Scaffold DIA マニュアル



1.	はじめに	4
1	1-1. Scaffold DIA とは	4
1	1-2. インストール・ライセンシングについて	5
1	1-3. 本資料の構成について	5
2.	データの準備	6
2	2-1. 取り込むデータの種類	6
2	2-2. DIA と DDA の比較	6
2	2-3. Scaffold DIA ワークフロー	7
2	2-4. Scaffold DIA データ処理の流れ	9
3.	データの取り込み	13
3	3-1. 取り込み可能な raw ファイルのフォーマット	13
3	3-2. コマンドラインでの利用	13
3	3-3. ライブラリマネージャー	14
3	3-4. 解析の開始とワークフローダイアログ	16
	3-4-1. ワークフローダイアログの Search タブ	17
	3-4-2. ワークフローダイアログの Analysis タブ	24
	3-4-3. ワークフローダイアログの Advanced タブ	25
4.	Main 画面	. 26
4.	Main 画面 4-1. Scaffod DIA メイン画面について	. 26 26
4. 2	Main 画面 4-1. Scaffod DIA メイン画面について 4-3. ツールバー(アイコン)	. 26 26 30
4. 2 2	Main 画面 4-1. Scaffod DIA メイン画面について 4-3. ツールバー(アイコン) 4-4. 表示タンパク質のフィルターリング	. 26 26 30 31
4. 2 2 2	Main 画面 4-1. Scaffod DIA メイン画面について 4-3. ツールバー(アイコン) 4-4. 表示タンパク質のフィルターリング 4-5. 各画面への切り替え	. 26 26 30 31 32
4.	Main 画面 4-1. Scaffod DIA メイン画面について 4-3. ツールバー(アイコン) 4-4. 表示タンパク質のフィルターリング 4-5. 各画面への切り替え 4-6. FDR 計算値表示	. 26 26 30 31 32 32
4.	Main 画面 4-1. Scaffod DIA メイン画面について 4-3. ツールバー(アイコン) 4-4. 表示タンパク質のフィルターリング 4-5. 各画面への切り替え 4-6. FDR 計算値表示 4-7. Summarization	. 26 26 30 31 32 32 33
4. 2 2 2 2 2 5.	Main 画面 4-1. Scaffod DIA メイン画面について 4-3. ツールバー(アイコン) 4-4. 表示タンパク質のフィルターリング 4-5. 各画面への切り替え 4-6. FDR 計算値表示 4-7. Summarization Organize 画面	. 26 26 30 31 32 32 33 . 34
4. 2 2 2 2 5.	Main 画面 4-1. Scaffod DIA メイン画面について 4-3. ツールバー(アイコン) 4-4. 表示タンパク質のフィルターリング 4-5. 各画面への切り替え 4-6. FDR 計算値表示 4-6. FDR 計算値表示 4-7. Summarization Organize 画面 5-1. Organize パネルについて	. 26 30 31 32 32 33 . 34
4. 2 2 2 5.	Main 画面 4-1. Scaffod DIA メイン画面について	. 26 30 31 32 32 33 . 34 34 34
4. 2 2 2 2 5. 5.	Main 画面	. 26 30 31 32 32 33 . 34 34 34 35
4. 2 2 2 2 5. 5 5	Main 画面 4-1. Scaffod DIA メイン画面について 4-3. ツールバー(アイコン) 4-4. 表示タンパク質のフィルターリング 4-5. 各画面への切り替え 4-6. FDR 計算値表示 4-7. Summarization Organize 画面 5-1. Organize パネルについて 5-2. Organize パネルの画面と設定箇所 5-3. Category, Attribute の定義設定 5-4. Sample ファイルと Attribute の関係性を設定・確認する表	. 26 30 31 32 32 33 . 34 34 34 35 38
4. 2 2 2 2 5. 5.	Main 画面 4-1. Scaffod DIA メイン画面について 4-3. ツールバー(アイコン) 4-4. 表示タンパク質のフィルターリング 4-5. 各画面への切り替え 4-6. FDR 計算値表示 4-7. Summarization Organize 画面 5-1. Organize パネルについて 5-2. Organize パネルの画面と設定箇所 5-3. Category, Attribute の定義設定 5-4. Sample ファイルと Attribute の関係性を設定・確認する表 5-5. Import Attributes File / Export Attributes File	. 26 30 31 32 32 33 . 34 34 34 35 38 41
4. 2 2 2 2 5. 5.	Main 画面 4-1. Scaffod DIA メイン画面について 4-3. ツールバー(アイコン) 4-4. 表示タンパク質のフィルターリング 4-5. 各画面への切り替え 4-6. FDR 計算値表示 4-7. Summarization Organize 画面 5-1. Organize パネルについて 5-2. Organize パネルの画面と設定箇所 5-3. Category, Attribute の定義設定 5-4. Sample ファイルと Attribute の関係性を設定・確認する表 5-5. Import Attributes File / Export Attributes File 5-6. 選択可能な 3 種類の Experimental Design について	. 26 30 31 32 32 33 . 34 34 34 35 38 41 41
4. 2 2 2 5. 5 5	Main 画面 4-1. Scaffod DIA メイン画面について 4-3. ツールバー(アイコン) 4-4. 表示タンパク質のフィルターリング 4-5. 各画面への切り替え 4-6. FDR 計算値表示 4-7. Summarization Organize 画面 5-1. Organize パネルについて 5-2. Organize パネルについて 5-3. Category, Attribute の定義設定 5-4. Sample ファイルと Attribute の関係性を設定・確認する表 5-5. Import Attributes File / Export Attributes File 5-6. 選択可能な 3 種類の Experimental Design について 5-7. Experimental Design の設定	. 26 30 31 32 32 32 33 . 34 34 34 35 38 41 43
4. 2 2 2 5. 5 5	Main 画面 4-1. Scaffod DIA メイン画面について 4-3. ツールバー(アイコン) 4-4. 表示タンパク質のフィルターリング 4-5. 各画面への切り替え 4-6. FDR 計算値表示 4-7. Summarization Organize 画面 5-1. Organize パネルについて 5-2. Organize パネルについて 5-3. Category, Attribute の定義設定 5-4. Sample ファイルと Attribute の関係性を設定・確認する表 5-5. Import Attributes File / Export Attributes File 5-6. 選択可能な 3 種類の Experimental Design について 5-7. I. Sample Hierarchy タブ	. 26 30 31 32 32 32 33 . 34 34 34 34 34 34 34 34 34 34 41 43 44

6. Samples 画面	50
6-1. Samples 画面について	50
6-2. Samples Table(Fig6-1の[1])	50
6-3. Display Type バー (Fig6-1の②)	
6-4. 「Summarization」と定量値	
6-5. 定量値データのまとめられ方	
7. Proteins 画面	
7-1. Proteins 画面について	
7-2. ペプチドフィルターリング(Fig.7-1 [1])	
7-3. ペプチド情報(Fig.7-1 [2])	
7-4. ペプチド – Query マッチ情報 (Fig.7-1 [3])	61
7-5. Visualization 情報(Fig.7-1 [4])	61
7-5-1. Protein Sequence タブ	
7-5-2. Protein Level Chart タブ	
7-5-3. Statistics タブ	
7-5-4. Protein Intensities タブ	
7-5-5. Fragment Intensities タブ	
7-5-6. Chromatograms タブ	
7-5-7. Fragmentation table タブ	
7-6. Similar Proteins タブ	67
8. Visualize 画面	
8. Visualize 画面 8-1. Visualize 画面について	
8. Visualize 画面 8-1. Visualize 画面について 8-2. Quantitation タブ	
8. Visualize 画面 8-1. Visualize 画面について 8-2. Quantitation タブ 8-2-1. Protein Sequence	
8. Visualize 画面 8-1. Visualize 画面について 8-2. Quantitation タブ 8-2-1. Protein Sequence 8-2-2. Quantitative Scatterplot	68
8. Visualize 画面 8-1. Visualize 画面について 8-2. Quantitation タブ 8-2-1. Protein Sequence	68
8. Visualize 画面 8-1. Visualize 画面について 8-2. Quantitation タブ 8-2-1. Protein Sequence 8-2-2. Quantitative Scatterplot 8-2-3. Quantitative Trend Chart & CVs 8-2-4. GO Term Pie Chart, GO Connections Graph	68
8. Visualize 画面 8-1. Visualize 画面について 8-2. Quantitation タブ 8-2-1. Protein Sequence 8-2-2. Quantitative Scatterplot 8-2-3. Quantitative Trend Chart & CVs 8-2-4. GO Term Pie Chart, GO Connections Graph 8-3. Principal Component Analysis タブ	
8. Visualize 画面 8-1. Visualize 画面について 8-2. Quantitation タブ 8-2-1. Protein Sequence 8-2-2. Quantitative Scatterplot 8-2-3. Quantitative Trend Chart & CVs 8-2-4. GO Term Pie Chart, GO Connections Graph 8-3. Principal Component Analysis タブ 8-3-1. Overview	68
8. Visualize 画面 8-1. Visualize 画面について 8-2. Quantitation タブ 8-2-1. Protein Sequence 8-2-2. Quantitative Scatterplot 8-2-3. Quantitative Trend Chart & CVs 8-2-4. GO Term Pie Chart, GO Connections Graph 8-3. Principal Component Analysis タブ 8-3-1. Overview 8-3-2. Scree Plot	68
8. Visualize 画面 8-1. Visualize 画面について 8-2. Quantitation タブ 8-2-1. Protein Sequence 8-2-2. Quantitative Scatterplot 8-2-3. Quantitative Trend Chart & CVs 8-2-4. GO Term Pie Chart, GO Connections Graph 8-3. Principal Component Analysis タブ 8-3-1. Overview 8-3-2. Scree Plot 8-3-3. Scores Plot	68
8. Visualize 画面 8-1. Visualize 画面について 8-2. Quantitation タブ 8-2-1. Protein Sequence 8-2-2. Quantitative Scatterplot	68 69 70 71 71 72 73 74 75 75 76 76
8. Visualize 画面 8-1. Visualize 画面について 8-2. Quantitation タブ 8-2-1. Protein Sequence 8-2-2. Quantitative Scatterplot 8-2-3. Quantitative Trend Chart & CVs 8-2-4. GO Term Pie Chart, GO Connections Graph 8-3. Principal Component Analysis タブ 8-3-1. Overview 8-3-2. Scree Plot 8-3-3. Scores Plot 8-3-4. Loading Plot 8-4. Heatmap タブ	68 69 70 71 71 72 73 74 75 75 76 76 77
8. Visualize 画面について 8-1. Visualize 画面について 8-2. Quantitation タブ 8-2-1. Protein Sequence 8-2-2. Quantitative Scatterplot	68 69 70 71 72 73 74 75 75 76 76 77
8. Visualize 画面 8-1. Visualize 画面について 8-2. Quantitation タブ 8-2-1. Protein Sequence 8-2-2. Quantitative Scatterplot 8-2-3. Quantitative Trend Chart & CVs 8-2-4. GO Term Pie Chart, GO Connections Graph 8-3. Principal Component Analysis タブ 8-3-1. Overview 8-3-2. Scree Plot 8-3-3. Scores Plot 8-3-4. Loading Plot 8-4.1. Heatmap タブ 8-4-2. Heatmap Details pane	68 68 69 70 71 72 73 74 75 76 77 77 77 78
8. Visualize 画面 8-1. Visualize 画面について 8-2. Quantitation タブ 8-2-1. Protein Sequence	68 686970717272737475757576767777777777777777

10. Publish 画面	80
10-1. Publish 画面について	80
10-2. Experiment Methods タブ	81
10-3. SQL Report タブ	83
11.定量の手法と検定機能	
11-1. 定量の手法について	84
11-1-1. Peptide Counting	84
11-1-2. ラベルフリー定量	84
11-2. 検定	86
11-2-1. Basic Design で実施可能な検定	86
11-2-2. Repeated Measures で実施可能な検定	87
11-2-3. Two-way で実施可能な検定	87
11-3. 多重検定補正	88
12. タンパク質のグループ化、クラスター化	89
12-1. 同定タンパク質のグループ化、クラスター化	89
13. Report	91
13-1. Scaffold DIA で出力可能なファイルについて	91
14. Prosit との連携によるライブラリフリーサーチ	93
14-1. Proteome Software 社が公開している DLIB ファイルを利用する	93
14-2. FASTA ファイルを元に Prosit を利用して DLIB ファイルを作成	95
14-2-1. FASTA ファイルを取得する(例:Uniprot サイト)	95
14-2-2. Scaffold DIA で Prosit 用のインプットファイルを作成する	100
14-2-3. Prosit 公開サイトにてペプチド配列情報から計算を行う	101
14-2-4. Scaffold DIA で Prosit 出力ファイルと FASTA ファイルから DLIB ファイルを作成する	3 105
14-3. DLIB 並びに FASTA ファイルを Scaffold DIA にセットする方法	106
15. Gene Ontology	108
15-1. Gene Ontology ファイルを Scaffold DIA にセットする方法	108
15-2. Scaffold DIA の各タンパク質に Gene Ontology 情報を表示させる方法	110
15-2-1. 表示する GO 項目を選択する	110
15-2-2. タンパク質と GO 情報との紐づけを行う	112
16. デモデータと追加情報のご案内	114
16-1. デモデータについて	114
16-2. Appendix について	114
16-3. EncyclopeDIA について	115

1. はじめに

1-1. Scaffold DIA とは

Scaffold DIA は、DIA 定性/定量プロテオミクス解析において、RAW データ読み込みから解析、結果 表示までの一連作業がワンパッケージで行えるソフトウェアです。Scaffold DIA の機能に ついて、以下の 4 つ(5 つ)のキーワードで説明いたします。

- Identify
- •Organize,Summarize
- •Quantify
- •Visualize

Identify

事前に作成した DDA ピークリストデータのライブラリと比較して DIA データからタンパク質を同定する 事ができます。また FASTA 配列から作成された仮想ピークパターンに対して検索を行う事もできます。 同定ペプチド・タンパク質は FDR を基に再評価されるほか、同定タンパク質はシェアペプチドの状況により グループ化やクラスタリングが行われます。

■Organize と Summarize

取り込んだデータに対して属性グループを定め、属性値を付与する事ができます。定められた属性を基に 定量値や検定結果をまとめる事ができます。

Quantify

各ペプチドの代表的なイオンフラグメントに基づいて行われます。サンプル間の正規化が行われ、 各ペプチドの定量値が合計されてタンパク質の定量値が算出されます。定量値に基づいて様々な検定・ 統計解析を実施する事ができます。

Visualize

表示するタンパク質はフィルター機能などで調整可能です。タンパク質のリストには、Gene Ontology の 情報を併記することができます。検定・統計解析を行った結果を表示する各種方法に対応しています。 定められた属性ごとに計算される定量値は、選択肢を変更する事でインタラクティブに表示内容を変更 する事ができます。

1-2. インストール・ライセンシングについて

インストール並びにライセンシングについては、別資料「ScaffoldDIA インストール手順.pdf」を ご覧ください。インストール DVD 内にも含まれています。

https://www.matrixscience.co.jp/supportpdf/scaffold/ScaffoldDIA_install.pdf

1-3. 本資料の構成について

本資料では Scaffold DIA の基本的な使い方について説明しています。

2,3 章では 解析対象のデータとその取り込みについて、

4章では ソフトウェア画面のメニューや全体像の説明、

5章~10章では画面左側にあるパネルの切り替えで表示させる各画面の説明、

11章では使用可能な検定と Significant Level の設定、

12章ではタンパク質のグループ化に関する設定、

13 章では出力可能な report、

14 章ではライブラリフリーサーチ作成の手順、

15 章では Gene Ontology について、

16章ではデモデータや参考資料などその他の情報について、

それぞれご紹介しています。

2. データの準備

この章では Scaffold DIA に取り込む測定データの特徴について説明しています。

2-1. 取り込むデータの種類

これから DIA 測定を始める方に対して解析のヒントを記した資料が「Designing a DIA Method」 <u>https://www.proteomesoftware.com/documentation/scaffold_dia_ms_methods/</u> です。ご興味がございましたらご参照ください。

2-2. DIA と DDA の比較

DDA ワークフローの場合、再現性とカバレージを両立させようとすると測定しなければならない データ数が膨大となります(Fig.2-1)。DIA は測定の数を減らしながらカバレージを落とさず網羅的に解析 ができる、バランスの良い測定手法と言えます(次頁 Fig.2-2)。



Fig.2-1 DDA 解析におけるデータ収集



Fig.2-2 DIA 解析におけるデータ収集

2-3. Scaffold DIA ワークフロー

Scaffold DIA は DIA データ内とライブラリにおけるフラグメントパターンを比較する事で、ペプチド/タンパク質の同定と定量を同時に行います。



Fig.2-3 Scaffold DIA のワークフロー

検索時に参照する「ライブラリ」については、主に以下の4つの方法に対応しています。

- **FASTA** ファイルを使う方法

FASTA フォーマットのタンパク質配列ファイルを使って解析できます。ピーク強度情報は予測 しません。

Prosit ライブラリを使う方法(DLIB を使う方法)

FASTA フォーマットのタンパク質配列情報から、プログラム"Prosit"を介して作成された スペクトルライブラリに対して検索が可能です。ファイル拡張子は .dlib です。ライブラリは、 Proteome Software 社サイト

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

から取得できるほか、ご自身で FASTA ファイル元に Prosit のサイト

https://www.proteomicsdb.org/prosit/

で作成し得られたファイル(拡張子が .csv または .spectronaut)を使用する事もできます。 ライブラリ作成について詳しい手順を 14 章(P.94〜)にまとめたので、作成の際はそちらをご参照 ください。

- BLIB を使う方法 (Skyline または Scaffold 5 経由)

DDA 解析時に得られたピークリストライブラリです。ソフトウェア Skyline や Scaffold 5 上で 作成できます。Skylineなどの他ソフトウェアから得た iRTDB ファイルも併せて取り込む事ができ、 RT のアライメントに利用する事ができます。blib 以外に、dlib, traML,MSP,SPTXT といった フォーマットにも対応しています。

- ELIB を使う方法

Scaffold DIA が作成したクロマトグラムのライブラリで、BLIB,DLIB や FASTA を使った解析の 結果得られた、ピークと RT なども含むクロマトグラム情報が組み合わさったエントリーの ライブラリです(Fig.2-4)。

ELIB と DLIB は同じフォーマットで実質中身が同じであり、相互互換があります。Seaffold の
 ELIB 用/DLIB 用オプションはどちらも利用可能です



Fig.2-4 ELIB 作成例

一方 Scaffold DIA では pool のサンプルと個別のサンプルを組み合わせた解析法にも対応しています。 例えば pool されたサンプルを FASTA や BLIB(DDA 由来のピークライブラリ)に対して検索を行い その 結果から ELIB を作成し、個々の DIA データをその ELIB に対して検索させることができます(Fig.2-5、 次頁図)。

Scaffold DIA では様々な解析方法に対応しています。データ取り込みについては「3 章データ取り込み」 にて詳しく説明しています。



Fig.2-5 個別の DIA データと pooled データを組み合わせた解析

2-4. Scaffold DIA データ処理の流れ

データ処理の流れを図示したのが、次頁図「Fig.2-6 Scaffold DIA データ処理の流れ」です。Scaffold DIA では raw データの変換に、外部プログラムの「**ProteoWizard**」を利用しています。その後、 読み 込んだデータを解析に中心的な役割を果たすのが「**EncyclopeDIA**」というプログラムです。 https://bitbucket.org/searleb/encyclopedia/wiki/Home

ペプチドの特定とスコアリング、保持時間のアライメント、Percolator によるスコアリング結果の再評価、 ペプチド同定の決定とタンパク質同定、各種計算結果のサンプル横断的な補正、定量、といった処理が EncyclopeDIA 上で行われます。EncyclopeDIA はワシントン大学の MacCoss 教授の研究室(旧所属)で Proteome Software 社の製品開発担当者 Brian Searle が作成・公開しているプログラムです。 次々項以降に、以下の処理・計算について詳しく説明しています。

- RAW ファイルの変換
- ペプチドの同定
- Decoy データの作成方法
- ペプチドのスコアリング
- 保持時間のアライメント
- ペプチドの定量



Fig.2-6 Scaffold DIA データ処理の流れ

■ RAW ファイルの変換

Scaffold DIA では raw データの変換に、ProteoWizard で提供されている「MSConvert」を 利用しています。Scaffold DIA で対応する RAW データは、MSConvert の対応に依存します。 MSConvert では raw データを共通 XML フォーマットである「mzML」フォーマットに変換します。 変換時の一時ファイルはソフトウェア内で「Processing directory」として定義されたフォルダに 保存されます。使用後は基本的にファイルを削除しますが、もしデータを残すよう指定していた 場合、データは削除されず残ります。

■ ペプチドの同定

ペプチドの同定などにはプログラム「EncyclopeDIA」が使われています。同定に際し測定 データとライブラリーデータとの比較が行われます。ライブラリーデータには前述のように 幾つかのバリエーションがあります。

・BLIB : DDA 測定データのピークリストライブラリ : Skyline や Scaffold から出力

・FASTA: タンパク質のアミノ酸配列情報から単純なピークパターンを計算

・Prosit Library:アミノ酸配列情報から計算された理論ピークリスト。RT や強度情報を含む

・ELIB : Scaffold DIA で検索された過去の解析データ。RT や強度情報を含む

■ Decoy データの作成方法

EncyclopeDIA の中で Percolator の計算を行う際、decoy データベースを使用します。decoy の配列は末端を固定して target エントリーのペプチドを逆向きにする事で作成します。ピークに ついては逆向きにした配列が来るべきフラグメントピークの位置に合うように m/z をシフトし、 強度についてはそのまま利用します。

■ ペプチドのスコアリング

EncyclopeDIA によって計算される主要なスコア、[primary peptide score]は、純粋に プレカーサー分離ウィンドウ内の実測ピークとライブラリ内のピークとのマッチングから評価 します。DIA では共溶出ペプチドの存在が当たり前である事を踏まえ、マッチしなかったイオンに 関してペナルティを与えていません。1つのウィンドウ内で複数のペプチドフラグメントピークとの マッチを評価します。[primary peptide score] はX!TandemのHyperScoreと概念が似ていて、 測定したスペクトルとライブラリのスペクトルにおけるフラグメントの強度の加重ドット積に、 マッチしたイオンの数の階乗をかけたものです。そのほか関連スコアとしてデルタ CN や フラグメントイオンの精度、プレカーサーマスの精度、保持時間精度などを評価したスコアも計算 されます。

[次頁に続きます]

■ 保持時間のアライメント

EncyclopeDIA では同定されたペプチドすべてについて、x 軸がライブラリ内にある保持時間、 y 軸が測定値の保持時間のプロットを取ります(Fig.2-7)。続いてカーネル密度推定(ヒストグラム から計算された曲線)により 1000x1000 グリッドデータに補正され線が引かれます。サンプル間 にわたりアライメントの調整が行われる段階では、構成要素のペプチド数が最も多いサンプルを リファレンスサンプルと定め、そのリファレンス内のペプチドを基準に他のサンプルのペプチドの 保持時間を調整します。保持時間のずれが大きく異常値と判定された各サンプルのペプチドに ついてはマッチの再評価が試みられます。



Fig.2-7 RT Alignment

■ ペプチド定量

ペプチドの定量はフラグメントのピーク強度のみを定量値として利用しており、プリカーサーピークの 強度の値は使っていません。

3. データの取り込み

この章では Scaffold DIA でのデータ取り込みについて説明いたします。

3-1. 取り込み可能な raw ファイルのフォーマット

Scaffold DIA では データ取り込みの際、ProteoWizard で提供されているプログラム 「msconvert.exe」を利用しており、取り込み可能な raw ファイルのフォーマットもそのプログラムの対応に 準じます。読み込み可能なファイルをまとめたのが以下の表です(Tab.3-1 Scaffold DIA で取り込み可能 なファイルフォーマット)。例として挙げたメーカーのファイル読み込みについてはよく確認していますが、 それ以外のメーカーでも raw データの読み込みが可能である他、mzML に変換されたデータについても 基本的に読み込みが可能です。

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			
AB/SCIEX	*.wiff (.wiff.scan も同じフォルダに置く事)			
Agilent	*.d (ディレクトリ)			
Thermofisher scientific *.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				

Tab.3-1 Scaffold DIA で取り込み可能なファイルフォーマット

3-2. コマンドラインでの利用

Scaffold DIA では、サーバーでの利用や自動化処理を目的として、コマンドラインでも利用できます。 実行時にワークフローの定義を行っているファイルと raw ファイルを読み込みます。コマンドラインの インターフェースを利用する場合、「ScaffoldDIABatch.exe」プログラムを利用します。「-help」で使い方に 関する説明文が現れます。また以下の HELP サイトを利用すると実行したい内容を引数として指定 するにはどうするか、より簡単に理解できます。

https://www.proteomesoftware.com/annotated-scafml

ご不明な点はサポートへお問い合わせください。

3-3. ライブラリマネージャー

検索を行う際に使用するライブラリについて、ダウンロード・ライブラリ作成と管理・選択を行うのが ライブラリマネージャーです。ライブラリマネージャーを通じて検索時に複数のライブラリを選択したり、 FASTA ファイルとの紐づけを予め指定したりすることも可能です。

🧱 Library Manager		×
Name	Status	FASTA
human_hcd_selected.msp	Ready	
saccharomyces_cerevisiae_prosit_generated_library.dlib	Pending	saccharomyces_cerev
human_coronavirus_combined_prosit_generated_library.dlib	Ready	human_coronavirus
Label-Free.blib	Ready	
Add Library Remove Library Creation	eate Library	Download OK

Fig.3-1 ライブラリマネージャー ダイアログ

ライブラリマネージャーは以下の3つの操作のいずれかで起動できます。

- 「File」メニューから「Open Library manager」を選択
- ツールバーの「ライブラリマネージャー」アイコン
- 「Workflow」ダイアログの「reference file」選択ボックスで「Choose...」をクリックすると表示される ドロップダウンから「Library manager」を選択

