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

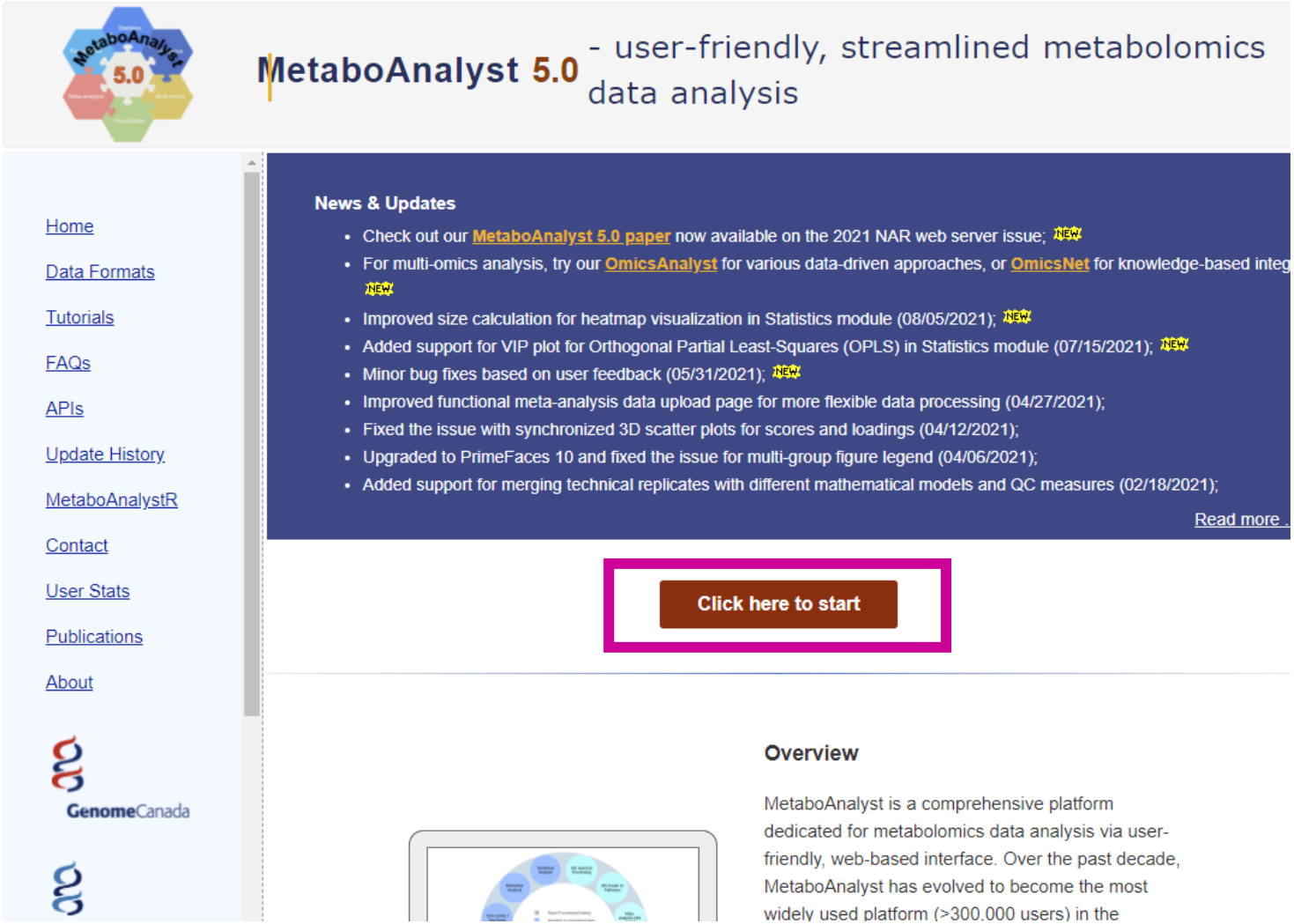
データファイルの作り方

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をクリック

Module Overview

Input Data Type	Available Modules (**click on** a module to proceed, or scroll down for more details)					
Raw Spectra (mzML, mzXML or mzData)			LC-MS Spectra Processing			
MS Peaks (peak list or intensity table)			Functional Analysis	Functional Meta-analysis		
Annotated Features (compound list or table)		Enrichment Analysis	Pathway Analysis	Joint-Pathway Analysis	Network Analysis	
Generic Format (.csv or .txt table files)	Statistical Analysis	Biomarker Analysis	Time-series/Two-factor Analysis	Statistical Meta-analysis	Power Analysis	Other Utilities

どちらでもOK

Statistical Analysis

...only used statistical and machine learning methods including t-tests, ANOVA, PCA, PLS-DA and Orthogonal PLS-DA. It also provides clustering and visualization tools to create dendrograms and heatmaps as well as to classify data based on random forests and SVM.

Biomarker Analysis

This module performs various biomarker analyses based on receiver operating characteristic (ROC) curves for a single or multiple biomarkers using well-established methods. It also allows users to manually specify biomarker models and perform new sample prediction.

Pathway Analysis (targeted)

This module supports pathway analysis (integrating enrichment analysis and pathway topology analysis) and visualization for 26 model organisms, including Human, Mouse, Rat, Cow, Chicken, Zebrafish, *Arabidopsis thaliana*, Rice, Drosophila, Malaria, *S. cerevisae*, *E.coli*, and others species.

Spectral Analysis

This module allows users to upload raw LC-MS spectra (mzML, mzXML or mzData) to be processed using our optimized workflow based on MetaboAnalystR - OptiLCMS. The module supports common LC-MS platforms. The result peak intensity table can be used for statistical and functional analysis.

Functional Analysis (MS Peaks)

This module accepts high-resolution LC-MS spectral peak data to perform metabolic pathway enrichment analysis and visual exploration based on the well-established **mummichog** algorithm. It currently supports 26 organisms including Human, Mouse, Zebrafish, *C. elegans*, and other species.

