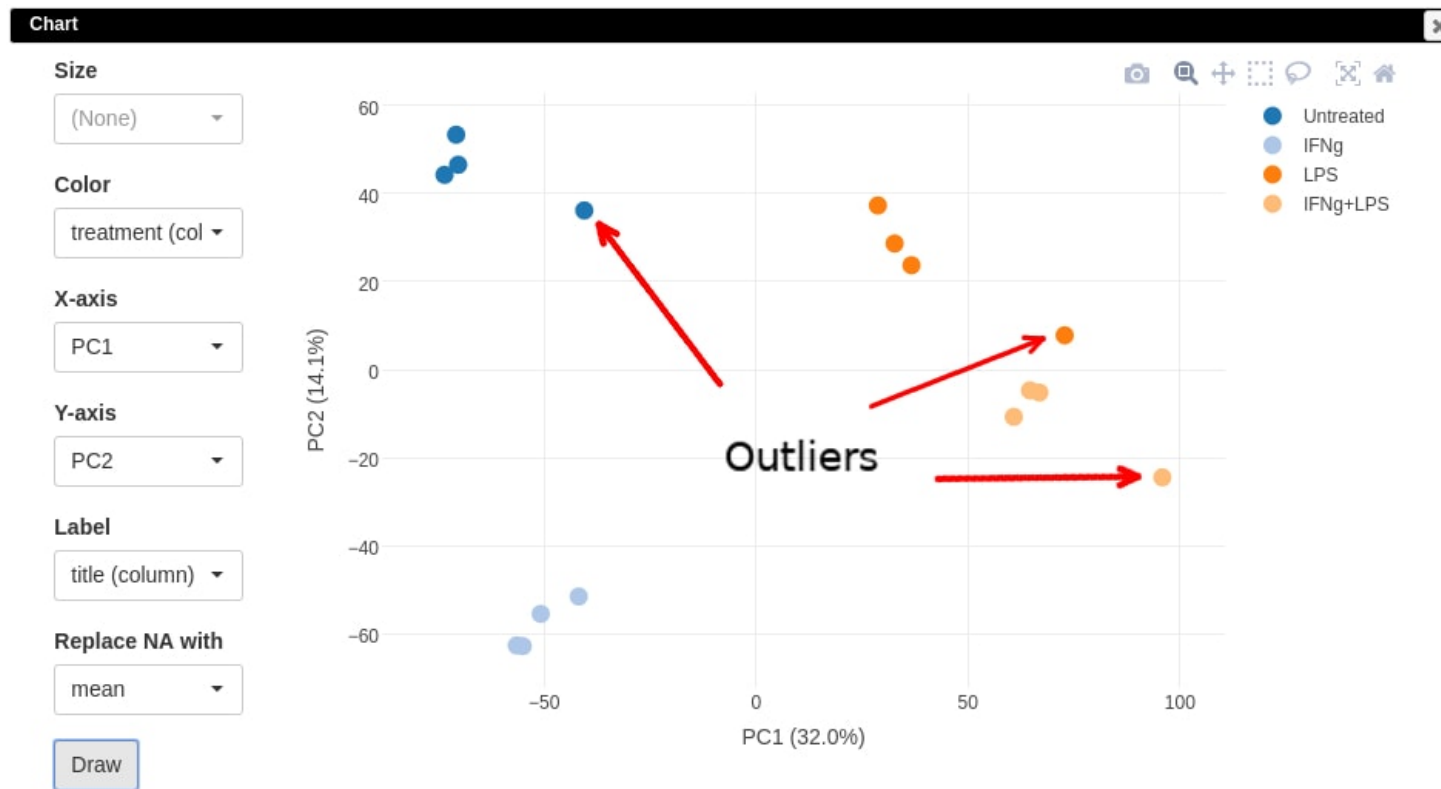
(None)

Replace NA with

mean

Draw

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.

