

# Skyline を使用したグループ化研究データの処理

---

## パート 1: メソッド最適化のための標識なし差分法

ターゲットプロテオミクスは様々なグループ間あるいは生命体の状態間でのタンパク質/ペプチドの発現の差異を測定するのに長けているため、研究アプローチの一つとして広く普及し始めています。元々、Skyline はターゲットプロテオミクスを効率的に応用して疾病バイオマーカー候補タンパク質を検出し、その臨床的検証を得ることを証明するための試みとして、NCI の癌に関する臨床プロテオミクス技術評価 (Clinical Proteomics Technology Assessment for Cancer, CPTAC) プログラムから出資を受けていました。

本チュートリアルでは Skyline を使用して、塩分感受性ラットモデルを用いた心不全の実験的ケースコントロール研究で得られた SRM データを解析します。本研究では、文献から得られた心不全に関連するタンパク質候補のリストを活用して、罹患群と健常群との間の血漿内タンパク質発現の差を同定しています。

本実験では、42 回の LC-MS/MS 注入を行い、49 種類のタンパク質に由来する 137 種類のペプチドを観察しました。本チュートリアルで説明されているデータ処理および可視化技術に一度慣れてしまえば、より多くの実験群でより多くのターゲットを扱う大規模な研究のデータ管理・評価を比較的簡単にできるようになります。しかし本チュートリアルなしでは、比較的小規模なデータセットを扱う作業でさえ、その煩雑さゆえ容易に圧倒されてしまうということが起こり得ます。

本チュートリアルは以下のセクションからなります：

- はじめに
- メソッド最適化後の差分解析
- 参照標準なしで複数の繰り返し測定データ処理を設定する
- スケジュール化 SRM およびピーク切断
- 複数の繰り返し測定データの処理を開始する
- グローバル正規化標準
- 複数の繰り返し測定データ処理を続行する
- 繰り返し注釈のある統計分析を作成する
- 問題ピークのあるペプチドに注釈を付ける
- Skyline 内での最初の複数の繰り返し測定検査
- 結論

## はじめに

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

<https://skyline.gs.washington.edu/tutorials/GroupedStudies1.zip>

この中のファイルを、次のようにコンピュータ上のフォルダで解凍します:

C:\Users\brendanx\Documents

これにより次の新しいフォルダが作成されます:

C:\Users\brendanx\Documents\GroupedStudies1

この新規フォルダのコンテンツを調べていくと、サブフォルダが含まれていることが分かります:

- Heart Failure – Michael J. Toth および Michael J. MacCoss のラボにより提供された、塩分感受性ラットモデルにおける心不全研究のファイルです。

以下のチュートリアルに従い、この種の研究データの処理方法を学習し、研究のタンパク質/ペプチドターゲットおよび全体的なデータ品質について理解を深めてください。

## メソッド最適化後の差分解析

作成した直後の「GroupedStudies1」フォルダの「Heart Failure」サブフォルダ内のデータは、[ターゲットメソッドの最適化](#)チュートリアルで説明されているアプローチの可能性や限界を調べる際に MacCoss ラボで収集されたものです。このデータは、さまざまな理由により論文に掲載されませんでした。一部の理由については、Skyline を利用して収集データを理解していくと明らかになってくるでしょう。しかし、このターゲットメソッドの最適化およびその他の類似実験の試みから生まれたアイディアの多くは、このデータセットとは別に論文で報告されました。<sup>1</sup>

元の研究では、文献上で既に示されている、心不全に関与する 109 のタンパク質が選択されました。これらのタンパク質は Skyline ドキュメントに追加されました。その一部は、タンパク質シーケンスの FASTA 形式のテキストをインポートし、Skyline がトリプシン消化を *in silico* で実行できるようになっています。長さ 6~30 アミノ酸で開裂欠損のないペプチドを全て含むためのパラメータが、Skyline 内で設定されました。その他のタンパク質については、Skyline を使用しない *in silico* 処理を実行しました。これらのタンパク質から得られたペプチドリストが、Skyline ドキュメントに直接加えられました ([ターゲットメソッドの編集](#)チュートリアルをご覧ください)。これにより、2,165 のターゲットペプチドを持つドキュメントが作成されました。2 価の各プリカーサーについては、 $y_3 \sim y_{(n-1)}$  に対応する一価プロダクトイオンが検討されました。最初の包括的メソッドでは 12,194 のトランジションをカバーし、プールした

ラット血漿試料を未スケジュール化 SRM メソッドで測定しました。この最初の分析は一度だけ実行され、151 回の個別の質量分析注入を要しました。

得られた raw データファイルを Skyline へとインポートし、試料処理をこれ以上行わずに血漿マトリックス内でどのターゲットペプチドが検出に適しているかを判定しました。元の 2,165 のターゲットペプチドのうち、ターゲット y-イオン（3 イオンで 2、4 イオンで 5、5 イオンで 27、および 6 イオンで 103）で十分に共溶出し、フル勾配クロマトグラムにおいてピークを有していたのは、135（49 のタンパク質由来）のみでした。残りの 780 のトランジションについて、90 分勾配で検出されたピーク前後の 5 分間の保持時間ウィンドウを利用して、一回の分析で定量化するようスケジュールしました。このスケジュール化メソッドを使って、14 体の塩分感受性ラット（7 体が健常、および 7 体が塩分過多の食事による心不全を罹患）の血漿試料を分析し、各試料について 3 回反復測定し、計 42 回の試料注入を実行しました。本チュートリアル全体を通して解析することになるこれらのデータは、135 種類の検出可能なペプチドのうち 2 グループ間の差の見込みを示すものを判定し、それらを今後の研究において潜在的バイオマーカーの候補としたものです。

本研究の全体的な目標は、ターゲットタンパク/ペプチドの予備知識をあまり多く持たず、又は安定同位体標識した参照ペプチドを使用せずに、SRM を使用したターゲット解析から、役立つ科学的洞察が得られるかどうかを判定することです。

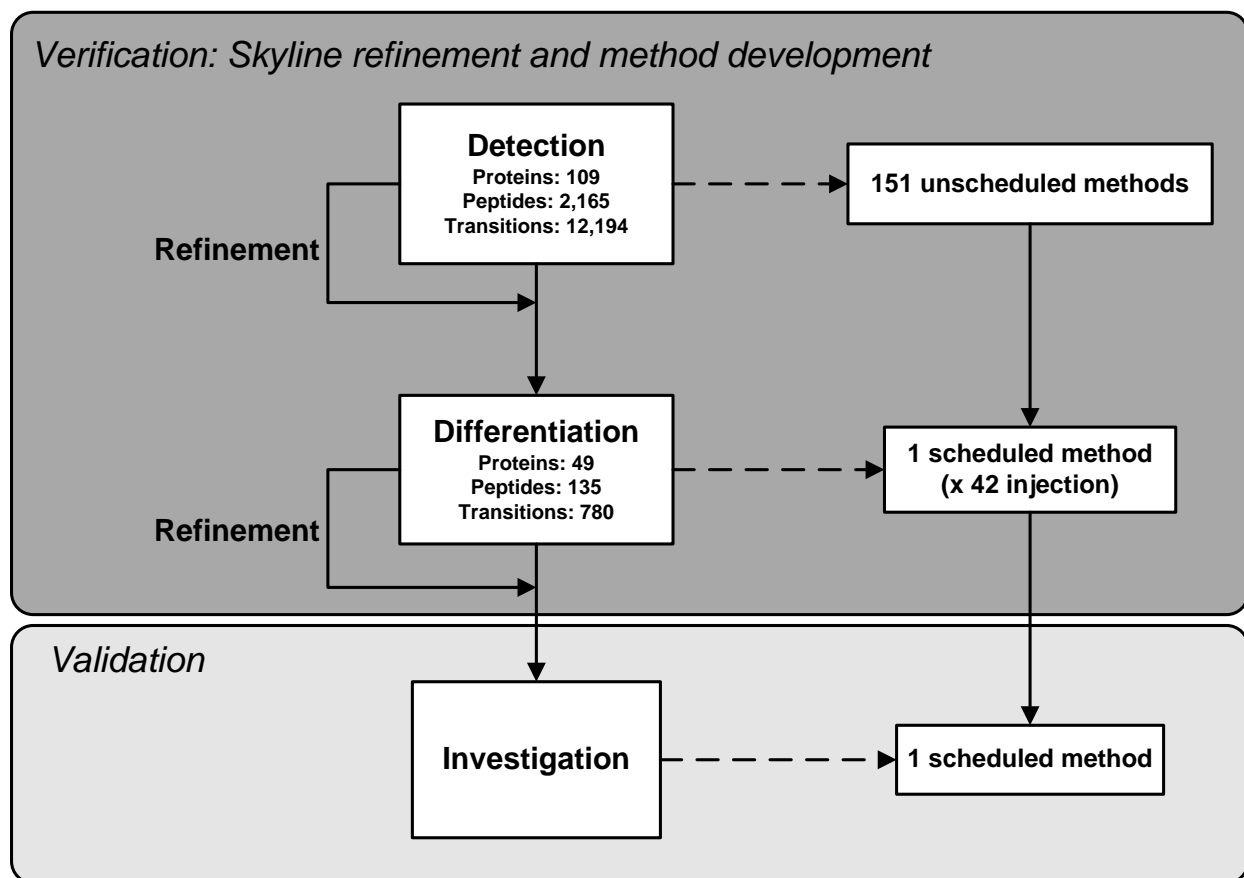


図 1: 塩分感受性ラットモデルの心不全研究において、ターゲットメソッド最適化の「検出」および「差分」フェーズを示す概略図。



このメソッド最適化研究の差分フェーズのため収集されたデータの処理を開始するには:

- 「GroupedStudies1\Heart Failure\raw」サブフォルダ内の「Rat\_plasma.sky」ファイルを開きます。

Skyline ウィンドウの右下角の数値をチェックすると、開いたファイルには 49 のタンパク質、137 のペプチド、および 789 のトランジションが含まれているのが分かります。



1/49 prot    1/137 pep    1/137 prec    1/789 tran ...

これは上記の図（論文発表されていない図）で示されたものとはかなり異なり、またドキュメントには追加のペプチドリストが 1 つ含まれています（下の方に「S」と名付けられているもの）。同リストにはグローバル正規化ペプチドが 3 つ含まれています。これらは、試料内のすべてのペプチドに影響を与えてしまう、分析における体系的変動の影響を低減させることを目的としています。これについては、後で詳しく触れます。また、どういうわけかペプチドを 1 つ持つタンパク質が 1 つ、上記で説明したファイルから無くなってしまっています。

また、一部のペプチドにはライブラリスペクトルと一致するものがあることにも気づかれるでしょう。これは、[ターゲット] ビューで見ることができます。右下角のスペクトルラインのあるペプチドアイコン（)には一致スペクトルがあり、ラインなしのペプチドアイコン（)には一致スペクトルはありません。当該スペクトルは 2 つの異なるスペクトルライブラリからのもので、1 つは NIST、もう 1 つはグローバルプロテオームマシン (Global Proteome Machine, GPM) からのものです。以下を行ってこれらの範囲を探索できます:

- [ビュー] メニューで、[ドキュメントグリッド] をクリックします。
- [ドキュメントグリッド] フォームの左上角の [ビュー] メニューで、[プリカーサー] をクリックします。
- [ドキュメントグリッド] の右側をスクロールして「ライブラリ名」列を見つけて、2 回クリックして降順に並べ替えます。

ここでグリッドをスクロールダウンすると、ライブラリスペクトルを用いたペプチドの数 (80 個) を見ることができます。これらは、ラット (NIST) ライブラリ (49 個) とラット (GPM) ライブラリ (31 個) の各スペクトルデータに由来します。注記: 選択したセルの行番号は、グリッドの上にあるツールバー内に表示されます。

ビュー -  80 of 137 

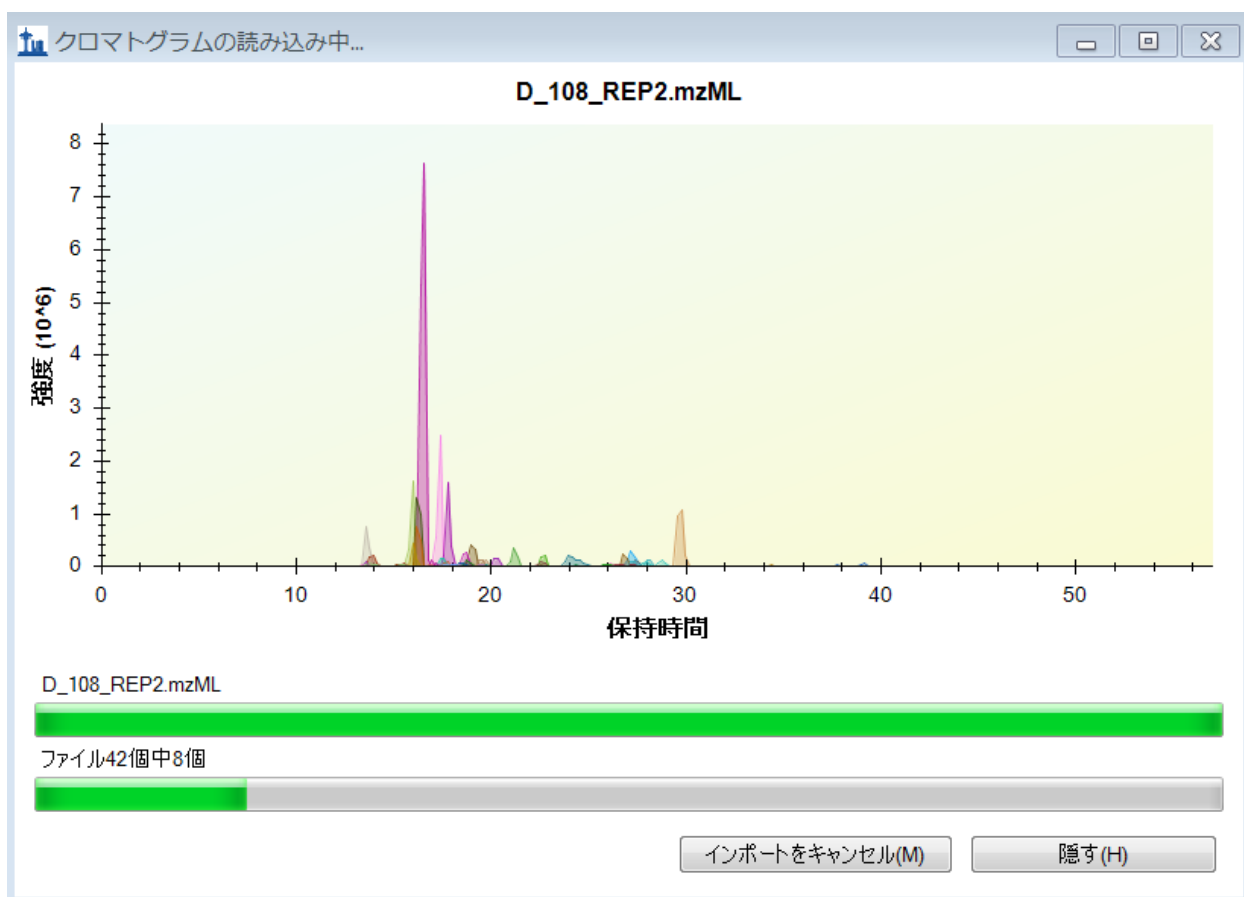
現在の例については、この情報により、この実験の基となった予備知識がわずかに得られます。

## 参照標準なしで複数の反復測定データ処理を設定する

14 個の被験体について技術的反復測定を 3 回行って取得した SRM データの処理を始めるには、以下を行います。

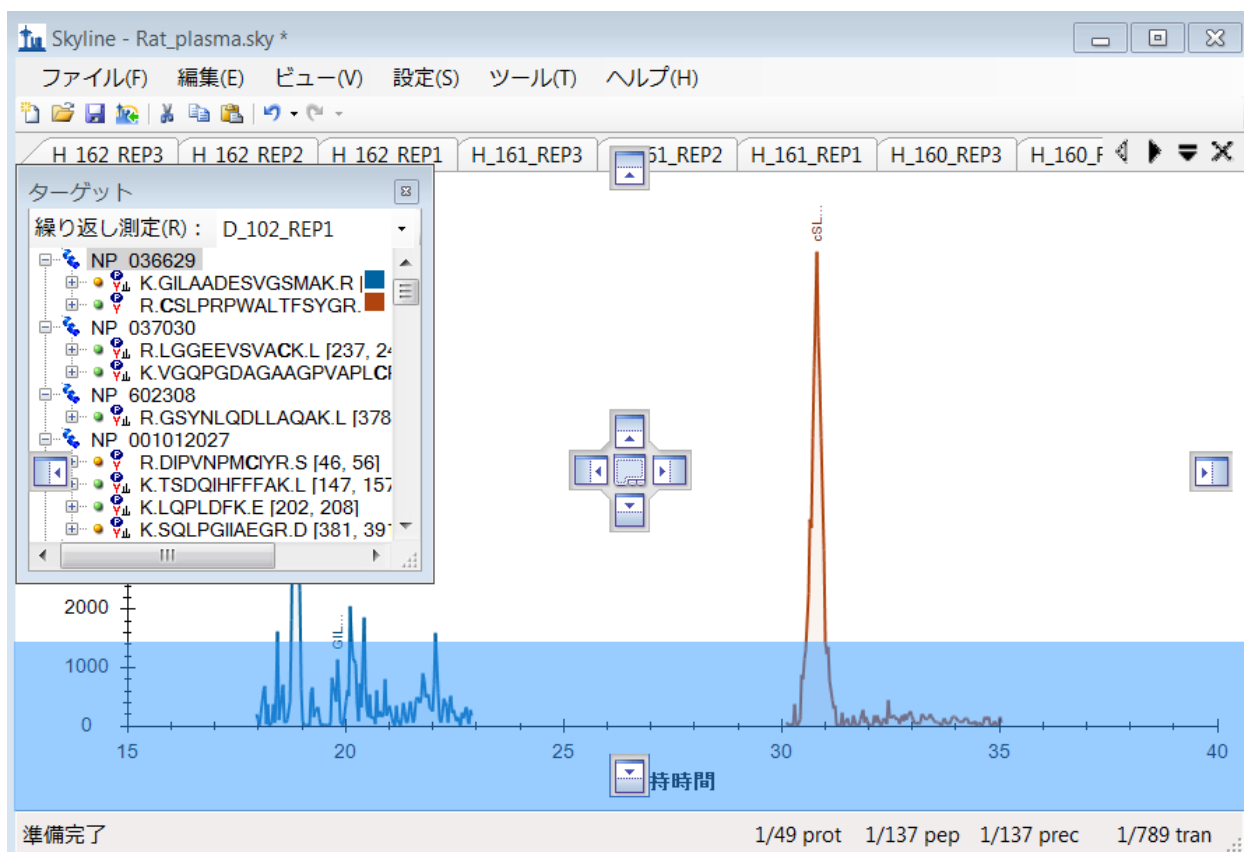
- [ドキュメントグリッド] フォームを閉じます。
- [ファイル] メニューで、[インポート] を選択して [結果] をクリックします。
- [結果をインポート] フォーム内の [OK] ボタンをクリックして、1 回注入された反復測定結果をインポートします。
- [結果ファイルをインポート] フォーム内の「D\_102\_REP1.raw」という名の最初の raw データファイルをクリックします。
- Shift キーを長押ししながら「H\_162\_REP3.raw」という名の最後の raw データ ファイルをクリックして、「GroupedStudies1\Heart Failure\raw」フォルダのすべての raw データ ファイルを選択します。
- [開く] ボタンをクリックします。

ファイルの読み込みが始まり、Skyline に以下のようなウィンドウが現れ進行状況が表示されます：



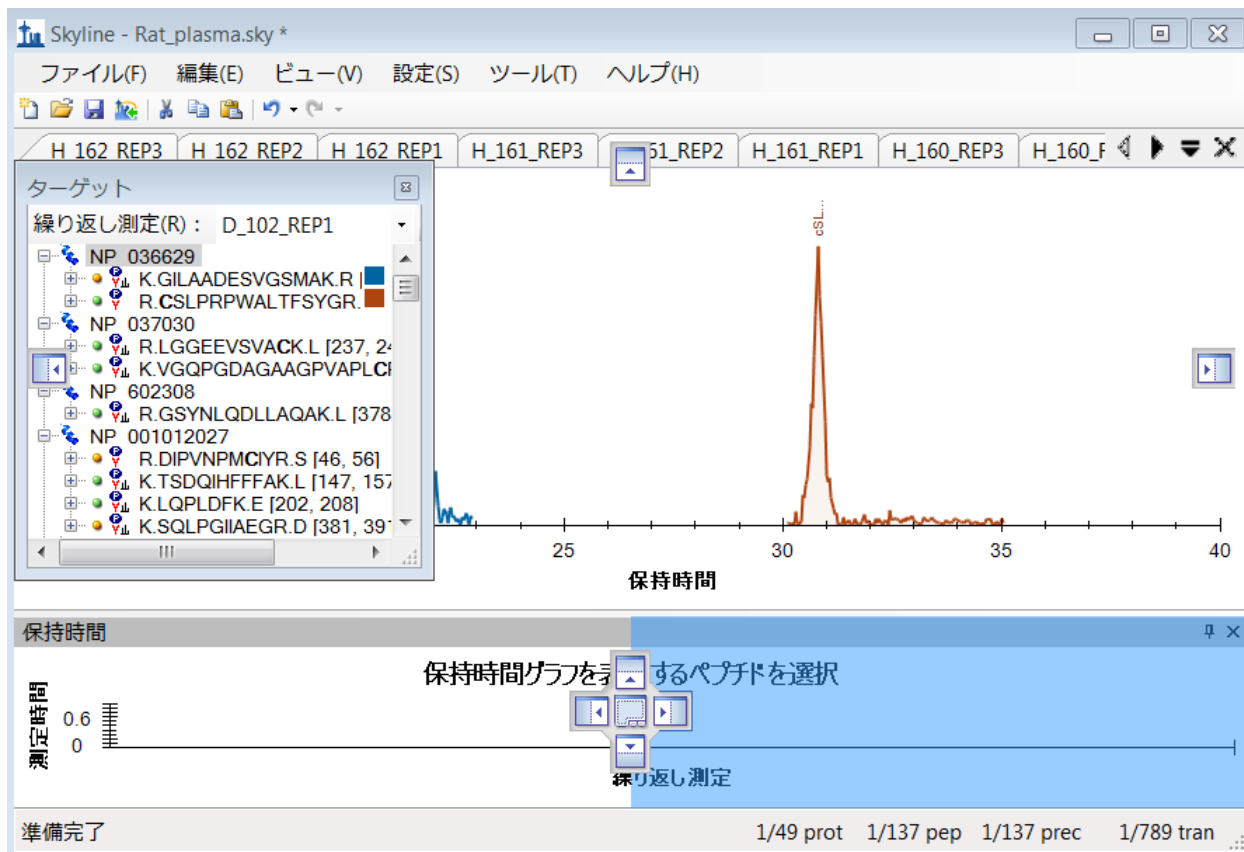
Skyline が SRM クロマトグラムをインポートしている間にデータ処理の準備を続行するには、以下の操作を行います：

- [隠す] ボタンをクリックします。
- [ライブラリー致] ビューを閉じます。
- [ターゲット] ビューのタイトルを、マウスのカーソルを追跡する可動式の青い長方形が見えるまで、クリック&ドラッグします。
- マウスの左ボタンから指を放し、[ターゲット] ビューが Skyline ウィンドウの上に表示されているままにします。
- [ビュー] 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。
- [保持時間] ビューのタイトルを、ドッキングインジケータが見えるまでクリック&ドラッグします。
- カーソルを下部ドッキングインジケータへと、以下のように、青い長方形の枠が Skyline ウィンドウの下端を覆うように見えるまでドラッグします。

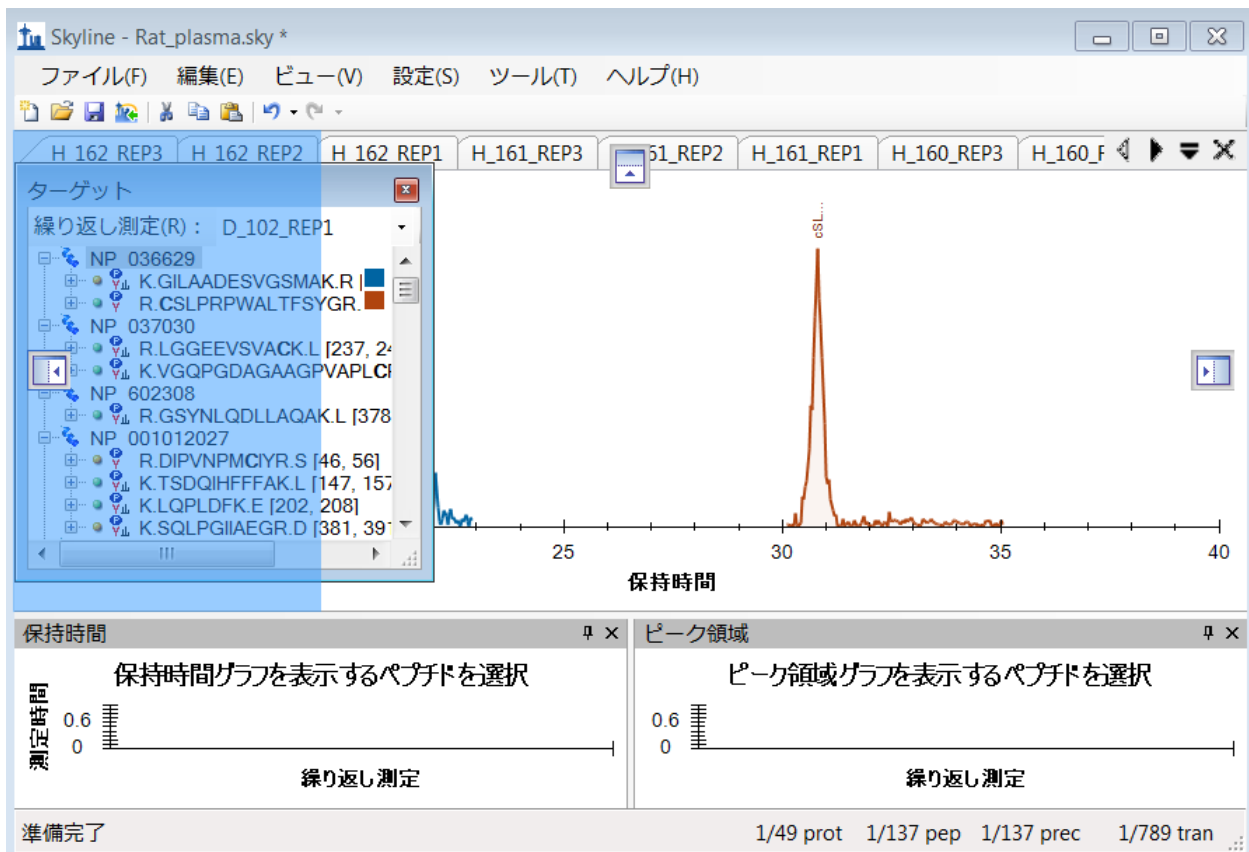


- マウスの左ボタンから指を放し [保持時間] ビューを ドッキングします。
- [ビュー] メニューで、[ピーク領域] を選択して [繰り返し測定比較] (F7) をクリックします。

- [ピーク領域] ビューを、マウスマウスカーソルがドッキングインジケーター内に入るまでドラッグすると、青い長方形枠が [ピーク領域] ビューの下端を共有することが示されます。



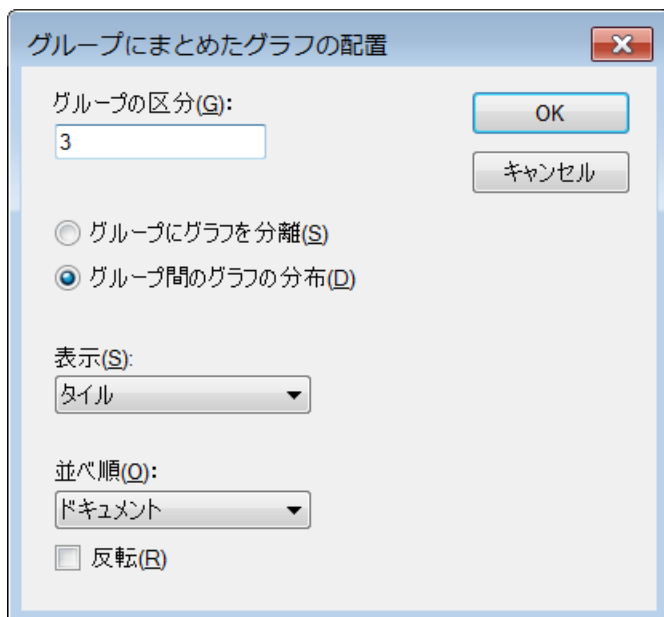
- マウスの左ボタンから指を放し [ピーク領域] ビューをドッキングします。
- [ターゲット] ビューを、マウスマウスカーソルがドッキングインジケーター内に入るまでドラッグすると、青い長方形枠が Skyline ウィンドウの左端とドッキングすることが示されます。



- メインの Skyline ウィンドウの外側で表示されている、その他のウィンドウを閉じます。
- [ビュー] 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。
- [グループにまとめたグラフの配置] フォーム上の [ペインのグループ化] に「3」と入力します。
- [グループ間のグラフの分布] オプションをクリックします。
- [並び順<sub>[M1]</sub>] ドロップダウンリストで、「ドキュメント」を選択します。

