

# DIA 定量プロテオーム解析ソフトウェア Scaffold DIA

Volcano plot

図の補足説明

→ P.34

日本語マニュアルの該当ページ  
×対応するスライド番号

スライド番号→

1

1

## Scaffold DIA 何ができるか

- DIAデータのタンパク質の同定、定量
- **検索対象:** 4種の検索に対応
  - FASTA
  - BLIB (DDA-ライブラリ),
  - DLIB (prositライブラリ、配列から理論DDAライブラリ作成)
  - ELIB (DIA解析結果ライブラリ)
- 定量解析/検定とグラフ表示
- Viewerによる結果ファイルのシェア

# 対応フォーマット: ProteoWizard Msconvert を使っている

## 各社メーカーのファイルフォーマット [読み込みなどがよく確認されているもの]

装置メーカー	ファイルフォーマット
SCIEX	*.wiff (.wiff.scanも同じフォルダに置く事)
Agilent	*.d (ディレクトリ)
Thermo	*.raw
オープンフォーマット (原理的には mzML フォーマットになっていればデータの読み込みが可能)	
HUPO Proteomics Standards Initiative mzML	*.mzML

\* mzMLはほぼすべてのメーカーで変換可能なフォーマット

## 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


The screenshot shows the Scaffold DIA software interface. The main window displays a table of identified proteins. The table has columns for: #, Visible, Star, Protein Name, Molecular Weight, Protein Group Score, Identified Peptide Count, Exclusivity, Next CL Treatment, Taxonomy, Biological Process, Cellular Component, Molecular Function, Control, and Insulin. A 'Samples' button is highlighted in the left sidebar. The table lists various proteins such as HUMAN E3 SUMO-protein ligase, HUMAN Desmoglein OS=Homo sa..., HUMAN Alpha-actinin-4 OS=Hom..., HUMAN DNA topoisomerase 2-alp..., HUMAN Vinculin OS=Homo sapiens..., HUMAN Alpha-actinin-1 OS=Hom..., HUMAN Nucleolar RNA helicase 2..., HUMAN Ribosome-binding protei..., HUMAN Kinectin OS=Homo sapien..., HUMAN Splicing factor 3B subunit..., HUMAN Heat shock 70 kDa protein..., HUMAN Threonine-tRNA ligase, cycl..., HUMAN Eukaryotic translation initi..., HUMAN Trifunctional purine biosyn..., HUMAN T-complex protein 1 subun..., HUMAN T-complex protein 1 subun..., HUMAN Eukaryotic translation initi..., HUMAN Endoplasmic OS=Homo sa..., HUMAN Keratin, type II cytoskeletal..., HUMAN Phosphoglycerate kinase 1..., HUMAN Alpha-enolase OS=Homo..., and HUMAN Telomere-associated protei...

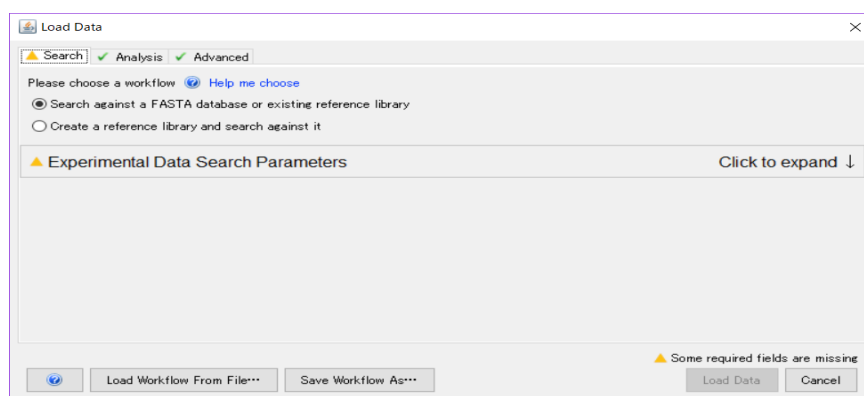
# データ取り込み時の画面

5

→ P.16

## 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



6

### ① workflow : elib作成と連動した2段階検索かそれ以外か

Search a reference library : 直接ライブラリー検索を実施

Create a chromatogram library and search against it : 2段階検索

### ② ライブラリ、タンパク質配列の指定

直接検索対象とするファイル(elib, dlib, blib, fasta)と、それに対応する配列データベース(fasta)を選択

### ③ parameter

Fragmentation :

CID/HCD/ETDから選択。考慮するイオンシリーズ

Precursor / Fragment / Library Fragment Tolerance :

理論値と実測スペクトルデータとの許容誤差範囲

Digestion Enzyme : ペプチド切断方法

Peptide Length : 検索対象とするペプチドの長さ

Peptide Charge : 検索対象とするペプチドの電荷

Max Missed Cleavages :

理論ペプチド作成の際、切断対象のアミノ酸を何回無視したペプチドを作成するかの設定

Modifications: 考慮する修飾

① → スライド13

② → スライド11

③

7

### ④ Peptide FDR Threshold :

ペプチドの同定基準値、FDRの値

### ⑤ Data Acquisition Type

DIAの測定でのPrecursor Isolation Window タイプについて

[次、次々スライドに補足説明図]

- Non-overlapping Windows :

両端のオーバーラップ領域がない

- Overlapping Margins :

両端のオーバーラップ領域がある(値も指定)

- Staggered Windows :

