

Statistical analysis with Metaboanalyst

Stephen Barnes, PhD

934-7117, sbarnes@uab.edu

Go to www.metaboanalyst.ca



MetaboAnalyst-- a comprehensive tool for metabolomics analysis and interpretation

[Home](#)

[Overview](#)

[Data Formats](#)

[FAQs](#)

[Tutorials](#)

[Troubleshooting](#)

[Resources](#)

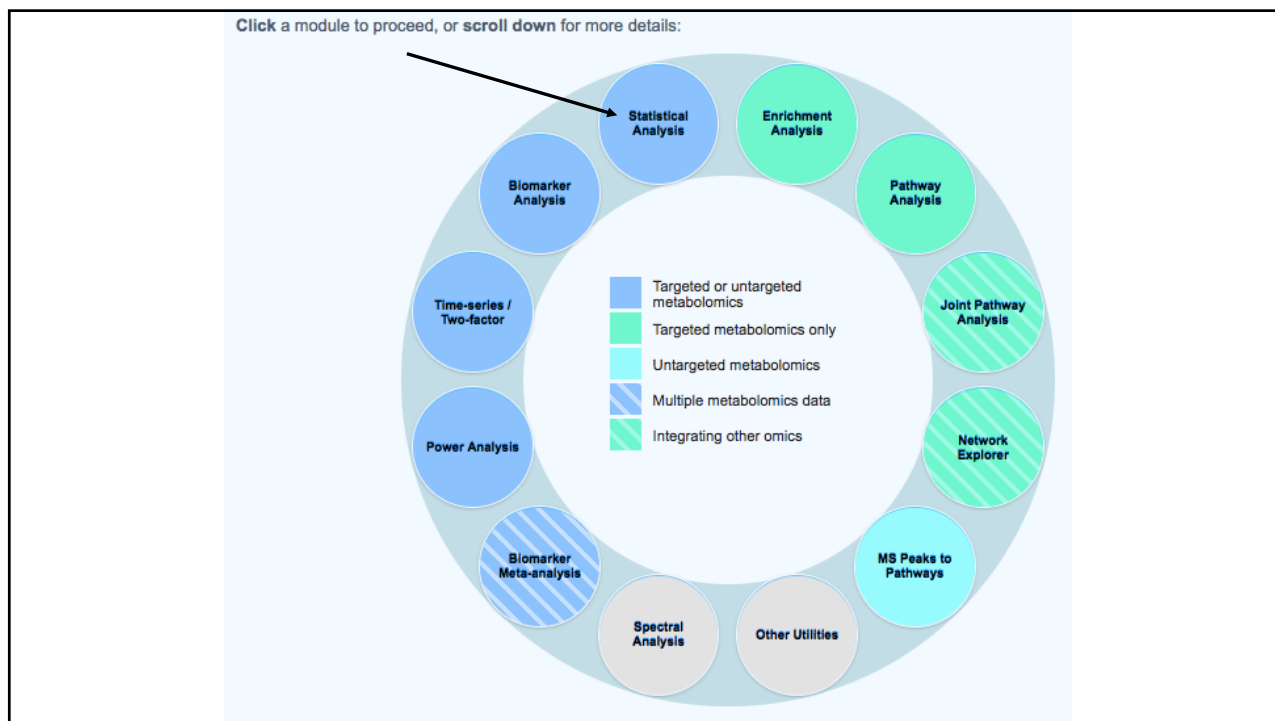
[Update History](#)

Welcome >> [click here to start](#) <<

News & Updates

- Minor bug fixes and feature enhancements based on user feedback (02/02/2018); **NEW**
- Release of **MetaboAnalyst 4.0** (01/29/2018); **NEW**
- Updated the interface for module selection (01/22/2018); **NEW**
- Updated the compound libraries based on the latest HMDB (v4.0) (01/18/2018); **NEW**
- Added a new module for network-based integrative analysis for lists of metabolites and genes (supporting KEGG Orthologs from metagenomics studies) (01/12/2018); **NEW**
- Added a new module for meta-analysis for multiple metabolomics datasets for robust biomarker identification (01/08/2018); **NEW**
- Added a new module for pathway enrichment analysis for untargeted metabolomics (mummichog) (12/22/2017); **NEW**
- Minor bug fixes and interface improvements based on user feedback (12/18/2017); **NEW**
- Added support for interactive enrichment network and SVG export (12/15/2017); **NEW**
- Release of MetaboAnalyst 3.5 together with a companion R package **MetaboAnalystR** for more flexible data analysis and batch processing; **NEW**
- Check out our **MicrobiomeAnalyst** tool for comprehensive analysis of microbiome data;

[Read more.....](#)



1) Upload your data

Tab-delimited text (.txt) or comma-separated values (.csv) file:

Data Type: Concentrations Spectral bins Peak intensity table

Format:

Data File: No file chosen

Zipped Files (.zip) :

Data Type: NMR peak list MS peak list MS spectra

Data File: Class_neg.zip

Pair File: No file chosen

MetaboAnalyst -- a comprehensive tool for metabolomics analysis and interpretation

Processing MS peak list data :

Peaks need to be matched across samples in order to be compared. For two-column format (retention time, and intensities), the program will further group peaks based on their retention time and intensities. For LC-MS peak, and 5 values: mass tolerance - 0.25 (m/z); retention time - 30 (seconds) for LC-MS peak, and 5 be replaced by their sum; some groups will be excluded if none of the classes has at least one sample occupies a row and each column represents a peak group identified by the medication.

Mass tolerance (m/z):

Retention time tolerance:

R Command History

 Keep collapsed

```
1. InitDataObjects("mspeak", "stat", FALSE)
2. UnzipUploadedFile("Replacing_with_your_file_path", "upload", T);
```

Set mass tolerance to 0.001 and retention time tolerance to 0.005. This stops Metaboanalyst grouping the features previously identified by XCMS. Click on 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 39096 peaks,
with an average of 3258 peaks per sample

A total of 3258 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.

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 39096 peaks,
with an average of 3258 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

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.](#)

Filtering features if their RSDs are % in QC samples >

None (less than 5000 features)

Interquartile 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

Data transformation

None

Log transform

Cube root transform

Sample normalization

None

Sample-specific normalization (i.e. weight, volume) [Click here to specify](#)

Normalization by sum

Normalization by median

Normalization by reference sample (PQN)

