nonlinear dynamics

Progenesis LC-MS Tutorial

Including Data File Import, Alignment, Filtering, Progenesis Stats and Protein ID

Introduction

This tutorial takes you through a complete analysis of 9 LC-MS runs (3 replicate runs per sample) using the unique Progenesis LC-MS workflow. It starts with LC-MS data file loading then Alignment, followed by Analysis that creates a list of interesting features (peptides) which are explored within Progenesis Stats using multivariate statistical methods then onto Protein identity and Reporting.

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. However, the document covers all the stages in the LC-MS workflow, therefore if you are using your own data files refer to Appendix 1 then start at page 5.

How to use this document

You can print this tutorial to help you work hands-on with the software. The complete tutorial takes about 50 minutes and is divided into two sections. 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. If you experience any problems or require assistance, please contact us at **download@nonlinear.com**

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

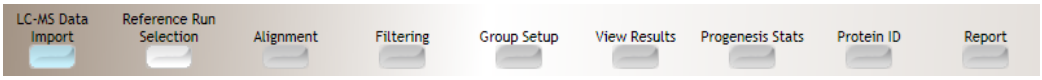
You can freely explore the quality of your LC-MS data using Image QC and then licence your own LC-MS runs using this evaluation copy of Progenesis LC-MS. Instructions on how to do this are included in a section at the end of the tutorial document. Alternatively if you would like to arrange a demonstration in your own laboratory contact **download@nonlinear.com** and we will help you.

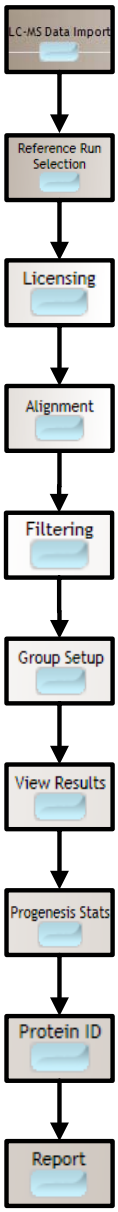
LC-MS Data used in this tutorial

NLD would like to thank Dr Robert Parker and Prof Haroun Shah at the Health Protection Agency, London, UK for providing the example data used in this tutorial as well as invaluable discussion on the handling of the data.

Workflow approach to LC-MS run analysis

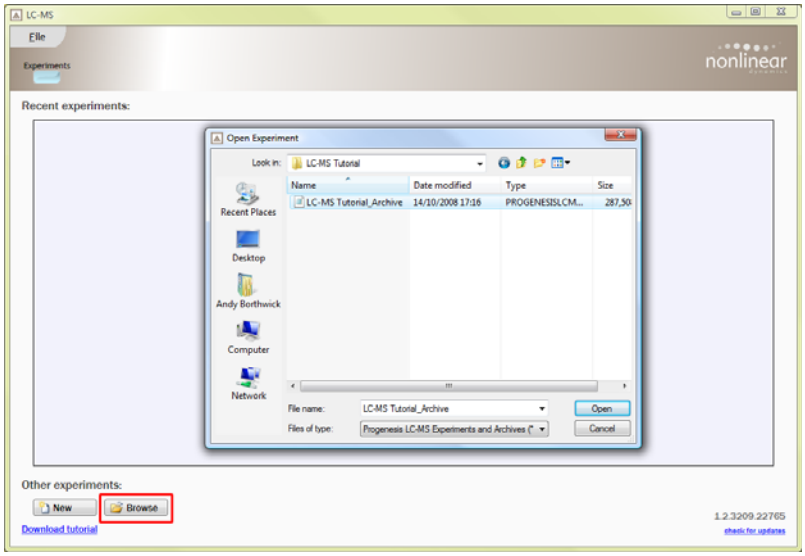
Progenesis LC-MS adopts an intuitive 'Workflow' approach to performing comparative LC-MS data analysis. The following tutorial describes the various stages of this workflow (see below) focusing mainly on the stages from Alignment to Reporting.



Stage	Description	Page
	**LC-MS Data Import:** Selection and review of data files for analysis.	4-5
	Reference Run Selection: Select run to align to.	6
	Licensing: appears when using unlicensed images or when there is no dongle attached (Appendix 2)	6
	Alignment: automatic and manual run alignment	7-15
	Filtering: defining filters for features based on Retention Time and m/z as well as Charge state and Number of Isotopes.	16-18
	Group Setup: defining one or more group set ups for analysed aligned runs	19-20
	View Results: review and validate results, edit feature detection, tag groups of features and select features for further analysis	20-28
	Progenesis Stats: performing multivariate statistical analysis on tagged and selected groups	28-36
	Protein ID: manages protein ID searches	36-38
	Report: generate a report on selected features	38-40

Restoring the LC-MS Tutorial

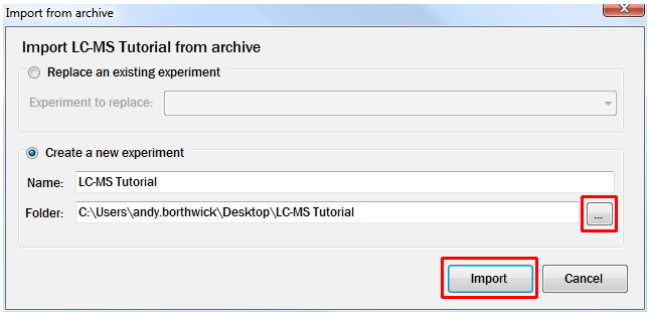
To restore the LC-MS tutorial from the 'Downloaded' Archive file, open LC-MS and locate the LC-MS Tutorial Archive file using the **Browse** function.



This opens the Import from Archive dialog. Set the folder option to the same folder you have just created on your Desktop.

Select the **Create a new experiment** option and select the folder in which you placed the archive.

Then press **Import.**

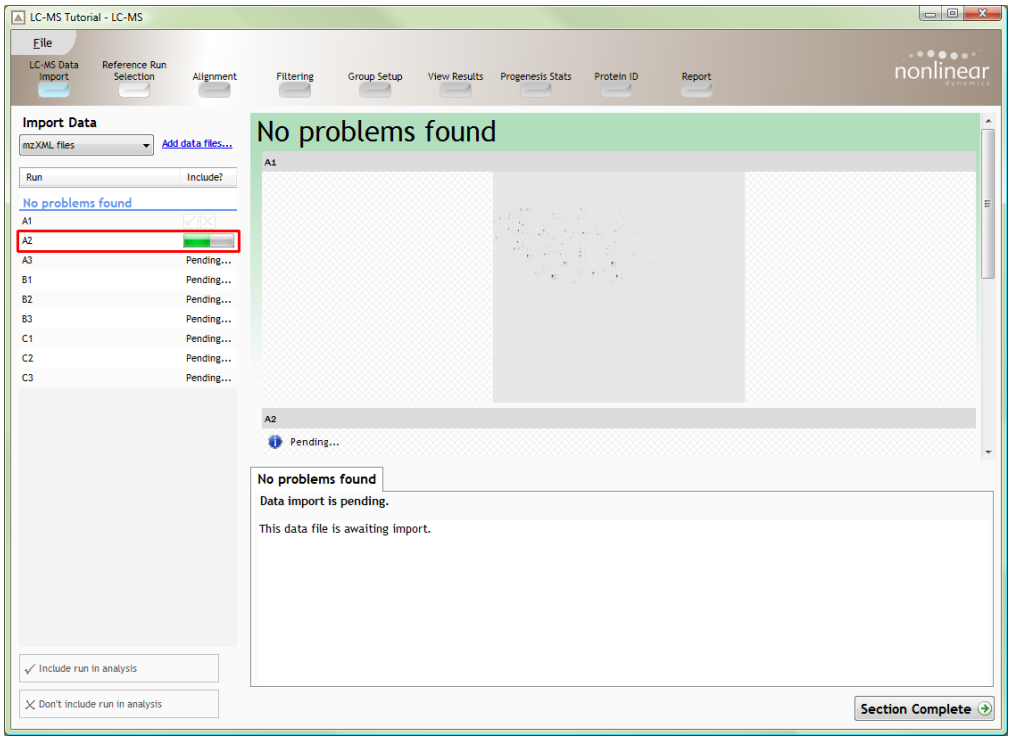


Note: use the **Replace an existing experiment** option if you want to over-write an existing version of the tutorial.

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

The LC-MS tutorial will now open at the LC-MS Data Import stage (see below).

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.

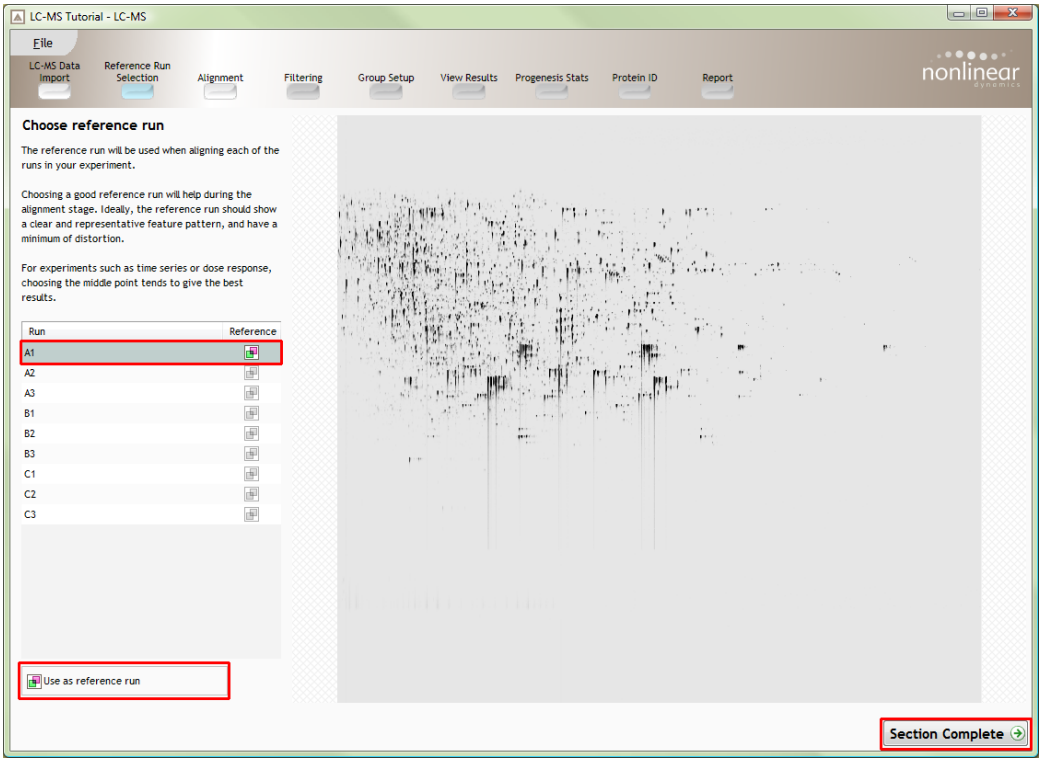


Note: as each data file is loaded the progress is reported in the **Import Data** list. The dialog below the image reports on the QC of the imported Data files. In this case 'No problems found' with the this data file.

Now move to the next stage in the workflow by clicking **Section Complete.**

Stage 2: Reference data file selection

This stage in the analysis workflow allows you to review and select the most appropriate Reference LC-MS run to align all the runs to.



To select a Reference run either click on the run in the list and then click **Use as reference run** or double click on the run in the list.

Now move to the next stage in the workflow by clicking **Section Complete.**

Stage 3: Licensing

This stage in the analysis workflow will **always** appear in the LC-MS workflow if you are using 'Unlicensed' data files to evaluate the software or have no dongle attached.

For details on how to use Licensing go to Appendix 2 (page 43)

Stage 4: Alignment

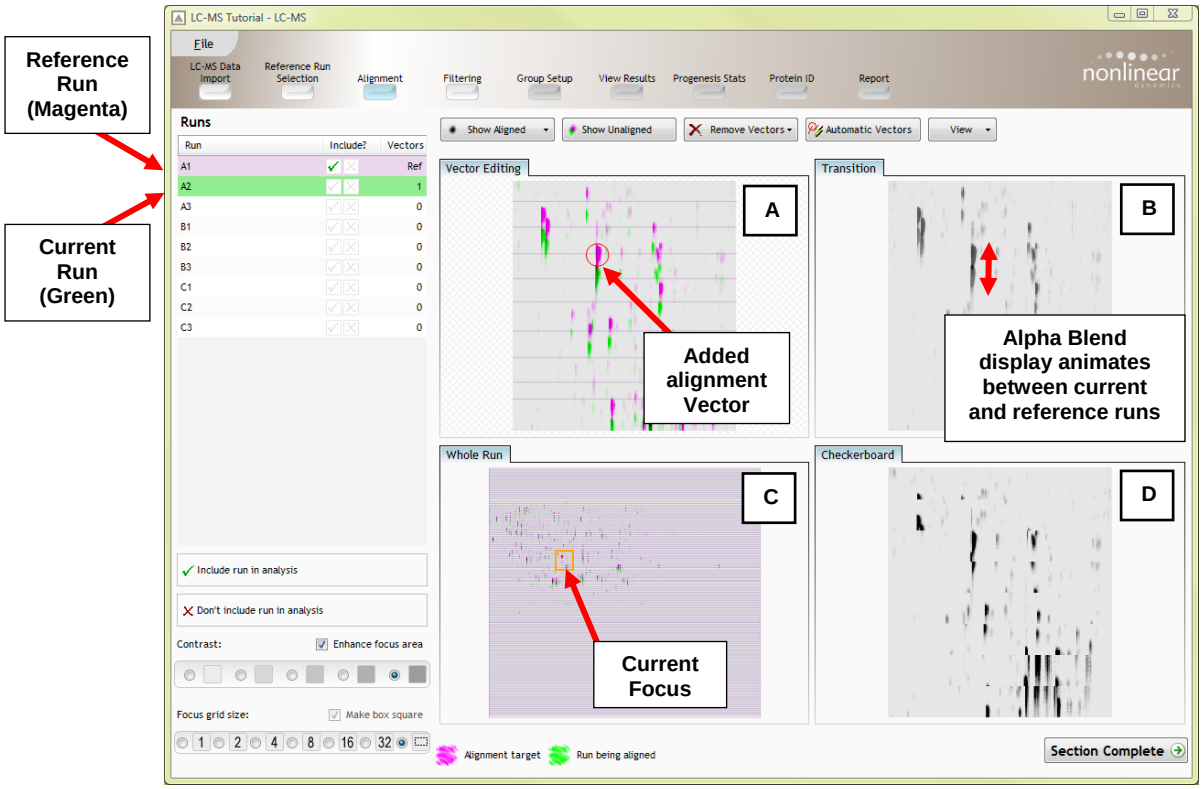
At this stage Progenesis LC-MS Alignment opens displaying your data.

