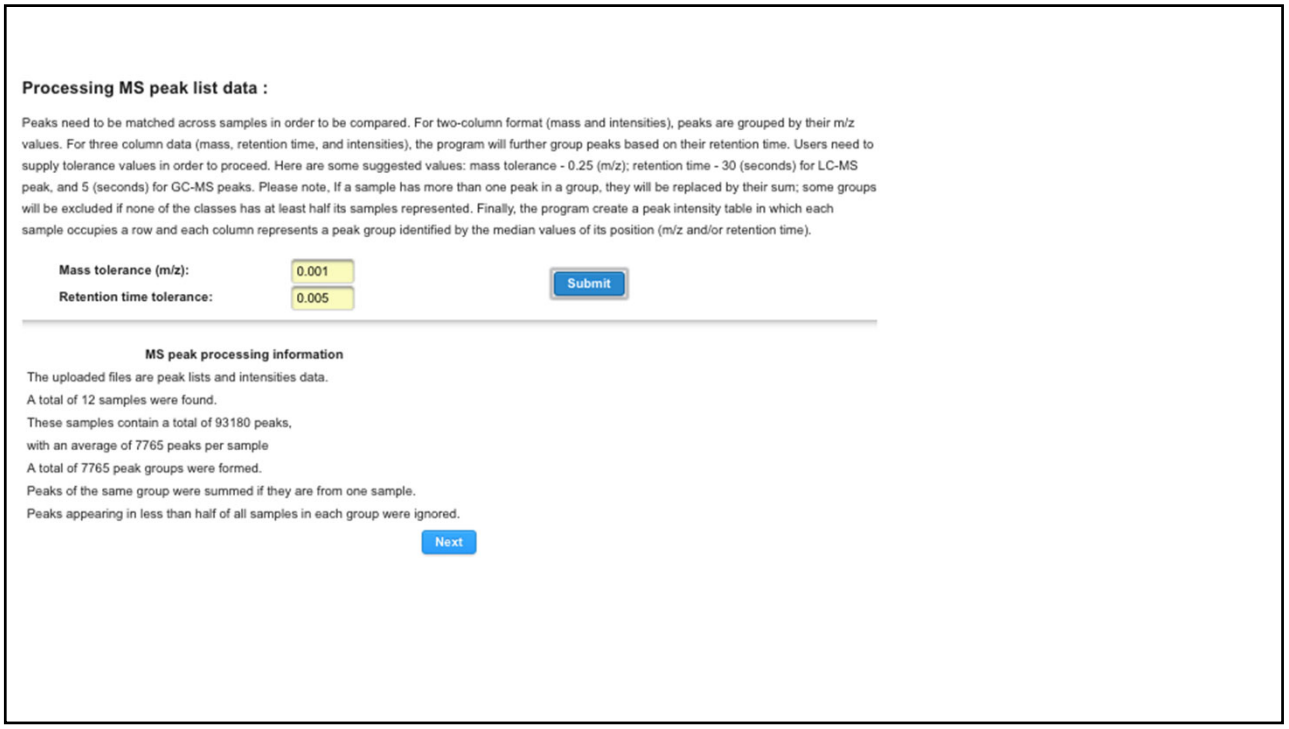
Statistical analysis with Metaboanalyst 4.0

Stephen Barnes



Processing MS peak list data :

Peaks need to be matched across samples in order to be compared. For two-column format (mass and intensities), peaks are grouped by their m/z values. For three column data (mass, retention time, and intensities), the program will further group peaks based on their retention time. Users need to supply tolerance values in order to proceed. Here are some suggested values: mass tolerance - 0.25 (m/z); retention time - 30 (seconds) for LC-MS peak, and 5 (seconds) for GC-MS peaks. Please note, If a sample has more than one peak in a group, they will be replaced by their sum; some groups will be excluded if none of the classes has at least half its samples represented. Finally, the program create a peak intensity table in which each sample occupies a row and each column represents a peak group identified by the median values of its position (m/z and/or retention time).

Mass tolerance (m/z): 0.001

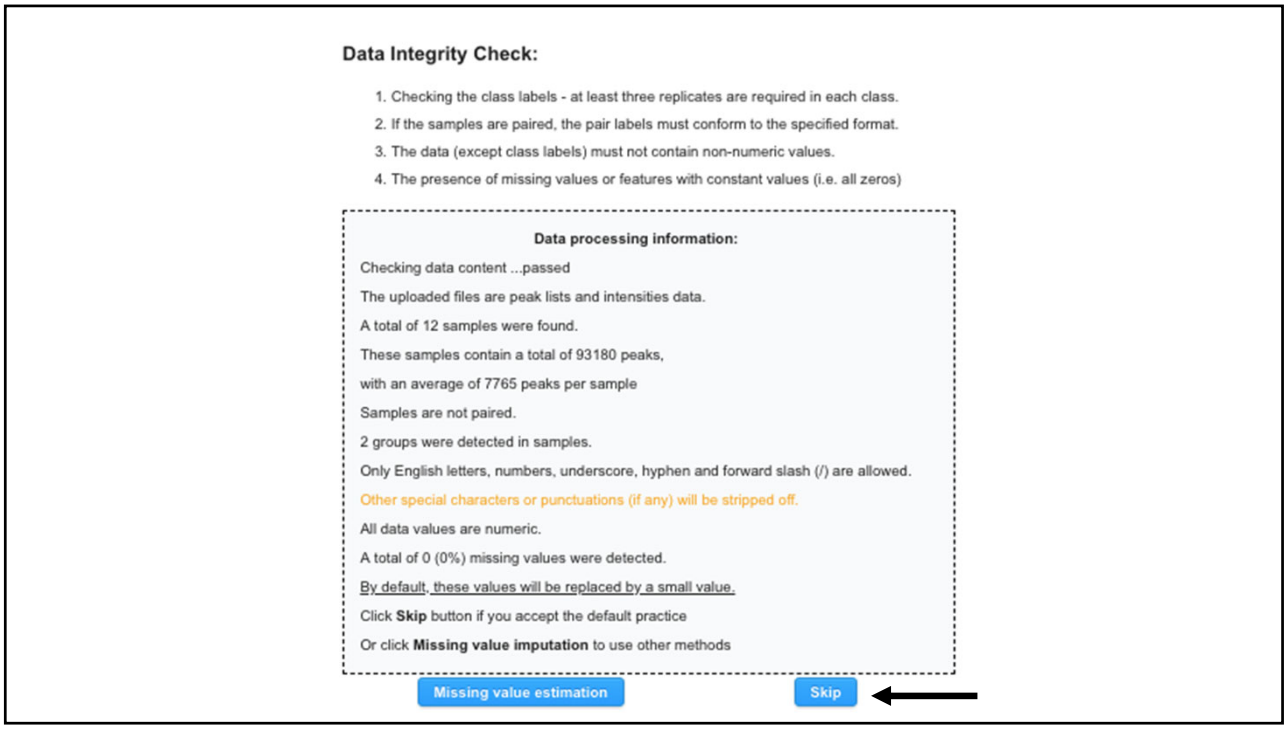
Retention time tolerance: 0.005

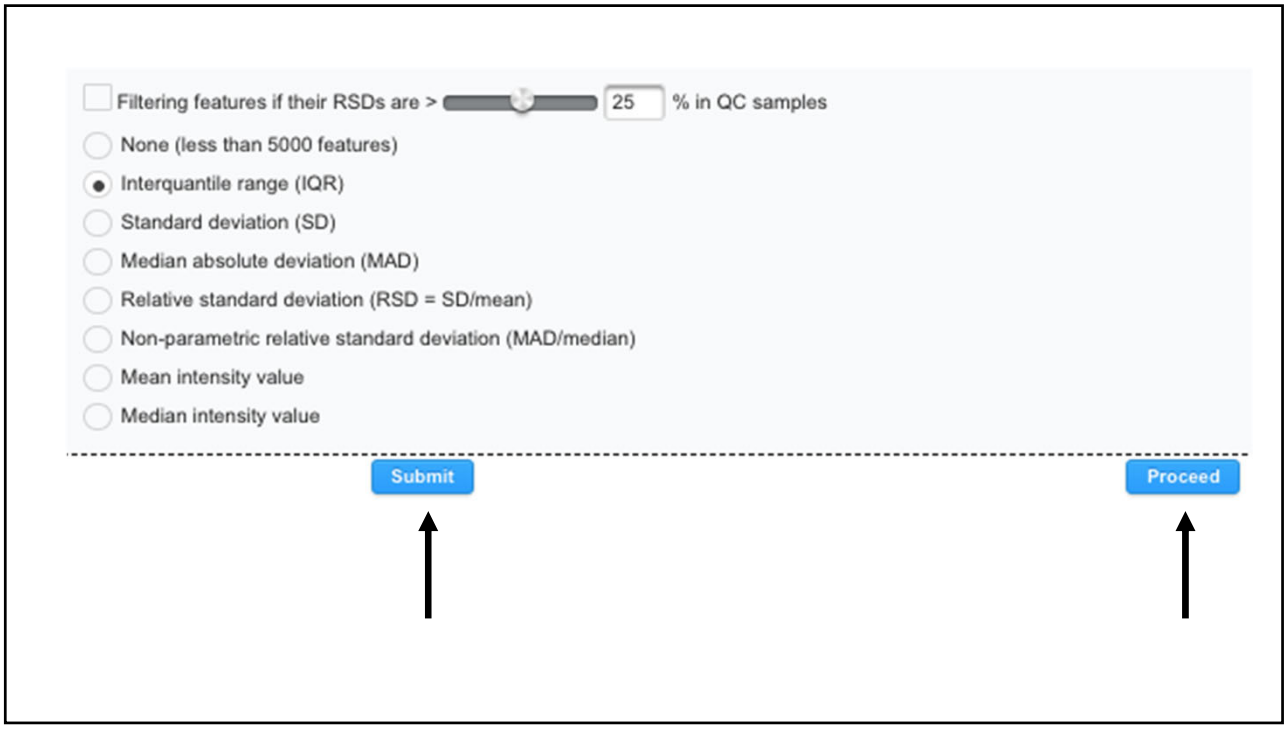
Submit

MS peak processing information

The uploaded files are peak lists and intensities data.
A total of 12 samples were found.
These samples contain a total of 93180 peaks,
with an average of 7765 peaks per sample
A total of 7765 peak groups were formed.
Peaks of the same group were summed if they are from one sample.
Peaks appearing in less than half of all samples in each group were ignored.

Next


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
The uploaded files are peak lists and intensities data.
A total of 12 samples were found.
These samples contain a total of 93180 peaks,
with an average of 7765 peaks per sample
Samples are not paired.
2 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, these values will be replaced by a small value.
Click Skip button if you accept the default practice
Or click Missing value imputation to use other methods
Missing value estimation
Skip


Filtering features if their RSDs are > 25 % in QC samples
None (less than 5000 features)
Interquantile range (IQR)
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

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

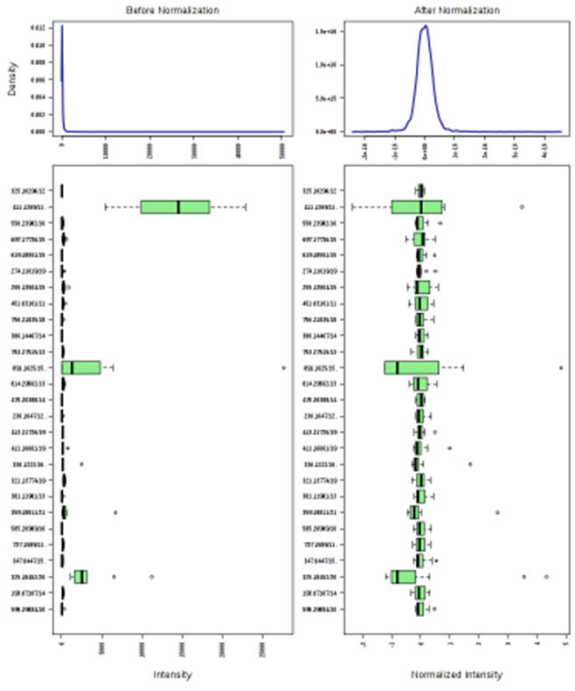
Data transformation

- **None**
- Log transformation (generalized logarithm transformation or glog)
- Cube root transformation (takes the cube root of data values)

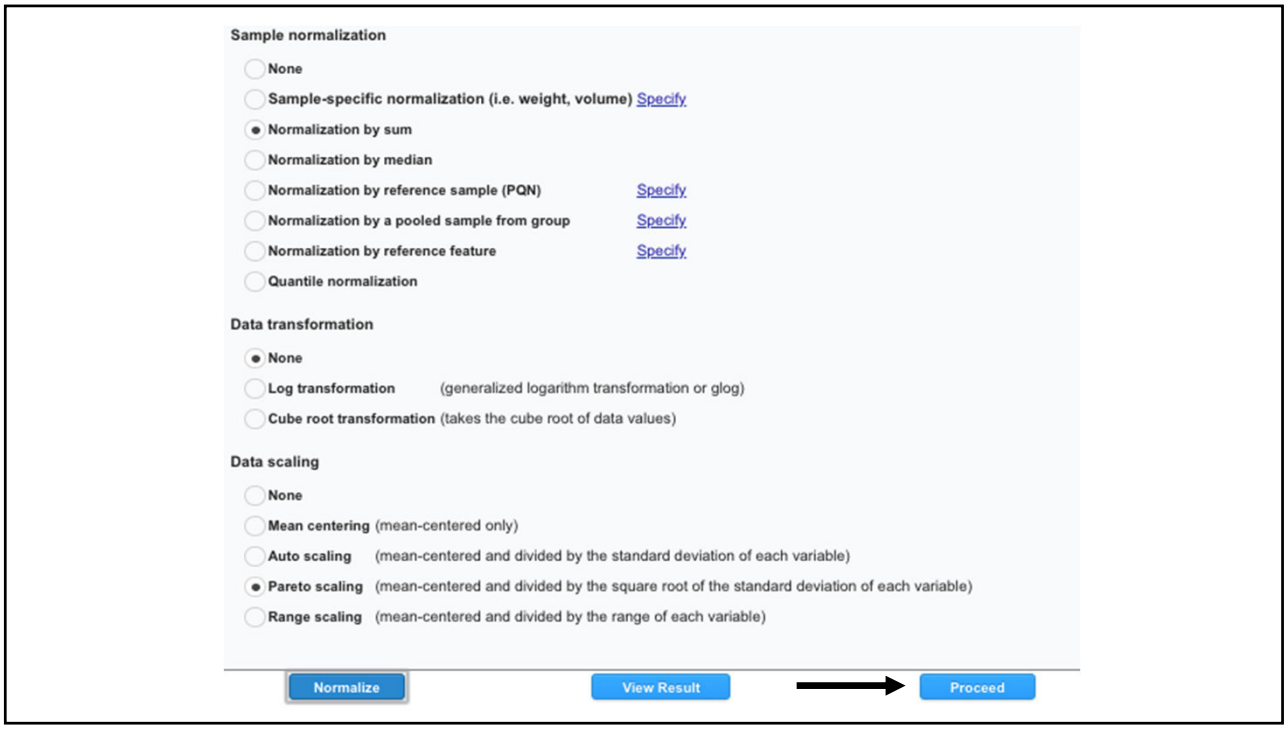
Data scaling

- 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)