Normalization by a pooled sample from group

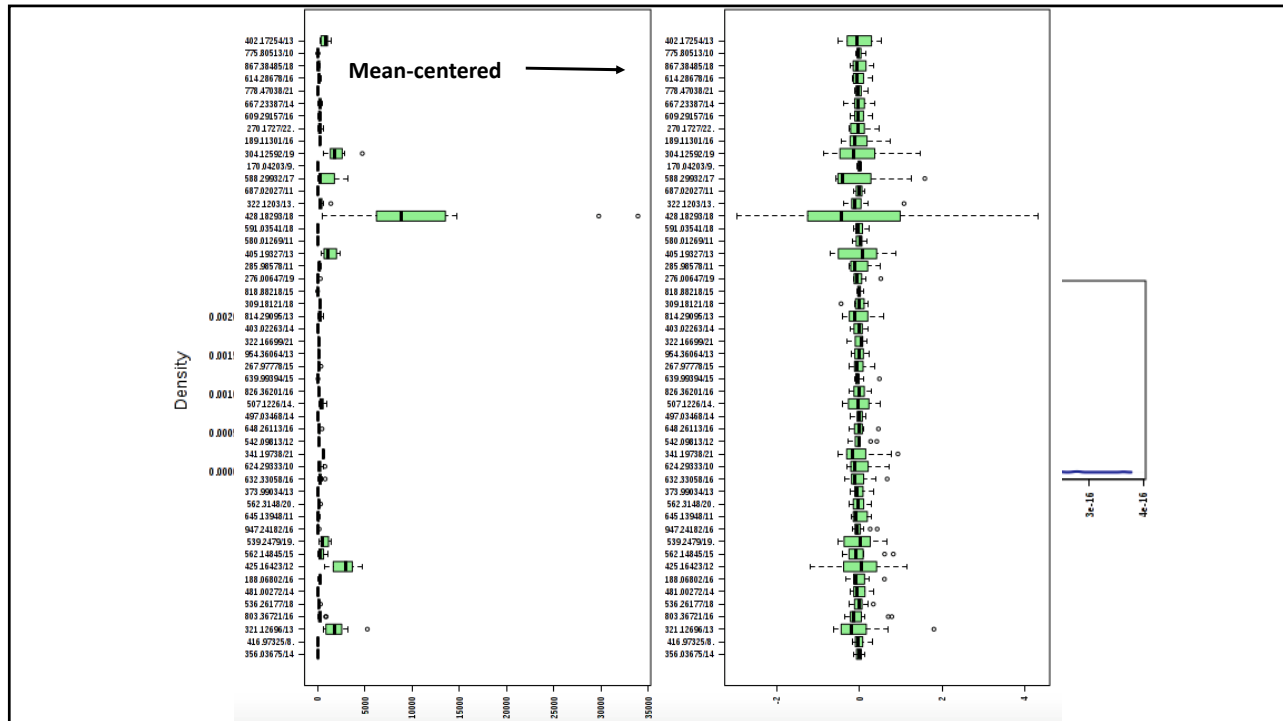
Normalization by reference feature

Quantile normalization

Normalize

View Result

Proceed



Ready for Stats?

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 \(orthoPLS-DA\)](#)

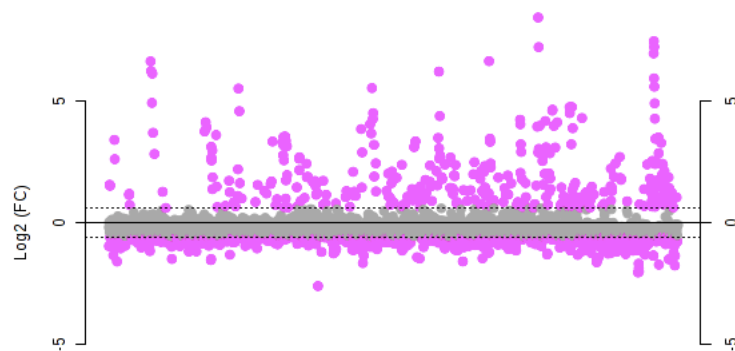
Univariate analysis

Analysis type:

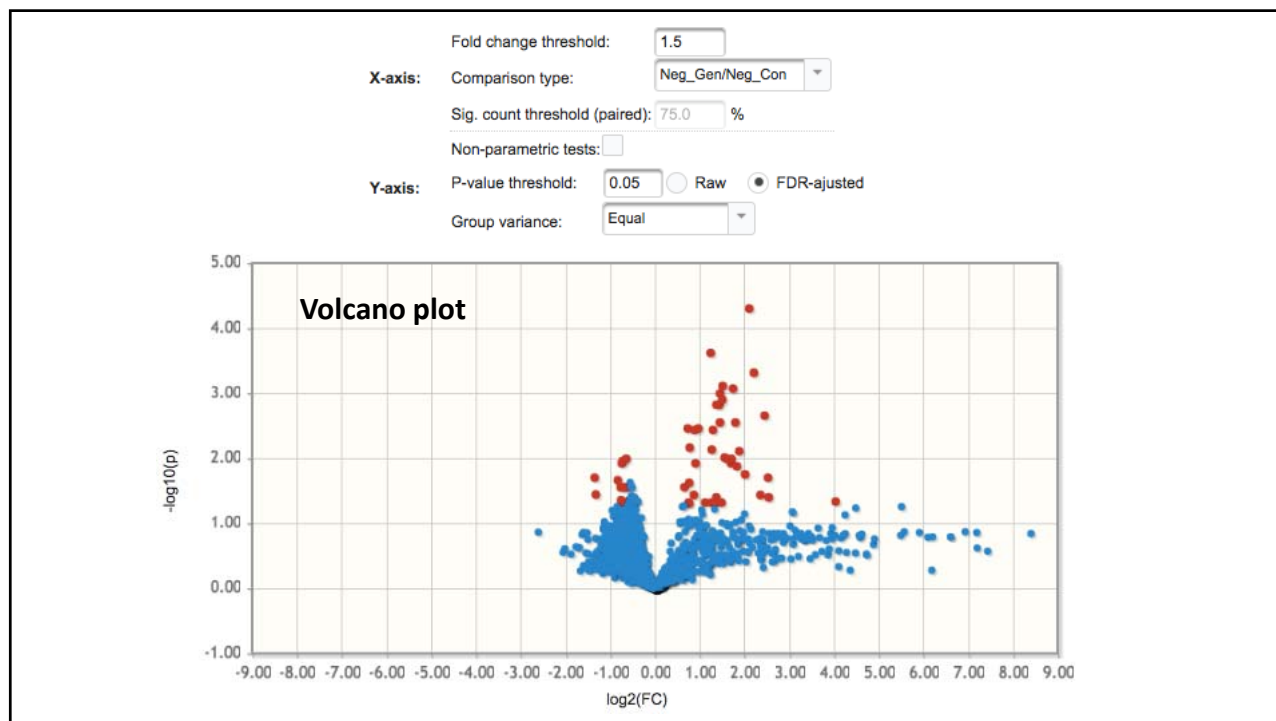
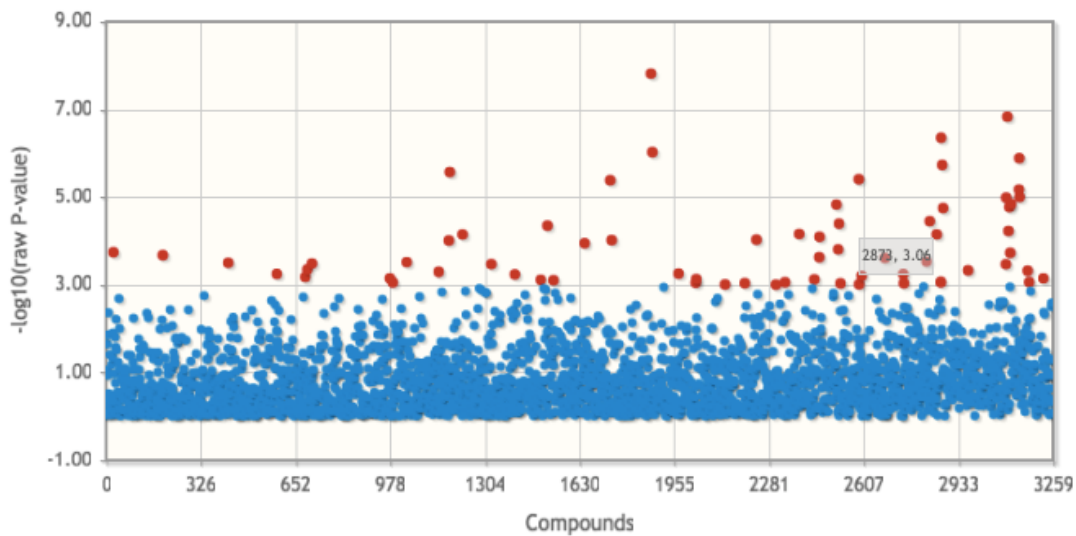
Fold change threshold:

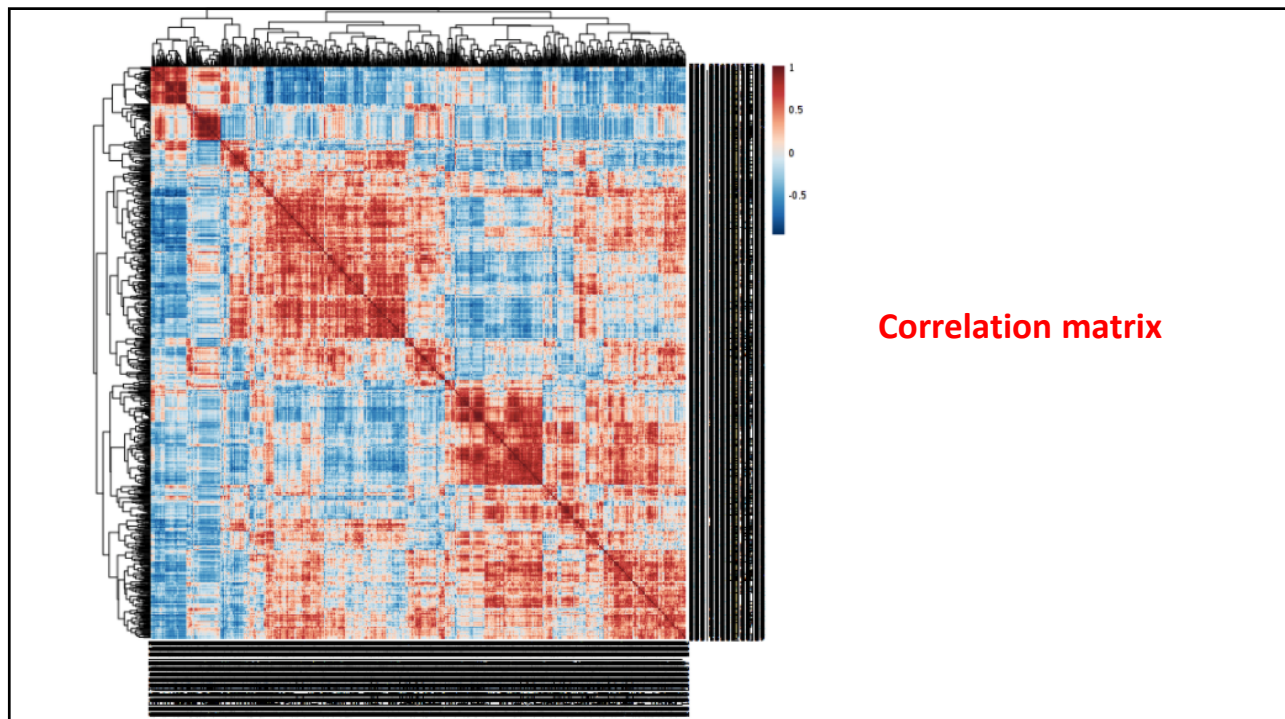
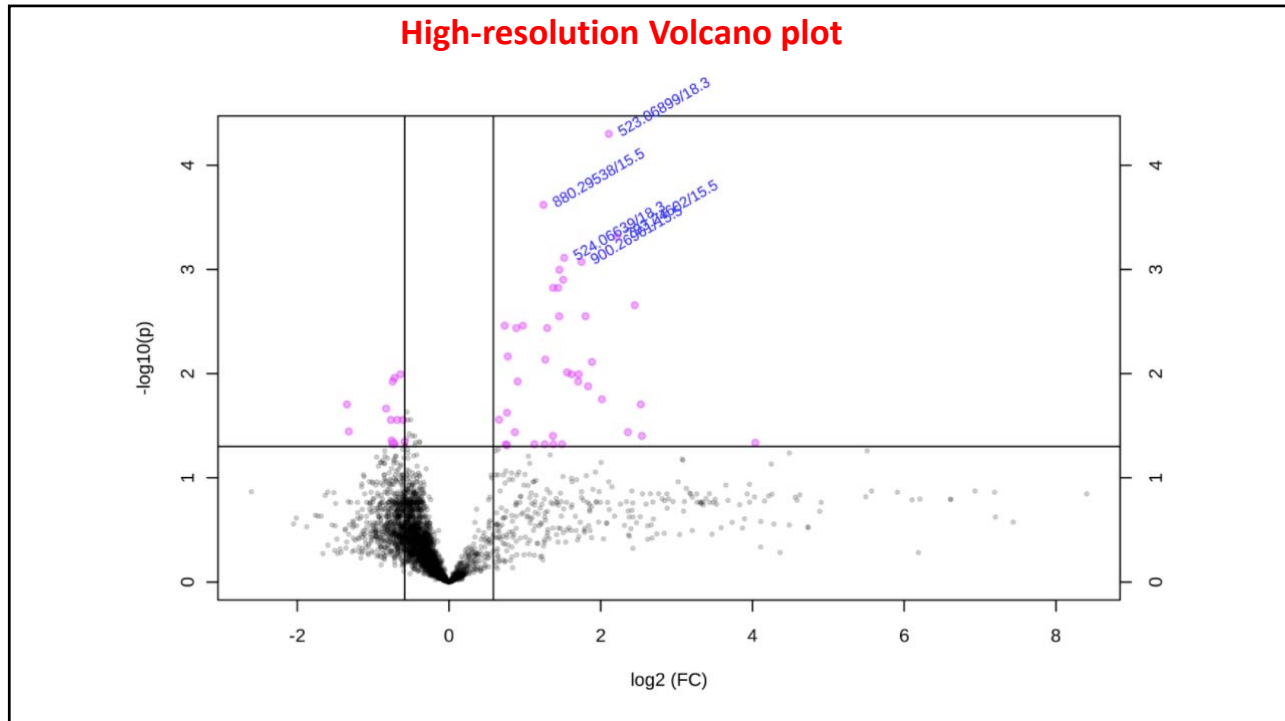
Comparison type:

