

# **Progenesis Q1 for proteomics User Guide**

**Analysis workflow guidelines for HDMse and MSe  
data**

# **Waters**

**THE SCIENCE OF WHAT'S POSSIBLE.™**

## Contents

Introduction .....	3
How to use this document .....	3
How can I analyse my own runs using LC-MS? .....	3
LC-MS Data used in this user guide.....	3
Workflow approach to LC-MS run analysis .....	4
Restoring a Tutorial Archive .....	5
Stage 1: Import Data and QC review of LC-MS data set .....	6
Stage 2: Automatic Alignment of your data .....	7
Stage 3: Licensing .....	9
Stage 4: Review Alignment .....	9
Reviewing Quality of Alignment.....	11
Stage 5A: Filtering .....	14
Stage 5B: Reviewing Normalisation .....	18
Stage 6: Experiment Design Setup for Analysed Runs.....	21
Stage 7: Validation, review and editing of results .....	24
Stage 8: Peptide Statistics on Selected Features .....	38
Stage 9: Identify peptides .....	42
Stage 10: Refine Identifications.....	45
Stage 11: Resolve Conflicts .....	47
Stage 12: Review Proteins .....	51
Stage 13: Protein Statistics .....	55
Stage 14: Exporting data .....	56
Stage 15: Reporting.....	57
Appendix 1: Stage 1 Data Import and QC review of LC-MS data set .....	59
Appendix 2: Manual assistance of Alignment.....	65
Appendix 3: Licensing runs (Stage 3) .....	70
Appendix 4: Within-subject Design.....	71
Appendix 5: Power Analysis (Peptide Stats) .....	73
Appendix 6: Waters Machine Specification .....	74

## Introduction

This user guide takes you through a complete analysis of 9 LC-MS runs with 3 groups (3 replicate runs per group) using the unique Progenesis QI for Proteomics workflow. It starts with LC-MS data folders loading then Alignment, followed by Peak Detection that creates a list of interesting features (peptides) which are explored within Peptide Stats using multivariate statistical methods then onto Protein identity and Protein Stats.

To allow ease of use the tutorial is designed to start with the restoration of an Archived experiment where the data files have already been loaded. The document covers all the stages in the workflow, so if you are using your own data files please refer to Appendix 1 (page 59) then start at page 6.

## How to use this document

You can print this user guide to help you work hands-on with the software. The complete user guide takes about 60 minutes. This means you can perform the first half focused on LC-MS run alignment and analysis then complete the second half of analysis exploring comparative differences and Protein identity at a convenient time.

## How can I analyse my own runs using LC-MS?

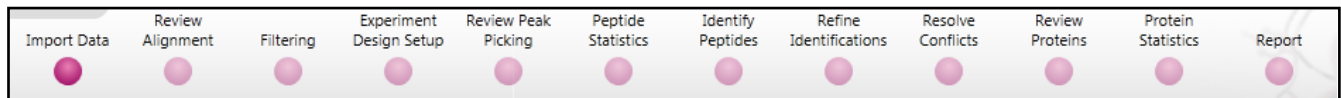
You can freely explore the quality of your LC-MS data using Data Import and then licence your own LC-MS runs using this evaluation copy of Progenesis LC-MS. Instructions on how to do this are included in a section at the end of the user guide document.


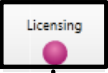
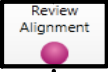


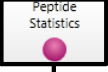

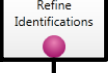

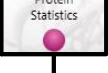

## LC-MS Data used in this user guide

For the purposes of this data set the MS<sup>E</sup> parameters were set to 250:125:1000 instead of the default settings as defined in Appendix 1 (page 59). This was to done to reduce the time taken to demo the data analysis.

## Workflow approach to LC-MS run analysis

Progenesis QI for Proteomics adopts an intuitive **Workflow** approach to performing comparative LC-MS data analysis. The following user guide describes the various stages of this workflow (see below) focusing mainly on the stages from Alignment to Report.

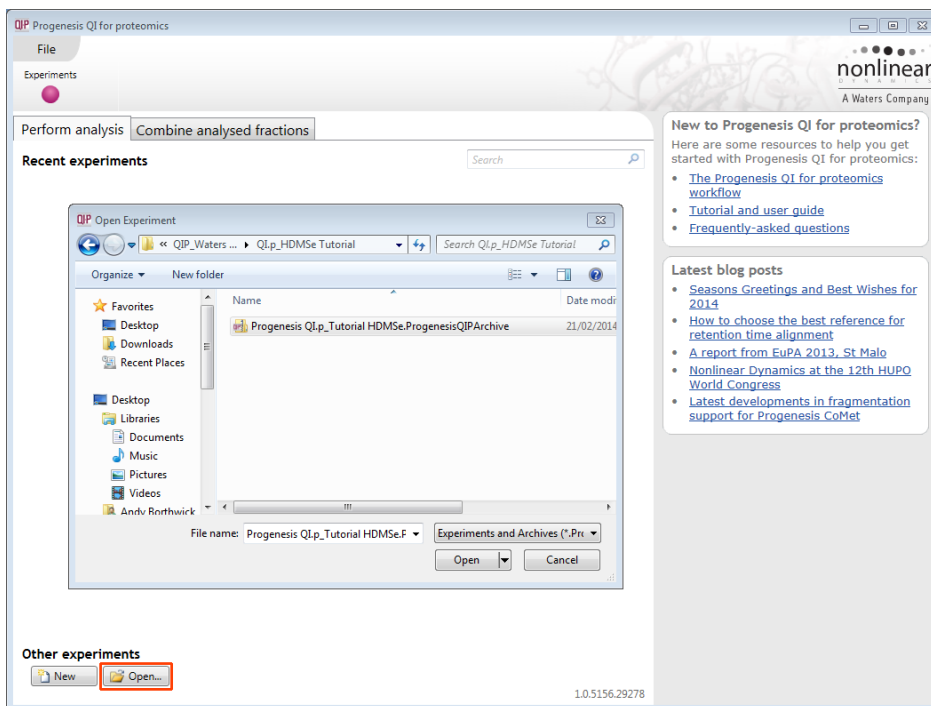


Stage	Description	Page
	<b>LC-MS Import Data:</b> Selection and review of data files for analysis	6
	<b>Automatic Alignment:</b> Automatic Reference selection and Alignment	7
	<b>Licensing:</b> allows licensing of individual data files when there is no dongle attached (Appendix 3)	9
	<b>Review Alignment:</b> review of automatic and manual run alignment	9
	<b>Filtering:</b> defining filters for peaks based on Retention Time, m/z , Charge and Number of Isotopes.	14
	<b>Review Normalisation:</b> explaining LC-MS normalisation	18
	<b>Experiment Design Setup:</b> defining one or more group set ups for analysed aligned runs	21
	<b>Review Peak Picking:</b> review and validate results, edit peak detection, tag groups of peaks and select peaks for further analysis	24
	<b>Peptide Statistics:</b> performing multivariate statistical analysis on tagged and selected groups of peptides	38
	<b>Identify Peptides:</b> managing export of MS/MS spectra to, and import of peptide ids from Peptide Search engines	42
	<b>Refine Identifications:</b> manage peptide ids and filters	45
	<b>Resolve Conflicts:</b> validation and resolution of peptide id conflicts for data entered from Database Search engines	47
	<b>Review proteins:</b> review protein and peptide identity	51
	<b>Protein Statistics:</b> multivariate statistical analysis on proteins	55
	<b>Report:</b> generate a report for proteins and/or peptides	57

## Restoring a Tutorial Archive

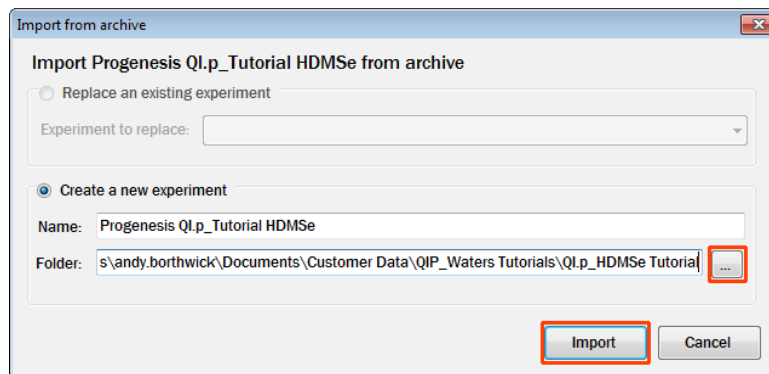
The data in the tutorial archive (Progenesis QI.P\_Tutorial HDMSe.ProgenesisQIArchive) is part of the Progenesis QI for Proteomics demo suite. Download and uncompress the zipped archive from the link on the website.

Now restore the uncompressed tutorial archive file. To do this, first locate the Tutorial Archive file from Progenesis QI for proteomics using the **Open** button and press Open.

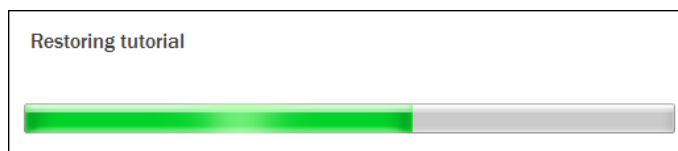


This opens the 'Import from archive' dialog.

Select the **Create a new experiment** option and select the folder in which you placed the archive, using the icon (to the right).



Then press **Import**.



**Note:** use the **Replace an existing experiment** option if you want to over-write an existing version of the tutorial.

**Tip:** at each stage in the software there are links to more information and help on the website.

## Stage 1: Import Data and QC review of LC-MS data set

The tutorial will now open at the **Import Data** stage (see below).

The screenshot displays the Progenesis Q1 software interface during the 'Import Data' stage. The 'File' menu is highlighted, and the 'Import Data' section is active. The main window shows a large 2D plot labeled 'A\_01' with a retention time axis (min) ranging from 20 to 90 and an m/z axis ranging from 500 to 1500. The plot shows a dense distribution of data points. The interface includes a 'Section Complete' button and a 'Help' icon.

Each data file appears as a 2D representation of the run. At this stage you will be warned if any of the data files have been 'centroided' during the data acquisition and conversion process.

**Note:** the **Experiment Properties** are available from the File menu. These were selected when the experiment was created (see Appendix 1, page 59).

The 'Experiment Properties' dialog box shows the following information:

- Runs in this experiment: 9
- Machine resolution: High resolution
- Peak processing: Profile data

A 'Close' button is visible at the bottom right.

**Tip:** the **'Exclude areas from selected run'** facility allows you to examine and exclude areas (usually early and/or late in the LC dimension (Retention Time)) that appear excessively noisy due to capture of data during column regeneration. This is not required for this data set.

The 'Actions' menu is shown with the following options:

- Mask areas for peak picking...
- Remove run
- Total ion intensity: 6.096e+008
- Masked areas: none

**Note:** use the **Remove Run** to remove run(s) from the current experiment.

Now start the Alignment process

## Stage 2: Automatic Alignment of your data

The process of alignment of your data can be started before the import of your data has been completed by clicking on **Start alignment process**.

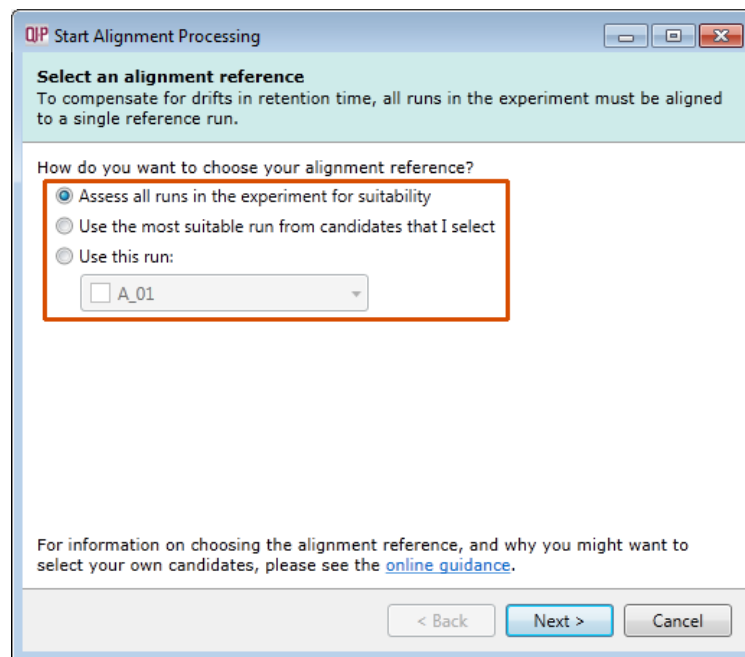
**2 Start the alignment process**  
While your runs are importing, click the button below to:

- Select [alignment reference](#) candidates
- Determine the best of the candidates
- Align all runs to that reference run

Start alignment process

During this process the software will Align all your runs to a Reference run which can either be selected automatically by the software or manually selected by you.

In this tutorial example you have 9 runs which have already been imported so to start the process click on **Start alignment process**



Progenesis QI for proteomics provides three methods for choosing the alignment reference run, as seen below:

### 1. Assess all runs in the experiment for suitability

This method compares every run in your experiment to every other run for similarity.

The run with the greatest similarity to all other runs is chosen as the alignment reference.

If you have no prior knowledge about which of your runs would make a good reference, then this choice will normally produce a good alignment reference for you. This method can take a long time

### 2. Use the most suitable run from candidates that I select

This method asks you to choose a selection of reference candidates, and the automatic algorithm chooses the best reference from these runs.

When you have some prior knowledge of your runs suitability as references:

runs from pooled samples

runs for one of your experimental conditions will contain the largest set of common peptides.

### 3. Use this run

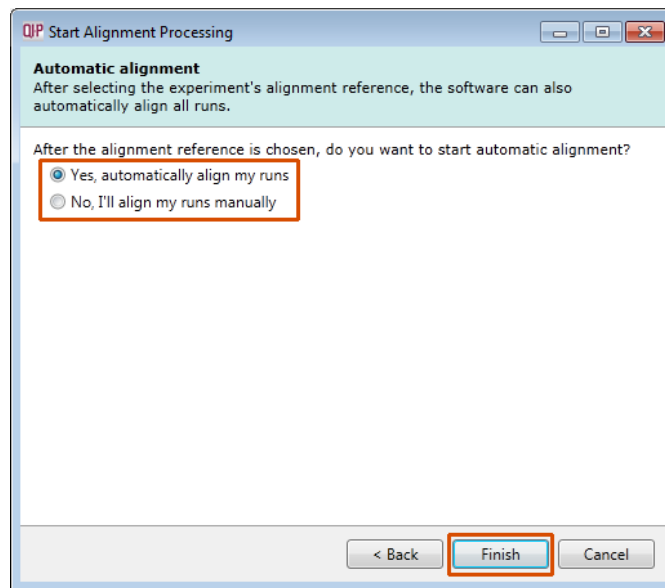
This method allows you to manually choose the reference run.

Manual selection gives you full control, but there are a couple of risks to note:

- If you choose a pending run which subsequently fails to load, alignment will not be performed.
- If you choose a run before it fully loads, and it turns out to have chromatography issues, alignment will be negatively affected (for this reason we recommend that you let your reference run fully load and assess its chromatography before loading further runs).

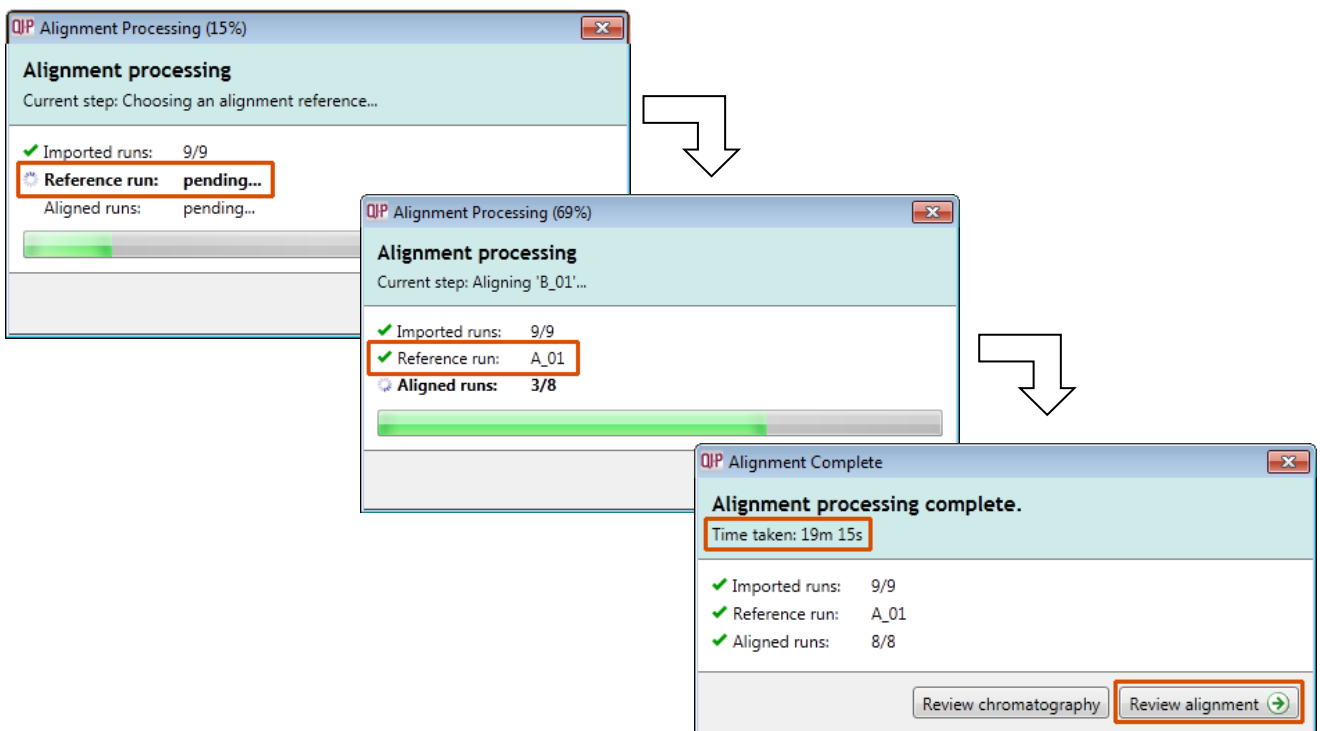
For this tutorial we will select the first option (See Appendix 1, page 59 for more details on using the other options).

You will now be asked if you want to align your runs automatically or manually.



Select automatically and click finish.

The Alignment process starts with the automatic selection of **A\_01** as the reference



Once the Reference run has been chosen the automatic alignment is then performed. As the whole process proceeds you get information on what stage has been performed and also the % of the process that has been completed.

When the Alignment completes you can either review the chromatography or go to the Review Alignment using the options on the Alignment Dialog.

Click **Review Alignment**.



## Stage 3: Licensing

This stage in the analysis workflow will **only** appear if you are using 'Unlicensed' data files to evaluate the software and have no dongle attached.



For details on how to use Licensing go to Appendix 3 (page 70)

If you are using the tutorial archive, this page will not appear as the data files are licensed.

## Stage 4: Review Alignment

At this stage Progenesis QI for proteomics **Review Alignment** opens displaying the alignment of the runs to the Reference run (A\_01).

**Table of Alignment Vectors and Scores**

Run	Include?	Vectors	Score
A_01	<input checked="" type="checkbox"/>	Ref	
A_02	<input checked="" type="checkbox"/>	1974	97.5%
A_03	<input checked="" type="checkbox"/>	2043	97.4%
B_01	<input checked="" type="checkbox"/>	1990	96.6%
B_02	<input checked="" type="checkbox"/>	1901	96.0%
B_03	<input checked="" type="checkbox"/>	1930	96.2%
C_01	<input checked="" type="checkbox"/>	1962	95.2%
C_02	<input checked="" type="checkbox"/>	2031	96.0%
C_03	<input checked="" type="checkbox"/>	1955	94.0%

Ion maps: ■ Alignment target ■ Run being aligned

Alignment quality: ■ Good ■ OK ■ Needs review

Section Complete

**Vector Editing (Window A):** is the main alignment area and displays the area defined by the current **focus** rectangle shown in Window C. The current run is displayed in green and the reference run is displayed in magenta. Here is where you can review in detail the vectors and also place the manual alignment vectors when required.

**Transition (Window B):** uses an **alpha blend** to animate between the current and reference runs. Before the runs are aligned, the features appear to move up and down. Once correctly aligned, they will appear to

pulse. During the process of adding vectors, this view can be used to zoom in and also navigate thus helping with accurate placement of manual vectors.

**Ion Intensity Map (Window C):** shows the **focus** for the other windows. When you click on the view the blue rectangle will move to the selected area. The focus can be moved systematically across the view using the cursor keys. The focus area size can be altered by clicking and dragging out a new area with the mouse. This view also provides a visual quality metric for the Alignment of the runs, from green through yellow to red.

**Total Ion Chromatograms (Window D):** shows the current **total ion** chromatogram (green) overlaid on the Reference chromatogram (magenta). As the features are aligned in the **Vector Editing** view the chromatograms become aligned. The retention time range displayed is the vertical dimension of the Focus Grid currently displayed in the **Ion Intensity map** (Window C).

### Layout of Alignment

To familiarize you with Progenesis QI for proteomics Alignment, this section describes the various graphical features used in the alignment of the LC-MS runs

To setup the display so that it looks similar to the one above:

- In the Run table click on Run B\_02 to make it current. You will now be looking at the alignment of B\_02 to A\_01 in the Unaligned view. Now drag out an area to review on the **Image intensity map**. The other 3 views will update to reflect the new focus.

Run	Include?	Vectors	Score
A_01	<input checked="" type="checkbox"/> <input type="checkbox"/>	Ref	
A_02	<input checked="" type="checkbox"/> <input type="checkbox"/>	1974	97.5%
A_03	<input checked="" type="checkbox"/> <input type="checkbox"/>	2043	97.4%
B_01	<input checked="" type="checkbox"/> <input type="checkbox"/>	1990	96.6%
B_02	<input checked="" type="checkbox"/> <input type="checkbox"/>	1901	96.0%
B_03	<input checked="" type="checkbox"/> <input type="checkbox"/>	1930	96.2%
C_01	<input checked="" type="checkbox"/> <input type="checkbox"/>	1962	95.2%
C_02	<input checked="" type="checkbox"/> <input type="checkbox"/>	2031	96.0%
C_03	<input checked="" type="checkbox"/> <input type="checkbox"/>	1955	94.0%

### Reviewing quality of alignment vectors

After **Automatic alignment** the number of vectors and Quality Scores will be updated on the **Runs** panel and the vectors will appear (in blue) on the view.

The screenshot displays the Progenesis QI software interface for proteomics. The top menu bar includes options like File, Review Alignment, Filtering, Experiment Design Setup, Review Peak Picking, Peptide Statistics, Identify Peptides, Refine Identifications, Resolve Conflicts, Review Proteins, Protein Statistics, and Report. The main workspace is divided into several panels:

- Review Alignment:** Contains instructions for aligning retention times and reviewing alignment quality. It includes a 'Show Aligned' dropdown, 'Show Unaligned', and 'Remove Vectors' buttons.
- Vector editing:** A scatter plot showing alignment vectors as blue dots, with axes for Retention Time (min) and m/z.
- Transition:** A scatter plot showing transition vectors as black dots, with axes for Retention Time (min) and m/z.
- Ion intensity map:** A heatmap showing ion intensity across retention time and m/z, with a green focus area.
- Total ion chromatogram:** A line graph showing intensity versus retention time, with a green line for the current run and a magenta line for the reference run.

At the bottom left, there is a 'Run' table with columns for Run, Include?, Vectors, and Score. Run B\_02 is highlighted in green. At the bottom right, there is a 'Section Complete' button.

Where the alignment has worked well, the alignment views will look as below with the Ion Intensity Map showing green indicating good quality alignment and the Transition view showing features pulsing slightly but not moving up and down.

To simulate poor alignment, place a single manual vector on the Vector editing view (Window A). To do this click and drag out a single vector then release the mouse button. By doing this a single manual vector will appear with a length corresponding to the 'drag'.

**Note:** the manual vector is **red**, to distinguish it from the automatic vectors (blue)

The effect of adding this incorrect manual vector is to reduce the Alignment score and also cause a significant proportion of the Alignment quality squares to turn red on the Ion Intensity Map (as shown below).

Using a **Simulated** miss-aligned example to explain the review process for alignment, the alignment looks as below with a region of poor alignment (highlighted in red).

**Review Alignment**

Sample ions are aligned to compensate for drifts in retention time between runs.

[Learn about the visualisations shown here](#)

- Align retention times automatically**  
For maximum reproducibility, the software can automatically align your runs.
- Review the alignment**  
Using the quality control measures, review and edit the runs' alignment:
  - Order the runs by alignment score and start by selecting the first run
  - Within each run, inspect and edit any areas rated as Needs Review

[Learn about the review and editing process](#)

Run	Include?	Vectors	Score
A_01	<input checked="" type="checkbox"/>	Ref	
A_02	<input checked="" type="checkbox"/>	1974	97.5%
A_03	<input checked="" type="checkbox"/>	2043	97.4%
B_01	<input checked="" type="checkbox"/>	1990	96.6%
B_02	<input checked="" type="checkbox"/>	1901	96.0%
B_03	<input checked="" type="checkbox"/>	1930	96.2%
C_01	<input checked="" type="checkbox"/>	1891	91.9%
C_02	<input checked="" type="checkbox"/>	2031	96.0%
C_03	<input checked="" type="checkbox"/>	1955	94.0%

Ion maps: ■ Alignment target ■ Run being aligned

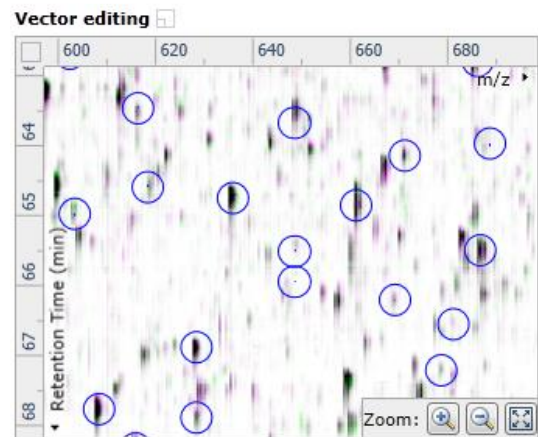
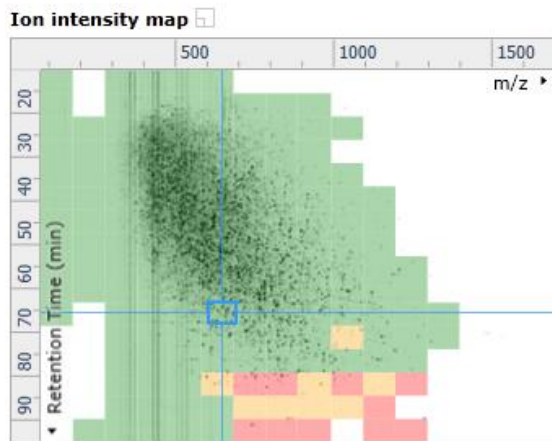
Alignment quality: ■ Good ■ OK ■ Needs review

Section Complete

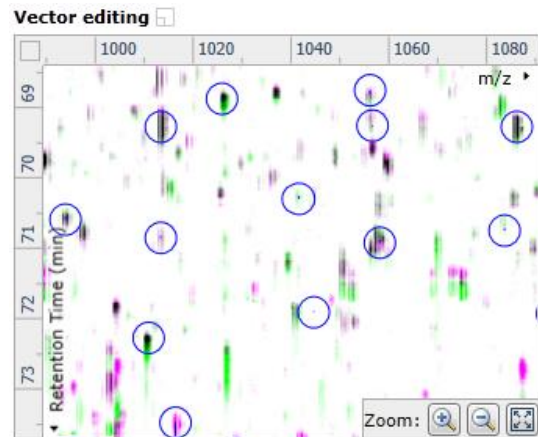
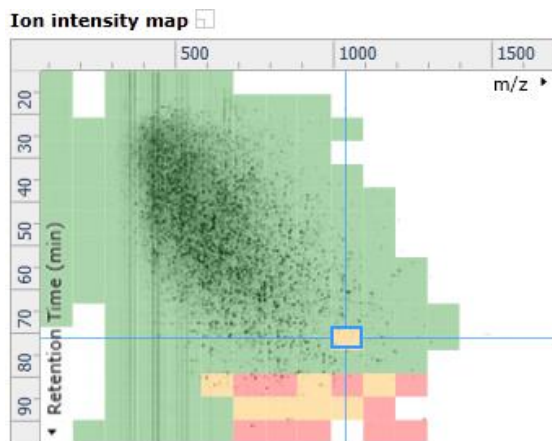
## Reviewing Quality of Alignment

At this point the quality metric, overlaid on the Ion Intensity Map as coloured squares, acts as a guide drawing your attention to areas of the alignment. These range from Good (Green) through OK (Yellow) to Needs review (Red). Drag out a 'Focus' area that corresponds to one of the coloured squares. Three example squares are examined here.

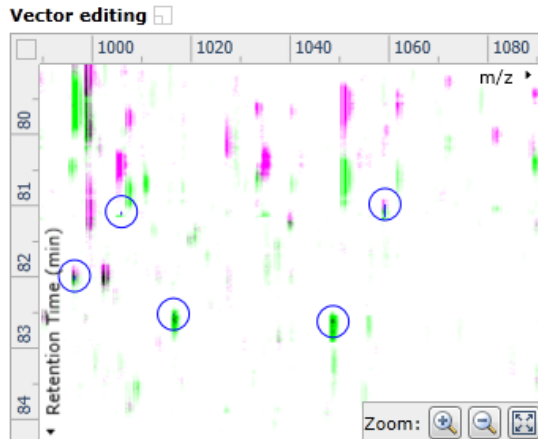
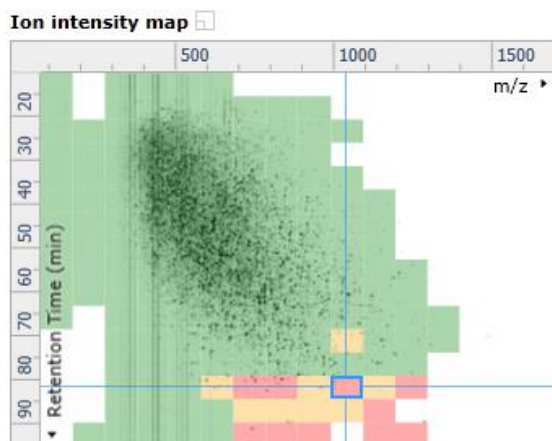
For a 'green' square the majority of the data appears overlapped (black) indicating good alignment. When viewed in the Transition view the data appears to pulse.



For a 'yellow' square some of the data appears overlapped (black) indicating OK alignment. When viewed in the Transition view some of the data appears to pulse.

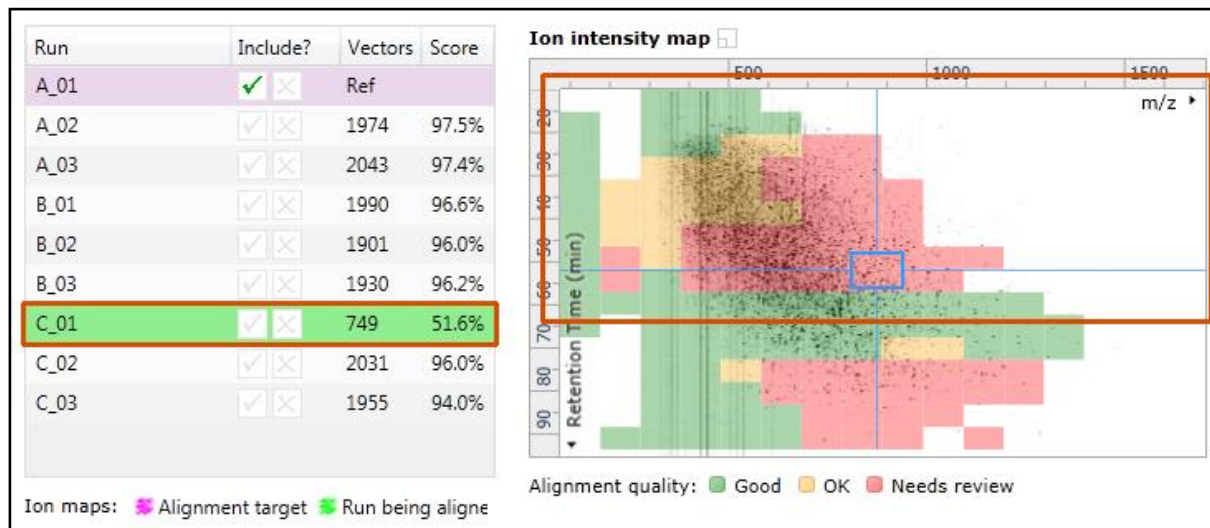


For a 'red' square little if any of the data appears overlapped (black) indicating questionable alignment. When viewed in the Transition view little data appears to pulse.