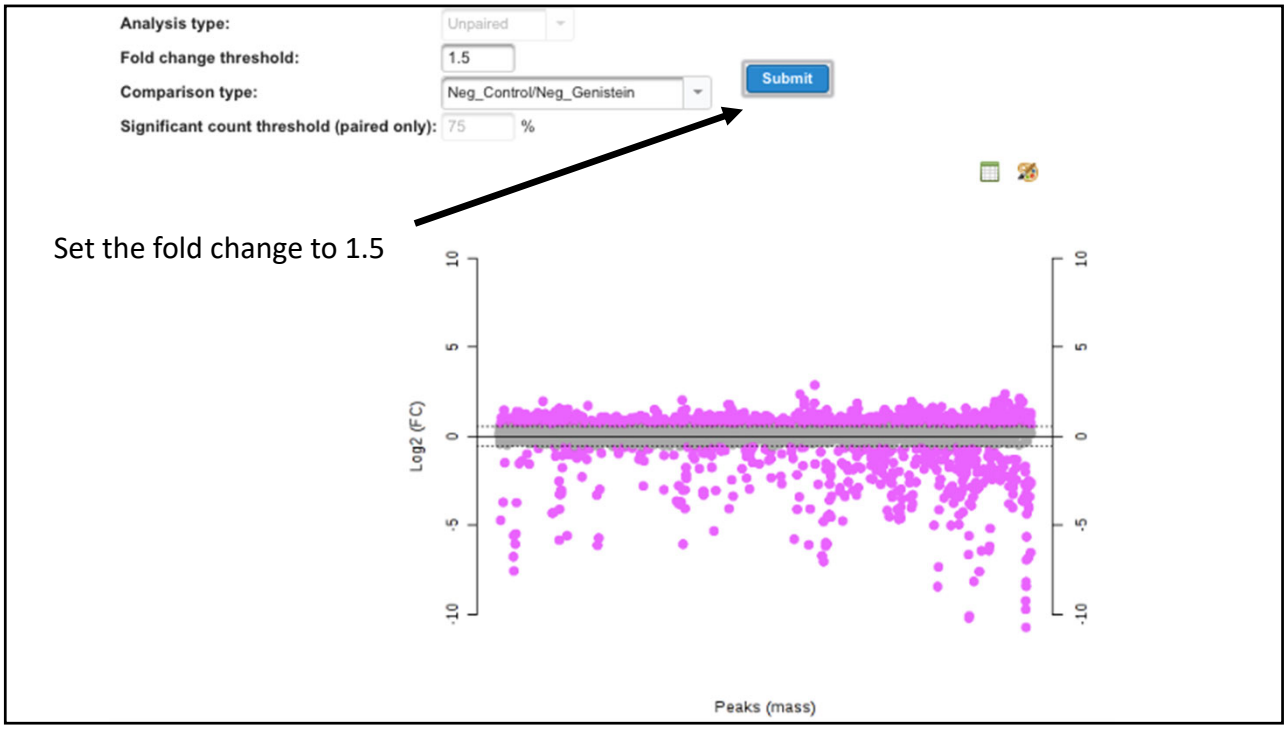
Normalize → View Result → Proceed



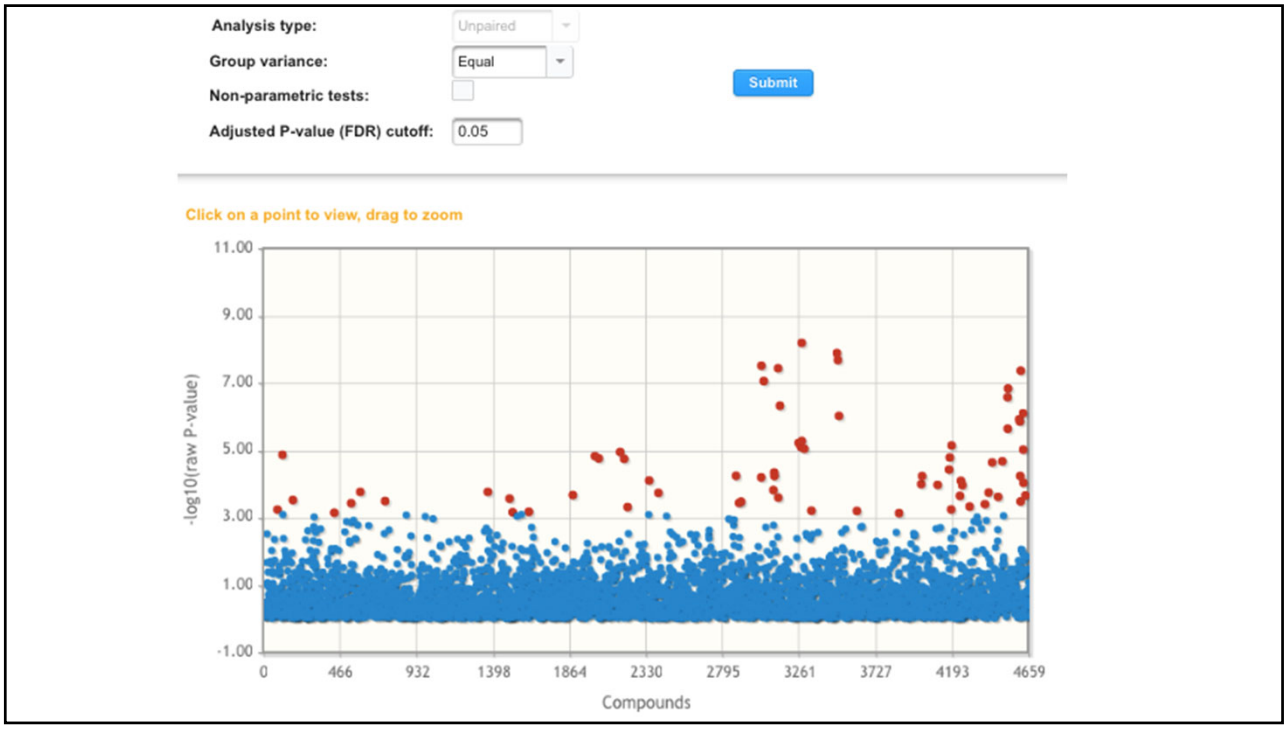
Exit this page


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

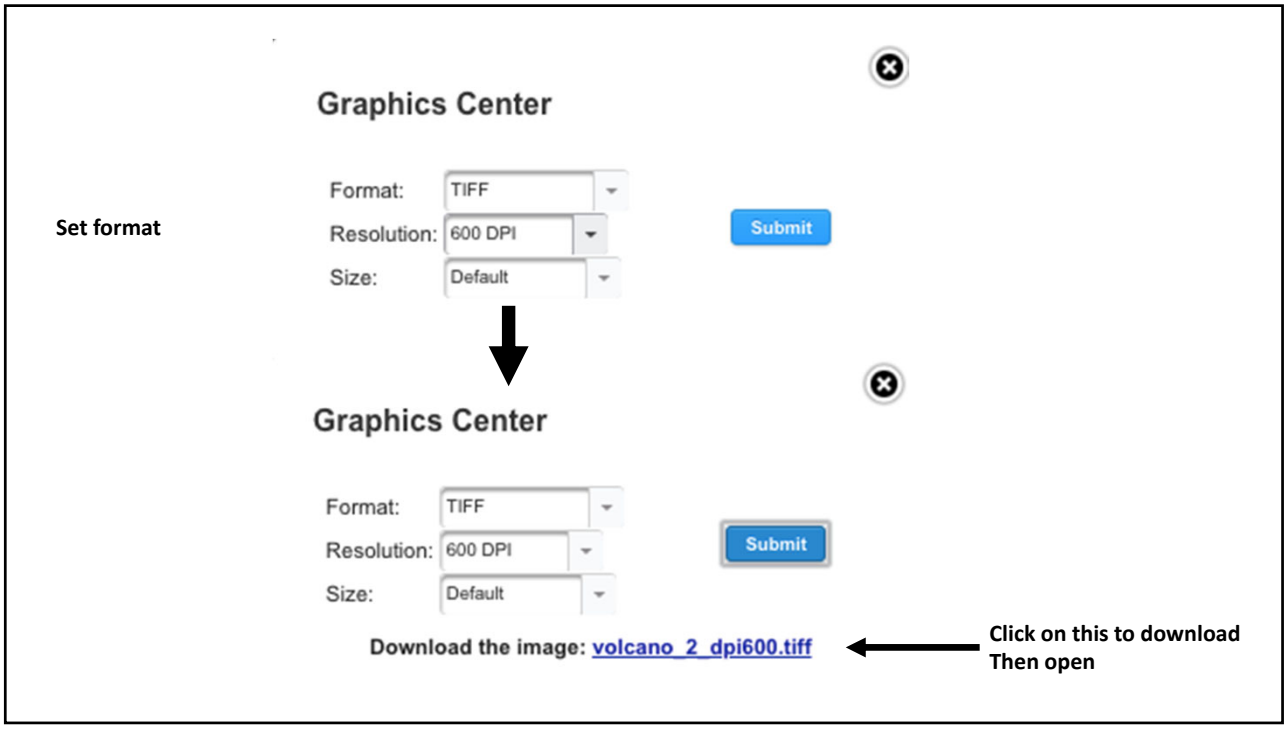
Univariate Analysis
Fold Change Analysis T-tests Volcano plot
One-way Analysis of Variance (ANOVA)
Correlation Analysis Pattern Searching
Chemometrics Analysis
Principal Component Analysis (PCA)
Partial Least Squares - Discriminant Analysis (PLS-DA)
Sparse Partial Least Squares - Discriminant Analysis (sPLS-DA)
Orthogonal Partial Least Squares - Discriminant Analysis (orthoPLS-DA)
Feature Identification
Significance Analysis of Microarray (and Metabolites) (SAM)
Empirical Bayesian Analysis of Microarray (and Metabolites) (EBAM)
Cluster Analysis
Hierarchical Clustering: Dendrogram Heatmaps
Partitional Clustering: K-means Self Organizing Map (SOM)
Classification & Feature Selection
Random Forest
Support Vector Machine (SVM)

