

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

Header

Separator

Comma
 Semicolon
 Tab

Choose transcriptomics data file

Browse... No file selected

Header

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 luhuan@salk.edu or mshokhirev@salk.edu for any question.

Tutorial: [summer tutorial](#)

download sample dataset

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 \(transformed from gene test dataset\) \[1\]](#)

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 format:

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)

Data upload panel

To run a test dataset:

SUMMER - Shiny Utility for Metal

1. Check this box
(on top left
corner)

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

Use transcriptomes or proteomes?

Transcriptome

2. Click this button
(bottom left corner)

Choose transcriptomics data file

Browse...

No file selected

Header

Separator

Comma

Semicolon

Tab

Check this box for microarray data.

Run analyses

After clicking the button, please go to the right panel for detailed instruction.

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".

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

- Header

Separator

- Comma
- Semicolon
- Tab

Choose transcriptomics data file

Browse... No file selected

- Header

Separator

- Comma
- Semicolon
- Tab

Check this box for microarray data.

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.

Select Info for Metabolite:

MetaboliteID:

KEGG

Group A:

Old_rep1 Old_rep2 Old_rep3 |

Old_rep4

Old_rep5

Old_rep6

Young_rep1

Young_rep2

Young_rep3

Young_rep4

Young_rep5

Group A:

DD6_old_rep1 DD7_old_rep2 DD8_old_rep3

DD9_old_rep4 DD10_old_rep5 DD11_old_rep6

Group B:

DD1_young_rep1 DD2_young_rep2 DD3_young_rep3

DD4_young_rep4 DD5_young_rep5

Run analyses

8. click "Run analyses" to start
the analysis.

After clicking the button, please go to the right panel for detailed
instruction.

