MATRIX SCIENCE K.K.

DIA 定量プロテオーム解析ソフトウェア

Scaffold DDA

MATRIX SCIENCE K.K.

Scaffold DDA　何ができるか

- DDAデータ：同定タンパク質の定性/定量データの取り込み
  ＊MASCOT Server/Distiller,
  Scaffold, etc
- MSFragger と連動したタンパク質の同定・定量
- 定量解析データに基づいた検定とグラフ表示
- GO 情報の表示
- Viewerによる結果ファイルのシェア

解析操作の流れ

- ソフトウェア起動
- 新規作成、Ctrl + N
  （menuの「File」-> New やアイコンクリック）
- 検索パラメータ指定、検索開始
- データ取り込み完了
- 属性付与、階層構造化
- データ解析（定性解析、定量解析、検定）
- レポート

対応フォーマット：

定性[同定]	.dat
	.mzIdentML [Scaffold 5含む]
	.msf
	etc（.mzXML, .d, .tsv …）
定量：isobalic	定性と同じファイル
定量：ラベルフリー	.XML(MASCOT Distiller)
	mzIdentML [Scaffold Q+S 含む]
	.msf

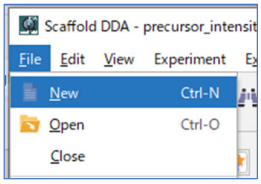
その他のソフトウェアをご利用の場合、英語マニュアルChapter 9 を参照の事

データ取り込み時の画面

データ取り込み開始

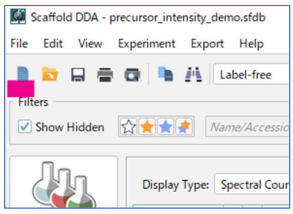
・以下のいずれかの操作を実施

- メニューのFile -> New



- Ctrl + N

- メニューバー下にあるアイコン　をクリック



Load Data

MATRIX SCIENCE K.K.

[1]Probability Mode

取り込んだデータに対してPercolatorの再評価を行うかどうかを選択。

[2] Quantitation Type

- Label Free ［定量計算を行わない時もこちらを選択］
- TMT X-plex (X=2,6,10,11,16)
- iTRAQ Y-plex (Y=4,8)
- SILAC Z-plex (Z=2,3)

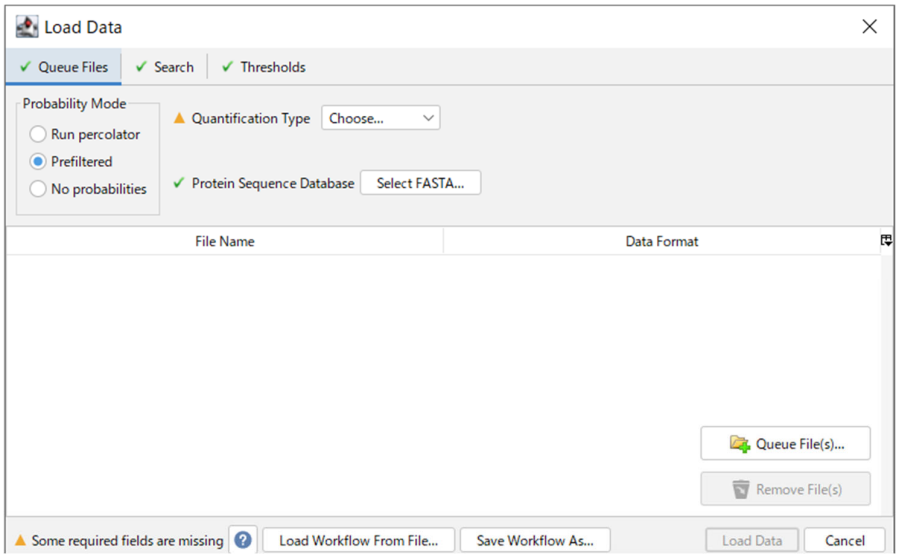
＊Customize可能

[3] Protein Sequence Database

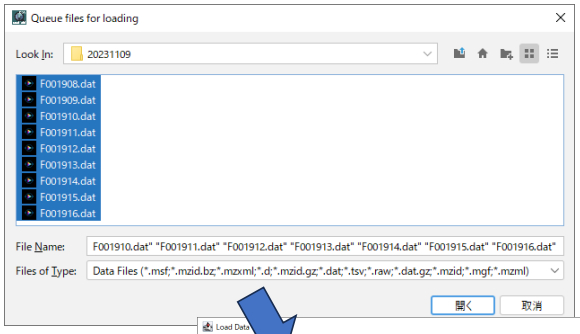
FASTAファイルを選択。

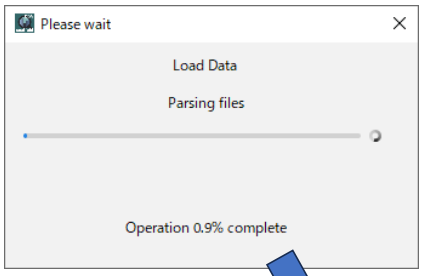
[4] Que Files

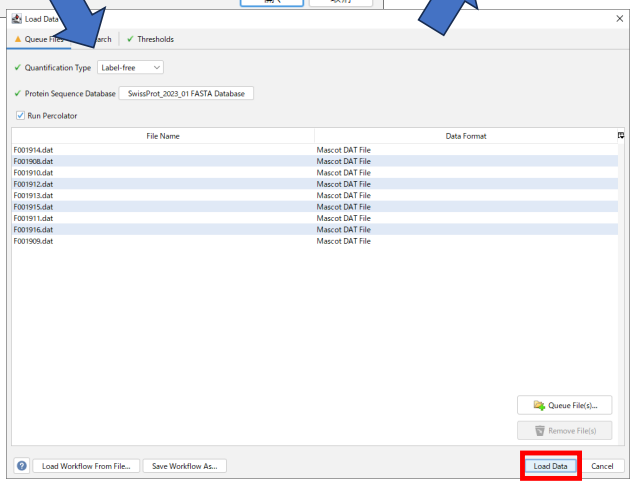
取り込みたいファイルを選択。

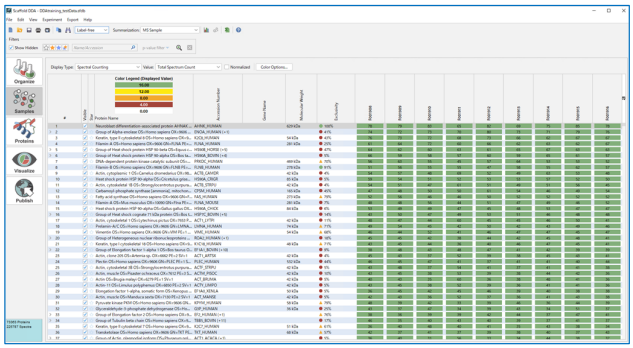


MATRIX SCIENCE K.K.

