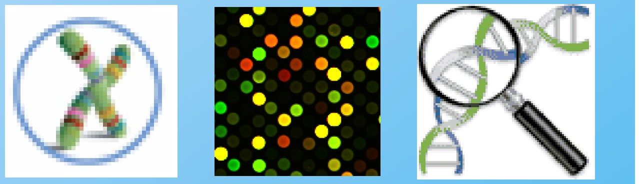
LC/MS、GC/MSデータ解析セミナー

- GeneSpring MSトレーニングセミナー



本日の内容

- はじめに

  - 多変量解析の簡単な説明

  - RT、**m/z**、強度（**Intensity**）の関係

  - 解析を行う際に考慮すべき注意点

- **BLB**に感染した**Rice**の解析例（差分解析の例）

  - データ正規化（**Normalization**）

  - Fold Analysis（倍率変化を用いた解析）

  - 統計的有意差（**T-test**による解析）

- ビール**6**銘柄の解析例（パターン解析の例）

  - 主成分分析（**PCA**）

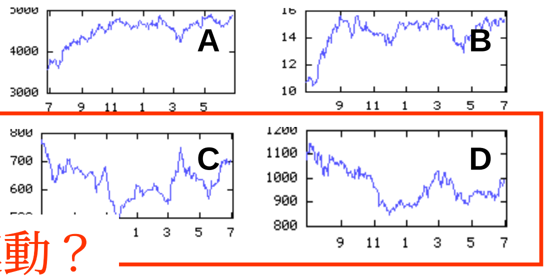
  - クラスタリング解析

多変量解析(Multivariate Analysis)とは？

多変量解析は、観測値が複数の値からなる多変量データを統計的に扱う手法。因子分析・クラスター分析・主成分分析などがある。
（ウィキペディアフリー百科事典から）

1.膨大なデータを要約して、特徴を掴みやすくするツール
2.予測に有効なツール

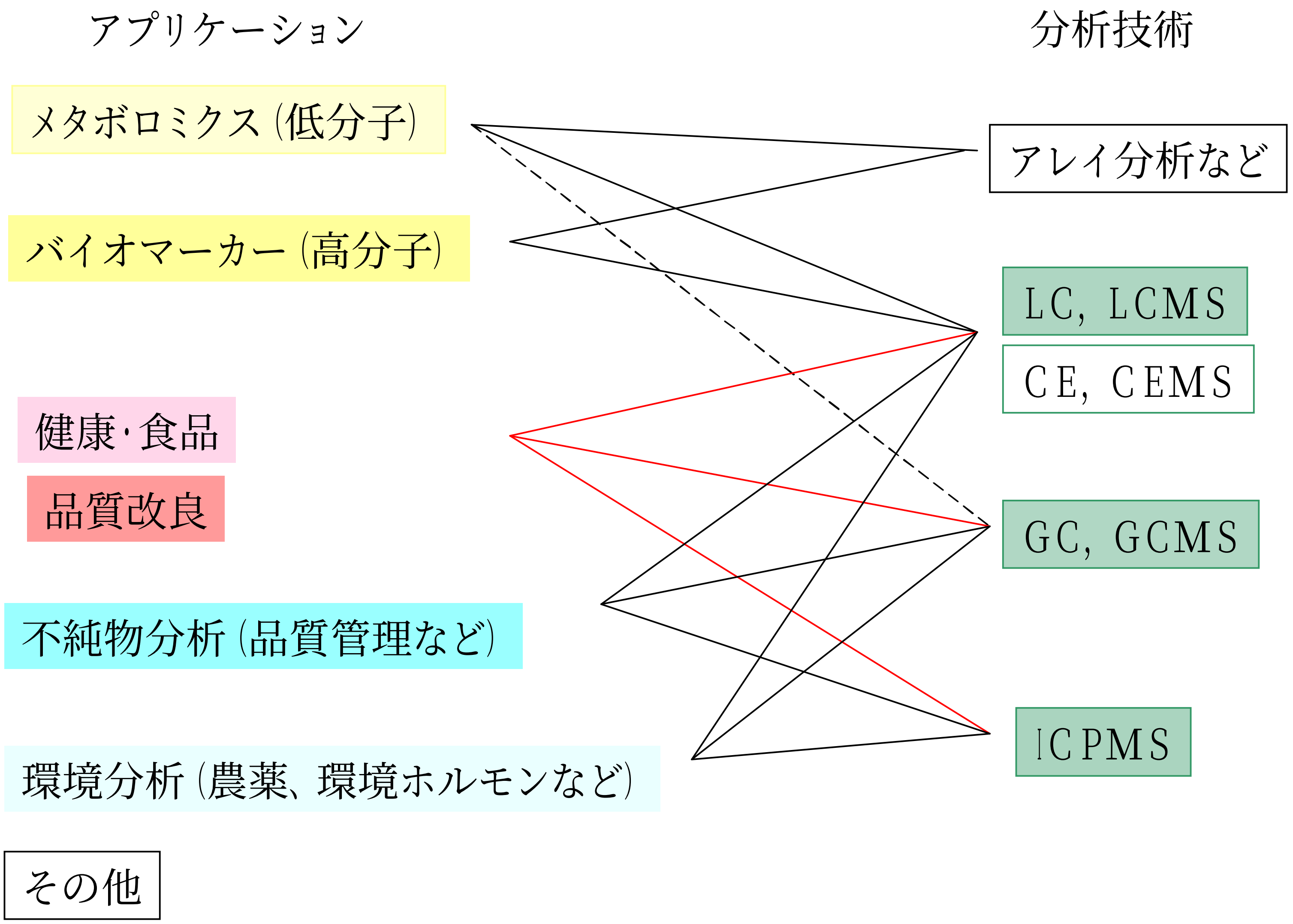
具体的には、
金融分析、株価の予測、マーケティング、マイクロアレイ解析などで用いられています。



CとDは連動？

[イメージ]　株価の例

多変量解析(Multivariate Analysis)の適用



アプリケーション

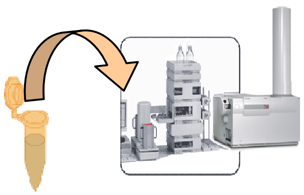
- メタボロミクス (低分子)
- バイオマーカー (高分子)
- 健康・食品
- 品質改良
- 不純物分析 (品質管理など)
- 環境分析 (農薬、環境ホルモンなど)
- その他

分析技術

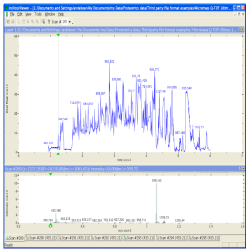
- アレイ分析など
- LC, LCMS
- CE, CEMS
- GC, GCMS
- ICPMS

GeneSpring MSを用いたデータ解析ワークフロー

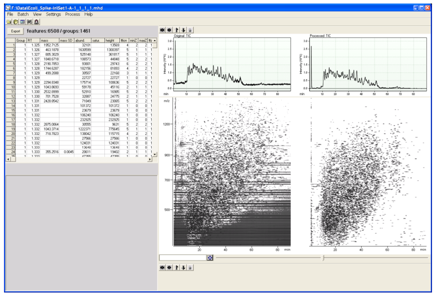
LC/MS
TOF



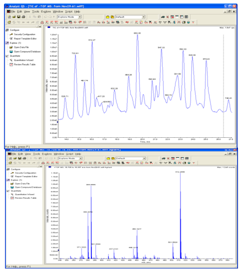
.WIFF files (TOF)
.D files (QTOF)
その他



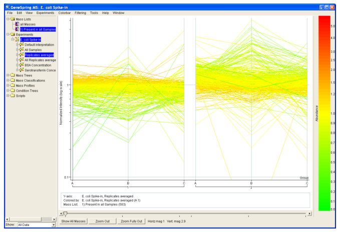
MFE
その他



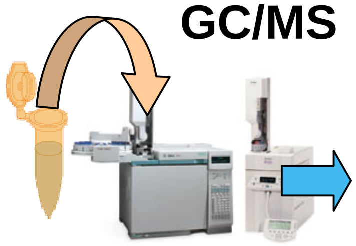
mzXML Files
*ファイル変換、データインポートに時間がかかります。
csv Files
*特定のフォーマットに書き換える必要があります。
MHD Files



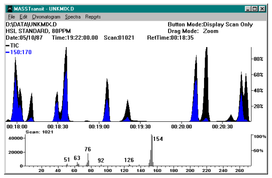
GeneSpring MS



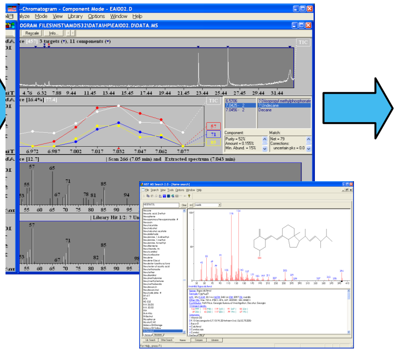
GC/MS



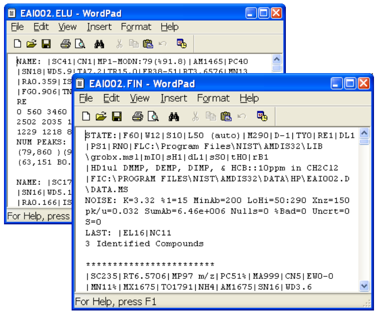
.D files
その他



AMDIS
NIST



ELU/FIN files



EAI002.ELU - WordPad

```
NAME: |SC41|CN1|MP1-MODN:79(%91.8)|AM1465|PC40
|SN18|WD5.9
|RA0.359|IS
|FG0.906|TN
RE
0 560 3460
2502 2035 1
1229 1218 8
NUM PEAKS:
(79,860 )(9
(63,151 BO.

NAME: |SC17
|SN16|WD5.1
|RA0.166|IS
```

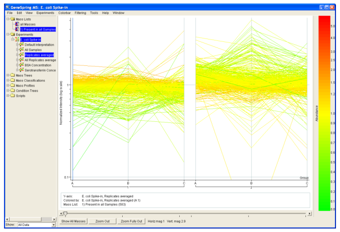
EAI002.FIN - WordPad

```
STATE:|F60|W12|S10|L50 (auto)|M290|D-1|TY0|RE1|DL1
|PS1|RN0|FLC:\Program Files\NIST\AMDIS32\LIB
\grobx.msl|mI0|sH1|dL1|sS0|tH0|rB1
|HD1ul DMMP, DEMP, DIMP, & HCB::10ppm in CH2Cl2
|FIC:\PROGRAM FILES\NIST\AMDIS32\DATA\HP\EAI002.D
\DATA.MS
NOISE: K=3.32 %1=15 MinAb=200 LoHi=50:290 Xnz=150
pk/u=0.032 SumAb=6.46e+006 Nulls=0 %Bad=0 Uncrt=0
S=0
LAST: |EL16|NC11
3 Identified Compounds

*************************
|SC235|RT6.5706|MP97 m/z|PC51%|MA999|CN5|EW0-0
|MN11%|MX1675|TO1791|NH4|AM1675|SN16|WD3.6
```

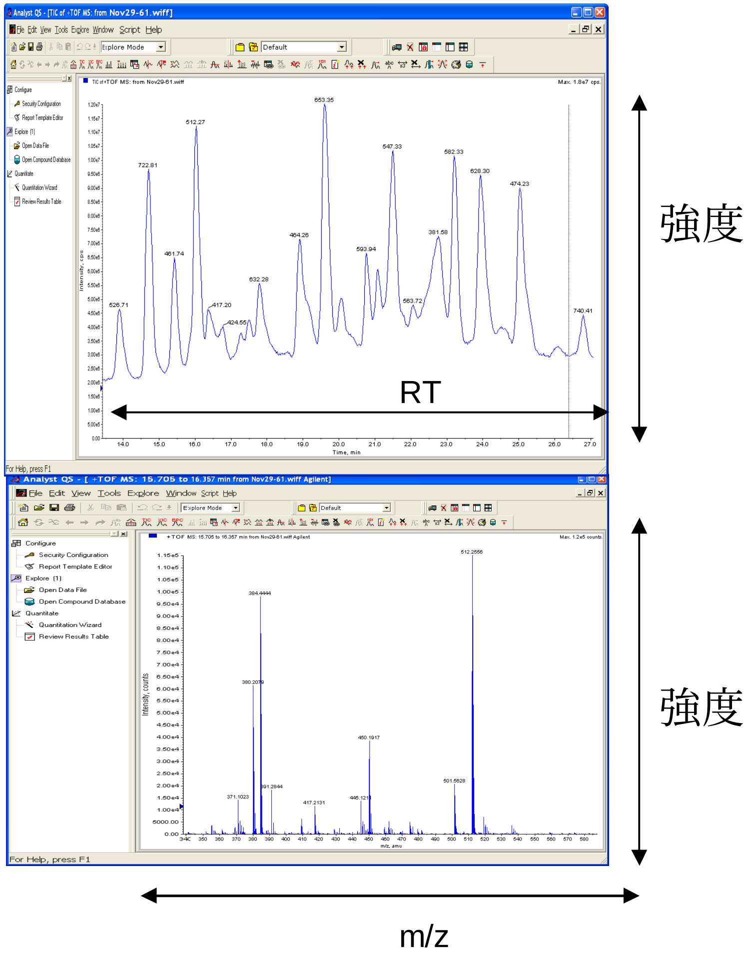
For Help, press F1

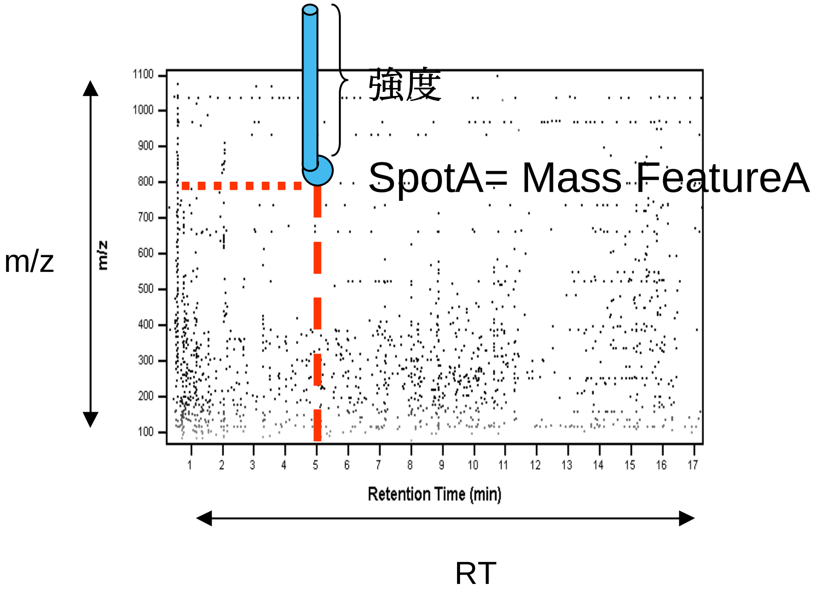
GeneSpring MS



CE/MSデータ解析も可能

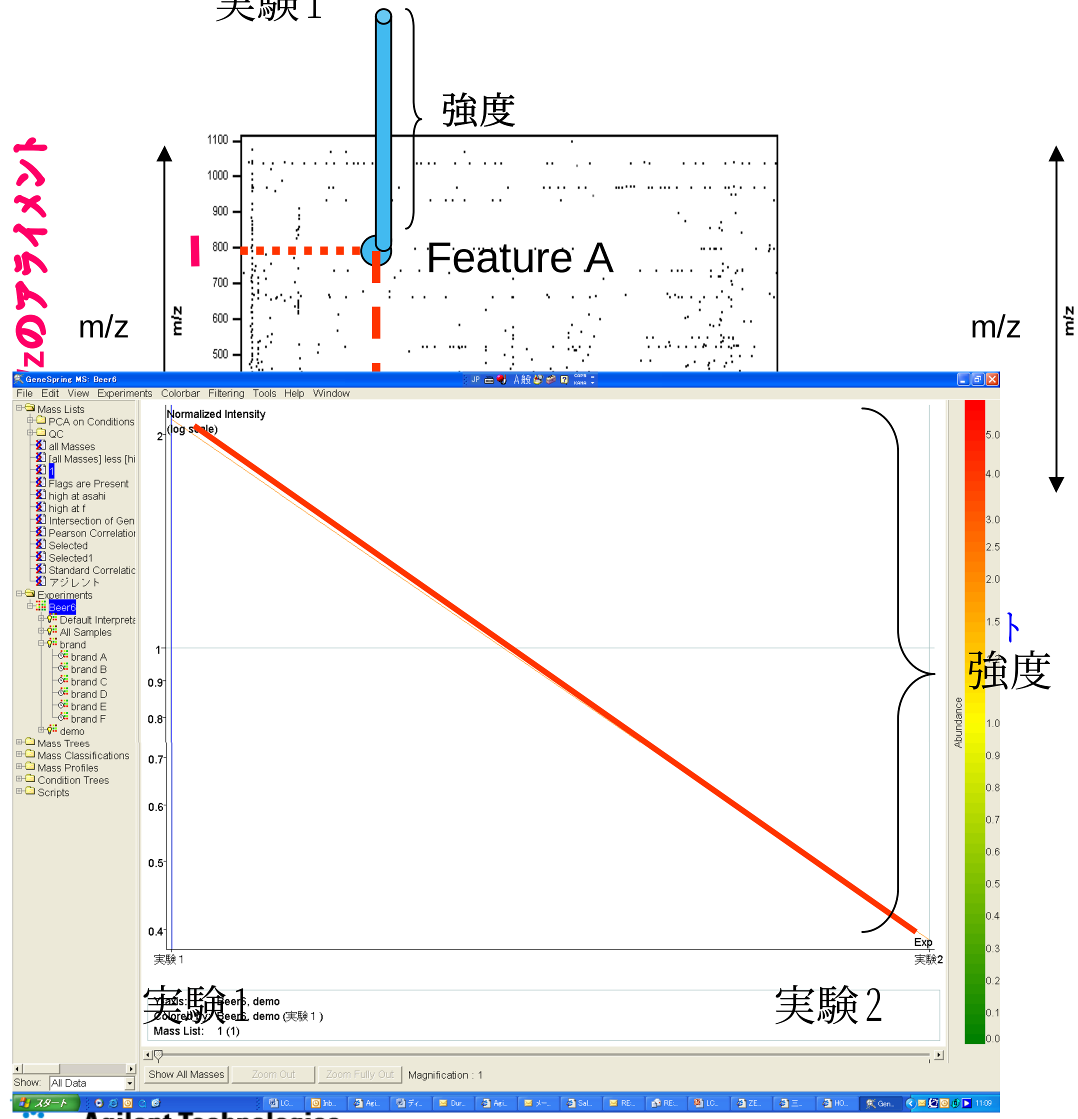
RT⇔ m/z ⇔ 強度の関係


Analyst QS - [TIC of +TOF MS: from Nov29-61.wiff]
RT
強度
Analyst QS - [+TOF MS: 15.705 to 16.357 min from Nov29-61.wiff Agilent]
強度
m/z

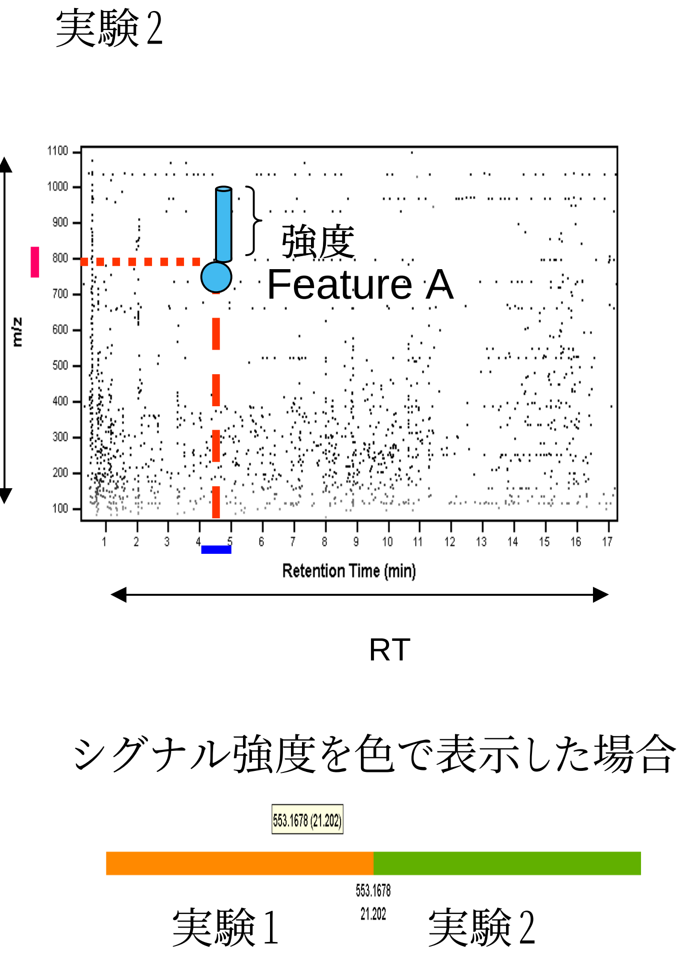

強度
SpotA= Mass FeatureA
m/z
m/z
Retention Time (min)
RT

2実験のアライメント→強度比較

zのアライメント

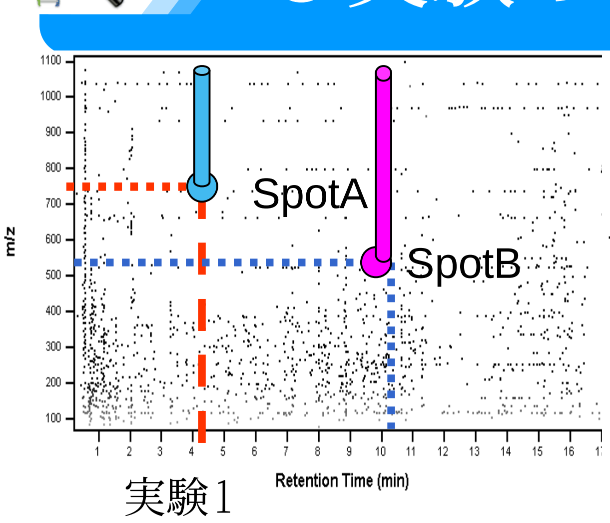


Y axis: Beer6, demo
Colored by: Beer6, demo (実験１)
Mass List: 1 (1)

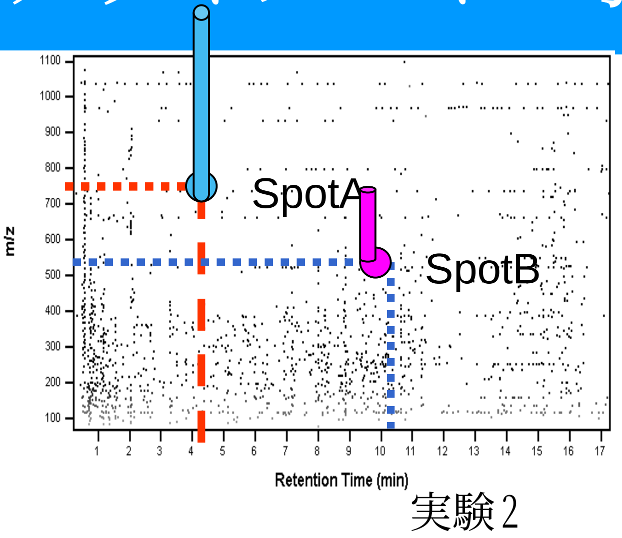


シグナル強度を色で表示した場合

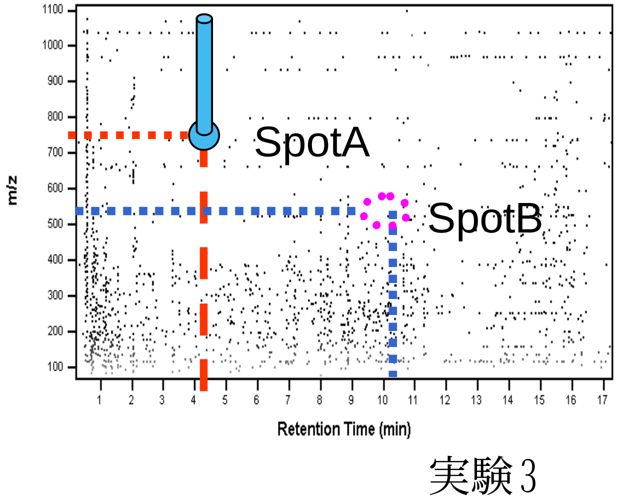
３実験のアライメント→強度比較



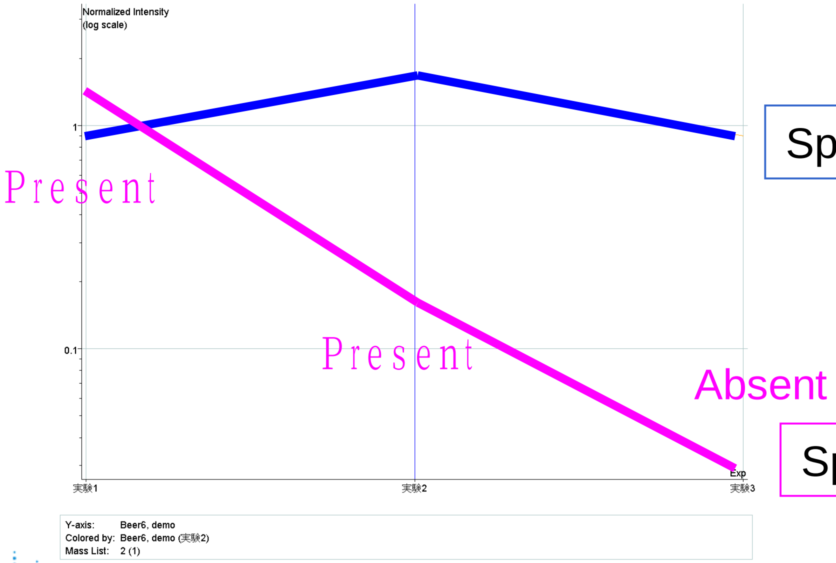
実験1



実験2

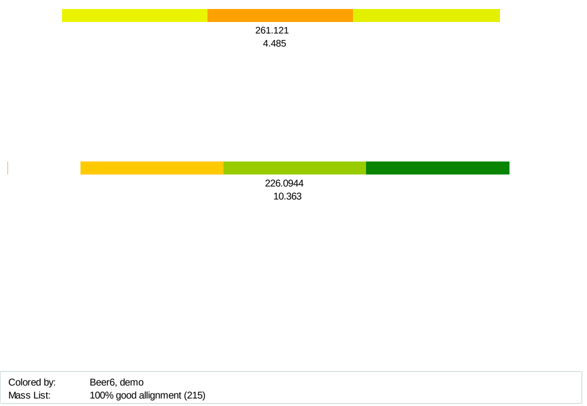


実験3



Spot　A のパターン

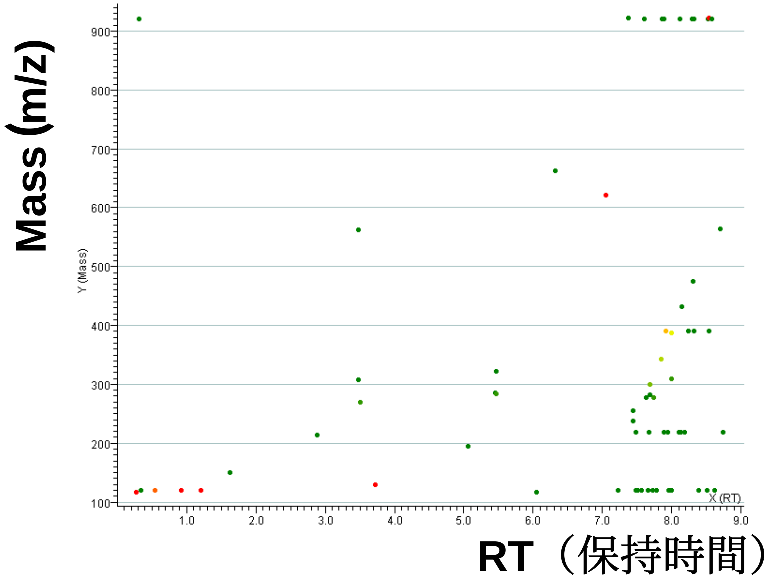
Spot Bのパターン



解析の際に注意する点
- 4成分の標準サンプルを用いた例

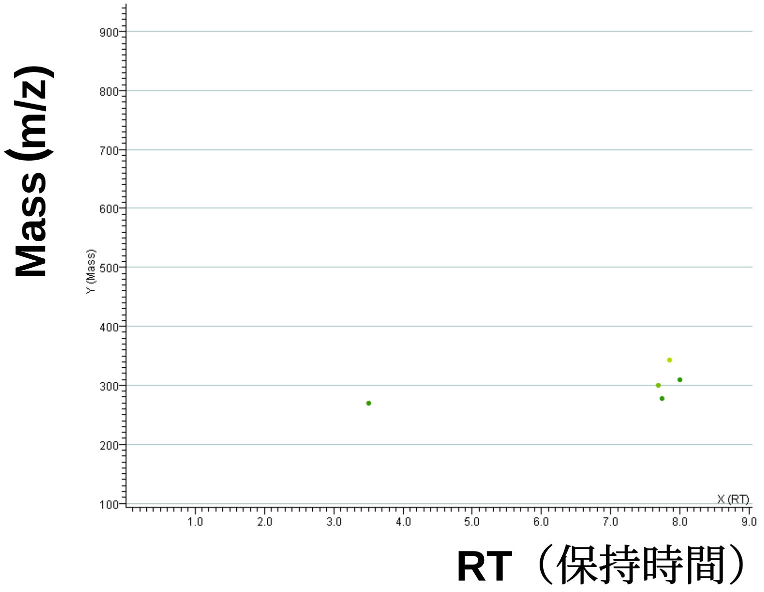
何も考えずにデータをインポートしてみると…

6実験のデータをインポートした結果
65個のMassが出現



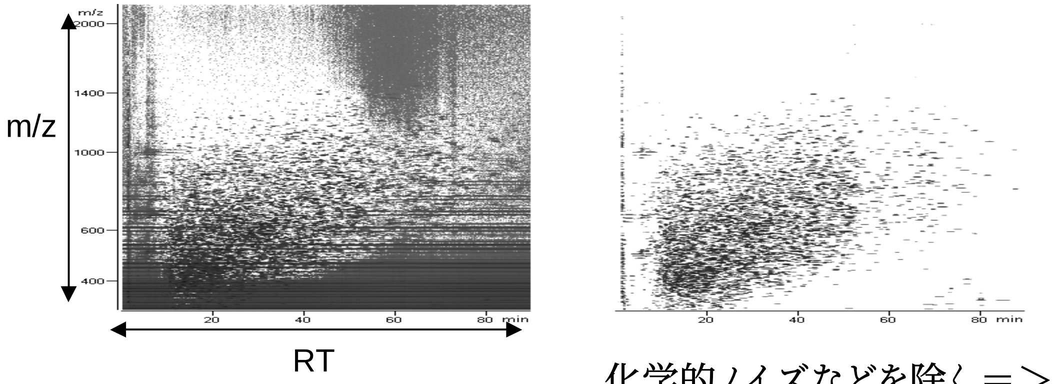
フィルタリング
（ノイズの除去？）

全てのサンプルで存在するMass
5個のMassが抽出



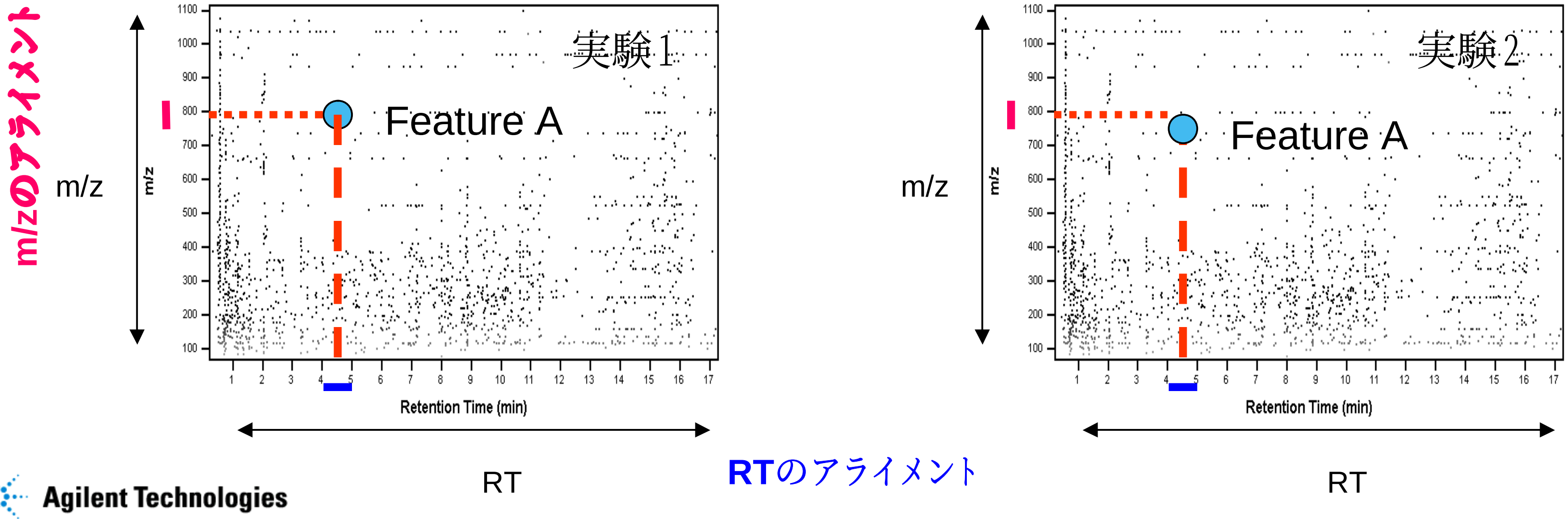
解析の際に注意する点
- 要因を考察

1. ノイズを**Mass**として捉えている？



化学的ノイズなどを除く ＝＞質の良いデータに絞込む

2.アライメントが完璧ではない？



解析の際に注意する点

- アライメントの難しさ

  - RT、m/zのずれをどこまで許容（Tolerance設定）

  - 測定されたノイズをMassとして認識しまう問題

- ノイズとなるMassを除いて解析をすることが必要

  - データインポート前の処理

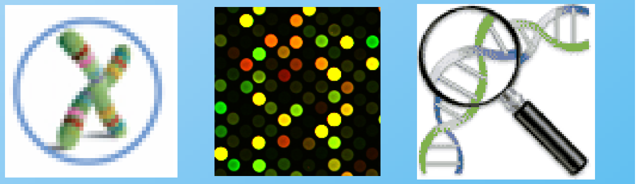
  - データインポート後にフィルタリング操作

    例：Intensityが高いMass、SN比が高いMass

※どちらかといえば、データインポート前での処理が推奨できる。

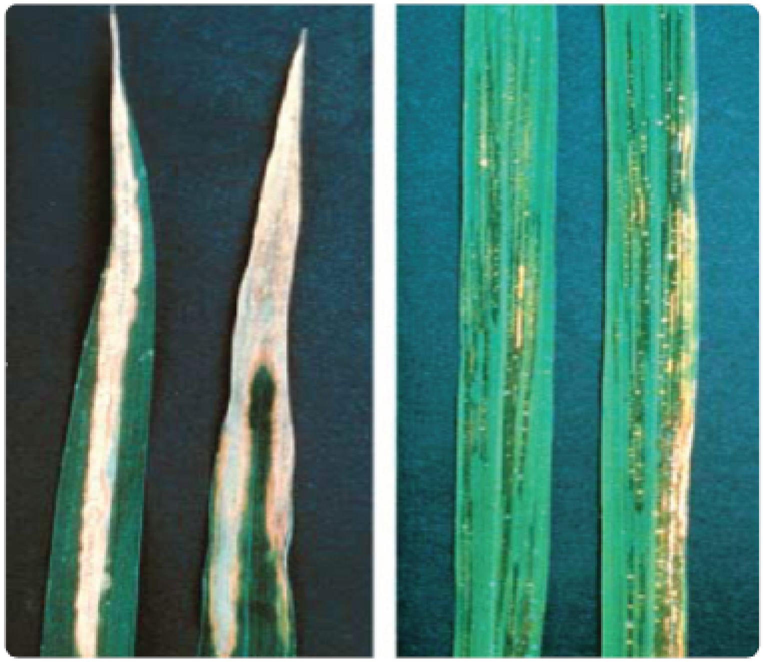
Agilent Technologies

BLBに感染したRiceの解析例
- 差分解析の例



イネの野生株とBLB耐性株

Bacterial Leaf Blight (BLB)
caused by *Xanthomonas oryzae pv. Oryzae* (*Xoo*)



