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

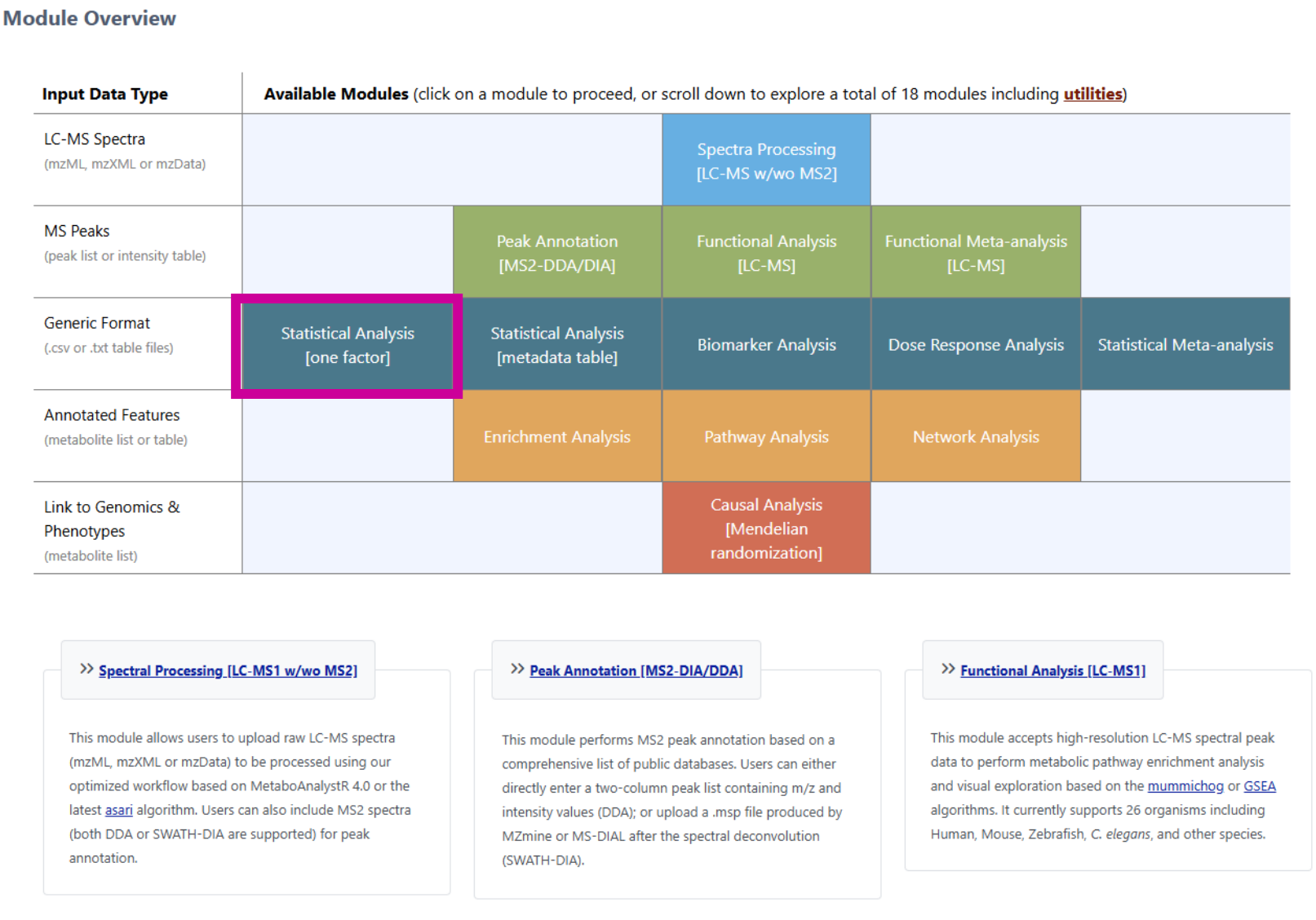
カンマ区切りファイル（csv）にしたいデータだけを記載した、エクセルシートを作る。