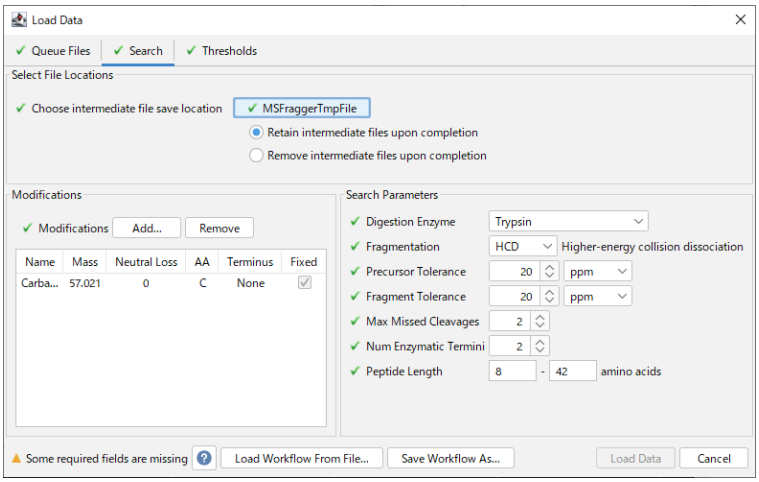


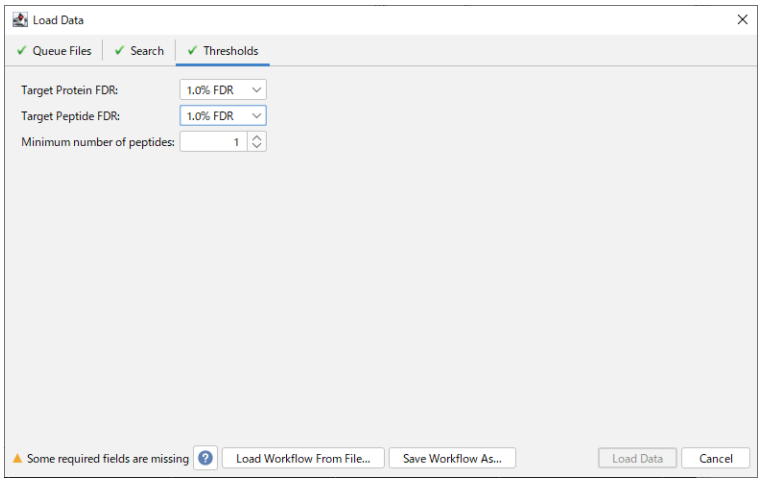






Search タブ、Threshold タブ





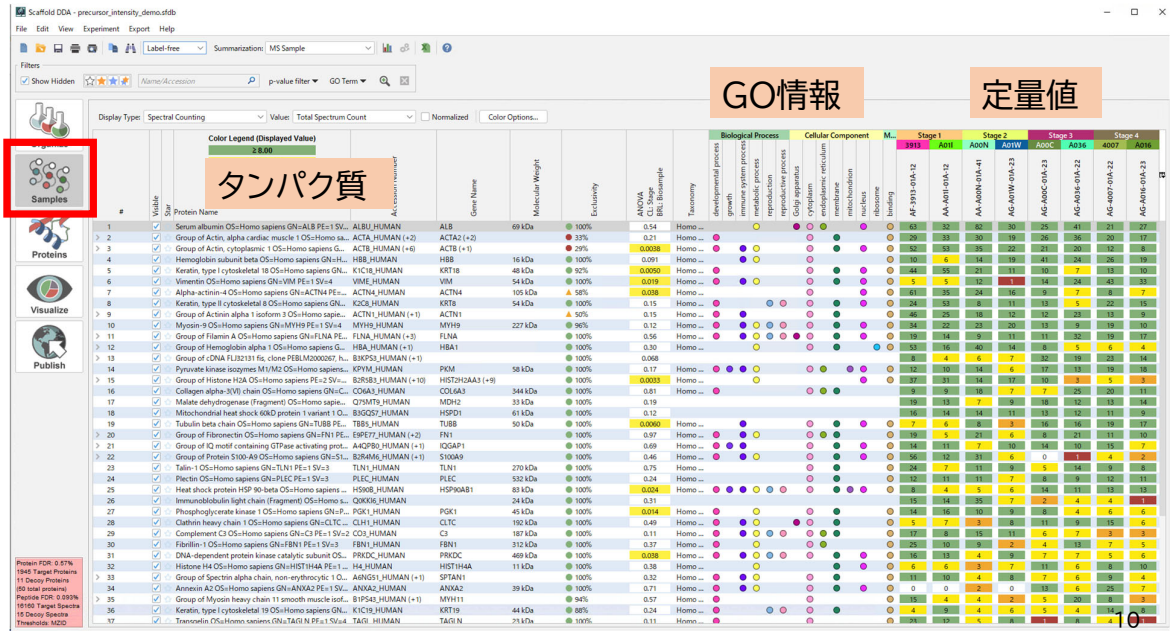
Search タブ
Rawデータの取り込みから行う場合、パラメーターを指定

Thresholds タブ
タンパク質・ペプチドの同定基準

Samples 画面

同定されたタンパク質の一覧並びに定量値をはじめとするタンパク質の関連情報が表示

解析結果の概要を確認する上で主体となる画面



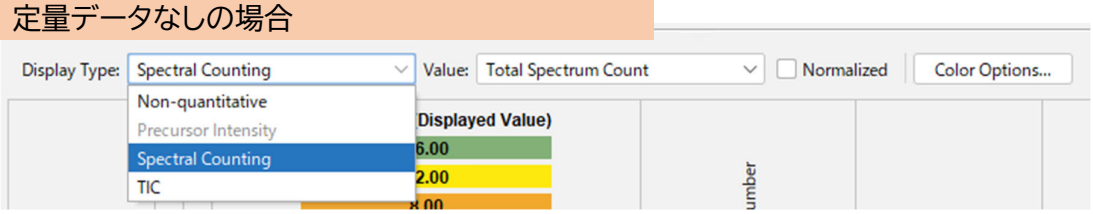
ペプチド・タンパク質の定性(同定)

- ペプチド
  - 各検索エンジンでの評価 [ファイル出力の段階で選別済みにしておく]
  - Percolator 適用 → peptide FDR 1%

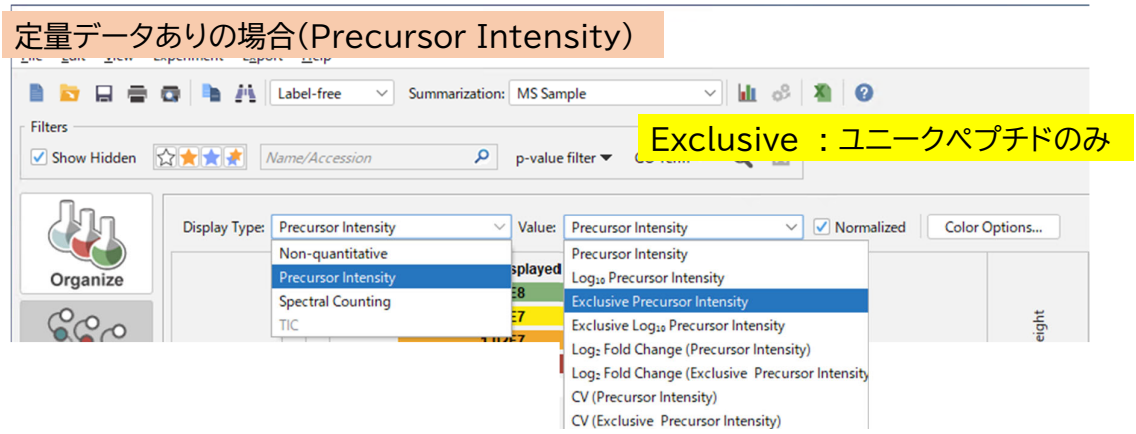
- タンパク質
  - 同定ペプチド N以上アサインされている
  - protein FDR 設定値を満たす
  - clustering ルールでグループ化

