

Skyline Processing Grouped Study Data

Part 1: Label-Free Differentiation for Method Refinement

Targeted proteomics as an approach has become popular largely because of its advantages in measuring differences in protein and peptide expression across multiple groups or states of organisms. Skyline was originally funded by the NCI Clinical Proteomics Technology Assessment for Cancer (CPTAC) program in its efforts to show targeted proteomics could be effectively applied to detect and achieve clinical verification of candidates for disease biomarkers.

In this tutorial, you will use Skyline to analyze SRM data from an experimental case-control study of heart failure using a salt-sensitive rat model. This study sought to identify differences in protein expression in plasma between diseased and healthy groups utilizing a list of candidate proteins associated with heart failure derived from the literature.

This experiment monitored 137 peptides from 49 proteins over the course of 42 LC-MS/MS injections. Once familiar with the data processing and visualization techniques described in this tutorial, you should find it relatively straightforward to manage and assess data from larger studies with even more targets across more experimental groups. Without this tutorial, however, one can easily get overwhelmed by the details of working with comparatively smaller data sets.

This tutorial is comprised of the following sections:

- Getting Started
- Differential analysis after method refinement
- Setting up to process multi-replicate data without reference standards
- Scheduled SRM and peak truncation
- Beginning multi-replicate data processing
- Global normalization standards
- Continuing multi-replicate data processing
- Preparing for statistical analysis with replicate annotations
- Annotating peptides with problem peaks
- Initial multi-replicate inspection in Skyline
- Conclusion

Getting Started

To start this tutorial, download the following ZIP file:

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

Extract the files in it to a folder on your computer, like:

C:\Users\brendanx\Documents

This will create a new folder:

C:\Users\brendanx\Documents\GroupedStudies1

If you inspect the contents of this new folder, you will find that it contains a subfolder:

- Heart Failure – files from a heart failure study in a salt sensitive rat model, provided by the labs of Michael J. Toth and Michael J. MacCoss.

Follow the tutorial below to learn how to process data for a study of this nature, and to gain a better understanding of a study's protein/peptide targets as well as its overall data quality.

Differential analysis after method refinement

The data in the “Heart Failure” subfolder of the “GroupedStudies1” folder you just created was collected at the MacCoss lab while investigating the potential and limitations of the approach described in the [Targeted Method Refinement](#) tutorial. The data were never published for a variety of reasons, some of which will become apparent when you use Skyline to understand the collected data. However, many of the ideas that came from this trial of targeted method refinement and other experiments like it were published separate from this data set.¹

In the original study, 109 proteins were selected from the literature based on their proposed involvement in heart failure. These proteins were added to a Skyline document. Some were added by importing FASTA format text for the protein sequences and allowing Skyline to perform *in silico* tryptic digestion. Parameters were set in Skyline to include all peptides between 6 and 30 amino acids long with no missed cleavages. For other proteins, the *in silico* processing was performed outside Skyline. The resulting peptide lists from these proteins were pasted directly into a Skyline document (see [Targeted Method Editing](#) tutorial). This created a document with 2,165 target peptides. For each doubly-charged precursor, singly-charged product ions corresponding to y_3 through $y_{(n-1)}$ were considered. This initial, exhaustive method covered 12,194 transitions and was measured with an unscheduled SRM method on a sample of pooled rat plasma. This initial analysis was only performed once and required 151 separate mass spec injections.

The resulting raw data files were imported into Skyline and used to determine which of the target peptides showed any promise of being detected in the plasma matrix without any further sample preparation. Of the original 2,165 target peptides, only 135 (from 49 proteins) had a peak in the full-gradient chromatograms with sufficient co-elution of the targeted y -ions (2 with 3 ions, 5 with 4 ions, 27 with 5 ions and 103 with 6 ions). These remaining 780 transitions were scheduled for quantification in a single run using 5 minute wide retention time windows around the detected peaks in a 90 minute gradient. This scheduled method was then run on plasma digests from 14 salt sensitive rats (7 healthy and 7 that had experienced heart failure induced by a high salt diet) in triplicate for a total of 42 injections. These injections, which will be analyzed throughout the course of the current tutorial, were intended to

determine which of the 135 detectable peptides demonstrated any potential of differentiating between the two groups, making them candidates for further study as potential biomarkers.

The overall goal of this study was to determine whether useful scientific insight could come from a targeted investigation, using SRM, starting without very much prior knowledge of the targeted proteins or peptides and without aid of stable isotope labeled reference peptides.

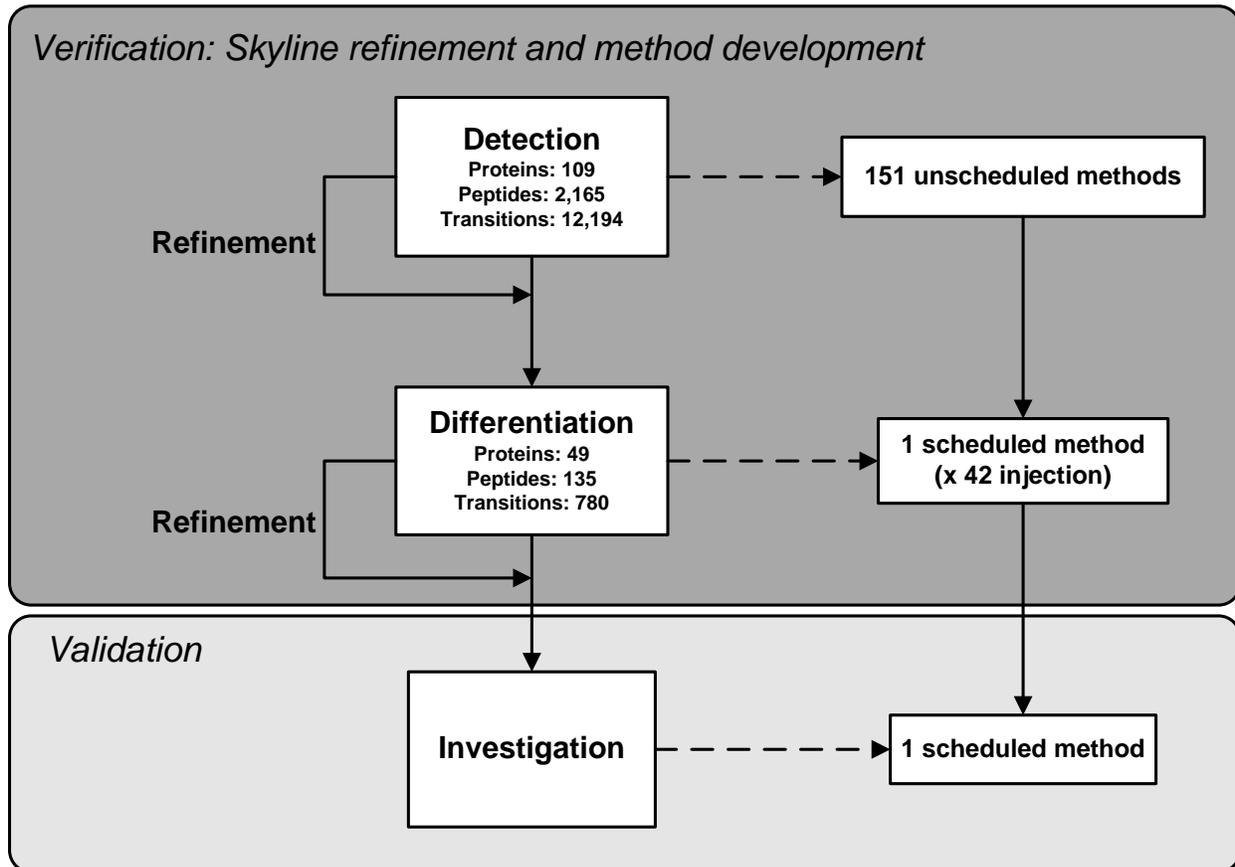


Figure 1: A schematic showing proposed “Detection” and “Differentiation” phases of targeted method refinement in a heart failure study on a salt sensitive rat model.

To begin processing the data collected for the Differentiation phase of this method refinement study:

- Open the “Rat_plasma.sky” file in the “GroupedStudies1\Heart Failure\raw” subfolder.

By inspecting the indicators at the bottom right corner of the Skyline window, you will see that the file you opened contains 49 proteins, 137 peptides and 789 transitions.



This is not quite what is promised in the figure above (from a manuscript draft that was never published), but the document also contains one extra peptide list (at the bottom, named “S”), which contains 3 global normalization peptides. These are intended to reduce the impact of systematic variance in measurement

that impacts all peptides in the sample. You will learn more about this later. And, somehow, this file must have lost 1 protein with 1 peptide from what was described above.

You will also notice that some of the peptides have matching library spectra. You can see this in the **Targets** view. The peptide icons with spectrum lines in the lower right corner (Y₁) have matching spectra and those without these lines (Y₂) do not. The spectra come from two different spectral libraries: one from NIST and one from the Global Proteome Machine (GPM). You can explore their coverage by doing the following:

- On the **View** menu, click **Document Grid**.
- On the **Views** menu in the upper-left corner of the **Document Grid** form, click **Precursors**.
- Scroll to the right in the **Document Grid** to find the “Library Name” column, and click it twice to sort descending.

You can now scroll down the grid to see the number of peptides with a library spectrum (80), from the Rat (NIST) library (49) and from the Rat (GPM) library (31). Note that the row number of the selected cell is shown in the toolbar above the grid.



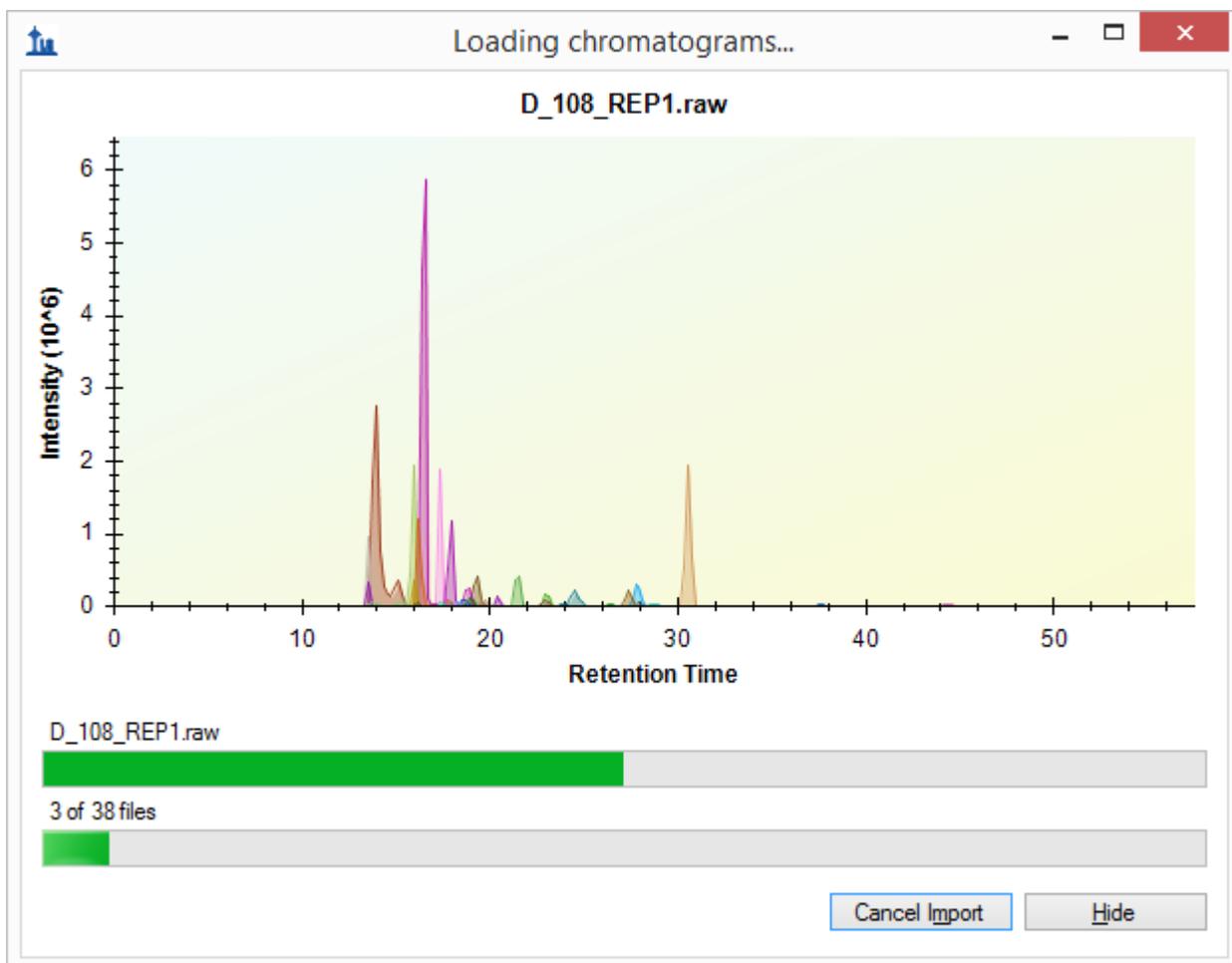
For the current example, this information gives you a sense of the sparse prior knowledge on which the experiment was based.

Setting up to process multi-replicate data without reference standards

To begin processing the SRM data acquired for the 14 subjects in technical triplicate, do the following:

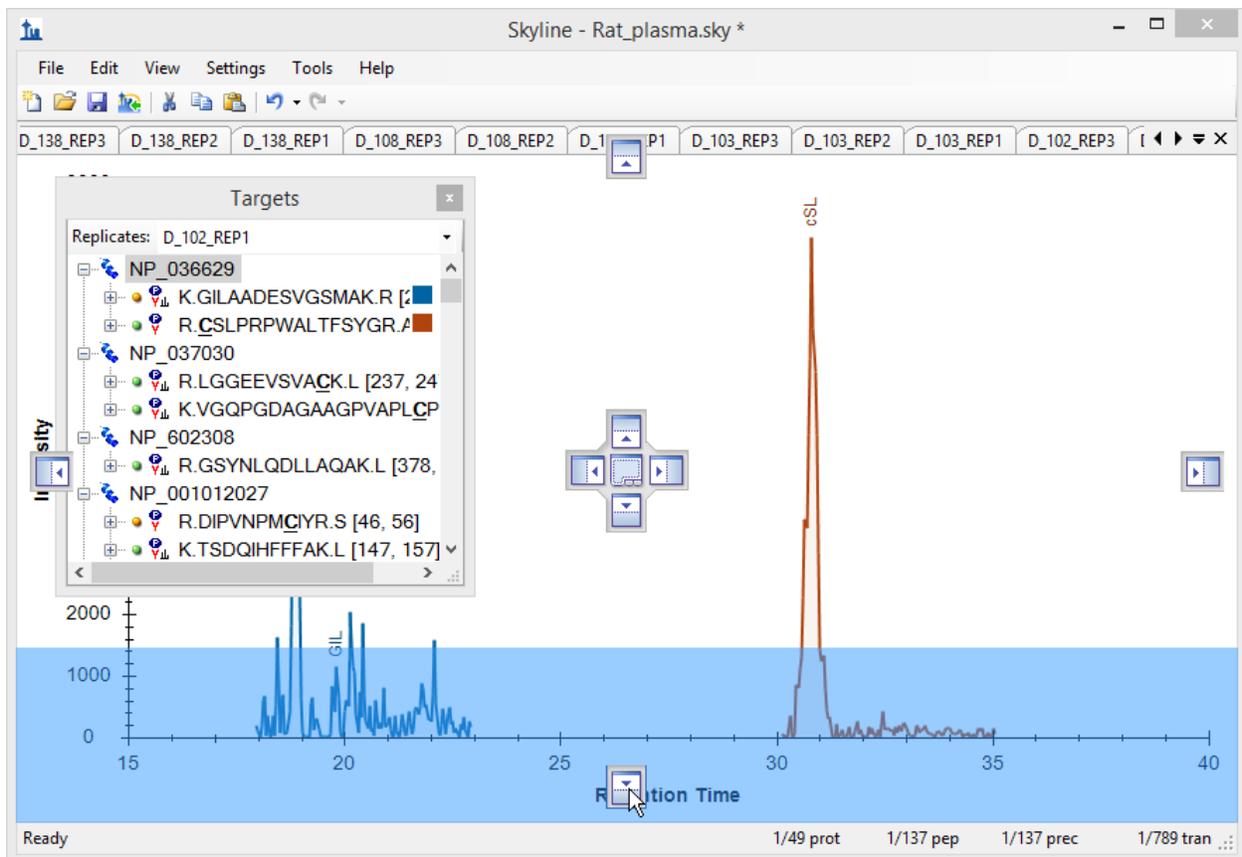
- Close the **Document Grid** form.
- On the **File** menu, choose **Import** and click **Results**.
- Click the **OK** button in the **Import Results** form to import single injection replicates.
- Click on the first raw data file named “D_102_REP1.raw” in the **Import Results Files** form.
- Hold down the Shift key and click the last raw data file named “H_162_REP3.raw” to select all the raw data files from the “GroupedStudies1\Heart Failure\raw” folder.
- Click the **Open** button.

The files should begin loading, and Skyline will display progress in a window like the one shown below:

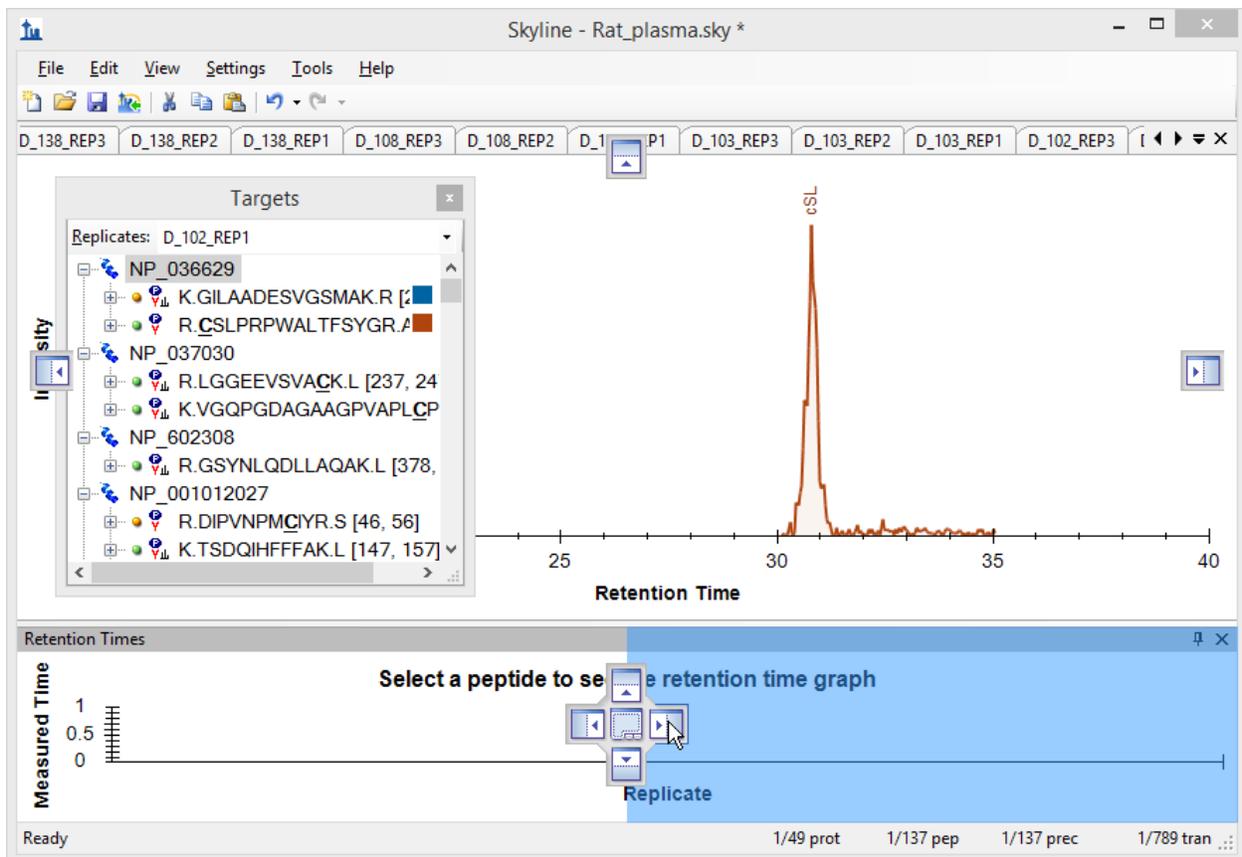


To continue preparing for data processing while Skyline imports the SRM chromatograms, do the following:

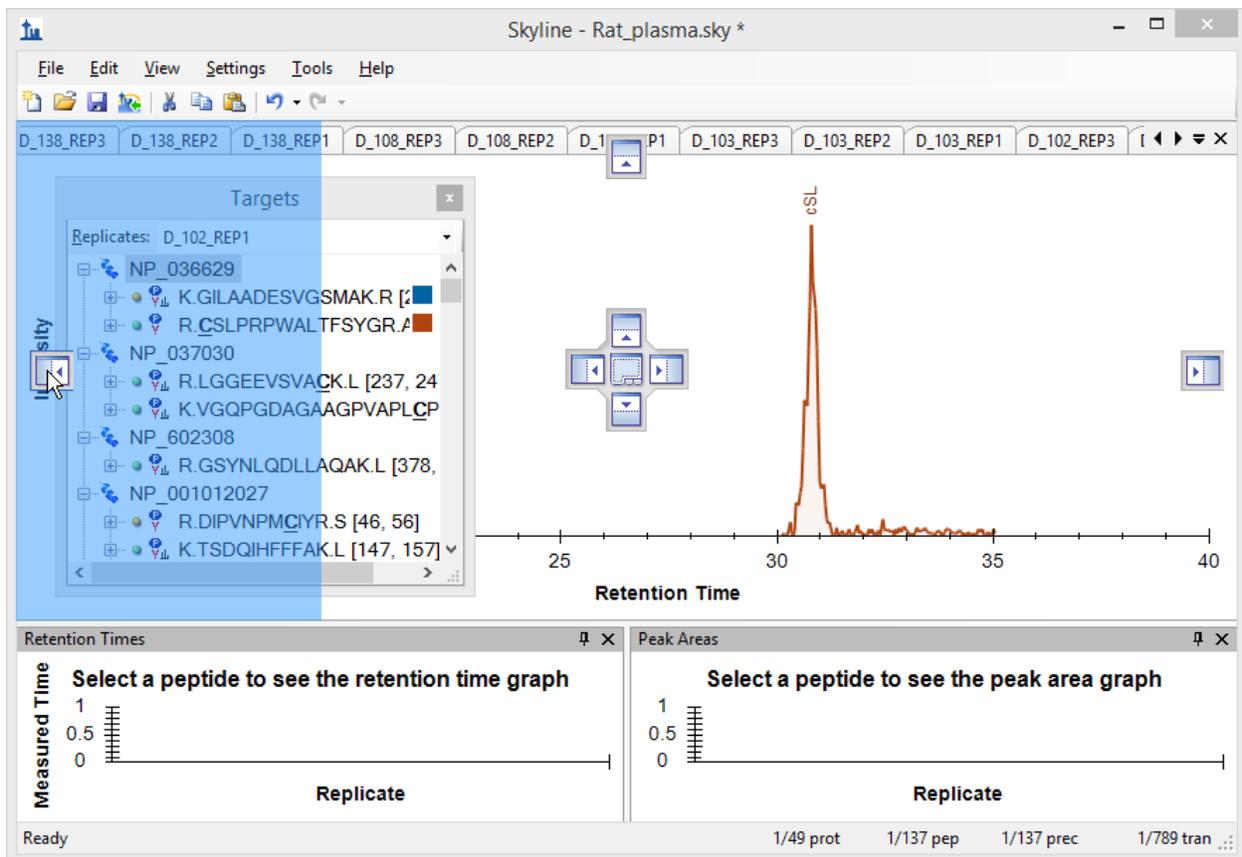
- Click the **Hide** button.
- Close the **Library Match** view.
- Click and drag the title of the **Targets** view until you see a floating blue rectangle tracking the mouse cursor.
- Release the left mouse button to leave the **Targets** view floating above the Skyline window.
- On the **View** menu, choose **Retention Times** and click **Replicate Comparison (F8)**.
- Click and drag the title of the **Retention Times** view until you see the docking indicators.
- Drag the cursor into the bottom docking indicator to see a blue rectangle fill the bottom edge of the Skyline window, as shown below.



- Release the left mouse button to dock the **Retention Times** view.
- On the **View** menu, choose **Peak Areas** and click **Replicate Comparison (F7)**.
- Drag the **Peak Areas** view until the mouse cursor is inside a docking indicator, and a blue rectangle indicates it will share the bottom edge with the **Peak Areas** view.

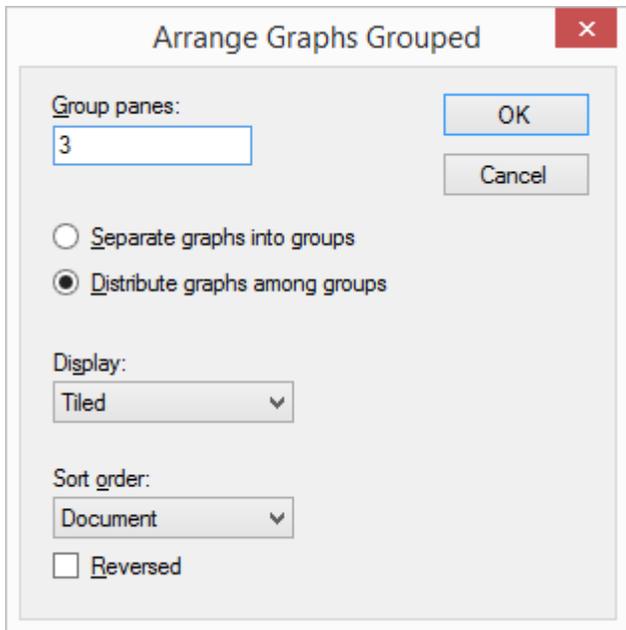


- Release the left mouse button to dock the **Peak Areas** view.
- Drag the **Targets** view until the mouse cursor is inside a docking indicator, and a blue rectangle indicates it will dock to the left edge of the Skyline window.



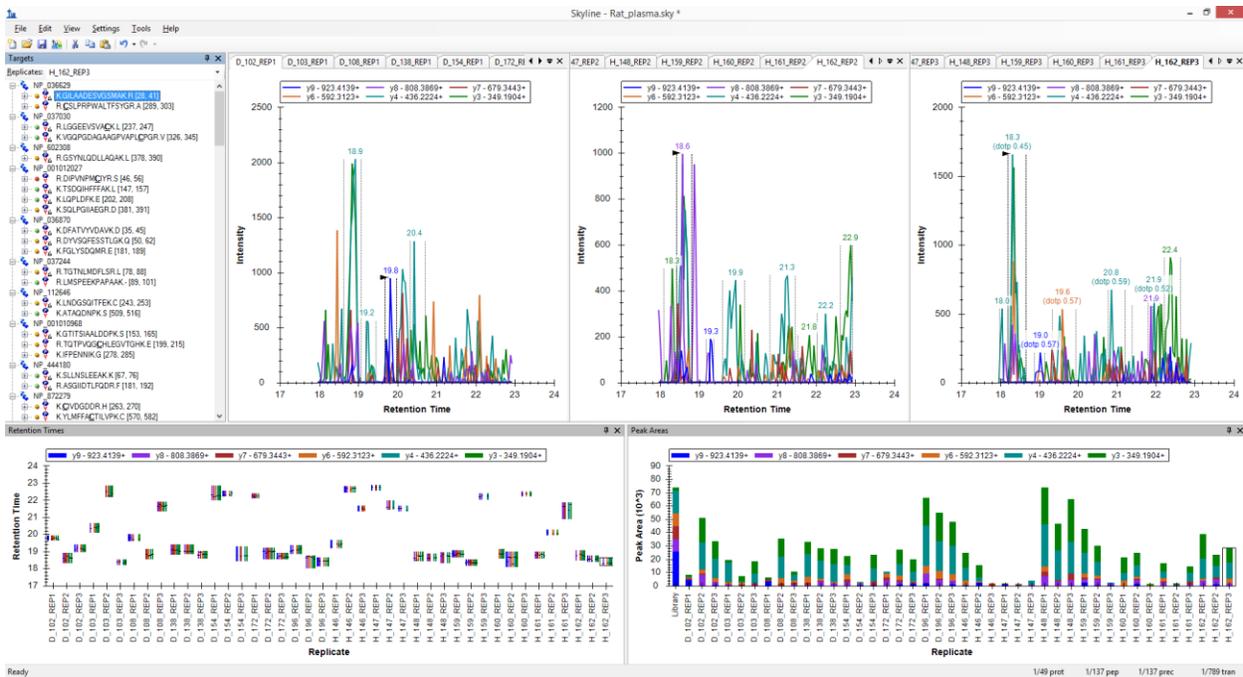
- Close any other windows left floating outside the main Skyline window.
- On the **View** menu, choose **Arrange Graphs** and click **Grouped**.
- In the **Group panes** field enter "3" on the **Arrange Graphs Grouped** form.
- Click the **Distribute graphs among groups** option.
- In the **Sort order** dropdown list choose "Document".