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）を選んでで保存する。

MetaboAnalystへアクセス

Click here to startをクリック

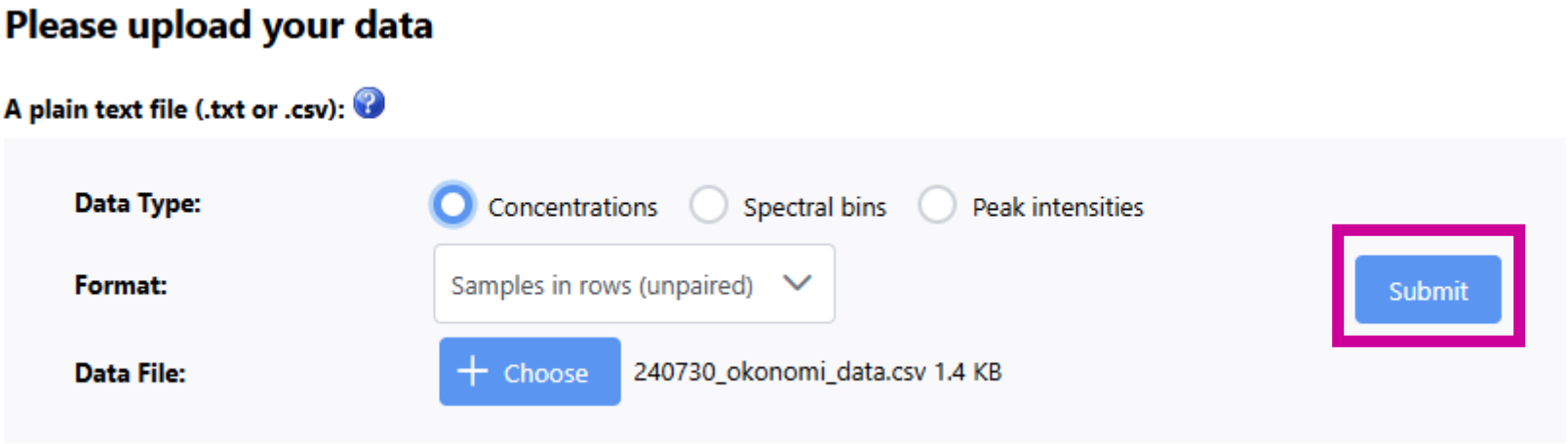
Statistical Analysis [one factor]をクリック



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



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 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

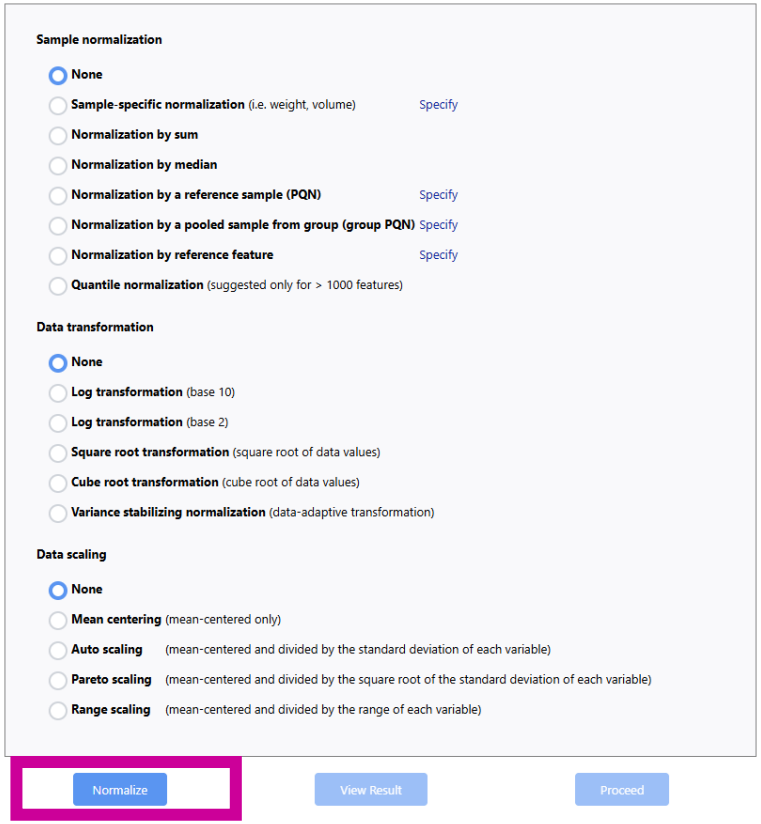
必要に応じて、データの標準化（normalization）、変換（transformation）、スケーリング方法を選択する

Normalizeボタンを押す

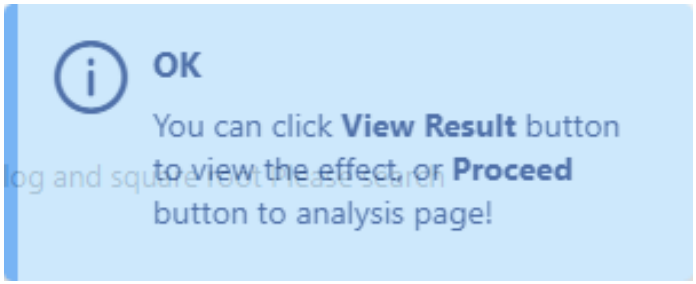
Normalization Overview:

