

Skyline を使用したグループ化研究データの処理

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生物のタンパク質並びにペプチドの発現解析ができるため、有用な研究手法の 1 つとして広く普及し始めています。元々 Skyline は 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ターゲットプロテオミクスを用いた疾病バイオマーカー候補タンパク質の発見と検証に尽力してきました。

本チュートリアルでは塩分感受性ラットモデルを用いた心不全の症例対照研究の SRM（Selected Reaction Monitoring）データを Skyline で解析します。この研究では過去文献の心不全関連するタンパク質のリストをもとに血漿内タンパク質発現を疾患群と健常群で比較しています。

本実験は LC-MS/MS での測定を計 42 回行い、49 種類のタンパク質由来の 137 種類のペプチドを検出しました。本チュートリアルでデータ処理と解析に一度慣れてしまえば、より多くの実験群でより大規模で多くのターゲットを扱う研究でも比較的簡単にデータを取り扱えるようになります。チュートリアルなしでは小規模なデータセットでさえ、データ解析に戸惑うことは大いにありえます。ここでの知識はよビッグデータセットや、Skyline が使用可能な他の測定手法（PRM、DIA、DDA）にも適用可能です。

本チュートリアルのセクション

- [はじめに](#)
- [メソッド最適化後の比較解析](#)
- [標準ペプチドなしでの繰り返し測定データ処理設定](#)
- [スケジュールモードの SRM 時のピークの誤同定](#)
- 複数の繰り返し測定データ処理 1
- [正規化のための標準合成ペプチドの利用](#)
- [複数の繰り返し測定データ処理 2](#)
- [測定データの統計解析](#)
- ピーク形状が悪いペプチドへの注釈
- [Skyline での複数の繰り返し測定検査](#)
- [Skyline でのグループ間比較解析](#)
- [技術的繰り返し測定の指定](#)
- [群比較のための最適化](#)
- [まとめ](#)

はじめに

チュートリアルを始めるため以下サイトよりの zip ファイルをダウンロードしてください。

<https://skyline.ms/tutorials/GroupedStudies1.zip>

ファイルをコンピュータ内のフォルダに解凍します（以下は Path の例）。

C:\Users\brendanx\Documents

新しいフォルダが作成されます。

C:\Users\brendanx\Documents\GroupedStudies1

この新規フォルダを開くとサブフォルダ見つかります。

- Heart Failure – Michael J. Toth および Michael J. MacCoss のラボより提供された「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でターゲットプロテオミクス研究データの処理方法を学習し、タンパク質／ペプチドターゲットおよび実験データの質についても理解を深めてください。

メソッド最適化後の比較解析

「GroupedStudies1」フォルダ内の「Heart Failure」サブフォルダに格納されているデータは、[ターゲットメソッドの最適化](#)チュートリアルでも記載されています。これはもともとアプローチの可能性や限界を調べるため MacCoss ラボで測定されたものです。このデータはさまざまな理由で論文未発表のままとなっています。Skyline を利用してデータを解析していくとその理由がわかってくると思われます。もちろんこのターゲットメソッドの最適化および関連実験については、このデータセットとは別に論文発表がすでになされています。¹

初期の研究では心不全に関与するとされる 109 のタンパク質が過去文献から選択されました。これらのタンパク質は Skyline ドキュメントにすでに追加されています。そのタンパク質シーケンスの一部は FASTA 形式のテキストからインポートされており、Skyline 上で *in silico* トリプシン消化が実行できます。長さ 6~30 のアミノ酸から構成される（非特異的開裂などがない）ペプチドから Skyline 内にパラメータが設定されますが、その他のタンパク質については、*in silico* 処理は Skyline を使用せずに実行されています。これらタンパク質から得られたペプチドリストは 2,165 のターゲットペプチドを持つドキュメントとして Skyline に加えられます（[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）。 $y_3 \sim y_{(n-1)}$ に対応する電荷+1 のプロダクトイオンに対応する電荷が+2 のプリカーサーペプチドが検討対象となります。初期のメソッドは集められたラット血漿試料をスケジュール化無しの SRM メソッドで測定し得られた 12,194 のトランジションをカバーしています。この分析は 2009 年にわずか一回のみ実行されただけで、151 回の LC-MS/MS 測定（151 回のインジェクション）をしています。最近では質量分析装

置の 1 回の測定で 2,165 のペプチドがプロファイルできることもあり DIA が使用される傾向にあります。

得られた 151 個の生データファイルは Skyline へとインポートし、試料の前処理なしでも（血漿マトリックス存在下でも）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グラジエントを用いた LC 分離で複数のターゲットプロダクト y-イオンの共溶出が確認できたのは 135 のペプチド（49 のタンパク質由来）でした（その内訳は y イオン 3 つ検出できたのは 2 個のペプチド、y イオン 4 つ検出できたのは 5 個のペプチド、y イオン 5 つ検出できたのは 27 個のペプチド、y イオン 6 つ検出できたのは 103 個のペプチド）。残りの 780 のトランジションについては、1 回の分析で定量化できるようスケジュールを設定しました。すなわち 90 分のグラジエント溶出で検出されたピークについて前後 5 分の保持時間ウィンドウを利用します。このスケジュール化メソッドを使って 14 個体の塩分感受性ラット（健常群が 7 個体。疾患群が塩分過多の食事による心不全 7 個体）の血漿試料を各試料について 3 回繰り返し（計 42 回の試料注入）で測定しました。本チュートリアルでの研究はもともとバイオマーカー研究で 135 種類の検出可能なペプチド中 2 グループ間（健常群と疾患群）で異なる、バイオマーカーの候補の探索となります。

本研究の目標はターゲットタンパク質／ペプチドの予備知識があまりなくかつ、安定同位体標識したペプチドを使用せずともターゲットプロテオミクスで有益な解析結果が得られるかどうかを判定することです。

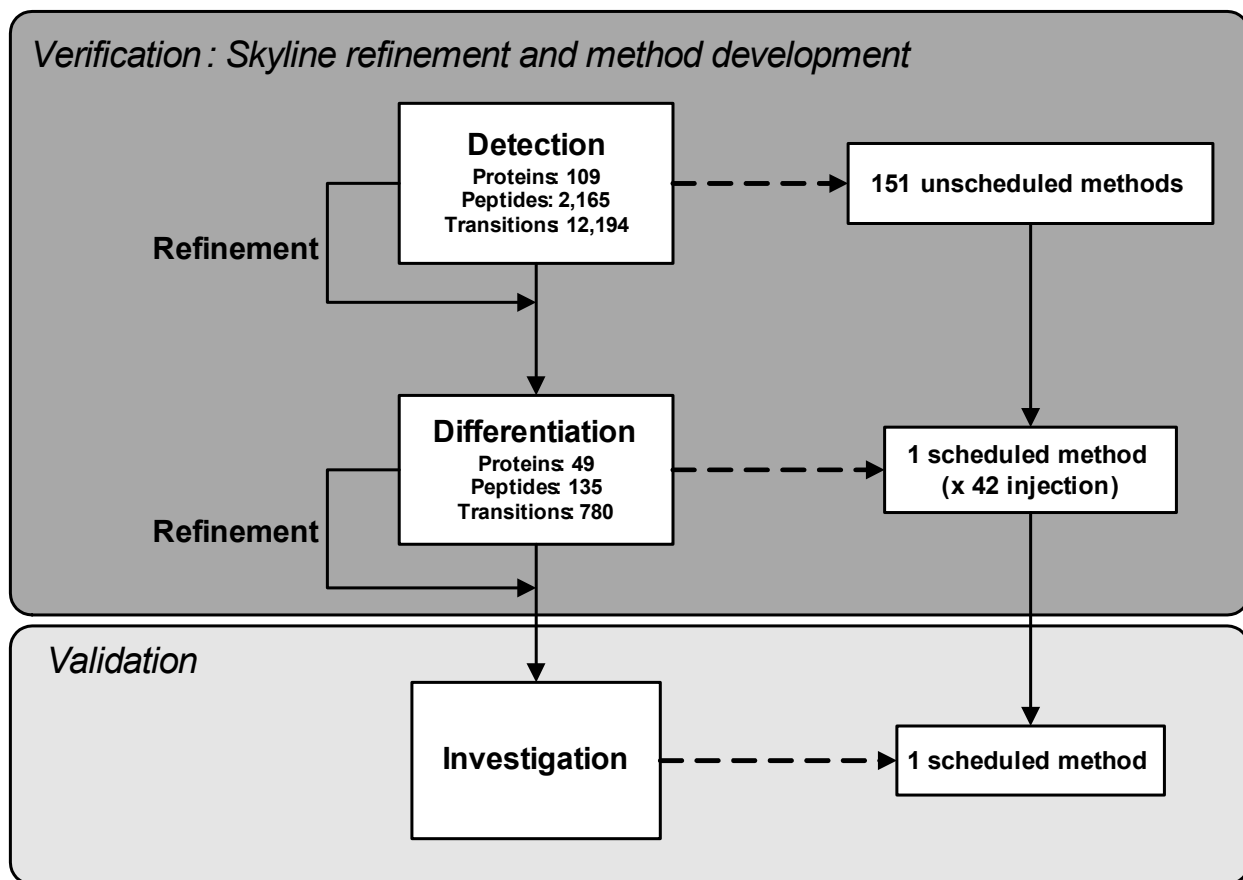


図 1 : 塩分感受性ラットモデル心不全研究のターゲットメソッド最適化における「検出」および「比較」のフェーズを示す概略図。

このメソッドの最適化研究の健常-疾患比較フェーズ向けに収集されたデータの処理を開始するため以下の操作を行います。

- 「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、789 のトランジションが含まれていることが分かります。

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

これは上記の図（未発表データ）で示されたものとは異なります。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（リストの下の方で「S」と名付けられているもの）。このリストには測定間の差を抑えるため正規化のための標準ペプチドが 3 つ含まれています。これについては後で詳しく触れます。またペプチドが 1 つ（タンパク質としても 1 つ）、上記で説明したファイルからなくなっています。

一部のペプチドはライブラリスペクトルと一致するものがあることに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となります。1 つは NIST 由来。そしてもう 1 つは Global Proteome Machine (GPM) 由来のもので、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]メニューの[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「ライブラリ名」列を見つけ、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)を確認できます。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。選択したセルの行番号がグリッドの上にあるツールバー内に表示されます。



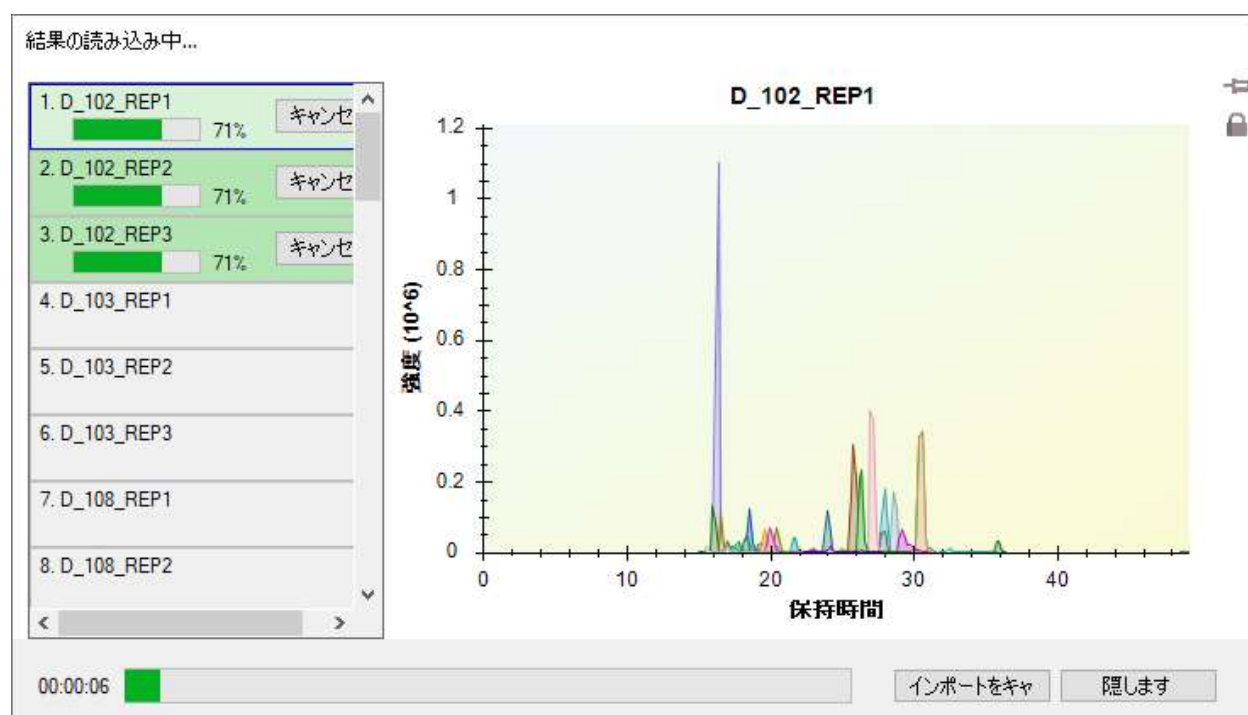
現在の例からこの実験に基づく予備知識が得られます。タンパク解析技術は進歩を続けています。実際 Skyline にも現在統合されておる Prosit というスペクトル予測ツールを利用すればターゲットペプチドの完全網羅も可能となってきています。

標準ペプチドなしの繰り返し測定データ処理設定

14 個のサンプルの繰り返し測定を 3 回行い取得した SRM データの処理を始めるため以下の操作を行います。

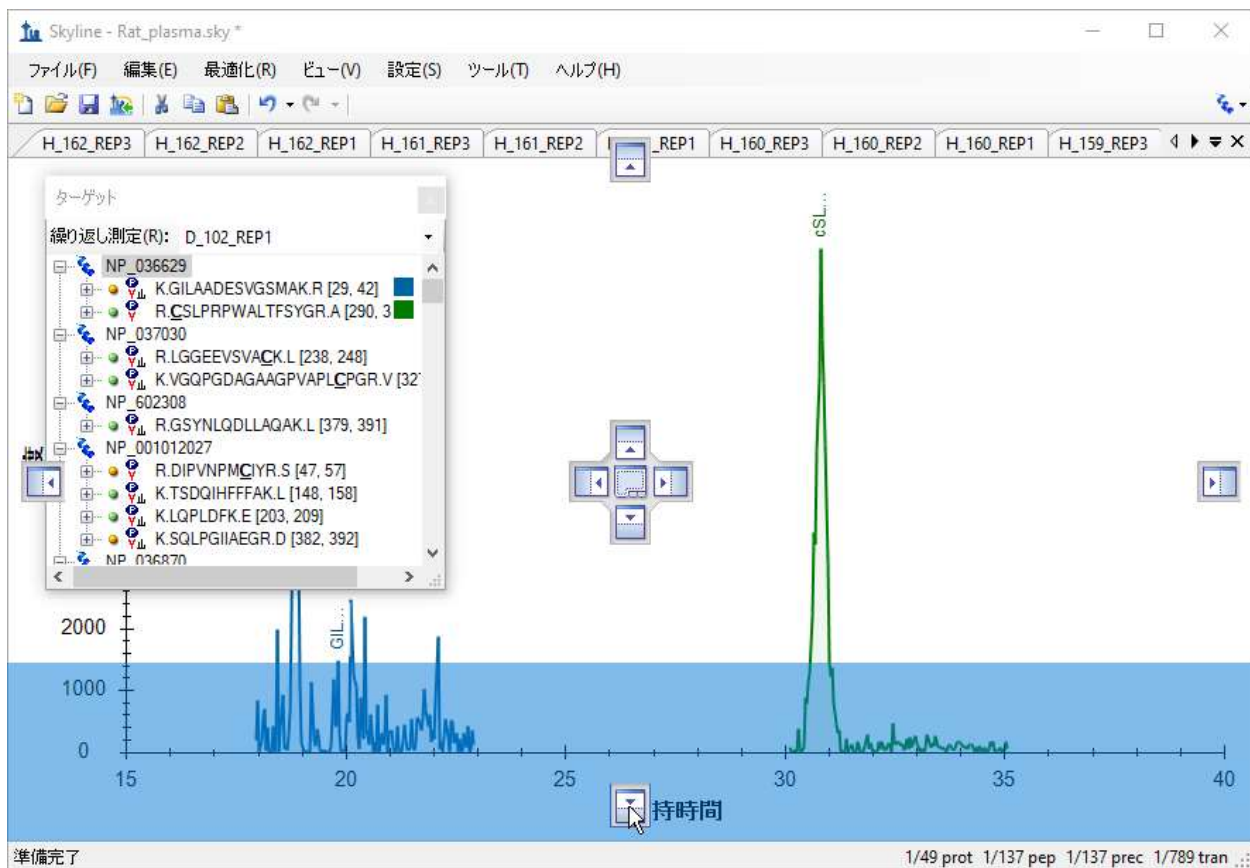
- [ドキュメントグリッド]フォームを閉じます。
- [ファイル]メニューで、[インポート]を選択して[結果]をクリックします。
- [結果をインポート]フォームの[OK]ボタンをクリックして、3 回の繰り返し測定の 1 回目の繰り返し測定をインポートします。
- [結果ファイルをインポート]フォーム内の「D_102_REP1.raw」というデータファイルをクリックします。
- Shift キーを長押ししながら 3 回繰り返しの最後にあたるデータファイル「H_162_REP3.raw」をクリックします。これで「GroupedStudies1\Heart Failure\raw」フォルダのすべてのデータファイル (1-3 回目までの繰り返し測定すべて) が選択されます。
- [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 はウィンドウには進行状況が以下のように表示されます。

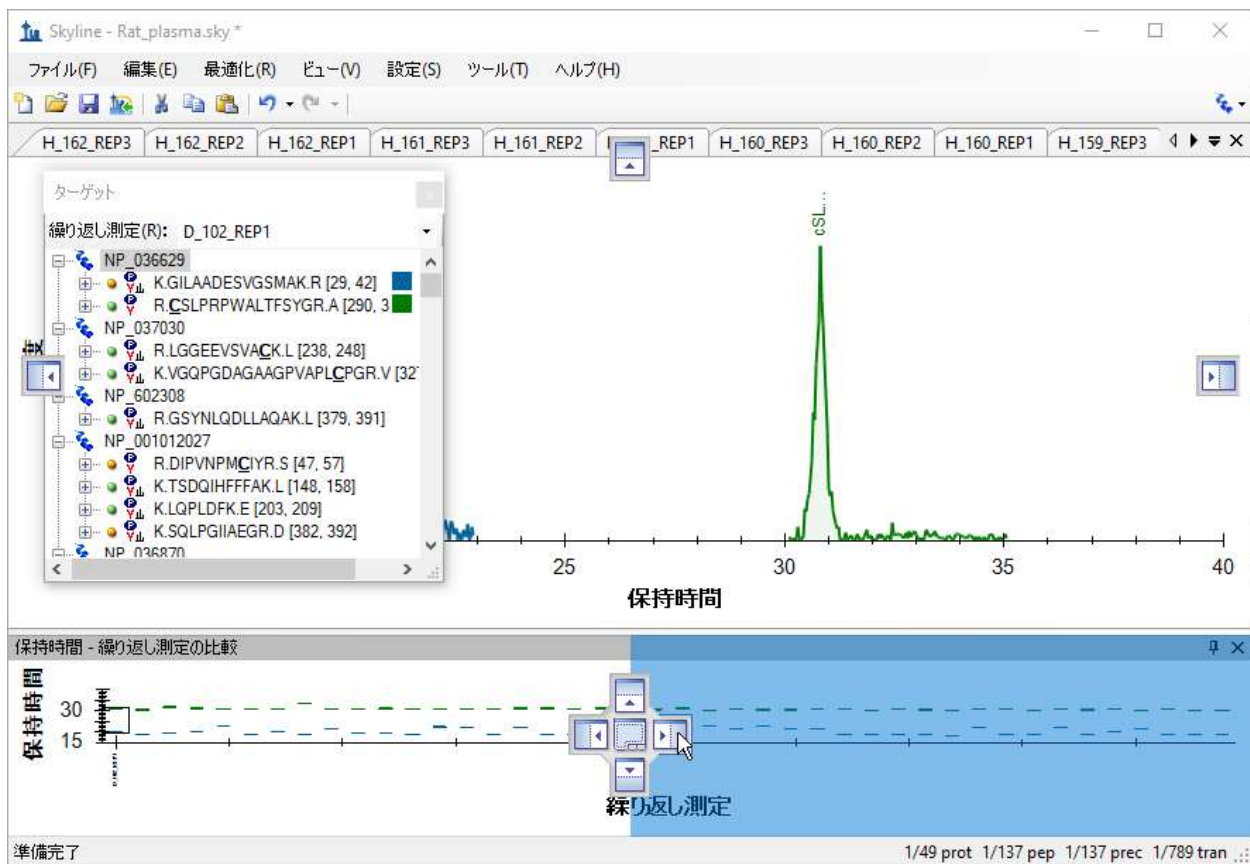


Skyline で SRM クロマトグラムをインポートしている間にデータ処理の準備を続行するため以下の操作を行います。

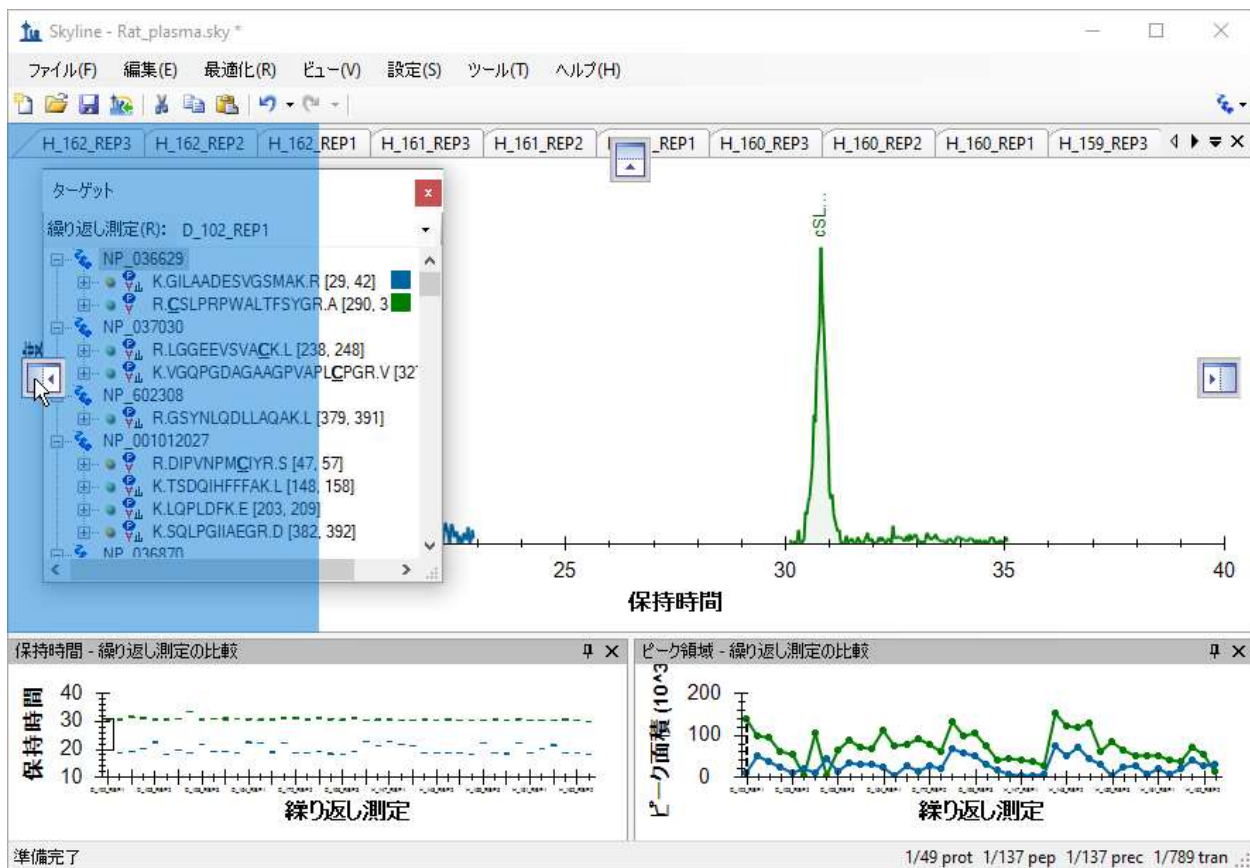
- [詳細を隠す] ボタンをクリックします。
- [ライブラリー一致] ビューを閉じます。
- [ターゲット] ビューのタイトルをクリック & ドラッグしてマウスカーソルを追跡する青い長方形が見えるようにします。
- マウスの左ボタンから指を離し、[ターゲット] ビューを Skyline ウィンドウ上に表示されたままにします。
- [ビュー] メニューで、[保持時間] を選択して [繰り返し測定の比較] をクリックします (F8)。
- [保持時間] ビューのタイトルを、ドッキングインジケータが見えるまでクリック & ドラッグします。
- 以下のように、青い長方形の枠が Skyline ウィンドウの下端を覆うまでカーソルを下部ドッキングインジケータへドラッグします。



- マウスの左ボタンから指を離し[保持時間]ビューをドッキングします。
- [ビュー]メニューから[ピーク領域]を選択して[繰り返し測定の比較]をクリックします (F7)。
- [ピーク領域]ビューをマウスカーソルがドッキングインジケータ内に入るまでドラッグすると、青い長方形枠が表示されて[ピーク領域]ビューの下端が共有されます。



- マウスの左ボタンから指を離し[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、マウスカーソルがドッキングインジケータ内に入るまでドラッグすると青い長方形枠が表示され Skyline ウィンドウの左端とドッキングされます。



- メインの Skyline ウィンドウの外側で表示されている他のウィンドウを閉じます。
- [ビュー]メニューから[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]フォームの[枠のグループ化]内に「3」と入力します。
- [グループ間の分散グラフ]オプションをクリックします。
- [並べ順]ドロップダウンリストから「ドキュメント」を選択します。

フォームは以下ようになります。

グループ化されたグラフを配置する ✕

枠のグループ化(G): OK

キャンセル

グループにグラフを分割する(S)

グループ間の分散グラフ(D)