The form should look like this:



- Click the **OK** button.
- Maximize the Skyline main window.
- Select the first peptide in the document.

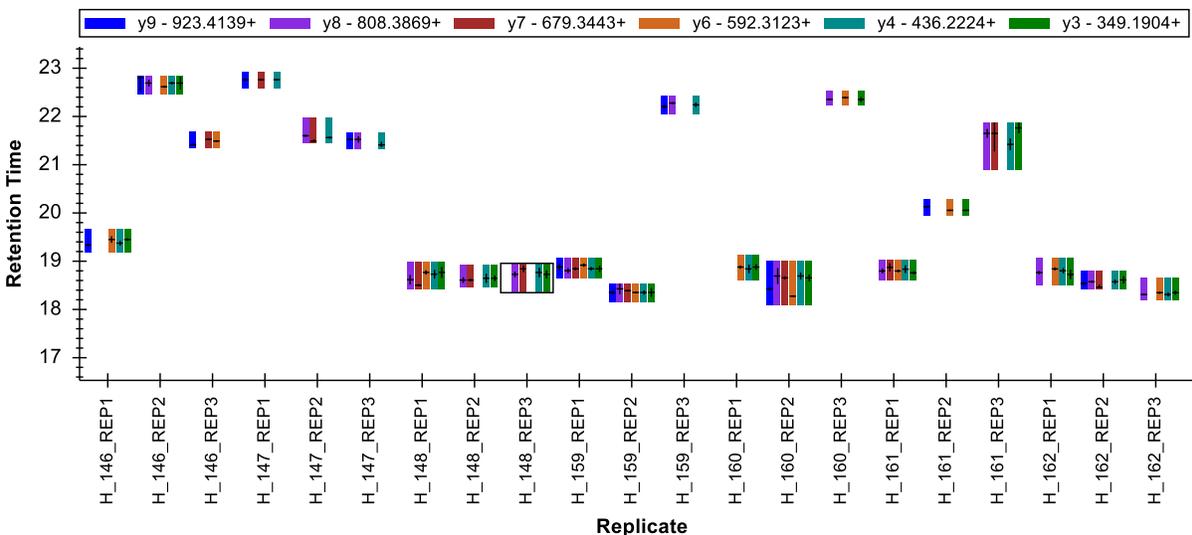
Your Skyline window should now look something like this:



Note: This image was taken on a 24" monitor at 1920x1080. If you are viewing this tutorial digitally, you may want to zoom in to 200% or more to view this image. After this, the tutorial will return to using

images better suited to the 8½ x 11” page layout. While many Skyline workshops have been taught with the 1024x768 screen resolution, you will likely enjoy using Skyline more with a larger monitor.

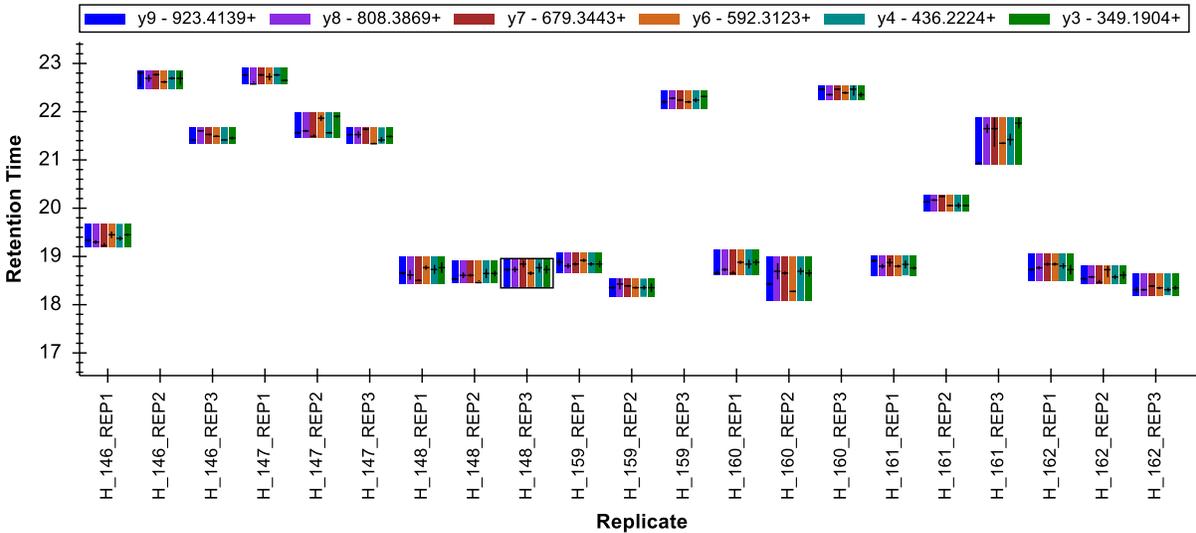
Looking at the **Retention Times** view for this first peptide, K.GILAADESVGSMAR [28, 41], you can see that the integration is not very consistent. Most of the chosen peaks elute around 19 minutes. However, nearly one third (12) elute closer to 22 minutes. You might also notice that there are many missing colored bars in this plot:



This is because Skyline is still in **Refinement Mode**. While in this mode, Skyline attempts to decide which peaks have similar enough shape and apex times to be considered co-eluting. The missing bars are an indication that the peaks within the integrated range are not matching very well on all monitored transitions. For this phase of the experiment, you want to move from **Refinement Mode**, where you are choosing which transitions to measure, to **Quantification Mode**, where Skyline performs integration on all peaks, reducing variance in your measurements. To make this change, do the following:

- On the **Settings** menu, click **Integrate All**.

This should change the plot to look like the one below:

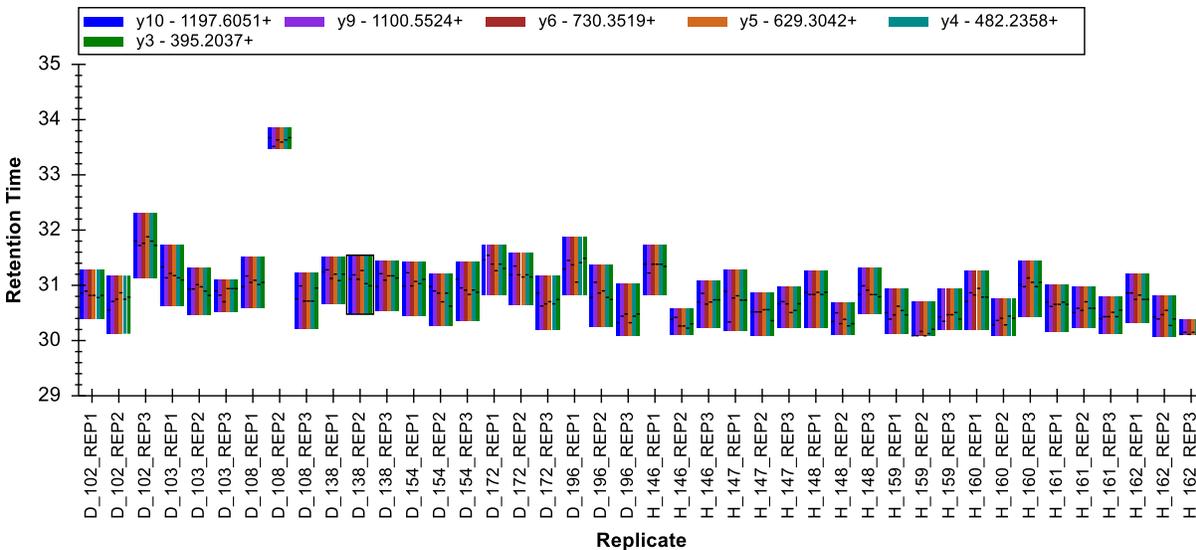


Still, the integration for this peptide is quite poor. While the peak at around 19 minutes appears to be the better peak, it is hard to have much confidence that it is caused by a single peptide. Furthermore, it is unlikely that any integration adjustment will result in a single peptide that can be measured consistently across all runs. Therefore, instead of wasting more time here making additional adjustments, you should:

- Delete the first peptide.

Scheduled SRM and peak truncation

Just by looking at the retention time plot for the second peptide, you can see that it is much more consistently integrated.



However, you can still see that the peak apexes for all transitions (shown by the horizontal lines in the bars) are not overly consistent. If you look at a few of the chromatogram graphs, you can also see that the peaks are jagged and of medium- to low-intensity.

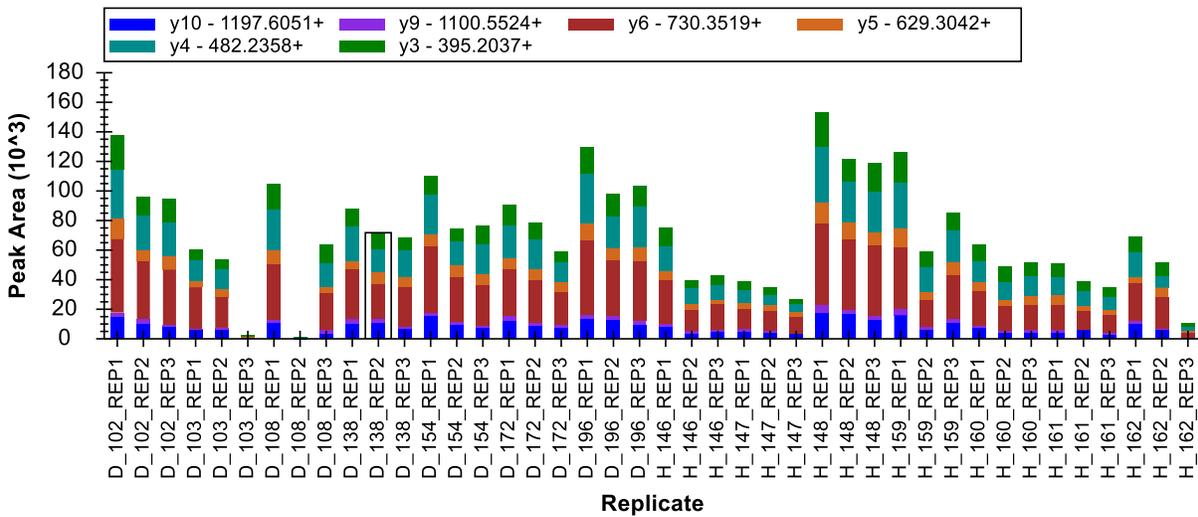
If you do not see the order shown above, with all of the diseased replicates (D_) on the left and all of the healthy replicates (H_) on the right, do the following:

- Right-click the **Retention Times** view, choose **Order** and click **Document**.

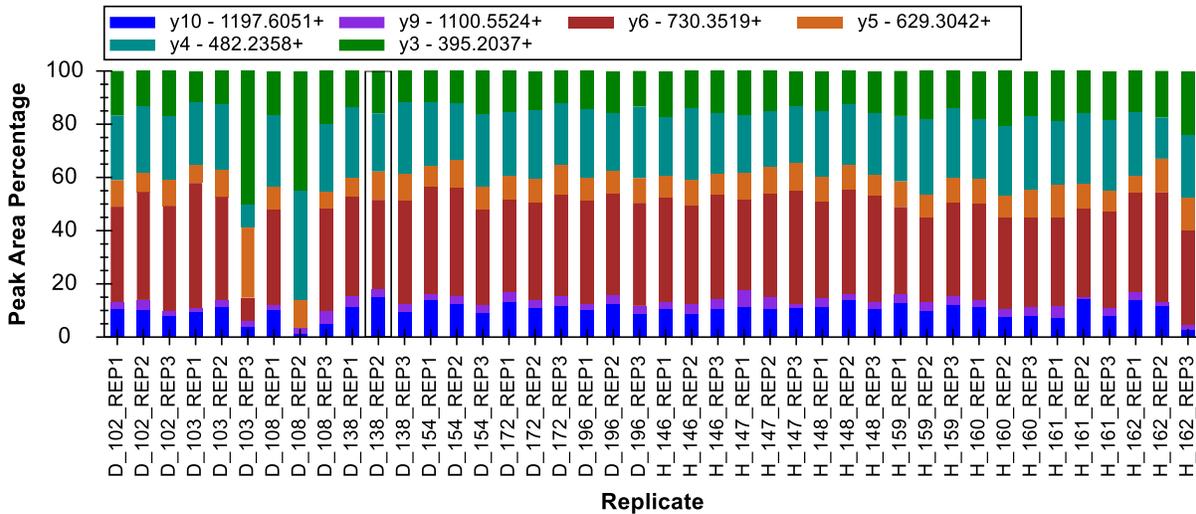
Before continuing, make one final adjustment to the plots you are viewing.

- Right-click on the **Peak Areas** view, choose **Normalize To** and click **Total**.

This changes the **Peak Areas** plot from:



To:

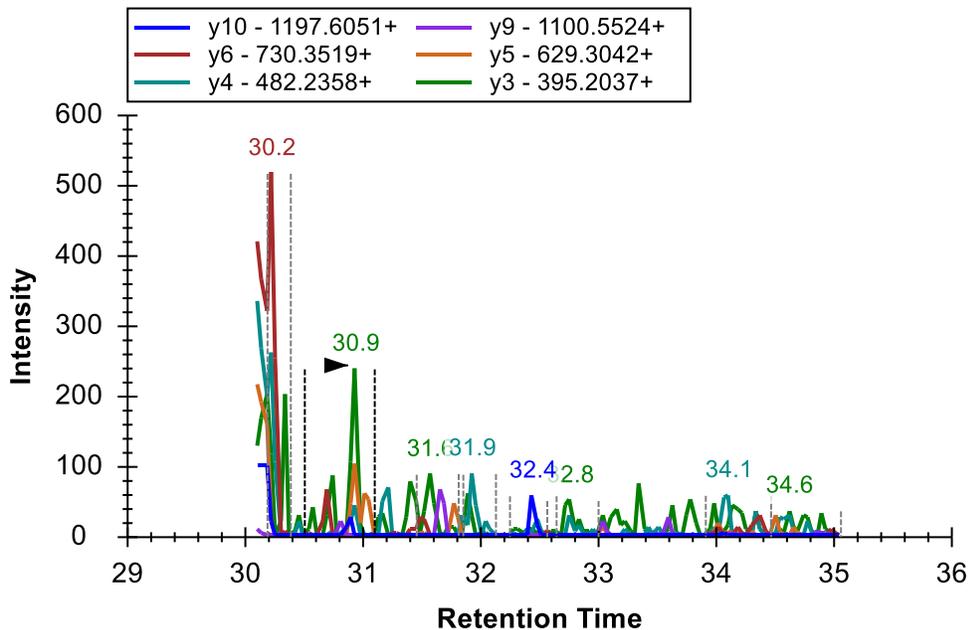


This makes it possible to see differences in relative ion abundance across multiple runs in a single glance and adds to the information you are already getting from the **Retention Times** view. You might notice that D_103_REP3 and D_108_REP2 are quite different, and that H_162_REP3 is somewhat different from all other replicates. Note that only the D_108_REP2 sample looked suspicious when only using the **Retention Times** view.

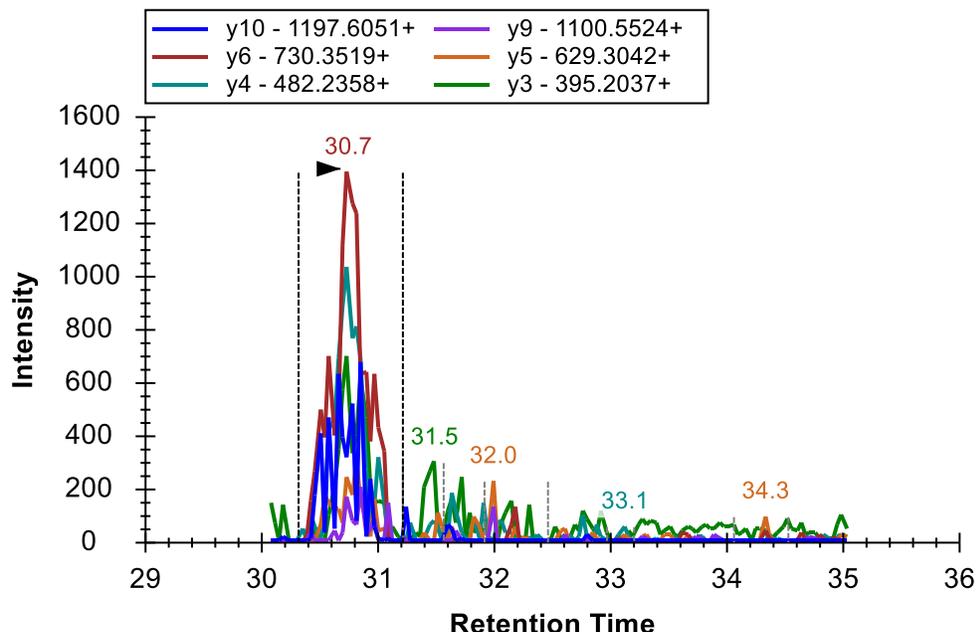
Do the following to inspect and correct these peaks:

- Click the D_103_REP3 bar in the **Peak Areas** view.

This will activate the chromatogram graph for the peak, which looks like:



While plots for the correctly integrated peaks look more like this:



Since the scheduled acquisition window did not capture the entire elution profile for the peptide, Skyline has picked the wrong peak. However, you can still see the correct peak ending around 30.2 minutes. This is called a “truncated peak”. You can correct the integration for this peptide to use the truncated peak by doing the following:

- Click-and-drag beneath the x-axis at about 30.4 minutes, just after the truncated peak.
- Drag left until the integration boundary stops moving, and release the left mouse button.
- Click the bars for D_108_REP2 in the **Retention Times** view to activate its chromatogram graph.
- Click-and-drag beneath the x-axis of the graph from about 30.5 minutes to the far left.
- Click the bars for D_162_REP3 in the **Retention Times** view to activate its chromatogram graph.

The peak for D_162_REP3 is also truncated. However, Skyline has already chosen it correctly, as it has in several other cases. You might be able to deduce peak truncations by closely inspecting the **Retention Times** plot. Pay close attention to the lengths of the bars, as well as the proximity of the horizontal lines to the bar edges.

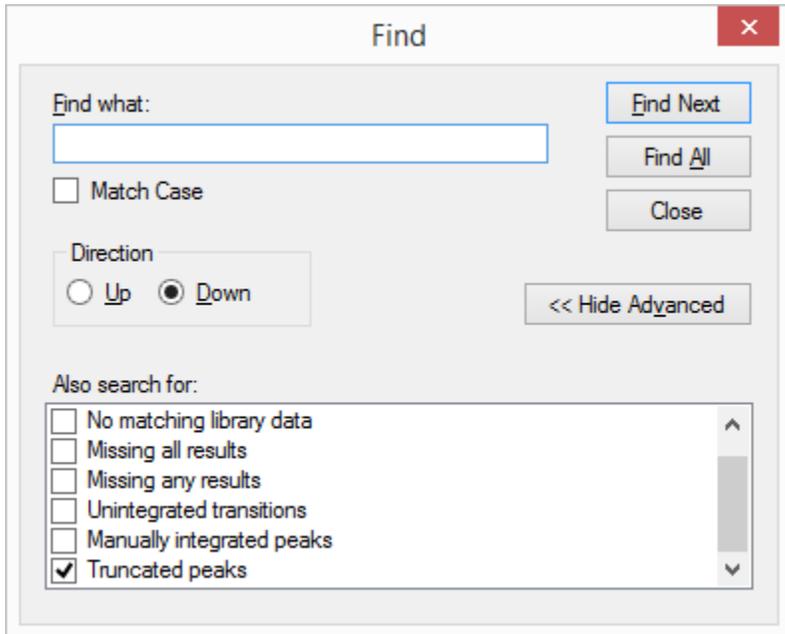
Peak truncation is a considerable problem for label-free data such as this. With isotope labeled reference peptides, where every point in the peak provides a valid ratio between light and heavy peptide precursors, you will lose precision with a truncated peak but the ratio between light and heavy should still be valid. With label-free data, conversely, you cannot count a truncated peak as a valid measurement of the peptide. You must either calculate your differential statistics with missing data or omit the peptide from consideration entirely. Skyline keeps track of truncated peaks in order to make that possible for downstream tools.

Skyline defines a truncated peak as any peak where one of its boundaries is at a terminal point of the chromatogram and the intensity at that terminal end is more than 1% of the peak height higher than the intensity at the other integration boundary.

To see other automatically integrated peaks which Skyline has identified as truncated, perform the following steps:

- On the **Edit** menu, click **Find** (Ctrl-F).
- Clear the **Find what** field.
- Choose **Down** for **Direction**.
- Click the **Show Advanced** button, advanced options are not already showing.
- Check **Truncated peaks** (and uncheck anything else that may be checked).

The **Find** form should look like this:



- Click the **Find All** button.
- Click the **Close** button.

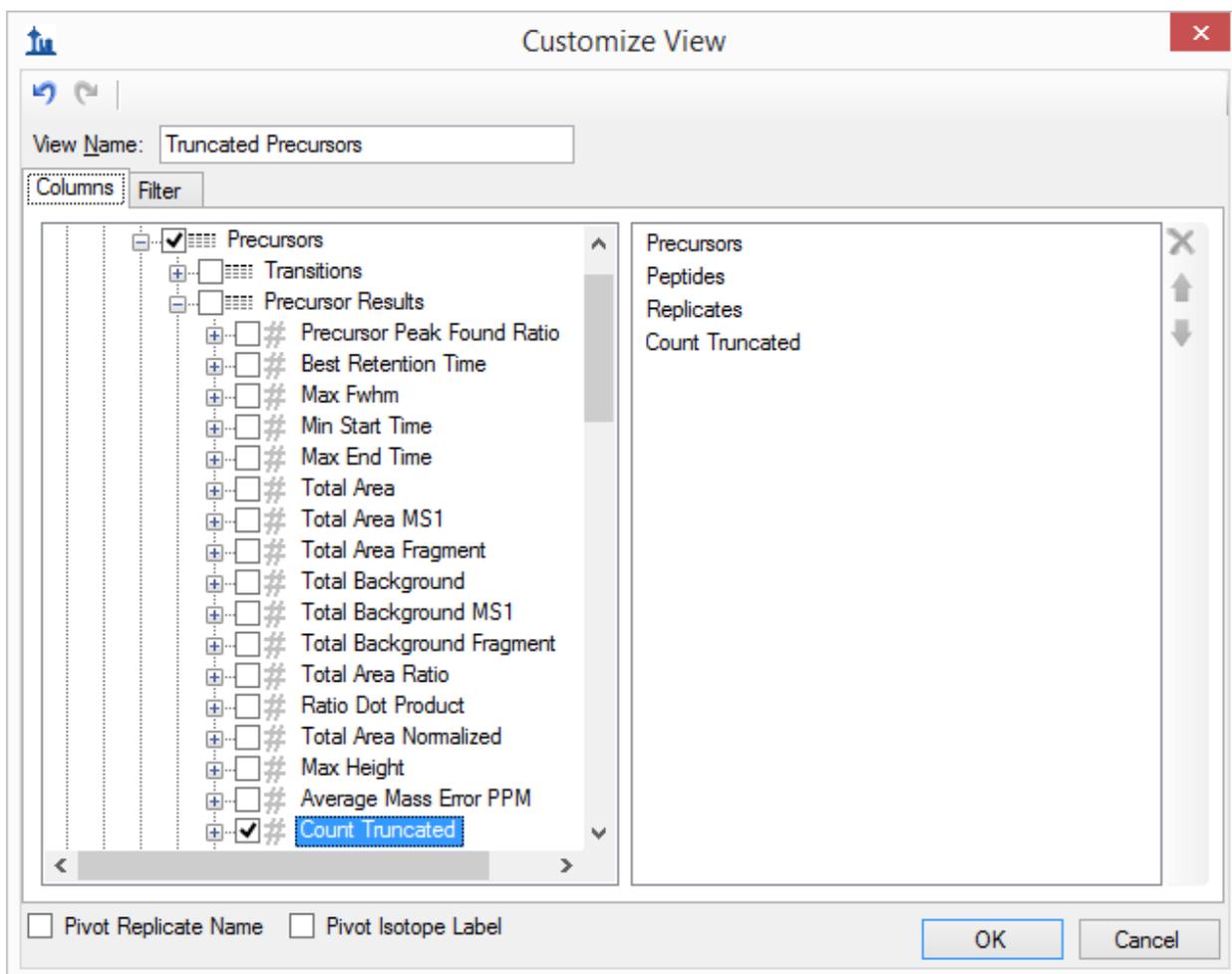
Skyline shows a long list of truncated precursors and transitions, starting with the three peaks you just inspected. You can double-click on any line to see the corresponding chromatogram graph activated.

Find Results		
Location	Type	Text
905.9565++ (D_103_REP3)	Precursor Results	6 truncated peaks
905.9565++ (D_108_REP2)	Precursor Results	6 truncated peaks
905.9565++ (H_159_REP2)	Precursor Results	6 truncated peaks
P [y10] - 1197.6051+[5] (H_162_REP3)	Transition Results	Truncated peak
W [y9] - 1100.5524+[6] (D_196_REP3)	Transition Results	Truncated peak
T [y6] - 730.3519+[1] (D_196_REP3)	Transition Results	Truncated peak
T [y6] - 730.3519+[1] (H_146_REP2)	Transition Results	Truncated peak
T [y6] - 730.3519+[1] (H_148_REP2)	Transition Results	Truncated peak
T [y6] - 730.3519+[1] (H_159_REP1)	Transition Results	Truncated peak
T [y6] - 730.3519+[1] (H_160_REP2)	Transition Results	Truncated peak
T [y6] - 730.3519+[1] (H_162_REP2)	Transition Results	Truncated peak
T [y6] - 730.3519+[1] (H_162_REP3)	Transition Results	Truncated peak
F [y5] - 629.3042+[4] (H_160_REP2)	Transition Results	Truncated peak
F [y5] - 629.3042+[4] (H_162_REP2)	Transition Results	Truncated peak

You can also use the **Document Grid** to create a list of all precursors with any peak truncation by doing the following:

- On the **View** menu, click **Document Grid**.
- On the **Views** menu in the **Document Grid** view, click **Precursors**.
- On the **Views** menu in the **Document Grid** view, click **Customize View**.
- In the **View name** field, enter “Truncated Precursors”.
- In the right side list, select all column names (using click and shift-click) after the first two (Precursor and Peptide).
- Click the **X** button near the right edge to delete all but the first two columns.
- Check “Replicates” in the tree on the left side.
- Expand the tree to check Proteins > Peptides > Precursors > PrecursorResults > Count Truncated.

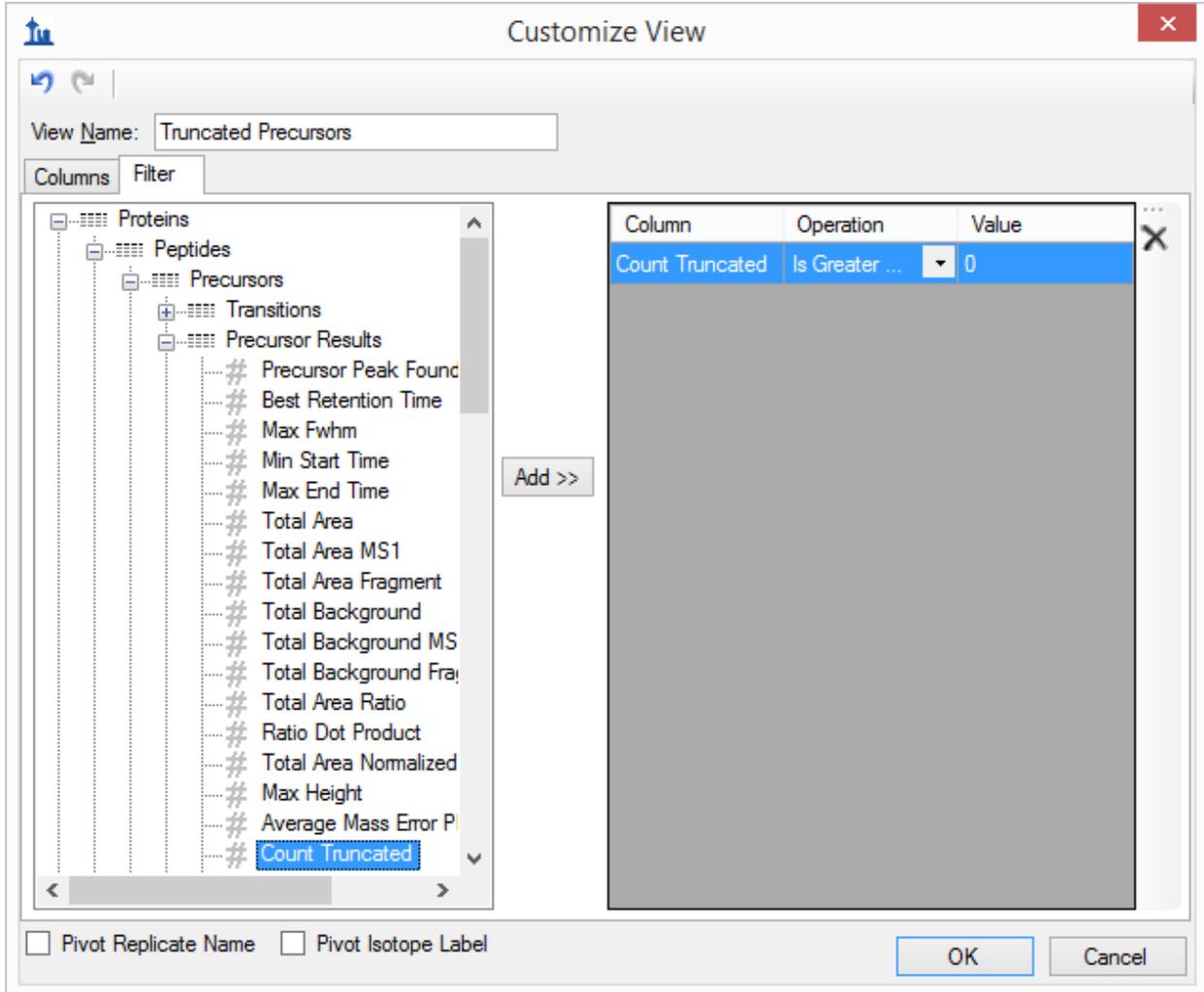
The **Customize View** form should look something like this:



- In the right side list, double-click “Count Truncated”.
- Click the **Filter** tab.
- Click the **Add** button.