window領域が半分ずつずれた測定を実施

### ⑥ Precursor Window Size

DIAの測定Window に対する設定。通常はrawファイルに書かれた情報をそのまま利用

### ⑦ Experimental Data File

検索対象となるrawファイルを選択

④ → スライド 9,10

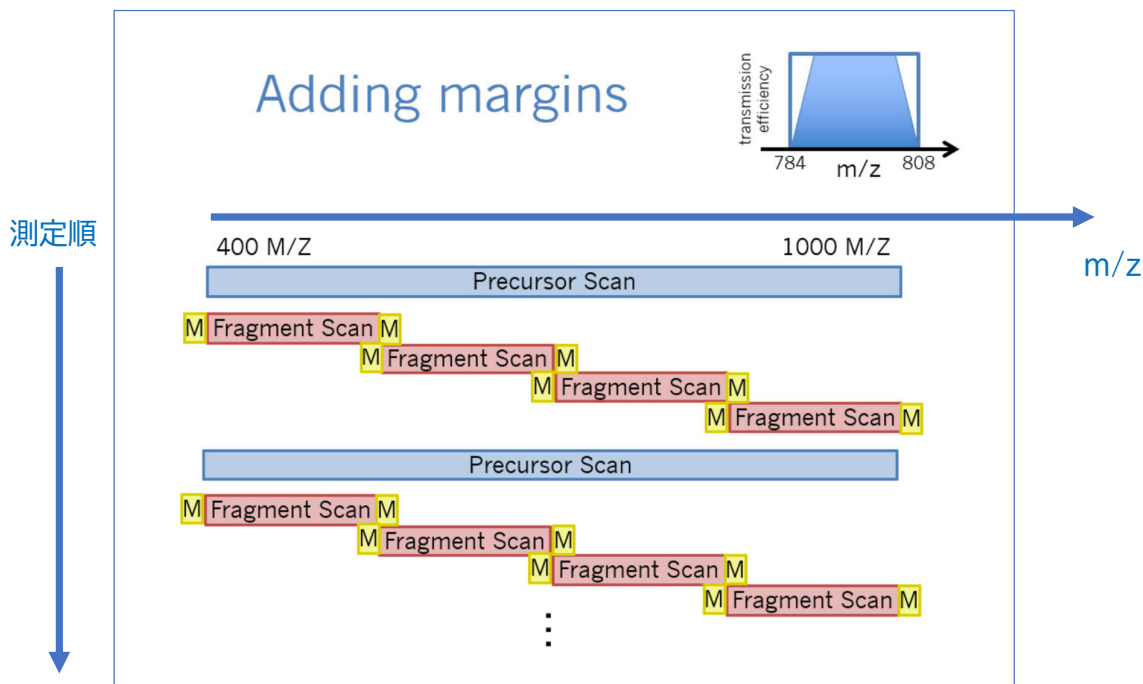
⑤

⑥

⑦

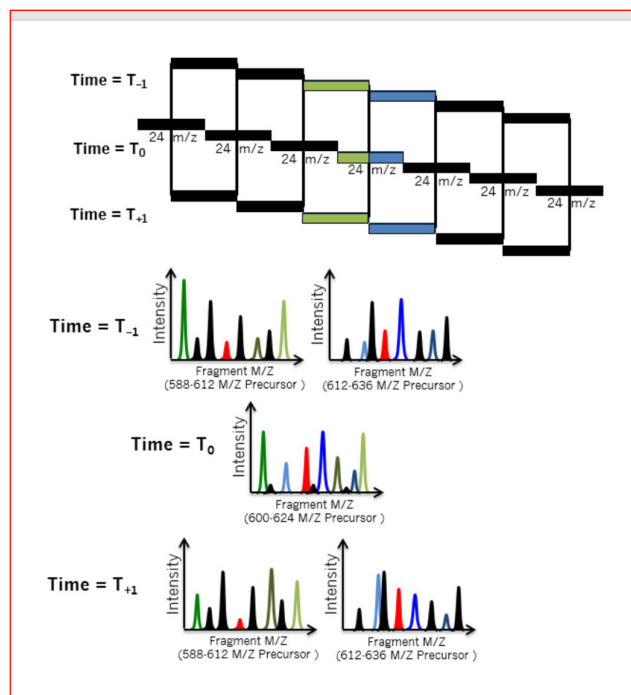
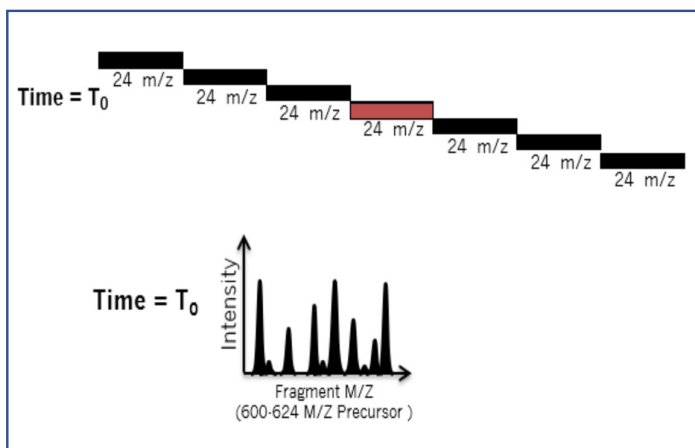
8

# 補足説明：対応するDIA解析 -- marginあり



9

# 補足説明：対応するDIA解析-- Marginなし、staggered



10

# 補足説明：検索対象 FASTA、BLIB、ELIB (+DLIB)

```

>sp|A2Z0B0|ACTO_XENLA Actin, cytoplasmic 2 OS=Xenopus laevis GN=ac
MEE EEAALVIIDGSSGMKAFAGADAPRAVFP31VGRPRHGVMMGMKDSYVGEADSRGILT
DMEKIHWHITINELNVAPEEHPVLLTEAPLNPKAREIMTODMEFTNTPMNYVAQVLSLVAQ
HTVPIYEGYALPHALRLDLADROLTDVLMKILTERGYSFTTTAEREIVRDIKKELCYVALDFEQ
ELPGDGVITLNERFRPEALFQPSFLMESQ31HEITFNSIMKDDVDIKDLVANTVLSGGTMM
AFSTMKIKITLAPPERKYSVNIIGSSILASLSTFQDMMSKQYDESSPSTVHRVCF
>sp|A20Z0Z|EF1AI HORSE Elongation Factor 1-alpha I OS=Equus caball
MGEKTHINIVVIGHVDSGKSTTTOHLIVKCGDIDKRTIEKFEKAAEMKGSFKYAWLQKKA
ETSKYYTITIDAPGHDFIKNMTIGTSQADICLVAAAGVGEFAGISKNGDTREHALAYTLGVP
PSSDARYEEIWEVSYTYIKKIDVNPDTVAFYFISDWDNLEPSAMPHKQWVTRKQNASQI
PTDKPLRLKLDVYKIGGIITVYVGRVETGVLKPSMVFYAPVNYITEKVSVMHEALSEALPGI
RRGVAGDSKNDPMEAGFTADVIILNHPGQISAGYAPVDDOHTAHIAKFAELKEKIDRRSGKP
IVDMVPGKPMVEFSQYFPLGRFAVRMRQTVAVGVYKAVDKKAAGAGKYTKSAKKAQAK

```

## FASTA

- 配列ファイル
- ※配列から測定条件などを考慮して計算されたスペクトルデータ (DLIB) も使用できる。

## BLIB

- (DDA) マススペクトル ピークリスト

## ELIB

- (DIA) RT + m/zのピーク位置
- ※DLIBフォーマットはELIBとほぼ同じ

# 補足説明：Prositで作成されたライブラリ、DLIB

**Library Creation Parameters**

The following parameters were used in the conversion process and thus should match your instrumentation settings. If they vary dramatically from your settings please contact us.

Parameter	Setting
Charge Range	2 - 3
Maximum Missed Cleavages	1
m/z Range	396.4 - 1002.7
Default NCE	33
Default Charge	3

**Featured Libraries**

*Coronavirus reference*  
Download Coronavirus only DLIB (1.9 MB)  
Download Coronavirus plus Human (pan human) DLIB (259 MB)  
Download Coronavirus only FASTA - 13 entries  
Download Coronavirus plus Human FASTA - 20,350 entries

**Available Libraries**

*Arabidopsis thaliana*  
Download Arabidopsis thaliana DLIB (1.2 GB)  
Download FASTA file accessed 10/22/19 - 15,896 entries

*Caenorhabditis elegans*  
Download Caenorhabditis elegans DLIB (360 MB)  
Download FASTA file accessed 10/23/19 - 4,089 entries

*Danio rerio*  
Download Danio rerio DLIB (250 MB)  
Download FASTA file accessed 10/22/19 - 3,125 entries

*Drosophila melanogaster*  
Download Drosophila melanogaster DLIB (365 MB)  
Download FASTA file accessed 10/22/19 - 3,586 entries

*Escherichia coli (strain K12)*

<https://support.proteomesoftware.com/hc/en-us/articles/360035151172-Prosit-Derived-Spectral-Libraries-for-Scaffold-DIA-Searches>

**Prosit**

- 配列からdeep Neural Networkアルゴリズムで計算された理論スペクトルで、保持時間やintensityも予測
- Proteome Software社側で準備したDLIBを公開している
- 多くのケースでBLIBの代わりになる、手軽に使える検索対象として利用してください
- Scaffold DIA上にFASTAからprositのインプットファイルを作成するツールあり
- 修飾が苦手