表示される「定量値」[Display Type]

定量データなしの場合

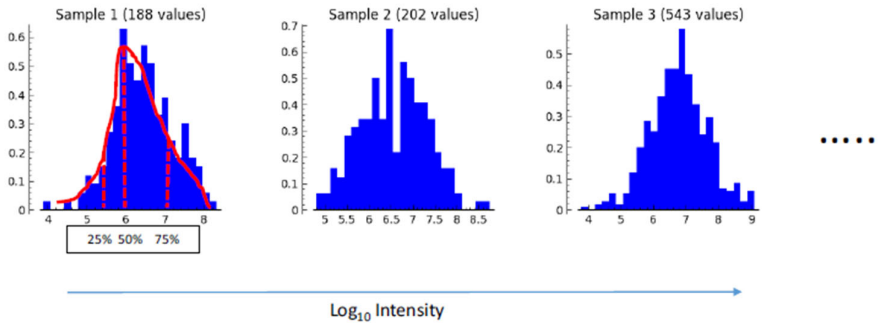


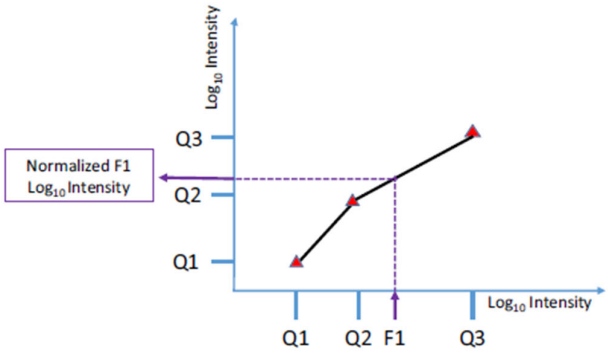
定量データありの場合(Precursor Intensity)



Exclusive ：ユニークペプチドのみ

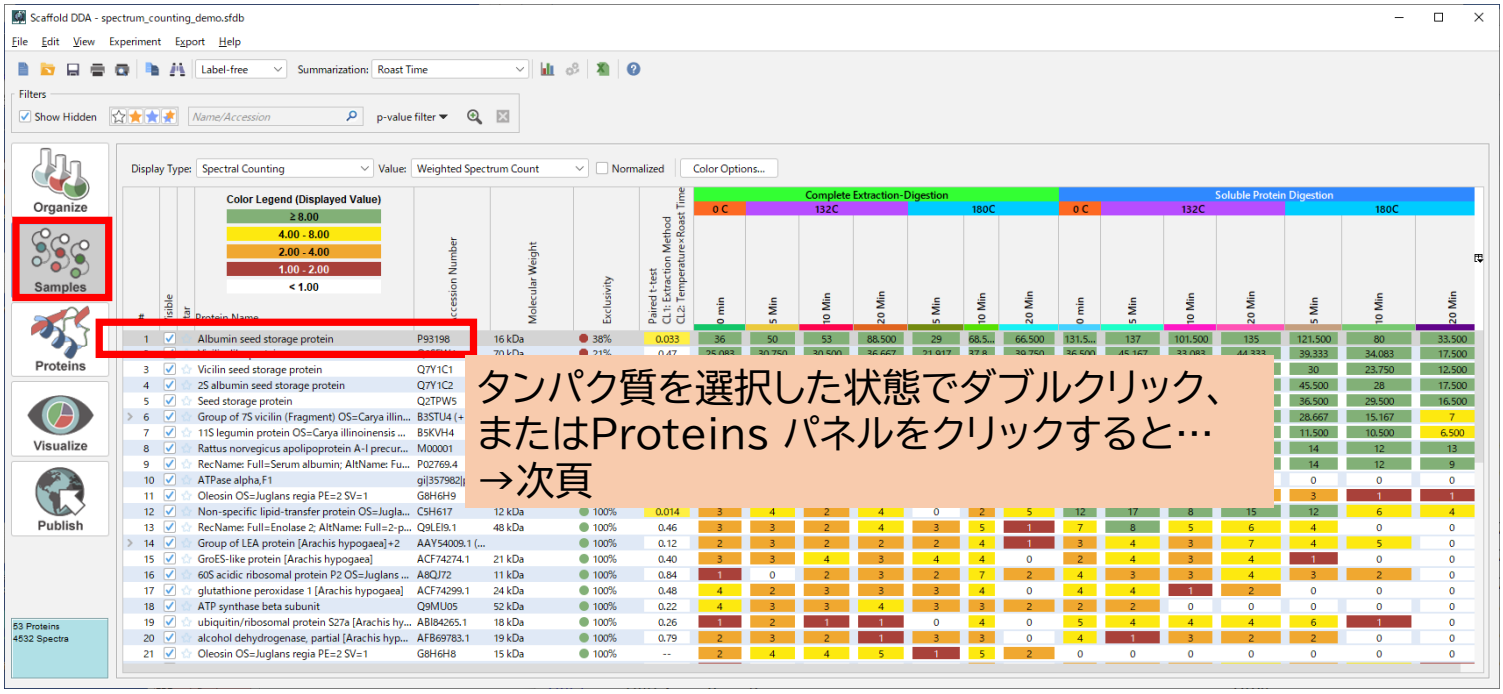
Normalization





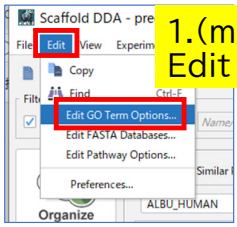
サンプル強度分布によるNormalization

・全データ vs 各sample 、それぞれ強度分布を作成
・各サンプルにおいて、25%,50%,75%にあたる定量値を、全データ平均にそろえる
・各四分位数の間は点間を結ぶ直線に従って変換する