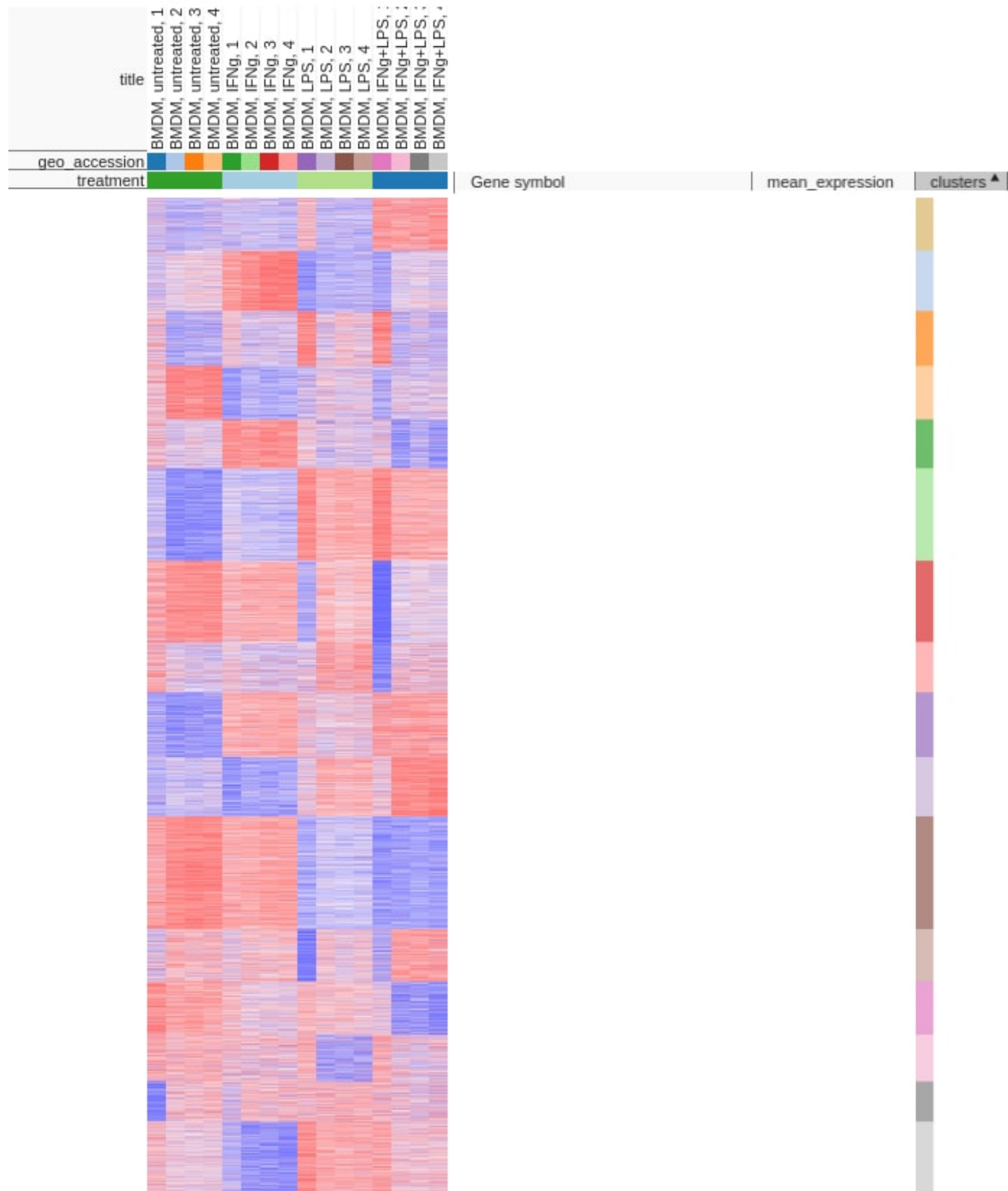
k-means

Number of clusters: 16

Replace NA with: mean

OK Cancel

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.



Hierarchical clustering

Use Tool/Hierarchical clustering to cluster samples.