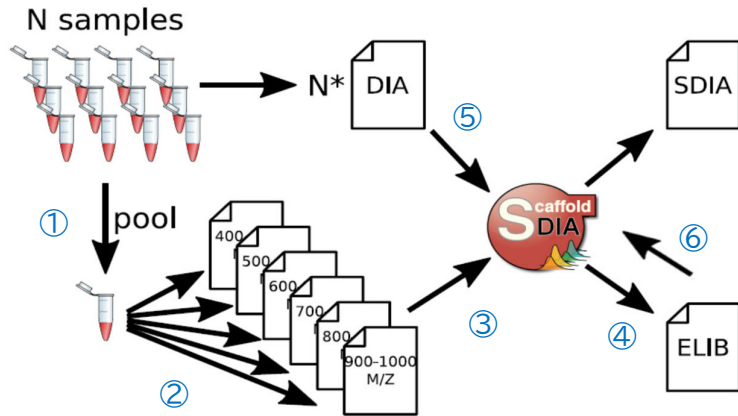
# 補足説明：ELIBと「2段階」検索

「pool」データへの検索でELIBを作成し、そのELIBに対して元のデータを検索するプロトコル例

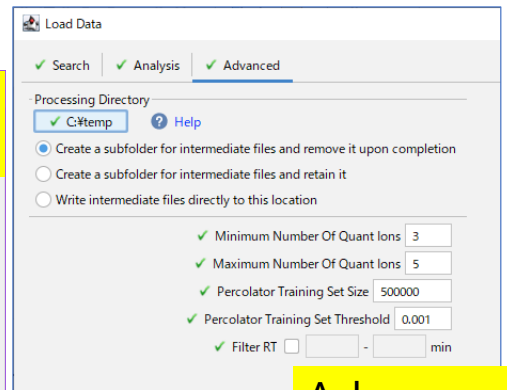
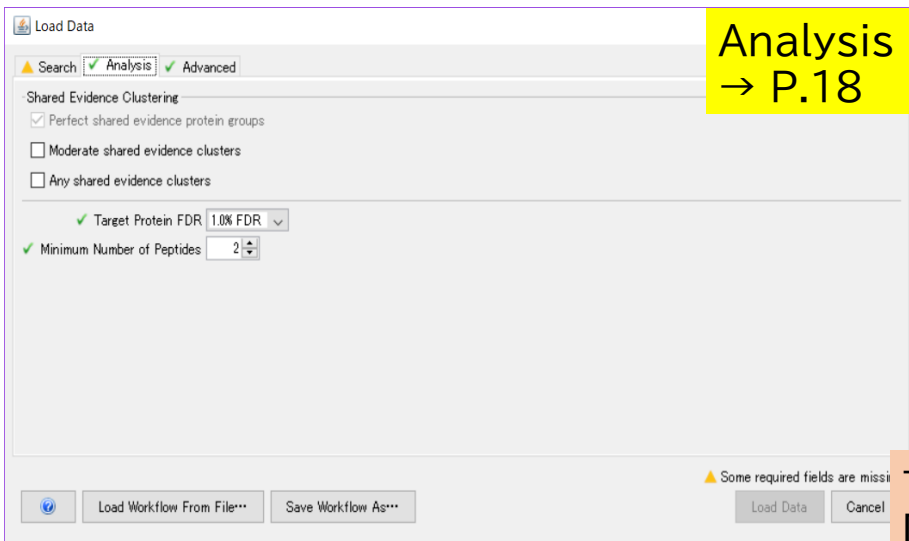
- ① sampleから少量集めプール
- ② 100Daなど大きさを分ける
- ③ fasta,blibを使って解析
- ④ 解析結果を「ELIB」登録
- ⑤ 元のサンプルの解析開始 (N回繰り返し実験)
- ⑥ ④で作ったELIBを使って同定・定量力アップ

## The Combined Workflow

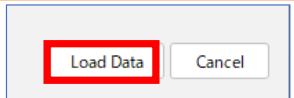
For the highest quality identifications, Scaffold DIA can create a library. A pooled sample is searched with XCorDIA and the resulting ELIB is used to extract and quantify peptides from experimental samples.



# Analysisタブ、Advanced タブ、取り込み開始



すべての項目を埋めてから「Load Data」ボタンを押す



# ペプチド・タンパク質の定性(同定)

## • ペプチド

- encyclopeDIA スコア評価
- peptide FDR 設定値 (percolator も使用)を満たす

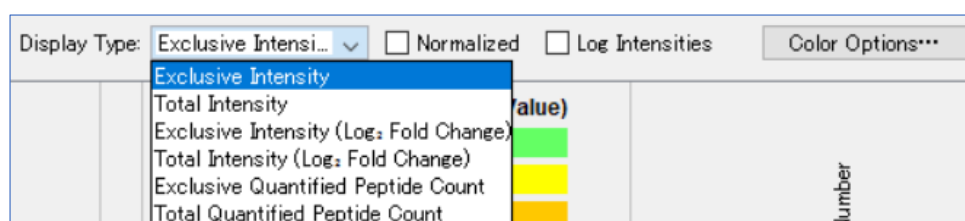
## • タンパク質

- 同定ペプチド N(デフォルト2) 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

15

→ P.53

## 表示される「定量値」 [Display Type]



Exclusive : ユニークペプチドのみ

Total : シェアペプチド含む

### • Exclusive/total Intensity

ペプチド定量値(フラグメントXICのピーク強度から算出) の和

### • Exclusive/total Intensity (Log<sub>2</sub> Fold Change)

ペプチド定量値について、Referenceとの比を取りその数字に対して2を底にするLogに変換

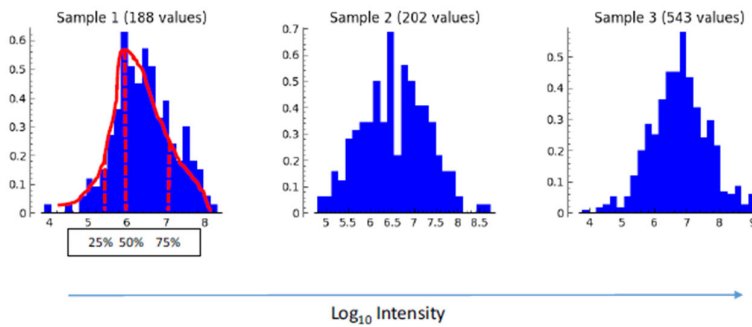
### • Exclusive/total Quantified Peptide Count

タンパク質あるいは表示グループにアサインされたペプチドで、かつ定量に使われたフラグメントを持つペプチドに限定し数え上げた数(標準化处理も実施)。

16

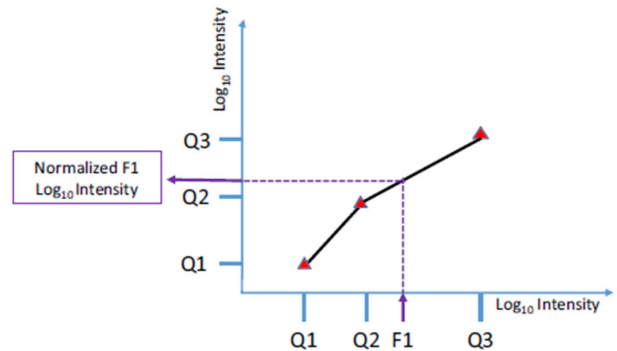


# Normalization



## 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



Scaffold DIA - Demo2\_HeLa\_insulin\_6\_files\_30\_proteins.sdia

File Edit View Experiment Export Tools Help

Summarization: Treatment Protein FDR: 1.0% FDR Min # Peptides: 2

Filters: Show Hidden Name/Accession p-value filter GO Term

Display Type: Exclusive Intensity Normalized Log Intensities Color Options...