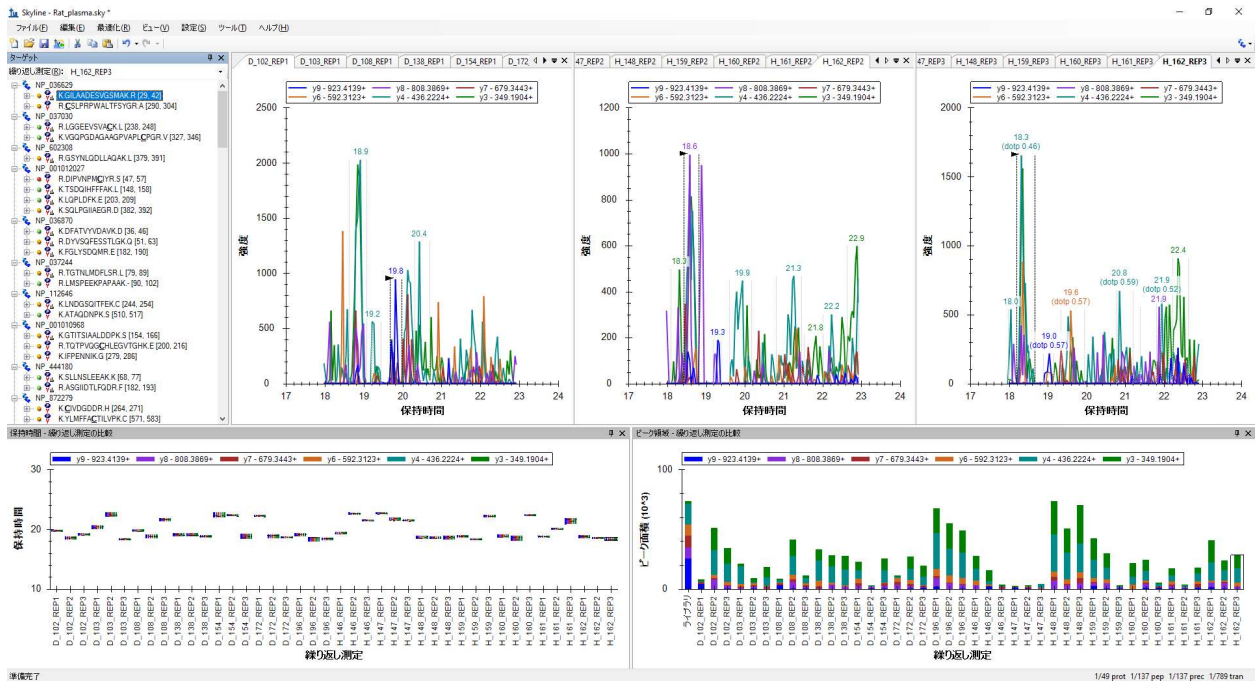
表示(S): タイル

並べ順(O): ドキュメント

反転(B)

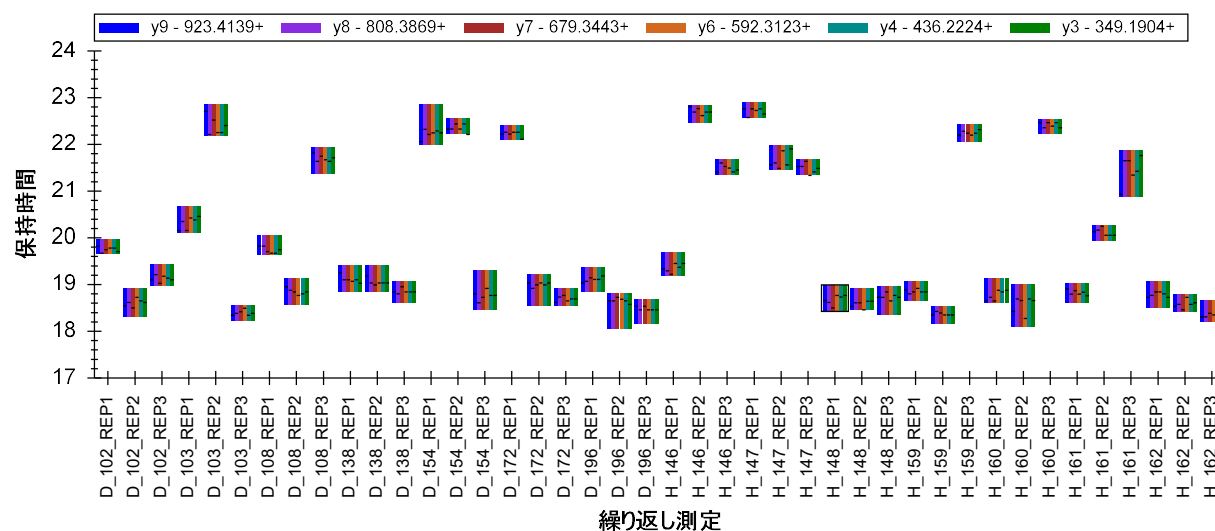
- [OK] ボタンをクリックします。
- Skyline メインウィンドウを最大化します。
- ドキュメント内の最初のペプチドを選択します。

Skyline ウィンドウは以下ようになります。



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。そのため本チュートリアルをデジタルでご覧になっている場合は 200%以上に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は元通り、8½ x 11 インチのページレイアウト用の解像度の画像となります。ほとんどの Skyline ワークショップは 1024x768 のスクリーン解像度なので、より大きいモニタを使用してより Skyline の利用を見やすくするのもよいでしょう。

この測定での最初のペプチド (K.GILAADESVGSMK.R) [28, 41]の [保持時間] ビューを見るとピークがあまりそろっ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 分で溶出されていますが、3分の1ほど (12) は 22 分付近で溶出しています。

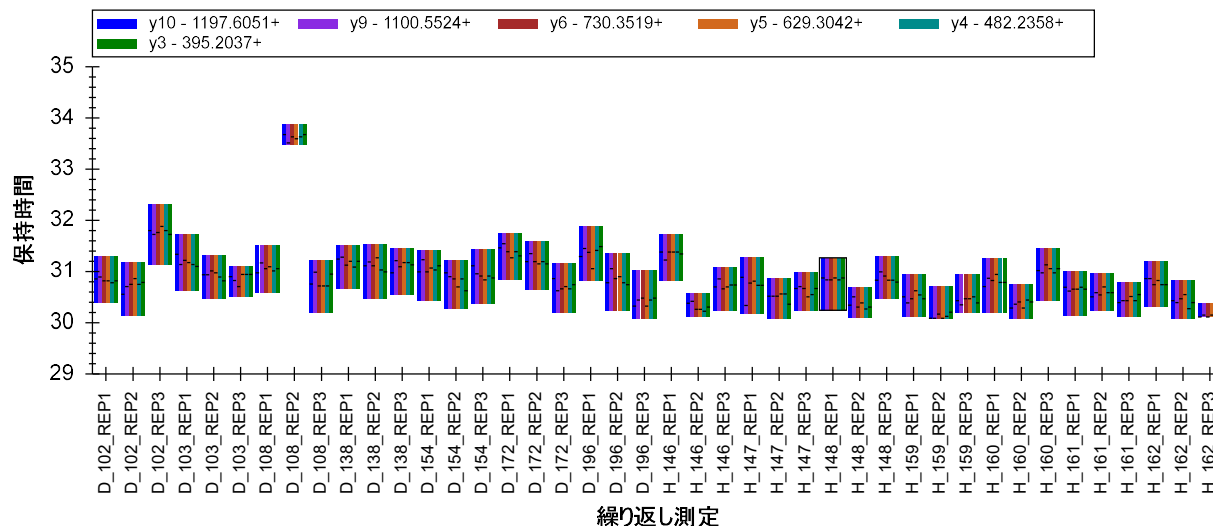


確かに 19 分前後のピークはより良いピークに見えますが、全体としてペプチドのピーク積分結果はあまりそろっていません。これでは単一ペプチドの結果と結論づけることが難しいです。しかしいくらピーク積分条件を調整してもすべての測定データがそろい単一ペプチドと見なせるような可能性は低いです。調整は行わうのは時間の無駄となるので次の操作を行います。

- 最初のペプチドを削除します。

・スケジュールモード SRM 時のピークの誤同定

2 番目のペプチドの保持時間プロットを見ると同ペプチドの溶出時間は 1 番目と比較してよりそろっています。



しかし、すべてのトランジションのピークトップ（各測定のパール上では水平ライン表示）はやはりずれが見られます。そこでクロマトグラムデータを見てみると強度が中～低のピークがありこれらはピーク形状もギザギザとなっています。

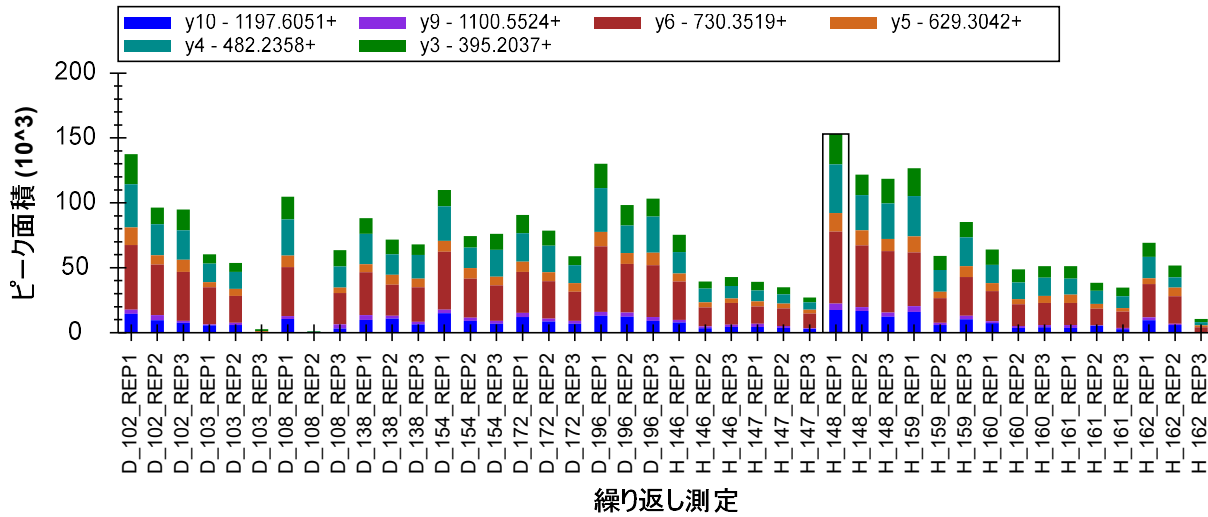
疾患と健常群を分けて比較すべく疾患群の繰り返し測定（D_）を左側健常群の繰り返し測定（H_）を右側に表示した場合は以下の操作を行います。

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

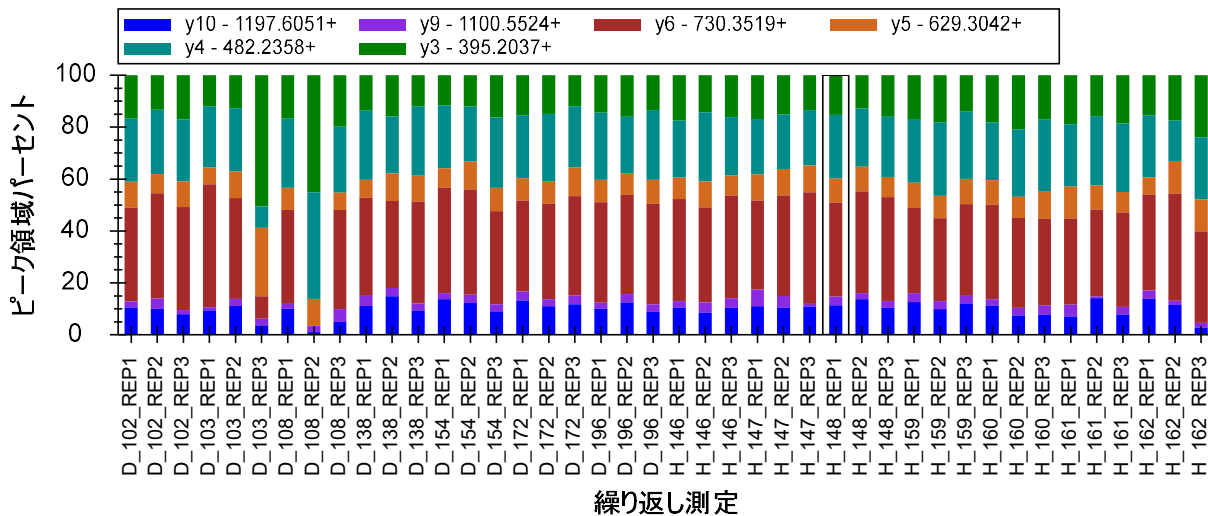
続行する前に表示中プロットの最終調整を行います。

- **[ピーク領域]** ビューで右クリックして **[正規化]** を選択したら **[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 (ピーク強度) から



次のプロット (ピーク面積%) へと変わります。

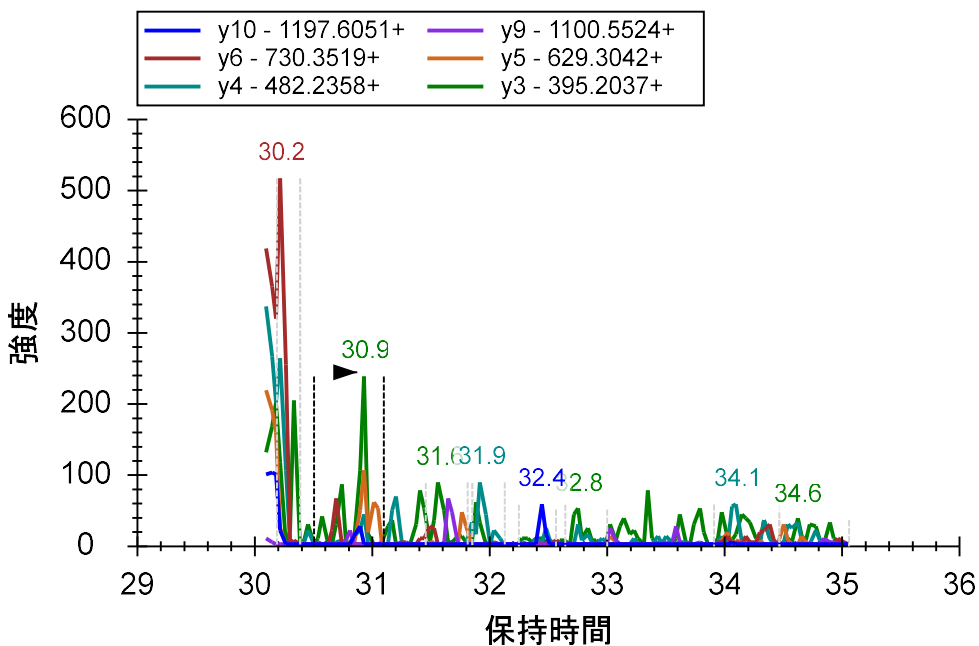


ピーク面積の%表示にすることで測定間のイオンの相対存在量を一目で比較できかつ、[保持時間] ビューからすでに取得している情報に追加されます。この測定間の比較では D_103_REP3 と D_108_REP2 が例えばは他の測定とはかなり違ったプロファイルになっていること、H_162_REP3 についても他と若干のプロファイルの違いが見えてきます。上記で[保持時間] ビューだけを使用したときは、D_108_REP2 のみが他との違いが顕著だったことには注意してください。

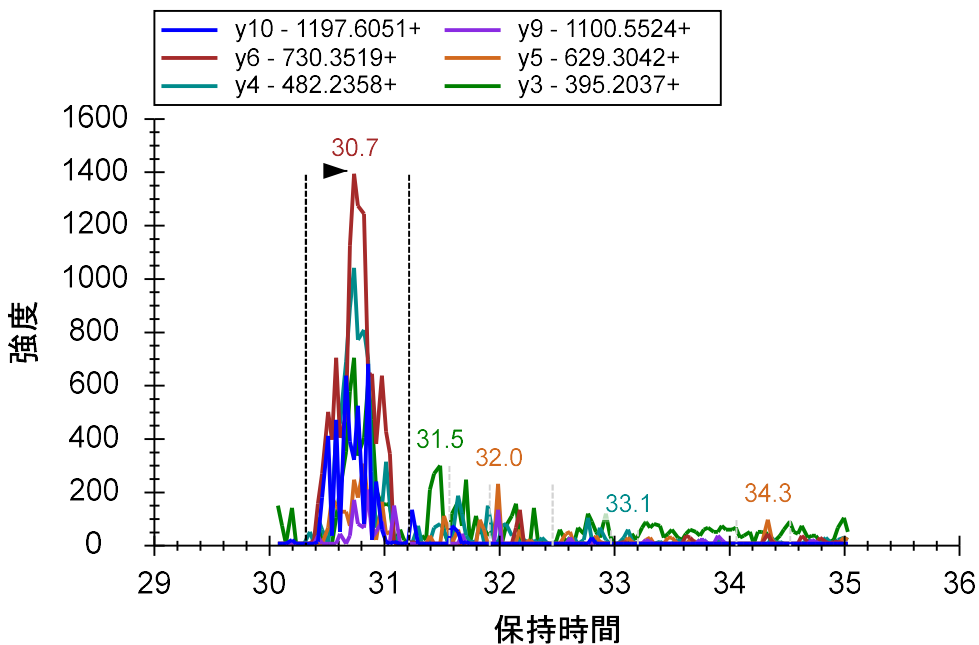
次にこれらのピークを検査・修正します。

- [ピーク領域] ビュー内の D_103_REP3 バーをクリックします。

これで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表示されます。



補正されたピークのプロットは以下ようになります。



スケジュールモードのウィンドウはターゲットペプチドのクロマトグラム全体を表示しておらず、Skyline の表示では誤ったピークが選択されてしまいました。実は正しいピークが 30.2 分付

近で終わっています。これは「ピークの分断」と呼ばれます。このペプチドの分断されたピークは以下のやり方で補正できます。

- 分断されている 30.4 分付近のピーク直後付近で x 軸下をクリック & ドラッグします。
- 左端までドラッグしてからマウスボタンから指を離します。
- [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図を有効化します。
- 30.5 分付近から左端までグラフの 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図を有効化します。

D_162_REP3 のピークも同様に分断されていますが、他同様 Skyline ではピークの実選択は正しくなされています。このピークの実断については[保持時間] プロットを綿密に調べることである程度推定できます。特にバーの長さおよびバーの端と x 軸の距離に注意して観察してみてください。

ピークの実断は今回のように同位体標識（内部標準）なし測定の場合無視できない問題となります。ピークの各ポイントでライトペプチドプリカーサーとヘビーペプチドプリカーサーの相対比が比較できるように同位体標識した内部標準ペプチドを利用すると今度は分断されたピークの実断精度は失われますが、ライトペプチドとヘビーペプチド間の比は有効なままです。同位体標識された内部標準ペプチドなしの場合は当然分断されたピークはペプチドデータとして信頼できません。欠損データを統計学的に処理してみるか、当該ペプチドをデータ解析対象から外すかのどちらかです。今回のチュートリアルでは以降の操作で使うツールで有効になるよう分断されたピークを解析してみます。

Skyline ではピーク実断を以下のように定義しています。片方のピーク端のピークの高さがもう片方のピークの高さの 1%以上高い場合。

Skyline がピーク実断として判定したピークを見るため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。
- [ピーク実断] のチェックボックスにチェックを入れてください（その他の項目のチェックは外しておきます）。

[検索] フォームは以下ようになります。

- [すべて検索] をクリックします。
- [閉じる] をクリックします。

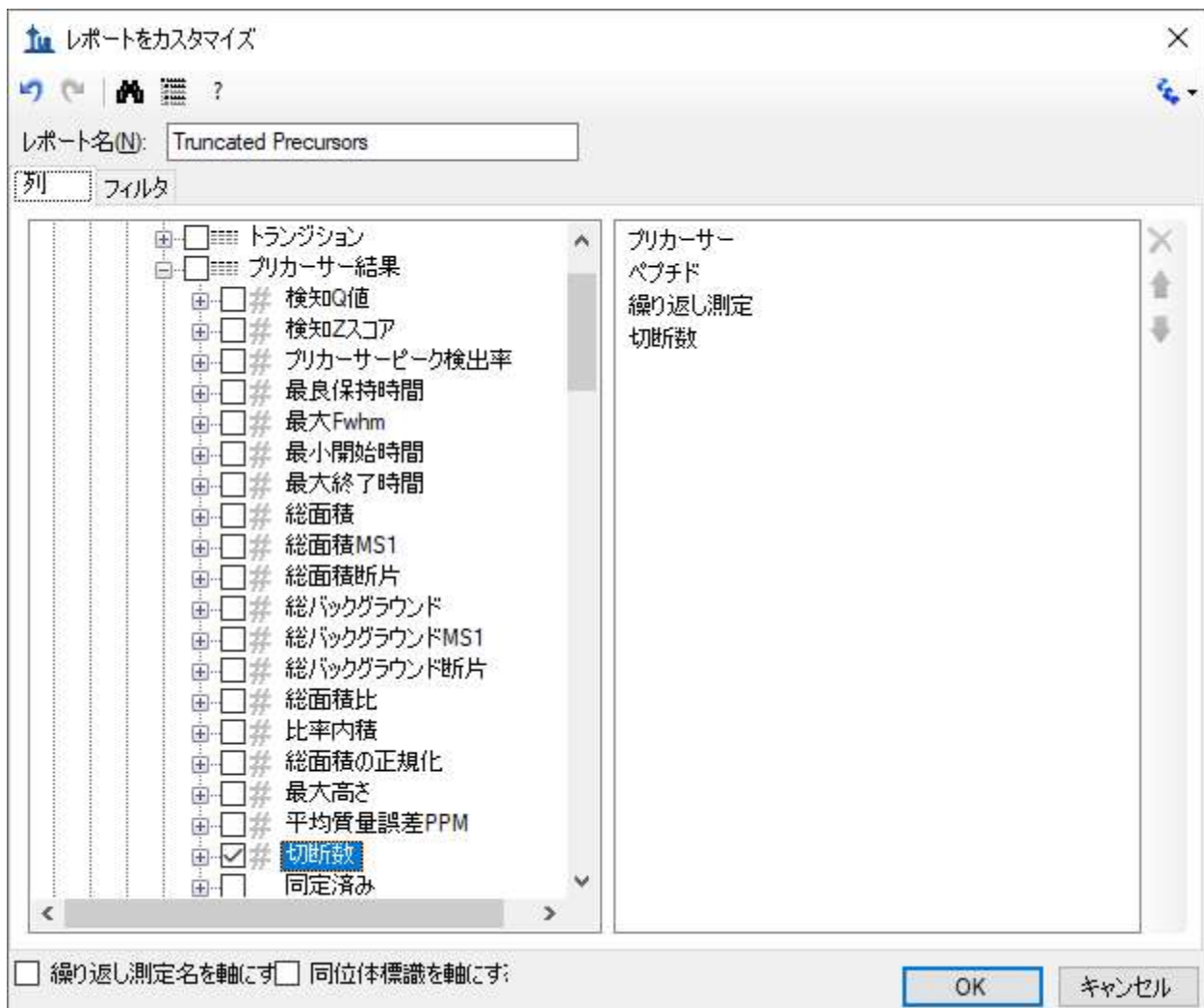
Skyline に分断されたプリカーサーおよびトランジションの長いリストが表示され、すでに調べたプリカーサーピーク 3 つが冒頭に記載されているのが見えます。ラインをダブルクリックすると対応するクロマトグラムグラフが有効化されます。

場所	タイプ	テキスト
905.9565++ (D_103_REP3)	プリカーサー 結果	6個の切断ピーク
905.9565++ (D_108_REP2)	プリカーサー 結果	6個の切断ピーク
905.9565++ (H_159_REP2)	プリカーサー 結果	6個の切断ピーク
P [y10] - 1197.6051+[5](比率296.28) (H_1...	トランジション 結果	切断ピーク
W [y9] - 1100.5524+[6](比率3058.92) (D_...	トランジション 結果	切断ピーク
T [y6] - 730.3519+[1](比率39951.41) (D_1...	トランジション 結果	切断ピーク
T [y6] - 730.3519+[1](比率14467.67) (H_1...	トランジション 結果	切断ピーク
T [y6] - 730.3519+[1](比率47677.84) (H_1...	トランジション 結果	切断ピーク
T [y6] - 730.3519+[1](比率41536.56) (H_1...	トランジション 結果	切断ピーク
T [y6] - 730.3519+[1](比率16914.63) (H_1...	トランジション 結果	切断ピーク
T [y6] - 730.3519+[1](比率21305.15) (H_1...	トランジション 結果	切断ピーク
T [y6] - 730.3519+[1](比率3739.68) (H_16...	トランジション 結果	切断ピーク
F [y5] - 629.3042+[4](比率3928.79) (H_16...	トランジション 結果	切断ピーク
F [y5] - 629.3042+[4](比率6538.54) (H_16...	トランジション 結果	切断ピーク

またドキュメントグリッドを使用すれば、すべての分断されたプリカーサーイオンのピークリストが作成できます。

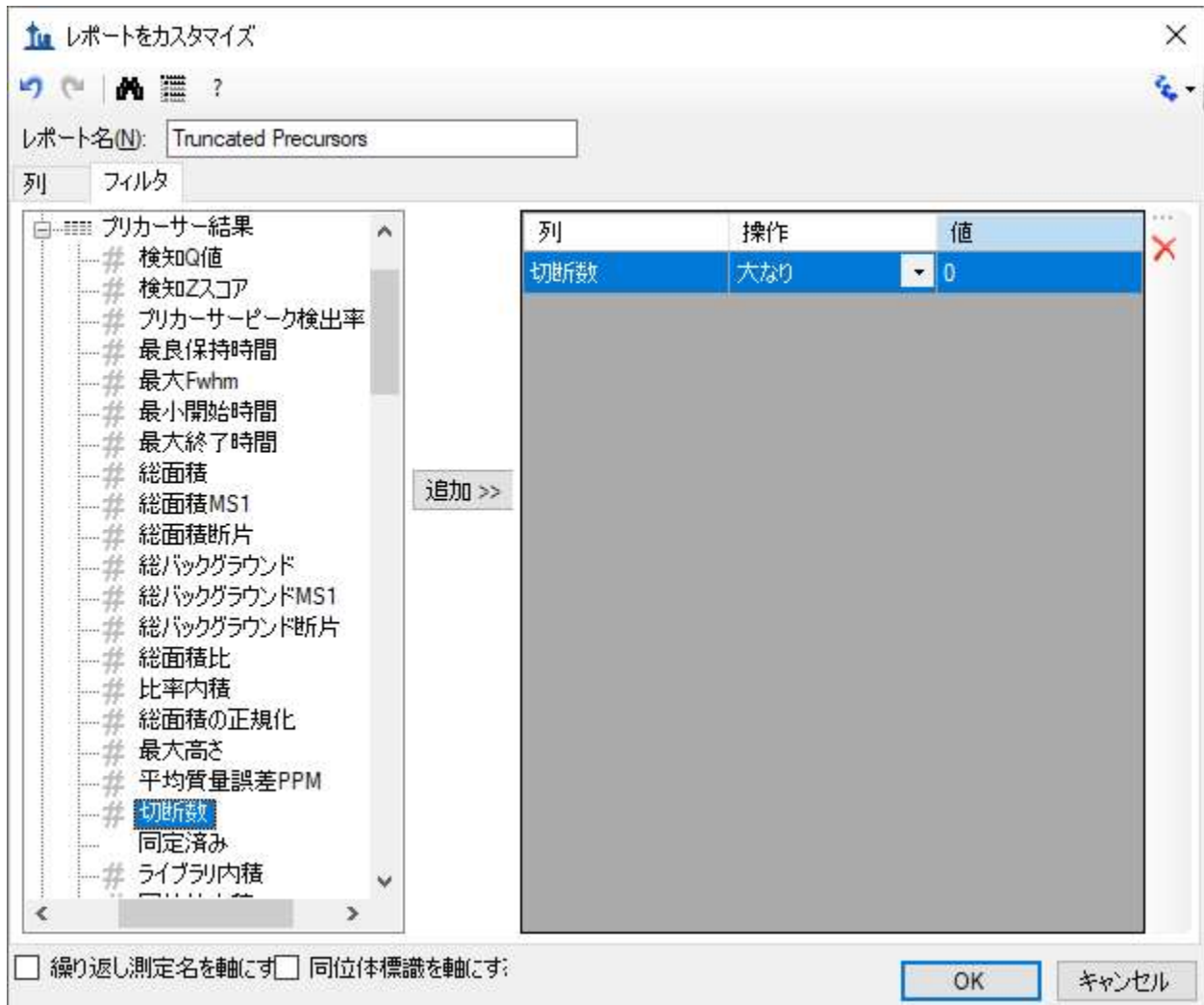
- [ビュー]メニューで[ドキュメントグリッド]をクリックします。
- [ドキュメントグリッド]ビューの[レポート]メニューで[プリカーサー]をクリックします。
- [ドキュメントグリッド]ビューの[レポート]メニューで[レポートのカスタマイズ]をクリックします。
- [レポート名]フィールドに「Truncated Precursor (分断されたプリカーサーイオン)」と入力します。
- リストの右側で、最初の2つ(プリカーサーおよびペプチド)を除くすべての列を選択します(クリックとシフト-クリックを利用)。
- 右縁付近の✖ボタンをクリックして最初の2つ以外のすべての列を削除します。
- 左側のツリー内の「繰り返し測定」のチェックボックスをオンにします。
- ツリーを展開しタンパク質>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。

