SUMMER Tutorial

07-29-2020

Salk IGC

Overview of SUMMER web server

SUMMER - Shiny Utility for Metabolomics and Multiomics Exploratory Research

run a test sample dataset. (Then click 'Run analyses' button at the bottom)

Use transcriptomes or proteomes?

- (•) Transcriptome
- ( ) Proteome

Species

- ( ) Human
- (•) Mouse

Choose metabolomics data file

Browse... No file selected

[x] Header

Separator

- ( ) Comma
- ( ) Semicolon
- (•) Tab

Choose transcriptomics data file

Browse... No file selected

[x] Header

Data upload panel

```
Welcome to SUMMER.
Please start by uploading your data at the left panel.
accepted plain text input, separated by tab/comma/semicolon
```

System information box: Used for instructions and warning messages.

Intro | PCA | Differential Expression | Pathway Analysis | Network Graph

Data analysis tabs

Background

SUMMER is an R shiny application for multiomics analysis (metabolomics + transcriptomics/proteomics).

SUMMER was developed by the Bioinformatics Core at the Salk Institute. Please email lhuang@salk.edu or mshokhirev@salk.edu for any question.

Tutorial: summer tutorial

Test Datasets: Download the test input file for metabolites [1] | Download the test input file for genes [1]

Download the test input file for proteins (transformed from gene test dataset) [1]

download sample dataset

Note: the raw intensities can be shown in '1000' or '1,000' format, but the latter one needs to be delimited by tab or space rather than comma. The gene expression table does not allow comma in the numbers.

More about input data formats:

SUMMER accepts the following input fomrat:

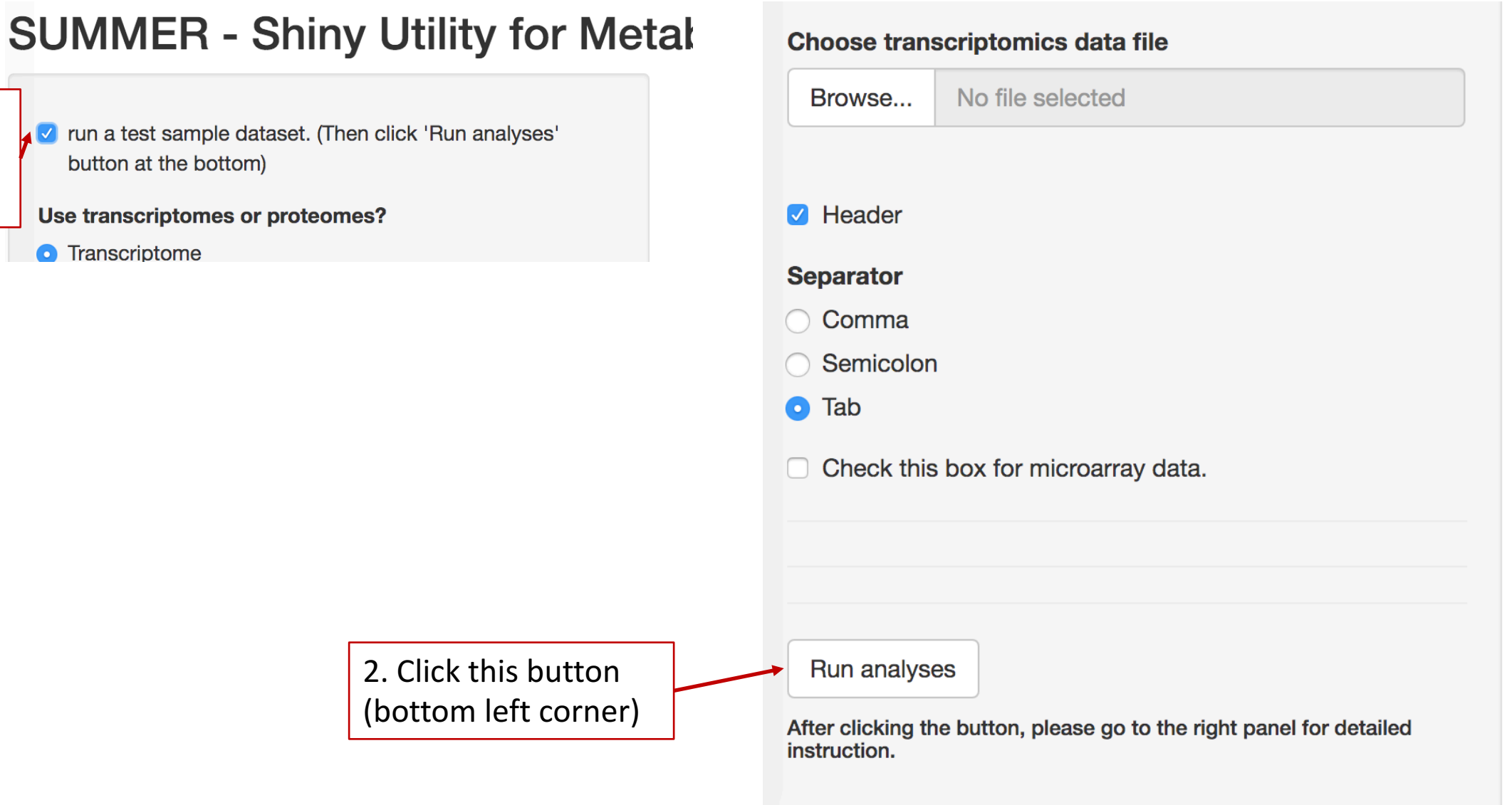
1. Metabolomics: KEGG [2] compound ID with raw intensity from Mass Spec.