Color Legend (Displayed Value): 6.36E8, 1.22E8, 2.33E7

Proteins: 0.0% FDR (attained) 28 Targets 0 Decoys

Peptides: 0.0% FDR (attained) 880 Targets 0 Decoys

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

#	Star	Protein Name	Score	Count	Biological Process	Cellular Component	Molecular Function	Control	Insulin
1	✓	spjP4979[R8P2_HUMAN]	1	43	63%			2.45E8	2.24E8
2	✓	spjP1826[VINC_HUMAN Vinculin OS=Homo sapiens]	1	43	0.49			1.32E8	1.39E8
3	✓	spjP12814[ACTN1_HUMAN Alpha-actinin-1 OS=Homo sapiens]	1	43	0.23			6.17E8	5.81E8
4	✓	spjQ9NR30[DDX21_HUMAN Nucleolar RNA helicase 2-like protein 1 OS=Homo sapiens]	1	41	0.090			8.93E7	8.06E7
5	✓	spjP11388[TOP2A_HUMAN Topoisomerase II alpha OS=Homo sapiens]	1	41	0.18			4.16E8	4.49E8
6	✓	spjP18206[VINC_HUMAN Vinculin OS=Homo sapiens]	1	43	0.31			6.58E8	5.44E8
7	✓	spjP12814[ACTN1_HUMAN Alpha-actinin-1 OS=Homo sapiens]	1	43	0.49			1.77E8	1.68E8
8	✓	spjQ9NR30[DDX21_HUMAN Nucleolar RNA helicase 2-like protein 1 OS=Homo sapiens]	1	41	0.090			5.71E8	5.91E8
9	✓	spjQ86UP2[KTN1_HUMAN Kinectin OS=Homo sapiens]	1	41	0.18			1.22E8	1.04E8
10	✓	spjQ86UP2[KTN1_HUMAN Kinectin OS=Homo sapiens]	1	41	0.18			1.24E8	1.09E8
11	✓	spjP11388[TOP2A_HUMAN Topoisomerase II alpha OS=Homo sapiens]	1	40	0.14			2.66E8	2.42E8
12	✓	spjP34932[HS74_HUMAN Heat shock 70 kDa protein 1 OS=Homo sapiens]	1	38	0.65			5.61E8	5.55E8
13	✓	spjP26639[SYT3_HUMAN Threonine--RNA ligase, cytosolic OS=Homo sapiens]	1	38	0.51			8.78E8	9.20E8
14	✓	spjQ14152[EIF3A_HUMAN Eukaryotic translation initiation factor 3 subunit A OS=Homo sapiens]	1	38	0.83			4.83E8	4.82E8
15	✓	spjP22102[PUR2_HUMAN Trifunctional putative biosynthetic protein OS=Homo sapiens]	1	37	0.65			5.69E8	5.46E8
16	✓	spjP90990[TCQP_HUMAN T-complex protein 1 subunit OS=Homo sapiens]	1	36	0.25			1.20E9	1.27E9
17	✓	spjP78371[TCPB_HUMAN T-complex protein 1 subunit OS=Homo sapiens]	1	35	0.39			7.59E8	8.82E8
18	✓	spjQ04637[IFAG1_HUMAN Eukaryotic translation initiation factor 1 OS=Homo sapiens]	1	34	0.51			4.85E8	4.79E8
19	✓	spjP14625[ENPL_HUMAN Endoplasmic reticulum protein OS=Homo sapiens]	1	34	0.89			1.19E9	1.17E9
20	✓	spjP06729[KCC7_HUMAN Keratin, type II cytoskeletal class 7 OS=Homo sapiens]	1	33	0.61			2.17E9	2.18E9
21	✓	spjP00553[CK1_HUMAN Phosphoglycerate kinase 1 OS=Homo sapiens]	1	30	0.52			4.58E9	4.82E9
22	✓	spjP06733[ENO4_HUMAN Alpha-enolase OS=Homo sapiens]	1	30	0.13			7.42E9	1.01E10
23	✓	spjQ5UIP0[RIF1_HUMAN Telomere-associated protein 1 OS=Homo sapiens]	1	25	0.71			3.07E7	2.96E7

# 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

→ ①/②スライド 19

Quantified	Peptide Sequence	Quantified Mat...	Fixed Modifications	Variable Modificati...	Start	Stop	Protein Accessi...	Proba...	Mass
<input type="checkbox"/>	AAPPPPPPPPLESSPR	0 of 6			606	622	sp Q5T4S7 UBR4...	86%	1,702.904
<input checked="" type="checkbox"/>	AEHASSLLELASTTK	0 of 6			1065	1079	sp Q5T4S7 UBR4...	100%	1,556.805
<input type="checkbox"/>	ALGTLGTTNEK	0 of 6			4803	4814	sp Q5T4S7 UBR4...	99%	1,234.623
<input type="checkbox"/>	APSYIEIFGR	0 of 6			2364	2373	sp Q5T4S7 UBR4...	98%	1,151.598
<input checked="" type="checkbox"/>	AGQALSELHVEK	2 of 6			1828	1850	sp Q5T4S7 UBR4...	100%	1,452.757

① 同定ペプチド

Sequence	Modifications	Charge	Sample	Quant. Intensity	# Qua...	RT Start (min)	RT Center (min)	RT Stop (min)	Precur...
AEHASSLLELASTTK		2	Control_1	—	0	—	—	—	779.410
AEHASSLLELASTTK		2	Control_2	—	0	—	—	—	779.410
AEHASSLLELASTTK		2	Control_3	—	0	—	—	—	779.410
AEHASSLLELASTTK		2	Insulin_1	—	0	—	—	—	779.410

② 同定スペクトル

Protein Sequence | Protein Level Charts | Chromatograms | Fragment Intensities | Fragmentation Table

sp|Q5T4S7|UBR4\_HUMAN E3 ubiquitin-protein ligase UBR4 OS=Homo sapiens GN=UBR4 PE=1 SV=1

→ ③ スライド 20

Q H N L L S P P F G W A S G S Q D S N S R R A T T P L ...

**H F S S D A V P H P R** F Y C V L S P E A S E D D L N R L D S V A C D V L F S K L 96

V K Y D E L Y A A L T A L L A A G S Q L D T V R R K E N K N V T A L E A C A ...

Y Y F L I L W R I L G I L P P S K **TYI N Q L S M N S P E M S E C D I L H T L R** 1040

③ 関連グラフ

756 out of 5183 (14.58%) amino acids identified with 23 modifications

①  
同定ペプチド

Quantified	Peptide Sequence	Quantified Matches	Fixed Modif...	Variable Mo...	Start	Stop	Protein Accessi...	Proba...	Mass
<input type="checkbox"/>	AAPPPPPPPPLESSPR	0 of 6			606	622	sp Q5T4S7 UBR4...	86%	1,702.904
<input checked="" type="checkbox"/>	AEHASSLLELASTTK	0 of 6			1065	1079	sp Q5T4S7 UBR4...	100%	1,556.805
<input type="checkbox"/>	ALGTLGTTNEK	0 of 6			4803	4814	sp Q5T4S7 UBR4...	99%	1,234.623
<input type="checkbox"/>	APSYIEIFGR	0 of 6			2364	2373	sp Q5T4S7 UBR4...	98%	1,151.598
<input checked="" type="checkbox"/>	AGQALSELHVEK	2 of 6			1828	1850	sp Q5T4S7 UBR4...	100%	1,452.757
<input type="checkbox"/>									485.677
<input checked="" type="checkbox"/>									520.342
<input checked="" type="checkbox"/>									521.247

## Quantified

定量計算に使用される条件(一定数以上のフラグメントの検出など)を満たしたタンパク質の定量計算に使用されているかどうか。手動でチェックを外すと定量計算から外される。

## Quantified Matches

全サンプルに対して定量計算に使用されたサンプルがいくつあるのか

②

同定スペクトル

Sequence	Modifications	Charge	Sample	Quant. Intensity	# Qua...	RT Start (min)	RT Center (min)	RT Stop (min)	Precursor MZ	Attribute
3ALQYDTLISLMEHL...		3	Control_1	2.72E5	5	68.91	69.16	69.36	874.455	control
3ALQYDTLISLMEHL...		3	Control_2	—	0	68.55	—	69.32	874.455	control
ASVVTASSGSALQYDTLISLMEHL...		3	Control_3	4.455E5	5	68.65	68.89	69.13	874.455	control
ASVVTASSGSALQYDTLISLMEHL...		3	Control_2	3.8E5	5	68.97	69.21	69.41	874.455	insulin

## Quant. Intensity

定量計算に利用したフラグメントピーク強度の和

## # of Quant. fragment

定量計算に利用したフラグメントピーク数

## RT Start, RT Center, RT Stop

該当ペプチドを検出し定量計算に利用したRTの開始から終了までの時間

全長とcoverage、修飾残基

Protein Sequence | Protein Level Charts | Chromatograms | Fragment Intensities | Fragmentation Table

sp|Q5T4S7|UBR4\_HUMAN E3 ubiquitin-protein ligase UBR4 OS=Homo sapiens GN=UBR4 PE=1 SV=1

Q H N L L S P P F G W A S G S Q D S N S R R A T T P L Y H G F K E V E E N W S K 920

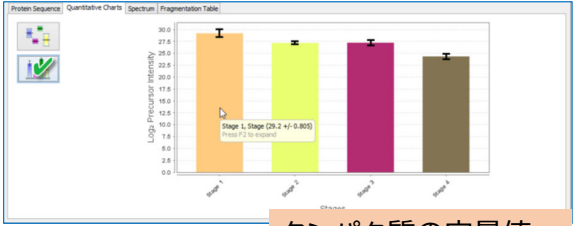
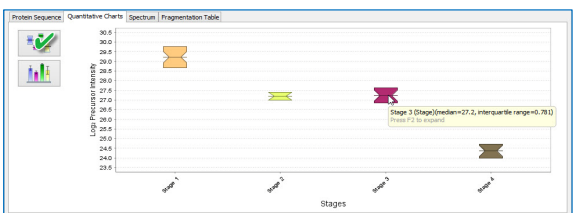
**H F S S D A V P H P R** F Y C V L S P E A S E D D L N R L D S V A C D V L F S K L 960

V K Y D E L Y A A L T A L L A A G S Q L D T V R R K E N K N V T A L E A C A L Q 1000

Y Y F L I L W R I L G I L P P S K **T Y I N Q L S M N S P E M S E C D I L H T L R** 1040

756 out of 5183 (14.59%) amino acids identified with 23 modifications

③ 関連グラフ



タンパク質の定量値

ペプチド/フラグメントのクロマトグラム

Peptide Sequence	Spectral Matches	Modifications	Start	Stop	Protein Accession	Probab...	Mass
IITLAKDQYR	3	C3 Carbamidomethyl...	1071	1079	sp Q13423 NTM...	100%	1,033.523
QGFNWEVSGAGEAK	3		65	107	sp Q13423 NTM...	100%	1,577.769
SGAPELEQL	3		268	279	sp Q13423 NTM...	100%	1,269.692

Valid Sequence Modifications Charge Sample Quant. Intensity # Qua... RT Start (min) RT Center (min) RT Stop (min) Precu... Attrib...