Functional Meta-analysis (MS peaks)

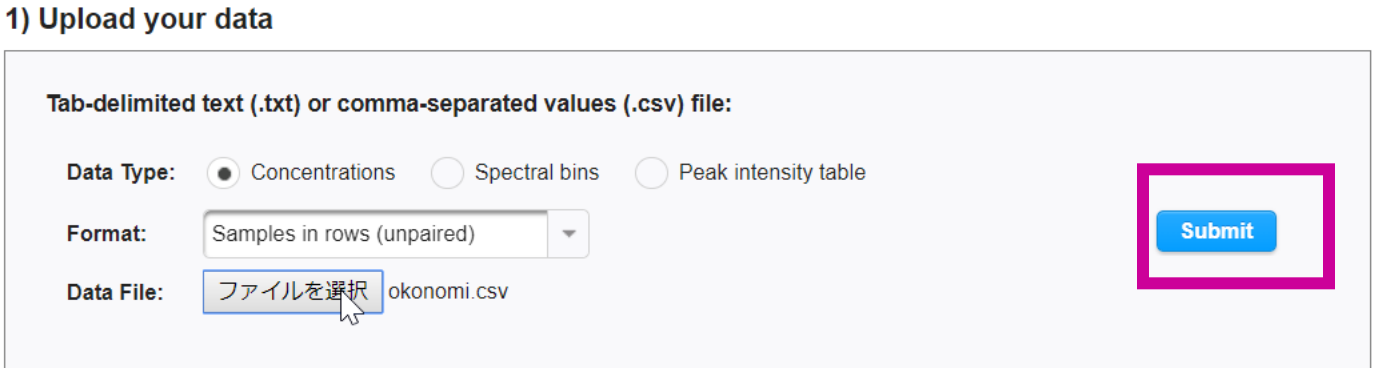
This module aims to identify robust **functional profiles** across multiple global metabolomics datasets via two approaches: 1) integrating functional profiles from independent studies conducted under compatible LC-MS conditions; or 2) pooling peaks from complementary instruments within the same studies.

Time-series/Two-factor Analysis　Enrichment Analysis　Joint-Pathway Analysis

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



自分のデータがなくても、デモデータが用意されているので、それを選んで試すのもおすすめ。

Try our test data

	Data Type	Description
●	Concentrations	Metabolite concentrations of 77 urine samples from cancer patients measured by 1H NMR (Eisner R, et al.). Group 1- cachexic; group 2 - control
○	Concentrations	Metabolite concentrations of 39 rumen samples measured by proton NMR from dairy cows fed with different proportions of barley grain (Ametaj BN, et al.). Group label - 0, 15, 30, or 45 - indicating the percentage of grain in diet.
○	NMR spectral bins	Binned 1H NMR spectra of 50 urine samples using 0.04 ppm constant width (Psihogios NG, et al.) Group 1- control; group 2 - severe kidney disease.
○	NMR peak lists	Peak lists and intensity files for 50 urine samples measured by 1H NMR (Psihogios NG, et al.). Group 1- control; group 2 - severe kidney disease.
○	Concentrations (paired)	Compound concentrations of 14 urine samples collected from 7 cows at two time points using 1H NMR (unpublished data). Group 1- day 1, group 2- day 4.
○	MS peak intensities	LC-MS peak intensity table for 12 mice spinal cord samples (Saghatelian et al.). Group 1- wild-type; group 2 - knock-out.
○	MS peak lists	Three-column LC-MS peak list files for 12 mice spinal cord samples (Saghatelian et al.). Group 1- wild-type; group 2 - knock-out.
○	LC-MS mzTab	LC-MS mzTab file of 15 mouse liver samples collected using LTQ Orbitrap Velos by (Hartler et al.) Group 1 - mouse liver 1; group 2 - mouse liver 2; group 3 - mouse liver 3.
○	GC-MS mzTab	GC-MS mzTab file of 6 *Arabidopsis* samples obtained using (MS-DIAL). Group 1 - cont; group 2 - MeKo.

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:

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.

Samples are in rows and features in columns

The uploaded file is in comma separated values (.csv) format.

The uploaded data file contains 33 (samples) by 29 (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.

A total of 0 (0%) missing values were detected.

By default, missing values will be replaced by 1/5 of min positive values of their corresponding variables

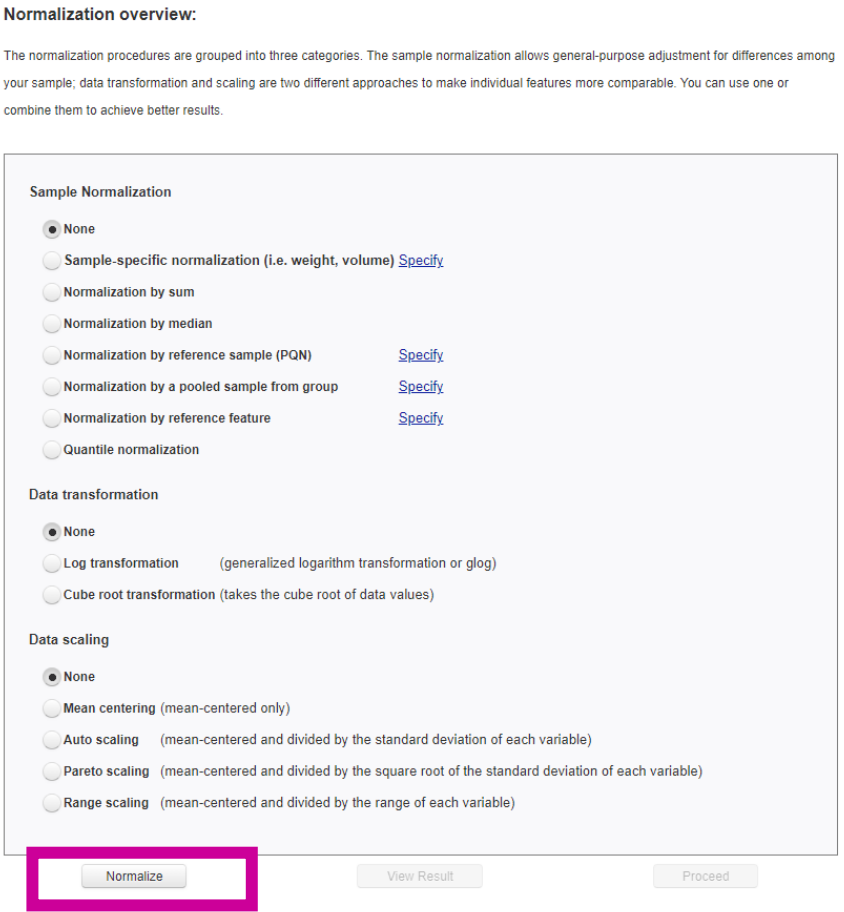
Click the **Skip** button if you accept the default practice;