Significant count threshold (paired only): %

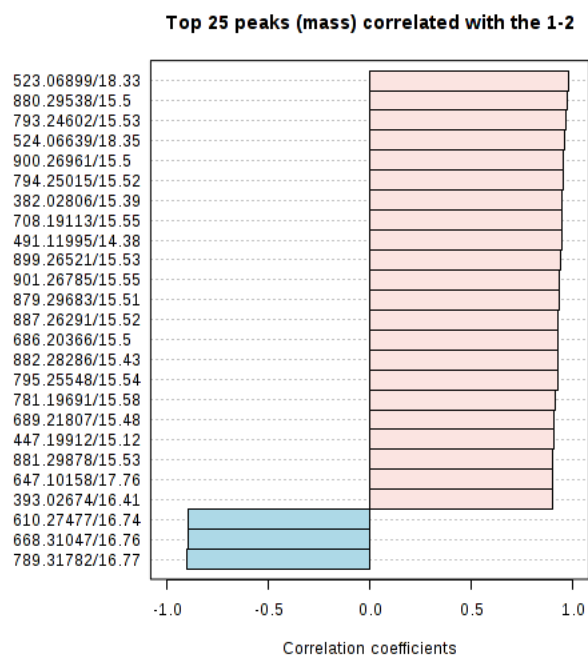


T-test = FDR adjusted





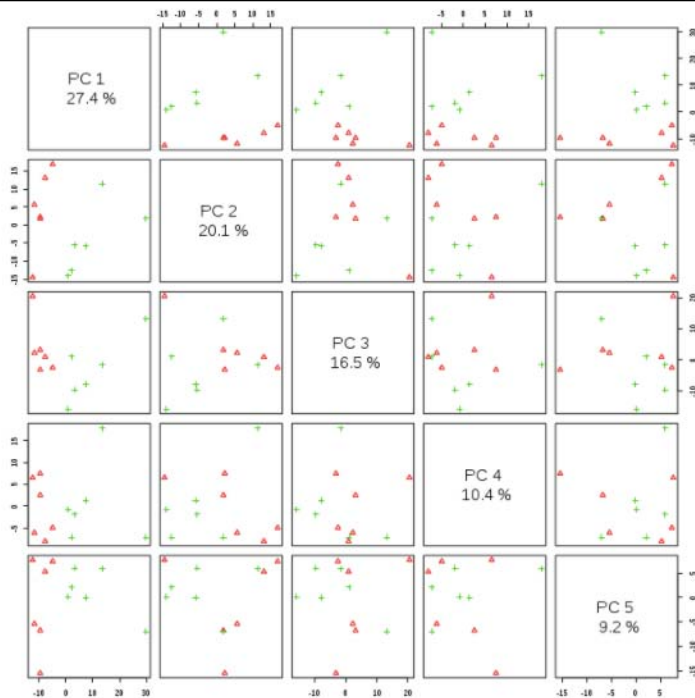
**Top features correlated
with Genistein vs Control**



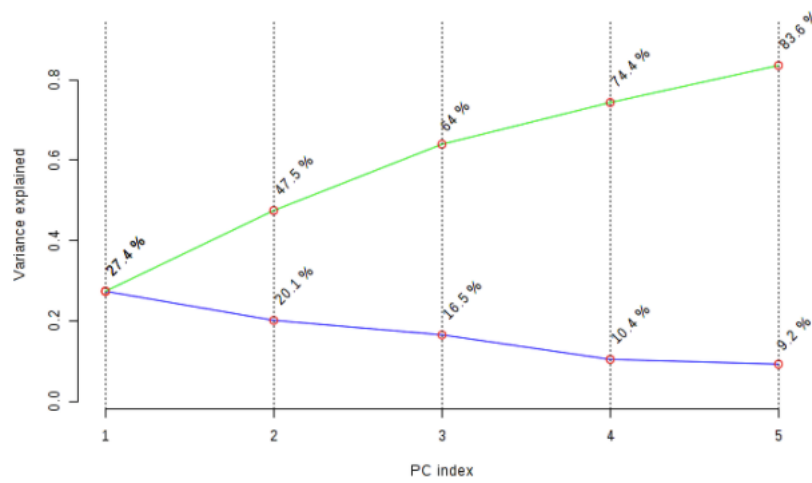
Multivariate analyses

Principal Components Analysis (PCA)

Plots of 5 principal components, taking two at a time.



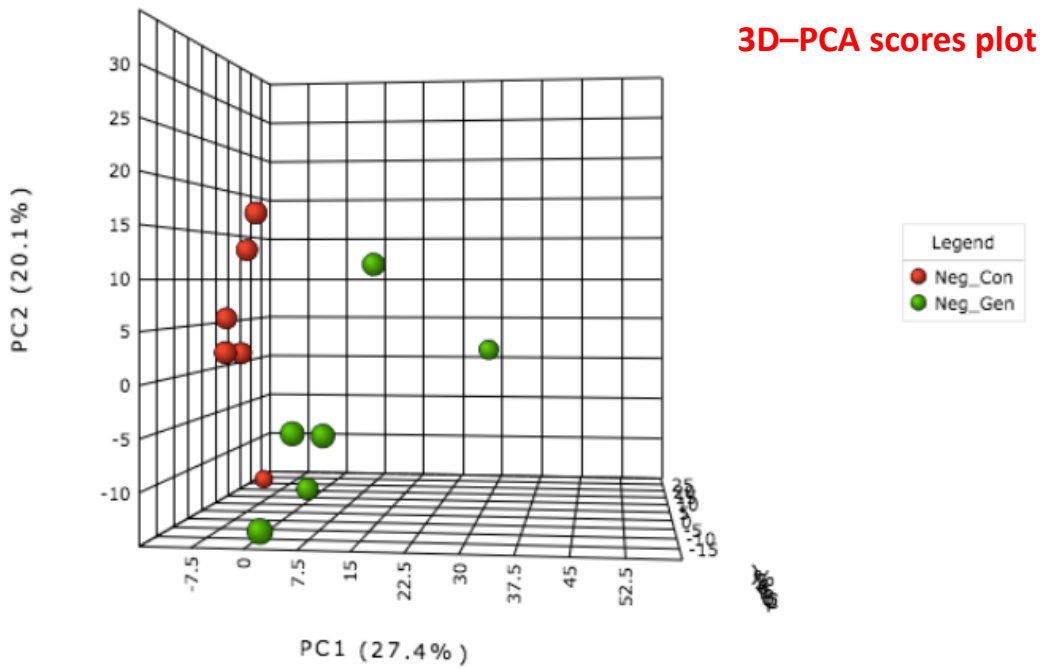
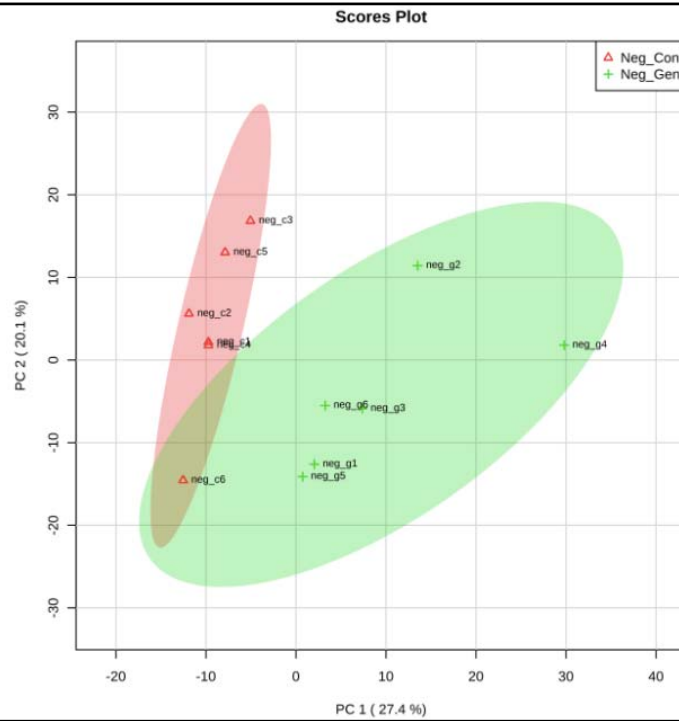
Scree plot