[レポートをカスタマイズ]フォームは以下のように表示されます。



- 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「(指定の値) より大きい」を選択します。
- [値] フィールドに「0」と入力します。

[レポートをカスタマイズ] フォームは以下のように表示されます。



- [OK] ボタンをクリックします。
- 「切断数」列ヘッダーで[降順]をクリックします。

[ドキュメントグリッド]に自動的に選択された 221 の分断されたピークがリストアップされます。

ドキュメントグリッド: Truncated Precursors

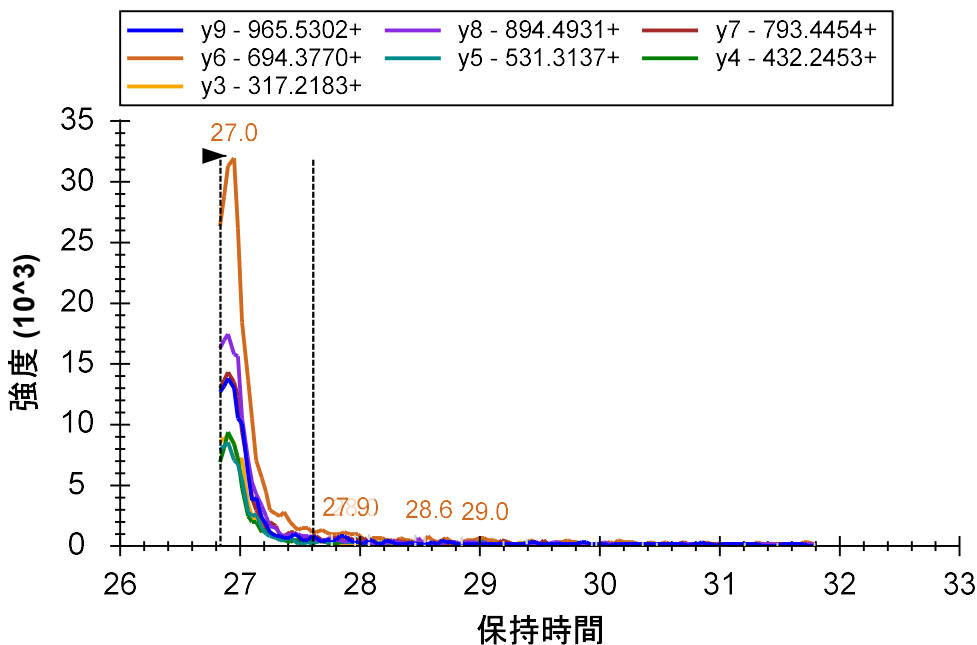
レポート ▾ | 1 of 221 | エクスポート...

	プリカーサー	ペプチド	繰り返し測定	切断数
▶	614.3164++	DFATVYVDAVK	D_196 REP3	7
	614.3164++	DFATVYVDAVK	H_147 REP2	7
	614.3164++	DFATVYVDAVK	H_148 REP2	7
	614.3164++	DFATVYVDAVK	H_159 REP1	7
	614.3164++	DFATVYVDAVK	H_159 REP3	7
	614.3164++	DFATVYVDAVK	H_160 REP2	7
	614.3164++	DFATVYVDAVK	H_162 REP3	7
	730.8488++	DYVSQFESSTL...	D_196 REP2	7
	730.8488++	DYVSQFESSTL...	D_196 REP3	7
	730.8488++	DYVSQFESSTL...	H_147 REP2	7

これらのピークを調べるため以下の操作を行います。

- DFATVYVDAVK（「ペプチド」リンク）をクリックします。
- このペプチドの各「繰り返し測定」リンクをそれぞれクリックします。

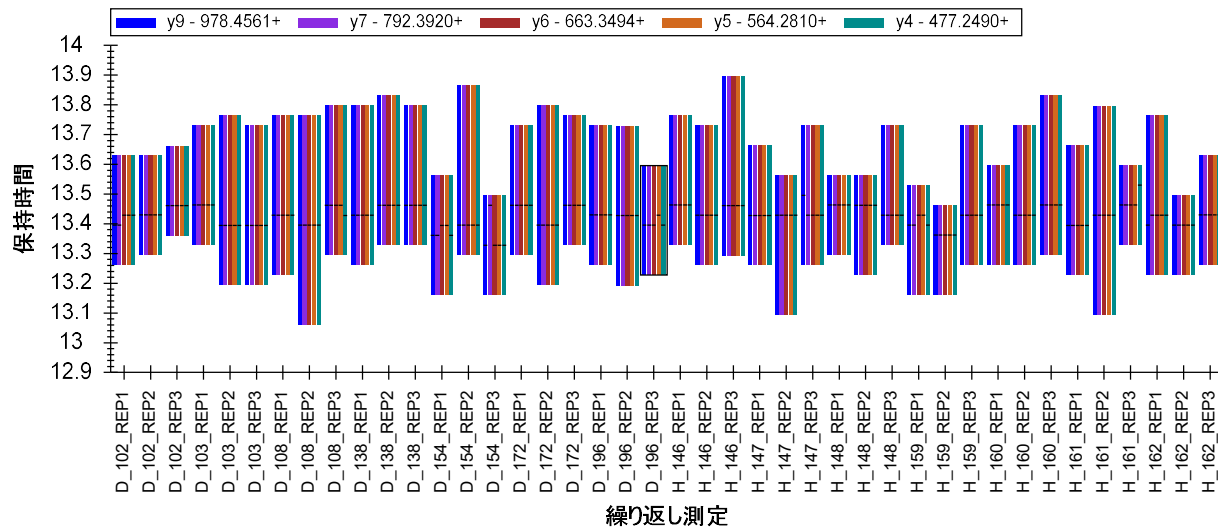
Skyline で以下のクロマトグラムグラフが有効化されます。



- [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.LGGEVSVACK.L [237, 247]を選択します。

複数の繰り返し測定データ処理 1

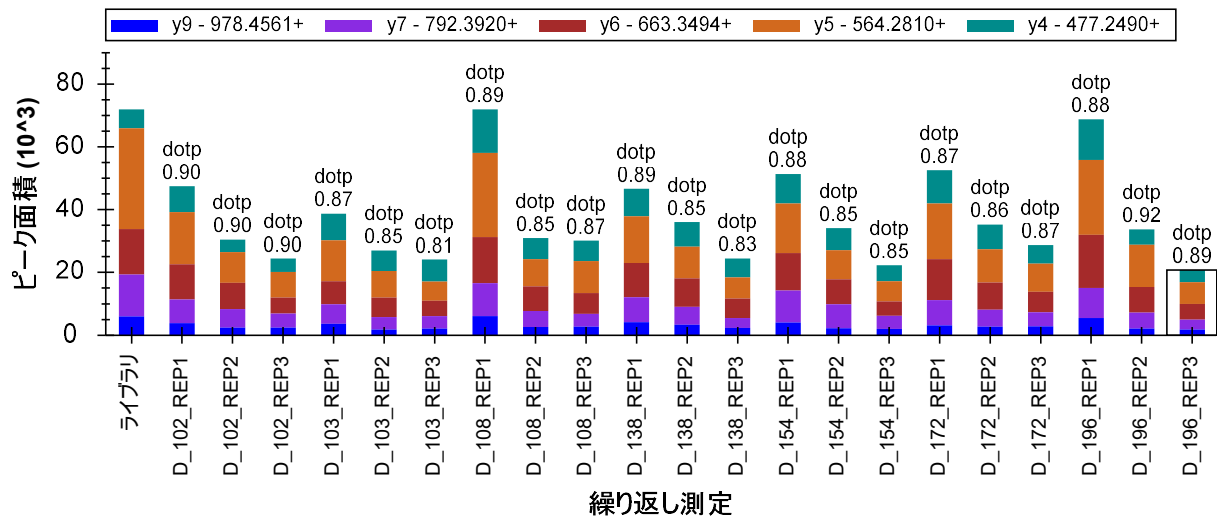
ペプチド LGGEVSVACK はこれまで見てきた他の 2 つのペプチドよりもかなり良い結果となっています。保持時間は非常に安定していますしピークトップも各繰り返し測定間でほとんど一致しています。



[ピーク領域]ビューに表示される相対イオン強度も比較的安定しています。このペプチドには対応するライブラリスペクトルがあり、ペプチドの相対イオン強度（プロット内の一番左側のバーに表示）が測定ピークと類似していることがわかるでしょう。ペプチドが対応するライブラリスペクトルとどの程度一致しているかを詳しく見るため以下の操作を行います。

- [ピーク領域] グラフを右クリックして [正規化] を選択し [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大部分をカバーできるようにします。もしスクリーンスペースが限られている場合は疾患群 (D_) 繰り返し測定付近の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と測定ピーク間のピーク強度類似性を示すドット積 (dotp) が表示されます。



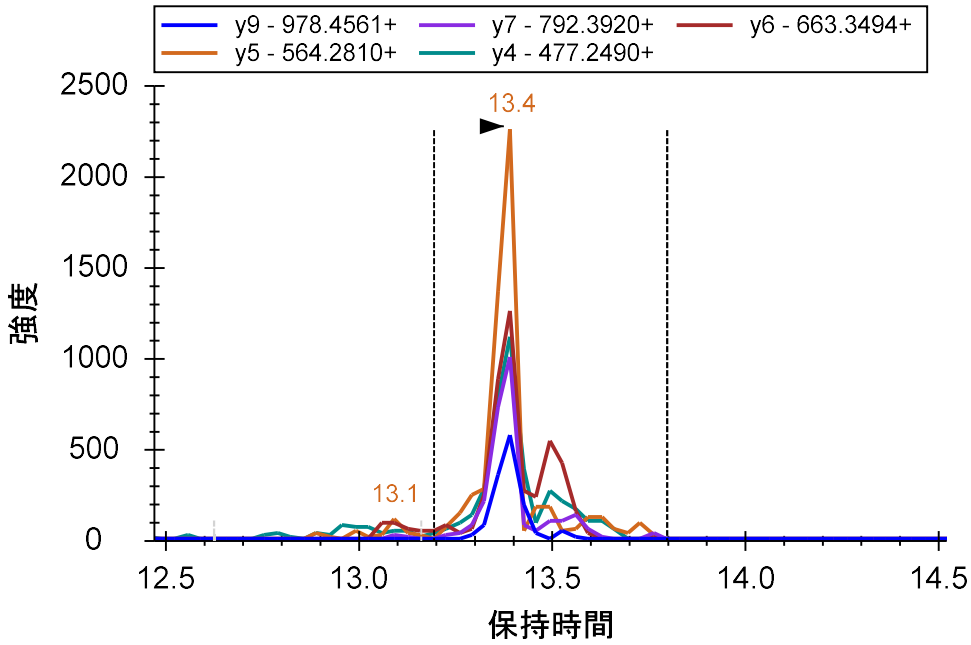
以下の操作を行って [ピーク面積] ビューを元の状態に戻します。

- ズームを使用している場合、右クリックから [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。

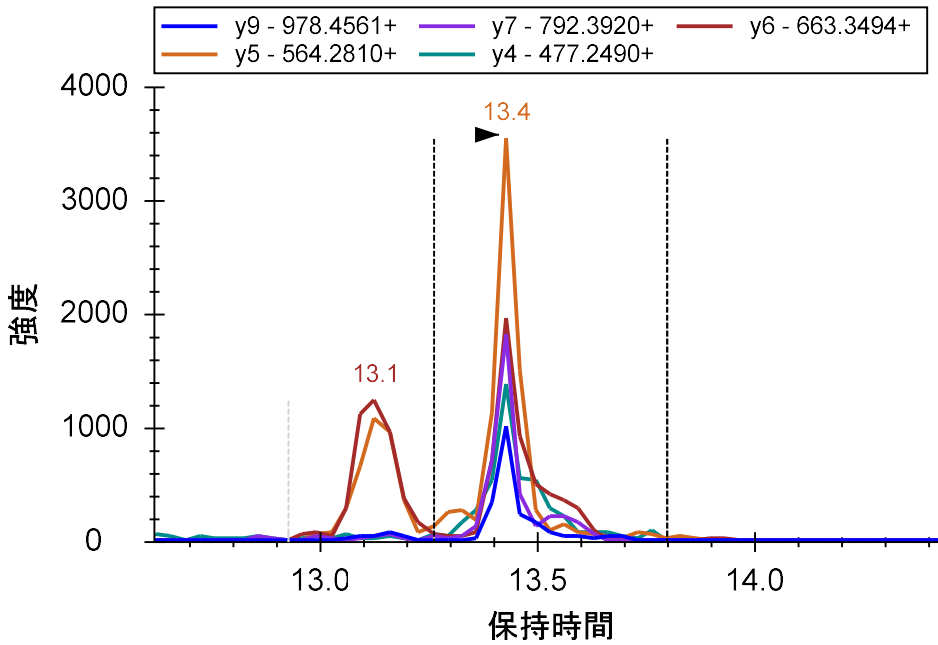
これらのピークのクロマトグラムを詳しく見るため以下の操作を行います。

- [ビュー] メニューから [自動ズーム] を選択し [最適ピーク] (F11) をクリックします。
- [ターゲット] ビューのドロップダウンリストから [繰り返し測定] 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をスキャンします。

いくつかのプロダクトイオン (y ion) はメインピークとは別の位置にピークがあることに気付くでしょう。

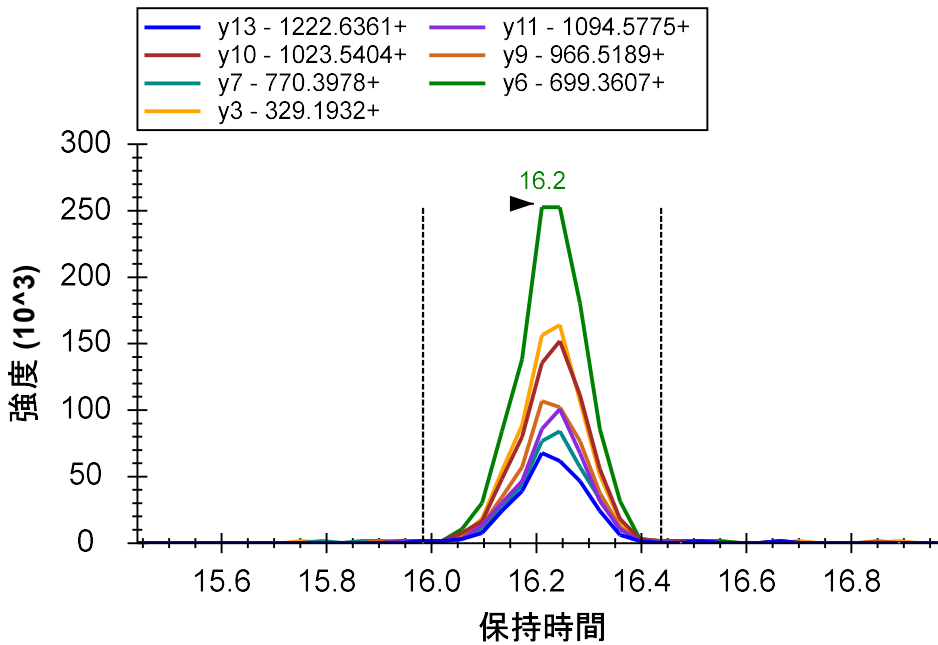


13.1 分付近の y5 および y6 には明確なピークシグナルが見られますが、これは明らかに別のペプチド由来のものと思われます。

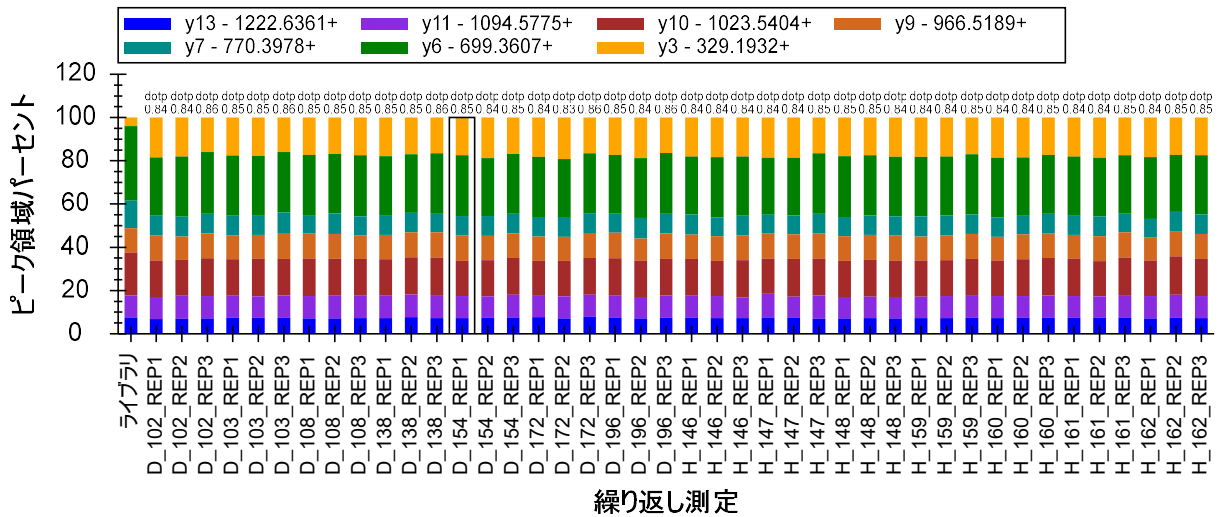


このような干渉シグナルが積分境界内で起こると定量エラーも当然増えてしまいます。定量の際にはこのペプチドの 5 つのトランジションすべてが本当に必要かどうかまず検討してみてください。

次のペプチドのデータを見てみましょう。今度は 7 個のプロダクトイオンが高い強度でかつピーク形状も整っています。また保持時間もそろっています。



プロダクトイオン同士の相対イオン比率も非常に一定しています。

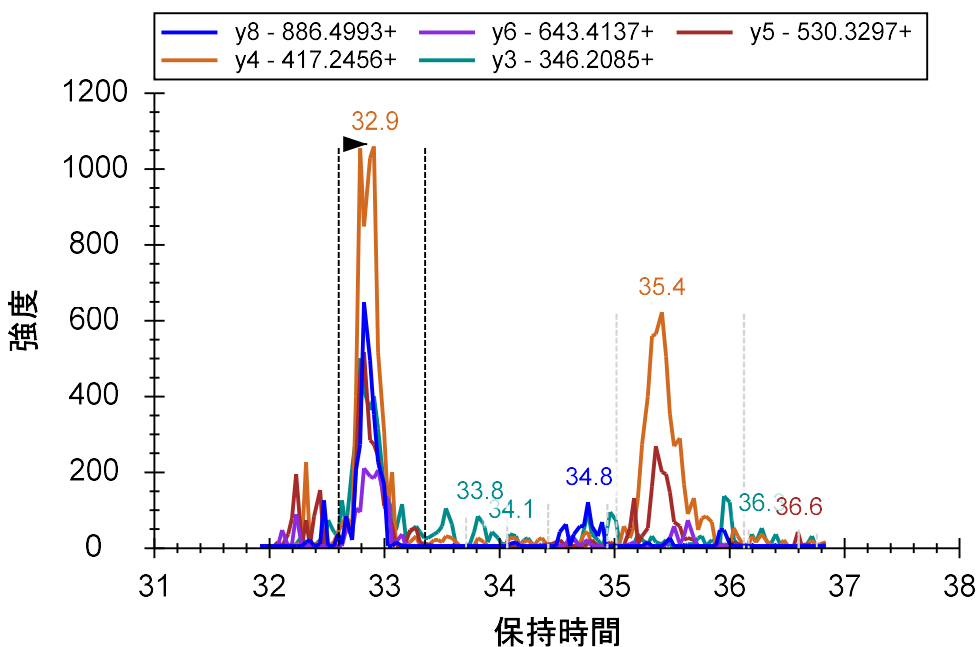


プロダクトイオン y3 はライブラリスペクトルと比較して相対的に強く検出されています。これは質量分析装置の違いに起因するのかもしれません。

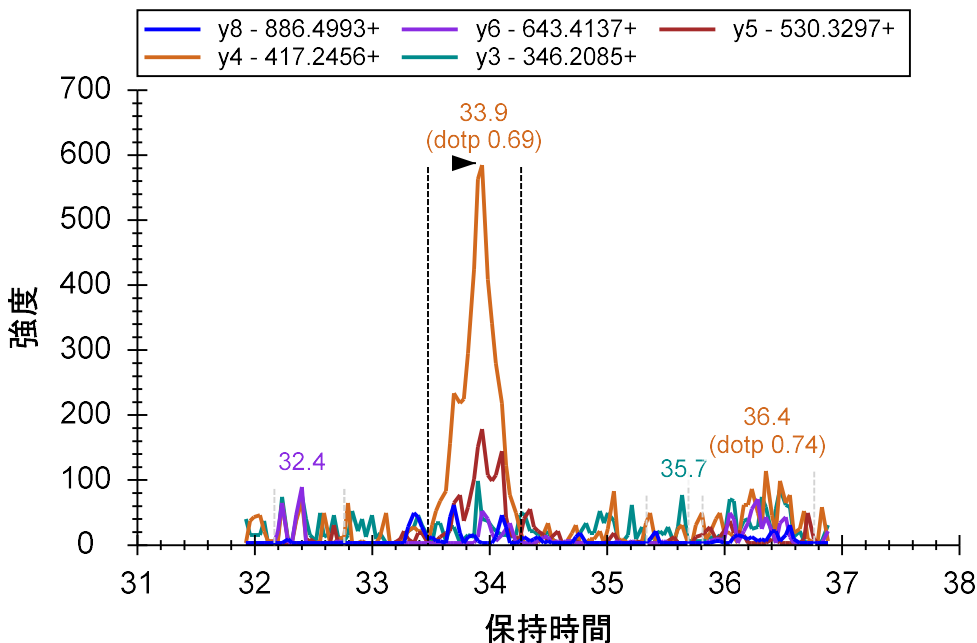
この積分は十分に良好なので次のペプチド R.GSYNLQDLLAQAK.L [378, 390]に移行します。この次のペプチドについては[保持時間]ビューと[ピーク領域]ビューの両方に問題が示されています。何が問題なのかを理解するため以下の操作を行います。

- [ビュー]メニューで[自動ズーム]を選択し[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 D_103_REP3 まで冒頭から繰り返し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見てみると選択されたピークが約32.5分~33.5分にあり、5つのトランジションすべてにシグナルがあります。しかし、35.4分付近で主にy4およびy5イオンのピークが見られます。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。このような場合はターゲットペプチドと共溶出しない干渉シグナルの情報は表示中のクロマトグラム「環境」に追加できます。



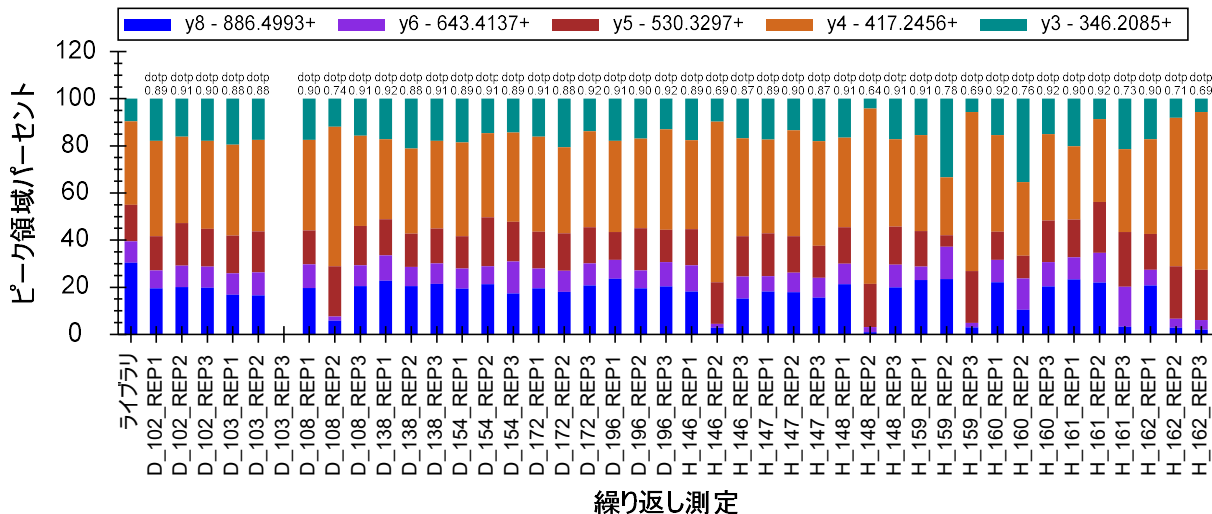
D_103_REP3 では 33.9 分に y4 および y5 のプロダクトイオンピークが見られますがその他のプロダクトイオンピークはほとんど見えない状態です。