画面下の各ボタンについて、以降より詳しく説明します。

"Add Library"

新たにライブラリを追加します。ボタンを押すとファイル選択ダイアログが現れます。ライブラリファイル (FASTA 以外)を選択して Scaffold DIA に取り込みますが、この段階では対応する FASTA ファイルとの 紐づけが行われていません。紐づけを行うには、ライブラリを選択して右クリックし、「Associate with FASTA」を選択して、対応する FASTA ファイルを選んでください(次頁 Fig3-2)。紐づけを行う事で、 ライブラリ検索におけるペプチドとタンパク質の連結が行われ、タンパク質全長の配列や分子量の情報を 取得できるようになります。

🚛 Library Manager		×
Name	Status	FASTA
arabidopsis_thaliana_prosit_gener	Readv	
	Associate with FASTA	
	Show in file browser	
Add Library	Remove Library Create Li	brary Download OK

Fig.3-2 ライブラリマネージャー ライブラリと FASTA の紐づけ操作

"Remove Library"

作成したライブラリを Library Manager から削除します。

"Create Library"

検索を行いその結果を DIA の検索結果ライブラリ(ELIB)として保存する事ができます。検索後作成 されたライブラリは自動的に Library Manager に、対応する FASTA ファイル情報と共に登録されます。

"Download"

Proteome Software 社が作成した、Prosit 由来のピークリストファイルをダウンロードしセットする事が できます。ボタンを押すと Fig3-3 のようなダイアログが現れます。①で生物種を選ぶと、データベースの 詳細情報と共に URL が③に自動的に入力されます(次頁 Fig.3-4)。④でファイルダウンロード先並びに ダウンロード後のファイル名を選択すると"Download"ボタンを押すことができるようになります。この時、 ライブラリに対応する FASTA ファイルも同時に取得します。

Name		Status		FASTA
🧱 Download Lib	rary			×
Public libraries	Choose 🚺		✓ Reque	st Prosit Library.
URL 🔺	3			
Destination 📥		4		Choose
				1

Fig.3-3 ライブラリマネージャー "Download" ダイアログ

Public libraries	Arabidopsis thaliana	Request Prosit Library		
	DLIB size	1.2gb		
	FASTA size	8.7mb		
	FASTA	arabidopsis_thaliana_reviewed_unipr.		
	Min charge	2		
	Max charge	3		
	Missed cleavages	1		
	Min m/z	396.4		
	Max m/z	1002.7		
	Normalized collision energy	33		
	Default charge	3		
	Fixed mod(s)	["carbamidomethylation"]		
	Variable mod(s)	0		
	Source	Proteome Software		
	Туре	prosit derived		
URL 🗸	esoftware.net/prosit_dlibs/arabidopsis_thaliana_prosit_generated			
Destination 🖌		Choose		

Fig.3-4 ライブラリマネージャー "Download" ダイアログ ライブラリの詳細情報

3-4. 解析の開始とワークフローダイアログ

新しく解析を始めるには、ソフトウェアを起動後以下のどちらかの操作を行います。

- メニューの File -> New
- Ctrl + N
- メニューバー下にあるアイコン
 シ をクリック

新しいファイルが作成されると最初に現れるのが「ワークフローダイアログ」です(Fig.3-5)。ダイアログの タイトル欄には「Load Data」と表示されています。

kan Load Data	×
▲ Search ✓ Analysis ✓ Advanced	
Please choose a workflow 🕐 Help me choose	
 Search a reference library 	
Create a chromatogram library and search against it	
Save DIA Files in Choose	
Experimental Data Search Parameters	Click to expand \downarrow
Load Workflow From File Save Workflow As	Some required fields are missing

Fig.3-5 ワークフローダイアログ 初期表示画面

ワークフローダイアログでは検索実行に必要な各種パラメータを設定します。パラメータの多くは デフォルト値が定められていますが一部パラメータはユーザーが指定する必要があります。ユーザーが 指定しなければならないパラメータが残っている項目については、▲で警告が表示されています。 ワークフローダイアログは以下3つのタブから構成されています。

- Search タブ (3-4-1,P.17 ~) 検索のタイプ、検索対象のライブラリや配列ファイル、許容可能な誤差範囲などを指定します。
- Analysis タブ (3-4-2, P.24 ~) 検索後、どのような条件のタンパク質を結果に表示するかに関するパラメータを指定します。
- Advanced タブ (3-4-3, P.25 ~)
 定量計算の条件や一時ファイルの置き場所などを指定します。

使用初期でよくわからない場合、とりあえずは Search タブで▲が表示され入力を促されている項目の み記入してください。また各タブのフッター部分に共通して表示される「Load Workflow From File ...」 と「Save Workflow As ...」は、このワークフローダイアログで定義する内容について、予めファイルと して保存した内容を呼び出したり、逆に定義した内容をファイルとして保存したりする際に利用します。

以降、ワークフローダイアログの各タブの内容について、詳細を説明します。

3-4-1. ワークフローダイアログの Search タブ

Search タブでは検索対象のライブラリ(FASTA,BLIB,DLIB,ELIB など)、配列ファイル、許容可能な 誤差範囲などを指定します(Fig.3-3)。

ダイアログ上部には2つの選択肢があります。

- Search a reference library
- Create a chromatogram library and search against it

「Search a reference library」は、FASTA あるいは BLIB、DLIB,ELIB(この中のいずれか 1 種類)に 対する検索を行います。開発元が特に推奨している手軽さと同定力を兼ね備えたライブラリが Prosit に よって作成されたライブラリ(DLIB)です。Prosit は Wilhelm と Kuster のグループによって開発された、 ペプチドのフラグメントと保持時間を予測する深層学習アルゴリズムです。Proteome Software 社では 多くの生物の Prosit ライブラリを提供しており、Library Manager からファイルをダウンロードできます し、以下 WEB サイトからダウンロードする事も可能です。作成に関する詳しい内容は、以下 URL または 本資料の 14 章(P.93~)をご参照ください。

https://support.proteomesoftware.com/hc/en-us/ articles/360035151172-Prosit-Derived-Spectral-Libraries-for-Scaffold-DIA-Searches もう 1 つの選択肢、「Create a chromatogram library and search against it」は最初に FASTA あるいは BLIB, DLIB に対する検索を行い、そこで得られた情報を元に作成された DIA の解析結果を ライブラリ化した ELIB に対して別のデータを検索する、2 段階検索です (P.8 の Fig2-5 の検索方法に 対応しています)。1 回目と 2 回目の検索させるデータとの間で、保持時間に関する関連性がある事が 求められます。

2 種類の検索のどちらかを選ぶかに伴って指定が必要なパラメータが連動してユーザーに提示 されます。前出の Fig3-5 は「Search a reference library」を選んだ場合、以下 Fig.3-6 が 「Create a chromatogram library and search against it」を選んだ場合です。

以降、2つの選択肢を選んだ場合それぞれにおけるパラメータ入力画面について説明いたします。

Load Data	×
▲ Search ✓ Analysis ✓ Advanced	
Please choose a workflow 😮 Help me choose	
○ Search a reference library	
 Create a chromatogram library and search against it 	
▲ Save Chromatogram Library As ▲ Choose Please specify a save location	
Save DIA Files in Choose	
Chromatogram Library Creation Parameters	Click to expand \downarrow
Experimental Data Search Parameters	Click to expand \downarrow
Load Workflow From File Save Workflow As	Some required fields are missing

Fig.3-6 ワークフローダイアログ Search タブ初期表示画面

■ Search a reference library を選択した場合

ダイアログ内に「Experimental Data Search Parameters」という領域が現れます。その行を クリックするとパラメータ入力画面が現れます(次頁 Fig.3-7)。各パラメータを入力後、画面下の 「Load Data」ボタンを押すことで解析が開始されます。 この領域について、画面を3つのパーツに分けて説明いたします。

🛃 Load Data			×		
▲ Search ✓ Analysis ✓ Advanced					
Please choose a workflow 😢 Help me choose	Please choose a workflow 😗 Help me choose				
 Search a reference library 					
O Create a chromatogram library and search against it					
Save DIA Files in Choose					
Experimental Data Search Parameters			Click to collapse ↑		
Experimental Data Search Parameters		A Experimental Data Files	Add Remove		
🔺 Reference Library 🚺 A Choose 🛛 Ple	ase select a reference library				
🔺 Protein Sequence Database 🚺 Choose Ple	ase select a fasta file	2			
iRT Database File: Choose		ى ك			
✓ Fragmentation HCD ✓ Higher-	energy collision dissociation				
✓ Precursor Tolerance 10 🗘 ppm	~				
✓ Fragment Tolerance 10 🗘 ppm	~				
✓ Library Fragment Tolerance 10 <i> ppm</i>	~				
✓ Digestion Enzyme Trypsin ✓	1				
✓ Peptide Length 6 - 30	amino acids				
✓ Peptide Charge 2 - 3					
✓ Max Missed Cleavages 1 😳					
Modifications					
	Add Remove	No English and I D	to Films I and ad		
Name Mass Neutral L AA	None Fixed	No Experimental Da	ata Files Loaded		
Pentide EDR Threshold 0.01					
- Data Acquisition Type					
▲ Please choose an acquisition type 💡 Help me ch	oose				
Non-overlapping Windows	2				
Overlapping Margins	Da				
Staggered Windows					
Precursor Window Size Determine from raw file					
Precursor window of Da					
Some required fields are missing					
Load Workflow From File Save Workflow	v As		Load Data Cancel		

Fig.3-7 ワークフローダイアログ Search タブ

Experimental Data Search Parameters					
🔺 Reference Library	🔺 Cho	ose.	P	lease se	lect a reference library
🔺 Protein Sequence Database	🔺 Cho	ose.	P	lease se	lect a fasta file
iRT Database File:	Choose				
 Fragmentation 	HCD	\sim	Higher	-energy	collision dissociation
 Precursor Tolerance 	10	$\hat{}$	ppm	~	
 Fragment Tolerance 	10	$\hat{}$	ppm	~	
✓ Library Fragment Tolerance	10	\Diamond	ppm	\sim	1
 Digestion Enzyme 	Trypsin		~	·]	
🗸 Peptide Length	6	- 3	30	amino	acids
🗸 Peptide Charge	2	- 3	3]	
 Max Missed Cleavages 	1 <	2			

Fig.3-8 ワークフローダイアログ Search タブの 1 部分

• Reference Library **& Protein Sequence Database**

「Reference」検索対象となる Library を選択します。ライブラリマネージャーから選ぶ方法と、 DLIB や BLIB などのファイルを直接選択する方法があり、後者の場合はピークリストリファレンスに 対応する FASTA ファイルを下の「Protein Sequence Database」で選択する必要があります。

·iRT Database File [必須ではない]

BLIB を検索対象とし、かつ Skyline から該当ファイルを出力している場合、Scaffold DIA の計算に そのファイルを保持時間情報として利用できます。" Choose" でファイルを選択します。

Fragmentation

CID / HCD / ETD から選択します。ペプチドの同定並びに定量の際に対象とするイオンシリーズが 変わります。CID,HCD は b,y イオン、ETD は c と z イオンを同定や定量時に利用します(CID は特定の ケースにおいて同定時 y イオンしか考慮しない事がありますが、その時も定量には b イオンを使用します)。

• Precursor / Fragment / Library Fragment : Tolerance

測定値と、理論値またはライブラリ値 との許容誤差範囲です。単位として Da と ppm どちらでも 選択可能です。

•Digestion Enzyme

実験で用いたペプチド切断方法(酵素)です。None Specific の場合任意の場所での切断パターンを考慮 しますが、その場合は missed cleavage の設定を、(考慮したい最大ペプチド残基長-1) の数値を設定する ようにしてください。

•Peptide Length

検索対象とするペプチドの長さです。

·Peptide Charge

検索対象とするペプチドの電荷です。

•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です。 数が 多いと理論ペプチド数が増えます。

Modifications								
			[Add···	Remove			
Name	Mass	Neutral L	AA	Terminus				
Carbamid	57.021464	0	С	None	Fixed			
✓ Peptid -Data Acquisit	✓ Peptide FDR Threshold 0.01 Data Acquisition Type							
O Non-over	lapping Wind	lows	0					
⊖ Overlappi	ng Margins			Da				
⊖ Staggered	◯ Staggered Windows							
Precursor Win O Determine	Precursor Window Size 2							
U Precursor	window of	Va						

Fig.3-9 ワークフローダイアログ Search タブ の 2 部分

Modifications

考慮する修飾です。FASTAを対象とする検索の場合、このリストで指定した修飾を考慮した理論値を 作成して検索します。一方 BLIB や ELIB など、ピークリストへの検索の際は、ライブラリに含まれる修飾に 関する情報を極力拾い上げてこの欄に修飾情報を表示します。

•Peptide FDR Threshold

Percolator 適用によって計算されるペプチド FDR の閾値です。

·Data Acquisition Type

DIA の測定での Precursor Isolation Window の幅に関する設定です(Fig.3-9)。

- Non-overlapping Windows:両端のオーバーラップ領域がない
- Overlapping Margins:オーバーラップ領域がある場合、ウィンドウ幅と合わせてマージン幅を 指定
- Staggered Windows:一度 MS/MS の測定をした後、ウィンドウ幅を半分ずらして再度 MS/MS の測定を行う手法。



Precursor Window Size

ウィンドウ幅を、ファイル情報から取得し設定するかユーザーが特定するか選びます。通常は "Determine from raw file"を選択してください。

	Click to collapse 1	
	▲ Experimental Data Files <u>A</u> dd <u>R</u> emove	
	3	
Fig.3-1	1 ワークフローダイアログ(Search タブ)	 の 3 部分

Experimental Data Files

検索対象となる raw ファイルを選択します。

■ Create a reference library and search against it を選択した場合

ダイアログ内に「Chromatogram Library Creation Parameters」と「Experimental Data Search Parameters」の2つのセクションが現れます。最初の検索で ELIB を作成し、その ELIB に対して 本命のサンプルデータを検索します。「Chromatogram Library Creation Parameters」については 前述の通常検索とパラメータの内容が同じです。後段にあたる、「Experimental Data Search Parameters」の「Experimental Data Files」で、検索対象とする本命のサンプルデータを指定して ください(Fig.3-12)。

Load Data		>
▲ Search ✓ Analysis ✓ Advanced		
Please choose a workflow 🕜 <u>Help me choose</u>		
Search a reference library		
 Create a chromatogram library and search against it 		
A Save Chromatogram Library As A Choose Please specify a sa	ve location	
Save DIA Files in Choose		
Chromatogram Library Creation Parameters		Click to expand \downarrow
Experimental Data Search Parameters		Click to collapse ↑
Experimental Data Search Parameters		
Using the following from 'Chromatogram Library Creation Parameters':	A Experimental Data Files	<u>A</u> dd <u>R</u> emove
Protein Sequence Database: Arabidopsis thaliana_TAIR10.fasta		
Digestion Enzyme: Trypsin		
Fragmentation: HCD		
Precursor Tolerance: 10.0 ppm	No Experimental Data	Files Londod
Fragment Tolerance: 10.0 ppm	No Experimental Data	Files Loaded
Peptide FDR Threshold: 0.01		
Peptide FDR Threshold: 0.01 Data Acquisition Type:		

Fig.3-12 Search タブ Experimental Data Search Parameters 画面

3-4-2. ワークフローダイアログの Analysis タブ

Analysis タブでは、ペプチドに関する検索後、その結果から同定タンパク質に選定しグループ化する際に利用されるパラメータです。

🛃 Load Data	×
▲ Search 🗹 Analysis 🗸 Advanced	
Shared Evidence Clustering	
Moderate shared evidence clusters	
Any shared evidence clusters	
 ✓ Target Protein FDR 1.0% FDR ↓ ✓ Minimum Number of Peptides 2 ÷ 	
	🔺 Some required fields are missing
Load Workflow From File*** Save Workflow As***	Load Data Cancel

Fig.3-13 ワークフローダイアログ Analysis タブ

Shared Evidence Clustering

シェアペプチドを持つ類似タンパク質のグループ化についての設定です。以下3段階があります。

- Perfect shared evidence protein groups
- Moderate shared evidence clusters
- Any shared evidence clusters

上に行くほどクラスター化させるためにペプチドセットの組み合わせ条件が厳しくなり、逆に下は緩く なります。一番上の「Perfect〜」なら、完全にペプチドセットが同じタンパク質でないとグループ化されませ ん。逆に一番下の「any〜」では1つでもシェアペプチドがあればグループ化されます。

詳細は本資料 12 章「タンパク質のグループ化、クラスター化」(P.89~)をご覧ください。

•Target Protein FDR

表示させるタンパク質について FDR の閾値を設定します。ペプチドの FDR 設定とは異なる、タンパク質の FDR であることにご注意ください。

•Minimum Number of Peptides

表示させるタンパク質の条件として、アサインされているユニークペプチドの最低数を設定します。

3-4-3. ワークフローダイアログの Advanced タブ

Advanced タブでは、データ処理に関する基本的な部分に関する設定を行います(Fig.3-14)。

🕌 Load Data	\times
▲ Search 🗸 Analysis 🖌 Advanced	
Processing Directory	
• Create a subfolder for intermediate files and remove it upon completion	
○ Create a subfolder for intermediate files and retain it	
○ Write intermediate files directly to this location	
 ✓ Minimum Number Of Quant Ions 3 ✓ Maximum Number Of Quant Ions 5 	
✓ Percolator Training Set Size 500000	
 Percolator Training Set Threshold 0.001 	
✓ Filter RT 🔲 – min	
▲ Some required fields are mis:	sing
Load Workflow From File*** Save Workflow As*** Load Data Cance	1

Fig.3-14 ワークフローダイアログ Advanced タブ

Processing Directory

Scaffold DIA がデータ処理を行う際に利用する一時的なファイルの保管場所です。フォルダの選択と、 ラジオボタンで計算に利用した一時ファイルの扱いを決定します。デフォルトでは計算の度にサブフォルダ を作成してその下に一時的なファイルを作成し、計算後にそのファイルを消すように設定しています。 ファイルを消さないようにしたり、サブフォルダでなく作成フォルダに直接ファイルを書き込むように設定 したりする事ができます。

·Minimum/Maximum Number Of Quant Ions

DIA の定量計算に使うフラグメントイオンの数について上限と下限を設定しています。Scaffold DIA の 定量計算はフラグメントピークのみを対象とします。共溶出の影響による精度の低下を避ける目的で、 すべてのフラグメントを使うわけでなく設定した上限個数までしか使用しません。Scaffold DIA では デフォルトで使用するフラグメントピーク数を3以上5以下として設定しています。

·Percolator Training Set Size, Percolator Training Set Threshold

Percolator 計算に関するパラメータです。通常はデフォルト値を使用してください。

·Filter RT

RT のフィルターです。明示的に解析に不要な領域がある場合に指定してください。なお、データ取り込みの最初の段階で raw ファイルから .dia ファイルを作成する際にのみ機能するため、input データが.dia ファイル(raw や mzML から Scaffold DIA 用に変換したファイル)の場合は機能しません。

4. Main 画面

4-1. Scaffod DIA メイン画面について

Scaffold DIA は、以下 6 つのパネル(画面左側)から構成されています。

- **Organize** (5章)
- **Samples** (6章)
- **Proteins** (7章)
- Visualize (8章)
- **Analysis** (9章)
- **Publish** (10章)

以降に続く5章~10章ではそれぞれのパネルについて説明しています。

この章では、どのパネルを選択していても表示される、画面の周辺にあるタイトルバーやメニューバーな どについて詳しく説明しています(Fig.4-1)。



Fig.4-1 Scaffold DIA メイン画面

メニューバーの各項目(Fig.4-2) について説明します。

File Edit View Experiment Export Tools Help

Fig.4-2 Scaffold DIA メニューの内容

メニュー	各項目の内容
File New Ctrl-N Open Ctrl-O Close Save Ctrl-S Save As New Chromatogram Library Open Library Manager Open Library Manager Print Ctrl-P Print Preview Exit	 New - 新たな解析ファイルを作成します。詳細は 3 章「データの取り込み」を ご覧ください。 Open - Scaffold のファイル「*.sdia」を開きます。 Close - 現在開いているファイルを閉じます。 Save - 現在開いているファイルを別名で保存します。 Save As - 現在開いているファイルを別名で保存します。 New Chromatogram Library - 検索時に使用する新たな Chromatogram ライブラリ(ELIB ファイル)を作成するダイアログを開きます。 Open Library Manager - Library Manager ダイアログを開き、現在登録 されているライブラリの調整ができます。 Convert Raw Files to DIA Files - raw ファイルを dia ファイルに変換する 処理を行うためのダイアログを開きます。 Print - 現在開いている画面を印刷します。 Print Preview - 現在開いている画面を印刷します。 Exit - ソフトウェアを終了します。
Edit Copy Ctrl-C Find Ctrl-F Edit GO term options Preferences	 Copy - 現在開いている画面の主な情報をまとめてクリップボードにコピー します。データはタブ区切りです。他ソフトウェアにテキスト情報として 貼り付けることができます。 Find - 現在開いている画面内でキーワード検索をするダイアログを表示 します。 Edit Go Terms Options - GO 情報について表示項目を選択したり、参照の ファイルを設定したりします。 Preferences - Scaffold DIA を使用する際に必要な各種基本設定 (メモリ、プロセッサー数、proxy、msconvert プログラムの指定など)を 確認・変更します。

	Color Options – 定量値の表示数字の大小を表す色について、設定を変更
View	するダイアログを表示します。
View Experiment Export To	・ Advanced Filter -メイン画面にあるフィルター画面では指定しきれない
Color Options	細かい条件を指定したフィルターリングを実行します。
Advanced Filter	 Show/Hide Columns - 各画面で表示されている列の内容を調整します
Show/Hide Columns	• Navigate - 表示されているパネル画面内のタブを切り替えます(タブを直接
Navigate >	クリックしても切り替わります)
Summarization >	Summarization - メイン画面内の[Summarization]と同じく、データを
Display Type >	
Cluster value Suppression 7	
Log Intensities	Display Type - ハーン回面P307 Display type」と同じて、私示する数子の
Organize View Ctrl-1	
Samples View Ctrl-2	Cluster Value Suppression - 知似タノハク質をまどのるクラスター化を
Proteins View Ctrl-3	した際、一覧表にクラスター並びに各構成タンパク質の定量値をどのように
Visualize View Ctrl-4	表示するか設定します。
Analysis View Ctrl-5	 Log Intensities – 定量値表示を底が 10 の Log 表示に切り替えたり元に戻
Publish View Ctrl-6	したりします。
	 Organize View ~ Publish View – メイン画面の左側のパネルと同じく、
	各パネルへ切り替え・移動できます。
Evenoviment	Apply Thresholding -各ペプチドについて同定の信頼度を求めるパラメー
Experiment	タです。設定値の Percolator q-value 閾値を上回らないペプチドは定量計算
Experiment Export Tools Hel	に使用されません。デフォルトは No bound で閾値設定はありません。
Apply Thresholding >	Protein Clustering - 類似タンパク質をグループ化してまとめる
Protein Clustering >	「Clustering」について、まとめる基準の設定をします(→12 章、P.89~)。
Apply GO Terms >	・ Apply GO Terms – (Samples 画面) 予め設定した GO データベース
🝺 Import Attributes File	ファイルに照合しタンパク質に GO 情報を付与します(→15 章、P.108~)。
Quantitative Analysis	Import Attributes Files - 各サンプルデータ間の階層・属性を定義した
Clear User Peptide Validation	「Attribute file」を読み込みます。
	Ouantitative Analysis - 定量値を使って検定を行ったり、検定の有意基準
	についての設定を行ったりします。
	Clear User Peptide Validation - 定量計算を行う際に行った各種設定を
	リセットして、デフォルト値に戻します。
Export	
	Attribute file を設定ファイルとして出力します。
Export Attributes File	・ Fynort Samples Peport to Fynol ~ Fynort Workflow - 解析データ
Export Samples Report to Excel	
 Export Peptide Report to Excel Export Peptide Quant Report to Excel 	
 Export Peptide Match Report to Excel Export Publish Report to Excel 	
Export Workflow	・ Kun SQL Query for Export - SQL コマノトを入力美行するダイアロクか
Run SQL Query for Export Ctrl+Shif	現れます。

Tools

 Tools
 Help

 Windowing Scheme Wizard...

 DIA Structure Viewer (mzML only)...

 Launch Peptide Browser...

 Convert BLIB to ELIB...

 Convert MSP/SPTXT to ELIB...

 Convert Prosit/Spectronaut CSV to DLIB...

 Convert TraML to ELIB...

 Convert ELIB to BLIB...

 Convert ELIB to OpenSWATH TSV...

 Combine ELIB libraries...

 Subset ELIB library...