Infected **Uninfected**

Xa21

- TP309に導入すると、BLB耐性を示す
- Leucine-rich repeat motifとserine-threonine kinase-like domainを持つ
- 細胞表面で病原体由来のリガンド（AvrXa21 peptide）を認識して、免疫反応を誘導？

TP309 **(WT)**

TP309-Xa21 **(TG, transgenic)**

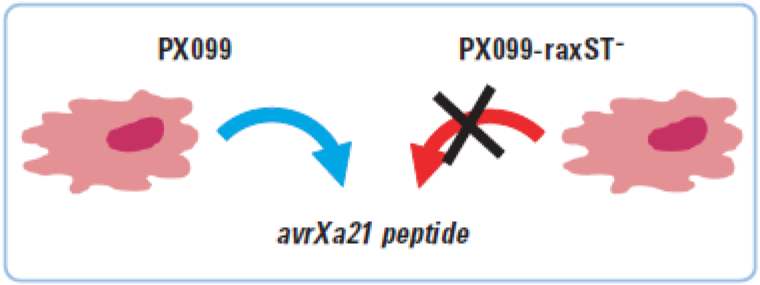
*Xoo*の野生株とノックアウト株

AvrXa21 peptide

- Xa21により認識される*Xoo*由来のペプチド
- BLB耐性のTP309-Xa21に存在するが、 感染はしない

raxST

- AvrXa21ペプチドの生産に必要な遺伝子
- Sulfotransferase-like protein



→

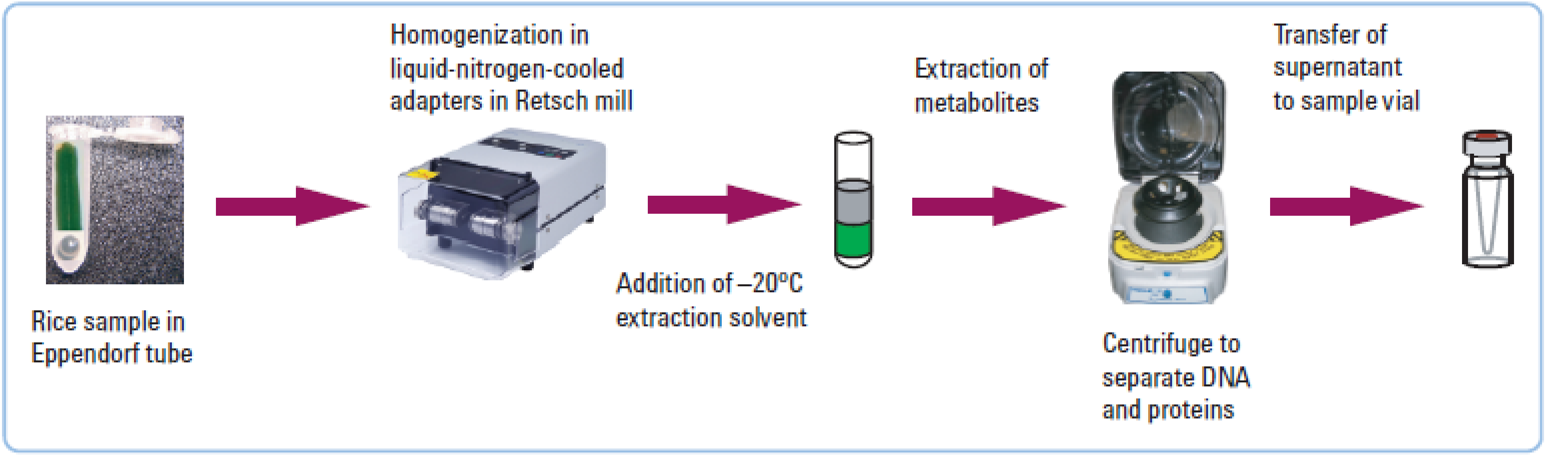
PXO99 **(WT)**

PXO99-raxST$^-$ **(KO)**

サンプルの概要

Condition (Class)	**TP309 (Wild-type)**		**TP309-Xa21 (Transgenic)**	
PXO99 (wild-type)	6	感染	6	耐性
Mock treatment control	6	Ctr	6	Ctr
No treatment (NT) control	6		6	
PXO99-raxST$^-$ (*raxST* knock-out)	NA	感染	6	感染

実験手順



Rice sample in Eppendorf tube

Homogenization in liquid-nitrogen-cooled adapters in Retsch mill

Addition of –20°C extraction solvent

Extraction of metabolites

Centrifuge to separate DNA and proteins

Transfer of supernatant to sample vial

Instrument Conditions – LC/TOF MS

LC Conditions

Column: ZORBAX SB-Aq column 2.1 x 150 mm, 3.5 µm

Mobile phase:
- A = 0.1% formic acid in water
- B = 0.1% formic acid in acetonitrile

Gradient:
- 2% B at 0 min
- 98% B at 46 min
- 98% B at 54.9 min
- 2% B at 55 min

MS stop time: 54.9 min

LC stop time: 55 min

Column temperature: 20°C

Flow rate: 0.3 mL/min

Injection volume: 2 µL + 3 sec flush

MS Conditions

Ionization mode: Electrospray

Ionization polarity: Positive ionization*

Drying gas flow: 10 L/min

Drying gas temperature: 250°C

Nebulizer pressure: 35 psi

Scan range: *m/z* 50–950

Fragmentor voltage: 170 V

Capillary voltage: 4000 V

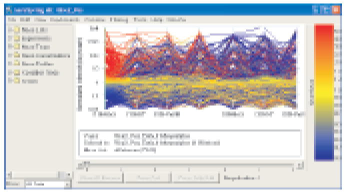
Reference masses: *m/z* 121, 922

Reference mass flow: 10 µL/min

*Both positive- and negative-ionization data were successfully acquired, but this note deals only with processing of the positive ion data.

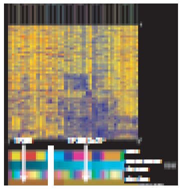
GeneSpring MSによる解析（概要）

1. アラインメントとノーマライズ



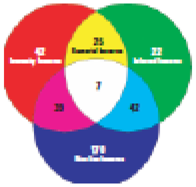
Align and normalize features

2. 階層型クラスタリングによる、データの品質チェック



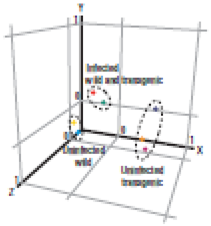
Hierarchical clustering to check for reproducibility of biological replicates

3. 1-/2-way ANOVA による、統計学的有意な量的変化を示すFeatureの抽出



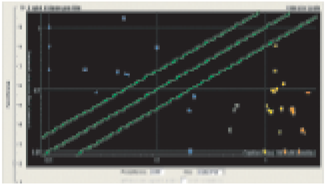
Analysis of variance to find features with statistically significant differences between classes

4. PCAによる、分割可能なクラスの探索



Principle component analysis of significantly different features representing class differences

5. fold changeによる量的変化の確認

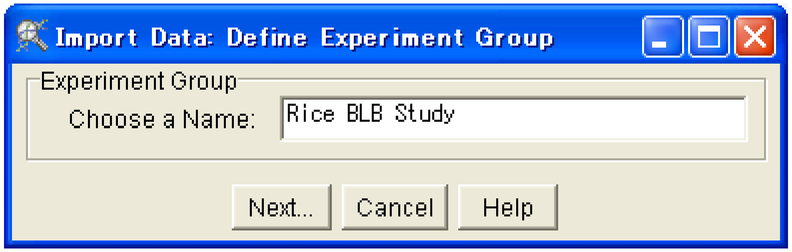


Fold-change filtering to select the most statistically significant features

6. 抽出したターゲット代謝物をDBで検索

Create inclusion lists for database searching and MS/MS analysis

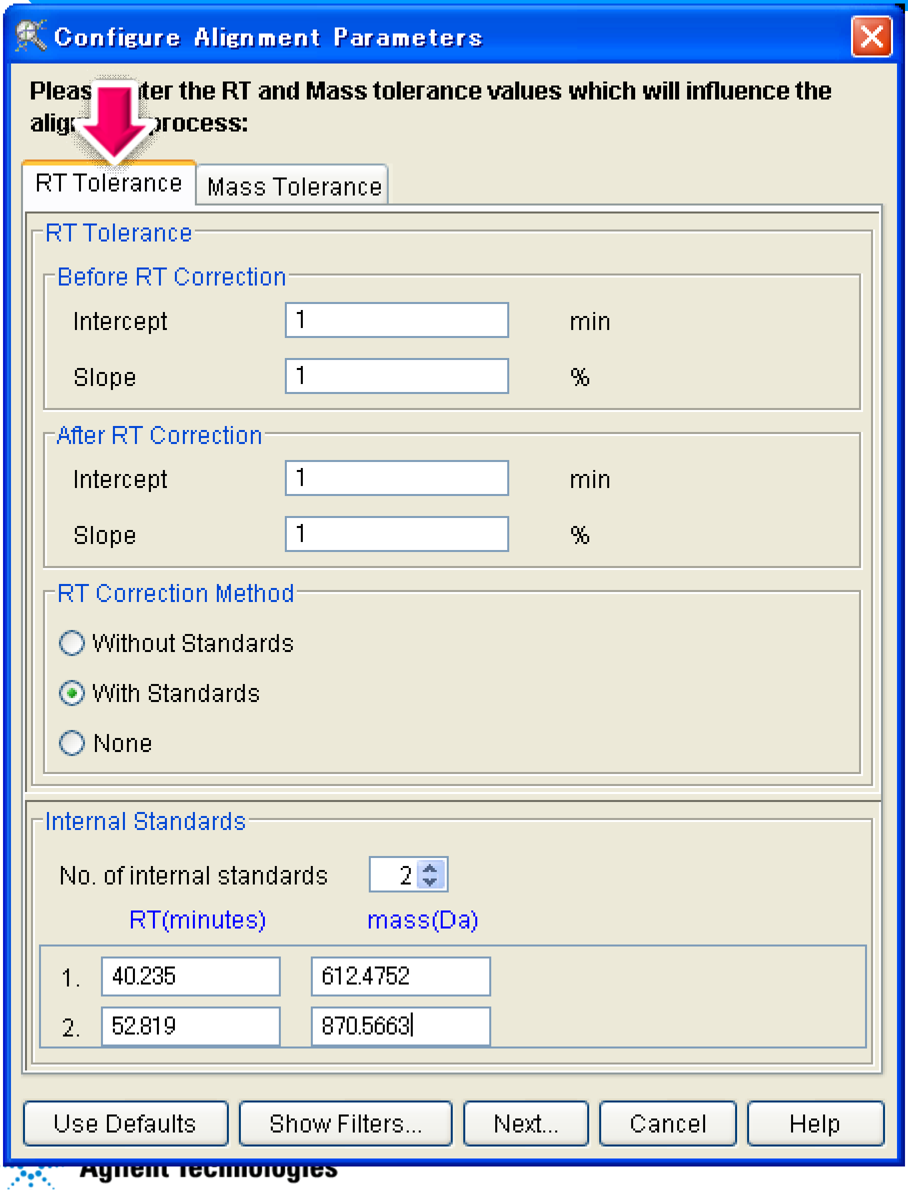
実験データのインポート



Experiment Group

Choose a Name: Rice BLB Study

アラインメントの設定



Please enter the RT and Mass tolerance values which will influence the alignment process:

RT Tolerance

Before RT Correction

Intercept	1	min
Slope	1	%

After RT Correction

Intercept	1	min
Slope	1	%

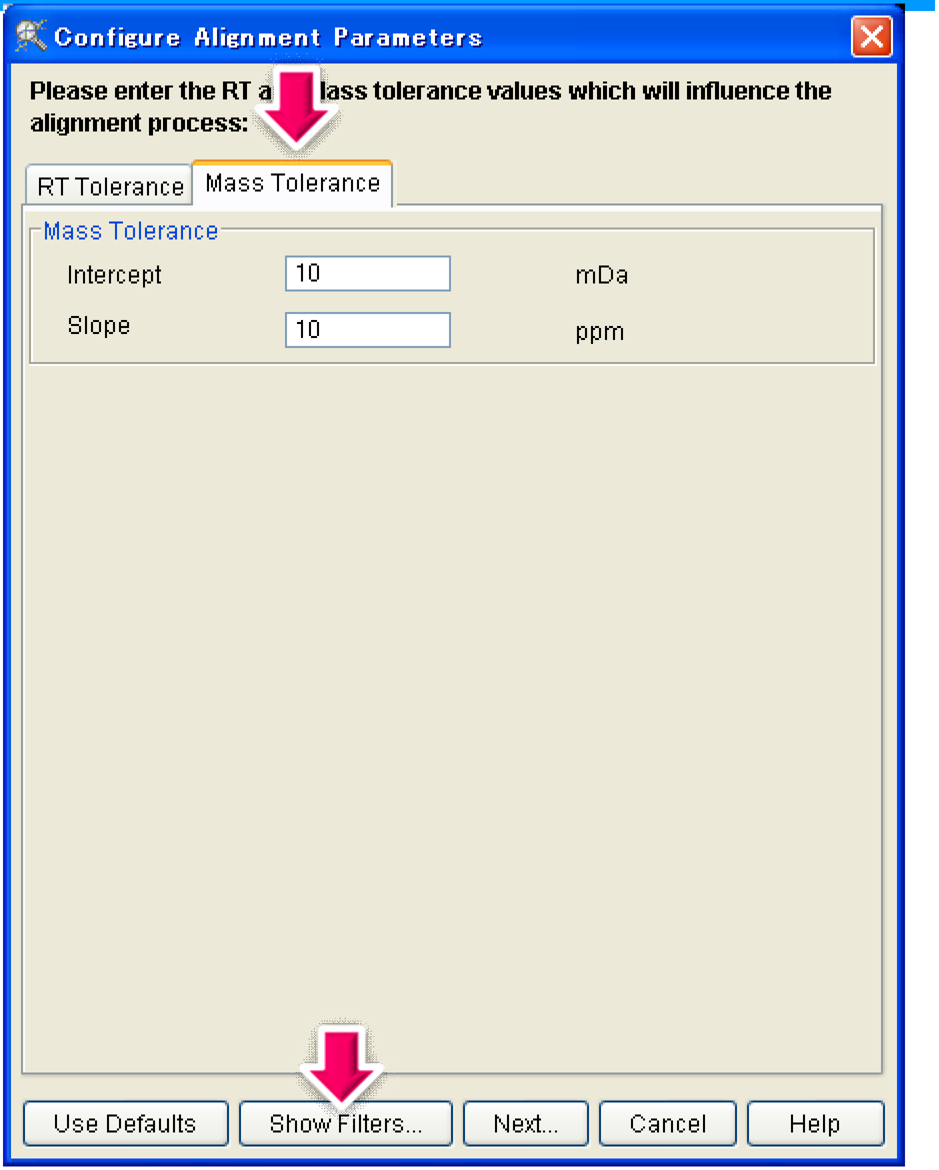
RT Correction Method

○ Without Standards
● With Standards
○ None

Internal Standards

No. of internal standards: 2

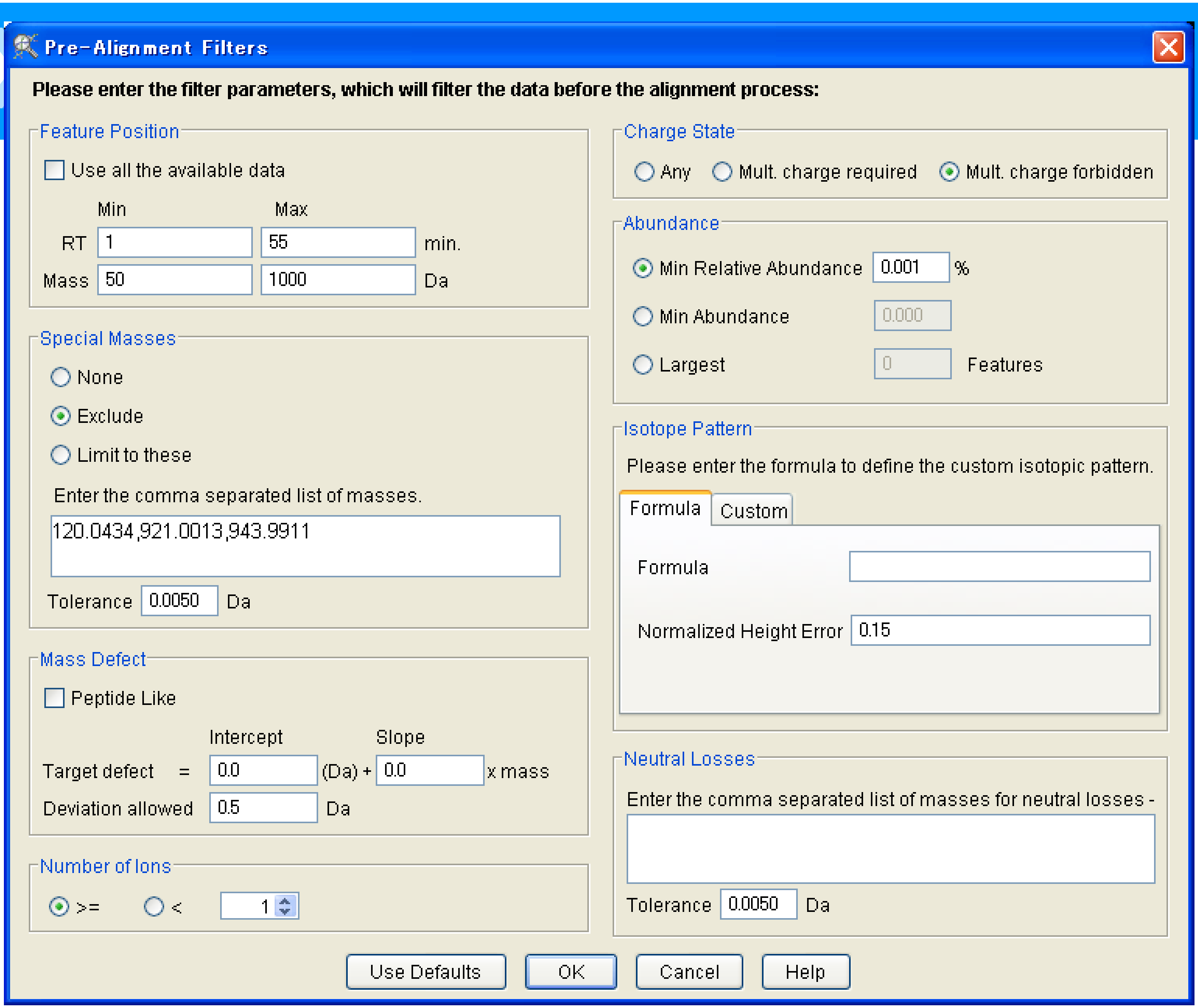
	RT(minutes)	mass(Da)
1.	40.235	612.4752
2.	52.819	870.5663



Please enter the RT and Mass tolerance values which will influence the alignment process:

Mass Tolerance

Intercept	10	mDa
Slope	10	ppm



Pre-Alignment Filters

Please enter the filter parameters, which will filter the data before the alignment process:

Feature Position

☐ Use all the available data

	Min	Max	
RT	1	55	min.
Mass	50	1000	Da

Special Masses

○ None

● Exclude

○ Limit to these

Enter the comma separated list of masses.

120.0434,921.0013,943.9911

Tolerance 0.0050 Da

Mass Defect

☐ Peptide Like

		Intercept		Slope	
Target defect	=	0.0	(Da) +	0.0	x mass
Deviation allowed		0.5	Da		

Number of Ions

● >= ○ < 1

Charge State

○ Any ○ Mult. charge required ● Mult. charge forbidden

Abundance

● Min Relative Abundance 0.001 %

○ Min Abundance 0.000

○ Largest 0 Features

Isotope Pattern

Please enter the formula to define the custom isotopic pattern.

Formula

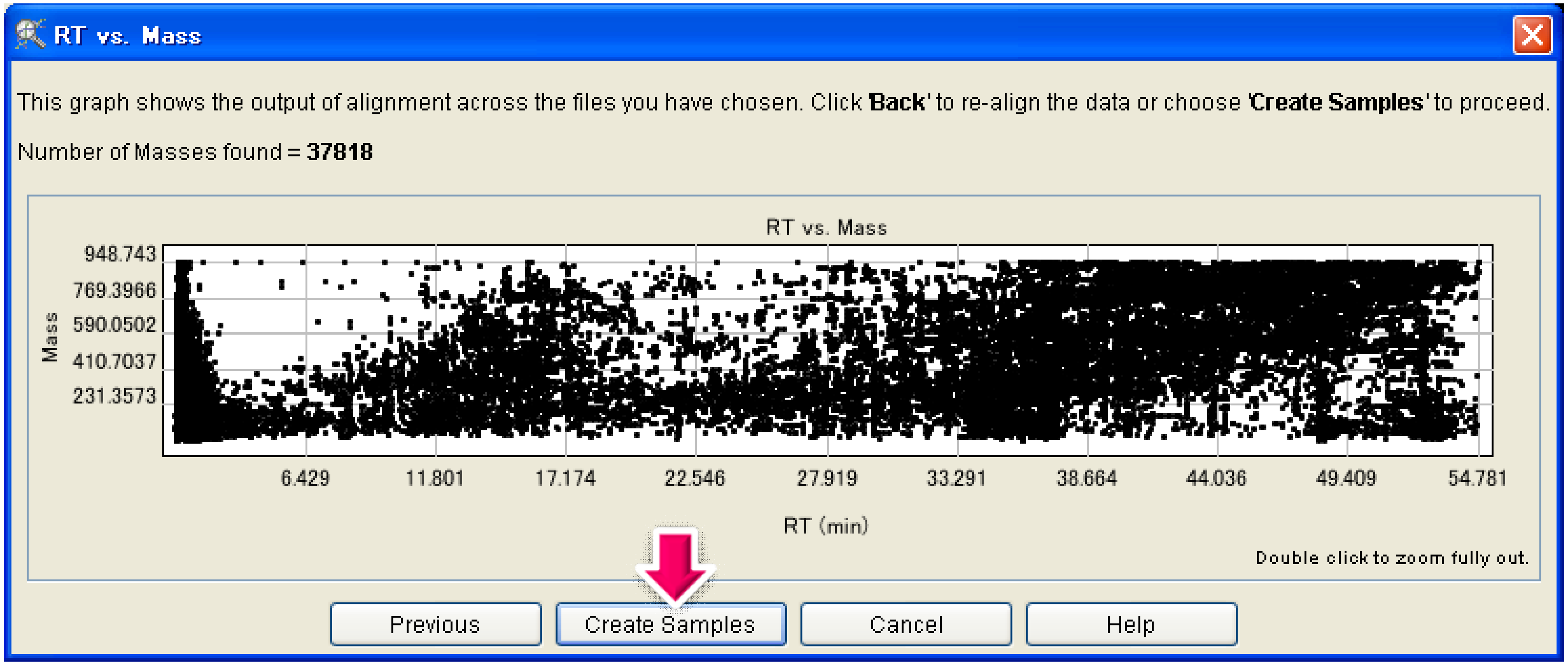
Normalized Height Error 0.15

Neutral Losses

Enter the comma separated list of masses for neutral losses -

Tolerance 0.0050 Da

アラインメントされたピーク

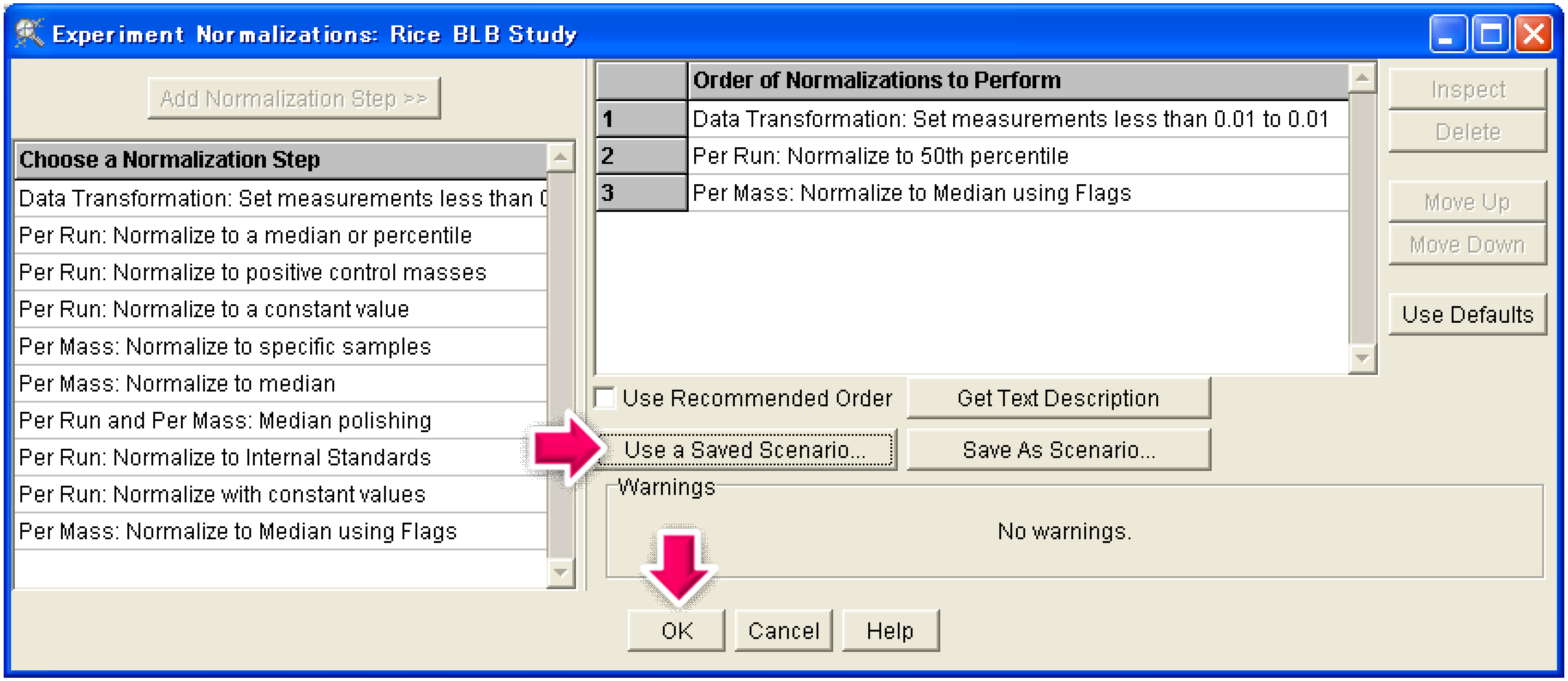


RT vs. Mass

This graph shows the output of alignment across the files you have chosen. Click **'Back'** to re-align the data or choose **'Create Samples'** to proceed.

Number of Masses found = **37818**

ノーマライズ（正規化）



Experiment Normalizations: Rice BLB Study

Choose a Normalization Step

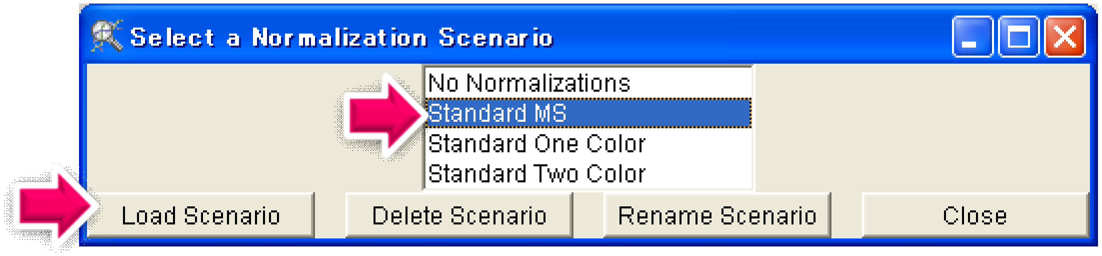
- Data Transformation: Set measurements less than 0
- Per Run: Normalize to a median or percentile
- Per Run: Normalize to positive control masses
- Per Run: Normalize to a constant value
- Per Mass: Normalize to specific samples
- Per Mass: Normalize to median
- Per Run and Per Mass: Median polishing
- Per Run: Normalize to Internal Standards
- Per Run: Normalize with constant values
- Per Mass: Normalize to Median using Flags

	Order of Normalizations to Perform
1	Data Transformation: Set measurements less than 0.01 to 0.01
2	Per Run: Normalize to 50th percentile
3	Per Mass: Normalize to Median using Flags

☐ Use Recommended Order

Warnings

No warnings.

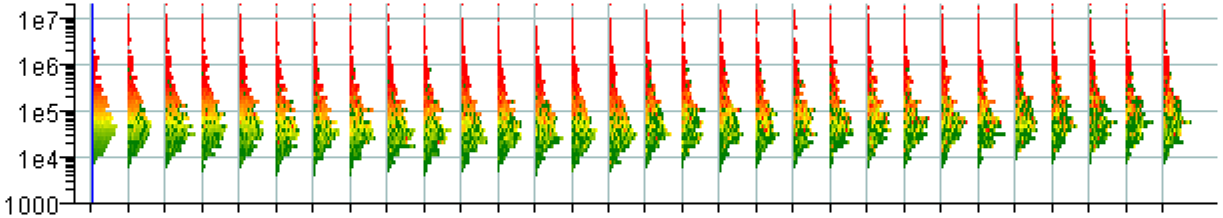


Select a Normalization Scenario

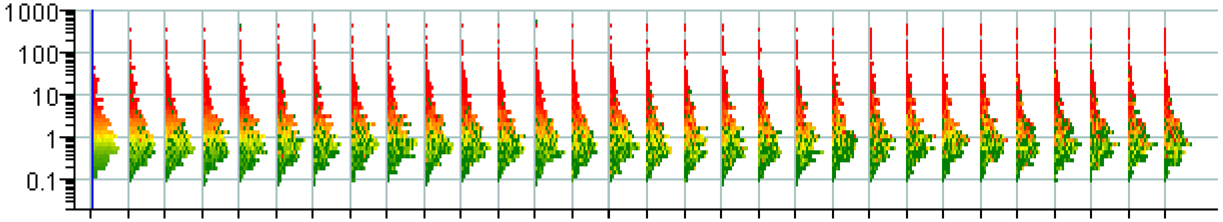
- No Normalizations
- Standard MS
- Standard One Color
- Standard Two Color

ノーマライズの意味

Raw data
(ノーマライズ前)

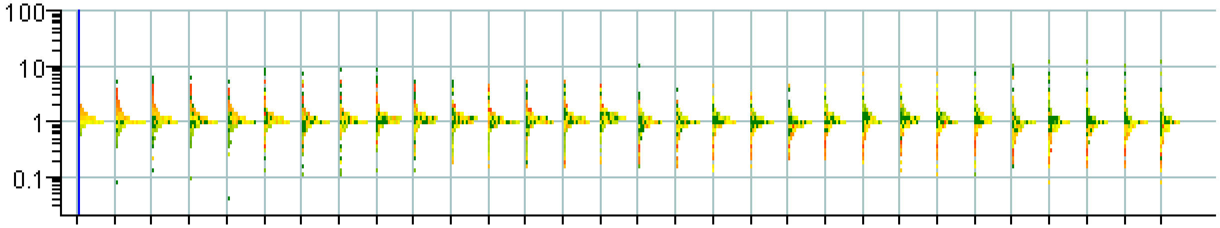


Per Run Normalize



実験ごとのばらつきをキャンセル

Per Run Normalize
+ Per Mass Normalize



Abundanceの変化を強調

Normalized値が1付近 = Abundanceの変化のまん中くらい

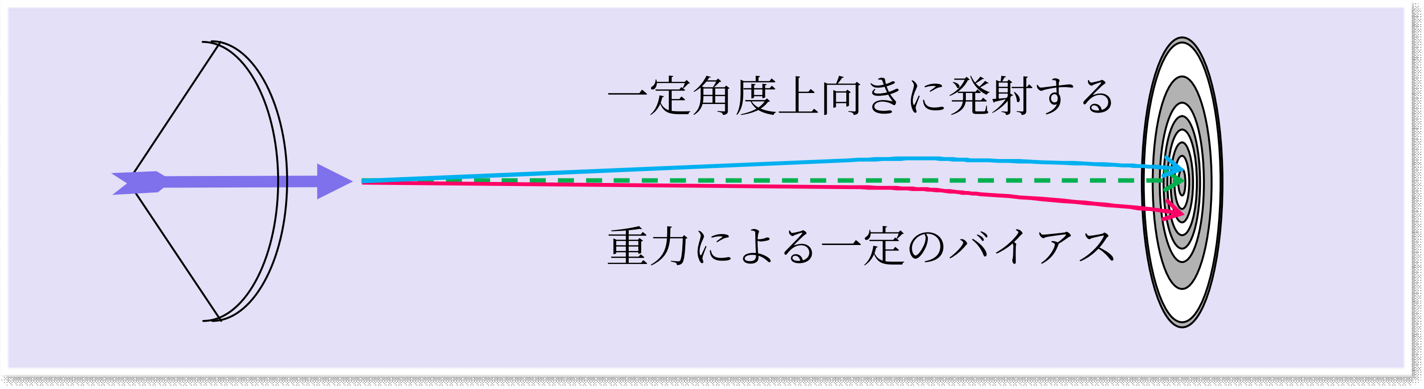
tips

バイアスの補正（ノーマライズ）

- 正規化（normalization）とは、データセットを相互比較できるように変換する操作
- 正規化の一般的な前提
  - 測定誤差は線形で系統的なものと考える

