Last updated: 16/04/2021

GCMS Data Analysis

Introduction

Gas chromatography-mass spectrometry (GCMS) is the premier analytical technique for identification and quantitation of volatile and semi-volatile compounds. Low cost of the instruments, high reproducibility, superior chromatographic resolution, and extensive spectral libraries make GCMS essential for analysis of compounds and contaminants in foods, fuels, and environmental samples. With the use of derivatisation agents, many non-volatile compounds become amenable to GCMS making this technique compatible with metabolic phenotyping studies in biological samples, complementary to LCMS. Because GCMS-based quantitation is strictly chromatography based, GCMS data are ideal for analysis using Skyline.

In this tutorial, you will learn how to analyse untargeted metabolomics data corresponding to central carbon metabolism of fruit fly fed or unfed with zingerone. Mostly polar metabolites were enriched in the samples using a standard methanol:water:chloroform extraction, dried, and converted to their trimethylsilyl (TMS) derivatives using BSTFA reagent. The data was then acquired on Shimadzu GCMS QQQ instrument in full MS scan mode (untargeted approach). To bypass the challenge associated with correct identification of hundreds of metabolites, each metabolite will be considered an unknown chromatographic feature and processed in Skyline without knowing its identity. A simple statistical analysis, also performed in Skyline, will determine which chromatographic features consist of ions corresponding to differentially regulated metabolites and this will guide metabolite identification conducted outside Skyline.

Getting Started

To start this tutorial, download the *GCMSmetabolomics.zip* file (~1GB):

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

C:\Users\sadowskp\Downloads

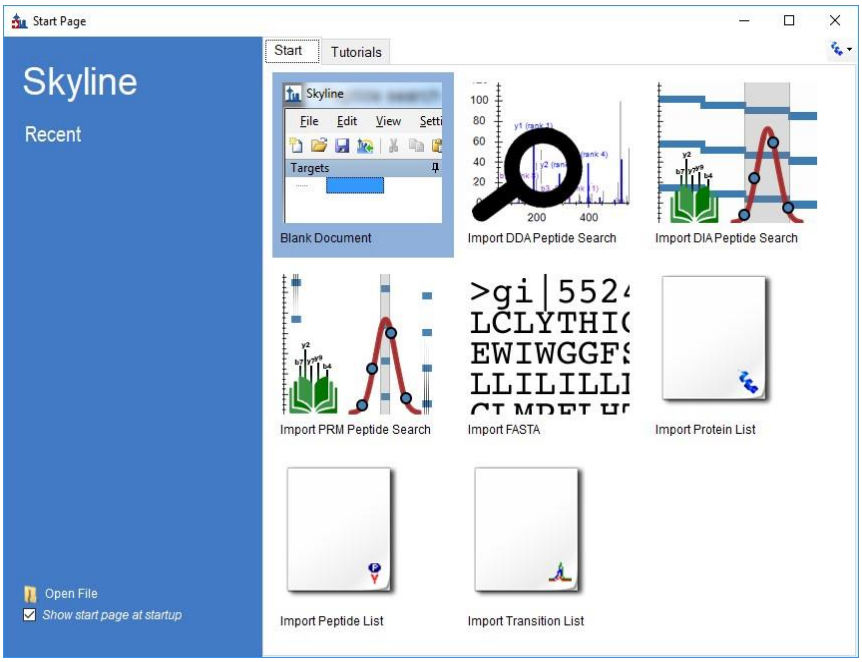
This will create a new folder:

C:\Users\sadowskp\Downloads\GCMSMetabolomics

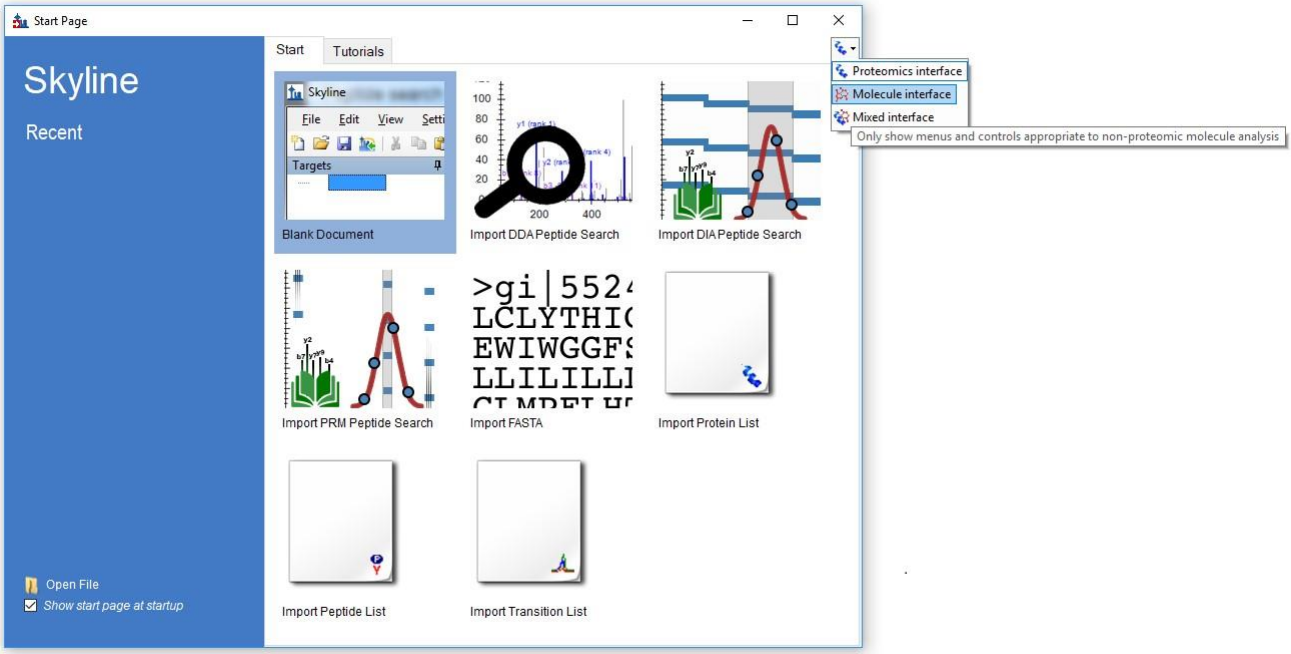
The folder will contain all files necessary for this tutorial including raw data in .mzXML format and transition list as .csv file (instructions how to convert raw data into .mzXML format can be found In **Appendix**).

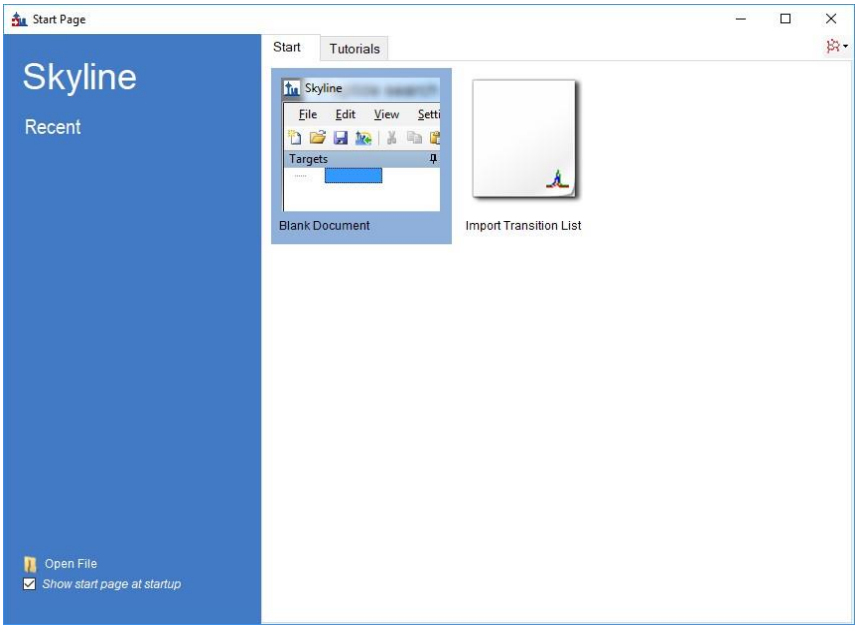
Now:

- Open Skyline. You should see the **Start page** like below:

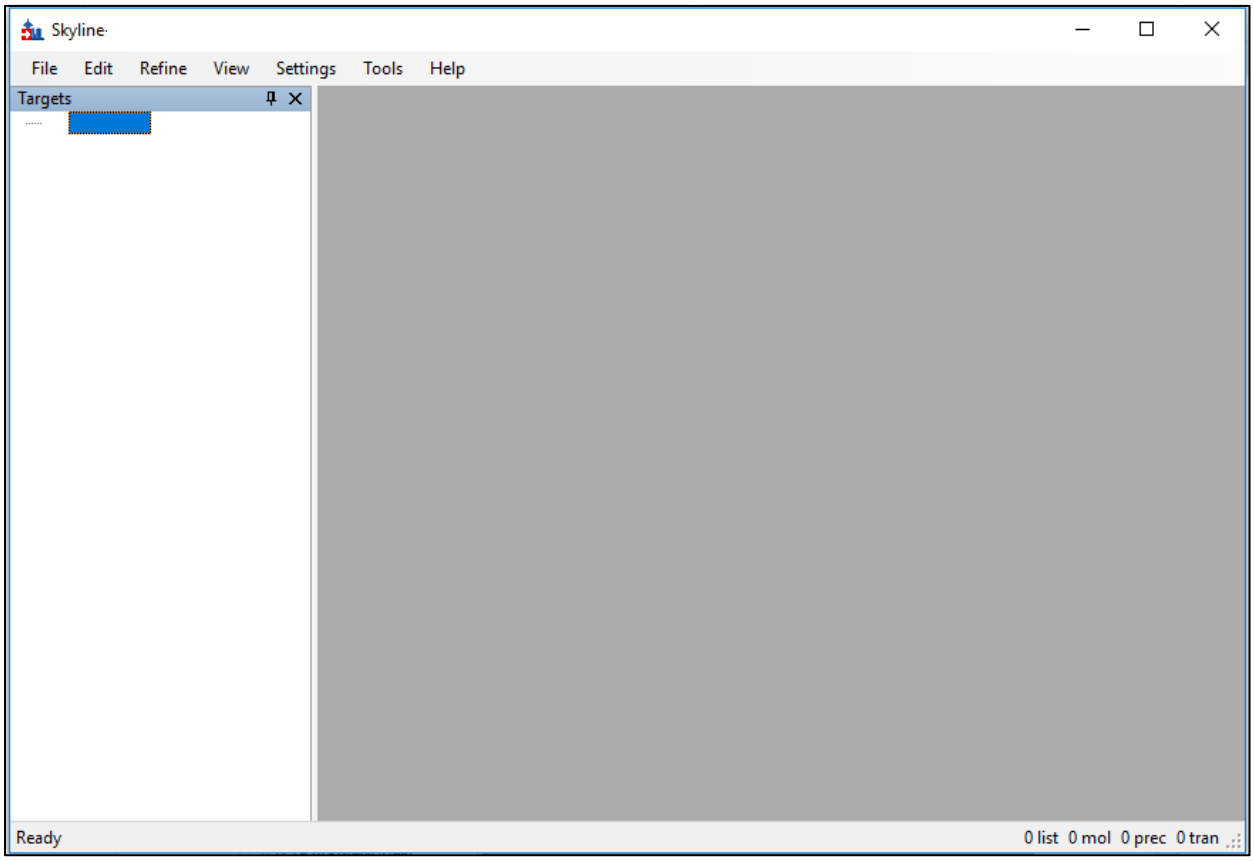


- Switch user interface to **Molecule interface** by selecting this option in the drop-down menu accessible in the top right corner of the software window. This will turn off any protein-related options and descriptions in Skyline.





- Click on **Blank Document**. You should see an empty document with 0 lists, 0 molecules, 0 precursors and 0 transitions like below:

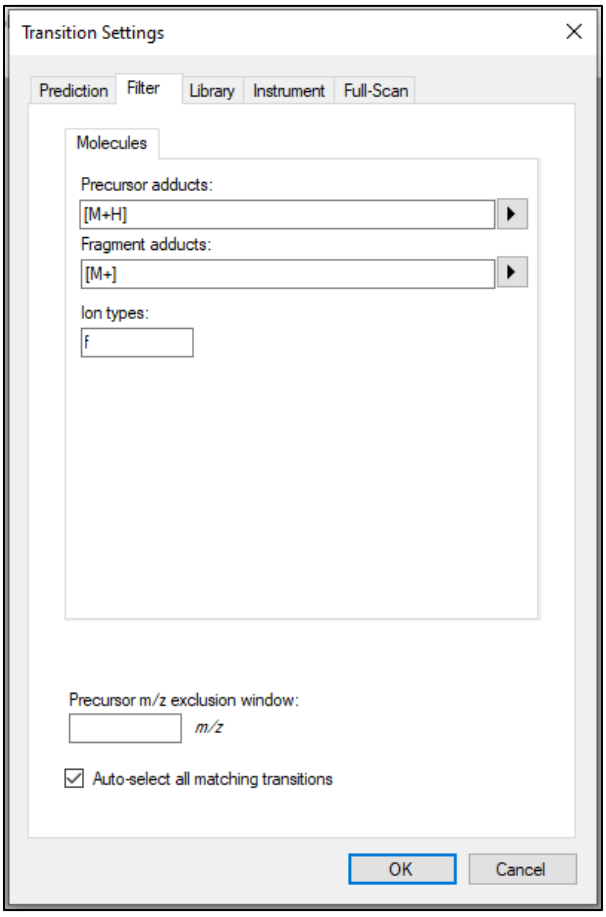


Inserting a Small Molecule Transition List

The below applies to Shimadzu and Thermo data only. Instructions how to create and then insert a small molecule transition list that works for Agilent data can be found in **Appendix**. To insert a list of target *m/z* to be extracted from GCMS data inside Skyline document, do the following:

- Open the provided *Shimadzu_TransitionList.csv* file in any text editor or any office productivity software suite (i.e. Excel).
- Highlight all rows and copy the content of .csv file to clipboard.
- Switch back to Skyline and there in **Settings** menu, choose **Transition settings,** and click **Filter** tab. In **Filter** tab ensure that only '**f**' (fragment) is used as **Ion types** and that **Auto-select all matching transitions** option is checked. Delete '**p**' (precursor) in **Ion types**, if necessary.

The **Transitions Settings** form should look like below:



- Click **OK** button
- Next, in **Edit** menu, choose **Insert** and click **Transition List**.

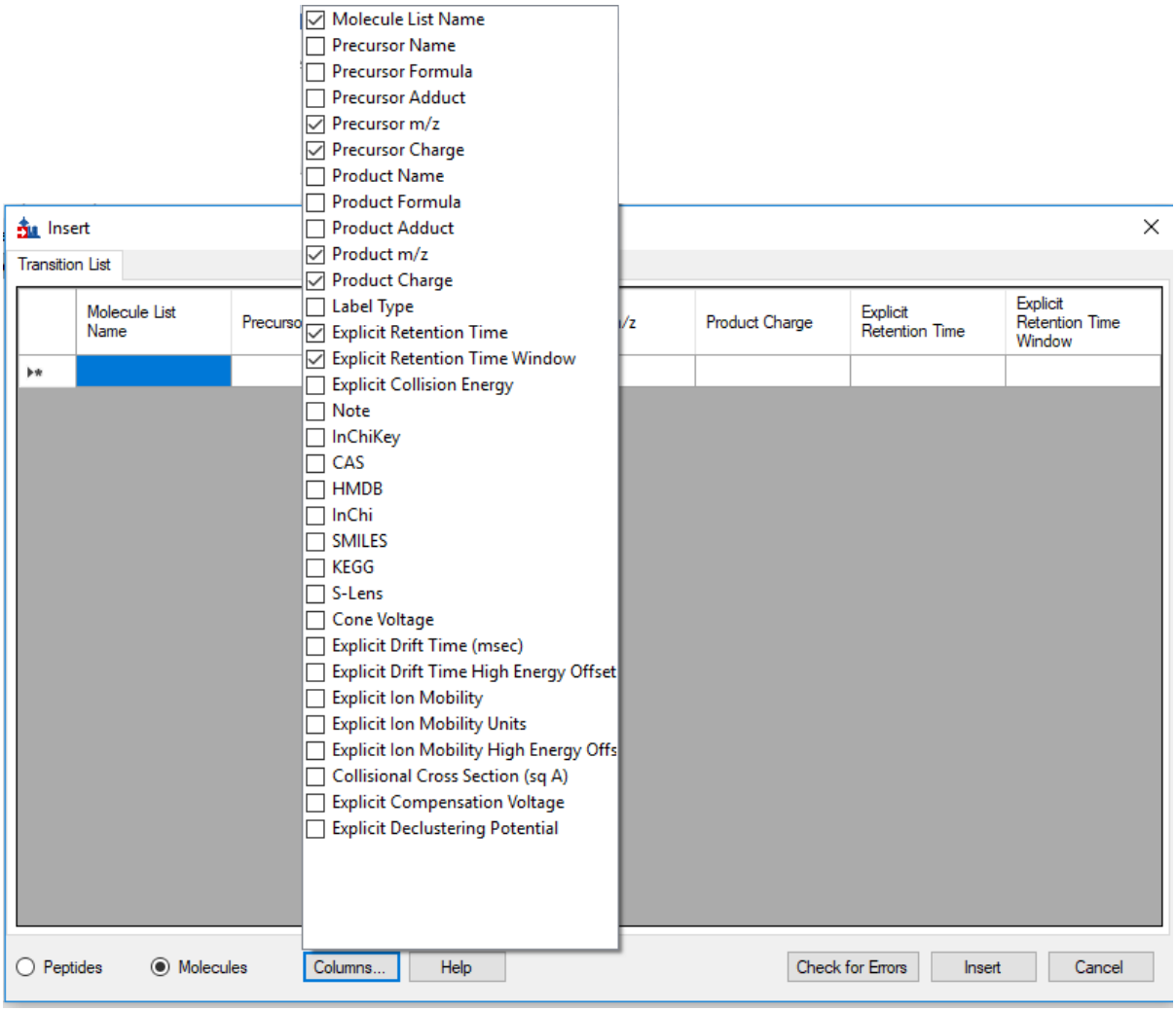
You should see the **Insert** form like below (the column order may differ if you modified it before):

There may be some extra column headers, or columns may be in a different order when compared to the columns present in the .csv file the content of which you copied into clipboard in earlier steps.

To remove extra column headers:

- Click **Columns** button and uncheck the columns that do not appear in the .csv file.

