

MetaboAnalystを用いた PCA, PLS-DA

<https://www.metaboanalyst.ca/>

データファイルの作り方

各サンプルが各行になるようにする。行と列を入れ替えることも可能

1列目にID、2列目にサンプルの分類情報、3列目以降に変数データを記入する

id	class	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
1	M	100	100	50	70	10	100	100	
2	F	80	100	90	100	0	90	100	
3	F	50	100	100	100	0	70	80	
4	F	100	100	50	50	0	70	100	
5	F	94	80	98	30	60	98	99	
6	F	99	100	80	70	0	10	60	
7	M	50	40	80	50	0	80	40	
8	F	55	50	69	100	0	0	15	
9	F	25	0	100	100	0	25	75	
10	F	30	75	80	60	0	90	80	
11	M	85	100	85	90	0	100	40	
12	M	23	74	87	72	0	68	82	
13	F	70	100	80	100	10	60	60	
14	F	55	90	100	80	30	50	42	
15	F	50	80	100	80	0	30	100	
16	F	90	50	70	70	0	0	40	
17	M	80	30	100	100	2	30	95	

転置可能



データファイルの作り方

カンマ区切りファイル(csv)にしたいデータだけを記載した、エクセルシートを作る。

そのシートを閲覧している状態で、「名前を付けて保存」から、カンマ区切り(csv)を選んで保存する。

MetaboAnalystへアクセス

Click here to startをクリック



MetaboAnalyst 6.0 - from raw spectra to biomarkers, patterns, functions and systems biology

[Home](#)

[Data Formats](#)

[Tutorials](#)

[User Forum](#)

[MetaboAnalystR](#)

[Publications](#)

[Update History](#)

[Databases](#)

[APIs](#)

[User Stats](#)

[Data Policy](#)

[About](#)

[Contact](#)

News & Updates

- Fixed the issue with two-way ANOVA for large data size (08/01/2025) **NEW**;
- Added support for computing partial correlation for Pattern Search and correlation heatmaps (07/18/2025) **NEW**;
- Enhanced LC-MS and MS/MS result integration to enable simultaneous assessment of quantitative differences and annotation quality (07/15/2025) **NEW**;
- Pair-wise PCA plot now offer p-values to help assess patterns w.r.t discrete or continuous responses (07/12/2025) **NEW**;
- Fixed the issue with PDF image generation and PCA analysis when data contains very small number of features (07/10/2025) **NEW**;
- Added two missing value imputation methods: quantile regression imputation of left-censored data (QRILC) and MissForest (07/08/2025) **NEW**;
- You can now set group-wise threshold to perform data filtering based on missing values (07/06/2025) **NEW**;
- Added support for Steiger filtering and literature evidence for reverse causality check in MR analysis (07/04/2025) **NEW**;
- Added two normalization options: Log2 normalization and variance stabilizing normalization (07/02/2025) **NEW**;
- T-tests will automatically try non-parametric t-tests when regular t-tests failed (06/30/2025) **NEW**;
- Added diagnostic graphics for missing values and RSD distributions for data integrity check and processing (06/25/2025) **NEW**;
- Added support for enrichment network to explore pathway analysis results for both targeted and untargeted metabolomics (06/18/2025) **NEW**;
- Improved the default graphical resolution to 150 dpi for better user experience (06/12/2025) **NEW**;

[Read more](#)

[Click here to start](#)



Overview

MetaboAnalyst is a web-based platform dedicated for comprehensive metabolomics data analysis, interpretation and integration with other omics data. Over the past decade, MetaboAnalyst has evolved from statistical and functional analysis for targeted metabolomics data, towards more streamlined analysis for both quantitative and untargeted metabolomics data. In addition to many feature enhancements, the version 6.0 contains three new modules - tandem MS spectral processing and **compound annotation**, dose response analysis for **chemical risk assessment**, and leveraging metabolite-genome wide association analysis and Mendelian randomization for **causal analysis**.



Statistical Analysis [single factor]

Statistical Analysis [one factor]をクリック

Module Overview

Input Data Type	Available Modules (click on a module to proceed, or scroll down to explore a total of 18 modules including utilities)				
LC-MS Spectra (mzML, mzXML or mzData)			Spectra Processing [LC-MS w/wo MS2]		
MS Peaks (peak list or intensity table)		Peak Annotation [MS2-DDA/DIA]	Functional Analysis [LC-MS]	Functional Meta-analysis [LC-MS]	
Generic Format (.csv or .txt table files)	Statistical Analysis [one factor]	Statistical Analysis [metadata table]	Biomarker Analysis	Dose Response Analysis	Statistical Meta-analysis
Annotated Features (metabolite list or table)		Enrichment Analysis	Pathway Analysis	Network Analysis	
Link to Genomics & Phenotypes (metabolite list)			Causal Analysis [Mendelian randomization]		

>> [Spectral Processing \[LC-MS1 w/wo MS2\]](#)

This module allows users to upload raw LC-MS spectra (mzML, mzXML or mzData) to be processed using our optimized workflow based on MetaboAnalystR 4.0 or the latest [asari](#) algorithm. Users can also include MS2 spectra (both DDA or SWATH-DIA are supported) for peak annotation.