QGFNWEVSGAGEAK 2 20170430\_HeLa\_DIA\_control\_2... 6.96185 5 41.04 41.37 41.63 789.892 Control

QGFNWEVSGAGEAK 2 20170430\_HeLa\_DIA\_insulin\_1... 7.94665 5 41.53 41.87 42.05 789.892 Insulin

QGFNWEVSGAGEAK 2 20170430\_HeLa\_DIA\_insulin\_2... 8.57665 5 41.55 41.88 42.14 789.892 Rapa...

Protein Sequence | Quantitative Charts | Chromatograms | Fragmentation Table

**Precursor** **Fragments**

定量にはPrecursorのピーク強度情報を使っていない

# 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

Scaffold DIA Viewer - Demo2\_HeLa\_insulin\_6\_files\_30\_proteins.sdia

File Edit View Experiment Export Help

Summarization Condition Protein FDR 1.0% FDR Min # Peptides 2

Filters Show Hidden Name/Accession p-value filter GO Term

Organize Samples Proteins **Visualize** Analysis Publish

Proteins 0.0% FDR (attained) 29 Targets 0 Deays

Peptides 0.0% FDR (attained) 1002 Targets 0 Deays

**Volcano plot**

Principal Component Analysis (Log<sub>2</sub> Intensity)

Significance Marker 0-Fold Change Marker Multiselect Action Zoom Reference control

Log<sub>2</sub> p-Value vs Log<sub>2</sub> Fold Change (Precursor Intensity) (insulin/control)

Significant Sig. w/o correction Not significant

**Scatter plot**

Quantitative Scatterplot

contr vs insulin 45° -line Regression Line Multiselect Action Zoom

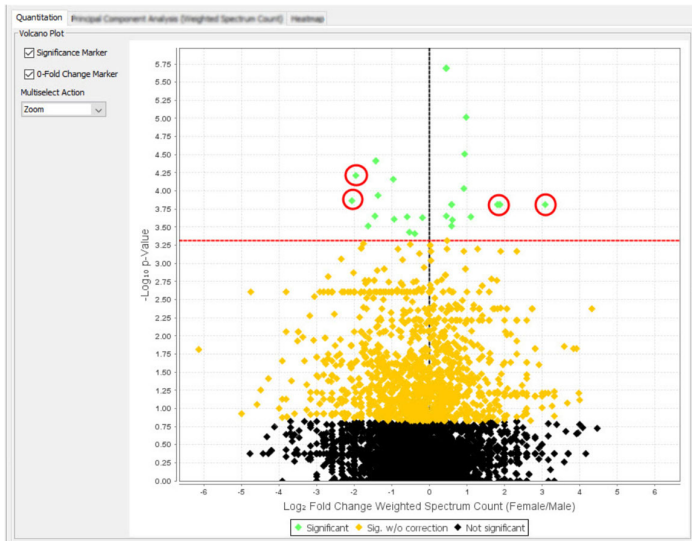
Log<sub>2</sub> Intensity (insulin) vs Log<sub>2</sub> Intensity (control)

Biological Process 3D Values Legend

Cellular process (28) Growth (1) Enzyme regulator activity (2) Biological adhesion (4) Developmental process (13) Establishment of localization (14) Biological regulation (21)

● Growth ● Enzyme regulator activity ● Biological adhesion ● Developmental process ● Biological regulation ● Cellular process

