



# **Progenesis CoMet User Guide**

**Analysis workflow guidelines**

**for version 1.0**



## Contents

|   |    |
|---|----|
| Introduction .....  | 3  |
| How to use this document .....  | 3  |
| How can I analyse my own runs using CoMet? .....                        | 3  |
| LC-MS Data used in this user guide.....                                 | 3  |
| Workflow approach to data analysis.....                                 | 4  |
| Restoring the CoMet Tutorial.....                                       | 5  |
| Stage 1: Data import and QC review of CoMet data set .....              | 6  |
| Stage 2: Reference Run selection.....                                   | 7  |
| Stage 3: Licensing .....  | 7  |
| Stage 4: Alignment .....  | 8  |
| Stage 5A: Filtering .....   | 14 |
| Stage 5B: Reviewing Normalisation .....                                 | 17 |
| Stage 6: Experiment Design Setup for Analysed Runs.....                 | 20 |
| Stage 7: Validation, review and editing of results .....                | 23 |
| Stage 8: Identify Compounds .....                                       | 31 |
| Stage 9: Review Compounds .....   | 33 |
| Stage 10: Multivariate Statistics .....                                 | 40 |
| Congratulations! .....  | 44 |
| Appendix 1: Stage 1 Data Import and QC review of LC-MS data set .....   | 45 |
| Appendix 2: Stage 1 Data QC review and addition of exclusion areas..... | 47 |
| Appendix 3: Licensing runs (Stage 3) .....                              | 48 |
| Appendix 4: Manual assistance of Alignment.....                         | 49 |
| Appendix 5: Within-subject Design.....                                  | 54 |
| Appendix 6: Deciding which identifications are correct .....            | 56 |
| Appendix 7: Power Analysis .....  | 58 |

## Introduction

This user guide takes you through a complete analysis of 12 LC-MS runs of metabolite containing samples for 3 biological groups (4 replicate runs per group) using the unique Progenesis CoMet workflow. It starts with LC-MS data file loading then Alignment, followed by Analysis that creates a list of interesting features (molecules) which are identified with compound search and then explored within Progenesis Stats using multivariate statistical methods.

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 CoMet workflow, therefore if you are using your own data files please refer to Appendix 1 (page 45) then start at page 7.

## 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 50 minutes and is divided into two sections. This means you can perform the first half focused on run alignment and analysis then complete the second half of analysis exploring comparative differences and Compound identity at a convenient time. If you experience any problems or require assistance, please contact us at [support@nonlinear.com](mailto:support@nonlinear.com)

## How can I analyse my own runs using CoMet?

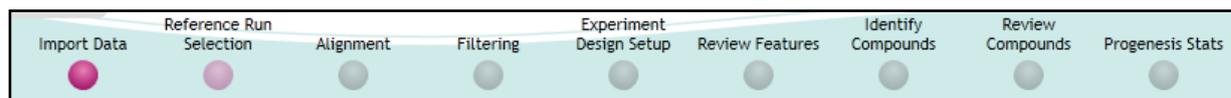
You can freely explore the quality of your data using Data Import and then licence your own runs using this evaluation copy of Progenesis CoMet. Instructions on how to do this are included in a section at the end of the user guide document. Alternatively if you would like to arrange a demonstration in your own laboratory contact [support@nonlinear.com](mailto:support@nonlinear.com) and we will help you.

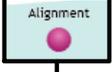
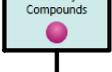
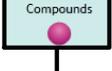
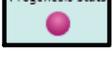
## LC-MS Data used in this user guide

NLD would like to thank Dr Matthew DiLeo of the Boyce Thompson Institute (now at Keygene) and Meghan den Bakker for providing the example data used in this user guide as well as invaluable discussion on the handling of the data.

## Workflow approach to data analysis

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

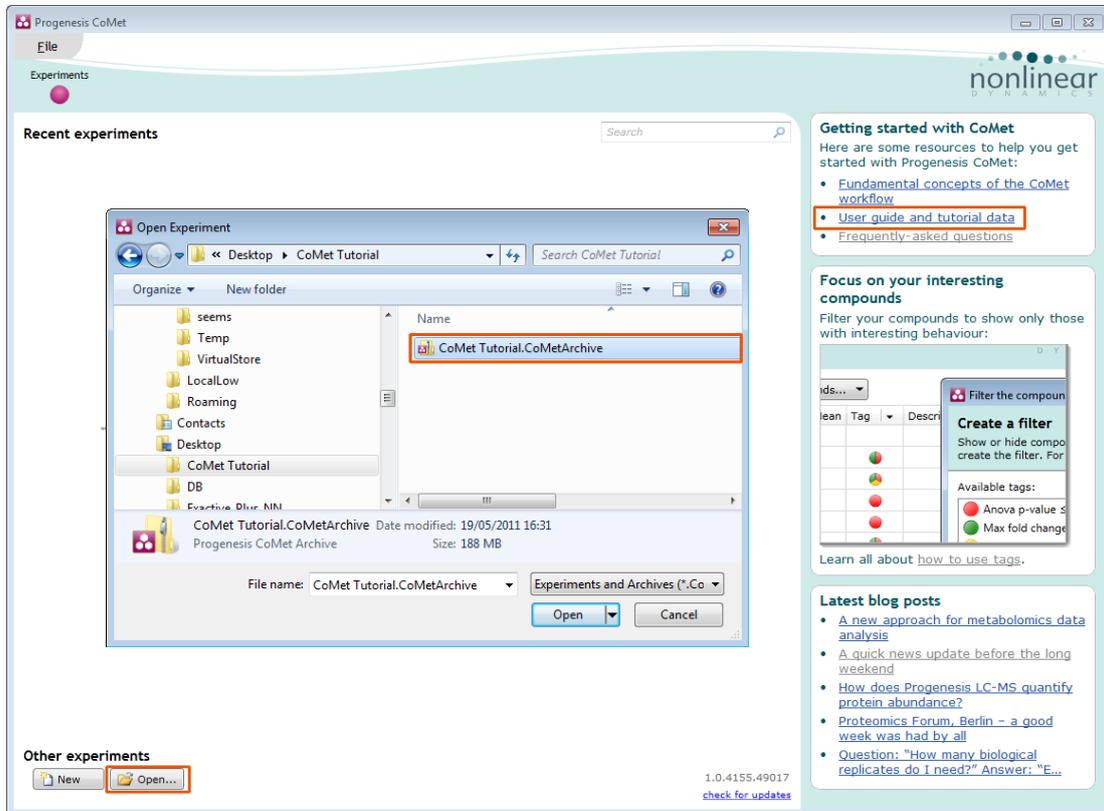


| Stage   | Description  | Page |
|---|--|------|
|    | <b>Import Data:</b> Selection and review of data files for analysis  | 6    |
|    | <b>Reference Run Selection:</b> Select run to align to.  | 7    |
|    | <b>Licensing:</b> allows licensing of individual data files when there is no dongle attached (Appendix 3)                | 7    |
|    | <b>Alignment:</b> automatic and manual run alignment   | 8    |
|  | <b>Filtering:</b> defining filters for features based on Retention Time, m/z , Charge and Number of Isotopes.            | 14   |
|  | <b>Review Normalisation:</b> examine data normalisation methods  | 17   |
|  | <b>Experiment Design Setup:</b> defining one or more group set ups for analysed aligned runs                             | 20   |
|  | <b>Review Features:</b> review and edit feature detection, tag groups of features and select features for identification | 23   |
|  | <b>Identify Compounds:</b> search identity of compounds using tagged and selected groups                                 | 31   |
|  | <b>Review Compounds:</b> managing possible compound identities of peptide ids from Peptide Search engines                | 33   |
|  | <b>Progenesis Stats:</b> performing multivariate statistical analysis on tagged and selected groups of compounds         | 40   |

## Restoring the CoMet Tutorial

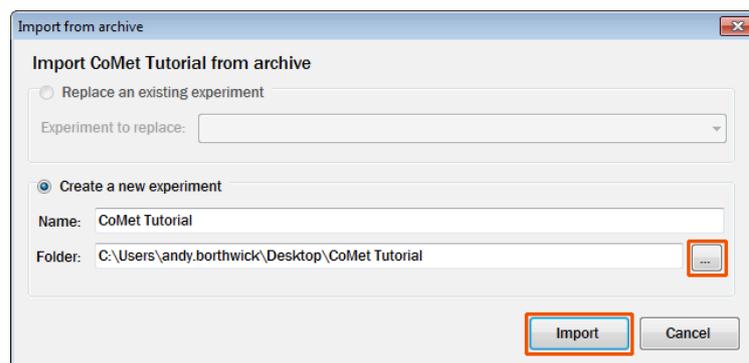
Open Progenesis CoMet and download the Compressed (.zip) Tutorial Archive file from the **'User guide and tutorial data'** link shown below, placing it in a **new folder** on your desktop. Before restoring the tutorial in the software **you must** first right click on the downloaded (.zip) file and extract it to the a folder, i.e. CoMet Tutorial.

Now you can restore the uncompressed CoMet tutorial archive file. To do this, first locate the CoMet Tutorial Archive file using the **Open** button.

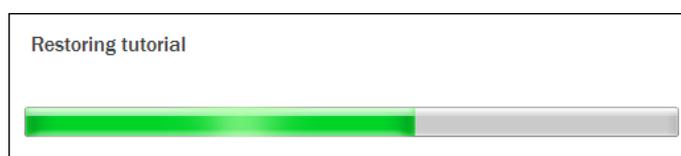


This opens the 'Import from archive' dialog.

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.

## Stage 1: Data import and QC review of CoMet data set

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

The screenshot shows the Progenesis CoMet software interface during the 'Import Data' stage. The main window displays a grid of 12 2D ion intensity maps (A1 through P4) and a large 2D plot of ion intensity versus Retention Time (mins) and m/z. The 'Import Data' section on the left provides instructions for importing and reviewing chromatography. The 'About this run' box on the right displays statistics for the selected run (A1), including MS/MS count, Peak count, Total ion intensity, and Masked areas. The 'Section Complete' button is visible at the bottom right.

Each run appears as a 2D ion intensity map in the list. The current Run is displayed in the main window. At this stage you will be warned if there are any data import errors for any of the files. The files will be highlighted in red and the error will appear to the right of the screen.

**Note:** if you have imported one or more runs that are either not required for the experiment or are displaying data import errors (such as incorrect polarity) these runs can be deleted by right clicking on the run in the list and selecting **Remove run**.

Details about the experiment are displayed underneath the run list. This records the selections made when the experiment was created.

**Note:** the options for '**Peak processing, Feature detection Methods and Ionisation Polarity**', selected when the experiment was created, are detailed in Appendix 1, page 45).

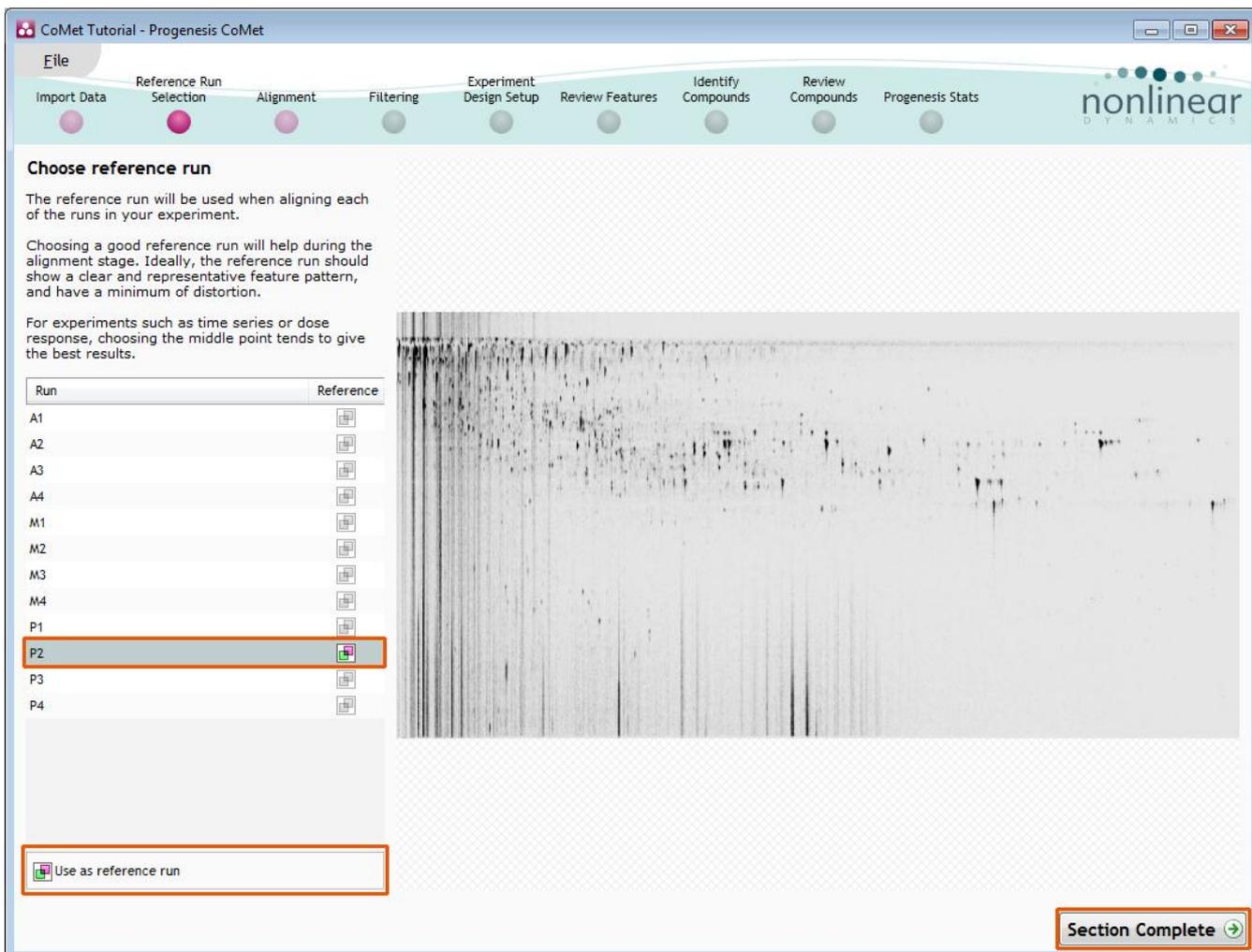
**Tip:** the '**Mask areas for peak picking**' facility (under **Actions**) allows you to exclude areas (usually early and/or late in the LC dimension (Retention Time) that appear excessively noisy due to capture of data during column regeneration (see Appendix 2, page 47). This is not used for this data set.

Examine the quality of the run(s) using the Zoom facility on the main window.

Once all the files have been imported move to the next stage in the workflow by clicking **Section Complete**.

## Stage 2: Reference Run selection

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

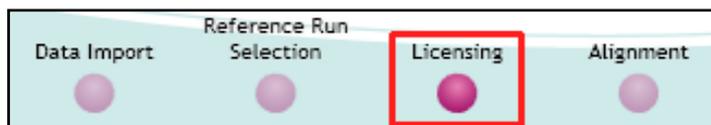


To select a Reference run either click on the run in the list and then click **Use as reference run** or double click on the run in the list. In this example **P2** is selected as the Reference Run

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

## 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..

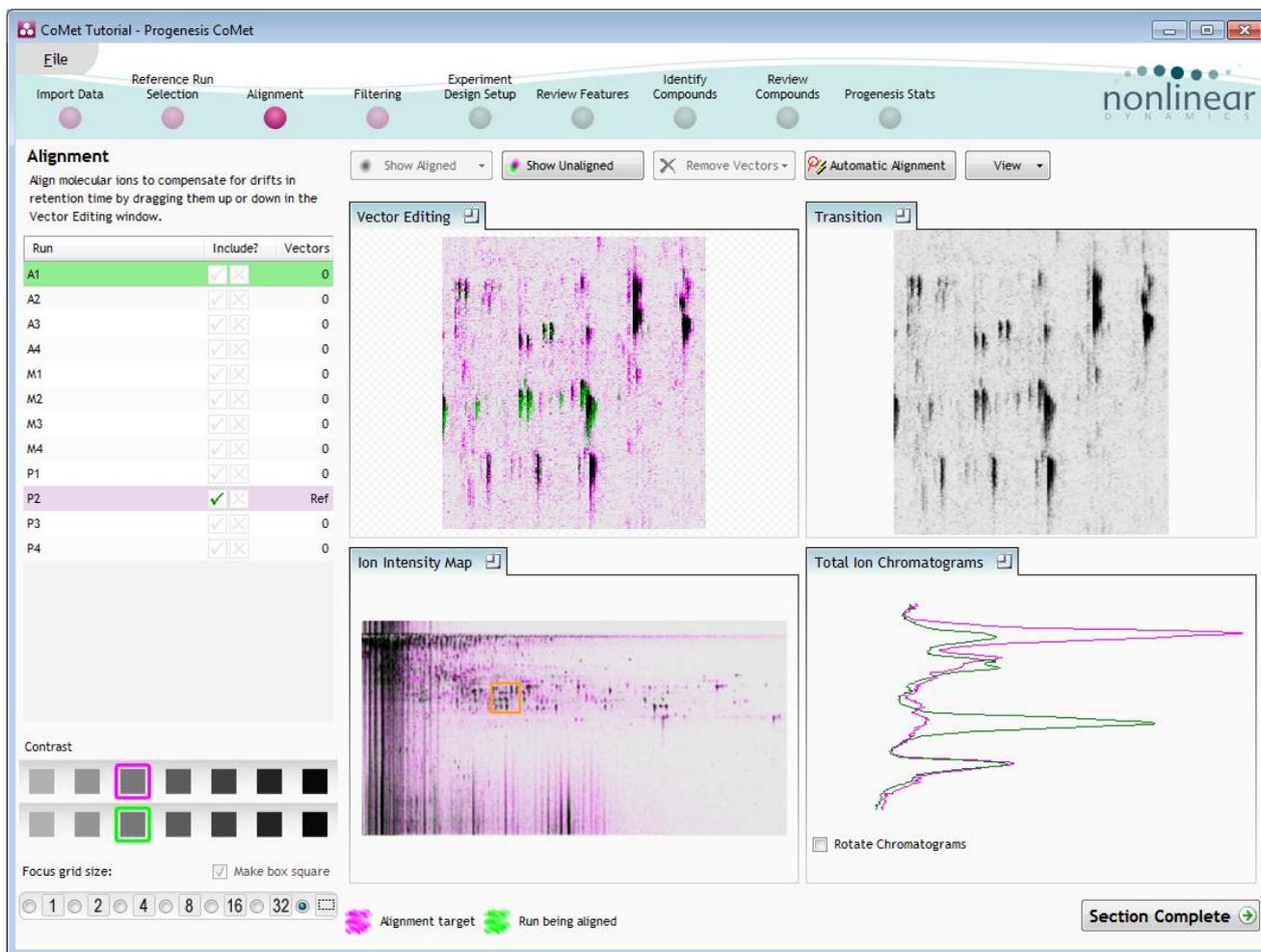


For details on how to use Licensing go to Appendix 3 (page 48)

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

## Stage 4: Alignment

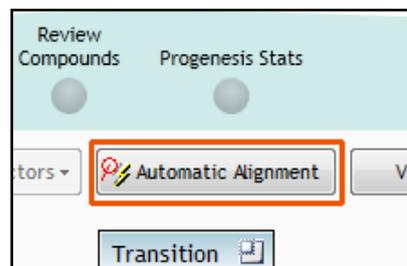
At this stage Progenesis CoMet Alignment opens displaying your data.



### Generation of alignment vectors

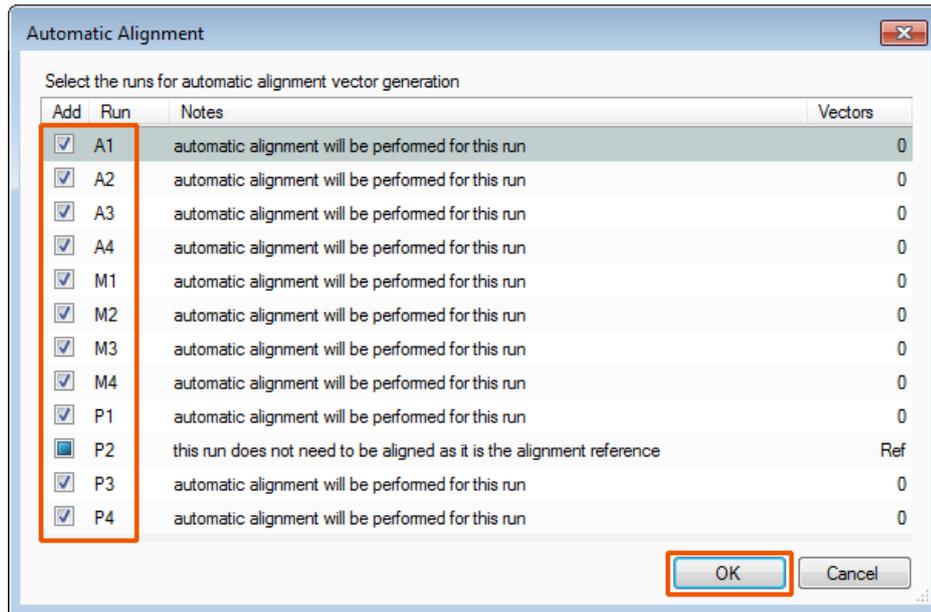
The alignment of the runs is required in the LC (retention time) direction, this is key to correcting for the variable elution of ‘compound ions’ during the chromatographic separation.

The Alignment algorithm will generate ‘Automatic’ vectors, in the retention time direction for each run, to enable the alignment of all the runs to the ‘Reference Run’.



Click **Automatic Alignment** on the top tool bar.

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



The following pages in this user guide explain in more detail the views and functions of the Alignment stage in the Progenesis CoMet, focusing on the Program layout.

These pages act as a useful guide and reference to the Alignment Stage that you can return to after having generated the Alignment vectors automatically.

## Taking a detailed approach to alignment

In some cases, where the misalignment is severe, 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 give better results.

For more details on manual assistance of Alignment refer to Appendix 4 page 49

## Layout of Alignment

To familiarize you with Progenesis CoMet alignment, this section describes the various graphical features used in the alignment of the 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.
- If the green and magenta features (immediately above) have not aligned automatically then **drag** the green feature over the magenta feature and **release** the mouse button.
- The view will 'bounce' back and a red vector, starting in the green feature and finishing in the circled magenta feature will now appear as shown below in window A.

The experiment structure is displayed on the left of the screen in the **Run** panel.

**Current Run (Green)**

**Reference Run (Magenta)**

**Added alignment Vector**

**Alpha Blend display animates between current and reference runs**

**Current Focus**

**Section Complete**

The **Runs**: panel shows the run that is currently being aligned in green, and the run it is being aligned to in magenta. This is the reference run you chose at the previous stage, in this case **P2**.

| Run | Include?   | Vectors |
|-----|--|---------|
| A1  | <input checked="" type="checkbox"/> <input type="checkbox"/> | 0       |
| A2  | <input checked="" type="checkbox"/> <input type="checkbox"/> | 0       |
| A3  | <input checked="" type="checkbox"/> <input type="checkbox"/> | 0       |
| A4  | <input checked="" type="checkbox"/> <input type="checkbox"/> | 0       |
| M1  | <input checked="" type="checkbox"/> <input type="checkbox"/> | 0       |
| M2  | <input checked="" type="checkbox"/> <input type="checkbox"/> | 0       |
| M3  | <input checked="" type="checkbox"/> <input type="checkbox"/> | 1       |
| M4  | <input checked="" type="checkbox"/> <input type="checkbox"/> | 0       |
| P1  | <input checked="" type="checkbox"/> <input type="checkbox"/> | 0       |
| P2  | <input checked="" type="checkbox"/> <input type="checkbox"/> | Ref     |
| P3  | <input checked="" type="checkbox"/> <input type="checkbox"/> | 0       |
| P4  | <input checked="" type="checkbox"/> <input type="checkbox"/> | 0       |

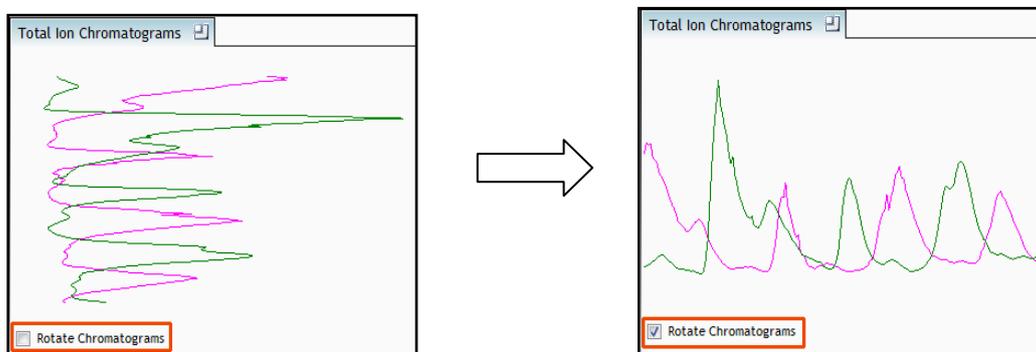
**Vector Editing (Window A):** is the main alignment area and displays the area defined by the current **focus** rectangle shown in Window C. The current run is displayed in green and the chosen reference run is displayed in magenta. This is where you place the alignment vectors.

**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 will change to show a zoomed view of the area being aligned to help accurate placement.

**Ion Intensity Map (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 cursor keys. The focus area size can be altered using the controls in the bottom left of the screen or by clicking and dragging out a new area with the mouse.

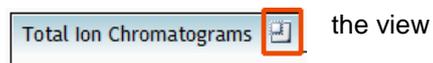
**Total Ion Chromatograms (Window D):** shows the current **total ion** chromatogram (green) overlaid on the Reference chromatogram (magenta). As the features are aligned in the **Vector Editing** view the chromatograms become aligned. The retention time range displayed is the vertical dimension of the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

**Note:** the orientation of the TIC view can be changed according to individual preference



This view assists in the verification of the feature alignment .