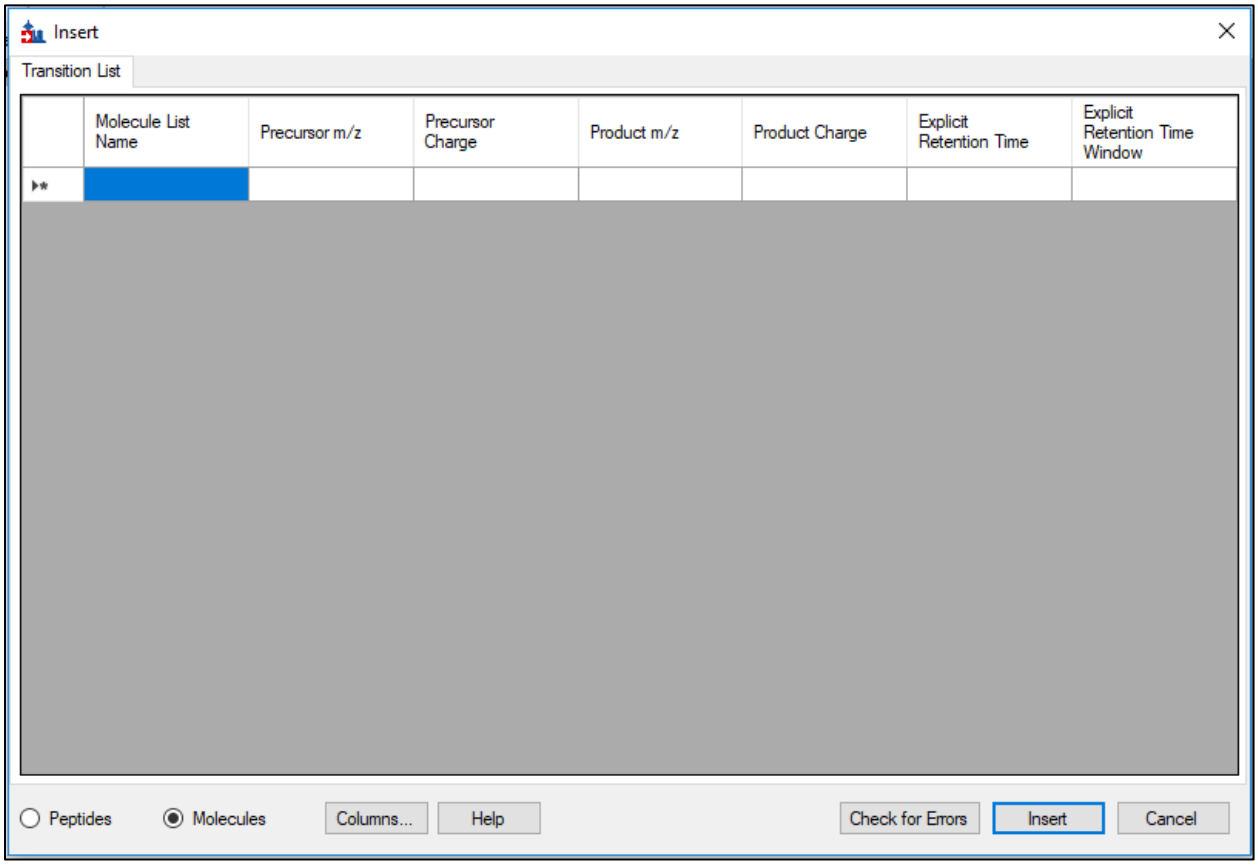
This should result in a column-picking menu like the one shown below:



To reorder the columns:

- Click and drag each column header you want to move to the order matching the .csv file

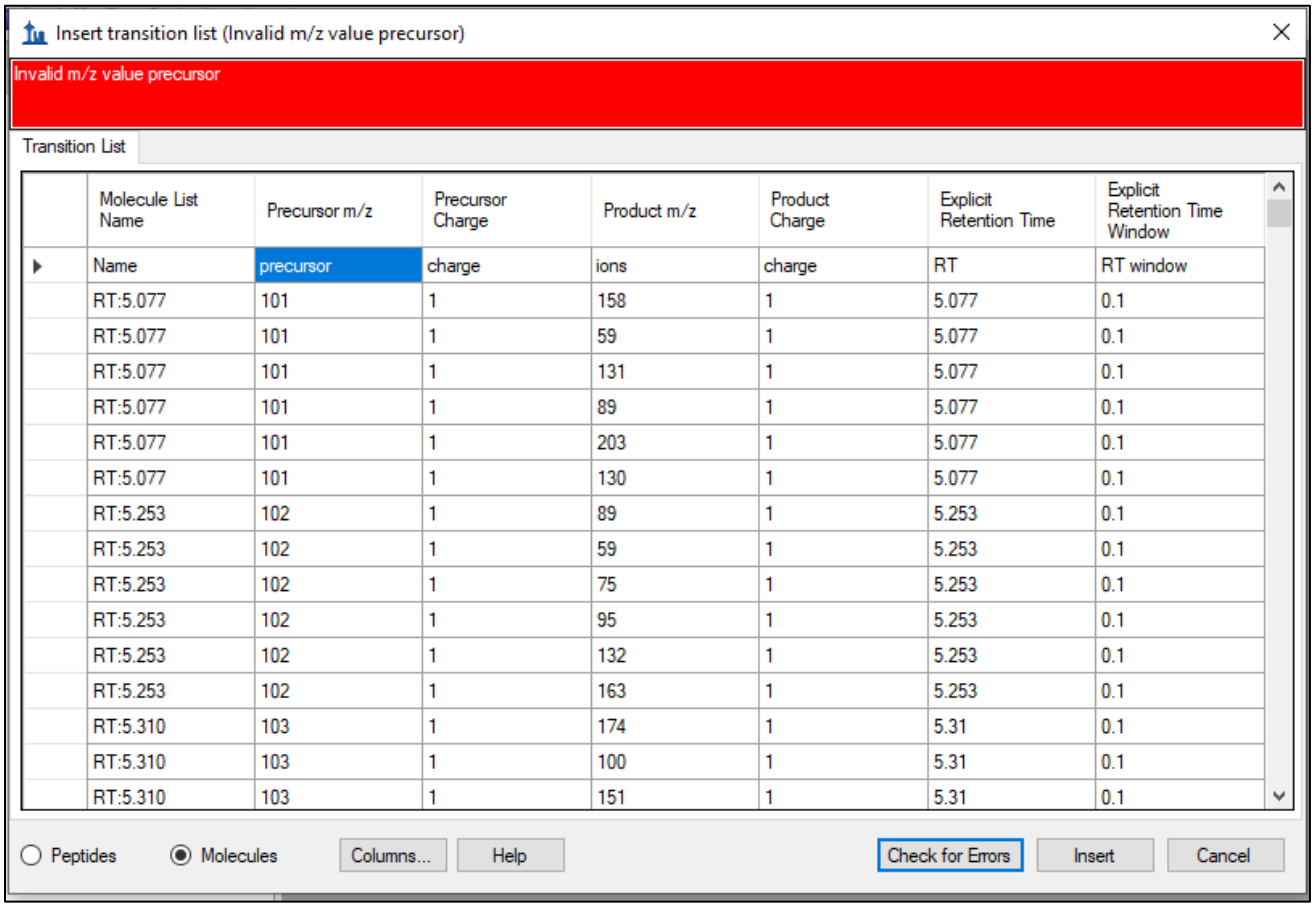
Once you have selected the required columns and rearranged them to match their order in .csv file, the **Insert** form should look like below:



Now:

- Paste the clipboard content into the **Insert** form by pressing **CTRL** and **V** keys simultaneously on the keyboard. Note that for the paste function to work correctly, the first cell of the form needs to be highlighted in blue as shown above. You can achieve this by clicking on that cell and then pressing **ESC** key on the keyboard. If the cell is not highlighted in blue but rather contains a black flashing line (text cursor) that allows you to type in that cell, pasting function will insert the entire content of the .csv file into that cell instead of populating it across the entire form.
- Click **Check for errors** button**.**

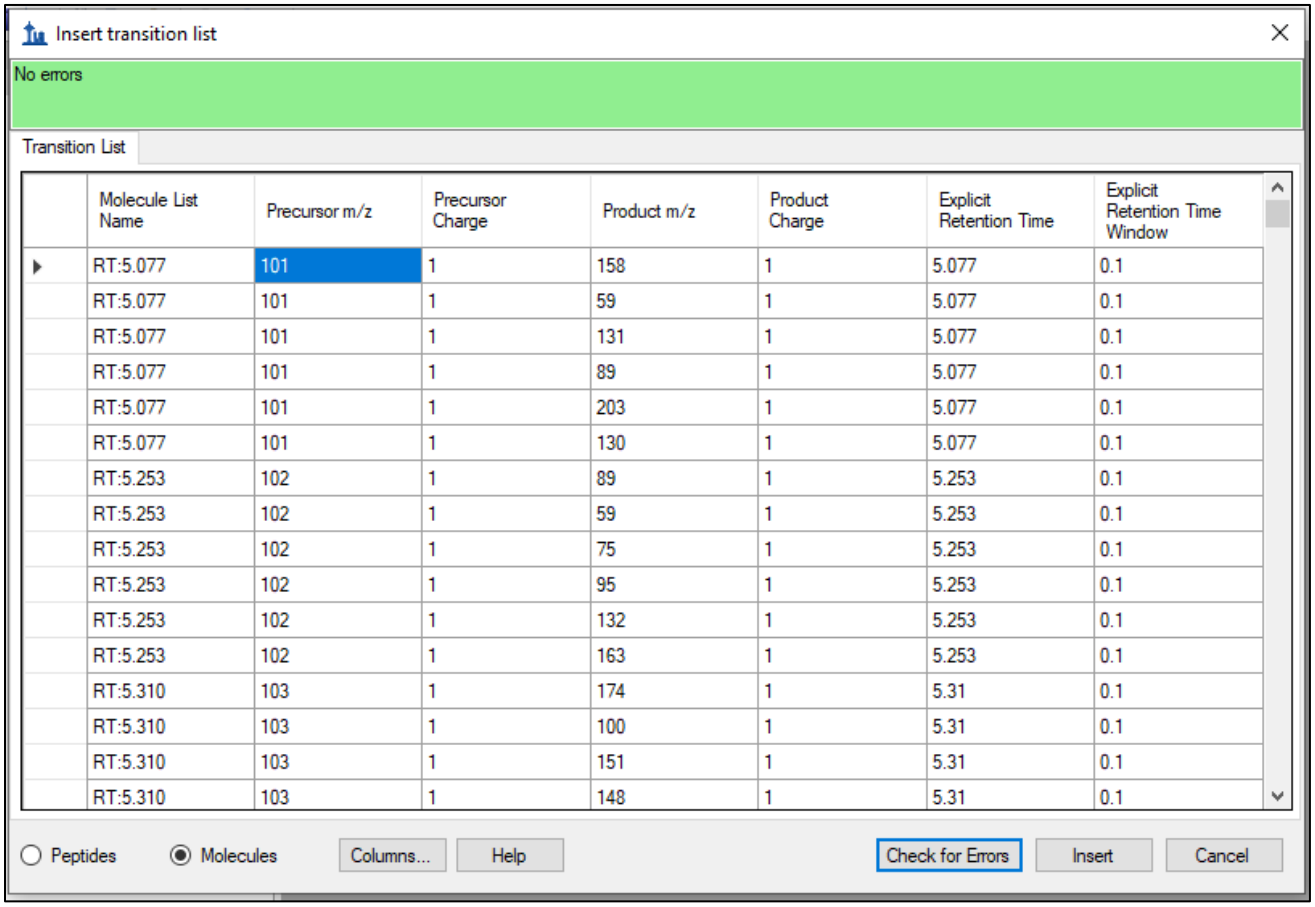
Because the .csv file contained the names of the columns in the first row, you will see an error message “Invalid *m/z* value precursor” like below:



To fix this error:

- Highlight the first row in the **Insert** form and press **DEL** key on the keyboard to delete it.
- Click **Check for Errors** button again.

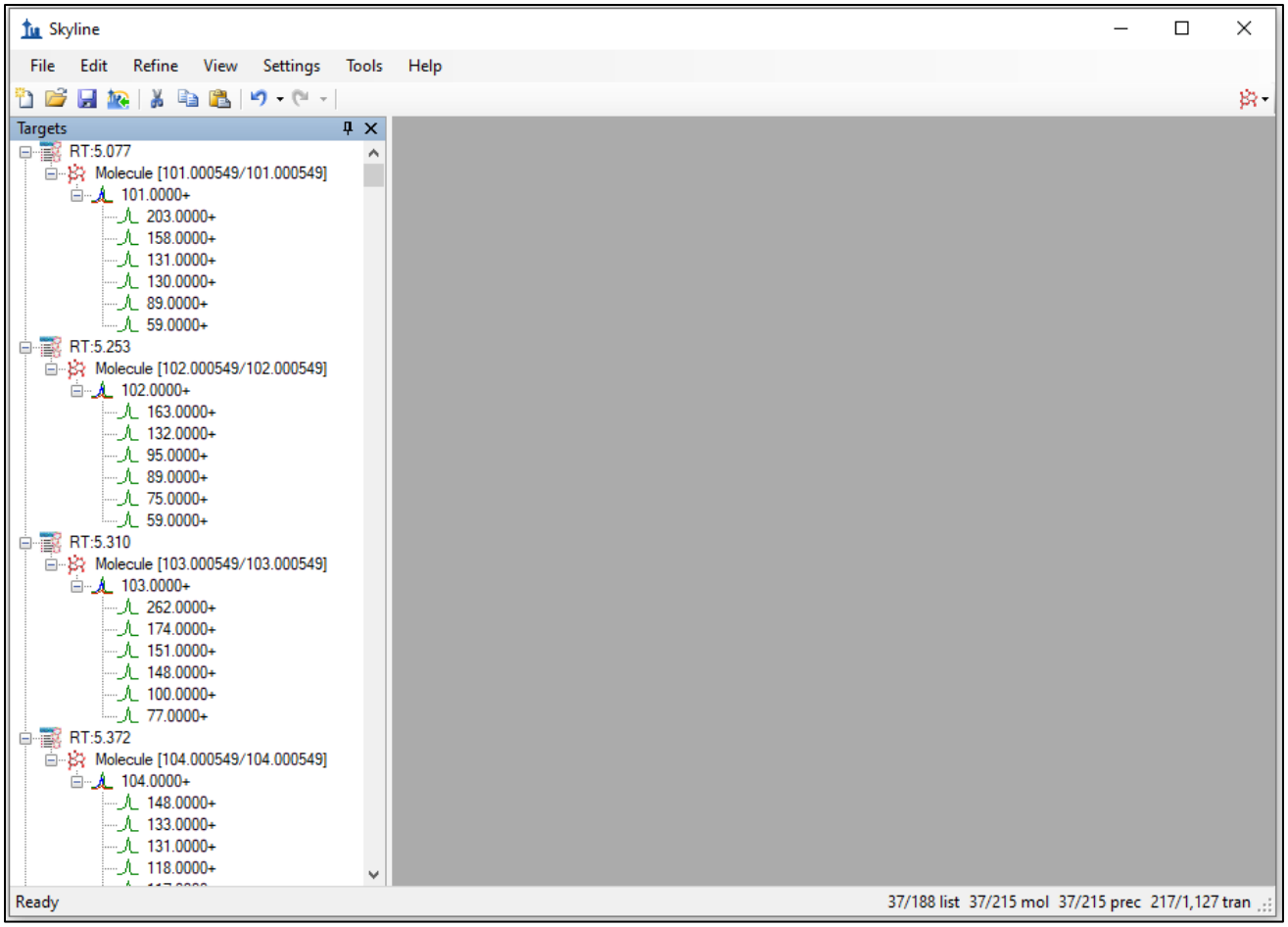
You should see “No errors” message like below:



Once you have confirmed that there are no more errors:

- Click **Insert** button in the bottom right corner of the **Insert** form.
- In **Edit** menu, choose **Expand All,** and click **Precursors**.

Your Skyline window should look like below, displaying a tree of chromatographic features (corresponding to unknown compounds) named according to the retention time (RT) at which a particular feature was integrated together with the *m/z* of fragment ions that were the most abundant at that RT. In the next steps, you will attempt to quantify these features by extracting the above *m/z* from GCMS data and plotting their abundance over time (XIC- based quantitation).



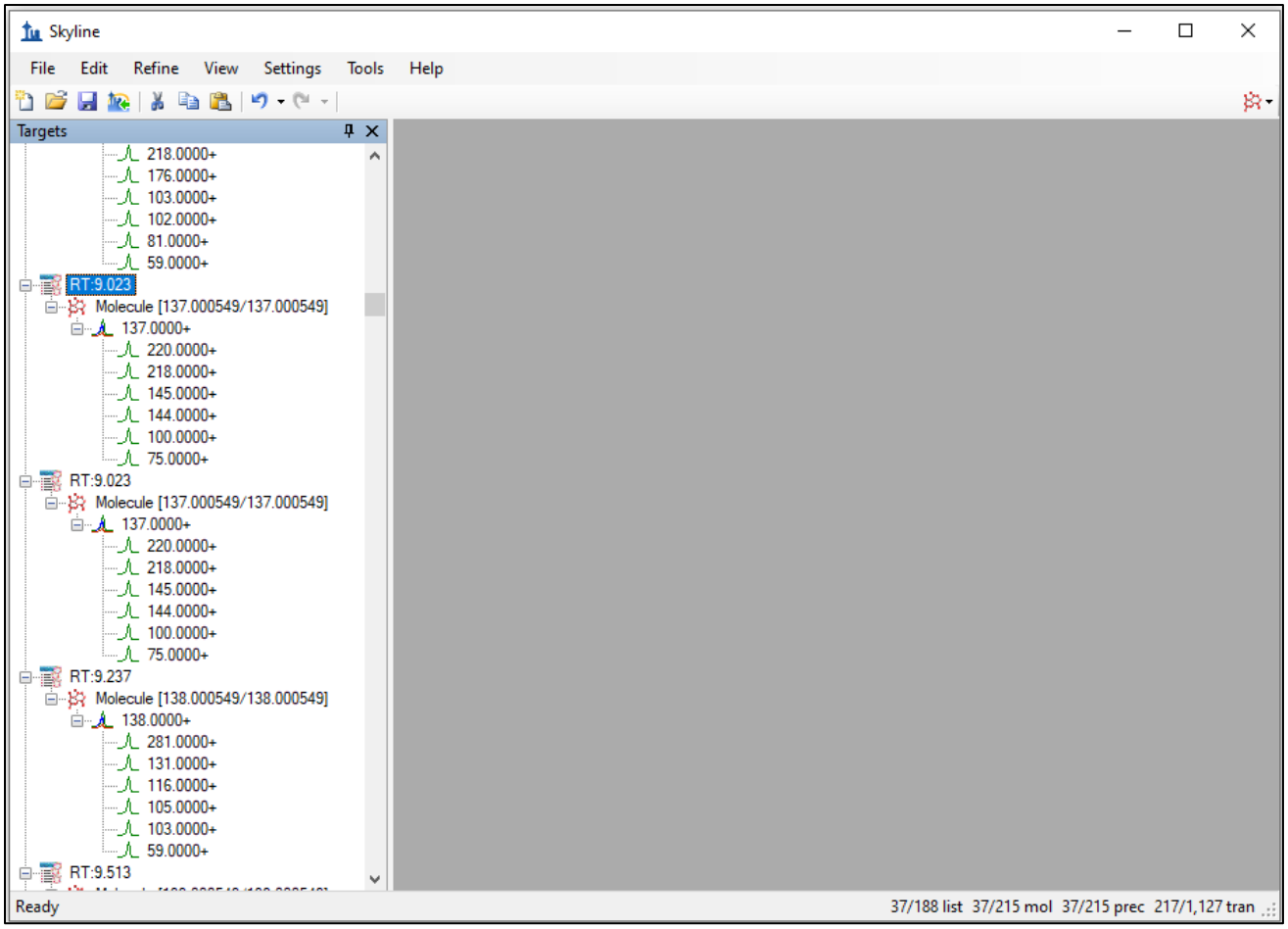
Note that the mass of a precursor ion is meaningless. It was made up to ensure compatibility of GCMS data with Skyline software and the later requires precursor ion information to build a transition list. Here 101, 102, … etc were used.