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.
```

Intro

PCA

Differential Expression

Pathway Analysis

Network Graph

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 univariate analysis. Once the input dataset is successfully 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 works on total metabolites and top 10% most variable genes/proteins to minimize noise in datasets.

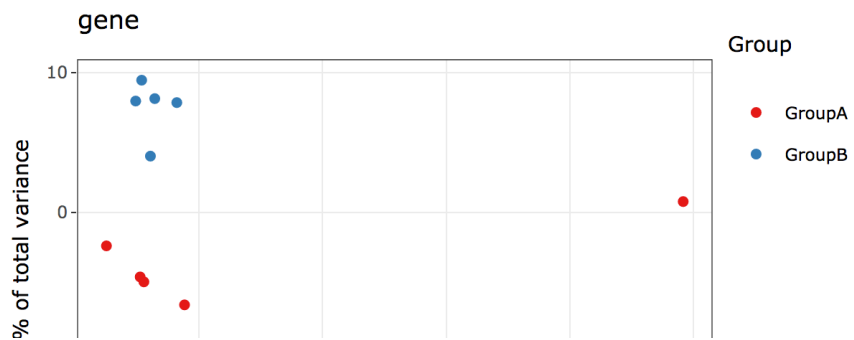
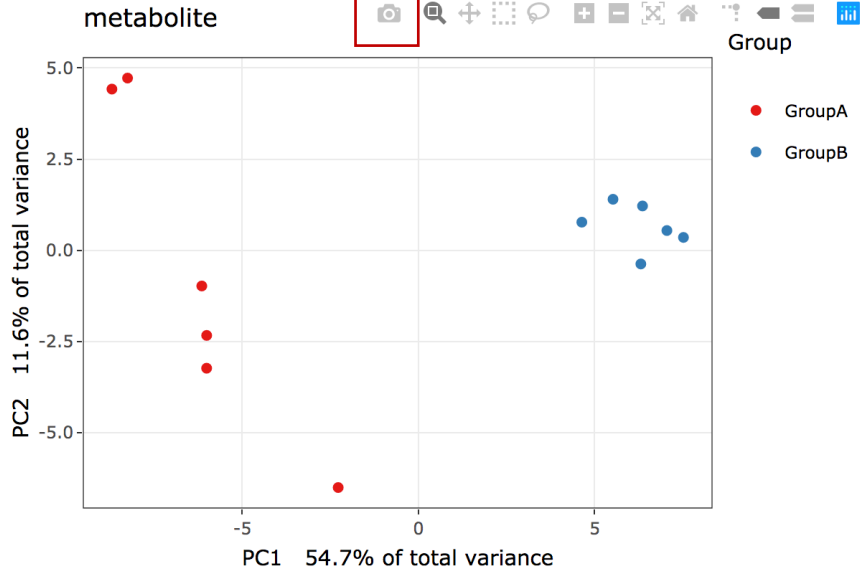
Run PCA

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

Principle Component Analysis (PCA)

Run PCA

Plots can be downloaded



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

Click Differential Expression tab.

PCA analysis done

Intro

PCA

Differential Expression

Pathway Analysis

Network Graph

Once the input dataset is successfully uploaded and mapped to KEGG database, a button to run DE analysis will be shown below.

Metabolites associated with KEGG reactions will be analyzed by limma to identify DE metabolites.

Expressed reactions that have at least one measured substrate and one measured product will be tested by bootstrap method to identify DE reactions.

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.

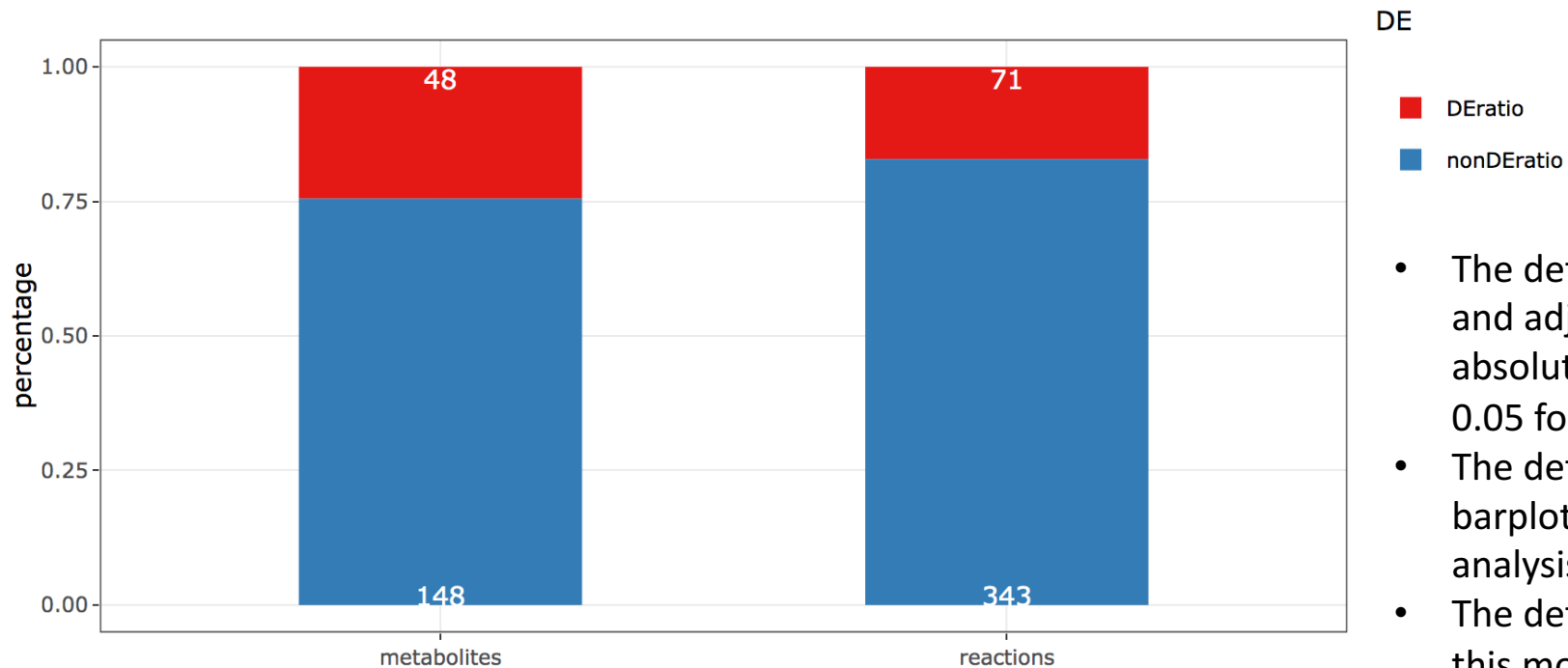
Run Differential Expression Analysis

This button will appear if the uploaded data looks good.
Note: bootstrap takes about 2-5 mins to run depending on data size.