**Note:** the icon to the right of the 'Window' titles expands



the view

## Reviewing generation of alignment vectors

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

If the alignment has worked well then in Windows A and C the grid lines should show minimal distortion, Window B (Transition) will show features pulsing slightly but not moving up and down.

**Alignment**

Alignment is used to compensate for retention time drift between your runs.

Use the Automatic Alignment button to begin then review the alignment on each of your runs before moving on.

| Run | Include?                            | Vectors |
|-----|-------------------------------------|---------|
| A1  | <input checked="" type="checkbox"/> | 189     |
| A2  | <input checked="" type="checkbox"/> | 251     |
| A3  | <input checked="" type="checkbox"/> | 211     |
| A4  | <input checked="" type="checkbox"/> | 198     |
| M1  | <input checked="" type="checkbox"/> | 192     |
| M2  | <input checked="" type="checkbox"/> | 201     |
| M3  | <input checked="" type="checkbox"/> | 207     |
| M4  | <input checked="" type="checkbox"/> | 215     |
| P1  | <input checked="" type="checkbox"/> | 241     |
| P2  | <input checked="" type="checkbox"/> | Ref     |
| P3  | <input checked="" type="checkbox"/> | 275     |
| P4  | <input checked="" type="checkbox"/> | 257     |

Contrast:

Focus grid size:  Make box square

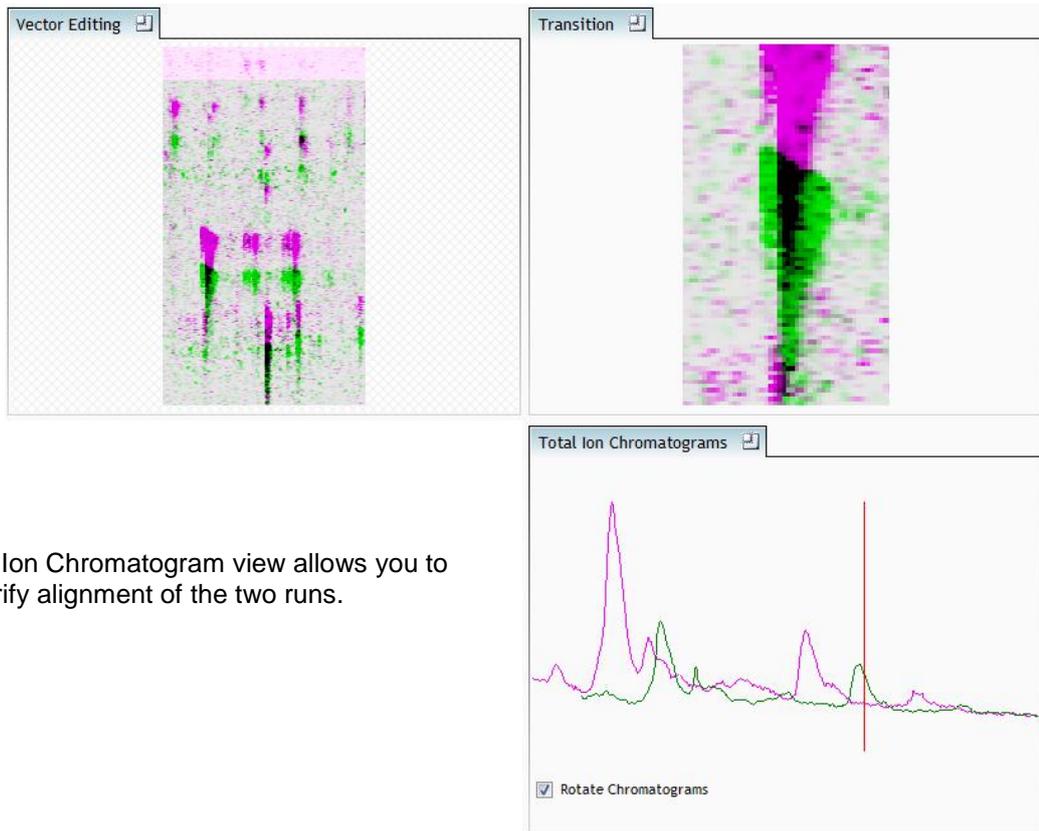
1  2  4  8  16  32  64

Alignment target  Run being aligned

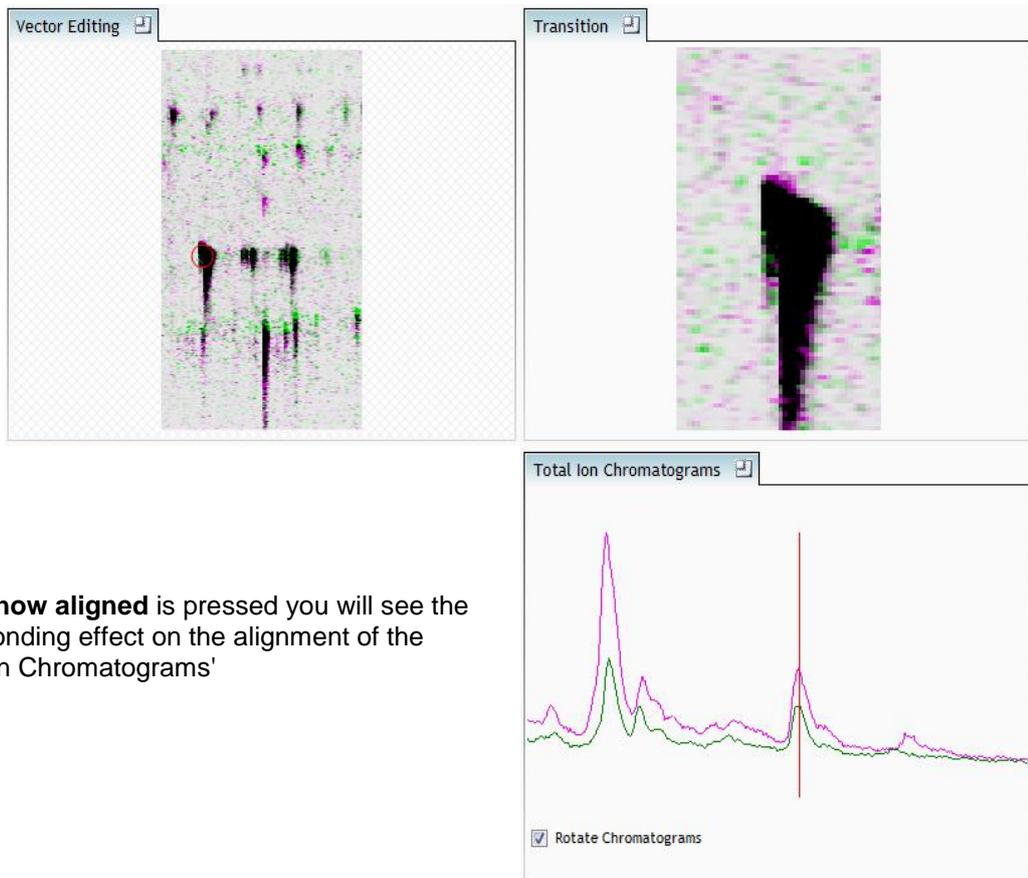
Section Complete →

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 run alignment (for more information refer to Appendix 4 (page 49)).



The Total Ion Chromatogram view allows you to further verify alignment of the two runs.



When **show aligned** is pressed you will see the corresponding effect on the alignment of the 'Total Ion Chromatograms'

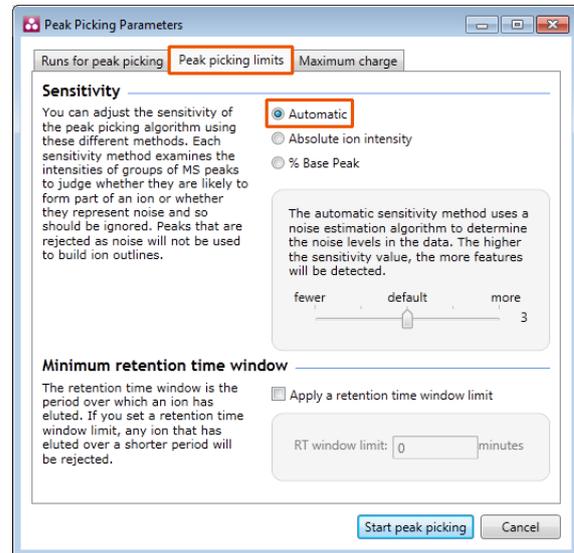
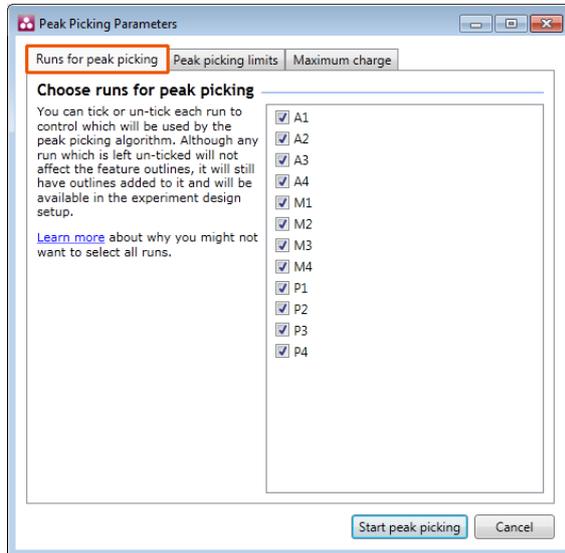
## Stage 5A: Filtering

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, showing all the runs in the current experiment and a tick against each run, opens. 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:** feature outlines will be added to 'un-ticked' runs; however, 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.

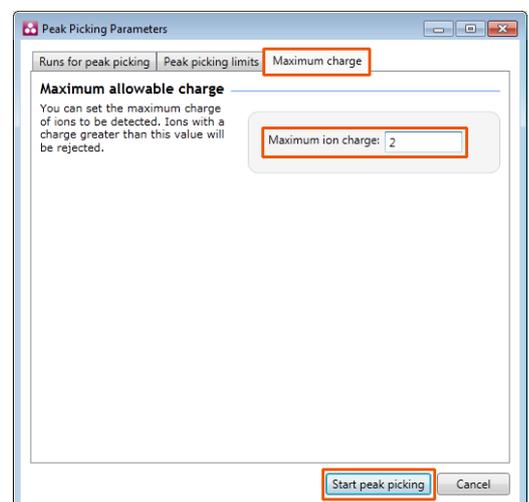
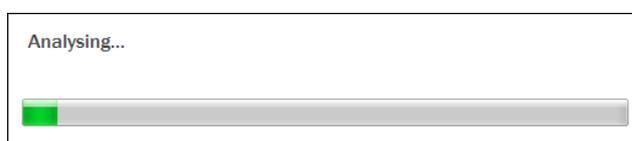
The sensitivity of the peak picking can be controlled by adjusting settings under the **Peak picking limits** tab.

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 should be detected. The default setting is a charge state of 2.

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, 3568.

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



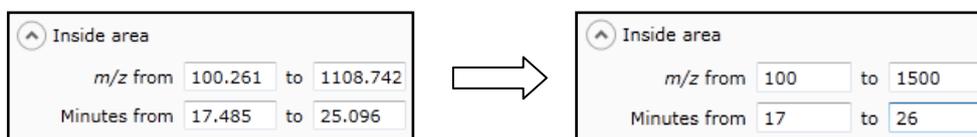
For example, try deleting features with late 'Retention times' by dragging out an area as shown.

All features contained within the mask will be selected

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



**Tip:** the limits can be adjusted by entering the required values in the boxes



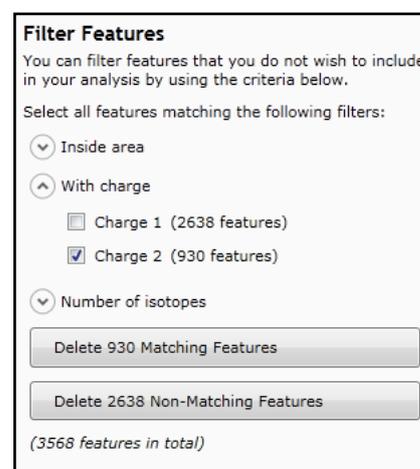
To remove the (in this case 1397) features inside (and touching) the selected area, press the **Delete 1397 Matching Features** button.

As we require all the detected features press **Undo** (located above the graph).

In addition to setting limits for 'Retention time and m/z', features can also be selected on the basis of charge or the number of isotopes present. This allows 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.



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

**As an example** try filtering out all the features with a charge state of 2 as shown above.

All features with a charge state of 2 will appear **red**.

To remove these features press **Delete 930 Matching Features**.

As mentioned earlier 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 should review the normalisation of the experiment.

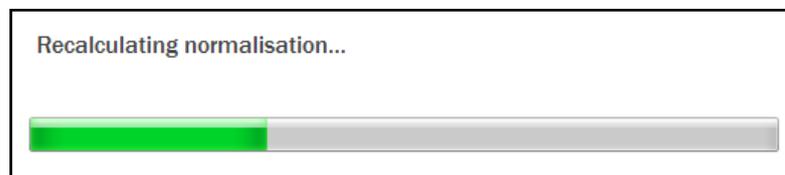
**After reviewing Normalisation restore the 930 features with Charge State 2 using the Undo button**

## Stage 5B: Reviewing Normalisation

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



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



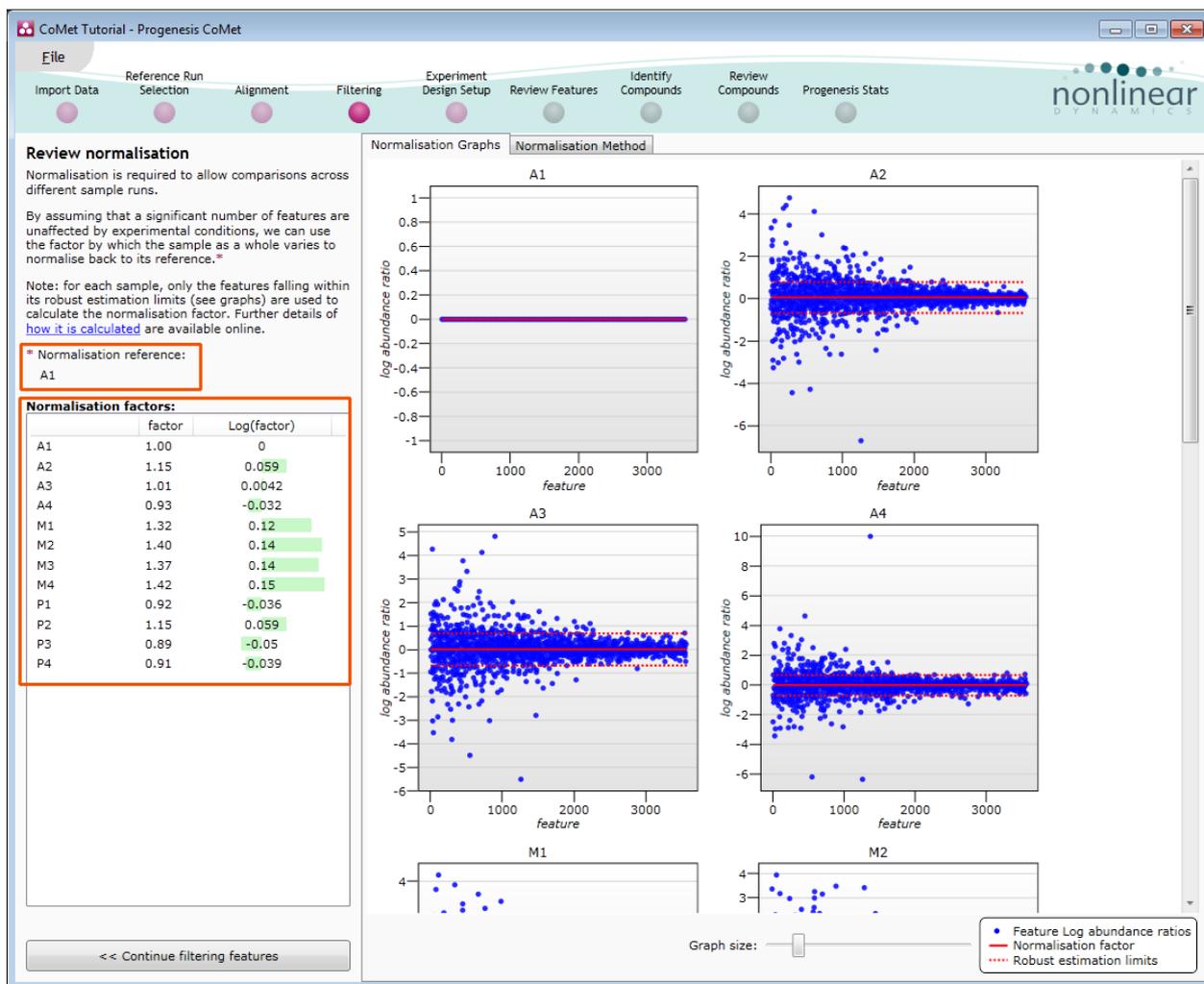
The **Review Normalisation** page will open displaying plots for the normalisation of all the features on each run.

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 by Changing the Normalisation method.

The default method is to **Normalise to all features**.

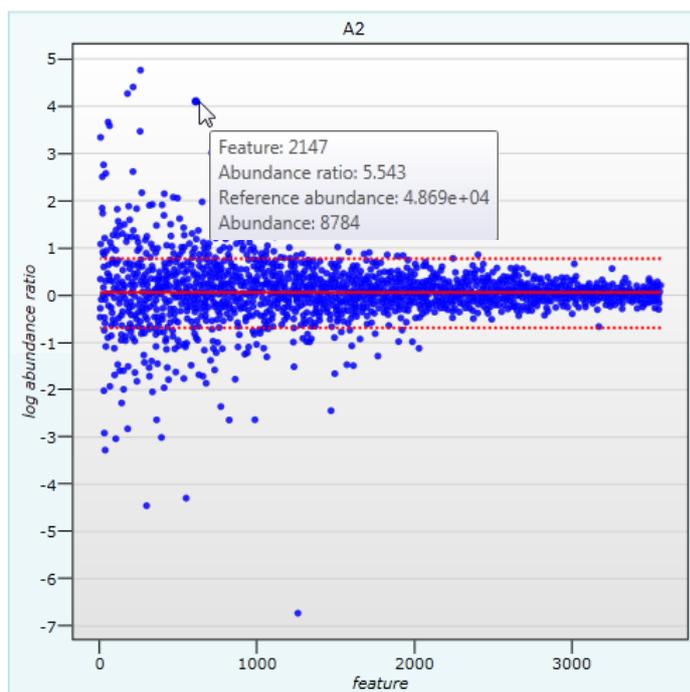
Normalisation factors are reported in the table to the left of the plots.



**Calculation of Normalisation Factor:**

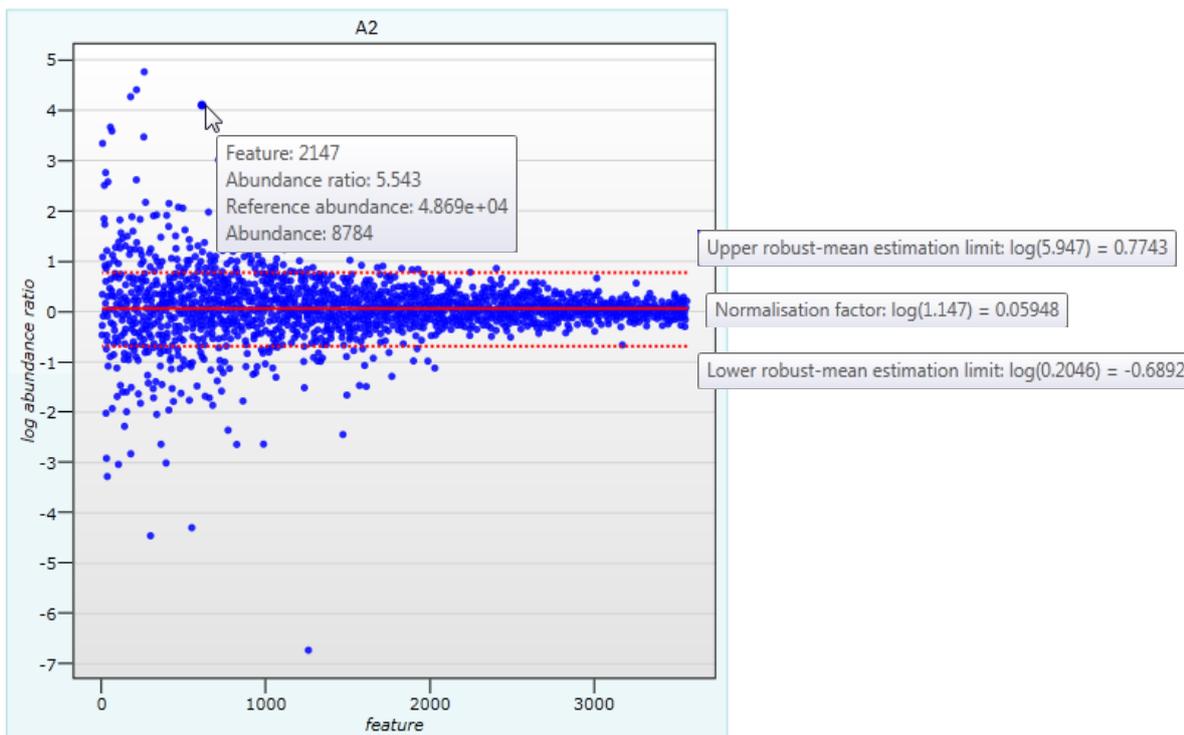
Progenesis CoMet will automatically select one of the runs that is 'least different' from all the other runs in the data set to be the 'Normalising reference'. The run used is shown above the table of Normalisation factors.

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

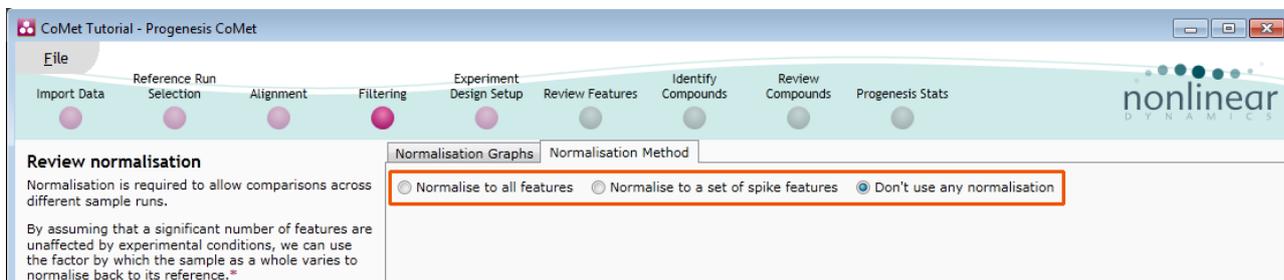


The details for individual features can be viewed as you hold the cursor over the dots on the plot.

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



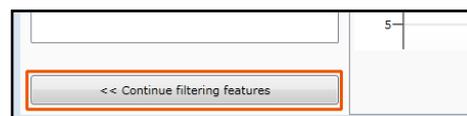
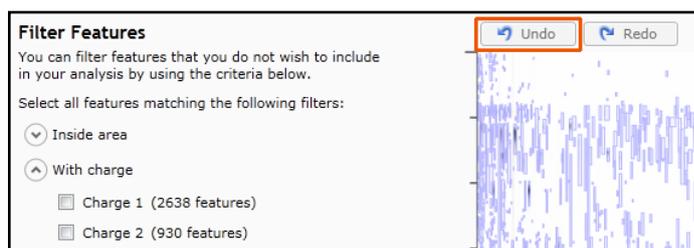
Finally, if you do not wish to work with normalised data then you can **use the raw abundances** by switching off the normalisation.



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

For this example experiment you should leave the **Normalise to all features** option selected.