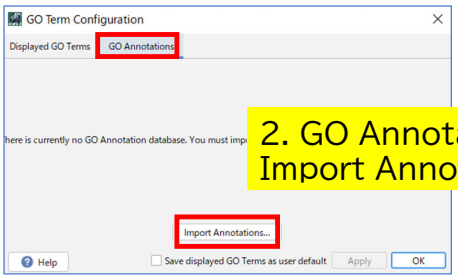


タンパク質を選択した状態でダブルクリック、またはProteins パネルをクリックすると…
→次頁

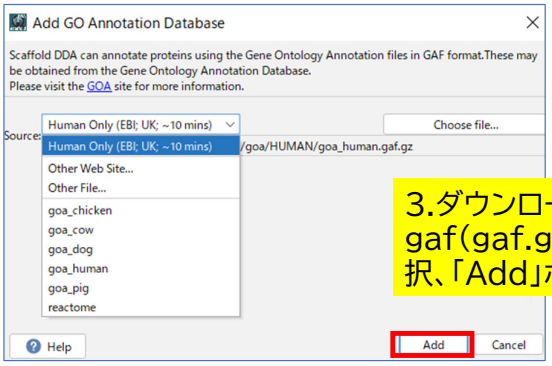
GOファイルのセット方法



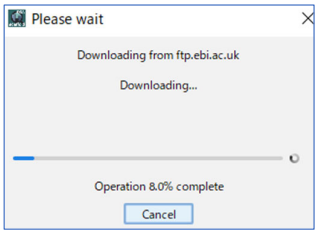
1.(menu) Edit ->
Edit GO Term Options



2. GO Annotations タブ、
Import Annotations

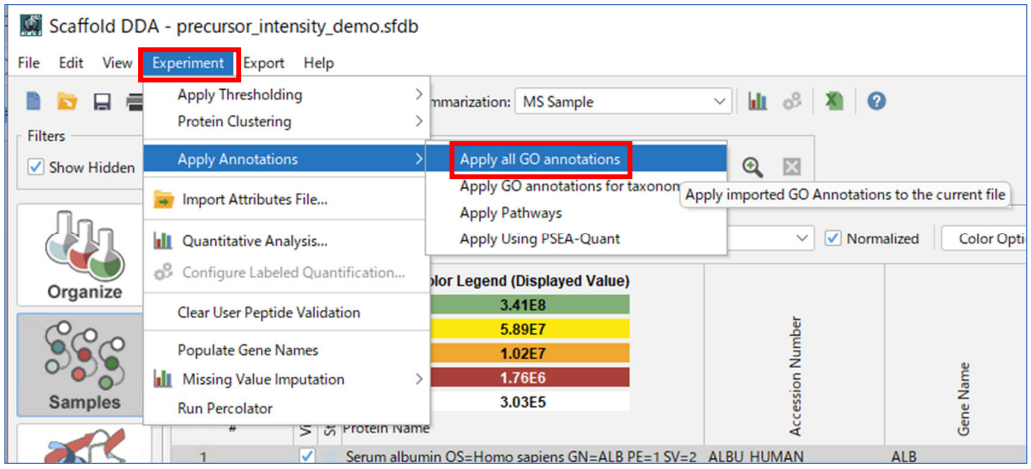


3.ダウンロード先または
gaf(gaf.gz)ファイルを選
択、「Add」ボタン



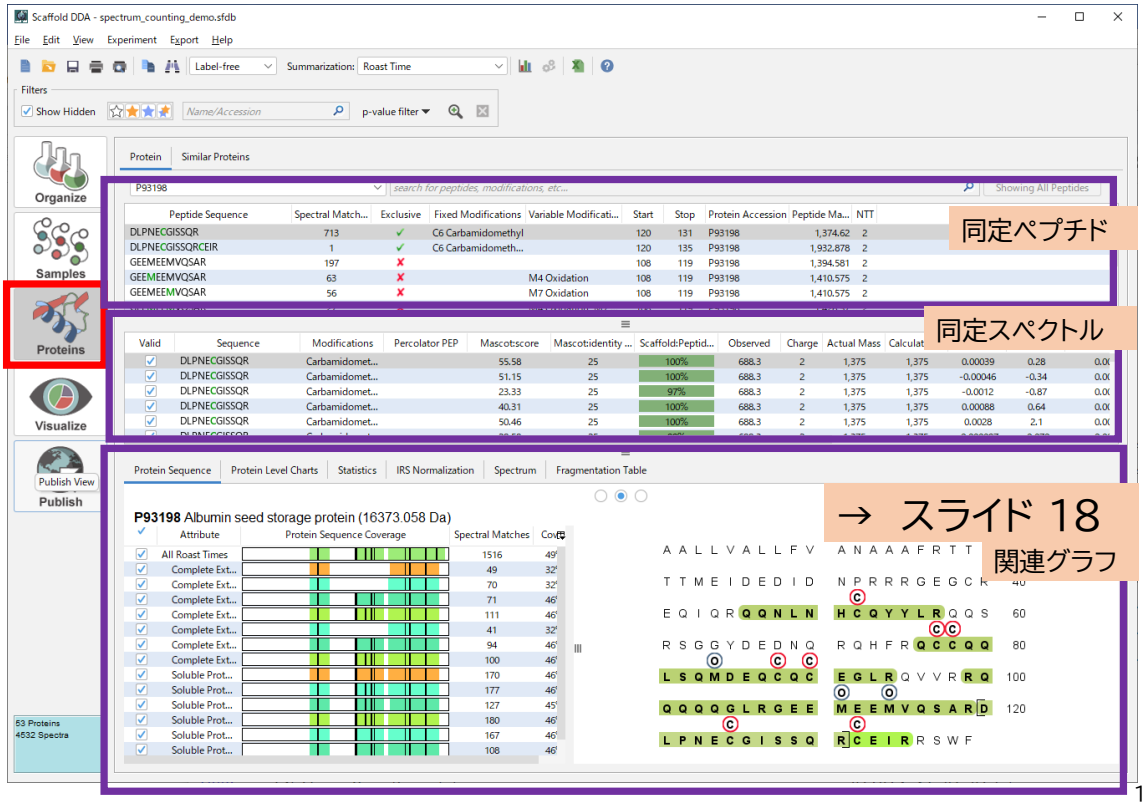
GO情報のannotation付与方法

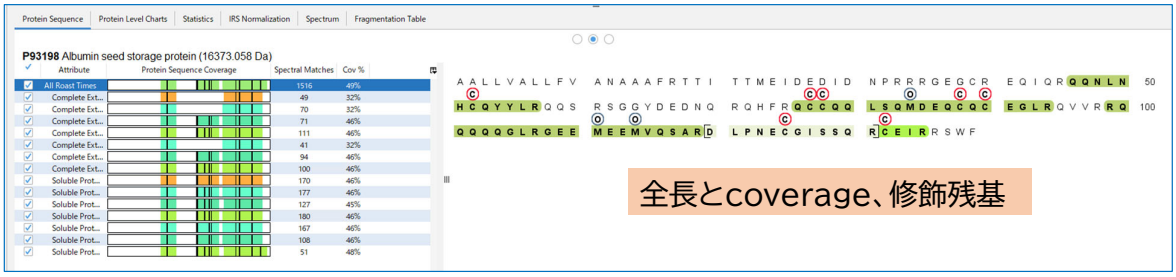
1.(menu) Experiment
-> Apply Annotations
-> Apply all GO annotations