>> [Peak Annotation \[MS2-DIA/DDA\]](#)

This module performs MS2 peak annotation based on a comprehensive list of public databases. Users can either directly enter a two-column peak list containing m/z and intensity values (DDA); or upload a .msp file produced by MZmine or MS-DIAL after the spectral deconvolution (SWATH-DIA).

>> [Functional Analysis \[LC-MS1\]](#)

This module accepts high-resolution LC-MS spectral peak data to perform metabolic pathway enrichment analysis and visual exploration based on the [mummichog](#) or [GSEA](#) algorithms. It currently supports 26 organisms including Human, Mouse, Zebrafish, *C. elegans*, and other species.

準備したCSVファイルをアップロード

必要に応じて、Formatなどを選択する

Submitボタンを押す

Please upload your data

A plain text file (.txt or .csv): ?

Data Type:



Concentrations



Spectral bins



Peak intensities

Format:

Samples in rows (unpaired) ▼

Data File:



Choose

240730_okonomi_data.csv 1.4 KB

Submit

データチェックが行われる 問題なければskipで先に進む

Data Integrity Check:

- Checking sample names - spaces will be replaced with underscore, and special characters will be removed;
- Checking the class labels - at least three replicates are required in each class.
- The data (except class labels) must not contain non-numeric values.
- If the samples are paired, the pair labels must conform to the specified format.
- 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 29 (samples) by 12 (compounds) data matrix.

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.

No missing values were detected. Click the **Proceed** button to the next step.

Edit Groups

▶ Proceed

必要に応じて、データの標準化(normalization)、変換(transformation)、スケーリング方法を選択する

Normalizeボタンを押す

Normalization Overview:

The normalization procedures are grouped into three categories. You can use one or combine them to achieve better results.

- Sample normalization is for general-purpose adjustment for systematic differences among samples;
- Data transformation applies a mathematical transformation on individual values themselves. A simple mathematical approach is used to deal with OmicsForum using "normalization #metaboanalyst" to find more information.
- Data scaling adjusts each variable/feature by a scaling factor computed based on the dispersion of the variable.

Sample normalization

☒ None

☐ Sample-specific normalization (i.e. weight, volume) [Specify](#)

☐ Normalization by sum

☐ Normalization by median

☐ Normalization by a reference sample (PQN) [Specify](#)

☐ Normalization by a pooled sample from group (group PQN) [Specify](#)

☐ Normalization by reference feature [Specify](#)

☐ Quantile normalization (suggested only for > 1000 features)

Data transformation

☒ None

☐ Log transformation (base 10)

☐ Log transformation (base 2)

☐ Square root transformation (square root of data values)

☐ Cube root transformation (cube root of data values)

☐ Variance stabilizing normalization (data-adaptive transformation)

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)

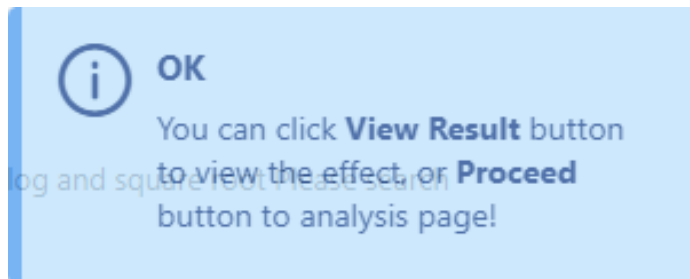
☐ Range scaling (mean-centered and divided by the range of each variable)

Normalize

View Result

Proceed

右上にこのウィンドウが出ればOK



アクティブになったProceedボタンを押す

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**

主成分分析 Principal Component Analysis (PCA) をクリック

Select an analysis path to explore :

Univariate Analysis

[Fold Change Analysis](#) [T-tests](#) [Volcano plot](#)

One-way Analysis of Variance (ANOVA)

[Correlation Heatmaps](#) [Pattern Search](#) [Correlation Networks \(DSPC\)](#)

Advanced Significance Analysis

[Significance Analysis of Microarray \(and Metabolites\) \(SAM\)](#)