Or click the **Missing value imputation** to use other methods.

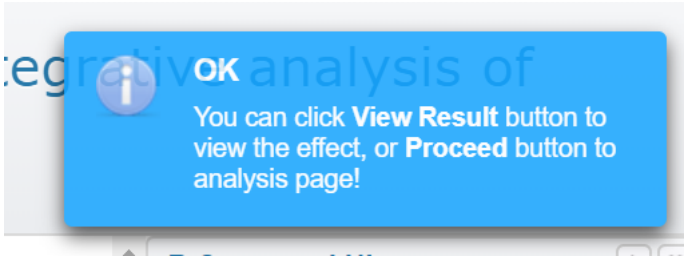
Edit Groups | Missing Values | → Proceed

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

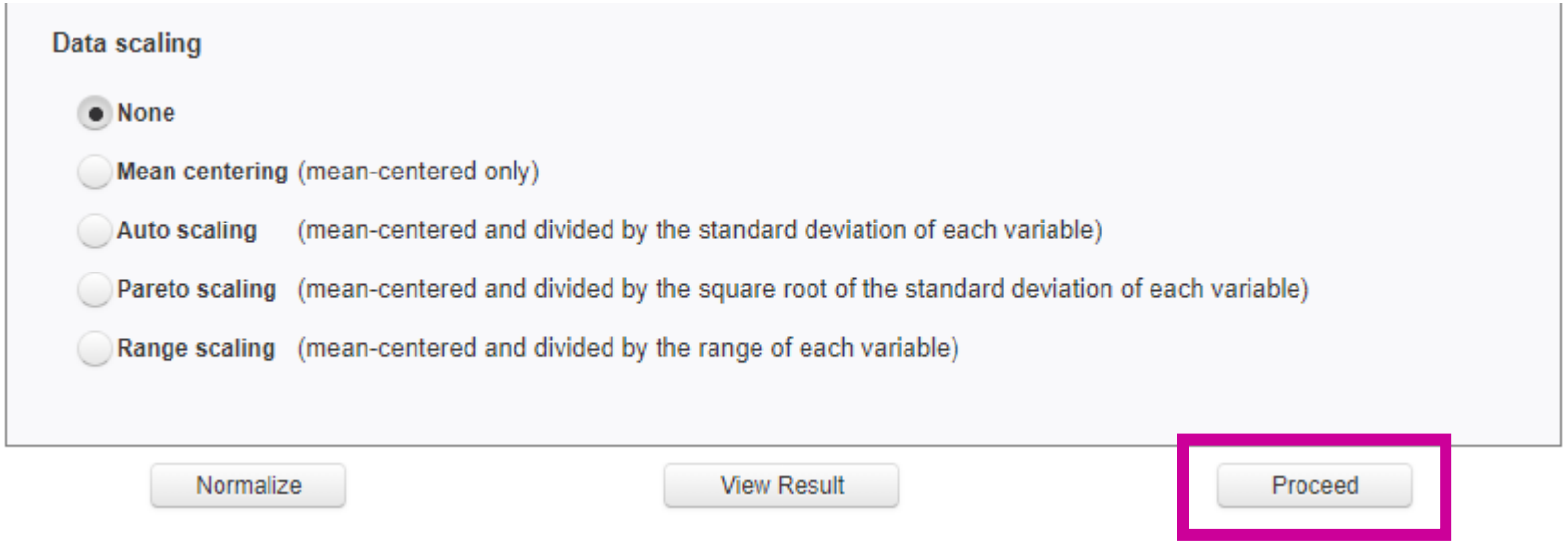
Normalizeボタンを押す



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)

