

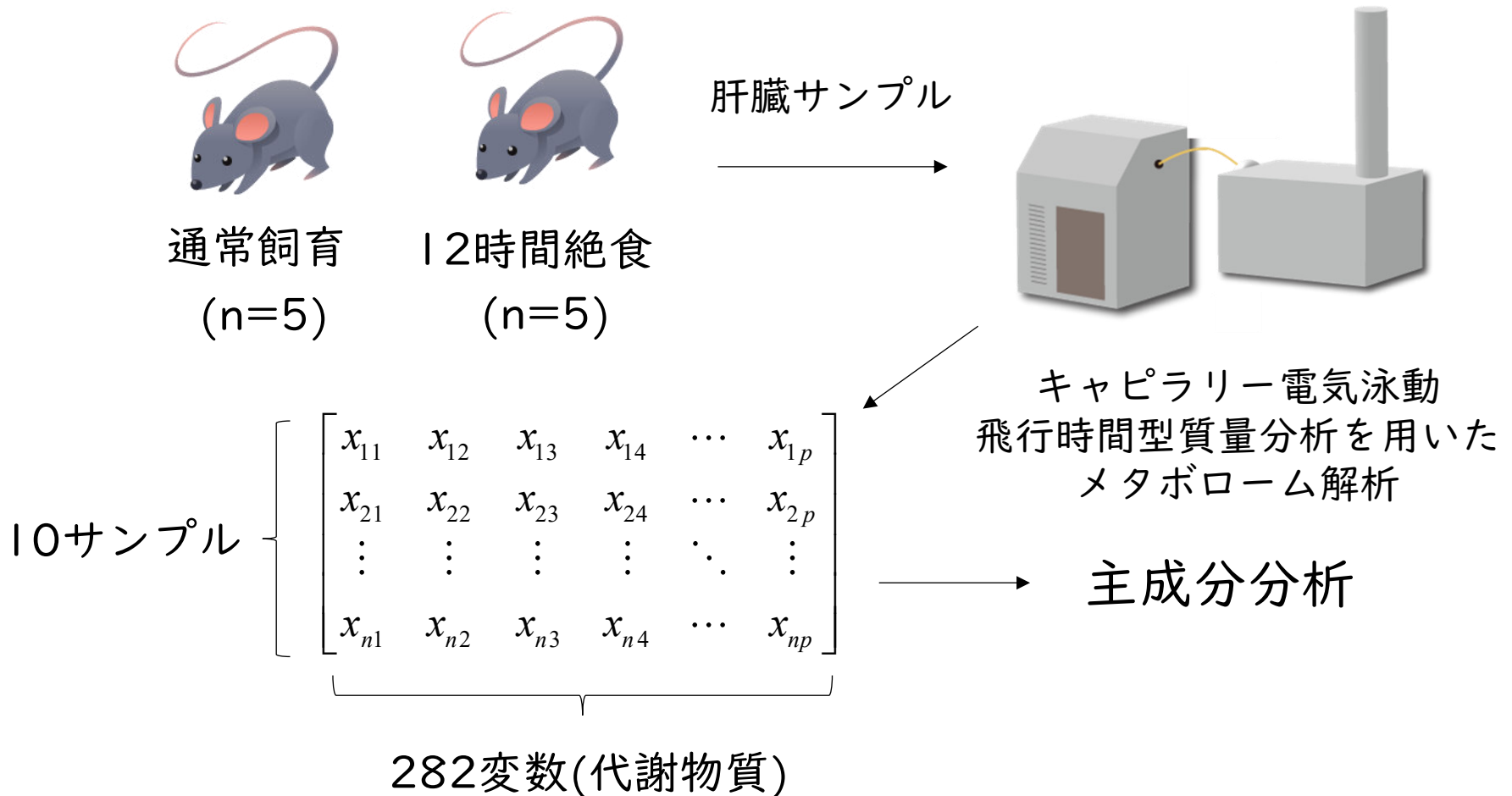
# 応用事例にみる多変量解析

# メタボロミクス

- メタボロミクス、メタボローム解析とは
  - メタボロミクスとは、アミノ酸、有機酸、脂肪酸、糖など、分子量が約1000以下の低分子の代謝物(メタボライト)を網羅的に解析すること
- サンプル
  - 動物の臓器、微生物の培養サンプル、人の血液、尿など様々な種類のサンプルを、質量分析装置で測定を行い、メタボロームデータを取得する
- 応用範囲
  - 薬剤の作用機序の解明や、疾患バイオマーカーの探索、食品の機能性研究など様々な研究に利用されている
- 多変量解析の利用
  - 主成分分析やPLSを用いたデータの可視化とローディングを用いて重要な代謝物を選ぶ

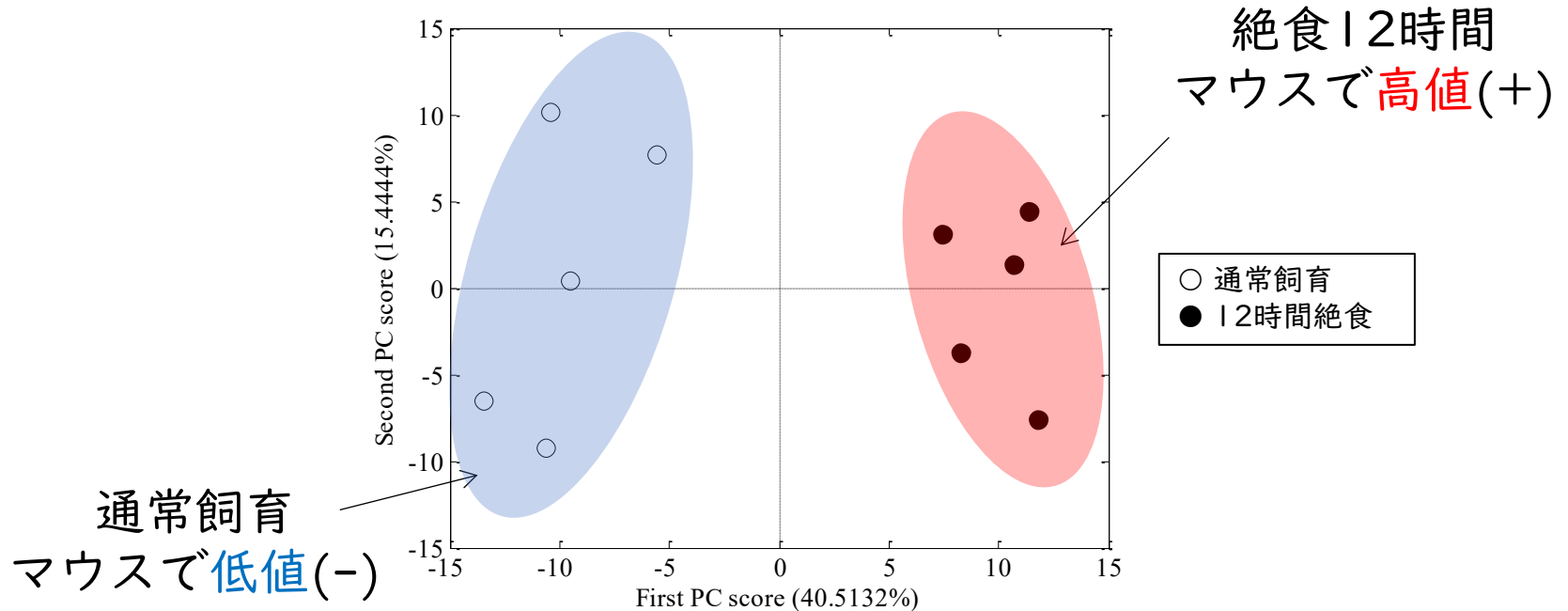
# メタボロミクスの研究例(1)

## 絶食マウス肝臓のメタボロームデータ



Yamamoto et al., "Statistical hypothesis testing of factor loading in principal component analysis and its application to metabolite set enrichment analysis" BMC Bioinformatics, (2014) 15(1):51.

# 主成分分析を用いたデータの可視化の例



身長・体重での合成変数「体の大きさ」と同様に合成変数を計算

↓  
実際のデータでは変数の組み合わせが多く、身長と体重を体の大きさに代表するというように、直感的に考えることは難しい

データを可視化した後に着目する主成分スコア(合成変数)を決める

# 重み・主成分係数を用いた代謝物の選び方

メタボロームデータ

10サンプル

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} & \cdots & x_{1p} \\ x_{21} & x_{22} & x_{23} & x_{24} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & x_{n3} & x_{n4} & \cdots & x_{np} \end{bmatrix}$$

282変数(代謝物質)