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



DE

- DEratio
- nonDEratio

- The default DE cutoff is absolute $\log_{FC} > 0.5$ and adjusted p-val < 0.05 for metabolites, absolute $\log_{FC} > 0.5$ 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.

DE enzyme analysis finished.
DE reaction analysis finished.
Now you can click the Pathway Analysis or Network Graph panel for downstream analysis

[Intro](#) [PCA](#) [Differential Expression](#) **Pathway Analysis** [Network Graph](#)

A button to run pathway analysis will be shown after the DE analysis is performed.

Currently, only over-representation analysis is supported.

A network for each pathways included in KEGG reaction pathway database can be constructed regardless of the significance of that pathway in order to provide an unbiased overview of what happens at pathway level. However, care should be taken to properly interpret the information included in the pathway network. The pathway network will be shown at the bottom of the page, please scroll down a bit to find it.

Method for Pathway Analysis

Over-Representation Analysis

FDR cutoff for Pathway Analysis

0.01

0.05

0.1

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

Run KEGG Pathway Enrichment Analysis

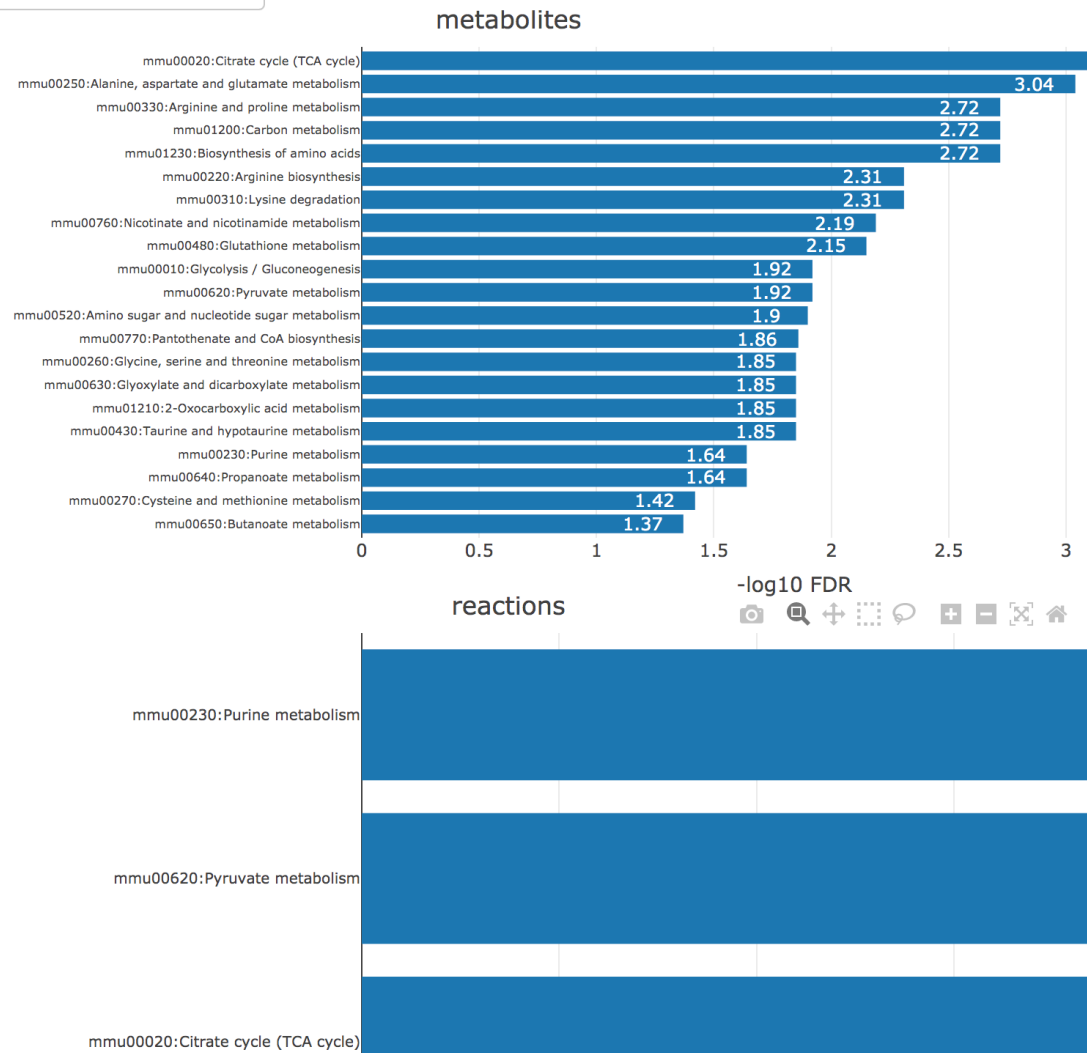
This button will appear if the DE analysis is performed.

Pathway Analysis