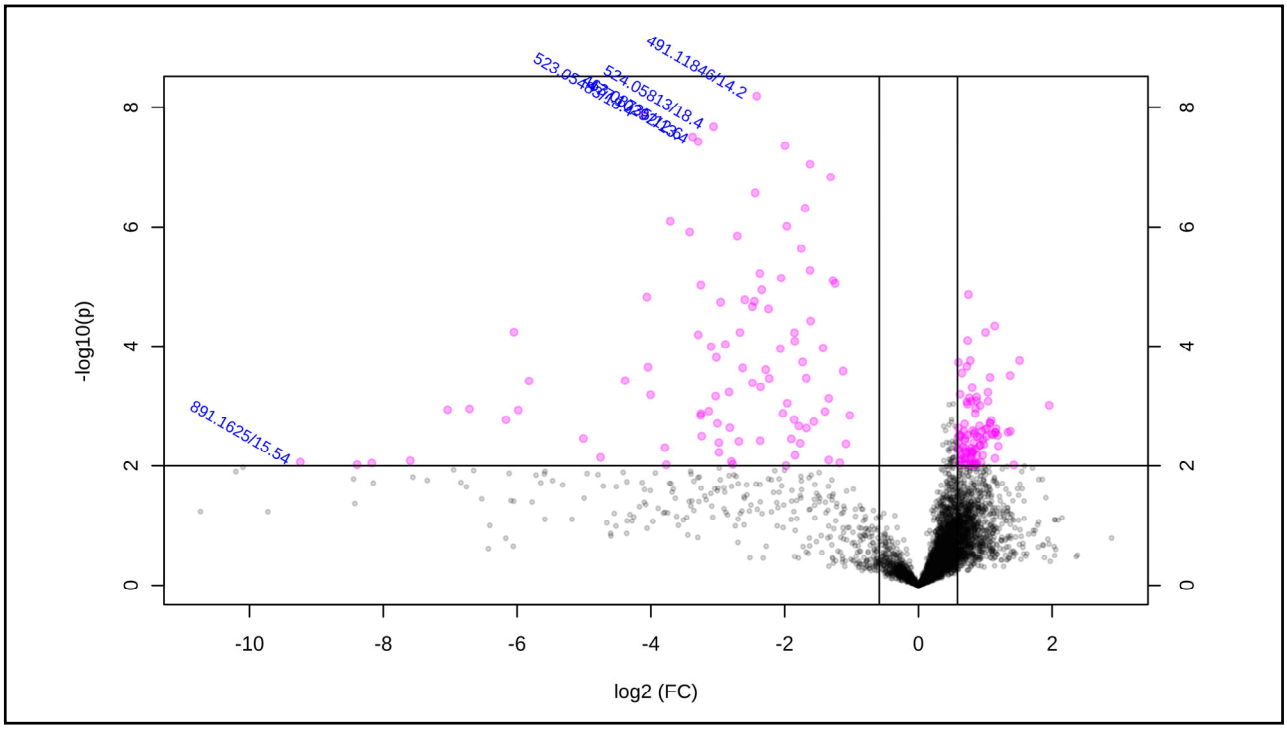


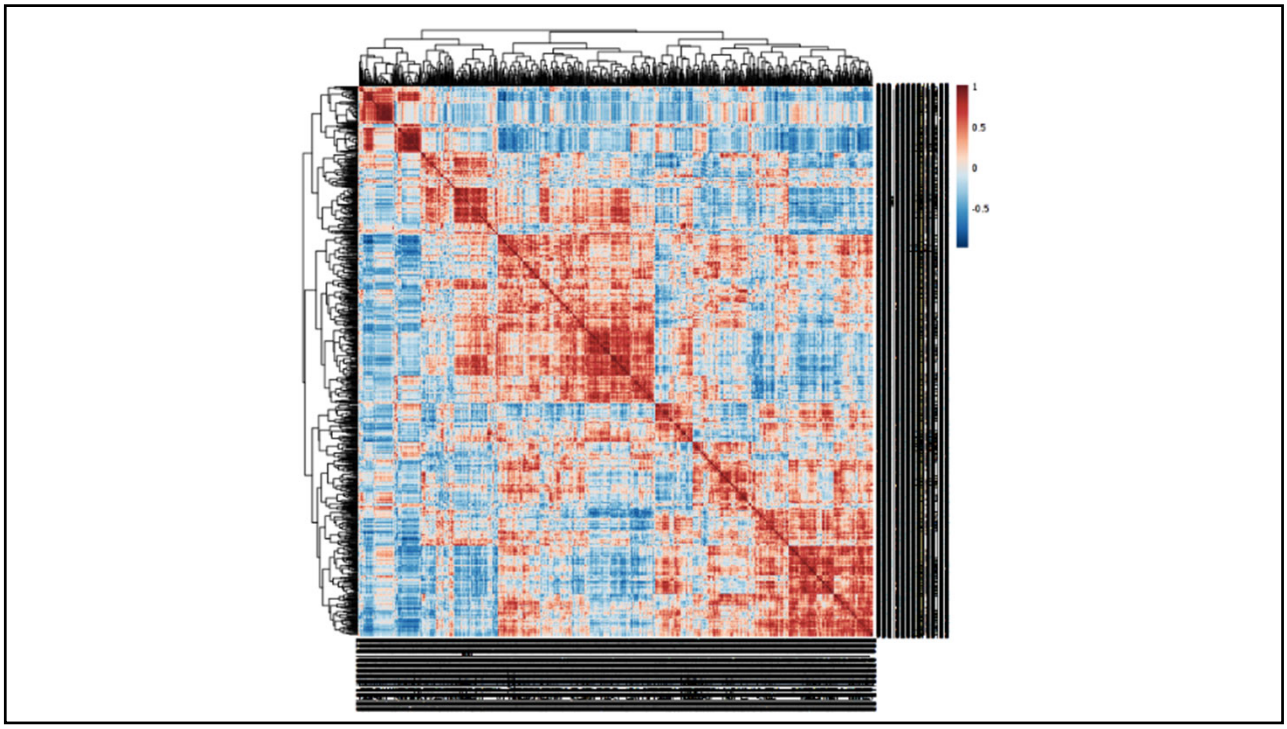
Set the fold change to 1.5

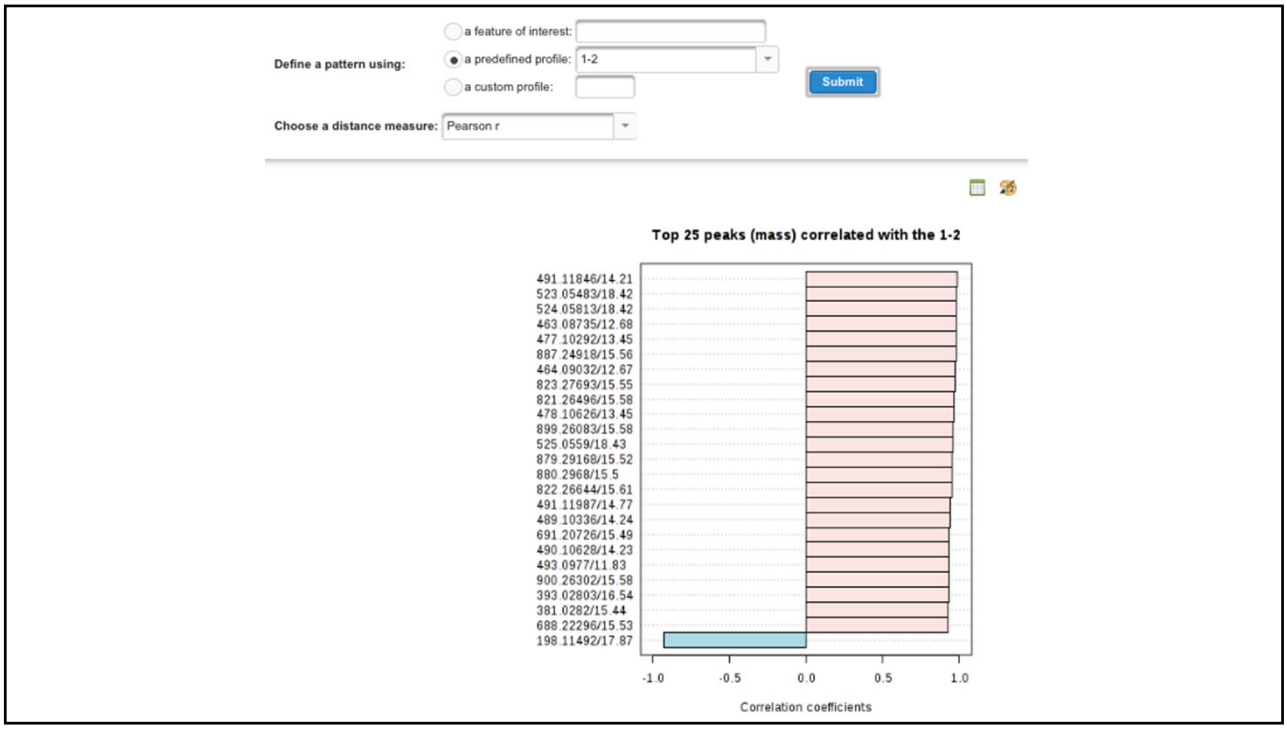


Analysis:
Unpaired
Plot label:
Yes
No
(used for download image only)
Fold change threshold:
1.5
X-axis:
Comparison type:
Neg_Control/Neg_Ger
Sig. count threshold (paired):
75.0
%
Non-parametric tests:
Y-axis:
P-value threshold:
0.01
Raw
FDR-ajusted
Group variance:
Equal
Submit
Reset these
Click on a point to view, drag to zoom
Reset
log10(p)
log2(FC)

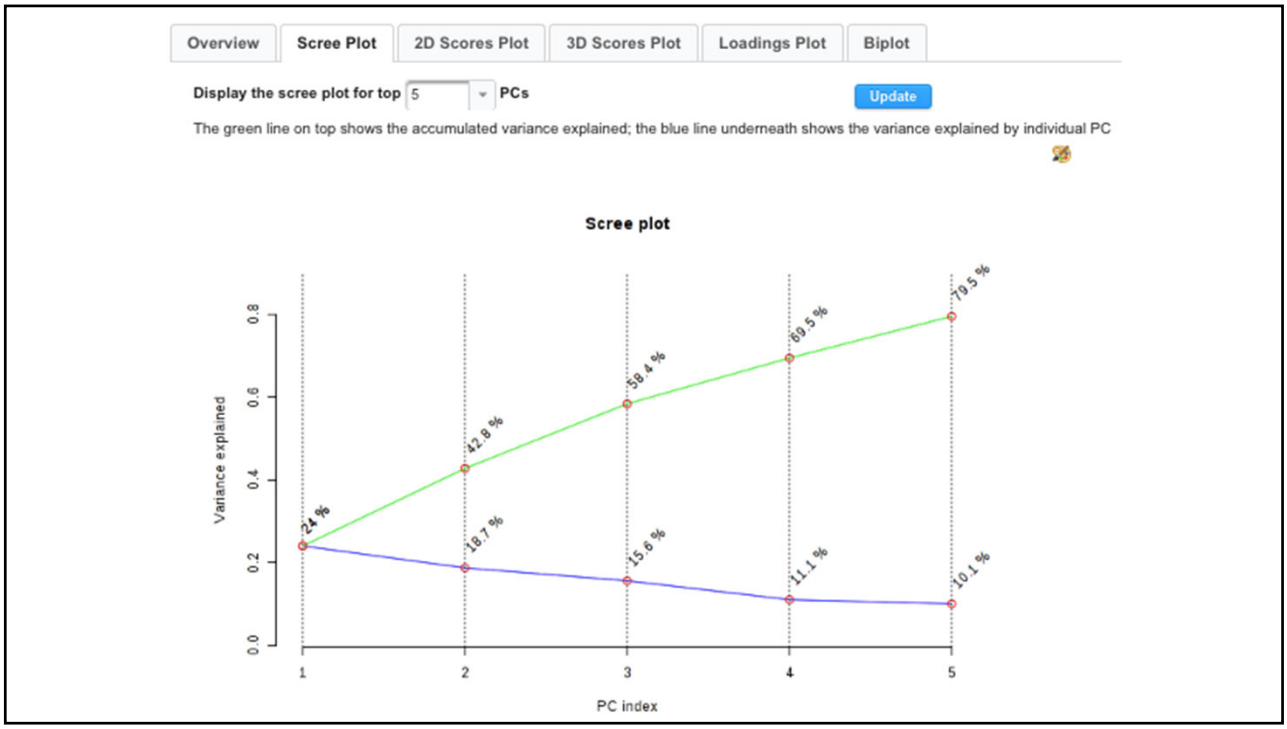

Set format
Graphics Center
Format:
TIFF
Resolution:
600 DPI
Size:
Default
Submit
Graphics Center
Format:
TIFF
Resolution:
600 DPI
Size:
Default
Submit
Download the image: volcano_2_dpi600.tiff
Click on this to download
Then open

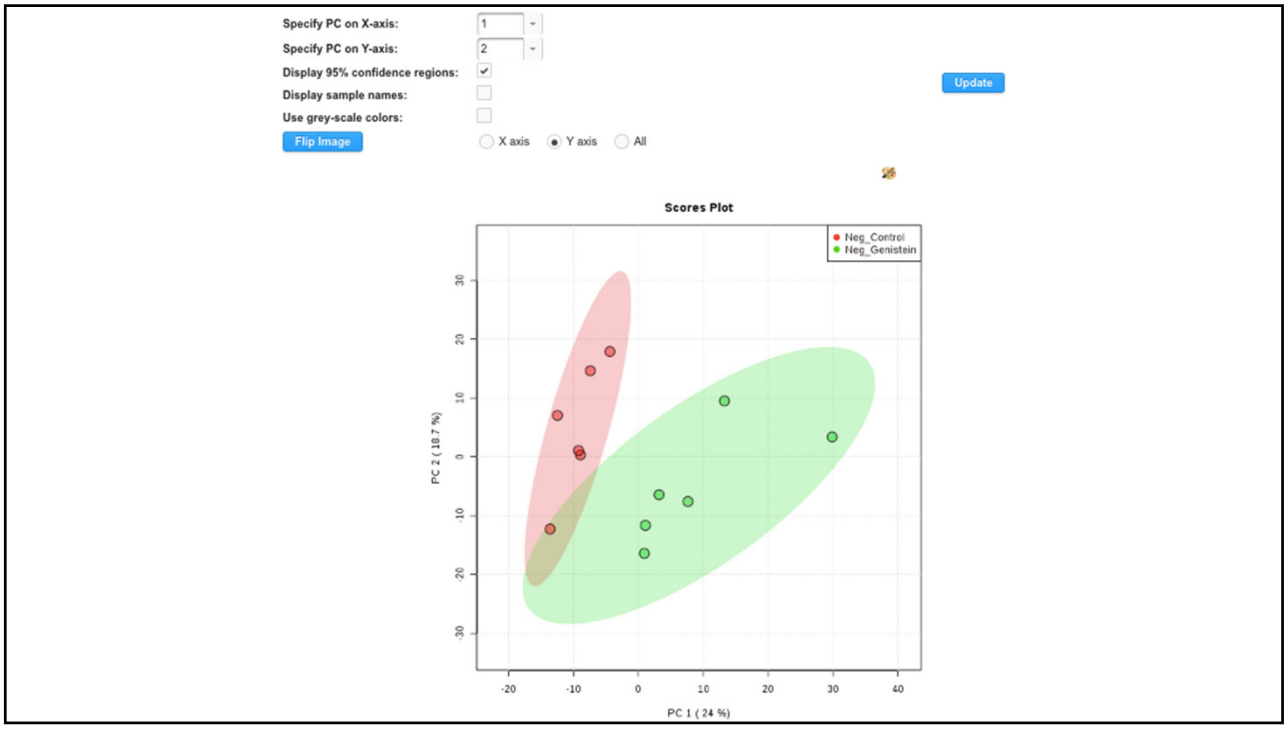

491.11846/14.2
524.05813/18.4
891.1625/15.54
-log10(p)
log2 (FC)
0
2
4
6
8
-10
-8
-6
-4
-2
0
2

