Quantitation in Scaffold Q+/Q+S

Proteome Software Users Group 2013 Tutorial Session

Copyright 2013 Proteome Software
# Quantitation in Scaffold Q+/Q+S

*User’s Tutorial - Proteome Software Users Group 2013*

## Contents

- **What data can Scaffold Q+ analyze?** ................................................................. 2
- **What data can Scaffold Q+S analyze?** .......................................................... 2

**Task 1: Load Quantitative data into Scaffold Q+/Q+S** ........................................... 2

**Task 2: Setting up Q+S for SILAC Analysis** ....................................................... 4

**Task 3: Setting up the Q+ Multiplex Quantitation Window for iTRAQ** .................. 7

**Task 4: Detecting Changes over entire experiment** ............................................. 9

  - **Peptide Pane** ................................................................................................. 10
  - **Spectra Pane** ............................................................................................... 11
  - **Peptide Level Charts** .................................................................................. 14
  - **Protein Sequence** ........................................................................................ 15
  - **Spectrum** .................................................................................................... 15

**Task 6: Statistics** ................................................................................................ 16

  - **Sample Wide fold Changes** ........................................................................ 16
  - **Sample Wide Scatterplot** ........................................................................... 16
  - **Intensity Weighting Graph** ......................................................................... 16
  - **Pre/Post Normalization Graph** ..................................................................... 17
  - **Raw Quantifiable Intensity** .......................................................................... 17
  - **Signal to Noise** .......................................................................................... 18

**Task 7: Quantitative Settings** .............................................................................. 19

  - **Advanced Settings** ..................................................................................... 19
  - **Quantitative Settings** ................................................................................ 19
  - **Calculation Type** ....................................................................................... 19
    - **Median or Mean** .................................................................................... 19
Reference Type ................................................................................................................................................... 21
Normalization Type ............................................................................................................................................ 22
Minimum Dynamic Range ................................................................................................................................ 22
Other Settings ..................................................................................................................................................... 22

Task 8: Quantitative Testing .............................................................................................................................. 23
Average-Based Tests ........................................................................................................................................ 23
Median-Based Tests ........................................................................................................................................ 28
What data can Scaffold Q+ analyze?
Scaffold Q+ is Proteome Software’s labeling quantitation software package. It loads iTRAQ (Applied Biosystems) and Tandem Mass Tagged (TMT, Thermo Scientific) labeled data. Scaffold Q+ can load SEQUEST, Mascot, X!Tandem, Phenyx, OMSSA, Spectrum Mill, Waters Identity and Proteome Discoverer data.

What data can Scaffold Q+S analyze?
Scaffold Q+S can handle SILAC data from MaxQuant, Mascot Distiller, Spectrum Mill and Proteome Discoverer.

Task 1: Load Quantitative data into Scaffold Q+/Q+S
Loading a Sample Dataset

This task will illustrate the process of loading SILAC data into Scaffold Q+S, but loading iTRAQ or TMT data is similar. The key is to select the proper Quantitative Technique for the data you are loading.

You will need to download a sample set of SILAC data from our website at http://www.proteomesoftware.com/products/demo-data#qplus. Download “Mascot Distiller Data” and “FASTA File for SILAC Data”.

When you first open Q+S, the welcome screen will present itself.

Select "New" and Scaffold Q+S will launch the load data wizard. You can skip any Welcome pages that appear and move on to the following window.

Select the SILAC (Multiplex) option and click “Next.”

Load the file Tutorial_Distiller_SILAC.xml

Proceed with loading as you would in Scaffold. For this exercise, do not select MuDPIT but you can condense data during loading to speed up the process.
Add the SILAC-demo.fasta database from the Q+S folder in your demo data and use Auto Parse.

Set the other parameters as shown.

Click “Load and Analyze Data” to begin analysis. Once the SILAC data has finished loading into Scaffold Q+S, the user is presented with the traditional Scaffold Samples view. You may wish to save this file for use in subsequent tutorials.
Task 2: Setting up Q+S for SILAC Analysis

This task uses the file created in Task 1.