- From the **Operation** dropdown list choose “Is Greater Than”.
- In the **Value** field, enter “0”.

The **Customize View** form should look something like this:



- Click the **OK** button.
- Click the “Count Truncated” column header twice to sort descending.

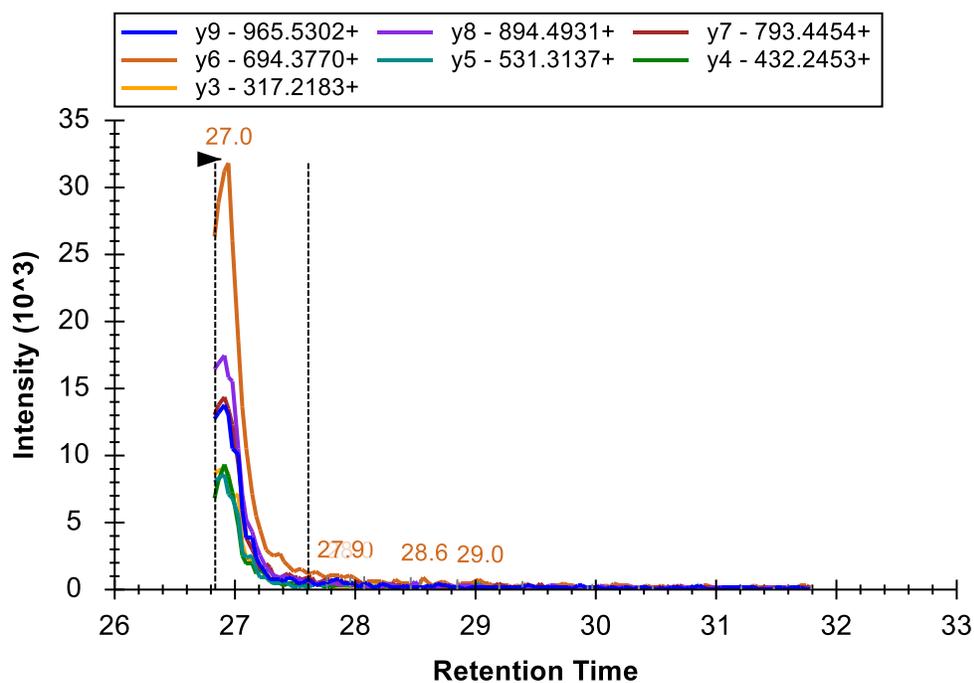
The **Document Grid** will show 222 truncated peaks that Skyline has already chosen automatically:

Document Grid				
Views	of 222		Export...	Find:
	Precursor	Peptide	Replicate	Count Truncated
▶	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

To view these peaks:

- Click the **Peptide** link for DFATVYVDAVK.
- Click each **Replicate** link for the peptide.

Skyline will activate chromatogram graphs that look like the following:

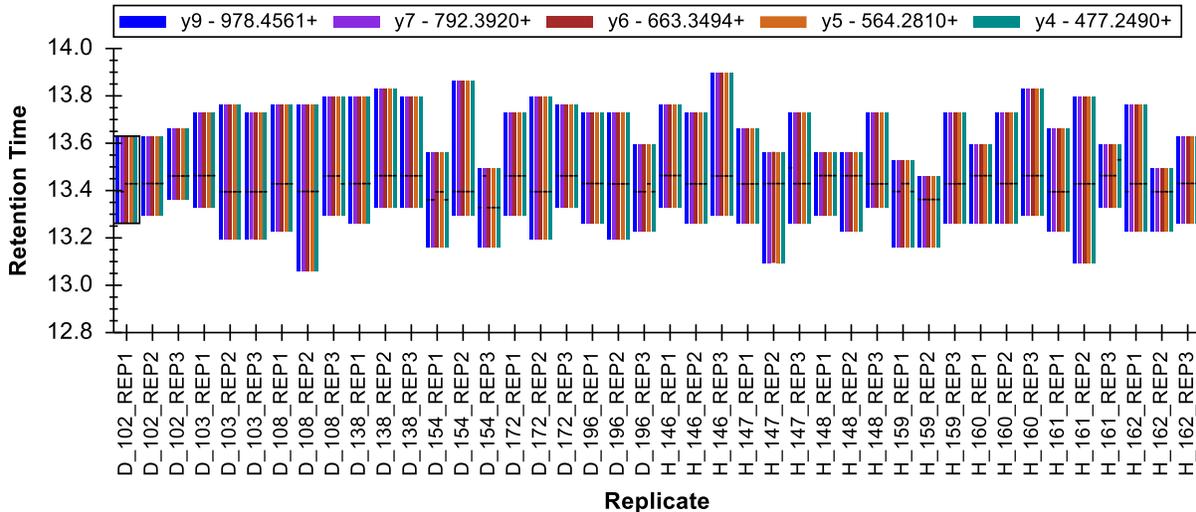


- Close the **Document Grid**.

- Close the **Find Results** view.
- Select the second peptide R.LGGEEVSVACK.L [237, 247]

Beginning multi-replicate data processing

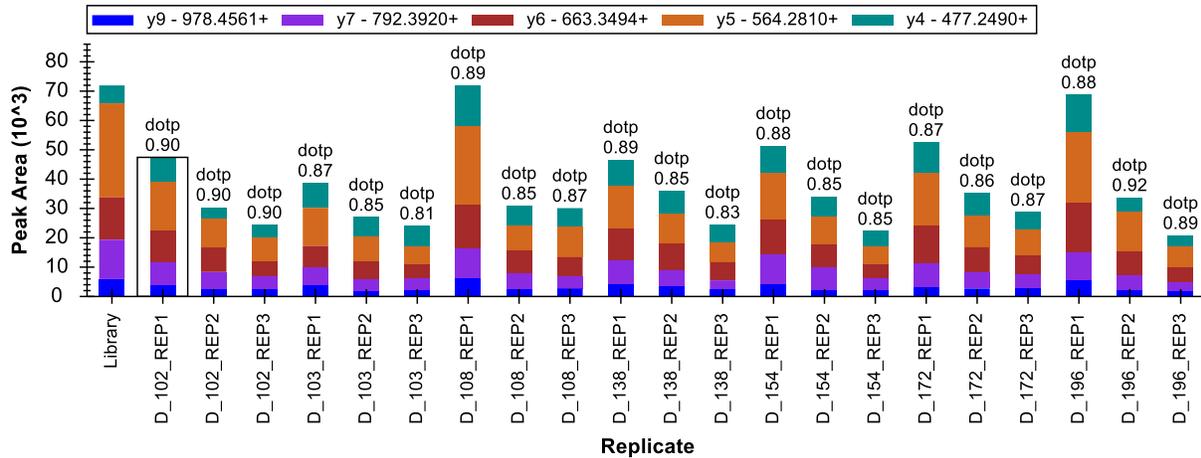
The peptide LGGEEVSVACK looks much better than the other two we have looked at so far. The retention times are quite stable and the horizontal peak apex lines match up for almost all transitions in each replicate.



The relative ion abundances shown in the **Peak Areas** view also appear relatively stable. You might notice that there is a corresponding library spectrum for this peptide and that its relative ion abundance (shown in the left-most bar in the plot) is similar to that of the measured peaks. To get a better sense of just how closely a peptide matches its corresponding library spectrum, do the following:

- Right-click the **Peak Areas** view, choose **Normalize To** and click **None**.
- Move the splitter bar to give the **Peak Areas** view most of the width of the screen. Conversely, if your screen space is limited, click-and-drag a zoom rectangle around the diseased (D_) replicates.

Skyline will show you the dot-product (dotp) relationship, a measure of similarity between the library spectrum peaks and the measured peak areas ranging from 1 (best) to zero (worst).



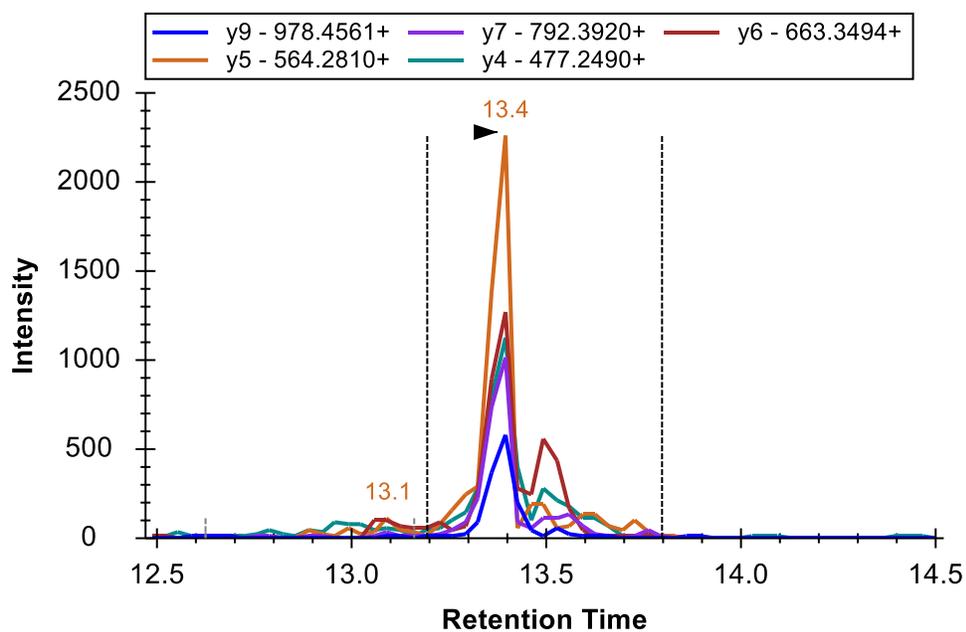
Return the **Peak Areas** view to its former state by doing the following:

- If you used zooming, right-click and click **Undo All Zoom/Pan**.
- Right-click, choose **Normalize To** and click **Total**.

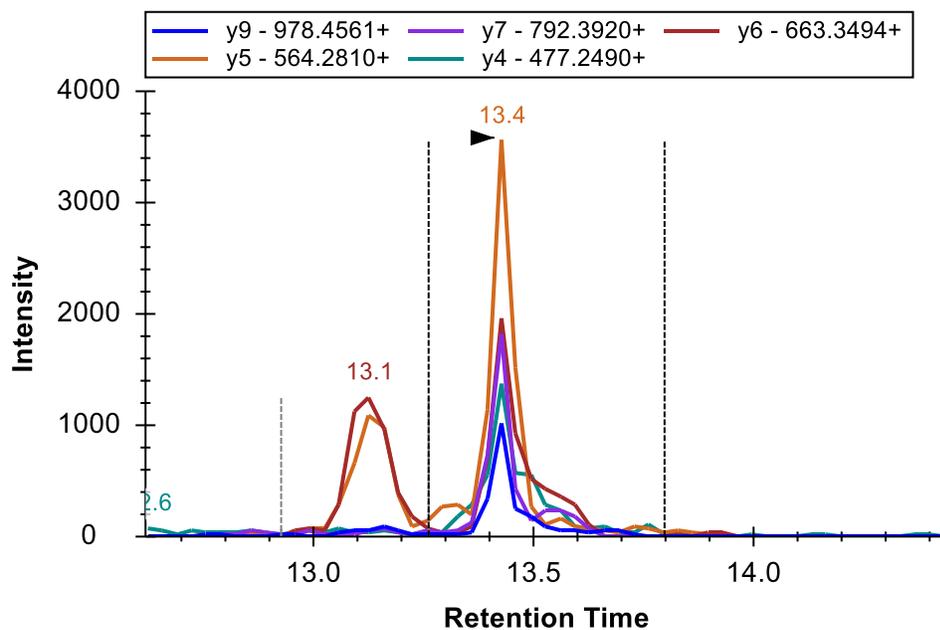
Now take a closer look at the chromatograms for these peaks by doing the following:

- On the **View** menu, choose **Auto-Zoom** and click **Best Peak (F11)**.
- Click the **Replicates** dropdown list in the **Targets** view.
- Use the down arrow key to scan through several of the chromatogram graphs.

You might notice that some of the transitions show a signal around the edges of the main peak that is not well matched by the other peaks.

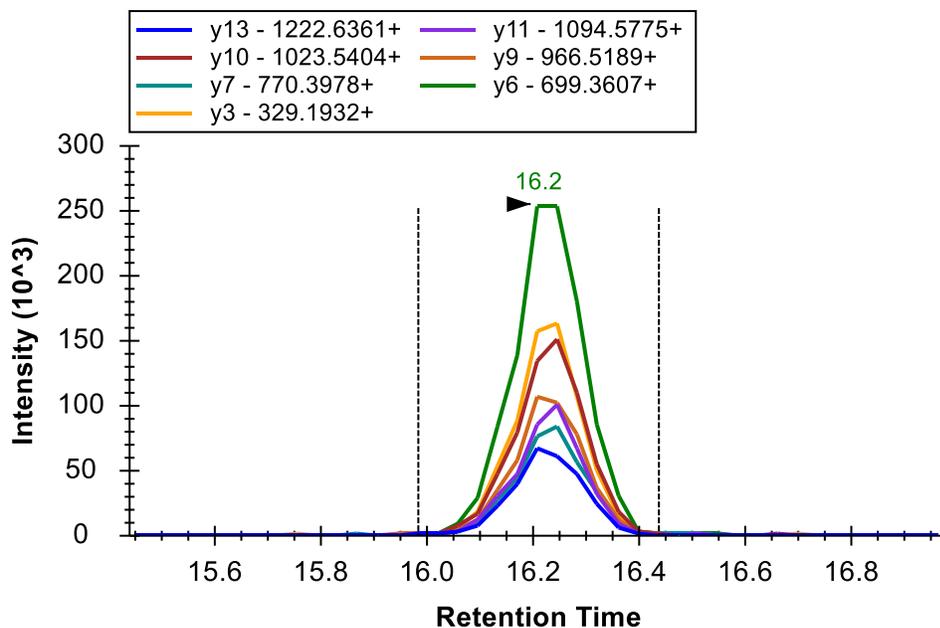


There is frequently signal on y5 and y6 around 13.1 minutes, which is clearly not from the same peptide.

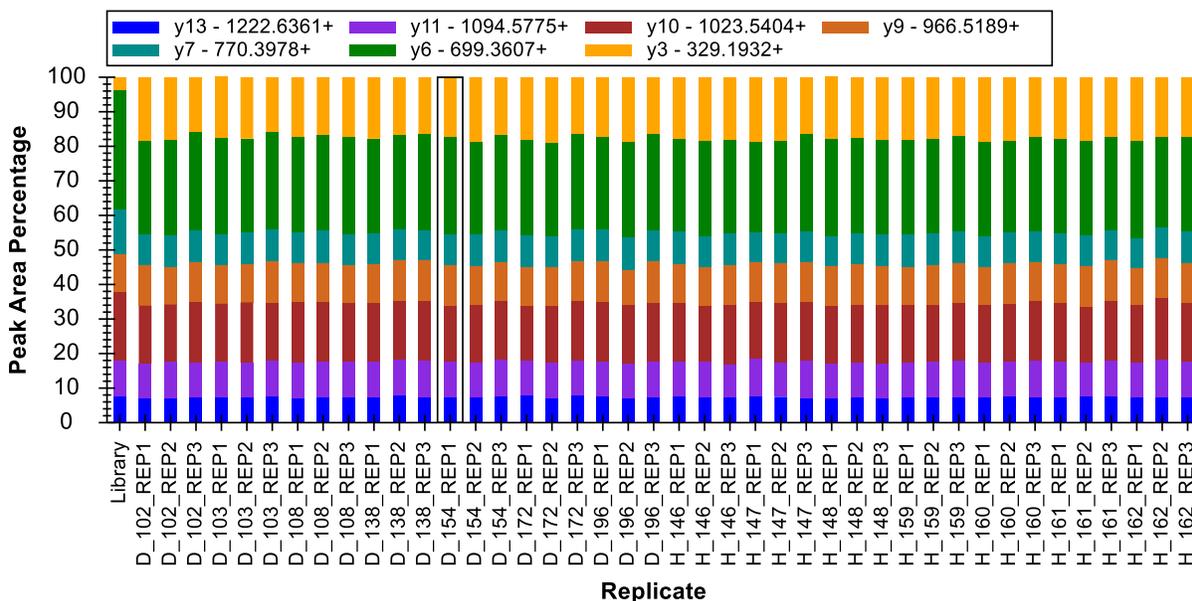


When an interfering signal like this occurs inside the integration boundaries, it will add error to the quantitative measurements. For real quantitative data, you might consider whether you need all 5 transitions for this peptide.

For the current tutorial, however, continue to the next peptide. There you will find strong, clean, and co-eluting peaks for all 7 transitions.



Relative ion abundance is also extremely consistent.



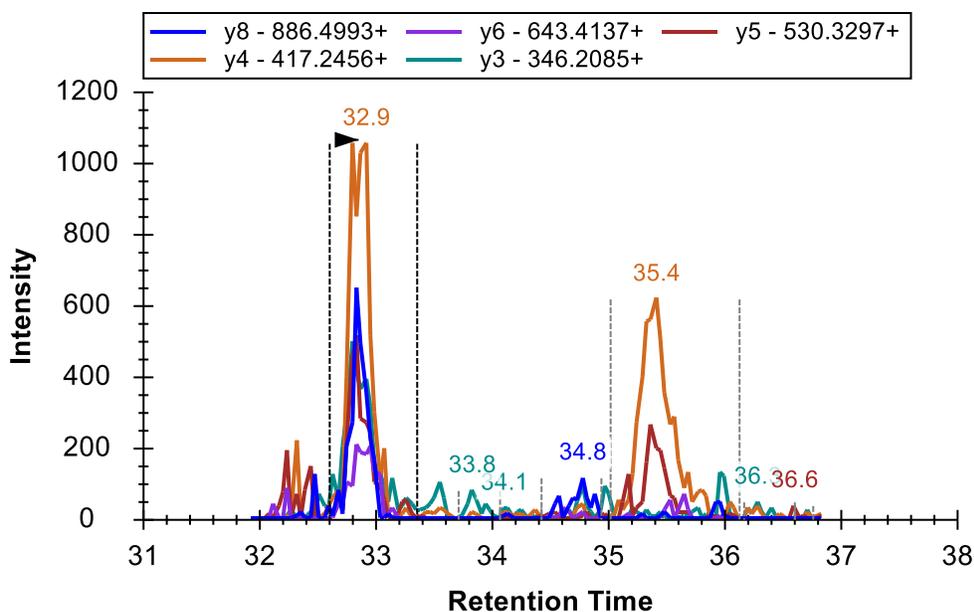
The peak area in y3 is stronger than what was seen for the library spectrum. This might be explained by secondary fragmentation occurring in a triple quadrupole that might not occur with resonance excitation

in an ion trap instrument, if an ion trap generated the library spectrum. In this case, the spectrum comes from a NIST public library, and we have no information on the instrument type that generated it.

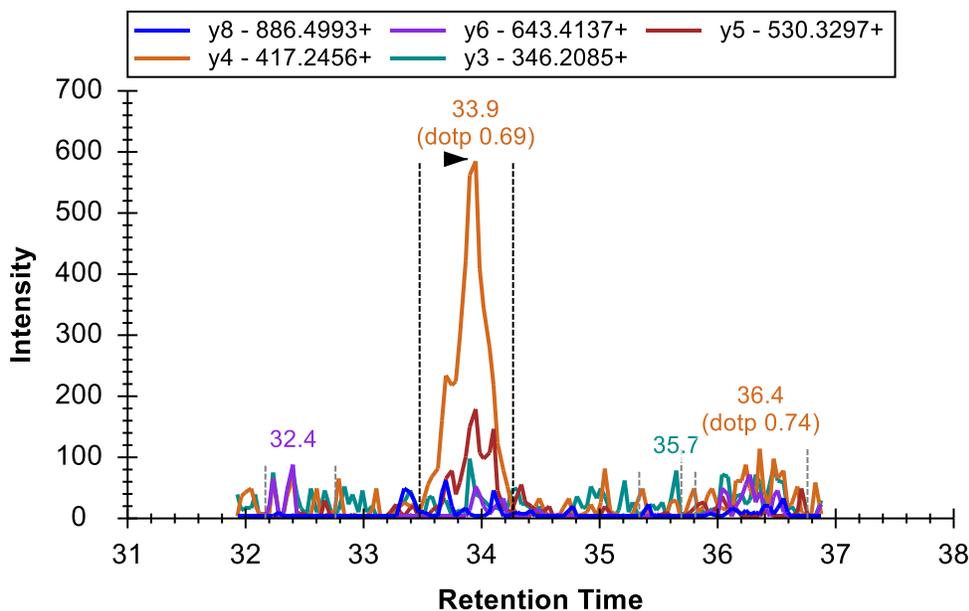
At a glance, this integration looks good enough that you should be able to continue to the next peptide, R.GSYNLQDLLAQAK.L [378, 390], without further inspection. For this peptide, both the **Retention Times** and **Peak Areas** views show issues. To understand what is going on with these measurements, do the following:

- On the **View** menu, choose **Auto-Zoom** and click **None** (Shift-F11).
- Use the **Replicates** dropdown list in the **Targets** view to review the first few replicates until you come to the first problem, D_103_REP3.

For the first 5 replicates the chosen peak is around 32.5 to 33.5 minutes and has signal on all 5 transitions. However, there is also another peak around 35 minutes with signal mostly on y4 and y5 that is surely caused by another peptide. In cases like this, interference that does not coincide with the peptide of interest can add information to the chromatogram “landscape” you see.



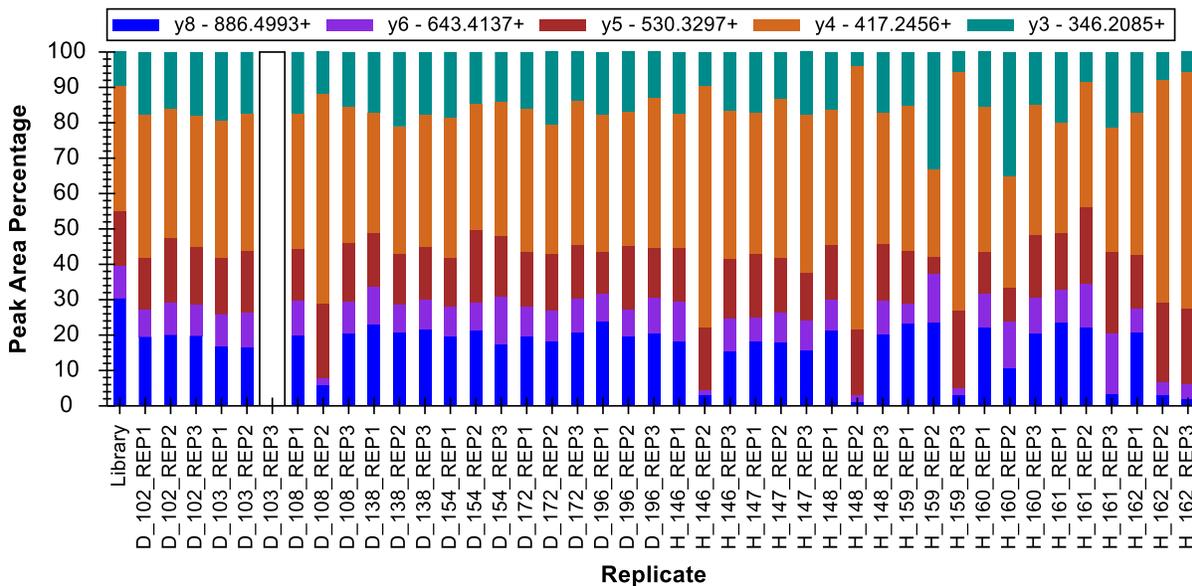
In D_103_REP3, you will see the integrated peak at 33.9 minutes with signal mostly on y4 and y5, but no other candidate peak to the left.



In this case, the scheduling window has failed to capture any signal from the candidate peptide. To remove the incorrect peak do the following:

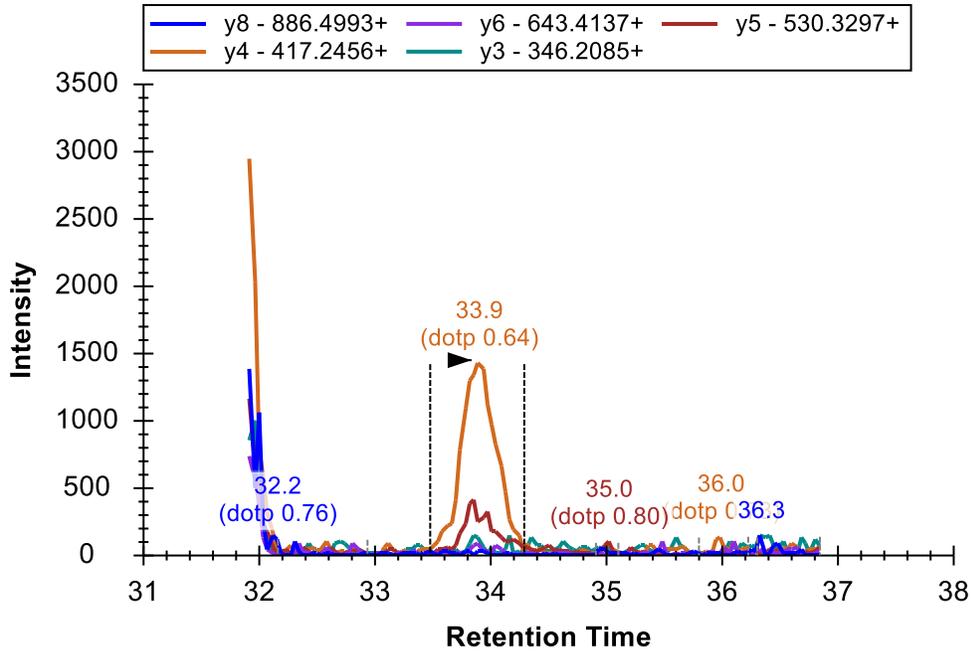
- Right-click on the chromatogram graph, choose **Remove Peak** and click **All**.

This should leave a blank spot in the **Peak Areas** graph.