**GO情報**



### Volcano plot

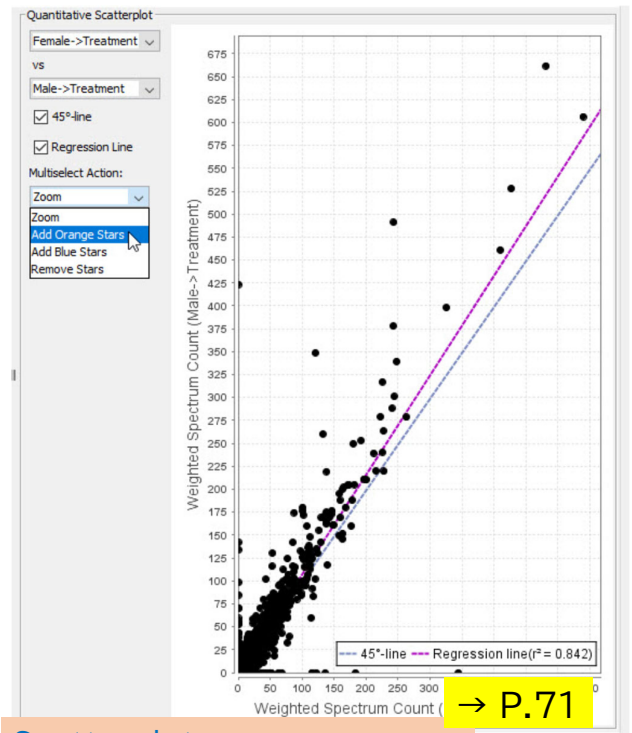
縦軸  $-\log_{10}(p)$

横軸  $\log_2(\text{Fold Change})$

赤点線 多重検定検証の  $p(q, \alpha)$

変動タンパク質をratio, p-value両面から探す

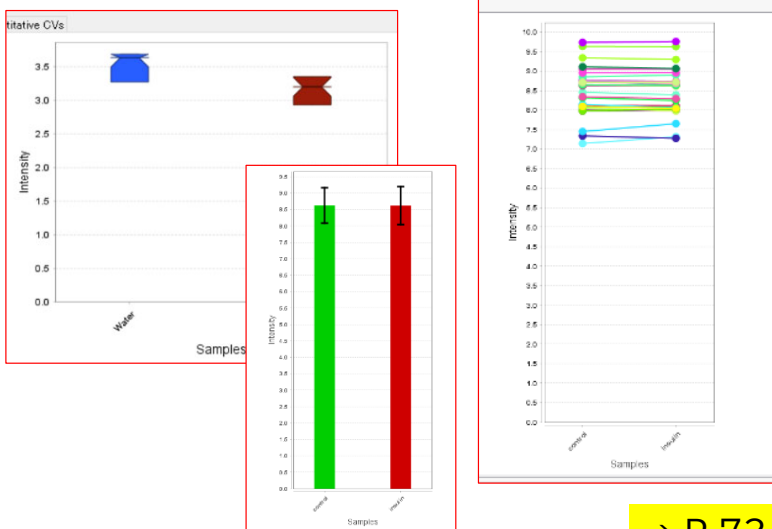
→ P.70



### Scatterplot

$y=x$  や回帰直線から大きく外れるタンパク質を探す

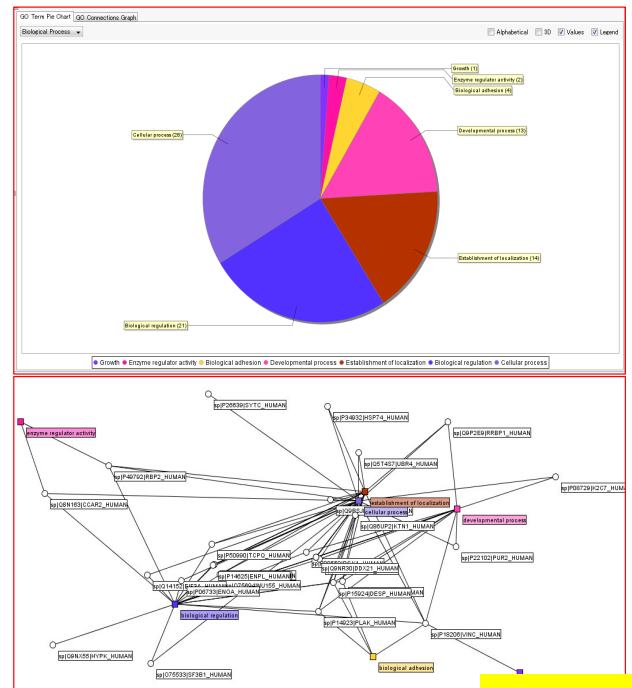
→ P.71



### 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

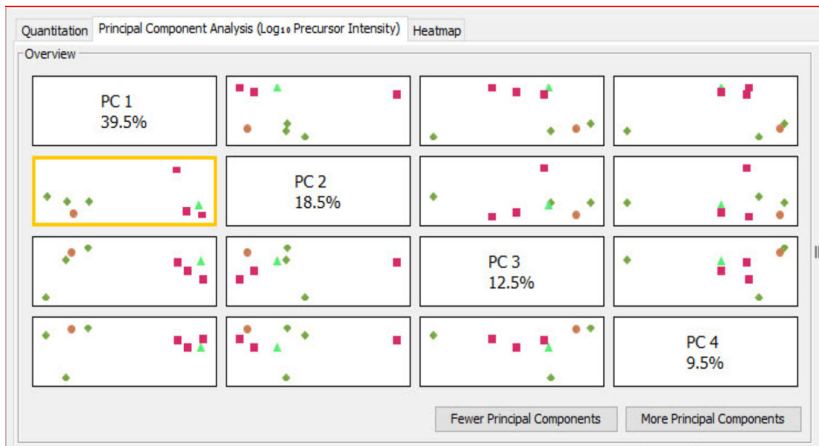
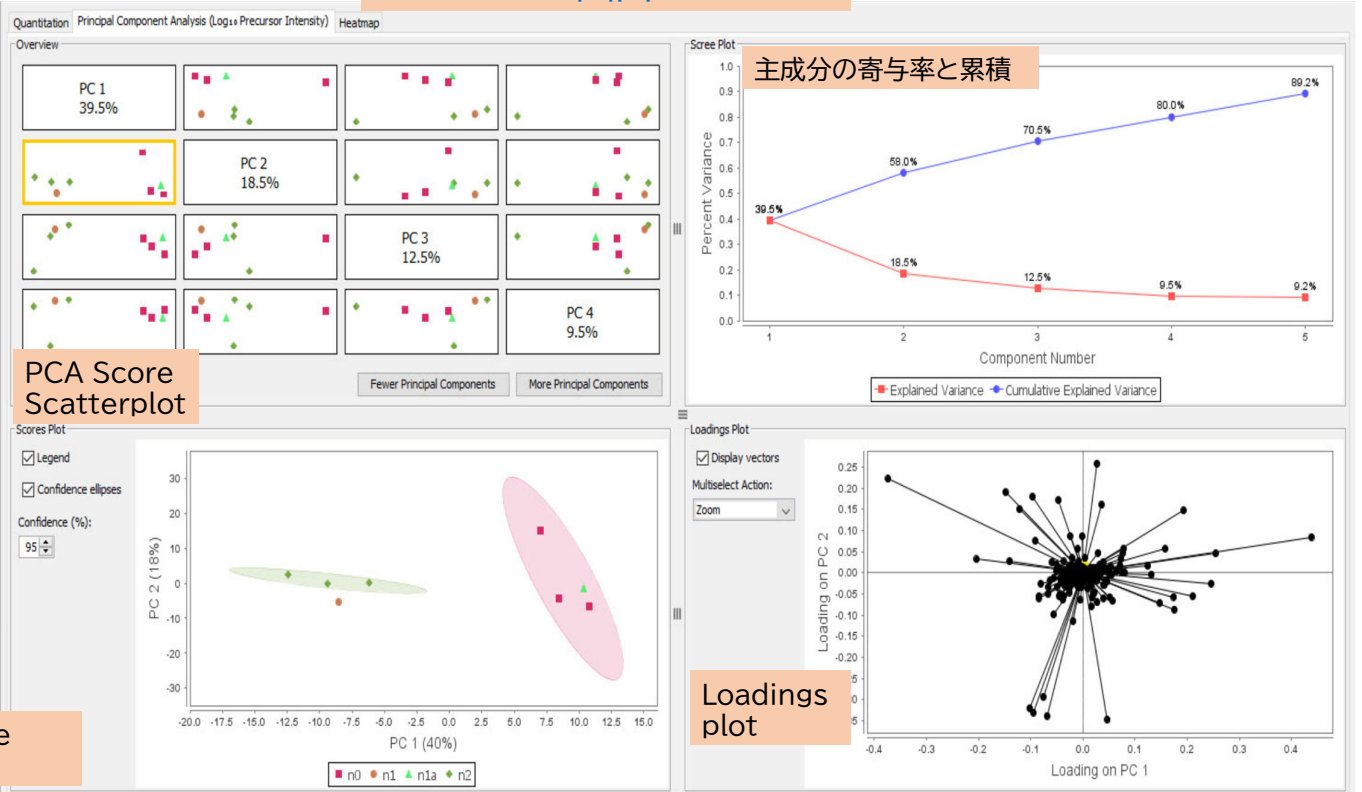
→ P.72



### GO情報

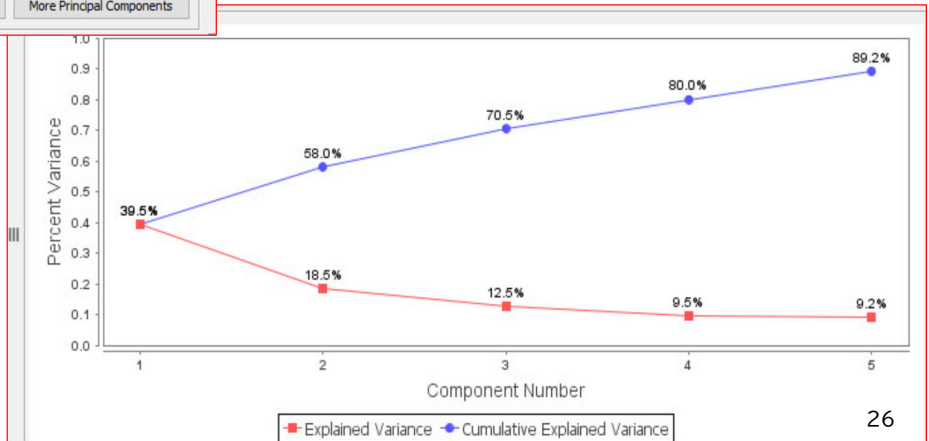
円グラフ(上) ネットワーク図(下)