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

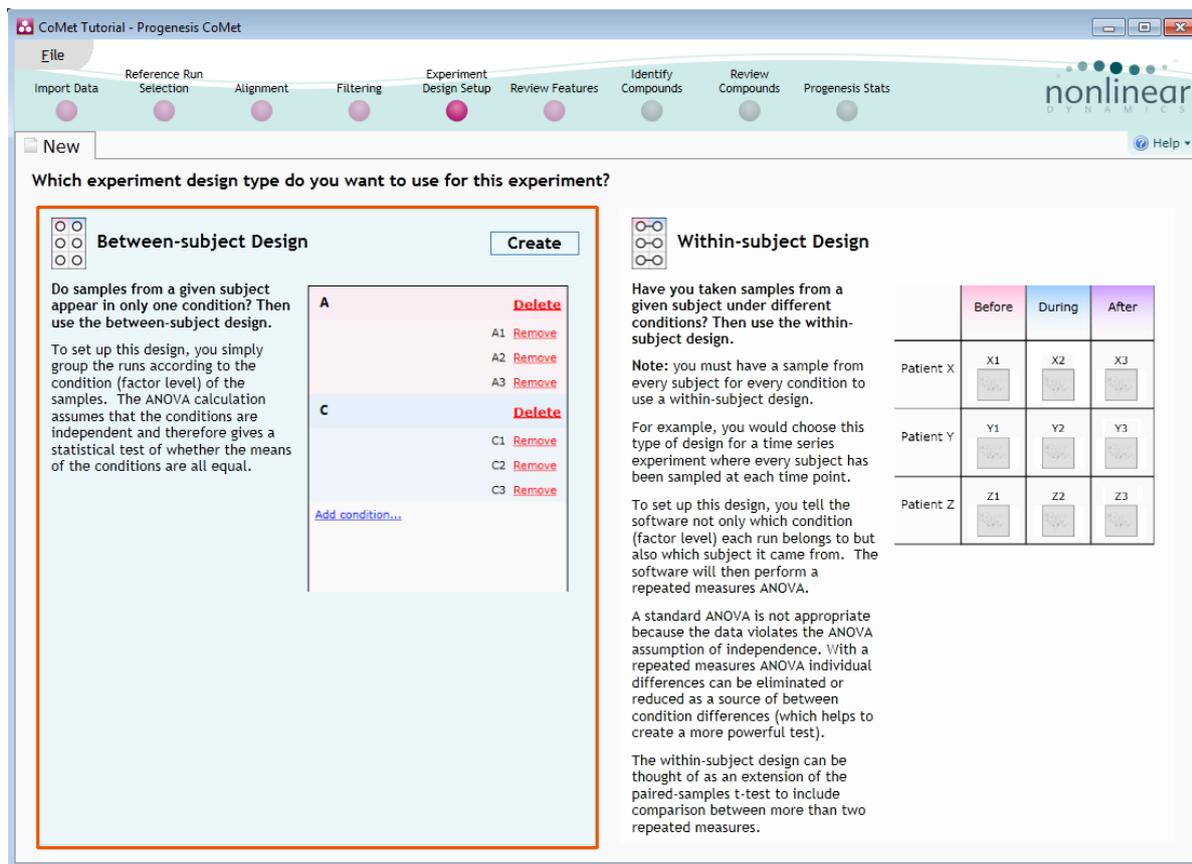


For this example we **DO NOT** do any Filtering so click Undo and then **Section complete.**

## 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:



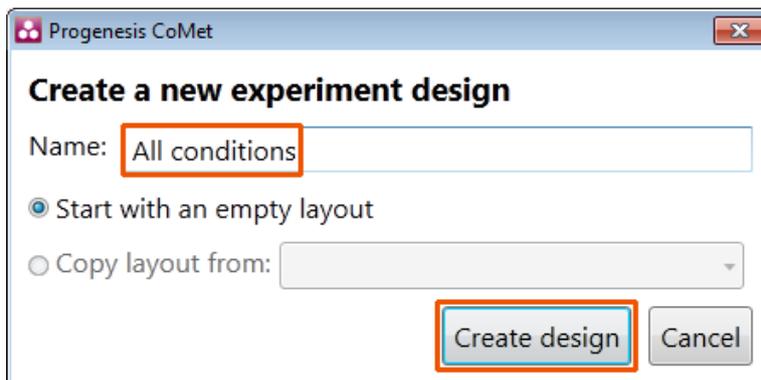
**Between-subject design:** here samples from any given subject appear in only one condition. (i.e. control versus various drug treatments). The ANOVA calculation assumes that the conditions are independent and applies the statistical test that assumes the means of the conditions are equal.

**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.

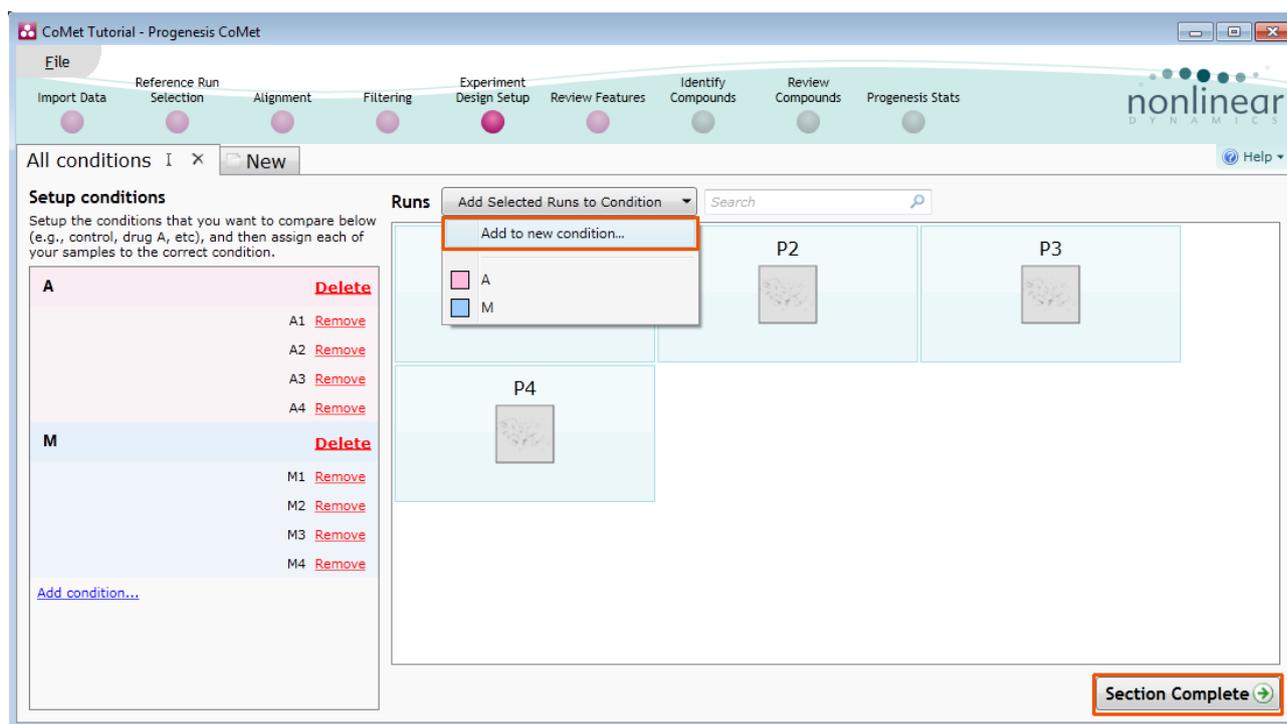
Additional information on how to apply the Within-subject Design is in Appendix 5 page 54

This experiment contains 3 conditions: A, M and P 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** move the cursor over this option and click to open the dialog.



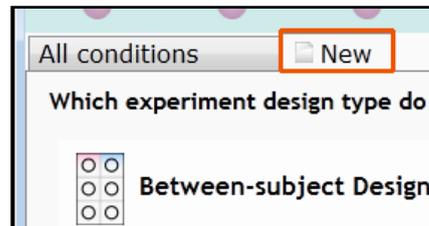
Give the new experimental design a name and then click **Create design**.



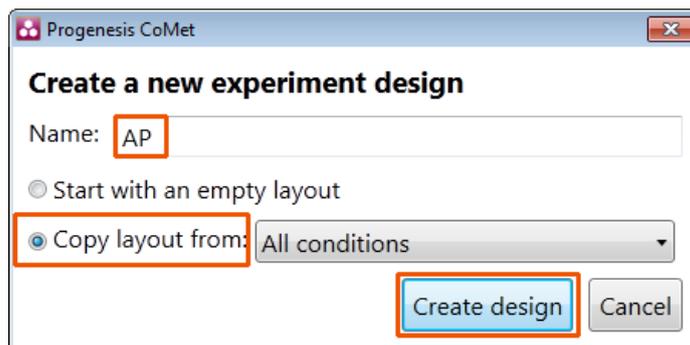
## To create a new condition

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. P) by over typing the default name.
6. Repeat steps 1 to 5 until all the runs are grouped into conditions.

To create another Design, for example comparing only two conditions, A and P, click on the **New** tab and click on Between-subject Design again.

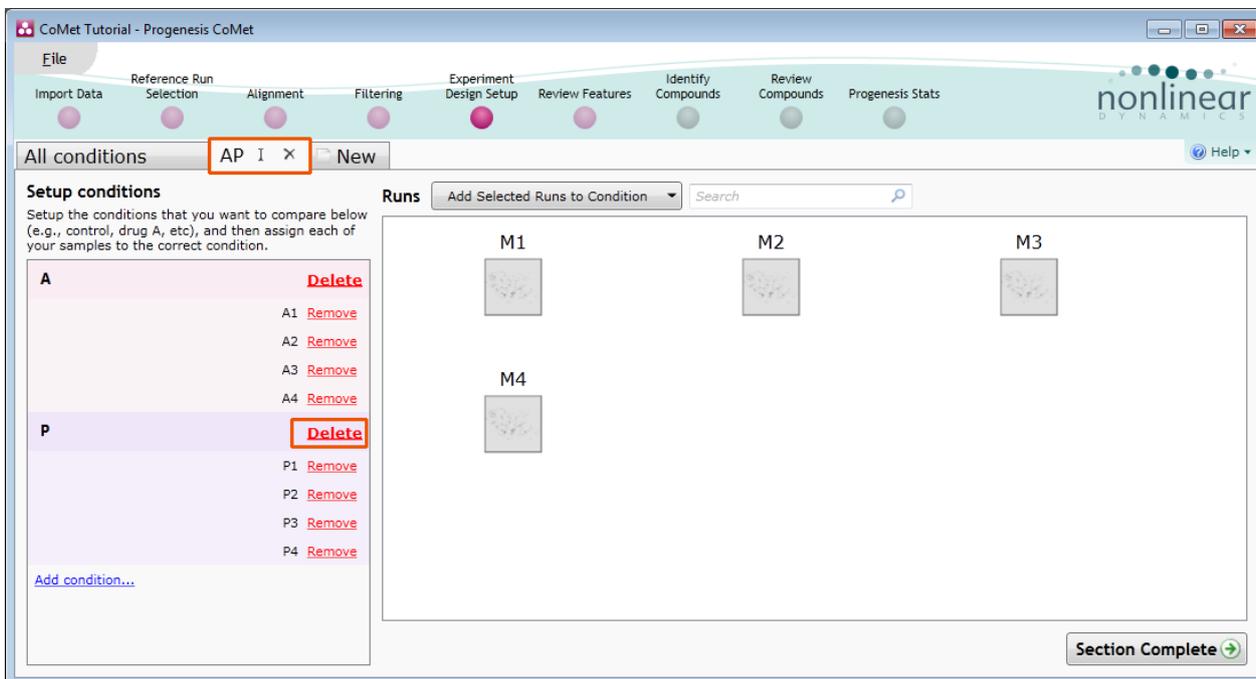


Give the new design a name, then tick the **Copy layout from** option and select the **AC** design.



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.



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 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 3 display modes: 1D, 2D and 3D and Feature Details controlled by the tabs on 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.

The screenshot displays the Progenesis CoMet software interface during the 'Review Features' stage. The top navigation bar includes steps: Import Data, Reference Run Selection, Alignment, Filtering, Experiment Design Setup, Review Features (active), Identify Compounds, Review Compounds, and Progenesis Stats. The 'Review Features' section shows a table of features (labeled A) with columns for #, Anova (p), Fold, Tag, and Notes. The 'Mass spectrum (5.132 min)' plot (labeled B) shows intensity vs m/z with a peak at 394. The 'Chromatogram (m/z = 394.0095)' plot (labeled C) shows retention time vs intensity. The 'Run: Abundance Profile' plot (labeled D) shows retention time vs m/z with a peak at 394. The 'Retention time (min)' vs 'm/z' plot (labeled E) shows retention time vs m/z with a peak at 394. The bottom left has tabs for '1D Display', '2D Montage', and '3D Montage'. A 'Section Complete' button is at the bottom right.

### 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 grouping.

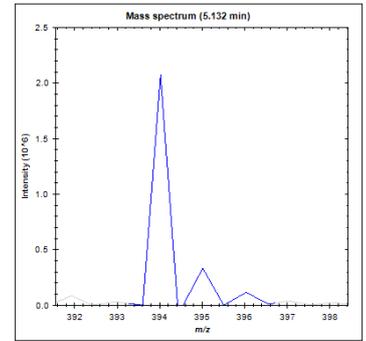
**Note:** A value of 'Infinity' in the **Fold** column indicates 'Presence/Absence'

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

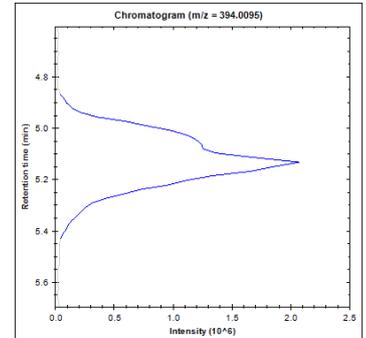
To highlight a group of features drag out a selection on the table.

## 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).

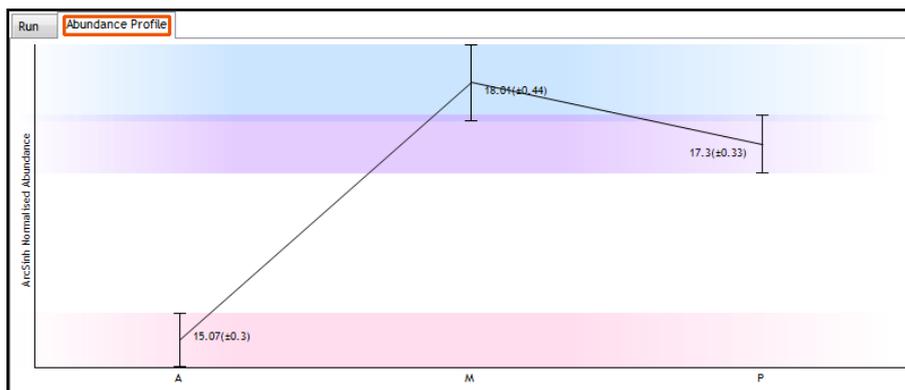
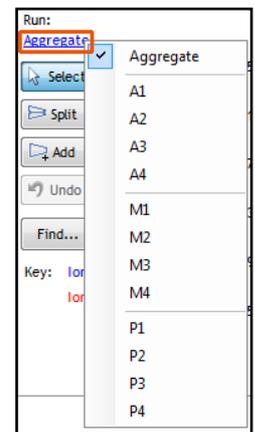


**Window D:** displays the details of the currently selected run. By default the selected run is an Aggregate of all the aligned runs.

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 27 for functional explanation).

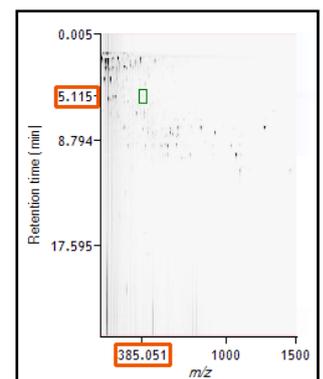
Clicking on the Abundance Profile tab in Window D shows the comparative behaviour of the feature across the various biological groups based on group average normalised volume. The error bars show +/- 3 standard errors.



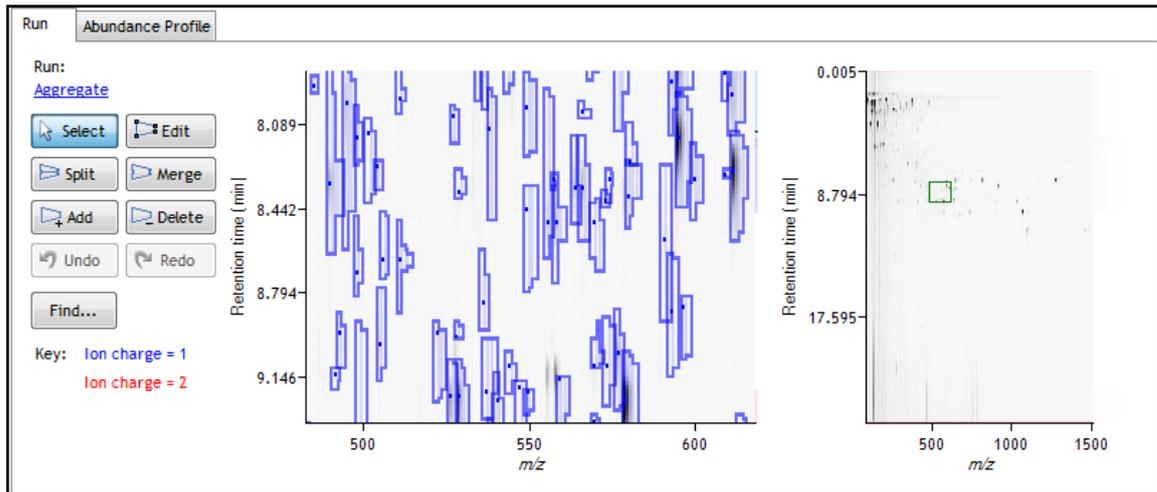
**Window E:** shows where the current feature is located on the run by means of the 'Green' rectangle.

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 a larger area on this view that will refocus the other windows



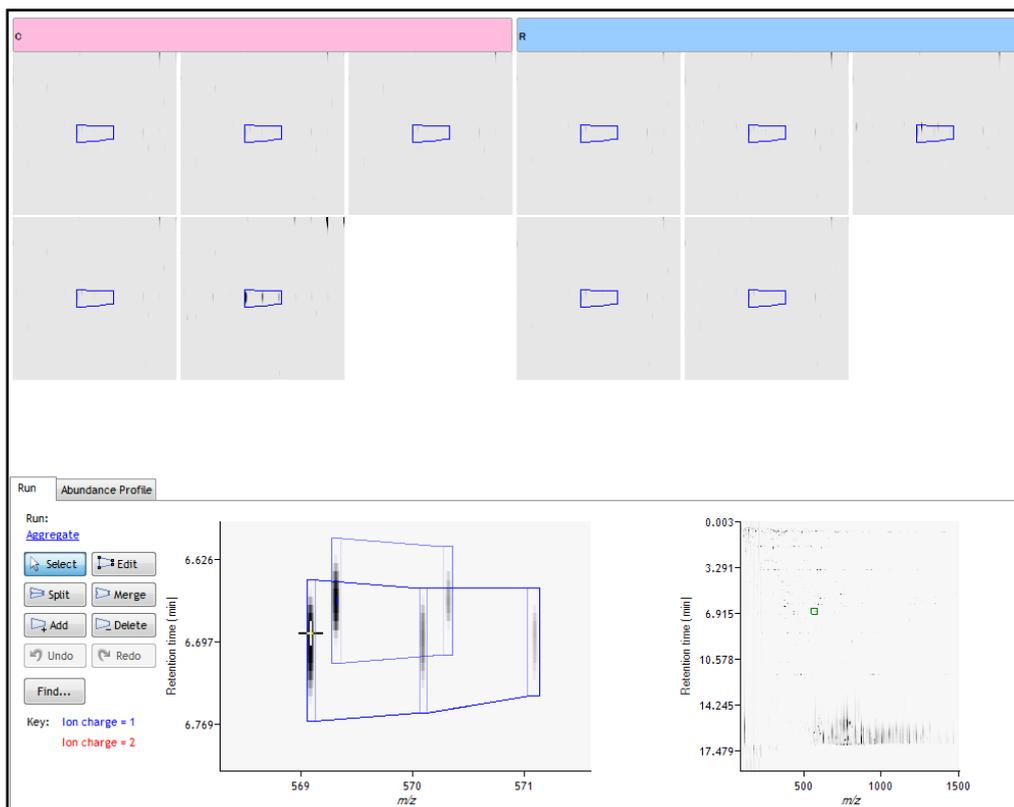
### The 2D Display

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

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

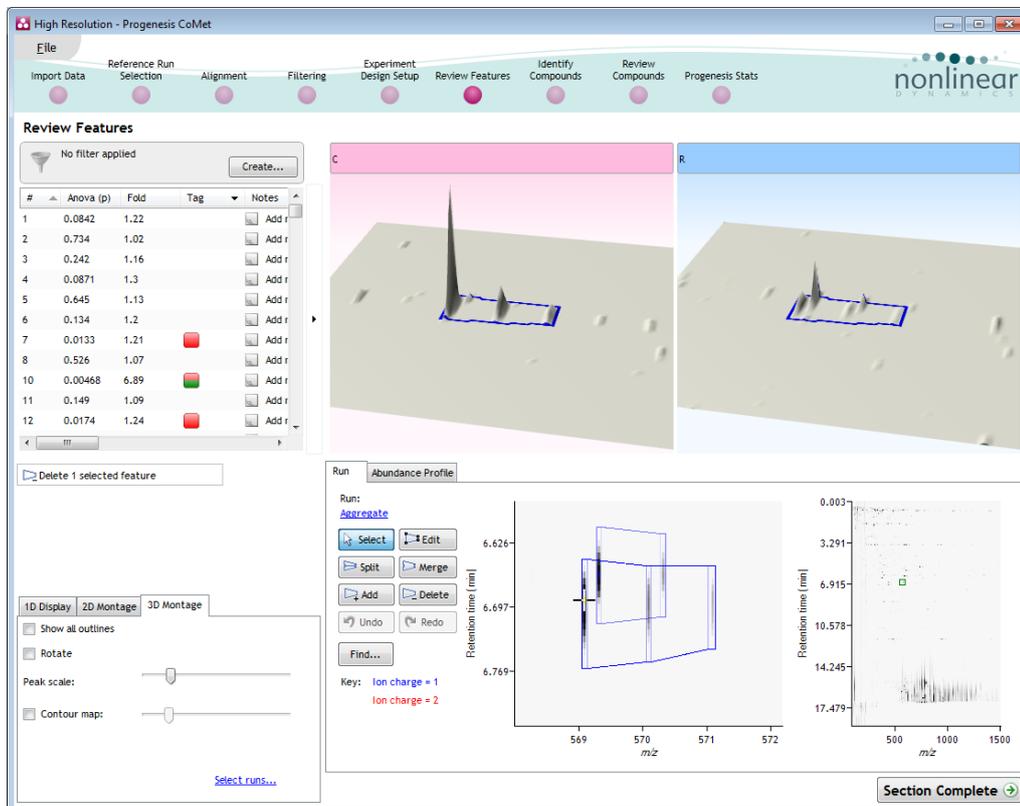
The appearance of the Montage (window B) is controlled by the panel on the bottom left of the display.

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 for an experiment using High resolution data..



### The 3D Display

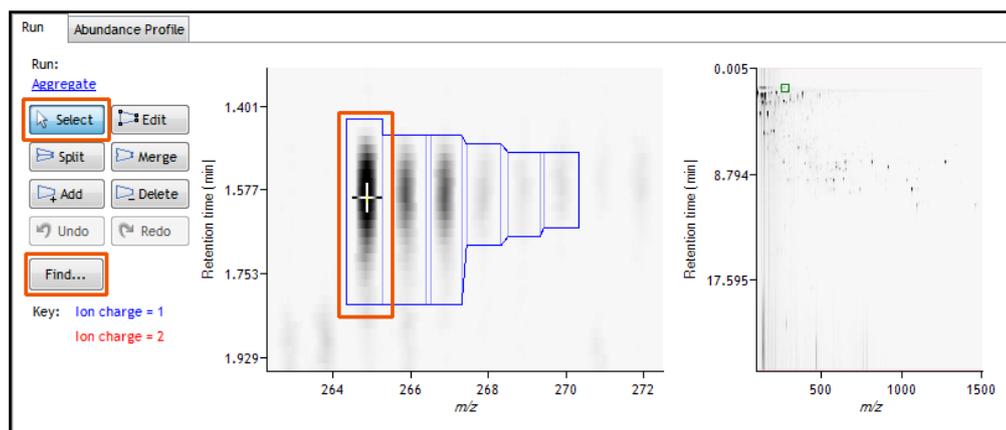
Window B changes into a 3D view by selecting the 3D Montage tab on the bottom left of the display.



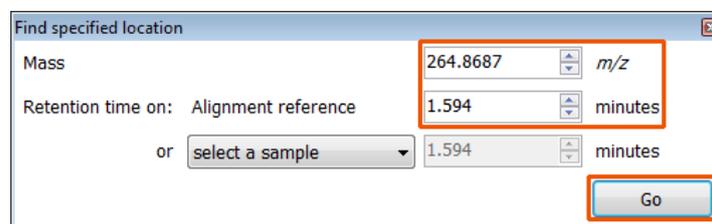
The number of 3D views displayed in the montage is controlled using the [Select runs](#) link on the 3D Montage tab. The views can be set to **Rotate** automatically or you can rotate them manually by clicking and dragging them with the mouse.

## Editing of features in the View Results stage

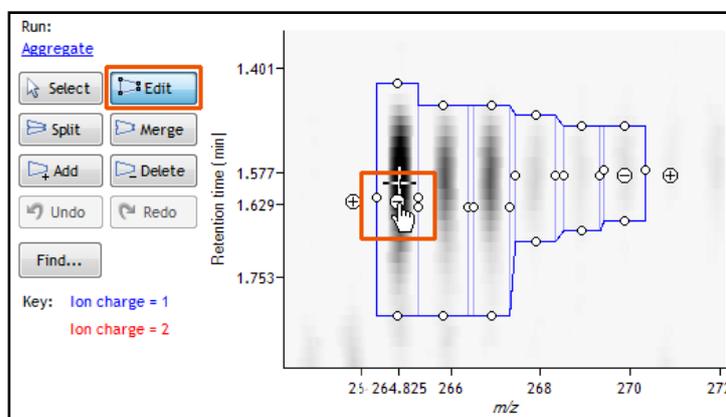
As an **example** of using the editing tools which are located on the left of the Run view, we will remove and add back the 'monoisotopic peak' for the detected feature selected below. A feature can be selected from the 'Features' list or located using the various views.



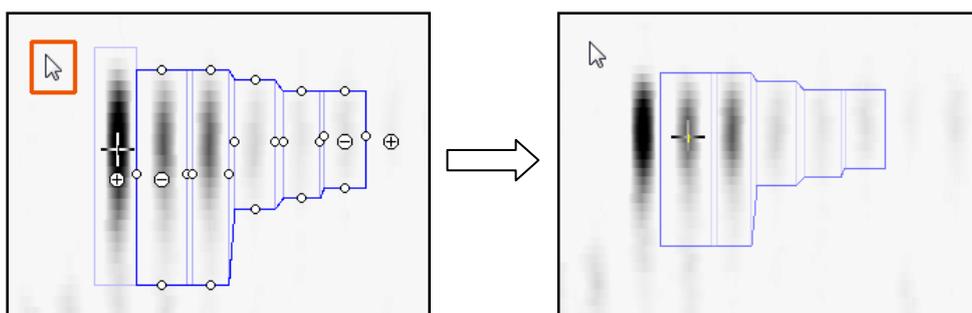
1. Locate the feature at approx 264.8687 m/z and 1.594 min using the **Find** tool.