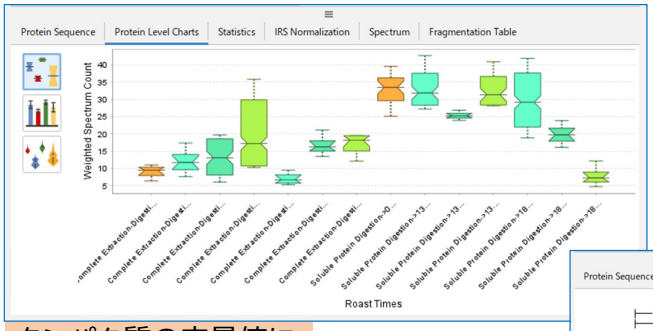


Proteins 画面

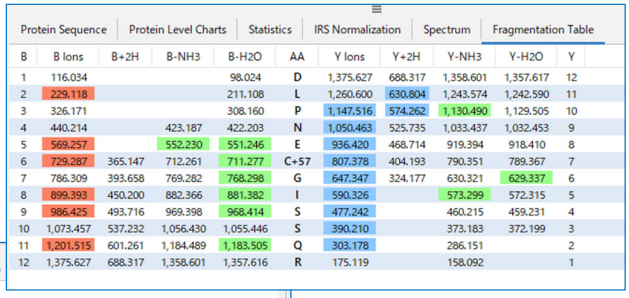
ペプチド/タンパク質の
定性/定量　結果を
より詳細に検証

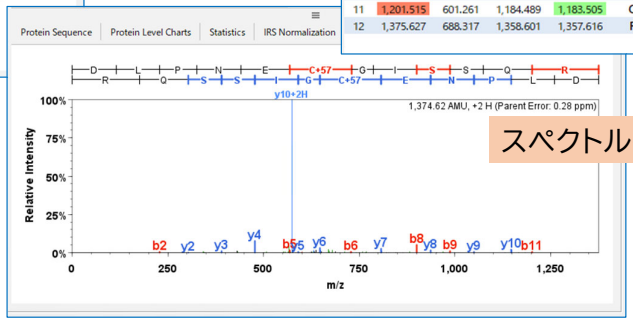






タンパク質の定量値に
基づいたグラフ

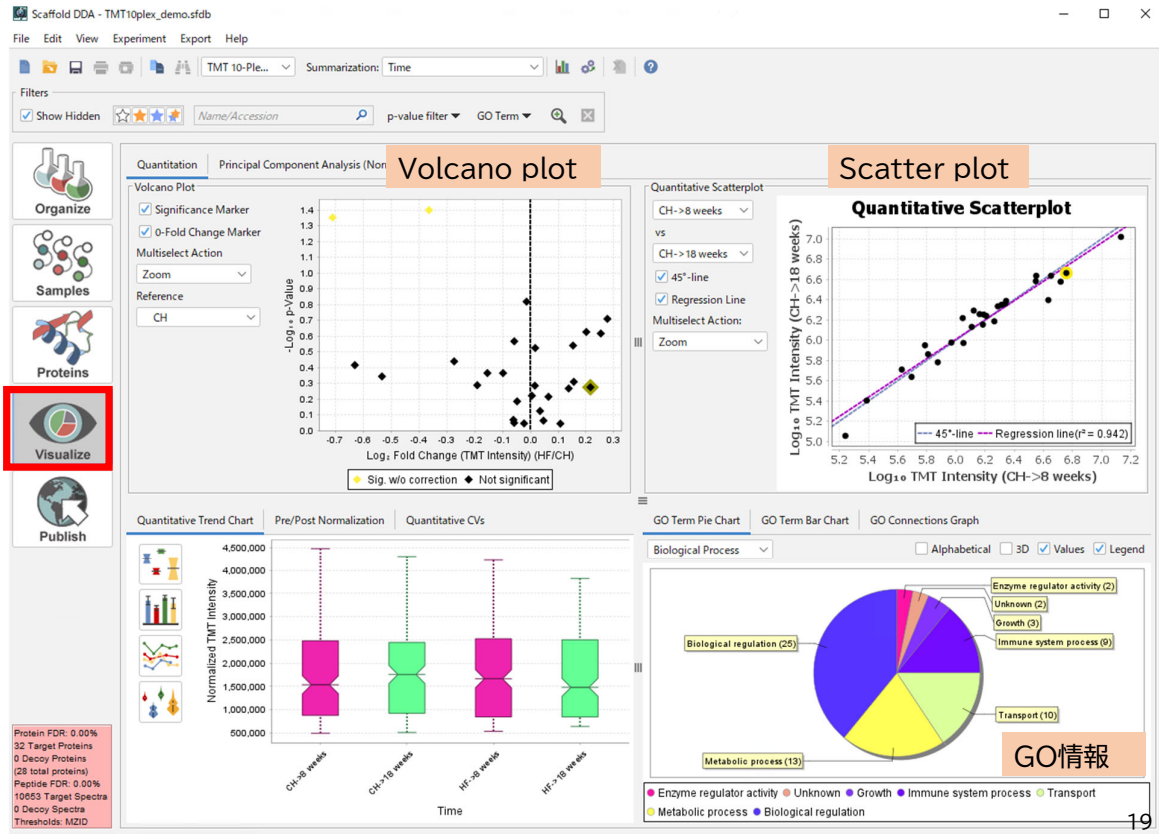


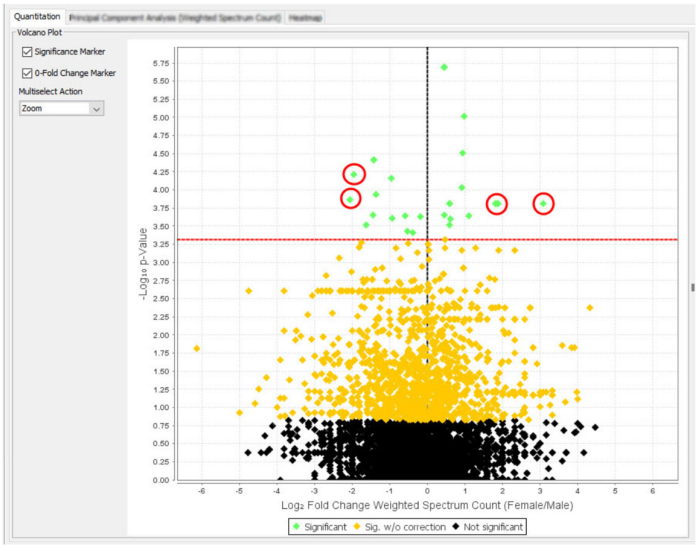


スペクトル-理論値 マッチング

Visualize 画面

有意な変動をしているタンパク質の解析＆データのバラつきをはじめとした確からしさを検証するグラフを提供

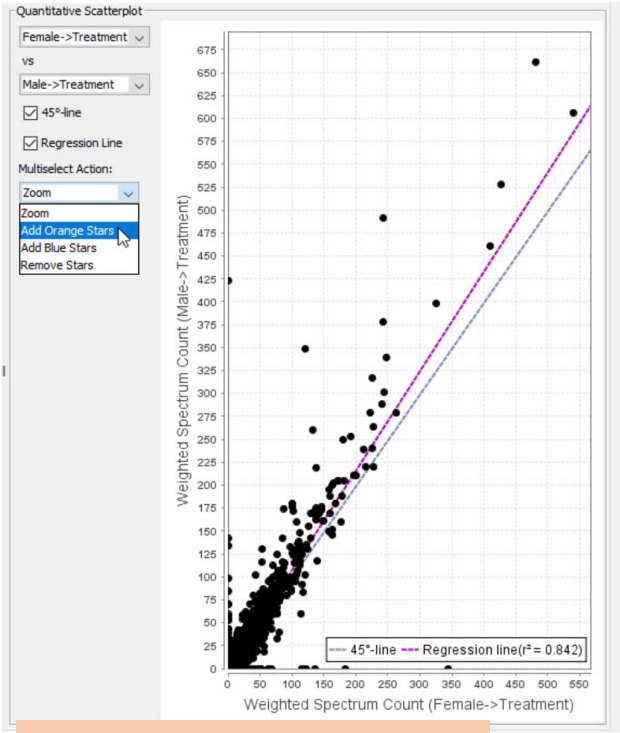




Volcano plot

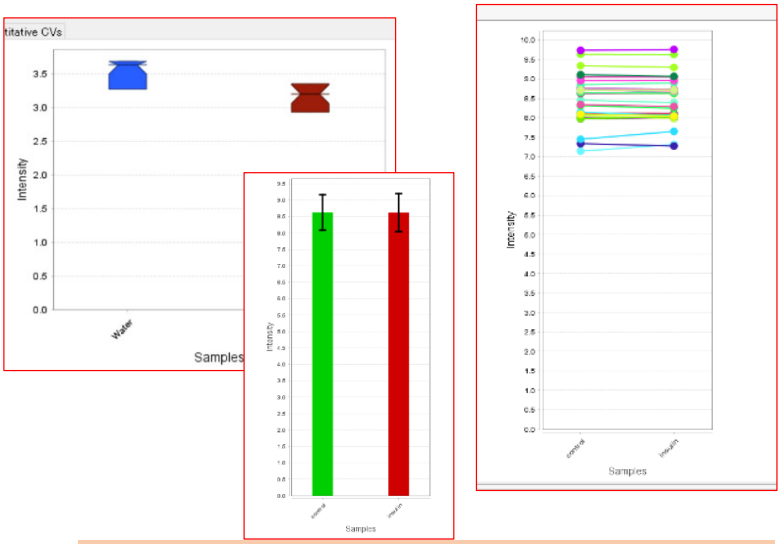
縦軸　$-\mathrm{Log}_{10}$ (p)
横軸　Log_2 (Fold Change)
赤点線　多重検定検証の p(q,α)