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

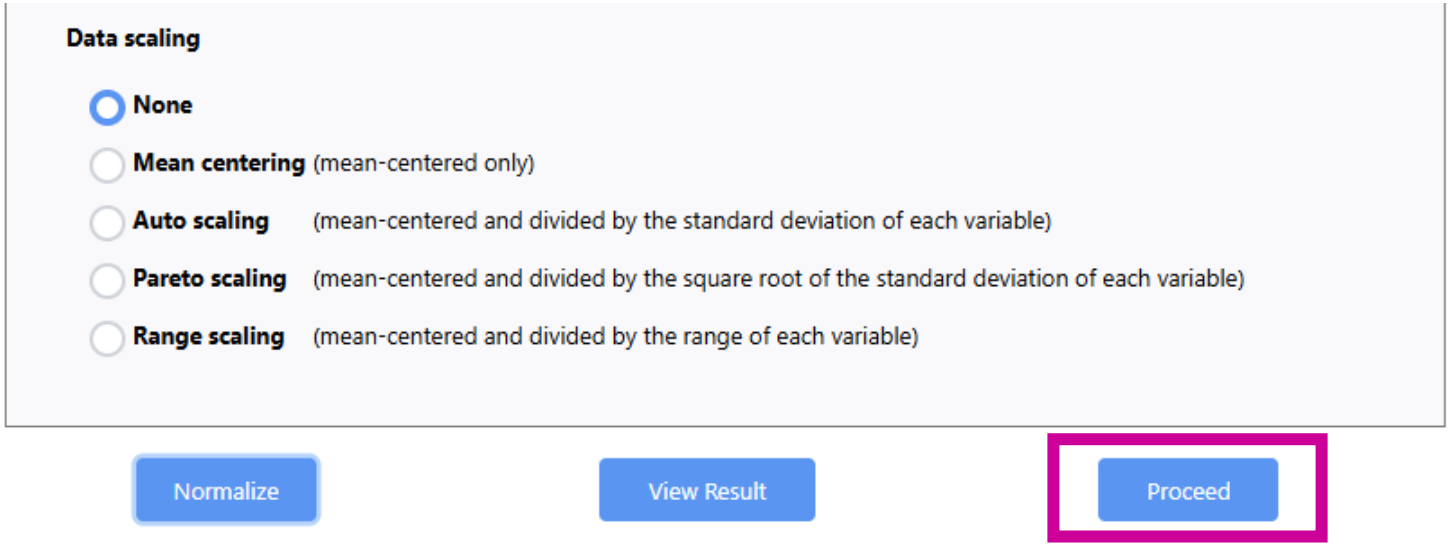
- Sample normalization is for general-purpose adjustment for systematic differences among samples;
- Data transformation applies a mathematical transformation on individual values themselves. A simple mathematical approach is used to deal with OmicsForum using "normalization #metaboanalyst" to find more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



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



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

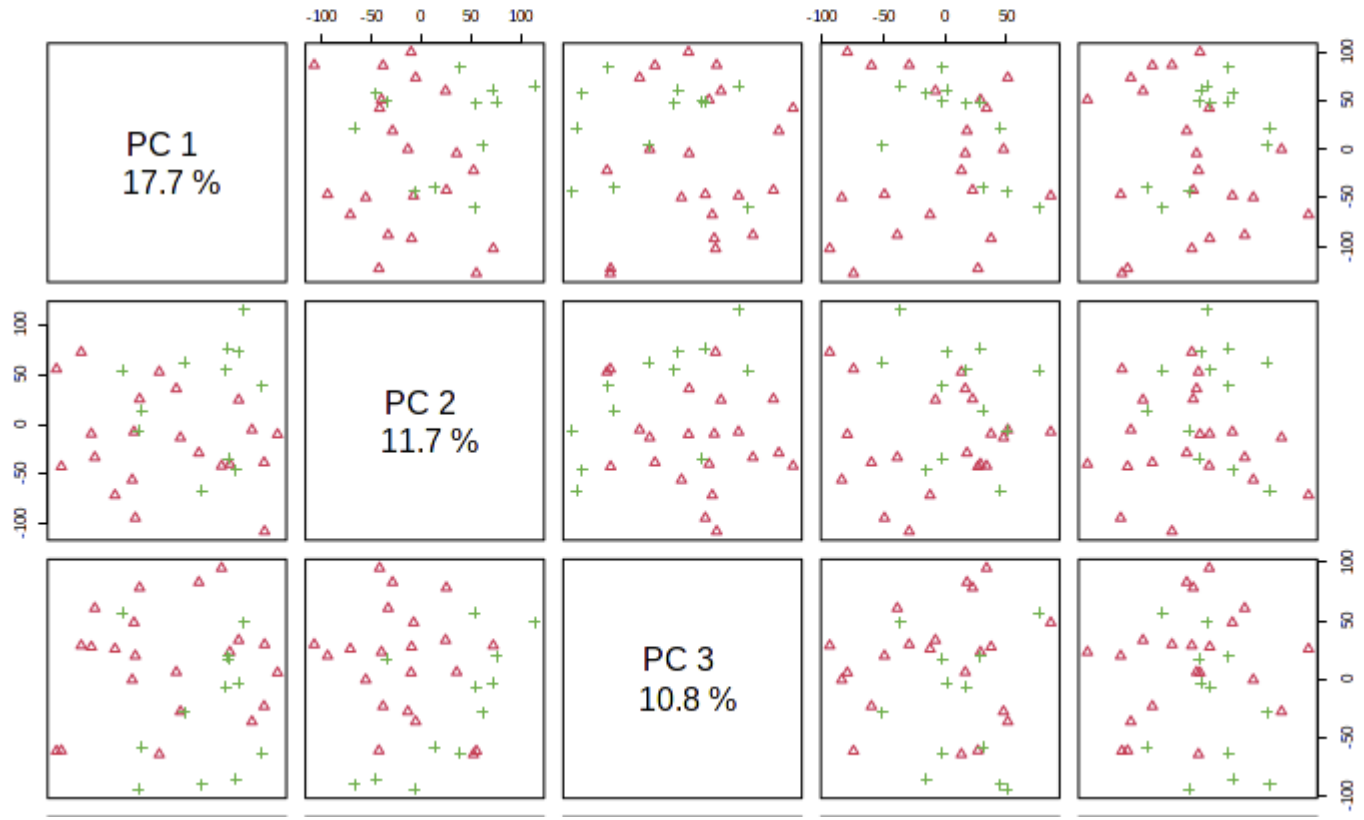
Hierarchical Clustering: Dendrogram　Heatmaps