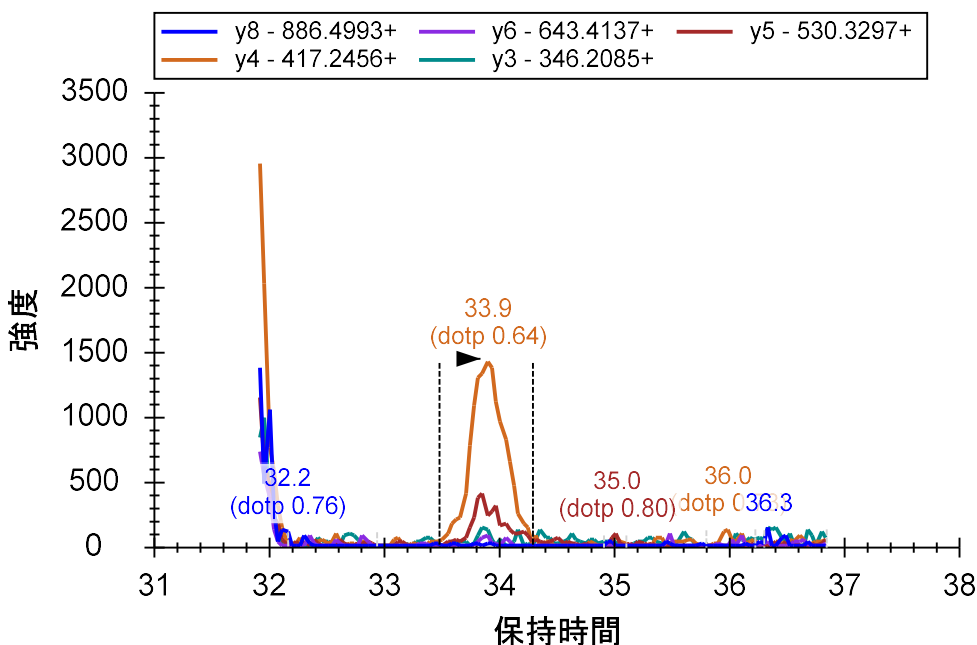
これはスケジュールウィンドウがターゲットペプチドのシグナルキャプチャに失敗したためです。不正確なピークを削除するため以下の操作を行います。

- クロマトグラムを右クリックし[ピークを削除]をクリックします。

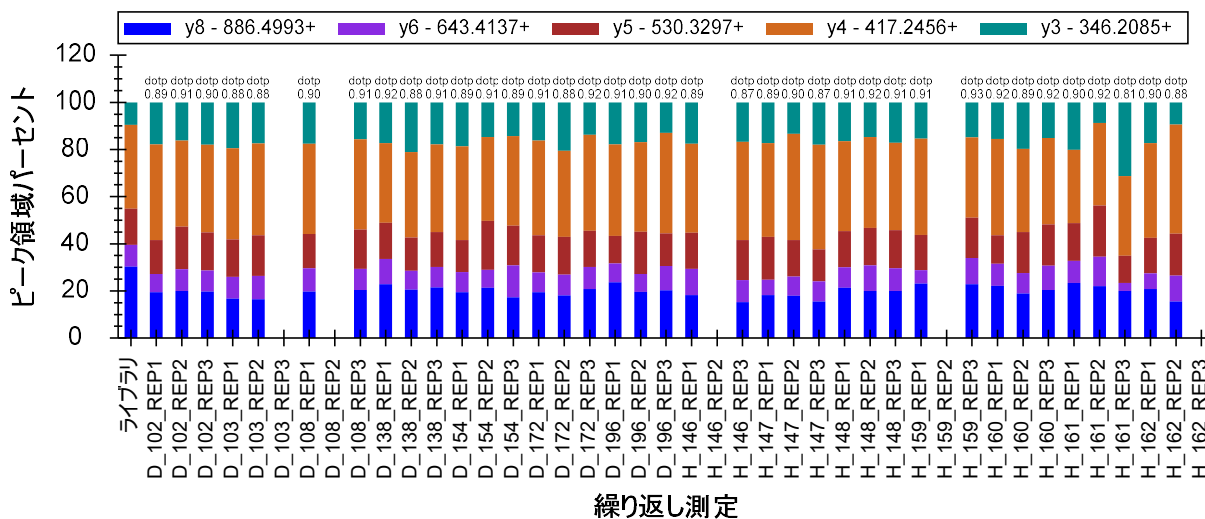
[ピーク領域パーセント]のグラフで「D_103_REP3」は空白になりました（下図参照）。



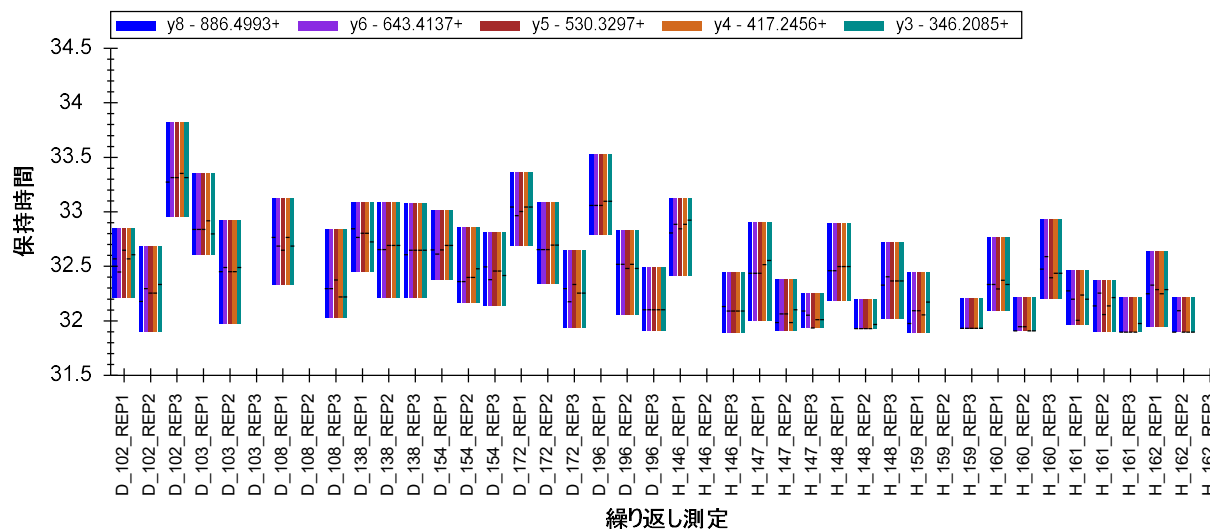
クロマトグラムプロットを続行すると同様の問題が生じていることがわかるので上記のように削除をしてください。またピークが分断されたケースも一部見られます。そこで H_148_REP2 に注目すると 33.9 分のピークが選択されていますが実は正しい（メインピーク）は 32 分のあたりのもので、33.9 分のピークは分断されたピークとなっています。



上記で行ったように x 軸の下をクリック & ドラッグすれば分断されたピークとして積分可能です。ピーク領域プロットは以下ようになります。



また保持時間プロットは以下のようになります。



- ペプチド R.DIPVNPMCIYR.S [46, 56]へと続けていくと、

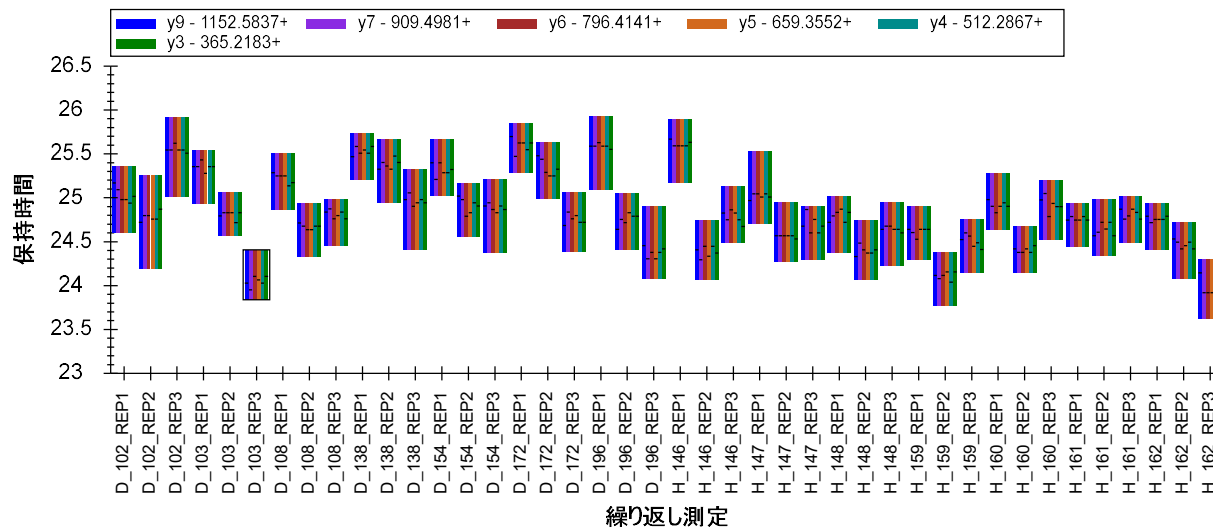
5分のスケジュールウィンドウのどこにも一致しているピークはありません。

- したがってこのペプチドは消去します。

NP_001012027 タンパク質由来の次のペプチド 3 つはすべて保持時間も安定していて相対イオン強度もそろっており良好なデータに見えます。

測定間の保持時間のずれは予想よりそろいが悪いと思われるかもしれませんが。のではなく、表示されている並び順は繰り返し測定をドキュメントへとインポートした順で測定順に並んでいるわけではありません。この種の研究ではすべての疾患検体を最初にリストしその後すべての健常検体をリストするとよいでしょう。各検体のすべての技術的繰り返し測定もまたグループ化できます。

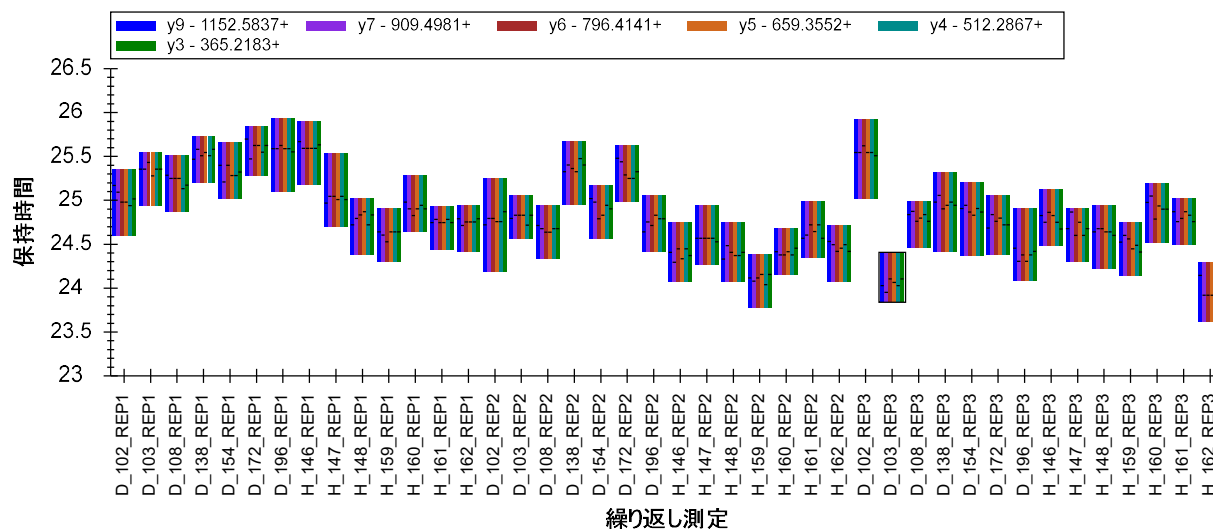
ペプチド TSDQIHFFFAK の保持時間は以下ようになります。



測定された順ごとに分析結果を見る方がよい場合もあります。これは以下の操作で実行可能です。

- [保持時間] グラフを右クリックして [並べ順] を選択し [取得時間] をクリックします。

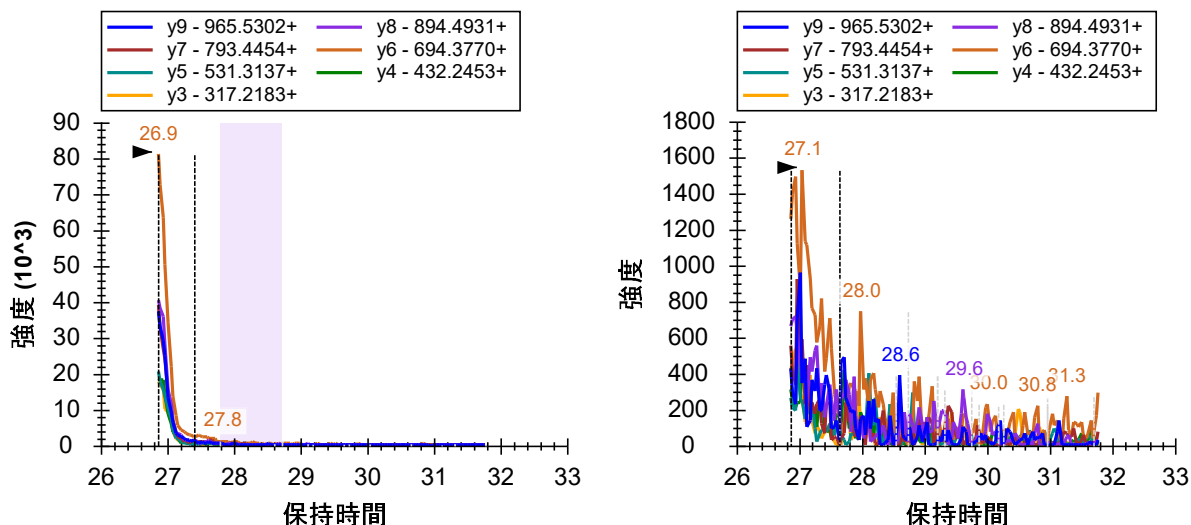
以下のように表示は変更されます。



今度は保持時間の一貫性がやや向上しました。次に 2 つのペプチド (LQPLDFK、SQLPGIIAEGR) を選択する、パターンの一貫性を確認できます。

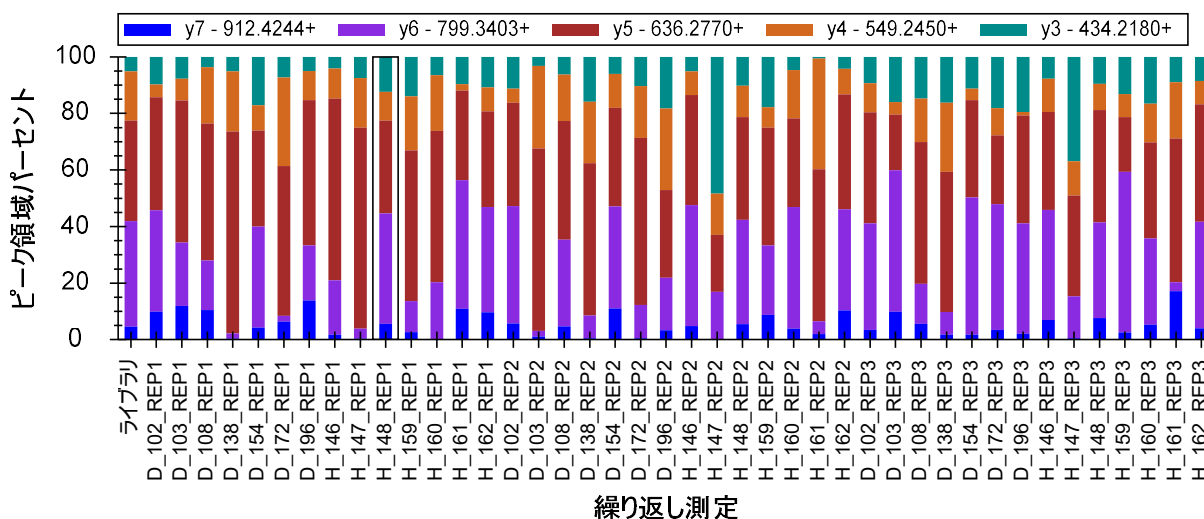
DFATVYVDAVK はピークが分断されています。そこで一致していない [ピーク領域] ビュー上のバーをクリックして対応するクロマトグラム x 軸の下をクリック&ドラッグして修正します。

またノイズが数多く見えるクロマトグラムがあります（右下図）。要するにこれはピークの選択が間違っていたためです（目的のピークは27分より前に存在する）。



このようなケースではノイズを積分しないことが重要です。この場合は繰り返し測定のピークそのものを削除してしまう方が楽でしょうね。

このセクションで最後にペプチド FGLYSDQMR を見てみます。[保持時間] ビューで、Skyline は19分に近いピークを選択しているのがわかります。しかし [ピーク領域パーセント] ビューを見ると相対イオンの比率はで測定間で一貫性がないことがわかります。



いくつかのクロマトグラムを分析すると 18.8 分のピークはすべてが同じペプチドに由来しているとは考えにくいですね。そこでこのペプチドデータを消去して先に進みます。

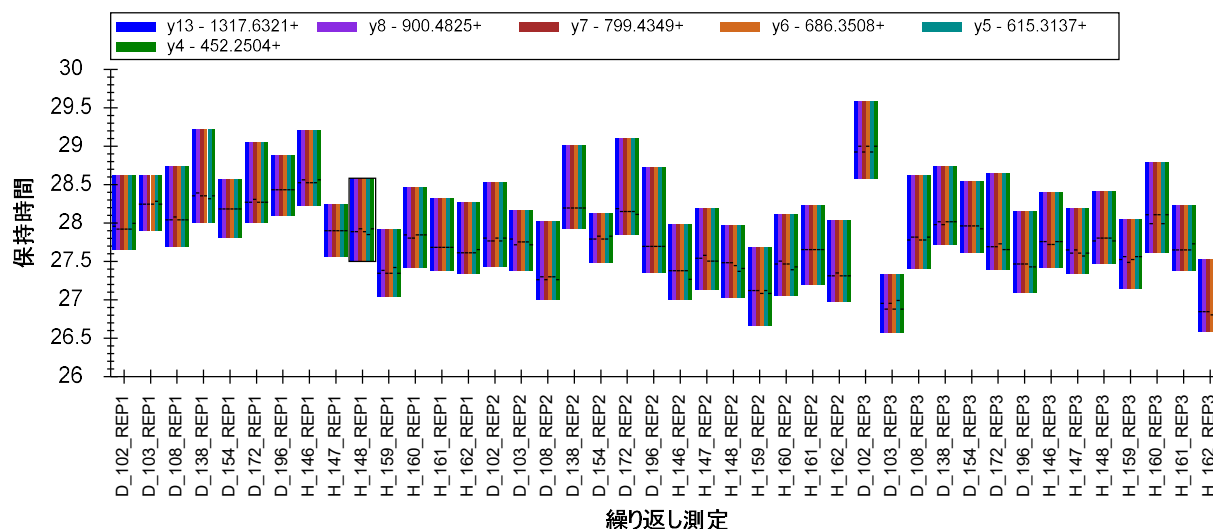
正規化のための標準合成ペプチドの利用

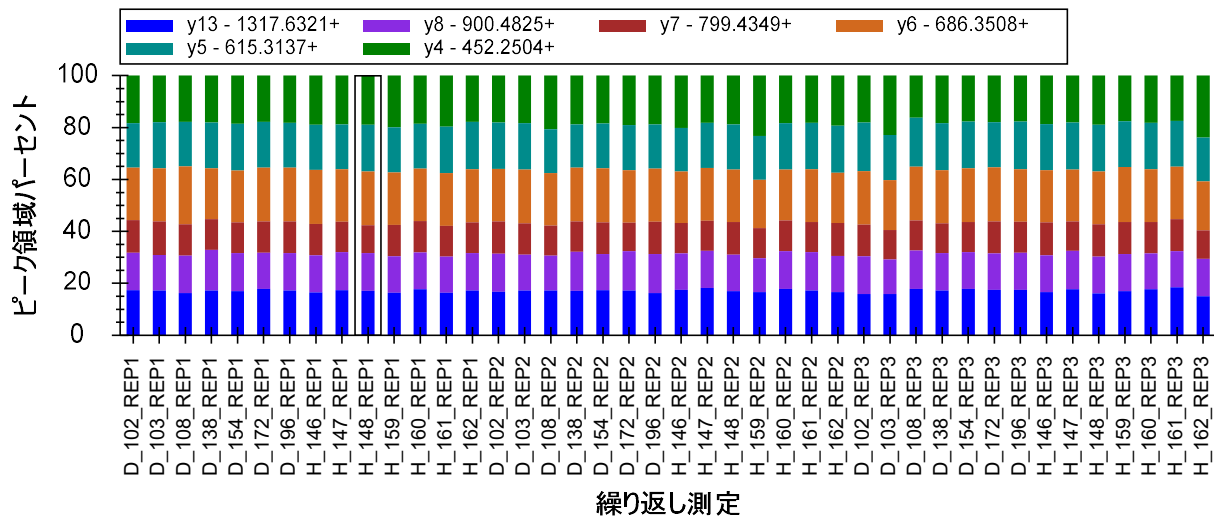
このデータセットには安定同位体標識 labeled (SIL: stable isotope labelled peptide) ペプチドが使用されていませんが、合成グローバル標準ペプチドが試料にスパイクされています。これらのペプチドを使用する目的は LC-MS/MS 分析時の測定間のブレを低減するため、そして測定

ペプチドのピーク面積の正規化をするためです。²そのため以下の操作を行います。

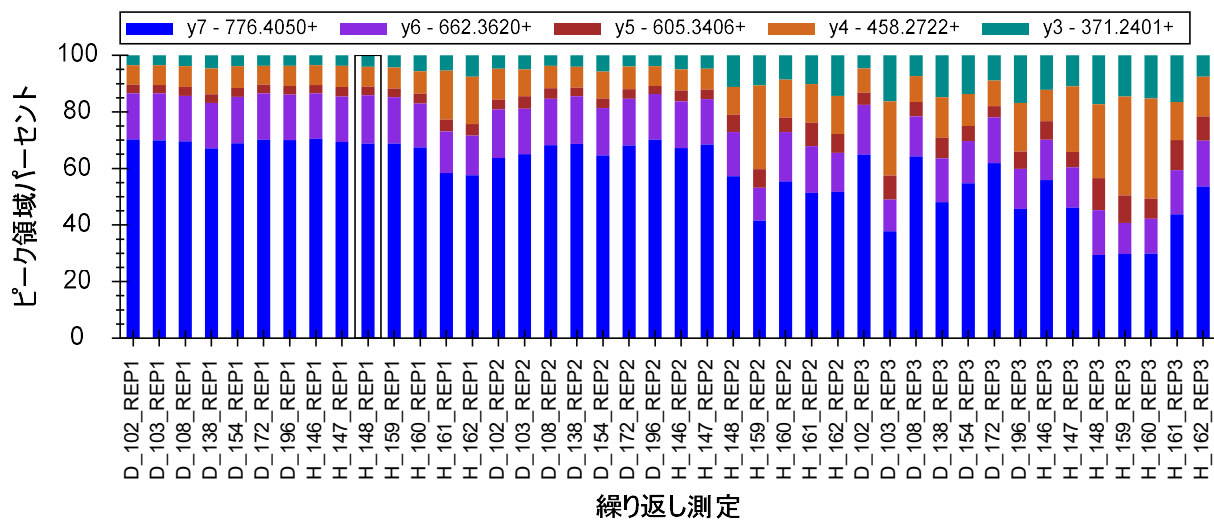
- [ターゲット]ビューをクリックします。
- キーボードの End キーを押し、上向き矢印キーを押します。

これで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、VVLGSGDATLAYSAFK と AFGLSSPR は、本実験で使用した合成ペプチドです。これまで学習してきた 2 種のプロットで見ると最後のペプチド 2 つについては自動積分の精度が高いと確信できます。保持時間とピーク面積は比較的そろっています。



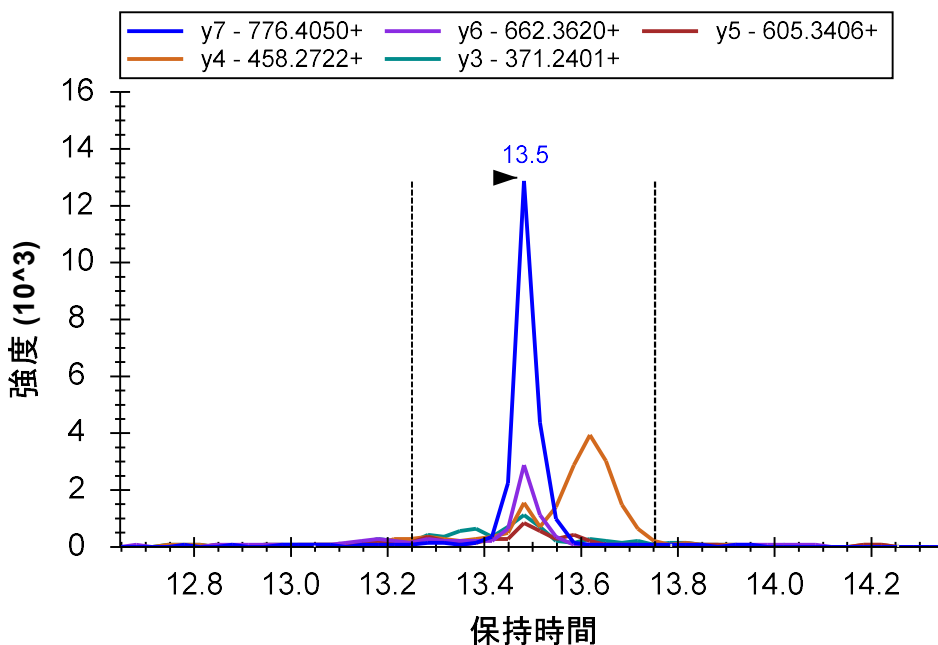


ペプチド HLNGFSVPR を見てみましょう。保持時間は安定していますが、ピーク面積はばらつきが大きいです。特に y4 と y3 の面積比はデータ全体で見ればばらつきが目立ちます。



y4 の面積比が大きい検体についてピーク面積バーをクリックすると、y4 そして y3 でのもめらかな干渉が見られます。

- [ビュー]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。




注入したペプチドが正規化に使用できるか精査するため以下の操作を行います。

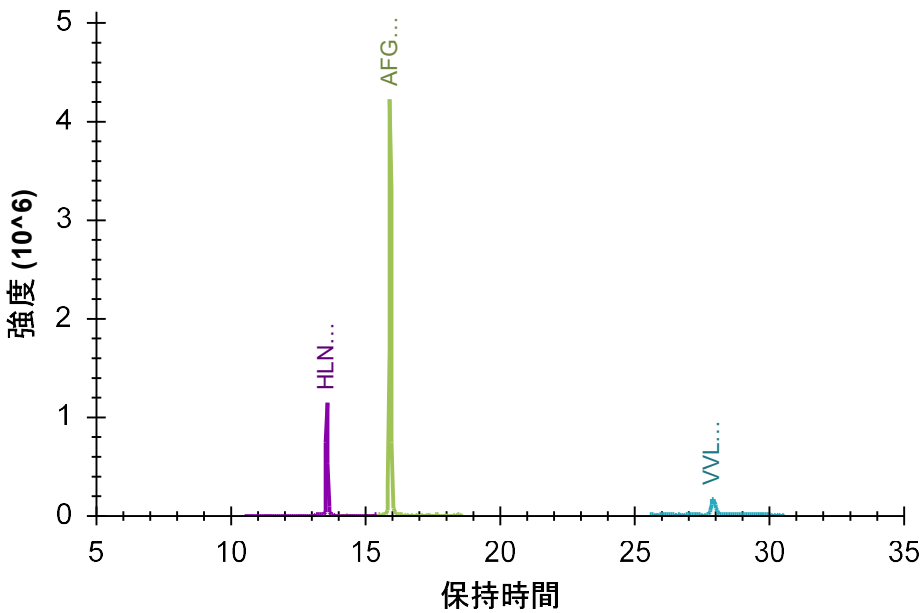
- [ピーク面積] グラフ上で右クリックして [正規化] を選択し [なし] をクリックします。

