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

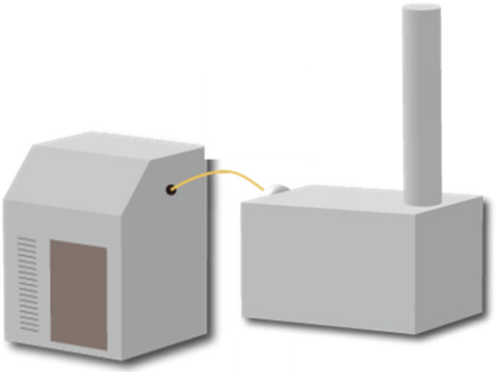
メタボロミクスの研究例(1)

絶食マウス肝臓のメタボロームデータ

通常飼育
(n=5)

12時間絶食
(n=5)

肝臓サンプル



キャピラリー電気泳動
飛行時間型質量分析を用いた
メタボローム解析

10サンプル

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

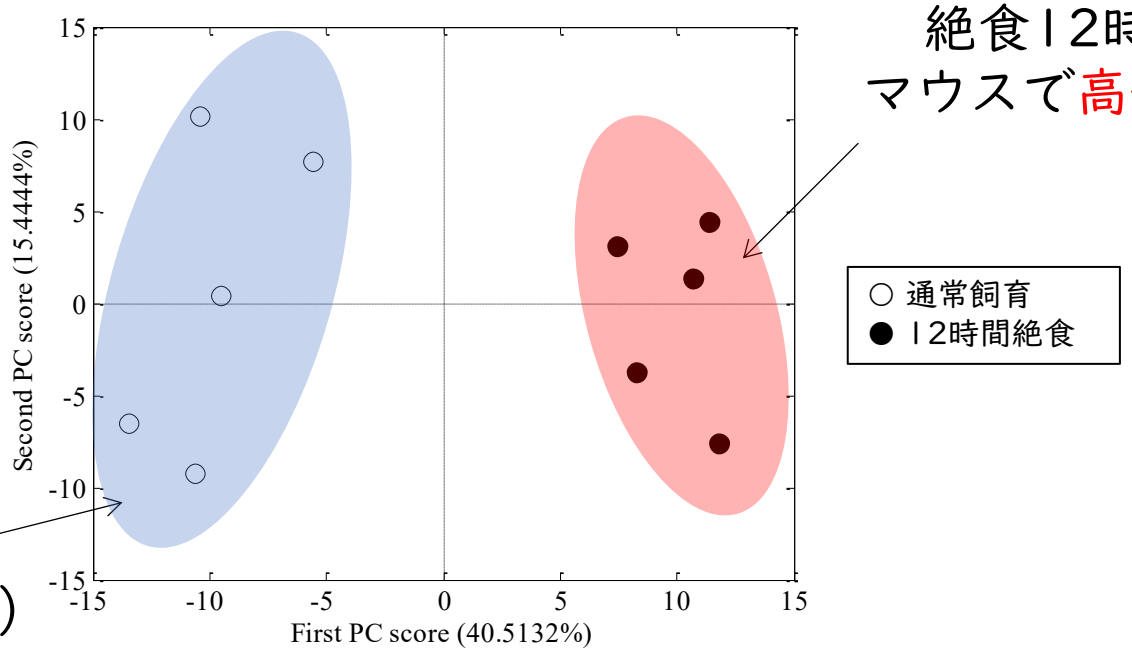
282変数(代謝物質)

主成分分析

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

絶食12時間
マウスで高値(+)



通常飼育
マウスで低値(−)

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で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

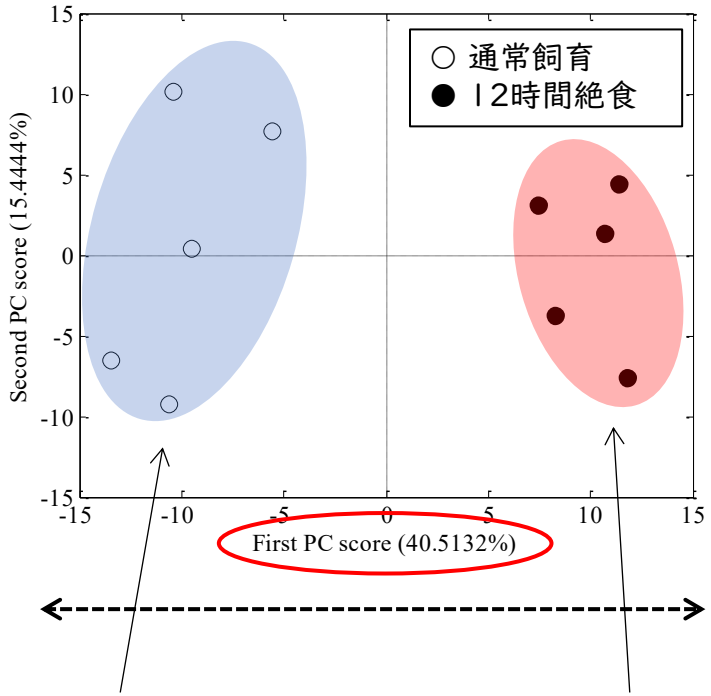
$$\begin{bmatrix} t_{11} \\ t_{21} \\ \vdots \\ t_{n1} \end{bmatrix}$$

主成分スコア＝(代謝物1)×w_1＋(代謝物2)×w_2＋・・・＋(代謝物P)×w_p

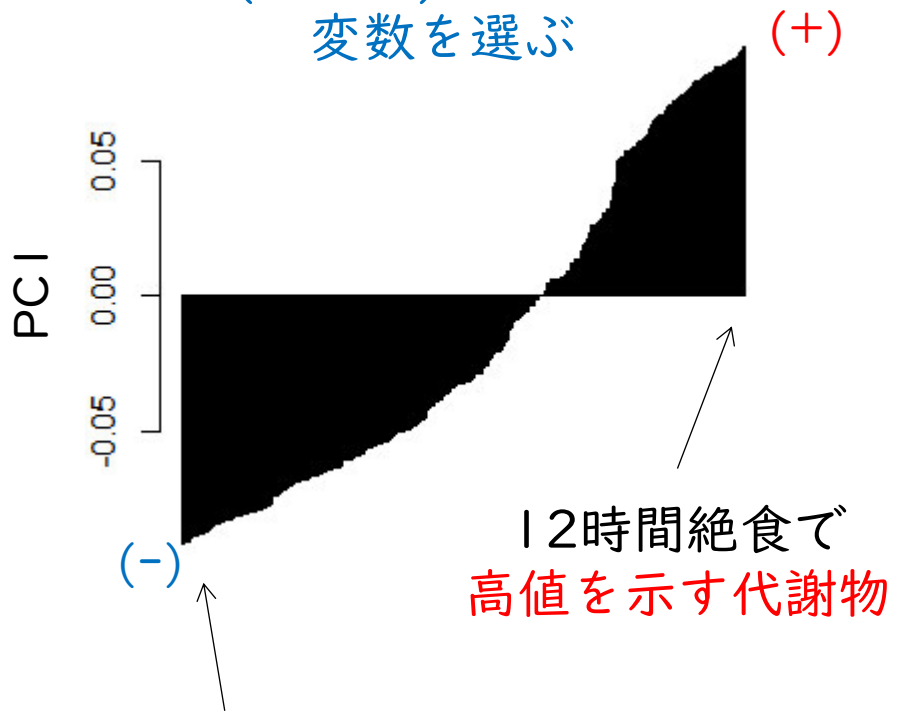
$$t = x_1 w_1 + x_2 w_2 + \cdots + x_p w_p$$

絶食マウスでの主成分分析の解析例

データの可視化



重み(主成分)係数を用いて
変数を選ぶ



12時間絶食で
高値を示す代謝物

通常飼育
マウスで低値(−)

絶食12時間
マウスで高値(+)

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)

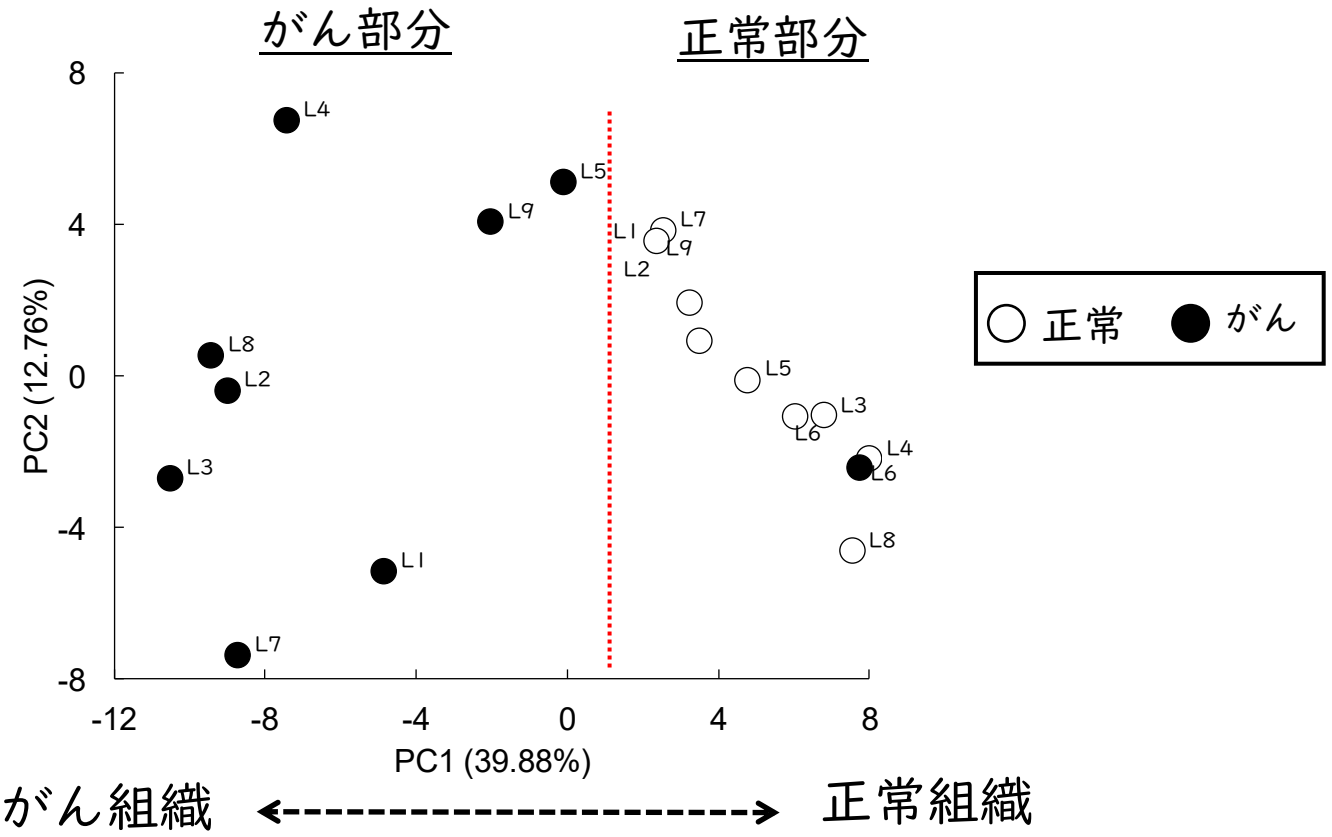
メタボロミクスの研究例(2)