**Note:** the coloured metric **should be used as a guide**. In cases where there are a few 'isolated' red squares this can also be indicative of 'real' differences between the two runs being aligned and should be considered when examining the overall score and surrounding squares in the current alignment.

The weighted average of the individual squares gives the overall percentage score for each run's alignment.



**Note:** a marked red area combined with a low score clearly indicates a 'miss alignment' and may require some manual intervention (see Appendix 2, page 65).

The alignment quality of this data set does not require any manual intervention so before going to the next section make sure you have Removed all manual vectors and re-performed the Automatic alignment. To do this for C\_01 first select Remove 'All vectors in the whole run' and then click Align runs automatically.

**Remove Vectors**

- Automatic vectors in the current area
- Automatic vectors in the whole run
- All vectors in the current area
- All vectors in the whole run

**Section Complete**

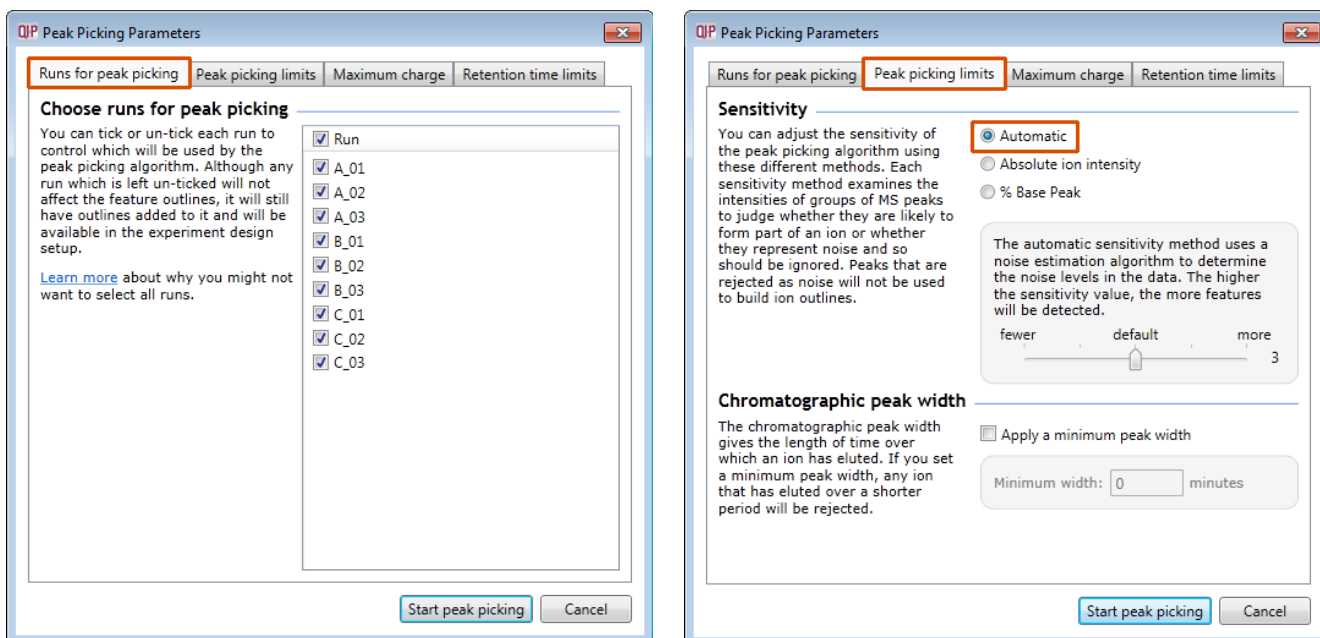
Having aligned the runs automatically, click **Section Complete** to move to Filtering.

## Stage 5A: Filtering

Now that you have reviewed your aligned Runs, you are ready to analyse them. Move to the **Filtering** stage, by either clicking on **Section Complete** (bottom right) or on Filtering on the workflow.

### Peak Picking Parameters

The Peak Picking Parameters dialog opens, showing all the runs in the current experiment and a tick against each run. This is the default setting, where the peak picking algorithm uses information from all of the runs to contribute to the pattern of feature outlines.



**Tip:** It may be appropriate **only** to pick peaks that are present in a limited number of your runs. In which case un-tick the runs that you do **NOT** want to contribute to the feature detection pattern. This may be important when one or more of the runs appear noisy due to non-optimal chromatography or sample handling.

**Note:** features outlines will be added to 'un-ticked' runs; although these runs will not contribute to the peak picking pattern.

**Tip:** depending on run quality, a suggested minimum number of ticked runs should include at least one replicate of each experimental condition.

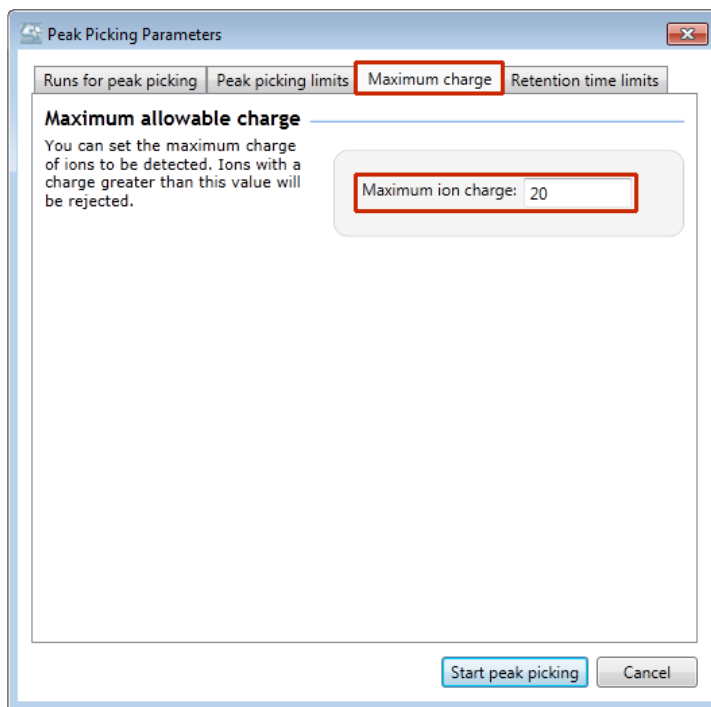
The sensitivity of the detection can be controlled by adjusting settings under the **Peak picking limits** tab.

**Note:** for all 3 **Sensitivity** modes a Chromatographic peak width (Retention time window) for the peaks **can** be set by applying a minimum retention window or peak width in minutes.

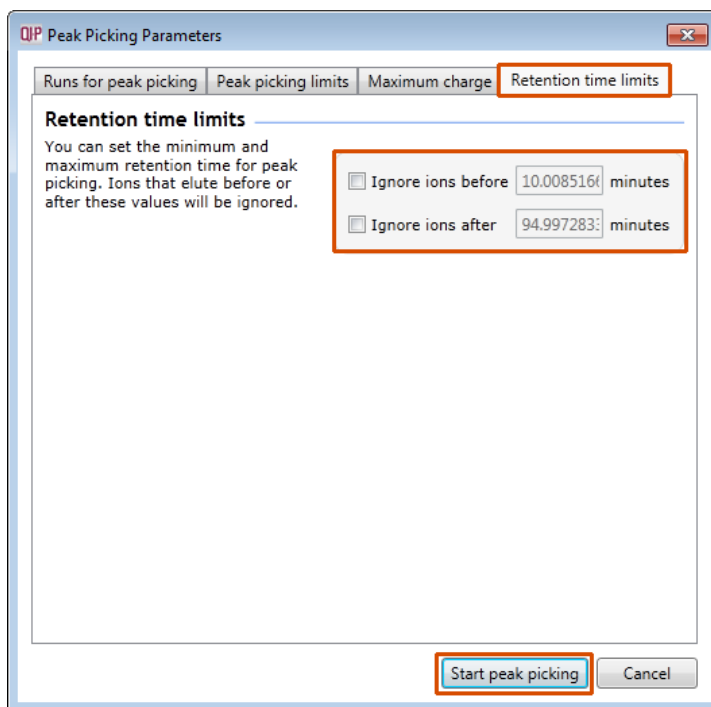


For the runs in this user guide, we will use the default settings for the **Automatic** method.

The third tab allows you to set the **Maximum charge** of the ions, which will be detected. The default setting is a charge state of 20.



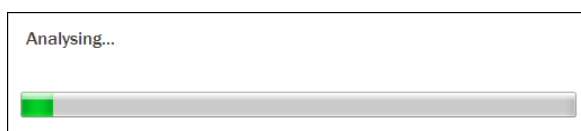
Finally, you can set **Retention time limits** for the detection. The default limits are displayed in the unticked boxes and detection will use these values whenever the box(s) are unticked.



For this data set we will use the default settings.

Press **Start peak picking** to start the detection process.

During the few minutes that the automatic analysis requires, a progress bar will appear telling you that it is Analysing.



On completion of analysis, the Filtering stage will open displaying the number of features detected, in this example 50790.

The screenshot displays the Progenesis QI software interface during the Filtering stage. The window title is "QIP Progenesis QI.P\_Tutorial HDMSe - Progenesis QI for proteomics". The top menu bar includes "File", "Import Data", "Review Alignment", "Filtering", "Experiment Design Setup", "Review Peak Picking", "Peptide Statistics", "Identify Peptides", "Refine Identifications", "Resolve Conflicts", "Review Proteins", "Protein Statistics", and "Report". The "Filtering" stage is active, indicated by a pink dot.

The "Filter Features" panel on the left contains the following information:

- Filter Features:** You can filter features that you do not wish to include in your analysis by using the criteria below.
- Select all features matching the following filters:
  - Inside area
  - With charge
  - Number of isotopes
- Buttons: "Delete 0 Matching Features" and "Delete 50790 Non-Matching Features".
- Summary: "(50790 features in total)"
- Normalisation:** Note that all remaining features will be used in the normalisation calculation. If deleting a substantial number of features, you should review the normalisation afterwards.
- Button: "Review normalisation >>"

The main plot is a scatter plot showing Retention time (min) on the y-axis (ranging from 10 to 79.4) and m/z on the x-axis (ranging from 0 to 1500). The plot contains a dense cloud of blue data points. At the top of the plot, there are "Undo" and "Redo" buttons. At the bottom right of the plot area, there is a "Section Complete" button with a green arrow.

If required you can remove features based on position, charge state, number of isotopes or combinations of these feature properties.

For example if the Retention time limits you set before performing Peak Detection were not 'stringent' enough, you can delete detected features with early and late 'Retention times' selecting and setting the **Inside area** filter.

Set the thresholds in minutes, for example 15 and 90

All features contained outside the new retention time limits can then be deleted.



**Filter Features**

You can filter features that you do not wish to include in your analysis by using the criteria below.

Select all features matching the following filters:

- Inside area
  - m/z from [ ] to [ ]
  - Minutes from 15 to 90
- With charge
- Number of isotopes

Delete 50734 Matching Features

Delete 56 Non-Matching Features

(50790 features in total)

**Normalisation**

Note that all remaining features will be used in the normalisation calculation. If deleting a substantial number of features, you should review the normalisation afterwards.

Review normalisation >>

Section Complete

To remove the features outside the retention time limits, as shown in blue, (in this case 56), press the **Delete 56 Non-Matching Features** button.

In addition to setting limits for 'Retention time and m/z', features can also be selected based on charge or the number of isotopes present. This allows you to refine the selection through a combination of feature properties.

For example, when **With charge** is selected the number of features present at each charge state is displayed, these can be selected accordingly.

Area limits, charge state and number of isotopes can be combined to refine the feature selection.

**Filter Features**

You can filter features that you do not wish to include in your analysis by using the criteria below.

Select all features matching the following filters:

- Inside area
- With charge
  - Charge 1 (12147 features)
  - Charge 2 (25129 features)
  - Charge 3 (11224 features)
  - Charge 4 (1208 features)
  - Charge 5 (206 features)
  - Charge 6 (290 features)
  - Charge 7 (151 features)
  - Charge 8 (68 features)
  - Charge 9 (46 features)
  - Charge 10 (14 features)
  - Charge 11 (19 features)
  - Charge 12 (36 features)

**Tip:** when filtering on one property of the feature i.e. charge state, make sure you have 'collapsed' the other filters (see right)

**For this user guide**, we will remove features appearing before 15 min and after 90 min (as described above). Also delete a further 12526 features with a charge state of 1 or 8 and above by ticking the various options.

Hence all features with a charge state of 1 or 8 and above will appear **red** on the main view as you hold the cursor over the appropriate delete button.

To remove these features press **Delete 12526 Non Matching Features**.

The screenshot shows the Progenesis QI software interface during the filtering stage. The main window displays a scatter plot of Retention time (min) vs m/z. The left sidebar shows a list of charge states from 1 to 20, with Charge 1 and Charge 8 selected. Below the list, there are buttons for 'Delete 38208 Matching Features' and 'Delete 12526 Non-Matching Features', with the latter highlighted by a red box. An 'Undo' button is also highlighted by a red box. At the bottom right, a 'Section Complete' button is visible.

You can use the **Undo** button to bring back deleted features, however, when you move to the next section you will lose the capacity to undo the filter. Before moving on from filtering, you can review the normalisation of the experiment.

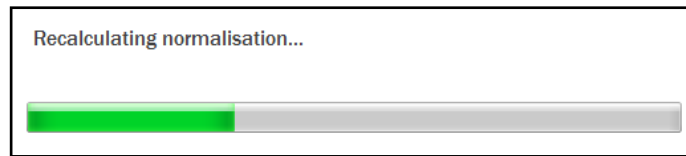
**Tip:** When you have reached the filtering stage, it is good practice to close the experiment and save an archive. This can be used to restore the unfiltered state if the filtering you have performed is too 'stringent'.

## Stage 5B: Reviewing Normalisation

Normalisation review is accessed from the button at the bottom right corner of the filtering page

The screenshot shows the Normalisation review dialog box. The dialog box contains the text: "Normalisation Note that all remaining features will be used in the normalisation calculation. If deleting a substantial number of features, you should review the normalisation afterwards." Below the text is a button labeled "Review normalisation >>" which is highlighted by a red box.

If you have filtered out a number of features from the original detection pattern then the normalisation will update.

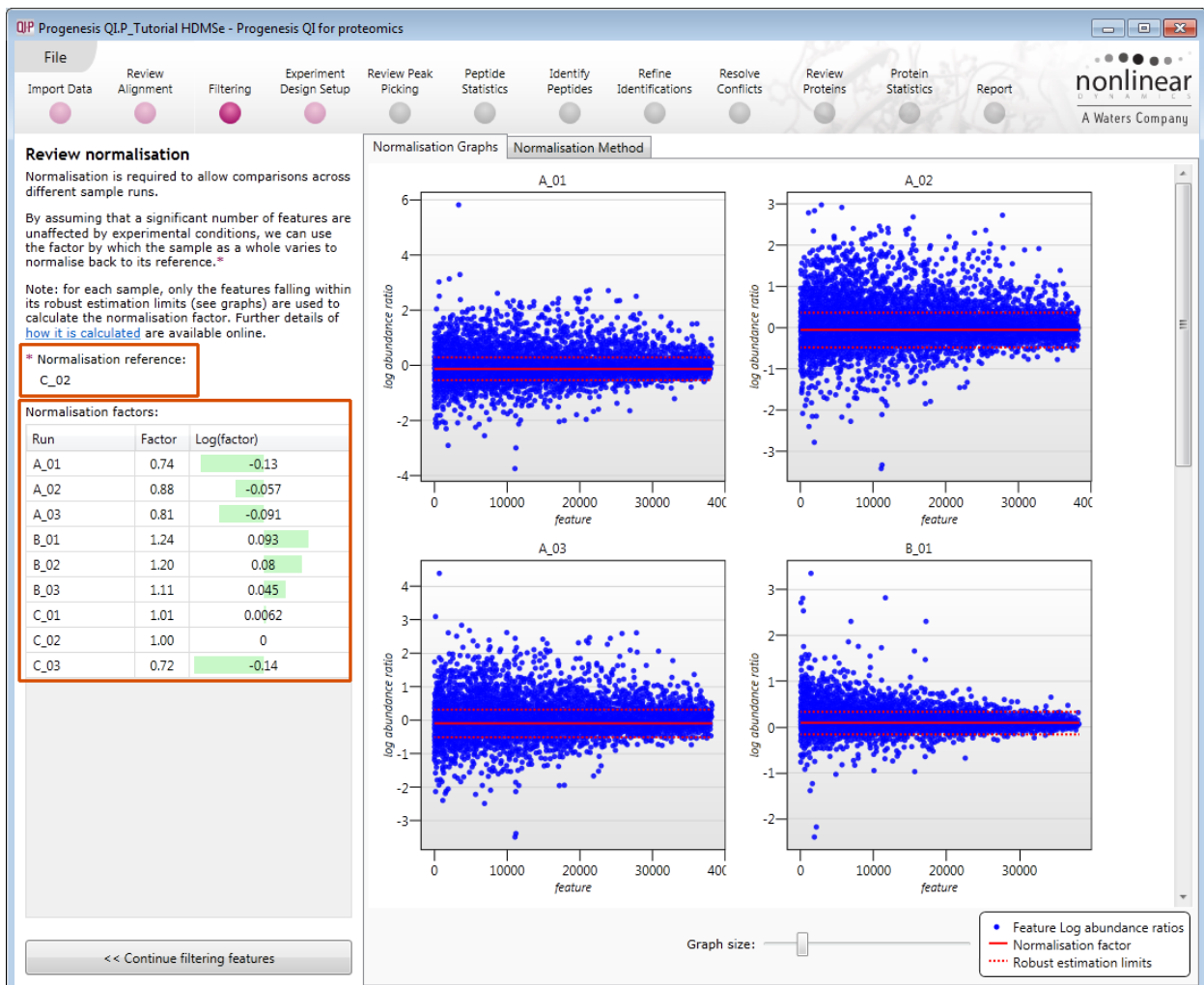


The **Review Normalisation** page will open displaying plots for the normalisation of all the features on each run.

This page in the workflow **does not** allow you to alter the Normalisation of your data but provides you with individual views for each run showing the data points used in the calculation of the normalisation factor for the run.

Alternatively, if you do not believe normalisation is necessary, you can opt to use un-normalised feature abundances for the rest of the analysis.

Normalisation factors are reported in the table to the left of the plots.

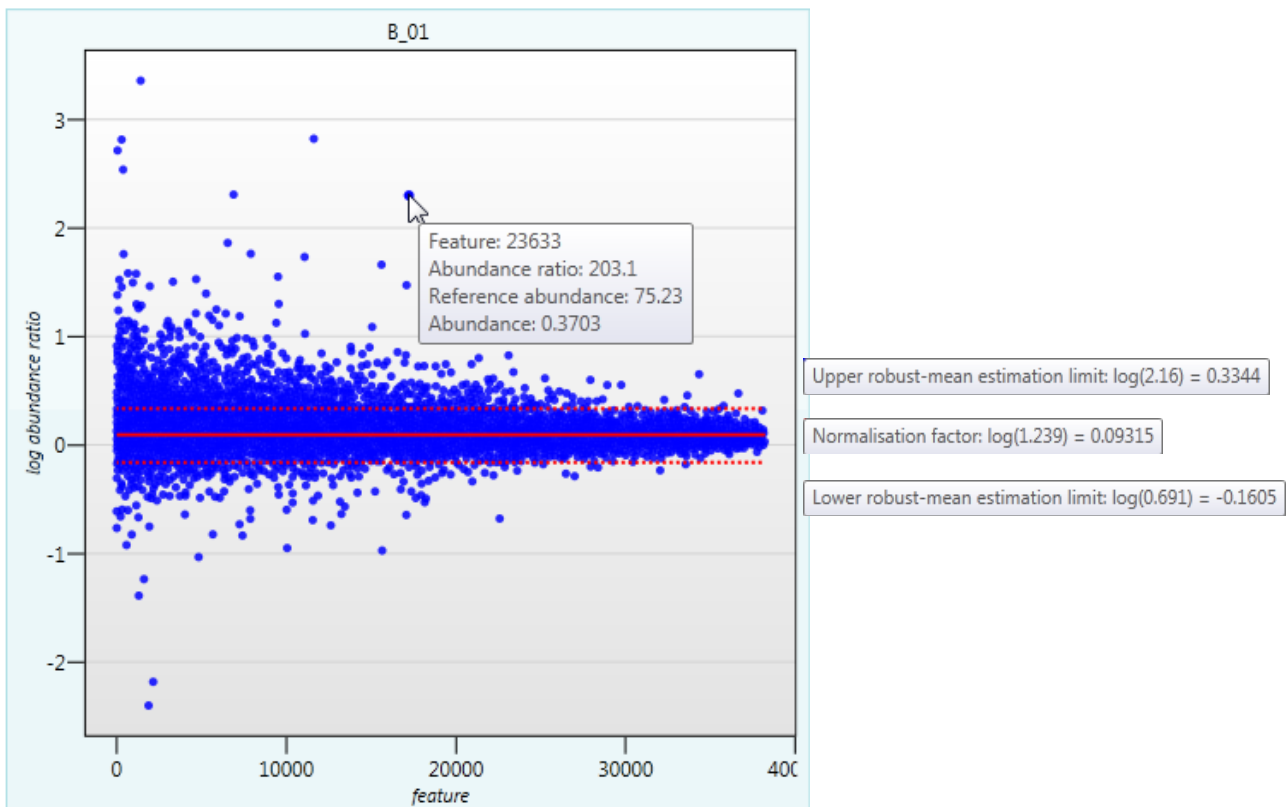


**Calculation of Normalisation Factor:**

Progenesis QI for proteomics will automatically select one of the runs that is 'least different' from all the other runs in the data and then set this to be the 'Normalising reference'. The run used, is shown above the table of Normalisation factors (in this example C\_02).

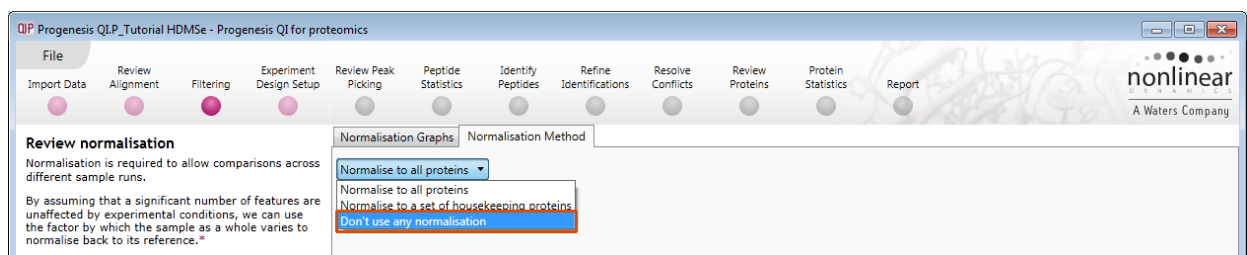
For each sample run, each blue dot shows the log of the abundance ratio for a different feature (normalisation target abundance/run abundance).

The details for individual features can be viewed as you hold the cursor over the dots on the plot.



On the graph the features are shown ordered by ascending mean abundance. The normalisation factor is then calculated by finding the mean of the log abundance ratios of the features that fall within the 'robust estimated limits' (dotted red lines). Features outside these limits are considered to be outliers and therefore will not affect the normalisation.

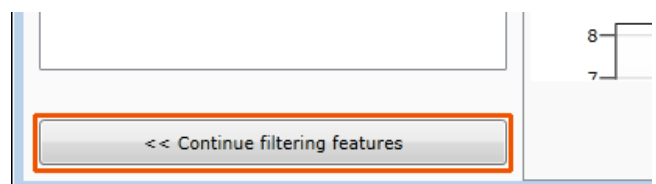
Finally, if you do **not** wish to work with normalised data then you can **use the raw abundances** by switching off the normalisation.



**Note:** once you have identified the features, you can then apply the **Normalise to a set of house keeping proteins** by using this option to locate and select the features.

For this experiment, you should leave the **Normalise to all proteins** option selected.

Now return to filtering by clicking on the button on the bottom left of the screen



For this example, we **DO NOT** do any additional Filtering so click on **Section complete**.

**Note:** if you do any extra filtering then **Normalisation recalculates** as you move to the next stage in the Workflow.

## Stage 6: Experiment Design Setup for Analysed Runs

At this stage in the workflow you can setup one or more experimental designs for your data.

There are two basic types of experimental designs:

**Between-subject design:** here samples from any given subject appear in only one condition. (i.e. control versus various drug treatments). The ANOVA calculation assumes that the conditions are independent and applies the statistical test that assumes the means of the conditions are equal.

QIP Progenesis QI.P\_Tutorial HDMSe - Progenesis QI for proteomics

File Import Data Review Alignment Filtering Experiment Design Setup Review Peak Picking Peptide Statistics Identify Peptides Refine Identifications Resolve Conflicts Review Proteins Protein Statistics Report

nonlinear DYNAMICS A Waters Company

New Help

Which experiment design type do you want to use for this experiment?

**Between-subject Design**  **Within-subject Design**

**Between-subject Design**

Do samples from a given subject appear in only one condition? Then use the between-subject design.

To set up this design, you simply group the runs according to the condition (factor level) of the samples. The ANOVA calculation assumes that the conditions are independent and therefore gives a statistical test of whether the means of the conditions are all equal.

**A** **Delete**  
A1 **Remove**  
A2 **Remove**  
A3 **Remove**

**C** **Delete**  
C1 **Remove**  
C2 **Remove**  
C3 **Remove**

[Add condition...](#)

**Within-subject Design**

Have you taken samples from a given subject under different conditions? Then use the within-subject design.

**Note:** you must have a sample from every subject for every condition to use a within-subject design.

For example, you would choose this type of design for a time series experiment where every subject has been sampled at each time point.

To set up this design, you tell the software not only which condition (factor level) each run belongs to but also which subject it came from. The software will then perform a repeated measures ANOVA.

A standard ANOVA is not appropriate because the data violates the ANOVA assumption of independence. With a repeated measures ANOVA individual differences can be eliminated or reduced as a source of between condition differences (which helps to create a more powerful test).

The within-subject design can be thought of as an extension of the paired-samples t-test to include comparison between more than two repeated measures.

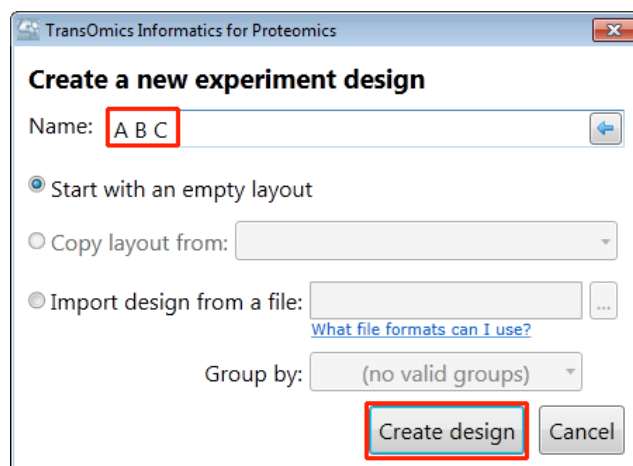
	Before	During	After
Patient X	X1	X2	X3
Patient Y	Y1	Y2	Y3
Patient Z	Z1	Z2	Z3

**Within-subject design:** here samples have been taken from a given subject under different conditions (i.e. the same subject has been sampled over a period of time or after one or more treatments). Here a standard ANOVA is not appropriate as the data violates the ANOVA assumption of independence. Therefore by using a *repeated measures* ANOVA, individual differences can be eliminated or reduced as a source of between condition differences. This within-subject design can be thought of as an extension of the paired samples t-test, including comparison between more than two repeated measures.

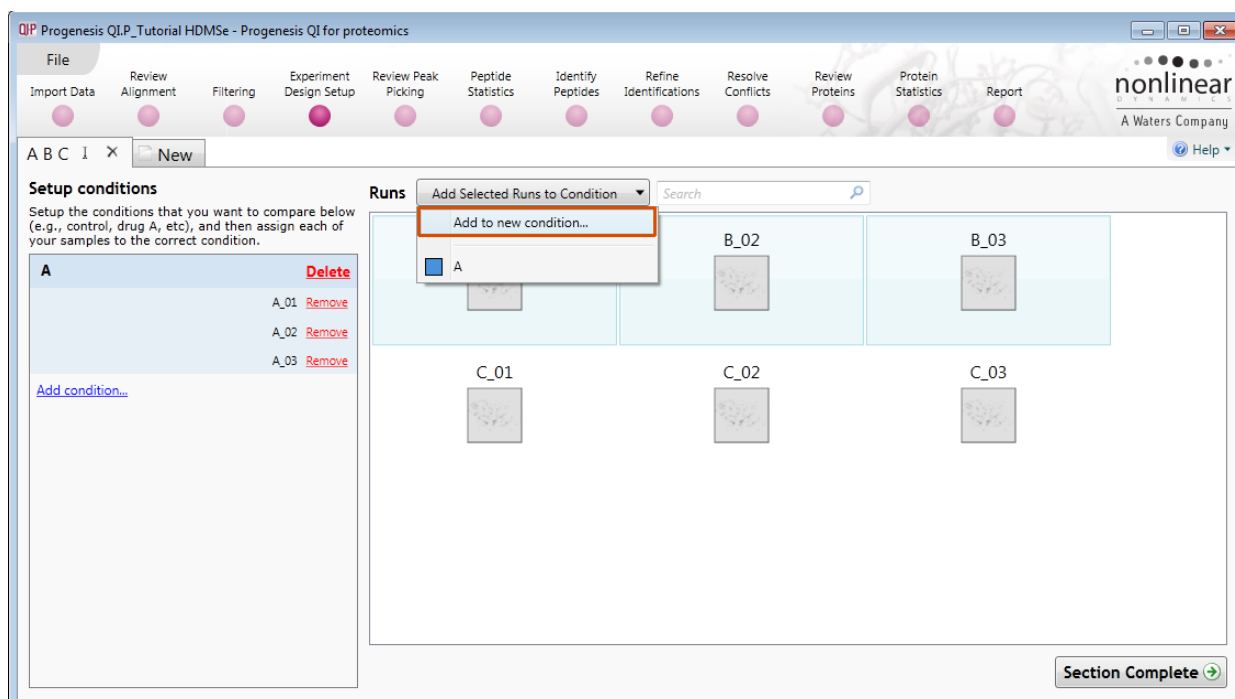
Additional information on how to apply the Within-subject Design is in **Appendix 4** page 71

This experiment contains 2 conditions: Control and Treated and uses the **Between-subject design** to group the analysed runs to reflect the Biological conditions in the original study.

To create a new **Between-subject Design** hold the cursor over this option and click to open the dialog.



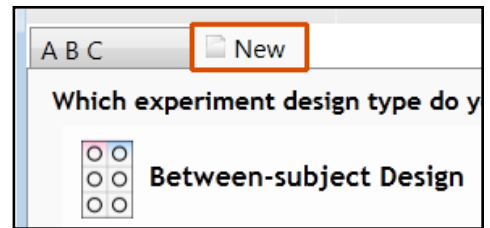
Give the new experimental design a name and then click **Create design**.



## To create a new condition

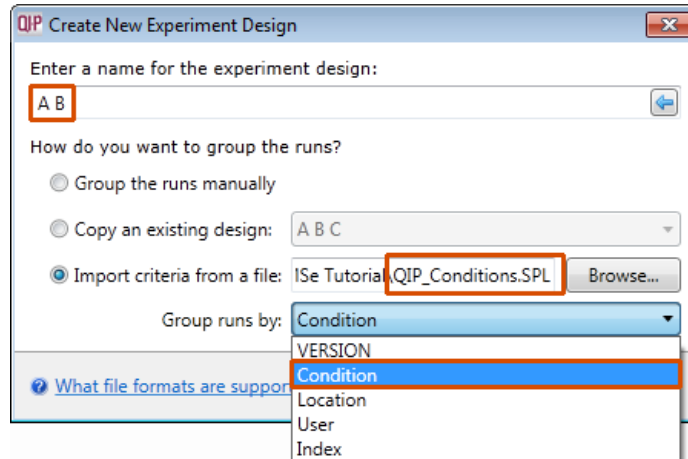
1. Select the runs for the condition by clicking on the required icon in the **Runs** panel, as shown.
2. Press the 'black triangle' next to the **Add Selected Runs to Condition** button on the main toolbar.
3. Select **Add to new condition...** from the drop down menu.
4. A new condition will appear in the **Conditions** panel on the left.
5. Rename the condition (e.g. C) by over typing the default name
6. Repeat steps 1 to 5 until all the runs are grouped into conditions.