ここでペプチド HLNGFSVPR のピーク面積は測定の最初 (6,000,000) から最後 (30,000) に移行するにつれて劇的に減少しているのが分かります。ピーク面積は最初と最後でおよそ 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++ について [結果グリッド] ([ビュー] > [その他のグリッド] > [結果グリッド]) を使用し、最大および最小ピーク面積を計算できるか試してみてください。次に他の 2 つの標準ペプチドを見てみると、こちらも経時的に減少している (VVLSGSDATLAYSFAFK は 2,300,000 が 1,100,000、AFGLSSPR は 23,000,000 が 1,000,000) ことがわかります。42 回のすべての測定はこれら 3 つのペプチドに対する技術的繰り返し測定であるはずですが、この測定ではすべてのペプチドについて経時的なシグナルの低下が見られますが、200 分の 1 または 20 分の 1 に減少する例はまれです。

これらのペプチドのそれぞれの溶出時間を比較するため以下の操作を行います。

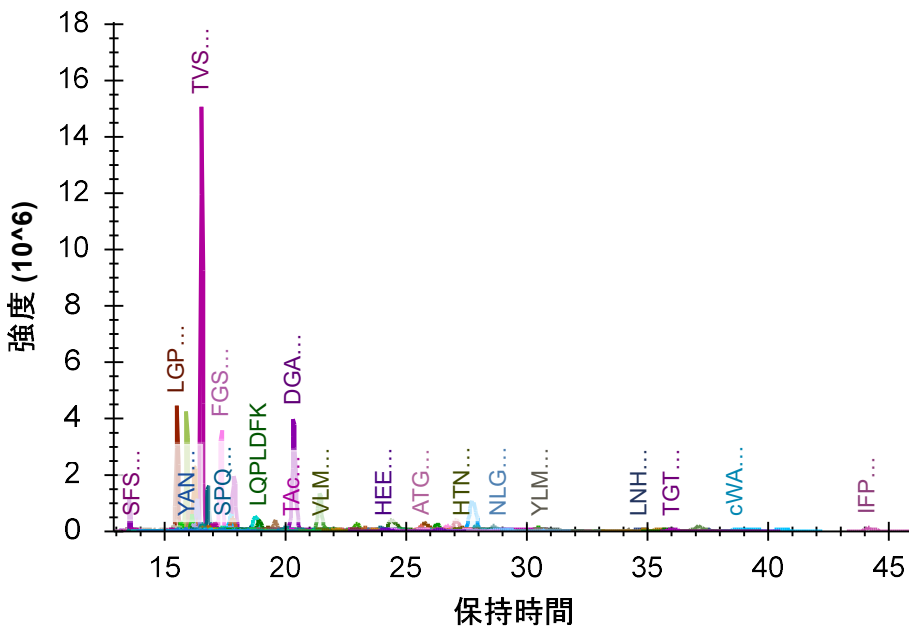
- この  アイコンが先頭にある「S」エレメントを選択します。

クロマトグラムグラフが変更され、「S」リスト内の 3 つのペプチドすべて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ため以下の操作を行います。

- [編集]メニューで[すべて選択] (Ctrl+A) をクリックします。



これで問題の 2 つのペプチドはどちらも親水性であり、溶出時間も早かったことがわかります。実際これらはクロマトグラム上の 3 番目および 10 番目の溶出ペプチドです。これを Skyline で見るため以下の操作を行います。

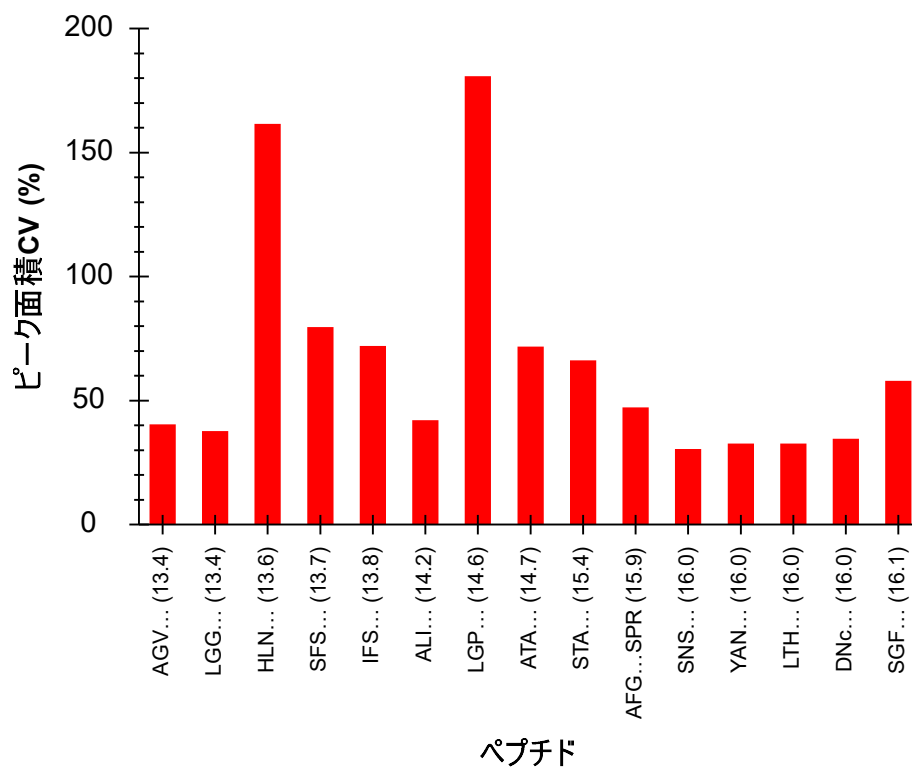
- [ビュー]メニューで[保持時間]を選択し、[ペプチドの比較] (Ctrl+F8) をクリックします。
- [保持時間]ビューを右クリックして[並べ順]を選択し[保持時間]をクリックします。
- クロマトグラムで溶出される最初の 15 ペプチド程度が [保持時間] グラフ内に入るように長方形をクリック&ドラッグします。

早く溶出するペプチドは測定間のブレが大きい傾向があります。したがってこれらは標準化のためのペプチド候補としては最適ではありません。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など、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と同様に振舞うかを評価するに以下の操作を行います。

- [ビュー]メニューでから[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]メニューから[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。
- クロマトグラムで溶出される最初の 15 ペプチド程度が [ピーク面積] グラフ内に入るよう長方形をクリック&ドラッグします。

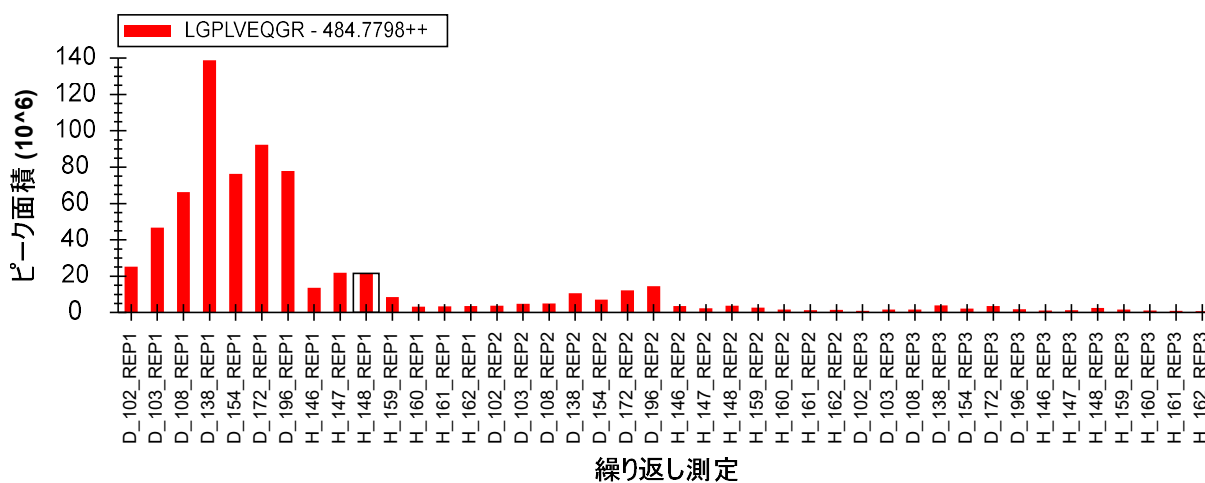
14 サンプルで測定された標準ペプチドではない他のすべてのペプチドは標準ペプチドよりも大きなブレ (測定誤差) があることを想定してください。また積分はすべてのペプチドについて完全に調整できているわけではありません。現在のデータセットについては標準ペプチドについて自動選択されたピークは正しいように見えます。しかし HLNGFSVPR ペプチドは約 160%で 2 番目に高い CV 値を持ち、また AFGLSSPR ペプチドの CV 値は約 50%で、初めに溶出される 15 個のペプチドの中央値と似たような数字となっています。(より疎水性で 108 番目に溶出される VVLSGSDATLAYSAFK ペプチドの CV 値は 18.6%で 80 番目となります。)



以下の操作で LGLP ペプチドには HLN ペプチドと非常に似た問題があることがわかります。

- [ピーク面積] ビューの LGLP ペプチドのバーをクリックします。
- [ビュー] メニューで、[ピーク面積] を選択して [繰り返し測定の比較] をクリックします (F7)。

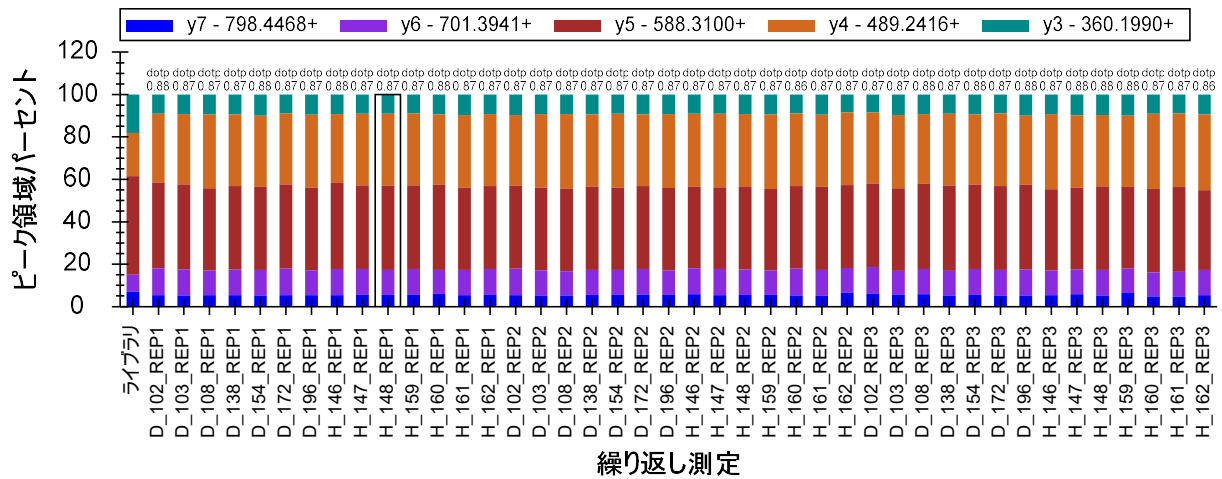
技術的繰り返し測定は 3 回であることを留意してください。



ピークが正しく積分されていることを確認するため以下の操作を行います。

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

最小値と最大値は 200 倍の開き (140,000,000~700,000) がありますが、測定間のプロダクトイオン存在比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正規化のための標準合成ペプチドの対象にはしません。こうした標準ペプチドの選択は重要な測定前に実行すべきことです。この場合は VVLSGSDATLAYSFAK のみが正規化のための標準合成ペプチドの候補となりえます。詳細は本チュートリアル範囲外であるため割愛します。

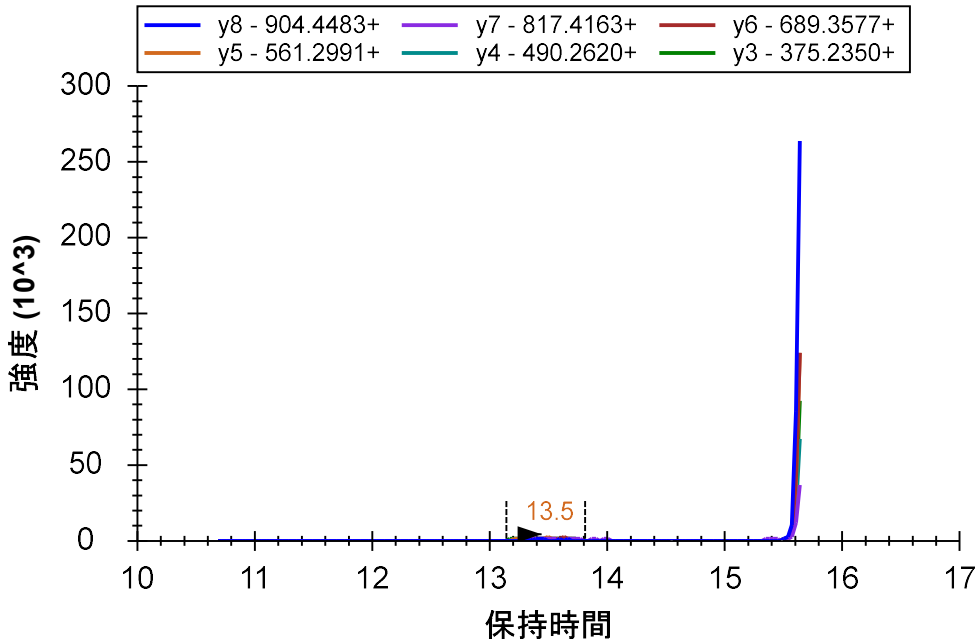
複数の繰り返し測定データ処理 2

今度は「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についての確認と修正をしていきます。

- [ビュー]メニューで、[保持時間]を選択して[繰り返し測定の比較]をクリックします。

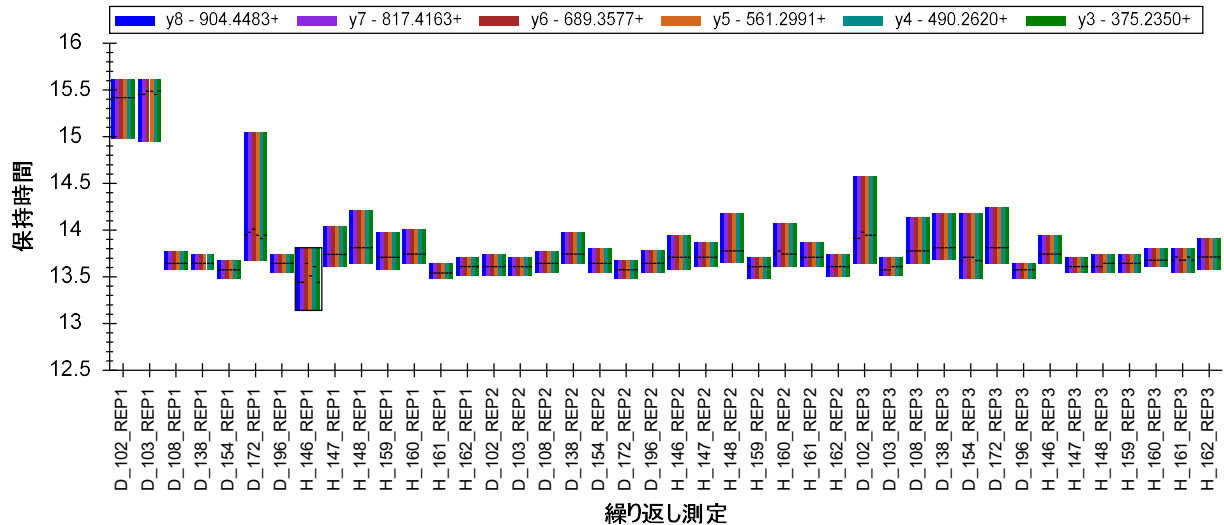
グラフ中で見られる**保持時間**や**ピーク面積**の一貫性から、すべての測定においてペプチドが一定してピーク積分されているという信頼性が得られるはずですが、このペプチドについてはこれ以上精査せず[ターゲット]リストを続行して構いません。他の2つのペプチドも同様となります。

この作業をペプチド (IFSQQADLSR) に達するまで続行します。[ピーク面積]のグラフを見ると H_146_REP1 を除くすべての繰り返し測定は相対イオン強度が一貫していることがわかります。この測定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収まっていないことがわかります。(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。)

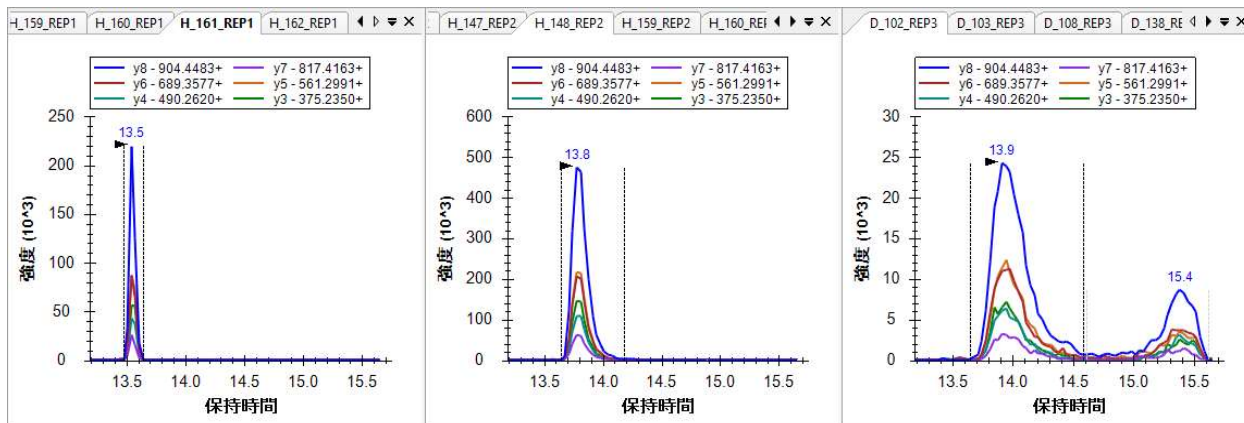


この問題については以前行ったように、x 軸の下をクリック&ドラッグして分断されたピークを積分するか、または右クリックメニューからピークを完全に削除します。

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に通常とは違う測定がいくつかあります。



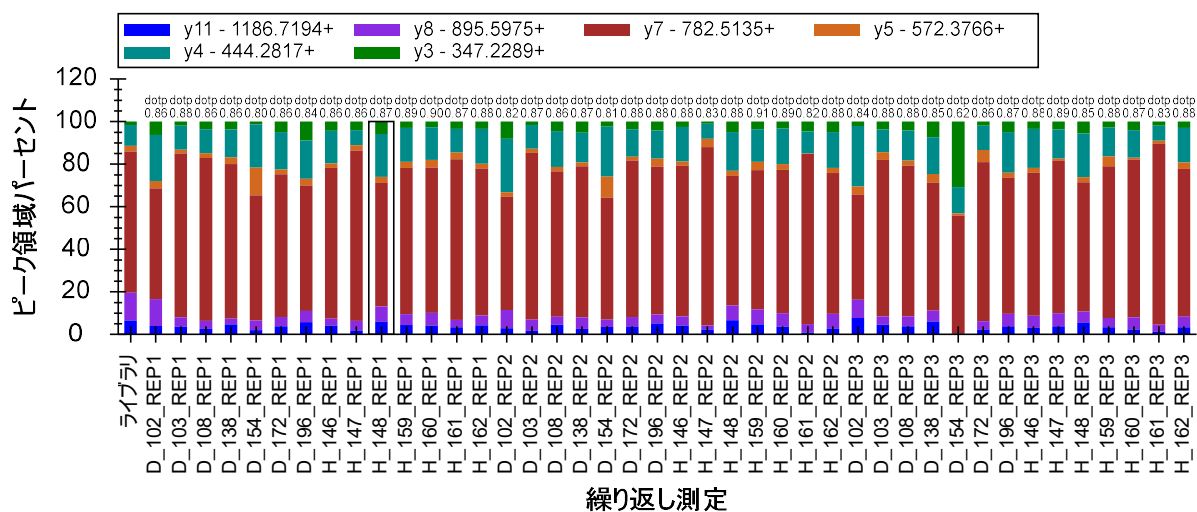
そこでこのペプチドに関するいくつかのクロマトグラムを精査すると、ほとんどが単一の非常に優れた 0.2 分間に溶出するピークであるのが分かります。しかしその他については 0.5 分間にわたってテーリング溶出しているピークであったり、保持時間 2 分間程度の中で 2 つの明白なピークに分断されたものであったりします。



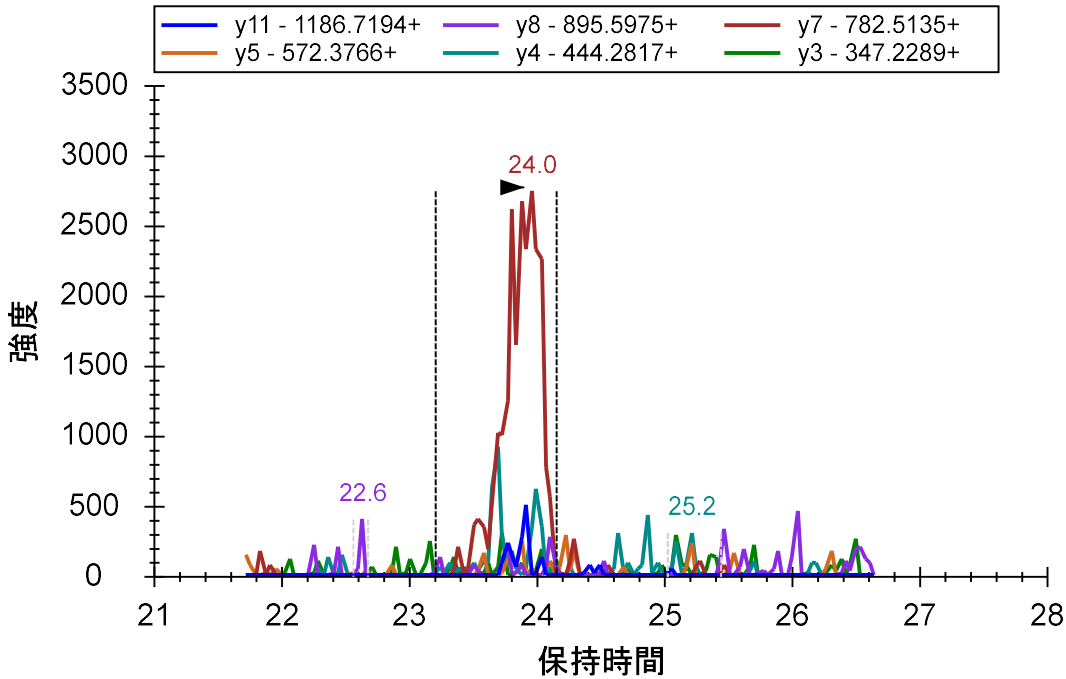
このような場合 SIL (安定同位体: Stable isotope labelled peptide) ペプチドなしに定量することは困難です。そのため本実験ではこのペプチドは消去した方がよいかもしれません。少なくとも、二重溶出プロファイル (ピークが分断されている) ピークは削除すべきでしょう。

続いてペプチドリストを見てみると次の 7 つのペプチドは一貫性が十分にあることがサマリープロットから確認できます。Skyline によるピーク修正は不要で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ります。

他方、ペプチド MLSGIPLKPTVK は大きな変動があることが [ピーク面積比] グラフから分かります。



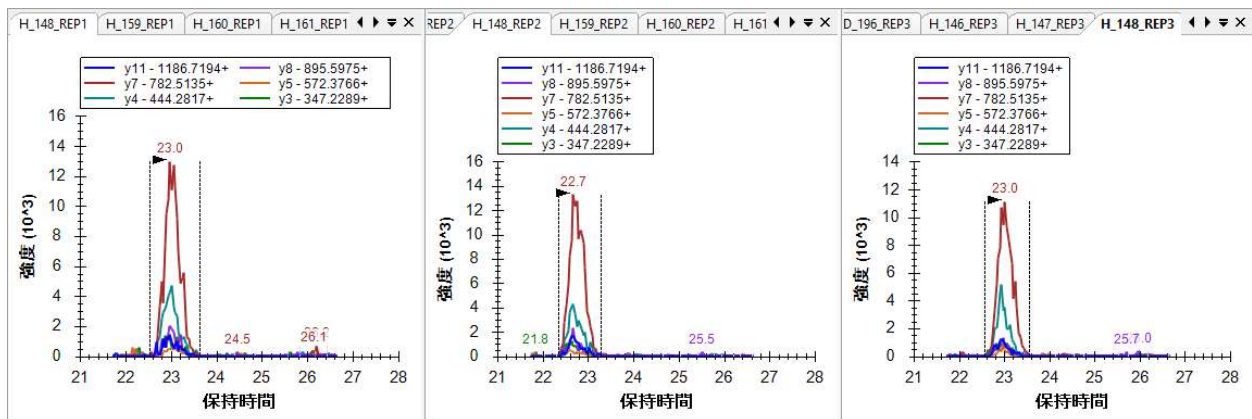
このペプチドでは y7 プロダクトイオンがもっとも多く存在しています。これは上記グラフの左端のライブラリスペクトルでも同じです。クロマトグラムをみただけでは y7 プロダクトイオンと共溶出するイオンがあるかどうかについて識別することは困難です (以下を参照)。



