

# Skyline での DIA/SWATH データ分析

注：本チュートリアルでは Sciex TripleTOF 6600 装置のデータを使用しています。Thermo Q-Exactive Plus のデータを使用したい場合は、Skyline DIA QE というタイトルのバージョンをご覧ください。

本チュートリアルでは、質量範囲の異なる 64 のウィンドウでプレカーサーを単離する枠組設定（スキーム）と 1 時間のグラジエントにより、QqTOF 装置 (6600 TripleTOF、Sciex) で取得された SWATH-MS データセットを用いて、Skyline を使用したペプチドと推測タンパク質の検出および定量化の取得後ターゲット分析を実施する方法を学びます。

使用するデータセットは、[LFQBench 調査](#) で用いられた定量的ベンチマーク試料と同様に 3 種の生物プロテオームを一定比率で混合した定量的ベンチマーク試料が有機体 3 種のプロテオームを定義された比率で混合した試料から得られたものです (図)。

最初は、データ非依存性測定 (DIA) のデータを解析するため、Skyline のパラメータを設定し、次に、生データファイルから定量化情報を抽出します。また、データ依存性解析 (DDA) の結果をインポートしてスペクトルライブラリを作成し、DIA のデータを解析するための設定を行います。



[図の出典：Navarro, P. et al. A multicenter study benchmarks software tools for label-free proteome quantification. *Nature Biotech* 34, 1130–1136 (2016)]

## はじめに

チュートリアルを始める前に、以下の zip ファイルをダウンロードしてください。

<https://skyline.ms/tutorials/DIA-TTOF.zip>

この中のファイルを、以下のコンピュータ上のフォルダに解凍します。

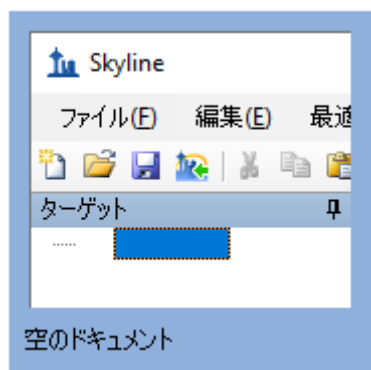
C:\Users\brendanx\Documents

これにより以下の新しいフォルダが作成されます。

C:\Users\brendanx\Documents\DIA-TTOF

本チュートリアルを始める前に Skyline を使用していた場合には、Skyline をデフォルト設定に戻すことをお勧めします。デフォルト設定に戻すには、以下の操作を行います。

- Skyline を起動します。
- 開始ページで、以下の空のドキュメントをクリックします。

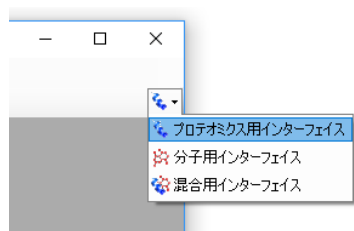



- [設定]メニューで、[デフォルト]をクリックします。
- 現在の設定を保存するかどうかを尋ねるフォームで[いいえ]をクリックします。

Skyline の現在のインスタンスの設定がデフォルトにリセットされました。

本チュートリアルではプロテオミクスのトピックを扱うため、ユーザーインターフェイスを必ず「プロテオミクス用インターフェイス」に設定してください。

- 開始ページの右上隅にあるユーザーインターフェイスボタンをクリックし、[プロテオミクス用インターフェイス]を選択します。



これで、Skyline は、右上隅のプロテインアイコン  で示されるプロテオミクスモードで動作します。

DIA 分析の場合では、Skyline に抽出された定量トランジションすべてのピーク領域を積分するようにします。このためには、以下のように操作して下さい。

- [設定]メニューで、[すべてを積分]をクリックします。

## ペプチド検索のインポートウィザードの使用

Skyline の操作に慣れているユーザーは、以下のすべての手順はメニューオプションから行うことができますが、DIA の通常のワークフローをより簡単に実行するための「ウィザード」

を利用することにより、選択肢を選ぶことで順をおって解析することが可能になります。最初に、DDA 解析で得られたペプチド・スペクトル・マッチの結果からスペクトルライブラリを構築します。次に、クロマトグラムを抽出する対象となる DIA 測定のデータセットを指定します。ついで条件設定を行い、最後にクエリの対象を選びます。

ここでは、試料 A と B のそれぞれについて測定した 1 回の DDA ランの解析結果からライブラリを構築します。この解析結果は、Comet 検索エンジンを用いて検索し、PeptideProphet により検証したものです。Skyline が対応しているその他の DDA 検索エンジンについては、[https://skyline.ms/wiki/home/software/Skyline/page.view?name=building\\_spectral\\_libraries](https://skyline.ms/wiki/home/software/Skyline/page.view?name=building_spectral_libraries) を参照して下さい。使用するファイルは interact.pep.xml です。このファイルは PeptidePrphet で作成されたもので、2 つの DDA ファイルのデータベース検索結果を含みます。

開始するには、次の操作を行います。

- [ファイル]メニューで、[インポート]を選択して[ペプチド検索]をクリックします。

Skyline は現在のドキュメントを保存するか尋ねてきます。これは、コンピュータのどこのディスクドライブのどのフォルダで解析を実行するのかを決めておくためです。

- [OK] ボタンをクリックします。
- 先ほど作成した DIA-TTOF フォルダに移動します。
- [ファイル名] フィールドに、「DIA-TTOF-tutorial」と入力します。
- [保存] ボタンをクリックします。

すると、ペプチド検索のインポートウィザードが表示されます。

ペプチド検索のインポート

スペクトルライブラリ

構築(B)     既存を使用する(U)

カットオフスコア(C):  
0.95

開始(M):  
検索結果 (ライブラリを直接構築)

結果ファイル:  
[Empty Text Box]    [ファイルを追加(A)...]  
[ファイルを削除(O)]

iRT標準ペプチド(I):  
なし     曖昧な一致を含める(O)

ワークフロー  
 MS1フィルタを用いたDDA  
 DIA  
 PRM

[完了(F)]    [次へ(N) >]    [キャンセル]

注：0.95 は、DDA データベース検索において、すべてのペプチド・スペクトル・マッチ（PSM）に対して PeptideProphet で設定した確率の閾値です。このチュートリアルで用いるデータセットでは、偽陽性率（false discovery rate, FDR）0.2%に相当します。データセットが異なればこの値は変更する必要があります。ユーザーが設定した FDR 値を与える閾値をここに入力してください

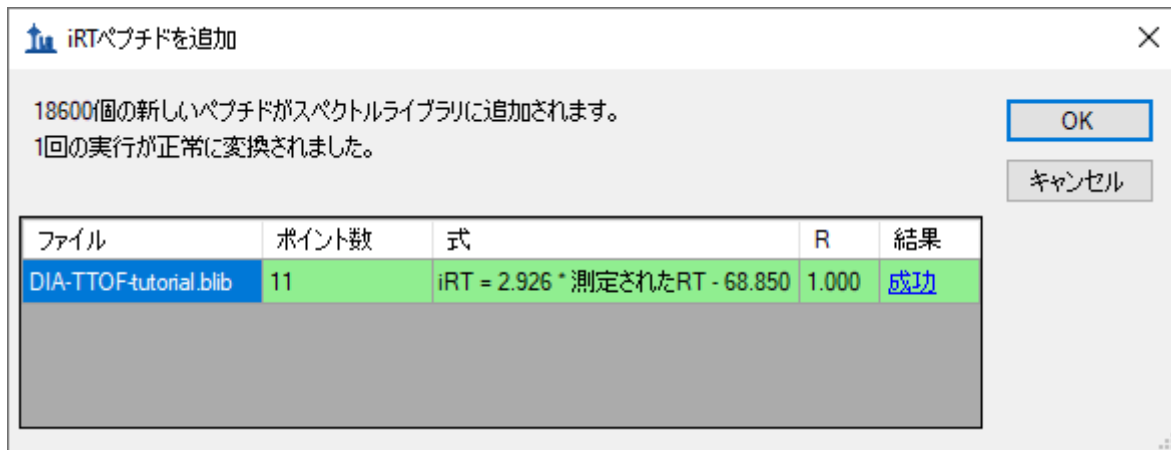
- [ファイルを追加] ボタンをクリックします。
- 先ほど作成した DIA-TTOF フォルダに移動し、DDA\_search サブフォルダに移動します。
- 「interact.pep.xml」ファイルをダブルクリックします。
- [iRT 標準ペプチド] ドロップダウンリストで、「Biognosys-11 (iRT-C18)」をクリックします。
- [ワークフロー] ボックスで [DIA] オプションをクリックします。

ウィザードの [ スペクトルライブラリ ] ページは以下ようになります。

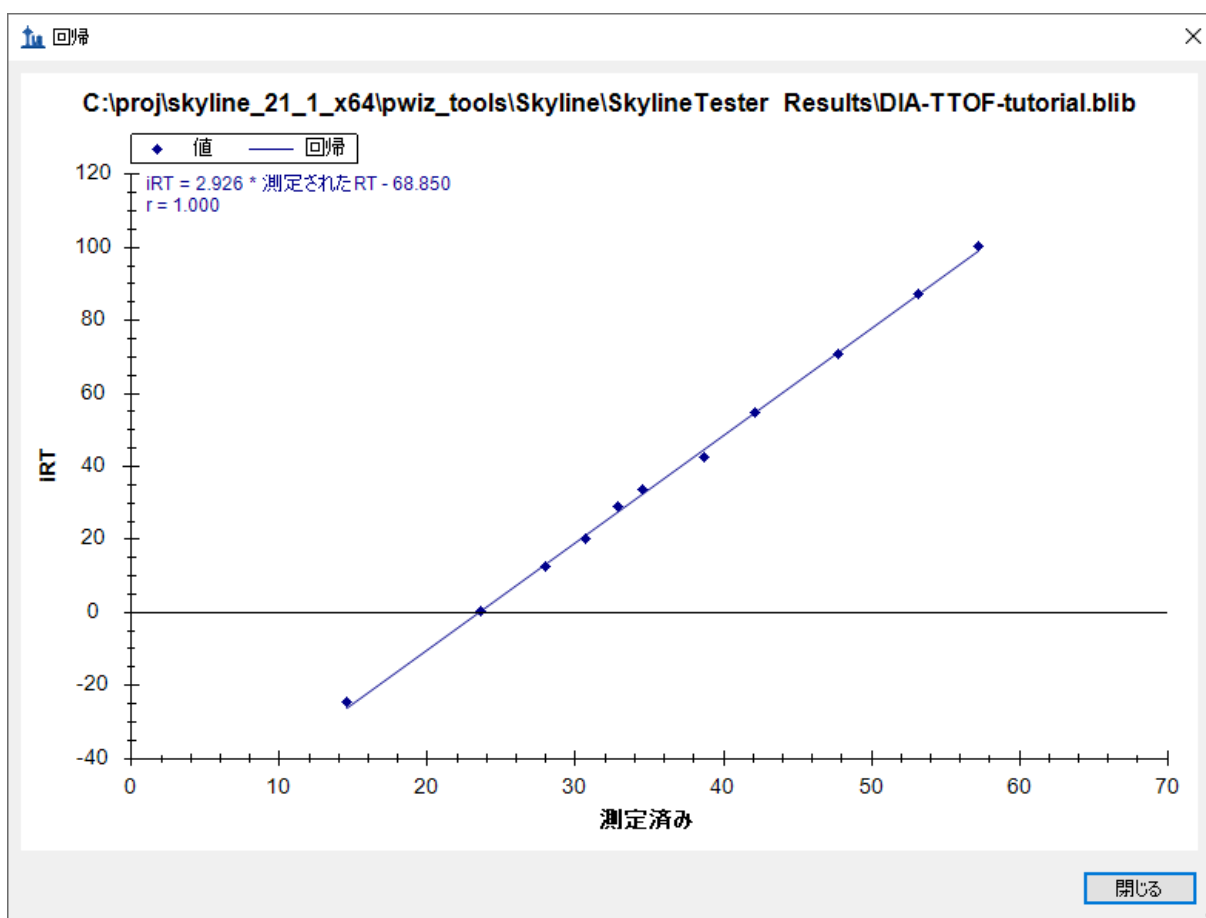
The screenshot shows a dialog box titled 'ペプチド検索のインポート' (Import Peptide Search) with a close button (X) in the top right corner. The main title is 'スペクトルライブラリ' (Spectral Library). Below the title, there are two radio buttons: '構築(B)' (Build) which is selected, and '既存を使用する(U)' (Use Existing). Underneath, there is a label 'カットオフスコア(C):' (Cut-off Score) and a text input field containing '0.95'. The next section is '開始(M):' (Start) with a dropdown menu showing '検索結果 (ライブラリを直接構築)' (Search Results (Build Library Directly)). Below that is '結果ファイル:' (Result File) with a text input field containing 'interact.pep.xml' and two buttons: 'ファイルを追加(A)...' (Add File...) and 'ファイルを削除(D)' (Remove File). The 'iRT標準ペプチド(I):' (iRT Standard Peptide) section has a dropdown menu showing 'Biognosys-11 (iRT-C18)' and a checkbox '曖昧な一致を含める(U)' (Include Ambiguous Matches) which is unchecked. The 'ワークフロー' (Workflow) section has three radio buttons: 'MS1フィルタを用いたDDA' (DDA with MS1 filter), 'DIA' (selected), and 'PRM'. At the bottom, there are three buttons: '完了(F)' (Finish), '次へ(N) >' (Next), and 'キャンセル' (Cancel). The '次へ(N) >' button is highlighted with a blue border.