第1主成分

10サンプル

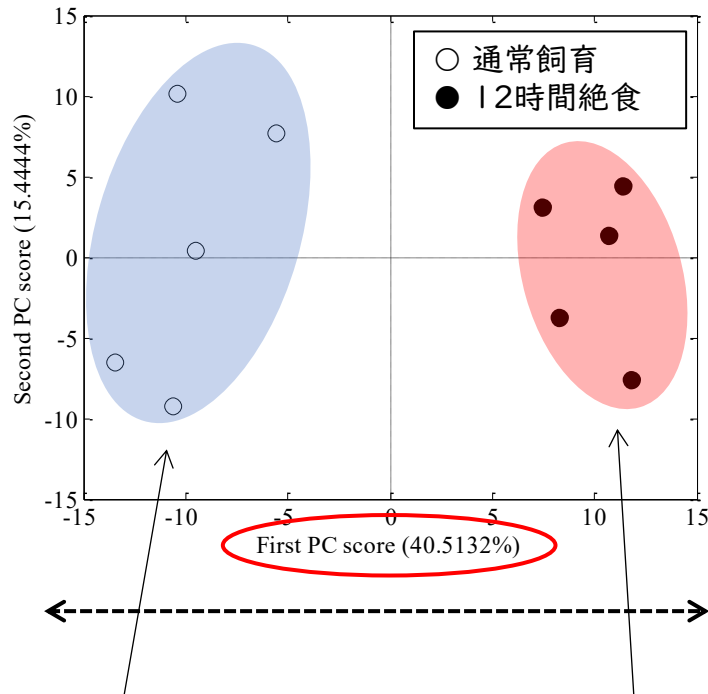
$$\begin{bmatrix} t_{11} \\ t_{21} \\ \vdots \\ t_{n1} \end{bmatrix}$$

$$\text{主成分スコア} = (\text{代謝物1}) \times \underline{w_1} + (\text{代謝物2}) \times \underline{w_2} + \cdots + (\text{代謝物P}) \times \underline{w_p}$$

$$t = x_1 w_1 + x_2 w_2 + \cdots + x_p w_p$$

# 絶食マウスでの主成分分析の解析例

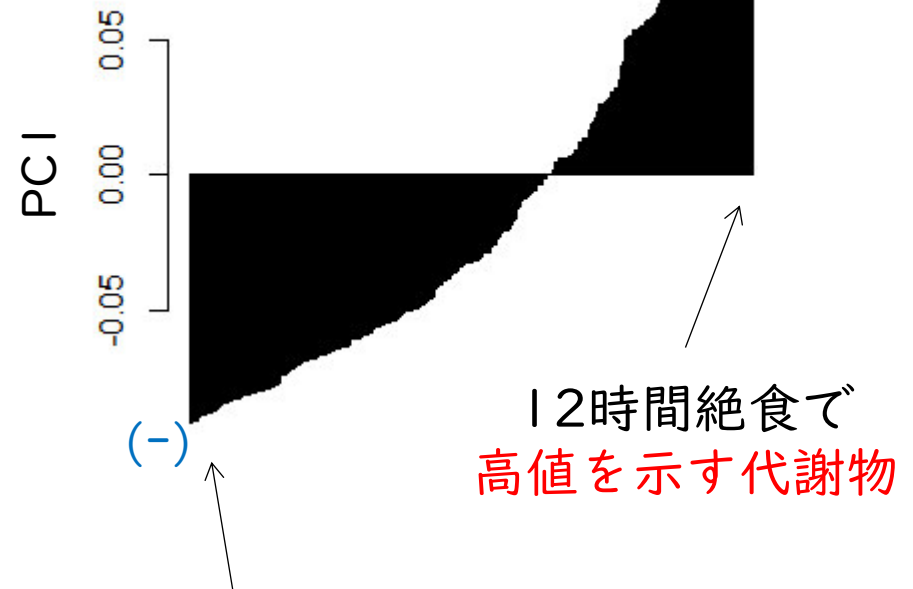
## データの可視化



通常飼育  
マウスで低値(-)

絶食12時間  
マウスで高値(+)

## 重み(主成分)係数を用いて 変数を選ぶ (+)



12時間絶食で  
低値を示す代謝物

合成変数に対する主成分係数 $w$ から変数を選ぶ

⇒ 実際は、主成分負荷量を用いて重要な変数を選ぶ (第2部後半で紹介)

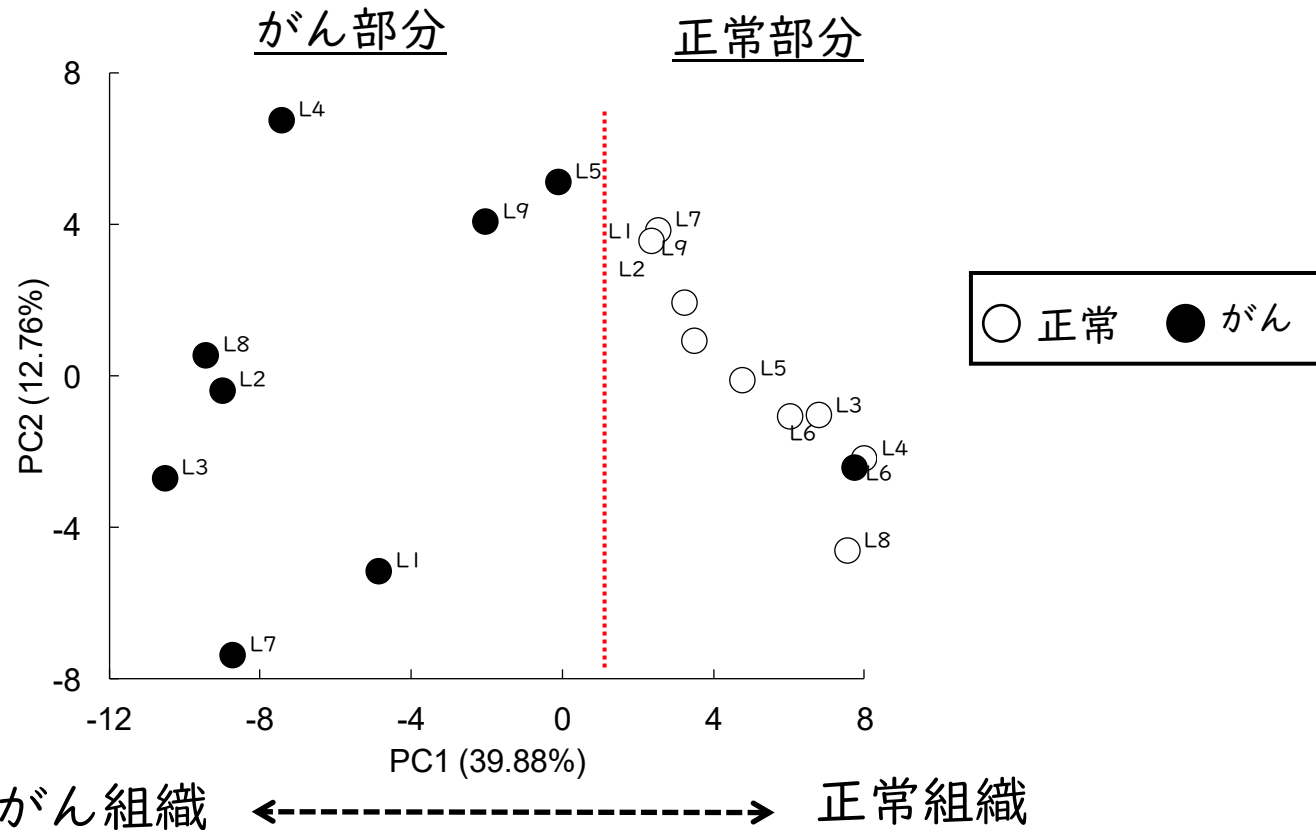
# メタボロミクスの研究例(2)

## メタボロミクスを用いたがん代謝機構の解明

- 試験内容
  - 肺がん患者の組織のメタボローム解析
    - 同一患者(N=9)の癌部分、正常部分
  - 前立腺がん患者の組織のメタボローム解析
    - 同一患者(N=7)の癌部分、正常部分

Kami K. et al., "Metabolomic profiling of lung and prostate tumor tissues by capillary electrophoresis time-of-flight mass spectrometry" Metabolomics, 9(2) (2013) 444-453.

