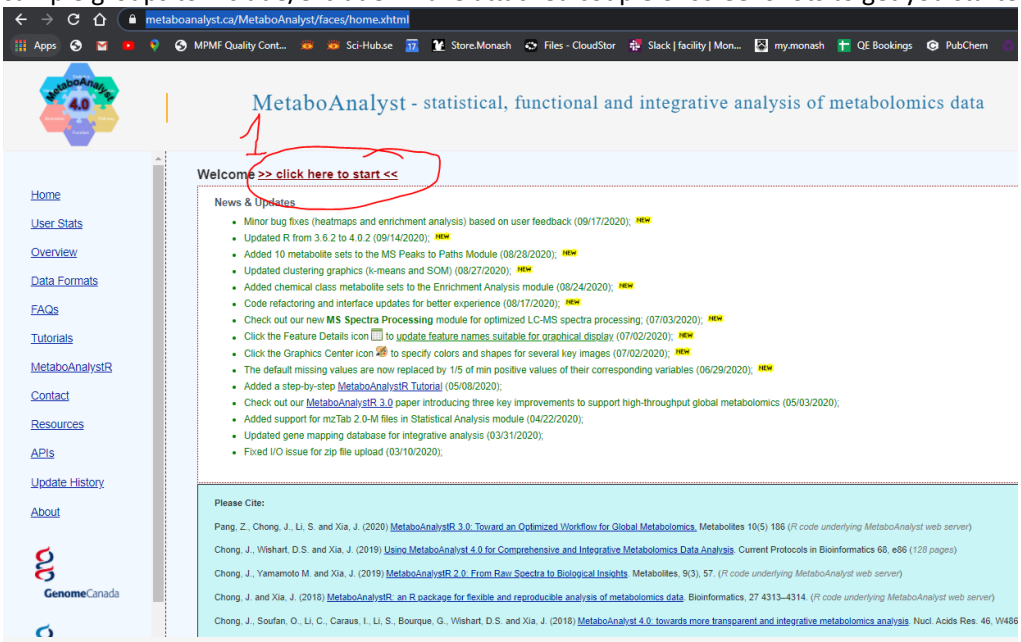
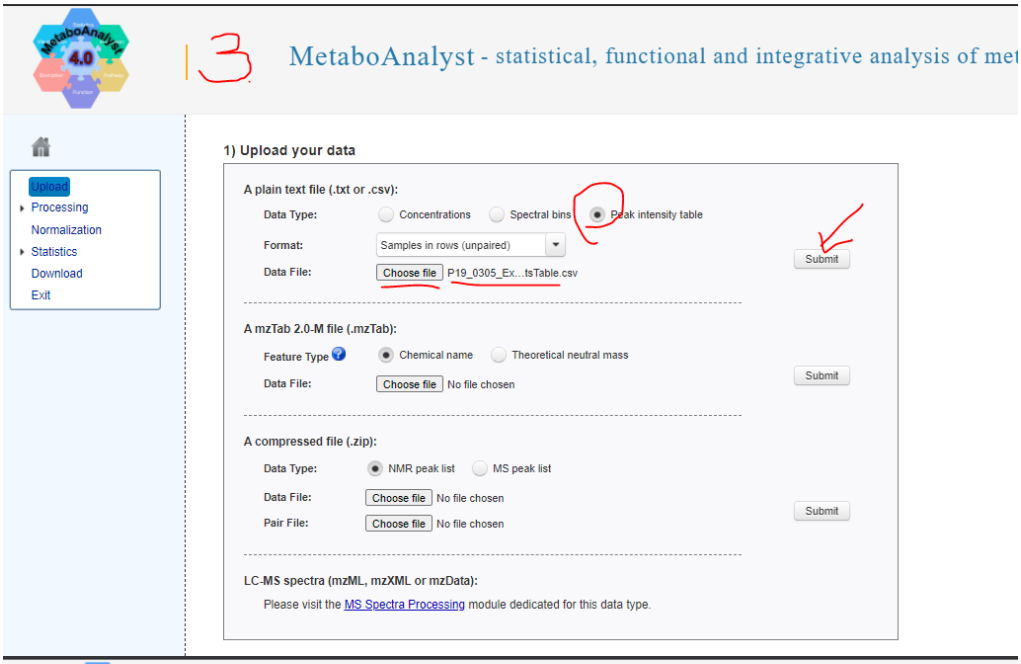
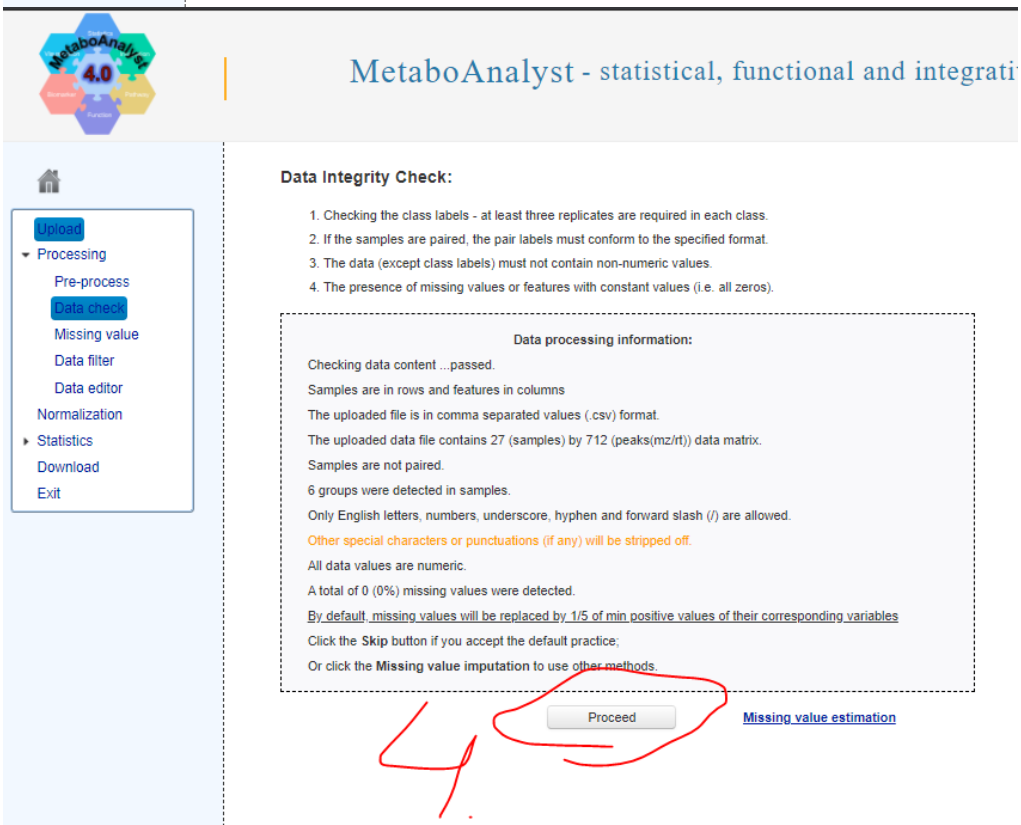
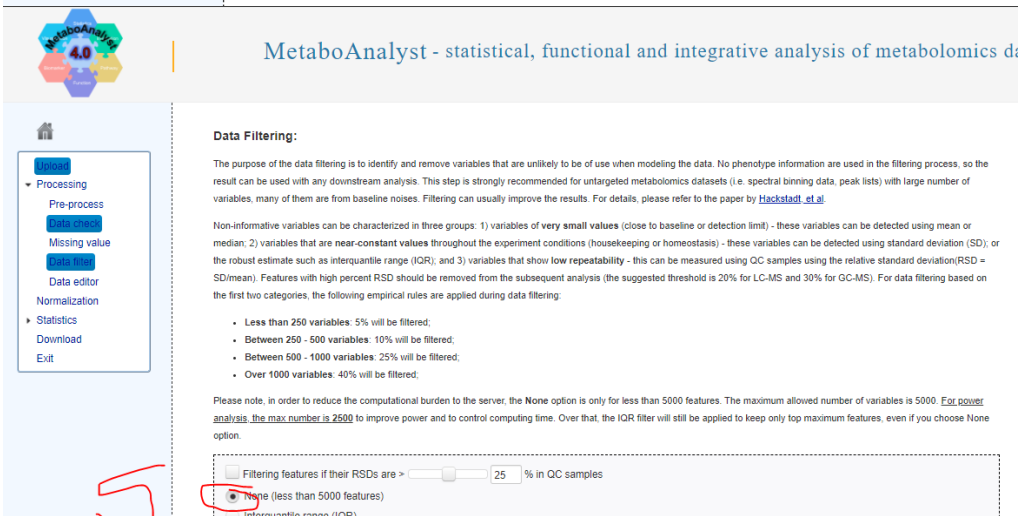
Subject: Re: P19_0305_Exp1 RESULTS and REPORT
Date: Friday, 18 September 2020 at 1:22:44 pm Australian Eastern Standard Time
From: Dovile Anderson
To: Sarah Best
Attachments: image.png, image.png, image.png, image.png, image.png, image.png, image.png, image.png