2. Centre and enlarge the feature by dragging out an area around it. Then select the **Edit** tool and click on the feature to reveal the 'edit handles'

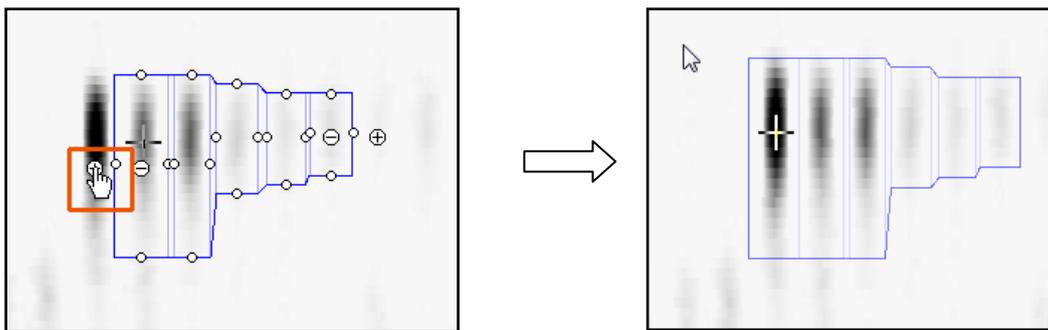


3. Click on the 'minus' handle over the monoisotopic peak to remove it.

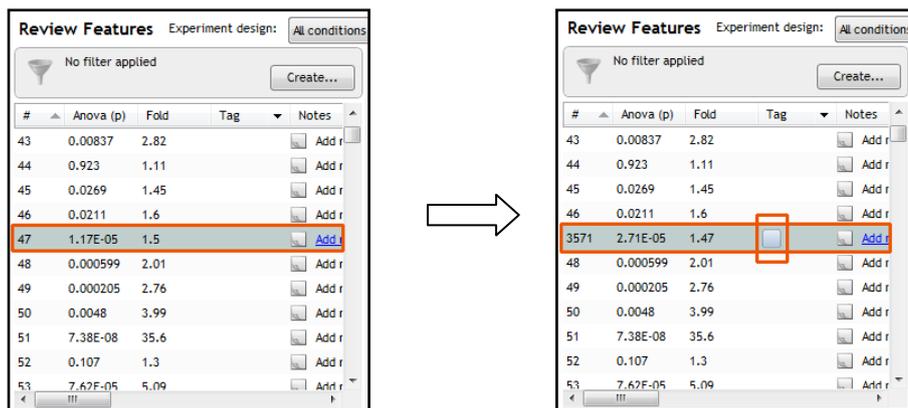


4. Click outside the boundary of the feature to update the view.

- 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.



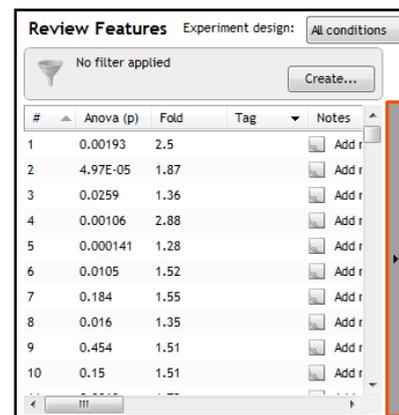
The other tools: split, merge, add and delete behave in a similar fashion and their use can be combined to achieve the desired results.

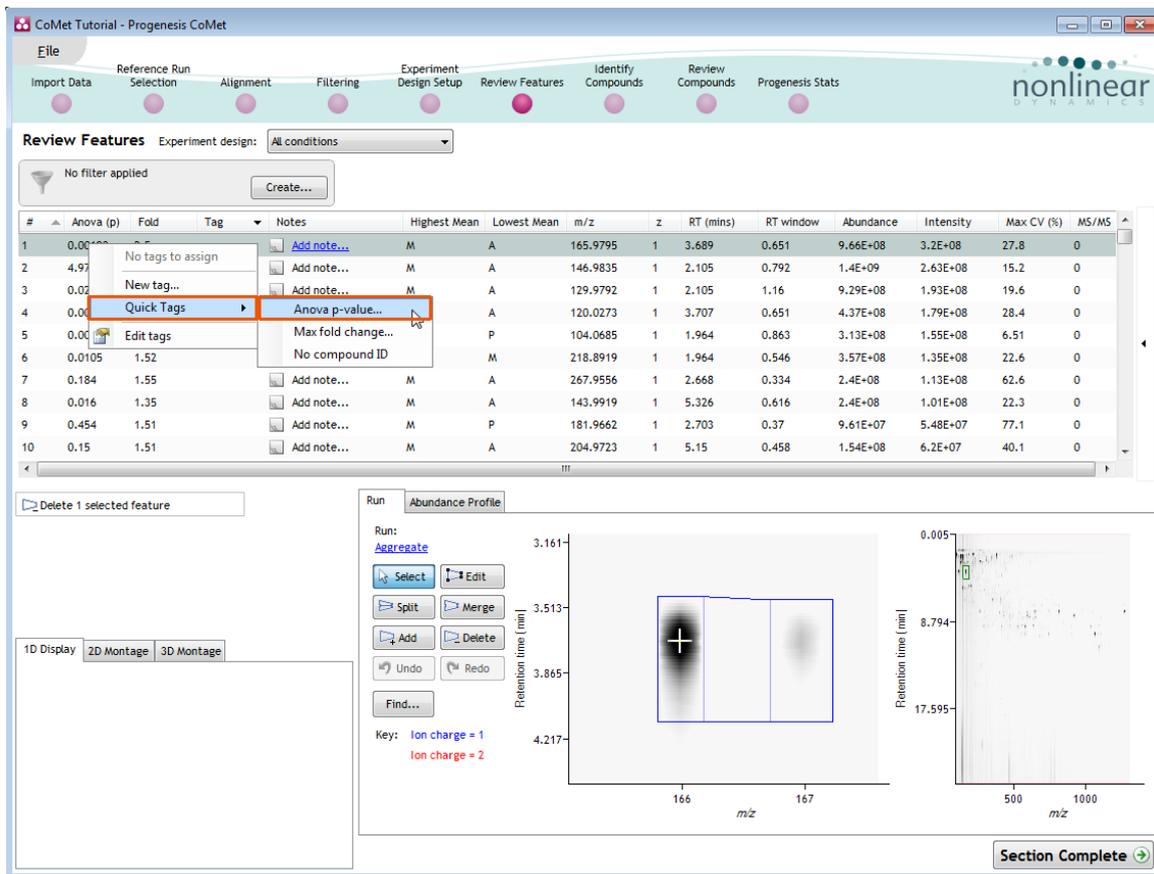
### Using 'Tagging' to filter data for identification

There are a number of ways to 'refine' your 'Ranked List' of analysed features before identifying the compounds. These make use of simple 'Tagging' tools that can be applied to the various Groupings created in Stage 6 (page 20).

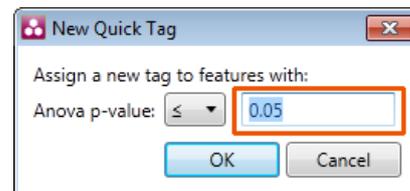
As an example of the use of tags we will set out to identify the compounds that are associated with a specific population of detected features. This population of features are required to show a significant (Anova p value < 0.05), 2 fold or greater increase in abundance for one of the conditions (A, M or P).

First expand the 'Features' table to show all the details by clicking on the 'Expander bar' to the right of the Review features table.

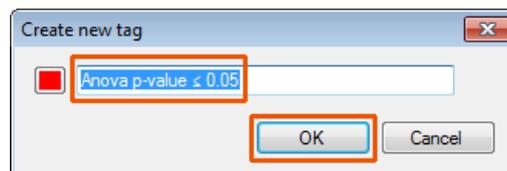




Right click on the table to open the 'Tag' menu. Select **Quick Tags** and then **Anova p-value** and accept the default value ( $\leq 0.05$ ).

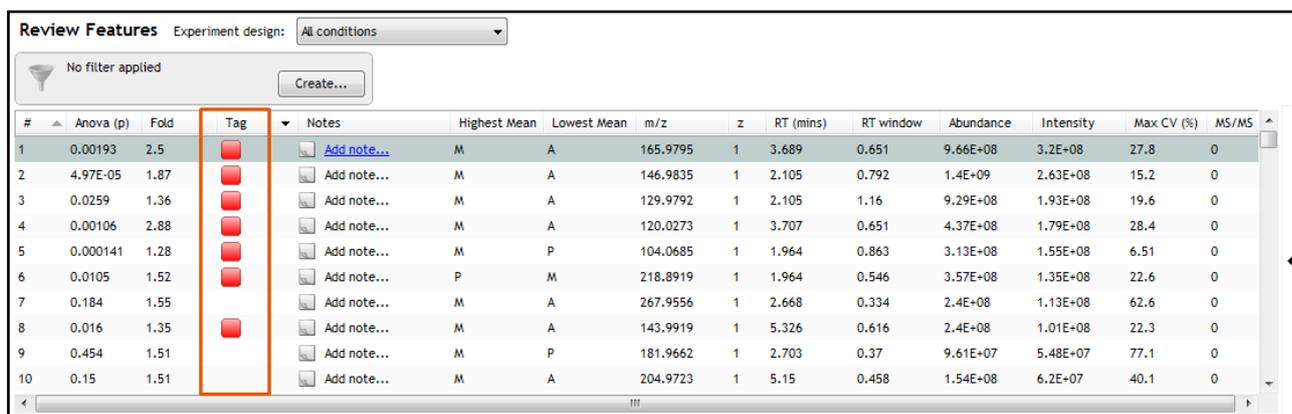


You will be offered a tag named **Anova p-value  $\leq 0.05$** .

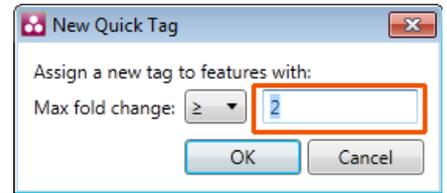


**Note:** you can change the name by overtyping.

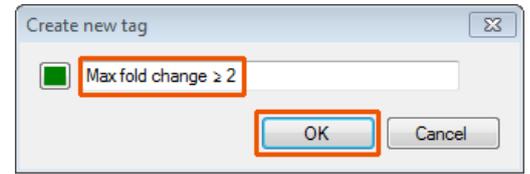
On pressing **OK** a tag appears in the table against all the features with an Anova p-value  $\leq 0.05$



To add a second 'Quick Tag' for those features with a Fold difference of 2 or greater, right click on the table to open the 'Tag' menu. Select **Quick Tags** and then **Max fold change** and accept the default value ( $\geq 2$ ).



You will be offered a tag named **Max fold change  $\geq 2$** .

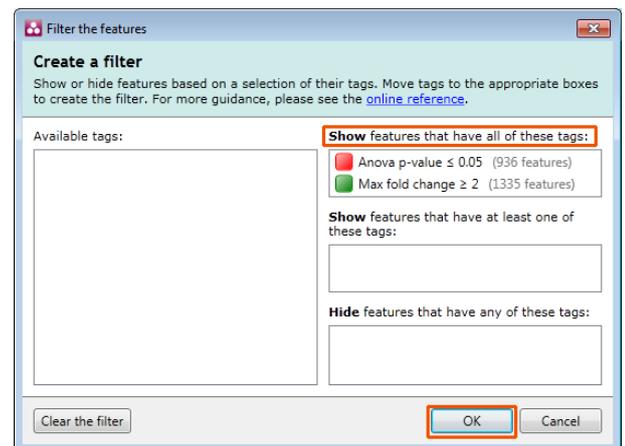
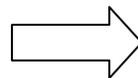
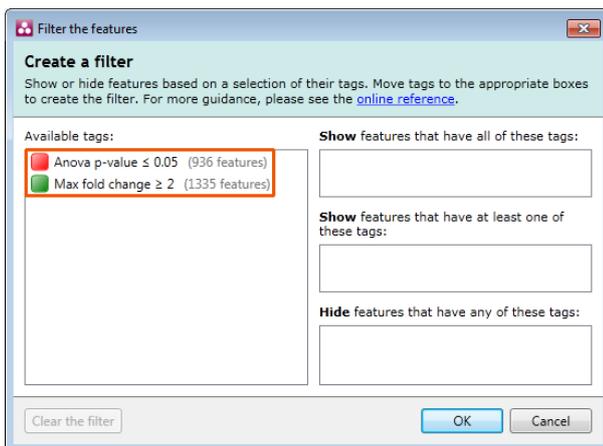


On pressing **OK** a second tag appears in the table against all the features with a Max fold change  $\geq 2$ .

| #  | Anova (p) | Fold | Tag | Notes       | Highest Mean | Lowest Mean | m/z      | z | RT (mins) | RT window | Abundance | Intensity | Max CV (%) | MS/MS |
|----|-----------|------|-----|-------------|--------------|-------------|----------|---|-----------|-----------|-----------|-----------|------------|-------|
| 1  | 0.00193   | 2.5  |     | Add note... | M            | A           | 165.9795 | 1 | 3.689     | 0.651     | 9.66E+08  | 3.2E+08   | 27.8       | 0     |
| 2  | 4.97E-05  | 1.87 |     | Add note... | M            | A           | 146.9835 | 1 | 2.105     | 0.792     | 1.4E+09   | 2.63E+08  | 15.2       | 0     |
| 3  | 0.0259    | 1.36 |     | Add note... | M            | A           | 129.9792 | 1 | 2.105     | 1.16      | 9.29E+08  | 1.93E+08  | 19.6       | 0     |
| 4  | 0.00106   | 2.88 |     | Add note... | M            | A           | 120.0273 | 1 | 3.707     | 0.651     | 4.37E+08  | 1.79E+08  | 28.4       | 0     |
| 5  | 0.000141  | 1.28 |     | Add note... | M            | P           | 104.0685 | 1 | 1.964     | 0.863     | 3.13E+08  | 1.55E+08  | 6.51       | 0     |
| 6  | 0.0105    | 1.52 |     | Add note... | P            | M           | 218.8919 | 1 | 1.964     | 0.546     | 3.57E+08  | 1.35E+08  | 22.6       | 0     |
| 7  | 0.184     | 1.55 |     | Add note... | M            | A           | 267.9556 | 1 | 2.668     | 0.334     | 2.4E+08   | 1.13E+08  | 62.6       | 0     |
| 8  | 0.016     | 1.35 |     | Add note... | M            | A           | 143.9919 | 1 | 5.326     | 0.616     | 2.4E+08   | 1.01E+08  | 22.3       | 0     |
| 9  | 0.454     | 1.51 |     | Add note... | M            | P           | 181.9662 | 1 | 2.703     | 0.37      | 9.61E+07  | 5.48E+07  | 77.1       | 0     |
| 10 | 0.15      | 1.51 |     | Add note... | M            | A           | 204.9723 | 1 | 5.15      | 0.458     | 1.54E+08  | 6.2E+07   | 40.1       | 0     |

Now to get the table to display **only** those features which satisfy the required criteria, i.e. those features that display a significant 2 fold or greater increase in Mean abundance, apply a filter..

To set up a filter click **Create** to open the Filter dialog and drag the new tags on to the **Show** features with all of these tags.



On clicking **OK** the table now displays only those features with the 2 tags.

**Note:** the Tag filter panel has changed, informing you that a filter is currently applied.

**Review Features** Experiment design: All conditions

Tag filter applied features may be hidden Edit...

| #  | Anova (p) | Fold | Tag | Notes       | Highest Mean | Lowest Mean | m/z      | z | RT (mins) | RT window | Abundance | Intensity | Max CV (%) | MS/MS |
|----|-----------|------|-----|-------------|--------------|-------------|----------|---|-----------|-----------|-----------|-----------|------------|-------|
| 1  | 0.00193   | 2.5  | 🟢🔴  | Add note... | M            | A           | 165.9795 | 1 | 3.689     | 0.651     | 9.66E+08  | 3.2E+08   | 27.8       | 0     |
| 4  | 0.00106   | 2.88 | 🟢🔴  | Add note... | M            | A           | 120.0273 | 1 | 3.707     | 0.651     | 4.37E+08  | 1.79E+08  | 28.4       | 0     |
| 13 | 0.000706  | 2.16 | 🟢🔴  | Add note... | A            | M           | 380.9249 | 1 | 2.034     | 0.863     | 3.44E+08  | 3.93E+07  | 27.9       | 0     |
| 15 | 3.68E-06  | 2.31 | 🟢🔴  | Add note... | A            | M           | 159.9706 | 1 | 2.914     | 0.528     | 8.75E+07  | 4.36E+07  | 13.6       | 0     |
| 16 | 0.00602   | 2.12 | 🟢🔴  | Add note... | M            | A           | 187.9568 | 1 | 5.15      | 0.458     | 9.95E+07  | 3.24E+07  | 32.1       | 0     |
| 19 | 0.000224  | 2    | 🟢🔴  | Add note... | A            | M           | 135.0651 | 1 | 23.23     | 5.14      | 1.19E+09  | 3.74E+07  | 17.7       | 0     |
| 25 | 0.0307    | 2.85 | 🟢🔴  | Add note... | A            | M           | 354.9437 | 1 | 6.523     | 0.51      | 1.27E+08  | 2.36E+07  | 52.5       | 0     |
| 26 | 8.26E-07  | 13   | 🟢🔴  | Add note... | A            | M           | 578.2682 | 1 | 9.234     | 0.493     | 2.79E+08  | 3.49E+07  | 28.9       | 0     |
| 35 | 4E-06     | 2.42 | 🟢🔴  | Add note... | A            | M           | 176.991  | 1 | 2.932     | 0.528     | 3.65E+07  | 1.77E+07  | 16.9       | 0     |
| 36 | 0.000282  | 2.83 | 🟢🔴  | Add note... | A            | P           | 118.031  | 1 | 2.14      | 0.669     | 3.9E+07   | 1.83E+07  | 25.9       | 0     |

With the filter still active click **Section Complete** to move to the Identify Compounds stage.

## Stage 8: Identify Compounds

The user guide now describes how to identify compounds from all or a subset of features.

CoMet Tutorial - Progenesis CoMet

File Reference Run Selection Alignment Filtering Experiment Design Setup Review Features Identify Compounds Review Compounds Progenesis Stats

**Identify Compounds** Help

Select your identification method:  
ms Exact mass and time (.csv)

- Export search details**  
Export the feature data to a spreadsheet to use when searching for compound identifications.  
Export CSV file...
- Perform the search**  
Open the file in Excel (or equivalent) then, using the mass and retention time information, enter compound data into the empty cells.
- Import search results**  
Once you have saved the CSV file, import that file to associate the data with the features.  
Import results CSV file...

The selected search method does not use MS/MS data  
If you want to use your MS/MS data as part of your search, please choose another search method using the options on the left

Tag filter applied features may be hidden Edit...

| #  | IDs | Mass     | Tag | Notes |
|----|-----|----------|-----|-------|
| 1  | 0   | 165.9795 | 🟢🔴  |       |
| 4  | 0   | 120.0273 | 🟢🔴  |       |
| 13 | 0   | 380.9249 | 🟢🔴  |       |
| 15 | 0   | 159.9706 | 🟢🔴  |       |
| 16 | 0   | 187.9568 | 🟢🔴  |       |
| 19 | 0   | 135.0651 | 🟢🔴  |       |
| 25 | 0   | 354.9437 | 🟢🔴  |       |
| 26 | 0   | 578.2682 | 🟢🔴  |       |

0 of 3568 features have been identified.

Clear all compound identifications Section Complete

**Note:** that the filter you applied at the previous stage is currently active.

This means that when you export the feature data to a spreadsheet, to use when searching compound identifications, only the features currently displayed in the table will be exported. In this case, only the features that display a significant 2 fold or greater increase in Mean abundance will be exported.

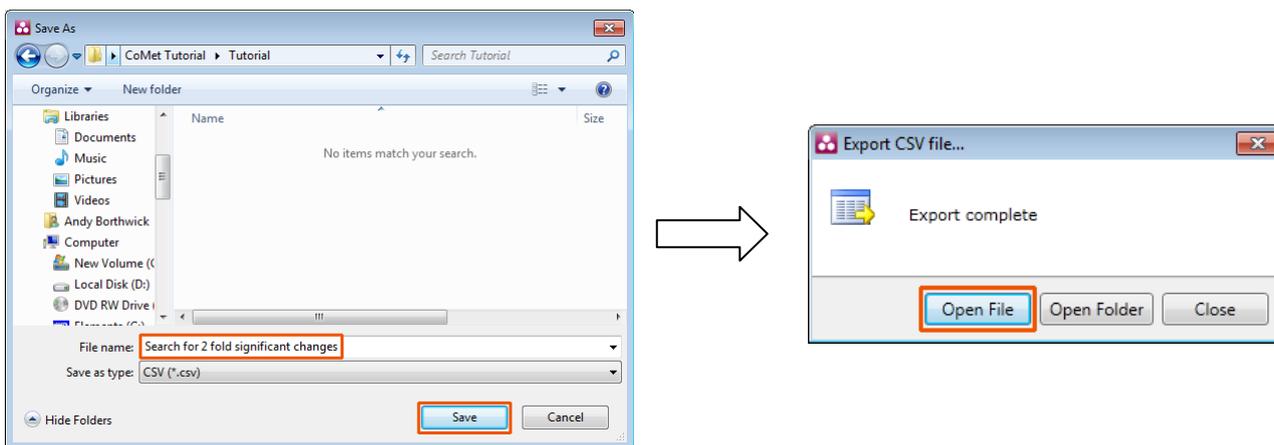
With the identification method set as **Exact mass and time (.csv)** click **Export CSV file...**

**Identify Compounds**

Select your identification method:  
ms Exact mass and time (.csv)

- Export search details**  
Export the feature data to a spreadsheet to use when searching for compound identifications.  
Export CSV file...

When you save the file you will get the option to open it.



The file opens displaying the list of m/z, retention time etc. for the exported features. Search results can be entered against these as they are acquired from various 'search engines' and the file saved.

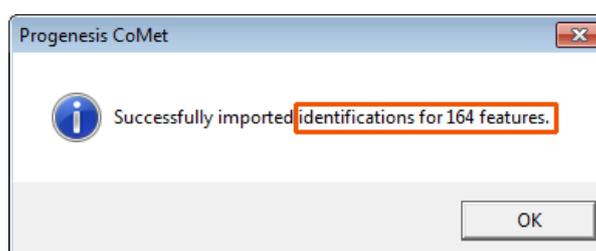
|   | A        | B                        | C            | D      | E     | F         | G       | H         | I       | J      | K   |
|---|----------|--------------------------|--------------|--------|-------|-----------|---------|-----------|---------|--------|-----|
| 1 | m/z      | Retention Time (minutes) | Feature Mass | Charge | Score | Theoretic | Compoun | Descripti | Formula | Adduct | URL |
| 2 | 165.9795 | 3.689                    | 165.9795028  | 1      |       |           |         |           |         |        |     |
| 3 | 120.0273 | 3.7066                   | 120.0273374  | 1      |       |           |         |           |         |        |     |
| 4 | 380.9249 | 2.034366667              | 380.9248998  | 1      |       |           |         |           |         |        |     |
| 5 | 159.9706 | 2.914483333              | 159.9705553  | 1      |       |           |         |           |         |        |     |
| 6 | 187.9568 | 5.149933333              | 187.9568304  | 1      |       |           |         |           |         |        |     |
| 7 | 135.0651 | 23.22976667              | 135.065065   | 1      |       |           |         |           |         |        |     |
| 8 | 354.9437 | 6.522816667              | 354.9437265  | 1      |       |           |         |           |         |        |     |

For the purposes of this User Guide an example file **Search\_results.csv** is available in the same folder as you restored the CoMet tutorial. Opening the file (see below) it shows data the entered details for Theoretical Mass, Compound ID, Description, Formula and Adduct.

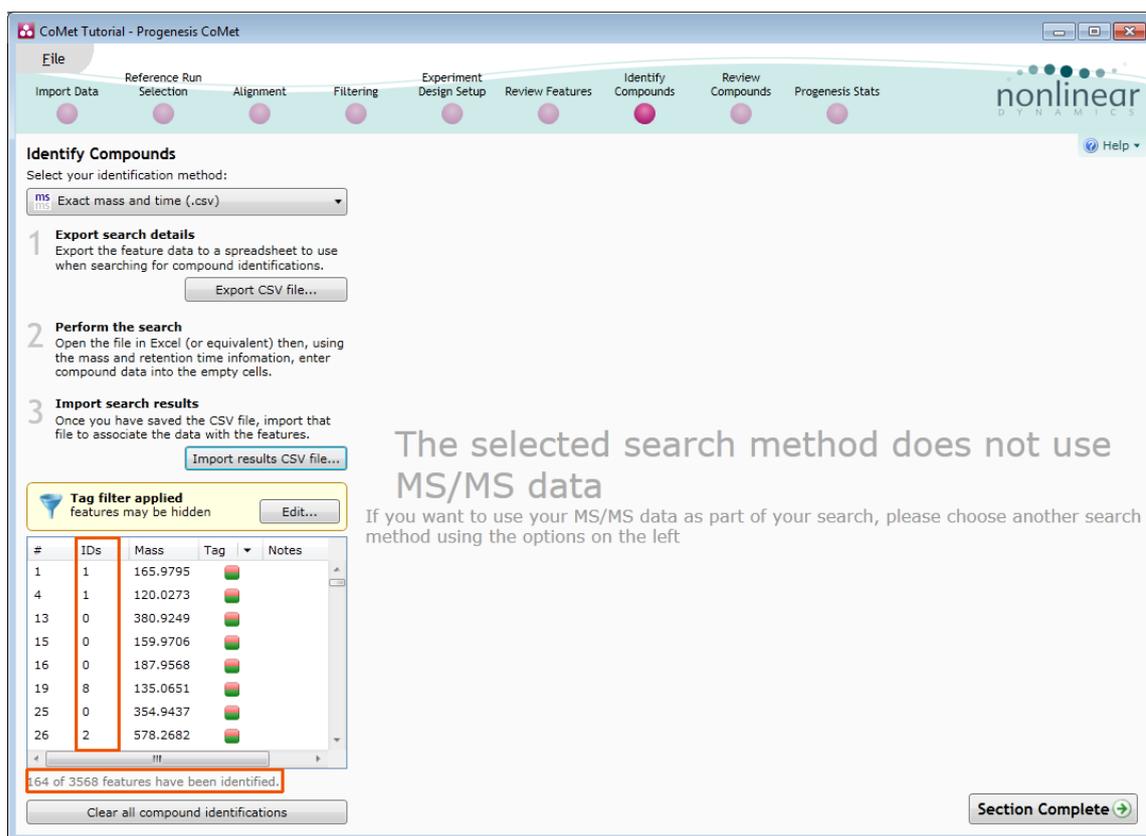
|     | A        | B         | C         | D      | E        | F         | G          | H         | I       | J       | K   | L | M |
|-----|----------|-----------|-----------|--------|----------|-----------|------------|-----------|---------|---------|-----|---|---|
| 1   | m/z      | Retention | Feature M | Charge | Score    | Theoretic | Compoun    | Descripti | Formula | Adduct  | URL |   |   |
| 96  | 135.0651 | 23.22977  | 135.0651  | 1      | 134.0592 | cmid:6249 | Dihydropt  | C6H6N4    |         | [M+H]+  |     |   |   |
| 97  | 135.7365 | 11.36352  | 135.7365  | 1      |          |           |            |           |         |         |     |   |   |
| 98  | 135.7858 | 10.32492  | 135.7858  | 1      |          |           |            |           |         |         |     |   |   |
| 99  | 136.9548 | 24.25112  | 136.9548  | 1      | 113.9639 | cmid:2923 | Dichloroe  | C2H4Cl2O  |         | [M+Na]+ |     |   |   |
| 100 | 136.9704 | 24.88513  | 136.9704  | 1      |          |           |            |           |         |         |     |   |   |
| 101 | 137.0383 | 22.66635  | 137.0383  | 1      | 114.0503 | cmid:4410 | DIALLYL S  | C6H10S    |         | [M+Na]+ |     |   |   |
| 102 | 137.0383 | 24.79705  | 137.0812  | 1      | 98.07316 | cmid:3652 | 3-propyl a | C6H10O    |         | [M+K]+  |     |   |   |
| 103 | 137.0383 | 24.79705  | 137.0812  | 1      | 98.07316 | cmid:3652 | 3-hexenal  | C6H10O    |         | [M+K]+  |     |   |   |

To import these search results click **Import results CSV file...** and select the **Search\_results.csv** file and then **open**.