[Empirical Bayesian Analysis of Microarray \(and Metabolites\) \(EBAM\)](#)

Chemometrics Analysis

[Principal Component Analysis \(PCA\)](#)

[Partial Least Squares - Discriminant Analysis \(PLS-DA\)](#)

[Sparse Partial Least Squares - Discriminant Analysis \(sPLS-DA\)](#)

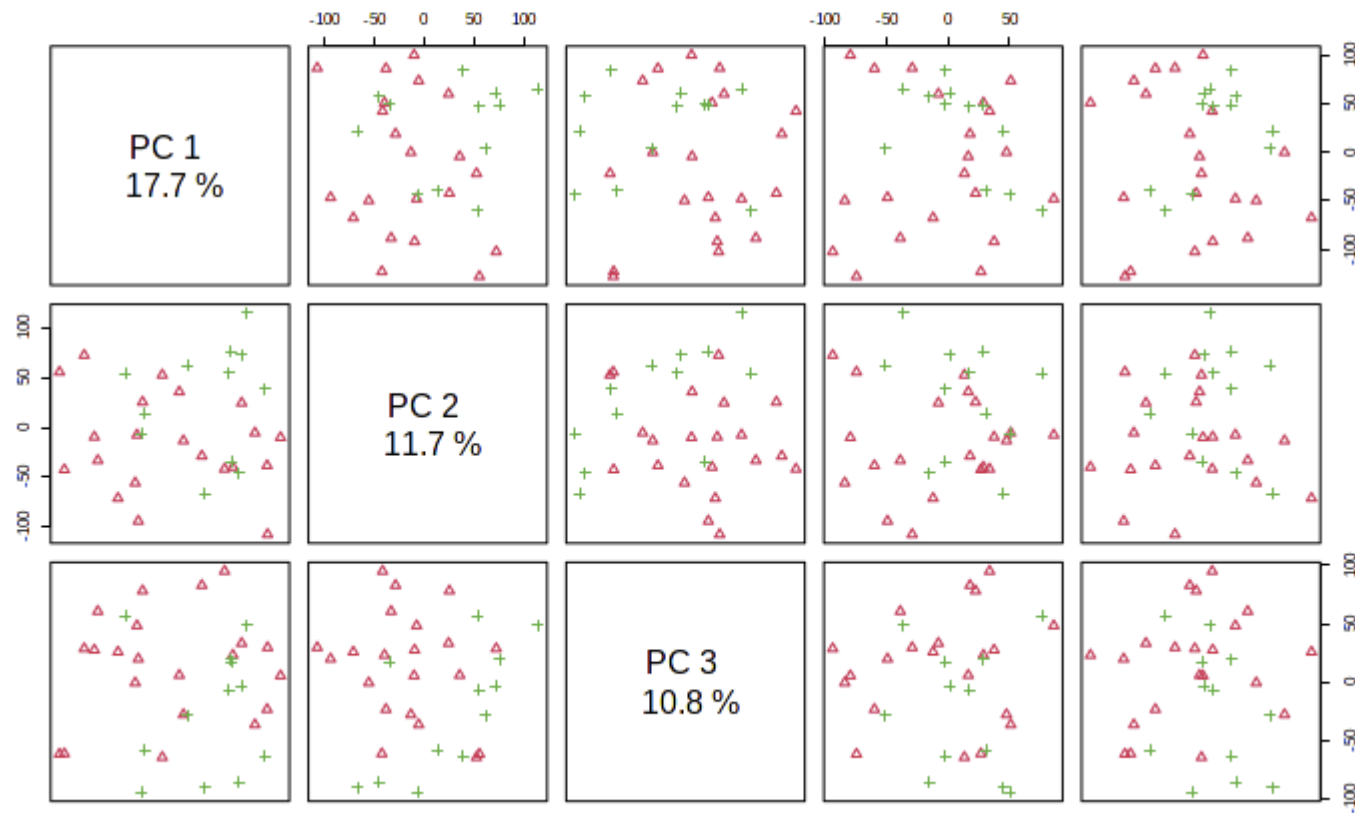
[Orthogonal Partial Least Squares - Discriminant Analysis \(orthopLS-DA\)](#)