An alternative way to handling the grouping of this set and other larger (and more complex) experimental designs is to make use of **sample tracking information** that has been stored in a spreadsheet at the time of sample collection and/or preparation.



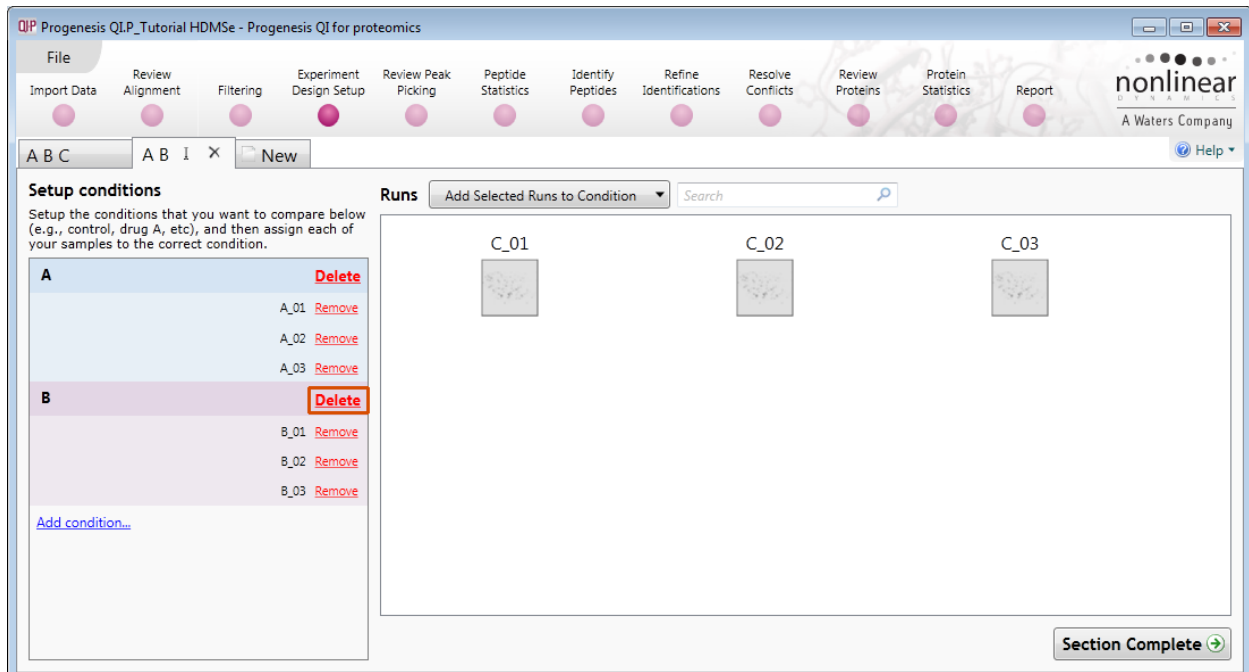
For this example there is a **QIP\_Conditions.spl** file available in the Experiment Archive you restored at the beginning of this tutorial exercise.

To use this approach select the **Import design from file** option from the New Experiment Design dialog. Then locate the QIP\_Conditions file and select what to **Group by**, for example: **Condition**.



When **Create design** is pressed the new tab refreshes to allow you to adjust the conditions.

Use **Delete** on the Conditions panel to remove conditions that are not required in this particular design.



**Note:** On deleting a condition the runs will reappear in the Runs window.

**Note:** both designs are available as separate tabs.

To move to the next stage in the workflow, **Review Peak Picking**, click **Section Complete**.

## Stage 7: Validation, review and editing of results

The purpose of this stage in the Workflow is to review the list of features using the visual tools provided and edit features if required.

The review stage has 5 display modes: 1D, 2D, 3D, Drift Time and Feature Details controlled by the tabs at the bottom left of the display and the expander bar to the right of the table. Each display has multiple views to allow comparative exploration of the detected features on the aligned LC-MS runs.

### Exploring analysed data using the Data displays

**Window A:** shows the list of features ranked by the p value for the one way **Anova** using the current Experiment Design (A B C).

**Note:** a value of 'Infinity' in the **Fold** column indicates 'Presence/Absence'

The screenshot shows the Progenesis QI software interface. At the top, there is a menu bar with options like File, Review, Alignment, Filtering, Experiment Design Setup, Review Peak Picking, Peptide Statistics, Identify Peptides, Refine Identifications, Resolve Conflicts, Review Proteins, Protein Statistics, and Report. Below the menu bar, there is a 'Review Features' section with a table of features. The table has columns for #, Anova (p), Fold, Tag, and Notes. The feature with # 27210 is highlighted in blue. To the right of the table, there are two plots: a Mass spectrum (38.737 min) and a Chromatogram (m/z = 906.4226). Below the plots, there are two more plots: a Run plot (D) and a 2D plot (E). At the bottom left, there are tabs for 1D Display, 2D Display, 3D Display, and Drift Time. At the bottom right, there is a 'Section Complete' button.

#	Anova (p)	Fold	Tag	Notes
36962	1.57E-10	Infinity		
45663	1.86E-10	Infinity		
27210	3.89E-10	8.16		Add a note...
23520	8.44E-10	12.8		
37291	1.06E-09	Infinity		
30088	1.27E-09	3.39		
46000	1.39E-09	Infinity		
44550	2.04E-09	Infinity		
48595	2.8E-09	Infinity		
8204	3.97E-09	2.94		
49503	4.49E-09	Infinity		

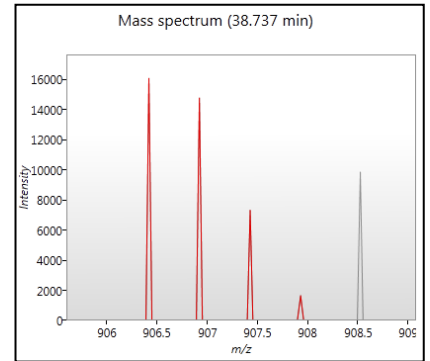
**Note:** by default all the features are included in the selection for the next section of the analysis.

To highlight a group of features drag out a selection on the table.

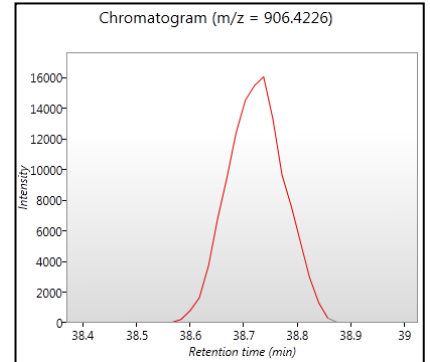


## The 1D Display

**Window B:** displays the Mass spectrum for the current feature on the selected Run (in window D).



**Window C:** displays the Chromatogram for the current feature on the selected Run (in window D).

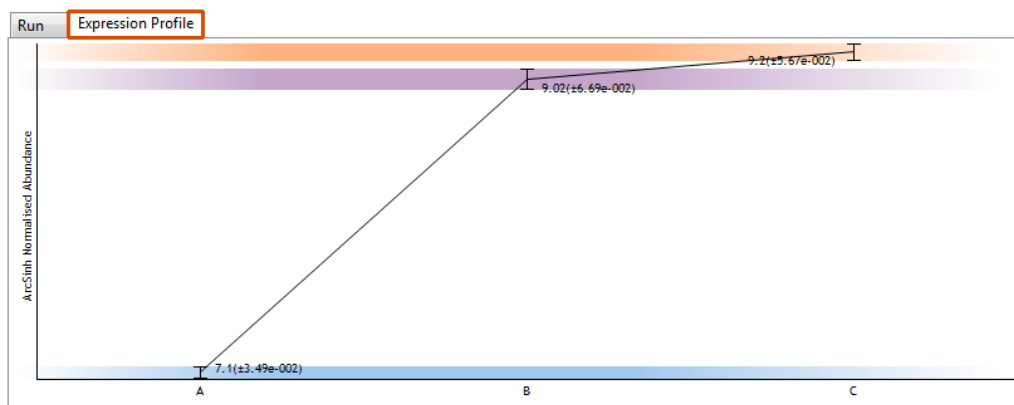
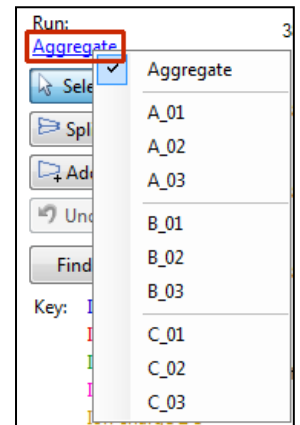


**Window D:** displays the details of the currently selected run. By default the selected run is an Aggregate view of all the aligned runs.

Details of individual runs can be viewed by using the 'Run' link and selecting the run you wish to view.

The feature editing tools are located in this window (see page 34 for functional explanation).

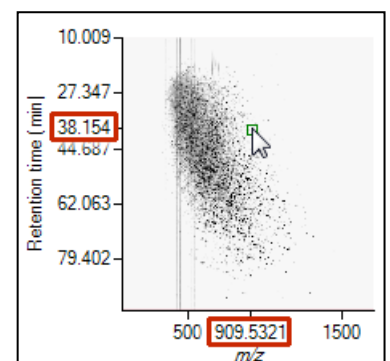
Clicking on the Expression Profile tab in Window D shows the comparative behaviour of the feature across the various biological groups based on group average normalised volume. The error bars show +/- 3 standard errors.



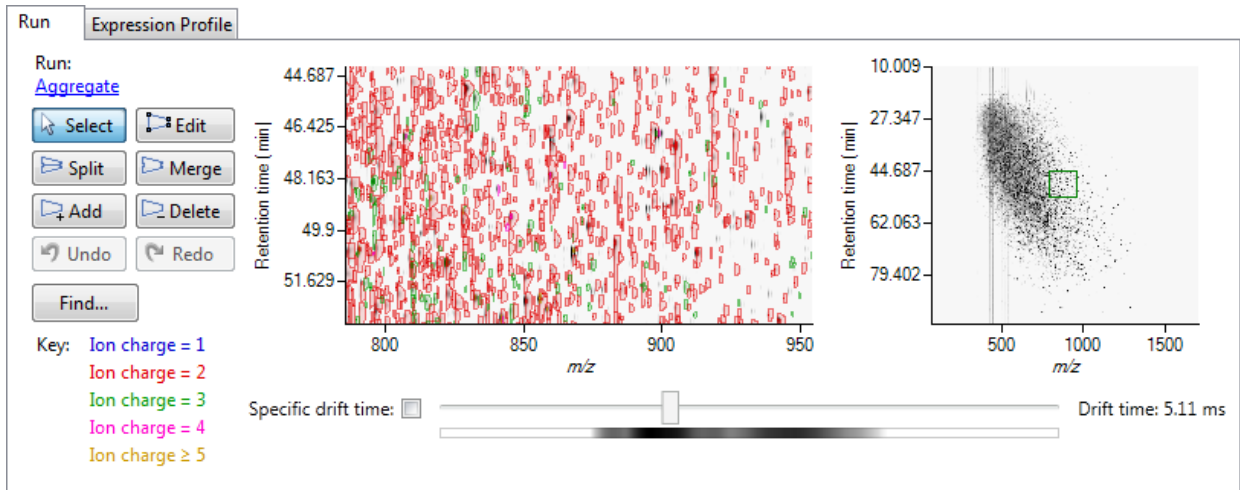
**Window E:** shows where the current feature is located on the LC-MS run by means of the 'Green' rectangle.

To change the current location, click on the image of the run (**note:** the retention time and m/z values update as you move the cursor around this view).

**Note:** doing this updates the focus of all the other windows.



You can also drag out an area (green square) on this view that will re-focus the other windows.



### The 2D Display

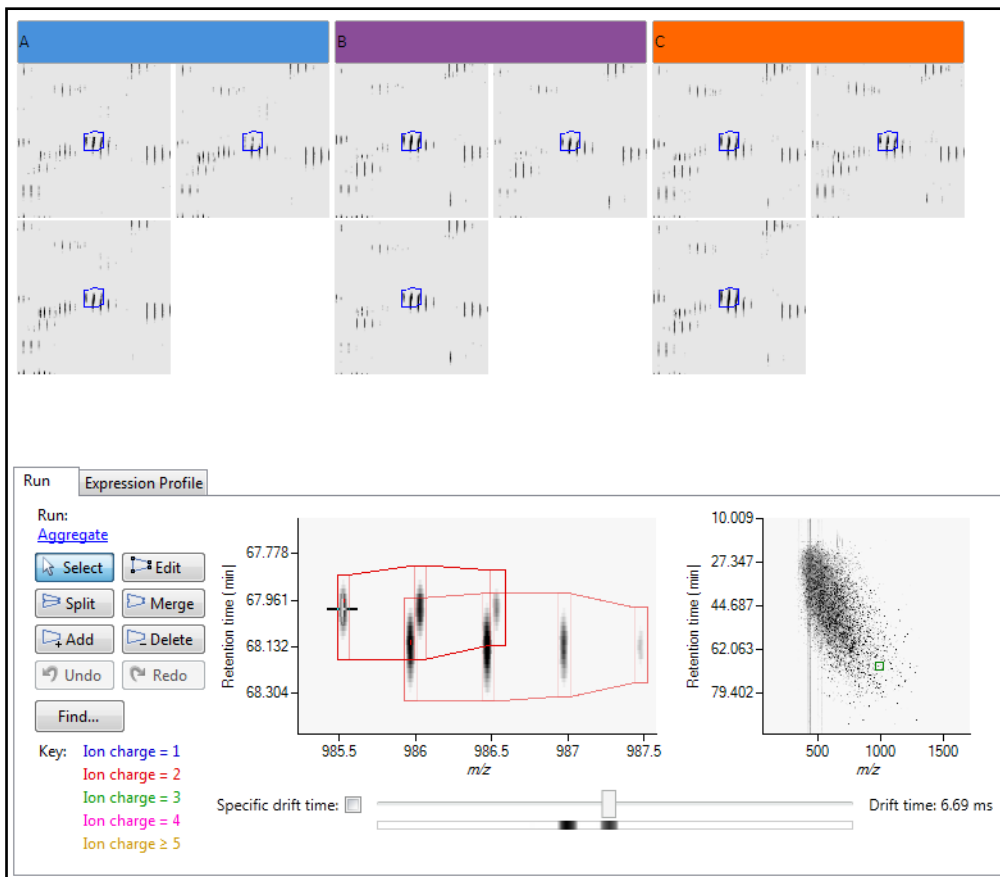
Windows **A**, **D** and **E**: perform the same functions across all 4 display modes.

In the 2D Montage mode, Window B displays a montage of the current feature across all the aligned LC-MS runs.



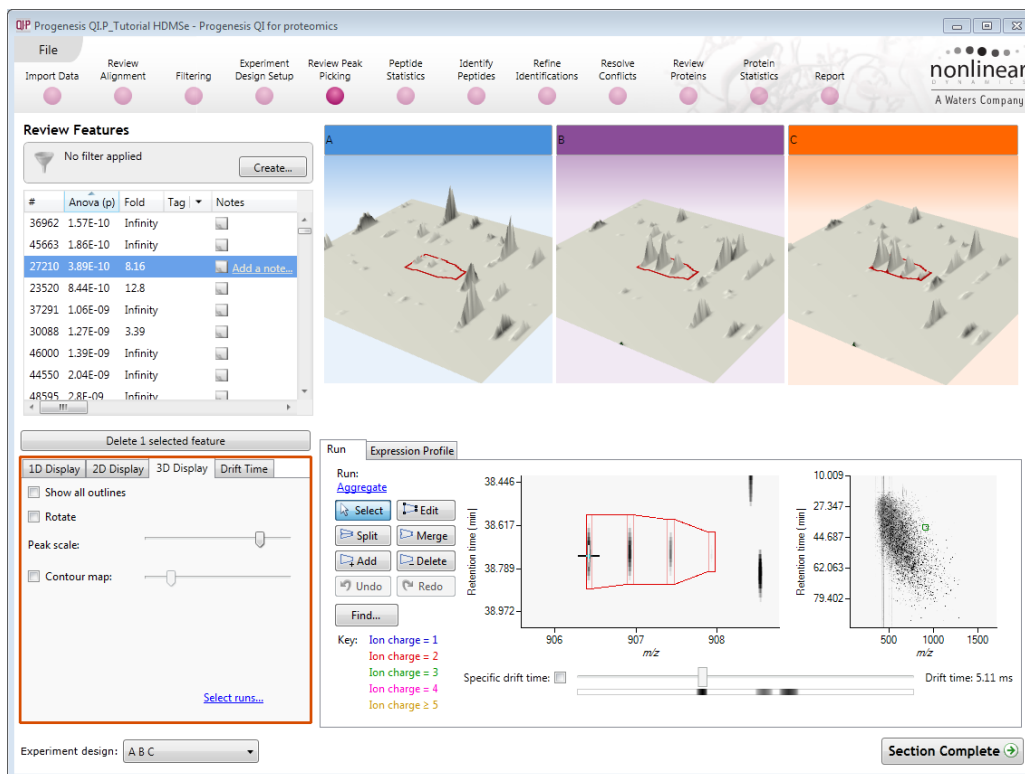
The appearance of the Montage (window B) is controlled by the panel on the bottom left of the display.

Using the the various views in the 2D display one can examine the feature detection in detail to validate the correct detection of even fully overlapping features as shown above.



### The 3D Display

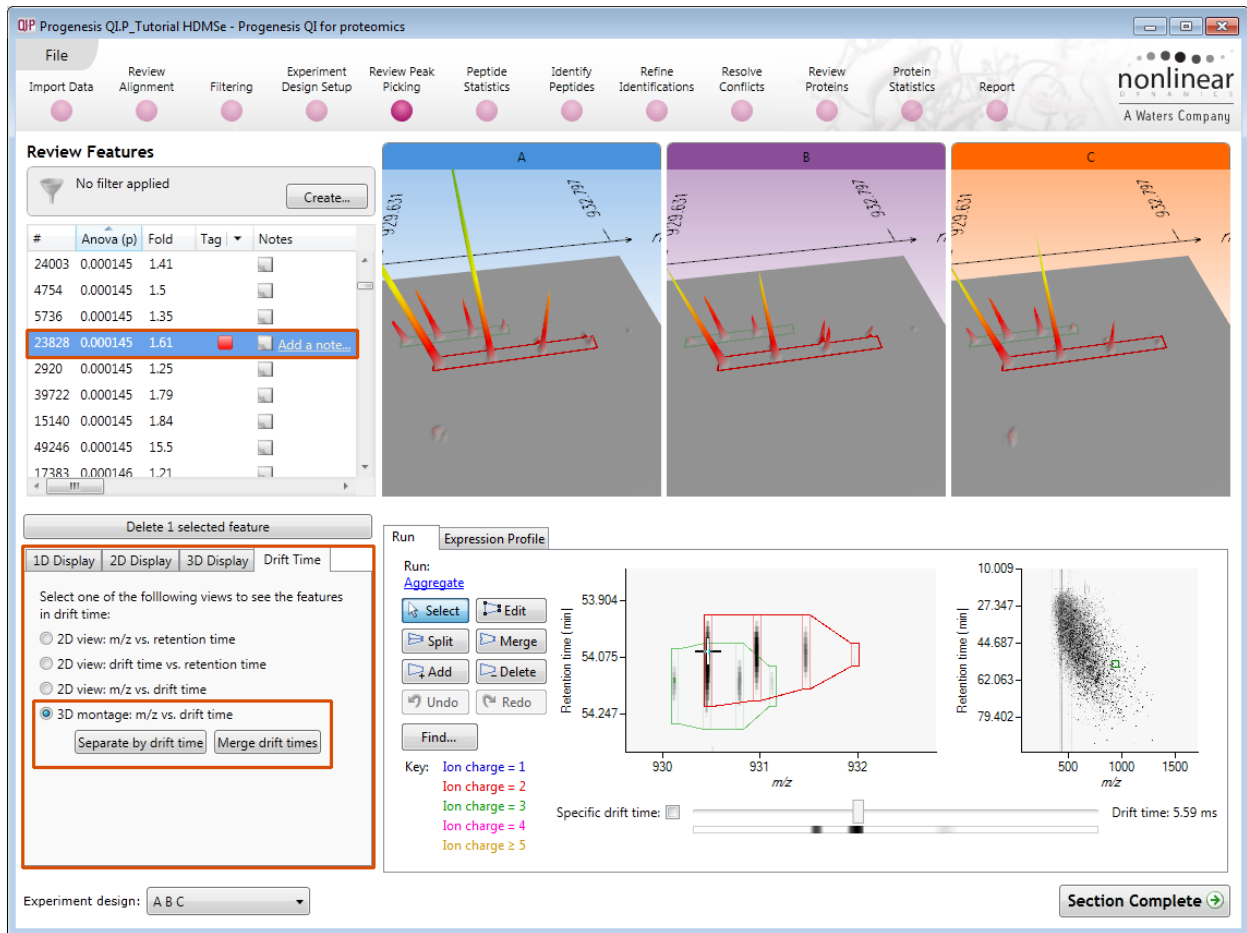
Window B changes into a 3D view by selecting the 3D Montage tab on the bottom left of the display.



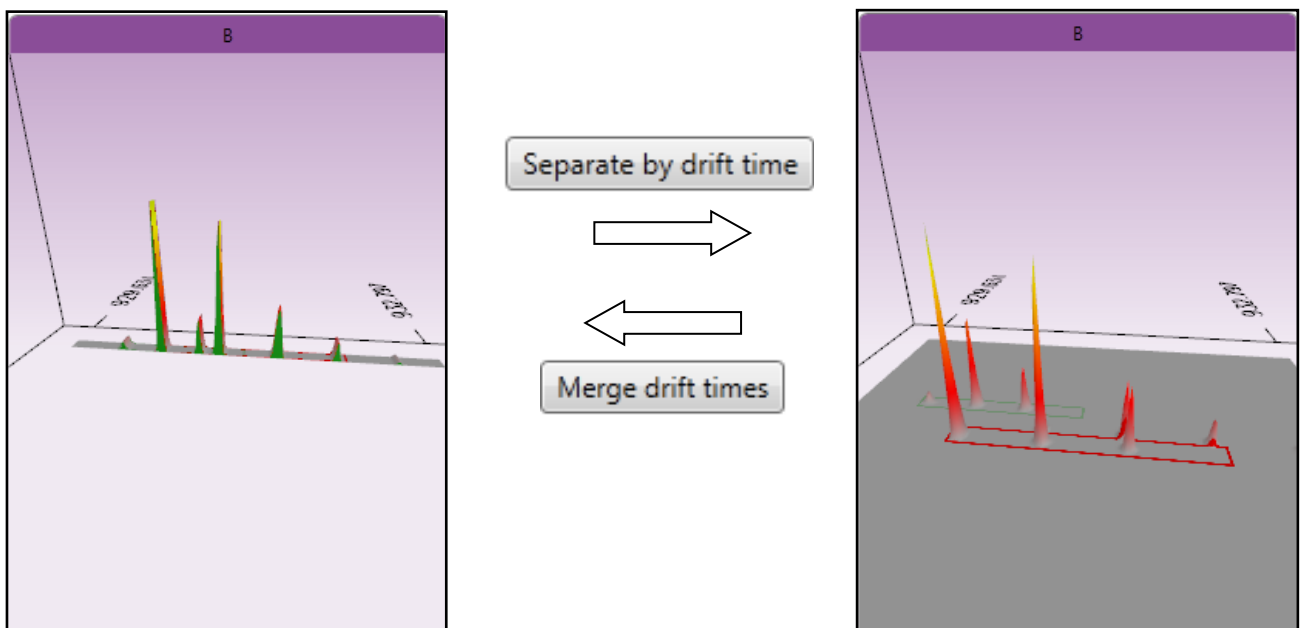
The number of 3D views displayed in the montage is controlled using the [Select runs](#) link on the 3D Montage tab. The views can be set to **Rotate** automatically or you can rotate them manually by clicking and dragging them with the mouse.

### Drift Time Display

When the Drift Time tab (F) is selected in the bottom left of the display Window B changes into a 3D view displaying the Drift Time dimension..



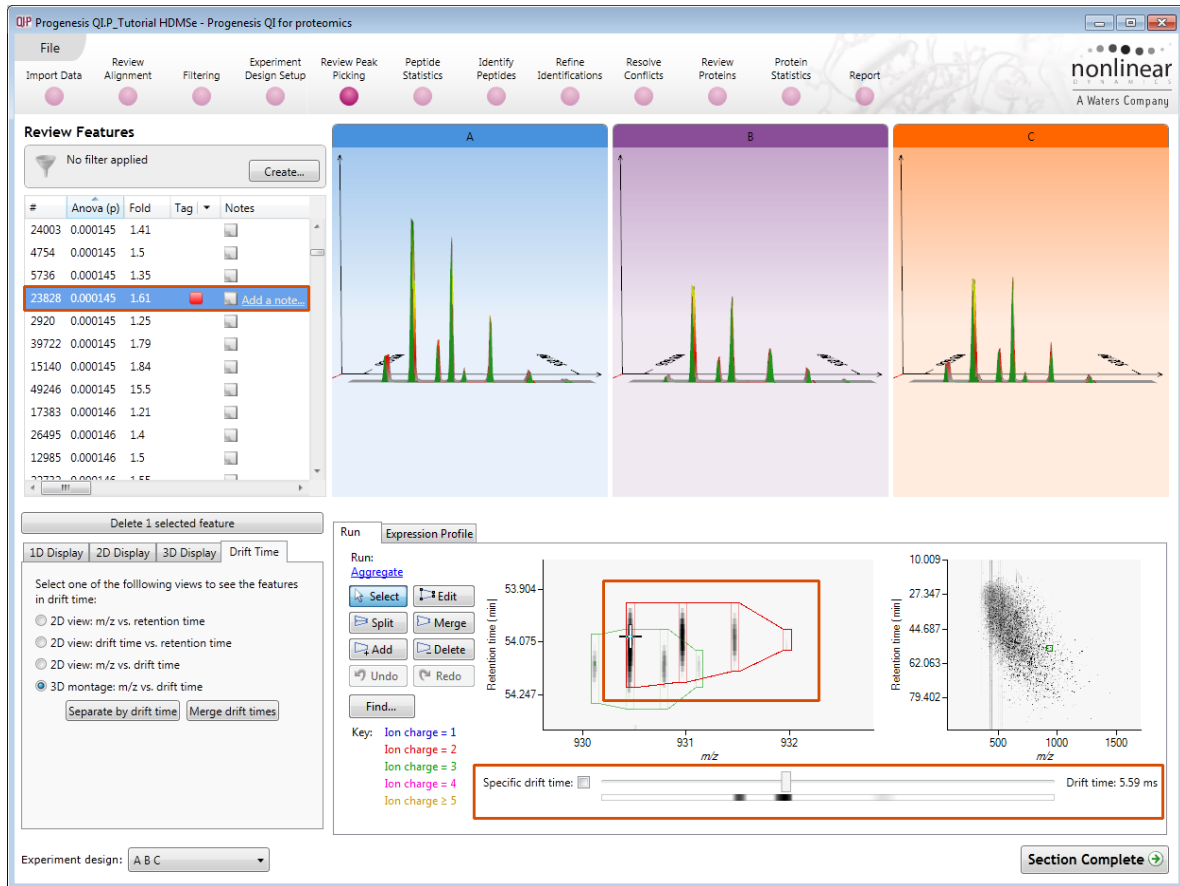
To view the drift time dimension for the current feature, click on **Separate by drift time** on the Drift Time tab. This will expand the view showing the drift time separation between the detected features (below right)



The left hand view displays the merged Drift Times.

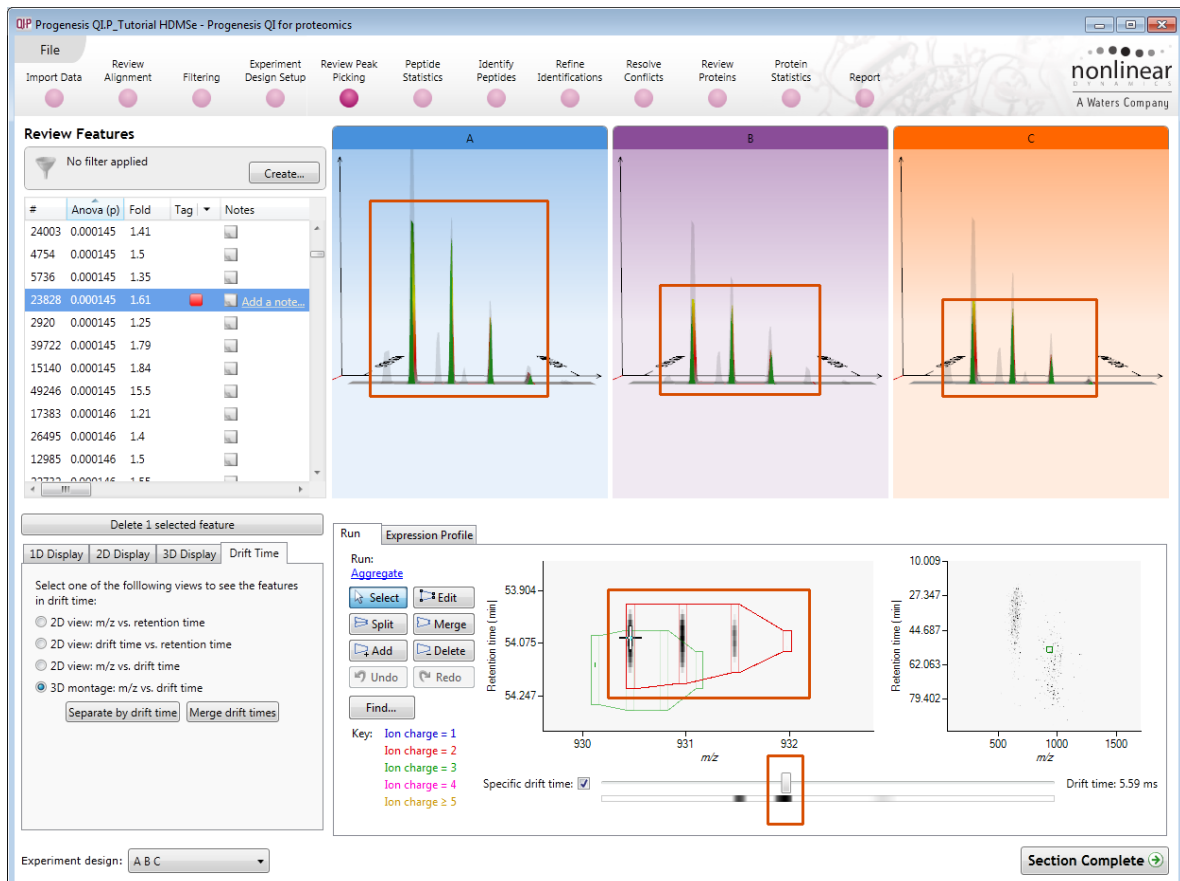
**Note:** you can manipulate the orientation of the views by clicking on them and dragging the display to the required orientation. You can also zoom in and out of the panels by using the 'scroll' wheel on your mouse

**Note:** you can step through the specific **Drift Times** (measured in milli seconds) for the current feature by clicking on the **Specific drift time** tick box at the bottom of the display.