In the experiment that you will be analysing in Skyline, a heavy-labelled variant of Valine was used as surrogate internal standard (a chemical spiked into every sample at the same concentration prior sample preparation). This compound co-elutes with unlabelled Valine at 9.023 min and therefore, the compound at RT: 9.023 contains fragment ions characteristic to two different metabolites (labelled and unlabelled Valine). To quantify them separately, you need to assign fragment ions specific to each compound to a separate compound. You will achieve this by duplicating the compound at RT: 9.023, making the masses of a precursor ion in each compound different from each other, and then for each compound, you will delete fragment ions that are specific to the other compound.

To do this:

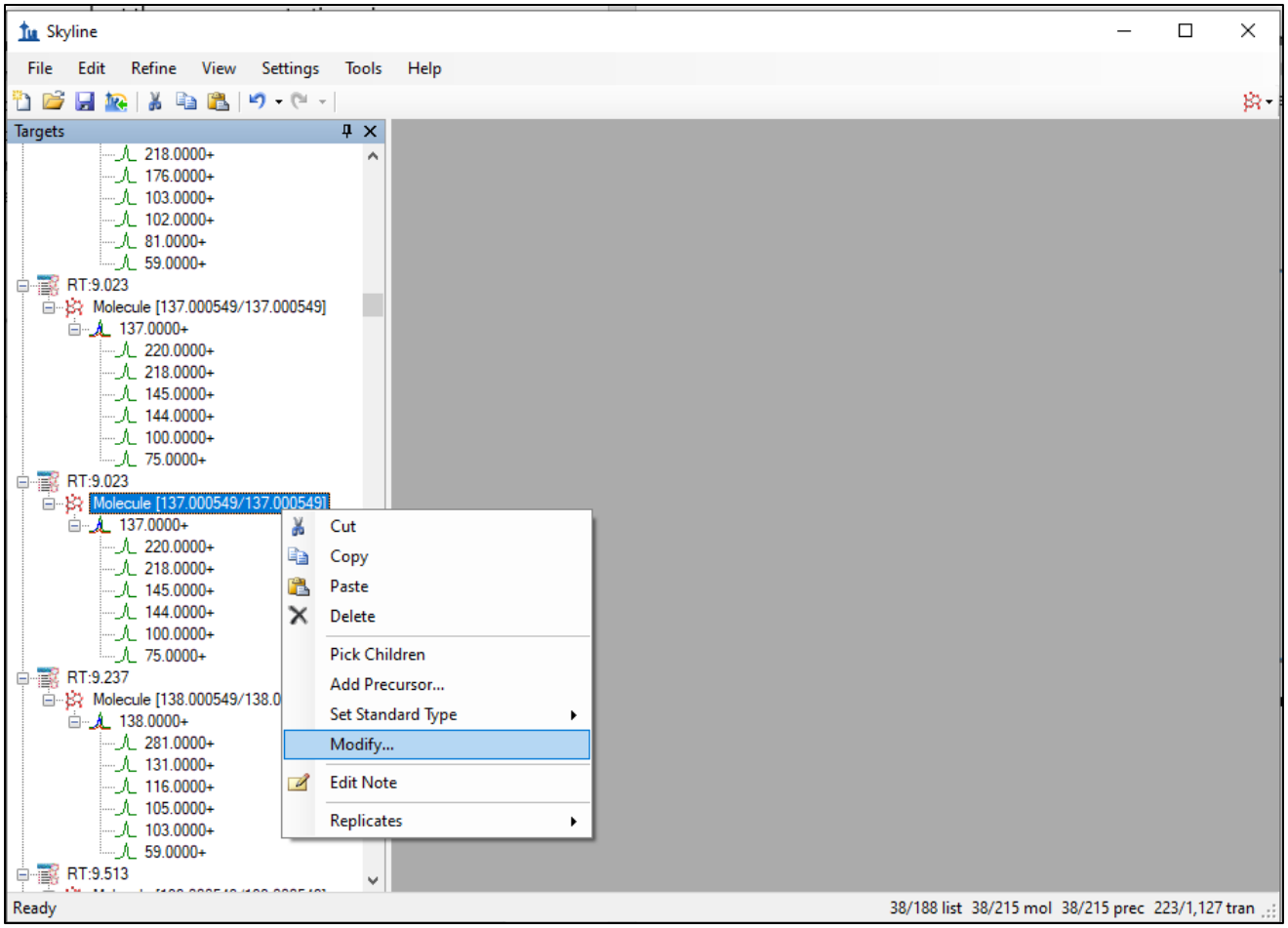
- Scroll down to RT: 9.023 and click on it
- In **Edit** menu, click **Copy** then **Paste**.

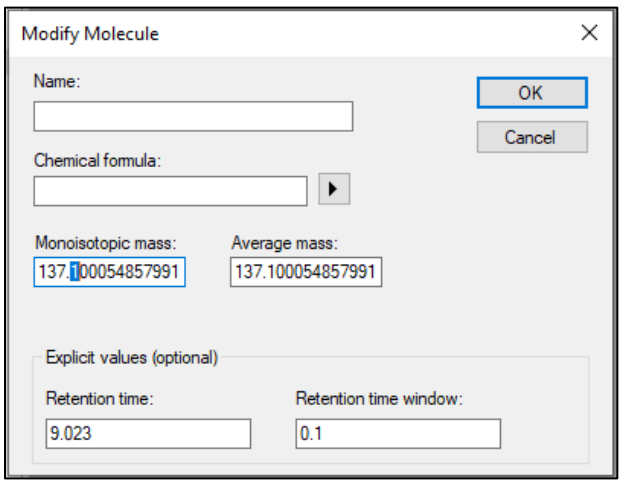
You will see two identical instances of RT: 9.023 like below:



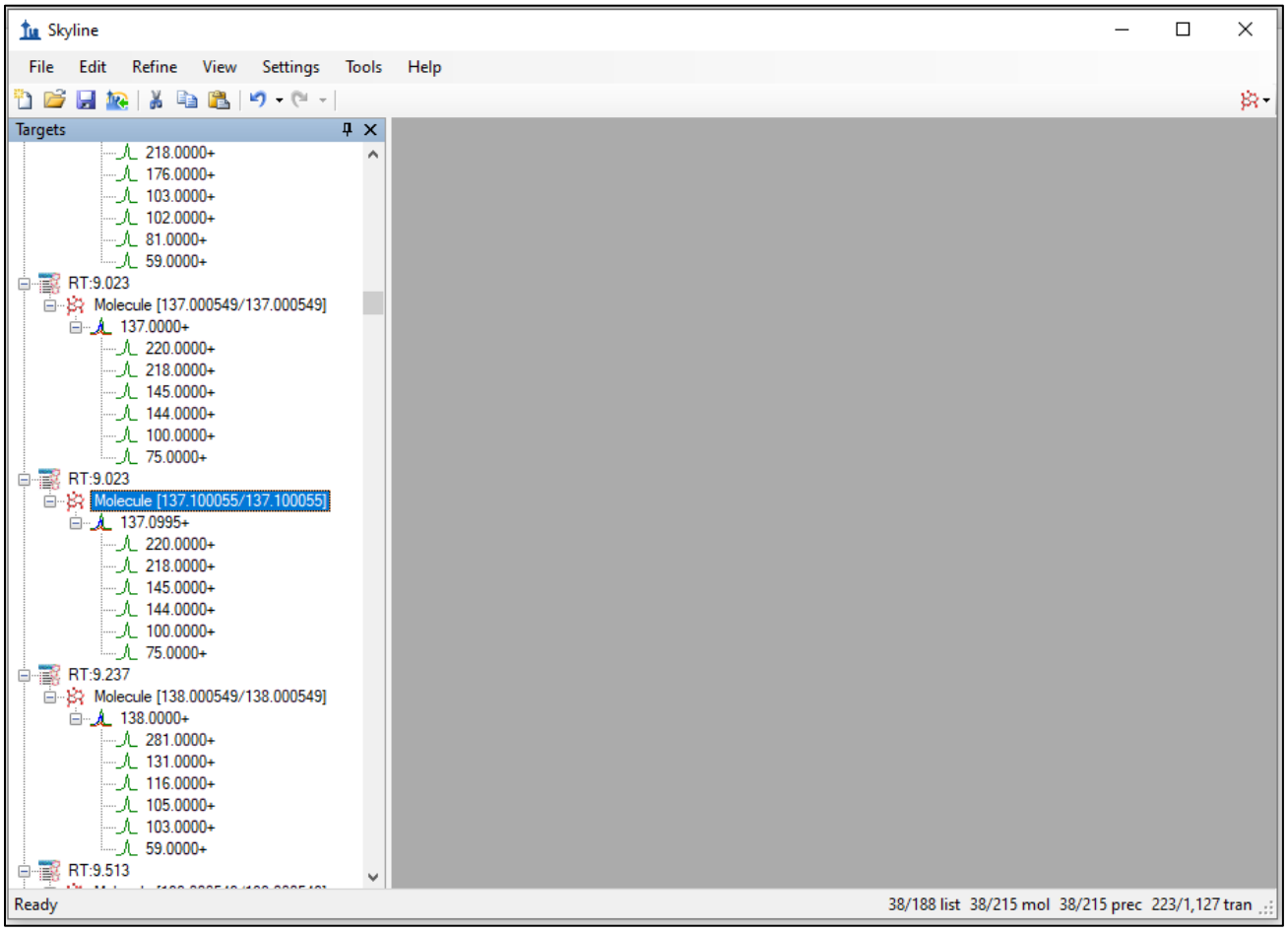
Next:

- In the case of the second instance of RT: 9.023, right click on **Molecule (137.000549/137.000549)** to open a pop-up menu and choose **Modify** to open **Modify Molecule** form. In this form, increase the **Monoisotopic mass** and **Average mass** by 0.1 as shown below and click **OK** button.


Skyline
File Edit Refine View Settings Tools Help
Targets
218.0000+
176.0000+
103.0000+
102.0000+
81.0000+
59.0000+
RT:9.023
Molecule [137.000549/137.000549]
137.0000+
220.0000+
218.0000+
145.0000+
144.0000+
100.0000+
75.0000+
RT:9.023
Molecule [137.000549/137.000549]
137.0000+
220.0000+
218.0000+
145.0000+
144.0000+
100.0000+
75.0000+
RT:9.237
Molecule [138.000549/138.0
138.0000+
281.0000+
131.0000+
116.0000+
105.0000+
103.0000+
59.0000+
RT:9.513
Cut
Copy
Paste
Delete
Pick Children
Add Precursor...
Set Standard Type
Modify...
Edit Note
Replicates
Ready
38/188 list 38/215 mol 38/215 prec 223/1,127 tran


Modify Molecule
Name:
OK
Cancel
Chemical formula:
Monoisotopic mass:
Average mass:
137.100054857991
137.100054857991
Explicit values (optional)
Retention time:
Retention time window:
9.023
0.1

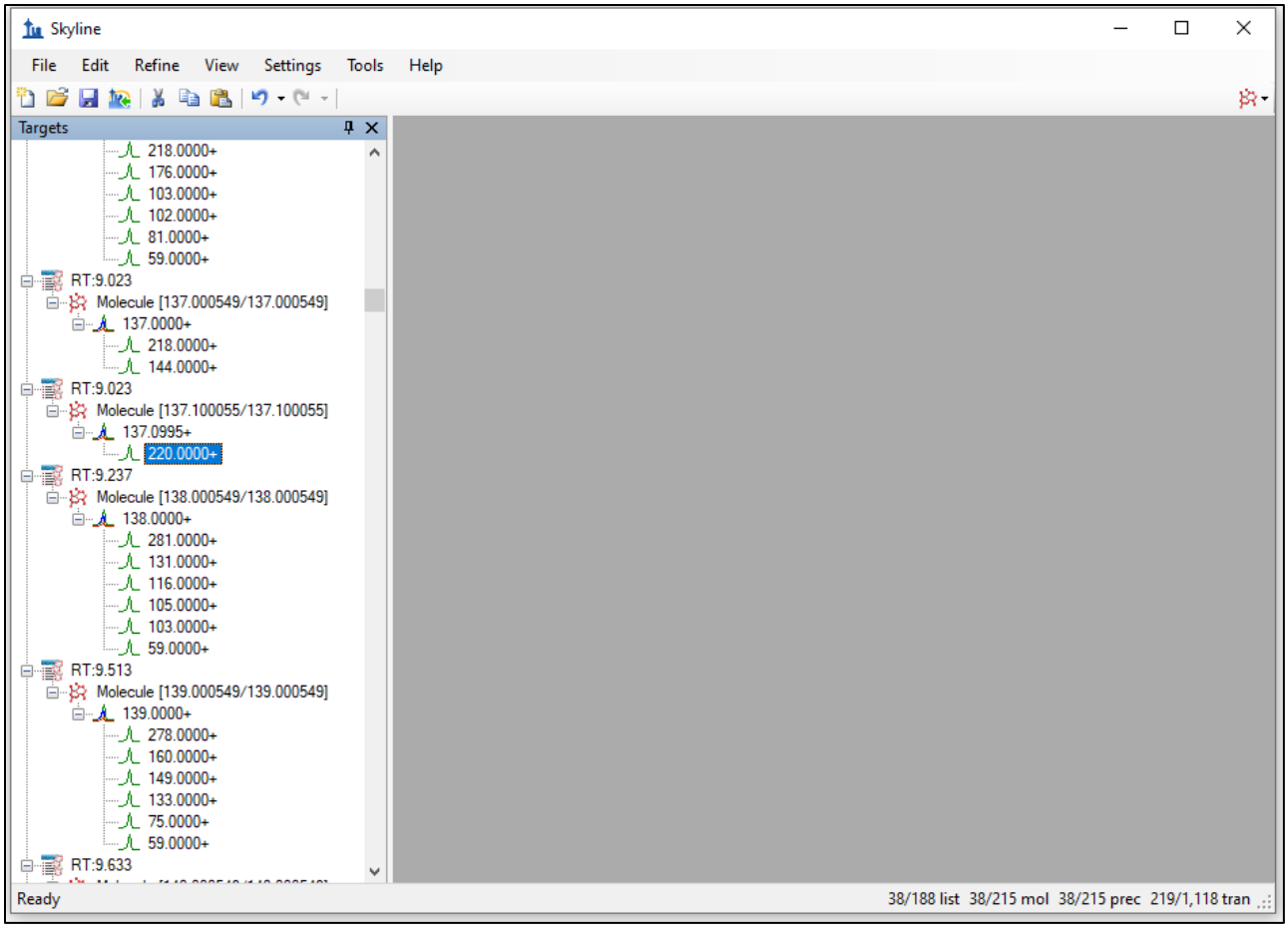
You will see that the precursor ion mass in the case of the second instance of RT: 9.023 has changed like below:



Now:

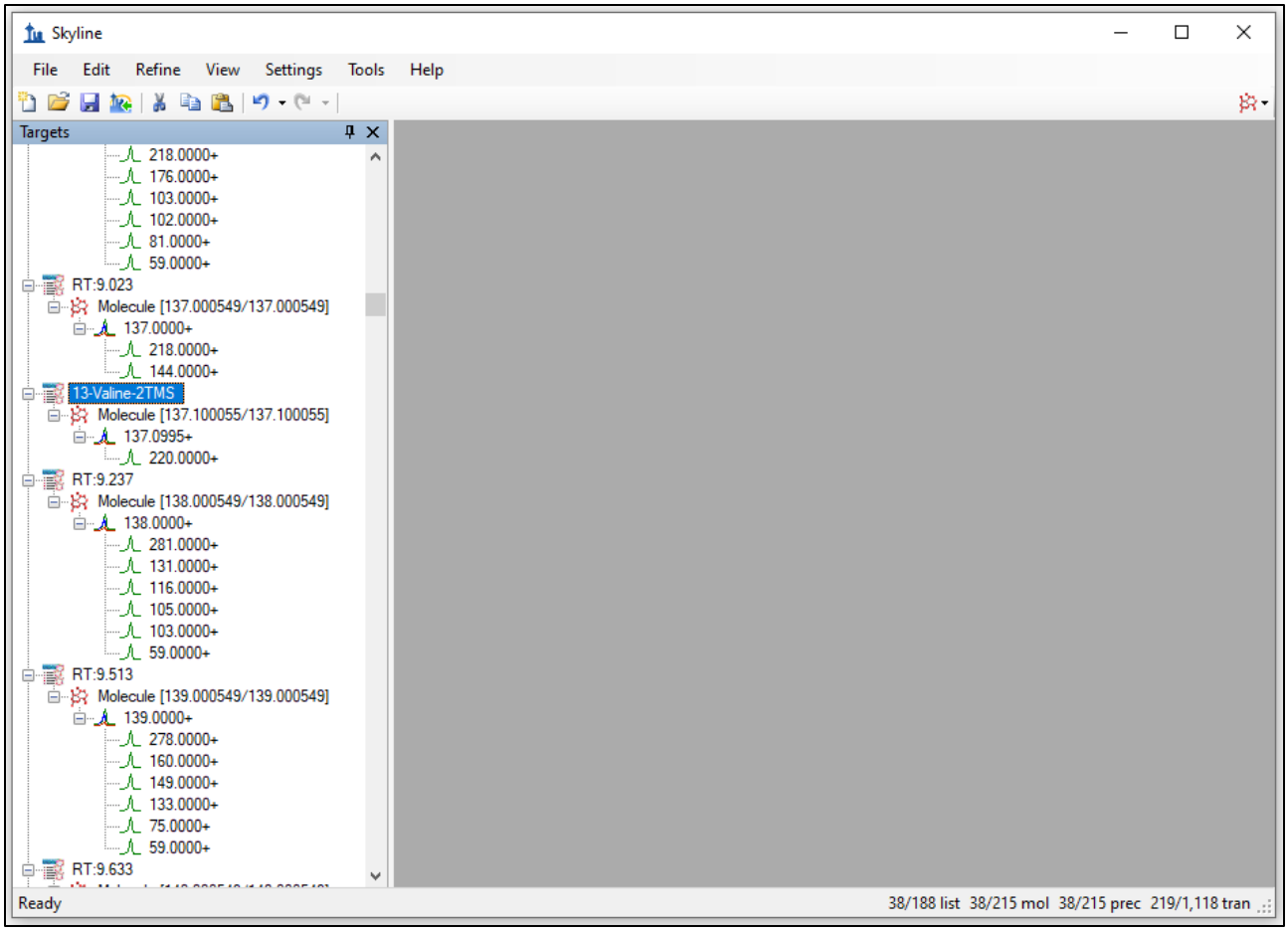
- In the case of the first instance of RT: 9.023 (the one with the original *m/z*), delete all but 218 and 144 ions by right clicking on them and choosing **Delete** in the pop-up menu.
- In the case of the second instance of RT: 9.023 (the one with modified *m/z*), delete all but 220 ion.