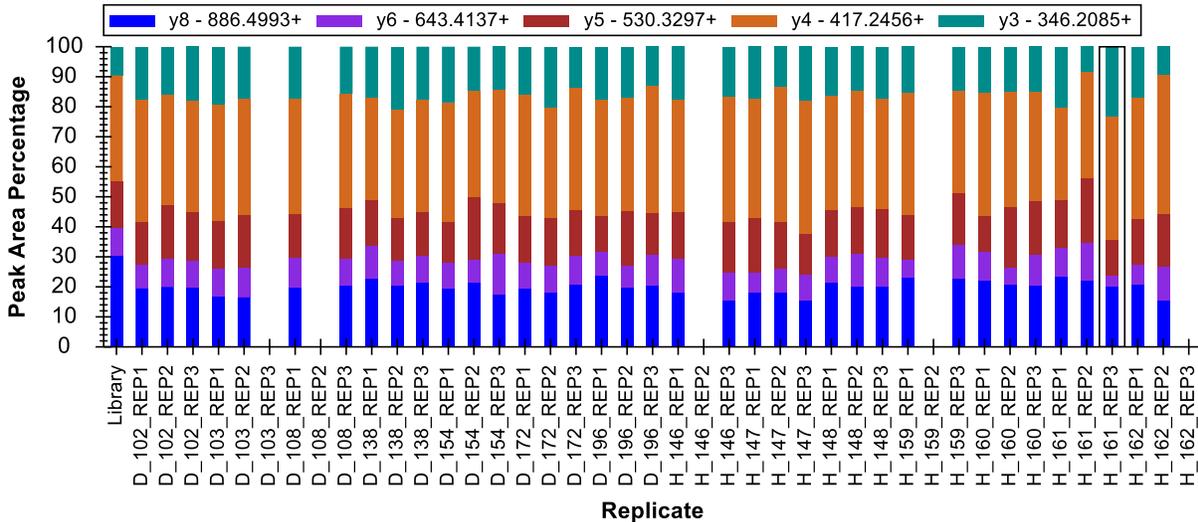


Continuing through the chromatogram plots, you will see this pattern repeated several times. Keep removing the missing peaks. You will even start to see some cases of peak truncation. When you get to

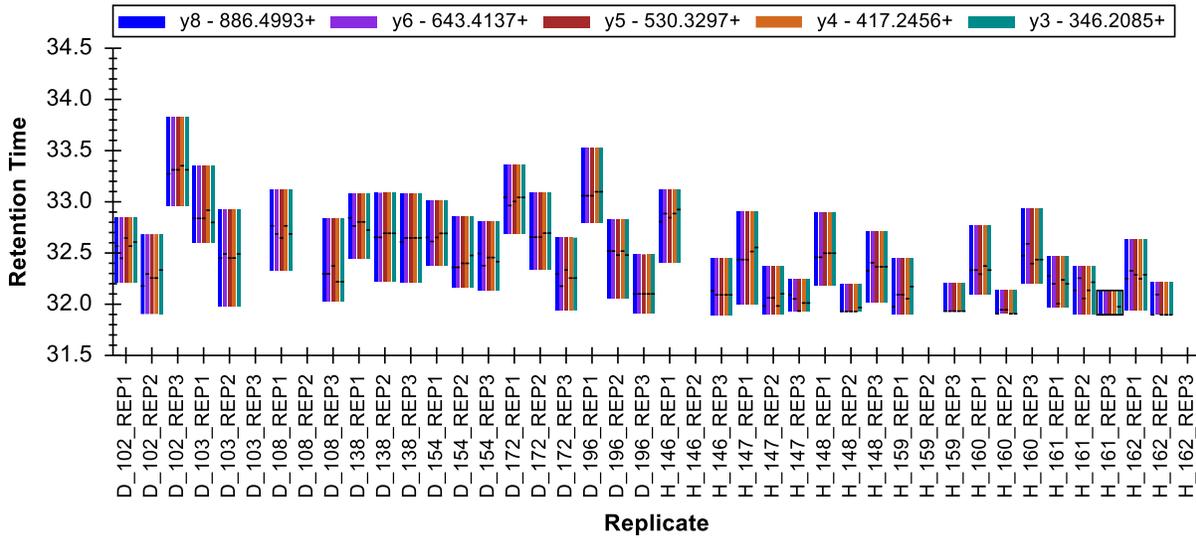
H_148_REP2, you will see a case where Skyline has chosen the peak at 33.9 minutes, but part of the correct peak is still visible.



You can simply integrate this as a truncated peak, as you did above, by clicking and dragging beneath the x-axis. You should end up with a **Peak Areas** plot that looks like this:



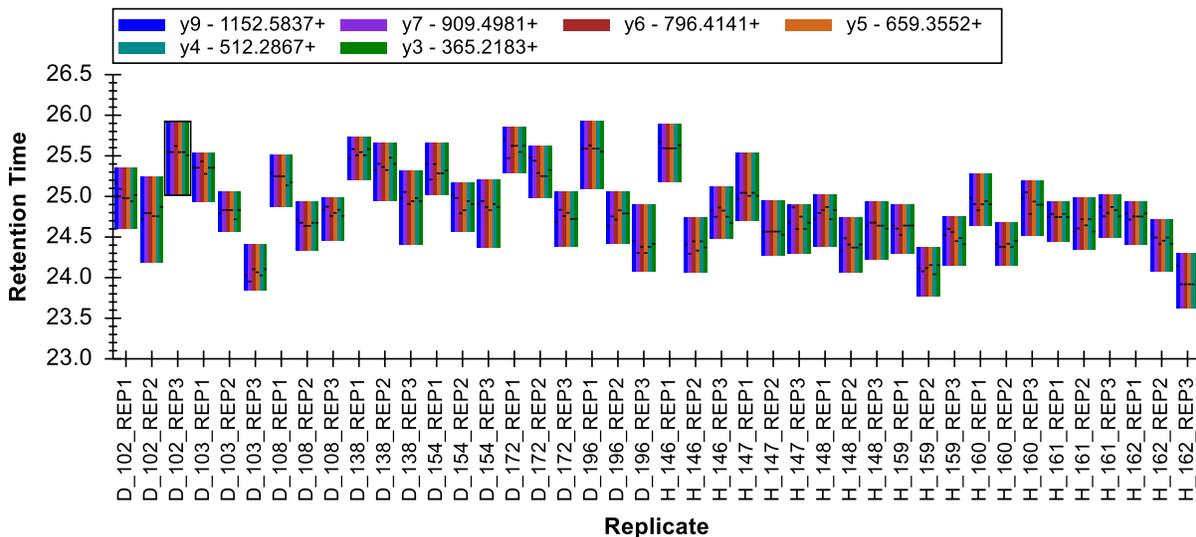
And a retention times plot that looks like this:



Continuing to the peptide R.DIPVNPMLIYR.S [46, 56], you will find no real consistent signal with integrated peaks all over the 5 minute scheduling window. Thus, you may simply delete this peptide.

The next three peptides in the protein NP_001012027 all look fine with relatively stable retention times and relative ion abundances.

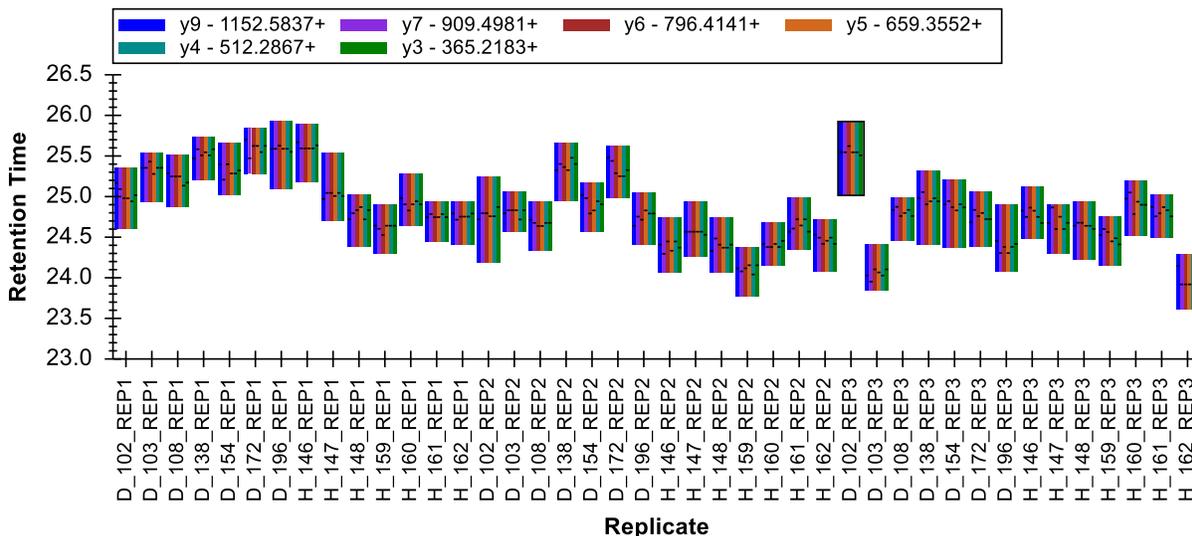
You may feel the variance you observe in the retention times is more erratic than you would expect. This is often the case when the runs are logically ordered. The current order is a result of the way we imported the replicates into the document. For a study of this nature, it was useful to list all diseased subjects first and follow them with all the healthy subjects. All technical replicates for each subject are also grouped together. For the peptide TSDQIHFFFAK the retention times look like:



However, it may also be quite helpful to see the runs in the order they were acquired on the instrument. You can achieve this by doing the following:

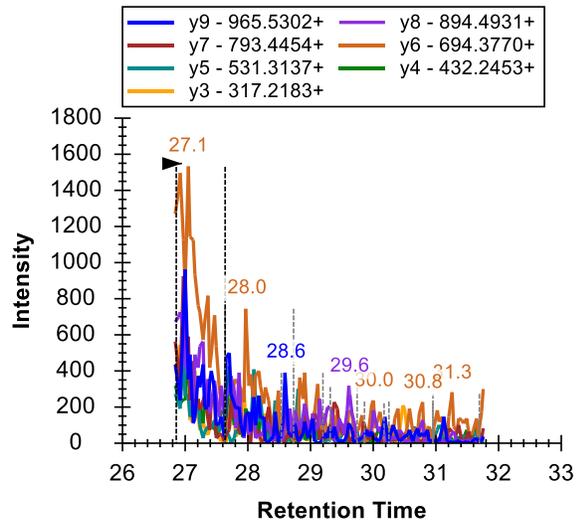
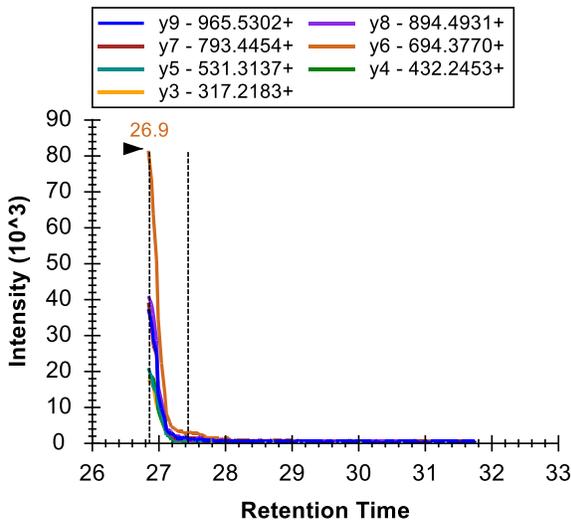
- Right-click the **Retention Times** graph, choose **Order By** and click **Acquired Time**.

The graph will change to look like the following:



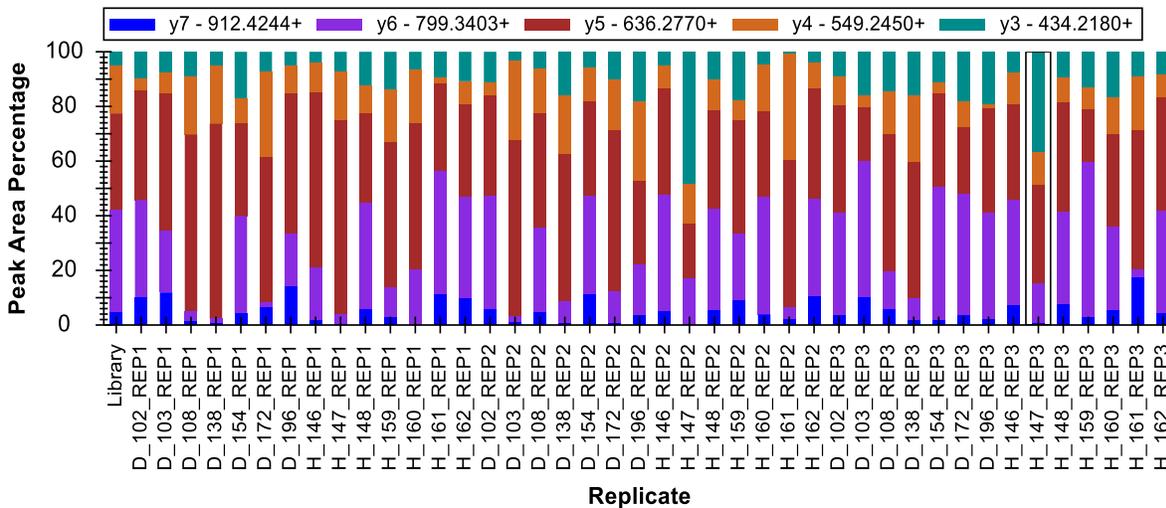
This retention time pattern is now a little less erratic. If you now select through the next two peptides (LQPLDFK, SQLPGIIAEGR), you will see how consistent the pattern is.

The next peptide DFATVYVDAVK has some peak truncation. You should now be able to correct this fairly quickly by clicking on the bars in the **Peak Areas** view that do not match well and clicking and dragging beneath the x-axis in the corresponding chromatogram graphs. There are also some peaks where the chromatography looks very noisy. An example of this is on the right in the chromatograms below. This is because only the very right edge of the peptide peak was measured, and the higher intensity portion is missing.



In a case like this, you want to be very sure you avoid integrating some noise that Skyline fails to identify as a truncated peak. It may just be easier to remove the peaks for these replicates.

Finishing this section with the peptide FGLYSDQMR, you can see in the **Retention Times** view that Skyline has mostly picked a peak close to 19 minutes. However, the **Peak Areas** view shows very poor consistency in relative ion abundance.



If you examine a few of the chromatogram graphs, you will find it hard to imagine consistently integrating the signal at 18.8 minutes with any confidence that all integrated signals come from a single peptide. In this case, you should simply delete the peptide and move on.

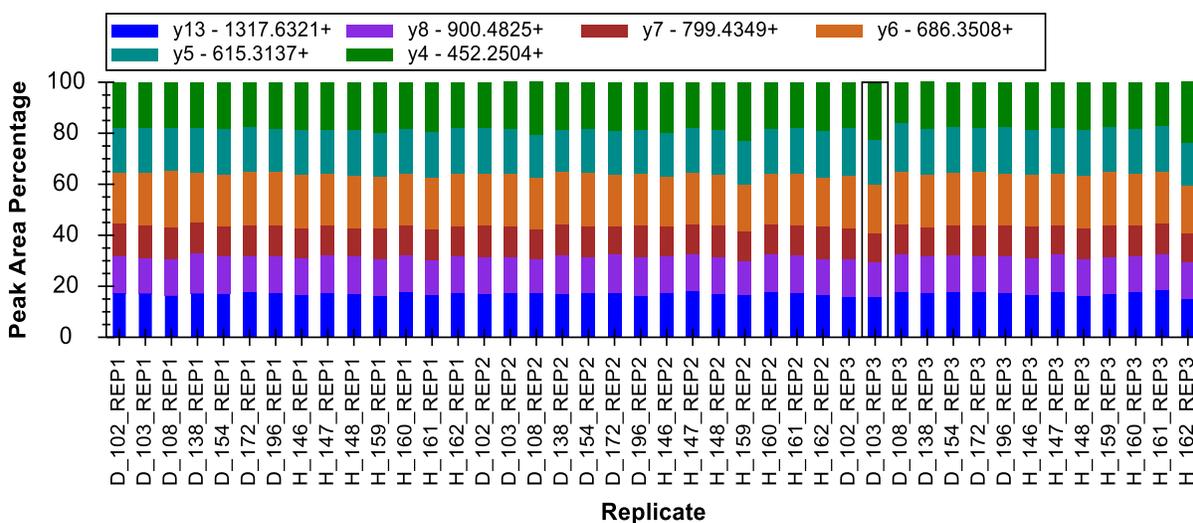
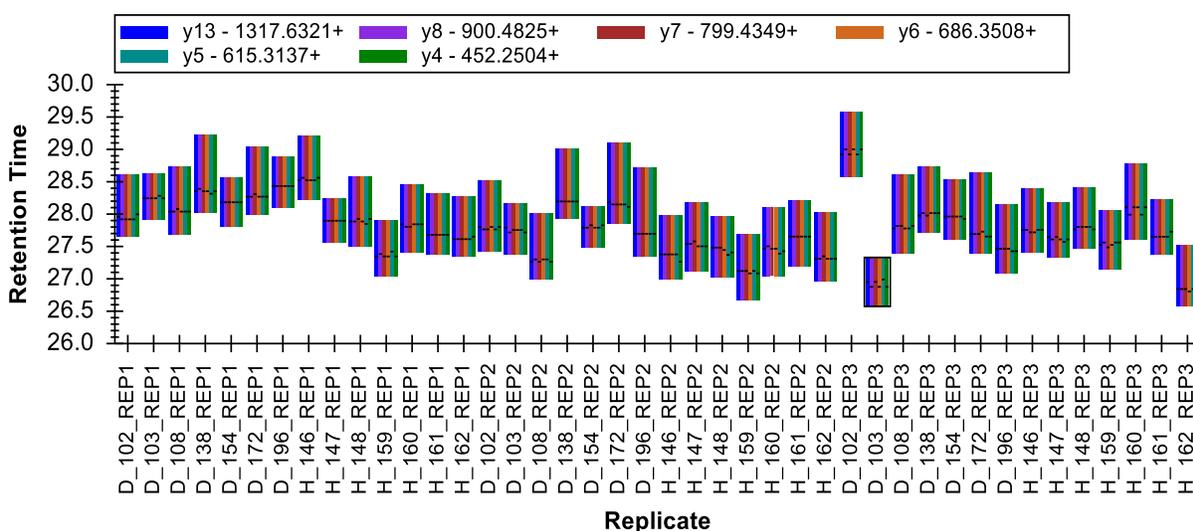
Global normalization standards

This data set lacks stable isotope dilution (SID) peptides but it does contain global standard peptides (from *C. elegans*) that were spiked into the samples. The intent was to use these peptides for peak area

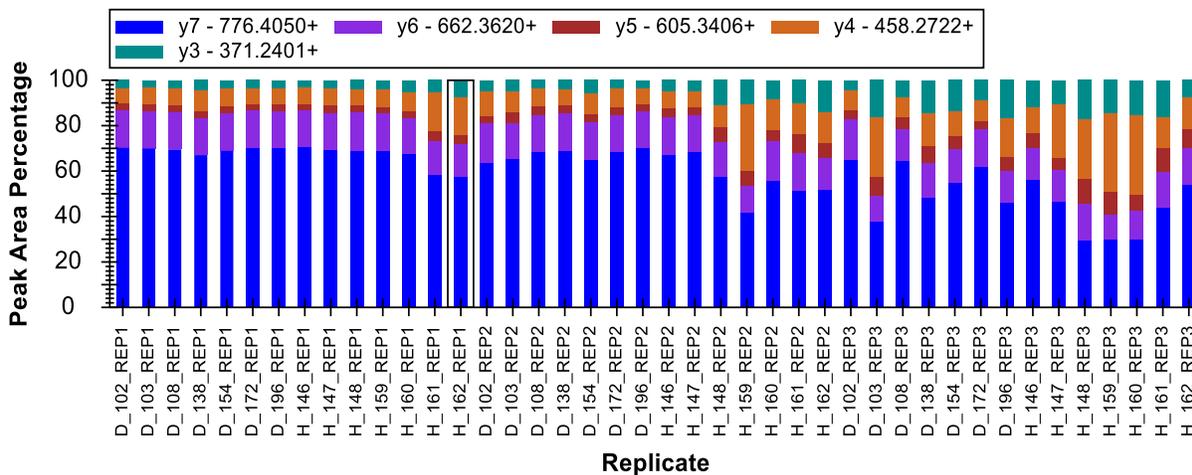
normalization for all other peptides in an effort to reduce the effect of systematic variance in the LC-MS/MS phase of the experiment.² To examine these peptides now, do the following:

- Click in the **Targets** view.
- Press the End key and then the Up-Arrow key on the keyboard.

This should move the selection to AFGLSSPR, the last peptide in the document. The three peptides HLNGFSVPR, VVLGSDATLAYSAFK and AFGLSSPR grouped in the list named “S” are the injected *C. elegans* peptides for this experiment. You can gain immediate confidence in the automatic integration of the last two peptides from the two summary plots we have been examining. The retention times and peak areas are relatively consistent.

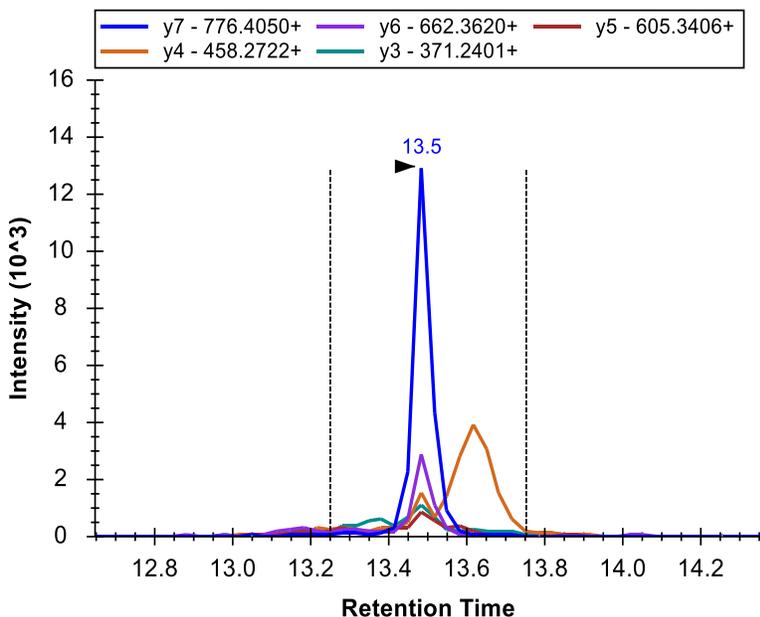


Moving to the peptide HLNGFSVPR, you will observe consistent retention times. However, there is much more variance in the peak areas. The relative area for y4 and sometimes y3 seems to vary throughout the data set.



If you click on some of the peak area bars where y4 is most abundant, you will see obvious interference on y4 and sometimes y3.

- On the **View** menu, choose **Auto-Zoom** and click **Best Peak (F11)**.



To gain an even clearer understanding of the performance of these injected peptides as potential normalization standards, you should first to the following:

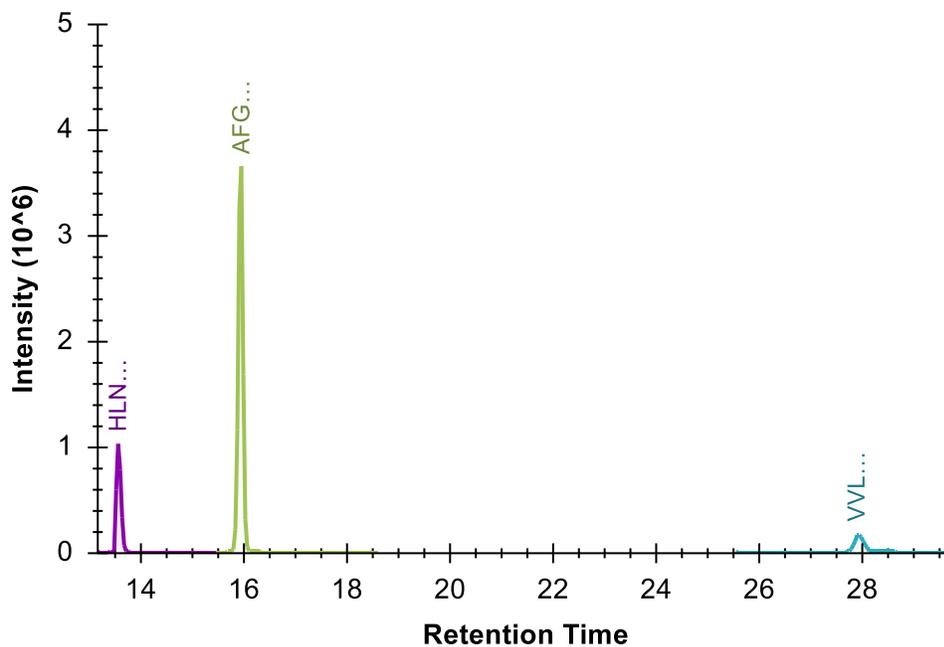
- Right-click the **Peak Areas** view, choose **Normalize To** and click **None**.

Now you can see that the peak area for peptide HLNGFSVPR decreases dramatically from the results acquired for the first sample to the last. The total peak area goes from roughly 6,000,000 to 30,000, spanning a 200 fold difference. See if you are able to use the **Results Grid** (under **View > Results Grid**) with the HLNGFSVPR peptide precursor 513.7776++ selected to determine the exact maximum and minimum total peak areas. Looking at the other two standard peptides, you will see that they too decrease over time (VVLGSDATLAYSAFK 2.3 to 1.1 million and AFGLSSPR 23 to 1 million). All 42 runs are supposed to be essentially technical replicates. While there is clearly systematic signal degradation for all peptides over time in this data set, it is unlikely that very many of them experience 200x or even 20x decreases.

To see where these peptides elute relative to each other do the following:

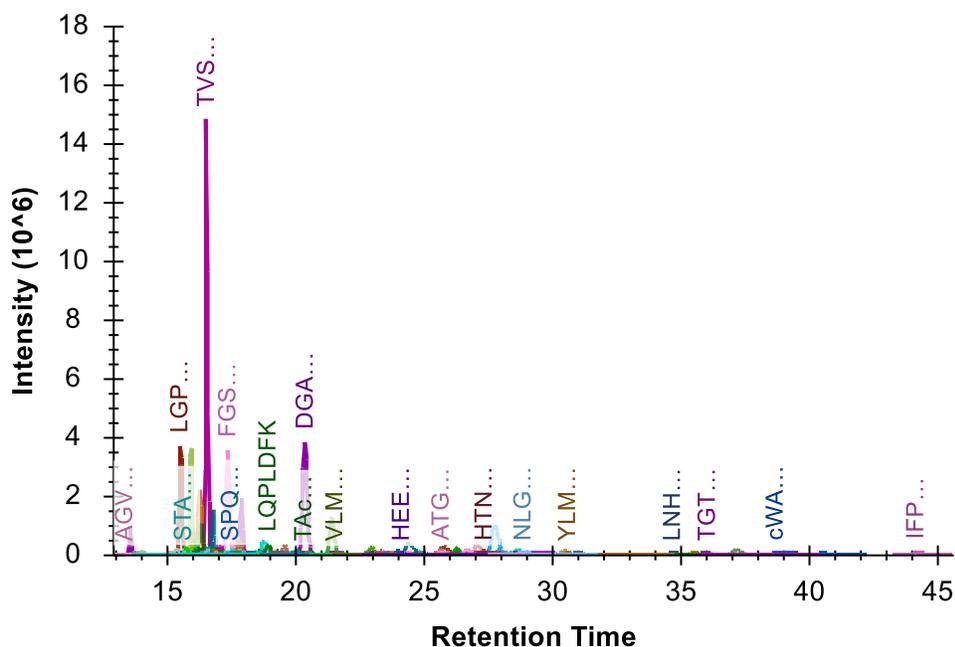
- Select the “S” element preceded by the  icon.

The chromatogram graphs will change to show all three peptides in the “S” list together.



Now to see all of the peptides in the document together:

- On the **Edit** menu, click **Select All** (Ctrl-A).



This makes it pretty clear that the two problematic peptides are both hydrophilic and early eluting. In fact, they are the 3rd and 10th eluting peptides in this experiment. You can do the following to see this in Skyline:

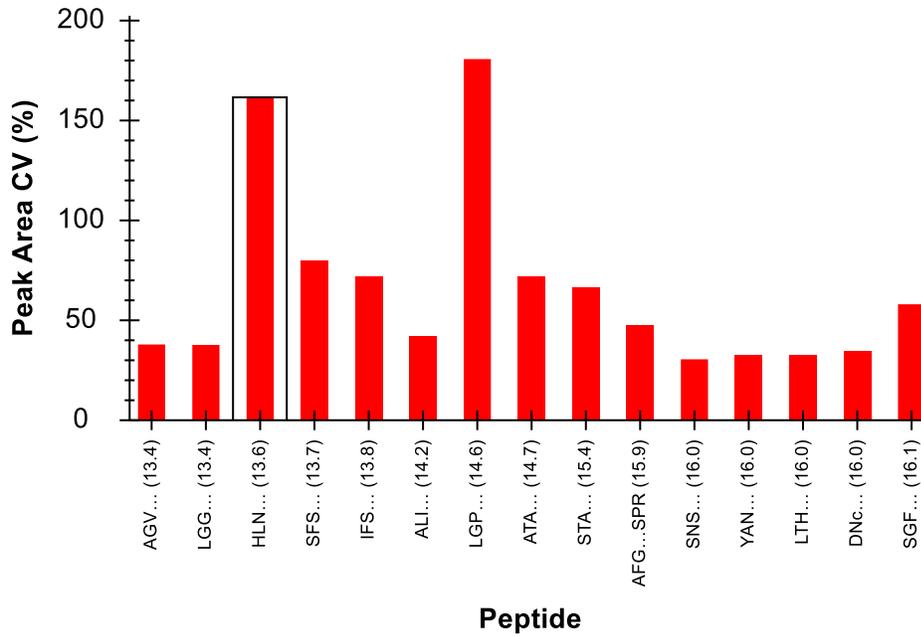
- On the **View** menu, choose **Retention Times** and click **Peptide Comparison** (Ctrl-F8).
- Right-click the **Retention Times** view, choose **Order** and click **Retention Time**.
- Click-and-drag a rectangle around the first 15 or so peptides in the **Retention Times** graph.

Early eluting peptides tend to have more variance. Therefore, they are often not the best candidates for global normalization standards. These peptides may also be impacted by some other factor such as degradation while queuing in the auto-sampler.