A dialog will tell you the number of features that you have imported identifications for.



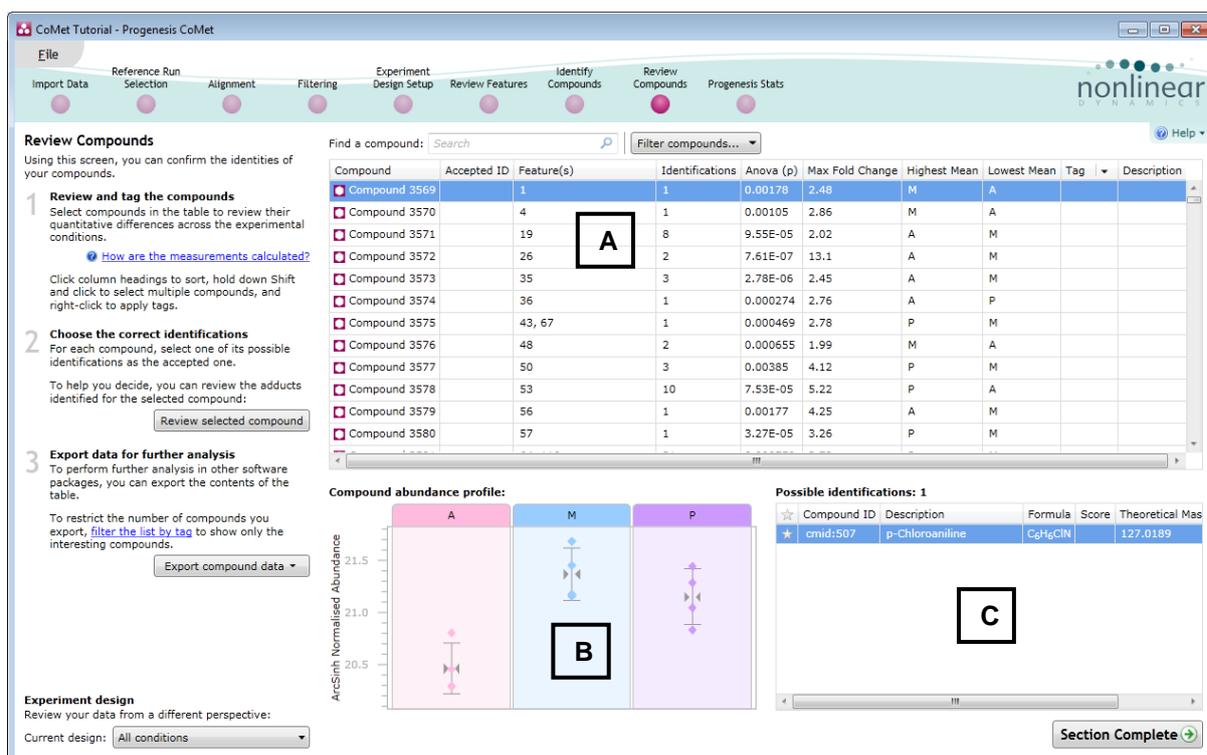
Also the Table will update the IDs column displaying the number of identifications imported for each feature.



Having imported the search results click **Section Complete** to move to the **Review Compounds** stage.

## Stage 9: Review Compounds

The Review Compounds opens allowing you to examine and confirm identities of the compounds.



**Window A** displays the main table of Compounds with identifications and those that remain Unknown after the identifications have been imported.

To concentrate on only the compounds with identifications, right click on the table and select **'Not identified'** from the **Quick Tags** menu.

**Note:** tags created at this stage are at the level of compounds not features

| Compound      | Accepted ID | Feature(s) | Identifications | Anova (p) | Max Fold Change | Highest Mean | Lowest Mean | Tag | Description |
|---------------|-------------|------------|-----------------|-----------|-----------------|--------------|-------------|-----|-------------|
| Compound 3725 |             | 3315       | 5               | 0.0442    | 2.1             | A            | M           |     |             |
| Compound 3726 |             | 3323       | 15              | 0.00225   | 2.44            | A            | M           |     |             |
| Compound 3727 |             | 3340       | 1               | 0.0324    | 2.22            | P            | M           |     |             |
| Compound 3728 |             | 3473       | 1               | 0.00794   | 2.95            | P            | M           |     |             |
| Compound 3729 |             | 3522       | 11              | 0.0288    | 3.04            | P            | M           |     |             |
| ? Unk         |             |            | 0               | 4.47E-05  | 1.85            | M            | A           |     |             |
| ? Unk         |             |            | 0               | 0.0361    | 1.35            | M            | A           |     |             |
| ? Unk         |             |            | 0               | 0.000366  | 1.24            | A            | P           |     |             |
| ? Unk         |             |            | 0               | 0.00711   | 1.57            | P            | M           |     |             |
| ? Unknown 7   | 7           |            | 0               | 0.186     | 1.54            | M            | A           |     |             |
| ? Unknown 8   | 8           |            | 0               | 0.0161    | 1.34            | P            | A           |     |             |
| ? Unknown 9   | 9           |            | 0               | 0.485     | 1.45            | M            | P           |     |             |

Accept the name, 'Not identified' offered by the software

| Compound      | Accepted ID | Feature(s) | Identifications | Anova (p) | Max Fold Change | Highest Mean | Lowest Mean | Tag | Description |
|---------------|-------------|------------|-----------------|-----------|-----------------|--------------|-------------|-----|-------------|
| Compound 3725 |             | 3315       | 5               | 0.0442    | 2.1             | A            | M           |     |             |
| Compound 3726 |             | 3323       | 15              | 0.00225   | 2.44            | A            | M           |     |             |
| Compound 3727 |             | 3340       | 1               | 0.0324    | 2.22            | P            | M           |     |             |
| Compound 3728 |             | 3473       | 1               | 0.00794   | 2.95            | P            | M           |     |             |
| Compound 3729 |             | 3522       | 11              | 0.0288    | 3.04            | P            | M           |     |             |
| ? Unknown 2   | 2           |            | 0               | 4.47E-05  | 1.85            | M            | A           | ●   |             |
| ? Unknown 3   | 3           |            | 0               | 0.0361    | 1.35            | M            | A           | ●   |             |
| ? Unknown 5   | 5           |            | 0               | 0.000366  | 1.24            | A            | P           | ●   |             |
| ? Unknown 6   | 6           |            | 0               | 0.00711   | 1.57            | P            | M           | ●   |             |
| ? Unknown 7   | 7           |            | 0               | 0.186     | 1.54            | M            | A           | ●   |             |
| ? Unknown 8   | 8           |            | 0               | 0.0161    | 1.34            | P            | A           | ●   |             |
| ? Unknown 9   | 9           |            | 0               | 0.485     | 1.45            | M            | P           | ●   |             |

The tag appears next to the compounds without identifications. To hide these 'Unknown' compounds click on **Filter compounds...**

**Filter the compounds**

**Create a filter**  
Show or hide compounds based on a selection of their tags. Move tags to the appropriate boxes to create the filter. For more guidance, please see the [online reference](#).

Available tags:

Show compounds that have all of these tags:

Show compounds that have at least one of these tags:

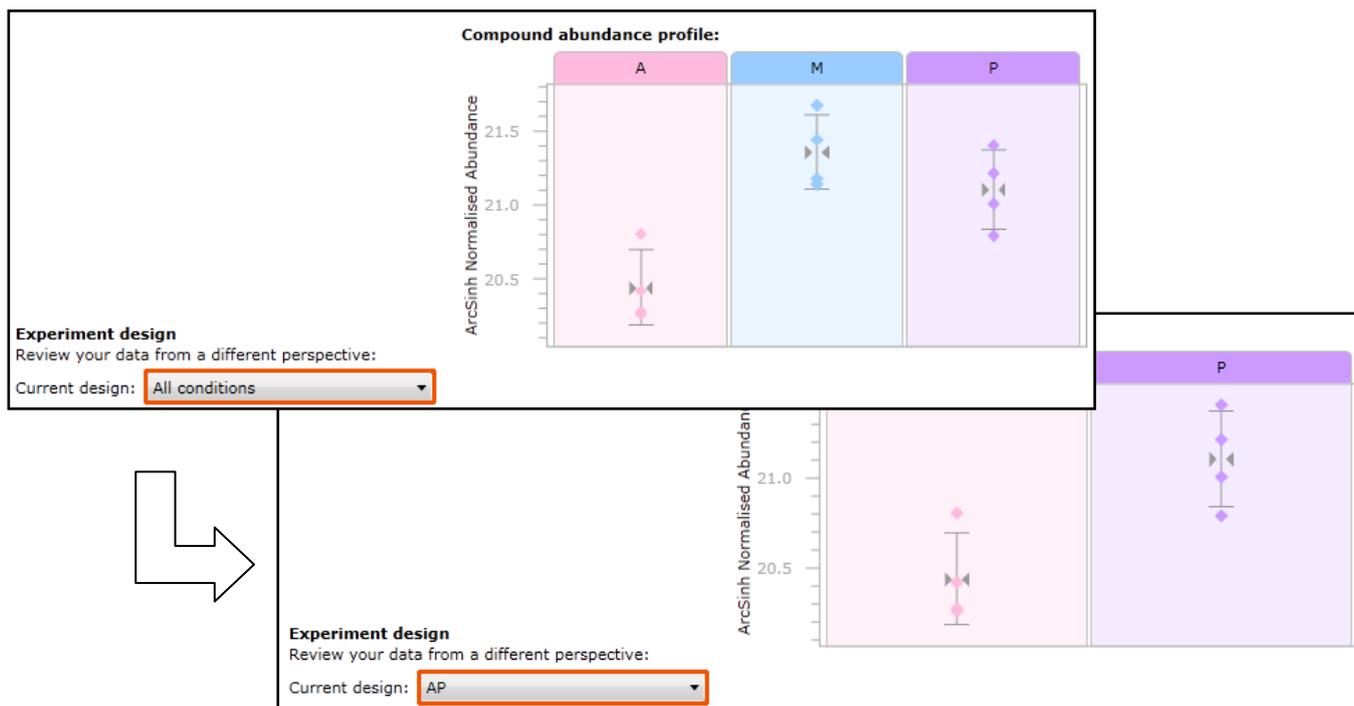
Hide compounds that have any of these tags:

● Not identified (3404 compounds)

Clear the filter OK Cancel

Drag the tag for **Not identified** on to the panel **Hide** compounds that have any of these tags:

**Window B** displays the Abundance profile for the selected Compound in Window A across the conditions for the current design. Changing the Design will update windows A and B.



**Window C** displays the **Possible identifications** for the selected Compound in Window A for example compare the possible identifications for compounds 3569 and 3573

| Compound      | Accepted ID | Feature(s) | Identifications | Anova (p) | Max Fold Change | Highest Mean | Lowest Mean | Tag | Description |
|---------------|-------------|------------|-----------------|-----------|-----------------|--------------|-------------|-----|-------------|
| Compound 3569 |             | 1          | 1               | 0.00178   | 2.48            | M            | A           |     |             |
| Compound 3570 |             | 4          | 1               | 0.00105   | 2.86            | M            | A           |     |             |
| Compound 3571 |             | 19         | 8               | 9.55E-05  | 2.02            | A            | M           |     |             |
| Compound 3572 |             | 26         | 2               | 7.61E-07  | 13.1            | A            | M           |     |             |
| Compound 3573 |             | 35         | 3               | 2.78E-06  | 2.45            | A            | M           |     |             |
| Compound 3574 |             | 36         | 1               | 0.000274  | 2.76            | A            | P           |     |             |

**Possible identifications: 1**

| ☆ | Compound ID | Description     | Formula                           | Score | Theoretical |
|---|-------------|-----------------|-----------------------------------|-------|-------------|
| ★ | cmid:507    | p-Chloroaniline | C <sub>6</sub> H <sub>6</sub> ClN |       | 127.0189    |

**Possible identifications: 3**

| ☆ | Compound ID | Description         | Formula  | Score | Theoretical |
|---|-------------|---------------------|--|-------|-------------|
| ★ | cmid:63256  | Propanoyl phosphate | C <sub>3</sub> H <sub>7</sub> O <sub>5</sub> P |       | 154.0031    |
| ☆ | cmid:45437  | 2-Pyridylthioamide  | C <sub>6</sub> H <sub>6</sub> N <sub>2</sub> S |       | 138.0252    |
| ☆ | cmid:45441  | 4-Pyridylthioamide  | C <sub>6</sub> H <sub>6</sub> N <sub>2</sub> S |       | 138.0252    |

### Setting a Compound's identity

Where a number of different compound identities have been found as possible identifications for the same set of features, those features are considered to be the same compound where the true identity of the compound has yet to be determined.

To select the 'true' identity from the possible identifications click on the 'star' icon to the left of the compound ID in the Possible identifications (Window C)

**Possible identifications: 3**

| ☆ | Compound ID | Description         | Formula  | Score | Theoretical Mas |
|---|-------------|---------------------|--|-------|-----------------|
| ★ | cmid:63256  | Propanoyl phosphate | C <sub>3</sub> H <sub>7</sub> O <sub>5</sub> P |       | 154.0031        |
| ☆ | cmid:45437  | 2-Pyridylthioamide  | C <sub>6</sub> H <sub>6</sub> N <sub>2</sub> S |       | 138.0252        |
| ☆ | cmid:45441  | 4-Pyridylthioamide  | C <sub>6</sub> H <sub>6</sub> N <sub>2</sub> S |       | 138.0252        |

This is the accepted identification of the compound

As an example select the true identity for Compound 3573.

Once you have clicked on the 'star' holding the cursor over the 'star' will confirm that this is the 'accepted' identity of the compound. To remove or change the compound identity, click on it again or the star for another identity.

**Note:** once you have selected the 'true' identity, the **Accepted ID** and **Description** updates in the Compound table (Window A)

Find a compound:  Filter compounds... **Filter is active** Help

| Compound  | Accepted ID | Feature(s) | Identifications | Anova (p) | Max Fold Change | Highest Mean | Lowest Mean | Tag | Description |
|---|-------------|------------|-----------------|-----------|-----------------|--------------|-------------|-----|-------------|
| <input type="checkbox"/> Compound 3569            |             | 1          | 1               | 0.00178   | 2.48            | M            | A           |     |             |
| <input type="checkbox"/> Compound 3570            |             | 4          | 1               | 0.00105   | 2.86            | M            | A           |     |             |
| <input type="checkbox"/> Compound 3571            |             | 19         | 8               | 9.55E-05  | 2.02            | A            | M           |     |             |
| <input type="checkbox"/> Compound 3572            |             | 26         | 2               | 7.61E-07  | 13.1            | A            | M           |     |             |
| <input checked="" type="checkbox"/> Compound 3573 | cmid:63256  | 35         | 3               | 2.78E-06  | 2.45            | A            | M           |     | Propanoyl   |
| <input type="checkbox"/> Compound 3574            |             | 36         | 1               | 0.000274  | 2.76            | A            | P           |     |             |
| <input type="checkbox"/> Compound 3575            |             | 43, 67     | 1               | 0.000469  | 2.78            | P            | M           |     |             |

**Note:** the red text to the right of the **Filter compounds...** button indicating that a **Filter is active** this is warning you that you have an active 'Compounds' filter. In this case, the 'Not identified' tag filter that you applied in the previous section which hides all 'Unknown' compounds from the table.

## Reviewing identified Compound adducts and their Abundance profiles

To examine the compound adducts and their abundance profiles select the compound and click **Review selected compound**.

**Review Compounds**  
Using this screen, you can confirm the identities of your compounds.

**1 Review and tag the compounds**  
Select compounds in the table to review their quantitative differences across the experimental conditions.  
[How are the measurements calculated?](#)  
Click column headings to sort, hold down Shift and click to select multiple compounds, and right-click to apply tags.

**2 Choose the correct identifications**  
For each compound, select one of its possible identifications as the accepted one.  
To help you decide, you can review the adducts identified for the selected compound:  
**Review selected compound**

Find a compound:  Filter compounds... **Filter is active** Help

| Compound  | Accepted ID | Feature(s) | Identifications | Anova (p) | Max Fold Change | Highest Mean | Lowest Mean | Tag | Description |
|---|-------------|------------|-----------------|-----------|-----------------|--------------|-------------|-----|-------------|
| <input type="checkbox"/> Compound 3569            |             | 1          | 1               | 0.00178   | 2.48            | M            | A           |     |             |
| <input type="checkbox"/> Compound 3570            |             | 4          | 1               | 0.00105   | 2.86            | M            | A           |     |             |
| <input type="checkbox"/> Compound 3571            |             | 19         | 8               | 9.55E-05  | 2.02            | A            | M           |     |             |
| <input type="checkbox"/> Compound 3572            |             | 26         | 2               | 7.61E-07  | 13.1            | A            | M           |     |             |
| <input checked="" type="checkbox"/> Compound 3573 | cmid:63256  | 35         | 3               | 2.78E-06  | 2.45            | A            | M           |     | Propanoyl p |
| <input type="checkbox"/> Compound 3574            |             | 36         | 1               | 0.000274  | 2.76            | A            | P           |     |             |
| <input checked="" type="checkbox"/> Compound 3575 |             | 43, 67     | 1               | 0.000469  | 2.78            | P            | M           |     |             |
| <input type="checkbox"/> Compound 3576            |             | 48         | 2               | 0.000655  | 1.99            | M            | A           |     |             |
| <input type="checkbox"/> Compound 3577            |             | 50         | 3               | 0.00385   | 4.12            | P            | M           |     |             |
| <input type="checkbox"/> Compound 3578            |             | 53         | 10              | 7.53E-05  | 5.22            | P            | A           |     |             |
| <input type="checkbox"/> Compound 3579            |             | 56         | 1               | 0.00177   | 4.25            | A            | M           |     |             |
| <input type="checkbox"/> Compound 3580            |             | 57         | 1               | 3.27E-05  | 3.26            | P            | M           |     |             |

This will open the Reviewing for Compound 3575

**Reviewing:  3575**  
Using this screen, you can assess the reliability of your compound identifications.

**1 Select a compound identification**  
Select compound identifications below to view the identified adducts and their abundance profiles.

**2 Review adducts and their expression**  
Any adduct whose abundance profile does not match the majority for the selected compound may represent a misidentification.

This compound has 1 possible identifications:

| Compound ID                                    | Description   |
|--|---------------|
| <input checked="" type="checkbox"/> cmid:58309 | Endomorphin-2 |

Formula: C<sub>32</sub>H<sub>37</sub>N<sub>5</sub>O<sub>5</sub>

**Experiment design**  
Review your data from a different perspective:  
Current design:

**Back to all compounds** **Export compound adducts...** **Remove feature from compound**

| Adduct              | Feature | Tag | Score | Anova (p) | Max Fold Change | Highest Mean | Lowest Mean | Abundance | Max CV (%) | Retention Time | m/z      | Charge |
|---------------------|---------|-----|-------|-----------|-----------------|--------------|-------------|-----------|------------|----------------|----------|--------|
| [M+Na] <sup>+</sup> | 43      |     |       | 0.00635   | 2.9             | P            | M           | 1.371E+08 | 39.29      | 8.142          | 594.2613 | 1      |
| [M+Na] <sup>+</sup> | 67      |     |       | 4.49E-07  | 2.72            | A            | M           | 5.617E+07 | 14.62      | 7.79           | 594.2603 | 1      |

**Standardised Abundance Profiles**

The table currently displays the abundance profiles for the compound adducts. In order to provide an accurate measure of a compounds abundance features corresponding to Adducts with 'unexpected' abundance profiles may be removed from the compound by selecting the adduct and then clicking **Remove feature from compound**.

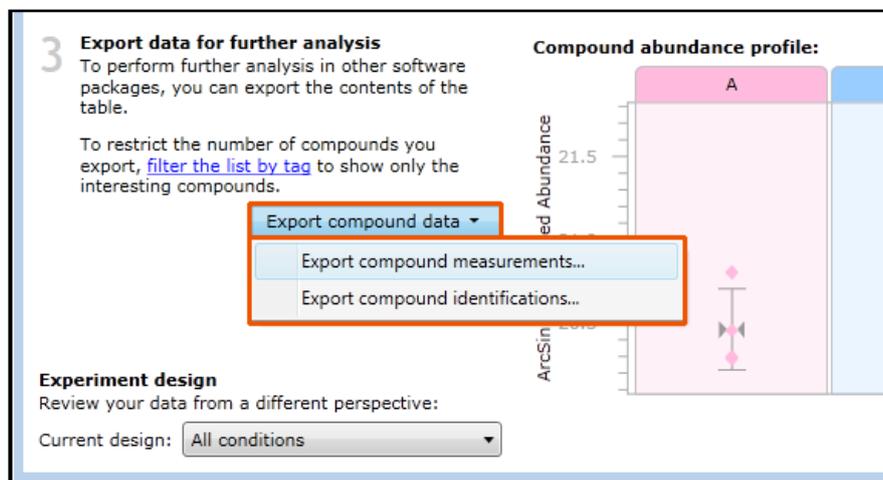
**Note:** if you cannot resolve the behaviour of the adducts, then delete all of the adducts.

Additional help on determining the identity of a compound is provided Appendix 6 page 56

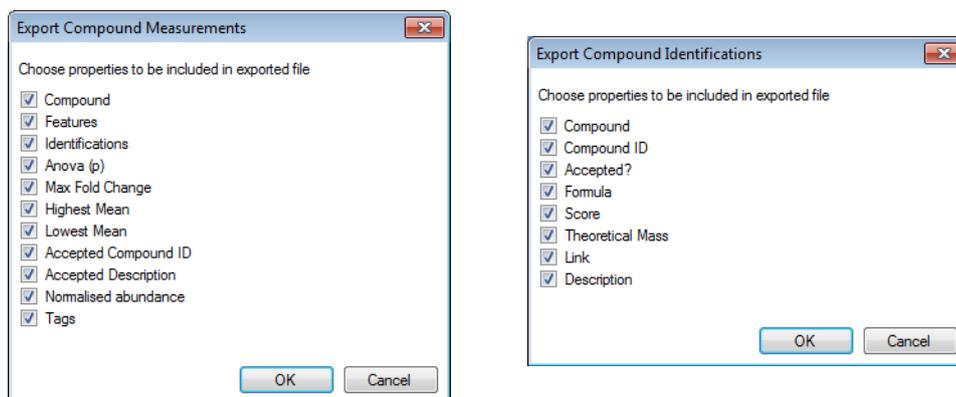
**Note:** You can also assign the compounds 'true' identity in this view using the table of possible identifications to the left of the Abundance profile

## Exporting compound data

Click **Export compound data** to export the data in a csv file format. You can either export the **compound measurements** or the **compound identifications**.