変動タンパク質をratio, p-value両面から探す

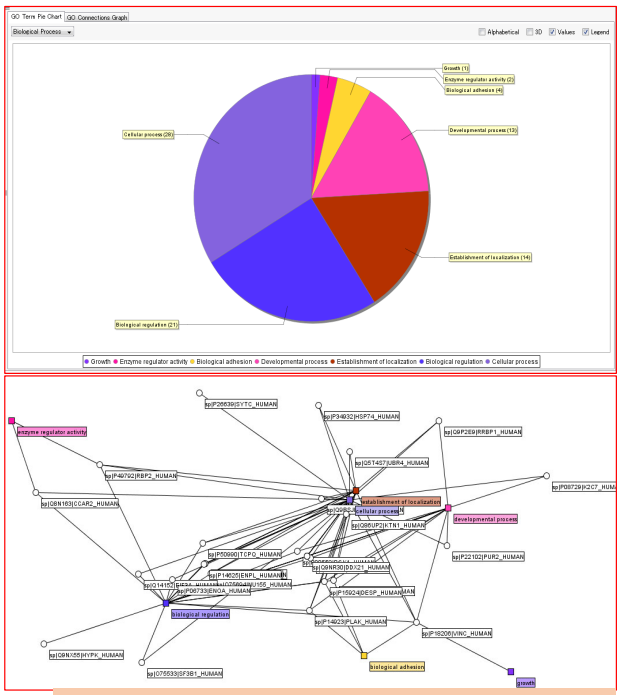


Scatterplot

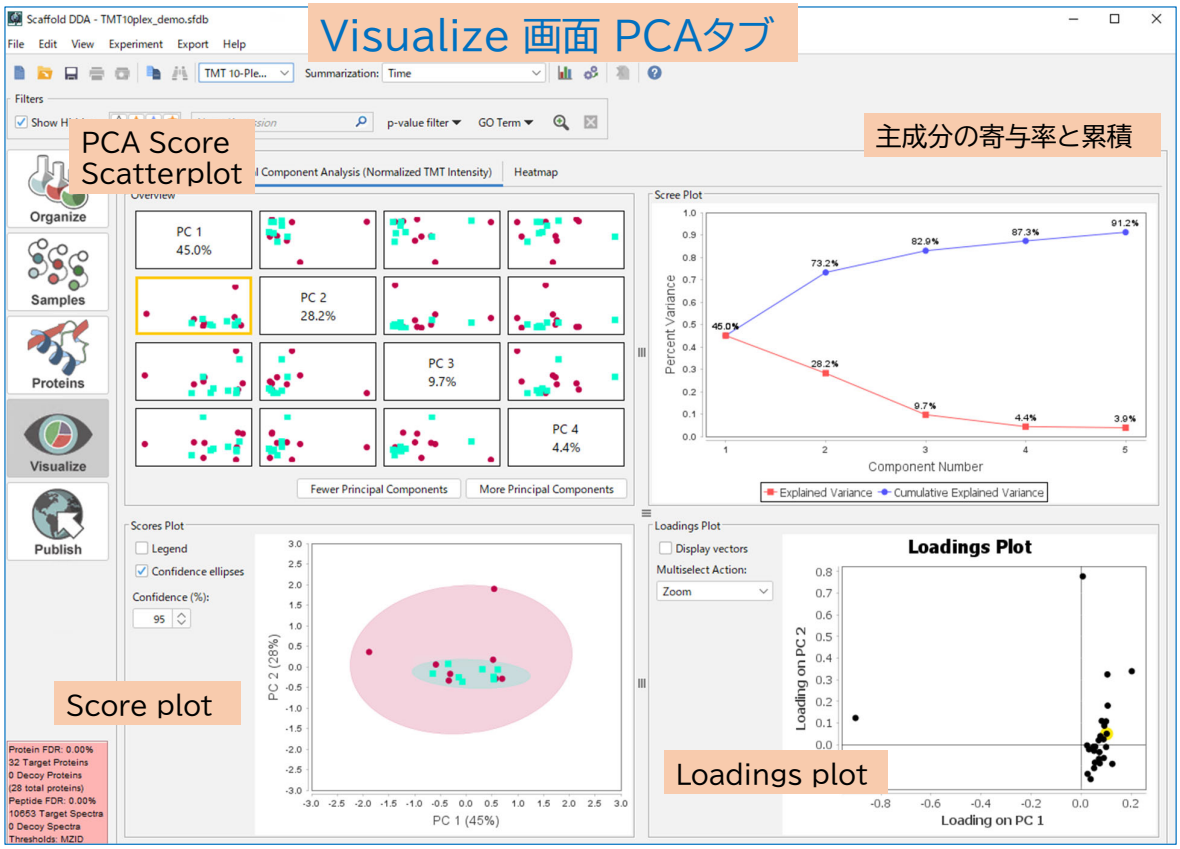
y=x や回帰直線から大きく外れるタンパク質を探す



サンプル毎の定量値に関するグラフ
箱ひげ図(左)、棒グラフ(中)、Trend line(右)
サンプル毎の定量値ばらつきをチェックする

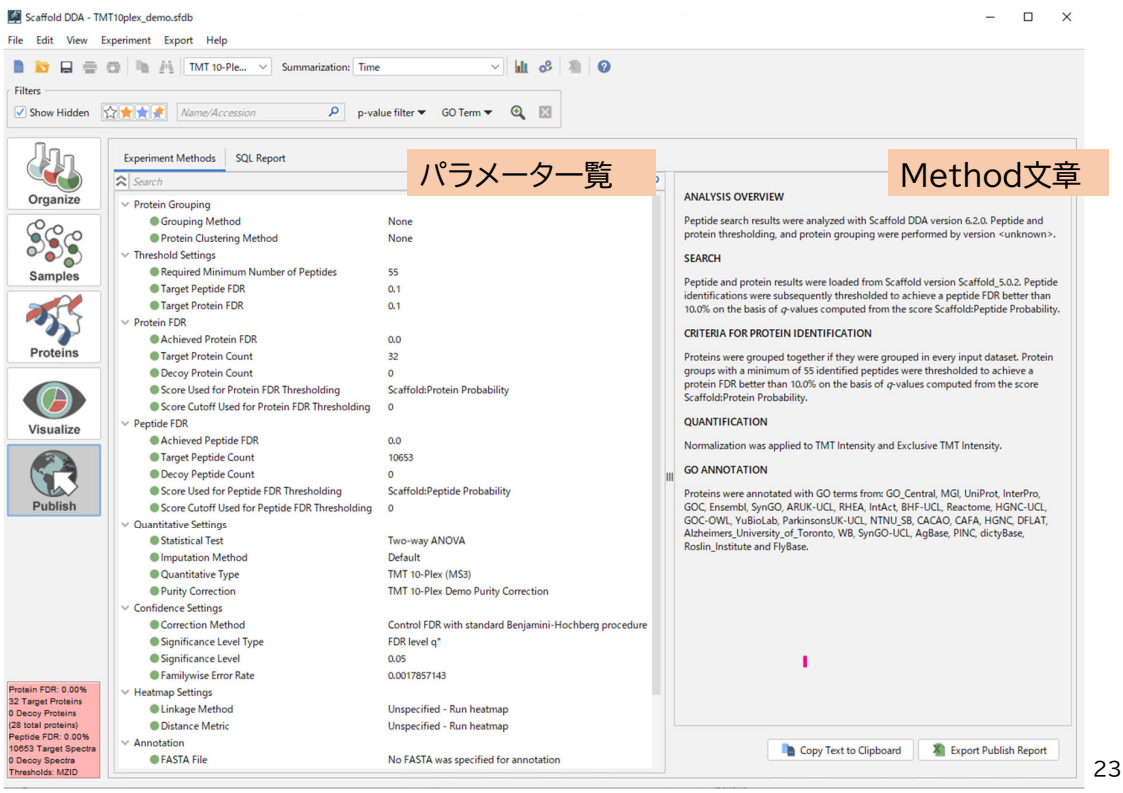


GO情報
円グラフ(上) ネットワーク図(下)