- [次へ] ボタンをクリックします。

Skyline はライブラリ構築を開始し、[ペプチド検索ライブラリの構築] フォームで進行捗状況が表示されます。ライブラリ構築が完了すると、これらの試料にスパイクされた Biognosys iRT 標準ペプチドに基づいて、ペプチド検索結果の iRT 値が校正されます。校正が成功すると、以下のようなライブラリ構築の要約が表示されます。



- 成功(青でハイライト)をクリックすると線形回帰の結果が表示されます。



- [閉じる] ボタンをクリックします。
- [OK] ボタンをクリックします。

Skyline は、上記の回帰直線に追加されたペプチドに対して iRT 標準値を再校正するかどうかをユーザーに尋ねます。

- [いいえ] ボタンをクリックします。

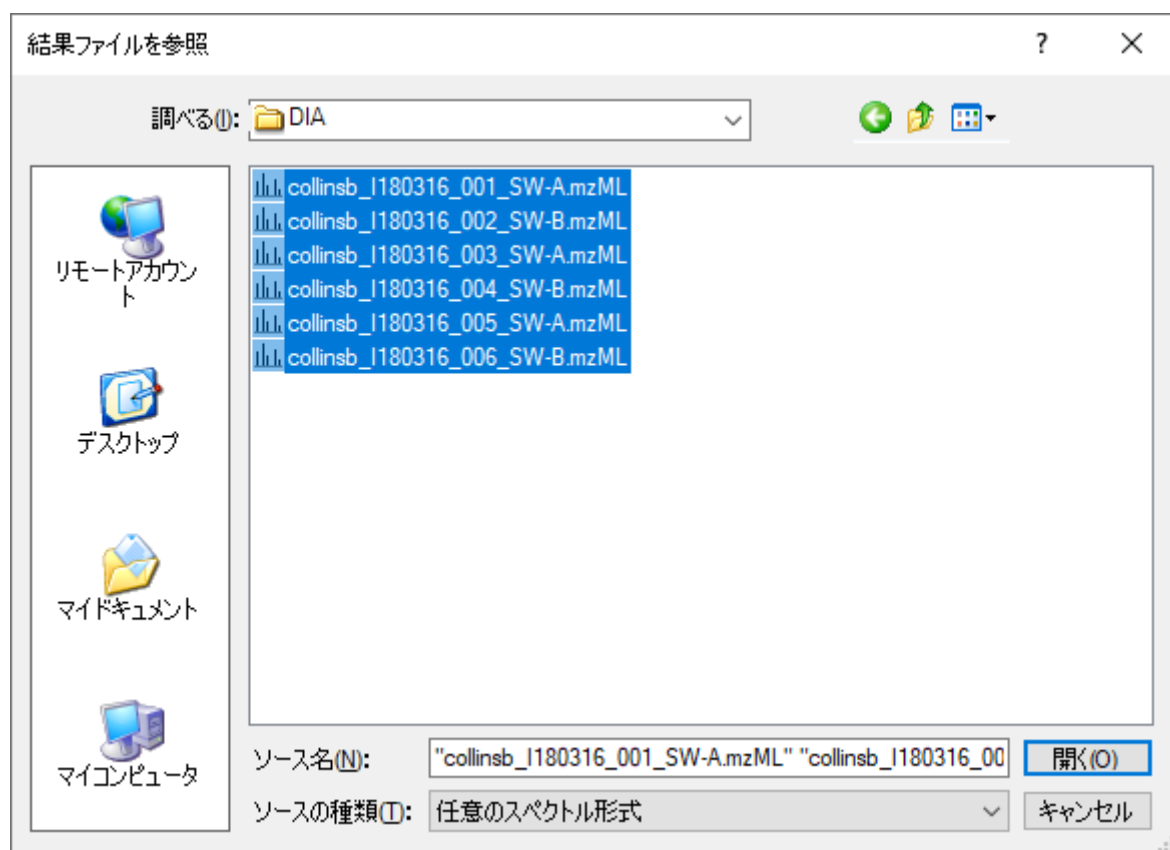
Skyline が曖昧に一致したペプチドの長いリストを表示します。これらのペプチドは、複数のペプチドにマッチするスペクトルが存在することから表示されたものであり、これらのスペクトルとマッチしたペプチドは定量的ターゲット解析には不適切なものです。

- [OK] ボタンをクリックします。

[ クロマトグラムを抽出 ] ページが表示されます。このページでは、クロマトグラムの抽出、ピーク検出、ピーク領域計算に使用する DIA データファイルがどこにあるかを指定します。

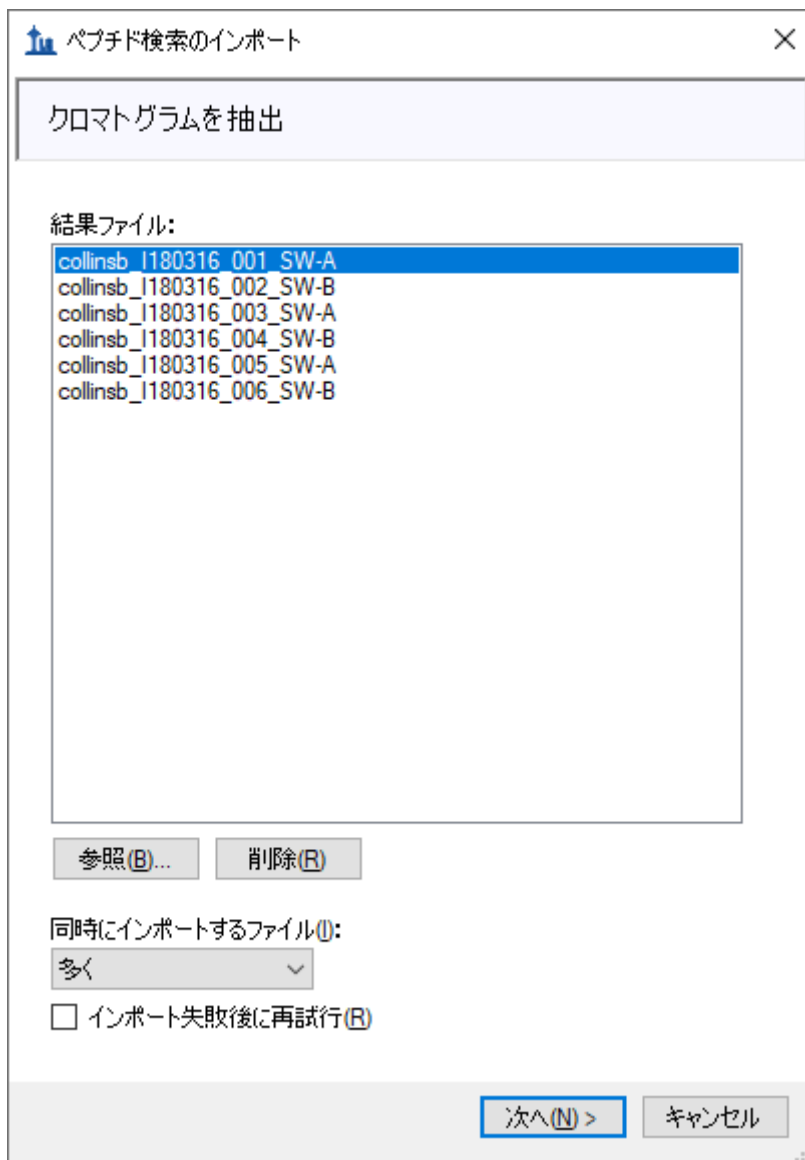
- [参照] ボタンをクリックします。
- DIA-TTOF フォルダ内の DIA サブフォルダをダブルクリックします。
- このファイルにある 6 個のファイルをすべて選択します。

[ 結果ファイルを参照 ] は以下のようにになります。



- [開く] ボタンをクリックします。

ペプチド検索のインポートフォームは以下のようになります。



- [次へ] ボタンをクリックします。

Skyline は、データファイル名を短縮して作業しやすくするため、ファイル名に共通するプリフィックス「collinsb\_X1803\_17」を削除するかどうかユーザーに尋ねます。

- [OK] ボタンをクリックします。

次の操作で [修飾を追加] ページをスキップします。

- [次へ] ボタンをクリックします。

[トランジション設定を行う] ページでデフォルト値を以下のように変更します。

- [プリカーサーの電荷] を「2,3,4」に設定します。
- [イオンタイプ] を「y,b」に設定します。

注：イオンタイプは、プリカーサーを意味する「p」を残してもかまいません。この設定



では、Skyline は DIA データファイルの MS1 スペクトルから同位体を含むプリカーサーのクロマトグラムを抽出します。ただし、このチュートリアルを簡潔にし、処理時間を短縮して出力ファイルサイズを小さくするため、以下の説明では DIA の MS/MS スペクトルから抽出したフラグメントイオンのクロマトグラムのみを対象とします。

- [DIA のプリカーサーウィンドウを exclusion に使用] チェックボックスをオンにします。

[ペプチド検索のインポート] フォームは以下のようになります。

ペプチド検索のインポート

トランジション設定を行う

プリカーサーの電荷 (P):     イオン電荷 (I):     イオンタイプ (T):

プロダクトイオンのソース (R):     プロダクトイオンの移動先 (Q):

最小 m/z (N):  m/z    最大 m/z (X):  m/z

DIA のプリカーサーウィンドウを exclusion に使用 (U)

イオン許容誤差 (M):  m/z    選択 (S):  プロダクトイオン  
 プロダクトイオンの最小数

< 戻る (B)    次へ (N) >    キャンセル

- [次へ] ボタンをクリックします。

Skyline は [フルスキャン設定を行う] ページを表示します。

注：これらの抽出設定は、データ取得に使用した装置の種類や設定により変わります。至適な抽出設定は、それぞれのデータセットによって多少異なることがあります。この分析では、スペースを小さくするためにセントロイド化されたデータを使用します。したがって、

「Centroided」を選択し、抽出の質量精度は「20」 ppm を指定します。プロファイルモードデータでは、装置の分解能を指定できます。

デフォルト値を以下のように変更します。

- 両方の [質量精度] フィールドで「20」 ppm を使用します。

次に、DIA 測定に使用した装置で定義されているパラメータに従って新規に単離スキームを定義する必要があります。

注：この例では、400～1200 m/z の質量範囲をカバーする 64 に分割した可変幅のウィンドウを使用し、重複を 1 m/z として四重極の質量範囲の両端で 0.5 m/z の除外マージンを設けるように設定してあります。

- [単離スキーム] ドロップダウンリストで、「<追加...>」をクリックします。

この操作で [単離スキームを編集] フォームが表示されます。ここでは以下の操作を行います。

- [名前] フィールドに「ETH TTOF (64 variable)」と入力します。
- [事前に指定した単離ウィンドウ] オプションをクリックします。
- [インポート] ボタンをクリックします。
- 先ほど作成した DIA-TTOF フォルダに移動し、DIA サブフォルダに移動します。
- 「collinsb\_l180316\_001\_SW-A.mzML」 ファイルをダブルクリックします。

[単離スキームを編集] フォームは以下のようになります。

開始	終了	マージン
399.5000	408.2000	0.5
407.2000	415.8000	0.5
414.8000	422.7000	0.5
421.7000	429.7000	0.5
428.7000	437.3000	0.5
436.3000	444.8000	0.5
443.8000	451.7000	0.5

Skyline は、質量分析計で測定された DIA 単離範囲の [開始] および [終了]  $m/z$  値を表示しています。これらの値は  $1.0m/z$  ずつ重複しています。この重複を相殺するため、Skyline は単離ウィンドウの各端に  $0.5m/z$  のマージンを追加しています。このマージンは、クロマトグラム抽出時に除外されます。これを表示するには、以下の操作を行います。