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

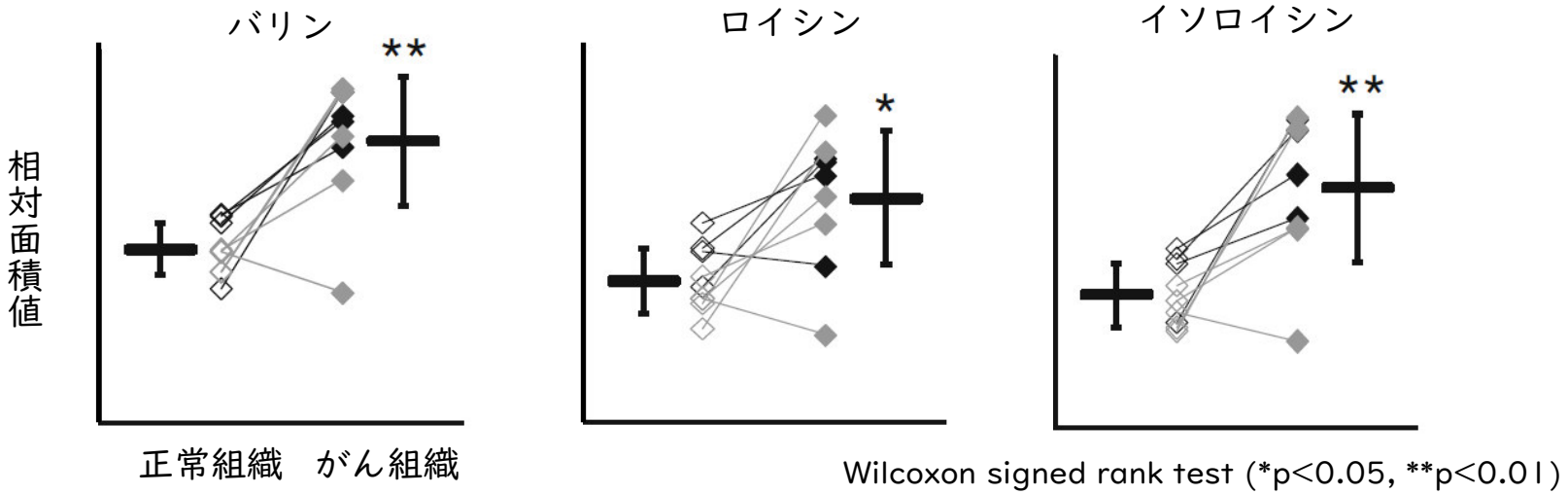


横軸はがん部分と正常部分の違いを表している

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

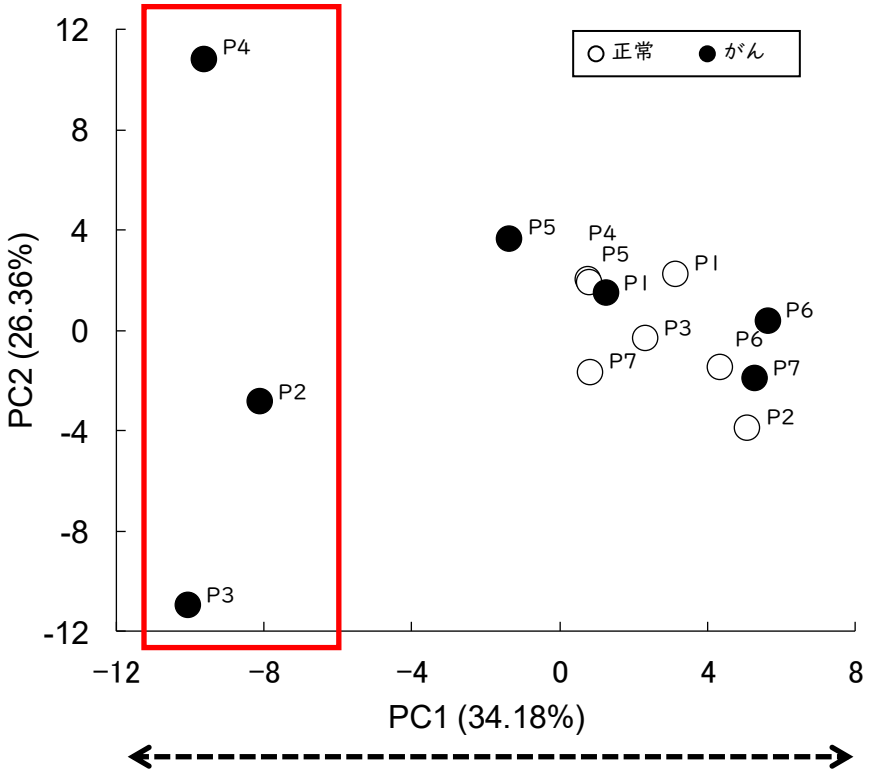


分岐鎖アミノ酸は、がん組織によって吸収され、がん患者でほとんど酸化されることが知られている(Baracos and Mackenzie, 2006)

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

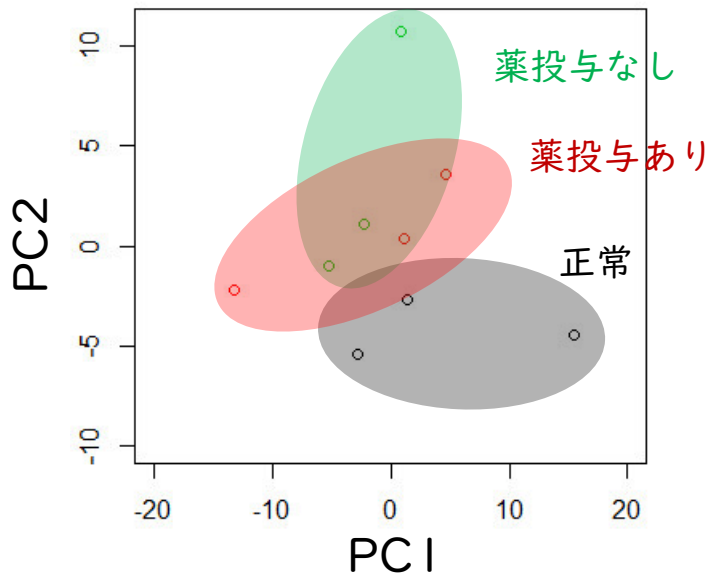
Ooga T, Sato H, Nagashima A, Sasaki K, Tomita M, Soga T, Ohashi Y., “Metabolomic anatomy of an animal model revealing homeostatic imbalances in dyslipidaemia.”, Mol Biosyst. 2011 Apr;7(4):1217-23.

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

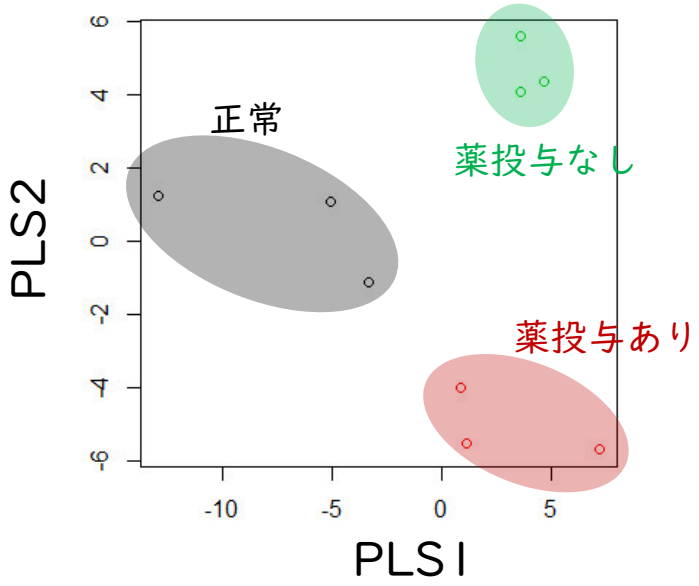
主成分分析とPLSの解析例

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



PLS



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

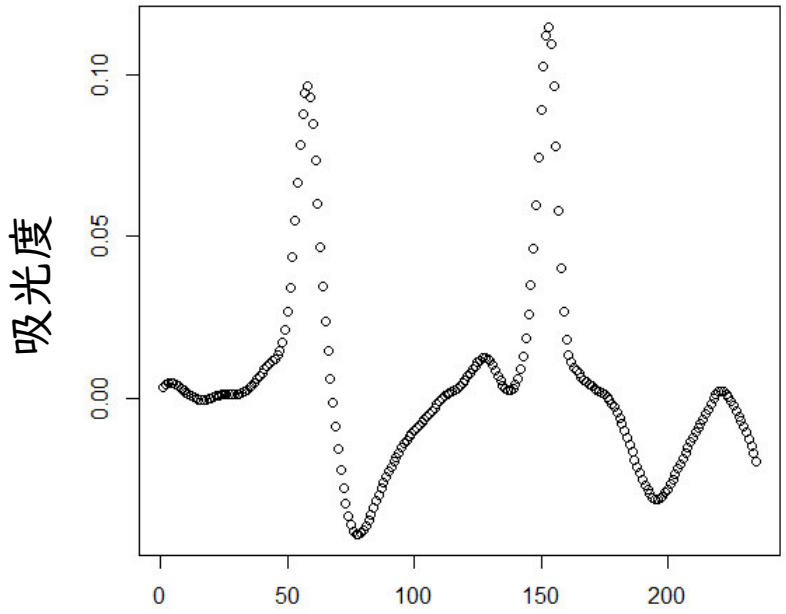
- 食品
  - 小麦粉、スターチ、食用油、食肉等の材料系から、乳製品等の加工食品系の成分分析に用いられており、日本酒、ワイン、醤油などの液体の測定にも多く用いられている。
- 農業
  - お茶の成分測定(窒素、タンニン、水分など)や、野菜の硝酸イオン濃度、ミカンなどの糖度評価／選別に用いられる。
- 医薬品
  - 原材料の受け入れ検査や工程管理(混合均一性の確認)に用いられることが多い。また、その他にも、結晶形や結晶化度のチェックに用いることが出来ることが知られている。