You should see them like this:



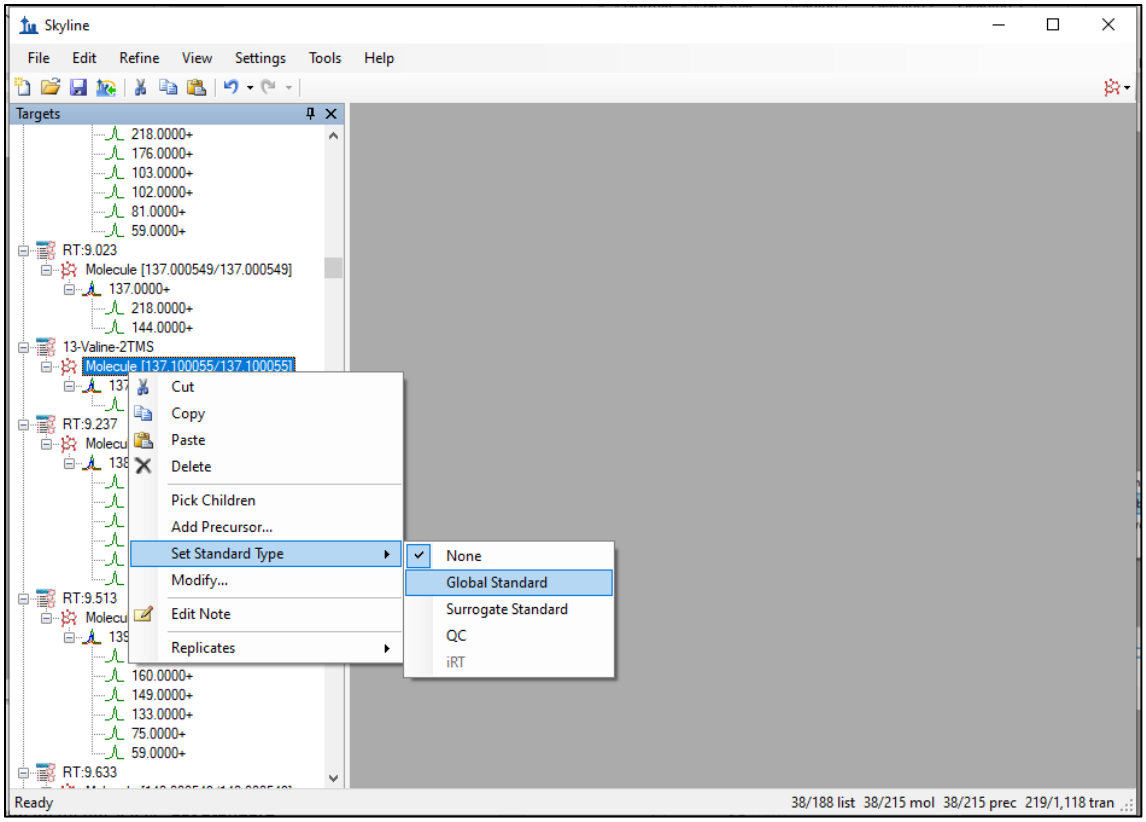
Next:

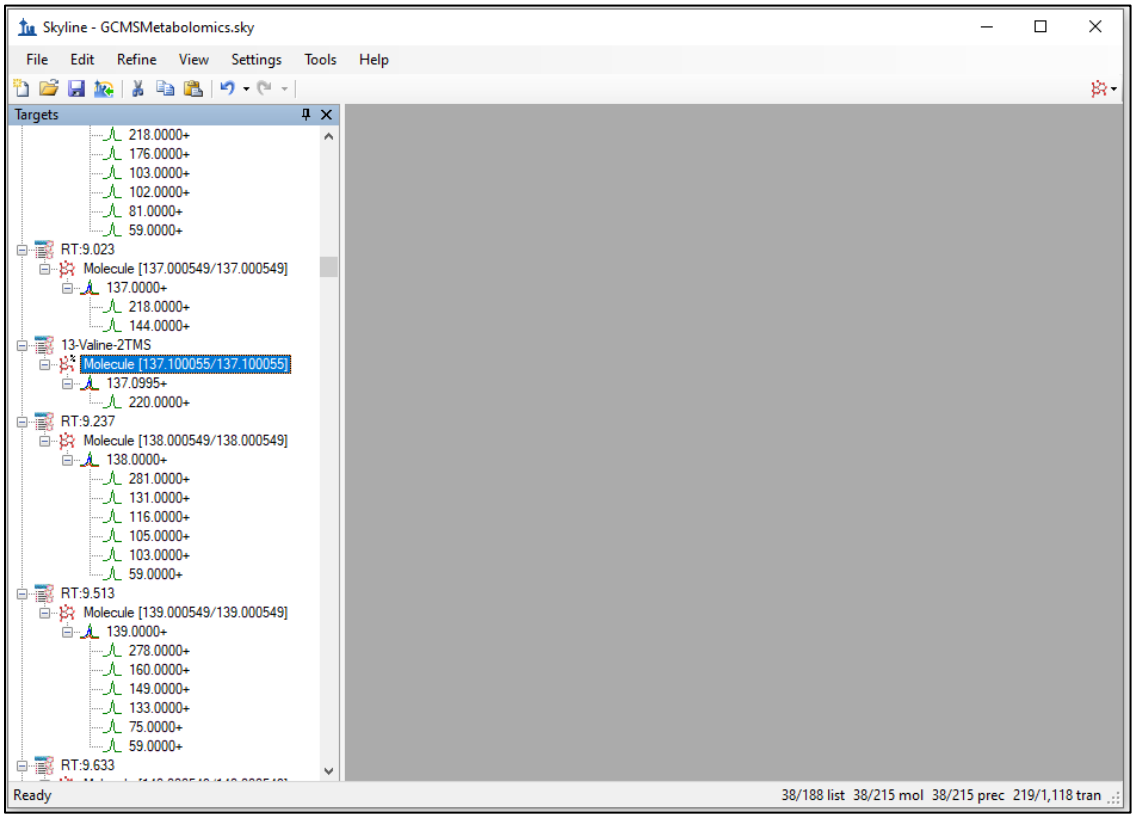
- Rename the second instance of RT: 9.023 to 13C-Valine-2TMS by left clicking on RT: 9.023, typing the new name, and then pressing **ENTER** key on the keyboard.



Finally, set 13C-Valine-2TMS as a global standard. To do this:

- Right click on **Molecule (137.100055/137.100055)** to open a pop-up menu, choose **Set Standard Type**, and then click **Global Standard** option. A small "**%**" symbol should be now visible next to the **Molecule (137.100055/137.100055)**.


Skyline
File Edit Refine View Settings Tools Help
Targets
218.0000+
176.0000+
103.0000+
102.0000+
81.0000+
59.0000+
RT:9.023
Molecule [137.000549/137.000549]
137.0000+
218.0000+
144.0000+
13-Valine-2TMS
Molecule [137.100055/137.100055]
Cut
Copy
Paste
Delete
Pick Children
Add Precursor...
Set Standard Type
Modify...
Edit Note
Replicates
None
Global Standard
Surrogate Standard
QC
iRT
RT:9.237
RT:9.513
160.0000+
149.0000+
133.0000+
75.0000+
59.0000+
RT:9.633
Ready
38/188 list 38/215 mol 38/215 prec 219/1,118 tran


Skyline - GCMSMetabolomics.sky
File Edit Refine View Settings Tools Help
Targets
218.0000+
176.0000+
103.0000+
102.0000+
81.0000+
59.0000+
RT:9.023
Molecule [137.000549/137.000549]
137.0000+
218.0000+
144.0000+
13-Valine-2TMS
Molecule [137.100055/137.100055]
137.0995+
220.0000+
RT:9.237
Molecule [138.000549/138.000549]
138.0000+
281.0000+
131.0000+
116.0000+
105.0000+
103.0000+
59.0000+
RT:9.513
Molecule [139.000549/139.000549]
139.0000+
278.0000+
160.0000+
149.0000+
133.0000+
75.0000+
59.0000+
RT:9.633
Ready
38/188 list 38/215 mol 38/215 prec 219/1,118 tran

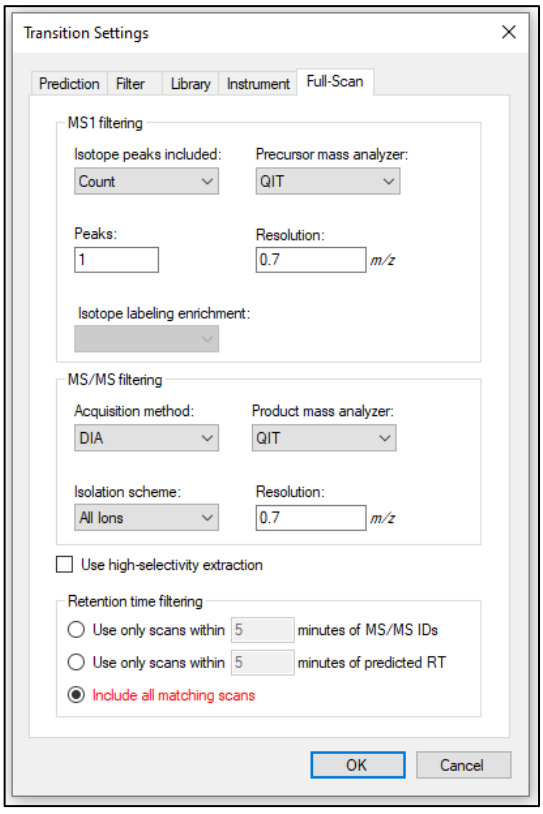
Note that the above steps allowed you to conduct a simple deconvolution of GCMS data inside Skyline. More on deconvolution later.

Also note that in this experiment there were in fact two internal standards spiked, 13C-Valine and 13C-Sorbitol. The former for normalisation of peak areas of earlier eluting compounds (mostly organic acids and amino acids) and the later for normalisation of peak areas of later eluting compounds (mostly mono-, di- and trisaccharides). Fatty acids and sterols enriched in the chloroform phase were discarded during sample preparation. Normalisation of different compounds to a different internal standard can be enabled by selecting **Surrogate Standard** option instead of **Global Standard** option in the **Set Standard Type** menu, however, beyond the scope of this tutorial.

Preparing for Mass Spectrometer Data Import

To set up Skyline for importing full MS scan GCMS data, do the following:

- In **Settings** menu, choose **Transition Settings** and then **Full-Scan** tab.
- Change settings as necessary to match the following and click **OK** button



Note that it is possible that at this stage Skyline automatically changed the settings in the **Filter** tab and added '**p**', precursor, in **Ion types**. If this happened, delete it again, and click **OK** button. It is also possible that depending on Skyline version, enabling **MS1 filtering** may or may not be required.

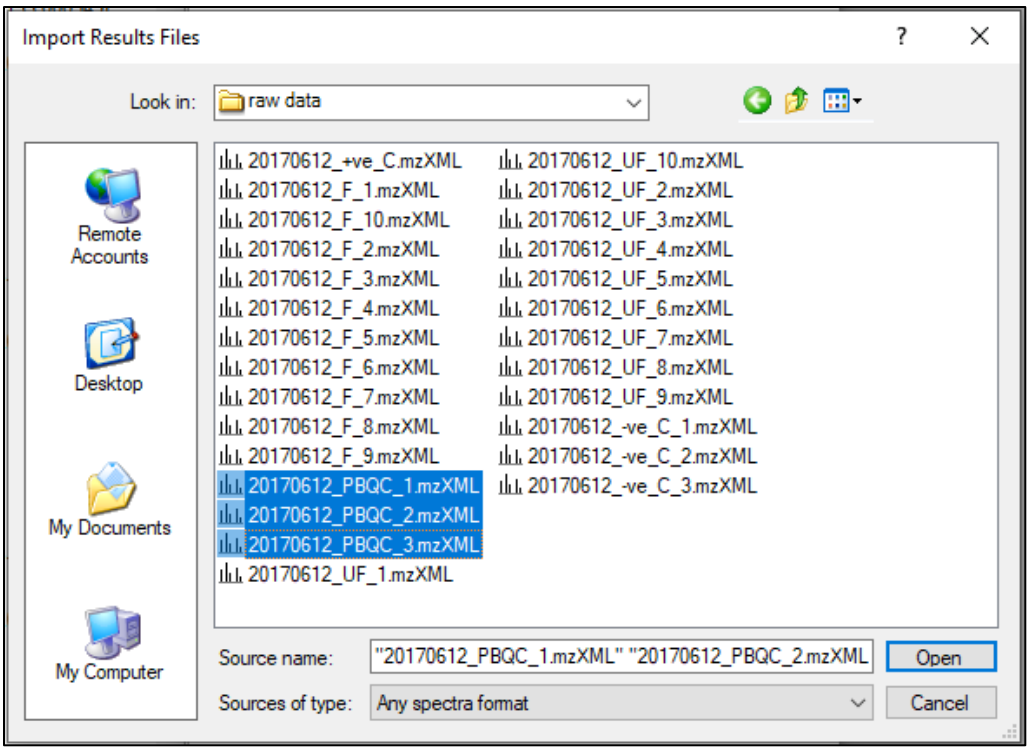
Now you are ready to import the experimental mass spectrometer results.

Importing Mass Spectrometer Data Corresponding to PBQC

Perform the following steps:

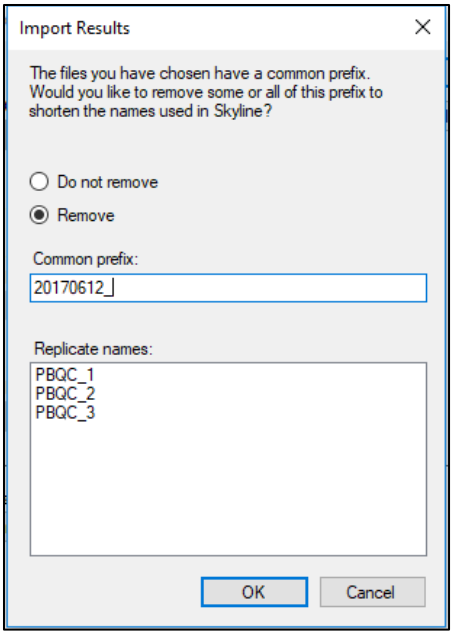
- In **File** menu, click **Save As** and save the current Skyline document as "*GCMSMetabolomics.sky*".
- In **File** menu, select **Import** and click **Results**.
- In **Import Results** form, choose **Add single-injection replicates in files** option. For best performance, be sure to select **Many** option in **Files to import simultaneously** control at the bottom of the form.
- Click **OK** button.
- Browse to a folder with tutorial data. Select 3 injection replicates of PBQC sample by clicking the first PBQC file and then holding down the **SHIFT** key and clicking the last PBQC file.

This will result in the **Import Results Files** form look like this:

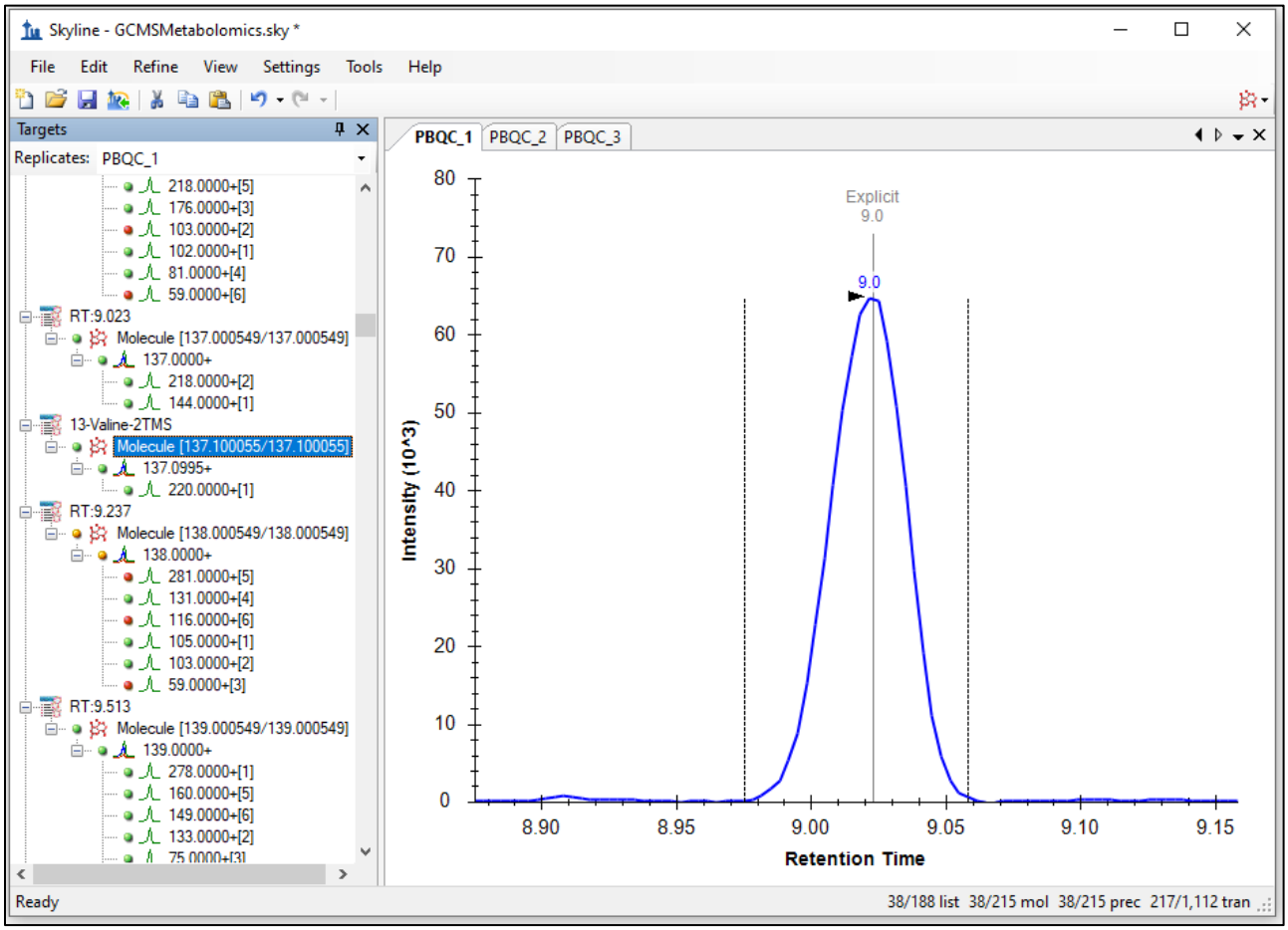


Now:

- Click **Open** button.
- Skyline will present you with an option to remove common prefix that allows you to shorten the names of the files and therefore, makes it easier to analyse the data on smaller screens. Select **Remove** option, and type “20170612_” in **Common prefix** box, and then click **OK** button.