結果画面が表示される（出ない場合もあるようです）

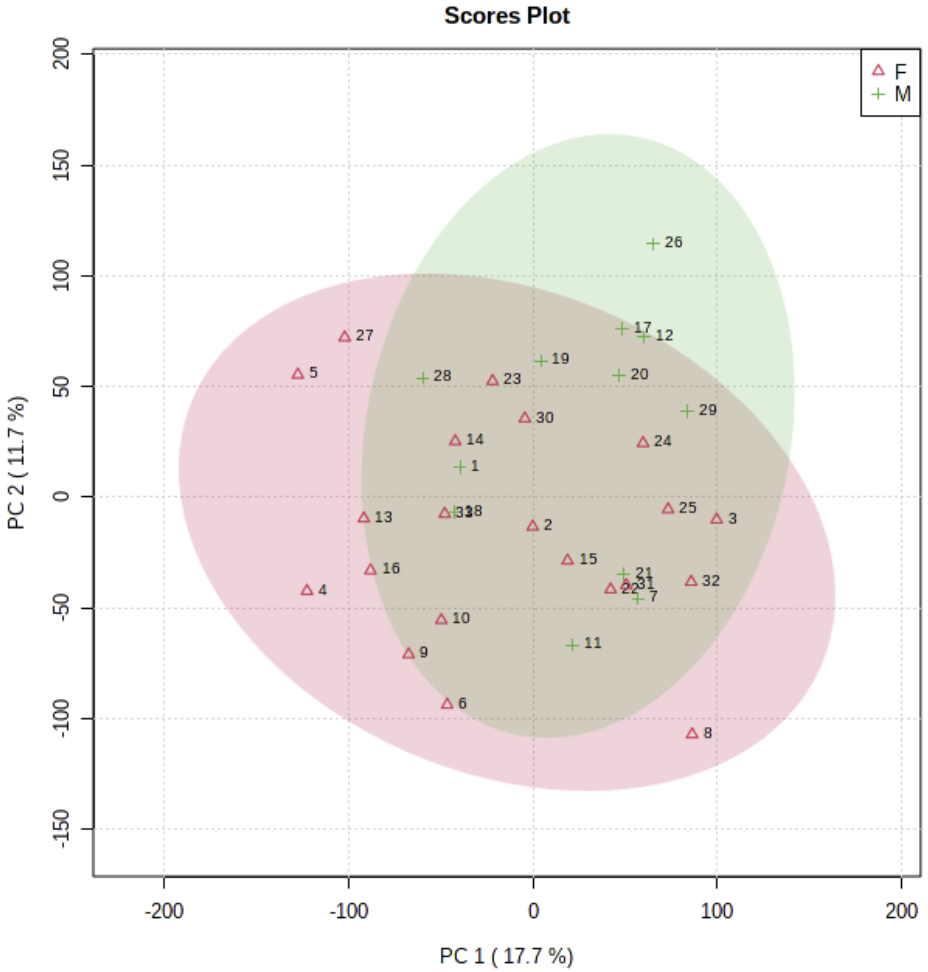
Principal Component Analysis (PCA)

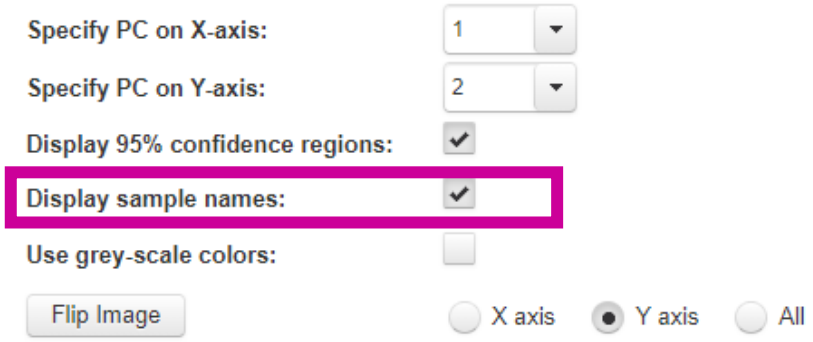
Overview | Scree Plot | 2D Scores Plot | Loadings Plot | Synchronized 3D Plots | Biplot