Hierarchical Clustering

MetricOne minus pearson correlation▼

Linkage methodComplete▼

ClusterColumns▼

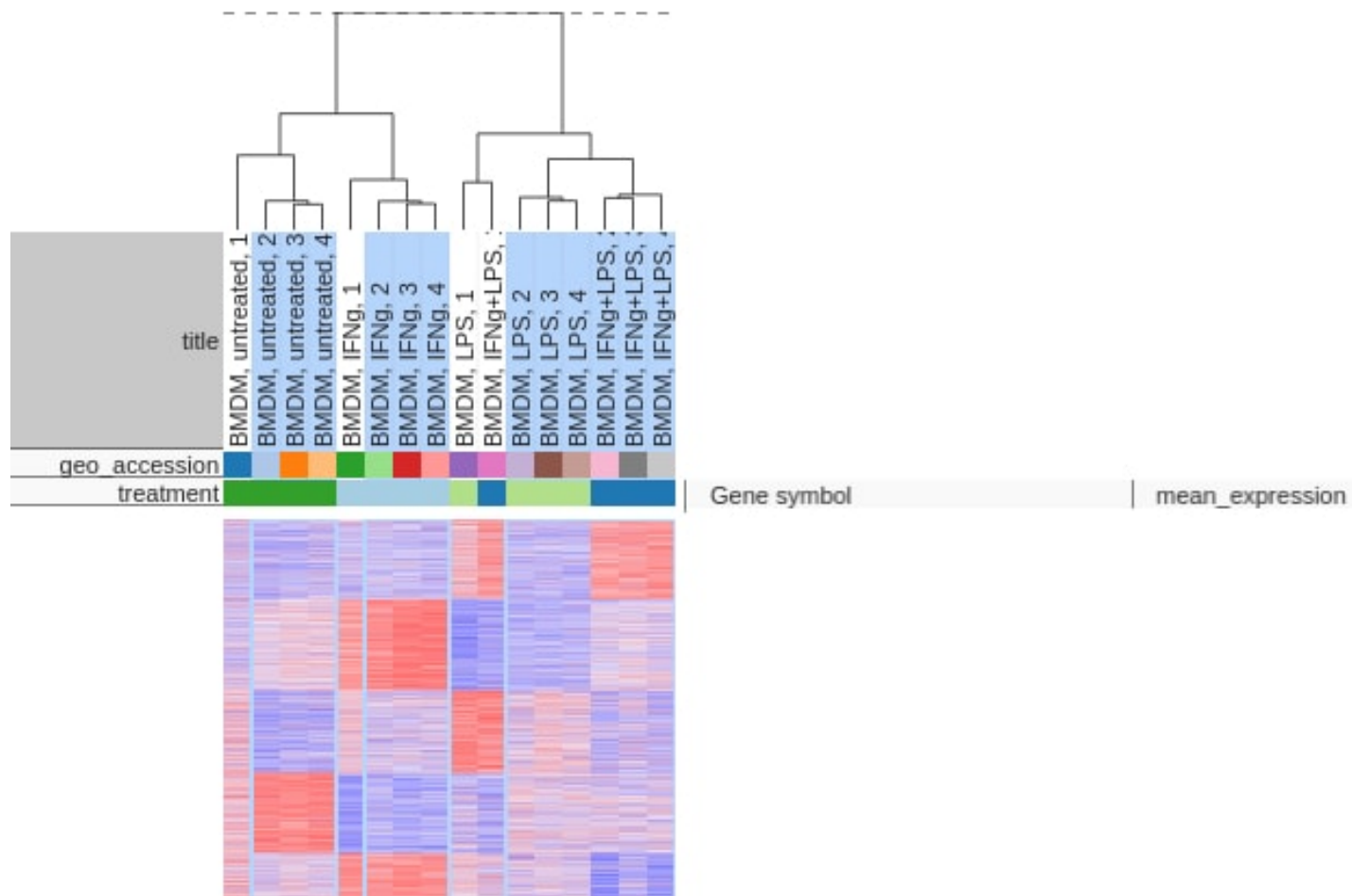
Group columns byNothing selected▼

☐ Cluster columns in space of selected rows only

OKCancel

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.

limma

Field

geo_accession

strain

tissue

title

treatment

Class A

Search

1/4

☐ IFNg

☐ IFNg+LPS

☐ LPS

☒ Untreated

Class B

Search

1/4

☐ IFNg

☐ IFNg+LPS

☒ LPS

☐ Untreated

OK

Cancel

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