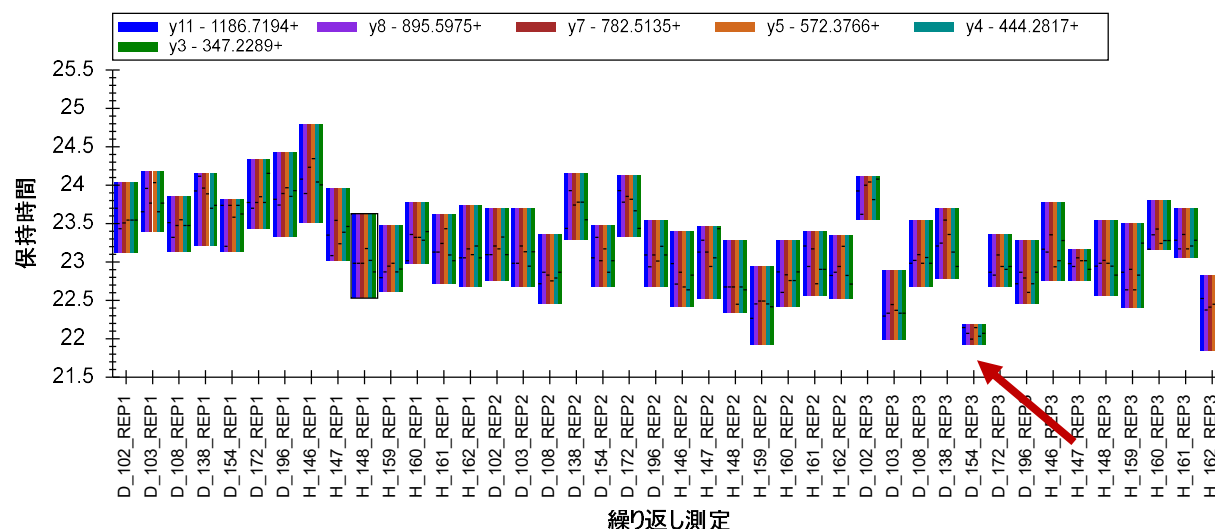
しかしほとんどの場合各測定において同一ピークの同じプロダクトイオンを積分していることは明らかです。y7 プロダクトイオンピーク以外のものがあるかを確認するため最高強度の測定値に注目してください。以下の手順でこれを行うことができます。

- [ピーク面積] グラフを右クリックして[正規化]を選択し[なし]をクリックします。
- 3つの明らかに最高強度であるピーク（すべてサンプル H_148 のピーク）をクリックします。

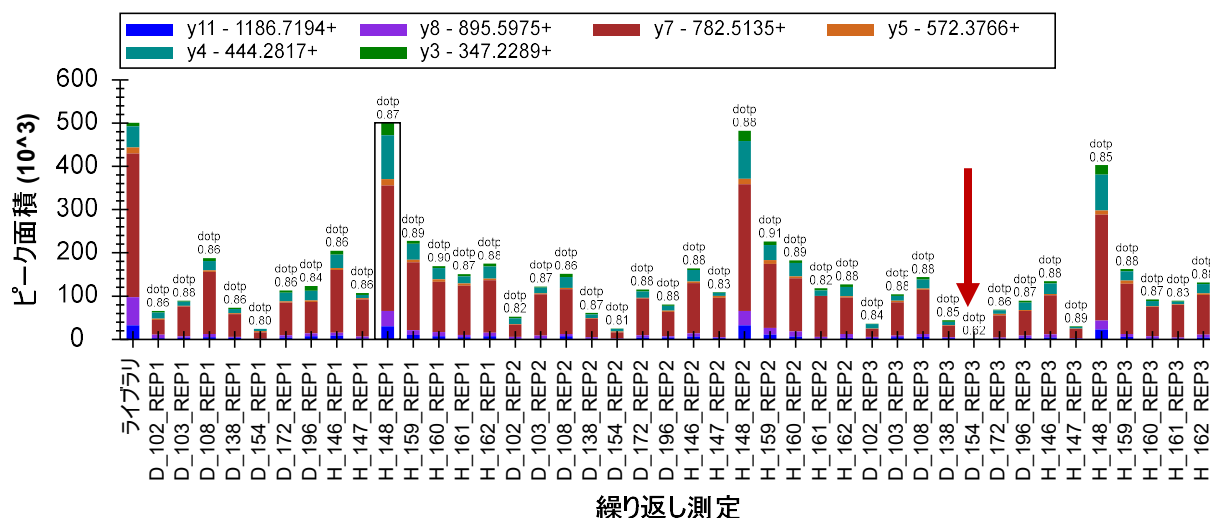
3つのプロットで見られるクロマトグラムのピークはペプチド量が十分な場合、23~24分付近に強度の高いy7プロダクトイオンピークが検出できかつ、他の5つのトランジションもすべて共溶出していること（6つのプロダクトイオンがすべて同じプリカーサー由来であること）が確認できます。



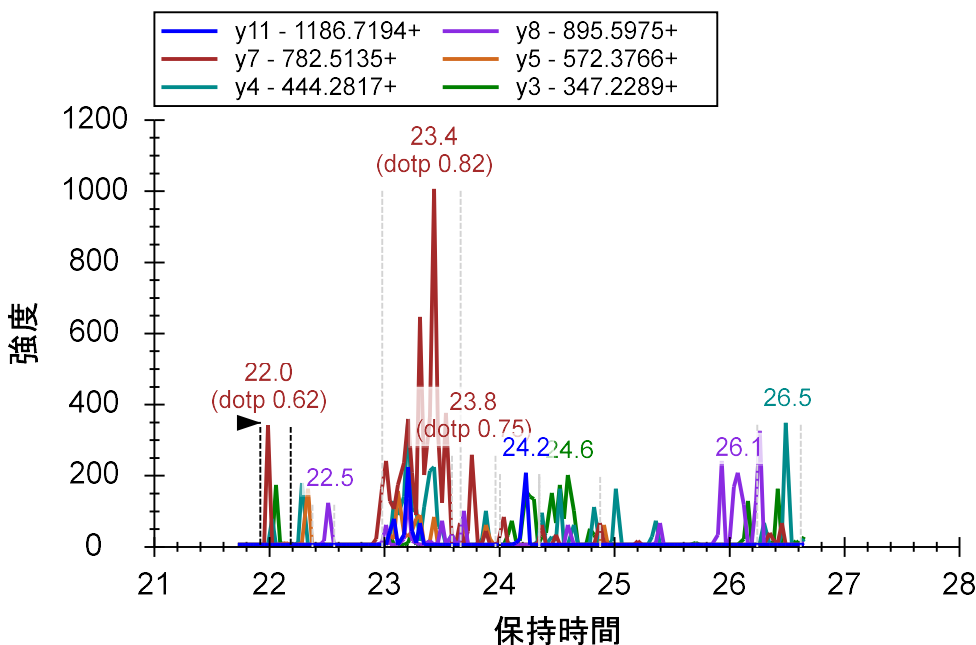
この場合消去するより誤認識されたピークをデータ修正する方が良いでしょう。これは[保持時間]プロットの22.5分未満に現れる短いバンドとして確認できます。



[ピーク面積]プロットを見てもこの測定のシグナル強度は低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、この測定のクロマトグラムを有効化します。この測定ではy7プロダクトイオンの強度が非常に低いことが分かります。その前後の測定データを確認せずにトランジションリストのプロダクトイオンとして選択することはまずないでしょう。



この測定のピーク積分を修正するため以下の操作を行います。

- 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、プロダクトイオンの相対量を示す [ピーク面積] グラフに戻ります。

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

修正されたピークはサマリープロット内のその他の測定と似通っ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 GMYESLPVVAVK ではサマリープロットに一貫性がないので、このペプチドは消去してピーク積分の修正は行いません。ペプチド ETGLMAFTNLK は1つの測定だけ調整が必要です。この測定は [ピーク面積] グラフを見ればどれかすぐわかります。ここまでの学習で何が間違いで何をどのように修正すべきか理解が深まったことと思います。

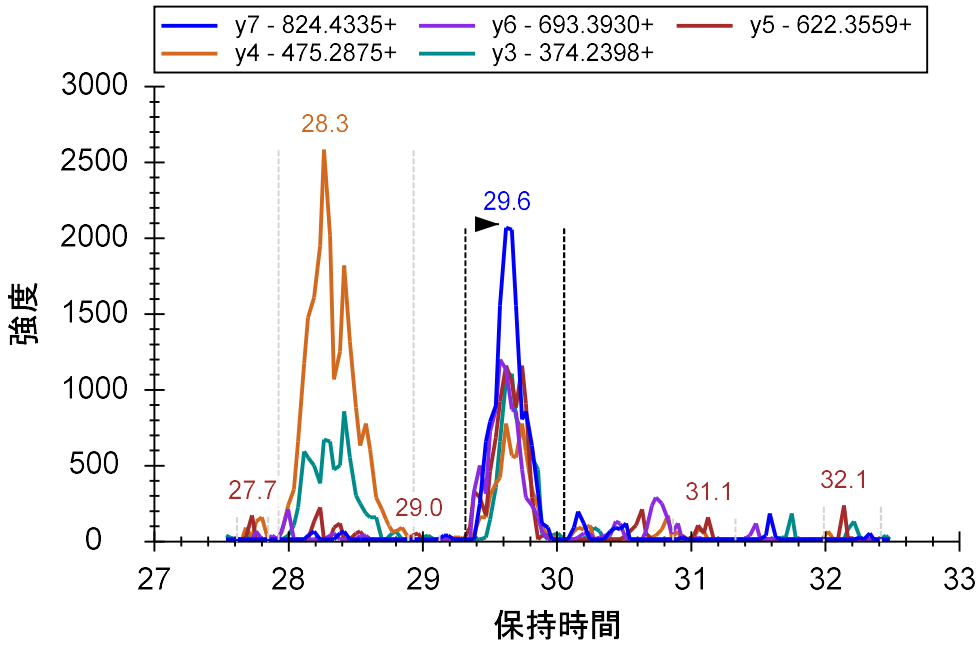
この修正を行った後でも、ETGLMAFTNLK ペプチドのすべてのケースにおいて同一ペプチドの同一ピークが積分されていると確信できない場合、以下のいずれかの操作を行って、すべてのクロマトグラムを迅速に再確認できます。

- [ターゲット] ビューの上部にある [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。
- 上向きおよび下向き矢印キーを使用して測定を横送りし、クロマトグラムを有効化します。

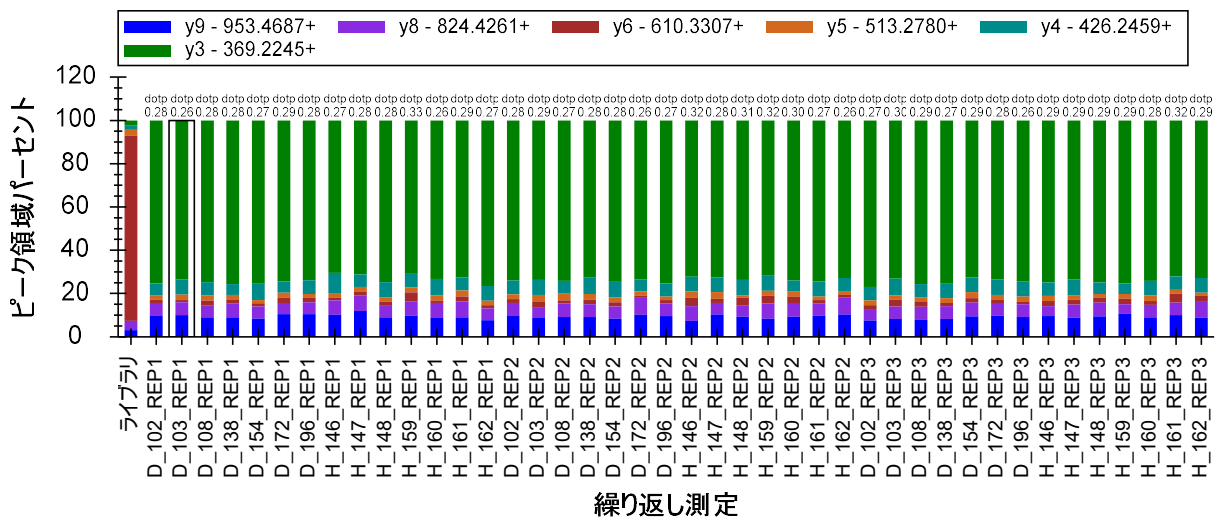
あるいは

- Ctrl+上向き矢印または Ctrl+下向き矢印を押して測定を横送りします。

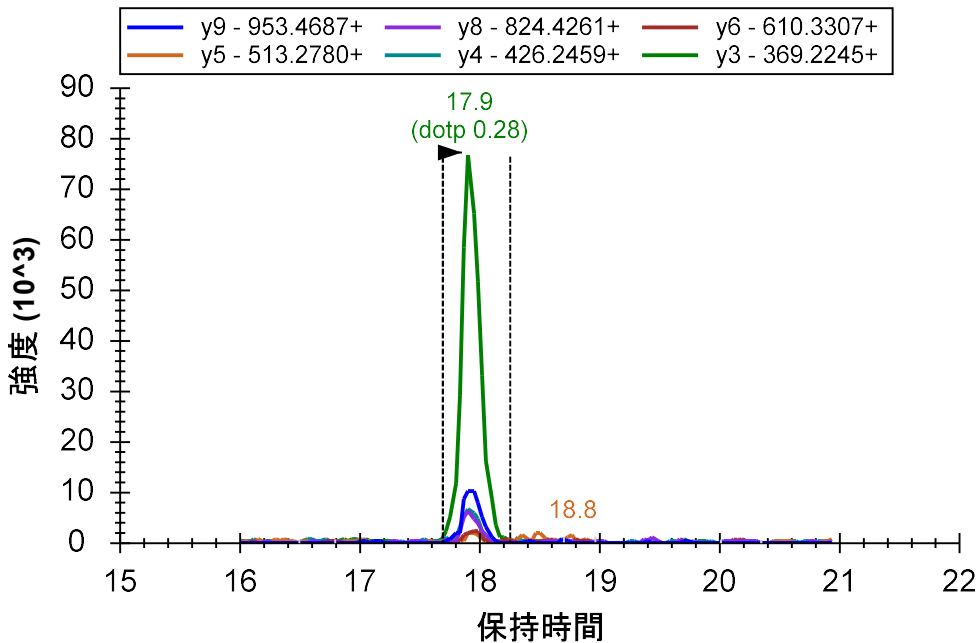
多くのクロマトグラムで選択したピークの約 1.5 分前の y3 および y4 プロダクトイオンのピークがある (干渉) のが分かります。ピーク積分の一貫性がここで認識できます。



[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) で測定ピークの相対イオン存在比とライブラリスペクトルとの違いが明白となっていることがわかります。これらの差は [ピーク面積] 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ます。Skyline 上ではピーク保持時間の下に低いドット積値「(dotp 0.28)」が表示されます。



同じ保持時間上で強度の高い別のピークがある場合は、このように「dotp」値が表示されます。上記のケースでは dotp が表示されているピーク付近にはノイズレベルの低いピークしか見られません。そこで他のピークを解析するため以下の操作を行います。

- 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скар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てズームします。

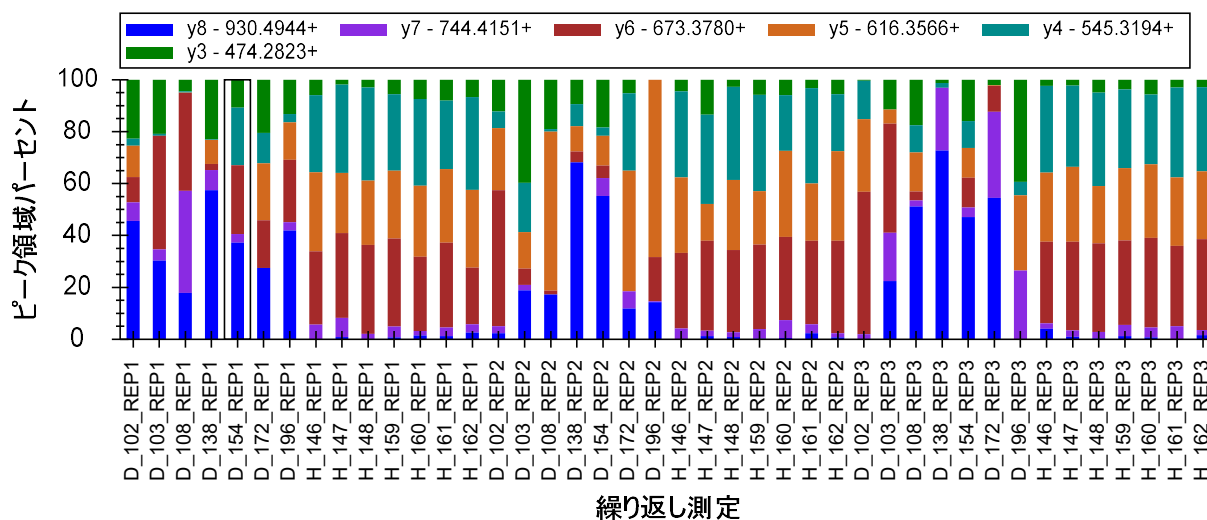
より良い dotp 値を持つ小さいピークが見えるようにグラフの y スケールをズーム調整します。

本実験で使用した質量分析計装置がライブラリスペクトルで使用された装置を類似もしくは同じであれば、このピーク積分はターゲットのペプチドとは異なるものだったのではないかと思います。しかしこの場合このスペクトルの詳細は不明で他にオプションはありません。

[ピーク面積] グラフおよび [保持時間] グラフにおける測定間の一貫性から、6 つのプロダクトイオンが共溶出していて同一ペプチド由来であることがわかります。最も強度の高いイオンは y3 プロダクトイオンですが選択性の低いいため、削除して他のプロダクトイオン (y9 イオンは 2 番目に強度が高いプロダクトイオン) についてより詳しく分析することが重要です。一方でもしの y3 ピークの方がやはり適していると確信が持てた場合は、[元に戻す] ボタン (Ctrl+Z) を使用し削除された y3 プロダクトイオンを Skyline ドキュメントへ戻せます。

TDEDVPSGPPR の上の 7 つのペプチドの中では、ペプチド SPQGLGASTAEISAR について誤ってピーク積分されたピークが 1 つだけ見つかります。ここでは問題が簡単に見つけられ修正できるはずです。

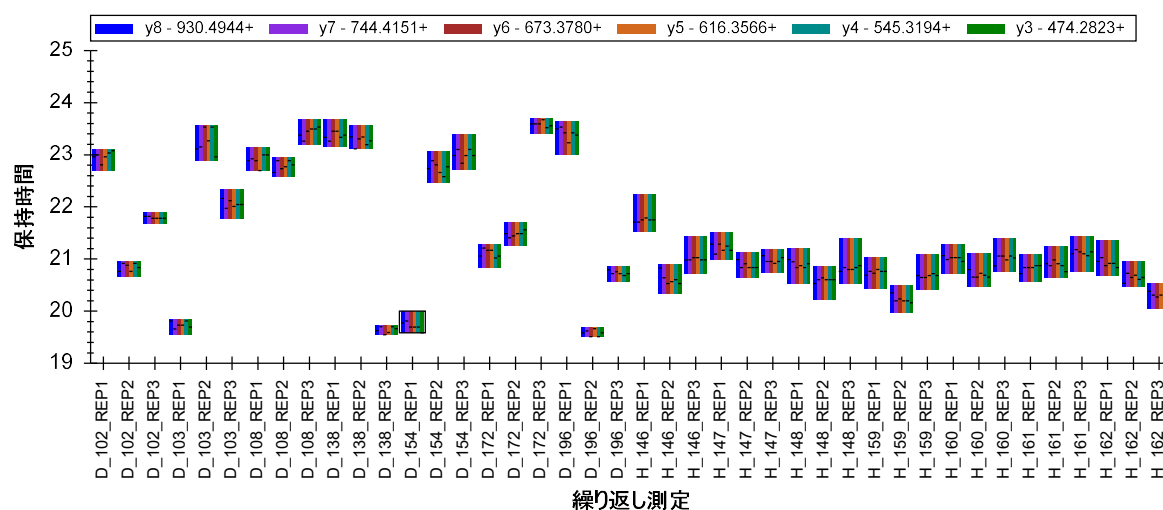
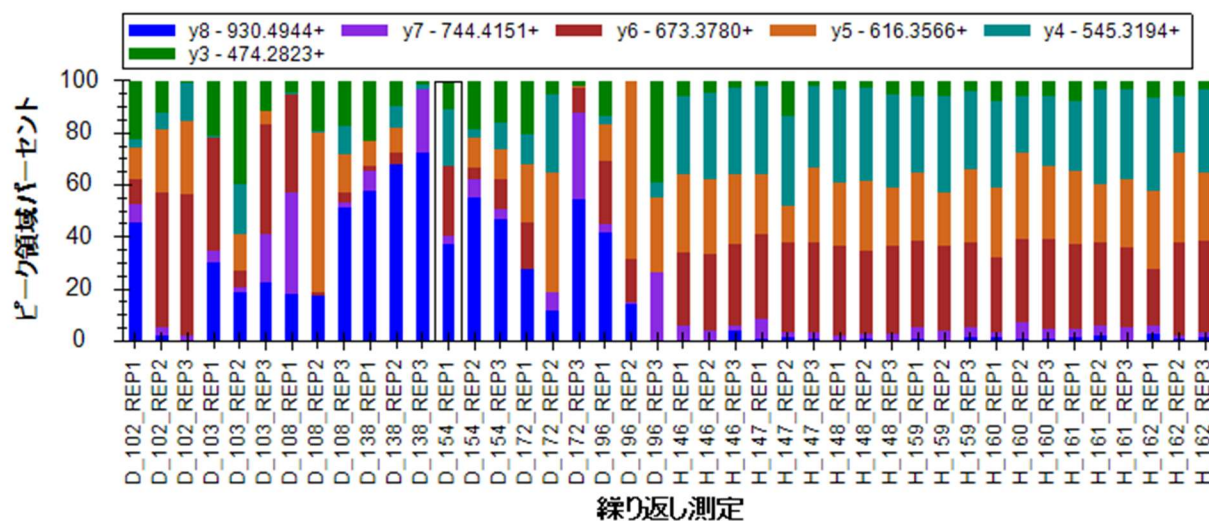
ペプチド CSSLLWAGAAWLR へと移ります。サマリープロットで観察されるブレを見るとこのペプチドは削除した方がよいと考えるかもしれません。



しかし健常検体群に注目してみると実は一貫した部分があります。これを知るべく以下の操作を行います。

- [ピーク面積] グラフを右クリックし、[並べ順] を選択して [ドキュメント] をクリックします。

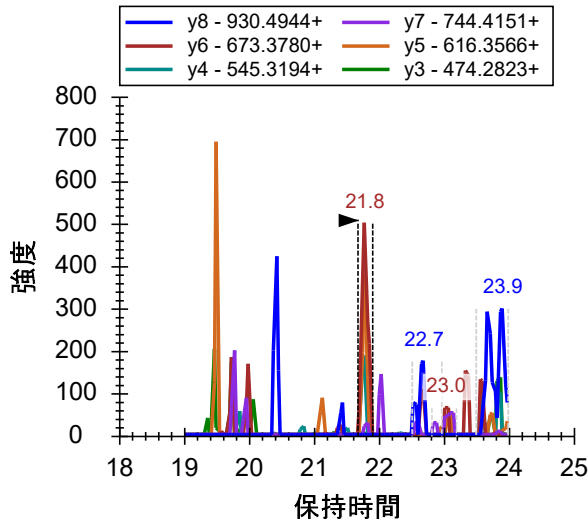
そこでグラフは以下のように表示されます。



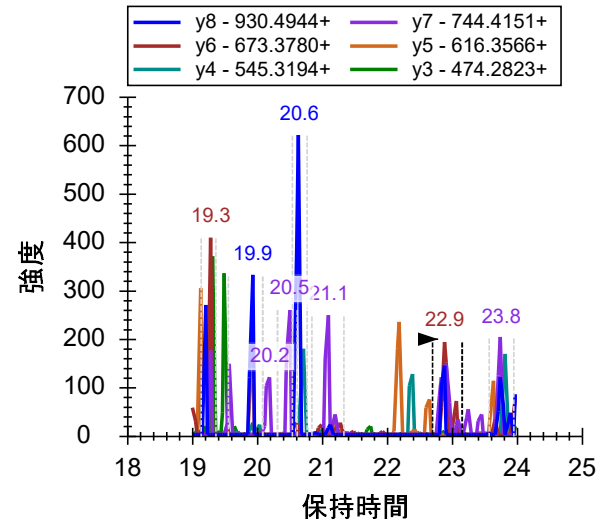
健常群はピーク積分が安定していますが、疾患群についてはそうではないことが明確になります。健常群のピークを再確認すると、強度はかなり低いので y4、y5、および y6 イオンの共溶出が関係しているのが分かります。この共溶出はおそらく、1つのペプチドに起因しています。しかし、このペプチドが実際に CSSLLWAGAAWLR であるかどうかを判定することは困難です。

疾患群のクロマトグラムを見ると、y6 プロダクトイオンが溶出されている 21 分付近にピークと思われるものがいくつか見られます。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。

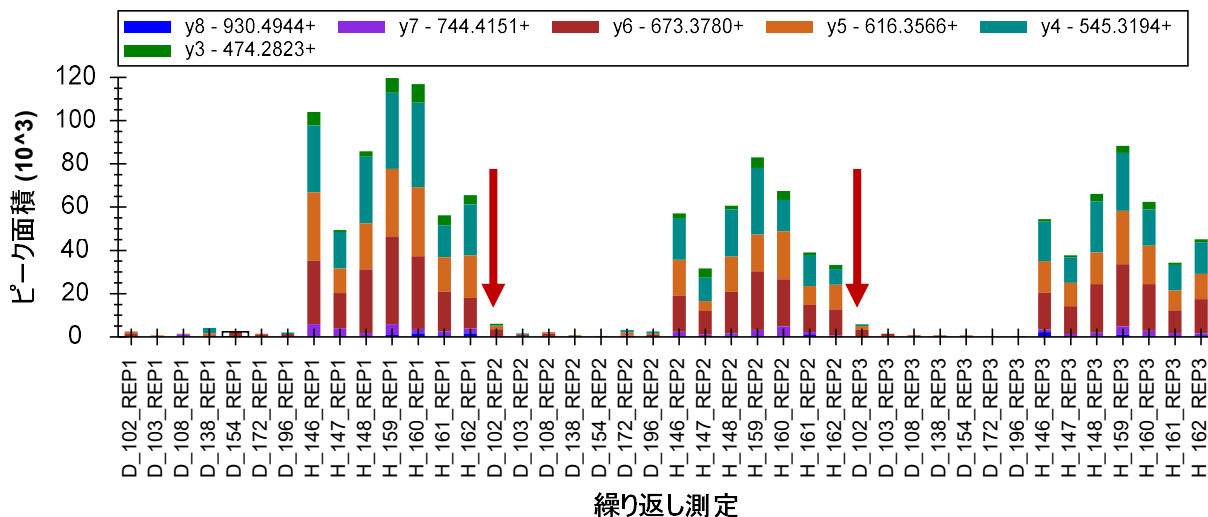
D_102_REP3



D_108_REP1



しかし、疾患検体のクロマトグラム内には、健常検体で観察されたのと同じようなピークは見つかりません。これまでにピークドリフトや分断ピークの操作で得てきた知識、経験からすれば、ピークがすべての疾患群で測定範囲外にシフトしたわけではないと確信できます。こうした問題はランダムにサンプルを測定する（測定順をランダム化する）ことで解決できる可能性もあります。しかし今回の研究では、3 回繰り返し測定サイクルで十分です。可能な限りピーク積分を修正する段階で健常群の測定の直後に続く疾患群の測定 2 つで、ピークが多くみられます（D_102_REP2 および D_102_REP3）。これはキャリーオーバーである可能性があります。