Cluster Analysis

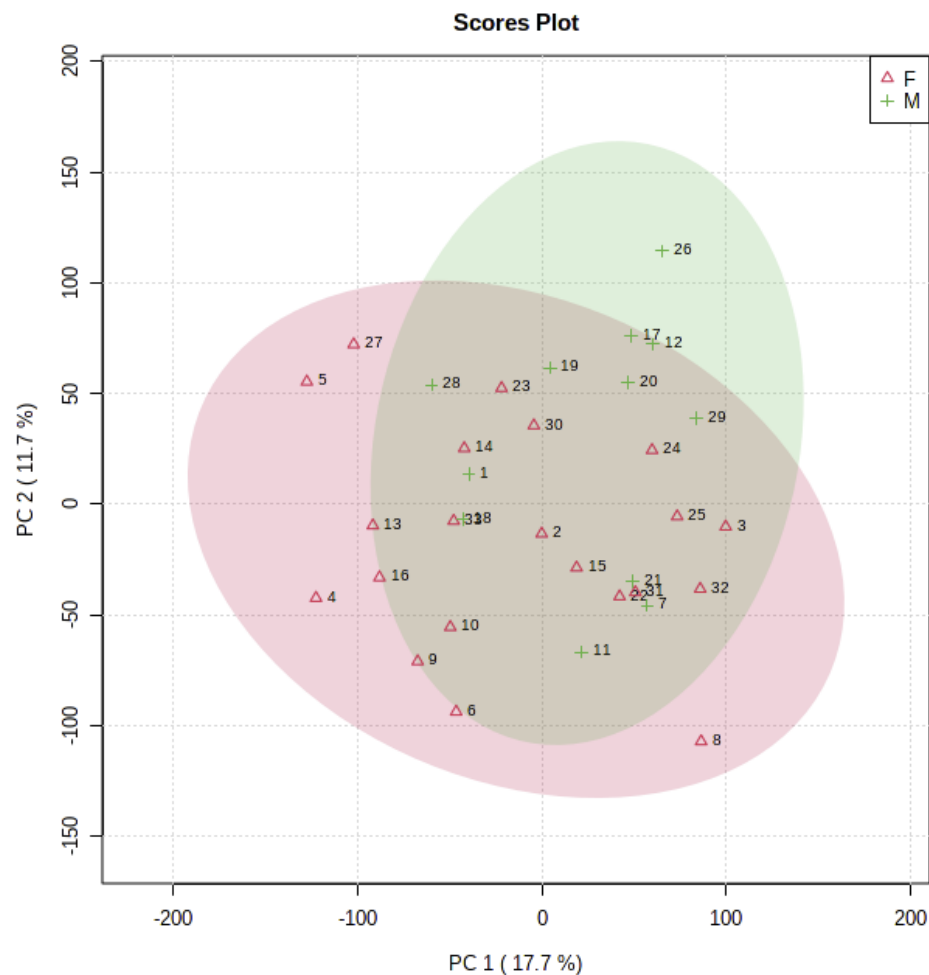
Hierarchical Clustering: [Dendrogram](#) [Heatmaps](#)

結果画面が表示される (出ない場合もあるようです)

Principal Component Analysis (PCA)

[Overview](#)[Scree Plot](#)[2D Scores Plot](#)[Loadings Plot](#)[Synchronized 3D Plots](#)[Biplot](#)Display pairwise score plot for top PCs

2D Scores Plot



Specify PC on X-axis:

Specify PC on Y-axis:

Display 95% confidence regions:

☒

Display sample names:

☒

Use grey-scale colors:

☐

Flip Image

☐

X axis

☒

Y axis

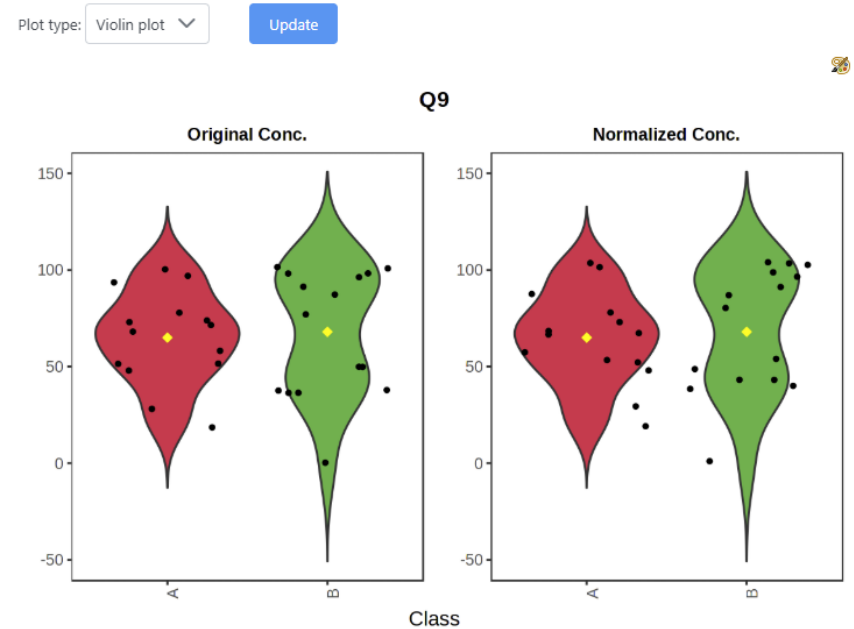
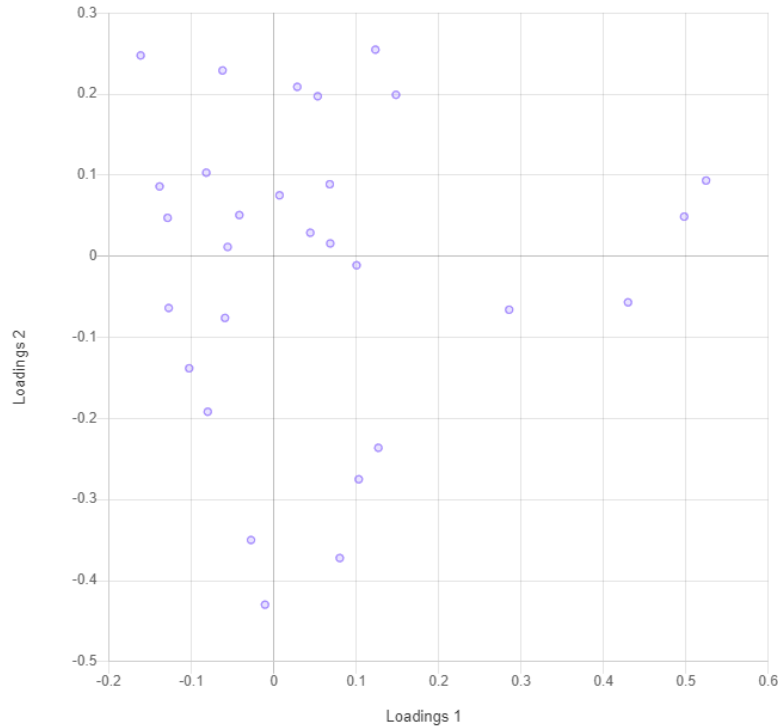
☐