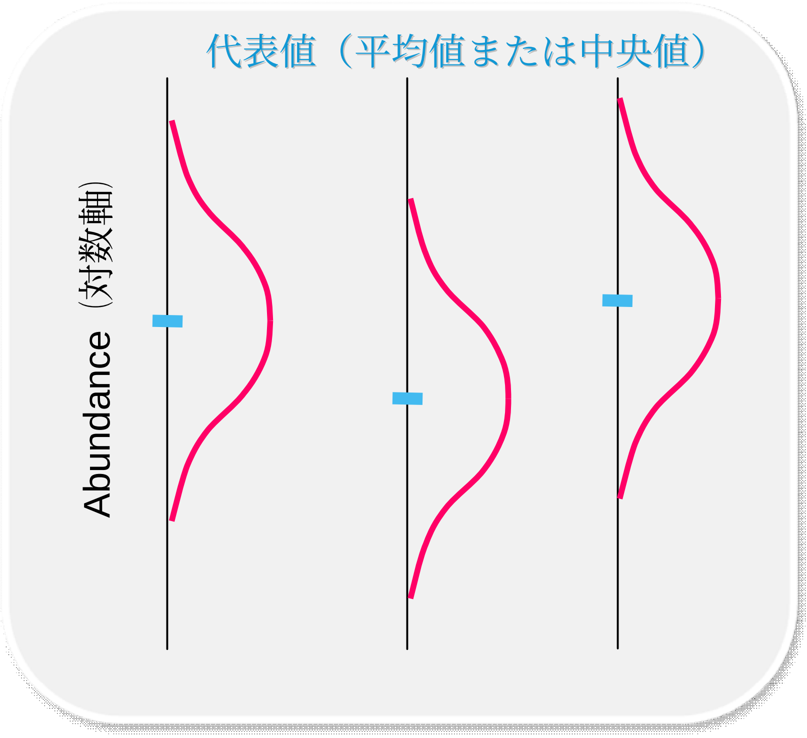
tips

線形な系統誤差と補正

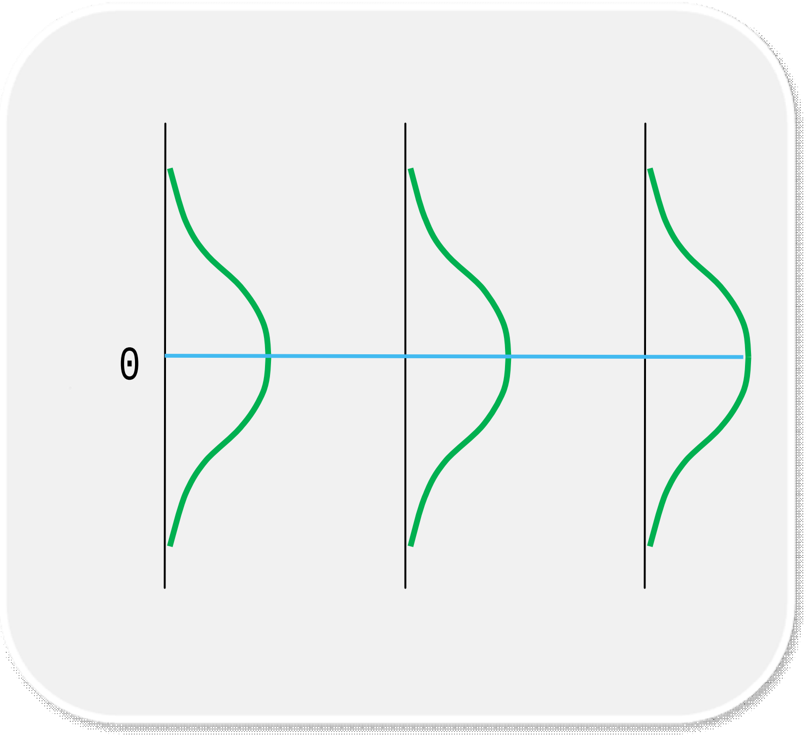


偶然誤差のみ　　偶然誤差＋線形の系統誤差

Per Run Normalization

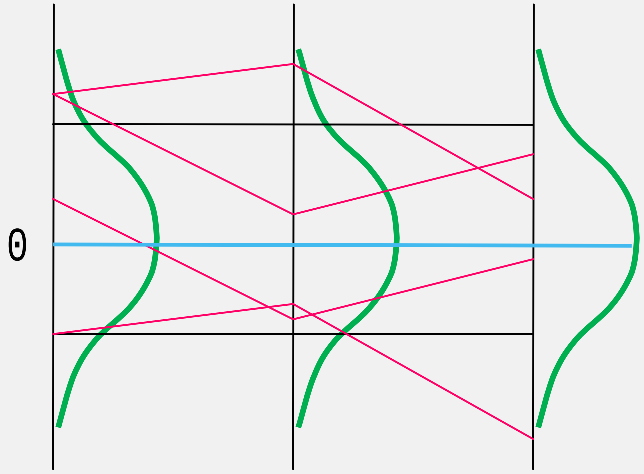


補正前

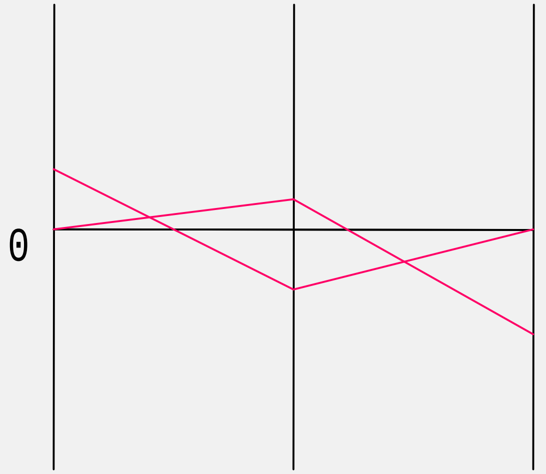


Per Run Normalization後

Per Mass Normalization



Abundanceが変化しない成分
Abundanceが変化する成分



Abundanceの変化が明確に見えるようになる。

Per Run Normalization

Per Run Normalization

\+ Per Mass Normalization

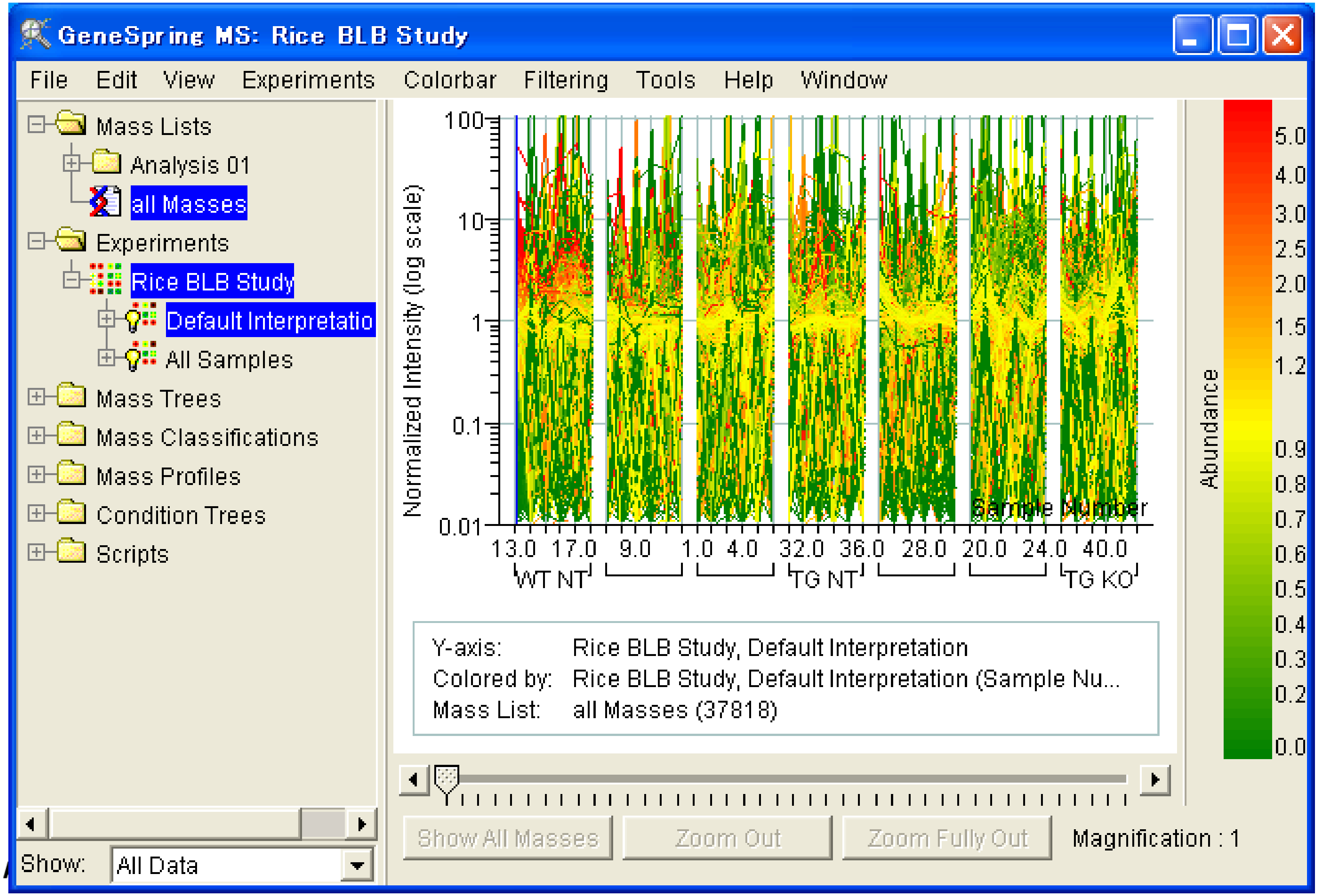
tips

実験データの属性情報

File	Strain	Infection	Status
1A	WT	WT	Infected
1B	WT	Mock	Uninfected
1C	WT	NT	Uninfected
2A	TG	WT	Resistant
2B	TG	Mock	Uninfected
2C	TG	NT	Uninfected
2D	TG	KO	Infected

※それぞれ6レプリケートを測定
（7 x 6 ＝42サンプルを測定）

それぞれのサンプルでの再現性を確認するグラフ

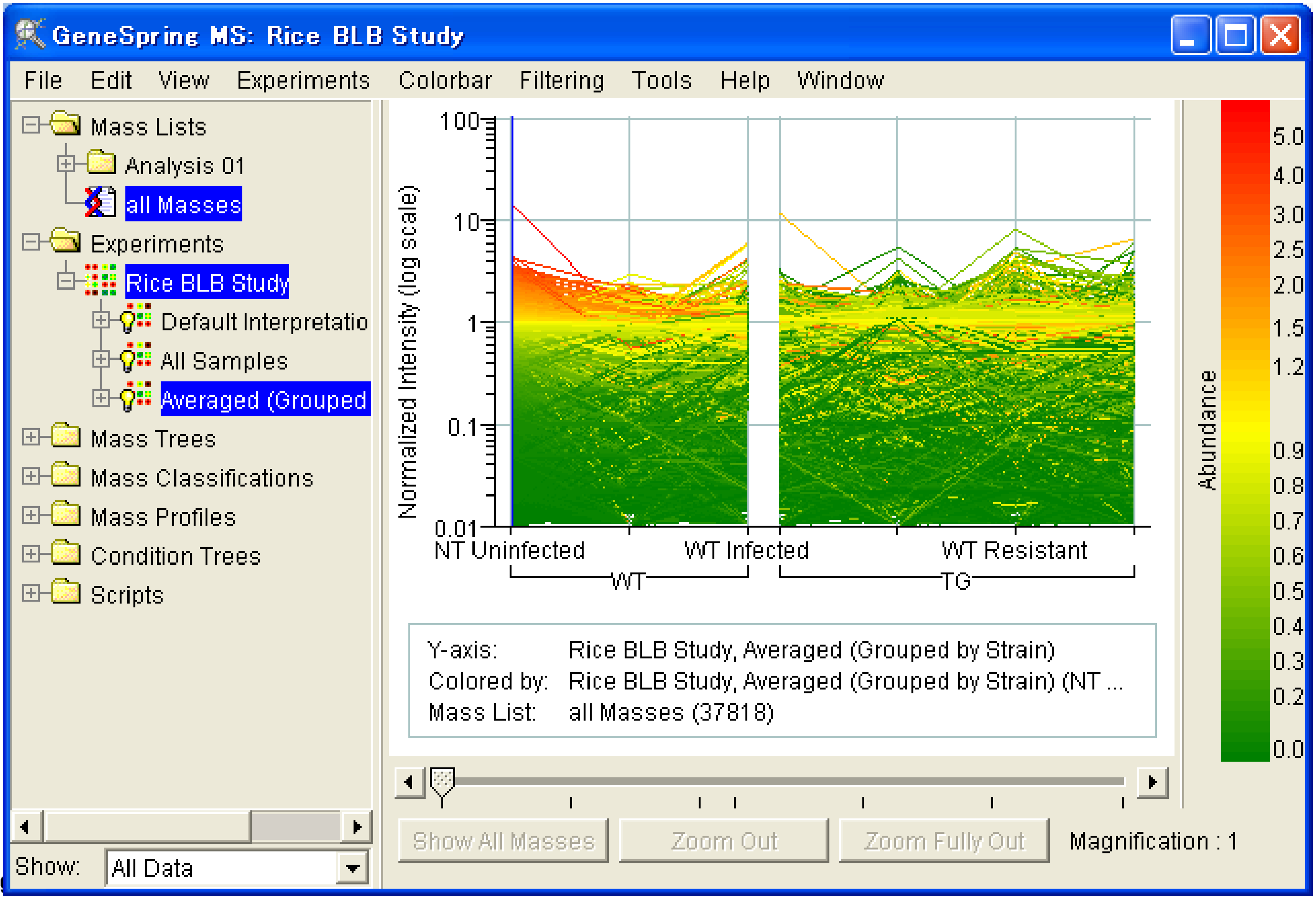


Y-axis: Rice BLB Study, Default Interpretation
Colored by: Rice BLB Study, Default Interpretation (Sample Nu...
Mass List: all Masses (37818)

Magnification : 1

Show: All Data

Strainのパラメーターを考慮したグラフ

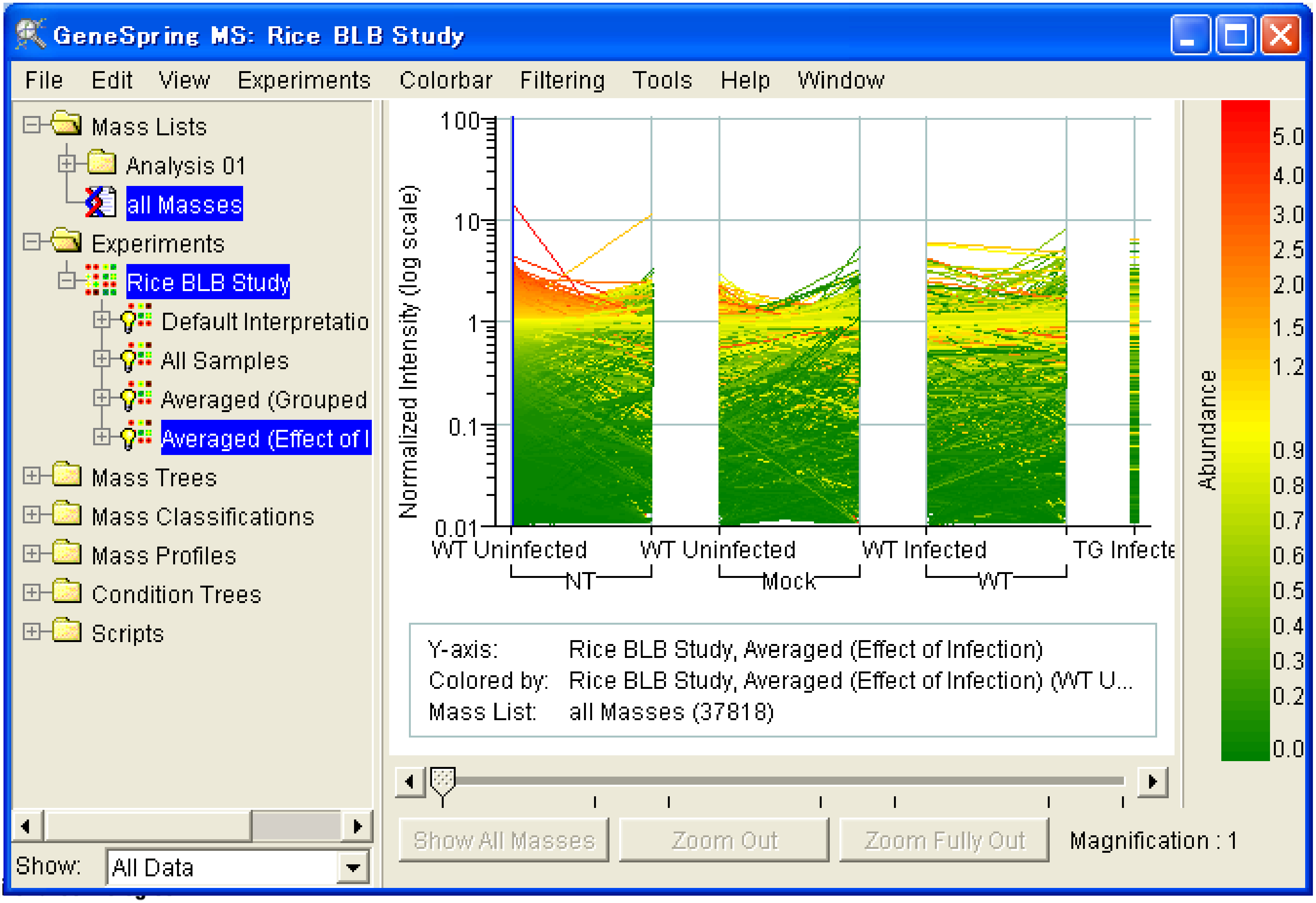


Y-axis: Rice BLB Study, Averaged (Grouped by Strain)
Colored by: Rice BLB Study, Averaged (Grouped by Strain) (NT ...
Mass List: all Masses (37818)

Show: All Data

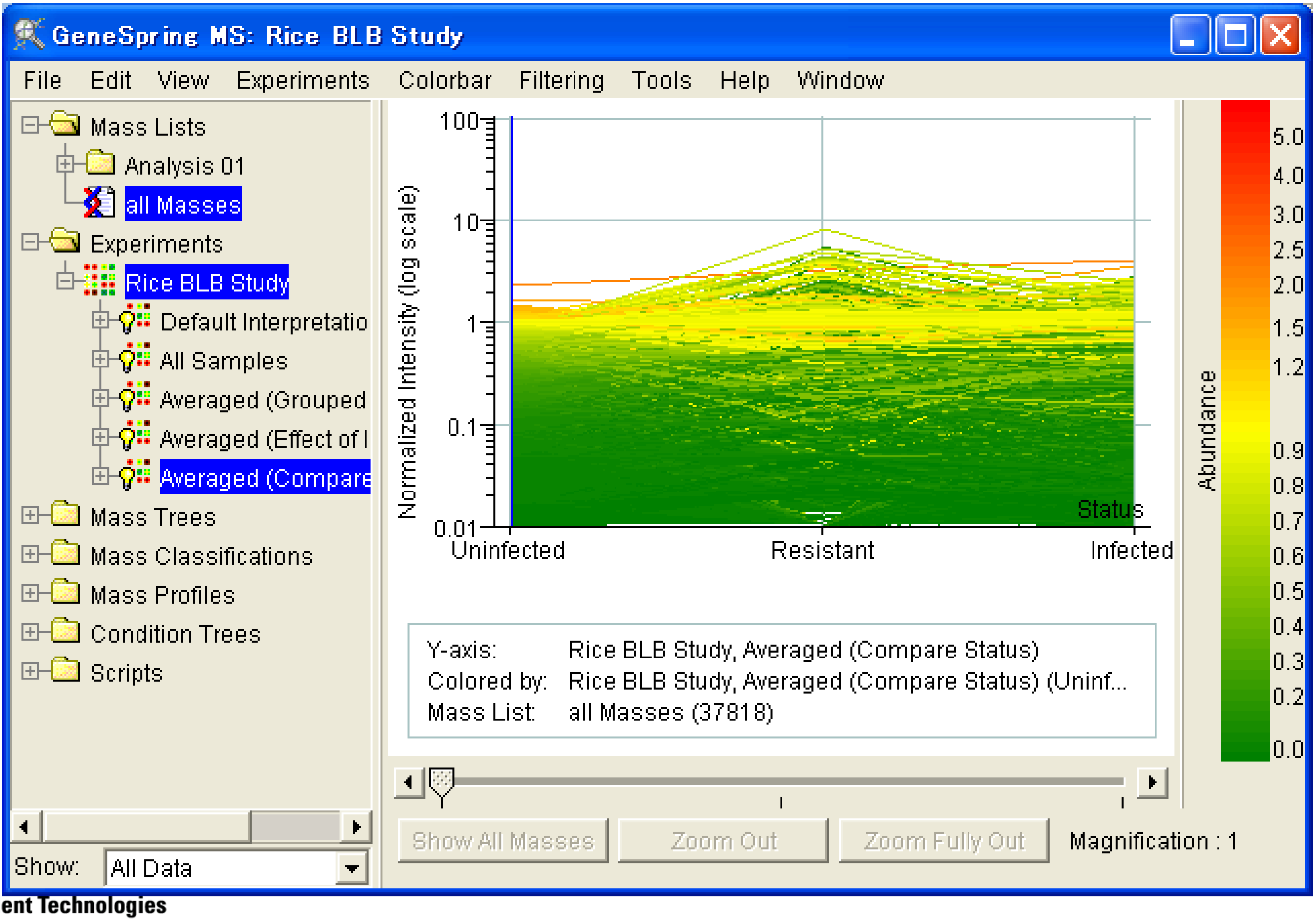
Magnification : 1

Infectionのパラメーターを考慮したグラフ



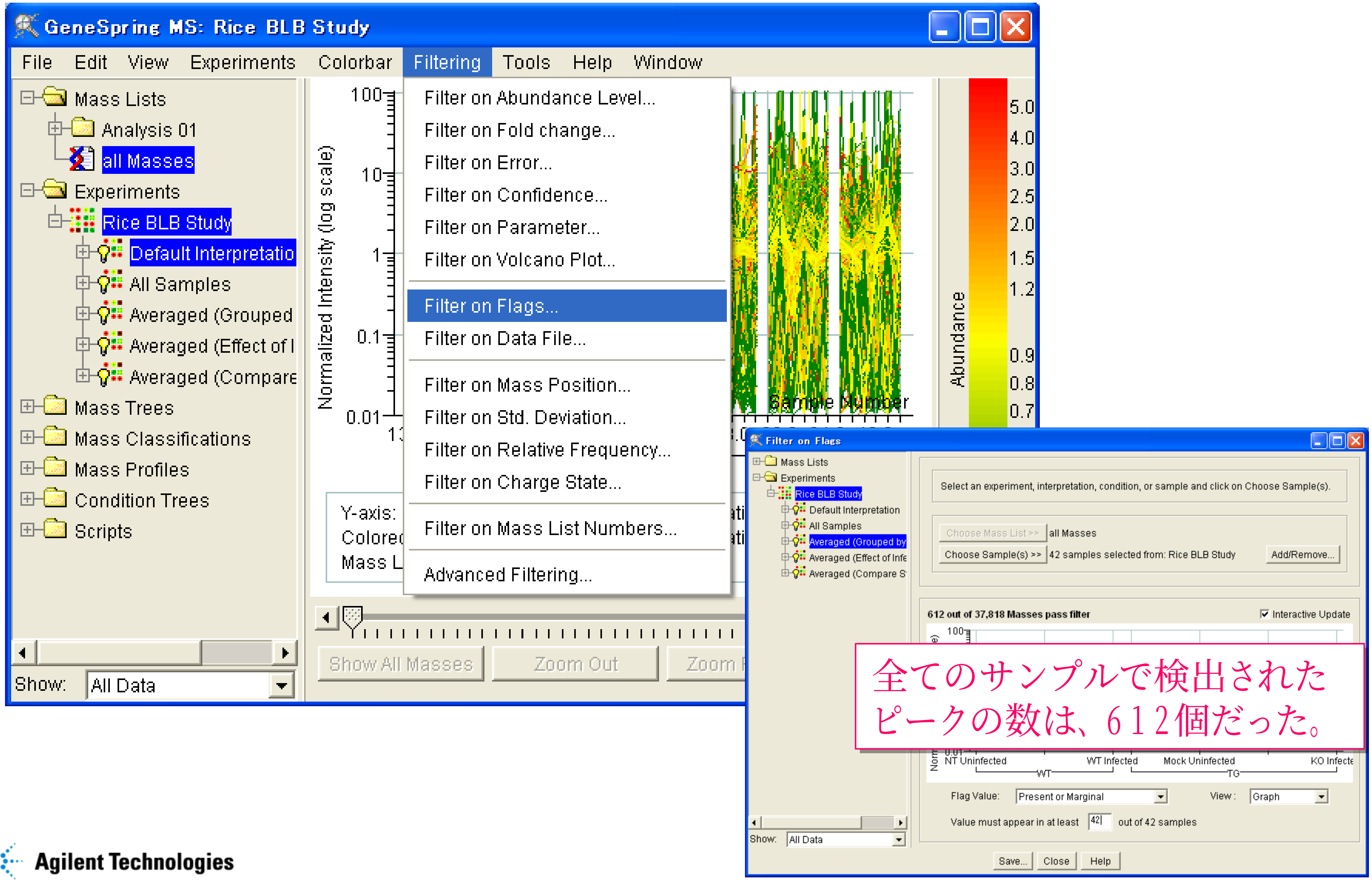
Y-axis: Rice BLB Study, Averaged (Effect of Infection)
Colored by: Rice BLB Study, Averaged (Effect of Infection) (WT U...
Mass List: all Masses (37818)

Statusのパラメーターを考慮したグラフ



Y-axis: Rice BLB Study, Averaged (Compare Status)
Colored by: Rice BLB Study, Averaged (Compare Status) (Uninf...
Mass List: all Masses (37818)

アラインメントができている成分を抽出



Select an experiment, interpretation, condition, or sample and click on Choose Sample(s).

Choose Mass List >> all Masses

Choose Sample(s) >> 42 samples selected from: Rice BLB Study

612 out of 37,818 Masses pass filter

☑ Interactive Update

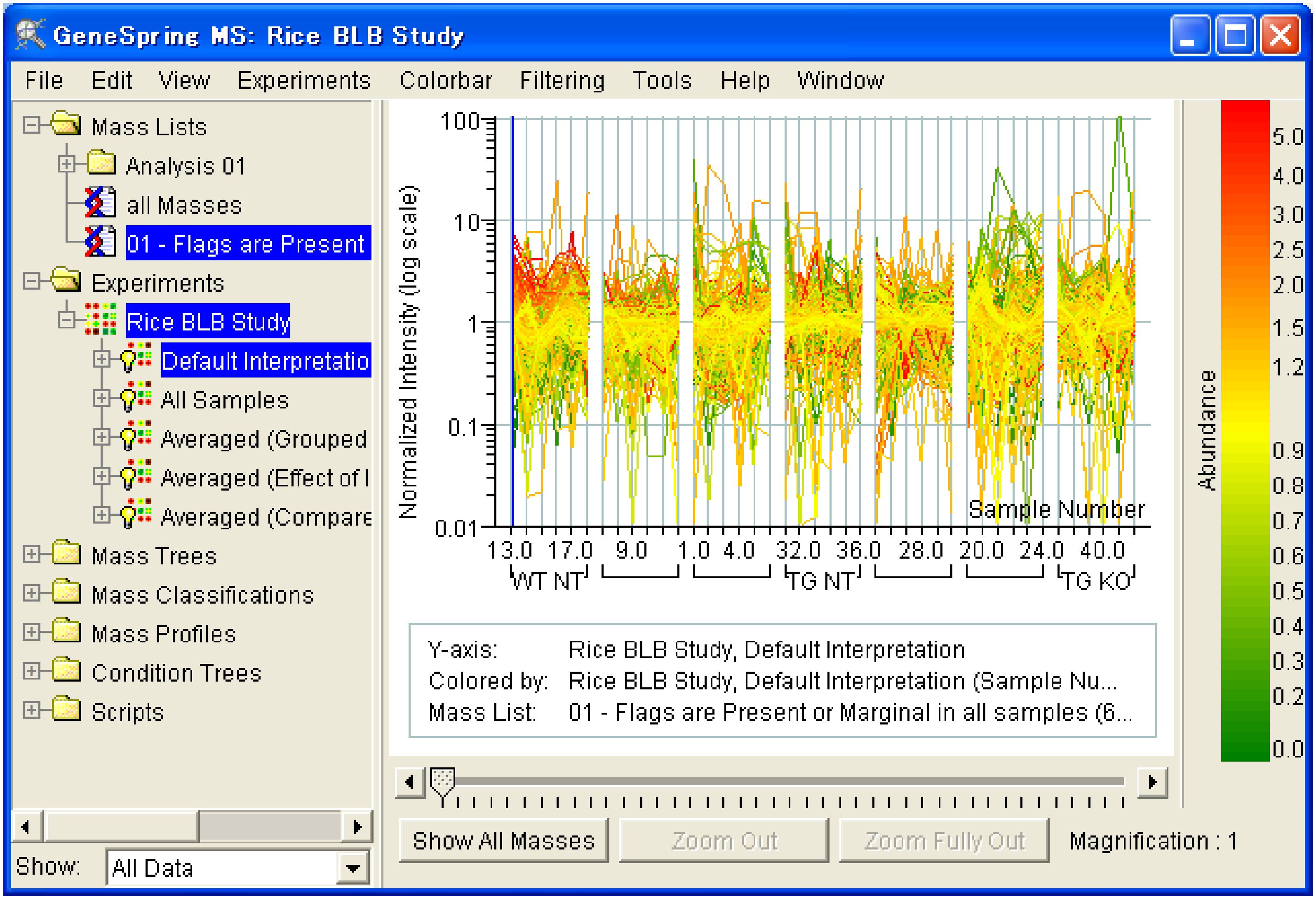
Flag Value: Present or Marginal

View: Graph

Value must appear in at least 42 out of 42 samples

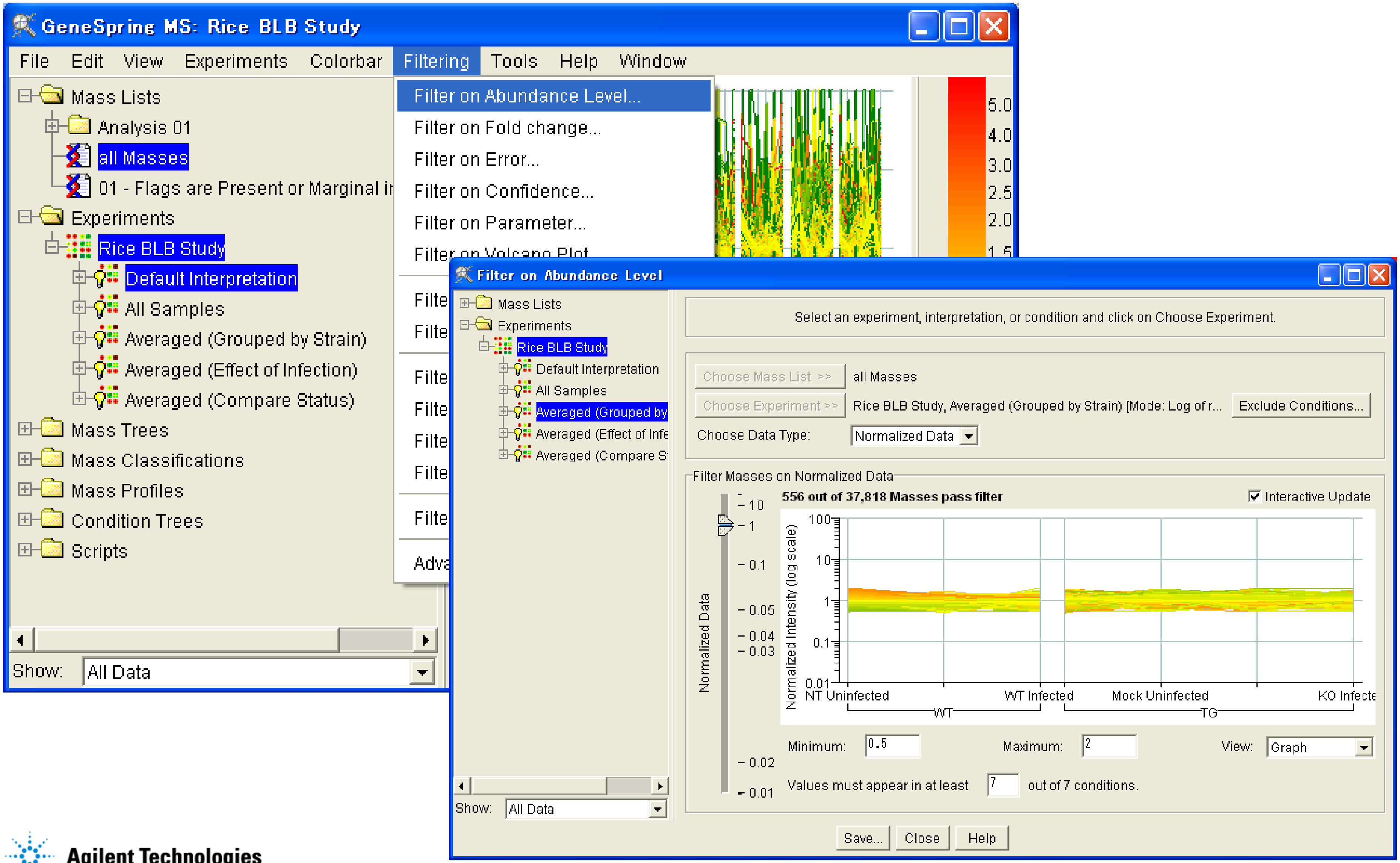
全てのサンプルで検出されたピークの数は、612個だった。

確実にアラインメントされた成分



Y-axis: Rice BLB Study, Default Interpretation
Colored by: Rice BLB Study, Default Interpretation (Sample Nu...
Mass List: 01 - Flags are Present or Marginal in all samples (6...

Abundance Level の変化しない成分を除去



Select an experiment, interpretation, or condition and click on Choose Experiment.

all Masses

Rice BLB Study, Averaged (Grouped by Strain) [Mode: Log of r...