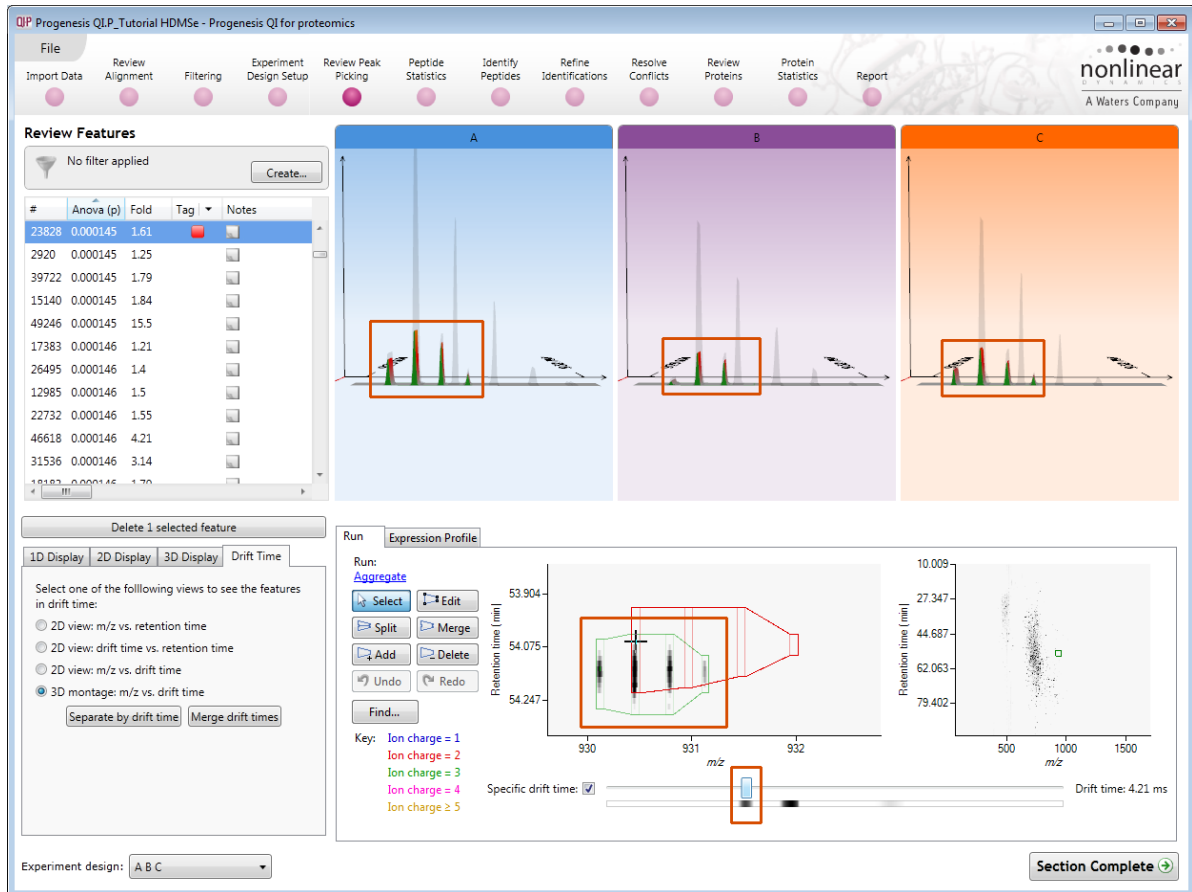


The 'crosshairs' on the feature in the **Run** view identifies it as the current feature in the table.

As you move the slider over the intense areas, indicated below, all the views update to the corresponding drift time.



**Note:** the crosshairs will remain on the original feature in the table as you explore the Specific Drift times



When you un-tick the 'Specific drift time' tick box the 3D views will return to showing the Merged Views for the current feature in the table.



Click **Separate by drift time** to view the drift times in 3D.

The screenshot displays the Progenesis QI software interface. On the left, the 'Review Features' table lists various features with columns for '#', 'Anova (p)', 'Fold', 'Tag', and 'Notes'. The 'Tag' column contains red and grey icons. Below the table, there are options for '1D Display', '2D Display', '3D Display', and 'Drift Time'. Under 'Drift Time', several view options are listed, with 'Separate by drift time' highlighted in a red box. The main workspace shows three 3D chromatograms labeled A, B, and C. Below these, there is an 'Expression Profile' section with a 'Run' dropdown set to 'Aggregate' and various manipulation buttons like 'Select', 'Edit', 'Split', 'Merge', 'Add', 'Delete', 'Undo', and 'Redo'. Two plots are visible: a retention time vs. m/z plot and a retention time vs. m/z scatter plot. A 'Specific drift time' slider is set to 4.21 ms. At the bottom right, a 'Section Complete' button is visible.

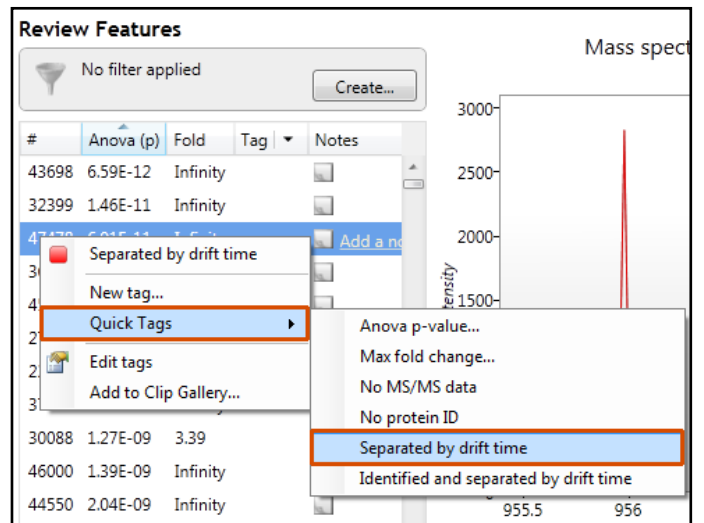
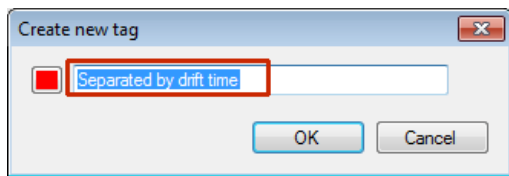
**Note:** you can use the arrow keys to review the features while maintaining the current view

## Using Quick Tags to locate examples of Drift Time

In the previous section, describing how to view Drift Time, you will have noticed the presence of a red 'Tag' in the table next to the feature that we examined. Progenesis QI for proteomics allows you to assign tags based on the properties of detected features either through the manual sorting of the table or making use of the 'Quick Tags'. These tags can be used to filter the list of displayed data in order to aid exploration of the data.

To create a Quick tag for all features demonstrating separation by Drift time, right click on the table. Select **Quick Tags** then **Separated by drift time**.

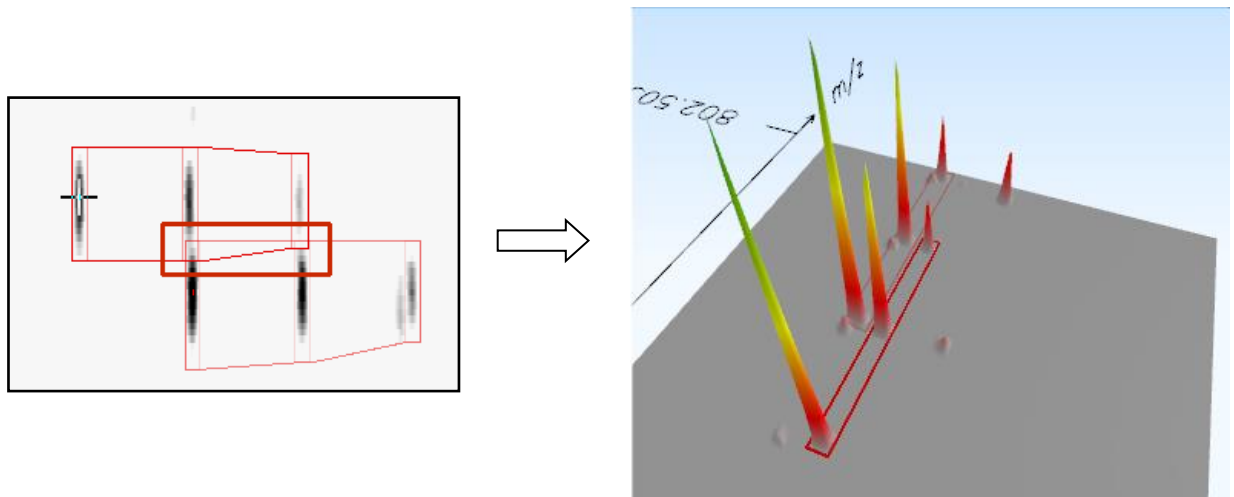
In the new tag dialog either accept or overwrite the tag name.



When the tag is created it will appear against those features that meet the criteria for the creation of the tag, in this case:

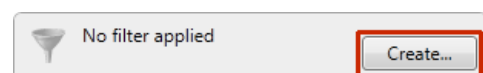
***It tags features that overlap in both m/z and retention time but do not show an overlap in the drift time dimension i.e. those features that drift time has separated***

For example the features below is overlapping at the same m/z and RT but are **separated in drift time**



Now filter the table so that it currently only displays a list of features containing the **separated by drift time** tag.

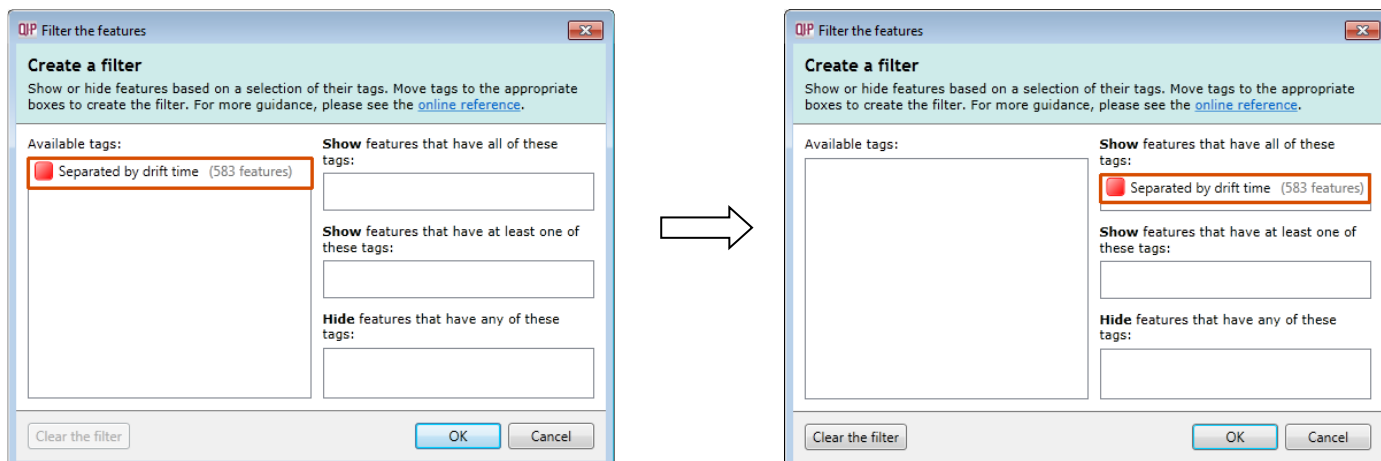
Click on **Create** on the filter panel above the table.



This will open a Tag filter dialog, in this example, displaying that you have created assigned the **Separated by drift time** tag to 583 features in your experiment.



To display only those features containing this tag drag the **Separated by drift time** tag on to the **Show** panel and click OK.



When you apply the tag filter the table will now only display the features with the appropriate tag(s).

**Review Features**

Tag filter applied  
features may be hidden

#	Anova (p)	Fold	Tag	Notes
23337	0.161	1.24	<input checked="" type="checkbox"/>	
23374	0.117	1.33	<input checked="" type="checkbox"/>	
23496	8.45E-06	1.87	<input checked="" type="checkbox"/>	
23520	8.44E-10	12.8	<input checked="" type="checkbox"/>	
23801	0.000654	2.05	<input checked="" type="checkbox"/>	
23828	0.000145	1.61	<input checked="" type="checkbox"/>	Add a note
23875	0.00467	4.26	<input checked="" type="checkbox"/>	
23953	0.0529	7.35	<input checked="" type="checkbox"/>	
24077	0.00678	1.42	<input checked="" type="checkbox"/>	
24125	0.633	1.45	<input checked="" type="checkbox"/>	
24161	0.0654	1.28	<input checked="" type="checkbox"/>	
24536	0.000000	2.08	<input checked="" type="checkbox"/>	

Delete 1 selected feature

1D Display | 2D Display | 3D Display | **Drift Time**

Select one of the following views to see the features in drift time:

- 2D view: m/z vs. retention time
- 2D view: drift time vs. retention time
- 2D view: m/z vs. drift time
- 3D montage: m/z vs. drift time

Separate by drift time | Merge drift times

Run: Aggregate

Key: Ion charge = 1 (blue), Ion charge = 2 (red), Ion charge = 3 (green), Ion charge = 4 (magenta), Ion charge ≥ 5 (yellow)

Retention time (min) vs. m/z plot showing a peak at 5.59 ms drift time.

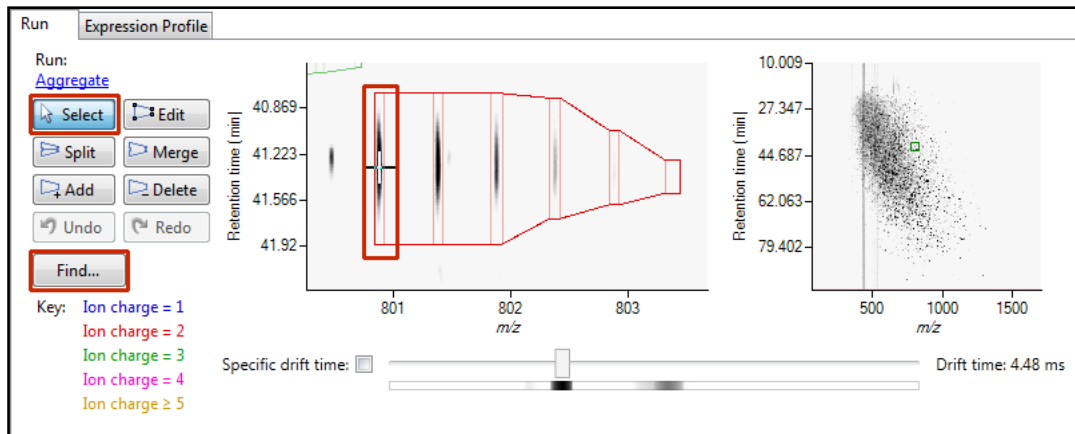
Section Complete

**Note:** with this **Tag filter applied** you can easily review the effect of Drift time separation for the features.

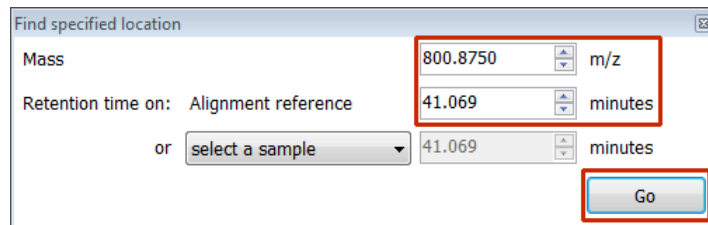
To remove the filter click on **Edit**, above the table, and **Clear the filter** followed by **OK**.

## Editing of features in the View Results stage

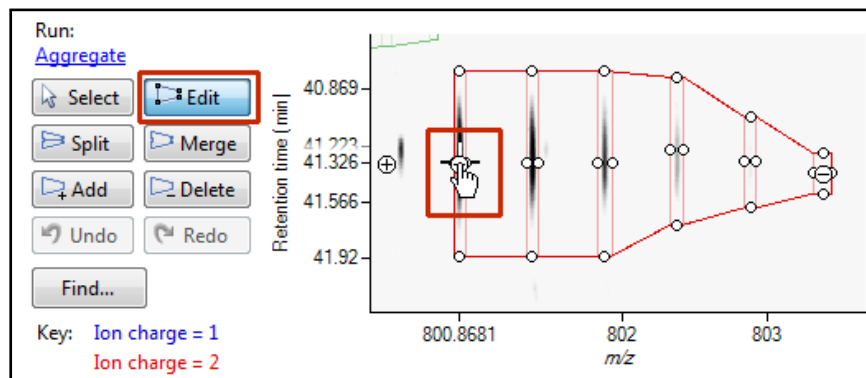
As an example of using the editing tools which are located on the left of the LC-MS Run view, we will remove and add back the 'monoisotopic peak' for the detected feature selected below. A feature can be selected from the 'Features' list or located using the various views.



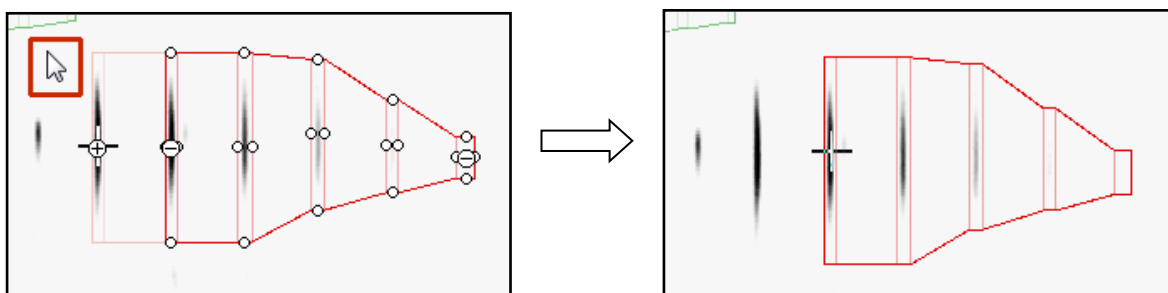
1. Locate the feature at approx 880.87 m/z and 41.07 min using the **Find** tool then click the cross hairs on the monoisotopic peak to set the zoom.



2. Select the **Edit** tool and click on the feature (in the Run view) to reveal the 'edit handles'

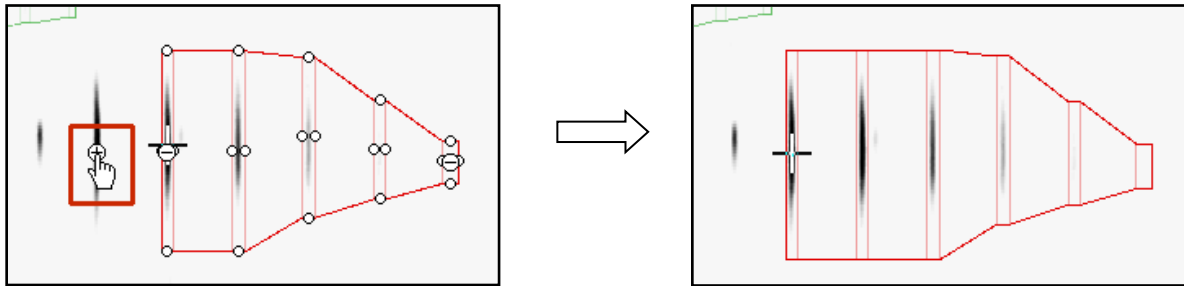


3. Click on the 'minus' handle over the monoisotopic peak to remove it.

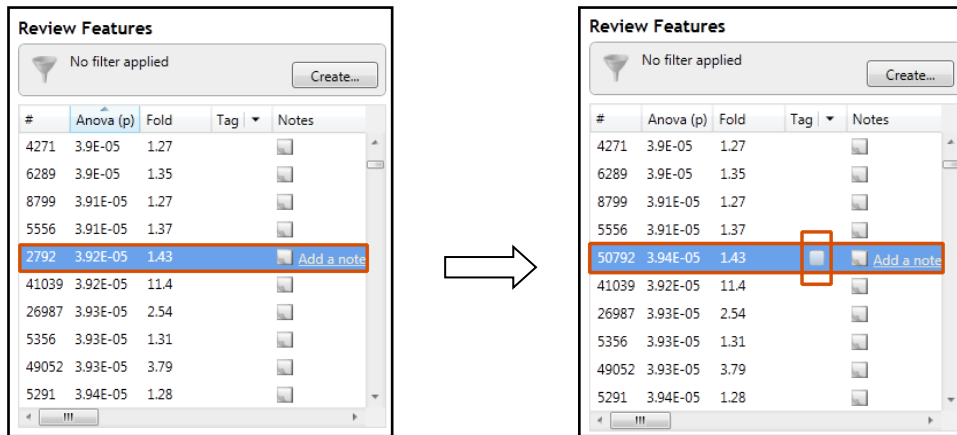


4. Click outside the boundary of the feature to update the view.

- To add a peak to an existing feature, ensure that **Edit** is selected then click inside the feature to reveal the handles.



- Click on the 'plus' handle on the peak to add it.
- Then click outside the feature to update the view.
- Note:** If you are not satisfied with the editing use the **Undo** button and retry.
- Finally note: that a tag is automatically added to the edited feature in the table and the features id. number is changed to the next available one at the end of the list.



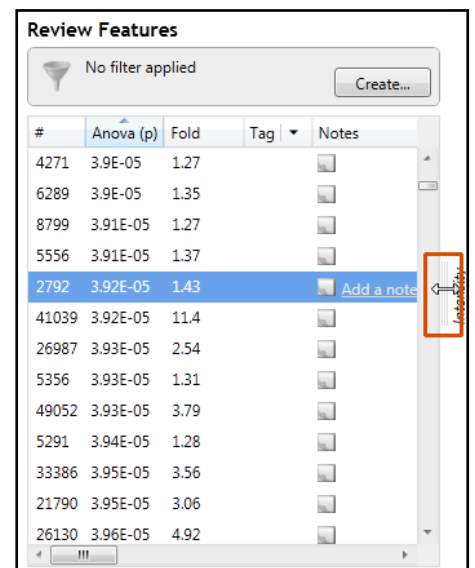
The other tools: **split, merge, add and delete** behave in a similar fashion and their use can be combined to achieve the desired results.

### Selecting and tagging features for Peptide Statistics

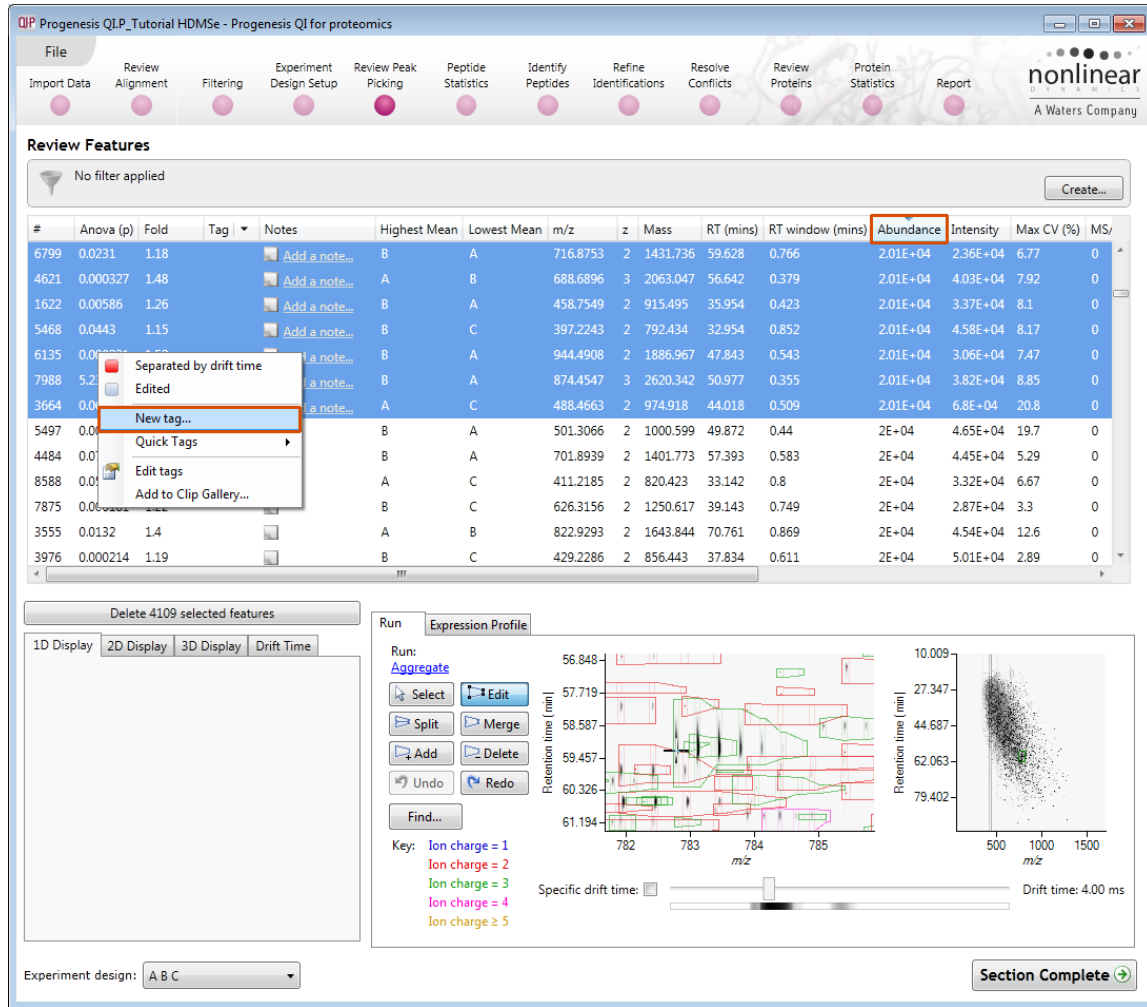
There are a number of ways to 'refine' your 'Ranked List' of analysed features before examining them with the Statistical tools in **Peptide Statistics**. These make use of simple 'Selection' and 'Tagging' tools that can be applied to the various groupings created in Stage 5 (page 21). An example is described below.

First expand the 'Features' table to show all the details by double clicking on the 'Splitter Control' to the right of the Review Features table.

Then order on Abundance and select all features with an Abundance > 2x10<sup>4</sup>.

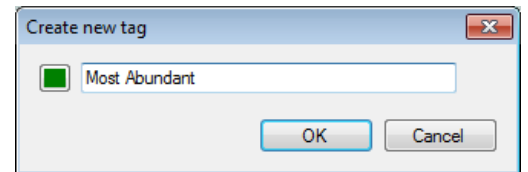


Right click on the highlighted features and select 'New Tag'.

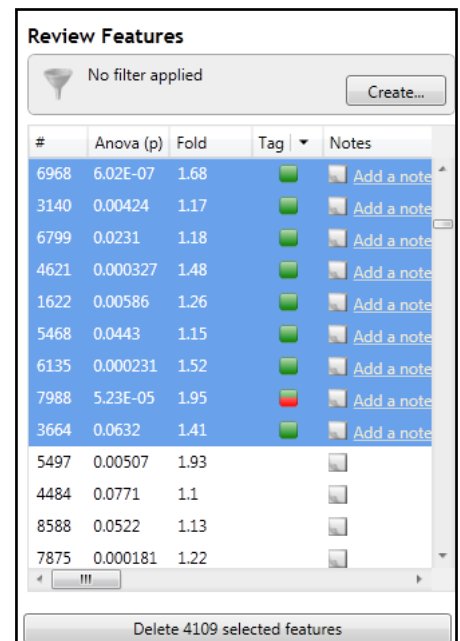


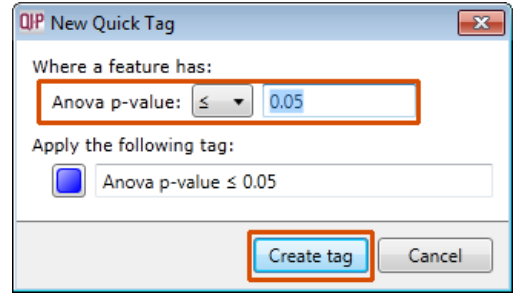
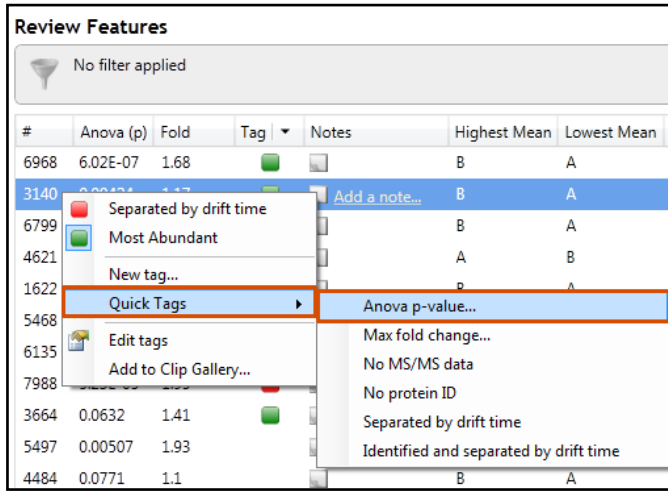
Give the Tag a name. i.e. 'Most Abundant'.

**Note:** there is already a read tag present that was assigned to those features that are **Separated by drift time**, which you created in the previous section



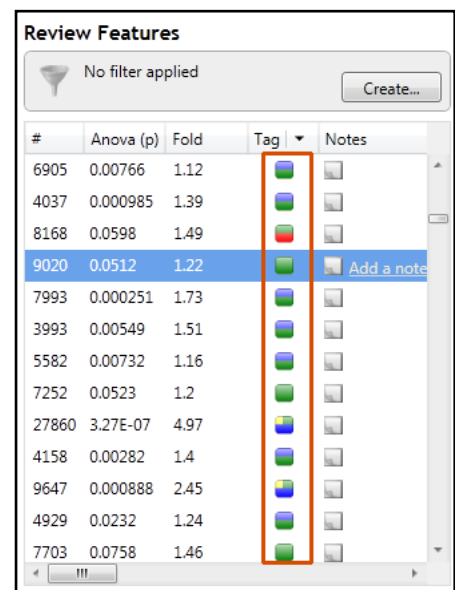
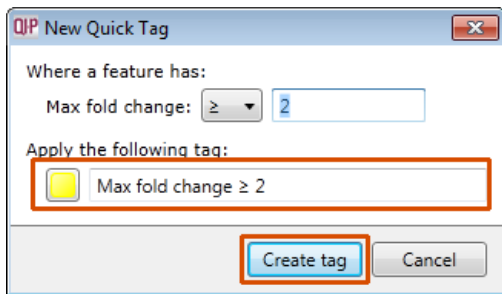
On clicking **OK** the Tag is added to the features highlighted in the table (signified by a coloured square, green in this example).





Now right click on any feature in the table and select **Quick Tags** this will offer you a number of standard tag options. Select **Anova p-value....** Then set the threshold as required and adjust default name as required and click **Create Tag**.

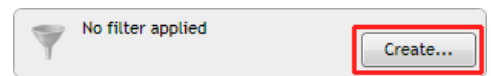
Once this tag appears against features in the table right click on the table again and create another Quick Tag, this time for features with a **Max fold change ≥ 2**



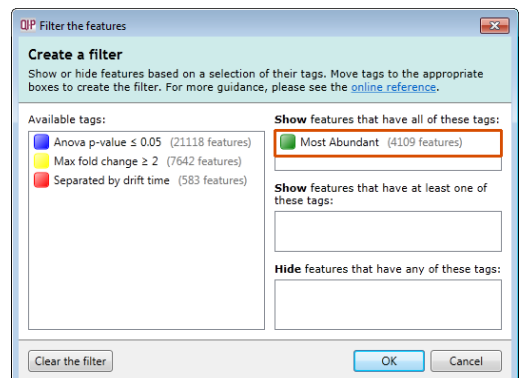
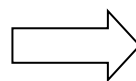
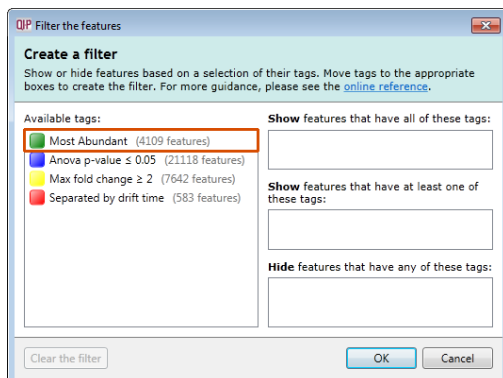
The table now displays features with multiple tags.

The tags can be used to quickly focus the table on those features that display similar properties.

For example: to focus the table on displaying those features that are **Most Abundant** click on **Create** on the filter panel above the table.



Drag the tag on to the panel **Show features that have all of these tags** and press OK.



To move to the next stage in the workflow, Peptide Statistics, click **Section Complete**.