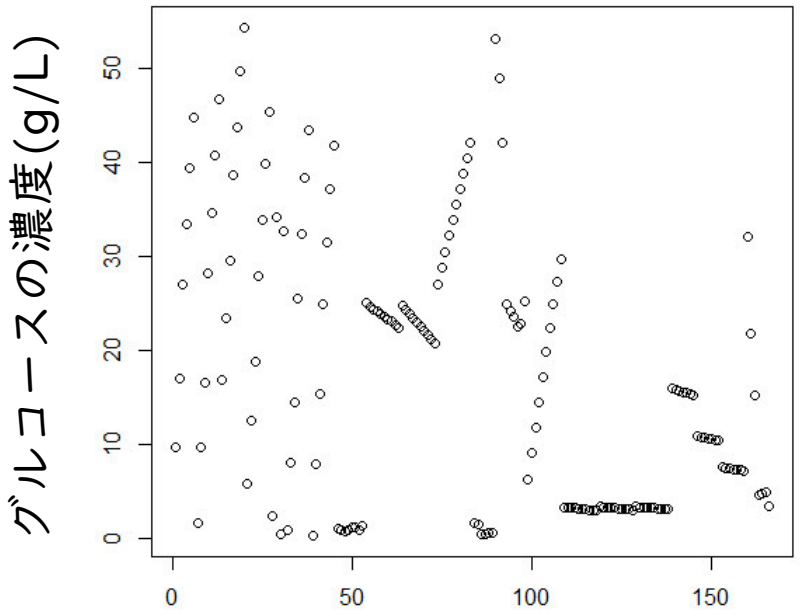
“https://ja.wikipedia.org/wiki/近赤外線分光法”より一部改変

ケモメ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



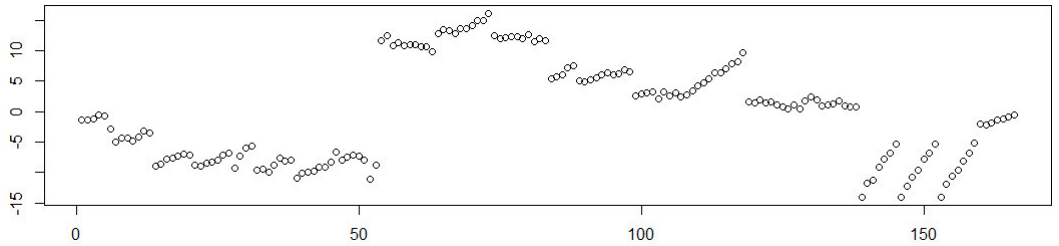


同様のスペクトルが166サンプル分ある

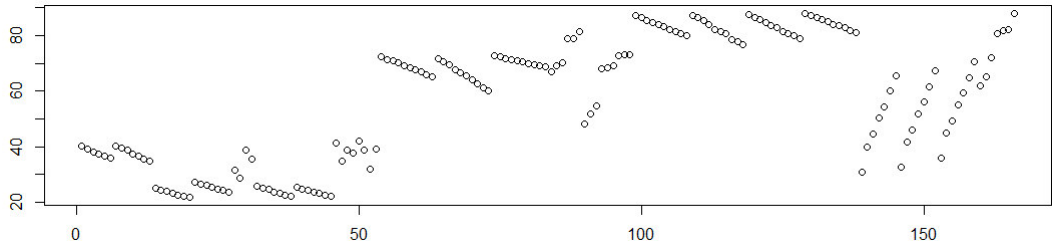
エタノールについても同様のデータ

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

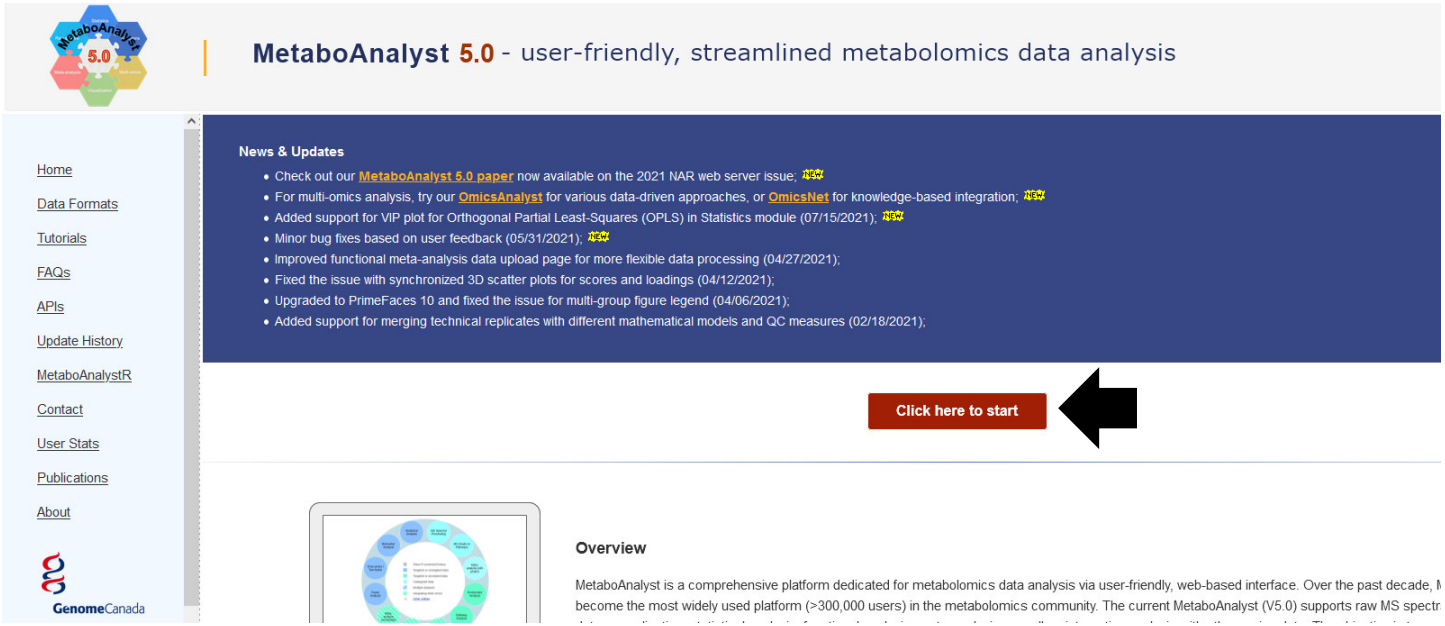
MetaboAnalystを用いた解析

MetaboAnalystを用いた解析

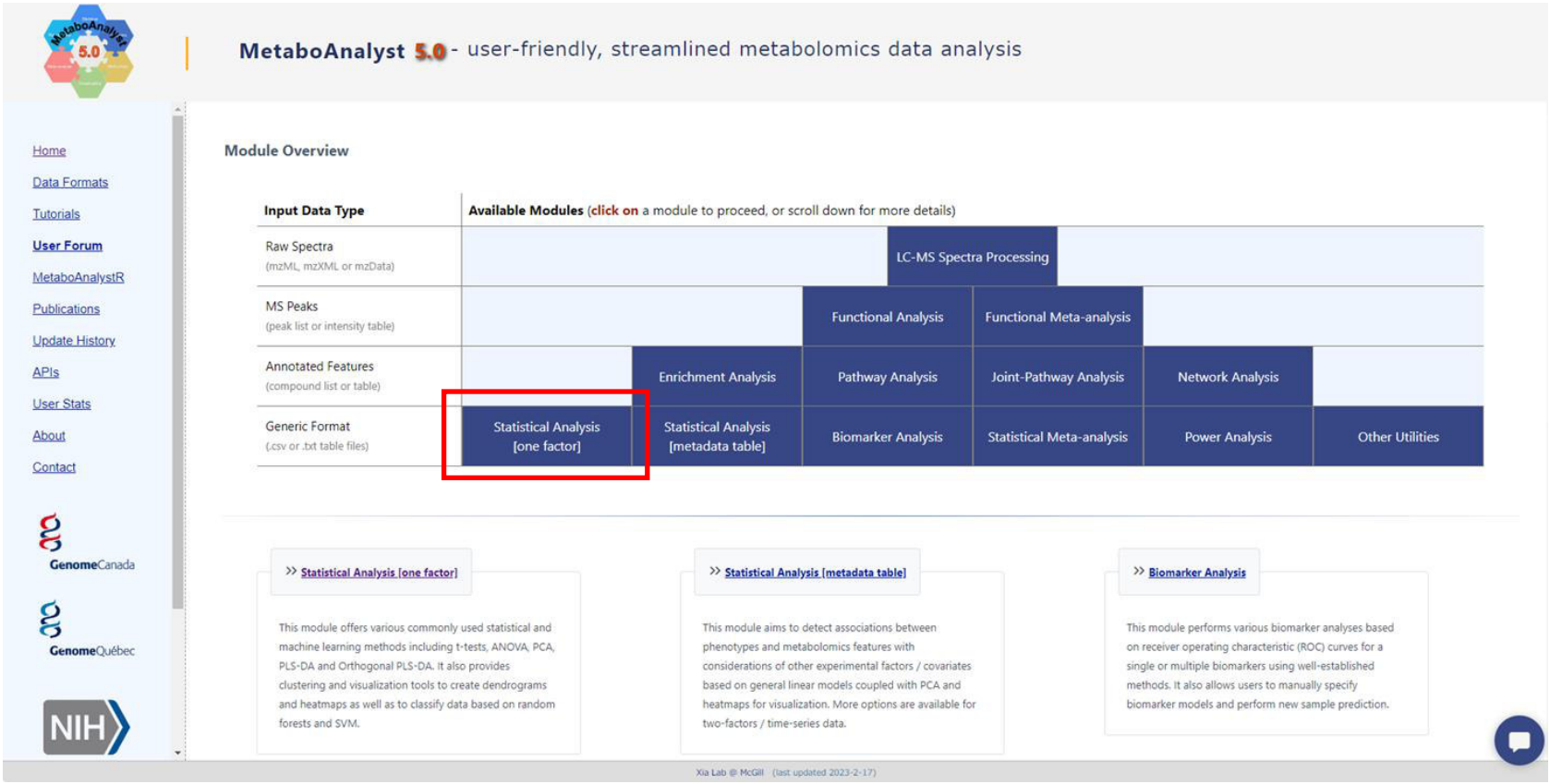
- MetaboAnalystとは
  - カナダのアルバータ大学のDavid Wishartとマギル大学のJianguo Xiaの研究グループにより運営されているフリーのWebサービス