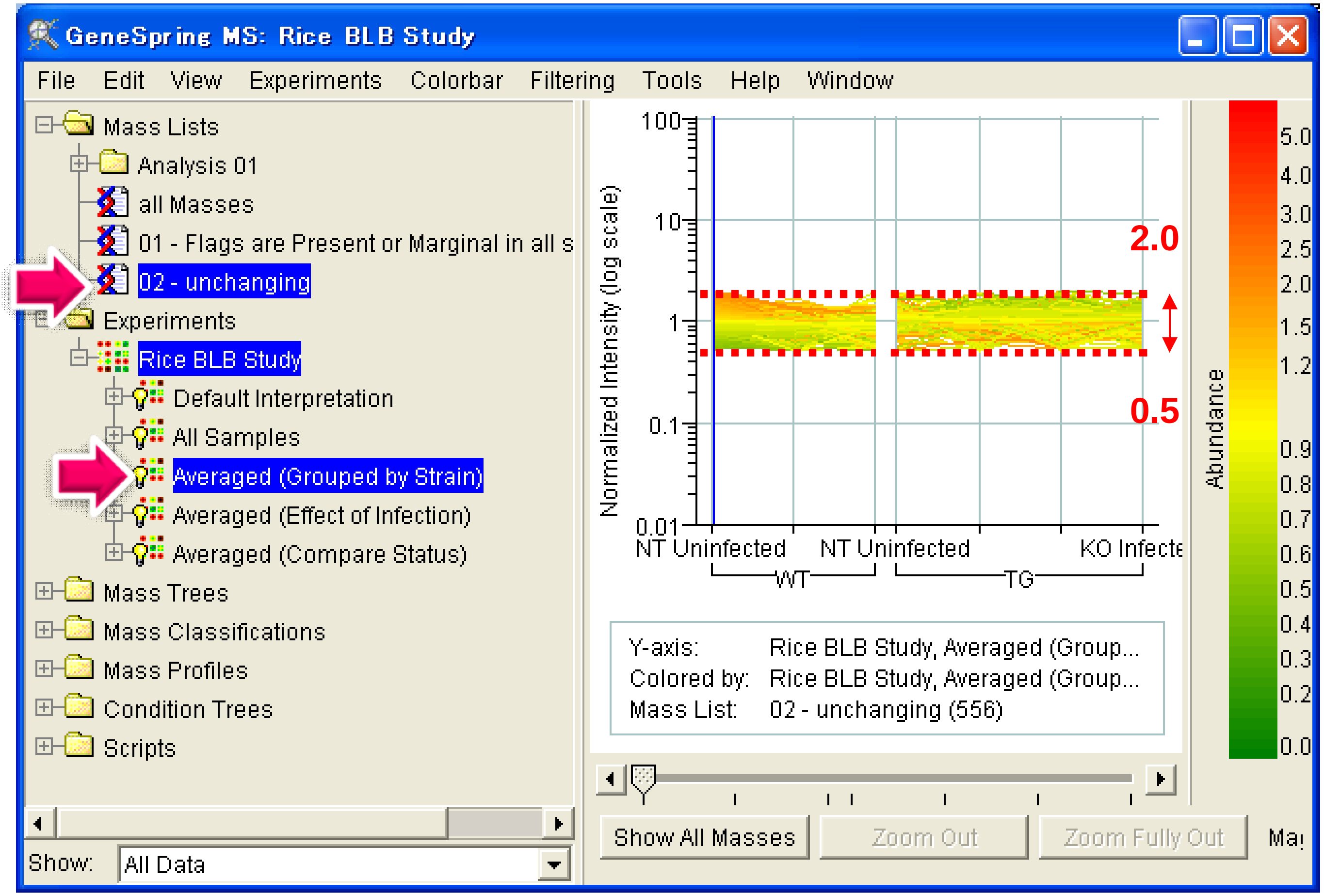
Choose Data Type: Normalized Data

556 out of 37,818 Masses pass filter

Minimum: 0.5　Maximum: 2　View: Graph

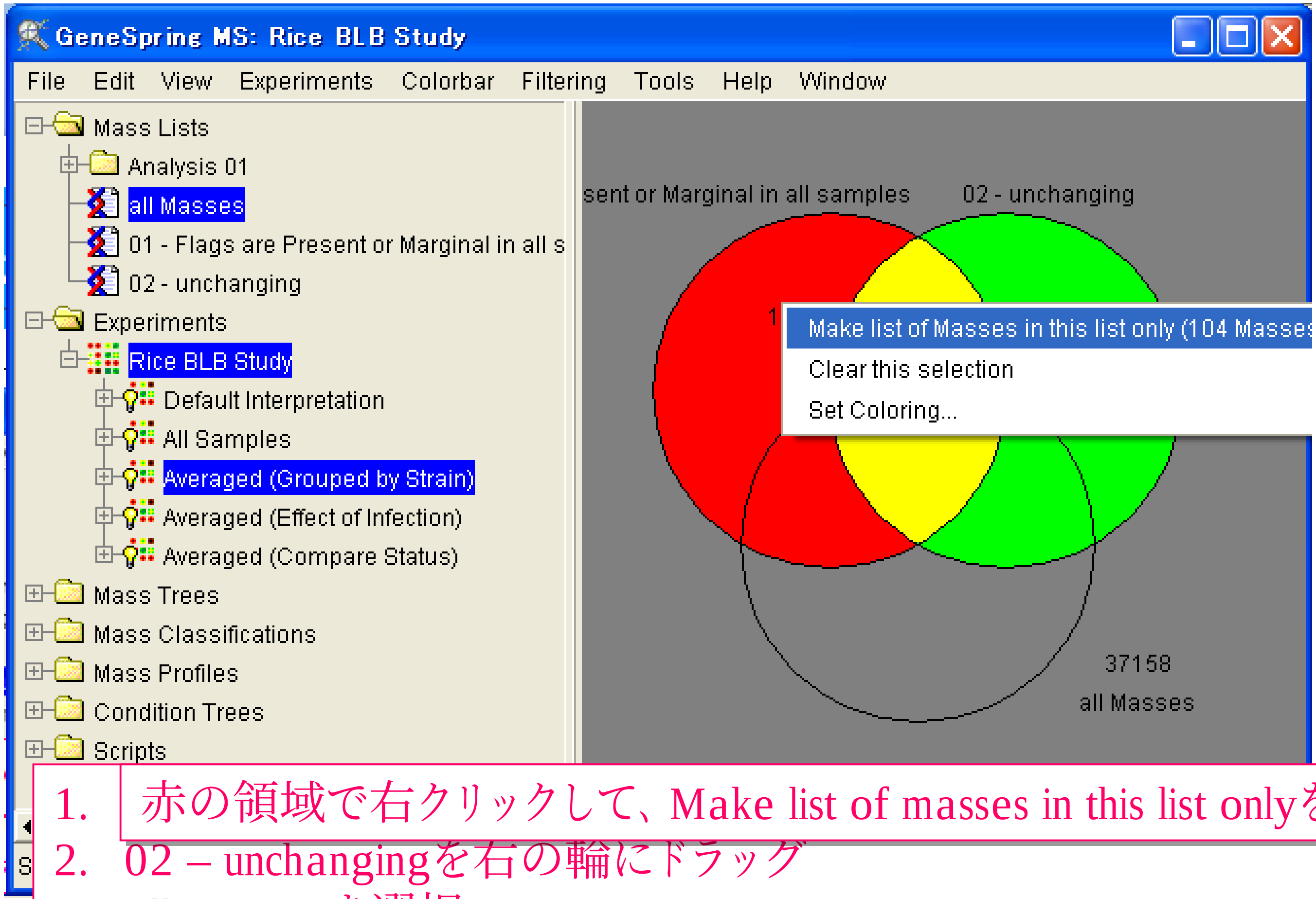
Values must appear in at least 7 out of 7 conditions.

Abundance Levelの変化しない成分



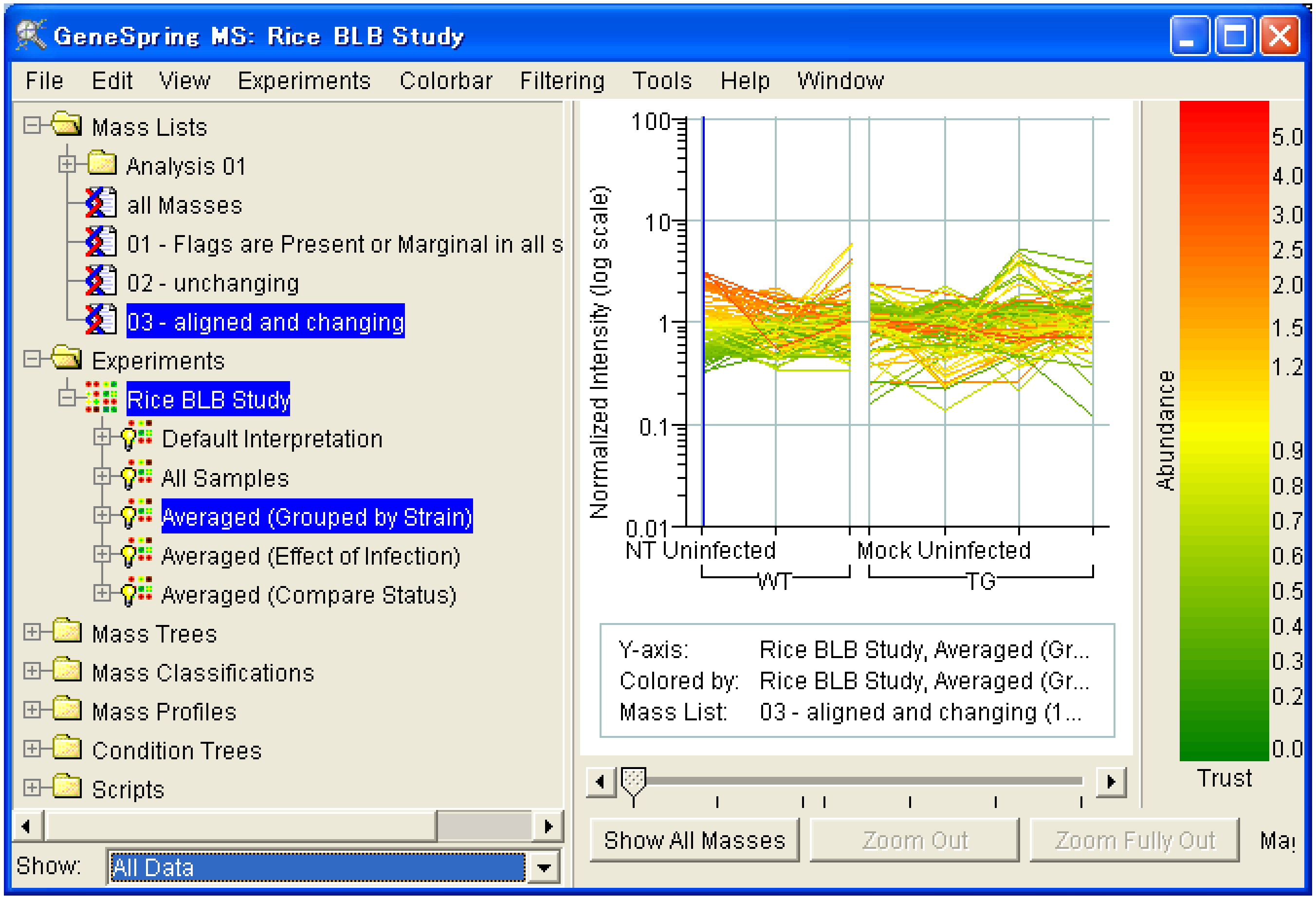
Y-axis: Rice BLB Study, Averaged (Group...
Colored by: Rice BLB Study, Averaged (Group...
Mass List: 02 - unchanging (556)

ベン図による選別



1. 赤の領域で右クリックして、Make list of masses in this list onlyを選択
2. 02 – unchangingを右の輪にドラッグ
3. All Masses を選択

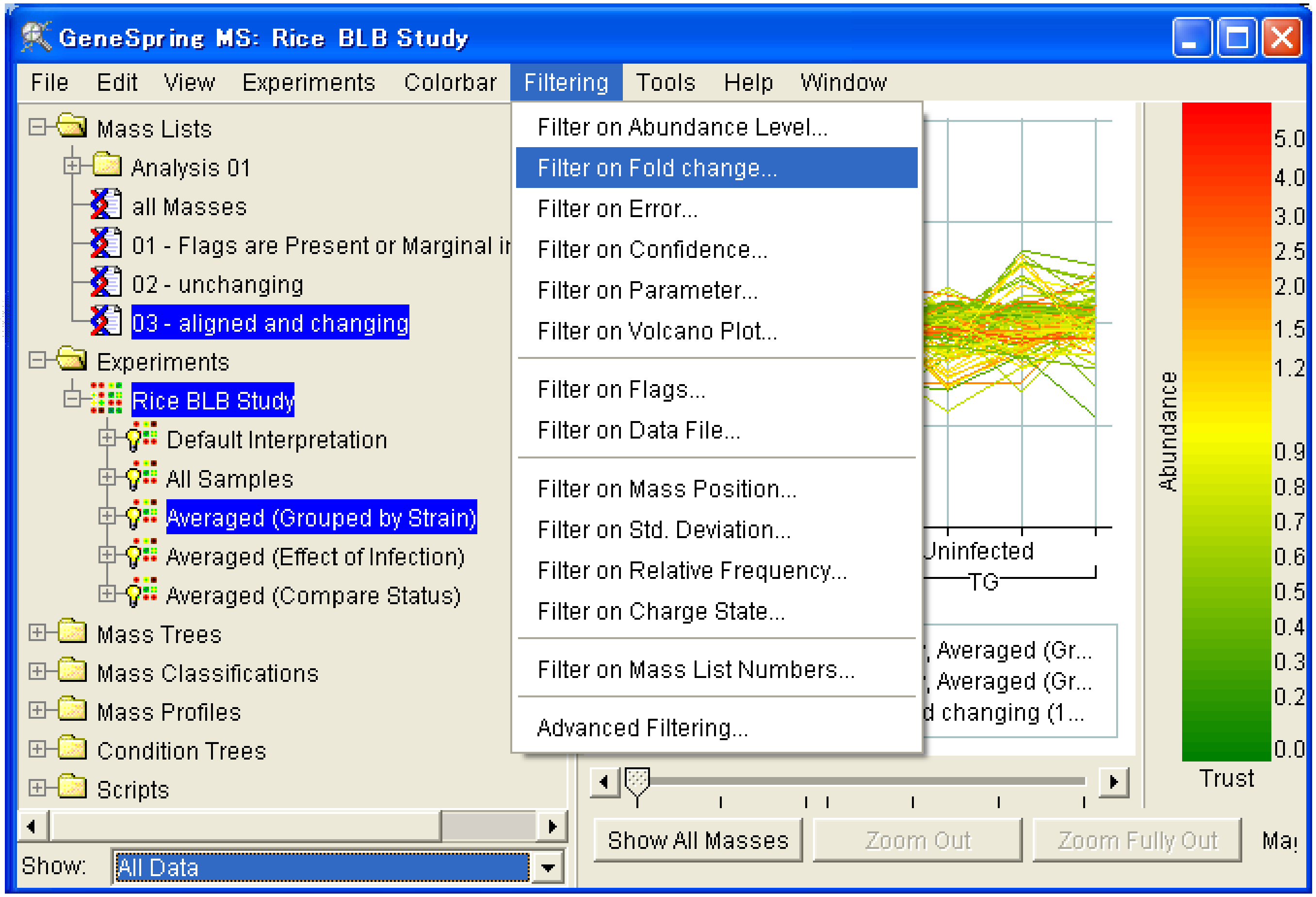
アラインされ、かつ動きのある成分



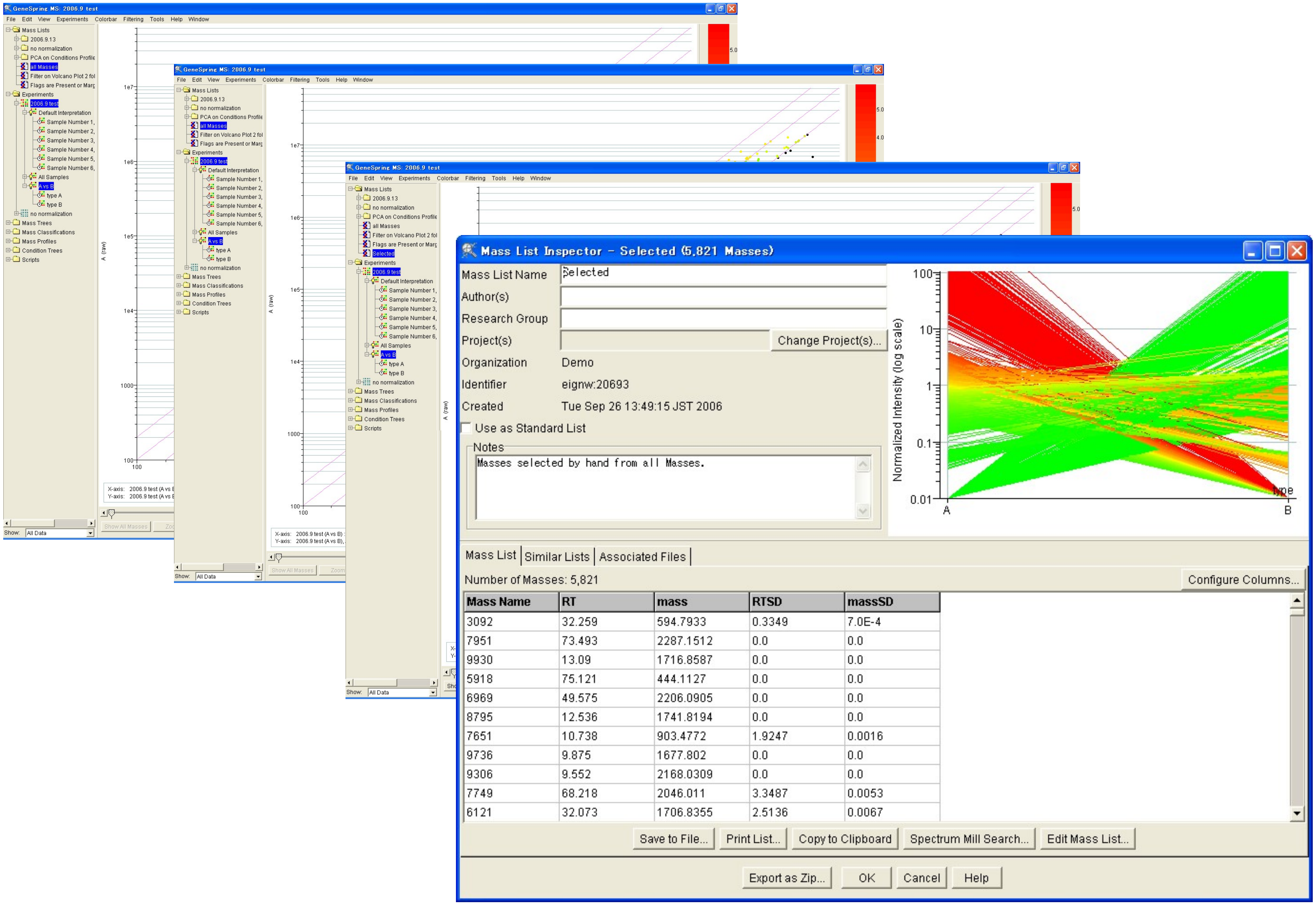
Show: All Data

Y-axis: Rice BLB Study, Averaged (Gr...
Colored by: Rice BLB Study, Averaged (Gr...
Mass List: 03 - aligned and changing (1...

WTのMockとInfectedでAbundanceが変化する成分



2群比較の単純な平均差比較　Fold Change解析



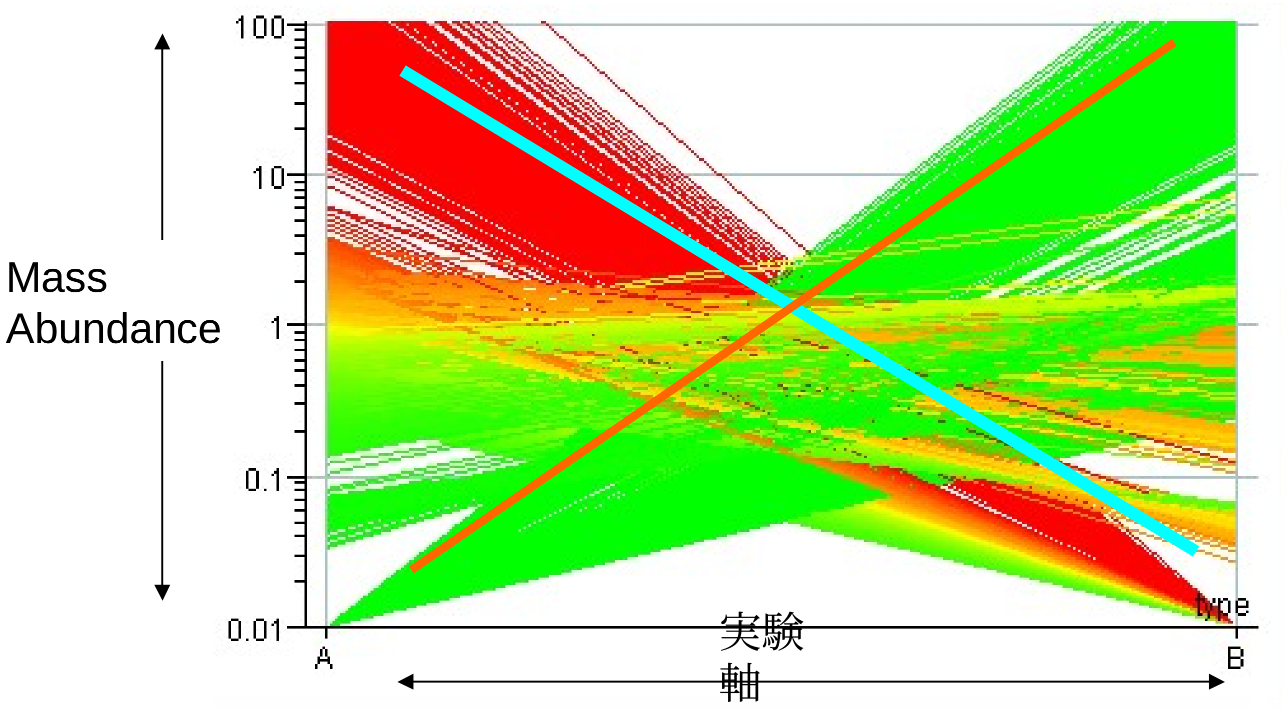
Mass List Name: Selected
Author(s):
Research Group:
Project(s):
Organization: Demo
Identifier: eignw:20693
Created: Tue Sep 26 13:49:15 JST 2006
☐ Use as Standard List
Notes: Masses selected by hand from all Masses.

Number of Masses: 5,821

Mass Name	RT	mass	RTSD	massSD
3092	32.259	594.7933	0.3349	7.0E-4
7951	73.493	2287.1512	0.0	0.0
9930	13.09	1716.8587	0.0	0.0
5918	75.121	444.1127	0.0	0.0
6969	49.575	2206.0905	0.0	0.0
8795	12.536	1741.8194	0.0	0.0
7651	10.738	903.4772	1.9247	0.0016
9736	9.875	1677.802	0.0	0.0
9306	9.552	2168.0309	0.0	0.0
7749	68.218	2046.011	3.3487	0.0053
6121	32.073	1706.8355	2.5136	0.0067

tips

Fold Change解析

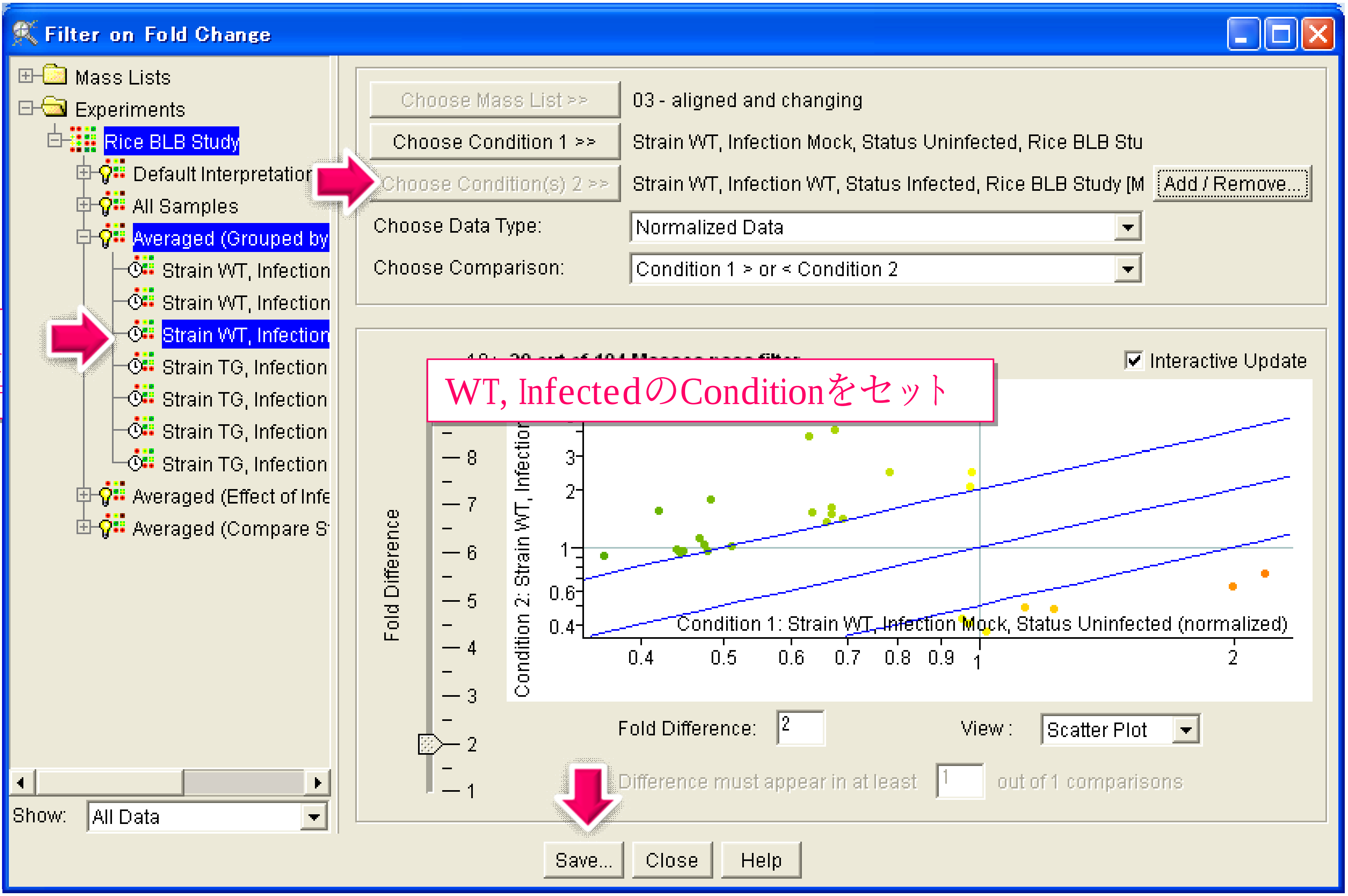


▬ Blue Line Mass A実験で高い、B実験で低い

▬ Red Line Mass A実験で低い、A実験で低い

tips

Filter on Fold Change



03 - aligned and changing

Strain WT, Infection Mock, Status Uninfected, Rice BLB Stu

Strain WT, Infection WT, Status Infected, Rice BLB Study [M

Choose Data Type: Normalized Data

Choose Comparison: Condition 1 > or < Condition 2

☑ Interactive Update

Fold Difference: 2

View : Scatter Plot

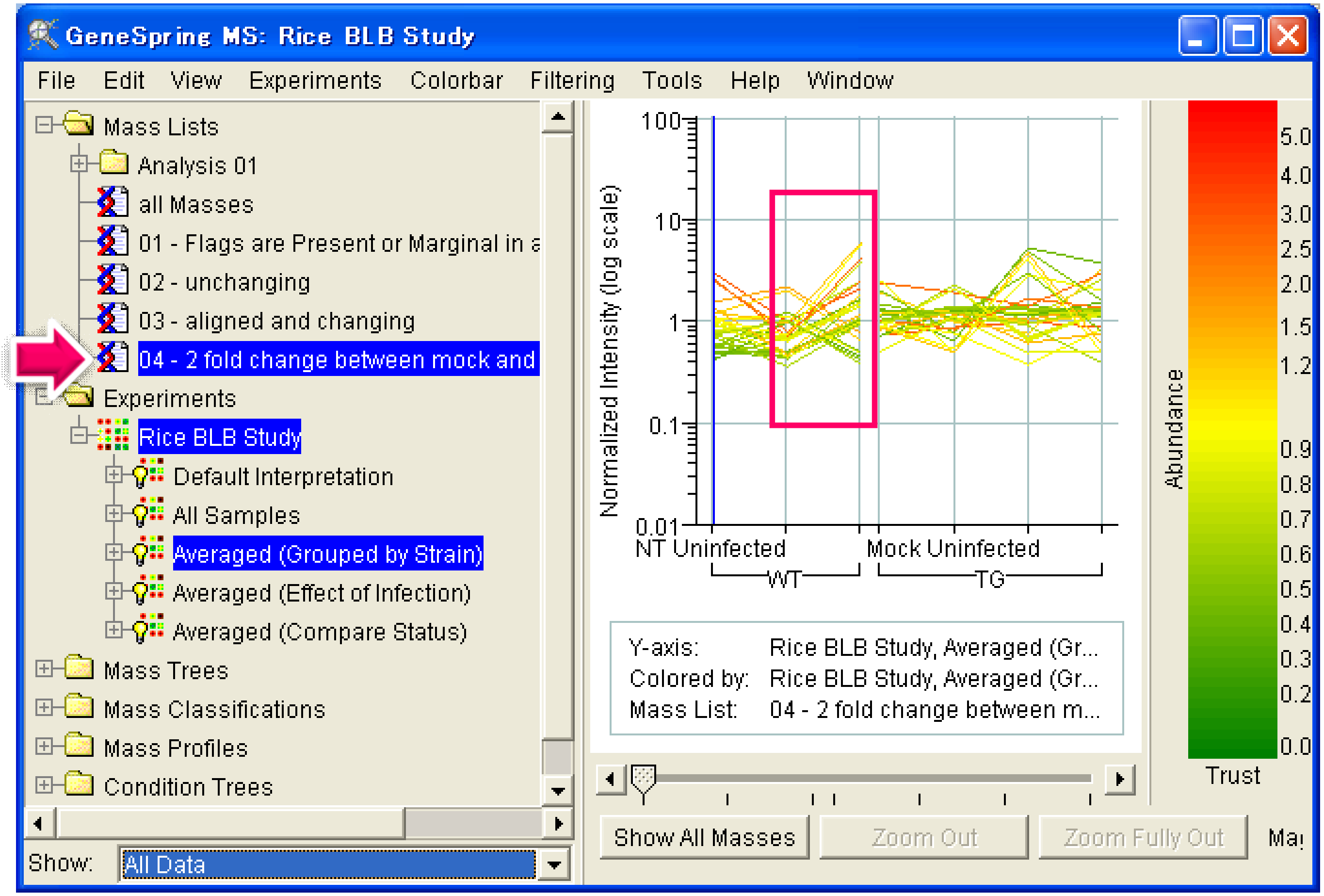
Difference must appear in at least 1 out of 1 comparisons

Show: All Data

Avera
を展開

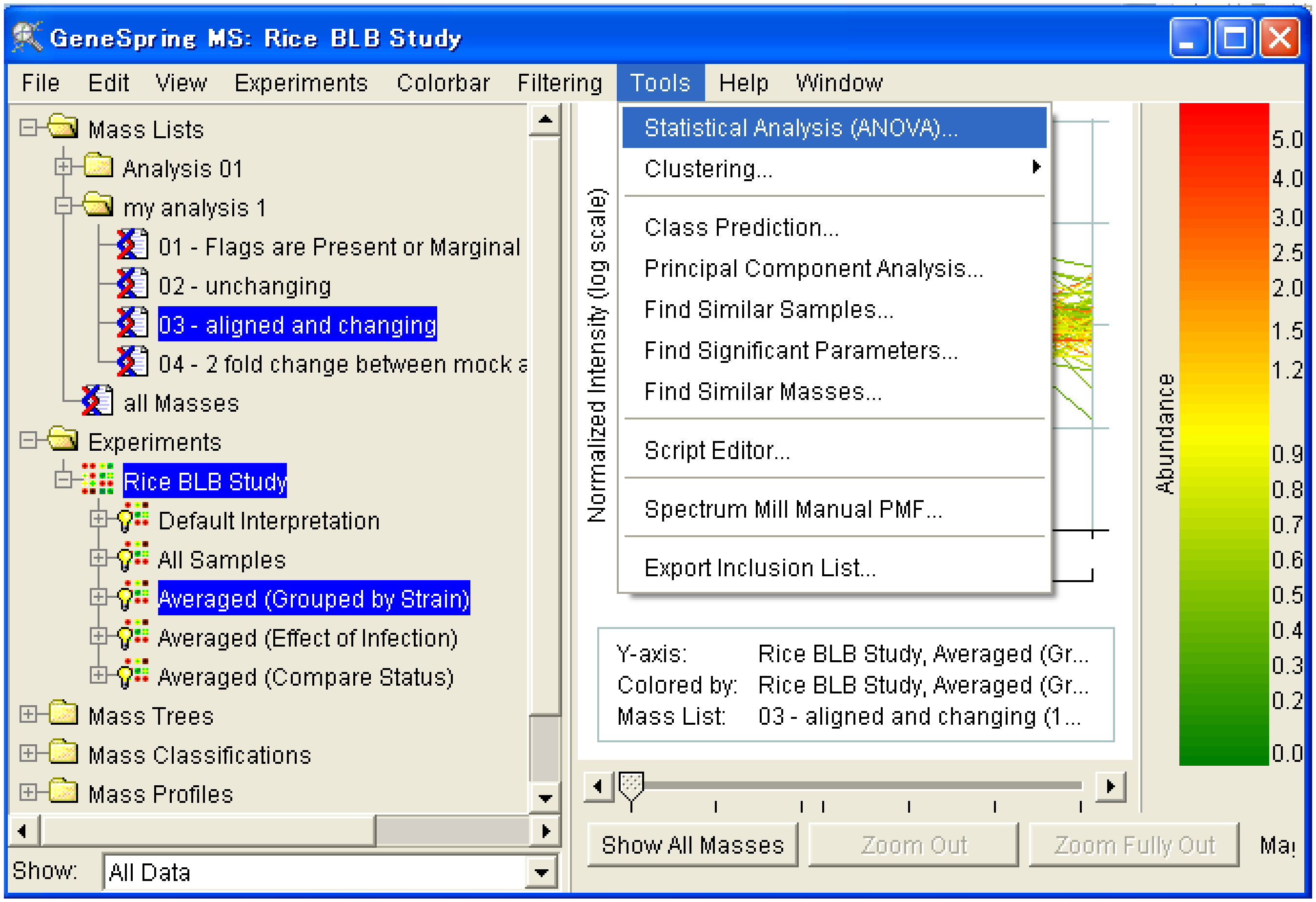
WT, InfectedのConditionをセット

MockとInfectedで差がある成分
- Fold Change解析



Y-axis: Rice BLB Study, Averaged (Gr...
Colored by: Rice BLB Study, Averaged (Gr...
Mass List: 04 - 2 fold change between m...

統計学的有意な変化を示す成分の抽出

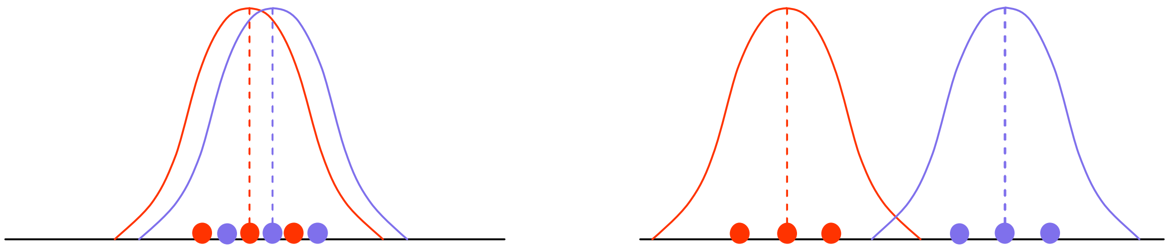


Show: All Data

Y-axis: Rice BLB Study, Averaged (Gr...
Colored by: Rice BLB Study, Averaged (Gr...
Mass List: 03 - aligned and changing (1...

