**Analyzing data with Metaboanalyst 4.0**

Ion features in liquid chromatography-mass spectrometry (LC-MS) datasets are aligned with regard to their retention time and the mass-to-charge (*m/z*) ratio. This analysis of datasets is carried out using MZmine (<http://mzmine.github.io/>) (as in Dr. Xiuxia Du’s presentation) or by XCMS. The latter can be run in R, or (in many labs) using an online version of XCMS (<https://xcmsonline.scripps.edu>). Details of the features of XCMSonline can be obtained after setting up a registration account.

Information about the peaks identified using XCMS can be downloaded as a .zip file. Once unzipped, there is an Excel file that contains information on the *m/z* and retention time values of each ion feature, univariate statistical analysis (p-values and q-scores), and the area under the curve of each ion feature in each individual experimental sample.

To prepare for Metaboanalyst, .csv files are created for each experimental sample. These contain three columns – the median *m/z* values, the median retention time values and the areas under the peaks. IMPORTANT: - the names used for each of the files should not contain spaces (an underscore should be used instead). The .csv files corresponding to a group are next placed in a folder (again, no spaces in the name used). This should be repeated for all groups. The folders are then zipped and the name given to the .zip file should again not contain a space. This is the file to be used by Metaboanalyst.

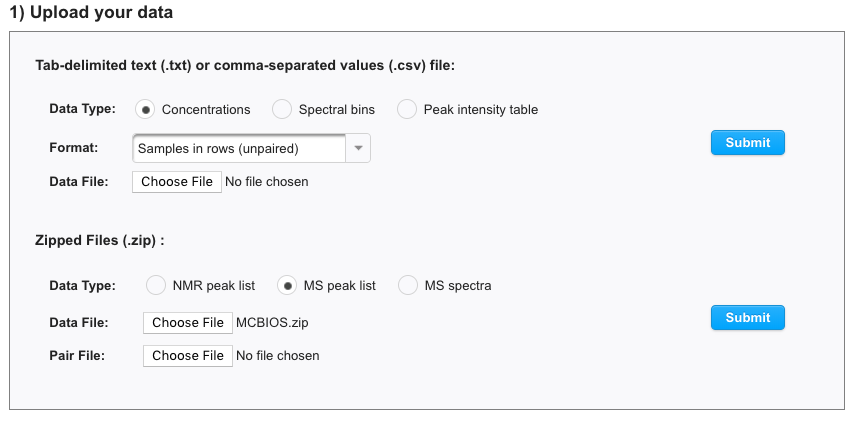
**Uploading a .zip file to Metaboanalyst 4.0**

Set your browser to <https://www.metaboanalyst.ca/faces/ModuleView.xhtml>



and then select the statistical analysis module –

<https://www.metaboanalyst.ca/faces/upload/StatUploadView.xhtml>.

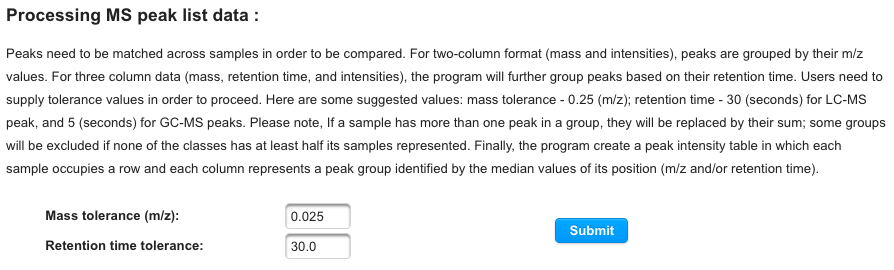


Use the Zipped files option – first select “MS peak list” and then from “Choose file” browse to locate the .zip file. Then press **submit**.

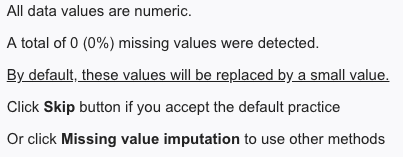
Once loaded, entries are needed for mass and retention time tolerance. Since we set these to the same values for each .csv file, enter 0.001 for mass tolerance and 0.005 min for retention time tolerance (do not use zeros since these are not allowed entries). **Submit**.