2. Transcriptomics: Entrez Gene ID with FPKM/TPM values or microarray raw intensity.

3. or Proteomics: Entrez Gene ID with raw intensity from Mass Spec. Imputation and quantile normalization will be performed on the input data. If you have normalized proteomics data, you can upload it in the 'Transcriptomics panel'. (Currently either transcriptomics or proteomics data is accepted as input)

To run a test dataset:

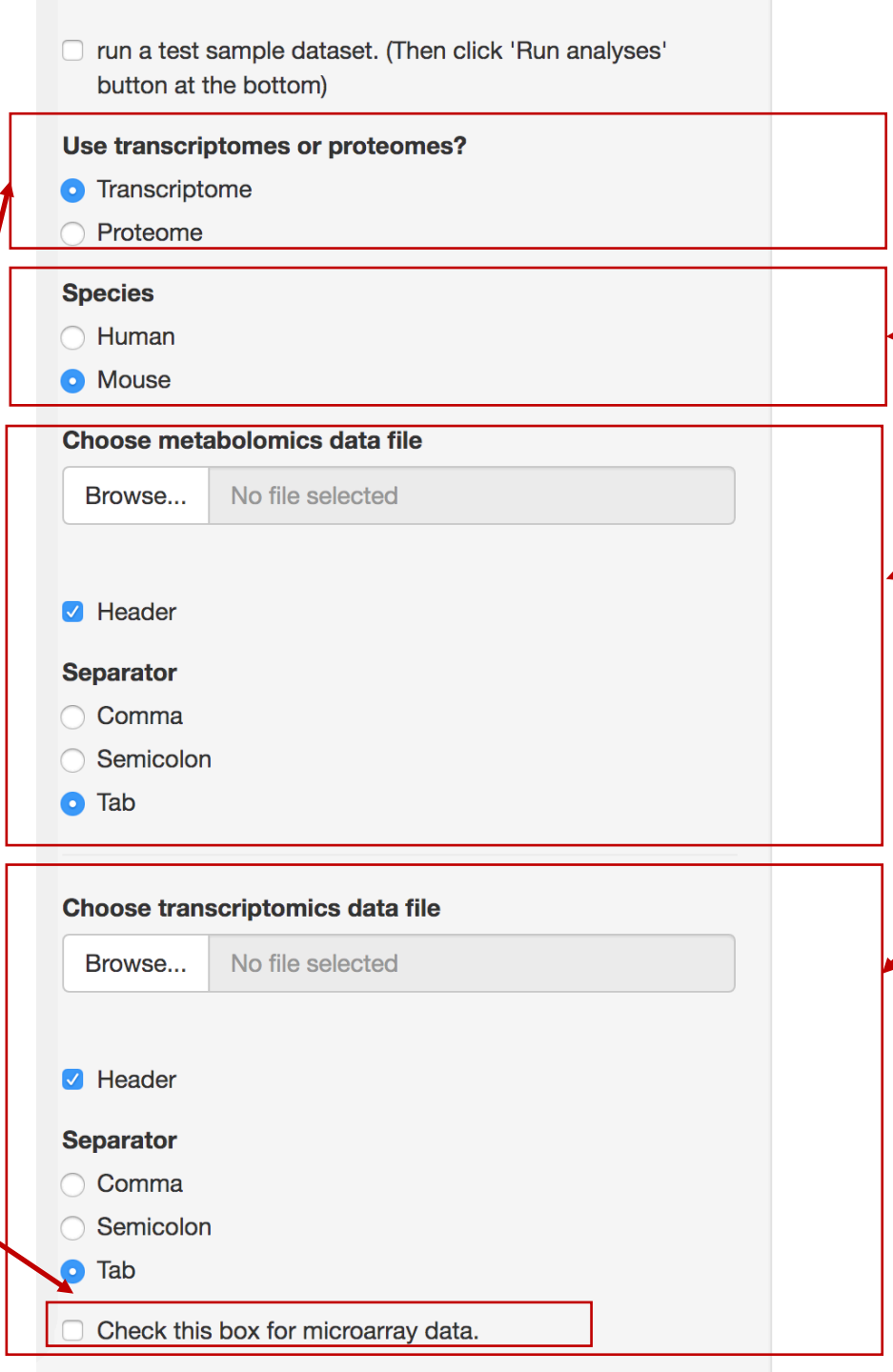
1. Check this box (on top left corner)



2. Click this button (bottom left corner)

To run a new analysis:

1. choose to use transcriptome or proteome data.
Note: if you don't have either, leave the selection as is, skip the upload of transcriptomics data, click "Run analyses".



2. choose species.