## Stage 8: Peptide Statistics on Selected Features

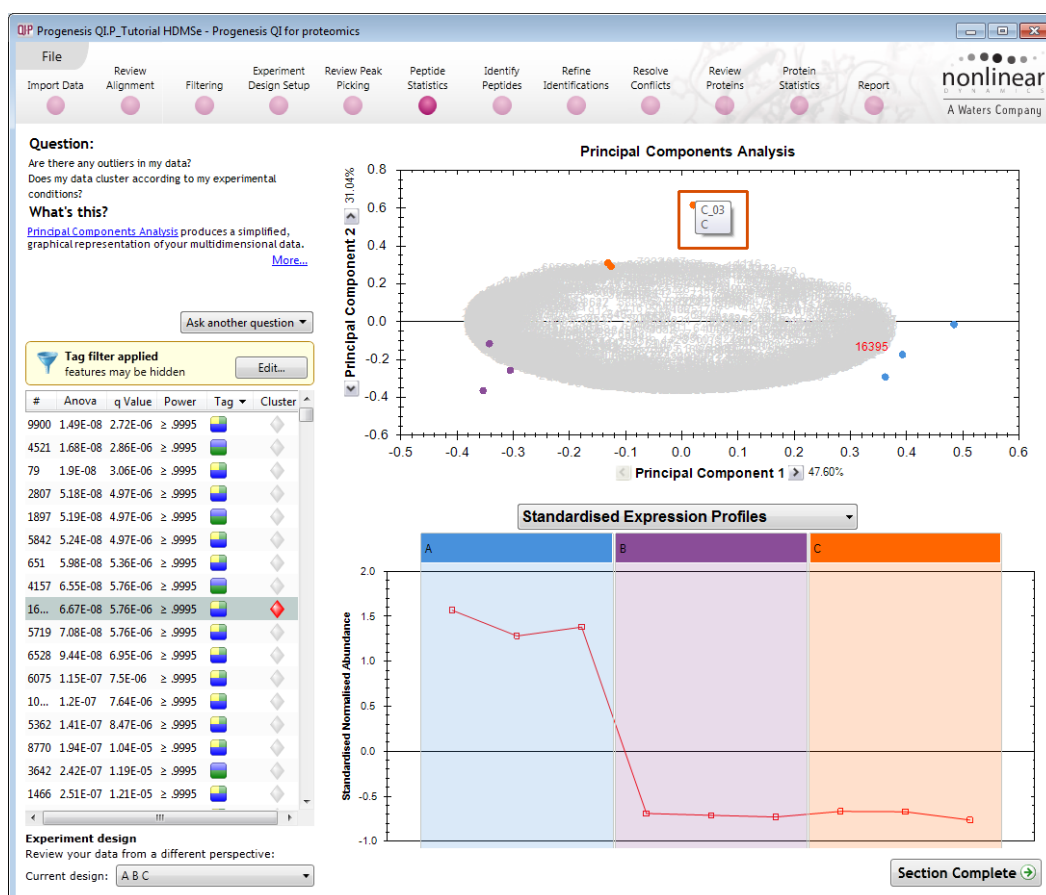
The user guide now describes the functionality of the Multivariate Statistics.

Peptide Statistics opens calculating the Principal Components Analysis (PCA) for the active 'tag' in this case the **Most Abundant** features.



As an example we will start by examining the behaviour of the **Most Abundant** features from the previous stage, **Review Peak Picking**.

The statistical analysis of the selected data is presented to you in the form of interactive graphical representation of answers to questions asked of the analysed data.



**Note:** the LC-MS runs (samples) are displayed as solid coloured circles on the plot. To identify the runs, a tooltip is displayed when the cursor is held over each circle.

### Principal Component Analysis (PCA)

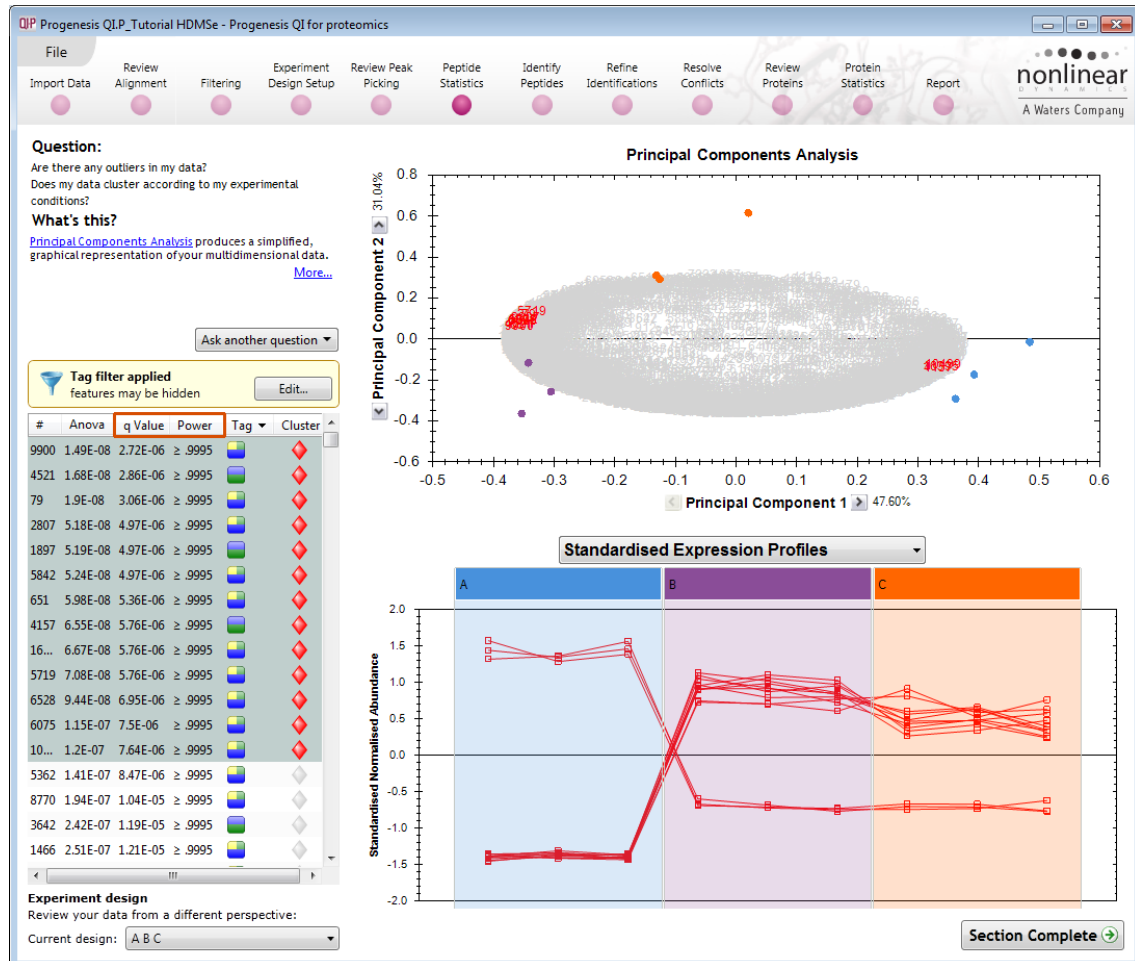
In **Peptide Stats** the first statistically based question asked of the data takes the form of a Quality Control assessment:

*Are there any outliers in my data? And does my data cluster according to my experimental conditions?*

It answers this question by:

*'Using Principal Components Analysis (PCA) to produce a simplified graphical representation of your multidimensional data'.*

PCA can be used to determine whether there are any outliers in the data and also look at how well the samples group. The groupings that can be observed on the 2D PCA plot can be compared to your experimental conditions and conclusions can be drawn regarding possible outliers in your data. Selecting features in the table will highlight the features on the 'Biplot' and their expression profiles will appear in the lower panel.



**Note:** the Table in the Stats view contains additional columns:

**q value:** tells us the expected proportion of false positives if that feature's p-value is chosen as the significance threshold.

**Power:** can be defined as the probability of finding a real difference if it exists. 80% or 0.8 is considered an acceptable value for power. The Power Analysis is performed independently for each feature, using the expression variance, sample size and difference between the means.

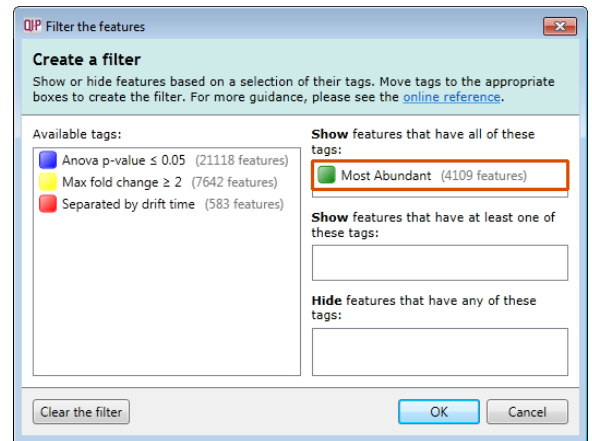
Also, for a given power of 80% we can determine how many samples are required to ensure we find a difference if it actually exists.

**Note:** Power analysis is discussed in **Appendix 4** (page 73)

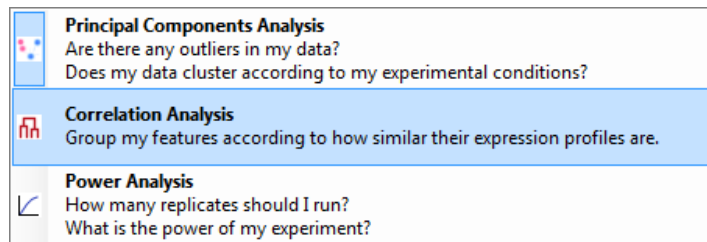
## Correlation Analysis

With the tag filter still set to display only the top 4109 **Most Abundant features**, we are going to explore the Correlation Analysis of these features.

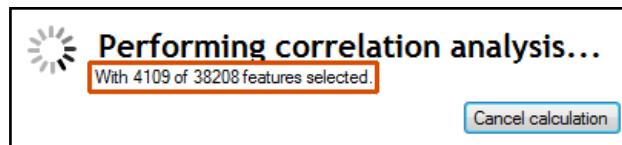
To set up the **Correlation Analysis** using this filtered data set click on **Ask another question** (above the table)



A selection of 3 tools will appear in the form of questions.



Select the second option to explore 'feature correlation based on similarity of expression profiles'



This time the statistically based question(s) being asked is:

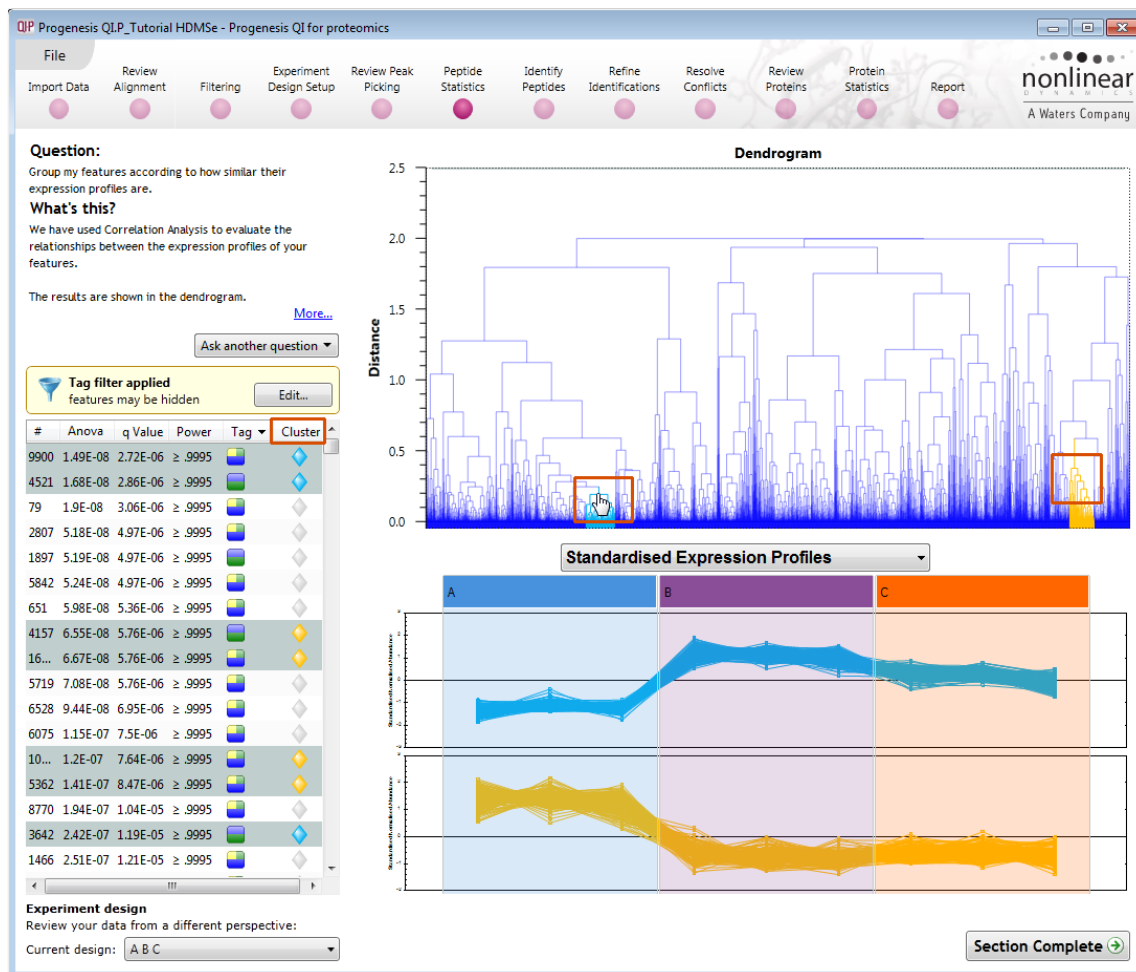
*'Group my (selected) features according to how similar their expression profiles are'*

The question is answered by:

*'Using Correlation analysis to evaluate the relationships between the (selected) features' expression profiles'*

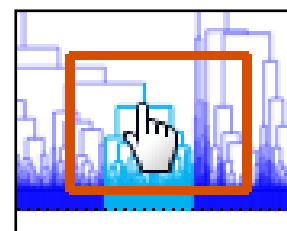


The answer is displayed graphically in the form of an interactive dendrogram where the vertical distance, between each feature can be taken as indicative of how similar the expression profiles of each cluster of features are to each other.



**Correlation Analysis** enables the grouping of features together according to how similar their expression profiles are.

**For example:** To highlight groups of features demonstrating **different expression profiles** click on a 'node' for a branch of the Dendrogram (as shown above) while holding the mouse button, hold down the **Ctrl** key and then click on another node as shown.



If you have selected 2 nodes then there will be two expression profile graphs

**Note:** by highlighting a group of features with similar expression the features are identified with the same colour of cluster flag in the table. This allows the table to be sorted on cluster and tagged accordingly

**Tip:** when reviewing the tags (see above) if you are not applying a new filter then use the **Cancel** button to return to the main view, this prevents unnecessary recalculation of your data.

Before moving to the Identify Peptides stage in the Workflow, first return to the PCA display and clear all tag filters **Clear all Tag filters**.

To move to the next stage in the workflow, Peptide Search, click **Section Complete**.

## Stage 9: Identify peptides

Progenesis QI for proteomics is designed to perform peptide identifications either directly or by allowing you to export MS/MS peak lists in formats which can be used to perform peptide searches by various search engines. The resulting identifications can then be imported back into Progenesis QI for proteomics, using a number of different file types, and matched to your detected features.

The screenshot shows the Progenesis QI software interface. The 'Identify Peptides' step is active. The 'MS<sup>E</sup> Search' method is selected. A table of detected features is displayed, with the first feature highlighted. A 'No filter applied' message is visible in the top right. The 'Identifications for feature 1' section is empty, and a message states 'No identification selected'.

#	Identifications	m/z	Charge	Retention time	Drift time	Tag
1	0	828.4190	2	58.67	4.90	
3	0	822.4427	2	69.28	4.76	
4	0	962.7953	3	76.54	5.04	
5	0	894.4742	3	67.50	4.48	
6	0	964.1703	3	63.94	5.45	
7	0	763.7623	3	69.69	4.07	
8	0	823.4387	2	75.96	5.04	
9	0	913.5254	2	71.78	6.35	
10	0	820.1444	2	74.48	4.55	

For this example we are using the direct method **MS<sup>E</sup> Search**.

The Identify Peptides page currently displays the full list of the detected features in your experiment and some of their attributes, including the number of **Identifications**. If search results exist these can be cleared by clicking **Clear all identifications**.

## Entering Search Parameters

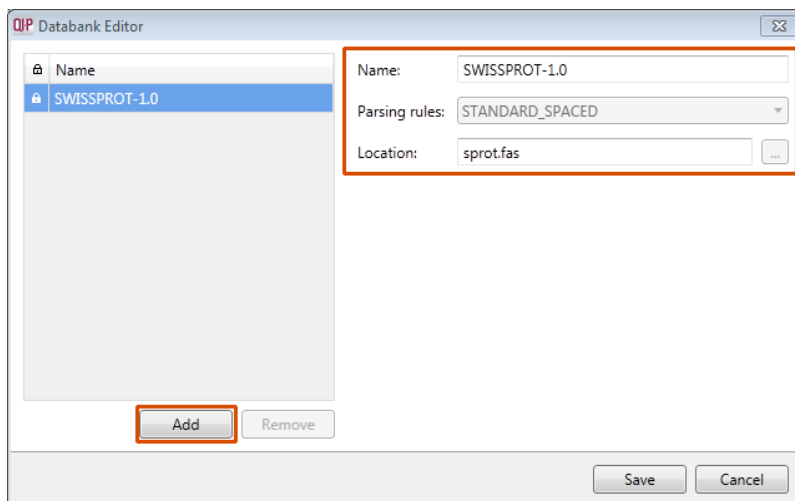
Firstly you need to select the FASTA file containing peptide and protein identifications.

SWISSPROT-1 is provided with the installation of the software.

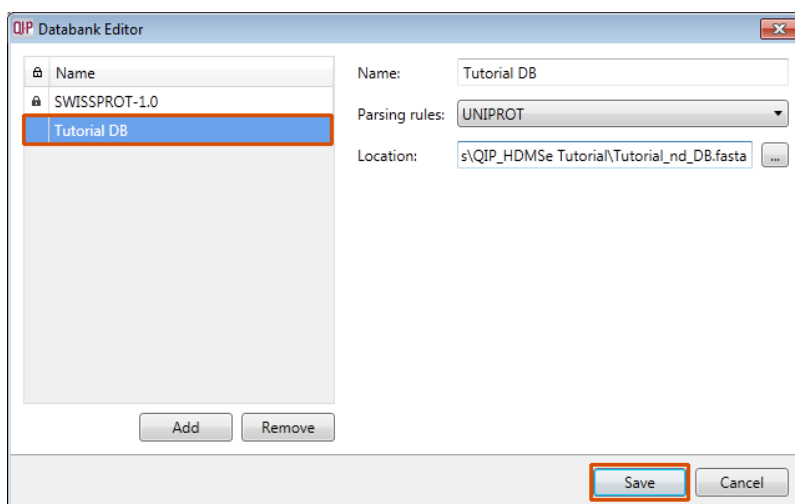
To add new Databanks in the form of FASTA files click on **Edit...** to open the Databank editor

**Note:** the SWISSPROT-1.0 is locked

The screenshot shows the 'Identify Peptides' step in Progenesis QI. The 'Enter the search parameters' section is highlighted. The 'SWISSPROT-1.0 STANDARD\_SPACED' database is selected, and the 'Edit...' button is highlighted.



For a new Databank you need to give it name, select the parsing rules and specify the location of the FASTA file, see the example below.



The new Data bank will appear in the left panel now click **Save** to return to the Search parameters.

If your databank is not already displayed then select it from the drop down list.

**Expand the Common search parameters**

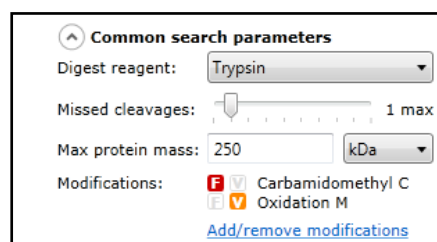
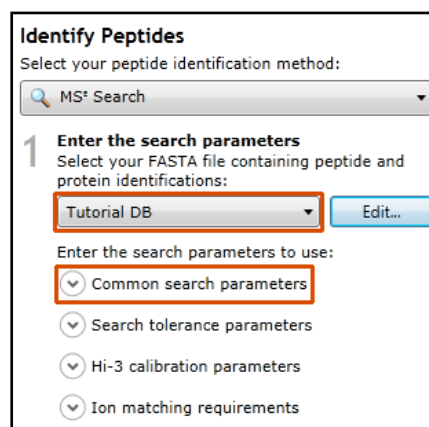
The default settings are displayed:

**Digest reagent:** is set as Trypsin. Alternative Digest reagents are available from the list and additional ones can be added to the list using the **Reagent editor...**

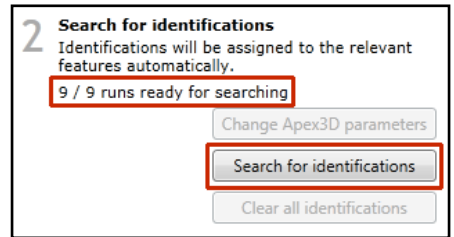
**Missed cleavages:** is set as 1.

**Maximum protein mass:** is set at 250kDa

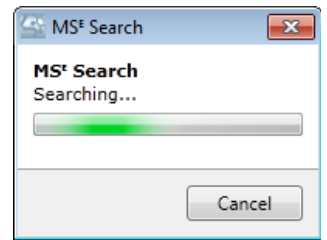
**Modifications:** are set Carbamidomethyl C (Fixed) and Oxidation M (variable). More modifications are available from the list and additional ones can be added to the list using the **Modification editor...**



Having selected the Databank and set the parameters, before Searching for identifications make sure that all of the runs are available for searching.



Depending on the search parameters the MS<sup>E</sup> Search can take some time



Once the Ion Accounting is complete, features with identifications are identified with a solid grey symbol and the number of identifications appears in the next column.

**Identify Peptides**  
Select your peptide identification method:  
MS<sup>E</sup> Search

**1 Enter the search parameters**  
Select your FASTA file containing peptide and protein identifications:  
Tutorial DB [Edit...]  
Enter the search parameters to use:  
**Common search parameters**  
Digest reagent: Trypsin  
Missed cleavages: 1 max  
Max protein mass: 250 kDa  
Modifications: Carbamidomethyl C, Oxidation M  
Add/remove modifications

**2 Search for identifications**  
Identifications will be assigned to the relevant features automatically.  
9 / 9 runs ready for searching  
Change Apex3D parameters  
Search for identifications  
Clear all identifications

**Features (18622 identified)**

#	Identifications	m/z	Charge	Retention time	Drift time	Tag
1	1	828.4190	2	58.67	4.90	
3	1	822.4427	2	69.28	4.76	
4	1	962.7953	3	76.54	5.04	
5	1	894.4742	3	67.50	4.48	
6	1	964.1703	3	63.94	5.45	
7	2	763.7623	3	69.69	4.07	
8	5	823.4387	2	75.96	5.04	
9	0	913.5254	2	71.78	6.35	

**Identifications for feature 1**

Peak Mass	Peptide Mass	Mass Error (Da)	Mass Error (ppm)	Score	Start	End	Sequence
1654.823	1654.809	0.0145	8.7362	9.663	52	66	DLSEASVYAEALPK

**Fragment ions for: DLSEASVYAEALPK**

Intensity (counts) vs m/z chromatogram showing peaks at various m/z values.

Section Complete

Details for the current features identifications are displayed in the table below and the Fragment ions for the current identification are displayed in the bottom panel.

**Note:** if you want to perform the search with a new set of parameters then first select **Clear all identifications**

In order to review, and refine the quality of the **Peptide Search** results click on the next stage in the workflow, **Refine Identifications**.

## Stage 10: Refine Identifications

In this example we are going to apply a number of filters to 'refine' the quality of the Databank search.

As an **example** 'Acceptance Criteria' on which to base the sequential filtering of the Peptide results, the following thresholds will be applied:

- Remove identifications with a Score less than 4
- Remove identifications where less than 2 hits were returned
- Remove all identifications where the Protein Description **Contains** the following: 'Putative', 'Probable', 'Like', 'Potential' and 'Predicted'

To perform these filters, on the Batch detection options panel, set the Score to less than 4, then **Delete matching search results**.

**Refine Identifications**

If your peptide identifications include unwanted or irrelevant results, you can remove them here.

- Specify a set of deletion criteria**  
In the batch deletion criteria, enter the property values for a set of identifications you want to delete.  
Any identifications matching the criteria will be highlighted in pink.
- Delete the unwanted identifications**  
To delete the identifications you don't want, click either:
  - Delete Matching Search Results, to delete the highlighted IDs
  - Delete Non-matching Search Results, to delete the IDs that are not highlighted
- Reset the criteria to start again**  
To specify another batch of identifications to delete, click Reset the Criteria and then return to step 1 above.

Batch deletion criteria

Score: less than 4

Hits: less than

Mass: less than

Mass error (ppm): less than

M/Z: less than

Retention Time: less than

Sequence Length: less than

Charge: less than

Sequence: contains

Accession: contains

Description: contains

Modifications: contains

Delete matching search results  Delete non-matching search results

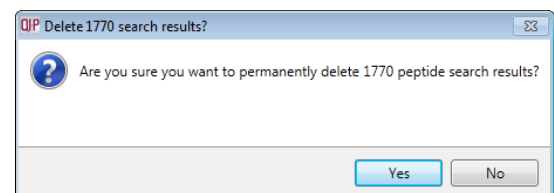
#	Score	Hits	m/z	RT(mins)	Charge	Mass	Mass err.	Sequence	Accession	Modifications
8296	8.37	8	687.38	49.73	2	1372.74	4.54	WQSLVEANV	P40910	40S i
8298	8.16	9	570.29	42.59	2	1138.51	5.17	SVTDISFDAG	Q59KZ1	Amir
8299	3.86	1	559.79	49.99	2	1117.51	29.68	DPMGLSEIEK	Q5A2N5	Puta
8299	6.80	9	559.79	49.99	2	1117.51	-9.96	SVWGDVTAC	C4YSW5	Puta
8299	0.00	1	559.79	49.99	2	1117.51	9.00	GFDHPFASLK	Q5AIA2	Hom
8299	6.80	9	559.79	49.99	2	1117.51	-9.96	SVWGDVTAC	Q5A5F5	Puta
8302	0.00	5	449.29	60.50	2	896.56	18.89	GLISSILGR	Q5AG68	[C-term] neutral loss Nucl
8302	0.00	7	449.29	60.50	2	896.56	-9.22	PIQLVVTK	P46598	Heat
8302	4.64	1	449.29	60.50	2	896.56	-9.22	QLLPLVSK	Q5A5R9	Likel
8303	7.92	4	391.75	36.45	2	781.49	7.95	LNAPLVR	C4YE49	Puta
8304	4.93	2	689.72	52.04	3	2066.11	22.03	DSVNAVIVYK	C4Y197	Nucl
8304	5.16	2	689.72	52.04	3	2066.11	-1.64	HPITLVRPRG	Q5APD2	Mal
8304	4.56	2	689.72	52.04	3	2066.11	11.19	THAIVLSSEI	C4YS35	Puta
8304	4.85	1	689.72	52.04	3	2066.11	69.42	FGSFEEISPSE	O94201	6-ph
8304	4.23	1	689.72	52.04	3	2066.11	4.41	LKHPQLLQVY	C4YEG0	Puta
8305	8.23	9	1086.91	72.69	3	3257.61	4.14	NGQLVEIPAN	P28877	Plasr
8306	4.71	1	593.63	62.80	3	1777.81	-1.29	LNSLTKDFDN	Q5A044	Pote
8306	4.71	1	593.63	62.80	3	1777.81	-1.29	LNSLTKDFDN	C4YLH0	Puta
8307	7.21	7	920.81	64.02	3	2759.41	3.36	FEKPVNFDDI	Q5AHH4	Puta
8307	5.69	3	920.81	64.02	3	2759.41	-1.94	VNNVEIVAFV	Q59KD6	Chor
8308	7.85	9	896.55	70.23	2	1795.01	2.07	EVLFPFLKPE	Q5ADU2	Lysyl
8309	4.95	1	738.88	43.37	2	1475.74	5.33	KNPESYYTIF	C4YNM6	Puta

28887 search results. 1770 matching batch delete options.

Section Complete

**Note:** the search results matching the filter criteria turn pink and the total is displayed at the bottom of the table (1770 matching out of 28887)

**Note:** a dialog warns you of what you are about to delete



Now **Clear all filters** and then apply the next filter (Hits: less than 2)

**Refine Identifications**

If your peptide identifications include unwanted or irrelevant results, you can remove them here.

- Specify a set of deletion criteria**  
In the batch deletion criteria, enter the property values for a set of identifications you want to delete.  
Any identifications matching the criteria will be highlighted in pink.
- Delete the unwanted identifications**  
To delete the identifications you don't want, click either:
  - Delete Matching Search Results, to delete the highlighted IDs
  - Delete Non-matching Search Results, to delete the IDs that are not highlighted
- Reset the criteria to start again**  
To specify another batch of identifications to delete, click Reset the Criteria and then return to step 1 above.

Batch deletion criteria

Score: less than  
 Hits: less than 2  
 Mass: less than  
 Mass error (ppm): less than  
 M/Z: less than  
 Retention Time: less than

Sequence Length: less than  
 Charge: less than  
 Sequence: contains  
 Accession: contains  
 Description: contains  
 Modifications: contains

Delete matching search results  Delete non-matching search results

QIP Delete 14289 search results?  
 Are you sure you want to permanently delete 14289 peptide search results?

#	Score	Hits	m/z	RT(min)	Charge	Mass	Mass err	Sequence	Accession	Modifications
27550	4.70	1	616.83	51.71	2	1231.6	5.05	LGEDSFGGLK	Q5A5R9	Like
27550	4.46	1	616.83	51.71	2	1231.6	10.33	LHCLNTLFSK	C4YDW1	[3] Carbamidomethyl C
27551	4.97	1	672.33	41.69	2	1342.6	-64.97	MSLNILVIGN	C4YDN4	Phos
27551	4.97	1	672.33	41.69	2	1342.6	-64.97	MSLNILVIGN	Q5A7H5	Put
27551	5.87	1	672.33	41.69	2	1342.6	8.46	LDVYFNEATS	P10875	Tub.
27553	4.85	2	552.24	55.67	3	1653.7	17.53	SWDDEDFDII	Q5ACM9	Euka
27559	5.77	1	765.89	62.80	2	1529.7	15.75	TFTTSTINFAQ	C4YR87	Fum.
27560	6.08	5	449.74	53.39	2	897.46	-3.80	GSFFAELK	C4YTC6	DNA
27564	4.03	1	870.42	67.16	2	1738.8	-1.70	MNNSMLSSI	C4YL16	Put
27564	6.90	1	870.42	67.16	2	1738.8	-6.65	DWFDIKAPTI	P40910	40S i
27564	6.90	1	870.42	67.16	2	1738.8	-6.65	DWFDIKAPTI	Q5A500	40S i
27567	5.05	2	572.82	45.86	2	1143.6	-2.96	EIQSAQISR	Q59N33	Put
27567	4.43	1	572.82	45.86	2	1143.6	-9.43	AMLVVVEQA	C4YK8	Put
27570	5.43	1	589.82	43.35	2	1177.6	-1.94	LSIGQEYK	P84149	mRN

27117 search results. 14289 matching batch delete options.

Section Complete

Now in the Description first enter 'Like' and delete matching search results. Then enter the 'regular expression': **regex:Put|Prob|Pote|Pred** and delete matching search results.

**Refine Identifications**

If your peptide identifications include unwanted or irrelevant results, you can remove them here.

- Specify a set of deletion criteria**  
In the batch deletion criteria, enter the property values for a set of identifications you want to delete.  
Any identifications matching the criteria will be highlighted in pink.
- Delete the unwanted identifications**  
To delete the identifications you don't want, click either:
  - Delete Matching Search Results, to delete the highlighted IDs
  - Delete Non-matching Search Results, to delete the IDs that are not highlighted
- Reset the criteria to start again**  
To specify another batch of identifications to delete, click Reset the Criteria and then return to step 1 above.

Batch deletion criteria

Score: less than  
 Hits: less than  
 Mass: less than  
 Mass error (ppm): less than  
 M/Z: less than  
 Retention Time: less than

Sequence Length: less than  
 Charge: less than  
 Sequence: contains  
 Accession: contains  
 Description: contains regex:Put|Prob|Pote|Pred  
 Modifications: contains

Delete matching search results  Delete non-matching search results

QIP Delete 4280 search results?  
 Are you sure you want to permanently delete 4280 peptide search results?