All

Display sample
nameにチェックが入っ
ていると、サンプル名が
一緒に表示される。

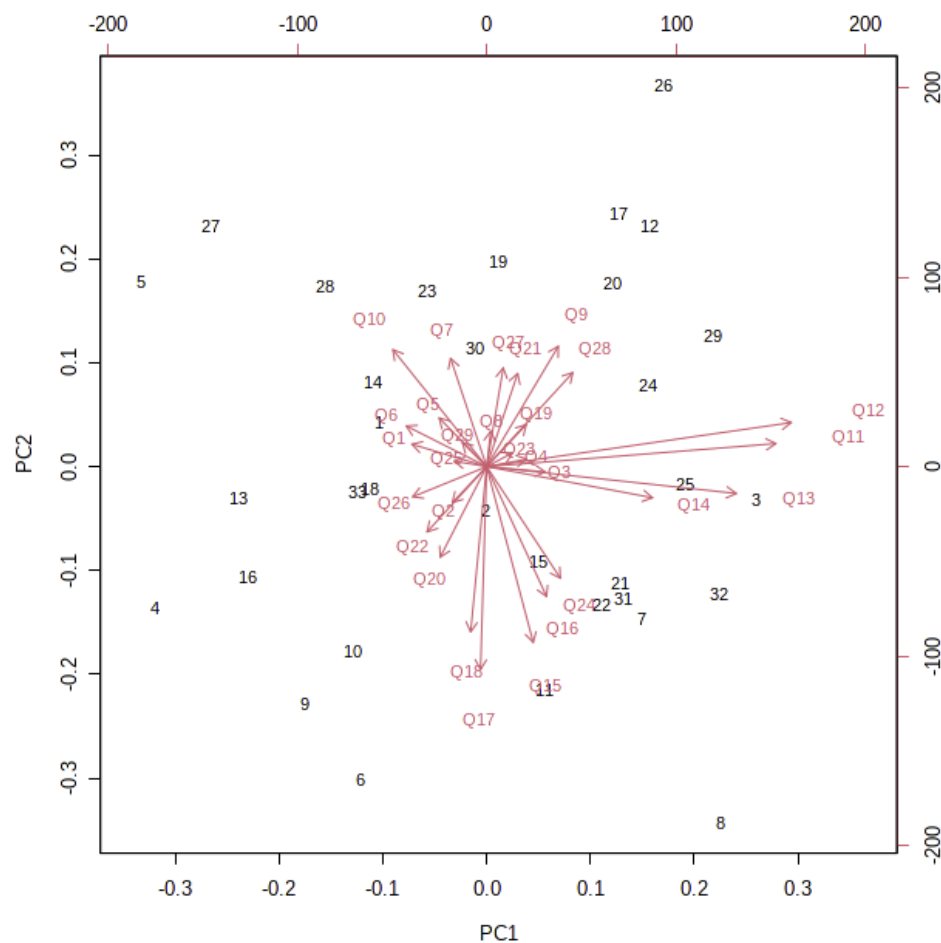
※条件を変えた後に「Update」ボタン
を押して反映させます。

Loadings Plot



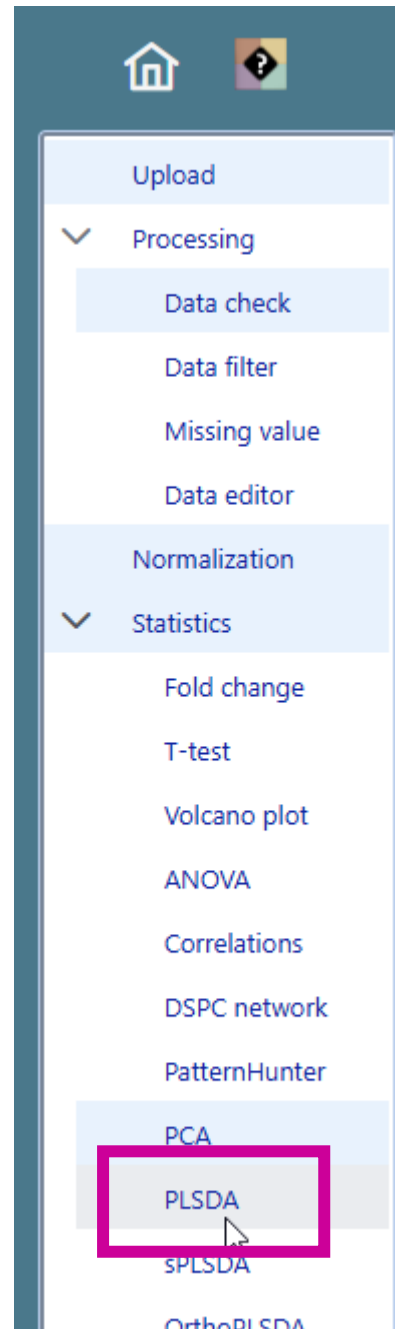
プロットをクリックすると、その変数の特徴が表示される。