Display pairwise score plot for top 5 PCs Update



2D Scores Plot

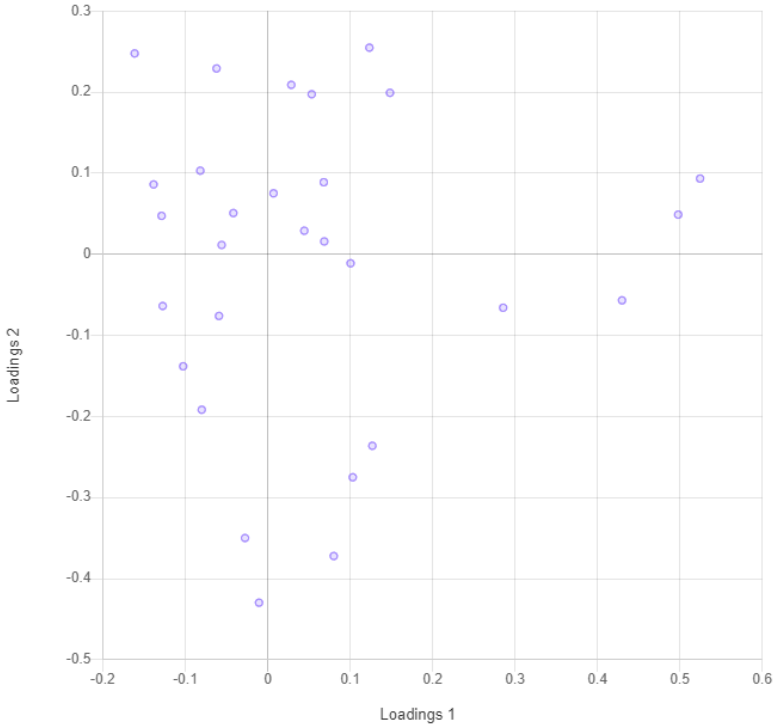


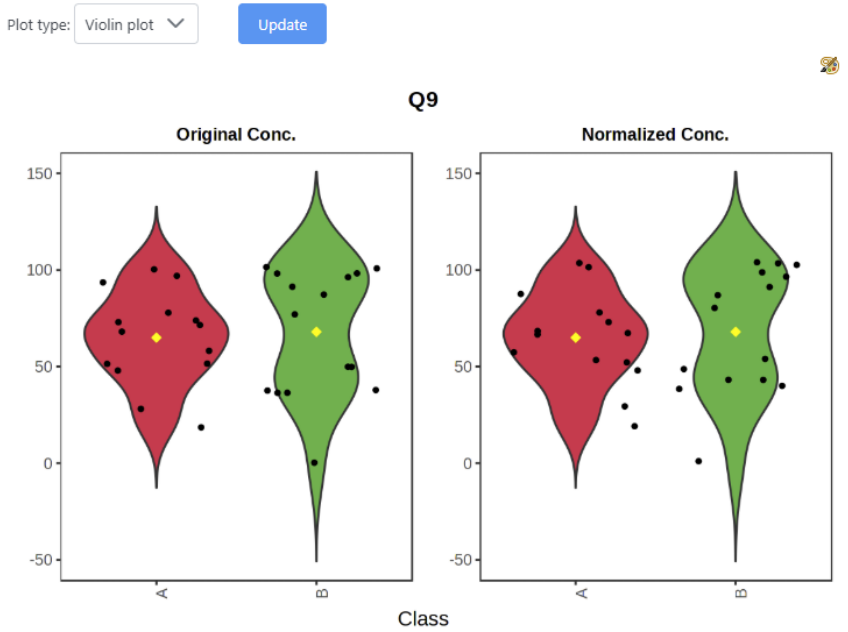


Display sample nameにチェックが入っていると、サンプル名が一緒に表示される。