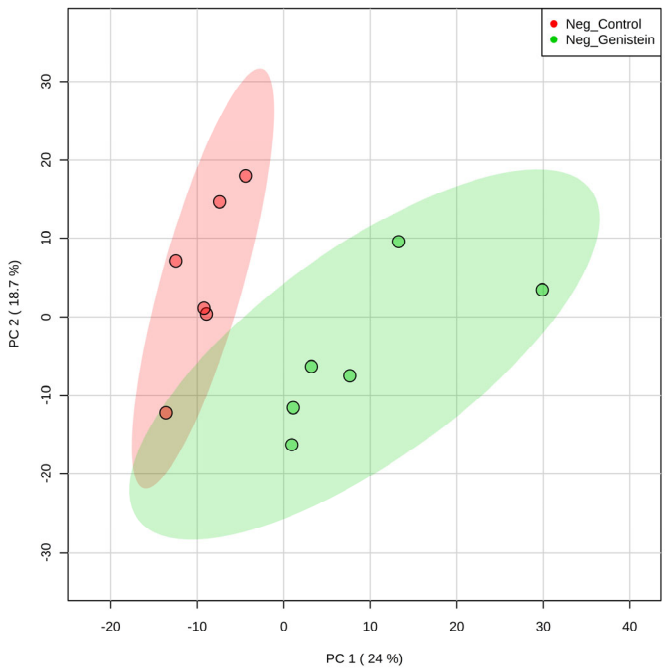



Define a pattern using:
a feature of interest:
a predefined profile: 1-2
a custom profile:
Submit
Choose a distance measure: Pearson r
Top 25 peaks (mass) correlated with the 1-2
491.11846/14.21
523.05483/18.42
524.05813/18.42
463.08735/12.68
477.10292/13.45
887.24918/15.56
464.09032/12.67
823.27693/15.55
821.26496/15.58
478.10626/13.45
899.26083/15.58
525.0559/18.43
879.29168/15.52
880.2968/15.5
822.26644/15.61
491.11987/14.77
489.10336/14.24
691.20726/15.49
490.10628/14.23
493.0977/11.83
900.26302/15.58
393.02803/16.54
381.0282/15.44
688.22296/15.53
198.11492/17.87
-1.0
-0.5
0.0
0.5
1.0
Correlation coefficients

Overview
Scree Plot
2D Scores Plot
3D Scores Plot
Loadings Plot
Biplot
Display pairwise score plot for top 5 PCs
Update
PC 1
24 %
PC 2
18.7 %
PC 3
15.6 %
PC 4
11.1 %
PC 5
10.1 %

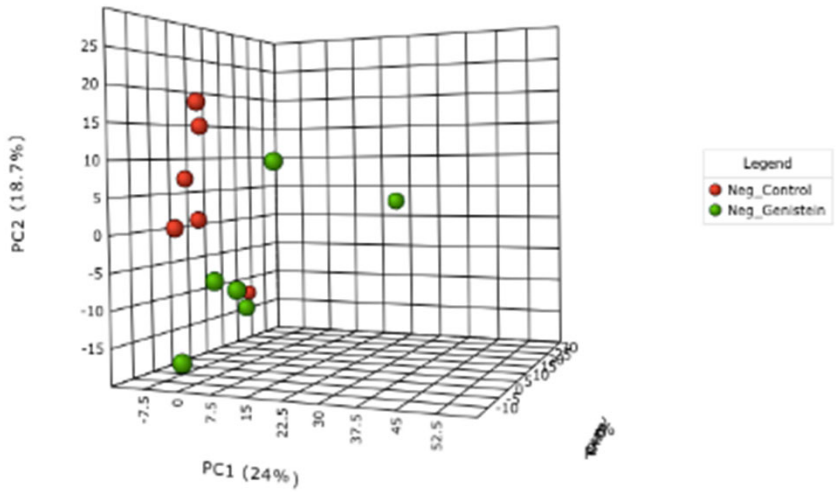

Overview
Scree Plot
2D Scores Plot
3D Scores Plot
Loadings Plot
Biplot
Display the scree plot for top 5 PCs
Update
The green line on top shows the accumulated variance explained; the blue line underneath shows the variance explained by individual PC
Scree plot
Variance explained
PC index
24 %
42.8 %
58.4 %
69.5 %
79.5 %
18.7 %
15.6 %
11.1 %
10.1 %

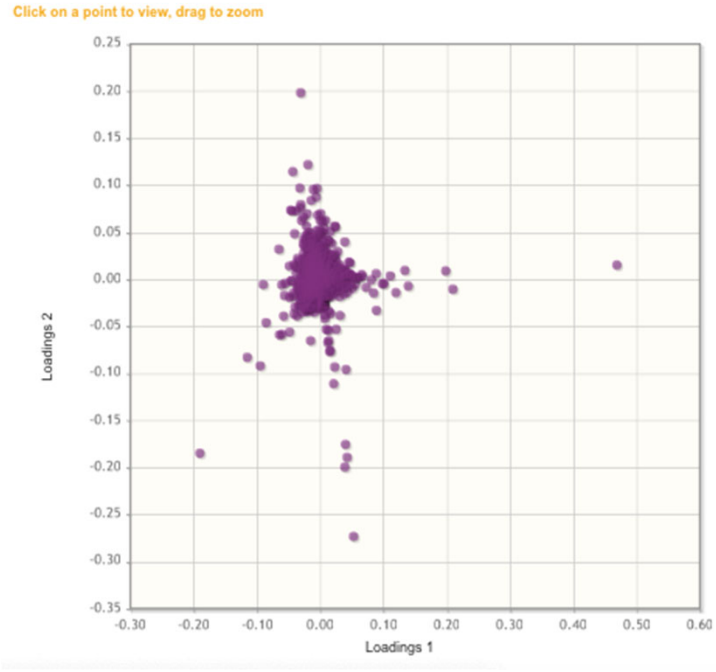

Specify PC on X-axis: 1
Specify PC on Y-axis: 2
Display 95% confidence regions:
Display sample names:
Use grey-scale colors:
Flip Image
X axis
Y axis
All
Update
Scores Plot
Neg_Control
Neg_Genistein
PC 2 (18.7 %)
PC 1 (24 %)



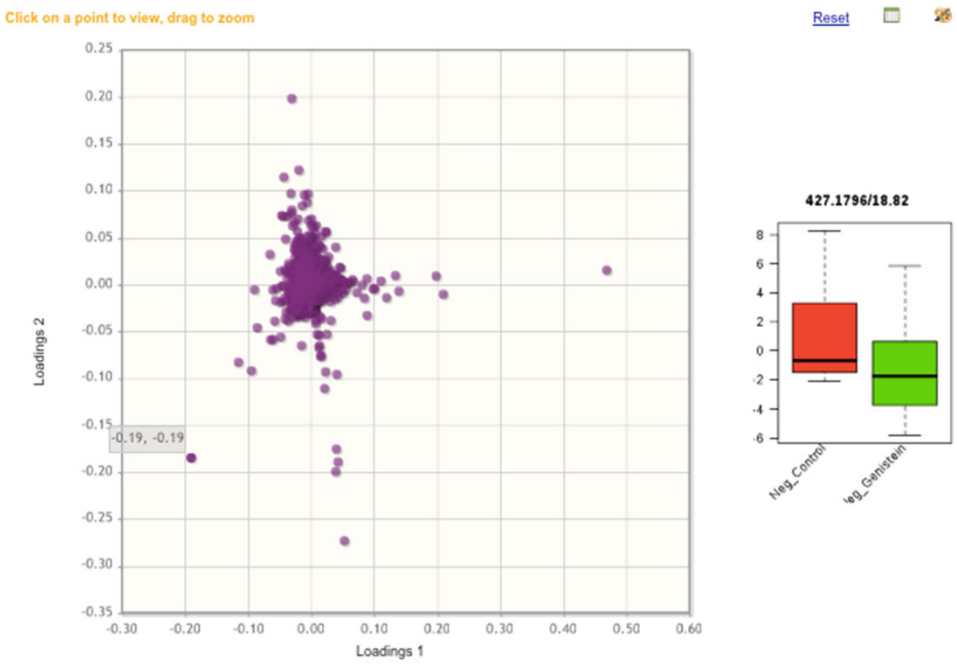
This is the TIFF file for publication quality figures

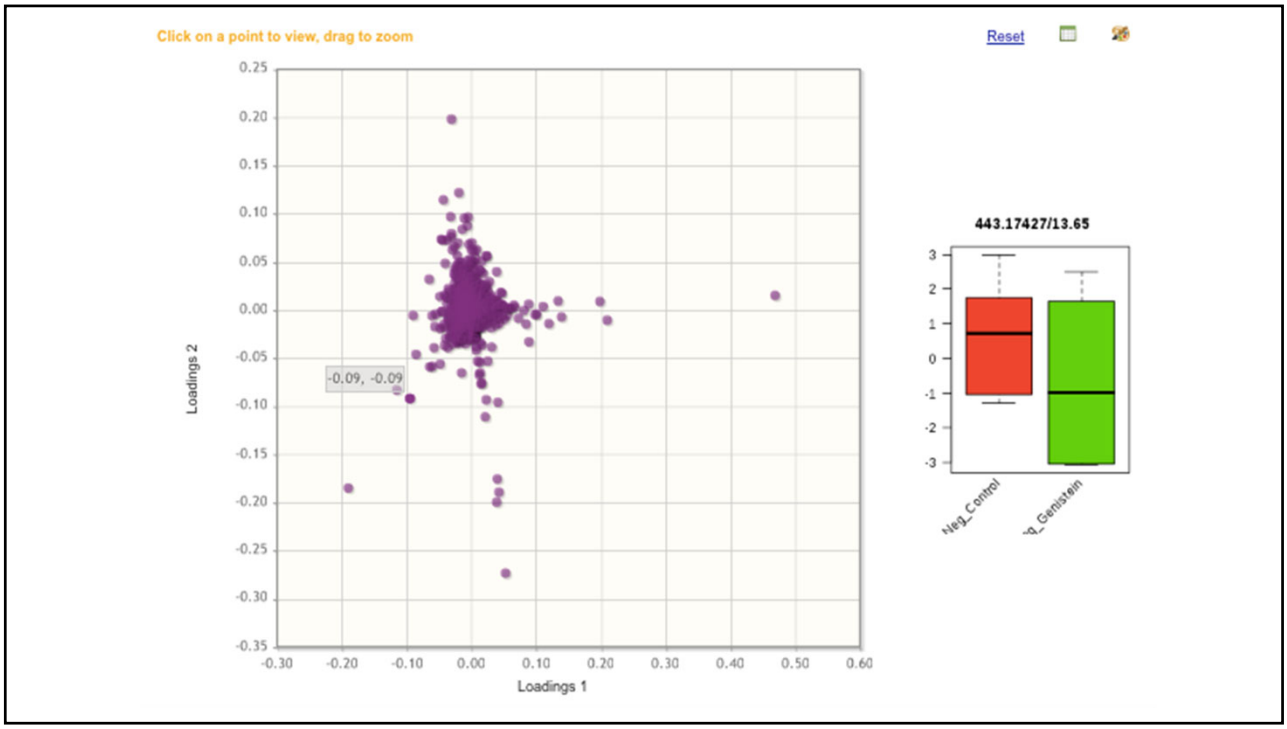
This is the 3D-PCA plot – it was screen captured