To begin working with your SILAC quantitative data, click the Q+ icon at the top of the screen. Scaffold will launch the Sample Organization window. Here you organize SILAC Quant samples into categories for comparison. To organize your Quant samples first select Quant 1 as the Reference. To add Quant 1 as the reference, select it and click “Add.”

You may find it useful to edit your Sample Names and give them more meaningful names such as Light, Medium, and Heavy. Scaffold does not explicitly identify the Quant samples, but they are ordered from lightest to heaviest.
Add Quant2 to Category 1 and then add a new category using the “+” icon.

You may rename the category if you wish by clicking Rename Category.” Enter the new name and click “Apply.”
Once the category is named it will appear in the Sample Organization window.

Add a Quant sample to the new category by highlighting it and clicking “Add” and edit it as desired.

You will then need to add a new category for the third sample. Click the green plus sign to create a new category, give it a name and click “Apply.”

Add the last quant sample to the new category and rename it. When the organization is complete, click “Apply” at the bottom of the screen. The Multiplex Quantitation window opens to the Scaffold Q+ Samples view.

Quantitation with SILAC data is identical to quantitation with iTRAQ/TMT data.
Task 3: Setting up the Q+ Multiplex Quantitation Window for iTRAQ

Open the iTRAQ.sf3 file available under Help>Open Demo Files.

To access Q+ features select the Q+ icon from the taskbar.

The first time the Q+ icon is selected after data is loaded, Scaffold will launch the Sample Organization window. Here you organize iTRAQ Quant samples into categories for comparison.

In the tutorial file, samples have already been organized, but you can open the Organize Samples dialog through the Quant menu in Q+ and reorganize them at any time. To restore Q+ to its original state, select both Quant samples under Reference and click Remove, then go to the Category 1 tab and select both Quant samples under Category 1 and click Remove. Click on the Quant ID column to reorder the samples. The screen should now appear as it did upon loading so you can work through the organization process yourself.

To organize your Quant samples first select Quant 1 (iTRAQ-114 label) as the Reference. To add Quant 1 as the reference select it and click “Add.”
Add Quant2 to Category 1 and then add a new category using the "+" icon.

Name the new category and click “Apply.”

Once the new category is made it will appear in the Sample Organization window. Add Quant 3. Repeat the process for each new category. In this case, because this is an iTRAQ 4 plex experiment there will be three categories and one reference. After the samples are organized click “Apply” and Scaffold Q+ will launch the Multiplex Quantitation window.
Task 4: Detecting Changes over entire experiment

Quantitation in ITRAQ, TMT or SILAC is similar. For this task, we will use Scaffold’s iTRAQ demo file for purposes of illustration.

Go into Scaffold, and select Help>Open Demo Files, and open iTRAQ.sf3. Click the Q+ icon.

Open Quant>Organize Samples, and rearrange the Quant Samples so that Quant 1 is in the Reference Category, Quant 2 is in Category 1, Quant 3 is in Category 2, and Quant 4 is in Category 3.

In the Q+ Samples view proteins are listed with their log2 intensity ratios displayed. log2 ratios are color coded. Ratios indicating decreased protein levels are colored in red and increases are colored in green. In this sample iTRAQ label 114 was the reference and all other labels are compared to 114. For protein Apolipoporins, there is decreased protein detection in Quant sample 2 and Quant sample 4.

It is possible to change the display from log2 Fold Change to “Protein Identification Probability,” “Number of Assigned Spectra,” “Fold Change Ratio,” Normalized Intensity Value” which is the log2 of the Intensity, or “Coefficient of Variance.”

Change the display option to “Number of Assigned Spectra.” Scaffold Q+ still shows fold change color coding based on the ratios even when the display is set to “Number of Assigned Spectra.”
Task 5: Analyzing changes for specific proteins

The Proteins view of Scaffold Q+ contains information regarding overall protein label detection and the protein’s peptide label detection. Select protein Apolipoporphins and double click on the protein or click the Protein tab.