# データの可視化(肺がんデータ)：簡単な例



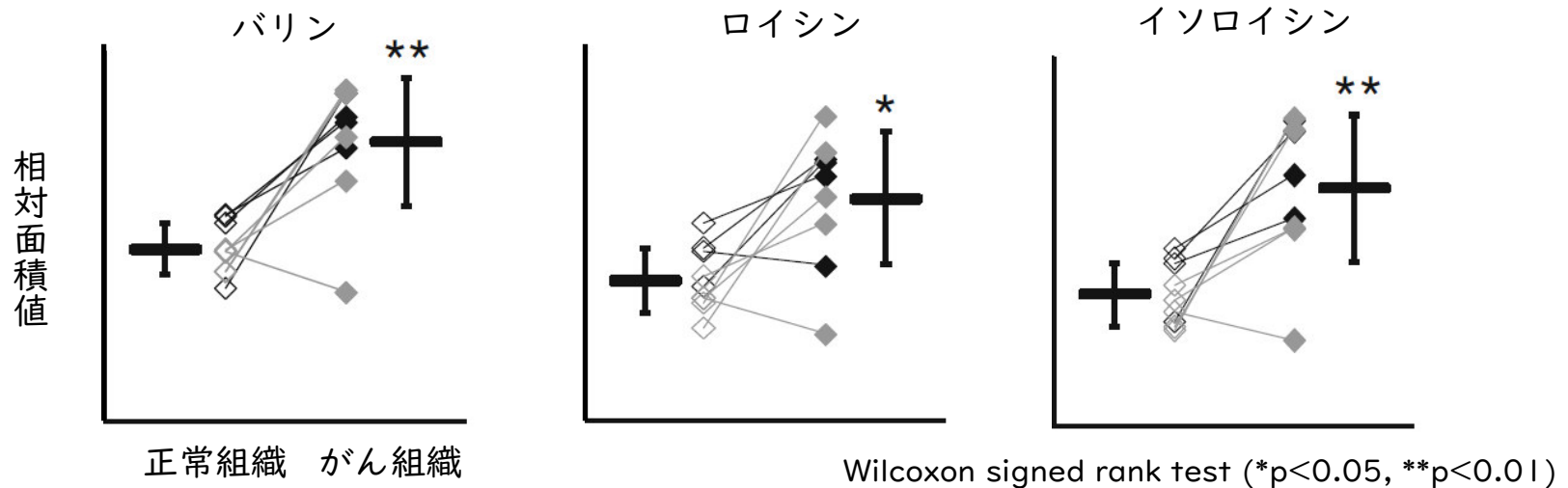
横軸はがん部分と正常部分の違いを表している

サンプルの様子は主成分スコアプロットを用いて  
可視化することができる



# 代謝物質の選択(肺がんデータ)

- 分岐鎖アミノ酸に着目



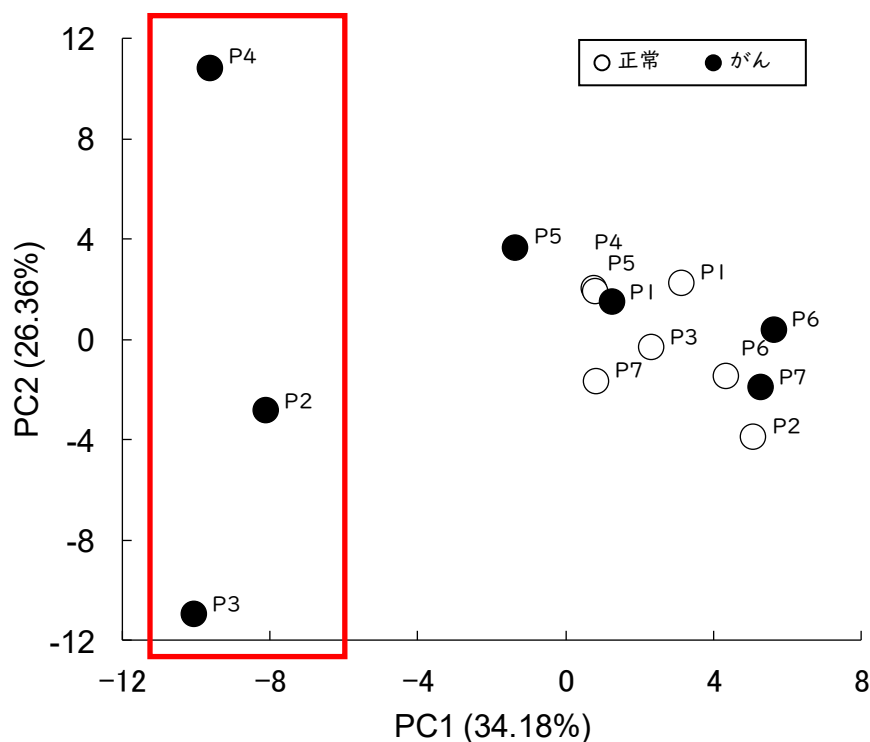
分岐鎖アミノ酸は、がん組織によって吸収され、がん患者でほとんど酸化されることが知られている (Baracos and Mackenzie, 2006)

着目する代謝物を選び出すために、

主成分負荷量が利用できる

PC1(がん部分で低く、正常部分で高い)とバリン ( $R = -0.97$ )、ロイシン ( $R = -0.89$ )、イソロイシン ( $R = -0.97$ )は有意な負の相関を示す

# データの可視化(前立腺がん)：難しい例



サンプル名	分化型
P1	中分化
P2	低分化
P3	低分化
P4	低分化
P5	中分化
P6	中分化
P7	中分化

Kami & Fujimori+(2013)  
より改変

低分化型がん組織

中分化型がん組織、正常組織

PC1: 分化型

さらにPC1の因子負荷量を確認し、低分化型  
と中分化型・正常組織と関連する代謝物を探すことが可能

# メタボロミクスの研究例(3)

## 高脂血症モデル動物に対する薬剤投与の影響をメタボローム解析により確認する

- 試験内容
  - 通常のウサギ、高脂血症のウサギ、コレステロールを低下させる薬剤を投与した高脂血症のウサギの3つの群
  - それぞれ3匹ずつの肝臓サンプル

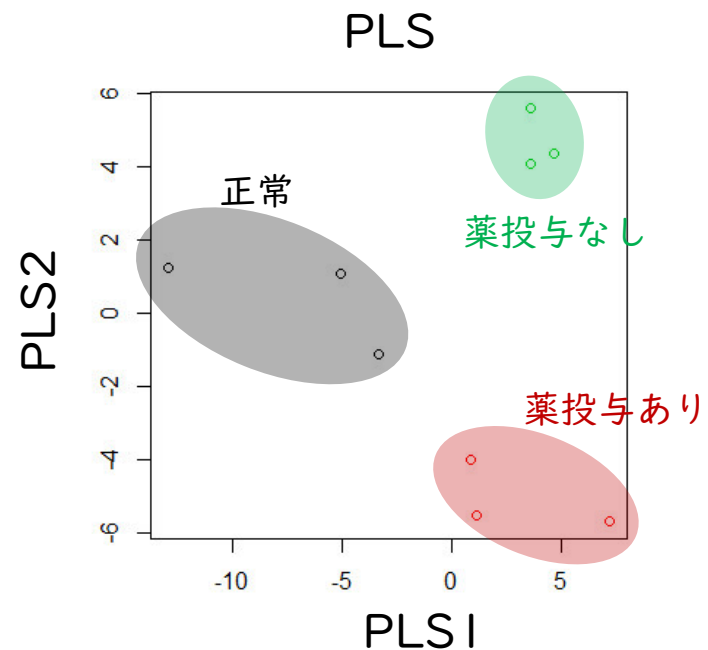
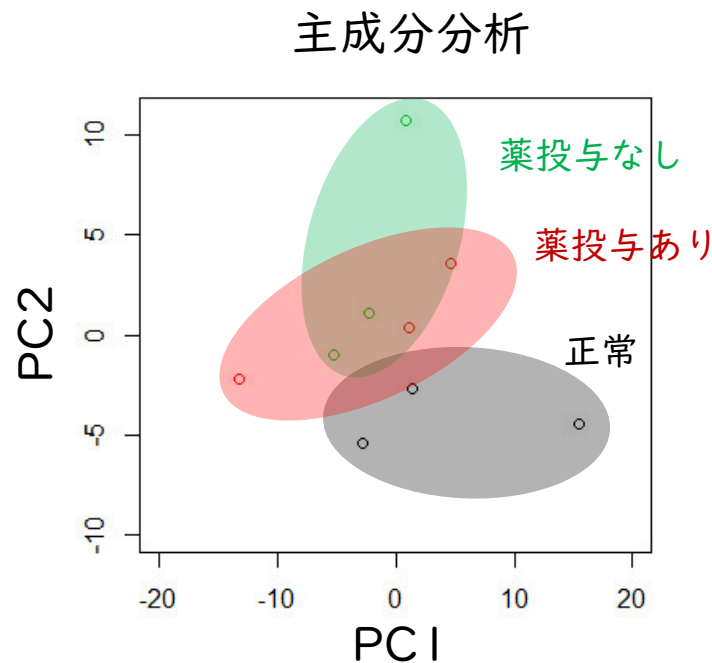
Ooga T, Sato H, Nagashima A, Sasaki K, Tomita M, Soga T, Ohashi Y., "Metabolomic anatomy of an animal model revealing homeostatic imbalances in dyslipidaemia." , Mol Biosyst. 2011 Apr;7(4):1217-23.

Yamamoto H., "PLS-ROG: Partial least squares with rank order of groups.", Journal of Chemometrics, 31(3) (2017) e2883.

# 主成分分析とPLSの解析例

高脂血症ウサギの肝臓のメタボローム解析

3群比較：Wild type、**高脂血症ウサギ**、**薬剤投与後の高脂血症ウサギ**



主成分分析の結果、主成分スコアで群間の差が表れなかったとき、  
PLSが用いられることが多い  
(2日目にご説明)

# 近赤外分光法の応用分野

- 食品

- 小麦粉、スターチ、食用油、食肉等の材料系から、乳製品等の加工食品系の成分分析に用いられており、日本酒、ワイン、醤油などの液体の測定にも多く用いられている。

- 農業

- お茶の成分測定(窒素、タンニン、水分など)や、野菜の硝酸イオン濃度、ミカンなどの糖度評価／選別に用いられる。

- 医薬品

- 原材料の受け入れ検査や工程管理(混合均一性の確認)に用いられることが多い。また、その他にも、結晶形や結晶化度のチェックに用いることが出来ることが知られている。