For another way to assess how similarly these peptides are behaving to the others around them:

- On the **View** menu, choose **Peak Areas** and click **Peptide Comparison** (Ctrl-F7).
- On the **View** menu, choose **Transitions** and click **Total** (Ctrl-F10).
- Right-click the **Peak Areas** view, and click **CV Values**.
- Click-and-drag a rectangle around the first 15 or so peptides in the **Peak Areas** graph.

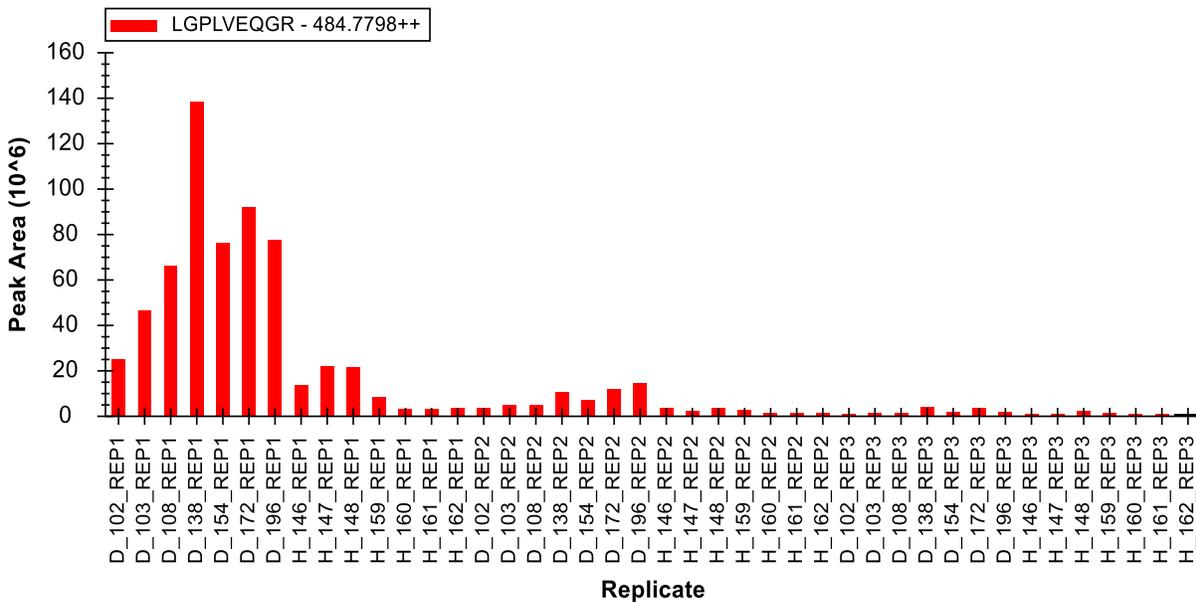
Before trying to understand what you see, you should remember that all of the non-standard peptides are expected to have much higher variance due to the fact that they are measured across 14 subjects. Also, the integration is not fully adjusted for all peptides. For the current data set, the automatically picked peaks looked correct for the standards. However, the HLNGFSVPR peptide has the second highest CV at about 160% while the AFGLSSPR peptide is close to the median for the first 15 peptides with a CV of about 50%. (The VVLSGSDATLAYSFAFK peptide, which is more hydrophobic and later eluting, is 108th or in the 80th percentile with 18.6% CV.)



If you do the following, you should be able to convince yourself that the LGP peptide has a problem very similar to the HLN peptide.

- Click the bar for the LGP peptide in the **Peak Areas** view.
- On the **View** menu, choose **Peak Areas** and click **Replicate Comparison (F7)**.

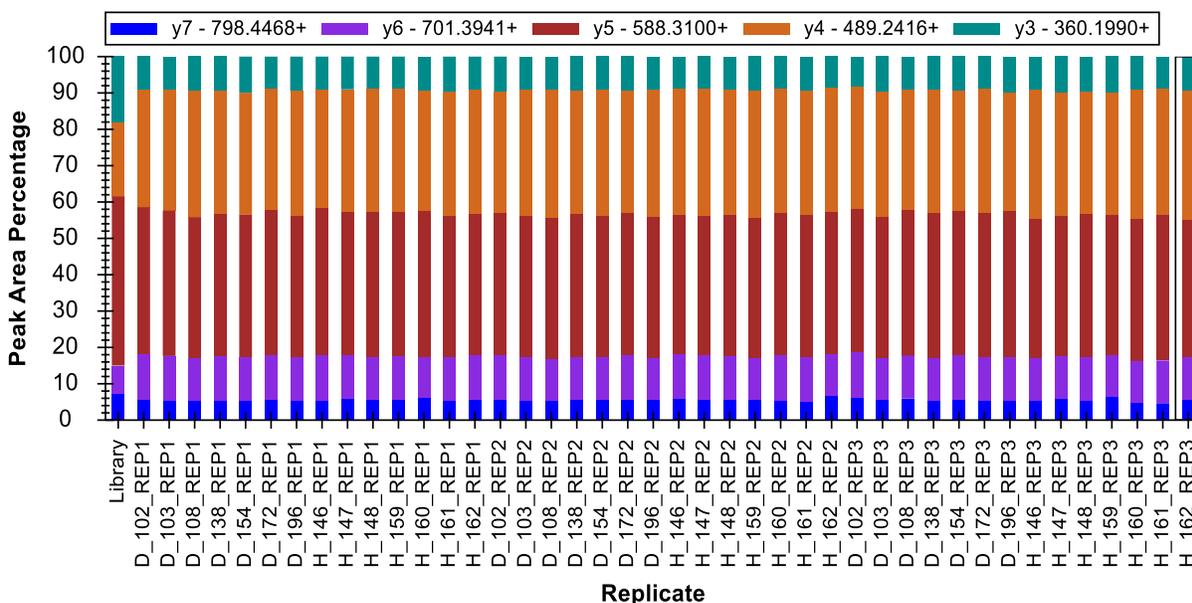
Remember that there are supposed to be three cycles or technical replicates:



To convince yourself that the peaks are correctly integrated, do the following:

- On the **View** menu, choose **Transitions** and click **All** (F10).
- Right-click the **Peak Areas** view, choose **Normalize To** and click **Total**.

You can see that despite the 200x range (140 to 0.7 million) from minimum to maximum, the relative product ion abundance between the runs is extremely stable.



Although this one peptide seems to have similar issues, you should exclude a peptide like HLNGFSVPR from consideration as a global normalization standard. Ideally, this takes place long before you inject your standard peptides into important quantitative data. In this case, you should really only use VVLSGSDATLAYSFAK as a global normalization standard for all other peptides in the document. Further proving this notion is beyond the scope of this tutorial.

Continuing multi-replicate data processing

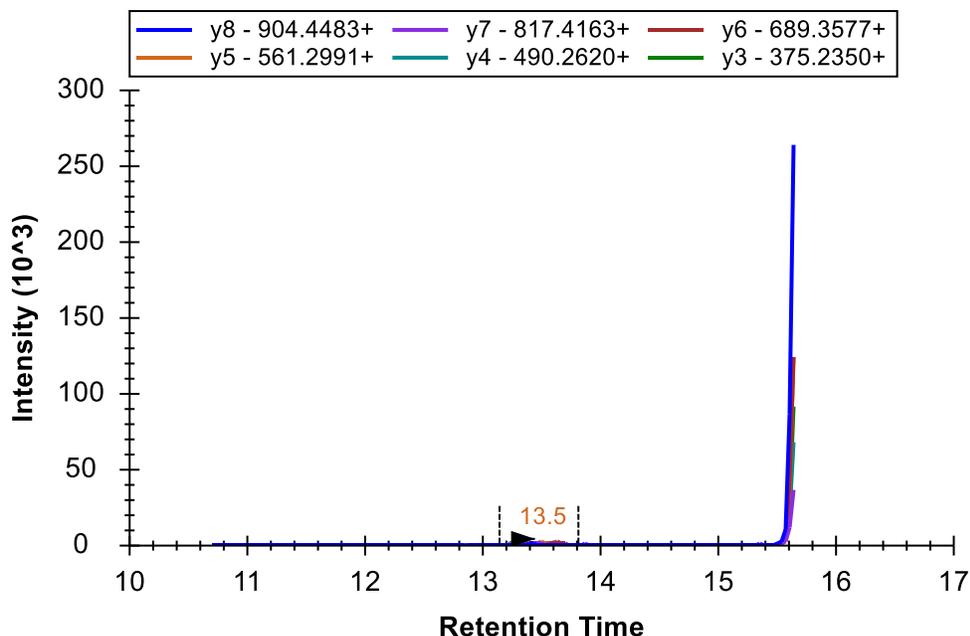
Now you will continue to review and correct peak integrations for this experiment by selecting the peptide DVFSQQADLSR above the “S” standard peptides list.

- On the **View** menu, choose **Retention Times** and click **Replicate Comparison**.

The consistency you will observe in the graphs for both the **Retention Times** and **Peak Areas** should give you the confidence that the peptide has been consistently integrated across all runs. You can continue up the **Targets** list without further review of this peptide. This should also be true of the next two peptides.

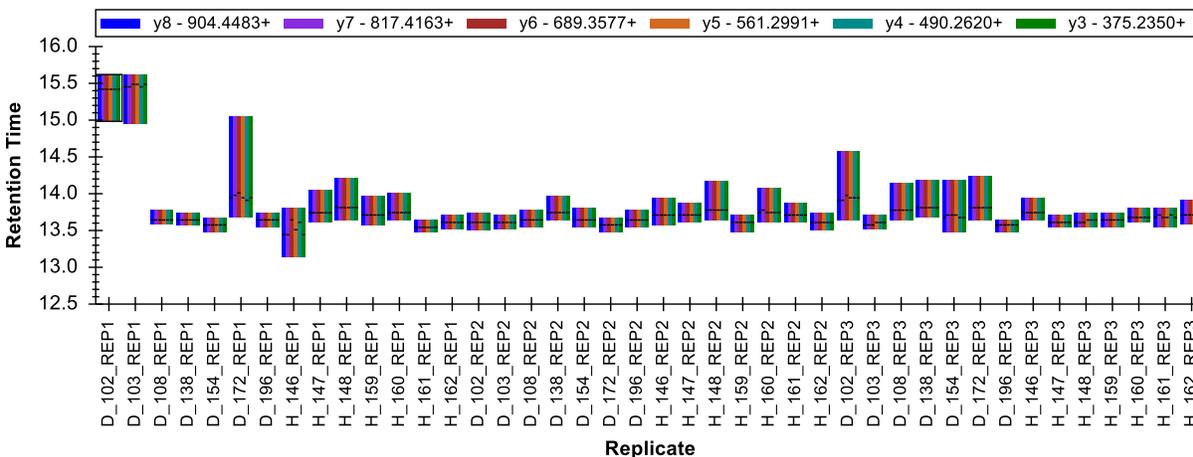
Continue up the list until you reach the peptide IFSQQADLSR. You should notice that every replicate except H_146_REP1 has consistent relative ion abundances in the **Peak Areas** graph. Again, when you

click on the bar for this run, you will see that the peak in H_146_REP1 was not fully captured by the 5-minute scheduling window. (You might need to turn off x-axis zooming – Shift-F11 – to see this.)

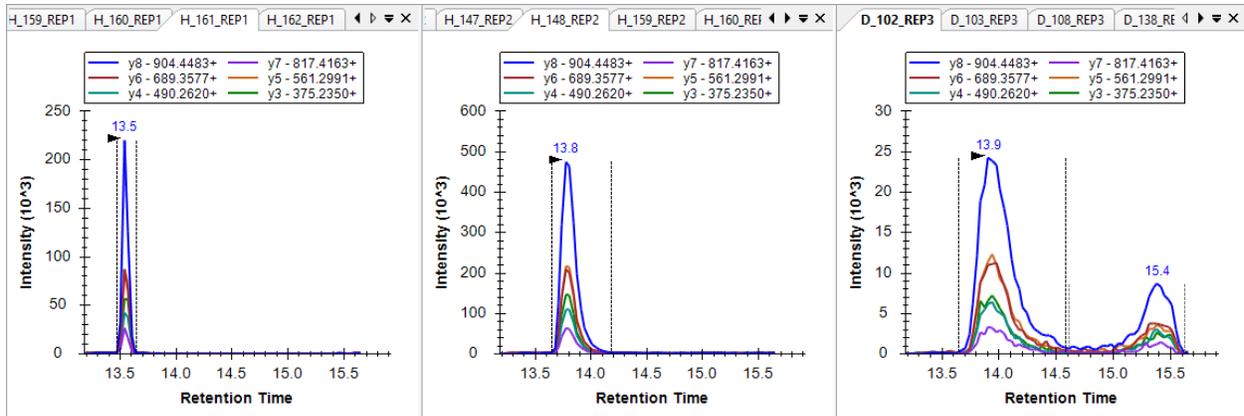


As you did before, you can correct this by clicking and dragging beneath the x-axis to integrate the truncated peak, or by using the right-click menu to remove the peak entirely.

However, this is not the only issue with this peptide. When you look at the **Retention Times** graph, you will observe some unusual differences in the ranges of time being integrated.



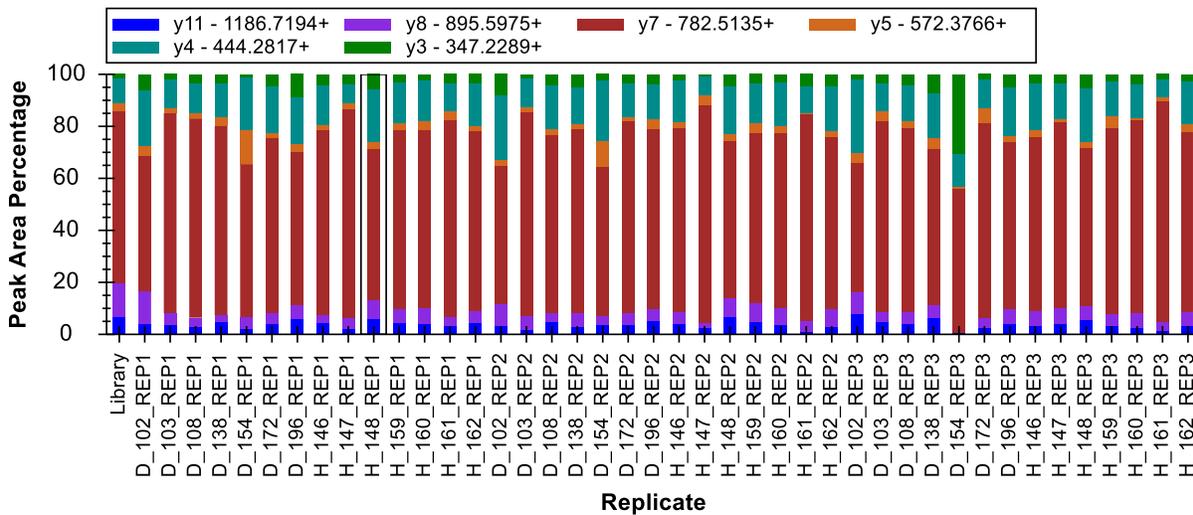
If you review some of the chromatograms for this peptide, you will see that many show a single, very nice peak eluting over 0.2 minutes. Others, however, have some tailing that extends out to about 0.5 minutes while still others show the peptide eluting over 2 minutes with two distinct peaks.



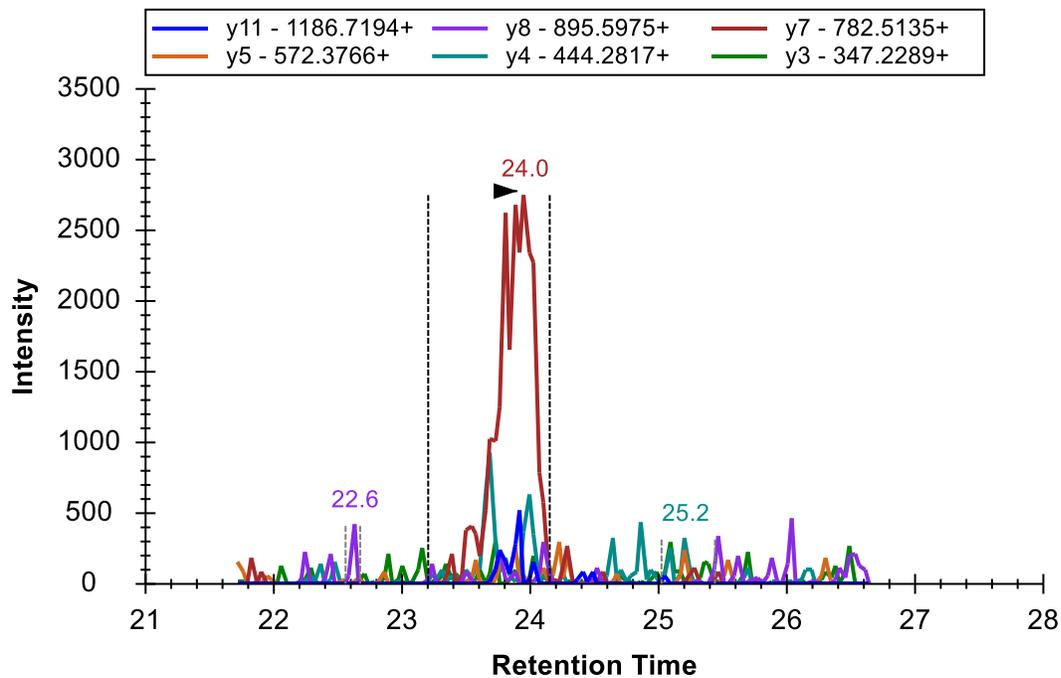
A peptide like this may be very difficult to use for quantification, especially without a matching SID peptide. You should probably delete this peptide from this experiment. At the very least, you should remove the peaks from the runs that show the double elution profile.

Continuing up the peptide list, you should find that the next 7 peptides show good enough consistency in the summary plots that you should not need to correct any of the default peak picking performed by Skyline or even give these peptides much more than just a glance.

When you reach the peptide **MLSGFIPLKPTVK**, however, you will see much more variance in the **Peak Areas** graph.



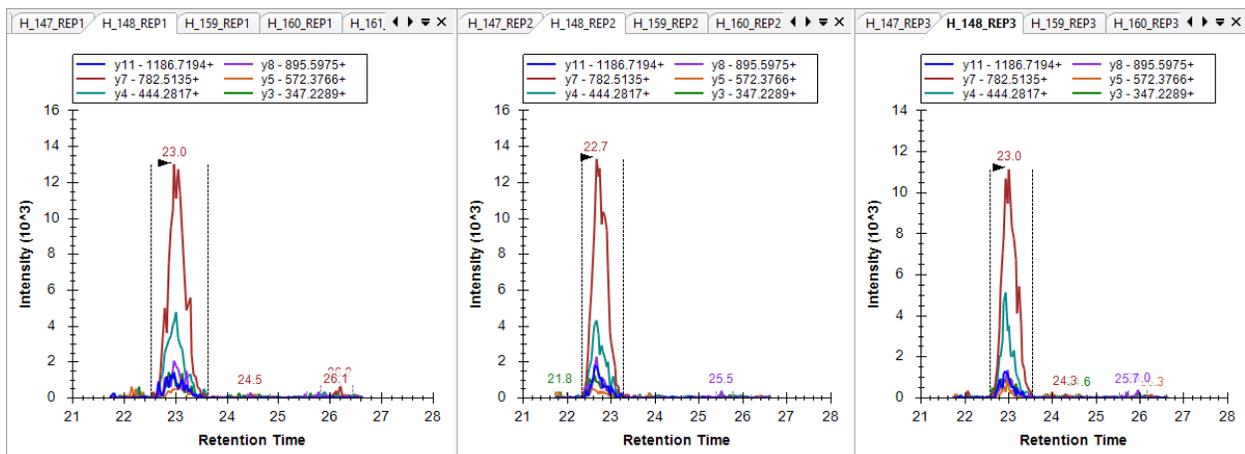
The total peak area for this peptide is dominated by the y7 ion. Although this is also reflected in the library spectrum (as shown at the far left of the above graph), visual inspection of the chromatograms makes it difficult to discern if anything is actually co-eluting with the y7 ion (see below).



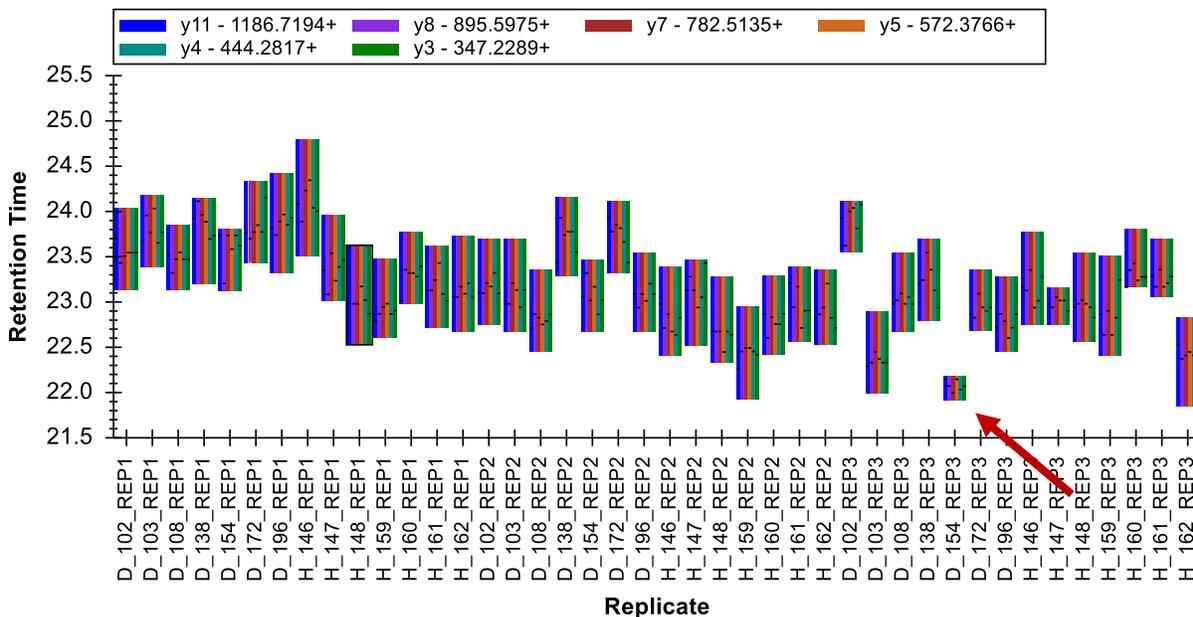
However, it is also clear that Skyline is, for the most part, integrating the same peak for the same molecule in each run. To convince yourself there is evidence of more than just the y7 peak, you should look at the most intense measurements. You can do this using the following steps:

- Right-click the **Peak Areas** graph, choose **Normalize To** and click **None**.
- Click on the three obviously most intense peaks (which all come from the subject H_148).

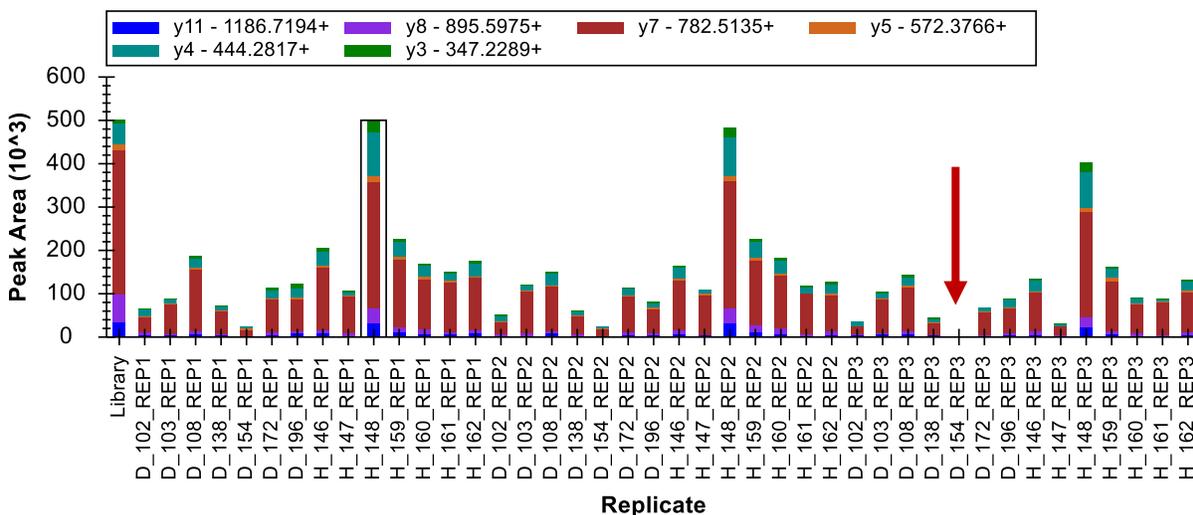
The chromatogram peaks you see in these three plots should be enough to convince you that the peptide causing the peak in y7 between 23 and 24 minutes does, in fact, cause the co-eluting signals on all 6 monitored transitions when peptide abundance is high enough.



Rather than deleting this peptide, you will want to correct the one case where Skyline has chosen the wrong peak. You can see this in the **Retention Times** plot, as a narrower band occurring below 22.5 minutes.

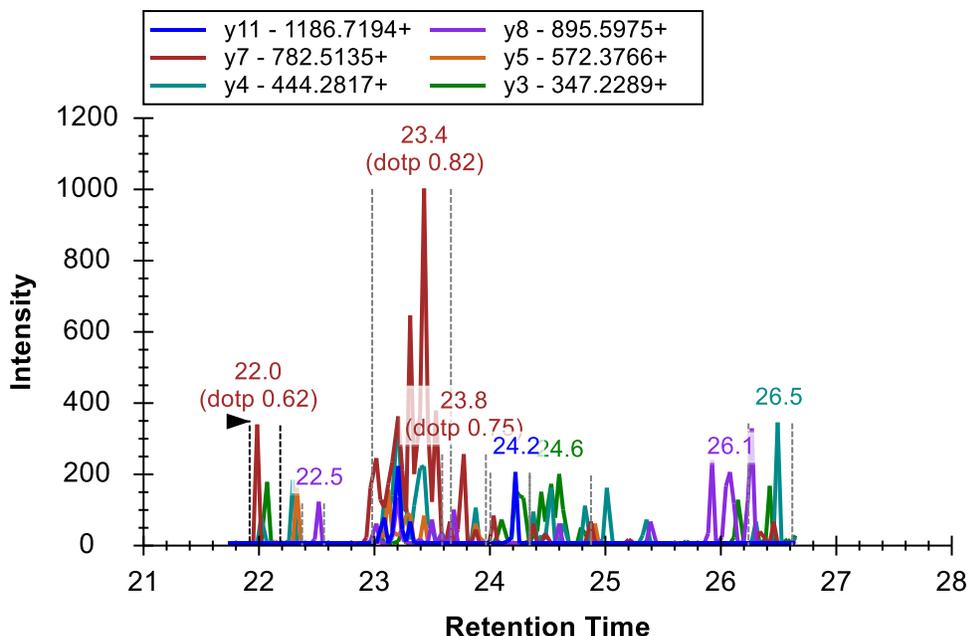


And also in the **Peak Areas** plot, as the run with no visible integrated signal.



You can click on the bar in the **Retention Times** view to activate the chromatogram graph for this run. You will see that the run retains some very low signal on the y7 transition. You would never choose this

transition alone to represent a peptide without the knowledge you have gained from the surrounding runs.



To correct the integration of this run, do the following:

- In the graph for D_154_REP3, click the peak annotation “23.4 (dotp 0.82)”.

Before continuing your review of peak integration, do the following to return the **Peak Areas** graph to showing the relative ion abundances:

- Right-click the **Peak Areas** view, choose **Normalize To** and click **Total**.

You will see the corrected peak now appears much more consistent with the other runs in the summary plots.

Continue to the peptides above. At the peptide GMYESLPVAVK, you will see such poor consistency in the summary plots that you should likely just delete this peptide and not bother adjusting the integration. The peptide ETGLMAFTNLK requires adjustment only in one run. The run should be very obvious in the **Peak Areas** graph. By this point in the tutorial, you should be well equipped to understand what went wrong, and how to fix the problem.