The default protein view shows information about the selected protein in three vertical panes. All of the peptides present in the selected protein appear in the top Peptides pane. Spectra for the selected peptide are seen in the middle Spectra pane. The lowest pane has multiple tabs providing a variety of charts illustrating the quantitative calculations as well as the protein sequence and the spectrum.

**Peptide Pane**

Protein peptides are in the top pane. Their respective quantitative values are displayed for each Quant sample. The value displayed depends on the display option selected in the Q+ Samples View. The log₂ fold change is displayed if you select “Protein Identification Probability,” “Number of Assigned Spectra,” or “Coefficient of Variance.” Here we see peptide AISEILEK has log₂ fold change scores of 0.28, -1.1, 0.14 and -1.2. Color coding from the samples view is retained in the Peptides Pane of the Proteins view.
The Peptide Pane also contains information regarding the number of spectra assigned to a peptide. This is indicated under the charge state +1H, +2H, +3H, and +4H. If the peptide could have been assigned to any other protein, the accession number of the other protein will be listed in the Other Proteins column. Categories are color-coded. These colors are retained in the charts and can be changed through “Quant>Organize Samples.”

Spectra Pane

The Spectra pane is located in the middle of the Peptide and Protein Sequence panes. The Spectrum pane in Scaffold Q+ contains information regarding the spectra assigned to a particular peptide sequence. Unlike standard Scaffold, Scaffold Q+ does not display redundant peptides in the Peptides Pane. In this example, select peptide AISEILEK (second from top).
Note that two spectra are assigned to peptide AISEILEK. In cases where multiple spectra are assigned to the same peptide, by default Scaffold Q+ calculates a weighted median value for the peptide. Details of this calculation and instructions for changing the quantitative settings can be found later in this document.

The Spectrum pane can be used to investigate multiple spectra assigned to one peptide. This is particularly useful for verification purposes. If multiple spectra are assigned to the same peptide and their profiles are consistent across Quant samples then confidence is gained in the detected change. However the opposite is also true, and if the profile for multiple spectra assigned to the same peptide differ then confidence in the detected change is lowered.

**Category Level Chart**

The Category Level Chart provides a visual representation of the distributions of the normalized intensity values for the various categories. The outer line graph shows the distribution of the \( \log_2 \) normalized intensities for all spectra corresponding to the selected protein. The distribution is calculated using a kernel density estimation to account for variation in the intensity measurements.

The solid graph represents the distribution of sample means derived by randomly sampling these range distributions. The degree of separation of these median density distributions is an indication of the statistical significance of the difference in protein levels and also helps to clarify the patterns of differences that might be detected by ANOVA. When the mean distributions of two categories are clearly separated, as these are, it indicates a greater likelihood that there is a meaningful difference in the level of this protein in these two categories.

**Protein Level Charts**

Protein Level Charts display the overall peptide label information for each protein in your sample. There are three different displays: Bar Chart, Box Plot, and Trend Line.
Box Plot

The legend indicates which category each Quant sample is assigned.

Bar Chart

Trend Line

It is important to note that the Trend Line display is interactive. Mousing over the circular point for peptides in each Quant sample highlights that peptide in the above Peptide pane.
Selected Peptide Highlighted

Peptide Level Charts
The Peptide Level Charts tab displays information for a specific peptide in a bar chart, box plot, or trend line. This information is for only one peptide as opposed to the many peptides that are displayed in Protein Level Charts. For this example select peptide AISEILEK.

Peptide Trend Line
Protein Sequence

In Q+, a dropdown gives the option to display the full protein coverage map colored yellow for identified peptide sequences, with modifications (including label) colored in green as in Scaffold.

Any individual quant sample’s peptides may also be displayed. Peptides are outlined in yellow and increased or decreased protein level is indicated in green or red.

Spectrum

The Spectrum tab contains a view of the spectral peaks.