0.005

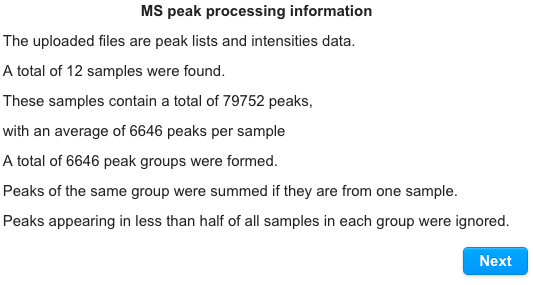
0.001



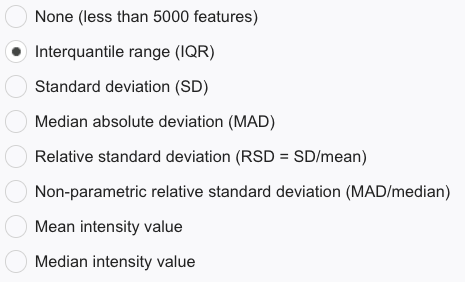
The summary will indicate that 12 sample files were loaded with 6,646 peak groups per file. A check is made for missing values – XCMS always puts a number into the Excel output file, so there is no need to do anything and therefore press SKIP.





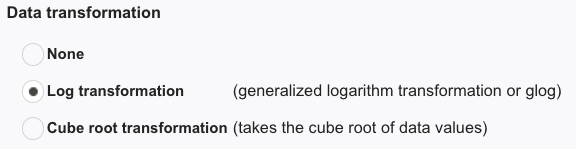


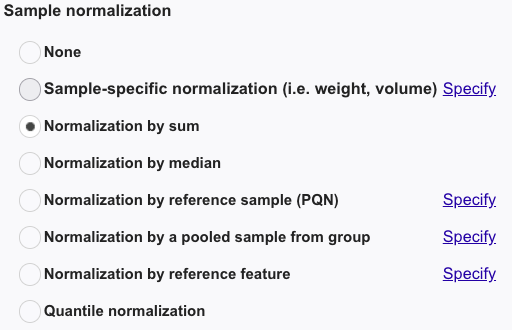
Since the number of features exceeds 5000, interquartile range (IQR) filtering is necessary. After selecting the IQR button, press SUBMIT. Then press PROCEED.

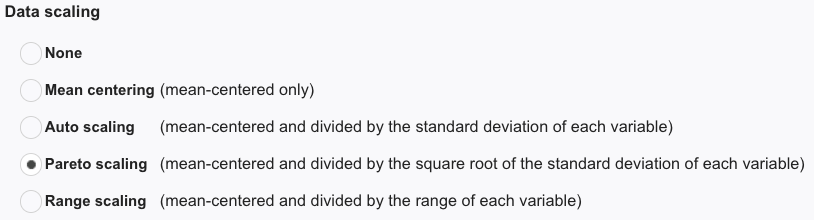




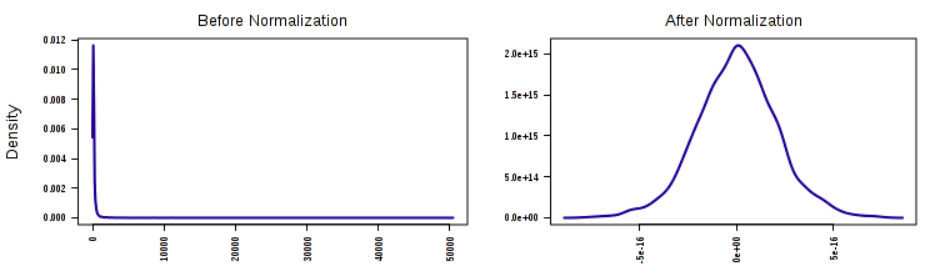
The next steps involve data normalization. First use “normalization by sum”. Second, transform the data using “log transformation”. Thirdly, use “Pareto scaling”. These three selections deal with (1) variation (in this case) with urine volume differences between animals, (2) a need to ensure the data have a normal distribution (required for statistical analysis) and (3) to mean center the data to overcome large differences between individual ion features.