#	Score	Hits	m/z	RT(min)	Charge	Mass	Mass err	Sequence	Accession	Modifications
10	9.57	9	920.14	74.48	3	2757.4	5.54	SIVTLV/KPW	Q5A652	Put
11	6.02	2	657.37	36.18	2	1312.7	63.24	AAREEAANI	C4YLD5	Asp
11	9.90	8	657.37	36.18	2	1312.7	16.31	AVGGEVGAS	Q5AJF7	60S i
13	9.18	9	828.45	45.13	2	1654.8	16.54	SINPNYTPVP	Q5A1E2	Pyr
16	7.79	4	761.47	60.92	2	1520.9	17.02	AVANVNDIA	P30575	Enol
17	9.31	9	845.77	70.80	3	2534.2	10.07	SINPDEAVY	P46587	Heat
18	9.39	9	740.48	79.13	2	1478.9	15.49	ASLVPGLVLI	Q9R834	60S i
22	9.91	9	776.41	51.90	2	1550.8	17.03	TLAETAQEVV	C4YR96	Put
22	9.91	9	776.41	51.90	2	1550.8	17.03	TLAETAQEVV	Q59Y00	Put
23	9.09	9	672.34	55.08	3	2014.0	9.67	AGLKPVEPEY	O13430	Elon
25	4.78	2	782.78	59.25	3	2345.3	26.59	TVHFQKPSVE	C4YJ04	Put
25	9.40	6	782.78	59.25	3	2345.3	12.82	TLLAIDAIEPI	P16017	Elon
25	4.96	2	782.78	59.25	3	2345.3	6.30	IAIQYSTYGH	C4YL21	Prot
26	8.47	9	983.15	61.47	3	2946.4	5.33	ASEVASFEQV	Q59931	60S i

12261 search results. 4280 matching batch delete options.

Section Complete

Having applied all the filters there will be **7981** search results remaining

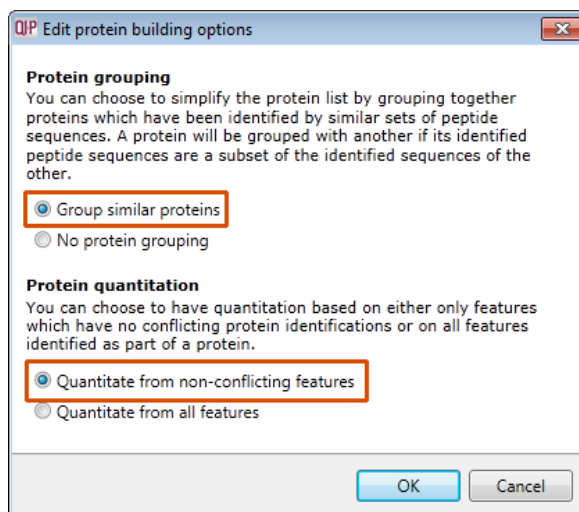
To validate the Peptide search results at the protein level select the next stage in the workflow by clicking on **Review Proteins**.

## Stage 11: Resolve Conflicts

This stage allows you to examine the behaviour of the identified peptides and choose to resolve any conflicts for the various peptide assignments at the protein level.

The number of conflicts you have to resolve will depend on the scope and stringency of the filters you apply at the **Refine Identifications** stage.

**Note:** the default **Protein options** for protein grouping and Protein quantitation are set as shown



This means that if you choose **not** to resolve the conflicts then a protein's quantitation will be based on the non-conflicting or unique features (peptides).

If you choose **not** to resolve conflicts then click **Section Complete** to **Review Proteins**.

To proceed with resolving conflicts there are some simple rules that you can apply using this stage in the workflow

With **Group similar proteins** selected the additional members are indicated by a bracketed number located after the Accession number. As an example, when the cursor is held over the accession number the group members appear in a tool tip.

The bracketed number in the **Peptides** column indicates the number of unique peptides used for quantitation.

QIP Progenesis QI.P\_Tutorial HDMSe - Progenesis QI for proteomics

File Import Data Review Alignment Filtering Experiment Design Setup Review Peak Picking Peptide Statistics Identify Peptides Refine Identifications Resolve Conflicts Review Proteins Protein Statistics Report

Experiment design: ABC

Proteins No filter applied Create...

Peptides of C4YQN7 No filter applied Create...

Accession	Peptides	Conflict	Score	Tag	#	Score	Hits	Mass	Mass error (p...)	RT (mins)	Charge	Tag	Abundance	Conflict	Peptide
P34731	75 (5)	90	696		111	6.98	6	1661.866	-0.234	52.1	2		1.06E+05	2	LI I
C4YMC3	73 (72)	1	566		39195	7.21	2	2154.007	3.05	67.3	2		1.92E+03	2	SW I
C4YQ7	70 (1)	89	644		2766	8.15	8	1414.679	4.09	46.2	2		3.77E+04	1	YE I
P43098	55 (4)	62	471		2503	8.23	9	1145.602	7.82	48.7	2		4.55E+04	1	TD I
C4YQN7 (+1)	54 (3)	62	472		2906	7.78	9	1277.678	9.65	57.4	2		5.36E+04	1	WI I
C4YR46								2685.265	6.95	56	4		6.9E+04	1	YV I
Q96VB9								2685.25	1.31	56	3		3.47E+04	1	YV I
Q13430	42 (38)	5	449		3247	7.94	9	866.468	16.9	55.7	2		2.59E+04	1	FD I

Protein: C4YQN7 Fatty acid synthase alpha subunit reductase GN=CAWG\_02796 PE=4 SV=1

Peptides of selected protein

Peptide Views Protein Resolution

Standardised Expression Profiles

Retention time (min)

Section Complete

The **Resolve Conflicts** stage provides a number of interrelated graphical and tabular views to assist you in the validation of the peptides that have been assigned to proteins and also to review the relevance of the data returned from the search.

Open **Resolve Conflicts** and order the data in the Proteins table (A) on the basis of **Conflicts**.

**Note: the look of the tables (with regards to ordering) in the following section may vary slightly.**

The screenshot displays the Progenesis QI software interface. At the top, the 'Resolve Conflicts' menu item is highlighted. Below the menu, the 'Proteins' table (A) is shown, sorted by 'Conflict' count. The first protein, P34731, is selected. To its right, the 'Peptides of P34731' table (B) is displayed, showing details for each peptide. Below these tables, the 'Protein Resolution' section (C) is active, showing a 'Standardised Expression Profiles' plot for the selected protein. The plot shows normalized abundance across three conditions (A, B, C). To the right of the plot, a 2D plot (D) shows retention time (min) versus m/z for the selected peptide, with a green box highlighting a specific peak.

Select the first protein in list A (in this case it has 82 conflicts) the panel to the right (B) lists the peptides for this protein and the conflicting protein for each peptide.

Panel C shows the expression profile(s) for the peptide(s) selected in list B

Panel D shows the details for the selected peptide.

Now click on the **Protein Resolution** in Panel C to display the proteins that are conflicting.

The screenshot displays the 'Protein Resolution' view in Progenesis QI. Panel E shows a table of 'Conflicting proteins for feature 6734', listing proteins like P34731, C4YQR7, and Q5A6R2 with their respective peptide counts, conflict counts, and scores. Panel F shows the 'Peptides of C4YQR7' table, listing peptides with their scores, hits, masses, mass errors, retention times, charges, tags, abundances, and conflict counts.



The lower left panel (E) displays the Conflicting proteins for the feature highlighted in panel (B) this includes the current protein in panel A as indicated by the orange ball to the right of the accession.

The Accession and description for the 2 proteins highlighted in Panels A and E are shown in the middle margin. As most of the features are conflicting between the 2 closely related proteins one simple way to resolve these conflicts is to favour the protein with the higher score and greater number of non-conflicting peptides.

One way to do this is to right click on the lower scoring protein in panel E which only has one unique peptide and turn off all its peptides

Conflicting proteins for feature 6734

Accession	Peptides	Conflict	Protein Score	Peptide
P34731	75 (5)	90	696	5.62
C4YQR7	70 (1)	89	644	5.62
Q5A6R2	25			8.49

Turn off all peptides

All the peptides are now switched off in panel B and all the entries for the lower scoring protein are set to zero. The higher scoring protein now has 66 non-conflicting peptides and only 3 conflicts

The screenshot shows the Progenesis QI interface with several panels labeled A, B, and E. Panel A shows a list of proteins with P34731 highlighted. Panel B shows a list of peptides for feature 6734, with the lower scoring protein Q5A6R2 having its peptides switched off. Panel E shows the conflicting proteins for feature 6734, with P34731 and C4YQR7 highlighted. The middle margin shows protein descriptions for P34731 and C4YQR7.

To resolve these remaining conflicts first order the conflicts in panel B and select the top one (which may still be selected) Panel B will display the peptides for this protein and the number of conflicts for each peptide. Panel E will also update to show the conflicting protein.

The screenshot shows the Progenesis QI interface after resolving a conflict. Panel B shows the peptides for feature 6734, with the top peptide selected. Panel E shows the conflicting proteins for feature 6734, with Q5A6R2 highlighted. A red arrow points to the 'Conflict' column in the peptide list, indicating the resolution of the conflict.

Again, favouring the protein with the higher score, but this time resolve the conflict by switching off (or un-assigning) the peptide in panel F for the protein with the lower score. By doing this the other 3 panels

The screenshot shows the Progenesis QI interface after resolving another conflict. Panel B shows the peptides for feature 1992, with the top peptide selected. Panel E shows the conflicting proteins for feature 1992, with C4YQD9 highlighted. A red arrow points to the 'Conflict' column in the peptide list, indicating the resolution of the conflict.

update to show the change in conflicts.

Repeat this process until there are no conflicts remaining for the current protein in Panel A.

**Protein: P34731 Fatty acid synthase subunit beta GN=FAS1 PE=3 SV=1**  
**Protein: C4YL05 Leucyl-tRNA synthetase GN=CAWG\_01519 PE=4 SV=1**

**Conflicting proteins for feature 2538**

Accession	Peptides	Conflict	Protein Score	Peptide
P34731	75	0	696	8.07
C4YL05	31	2	242	5.51

**Peptides of C4YL05**

#	Score	Hits	Mass	Mass error (p...)	RT (mins)	Charge	Tag	Abundance	Conflict	Peptide
2538	5.51	2	1391.677	-12.3	55.2	2		3.12E+04	0	DV1

Now repeat using a similar approach for the next protein in Panel A, here the situation is similar.

**Protein: P43098 Fatty acid synthase subunit alpha GN=FAS2 PE=3 SV=1**  
**Protein: C4YQN7 Fatty acid synthase alpha subunit reductase GN=CAWG\_02796 PE=4 SV=1**

**Conflicting proteins for feature 111**

Accession	Peptides	Conflict	Protein Score	Peptide
P43098	55 (4)	62	471	6.98
C4YQN7 (+1)	54 (3)	62	472	6.98
Q5A017	19 (18)			

**Peptides of C4YQN7**

#	Score	Hits	Mass	Mass error (p...)	RT (mins)	Charge	Tag	Abundance	Conflict	Peptide
111	6.98	6	1661.866	-0.234	52.1	2		1.06E+05	2	L1E

Resolution of conflicts for this protein

**Protein: P43098 Fatty acid synthase subunit alpha GN=FAS2 PE=3 SV=1**  
**Protein: P53698 Cytochrome c GN=CYC1 PE=3 SV=3**

**Conflicting proteins for feature 39195**

Accession	Peptides	Conflict	Protein Score	Peptide
P43098	55	0	471	7.21
P53698 (+1)	5	0	35.6	5.23
C4YQN7 (+1)	0	0	0	7.21

**Peptides of P53698**

#	Score	Hits	Mass	Mass error (p...)	RT (mins)	Charge	Tag	Abundance	Conflict	Peptide
12117	6.46	4	1151.622	1.71	44.8	3		2.11E+04	0	VG E

Adopting a similar approach to the next protein favouring the protein with the highest score as each conflict is assessed.

In this case the first peptide for protein (P41797) has 5 conflicting proteins therefore move down the 5 conflicting proteins in panel E resolving the conflict in favour of this protein before moving on to the next peptide (which has 4 conflicts) in Panel B.

**Protein: P41797 Heat shock protein SSA1 GN=SSA1 PE=1 SV=2**  
**Protein: P46587 Heat shock protein SSA2 GN=SSA2 PE=1 SV=3**

**Conflicting proteins for feature 177**

Accession	Peptides	Conflict	Protein Score	Peptide
P46587	38 (21)	49	454	8.27
P41797	33 (14)	57	396	8.27
Q5A0K5	26 (25)	5	295	9.56
Q5A516	19	5	190	7.68

**Peptides of P46587**

#	Score	Hits	Mass	Mass error (p...)	RT (mins)	Charge	Tag	Abundance	Conflict	Peptide
177	8.27	2	1182.636	-2.75	52.3	2		1.64E+05	5	FE1

Using this approach, resolve this protein's conflicts.

Protein: P41797 Heat shock protein SSA1 GN=SSA1 PE=1 SV=2  
 Protein: P10592 Heat shock protein SSA2 OS=Saccharomyces cerevisiae GN=SSA2 PE=1 SV=3

Conflicting proteins for feature 177

Accession	Peptides	Conflict	Protein Score	Pep
Q5AS16	19	0	190	7.68
P10591	18 (3)	50	186	8.27
P10592	15 (0)	48	157	8.27

Peptides of P10592

#	Score	Hits	Mass	Mass error (p...)	RT (mins)	Charge	Tag	Abundance	Conflict	Peptide
177	8.27	2	1182.636	-2.75	52.3	2		1.64E+05	2	F E I
3367	7.5	2	960.492	-21.3	27.3	2		3.25E+04	2	I S F
35462	7.57	3	960.534	22.3	21.2	3		1.68E+03	2	I S F
8508	7.57	3	960.522	9.04	21.2	2		1.02E+04	2	I S F

**Note:** where a protein has no or very few unique peptides and shares a large number of conflicts with a similarly named protein from a different species then right click and turn off all peptides

By adopting a combination of these rules you can work through the resolution of conflicts for the whole data set. Having completed **Conflict resolution**, move forward to **Review Proteins**.

## Stage 12: Review Proteins

The **Review Proteins** stage opens displaying details for all proteins. You can now create tags at the level of the proteins. Right click on the table and create Quick Tags for proteins with an Anova p value  $\leq 0.05$  and Max Fold change  $\geq 2$ .

File | Review Alignment | Filtering | Experiment Design Setup | Review Peak Picking | Peptide Statistics | Identify Peptides | Refine Identifications | Resolve Conflicts | Review Proteins | Protein Statistics | Report

Protein options... A B C

Accession	Peptide count	Confidence score	Anova (p)	Tag	Max fold change	Highest Mean	Lowest Mean	Description
C4YDY3	14	128	3.55E-08		2.03	B	A	Peroxisomal catalase GN=CAWG_00732 PE=4 SV=1
Q9Y872			1.72E-07		1.91	C	A	Sulfate adenylyltransferase GN=MET3 PE=3 SV=2
C4YFY2			2.65E-07		2.18	C	A	Vesicular-fusion protein SEC18 GN=CAWG_00094 PE=4 SV=1
Q9UVL1			1.18E-07		2.15	A	C	Non-histone chromosomal protein 6 GN=NHP6 PE=3 SV=1
Q7Z8E8			1.18E-07		2.15	B	A	Cell surface hydrophobicity-associated protein (Fragment) OS=Candida d
Q5AEN1			1.18E-07		2.15	B	A	Cytochrome c peroxidase, mitochondrial GN=CCP1 PE=3 SV=1
Q9P8Q7			1.18E-07		2.15	B	A	Isocitrate lyase GN=ICL1 PE=3 SV=1
Q59W33	20	216	1.18E-07		2.15	C	A	Glycerol-3-phosphate dehydrogenase GN=GPD2 PE=3 SV=1
C4YVY9	2	16.9	1.7E-06		2.33	A	C	60S ribosomal protein L38 GN=CAWG_03801 PE=4 SV=1
P87219	5	44.4	2.13E-06		3.82	C	A	Sorbose reductase SOU1 GN=SOU1 PE=1 SV=1
C4YRH4	8	58.3	2.39E-06		1.38	B	A	NAD(P)H-dependent D-xylose reductase [II] GN=CAWG_04674 PE=4 SV=1
P53707	9	70.1	2.74E-06		1.83	B	A	37 kDa cell surface protein GN=CSP37 PE=1 SV=2
P43071 (+1)	11	56.5	3.03E-06		1.75	B	A	Multidrug resistance protein CDR1 GN=CDR1 PE=3 SV=1

Selected protein: Sulfate adenylyltransferase GN=MET3 PE=3 SV=2

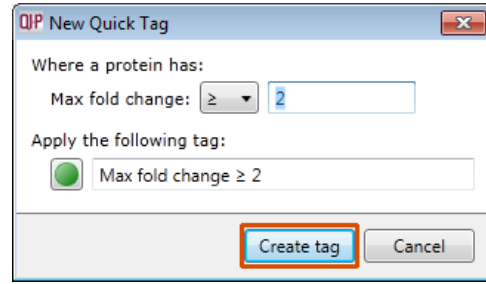
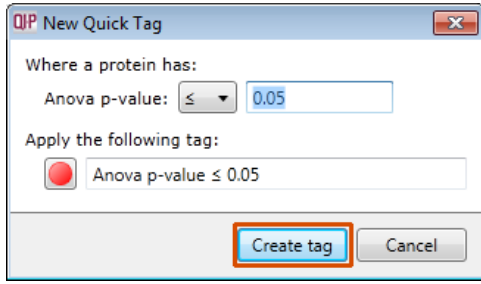
View peptide measurements

ArcSinh Normalised...

Quantifiable proteins displayed: 630

Section Complete

Create the 2 Quick Tags



Corresponding protein tags are now displayed for the proteins.

To view the corresponding peptide measurements for the current protein either double click on the protein in the table or use the **View peptide measurements** below the table.

QIP Progenesis QI.P\_Tutorial HDMSe - Progenesis QI for proteomics

File Import Data Review Alignment Filtering Experiment Design Setup Review Peak Picking Peptide Statistics Identify Peptides Refine Identifications Resolve Conflicts Review Proteins Protein Statistics Report

nonlinear A Waters Company

No filter applied Create... Search Protein options... A B C Help

Accession	Peptide count	Confidence score	Anova (p)	Tag	Max fold change	Highest Mean	Lowest Mean	Description
C4YDY3	14	128	3.55E-08		2.03	B	A	Peroxisomal catalase GN=CAWG_00732 PE=4 SV=1
Q9YB72	5	36.6	1.72E-07		1.91	C	A	Sulfate adenylyltransferase GN=MET3 PE=3 SV=2
C4YFY2	7	38.6	2.65E-07		2.18	C	A	Vesicular-fusion protein SEC18 GN=CAWG_00094 PE=4 SV=1
Q9UVL1	3	29.8	4.65E-07		2.45	A	C	Non-histone chromosomal protein 6 GN=NHP6 PE=3 SV=1
Q7Z8E8	6	43.2	5.88E-07		1.98	B	A	Cell surface hydrophobicity-associated protein (Fragment) OS=Candida d
Q5AEN1	6	44.7	6.62E-07		1.47	B	A	Cytochrome c peroxidase, mitochondrial GN=CCP1 PE=3 SV=1
Q9P8Q7	10	91.7	6.89E-07		1.75	B	A	Isocitrate lyase GN=ICL1 PE=3 SV=1
Q59W33	20	216	1.37E-06		2.47	C	A	Glycerol-3-phosphate dehydrogenase GN=GPD2 PE=3 SV=1
C4VIY9	2	16.9	1.7E-06		2.33	A	C	60S ribosomal protein L38 GN=CAWG_03801 PE=4 SV=1
P87219	5	44.4	2.13E-06		3.82	C	A	Sorbose reductase SOU1 GN=SOU1 PE=1 SV=1
C4YRH4	8	58.3	2.39E-06		1.38	B	A	NAD(P)H-dependent D-xylose reductase I,II GN=CAWG_04674 PE=4 SV=1
P53707	9	70.1	2.74E-06		1.83	B	A	37 kDa cell surface protein GN=CSP37 PE=1 SV=2
P43071 (+1)	11	56.5	3.03E-06		1.75	B	A	Multidrug resistance protein CDR1 GN=CDR1 PE=3 SV=1

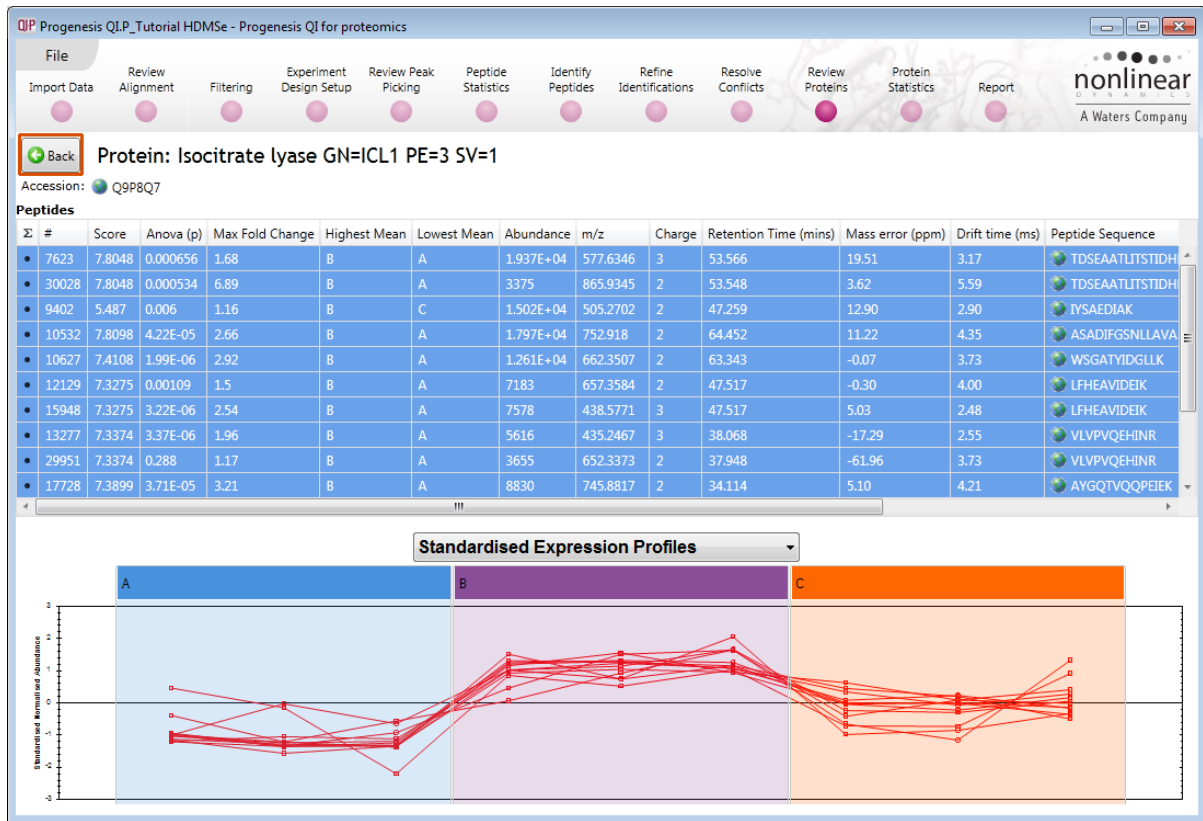
Selected protein: Non-histone chromosomal protein 6 GN=NHP6 PE=3 SV=1

[View peptide measurements](#)

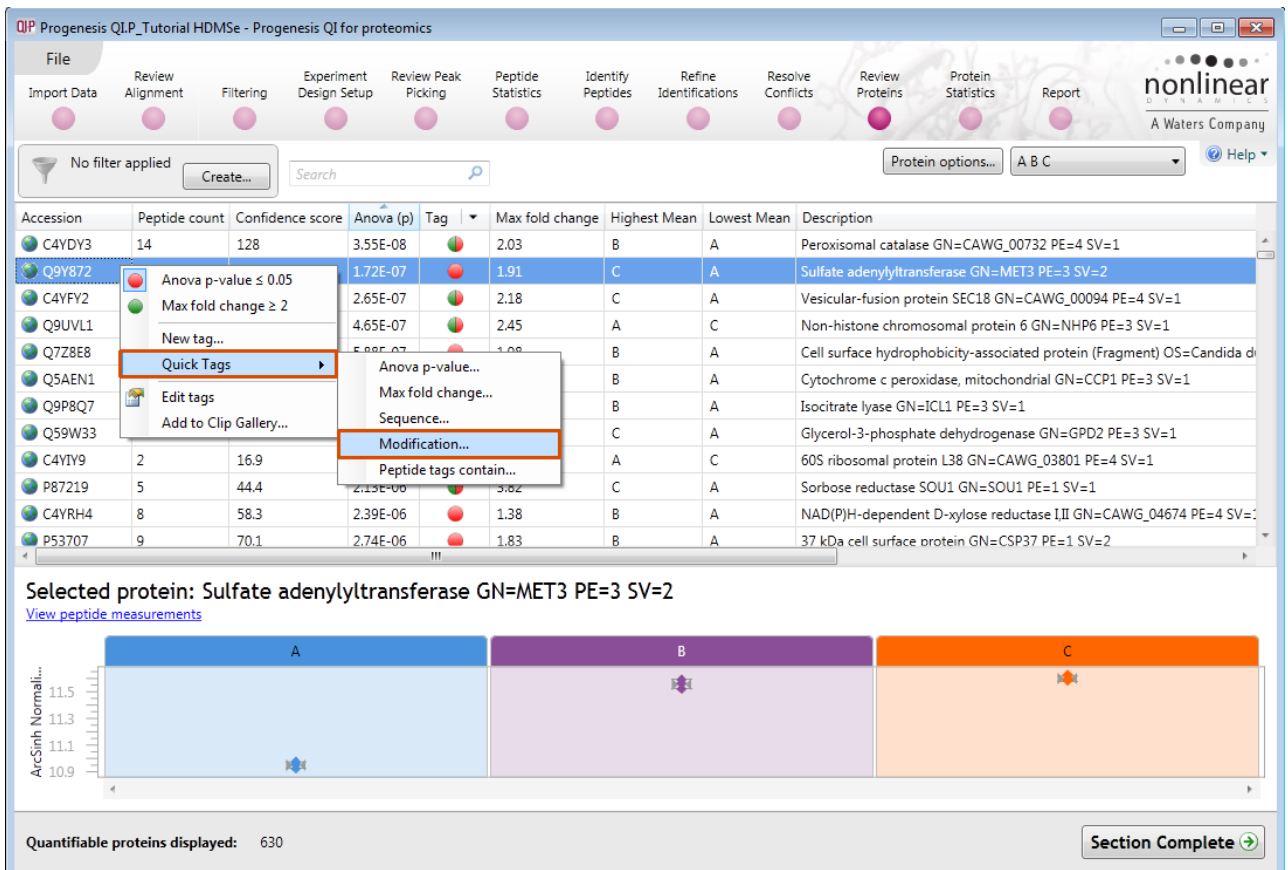
Quantifiable proteins displayed: 630

Section Complete

**Note:** by selecting all the peptides you can compare the pattern of expression across all the samples allowing you to identify 'atypical' behaviour of peptides assigned to the current protein.

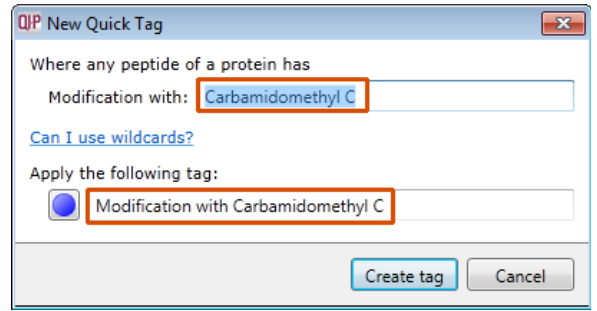


Modified proteins can be located by specifically searching for proteins containing modified peptides. Use the **Back** button to return to the Proteins List and right click on it and select **Modification** from the list of **Quick Tags**.

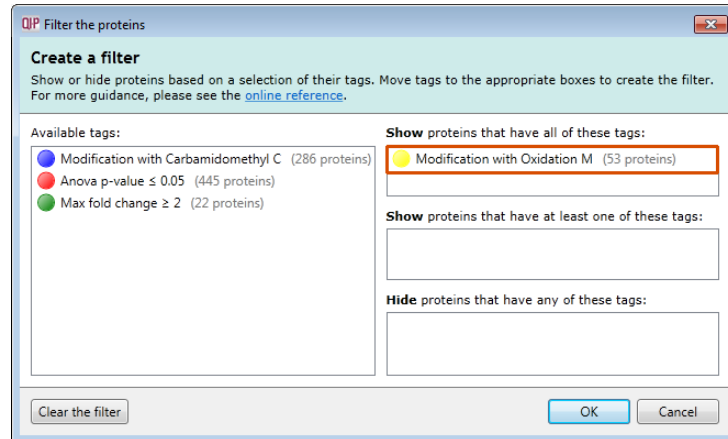


To find those proteins containing peptides with Carbamidomethyl on cysteine residues enter **Carbamidomethyl C**. This will automatically provide a named tag when you click **Create tag**.

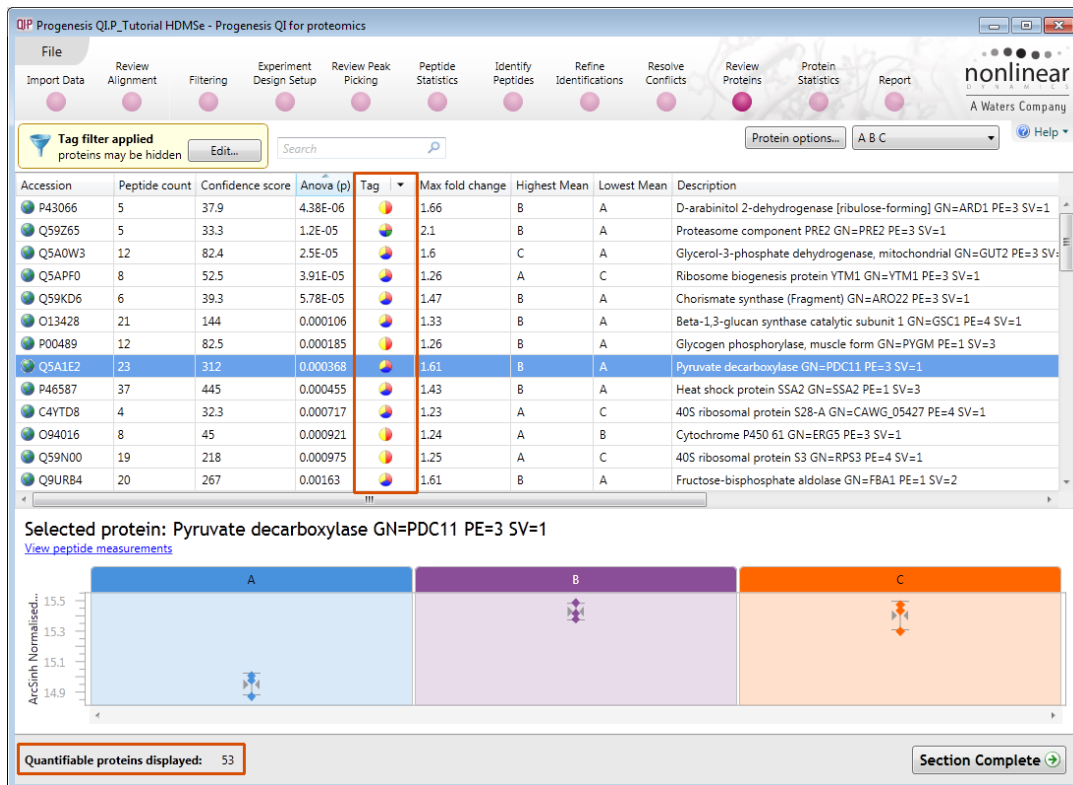
Repeat this for Oxidation of methionine residues.



To reduce the table to displaying only these proteins with modified peptides (oxidation on methionine) use the tag filter to focus on these proteins.



The proteins table will only display those proteins containing modified peptides.