This differs from standard Scaffold’s spectrum view in that it also displays a dashed cutoff line. The default cutoff is 1% of the highest peak in this spectrum. The level can be adjusted through Quantitative Settings. For this example select peptide CTAIEFDNNQF (third from the top) and click on the Spectrum tab.

In addition to the spectral peaks, this view also shows a bar chart of the log2 intensity values of the reporter ions in the spectrum. By checking and unchecking the box under this chart, you can display the normalized or un-normalized values.
Task 6: Statistics

The top half of the Scaffold Q+ Statistics View contains two displays of the data. Use the Primary and Secondary sample dropdown menus to change which Quant samples are compared in the two graphs. The protein selected in the Protein dropdown is displayed in yellow in these graphs.

Sample Wide Fold Changes

The Sample Wide Fold Changes are useful in displaying the overall fold change of the protein selected. For this task select protein Apolipoporphins from the drop down list above the Sample Wide Fold change view.

Note that the selected protein is highlighted in yellow at the far left of the plot. Apolipoporphins has the largest ratio drop between these two Quant samples. It is possible to compare any two Quant samples using the drop down lists above the plot. You can also use the drop down list to select any protein in the experiment.

Sample Wide Scatterplot

The Sample Wide Scatterplot compares the values for all proteins detected between two Quant samples. The axis values represent the ion counts of reporter tags for the spectra assigned to the protein in question.

There appears to be a deviation from the expected favoring Quant Sample 1 at high ion counts. This plot is useful in identifying anomalies and/or apparent biases in datasets. If all of the ion counts for a particular label are much higher or lower than another then perhaps there is an issue with the labeling reaction, or the order in which samples are being tagged.

The lower portion of the Statistics View displays graphs that provide insight into the normalization calculations performed by Scaffold Q+. These graphs represent the dataset as a whole, rather than an individual protein.

Intensity Weighting Graph

On the left is the Intensity Weighting graph.

Scaffold Q+ uses a weighting function in its normalization algorithm
to accommodate the fact that variation in the measurement of intensities is not consistent across the intensity distribution.

For example, in many cases higher intensity measurements tend to be more accurate than lower ones. The purple line graphs the standard deviation of the logs of the raw intensities at each intensity level. The green line shows the weighting function that Scaffold Q+ has calculated for each intensity level based on these standard deviations.

**Raw Intensity Distribution**

In the right pane, the Raw Intensity Distribution has three tabs. These views of the data can help inform your decisions about how you should organize and normalize your data. They can show if there are any problematic samples which may need to be discarded from the analysis, for example if a sample has so much missing data that it is unreliable for quantitation. They can also show whether your data is well-suited to normalization, and which quantitative settings will give you the best results.

**Pre/Post Normalization Graph**

The Pre/Post Normalization graph shows how much each sample is adjusted by normalization. Here you can see that Quant 3 is adjusted more than the other samples.

The percentage figures in the Pre-Normalization figure indicate the percentage of spectra which are missing quantitative data. If there is too much missing data, quantitation may be unreliable. The sample dataset you are using for this task is unusual in that it missing very little data.

**Raw Quantifiable Intensity**

The second tab in the Raw Intensity Distribution pane displays the Intensity Distribution histogram for a specific quantitative sample. The sample is chosen by changing the selection in the Primary Sample dropdown at the top of the screen.
This histogram shows the distribution of intensities in a sample. Ideally, it should resemble a normal distribution. You can toggle whether or not the intensities are displayed in log form. Scaffold Q+ converts all intensities to logs before performing its calculations in order to achieve a more symmetric distribution.

Signal to Noise

The final tab in the Raw Intensity Distribution shows the Signal to Noise histogram. This figure shows the relative intensity of the reporter ions compared to the largest peak in the spectrum.