3. upload the metabolomics data (must). Choose if the input table has a header line and whether the table is delimited by comma, semicolon, or tab.

4. upload the transcriptomics or proteomics data (optional).

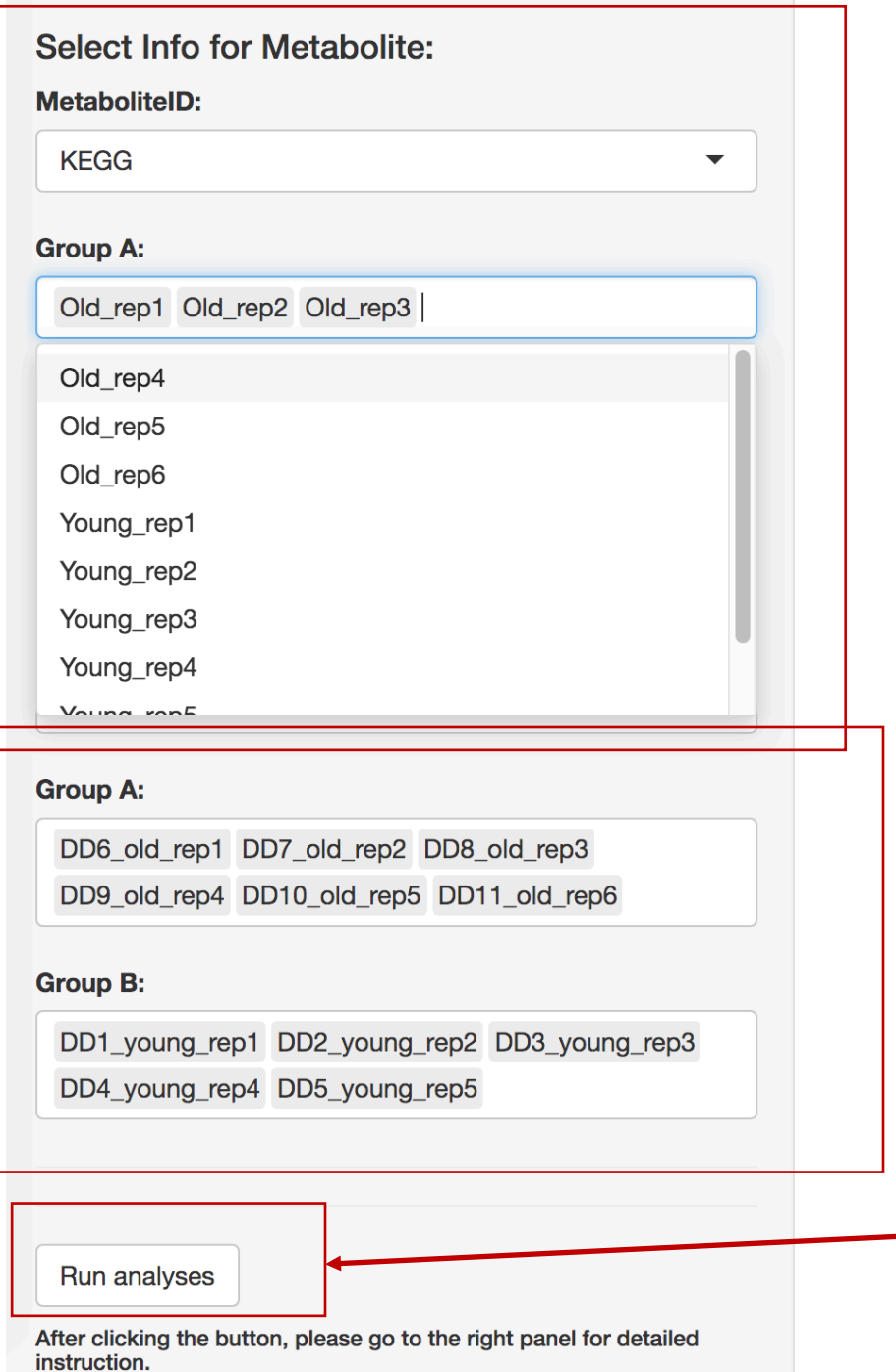
A panel to upload proteomics data will appear if "Proteome" is selected in step1.

5. Don't forget to check this box for microarray data.
Data will be quantile-normalized upon this selection.

To run a new analysis (continued):

6. set up experiment design for metabolomics data.
First, choose the column headers from drop-down menu that contain KEGG metabolite IDs. Then, choose the column headers that contain replicates for each condition

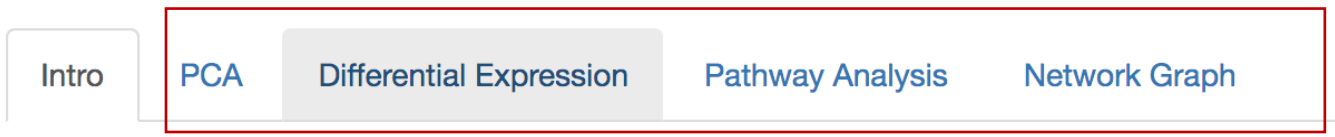
7. set up experiment design for transcriptomics/proteomics data.