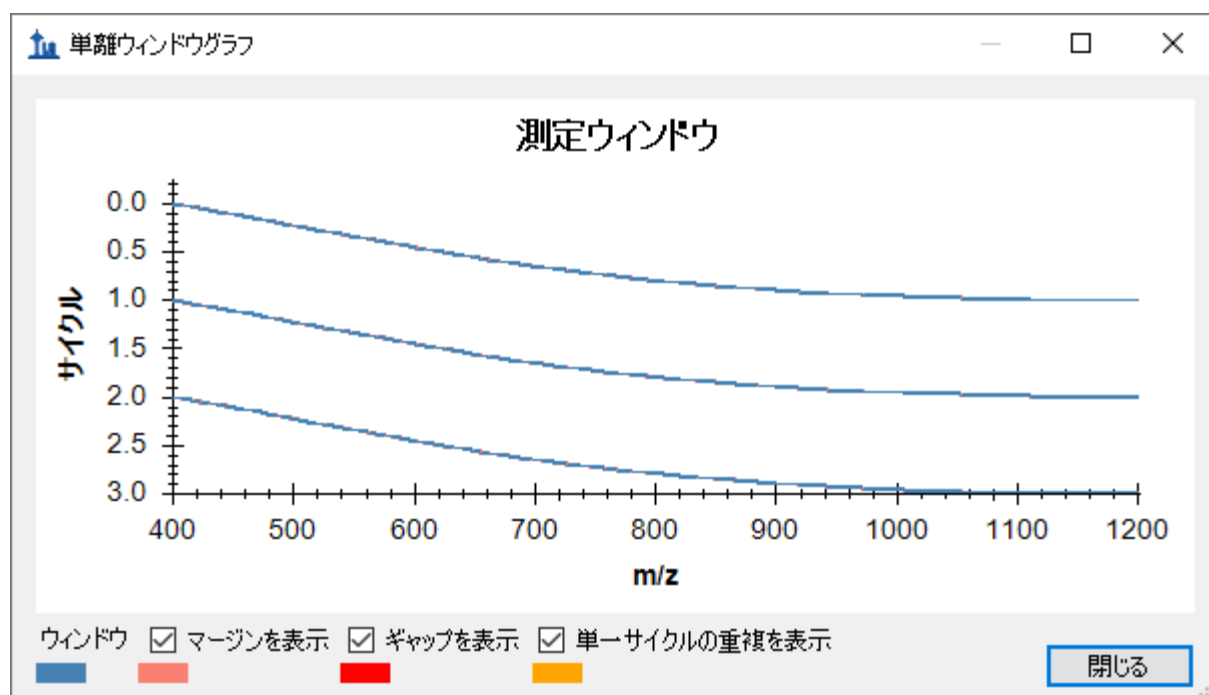
- [事前に指定した単離ウィンドウ] の下にあるドロップダウンリストで、「抽出」をクリックします。

[開始] および [終了] 値が  $0.5m/z$  内側にシフトし、範囲が重複しなくなります。これらは、フラグメントイオンクロマトグラムの抽出するために、ペプチドプリカーサー  $m/z$  値を取得したスペクトルとマッチさせるときに Skyline が使用する範囲です。

単離スキームを視覚化するには、以下の操作を行います。

- [グラフ化] ボタンをクリックします。

以下のようなフォームが表示されます。



クリック・アンド・ドラッグして囲んだ長方形を使い部分拡大するか、マウスのスクロールホイールを使用して拡大すると、マージンの位置の詳細を見ることができます。[マージンを表示] チェックボックスをオフにし、再びオンにすると、グラフのタイトルが [測定ウィンドウ] と [抽出ウィンドウ] の間で切り替わり、ピンクのマージンバーが非表示になったり再表示されたりします。[ギャップを表示] の赤や、[単一サイクルの重複を表示] の黄色は、どちらも DIA 単離スキームデザインのエラーであるため、この例では表示されません。

- [閉じる] ボタンをクリックします。
- [単離スキームを編集] フォームで [OK] ボタンをクリックします。

ペプチド検索のインポートウィザードは以下のようになります。

- [次へ] ボタンをクリックします。

[FASTA をインポート] ページが表示されます。このページでは以下の操作を行います。

- [参照] ボタンをクリックします。
- 最初に作成した DIA-TTOF フォルダに移動し、DIA サブフォルダに移動します。
- 「target\_protein\_sequences.fasta」ファイルをダブルクリックします。

注：ここで、前に使用した「interact.pep.xml」ファイルがある DDA\_search サブフォルダに移動して、ペプチド検索で使用した、検索対象のすべての FASTA シークエンスファイルから構成される「napedro\_3mixed\_human\_yeast\_ecoli\_20140403\_iRT\_reverse.fasta」をダブルクリックすると、検索対象とした全プロテオームの解析が可能になります。この選択によってずっと多くのターゲットペプチドが見つかるようになります。処理により長い時間がかかりますが、ほとんどの最新ラップトップで実行可能です。

- [デコイ生成法] ドロップダウンリストで、「シーケンスのシャッフル」をクリックします。
- [mProphetのモデルを自動的にトレーニングする] チェックボックスをオンにします。

ペプチド検索のインポートウィザードは以下のようになります。

ペプチド検索のインポート

**FASTAをインポート (必須)**

酵素: Trypsin [KR | P]      最大未切断数: 0

FASTAレコードは>で始まり、タンパク質名の後にオプションのタンパク質表記が続きます。

```
>sp|P63284|CLPB_ECOLI Chaperone protein ClpB OS=E
MRLDRLTNKFQLALADAQSLALGHDNQFIEPLHLSAL
RTDINQALNRLPQVEGTGGDVQPSQDLVRLNLCDKL
TLADILKAAGATTANITQAIEQMRGGESVNDQGAEDQF
IGRDEEIRRTIQVLQRRTKNNPVLIGEPGVGKTAIVEGL
DMGALVAGAKYRGEFEERLKGVLNDLAKQEGNVILFID
PALARGELHCVGATTLDYRQYIEKDAALERRFQKVVFV
HHVQITDPAIVAAATLSHRYIADRQLPDKAIDLIDEAASS
QLKLEQQALMKESDEASKKRLDMLNEELSDKERQYSE
QAKIAIEQARRVGD LARMSELQYGKIPELEKLEAATQL
ARWTGIPVSRMMESEREKLLRMEQELHHRVIGQNEA\
FLFLGPTGVGKTELCKALANFMFDSEAMVRIDMSEFM
YLTEAVRRRPYSVILLDEVEKAHPDVFNILLQVLDDGRL
GSDLIQERFGELDYAHMKELVLGVVSHNFRPEFINRIDE
```

参照(O)...

クリア(C)

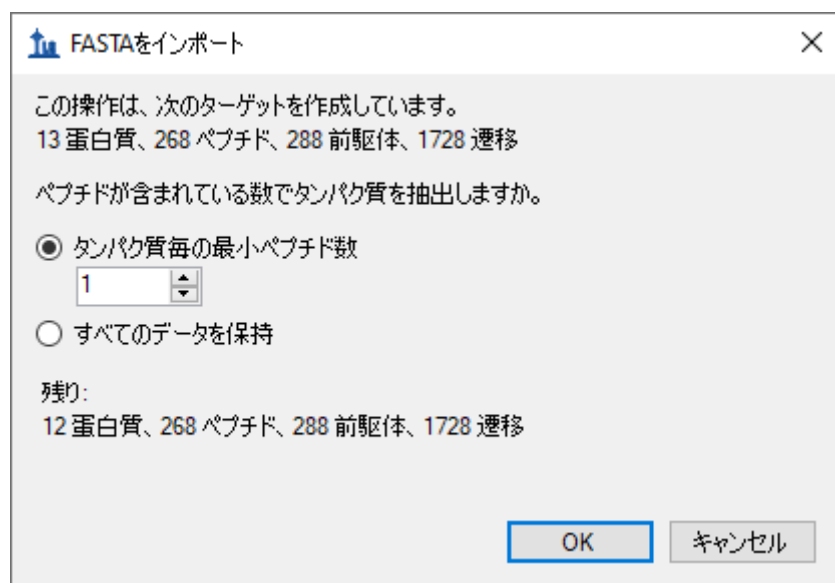
デコイ生成法: シーケンスのシャッフ      ターゲット毎のデコイ: 1

mProphetのモデルを自動的にトレーニングする

< 戻る(B)      完了(F)      キャンセル

- [完了] ボタンをクリックします。

設定条件と使用した FASTA シークエンステキストを基に計算されたターゲットを記述したフォームが、以下のように表示されます。



FASTAをインポート

この操作は、次のターゲットを作成しています。  
13 蛋白質、268 ペプチド、288 前駆体、1728 遷移

ペプチドが含まれている数でタンパク質を抽出しますか。

タンパク質毎の最小ペプチド数  
1

すべてのデータを保持

残り:  
12 蛋白質、268 ペプチド、288 前駆体、1728 遷移

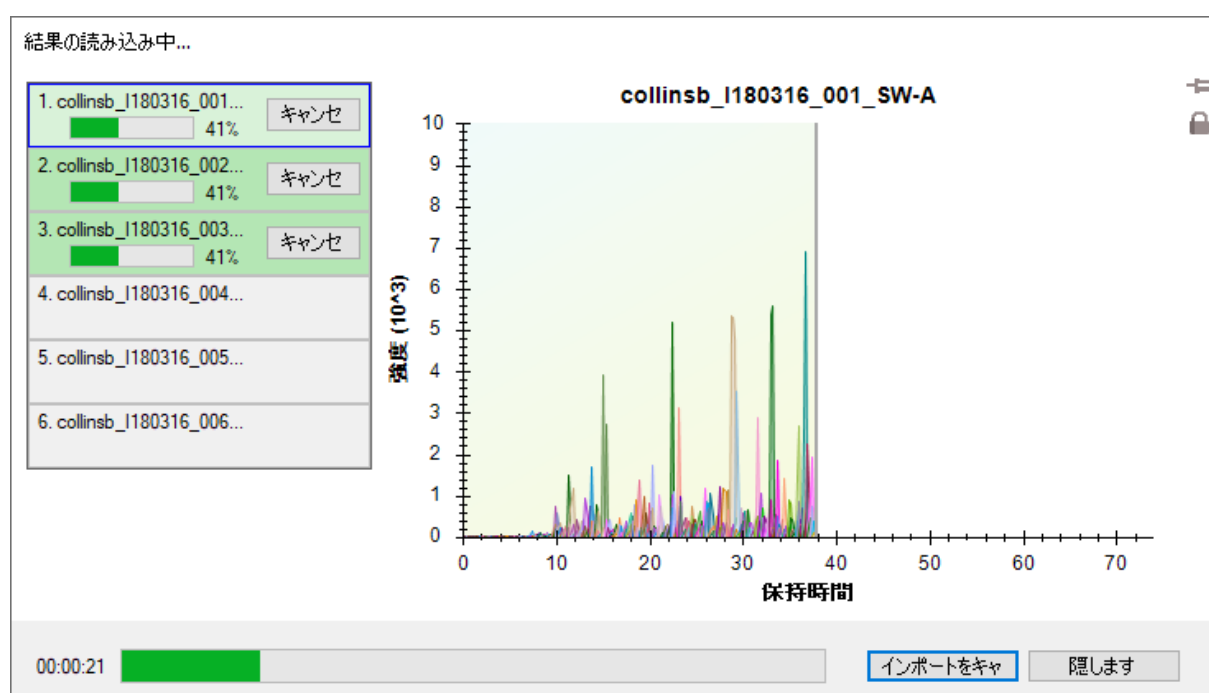
OK キャンセル

[タンパク質毎の最小ペプチド数] フィルタを「1」にした場合、FASTA ファイルで見つかった 13 個のタンパク質のうち、残りで表示されるタンパク質は 12 個になります。これは 1 つのタンパク質に帰属するペプチドがスペクトルライブラリの中に見つからないためです。

[すべてのデータを保持] をクリックすると、「1 個の空欄のタンパク質が追加されます。」という警告が表示されます。ここでは前の設定に戻り、空欄のタンパク質がターゲットリストに追加されないようにして下さい。