After making this correction, if you are still not fully convinced that the same peak for the same peptide is being integrated in all cases for the ETGLMAFTNLK peptide, you can quickly review all chromatogram graphs, by doing either of the following:

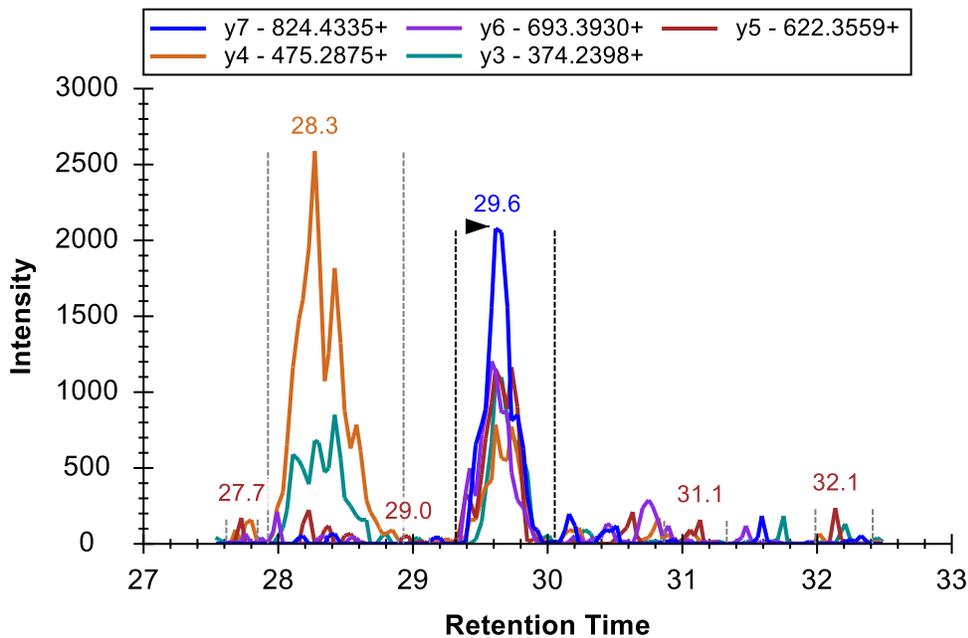
- Click on the **Replicates** dropdown list at the top of the **Targets** view.

- Use the up and down arrow keys to traverse the runs, activating chromatogram graphs.

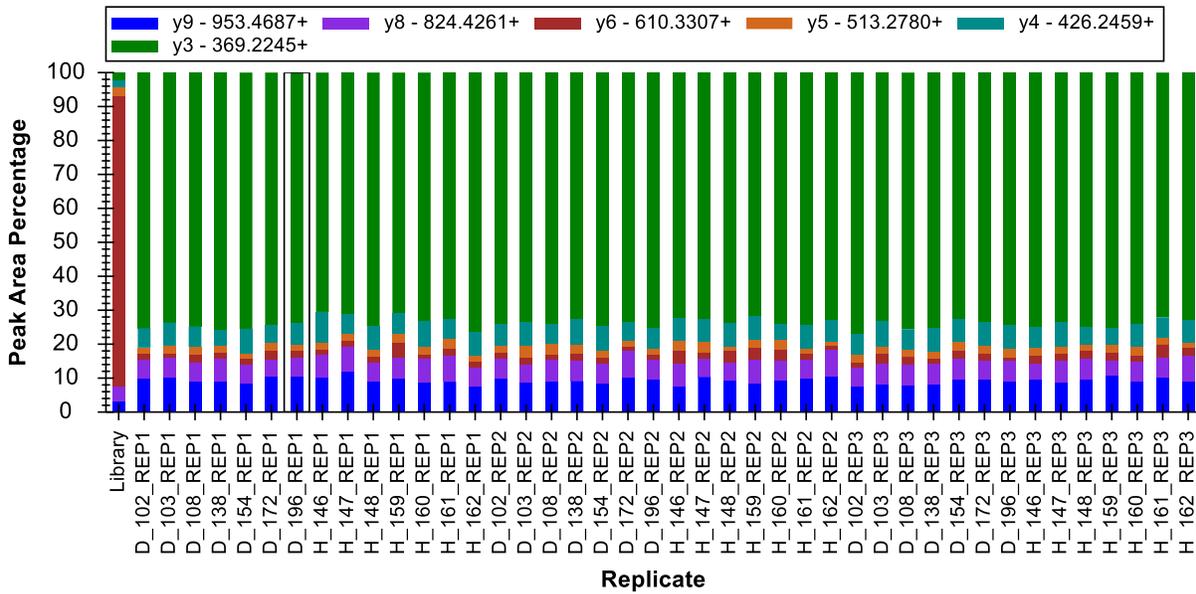
Or

- Press Ctrl-Up-Arrow or Ctrl-Down-Arrow to traverse the runs.

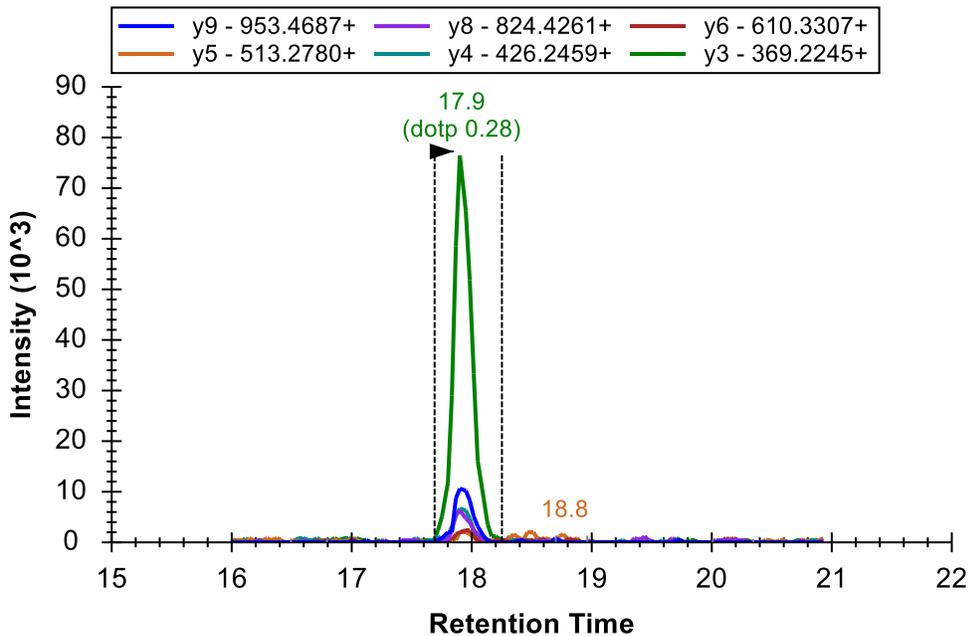
You will see in many of the graphs that there is a peptide interfering on the y3 and y4 transitions about 1.5 minutes before the chosen peak. This gives you added confidence in the consistency of the integrated peak.



Continuing up the **Targets** list, you will find several peptides (YANVIAYDHSR and TDEDVPSGPPR) where the difference between the relative ion abundances of the measured peak and the matching library spectrum is pronounced. You can see these differences in the **Peak Areas** graph:



You can also observe this in the chromatogram graphs where Skyline displays the low dot product values “(dotp 0.28)” beneath the peak retention times.



Skyline displays the “dotp” value like this when there is another peak on the graph with a higher dotp value. In the case above, you cannot see the peak with the better dotp value because it is down near the noise level for the chosen peak. To see it, you can do the following:

- Click the chromatogram graph of interest to activate it.

- Place the mouse cursor to the left of the y-axis.
- Use a mouse scroll wheel or a scrolling gesture on a track pad to scroll upward.

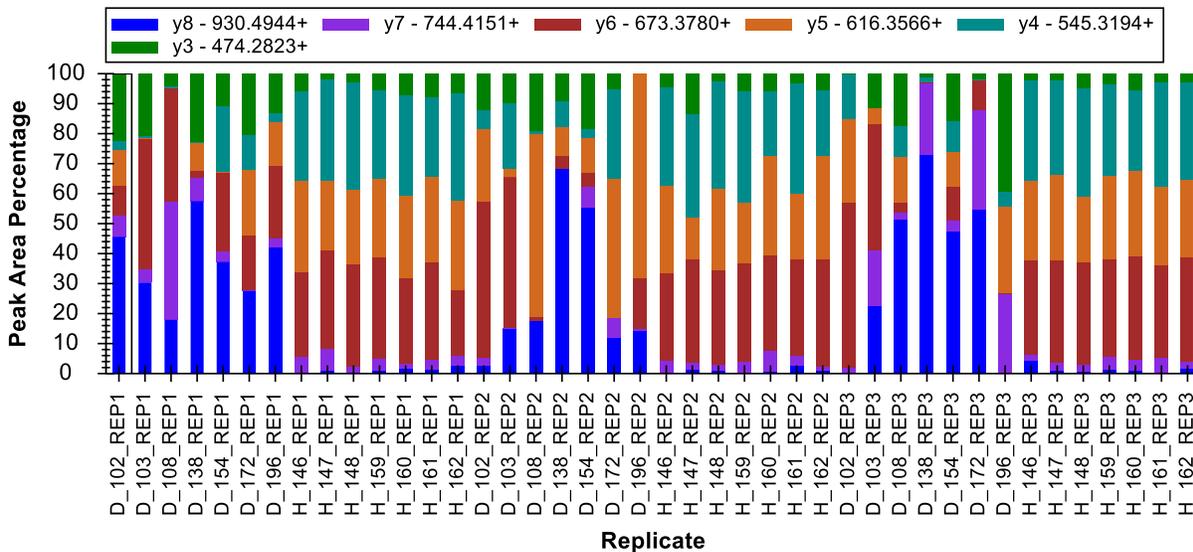
This will zoom the graph y-scale until you can see the smaller peak with a better dotp value.

If you had high confidence that the library spectrum was collected on a mass spectrometer similar to the one used in this experiment, this would certainly cause you to suspect that the integrated peak is not actually measuring the peptide in question. In this case, however, you don't really know the source of this spectrum and there is likely no other option for measuring this peptide with this sample prep.

From the run-to-run consistency in the **Peak Areas** and **Retention Times** graphs, you should be confident the same peptide is being measured in all runs via the 6 co-eluting transitions. Despite the fact that the most abundant ion is the y3 ion with low selectivity, you can delete this transition to more closely examine the consistency and co-elution from the other ions (where y9 is the second most abundant ion). When you have convinced yourself of the quality of this peak, you can use the **Undo** button (Ctrl-Z) to return the deleted y3 transition to the Skyline document.

Among the 7 peptides above TDEDVPSGPPR, you should see only one incorrectly integrated peak for the peptide SPQGLGASTAEISAR. This issue should now be relatively easy for you to identify and correct.

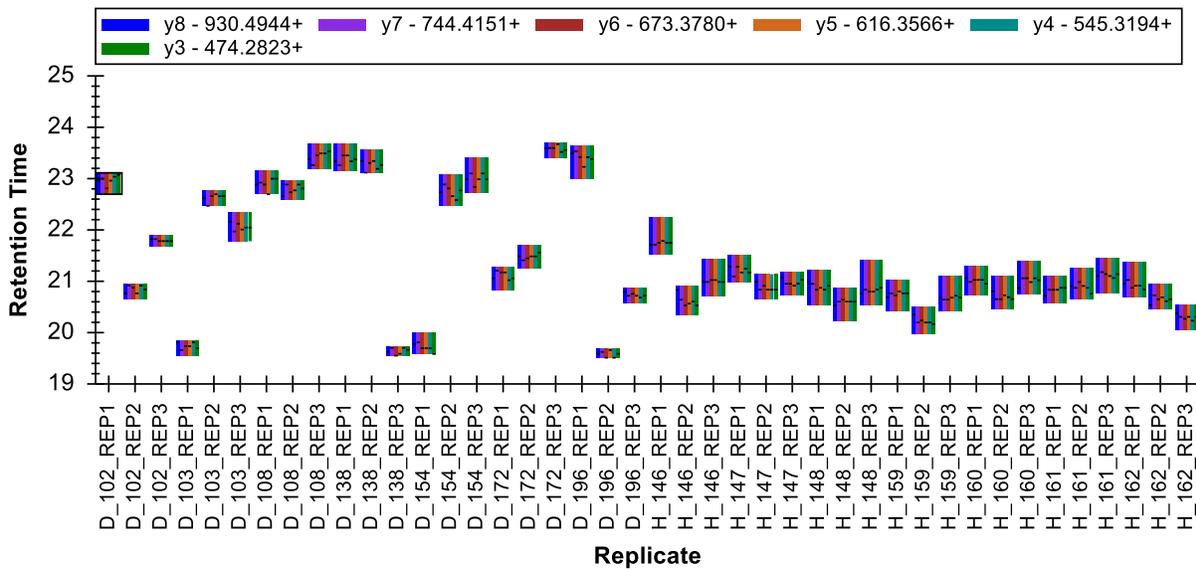
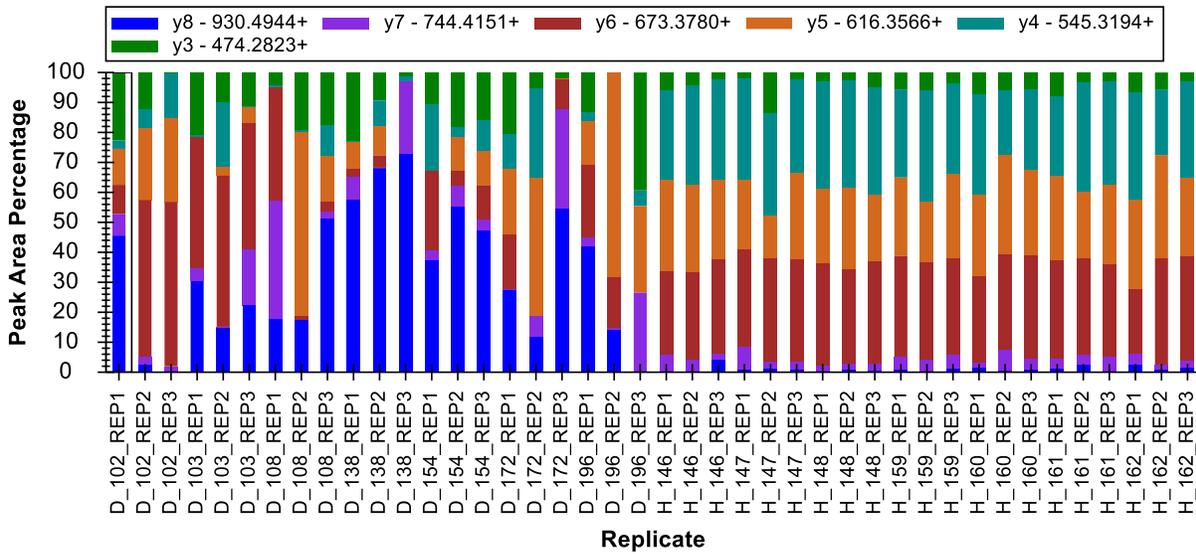
This will bring you to the peptide CSSLWAGAAWLR. Based on the variance you will observe in the summary plots (see below), you may initially be tempted to delete this peptide and move on:



However, if you look closely you will see that there are regions of consistency among the healthy subjects. To see this expressed more clearly, do the following:

- Right-click the **Peak Areas** graph, choose **Order By** and click **Document**.

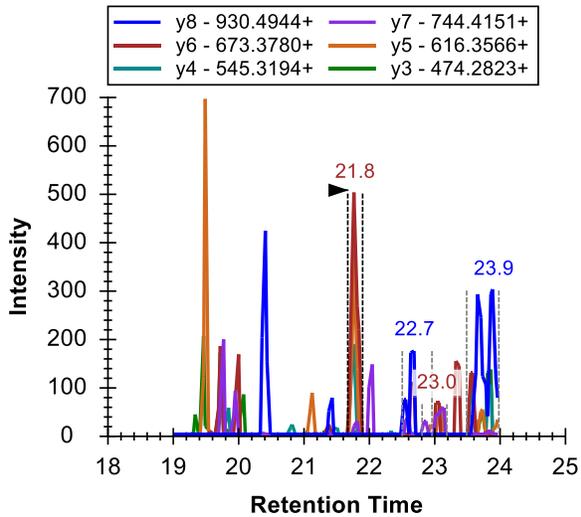
This will change the graphs to look like:



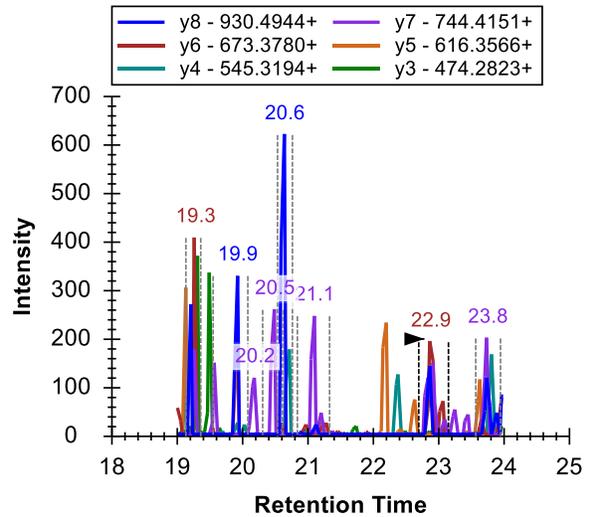
Making it clearer that a consistent peak is being integrated for the healthy subjects but not for the diseased subjects. If you review the peak for the healthy subjects, you can see that it is fairly low intensity and mostly relies on co-elution of the y4, y5 and y6 ions. This co-elution is probably caused by a peptide. However, it is difficult to determine if this peptide is actually CSSLLWAGAAWLR or not.

Looking at the chromatograms for the diseased subjects, you will sometimes find something that may be a peak around 21 minutes for the y6 transition. However, more often than not, it is difficult to observe any consistent elution pattern:

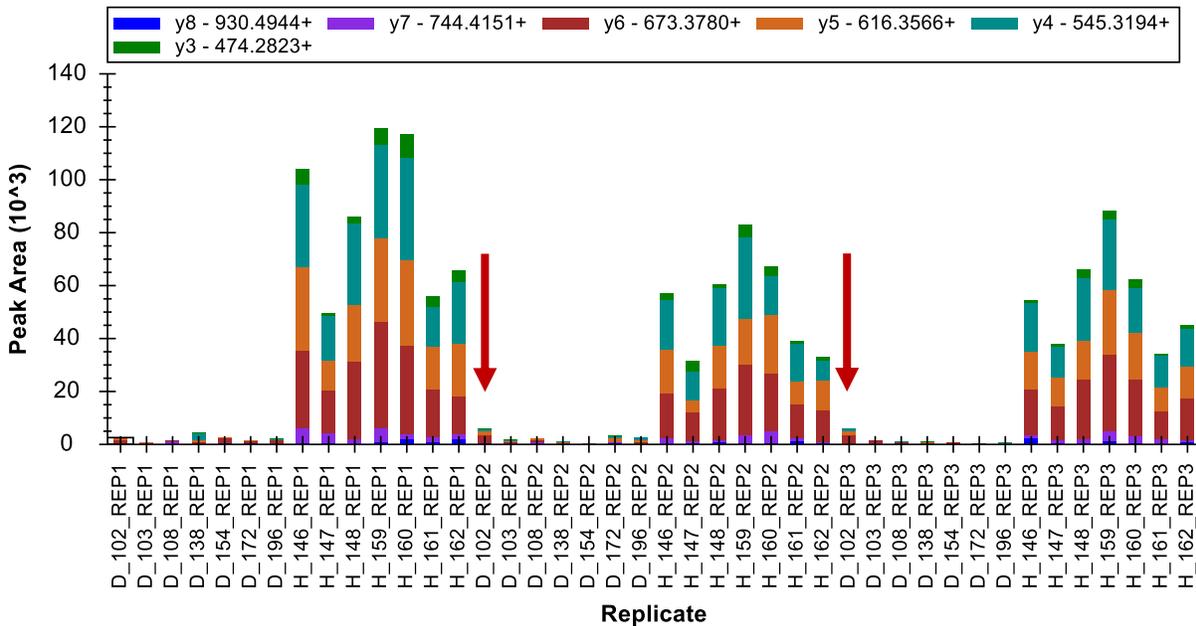
D_102_REP3



D_108_REP1



There is, however, clearly no visible peak in the chromatograms for the diseased subjects that could produce a peak area like those peaks observed in the healthy subjects. With the experience you have now acquired with peak drift and truncation, you should feel confident that the peak is not simply drifting outside the measured range for all diseased subjects, but not healthy subjects. Randomization of injections might also increase your confidence in this. For the current study, however, the three replicate cycles should be enough. Once you have done your best to correct integration, you might also note that two of the diseased runs with more visible peaks for this peptide follow immediately after runs of healthy subjects (D_102_REP2 and D_102_REP3). This observation might suggest a carryover effect.



In the end, these transitions may represent one of the strongest biomarker candidates in this document. You could now order a synthetic peptide to increase confidence that you are, in fact, measuring the CSSLWAGAAWLR peptide. You could also try acquiring MS/MS spectra across this peak (using parallel reaction monitoring – PRM) to see if it can be positively identified by a peptide search engine. Finally, if that failed, you could try *de novo* sequencing on the MS/MS spectra to identify what is causing the peak in the healthy subjects. In targeted proteomics, it is not always necessary to start with positive identification. Perhaps more similar to SDS-PAGE gel bands, finding a difference can initiate discovery.

At this point, you should be able to make your way through the remaining peptides, correcting integration and removing poorly behaving peptides. This should take you less than an hour. You will find peptides that perform well and which were entirely captured by the scheduled retention time window need very little correction. You can often tell this at a glance from the **Peak Areas** and **Retention Times** graphs. For peptides that perform poorly or where many peaks were truncated or missing from the scheduling window, you may simply wish to delete the peptide. However, be careful not to jump to this conclusion too quickly or you might miss some interesting observations.

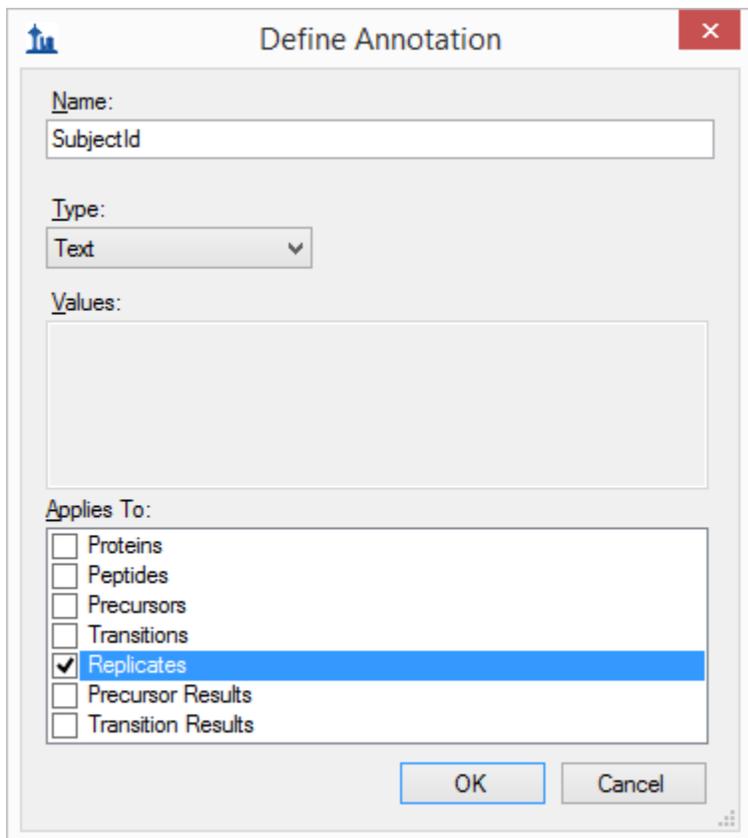
Preparing for statistical analysis with replicate annotations

Once you have completed careful validation and correction of any integration issues in your Skyline document, you will want to turn your attention to gaining a better understanding of what sort of differences in peptide abundance you might find between your experimental conditions. Making such inquiries, whether in Skyline or its external statistical tools, requires some further classification of the measured samples, generally referred to as “replicates” in Skyline. For such classification Skyline provides replicate annotations. In this tutorial, you will use three replicate annotations: SubjectId, BioReplicate and Condition.

To define the SubjectId annotation, perform the following steps:

- On the **Settings** menu, click **Document Settings**.
- Click the **Edit List** button in the **Annotations** tab of the **Document Settings** form.
- Click the **Add** button in the **Define Annotations** form.
- In the **Name** field of the **Define Annotation** form enter 'SubjectId'.
- In the **Applies to** list, check **Replicates**.

The form should look like this:

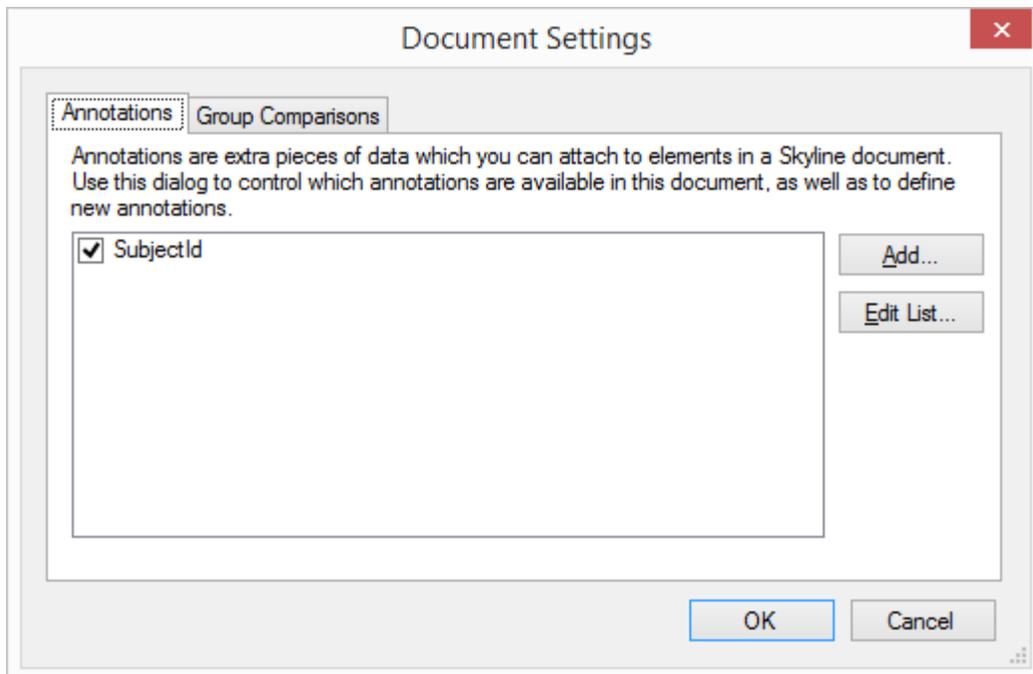


The screenshot shows a dialog box titled "Define Annotation" with a close button (X) in the top right corner. The dialog contains the following fields and controls:

- Name:** A text input field containing "SubjectId".
- Type:** A dropdown menu currently showing "Text".
- Values:** An empty text area.
- Applies To:** A list of checkboxes with the following items:
 - Proteins
 - Peptides
 - Precursors
 - Transitions
 - Replicates (highlighted in blue)
 - Precursor Results
 - Transition Results
- Buttons:** "OK" and "Cancel" buttons at the bottom right.

- Click the **OK** button in the **Define Annotation** form.
- Click the **OK** button in the **Define Annotations** form.
- Check the 'SubjectId' annotation you just created.

The **Document Settings** form should look like:



- Click the **OK** button.

Although this tutorial will not go so far as to cover statistical analysis methods available in the Skyline External Tool “MSstats”, it is well suited to such analysis. This particular data set has been used to demonstrate the use of MSstats in many courses and workshops. If you are interested in exploring what MSstats has to offer for this type of analysis, you can get the two other annotations it requires by installing it into Skyline as follows:

- On the **Tools** menu, click **Tool Store**.
- Select ‘MSstats’ from the list of tools.

The **Install from Tool Store** form should look like:

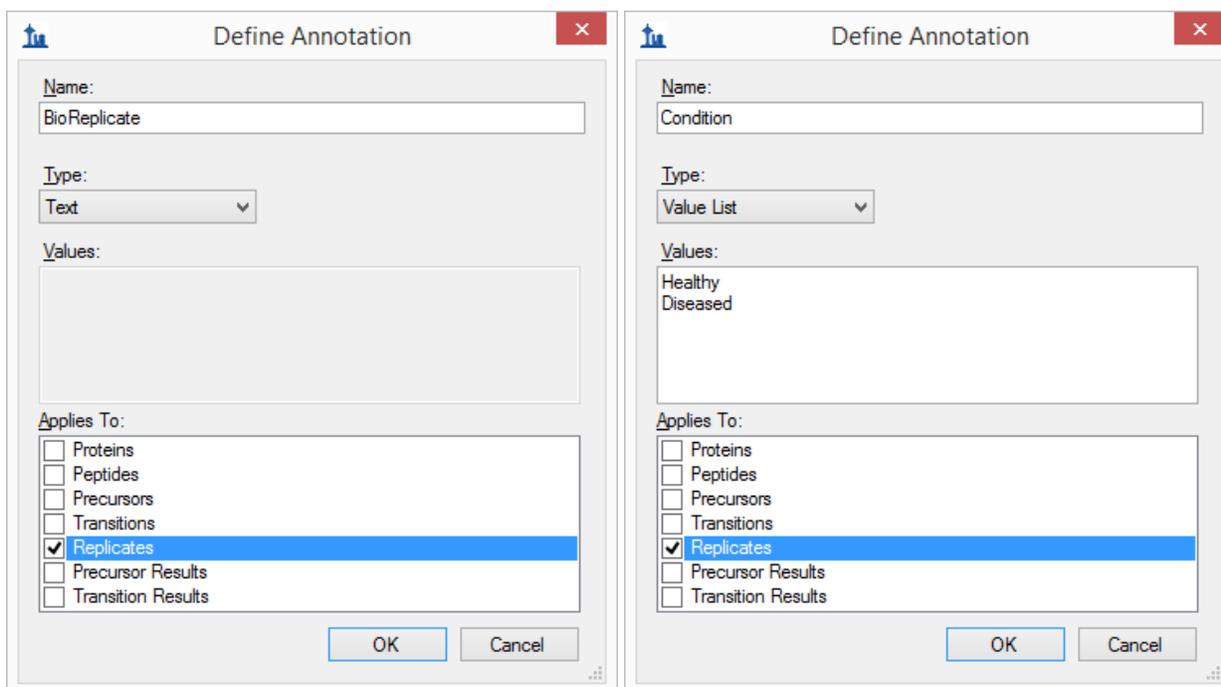
The screenshot shows a window titled "Install from Tool Store" with a close button in the top right corner. On the left is a list of tools: MPPReport, MS1Probe, MSstats (highlighted), Population Variations, Protter, QuaSAR, and SProCoP. The main area displays the MSstats logo, a "More Information" link to <http://www.msstats.org/>, and a "Status" field indicating it is not installed and version 3.1.2 is available. The "Description" field contains text about MSstats as an R package for proteomics. On the right, there are input fields for "Organization" (Vitek Lab, Purdue University), "Authors" (Meena Choi, Ching-Yun Chang, Dr. Timothy Clough, Dr. Olga Vit), and "Languages" (R(3.0.3), C#). At the bottom are buttons for "Tool Store...", "Install", and "Close".