これらのトランジションが本研究におけるバイオマーカーの最有力候補の 1 つの可能性が出てきました。そこで合成ペプチドを発注して使用して **CSLLWAGAAWLR** ペプチドの測定の精度は向上させていきます。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、健常群で見られるピークがどのようなペプチドか確認するため、MS/MS スペクトル上で *de novo* シーケンシングを行うこともできます。ターゲットプロテオミクスでは、必ずしも標的ペプチドの同定作業から開始しなければならないわけではありません。SDS-PAGE ゲルでの古典的解析のように差異を見つけてから、その差異について解析を進めることもできます。

ここまでピーク積分を修正したり解析に適さないペプチドを削除したことで、他の残りのペプチドについても処理していくことができるかと思います。この操作は 1 時間程度でできるでしょう。ピーク形状が良好で保持時間ウィンドウ内に溶出されたペプチドは、ほとんど修正する必要がないということも確認できます。これは多くの場合[**ピーク面積**] グラフおよび[**保持時間**] グラフを一目見れば分かります。ピーク形状が悪い、または多くのピークがスケジュールウィンドウ内でピークが分断されているようなペプチドについては、そのペプチドを削除すればいいと思うかもしれません。しかし興味深いデータがある可能性もあるため即断しない方がよいでしょう。

測定データの統計解析

Skyline のピーク積分についての問題を注意深く検証し修正した後は、実験条件間でペプチド量にどのような差異があるかを知りたいと思われるかもしれません。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ため Skyline で通常「繰り返し測定」とされている、測定サンプルを更に分類する操作が必要です。当該の分類に向け、Skyline は繰り返し測定注釈を用意しています。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 注釈を定義するため以下の手順を実施します。

- [**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**] タブで[**リストを編集**] ボタンをクリックします。
- [**注釈を定義**] フォームで[**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「SubjectId」と入力します。
- [**適用先**] 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。

フォームは以下ようになります。

注釈を定義

名前(N):
SubjectId

編集可能 計算済み

タイプ(T):
テキスト

値(V):

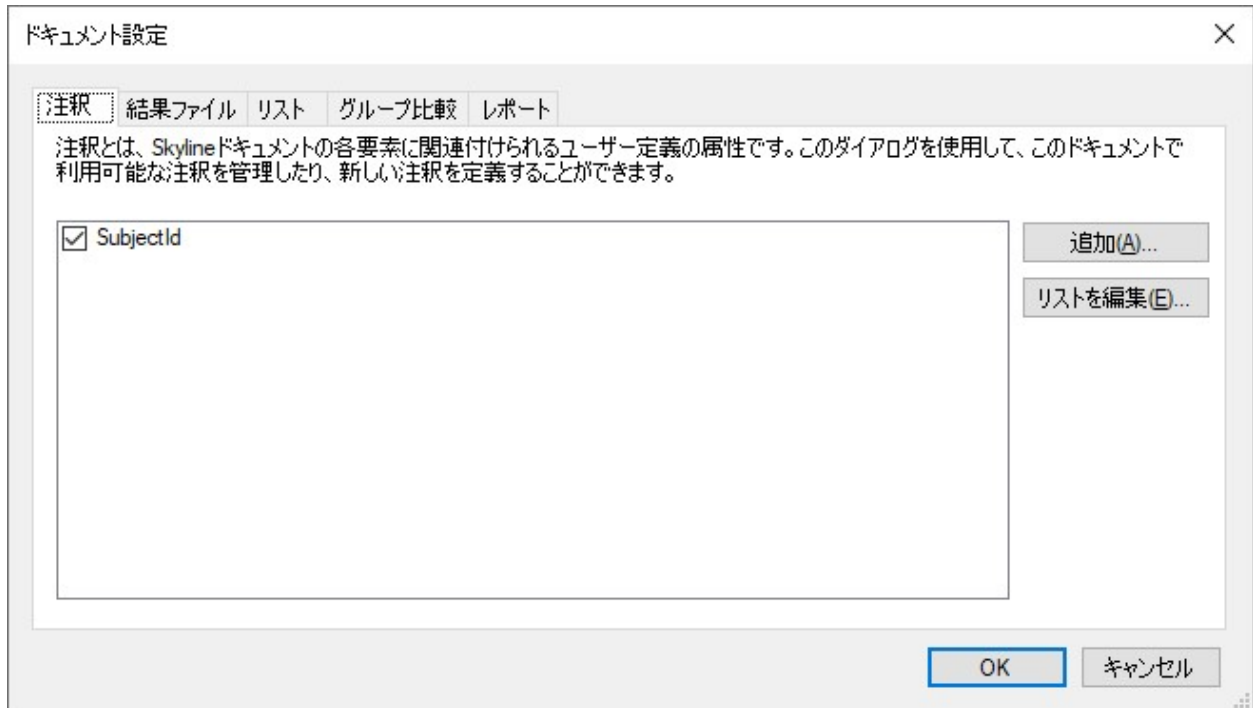
適用先(A):

- 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] フォームで [OK] ボタンをクリックします。
-
- 作成した「SubjectId」注釈のチェックボックスをオンにします。

[ドキュメント設定] フォームは以下のようになります。



- [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までは詳しく取り上げません。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 の使い方を説明する際に利用されています。MSstats で一体何ができるか知るべく以下のように Skyline にインストールすると 2 つの異なる注釈が得られます。

- [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ります。



- [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よう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、または Msstats をインストールせずに注釈を直接追加する場合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] の [注釈] タブで [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 (あるいは [追加] ボタンをクリックします)。
- Condition 注釈について同様の作業を繰り返します。

「注釈を定義」はそれぞれ以下のように表示されます。

The image shows two side-by-side dialog boxes titled 「注釈を定義」 (Define Annotation).
The left dialog box has the following fields:
- 名前(N): BioReplicate
- タイプ(T): テキスト (Text)
- 値(V): (empty text box)
- 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
- Buttons: OK, キャンセル

The right dialog box has the following fields:
- 名前(N): Condition
- タイプ(T): 値のリスト (Value List)
- 値(V): Healthy, Diseased
- 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
- Buttons: OK, キャンセル

- [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 注釈のチェックボックスをオンにして、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なります。

The image shows the 「ドキュメント設定」 (Document Settings) dialog box. It has tabs for 「注釈」 (Annotations), 「結果ファイル」 (Result Files), 「リスト」 (List), 「グループ比較」 (Group Comparison), and 「レポート」 (Report). The 「注釈」 tab is selected.
The text below the tabs reads: 「注釈とは、Skylineドキュメントの各要素に関連付けられるユーザー定義の属性です。このダイアログを使用して、このドキュメントで利用可能な注釈を管理したり、新しい注釈を定義することができます。」
Below the text is a list of annotations with checkboxes:
- SubjectId
- BioReplicate
- Condition
- Buttons: 追加(A)... (Add...), リストを編集(E)... (Edit List...)
- Bottom buttons: OK, キャンセル

- [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、以下の操作を行います。

- [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]の左上隅にある[レポート]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]は以下のように表示されます。

	繰り返し測定	試料タイプ	アナライタ濃度	SubjectId	BioReplicate	Condition
▶	D_102_REP1	不明				
	D_102_REP2	不明				
	D_102_REP3	不明				
	D_103_REP1	不明				
	D_103_REP2	不明				
	D_103_REP3	不明				
	D_108_REP1	不明				
	D_108_REP2	不明				
	D_108_REP3	不明				
	D_138_REP1	不明				


これで 42 個の繰り返し測定すべての注釈をこのグリッドに手動入力できます。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。これを行うため以下の操作を行います。

- Excel で「GroupedStudies1\Heart Failure\raw」フォルダ内の「Annotations.xlsx」ファイルを開きます。
- 3 つの注釈列に対してヘッダ行の下 42 行の値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：このステップ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]は以下のように表示されます。

ドキュメントグリッド: 繰り返し測定

レポート ▾ | 1 of 42 | エクスポート... 操作 ▾ | 検索: 

	繰り返し測定	試料タイプ	アナライト濃度	SubjectId	BioReplicate	Condition
▶	D_102_REP1	不明				
	D_102_REP2	不明				
	D_102_REP3	不明				
	D_103_REP1	不明				
	D_103_REP2	不明				
	D_103_REP3	不明				
	D_108_REP1	不明				
	D_108_REP2	不明				
	D_108_REP3	不明				
	D_138_REP1	不明				

- Ctrl+V を押して [ドキュメントグリッド] へ貼り付けます。

スプレッドシートの値が [ドキュメントグリッド] 内のセルに以下のように入力されます。

ドキュメントグリッド: 繰り返し測定

レポート ▾ | 42 of 42 | エクスポート... 操作 ▾ | 検索: 

	繰り返し測定	試料タイプ	アナライト濃度	SubjectId	BioReplicate	Condition
	H_160_REP1	不明		H160	160	Healthy
	H_160_REP2	不明		H160	160	Healthy
	H_160_REP3	不明		H160	160	Healthy
	H_161_REP1	不明		H161	161	Healthy
	H_161_REP2	不明		H616	161	Healthy
	H_161_REP3	不明		H161	161	Healthy
	H_162_REP1	不明		H162	162	Healthy
	H_162_REP2	不明		H162	162	Healthy
▶	H_162_REP3	不明		H162	162	Healthy

ピーク形状が悪いペプチドへの注釈

削除されたピークまたは分断されたピークのあるペプチドには注釈を付けた方がよいでしょう。なぜなら、どちらの場合も統計分析中に問題とな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ため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]タブで[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」と入力します。
- [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] チェックボックスをオンにします。

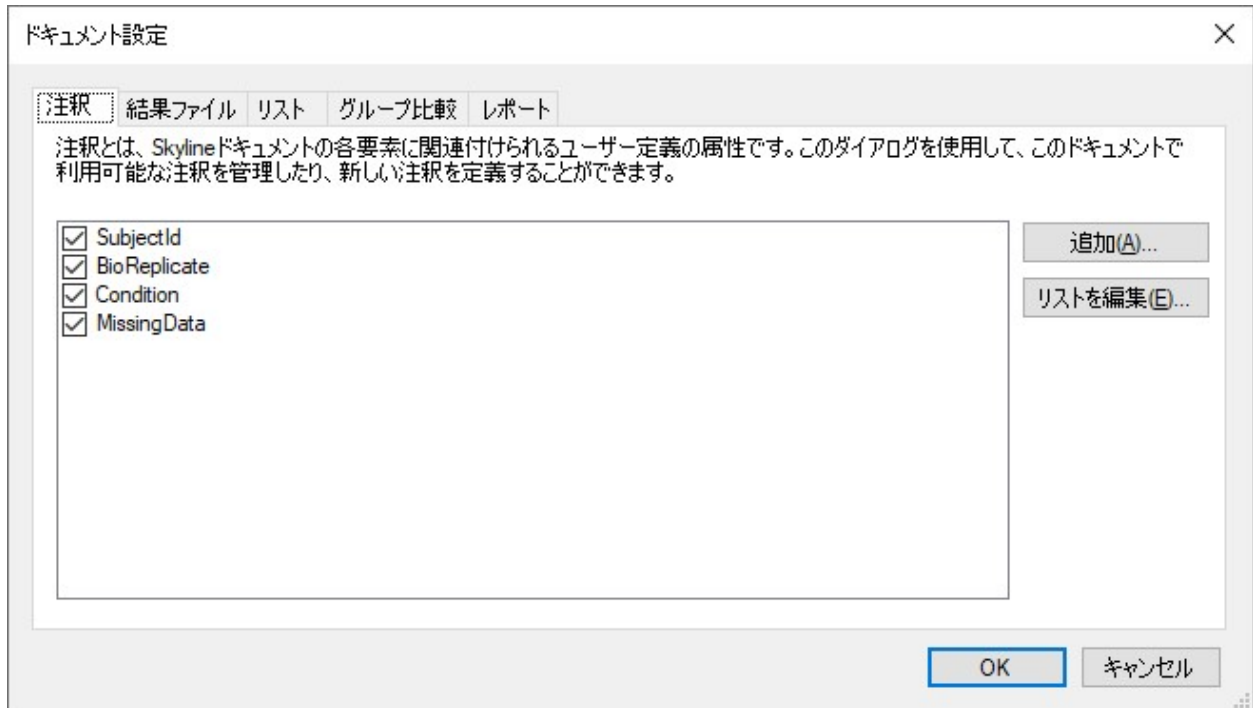
[注釈を定義] フォームは以下のように表示されます。

The screenshot shows a dialog box titled "注釈を定義" (Define Annotation). It contains the following fields and controls:

- 名前(N):** A text input field containing "MissingData".
- 編集可能** and **計算済み**: Two buttons located below the name field.
- タイプ(T):** A dropdown menu currently showing "True/False".
- 値(V):** An empty text input field.
- 適用先(A):** A list of checkboxes:
 - 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** and **キャンセル**: Two buttons at the bottom of the dialog.

- [OK] ボタンをクリックします。
- [注釈を定義] フォームで[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表示されます。

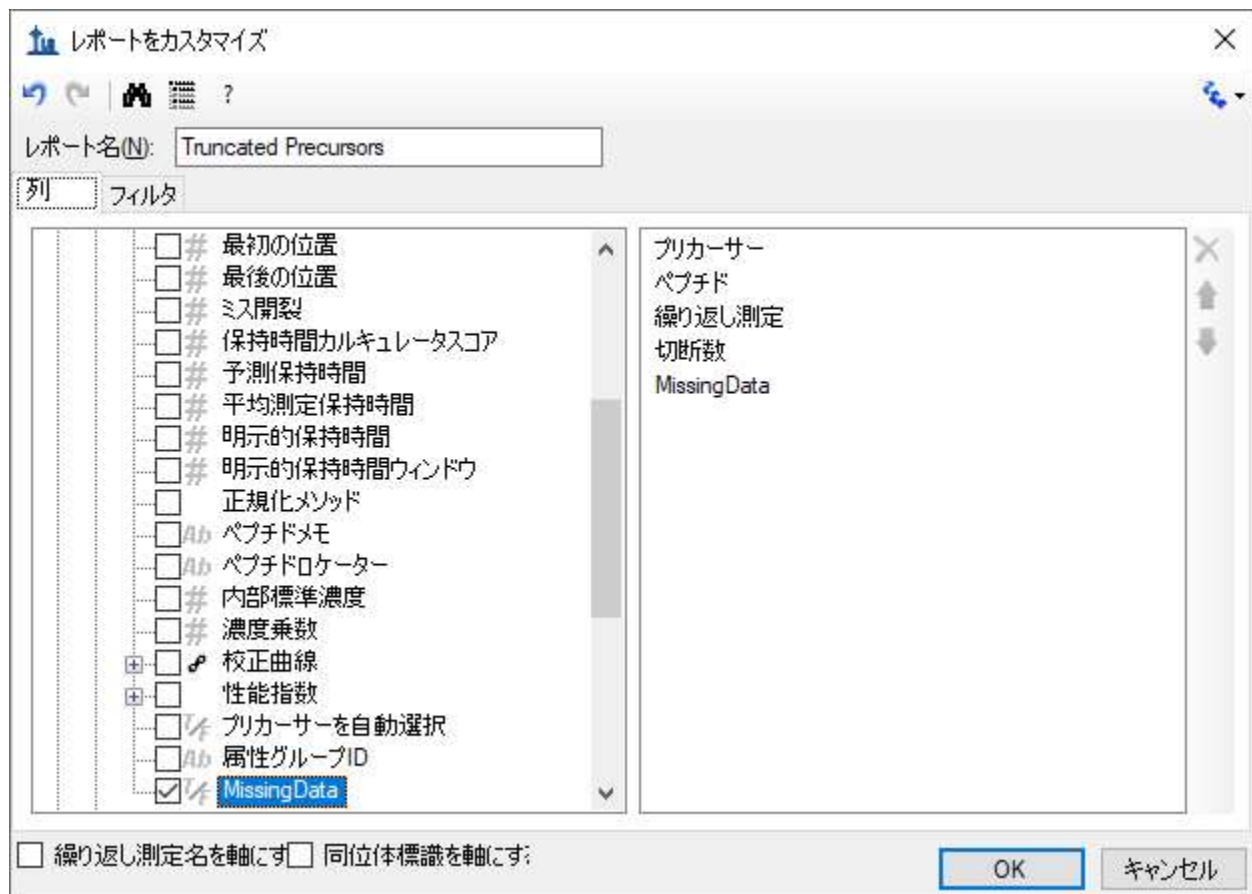


- [OK] ボタンをクリックします。

分断されたピークを示すすべてのペプチドに新しい「MissingData」注釈を設定するため以下の操作を行います。

- [ドキュメントグリッド] の [レポート] メニューで、先に作成した「分断されたプリカーサーピーク」レポートをクリックします。
- 同じ [レポート] メニューで、[レポートを編集] をクリックします。
- [レポートをカスタマイズ] 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。

[レポートをカスタマイズ] フォームは以下のように表示されます。



- [OK] ボタンをクリックします。
- 「切断数」列ヘッダー、[降順]の順にクリックします。

[ドキュメントグリッド]は以下のように表示されます。



The screenshot shows a window titled 'ドキュメントグリッド: Truncated Precursors'. The window contains a table with the following columns: 'プリカーサー', 'ペプチド', '繰り返し測定', '切断数', and 'MissingData'. The table lists several entries, with the first one highlighted in blue. The 'MissingData' column contains checkboxes, all of which are currently unchecked.

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

残りのドキュメントの処理方法によって分断されたプリカーサーピークグループの数は 223 個（測定サンプルの合計数）と同じになったり、ならなかったりする場合があります。ペプチドの「MissingData」チェックボックスは 1 個ずつオンにすることが可能です。任意ペプチドのボックスをオンにすると、残りは自動的にオンになります。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]は以下のように表示されます。

ドキュメントグリッド: Truncated Precursors

レポート ▾ | 2 of 222 | エクスポート... 操作 ▾ | 検索:

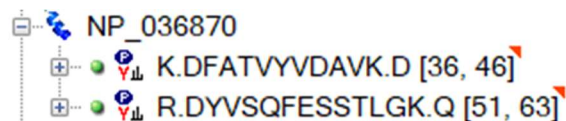
	プリカーサー	ペプチド	繰り返し測定	切断数	MissingData
	614.3164++	DFATVYVDAVK	D_108_REP2	7	<input checked="" type="checkbox"/>
▶	614.3164++	DFATVYVDAVK	D_196_REP3	7	<input checked="" type="checkbox"/>
	614.3164++	DFATVYVDAVK	H_147_REP2	7	<input checked="" type="checkbox"/>
	614.3164++	DFATVYVDAVK	H_148_REP2	7	<input checked=""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

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 1 回の貼り付けで 163 行すべてをまとめて設定できます。

- Excel の空白のスプレッドシート内で、163 行目までスクロールします。
- 163 行目の最初の列に「TRUE」と入力します。
- 今入力したセルを選択します。
- 選択長方形の右下角にある範囲を拡大する四角をクリック&ドラッグして、163 行目から 1 行目までの最初の列のセルに値「TRUE」を入力します。
- Ctrl+C を押してセルをコピーします。
- Skyline に戻り、「MissingData」列内の最初のセルを選択します。
- 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思われるかもしれませんが、各ペプチドについて、MissingData 注釈を TRUE に設定するのは 1 回だけです。しかし上記を行うと、分断されたピークを持つ 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簡単に処理できます。

上記手順を実行すると、[ターゲット] リスト上位の多くのペプチドの名前の右上に小さな赤い三角形マークが付きます。

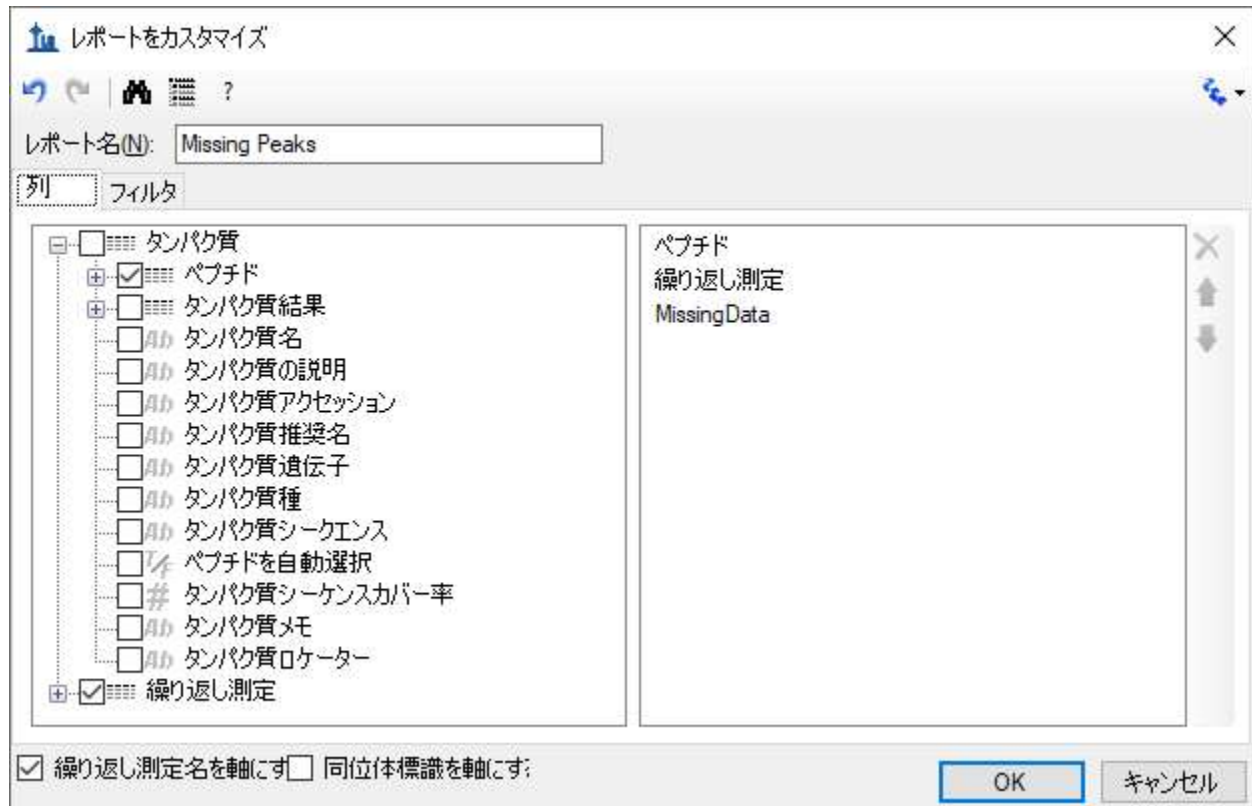


これらの三角形の上にマウスをポイントすると、テキスト「Missing Data True」とともに説明が表示されます。

ここでは分断されたピークは積分せず、以下の操作で完全に削除したすべてのペプチドについて処理を繰り返します。

- [ドキュメントグリッド]の[レポート]メニューで、[レポートを管理]をクリックします。
- これまで使用していた「分断されたプリカーサーイオンピーク」レポートを選択します。
- [コピー]ボタンをクリックします。
- [レポートをカスタマイズ]フォームの[レポート名]フィールドに「Missing Peaks」と入力します。
- [プリカーサー]列および[分断ピーク数]列を削除するには、列リストの右側の✕ボタンを使用します。
- [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。

[レポートをカスタマイズ]フォームは以下のように表示されます。



- [フィルタ]タブをクリックします。
- 次に✕ボタンをクリックして、[分断ピーク数]の既存のフィルタを削除します。
- 列ツリーを展開して、タンパク質>ペプチド>プリカーサー>プリカーサー結果をチェックした後、[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。

[レポートをカスタマイズ] フォームは以下のように表示されます。



- [OK] ボタンをクリックします。
- [レポートを管理] フォームで [OK] ボタンをクリックします。
- [ドキュメントグリッド] の [レポート] メニューで、[Missing Peaks] をクリックします。

[ドキュメントグリッド] は以下のように表示されます。

ドキュメントグリッド: Missing Peaks

レポート ▼ | 1 of 10 | エクスポート... | 操作 ▼ | 検索:

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
GSYNLQDLLAQ...	<input checked="" type="checkbox"/>				D_103_REP3	D_108_REP2
DYVSQFESSTL...	<input checked="" type="checkbox"/>				D_103_REP3	D_108_REP2
TGTNLMDFLSR	<input checked="" type="checkbox"/>					D_108_REP2
GTITSIAALDDPK	<input checked="" type="checkbox"/>				D_103_REP3	
JFPENNIK	<input checked="" type="checkbox"/>				D_103_REP3	
ALIHCLHMS	<input checked="" type="checkbox"/>					
AGDQILAINEN...	<input checked="" type="checkbox"/>					
FAEDHFAHEATK	<input checked="" type="checkbox"/>					
TLNSINIAVFSK	<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 GSYNLQDLLAQ1 行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手順については、次の注で説明します。

このレポートで分断されたピークがある測定のパプチド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。またピークが分断されていないペプチドは 2 個のみで、8 個にはすでに「MissingData」注釈セットがあります。問題のあるピークを持つ繰り返し測定名は、[ペプチド] 列と [MissingData] 列の右側にそれぞれ表示されます。問題のあるペプチドすべてにデータ標識を付けるため以下の操作を行います。

- 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。データ品質に影響を与える問題をより明確に理解し、取得データに対して可能な限りピーク積分を最適化すれば、データ品質を評価するための高レベル統計処理の準備は完了です。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も理解ができていると思います。

注：チュートリアルは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、以下の手順で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。