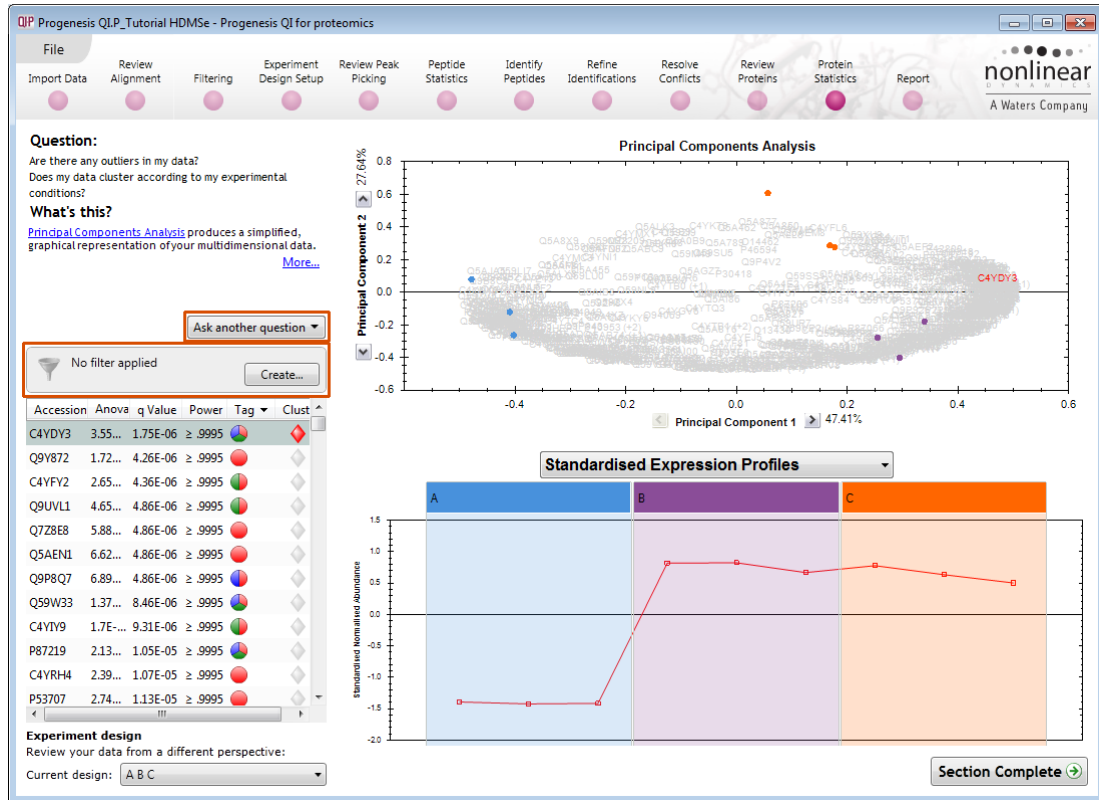


**Note:** the **Sequence** Quick tag can be used to locate Proteins containing peptides with specific motifs. Before leaving this stage in the workflow clear active Tag Filters.

Now move to the Protein Statistics section by clicking on **Protein Statistics** icon on the workflow at the top of the screen.

## Stage 13: Protein Statistics

Protein Statistics opens with a Principal Components Analysis (PCA) for all the proteins displayed.



The Multivariate Stats can now be applied to all or subsets of proteins as determined by the current Tag filters. Allowing you to identify similar patterns of expression using the Correlation Analysis. Click on 2 of the branches (holding the Ctrl key down) to see differing patterns of expression.

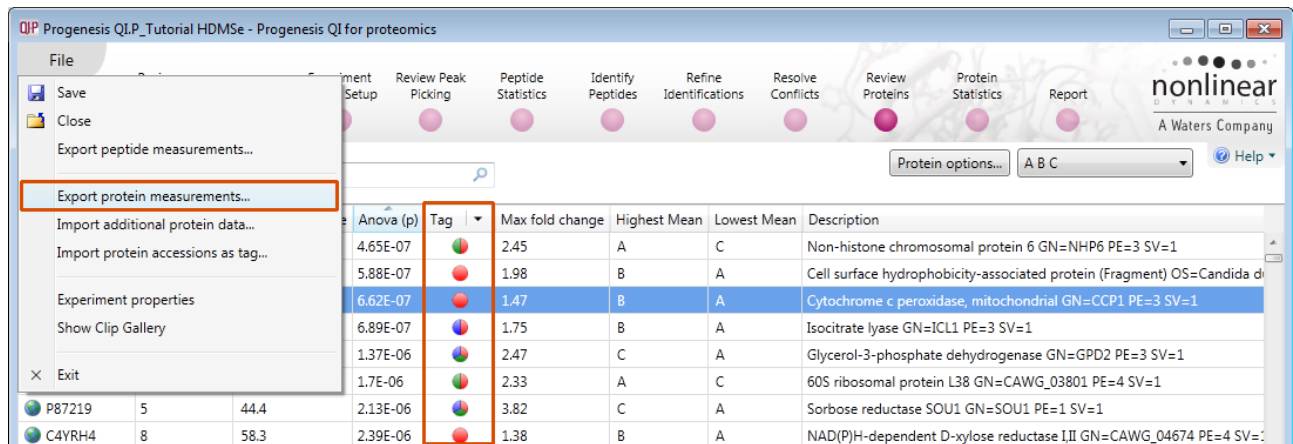


Now move to the **Report** section to report on Proteins and /or peptides.

## Stage 14: Exporting data

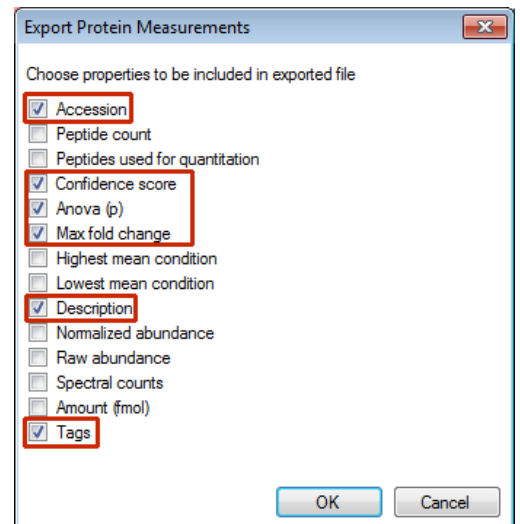
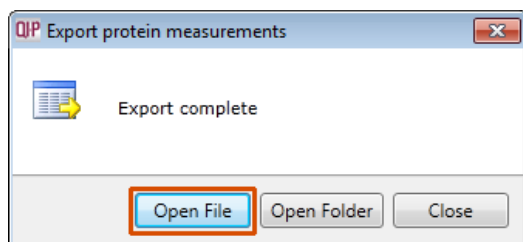
Protein measurement data can be exported from Progenesis QI for proteomics Workflow from the **Review Peak Peaking** stage onwards. Various data can be output and imported during the LC-MS workflow using the File Menu options. The actual data available depends on the current stage of the Workflow.

As an example, try exporting data at the **Review Proteins** stage.



This opens a dialog that controls the content of the exported file. In this example we are only exporting a limited number of fields.

On saving the file the option to Open File is available.



The csv file opens displaying the exported Protein Measurements

	A	B	C	D	E
1					
2	Accession	Confidence score	Anova (p)	Max fold change	Description
3	C4YD3	127.8477	3.55E-08	2.026682963	Peroxisomal catalase GN=CAWG_00732 PE=4 SV=1
4	Q9Y872	36.5644	1.72E-07	1.9125987	Sulfate adenylyltransferase GN=MET3 PE=3 SV=2
5	C4YFY2	38.5924	2.65E-07	2.180571648	Vesicular-fusion protein SEC18 GN=CAWG_00094 PE=4 SV=1
6	Q9UVL1	29.789	4.65E-07	2.448480165	Non-histone chromosomal protein 6 GN=NHP6 PE=3 SV=1
7	Q7Z8E8	43.2056	5.88E-07	1.976380131	Cell surface hydrophobicity-associated protein (Fragment) OS=Candida dubliniensis GN=CSH1 PE=4 SV=1
8	Q5AEN1	44.6778	6.62E-07	1.471490415	Cytochrome c peroxidase, mitochondrial GN=CCP1 PE=3 SV=1
9	Q9P8Q7	91.7489	6.89E-07	1.751639741	Isocitrate lyase GN=ICL1 PE=3 SV=1
10	Q59W33	215.5667	1.37E-06	2.466135008	Glycerol-3-phosphate dehydrogenase GN=GPD2 PE=3 SV=1
11	C4YIY9	16.9005	1.70E-06	2.32820763	60S ribosomal protein L38 GN=CAWG_03801 PE=4 SV=1
12	P87219	44.3802	2.13E-06	3.824236542	Sorbose reductase SOU1 GN=SOU1 PE=1 SV=1
13	C4YRH4	58.2748	2.39E-06	1.38424931	NAD(P)H-dependent D-xylose reductase I,II GN=CAWG_04674 PE=4 SV=1
14	P53707	70.076	2.74E-06	1.832246334	37 kDa cell surface protein GN=CSP37 PE=1 SV=2
15	P43071;C4YQP2	56.4782	3.03E-06	1.752950291	Multidrug resistance protein CDR1 GN=CDR1 PE=3 SV=1
16	C4YL44	43.4056	3.71E-06	1.539504242	Extracellular matrix protein 4 GN=CAWG_01559 PE=4 SV=1

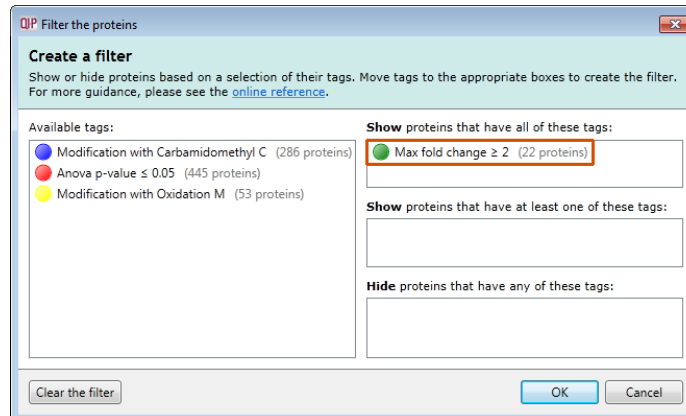
**Note:** you can add additional columns to the csv file (shown above) containing numerical or descriptive information about the proteins. Then using the **Import additional Protein Data...** option on the File menu selectively import these extra columns into your experiment.



## Stage 15: Reporting

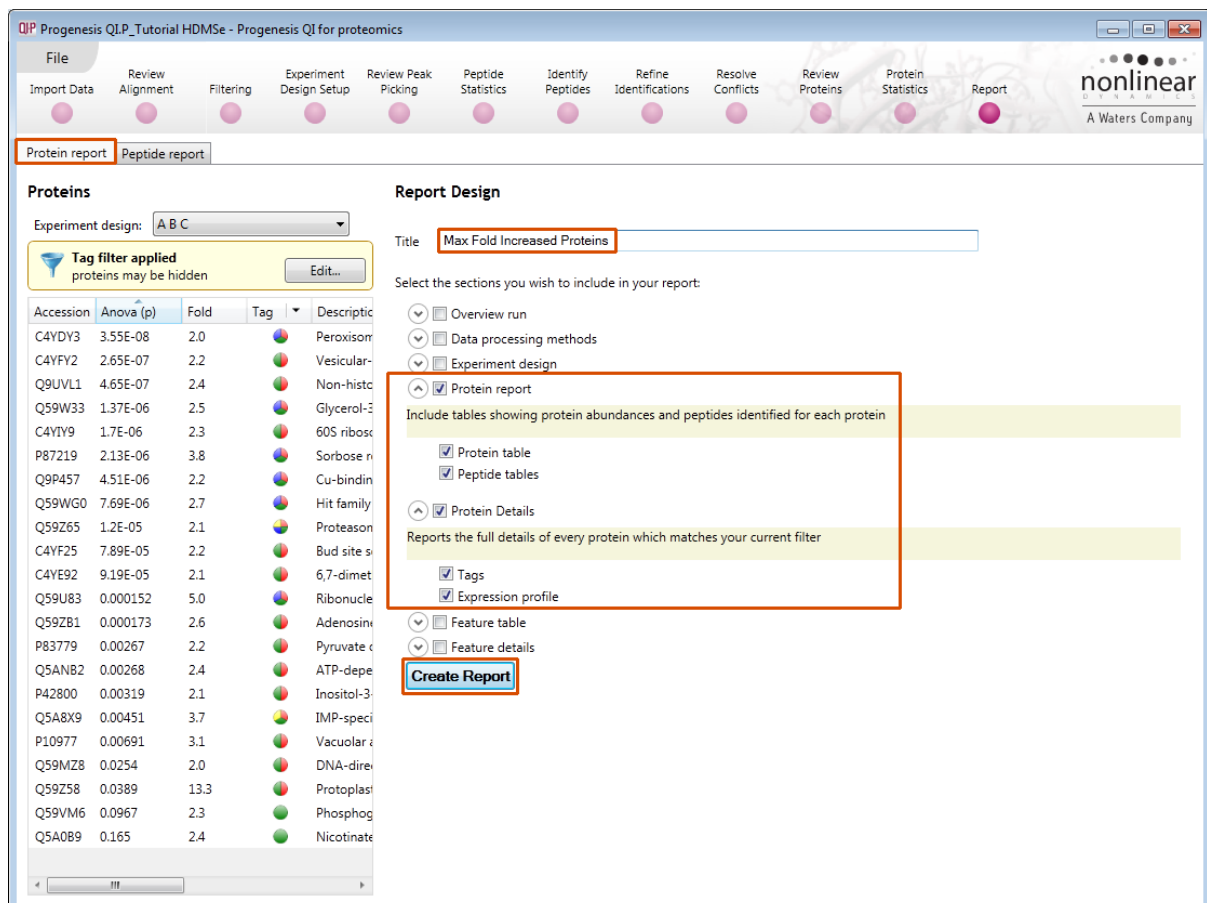
The **Report Design** stage allows you to select what views you want to include in a report based on the list of **currently selected proteins**.

**Note:** this facility is used to generate Html reports on a limited selection of Proteins in your data. Creating a report on all the data in your experiment can take a long time



As an example we will create a report for **only** the proteins showing a Max Fold change of greater than 2.

1. First reduce the proteins to report on by selecting the '**Max fold change  $\geq$  2**' tag. In this example it reduces the number of proteins in the table to 22.
2. Expand the various Report Design options (by default they are all selected)
3. Un-tick as shown below
4. Click **Create Report**



This opens a dialog to allow you to save the report, after which it will be opened in the form of a web page.

Click on the **Accession No.** in the proteins section of the Report and this will take you to the Assigned peptides for this protein

Max Fold Increased Proteins

**Experiment:** Progenesis QI\_P\_Tutorial HDMSe

**Report created:** 23/01/2014 15:22:39

Proteins

Protein building options

Protein grouping **Group similar proteins**

Protein quantitation **Using only features with no protein conflicts**

Accession	Peptides	Score	Anova (p)	Fold	Tags	Description	Average Normalised Abundances		
							A	B	C
Q59V33	20	215.57	1.37e-006	2.47		Glycerol-3-phosphate dehydrogenase GN=GPD2 PE=3 SV=1	2.32e+005	5.60e+005	5.71e+005
C4YD3									
P42800									
P87219									
C4YFY2									
C4YE92									
Q59Z65									
Q9P457									
Q9UVL1									
P10977									
Q59V30									
Q59VM6									
Q59U83									

P87219

Sorbose reductase SOU1 GN=SOU1 PE=1 SV=1  
5 peptides

Sequence	Feature	Score	Hits	Mass	Charge	Tags	Conflicts	Modifications	In quantitation	Drift time (ms)	Average Normalised Abundances		
											A	B	C
APQLPSNVLDLFLSK	1871	8.59	9	1640.9272	2		0		yes	5.04	1.11e+004	5.32e+004	6.31e+004
IVDCDLNGVYYCAHTVGQIFKK	34858	5.76	2	2599.2644	3		0	[4] Carbamidomethyl C [12] Carbamidomethyl	yes	4.55	1225.09	1110.45	1409.60

Accession P87219

**Description** Sorbose reductase SOU1 GN=SOU1 PE=1 SV=1

**Peptides** 5

**Score** 44.38

**Anova** 2.13e-006

**Fold** 3.82

- Anova p-value ≤ 0.05
- Max fold change ≥ 2
- Modification with Carbamidomethyl C

Condition	ArcSinh Normalised Abundance
A	~11.1
B	~12.2
C	~12.3

Tags

- Separate
- Most Ab
- Anova p
- Max fold

Having closed the report it can be reopened by double clicking on the saved html file.

**Note:** you can also copy and paste all or selected sections of the report to Excel and/or Word.

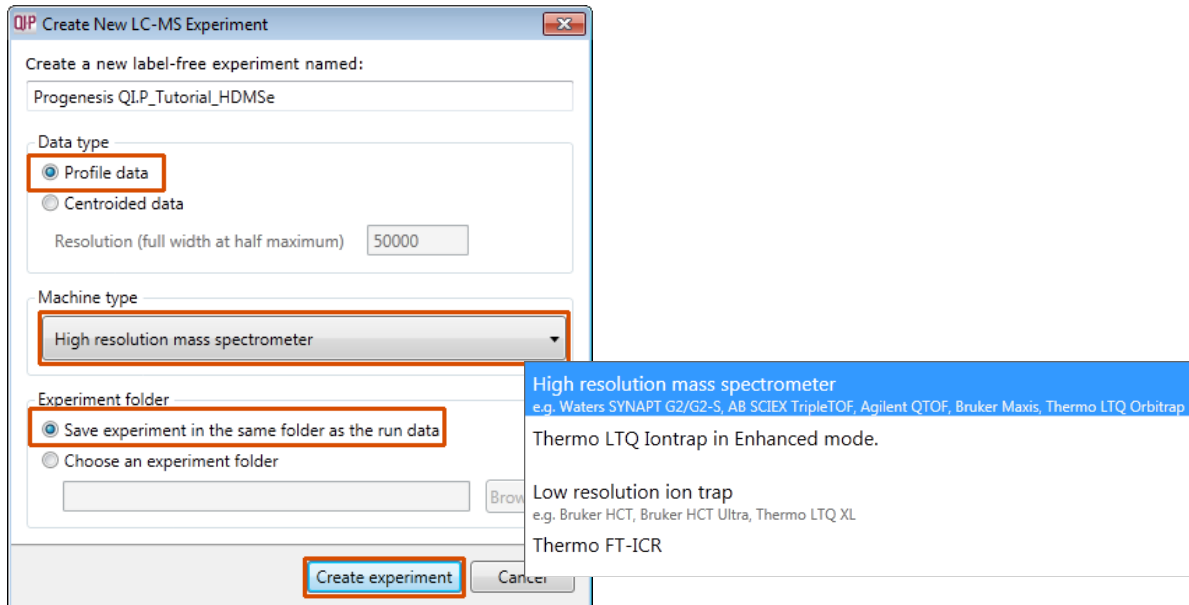
**Note:** there are separate panels for reporting on Proteins and Peptides.

## Appendix 1: Stage 1 Data Import and QC review of LC-MS data set

You can use your own data files, either by directly loading the raw files (Waters, Thermo, Bruker, ABSciex and Agilent) or, for other Vendors, convert them to mzXML or mzML format first.

To create a new experiment with your files select **New** give your experiment a name. Then select data type, the default is 'Profile data'.

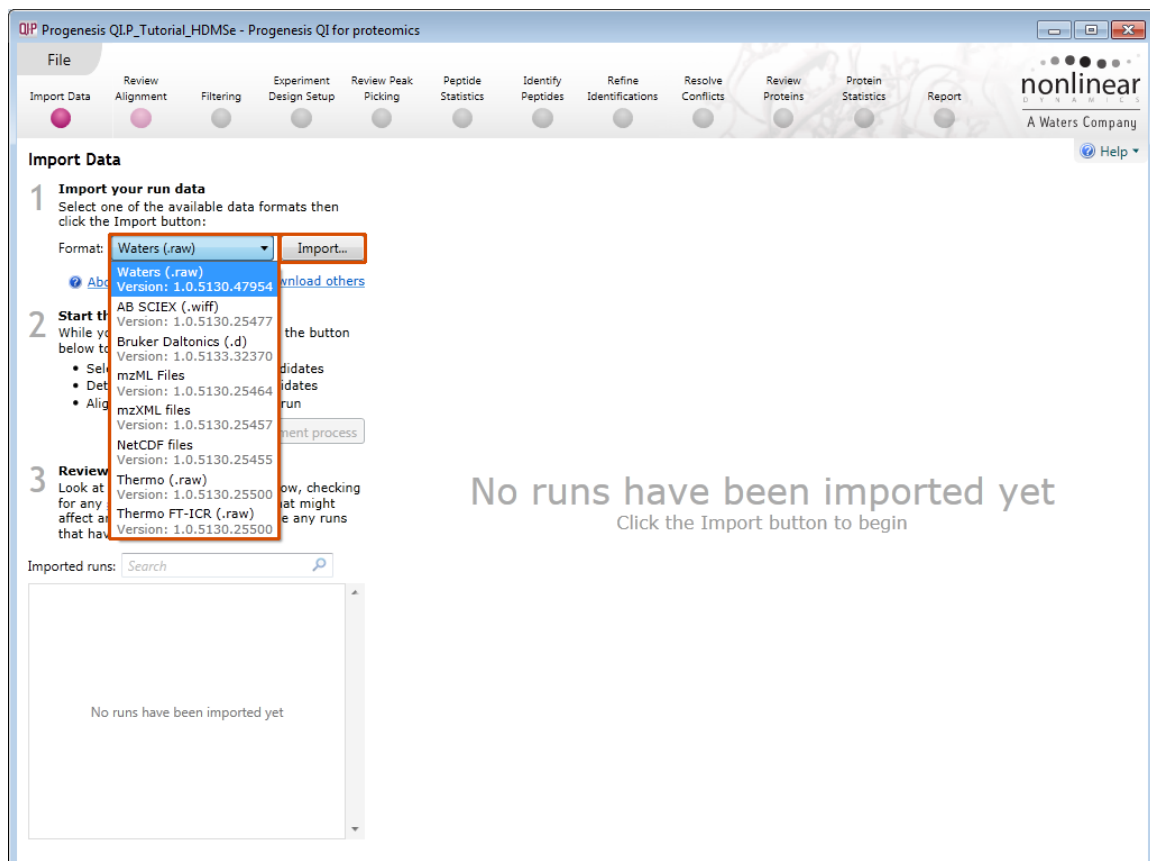
**Note:** if you have converted or captured the data as centroided then select Centroided data and enter the Resolution for the MS machine used.



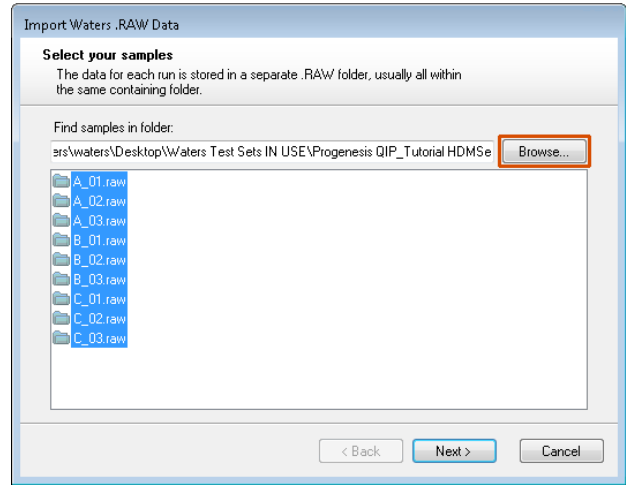
Click **Create experiment** to open the LC-MS Data Import stage of the workflow.

Select the 'Import Data file format', in this example they are Waters/SYNAPT data

Then locate your data files using **Import...**

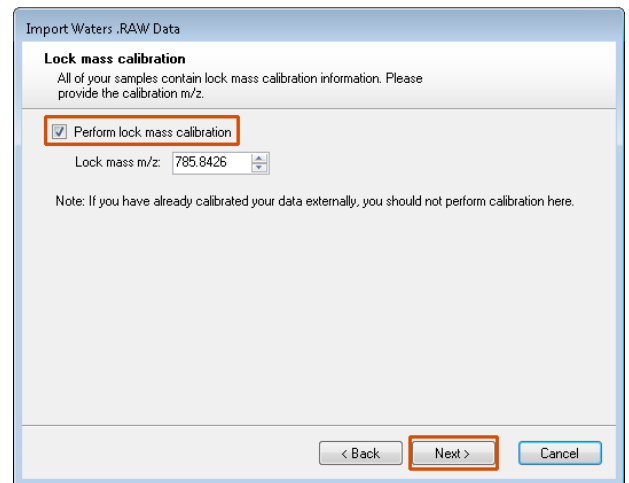


Locate and select all the .RAW folders (A\_01 to C\_03).



On importing, the lock mass calibration is read and presented on this dialog

You can, if required, alter the lock mass calibration at this step.



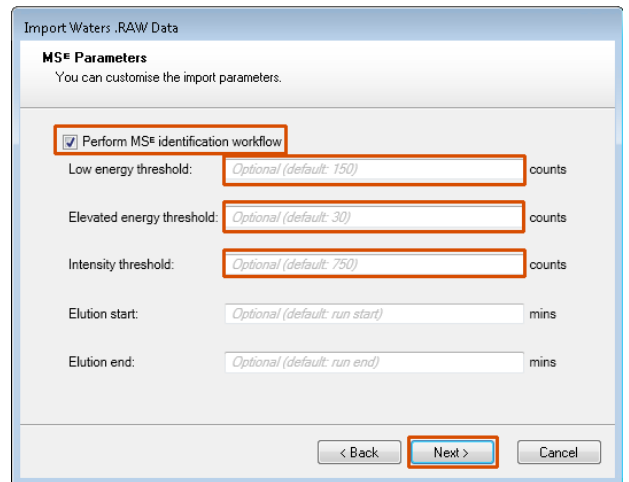
You can set the MS<sup>E</sup> parameters depending on how your data was acquired.

For HDMS<sup>e</sup> the settings are 150, 30 and 750

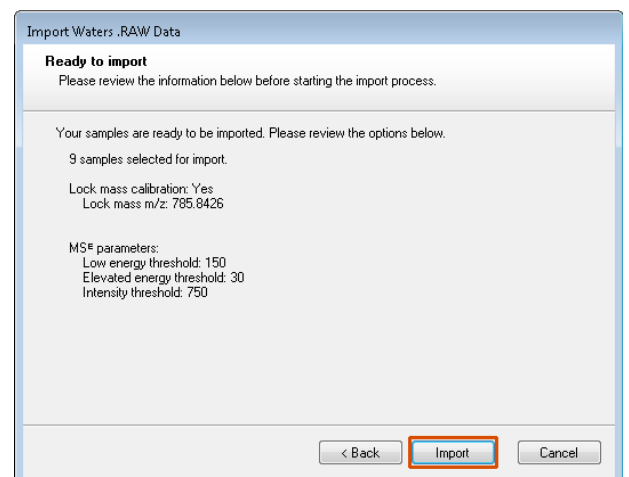
For MSe the settings are 135, 30 and 500.

The default is set to HDMS<sup>e</sup>.

**Note:** There is an option to switch off the MS<sup>E</sup> identification workflow



A summary of the loading parameters is provided before you click Import



On loading the selected runs your data set will be automatically examined and the size of each file will be reduced by a 'data modelling routine', which reduces the data by several orders of magnitude but still retains all the relevant quantitation and positional information.

**Note:** For a large number of files this may take some time.

Each data file appears as a 2D representation of the run. At this stage you will be warned if any of the data files have been 'centroided' during the data acquisition and conversion process.

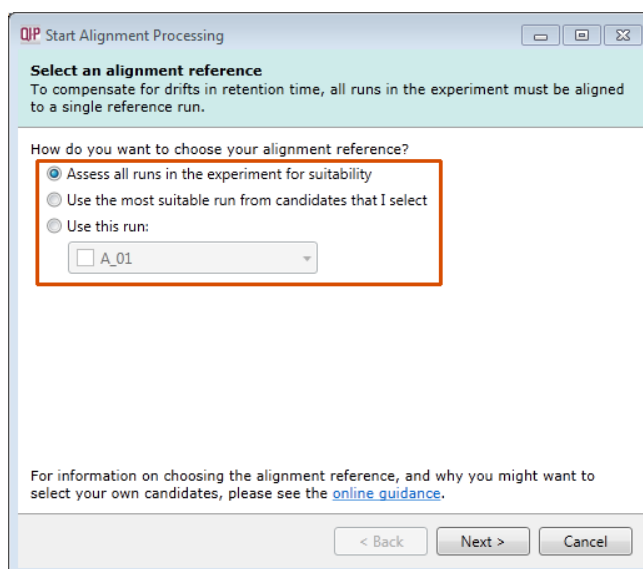
**Note:** as each data file is loaded the progress is reported in the **Import Data** list. The dialog below the run reports on the QC of the imported Data files. In this case 'No problems found' with the this data file

**Note:** details of the current run appear on the top right of the view.

**Note:** as the loading process starts you can also start the automatic alignment before the loading has completed. This is a 2 stage process that involves the selection of an Alignment Reference (either automatically or manually then the automatic alignment of all your runs to this Reference run.

Click **Start alignment process** to start the automatic alignment of your runs.

**Progenesis QI for proteomics provides three methods** for choosing the alignment reference run, as seen below:



### 1. Assess all runs in the experiment for suitability

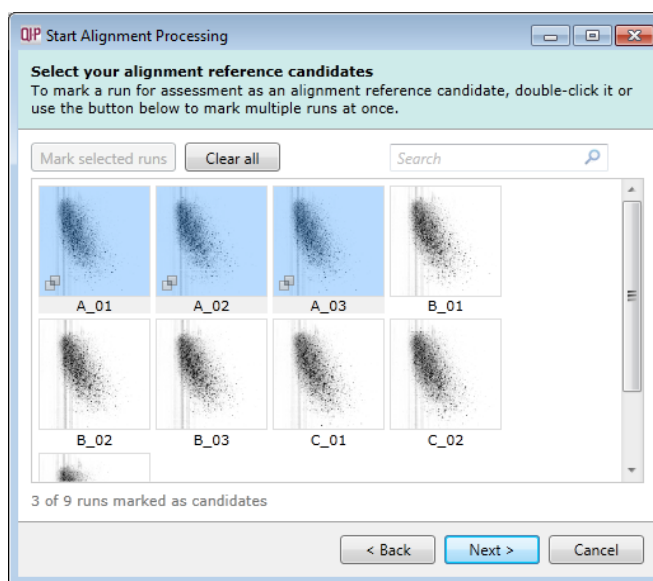
This method compares every run in your experiment to every other run for similarity.

The run with the greatest similarity to all other runs is chosen as the alignment reference.

If you have no prior knowledge about which of your runs would make a good reference, then this choice will normally produce a good alignment reference for you. This method can take a long time

### 2. Use the most suitable run from candidates that I select

This method asks you to choose a selection of reference candidates, and the automatic algorithm chooses the best reference from these runs.



When you have some prior knowledge of your runs suitability as references:

runs from pooled samples

runs for one of your experimental conditions will contain the largest set of common peptides.

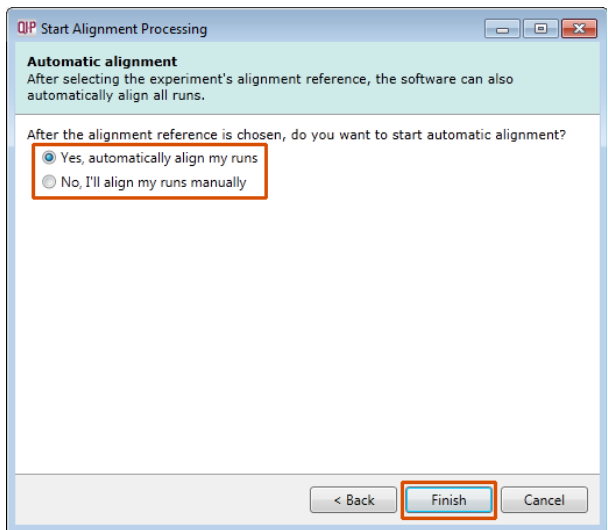
### 3. Use this run

This method allows you to manually choose the reference run.