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

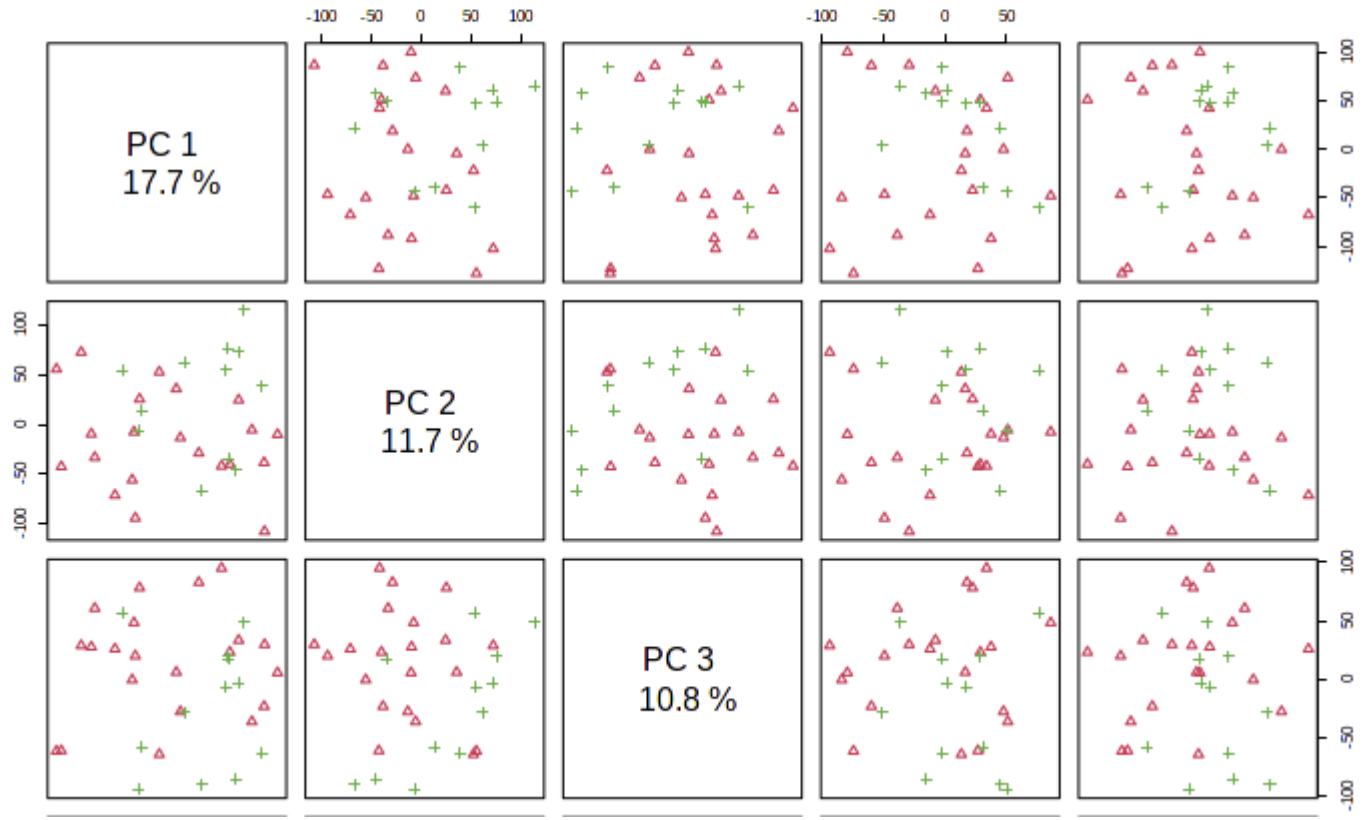
Hierarchical Clustering: Dendrogram Heatmaps

結果画面が表示される

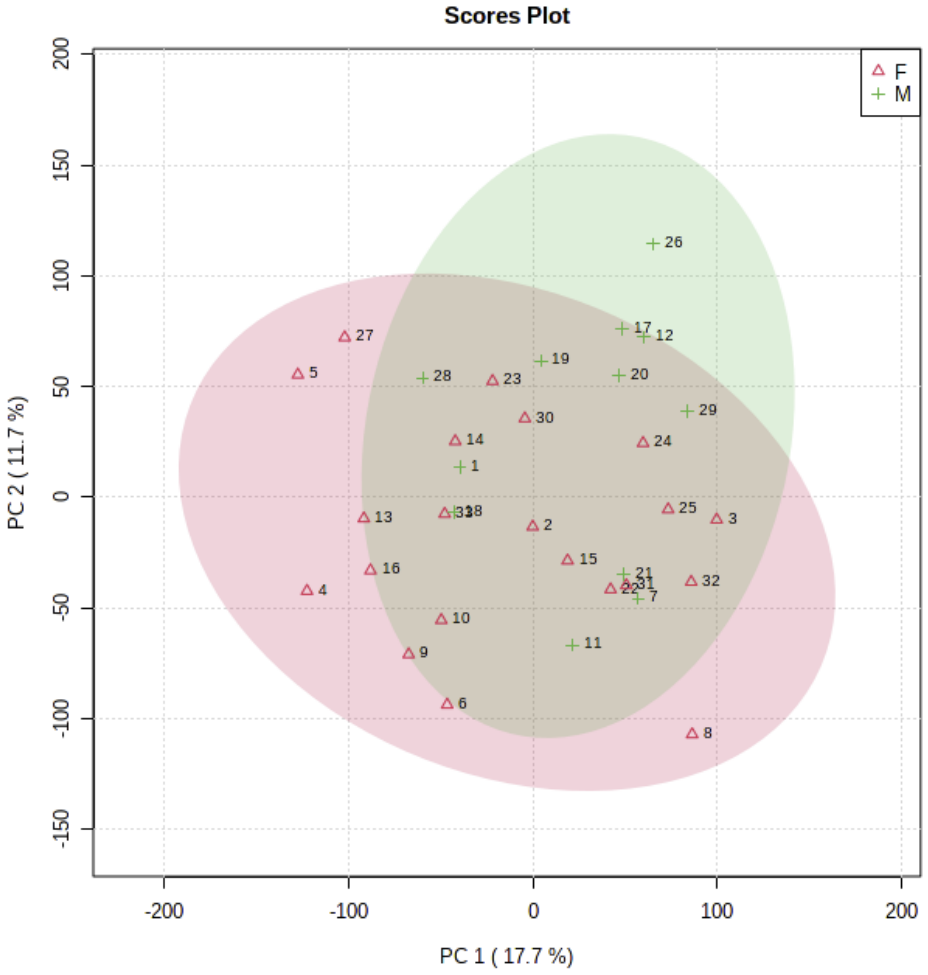
Principal Component Analysis (PCA)