8. click “Run analyses” to start the analysis.

Run analyses

```
Checking metabolomics data...
If there are duplicated IDs, only the first one will be used for analysis.
284 mapped compound ID in input metabolites data
6 cpd were removed due to too many NAs in a row
Checking transcriptomics data...
24531 mapped gene ID in input transcriptomics data
0 transcripts were removed due to too many NAs in a row
Replicates number >= 3 for all datasets. Good.
Replicates number is not the same for all datasets.
It will cause problem if we want to integrate protein and RNA data. Since that module is currently not
supported, you can go ahead and try to use these samples in the analysis. Now you can click the next
panel button to start the analysis, e.g. PCA or DE analysis.
```



After clicking “Run analyses”, go to the information box. It will show how many metabolites, transcripts/proteins are mapped and filtered. Replicate numbers are allowed to be different for metabolomics and transcriptomics/proteomics for now.

Now it is ready to actually run the analyses. Note: Pathway analysis and network graph can only be run after differential expression analysis is performed.

Principle Component Analysis (PCA)

Click PCA tab.

supported, you can go ahead and try to use these samples in the analysis. Now you can click the next panel button to start the analysis, e.g. PCA or DE analysis.

Intro | PCA | Differential Expression | Pathway Analysis | Network Graph

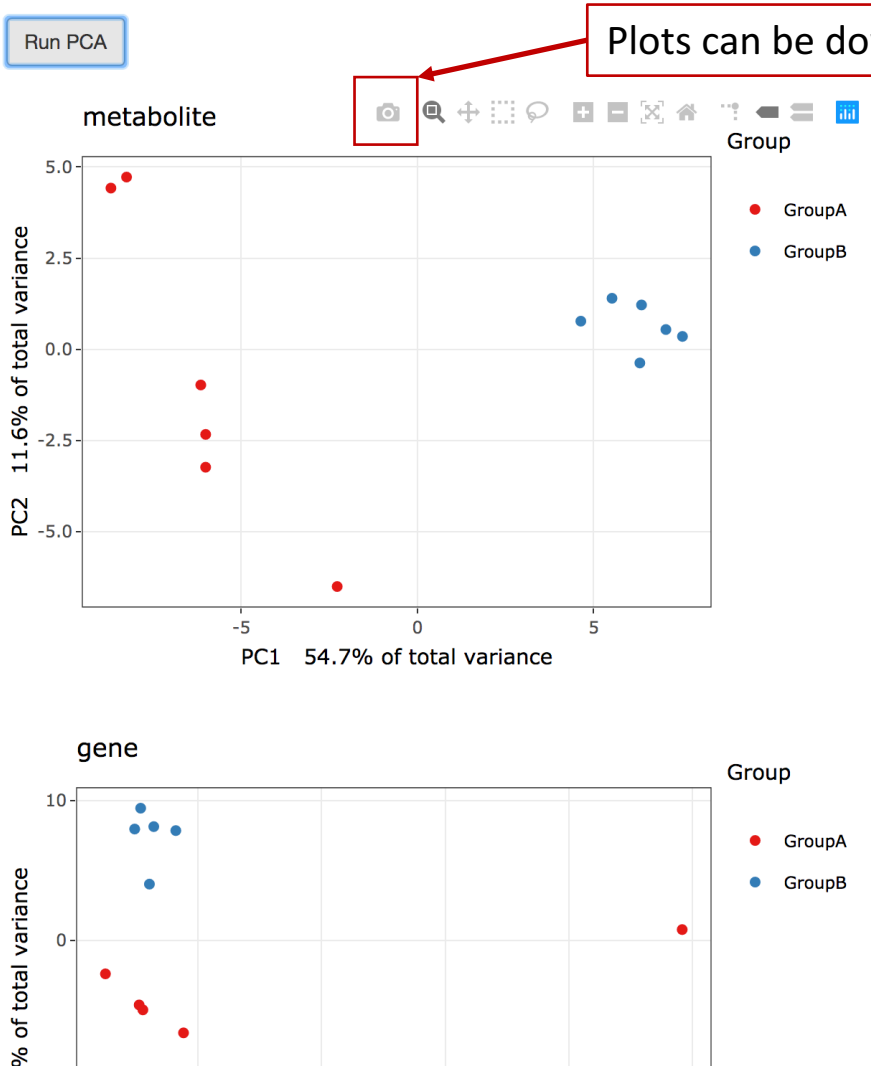
Principle Component Analysis is a great tool to check sample quality through univariant analysis. Once the input dataset is succesfully uploaded and mapped to KEGG database, a button to run PCA will be shown below.

The PCA plots will be shown in the order of metabolite, gene, and protein (at the bottom) when multiple inputs are detected. The PCA work on total metabolites and top 10% most variable genes/proteins to minimize noise in datasets.

Run PCA

This button will appear if the uploaded data looks good. Click to run PCA.