For iTRAQ data the largest peak is often one of the peaks identifying the peptide rather than one of the reporter ions. If most of the reporter ions are very small, it may indicate that the collision energy was tuned to fragment the peptide backbone rather than the reporter ions. Low reporter ions may be less accurate if the instrument has a limited dynamic range.
Task 7: Quantitative Settings

To access Scaffold the Q+ Quantitative options go to “Edit” > “Quant” in the menu.

There are four options under the “Quant” menu option: Quantitative Settings, Advanced Settings, Quantitative Testing, and Organize Samples.

Advanced Settings

The Advanced Setting tab should only be used if you need to switch to “Ratio-Based Normalization.” The more advanced statistical concepts in Q+ have all been implemented only for “Intensity Based Normalization.” In versions of Q+ prior to 3.4, however, the data was normalized using the ratio-based method. This option is still available for looking at datasets that were analyzed with these earlier versions, but we recommend not using this option for new data.

Quantitative Settings

The Quantitative Settings option contains three tabs:

Normalization tab

It is almost always a good idea to normalize the data. Scaffold Q+ offers several normalization options, and depending on your experimental design and the nature of your data you will need to select the ones that best suit your dataset.

Calculation Type

Median or Mean

The Intensity-Based Normalization algorithm can calculate the normalization, roll up the results from the reporter-ion level to the protein level and do statistical tests on the data using either medians or means. In general, we recommend using the median option since it is more robust,
but in this task, we will see how to use the tools provided in Scaffold Q+ to make the proper decision for your data.

The median is more appropriate than the mean unless the data is well behaved in the following ways:

- Little missing data
- Few outliers
- Little data below the limit of detection
- Little distortion due to saturation at high intensities
- Little variation in the spread of the data between datasets
- Little skew - the dataset looks symmetric.
- No fat, or thin, tails – the dataset looks like a normal distribution.

Choosing the mean normalization implies that the data will be more or less normal and that Scaffold Q+ should use the standard statistical tests, the t-test and ANOVA test. Missing data is assigned an arbitrary low number for the calculations.

Choosing the median normalization implies that you want to use the non-parametric statistical tests that are more robust to missing and non-normal data. To examine these properties in this dataset, click “Cancel” to exit Quantification Setup and return to Scaffold Q+.

Click on the Statistics View icon and select the Pre/Post Normalization tab in the Raw Intensity Distribution pane.

Notice that in this case, there is very little missing data, that all of the quantitative samples have approximately the same variance, and that relatively little adjustment would be required to align the samples.

Now, select the Raw Quantifiable Intensity tab.

Notice in this graph whether the data looks like a “normal” distribution. That is, is it symmetric or is it skewed? Does it have fat tails or no tails? Select each quantitative sample in turn from the Primary Sample dropdown and examine its histogram.

Some of the samples show longer and less symmetric tails than the first, indicating a degree of deviation from the normal distribution.
Select the Signal to Noise Tab and cycle through the samples Primary Sample dropdown.

If the distribution shows a good dynamic range, this indicates that the data is “well-behaved.”

If there are many low values (as in the following example from a different dataset), it may indicate that the collision energy was tuned to fragment the peptide backbone rather than the reporter ions. Low reporter ions may be less accurate if the instrument has a limited dynamic range.

Because this data looks quite “well-behaved,” you may wish to try changing the “Calculation Type” to “Mean” and observing how this changes the normalized values. To do this Select Quant>Quantitative Settings and select “Mean” from the first dropdown on the “Normalization” tab.

Then click “Apply” to reanalyze the data.

Reference Type

Scaffold Q+ shows the results in terms of fold change, or log base 2 of the fold change. A fold change implies that a reference value is in the denominator. Scaffold Q+ gives you two options for this denominator.

The default reference is “Average Protein Reference.” This means that all the spectra in all of the reference channels are averaged (median or mean) together to give one value as a reference. This allows the graphs to show how much variation there is in this reference channel and it allows all the data to be divided by the same reference. In particular this reference denominator is the same for all the samples irrespective of which n-plex (for example iTRAQ 4-plex, TMT 6-plex, or SILAC 2-plex).