“<https://ja.wikipedia.org/wiki/近赤外線分光法>” より一部改変

# ケモトリックスの研究例

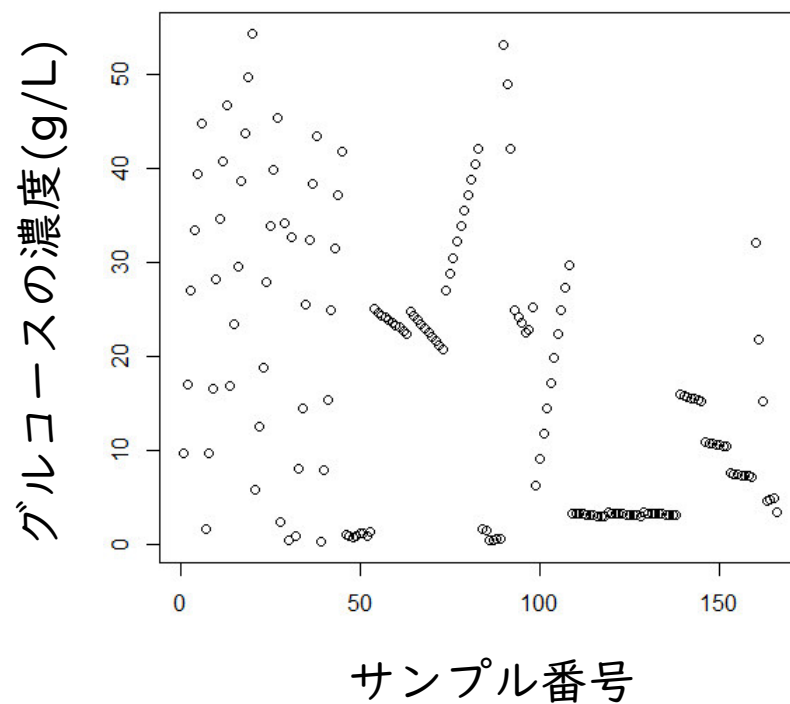
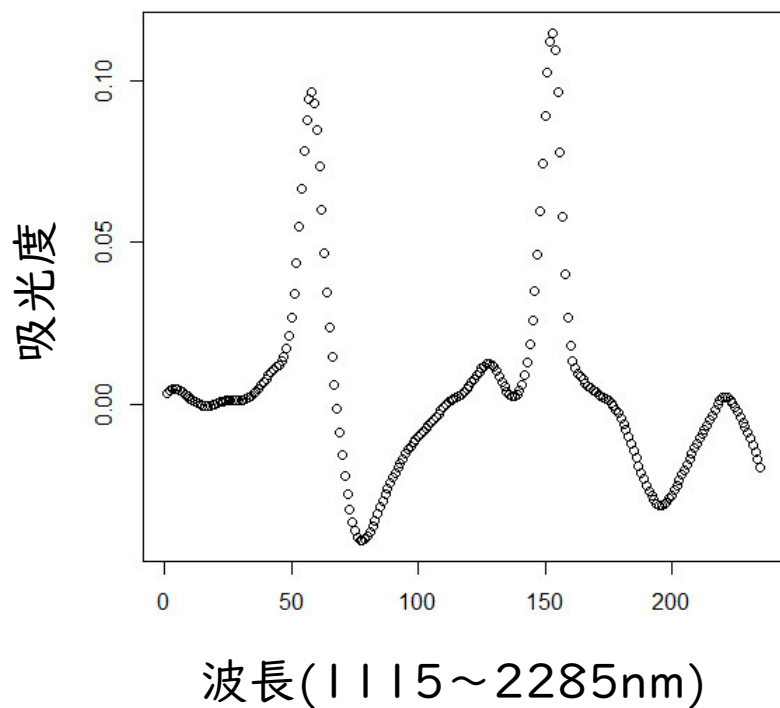
## • 試験内容の概要

- 酵母によるバイオエタノール(サトウキビなどのバイオマスから生成されるエタノール)生産において、グルコースとエタノール濃度を、近赤外スペクトル(NIRS)データから推定する
- グルコースは酵母の栄養源であり、発酵生成物がエタノールである
- 詳しくは、B.Liebmann, A. Friedl, K.Varmuza, “Determination of glucose and ethanol in bioethanol production by near infrared spectroscopy and chemometrics”, *Analytica Chimica Acta*, 642(1-2), pp 171-178(2009)を参照

## • データの説明

- 説明変数：ライ麦、小麦、トウモロコシを材料としてアルコール発酵を行い、波長領域が1115から2285nmを近赤外分光分析装置で測定し、得られた吸光度の値の1次微分をデータとしている
  - 166サンプル、 235変数のデータ
- 目的変数：グルコースとエタノールの濃度(g/L)を液体クロマトグラフィーで測定した得られたデータ

# NIRSの一例とグルコースの濃度

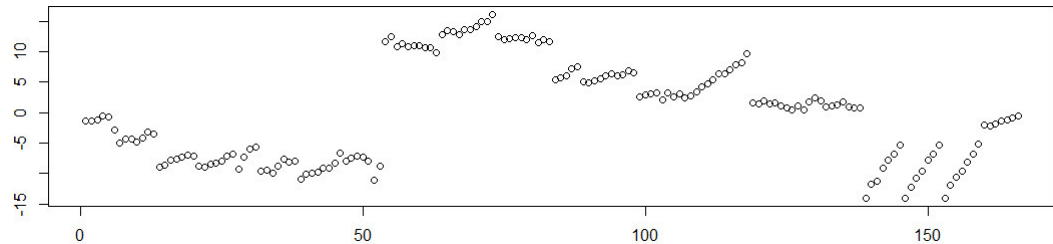


同様のスペクトルが166サンプル分ある

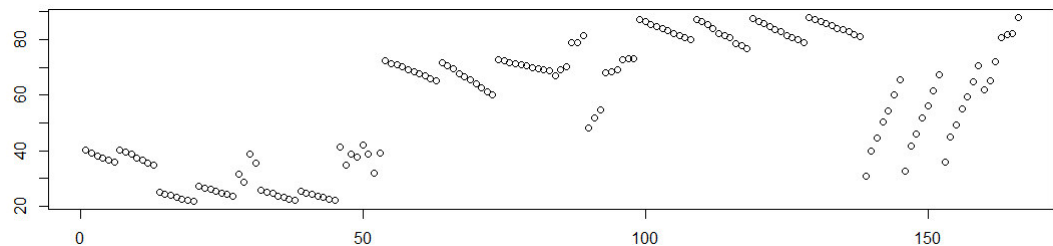
エタノールについても同様のデータ

# 第2主成分スコアとエタノールの濃度との相関

第2主成分スコア  
(正負逆)



エタノール濃度(g/L)



サンプル番号

グラフを見てわかる通り、第2主成分スコアとエタノール濃度の傾向は部分的に類似しており、その相関係数は0.683であった。  
一方、グルコース濃度と相関の高い主成分は見当たらなかった



論文では、**PLS回帰分析**を用いてグルコース、エタノールそれぞれの濃度を予測しているが、グルコースよりも**エタノール濃度の予測の方が精度が高い**



# MetaboAnalystを用いた解析

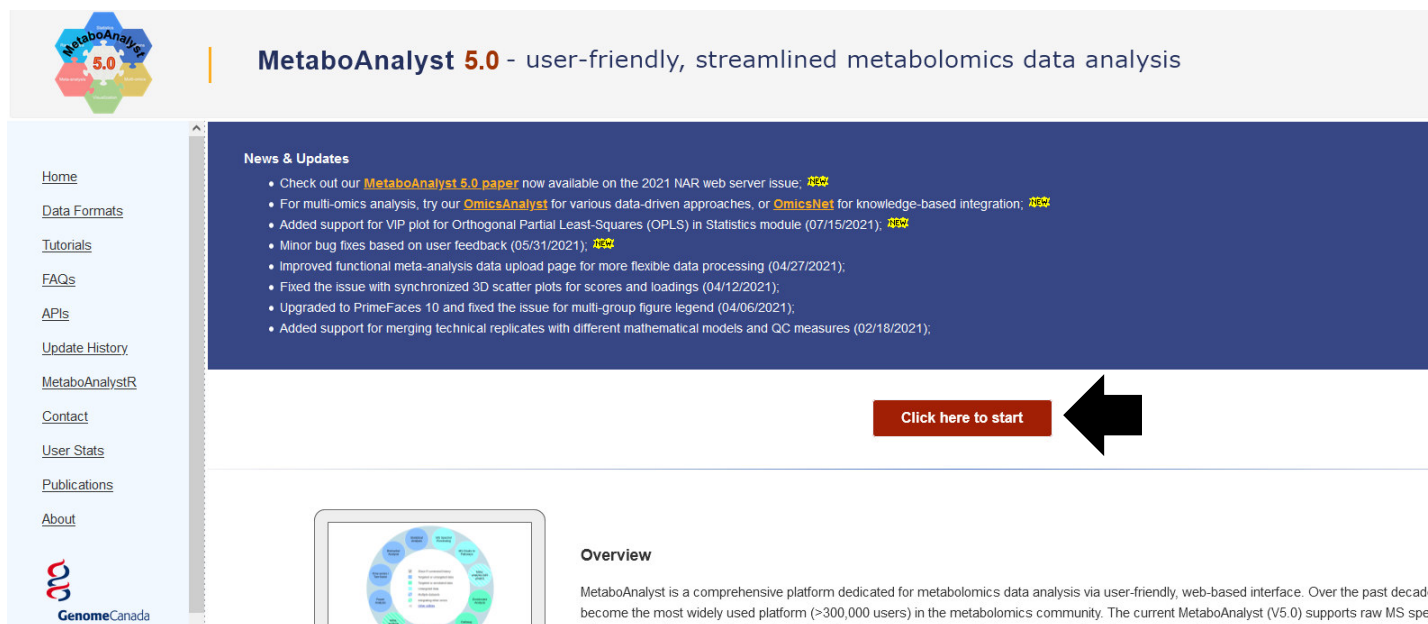
# MetaboAnalystを用いた解析

- MetaboAnalystとは

- カナダのアルバータ大学のDavid Wishartとマギル大学のJianguo Xiaの研究グループにより運営されているフリーのWebサービス





以下のwebサイトにアクセス

- <http://www.metaboanalyst.ca/> (最新版、2024.9時点では、6.0)
- <https://genap.metaboanalyst.ca/> (5.0)



**MetaboAnalyst 5.0** - user-friendly, streamlined metabolomics data analysis

**News & Updates**


- Check out our [MetaboAnalyst 5.0 paper](#) now available on the 2021 NAR web server issue; 
- For multi-omics analysis, try our [OmicsAnalyst](#) for various data-driven approaches, or [OmicsNet](#) for knowledge-based integration; 
- Added support for VIP plot for Orthogonal Partial Least-Squares (OPLS) in Statistics module (07/15/2021); 
- Minor bug fixes based on user feedback (05/31/2021); 
- Improved functional meta-analysis data upload page for more flexible data processing (04/27/2021);
- Fixed the issue with synchronized 3D scatter plots for scores and loadings (04/12/2021);
- Upgraded to PrimeFaces 10 and fixed the issue for multi-group figure legend (04/06/2021);
- Added support for merging technical replicates with different mathematical models and QC measures (02/18/2021);

[Click here to start](#)

**Overview**

MetaboAnalyst is a comprehensive platform dedicated for metabolomics data analysis via user-friendly, web-based interface. Over the past decade, it has become the most widely used platform (>300,000 users) in the metabolomics community. The current MetaboAnalyst (V5.0) supports raw MS spectra...