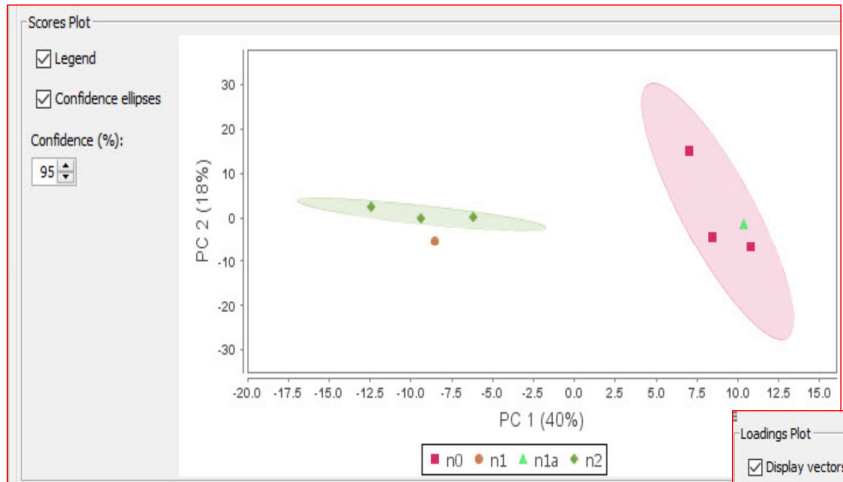
→ P.73



PCA score scatterplot  
主成分別  
プロットはsampleに該当

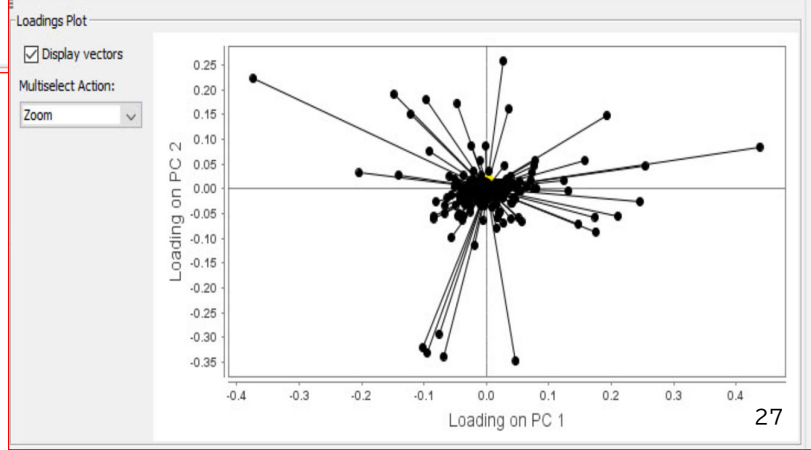
各主成分の寄与率と累積





### Loadings Plot

各タンパク質の主成分への寄与を表すプロット  
左上で選択した主成分組み合わせと連動

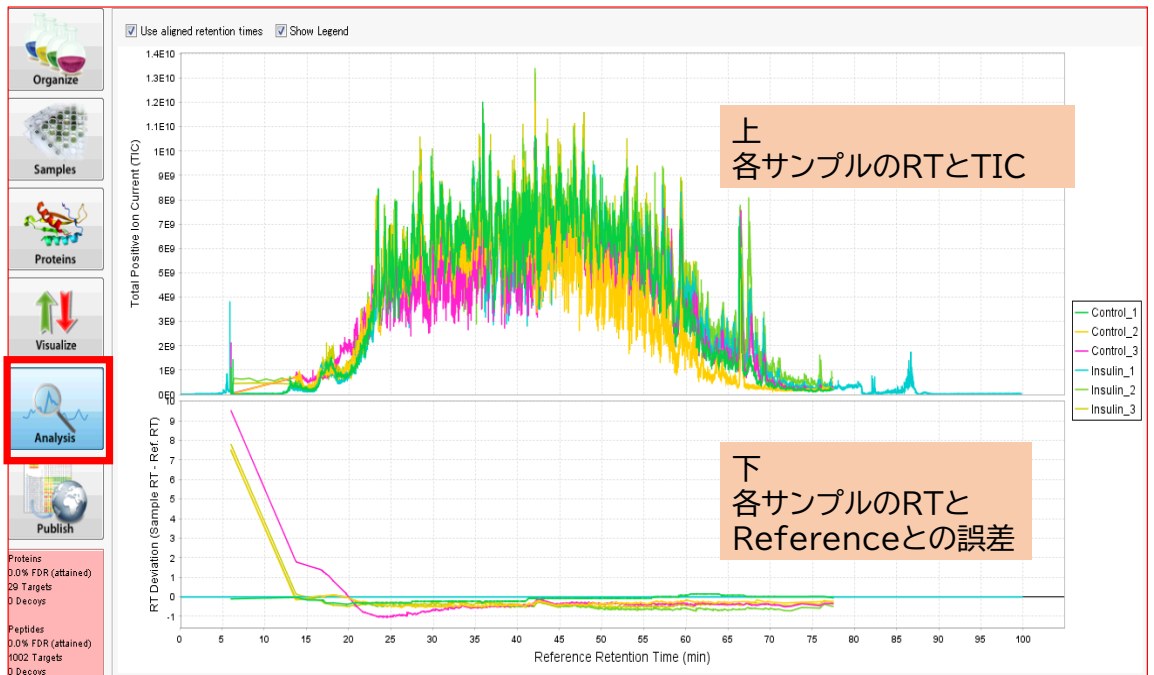


### Score plot

左上で選択した主成分組み合わせのスコア分布図  
拡大版

## Analysis 画面

TICとLC保持時間の  
サンプル間誤差を確認  
できる



Organize

Samples

Proteins

Visualize

**Analysis**

Publish

Proteins  
0.0% FDR (attained)  
29 Targets  
0 Decoys

Peptides  
0.0% FDR (attained)  
1002 Targets  
0 Decoys

# Publish 画面

検索パラメータの確認、論文のMethodのような文章の作成

Experiment Methods | SQL Report

Search

- Search Library: demo2elib
- Processing Directory: #work
- Protein Sequence Database: uniprot-swissprot-humanfasta
- Perform RT alignment between reference: false
- Fragmentation: CID
- Precursor Tolerance: 10.0 ppm
- Fragment Tolerance: 10.0 ppm
- Library Fragment Tolerance: 10.0 ppm
- Peptide FDR Threshold: 0.01
- Data Acquisition Type: Overlapping DIA
- Digestion Enzyme: Trypsin
- Peptide Length: [6 - 30]
- Peptide Charge: [2 - 3]
- Max Missed Cleavages: 1
- Modifications: Carbamidomethylation C 57.0214635 Non-termin...

Analysis

- Shared Evidence Clustering: Perfect
- Target Protein FDR: 0.01
- Minimum Number of Peptides: 2
- Grouping Applied in Version: 1.0.0
- Thresholding Applied in Version: 1.0.0
- Clustering Applied in Version: 1.0.0
- Quantify on Exclusive Peptides: true

Advanced

- Precursor Window Size: Deduced from file
- Minimum Number of Quant Ions: 3
- Maximum Number of Quant Ions: 5

Version

- Scaffold DIA: 1.0.0
- Encyclopedia: 0.6.12
- ProteoWizard: 3.0.11748
- Percolator: 3.0.1nightly-13-655e4c7-dirty

パラメーター一覧

**ANALYSIS OVERVIEW**

DIA data were analyzed using Scaffold DIA (1.0.0).

**RAW DATA PROCESSING**

Raw data files were converted to mzML format using ProteoWizard (3.0.11748). Deconvolution of overlapping windows was performed.

**SPECTRAL LIBRARY SEARCH**

Analytic samples were aligned based on retention times and individually searched against demo2.lib with a peptide mass tolerance of 10.0 ppm and a fragment mass tolerance of 10.0 ppm. Fixed modifications considered were: Carbamidomethylation C. The digestion enzyme was assumed to be Trypsin with a maximum of 1 missed cleavage site(s) allowed. Only peptides with charges in the range [2 - 3] and length in the range [6 - 30] were considered. Peptides identified in each sample were filtered by Percolator (3.0.1nightly-13-655e4c7-dirty) to achieve a maximum FDR of 0.01. Individual search results were combined and peptide identifications were assigned posterior error probabilities and filtered to an FDR threshold of 0.01 by Percolator (3.0.1nightly-13-655e4c7-dirty).

**QUANTIFICATION**

Peptide quantification was performed by Encyclopedia (0.6.12). For each peptide, the 5 highest quality fragment ions were selected for quantitation. Only peptides exclusive to each protein or cluster were used for quantitation.

**CRITERIA FOR PROTEIN IDENTIFICATION**

Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis were grouped to satisfy the principles of parsimony. Proteins with a minimum of 2 identified peptides were thresholded to achieve a protein FDR threshold of 1.0%.