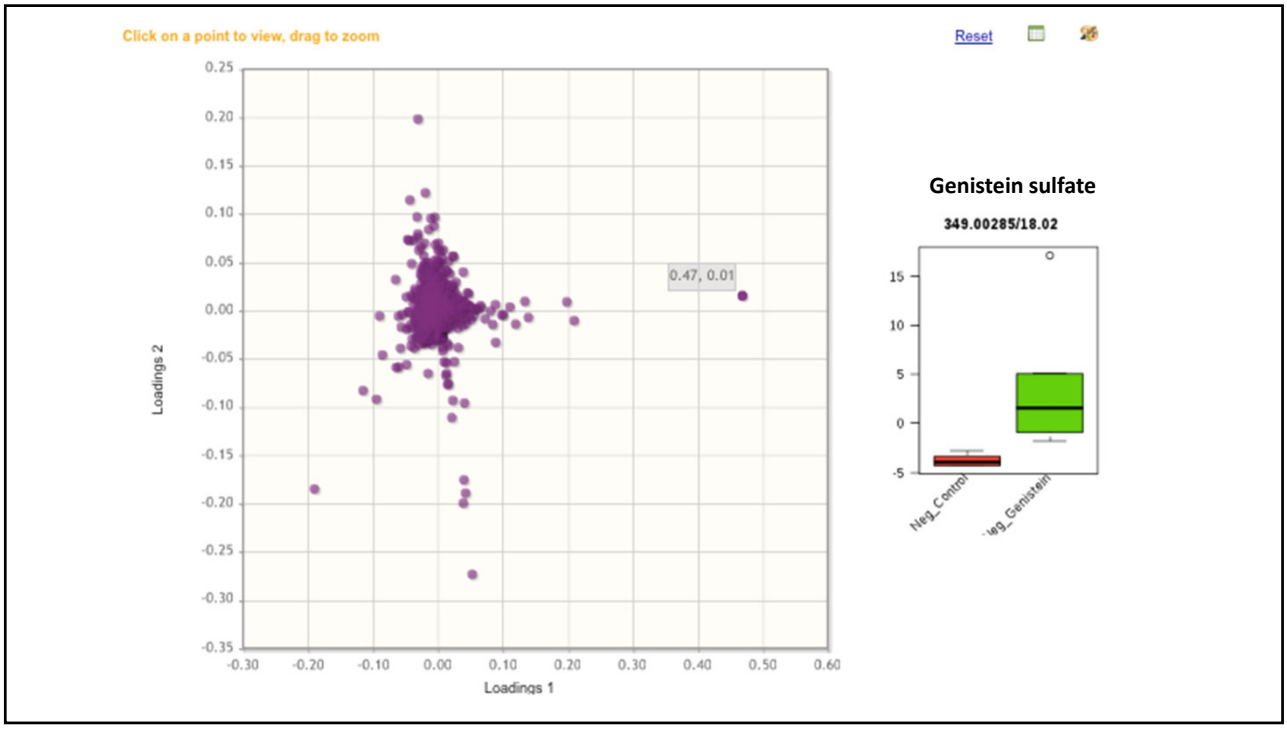


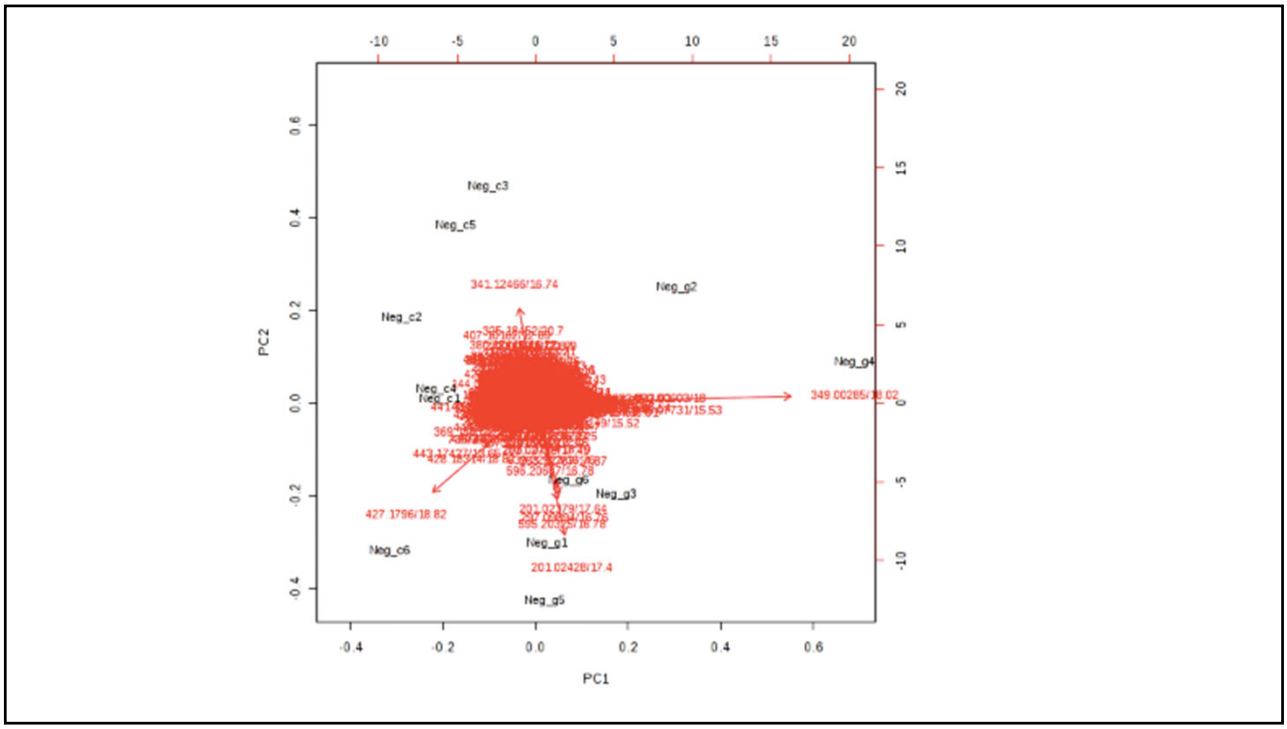


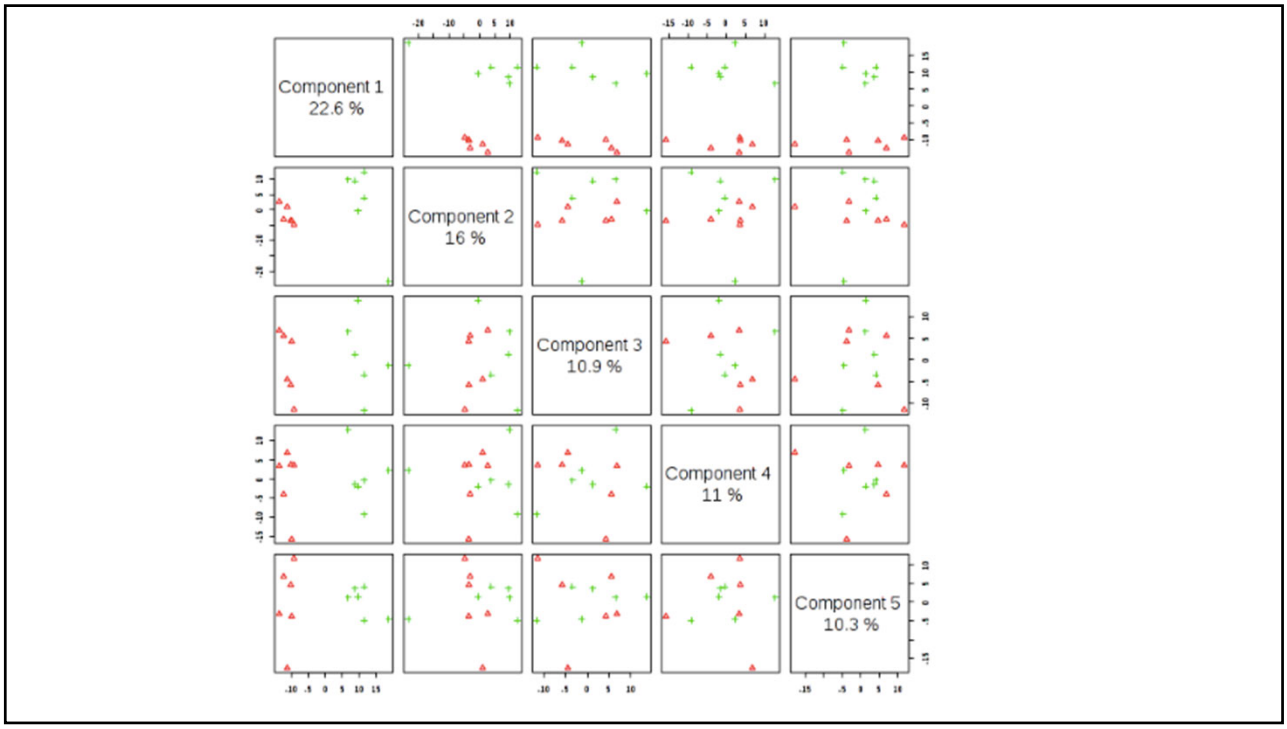
Loadings plot

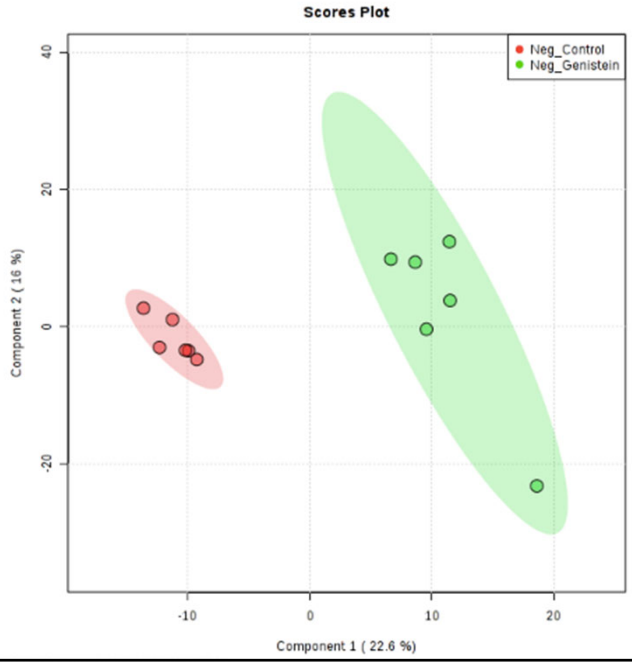



Click on a point to view, drag to zoom
Reset
Loadings 2
Loadings 1
-0.09, -0.09
443.17427/13.65
Neg_Control
Neg_Genistein

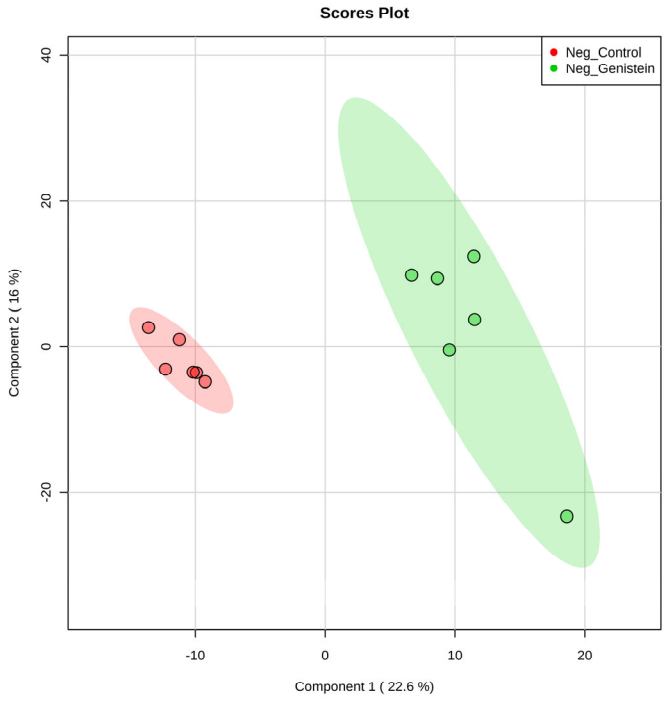

Click on a point to view, drag to zoom
Reset
Loadings 2
Loadings 1
0.47, 0.01
Genistein sulfate
349.00285/18.02
Neg_Control
Neg_Genistein


PC1
PC2
Neg_c1
Neg_c2
Neg_c3
Neg_c4
Neg_c5
Neg_c6
Neg_g1
Neg_g2
Neg_g3
Neg_g4
Neg_g5
Neg_g6
341.12466/16.74
427.1796/18.82
201.02428/17.4
349.00285/18.02

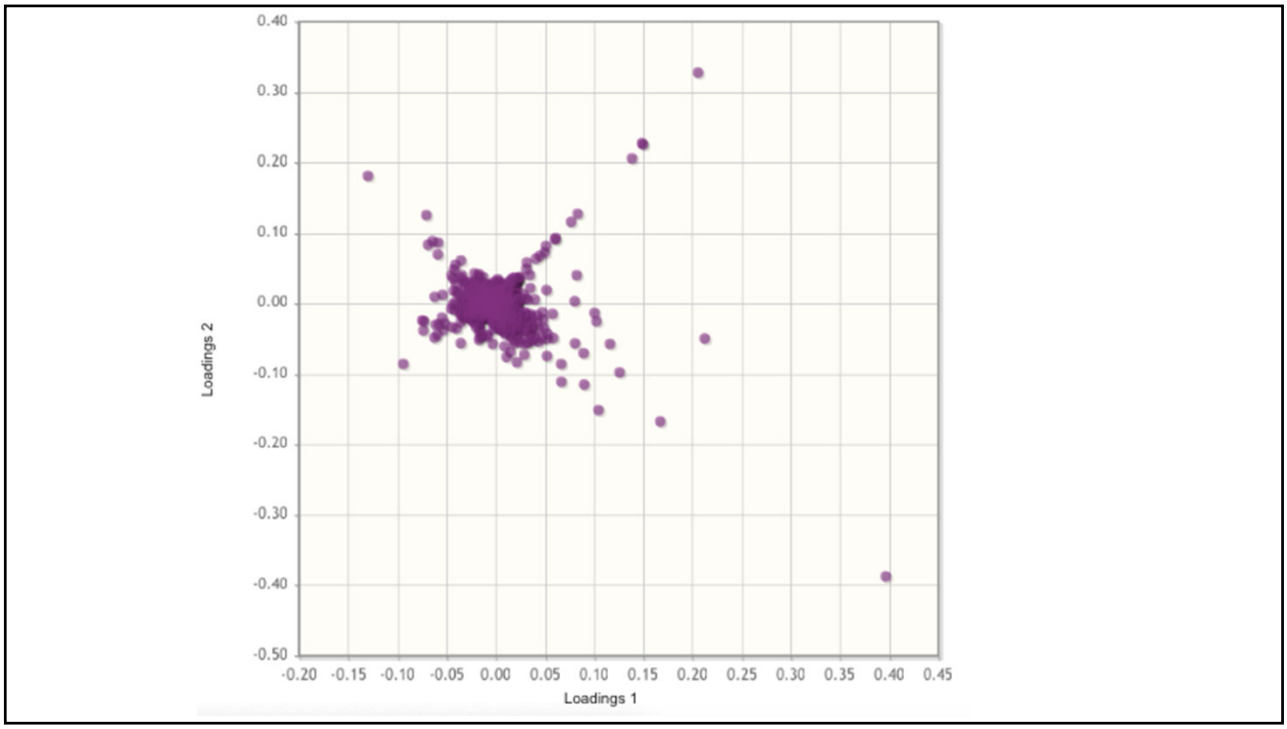

Component 1
22.6 %
Component 2
16 %
Component 3
10.9 %
Component 4
11 %
Component 5
10.3 %