The Program Layout

To familiarize you with Progenesis LC-MS 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 below:

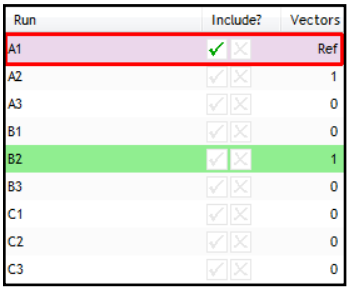
- Click on the features shown in the current focus (orange rectangle) in Window C, this will update windows A,B and D as shown below.
- In window A **click and hold** the left mouse button on a green feature, it can auto align.
- If the green and magenta spot (immediately above) have not aligned automatically then **drag** the green feature over the magenta feature and **release** the mouse button.
- The image will 'bounce' back and a red vector, starting in the green feature and finishing in the magenta feature as a circle will now appear as shown below in window A .



The experiment structure is displayed on the left of the screen in the **Image** panel based on the information used in the setup described in Stage 1.

Runs: panel shows the run that is currently being aligned in green, and the run it is being aligned to in magenta.

The **Ref** run for any experiment is the run that you chose to align all the runs to, in this case **A1** highlighted in magenta.



Vector Editing (Window A): is the main alignment area and shows the current **focus** rectangle as shown in Window C. The current image is displayed in green and the chosen reference image is displayed in magenta. Here is where you place the alignment vectors. Retention time alignment is performed in the vertical dimension

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 back and forwards. After alignment, they will appear to pulse if correctly aligned. During the process of adding vectors, this view will change to show a zoomed view of the area being aligned to help accurate placement.

Whole Data File (Window C): shows the **focus** for the other windows. When you click on the view the orange rectangle will move to the selected area. The focus can be moved systematically across the view using the left and right cursor keys. The focus grid size can also be altered using the controls in the bottom left of the screen.

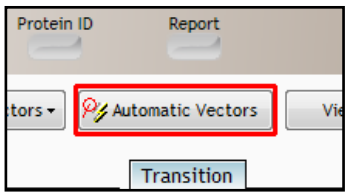
Checkerboard (Window D): shows a checkerboard view of the current run interleaved with the reference run. When the area is aligned the edges of the squares start to merge.

The unique use of the alpha blend and checkerboard views helps to make highly accurate alignment of feature borders more obvious to the human eye.

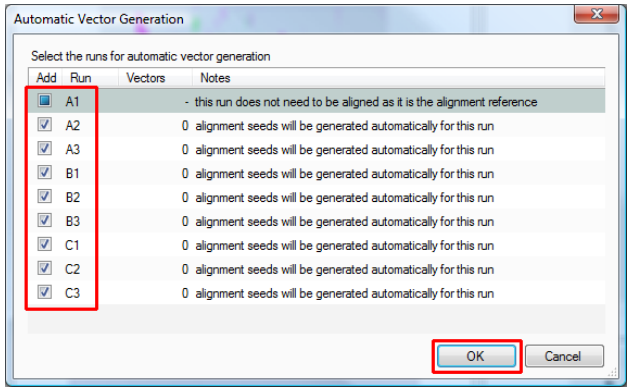
Generation of alignment vectors

The alignment of LC-MS runs is required in the LC (retention time) direction, this is key to correcting for the variable elution of peptides during the chromatographic separation.

The Alignment algorithm at this stage in the workflow will generate 'Automatic' vectors, in the retention time direction for each run, to enable the alignment of all the LC-MS runs to the 'Reference Run' that was chosen in the previous section.



The alignment vectors are generated automatically for all the LC-MS runs by using the 'Automatic vector wizard' accessed by clicking on **Automatic Vectors** on the top tool bar.



Select (tick) the runs you require to generate vectors for and click **OK**.

Alternatively, in some cases, using a combination of a 'few' manually placed vectors on each run and then using the Automatic vector wizard to generate the rest of the vectors for each run can be used depending on the severity of the misalignment.

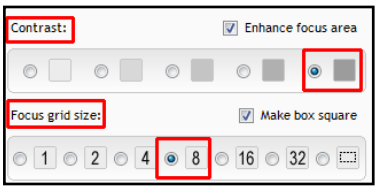
In this example try placing some manual vectors before generating the automatic vectors.

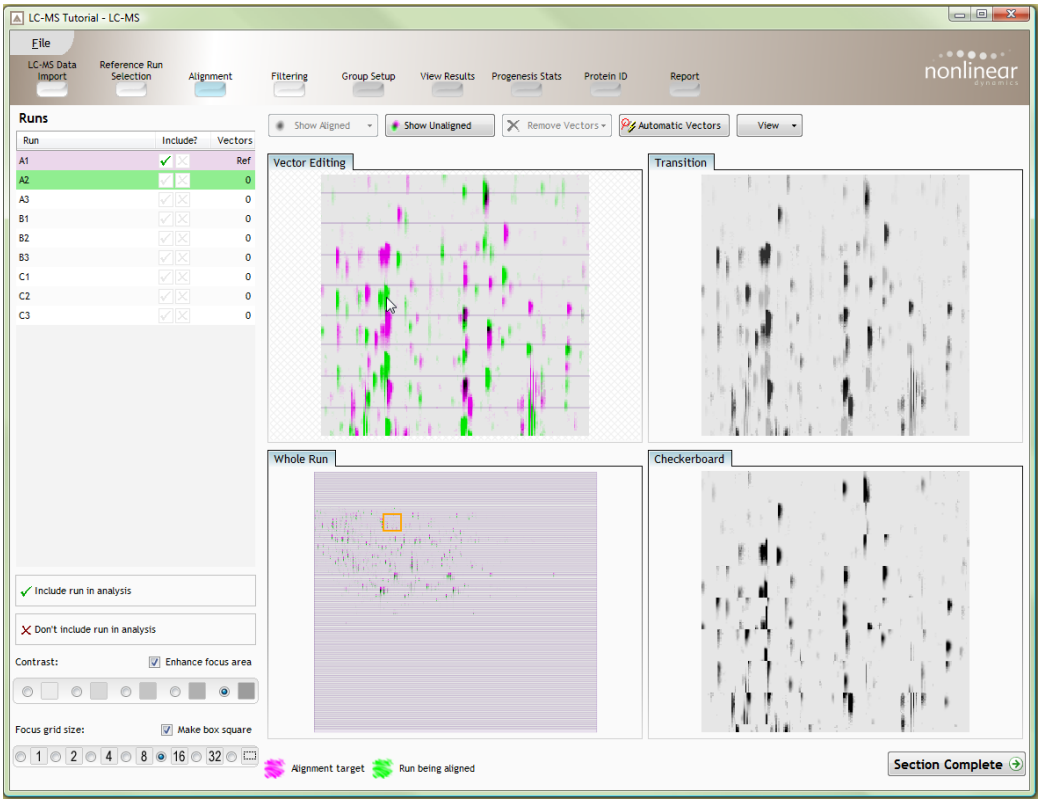
The following section describes how to manage the placement of manual vectors on your LC-MS runs

Approach to alignment

To place manual alignment vectors on your current run (A2 in this example):

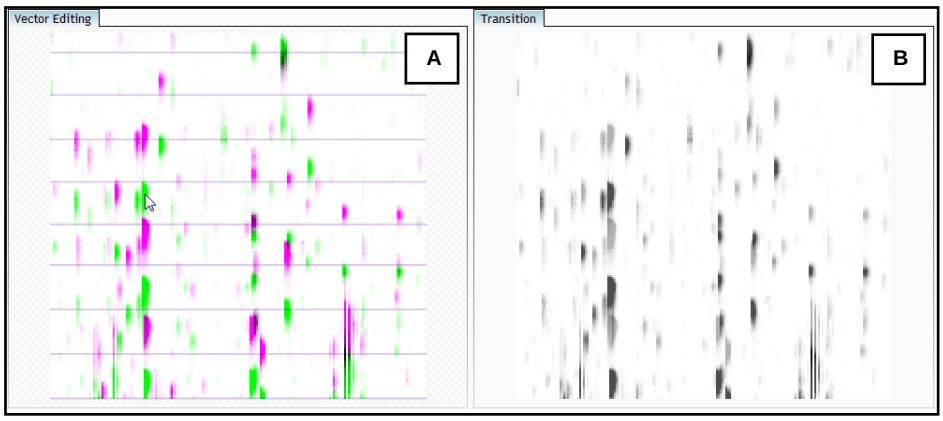
1. Click on Run A2 in the **Runs** panel, this will be highlighted in green and the reference run (A1) will be highlighted in magenta.
2. You will need approximately five **alignment vectors** evenly distributed from top to bottom of the whole run.
3. First ensure that the size of the focus area is set to **8 or 16** in the Focus grid size on the bottom left of the screen.
4. Click on an area (see below) in the **Whole Run** window (C) to refocus all the windows. Adjust Contrast as required.



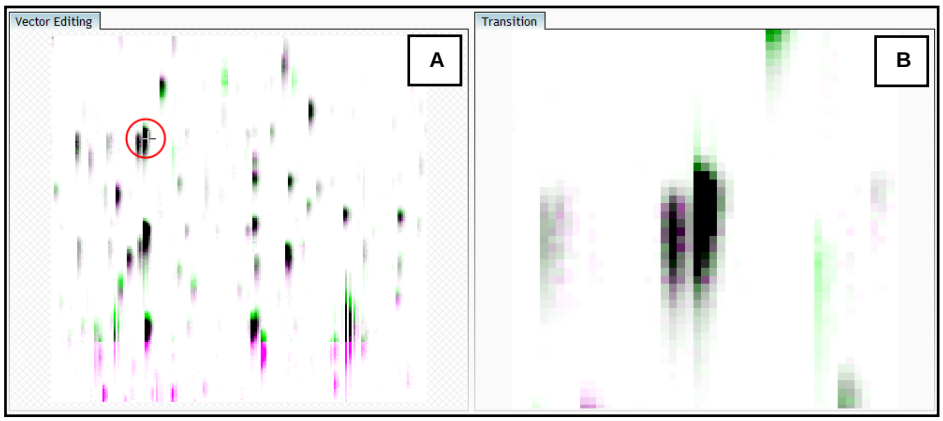


Note: the features moving back and forwards between the 2 runs in the **Transition** view indicating the misalignment of the two LC-MS runs

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

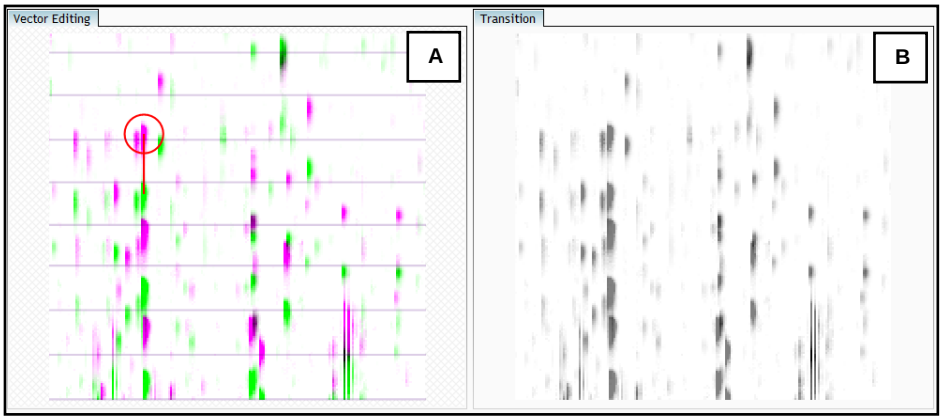


6. As you are holding down the left mouse button drag the green feature over the corresponding magenta feature of the reference image. The vector will appear as shown below as a red circle with a 'cross hair' indicating that a positional lock has been found for the overlapping features.



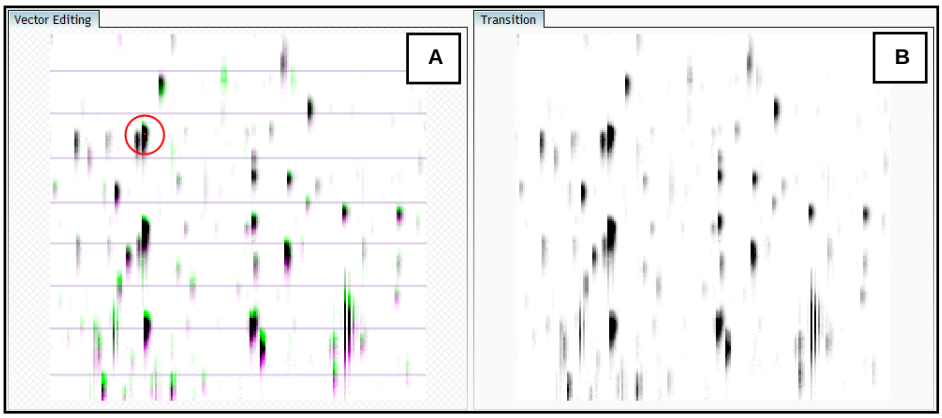
Note: as you hold down the mouse button, window B zooms in to help with the alignment.