以下のwebサイトにアクセス

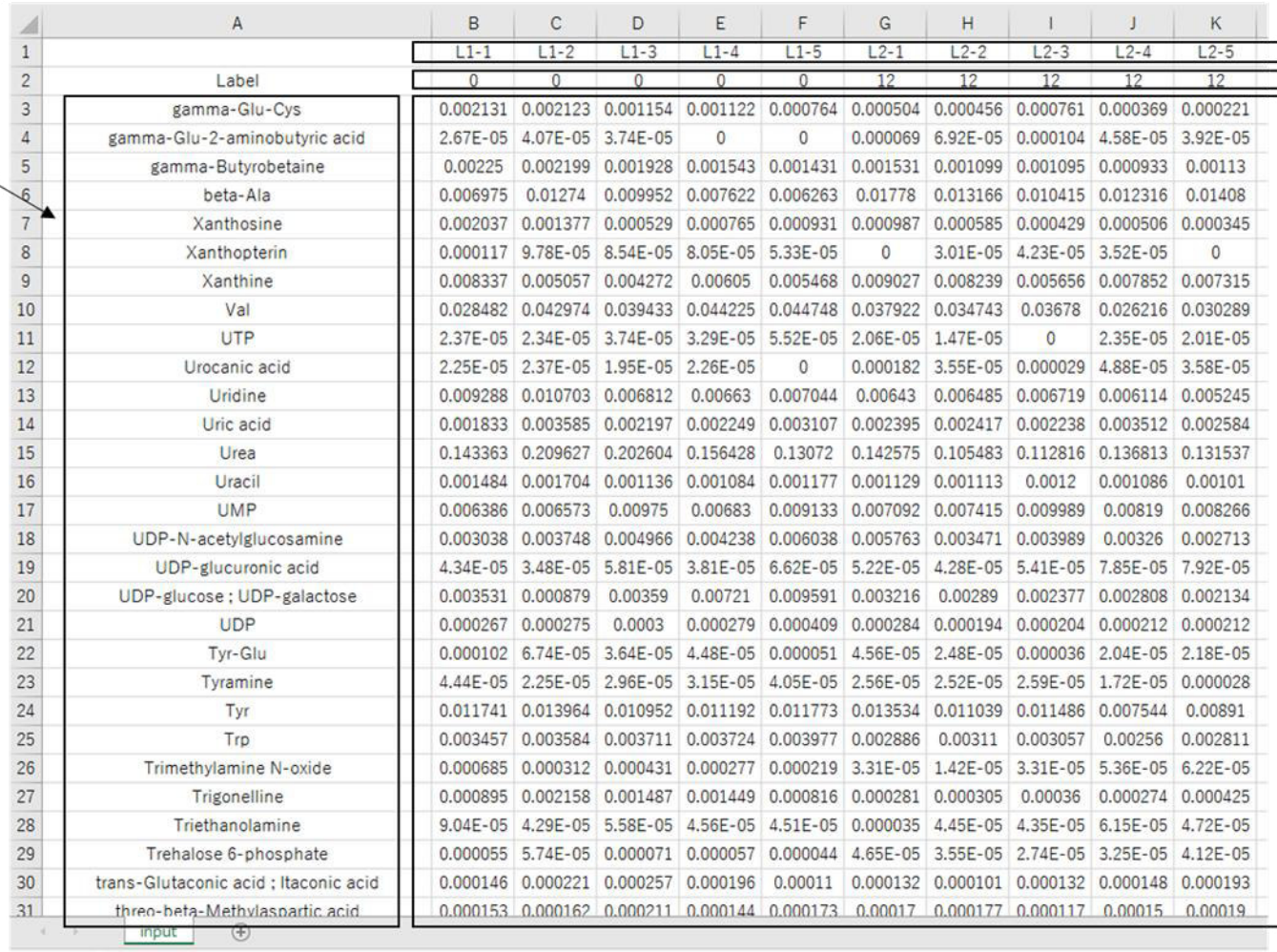
- http://www.metaboanalyst.ca/ (最新版、2024.9時点では、6.0)
- https://genap.metaboanalyst.ca/ (5.0)