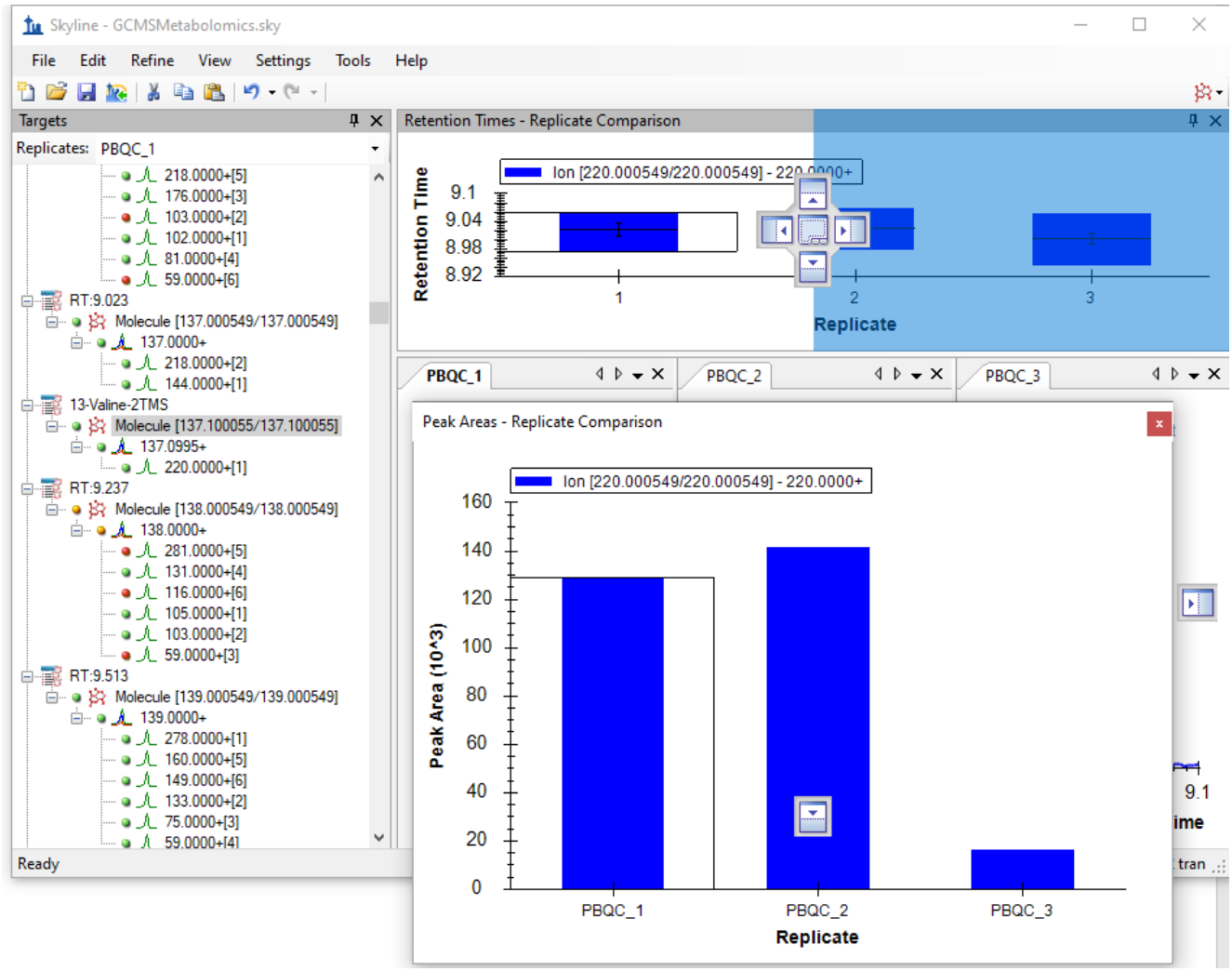
The files should import within 2 minutes or so, leaving your Skyline window look like this:

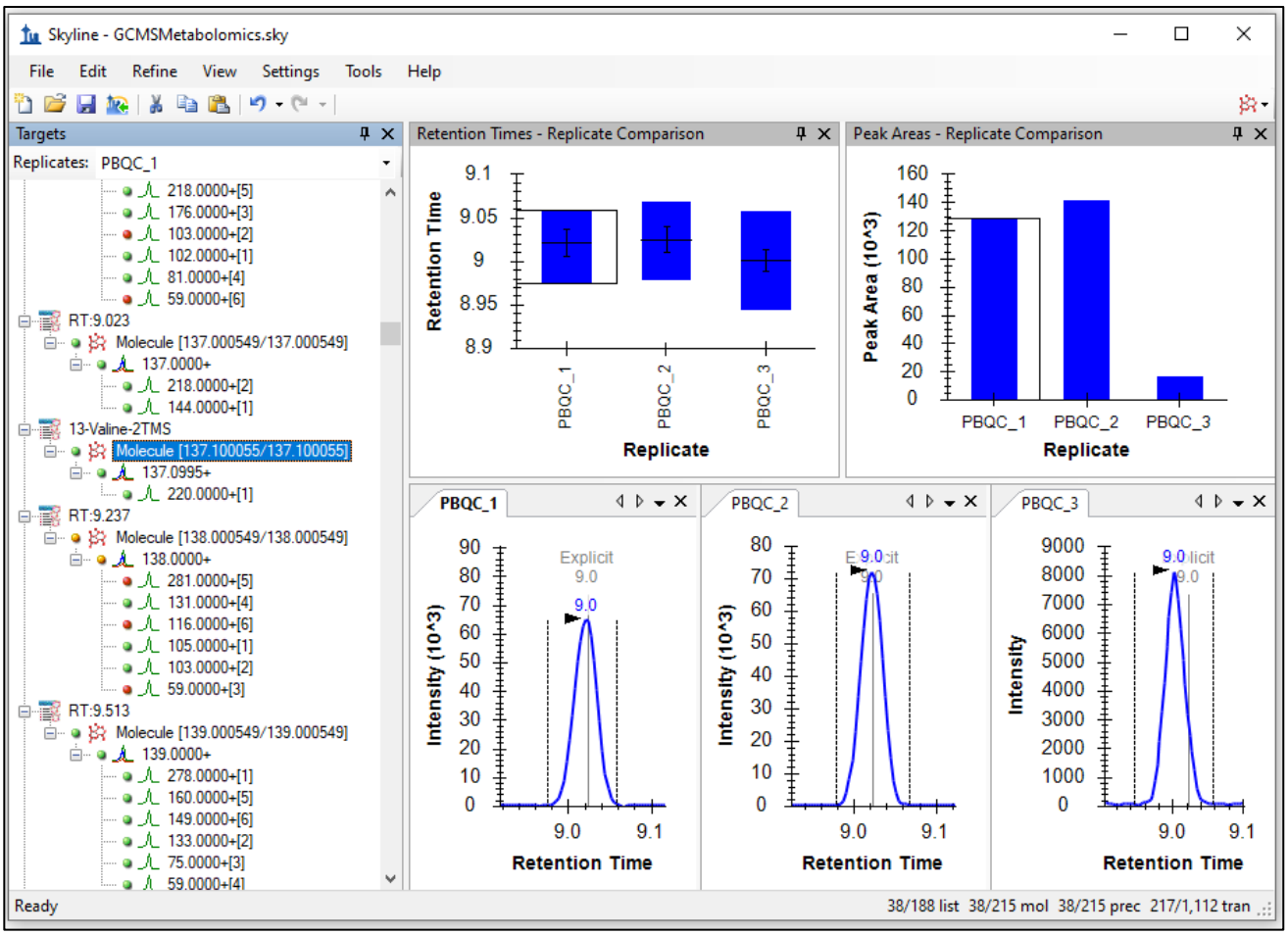


Once the files have been imported, take advantage of view customization options and summary graphs:

- In **View** menu, choose **Arrange Graphs** and click **Row**.
- Right click on any chromatogram graph and choose **Auto-zoom X-axis** and then click **Best Peak.**
- Again, right click on any chromatogram graph and choose **Properties** then type "0.2" in **Best peak time range,** untick **Peak width relative** and click **OK** button**.**
- In **View** menu, choose **Peak Areas** and click **Replicate Comparison**.
- In **View** menu, choose **Retention Times** and click **Replicate Comparison**.
- Left click and drag these plots to the top and drop them onto the right arm of a blue cross to dock them above the chromatogram graphs. The view is fully customizable. You can use the separating bars between different panels to adjust the view further to make certain panels appear wider or taller.

Below are Skyline windows before and after view customization:


Skyline - GCMSMetabolomics.sky
File Edit Refine View Settings Tools Help
Targets
Replicates: PBQC_1
218.0000+[5]
176.0000+[3]
103.0000+[2]
102.0000+[1]
81.0000+[4]
59.0000+[6]
RT:9.023
Molecule [137.000549/137.000549]
137.0000+
218.0000+[2]
144.0000+[1]
13-Valine-2TMS
Molecule [137.100055/137.100055]
137.0995+
220.0000+[1]
RT:9.237
Molecule [138.000549/138.000549]
138.0000+
281.0000+[5]
131.0000+[4]
116.0000+[6]
105.0000+[1]
103.0000+[2]
59.0000+[3]
RT:9.513
Molecule [139.000549/139.000549]
139.0000+
278.0000+[1]
160.0000+[5]
149.0000+[6]
133.0000+[2]
75.0000+[3]
59.0000+[4]
Ready
Retention Times - Replicate Comparison
Retention Time
9.1
9.04
8.98
8.92
1
2
3
Replicate
PBQC_1
PBQC_2
PBQC_3
Peak Areas - Replicate Comparison
Ion [220.000549/220.000549] - 220.0000+
Peak Area (10^3)
160
140
120
100
80
60
40
20
0
PBQC_1
PBQC_2
PBQC_3
Replicate



Verifying Batch Using PBQC Sample

In metabolomics experiments, pooled sample (PBQC) is injected several times during the acquisition of a batch and provides a means to monitor the sensitivity and stability of the GCMS system and thus allows one to assess (and potentially correct for) any bias in quantitative data.

Click on a few different compound entries to get familiar with how PBQC data looks in Skyline:

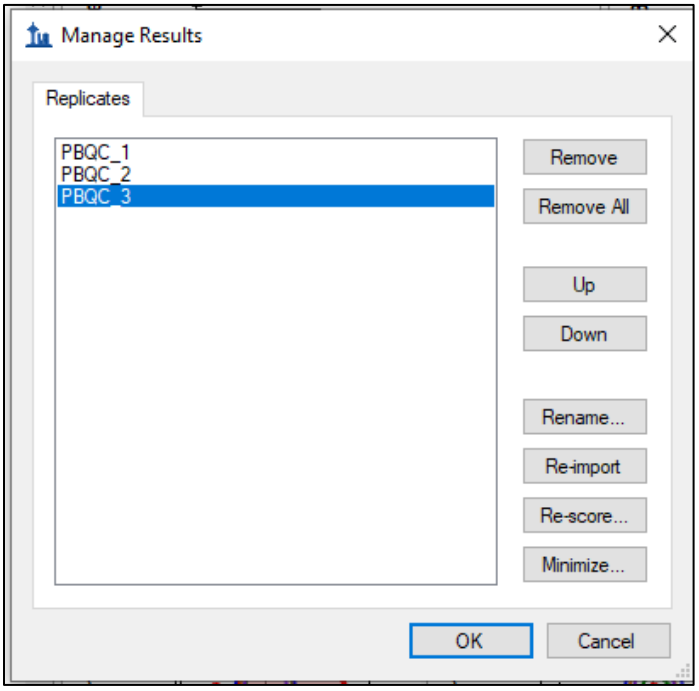
- Use **Retention Times – Replicate Comparison** plot to verify if the compounds are integrated consistently across all 3 PBQC injections.
- Use **Peak Areas – Replicate Comparison** plot to verify if the ion intensities are consistent across all 3 PBQC injections.

Note that there is something different about the third injection of PBQC sample (PBQC_3). Peaks tend to have a significantly different intensity and slightly different elution time, and the later

occasionally confuses peak picking algorithm in Skyline. In this experiment, a single vial with PBQC sample was prepared and analysed at three different times during acquisition of the batch and likely evaporation affected the concentration (ideally, PBQC should have been aliquoted into several vials to ensure that each injection is conducted from a new vial to avoid issue with evaporation). It is also possible that there was a problem with instrument performance when running later samples. Therefore, one should be cautious about analysing samples that were run shortly before or after PBQC_3. For this tutorial, you will assume that the problem was associated with evaporation and simply delete the third data file. In other case, it is possible to model PBQC data and use it for normalization of samples however, beyond the scope of this tutorial.

To remove the last PQBC sample from Skyline document:

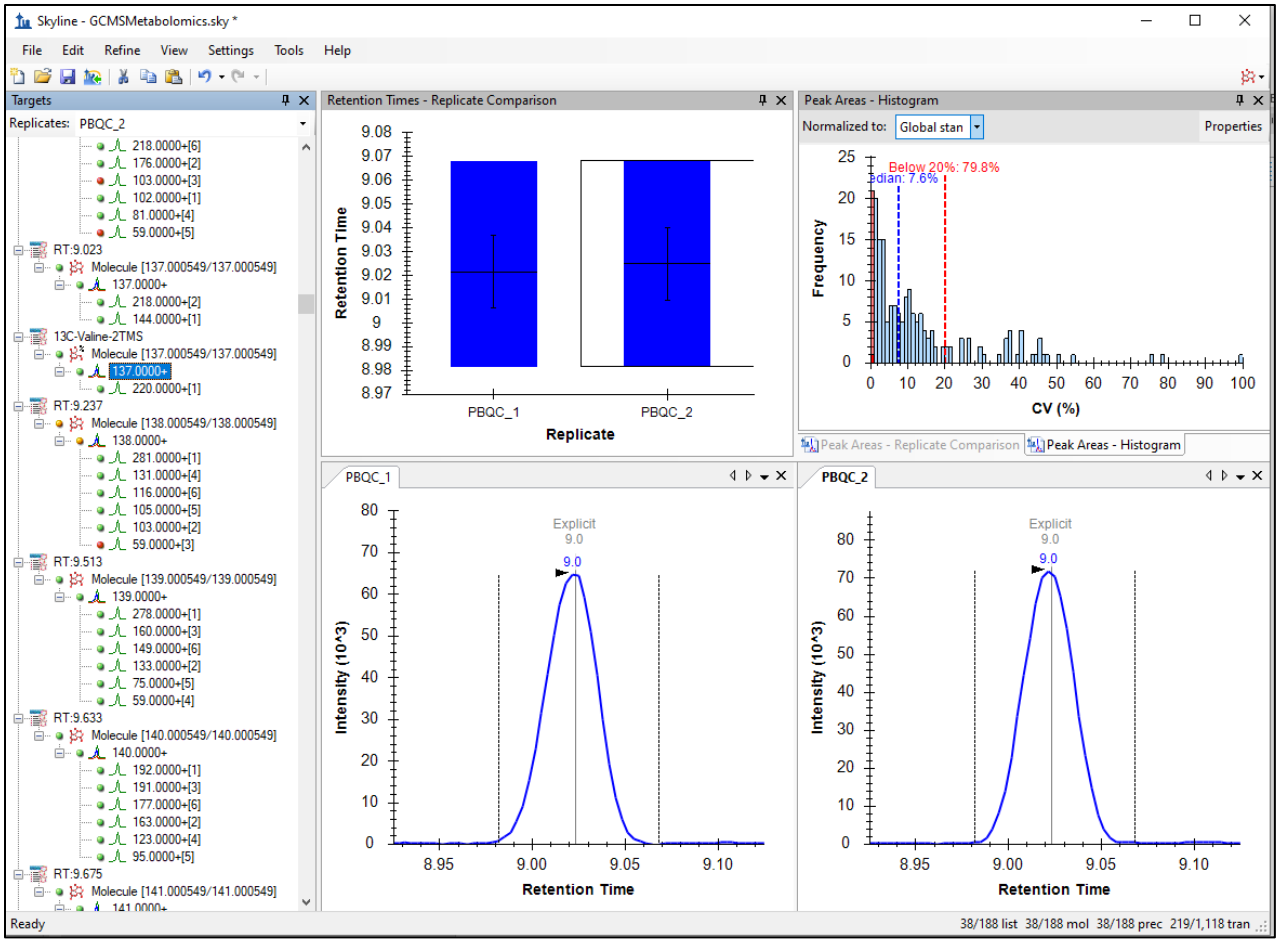
- In **Edit** menu choose **Manage Results**
- Highlight "PBQC_3" sample and click **Remove** button



Now:

- Use **Peak Areas – CV Histogram** window (available from **View** menu > **Peak areas**) for an overview of PBQC data.

You should see the **Peak Areas – Histogram** plot in the top right corner like below:



Note that ~ 80% data have CV below 20% (when **Normalized to: Global standard** option is selected), and that the median CV is ~7%, which suggest that the instrument performance was acceptable during the acquisition of samples run between the 2 PBQC injections that were kept in Skyline document. Also, the plot is clickable and allows one for additional refinement of data (i.e. removal of compounds with CV greater than certain value) however, in this tutorial you will leave data as is and let statistics to take care of everything.