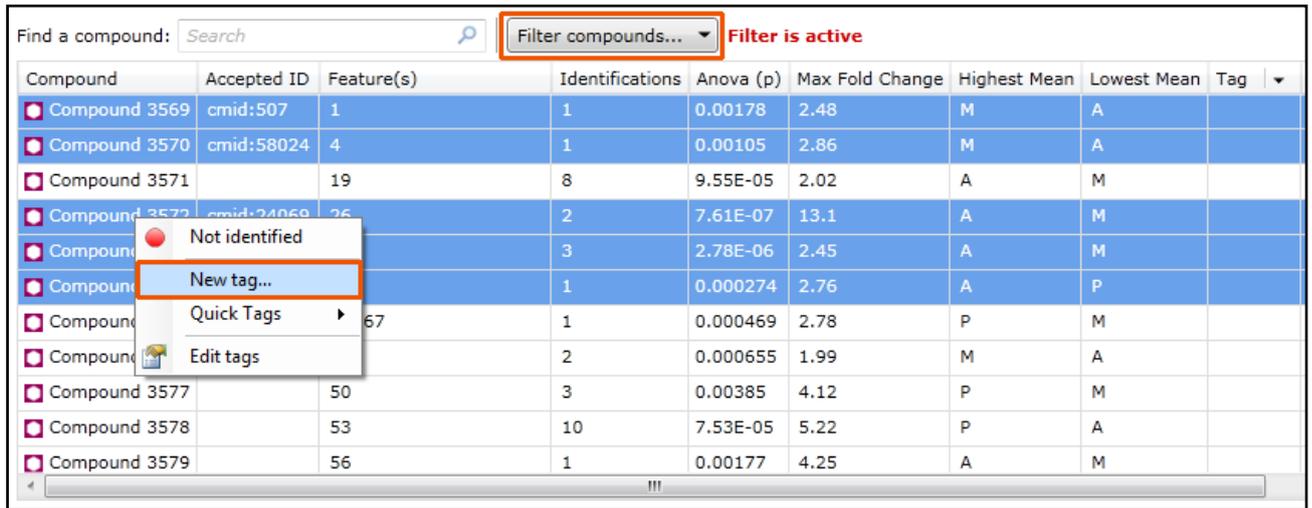


Dialogs allow you to control the export content and the use of tag filters will control the Compounds that you report on.

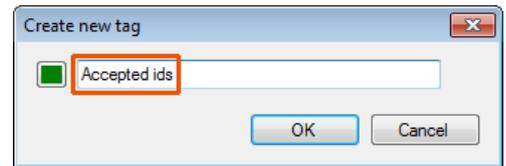


For example let us create a tag for the 5 compounds where the identity has been accepted.

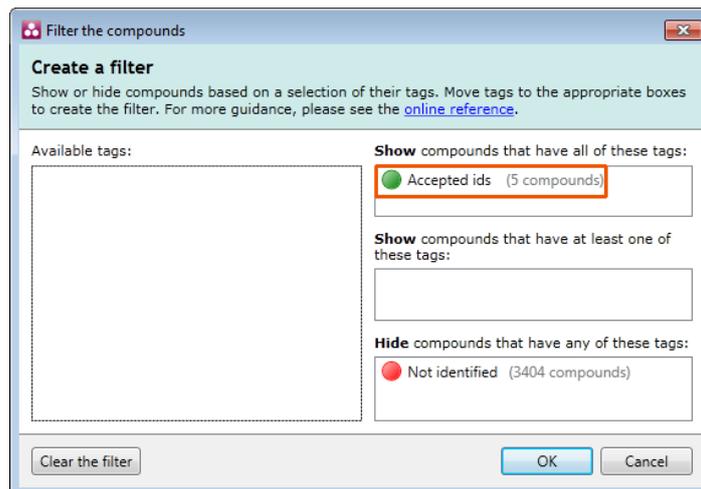
First highlight the 5 compounds in the list and right click. Select



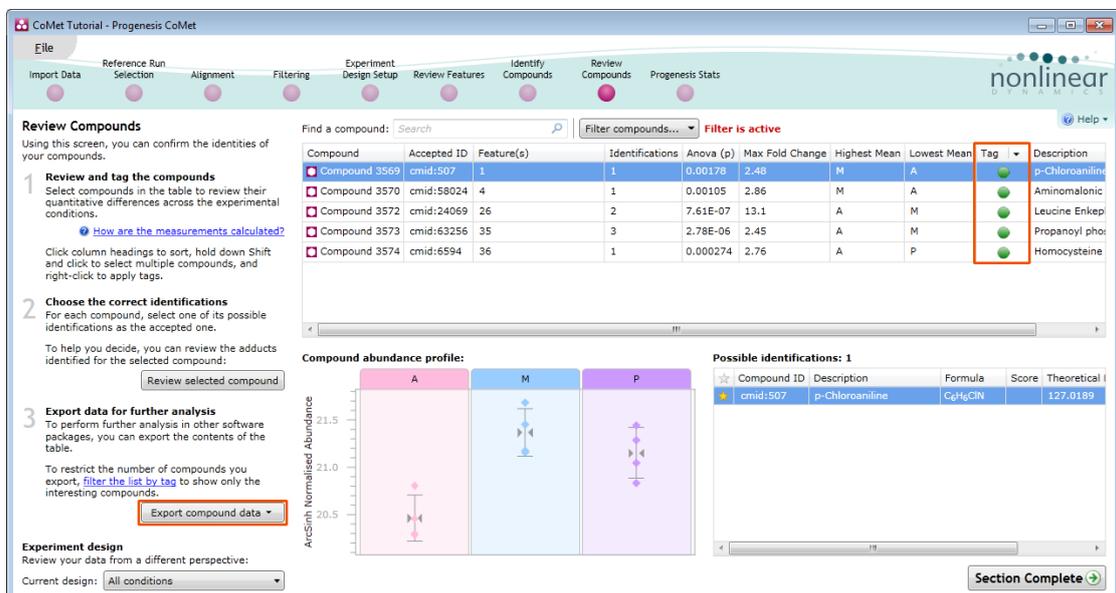
Give the new 'Compound' tag a name for example **Accepted ids**



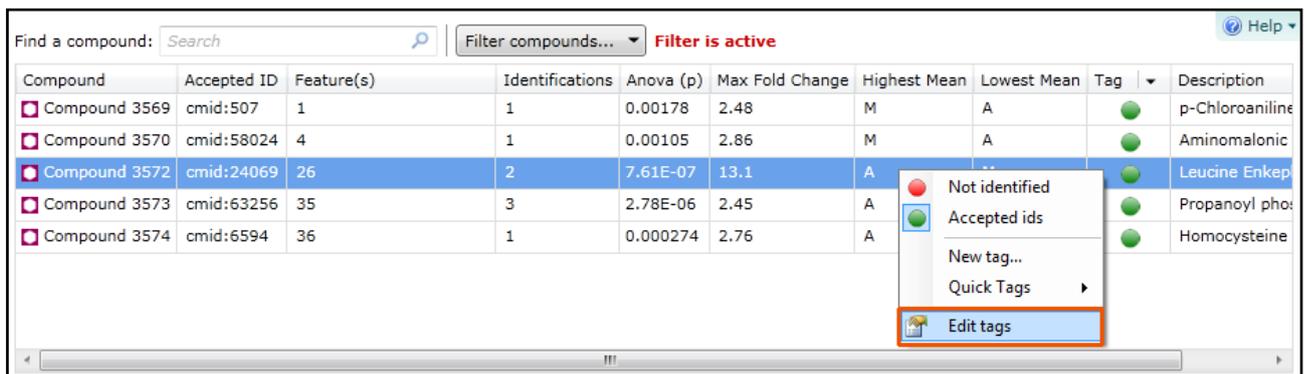
Now **Filter compounds** so that the table is only displaying the 5 compounds that have Accepted ids.



Now, when you click **Export compound data** you only export details for these tagged compounds..

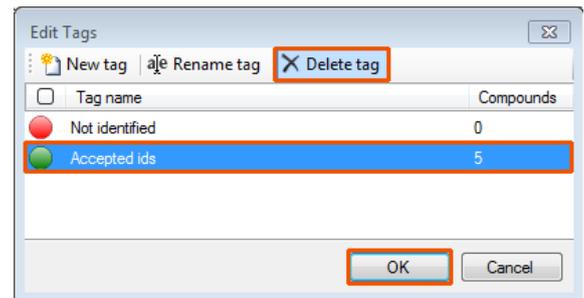


To delete a tag right click on the table and select **Edit tags**.

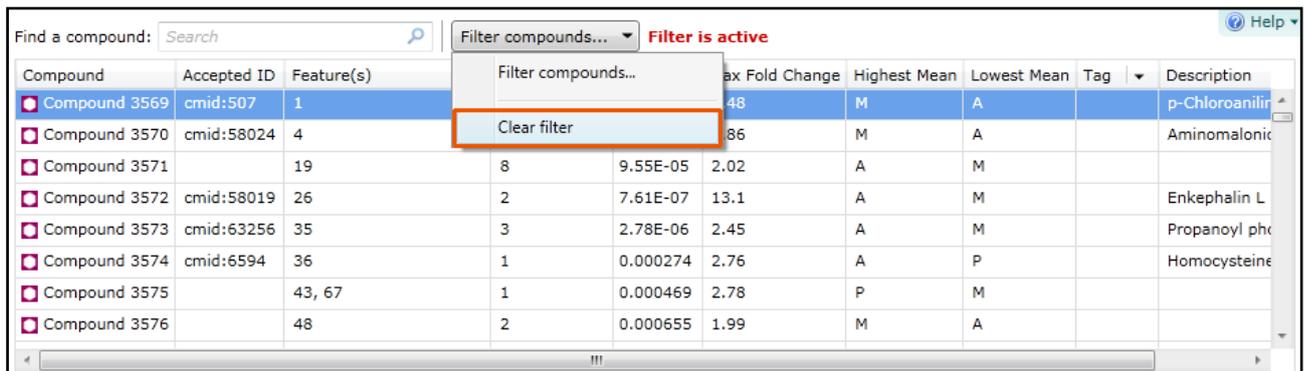


Highlight the tag you wish to delete and click on Delete tag selected tags.

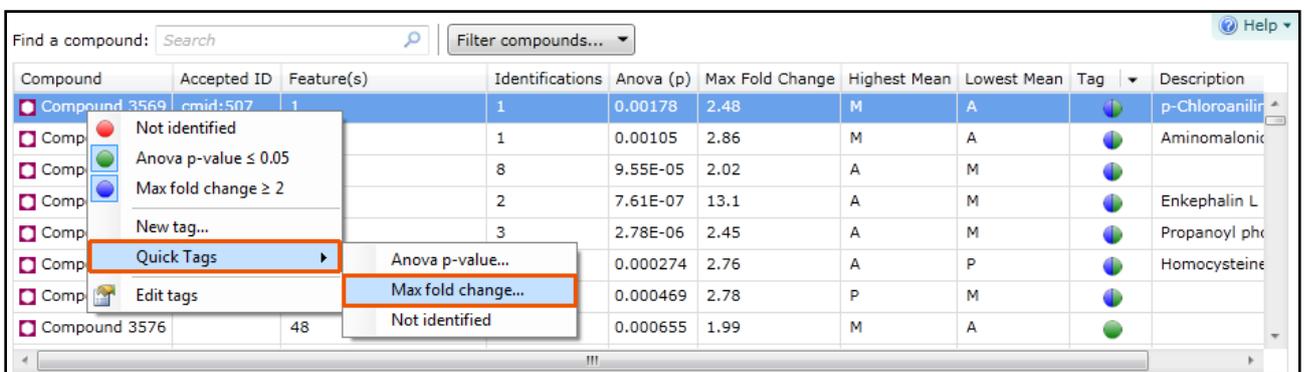
**Note:** doing this with a tag filter active for that tag will cause the hidden rows to reappear



Now clear any active filters by clicking on **Filter compounds** and selecting **Clear filter**.



Create two **Compound** tags using the Quick Tags for those compounds that have an **Anova** p-value  $\leq 0.05$  and a **Max fold change**  $\geq 2$  as shown below.

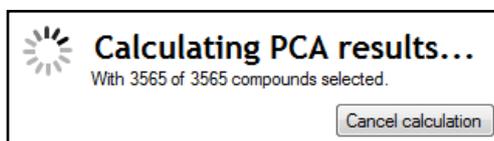


Now move forward to Progenesis Stats by clicking **Section Complete**

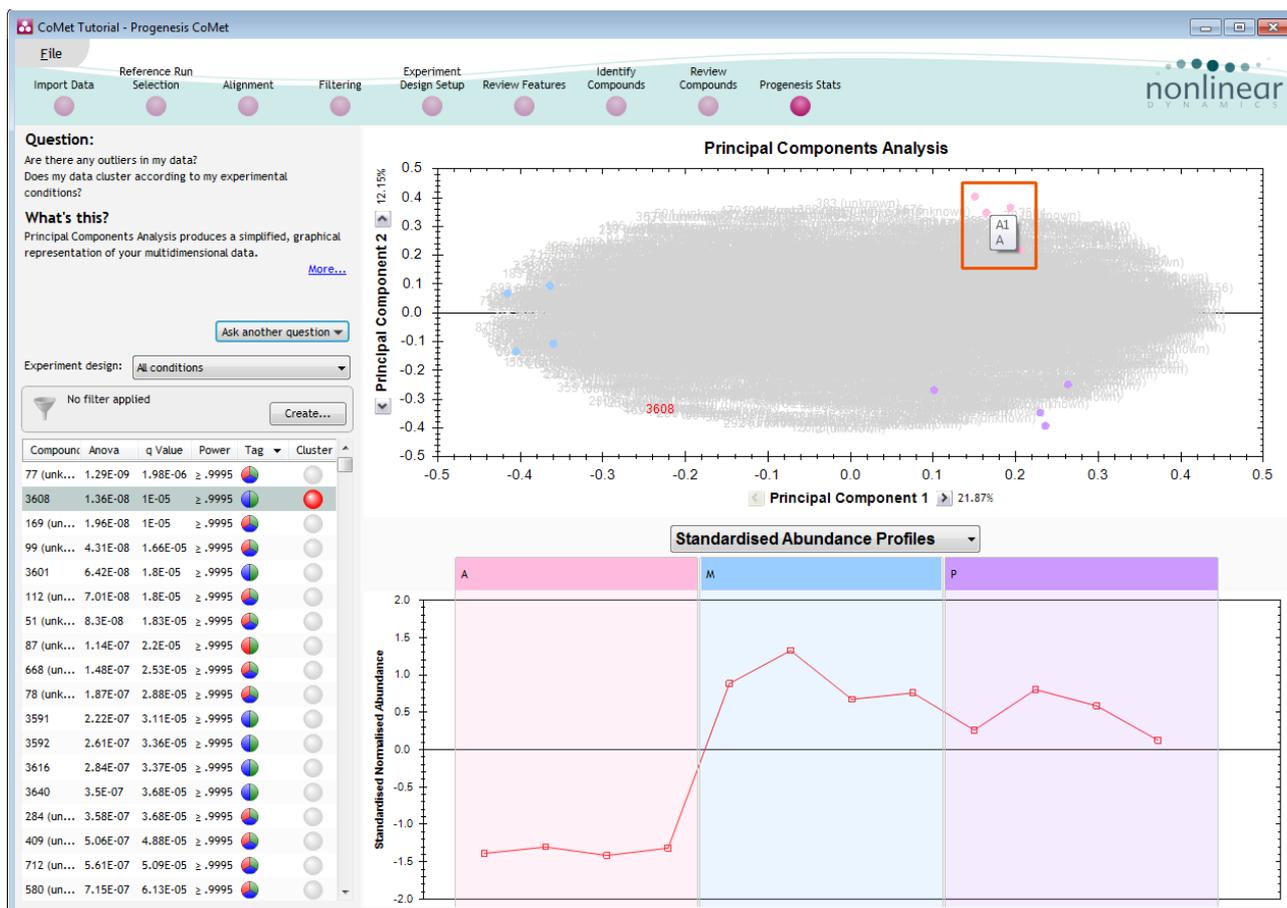
## Stage 10: Multivariate Statistics

The user guide now describes the functionality of the Multivariate Statistics. This section is only available if Progenesis Stats is licensed.

Progenesis Stats opens calculating the Principal Components Analysis (PCA) for the active 'tag' filter if one exists.



As an example we will start by examining the PCA for all of the 3565 compounds.



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

**Note:** the 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)

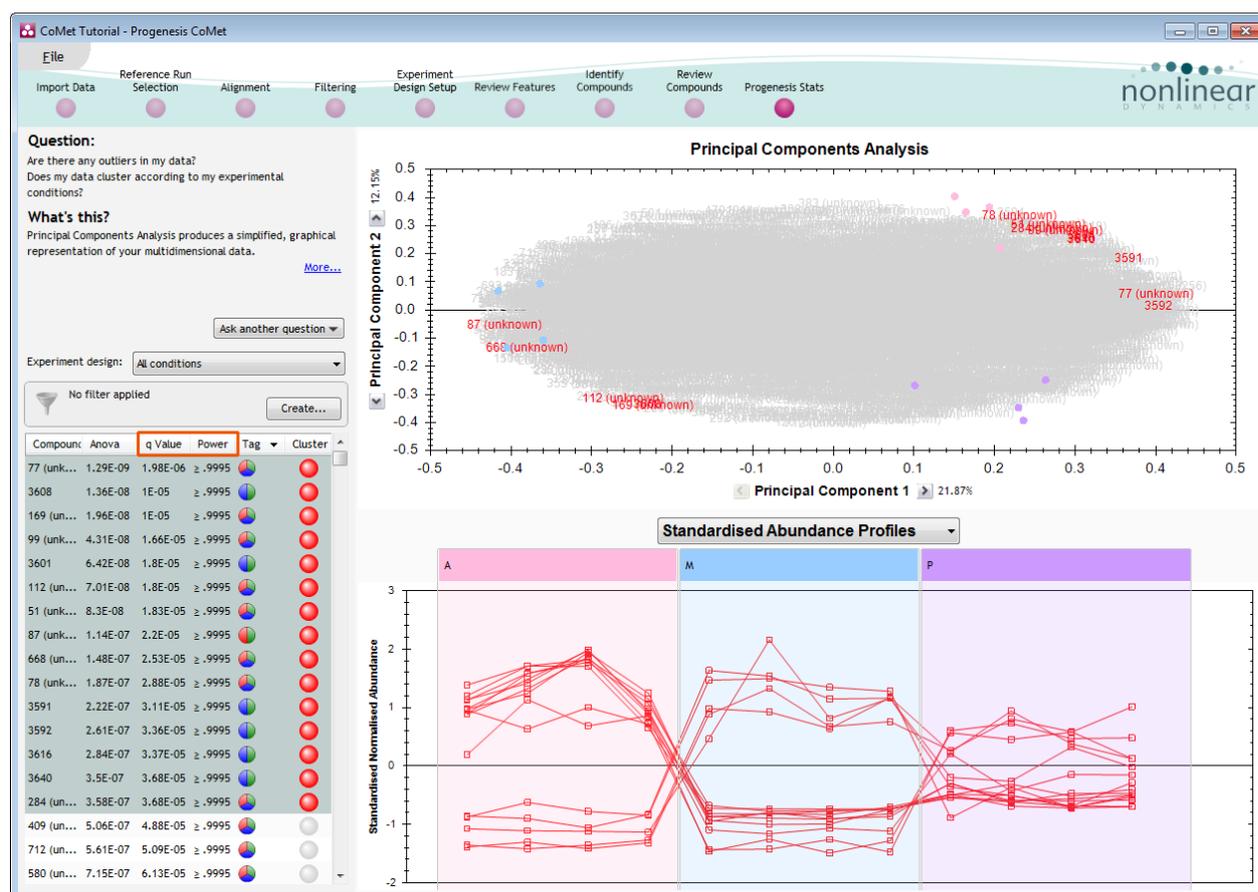
In **Progenesis Stats** the first statistically based question asked of the data takes the form of a Quality Control assessment:

*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 compounds in the table will highlight the compounds on the 'Biplot' and their abundance 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 compound, using the abundance variance, sample size and difference between the means.

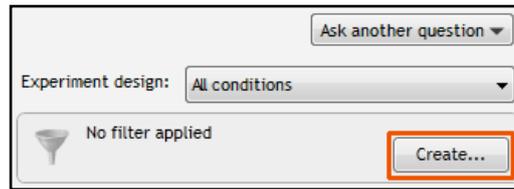
Also, for a given power of 80% we can determine how many samples are required to ensure we find a difference if it actually exists.

**Note:** Power analysis is discussed in Appendix 7 (page 58)

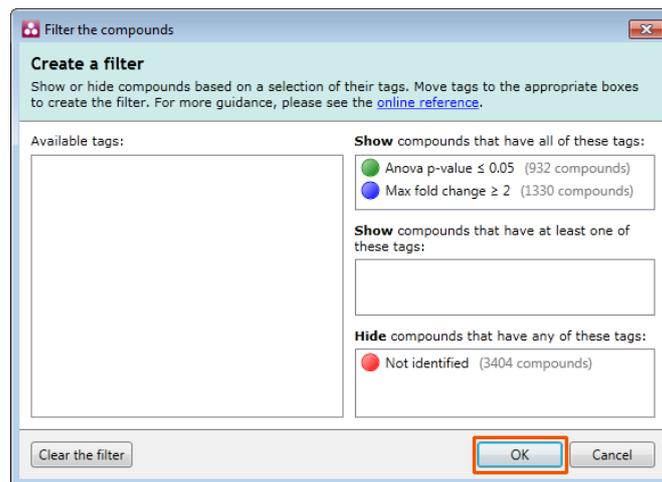
## Correlation Analysis

Use the tags created in Review Compounds to filter the compounds displayed in the table. We are going to explore the Correlation Analysis for all the Compounds (with possible identifications) that display a significant 2 fold or greater difference in abundance.

To filter the data click **Create**

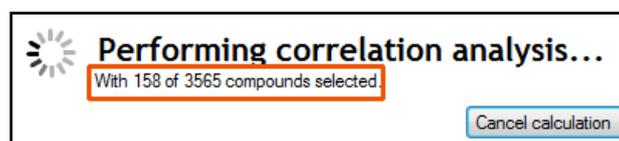
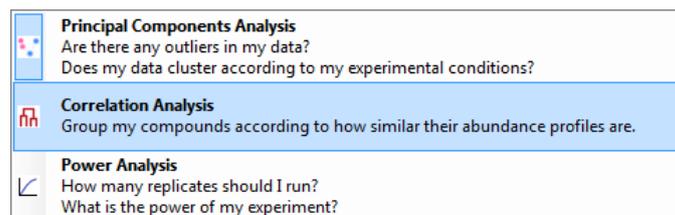


The set up the filter as shown below



On pressing OK the PCA will recalculate.

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 abundance profiles'

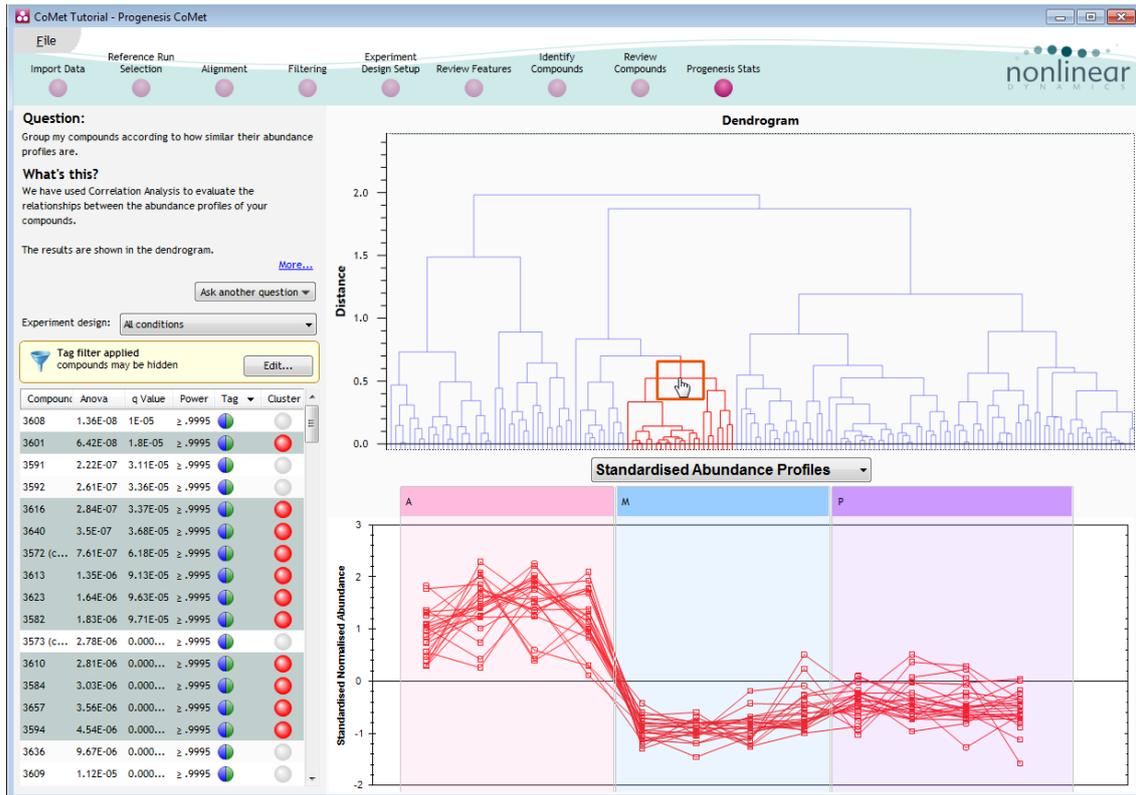
This time the statistically based question(s) being asked is:

*'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'*

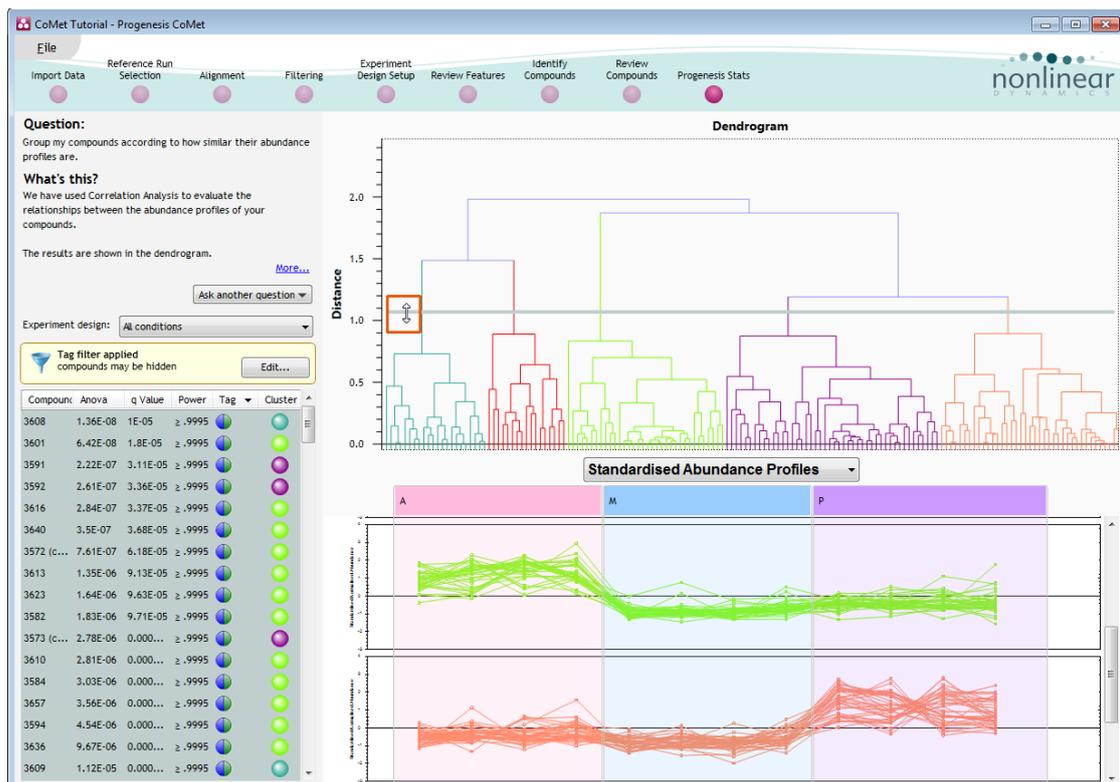
The answer is displayed graphically in the form of an interactive dendrogram where the vertical distance, between each branch can be taken as indicative of how similar the abundance profiles of each cluster of compounds are to each other.