Statistical Analysis [one factor]を選択



解析用のデータ(csvファイル)の準備



	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

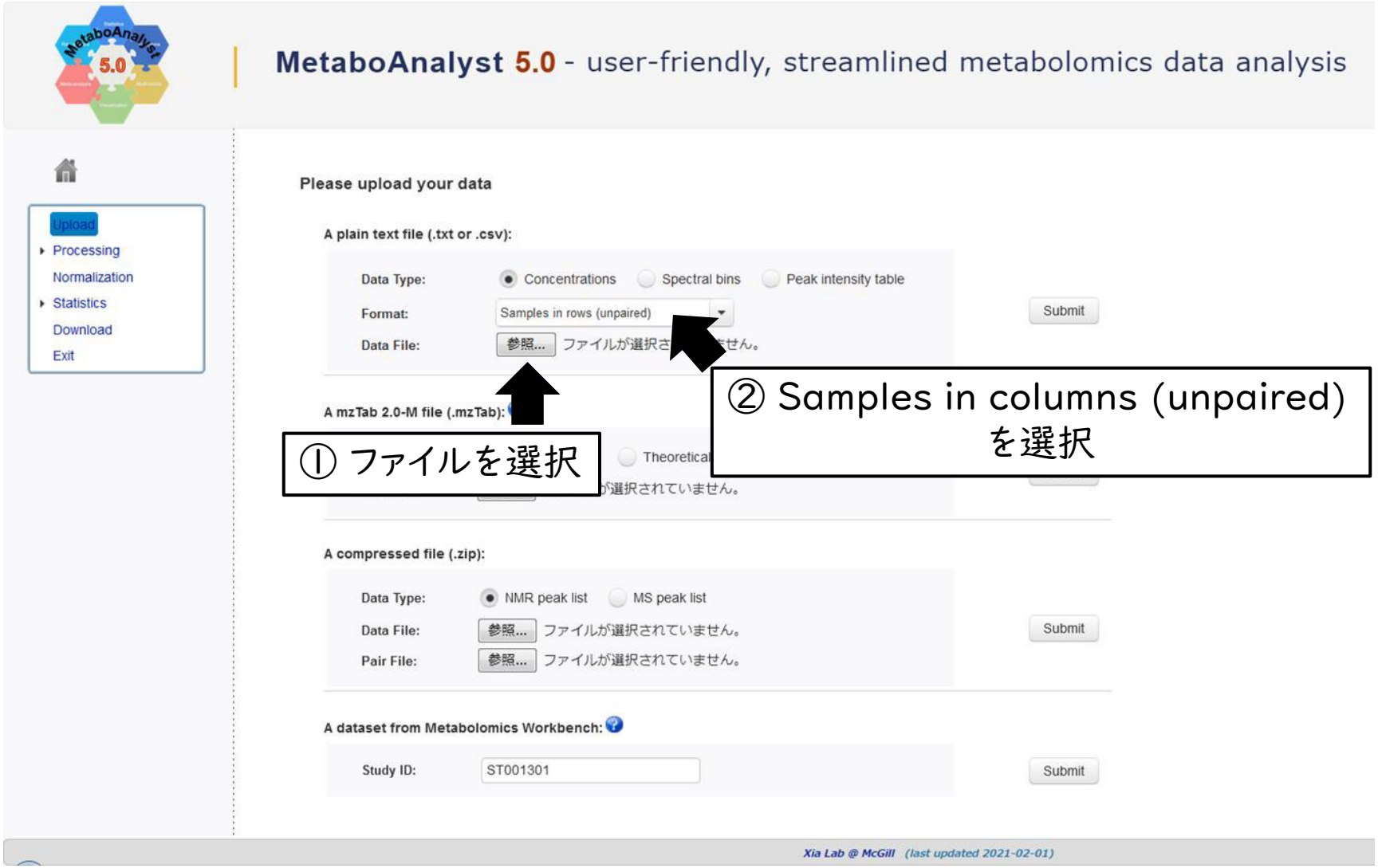
サンプル名

群名

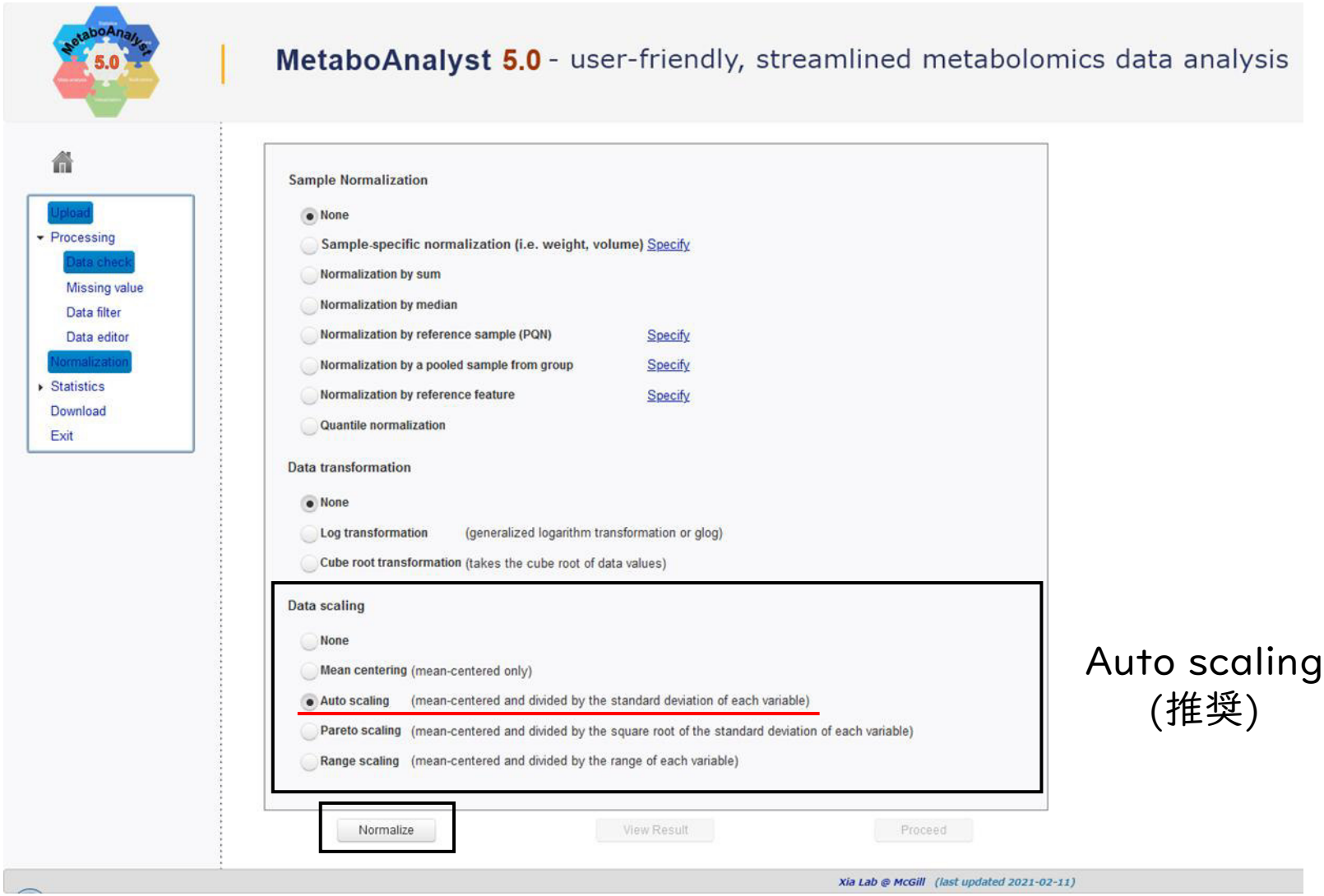
代謝物名

データ行列

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: Samples in rows (unpaired)
Data File: 参照... ファイルが選択さ せん。
Submit
A mzTab 2.0-M file (.mzTab):
Theoretical
が選択されていません。
A compressed file (.zip):
Data Type: NMR peak list MS peak list
Data File: 参照... ファイルが選択されていません。
Pair File: 参照... ファイルが選択されていません。
Submit
A dataset from Metabolomics Workbench:
Study ID: ST001301
Submit
Xia Lab @ McGill (last updated 2021-02-01)
① ファイルを選択
② Samples in columns (unpaired) を選択

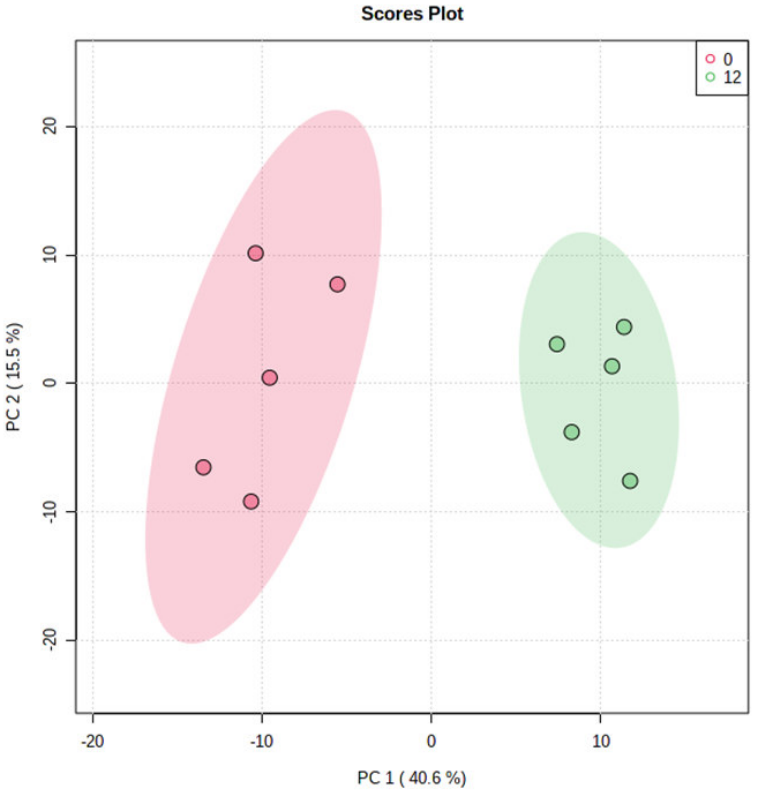
スケーリングの実行



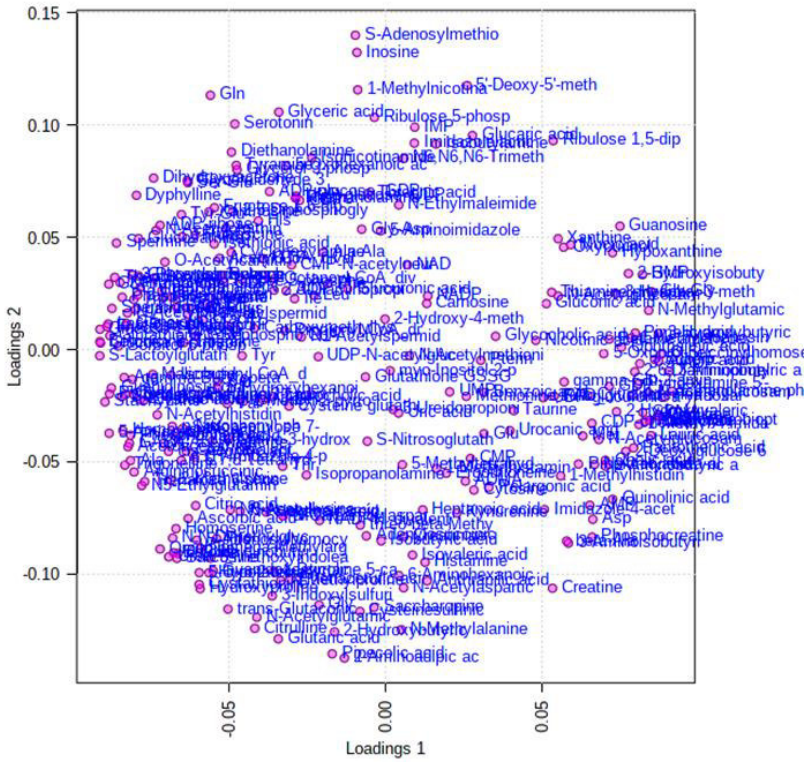
Auto scaling
(推奨)