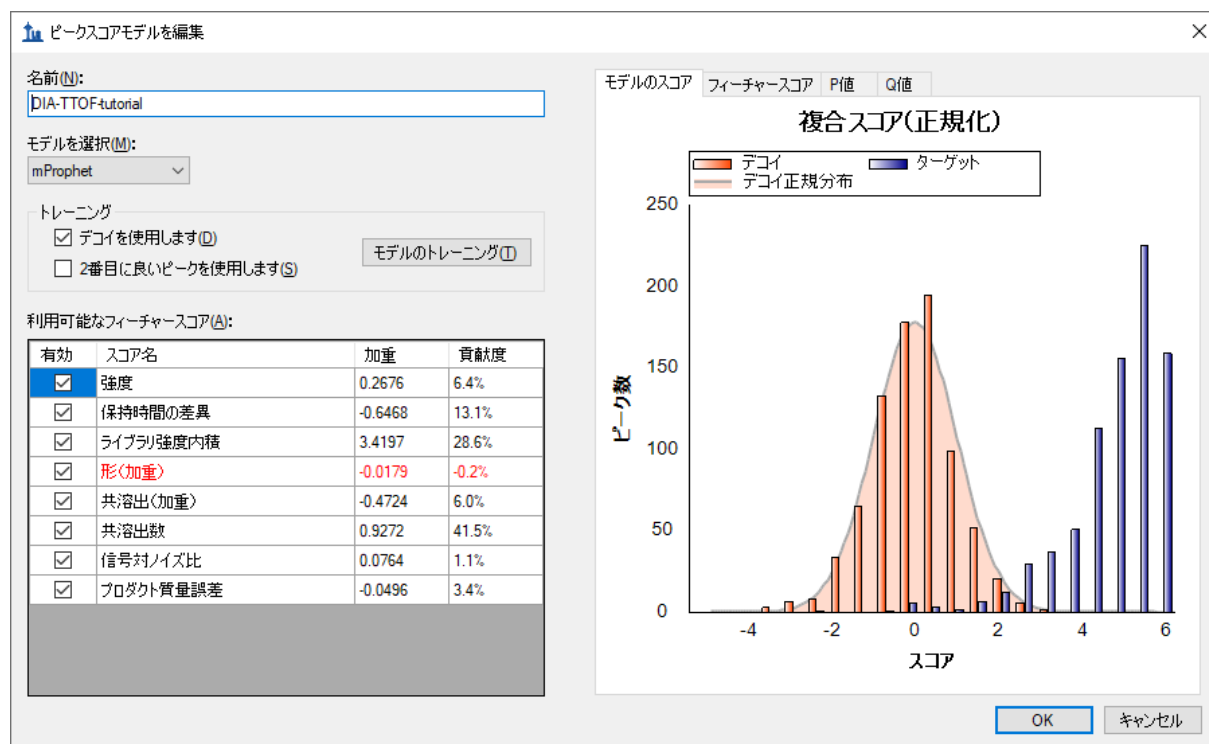
- [OK] ボタンをクリックします。

Skyline はクロマトグラムの抽出を開始します。標準の i74 コアプロセッサでは以下のようになります。



6 コア以上ではすべてのファイルが並列処理され、2 コアのほとんどのラップトップでは、インポートは2 ファイルずつ処理されます。

インポートが完了すると、Skyline はペプチド検索のインポートウィザードの [FAST をインポート] ページで選択した mProphet モデルを表示します。これは以下ようになります。



mProphet モデルでは、それぞれのクロマトグラムセットの中から、Skyline がデフォルトとして使用するスコアに基づき選んだ 10 個のベストピークを見直し、最も高い mProphet スコアを与える 10 個のピークを選びます。これらのピークの mProphet スコア (Skyline では **Detection Z Score** という名称になっています) は、用いたデコイペプチドシーケンスに対して同様な方法で得られる 10 個のベストピークの mProphet スコアの平均値の標準偏差が 1.0 になるように再計算されます。この値は q 値となります (Skyline では **Detected Q Value** という名称になっています)。

## 試料の注釈

次に、どの試料がどの実験グループに属するかを定義する必要があります。

- [設定]メニューで[ドキュメント設定]をクリックします。
- アクティブになっていない場合は、[注釈]タブをクリックします。
- [追加]ボタンをクリックして[注釈を定義]ウィンドウを表示します。
- [名前]フィールドに「Condition」と入力します。
- [タイプ]フィールドで「値のリスト」を選択します。
- リストの下にある[適用先]で、[繰り返し測定]チェックボックスをオンにします。

この実験では、2つの条件があります。条件 A は E.coli が 20%、酵母が 15%、ヒトが 65%の proteome 構成である試料、条件 B は E.coli が 5%、酵母が 30%、ヒトが 65%の構成です。



[注釈を定義] ウィンドウは以下のようになります。

注釈を定義

名前(N):  
Condition

編集可能 計算済み

タイプ(T):  
値のリスト

値(V):  
A  
B

適用先(A):  
 タンパク質  
 ペプチド  
 プリカーサー  
 トランジション  
 繰り返し測定  
 プリカーサー結果  
 トランジション結果

OK キャンセル

- [OK] ボタンをクリックします。
- 上記ステップに従い、「BioReplicate」という名前の2つ目の注釈を作成します。
- [タイプ] ドロップダウンリストで、BioReplicate 注釈に「テキスト」を選択します。
- [適用先] の下にあるリストで、[繰り返し測定] チェックボックスをオンにします。
- [OK] ボタンをクリックします。

[ドキュメント設定] ウィンドウに戻ります。

- [Condition] チェックボックスと [BioReplicate] チェックボックスがオンになっていることを確認します。
- [OK] ボタンをクリックします。

これでインポートした繰り返し測定に注釈を付ける準備ができました。

- [表示] メニューで、[ドキュメントグリッド] (Alt+3) を選択します。

[ドキュメントグリッド] ウィンドウが表示され、このウィンドウを初めて使用する場合はタンパク質レポートが表示されます。

- [ドキュメントグリッド: タンパク質] ウィンドウで、[レポート] メニューをクリックし、[繰り返し測定] をクリックします。

キーボードの Enter、矢印キーを使って移動し、A、B、1、2、3 キーを使って以下のスクリーンショットで示すように試料に注釈を付けます。

ドキュメントグリッド: 繰り返し測定

	繰り返し測定	試料タイプ	アナライト濃度	Condition	BioReplicate
	<a href="#">1 SW-A</a>	不明		A	1
	<a href="#">2 SW-B</a>	不明		B	1
	<a href="#">3 SW-A</a>	不明		A	2
	<a href="#">4 SW-B</a>	不明		B	2
	<a href="#">5 SW-A</a>	不明		A	3
▶	<a href="#">6 SW-B</a>	不明		B	3

- フォームの右上隅にある赤い X をクリックして [ドキュメントグリッド: 繰り返し測定] ウィンドウを閉じます。

見やすくするため、Skyline ウィンドウでデータを条件毎に分割して 2 つの枠で表示できるようになりました。

- [表示] メニューで、[グラフを配置] を選択して [グループ化] をクリックします。
- [グループ化されたグラフを配置する] フォームで [枠のグループ化] を「2」に設定します。
- [グループ間の分散グラフ] オプションを選択します。
- [表示] フィールドで「タイトル」を選択します。
- [並べ順] フィールドで「ドキュメント」を選択します。
- [OK] ボタンをクリックします。

注：この時点でデータがまだインポート中である場合は、インポートが完了するまで待つからドキュメントを保存します。

- Skyline ドキュメントを保存します。

## データのマニュアル検証

次に、クロマトグラフィーやその基になるスペクトルをマニュアルで検証します。この操作を行うには、**保持時間**と**ピーク領域**表示を表示させる必要があります。

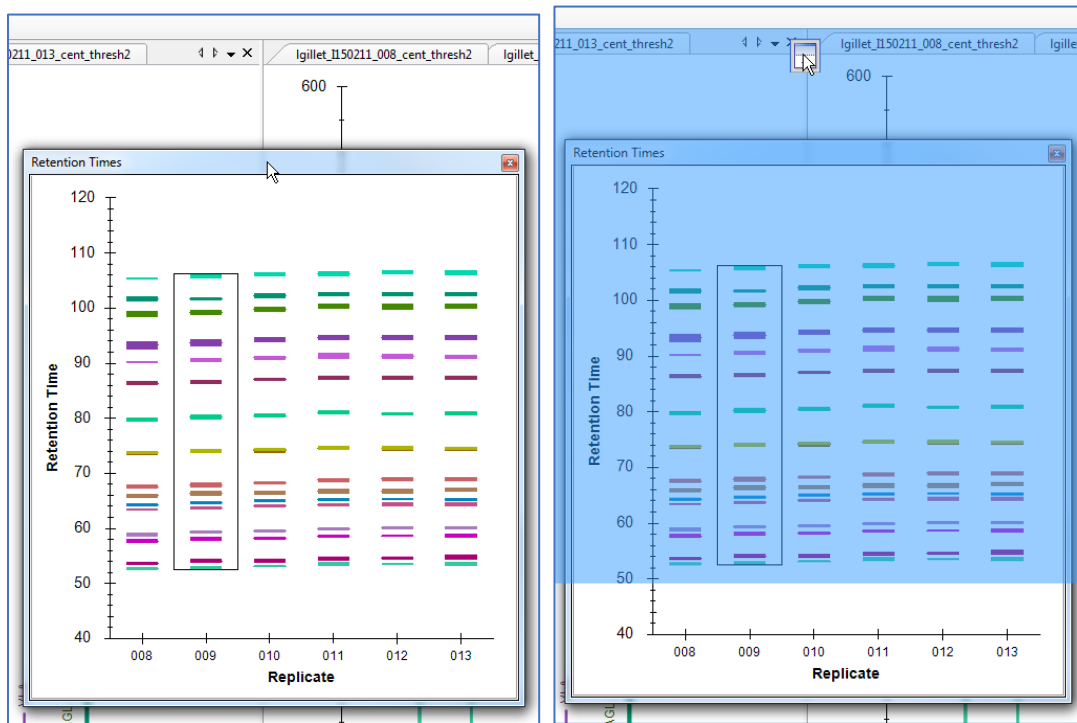
- [表示] メニューで、[保持時間] を選択して [繰り返し測定比較] (F8) をクリックします。
- [表示] メニューで、[ピーク領域] を選択して [繰り返し測定比較] (F7) をクリックします。

これで両方が表示されます。以下の操作を行うと、画面上でこの表示を最適に配置できます。

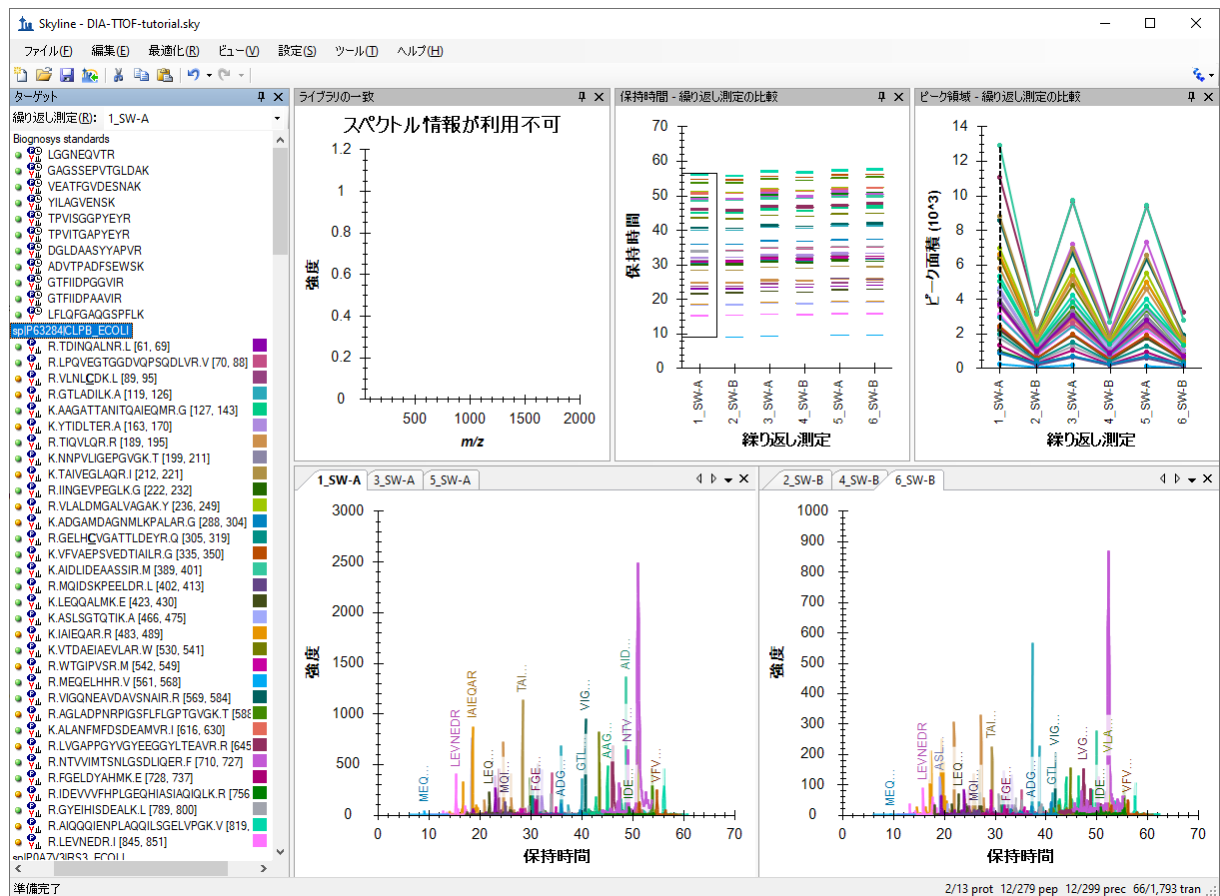
- この新しいウィンドウがフロート表示になっている場合は、フロート表示になっているウィンドウの上枠をクリックし、マウスの左ボタンを押したままウィンドウをドラッグするとメインウィンドウにドッキングできます。