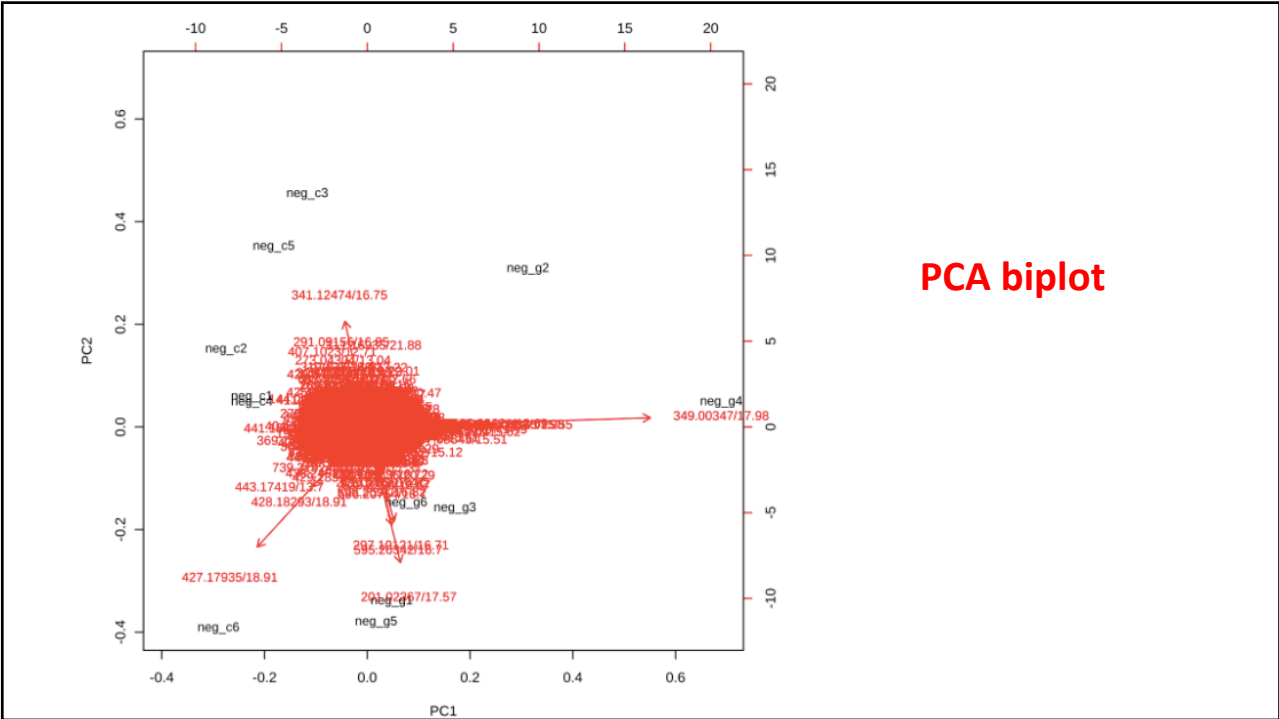
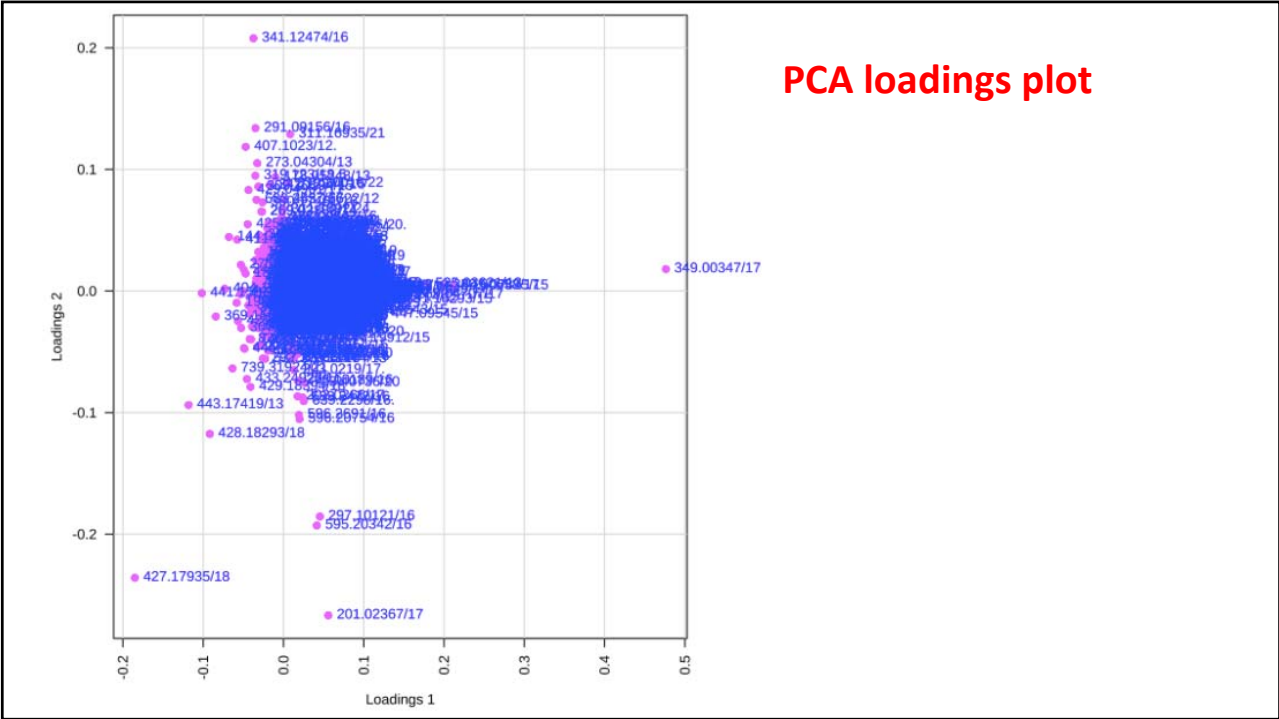