主成分分析の結果

主成分スコア

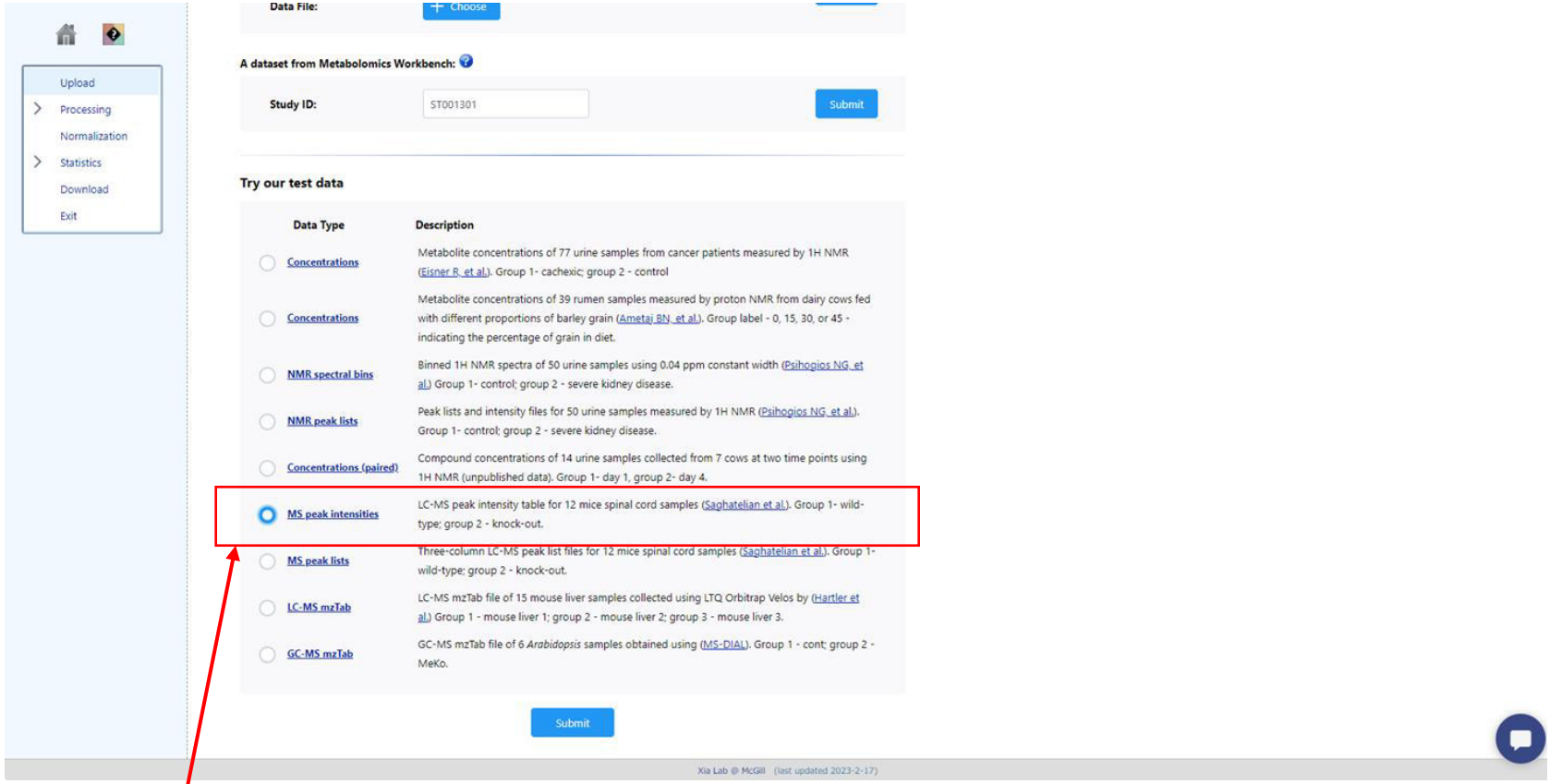


ローディング(固有ベクトル)



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

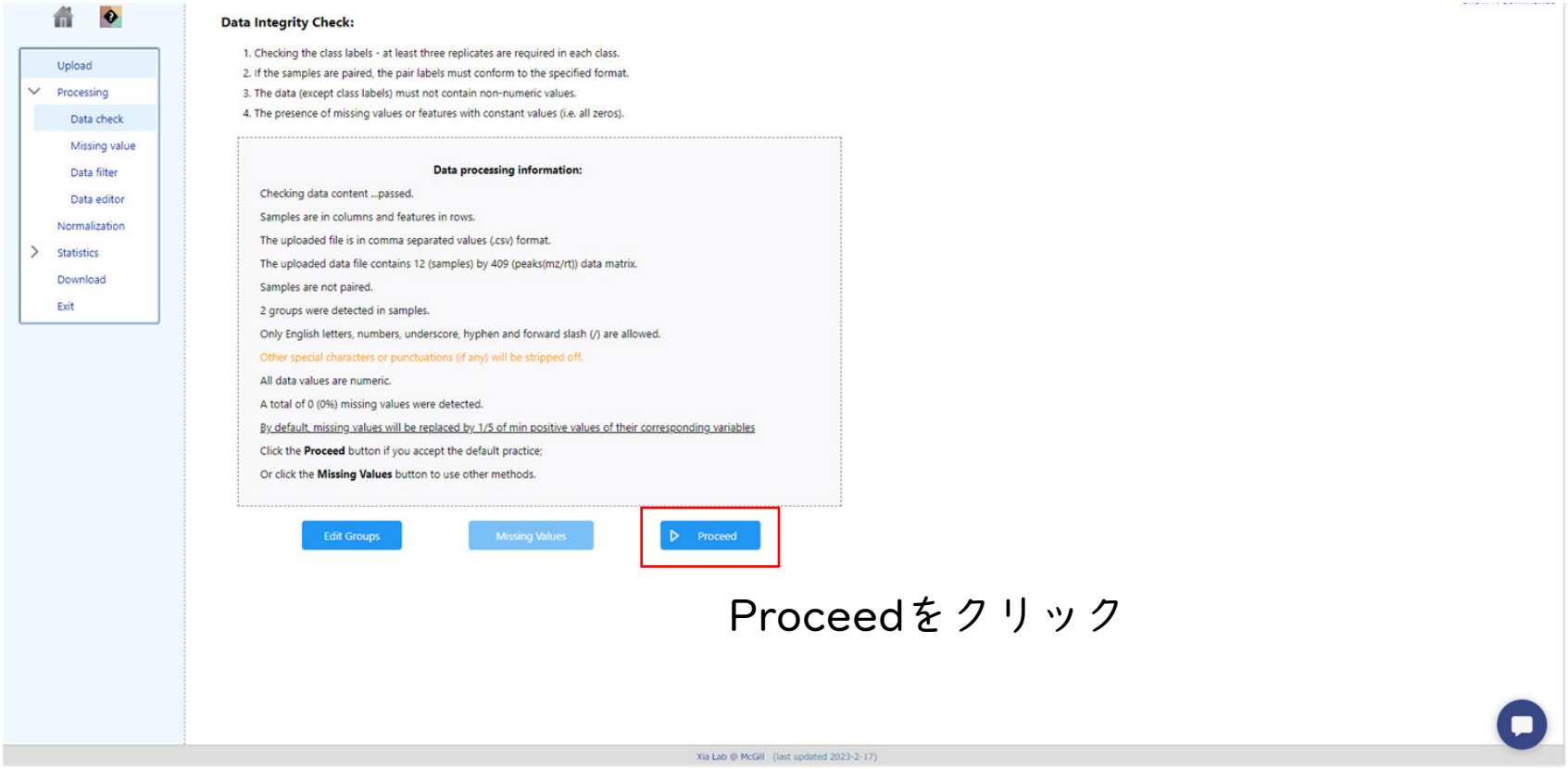
MetaboAnalystの実習(デモデータ)



MS peak intensities

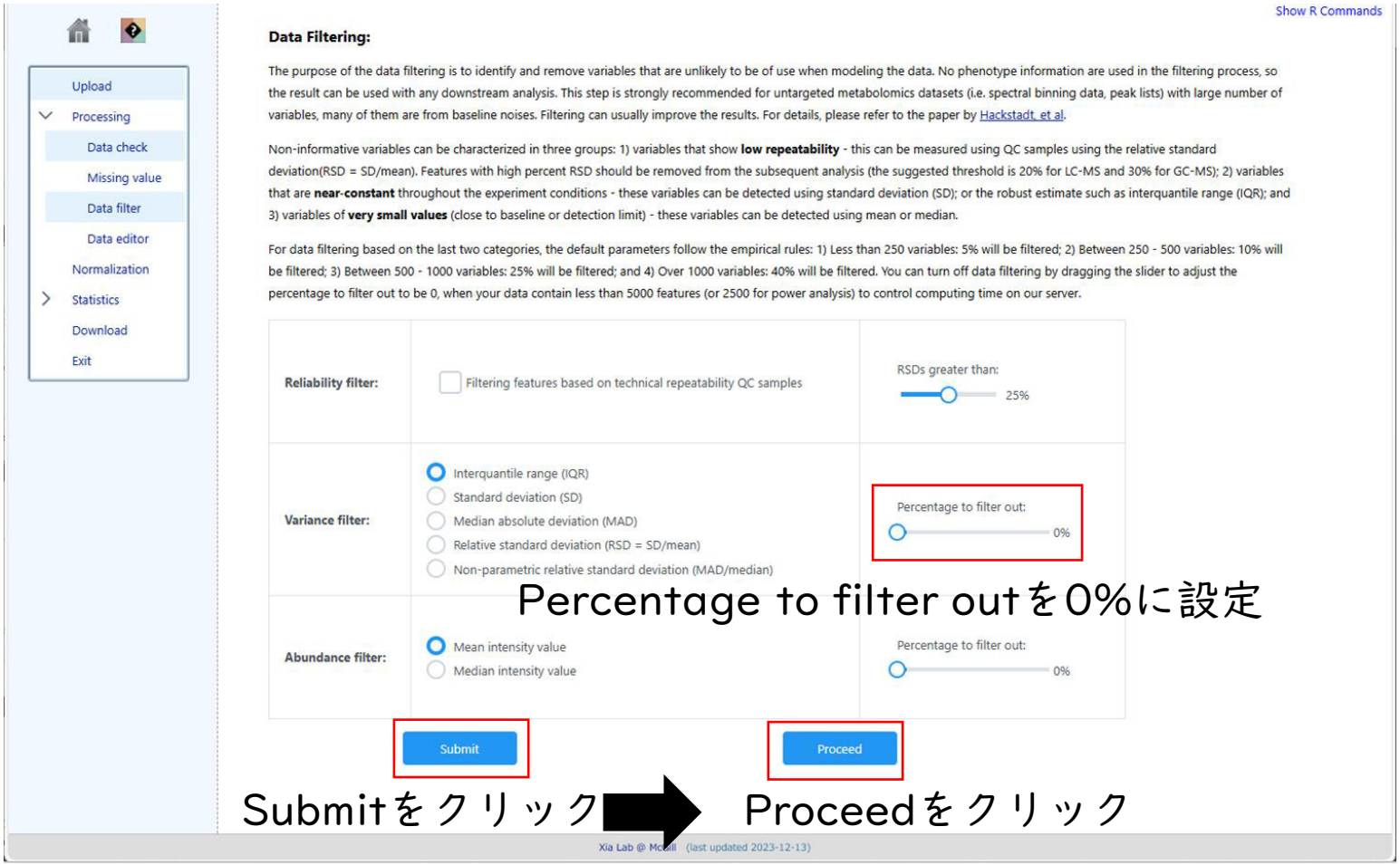
LC-MS peak intensity table for 12 mice spinal cord samples (Saghatelian et al.). Group 1- wild-type; group 2 - knock-out.