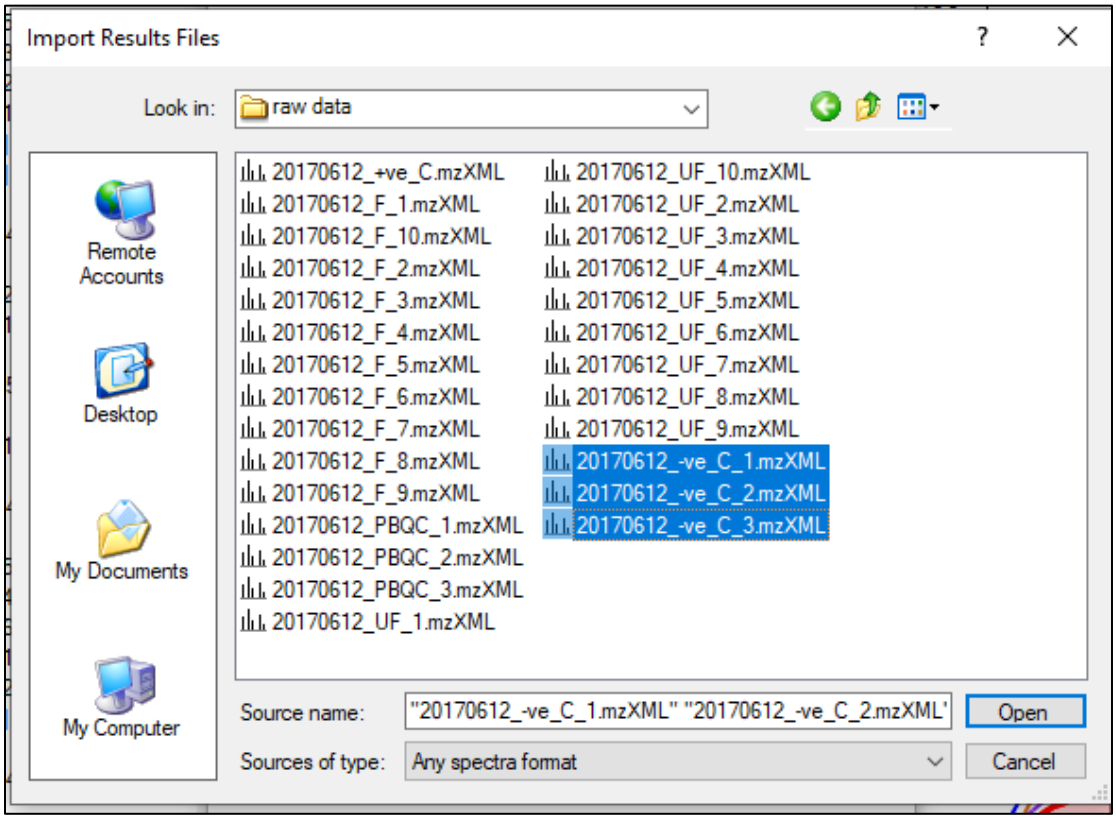
Removing Compounds Present in Negative Control Sample

Experimental data of the negative control sample (procedural blank) are used to monitor the background signal intensity during metabolomics data analysis and thus allow one to identify compounds whose abundance does not correlate with a biological phenomenon under study.

To check which compounds are present in negative control sample, perform the following steps:

- In **File** menu, select **Import** and click **Results**.
- In **Import Results** form, choose **Add single-injection replicates in files** option.
- Click **OK** button.
- Browse to a folder with tutorial data. Select 3 injection replicates of negative control sample by clicking the first -ve_C file and then holding down the **SHIFT** key and clicking the last -ve_C file.

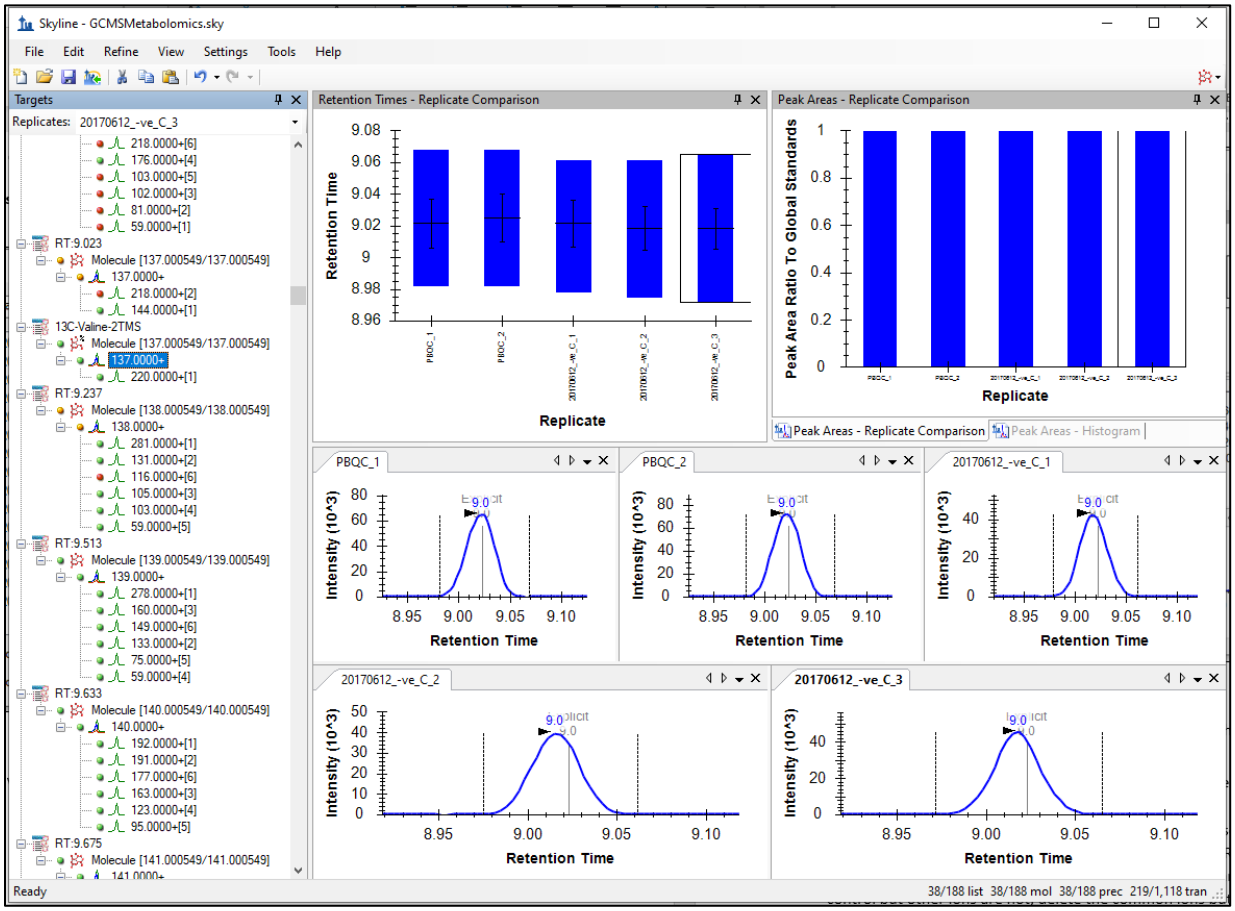
The **Import Results Files** form should look like this:



Now:

- Click **Open** button.
- Select **Remove** option, type “20170612_” in the **Common prefix** box and click **OK** button.

The files should import within 2 minutes or so, leaving your Skyline window look like this:



Note that the internal standard (13C-Valine) was spiked into negative control sample.

Click on each compound to identify metabolites that are also present in negative control sample:

- If all ions assigned to a particular compound are present in both PBQC and negative control, delete such compound (i.e. delete compounds at RT:21.203 and RT:21.285)
- If only some ions assigned to a particular compound are present in both PBQC and negative control but other ions assigned to that same compound are not, delete the common ions but keep the compound (i.e. in the case of compound at RT:20.527, delete *m/z* 204 and 131; in the case of compound at RT:20.682, delete *m/z* 218, 206 and 59)
- If none of the ions assigned to a particular compound are present in negative control, keep such compound (i.e. keep compounds at RT:20.810 and RT:20.865)

Note that there are many more compounds and ions that could be deleted but for this tutorial you will just focus on the ones mentioned above and let statistics to take care of everything assuming that column and septum bleed peaks, carry overs and others contaminants will not be statistically

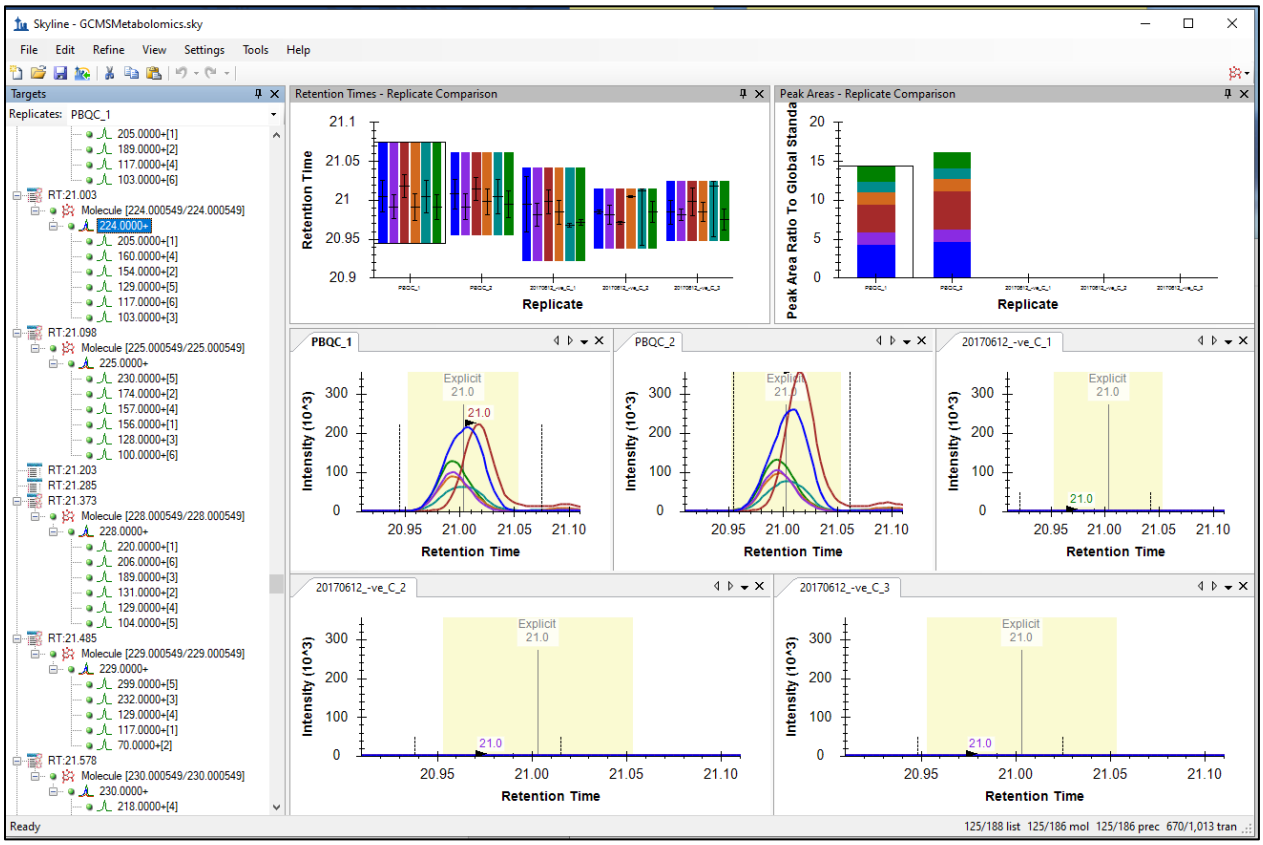
significantly different between experimental conditions because the samples were run in a random order.

Additional Refinement and Importing Remaining Data

As you worked with tutorial data, you most likely noticed that, occasionally, not all fragment ions assigned to a particular compound co-eluted perfectly and, most likely, they represented different compounds. Moreover, several fragment ions consistently had low signal-to-noise and did not seem specific to any compound and, in particular, these were 73 (TMS), 75 (septum bleed), 147 (TMS) and 207 (column bleed). In the next steps, you will perform limited deconvolution of closely eluting compounds and delete problematic ions to improve specificity and precision of quantitative analysis.

To deconvolute closely eluting compounds, perform the following steps:

- Scroll down to RT: 21.003 and click on its precursor ion 224.0000+
- Note that its fragment ion peaks can be assigned to 3 groups depending on their apex RT.



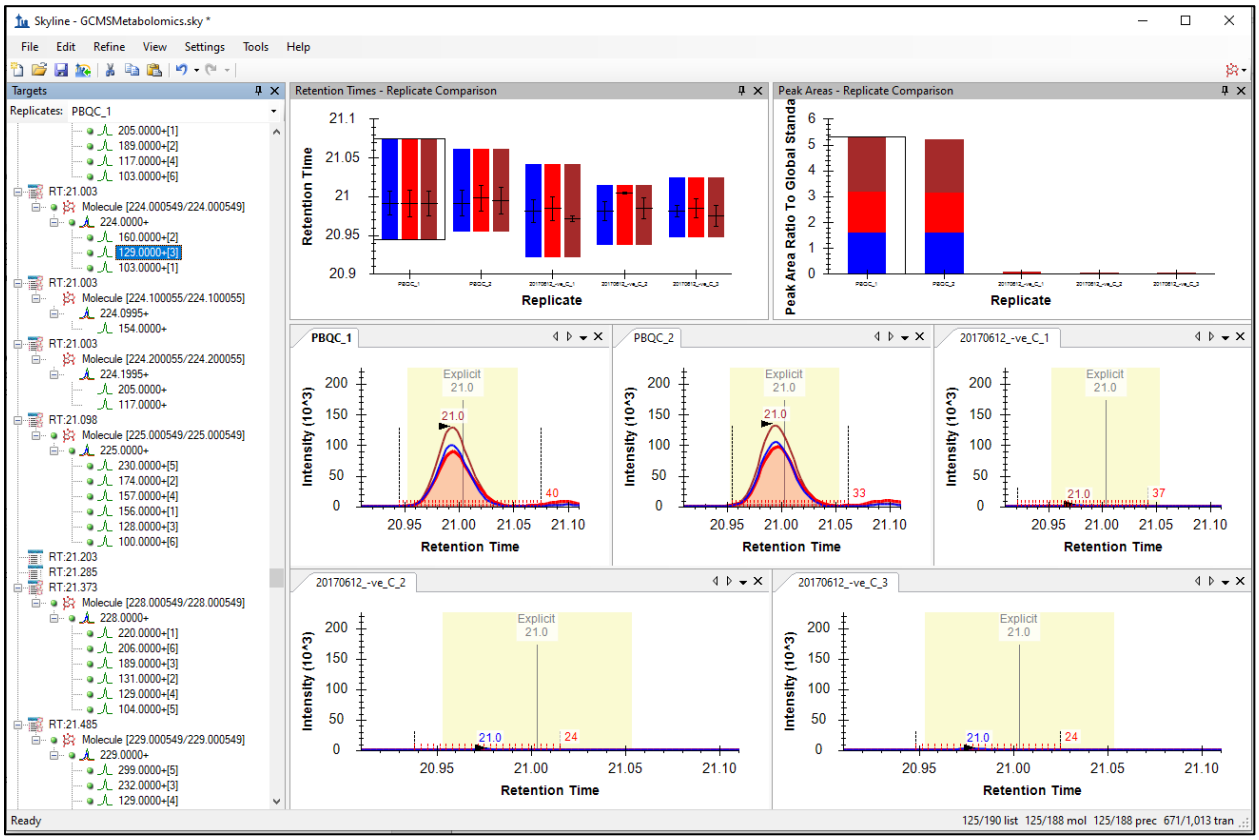
Now replicate steps conducted previously for Valine:

- Click on RT: 21.003
- In **Edit** menu, click **Copy** then **Paste**
- Again**,** in **Edit** menu, click **Paste**

You will see three identical instances of RT: 21.003. Next:

- In the case of the second instance of RT: 21.003, right click on **Molecule (224.000549/224/000549)** to open a pop-up menu and choose **Modify** to open **Modify Molecule** form. In this form, increase the **Monoisotopic mass** and **Average mass** by 0.1 and click **OK** button.
- In the case of the third instance of RT: 21.003, increase the **Monoisotopic mass** and **Average mass** by 0.2.
- In the case of the third instance of RT: 21.003, delete all but 117 and 205 ions by right clicking on them and choosing **Delete** in the pop-up menu.
- In the case of the second instance of RT: 21.003, delete all but 154 ion.
- In the case of the first instance of RT: 21.003, delete all but 160, 129 and 103 ions.

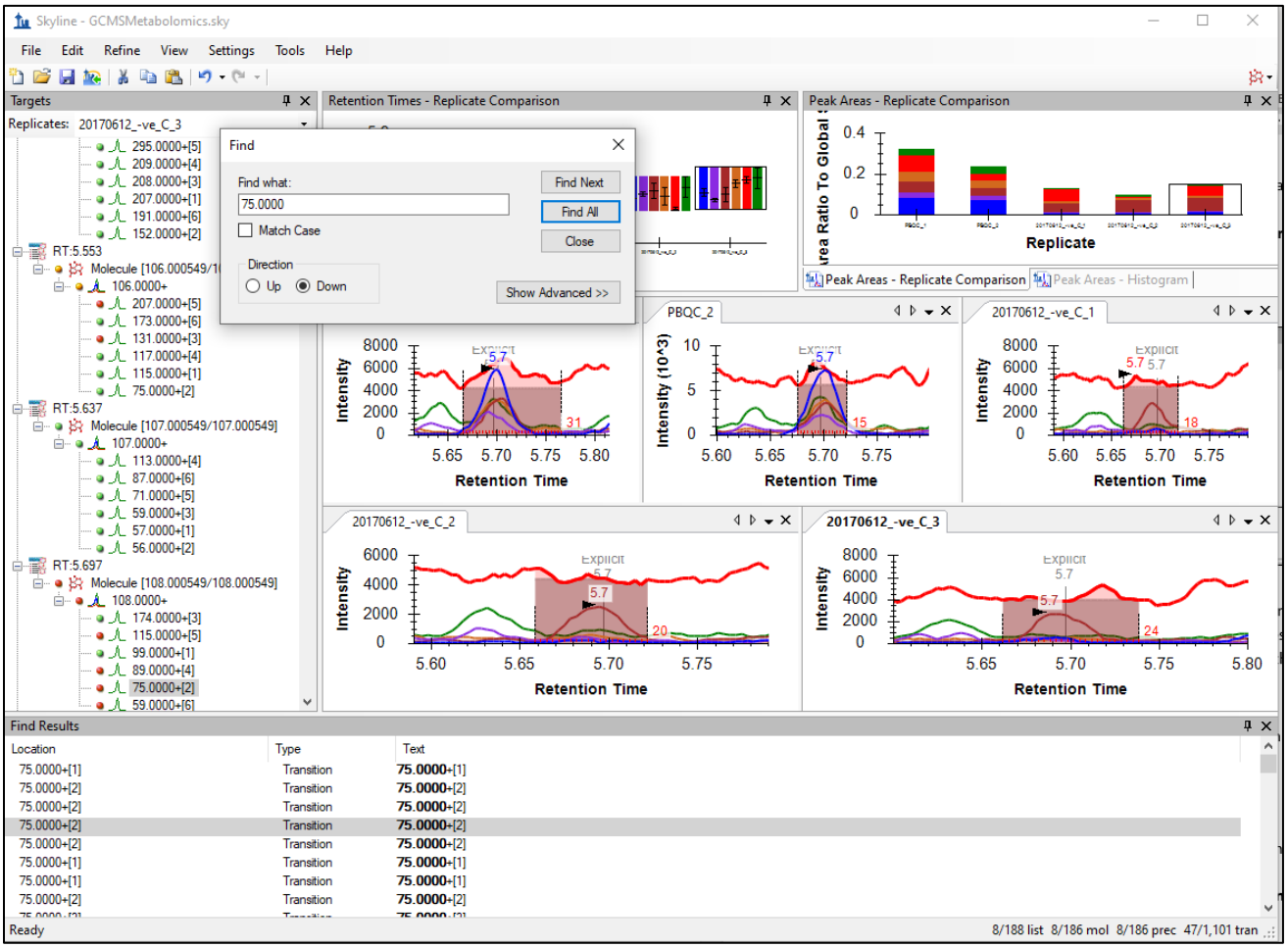
You should see them like this:



To delete non-specific fragment ions, perform the following steps:

- In **Edit** menu, click **Find**.
- Type '75.0000' in **Find what** field, and click **Find All**

The **Find Results** output should look like this:



Now, one-by-one, double click on each '75.0000' instance in the **Find Results** and press **DEL** key on the keyboard. Be careful not to delete any similar entries that are not of interest here such as '175.0000' or a precursor with that specific *m/z*. Do the same for 207 ion and close **Find Results** view by clicking 'x' sign.