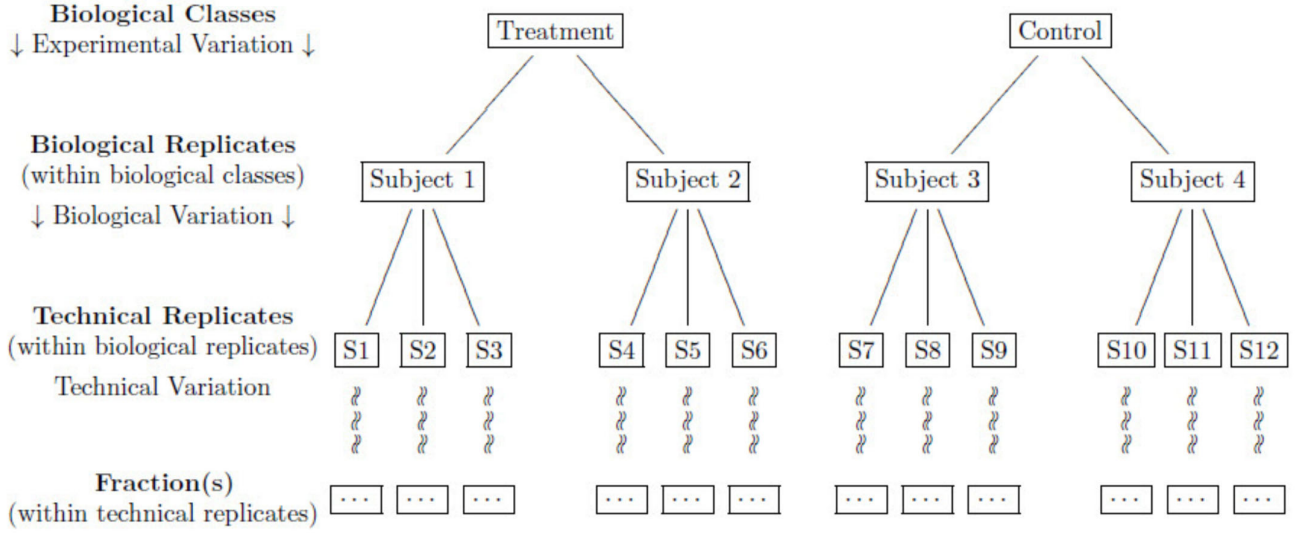


Publish 画面

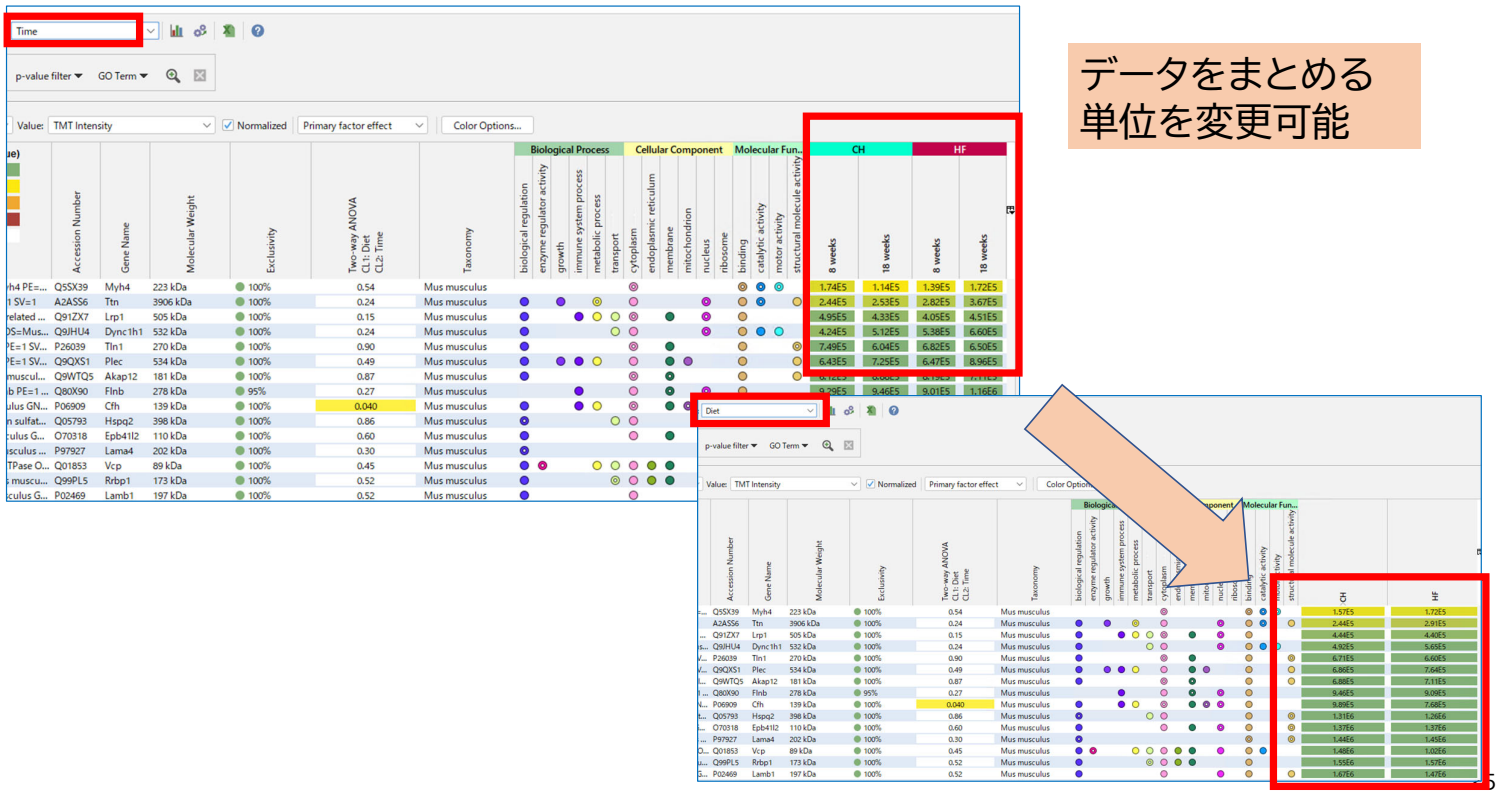
検索パラメータの確認、
論文のMethodに該当する
文章の作成



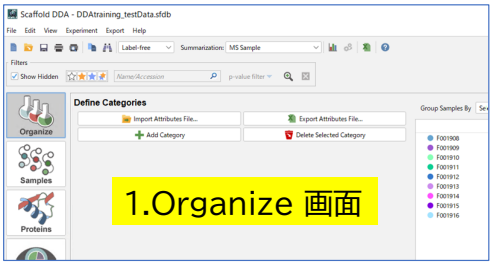
定量：階層構造化と属性付与の例



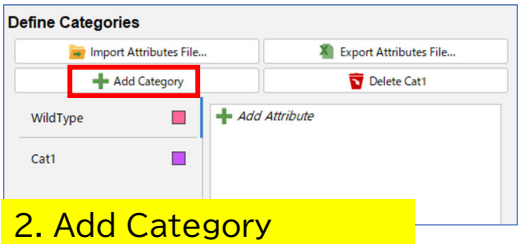
データをまとめる
単位を変更可能



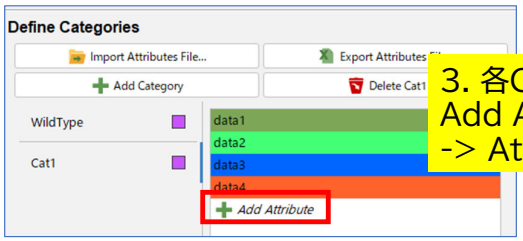
階層設定の操作



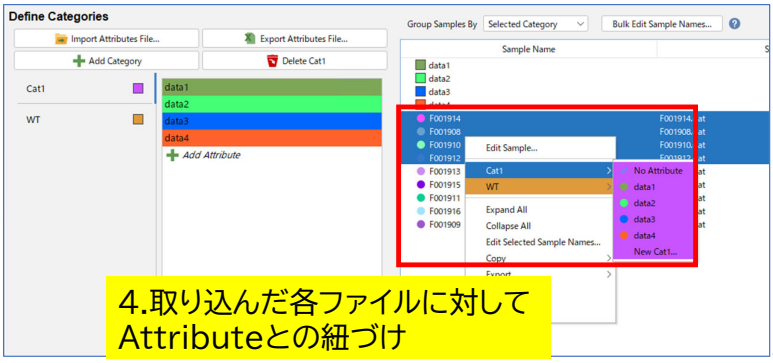
1.Organize 画面



2. Add Category
-> Category（階層）の作成

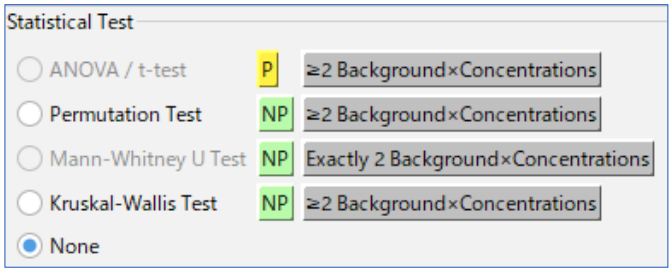


3. 各Categoryにおいて、
Add Attribute
-> Attribute（階層）の作成



4.取り込んだ各ファイルに対して
Attributeとの紐づけ

検定



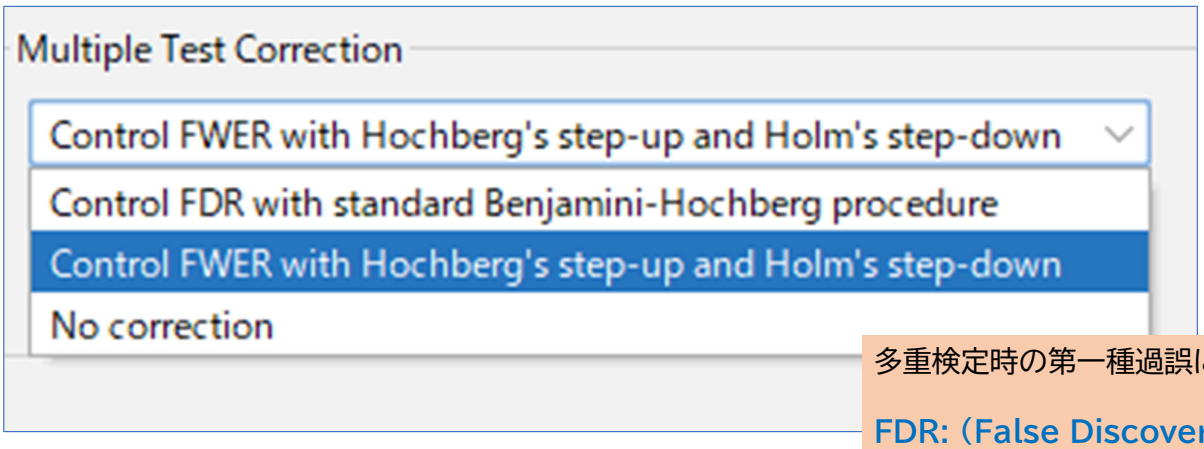
Permutation Test
- ノンパラメトリック
- 2群 ○ 3群以上 ○
- ベースはF検定、群間のランダムなデータを入れ替え
  F値を計算し続ける
- 10000回の 入れ替え計算を行い、入れ替え前のF値より
  有意差以上に差があった回数を10000
  （データ入れ替えの試行回数）で割った値をp-value とする。

t-test / ANOVA
- パラメトリック
- 2群 ○ 3群以上 ○
- ANOVA両側検定を行う。（2群しかない場合 t検定と同じ。）

Mann Whiteney U test / Kruskal-Wallis test
- ノンパラメトリック
- 2群→Mann Whiteney U，3群以上→Kruskal-Wallis
- 同じ分布の形、スケールである事を前提とする

→ P.88

多重比較補正



多重検定時の第一種過誤に対応する補正

FDR: （False Discovery Rate）
BH法