Principle Component Analysis (PCA)

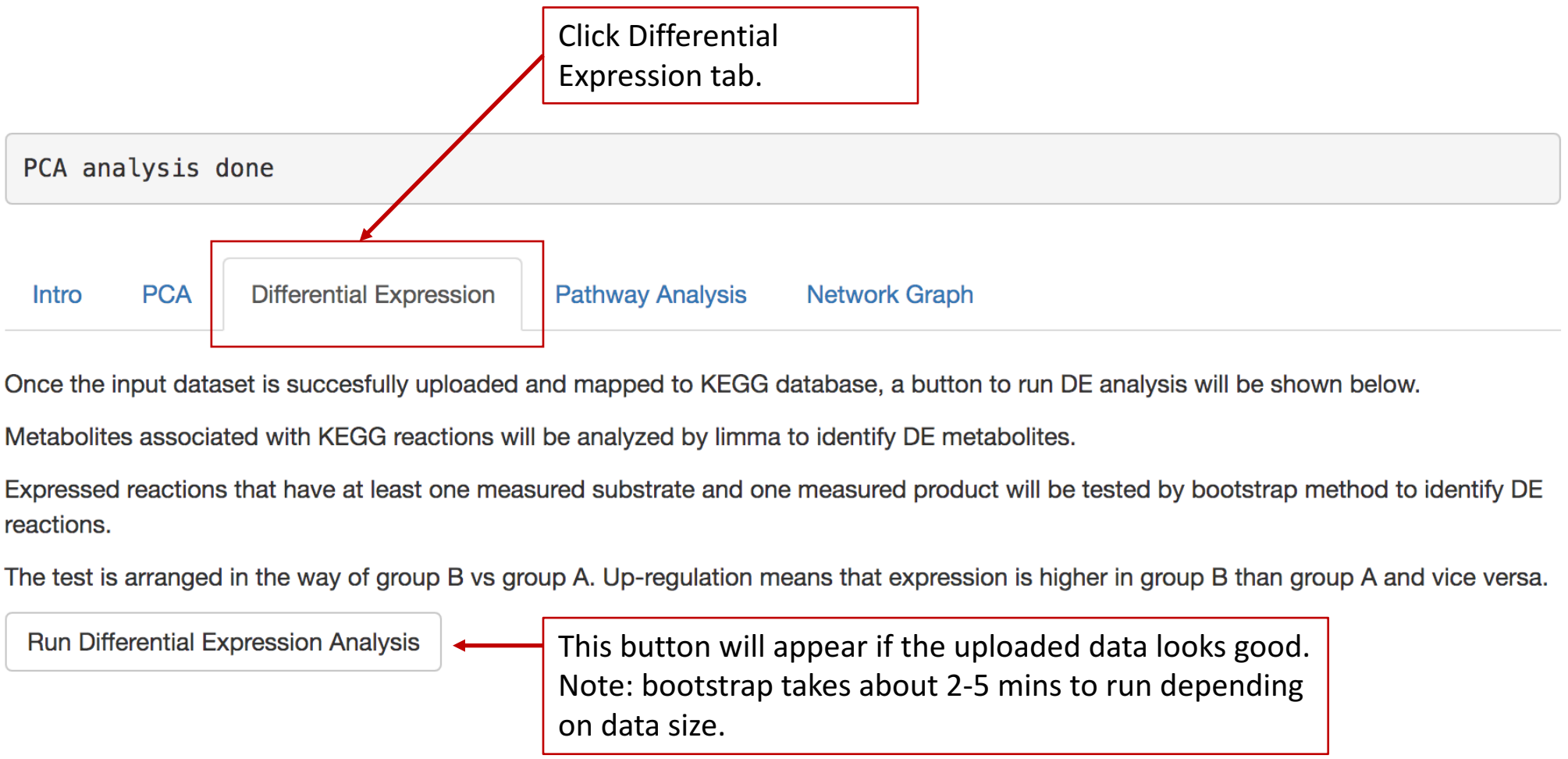


After clicking the “Run PCA” button, the PCA plots will be generated to visualize sample separations. Details will be shown upon mouse hovering.

It takes about half a minute to render the plots, please be patient.

PCA panels will be displayed in the order of metabolite (top), gene (middle), and protein (bottom).