•

* ELIB と DLIB は同じ フォーマットで、通常の解析において は実質中身が同じものと考えて問題 ありません。Scaffold DIA 上では ELIB 用/DLIB 用 オプションは どちらのフォーマットも利用可能で す。(例:DLIB の結合に「combine ELIB libraries」を使用する、など)

ELIB と DLIB について

https://bitbucket.org/searleb/encyclopedia/wiki/

EncyclopeDIA%20File%20Formats

Windowing Scheme Wizard – Windowing Scheme Wizard (EncyclopeDIA の機能)を起動します。質量分析装置の DIA 測定に 際して装置の設定に必要なウィンドウ幅の設定数値を計算します。

- **DIA Structure Viewer (mzML only)** DIA Structure Viewer (EncyclopeDIA の機能)を起動します。mzML を開いて windowing scheme に関する設定を確認する事ができます。Data Acquisition type の確認などに利用できます。Scaffold DIA 上では mzML への 変換ツールを利用できないので、手元の raw ファイルを見たい時には 別途 msconvert を起動し変換を実施した上でツールに読み込ませる 必要があります。
- Launch Peptide Browser Peptide/DIA Detection Viewer (EncyclopeDIA の機能)を起動します。mzML または dia ファイル から、特定ペプチドを検索します。実行後、ダイアログ内の上のパネル にはフラグメントイオンのピーク強度が、下のパネルにはスコアの情報 が表示されます(どちらも横軸は RT)。
- Convert BLIB to ELIB- BLIB ファイルを ELIB ファイルに変換します。
- Convert MSP/SPTXT to ELIB MSPや SPTXT を ELIB に変換 します。
 - Convert Prosit/Spectronaut CSV to DLIB Prosit のライブラ リ(Prosit から出力されたファイル)、または Spectronaut のライブラリ である CSV ファイルを DLIB に変換します。出力ファイルは入力ファイ ルと同じ場所に作成されます。
- Convert TraML to ELIB TraML ファイルを ELIB に変換します。 Convert ELIB to BLIB – ELIB ファイルを BLIB に変換します。
- Convert ELIB to OpenSWATH TSV ELIB ファイルを OPENSWATH TSV フォーマットに変換します。出力ファイルは 入力ファイルと同じ場所に作成されます。
- Convert FASTA to Prosit CSV FASTA ファイルから、Prosit 検索の入力データ(CSV 形式)を作成します。作成されたファイルは Prosit サイトにて入力データとして使用でき、ライブラリを作成する事 ができます。この機能を利用した場合は 以下論文への Citation をつ けてください。

https://www.nature.com/articles/s41467-020-15346-1

- Combine ELIB libraries 複数の ELIB ファイルを結合します。
- Subset ELIB library ELIB を開き、特定条件に合う結果のみを 抜き出した新たな ELIB ファイルに出力します。



4-3. ツールバー(アイコン)

画面上部に、頻繁に使うコマンドに対応するアイコンが準備されています(下図)。



アイコンが表すコマンドは以下の通りです。

アイコン	機能			
	New - 新たなファイルを作成します。詳細は3章「データの取り込み」をご覧ください。			
	Open - Scaffold DIA のファイル「*.sdia」を開きます。			
	Save - 現在開いているファイルを上書き保存します。			
ell.	≧ Open Library Manager – Library Manager を開きます。			
	Print - 現在開いている画面を印刷します。			
ß	Print preview - 現在開いている画面を印刷する前に、印刷後のイメージを表示します。			

	Copy - 現在開いている画面の主な情報をまとめてクリップボードにコピーします。他ソフトウェアに
	テキスト情報として貼り付けることができます。
25	Find - 現在開いている画面内で、キーワード検索をするダイアログを表示します。
lalla	Quantitative Analysis - 表示されている定量値を元に検定を行ったり、タンパク質の有意基準につい
	ての設定を行ったりします(詳しくは「5-7 Experimental Designの設定」(P.43~)をご覧ください)。
*	Excel - 現在開いている画面に適したレポートを EXCEL で読める CSV 形式で出力します。
	Help - オンラインヘルプを開きます

4-4. 表示タンパク質のフィルターリング

表示タンパク質について、各種条件によるフィルターリングを行うことができます。

🗎 🗖 🗟 🗎	• •		Summarization:	Treatment	~	hit Pr	rotein l	DR	1.0% FDR	~	Min # Peptides	2 🗘	*	0
Filters														
🗹 Show Hidden		*	Name/Accession	٩	p-value filter 🔻	GO Te	rm 🔻	⊙	×					

Fig.4-4 Scaffold DIA 表示タンパク質のフィルターリング設定項目

アイコン	機能
	Show Hidden - 各種絞り込み条件に合わないタンパク質を表示する / しない を切り
	替えます。
	Star Filter – Samples 画面内にて設定された項目「Star」の内容でフィルターリングを
	行います。Star には、/星なし/黄色/青色/黄+青色/ の4種類があります。
Name/Accession	Name/Accession - キーワード検索により Accession やタンパク質の名称に対して
	絞り込みを行うことができます。
D. Value Filter	P-value filter - P value が、検定の p 値、またはさらに別途定められた Family-Wise
	Error Rate(または FDR)の閾値を超えているかどうかでフィルターリングを行います。
	検定を実施した後のみで選択できます。
GO Term 🕶	GO term - アサインされた GO 情報によるフィルターリングを行います。
0	Advanced filter - メイン画面にあるフィルター画面では指定しきれないような細かい
	条件を指定したフィルターリングを実行します。
Protein FDR 1.0% FDR	Protein FDR – Protein FDR の値によるフィルターリングを行います。
Min # Peptides 2	Min # Peptides - アサインされたペプチド数によるフィルターリングを行います。

4-5. 各画面への切り替え

画面左の「Navigation pane」(Fig.4-5) にて、Scaffold DIA の解析画面を 切り替えることができます。

- **Organize** (5章)
- **Samples** (6章)
- **Proteins** (7章)
- Visualize (8章)
- **Analysis** (9章)
- **Publish** (10章)

以降に続く5章~10章ではそれぞれのパネルについて 説明しています。

Fig.4-5 Scaffold DIA Navigation pane

4-6. FDR 計算値表示

Navigation pane の下に、FDR の情報がまとめられた表示欄があります(Fig.4-6)。ペプチド、タンパク質 それぞれの FDR 設定値並びに閾値を超える Target / Decoy エントリーの数が表示されます。

このダッシュボード内の情報は、エディター操作のようにドラッグ&ドラッグを行ってコピーする事ができます。

Fig.4-6 Scaffold DIA FDR Info ボックス





4-7. Summarization

Summarization バー(Fig.4-7)では、着目するサンプルの階層構造を切り替えることができます。結果 画面に表示されるサンプル、並びにその定量値も連動して変更されます。

Summarization:	Condition 🗸
	Condition
	MS Sample
Name/Accessio	Edit···

Fig.4-7 Scaffold DIA Summarization //-

項目の一番下、「Edit」を選ぶと、「Configure Sample Organization and Statistical Analysis」ダイア ログの「Sample Hierarchy」が表示されます。このダイアログについては「5-6.選択可能な3種類の Experimental Design について(P.41~)」にて詳しく説明しています。

5. Organize 画面

5-1. Organize パネルについて

Scaffold DIA ではデータの階層につける名称を"Category"と呼び、Category 内の各変数(属性)の名称 を"Attribute"と呼びます。Organize パネルではデータの Category(階層)や Attribute(属性)の定義や、 各 Sample データに対して Attribute との紐づけを行う事ができます。

例えば、"Treatment Group "という Category を定義し、そこに "Control "と "Treated "という Attribute を配置する事ができます。あるいはタイムコース、薬物の投与に対する反応を測定する研究で は、"Time "と呼ばれる Category を作成して、そこに属する Attribuite として「Omin」「20min」 「40min」「60min」などを設定する事もできます。臨床サンプルには多くの場合、年齢、性別、病歴などの 多くの属性が付与されていますが、Scaffold DIA 上においても1つのデータに複数の Category の Attribute を割り当てる事が可能です。Category と Attribute は、Organize パネルのグラフィカル ユーザーインターフェースで定義する事ができるほか、予め設定が完了した情報を内包するファイルを 読み込んで利用する事ができます。

最初に Scaffold DIA にロードされたばかりの段階ではこれらの定義が全くされておらず、取り込んだ データ(raw 又は mzML)の単純なリストが存在するのみです。Fig.5-1 は、サンプルがロードされた直後の Organize パネル 画面です。

Define Categories		Group Samples By Sel	lected Category	Bulk Edit Sample N	Names		P
🖮 Import Attributes File	X Export Attributes File			male Name		属性情報なし	Æ
+ Add Category	Delete Selected Category	 SD_1 from F001466 SD_1 from F001467 SD_1 from F001472 SD_1 from F001473 	,				

Fig.5-1 データと階層構造の例

以降、Organize パネルの画面内で Category や Attribute の定義を作成し各データと紐づけをする 作業についてご案内します。

5-2. Organize パネルの画面と設定箇所

Organize パネルの概要です(次頁図)。以下の4つのポイントに分けてご説明します。

- ① Category, Attribute の定義設定 [5-3]
- ② Sample ファイルと Attribute の関係性を設定・確認する表 [5-4]
- **③ Import Attribute File / Export Attribute File [5-5]**
- ④ Experimental Design の定義 [5-6,5-7]

Scaffold DIA - Dem <u>F</u> ile <u>E</u> dit <u>V</u> iew E	102_HeLa_insulin_6_files_30_proteins.sc xperiment E <u>x</u> port Tools <u>H</u> elp	ia				- 0	×
■ ■ ■ ■ Filters ✓ Show Hidden	Summarization:	 Treatment If Attribute File / Expo 	Protein FDR 1.0%	FDR V Min # Peptides	2 🗘 📓 🛛		
Organize	Define Categories	Export Attributes File	Group Samples E	By Selected Category V	Bulk Edit Sample Names		٩
	Add Category	Delete Treatment	Control	Sample Name	Treatment		E.
Samples Proteins	Treatment 📃	Control Insulin Add Attribute	Control Control Control Insulin Insulin	_1 _2 _3 _1	 Control Control Control Insulin 		
Visualize			 Insulin_ Insulin_ 	2 3	InsulinInsulin		
Analysis	Category, Attribute	の定義設定					
Publish		rimeratel Designe の完美		② Sample ファ-	イルとAttribute の エ		
Proteins 0.0% FDR (attained) 29 Targets 0 Decoys	(4) Expe			実 氷 土を唯認9る	衣		
Peptides 0.0% FDR (attained) 988 Targets 0 Decoys	and Statist	Experimental Design ical Analysis	Sample Informa No Sample Sel	ected			

Fig.5-2 Organize 画面内の各機能に関する説明

5-3. Category, Attribute の定義設定

画面左側、「**Define Categories**」では category と、それに属する Attribute の設定を行う事が できます(Fig.5-3)。例として、①のエリアで「**Treatment**」というカテゴリーを作成し、そこに「**Control**」と 「**Insulin**」という2つの Attribute を作成する操作をご案内します。

🖳 Import Attributes File		1	Export	t Attribute	es File…	
🕂 Add Category	1	Delete T	Treatment			
Treatment		ntrol				
	Ins	ulin				
	+	Add Attr	ibute			
						A-Z

Fig.5-3 Organize 画面 View Table 拡大

最初に、"Add Category"ボタンを押します(Fig.5-4)。



Fig.5-4 Organize 画面 Add Category ボタンの位置

Category 名として「**Treatment**」と入力します。続いて、右側にある「**Add Attribute**」をクリックします (Fig.5-5)。

Define Categories	
🖳 Import Attributes File…	🕙 Export Attributes File…
🕂 Add Category	📸 Delete New Category
Treatment	+ Add Attribute

Fig.5-5 Organize 画面 Add Attribute ボタンの位置

Attribute 名として「**Insulin**」と入力します(Fig.5-6)。

Define Categories	
😂 Import Attributes File…	🐔 Export Attributes File…
🕂 Add Category	💕 Delete Treatment
Treatment	Insulin

Fig.5-6 Organize 画面 Attribute 設定
必要に応じて色のアイコンをクリックし、色の設定を変更します(Fig.5-7)。



Fig.5-7 Organize 画面 Attribute の色設定

続いて、作成した Attribute の下に新たに現れる「+Add Attribute」をクリックします(Fig.5-8)。

-		
	Add Attr	ibute
	T	A-Z
		Add Attr

Fig.5-8 Organize 画面 Attribute 追加

同様の操作で、もう1つの Attribute である「Control」を作成します(Fig.5-9)。

Treatment	Control
	Insulin + Add Attribute

Fig.5-9 Organize 画面 Attribute 追加完了

以上で作成の操作は終了です。

もし作成後の Category を削除する時は、画面内のボタン「Delete (カテゴリー名)」をクリックします。 作成後の Attribute を削除する時は、Attribute を選択してから右クリックして「delete」を選択します。

5-4. Sample ファイルと Attribute の関係性を設定・確認する表

Fig.5-2 画面、右側の②の領域では、取り込んだ データ の各 Sample と定義した Attribute との 紐づけの定義をしたりその内容を確認したりする事ができます(Fig.5-10)。

Group Samples By Selected Categ	ory 🗸 🛛 Bulk Edit Sample Names… 🕜 🖉 🔎
Sample Name	Treatment
🖃 🚾 Control	· · · · · · · · · · · · · · · · · · ·
Oontrol_1	😑 Control
Oontrol_2	😑 Control
🖳 😑 Control_3	Control
🗄 🔚 Insulin	
- O Insulin_1	🔵 Insulin
- 🦲 Insulin_2	😑 Insulin
🔲 🌑 Insulin 3	Insulin

Fig.5-10 Organize 画面 Sample-Attribute 対応表

以下、データと属性の紐づけ操作に関するご案内です。

[sample と attribute を紐づけする操作について]

Sample を選択した状態で右クリックすると、表示リストの中に Category 名が現れます。Category 名に カーソルを合わせると、さらに Attribute が表示されます。Attribute を選択する事で、該当 Sample に 選択した Attribute 情報が紐づけされます。

なお Sample をドラッグ&ドロップや Ctrl キーを使って複数選択し、同時に Attribute 紐づけ設定する事 もできます(Fig.5-11)。

Group Samples By	Selected Category 🗸 🗸	Bulk Edit Sar	nple Names…	0				
Sample Name						Trea	tment	t
Control Insulin Control_1								
Control_2 Control_3 Insulin_1					Edit sample Treatment	>		No Treatment
● Insulin_3					Expand All Collapse All Edit Selected Sample Names		•	Insulin Control Ne Control er

Fig.5-11 Organize 画面 Sample に対して Attribute 情報を付与している操作

Attribute 情報を付与後、その情報が表内に表示されます(Fig.5-12)。

Group Samples By Selected Category 🧹	Bulk Edit Sample Names… 🕜	۶
Sample Name	Treatment	
₽- <mark></mark> Control		
Control_1	😑 Control	
 Ontrol_2 	😑 Control	
Control_3	🛑 Control	
Insulin		
- Insulin_1		
Insulin_2		
Insulin_3		

Fig.5-12 Organize 画面 Sample に Attribute 情報が付与された画面

[Group Samples By]の選択肢について

3つの選択肢、「Sample Name」「Selected Category」「Experimental Design」があります。

Group Samples By	Selected Category 🗸 Sample Name	Bulk	. Edit Sample Names…	0
Sample Name	Selected Category Experimental Design		Treatment	
- 🖶 Control_1			🛑 Control	
Oontrol_2			🔴 Control	
🖳 😑 Control_3			🔴 Control	
insulin				

Fig.5-13 Organize 画面 「Group Samples By」項目

表内のデータの並びについてどの項目に着目するかを選びます。

Sample Name: Sample 名のアルファベット順に並んだリストになります。 **Selected Category**: Category、Attribute 名の順に並んだリストになります。 **Experimental Design**: Experimental Design(詳細は **5-6.**をご参照ください)の階層構造と Category, Attribute 情報に基づきデータがまとめられ表示されます。

[Bulk Edit Sample Names] と選択肢について



取り込んだ Sample 名が長すぎる場合、Scaffold DIA の中でデータを扱いやすくするために、表示 される Sample 名を短く変更すると便利です。一つ一つを選択して表示される Sample 名を編集する事も できますが、データ数が多い時はその操作の実施も非常に煩雑です。そういった場合に**一括してファイル名** を変更できるのが、「Bulk Edit Sample Names」ダイアログです (Fig.5-14)。

ファイル名について、全ファイル名に共通する特定文字を一括して削除したり(Remove prefix, Remove suffix*)、ファイル名の前/後ろからの特定文字数を強制的に削除する事ができます(Remove characters at beginning, Remove characters at end)。また単純に文字を削除するだけでなく、選択した Attribute 情報に基づいてファイル名を変更するオプションもあります(Rename based on Attributes)。 * prefix -> 接頭語、suffix -> 接尾語

🐁 Edit Sample Names					×
O Remove prefix	20121130	_MD_			
O Remove suffix)				
0		•			
Remove characters at beginning	10	Ŧ			
O Remove characters at end	10	÷			
Rename based on Attributes	Extra	action Method 🔀		loast Time 🗸 🗸	+
◯ Custom					
Current Name		New Name			Ę
20121130_MD_10A (20121130_MD_1	0A.mgf-a	Complete Extractio	n-Di	gestion_10 Min_1	1
20121130_MD_10B (20121130_MD_1	0B.mgf-al	Complete Extraction	n-Di	gestion_10 Min_2	
20121130_MD_10C (20121130_MD_1	0C.mgf-a	Complete Extraction	n-Di	gestion_10 Min_3	
20121130_MD_10D (20121130_MD_1	0D.mgf-a	Complete Extraction	n-Di	gestion_10 Min_4	
20121130_MD_11A (20121130_MD_1	1A.mgf-a	Complete Extraction	n-Di	gestion_20 Min_1	
20121130_MD_11B (20121130_MD_1	1B.mgf-al	Complete Extraction	n-Di	gestion_20 Min_2	
20121130_MD_11C (20121130_MD_1	1C.mgf-a	Complete Extraction	n-Di	gestion_20 Min_3	
20121130_MD_11D (20121130_MD_1	1D.mgf-a	Complete Extractio	n-Di	gestion_20 Min_4	
20121130_MD_12A (20121130_MD_1	2A.mgf-a	Complete Extractio	n-Di	gestion_5 Min_5	
20121130_MD_12B (20121130_MD_1	2B.mgf-al	Complete Extractio	n-Di	gestion_5 Min_6	
20121130_MD_12C (20121130_MD_1	2C.mgf-a	Complete Extractio	n-Di	gestion_5 Min_7	
20121130_MD_12D (20121130_MD_1	2D.mgf-a	Complete Extractio	n-Di	gestion_5 Min_8	
20121130_MD_13A (20121130_MD_1	3A.mgf-a	Complete Extraction	n-Di	gestion_10 Min_5	
20121130_MD_13B (20121130_MD_1	3B.mgf-al	Complete Extraction	n-Di	gestion_10 Min_6	
20121130_MD_13C (20121130_MD_1	3C.mgf-a	Complete Extraction	n-Di	gestion_10 Min_7	
20121130_MD_13D (20121130_MD_1	3D.mgf-a	Complete Extractio	n-Di	gestion_10 Min_8	
20121130 MD 14A (20121130 MD 1	4A.mgf-a	Complete Extractio	n-Di	gestion 20 Min 5	
20121130 MD_14B (20121130 MD 1-	4B.mgf-al	Complete Extractio	n-Di	gestion_20 Min_6	
20121130 MD 14C (20121130 MD 1	4C.mgf-a	Complete Extractio	n-Di	gestion 20 Min 7	
20121130_MD_14D (20121130_MD_1	4D.mgf-a	Complete Extractio	n-Di	gestion_20 Min_8	,
			_		
🕡 Help		0	(Cancel	Apply

Fig.5-14 Organize 画面 「Bulk Edit Sample Names」ボタンとダイアログ

5-5. Import Attributes File / Export Attributes File

同じデータを取り込むもののパラメータを変えて再検索したい場合、現バージョンの Scaffold DIA で はデータの読み込みから再度実施する必要があります。その際、Sample と Attribute の紐づけを再度 設定するのは少々面倒です。

そのような時には Sample ファイルと Attribute 情報の紐づけ状況をファイルとして保存しておき、後で 再び読み込んで利用する事ができます。ファイルはカンマで区切られた CSV ファイルです(Fig.5-15)。

1	Sample Name Biosample Categor	y Ethnicity	Anticoa	gulant	HPLC
2	CAK-20-400-900.RAW (F002703)	PAe000817 La	ab-1 b1	edta	20-400-900
3	CAS-40-900-1200.RAW (F002698)	PAe000810 La	ab-1 b1	serum	40-900-1200
4	CAH-40-900-1200.RAW (F002760)	PAe000862 La	ab-1 b1	heparin	40-900-1200
5	CAH-20-900-1200.RAW (F002757)	PAe000862 La	ab-1 b1	heparin	20-900-1200
6	CAC-10-400-900.RAW (F002743)	PAe000859 La	ab-1 b1	citrate	10-400-900
7	CAC-40-900-1200.RAW (F002751)	PAe000859 La	ab-1 b1	citrate	40-900-1200
8	CAH-20-400-900.RAW (F002756)	PAe000862 La	ab-1 b1	heparin	20-400-900
9	CAS-20-1200-200.RAW (F002696)	PAe000810 La	ab-1 b1	serum	20-1200-200
10	CAK-40-400-900.RAW (F002706)	PAe000817 La	ab-1 b1	edta	40-400-900
11	AAS-10-400-900.RAW (F002734)	PAe000797 La	ab-1 b3	serum	10-400-900
12	CAS-40-400-900.RAW (F002697)	PAe000810 La	ab-1 b1	serum	40-400-900
13	CAC-40-1200-2000.RAW (F002752)	PAe000859 La	ab-1 b1	citrate	40-1200-2000
14	CAC 20 400 000 DAM (E002747)	D1-000950 T-	h_1 h1	aitrata	20-400-000

Fig.5-15 Attributes text file

設定した内容を CSV ファイルとして出力するには"Export Attributes File"を、いったん保存した CSV ファイルを読み込むには "Import Attributes File"ボタンを押します (Fig.5-16)。

Define Categories			
🔍 Import Attributes File…	街 Export Attributes File…		
🕂 Add Category	Delete Selected Category		
Configure Experimental Design and Statistical Analysis…			

Fig.5-16 Import Attribute Files / Export Attributes File

5-6. 選択可能な 3 種類の Experimental Design について

Scaffold DIA では実験計画法に基づいた定量解析を行う事ができます。データの構造について以下3つのオプションが選択可能です。

- Basic Design
- Repeated Measures
- Two-way Design

以降それぞれのオプションについて詳しく説明します。

Basic Design



Fig.5-17 experimental design 構成例

残り2つの選択肢が特殊な状況を想定しているのに対し、「Basic Design」はより一般的なデータ構造 を想定しており、Experimental Designの選択肢がよくわからない場合はこの「Basic Design」を選択し て下さい。

まず比較を行いたい一番大きな分類として、2つ以上の「Biological class」を設定します。「Biological class」は実験条件の違いを反映したグループから構成されます(例:Control と Treatment など)。

それぞれの「Biological class」に「Biological Replicates」を属させる事が可能です。これは同一実験で ありながらも生物的な差異があるケースなどが当てはまります。さらにそれぞれの「Biological Replicates」に対して、繰り返し実験である「Technical Replicates」を対応付ける事ができます。データ 構造の最下層として、各「Technical Replicates」が複数の画分から構成されている場合それを 「Fraction」として紐づけする事ができます。

Fig.5-17 の状態を Scaffold DIA の Category 情報を使って以下のように定義するとします。 Group -> { Treatment, Control }

Subject -> { Subject 1, Subject 2, Subject 3, Subject 4 }

すると Scaffold DIA の summarization 選択では以下のように階層構造が表示されます。

Group

Subject

(MS Sample)

結果画面の Summarization で Group を選んだ場合は Treatment と Control にまとめた結果が、 Subject を選んだ場合は Subject の単位でまとめられた結果が表示されます。(MS Sample)を選択した 場合はデータがまとめられず、各 MS データ単位(測定データの単位)で結果が表示されます。

Repeated Measures

同一の生物種個体からデータを取得していることを前提としており、各個人のレベルがどのように変化 するかを分析する際に選ぶ選択肢です。例として、**経時的変化を追った実験やクロスオーバー法**が挙げら れます。前述の項目「Basic Design」のように、Biological Replicates や technical replicates、fraction などの分類を定義する事もできます。

Two-way Design

2つの独立変数の効果を比較し、2つの変数の間に交互作用があるかどうかを判断する際に選ぶ 選択肢です。例えばある薬剤を投与した時の男性と女性の反応の違いについての比較です。Two-way ANOVA を実施すると、3つの尺度で評価する事ができます。3つの尺度とは、2つの要素が交互作用する 程度(Interaction)並びに Primary として指定した要素の影響度、並びに Secondary として指定した要素 の影響度です。前述の項目「Basic Design」のように、Biological Replicates や technical replicates、 fraction などの分類を定義する事もできます。

なお、特殊な状況での Two-way Design である Randomized Block Design (乱塊法)にも対応してい ます(乱塊法 : <u>Randomized Blocks in Analysis of Variance (ANOVA) - StatsDirect</u>)。サンプルを ブロックと呼ばれるグループに分割(ブロック化)し、処理グループの状態以外の既知の要因(年齢・性別等) と、測定データ側の各要素とが与える影響の度合いを確認します。Randomized Block ANOVA は ブロック化したカテゴリーの効果を最小化しながら治療効果 (treatment effect)を見るための解析方法 です。Two-way ANOVA と異なり、ブロッキング・カテゴリーの効果の評価は行いません。

5-7. Experimental Design の設定

以降、Scaffold DIA 上でデータから実験計画の内容を定義する操作方法についてご紹介します。 設定は「Configure Sample Organization and Statistical Analysis」ダイアログ(Fig.5-18)から行います。 ダイアログを表示させる方法は以下の4通りあります。

- 「Organize」Viewの画面左下にある「Configure Experimental Design and Statistical Analysis」ボタンをクリック
- menuの「Experiment」->「Quantitative analysis」を選択
- ツールバーにある「Quantitative analysis」アイコンをクリック
- ツールバーの「Summarization」選択肢内にある「Edit」をクリック

Sample Hierarchy Statist	tical Analysis	
After providing information about the sample acquisition and experimental design, drag and drop	Sample Acquisition Samples were fractionated Technical replicates were acquired	Î
Categories from the pool of Available Categories to build the summarization	Experimental Design	
hierarchy. If desired, switch to the Statistical Analysis	Summarization Hierarchy Available Categories @ Which Categories should be studied ? @	
tab to apply a statistical test.	Roast-Time Extraction Method Temperature Particular	
	Which Category identifies the biological samples? @ MS Sample	
	<< Clear Summarization	
Design Matrix		
	All Samples 2012/11/30 MD_108 (2012/11/30 MD_108.mgf-allergens-9860)	^
	20121130 MP - MB (20121130 MP - MB (2019) 20121130 MP - MB (20121130 MP - MB (2019) 20121130 MD 20 (20121130 MD 20 mg failergene-9828)	
-	20121130_MD_5A (20121130_MD_5A.mgf-allergens-9838) 20121130_MD_78 (20121130_MD_78.mgf-allergens-9848)	~
0		Apply Cancel