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

Clicking on a branch on the Dendrogram selects the compounds on the table. You can then tag this group, of 'potentially' related compounds, by right clicking and creating a 'New tag' for them. Then use the tag to focus on them at the **Review Compounds** stage.

Finally use the slider (located at the top of the dendrogram) to highlight multiple branches by dragging it down.



Each highlighted branch will have an Abundance Profile coloured the same.

## **Congratulations!**

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

Hopefully our example has shown you how this unique technology can deliver significant benefits with

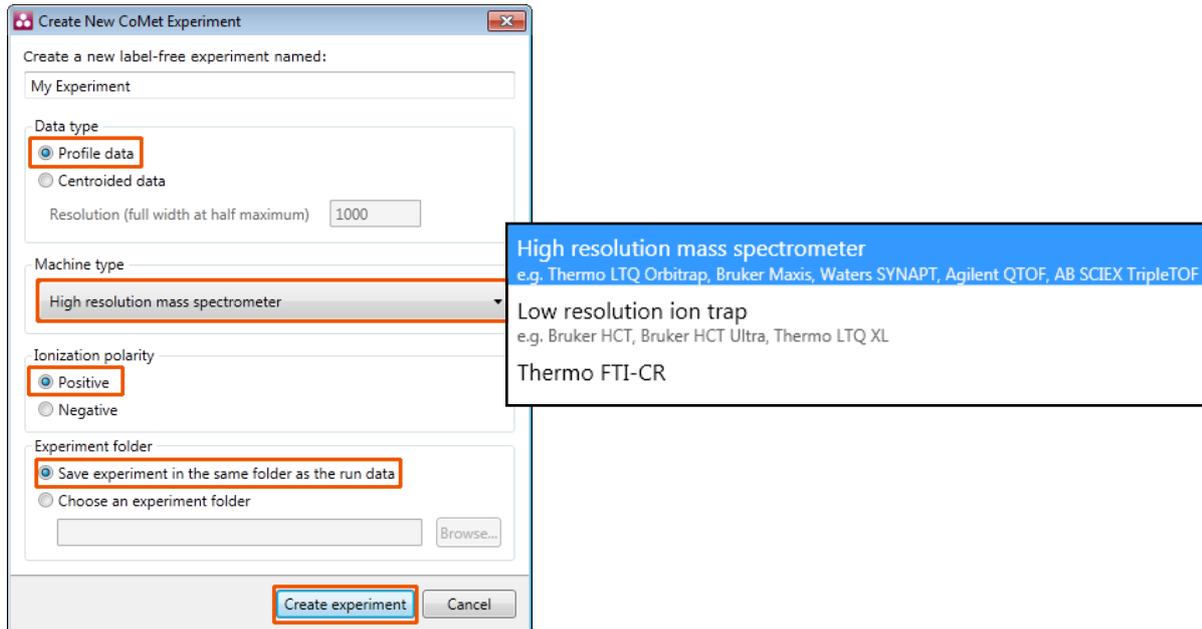
- Speed
- Objectivity
- Statistical Power

If you would like to see the benefits of running Progenesis CoMet using your own runs and explore the Progenesis CoMet workflow please go to Appendix 3: Licensing Runs (page 48).

## 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' Waters, Buker and Agilent) or, for other Vendors, convert them to mzXML 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'.

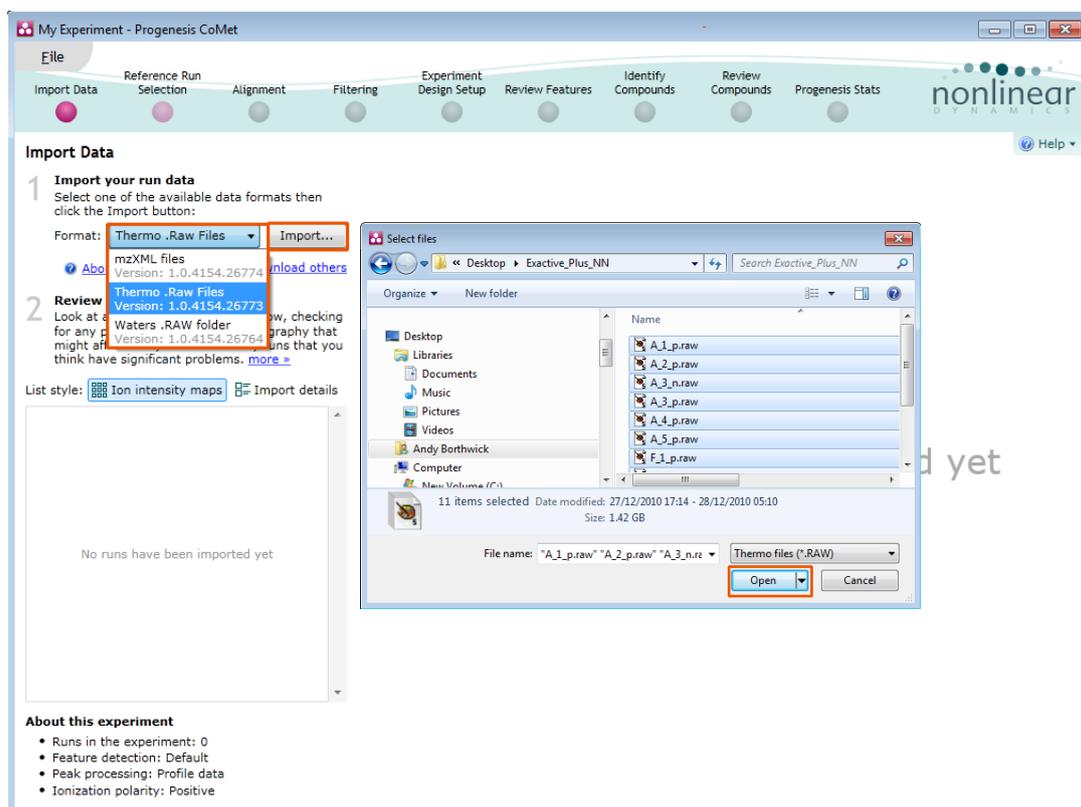


**Note:** if you have converted or captured the data as centroided then select Centroided data and enter the Resolution for the MS machine used.

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

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

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

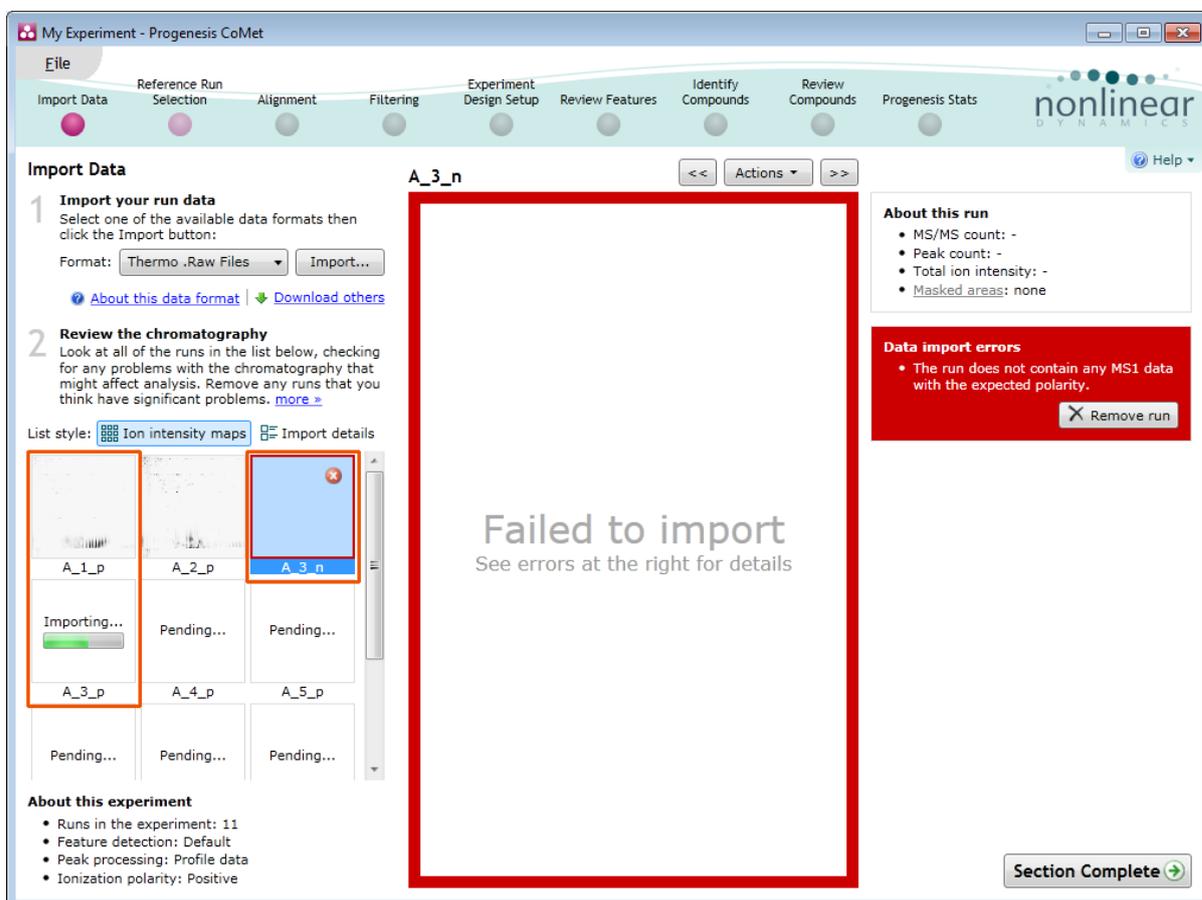


Locate and select all the Data files, in this example (A1 to F5).

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 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 failed to be imported and a possible reason for the failure.



**Note:** as each data file is loaded the progress is reported in the **Import Data** list with a progress bar for the file currently being imported and a 2D representation of the successfully imported files.

The dialog to the right of the main panel provides details of the current file and if it has failed import it can be removed by clicking **Remove run** in the Data import errors panel. In this example it had the incorrect polarity.

On the bottom left of the Import data list are the experiment details.

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

## Appendix 2: Stage 1 Data QC review and addition of exclusion areas

During the process of Data QC you may identify areas of the raw data for a particular run that appear 'noisy' yet still have identifiable 'isotopic patterns'.

For example if the run is part of a 'replicate set' of runs it is possible to exclude such areas on the noisy run by applying a mask to the area. By doing so this area is excluded during the initial part of the detection process in order that it does not 'interfere' with the detection of the features in the replicate group.

To do this select Exclude areas from selected run on the bottom left of the screen.

Drag out an area over the noisy part of the run to create the mask.

**Note:** if you now zoom into the masked area using the **Zoom** tool you will see the isotopic features in the noise.

**Import Data**

Thermo .RAW files Add Files...

[About this data format...](#)

Include?

---

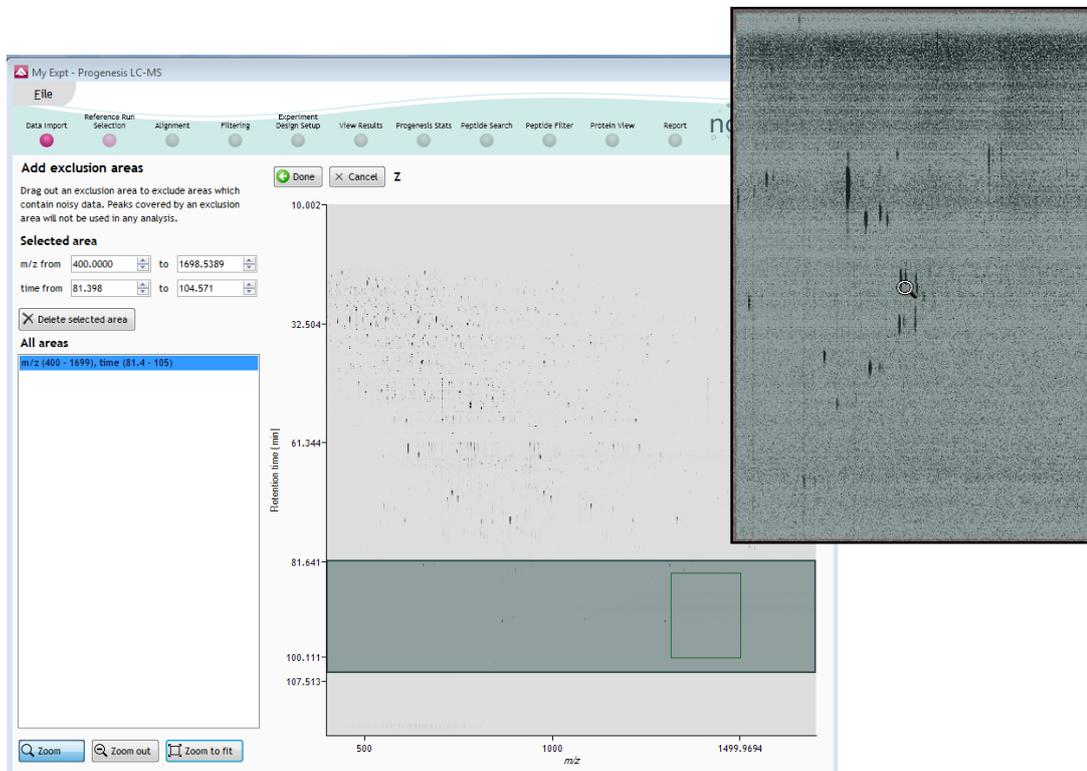
**No problems found**

|   |                                     |                          |
|---|-------------------------------------|--------------------------|
| X | <input checked="" type="checkbox"/> | <input type="checkbox"/> |
| Y | <input checked="" type="checkbox"/> | <input type="checkbox"/> |
| Z | <input checked="" type="checkbox"/> | <input type="checkbox"/> |

Include run in analysis

Don't include run in analysis

Exclude areas from selected run



**Note:** if the level of noise is high and affecting many of your runs a preferred approach would be to re-optimize the chromatography to improve the levels of noise in your data

## Appendix 3: Licensing runs (Stage 3)

When setting up a **New experiment** if you are evaluating Progenesis CoMet 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 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 **Nonlinear Sales Person** or **purchase a licence code directly from Nonlinear**.

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**.

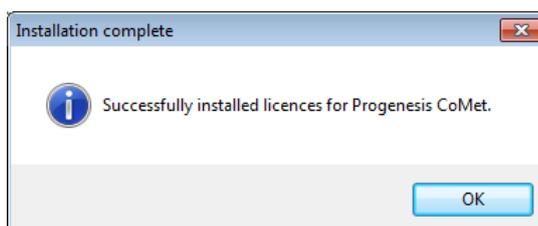
**Note:** you will need an internet connection to use this method.

The screenshot shows the 'License Runs' tab in the Progenesis CoMet software. The main content area contains a table with the following data:

| Run name   | License state | License this run                    |
|--|---------------|-------------------------------------|
| C:\Users\andy.borthwick\Desktop\Tutorial Lic\A1.C1k-P.mznlid | Unlicensed    | <input checked="" type="checkbox"/> |
| C:\Users\andy.borthwick\Desktop\Tutorial Lic\A2.C1k-P.mznlid | Unlicensed    | <input checked="" type="checkbox"/> |
| C:\Users\andy.borthwick\Desktop\Tutorial Lic\A3.C1k-P.mznlid | Unlicensed    | <input checked="" type="checkbox"/> |
| C:\Users\andy.borthwick\Desktop\Tutorial Lic\A4.C1k-P.mznlid | Unlicensed    | <input checked="" type="checkbox"/> |
| C:\Users\andy.borthwick\Desktop\Tutorial Lic\M1.C1k-P.mznlid | Unlicensed    | <input checked="" type="checkbox"/> |
| C:\Users\andy.borthwick\Desktop\Tutorial Lic\M2.C1k-P.mznlid | Unlicensed    | <input checked="" type="checkbox"/> |
| C:\Users\andy.borthwick\Desktop\Tutorial Lic\M3.C1k-P.mznlid | Unlicensed    | <input checked="" type="checkbox"/> |
| C:\Users\andy.borthwick\Desktop\Tutorial Lic\M4.C1k-P.mznlid | Unlicensed    | <input checked="" type="checkbox"/> |
| C:\Users\andy.borthwick\Desktop\Tutorial Lic\V1.C1k-P.mznlid | Unlicensed    | <input checked="" type="checkbox"/> |
| C:\Users\andy.borthwick\Desktop\Tutorial Lic\V2.C1k-P.mznlid | Unlicensed    | <input checked="" type="checkbox"/> |
| C:\Users\andy.borthwick\Desktop\Tutorial Lic\V3.C1k-P.mznlid | Unlicensed    | <input checked="" type="checkbox"/> |
| C:\Users\andy.borthwick\Desktop\Tutorial Lic\V4.C1k-P.mznlid | Unlicensed    | <input checked="" type="checkbox"/> |

Below the table, there is a text input field for 'Run licence code' with the placeholder 'xxx-xxxx-xxxx-xxx' and a 'Use Licence Code' button. A 'Section Complete' button is located at the bottom right of the main content area.

A message confirming successful installation of your licences will appear.



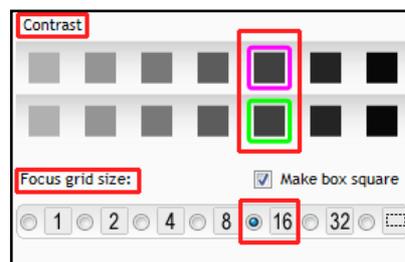
Click **OK**, the view will update and Alignment, the next stage in the workflow, will open with the licensed files.

## Appendix 4: Manual assistance of Alignment

### Approach to alignment

To place manual alignment vectors on a 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 5 - 10 **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.



Click on an area (see below) in the **Ion Intensity Map** window to refocus all the windows. Adjust Contrast as required

**Alignment**

Alignment is used to compensate for retention time drift between your runs.

Use the Automatic Alignment button to begin then review the alignment on each of your runs before moving on.

| Run | Include?                            | Vectors |
|-----|-------------------------------------|---------|
| A1  | <input checked="" type="checkbox"/> | Ref     |
| A2  | <input checked="" type="checkbox"/> | 0       |
| A3  | <input checked="" type="checkbox"/> | 0       |
| B1  | <input checked="" type="checkbox"/> | 0       |
| B2  | <input checked="" type="checkbox"/> | 0       |
| B3  | <input checked="" type="checkbox"/> | 0       |
| C1  | <input checked="" type="checkbox"/> | 0       |
| C2  | <input checked="" type="checkbox"/> | 0       |
| C3  | <input checked="" type="checkbox"/> | 0       |

Contrast

Focus grid size:  Make box square

1 2 4 8 16 32

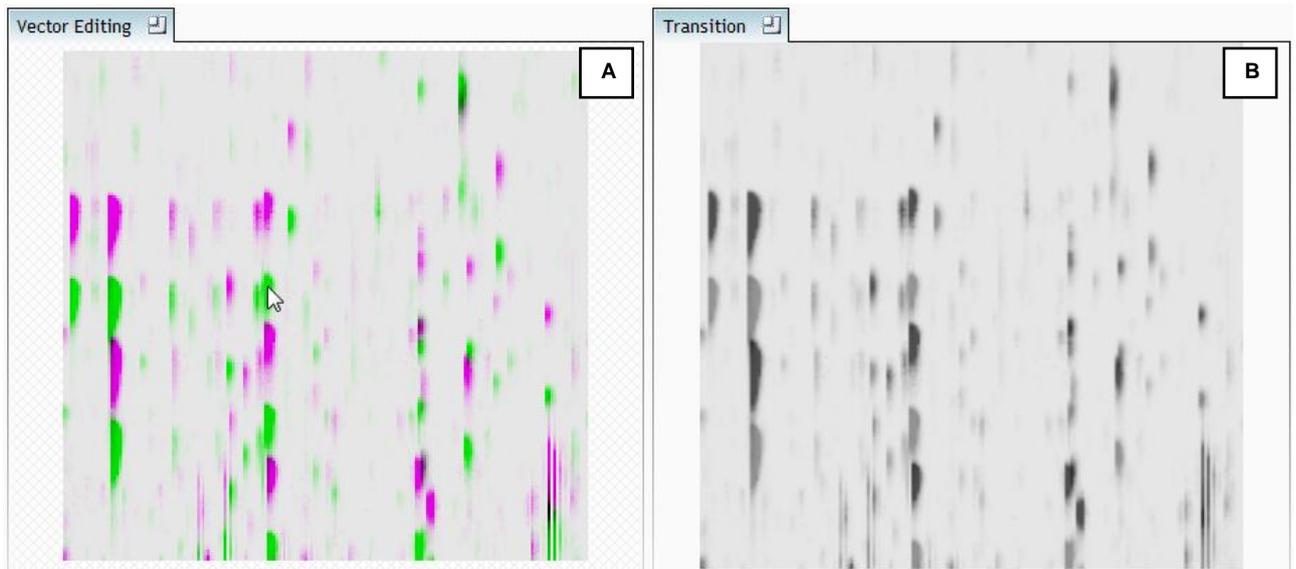
Alignment target Run being aligned

Section Complete

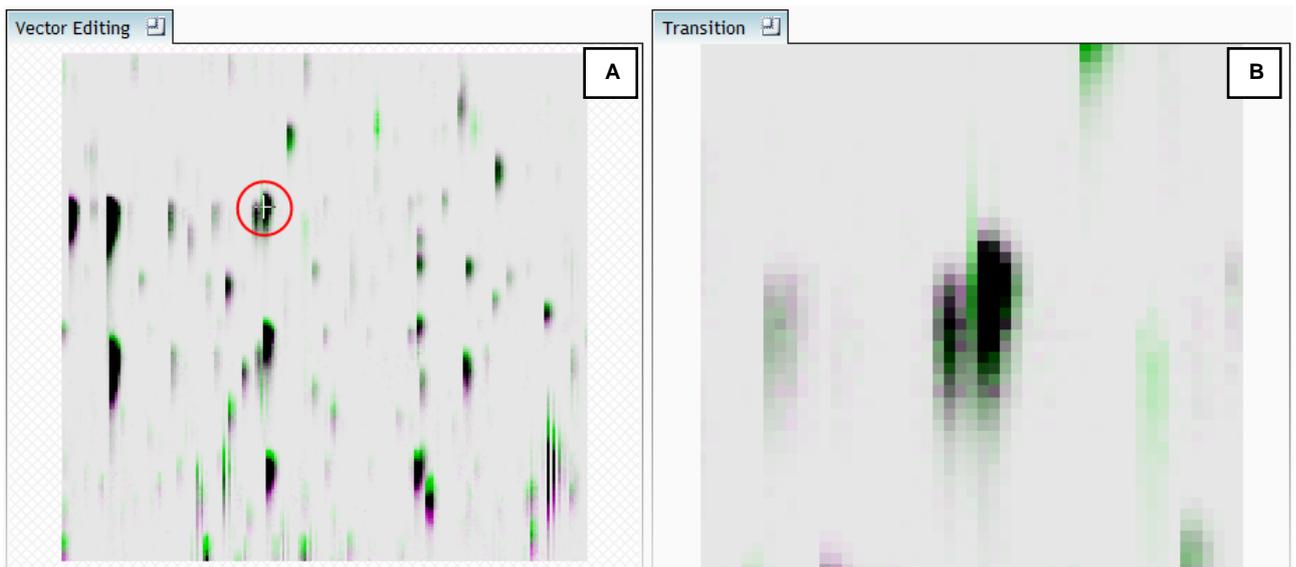
**Note:** the features moving back and forwards between the 2 runs in the **Transition** window indicating the misalignment of the two runs

**Note:** The **Total Ion Chromatograms** window 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).

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

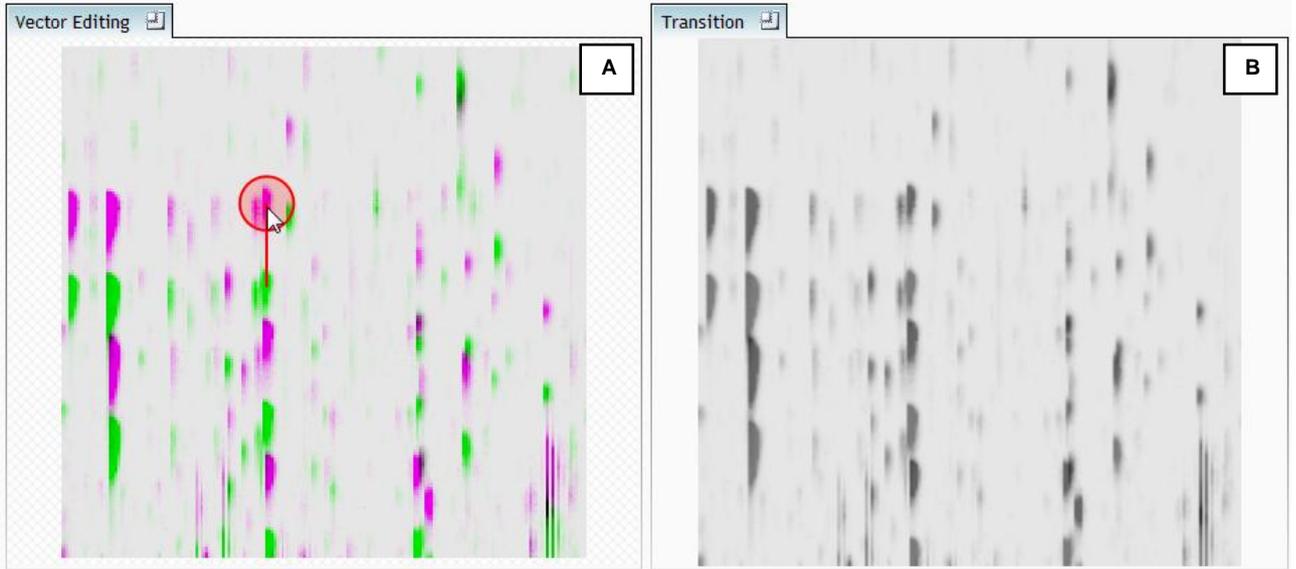


- As you are holding down the left mouse button drag the green feature over the corresponding magenta feature of the reference run. The red circle will appear as shown below indicating that a positional lock has been found for the overlapping features.