The alternative reference option is “Individual Spectrum Reference.” This option might be useful for a design something like the following. Suppose the experiment has 2 mice. Each mouse was measured at 4 time points. The first iTRAQ 4-plex has all the measurements for mouse 1. The second iTRAQ 4-plex has all the measurements for mouse 2. Now using Individual Spectrum Reference compares each mouse’s measurements back to its first time point.

Select an option, then click “Apply.” Scaffold Q+ will reanalyze the data.
Normalization Type

The final option on the Normalization Tab of the Quantification Setup screen is “Normalization Type.” Regardless of which option is selected, Scaffold Q+ will perform some normalization. It will make adjustments for different levels of reporter ions within the same spectra, and adjust from one protein to another within a sample. The user has the option to decide whether or not the program should normalize between samples. The default option is “Normalize Between Samples,” but there is an option to turn this function off.

If your experimental design involved enriching for different proteins in different samples or using different dilutions in each sample, for example, you might need to turn off normalization between samples.

Minimum Dynamic Range

Select the second tab of the Quantification Setup Screen, “Minimum Dynamic Range.” This displays a histogram similar to the Signal to Noise graph in the Statistics View but in this case it includes all channels combined together.

This screen is used for selecting the minimum value that will be considered valid for quantitation. Slide the control at the bottom of the graph and watch it adjust the minimum dynamic range value in the box to the right, and also move the vertical dotted line to indicate which values are being excluded.

Other Settings

The remaining tab on the Quantitation Setup screen is “Other Settings.” There is only one option here. By default, Scaffold Q+ uses only peptides that are exclusive to a protein for quantitation, since it is impossible to determine the relative contributions of the different proteins to the intensity of a shared peptide. If you choose to include shared peptides by checking “Use Non Unique Peptides” Scaffold Q+ will use the full intensity values in each of the proteins. It will not attempt to apportion them.
Task 8: Quantitative Testing

The goal of quantitative analysis is to detect differential expression of a protein between categories of samples. While the fold-change values can indicate differential expression, statistical testing is needed to determine whether the evidence for the presence or absence of differential expression is conclusive.

Scaffold Q+ provides a variety of quantitative tests, each of which is appropriate to use for certain types of experiments. The set of Quantitative Tests available for selection depends on the “Calculation Type” selected in the “Quantitative Settings.”

Average-Based Tests

We will begin with the Parametric tests, which are only available for calculation type “Mean” so you will need to prepare for this task by first setting the “Calculation Type” to “Mean.” Select “Quant>Quantitative Settings,” go to the “Normalization” tab and select “Mean” and click “Apply.”
**T-Test**

The T-Test is a standard statistical test that assesses whether the means of two groups are statistically different from each other. This is useful for determining whether a protein is differentially expressed between two categories. The result of a T-Test is a p-value which is the probability that there is no real difference between the two groups. In order to perform a T-test, then, we need to specify exactly two categories to compare. As a result you should remove two of the samples from the “Selected Samples” leaving two samples, each in its own category.

Select the two samples, and click “Remove.” The T-Test is automatically selected. Click “Apply.”

The Samples View opens, with the T-Test p-values displayed. Significant values are colored. In addition to the T-Test, Scaffold Q+ also performs a Random Permutation test, which is described below. The p-value is shown in green if it is significant in both tests, and in yellow if it is only significant in the currently displayed test, in this case the T-Test. The yellow color is to caution the user that the p-value may be based on a small number of measurements or unreliable data, and that it should be examined in more depth.

**ANOVA**

If you add the two “Removed Samples” again, The ANOVA TEST is automatically selected. Click “Apply” and the Samples View returns, showing ANOVA results.

The ANOVA calculates the probability that all of the categories are the same, so a significant p-value only tells us that at least one of the categories is different from another, but it does not tell us which categories are different. The “Mean-Density Chart” on the Proteins Page can be helpful in...
understanding the differences. *Randomized permutation test*

The Permutation Test will work on two or more categories. Select “Permutation Test” and click “Apply”

The permutation test between categories assumes that there is no differential expression between the categories when the differential expression is calculated using the f-statistic and then sets out to calculate a p-value to find out how likely it is that this assumption is wrong.