- Click the **Install** button and follow the prompts until MSstats installation is completed.

Perform the following steps to see the annotations added during MSstats installation or to add them directly without MSstats installation:

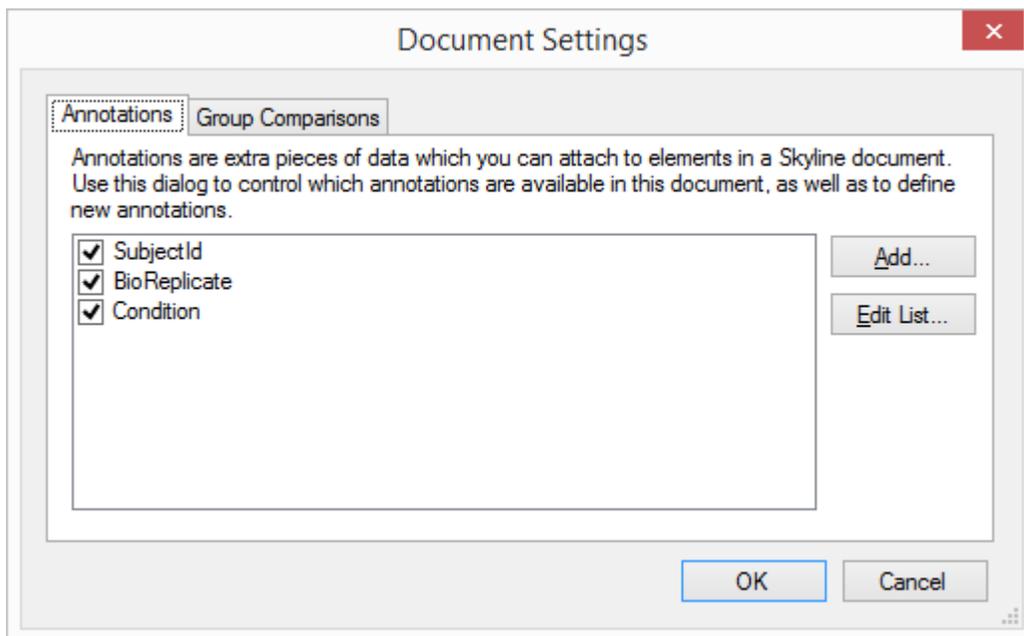
- On the **Settings** menu, click **Document Settings**.
- Click the **Edit List** button in the **Annotations** tab of the **Document Settings** form.
- If you installed MSstats, select the BioReplicate annotation, and click the **Edit** button. Otherwise, click the **Add** button.
- Repeat for the Condition annotations.

These annotation definitions should look like:



- Click **OK** to return to the **Document Settings** form.
- Check the BioReplicate and Condition annotations to add them to your document.

The **Document Settings** form should look like this:

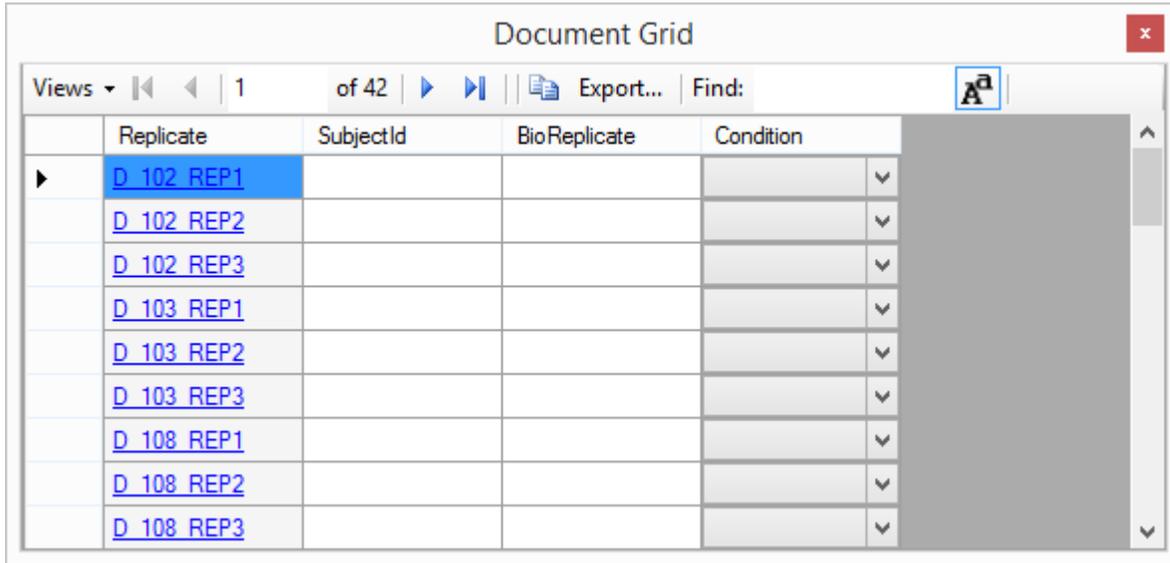


- Click the **OK** button.

To set the annotations you have added to the document, do the following:

- On the **View** menu, click **Document Grid**.
- On the **Views** menu, in the upper left corner of the **Document Grid**, click **Replicates**.

The **Document Grid** should look something like this:



	Replicate	SubjectId	BioReplicate	Condition
▶	D_102 REP1			▼
	D_102 REP2			▼
	D_102 REP3			▼
	D_103 REP1			▼
	D_103 REP2			▼
	D_103 REP3			▼
	D_108 REP1			▼
	D_108 REP2			▼
	D_108 REP3			▼

You can now manually enter the annotations for all 42 replicates into this grid. You can also paste values from a spreadsheet directly into this form. You can do this now by performing the following steps:

- In Excel, open the “Annotations.xlsx” file in the “GroupedStudies1\Heart Failure\raw” folder.
- Select all 42 rows of value cells beneath the header row for the 3 annotation columns.
- Press Ctrl-C to copy the cells.
- In Skyline select the top cell in the SubjectId column.

Note: Avoid entering cell edit mode on this step. If you see a blinking cursor inside the top SubjectId cell, press the Esc key. The **Document Grid** should look like this:

	Replicate	SubjectId	BioReplicate	Condition
▶	D_102 REP1			▼
	D_102 REP2			▼
	D_102 REP3			▼
	D_103 REP1			▼
	D_103 REP2			▼
	D_103 REP3			▼
	D_108 REP1			▼
	D_108 REP2			▼
	D_108 REP3			▼

- Press Ctrl-V to paste into the **Document Grid**.

Skyline should fill in the cells in the **Document Grid** with the values from the spreadsheet as follows:

	Replicate	SubjectId	BioReplicate	Condition
	H_160 REP2	H160	160	Healthy ▼
	H_160 REP3	H160	160	Healthy ▼
	H_161 REP1	H161	161	Healthy ▼
	H_161 REP2	H616	161	Healthy ▼
	H_161 REP3	H161	161	Healthy ▼
	H_162 REP1	H162	162	Healthy ▼
	H_162 REP2	H162	162	Healthy ▼
▶	H_162 REP3	H162	162	Healthy ▼

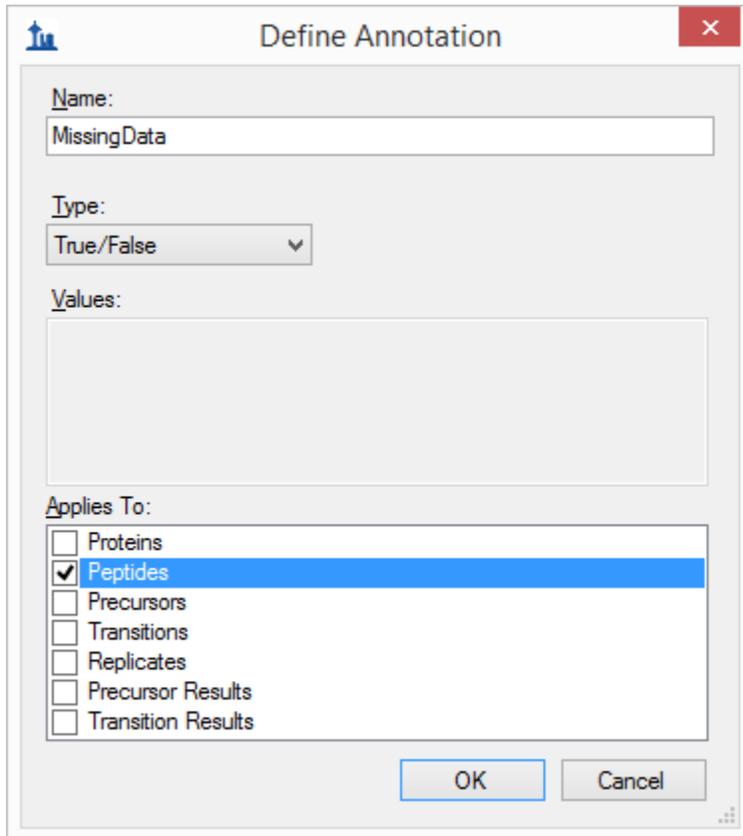
Annotating Peptides with Problem Peaks

You may also wish to annotate the peptides that either had peaks removed or truncated peaks, as both of these may cause issues during statistical analysis. To define the annotation you will use for this purpose, perform the following steps:

- On the **Settings** menu, click **Document Settings**.
- Click the **Edit List** button in the **Annotations** tab of the **Document Settings** form.
- Select the last annotation in the **Annotations** list.
- Click the **Add** button.

- In the **Name** field of the **Define Annotation** form, enter “MissingData”.
- From the **Type** dropdown list, choose “True/False”.
- In the **Applies To** list, check **Peptides**.

The **Define Annotation** form should look like:



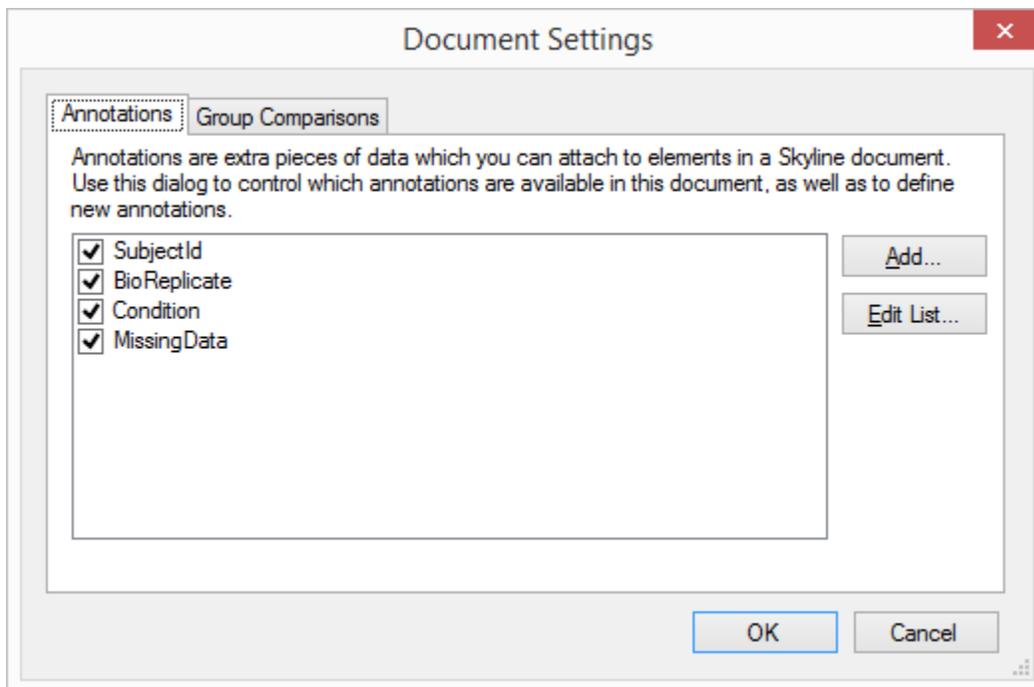
The screenshot shows a dialog box titled "Define Annotation". It has a standard Windows window title bar with a maximize icon, a close button (red X), and a small icon in the top left corner. The dialog is divided into several sections:

- Name:** A text input field containing the text "MissingData".
- Type:** A dropdown menu currently showing "True/False".
- Values:** A large, empty text area.
- Applies To:** A list of checkboxes with the following items:
 - Proteins
 - Peptides (highlighted in blue)
 - Precursors
 - Transitions
 - Replicates
 - Precursor Results
 - Transition Results

At the bottom right of the dialog, there are two buttons: "OK" and "Cancel". The "OK" button is highlighted with a blue border.

- Click the **OK** button.
- Click the **OK** button in the **Define Annotations** form.
- Check the “MissingData” annotation you just defined.

The **Document Settings** form should look like:

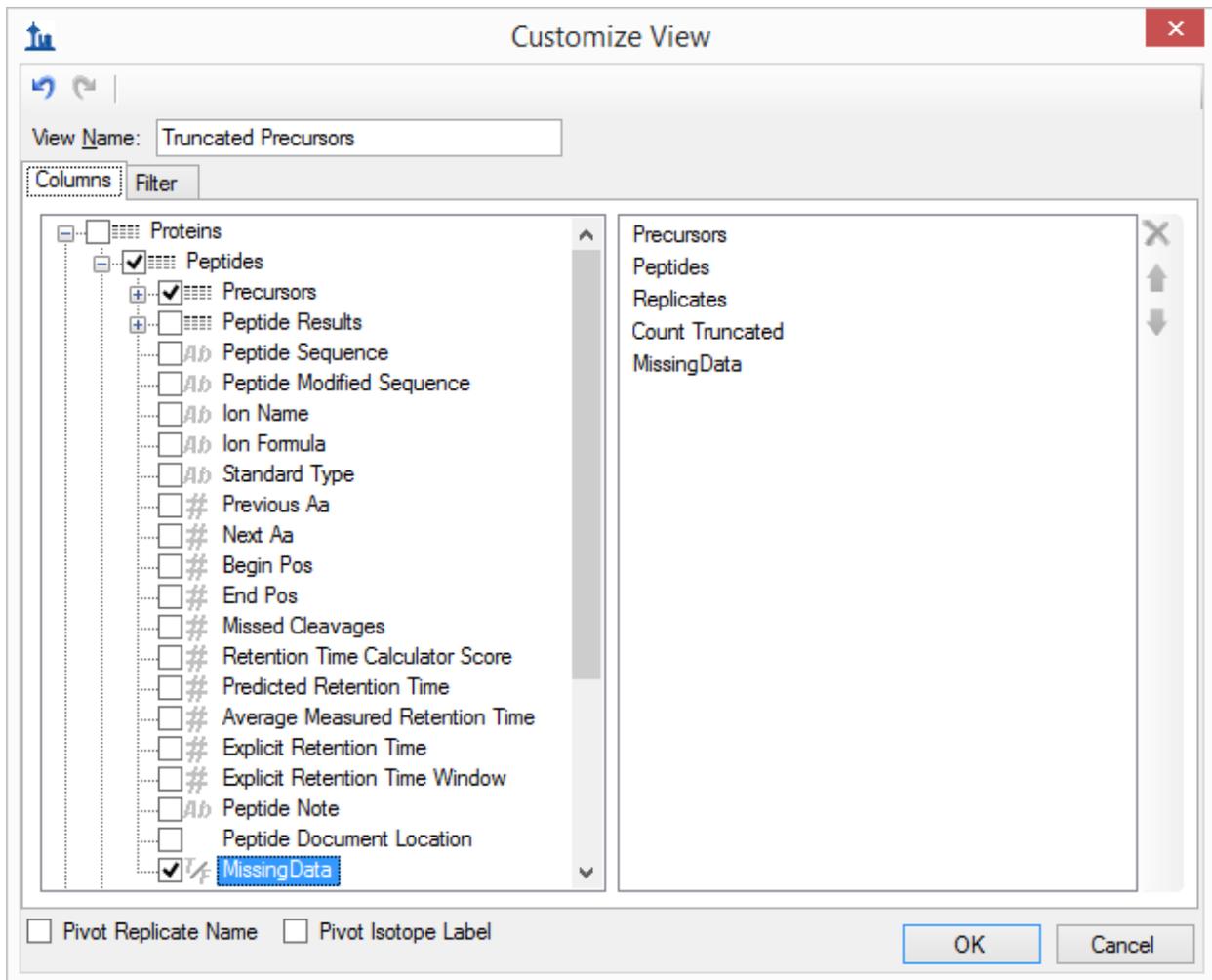


- Click the **OK** button.

To prepare for setting your new 'MissingData' annotation on all peptides with truncated peaks, do the following:

- In the **Document Grid**, on the **Views** menu, click the "Truncated Precursors" view you created before.
- On the same **Views** menu, click **Edit View**.
- On the **Customize View** form, expand the **Columns** tree to check Proteins > Peptides > MissingData, to add a column for the annotation you just created.

The **Customize View** form should look like this:



- Click the **OK** button.
- Click the “Count Truncated” column header twice to sort descending.

The **Document Grid** should look like this:

Document Grid

Views ▾ | 1 of 223 | Export... | Find:

	Precursor	Peptide	Replicate	Count Truncated	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"/>

Depending on how you processed the rest of your document, there may or may not be exactly 223 truncated precursor peak groups. You could check the “MissingData” check boxes one peptide at a time. Once you check a box for any peptide, the rest will check automatically, because the annotations apply only once to any peptide. Try this now by doing the following:

- Check the checkbox in the first row of the grid, immediately under the “MissingData” header.
- Press the Enter key.

The **Document Grid** should now look like this:

Document Grid

Views ▾ | 2 of 223 | Export... | Find:

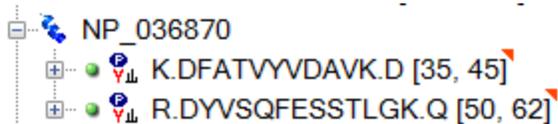
	Precursor	Peptide	Replicate	Count Truncated	MissingData
	614.3164++	DFATVYVDAVK	D_108_REP2	7	<input checked="" type="checkbox"/>
▶	614.3164++	DFATVYVDAVK	D_196_REP3	7	<input checked="" type="checkbox"/>
	614.3164++	DFATVYVDAVK	H_147_REP2	7	<input checked="" type="checkbox"/>
	614.3164++	DFATVYVDAVK	H_148_REP2	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"/>

You can also use Excel and the **Document Grid** support for pasting values to set all 163 rows in a single paste as follows:

- In a blank spreadsheet in Excel, scroll down to row 163.
- Enter “TRUE” in the first column of row 163.
- Select the cell you just entered.
- Click-and-drag the range extension square in the lower right corner of the selection rectangle to fill cells in the first column from row 163 to 1 with the value “TRUE”.
- Press Ctrl-C to copy these cells.
- Return to Skyline and select the first cell in the “MissingData” column.
- Press Ctrl-V to paste ‘TRUE’ values into all 163 rows from “MissingData”.

You may realize that this seems a little redundant. You only need to set the MissingData annotation to true once for every peptide. However, doing this achieves the desired result more quickly than by clicking a check box for each of the 31 peptides with truncated peaks.

Once you have performed the steps above, you should notice that many of the peptides at the top of the **Targets** list are now marked with a small red triangle above and right of the peptide name.

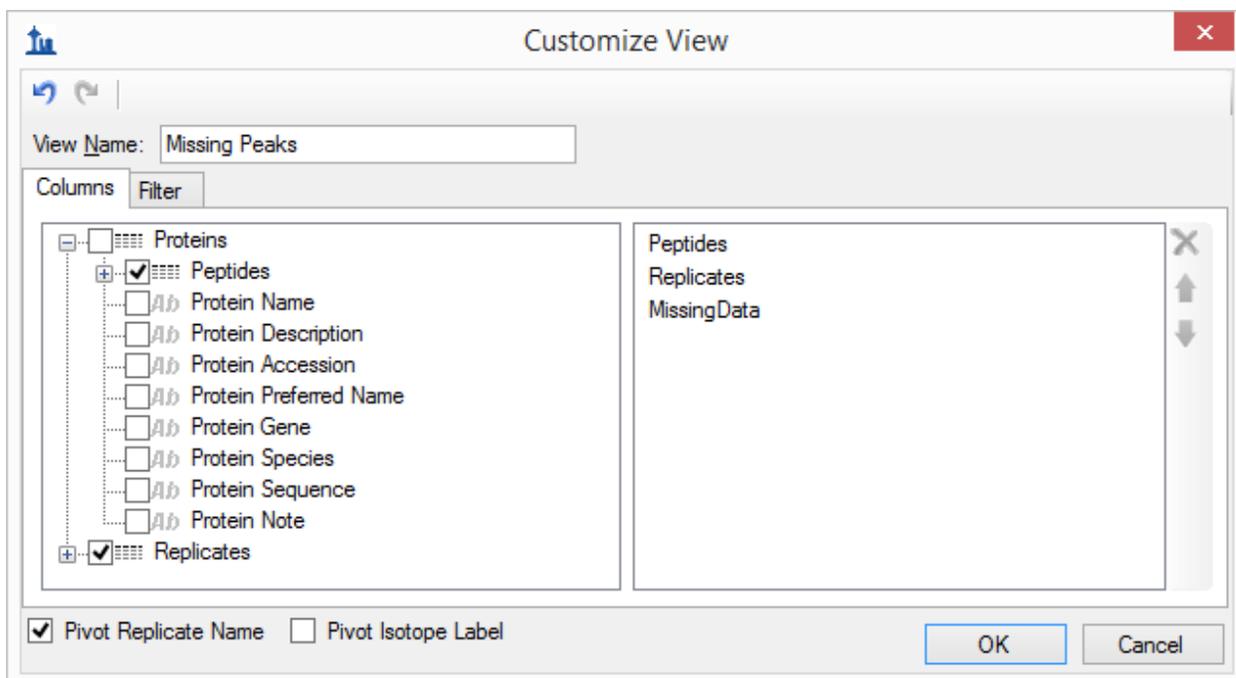


If you hover the mouse cursor over one of these triangles, you will see a tip appear with the text “**Missing Data True**”.

Now repeat this process for all of the peptides where you removed the peak entirely instead of integrating a truncated peak by doing the following:

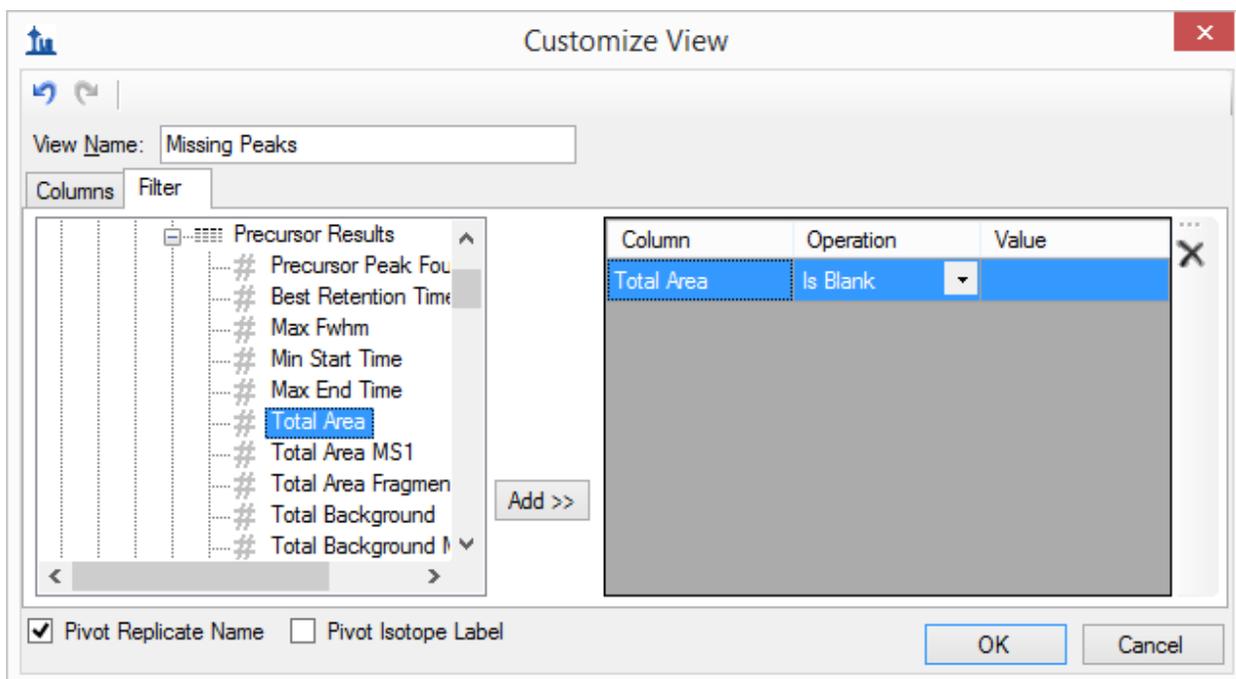
- On the **Views** menu of the **Document Grid**, click **Manage Views**.
- Select the “Truncated Precursors” view you have been using.
- Click the **Copy** button.
- In the **View Name** field of the **Customize View** form, enter “Missing Peaks”.
- Remove the **Precursors** and **Count Truncated** columns using the **X** button to the right of the columns list.
- Check the **Pivot Replicate Name** checkbox.

The **Customize View** form should look like this:



- Click the **Filter** tab.
- Click the **X** button to remove the existing filter on **Count Truncated**.
- Expand the columns tree to check Proteins > Peptides > Precursors > Precursor Results, to add a filter for the **Total Area** column.
- In the **Operation** column, choose “Is Blank”.

The **Customize View** form should look like this:



- Click the **OK** button.
- Click the **OK** button in the **Manage Views** form.
- On the **Views** menu in the **Document Grid**, click **Missing Peaks**.

The **Document Grid** should look like the following:

Peptide	MissingData	D_102_REP1 Replicate	D_102_REP3 Replicate	D_103_REP1 Replicate	D_103_REP3 Replicate	D_108_REP2 Replicate
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	
IFPENNIK	<input checked="" type="checkbox"/>				D_103_REP3	
ALIHCLHMS	<input checked="" type="checkbox"/>					
AGDQILAIN...	<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		

NOTE: If you did not fully process your document, you may instead see only one line for the peptide GSYNLQDLLAQ. The next note explains how to open a fully processed document.

In this view, you get a simple list of peptides with a missing peak for any run. You can easily see that there are 10 peptides. Only 2 of these do not also have truncated peaks and 8 already have the “MissingData” annotation set. The replicate names with missing peaks are pivoted to the right of the **Peptide** and **MissingData** columns. To finish labeling all peptides with missing data:

- Check the **MissingData** check boxes for TLNSINIAVFSK and IFSQQADLSR.
- Close the **Document Grid**.

Congratulations! You have completed initial data processing on this data set. After gaining a clearer understanding of issues that may impact data quality and optimizing peak integration to the extent possible for the acquired data, you are ready to begin running higher level statistics to further assess data quality. Hopefully, you will begin to get a sense for which peptides or proteins may actually prove useful as biomarkers for the conditions under study.

NOTE: If you have not fully processed your document to this point in the tutorial or you would like to compare your processing to that of the tutorial author, you can open the fully processed file included with this tutorial as follows:

- On the **File** menu, click **Save** (Ctrl-S) to save your document.
- On the **File** menu, click **Open** (Ctrl-O).