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

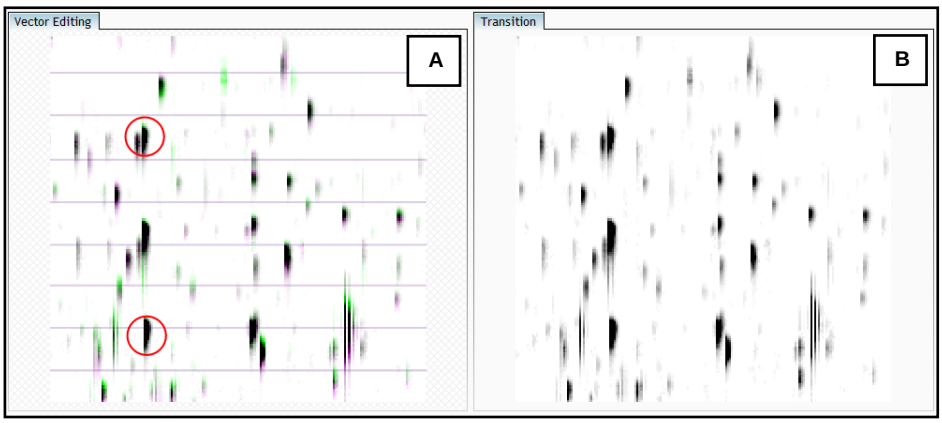


Note: an incorrectly placed vector is removed by right clicking on it in the **Vector Editing** window

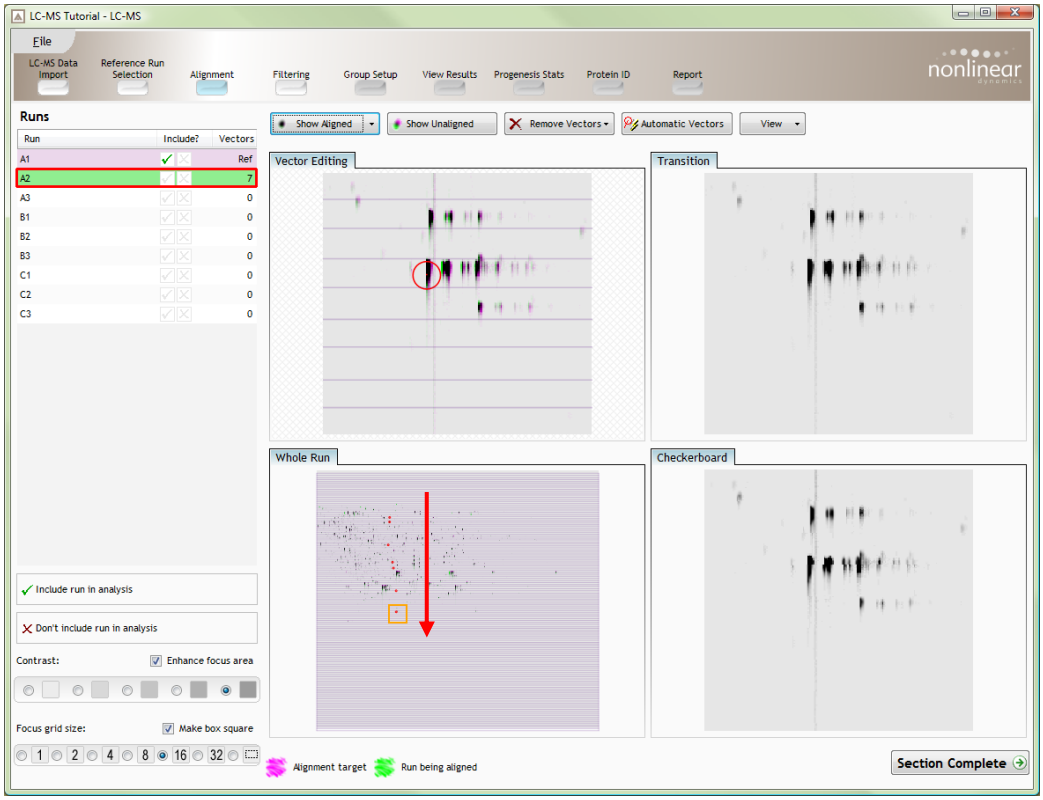
8. Now click Show Aligned on the top tool bar to see the effect of adding a single vector.



9. Additing an additional vector will improve the alignment further. **Note** this time as you click to add the vector it 'jumps' automatically to the correct position using the information from the existing alignment vector



10. Repeat this process moving the focus from top to bottom on the **Whole Run** view



Note: the number of vectors you add is recorded in the **Runs** table

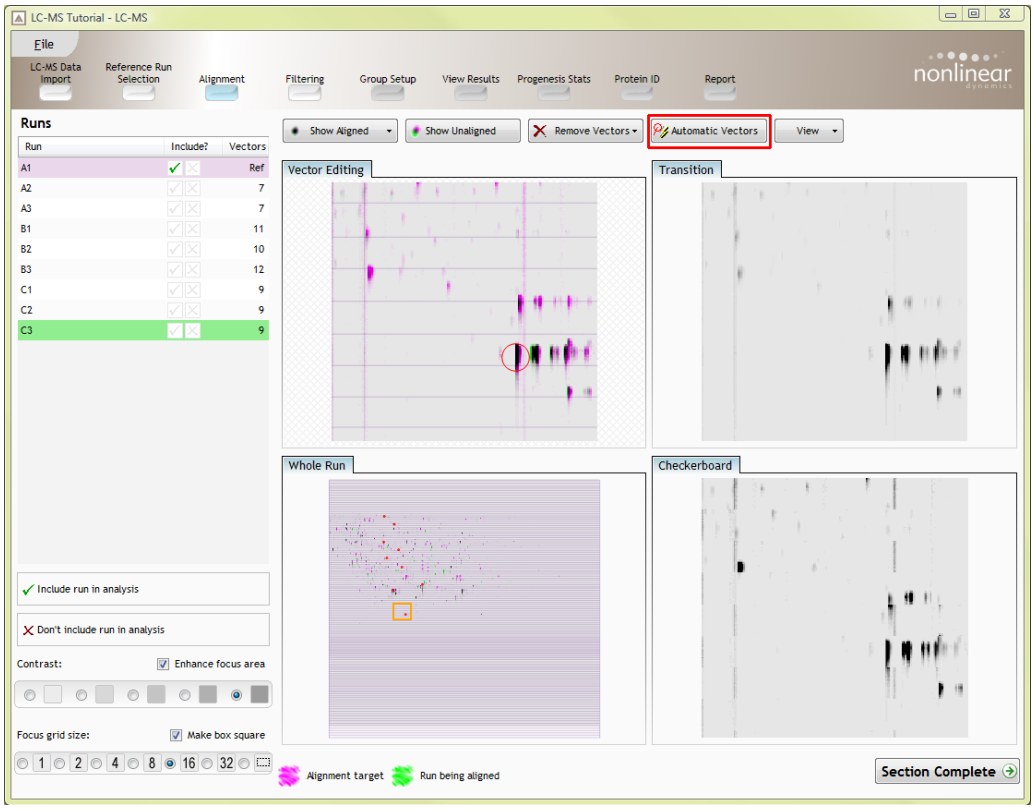
11. Now move on to the next run to align and repeat the addition of a few manual vectors

The number of manual vectors that you add at this stage is dependant on the misalignment between the current run and the Reference run. In a number of cases only using the Automatic vector wizard will achieve the alignment.

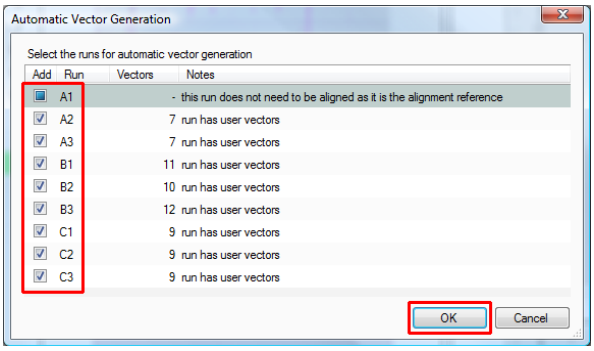
Also the ‘ease’ of addition of vectors is dependant on the actual differences between the LC-MS runs being aligned

In this tutorial the ‘B’ sample replicates are markedly different from the A and C samples so this will require that you make full use of the Transition Window to aid in the correct positioning of your initial manual alignment vectors

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



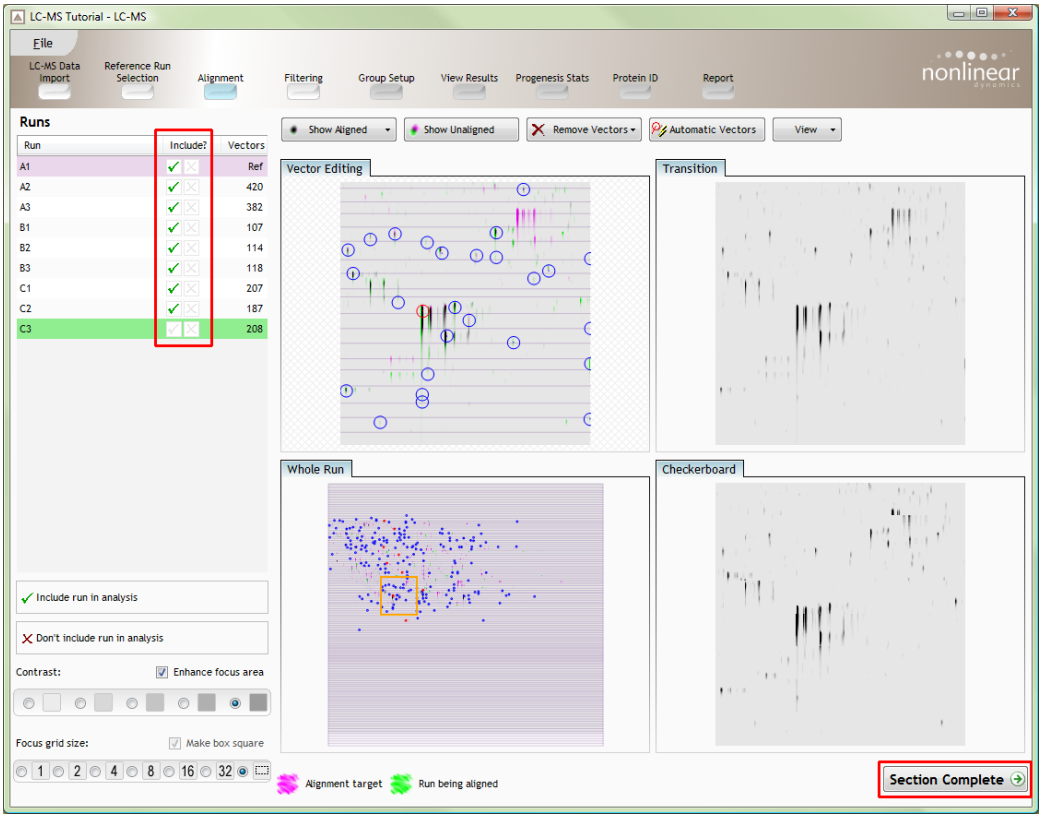
13. Then select Automatic vectors and click **OK**.



Note: the tick boxes next to the 'Run name control' which indicates the runs to generate vectors for.

On completion, the number of vectors will be updated on the **Runs** panel and the vectors will appear (in blue) on the image.

If the alignment has worked well then in Windows A and C the grid lines should show minimal distortion, Window B will show features pulsing slightly but with no movement of the features as it moves between the images.

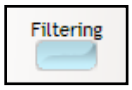


At this point, you should check the automatically placed (blue) vectors. This will be easier with a larger grid size. Make sure the grid size is set to 4 using the **'Focus grid size'** control at the bottom left of the window.

In each square, you can, if required edit the vectors to improve the image alignment.

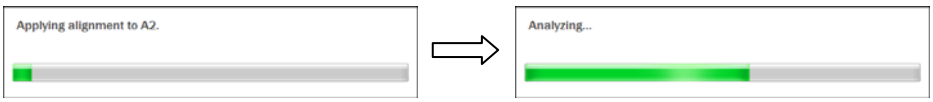
To indicate that alignment of the run is complete and ready for Analysis click **Include Run in analysis** located underneath the image list.

Stage 5: Filtering

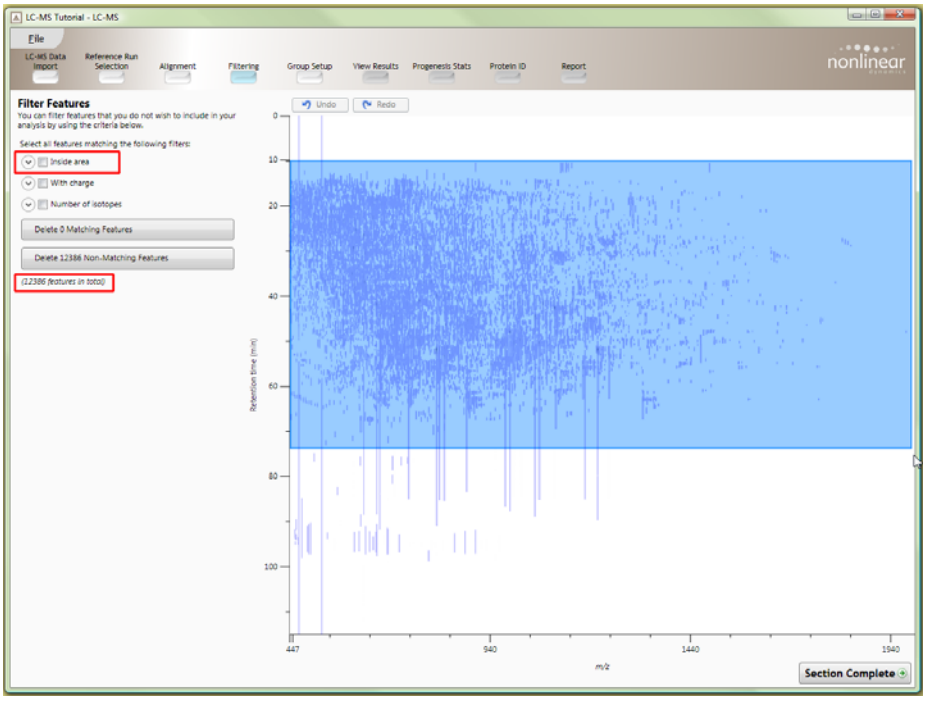


Now that you have reviewed your aligned Runs, you are now ready to analyse them. Move to the next stage, **Filtering**, by either clicking on **Section Complete** (bottom right) or on Filtering on the workflow.

During the few minutes that the automatic analysis requires, a progress bar will appear telling you first that it is applying alignment to the Runs and then secondly that it is Analysing.

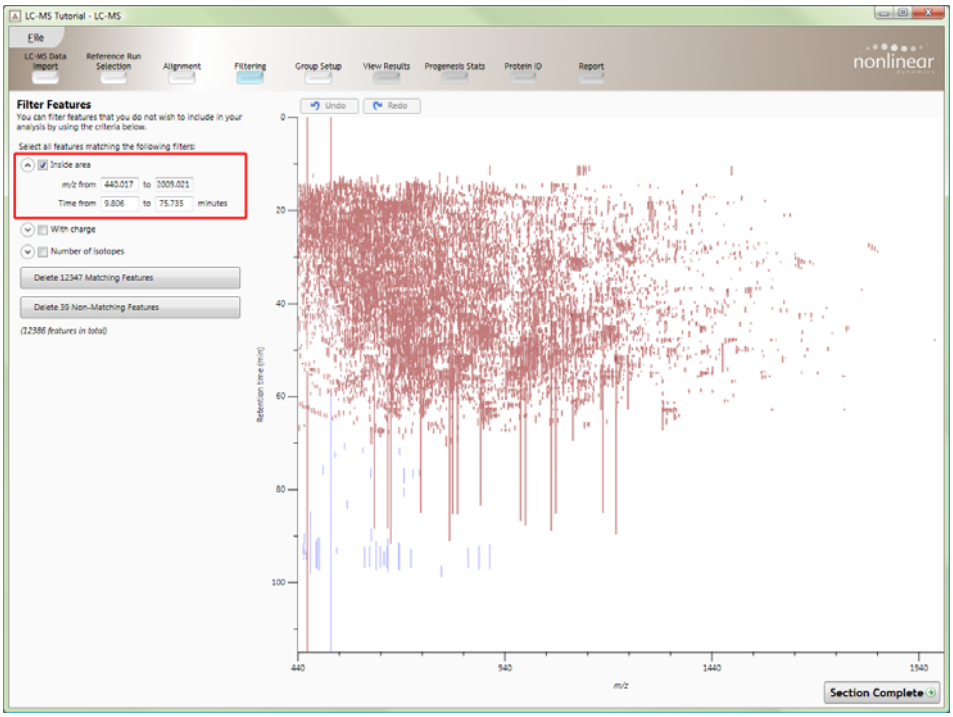


On completion of analysis the Filtering stage will open displaying the number of features detected (12386 as shown on the left). If required you can remove features based on position, charge state, number of isotopes and combinations of these feature properties.