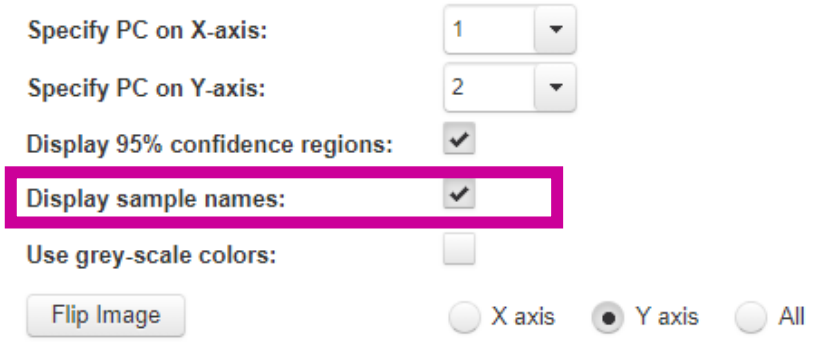
Overview | Scree Plot | 2D Scores Plot | Loadings Plot | Synchronized 3D Plots | Biplot

Display pairwise score plot for top 5 PCs Update



2D Scores Plot

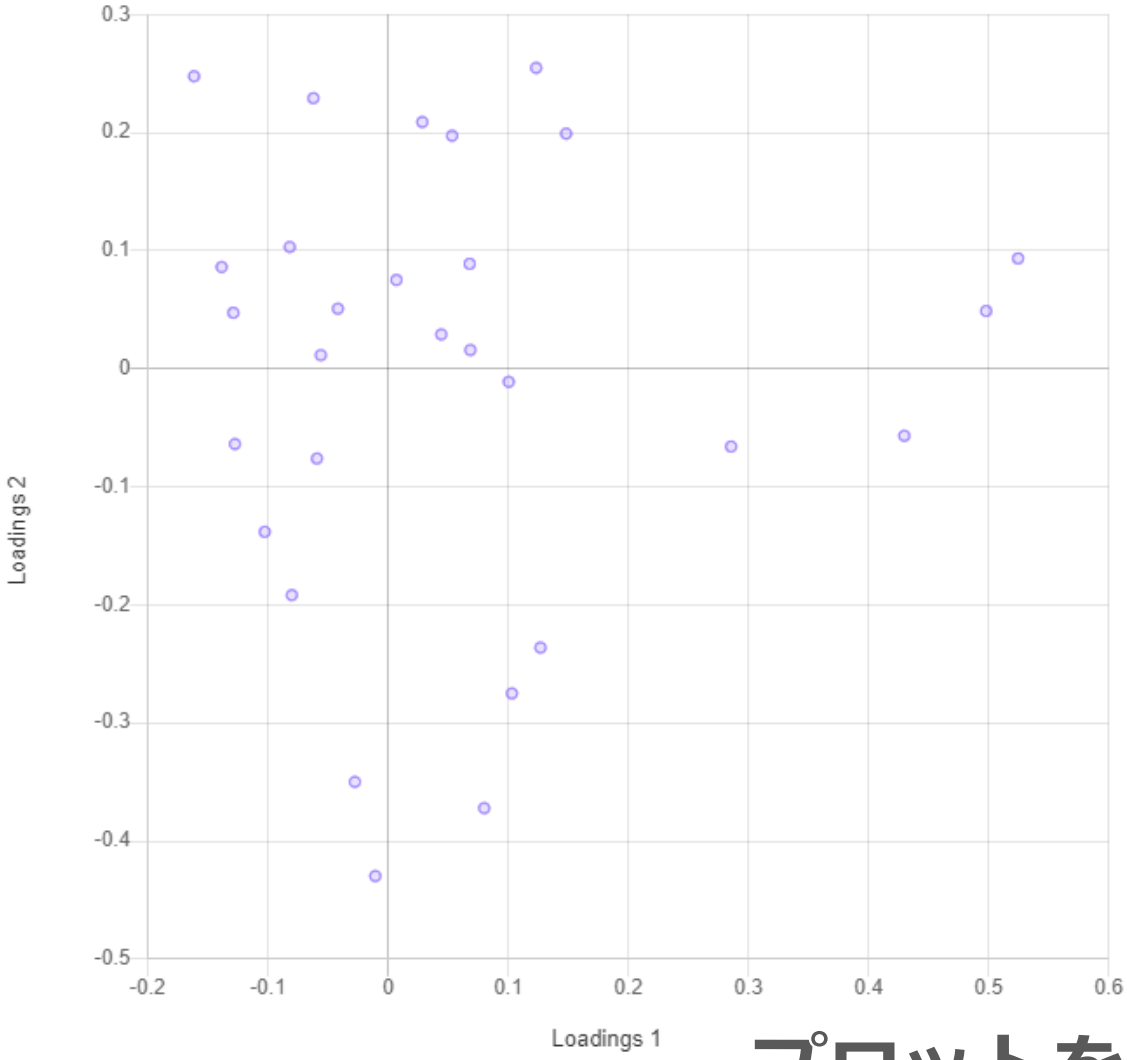


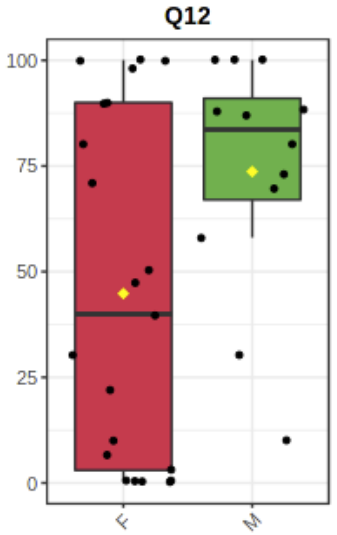