[ライブラリの一致] ウィンドウにも、以下に示すように同じ操作を行えます。

注：このようにウィンドウをドッキングさせるには、マウスの矢印カーソルをドッキングアイコンの内部に置く必要があります。この場合は上部にあります。マウスの左ボタンを押したままドラッグし始めると、Skyline がドッキングアイコンと、マウスの左ボタンから手を放したときにウィンドウが配置される位置を示す透明な青い長方形を表示します。



- 同様に、[ピーク領域: 繰り返し測定比較] ウィンドウと[ライブラリの一致] ウィンドウをドッキングし、以下のようにすべての情報が見やすくなるようにします。



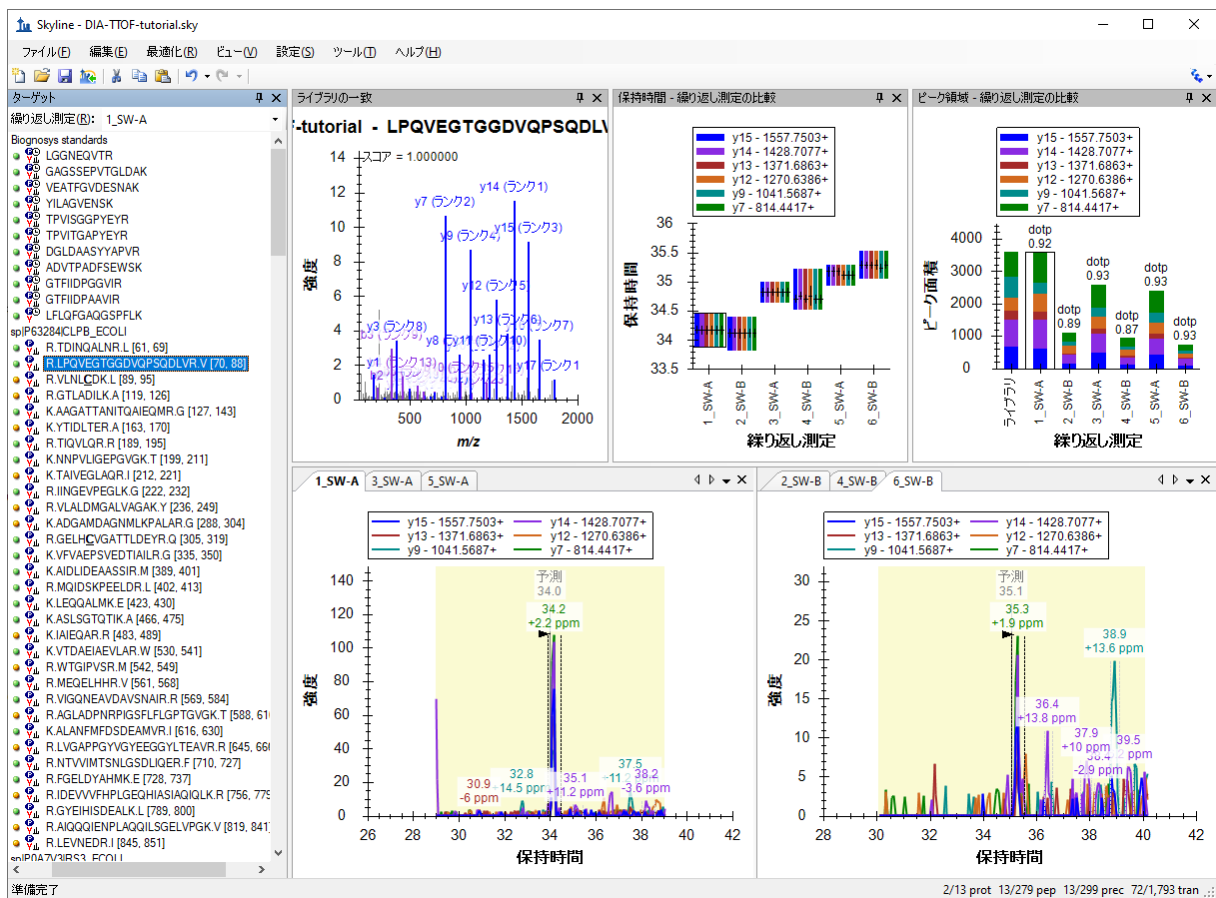
- 最初の E. coli タンパク質である「sp|P63284|CLPB\_ECOLI」をクリックします。

このタンパク質のすべてのペプチドがさまざまなプロット（クロマトグラム、ピーク領域、保持時間繰り返し測定グラフ）で表示されます。上記スクリーンショットは、あるタンパク質が選択された場合の例であり、このタンパク質のすべてのペプチドが各表示で要約されています（何も表示されない[ライブラリの一致]ウィンドウを除く）。

[ピーク領域 - 繰り返し測定比較]のプロットを見て、このタンパク質の発現量が異なっているように見えるでしょうか？ E. coli の場合、サンプル A と B では、E. coli タンパク質の量比が 4:1 になるように調整されていることを思い出してください。

- このタンパク質でペプチド LPQVEGTGGDVQPSQDLVR を選択します。

下に示すように、すべての表示でこのペプチドに関する情報が表示されます。



このタンパク質に属する残りのペプチドのピーク領域パターンを調べて下さい。このタンパク質のペプチドの定量パターンは、予想される差異発現パターン (4:1) と一致していますか？

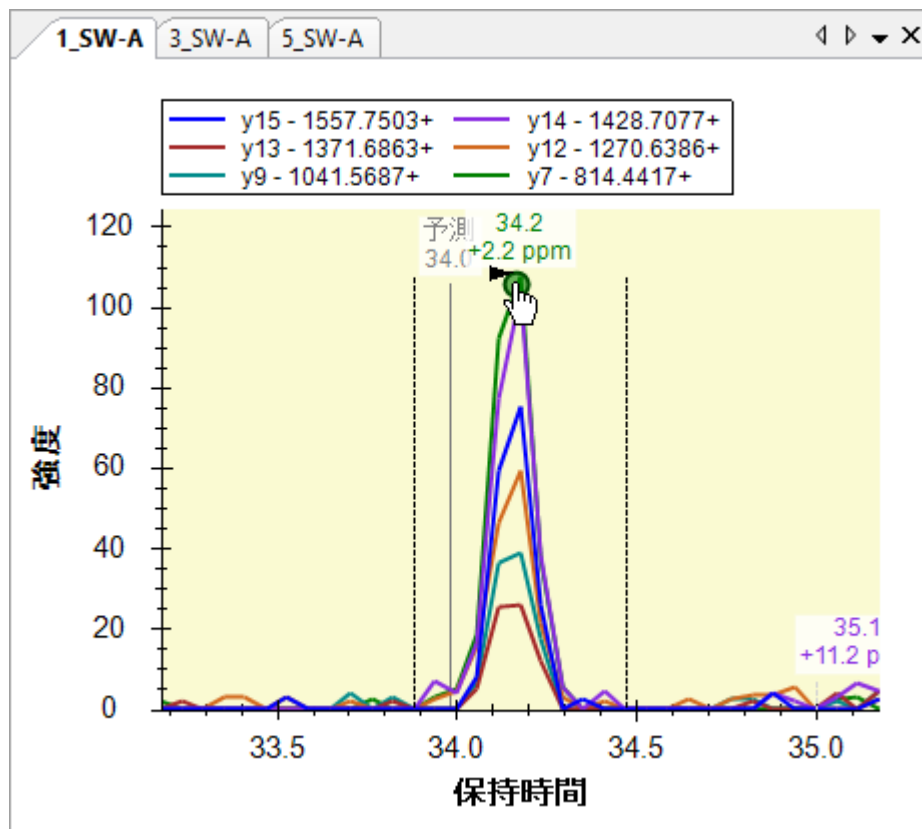
注：ペプチドシーケンスに対して2つ以上の電荷状態がある場合、それらは別々に抽出され、スコアが付けられています。これらは、ペプチドシーケンスの横にある+をクリックしてから個々の電荷が異なるプリカーサーをクリックすれば参照できます。

- ドキュメント内のヒトタンパク質をいくつかクリックします。

タンパク質レベルと、ペプチドレベルのそれぞれの繰り返し測定ピーク領域を調べます。ピーク領域は予想される比率 (1:1) と一致していますか？酵母タンパク質 (1:2) ではどうでしょうか？

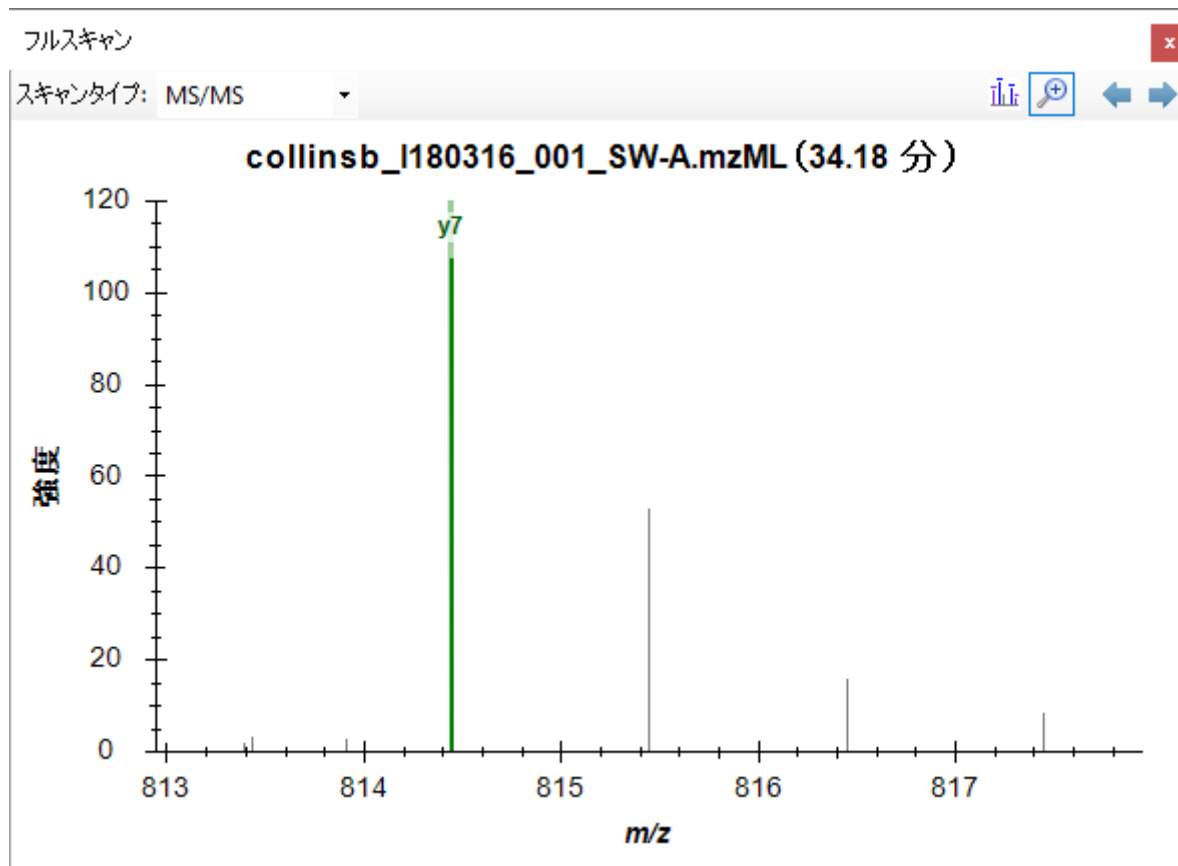
- 最初のタンパク質の LPQVEGTGGDVQPSQDLV ペプチドをもう一度クリックします。
- [表示]メニューで、[自動ズーム]を選択し、[最適ピーク] (F11) をクリックします。