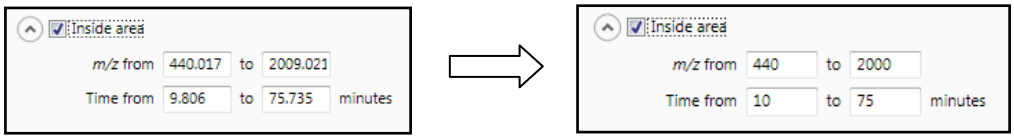
For example, to delete features with early and late ‘Retention times’ drag out an area as shown.

All features with their reported ‘Retention Time and m/z’ values contained within the mask will be selected.



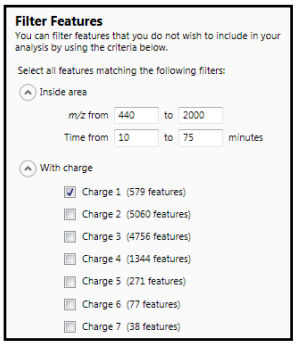
As you release the mouse button the values for the masked area will appear on the top left

Note: the limits can be adjusted by entering the required values in the boxes provided



To remove the 39 features press the **Delete 39 Non-Matching Features** button

In addition to setting limits for 'Retention time and m/z' for feature selection, features can also be selected on the basis of charge as well as the number of isotopes present. Thus allowing you to refine the selection through a combination of feature properties

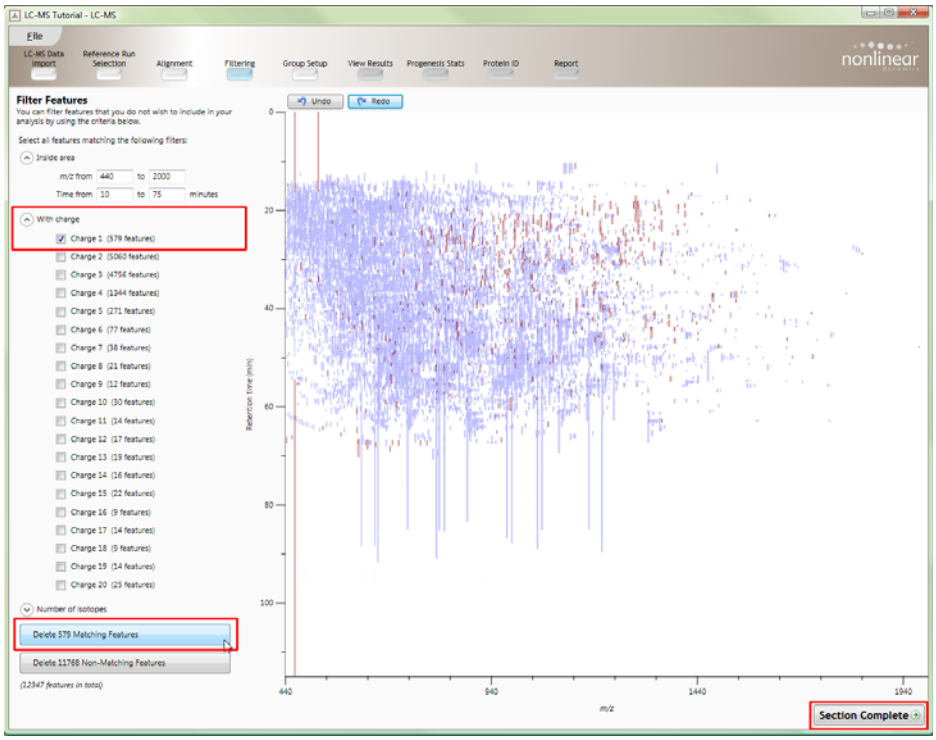


For example: when charge state is selected the number of features present at each charge state is displayed, these can be selected accordingly.

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

For this tutorial, we will filter the area as shown and remove features with a charge state of 1.

Having already delete 39 features outside the retention time and m/z limits we will now delete a further 579 with a charge state of 1 by ticking the ‘Charge State 1’ option.



Hence all features with a charge state of 1 will appear red (see above).

To remove the features with ‘Charge State 1’ press **Delete 579 Non-Matching Features**

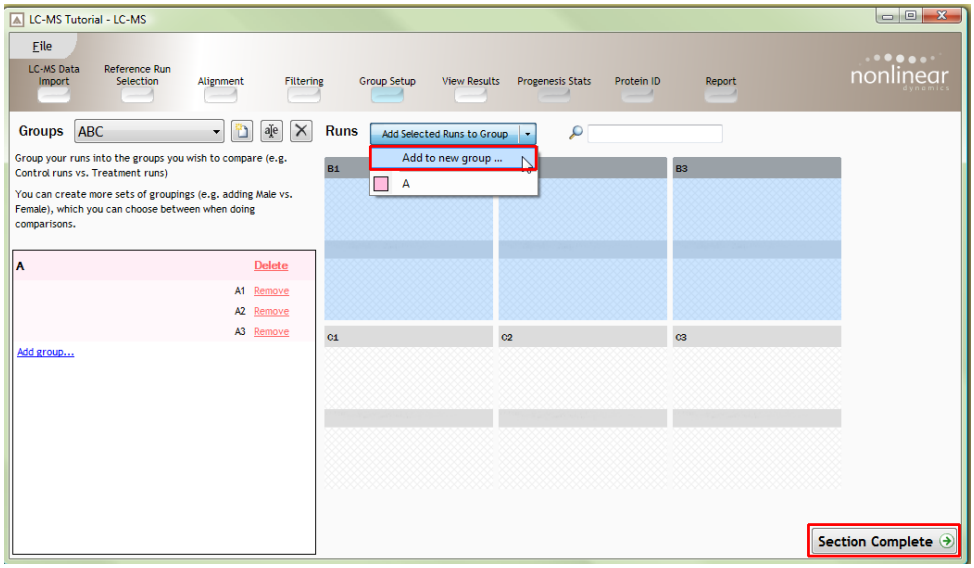
You can use the **Undo** button to recover from a filter, however, when you move to the next section you will lose the capacity to undo the filter.

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

Stage 6: Group Setup for Analysed LC-MS runs

At this stage in the workflow you can setup one or more groupings of your sample data.

For this example, group the analysed LC-MS runs to reflect the Biological groupings in the original study. This tutorial contains 3 groups: A, B, and C with 3 replicates each.

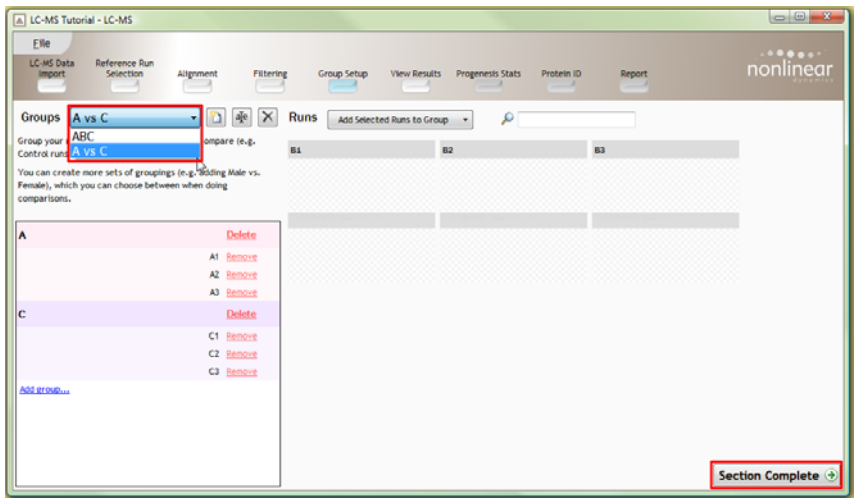


Creating a group

1. Select the Runs in a group (based on names at top of image) by clicking on the required thumbnails (Highlighted as above)
2. Press the 'black triangle' next to the **Add Selected Runs to Group** button on the main toolbar.
3. Select **Add to new group…** from the **drop down menu**.
4. A new group will appear in the **Groups** pane on the left panel
5. Rename the group by over typing the new group name (e.g. B)
6. Repeat steps 1 to 5 until all the Runs are grouped.

To create another Group Setup, for example comparing only the A and C groups, click on Create a new group setup (see right).

Give it a new name (i.e. A vs C). The Runs will reappear in the main window. Create the new groups as described on the previous page.

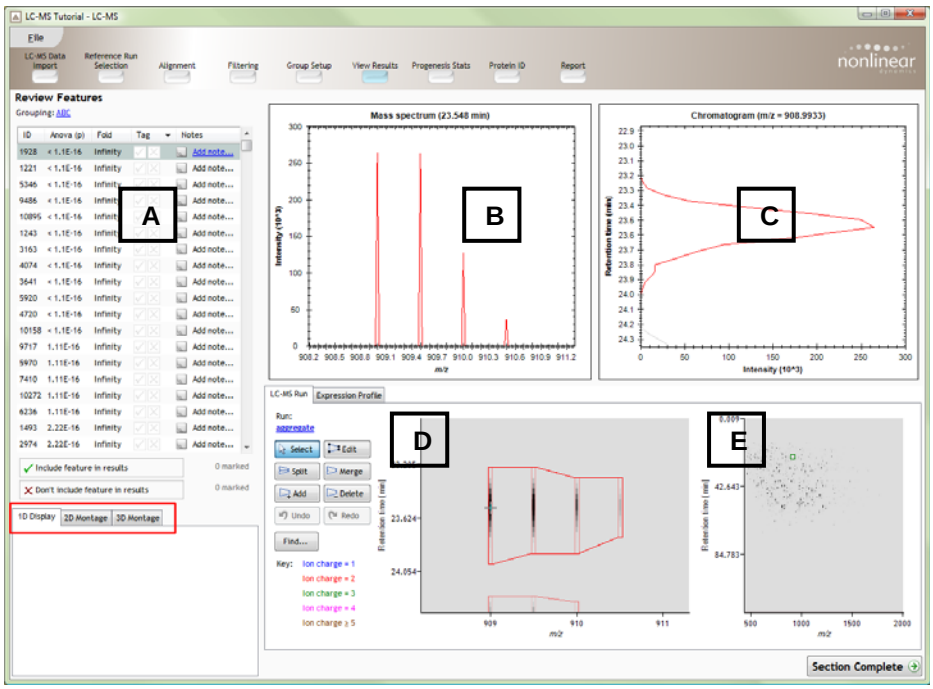


Note: the **Group set up** drop down will now contain both set ups and the ungrouped Data files (B) for the new setup will remain in the main window.

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

Stage 7: Validation, review and editing of results

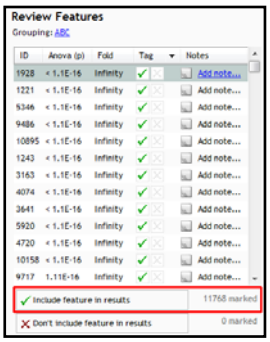
The purpose of this stage in the Workflow is to review the statistically ranked (by Anova values) list of features, for the current Group Setup (ABC is selected here) using the visual tools provided and edit features as required.



The review stage has 3 display modes: 1D, 2D and 3D controlled by the tabs on the bottom left of the display. 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

The 1D Display

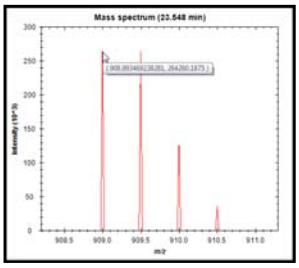


Window A: shows a ranked view of the features based on the p value for the one way **Anova** analysis of the specified sample groups across all the groups being compared.

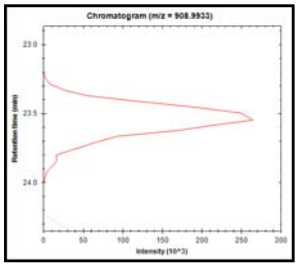
Note: A value of 'Infinity' in the **Fold** column indicates 'Presence/Absence'

To include a feature in the selection for the next section of the analysis, click on the **Include features in results** button at the bottom of the table. On clicking the button it will move on to the next feature on the list.

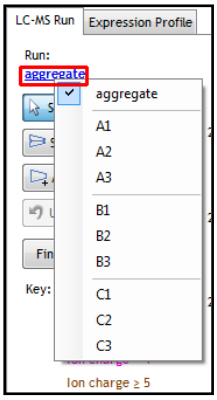
To select a group of features drag out a selection on the table and click on the **Include feature in results** button (see right)



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).

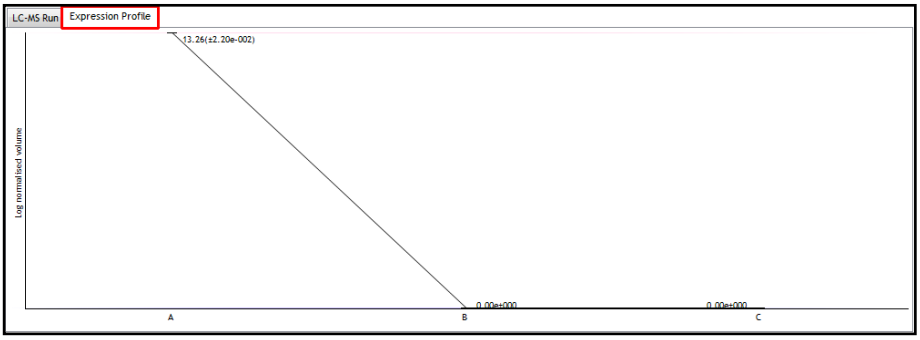


Window D: displays the details of the LC-MS currently selected run. By default the information is displayed as an Aggregate for the feature across all the aligned runs.

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 25 for functional explanation).

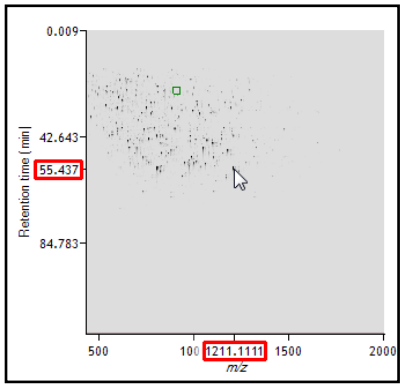
Clicking on the Expression Profile tab in Window D shows the comparative behaviour of the feature across the various biological groups based on average normalised volume for the groups where the error is shown as +/- 3 standard deviations.