# Statistical Analysis [one factor]を選択



## MetaboAnalyst 5.0 - user-friendly, streamlined metabolomics data analysis

[Home](#)  
[Data Formats](#)  
[Tutorials](#)  
[User Forum](#)  
[MetaboAnalystR](#)  
[Publications](#)  
[Update History](#)  
[APIs](#)  
[User Stats](#)  
[About](#)  
[Contact](#)

### 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 [one factor]	Statistical Analysis [metadata table]	Biomarker Analysis	Statistical Meta-analysis	Power Analysis	Other Utilities

[Statistical Analysis \[one factor\]](#)

This module offers various commonly 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.


[Statistical Analysis \[metadata table\]](#)

This module aims to detect associations between phenotypes and metabolomics features with considerations of other experimental factors / covariates based on general linear models coupled with PCA and heatmaps for visualization. More options are available for two-factors / time-series data.

[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.

Xia Lab @ McGill (last updated 2023-2-17)



# 解析用のデータ(csvファイル)の準備

代謝物名

サンプル名

群名

データ行列

	A	B	C	D	E	F	G	H	I	J	K
1		L1-1	L1-2	L1-3	L1-4	L1-5	L2-1	L2-2	L2-3	L2-4	L2-5
2	Label	0	0	0	0	0	12	12	12	12	12
3	gamma-Glu-Cys	0.002131	0.002123	0.001154	0.001122	0.000764	0.000504	0.000456	0.000761	0.000369	0.000221
4	gamma-Glu-2-aminobutyric acid	2.67E-05	4.07E-05	3.74E-05	0	0	0.000069	6.92E-05	0.000104	4.58E-05	3.92E-05
5	gamma-Butyrobetaine	0.00225	0.002199	0.001928	0.001543	0.001431	0.001531	0.001099	0.001095	0.000933	0.00113
6	beta-Ala	0.006975	0.01274	0.009952	0.007622	0.006263	0.01778	0.013166	0.010415	0.012316	0.01408
7	Xanthosine	0.002037	0.001377	0.000529	0.000765	0.000931	0.000987	0.000585	0.000429	0.000506	0.000345
8	Xanthopterin	0.000117	9.78E-05	8.54E-05	8.05E-05	5.33E-05	0	3.01E-05	4.23E-05	3.52E-05	0
9	Xanthine	0.008337	0.005057	0.004272	0.00605	0.005468	0.009027	0.008239	0.005656	0.007852	0.007315
10	Val	0.028482	0.042974	0.039433	0.044225	0.044748	0.037922	0.034743	0.03678	0.026216	0.030289
11	UTP	2.37E-05	2.34E-05	3.74E-05	3.29E-05	5.52E-05	2.06E-05	1.47E-05	0	2.35E-05	2.01E-05
12	Urocanic acid	2.25E-05	2.37E-05	1.95E-05	2.26E-05	0	0.000182	3.55E-05	0.000029	4.88E-05	3.58E-05
13	Uridine	0.009288	0.010703	0.006812	0.00663	0.007044	0.00643	0.006485	0.006719	0.006114	0.005245
14	Uric acid	0.001833	0.003585	0.002197	0.002249	0.003107	0.002395	0.002417	0.002238	0.003512	0.002584
15	Urea	0.143363	0.209627	0.202604	0.156428	0.13072	0.142575	0.105483	0.112816	0.136813	0.131537
16	Uracil	0.001484	0.001704	0.001136	0.001084	0.001177	0.001129	0.001113	0.0012	0.001086	0.00101
17	UMP	0.006386	0.006573	0.00975	0.00683	0.009133	0.007092	0.007415	0.009989	0.00819	0.008266
18	UDP-N-acetylglucosamine	0.003038	0.003748	0.004966	0.004238	0.006038	0.005763	0.003471	0.003989	0.00326	0.002713
19	UDP-glucuronic acid	4.34E-05	3.48E-05	5.81E-05	3.81E-05	6.62E-05	5.22E-05	4.28E-05	5.41E-05	7.85E-05	7.92E-05
20	UDP-glucose ; UDP-galactose	0.003531	0.000879	0.00359	0.00721	0.009591	0.003216	0.00289	0.002377	0.002808	0.002134
21	UDP	0.000267	0.000275	0.0003	0.000279	0.000409	0.000284	0.000194	0.000204	0.000212	0.000212
22	Tyr-Glu	0.000102	6.74E-05	3.64E-05	4.48E-05	0.000051	4.56E-05	2.48E-05	0.000036	2.04E-05	2.18E-05
23	Tyramine	4.44E-05	2.25E-05	2.96E-05	3.15E-05	4.05E-05	2.56E-05	2.52E-05	2.59E-05	1.72E-05	0.000028
24	Tyr	0.011741	0.013964	0.010952	0.011192	0.011773	0.013534	0.011039	0.011486	0.007544	0.00891
25	Trp	0.003457	0.003584	0.003711	0.003724	0.003977	0.002886	0.00311	0.003057	0.00256	0.002811
26	Trimethylamine N-oxide	0.000685	0.000312	0.000431	0.000277	0.000219	3.31E-05	1.42E-05	3.31E-05	5.36E-05	6.22E-05
27	Trigonelline	0.000895	0.002158	0.001487	0.001449	0.000816	0.000281	0.000305	0.00036	0.000274	0.000425
28	Triethanolamine	9.04E-05	4.29E-05	5.58E-05	4.56E-05	4.51E-05	0.000035	4.45E-05	4.35E-05	6.15E-05	4.72E-05
29	Trehalose 6-phosphate	0.000055	5.74E-05	0.000071	0.000057	0.000044	4.65E-05	3.55E-05	2.74E-05	3.25E-05	4.12E-05
30	trans-Glutaconic acid ; Itaconic acid	0.000146	0.000221	0.000257	0.000196	0.00011	0.000132	0.000101	0.000132	0.000148	0.000193
31	threo-beta-Methylaspartic acid	0.000153	0.000162	0.000211	0.000144	0.000173	0.00017	0.000177	0.000117	0.00015	0.00019

input

# データの読み込み



**MetaboAnalyst 5.0** - user-friendly, streamlined metabolomics data analysis



Upload

Processing

Normalization

Statistics

Download

Exit

Please upload your data

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

Data Type: ☒ Concentrations ☐ Spectral bins ☐ Peak intensity table

Format:

Data File:  ファイルが選択されていません。

A mzTab 2.0-M file (.mzTab):

① ファイルを選択

☐ Theoretical

が選択されていません。

② Samples in columns (unpaired)  
を選択

A compressed file (.zip):

Data Type: ☒ NMR peak list ☐ MS peak list


Data File:  ファイルが選択されていません。

Pair File:  ファイルが選択されていません。


A dataset from Metabolomics Workbench: ?