Fig.5-18 Configure Sample Organization and Statical Analysis ダイアログ

このダイアログは2つのタブから構成されています(前頁図赤枠)。

・「Sample Hierarchy」タブ:実験の種類の選択や各カテゴリーの位置づけの定義 [5-7-1]

・「Statistical Analysis」タブ:検定手法や多重検定の補正についての設定 [5-7-2]

以降、各タブでの設定内容について詳しく説明します。

5-7-1. Sample Hierarchy タブ

4つのパーツから構成されています(Fig.5-19)。

- Sample Acquisition
- Experimental Design
- Summarization Hierarchy

and the standard states of the		
Samples were fractionated		
Technical replicates were acquired		
xperimental Design		
Basic Design	Two-Wa	ay Design
ummarization Hierarchy		
		which Categories should be studied ?
Roast Time Extraction Method		(Optional)
Temperature		-
Replicate		
	<	Which Category identifies the biological samples ?
		MS Sample
		>

Fig.5-19「Simple Hierarchy タブ」の画面上部拡大

Sample Acquisition

Samples were fractionated

測定データに分画データが含まれている時にチェックを入れます。定性・定量において分画されたデータ は解析時に結合して1つのファイルであるかのように取り扱います。チェックを入れると、カテゴリーのどの 分類の属性情報が分画前の属性であるかを指定する設定欄が「Summarization Hierarchy」に現れま す。

Technical replicates were acquired

測定データについて、繰り返し実験(同一サンプル・同一条件のサンプルについて測定した)データを含む 時にチェックを入れます。定量において繰り返し実験の単位で標準化を行います。チェックを入れると カテゴリーのどの分類の属性情報を使って Technical Replicates データを認識するか指定する設定欄が 「Summarization Hierarchy」に現れます。

Experimental Design

Basic Design / Repeated Measures / Two-way Design の3択です。選択肢の詳細は以下の通りです。詳細については「**5-6.選択可能な3種類の Experimental Design について**」(P.41~)を ご覧ください。

Basic Design

下記2手法以外の場合に選択。データを階層構造化し様々な分類でデータを見る事が できます。

Repeated Measures

経時的な変化やクロスオーバー測定など、同一個体種からとったデータの比較をする

Two-way Design

2つの独立変数の効果を比較し、変数の間に交互作用があるかどうかを判断する

Summarization Hierarchy :

カテゴリーに対して分類・階層構造化を行う設定欄です。画面上部の「Sample Acquisition」並びに 「Experimental Design」で何を選んだかによって選択肢が変わります。基本的には、画面左側の 「Available Categories」に表示されるカテゴリーを画面右側の分類や階層構造を定める箇所に移動 (ドラッグ&ドロップまたは矢印を使用)させて設定します。

Experimental Design				
Basic Design Repeated Measures Design Two-Way Design				
Summarization Hierarchy				
Available Categories 💡	Which Categories should be studied? 💡			
(None)	> Condition > sex			
	Which Category identifies the biological samples? ② MS Sample			
<< Clear Summarization				

Fig.5-20 Summarization Hierarchy 設定 / Basic Design

Available Categories [すべての状況で現れる]

Organize 画面で設定したカテゴリーの一覧です。カテゴリーの設定やデータとの紐づけ方法については 5-3,5-4 をご覧ください。左側にあるカテゴリーを右側の適切な分類に当てはめていきます。

Categories to be studied [Experimental Design で Basic Design を選択した時]

統計解析を行う対象となるカテゴリーで、統計解析を行う場合は必ず何かしらのカテゴリーをこの分類に アサインする必要があります。ここで指定されるカテゴリーが 1 つである場合、そのカテゴリーの属性値毎に データがまとめられます。2 つ以上指定された場合(Fig.5-20)、各カテゴリーの属性についてすべての 組み合わせがデータまとめの対象となります。例えば「**Condition**」カテゴリーに属性値{Control,Treated} が、 「**sex**」カテゴリーに{male,female} が存在し Fig5-20 のように Condition と sex どちらも「Categories to be studied」に指定していた場合、データまとめの 対象は「Control,male」、「Control,female」、 「Treated,male」、「Treated,female」の4種類となります。

Biological Samples [すべての状況で現れる]

「Biological Replicates」にあたる項目を選択する事ができます。当てはまるものがない場合は特に気にせず 「MS Sample」が選ばれている状態であれば OK です。

Technical Replicates [Sample Acquisition で Technical replicates were acquired を選択した時]

Fig.5-17 データの説明における「Technical Replicates」にあたる項目を選択する事ができます。該当する ものがなければ「Technical replicates were acquired」のチェックを外してください。該当するものがある ものの特に当てはまるカテゴリーがない場合、「MS Sample」が選ばれている状態であれば OK です。

Fractions [Sample Acquisition で Samples were fractionated を選択した時]

Fig.5-17 データの説明における「Fractions」にあたる項目を選択する事ができます。該当するものがなければ「Samples were fractionated」のチェックを外してください。該当するものがあるものの当てはまるカテゴリーが特にない場合、「MS Sample」が選ばれている状態であれば OK です。

[Experimental Design で Repeated Measures Design を選択した時]

The Time Category

Repeated Measures Design では同じ解析対象の経時変化や条件を変えたデータ測定を行う実験を想定 しており、そのようなカテゴリーを選択します。例えば「TimePoint」カテゴリー{属性値(0hr,1hr,2hr,3hr)}や、 「Treatment」カテゴリーなどです(Fig.5-21)。

Experimental Design 💿 Repeated Measures Design 🔿 Two-Way Design										
Summarization Hierarchy										
Available Categories 💡	What is the time (rep. measures) Category? 💡									
(None)	< <tr> Timepoint What is the biological subject Category? ? > species</tr>									
< Clear Summarization										

Fig.5-21 Summarization Hierarchy 設定 / Repeated Measures Design

[Experimental Design で Two-Way Design を選択した時]

Experimental Design	Design (a) Two-Way Design
Summarization Hierarchy	besign 🕑 two way besign
Available Categories 💡	What is the primary analysis Category? 💡
(None)	Treatment What is the secondary analysis Category? ? sex
	Which Category identifies the biological samples? MS Sample

Primary Analysis Category ,Secondary Analysis Category

Fig.5-22 Summarization Hierarchy 設定 / Two-way Design

Two-way Design は2つの独立変数の効果を比較し、2 つの変数の間に交互作用があるかどうかを判断 する際に選ぶ選択肢で、例えばある薬剤を投与した時の男性と女性の反応の違いについての比較などが 例に挙げられます。この時実験の主な焦点となるカテゴリーを **Primary Analysis Category** として、 結果に影響を及ぼす可能性があるカテゴリーを **Secondary Analysis Category** で指定します。例では Primary Analysis カテゴリーに「**Treatment**」{control,treated}を、Secondary Analysis カテゴリーに 「**sex**」{male,female}を指定しています。

Design Matrix

設定した Experimental Design の内容に従って、カテゴリーの属性の分類とその属性を持つサンプル データを確認する事が表形式でまとめられます。

Design Matrix	
	All Samples
	20121130_MD_10B (20121130_MD_10B.mgf-allergens-9860)
	20121130_MD_14B (20121130_MD_14B.mgf-allergens-9878)
	20121130_MD_10D (20121130_MD_10D.mgf-allergens-9862)
	20121130_MD_2D (20121130_MD_2D.mgf-allergens-9828)
	20121130_MD_5A (20121130_MD_5A.mgf-allergens-9838)
	20121130_MD_78 (20121130_MD_78.mgf-allergens-9848)

Fig.5-24 Design Matrix

5-7-2. Statistical Analysis タブ

検定手法や多重検定の補正についての設定を行います(Fig.5-24)。 各統計手法や補正に関する詳細は、「**11-2.検定**(P.85~)」をご覧ください。ここでは設定項目について 簡単に説明します。

🛃 Configure Sample O	rganization and Statistical Analysis	×
Sample Hierarchy S	itatistical Analysis	
Choose and configure a statistical test. The selection of displayed tests depends on your experimental design (set on the Sample Hierarchy tab). If a displayed test is unavailable, its tooltip will explain why.	Statistical Test P ≥2 Treatments ANOVA / t-test P ≥2 Treatments Permutation Test NP ≥2 Treatments Mann-Whitney U Test NP Exactly 2 Treatments Kruskal-Wallis Test NP ≥2 Treatments None None None	Test Hypothesis H _g : Treatment does not have an effect on group mean H _a : Treatment has a significant effect on group mean
Check/uncheck columns or rows in the Design Matrix to determine which samples are included in the statistical analysis.	Multiple Test Correction Control FWER with Hochberg's step-up and Holm's step-down Significance level α 0.5	Reference Samples Control Use matched reference samples Use all reference samples
Design Matrix		
	✓ Control ✓ Insulin Control_2 Insulin_3 Control_1 Insulin_2 Control_3 Insulin_1	
0		Apply Cancel

Fig.5-24 Statistical Analysis タブ

Statistical Test

選択された実験計画の種類や数に基づき、実施/結果表示 可能な検定の選択肢が現れます (Fig.5-25) 。



Fig.5-25 Statistical Test

Test Hypothesis

選択された検定に合わせて帰無仮説(H₀)と対立仮設(H_a)の内容が表示されます。

Multiple Test Correction

多重検定の補正、すなわち多重検定に伴う擬陽性の上昇を抑えるために新たに設ける基準についての 設定です。補正方法を選び基準値を指定します。詳細は「11-2.検定」をご覧ください。

Reference Samples

表示する Intensity 情報を Fold Change とした場合、Reference とするサンプルをどれにするか指定 します。さらに対象とする Reference を、マッチしたデータに対応するサンプルに限定か、カテゴリー内 全てのサンプルを対象としたものにするかを選ぶことができます。

Design Matrix

Sample Hierarchy タブでも表示されているのと同様、設定した Experimental Design の内容に 従って、カテゴリーの属性の分類とその属性を持つサンプルデータを確認する事が表形式で まとめられます。こちらのタブではさらに、解析対象とする属性などをチェックボックスで ON/OFF し、 検定の対象を変更する事も可能です。



Fig.5-26 Design Matrix (Statistical Analysis タブ)

6. Samples 画面

6-1. Samples 画面について

Samples 画面では、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 表示されます(Fig.6-1)。解析結果を確認する上で主体となる画面です。



Fig.6-1 Scaffold DIA Samples 画面

この章では Samples 画面の表の見方(図[1] Samples Table)と、表示する定量値の内容(図[2] DisplayType バー)について詳しく説明します。

6-2. Samples Table (Fig6-1 $\mathcal{O}[1]$)

Samples Table では、同定タンパク質の一覧並びにそのタンパク質の関連情報、定量値が表示されます。

Samples Table で表示されるタンパク質は、各種フィルターリング機能で表示対象を変更する事が できます。詳しくは4章「メイン画面」の「4-4.表示タンパク質のフィルターリング」(P.31~)を ご覧ください。また各列をクリックする事で降順・昇順に並び替えることができます。

次頁以降で、tableに表示されている各列の内容について説明いたします。

				C	olor Lege	nd (Disp	olayed Value)									
						6.36E8	3	11				7	t.	9	1(5
						1.22E8	3						0	U		
						2.33E7	7		Ser		6	Core	<mark>_0</mark>		tz	
						4.46E6	5		- m		le le	s dr	ptic		, ⊢ +	
1	L	2	3			8.54E	5		Z		ar M	Grot	d Pe	₹.	tion	
	#	Visible	Star	Pro	tein Name	4			Accessio	5	Molecul	Protein (ldentifie	Exclusivi	Permuta CI - Treat	
	1	\checkmark		sp	O43707 A	CTN4_H	JMAN Alpha	sp O4	3707 ACT	N4_HU	105 kDa	1	51	▲ 69%	0.50)
	2	 Image: Image: Ima		sp	O75533 SF	F3B1_HU	MAN Splicin	sp 07	'5533 SF3E	31_HUM	146 kDa	1	40	0100%	0.30)
	3	✓		sp	075694 N	U155_HU	JMAN Nucle	sp 07	5694 NU1	55_HU	155 kDa	1	20	100%	0.20)
	4			sp	P00558 P0	GK1_HUN	MAN Phosph	sp P0	0558 PGK	1_HUM	45 kDa	1	30	0100%	0.50)
	5	✓		sp	P06733 EN	NOA_HU	MAN Alpha	sp P0	6733 ENO	A_HUM	47 kDa	1	30	100%	0.20)
	6			sp	P08243 AS	SNS_HU	MAN Aspara	sp P0	8243 ASN	S_HUM	64 kDa	1	10	100%	0.10	
	7	 Image: Image: Ima	☆	sp	P08729 K2	2C7_HUN	IAN Keratin,	sp P0	8729 K2C7	_HUMAN	51 kDa	1	33	100%	0.60)

Fig.6-2 Samples table 左側

番号	項目	説明
1	#	上位から順に割り当てた番号
2	Visible	「Hidden」フィルターにより該当タンパク質をリストに表示するか
		しないかを定めます
3	Star	タンパク質に付与できるマーク。無色/黄色/青/黄色+青の4種類を
		設定可能
4	Protein Name	タンパク質の名称
5	Accession Number	タンパク質の ID
6	Molecular Weight	タンパク質の質量
7	Protein Group	Clustering グループに属するタンパク質にアサインされた全ペプチド
	Score	の中で最も高いペプチドの同定確率の値。1、すなわち 100%、が
		最大値で数値が大きいほど信頼度の高いペプチドがアサインされて
		いる事を示しています。
8	Identified Peptide	同定ペプチド数
	Count	
9	Exclusivity	各タンパク質の全アサインペプチドにおける、Exclusive ペプチドの
		割合。数値が高いほどリスト上の他タンパク質とシェアされている
		ペプチドが少ない事を示し、Exclusivity が 100%のタンパク質は
		すべて他タンパク質とはシェアされていないペプチドだけがアサイン
		された状態である事を示しています。
10	検定に関する値	検定(Experiment -> Quantitative Analysis)で実施した検定に関連
		する数値。p値、p値や変動係数など
11	Color Legend	項目 14(次頁)の定量値表示の際の凡例。Display type の切り替えに
		より凡例表示も切り替わります。

								1	3											
	В	io la	gio	al	Pro	ce	SS	Ce	llu	lar	Co	m	Mo	lec	:u	٦				ſ
12 Amouooxe	biological adhesion	biological regulation	cellular process	developmental process	enzyme regulator activity	establishment of localization	growth	cytoplasm	extracellular region	membrane	nucleus	ribosome	binding	catalytic activity	structural molecule activity	ſ	control	14 /) [
Homo sapiens		•	•			•		٠	•		٠		•				8.76	8.	73	
Homo sapiens		٠	•		٠	•		•		•	٠		•	•			8.46	8.	39	
Homo sapiens			•			•		•		•	٠		•	•			7.98	7.	99	
Homo sapiens	٠	٠	•	•		•		•	٠	•	٠		•		٠		7.99	8.	02	
Homo sapiens			•	•		•		•		•		•	•				8.09	8.	04	
Homo sapiens		٠	٠	٠				•			٠		•	٠			8.62	8.	64	
			F	~ (a-	3	9	m	n	00	+	b.		古	伯山					

Fig.6-3 Samples table 右側

番号	項目	説明
12	Taxonomy	生物種
13	GO info	Gene Ontology 情報
14	Quantitation	定量値。表示する数値は Display type バーの設定内容に基づきます。また
		「Summarization」の選択項目により、データをまとめる単位が変更され、
		定量値を表示する列数が変化します。

6-3. Display Type バー (Fig6-1 の2)

	1	2	3	}		4	
Display Type:	Exclusi	ive Intensi 🗸 🗌 Normalize	d 🗌 Log	Inter	nsities C	olor Opt	tions····
	Total In Exclusi Total In Exclusi Total Q	itensity ve Intensity (Log: Fold Change) itensity (Log: Fold Change) ve Quantified Peptide Count uantified Peptide Count	'alue)			lumber	

Fig.6-4 Display Type //-

上記 14 の定量値について、表示内容を切り替えることができるのが「Display Type」 バーです (Fig.6-4)。以降、各項目について説明します。

1. Display Type

表示する数字を以降の8種類から選択する事ができます。

- Exclusive Intensity

share ペプチドを除く Exclusive(ユニーク)なペプチドを計算対象としてタンパク質の定量値を 算出

- Total Intensity

share ペプチドも含む全ペプチドを計算対象としてタンパク質の定量値を算出

- Log₂ Fold Change (Exclusive Intensity)

share ペプチドを除く Exclusive な(ユニークな)アサインペプチドを対象としてタンパク質の定量値 を算出し、Reference との比を取りその数字に対して 2 を底にする Log をとった値。Scaffold シリーズでは 比を取り 2 を底にする Log を「Fold Change」 として呼んでいます。

Fold Change を選択した場合、Reference を指定する欄が新たに表れます(下図 Fig.6-5)

Display Type:	Exclusive Intensi 🗸	Ref:	Water	~	Normalized

Fig.6-5 Fold Change を選択した時は reference を選択する項目が現れる

Log₂ Fold Change (Total Intensity)

share ペプチドを含む全アサインペプチドを対象としてタンパク質の定量値を算出し、Referenceとの比を取りその数字に対して2を底にする Log をとった値。Fig.6-5 Reference を指定する欄が現れます。

- CV (Exclusive Intensity)

share ペプチドを除く Exclusive な(ユニークな)ペプチドを対象とした定量値の変動係数

- CV (Total Intensity)

share ペプチド含むすべてのペプチドを対象とした定量値の変動係数

- Exclusive Quantified Peptide Count

share ペプチドを除く Exclusive な(ユニークな)ペプチドでかつ定量に使われたフラグメントを持つ ペプチドに限定して数え上げた数に対して、標準化処理を行った数字 (**11-1-1**, P.84~)。

- Total Quantified Peptide Count

share ペプチドを含む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て数え 上げた数に対して、標準化処理を行った数字(**11-1-1**, P.84~)

2. Normalized

Normalization を実施する/しない を切り替えることができます。サンプルの階層構造定義時に TR(Technical Replicate) として定義されたグループの単位毎に Normalization の処理が 行われ、平均の4分位数と各データの4分位数の値を揃えるように調整されます。詳細は「**11-1-2. ラベルフリー定量**」の「normalization」(P.85~)をご覧ください。

3. Log Intensities

タンパク質の定量値を、底を10としたログに変換して表示

4. Color Options

凡例の色分けについて、その境値と色を定義します。ユーザーにより自由に設定変更ができます。

🚂 Edit Coloring for Display Type "Total Log10 Intensity"	×
Adjust Colors Double-click on a color to change it, or on the track to add a color. To remove a	color, select its handle and press the delete key.
853,702.124 🗘 Selected Value:	635,899,664.106 🗘
✓ Use color gradient	
Help	Restore ScaffoldDIA defaults OK Cancel



6-4. 「Summarization」と定量値

データを取り込んだ当初の Samples 画面では、各データの定量値がそのまま表示されている状態です (Fig.6-7)。

🧱 Scaffold DIA - Den	no2_HeLa_i	nsulin_	6_files_30_proteins.sdia												12		×
Filters			Summarization: MS Sample	Protein FDR 1.0% F		Min # Peptide	es 2	0									
	Display	Туре:	Exclusive Intensity V Normalized	Log Intensities Color Op	tions			unt									
Samples	#	Visible	5.37E8 9.96E7 1.84E7 3.42E6 6 3.32E5 § Protein Name	Accession Number	Molecular Weight	Protein Group Score	Identified Peptide Count	Total Unique Peptide Co	Exclusivity	Taxonomy	Control_1	Control_2	Control_3	Insulin_1	Insulin_2	Insulin_3	
Proteins	1 2 3		sp[P49792]RBP2_HUMAN E3 SUMO-prote sp[Q5T4S7]UBR4_HUMAN E3 ubiquitin-pr sp[P15924]DESP_HUMAN Desmoplakin O	sp P49792 RBP2_HUMAN sp Q5T4S7 UBR4_HUMAN . sp P15924 DESP_HUMAN	358 kDa 574 kDa 332 kDa	0.996 0.995 0.994	58 53 53	58 53 53	100%100%100%	Homo sapiens Homo sapiens Homo sapiens	2.92E8 9.62E7 1.01E8	3.23E8 9.10E7 1.08E8	3.34E8 9.81E7 9.54E7	2.88E8 1.00E8 1.08E8	2.29E8 1.26E8 1.09E8	2.73E8 9.84E	2
Visualize	4 5 6 7 8		spIP43/0//ACINA_HUMAN Aipna-actinin. spIP18206/VINC_HUMAN Vinculin OS=Ho. spIP11388/TOP2A_HUMAN DNA topoiso spIP12814/ACTN1_HUMAN Alpha-actinin spIP918210/DX21 HIMAN Nucleolar RN	 sp[043/0/]ACIN4_HUMAN sp[P18206]VINC_HUMAN sp[P11388]TOP2A_HUMAN sp[P12814]ACTN1_HUMAN sp[O9NB30IDDX21_HUMAN 	105 kDa 124 kDa 174 kDa 103 kDa 87 kDa	0.997 0.996 0.996 0.994 0.997	50 45 45 43 41	50 45 45 43 41	 68% 100% 100% 63% 100% 	Homo sapiens Homo sapiens Homo sapiens Homo sapiens	5.51E8 6.33E8 4.01E8 1.37E8 5.25E8	5.88E8 6.86E8 4.28E8 1.33E8 5.25E8	5.09E8 5.15E8 4.16E8 1.23E8 5.34E8	5.63E8 4.41E8 1.35E8 5.76E8	5.02E8 5.02E8 4.32E8 1.21E8 5.09E8	4.87E8 4.28E0 1.27E1	
M	9 10 11 12		splQ9P2E9[RRBP]_HUMAN Ribosome-bin. splQ86UP2[KTN1_HUMAN Kinectin OS=H. splQ75533[SF381_HUMAN Splicing factor splP22102[PUR2_HUMAN Trifunctional p	sp Q9P2E9 RRBP1_HUMAN sp Q86UP2 KTN1_HUMAN sp Q75533 SF3B1_HUMAN sp P22102 PUR2_HUMAN	152 kDa 156 kDa 146 kDa 108 kDa	0.994 0.993 0.998 0.997	41 41 40 38	41 41 40 38	 100% 100% 100% 100% 	Homo sapiens Homo sapiens Homo sapiens Homo sapiens	1.31E8 1.39E8 2.11E8 5.15E8	1.50E8 1.40E8 2.48E8 5.58E8	1.17E8 1.27E8 2.22E8 5.49E8	1.14E8 1.04E8 2.07E8 5.58E8	1.12E8 1.13E8 1.87E8 5.08E8	1.14E8 1.15E8 2.08E1 4.94E	
Analysis	13 14 15 16 17		spIP34932[HSP74_HUMAN Heat shock 70 . spIQ14152[EIF34_HUMAN Eukaryotic tran. spIP26639[SYTC_HUMAN ThreoninetR spIP78371ITCPB_HUMAN T-complex_prot.	sp P34932 HSP74_HUMAN sp Q14152 EIF3A_HUMAN sp P26639 SYTC_HUMAN sp P50990 TCPQ_HUMAN sp P78371 TCPB_HUMAN	94 kDa 167 kDa 83 kDa 60 kDa 57 kDa	0.996 0.994 0.994 0.993 0.995	38 38 38 36 35	38 38 38 36 35	 100% 100% 100% 100% 100% 	Homo sapiens Homo sapiens Homo sapiens Homo sapiens Homo sapiens	5.44E8 3.55E8 8.47E8 1.11E9 8.62F8	5.89E8 3.69E8 8.90E8 1.15E9 7.70E8	5.76E8 3.56E8 9.18E8 1.16E9 6.95E8	5.50E8 3.45E8 9.29E8 1.11E9 8.42E8	4.78E8 3.25E8 7.85E8 1.04E9 8.27F8	5.38E8 3.32E8 8.56E0 1.08E9	
Publish Proteins 0.0% FDR (attained) 29 Targets 0 Decoys	18 19 20 21		sp Q04637 IF461_HUMAN Eukaryotic tran. sp P08729 K2C7_HUMAN Keratin, type II c. sp P14625 ENPL_HUMAN Endoplasmin O. sp P00558 PGK1_HUMAN Phosphoglycer.		175 kDa 51 kDa 92 kDa 45 kDa	0.997 0.996 0.994 0.996	34 34 34 30	34 34 34 30	 100% 100% 100% 100% 100% 	Homo sapiens Homo sapiens Homo sapiens Homo sapiens	3.89E8 1.90E9 1.21E9 3.76E9	4.52E8 2.31E9 1.34E9 3.56E9	4.32E8 2.19E9 1.26E9 4.09E9	4.37E8 2.09E9 1.15E9 3.67E9	4.06E8 1.92E9 1.15E9 3.25E9	4.21E8 1.91E9 1.12E9 3.28E4	
Peptides 0.0% FDR (attained) 1000 Targets	22 23 24 25		 splPub/33jENOA_HUMAN Alpha-enolase . splQ5UIP0[RIF1_HUMAN Telomere-associ splQ9BSJ8JESYT1_HUMAN Extended syna splP14923]PLAK_HUMAN Junction plako 	spjP06/33jENOA_HUMAN spjQ5UIP0jRIF1_HUMAN spjQ9BSJ8jESYT1_HUMAN spjP14923jPLAK_HUMAN	4/ kDa 274 kDa 123 kDa 82 kDa	0.994 0.997 0.998 0.995	30 25 23 22	30 25 23 22	 100% 100% 100% 100% 	Homo sapiens Homo sapiens Homo sapiens Homo sapiens	6.15E9 2.25E7 1.07E8 1.09E8	4.76E9 3.59E7 1.13E8 1.17E8	7.10E9 2.21E7 1.01E8 1.03E8	7.66E9 2.26E7 9.56E7 9.23E7	6.43E9 1.78E7 9.37E7 9.38E7	9.49E	7