Variation explained by the components

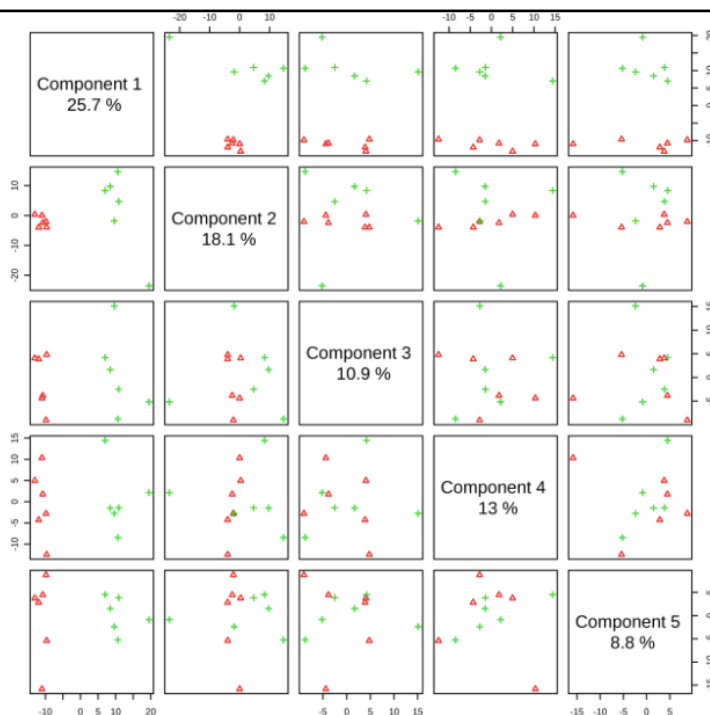
Principal components analysis (PCA) – components 1 and 2

This analysis is **unsupervised**

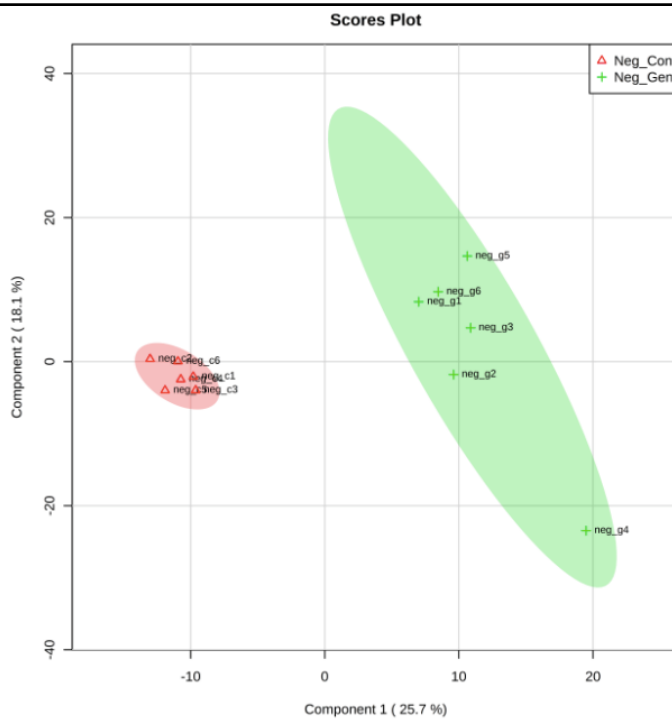


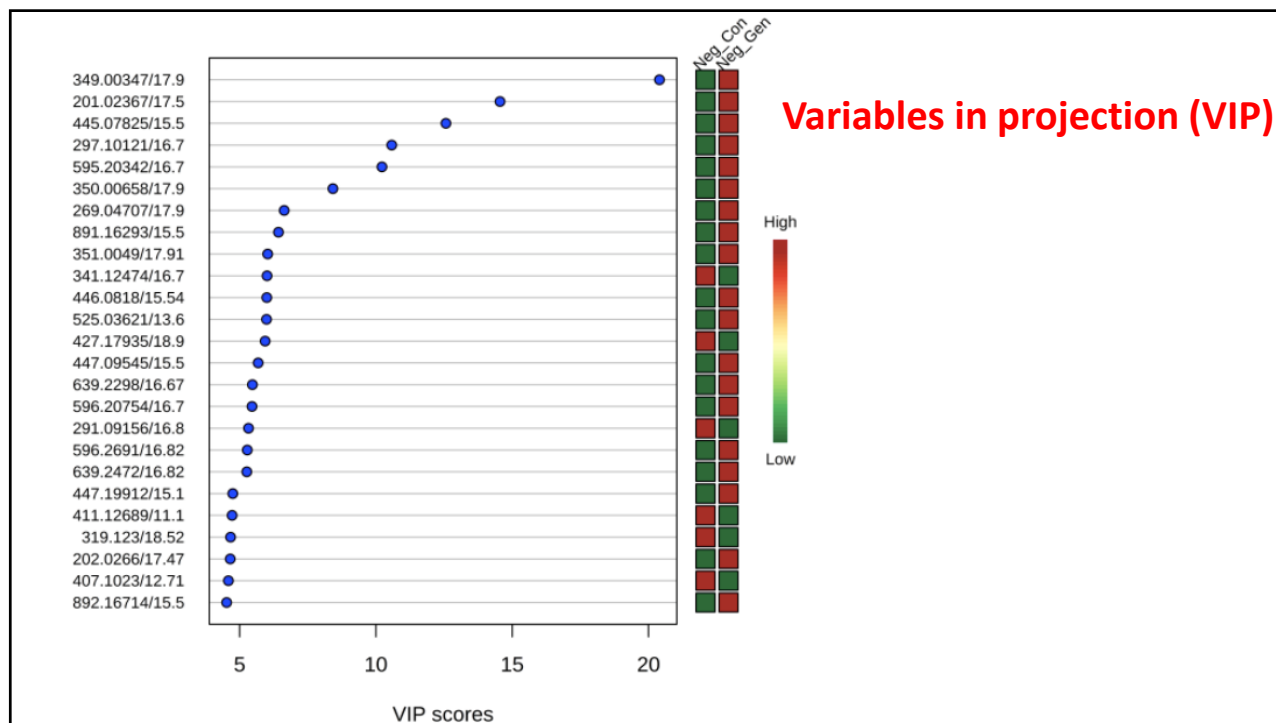
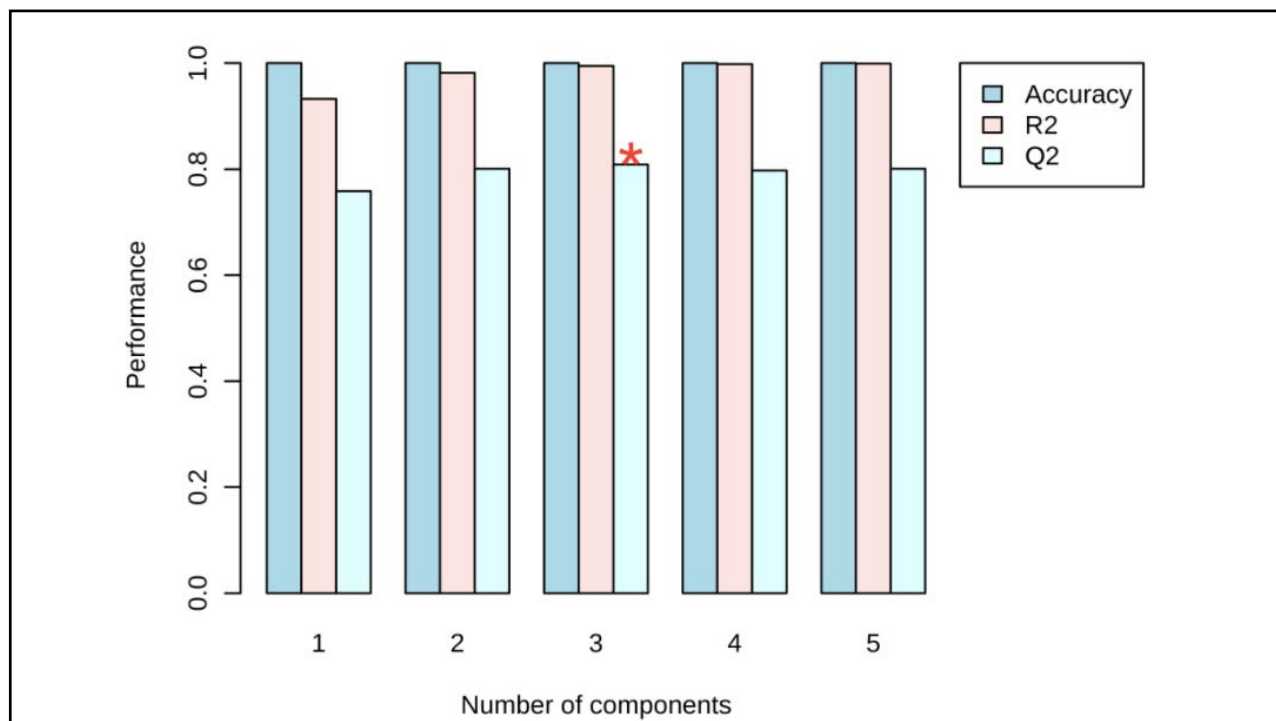