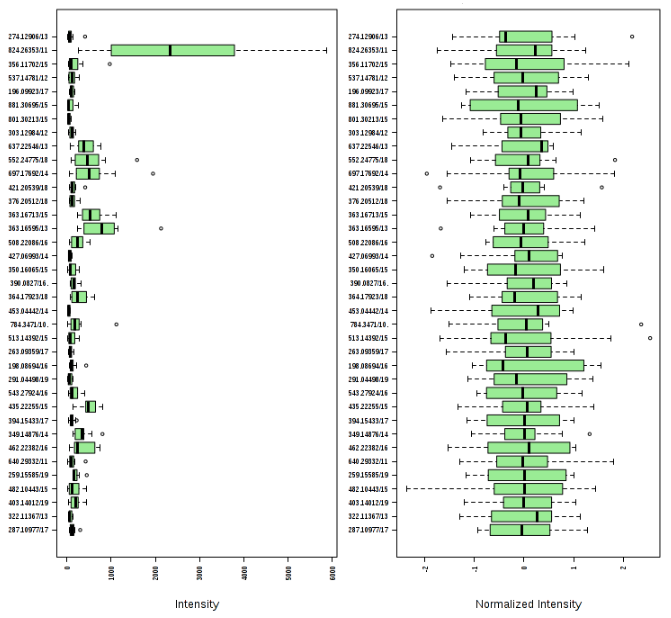






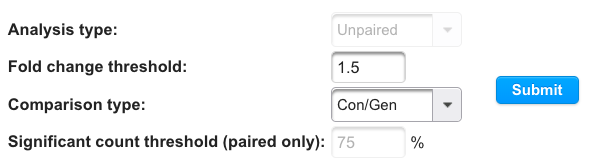
Now press **NORMALIZE**, **VIEW RESULT** and **PROCEED**.

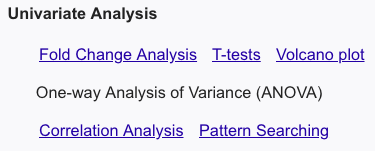




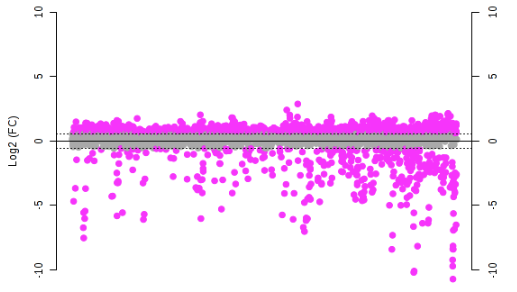
Clicking onto PROCEED brings you into the statistical analysis section of Metaboanalyst.

Univariate Analysis



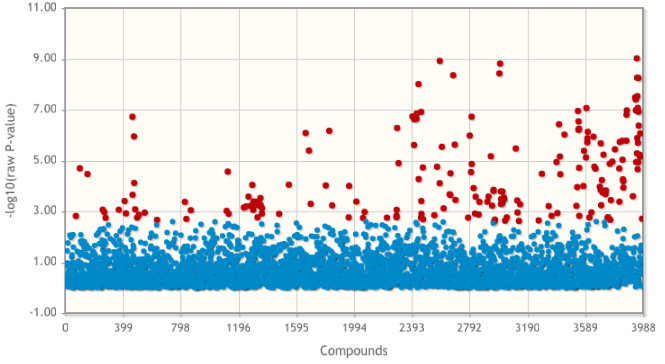


*Fold change* – update fold change threshold to 1.5 and **SUBMIT**.

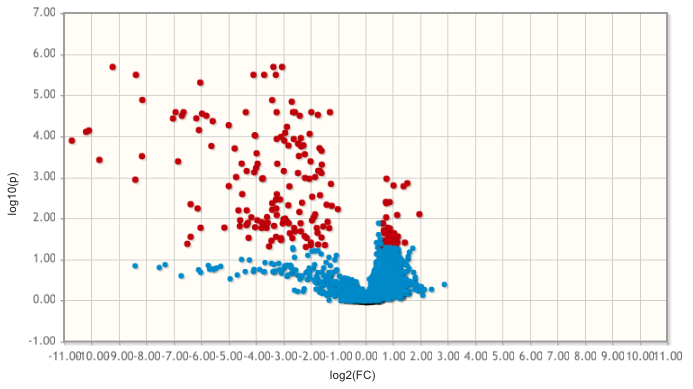


*T-test* – the assumption is that group variance is equal and adjusted P-value (for multiple testing) is 0.05.