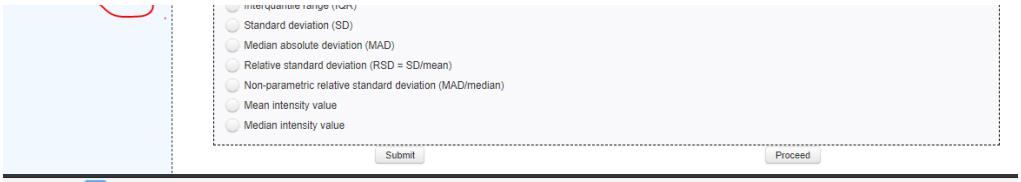
Hi Sarah,
there is an easier way to generate statistics plots
- https://www.metaboanalyst.ca/MetaboAnalyst/faces/home.xhtml
use the StatsTable.cvs file which should be available with your other files.
import this file into MetaboAnalyst and run your statistical analysis. In the data editor you can choose which sample groups to include/exclude. I have attached couple of screenshots to get you started:

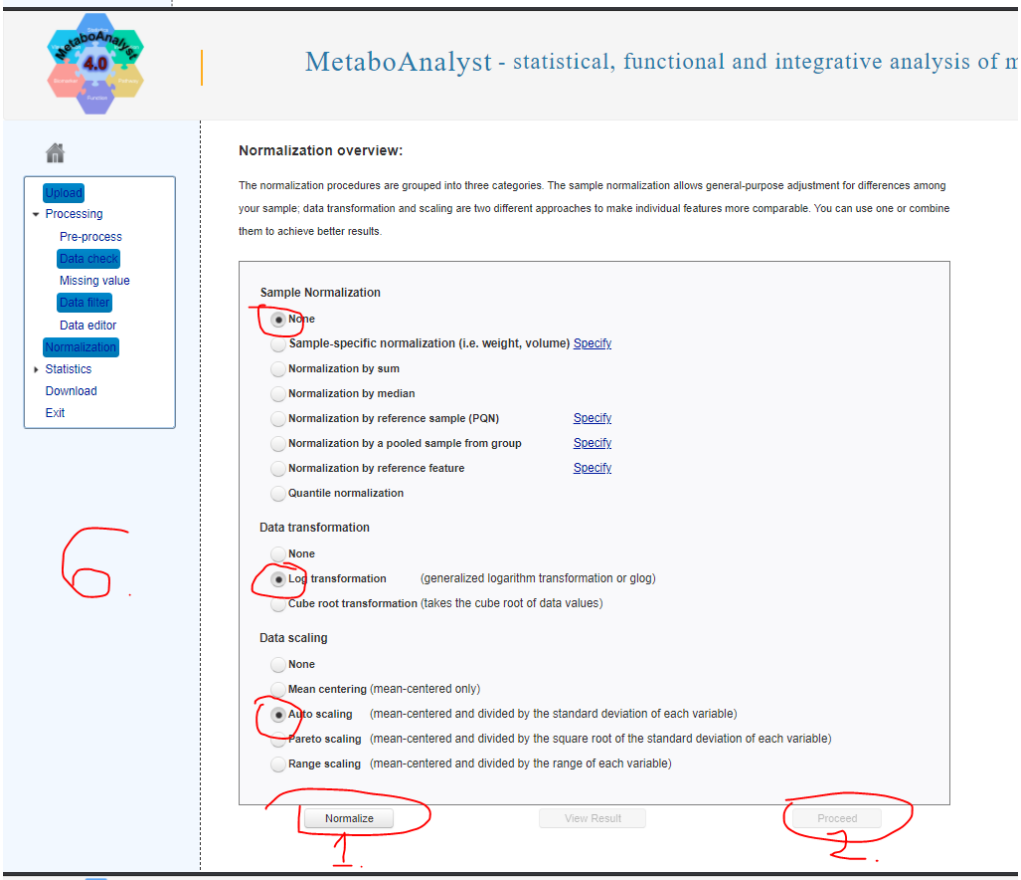


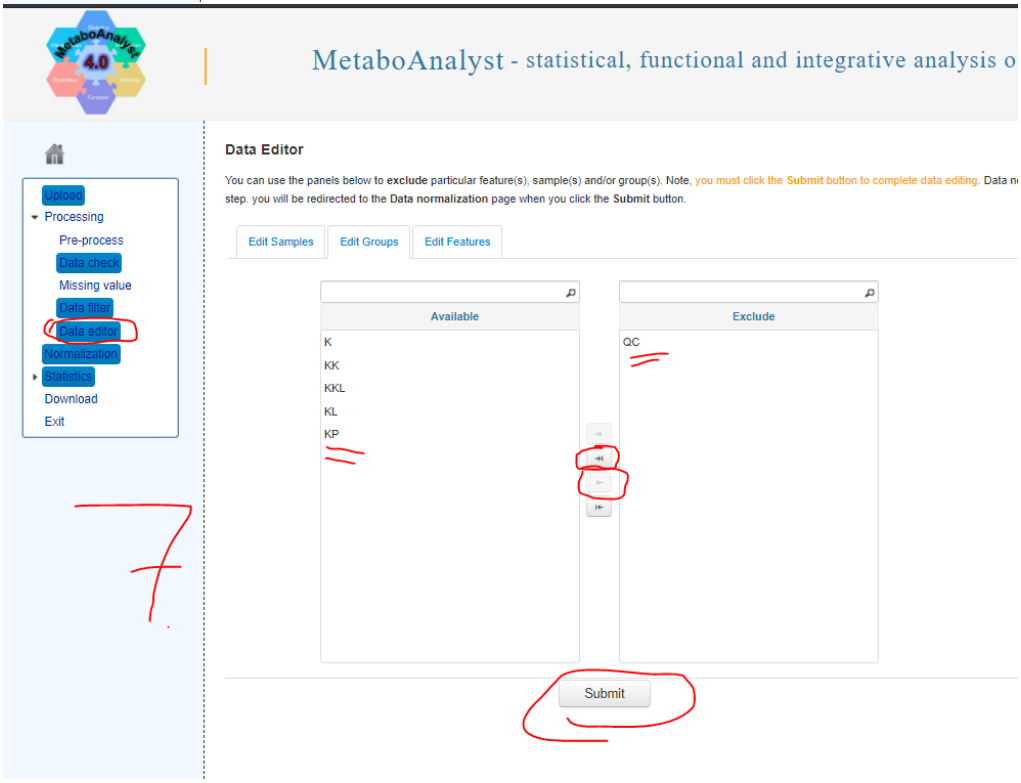

MetaboAnalyst - statistical, functional and integrative analysis of met
Upload
Processing
Normalization
Statistics
Download
Exit
1) Upload your data
A plain text file (.txt or .csv):
Data Type: Concentrations Spectral bins Peak intensity table
Format: Samples in rows (unpaired)
Data File: Choose file P19_0305_Ex...tsTable.csv
Submit
A mzTab 2.0-M file (.mzTab):
Feature Type Chemical name Theoretical neutral mass
Data File: Choose file No file chosen
Submit
A compressed file (.zip):
Data Type: NMR peak list MS peak list
Data File: Choose file No file chosen
Pair File: Choose file No file chosen
Submit
LC-MS spectra (mzML, mzXML or mzData):
Please visit the MS Spectra Processing module dedicated for this data type.