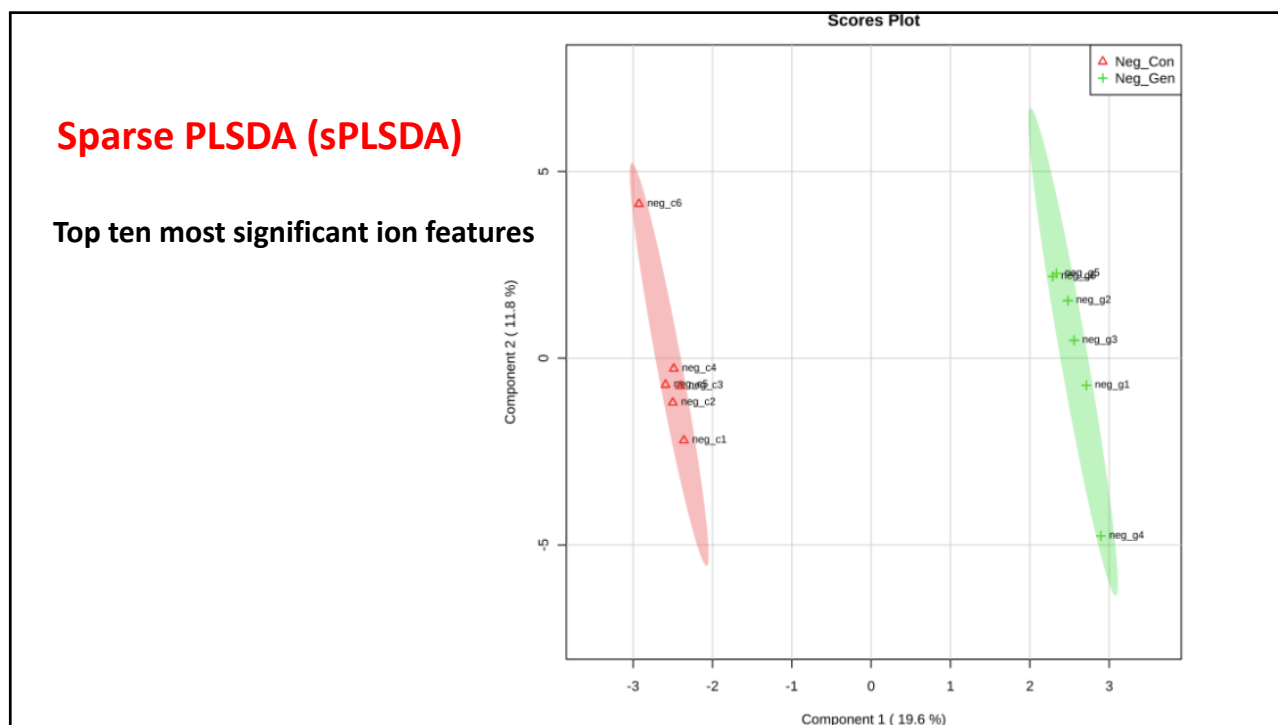
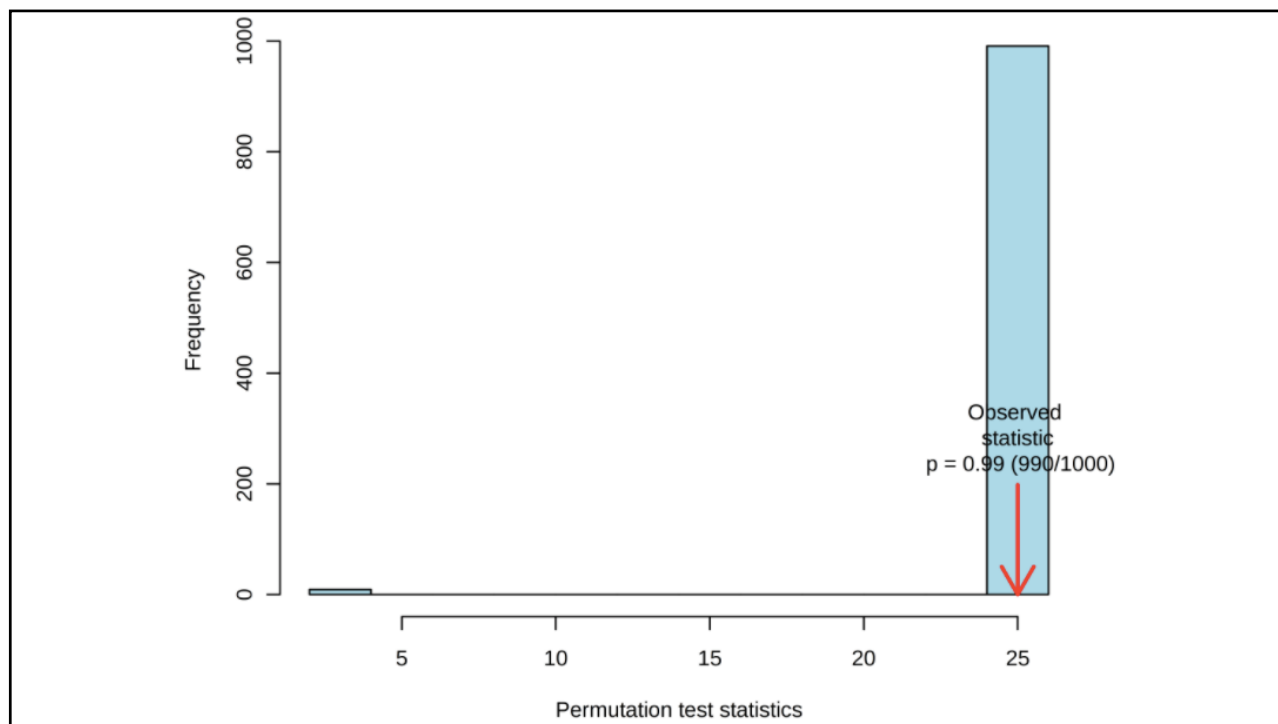
Partial Least Squares Discriminant Analysis (PLSDA)