The trick the permutation test uses to calculate the p-value depends upon building a reference histogram of f-statistic values. This reference histogram is made by permuting the measurements between the categories.

Let’s look at an example to see how this works. Figure 1 shows some sample data in two categories that will be used for the permutation example. From Figure 2, the probability density chart, you can see that the distributions overlap a fair amount and the confidence intervals of the medians overlap a smaller amount.

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>R2</td>
<td>T1</td>
<td>T2</td>
</tr>
<tr>
<td>9.86</td>
<td>9.72</td>
<td>10.0</td>
<td>10.1</td>
</tr>
<tr>
<td>9.75</td>
<td>9.90</td>
<td>9.95</td>
<td>10.1</td>
</tr>
</tbody>
</table>

**Figure 1:** Sample data for calculating a reference distribution.

**Figure 2.** Are these two distributions significantly different or not?

The f-statistic calculated for the distance between these two distributions is 0.34. We will now build a reference distribution to see if this 0.34 is statistically significant.

The assumption we are setting up to refute is that there is no differential expression. Another way of looking at this is that the data in the “Reference” samples and the data in the “treated” samples are all part of the same distribution. Since all this data comes from the same distribution, it makes no difference if we swap some of the data between the categories. This swap is shown in Figure 3.
Figure 3. The bottom table results from swapping 9.86 from the reference category with 10.0 in the treated category. This lower configuration will show a slightly smaller f-statistic measuring the differential expression.

Since the values in the Reference category are now bigger and the values in treated category are smaller, it is easy to see that the f-statistic will be in the swapped case. The value of this f-statistic is calculated and plotted on a histogram.

Now this procedure is repeated with some other values swapped between the treated and the Reference categories, the f-statistic calculated and plotted on the histogram. This is repeated over and over swapping not just one value but any random combination of one to all six values between categories. This repeated swapping, or permutation, gives the test its name. The histogram that results from permuting the values 1,000 times is plotted in Figure 4. This is the reference f-statistic histogram for this protein.

**Randomized Permutation Test**

P-Value calculated from percentages:

\[
\text{P-Value} = \frac{33}{33 + 967}
\]
Figure 4: Reference histogram resulting from permuting the values between the categories and calculating the f-statistic for each permutation.

The reference f-statistic histogram shows how the calculated differential expression varies under random permutations when it is assumed that there is no real differential expression.

The f-statistic that was calculated for the actual measured intensities is 0.34. This value is plotted as a red dashed line in Figure 6. Those 33 histogram values which are greater than the red line in Figure 6 arose quite by chance. This says that 33 out of 1000 times the measured differential expression (0.34) could be explained merely by random permutations. That is the p-value, the probability that the assumption is true, is 33/1000.

The reference f-statistic histogram shows how the calculated differential expression varies under random permutations when it is assumed that there is no real differential expression.

The f-statistic that was calculated for the actual measured intensities is 0.34. This value is plotted as a red dashed line in Figure 6. Those 33 histogram values which are greater than the red line in Figure 6 arose quite by chance. This says that 33 out of 1000 times the measured differential expression (0.34) could be explained merely by random permutations. That is the p-value, the probability that the assumption is true, is 33/1000.

A minor extension to the permutation test is to extrapolate the reference f-statistic distribution to more extreme values. As can be seen in Figure 5 this can give estimates of the p-value when the measured f-statistic is more extreme than any of the 1,000 permutations. As can also be seen in Figure 5, the data used for the extrapolation doesn't fit very tightly to the exponentially decreasing line which means that the resulting p-values are very approximate – they may be off by an order of magnitude.

Figure 5. When the measured f-statistic, red line, is more extreme than the 1,000 permutations, the reference distribution can be extrapolated.
Significant Permutation Test p-values are indicated in green or yellow depending on whether the corresponding standard test (T-Test or ANOVA) also gives a significant result. Green means that both tests found significant differential expression, yellow means that the standard test disagreed.