**GO ANNOTATION**

Proteins were annotated with GO terms from: UniProt, InterPro, GO\_Central, Reactome, GOC, HPA, Ensembl, IntAct, Pfam, UniProt, NTLU\_S6, LIPEDb, FlyBase, BHF-UCL, HNFIC, IKG1, SYSCILIA\_CONET, CACAO, AgBase, PINC, ARUK-UCL, CAFA, MTEBASE, Alzheimers\_University\_of\_Toronto, WormBase, GOB, SynGO-UCL, DFLAT, SGD, dictyBase and SynGO

**CITATIONS**

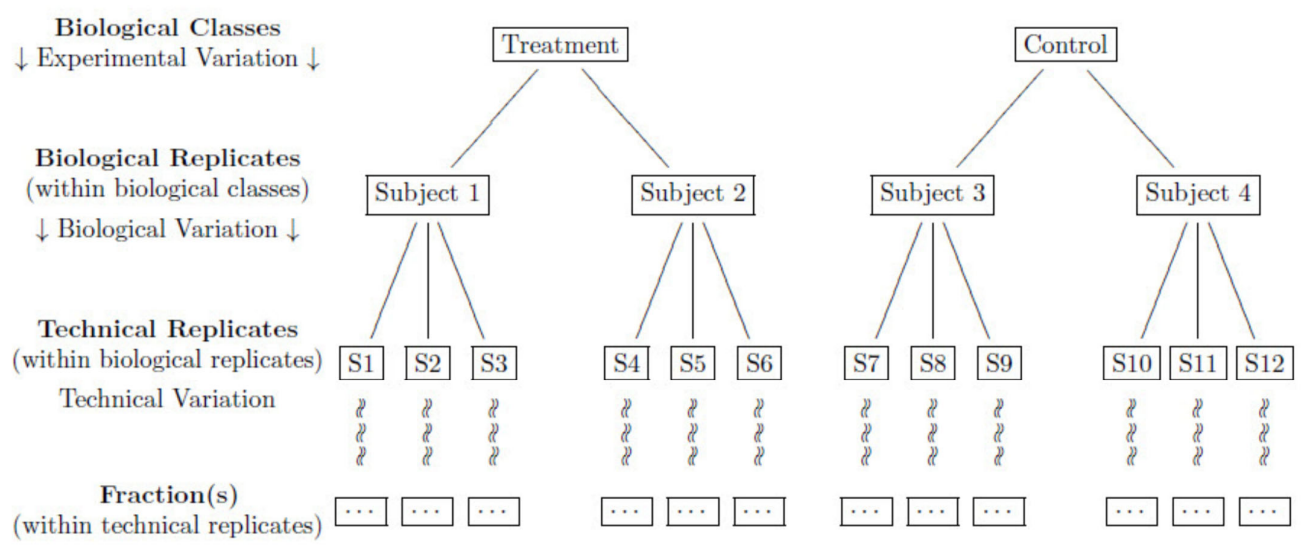
**ProteoWizard**  
A cross-platform toolkit for mass spectrometry and proteomics.  
Chambers, M.C., MacLean, B., Burke, R., Amodei, D., Ruderman, D.L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., Hoff, K., Kessner, D., Toman, N., Shulman, N., Frewen, B., Baker, T.A., Brusniak, M.-Y., Paulse, C., Oresny, D., Flashner, L., Kani, K., Moulden, C., Seymour, S.L., Nuwaysir, L.M., Lefebvre, B., Kuhlmann, F., Roark, J., Rainer, P., Detlev, S., Hemenway, T., Huhmer, A., Langridge, J., Connolly, B., Chadiack, T., Holly, K., Eckels, J., Deutsch, E.W., Moritz, R.L., Katz, J.E., Agus, D.B., MacCoss, M., Tabb, D.L. & Mallick, P. Nature Biotechnology 30, 918-920 (2012) <http://www.nature.com/nbt/journal/v30/n10/full/nbt.2377.html>

**Percolator**  
Semi-supervised learning for peptide identification from shotgun proteomics datasets  
Fujiwara, T., Koehn, C., Grotzer, T., Watanabe, M., Wilson, S., Bradford, A., and Mochly, J. MolCell

Copy Text to Clipboard | Export Publish Report | Export Supplementary Data

Method文章

# 定量：階層構造化と属性付与の例



Sample	Rep1	Rep2	Rep3
S1	2.21	2.31	2.33
S2	2.75	2.74	2.74

データをまとめる  
単位を変更可能

Sample	S1	S2
S1	2.31	2.75
S2	2.35	2.65

31

→ P.85

# 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 Whitney U test / Kruskal-Wallis test

- ノンパラメトリック
- 2群→Mann Whitney 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

**Statistical Test**

ANOVA / t-test P ≥2 Background×Concentrations

Permutation Test NP ≥2 Background×Concentrations

Mann-Whitney U Test NP Exactly 2 Background×Concentrations

Kruskal-Wallis Test NP ≥2 Background×Concentrations

None



# 多重比較補正

## Multiple Test Correction

Control FWER with Hochberg's step-up and Holm's step-down ▾  
Control FDR with standard Benjamini-Hochberg procedure  
Control FWER with Hochberg's step-up and Holm's step-down  
No correction

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**  
ホッホバルクのステップアップ手順&ホルムの  
ステップダウン

33

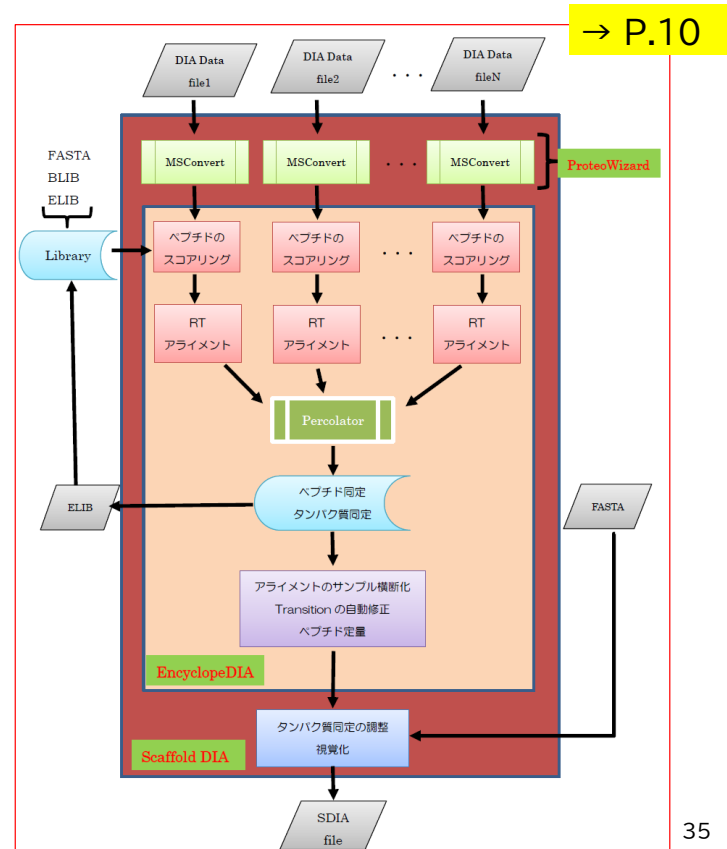
## 解析操作の流れ

- 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

34

# 内部プログラム

- **MSConvert**
  - データ変換
- **EncyclopeDIA**
  - ペプチドピックアップ
  - RTアライメント
  - ペプチド同定
  - タンパク質同定
  - ペプチド定量



## その他のトピックス(+日本語マニュアル対応ページ)

- **タンパク質のグループ化、クラスター化 (→P.89)**  
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.91)**  
Scaffold DIAで出力可能なファイル
- **Prositとの連携によるライブラリフリーサーチ (→P.93)**  
Prositでライブラリを作成しScaffold で使用可能にするための操作方法
- **Gene Ontology**  
Scaffold DIAでGene Ontology情報を付与させるために必要な前処理操作

## 英語マニュアルのAppendix

- **Appendix A.** Structure of Scaffold DIA files (\*.sdia)
- **Appendix B.** Computation of FDR in Scaffold DIA
- **Appendix C.** Summarization: Rolling up Values
- **Appendix D.** Missing Values
- **Appendix E.** Shared Evidence Clustering Algorithm
- **Appendix F.** Heat map clustering
- **Appendix G.** Techniques to Control the Family-wise Error Rate
- **Appendix H.** Using Principal Component Analysis
- **Appendix I.** How PCA is Performed in Scaffold DIA
- **Appendix J.** Description of Mouse Right Click Context Menu Commands

37

## 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やLinux利用の場合、mzMLを自身で準備

#### ・メモリ

最低 4GB以上

しかし64GB以上を強く推奨(大規模解析なら128GB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

38