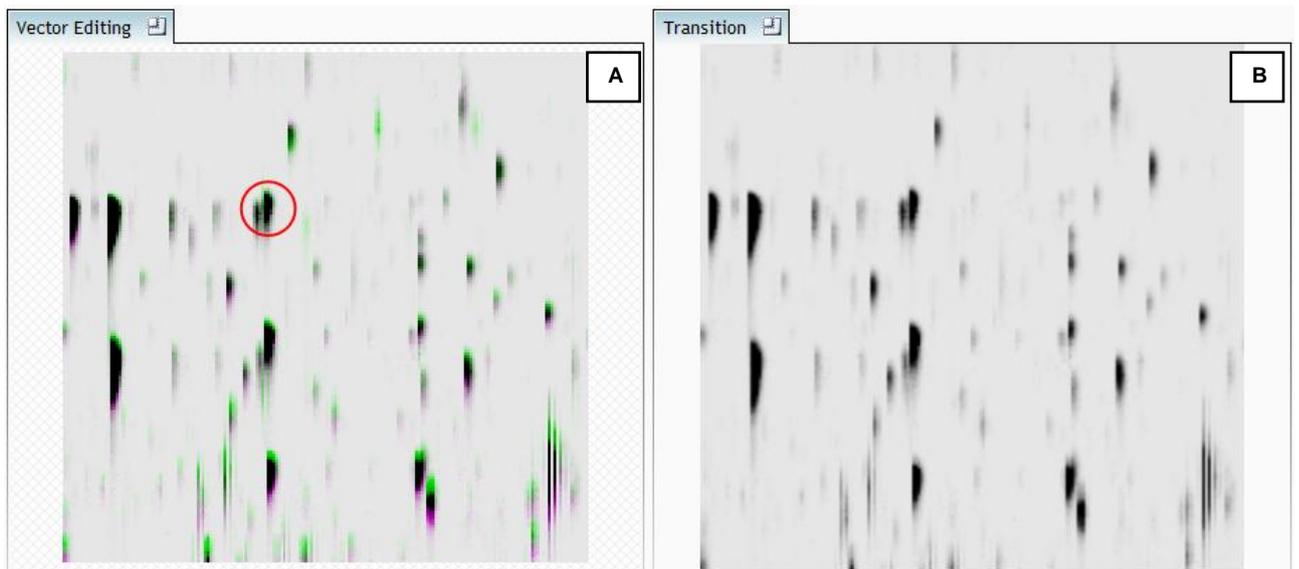
**Note:** as you hold down the mouse button, window B zooms in to help with the alignment.

- 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.

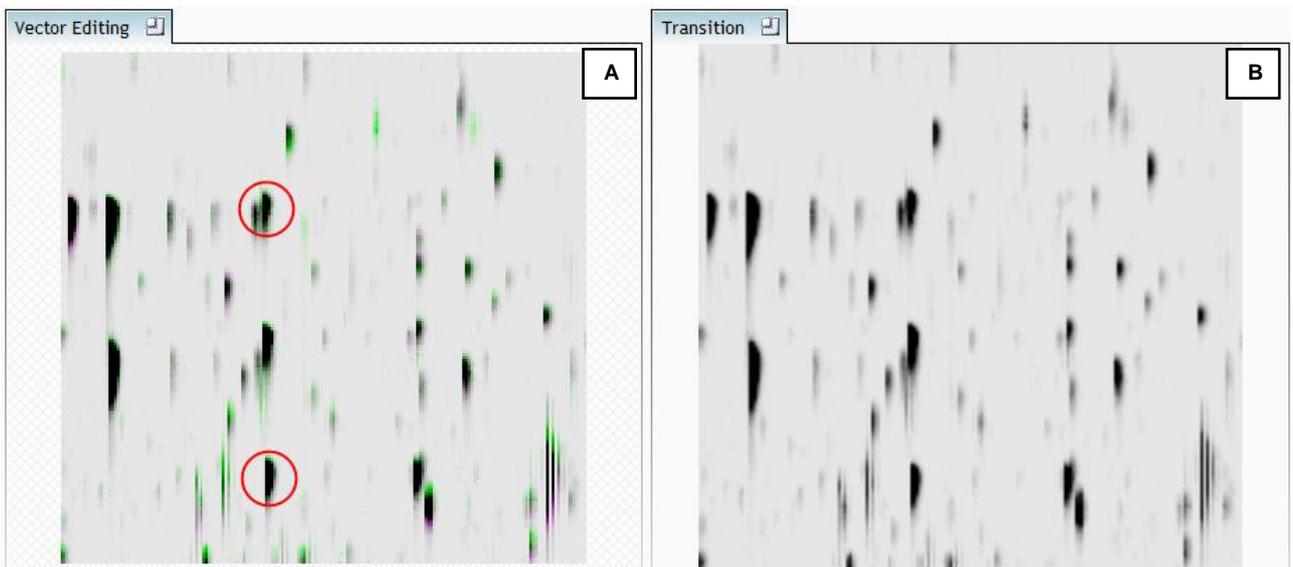


**Note:** an incorrectly placed vector is removed by right clicking on it in the **Vector Editing** window

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



8. Adding 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.



9. Repeat this process moving the focus from top to bottom on the **Ion Intensity Map** view

**Alignment**

Alignment is used to compensate for retention time drift between your runs.

Use the Automatic Alignment button to begin then review the alignment on each of your runs before moving on.

| Run | Include?                            | Vectors |
|-----|-------------------------------------|---------|
| A1  | <input checked="" type="checkbox"/> | Ref     |
| A2  | <input checked="" type="checkbox"/> | 5       |
| A3  | <input checked="" type="checkbox"/> | 0       |
| B1  | <input checked="" type="checkbox"/> | 0       |
| B2  | <input checked="" type="checkbox"/> | 0       |
| B3  | <input checked="" type="checkbox"/> | 0       |
| C1  | <input checked="" type="checkbox"/> | 0       |
| C2  | <input checked="" type="checkbox"/> | 0       |
| C3  | <input checked="" type="checkbox"/> | 0       |

Contrast

Focus grid size:  Make box square

1 2 4 8 16 32

Alignment target Run being aligned

Section Complete

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

## 10. Repeat this process for all the runs to be aligned.

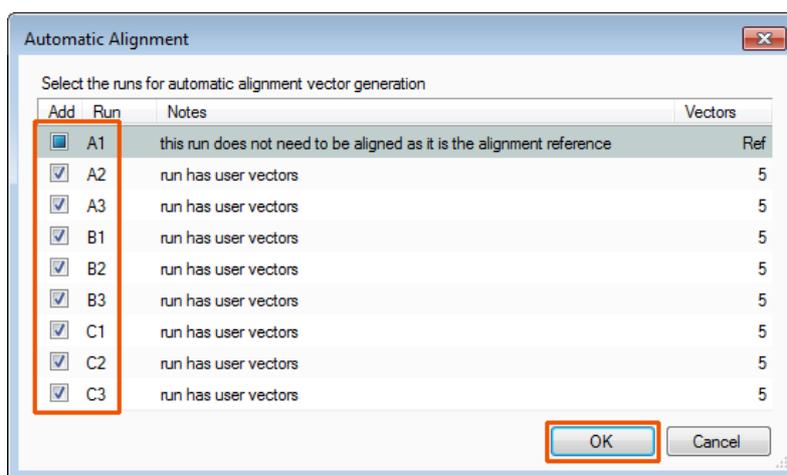
The number of manual vectors that you add at this stage is dependent on the misalignment between the current run and the Reference run.

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

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



11. Then select Automatic Alignment to bring up the Automatic Alignment dialog and click **OK**. The automatic alignment process will begin, using the manual vectors you have added to aid in automatic vector placement.



**Note:** the tick boxes next to the **Run** name control whether vectors will be generated for each run.

## Appendix 5: 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 Treatment:Day1 and then at 3 times following treatment: Day2, Day3 and Day4 .

My Compound Timecourse - Progenesis CoMet

File | Import Data | Reference Run Selection | Alignment | Filtering | Experiment Design Setup | Review Features | Identify Compounds | Review Compounds | Progenesis Stats

nonlinear DYNAMICS

New | Help

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.

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

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    |

Create a new experiment design

Name: 3 day treatment of Patients

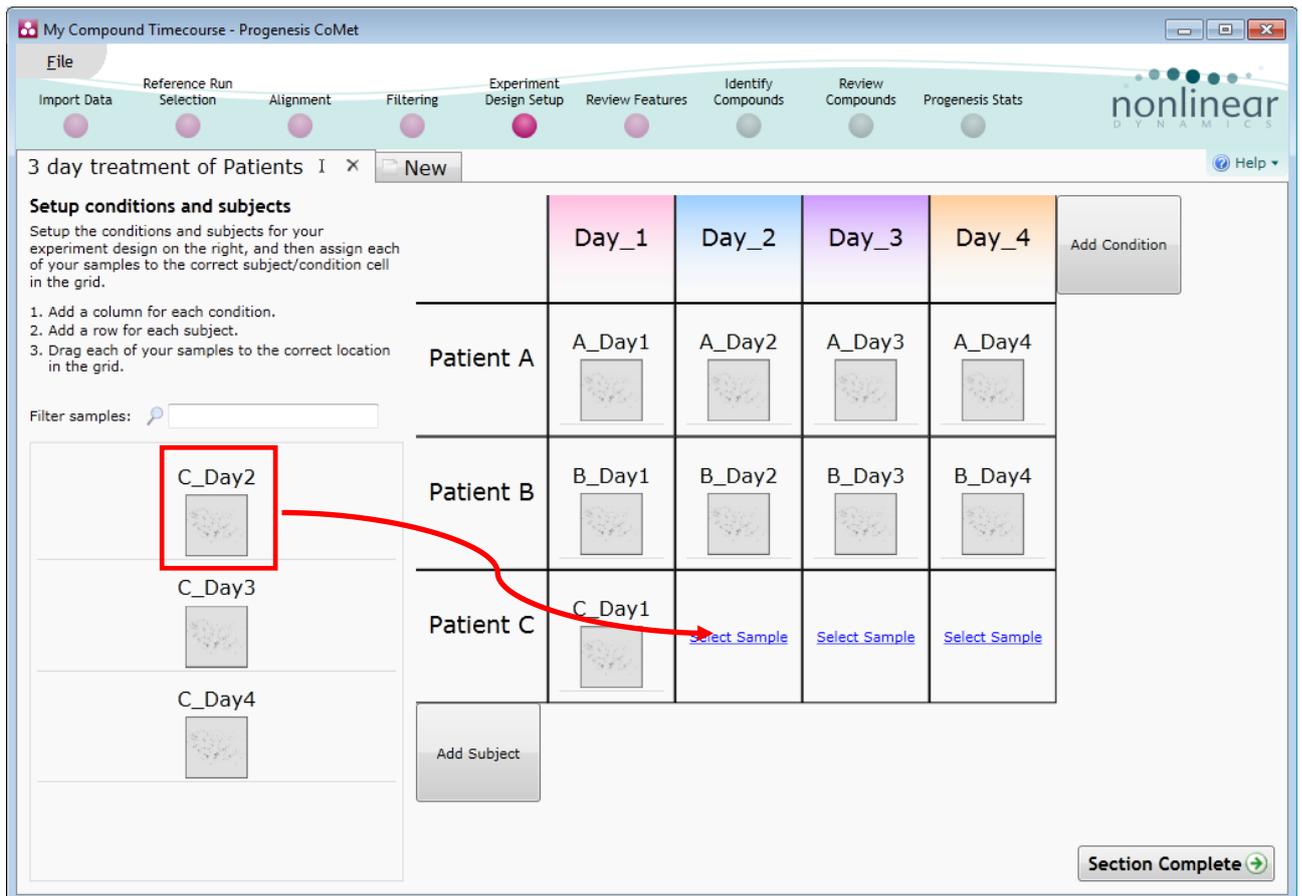
Start with an empty layout

Copy layout from:

Create design | Cancel

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.

Then Drag and drop the Samples on to the correct 'cell' of the matrix.



You can create additional Experimental Designs using the New tab

All of these Experimental Designs are available at all the following stages in the CoMet workflow

## Appendix 6: Deciding which identifications are correct

After importing identifications for your features, you may find that some compounds have more than one possible identification. In the **Review Compounds** screen, the possible identifications for the selected compound are listed at the bottom-right. To confirm the compound's 'true' identification, you need only click the star alongside the Compound ID, marking it as the *accepted identification*.

However, how do you decide which identification is the correct one? Here are a few considerations that can influence your decision:

### Highest identification score

If you're identifying compounds using a metabolite database that assigns each identification a score, with the highest value for a given compound representing the most likely identification, this appears in the Score column of the Possible identifications list.

You *could* simply accept the identification with the highest score, but even if you have score values, it's still recommended that you validate the identifications with some of the other considerations listed here.

### The presence of peptides

If you can safely ignore amino acids and peptides for your compound identifications, this may provide a way to reduce the number of identifications under consideration for a given compound. This is especially true with peptide chains, as the same set of amino acids could be present in different orders to produce the same compound mass, as seen in this example:

**Possible identifications: 10**

| ☆ | Compound ID | Description | Formula  | Score | Theoretical Mass | Link |
|---|-------------|-------------|--|-------|------------------|------|
| ☆ | cmid:22223  | Asp Glu Asp | C <sub>13</sub> H <sub>19</sub> N <sub>3</sub> O <sub>10</sub>               |       | 377.1070         |      |
| ☆ | cmid:17334  | Asp Asp Glu | C <sub>13</sub> H <sub>19</sub> N <sub>3</sub> O <sub>10</sub>               |       | 377.1070         |      |
| ☆ | cmid:17126  | Glu Asp Asp | C <sub>13</sub> H <sub>19</sub> N <sub>3</sub> O <sub>10</sub>               |       | 377.1070         |      |
| ☆ | cmid:22345  | Met Asp Pro | C <sub>14</sub> H <sub>23</sub> N <sub>3</sub> O <sub>6</sub> S <sub>1</sub> |       | 361.1308         |      |
| ☆ | cmid:21062  | Asp Met Pro | C <sub>14</sub> H <sub>23</sub> N <sub>3</sub> O <sub>6</sub> S <sub>1</sub> |       | 361.1308         |      |
| ☆ | cmid:19874  | Pro Asp Met | C <sub>14</sub> H <sub>23</sub> N <sub>3</sub> O <sub>6</sub> S <sub>1</sub> |       | 361.1308         |      |
| ☆ | cmid:19649  | Asp Pro Met | C <sub>14</sub> H <sub>23</sub> N <sub>3</sub> O <sub>6</sub> S <sub>1</sub> |       | 361.1308         |      |
| ☆ | cmid:19013  | Pro Met Asp | C <sub>14</sub> H <sub>23</sub> N <sub>3</sub> O <sub>6</sub> S <sub>1</sub> |       | 361.1308         |      |
| ☆ | cmid:17670  | Met Pro Asp | C <sub>14</sub> H <sub>23</sub> N <sub>3</sub> O <sub>6</sub> S <sub>1</sub> |       | 361.1308         |      |
| ☆ | cmid:1347   | Bisacodyl   | C <sub>22</sub> H <sub>19</sub> NO <sub>4</sub>                              |       | 361.1314         |      |

This is the accepted identification of the compound

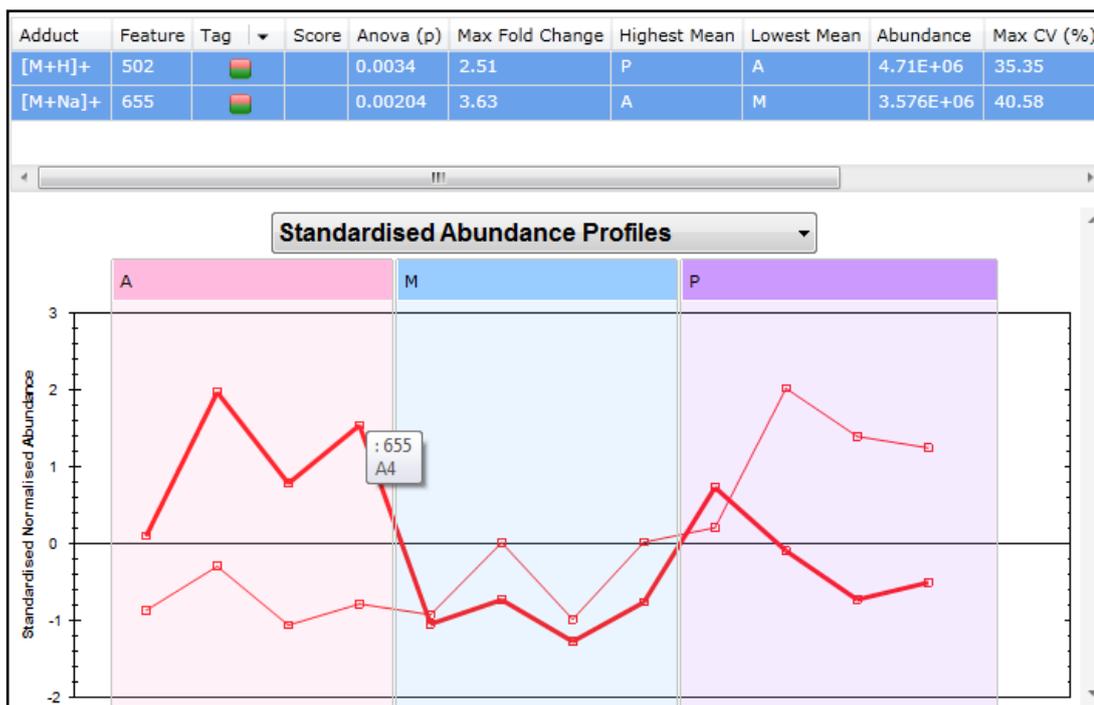
### Mismatch in the features' abundance profiles

Typically, you would expect each adduct or charge state of a given compound to have similar abundance levels in the various runs of each experimental condition. These abundance levels can be viewed easily by

For example, in an experiment with multiple conditions A, M and P conditions, all adducts of a given compound might be more abundant in the A samples than in the M and P samples.

However, if the compound comprised ions whose abundance profiles were different, that could indicate that one or more of the identifications is unreliable. The screen below, displayed by double-clicking a compound in the Review Compounds screen, shows one such situation:

Here, the compound under review is composed of two compound ions, the two adducts being a hydrogen ion and a potassium ion. By selecting both adducts' rows in the table, displays their abundance profiles overlaid on the same graph. We can clearly see that they are not consistent. One adduct has a relatively high average abundance in A whereas the other has a low abundance; and the same is true for the P condition.



**Note:** the profiles will be exactly the same, no matter which of the compound's identifications is selected in this screen. That's because the set of features listed isn't dependent on the identification; it's dependent on the compound itself. Consequently, all possible identifications for this compound — in other words, the compound as a whole — may have to be rejected.

**Adducts that don't reflect the separation technique**

If you're using GC/MS, it's unlikely that you'll see compound ions with acetonitrile (ACN) adduct, a common solvent used in LC/MS. If one of the possible identifications for a given compound has been assigned based on the ion being the original compound with an ACN adduct, that identification could then be rejected.

**Adduct retention times that don't make sense**

If one of your possible identifications is based on a pair of adducts (i.e. the compound is composed of two features) that you know should produce the same retention times (e.g. Na<sup>+</sup> and K<sup>+</sup>), but you can see that the retention times for the features are significantly different, that can be a reason to reject the identification.

An example scenario can be seen here:

| Adduct              | Feature | Tag | Score | Anova (p) | Max Fo | Highest | Lowest | Abundance | Max CV (%) | Retention Time |
|---------------------|---------|-----|-------|-----------|--------|---------|--------|-----------|------------|----------------|
| [M+Na] <sup>+</sup> | 144     |     |       | 0.00494   | 4.89   | M       | A      | 9.847E+06 | 66.23      | 6.683          |
| [M+K] <sup>+</sup>  | 200     |     |       | 0.0105    | 3.03   | A       | M      | 6.644E+06 | 43.54      | 9.411          |

**Knowledge about the samples**

Finally, if your knowledge of how the samples were obtained leads you to believe that certain compounds are simply unlikely to be present, you can use that to inform your decision of which is the correct identification.

## Appendix 7: Power Analysis

Power analysis is a statistical technique, which is used to gauge how many replicates are need to reliable observe the abundance differences in your data. It is available through the Progenesis 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 Progenesis 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 compounds according to how similar their abundance profiles are.

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

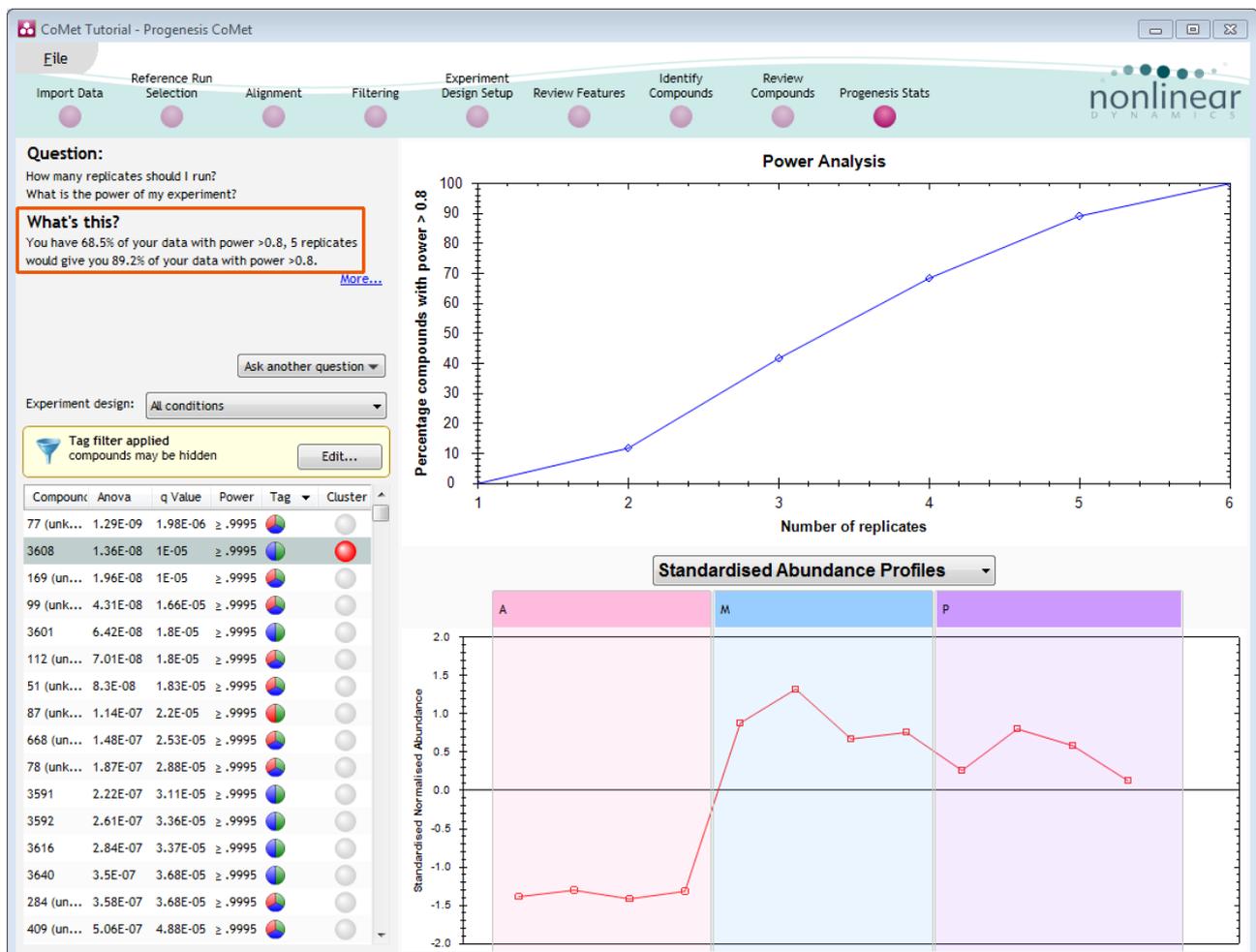
Select the option

*'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 compounds with a power >0.8'*

Using the **Significant p<0.05 features (932)**, as an example, view the power analysis.



This is displayed graphically showing that 68.5% of the 932 features have a power of 80% or that 5 replicates would give you 89.2% of your data with power > 0.8.