Display sample nameにチェックが入っていると、サンプル名が一緒に表示される。

※条件を変えた後に「Update」ボタンを押して反映させます。

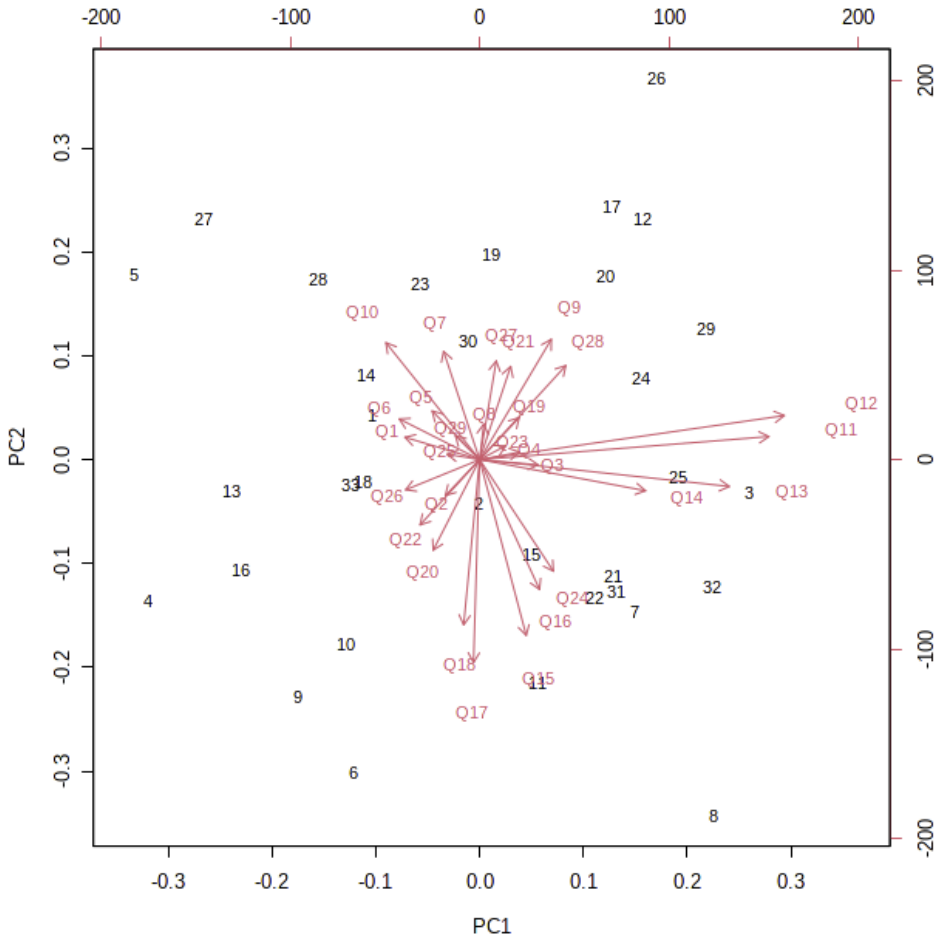
Loadings Plot





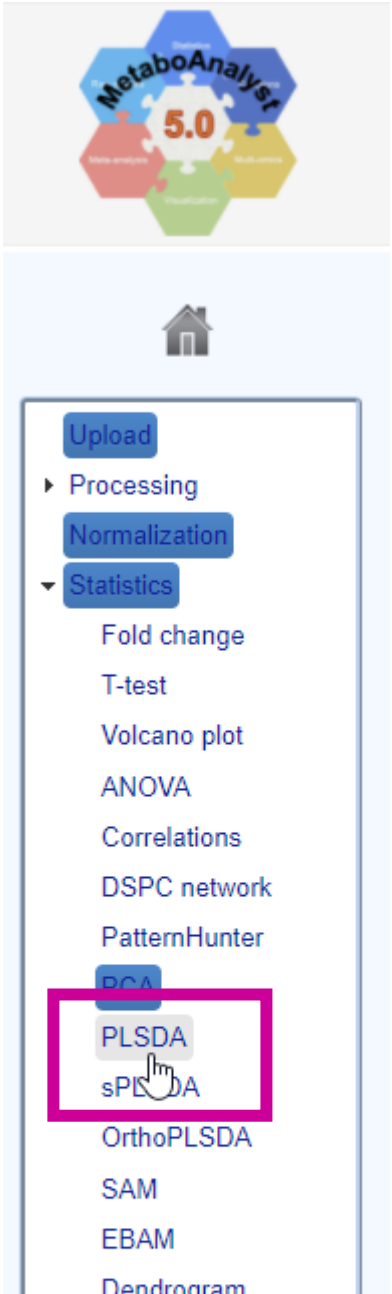
プロットをクリックすると、
その変数の特徴が表示される。

Biplot

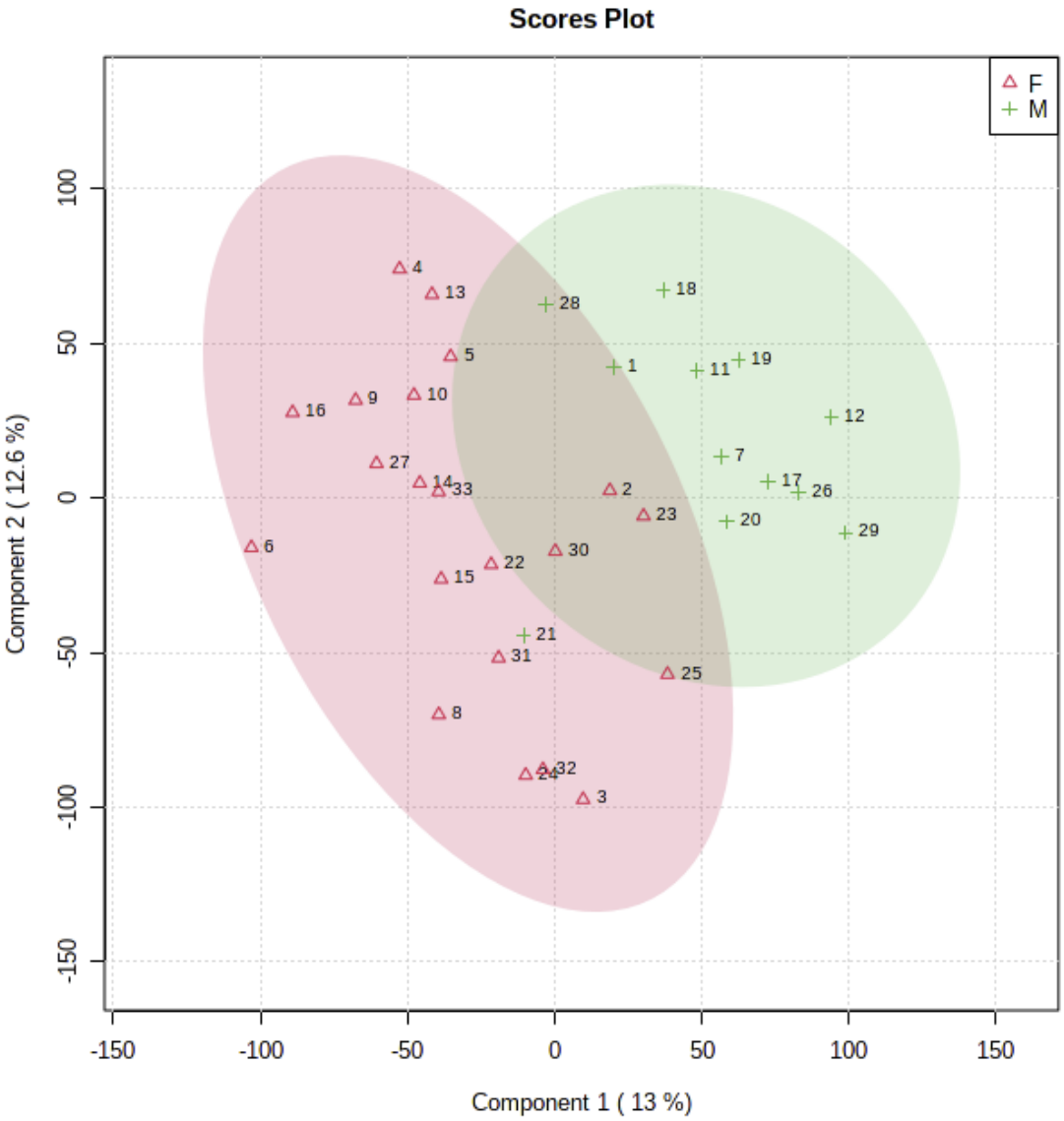