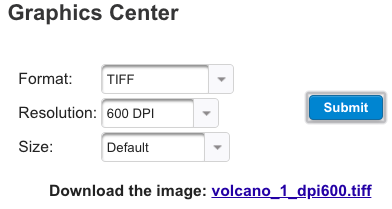


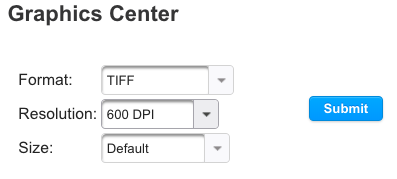


*Volcano plot* – reset to fold-change to 1.5 and p-value to 0.05. Then **Submit**.



A publication quality figure can be generated by clicking on the palette on the upper right of the Volcano plot. Set format to TIFF and resolution to 600 dpi – press **SUBMIT**. Click on volcano\_1\_dpi600.tiff to download figure.





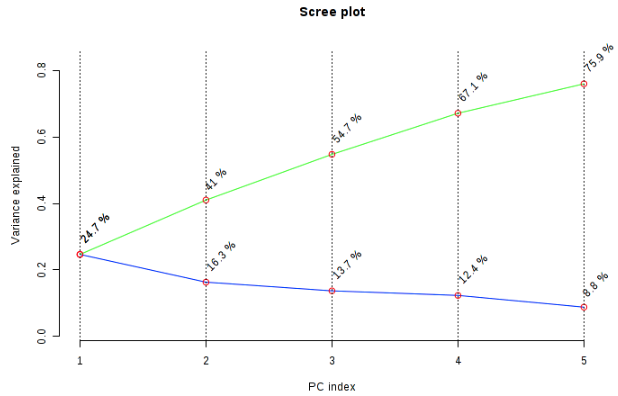


Multivariate Analysis

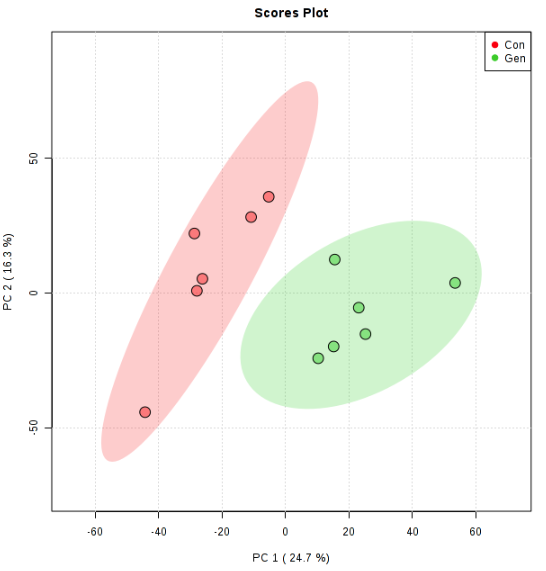
In univariate analysis, the ion features are considered to be independent. However, this is very unlikely in a biological system and therefore multivariate analysis is needed.

The data are first analyzed by *Principal Components Analysis* (PCA). This is an unsupervised analysis – taking the data from both groups as if they were one, principal component 1 (PC1) comes from a line that bisects the data to produce to the maximum error differences between the line and the data. Subsequent principal components represent lines drawn through the data orthogonal to PC1.

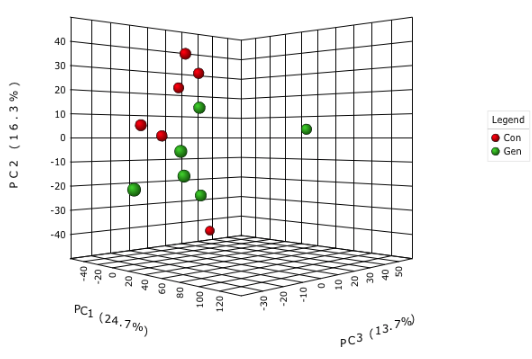
A Scree plot summarizes the accumulation of explained error.