※条件を変えた後に「Update」ボタンを押して反映させます。

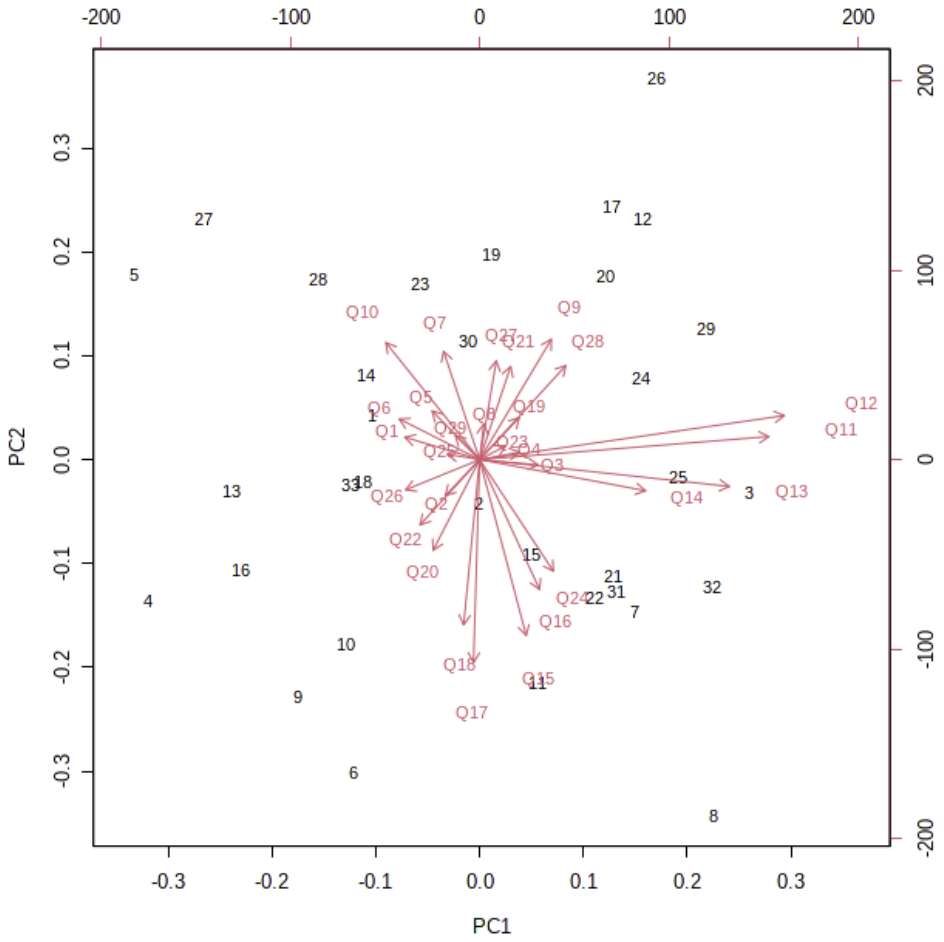
Loadings Plot





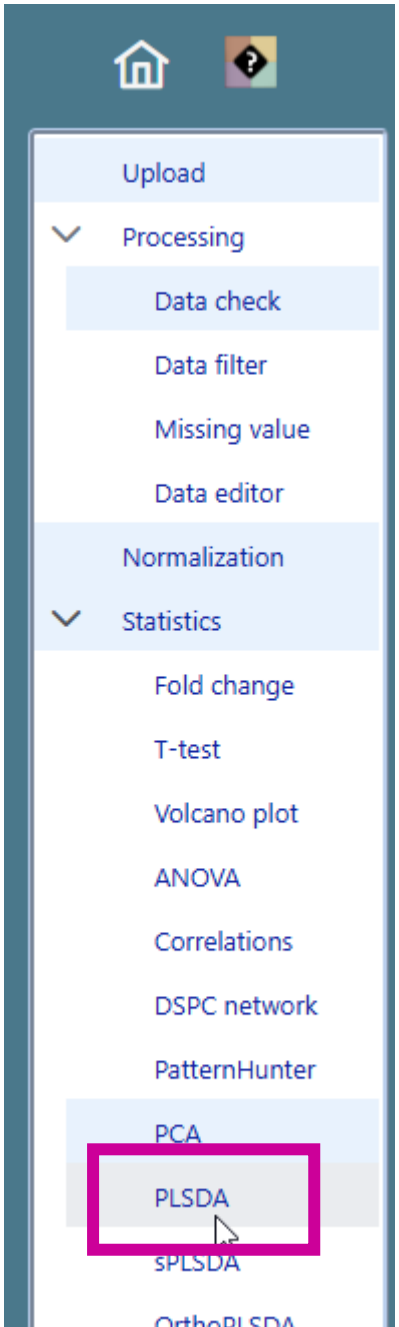
プロットをクリックすると、その変数の特徴が表示される。

Biplot

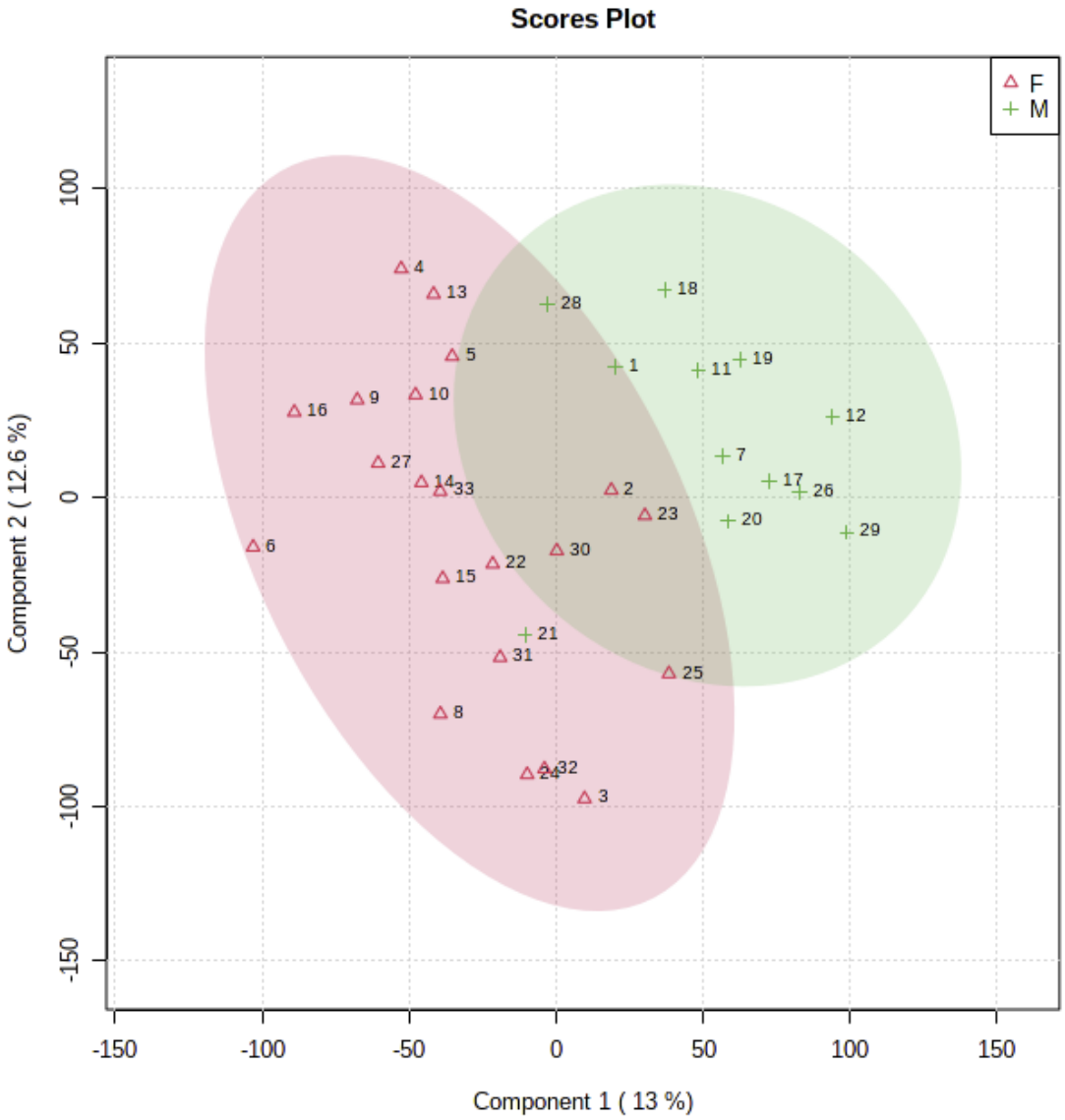