Fig.6-7 データ取り込み直後の Samples 画面

「**Organize**」画面で各データの階層構造と属性を定義すると、その定義内容が Samples 画面にも反映 されます(Fig.6-8)。表右側の定量値表示で、データ名の上部に階層の属性値が表示されます。

Scaffold DIA - Den	no2_HeLa_insulin_6_files_30_proteins.sdia												3200	E	i ×
File Edit View E	xperiment Export Tools Help														
Filters	Kummarization: MSS	p-value filter ▼	GO Term	FDR 1.0%	FDR ~	Min # P	eptides 2	0	0						
Organize Samples	Display Type: Exclusive Intensity tor Legend (Displayed V 5.37E8 9.96E7 1.84E7 3.42E6 6.33E5 # S Protein Name	Vormalized Log Inten	wolecular Weight	Color Opti	identified Peptide Count	Total Unique Peptide Count	Exclusivity	t-test CL: Treatment	Taxonoury	Control_1	Control_2	Control_3	Insulin_1	Insulin Z ⁻ ujinsul	Insulin_3
1	1 spjP49792jRBP2_HUMAN	sp/P49792/RBP2_HUMAN 35	8 kDa	0.996	58	58	• 100%	0.080	Homo sapiens	2.0	3.2	3.3	2.8	2.29	2.7
Proteins	2 ✓ spjQ51457/UBR4_HUMAN 3 ✓ spjP15924[DESP_HUMAN 4 ✓ spjQ43707[ACTN4_HUMA	sp P15924 DESP_HUMAN 33 sp P15924 DESP_HUMAN 33 sp O43707 ACTN4_HU 10	4 kDa 2 kDa 5 kDa	0.995 0.994 0.997	53 53 50	53 50	 100% 100% 68% 	0.21 0.49 0.16	Homo sapiens Homo sapiens Homo sapiens	9.6 1.0 5.6	1.0 5.8	9.5 5.6	1.0 5.7	1.09 5.06	9.8 9.8 5.2
	5 SpjP18206jVINC_HUMAN 6 SpjP11388jTOP2A_HUMAA	sp P18206 VINC_HUMAN 12- sp P11388 TOP2A_HU 17-	4 kDa 4 kDa	0.996	45 45	45 45	 100% 100% 53% 	0.17	Homo sapiens Homo sapiens	6.3 4.0	6.8	5.1 4.1	5.6	5.02	4.8
Visualize	8 ✓ splQ9NR30 DDX21_HUMA. 9 ✓ splQ9PE9 RRBP1_HUMA	sp[Q9NR30 DDX21_HU 87 sp[Q9P2E9 RRBP1_HU 15.	kDa 2 kDa	0.994 0.997 0.994	43 41 41	43 41 41	 03% 100% 100% 	0.62	Homo sapiens Homo sapiens Homo sapiens	5.2 1.3	5.2 1.5	5.3 1.1	5.7 1.1	5.09 1.12	5.2 1.1
MA	10 ✓ spjQ86UP2 KTN1_HUMA 11 ✓ spjO75533 SF3B1_HUMA 12 ✓ spjD75533 SF3B1_HUMA	 sp Q86UP2 KTN1_HUM 15 sp O75533 SF3B1_HUM 14 sp D22102 D1P2_HUM10 	6 kDa 6 kDa	0.993	41 40	41 40	 100% 100% 100% 	0.011	Homo sapiens Homo sapiens	1.3 2.1	1.4 2.4	1.2 2.2	1.0 2.0	1.13 1.87	1.1 2.0
Analysis	13 ✓ ↔ spjP22102[P0K2_HUMAN 13 ✓ ↔ spjP34932]HSP74_HUMA. 14 ✓ ↔ spjQ14152[EIF3A_HUMA	sp P34932 HSP74_HUM 94 sp Q14152 EIF3A_HUM 16	kDa 7 kDa	0.996	38 38	38 38	 100% 100% 	0.42	Homo sapiens Homo sapiens	5.4 3.5	5.8 3.6	5.7 3.5	5.5 3.4	4.78	5.3 3.3
	15 ✓ ☆ splP26639 SYTC_HUMAN 16 ✓ ☆ splP50990 TCPO_HUMAN	sp P26639 SYTC_HUMAN 83 sp P50990 TCPQ_HUM 60	kDa kDa	0.994	38 36	38 36	100%100%	0.56	Homo sapiens Homo sapiens	8.4 1.1	8.9 1.1	9.1 1.1	9.2 1.1	7.85	8.5
Publish	17 ✓ ☆ sp P78371 TCPB_HUMAN 18 ✓ ☆ sp Q04637 IF4G1_HUMA 19 ✓ ☆ sp P08729 K2C7 HUMAN	sp P/8371 TCPB_HUMAN 57 . sp Q04637 IF4G1_HUM 17 sp P08729 K2C7 HUMAN 51	kDa 5 kDa kDa	0.995 0.997 0.996	35 34 34	35 34 34	 100% 100% 100% 	0.57	Homo sapiens Homo sapiens Homo sapiens	8.6 3.8 1.9	7.7 4.5 2.3	6.9 4.3 2.1	8.4 4.3 2.0	8.27 4.06 1.92	4.2

Fig.6-8 Organize で階層構造と属性定義後の Samples 画面

Summarization の選択を切り替えて別の階層を選択すると、その階層における属性別にデータが まとめ直されます(Fig.6-9)。

🖬 🔤 🚍	0	1	Summarization: Treatment	t ~	Protein FD	R 1.0% FDR	Min # F	eptides	2 🗘 🕷	0			
Iters													
			[a	0		0							
Show Hidden			Name/Accession	p-value filter •	GO Term •	CL EAS							
0													
290	Displa		e: Exclusive Intensity V	Normalized Log Intens	ities C	olor Options							
		5 .5F		······································									
Organize			plor Legend (Displayed Val	u			*	uno					
			5.37E8				onu	e C					
COO			9.96E7	<u>er</u>	Ŧ	COL	le C	ptid					
			1.84E7	5	/eig	d S	btic	e Pe		¥			
			3.42E6	2	ar V	loc	dP	inbi	₽	mei	è		
Samples		a	6.33E5	SSIC	G	.u	ifie	5	NIS I	reat	Loc /	2	.=
	#	lisib	Protein Name	Acce	Aole	rote	dent	otal	xclu	-test	axo (out	Insu
	1	-	spiP49792IRBP2 HUMAN F	SDIP49792IRBP2 HUMAN	358 kDa	0.996	58	58	. 100%	0.080	Homo saniens	2 2358	2.75
	2	V	splO5T4S7IUBR4 HUMAN	spIO5T4S7IUBR4 HUMAN	574 kDa	0.995	53	53	. 100%	0.21	Homo sapiens	9.62E7	1.00E
Proteins	3		spIP15924IDESP HUMAN D	spiP15924IDESP HUMAN	332 kDa	0.994	53	53	• 100%	0.49	Homo sapiens	1.01E8	1.08E
	4		sp[043707]ACTN4_HUMAN	. sp[043707]ACTN4_HUM	105 kDa	0.997	50	50	A 68%	0.16	Homo sapiens	5.69E8	5.23E
	5	<	sp P18206 VINC_HUMAN V	sp P18206 VINC_HUMAN	124 kDa	0.996	45	45	• 100%	0.17	Homo sapiens	6.33E8	5.02E
	6	\checkmark	sp P11388 TOP2A_HUMAN	sp P11388 TOP2A_HUM	174 kDa	0.996	45	45	100%	0.11	Homo sapiens	4.16E8	4.32E
	7	\checkmark	Sp/P12814/ACTN1_HUMAN	sp P12814 ACTN1_HUM	103 kDa	0.994	43	43	6 3%	0.62	Homo sapiens	1.33E8	1.27E
Visualize	8	\checkmark	splQ9NR30 DDX21_HUMA	sp Q9NR30 DDX21_HUM	87 kDa	0.997	41	41	100%	0.78	Homo sapiens	5.25E8	5.20E8
	9	<	sp[Q9P2E9]RRBP1_HUMAN	. sp[Q9P2E9]RRBP1_HUM	152 kDa	0.994	41	41	100%	0.10	Homo sapiens	1.31E8	1.14E8
	10	\checkmark	sp Q86UP2 KTN1_HUMAN	splQ86UP2[KTN1_HUMAN	156 kDa	0.993	41	41	• 100%	0.011	Homo sapiens	1.39E8	1.1368
Addit	11	~	sp[075533]SF3B1_HUMAN	sp[075533]SF3B1_HUMAN	146 kDa	0.998	40	40	100%	0.11	Homo sapiens	2.22E8	2.07E
	12		sp P22102 PUR2_HUMAN T	. sp[P22102]PUR2_HUMAN	108 kDa	0.997	38	38	100%	0.42	Homo sapiens	5.49E8	5.08E8
Anaiysis	13	~	sp P34932 HSP74_HUMAN	. sp P34932 HSP74_HUMAN	94 kDa	0.996	38	38	• 100%	0.15	Homo sapiens	5.76E8	5.38E8
	14		splQ14152[EIF3A_HUMAN	sp[Q14152[EIF3A_HUMAN	167 kDa	0.994	38	38	100%	0.024	Homo sapiens	3.56E8	3.32E
	15		sp P26639 SYTC_HUMAN T	spIP26639[SYTC_HUMAN	83 kDa	0.994	38	38	• 100%	0.56	Homo sapiens	8.90E8	8.56E
	16		splP50990[TCPQ_HUMAN T	. splP50990[TCPQ_HUMAN	60 kDa	0.993	36	36	• 100%	0.059	Homo sapiens	1.15E9	1.08E
Bublich	17		sp[P78371]TCPB_HUMAN T	sp[P78371]TCPB_HUMAN	57 kDa	0.995	35	35	• 100%	0.57	Homo sapiens	7.70E8	8.27E
Fublish	18		sp[Q04637]IF4G1_HUMAN	sp[Q04637]IF4G1_HUMAN	175 kDa	0.997	34	34	• 100%	0.91	Homo sapiens	4.32E8	4.2168
eins	19	\checkmark	sp[P08729]K2C7_HUMAN K	sp[P08729]K2C7_HUMAN	51 kDa	0.996	34	34	100%	0.31	Homo sapiens	2.19E9	1.92E

Fig.6-9 Summarization 切り替えを行うと、定量値のデータまとめ方も変わる

表示の切り替えは Summarization 選択と連動して行われます。同じ属性の複数データがどのように まとまるかについては、次の項目「**6-5.定量値データのまとめられ方**」をご覧ください。

6-5. 定量値データのまとめられ方

【Rolling up, 代表値はメジアンです】

Summarization のレベルを変更し階層構造が上のレベルを選択すると、その階層に属する各定量値からメジアンが計算され代表値として表示されます。

【不足値と表記】



Fig.6-10 不足値がもたらす特別表記: Missing Value, No Values, Missing Ref

Log2 Fold Change あるいは Summarization 上位階層選択時の Intensity (または Log₁₀)をまとめる(複数サンプルから代表値を計算する)際、値の不足する状況に応じて以下のような表記がされます (Fig.6-10)。

Missing Value	ratio において参照先(分数の分子)の値がない
Missing Ref(又は Reference Missing)	ratio において参照元(分数の分母)の値がない
No Values	ratio において参照先も参照元も値がない

Summarization により Rolling up する時、50%以上のデータが不足している時には別途計算式によって代表値が計算されます(Fig.6-11,緑線)。またすべて Missing Value であった場合は、まとめた時にも Missing Value として表示されます(次頁 Fig.6-11,黒線)。



Fig.6-11 Rolled up & missing value

7. Proteins 画面

7-1. Proteins 画面について

Proteins 画面ではペプチド/タンパク質の同定並びに定量計算結果をより詳しく検証する事ができます。 タンパク質にアサインされているペプチドの情報、ペプチドが定量計算に使用されているかどうか、 グループに属するタンパク質のペプチドシェアの状況、等を確認する事ができます。Proteins 画面へ移行 するには、Samples 画面で気になるタンパク質が選択されている状態でダブルクリックするか、画面左の 「Proteins」pane(Fig.7-1,緑枠)をクリックします。

Proteins 画面は、2つのタブから構成されています。「**Protein**」タブと「**Similar Proteins**」タブです (Fig.7-1、青枠)。

Protein タブではペプチド/タンパク質の同定内容やペプチドの定量計算結果を詳しく検証する事ができます。それに対して「Similar Proteins」タブでは、配列が類似するタンパク質を集めたクラスターグループ に属する、各タンパク質のペプチドシェアの状況を確認する事ができます。



Fig.7-1 Proteins 画面 「Protein」 タブ

Protein タブは主に4つのパーツから構成されています(Fig.7-1 [1]~[4])。以降、各パーツで表示 される内容について説明します。

7-2. ペプチドフィルターリング (Fig.7-1 [1])

Protein タブの上部にあるのが「ペプチドフィルターリング」です(Fig.7-2)。



Fig.7-2 「Protein」タブ ペプチドフィルターリング

ペプチド情報(Fig.7-1 [2])で表示されるペプチドを絞り込む際に利用します。

A でグループ内のタンパク質を特定しており、B でアミノ酸配列あるいは修飾などのキーワードをもとに さらにペプチドを絞り込むことができます。絞り込みをリセットする場合は C「Showing All Peptides」 ボタンをクリックします。この「Showing All peptides」は Proteins 画面内での各種操作により表示 ペプチドが絞り込み表示になってしまった状態を元に戻す、フィルターリセットをしたい場合に利用します。 なおペプチドの絞り込みは、[4] Visualization でタンパク質のカバレージ情報を表示させている際、黄色 のハイライト箇所である同定ペプチド部分をクリックする事でも行われます。

7-3. ペプチド情報 (Fig.7-1 [2])

[2]の「ペプチド情報」欄では、選択したタンパク質あるいはクラスターグループにアサインされ、かつフィルターリング条件を満たしたすべての同定ペプチドが一覧表示されます(Fig.7-3)。

Quantified	Peptide Sequence	Quantified Matches	Pixed Modif	Variable Mo	Start	Stop	Protein Accessi	Proba
	AAPPPPPPPPPLESSPR	0 01 0			606	622	sp Q5T4S7 UBR4	86%
	AEHASSLLELASTTK	0 of 6			1065	1079	sp Q5T4S7 UBR4	100%
	ALGTLGMTTNEK	0 of 6			4803	4814	sp Q5T4S7 UBR4	99%
	APSYIEIFGR	0 of 6			2364	2373	sp Q5T4S7 UBR4	98%
V	AQQALSELHTVEK	3 of 6			1838	1850	sp Q5T4S7 UBR4	100%
	ASAQGDPDVPECLK	0 of 6	C12 Carbami		768	781	sp Q5T4S7 UBR4	100%
\checkmark	ASVVTASSGSALQYDTLISLM…	5 of 6			3253	3277	sp Q5T4S7 UBR4	100%
V	AVEMTDQLMVPTLGSQEGA…	6			1851	1874	sp Q5T4S7 UBR4	100%
				=				

Fig.7-3 「Protein」 タブ ペプチド情報

配列や修飾、ペプチドのタンパク質上の位置、同定確率や質量といった情報が表示されます。必要に 応じて各列の項目名称で降順や昇順の並び替えを行うことができます。またドラッグ&ドロップ操作 などで列の順番を入れ替えたり不要な項目を非表示にすることもできます。

「Quantified」のチェックの有無は、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 満たしタンパク質の定量計算に当該ペプチドが使用されているかどうかを表しています。

「**Quantified Matches**」は、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を 表しています。

行の後半に表示される情報はデータの確からしさを評価する指標です(次頁 Fig7-4)。

Probability \sim	Eragment Consistency Score	Peptide Profile Correlation	Peptide Mass
100%	0.841	0.254	2,578.286
100%	0.897	-0.059	2,769.387
100%	0.966	0.198	1,761.874
100%	0.969	-0.099	2,444.146
100%	0.612	0.488	1,888.957
100%	0.967	-0.152	1,697.734
4000/	0.05	0.245	4,005,0

Fig.7-4 「Protein」 タブ ペプチド情報

ペプチド情報の欄で表示されている指標は、サンプル全体にまたがって評価した内容です。

「**Probability**」:ペプチドの同定確率です。

「Fragment Consistency Score」: 強度値の算出に使用したフラグメントパターンの、サンプル間での 一貫性を表す指標です。同定結果が間違えている、あるいは定量計算への使用に適さない可能性がある ペプチドを評価する指標の一つです。「フラグメントのパターンの一貫性」とは Proteins 画面右下 の"Chromatograms"→"Fragments"のグラフで表現されている各 Intensity 曲線についての、サンプ ル間の一致度とも言えます。この指標の計算方法は以下の通りです。

n 個の DIA サンプルからなる実験で検出された各ペプチドについて、そのペプチドに割り当てられた 各フラグメントの強度を、各サンプルの強度の合計が 1 になるよう和で割ります。続いてサンプルと使用 したフラグメント、値が相対強度の行列を構成します。各セル(M_ij)には、i 番目のサンプルにおける j 番目のフラグメントの調整後の相対強度が含まれます。行列の各列は、サンプル間の単一のフラグメント の強度を表す「feature ベクトル」とみなす事ができ、各 feature ベクトルについて、その平均ベクトルから のユークリッド距離を算出します。ユークリッド距離の強度加重平均(以下式の avg)を使い、以下の式で Fragment Consistency Score が計算されます。

$$(fcs = \left(1 - avg/\sqrt{2}\right)^2$$

値が 1 に近いほどサンプル間の誤差が小さく、サンプル間の相関性がなく独立していると値が 0 に近くなります。サンプルが 2 つしかない場合は値の解釈に注意が必要です。

「Peptide Profile Correlation」:各ペプチドのペプチドプロファイルと平均ペプチドプロファイルとの ピアソン相関の係数です。Technical replicate 単位でピーク強度の和を取りその値を元に標準化処理を した後、すべてのペプチドから平均値をとった「average profile」を作成してピアソン相関の相関係数を 求めます。負の値の場合は逆相関があり、また0に近いと相関性がなく、他のタンパク質とペプチドが共有 されているなどといった要因が影響している可能性が考えられます。

この表内で選択しているペプチドについて、各サンプルにおける検出状況をまとめたのがすぐ下に表示 される「ペプチド-Query マッチ情報」(Fig.7-1 [3])です。またペプチドを選択しダブルクリックすると Samples 画面に戻り、該当ペプチドを含むデータ(タンパク質)のみがリストに表示されます。

7-4. ペプチド - Query マッチ情報 (Fig.7-1 [3])

Sequence ^	Modificatio	Anarge Global Probal	oility Gl	obal q-value Sa	ample Probability	Sample q-value	Sample				
AEAESLYQSK		2			92./0	0.002	Control-PEROXO_I	_rep1.mzML			
AEAESLYQSK		2			98%	0.002	Control-PEROXO_I	_rep2.mzML			
AEAESLYQSK		2			97%	0.004	Control-PEROXO_I	P_rep3.mzML			
AEAESLYQSK					444 /s	0.004	Control-PEROXO II	P ren4 mzMI			
AEAESLYQSK		Quant. Intensity	# Qua.	RT Start (min)	RT Center (min)	RT Stop (min)	Precursor MZ	C	Attributes)	
AEAESLYQSK		1.034E9	5	14.68	14.92	15.44	563.275	Control-PE	ROXO IP rep1.r	mzML,	
		2.825E8	5	14.99	15.27	15.79	563.275	Control-PE	ROXO_IP_rep2.r	mzML,	
		8.871E8	5	14.98	15.22	15.47	563.275	Control-PE	ROXO_IP_rep3.	mzML,	
		3.923E8	5	15.01	15.22	15.74	563.275	Control-PE	ROXO_IP_rep4.	mzML,	
		1.605E7	5	46.94	47.67	48.51	563.275	Control-PER	OXO_whole-ce	ll_rep	
		2.461E7	5	47.15	47.68	48.35	563.275	Control-PER	OXO_whole-ce	ll_rep	
		1.07467	c	47.04	47.60	40.02	E62 07E	Control DEE	OVO whole co	di ron	

Fig.7-5 「Protein」タブペプチド-Query マッチ情報 左側(上図)と右側(下図)

すぐ上のペプチド情報テーブルで選択しているペプチドについて、各サンプルにおける検出状況をまとめたのが「ペプチド – Query マッチ情報 [3]」です (Fig.7-5)。主な項目について、以下に説明をします。 「Global Probability」「Global q-value」「Sample Probability」「Sample q-value」

項目名の最初に「**Global**」がついている項目は、該当ペプチドがどこかのサンプル 1 つ以上に存在する 確率や Percolator で計算した時の q-value を表します。ペプチド-Query マッチ情報の表の中で最も 確率が高いサンプル 1 行にしか表示されません。一方「**Sample**」で始まる項目は、該当行のサンプルのみ を対象として求めた Probability や q-value が表示されます。

Quant. Intensity

ペプチドの定量値です。フラグメントピークの XIC のピーク面積(定量値)を基に算出されています。

[# Quant. fragment]

定量計算に使用したフラグメントの数です。

RT Start(min) RT Center(min) RT Stop(min)

該当ペプチドを検出し定量計算に利用した RT の開始・中央・終了の時間を示しています。

7-5. Visualization 情報 (Fig.7-1 [4])

画面の一番下、Visualization は7つのタブから構成されています(Fig.7-6)。

- Protein Sequence
- Protein Level Charts
- Statistics
- Peptide Intensities
- Fragment Intensities
- Chromatogram
- Fragmentation Table

Protein Sequence	Protein Level Charts	Statistics	Peptide Intensities	Fragment Intensities	Chromatograms	Fragmentation Table
------------------	----------------------	------------	---------------------	----------------------	---------------	---------------------

Fig.7-6 「Protein」 タブ Visualization 情報

7-5-1. Protein Sequence タブ

画面(Fig.7-7)上部にはタンパク質の全配列とマッチしたペプチド領域が、画面下部ではタンパク質全長を 1 行で表した状態で、ペプチドのマッチング状況を確認できます。

Protein Sequence Protein Level Charts Chromatogra	ams Fragment Intensities Fragmentation Table								
sp[Q5T4S7 UBR4_HUMAN E3 ubiquitin-protein li∉ase UBR4 OS=Homo sapiens GN=UBR4 PE=1 SV=1									
AIVELKNKPA	RWHKAKKVQL TPGQTEVKID LPLPIVASNL 3640	*							
MIEFADFYEN	YQASTETLQC PRCSASVPAN PGVCGNCGEN 3680								
VYQCНКСR <mark>SI</mark>	NYDEKDPFLC NACGFCKYAR FDFMLYAKPC 3720								
CAVDPIENEE	DRKKAVSNIN TLLDKADRVY HQLMGHRPQL 3760								
ENLLCK <mark>VNEA</mark>	APEKPQDDSG TAGGISSTSA SVNRYILQLA 3800								
QEYCGDCK NS	FDELSKIIQK VFASRK <mark>ELLE</mark> <mark>YDLQQR</mark> EAAT 3840								
K <mark>S S R T S V Q P T</mark>	FTASQYR <mark>ALS VLGCGHTSST KCYGCASAVT 3880</mark>	-							
756 out of 5183 (14.59%) amino acids identified with 23 modifications									

Fig.7-7 「Protein」 タブ Visualization 情報 Protein Sequence タブ

画面上部ではタンパク質のアミノ酸配列が表示されます。マッチしたペプチド部分が黄色のハイライトで 表記、修飾がある部分は色のついた〇で囲まれています。黄色のペプチドをダブルクリックすると画面上の ペプチド情報欄[2] でフィルターリングがかけられ、該当ペプチドのみが表示されます(この状況を解除 するためには「Show all peptides」をクリックします)。

画面下部のカバレージについては、表示の棒状の枠全体がタンパク質全長を表しており、左が N 未端 側、右が C 末端側です。枠の上には全長に対するペプチド残基長の割合(カバレージ)と、検出された修飾 の総数が表示されています。枠内で黄色のラインで塗られている箇所がペプチドマッチを、黒い線が引かれ た部分が修飾を表しています。

右クリックにより、表示されている情報を出力し各種フォーマットのファイルとして保存するメニューが 現れるほか、配列をコピーしたり BLAST 検索にかけたりする事もできます。

[次頁に続きます]

7-5-2. Protein Level Chart タブ

選択しているペプチドが属するタンパク質について、定量値をグラフ表示します。定量値の表示項目は Summarization で選択しているレベルと連動しています。表示形式には箱ひげ図と棒グラフ、 ヴァイオリン図の3種類があります(Fig.7-8)



Fig.7-8 「Protein」タブ Visualization 情報 Protein Level Chart タブ

棒グラフでは、メジアン値並びに誤差範囲が表示されます。

箱ひげ図では、四分位数並びに最大値と最小値が表示されます。

ヴァイオリン図は箱ひげ図と似ていますがデータの確率密度の情報も併せて表示されています。

それぞれグラフにカーソルを合わせると、ポップアップされたダイアログにグラフの主要な数字が表示 されます。

[次頁に続きます]

検定に関する図表です。

Statistics table :

検定に関連する数字をまとめた表で、実施した検定により表示される数字が異なります。例えば ANOVA/t-検定の場合、「平方和」「自由度」「平均の平方」「F-統計量」「F-統計量の有意性」です。 Interaction Chart:

ユーザーが一次変数と二次変数の間の交互作用効果の解釈を検討するための図です。

一次変数と二次変数の間に相互作用がない場合、線はほぼ平行になり、線の傾きは一次変数の効果を示します。線が平行ではないが交差しない場合、交互作用効果がないとは言えないため、二次変数がある 程度影響を及ぼしている事を理解した上で、一次変数の効果を検討することになります。線が交差する 場合2つの変数が有意な相互作用を示すことを意味し、一次変数の効果について一般的な結論を出す ことは不可能と言えます。