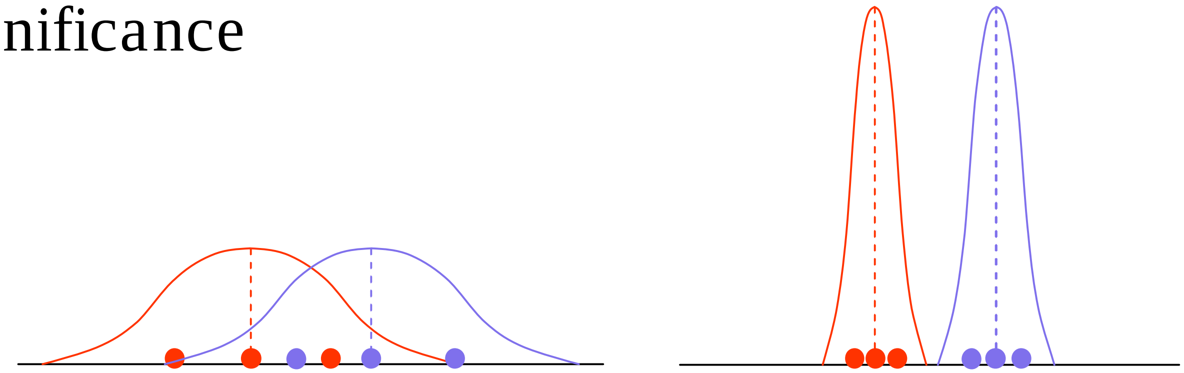
「統計学的有意差」のイメージ

Fold Analysis



距離

Significance

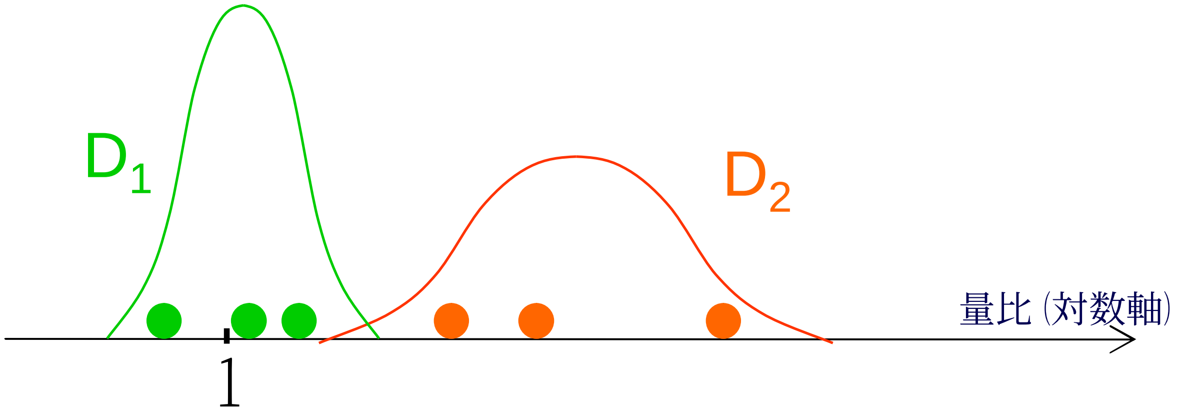


$$\frac{\text{距離}}{\text{誤差}}$$

tips

検定の意味と帰無仮説 – P値

たとえば、母集団をμ=1の正規分布と仮定したとき(帰無仮説)、偶然そのような観測値が得られる確率(p値)を計算する。
これが5%以下であれば、帰無仮説を棄却し、別の集団の観測値だという立場に立ち、μ≠1と結論付ける。



もし帰無仮説が真ならば、
D_1のような観測は珍しくない

→　P値が大きい

もし帰無仮説が真ならば、
D_2のような観測は起こりにくい

→　P値が小さい

tips

T-test / 1 way - ANOVA

適用するケース

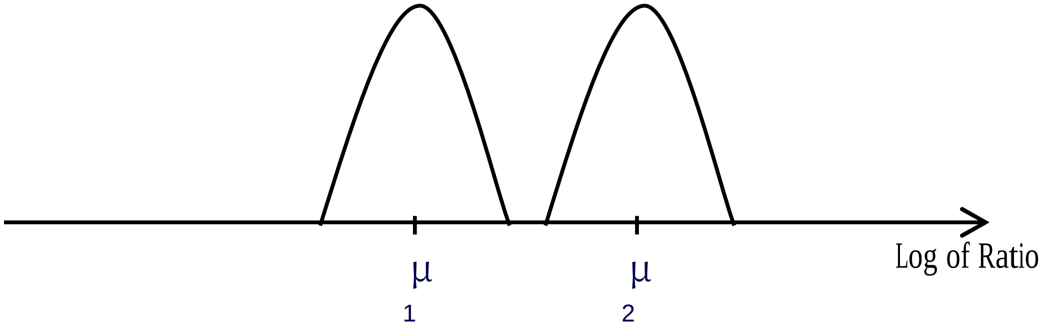
パラメーターによって分かれる2つ以上のグループ間で、
有意差のあるMassを抽出する。

例：　健常者と患者のサンプル群
帰無仮説：健常者と患者で一致する（＝両者間に差がない）。
対立仮説：健常者と患者で一致しない。
（P値→健常者と患者の間で差がない確率）

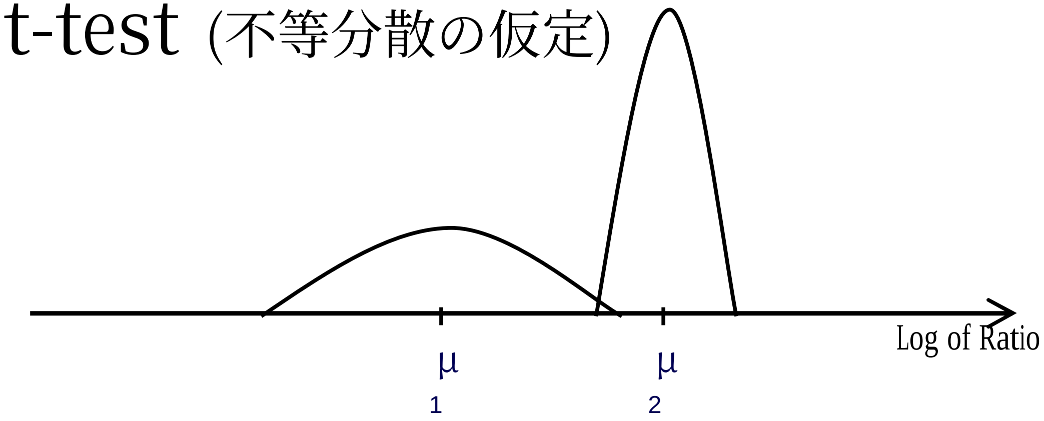
	2グループ	３グループ以上
パラメトリック (等分散と仮定)	Student's T-test	ANOVA
パラメトリック (等分散と仮定しない)	Welch t-test	Welch ANOVA
ノンパラメトリック	Wilcoxon rank test	Kruskal-Wallis test

パラメトリックテストの選択について

- Student’s t-test (等分散の仮定)

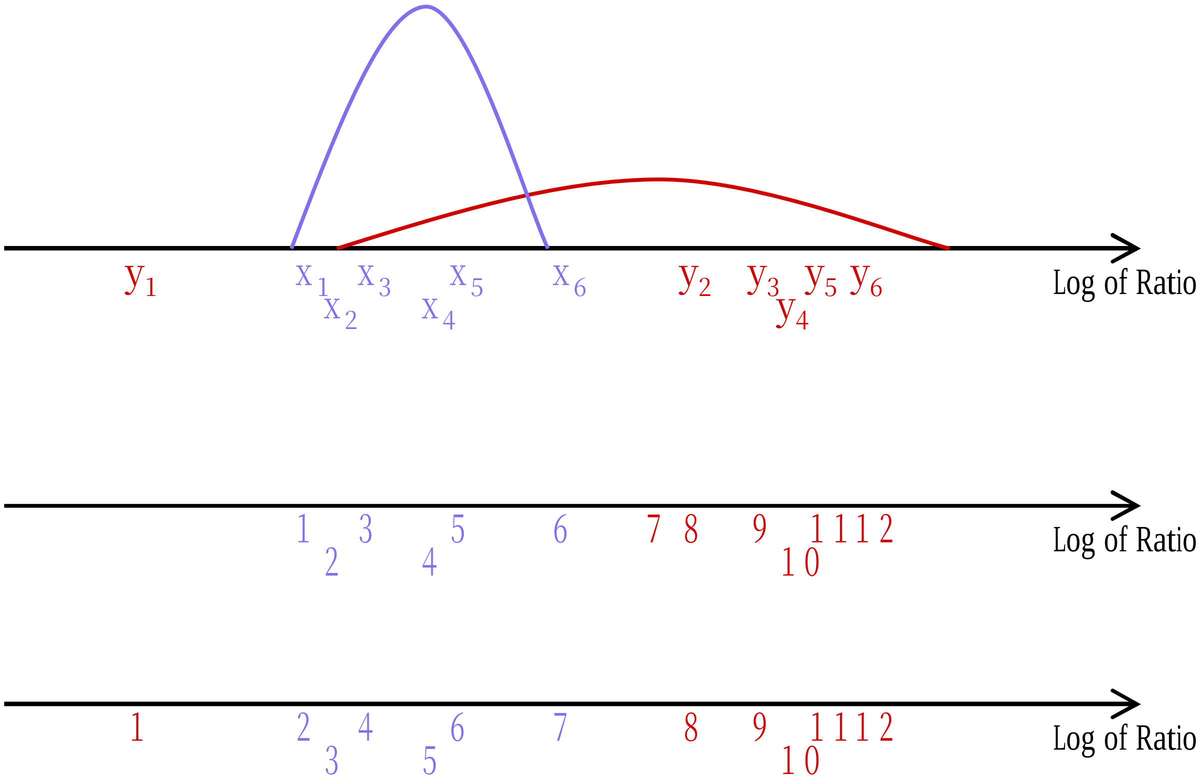


- Welch’s t-test (不等分散の仮定)



ノンパラメトリックの選択について

パラメトリックテストは外れ値の影響を受けやすい



順位検定は外れ値の影響を受けにくくなる

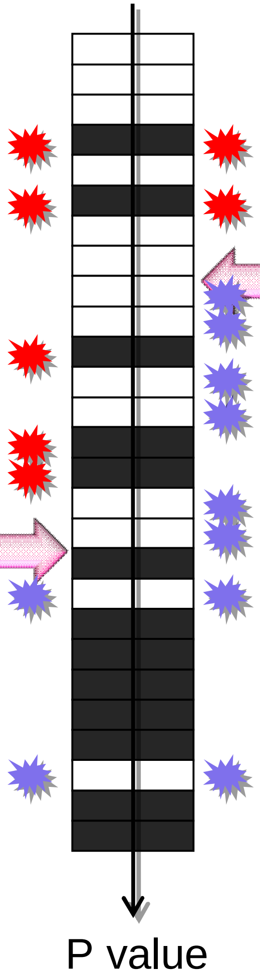
偽陽性と偽陰性

	有意差なし	有意差あり
真の差なし	True Negative	False Positive
真の差あり	False Negative	True Positive

False PositiveとFalse Negativeが0であれば、
完璧な解析と言える。

統計検定の問題１：トレードオフ

P値が小さければいいのか？



厳しい有意水準
（低いP値）

科学的発見や、商業的チャンスの見落としにつながる。

緩い有意水準
（高いP値）

偽陽性結果を検証するための別な実験を行える設備と費用がかかる。

☐ True Positive (TP)
■ True Negative (TN)
✸ False Positive (FP)
✸ False Negative (FN)

統計検定の問題 ２：分散の不確かさ

真の分布 （母集団, **population**）

μ　真の値、母平均、population mean

σ　真の分布のばらつき

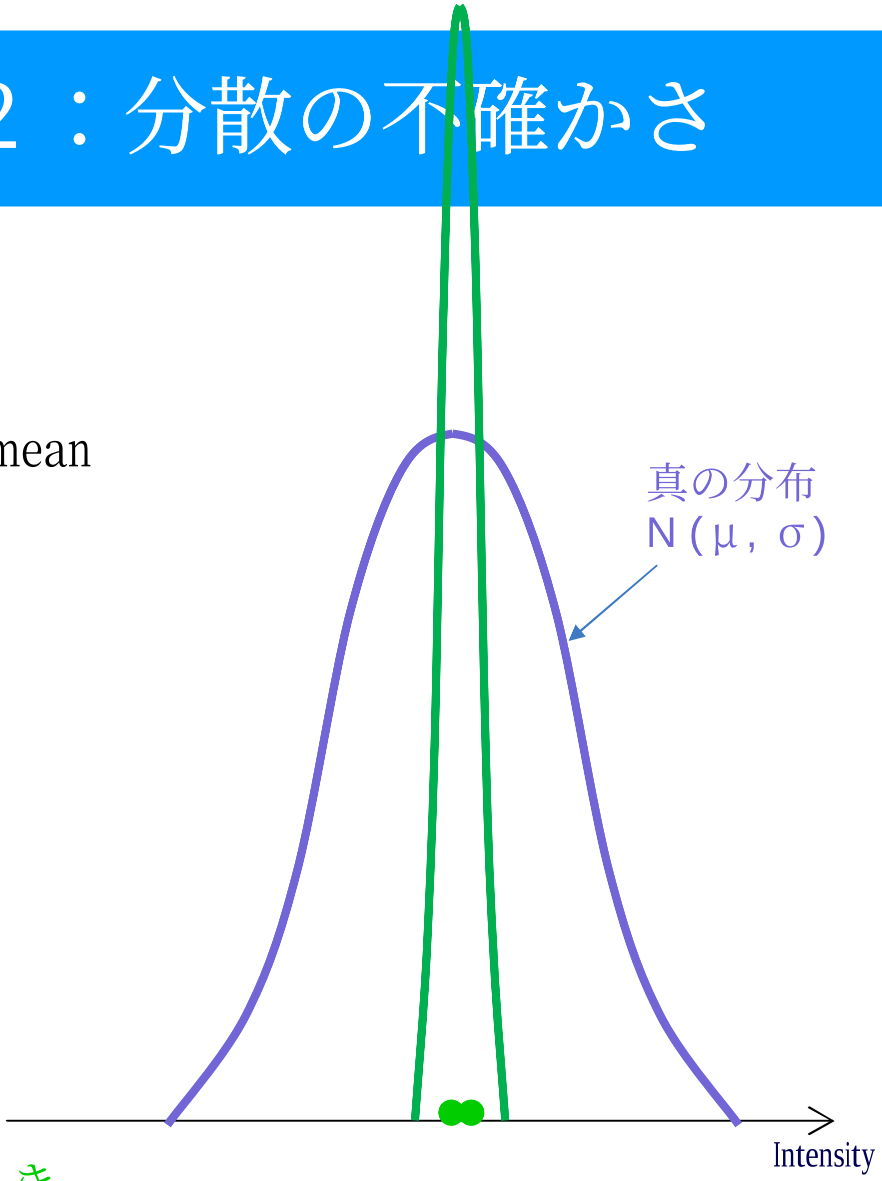
標本から推定された分布

m　標本平均、sample mean

s　不偏標準偏差

注意！

繰り返し実験の数が非常に少ないとき、
分散が極端に小さくなりやすい。



統計検定の問題３：多重比較

- １０,０００個のMassを有意水準５％で検定したとき、偽陽性の期待値は５００個にもなる。

検定によりMassを抽出するとき、抽出したMassの数だけでなく、その中に含まれる偽陽性の数を考慮しなければならない。

ちなみに、ある有名生物系の論文では、このような投稿規程が記載…

Large-scale experiments. In evaluating large-scale experiments such as transcript profiling, we will consider whether there is a clear and complete description of each experiment; whether biological and/or technical replicates should have been used; what statistical analysis has been performed; whether **a multiple comparison correction** has been used to control for **Type I family-wise error**, where necessary; and/or whether the need for statistical analysis to support the claims has been obviated through validation of claims by independent experiments.

- Type I Error （偽陽性）
- Type II Error （偽陰性）
- Multiplicity （多重性）

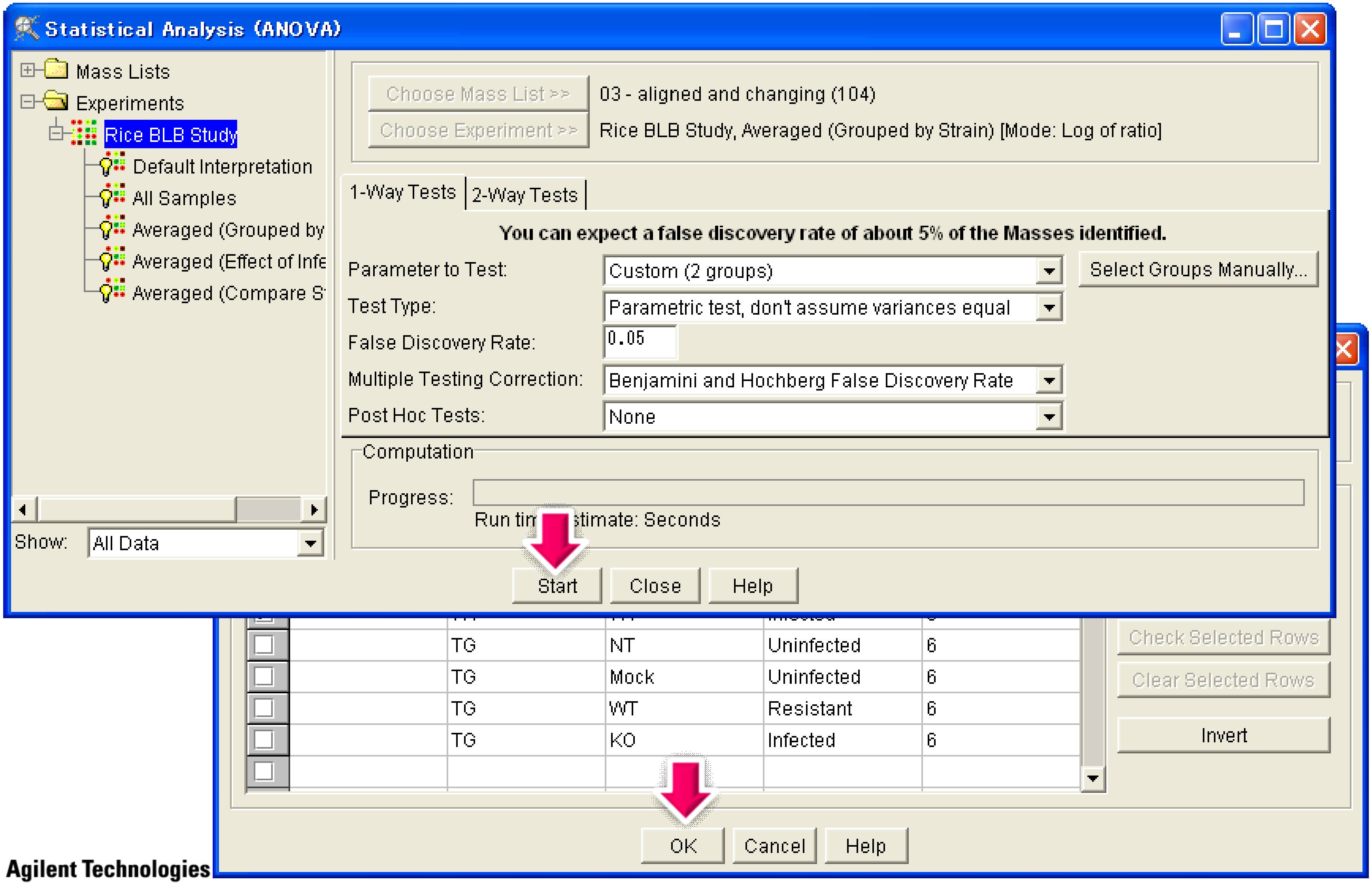
tips

有意水準の補正

- 仮説群全体の有意水準（FWER）
  - Bonferroni, Holm, Westfall&Young
  - きわめて保守的であり、多くの偽陰性の原因となる
- 陽性に対する偽陽性の割合の期待値（FDR）
  - Benjamini&Hochberg, SAM
  - FWERを弱くコントロール
  - 現実のデータ解析で頻用される

tips

Statistical Analysis (ANOVA)



Statistical Analysis (ANOVA)

Show: All Data

Choose Mass List >> 03 - aligned and changing (104)

Choose Experiment >> Rice BLB Study, Averaged (Grouped by Strain) [Mode: Log of ratio]

You can expect a false discovery rate of about 5% of the Masses identified.

Parameter to Test: Custom (2 groups)

Test Type: Parametric test, don't assume variances equal

False Discovery Rate: 0.05

Multiple Testing Correction: Benjamini and Hochberg False Discovery Rate

Post Hoc Tests: None

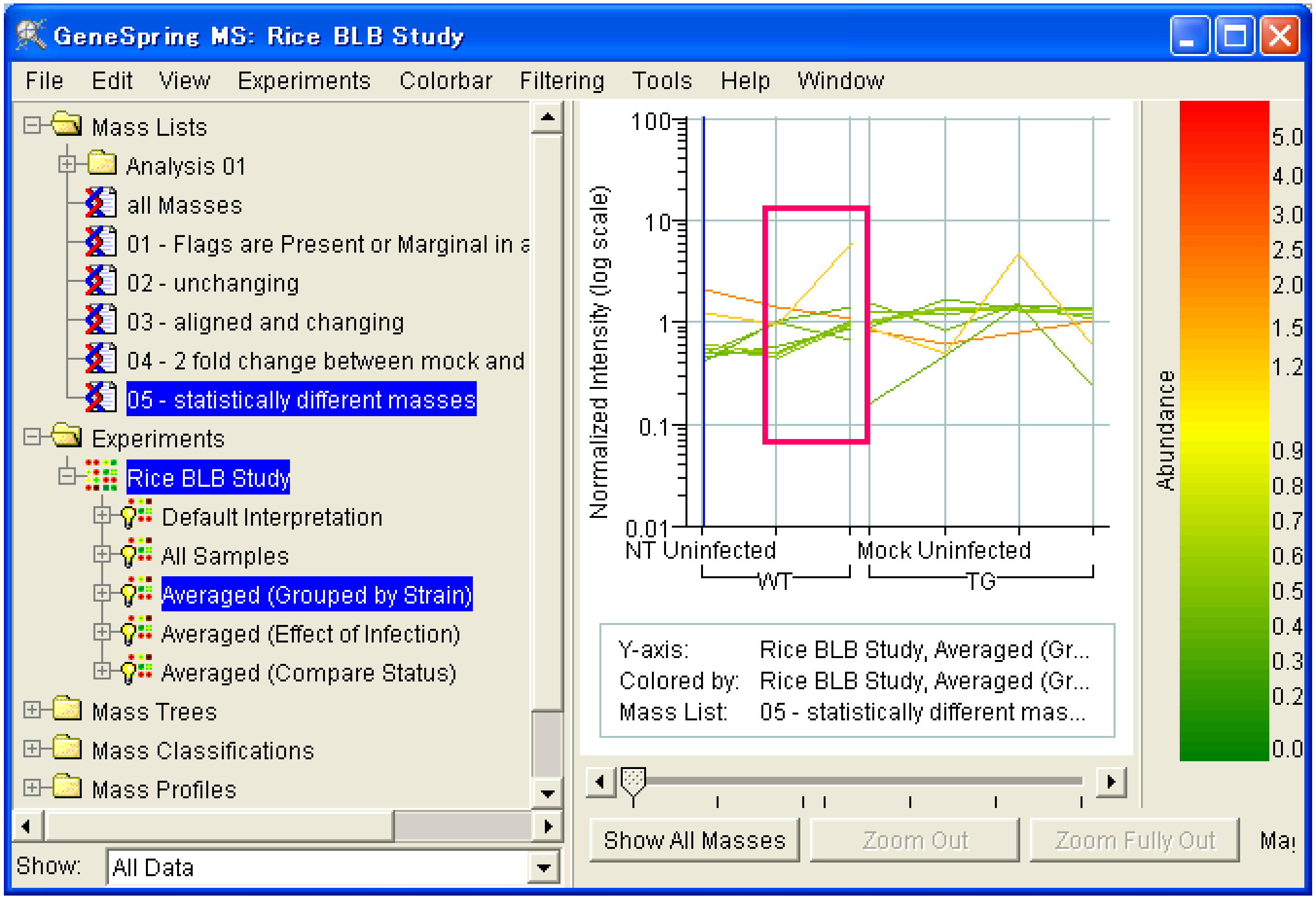
Computation

Progress:

Run tim stimate: Seconds

☐		TG	NT	Uninfected	6
☐		TG	Mock	Uninfected	6
☐		TG	WT	Resistant	6
☐		TG	KO	Infected	6
☐					

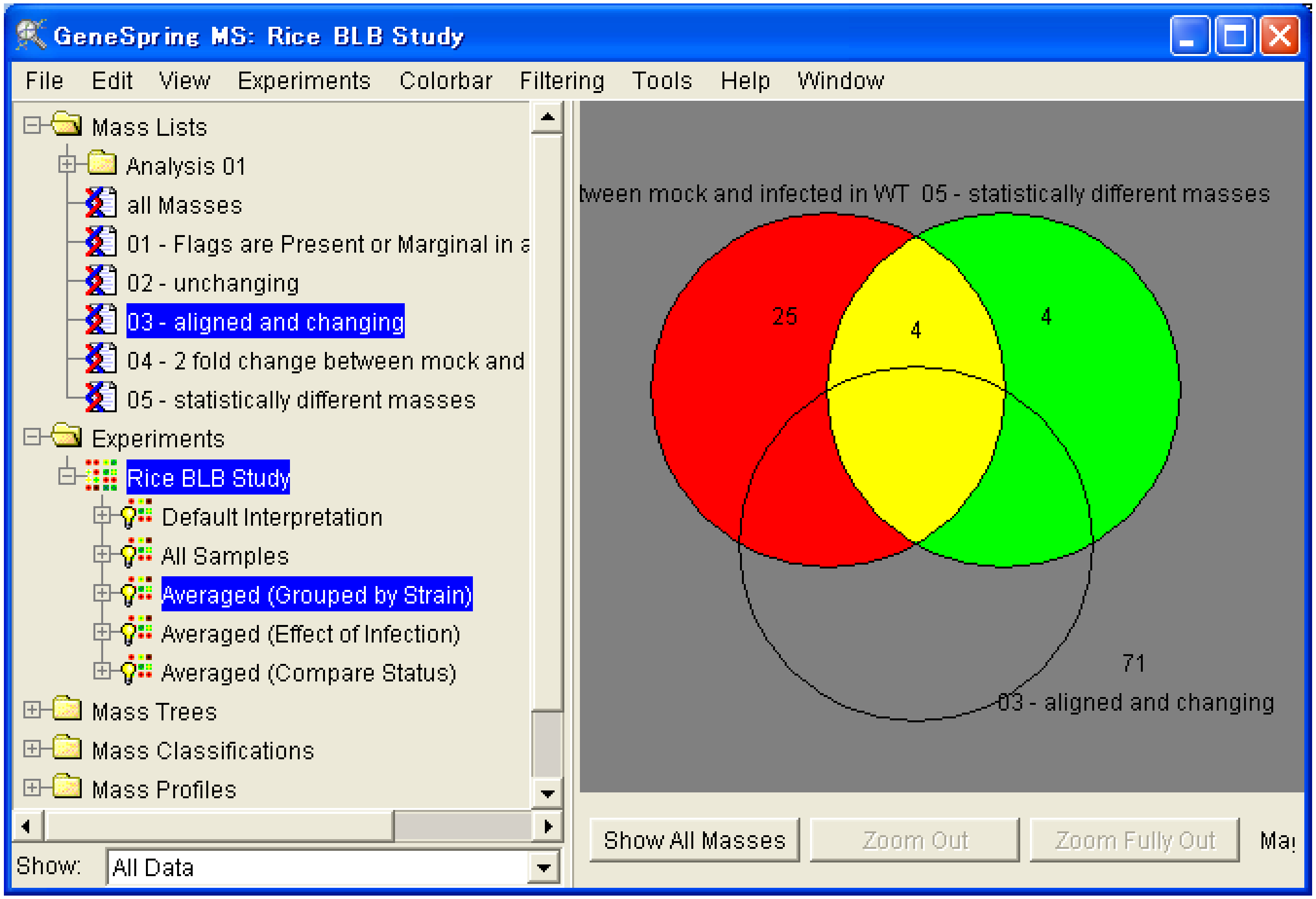
統計学的有意な変化を示す成分



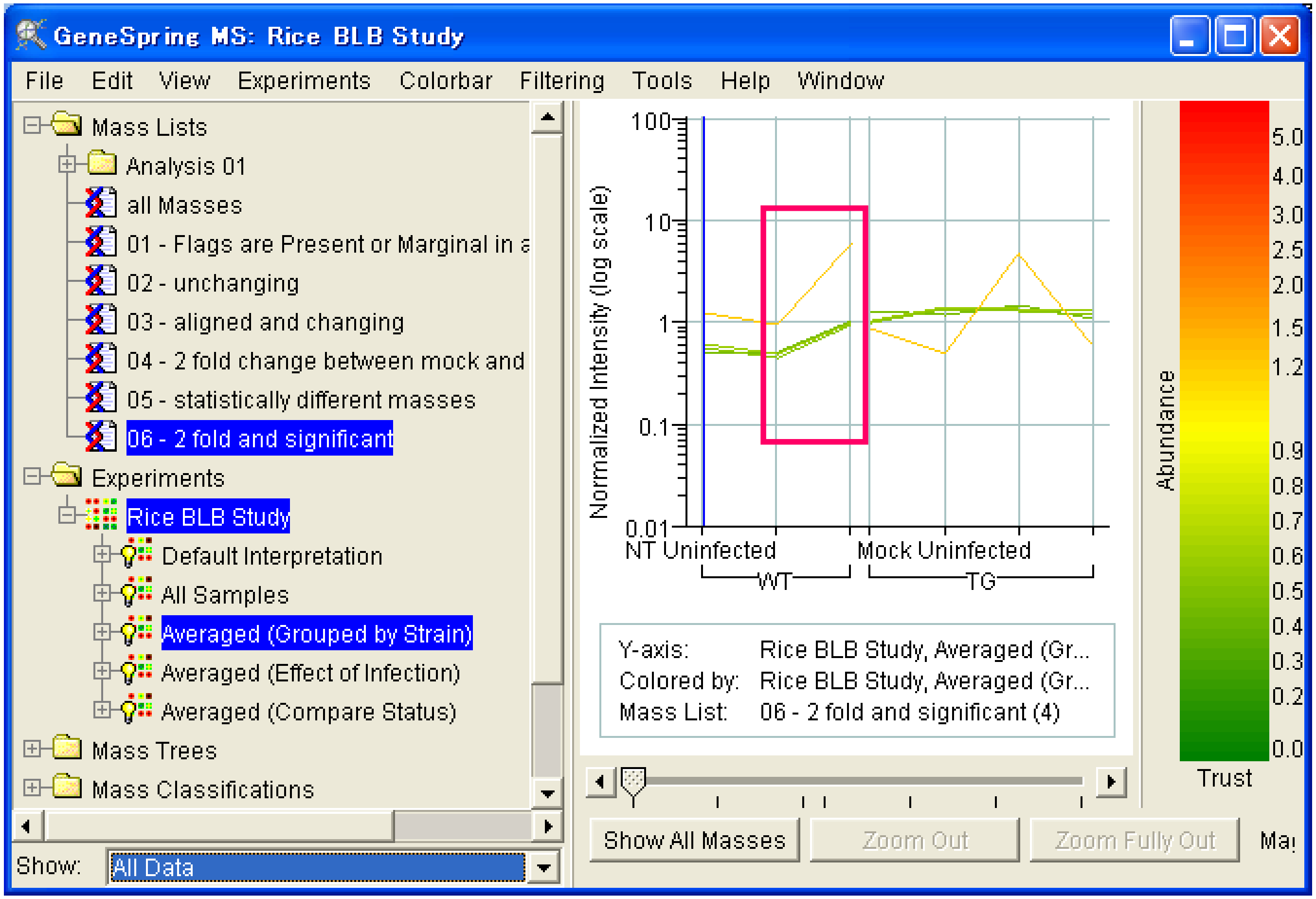
Show: All Data

Y-axis: Rice BLB Study, Averaged (Gr...
Colored by: Rice BLB Study, Averaged (Gr...
Mass List: 05 - statistically different mas...