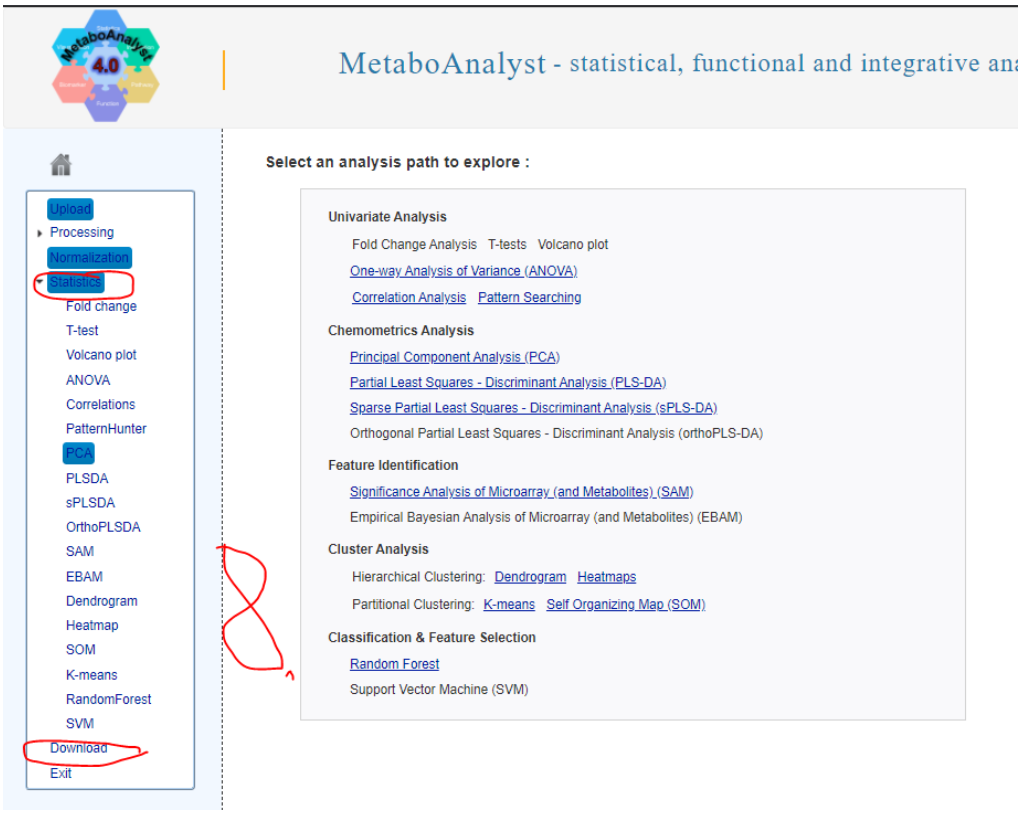

MetaboAnalyst - statistical, functional and integrati
Upload
Processing
Pre-process
Data check
Missing value
Data filter
Data editor
Normalization
Statistics
Download
Exit
Data Integrity Check:
1. Checking the class labels - at least three replicates are required in each class.
2. If the samples are paired, the pair labels must conform to the specified format.
3. The data (except class labels) must not contain non-numeric values.
4. The presence of missing values or features with constant values (i.e. all zeros).
Data processing information:
Checking data content ...passed.
Samples are in rows and features in columns
The uploaded file is in comma separated values (.csv) format.
The uploaded data file contains 27 (samples) by 712 (peaks(mz/rt)) data matrix.
Samples are not paired.
6 groups were detected in samples.
Only English letters, numbers, underscore, hyphen and forward slash (/) are allowed.
Other special characters or punctuations (if any) will be stripped off.
All data values are numeric.
A total of 0 (0%) missing values were detected.
By default, missing values will be replaced by 1/5 of min positive values of their corresponding variables
Click the Skip button if you accept the default practice;
Or click the Missing value imputation to use other methods.
Proceed
Missing value estimation


MetaboAnalyst - statistical, functional and integrative analysis of metabolomics da
Upload
Processing
Pre-process
Data check
Missing value
Data filter
Data editor
Normalization
Statistics
Download
Exit
Data Filtering:
The purpose of the data filtering is to identify and remove variables that are unlikely to be of use when modeling the data. No phenotype information are used in the filtering process, so the result can be used with any downstream analysis. This step is strongly recommended for untargeted metabolomics datasets (i.e. spectral binning data, peak lists) with large number of variables, many of them are from baseline noises. Filtering can usually improve the results. For details, please refer to the paper by Hackstadt, et al.
Non-informative variables can be characterized in three groups: 1) variables of very small values (close to baseline or detection limit) - these variables can be detected using mean or median; 2) variables that are near-constant values throughout the experiment conditions (housekeeping or homeostasis) - these variables can be detected using standard deviation (SD); or the robust estimate such as interquantile range (IQR); and 3) variables that show low repeatability - this can be measured using QC samples using the relative standard deviation(RSD = SD/mean). Features with high percent RSD should be removed from the subsequent analysis (the suggested threshold is 20% for LC-MS and 30% for GC-MS). For data filtering based on the first two categories, the following empirical rules are applied during data filtering:
Less than 250 variables: 5% will be filtered;
Between 250 - 500 variables: 10% will be filtered;
Between 500 - 1000 variables: 25% will be filtered;
Over 1000 variables: 40% will be filtered;
Please note, in order to reduce the computational burden to the server, the None option is only for less than 5000 features. The maximum allowed number of variables is 5000. For power analysis, the max number is 2500 to improve power and to control computing time. Over that, the IQR filter will still be applied to keep only top maximum features, even if you choose None option.
Filtering features if their RSDs are > 25 % in QC samples
None (less than 5000 features)


Standard deviation (SD)
Median absolute deviation (MAD)
Relative standard deviation (RSD = SD/mean)
Non-parametric relative standard deviation (MAD/median)
Mean intensity value
Median intensity value
Submit
Proceed


MetaboAnalyst - statistical, functional and integrative analysis of m
Upload
Processing
Pre-process
Data check
Missing value
Data filter
Data editor
Normalization
Statistics
Download
Exit
Normalization overview:
The normalization procedures are grouped into three categories. The sample normalization allows general-purpose adjustment for differences among your sample; data transformation and scaling are two different approaches to make individual features more comparable. You can use one or combine them to achieve better results.
Sample Normalization
None
Sample-specific normalization (i.e. weight, volume) Specify
Normalization by sum
Normalization by median
Normalization by reference sample (PQN) Specify
Normalization by a pooled sample from group Specify
Normalization by reference feature Specify
Quantile normalization
Data transformation
None
Log transformation (generalized logarithm transformation or glog)
Cube root transformation (takes the cube root of data values)
Data scaling
None
Mean centering (mean-centered only)
Auto scaling (mean-centered and divided by the standard deviation of each variable)
Pareto scaling (mean-centered and divided by the square root of the standard deviation of each variable)
Range scaling (mean-centered and divided by the range of each variable)
Normalize
View Result
Proceed