The 2D Scores plot plots PC2 versus PC1. Individual samples are color-coded to represent the two groups and a 95% confidence limit is drawn around the data in each group.



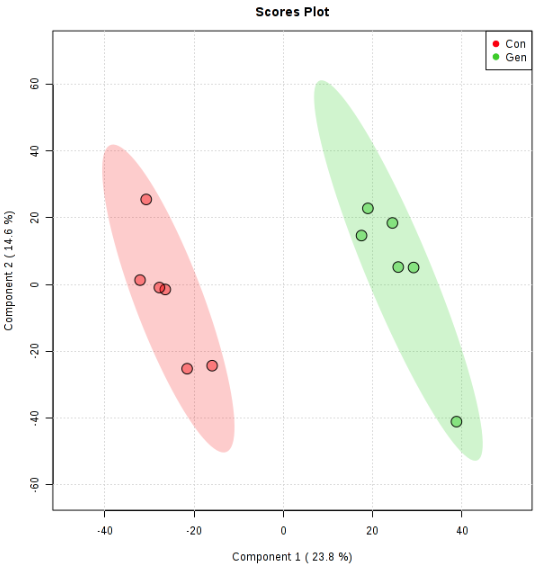
Use the palette to make a publication quality figure. Press **SUBMIT**.



A 3D-view of the PCA plot using PC1, PC2 and PC3 can be viewed – it’s rotatable! However, it cannot be saved into a publication quality figure.

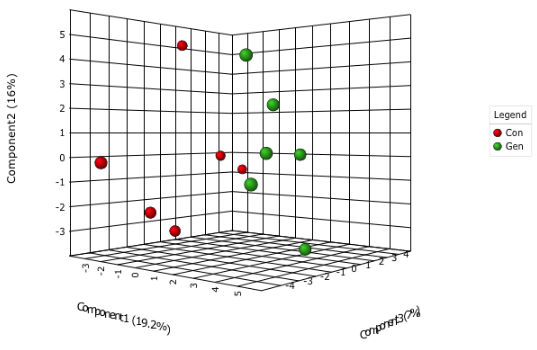
In the Loadings plot, most ions are centered around 0,0. Clicking on ions that are most away from 0,0 produces box-and-whisker plots for individual ions.

*Partial least squares-discriminant analysis* (PLSDA) is a supervised analysis that includes information that there are (in this case) two groups.

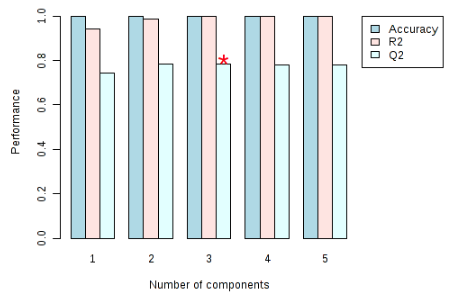


The *2D-scores plot* shows a complete separation of the two groups.

Use the palette to make a publication quality figure. Press **SUBMIT**.



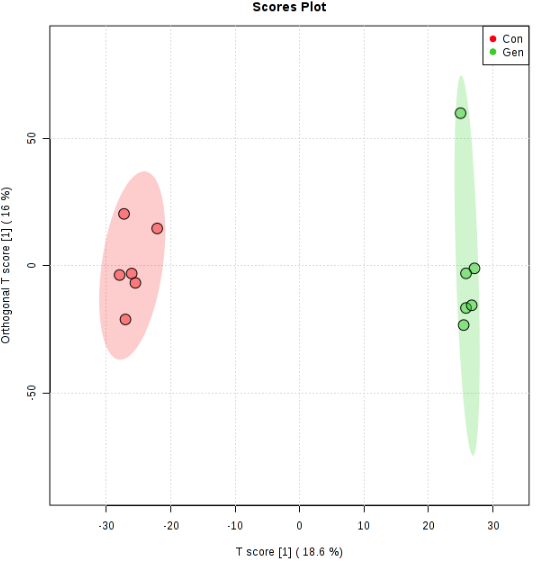
The *3D-PLSDA scores plot* shows improved resolution over the 3D-PCA plot.