クロマトグラム曲線上にカーソルを合わせると、曲線上に曲線と同じ色の丸が表示されます。

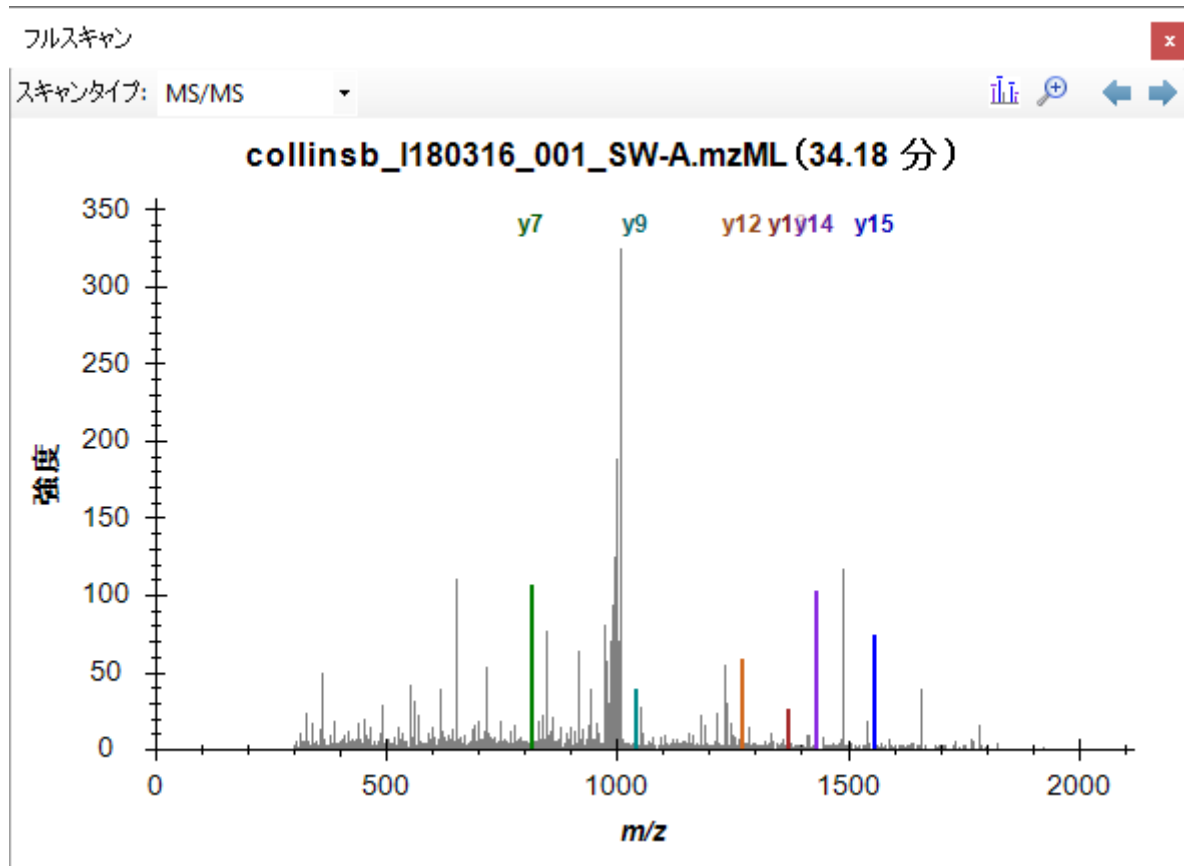


- クロマトグラムポイントにある丸をクリックします。

Skyline は選択したクロマトグラムポイントが抽出されたスペクトルを、選択したイオンを拡大して [フルスキャン] 表示を開きます。



- [フルスキャン] ツールバーの虫眼鏡ボタンをクリックしてフルスペクトルを表示します。



手作業でさらにデータを調べます (デコイを含む)。

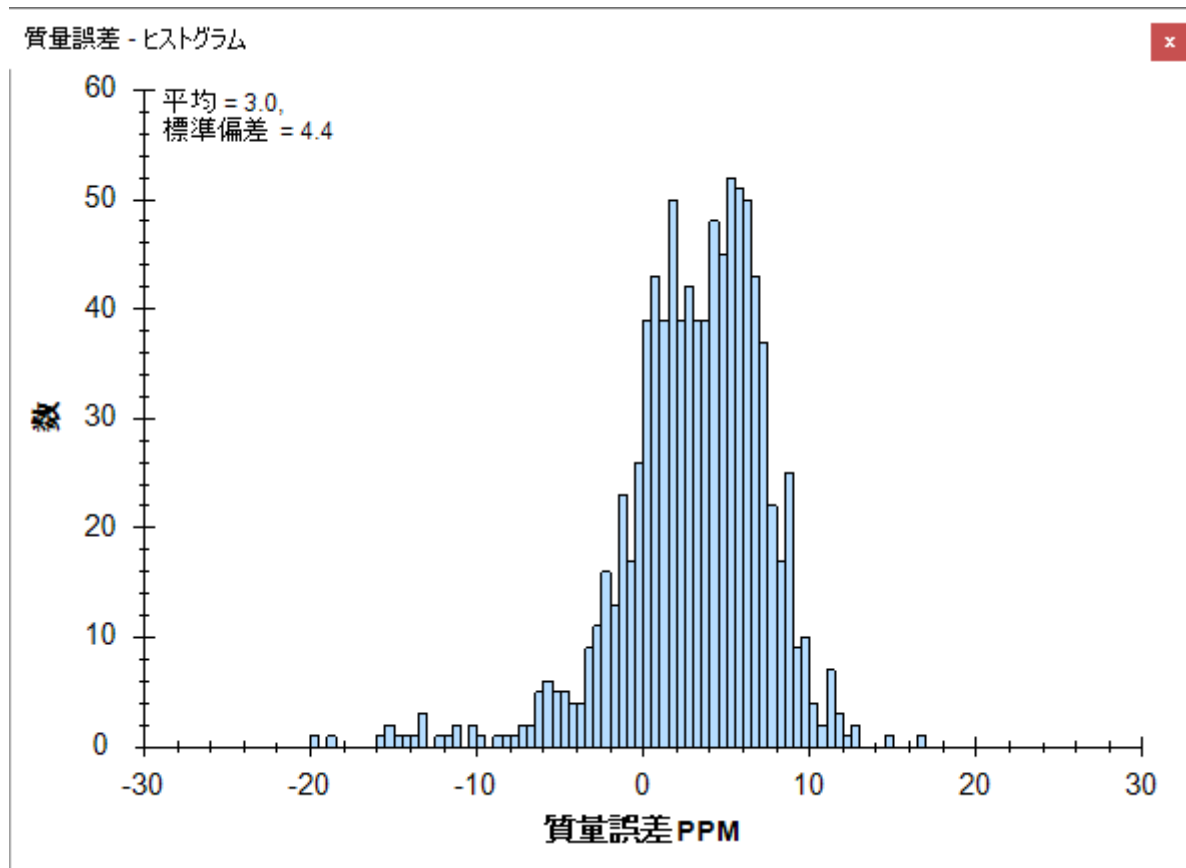
## 質量誤差と保持時間偏差

質量精度と保持時間予測精度を調べると、最適な抽出パラメータが使用されたか、調整によって結果の改善が期待できるかを判断できます。

- [表示]メニューで[質量誤差]を選択し、[ヒストグラム]をクリックします。

この操作によってデータセットの質量誤差分布が表示されます。抽出ウィンドウ ( $\pm 20$  ppm) はさらに最適化できるでしょうか？一般的には、平均差 $\pm 3$  の標準偏差で十分です。ただし、両端の数字が高く、ヒストグラムが途切れているような場合は、抽出ウィンドウを広げた方がいいかもしれません。





分布には、0 周辺と 5 周辺とピークが 2 つあるように見えることに注意してください。これをよりよく理解するには、これが 6 回のランすべてでの質量誤差値であることを知っていることが有用です。

- グラフを右クリックし、[繰り返し測定]を選択して[単一]をクリックします。
- [ターゲット]表示の上部にある[繰り返し測定]ドロップダウンリスト内をクリックします。
- キーボードの上下矢印キーを使用して、6 回の繰り返し測定すべてにおける質量誤差値を確認します。

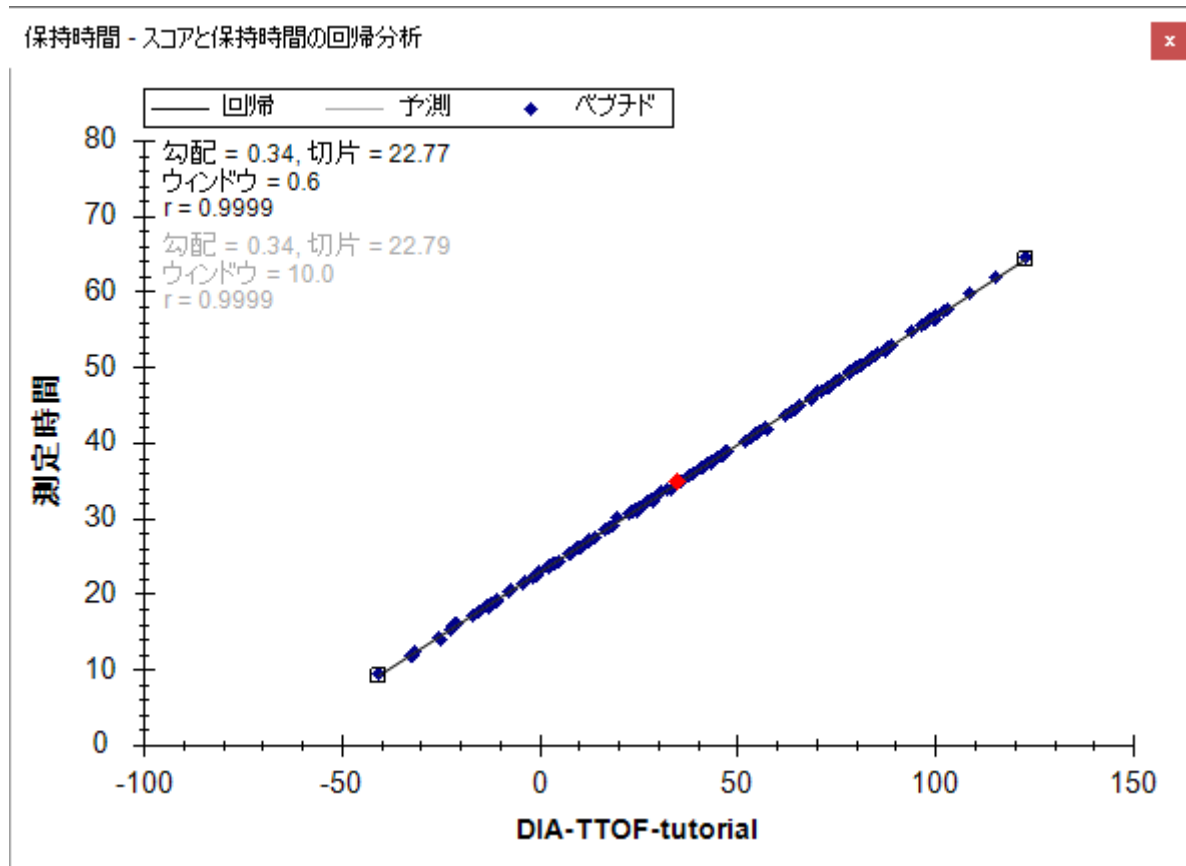
この確認操作により、平均値は-0.1~5.9、標準偏差値は 3.4~4.3 であることがわかります。最大値を選び、 $\text{平均} + 3 * \text{標準偏差} = 5.9 + 3 * 4.3 = 18.8\text{PPM}$  という簡単な計算により、20PPM の許容誤差がこれらのデータが要求するものはかなり優れた概算であるようです。

デコイの質量誤差を表示する場合は、以下の操作を行います。

- グラフを右クリックし、[点]を選択して[デコイ]をクリックします。
- グラフを右クリックし、[繰り返し測定]を選択して[すべて]をクリックします。  
(その後再び[ターゲット]に切り替えます)