Manual selection gives you full control, but there are a couple of risks to note:

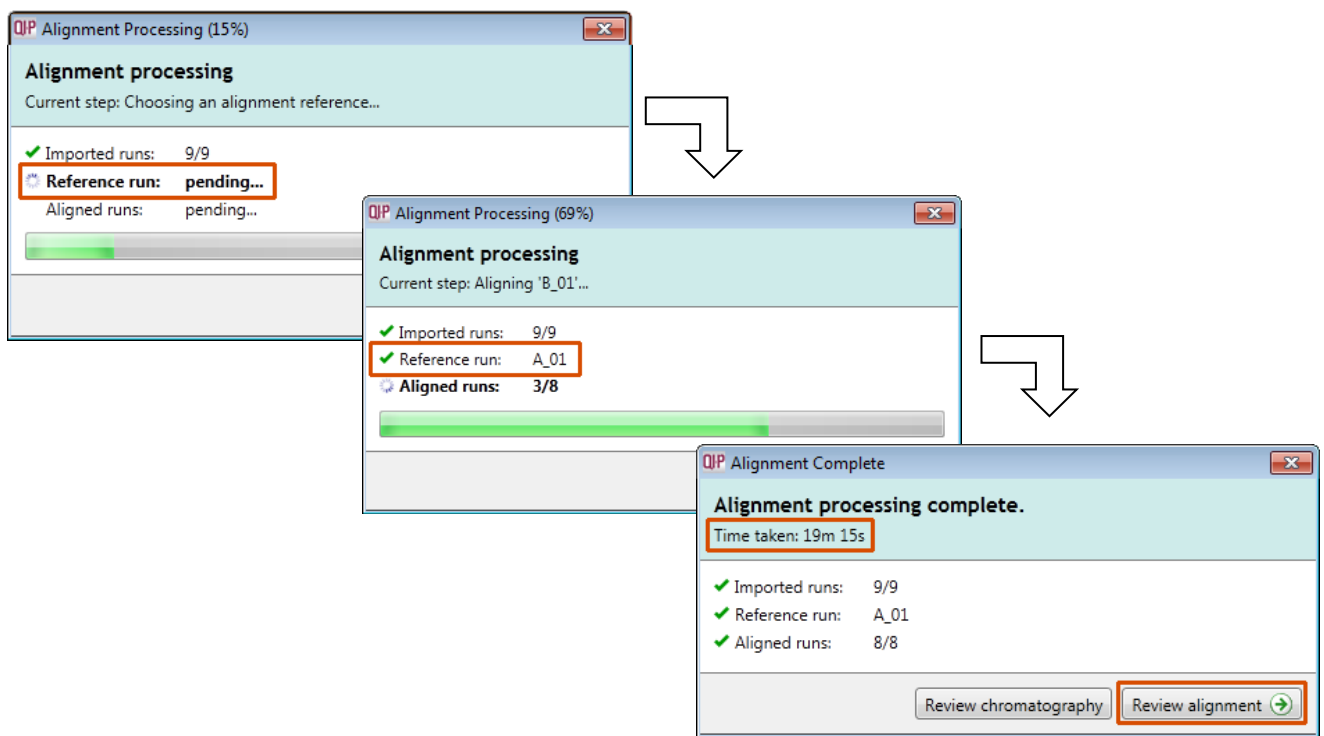
- If you choose a pending run which subsequently fails to load, alignment will not be performed.
- If you choose a run before it fully loads, and it turns out to have chromatography issues, alignment will be negatively affected (for this reason we recommend that you let your reference run fully load and assess its chromatography before loading further runs).

Once you have selected how to handle the choice of Reference run you will now be asked if you want to Align your runs automatically or manually.



Select automatically and click finish.

The Alignment process starts with the automatic selection of A\_01 as the reference



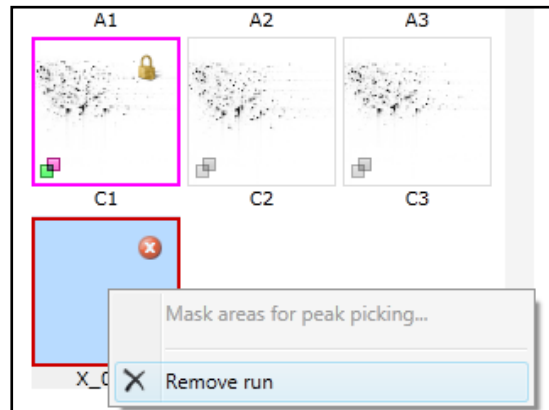
Once the Reference run has been chosen the automatic alignment is performed. As the whole process proceeds you get information on what stage has been performed and also the % of the process that has been completed.

**Note:** At this stage you have the option to Review the Chromatography or go straight to the review of the Automatic Alignment of your data.

## Review Chromatography

Each data file appears as a 2D representation of the run. At this stage you will be warned if any of the data files have been 'centroided' during the data acquisition and conversion process.

You can delete run(s) by left clicking on the run in the list.



At the Import Data stage you can examine the quality of the imported runs using the 2D representation of the runs

**Note:** details of the current run appear on the top right of the view.

Once you have reviewed the imported runs click on **Review Alignment** or **Section Complete** to move forward to the Review Alignment Stage.

**Note:** you will be offered the automatic alignment if you have not performed it automatically already.

Now move to the next stage in the workflow (page 9 in this user guide) by clicking **Section Complete**.



## Appendix 2: Manual assistance of Alignment

### Approach to alignment

To place manual alignment vectors on a run (B\_02 in this example):

**Review Alignment**  
Sample ions are aligned to compensate for drifts in retention time between runs.  
[Learn about the visualisations shown here](#)

- Align retention times automatically**  
For maximum reproducibility, the software can automatically align your runs.  
  
8 runs have no alignment vectors
- Review the alignment**  
Using the quality control measures, review and edit the runs' alignment:
  - Order the runs by alignment score and start by selecting the first run
  - Within each run, inspect and edit any areas rated as Needs Review[Learn about the review and editing process](#)

Run	Include?	Vectors	Score
A_01	<input checked="" type="checkbox"/>	Ref	
A_02	<input checked="" type="checkbox"/>	0	59.9%
A_03	<input checked="" type="checkbox"/>	0	82.9%
B_01	<input checked="" type="checkbox"/>	0	69.0%
B_02	<input checked="" type="checkbox"/>	0	44.2%
B_03	<input checked="" type="checkbox"/>	0	51.8%
C_01	<input checked="" type="checkbox"/>	0	39.5%
C_02	<input checked="" type="checkbox"/>	0	87.6%
C_03	<input checked="" type="checkbox"/>	0	64.9%

Ion maps: ■ Alignment target ■ Run being aligned

Alignment quality: ■ Good ■ OK ■ Needs review

**Section Complete**

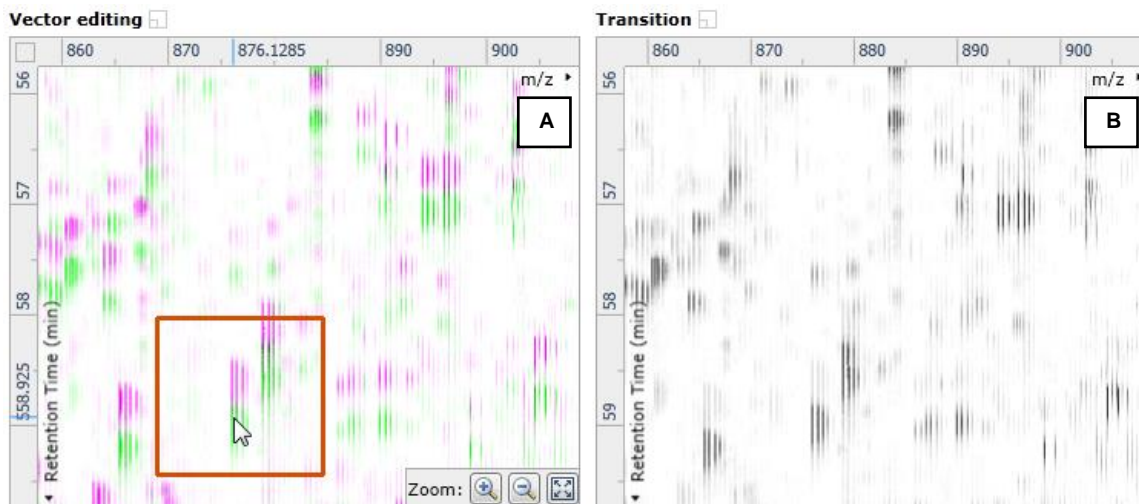
- Click on Run B\_02 in the **Runs** panel, this will be highlighted in green and the reference run (A\_01) will be highlighted in magenta.
- You will need approximately 5 - 10 **alignment vectors** evenly distributed from top to bottom of the whole run.
- First drag out an area on the **Ion Intensity Map** (C), this will reset the other 3 windows to display the same 'zoomed' area

**Note:** the features moving back and forwards between the 2 runs in the **Transition** window (B) indicating the misalignment of the two runs

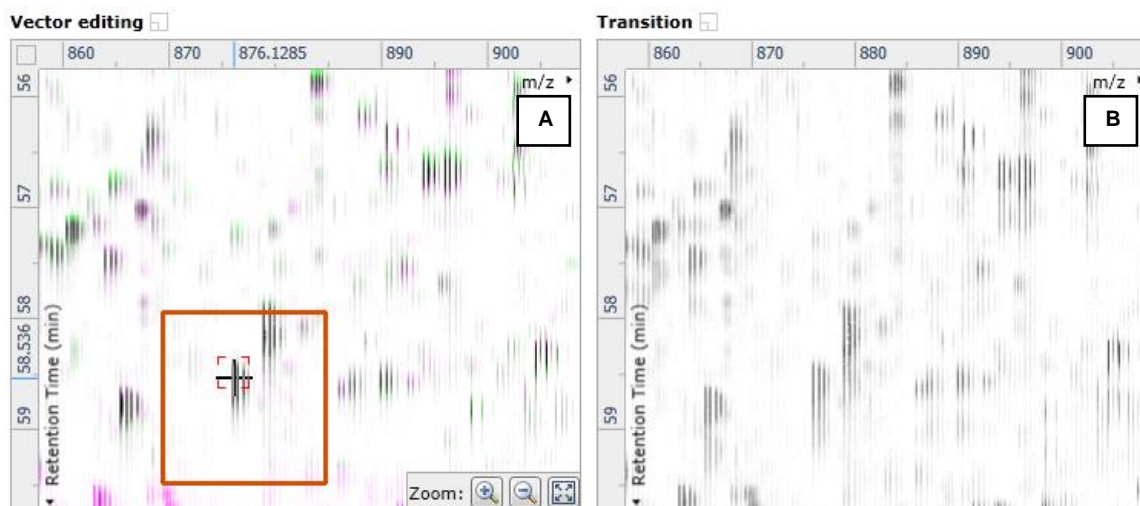
**Note:** the Ion Intensity Map gives you a colour metric, visually scoring of the current alignment and an overall score is placed next to the Vectors column in the table. With each additional vector this score will update to reflect the overall quality of the alignment. The colour coding on the Ion intensity Map will also update with each additional vector

**Note:** The **Total Ion Chromatograms** window (D) also reflects the misalignment of the 2 runs for the current Retention Time range (vertical dimension of the current Focus grid in the **Ion Intensity Map** window).

4. Click and hold on a green feature in Window A as shown below.

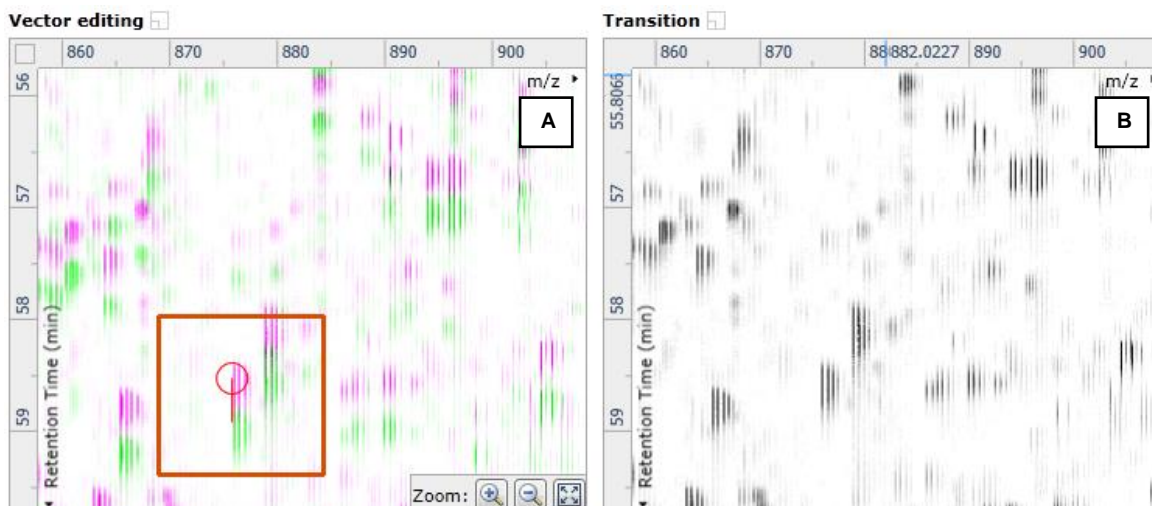


5. As you are holding down the left mouse button (depending on the severity of the misalignment), the alignment vector will automatically find the correct lock. If not, drag the green feature over the corresponding magenta feature of the reference run. The red box will appear as shown below indicating that a positional lock has been found for the overlapping features.



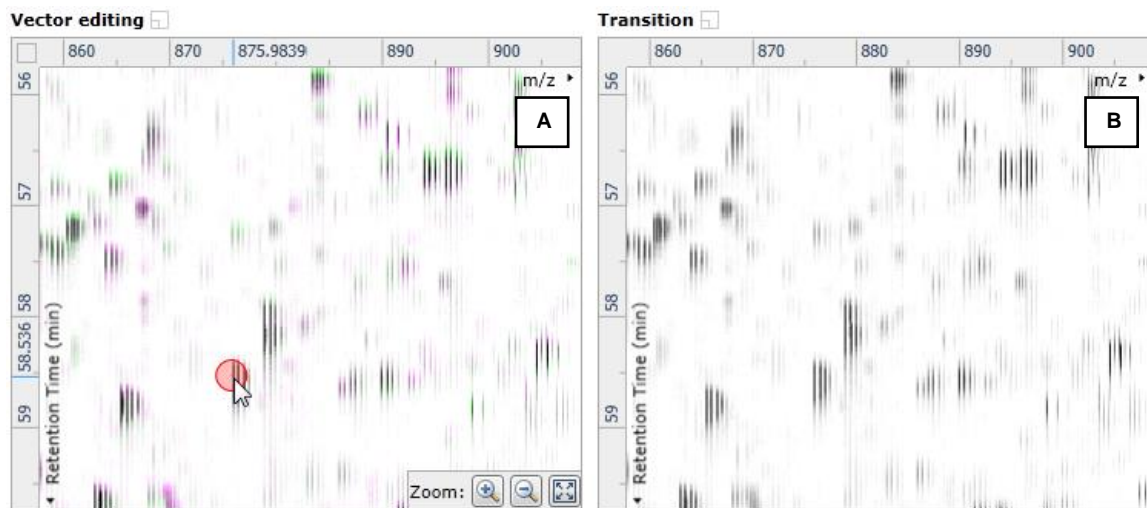
**Tip:** while holding down the mouse button hold down the **Alt** key. This will allow smooth movement of the cursor as the **Alt** key allows you to override the 'automatic alignment' performed as you depress the mouse button.

6. On releasing the left mouse button the view will 'bounce' back and a red vector, starting in the green feature and finishing in the magenta feature will appear.

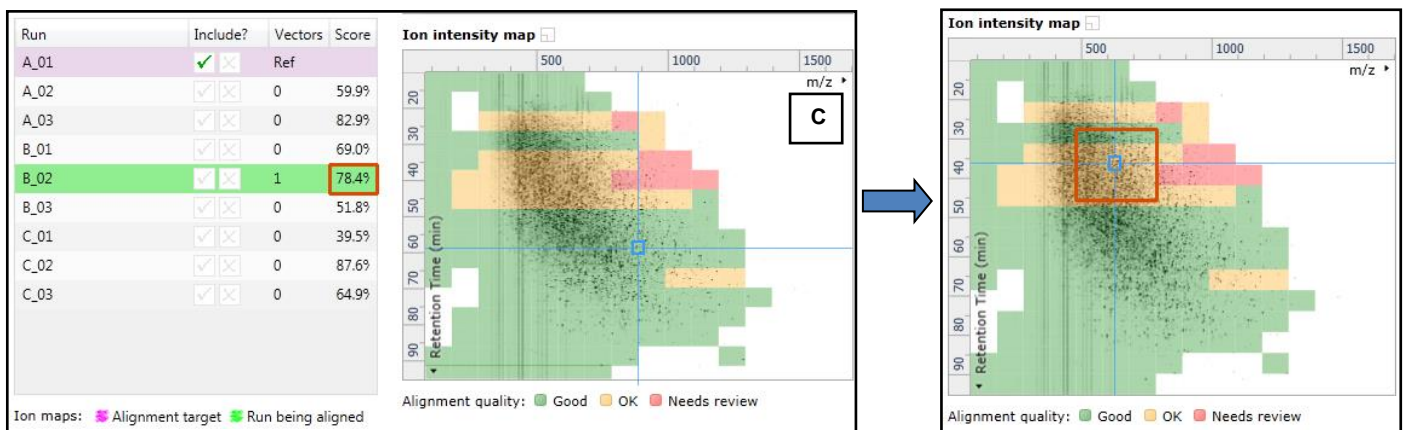


**Note:** an incorrectly placed vector is removed by right clicking on it in the **Vector Editing** window and selecting delete vector

7. Now click **Show Aligned** on the top tool bar to see the effect of adding a single vector.



8. With the placement of a single manual vector the increase in the proportion of the **Ion Intensity Map (C)** showing green is reflected in the improved alignment score in the table. Now click in the Ion Intensity Map to relocate the focus in order to place the next manual vector.



9. Adding an additional vector will improve the alignment further as shown below.



- The shift in the Retention Time (RT) is as a result of incorrect running of the chromatography. In many of these cases if the Automatic Alignment fails to generate a good alignment then removing all the alignment vectors for this run and placing a single manual vector to act as a 'seed' for the Automatic Alignment algorithm maybe all that is required to generate a good alignment.
- In the case of the example shown above placing a small number of vectors from the top to the bottom of the run is sufficient to markedly improve the alignment.

### Review Alignment

Sample ions are aligned to compensate for drifts in retention time between runs.

[Learn about the visualisations shown here](#)

- Align retention times automatically**  
For maximum reproducibility, the software can automatically align your runs.
- Review the alignment**  
Using the quality control measures, review and edit the runs' alignment:
  - Order the runs by alignment score and start by selecting the first run
  - Within each run, inspect and edit any areas rated as Needs Review

[Learn about the review and editing process](#)

Run	Include?	Vectors	Score
A_01	<input checked="" type="checkbox"/>	Ref	
A_02	<input checked="" type="checkbox"/>	0	59.9%
A_03	<input checked="" type="checkbox"/>	0	82.9%
B_01	<input checked="" type="checkbox"/>	0	69.0%
B_02	<input checked="" type="checkbox"/>	5	90.4%
B_03	<input checked="" type="checkbox"/>	0	51.8%
C_01	<input checked="" type="checkbox"/>	0	39.5%
C_02	<input checked="" type="checkbox"/>	0	87.6%
C_03	<input checked="" type="checkbox"/>	0	64.9%

Ion maps: ■ Alignment target ■ Run being aligned

Show Aligned
Show Unaligned
Remove Vectors

#### Vector editing

#### Transition

#### Ion intensity map

Alignment quality: ■ Good ■ OK ■ Needs review

#### Total ion chromatogram

- At this point you would redo the automatic alignment of this image by selecting automatic alignment. **Note:** if you are focusing only on the alignment of one run, then un-tick the other runs in the alignment dialog.

### Automatic Alignment

Choose which runs to automatically align:

Run	Notes	Vectors
<input type="checkbox"/> A_02	This run has not been automatically aligned	0
<input type="checkbox"/> A_03	This run has not been automatically aligned	0
<input type="checkbox"/> B_01	This run has not been automatically aligned	0
<input checked="" type="checkbox"/> B_02	run has user vectors	5
<input type="checkbox"/> B_03	This run has not been automatically aligned	0
<input type="checkbox"/> C_01	This run has not been automatically aligned	0
<input type="checkbox"/> C_02	This run has not been automatically aligned	0
<input type="checkbox"/> C_03	This run has not been automatically aligned	0

- On pressing OK the Automatic Alignment will run for the selected run. On completion the table and views will update to display the automatically generated vectors (shown in blue).

Waters  
THE SCIENCE OF WHAT'S POSSIBLE.™

68

14. Repeat this process for all the runs to be aligned.

The number of manual vectors that you add at this stage is dependent on the misalignment between the current run and the Reference run.

**Note:** In many cases only using the Automatic vector wizard will achieve the alignment.

**Review Alignment**  
Sample ions are aligned to compensate for drifts in retention time between runs.

[Learn about the visualisations shown here](#)

- Align retention times automatically**  
For maximum reproducibility, the software can automatically align your runs.
- Review the alignment**  
Using the quality control measures, review and edit the runs' alignment:
  - Order the runs by alignment score and start by selecting the first run
  - Within each run, inspect and edit any areas rated as Needs Review

[Learn about the review and editing process](#)

Run	Include?	Vectors	Score
A_01	<input checked="" type="checkbox"/>	Ref	
A_02	<input checked="" type="checkbox"/>	0	59.9%
A_03	<input checked="" type="checkbox"/>	0	82.9%
B_01	<input checked="" type="checkbox"/>	0	69.0%
<b>B_02</b>	<input checked="" type="checkbox"/>	<b>1841</b>	<b>95.9%</b>
B_03	<input checked="" type="checkbox"/>	0	51.8%
C_01	<input checked="" type="checkbox"/>	0	39.5%
C_02	<input checked="" type="checkbox"/>	0	87.6%
C_03	<input checked="" type="checkbox"/>	0	64.9%

Ion maps: ■ Alignment target ■ Run being aligned

Alignment quality: ■ Good ■ OK ■ Needs review

**Section Complete**

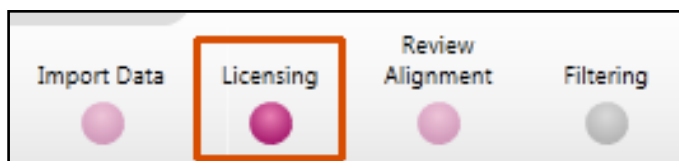
**Tip:** a normal alignment strategy would be: to run the automatic alignment first for all runs, then order the alignments based on score. For low scoring alignments remove all the vectors and place 1 to 5 manual vectors to increase the score then perform automatic alignment. Then review the improved alignment score.

Also the 'ease' of addition of vectors is dependent on the actual differences between the LC-MS runs being aligned.

To review the vectors, automatic and manual, return to page 10

## Appendix 3: Licensing runs (Stage 3)

When setting up a **New experiment** if you are evaluating Progenesis QI for proteomics with unlicensed runs then the licensing page will open after **Import Data section**.



**If you already have a programmed dongle attached to your machine then the License Runs page will not appear.**

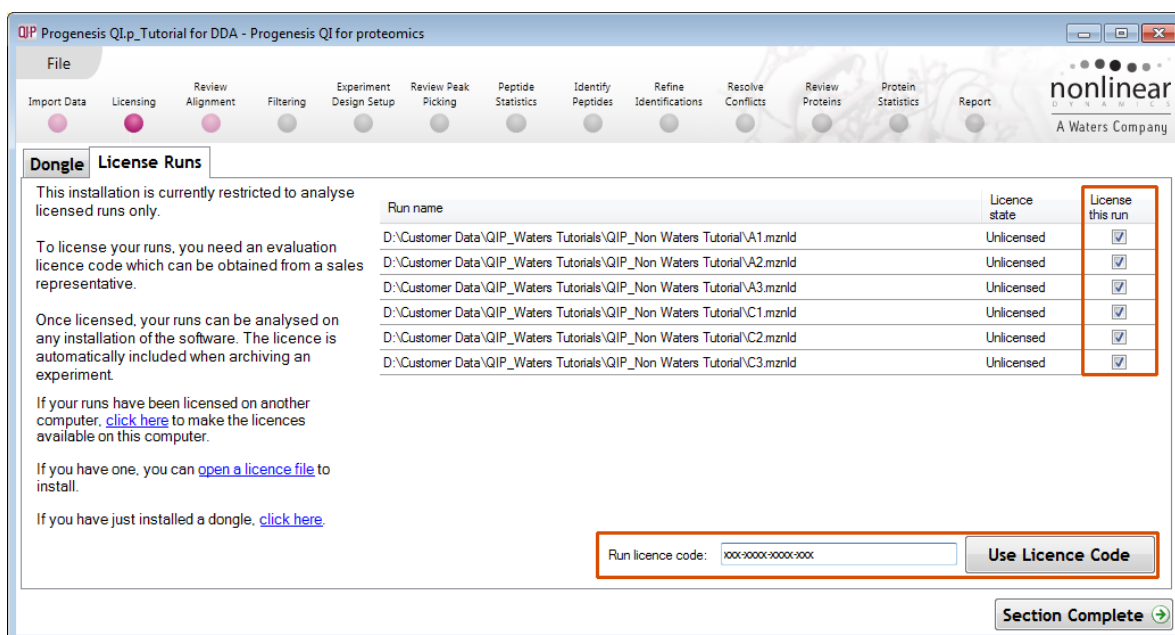
To use this page to License your Runs you must first either obtain an **'Evaluation' Licence Code** from a **Sales Person** or **purchase a licence code directly**.

Each code will allow you to license a set number of runs.

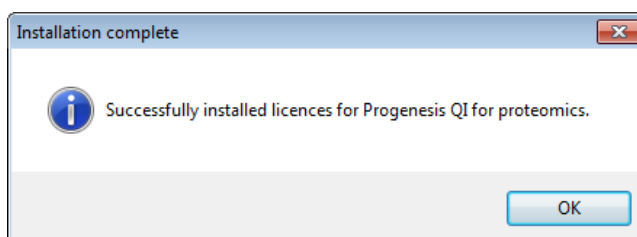
The runs in your experiment will be listed as shown below.

To activate license(s) for the selected runs enter the code in the space provided and click **Use Licence code**.

**Note:** you will need an internet connection to use this method.



A message confirming successful installation of your licences will appear.



Click **OK**, the view will update and Alignment, the next stage in the workflow, will open with the licensed files.

## Appendix 4: Within-subject Design

To create a **Within-subject Design** for your data set select this option on the **Experiment Design Setup** page and enter the name of the design.

In this example there are 3 Subjects (i.e. patients A, B and C) who have been individually sampled: Before(1), During (2) and After (3) treatment

Which experiment design type do you want to use for this experiment?

Between-subject Design

Do samples from a given subject appear in only one condition? Then use the between-subject design.

To set up this design, you simply group the runs according to the condition (factor level) of the samples. The ANOVA calculation assumes that the conditions are independent and therefore gives a statistical test of whether the means of the conditions are all equal.

**Within-subject Design**

Have you taken samples from a given subject under different conditions? Then use the within-subject design.

**Create New Experiment Design**

Enter a name for the experiment design:

How do you want to group the runs?  
 Group the runs manually  
 Copy an existing design:

	Before	During	After
Patient X	X1	X2	X3
Patient Y	Y1	Y2	Y3
Patient Z	Z1	Z2	Z3

A standard ANOVA is not appropriate because the data violates the ANOVA assumption of independence. With a repeated measures ANOVA individual differences can be eliminated or reduced as a source of between condition differences (which helps to create a more powerful test).

The within-subject design can be thought of as an extension of the paired-samples t-test to include comparison between more than two repeated measures.

When the design page opens use the **Add Subject** and **Add Condition** buttons to create the matrix that fits your experimental design, over typing the names as required.





## Appendix 5: Power Analysis (Peptide Stats)

Power analysis is a statistical technique that is used to gauge how many replicates are needed to reliably see expression differences in your data. It is available through the Peptide Stats section of the workflow.

To perform a power analysis of the data click on **Ask another question** at the top of the table in the Peptide Stats screen. A selection of 3 tools will appear in the form of questions.

**Principal Components Analysis**  
Are there any outliers in my data?  
Does my data cluster according to my experimental conditions?

**Correlation Analysis**  
Group my features according to how similar their expression profiles are.

**Power Analysis**  
How many replicates should I run?  
What is the power of my experiment?

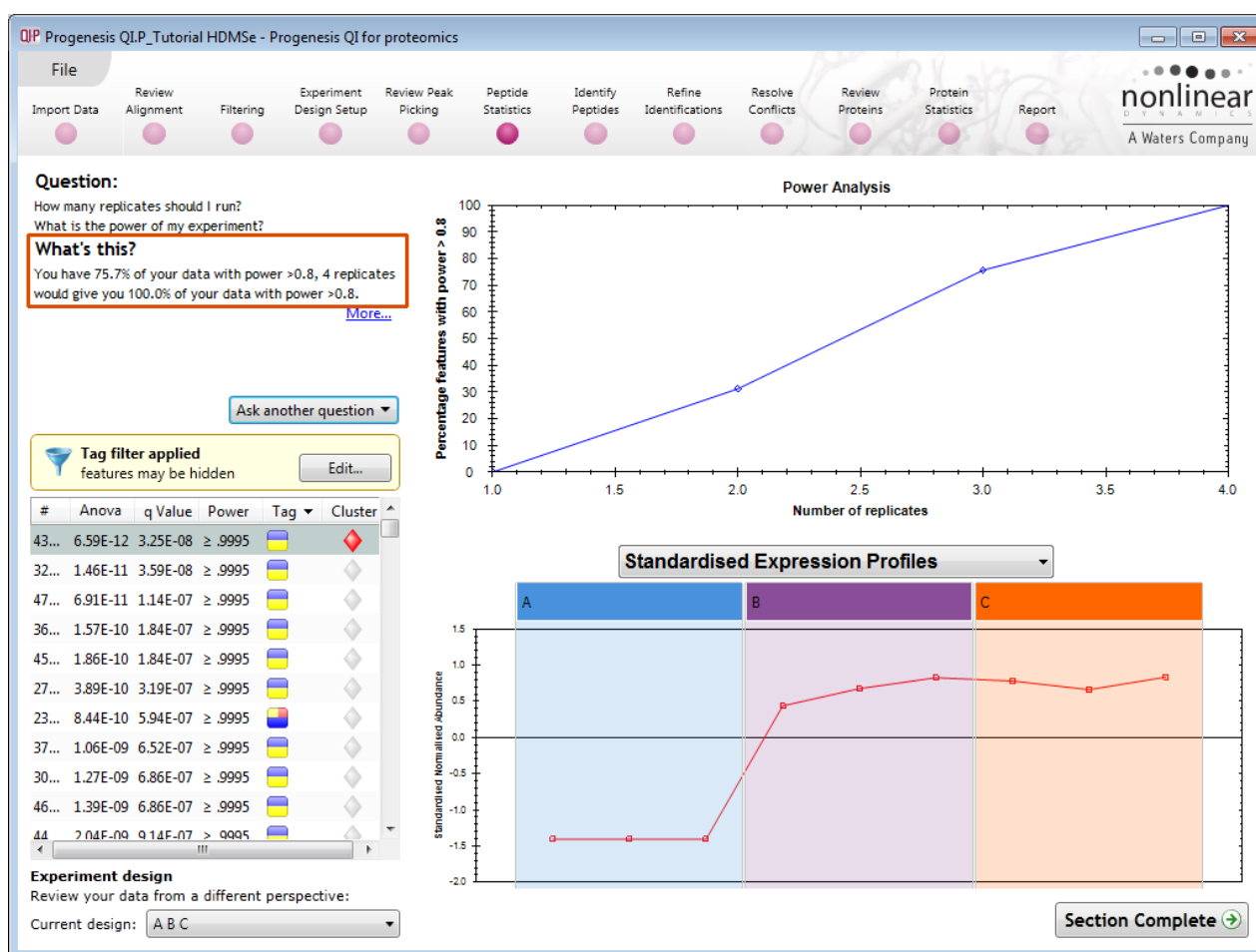
Select the option

*'How many replicates should I run and what is the power of my experiment?'*

It answers this question by informing you:

*'How many replicates you need so that at least 80% of your features with a power >0.8'*

Using the **Significant p<0.05 features (21118)**, as an example, view the power analysis.



This is displayed graphically showing that 75.7% of the 21118 features have a power of 80% or that 4 replicates would give you 100% of your data with power > 0.8.

- **The power of a statistical test** reflects our confidence in the experimental data's ability to find the differences that do actually exist
- The power is expressed as a percentage, where 80% power is an accepted level, therefore allowing you to assess the number of sample replicates that would be required to achieve a power of 80%.

## Appendix 6: Waters Machine Specification

This appendix provides information on the approximate time(s) taken at each stage and the total time taken to analyse a set of 9 (Phase 1) HDMSe runs on a Waters Demo Spec PC.

### Machine Spec: Lenovo

Processor: Intel® Xeon® CPU 2.66GHz 12core X5650 @ 2.67GHz

RAM: 24.0 GB

System Type: 64-bit Operating System

**File Folder Size:** Each file folder (.RAW): 40.9 Gig

<b>Analysis Stages:</b>		<b>Per file</b>	<b>Total</b>	
<b>Import Data:</b>	Loading of Raw data per file	12min	2hr 6min	Total for 9 files
	Apex Background processing (re-opening at Import Data)	1hr 05min(max)	9hr 45min 20s	Total for 9 files
	<b>Alignment:</b>	Automatic alignment of data (re-opening at Alignment)	8min 30s 10s	
<b>Peak Detection:</b>	Automatic Detection of data (re-opening at Peak Detection)		14min 15s 10s	
	<b>Identify Peptides:</b>	Performing MS <sup>E</sup> Search (re-opening at Identify Peptides)		14min 30s 10s
<b>Total Analysis Time</b>		Excluding Background Apex Processing		2hr 42min
	Including Apex processing assuming pause for Apex			9hr 58min
<b>Restoring Tutorial_A&gt;Loading HDMSe.TOIP Experiment from archive</b>				4min