Study ID:

# スケーリングの実行



MetaboAnalyst 5.0 - user-friendly, streamlined metabolomics data analysis



- Upload
- Processing
  - Data check
  - Missing value
  - Data filter
  - Data editor
  - Normalization
- Statistics
- Download
- Exit

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

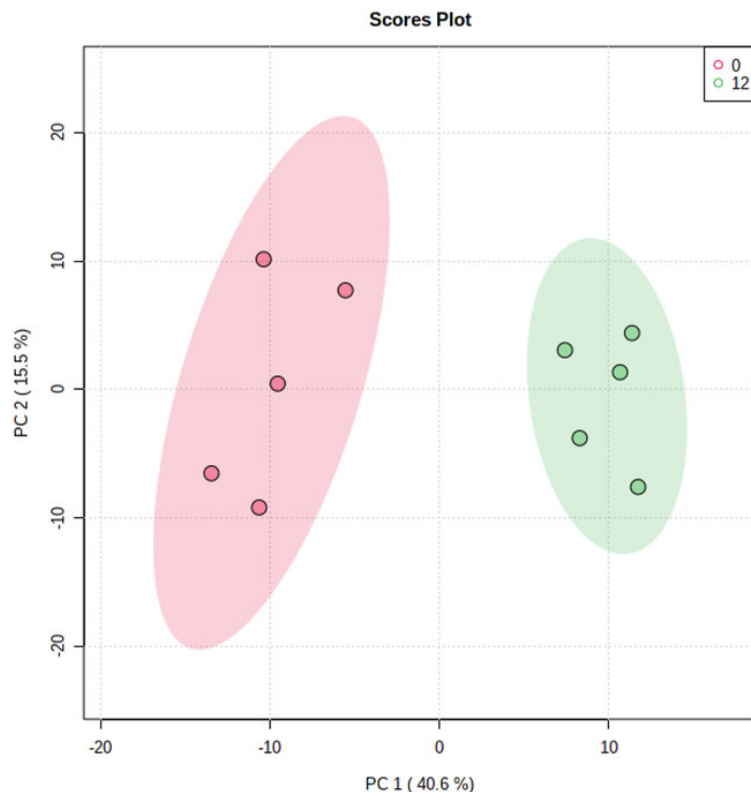
Xia Lab @ McGill (last updated 2021-02-11)

Auto scaling  
(推奨)

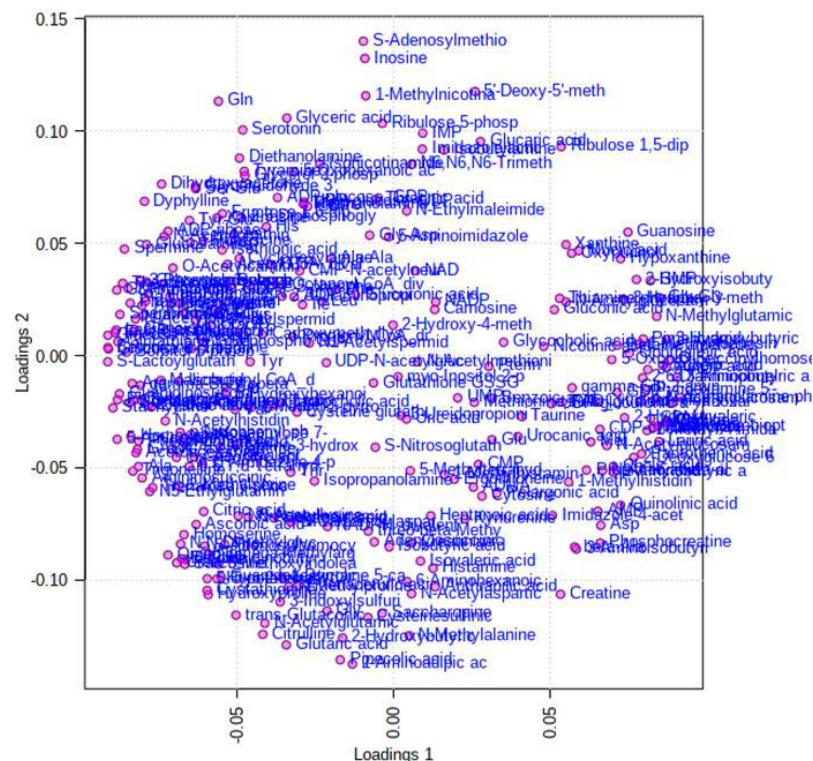


# 主成分分析の結果

## 主成分スコア



## ローディング(固有ベクトル)



上記の画像ファイル、主成分スコアとローディング(固有ベクトル)の計算結果は、それぞれダウンロードすることが出来る

# MetaboAnalystの実習(デモデータ)

The screenshot shows the MetaboAnalyst web interface. On the left is a sidebar with navigation links: Upload, Processing, Normalization, Statistics, Download, and Exit. The main area is titled 'Data File:' with a '+ Choose' button. Below this, it says 'A dataset from Metabolomics Workbench:' and 'Study ID:' with a text input field containing 'ST001301' and a 'Submit' button. The 'Try our test data' section contains a table with two columns: 'Data Type' and 'Description'. The 'MS peak intensities' option is selected with a radio button and is highlighted by a red rectangle. A red arrow points from the 'MS peak intensities' option in the table to the 'MS peak intensities' text in the sidebar. At the bottom of the table is a 'Submit' button. The footer of the page reads 'Xia Lab @ McGill (last updated 2023-2-17)'.

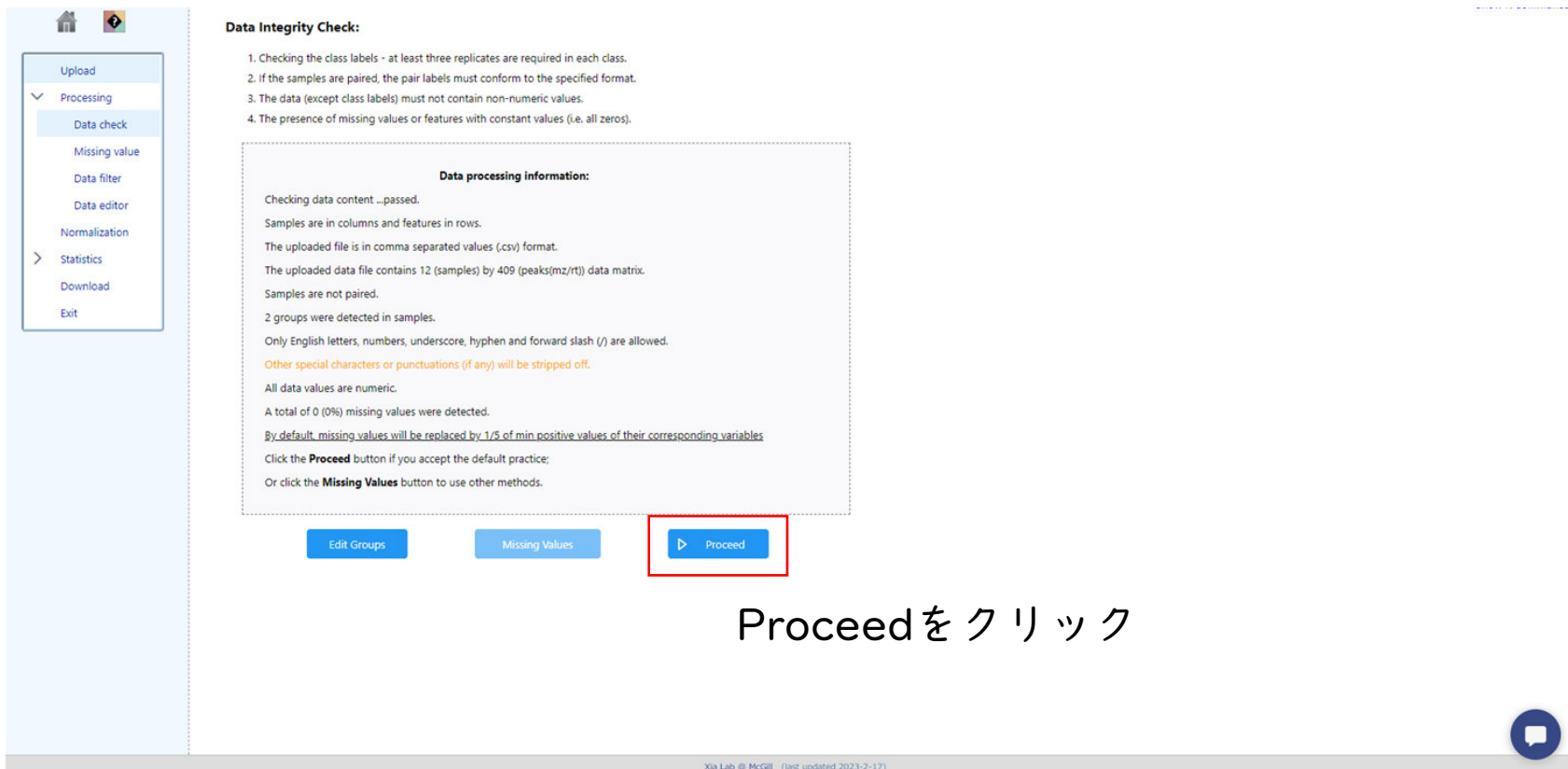
Data Type	Description
<input type="radio"/> Concentrations	Metabolite concentrations of 77 urine samples from cancer patients measured by 1H NMR (Eisner R. et al.). Group 1- cachexic; group 2 - control
<input type="radio"/> Concentrations	Metabolite concentrations of 39 rumen samples measured by proton 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.
<input type="radio"/> 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.
<input type="radio"/> 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.
<input type="radio"/> 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.
<input checked="" type="radio"/> 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.
<input type="radio"/> 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.
<input type="radio"/> 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.
<input type="radio"/> GC-MS mzTab	GC-MS mzTab file of 6 Arabidopsis samples obtained using (MS-DIAL). Group 1 - cont; group 2 - MeKo.

## 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.



# データの確認、欠損値の補完



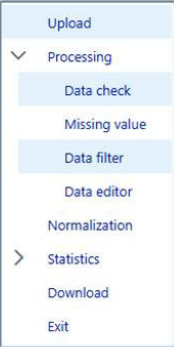
The screenshot shows a web-based interface for data processing. On the left is a sidebar with a menu: Upload, Processing (expanded), Data check (selected), Missing value, Data filter, Data editor, Normalization, Statistics, Download, and Exit. The main area is titled "Data Integrity Check:" and contains a list of four checks: 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).

Below this list is a box titled "Data processing information:" containing the following text: "Checking data content ...passed. Samples are in columns and features in rows. The uploaded file is in comma separated values (.csv) format. The uploaded data file contains 12 (samples) by 409 (peaks(mz/rt)) 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 **Proceed** button if you accept the default practice; Or click the **Missing Values** button to use other methods."