7-5-4. Protein Intensities タブ

Fig.7-9 「Protein」 タブ Visualization 情報 Peptide Intensities タブ

Peptide Intensities タブには、Fig.7-9 のようにタンパク質にアサインされた各ペプチドの定量値を 積み重ねた棒グラフが表示されます。タンパク質の定量値計算に使用された各ペプチドの全強度に対する 割合を表しており、各ペプチドの影響度(寄与の度合い)がわかります。

データをまとめる単位は選択中のカテゴリーに連動します。またサンプル横断的にデータを見ながら 組成の変わっている箇所に注目する事で、各ペプチドの定量評価における信頼度を検討する事も できます。例えばカテゴリー間で大きく割合が変動しているようであれば、該当ペプチドを信頼度の低い ものとして除外する事を検討する必要があります。なお検証の結果もし除外を行いたい場合、ペプチド 情報欄の Quantified チェックボックスをオフにしてください。

図のバンドにマウスを合わせると、ペプチドの強度値も含むペプチドの情報が表示されます。さらに クリックするとペプチドのフィルターリングが行われます。(フィルターリングを解除するには、タブ上部の Show All Peptides をクリックしてください)。

ペプチドを識別するための凡例表示切替オプションも準備されており、表示数が多い扱いにくい場合は 表示をオフにしてください。

7-5-5. Fragment Intensities タブ

「Fragment Intensities」タブでは、選択ペプチドについて定量計算に利用したフラグメントピークの 定量値の割合が表示されます(Fig.7-10)。ペプチドの定量データの信頼度を評価する際等に利用できま す。検証の結果もし除外を行いたい場合、ペプチド 情報欄の Quantified チェックボックスをオフにして ください。



Fig.7-10 「Protein」 タブ Visualization 情報 Fragment Intensities タブ

7-5-6. Chromatograms タブ



Precursor, Fragment それぞれにおける XIC、保持時間別の強度値です(Fig7-11)。

Fig.7-11 「Protein」 タブ Visualization 情報 Chromatgram タブ

左が Precursor、右が Fragment の図で、共に横軸は保持時間、縦軸が Intensity です。グラフ内で 青くなっている領域が定量計算に使用された時間帯です。左の Precursor 図では、+1,+2 の アイソトピックについても併せて表示されています。右の Fragments の図では定量計算に使用された フラグメントについて表示されています。検出されているものの定量計算には利用されていない フラグメントについては色のない点線で表示されています。Precursor, Fragments どちらの図に おいても、グラフ内の線にカーソルを合わせるとその線と凡例が強調表示されどの項目かわかります。 またドラッグ&ドロップの操作でグラフを拡大表示する事が可能であるほか、右クリックで表示を微調整 したり、画像やデータを保存したりする事も出来ます(次頁 Fig.7-12)。



Fig.7-12 Chromatgram タブ グラフ拡大

7-5-7. Fragmentation table タブ

ペプチドの配列から計算されるフラグメントの m/z 理論値一覧の表です。さらに検出された フラグメントを色分け表示しています (Fig.7-13)。

Prote	ein Sequence	e Protein	Level Charts	Chromat	ograms	Fragment	Intensities	Fragmentat	tion Table		
в	B Ions	B+2H	B-NH3	B-H2O	AA	Y Ions	Y+2H	Y-NH3	Y-H20	Y	₽
1	102.1			84.0	т	1,228.6	614.8	1,211.6	1,210.6	11	*
2	230.1		213.1	212.1	Q	1,127.6	564.3	1,110.6	1,109.6	10	
3	361.2		344.1	343.1	м	999.5	500.3	982.5	981.5	9	
4	432.2		415.2	414.2	Α	868.5	434.7	851.5	850.5	8	
5	561.2		544.2	543.2	Е	797.5	399.2	780.4	779.4	7	
6	660.3	330.7	643.3	642.3	v	668.4	334.7	651.4	650.4	6	
7	773.4	387.2	756.4	755.4	L	569.3		552.3	551.3	5	
8	870.4	435.7	853.4	852.4	Р	456.3		439.2	438.2	4	
9	957.5	479.2	940.4	939.5	s	359.2		342.2	341.2	3	
10	1,054.5	527.8	1,037.5	1,036.5	Р	272.2		255.1		2	
11	1,228.6	614.8	1,211.6	1,210.6	R	175.1		158.1		1	

Fig.7-13 「Protein」 タブ Visualization 情報 Fragment table タブ

7-6. Similar Proteins タブ

Similar Proteins タブでは、マッチしたペプチドの重なり具合を見てリスト内の同定タンパク質間の 類似度を確認する事ができます。類似タンパク質の候補として、protein inference にて良く使用される 概念の話で使用される用語でいえば same protein, sub-set protein,並びに family protein に 該当するタンパク質が表内に現れます。ユニークなペプチド(「Exclusive to」の列に色をつけて表示)や シェアされたペプチドがどれなのかをわかりやすく図示しています(Fig.7-14)。シェアペプチドは、その シェアされているタンパク質の数により円マークの色塗り部分が 1/2, 1/3, 1/4 ... と変わります。

Protein Similar Proteins	Protein Similar Proteins Search functions											
🗿 Protein	eptide Cluster of sp Q9	BQE3 TB	A1C_HUM	IAN Tubul	in alpha-1	LC chain C)S=Homo	sapiens				
Cluster of sp Q9BQE3 No Group												
Similarity Table		sp Q9	sp Q9	sp P6		_		~				
		MAI	MAN	MAN	MAN	MAN	NM N	MAI				
		Ξ,	로	Ē	로	로	≧	Ę				
Peptide Sequence	Exclusive To	AIO	4 1 8	.4₽	S,	Ъ	8	41A				
		<u> </u>	18	184	181	18/	118	TB				
		E E	823	366	748	EY2	1765	N36				
		8	동	.89	613	Ъ В	767	271				
		8	8	ds l	뷶	ds l	8	b				
TIGGGDDSFNTFFSETGAGK	sp Q9BQE3 TBA1C_H	0			0	0		0				
RTIQFVDWCPTGFK	sp Q9BQE3 TBA1C_H	0			0	0	0	0				
AVCMLSNTTAVAEAWAR	sp Q9BQE3 TBA1C_H	<u> </u>										
DVNAAIATIK	sp Q9BQE3 TBA1C_H	0			0	0		0				
AVFVDLEPTVIDEVR	sp Q9BQE3 TBA1C_H	<u> </u>						0				
AYHEQLTVAEITNACFEPANQ	sp Q9BQE3 TBA1C_H	0										
TIQFVDWCPTGFK	sp Q9BQE3 TBA1C_H	<u> </u>			0	0	0	0				
EIIDLVLDR	sp Q9BQE3 TBA1C_H	0						0				
YMACCLLYR				0				0				
AFVHWYVGEGMEEGEFSEAR				0	0		0	0				
LISQIVSSITASLR		<u> </u>	<u> </u>	0			0					
VGINYQPPTVVPGGDLAK				\circ	0	0	0	0				
IHFPLATYAPVISAEK				0	0	0		0				
FDGALNVDLTEFQTNLVPYPR		0		0	0	0	0	0				
FDLMYAKR		0		0	0		0	0				
QLFHPEQLITGKEDAANNYAR		0		0	0		0	0				
QLFHPEQLITGK		0		0	0	0	0	0				
RNLDIERPTYTNLNR		0		0	0	0	0	0				
FDLMYAK		0		\bigcirc	0	0	0	0				
NLDIERPTYTNLNR		0		0	0	0	0	0				
EDAANNYAR				\bigcirc	0		0	0				
LDHKFDLMYAK				\bigcirc	0		0	0				
QIFHPEQLITGK	sp Q9H853 TBA4B_HU		\bigcirc									
AVCMLSNTTAIAEAWAR	sp P68366 TBA4A_HU			\circ	0	0	0	0				
AYHEQLSVAEITNACFEPANQ	sp P68366 TBA4A_HU			\bigcirc	0	0		0				
DVNAAIAAIK	sp P68366 TBA4A_HU			0								

Fig.7-14 「Similar Proteins」 タブ

8. Visualize 画面

8-1. Visualize 画面について



Fig.8-1 Visualize 画面

Visualize 画面では、主に定量的なデータの傾向をつかむのに役立つグラフを提供しています。 Visualize 画面には、「Quantitation」タブと「Principal Component Analysis」タブ、そして 「Heatmap」タブの3つが存在します(Fig.8-1赤丸部分)。

以降、それぞれのタブ内にある図について説明します。

8-2. Quantitation タブ

Quantitation タブは4つのパーツから構成されます (Fig.8-2)。



Fig.8-2 Visualize 画面 Quantitation タブ

- ① Volcano plot (8-2-1, P.70)
- 2 Quantitative Scatterplot (8-2-2, P.71)
- 3 Quantitative Trend Chart, Quantitative CVs (8-2-3, P.72)
- 4 GO Term Pie Chart, GO Connections Graph (8-2-4, P.73)

以降、それぞれのグラフについて説明します。

8-2-1. Protein Sequence

Volcano plot は2つのグループを比較する検定を行った時のみ表示されます(Fig.8-3)。x 軸は Fold Change(ratio,かつ Log2 変換)、y 軸は検定時に計算された p-value の Log₁₀ にマイナスを 掛けた値です。プロット 1つ1つはタンパク質の定量データに対応しています。



Fig.8-3 Visualize 画面 Quantitation タブ Volcano plot

縦軸の値が上に行けば行くほど p-value が低く統計的に有意な差がある事を表し、また横軸については 0 から離れるほどサンプル間で発現量に差がある事になります。

プロットの色にも意味があります。Scaffold DIA 内で定めた「Significant Level」、例えば FDR や FWER の閾値を超える条件を持つデータは緑、その基準未満で $p \leq 0.05$ のデータは黄色で表示 されます。この色分けは、Samples 画面の統計数値(p-value)表示列にも同じ条件が適用されています。

グラフの左側にはいくつかのオプションがあります。縦軸では補正後の有意なラインをグラフ内に表示 したり、横軸では比較サンプルの定量値が同じ(値が0になるライン)を表示したりする事ができます。 また「Multiselect Action」の項目には星を付けるオプションがあり、グラフ内で左クリックによる四角形 ドラッグ描写を行うと、その四角形に含まれるタンパク質に星印がつきます(Samples 画面で確認 できます)。

8-2-2. Quantitative Scatterplot

「Quantitative Scatterplot」では選択した2つの要素の定量値をプロットした散布図が表示されます (Fig.8-4)。



Fig.8-4 Visualize 画面 Quantitation タブ Scatterplot

グラフ左の設定で、x,y 両軸に置くサンプル/選択項目を変更したり、 y=x の直線や回帰直線も表示 させることができます。また「Multiselect Action」の項目には星を付けるオプションがあり、グラフ内で 左クリックによる四角形ドラッグ描写を行うと、その四角形に含まれるタンパク質に星印がつきます (Samples 画面で確認できます)。

8-2-3. Quantitative Trend Chart & CVs

Visualize 画面の左下には 3 つのタブ「Quantitative Trend Chart」、 「Pre/Post Normalization」、そして「Quantitative CVs」の図が存在します。

「Quantitative Trend Chart」では、フィルター適用後のタンパク質すべてを対象に、定量値の分布に 関する情報を様々なグラフ(箱ひげ図・棒グラフ・Trend Line・ヴァイオリン図)で表示しています。

「**Pre/Post Normalization**」タブでは、Normalization 前後の定量値について、箱ひげ図または ヴァイオリン図で比較する事ができます(Fig.8-5)。



Fig.8-5 Visualize 画面 Quantitation タブ Quantitative Trend Chart

「Quantitative CVs」には定量値の信頼性を評価する2つのグラフが同時に表示されています (Fig.8-6)。グラフは Samples で選択されている Summary 単位による区分が適用されます。2 種類の 表示のうち1つは CV50%の信頼区間ならびにその中央値の線です。以下図の場合、線と色分け区間で 表現されているグラフの方になります。もう1つはカーネル密度推定値を使った分布となります。



Fig.8-6 Visualize 画面 Quantitation タブ Quantitative CVs
8-2-4. GO Term Pie Chart, GO Connections Graph

GO 情報に関しては2つのタブから構成されています。「GO Term Pie Chart」と「GO Connections Graph」です。利用するためにはデータ取り込み後に「GO 情報の付与」(メニューの experiment -> Apply GO Terms)を実行する必要があります。GO 情報について詳細は本資料の 15章(P.108~)を ご覧ください。

GO Term Pie Chart は、付与された Gene Ontology 情報を数え上げ円グラフにしたもの(Fig.8-7) で、Biological Process, Cellular Component, Molecular Function それぞれの分類で描写させる ことができます。グラフの右上にグラフ描写に関するいくつかのオプションがあります。



Fig.8-7 Visualize 画面 Quantitation タブ GO Term Pie Chart

「GO Connections Graph」は、GO の項目とタンパク質を線で繋いだもので、GO で共通項の多い タンパク質を可視化する図です(Fig.8-8)。色のついた項目が GO の項目で、白い項目がタンパク質です。



Fig.8-8 Visualize 画面 Quantitation タブ GO Connections Graph

8-3. Principal Component Analysis タブ

PCA 解析[主成分分析]とは、データのばらつきに影響を及ぼす「主成分」を見つけ、データの特徴を抽出 する解析手法です。PCA 解析の詳細については、英文マニュアルの「Appendix I. How PCA is Performed in Scaffold DIA(P.195~)」 をご覧ください。

Principal Component Analysis タブは4つのパネルから構成されています(Fig.8-9)。



Fig.8-9 Visualize 画面 Principal Component Analysis タブ

- Overview (左上,8-3-1)
- Scree Plot (右上,8-3-2)
- Scores Plot (左下,8-3-3)
- Loadings Plot (右下,8-3-4)

です。次頁以降、各グラフについて詳しく説明します。

8-3-1. Overview

主成分間における値のプロットです(Fig.8-10)。グラフ内の点はサンプルを表しています。グラフを選択 した状態でプロットにカーソルを合わせると、主成分の組み合わせについて確認する事ができます。また 選択中のグラフに対応して左下の Scores Plot や右下の「Loadings Plot」が連動して表示変更 されます。右下の「Fewer Principal Components」「More Principal Components」ボタンを押す事で、 解析対象の主成分数を増減する事ができます。



Fig.8-10 Visualize 画面 Principal Component Analysis タブ Overview

8-3-2. Scree Plot

赤い線が各主成分の寄与率、青い線が寄与率の累積を表します(Fig.8-11)。



Fig.8-11 Visualize 画面 Principal Component Analysis タブ Scree Plot

8-3-3. Scores Plot

Overview で選択したグラフを拡大した、主成分間のスコア分布図です(Fig.8-12)。色のついた楕円は 該当エリア内に同じ属性値を持ったデータが入るとみなした時の N%信頼区間で、N の数字はグラフ左の オプション(Confidence)で変更することができます。



Fig.8-12 Visualize 画面 Principal Component Analysis タブ Scores Plot

8-3-4. Loading Plot

各タンパク質の主成分への寄与を表す分布図(Fig.8-13)で、絶対値が大きいほどその主成分への影響が 大きいことを示しています。原点から離れているタンパク質に注目してください。「Multiselect Action」で 気になるタンパク質に星印をつけ、Samples 画面で確認する事もできます。



Fig.8-13 Visualize 画面 Principal Component Analysis タブ Loadings Plot

8-4. Heatmap タブ

ヒートマップは2次元の表や行列で構成された複雑なデータセットを可視化する有力な方法のひとつ です。類似性の高いデータを近さの基準に従って空間的にまとめつつ行と列を並び替えると同時に、 各セルにあたる定量値について量の大小を表す色分けをして傾向をわかりやすくしています。

ヒートマップはデータ数が多すぎると意味が薄れるため、Scaffold DIA ではタンパク質数の上限を 1000 に定めています。そのため 1000 を超えているケースでヒートマップを適用したい場合、ユーザーが 事前に Samples 画面でデータ数を制限して 1000 未満にする必要があります。制限をする際には キーワード検索や GO 並び替えで項目をピックアップするとともに、それらに対して星印を付して絞り込み 表示に活用してください。(→P.51)

ヒートマップタブは2つの画面から構成されます。

- Heatmap Landscape : ヒートマップの全体図 (8-4-1)
- Heatmap Details pane: ヒートマップのうち全体図で選択した内容についてのみ、より詳しい 情報を表示した図 (8-4-2)

2つの画面のうちそれぞれ一方のみ、あるいは両方を表示するかどうかで3つの選択肢(ラジオボタン)があります(Fig.8-14)。



Fig.8-14 Visualize 画面 Heatmap タブ

8-4-1. Heatmap Landscape pane

行がタンパク質、列がサンプルに該当します。行・列ともに Samples 画面で選択した内容が反映されて います。行列どちらもクラスタリングの結果を表すデンドログラムが記載され、その結果を基にデータが 並び直されています。デンドログラム内のブロックをクリックするとその塊が選択されます。選択状態にあ る領域は灰色で表示されます。ブロックの内容について詳しい情報は、「Heatmap Details pane」にて 確認できるほか、ブロックの内部における相同性が表示されます(Fig.8-14)。 ヒートマップのデータの内容を数字の CSV 出力したい時、あるいはグラフを図として出力したい時、 画面内で右クリックしてそれぞれの操作を実行できます(Fig.8-15)。



Fig.8-15 Visualize 画面 Heatmap タブ、コンテキスト

8-4-2. Heatmap Details pane

Landscape で選択されたブロックについて、タンパク質の名称や Description なども含めたより詳しい 情報を確認する事ができます(Fig.8-16)。



Fig.8-16 Visualize 画面 Heatmap タブ、details info.

9. Analysis 画面

9-1. Analysis 画面のグラフについて

Analysis 画面では TIC と補正前後の LC 保持時間の誤差を確認する事ができます(Fig.9-1)。2つの グラフを上下並べて表示しています。横軸は共に(Reference) Retention Time (min) です。



Fig.9-1 Analysis 画面

上のグラフは、縦軸が TIC(Total Ion Current)です。TIC は MS1/ペプチドレベルのデータから計算 しています。x 軸である保持時間についてはアライメントの前と後、どちらでも表示可能です。グラフ上に あるチェックボックスにより表示の切り替えができます。

下のグラフは元データの RT とアライメント後の値との誤差(元データとリファレンスの差)を表して います。

どちらも凡例ならびに実線の上にカーソルを合わせると強調表示され、凡例との連動によりどのデータ がどの曲線か、わかりやすくなります。

10. Publish 画面

10-1. Publish 画面について

Publish 画面は、検索パラメーター覧を確認したり、論文や Repository site で記載する Method の 文章にまとめた内容を表示したりする「**Experiment Methods**」タブと、SQL コマンドを実施して テーブル(ファイル)を出力する「**SQL Report**」タブから構成されています(Fig.10-1)。



Fig.10-1 Publish 画面

以降、Experiment Method タブ (10-2)、SQL Report タブ (10-3) についてご紹介します。

10-2. Experiment Methods タブ

Experiment Methods タブでは表示内容が大きく右側と左側に分かれています。 左側は Scaffold DIA で指定したパラメータが、ツリー上にまとめられて表示されています (Fig.10-2)。

Experiment Methods SQL Report	
Search	م
Search Library	demo2.elib
Processing Directory	.¥work
Protein Sequence Database	uniprot-swissprot-human.fasta
	efalse
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	[630]
Peptide Charge	[23]
- 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
i → Version	
Scaffold DIA	1.0.0
Encyclopedia	0.6.12
ProteoWizard	3.0.11748
Percolator	3.01 nightly=13=655e4c7=dirty

Fig.10-2 Experimental Method 左側

一方右側には、論文やレポートの Method 部分を意識した、解析手法についてまとめられた文章が表示 されます (Fig.10-3)。

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.elib* 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 $[2 \cdot 3]$ and length in the range $[6 \cdot 30]$ were considered. Peptides identified in each sample were filtered by Percolator (3.01. nightly-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 (3.01.nightly-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 quantification.

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, ParkinsonsUK-UCL, NTNU_SB, LIFEdb, FlyBase, BHF-UCL, HGNC, MGI, SYSCILIA_CCNET, CACAO, AgBase, PINC, ARUK-UCL, CAFA, MTBBASE, Alzheimers_University_of_Toronto, WormBase, GDB, 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., Amode, D., Ruderman, D.L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., Hoff, K., Kessner, D., Tasman, N., Shulman, N., Frewen, B., Baker, T.A., Brusniak, M.-Y., Paulse, C., Creasy, D., Flashner, L., Kani, K., Moulding, C., Seymour, S.L., Nuwaysir, L.M., Lefebvre, B., Kuhmann, F., Roark, J., Rainer, P., Detlev, S., Hemenway, T., Huhmer, A., Langridge, J., Connolly, B., Chadick, 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 Lukas Käll Jasse Canterbury Jasen Wosten William Stafford Noble and Wichael J. MacCoss Export Supplementary Data

Fig.10-3 Experimental Method 右側

これらの文章は、クリップボードにテキストとして保存させたり、CSV形式でデータを出力させたりする ことができます。 Scaffold DIA から、オープンソースのリレーショナルデータベース管理システム:SQLite のファイルを 出力する事ができます。SQL Report タブでは SQLite の GUI が使用可能です。ユーザーは SQLite の コマンドを実行する事で、データを好きなように集め、出力する事ができます。

🔐 Scaffold DIA Viewer - Demo2_HeLa_insulin_6_files_30_proteins.sdia	- • •
File Edit View Experiment Export Help	
Image: Summarization Condition Image: Protein FDR 10% FDR Image: Min # Peptides 2 Image: Summarization Condition Filters Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition Image: Summarization Condition <th></th>	
Experiment Methods SQL Report	
Saved Queries \$QL Very mark \$Sipply names of all tables in database */ Sylect Name EPON Solice Market WHEPE Twee*table*	
Samples	
Proteins	
Visualize Save query as Newr Save Run Query	
Results =	Ę
Publish	
Proteins 0.0% FDR (statuned) 20 Targete	
0 Decoys Peptides: 00% FDR (stained) 1002 Targets	Export Results
0 Decoye	

Fig.10-4 SQL Report タブ

出力される実験ファイル、sdia(sfdb)のスキーマについては、英文マニュアルの「Appendix A. Structure of Scaffold DIA files」(P.172~)をご覧ください。

11.定量の手法と検定機能

Scaffold DIA ではフラグメントピークの強度情報を基にしたラベルフリーの定量解析をサポートして います。値はサンプル間で Normalization する事ができ、因子レベルや治療法の組み合わせによる発現の 違いを確認するため、様々な検定を適用する事ができます。なお基本的に Normalization は実験間の 比較レベルにおいてタンパク質総量が同じである前提の下で行われており、総量が異なるケースにおいて Normalization を行う事は適切ではありません。

11-1. 定量の手法について

11-1-1. Peptide Counting

Scaffold DIA では定量解析の一環として、以下の2つのペプチド数を Samples 画面で表示可能です。 Display type を通じて表示内容を選択する事ができます。

•Total Quantified Peptide Count

EncyclopeDIA によって同定され定量に使われているペプチドについて、シェアペプチドかどうかに 関わらずタンパク質毎にカウントした数字です。

•Exclusive Quantified Peptide Count

EncyclopeDIA によって同定され定量に使われているペプチドについて、シェアペプチドを除きユニークペプチドのみをカウントした数字です。

もう1つ重要な点として、これらの表示におけるペプチドカウント数は正規化されているのでご注意 ください。Samples テーブルに表示される値を調整し、各列のユニークペプチドの合計数が同じになる ようにします。選択したまとめのレベルに応じ、Samples テーブルで表示される列の数は変動し、カウント も変動します。

11-1-2. ラベルフリー定量

■ ペプチドの定量、タンパク質の定量

Scaffold DIA では XIC におけるフラグメントイオンのピーク強度をベースとしたラベルフリー定量を 採用しています。プリカーサーイオンのピーク強度は定量計算そのものには利用されていません。

Scaffold DIA では定量の対象とするフラグメントの選定において、他のフラグメントイオンの干渉の 影響ができるだけないきれいなピークを選ぶため、干渉度合いを表すスコアによって評価しています。 選ばれたフラグメントピークについて、バックグラウンドの影響も加味して差し引いた面積強度を算出し、 その和がペプチドの定量値となります。場合によっては、検索したスペクトルライブラリのピークとしては含 まれていないフラグメントが定量されることがあります。これは、EncyclopeDIA がペプチドの理論的な フラグメンテーションパターンも考慮しており、理論的なスペクトルから追加のフラグメントイオンが同定 された場合、それらを選択して定量計算に加える事が可能である、という事です。 選択されたペプチドについては、全サンプルで見た干渉の程度や、ユーザーが指定したパラメータに合致 するかを判定したスコアを元に取捨選択され、タンパク質の定量値計算に利用されます。

Normalization

Normalization はユーザーが Experimental Design ダイアログで指定した Technical Replicates レベルで行われます(設定前は全データを1つのグループとして見ています)。

まず intensity の和が Log10 に変換され、Experimental Design で TR [Technical Replicate] の グループと認定されたデータ間で、各サンプルの四分位数における数字が全体平均の四分位数における 数字と一致するように調整され、四分位数間は直線で近似します。調整後、Log10 から通常の数字に戻し、 TR レベルで和を計算して結果画面に表示されます。

Normalization は画面にあるチェックボタンで実施する/しない を切り替える事ができます。

■ 欠損値

欠損値は Scaffold DIA の様々な計算に影響を与えるため、適切な処理をする必要があります。値を より高いレベルで要約したり、統計的なテストや主成分分析を行ったりするため、測定値が得られなかった 場合に適切な値を入力する必要があります。

DIA 分析における欠損値はサンプル中にペプチドが存在しないか、または非常に低い強度で存在するかのいずれかの結果であると仮定し、Scaffold DIA では「Quantile Regression for Imputation of Left-Censored data 法*」(QRILC)を使用した値の代入を行っています。

* Cosmin Lazar (2015): imputeLCMD v2.0". R package version 2.0

QRILC は, 観測値(非欠損値)と欠損値および観測値の数をもとに想定した正規分布から, 欠損値が どの観測値よりも小さいと仮定して値を推定します。

また欠損率が 50%を超えるケースでデータをまとめる場合、QRILC で推定した平均値を使用します。 なおその場合部分的に欠損しているものとして扱い、値を括弧で囲んで表示します。

欠損値の割合が全体データ数に対して X%の場合、推定値は推定分布の X パーセンタイル以下になる ようにします。

QRILC の参照実装を提供する R パッケージ "imputeLCMD v2.0 "もありますので、そちらの文章も 併せてご参照ください。

(<u>https://www.rdocumentation.org/packages/imputeLCMD/versions/2.0/topics/impute.QRI</u>LC).