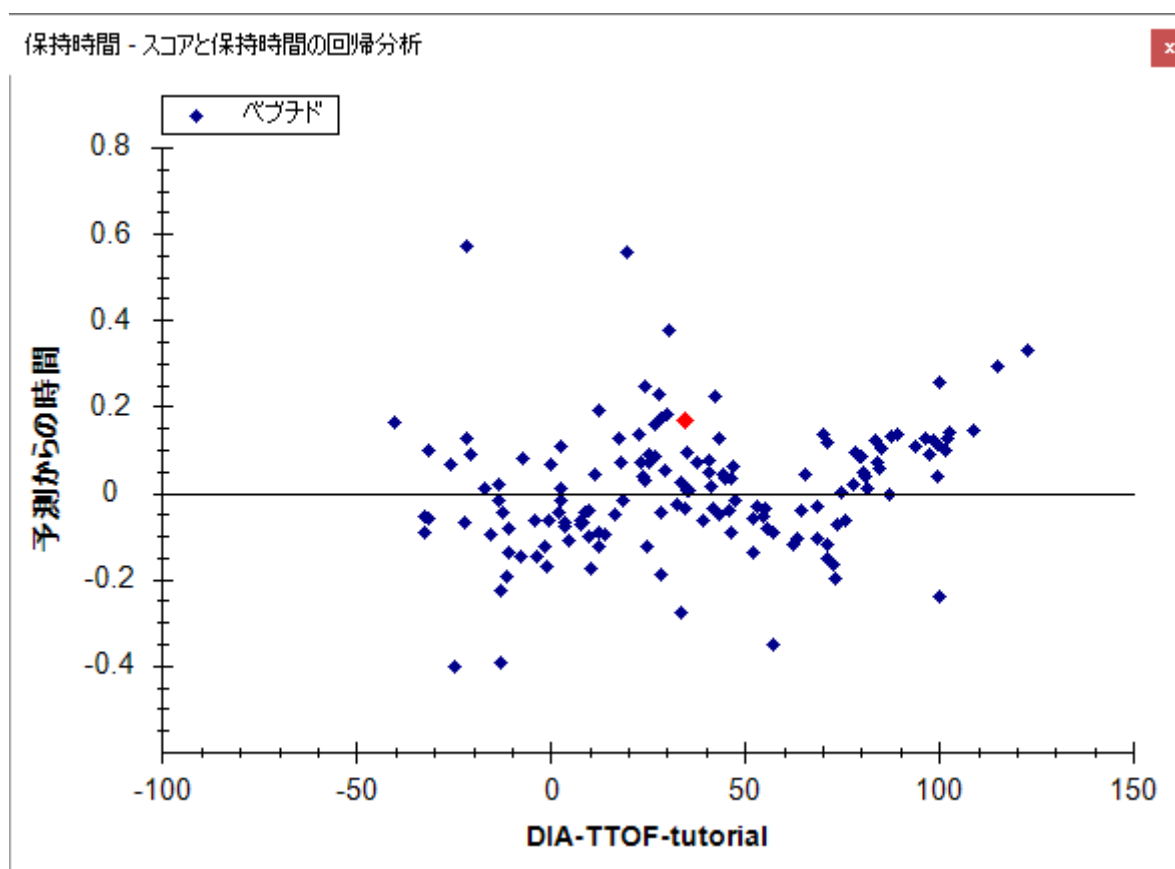
iRT ペプチドとそれらのライブラリにおける iRT の値に基づいてターゲットペプチド保持時間の予測に使用する線形回帰を表示するには、以下の操作を行います。

- [表示]メニューで、[保持時間]、続いて[回帰]を選択してから、最後に[実行するスコア]をクリックします。



- グラフを右クリックし、[プロット]を選択して[残余]をクリックします。

これにより、このデータセットでの予測保持時間からのずれが表示されます。この分析に対し、抽出ウィンドウ ( $\pm 5$  分) はさらに最適化できるでしょうか？デコイも確認してみましょう。



注：このチュートリアルの特クトルライブラリは同じ試料を同時に測定した結果を用いて構築しました。そのため、保持時間予測の精度は非常に高くなっています。異なる装置、異なる時間、あるいは試料が異なる場合は、構築したスペクトルライブラリの保持時間の誤差はもっと大きくなると考えられます。

## 定量比較

ここまでの Skyline によるデータ処理に全般的に大きな問題があるかどうかについて検証してきました。作成された mProphet スコアモデルには必要なすべての評価スコアが付き、ランダムマッチの程度を評価するのに用いられた標的ペプチドシーケンスとデコイペプチドシーケンスに対するマッチの分布も十分に分離していました。RT の抽出に使用された許容範囲を  $\pm 5$  から  $\pm 3$  に狭くできたのではないかと思うかも知れませんが、現在の設定でも明らかにうまくいっており、作成された mProphet モデルは許容できるように見えます。

ペプチドが数百種類にも及ぶ場合には、1つ1つのペプチドについてマニュアルで分析するのは時間がかかり、間違いを起し易くなります。このチュートリアルで行っている解析を、3つの種すべての FASTA ファイルを対象とすると、いわゆる proteome-wide 解析となり、数万のペプチドを解析することになります。このような場合は、個々のペプチドについて比較するより、何らかの方法で群間比較を行い、興味深い変化をしているペプチドあるいはタンパク質についてさらに詳細な解析を行うことが一般的です。

Skyline で簡単な 2 群間比較を実施するには、以下の操作を行います。

- [表示]メニューで、[その他のグリッド]、続いて[グループ比較]を選択し、[追加]をクリックします。
- [名前]フィールドに「By Condition」と入力します。
- [グループ注釈をコントロール]フィールドで「Condition」を選択します。
- [グループ値をコントロール]フィールドで「A」を選択します。
- [次に対して比較する値]フィールドで「B」を選択します。
- [詳細]ボタンをクリックします。
- [Q値のカットオフ]フィールドに、「0.01」と入力します。

[グループ比較を編集]フォームは以下のようになります。

グループ比較を編集

名前(N):  
By Condition

グループ注釈をコントロール(C)  
Condition

グループ値をコントロール(Q)  
A

次に対して比較する値(V):  
B

同一注釈(技術的繰り返し用)(D):  
BioReplicate

正規化メソッド(M):  
なし

信頼度レベル(E):  
95 %

適用範囲  
 ペプチド(P)  タンパク質(T)

要約手法:  
トランジション領域の合計

欠損しているピークはゼロを使用する

Q値のカットオフ  
カットオフより大きな検出のQ値を有するピークは欠けているものとして取り扱います。  
0.01

OK キャンセル

- [OK]ボタンをクリックします。

今作成したグループ間比較の結果を表示するには、以下の操作を行います。

- [表示]メニューで、[その他のグリッド]、続いて[グループ比較]を選択し、[By Condition]をクリックします。

A と B の試料混合物を比較した場合の、ペプチド倍率変化と調整された P 値（偽発見率 FDR の推定値）を表示する表が表示されます。

- タンパク質名とその生物種名が完全に表示されるよう、[タンパク質]列ヘッダーの幅を広げます。
- [倍率変化結果]列のヘッダーをクリックし、[昇順にソート]をクリックします。

どの生物種からのものか、またその予想比率（ヒト 1:1、酵母 1:2、E. coli 4:1）を念頭に、表内の複数のペプチドについて推定されている倍率変化を調べます。調整された P 値を調べ、予想される結果について考えてみましょう。

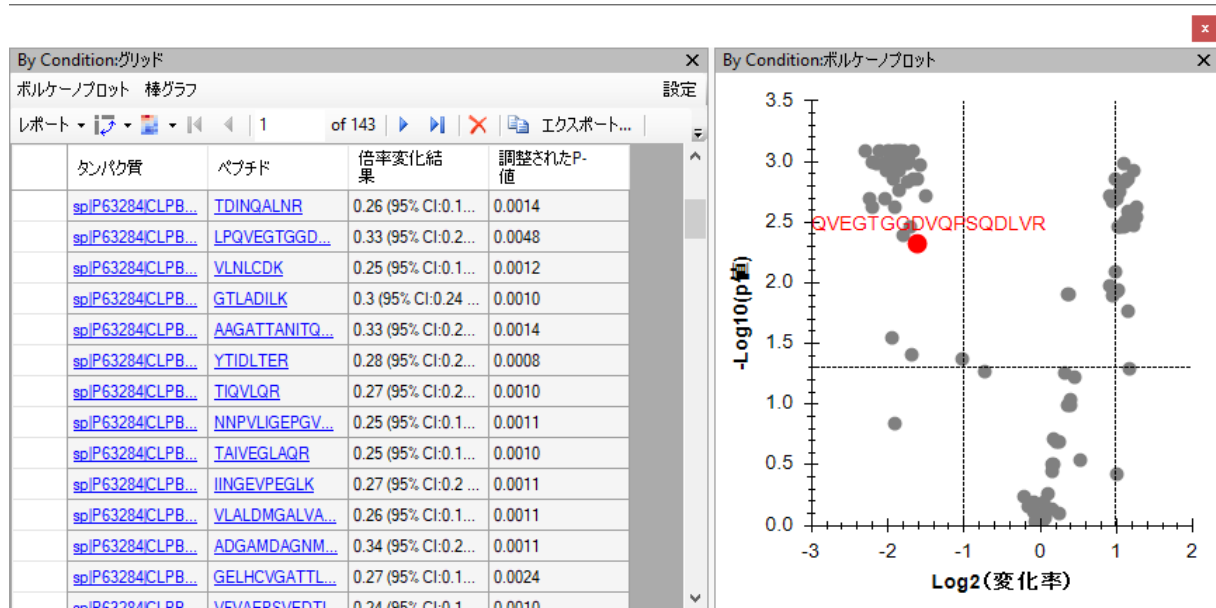
By Condition:グリッド

ボルケーノプロット 棒グラフ

レポート | 1 of 143 | エクスポート... | 検索:

タンパク質	ペプチド	倍率変化結果	調整されたP-値
<a href="#">sp P63284 CLPB_ECOLI</a>	<a href="#">TDINQALNR</a>	0.26 (95% CI:0.19 to 0.38)	0.0014
<a href="#">sp P63284 CLPB_ECOLI</a>	<a href="#">LPQVEGTGGD...</a>	0.33 (95% CI:0.21 to 0.52)	0.0048
<a href="#">sp P63284 CLPB_ECOLI</a>	<a href="#">VLNLCDK</a>	0.25 (95% CI:0.18 to 0.35)	0.0012
<a href="#">sp P63284 CLPB_ECOLI</a>	<a href="#">GTLADILK</a>	0.3 (95% CI:0.24 to 0.38)	0.0010
<a href="#">sp P63284 CLPB_ECOLI</a>	<a href="#">AAGATTANITQ...</a>	0.33 (95% CI:0.25 to 0.44)	0.0014
<a href="#">sp P63284 CLPB_ECOLI</a>	<a href="#">YTIDLTER</a>	0.28 (95% CI:0.23 to 0.35)	0.0008
<a href="#">sp P63284 CLPB_ECOLI</a>	<a href="#">TIQVLQR</a>	0.27 (95% CI:0.21 to 0.35)	0.0010
<a href="#">sp P63284 CLPB_ECOLI</a>	<a href="#">NNPVLIGEPGV...</a>	0.25 (95% CI:0.18 to 0.35)	0.0011
<a href="#">sp P63284 CLPB_ECOLI</a>	<a href="#">TAIVEGLAQR</a>	0.25 (95% CI:0.19 to 0.34)	0.0010
<a href="#">sp P63284 CLPB_ECOLI</a>	<a href="#">IINGEVPEGLK</a>	0.27 (95% CI:0.2 to 0.36)	0.0011
<a href="#">sp P63284 CLPB_ECOLI</a>	<a href="#">VLALDMGALVA...</a>	0.26 (95% CI:0.19 to 0.35)	0.0011
<a href="#">sp P63284 CLPB_ECOLI</a>	<a href="#">ADGAMDAGNM...</a>	0.34 (95% CI:0.26 to 0.43)	0.0011
<a href="#">sp P63284 CLPB_ECOLI</a>	<a href="#">GELHCVGATTL...</a>	0.27 (95% CI:0.18 to 0.41)	0.0024
<a href="#">sp P63284 CLPB_ECOLI</a>	<a href="#">VFVAEPSVEDTI...</a>	0.24 (95% CI:0.18 to 0.33)	0.0010

- グリッドウィンドウの左上隅にある [ ボルケーノプロット ] ボタンをクリックします。



表示されたボルケーノプロットでは、変化している点の多くが、予想される倍率変化の周辺に集まっており、また、5% FDR を示す水平線の上にあることがわかります。どの点がどの生物種に属するのかを明確にするため次の操作を行って下さい。