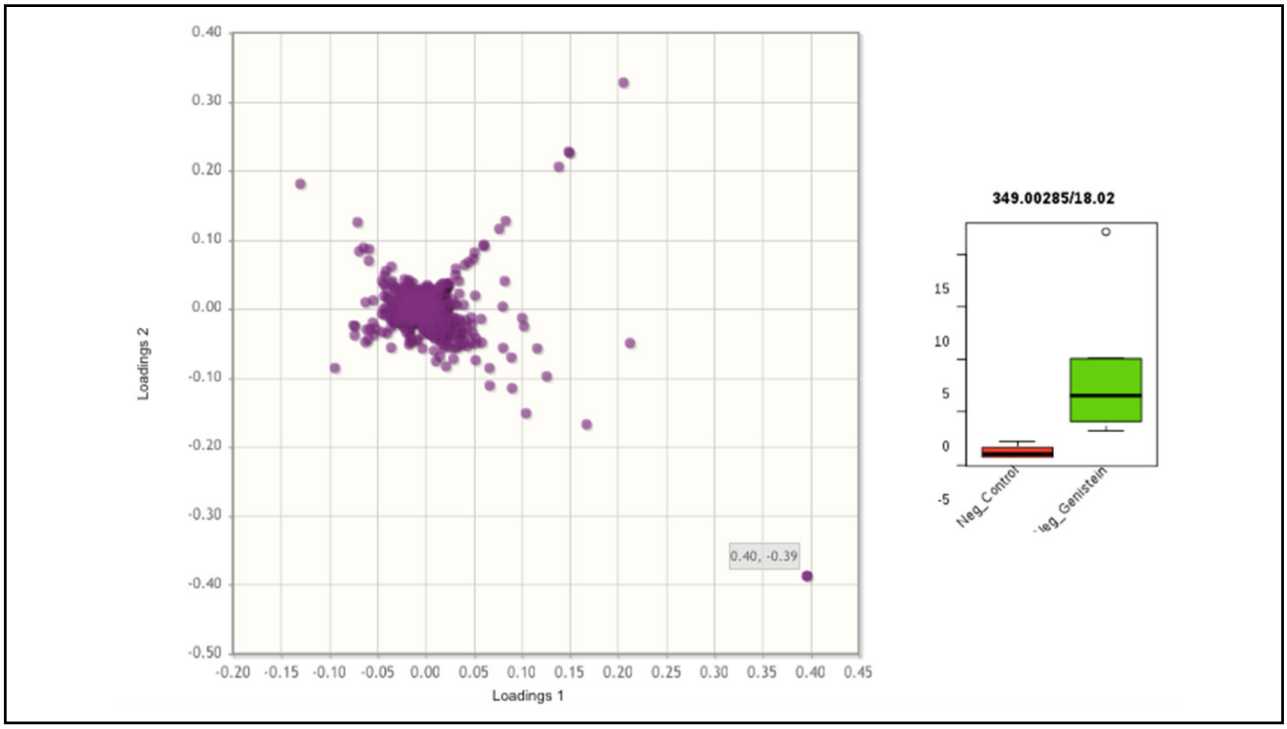


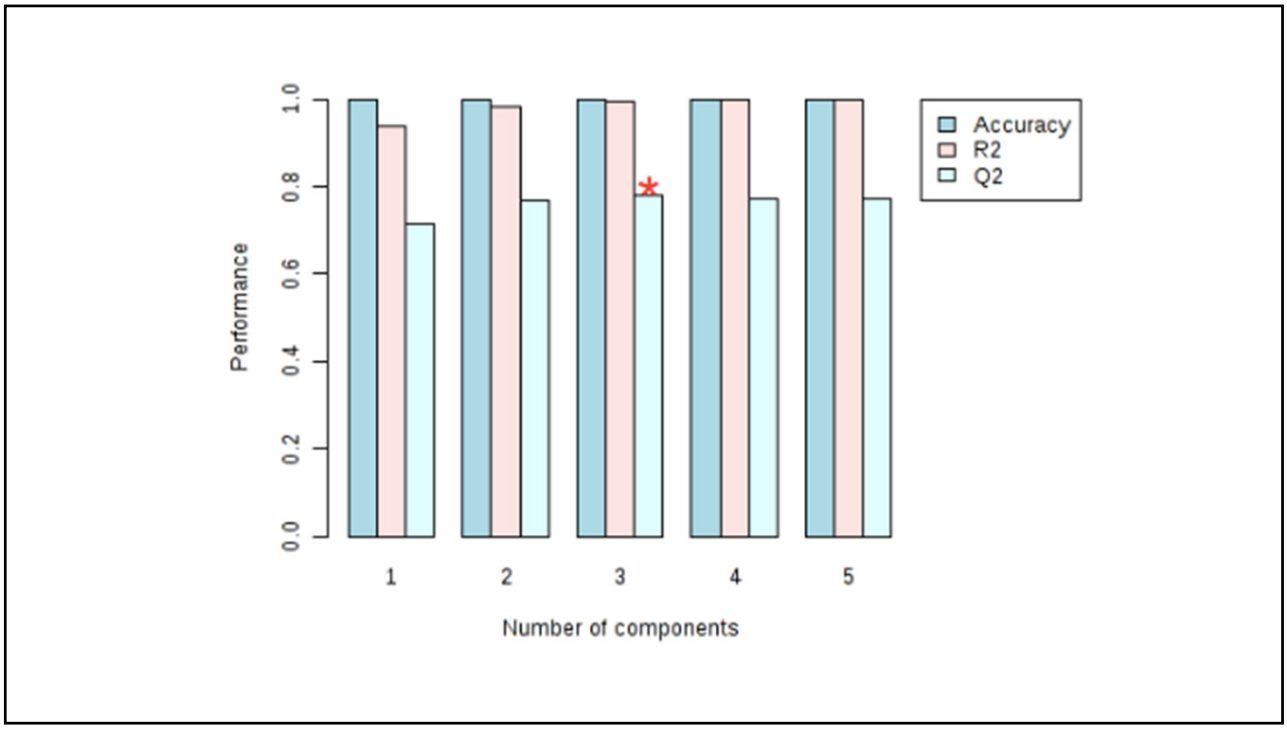
Screen captured 2D-PLS-DA plot

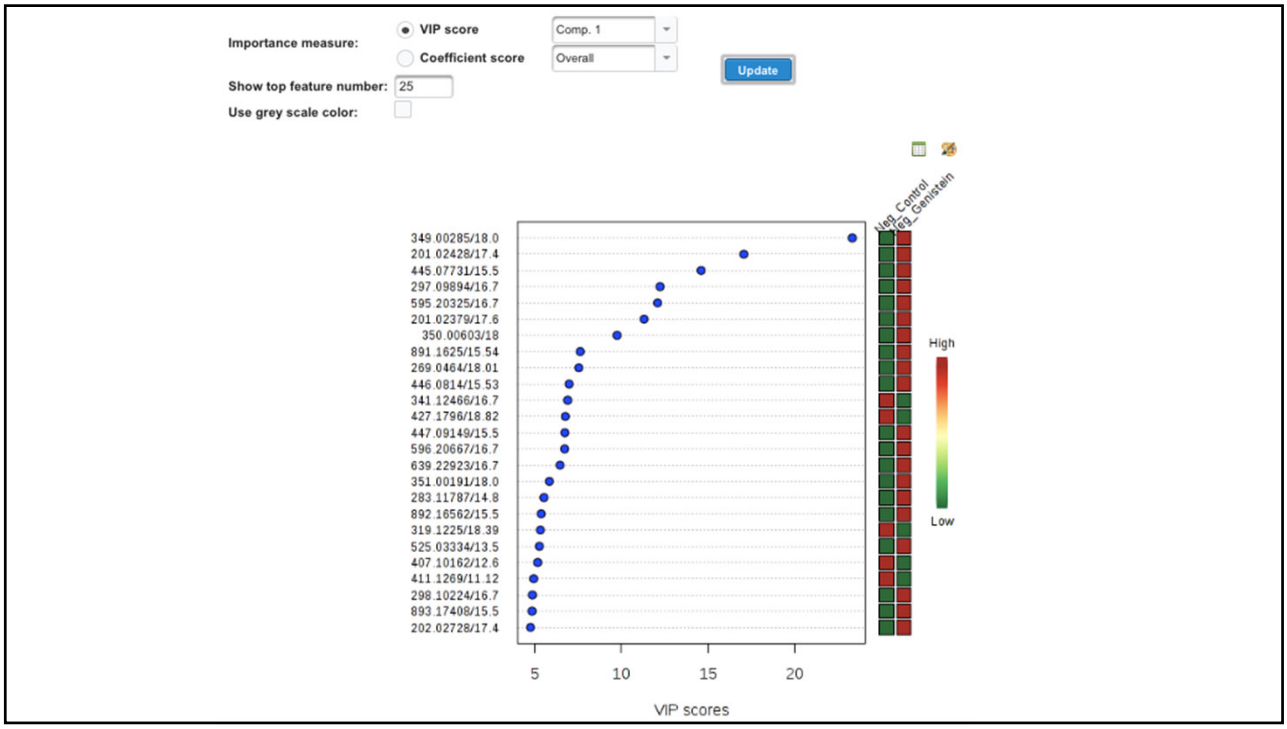


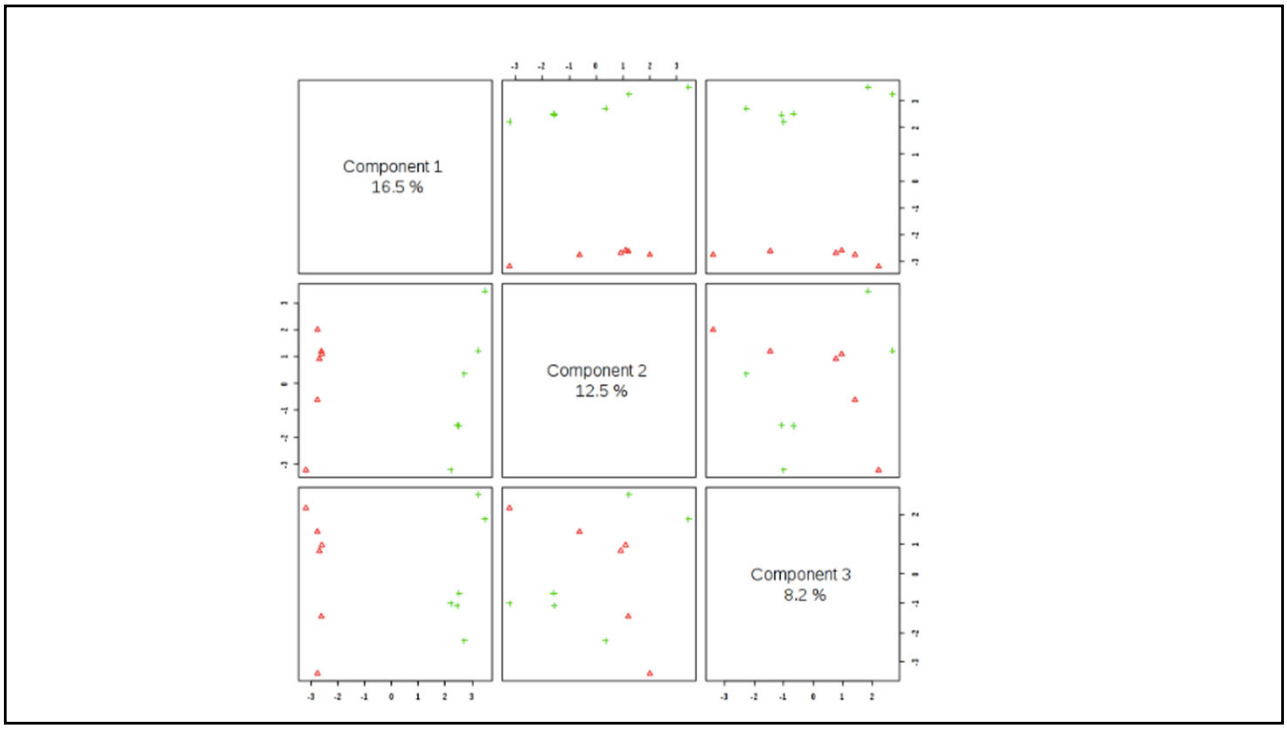
This is the TIFF file for publication quality figures

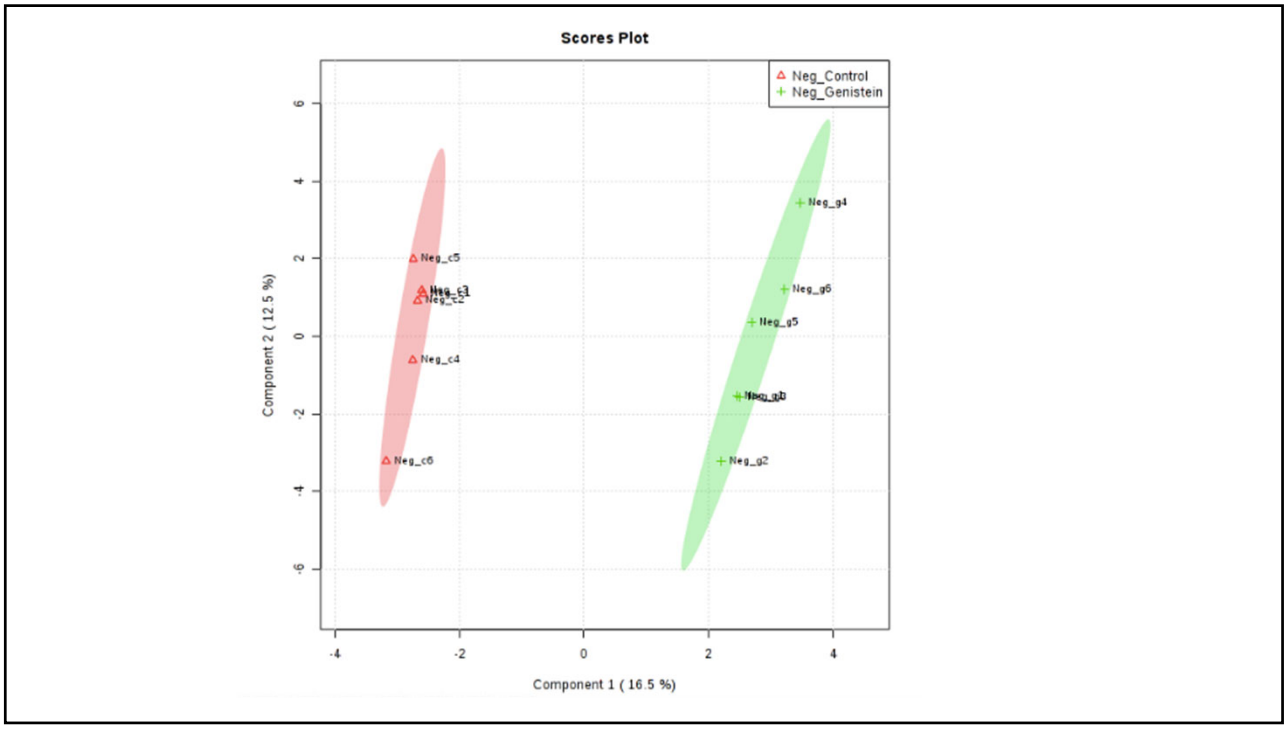

Loadings 2
Loadings 1
0.40
0.30
0.20
0.10
0.00
-0.10
-0.20
-0.30
-0.40
-0.50
-0.20 -0.15 -0.10 -0.05 0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45