Finally, on the “Quantitative Testing” screen, there is an option to apply the Bonferroni Correction to the p-values calculated. This is a correction factor that adjusts for repeated comparisons, and reduces the false positive rate, but may be overly stringent and may result in some actual instances of differential expression being overlooked. To see the effect of the Bonferroni Correction, choose the ANOVA, check the Bonferroni box, apply it and compare the results with those obtained previously.

**Median-Based Tests**

Now select “Quant>Quantitative Testing.” The Options for Statistical Testing with Median-Based analysis will be displayed.

Because medians are not normally distributed, non-parametric tests must be used. The Mann-Whitney Test is the non-parametric version of the T-Test. It compares two categories, so it requires that you select only two of the samples. To make the Mann Whitney test available, highlight two of the samples in the “Selected Samples” box and click “Remove.”

**Mann Whitney Test**

The Mann Whitney test will now appear in bold type and the radio button will be selected. Click “Apply.” The Samples View opens with the Mann Whitney p-values displayed.
Scaffold Q+ also performs a Randomized Permutation test in addition to the requested statistical tests. As before, the test result is shown in green if it is significant in both tests, and in yellow if it is only significant in the currently displayed test, in this case Mann Whitney. The yellow color is to caution the user that the p-value may be based on a small number of measurements or unreliable data, and that it should be examined in more depth.

If you return to “Quantitative Testing,” and add the two Removed Samples back to the “Selected Samples” box, the Kruskal-Wallis Test will be selected. This is a non-parametric version of the ANOVA and measures variation among three or more samples. It tells if at least one of the samples is different, but does not tell which.

The Randomized Permutation Test is also available in the Median-based Analysis screen. It is calculated similarly to the Mean-based test, but uses medians instead of means in its calculations. The Bonferroni Correction can be applied to any of the Median-based tests as well.
Release Information

This document is applicable for Scaffold, Release 4.0 or greater, and is current until replaced.

Copyright © 2013. Proteome Software, Inc., All rights reserved.

The information contained herein is proprietary and confidential and is the exclusive property of Proteome Software, Inc.. It may not be copied, disclosed, used, distributed, modified, or reproduced, in whole or in part, without the express written permission of Proteome Software, Inc.

Limit of Liability Proteome Software, Inc.. has used their best effort in preparing this guide. Proteome Software, Inc. makes no representations or warranties with respect to the accuracy or completeness of the contents of this guide and specifically disclaims any implied warranties of merchantability or fitness for a particular purpose. Information in this document is subject to change without notice and does not represent a commitment on the part of Proteome Software, Inc. or any of its affiliates. The accuracy and completeness of the information contained herein and the opinions stated herein are not guaranteed or warranted to produce any particular results, and the advice and strategies contained herein may not be suitable for every user.

The software described herein is furnished under a license agreement or a non-disclosure agreement. The software may be copied or used only in accordance with the terms of the agreement. It is against the law to copy the software on any medium except as specifically allowed in the license or the non-disclosure agreement.

Trademark - The name Proteome Software, the Proteome Software logo, Scaffold, Scaffold Q+, Scaffold Q+S, and the Scaffold, Scaffold Q+, and Scaffold Q+S logos are trademarks or registered trademarks of Proteome Software, Inc. All other products and company names mentioned herein may be trademarks or registered trademarks of their respective owners.

Customer Support - Customer support is available to organizations that purchase Scaffold, Scaffold Q+ or Scaffold Q+S and that have an annual support agreement.

Contact Proteome Software at:
Proteome Software, Inc.
1340 SW Bertha Blvd, Suite 10
Portland, OR 97219
1-800-944-6027 (Toll Free)
1-503-245-4910 (Fax)
www.proteomesoftware.com

Document Version Number Scaffold Q+Tutorial-002
Document Release Date June 5, 2013