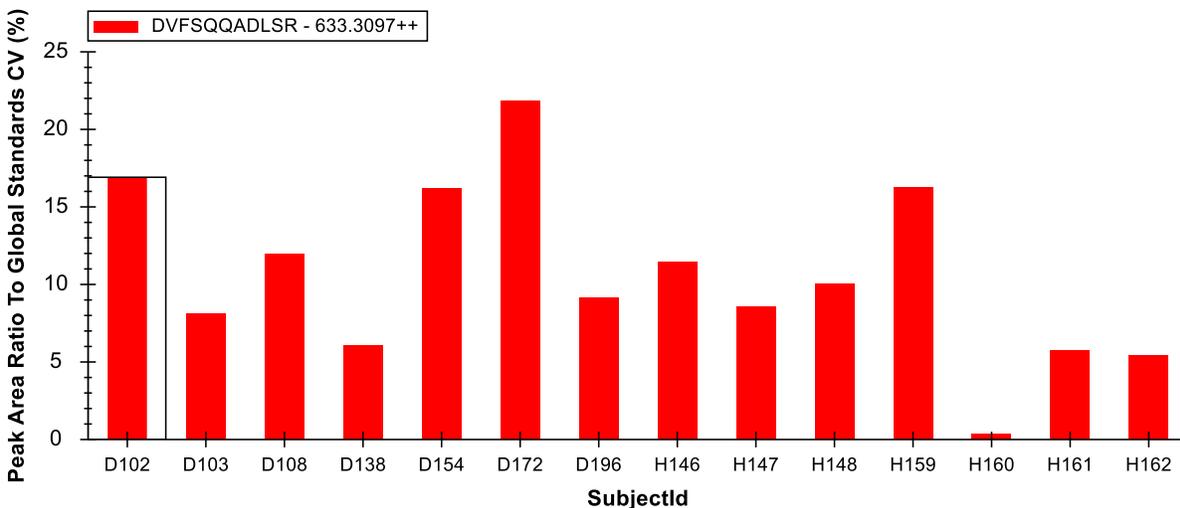
- Navigate up one folder to “GroupedStudies1\Heart Failure” and select “Rat_plasma.sky.zip”.
- Click the **Open** button.

Initial multi-replicate inspection in Skyline

While you will certainly want to do a more in-depth statistical analysis of any data set like this than Skyline currently offers, Skyline does provide some useful ways to perform an initial inspection of variance and group mean averages. To gain some insight into variance among the technical replicates for each subject, do the following:

- Select the peptide DVFSQQADLSR just above the global internal standard peptide list.
- On the **View** menu, choose **Transitions** and click **Total**.
- Right-click the **Peak Areas** view, choose **Normalize To** and click **Global Standards**.
- Right-click the **Peak Areas** view, choose **Group By** and click **SubjectId**.
- Right-click the **Peak Areas** view, and click **CV Values** if it is not already checked.

This should leave the **Peak Areas** graph looking something like:



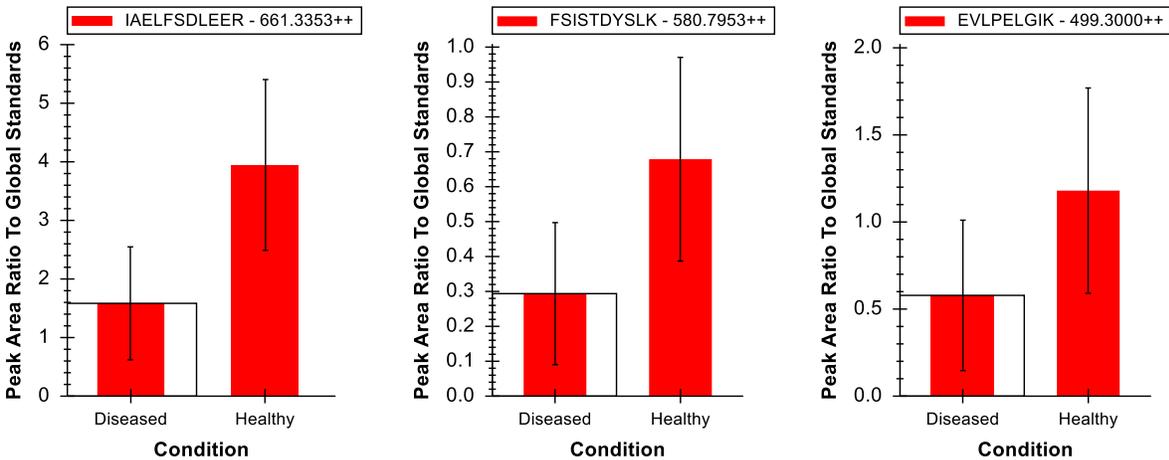
You can now select through several more peptides to review the coefficients of variation (CV) among the technical replicates for each subject. You will see that most are below 25%. Ideally, however, you would prefer to see lower CVs than in the ones observed in this experiment. It is worth noting that a lower CV derived from 3 measurements is more likely than the same CV derived from 10 measurements. This is due to a statistical tendency to underestimate standard deviations with small sample sizes.

Finally, you can use Skyline to get some initial insight into the difference in peptide expression between the two groups under study in this experiment. Perform the following steps to make a preliminary review of the differences in average peptide abundance between the healthy and diseased groups:

- Right-click the **Peak Areas** view, choose **Group By** and click **Condition**.
- Right-click the **Peak Areas** view, and click **CV Values** to uncheck this menu item.

- Select various peptides in the **Targets** view.

You will see graphs in the **Peak Areas** view like the following for the peptides of the protein NP_872280:



In these graphs, the bars represent the mean average value (in this case, peak area ratios to global standards) across all replicates and the whiskers show one standard deviation to either side of the mean. This gives you some idea of the sample distributions from which the means were derived.

In interpreting this graph, it is important to understand your goal in measuring differences across groups of samples. The two most common goals are:

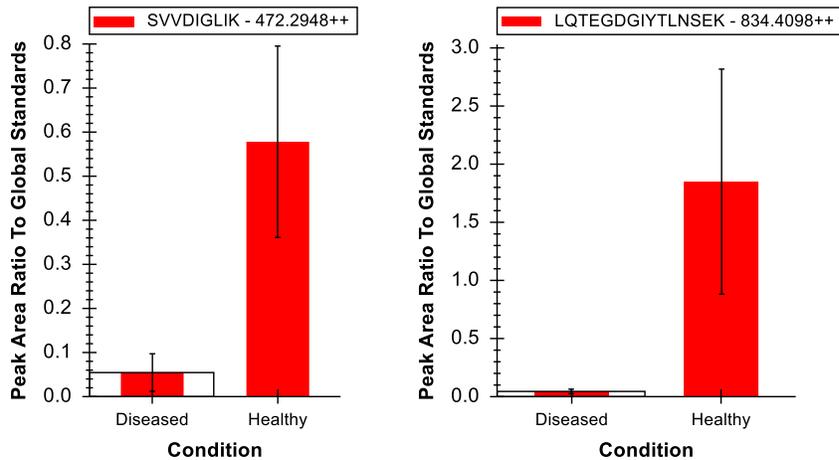
1. **Differential:** attempting to identify differences between groups. (e.g. Testing whether a treatment causes reduction in activity along some pathway of interest.)
2. **Prediction:** attempting to identify peptide expression that may be predictive of belonging to a particular group. (e.g. Testing whether a set of peptides might be used as biomarkers for inclusion in a certain group.)

If you want to simply detect a statistically significant difference between two groups, you would be more interested in the standard errors of the mean values than the standard deviations of the distributions. The p value from a T-test gives you a single number expressing the statistical significance of a difference in means related to the standard error values for those two means. The graphs above give you no insight into the standard error values. Therefore, they are not as useful for gaining insight on differential expression as they could be.

For prediction, the underlying distributions of the two populations are important. When you see a great deal of overlap between the two groups at just one standard deviation, the peptide alone is likely not a strong biomarker candidate. In that case, it would be hard to predict which distribution produced any single measurement. It is important to note here that it may still be possible to create a strong “biomarker panel” from a collection of peptides which are not individually predictive.

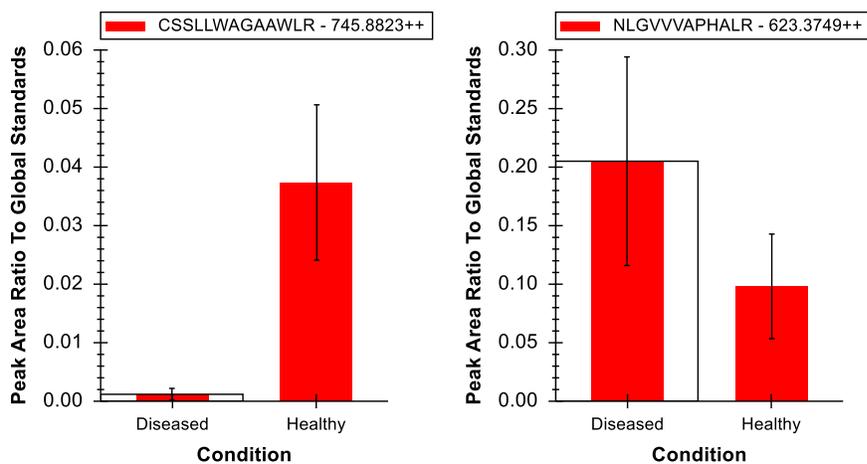
The peptides in the graphs above have statistically significant differences in means but are not individually predictive. This is illustrated by the overlap in distribution ranges at one standard deviation.

Continuing up the list of peptides / proteins, you will see that the protein immediately above, NP_036714, is a much stronger candidate for use as a biomarker:



Many of the peptides in this experiment show statistically significant differences in the means between the two groups. This should not be surprising given the source of our target list was the literature on differential protein expression in heart disease. Far fewer, however, appear to be strong biomarker candidates on their own merit.

When you reach the protein NP_001007697, which contains the peptide CSSLLWAGAAWLR, you will observe a case where the peptides assigned to a single protein experience very different relative expression levels between healthy and diseased groups:



This should act as a reminder that while you can achieve some confidence in measuring the same peptide molecule over many samples, you will often be far less confident in assigning multiple peptide precursors

to the same source protein form. The reasons for an observation of this nature are numerous. For instance, a peptide may have a post-translational modification (PTM) that makes it an important biomarker candidate. Whereas, the rest of the protein may not be affected.

Simple group comparisons in Skyline

As of Skyline 3.1, it is now possible to perform simple pairwise group comparisons of peptide or protein peak areas. The comparisons are performed by summing the available transition peak areas for a peptide or protein, optionally dividing by a normalization standard, taking the log, averaging any technical replicates and performing a T-test on the resulting values. Skyline automatically discards replicates with missing values, or truncated peaks in label-free data.

To try this now on the data set you have been processing, perform the following steps:

- On the **Settings** menu, click **Document Settings**.
- Click the **Add** button in the **Group Comparisons** tab of the **Document Settings** form.
- In the **Name** field of the **Edit Group Comparison** form, enter “Healthy v. Diseased”.
- For the **Control group annotations** choose “Condition”.
- For the **Control group value** choose “Healthy”.
- For the **Value to compare against** choose “Diseased”.
- For the **Identity annotation** choose “SubjectId”.
- For the **Normalization method** choose “Ratio to Global Standards”.
- In the **Confidence level** field, enter “99” %.
- For **Scope** choose **Protein**.

The **Edit Group Comparison Form** should look like this:

Edit Group Comparison

Name: Healthy v. Diseased

Control group annotation: Condition

Control group value: Healthy

Value to compare against: Diseased

Identity annotation (for technical replicates): SubjectId

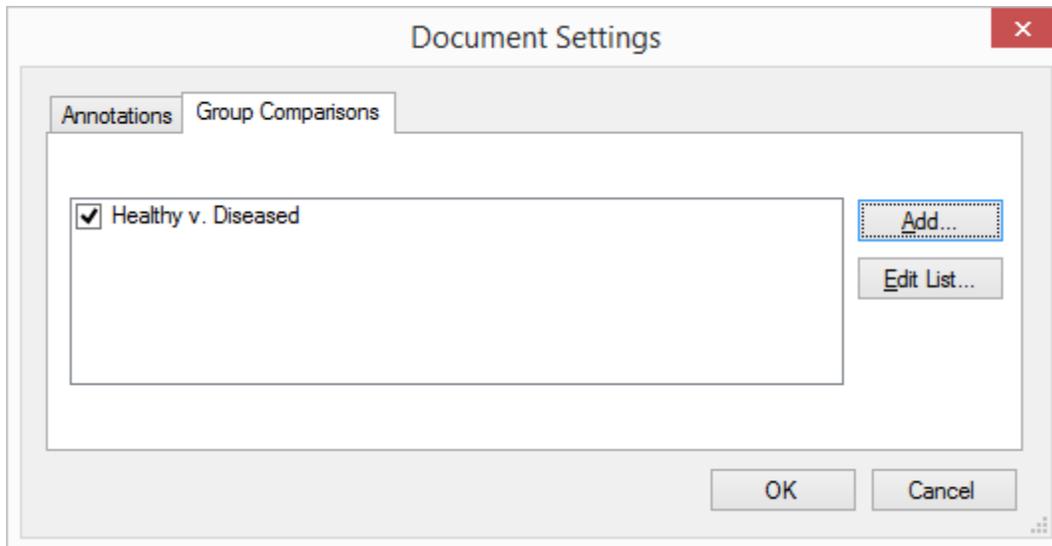
Normalization method: Ratio to Global Standards

Confidence level: 99 %

Scope
 Peptide Protein

- Click the **OK** button.

The **Document Settings** form should look like this:



- Click the **OK** button.

To inspect the group comparison you just defined do the following:

- On the **View** menu, choose **Group Comparisons** and click **Health v. Diseased**.

Skyline will show a grid view that looks like this:

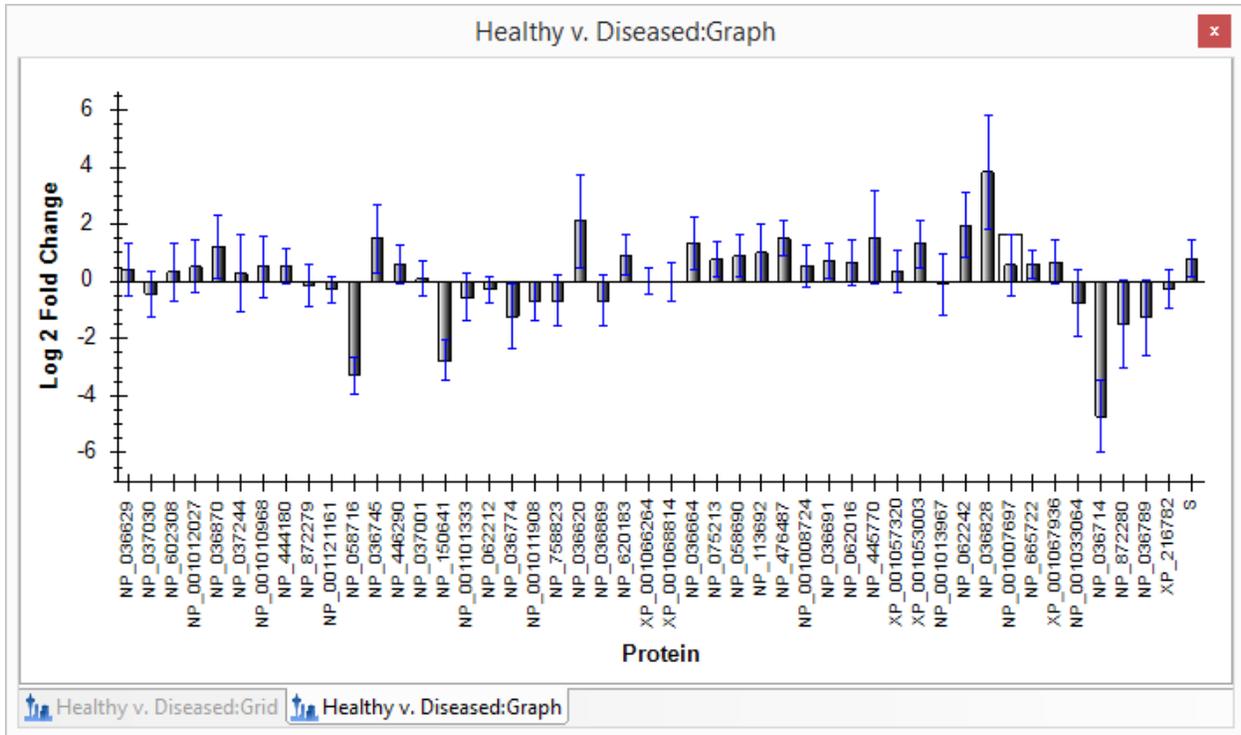
	Protein	Fold Change Result	Adjusted P-Value
▶	NP_036629	1.35 (99% CI:0.72 to 2.55)	0.2057
	NP_037030	0.74 (99% CI:0.42 to 1.3)	0.1685
	NP_602308	1.26 (99% CI:0.63 to 2.55)	0.3696
	NP_001012027	1.44 (99% CI:0.76 to 2.72)	0.1499
	NP_036870	2.31 (99% CI:1.08 to 4.95)	0.0163
	NP_037244	1.21 (99% CI:0.48 to 3.09)	0.5251
	NP_001010968	1.42 (99% CI:0.67 to 3.02)	0.2057

If the “Fold Change Result” column does not show the full confidence interval, double-click the vertical line between the Fold Change Result and Adjusted P-Value headers.

To see the logged fold change values and confidence intervals graphed:

- Click the **Show Graph** button in the menu above the toolbar in the grid view.

Skyline adds a graph tab to the **Healthy v. Diseased** view that looks like this:



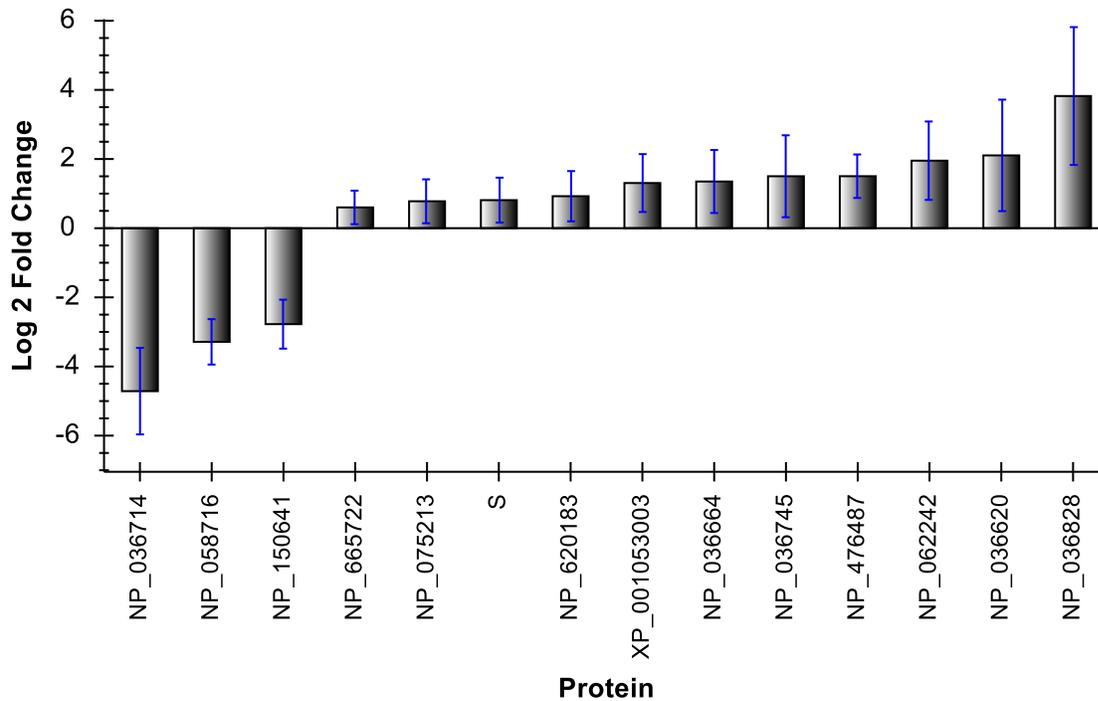
- Click and drag the “Healthy v. Diseased:Graph” tab at the bottom of the view until you see a blue floating rectangle.
- Release the mouse button and adjust the grid and graph views for side-by-side viewing.
- On the grid, click the “Fold Change Result” header, and click **Sort Ascending** in the menu that appears.

This will sort the grid, and if you switch back to the graph, you will see that it is also sorted. Notice that many of the confidence interval whiskers cross the zero line, indicating that at 99% confidence it would not be unusual for the observed data to occur by random chance. And, this is not even with any correction for multiple hypothesis testing.

To set a cut-off based on the Benjamini-Hochberg adjusted p-value, which estimates false discovery rate (FDR), do the following in the grid view:

- Click the “Adjusted P-Value” header and click **Filter**.
- In the **Filter Type** field, choose “Is Less Than”.
- In the field below this, enter “0.01”.
- Click the **OK** button.

You will see the number of rows indicated in the grid toolbar drop from 48 to 14, and the graph should now look like this:



Note that the list of significantly changing proteins includes the “S” protein that you inspected earlier, and which contains only the two problem degrading peptides and the normalization peptide. This gives some indication of the problems that failing to randomize sample order can cause. Because in all three cycles of measurement, the diseased subjects preceded the healthy subjects, you should expect that any degradation not accounted for in your normalization will cause fold change to appear as up-regulation in the diseased group. And so, along with the “S” protein, all proteins with just barely significant up-regulation in the diseased group should be viewed with some skepticism in this data set.

Correctly specifying technical replicates

If you collect technical replicates, as in this experiment, it is extremely important to your statistical inference that you specify them correctly in the **Edit Group Comparison** form. If you do not, then each measurement will be considered as coming from a distinct biological subject, incorrectly narrowing your standard errors and, hence, confidence intervals, and artificially lowering p values.

To see this in action, try the following:

- Click the **Change Settings** text in the menu above the toolbar in the grid view.
- For the **Identity annotation** choose the empty element at the top.

You will see the grid view show 37 rows (and the graph 37 bars) now with adjusted p values less than 0.01. This is because the statistics are now calculated as if you had 42 distinct subjects, instead of 14 with 3 measurements each. This is an important distinction, and as you can see making a mistake here could yield embarrassingly over-optimistic statistical inference. In this case, the 14 proteins showing

statistically significant differences in means between healthy and diseased groups, at an estimated 1% FDR, has ballooned to 37 proteins.

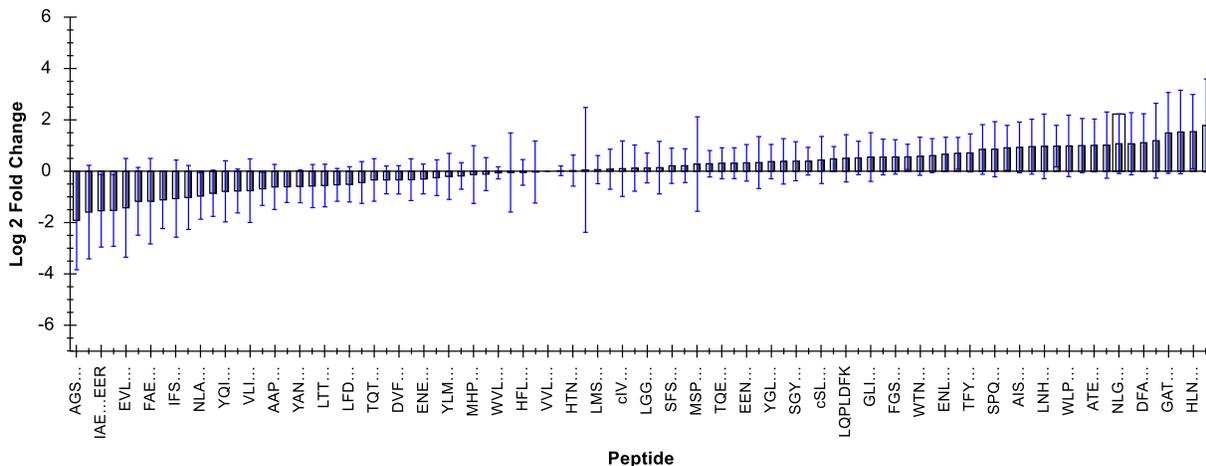
Refinement by differentiation

Returning to the original goal of this experiment, you can now reduce the document to only a subset of peptides with statistically significant differences in group means, at an estimated 1% FDR. Though, keep in mind that this data set may overstate the significance of up-regulation in the diseased group, due to the lack of randomization (diseased always measured before healthy) and systematic signal degradation over the runs.

Nonetheless, to reduce the targets list in this document to the set of peptides with differences in group means at an estimated 1% FDR, perform the following steps:

- On the **Healthy v. Diseased:Settings** form, change the **Identity annotation** field back to “SubjectId”.
- On the **File** menu, click **Save As**.
- Save the document to the name “Rat_plasma-diff.sky”.
- In the **Healthy v. Diseased:Settings** form, choose the **Peptide** option in the **Scope** box.
- Click the “Adjusted P-Value” header and click **Filter**.
- In the **Filter Type** field, choose “Is Greater Than or Equal To”.
- Click the **OK** button.

The **Healthy v. Diseased:Grid** should show 92 rows above the 1% FDR cut-off, and the graph should now look like this:



To delete these peptides from the document do the following:

- Click the gray rectangle in the upper left corner of the **Healthy v. Diseased:Grid**, where the row and column headers meet, to select all cells.
- Scroll down the grid until you find the VVLSGSDATLAYSFAK peptide in the “S” protein.

- Hold down the Ctrl key and click on the row header for this row to deselect it.

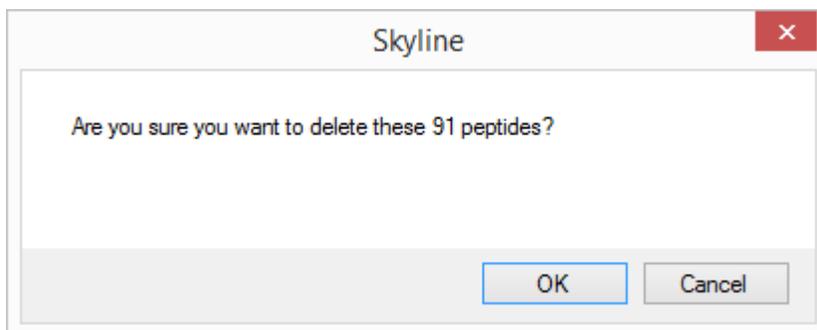
The grid should look like this:

	Protein	Peptide	Fold Change Result	Adjusted P-Value
	NP_872279	WVLTVAHCFEGR	0.96 (99% CI:0.8...	0.4948
	NP_058690	AFMDCCNYITK	0.97 (99% CI:0.3...	0.9451
	XP_001066264	HFLIETGPK	0.97 (99% CI:0.6...	0.8178
	NP_001013967	VTSAAFPSPIEK	0.98 (99% CI:0.4...	0.9474
	S	VVLSGSDATLA...	1 (99% CI:1 to 1)	0.9769
	NP_036745	ENSSNILDNLLSR	1.01 (99% CI:0.8...	0.8135
	NP_001121161	HTNNGMICTS...	1.02 (99% CI:0.6...	0.936
	NP_037244	TGTNLMDFLSR	1.04 (99% CI:0.1...	0.9474
	NP_037244	LMSPEEKPAPA...	1.04 (99% CI:0.7...	0.7959
	XP_001068814	SLVIQKPSEEN...	1.05 (99% CI:0.6...	0.8135
	NP_872279	CIVDGDNR	1.07 (99% CI:0.5...	0.8178

Note that the VVLSGSDATLAYSFAK peptide has a fold-change of 1, with a 99% confidence interval of 1 to 1. This is the single peptide used as the Global Standard for normalization. You do not want to delete this peptide from the document. That is why you deselected it.

- Click the  button in the grid toolbar.

Skyline will show the following message to confirm you really want to delete the selected rows:



- Click the **OK** button.

You can now close the group comparison windows and review the 34 peptides remaining in the document. They should all have pronounced differences in means between the two groups, though their distributions

will not always be disjoint even at only one standard deviation. That is, you will see the standard deviation whiskers overlapping in the **Peak Areas** graph that should still be showing the peak areas grouped by condition.

Conclusion

In this tutorial, you learned some of the most effective techniques for visual assessment and correction of the peak integration that Skyline automatically performs on all imported LC-MS/MS chromatography data. Although you only worked with an unlabeled SRM data set, the techniques you applied in this tutorial would apply equally well to data from other chromatography-based quantification methods. The manual nature of this inspection and correction makes it best suited for data sets containing not much more than two thousand peptides. In this case, 137 peptides took about 1 hour to process, which means a data set with 2,000 peptides might take upwards of 2 days of effort.

For much larger data sets you might want to first perform some very rough screening of peptides based on their initial promise in showing some difference between the conditions you are studying. In that case, you only spend time on this type of manual inspection and correction for potential peptides of interest. You should now be able to achieve this kind of filtering with the approach you learned using Skyline group comparison support. You may want to relax the limit on adjusted p value (or FDR cut-off) to something less stringent than the 1% used in this tutorial, but the approach would be similar.

You have now seen how this type of processing and expert use of the features Skyline provides can help you both understand and correct potential sources of error in your quantitative experiments. You have also dealt first-hand with mis-assigned peaks, interference, peak truncation, peptides at abundance levels that are difficult to detect, 200-fold peptide signal degradation, and doubly eluting peptides. With a mastery of the tools Skyline provides for understanding, correcting and annotating these issues, you should be able to reduce error in your quantitative targeted proteomics experiments and get to biological insights more effectively.

Bibliography

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).