ベン図を使った比較

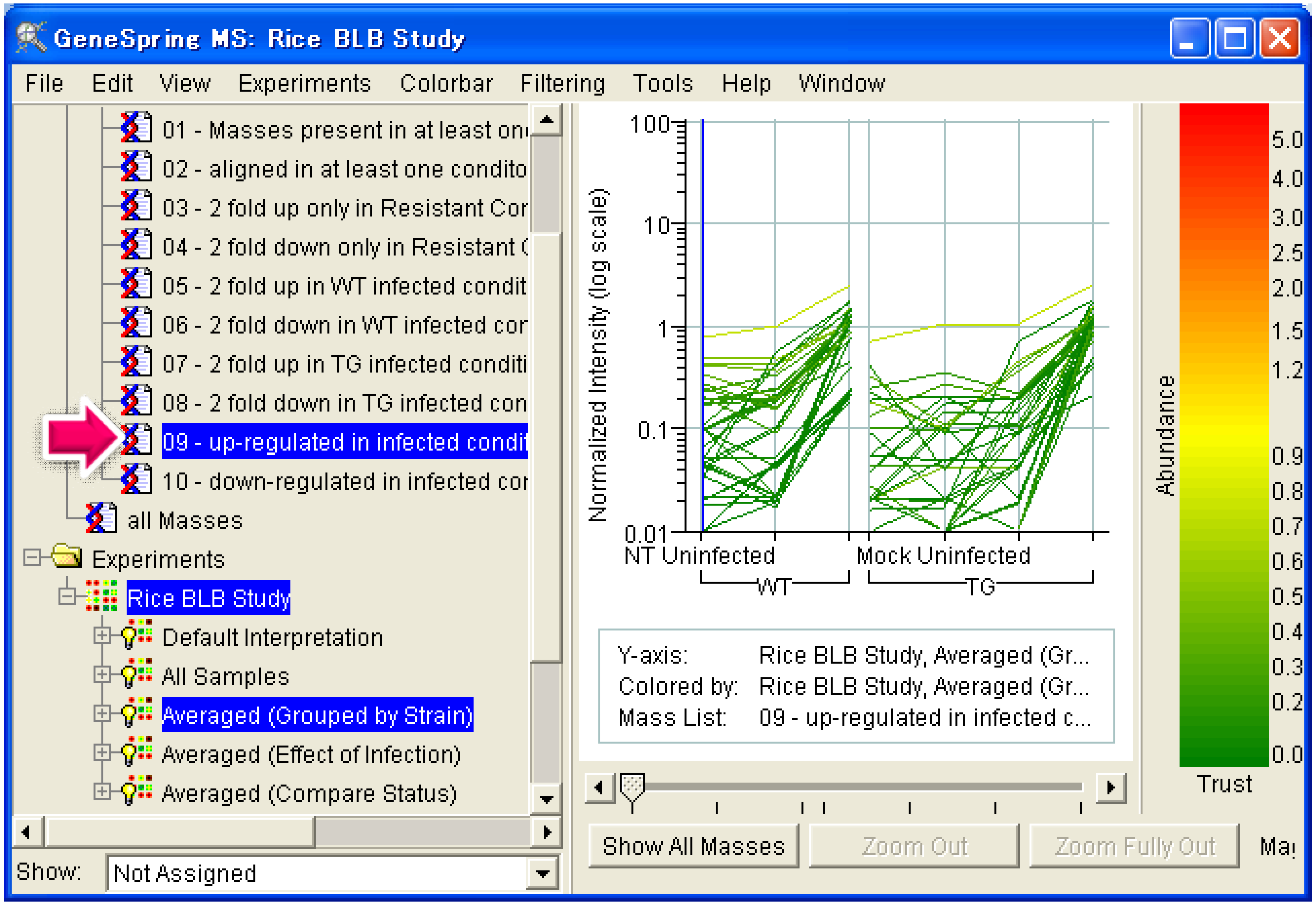

GeneSpring MS: Rice BLB Study
File
Edit
View
Experiments
Colorbar
Filtering
Tools
Help
Window
Mass Lists
Analysis 01
all Masses
01 - Flags are Present or Marginal in a
02 - unchanging
03 - aligned and changing
04 - 2 fold change between mock and
05 - statistically different masses
Experiments
Rice BLB Study
Default Interpretation
All Samples
Averaged (Grouped by Strain)
Averaged (Effect of Infection)
Averaged (Compare Status)
Mass Trees
Mass Classifications
Mass Profiles
Show:
All Data
tween mock and infected in WT
05 - statistically different masses
25
4
4
71
03 - aligned and changing
Show All Masses
Zoom Out
Zoom Fully Out
Ma

倍率と有意差により抽出した成分



Y-axis: Rice BLB Study, Averaged (Gr...
Colored by: Rice BLB Study, Averaged (Gr...
Mass List: 06 - 2 fold and significant (4)

感染状態を示すバイオマーカーの候補

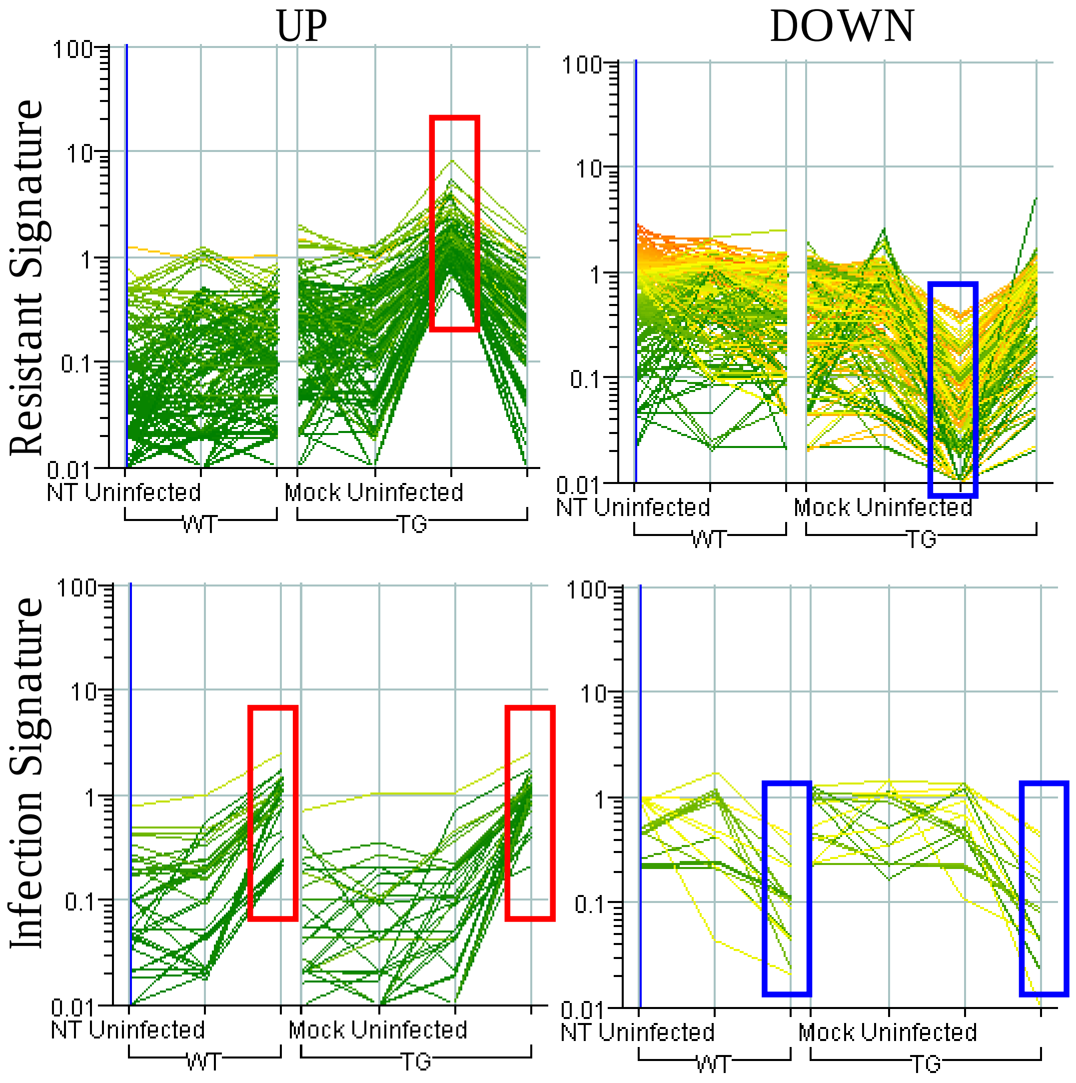


01 - Masses present in at least on
02 - aligned in at least one conditio
03 - 2 fold up only in Resistant Cor
04 - 2 fold down only in Resistant
05 - 2 fold up in WT infected condit
06 - 2 fold down in WT infected cor
07 - 2 fold up in TG infected conditi
08 - 2 fold down in TG infected con
09 - up-regulated in infected condit
10 - down-regulated in infected cor
all Masses
Experiments
Rice BLB Study
Default Interpretation
All Samples
Averaged (Grouped by Strain)
Averaged (Effect of Infection)
Averaged (Compare Status)

Show: Not Assigned

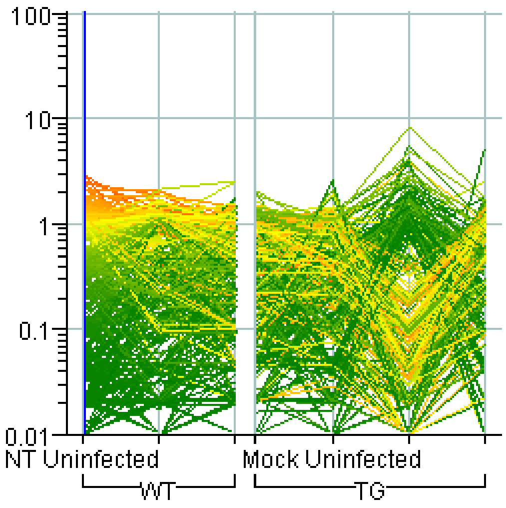
Y-axis: Rice BLB Study, Averaged (Gr...
Colored by: Rice BLB Study, Averaged (Gr...
Mass List: 09 - up-regulated in infected c...

バイオマーカー候補の Mass

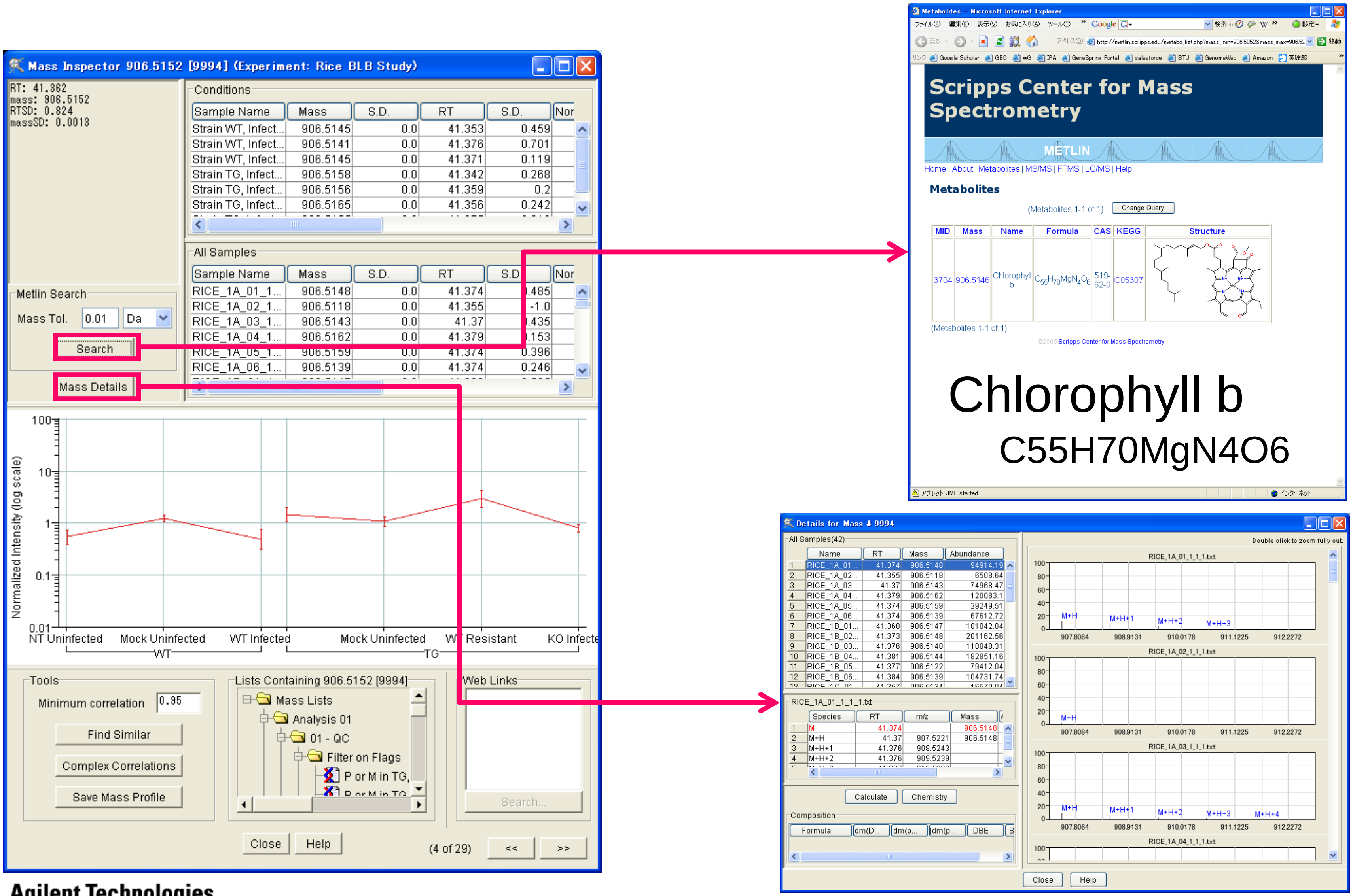


和集合

Profiling Mass の候補



ターゲット候補のMass Inspector



Mass Inspector 906.5152 [9994] (Experiment: Rice BLB Study)

```
RT: 41.362
mass: 906.5152
RTSD: 0.824
massSD: 0.0013
```

Conditions

Sample Name	Mass	S.D.	RT	S.D.	Nor
Strain WT, Infect...	906.5145	0.0	41.353	0.459	
Strain WT, Infect...	906.5141	0.0	41.376	0.701	
Strain WT, Infect...	906.5145	0.0	41.371	0.119	
Strain TG, Infect...	906.5158	0.0	41.342	0.268	
Strain TG, Infect...	906.5156	0.0	41.359	0.2	
Strain TG, Infect...	906.5165	0.0	41.356	0.242	

All Samples

Sample Name	Mass	S.D.	RT	S.D.	Nor
RICE_1A_01_1...	906.5148	0.0	41.374	.485	
RICE_1A_02_1...	906.5118	0.0	41.355	-1.0	
RICE_1A_03_1...	906.5143	0.0	41.37	.435	
RICE_1A_04_1...	906.5162	0.0	41.379	.153	
RICE_1A_05_1...	906.5159	0.0	41.374	0.396	
RICE_1A_06_1...	906.5139	0.0	41.374	0.246	

Metlin Search

Mass Tol. 0.01 Da

Metabolites - Microsoft Internet Explorer

http://metlin.scripps.edu/metabo_list.php?mass_min=906.5052&mass_max=906.52

Scripps Center for Mass Spectrometry

METLIN

Home | About | Metabolites | MS/MS | FTMS | LC/MS | Help

Metabolites

(Metabolites 1-1 of 1)

MID	Mass	Name	Formula	CAS	KEGG	Structure
3704	906.5146	Chlorophyll b	$C_{55}H_{70}MgN_4O_6$	519-62-0	C05307	

(Metabolites 1-1 of 1)

©2005 Scripps Center for Mass Spectrometry

Chlorophyll b

C55H70MgN4O6

Details for Mass # 9994

All Samples(42)

	Name	RT	Mass	Abundance
1	RICE_1A_01...	41.374	906.5148	94914.19
2	RICE_1A_02...	41.355	906.5118	6508.64
3	RICE_1A_03...	41.37	906.5143	74968.47
4	RICE_1A_04...	41.379	906.5162	120083.1
5	RICE_1A_05...	41.374	906.5159	29249.51
6	RICE_1A_06...	41.374	906.5139	67612.72
7	RICE_1B_01...	41.368	906.5147	101042.04
8	RICE_1B_02...	41.373	906.5148	201162.56
9	RICE_1B_03...	41.376	906.5148	110048.31
10	RICE_1B_04...	41.381	906.5144	182851.16
11	RICE_1B_05...	41.377	906.5122	79412.04
12	RICE_1B_06...	41.384	906.5139	104731.74

RICE_1A_01_1_1_1.txt

	Species	RT	m/z	Mass
1	M	41.374		906.5148
2	M+H	41.37	907.5221	906.5148
3	M+H+1	41.376	908.5243	
4	M+H+2	41.376	909.5239	

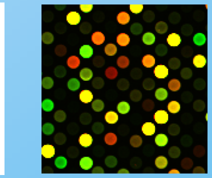
Composition

Formula	dm(D...	dm(p...	|dm(p...	DBE	S

Agilent Technologies

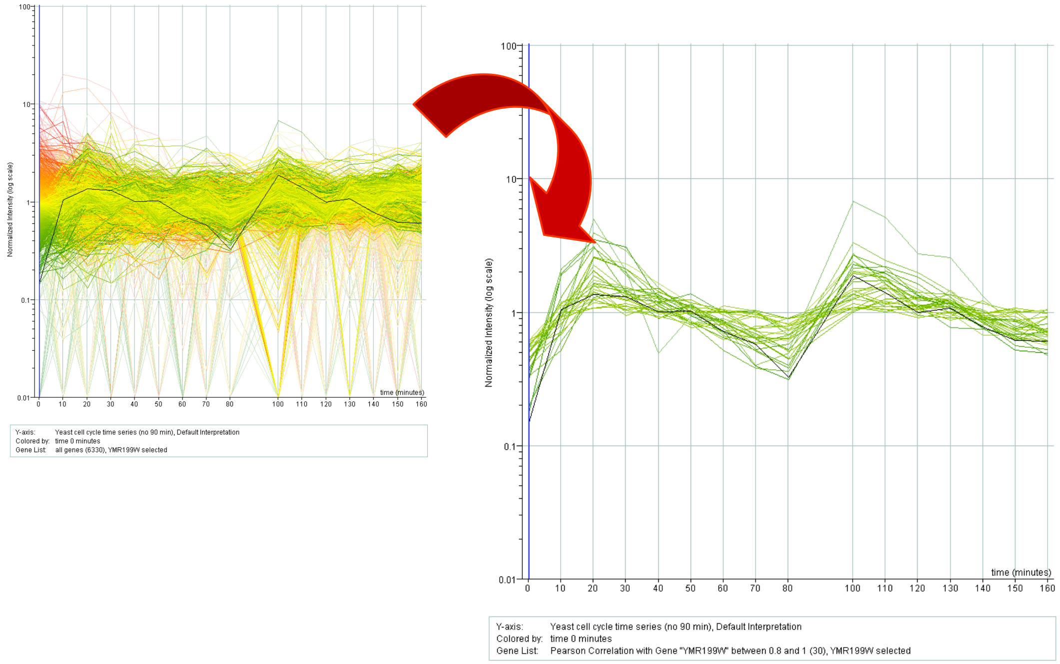
Agilent Technologies

ビール6銘柄の解析例 - パターン解析の例



パターンが「似ている」とは？

- パターン認識
  - 人間はほとんど考えなくても、視覚でパターンを認識できる。
  - コンピュータがパターンを認識して、似ているかどうかを判断するにはどのように（数学的に）表現すればいいのか？



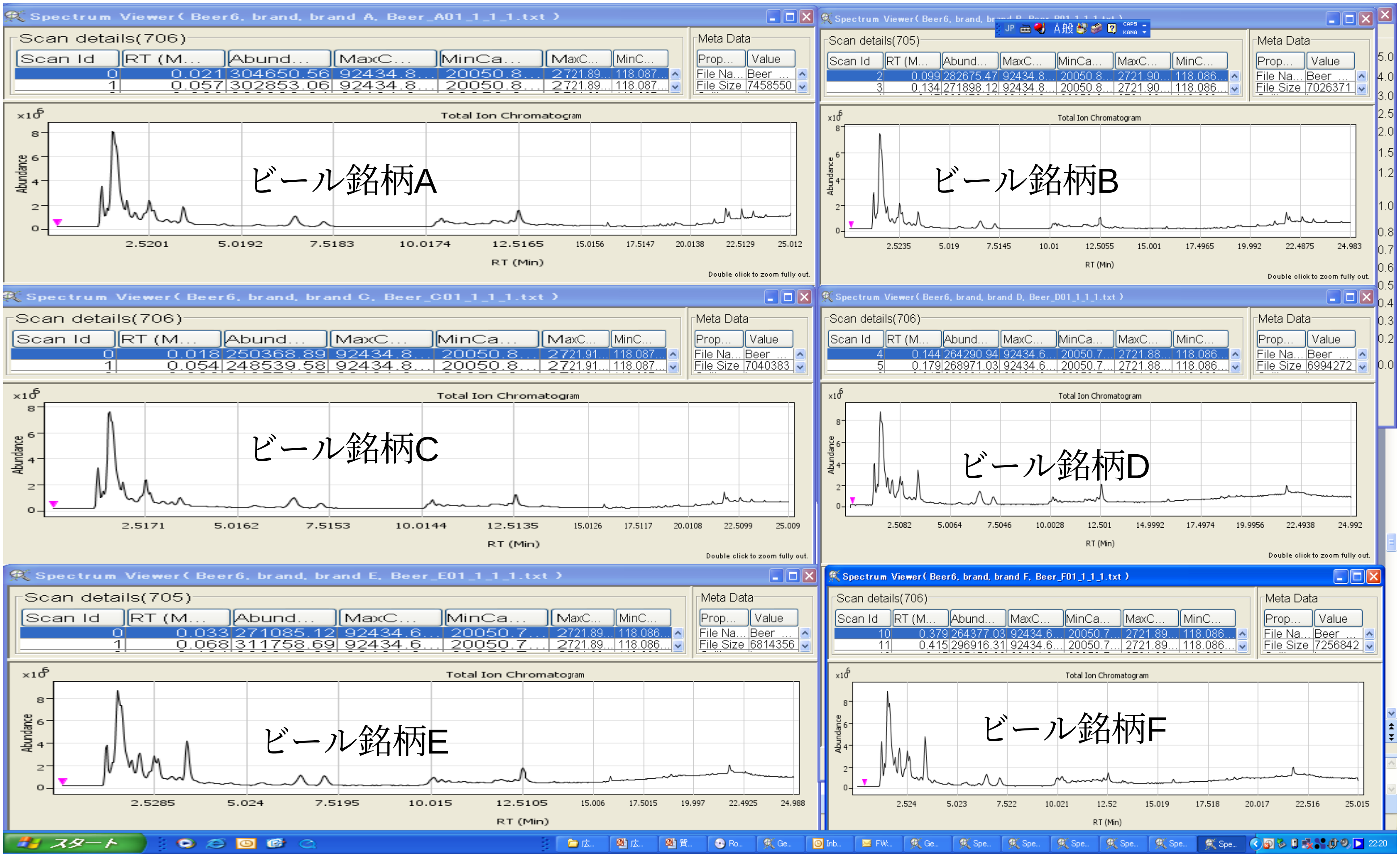
tips

ビール6銘柄の内訳

今回用いるビール6銘柄（レプリケートは3回測定）は、
1.A、B、Cが通常のビール（一般的なビール）
2.D、Eはプレミア系ビール（高級感を味わえるビール）
3.Fは黒ビール（黒ビール）
であることを覚えておいて下さい。

（6銘柄 x 3 Replicates = 18サンプルの結果）

Beer6銘柄(ｘ3replicates) トータルイオンクロマトグラム



Spectrum Viewer(Beer6, brand, brand A, Beer_A01_1_1_1.txt)

Scan details(706)

Scan Id	RT (M...	Abund...	MaxC...	MinCa...	MaxC...	MinC...
0	0.021	304650.56	92434.8...	20050.8...	2721.89...	118.087...
1	0.057	302853.06	92434.8...	20050.8...	2721.89...	118.087...

Meta Data

Prop...	Value
File Na...	Beer ...
File Size	7458550

Spectrum Viewer(Beer6, brand, br

Scan details(705)

Scan Id	RT (M...	Abund...	MaxC...	MinCa...	MaxC...	MinC...
2	0.099	282675.47	92434.8...	20050.8...	2721.90...	118.086...
3	0.134	271898.12	92434.8...	20050.8...	2721.90...	118.086...

Meta Data

Prop...	Value
File Na...	Beer ...
File Size	7026371

Spectrum Viewer(Beer6, brand, brand C, Beer_C01_1_1_1.txt)

Scan details(706)

Scan Id	RT (M...	Abund...	MaxC...	MinCa...	MaxC...	MinC...
0	0.018	250368.89	92434.8...	20050.8...	2721.91...	118.087...
1	0.054	248539.58	92434.8...	20050.8...	2721.91...	118.087...

Meta Data

Prop...	Value
File Na...	Beer ...
File Size	7040383

Spectrum Viewer(Beer6, brand, brand D, Beer_D01_1_1_1.txt)

Scan details(706)

Scan Id	RT (M...	Abund...	MaxC...	MinCa...	MaxC...	MinC...
4	0.144	264290.94	92434.6...	20050.7...	2721.88...	118.086...
5	0.179	268971.03	92434.6...	20050.7...	2721.88...	118.086...

Meta Data

Prop...	Value
File Na...	Beer ...
File Size	6994272

Spectrum Viewer(Beer6, brand, brand E, Beer_E01_1_1_1.txt)

Scan details(705)

Scan Id	RT (M...	Abund...	MaxC...	MinCa...	MaxC...	MinC...
0	0.033	271085.12	92434.6...	20050.7...	2721.89...	118.086...
1	0.068	311758.69	92434.6...	20050.7...	2721.89...	118.086...

Meta Data

Prop...	Value
File Na...	Beer ...
File Size	6814356

Spectrum Viewer(Beer6, brand, brand F, Beer_F01_1_1_1.txt)

Scan details(706)

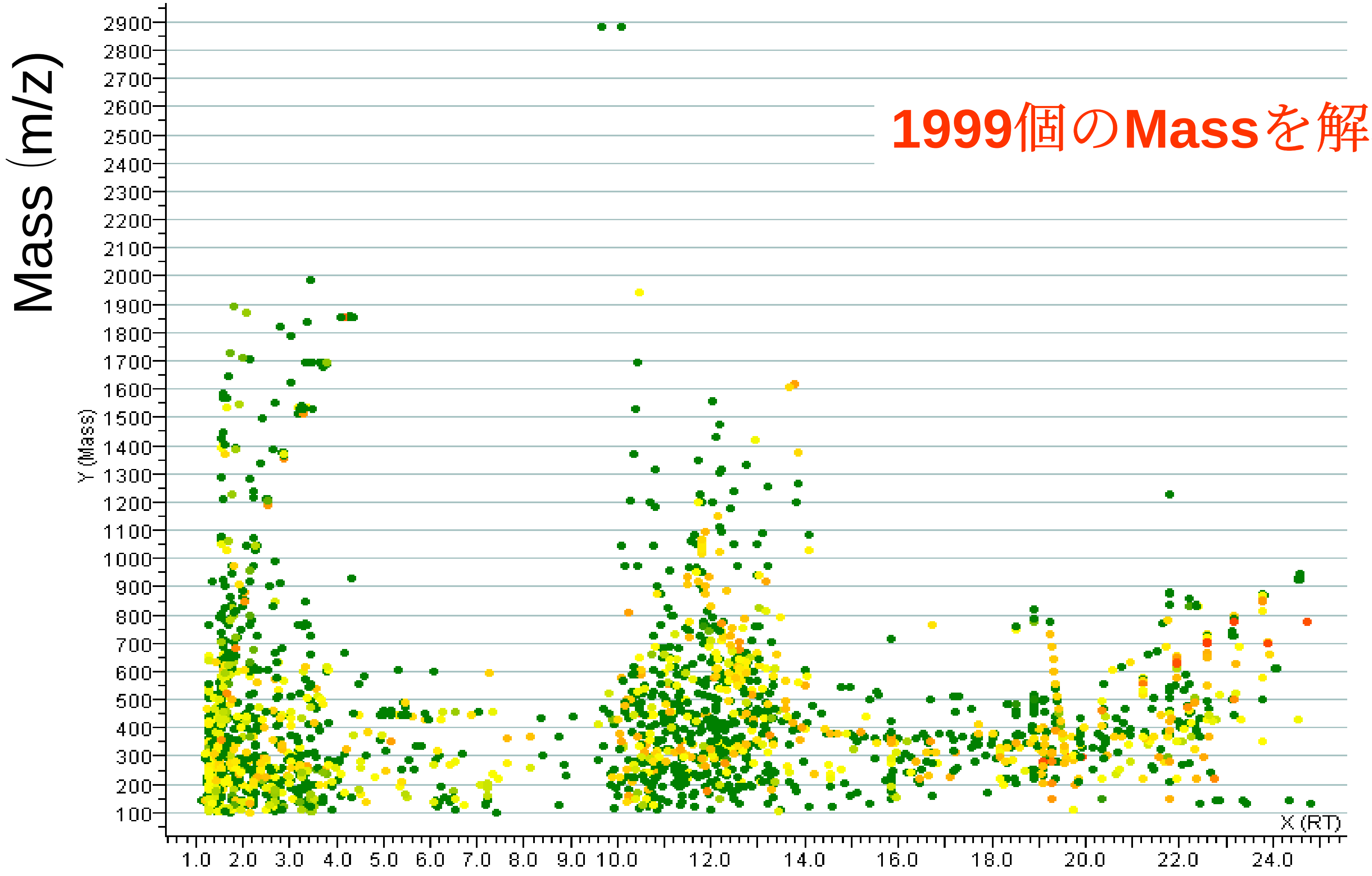
Scan Id	RT (M...	Abund...	MaxC...	MinCa...	MaxC...	MinC...
10	0.379	264377.03	92434.6...	20050.7...	2721.89...	118.086...
11	0.415	296916.31	92434.6...	20050.7...	2721.89...	118.086...

Meta Data

Prop...	Value
File Na...	Beer ...
File Size	7256842

Agilent Technologies

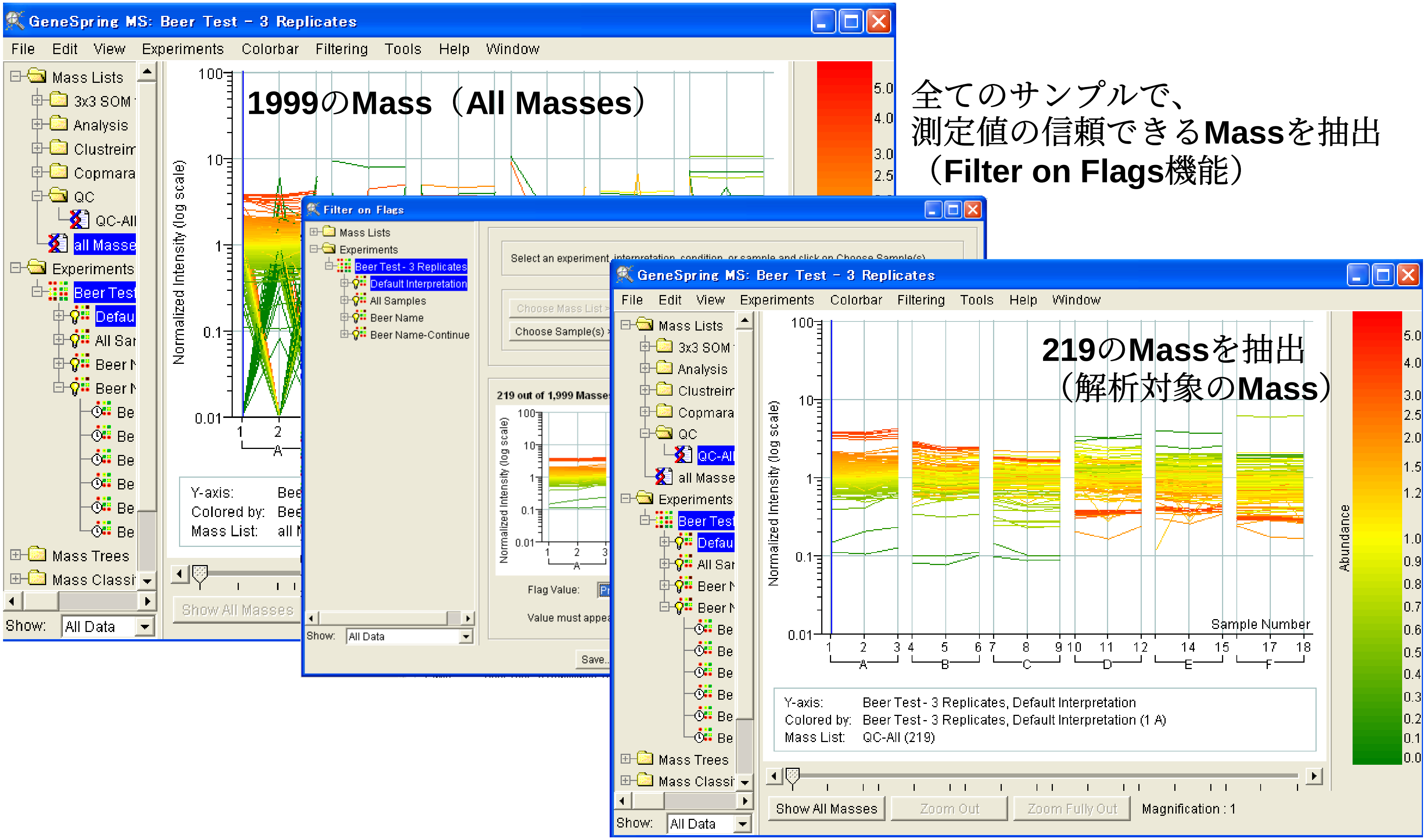
RT vs Mass Plot



1999個のMassを解析対象

RT（保持時間）

データ解析の準備（Quality Control）



1999のMass（All Masses）

Y-axis: Bee
Colored by: Bee
Mass List: all M

Select an experiment, interpretation, condition, or sample and click on Choose Sample(s)

219 out of 1,999 Masses

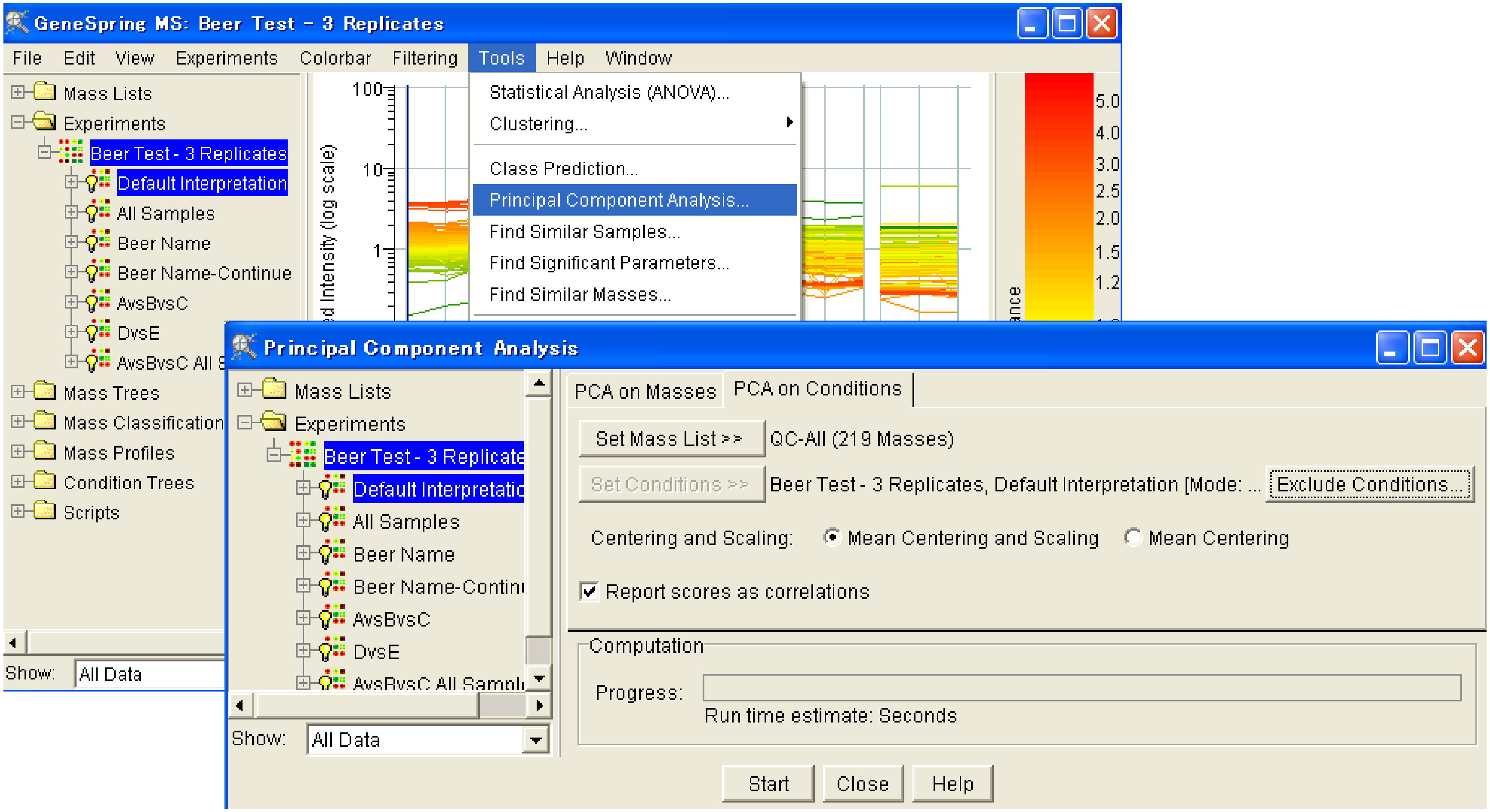
Value must appea

219のMassを抽出（解析対象のMass）

Y-axis: Beer Test - 3 Replicates, Default Interpretation
Colored by: Beer Test - 3 Replicates, Default Interpretation (1 A)
Mass List: QC-All (219)

全てのサンプルで、
測定値の信頼できる**Mass**を抽出
（**Filter on Flags**機能）

主成分分析 (PCA)



Set Mass List >> QC-All (219 Masses)

Set Conditions >> Beer Test - 3 Replicates, Default Interpretation [Mode: ...