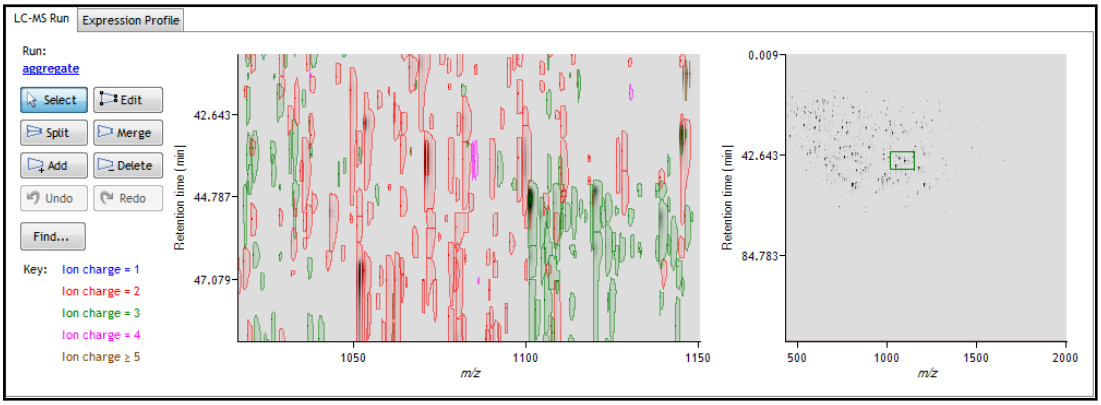
Window E: shows where the current feature is located on the LC-MS run by means of the 'Green' rectangle.

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

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



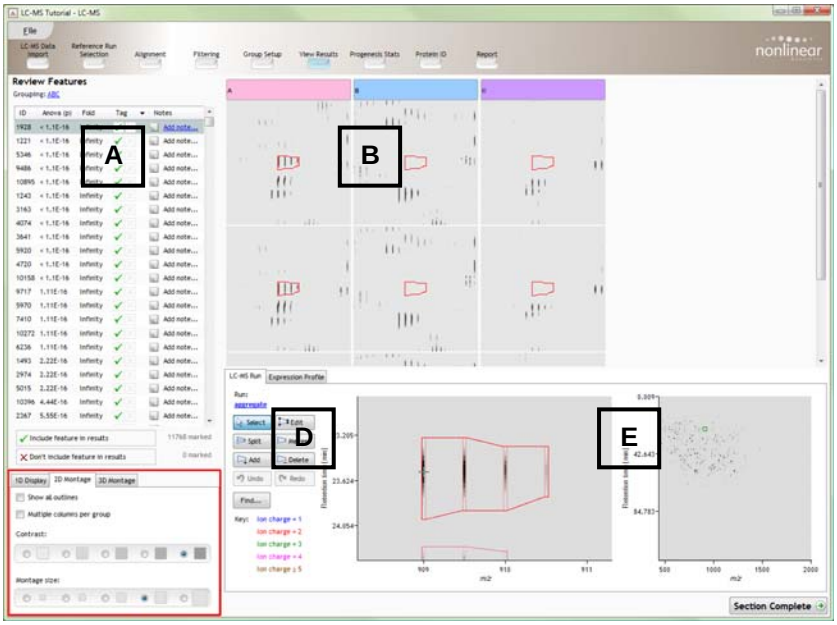
You can also drag out a larger area on this view that will refocus the other windows



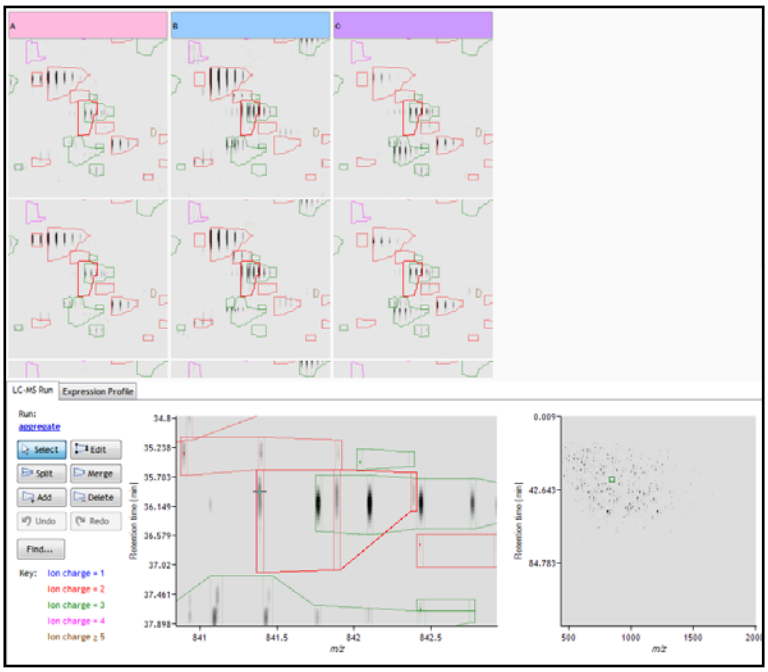
The 2D Display

Windows A, D and E: perform the same functions across all 3 display modes.

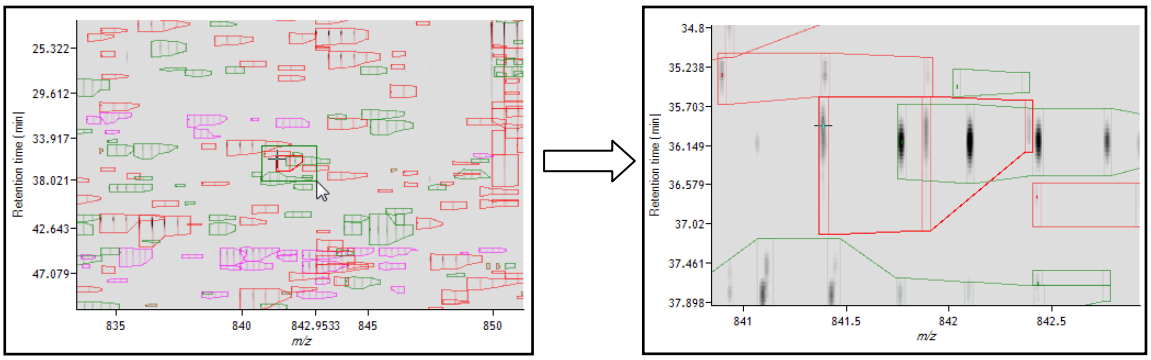
In the 2D display, Window B displays a montage of the current feature across all the aligned LC-MS runs with an Average expression profile available on the Expression Profile tab of Window D



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 at the bottom of the previous page.

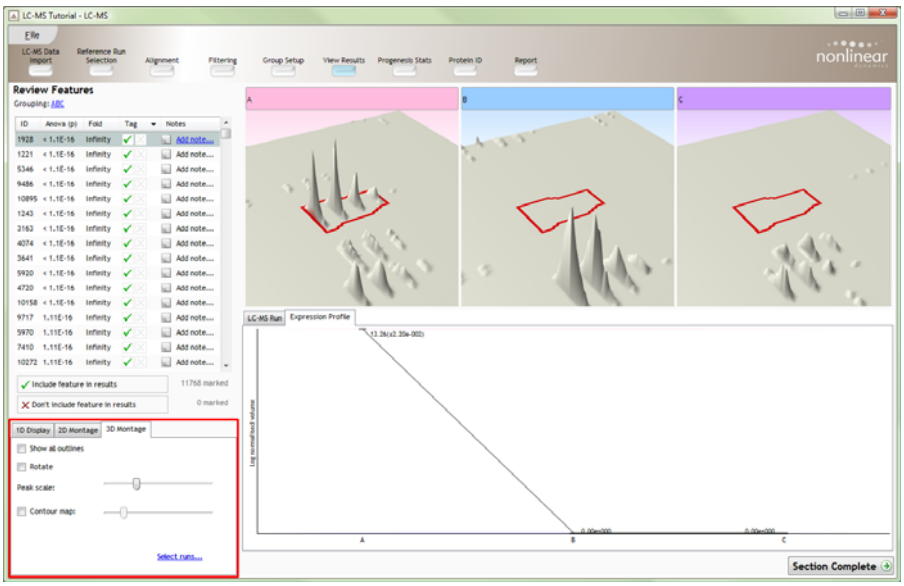


Drag out an area to zoom in to the view. To zoom out, right click on the view one or more times.

Note: when you click on a feature in the table the area displayed will return to default settings

The 3D Display

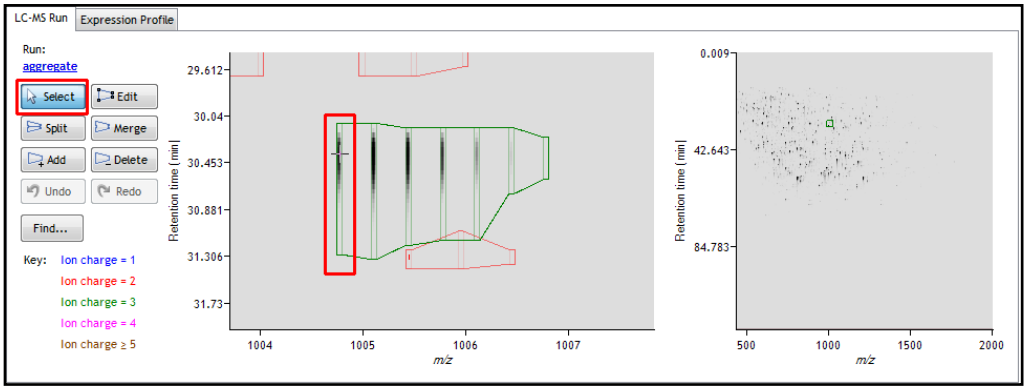
Window B changes into a 3D view by selecting the 3D tab on the bottom left of the display.



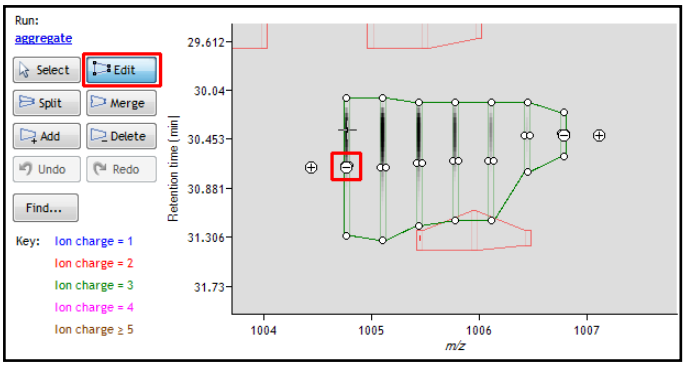
The number of 3D views displayed in the montage is controlled using the Select runs link on the 3D tab. Also the images can be set to **Rotate** automatically or you can rotate them manually by dragging the cursor over them.

Editing of features in the View Results stage

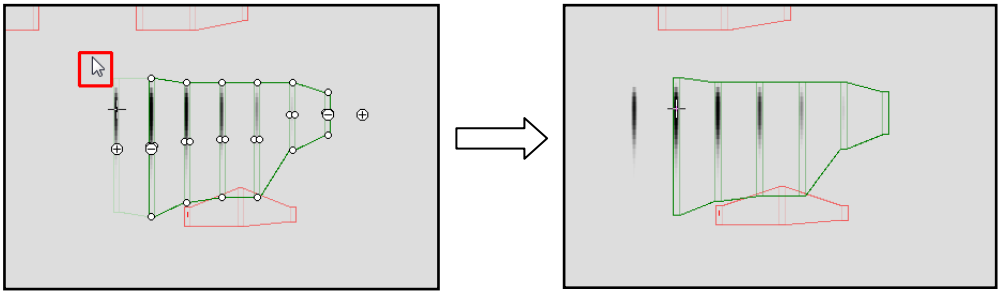
As an example of using the editing tools which are located on the left of the LC-MS Run view, we will remove and add back the 'monoisotopic peak' for the detected feature selected below. A feature can be selected from the 'Features' list or located using the views on the LC-MS tab.



1. Select a feature (286 in this example).
2. Select the **Edit** tool and click on the feature to reveal the 'edit handles'

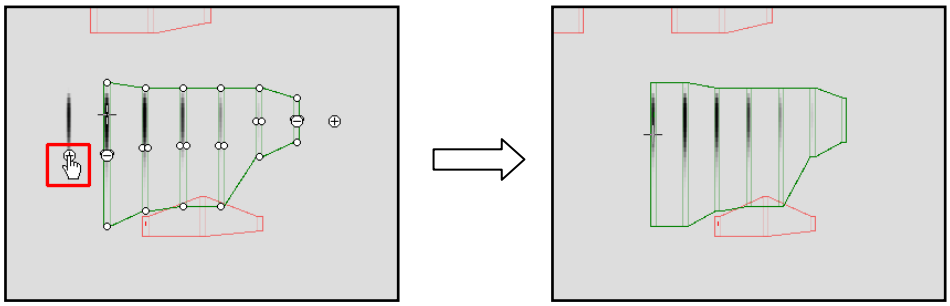


3. Click on the 'minus' handle over the monoisotopic peak to remove it.
4. Click outside the boundary of the feature to update the view.



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

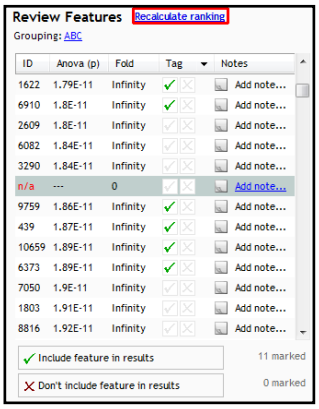
6. Click on the 'plus' handle on the peak to add it.



7. Then click outside the feature to update the view.

8. **Note**: If you are not satisfied with the editing use the **Undo** function and retry.

9. Finally to re-rank the new or edited features click on **Recalculate ranking** at top of the table.

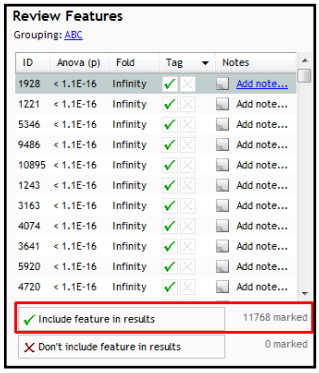


The other tools: split, merge, add and delete and add behave in a similar fashion and their use can be combined to achieve the desired editing of the selected features.

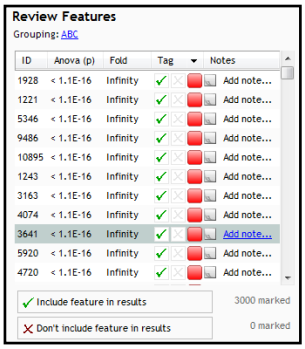
Selecting and tagging features for Progenesis Stats

There are a number of ways to 'refine' your 'Ranked List' of analysed features before examining them with the Statistical tools in Progenesis Stats. These make use of simple 'Selection' and 'Tagging' tools that can be applied to the various Groupings created in Stage 6 (p 19). Three examples are described below.