また英文マニュアルの「Appendix.D Missing Value」(P.178~)にも説明がございます。

[次頁に続きます]

11-2. 検定

Scaffold DIA ではいくつかの検定機能を有しています。検定機能を実行するためには、Experimental Design の設定が必要となります。検定ならびに Experimental Design の設定は、

「Configure Sample Organization and Statistical Analysis」ダイアログにて行います。この ダイアログを表示させるには以下 4 通りの操作があります。

- 「Organize」View の画面左下にある「Configure Experimental Design and Statistical Analysis」ボタンをクリック
- menu の「Experiment」->「Quantitative analysis」を選択
- ツールバーにある「Quantitative analysis」アイコンをクリック
- ツールバーの「Summarization」選択肢内にある「Edit」をクリック

選択時に開くダイアログは、「Sample Hierarchy」と「Statistical Analysis」2つのタブで構成されています。

Experimental Design で選択可能な項目の内容については「**5-6. 選択可能な 3 種類の Experimental Design について**(P.41~)」を、選択内容の設定については「**5-7. Experimental Design の設定**(P.44~)」をご覧ください。

検定を実行すると、Samples 欄に検定に関する数字を表示する列が現れます。選択した検定によって 表示される数字は異なります(p-value や q-value など)。セルは、有意で設定した Error rate を超えて いる場合は緑色で、また有意だが Error rate の設定を超えていない場合は<mark>黄色</mark>に色塗りされ、一方 有意でない場合は無色となります。

以下に、各データ構成タイプで選択可能な検定についてその内容を説明します。主な観点は「サンプル数」 と「パラメトリック/ノンパラメトリック」となります。

11-2-1. Basic Design で実施可能な検定

• t-test / ANOVA

取り扱うグループまたはカテゴリー間の平均値の等質性を評価する検定で、カテゴリー間に差があるか どうかを示します。両側検定を行います。3 群以上が ANOVA で、解析対象が 2 群しかない場合 t 検定と なります。

• Permutation Test

2 群あるいは 3 群以上のグループ比較で有意な差があるかを検定できる手法です。ベースは F 検定 ですが、群間のランダムなデータを入れ替えて F 値を計算し続けます。Scaffold DIA では 10000 回の 計算を行い、入れ替えた際に入れ替える前の F 値を超えた回数を 10000(試行回数)で割った値を p-value とします。

• Mann-Whitney U Test

独立した観測値の2つの標本のうち、一方の値が他方の値よりも大きい傾向にあるかどうかを評価する ノンパラメトリック検定です。2つの中央値が等しいかどうかに依存せず、2つの標本データの順位を使用 します。比較サンプル数がきっちり2つでないと使用できません。

• Kruskal-Wallis Test

順位別一元配置分散分析とも言え、標本が同じ分布に由来するかどうかを検定するノンパラメトリックな 手法です。しかし同じ分布の形とスケールである事を前提としています(中央値が同じである事を求めてい ません)。独立した3つ以上のサンプルに適用可能です。帰無仮説は「サンプルが由来する集団が同じ 中央値を持つ」であり、検定が有意な時、少なくとも1つのサンプルが他のサンプルとは異なる事を意味 します。この検定は差がどこで発生したのか、どれだけの差が発生したのかを特定する事はできません。

11-2-2. Repeated Measures で実施可能な検定

Repeated Measures ANOVA/Paired t-test

反復測定分散分析(rANOVA)はパラメトリックな手法です。3 つ以上の条件の違い(経時変化など)を 考慮した繰り返し測定が行われている場合、それらの測定の母平均が異なるかどうかを評価しています。 Paired-t 検定も同様ですがこちらはサンプルが2つの時にとられる手法です。正規分布でカテゴリー間 の分散が等しい場合に適用可能です。

• Wilcoxon Signed-rank test

ノンパラメトリックな手法でサンプル数は2つの時に限定されます。繰り返し測定の母平均の順位が 異なるかどうかを評価しており、2つのサンプルが独立で同一の分布であることを前提としています。 パラメトリックな手法である Paired t-test が適用できない際にご利用を検討してください。

• Friedman test

ノンパラメトリックな手法で3つ以上のサンプルがある時に適用する手法で、母平均順位が異なるか どうかを評価しています。サンプルが独立で同一の分布であることを前提としています。パラメトリックな 手法である rANOVA の代替方法であり、分析データが正規分布でない場合に適用してください。

11-2-3. Two-way で実施可能な検定

Two-way ANOVA

二元配置分散分析では2つの独立変数がタンパク質の量に与える影響を評価します。それぞれの独立 変数の主効果を測定し、さらに変数間に有意な相互作用があるかどうかも判定します。

Randomized Block ANOVA

Secondary 分析カテゴリー(この場合はブロック因子)の効果をコントロールしながら、Primary 分析 カテゴリーの効果を評価します。Secondary 分析カテゴリーそのものの効果を評価するものでは ありませんが、ブロッキング効果を評価するオプションとして、Primary 分析カテゴリーと Secondary 分析カテゴリーとの間に有意な相互作用があるかどうかを評価する事ができます。

• Friedman test

ノンパラメトリックな手法で3つ以上の条件で測定が行われた時、母平均順位が異なるかどうかを評価 しています。サンプルが独立で同一の分布であることを前提としています。分析データが正規分布でない 場合に適用してください。

11-3. 多重検定補正

多重検定において第一種の過誤(帰無仮説が真であるにもかかわらず偽として棄却してしまう誤り)を 防ぐため、p値と調整し直す処理を施す事があります。シンプルな方法としては Bonferroni 補正という 手法があります。これは p を検定数 N で割った p/N を新たな基準として使う方法ですが、この基準では 厳しすぎて False Negative が増えすぎてしまうのが問題です。Scaffold DIA ではその点を考慮した2つ の補正方法を採用しています。

- 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 (FWER)

- BH法 (FDR)

■ FWER とホッホベルクのステップアップ&ホルムのステップダウン

Bonferroniの不等式をベースとした、単純な Bonferroni 補正より積極的に同定タンパク質を拾える 補正を行って Family-wise error rate (FWER)をコントロールする手法です。「ホッホベルクのステップア ップ」と「ホルムのステップダウン」の2つをサポートしています。

詳細については 英文マニュアルの「Appendix G. Controlling the Family-wise Error Rate」 (P.184~)をご覧ください。

■ FDRとBH法

Scaffold DIA では FDR (False Discovery Rate)の計算を、Benjamini-Hochberg 法で計算して います。FDR ペプチドやタンパク質で使用している FDR とは基本コンセプトが同じあるものの中身は 大きく異なりますのでご注意ください。

12. タンパク質のグループ化、クラスター化

12-1. 同定タンパク質のグループ化、クラスター化

Scaffold DIA で表示される同定タンパク質は、シェアペプチドの情報を基に類似タンパク質がグループ 化、クラスター化されます。クラスター化における基準について、menu の **experiment** -> **Protein Clustering**-> **Shared Evidence clustering** を選択する事によって現れるダイアログで変更可能です (Fig.12-1)。

Cluster Proteins with Shared Evidence	×			
How similar do you wish proteins to be for clustering? You may check multiple boxes to achieve nested clustering. The first box must be checked because this clustering method requires perfect shared evidence protein grouping.				
Warning: Using Any shared evidence could result in meaningless clusters.				
✓ Perfect shared evidence protein groups				
Moderate shared evidence clusters				
Any shared evidence clusters				
Help Apply Cance				

Fig.12-1 Cluster Proteins with Shared Evience ダイアログ

類似度合いの評価については、二つのタンパク質 A,B のペプチドシェア度合いを表す L(A,B)と、閾値の 入によって定義されています。詳細は英語マニュアル Apendix .Shared Evidence Clustering Algorithm (P.180~)をご覧ください。パラメーターL (A,B)は、以下のように計算されています。

$$L(A, B) = \frac{\left[\sum_{MS-Sample} \# \text{ Shared peptides in A and B in MS-Sample}\right]}{\left[\sum_{MS-Sample} \# \text{ peptides in either A or B in MS-Sample}\right]}$$

L(A,B) を入と比較して基準を満たす時、タンパク質がクラスター化されます。

ダイアログ内各項目の選択肢と入はそれぞれ次頁の通りです。

Perfect shared evidence protein groups

ス= 1 、すなわち L(A,B)=1 で A と B が same-set でないとグループ化されません。最低限の設定レベル として常にオンになっているため、グレイアウトして設定が変更できないようになっています。

Moderate shared evidence clusters

 $\lambda = 1/3$ を基準として L(A,B)が 1/3 以上の時グループ化されます。「Perfect shared evidence protein groups」と「Any shared evidence clusters」の項目の間に定められた設定です。

Any shared evidence clusters

λ=0 を基準値としており、L(A,B)が0より大きい時、すなわち1つでもシェアペプチドがあればクラスター 化されます。MASCOT などで採用されている基準と同等です。

なお、「Moderate shared evidence clusters」と「Any shared evidence clusters」の両方にチェックが入った状態の場合、サブクラスターが表示されます。大きなクラスターは Moderate shared evidence clusters の基準が、その中のサブクラスターは Any shared evidence clusters の基準が適用されます。

13. Report

13-1. Scaffold DIA で出力可能なファイルについて

Scaffold DIA では様々な内容のファイル出力が可能です。メニューバーの「Export」にまとめられています。

• Export Attributes File

サンプルデータの属性設定値である Attribute 設定、並びに raw データと attribute との関連を 記した Scaffold DIA 用のフォーマットのテキストファイルを出力します。Organize 画面の「**Export Attributes File...**」ボタンと同じです。出力したファイルは Organize 画面で再び取り込む事が できます。同じデータで異なる条件を適用する 再検索などに利用します。

• Export Samples Report to Excel

Samples 画面で表示される内容、タンパク質の定量値のレポートを CSV 形式(コンマ区切りの テキストファイル)で出力できます。出力前には Fig.13-1 のような Export する定量値を選択する ダイアログが現れます。Samples 画面では一つしか表示されない定量値ですが、出力時には複数選択 する事も可能です。Normalization の設定も個別に行う事ができます。

Export Samples Report	×
Display Type	Normalization 🕐
 Exclusive Intensity 	On
Total Intensity	Off
Log ₂ Fold Change (Exclusive Intensity)	On
Log ₂ Fold Change (Total Intensity)	Off
CV (Exclusive Intensity)	On
CV (Total Intensity)	Off
Exclusive Quantified Peptide Count	n/a
Total Quantified Peptide Count	n/a
Export Options	
Include Experiment Parameters	
Collate Columns by Sample	
(2) Exp	oort Cancel

Fig.13-1 Export samples ダイアログ

• Export Peptide Report to Excel

Proteins 画面の上、Peptide Pane の内容に対する情報が、1 タンパク質でなく Samples 画面で 表示されているすべてのタンパク質について CSV 形式で出力されます。

• Export Peptide Match Report to Excel

Proteins 画面の中段、Peptide Match Pane の内容に対する情報が、1 タンパク質でなく Samples 画面で表示されているすべてのタンパク質について CSV 形式(コンマ区切りのテキストファイル)で 出力されます。「Export Peptide Report to Excel」と異なりファイル情報などは出力されません。

• Export Publish Report to Excel

Publish 画面、「**Experiment Methods**」タブの左側に表示されているパラメータの一覧情報が CSV 形式(コンマ区切りのテキストファイル)で出力されます。

• Export Workflow

検索時のパラメータファイルをまとめた「Workflow」ファイルを出力できます。New ボタンを押して データ取り込みを行う際に現れるダイアログ内にある「Save Workflow as ...」ボタンと同じです。 保存したファイルは別の解析で import する事もできます。別の解析で同じ検索条件を複数回適用 したい場合に利用します。

• Run SQL Query for Export

Publish 画面の「SQL Report」タブ画面を開きます。

14. Prosit との連携によるライブラリフリーサーチ

Scaffold DIA では検索の対象とするライブラリとして、Prosit により計算されたピークリスト並びに 保持時間情報を含むデータ(各関連資料にて"DLIB"と記載しています)の使用をお勧めしています。 Prosit は Wilhelm と Kuster のグループによって開発されたディープラーニングアルゴリズムです。 Scaffold 上で配列データベース(FASTA)を指定する検索と比較すると保持時間やピーク強度といった 予測値を持つことにメリットがあり、また DDA ライブラリ検索と比較すると事前の DDA 検索を必要 とせず、DDA 結果にとらわれない(DDA で同定できなかったペプチドもターゲットとできる)点で メリットがあります。ただし DLIB 検索では修飾パターンとして現段階ではシステインのカルバミドメチ ル化(固定)しか考慮していない事にご注意ください。

この章では、Scaffold DIA で使用する DLIB の取得方法について以下2つをご案内いたします。

14-1. Proteome Software 社が公開している DLIB ファイルを利用する 14-2. 準備した FASTA ファイルを元に Prosit を利用して DLIB ファイルを作成する また取得した DLIB ファイルについて、

14-3. DLIB 並びに FASTA ファイルを Scaffold DIA にセットする方法

もご紹介します。

14-1. Proteome Software 社が公開している DLIB ファイルを利用する

Proteome Software 社にて、いくつかの生物種については事前に計算を行った DLIB ファイルと 対応する FASTA ファイルを公開しています。作成パラメータは Tab.14-1 の通りです。 <u>https://support.proteomesoftware.com/hc/en-us/articles/360035151172-Prosit-Derived-</u> <u>Spectral-Libraries-for-Scaffold-DIA-Searches</u>

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

Tab.14-1 Proteome Software 社で公開している DLIB ファイルの Library 作成パラメータ

2022 年 12 月末現在、以下の生物種について準備しています (表記は公開サイトの内容に準ずる, Fig.14-1)

Coronavirus (または plus Human), Arabidopsis thaliana, Caenorhabditis elegans, Danio rerio, Drosophila melanogaster, Escherichia coli(strain K-12), Homo sapiens, Mus musculus, Rattus norvegicus, Saccharomyces cerevisiae

利用したい生物種について、"DLIB"と"FASTA" を両方ともダウンロードしてください。 取得したファイルを Scaffold DIA で利用する方法については、「14-3.DLIB 並びに FASTA を Scaffold DIA にセットする方法 (P.106)」をご覧ください。



Fig.14-1 Proteome Software 社が公開している DLIB と FASTA

14-2. FASTA ファイルを元に Prosit を利用して DLIB ファイルを作成

Proteome Software 社で予め準備されている DLIB は対象の生物種が限られています。また FASTA エントリーの対象を Uniprot 内の reviewed (=Swissprot に相当)に限定しています。必要に応じて公開 されているもの以外の生物種を利用したり unreviewed (=TrEMBL に相当)を加えたいなどといった 調整が必要な時、さらにはそれらとは全く別の FASTA ファイルを利用したい時は、ご自身で DLIB を作成 することができます。作成の例として、Uniprot-reveiewed にて生物種「Oryza sativa」に絞り込んだ データベースの作成についてご説明いたします。

DLIB 作成のステップは以下の通りです。 14-2-1. FASTA ファイルを取得する (例:Uniprot サイト) 14-2-2. Scaffold DIA で Prosit 用 のインプットファイルを作成する 14-2-3. Prosit 公開サイトにてペプチド配列情報から計算を行う 14-2-4. Scaffold DIA で Prosit 出力ファイルと FASTA ファイルから DLIB ファイルを作成する

以降、順にご説明します。

14-2-1. FASTA ファイルを取得する(例:Uniprot サイト)

FASTA ファイルを準備します。FASTA はどこから取得したファイルでも使用可能ですが、Gene Ontology 情報に紐づけができるデータベースとして、Uniprot データベースの使用をお勧めしており、 ここでは Uniprot サイトからファイルを取得する例をご紹介します。

Uniprot に登録されているタンパク質が表示されます。

Uniprot サイト(<u>https://www.uniprot.org/</u>)にアクセスし、Proteins をクリックします(Fig.14-2,赤丸)。



Fig.14-2 Uniprot サイト

検索対象を reviewed,しっかりしたアノテーションがついているデータのみにしたい場合は「SwissProt」 あたりをクリックします(Fig.14-2,黄色枠)。Unreviewed、幅広いデータのみを対象としたい場合は 「Unreviewed」のあたり((Fig.14-2,赤色枠))をクリックします。選択しない場合は両方のデータベースが 選ばれている状態(Uniprot データベース)です。

	ide search ID mappir	ng Sl	PARQL UniProtKB	•	Advanced List
Status Reviewed (Swiss-Prot) (568,363) Unreviewed (TrEMBL)	UniProtK BLAST Align Ma	B :	230,496,50 ⊥ Download ↔	3 results Add View: Cards⊙ Table €) 🖉 Customize columns 🖂
(229,928,140)	Entry .		Entry Name 🔺	Protein Names 🔺	Gene Names 🔺
Popular organisms Human (205,788)	A0A0C5B5G6	2	MOTSC_HUMAN	Mitochondrial-derived peptide MOTS-c[]	MT-RNR1
Rice (149,191)	□ A0A1B0GTW7	8	CIROP_HUMAN	Ciliated left-right organizer metallopeptidase[]	CIROP, LMLN2
Zebrafish (101,302)	A0JNW5	2	UH1BL_HUMAN	UHRF1-binding protein 1- like[]	UHRF1BP1L, KIAA0701, SHIP164
Mouse (86,431) Taxonomy	□ A0JP26	8	POTB3_HUMAN	POTE ankyrin domain family member B3	POTEB3
Filter by taxonomy	A0PK11	-	CLRN2_HUMAN	Clarin-2	CLRN2
Proteins with 3D structure (58,487)	□ A1A4S6	2	RHG10_HUMAN	Rho GTPase-activating protein 10[]	ARHGAP10, GRAF2

Fig.14-2 UniprotKB protein table

また、左フレームにある「Popular organisms」で表記されている生物種に選択したい内容をクリック すると、該当生物種に限定されたタンパク質のみが表示されます。目的の生物種がリストにない場合、 Filter by taxonomy をクリックします(Fig.14-3 緑色枠)。

Popular organisms
Human (205,788)
Rice (149,191)
A. thaliana (136,466)
Zebrafish (101,302)
Mouse (86,431)
Taxonomy
Filter by taxonomy

Fig.14-3 UniProtKB 左フレーム拡大

生物種名を入力すると、候補が現れます。適切な内容を選択して Search ボタンを押すことで生物種の 絞り込みが行われます(Fig.14-4)。

Advanced Search		>	< A
Searching in			^
UniProtKB			
	Taxonomy [OC]		
▼ Taxonomy [OC] 👻	Oryza sativa X	Remove	sto
	Oryza sativa subsp. japonica (Japanese rice/O. sativa/Japonica rice/Rice) [3994	1	Na
Add Field	Oryza sativa (Rice) [4530]	Search	٩R
i) Type * in the search box to search for all	• Oryza sativa endornavirus (OsEV) [362693]		
	Oryza sativa aus subgroup [1736659]		, L
	Oryza sativa aromatic subgroup [1736658]		1B
	Oryza sativa subsp. indica (Rice) [39946])70
	Oryza sativa Indica Group x Oryza sativa Japonica Group [1050722]		33
	Orviza sativa Japonica Group x Orviza sativa Indica Group [1080340]		

Fig.14-4 taxonomy filter で適切な生物種を選択

生物種別のタンパク質リストに変わります。変更後のエントリー数は生物種別のタンパク質リストに 変わります。変更後のエントリー数は左側の"Status"で()内に表示されている数字や一覧の上の数字で 確認ができます(Fig.14-5 赤色枠)。使用に適切かどうかは、想定される生物種の遺伝子数と比較 しながらご検討ください。

UniProt BLAST Align Peptide search ID mapping SPARQL UniProtKB (taxonomy_id:4530)							
Status Reviewed (Swiss-Prot)	UniProt	(B 195,603 r	esults				
 Unreviewed (TrEMBL) 	BLAST Align Map IDs 土 Download 🏛 Add View: Cards 🔿 Table 🖲 💆						
(190.483)	Entry 🔺	Entry Name 🔺	Protein Names 🔺	Gene			
Popular organisms Rice (149,191)	A0A0P0X9Z7	CWZF7_ORYSJ	Cysteine-tryptophan domain-containing zinc finger protein 7[]	CWZ Os07 LOC			
Taxonomy	□ B2ZX90	FAS1_ORYSJ	Chromatin assembly factor 1	FSM,			
4530 ×		_	subunit FSM[]	LOC			
Filter by taxonomy				P069 P069			
Proteins with	C7IW64	ROS1A_ORYSJ	Protein ROS1A[]	ROS			

Fig.14-5 Filter 適用後の UniProtKB

内容を確認後、ファイル取得を行います。画面上部の「Download」(Fig.14-6)をクリックし、Format として "FASTA(canonical)"を選択します。Compressed / Uncompressed はどちらでも結構です が、 圧縮されている形式が Windows 標準では解凍できない"gz"ですので、それを解凍できる アプリケーションを持っていない方は"Uncompressed"を選択して下さい。

全ての項目を選択後、「Download」ボタンを押すとファイル取得(ダウンロード)が実行されます。

UniProtKB 195,603 results					
BL	AST Align M	ap IDs	土 Download	☆ Add View: Cards ○ Table	e 💿 🗡 Cu
	Entry 🔺		Entry Name 🔺	Protein Names 🔺	Gene Na
	A0A0P0X9Z7	2	CWZF7_ORYSJ	Cysteine-tryptophan domain-containing zinc finger protein 7[]	CWZF7, Os07g06 LOC_Os0
	B2ZX90	2	FAS1_ORYSJ	Chromatin assembly factor 1	FSM, Os

Fig.14-6 UniProtKB の" Download"

Download		×
O Download selected (0)		
Download all (195,603)		
Format		
FASTA (canonical)		v
Compressed		
• Yes		
○ No		
	Generate URL for API Preview 10	Cancel Download

Fig.14-7 UniProtKB の"Download"オプション

[参考] 生物種情報を特定するために NCBIの Taxonomy サイトが利用できます。

上記操作で「Popular organisms」で生物種を指定する際、名称の類似性などからターゲットとする対象 を限定するのが難しいケースがあります。Popular organisms の候補選択肢として現れる文字列の後ろ に表示されている数字は NCBI の TaxID です。この数字について、NCBI の Taxonomy サイトにて 事前に調べておくと候補の中から自らのターゲットとなる生物種を選びやすくなります。以下その操作に ついてご紹介します。

NCBIの Taxonomy サイト https://www.ncbi.nlm.nih.gov/taxonomy ヘアクセスします。 画面上部の検索枠で、生物種名による検索を行います。キーワードを入れて「Search」ボタンを押します (Fig.14-8)。



Fig.14-8 NCBI Taxonomy 検索欄

検索結果が表示されます。目的の生物種名のハイパーリンクをクリックします(Fig.14-9)。

NIH National Library of Medicine National Center for Biotechnology Information	Log in	
Taxonomy Oryza sativa Create alert Limits		Search Help
Display Settings: - Summary <u>Oryza sativa</u> (Asian cunivated rice), species, monocots	Send to: 🗸	Related information
Nucleotide Protein		Assembly

Fig.14-9 NCBI Taxonomy 検索結果

遷移画面にて、該当生物種の Lineage(系統)が表示されます。選択項目が正しい階層・属・種・株である かを確認しながら、ターゲットとする項目のハイパーリンクをクリックします(Fig.14-10)。



Fig.14-10 NCBI Taxonomy 検索結果 Lineage 表示

遷移画面にて、NCBIの Taxonomy ID をチェックします。以下図(Fig.14-11)であれば「4530」です。 この数字を記録しておき、Uniprot で生物種を特定する際の識別情報として利用してください。

Search for PubMed Nucleolide Protein Genome Search for as complete name v Ick Display 3 levels using filter: none v	Ser PMC Go Clear	Taxonomy	3ioCollections
<u>Oryza sativa</u>	Entrez	records	
Taxonomy ID: 4530 (for eferences in articles please use NCBI:txid4530)	Database name	Subtree links	Direct links
current name	Nucleotide	2,289,219	322,013
Oryza sativa L., 1753	Protein	442,430	<u>62,025</u>
Genbank common name: Asian cultivated rice	Structure	<u>269</u>	<u>70</u>
NCBI BLAST name: monocots	Genome	1	1
Rank: species	Popset	1,227	<u>1,077</u>
Genetic code: <u>Translation table 1 (Standard)</u>	Conserved Domains	12	5
Mitochondrial genetic code: <u>Translation table 1 (Standard)</u>	GEO Datasets	21,846	16,106
Plastic genetic code: <u>Iranslation table II (Bacterial, Archaeal and</u> Plant Plastid)	PubMed Central	32,836	32,836
Other names:	Gene	95,351	149
_common name(s)	HomoloGene	9,787	<u>9,787</u>
red rice, rice	SRA Experiments	98,656	72,606

Fig.14-11 NCBI Taxonomy 検索結果 各生物種の情報表示画面

14-2-2. Scaffold DIA で Prosit 用のインプットファイルを作成する

FASTA を取得したら Scaffold DIA を起動し、menu の Tools -> Convert FASTA to Prosit CSV を選択します(Fig.14-12)。