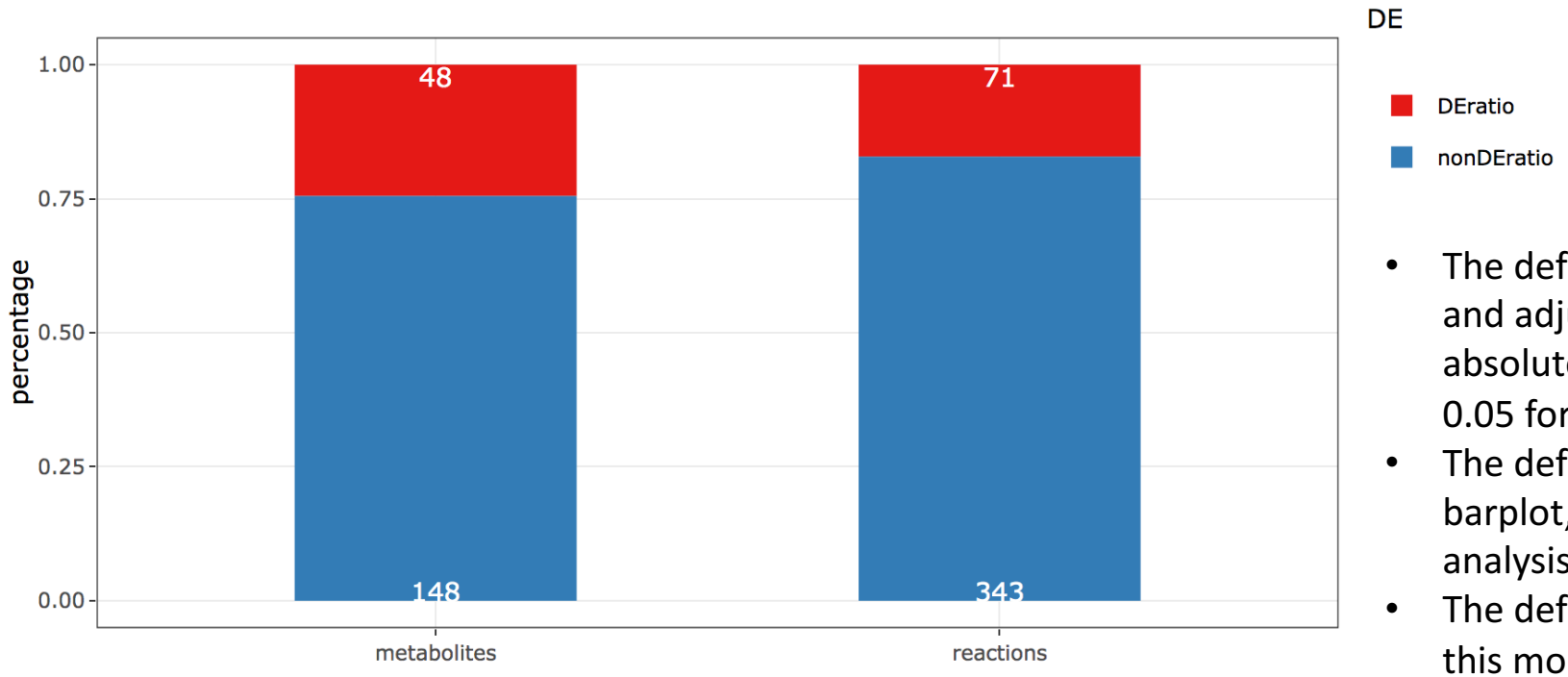
Differential Expression (DE) Analysis



Differential Expression (DE) Analysis

The test is arranged in the way of group B vs group A. Up-regulation means that expression is higher in group B than group A and vice versa.

Run Differential Expression Analysis



- The default DE cutoff is absolute logFC > 0.5 and adjusted p-val < 0.05 for metabolites, absolute logFC > 0.5 and ranking score < 0.05 for reactions.
- The default DE cutoff is used for this barplot, and the pathway enrichment analysis on the next slide.
- The default DE cutoff is not adjustable at this moment.

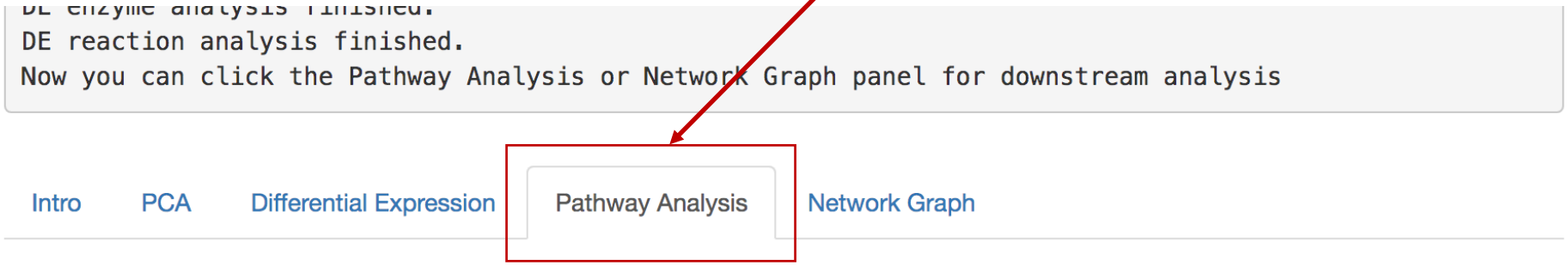
Download results for metabolites

Download results for reactions

Download button to download the DE results.