Note: the same population of features is available to all Group Setups that you have created. In this example we have two Group Setups (ABC and A vs C) each applied to the same 11768 features.



1. The simplest approach is to **select** a number of Anova ranked features and take them forward into Progenesis Stats (in this example 3000).

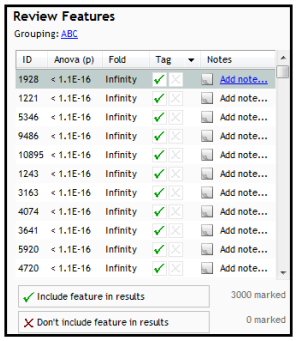


2. Alternatively select a group of features based on simple statistical thresholds; such as all the features with an Anova p value<0.05 and displaying an infinite fold change in expression and take these selected features into Progenesis Stats.

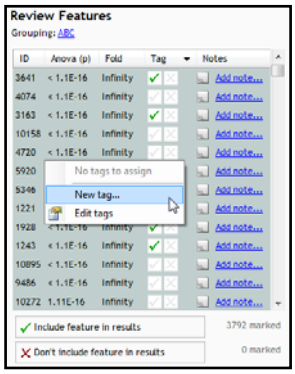
3. Or use a **combined approach**; incorporating 'Tagging' of interesting groups of features for one or more **Group set ups** combined with the selection of a number of features to take them forward to Progenesis Stats. Then make use of the 'Tags' to control the features of interest.

Setting of tags for ranked data

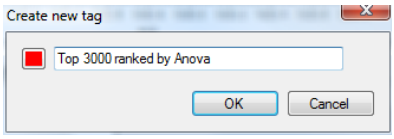
To set a Tag as shown in example 3 above, first make sure that the Grouping is set to **ABC** and the table is ordered on Anova (as shown to the right).



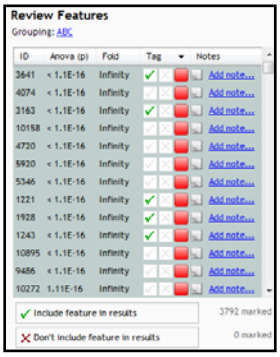
Now highlight the top 3000 ranked features for Group Set up **ABC**, right click on the highlighted features and click New Tag.



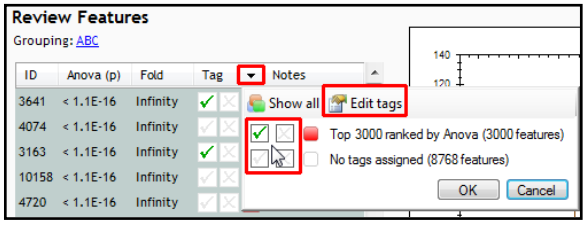
Give the Tag a name. i.e **'Top 3000 ranked by Anova'**.



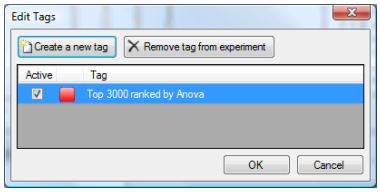
On clicking **OK** the Tag is applied to the features highlighted in the table (signified by a coloured square).



To view the Tags and also control the number of features displayed in the table, click on the drop down selection on the right of the Tag column header. By ticking one or more Tags you can control the number of features displayed.



To delete and/or create additional Tags click on **Edit tags** and Create/Remove Tags as required.



Before moving to the next section select all the features in the table by clicking on a row in the table then click **Ctrl A** and then press **Don't include feature in results** 2 times. This will deselect all features.

Now order the table by 'ID' by clicking on the 'ID' column header. Highlight up to feature 4000 and click **Include Peak in results.**

Due to the filtering you performed in the previous section this will mark 3792 features for further analysis.

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

Stage 8: Multivariate Statistics on Selected Features

The tutorial displays the **full** functionality of the Multivariate Statistics, however, only the PCA analysis functionality is available in the Standard release of Progenesis LCMS unless Progenesis Stats is licensed.

For this tutorial we start by examining the behaviour of the 3792 selected features from the previous stage, **View Results**, then making use of the Tags and Group Setups to explore specific groups of features that meet chosen statistical thresholds.

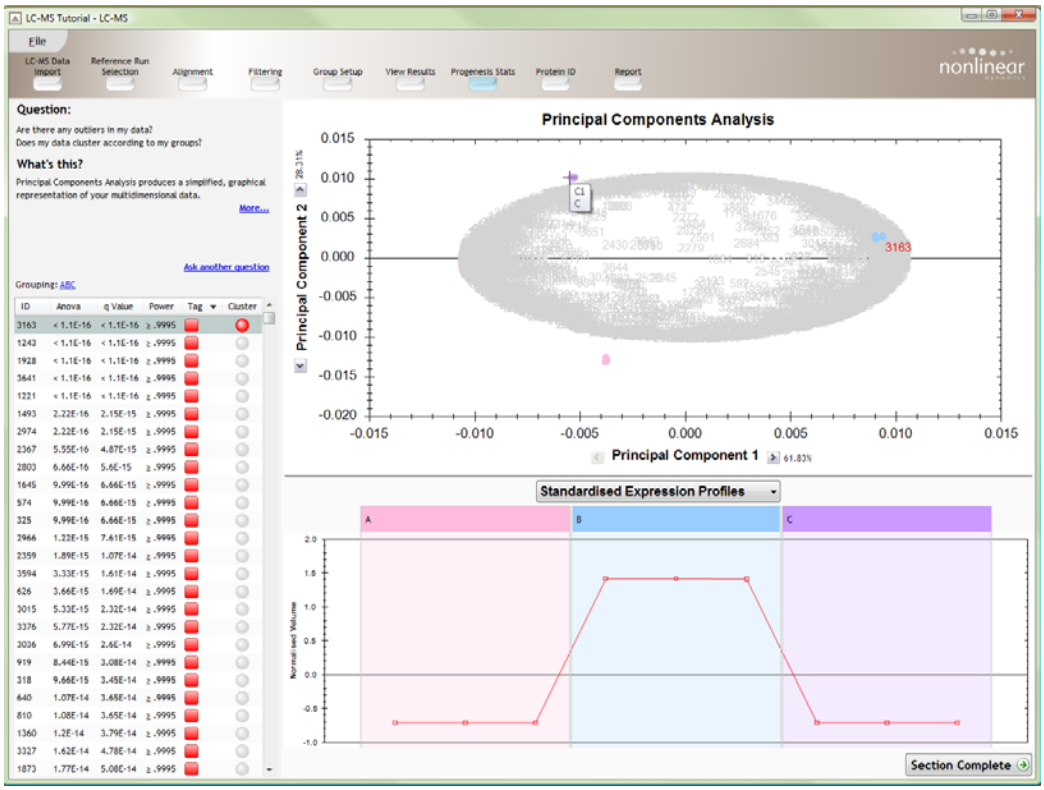
The statistical analysis of the selected data is presented to you in the form of interactive graphical representations of answers to questions asked of the analysed data.

For example: on opening **Progenesis Stats** a statistically based question asked of the application is:

Are there any outliers in my data? And does my data cluster according to my groups?

It answers this question by:

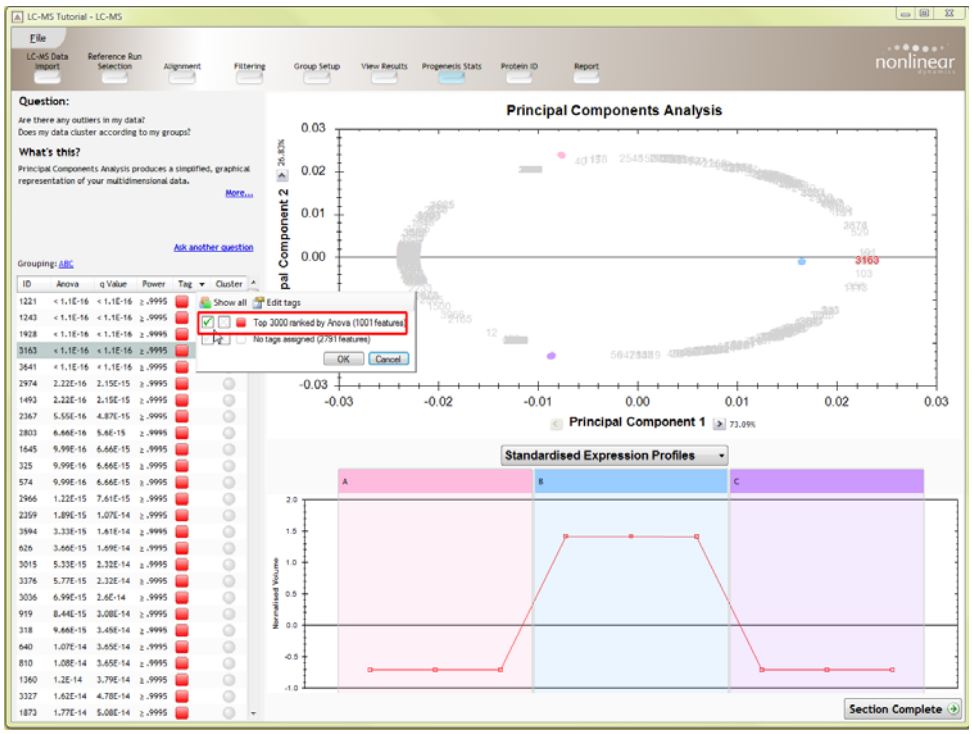
'Using Principal Components Analysis (PCA) to produce a simplified graphical representation of your multidimensional data'.



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

Principal Component Analysis (PCA)

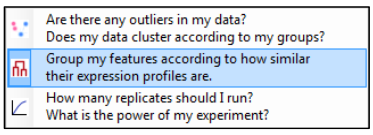
PCA can be used to determine whether there are any outliers in the data and also look at how well the samples group. The groupings that can be observed on the 2D PCA plot can be compared to your experimental design groupings and conclusions can be drawn regarding possible outliers in your original groupings.



Use the tags to filter the features being used to generate the PCA plot. In this example using 1001 of the top 3000 Anova ranked features (present in the top 3000 ranked on ID) this now shows a tighter plot of the samples on the PCA biplot.

To perform the Correlation Analysis click on the link 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'

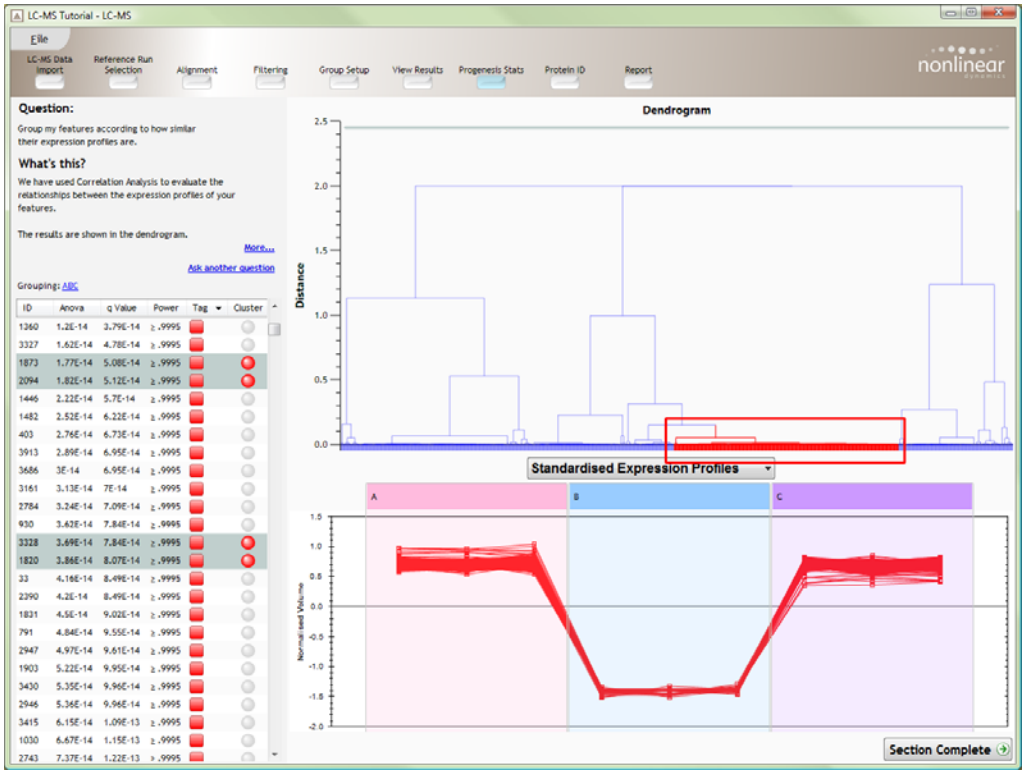
This time the statistically based question(s) being asked is:

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

It answers this question 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 distance, horizontally and vertically, between each feature can be taken as a measure of how similar the expression profiles of each feature are.



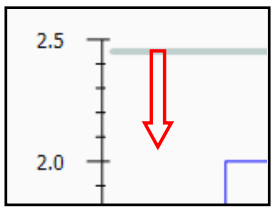
Correlation Analysis enables the grouping of features together according to how similar their expression profiles are.

Features with a high correlation value (i.e. close to 1) show similar expression profiles while features which a high negative correlation value (i.e. close to -1) show opposing expression profiles.

Clicking on a ‘node’ in the dendrogram causes the features under that node to become selected in the table. Multiple nodes can be selected by holding down **Ctrl** and clicking on the required nodes in the dendrogram as shown below.

Additional features can be selected by holding down **Ctrl** and then clicking on the feature in the table.

Multiple groups can be displayed by setting a ‘pruning threshold’ on the Dendrogram. To adjust this threshold (horizontal grey line located at the top of the y axis) click and drag the line down over the dendrogram.

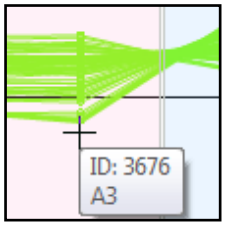


Note: as it moves over each branch, all the branches beneath become colour coded and selected

When you release the mouse button, the expression profiles for each branch of the dendrogram will be displayed. The number and composition of these ‘branches’ is controlled by sliding the threshold up and down.

To identify the shape of individual profiles in the expression graphs drag the mouse over the profiles. The current profile is highlighted with a tool tip showing the feature ID and relevant sample information.

Finally, the table can be ordered based on cluster by clicking on the ‘Cluster’ column.



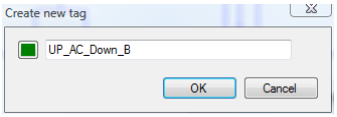
Additional tagging of features in Progenesis Stats

The Dendrogram view can be used to further categorise, and therefore tag, groups of features based on their expression profiles.

As you click on a 'branch' of the dendrogram the feature 'expression' profiles are displayed and all the features in the selected branch become selected in the table.