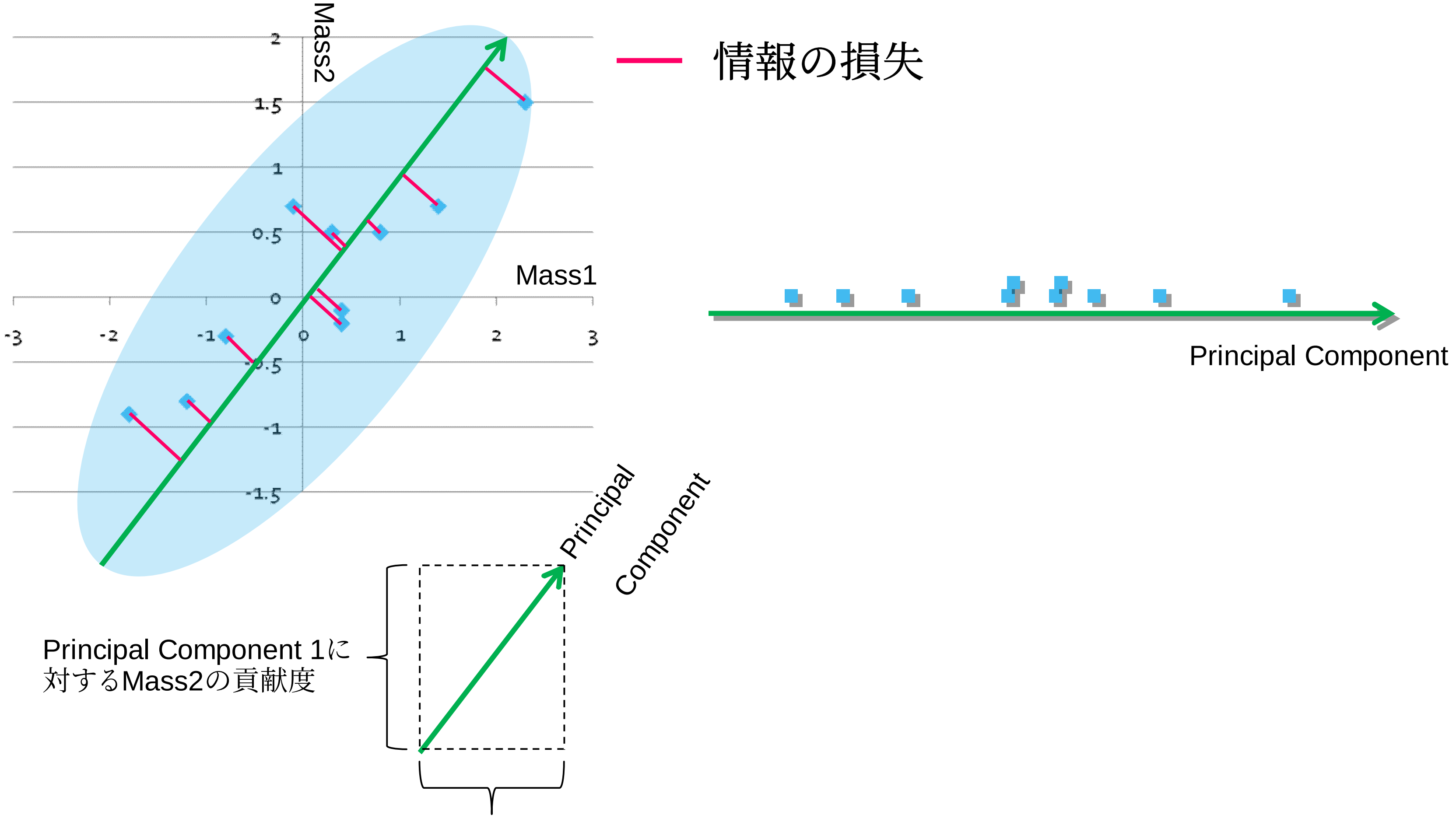
Centering and Scaling: ● Mean Centering and Scaling ○ Mean Centering

☑ Report scores as correlations

Progress:

Run time estimate: Seconds

主成分分析 (PCA)

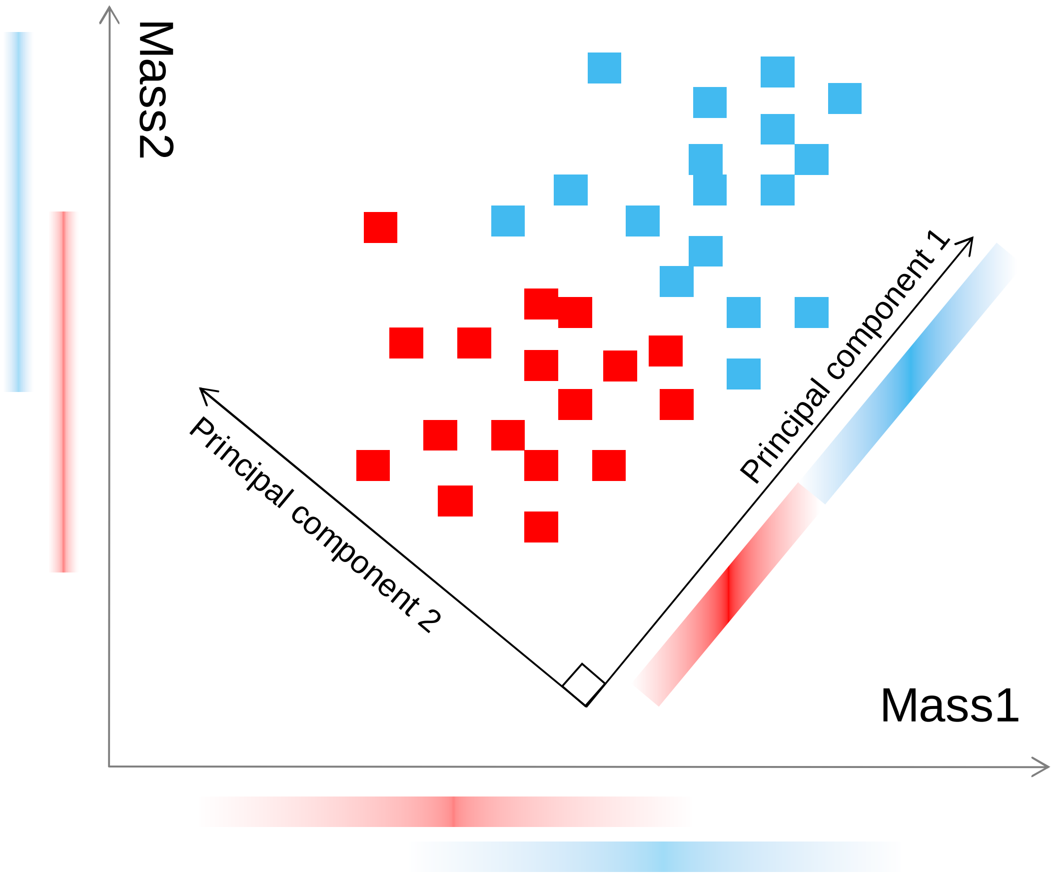


— 情報の損失

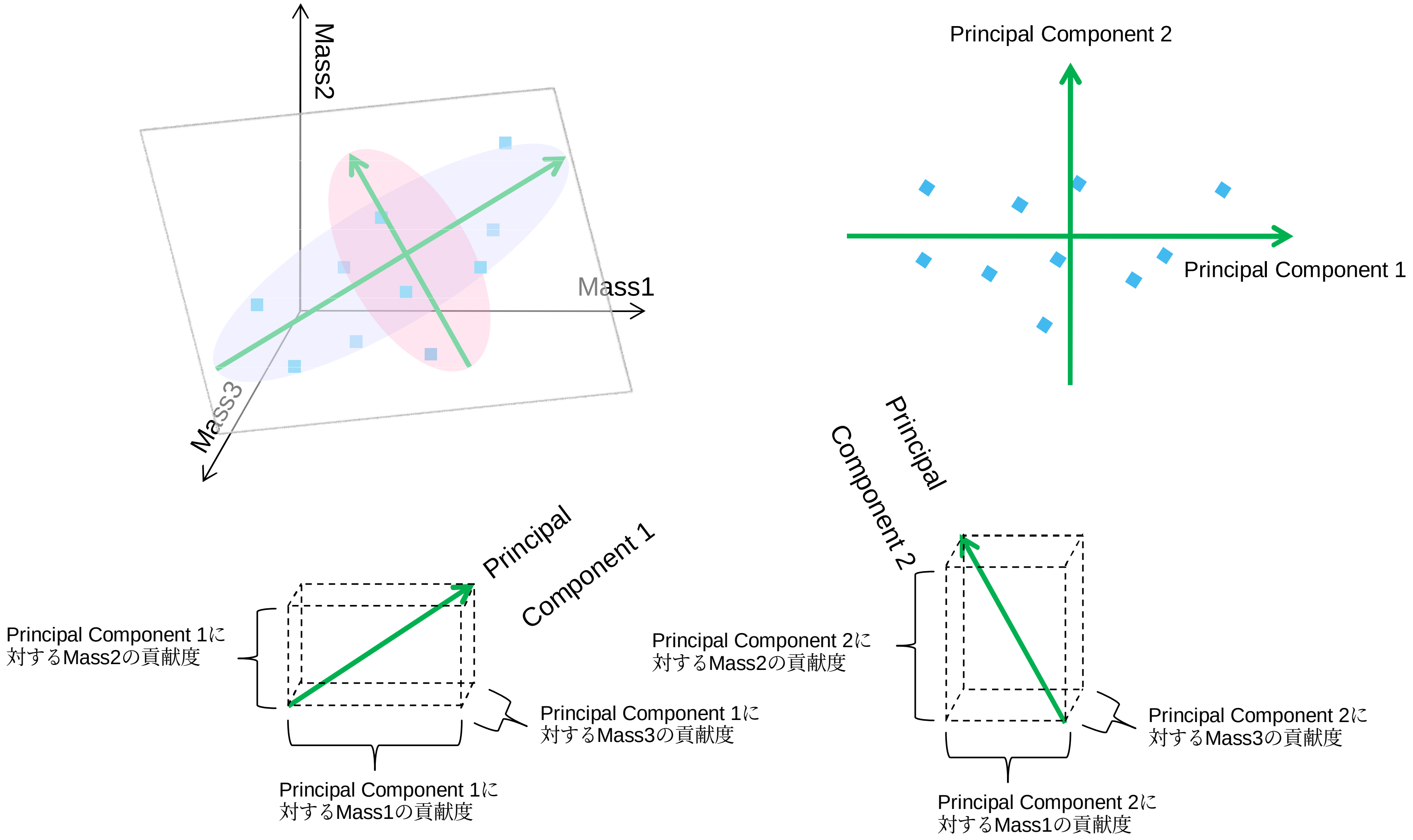
Principal Component 1に
対するMass2の貢献度

Principal Component 1に
たいするMass1の貢献度

PCAのメリット



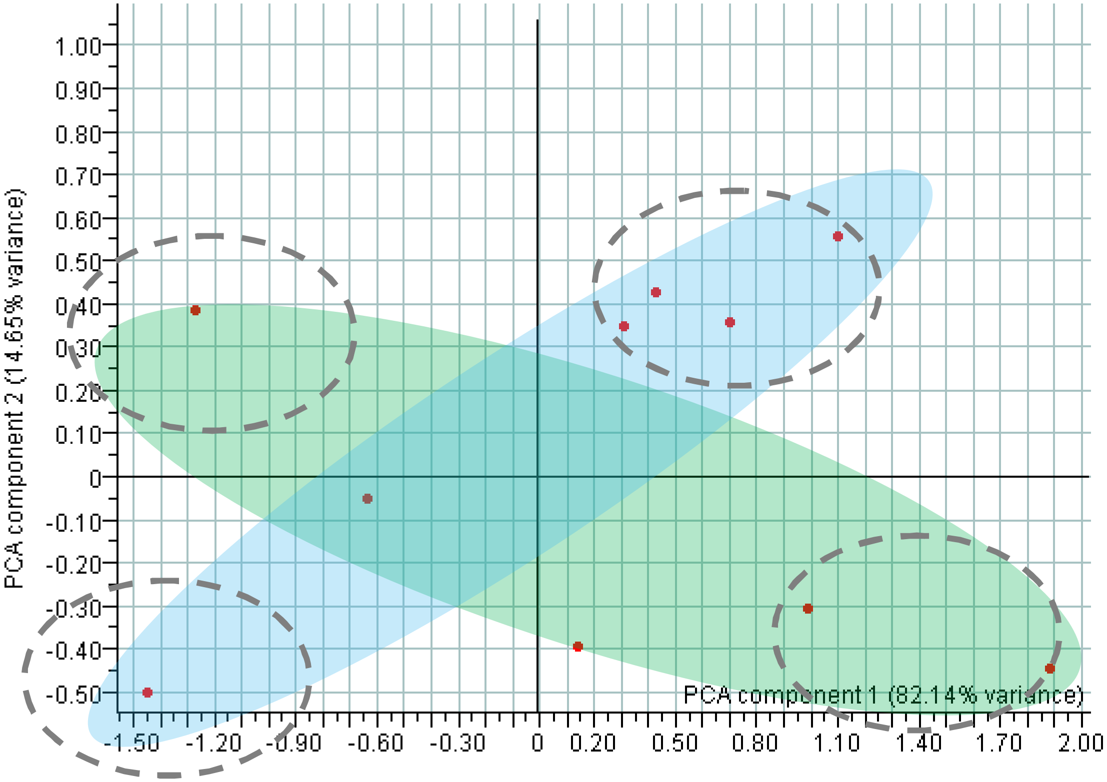
多次元空間におけるPCA



Agilent Technologies

tips

PCAの限界

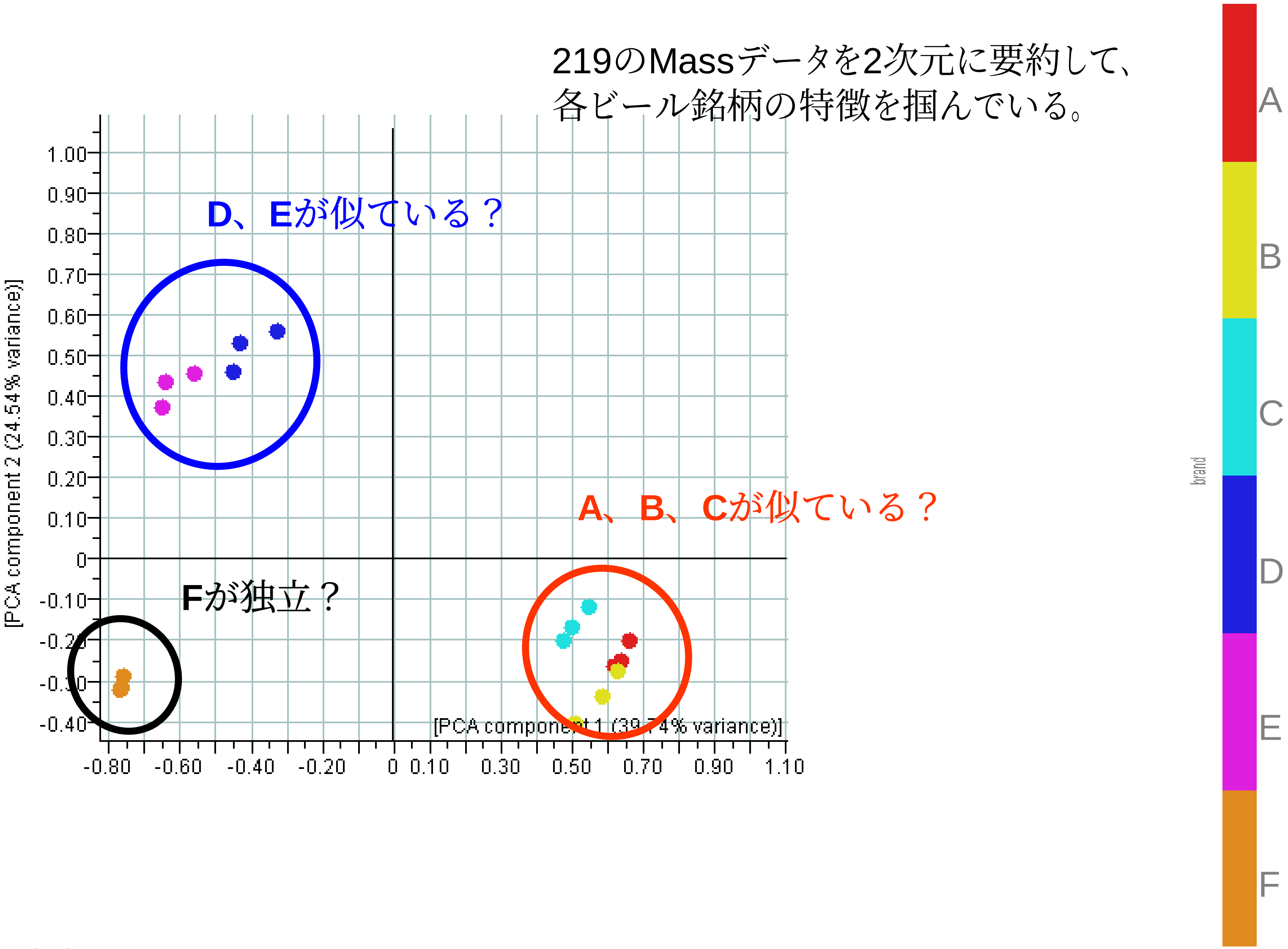


- 真のComponentが直行しない場合、PCAは本質的なcomponentを検出することに失敗する。
- クラスタリング解析によって4つのクラスターを検出することは可能。

tips

PCA（主成分分析）- 2D表示

219のMassデータを2次元に要約して、
各ビール銘柄の特徴を掴んでいる。



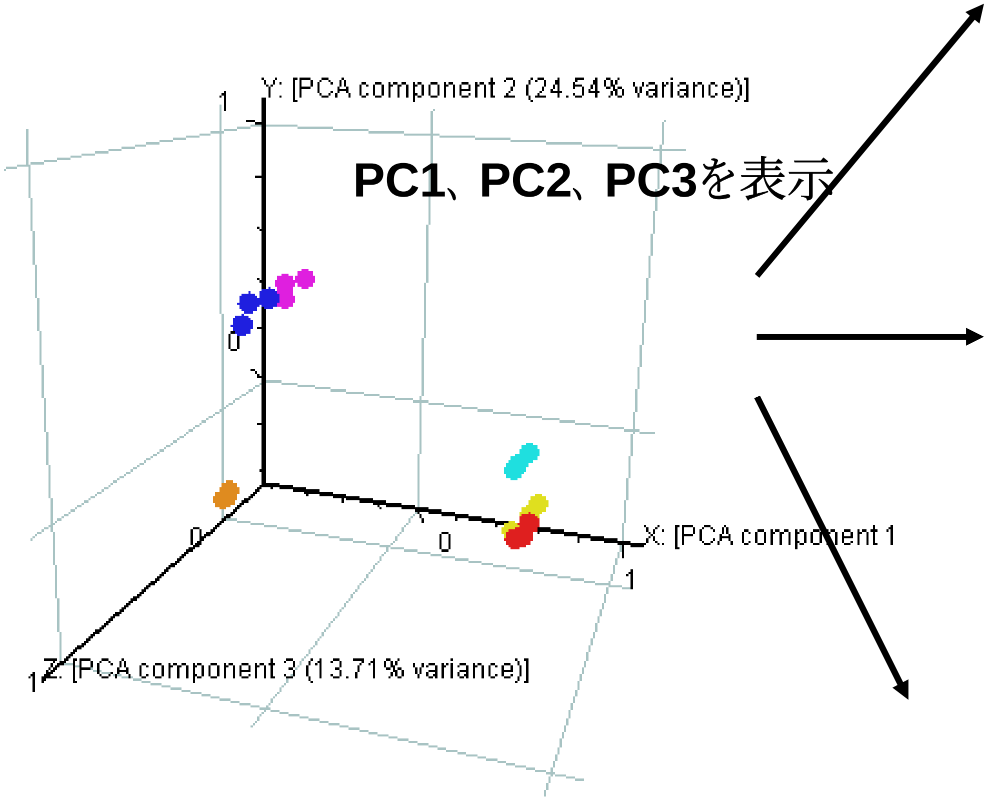
D、Eが似ている？

A、B、Cが似ている？

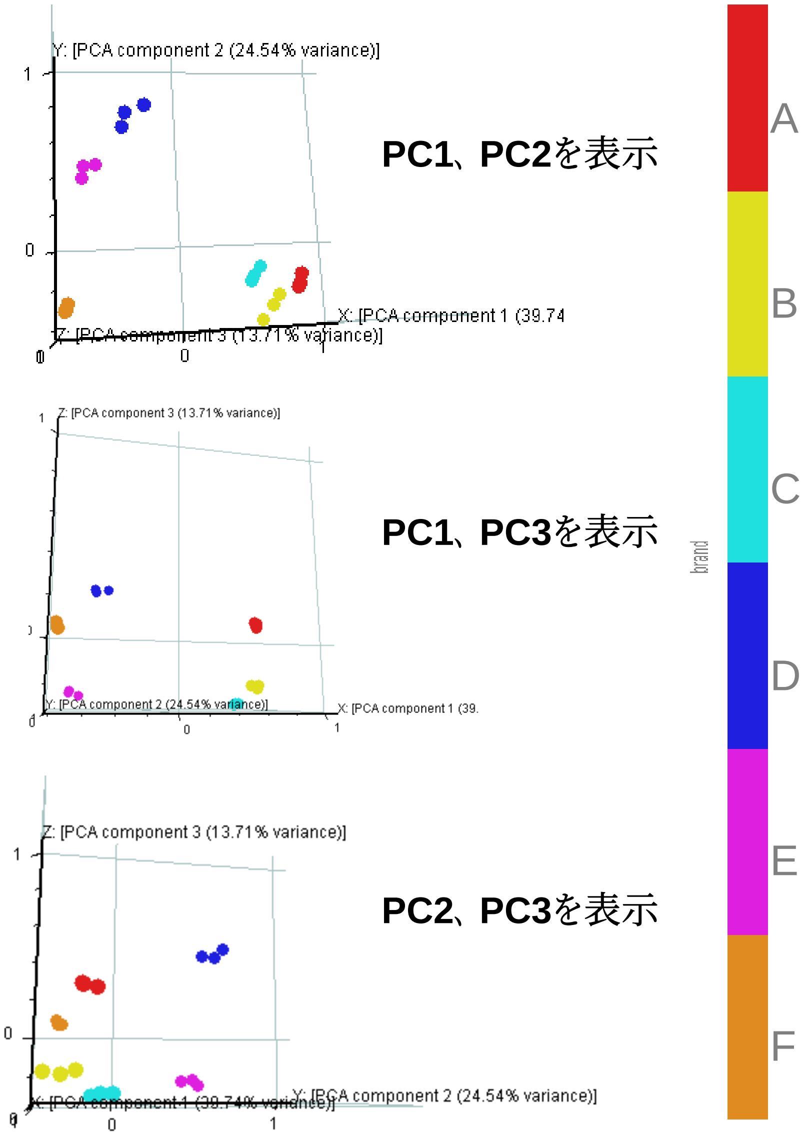
Fが独立？

PCA（主成分分析）- 3D表示

219のMassデータを3次元に要約して、
各ビール銘柄の特徴を掴んでいる。



PC1、PC2、PC3を表示

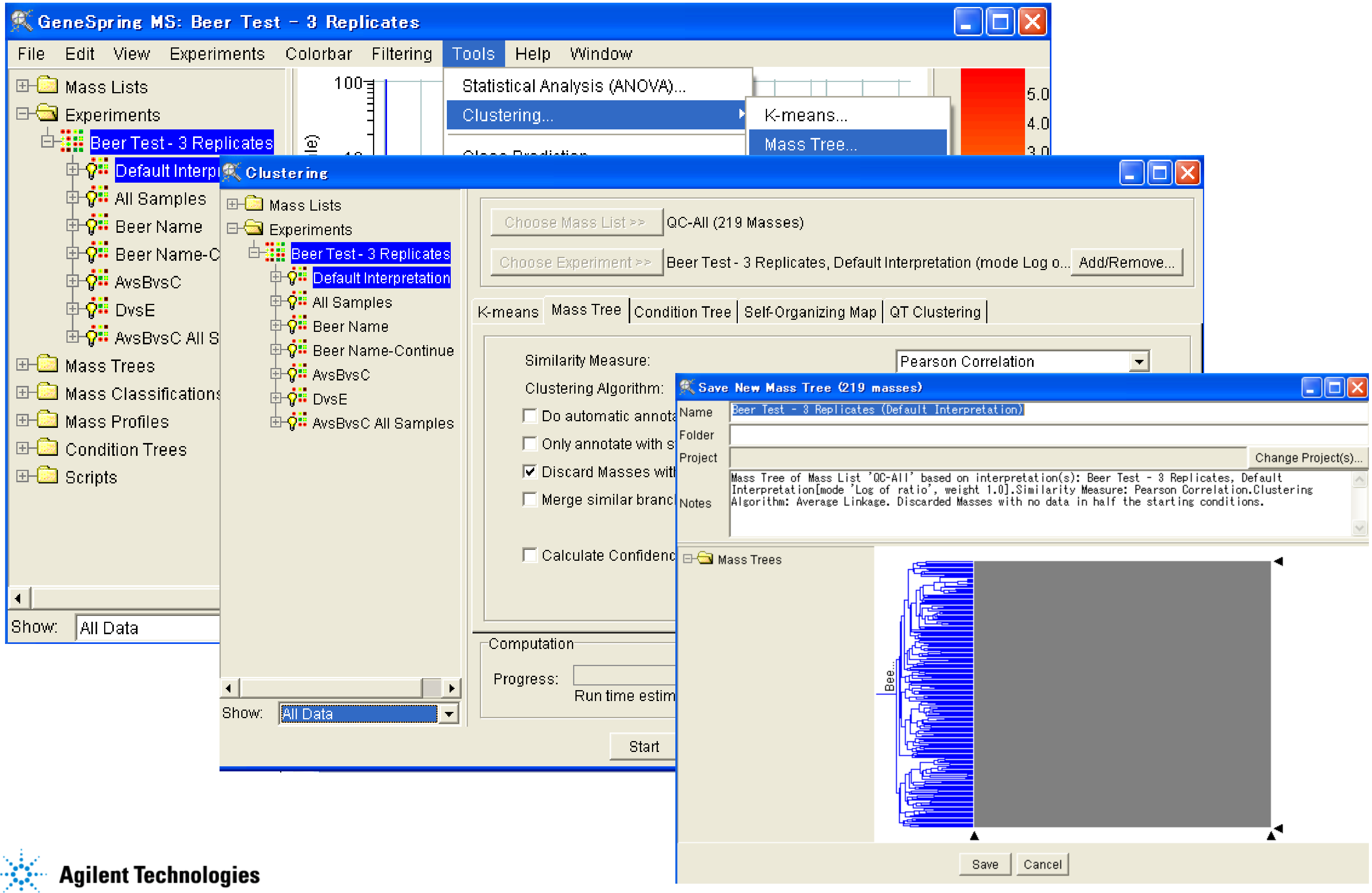


PC1、PC2を表示

PC1、PC3を表示

PC2、PC3を表示

Clustering



QC-All (219 Masses)

Beer Test - 3 Replicates, Default Interpretation (mode Log o...

Similarity Measure: Pearson Correlation

Clustering Algorithm:

☐ Do automatic annota

☐ Only annotate with s

☑ Discard Masses wit

☐ Merge similar branc

☐ Calculate Confidenc

Name: Beer Test - 3 Replicates (Default Interpretation)

Folder:

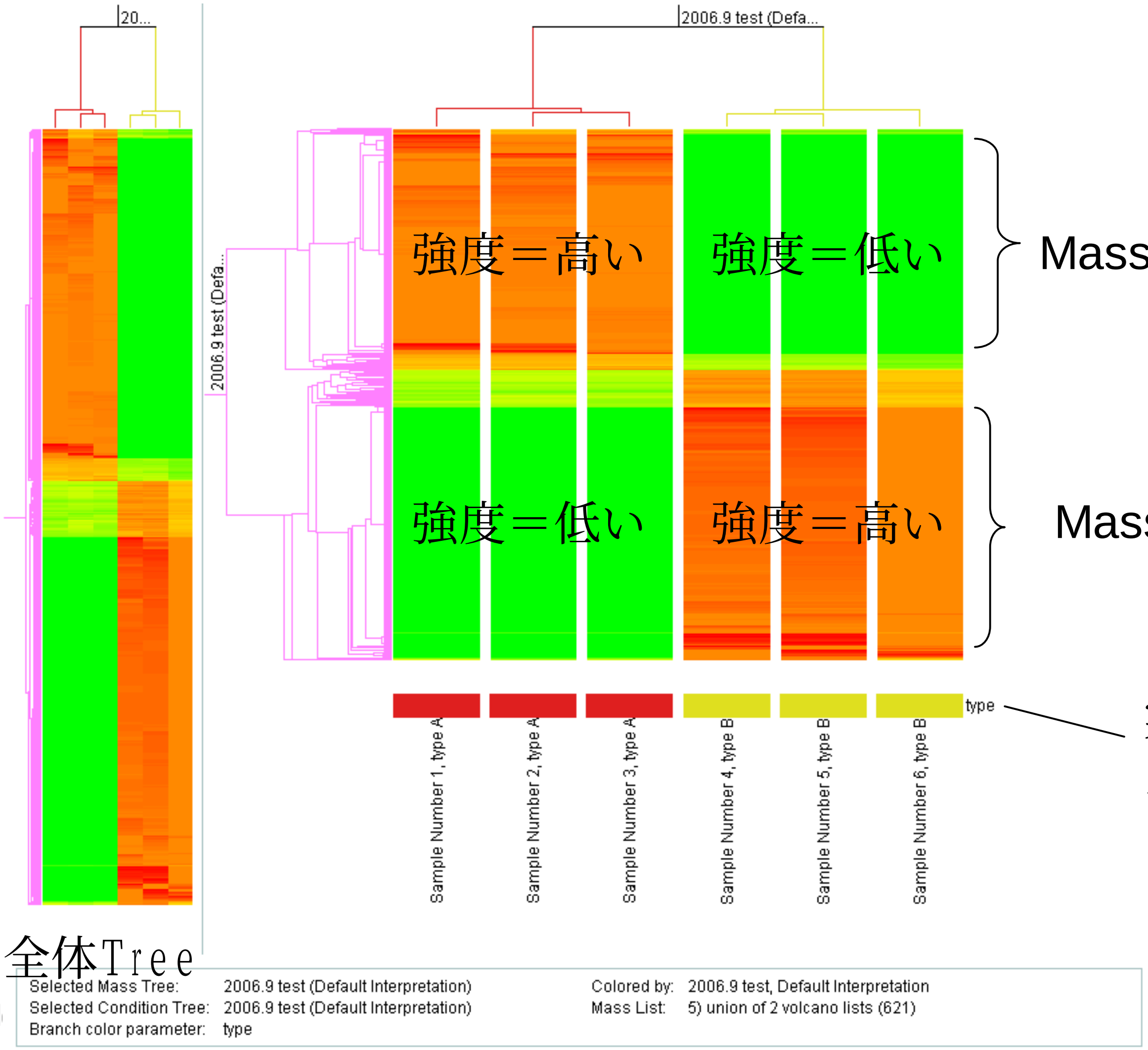
Project:

Notes: Mass Tree of Mass List 'QC-All' based on interpretation(s): Beer Test - 3 Replicates, Default Interpretation[mode 'Log of ratio', weight 1.0].Similarity Measure: Pearson Correlation.Clustering Algorithm: Average Linkage. Discarded Masses with no data in half the starting conditions.

Agilent Technologies

Clusteringの見方

実験セット全体に対して強度が似た動きをする（同じパターン）の化合物同士を分類する。



強度＝高い　強度＝低い

Mass Features Cluster

強度＝低い　強度＝高い

Mass Feature Cluster

実験のグルーピングを
カラータイルで表示

全体Tree

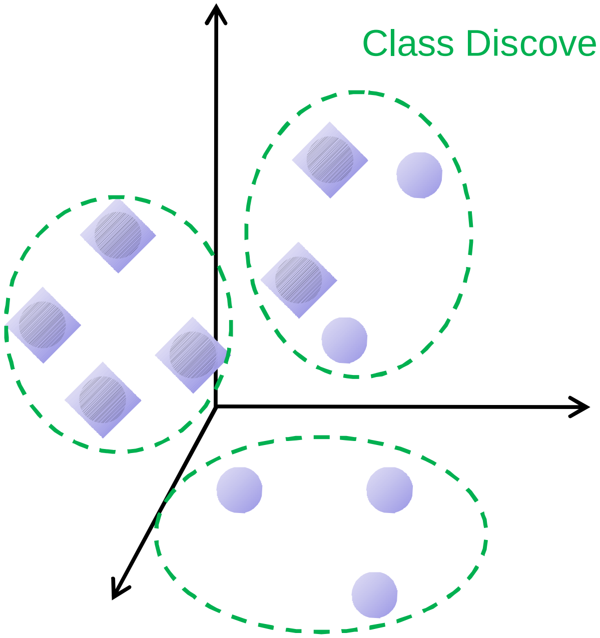
Selected Mass Tree: 2006.9 test (Default Interpretation)
Selected Condition Tree: 2006.9 test (Default Interpretation)
Branch color parameter: type
Colored by: 2006.9 test, Default Interpretation
Mass List: 5) union of 2 volcano lists (621)

tips

プロファイルの幾何学的イメージ

Data Matrix of Abundance
(i Masses x j samples)

	Sample 1	Sample 2	…	Sample j
Mass 1	5000	2700	…	1200
Mass 2	15000	0	…	60000
Mass 3	550	600	…	0
⋮	⋮	⋮	⋮	⋮
Mass i	1400	800	…	17000



サンプルの解析をする場合

i-次元の空間に、 j 個のベクトル(座標)

Mass（成分）の解析をする場合

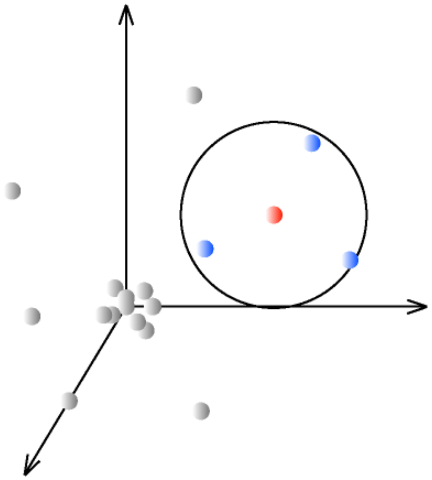
j-次元の空間に、 i 個のベクトル(座標)

tips

近似度の測定方法

1. 距離で測る

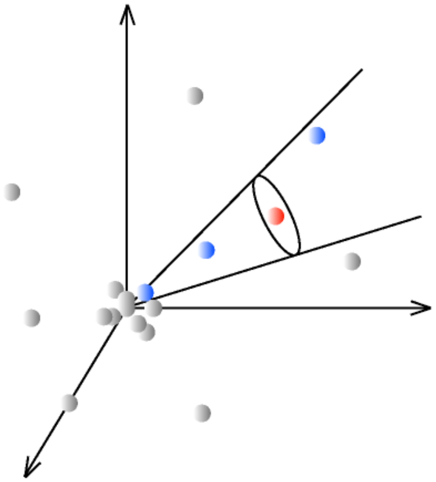
Smaller distance means more similar.