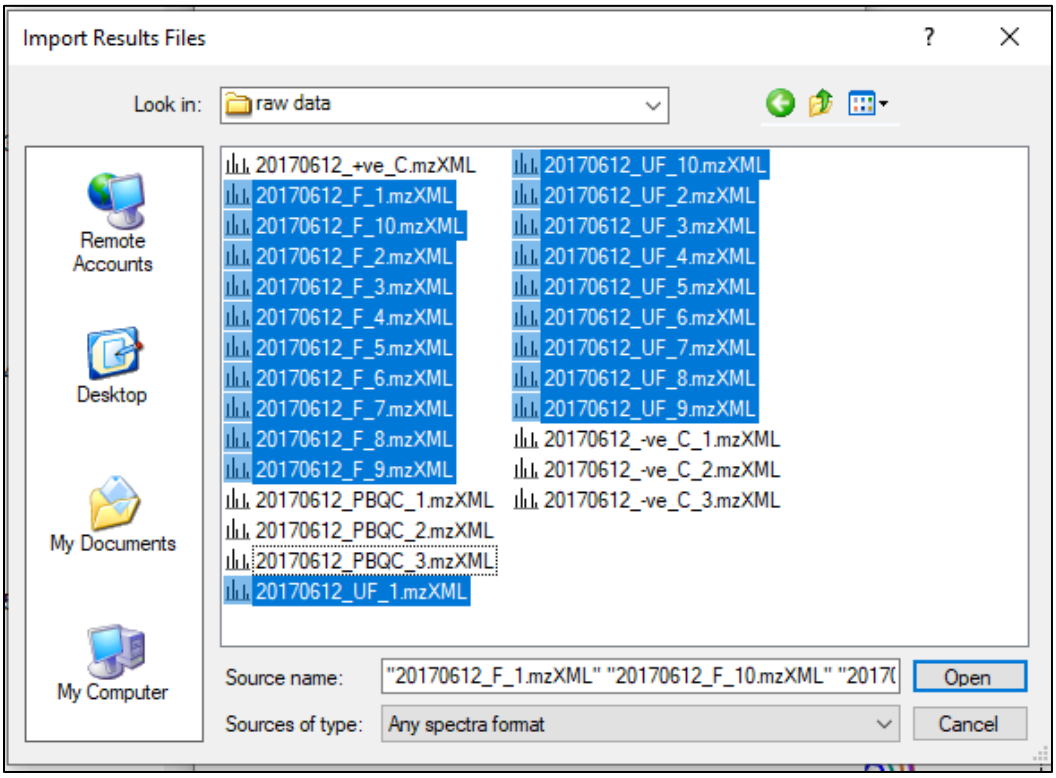
Note this process can be higher throughput if ions are deleted directly from *Shimadzu_TransitionList.csv* file before importing it into Skyline. In fact, TMS ions (73 and 147) were removed during creation of .csv file.

An additional refinement is possible however, in this tutorial you will leave data as are and let statistics take care of everything.

Once satisfied with how PBQC data look in Skyline, perform the following steps to import the remaining data files:

- In **File** menu, choose **Import** and click **Results**
- Click **OK** button.
- Browse to a folder with tutorial data. Select all remaining data files.

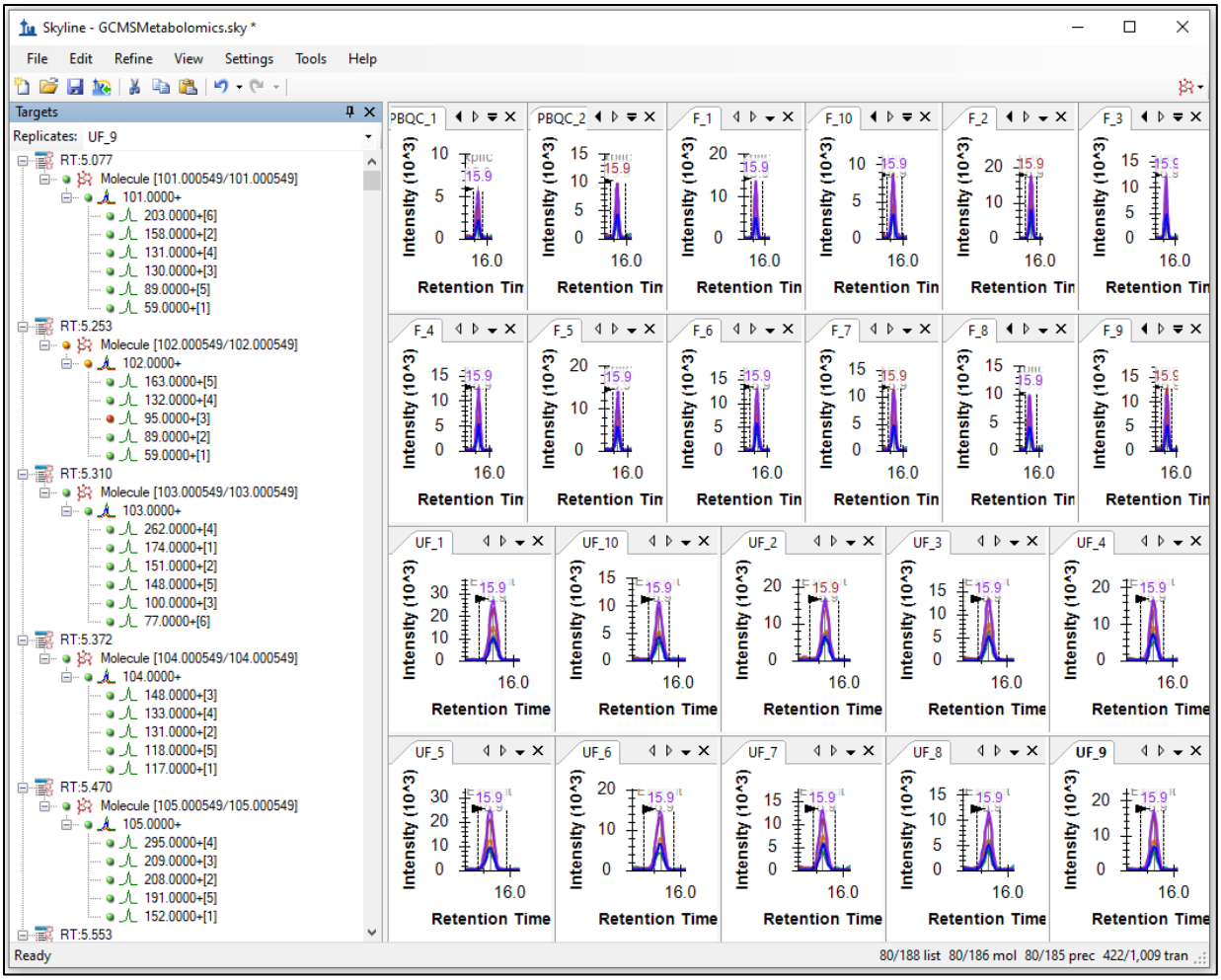
The **Import Results Files** form should look like this:



Now:

- Click **Open** button.
- Select **Remove** option, type “20170612_” in the **Common prefix** box and click **OK** button.

The files should import in just under 10 minutes, leaving your Skyline window look like this:



Preparing for Statistical Analysis with Replicate Annotation

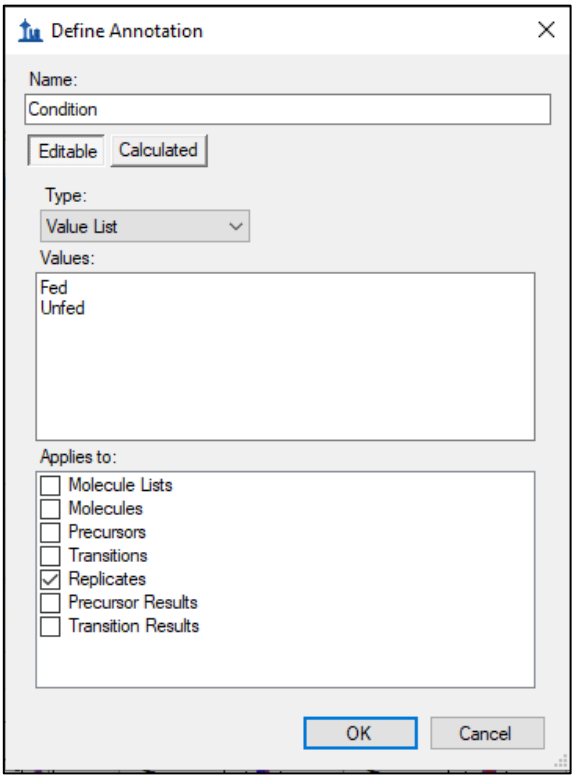
Once you have completed refining your Skyline document, you will want to identify what sort of difference in metabolite abundance there is between experimental conditions. Making such inquiries, whether in Skyline or in its external statistical tools (i.e. MSstats), requires some further classification of the measured samples (data files, injections), generally referred to as “replicates”. For such classification Skyline provides replicate annotations. In this tutorial, you will use one replicate annotation: Condition.

To define the ‘Condition’ annotation, perform the following steps:

- In **Settings** menu, click **Document Settings**.
- Click **Edit List** button in **Annotations** tab of **Document Settings** form.
- Click **Add** button in **Define Annotations** form.
- In **Name** field of **Define Annotation** form enter ‘Condition’.
- In **Type** drop down menu click ‘**Value List**’

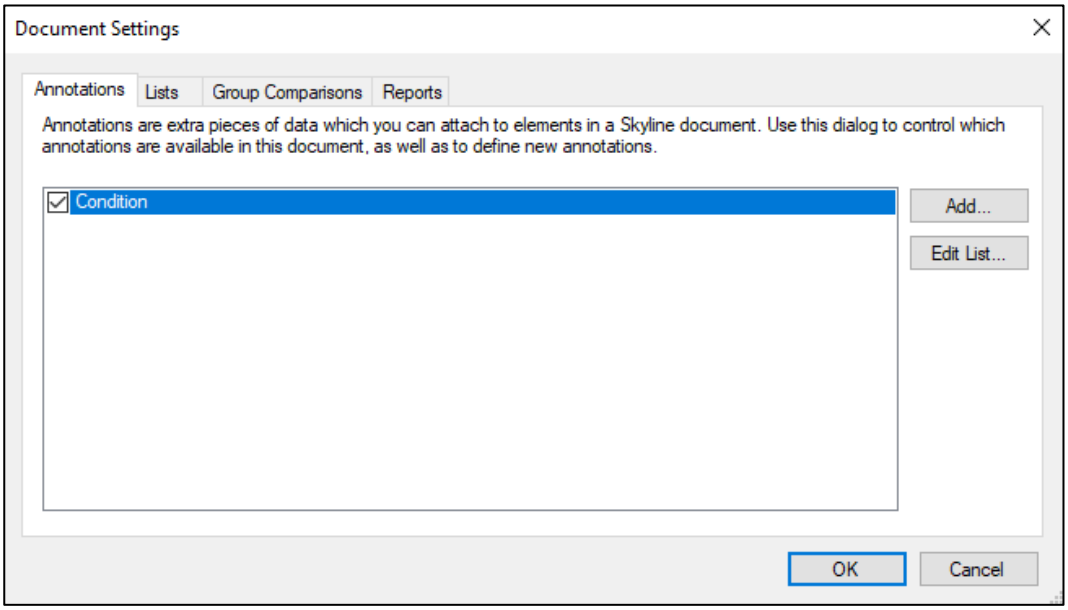
- In **Values** field type ‘Fed’, then press **ENTER** and type ‘Unfed’
- In **Applies to** list, check **Replicates**.

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



- Click **OK** button in **Define Annotation** form
- Again, click **OK** button in **Define Annotations** form
- Check ‘Condition’ annotation you just created.

The **Annotations** tab in **Document Settings** form should look like below:

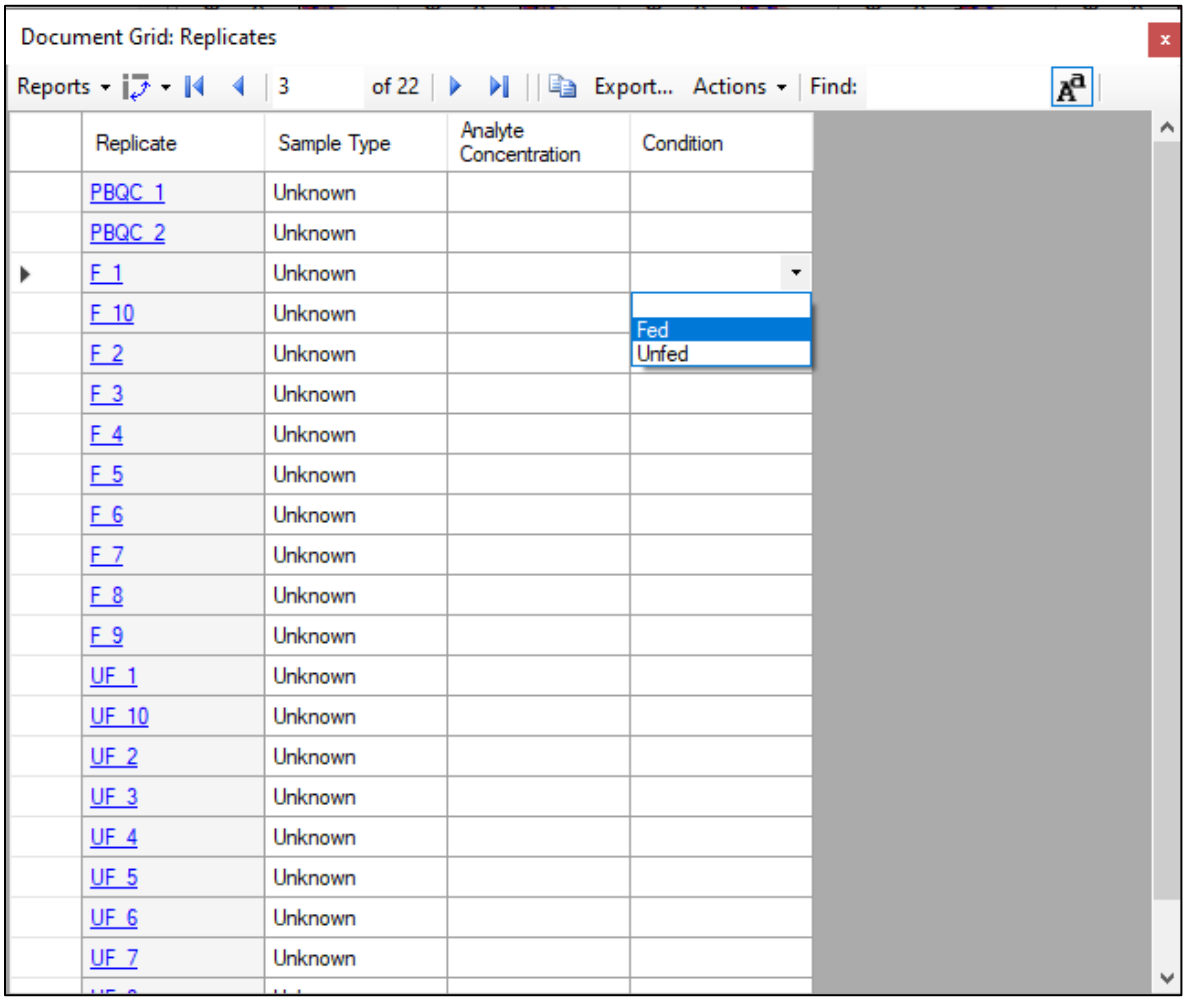


- Click **OK** button in **Document Settings** form

To apply 'Condition' annotation to the measured samples (replicates), perform the following steps:

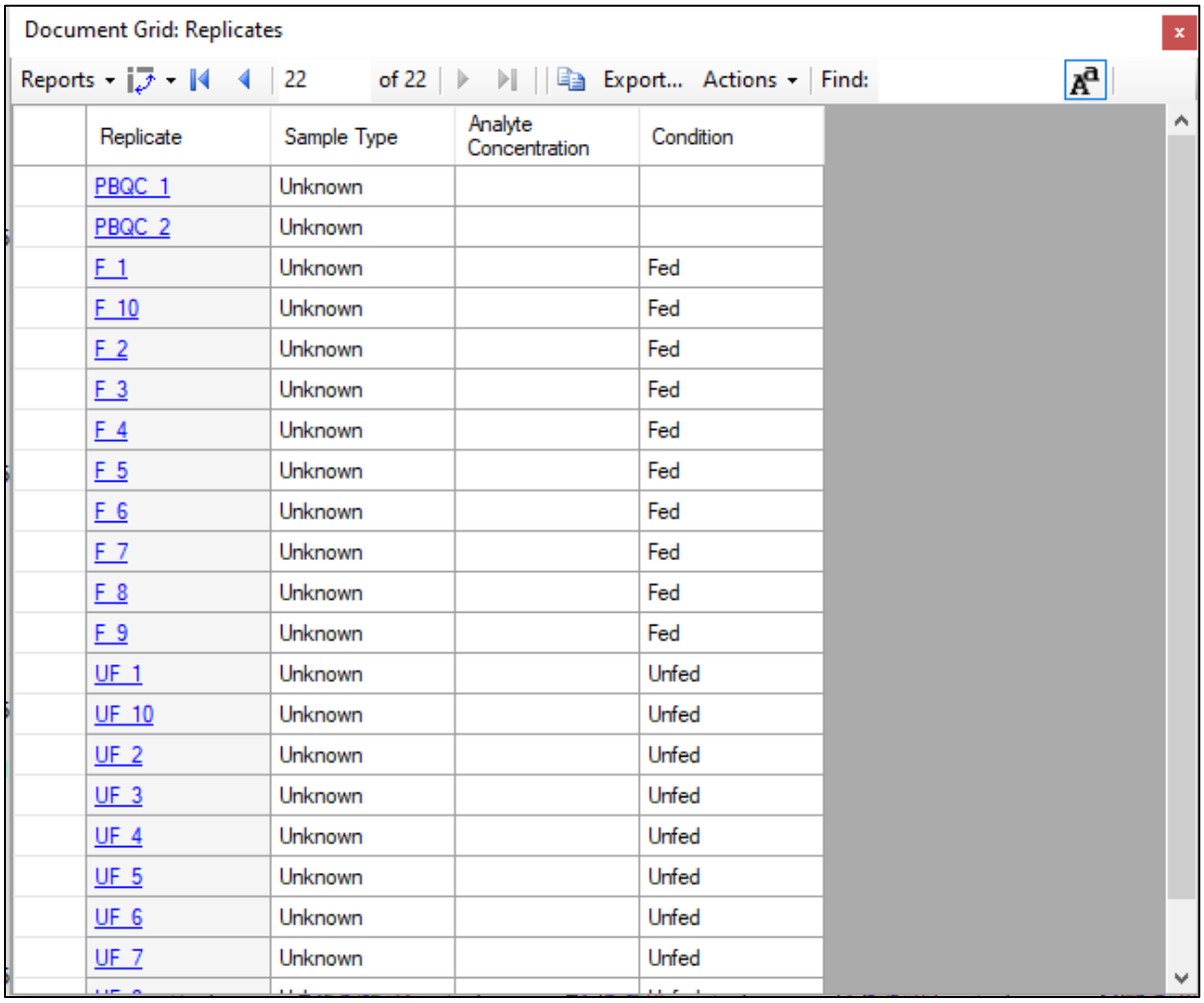
- In **View** menu, click **Document Grid**.
- In **Reports** menu that can be found in the upper left corner of the **Document Grid**, click **Replicates**.