To set a new Tag, for these feature with similar expression behavior, right click on a selected area of the table and select **New tag**

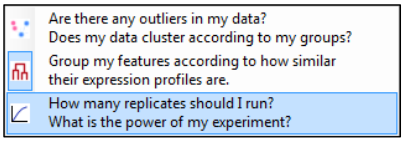
Give the Tag a name: **UP_AC_Down_B**



Select and name other branches of the dendrogram as required.

Power Analysis

To explore the third Statistical analysis of the data click on the blue link Ask another question at the top of the table. The selection of 3 tools will appear in the form of questions.



Select the third option to explore the number of replicates required and obtain a measure of the 'power' of the current experiment.

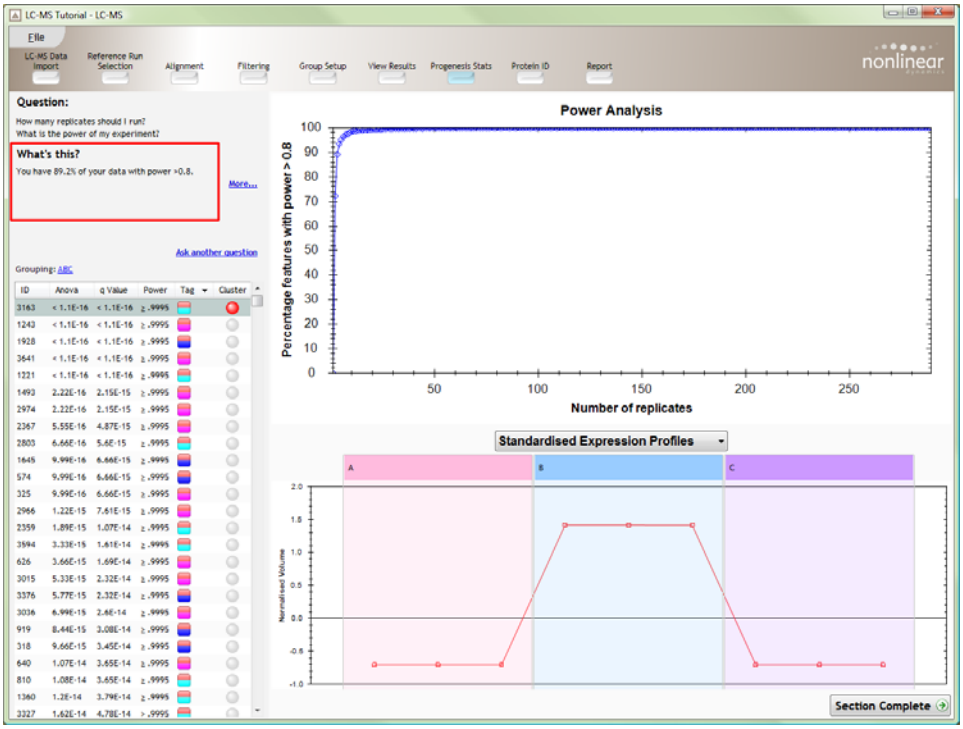
This time the statistically based question(s) being asked of the application is:

'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 all **3792 features** view the power analysis.



This is displayed graphically showing what percentage of the 3792 features (ranked on ID) used in the test, give a power of 80% or more and how this relates to the **number of replicates**.

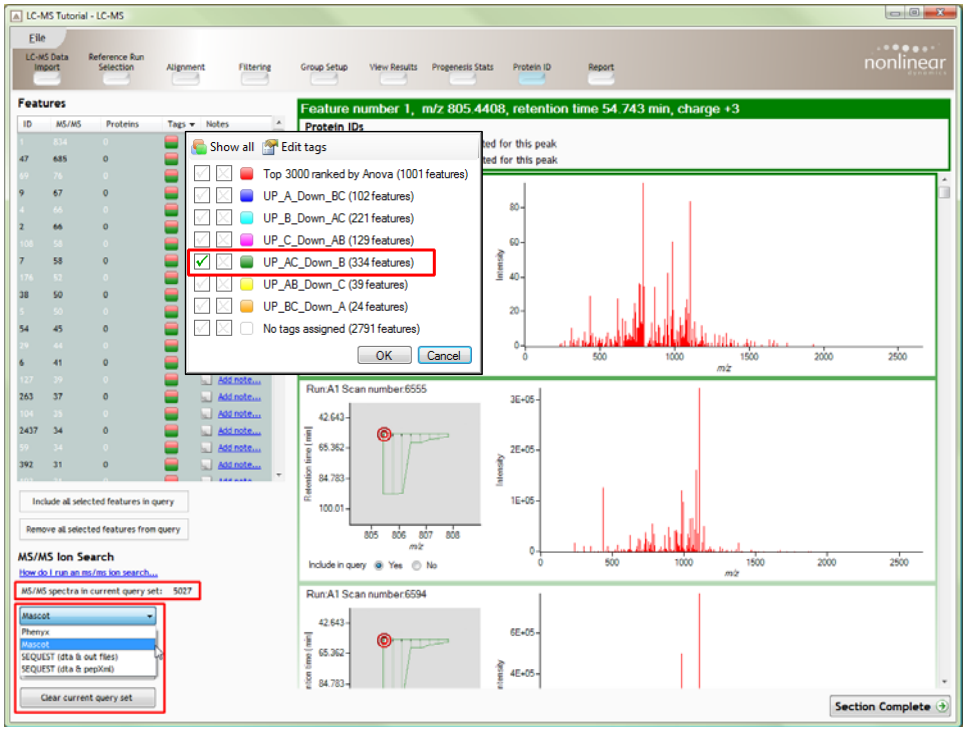
Power Analysis

- **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%.

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

Stage 9: Protein ID

Determining protein ID is dependant on the availability of MS/MS data for the LC-MS runs. This data may be available but limited if the LC-MS was performed in a data dependant MS/MS detection mode due to under sampling. Under these conditions MS/MS data acquisition is dependant on thresholds and parameters set prior to the LC-MS run.

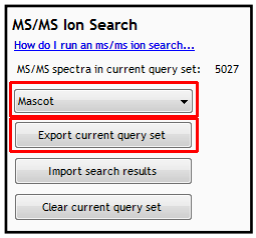


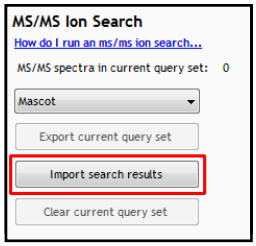
For this tutorial we are using LC-MS runs containing MS/MS data where the data was acquired in a data dependant mode.

Note: by default the table is ordered on the number of MS/MS spectra available for each Feature ID.

Now ensure that the table is only displaying those features showing the feature behaviour **UP_AC_Down_B** by setting the **Tags** as shown above, (approx 334 features are displayed).

Performing an MS/MS Ion Search





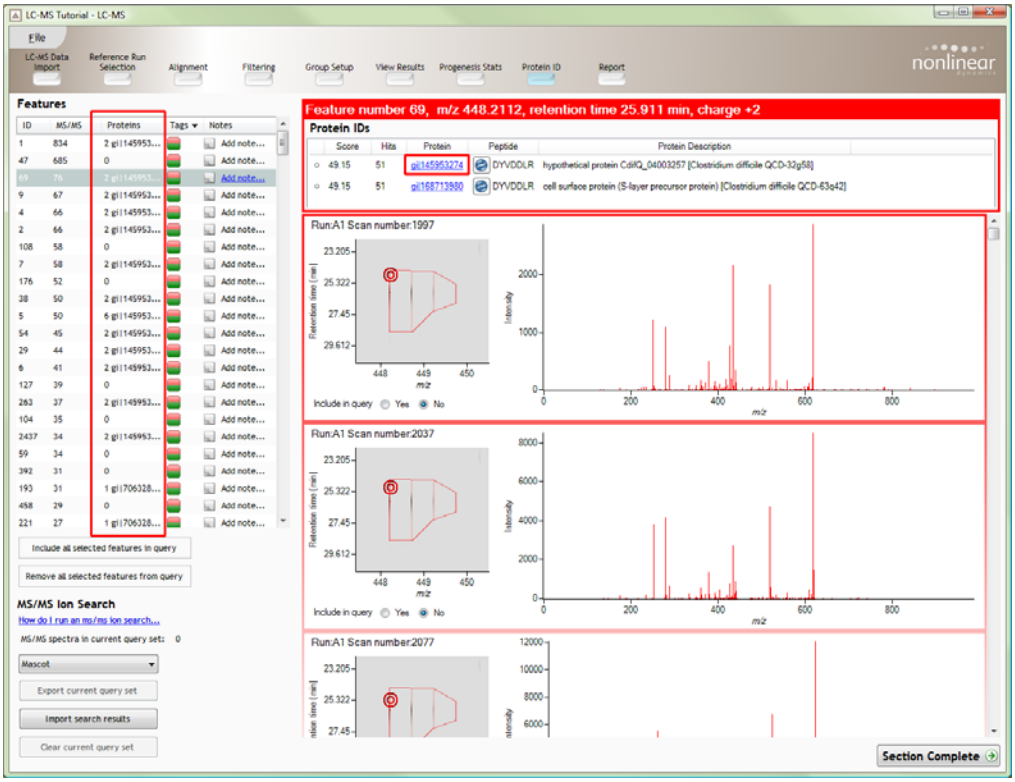
Using LC-MS data containing MS/MS spectra, protein ID can be determined for all or a selection of peaks which have available MS/MS spectra.

1. Highlight all the features with MS/MS spectra in the Features table
2. Click on 'Include all selected features in query'
3. In this case the current 'query set' contains 5,027 MS/MS spectra
4. Select appropriate search engine i.e. Mascot
5. Click 'Export current query set' to save search as file
6. Perform search on appropriate search engine and save results file
7. Click 'Import search results', locate results file and open

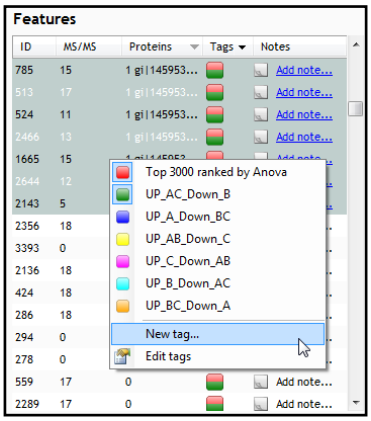
Note: an example Search Results file is available in the folder you restored the Archive to (Peptide search results.xml).

Note: the blue link tells you the appropriate formats for exporting ID results

On importing the Search results the Features table updates to reflect the identified proteins and lists the IDs for the currently selected feature.



Note: Additional information on the identified protein is available using the 'Accession Number Search' links in the Protein column.



Now create a new tag for 'Identified Proteins' by ordering the **Proteins** column, then highlighting all the peaks with protein id entries, and right clicking to create the New tag.

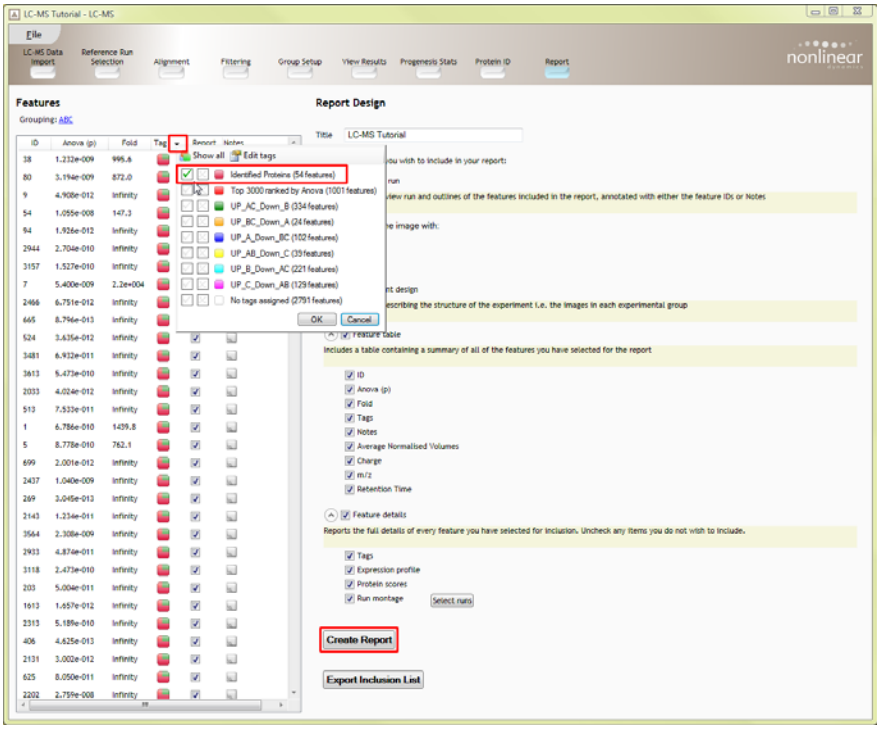
Click **section complete** to move to the Reporting section.

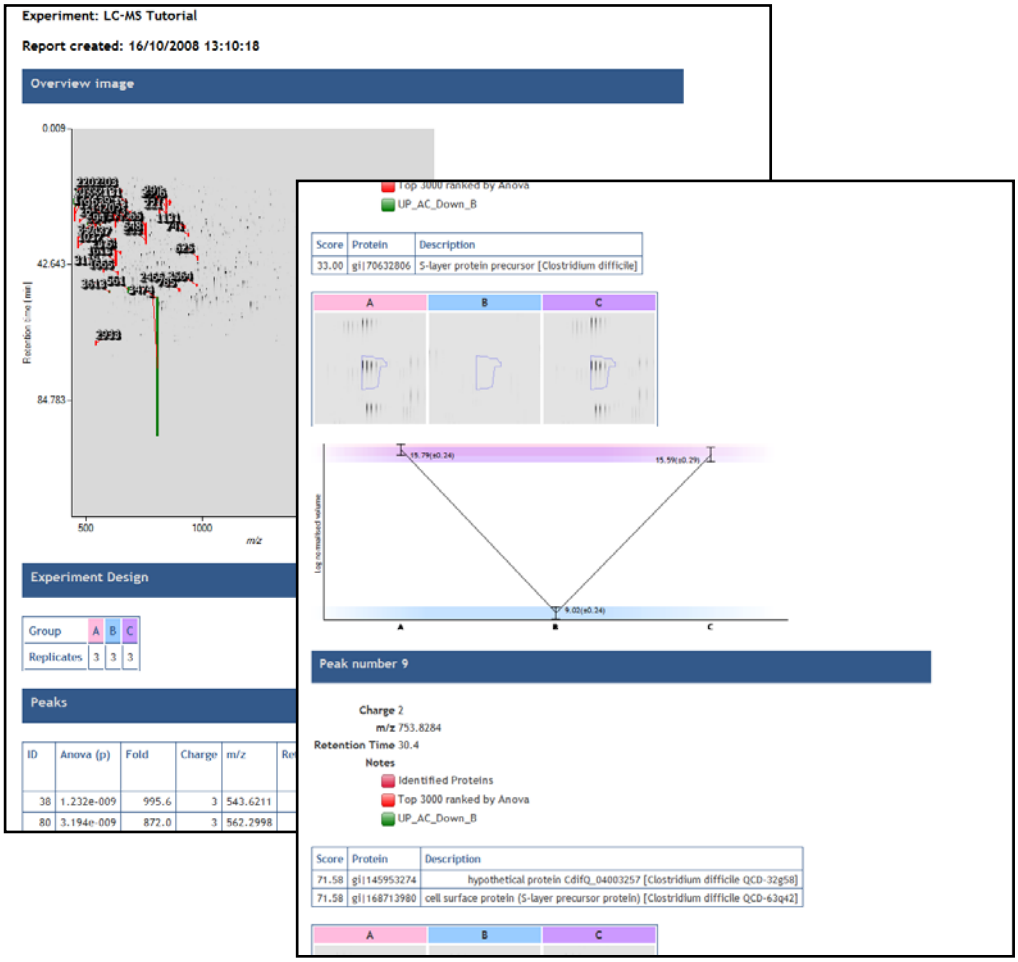
Stage 10: Reporting

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

As an example we will create a report for **only** the peaks with identified proteins using the tag created in the previous section.