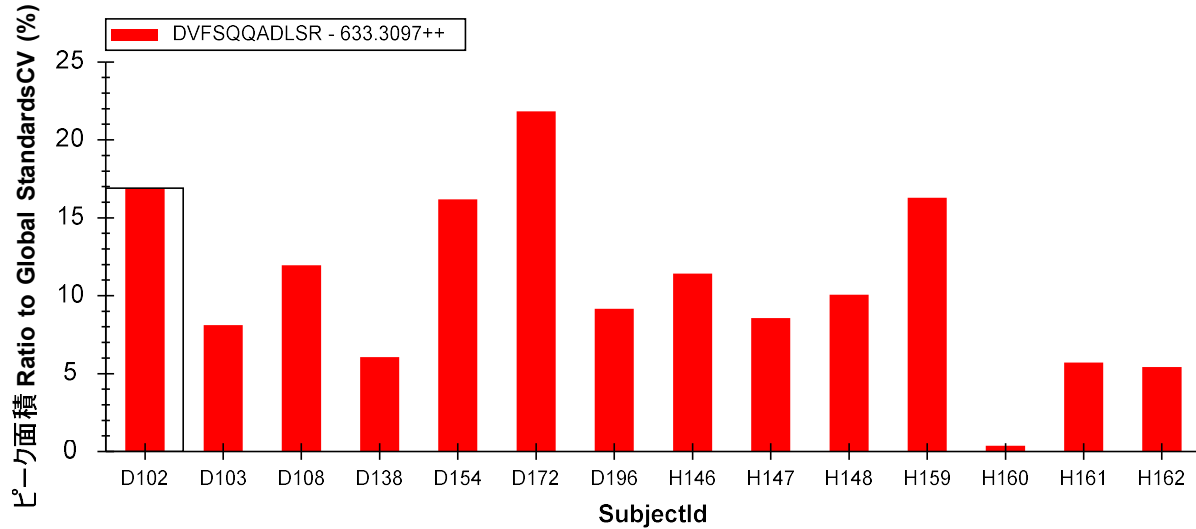
Skyline での複数の繰り返し測定検査

このようなデータセットの統計分析をより詳細に（Skyline で現在提供されている機能以上に）行うことももちろんできますが、変動やグループ平均の査定を行う有用な方法が Skyline にあります。各検体の技術的繰り返し測定間の変動について知るべく以下の操作を行います。

- 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。
- [ピーク面積] ビューを右クリックして [正規化] を選択し、[グローバル標準] をクリックします。
- [ピーク面積] ビューを右クリックして [グループ化] を選択し、[SubjectId] をクリックします。

- [ピーク面積]ビューを右クリックして、[CV値]がまだオンになっていない場合にはクリックしてオンにします。

、[ピーク面積]グラフは以下のように表示されます。

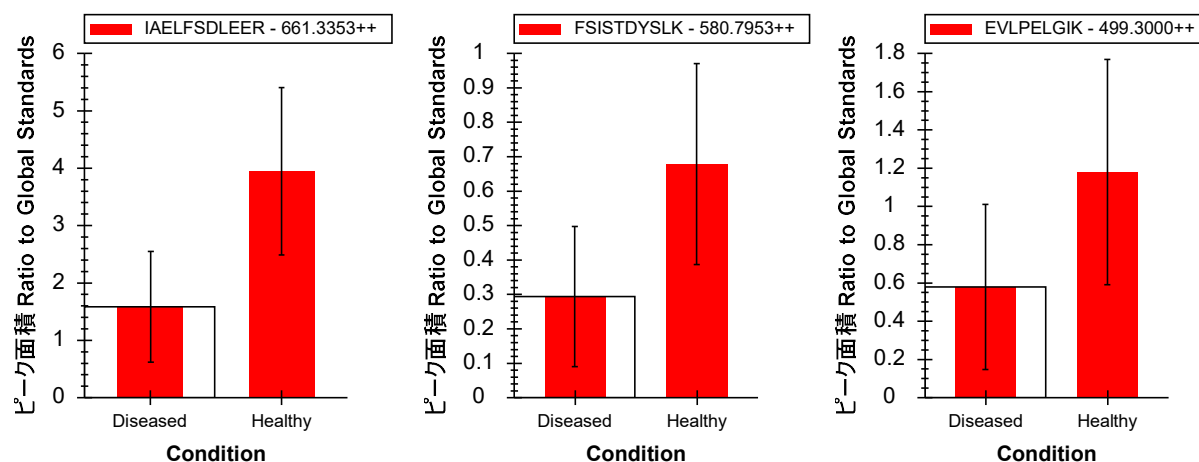


ここでさらに多くのペプチドを選択して、各サンプルの技術的繰り返し測定間の変動係数 (CV) を再確認できます。ほとんどが 25%未満であることが分かります。しかし理想的にはより低い CV が見られることが望ましいでしょう。3 個の測定値から得られた CV の方が 10 個の測定値から得られた CV より低いというのは注意すべきです。これは、サンプル数が少ない場合に標準偏差を過小評価するという統計の傾向を表しています。

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、すなわち健常群/疾患群間の平均ペプチドの差について確認をするため以下の手順を実行します。

- [ピーク面積]ビューを右クリックして[グループ化]を選択し、[Condition]をクリックします。
- [ピーク面積]ビューを右クリックして[CV値]をクリックし、このメニュー項目のチェックボックスをオフにします。
- [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 のペプチドに対し、以下のような [ピーク領域] のグラフが表示されます。



これらのグラフはすべての繰り返し測定の平均値で（この場合は正規化のための標準ペプチドに対するピーク面積比）、エラーバーは標準偏差となります。これによりサンプルの分布について理解できます。

グラフの解釈には、そもそもサンプル群を比較する目的を理解しなくてはなりません。典型的な目的は以下の2つです。

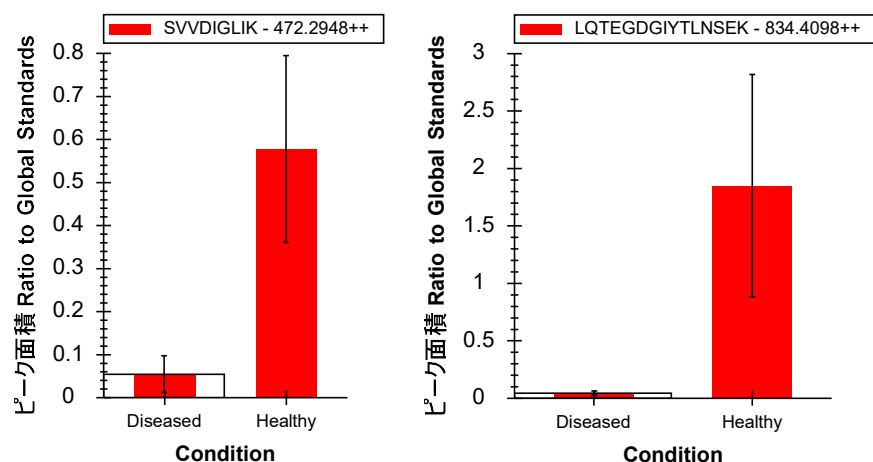
1. 差分：グループ間の差を同定する。（例：処理によって興味ある経路（例えばシグナル伝達経路）の流れに沿って活性低下が生じるかどうかを調べる）
2. 予測：ある特定グループに属すると予測されるペプチドの発現量を同定する。（例：ある一連のペプチド群が特定グループのバイオマーカーとして使用可能かどうかをテストする）

単に 2 グループ間での統計学的有意差を調べたいのであれば、分布の標準偏差よりも平均値の標準誤差を算出するかと思います。t検定で得られる p 値は 2 つのグループの統計的有意差を示す数値の 1 つですが、上記グラフでは標準誤差についての情報はまったく得られないため、差分的発現に関する情報を得るにはあまり有用ではありません。

予測については 2 集団の分布が重要です。標準偏差が 2 グループ間でオーバーラップしている場合、このペプチドはおそらく統計的優位差はなくバイオマーカー候補にはならないと予想されます。どの測定によってばらつき（分布）が生じているか予測することは難しいですが、ペプチドの集団から「バイオマーカーパネル」を作成することは可能性です。

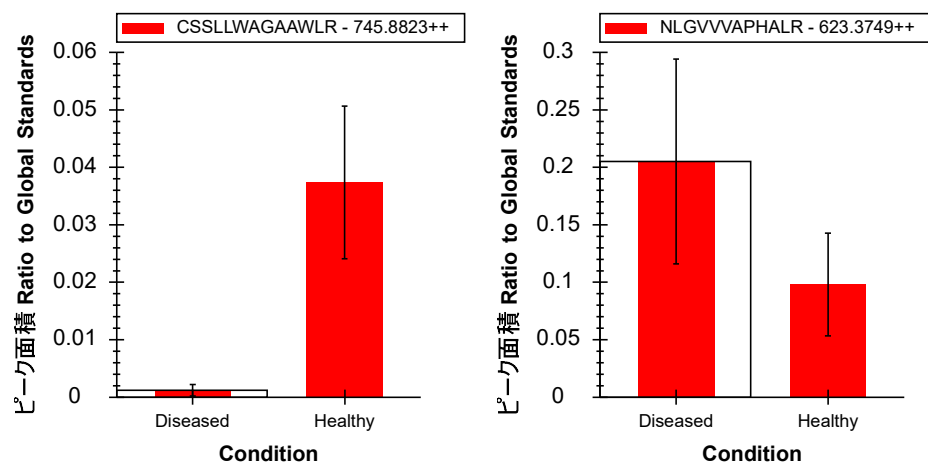
上記グラフ内のペプチドは実際には平均値に統計学的有意差がありますが、標準偏差の範囲がオーバーラップしているように見えるものがグラフでもあるため容易には予想できないことがあります。

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である NP_036714 はバイオマーカーとして使用できる可能性がありそうな有力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は、2 グループ間の各平均値における統計的有意差を示しています。このターゲットリストの文献が心疾患におけるタンパク質発現変化についての論文であったことを考えてみれば、これは驚くにあたりません。しかし有力なバイオマーカー候補と目されるものは実際にはかなり少数です。

タンパク質 NP_001007697 由来のペプチド CSSLWAGAAWLR を見ると、健常群と疾患群間でこのタンパク質由来のペプチドの相対発現レベルに顕著な違いが見られます。



このことから、多くの試料で同じペプチドの測定自体は信頼ができる一方、同一タンパク質に帰属する複数の異なるペプチドプリカーサイオンは安定していません。この特性が見られる理由は多数あります。例えば、ペプチドには翻訳後修飾 (PTM) されることがあり、それが重要なバイオマーカー候補となることもあります。その一方でタンパク質の他の領域は影響を受けない可能性があります。

Skyline でのグループ間比較解析

Skyline を利用してペプチドまたはタンパク質ピーク面積のグループ比較を容易に行うことができます。ペプチドまたはタンパク質の設定された SRM でのプロダクトイオンピーク面積の算出、正規化、ログの取得、技術的繰り返し測定の前平均化、t-検定を実施することで比較が可能です。Skyline は安定同位体ラベル標識ペプチドなし測定データの場合、数値がない測定データや分断されたピークデータ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ため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]タブで[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。
- [コントロールグループ]で「Healthy」を選択します。
- [比較対象]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「正規化のための標準ペプチドに対する比率」を選択します。
- [信頼度レベル]フィールドで、「99」と入力します。
- [適用対象]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表示されます。

グループ比較を編集

名前(N):
Healthy v. Diseased

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)
Healthy

次に対して比較する値(V):
Diseased

同一注釈(技術的繰り返し用)(D):
SubjectId

正規化メソッド(M):
グローバル標準に対する比率

信頼度レベル(E):
99 %

適用範囲
 ペプチド(P) タンパク質(T)

詳細 >>

OK キャンセル

- [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表示されます。

ドキュメント設定

注釈 結果ファイル リスト **グループ比較** レポート

グループ比較は、異なる繰り返し測定サブセットからの結果を相互比較するときに使用する設定を指定します。注釈は、どの繰り返し測定がどのサブセットに属するかを指定するのに使用します。

Healthy v. Diseased

追加(A)...

リストを編集(E)...

OK キャンセル

- [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調べるため以下の操作を行います。

- [ビュー] メニューで [グループ比較] を選択し [Health v. Diseased] チェックボックスをオンにします。

Skyline に以下のようなグリッドビューが表示されます。



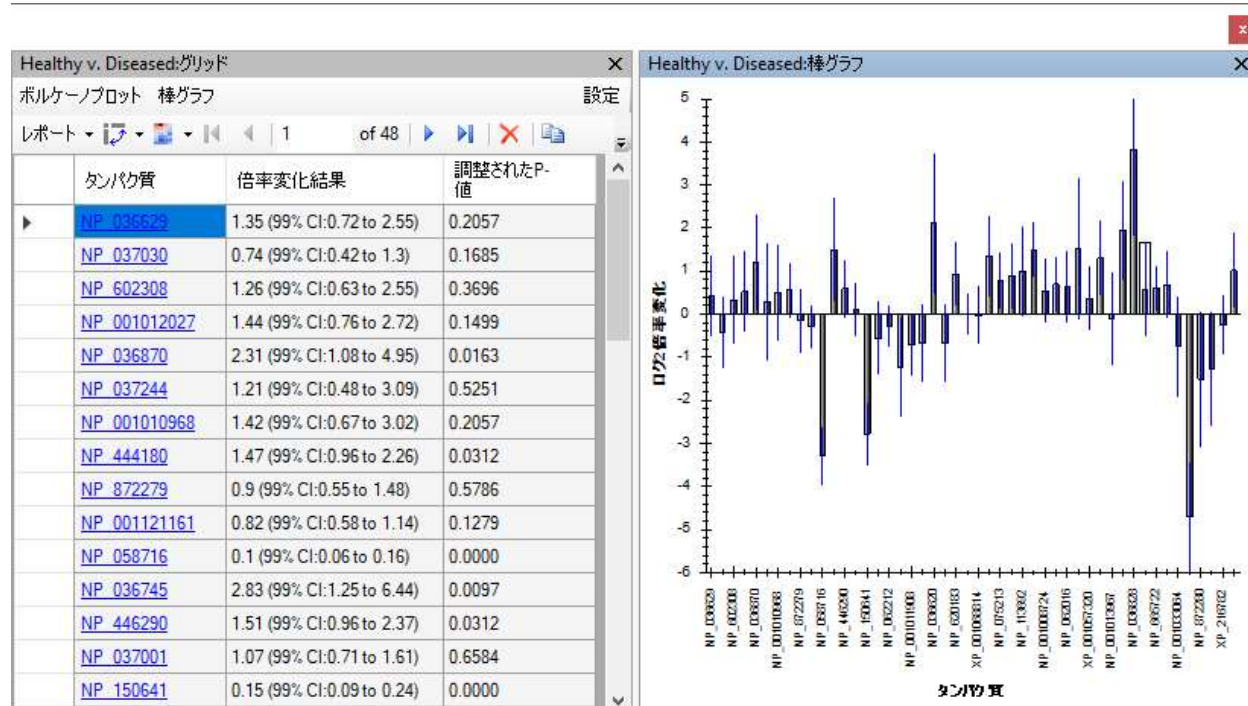
	タンパク質	倍率変化結果	調整されたP-値
▶	NP_036629	1.35 (99% CI:0.72 to 2.55)	0.2057
	NP_037030	0.74 (99% CI:0.42 to 1.3)	0.1685
	NP_602308	1.26 (99% CI:0.63 to 2.55)	0.3696
	NP_001012027	1.44 (99% CI:0.76 to 2.72)	0.1499
	NP_036870	2.31 (99% CI:1.08 to 4.95)	0.0163
	NP_037244	1.21 (99% CI:0.48 to 3.09)	0.5251
	NP_001010968	1.42 (99% CI:0.67 to 3.02)	0.2057

「比較群間変化の倍率」列に信頼区間が表示できていない場合、比較群間変化の倍率と P 値の両ヘッダー間の縦線をダブルクリックします。

ログ変換された比較群間変化の倍率と信頼区間グラフを見るため以下の操作を行います。

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

Skyline に以下のように [**Healthy v. Diseased : グリッド**] ビューの横にグラフパネルが追加されます。



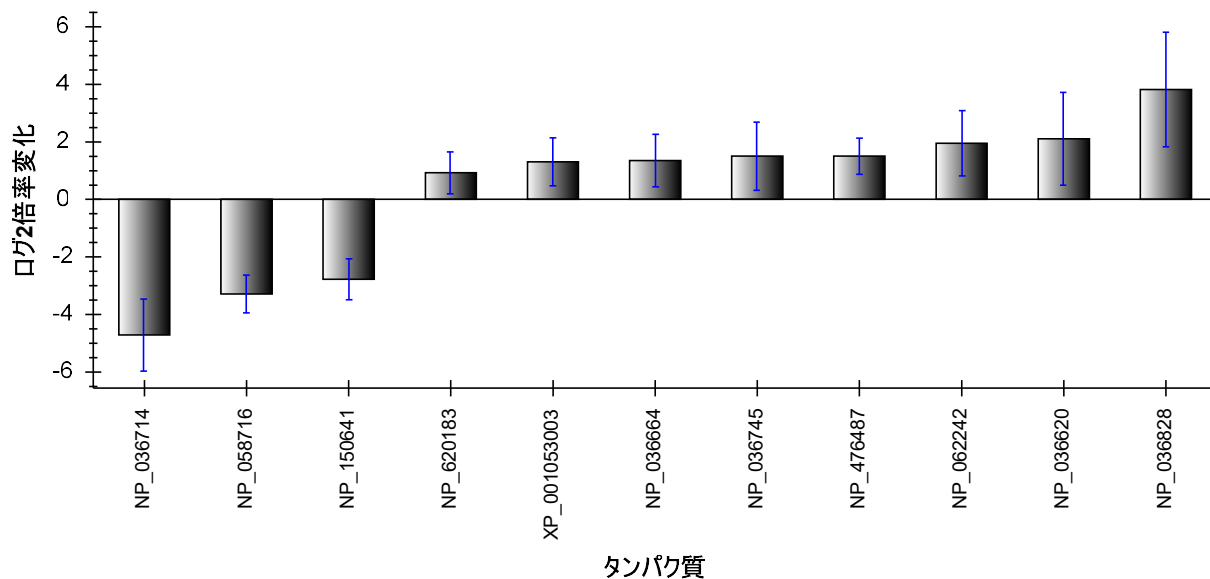
- 当該グリッドで「比較群間変化の倍率」ヘッダーをクリックして、表示されたメニュー内の [昇順に並べ替え] をクリックします。

これによりグリッド内とグラフのタンパク質が並べ替えられます。信頼区間を示すエラーバーの多くが x 軸を縦断していますが、99%の信頼度 (p 値 0.01) では生じうることに留意してください。また、複数の仮説を検証するための修正はまだ行っていません。

擬陽性、偽陰性 (FDR) 推定のため Benjamini-Hochberg 法で調整した p 値をカットオフ値に設定するためグリッドビューで以下の操作を行います。

- 「P 値」ヘッダー、[フィルタ] の順にクリックします。
- [フィルタタイプ] フィールドで「Is Less Than」を選択します。
- この下のフィールドに、「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に示す行数が 48 から 11 に減少し、当該グラフは以下ようになります。



ここで、有意の変化を示したタンパク質の多くには正の変化平均値、すなわち疾患群では高い値になることを意味します。ただし、試料順序のランダム化の失敗による影響を考慮することが重要です。測定の前 3 サイクルにおいて疾患検体は健常検体より先に測定していたため、正規化では考慮されなかった測定中の感度低下が補正されず、疾患群の方で値が上昇しているように見えてしまった可能性があります。

技術的繰り返し測定の指定

本実験のように技術的繰り返し測定データを収集する場合、測定データを [グループ比較を編集] フォーム内で正しく指定することが推計統計学的に極めて重要となります。これを行わないと各測定が生物学的に異なる検体と見なされて標準誤差や信頼区間に影響し p 値の低下につながります。

これを実際に確認するため以下の操作を行います。

- 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。

グリッドビューには 37 行 (及び 37 本のバーが示されたグラフ) が表示され、p 値が 0.01 未満になっています。しかしこれは 14 種類の生物学的検体の 3 回の技術的繰り返し測定とはならず、42 種類の生物学的に異なる検体と見なされています。これは明らかに間違いです。ご覧のように、ここで間違いを犯すと極度に楽観的な統計推論が得られてしまう場合さえあります。この場合健常群/疾患群間で平均値に統計学的有意差のあるのは 11 のタンパク質なのに推

定 1% FDR では一定濃度で注入された「S」タンパク質由来の 3 つの標準ペプチドさえ含む 37 個のタンパク質が有意差ありになってしまいます。

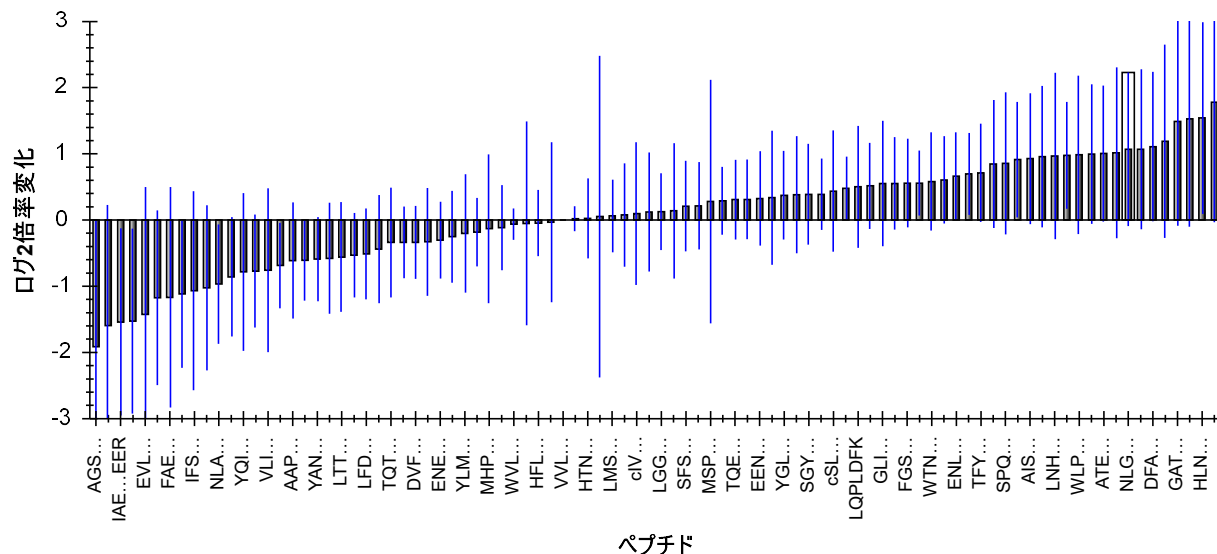
群比較のための最適化

本実験の本来の目標に立ち戻り、擬陽性、偽陰性（FDR） 1%でグループ平均値が統計学的有意差を示すペプチドのサブセットのみにドキュメントを減らすことができます。しかし測定ランダムに実施していないことと（疾患群がすべて健常群の前に測定されている）測定中の感度低下が考慮されていないため、疾患群で平均値が実際以上に高値である可能性に留意してください。

そこで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を精査し実際には差のない候補を減らすため以下の手順を実行します。

- [**Healthy v. Diseased : 設定**] 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 : 設定**] フォームで、[**適用対象**] ボックスで [**ペプチド**] オプションを選択します。
- 「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**] フィールドで「同じもしくはそれ以上」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、グラフは以下ようになります。



これらのペプチドをドキュメントから削除するべく以下の操作を行います。

- [**Healthy v. Diseased:グリッド**] の左上角、行ヘッダーと列ヘッダーが交わっているところ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」タンパク質で **VVLSGSDATLAYSFAFK** ペプチドが見えるまで、グリッドを下にスクロールします。
- **Ctrl** キーを押したまま行ヘッダーをクリックして、この行を非選択にします。

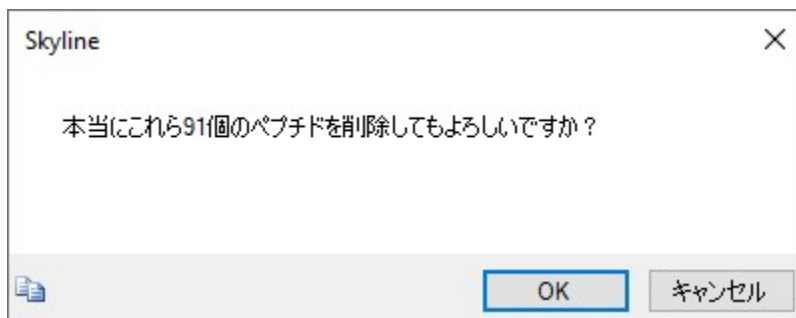
グリッドは以下のように表示されます。

タンパク質	ペプチド	倍率変化結果	調整されたP-値
NP_872279	WVLTVAHCFEGR	0.96 (99% CI:0.8...	0.4948
NP_058690	AFMDCCNYTK	0.97 (99% CI:0.3...	0.9373
XP_001066264	HFLJETGPK	0.97 (99% CI:0.6...	0.8111
NP_001013967	VTSAAFPSPIEK	0.98 (99% CI:0.4...	0.9399
S	VVLSGSDATLA...	1 (99% CI:1 to 1)	0.7689
NP_036745	ENSSNILDNLLSR	1.01 (99% CI:0.8...	0.8067
NP_001121161	HTNNGMCLTS...	1.02 (99% CI:0.6...	0.9284
NP_037244	TGTNLMDFLSR	1.04 (99% CI:0.1...	0.9399
NP_037244	LMSPEEKAPA...	1.04 (99% CI:0.7...	0.7890
XP_001066814	SLVIQKPSEEN...	1.05 (99% CI:0.6...	0.8067

VVLSGSDATLAYSFAFK ペプチドの倍率変化は 1 であり、信頼区間は 1 から 1 の 99%であることに注意してください。これは、正規化のための標準ペプチドとして使用されているものです。このペプチドはドキュメントから削除しない方がいいため非選択にしています。

- グリッドツールバー内の  ボタンをクリックします。

Skyline で選択した行の削除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。

これでグループ比較ウィンドウを閉じて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できます。残りのペプチドすべてについては 2 グループ間で平均値に有意差があるはずですが、標準偏差のばらつきが大きく有意差あるように見えないこともあります。実際条件別にグループ化されたピーク面積を表示している [ピーク面積] グラフを見ると標準偏差を示すエラーバーがオーバーラップしているのが分かります。

まとめ

本チュートリアルでインポートされた LC-MS/MS クロマトグラフィーデータを Skyline が自動ピーク積分した結果を視覚的に評価して修正する手法を学習しました。安定同位体標識ペプチドなしで SRM データセットのみを使用して作業を行いましたが、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でも同様に利用可能です。この精査と修正は手動で行うため、ペプチドが 2,000 個以下のデータセットでなければ現実的に無理かと思われれます。今回の場合 137 のペプチドの処理に約 1 時間かかりました。2,000 個のペプチドを含むデータセットでは 2 日以上かかることとなります。

より大規模なデータセットを扱う場合、まずは研究対象の条件間に存在する差異を示すような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。その場合、ターゲットのペプチドに対して査定と修正を手動で行うことに多くの時間を費やすことになるでしょう。Skyline のグループ比較サポートを使用すればこの種のスクリーニングもできるはずです。p 値（または FDR カットオフ）の設定は（本チュートリアルで設定した）1%である必要なく、必要に応じて変えることが可能です。

ここまで、Skyline での処理作業が、定量実験で生じる潜在的なエラーの原因の理解とその修正にどう役立つかを見てきました。ピークの誤判定、干渉、ピークの分断、検出限界下のペプチド、測定中における 200 倍もの感度低下、およびペプチドの共溶出など様々な問題についても扱ってきました。これらの問題を理解、修正し、そして注釈できるよう、Skyline のツールを使いこなせるようになりましょう。そうすればターゲットプロテオミクスの定量データ解析でのエラーも減り、より効率的に生物学的な考察をすることもできるようになるでしょう。

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).