Pathway Analysis

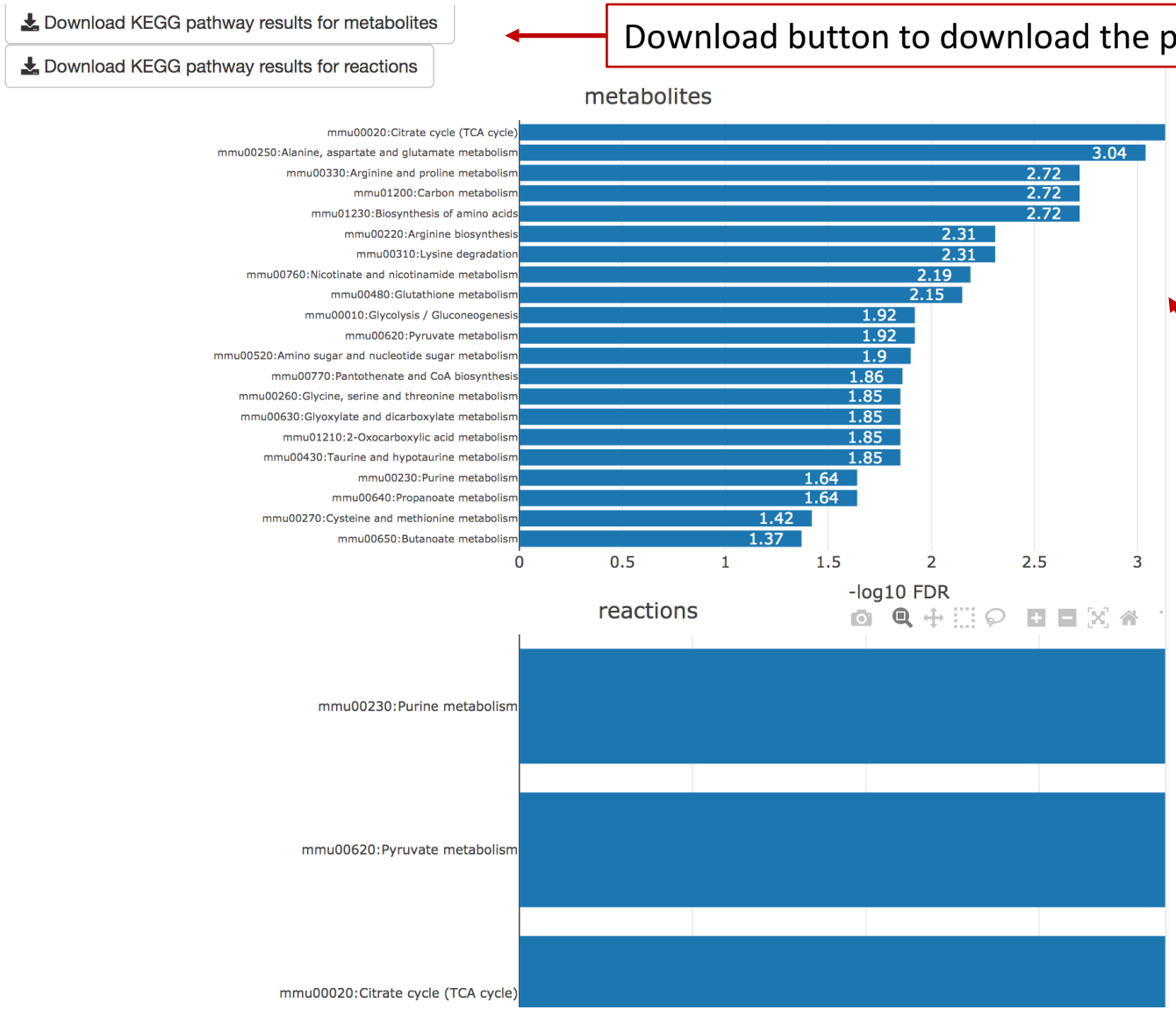
Click Pathway Analysis tab.



Currently, only Over-representation Analysis is supported.
Choose DE cutoff for the pathway enrichment result.

This button will appear if the DE analysis is performed.

Pathway Analysis



Download button to download the pathway results.

Barplots of significantly enriched KEGG metabolic pathway terms for DE metabolites and DE reactions.

Pathway Analysis

Select pathways to plot:

mmu00020:Citrate cycle (TCA cycle) - Mus musculus (mouse)

Render Pathway Network Graph

Select by label

Choose KEGG metabolic pathways from drop-down menus to generate a network graph for that pathway. All measured metabolites will be shown regardless of its statistical significance.