1. First reduce the features to report on by selecting the 'Identified Proteins' tag. In this example it reduces the number of features in the table to 54.
2. Tick all the optional details to be included in the report
3. Click **Create Report**

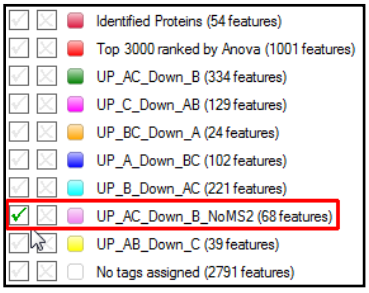




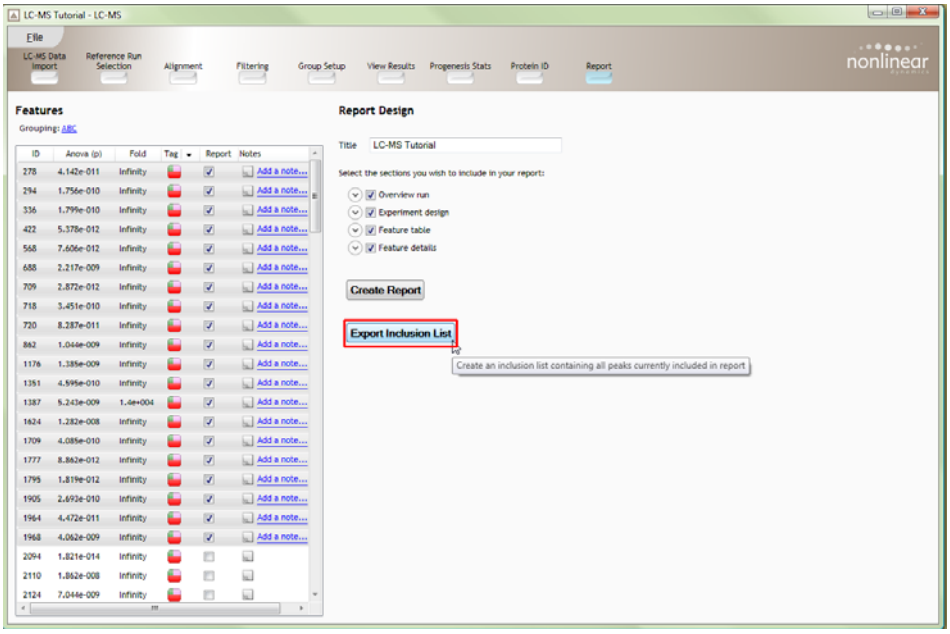
Scroll down to view details for each feature and montage. The report is printable and can also be saved.

You can combine the use of Tags to recall specific data and then use the tick boxes to reduce the number of features being reported on.

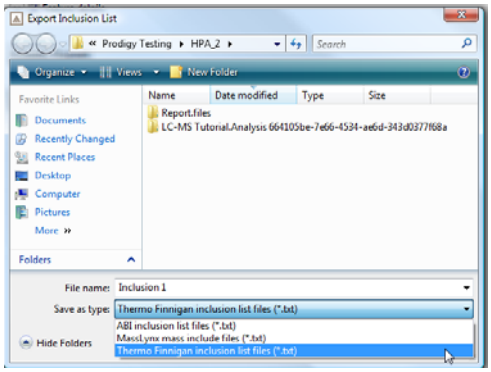
Alternatively the features can be exported as an ‘Inclusion list’.



For example recall those features showing reduced expression in strain B with no MS/MS spectra (68 features in this example). Then order the feature list using 'ID'. Now highlight the list using **Ctrl A** and untick all the features then select the top 20 features (as shown below)



Finally export the inclusion list in the appropriate MS machine format to use in the acquisition of additional MS/MS spectra from a new LC-MS run.



Note: The new LC-MS runs can then be added to the original experiment to increase the MS/MS coverage using the **Add data files** facility at the Data Import Stage.

Congratulations!

This document has taken you through a complete analysis using Progenesis LC-MS, from Alignment through Analysis to generating lists of interesting features using powerful Multivariate Statistical analysis of the data.

Hopefully our example has shown you how this unique technology, currently available within Progenesis LC-MS package, can deliver significant benefits with

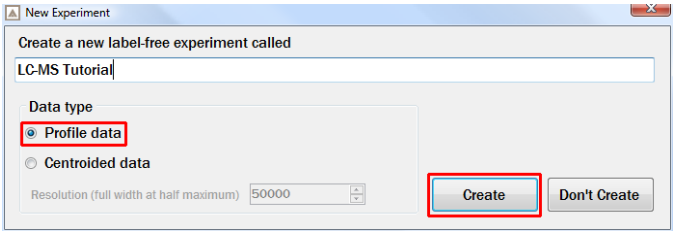
- Speed
- Objectivity
- Statistical Power

> If you would like to see the benefits of running Progenesis LC-MS using your own data and explore the LC-MS Data Import module as well as the rest of the workflow please go to the next section.

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 (Thermo and Waters) or, alternatively convert them to mzXML format first using the software managing the LC-MS acquisition.

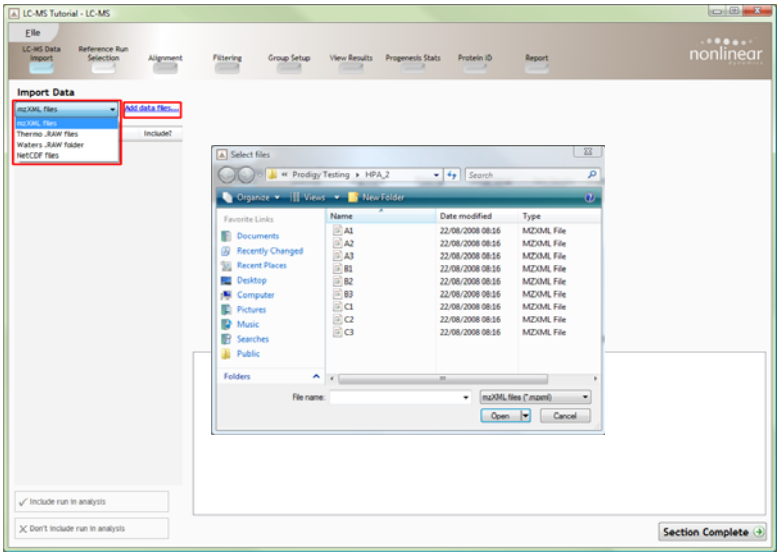
To create a new experiment with your files select **New** give your experiment a name. Then select data type, the default is 'Profile data'.



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

Select the 'Import Data file format', in this tutorial they are mzXML files

Then locate your data files using the Add data files.. link.

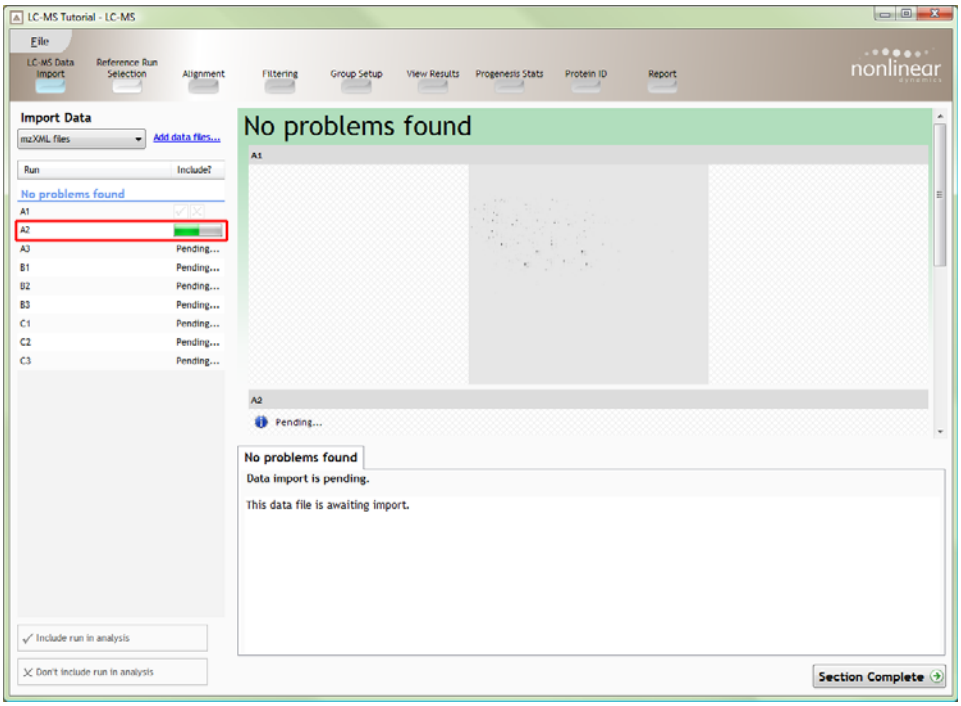


Locate and select all the Data files (A1 to C3).

On loading the selected runs your data set will be automatically examined and the size of each file will be reduced by a ‘data reduction routine’, which reduces 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.

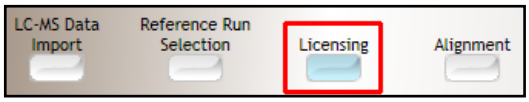


Note: as each data file is loaded the progress is reported in the **Import Data** list. The dialog below the image reports on the QC of the imported Data files. In this case ‘No problems found’ with the this data file

Now move to the next stage in the workflow (page 6 in this tutorial) by clicking **Section Complete.**

Appendix 2: Licensing Runs (Stage 3)

When setting up a **New experiment** if you are evaluating LC-MS with unlicensed Runs then the licensing page will open after **Reference Run Selection**



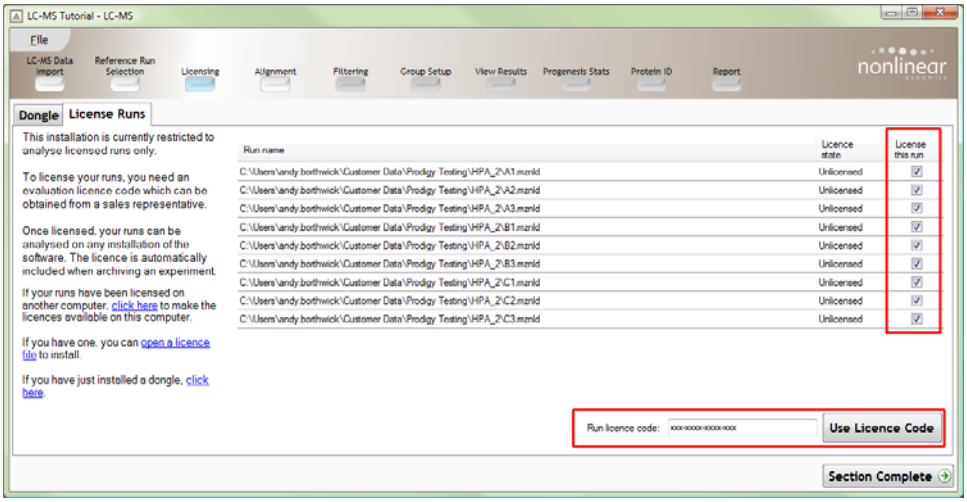
If you already have a programmed dongle attached to your machine then the following License Images page will not appear.

To use this page to License your Runs **you must first either obtain an 'Evaluation' License Code from a Nonlinear Sales Person or have purchased a license code directly from Nonlinear**.

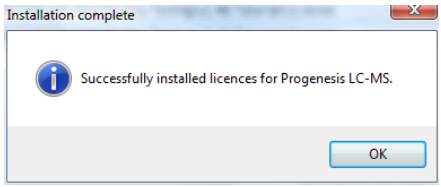
Each code will allow you to license a set number of Runs.

For the example the Runs you wish to License will be listed as shown below.

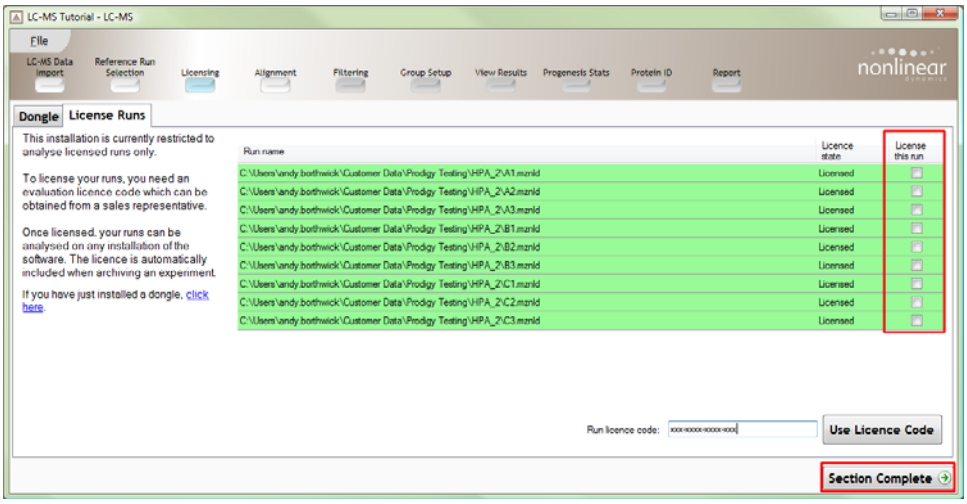
To activate license(s) for the selected images enter the code in the space provided and click **Use Licence code.**



A message confirming successful installation of your image licenses will appear.



Click **OK**, the view will update the display to show that the files are now licensed



Click **Section Complete** and **Alignment** will open.