- [ ボルケーノプロット ] を右クリックし、[ 形式 ] をクリックします。
- [ 表現 ] 列の後の最初の行にある [...] ボタンをクリックします。
- [ 一致式を作成します ] フォームで、[ 合致 ] ドロップダウンリストを「タンパク質名」に設定します。
- [ 正規表現 ] フィールドに「ECOLI」と入力します。

フォームには、以下のように一致するタンパク質名とともにペプチドの一覧が表示されます。

一致式を作成します

正規表現(R): ECOLI      合致(M): タンパク質名

フィルタ  
 倍率変化カットオフ(E): なし      P値カットオフ(P): なし

一致したルール

- TDINQALNR (sp|P63284|CLPB\_ECOLI)
- LPQVEGTGGDVQPSQDLVR (sp|P63284|CLPB\_ECOLI)
- VLNLCDK (sp|P63284|CLPB\_ECOLI)
- GTLADILK (sp|P63284|CLPB\_ECOLI)
- AAGATTANITQAIQMR (sp|P63284|CLPB\_ECOLI)
- YTIDLTER (sp|P63284|CLPB\_ECOLI)
- TIQVLQR (sp|P63284|CLPB\_ECOLI)

OK      キャンセル

- [OK] ボタンをクリックします。
- [RGB] 列の後の最初の行にある [...] ボタンをクリックします。
- 紫を選び、[OK] ボタンをクリックします。
- 「YEAS」について同じ処理を繰り返してオレンジにします。
- 「HUMAN」について同じ処理を繰り返して緑にします。

これで [ボルケーノプロット形式] フォームは以下ようになります。

ボルケーノプロット形式

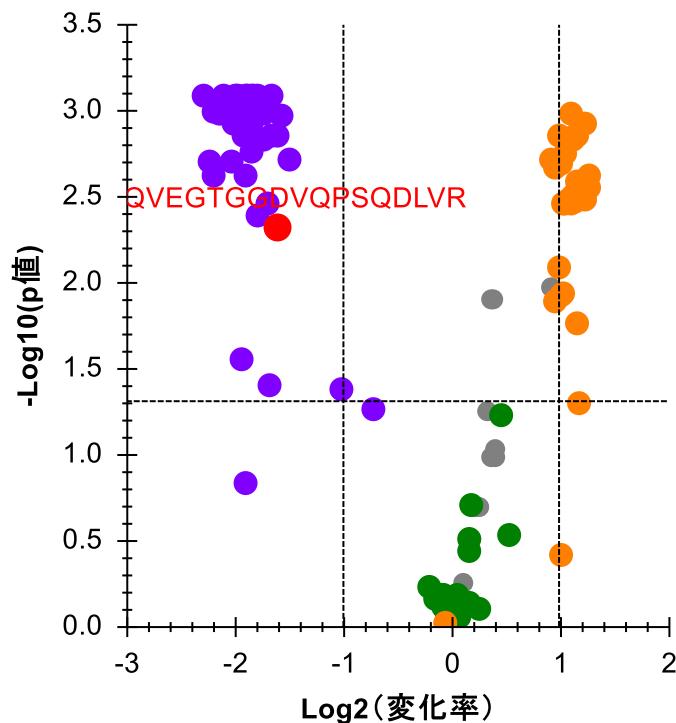
表現	RGB	記号	サイズ
ProteinName: ECOLI	128, 0, 255	円形	標準
ProteinName: YEAS	255, 128, 0	円形	標準
ProteinName: HUMAN	0, 128, 0	円形	標準

RGB       詳細

OK      キャンセル

- [OK] ボタンをクリックします。

これでボルケーノプロットは以下ようになります。



まだグレーのドットがいくつか残っていることに注意してください。これらのドットのどれかをクリックすると、[ターゲット]表示でそのペプチドが選択され、ボルケーノプロットでクリックしたドットが赤に変わります。また、任意のドットにマウスのカーソルを合わせると、そのドットに関する詳細情報が表示されます。これらの操作で、グレーのドットはIRT標準ペプチドであることがわかります。これらの標準ペプチドの情報は、以下の操作を行うと、グリッドとボルケーノプロットから除去できます。

- [タンパク質]列ヘッダーをクリックし、[フィルタ]をクリックします。
- [フィルタタイプ]ドロップダウンリストで、「含まない」を選びます。
- その下のフィールドに「standard」と入力します。
- [OK] ボタンをクリックします。

すると、ボルケーノプロットからグレーのドットが削除されます。同様にして、タンパク質名に他のフィルタを適用することで、ボルケーノプロットに1つの生物種のペプチドだけを表示させることもできます。

Skyline はまた、倍率変化値に対して（未補正）信頼区間の誤差バーを示す棒グラフを表示させることもできます。それらは多重仮説検定により補正されてはいませんが、測定のバラツキを知るのに役に立つでしょう。棒グラフで倍率変化値を表示させるには、以下の操作を行います。

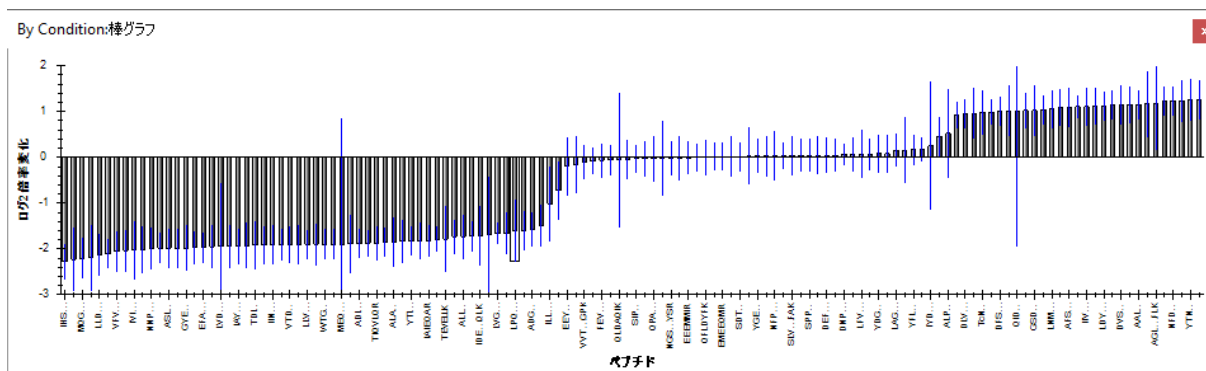
- グリッドウィンドウの左上隅にある[棒グラフ]ボタンをクリックします。



棒グラフがボルケーノプロットの上に表示されます。もうすこし表示が広がるように、以下の操作を行います。

- [ By Condition:棒グラフ ] タブをクリックしたまま新しい場所にドラッグしてマウスから指を離します。
- フロート表示になった [ By Condition:棒グラフ ] ウィンドウのサイズを変更します。
- [ By Condition:グリッド ] で、[ 倍率変化結果 ] 列ヘッダーをクリックし、[ 昇順にソート ] をクリックします。

グラフは以下のように表示されます。



ボルケーノプロット同様、個々の棒をクリックすると [ ターゲット ] 表示で選択されます (グラフ内のさまざまなペプチドがどの生物種からのものかを確認できます)。

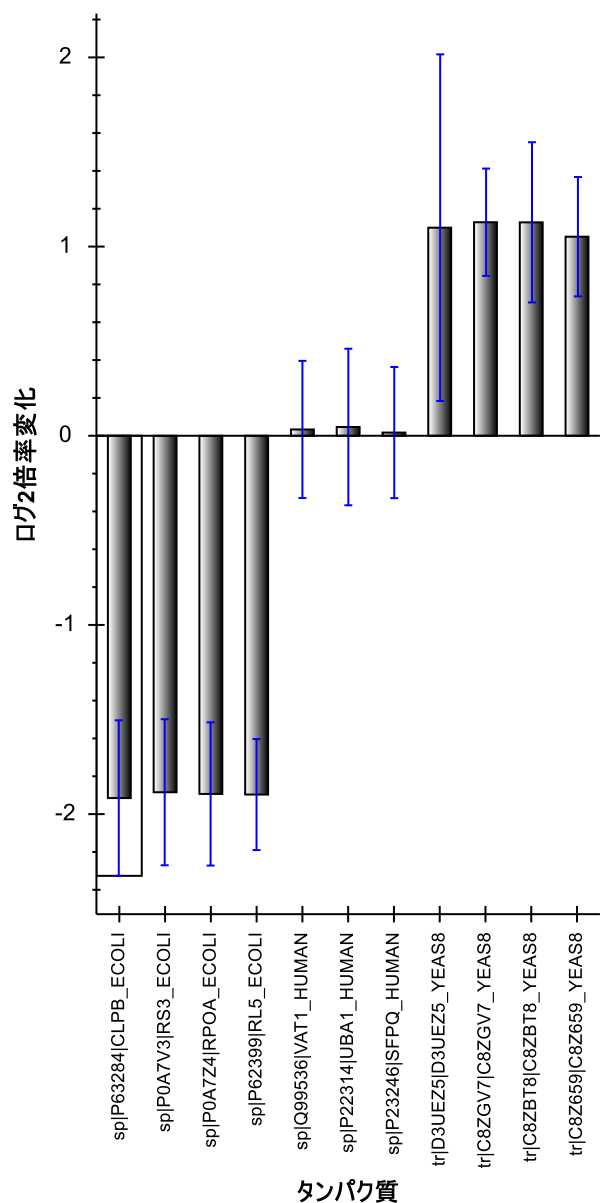
11 のターゲットタンパク質 (iRT 標準とデコイの結果を除いています) のタンパク質レベルの倍率変化結果を表示するには、以下の操作を行います。

- グリッドウィンドウの右上隅にある [ 設定 ] ボタンをクリックし、[ グループ比較を編集 ] フォームをもう一度開きます。
- [ 適用範囲 ] オプションを [ ペプチド ] から [ タンパク質 ] に変更します。

Skyline はすぐに棒グラフとボルケーノプロットを変更します。棒グラフには、棒の数が予想していた 11 本ではなく 9 本しかないことがわかります。これは、現在の設定では、t-検定を行うために必要な数の測定値が揃っていない状態であり、欠損値 (Q 値のカットオフ 0.01 による) に対応することができないためです。Skyline は欠損値を持つターゲットタンパク質を除去して表示しています。これを修正するには、以下の操作を行います。

- グリッドウィンドウの右上隅にある [ 設定 ] ボタンをクリックし、[ グループ比較を編集 ] フォームを開きます。
- [ 要約手法 ] を「チューキーの中央値分散分析」に変更します。

これで棒グラフが 11 のタンパク質すべての倍率変化推定値を、以下のように表示するようになります。



- 右上隅にある X をクリックして [グループ比較を編集] を閉じます。
- ドキュメントを保存します。

また、MSstats のような高機能の統計ツールも使用することができます。MSstats は、Skyline Tool Store からインストールできます (Skyline メニューで [ツール] > [ストア] を使用するか、Skyline インストールウェブページで [Tool Store] ボタンをクリックして下さい)。インストール後は、Skyline の [ツール] メニューから直接 MSstats を実行するか、以下の操作を行って、R または RStudio で MSstats を実行してください。

- [ファイル] メニューで、[エクスポート] を選択して [レポート] をクリックします。
- 外部ツールフォルダで **MSstats Input** を選択します。

- 「DIA\_TTOF\_MSstats\_Input.csv」として保存します。

## まとめ

おめでとうございます。これで、DIA データを使用して、Skyline でプロテオームの差異発現比較解析を行うチュートリアルは終了です。ここでは、ペプチド検索のインポートウィザードを利用して、DDA ペプチド検索データからスペクトルライブラリを構築する過程から始めて、定量的 DIA 測定結果からクロマトグラムを抽出し、mProphet の統計学的ピーク検出モデルを使用して、ピーク選択を改善する方法と群間比較に使用するピーク領域値で偽陽性率 (false discovery rate) を使用するための q 値を割り当てるところまで学びました。最初の段階で、DIA のデータの質を質量誤差と保持時間の回帰分析により評価することを学びました。最終段階では、Skyline を用いて単純な 2 群間比較を行う方法について学びました。Skyline は、インタラクティブな群間比較グリッド、ボルケーノプロット、棒グラフを表示することでユーザーの理解を容易にし、結果を検証することができます。また、MSstats input をエクスポートし、MSstats を使用して異なる統計検定を行うことも可能です。次の段階として、全 Proteome の FASTA ファイルを用いて、多くのターゲットを対象にした解析を同様に行うことや、プレカーサーイオンを対象に含めることで異なるレベルでの定量解析 (MS1 スペクトルを用いるプレカーサーイオン) を行うこともできます。あるいは、ご自分の測定結果を用いて定量解析を行ってもいいでしょう。