Fig.14-12 Tools -> Convert FASTA to Prosit CSV

FASTA を選択し(Fig.14-13)、パラメータはこだわりがない限り基本的にはデフォルト設定で「OK」 ボタンを押してください。FASTA ファイルと同じ場所に CSV ファイルが作成されます。

Convert FASTA to Prosit CSV				×
Parameters: "Cite" This function will in s create an input file for Searle et al. 2020.	<i>silic</i> o digest peptides fr or Prosit. If you use thi	rom your FAS s feature, plea	ΓA and ise cite	
FASTA: uniprot-organism_Ory	za+sativa+[4530]fasta		Edit	
Charge range:	2 🗘 to	3 🗘		
Maximum Missed Cleavage:			1	\diamond
m/z range:	396.4 🗘 to 1,0	02.7 🗘		
Default NCE:			33	\Diamond
Default Charge:			3	\Diamond
	OK Cancel			

Fig.14-13 Convert FASTA to Prosit CSV ダイアログ

*ダイアログ内にも記載がありますように、この機能を利用され論文公開・学会発表をされた場合、リンクが貼られている論文(Searle et.al 2020)の引用をお願いしています。

14-2-3. Prosit 公開サイトにてペプチド配列情報から計算を行う

Prosit のサイトヘアクセスし、「SPECTRAL LIBRARY」をクリックします(Fig.14-14 赤枠)。 https://www.proteomicsdb.org/prosit/

We now offer two new Pros fragment intensities predict prediction (Prosit_TMT_irt HCD fragmentation method We assume all the sequenc modification explicitly in you	it TMT models that will soon ion (Prosit_TMT_intensity :_ 2021) . The intensity mod Is but you need to add fragn es are fully labeled and you ur input files.	n be published. One is for 2021) and the other is for iRT el works with both CID and nentation column to the input. don't need to add the tmt
P··· PREDICT	LIBRARIES FAQ	STATUS 🗭 ?
Prosit offers high quality MS2 pre prediction. Prosit is part of the Pr trained on the project's high qual research, please cite "Gessulat, S	dicted spectra for any orgar oteomeTools (<u>www.proteon</u> ity synthetic dataset. When chmidt et al. 2019" <u>DOI 10</u>	nism and protease as well as iRT netools.org/) project and was using Prosit is helpful for your 1038/s41592-019-0426-7.
CE CALIBRATION	SPECTRAL LIBRARY	RESCORING
This task estimates the optimal	collision energy (CE) based	on a given search result. You

Fig.14-14 Prosit サイト

Scaffold DIA で作成した Prosit 用の CSV を指定するため「Next」ボタンを押します(Fig.14-15)。

•	CSV FASTA (comming soon)				
ſ	CSV Format				
	modified_sequence collisio	n_energy precurs	sor_charge		fragmentatio
	M(ox)CSDSDGLAPPQHLIR	15	2	т оlly	HC
	EMPQSDPSVEPPLSQETFSDLWK	28	2	r TM els (HC
	TCPVQLWVDSTPPPGTR	35	3	nod Fo	CI
	QSQHM(ox)TEVVR	35	5		CI
	 modified_sequence Use up indicate oxidized Methionin restricted to the range of 7 with carbamidomethylation amino acids. collision_energy Use interesting precursor_charge Use interesting 	oper case lette e with "M(ox) to 30. Each C n. Prosit does n eger values fro eger values fro o or CID, Use n	ers in the o ". Sequence C is treated not suppo om 10 and om 1 to 6. upper case	column ce leng d as Cy rt U or d 50. e letter	a and th is ysteine O as s. *
	*Only for TMT model				

Fig.14-15 「Settings」選択内容

遷移画面にて画面左側のボタンをクリックし、Scaffold DIA で作成した CSV ファイルを指定してから 「Next」ボタンを押してください(Fig.14-16)。

• • *Oi	precursor_charge Use integer values from 1 to 6. fragmentation Either HCD or CID, Use upper case letters.*
٩	uniprot-organismOryza+sativa+[4530]fasta.trypsin.z3_nce33. ×
	< BACK NEXT >

Fig.14-16 Upload Files 選択項目

続いてピーク強度並びに保持時間の予測モデルの選択を行います。適切な内容を選び「Next」ボタンを 押します。各モデルについての詳細は、論文<u>https://www.nature.com/articles/s41467-021-23713-9</u> などをご参照ください(概ね名前から内容をご判断いただいて問題ないと思います)。

0	Settings Indicate collision energy, the maximum number of missed cleavages, and number of oxidized methionines per peptide.
0	Upload Files Fasta or CSV with list of peptides
3	Model Select intensity and iRT model for prediction
	Intensity prediction model
	O Prosit_2019_intensity_hcd
	O Prosit_2020_intensity_preview
	Prosit_2020_intensity_hcd
	O Prosit_2020_intensity_cid
	O Prosit_TMT_intensity_2021
	iRT prediction model
	Prosit_2019_irt
	O Prosit_TMT_irt_2021
	< BACK NEXT >

Fig.14-17 Model 選択項目

出力のフォーマットを選びます。ここでは"Generic text"を選択して下さい(Fig.14-18)。 選択後"submit"ボタンを押すことで計算が実行されます。

*フォーマットを.MSP にして、scaffold の変換で .MSP→ELIB を選択した場合も、作成後のファイルは目的の DLIB と同じものができます。何らかのエラーなどにより以下ご案内する方法で DLIB が作成されない場合があ りましたらそちらの迂回策をお試しください。



Fig.14-18 Model 選択項目

計算には時間がかかります。Task の ID が表示されますが、この TaskID の情報をどこかに残しておく と、万が一ブラウザを閉じてしまった場合でも後でアクセスすることが可能です。 例えば以下のように Task ID が示されている場合、

	LIBRARIES	FAQ	STATUS	Ģ
Prosit offers high quality as well as iRT prediction. (<u>www.proteometools.org</u> quality synthetic dataset cite "Gessulat, Schmidt e	MS2 predicted spa . Prosit is part of th g/) project and wa . When using Pros et al. 2019" <u>DOI 10</u>	ectra for any or he ProteomeTo is trained on th it is helpful for 0.1038/s41592	rganism and pr ols e project's hig your research <u>-019-0426-7</u> .	rotease h , please
Task D907DA69F	AC698CA2D	23D4BDFB	BC5144	
This task is in progress. T depending on system load check back later. You can Resubmitting tasks will no	asks may take sev d. Please note dow download the resu bt lead to faster re	veral hours for t vn your Task ID ults here upon	full proteomes) or save this U completion.	JRL to

Fig.14-19 Prosit サイト Task ID 表示

URL は以下の通りとなります(後ろの部分がタスクの ID です)。

https://www.proteomicsdb.org/prosit/task/D907DA69FAC698CA2D23D4BDFBBC5144

計算が終わると以下のようにファイル取得が可能となります。「Download」ボタンを押してください。 ファイルは zip 形式で圧縮されていますので、(windows なら)右クリック→すべて展開、で解凍できます。



Fig.14-20 Prosit サイト Task ID 表示

14-2-4. Scaffold DIA で Prosit 出力ファイルと FASTA ファイルから DLIB ファイルを作 成する

*以降14章では図の番号を省略します。

Scaffold DIA を起動し、menuの Tools -> "Convert Prosit/Spectronaut CSV to DLIB"を 選択します。



時間がかかる可能性がある旨警告するダイアログが現れます。「OK」ボタンを押してください。



Prosit から得た CSV ファイルと、Uniprot サイトから取得した FASTA ファイルを選択して OK ボタン を押します。

Convert Prosit/Spectronaut CSV to Library	×
Parameters:	
Spectronaut CSV/XLS:myPrositLib.csv	Edit
FASTA: uniprot-organism_Oryza+sativa+[4530]_,fasta	Edit
OK Cancel	

ファイル作成が開始します。以下ダイアログが消えるとファイル作成が完了しており、Prosit から出力された CSV ファイルと同じ場所に DLIB ファイルが作成されています(拡張子 :_dlib)。

🛃 Please wait	×
Reading Spectronaut CSV File	

14-3. DLIB 並びに FASTA ファイルを Scaffold DIA にセットする方法

Menu の File -> Open Library manager またはアイコンなどで Library manager を 起動します。

s 🔊	caffol	d DIA					
<u>F</u> ile	<u>E</u> dit	View	Experiment	E <u>x</u> port	Tools	<u>H</u> elp	
			= o	ħ₽	Summ	arization:	
Filte	ers		Open Library	Manager			
	Show	Hidden		Nam	e/Acces	sion	۶

"Add Library" ボタンを押します。

Library Manager		×
Name	Status	FASTA
BosTaurus_myPrositLib.dlib	Ready	uniprot-reviewed_yes+AND+orga
whole-cell_IP_lib.blib	Ready	Human_Uniprot_Sprt_Trembl_Isof
BSA_67_myPrositLib.dlib	Ready	uniprot-bsa-filtered-reviewed_yes
arabidopsis_thaliana_prosit_gener	Ready	Arabidopsis thaliana_TAIR10.fasta
Add Library	Remove Library Create L	Library Download OK

準備した DLIB ファイルを選択します。続いて、新たに表示された DLIB ファイルの行を選択した状態で 右クリック→ "Associate with FASTA"を選択します。

ivame	Status	FASTA
OryzaeDB.dlib	Ready	
BosTaurus_myPrositLib.dlib	Keady	uniprot-reviewed_yes+AND+
whole-cell_IP_lib.blib	Ready	Human_Uniprot_Sprt_Trembl
BSA_67_myPrositLib.dlib	Ready	uniprot-bsa-filtered-reviewed
arabidopsis_thaliana_prosit_g	Ready	Arabidopsis thaliana_TAIR10.

DLIB と対になっている FASTA (あるいは Prosit で DLIB 作成時、最初に取得した FASTA)を 選びます。

🛃 開く							\times
Look In:	materialDLIBmaking	×	Û	ŧ	lin;	::	:=
📒 uniprot-o	organism_Oryza+sativa+[4530]fa	asta					
File <u>N</u> ame:	uniprot-organism_Oryza+sativa	+[45	30]1	fasta			
Files of <u>Type</u> :	FASTA						\sim
			開く			取消	

Library Manager に追加 DLIB がセットされました。

🚛 Library Manager		×
Name	Status	FASTA
OryzaeDB.dlib	Ready	uniprot-organism_Oryza+sat
BosTaurus_myPrositLib.dlib	Ready	uniprot-reviewed_yes+AND+
whole-cell_IP_lib.blib	Ready	Human_Uniprot_Sprt_Trembl
BSA_67_myPrositLib.dlib	Ready	uniprot-bsa-filtered-reviewed
arabidopsis_thaliana_prosit_g	Ready	Arabidopsis thaliana_TAIR10.f
Add Library Remov	ve Library Create Library	Download OK

検索時に Library manager をご利用頂く事で、DLIB と FASTA の選択がスムーズになります。

15. Gene Ontology

Scaffold DIA では、各タンパク質の Gene Ontology の情報を結果画面に表示させる事が出来ます。 また同定タンパク質リストの各タンパク質にアノテーションされた Gene Ontology 情報をまとめて グラフ表示させる事ができます。Gene Ontology 情報を適用させるためには、事前に関連ファイルを ダウンロードし Scaffold DIA に登録させる必要があります。

15-1. Gene Ontology ファイルを Scaffold DIA にセットする方法

Edit -> **Edit GO term options** で「GO Term Configuration」ダイアログを開きます。 「**GO Annotations**」タブをクリックして画面を開きます(Fig15-1)。

GO Term Configuration			×
Displayed GO Terms GO Annotations			
The GO Annotation database currently contains annotations from 1 files.		search	۶
	Name		₽.
goa_mouse.gaf.gz			
Im	port <u>A</u> nnotations		
<u>ℓ</u> elp <u>H</u> elp		<u></u> K	

Fig.15-1 GO Annotations ダイアログ、GO Annotations タブ

GO Annotations タブでは、現在 Scaffold DIA でセットされている GO ファイルが表示されます。 新たな項目を加えるため、画面下部の「**Import Annotations**」ボタンをクリックすると、新たなダイアログ [「Add GO Annotation Database」が現れます(Fig.15-2)。

Scaffo format	Id DIA can annotate proteins using the Gene Ontol t.These may be obtained from the Gene Ontology A wight the GOA site for more information	ogy Annotation files in GAF nnotation Database.	
- riedse	Human Only (EBI; UK; ~10 mins)	Choose file	
source:	ftp://ftp.ebi.ac.uk/pub/databases/GO/goa/HUM/	AN/goa_human.gaf.gz	

Fig.15-2 Add GO Annotation Database ダイアログ
Source の項目をクリックすると、一覧に取得可能な各種生物種/種類の GO ファイルが現れます。適切な項目を選び「Add」ボタンを押すことで、ファイルのダウンロードならびに Scaffold DIA へのセットが自動的に行われます。

同一生物種で複数の種類がありますが、どれを選べばよいか判断がつかない時はとりあえず後ろに何も ついていない項目を選んでください。例えば Fig.15-3 で Chicken であれば「goa_chicken」です。

🛼 Ad	Add GO Annotation Database									
Scaffold DIA can annotate proteins using the Gene Ontology Annotation files in GAF format. These may be obtained from the Gene Ontology Annotation Database. Please visit the <u>GOA</u> site for more information.										
Source:	Human Only (EBI; UK; ~10 mins)		Choose file							
	Human Only (EB); OK; ~10 mins) Other Web Site Other File	/goa/HUMAN/go	oa_human.gaf.gz							
	goa_chicken goa_chicken_complex goa_chicken_isoform goa_chicken_ma									
0	goa_cow goa_cow_complex goa_cow_isoform		Add <u>C</u> ancel							

Fig.15-3 Add GO Annotation Database ダイアログ 「Source」項目

上記オプションは Scaffold DIA がオンラインの時でかつセキュリティ的にこの操作が問題ない時のみ使用 可能です。上記オプションが適用できない場合でも、必要なファイルを別の場所で取得し、Scaffold DIA に セットする事も出来ます。取得したファイルをセットする時には「**Choose file...**」ボタンを押します。

取得するべきファイルの場所については、Source で取得したい項目を選んだ際にその下に表示される URLを参照してください(Fig.15-4、赤で囲われた部分)。



Fig.15-4 Add GO Annotation Database ダイアログ 「Source」項目

15-2. Scaffold DIA の各タンパク質に Gene Ontology 情報を表示させる方法

15-1 でセットした Gene Ontology ファイルの情報と同定タンパク質との紐づけと表示を行う事が 出来ます。Gene Ontology については非常に多数の項目がある中で、ユーザーが事前にどの項目を表示 させるのかを選択しておく必要があります。

15-2-1. 表示する GO 項目を選択する

Edit -> **Edit GO term options** で「GO Term Configuration」ダイアログを開き、「**Displayed GO Terms**」タブをクリックします。

画面上部が Gene Ontology の分類すべて、画面下部がその中でピックアップして Scaffold DIA で表示 する項目となります(Fig.15-5)。

GO Term Con	figuration				×
Displayed GO Te	erms GO Ann	otations			
Search:			Total Terms: 469	937	0
 O biological O cellular co O molecular 	l process omponent r function				
Salastad CO Tarm	A	Add Remove Reset to	User Default	Reset to Scaffold DIA Default	
Selected GO Term	5:49				-
Color ID	Head Node	Selected GO Terms		Definition	tê.
0 16209	Biological Pro	antioxidant activity	Inhibition of the	reactions brought about by dioxygen (O2) or peroxide	S
22610	Biological Pro	biological adhesion	Any process that	of a cell of organism to a substrate, another cell, of othe	er
0 1906	Biological Pro	cell killing	Any process in a	a organism that results in the killing of its own cells or	th
0097	Biological Pro	cellular process	Any process in al	is carried out at the cellular level, but not necessarily n	oct
42056	Biological Pro	chemoattractant activity	Providing the en	vironmental signal that initiates the directed movement	to
42030	Biological Pro	chemorenellent activity	Providing the en	vironmental signal that initiates the directed movemen	to
	n' i ' i n	i i i i	A L · · · ·	i in a indicate and the directed movement	
<u> ()</u> еlp				<u>0</u>	K

Fig.15-5 GO Term Configuration ダイアログ 「Displayed GO Terms」 タブ

ッリー構造となっている GO 情報から目的の項目を探し、その項目を選択した上で画面中央の"Add" ボタン(Fig.15-6,赤で囲われた部分)を押すと、下の"Selected GO Terms"項目へ移り、結果画面に該当 項目が表示されるようになります。

🜉 GO Term Configura	tion		×
Displayed GO Terms	GO Annotations		
Search:		Total Terms: 46937	0
 O biological proce O single-organ Hythmic pre localization actomyo localizati maintena cellular la cellular la	ess isism process sccess sin contractile ring localization on of cell ince of location scalization blecule localization ment of localization lishment of organelle localization oort tracellular transport ansporter activity tracellular transport ansport ansporter activity tracellular transport Add Remove Reset to	User Default Reset to Scaffold DIA Default	
Selected GO Terms: 49			
Color ID Head	Node Selected GO Terms	Definition	Ę
16209 Biologi	cal P antioxidant activity	Inhibition of the reactions brought about by dioxy	gen (O2) or
 22610 Biologie 65007 Biologie 	cal P biological adhesion	The attachment of a cell or organism to a substrate Any process that modulates a measurable attribute	e, another ce
 1906 Biologi 	cal P cell killing	Any process in an organism that results in the killin	a of its own
 9987 Biologi 	ral P cellular process	Any process that is carried out at the cellular level	but not nec
2 Help			<u>O</u> K

Fig.15-6 GO 項目の選択と Add ボタン

逆に現在設定されている表示項目を外したい場合、その項目を選び"**Remove**"ボタンを押すことで 実施できます(Fig.15-7)。



Fig.15-7 GO 項目の選択と Remove ボタン

15-2-2. タンパク質と GO 情報との紐づけを行う

Edit -> **Edit GO term options** で「GO Term Configuration」ダイアログを開き「**GO Annotations**」 タブをクリックして、適用したい GO ファイルを選択して「OK」を押してください(Fig.15-8)。

GO Term Configuration	×
Displayed GO Terms GO Annotations	
The GO Annotation database currently contains	search 👂
Name	4
goa_human.gaf.gz	
Import <u>A</u> nnot	ations
Help	<u><u>o</u>ĸ</u>

Fig.15-8 GOファイルの選択

続いて、Experiment -> Apply GO Terms ->Apply All annotations を選択すると、

Gene Ontology の情報が表示されます(Fig.15-9)。

Scaffold DIA - C	saffold DIA - Check03_16data2nd.sdia – — X														×																																	
<u>File Edit View</u>	e Edit View Experiment Export Tools Help																																															
🖹 🎦 🚍 🚍 🕼 🕂 Summarization: Type 🗸 🕍 Protein FDR 1.0% FDR 🗸 🖓 Min # Peptides 💈 🗘 🦹 🖉																																																
Filters																																																
Show Hidden	w Hidden 🕅 👷 Aune/Accession 👂 p-value filter 🔻 GO Term 🔻 🍳 🔣																																															
Лпа																																																
	Display	Type:	Exclus	sive Intensity	× 🗹	Norma	lized		og Inte	ensitie	s	Co	lor Op	otion	s																																	
															Biolo	gical	Proc	cess													C	ellul	ar Co	mpo	nent							N	lole	ular	Fund	tion		
Organize													c						, v																									≥				
Samples	up Score		eptide Count			activity	dhesion	egulation	Cess	ctant activity	llent activity	ntal process	unator activity ent of localizatio		stem process			process	ir organismal pri		e	e process	o stimulus ocess	regulator activit	activity	2		atus		ic reticulum		r region	r organelle		ion		embrane	art				uvity	mer acuvity	ransducer activi	ity	ervoir activity		utagila articitu
**	Protein Gro		Identified P	Exclusivity	t-test CL: A BRL: type	Taxonomy	biological a	biological r	cell killing cellular pro	chemoattra	chemorepe	developmen	establishme	growth	immune sy:	localization	locomotion	metabolic p	multicellula	pigmentatio	reproductio	reproductiv	response to rhythmic pr	translation	transporter	viral proces	transport	Golgi appar	cutoskeleto	endoplasmi	endosome	extracellula	intracellula	membrane	mitochond	nucleus	organelle n	organerie p	choromo	hinding	Guipuid	catalyuc ac	electron ca	molecular t	motor activ	nutrient res	protein tag	a landon da
Proteine	0.877	1	11	9%				•	0														•					(•		۲				•		(
Flotenis	0.825	1	10	10%				•	0				•			•					0	0	•				0	(0	٠	•	0	•	((
	0.993	1	16	12%	0.85			-	0			•	-			-											-	-			-	•	•	-			_			-								•
	0.976		8	• 12%				•				•				•		_	•				_				0					0	•	•						-								
	0.952		8	 12% 12% 	0.67			-				•	•			•		0		. •			•				0	• •			•	-		•						-								
Visualize	0.891		7	 12% 				•								•	•			-			•					-		5							- 2	5		-								2
4	0.867		7	• 14%				•	ŏ			ō				ŏ	ē											0		5		ŏ	ē			ō	-	5		Ċ								ć
	0.798		7	14%	0.47			•	•				٠			•		0		•					•		0	0					٠	•	•					(
	0.991	1	18	17%	0.49			•	0			•			۰								•					(0	٠			•	(()							(
Analysis	0.972		6	• 17%				•	0			•	0		-	0							•				0	0				•	•	•						0								
Analysis	0.985	1	10	20%	0.96			0	0			•	•		•	0							•				0	_				0	•	0		•												•
12 2	0.937		5	20%	0.32		0	•									•		•	0						•	0	• •			•	0	-			. '				-								
	0.936		о 0	20%	0.84											•											0					-													•			
	0.964		8	25%	0.29				ŏ			ŏ			•	0		0					ŏ		0		0	č		5		Ŭ	÷	ě	•	0				Ì	0				-			
Publish	0.957		8	25%	0.36			•	0			-			-	0		-					-		0		0	0					•	•	0	0				0	0							
	0.853		4	25%	0.52			•				•							•													0																
	0.832		4	25%	0.32			•	0				•		۰	0											0	• •			•	•	٠	٠						(
	0.983	1	19	26%	0.50		0	•	0			•		•				0			0	0	•			٠		(0	٠	•	•	•	((
	0.902		7	29%				•	0			•	0			0			•				•				0	0			•	•	•	•						(
	0.984	1	10	30%	0.98				0						•								0					(0	•			0	((9						(
Destrolog	0.896		9	• 33%				•	0			•											•				~			•		0	•			•												G
0.1% FDR (attained)	0.961		6	33%												0		0					_			۲	0													-		2						
804 Targets	0.98		3	33%	0.74		0	-					-		•	•				-			9				0	•			-	•	-	•		. '				-								
1 Decoys	0.963		3	33%	0.57			•	0																			4					-															×
Peptides	0.963		3	 33% 					ē																							0	•				6			-	n i							•
0.1% FDR (attained)	0.94		3	• 33%	0.96		0	•				0	0		•	0							0				0	Ċ				0					-	0		0								¢
2 Decoys								-																									-															

Fig.15-9 GO 情報の表示[Samples 画面]

GO 情報の表示では、円のマークが全塗りと中が空いているものがありますが、これは情報の信頼度を 表します。全塗りのものは信頼度が高い、人間の実験者やキュレーターの確認を経ている情報で、中が 空いているのはコンピューターのアルゴリズムによるアノテーションです。GO ファイル内にある Evidence code 情報に基づいています。Code とそれらを Scaffold DIA でどのように表示するかについては Tab.15-1 をご覧ください。

Type of evidence	Code	Type of Circle
Unrecognized Evidence Code	UNKNOWN	closed circle
Inferred from Mutant Phenotype	IMP	closed circle
Inferred by Curator	IC	closed circle
Inferred from Genetic Interaction	IGI	closed circle
Inferred from Physical Interaction	IPI	closed circle
Inferred from Sequence or Structural Similarity	ISS	closed circle
Inferred from Direct Assay	IDA	closed circle
Inferred from Expression Pattern	IEP	closed circle
Inferred from Electronic Annotation	IEA	open circle
Traceable Author Statement	TAS	closed circle
Non-traceable Author Statement	NAS	closed circle
Not Recorded	NR	closed circle
No biological Data available	ND	closed circle
Inferred from Reviewed Computational Analysis	RCA	open circle
Inferred from Sequence Orthology	ISO	open circle
Inferred from Sequence Alignment	ISA	open circle
Inferred from Sequence Model	ISM	open circle
Inferred from Genomic Context	IGC	open circle
Inferred from Experiment	EXP	closed circle

Tab.15-1 GO 情報の Evidence のコードと Scaffold DIA での表示(Type of Circle)

GO 情報については、Samples 画面で並べ直しを実施できるほか、Visualize で項目をカウントした 円グラフ、ならびにタンパク質と項目の関係性を表示した Connection graph を表示させる事も可能です (**8-2-4**, P.73~)。

16. デモデータと追加情報のご案内

16-1. デモデータについて

Scaffold DIA にはデモデータが準備されています。デモデータは

Help -> Open Demo Files

で開くことができます。



16-2. Appendix について

本ソフトウェアを使用するにあたり、特に計算方法について付録として英文マニュアルに説明が ございます。ご不明な点がございましたらそちらの資料もご参照ください。

- 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

16-3. EncyclopeDIA について

Scaffold DIA における、ペプチド同定や定量の中心的なプログラムは、「EncyclopeDIA」というフリーの プログラムが担っています。EncyclopeDIA については以下 URL をご参照ください。

https://www.biorxiv.org/content/early/2018/03/07/277822 https://bitbucket.org/searleb/encyclopedia/wiki/Home

ご不明な点がございましたら、ご遠慮なくお問い合わせください。

技術サポート 担当:高江洲(たかえす)

電子メール:support-jp@matrixscience.com 電話:03-5807-7897 ファックス:03-5807-7896