MetaboAnalyst - statistical, functional and integrative analysis o
Upload
Processing
Pre-process
Data check
Missing value
Data filter
Data editor
Normalization
Statistics
Download
Exit
Data Editor
You can use the panels below to exclude particular feature(s), sample(s) and/or group(s). Note, you must click the Submit button to complete data editing. Data n
step, you will be redirected to the Data normalization page when you click the Submit button.
Edit Samples
Edit Groups
Edit Features
Available
K
KK
KKL
KL
KP
Exclude
QC
Submit



Let me know if you have any questions and good luck!

Thanks,
Dovile

On Wed, 16 Sep 2020 at 17:47, Sarah Best <best@wehi.edu.au> wrote:

> Hi Dovile,
>
> I am starting to put a manuscript together with the data from our previous experiment – P19-0305.
>
> For this study, we are comparing the tumours of KK, KL and KKL. Additionally in the dataset are K and KP samples. I am trying to use IDEOM spreadsheet to generate heat maps and general overall data, but finding it difficult with the extra samples.
>
> If possible, could you please generate:
>
> -Top 100 metabolites heat map with dendrogram for just KK, KL, KKL samples
>
> -PCA plot for just KK, KL, KKL samples
>
> Or, could you point me in the direction so I can try to do myself?
>
> I hope the plasma samples have not caused any headaches for you!

Many thanks,

Sarah

From: Dovile Anderson <dovile.anderson@monash.edu>
Date: Tuesday, 13 August 2019 at 3:57 pm
To: Sarah Best <best@wehi.edu.au>, Kate Sutherland <sutherland.k@wehi.edu.au>, Darren Creek <Darren.Creek@monash.edu>
Subject: P19_0305_Exp1 RESULTS and REPORT

Hi Sarah and Kate,

Your recent metabolomics analysis has been successful and data processing has been completed. Please download IDEOM results and the report using the link below.

P19_0305_Exp1 RESULTS and REPORT.zip

Just a couple of tips on how to use IDEOM. You can browse the result in the "Comparison" sheet. Besides observing fold changes in metabolite responses (coloured cells), you can sort the columns by "pathway" or "map" using in-build Excel sorting functions. In this way you can see changes of all the metabolites involved in the specific pathway. Also you can filter according to t-test column for a particular sample group on the right side which will enable you to see only significant changes.

The metabolites highlighted in yellow are matched with the authentic standards by their retention times and thus have the highest confidence. Among them few from PPP pathway can be found. In the "Comparison" sheet the values in bold shows that the change is significant as indicated by p-value < 0.05.

By double-clicking on a metabolite name in the "Comparison" sheet you can see the average levels of chosen metabolite in sample groups. Error bars will be displayed on the chart. If you double-click on the “max intensity”, the chart showing the levels of this particulate metabolite across all samples will be displayed. It is a good practice to check variability of metabolite intensities across individual samples in a group before concluding that the change is real.

I hope you find it easy to navigate and if you have any questions please let me know. If you would like a data discussion meeting please let us know too.

Thanks,

Dovile

--

Dovile Anderson

Research Officer at Monash Proteomics and Metabolomics Facility (Parkville Node)

Monash Institute of Pharmaceutical Sciences

Monash University, Parkville Campus

Level 2, Building 3, Room 3.208

381 Royal Parade, Parkville

Victoria 3052, Australia

Phone: 0448671141

https://www.monash.edu/researchinfrastructure/proteomics

--

Dovile Anderson
Research Officer at Monash Proteomics and Metabolomics Facility (Parkville Node)
Monash Institute of Pharmaceutical Sciences
Monash University, Parkville Campus
Level 2, Building 3, Room 3.208
381 Royal Parade, Parkville
Victoria 3052, Australia

Phone: 0448671141

https://www.monash.edu/researchinfrastructure/proteomics