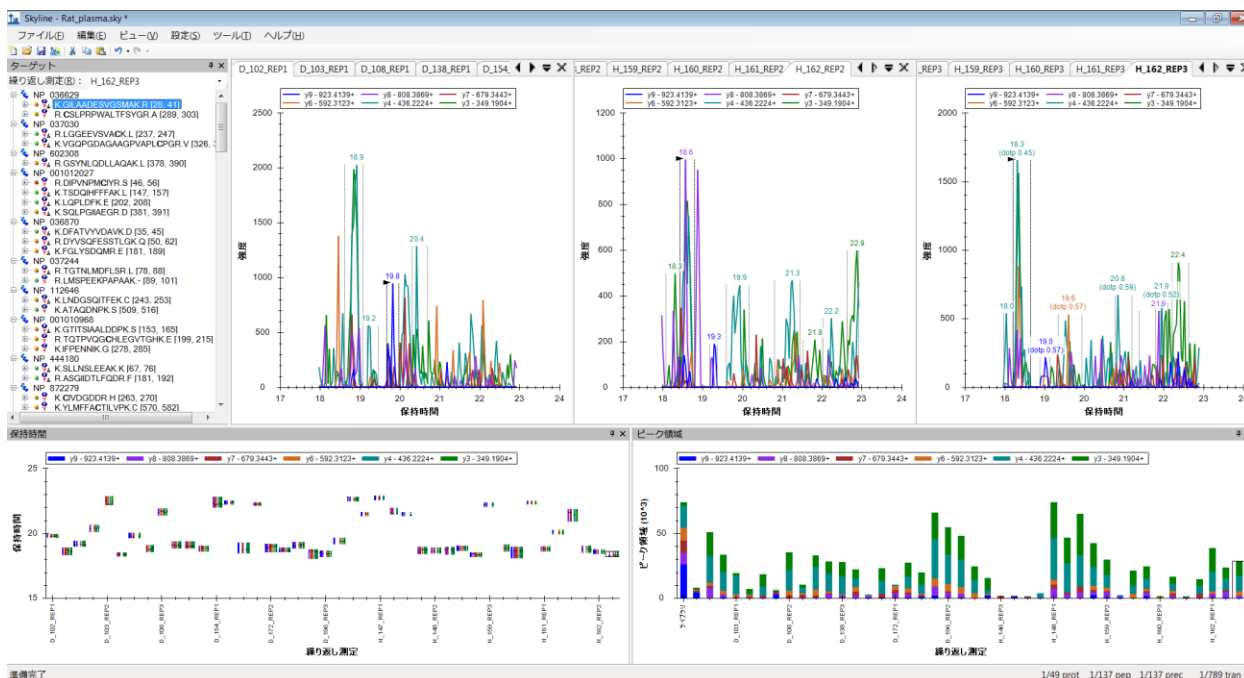
フォームは以下のように見えるはずです:





- [OK] ボタンをクリックします。
- Skyline メインウィンドウを最大化します。
- ドキュメント内の最初のペプチドを選択します。

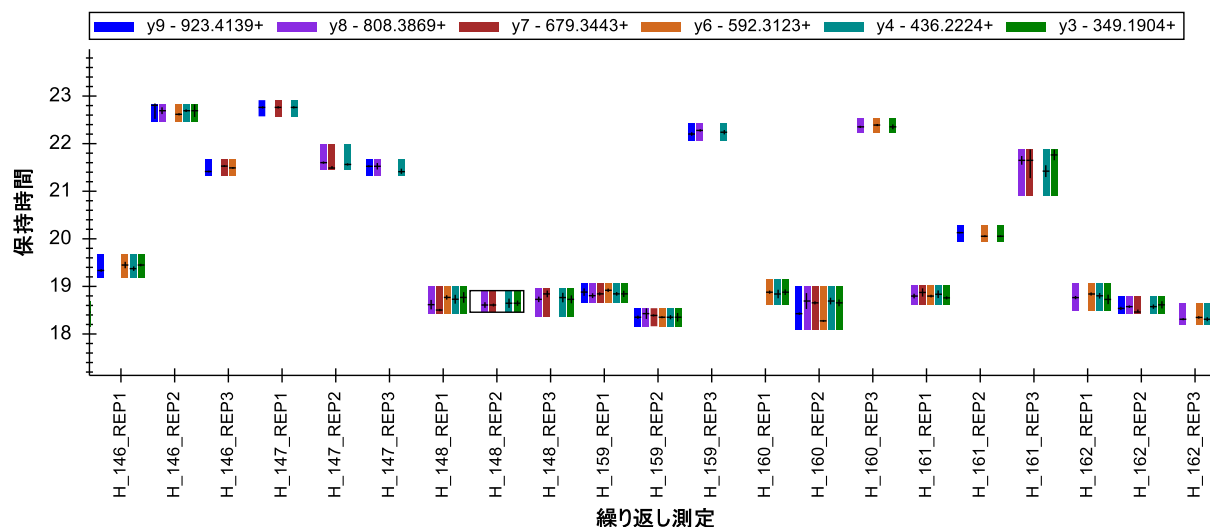
これで Skyline ドキュメントは以下のように見えるはずです:



注: この画像は、解像度 1920x1080 の 24 インチのモニタ上でキャプチャしたものです。本チュートリアルをデジタル方式でご覧になっている場合、200%以上ズームインしてこの画像を見ることをお勧めします。これ以降、本チュートリアルでは元通り、8½ x 11 インチのページ

レイアウトにより適した画像を使用していきます。多くの Skyline ワークショップは 1024x768 のスクリーン解像度で行われていますが、より大きいモニタの方が Skyline をもっと利用しやすくなるでしょう。

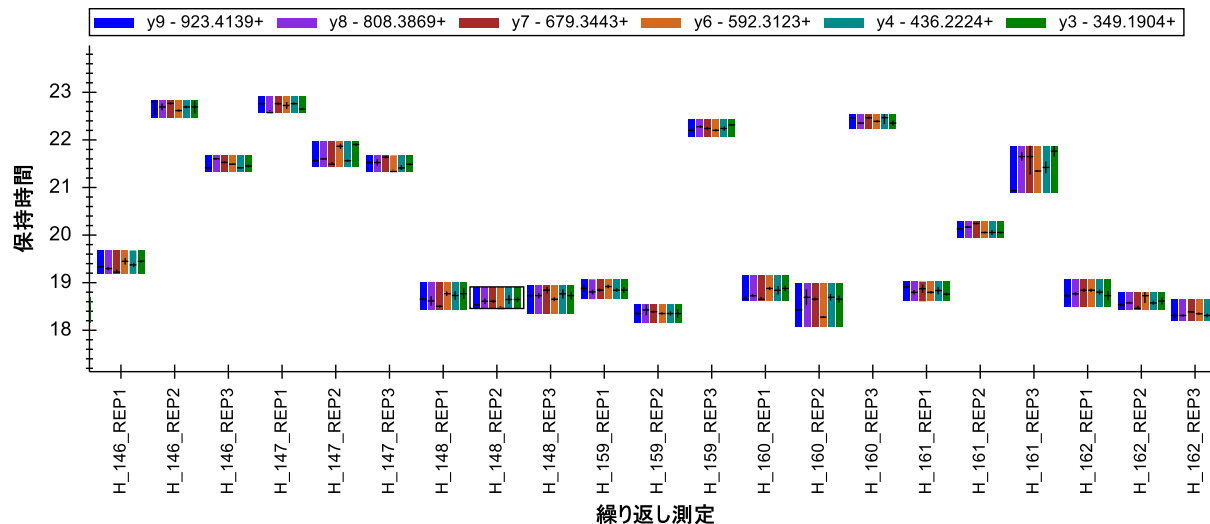
この最初のペプチド、K.GILAADESVGSMK.R [28, 41] の [保持時間] ビューに注目してみると、積分領域があまり一貫していないことが分かります。選択されたピークのほとんどは約 19 分で溶出します。しかし、3分の1ほど（12）は 22 分付近で溶出します。また、このプロットでは多くのカラーバーが欠けていることにも気づかれるでしょう：



これは、Skyline が未だ**最適化モード**であるからです。一方このモードでは、Skyline は、どのピークが共溶出と見なすのに十分類似した形状および頂点時間を有しているかを決定しようと試みます。欠けているバーは、積分範囲内のピークが、計測された全てのトランジションとあまり一致していないことを示します。実験のこの段階について、どのトランジションを測定するかを選択する [最適化モード] から、Skyline がすべてのピークで積分を実行し測定値の変動を少なくする [定量化モード]に変更することをお勧めします。この変更を行うには以下を行います：

- [設定] メニューで、[すべて統合] をクリックします。

これによりプロットが以下のように変わります：

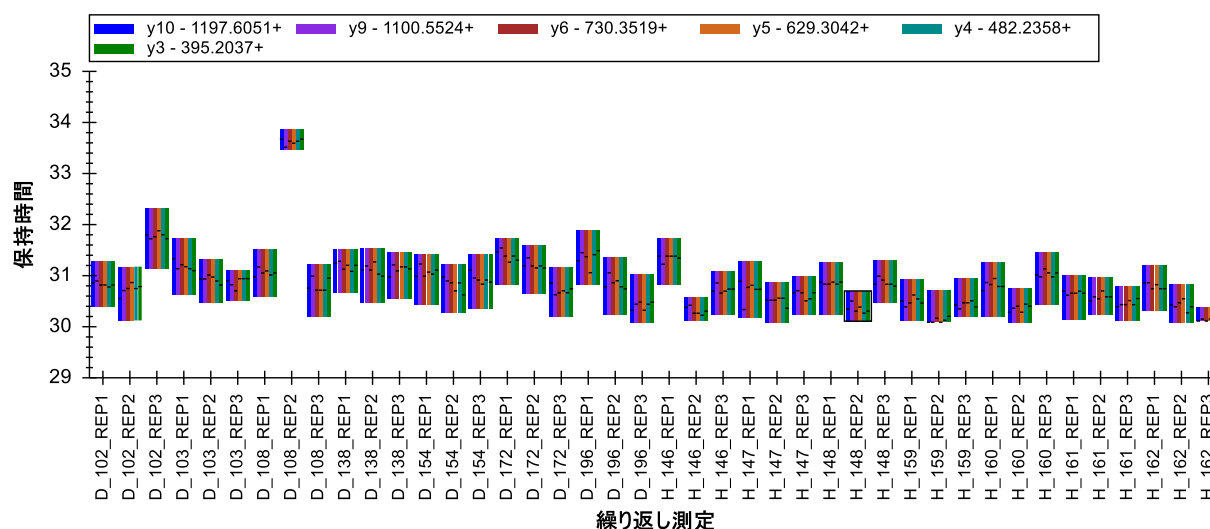


それでも、このペプチドの積分はあまり良くありません。19 分前後のピークはより良いピークに見えますが、単一ペプチドにより起こったものなのであまり信頼できないといったところです。更に、積分を調整しても全ての分析で一貫して測定可能な単一ペプチドとなる可能性は低いです。したがって、更に調整を行って時間を無駄にすることはせずに、次の操作を行ってください：

- 最初のペプチドを削除します。

### スケジュール化 SRM およびピーク切断

2 番目のペプチドの保持時間プロットを見るだけで、同ペプチドがより一致して積分されていることが分かります。



しかし、すべてトランジションのピーク頂点（バー内の水平ラインで表示）がうまく一定していないものも見られます。いくつかのクロマトグラムグラフに注目してみると、形状がギザギザで強度が中～低のピークも見られます。

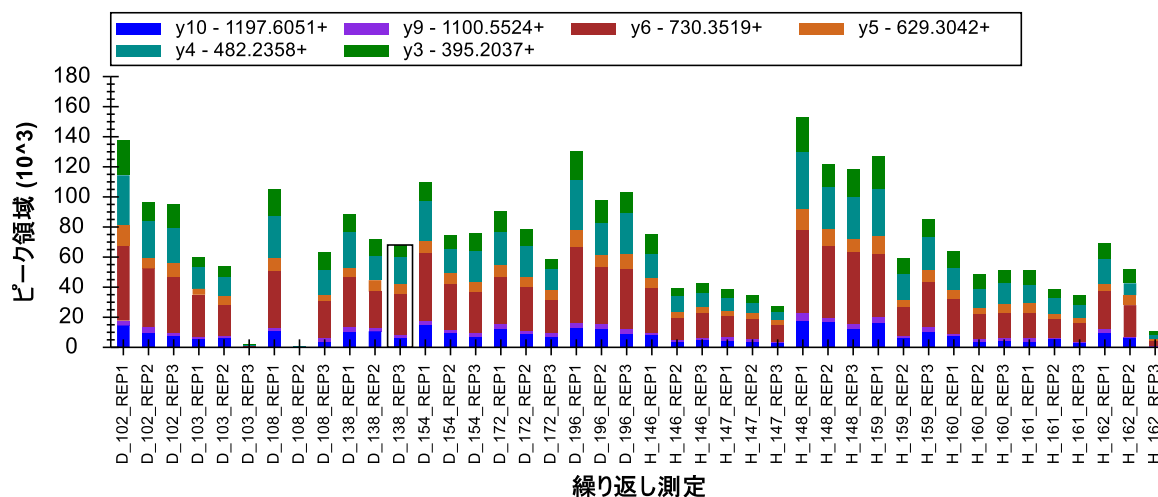
すべての罹患群反復測定（D\_）が左側で、および全ての健常群反復測定（H\_）が右側という、上記で示された並び順が見られない場合は、以下を行います：

- [保持時間] ビューを右クリックして、**[並び順]** を選択し [ドキュメント] をクリックします。

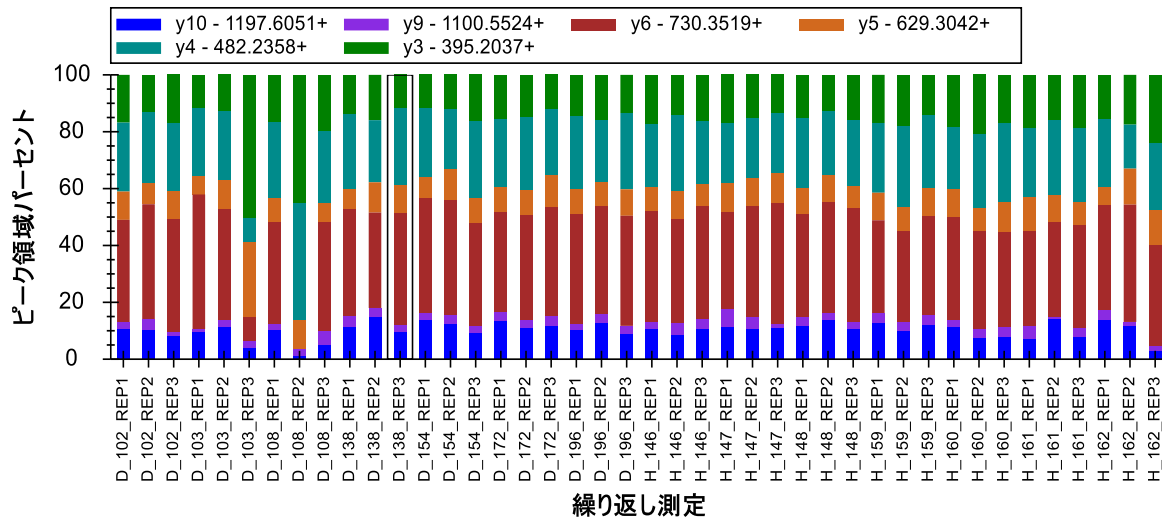
続行する前に、表示中のプロットに最後の調整を行います。

- [ピーク領域] ビューを右クリックし、[正規化] を選択して [合計] をクリックします。

これにより [ピーク領域] プロットが以下から：



次のプロットへと変わります：

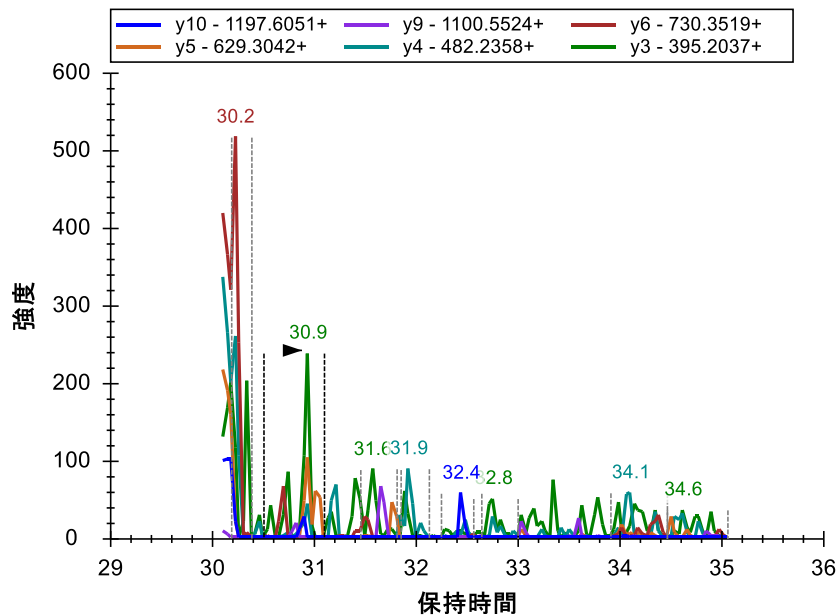


これにより、複数の分析にわたる相対イオン存在量の差を一瞥して見分けることが可能となり、[保持時間] ビューからすでに取得している情報を追加できます。他のすべての反復測定と比べて、D\_103\_REP3とD\_108\_REP2は顕著に異なっていること、そしてH\_162\_REP3はいくらか異なっていることに気付かれたかと思います。[保持時間] ビュー利用時には、D\_108\_REP2のみが疑わしく見えたことに注意してください。

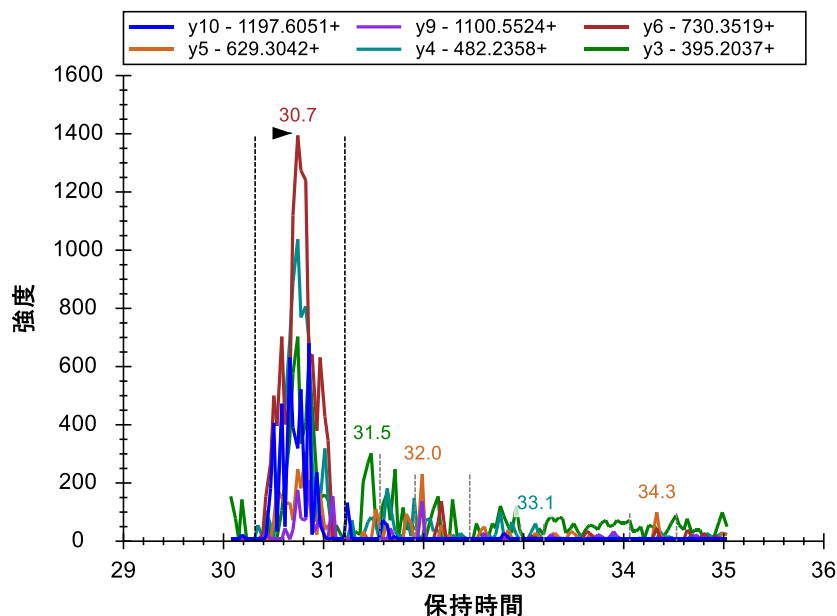
以下を行って、これらのピークを検査・修正します：

- [ピーク領域] ビュー内の D\_103\_REP3 バーをクリックします。

これにより当該ピークのクロマトグラムグラフが有効化され、以下のように見えます：



一方、正しく積分されたピークのプロットはこのように見えます:



スケジュール化取得ウィンドウが当該ペプチドの全溶出プロファイルをキャプチャしなかったため、Skylineにより誤ったピークが選択されています。しかし、正しいピークが 30.2 分付近で終わっているのが見られます。これは「**切断<sub>[M12]</sub>ピーク (truncated peak)**」と呼ばれます。このペプチドの積分の修正は、**切断ピーク**を使用し以下の操作を行うことで可能です:

- **切断ピーク**の直後の 30.4 分付近の x-軸の下をクリック&ドラッグします。
- 積分境界が動かなくなるまで左側をドラッグして、マウスボタンから指を放します。
- **[保持時間]** ビューの D\_108\_REP2 のバーをクリックして、クロマトグラムグラフを有効化します。
- 30.5 分付近から左端までのグラフの x-軸の下をクリック&ドラッグします。
- **[保持時間]** ビューの D\_162\_REP3 のバーをクリックして、クロマトグラムグラフを有効化します。

D\_162\_REP3 のピークも**切断 (truncated)**されています。しかし同ピークは別のケースに含まれていたため、Skylineによりすでに正しく選択されています。**[保持時間]** プロットを綿密に検査すると、ピーク**切断**を推定できる場合があります。バーの長さおよび水平ラインからバー端までの付近に、特に注意を払ってください。

ピーク**切断**は、このような非標識データにとって無視できない問題です。ピークの各ポイントで軽ペプチドプリカーサーと重ペプチドプリカーサーの間の相対比が得られる同位体標識参照ペプチドを利用すると、**切断ピーク**による精度は失われますが、軽と重の間の比は変わらず有効です。非標識データを用いる場合は逆に、**切断ピーク**は有効なペプチド測定値として考慮できません。欠損データとの差分統計を計算するか、当該ペプチドを考慮から完全に除かなくて

はなりません。Skyline は、以上のことがこれ以降の操作で使うツールで有効になるよう、**切断ピーク**を追跡します。

Skyline は**切断ピーク**を、その境界の 1 つがクロマトグラムの末端ポイントであり、同末端がピーク高さの 1% を超え、その他の積分境界での強度より高いピークとして定義します。

Skyline が**切断ピーク**として同定したその他の自動**積分ピーク**を見るには、以下の手順を実行します:

- **[編集]** メニューで **[検索]** (Ctrl-F) をクリックします。
- **[検索項目]** フィールドをクリアします。
- **[方向]** で **[下]** を選択します。
- 詳細オプションがすでに表示されていない場合は、**[詳細を表示]** ボタンをクリックします。
- **[切断ピーク]** のチェックを入れます (および、その他の項目のチェックをオフにします)。

[検索] フォームは以下のように見えるはずです:

検索

検索項目(F):

文字形式(大文字・小文字)の一致

方向

上(U)  下(D)

次も検索:

- 一致するライブラリデータがありません
- 欠損しているすべての結果
- 欠損している任意の結果
- 統合されていないトランジション
- 手動統合ピーク
- 切断ピーク

次を検索(F)

すべて検索(A)

閉じる

<< 詳細を隠す(V)

- [すべて検索] ボタンをクリックします。
- [閉じる] ボタンをクリックします。

Skyline に切断プリカーサーおよびトランジションの長いリストが表示され、先に検査したばかりの 3 つのピークが冒頭に記載されています。ラインをダブルクリックすると、対応するクロマトグラムグラフが有効化されます。



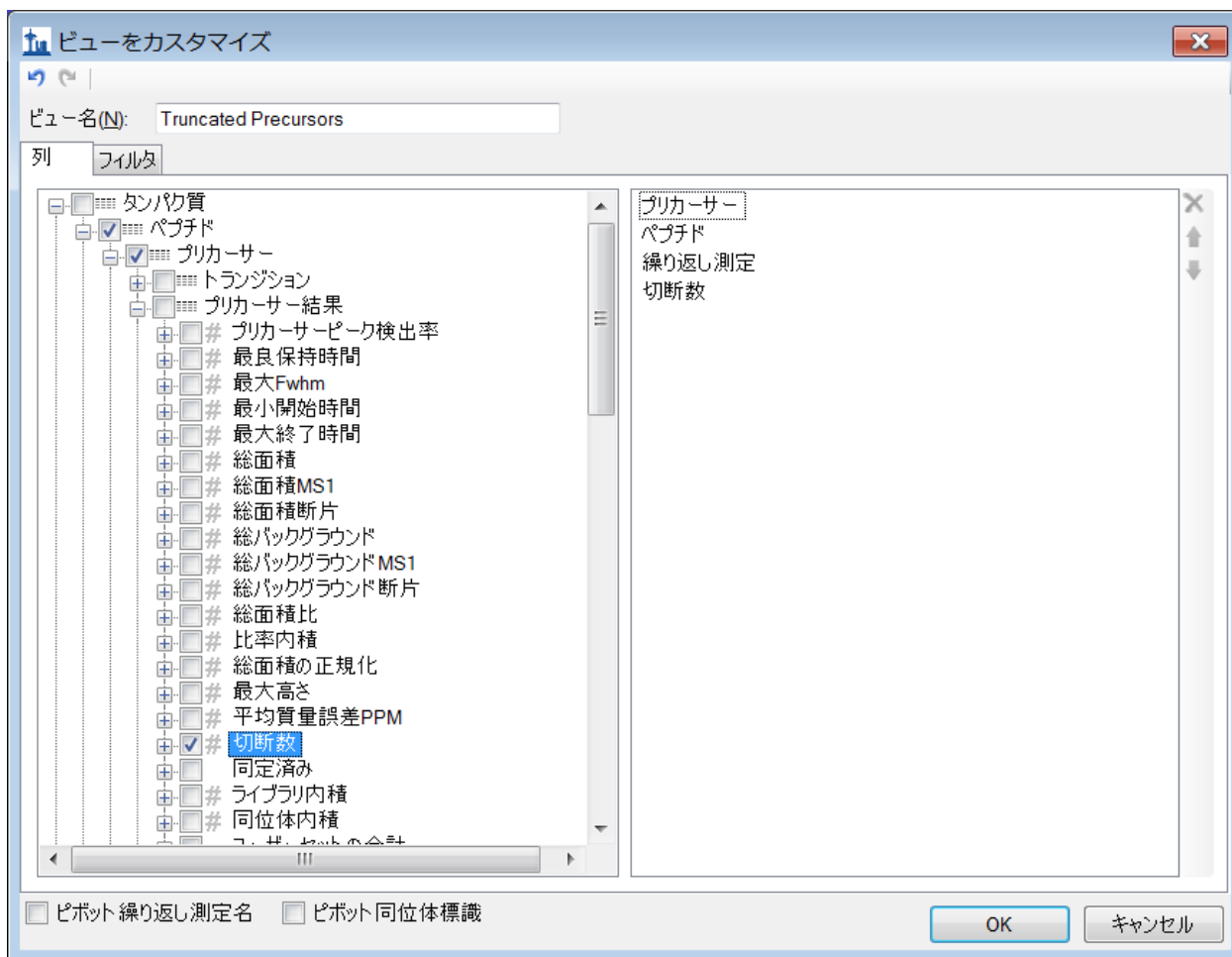
場所	タイプ	テキスト
905.9565++ (D_103_REP3)	プリカーサー ...	6個の切断ピーク
905.9565++ (D_108_REP2)	プリカーサー ...	6個の切断ピーク
905.9565++ (H_159_REP2)	プリカーサー ...	6個の切断ピーク
P [y10] - 1197.6051+[5] (H_162_RE...	トランジション ...	切断ピーク
W [y9] - 1100.5524+[6] (D_196_RE...	トランジション ...	切断ピーク
T [y6] - 730.3519+[1] (D_196_REP3)	トランジション ...	切断ピーク
T [y6] - 730.3519+[1] (H_146_REP2)	トランジション ...	切断ピーク
T [y6] - 730.3519+[1] (H_148_REP2)	トランジション ...	切断ピーク
T [y6] - 730.3519+[1] (H_159_REP1)	トランジション ...	切断ピーク
T [y6] - 730.3519+[1] (H_160_REP2)	トランジション ...	切断ピーク
T [y6] - 730.3519+[1] (H_162_REP2)	トランジション ...	切断ピーク
T [y6] - 730.3519+[1] (H_162_REP3)	トランジション ...	切断ピーク
F [y5] - 629.3042+[4] (H_160_REP2)	トランジション ...	切断ピーク
F [y5] - 629.3042+[4] (H_162_REP2)	トランジション ...	切断ピーク
F [y5] - 629.3042+[4] (H_162_REP3)	トランジション ...	切断ピーク

保持時間 結果を検索

またドキュメントグリッドを使用し以下の操作を行って、ピークが**切断**されているすべてのプリカーサーのリストを作成可能です:

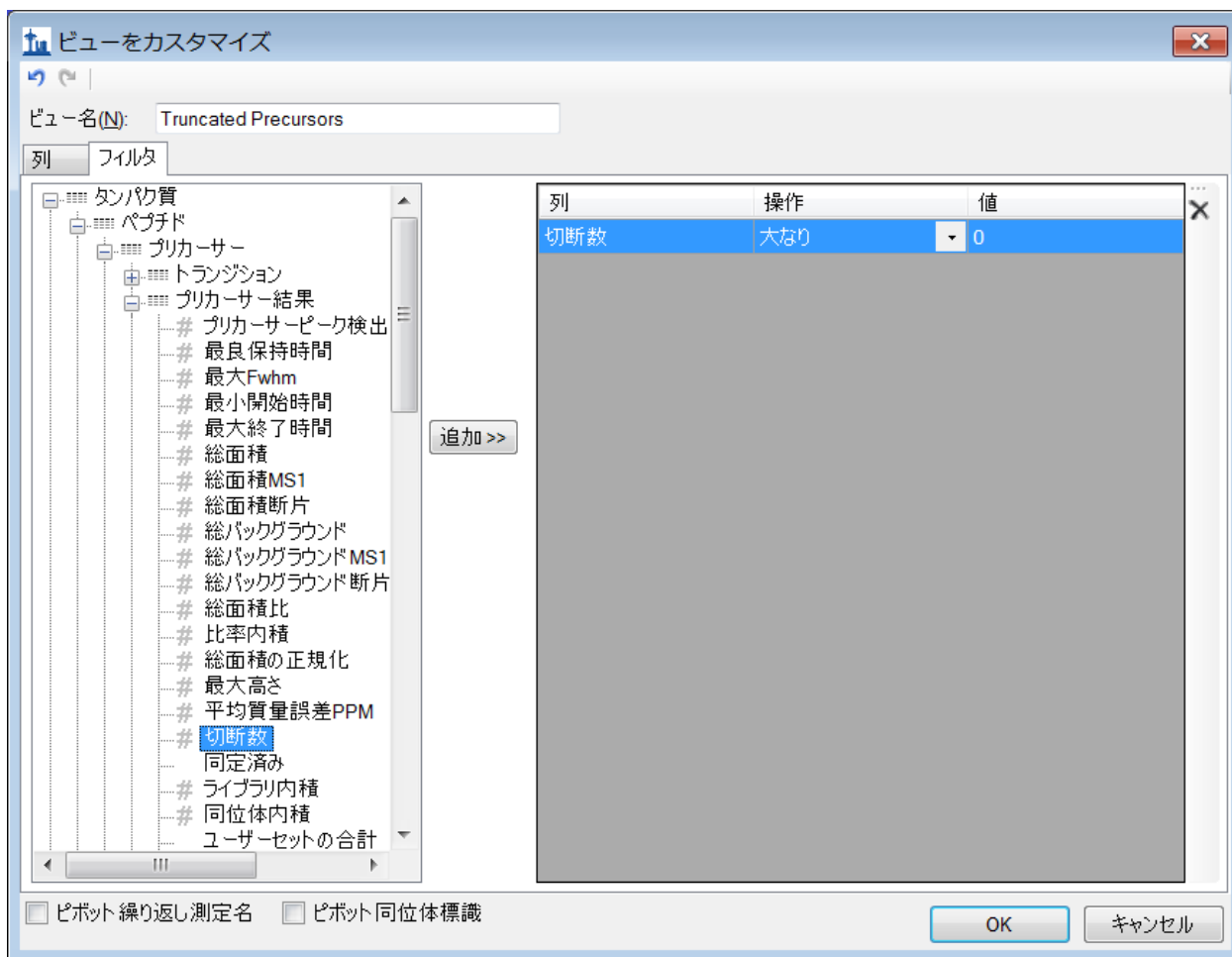
- [ビュー]メニューで、[ドキュメントグリッド]をクリックします。
- [ドキュメントグリッド]ビューの[ビュー]メニューで、[プリカーサー]をクリックします。
- [ドキュメントグリッド]ビューの[ビュー]メニューで、[ビューをカスタマイズ]をクリックします。
- [ビュー名]フィールドに「Truncated Precursors (切断プリカーサー)」と入力します。
- リストの右側で、最初の2つ(プリカーサーおよびペプチド)を除くすべての列名を選択します(クリックおよびシフト-クリックを利用)。
- 右端付近の✕ボタンをクリックして、最初の2つを除くすべての列を削除します。
- 左側のツリー内の「繰り返し測定」のチェックをオンにします。
- ツリーを展開して、タンパク質 > ペプチド > プリカーサー > プリカーサー結果 > **切断**数とチェックします。