2D-PLSDA plot







Top 50 features in the VIP

m/z	RT	Comp. 1			
349.0035	17.98	20.402			
201.0237	17.57	14.552			
445.0783	15.55	12.568			
297.1012	16.71	10.580			
595.2034	16.70	10.221			
350.0066	17.98	8.422			
269.0471	17.90	6.636			
891.1629	15.52	6.426			
351.0049	17.91	6.030			
341.1247	16.75	6.007			
446.0818	15.54	5.997			
525.0362	13.69	5.988			
427.1794	18.91	5.936			
447.0955	15.51	5.680			
639.2298	16.67	5.469			
596.2075	16.70	5.455			
291.0916	16.85	5.327			
596.2691	16.82	5.281			
639.2472	16.82	5.264			
447.1991	15.12	4.749			
411.1269	11.14	4.724			
319.1230	18.52	4.669			
202.0266	17.47	4.657			
407.1023	12.71	4.589			
892.1671	15.52	4.525			
404.1916	13.38	4.404			
298.1019	16.72	4.217			
893.1754	15.51	4.098			
429.0496	11.92	3.962			
441.1588	15.05	3.866			
275.0969	18.54	3.804			
683.2557	16.74	3.791			
309.1019	15.43	3.704			
425.1785	16.68	3.618			
443.1742	13.70	3.560			
203.0219	17.37	3.547			
823.2609	11.13	3.416			
640.2337	16.67	3.143			
269.0497	15.52	3.131			
415.1972	19.11	3.097			
640.2685	16.84	3.001			
428.1829	18.91	2.980			
380.1746	16.69	2.975			
373.1147	13.60	2.935			
526.0410	13.70	2.914			
537.2335	18.41	2.900			
269.0478	20.56	2.879			
273.0430	13.04	2.852			
369.1560	21.45	2.838			
144.0477	17.12	2.745			

m/z	RT	Comp. 1							
144.0477	17.12	2.745	349.0035	17.98	20.402	513.0725	13.83	2.345	
186.1152	17.75	2.582	350.0066	17.98	8.422	523.0690	18.33	2.267	
187.0829	16.87	2.624	351.0049	17.91	6.030	525.0362	13.69	5.988	
198.1149	18.69	2.340	352.0058	18.07	2.334	526.0410	13.70	2.914	
201.0237	17.57	14.552	367.1413	20.65	2.178	526.1134	13.34	2.114	
202.0266	17.47	4.657	369.1560	21.45	2.838	527.0433	13.70	2.157	
203.0219	17.37	3.547	373.1147	13.60	2.935	537.2335	18.41	2.900	
204.0679	16.05	2.058	380.1746	16.69	2.975	587.3044	17.35	2.039	
214.1105	17.43	2.034	404.1916	13.38	4.404	595.2034	16.70	10.221	
242.1407	14.95	2.595	405.1933	13.37	2.210	596.2075	16.70	5.455	
269.0211	14.06	2.549	407.1023	12.71	4.589	596.2691	16.82	5.281	
269.0388	14.03	2.486	408.1059	12.69	2.001	597.2109	16.70	2.652	
269.0471	17.90	6.636	411.1269	11.14	4.724	597.2637	16.82	2.517	
269.0478	20.56	2.879	412.1296	11.14	2.135	621.1138	12.61	2.152	
269.0497	15.52	3.131	415.1972	19.11	3.097	639.2298	16.67	5.469	
270.0494	17.90	2.723	425.1642	12.09	2.352	639.2472	16.82	5.264	
273.0430	13.04	2.852	425.1652	18.41	2.394	640.2337	16.67	3.143	
275.0969	18.54	3.804	425.1785	16.68	3.618	640.2685	16.84	3.001	
291.0916	16.85	5.327	427.1794	18.91	5.936	668.1241	15.50	2.634	
292.0935	16.88	2.121	428.1829	18.91	2.980	668.6207	15.51	2.212	
293.0992	17.00	2.004	429.0496	11.92	3.962	675.2739	16.80	2.249	
297.1012	16.71	10.580	432.2197	16.48	2.112	675.3014	16.75	2.277	
298.1019	16.72	4.217	433.0793	10.75	2.258	683.2557	16.74	3.791	
307.0770	17.22	2.037	433.2049	15.50	2.076	684.2595	16.74	2.322	
309.1019	15.43	3.704	433.2492	16.49	2.319	739.3192	21.46	2.186	
317.1071	18.45	2.021	435.2224	15.30	2.019	767.3151	16.74	2.267	
319.1230	18.52	4.669	441.1588	15.05	3.866	823.2609	11.13	3.416	
320.1281	18.49	2.092	443.1742	13.70	3.560	824.2636	11.13	2.078	
337.1327	15.13	2.341	445.0783	15.55	12.568	891.1629	15.52	6.426	
337.1331	16.70	2.245	446.0818	15.54	5.997	892.1671	15.52	4.525	
341.1247	16.75	6.007	447.0955	15.51	5.680	893.1754	15.51	4.098	
342.1284	16.73	2.571	447.1991	15.12	4.749	894.1822	15.52	2.461	
349.0035	17.98	20.402	449.1075	15.77	2.706				