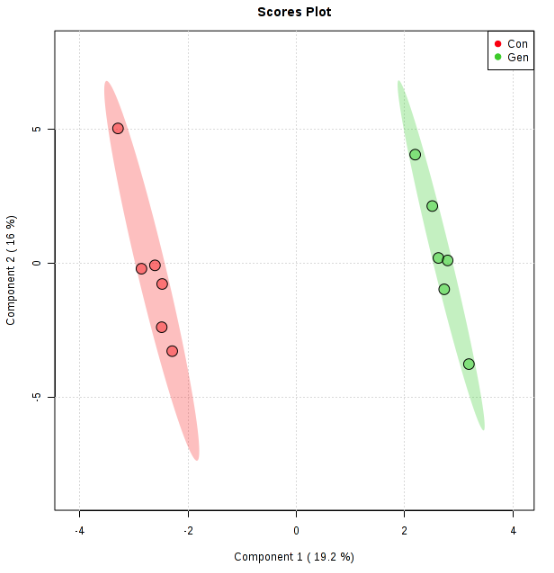


*Cross validation* reveals that there are 3 principal components and that 3D-plots are the best representation of the data.

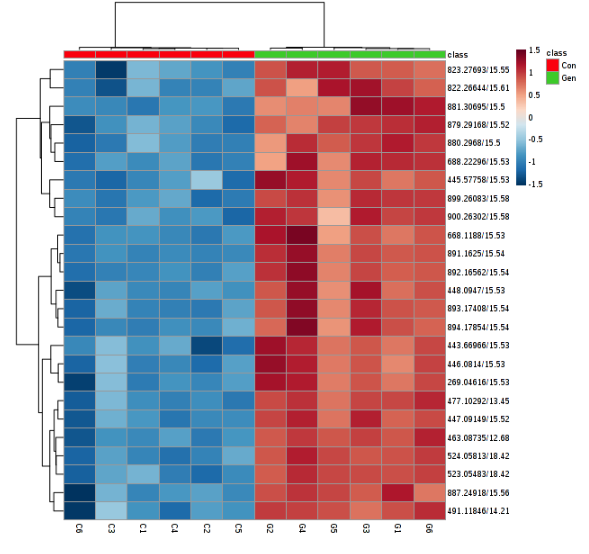
To identify the ion features most contributing to the differences between the two groups, the *Variable importance in projection* values are calculated (VIP scores). Usually these are selected from component 1. Increase the number to 25 shown over the default number (15) and press UPDATE. Use the palette to make a publication quality figure. Press **SUBMIT**.

Two other multivariate analyses that are frequently found in publications on metabolomics are sparse PLS-DA and orthogonal PLSDA. These are both available in Metaboanalyst 4.0.

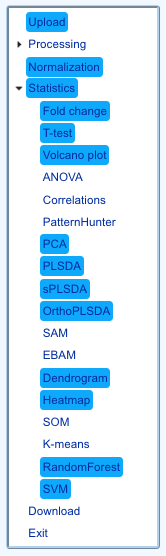




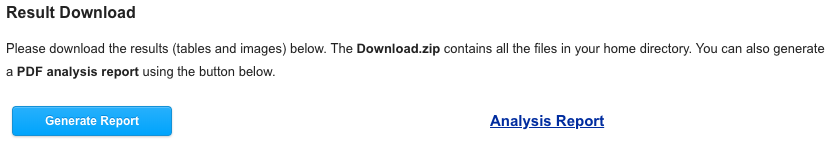
Heatmaps



Metaboanalyst allows for ion feature clustering and representation of the differences between the two groups in shades of blue and red. Select the option to show the top 25 ion features. These show the very clear differences between the two groups.



A nice feature of Metaboanalyst is that when you’ve finished the analyses you wanted to do, you’re able to automatically compile them into a downloadable file. A written compiled report in PDF is obtained by pressing “Generate Report”.





The report also contains the R-commands that were invoked in the Metaboanalyst session.

Finally, remember to logout.