スコアプロットとローディングプロットが一緒に描かれたもの。

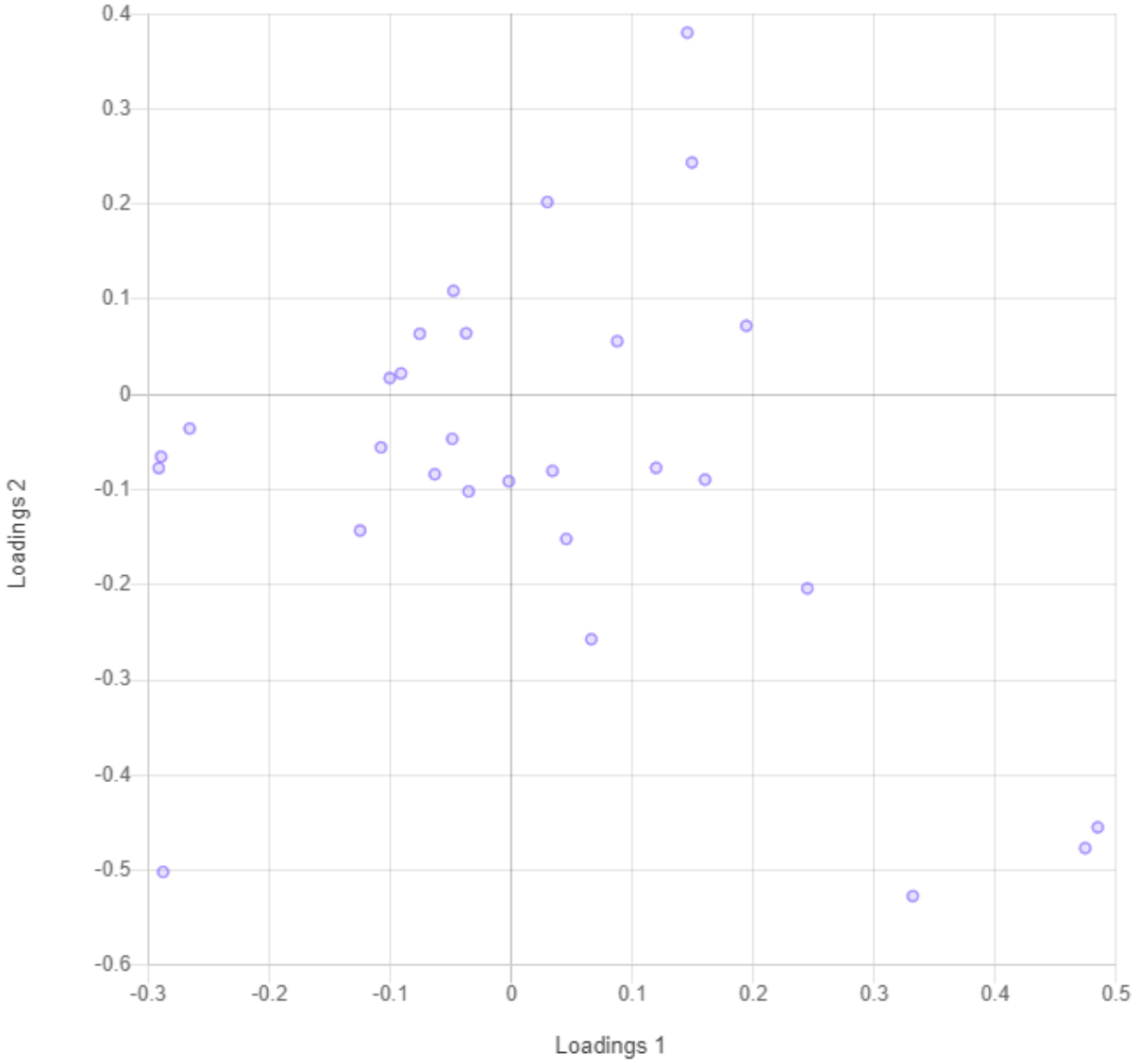
PLSDA

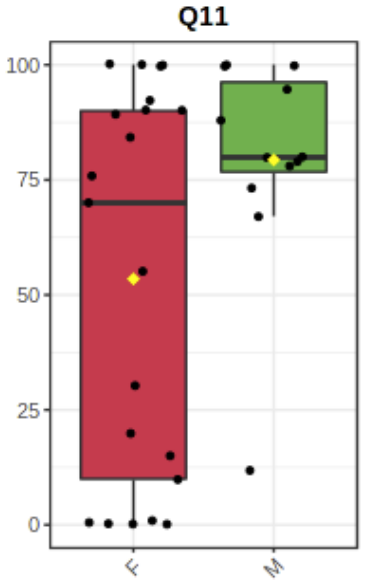

MetaboAnalyst
5.0
Upload
Processing
Normalization
Statistics
Fold change
T-test
Volcano plot
ANOVA
Correlations
DSPC network