FWER: Family Wise Error Rate
ホッホベルクのステップアップ手順＆ホルムのステップダウン

その他のトピックス(英語マニュアル対応ページ)

・Gene Ontology（→P.28）
Scaffold DDAでGene Ontology情報を付与させるために必要な前処理操作

・タンパク質のグループ化、クラスター化（→P.150）
シェアペプチドの存在とタンパク質のグループ化ルール

・Report (→P.177)
Scaffold DDAで出力可能なファイル

英語マニュアルのAppendix

- Appendix A. Computation of protein and peptide FDR in Scaffold DDA
- Appendix B.Rolling Up Intensity Values
- Appendix C. Shared Evidence Clustering Algorithm
- Appendix D. Distance Based Clustering
- Appendix E. Weighted Spectrum counts
- Appendix F. Terminology
- Appendix G. Heat map clustering
- Appendix H. Techniques to Control the Family-wise Error Rate
- Appendix I. Using Principal Component Analysis in Scaffold DDA
- Appendix J. How PCA is Performed in Scaffold DDA
- Appendix K. Description of Mouse Right Click Context Menu Commands

インストール環境、その他

■インストール環境

・対応OS
- Windows 10,11（64bit）
- Mac（最新版）
- Linux（Ubuntu 16.04 LTS 以降）

・メモリ
最低 4GB以上とあるが、製品版としては32GB以上を強く推奨
（大規模解析なら64GB以上）

・ストレージ
最低 数百 MB 以上。高速SSD の使用を推奨

・CPU
使用可能なコア数の上限が 64コア