データの確認、欠損値の補完



Proceedをクリック

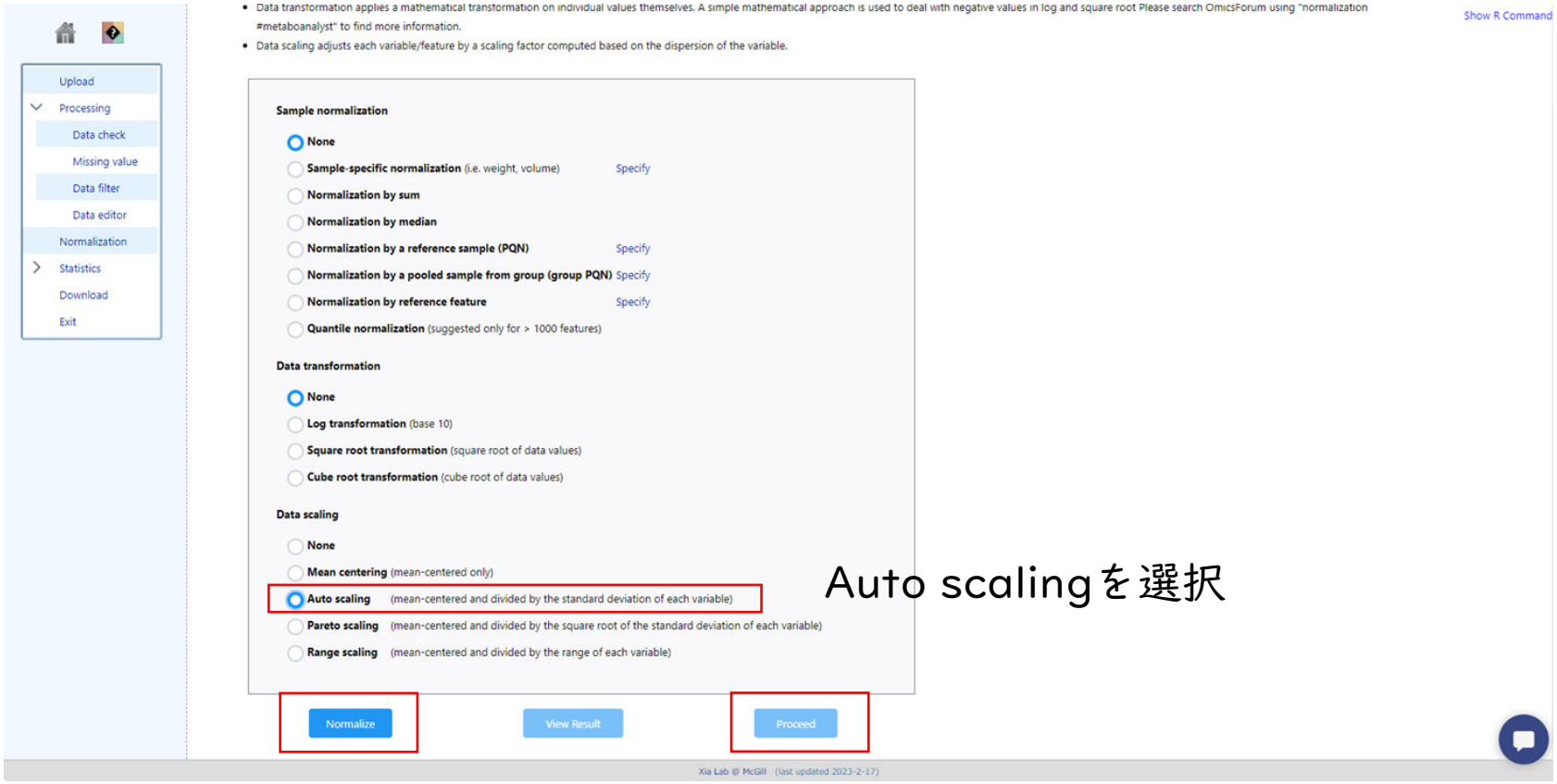
変数の除外



Percentage to filter outを0%に設定

Submitをクリック → Proceedをクリック

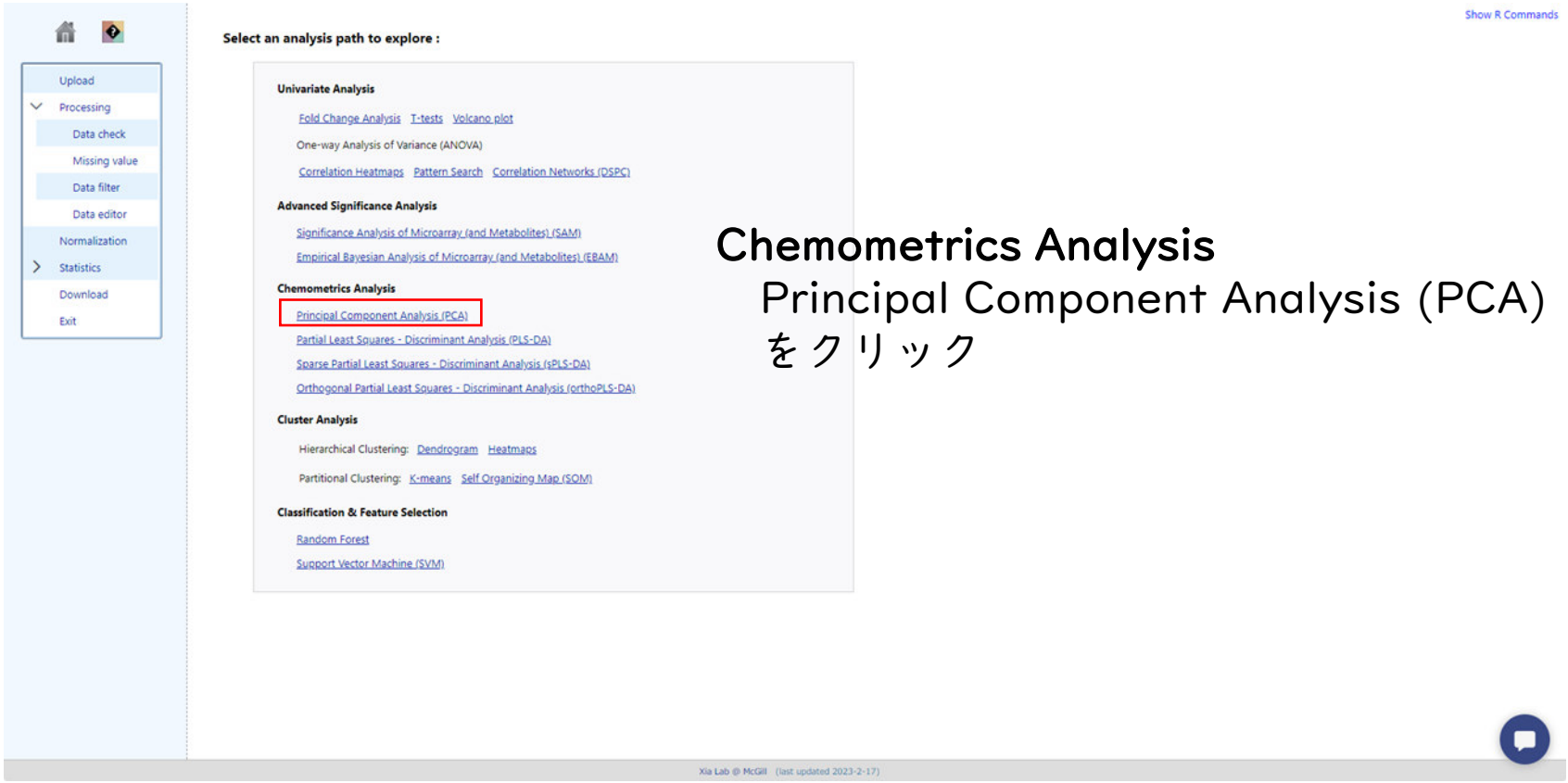
スケーリングの選択と実行



Auto scalingを選択

Normalizeをクリック → Proceedをクリック

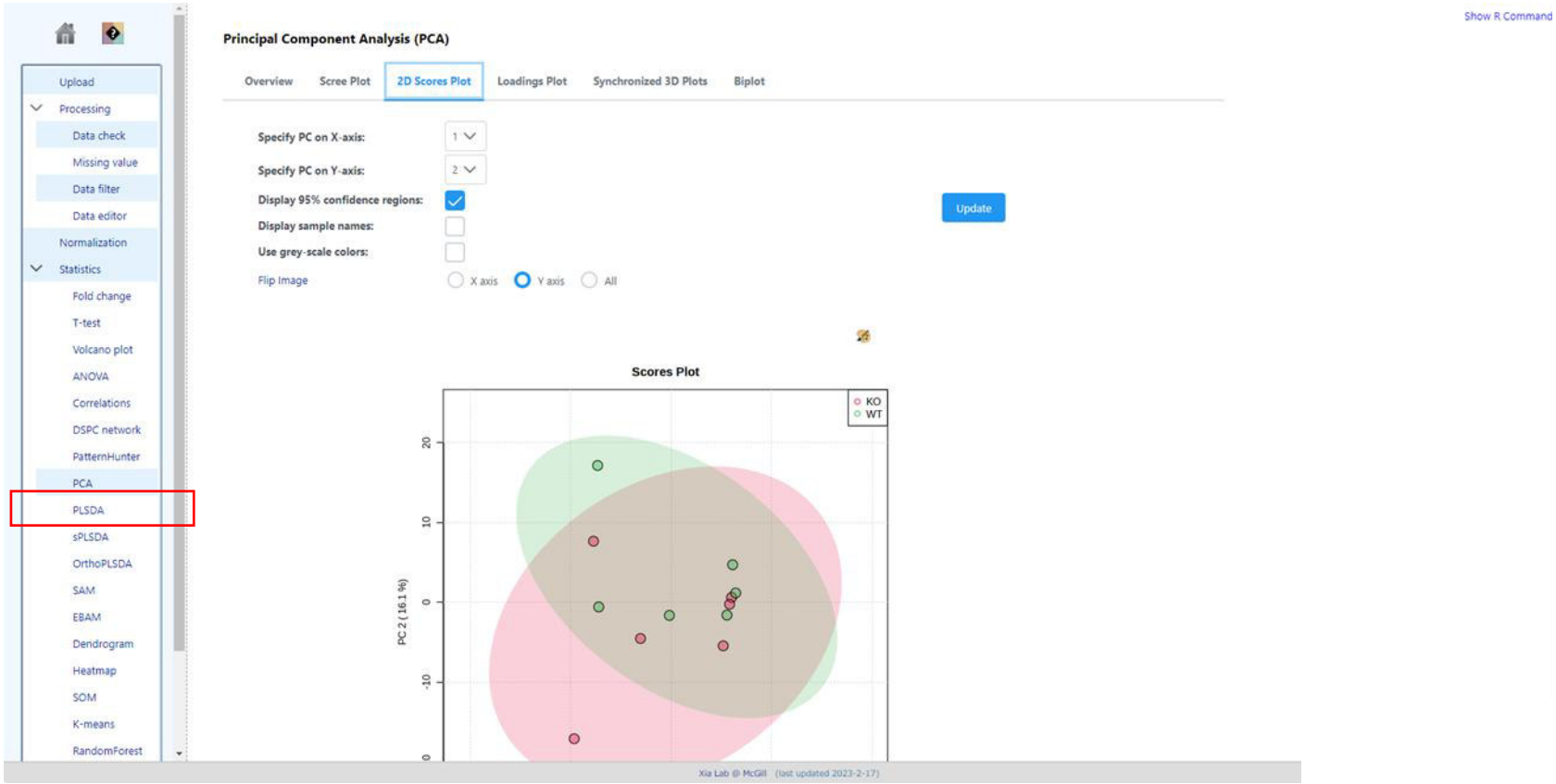
解析メニュー一覧



Chemometrics Analysis