Download KEGG pathway results for metabolites

Download KEGG pathway results for reactions

Download button to download the pathway results.



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

Pathway Analysis

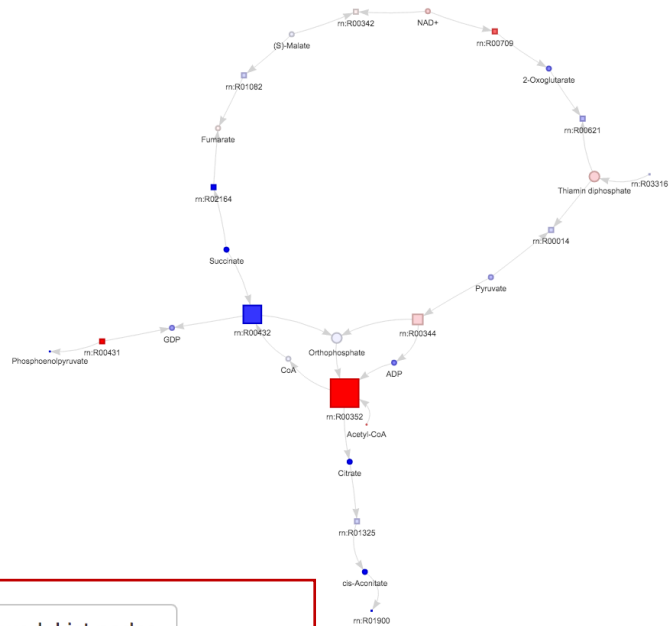
Select pathways to plot:

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.

Intro

PCA

Differential Expression

Pathway Analysis

Network Graph

A button to render the master network graph will be available once the DE analysis is performed. Please select the cutoff to control the size of the network. Stringer cutoffs will lead to small network with most significance whereas relaxed cutoff will lead to a bigger and more connected network.

Render Network Graph

This button will appear after the DE analysis is performed.

**cutoff: reaction and metabolite
absolute logFC >**



cutoff: reaction ranking score <

- 0.01
- 0.05
- 0.1

cutoff: metabolite adjusted p-value <

- 0.01
- 0.05
- 0.1

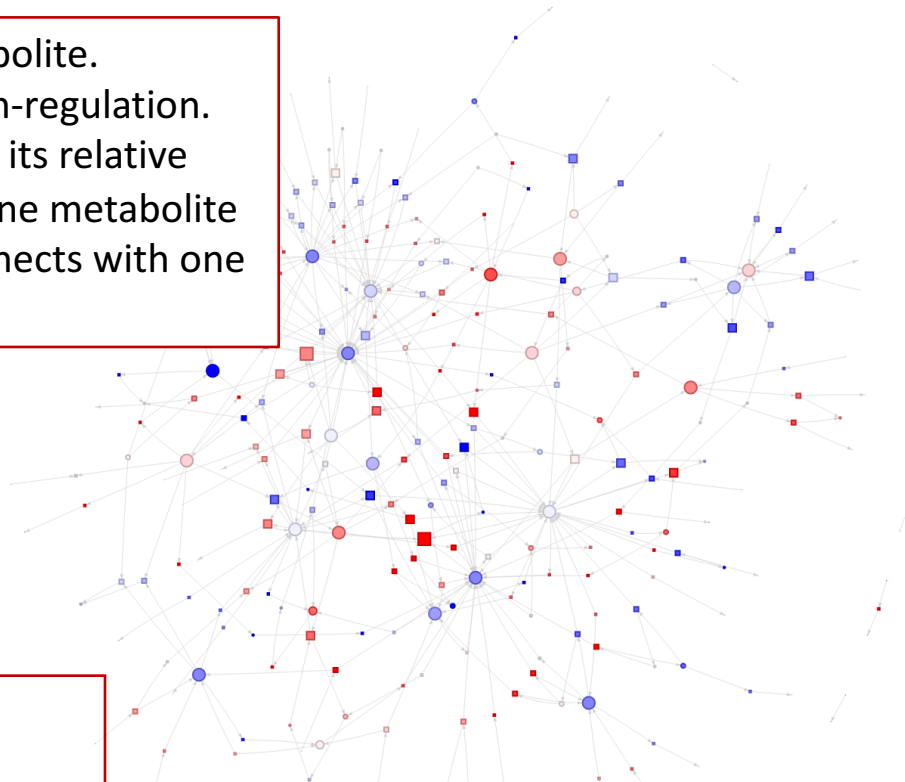
Choose the statistical significance cutoff to construct a network of DE metabolites and DE reactions plus their associated partners.

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.



reaction



cpd

Download button to download the network graph in different formats.

Download Network List nodes

Download Network List edges

Download Network Graph pdf

Download Network Graph HTML

Download Network GML for cytoscape