At the bottom of the main area are three buttons: "Edit Groups", "Missing Values", and "Proceed". The "Proceed" button is highlighted with a red rectangle. Below the buttons, the text "Proceedをクリック" is displayed.

Xia Lab @ McGill (last updated 2023-2-17)

# 変数の除外






**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.](#)

Non-informative variables can be characterized in three groups: 1) variables that show **low repeatability** - this can be measured using QC samples using the relative standard deviation ( $RSD = SD/mean$ ). Features with high percent RSD should be removed from the subsequent analysis (the suggested threshold is 20% for LC-MS and 30% for GC-MS); 2) variables that are **near-constant** throughout the experiment conditions - these variables can be detected using standard deviation (SD); or the robust estimate such as interquartile range (IQR); and 3) variables of **very small values** (close to baseline or detection limit) - these variables can be detected using mean or median.

For data filtering based on the last two categories, the default parameters follow the empirical rules: 1) Less than 250 variables: 5% will be filtered; 2) Between 250 - 500 variables: 10% will be filtered; 3) Between 500 - 1000 variables: 25% will be filtered; and 4) Over 1000 variables: 40% will be filtered. You can turn off data filtering by dragging the slider to adjust the percentage to filter out to be 0, when your data contain less than 5000 features (or 2500 for power analysis) to control computing time on our server.

<b>Reliability filter:</b>	<input type="checkbox"/> Filtering features based on technical repeatability QC samples	RSDs greater than:  25%
<b>Variance filter:</b>	<input checked="" type="radio"/> Interquartile range (IQR) <input type="radio"/> Standard deviation (SD) <input type="radio"/> Median absolute deviation (MAD) <input type="radio"/> Relative standard deviation ( $RSD = SD/mean$ ) <input type="radio"/> Non-parametric relative standard deviation ( $MAD/median$ )	Percentage to filter out:  0%
<b>Abundance filter:</b>	<input checked="" type="radio"/> Mean intensity value <input type="radio"/> Median intensity value	Percentage to filter out:  0%

Submit

Proceed

Percentage to filter outを0%に設定

Submitをクリック → Proceedをクリック

Xia Lab @ Meti (last updated 2023-12-13)

# スケーリングの選択と実行

The screenshot displays a web-based data processing interface. On the left is a sidebar menu with options: Upload, Processing (expanded), Data check, Missing value, Data filter, Data editor, Normalization, Statistics, Download, and Exit. The main panel contains instructions and configuration options. At the top, two bullet points explain data transformation and scaling. Below, three sections are visible: 'Sample normalization' with options like None, Sample-specific normalization, Normalization by sum, Normalization by median, Normalization by a reference sample (PQN), Normalization by a pooled sample from group (group PQN), Normalization by reference feature, and Quantile normalization; 'Data transformation' with options like None, Log transformation, Square root transformation, and Cube root transformation; and 'Data scaling' with options like None, Mean centering, Auto scaling (highlighted with a red box), Pareto scaling, and Range scaling. The 'Auto scaling' option is described as '(mean-centered and divided by the standard deviation of each variable)'. At the bottom of the main panel, three buttons are present: 'Normalize' (highlighted with a red box), 'View Result', and 'Proceed' (highlighted with a red box). A large text overlay 'Auto scalingを選択' points to the 'Auto scaling' option. In the top right corner, there is a link 'Show R Command'. The footer of the interface reads 'Xia Lab @ McGill (last updated 2023-2-17)'.

• Data transformation applies a mathematical transformation on individual values themselves. A simple mathematical approach is used to deal with negative values in log and square root. Please search 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)
- ☐ Square root transformation (square root of data values)
- ☐ Cube root transformation (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](#)

Xia Lab @ McGill (last updated 2023-2-17)

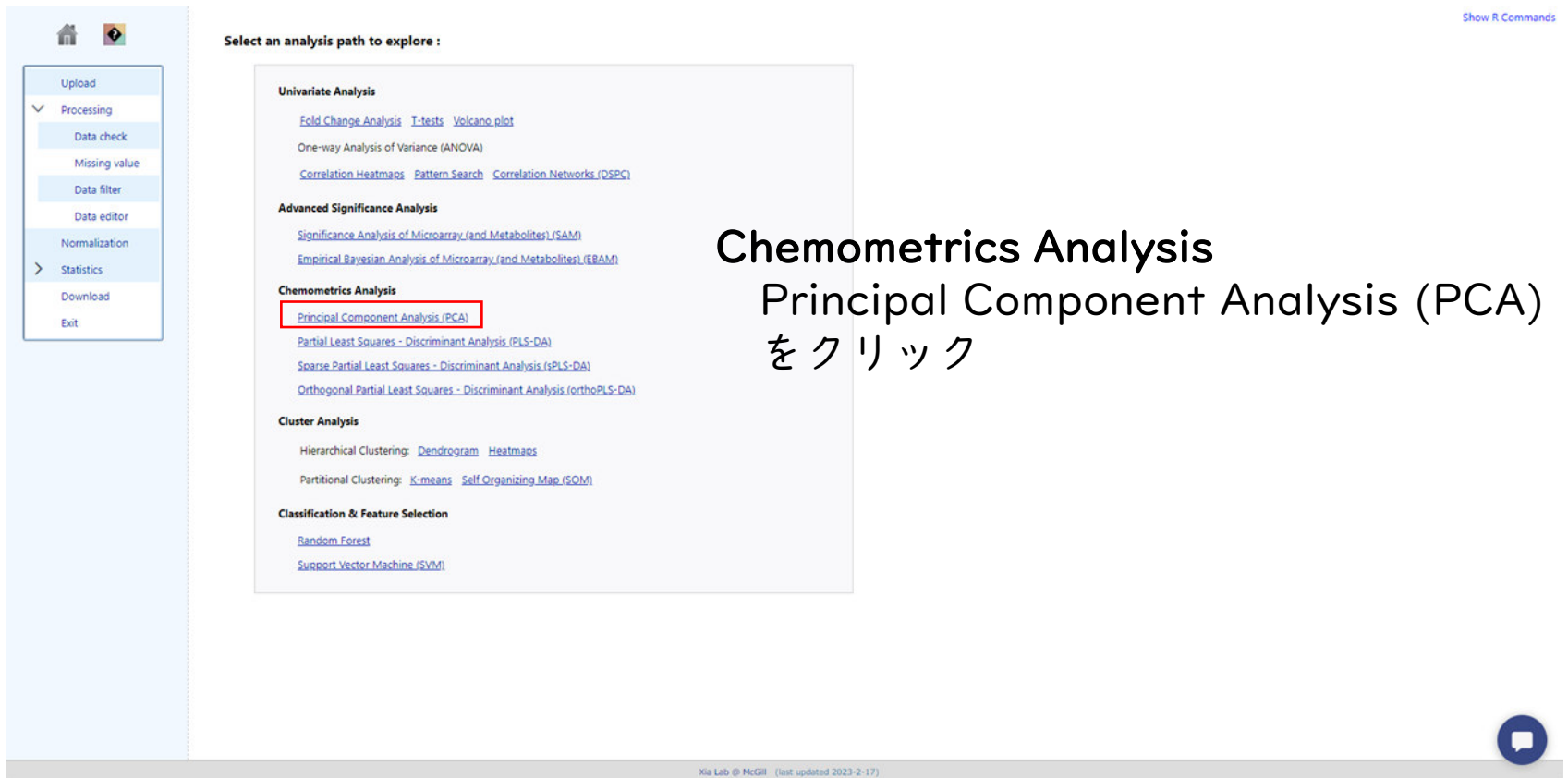
Show R Command

Normalizeをクリック



Proceedをクリック

# 解析メニュー一覧



The screenshot displays a software interface with a sidebar on the left and a main content area. The sidebar contains a menu with the following items: Upload, Processing (expanded), Data check, Missing value, Data filter, Data editor, Normalization, Statistics (expanded), Download, and Exit. The main content area is titled "Select an analysis path to explore :" and lists several analysis categories. Under the "Chemometrics Analysis" category, the "Principal Component Analysis (PCA)" option is highlighted with a red rectangle. Other categories include "Univariate Analysis", "Advanced Significance Analysis", "Cluster Analysis", and "Classification & Feature Selection".