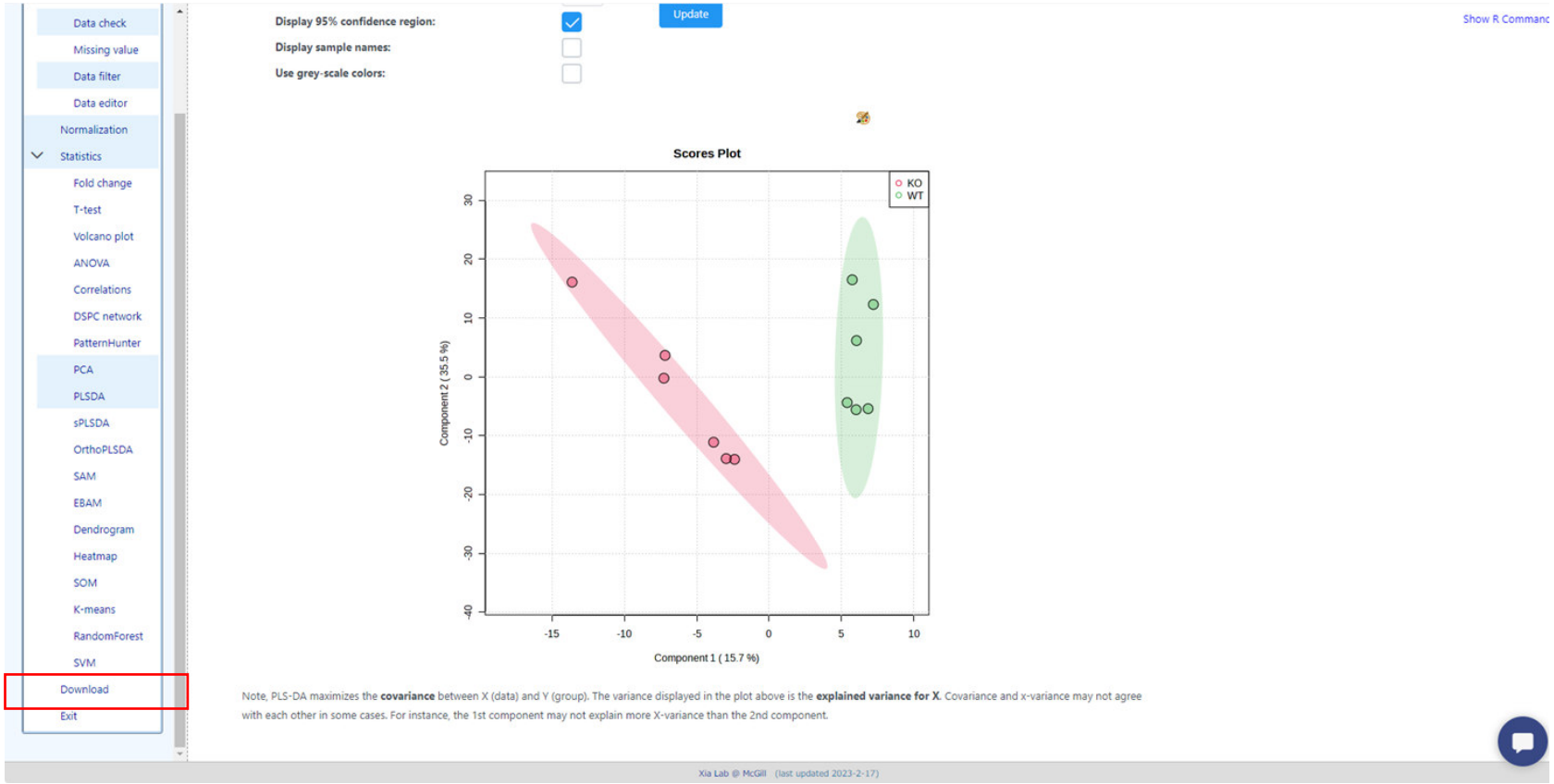
Principal Component Analysis (PCA)
をクリック

主成分分析の結果を確認



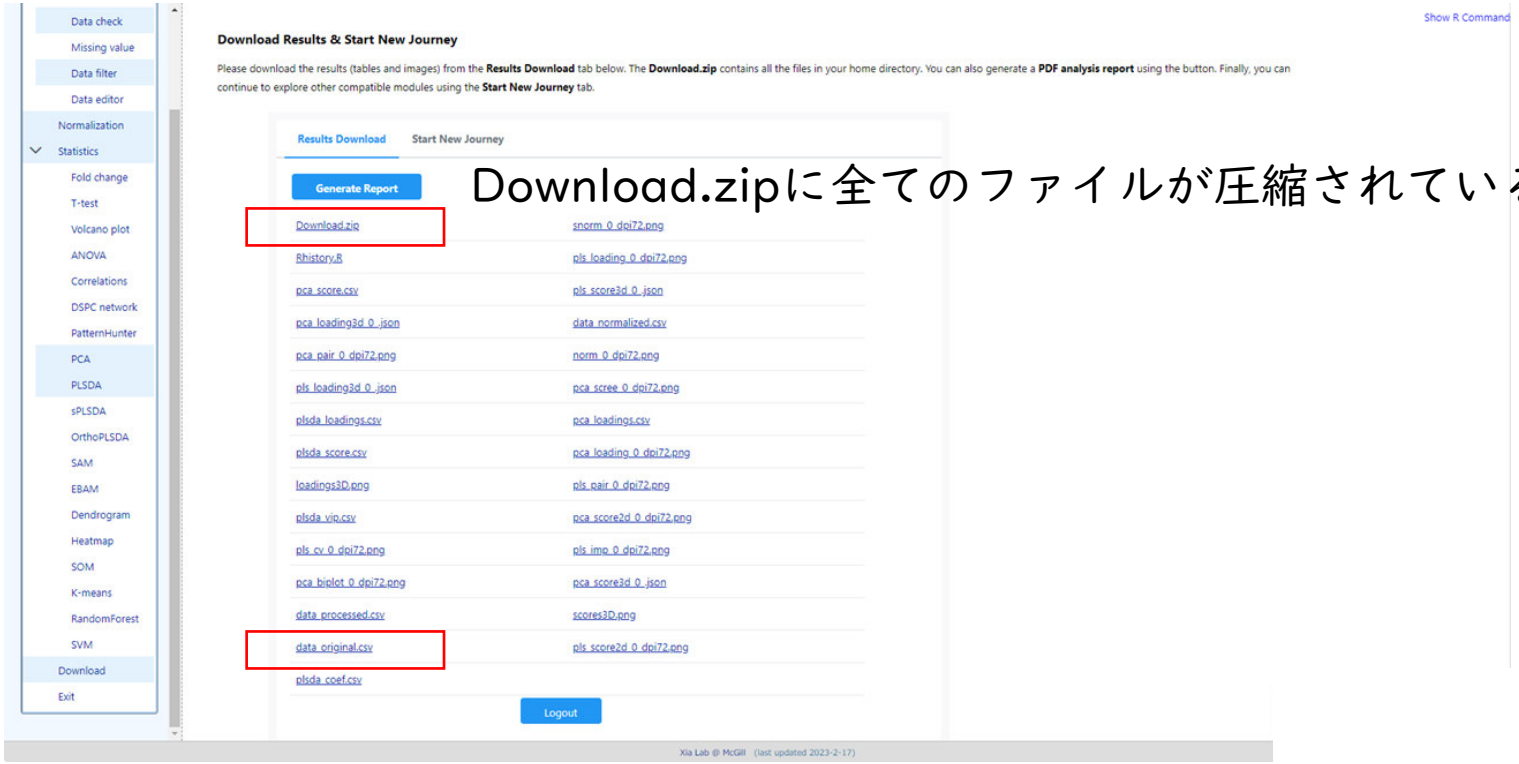
次にPLSDAをクリックし、PLS-DAの結果を確認

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.zipに全てのファイルが圧縮されている

data_original.csvは後で使用しますので、
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)