[ビューをカスタマイズ]フォームは次のようになります:



- 右側のリストで「**切断数**」をダブルクリックします。
- [フィルタ] タブをクリックします。
- [追加] ボタンをクリックします。
- [操作] ドロップダウンリストから「大なり」を選択します。
- [値] フィールドに「0」と入力します。

[ビューをカスタマイズ] フォームは次のようになります:



- [OK] ボタンをクリックします。
- 「**切断数**」列ヘッダーを2回クリックして、降順に並べ替えます。

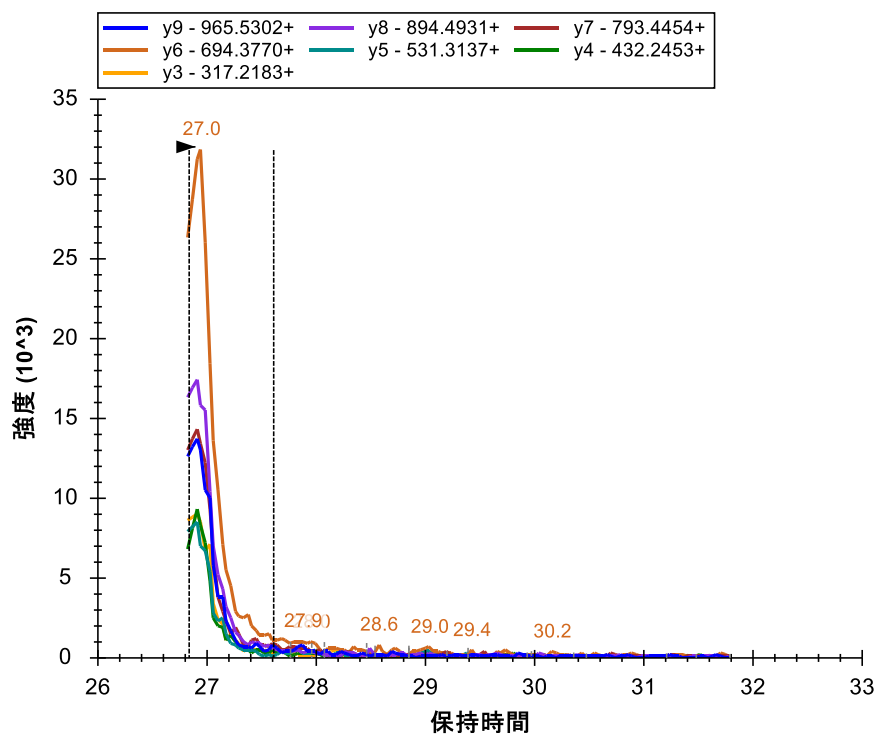
[ドキュメントグリッド] に、Skyline がすでに自動的に選択した 222 の**切断ピーク**が表示されます:

プリカーサー	ペプチド	繰り返し測定	切断数
<a href="#">614.3164++</a>	<a href="#">DFATVYVD...</a>	<a href="#">D 196 REP3</a>	7
<a href="#">614.3164++</a>	<a href="#">DFATVYVD...</a>	<a href="#">H 147 REP2</a>	7
<a href="#">614.3164++</a>	<a href="#">DFATVYVD...</a>	<a href="#">H 148 REP2</a>	7
<a href="#">614.3164++</a>	<a href="#">DFATVYVD...</a>	<a href="#">H 159 REP1</a>	7
<a href="#">614.3164++</a>	<a href="#">DFATVYVD...</a>	<a href="#">H 159 REP3</a>	7
<a href="#">614.3164++</a>	<a href="#">DFATVYVD...</a>	<a href="#">H 160 REP2</a>	7
<a href="#">614.3164++</a>	<a href="#">DFATVYVD...</a>	<a href="#">H 162 REP3</a>	7
<a href="#">730.8488++</a>	<a href="#">DYVSQFES...</a>	<a href="#">D 196 REP2</a>	7
<a href="#">730.8488++</a>	<a href="#">DYVSQFES...</a>	<a href="#">D 196 REP3</a>	7
<a href="#">730.8488++</a>	<a href="#">DYVSQFES...</a>	<a href="#">H 147 REP2</a>	7
<a href="#">730.8488++</a>	<a href="#">DYVSQFES...</a>	<a href="#">H 159 REP1</a>	7

これらのピークを表示するには:

- DFATVYVDAVK の [ペプチド] リンクをクリックします。
- ペプチドの各 [繰り返し測定] リンクをそれぞれクリックします。

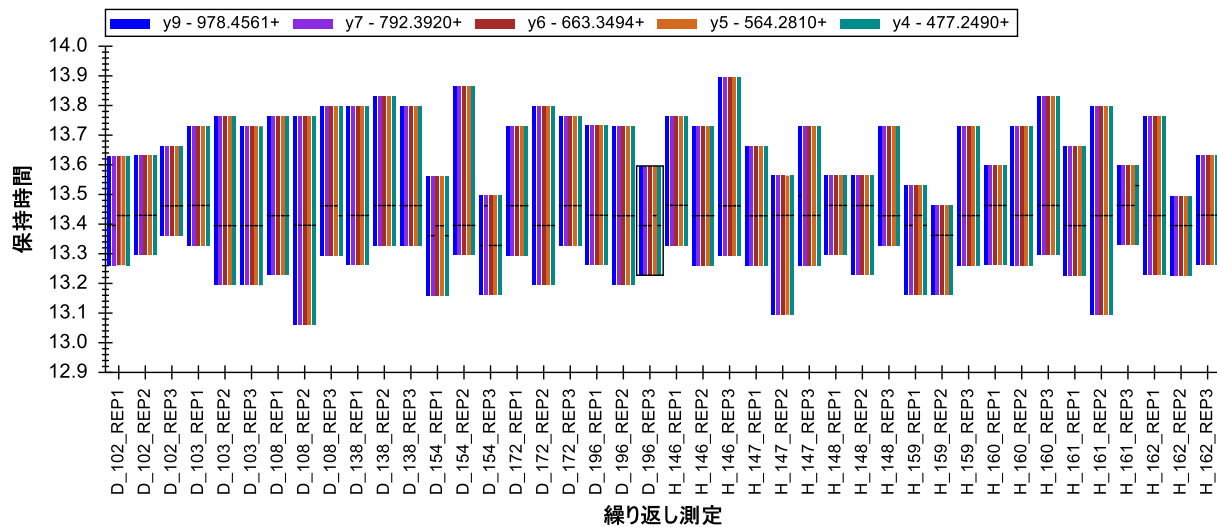
Skyline により以下のクロマトグラムグラフが有効化されます:



- [ドキュメントグリッド] を閉じます。
- [結果を検索] ビューを閉じます。
- 2 番目のペプチド R.LGGEEVSVACK.L [237, 247] を選択します

### 複数の反復測定データ処理を開始する

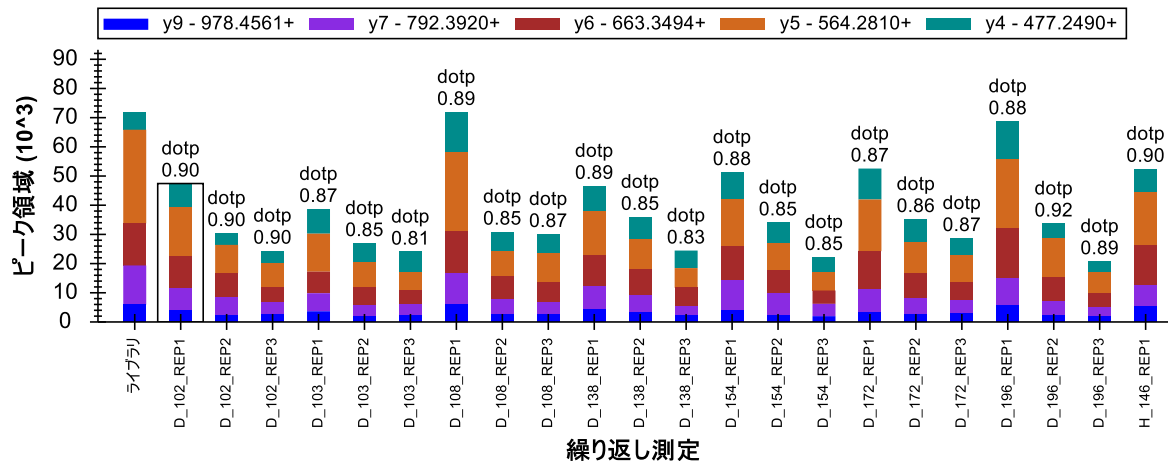
ペプチド LGGEEVSVACK は、これまで見てきた他の 2 つのペプチドよりもはるかに良く見えます。保持時間は非常に安定しており、水平ピーク頂点ラインは各反復測定内のほとんどすべてのトランジションに対して一致しています。



[ピーク領域] ビューに表示されている相対イオン存在量も比較的安定して見えます。このペプチドに対応するライブラリスペクトルがあること、そしてその相対イオン存在量（プロット内の一番左側のバーに表示）が測定ピークのものと同様であることに気付かれるかもしれません。ペプチドに対応するライブラリスペクトルにどれだけ厳密に一致しているか理解を深めるには、以下を行います：

- [ピーク領域] ビューを右クリックし、[正規化] を選択して [なし] をクリックします。
- スプリッターバーを移動させて、[ピーク領域] ビューがスクリーン幅のほとんどを占めるようにします。一方、スクリーンスペースが制限されている場合は、罹患群 (D\_) 反復測定付近にあるズーム長方形をクリック & ドラッグします。

Skyline に、1（最良）～0（最悪）の範囲のライブラリスペクトルピーク領域と測定ピーク領域との間の類似性測定値であるドット積 (dotp) 関係が表示されます。



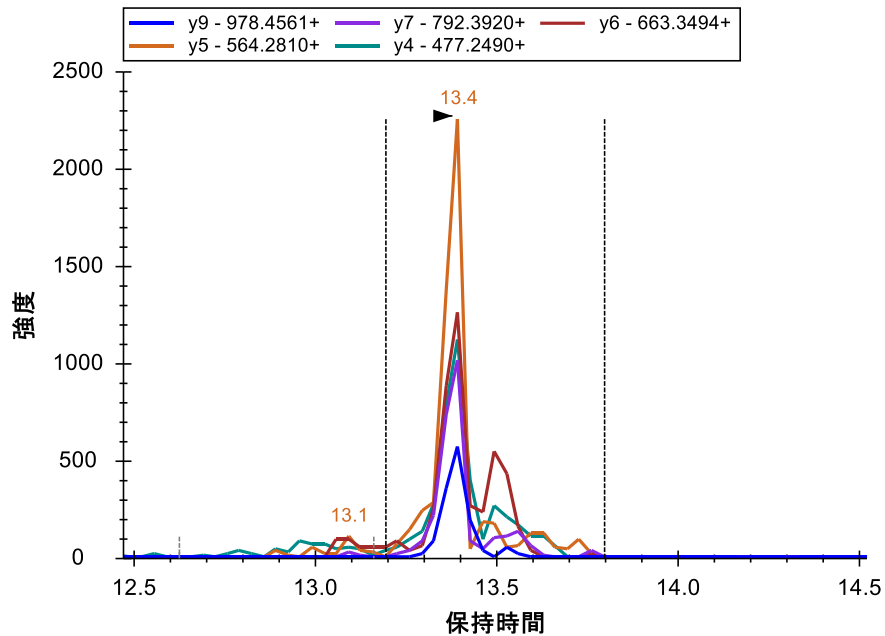
以下の操作を行って、[ピーク領域] ビューを基の状態に戻します:

- ズームを使用している場合、右クリックして[すべてのズーム/パンを元に戻す]をクリックします。
- 右クリックして、[正規化]を選択し[合計]をクリックします。

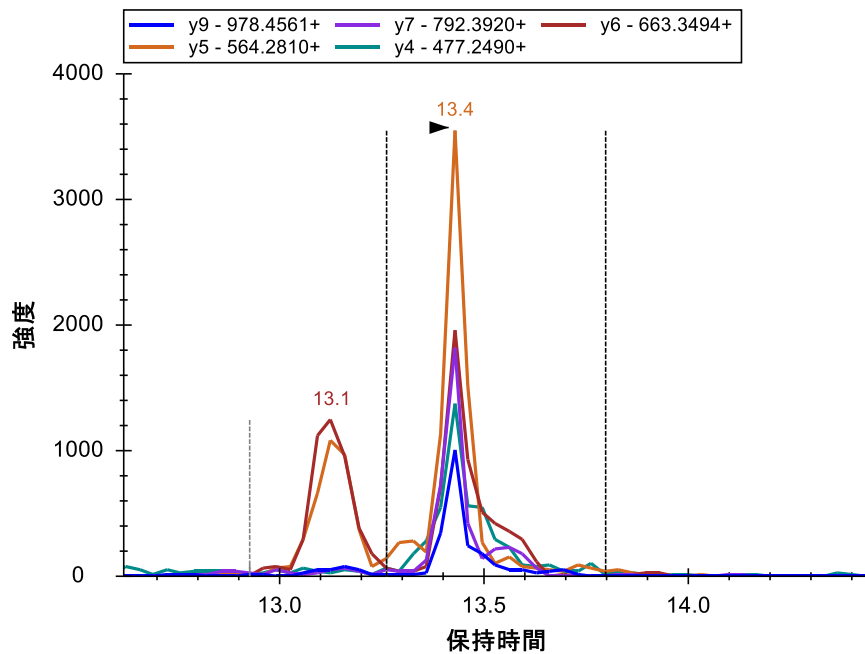
ここで、これらのピークのクロマトグラムを詳しく見るため、以下の操作を行います:

- [ビュー]メニューで、[オートズーム]を選択して[最良ピーク] (F11) をクリックします。
- [ターゲット]ビューの [繰り返し測定] ドロップダウンリストをクリックします。
- 下向き矢印キーを利用して、複数のクロマトグラムグラフをスキャンします。

一部のトランジションにおいて、メインピークの端付近に他のピークとあまり一致しないシグナルが表れていることに気付かれるかもしれません。

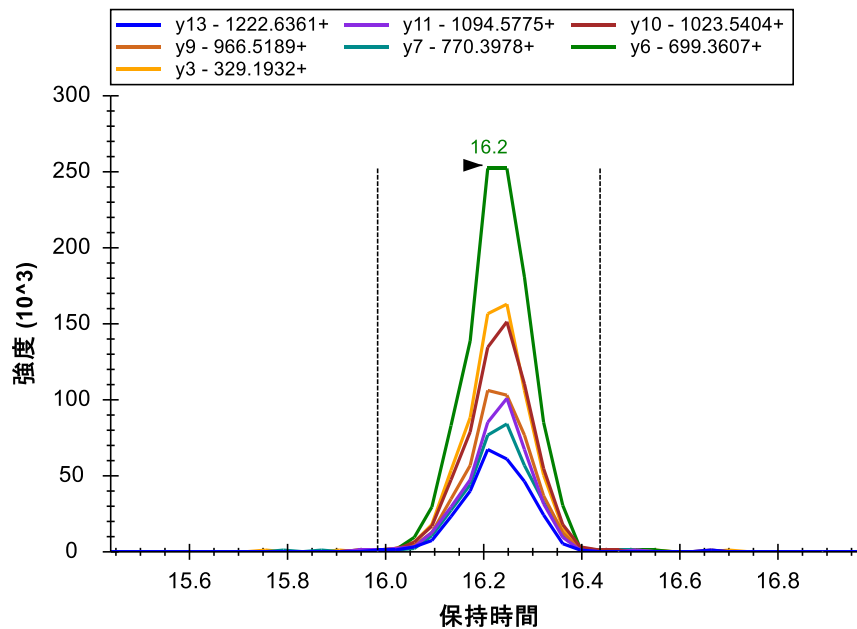


13.1 分付近の y5 および y6 に明確なシグナルが見られ、これは明らかに同一ペプチドからのものではありません。

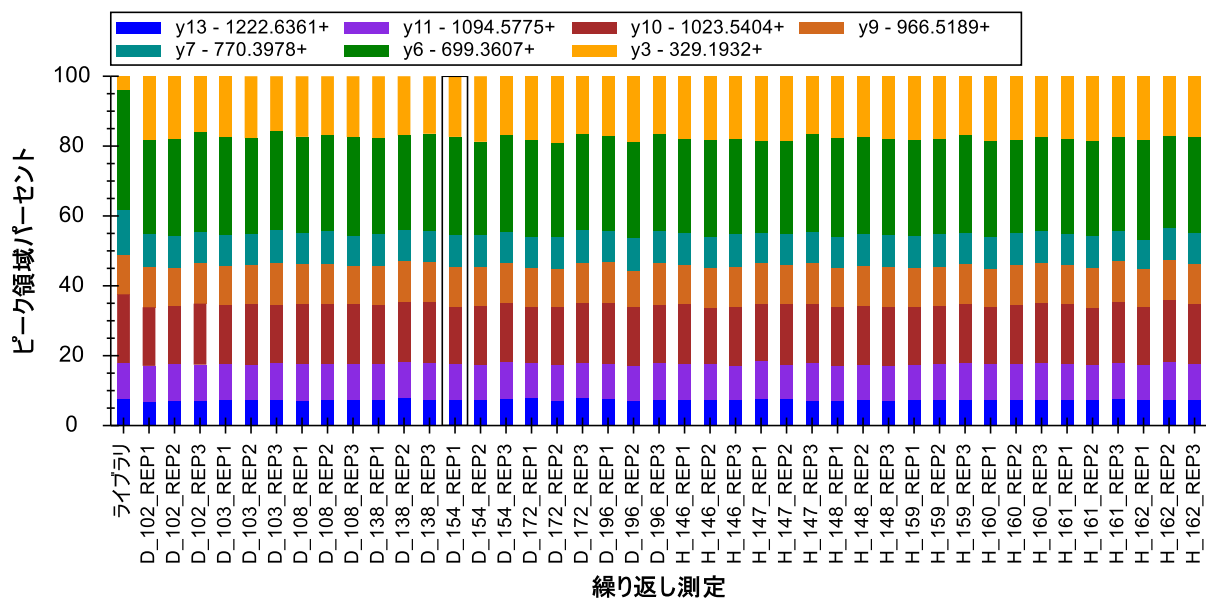


このような干渉シグナルが積分境界内で起こる場合、定量的測定にエラーが増えます。真の定量的データについては、このペプチドの 5 つのトランジション全てが必要かどうか検討してみてください。

しかし、本チュートリアルでは次のペプチドへと続行します。すべての 7 トランジションに対して、強いシグナルで形状が整っていてかつ共溶出のピークが見られます。



相対イオン存在量も非常に一定しています。



y3 のピーク領域は、ライブラリスペクトルで見たものより強力です。これは、トリプル四重極装置内で起こる二次断片化により、説明可能かもしれません。ライブラリスペクトルを生成した際に用いたイオントラップでは、装置内で起こる共鳴励起では二次断片化は起こらない場合

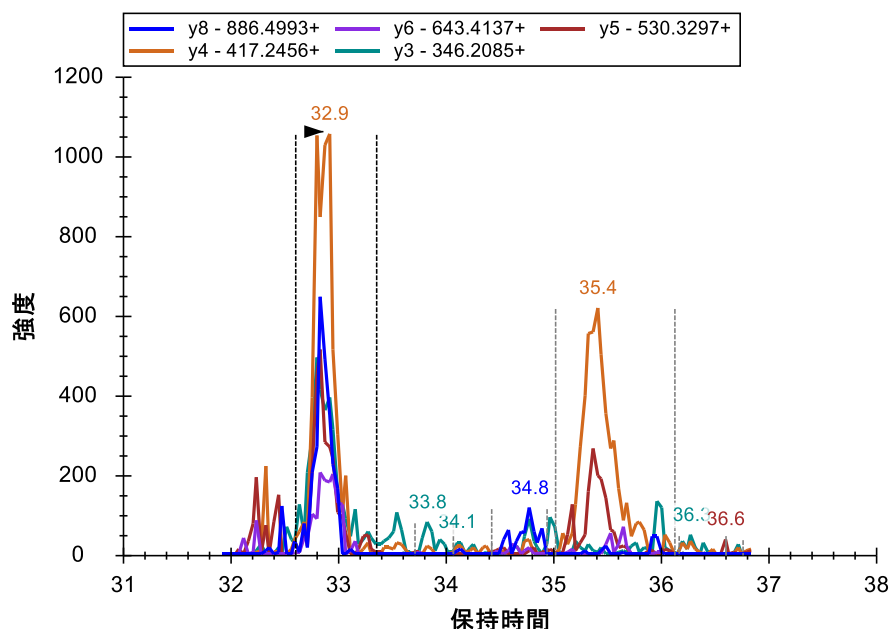


があります。このケースでは、スペクトルは NIST 公共ライブラリからのものですので、ライブラリ作成に使用した装置のタイプについての情報はありません。

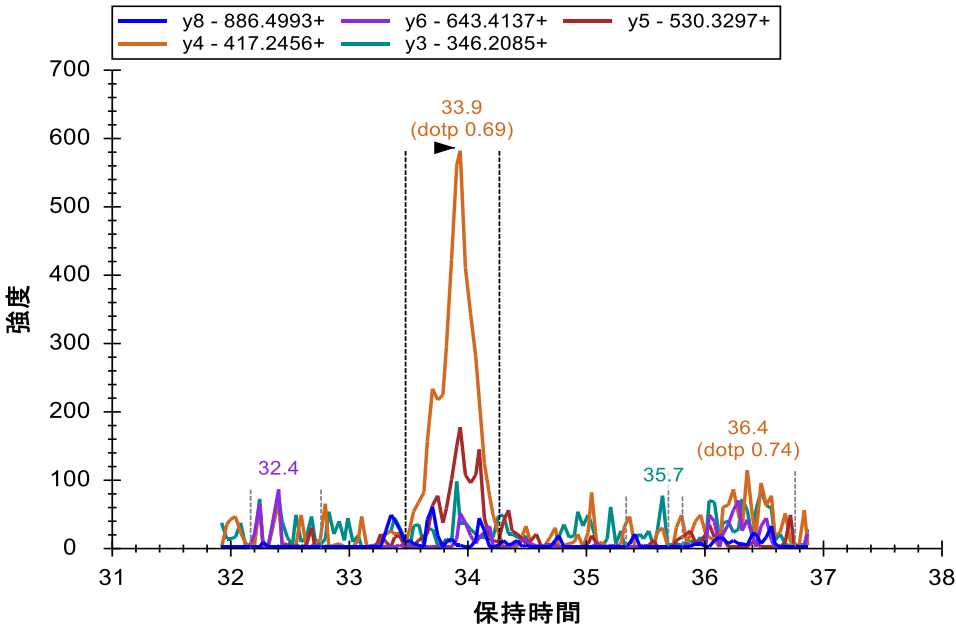
この積分は更に検査を行わずとも十分に良好であり、次のペプチド、R.GSYNLQDLLAQAK.L [378, 390] へと続行可能であることが、一目で分かります。このペプチドについては、[保持時間] ビューと [ピーク領域] ビューの両方に問題が示されています。これらの測定に何が起きているか理解するには、以下を行います：

- [ビュー] メニューで、[オートズーム] を選択して [なし] (Shift-F11) をクリックします。
- [ターゲット] ビューの [繰り返し測定] ドロップダウンリストを使用して、最初の問題データである D\_103\_REP3 が見つかるまで冒頭から反復測定を順に再確認します。

最初の 5 つの繰り返し測定については、選択されたピークが約 32.5 分～33.5 分にあり、すべての 5 つのトランジション上にシグナルがあります。しかし、35 分付近で主に y4 および y5 のシグナルを示すもう 1 つのピークがあり、これらは明らかに別のペプチドにより引き起こされたものです。このようなケースでは、目的ペプチドと共溶出しない干渉シグナルの情報が表示中のクロマトグラム「環境」に追加される場合があります。



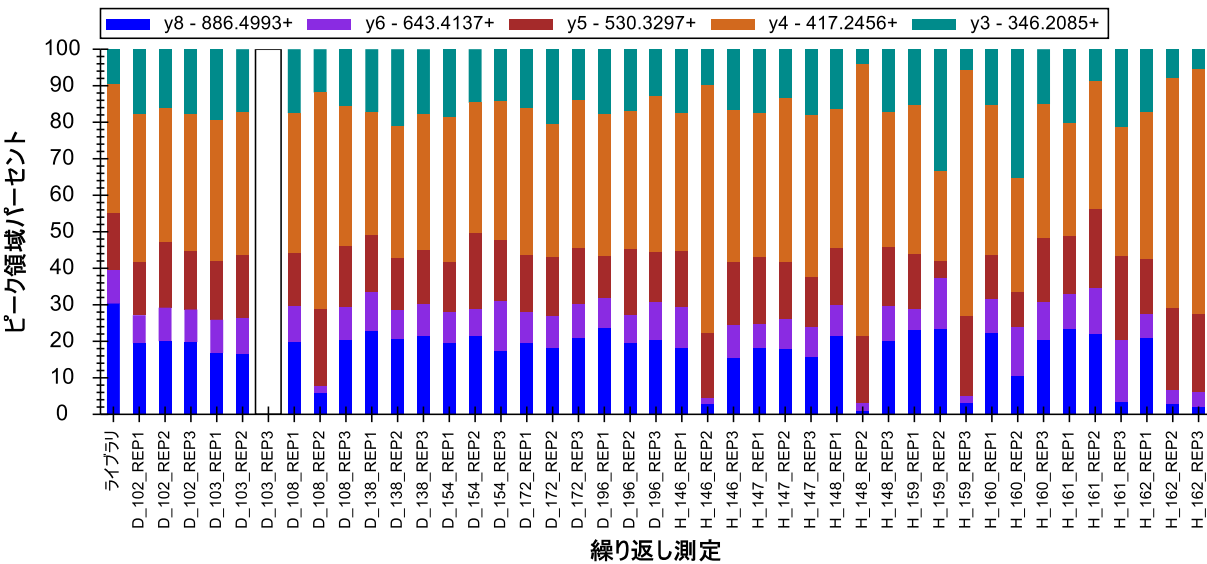
D\_103\_REP3 では、主に y4 および y5 上にシグナルを持つ積分ピークが 33.9 分に見られますが、左側にはその他の候補ピークは見られません。



このケースでは、スケジュールウィンドウによる候補ペプチドのシグナル捕捉が失敗に終わっています。不正確なピークを削除するには、以下の操作を行います：

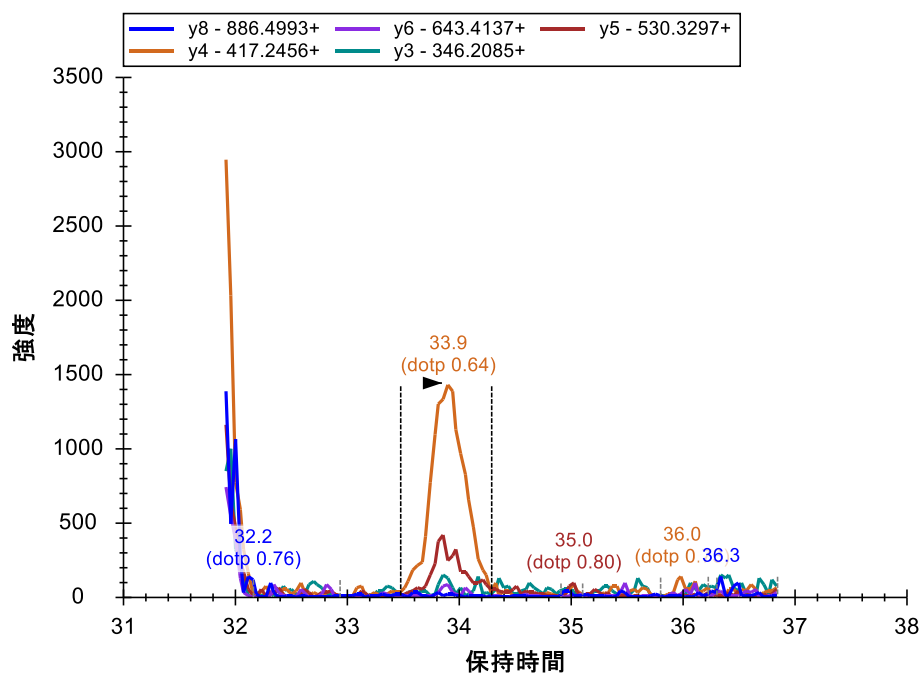
- クロマトグラムグラフを右クリックして、[ピークを削除] を選択し [すべて] をクリックします。