スコアプロットとローディングプロットが一緒に描かれたもの。

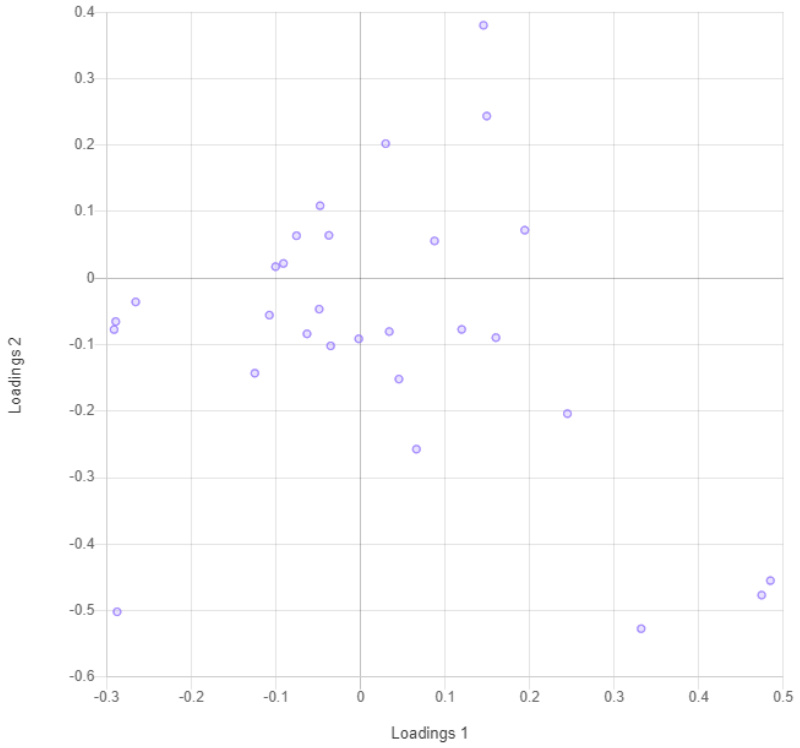
PLSDA

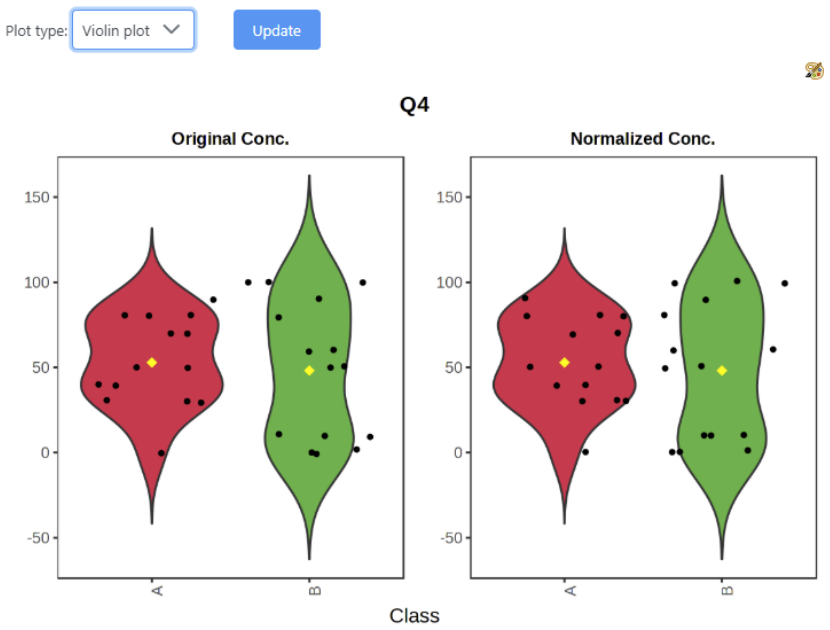

Upload
Processing
Data check
Data filter
Missing value
Data editor
Normalization
Statistics
Fold change