➡ ユークリッド距離など

2. 角度で測る

Smaller angle means more similar.
Usually calculate cosine of the angle.



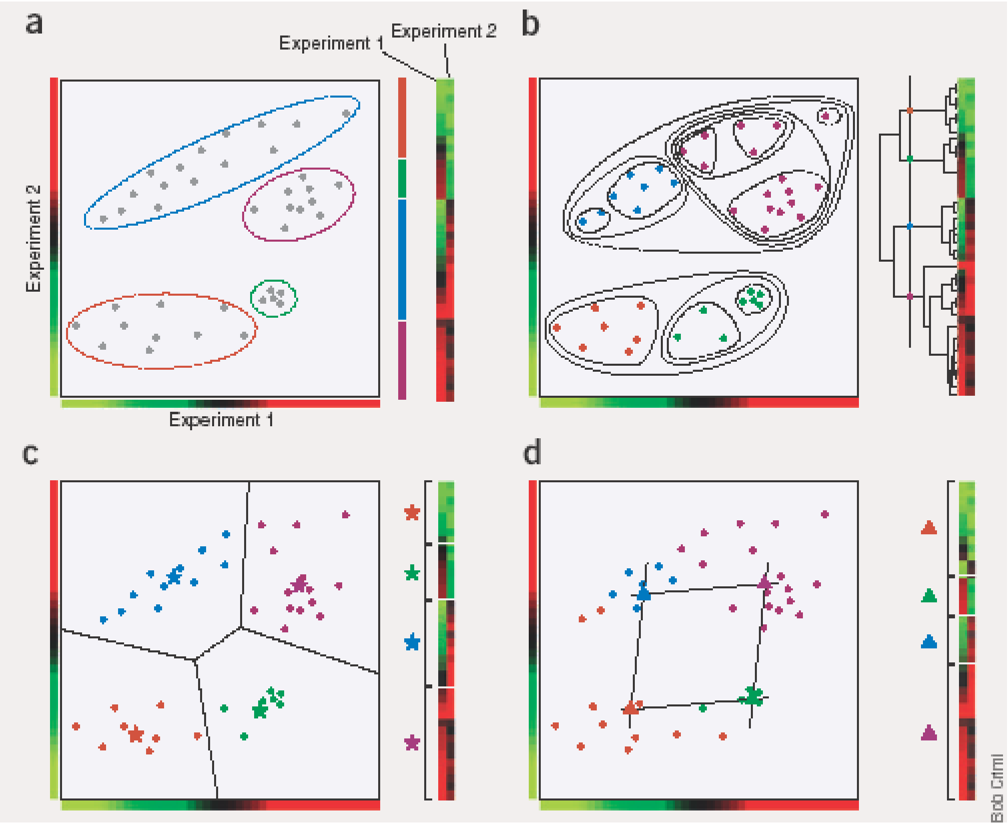
➡ 相関係数など

"How does gene expression clustering work?"
2005 D'haeseleer, *Nature Biotechnology* ; 23,12:1499-501

tips

Agilent Technologies

クラスタリング・アルゴリズム



Hierarchical clustering

K-means clustering

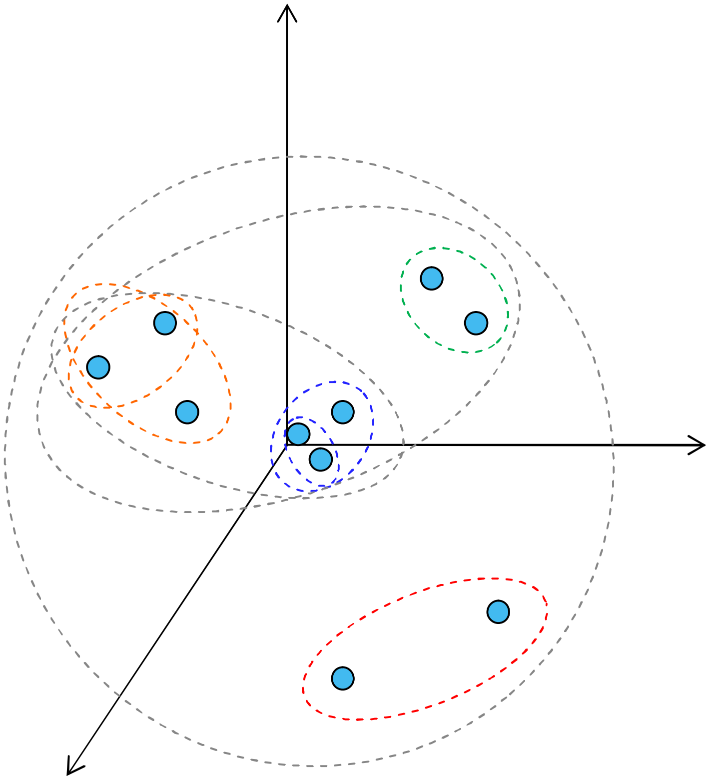
SOM

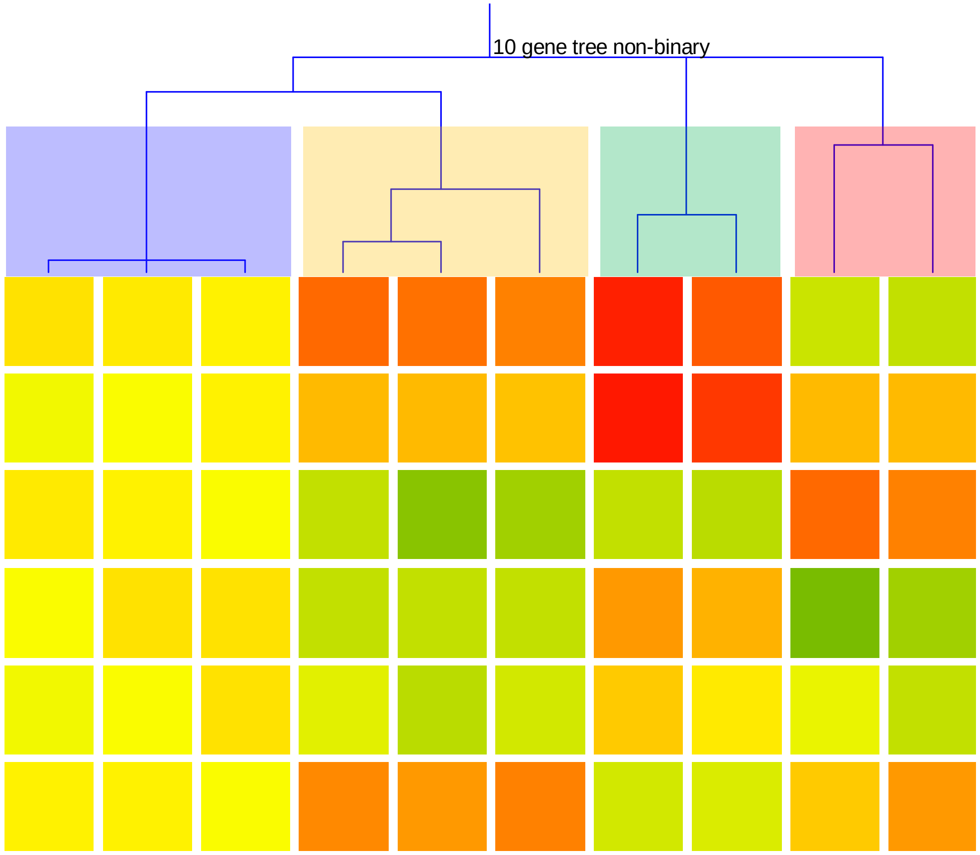
"How does gene expression clustering work?"
2005 D'haeseleer, *Nature Biotechnology* ; 23,12:1499-501

tips

Agilent Technologies

Hierarchical Clustering

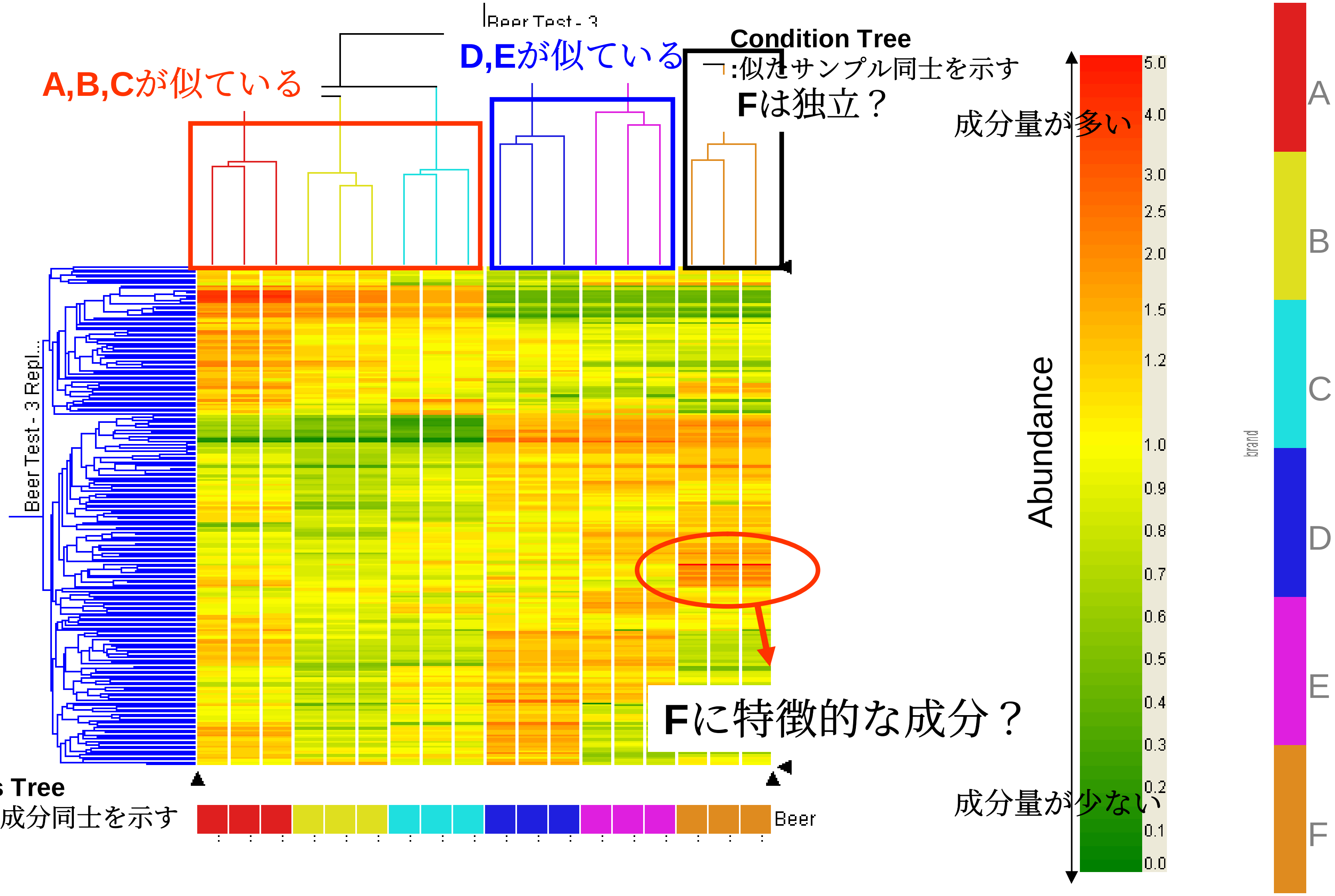




Agilent Technologies

tips

GeneSpring MS Tree-clusterig



A,B,Cが似ている

D,Eが似ている

Condition Tree
:似たサンプル同士を示す

Fは独立？

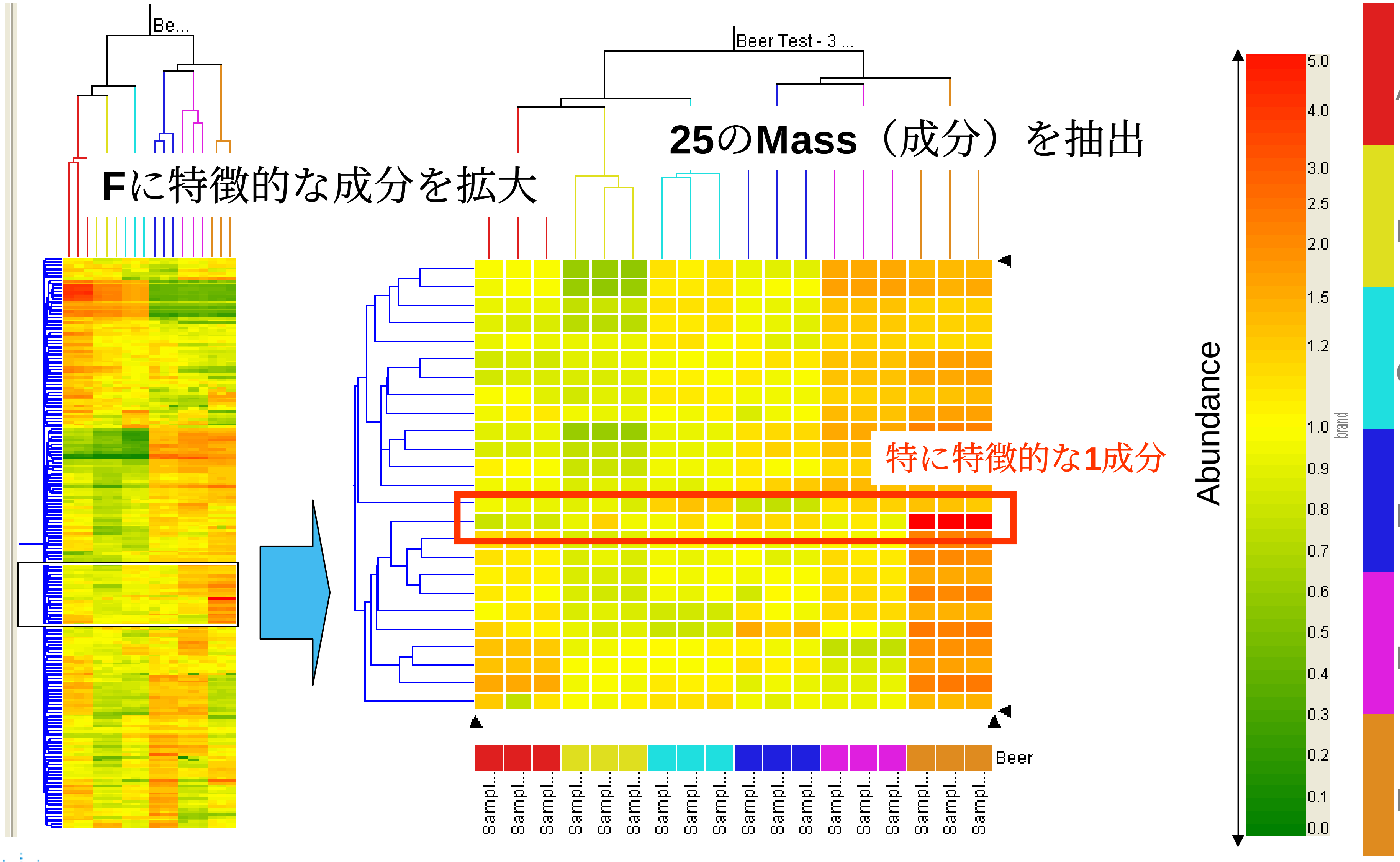
成分量が多い

成分量が少ない

Fに特徴的な成分？

Mass Tree
:似た成分同士を示す

GeneSpring MS Tree-clusterig



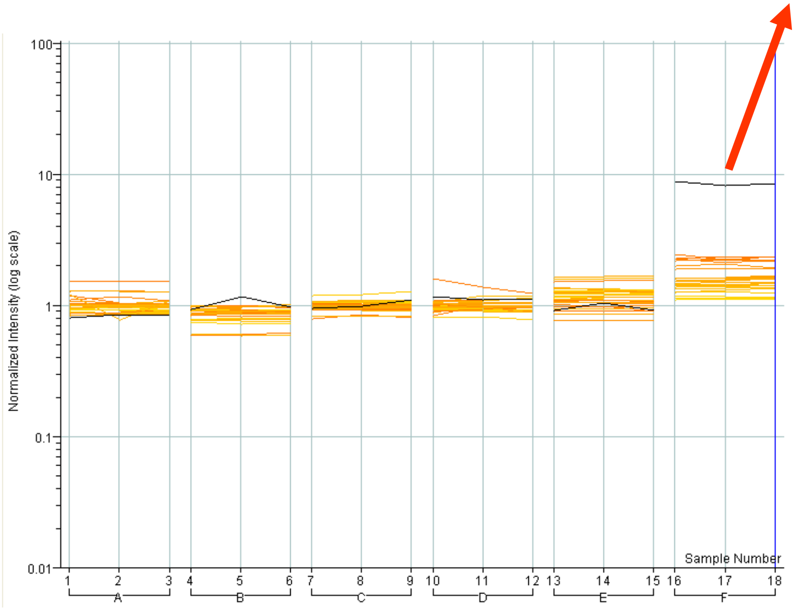
Fに特徴的な成分を拡大

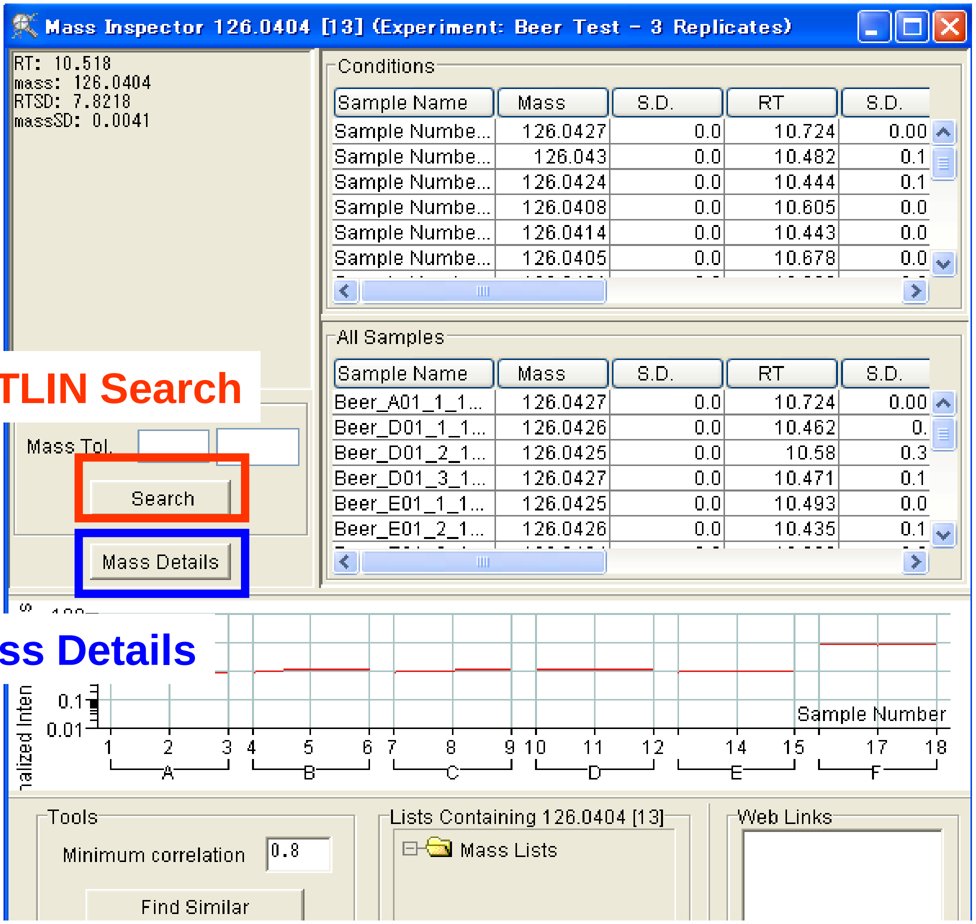
25のMass（成分）を抽出

特に特徴的な1成分

Fに特徴的な成分

特に特徴的な**1**成分はどのような成分なのか？





RT: 10.518
mass: 126.0404
RTSD: 7.8218
massSD: 0.0041

Conditions

Sample Name	Mass	S.D.	RT	S.D.
Sample Numbe...	126.0427	0.0	10.724	0.00
Sample Numbe...	126.043	0.0	10.482	0.1
Sample Numbe...	126.0424	0.0	10.444	0.1
Sample Numbe...	126.0408	0.0	10.605	0.0
Sample Numbe...	126.0414	0.0	10.443	0.0
Sample Numbe...	126.0405	0.0	10.678	0.0

All Samples

Sample Name	Mass	S.D.	RT	S.D.
Beer_A01_1_1...	126.0427	0.0	10.724	0.00
Beer_D01_1_1...	126.0426	0.0	10.462	0.
Beer_D01_2_1...	126.0425	0.0	10.58	0.3
Beer_D01_3_1...	126.0427	0.0	10.471	0.1
Beer_E01_1_1...	126.0425	0.0	10.493	0.0
Beer_E01_2_1...	126.0426	0.0	10.435	0.1

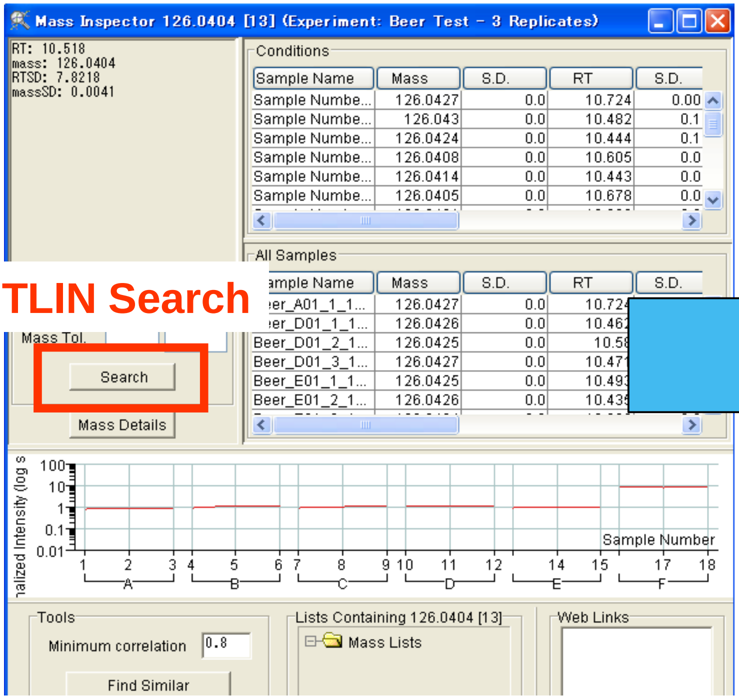
Minimum correlation 0.8

METLIN Search

Mass Details

Agilent Technologies

METLIN Searchで候補を調べる



Mass Inspector 126.0404 [13] (Experiment: Beer Test - 3 Replicates)

```
RT: 10.518
mass: 126.0404
RTSD: 7.8218
massSD: 0.0041
```

Conditions

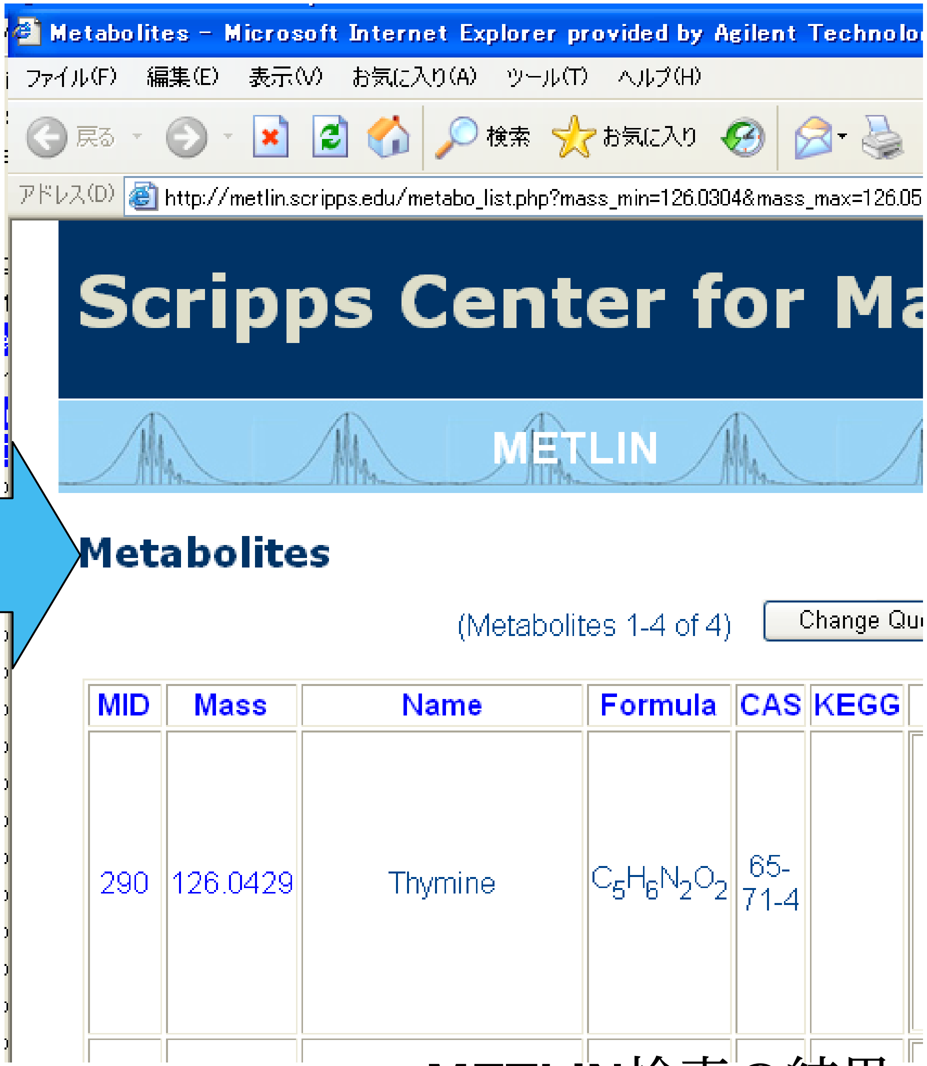
Sample Name	Mass	S.D.	RT	S.D.
Sample Numbe...	126.0427	0.0	10.724	0.00
Sample Numbe...	126.043	0.0	10.482	0.1
Sample Numbe...	126.0424	0.0	10.444	0.1
Sample Numbe...	126.0408	0.0	10.605	0.0
Sample Numbe...	126.0414	0.0	10.443	0.0
Sample Numbe...	126.0405	0.0	10.678	0.0

All Samples

ample Name	Mass	S.D.	RT	S.D.
eer_A01_1_1...	126.0427	0.0	10.724	
eer_D01_1_1...	126.0426	0.0	10.462	
Beer_D01_2_1...	126.0425	0.0	10.58	
Beer_D01_3_1...	126.0427	0.0	10.471	
Beer_E01_1_1...	126.0425	0.0	10.493	
Beer_E01_2_1...	126.0426	0.0	10.435	

Minimum correlation: 0.8

METLIN Search



http://metlin.scripps.edu/metabo_list.php?mass_min=126.0304&mass_max=126.05

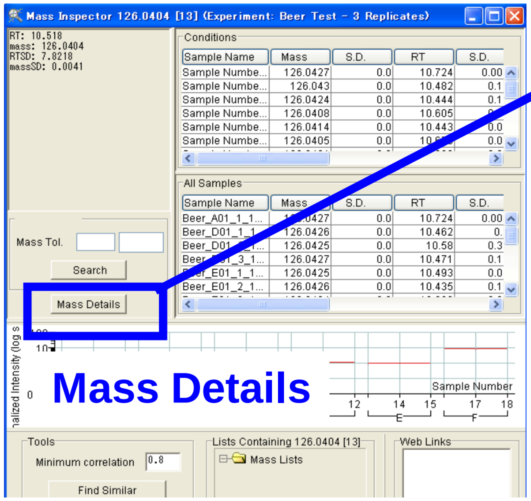
Metabolites

(Metabolites 1-4 of 4)

MID	Mass	Name	Formula	CAS	KEGG
290	126.0429	Thymine	$C_5H_6N_2O_2$	65-71-4	

METLIN検索の結果

Mass Detailsで詳細を調べる



RT: 10.518
mass: 126.0404
RTSD: 7.8218
massSD: 0.0041

Conditions

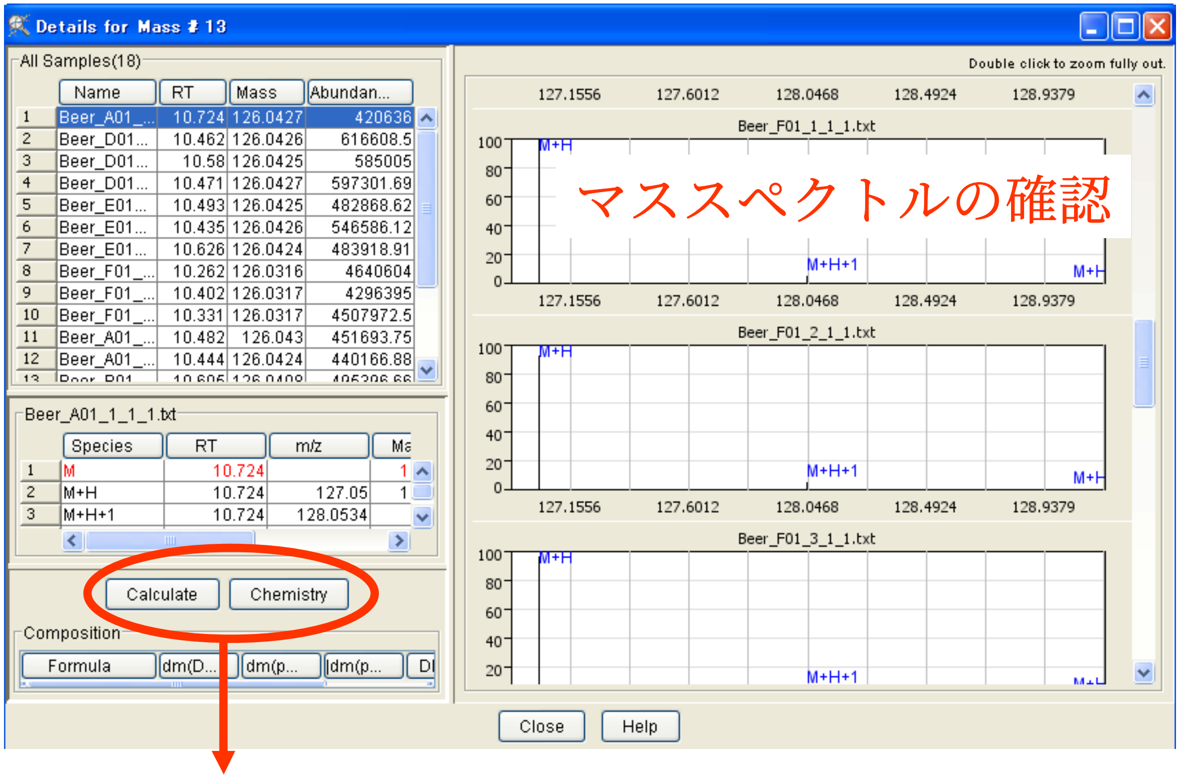
Sample Name	Mass	S.D.	RT	S.D.
Sample Numbe...	126.0427	0.0	10.724	0.00
Sample Numbe...	126.043	0.0	10.482	0.1
Sample Numbe...	126.0424	0.0	10.444	0.1
Sample Numbe...	126.0408	0.0	10.605	[illegible]
Sample Numbe...	126.0414	0.0	10.443	0.0
Sample Numbe...	126.0405	0.0	[illegible]	0.0

All Samples

Sample Name	Mass	S.D.	RT	S.D.
Beer_A01_1_1...	[illegible]	0.0	10.724	0.00
Beer_D01_1_1	126.0426	0.0	10.462	0.
Beer_D01_[illegible]	126.0425	0.0	10.58	0.3
Beer_[illegible]_3_1...	126.0427	0.0	10.471	0.1
[illegible]_E01_1_1...	126.0425	0.0	10.493	0.0
Beer_E01_2_1...	126.0426	0.0	10.435	0.1

Minimum correlation: 0.8

Mass Details



All Samples(18)

	Name	RT	Mass	Abundan...
1	Beer_A01_...	10.724	126.0427	420636
2	Beer_D01...	10.462	126.0426	616608.5
3	Beer_D01...	10.58	126.0425	585005
4	Beer_D01...	10.471	126.0427	597301.69
5	Beer_E01...	10.493	126.0425	482868.62
6	Beer_E01...	10.435	126.0426	546586.12
7	Beer_E01...	10.626	126.0424	483918.91
8	Beer_F01_...	10.262	126.0316	4640604
9	Beer_F01_...	10.402	126.0317	4296395
10	Beer_F01_...	10.331	126.0317	4507972.5
11	Beer_A01_...	10.482	126.043	451693.75
12	Beer_A01_...	10.444	126.0424	440166.88

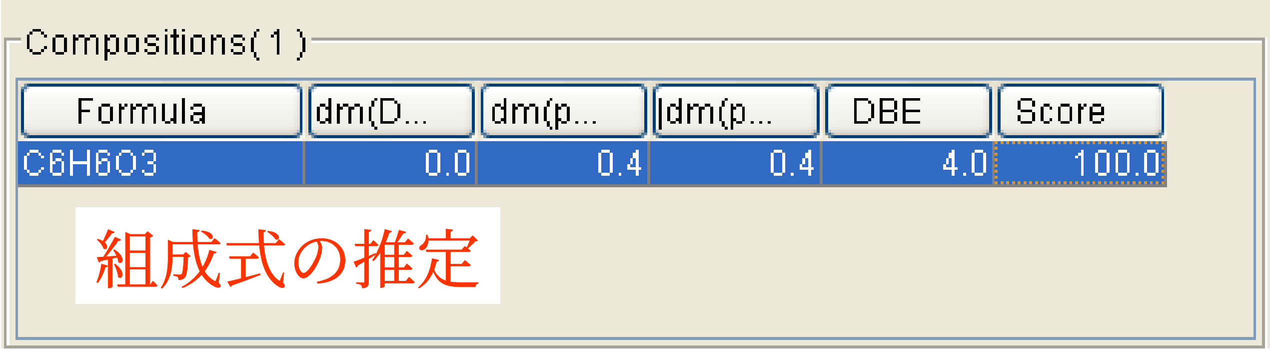
Beer_A01_1_1_1.txt

	Species	RT	m/z	Ma
1	M	10.724		1
2	M+H	10.724	127.05	1
3	M+H+1	10.724	128.0534	

Composition

Formula	dm(D...	dm(p...	|dm(p...	DI

マススペクトルの確認



Compositions(1)

Formula	dm(D...	dm(p...	|dm(p...	DBE	Score
C6H6O3	0.0	0.4	0.4	4.0	100.0

組成式の推定

GeneSpring MS詳細のお問い合わせ先

GeneSpring MS詳細のお問い合わせは、

安藤（**kohei_ando@agilent.com**）までお願い申し上げます。

- 補足情報 -

***GeneSpring MS** 無償デモ版のダウンロードサイト

http://www.chem.agilent.com/scripts/generic.asp?lpage=54770&indcol=Y&prodcol=Y

***GeneSpring MS**紹介サイト

http://www.chem.agilent.com/scripts/pds.asp?lpage=42556