Pathway view of Network Graph

Download Pathway Network List nodes

Download Pathway Network List edges

Download Pathway Network Graph HTML

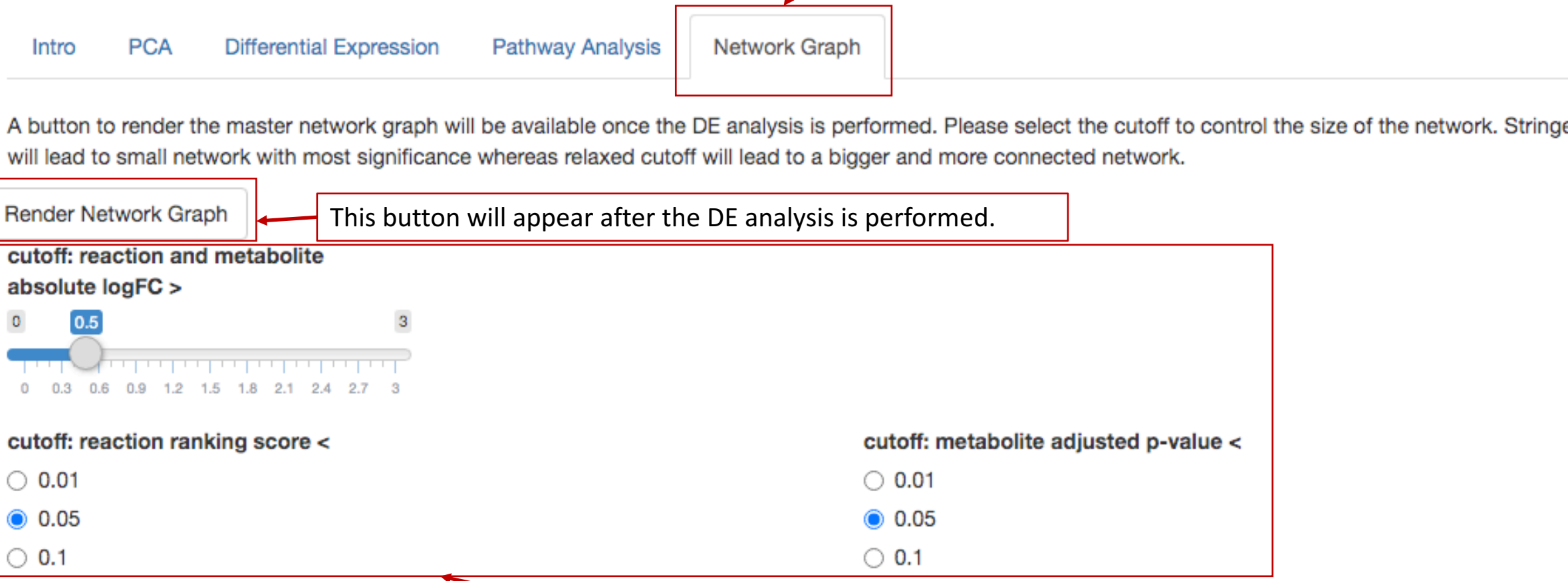
Download Pathway Network Graph PDF

Download Pathway Network GML for cytoscape

Download button to download the pathway network graph in different formats.

Network Graph

Click Network Graph tab.

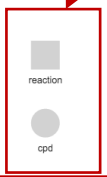


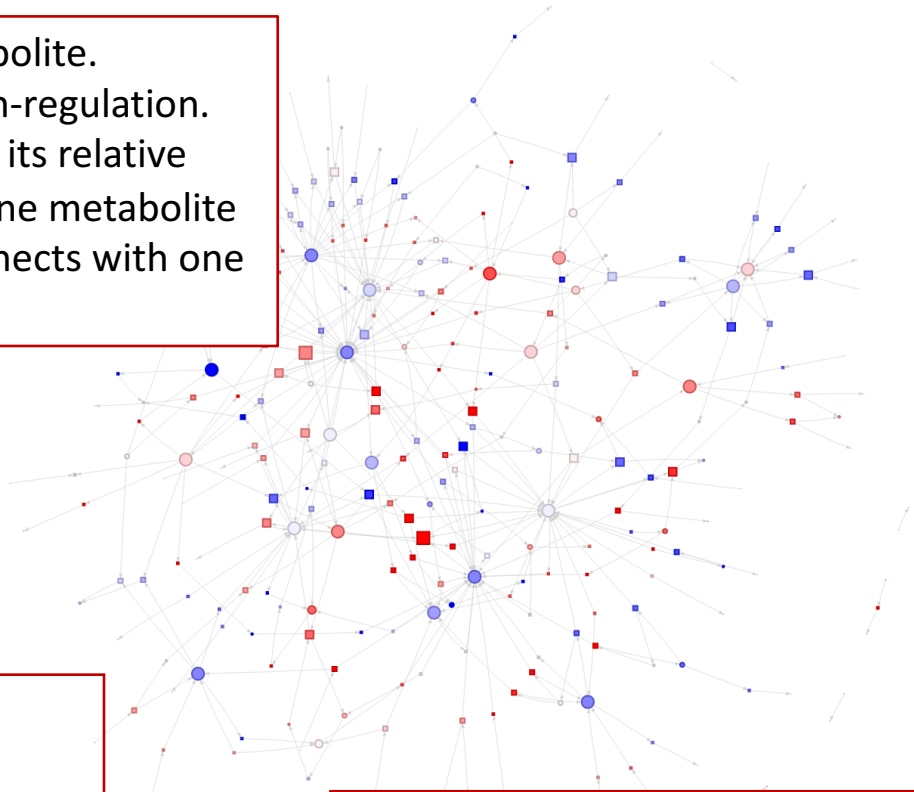
Network Graph

Select by label

Choose from drop-down menu the metabolite/reaction needs to be highlighted.

Square=reaction, circle=metabolite.
red=up-regulation, blue=down-regulation.
Network node size represents its relative connectivity with others. So one metabolite node can be tiny if it only connects with one reaction.





Global view of Network Graph

Can be magnified by mouse scrolling.

Download Network List nodes
Download Network List edges
Download Network Graph pdf
Download Network Graph HTML
Download Network GML for cytoscape

Download button to download the network graph in different formats.