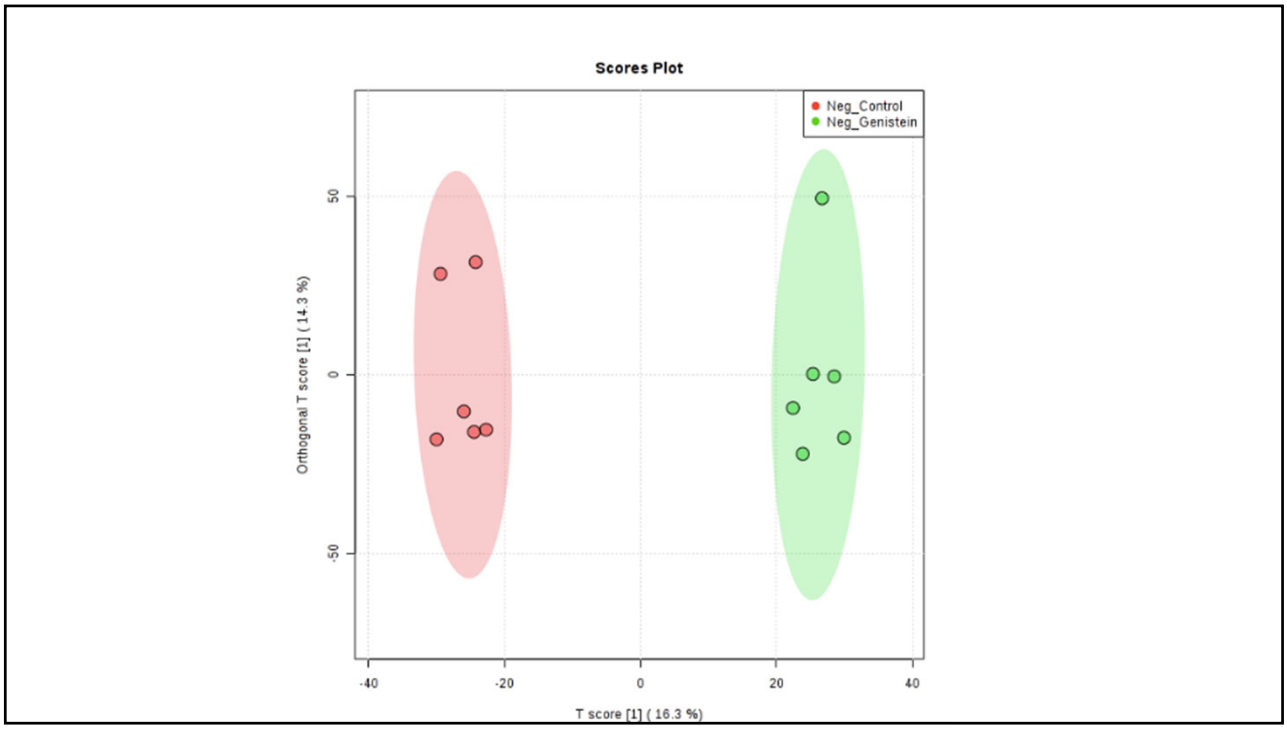

Loadings 2
Loadings 1
0.40, -0.39
349.00285/18.02
15
10
5
0
-5
Neg_Control
Neg_Genistein

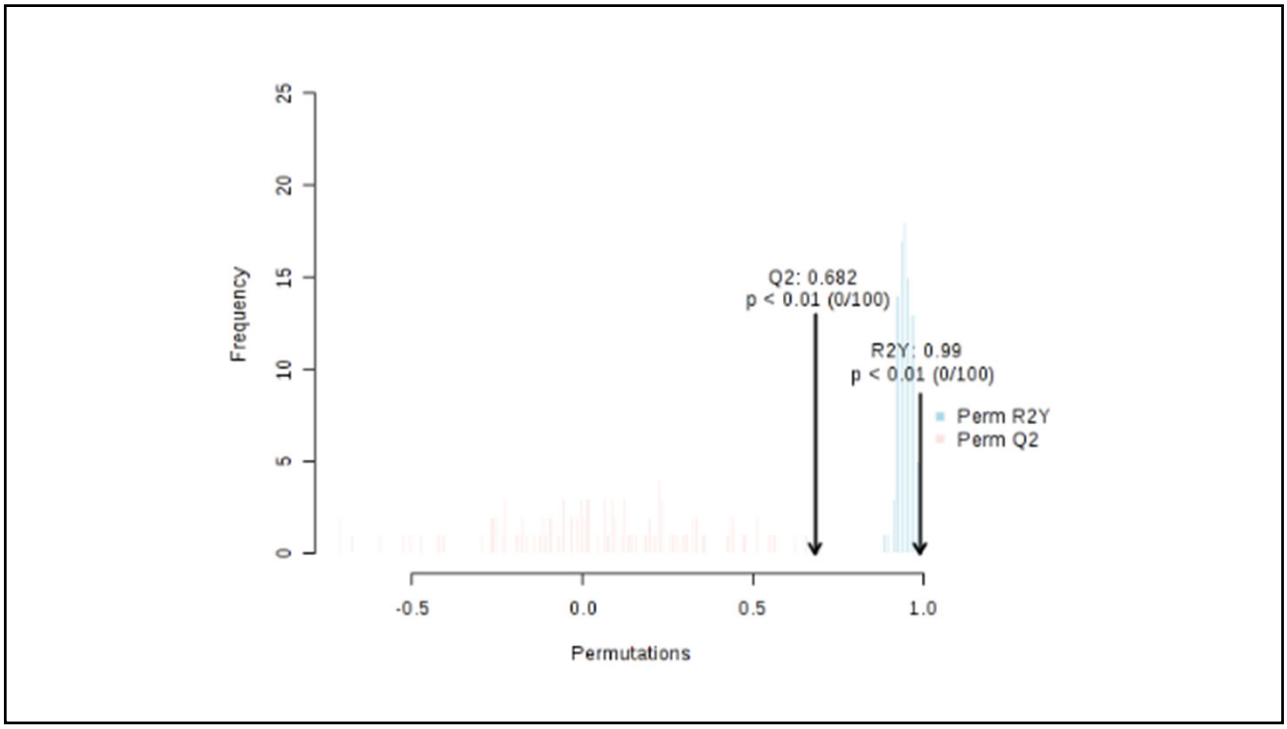

Performance
0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
5
Number of components
Accuracy
R2
Q2

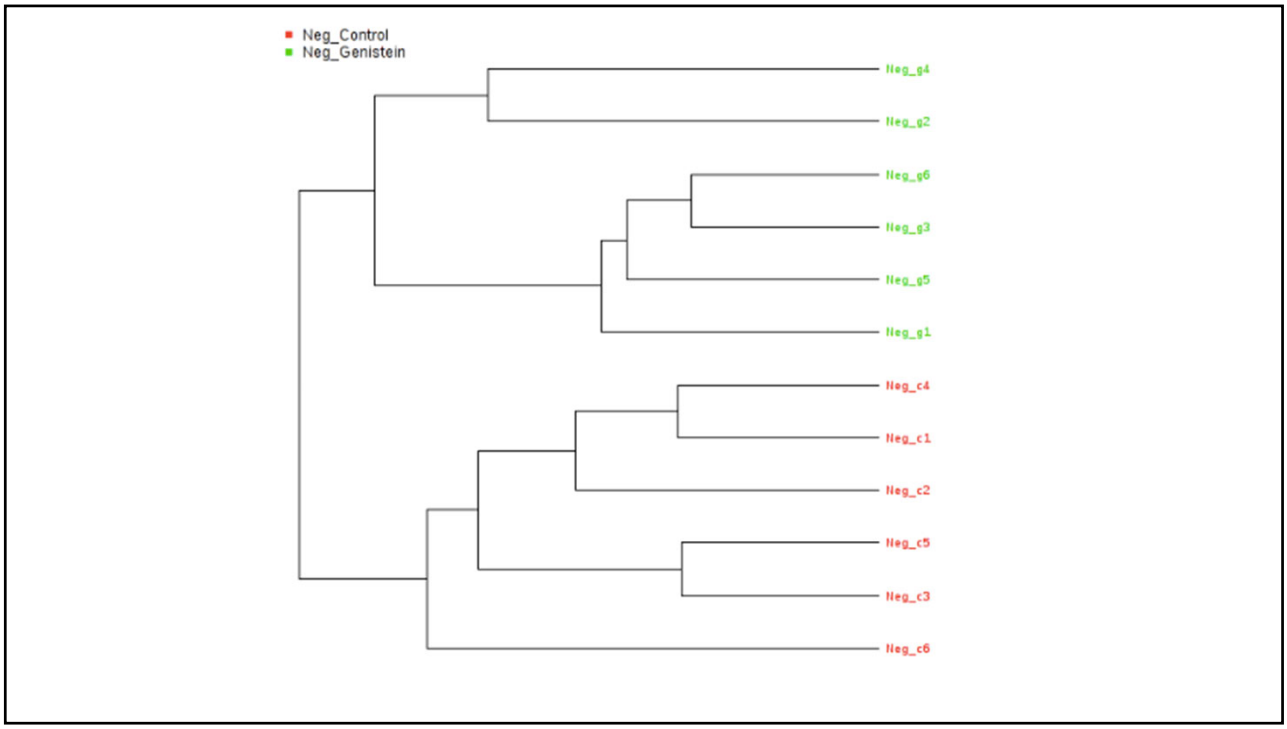

Importance measure:
VIP score
Coefficient score
Comp. 1
Overall
Update
Show top feature number:
25
Use grey scale color:
349.00285/18.0
201.02428/17.4
445.07731/15.5
297.09894/16.7
595.20325/16.7
201.02379/17.6
350.00603/18
891.1625/15.54
269.0464/18.01
446.0814/15.53
341.12466/16.7
427.1796/18.82
447.09149/15.5
596.20667/16.7
639.22923/16.7
351.00191/18.0
283.11787/14.8
892.16562/15.5
319.1225/18.39
525.03334/13.5
407.10162/12.6
411.1269/11.12
298.10224/16.7
893.17408/15.5
202.02728/17.4
5
10
15
20
VIP scores
Neg_Control
Pos_Genistein
High
Low