T-test
Volcano plot
ANOVA
Correlations
DSPC network
PatternHunter
PCA
PLSDA
sPLSDA

2D Scores Plot



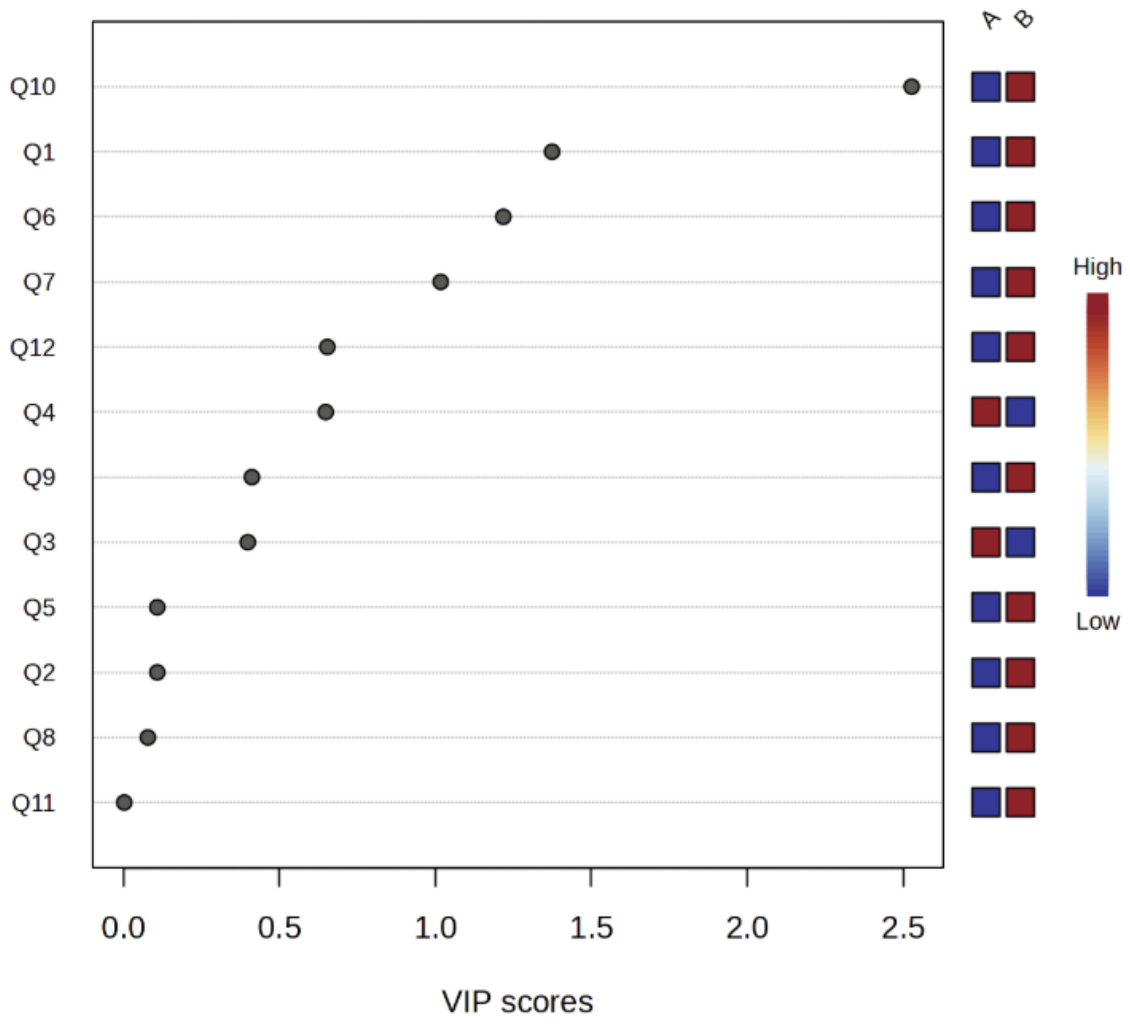
Loadings Plot





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)