Biplot

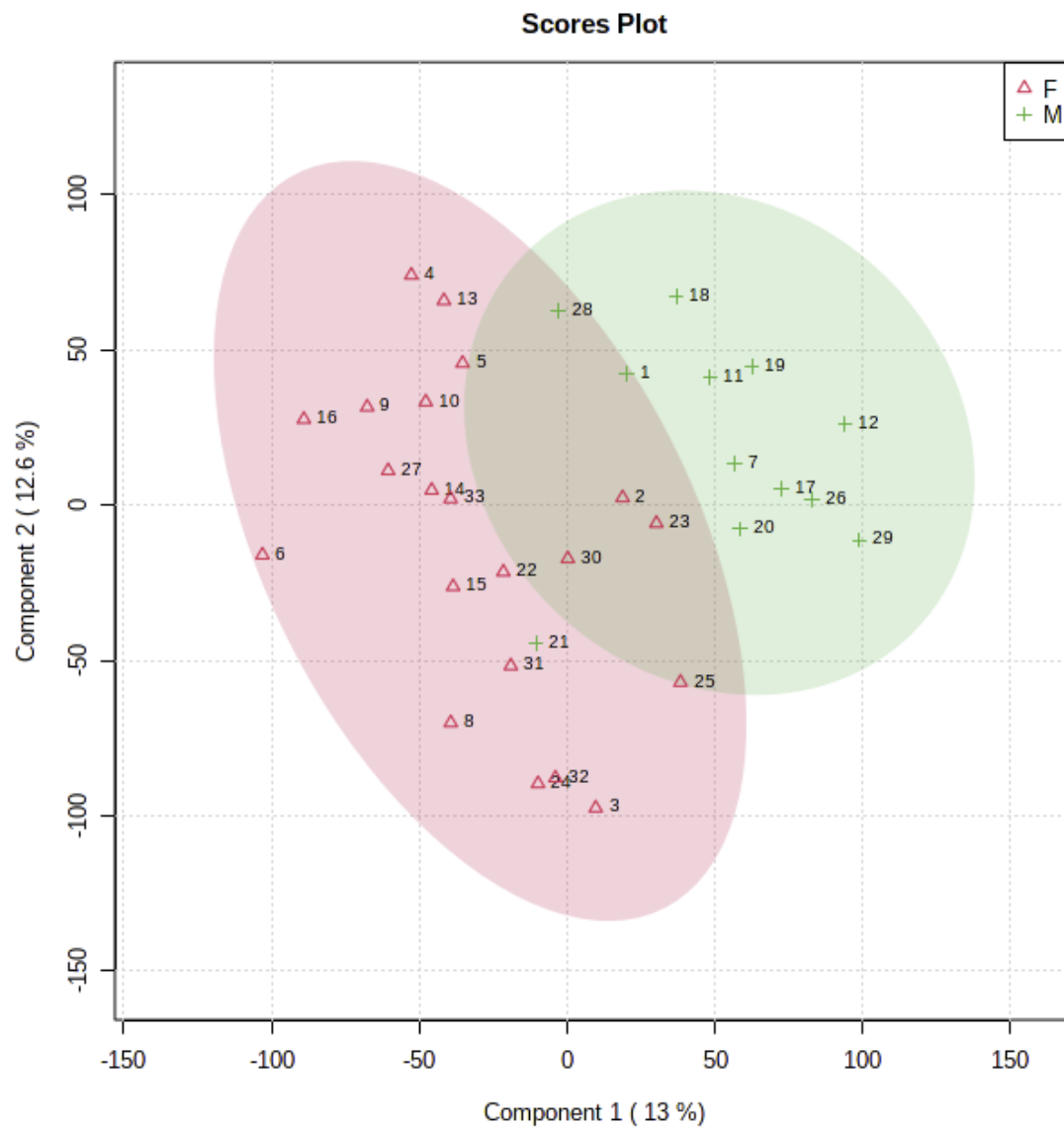


スコアプロットとローディングプロットが一緒に描かれたもの。

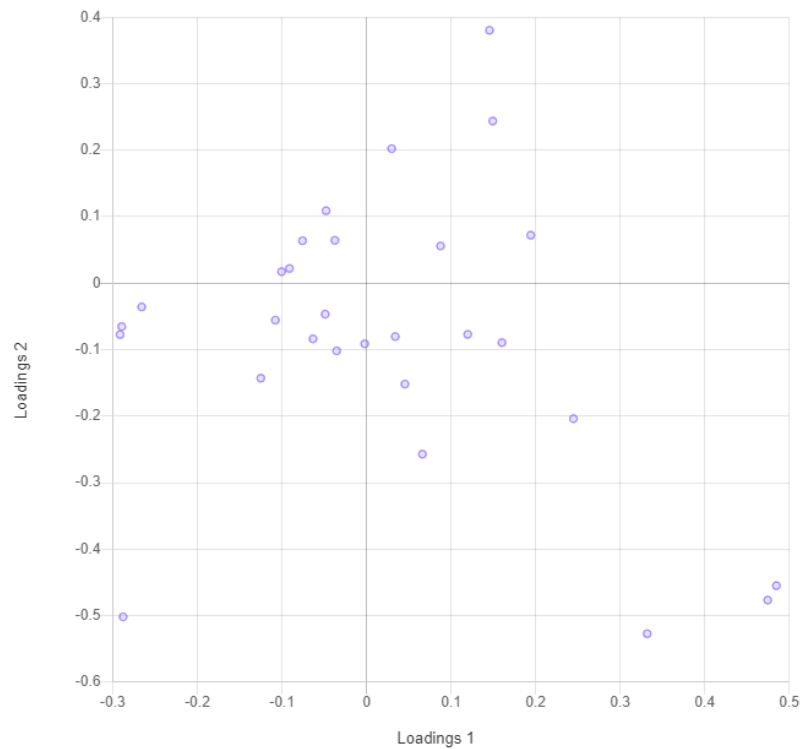
PLSDA



2D Scores Plot



Loadings Plot



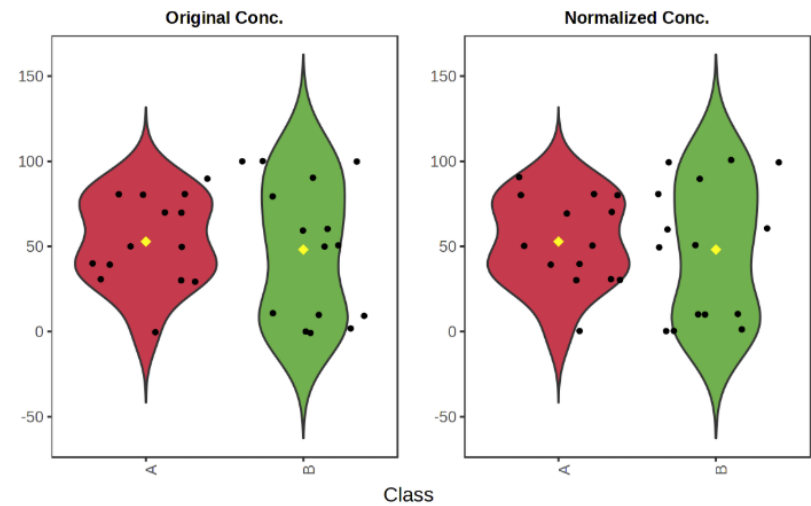
Plot type:

Violin plot ▾

Update

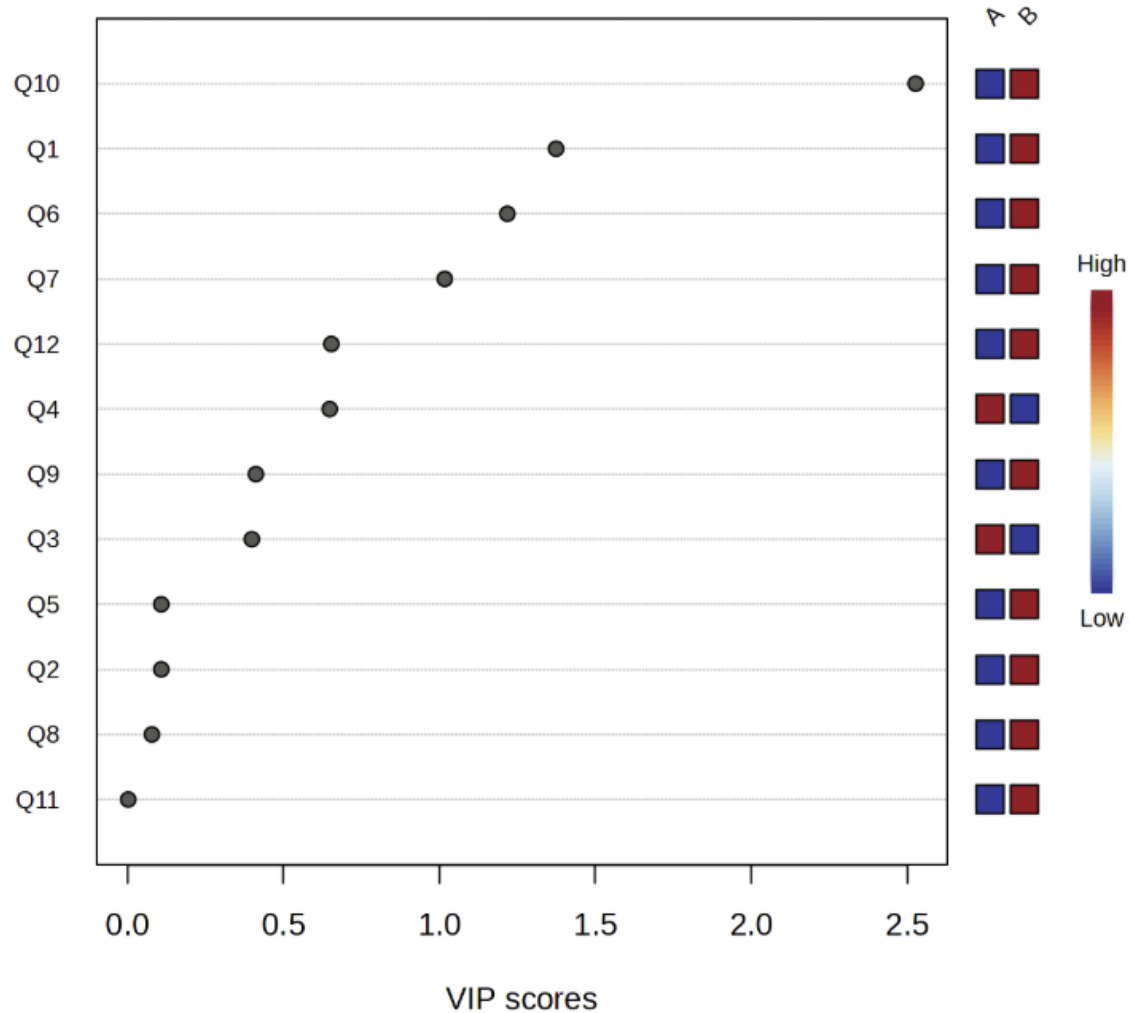


Q4



VIP値

AとBを特徴づけるのに寄与が大きい因子を探せる



累積寄与率と構築したモデルの精度(Q)