PatternHunter
PCA
PLSDA
sPLSDA
OrthoPLSDA
SAM
EBAM
Dendrogram

2D Scores Plot

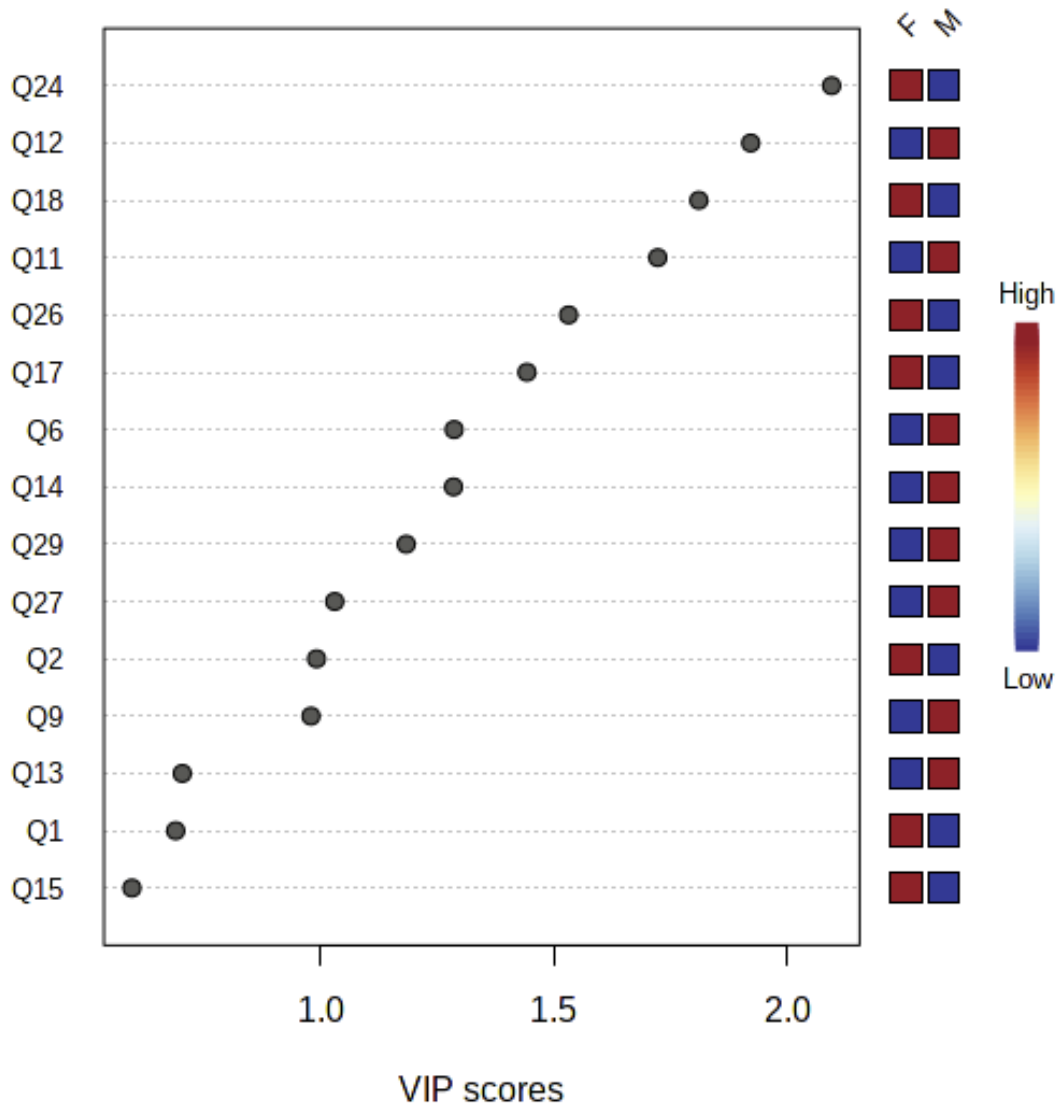


Loadings Plot





VIP値


F
M
Q24
Q12
Q18
Q11
Q26
Q17
Q6
Q14
Q29
Q27
Q2
Q9
Q13
Q1
Q15
High
Low
1.0
1.5
2.0
VIP scores

累積寄与率と構築したモデルの精度(Q)