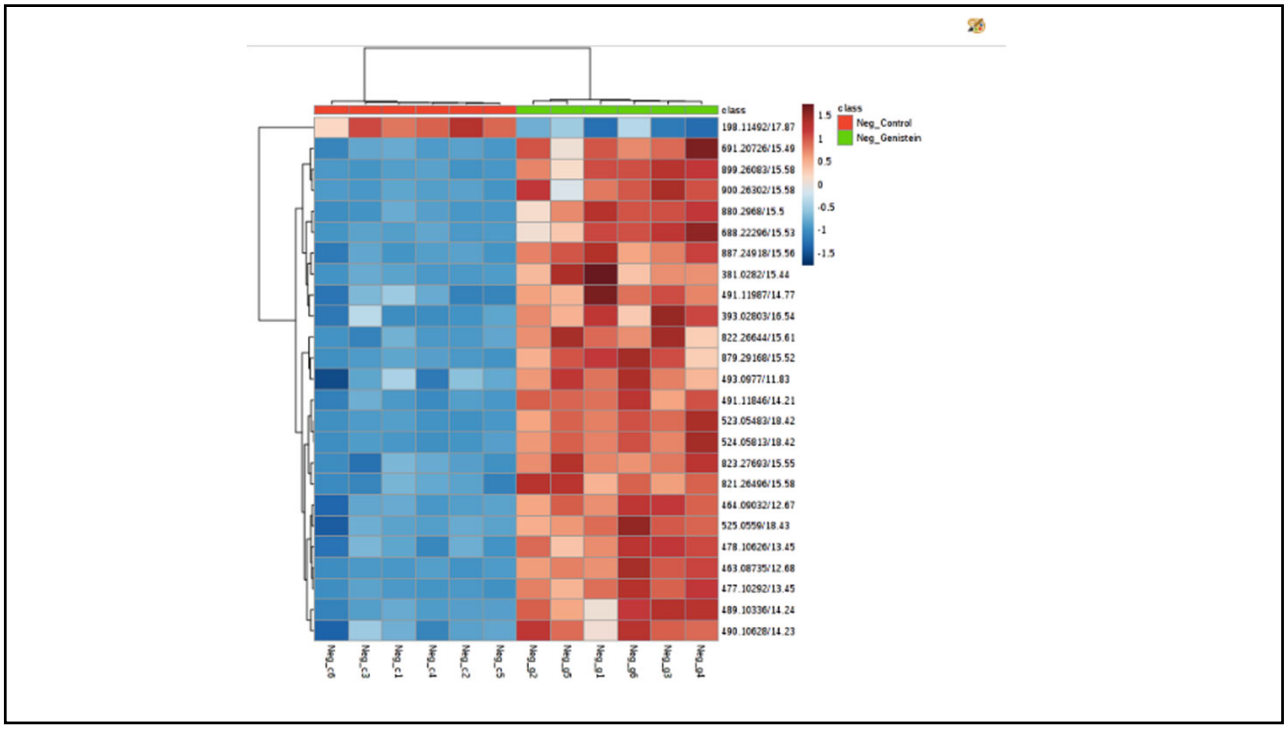

Component 1
16.5 %
Component 2
12.5 %
Component 3
8.2 %

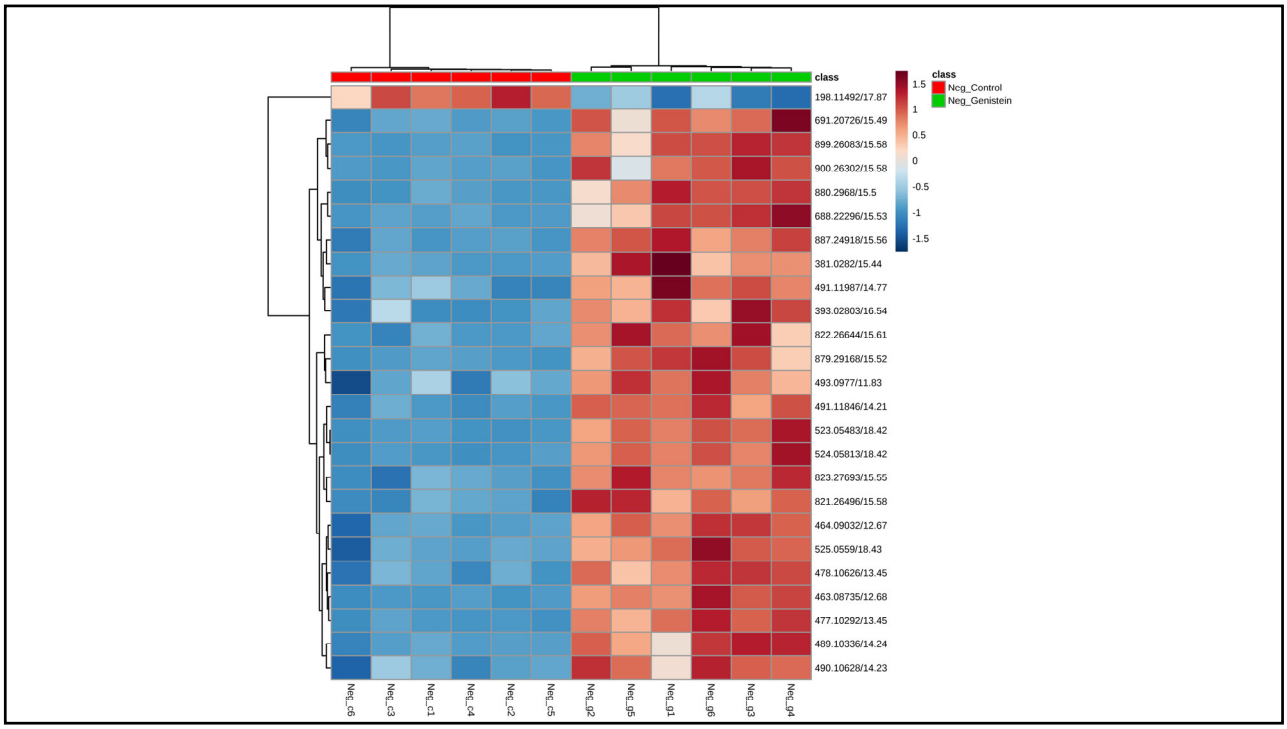

Scores Plot
Neg_Control
Neg_Genistein
Neg_c5
Neg_c4
Neg_c6
Neg_g4
Neg_g6
Neg_g5
Neg_g2
Component 2 (12.5 %)
Component 1 (16.5 %)

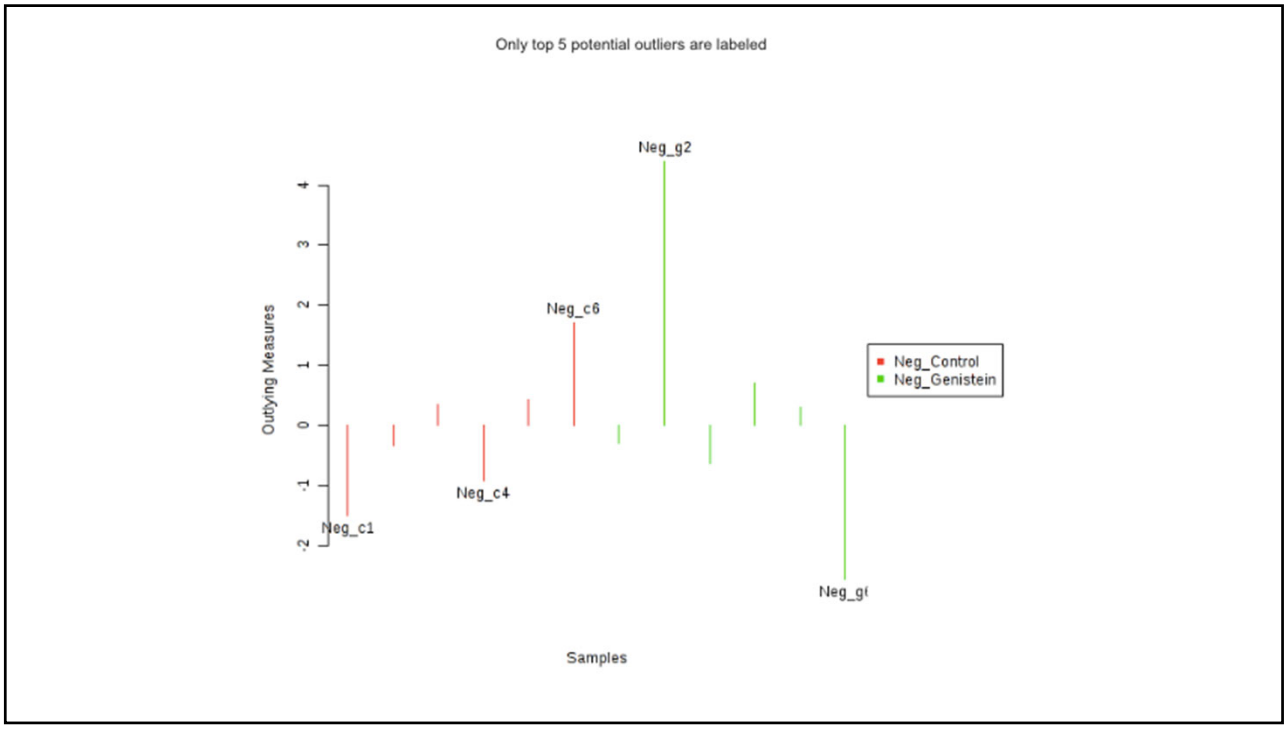

Scores Plot
Neg_Control
Neg_Genistein
Orthogonal T score [1] (14.3 %)
50
0
-50
-40
-20
0
20
40
T score [1] (16.3 %)


Frequency
25
20
15
10
5
0
Q2: 0.682
p < 0.01 (0/100)
R2Y: 0.99
p < 0.01 (0/100)
Perm R2Y
Perm Q2
-0.5
0.0
0.5
1.0
Permutations


Neg_Control
Neg_Genistein
Neg_g4
Neg_g2
Neg_g6
Neg_g3
Neg_g5
Neg_g1
Neg_c4
Neg_c1
Neg_c2
Neg_c5
Neg_c3
Neg_c6


class
198.11492/17.87
691.20726/15.49
899.26083/15.58
900.26302/15.58
880.2968/15.5
888.22296/15.53
887.24918/15.56
381.0282/15.44
491.11987/14.77
393.02803/16.54
822.26644/15.61
879.29168/15.52
493.0977/11.83
491.11846/14.21
523.05483/18.42
524.05813/18.42
823.27693/15.55
821.26406/15.58
464.09032/12.67
525.0559/18.43
478.10626/13.45
463.08735/12.68
477.10292/13.45
489.10336/14.24
490.10628/14.23
1.5
1
0.5
0
-0.5
-1
-1.5
class
Neg_Control
Neg_Genistein
Neg_c6
Neg_c3
Neg_c1
Neg_c4
Neg_c2
Neg_c5
Neg_g2
Neg_g5
Neg_g1
Neg_g6
Neg_g3
Neg_g4


class
198.11492/17.87
691.20726/15.49
899.26083/15.58
900.26302/15.58
880.2968/15.5
688.22296/15.53
887.24918/15.56
381.0282/15.44
491.11987/14.77
393.02803/16.54
822.26644/15.61
879.29168/15.52
493.0977/11.83
491.11846/14.21
523.05483/18.42
524.05813/18.42
823.27693/15.55
821.26496/15.58
464.09032/12.67
525.0559/18.43
478.10626/13.45
463.08735/12.68
477.10292/13.45
489.10336/14.24
490.10628/14.23
Neg_c6
Neg_c3
Neg_c1
Neg_c4
Neg_c2
Neg_c5
Neg_g2
Neg_g5
Neg_g1
Neg_g6
Neg_g3
Neg_g4
1.5
1
0.5
0
-0.5
-1
-1.5
class
Neg_Control
Neg_Genistein


Only top 5 potential outliers are labeled
Outlying Measures
4
3
2
1
0
-1
-2
Neg_g2
Neg_c6
Neg_c4
Neg_c1
Neg_g
Neg_Control
Neg_Genistein
Samples

Number of trees: 500

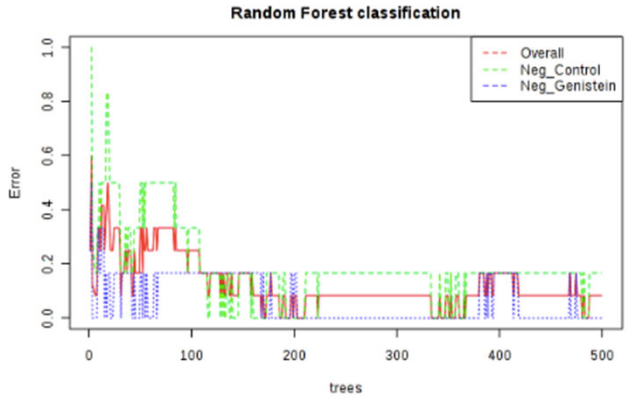
Number of predictors: 7

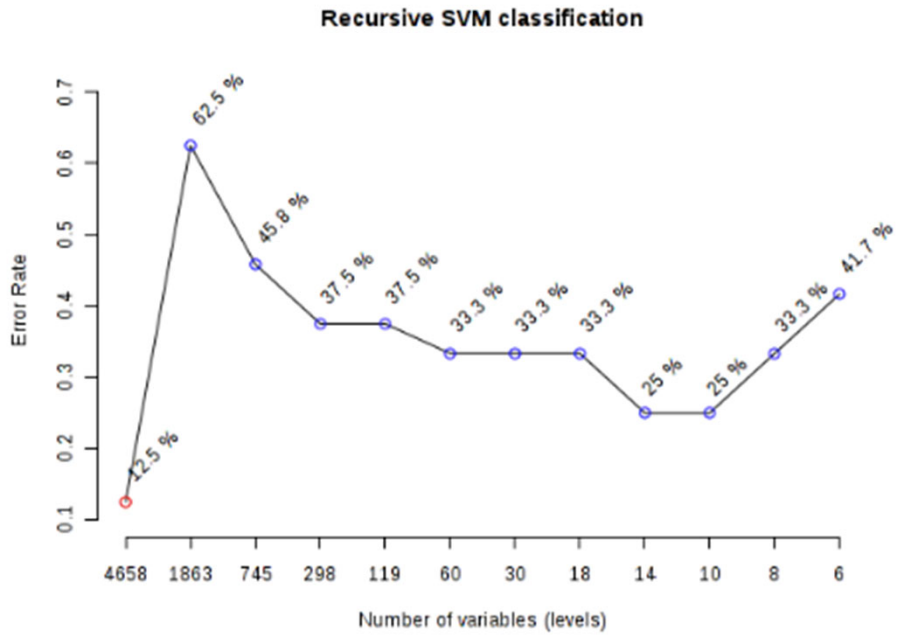
Randomness: On

Update

The OOB error is 0.0833

	Neg_Control	Neg_Genistein	class.error
Neg_Control	5	1	0.167
Neg_Genistein	0	6	0.0





Result Download

Please download the results (tables and images) below. The **Download.zip** contains all the files in your home directory. You can also generate a **PDF analysis report** using the button below.

Generate Report ←

Download.zip	svm_cls_0_dpi72.png
Rhistory.R	ptn_1_dpi72.png
spls_pair_1_dpi72.png	spls_cv_0_dpi72.png
ebam_sigfeatures.csv	opls_perm_1_dpi72.png
randomforests_sigfeatures.csv	opls_mdl_0_dpi72.png
plsda_coef.csv	pval_corr_table.csv
fold_change.csv	sam_view_0_dpi72.png
oplsda_score.csv	splsda_score.csv
349.00285_18.02_dpi72.png	fc_1_dpi72.png

Result Download

Please download the results (tables and images) below. The **Download.zip** contains all the files in your home directory. You can also generate a **PDF analysis report** using the button below.

Generate Report → Analysis Report

Download.zip	svm_cls_0_dpi72.png
Rhistory.R	ptn_1_dpi72.png
spls_pair_1_dpi72.png	spls_cv_0_dpi72.png
ebam_sigfeatures.csv	opls_perm_1_dpi72.png
randomforests_sigfeatures.csv	opls_mdl_0_dpi72.png
plsda_coef.csv	pval_corr_table.csv
fold_change.csv	sam_view_0_dpi72.png
oplsda_score.csv	splsda_score.csv
349.00285_18.02_dpi72.png	fc_1_dpi72.png

Metabolomic Data Analysis with MetaboAnalyst 4.0

Name: guest1405087504956705098

February 13, 2019

1 Data Processing and Normalization

1.1 Reading and Processing the Raw Data

MetaboAnalyst accepts a variety of data types generated in metabolomic studies, including compound concentration data, binned NMR/MS spectra data, NMR/MS peak list data, as well as MS spectra (NetCDF, mzXML, mzDATA). Users need to specify the data types when uploading their data in order for MetaboAnalyst to select the correct algorithm to process them. Table 1 summarizes the result of the data processing steps.

Result Download

Please download the results (tables and images) below. The **Download.zip** contains all the files in your home directory. You can also generate a **PDF analysis report** using the button below.

Generate Report

Analysis Report

Download.zip	svm_cls_0_dpi72.png
Rhistory.R	ptn_1_dpi72.png
spls_pair_1_dpi72.png	spls_cv_0_dpi72.png
ebam_sigfeatures.csv	opls_perm_1_dpi72.png
randomforests_sigfeatures.csv	opls_mdl_0_dpi72.png
plsda_coef.csv	pval_corr_table.csv
fold_change.csv	sam_view_0_dpi72.png
oplsda_score.csv	splsda_score.csv
349.00285_18.02_dpi72.png	fc_1_dpi72.png