これにより [ピーク領域] グラフが空白となります。

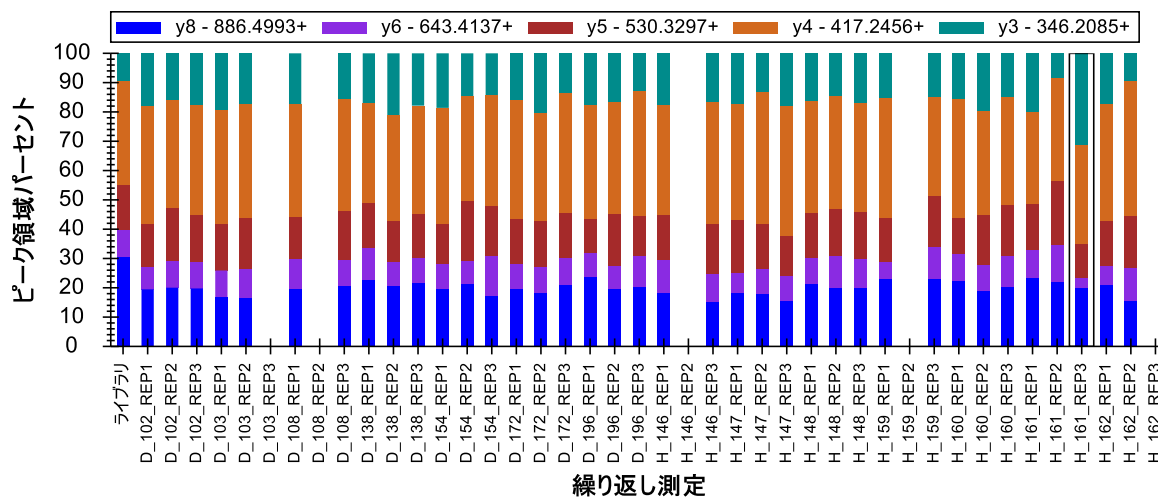


クロマトグラムプロットを続行すると、このパターンが何回か反復されているのが見られます。欠損ピークを削除し続けてください。ピーク切断のケースも一部見られます。H\_148\_REP2 に

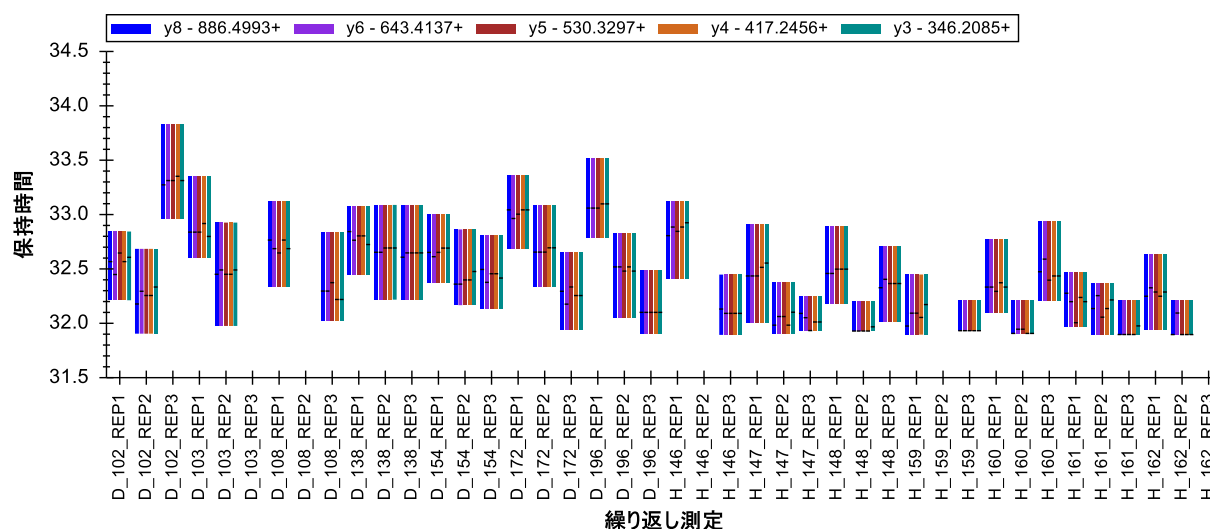
目を向けると、Skyline が 33.9 分のピークを選択しましたが正しいピークの一部が視認できません。



これは単に、上記で行ったように x 軸の下をクリック & ドラッグして、**切断** ピークとして積分可能です。以下のように見える [ピーク領域] プロットとなるはずですが:



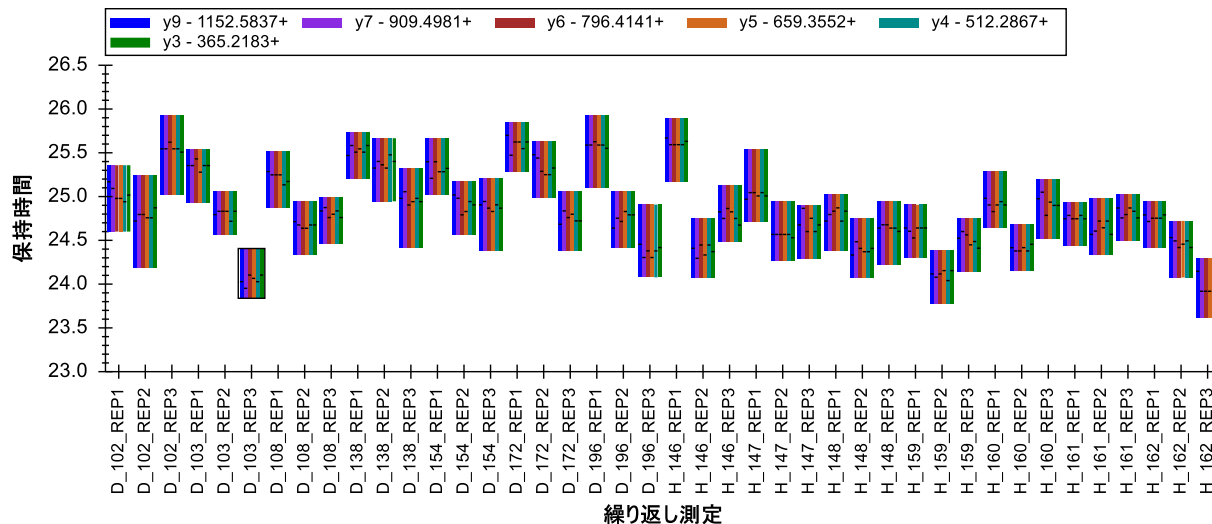
また [保持時間] プロットは以下のように見えるはずですが:



ペプチド、R.DIPVNPMLIYR.S [46, 56] へと続けていくと、5 分のスケジュールウィンドウにおいて、積分ピークと真に一致しているシグナルはないことに気付くでしょう。したがって、このペプチドは消去して構いません。

タンパク質 NP\_001012027 内の次の 3 つのペプチドはすべて、比較的安定した保持時間と相対イオン存在量で見たい目は良好です。

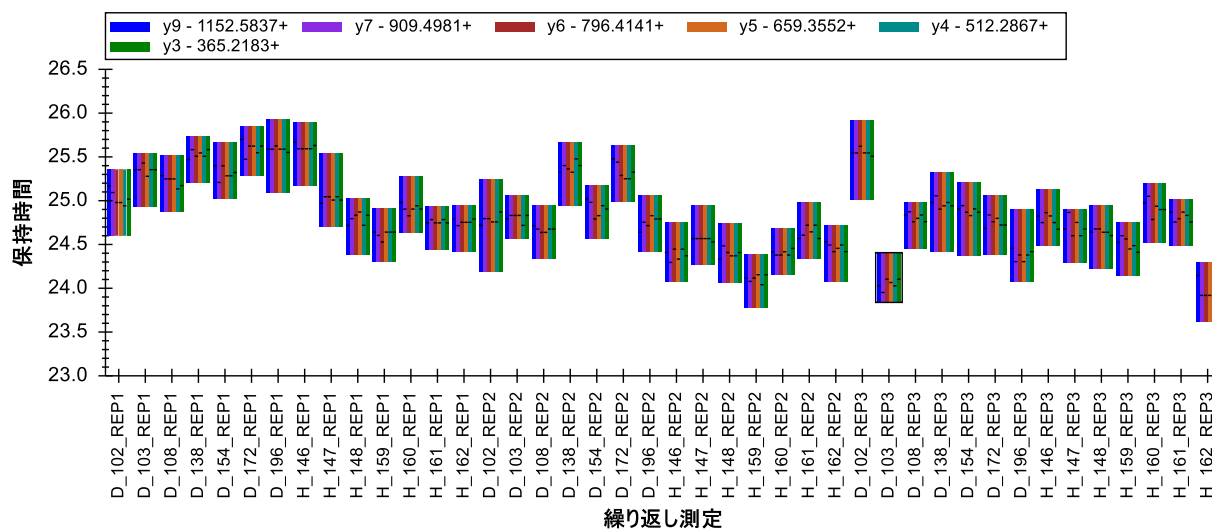
保持時間内で見られる変動が、予想より一貫性がないと思われるかもしれません。これは、分析が論理的に並んでいるときによく見られるケースです。現在の並び順は、反復測定をドキュメントへとインポートした際の結果です。この種の研究では、すべての罹患被験体を最初にリストして、その後すべての健常被験体をリストすることが便利です。各被験体のすべての技術的反復測定もまた、グループ化されます。ペプチド TSDQIHFFFAK については、保持時間はこのように見えます:



しかし、装置上で取得された順に分析をみると、かなり有用な場合もあります。これは、以下の操作によって実行可能です：

- [保持時間] グラフを右クリックし、[並び順種別] を選択して [取得時間] をクリックします。

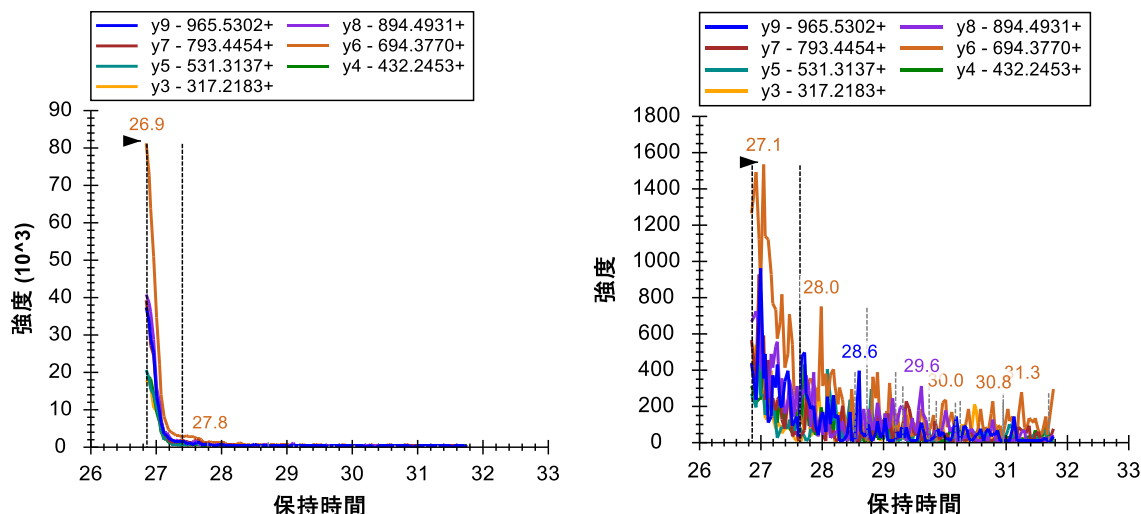
グラフが以下のように変更されます：



この保持時間パターンは、一貫性がやや向上しています。ここで、次の 2 つのペプチド (LQPLDFK、SQLPGIIAEGR) の中で選択すると、非常に一貫したパターンを見ることができます。

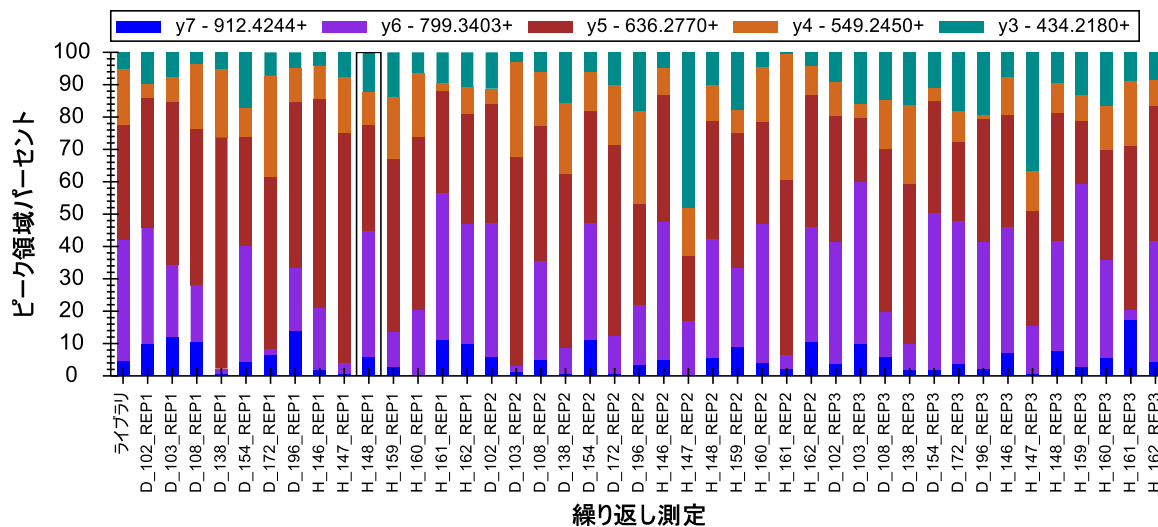
次のペプチド、DFATVYVDAVK には一部ピーク切断があります。これは、あまり良く一致していない [ピーク領域] ビュー上のバーをクリックして、および対応するクロマトグラムグラフ

の x-軸の下をクリック&ドラッグして、かなり迅速に修正可能です。また、ノイズ過多に見えるクロマトグラフィーにおいて、ピークがいくつかあります。この例は、以下のクロマトグラムの右側にあります。これは、最も右端のペプチドピークのみが測定され、強度がより高い部分が欠けているためです。



このようなケースでは、Skyline が切断ピークとして同定できなかった一部ノイズを絶対に積分しないことです。これらの反復測定のピークを削除する方が簡単かもしれません。

このセクションで最後に扱うのは、ペプチド FGLYSDQMR です。[保持時間] ビューで、Skyline が主に 19分に近いピークを選択しているのが見られます。しかし [ピーク領域] ビューには、相対イオン存在量における一貫性が非常に乏しいことが示されています。



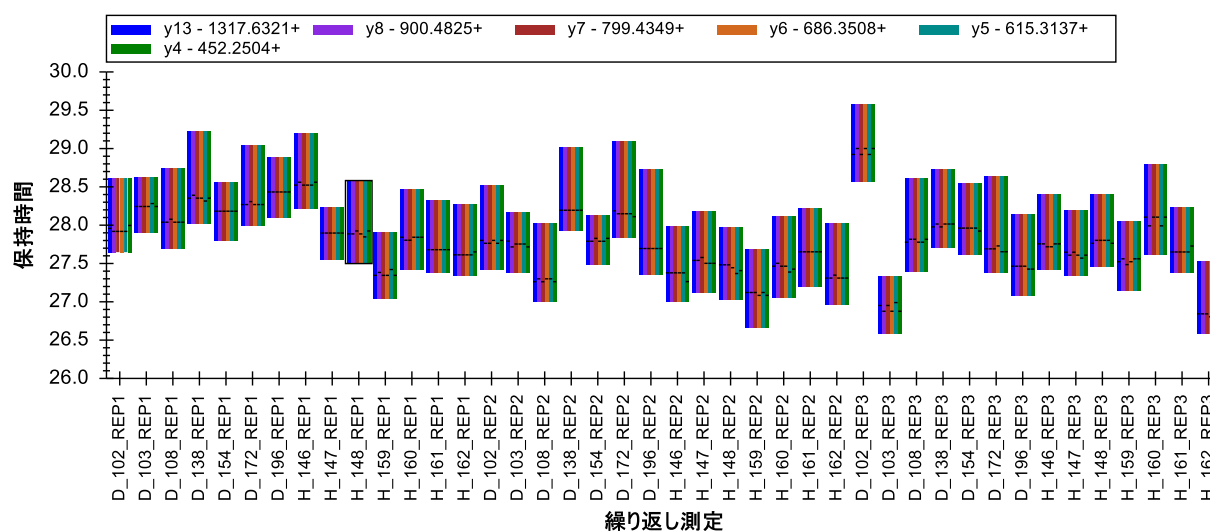
いくつかのクロマトグラムグラフを分析すると、18.8 分の信号は、積分シグナルすべてが単一ペプチドに由来していると信頼した上で、一定して積分されているとは考えにくいということが分かります。このケースでは、当該ペプチドを消去して先に進みます。

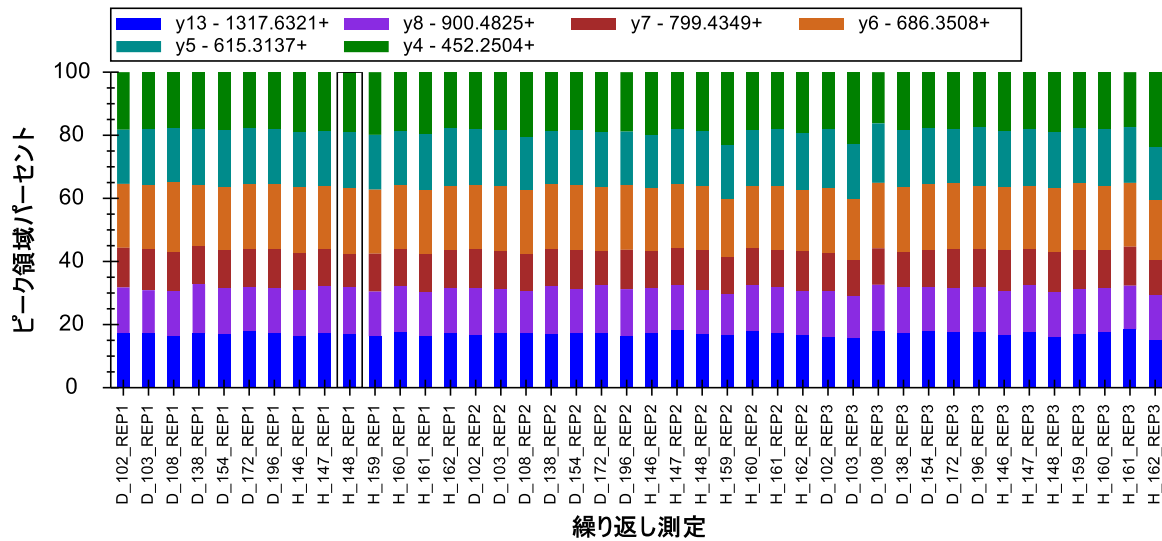
## グローバル正規化標準

このデータセットには安定同位体希釈 (stable isotope dilution, SID) ペプチドが欠如していますが、試料にスパイクしたグローバル標準ペプチド (*C. elegans* 由来) が含まれています。これらのペプチドを使用する目的は、LC-MS/MS 分析時のシステム変動の影響を低減させるために他のすべてのペプチドのピーク領域を正規化することです。<sup>2</sup> これらのペプチドを分析するには、以下を行います:

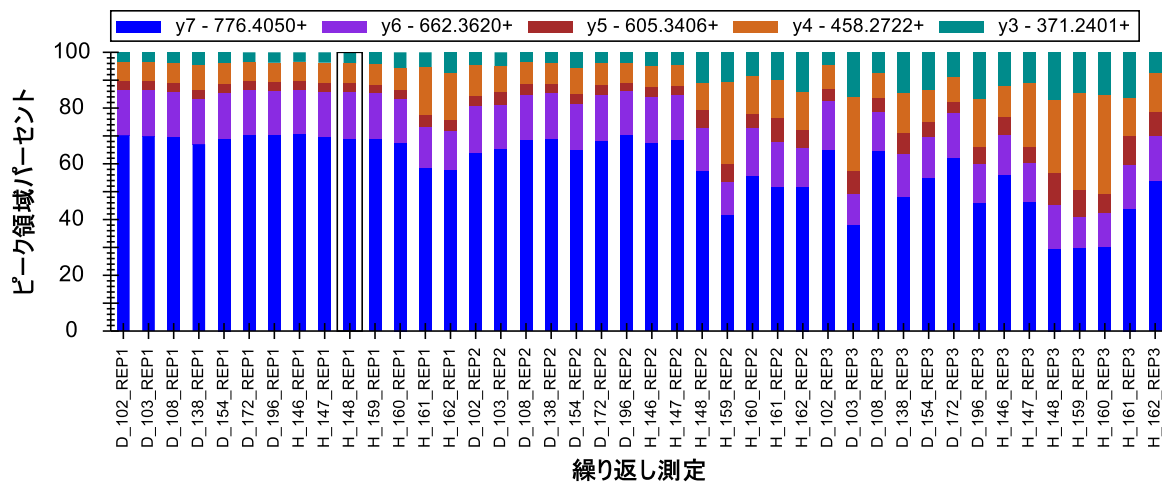
- [ターゲット] ビューをクリックします。
- キーボードの End キーを押してから、上向き矢印キーを押します。

これにより、ドキュメント内の最後のペプチドである AFGLSSPR が選択されます。「S」という名のリスト内にグループ化されている 3 つのペプチド、HLNGFSVPR、VVLSGSDATLAYSFAK と AFGLSSPR は、本実験で使用した *C. elegans* ペプチドです。これまで検査してきた 2 つの概要プロットでの最後の 2 つのペプチドの自動積分によって、即座に確信を得ることができます。保持時間およびピーク領域は比較的一貫しています。





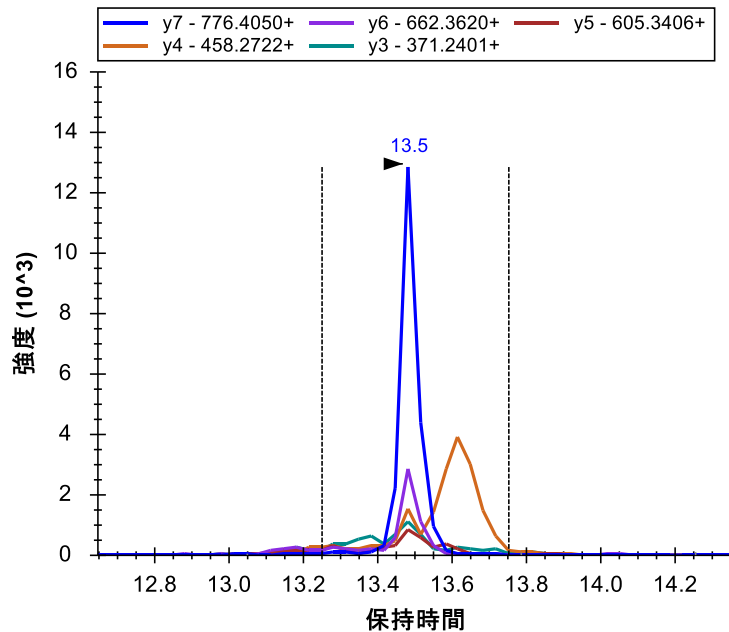
ペプチド HLNGFSVPR へと移ると、一貫した保持時間が観察されます。しかし、ピーク面積にはより多くの変動が見られます。y4 と時折 y3 の相対面積はデータセットを通して変わっているように見えます。



y4 が最も豊富ないくつかの [ピーク領域] バーをクリックすると、y4 と時折 y3 への明らかな干渉が見られます。

- [ビュー] メニューで、[オートズーム] を選択して [最良ピーク] (F11) をクリックします。





これら注入ペプチドが正規化標準として使用できるかについてよりはっきりと理解するには、まず以下を行います：

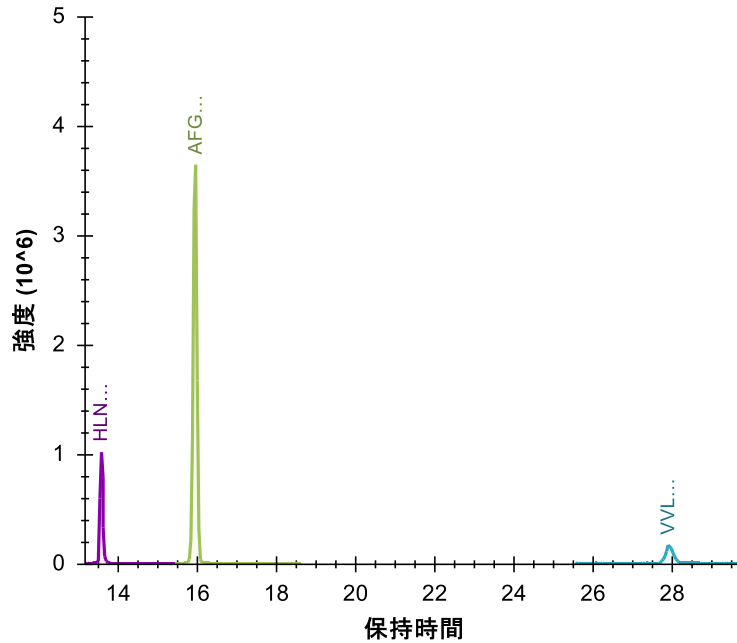
- [ピーク領域] ビューを右クリックし、[正規化] を選択して [なし] をクリックします。

ここでペプチド HLNGFSVPR のピーク領域が、最初の試料で取得された結果から最後の試料の結果になるにつれて劇的に減少しているのが見られます。総ピーク領域はおよそ 6,000,000 から 30,000 で、200 倍の差があります。選択した HLNGFSVPR ペプチドプリカーサー 513.7776++で [結果グリッド] ([ビュー] > [結果グリッド]) を使用して、正確な最大および最小総ピーク領域を判定できるか、やってみてください。その他の 2 つの標準ペプチドを見てみると、経時的に非常に減少している (VVLGSDATLAYSFAFK 230 万が 110 万に、AFGLSSPR 2300 万が 100 万に) のが見られます。42 ランの全ての分析は、基本的に技術的繰り返し測定であるはずですが、このデータセットでは経時的にすべてのペプチドに明らかなシステム的なシグナル劣化がありますが、200 分の 1 または 20 分の 1 に減少するものは非常に少数です。

これらのペプチドがお互いと比較して何時に溶出してくるかを見るには、以下を行います：

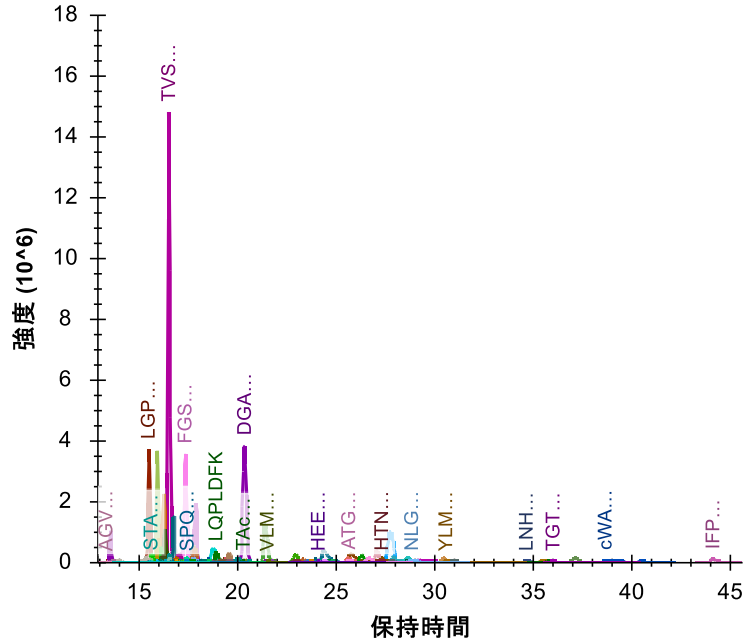
- アイコンが先頭にある「S」エレメントを選択します。

クロマトグラムグラフが変更され、「S」リスト内の 3 つのすべてのペプチドがまとめて表示されます。



ここでドキュメント内のすべてのペプチドをまとめて見るには:

- **[編集]**メニューで**[すべて選択]** (Ctrl-A) をクリックします。



これにより、問題があった 2 つのペプチドはどちらも親水性で早く溶出することがかなり明確となります。実際のところ、これらは本実験の 3 番目および 10 番目の溶出ペプチドです。これを Skyline で見るには、以下を行います:

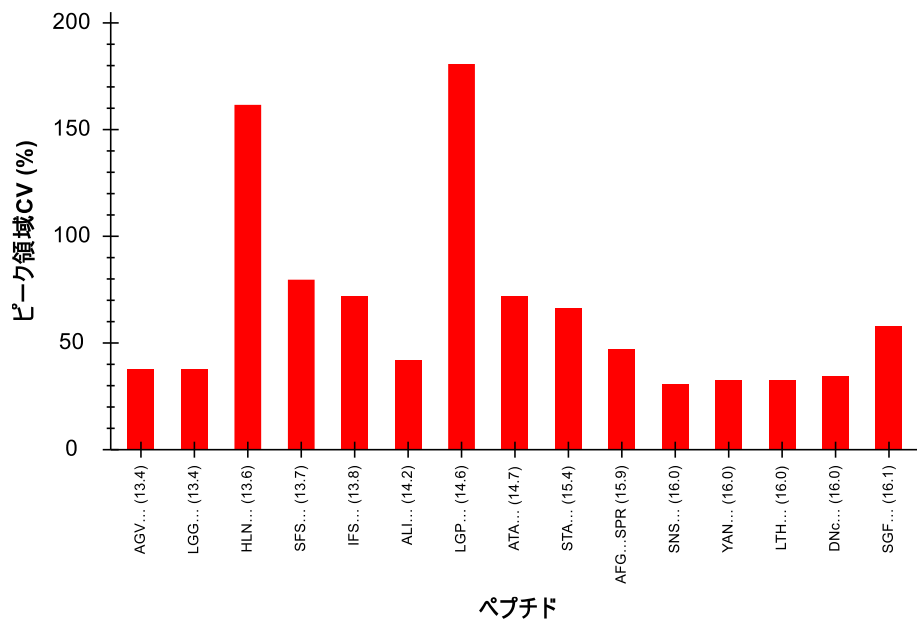
- [ビュー]メニューで、[保持時間]を選択して[ペプチド比較] (Ctrl-F8) をクリックします。
- [保持時間]ビューを右クリックして、[並び順]を選択し [保持時間] をクリックします。
- ペプチドが [保持時間] グラフ内に入るよう、長方形を最初の 15 分付近へとクリック & ドラッグします。