Univariate Analysis

- [Fold Change Analysis](#) [T-tests](#) [Volcano plot](#)
- One-way Analysis of Variance (ANOVA)
- [Correlation Heatmaps](#) [Pattern Search](#) [Correlation Networks \(QSPC\)](#)

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](#)
- Partitional Clustering: [K-means](#) [Self Organizing Map \(SOM\)](#)

Classification & Feature Selection

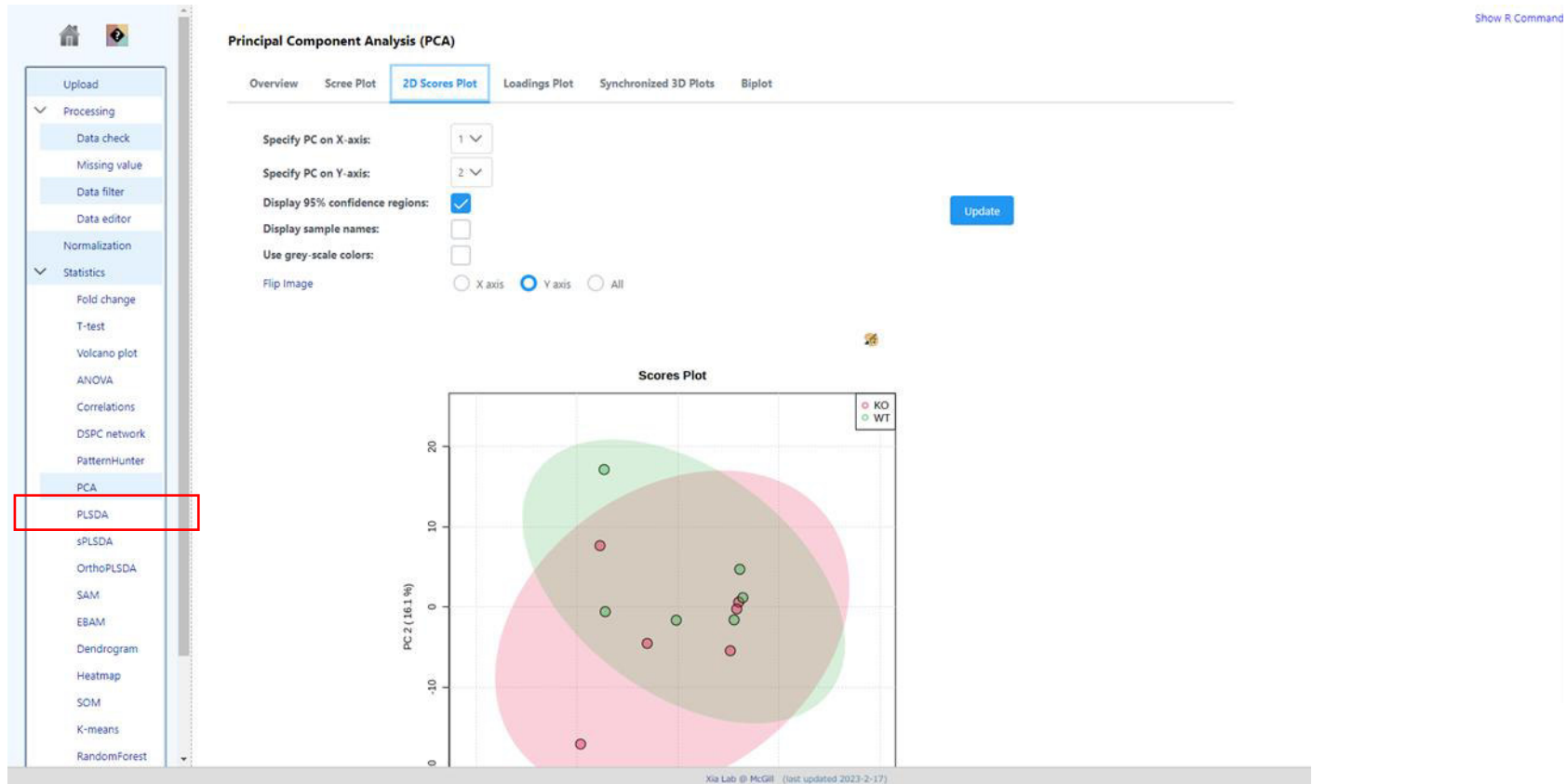
- [Random Forest](#)
- [Support Vector Machine \(SVM\)](#)

Chemometrics Analysis  
Principal Component Analysis (PCA)  
をクリック

Show R Commands

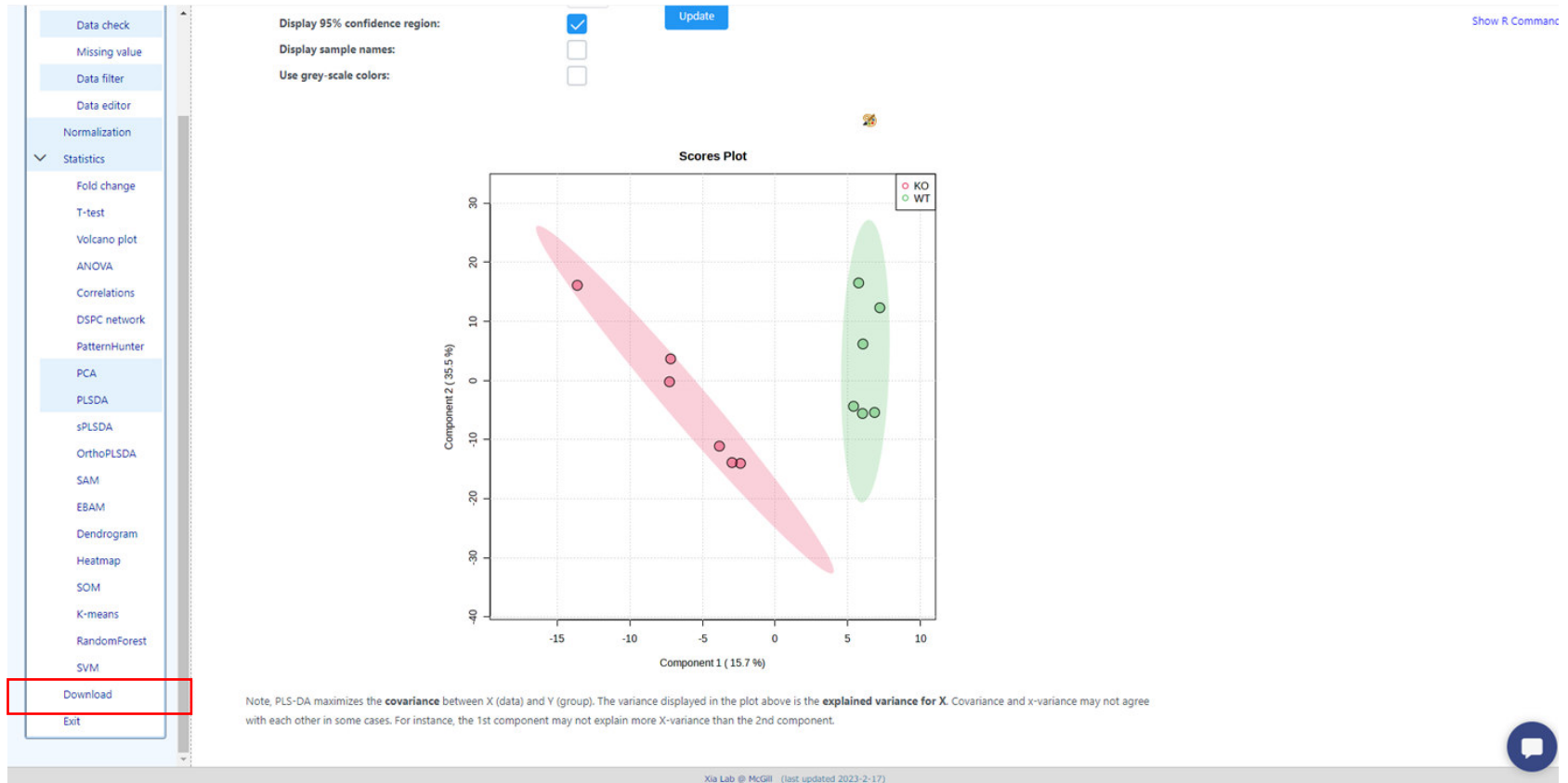
Xia Lab @ McGill (last updated 2023-2-17)

# 主成分分析の結果を確認



次にPLSDAをクリックし、PLS-DAの結果を確認

# PLS-DAの結果を確認



次にDownloadをクリックし、解析用データと結果をダウンロードする

# 結果のダウンロード

**Download Results & Start New Journey**

Please download the results (tables and images) from the **Results Download** tab below. The **Download.zip** contains all the files in your home directory. You can also generate a **PDF analysis report** using the button. Finally, you can continue to explore other compatible modules using the **Start New Journey** tab.

**Results Download** | **Start New Journey**

**Generate Report**

**Download.zip** | [snorm\\_0\\_dpi72.png](#)

[Rhistory.R](#) | [pls\\_loading\\_0\\_dpi72.png](#)

[pca\\_score.csv](#) | [pls\\_score3d\\_0.json](#)

[pca\\_loading3d\\_0.json](#) | [data\\_normalized.csv](#)

[pca\\_pair\\_0\\_dpi72.png](#) | [norm\\_0\\_dpi72.png](#)

[pls\\_loading3d\\_0.json](#) | [pca\\_scee\\_0\\_dpi72.png](#)

[plsda\\_loadings.csv](#) | [pca\\_loadings.csv](#)

[plsda\\_score.csv](#) | [pca\\_loading\\_0\\_dpi72.png](#)

[loadings3D.png](#) | [pls\\_pair\\_0\\_dpi72.png](#)

[plsda\\_vic.csv](#) | [pca\\_score2d\\_0\\_dpi72.png](#)

[pls\\_cv\\_0\\_dpi72.png](#) | [pls\\_imo\\_0\\_dpi72.png](#)

[pca\\_biplot\\_0\\_dpi72.png](#) | [pca\\_score3d\\_0.json](#)

[data\\_processed.csv](#) | [scores3D.png](#)

**data\_original.csv** | [pls\\_score2d\\_0\\_dpi72.png](#)

[plsda\\_coef.csv](#)

**Logout**

Xia Lab @ McGill (last updated 2023-2-17)

data\_original.csvは後で使用するので、  
ファイル名をmouse\_data\_original.csvに変更して、C:¥Rに保存する  
(上手く行かなかった方は、配布用ファイルをご利用ください)