早く溶出するペプチドは変動が大きい傾向があります。したがって多くの場合、これらはグローバル正規化標準の最良の候補ではありません。これらのペプチドは、オートサンプラー内での注入待ちの間の劣化といった、その他の要因から影響を受ける可能性もあります。

別の方法で、これらのペプチドが周辺の他のペプチドにどれだけ類似して振舞うかを評価するには:

- [ビュー]メニューで、[ピーク領域]を選択して[ペプチドの比較] (Ctrl-F7) をクリックします。
- [ビュー]メニューで、[トランジション]を選択して[合計] (Ctrl-F10) をクリックします。
- [ピーク領域]ビューを右クリックして、[CV 値] をクリックします。
- ペプチドが [ピーク領域] グラフ内に入るよう、長方形を最初の 15 分付近へとクリック & ドラッグします。

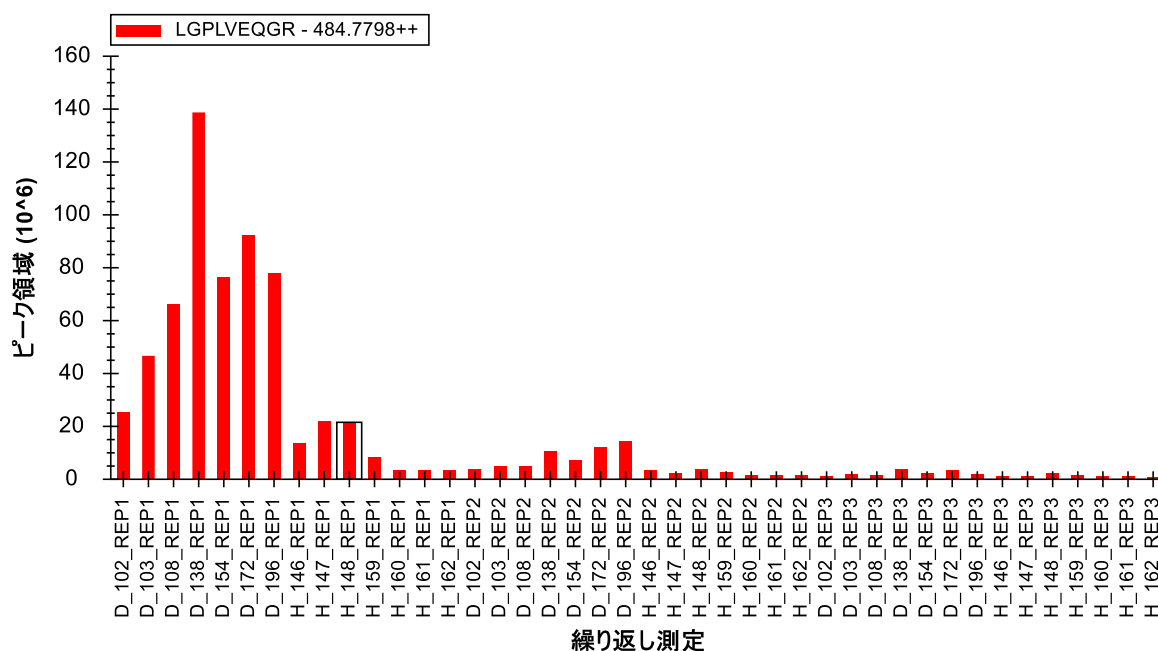
表示内容を理解しようとする前に、すべての非標準ペプチドは 14 個の被験体で測定されたためより大きな誤差を持つと予想されることを思い出してください。また積分は、すべてのペプチドについて完全に調整されているわけではありません。現在のデータセットについては、当該標準に関して自動的に選択されたピークは正しいように見えます。しかし、HLNGFSVPR ペプチドは約 160%の 2 番目に高い CV 値を持ち、また AFGLSSPR ペプチドの CV 値は約 50%で、最初の 15 個のペプチドの中央値に近くなっています。(より疎水性で遅く溶出する VVLSGSDATLAYSAFK ペプチドは 108 番目であり、または CV 値が 18.6%で 80 パーセントイルに入ります。)



以下を行うと、LGP ペプチドに HLN ペプチドと非常に似た問題があると確信できるはずですが。

- [ピーク領域] ビューの LGP ペプチドのバーをクリックします。
- [ビュー] 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。

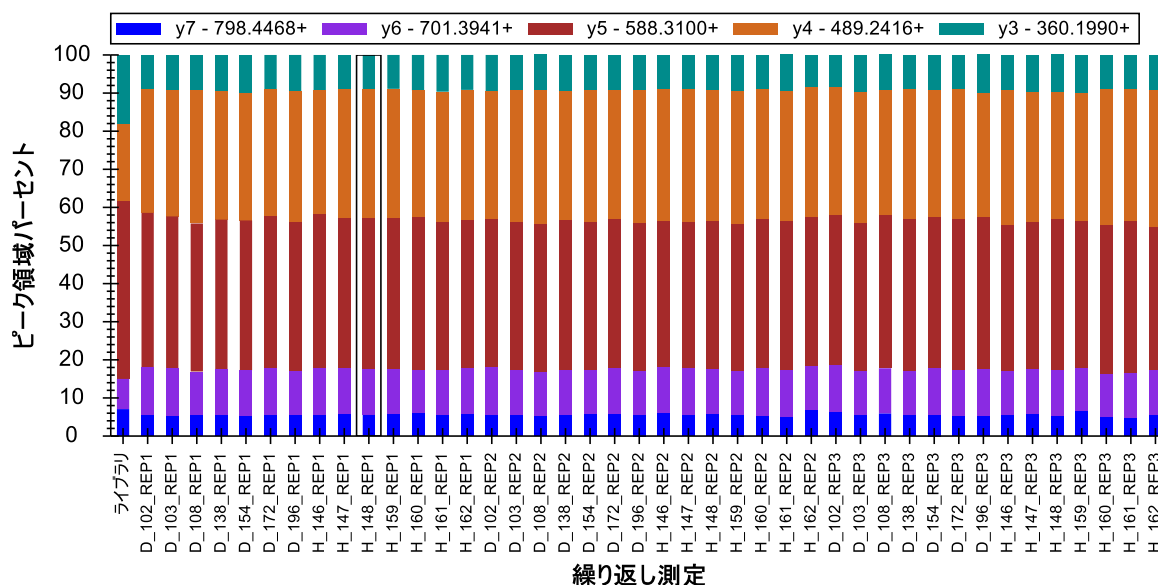
3 サイクル/技術的反復測定があるはずですので留意してください:



ピークが正しく積分されていることを確認するには、以下を行います：

- [ビュー] メニューで、[トランジション] を選択して [すべて] (F10) をクリックします。
- [ピーク領域] ビューを右クリックし、[正規化] を選択して [合計] をクリックします。

最小から最大まで 200 倍の範囲（1 億 4000 万～70 万）であるにもかかわらず、分析間の相対プロダクトイオン存在量は非常に安定しています。



このペプチドには同様の問題があるように見えますが、HLNGFSVPR のようなペプチドはグローバル正規化標準としての考察から除外すべきです。理想的には、標準ペプチドを重要な定量的データに注入する前に、この操作を実行します。この場合は実際、VVLGSDATLAYSFAK のみを、ドキュメント内の他の全てのペプチドに対するグローバル正規化標準として利用すべきです。この概念をさらに証明することは、本チュートリアル範囲外ですので割愛します。

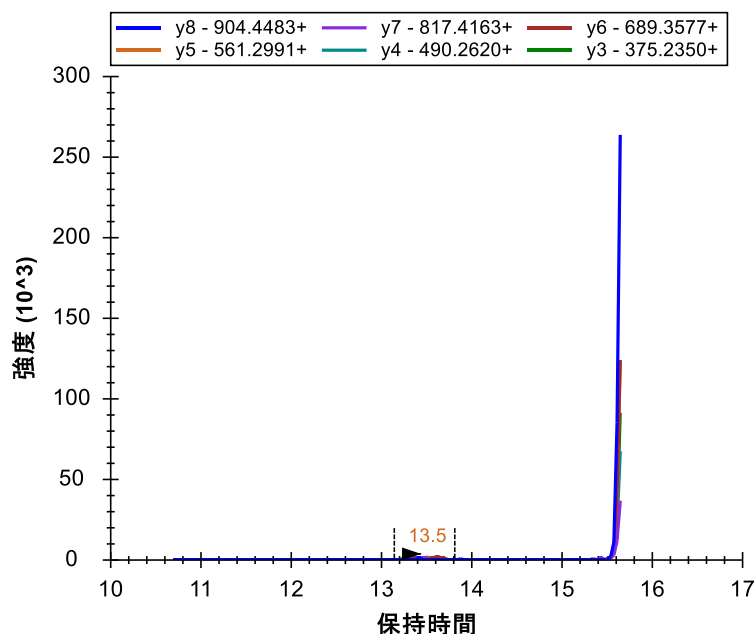
### 複数・繰り返し測定データ処理を続行する

ここでも、「S」と表示されている標準ペプチドリスト上のペプチド DVFSQQADLSR を選択して、本実験のピーク積分を引き続き再確認・修正していきます。

- [ビュー] メニューで、[保持時間] を選択して [繰り返し測定比較] をクリックします。グラフ中で見られる[保持時間]や[ピーク領域]の一貫性から、すべてのランにおいてペプチドが一定して積分されているという信頼性が得られるはずですが、このペプチドをさらに再確認しな

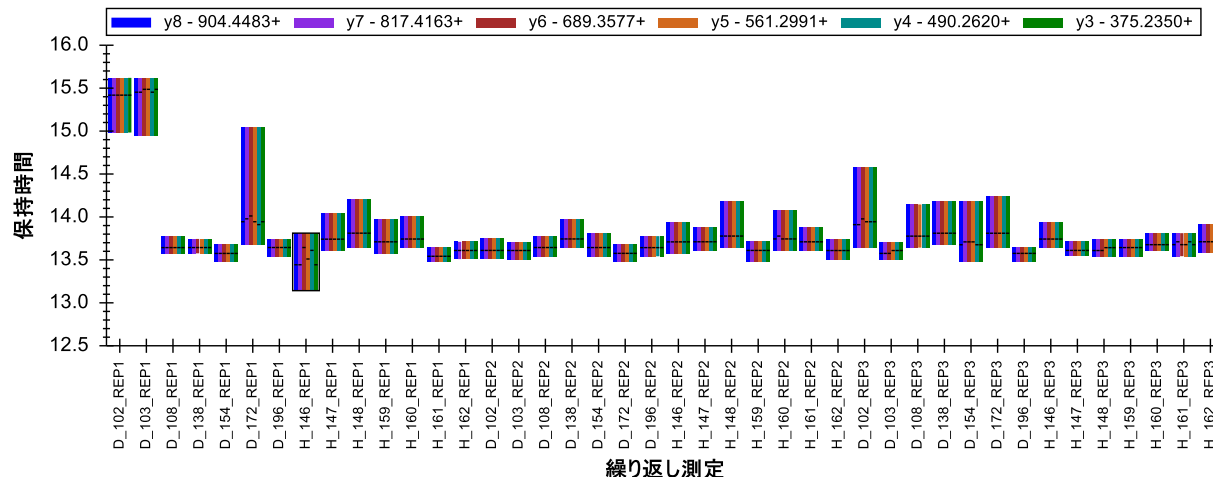
いで、[ターゲット] リストを続行して構いません。これは次の2つのペプチドにも当てはまりません。

ペプチド IFSQQADLSR に達するまでリストを続行します。H\_146\_REP1 を除くすべての繰り返し測定で[ピーク領域] グラフ中に一貫した相対イオン存在量があることに気付くでしょう。ここでも、このランでのバーをクリックすると、H\_146\_REP1 のピークが5分間のスケジュールウィンドウに完全に収まっていないことがわかります。（場合によっては、これを見るにはx-軸ズーム（Shift-F11）をオフにする必要があります。）

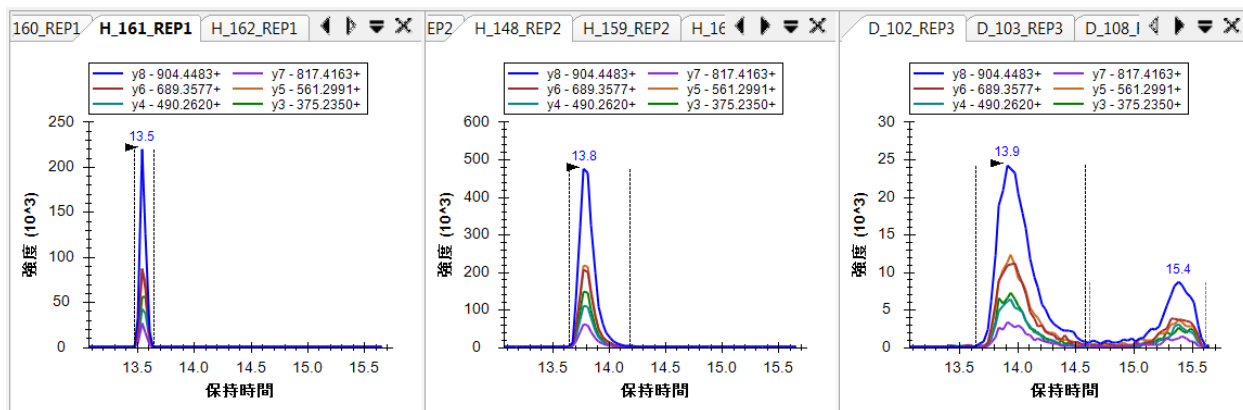


この問題については、以前行ったように、x-軸の下をクリック&ドラッグして切断ピーク（truncated peak）を積分するか、又は右クリックメニューを利用してピークを完全に削除することで修正することができます。

しかし、これはこのペプチドに限った問題ではありません。[保持時間] グラフに注目すると、積分される時間の範囲内に、通常とは異なった差異がいくつか見られます。



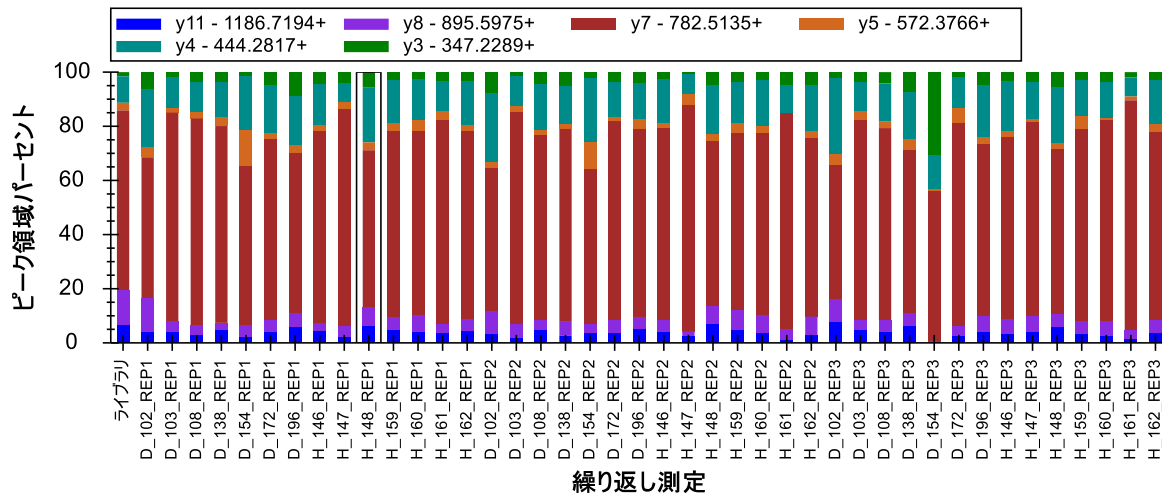
このペプチドに関するいくつかのクロマトグラムを再確認すると、多くのものが 0.2 分間に溶出する単一の非常に優良なピークを示しているのが分かります。しかしその他については、溶出が約 0.5 分間にわたってテーリングしているピークや、2 分間にわたる溶出で 2 つの明白なピークを持つペプチドが存在します。



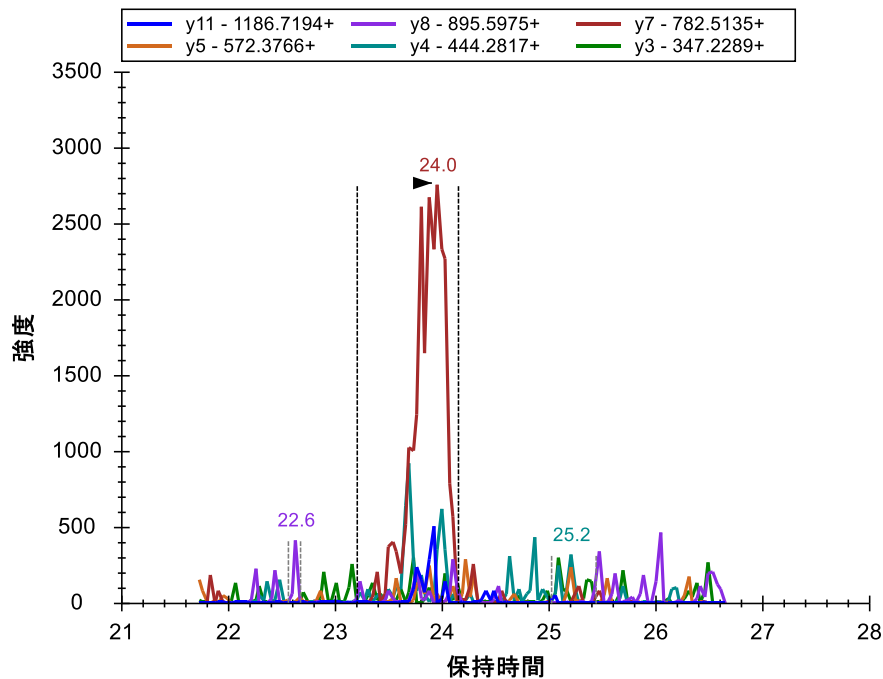
このようなペプチドは、特に一致 SID ペプチド (matching SID peptide) が無ければ、定量化に使用することは非常に困難です。本実験からこのペプチドを消去した方がよいかもかもしれません。せめて、二重溶出プロファイルを示すピークは削除した方が良いです。

続いてペプチドリストを見てみると、次の 7 つのペプチドでは十分に良好な一貫性があることがサマリープロットから示されており、Skyline を使ってピーク選択を修正する必要はなく、これらのペプチドを一瞥するだけで十分ということが分かると思います。

しかしペプチド MLSGFIPLKPTVK を見ると、大きな変動があることが[ピーク領域] グラフから分かります。



このペプチドの総ピーク領域では、y7 イオンが優位となっています。これは、ライブラリスペクトルにも反映されていますが（上記グラフの左端に表示）、実際に y7 イオンと共溶出するものがあるかどうかについてクロマトグラムを目で見ただけでは識別することは困難です（以下を参照）。

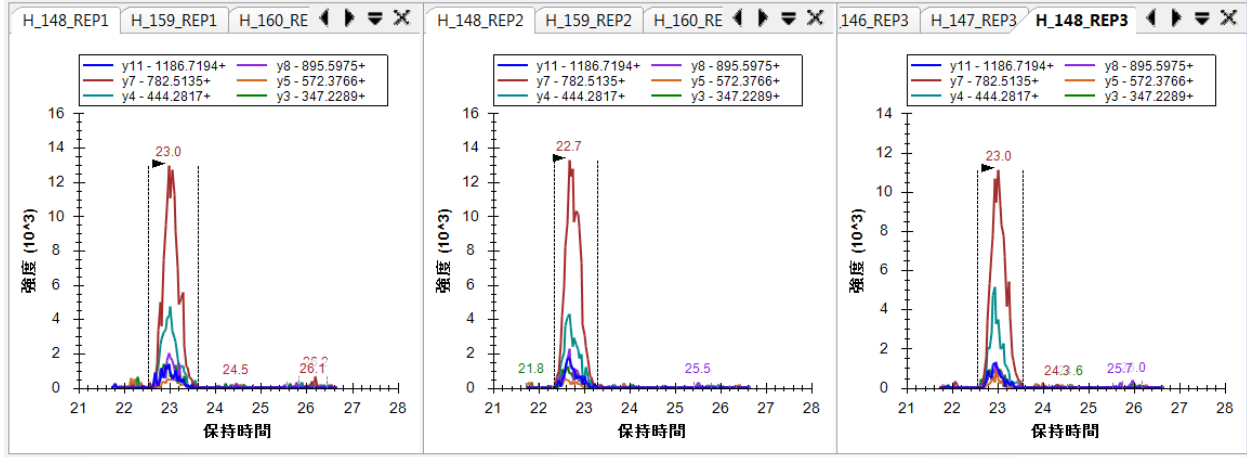


しかし Skyline はほとんどの場合、各ランにおいて同一分子に対する同一ピークを積分しているということも明らかです。y7 ピーク以外のものがあるかを確認するには、最高強度の測定値に注目してください。以下の手順でこれを行うことができます：

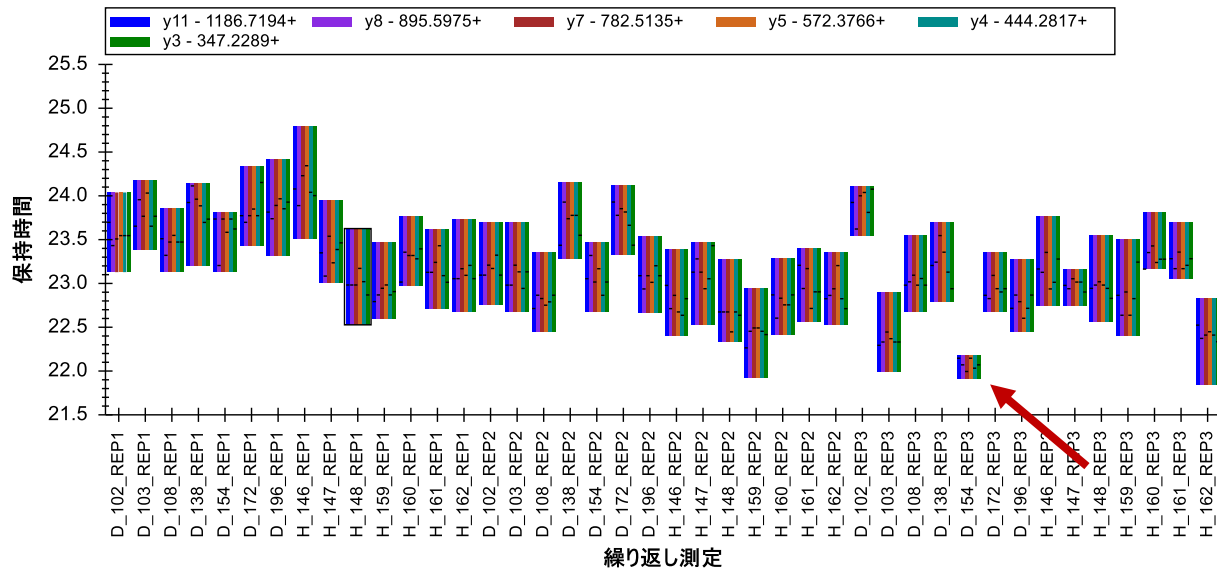
- [ピーク領域] グラフを右クリックし、[正規化] を選択して [なし] をクリックします。



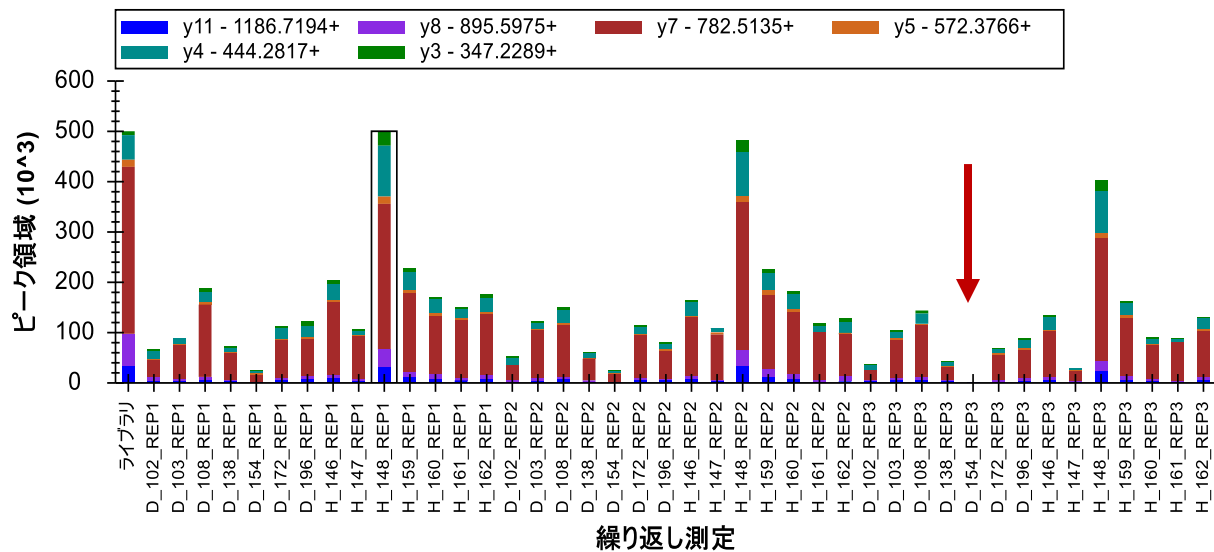
- 3つの明らに最高強度であるピーク（全てサンプル H\_148 のピーク）をクリックします。これらの3つのプロット内で見られるクロマトグラムピークから、ペプチド存在量が十分に高い場合、23~24分で y7 ピークを持つペプチドは6つのトランジション全てで共溶出信号が得られている、という事を十分に理解できます。



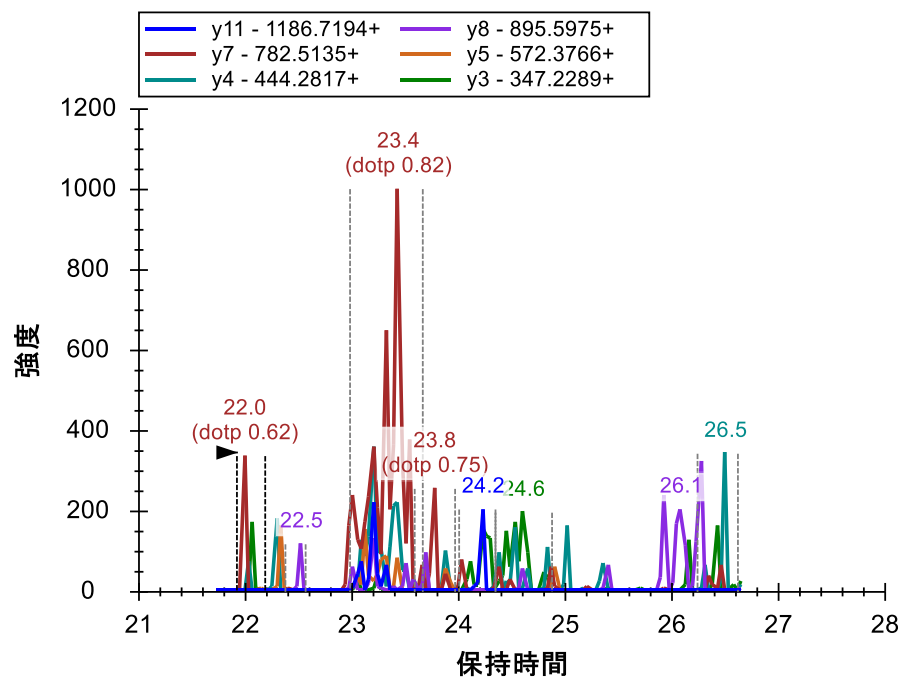
このペプチドを消去するよりは、Skyline によって誤ったピークが選択されているデータを修正するのが良いでしょう。これは、[保持時間] プロットにおいて、22.5分未満に現れる短いバンドとして確認できます。



また [ピーク領域] プロットでも、このランには視認できる積分信号がないことが分かります。



[保持時間] ビュー内のバーをクリックして、このランのクロマトグラムグラフを有効化します。このランが y7 トランジションにおいて非常に低い信号を有しているのが分かります。周辺のランから得たデータを確認せずに、ペプチドを代表するトランジションとしてこのトランジションのみを選択することはまずないでしょう。



このランの積分を修正するには、以下を行います:

- D\_154\_REP3 のグラフ内で、ピーク注釈「23.4 (dotp 0.82)」をクリックします。

ピーク積分の再確認を続ける前に、以下の操作を行って、相対イオン存在量を示す [ピーク領域] グラフに戻ります:

- [ピーク領域] ビューを右クリックし、[正規化] を選択して [合計] をクリックします。これで修正されたピークは、サマリープロット内のその他のランと前と比べて、より一致していることが分かります。

上記ペプチドへと続行します。ペプチド GMYESLPVAVK ではサマリープロットに一貫性がないので、このペプチドを消去するだけにしてピーク積分の調整は行わない方がよいでしょう。ペプチド ETGLMAFTNLK では、1 つのランのみで調整が必要です。このランは [ピーク領域] グラフ内で非常に明確であるはずですが、チュートリアル時点までで、何が間違っていて、その問題をどのように修正するかについて、理解が深まったことと思います。

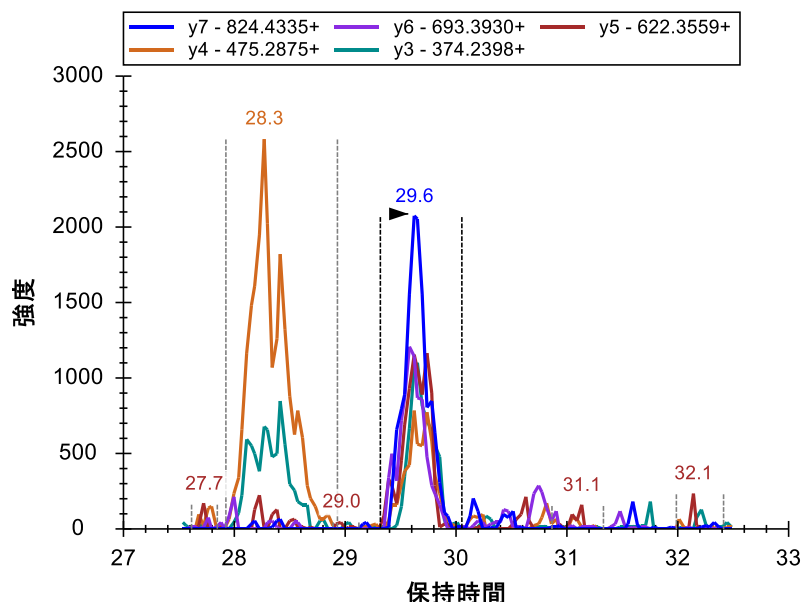
この修正を行った後でも、ETGLMAFTNLK ペプチドの全てのケースにおいて同一ペプチドの同一ピークが積分されていると確信できない場合、以下のいずれかの操作を行って、すべてのクロマトグラムグラフを迅速に再確認することが可能です:

- [ターゲット] ビュー上部の [繰り返し測定] ドロップダウンリストをクリックします。
- 上向きおよび下向き矢印キーを使用して、ランを横送りしクロマトグラムグラフを有効化します。

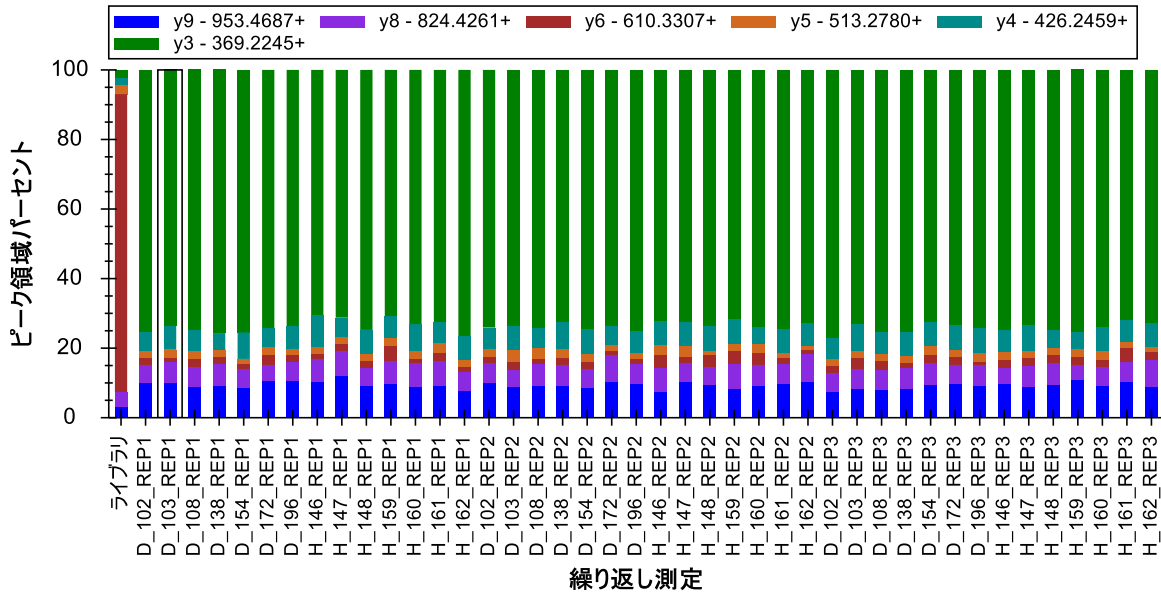
または

- Ctrl-上向き矢印または Ctrl-下向き矢印を押して、ランを横送りします。

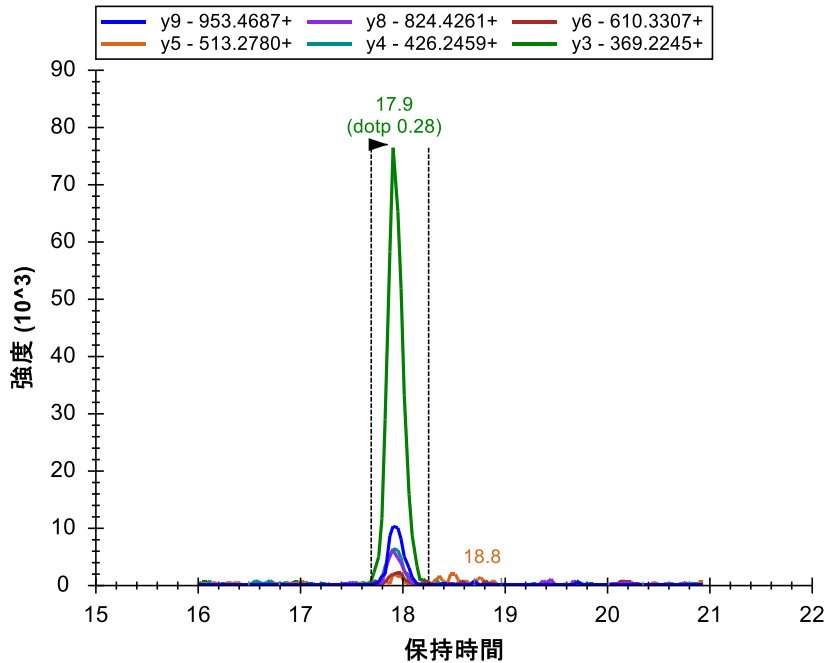
多くのグラフで、選択したピークの約 1.5 分前の y3 および y4 トランジションでペプチドが干渉しているのが見られます。これにより、積分ピークの一貫性における信頼度が高まります。



[ターゲット] リストを続行していくと、複数のペプチド（YANVIAYDHSR と TDEDVPSGPPR）が見られ、測定ピークの相対イオン存在量と一致ライブラリスペクトルとの間の差が明白となっています。これらの差は[ピーク領域] グラフ内で見ることができます：



これはクロマトグラムグラフ内でも観察でき、Skyline ではピーク保持時間の下にドット積の値「(dotp 0.28)」が表示されます。



Skyline は、dotp 値が高いグラフ上に別のピークがある場合に、このような「dotp」値を表示します。上記のケースでは、選択したピークに対してノイズレベル付近の低いピークしか見られないため、より良い dotp 値を持つピークを見つけることはできません。上記を見るためには、以下を行います：

- 解析対象のクロマトグラムグラフをクリックして有効化します。
- マウスマウスカーソルを y-軸の左側に置きます。
- マウスのスクロールホイールまたはトラックパッドのスクロールジェスチャーを使用して、前方にスクロールします。

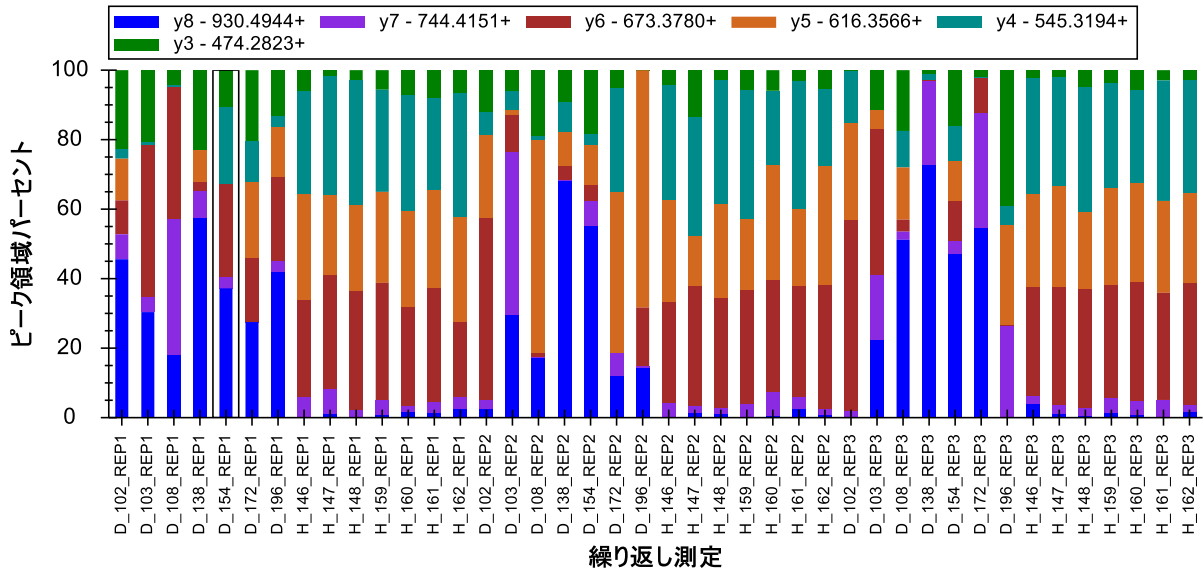
より良い dotp 値を持つ小さいピークが見えるようになるまで、グラフの y-スケールをズームします。

本実験で使用したものと同様の質量分析計で当該ライブラリスペクトルが収集されたという確信が強ければ、この積分ピークは目的のペプチドを正しく測定していないのではないかと疑念が起きます。しかしこの場合、このスペクトルのソースについてはあまりよく分かっておらず、このサンプルで当該ペプチドを測定する他の選択肢は無いでしょう。

[ピーク領域] グラフおよび [保持時間] グラフにおけるラン間の一貫性から、6 つの共溶出トランジションを介して全てのランで同一ペプチドが測定されていることが確信できるはずですが。最も豊富なイオンが選択性の低い y3 イオンであるにもかかわらず、このトランジションを消去して他のイオン（y9 が 2 番目に最も豊富なイオン）の一貫性および共溶出をより詳しく分析することが可能です。このピークの品質に確信が持てた場合は、[元に戻す] ボタン（Ctrl-Z）を使用して、消去した y3 トランジションを Skyline ドキュメントへと戻します。

TDEDVPSGPPR の上の 7 つのペプチドの中では、ペプチド SPQGLGASTAEISAR について誤って積分されたピークが 1 つだけ見られます。今となっては、この問題を比較的簡単に見つけ、修正できるでしょう。

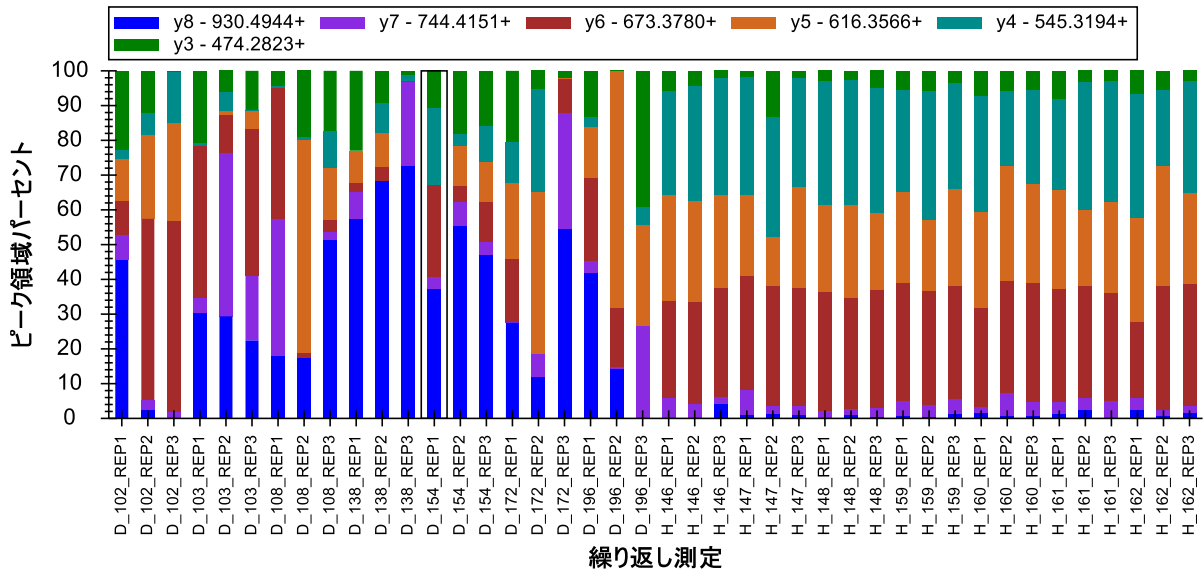
ペプチド CSSLLWAGAAWLR へと移ります。サマリープロットで観察される変動から判断して（以下を参照）、最初はこのペプチドを消去して先に進みたいと思われるかもしれません：

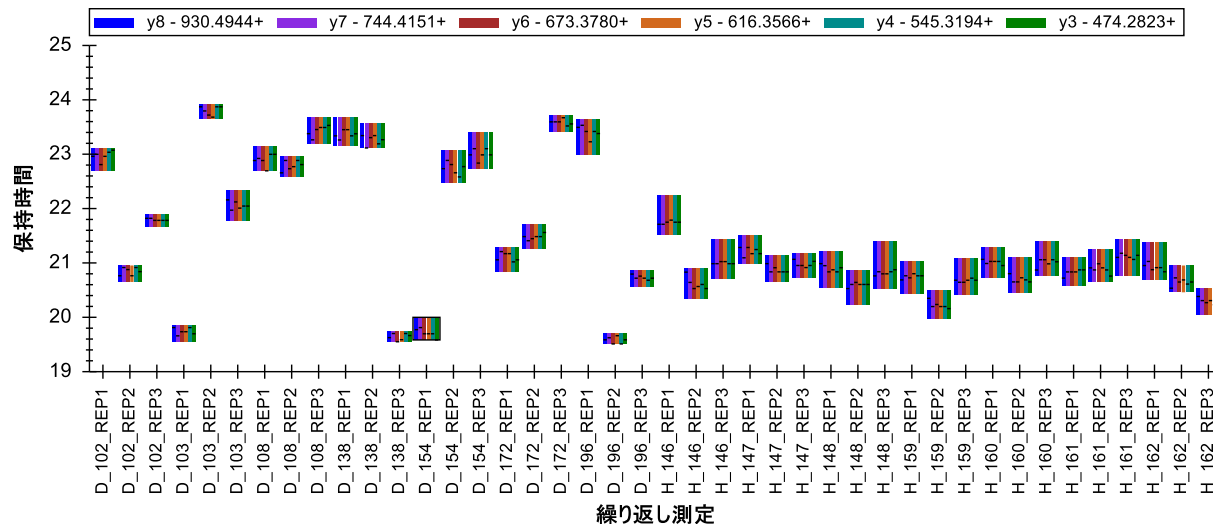


しかし注視してみると、健常被験体の中に一貫性のある領域が見られます。これをよりはっきりと示すため、以下を行います:

- [ピーク領域] グラフを右クリックし、[並び順種別] を選択して [ドキュメント] をクリックします。

これによりグラフは以下のように変更されます:

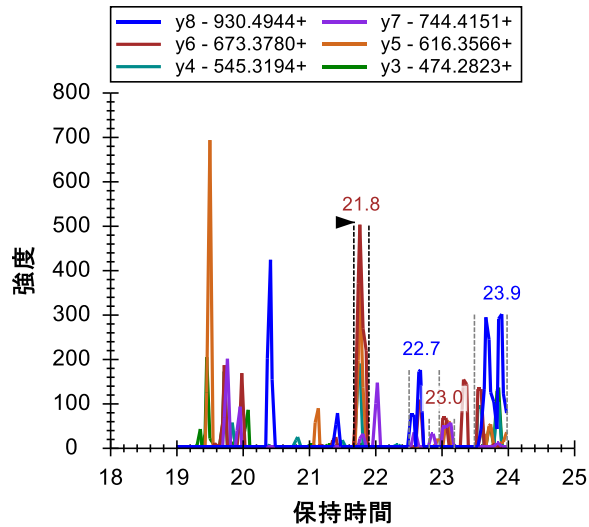




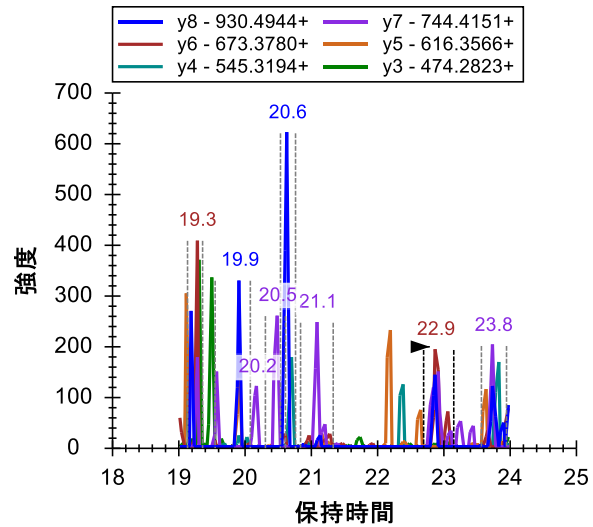
健常被験体では一定のピークが積分されていて、罹患被験体についてはそうでは無い事が明確になります。健常被験体のピークを再確認すると、かなり強度が低く主に y4、y5、および y6 イオンの共溶出に依存しているのが分かります。この共溶出はおそらく、1つのペプチドに起因しています。しかし、このペプチドが実際に CSSLWAGAAWLR であるか否かを判定することは困難です。

罹患被験体のクロマトグラムを見てみると、y6 トランジションの 21 分付近にピークと思われるものが時折見られます。しかし大抵の場合、一貫した溶出パターンを観察することは困難です：

D\_102\_REP3

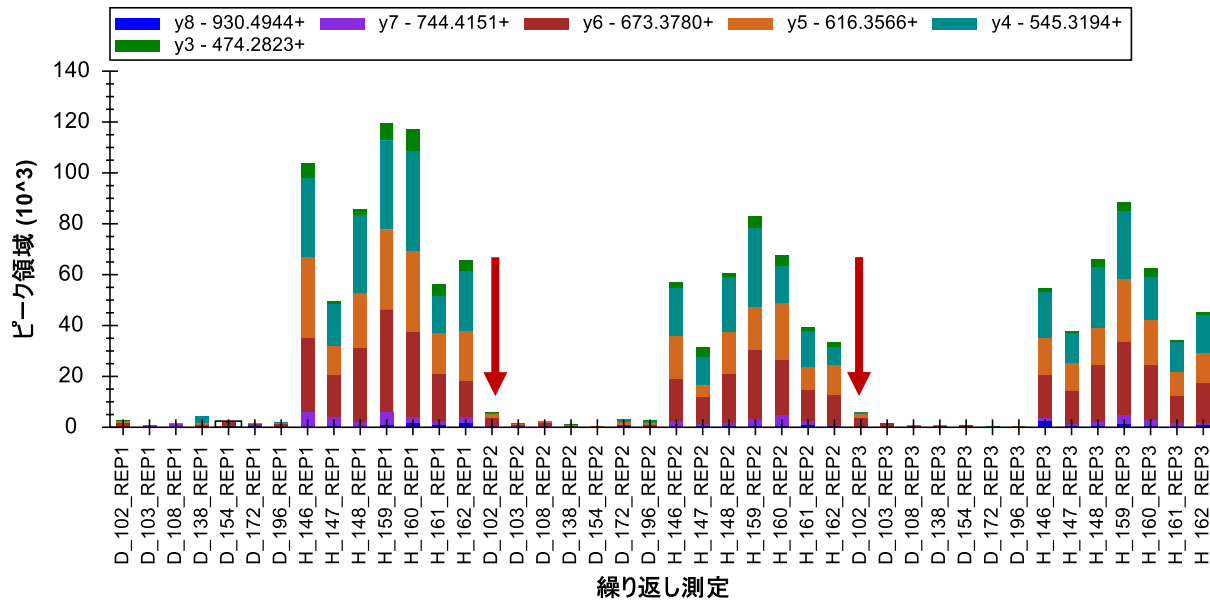


D\_108\_REP1



しかし明らかに、罹患被験体のクロマトグラム内には、健常被験体で観察されたのと同じようなピーク領域を作り出せるピークは見つかりません。これまでにピークドリフトや切断の操作で得てきた経験を用いれば、このピークは単に全ての罹患被験体の測定範囲外で点々に散っているわけではないと確信できるますが、健常被験体に関しては違います。注入のランダム化により、上記の信頼度が増す可能性もあります。しかし現在の研究については、3つの繰り返し測定サイクルで十分なはずですが、最善を尽くして積分を修正し終わったら、このペプチドについて視認できるピークを持つ罹患群のランのうち2つが、健常被験体のランのすぐ後に続いていることにも気付くかもしれません (D\_102\_REP2 および D\_102\_REP3)。これは、キャリアオーバー効果を示唆している可能性があります。





結局のところ、これらのトランジションは本研究におけるバイオマーカーの最有力候補の 1 つを表している可能性があります。合成ペプチドを使うことで、CSSLLWAGAAWLR ペプチドを正しく測定しているという信頼性を高めることができます。また、ペプチド検索エンジンでこのペプチドが同定されるかどうか確かめるために、このピークの MS/MS スペクトルを取得（並列反応モニタリング（PRM）を使用）することもできます。最後に、上記の操作が失敗した場合は、健常被験体で見られるピークがどのようなペプチドかを同定するため、MS/MS スペクトル上で *de novo* シーケンシングを行うことが可能です。ターゲットプロテオミクスでは、標的ペプチドの同定作業から開始することが常に求められるわけではありません。おそらく、SDS-PAGE ゲルバンドによく似ており、差異を見つける事が解析の始まりとなり得ます。

この時点で、積分を修正したり解析に適さないペプチドを削除したりすることで、残りのペプチドを処理していくことができるようになってはいるはずですが、この操作は 1 時間以内でできるでしょう。ピーク形状が良好で、予定された保持時間ウィンドウ内で完全に溶出するペプチドは、ほとんど修正する必要がないということが分かると思います。これは、[ピーク領域] グラフおよび [保持時間] グラフを一目見て分かることが多いです。ピーク形状が悪い、又は多くのピークがスケジュールウィンドウ内で切断もしくは欠損したペプチドについては、そのペプチドは単に消去することにして構いません。しかし、この結論は早く出し過ぎないでください。さもないと、興味深い発見を見逃す可能性があります。

### 繰り返し注釈のある統計解析の準備

Skyline ドキュメント内の積分についての問題を注意深く検証して修正した後は、実験条件間で見られるペプチド存在量においてどのような差異があるかをより深く理解することに関心を向けていきたいのではないのでしょうか。Skyline または外部統計ツールを使ってそのような解析を行うには、Skyline で通常「繰り返し測定」と呼ばれている、測定サンプルを更に分類する操作

が必要となります。そのような分類を行うために、Skyline は繰り返し測定注釈を用意していません。本チュートリアルでは、3 つの繰り返し測定注釈（SubjectId、BioReplicate 及び Condition）を使用します。

SubjectId 注釈を定義するには、以下の手順を実行します：

- [設定] メニューで [ドキュメント設定] をクリックします。
- [ドキュメント設定] フォームの [注釈] タブ内の [リストを編集] ボタンをクリックします。
- [注釈を定義] フォームの [追加] ボタンをクリックします。
- [注釈を定義] フォームの [名前] に「SubjectId」と入力します。
- [適用先] リストで、[繰り返し測定] のチェックをオンにします。

フォームは以下のように見えるはずです：

注釈を定義

名前(N):  
SubjectId

タイプ(T):  
テキスト

値(V):

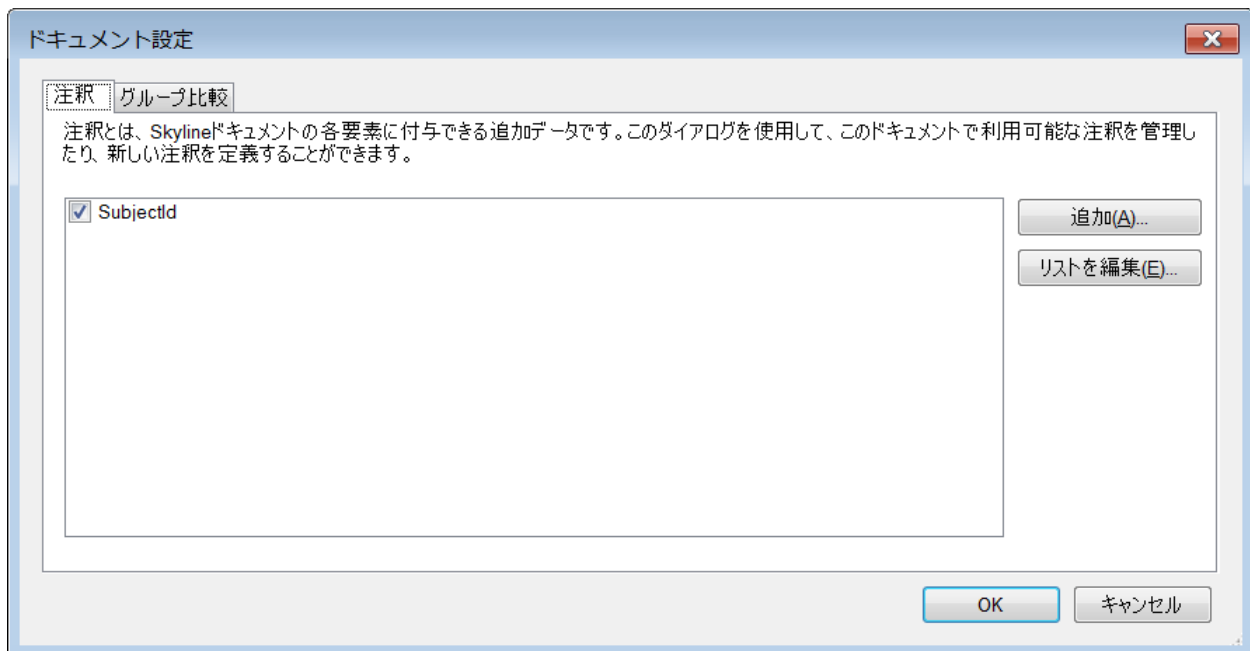
適用先(A):

- 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

- [注釈を定義] フォームの [OK] ボタンをクリックします。
- [注釈を定義] フォームの [OK] ボタンをクリックします。
- 今作成した「SubjectId」注釈のチェックをオンにします。

[ドキュメント設定] フォームは以下のように見えるはずです：



- [OK] ボタンをクリックします。

本チュートリアルは、Skyline 外部ツール「MSstats」で利用可能な統計分析メソッドに関しては詳しく取り上げませんが、このツールは今回のような解析に適しています。本データセットは、数多くのトレーニングコースやワークショップで MSstats の使用法を実地説明するのに利用されています。この種の解析において MSstats でできる事に関心がある場合、2 つの別の注釈が必要になりますので、以下のように Skyline へとインストールしてください:

- [ツール] メニューで、[ツールストア] をクリックします。
- ツールのリストから「MSstats」を選択します。

[ツールストアからインストール] フォームは次のようになっています:

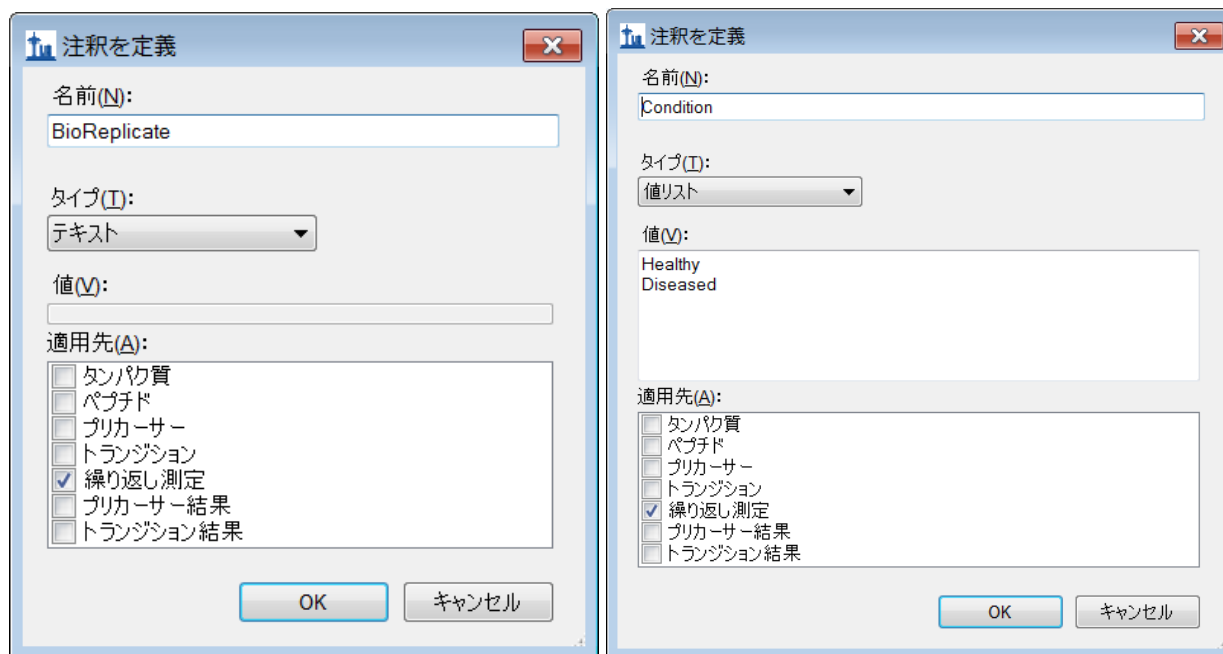


- [インストール] ボタンをクリックし、MSstats のインストールが完了するまでプロンプトに従います。

MSstats のインストール中に追加した注釈を見る、又は Msstats をインストールせずに注釈を直接追加するには、以下の手順を実行します:

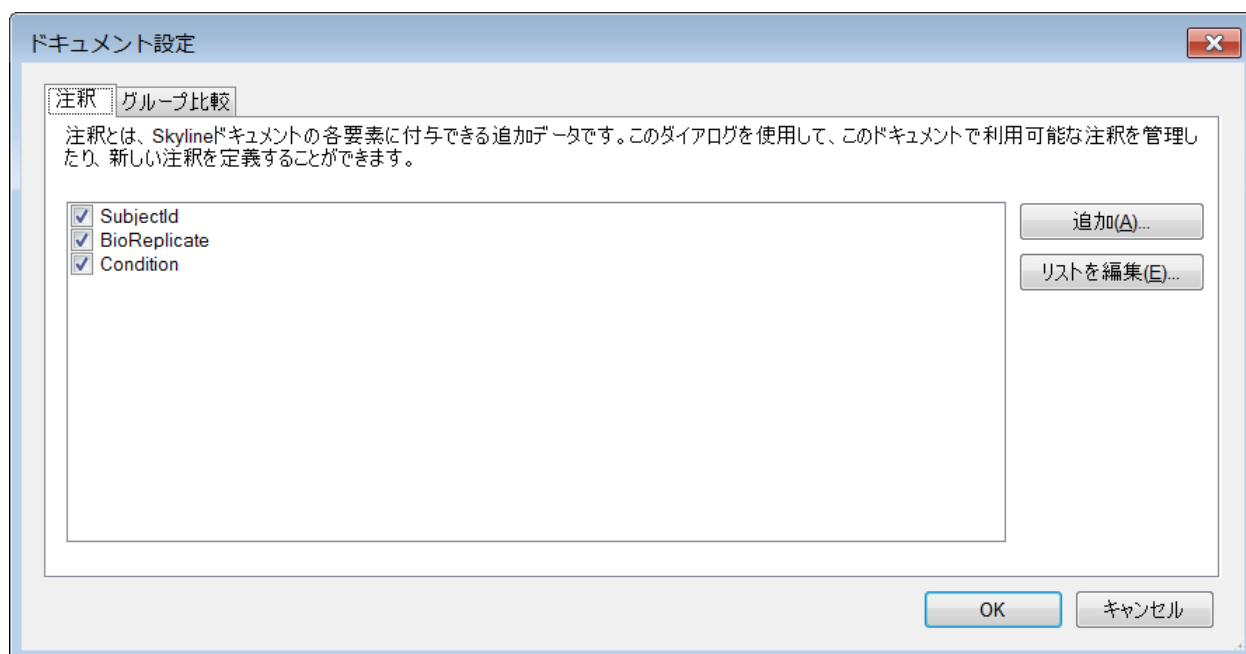
- [設定] メニューで [ドキュメント設定] をクリックします。
- [ドキュメント設定] フォームの [注釈] タブ内の [リストを編集] ボタンをクリックします。
- MSstats をインストールした場合は、BioReplicate 注釈を選択し [編集] ボタンをクリックします。それ以外の場合は、[追加] ボタンをクリックします。
- Condition 注釈についても同様に繰り返します。

これらの注釈の定義は以下のように見えるはずです:



- [OK] をクリックして、[ドキュメント設定] フォームに戻ります。
- BioReplicate 注釈と Condition 注釈のチェックをオンにして、これらをドキュメントに追加します。

[ドキュメント設定] フォームは以下のように見えるはずです:



- [OK] ボタンをクリックします。

ドキュメントに追加した注釈を設定するには、以下を行います：

- [ビュー] メニューで、[ドキュメントグリッド] をクリックします。
- [ドキュメントグリッド] の左上角の [ビュー] メニューで、[繰り返し測定] をクリックします。

[ドキュメントグリッド] は以下のように見えるはずですが：

	繰り返し測定	SubjectId	BioReplicate	Condition
▶	<a href="#">D_102 REP1</a>			▼
	<a href="#">D_102 REP2</a>			▼
	<a href="#">D_102 REP3</a>			▼
	<a href="#">D_103 REP1</a>			▼
	<a href="#">D_103 REP2</a>			▼
	<a href="#">D_103 REP3</a>			▼
	<a href="#">D_108 REP1</a>			▼
	<a href="#">D_108 REP2</a>			▼
	<a href="#">D_108 REP3</a>			▼

これで、すべての 42 個の繰り返し測定の注釈をこのグリッドへと手動で入力できます。スプレッドシートからこのフォームに直接貼り付けることも可能です。これを行うには、以下の手順を実行します：

- Excel で「GroupedStudies1\Heart Failure\raw」フォルダ内の「Annotations.xlsx」ファイルを開きます。
- 3つの注釈列について、ヘッダー行の下 42 行のセルをすべて選択します。
- Ctrl-C を押してセルをコピーします。
- Skyline で、SubjectId 列の一番上のセルを選択します。

注: この工程では、セル編集モードの利用は避けてください。一番上の SubjectId セル内のカーソルが点滅したら Esc キーを押してください。[ドキュメントグリッド] は以下のように見えるはずですが：

ドキュメントグリッド

ビュー ▾ | 1 of 42 | エクスポート... | 検索:

	繰り返し測定	SubjectId	BioReplicate	Condition
▶	<a href="#">D_102 REP1</a>			
	<a href="#">D_102 REP2</a>			
	<a href="#">D_102 REP3</a>			
	<a href="#">D_103 REP1</a>			
	<a href="#">D_103 REP2</a>			
	<a href="#">D_103 REP3</a>			
	<a href="#">D_108 REP1</a>			
	<a href="#">D_108 REP2</a>			
	<a href="#">D_108 REP3</a>			

- Ctrl-V を押して、[ドキュメントグリッド] へと貼り付けます。  
[ドキュメントグリッド] 内のセルにスプレッドシートの値が、以下のように入力されます:

ドキュメントグリッド

ビュー ▾ | 42 of 42 | エクスポート... | 検索:

	繰り返し測定	SubjectId	BioReplicate	Condition
	<a href="#">H_159 REP3</a>	H159	159	Healthy
	<a href="#">H_160 REP1</a>	H160	160	Healthy
	<a href="#">H_160 REP2</a>	H160	160	Healthy
	<a href="#">H_160 REP3</a>	H160	160	Healthy
	<a href="#">H_161 REP1</a>	H161	161	Healthy
	<a href="#">H_161 REP2</a>	H616	161	Healthy
	<a href="#">H_161 REP3</a>	H161	161	Healthy
	<a href="#">H_162 REP1</a>	H162	162	Healthy
	<a href="#">H_162 REP2</a>	H162	162	Healthy

### 問題ピークのあるペプチドに注釈を付ける

削除されたピーク又は切断ピークのあるペプチドに、注釈を付けることも推奨されます。なぜなら、どちらの場合も統計分析中に問題がおこる可能性があるからです。このような目的で使用する注釈を定義するには、以下の手順を実行します:

- [設定] メニューで [ドキュメント設定] をクリックします。
- [ドキュメント設定] フォームの [注釈] タブ内の [リストを編集] ボタンをクリックします。
- [注釈] リスト内の最後の注釈を選択します。
- [追加] ボタンをクリックします。
- [注釈を定義] フォームの [名前] フィールドに「MissingData」とタイプします
- [タイプ] ドロップダウンリストから「True/False」を選択します。

- [適用先] リストで、[ペプチド] のチェックをオンにします。  
[注釈を定義] フォームは以下のように見えるはずです:

注釈を定義

名前(N):  
MissingData

タイプ(T):  
True/False

値(V):

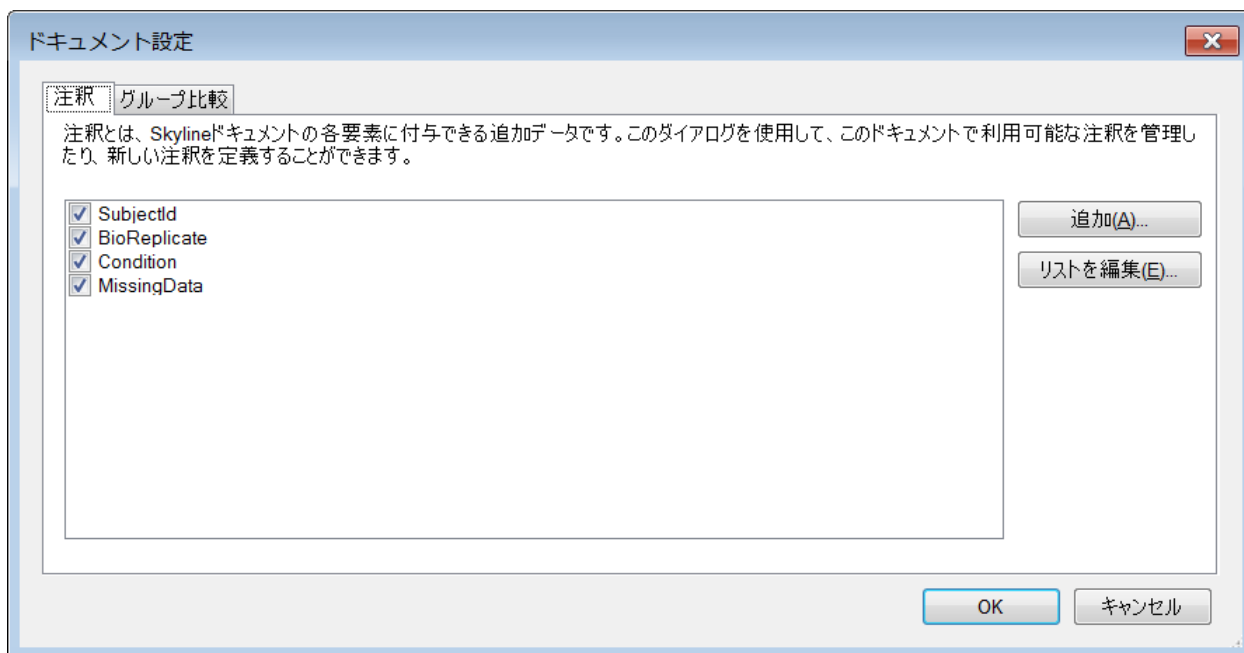
適用先(A):

- 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。
  - [注釈を定義] フォームの [OK] ボタンをクリックします。
  - 先ほど定義した「MissingData」注釈のチェックをオンにします。
- [ドキュメント設定] フォームは以下のように見えるはずです:



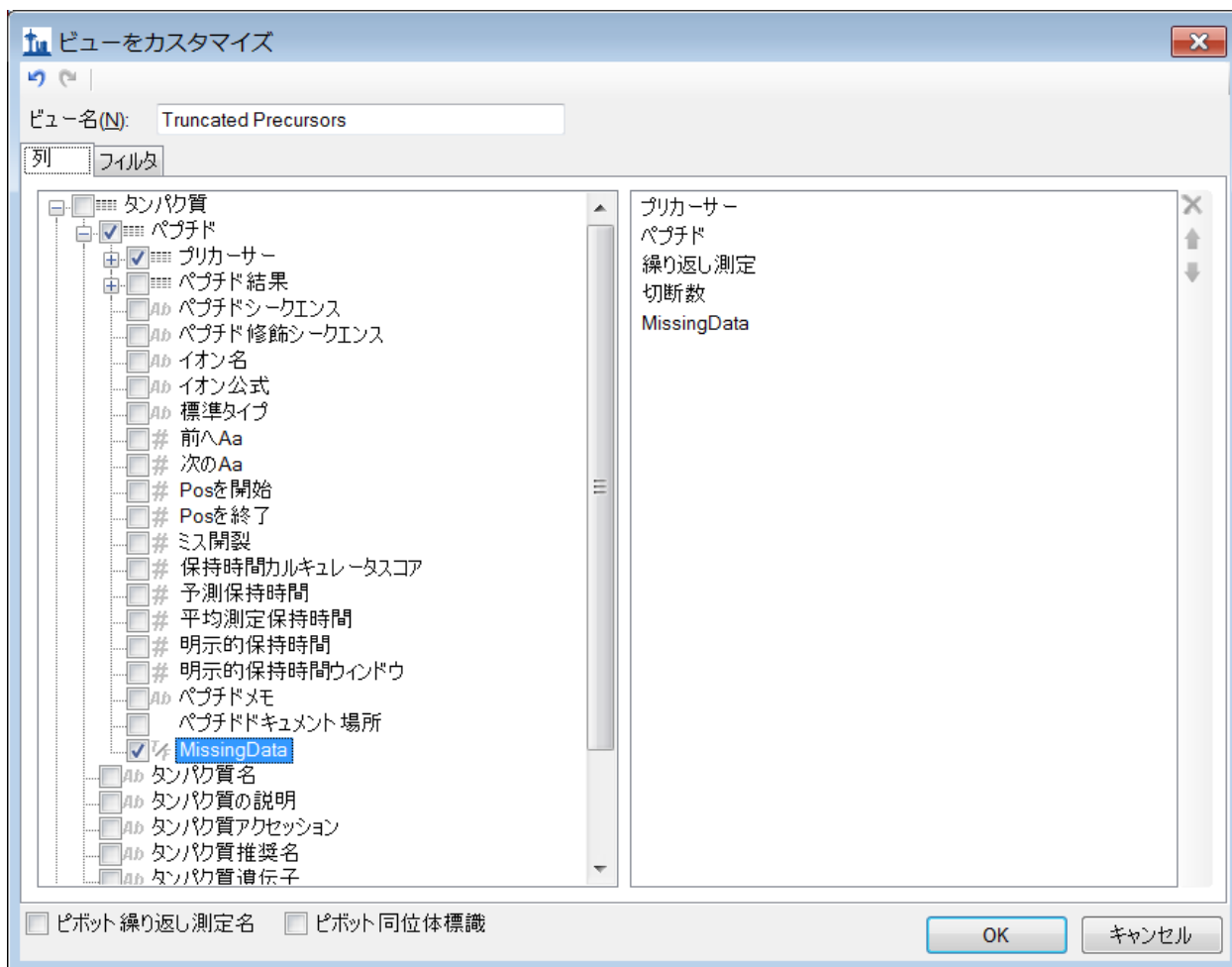


- [OK] ボタンをクリックします。

「MissingData」注釈を切断ピークのある全てのペプチドに設定する準備をするために、以下を行います：

- [ドキュメントグリッド] の [ビュー] メニューで、先に作成した「切断プリカーサー」ビューをクリックします。
- 同じ [ビュー] メニューで、[編集] ビューをクリックします。
- [ビューをカスタマイズ] フォームで、[列] ツリーを展開して、タンパク質 > ペプチド > MissingData とチェックして、先ほど作成した注釈の列を追加します。

[ビューをカスタマイズ] フォームは次のようになります：



- [OK] ボタンをクリックします。
- 「切断数」列ヘッダーを2回クリックして、降順に並べ替えます。

[ドキュメントグリッド] は以下のように見えるはずです:

ドキュメントグリッド

ビュー ◀ ▶ ◀ ▶ 1 of 223 | エクスポート... 検索: A

	プリカーサー	ペプチド	繰り返し測定	切断数	MissingData
▶	<a href="#">614.3164++</a>	<a href="#">DFATVYVDAVK</a>	<a href="#">D_108_REP2</a>	7	<input type="checkbox"/>
	<a href="#">614.3164++</a>	<a href="#">DFATVYVDAVK</a>	<a href="#">D_196_REP3</a>	7	<input type="checkbox"/>
	<a href="#">614.3164++</a>	<a href="#">DFATVYVDAVK</a>	<a href="#">H_147_REP2</a>	7	<input type="checkbox"/>
	<a href="#">614.3164++</a>	<a href="#">DFATVYVDAVK</a>	<a href="#">H_148_REP2</a>	7	<input type="checkbox"/>
	<a href="#">614.3164++</a>	<a href="#">DFATVYVDAVK</a>	<a href="#">H_159_REP1</a>	7	<input type="checkbox"/>
	<a href="#">614.3164++</a>	<a href="#">DFATVYVDAVK</a>	<a href="#">H_159_REP3</a>	7	<input type="checkbox"/>
	<a href="#">614.3164++</a>	<a href="#">DFATVYVDAVK</a>	<a href="#">H_160_REP2</a>	7	<input type="checkbox"/>
	<a href="#">614.3164++</a>	<a href="#">DFATVYVDAVK</a>	<a href="#">H_162_REP3</a>	7	<input type="checkbox"/>
	<a href="#">730.8488++</a>	<a href="#">DYVSQFESSTL...</a>	<a href="#">D_196_REP2</a>	7	<input type="checkbox"/>
	<a href="#">730.8488++</a>	<a href="#">DYVSQFESSTL...</a>	<a href="#">D_196_REP3</a>	7	<input type="checkbox"/>

残りのドキュメントの処理方法によっては、切断プリカーサーピークグループの数は223個ちょうどとなるかもしれませんが、ならないかもしれません。ペプチドの「MissingData」チェックボックスに1個ずつチェックを入れることが可能です。任意ペプチドのボックスのチェックを入れたら、残りは自動的にオンになります。なぜならこの注釈は、どのペプチドにも一度だけ適用されるからです。以下の操作を行って、これを試してみましょう:

- グリッドの最初の行（「MissingData」ヘッダーのすぐ下）のチェックボックスをオンにします。
- Enterキーを押します。

[ドキュメントグリッド] は以下のように見えるはずです:

ドキュメントグリッド

ビュー ◀ ▶ ◀ ▶ 2 of 223 | エクスポート... 検索: A

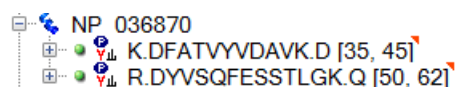
	プリカーサー	ペプチド	繰り返し測定	切断数	MissingData
	<a href="#">614.3164++</a>	<a href="#">DFATVYVDAVK</a>	<a href="#">D_108_REP2</a>	7	<input checked="" type="checkbox"/>
▶	<a href="#">614.3164++</a>	<a href="#">DFATVYVDAVK</a>	<a href="#">D_196_REP3</a>	7	<input checked="" type="checkbox"/>
	<a href="#">614.3164++</a>	<a href="#">DFATVYVDAVK</a>	<a href="#">H_147_REP2</a>	7	<input checked="" type="checkbox"/>
	<a href="#">614.3164++</a>	<a href="#">DFATVYVDAVK</a>	<a href="#">H_148_REP2</a>	7	<input checked="" type="checkbox"/>
	<a href="#">614.3164++</a>	<a href="#">DFATVYVDAVK</a>	<a href="#">H_159_REP1</a>	7	<input checked="" type="checkbox"/>
	<a href="#">614.3164++</a>	<a href="#">DFATVYVDAVK</a>	<a href="#">H_159_REP3</a>	7	<input checked="" type="checkbox"/>
	<a href="#">614.3164++</a>	<a href="#">DFATVYVDAVK</a>	<a href="#">H_160_REP2</a>	7	<input checked="" type="checkbox"/>
	<a href="#">614.3164++</a>	<a href="#">DFATVYVDAVK</a>	<a href="#">H_162_REP3</a>	7	<input checked="" type="checkbox"/>
	<a href="#">730.8488++</a>	<a href="#">DYVSQFESSTL...</a>	<a href="#">D_196_REP2</a>	7	<input type="checkbox"/>
	<a href="#">730.8488++</a>	<a href="#">DYVSQFESSTL...</a>	<a href="#">D_196_REP3</a>	7	<input type="checkbox"/>
	<a href="#">730.8488++</a>	<a href="#">DYVSQFESSTL...</a>	<a href="#">H_147_REP2</a>	7	<input type="checkbox"/>

また以下の操作を行うことで、Excel および [ドキュメントグリッド] の値貼り付け機能を利用して 163 行全てを一度に貼り付けることが可能です:

- Excel の空白のスプレッドシート内で、163 番目の行までスクロールダウンします。
- 163 番目の行の最初の列に「TRUE」と入力します。
- 今入力を行ったセルを選択します。
- 今選択されているセルの右下角をクリック & ドラッグして、163 行から 1 までの最初の列のセルに値「TRUE」と入力する。
- Ctrl-C を押してこれらのセルをコピーします。
- Skyline に戻り、「MissingData」列内の最初セルを選択します。
- Ctrl-V を押して、「MissingData」からの 163 行すべてに「TRUE」値を貼り付けます。

この手順は少し冗長であるとお気づきになったかもしれません。各ペプチドについて、MissingData 注釈を TRUE に設定しなければならないのは 1 回だけです。しかし上記を行うと、切断ピークを持つ 31 個のペプチドのチェックボックスをそれぞれクリックするよりも、求めていた結果を迅速に得られます。

上記手順を実行したら、[ターゲット] リスト上位の多くのペプチドの名前の上/右側に、小さな赤い三角形マークが付いているのに気付くでしょう。



これらの三角形の上にマウスカーソルを持っていくと、テキスト「Missing Data True」とともにヒントが表示されます。

ここで以下の操作を行って、切断ピークを積分せずに完全に削除した全てのペプチドについてこの処理を繰り返します:

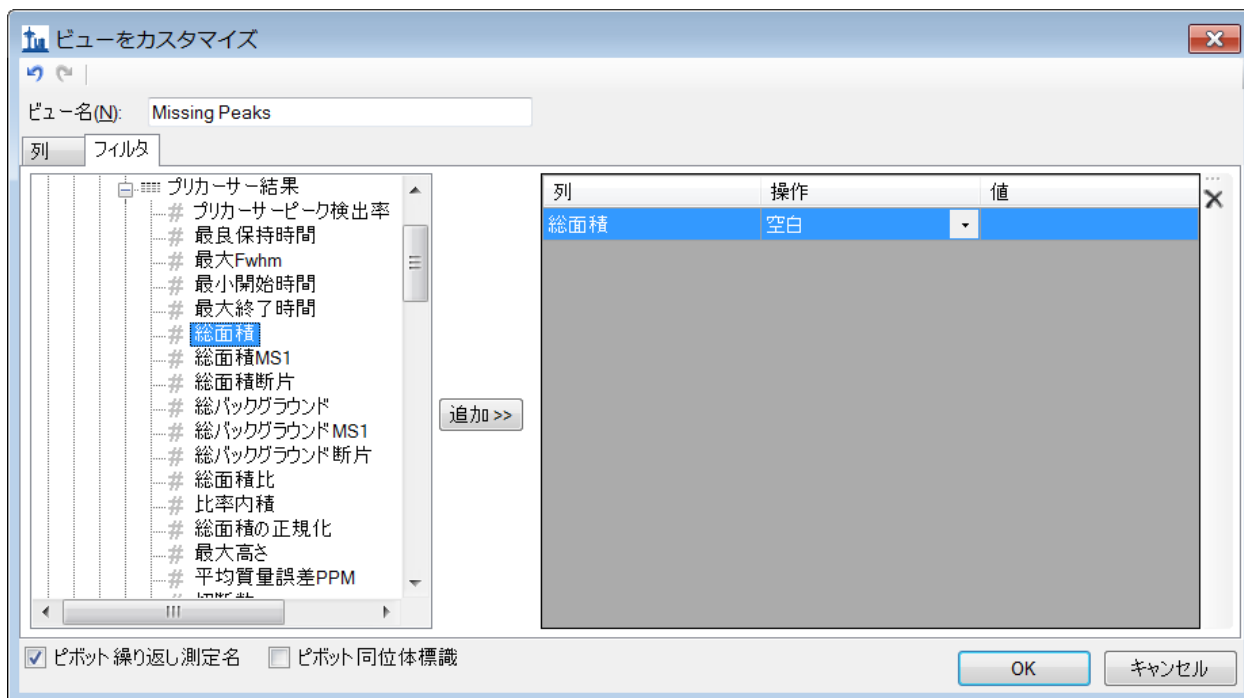
- [ドキュメントグリッド] の [ビュー] メニューで、[ビューを管理] をクリックします。
- 使用してきた「切断プリカーサー」ビューを選択します。
- [コピー] ボタンをクリックします。
- [ビューをカスタマイズ] フォームの [ビュー名] フィールドに、「Missing Peaks」と入力します。
- 列リストの右側の X ボタンを使用して、[プリカーサー] 列および [切断数] 列を削除します。
- [ピボット繰り返し測定名] チェックボックスをオンにします。

[ビューをカスタマイズ] フォームは次のようになります:



- [フィルタ] タブをクリックします。
- X ボタンをクリックして、[切断数] の既存のフィルターを削除します。
- 列ツリーを展開して、タンパク質 > ペプチド > プリカーサー > プリカーサー結果とチェックして、[総面積] 列のフィルタを追加します。
- [操作] 列で、「空白」を選択します。

[ビューをカスタマイズ] フォームは次のようになります:



- [OK] ボタンをクリックします。

- [ビューを管理] フォームの [OK] ボタンをクリックします。
- [ドキュメントグリッド] ビューの [ビュー] メニューで、[Missing Peaks (欠損ピーク)] をクリックします。

[ドキュメントグリッド] は以下のように見えるはずですが。

ペプチド	MissingData	D_102_REP1 繰り返し測定	D_102_REP3 繰り返し測定	D_103_REP1 繰り返し測定	D_103_REP3 繰り返し測定	D_108_REP2 繰り返し測定	D_108_REP3 繰り返し測定
GSYNLQDL...	<input checked="" type="checkbox"/>				D_103_REP3	D_108_REP2	
DYVSQFES...	<input checked="" type="checkbox"/>				D_103_REP3	D_108_REP2	
TGTNLMDF...	<input checked="" type="checkbox"/>					D_108_REP2	
GTITSIAAL...	<input checked="" type="checkbox"/>				D_103_REP3		
IFPENNIK	<input checked="" type="checkbox"/>				D_103_REP3		
ALIHCLHMS	<input checked="" type="checkbox"/>						D_108_REP2
AGDQILAIN...	<input checked="" type="checkbox"/>						
FAEDHFAH...	<input checked="" type="checkbox"/>						
TLNSINIAV...	<input type="checkbox"/>				D_103_REP3		
IFSQQADLSR	<input type="checkbox"/>	D_102_REP1	D_102_REP3	D_103_REP1			

注: ドキュメントを完全に処理しなかった場合、ペプチド GSYNLQDL... には 1 ライン [IM3] しか表示されない可能性があります。完全に処理されたドキュメントを開く方法は、次の注記で説明いたします。

このビューでは、任意のランに欠損ピークがあるペプチドの簡易リストが得られます。ペプチドが 10 個あることが容易に分かります。また、切断ピークを持たないのが 2 つのみで、8 つはすでに「MissingData」注釈セットを持っています。欠損ピークを持つ繰り返し測定名は、[ペプチド] 列および [MissingData] 列の右側に表示されます。すべてのペプチドに欠損データを標識付けるには:

- TLNSINIAVFSK および IFSQQADLSR の [MissingData] チェックボックスをオンにします。
- [ドキュメントグリッド] を閉じます。

おめでとうございます！このデータセットでの初期データ処理が完了しました。取得データについて、データ品質やピーク積分の最適化に影響を与え得る問題を最大限に理解できたら、データ品質をより評価するための更に高いレベルの統計を実行する準備が整っているといます。おそらく、どのペプチドまたはタンパク質がバイオマーカーとして有用なのかについて理解し始めていると思います。

注: チュートリアルこの時点でドキュメントをまだ完全に処理していない、あるいはご自身の処理とチュートリアル著者の処理を比較したい場合は、以下のようにして、本チュートリアルに含まれている完全に処理されたファイルを開いてください:

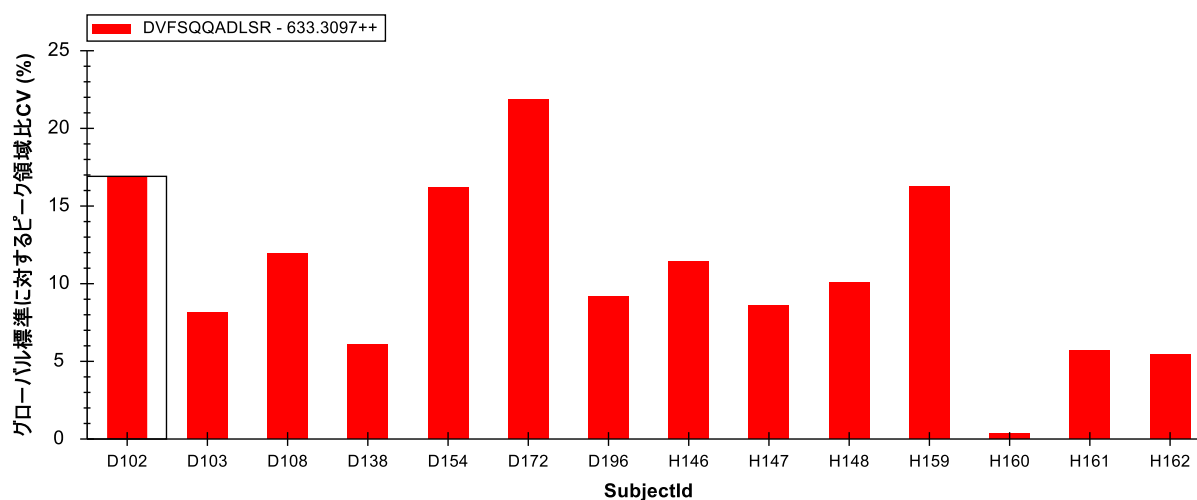
- [ファイル]メニューで[保存] (Ctrl-S) をクリックしてドキュメントを保存します。
- [ファイル]メニューで[開く] (Ctrl-O) をクリックします。
- 1つ上のフォルダ「GroupedStudies1\Heart Failure」に移動して、「Rat\_plasma.sky.zip」を選択します。
- [開く] ボタンをクリックします。

### Skyline を使った最初の複数繰り返し測定検査

このような任意データセットの統計分析を Skyline で現在行える以上に更に掘り下げて行うことはもちろんできますが、変動やグループ平均の初期検査を行う有用な方法が Skyline には用意されています。各被験体の技術的繰り返し測定の間の変動について情報を得るには、以下を行います:

- グローバル内部標準ペプチドリストのすぐ上にある、ペプチド DVFSQQADLSR を選択します。
- [ビュー]メニューで、[トランジション]を選択して[合計]をクリックします。
- [ピーク領域]ビューを右クリックし、[正規化]を選択して[グローバル標準]をクリックします。
- [ピーク領域]ビューを右クリックし、[グループ化]を選択して[SubjectId]をクリックします。
- [ピーク領域]ビューを右クリックして、チェックを入れていない場合は[CV 値]をクリックします。

これにより、[ピーク領域] グラフは以下ようになります:

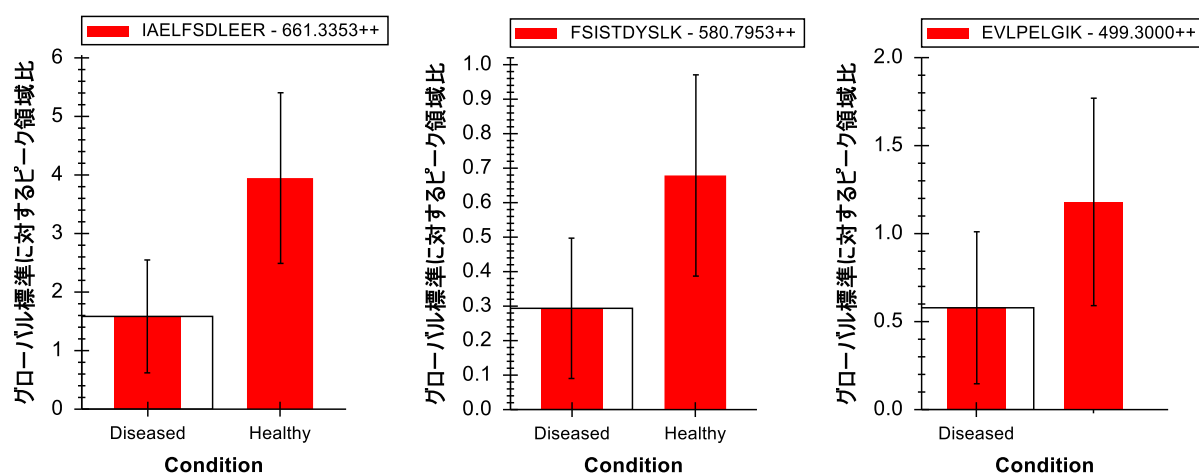


ここでさらに多くのペプチドを選択して、各被験体の技術的繰り返し測定間での変動係数 (coefficients of variation, CV) を再確認できます。ほとんどが 25%未満であることが分かります。しかし理想的には、本実験よりも低い CV が見られることが好ましいでしょう。3 個の測定値から得られた低い CV が、10 個の測定値から得られた同一 CV よりふさわしいということは、注目に値します。これは、小さい試料サイズでの標準偏差を過小評価している統計的傾向によるものです。

最後に、本実験における 2 グループ間のペプチド発現の差異について、Skyline を使用して基本的な情報を得ることができます。健常群/罹患群間の平均ペプチド存在量の差について予備的な確認を行うには、以下の手順を実行します:

- [ピーク領域] ビューを右クリックし、[グループ化] を選択して [Condition] をクリックします。
- [ピーク領域] ビューを右クリックし、[CV 値] をクリックして、このメニュー項目のチェックをオフにします。
- [ターゲット] ビューで様々なペプチドを選択します。

タンパク質 NP\_872280 のペプチドの [ピーク領域] のグラフは、以下のように見えます:



これらのグラフでは、バーが全ての繰り返し測定の平均値を表し（この場合はグローバル標準に対するピーク領域比）、エラーバーは平均値の両側の標準偏差を表します。これにより、平均値からサンプル分布についてある程度分かるようになります。

このグラフを解釈するには、試料群間の差を調べるにおいて自分の目的を理解していることが重要です。最も一般的な目的は以下の 2 つです:

1. 差分: グループ間の差を同定すること。（処理により、ある対象経路に沿って活性が低下するかどうかを調べる、など。）



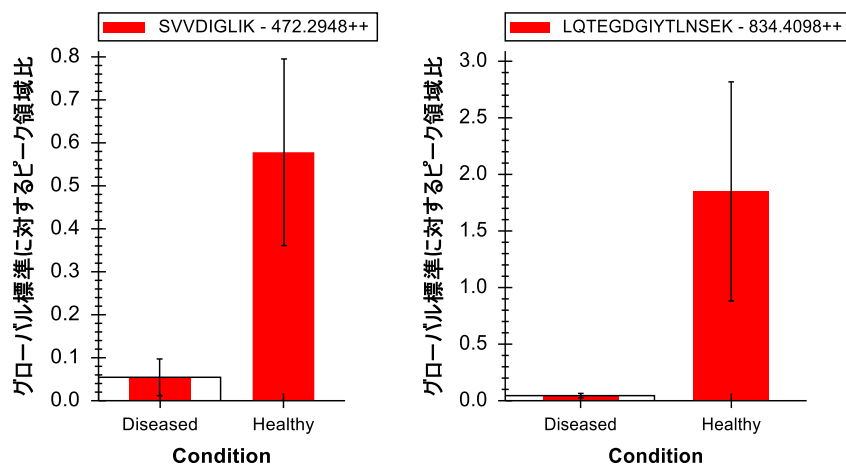
2. 予測: ある特定グループに属すると予測されるペプチドの発現量を同定すること。(ある一連のペプチド群が特定グループのバイオマーカーとして使用可能かどうかを試験する、など。)

単に 2 グループ間での統計学的に優位な差を検出したい場合、分布の標準偏差より平均値の標準誤差に関心を持っていることでしょう。T-検定から得た p 値により、これら 2 つの平均の標準誤差値に関連した平均値の差について統計的有意性を示す数値が 1 つ得られます。上記グラフでは、標準誤差値についての情報はまったく得られません。したがって、差分的発現に関する情報を得るには、これらはあまり有用ではありません。

予測については、2 グループの元となる分布が重要です。ある一つの標準偏差が 2 グループ間で大幅に重複している場合、このペプチドのみでは強力なバイオマーカー候補にならないと思われれます。このケースでは、どの分布によって単一測定値が得られたかを予測することは困難でしょう。ここでは、個別に予測されないペプチドの集団からでも強力な「バイオマーカーパネル」を作成することができる場合があることに、注意することが重要です。

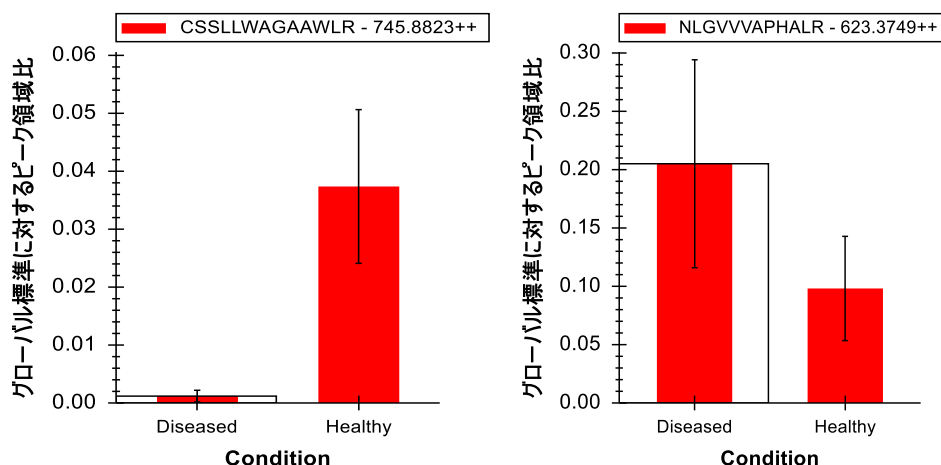
上記グラフ内のペプチドでは、平均値において統計学的有意差がありますが、個々について予測するものではありません。これは、1 つの標準偏差で分布範囲が重複しているグラフとして表されています。

ペプチド/タンパク質のリストを続行すると、すぐ上のタンパク質、NP\_036714 はバイオマーカーとして使用できる非常に強力な候補であることが分かります:



本実験内の多くのペプチドで、2 グループの各平均値の間に統計学的有意差が示されています。このターゲットリストの元が心臓疾患でのタンパク質発現の変化に関する論文であったことを考えてみれば、これは驚くことではありません。しかし、自力では強力なバイオマーカー候補と目されるものは、かなり少数です。

ペプチド **CSSLWAGAAWLR** を含むタンパク質 NP\_001007697 に到達すると、単一タンパク質に割り当てられているペプチドが健常群と罹患群間において相対発現レベルに大きな違いがあるケースが見られます:



このことから、多くの試料での同一ペプチド分子の測定についてある程度の信頼性が得られる一方で、複数のペプチドプリカーサーを元の同一タンパク質へと割り当てる事については信頼性はかなり低いことが多い、ということが思い出されるはずです。この特性が見られる理由は多数あります。例えば、ペプチドは翻訳後修飾 (post-translational modification, PTM) を持ち、それ自身が重要なバイオマーカー候補となる場合があります。その一方で、残りのタンパク質は影響を受けない可能性があります。

### Skyline でのシンプルなグループ比較

Skyline 3.1 では現在、ペプチド又はタンパク質ピーク領域の簡易なペアワイズグループ比較を行うことが可能です。ペプチド又はタンパク質のトランジションピーク領域を合計し、場合によっては正規化標準で分割し、ログを取り、技術的繰り返し測定を平均化し、結果得られた値について T-検定を実行することで、この比較を行う事ができます。Skyline は標識なしデータ内の値が欠けている繰り返し測定又は切断ピークを自動的に破棄します。

現在処理中のデータセットでこれを試すには、以下の手順を実行します:

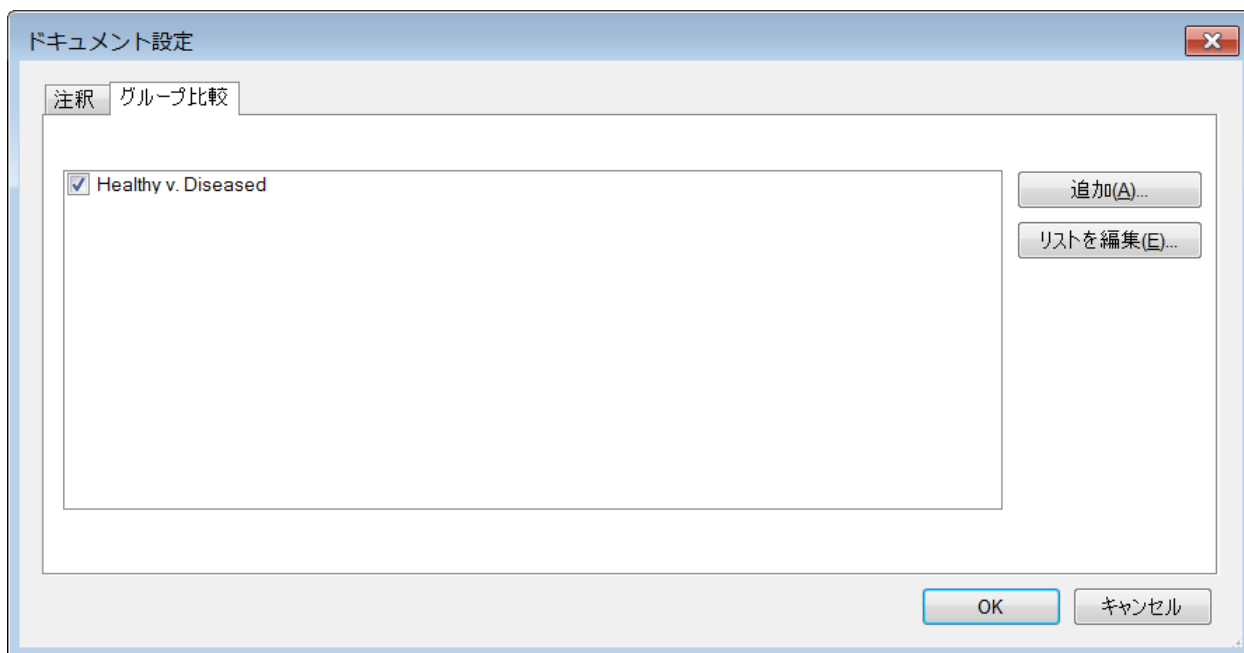
- [設定] メニューで [ドキュメント設定] をクリックします。
- [ドキュメント設定] フォームの [グループ比較] タブ内の [追加] ボタンをクリックします。
- [グループ比較を編集] フォームの [名前] フィールドに「Healthy v. Diseased」と入力します。
- [グループ注釈をコントロール] で「Condition」を選択します。
- [グループ値をコントロール] で「Healthy」を選択します。
- [次に対して比較する値] で「Diseased」を選択します。
- [同一注釈] で「SubjectId」を選択します。

- [正規化メソッド]で「グローバル標準に対する比率」を選択します。
- [信頼度レベル]に「99」%と入力します。
- [適用範囲]で「タンパク質」を選択します。

[グループ比較を編集] フォームは以下のように見えるはずですが:

- [OK] ボタンをクリックします。

[ドキュメント設定] フォームは以下のように見えるはずですが:



- [OK] ボタンをクリックします。

先ほど定義したグループ比較を検査するには、以下を行います：

- [ビュー] メニューで、[グループ比較] を選択して **Health v. Diseased** をクリックします。Skyline により、以下のようなグリッドビューが表示されます：

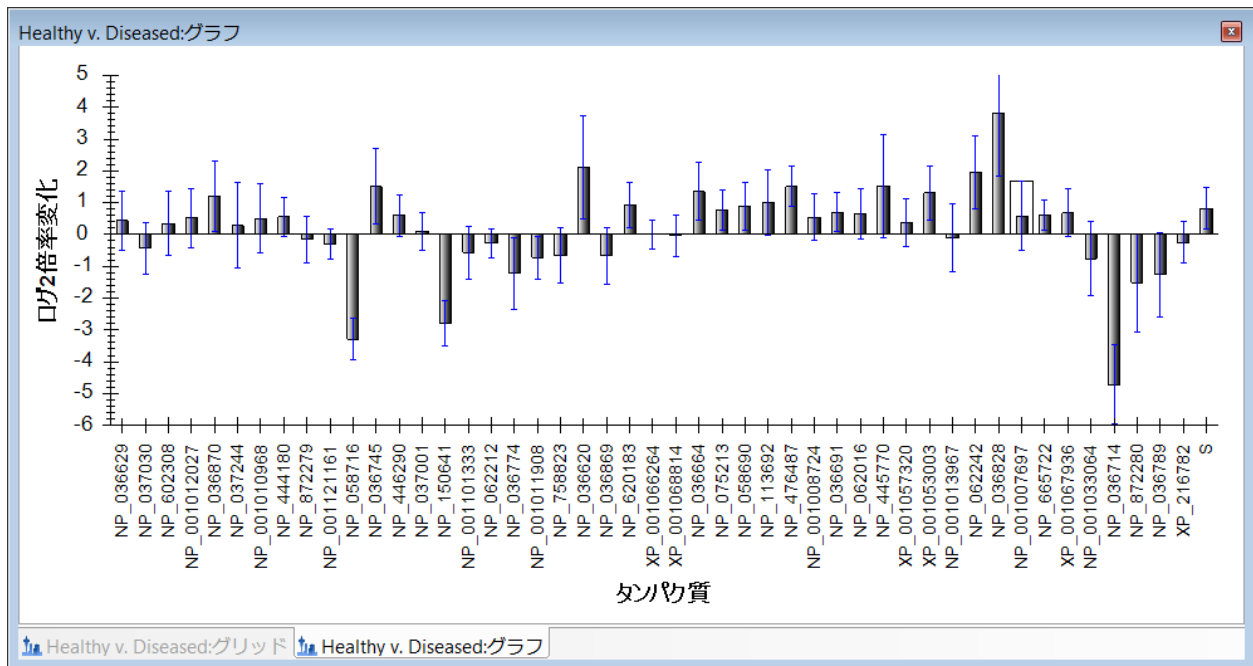
タンパク質	倍率変化結果	調整された P-値
<a href="#">NP_036629</a>	1.35 (99% CI:0.72 to 2.55)	0.2057
<a href="#">NP_037030</a>	0.74 (99% CI:0.42 to 1.3)	0.1685
<a href="#">NP_602308</a>	1.26 (99% CI:0.63 to 2.55)	0.3696
<a href="#">NP_001012...</a>	1.44 (99% CI:0.76 to 2.72)	0.1499
<a href="#">NP_036870</a>	2.31 (99% CI:1.08 to 4.95)	0.0163
<a href="#">NP_037244</a>	1.21 (99% CI:0.48 to 3.09)	0.5251
<a href="#">NP_001010...</a>	1.42 (99% CI:0.67 to 3.02)	0.2057

「倍率変化結果」列に信頼区間が表示されていない場合、倍率変化結果と調整 P 値の両ヘッダー間の垂直ラインをダブルクリックします。

ログ変換された倍率変化値と信頼区間グラフを見るには：

- グリッドビューのツールバーの上にあるメニュー内の [グラフを表示] ボタンをクリックします。

Skyline により、グラフタブが **Healthy v. Diseased** ビューに以下のように追加されます:



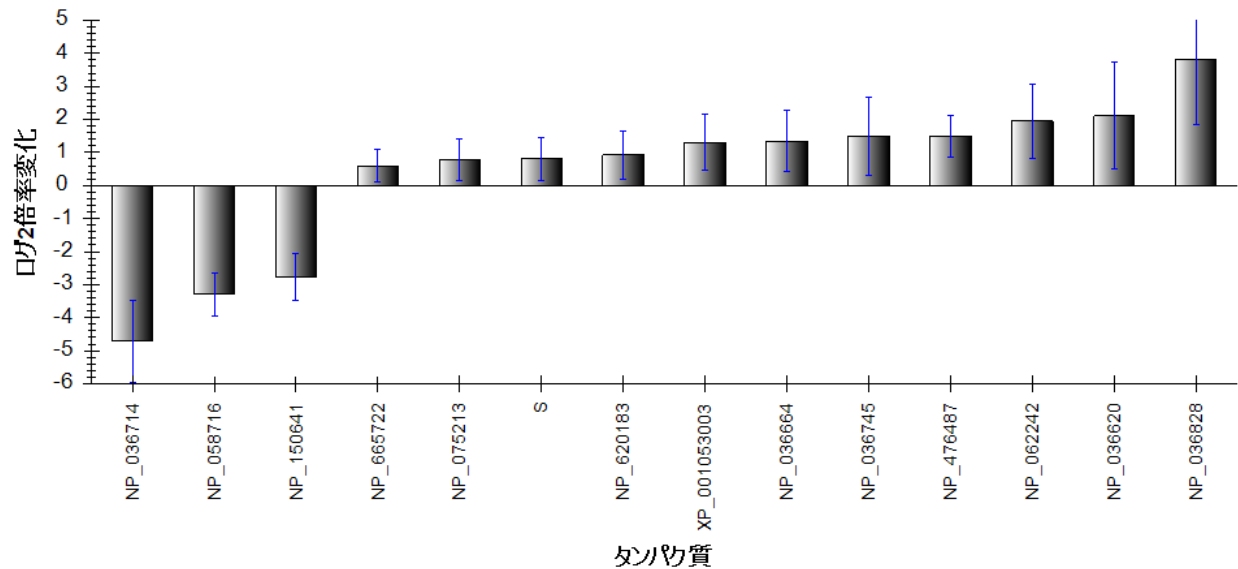
- ビュー下部にある「Healthy v. Diseased: グラフ」タブを青い長方形が見えるまでクリック & ドッグします。
- マウスボタンから指を放し、並列ビューとなるようにグリッドおよびグラフビューを調整します。
- 当該グリッドで「倍率変化結果」ヘッダーをクリックして、現れたメニュー内の [昇順にソート] をクリックします。

これによりグリッドが並べ替えられ、グラフに戻るとそちらも並べ替えられているのが分かります。多くの信頼区間を示すエラーバーがゼロラインを超えていることに留意して下さい。99%信頼度では、この観測データが無作為に生じるのは珍しいことではありません。そしてこれは、複数の仮説検証のために何の修正も行われていません。

不正発見率 (false discovery rate, FDR) を推定する Benjamini-Hochberg 法で調整した p 値を基にカットオフを設定するには、グリッドビューで以下を行います:

- 「調整 P 値」ヘッダーをクリックして、[フィルタ] をクリックします。
- フィルタタイプ [フィルタタイプ] フィールドで、「Is Less Than」を選択します。
- 以下のフィールドに、「0.01」と入力します。
- [OK] ボタンをクリックします。

グリッドツールバードロップで 48~14 の行番号が示されているのが見られ、当該グラフは以下のように見えるはずで:



有意に変化したタンパク質のリストには、先に分析した「S」タンパク質が含まれ、そこには分解ペプチドと正規化ペプチドの 2 つの問題のみが含まれる、ということに注意してください。これにより、試料順序のランダム化に失敗した可能性があるという問題が示唆されます。測定の全 3 サイクルにおいて罹患被験体は健常被験体に先行していたため、正規化の際に考慮しなかった分解が倍率変化を引き起こして罹患群が増加しているように見えてしまうと予測されます。そのため「S」タンパク質と併せて、罹患群で僅かに有意に増加している全てのタンパク質は、このデータセット内ではある程度懐疑の眼を向けるべきです。

### 技術的繰り返し測定を正しく指定する

本実験のように、技術的繰り返し測定を収集する場合、測定を [グループ比較を編集] フォーム内で正しく指定することは統計的推論にとって非常に重要です。これを行わないと、各測定は異なる生物学的主題に由来すると見なされて、標準誤差や信頼区間が誤って絞られ、人為的に p 値が低下してしまいます。

これを実際に見るには以下を行います:

- グリッドビューのツールバーの上にあるメニュー内の [設定を変更] テキストをクリックします。
- [同一注釈] で、一番上の空の要素を選択します。

グリッドビューに 37 行 (及び 37 本のバーが示されたグラフ) が表示されているのが見え、この時の調整された p 値は 0.01 未満です。これは、14 種類の被験体で各 3 回の測定ではなく、42 個の異なる被験体があると見なされて統計が計算されているからです。これは重要な違いであり、お分かりのように、ここで間違いを犯すとあきれるほど楽観的過ぎる統計推論が得られてしまう場合があります。この場合、推定 1% FDR で健常群/罹患群間の平均値で統計学的有意差のある 14 個のタンパク質が、37 個のタンパク質に増加しています。

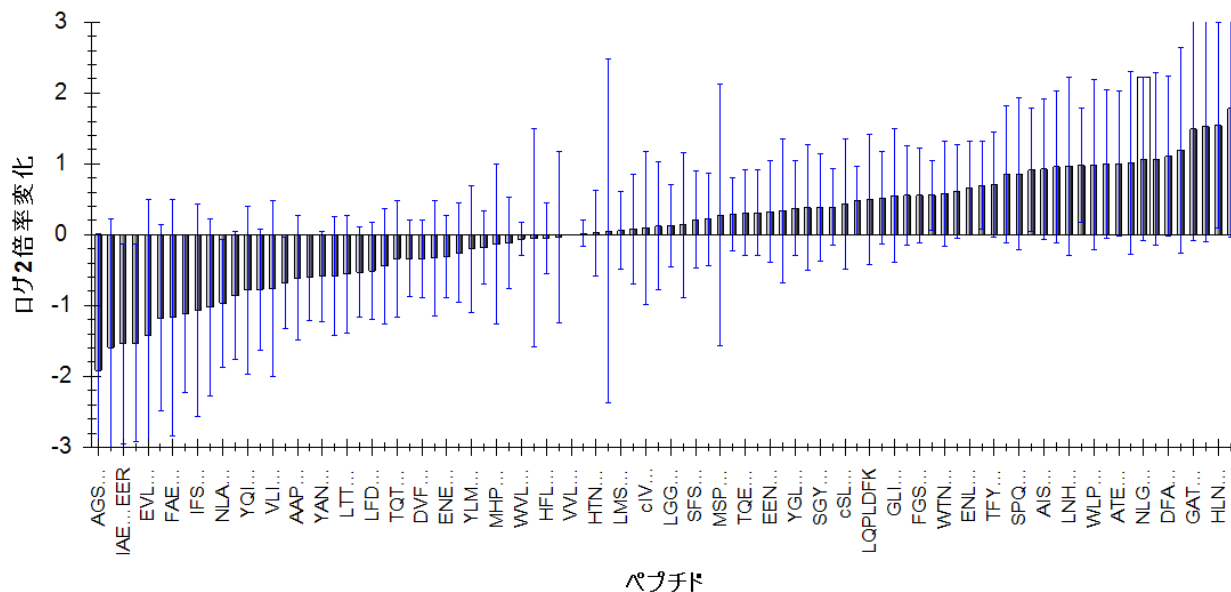
## 差分による最適化

本実験の元の目標に戻って、推定 1% FDR でグループ平均値で統計学的有意差を持つペプチドのある一つのサブセットのみにドキュメントを削減できます。しかしこのデータセットは、ランダム化の欠如（罹患群を常に健常群の前に測定している）及びラン間の体系的なシグナル劣化が原因で、罹患群での増加の程度を過大に表している可能性があるということに留意してください。

それでもなお、本ドキュメントのターゲットリストを、推定 1% FDR でグループ平均値に差のあるペプチドセットに変換する場合は、以下の手順を実行します:

- **Healthy v. Diseased:Settings** フォームで、[同一注釈] フィールドを「SubjectId」に戻します。
- [ファイル] メニューで、[名前を付けて保存] をクリックします。
- 当該ドキュメントを、「Rat\_plasma-diff.sky」の名前で保存します。
- **Healthy v. Diseased:Settings** フォームで、[適用範囲] ボックス内の [ペプチド] オプションを選択します。
- 「調整 P 値」ヘッダーをクリックして、[フィルタ] をクリックします。
- [フィルタータイプ] フィールドで、「Is Greater Than or Equal To (以上)」を選択します。
- [OK] ボタンをクリックします。

**Healthy v. Diseased:グリッド**には、1% FDR カットオフを超えている 92 行が表示されているはずであり、グラフはこのように見えるはずです:



これらのペプチドをドキュメントから消去するには、以下を行います:

- ヘッダーと列ヘッダーが交わっている、**Healthy v. Diseased:グリッド**の左上角列にあるグレーの長方形をクリックして、すべてのセルを選択します。
- 「S」タンパク質内に **VVLSGSDATLAYSAFK** ペプチドが見えるまで、グリッドをスクロールダウンします。
- Ctrl キーを押し下げながら行ヘッダーをクリックして、この行を非選択にします。

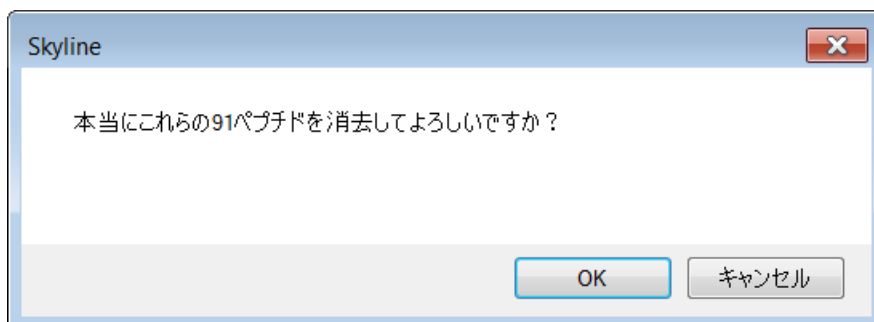
グリッドは以下のように見えるはずですが:

タンパク質	ペプチド	倍率変化結果	調整されたP-値
NP_001011...	DVNEAIQW...	0.92 (99% CI:0.59 to 1.44)	0.6708
NP_872279	WVLTVAHC...	0.96 (99% CI:0.81 to 1.13)	0.4948
NP_058690	AFMDCCNY...	0.97 (99% CI:0.33 to 2.81)	0.9451
XP_001066...	HFLIETGPK	0.97 (99% CI:0.68 to 1.37)	0.8178
NP_001013...	VTSAAFPS...	0.98 (99% CI:0.42 to 2.26)	0.9474
S	VVLSGSDA...	1 (99% CI:1 to 1)	0.9769
NP_036745	ENSSNILD...	1.01 (99% CI:0.89 to 1.15)	0.8135
NP_001121...	HTNNGMIC...	1.02 (99% CI:0.67 to 1.55)	0.936
NP_037244	TGTNLMDF...	1.04 (99% CI:0.19 to 5.59)	0.9474
NP_037244	LMSPEEKP...	1.04 (99% CI:0.71 to 1.53)	0.7959
XP_001068...	SLVIQKPE...	1.05 (99% CI:0.61 to 1.81)	0.8135

VVLSGSDATLAYSAFK ペプチドの倍率変化は 1 であり、99%信頼区間は 1 to 1 であることに注意してください。これは、正規化のグローバル標準として使用されている単一ペプチドです。このペプチドはドキュメントから削除してはいけません。だから非選択にしたのです。

- グリッドツールバー内の **X** ボタンをクリックします。

Skyline に、選択した行を消去してよいかを確認する、以下のメッセージが表示されます:



- [OK] ボタンをクリックします。



ここでグループ比較ウィンドウを閉じ、ドキュメント内の残りの 34 個のペプチドを再確認することが可能です。これらは全て 2 グループ間での平均値に明白な差があるはずですが、これらの分布はわずか 1 つの標準偏差でさえ常にばらばらになるわけではありません。すなわち、条件別にグループ化されたピーク領域を表示している [ピーク領域] グラフ内で、標準偏差を示すエラーバーが重複するのが見られます。

## 結論

本チュートリアルでは、全てのインポート済み LC-MS/MS クロマトグラフィーデータについて Skyline が自動的に実行するピーク積分を視覚的に評価して修正する最も効果的な手法を学習しました。標識なし SRM データセットに対してのみ作業を行いました。本チュートリアルで利用した手法はクロマトグラフィーを使った他の定量法で取得したデータにも同様に利用可能です。この検査と修正は手動で行うため、含まれているペプチドが 2 千個を超えないデータセットに最も適しています。本ケースでは、137 個のペプチドの処理に約 1 時間かかりました。すなわち、2,000 個のペプチドを含むデータセットでは 2 日以上かかることを意味します。

より大規模なデータセットについては、まず初めに、研究対象の条件間においてある差異を示す初期兆候に基づいてペプチドを大まかにスクリーニングするのが良いかもしれません。このケースでは、可能性のある対象ペプチドに対してこの種の検査と修正を手動で行うことだけに時間を費やしています。今ではこの種のフィルタリングを、Skyline のグループ比較サポートを使用して学習した方法により達成できるはずですが、調整 p 値（又は FDR カットオフ）の制限を、本チュートリアルで使用した 1%より厳しさを緩めても良いでしょう。しかしアプローチは同様です。

Skyline が提供するこの種の処理や専門的な使用法が、定量実験における潜在的なエラーの原因を理解して修正する際にどのように役立つかをここまで見てきました。誤って割り当てられたピーク、干渉、ピーク切断、検出困難な量のペプチド、200 倍ものペプチド信号の劣化、および二重溶出ペプチドについても、直に扱ってきました。これらの問題を理解し、修正し、そして注釈するために、Skyline が提供するツールを使いこなすことで、定量的ターゲットプロテオミクス実験でのエラーを削減し、より効率的に生物学的な情報を得ることができるでしょう。

## 参考文献

1. Bereman, M. S., MacLean, B., Tomazela, D. M., Liebler, D. C. & MacCoss, M. J. The development of selected reaction monitoring methods for targeted proteomics via empirical refinement. *PROTEOMICS* **12**, 1134–1141 (2012).
2. Zhang, H. *et al.* Methods for Peptide and Protein Quantitation by Liquid Chromatography-Multiple Reaction Monitoring Mass Spectrometry. *Mol. Cell. Proteomics MCP* **10**, (2011).