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



You can now manually enter the annotations for all replicates into this grid. All replicates named 'F' should be annotated as 'Fed', and all replicates named 'UF' should be annotated as 'Unfed'.

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


Document Grid: Replicates

Reports | 22 of 22 | Export... Actions | Find:

Replicate	Sample Type	Analyte Concentration	Condition
PBQC_1	Unknown		
PBQC_2	Unknown		
F_1	Unknown		Fed
F_10	Unknown		Fed
F_2	Unknown		Fed
F_3	Unknown		Fed
F_4	Unknown		Fed
F_5	Unknown		Fed
F_6	Unknown		Fed
F_7	Unknown		Fed
F_8	Unknown		Fed
F_9	Unknown		Fed
UF_1	Unknown		Unfed
UF_10	Unknown		Unfed
UF_2	Unknown		Unfed
UF_3	Unknown		Unfed
UF_4	Unknown		Unfed
UF_5	Unknown		Unfed
UF_6	Unknown		Unfed
UF_7	Unknown		Unfed

- Click the **X** button in the top right corner of **Document Grid** to close it.

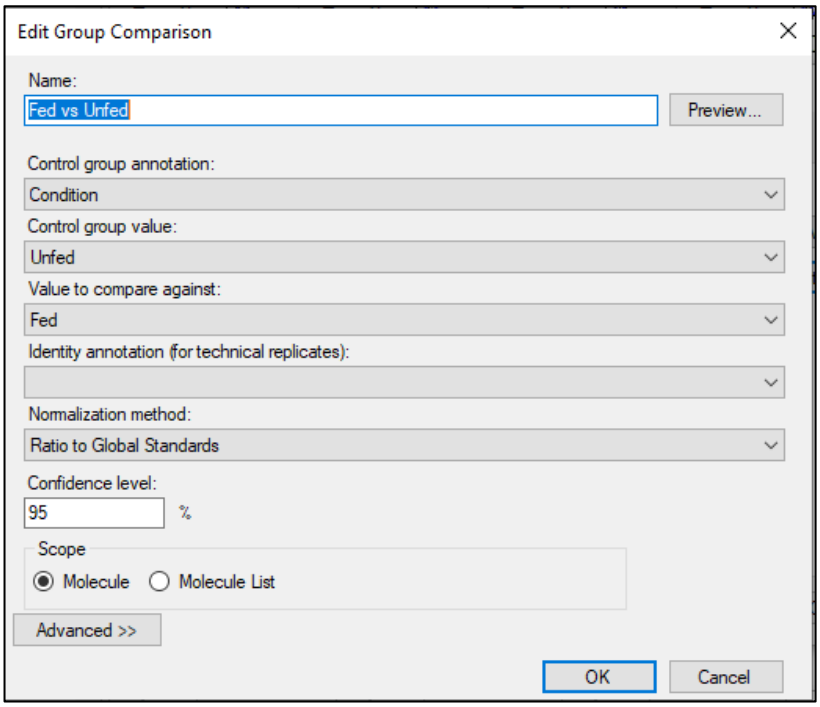
Group Comparison and Detection of Differentially Regulated Metabolites

Once you have completed annotating replicates, perform simple pairwise group comparison of metabolite peak areas. The comparison is performed by summing the available transition peak areas for a metabolite, optionally dividing by a normalization standard (13C-Valine), taking the log, and performing a t-test on the resulting values. Skyline automatically discards replicates with missing values, or truncated peaks.

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

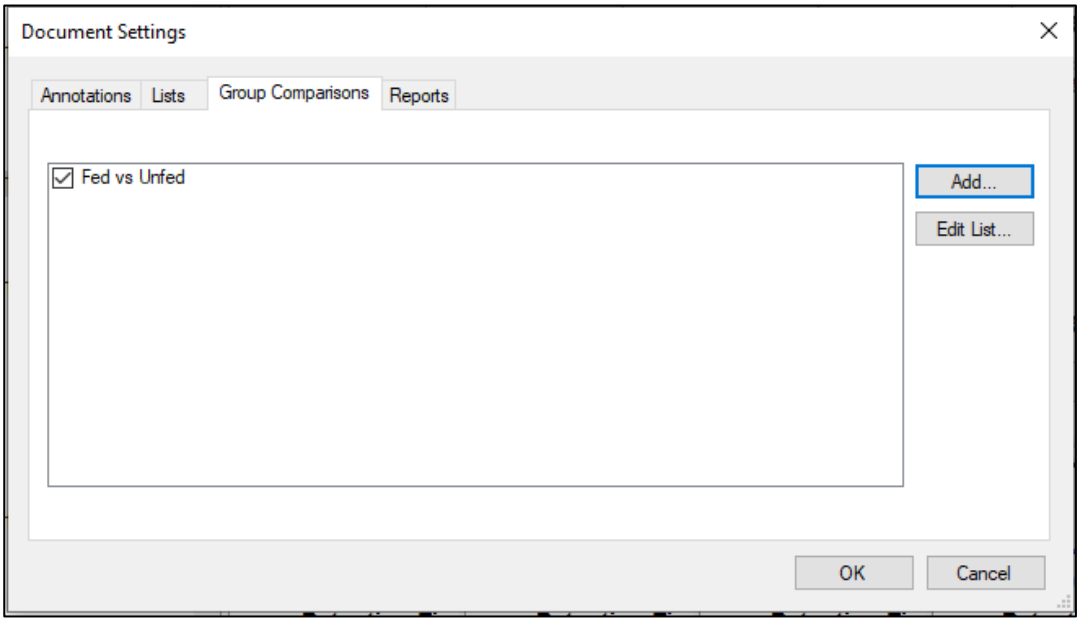
- In **Settings** menu, click **Document Settings**.
- Click **Group Comparisons** tab and then **Add** button.
- In **Name** field of **Edit Group Comparison** form, enter “Fed vs Unfed”.
- For **Control group annotation** choose “Condition”.
- For **Control group value** choose “Unfed”.
- For **Value to compare against** choose “Fed”.
- Leave **Identity annotation** blank because the experiment only had biological replicates and no technical replicates.
- For **Normalization method** choose “Ratio to Global Standards”.
- In **Confidence level** field, enter “95” %.
- For **Scope** choose **Molecule**.

The **Edit Group Comparisons** form should look like this:



- Click **OK** button in **Edit Group Comparison** form
- Click **OK** button in **Define Group Comparison** form
- Check ‘Fed vs Unfed’ group comparison you just created.

The **Group Comparisons** tab in **Document Settings** form should now look like this:

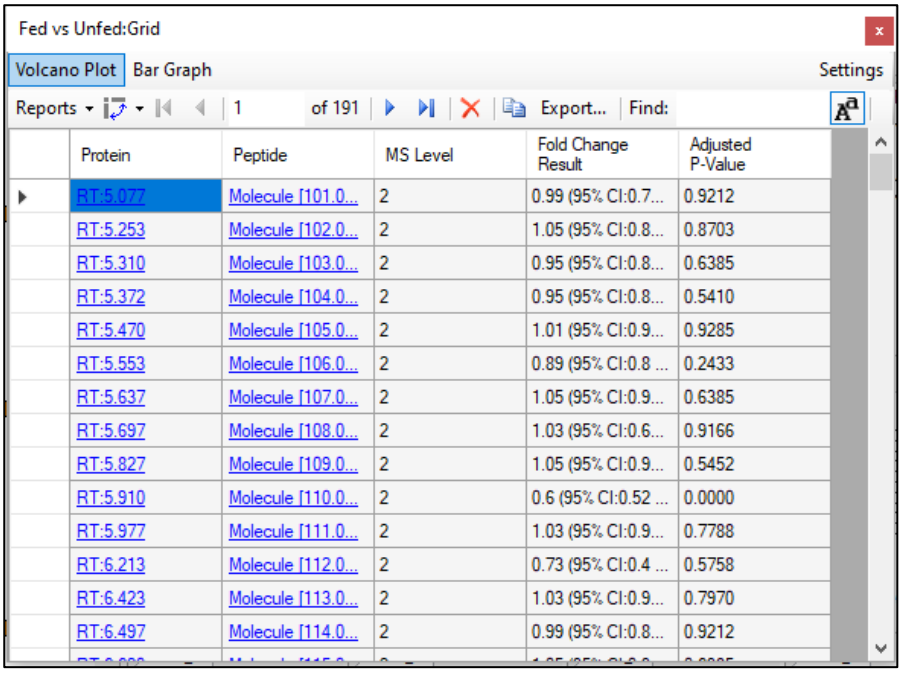


- Click **OK** button in **Document Settings** form

To see the output of group comparison, do the following:

- In **View** menu, choose **Other Grids**, then **Group Comparisons** and there click **Fed vs Unfed**.

Skyline will show a Grid that looks like this:



Protein	Peptide	MS Level	Fold Change Result	Adjusted P-Value
RT:5.077	Molecule [101.0...	2	0.99 (95% CI:0.7...	0.9212
RT:5.253	Molecule [102.0...	2	1.05 (95% CI:0.8...	0.8703
RT:5.310	Molecule [103.0...	2	0.95 (95% CI:0.8...	0.6385
RT:5.372	Molecule [104.0...	2	0.95 (95% CI:0.8...	0.5410
RT:5.470	Molecule [105.0...	2	1.01 (95% CI:0.9...	0.9285
RT:5.553	Molecule [106.0...	2	0.89 (95% CI:0.8 ...	0.2433
RT:5.637	Molecule [107.0...	2	1.05 (95% CI:0.9...	0.6385
RT:5.697	Molecule [108.0...	2	1.03 (95% CI:0.6...	0.9166
RT:5.827	Molecule [109.0...	2	1.05 (95% CI:0.9...	0.5452
RT:5.910	Molecule [110.0...	2	0.6 (95% CI:0.52 ...	0.0000
RT:5.977	Molecule [111.0...	2	1.03 (95% CI:0.9...	0.7788
RT:6.213	Molecule [112.0...	2	0.73 (95% CI:0.4 ...	0.5758
RT:6.423	Molecule [113.0...	2	1.03 (95% CI:0.9...	0.7970
RT:6.497	Molecule [114.0...	2	0.99 (95% CI:0.8...	0.9212

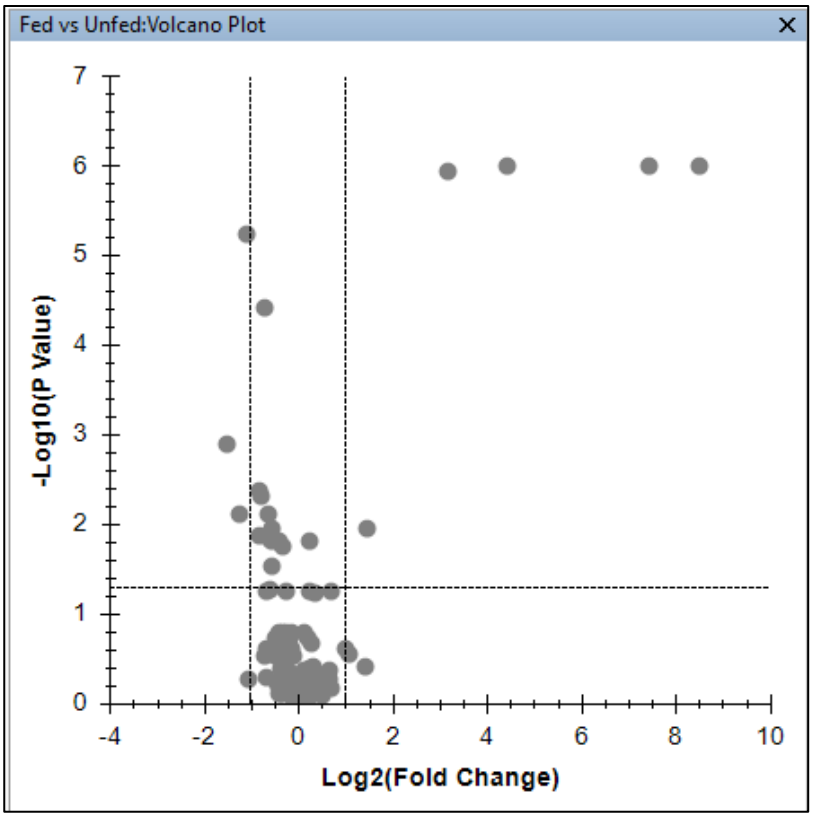
If **Fold Change Result** column does not show full confidence interval (CI):

- Double-click the vertical line between the **Fold Change Result** and **Adjusted P-Value** headers.

To see a Volcano Plot:

- Click **Volcano Plot** button in menu above the Grid toolbar.

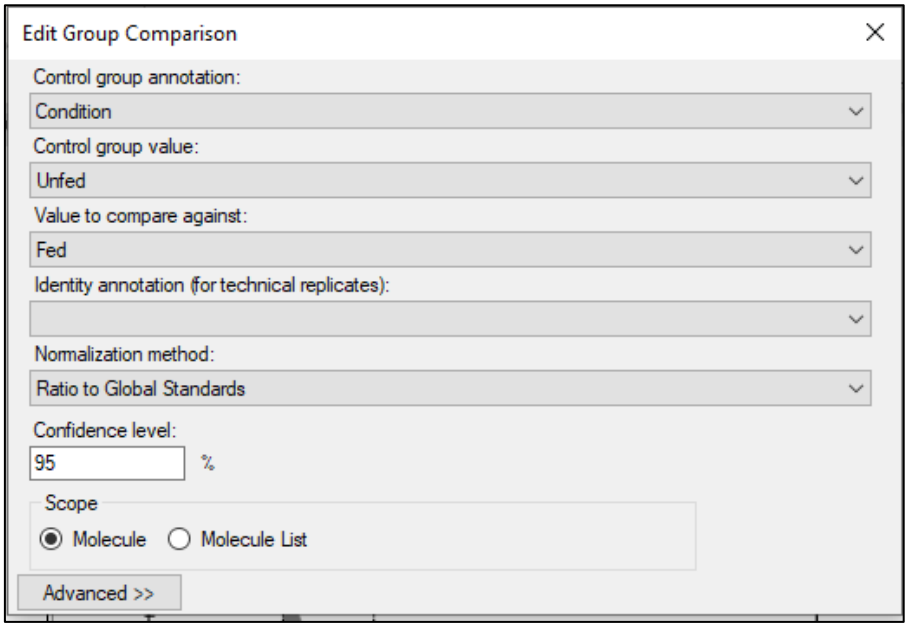
Skyline adds a **Volcano Plot** tab to the "Fed vs Unfed" grid view that looks like this:



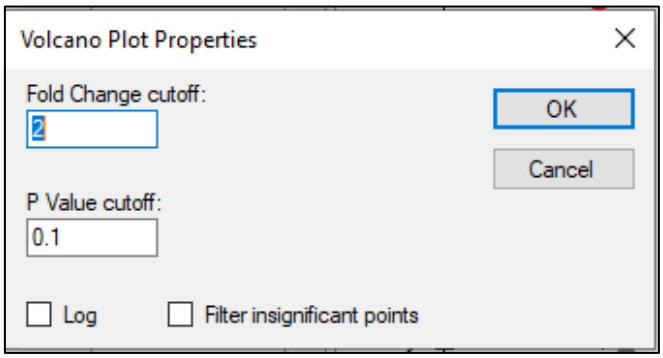
The X axis of the plot represents a biological significance whilst Y axis represents a statistical significance. Vertical and horizontal lines are user-defined cutoff lines. Each grey dot represents a metabolite and, at the same time, an independent statistical test (that is an independent hypothesis being tested).

Upon right click on the Volcano Plot, you can change **Settings**, **Properties** and **Formatting**.

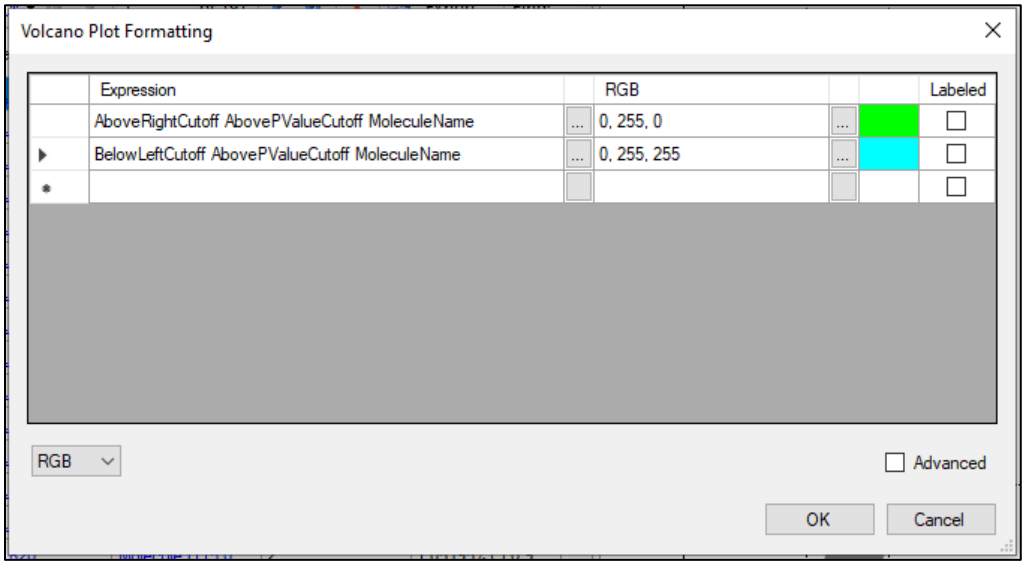
- Choosing **Settings** will allow you to re-define how the group comparison is performed as shown below:

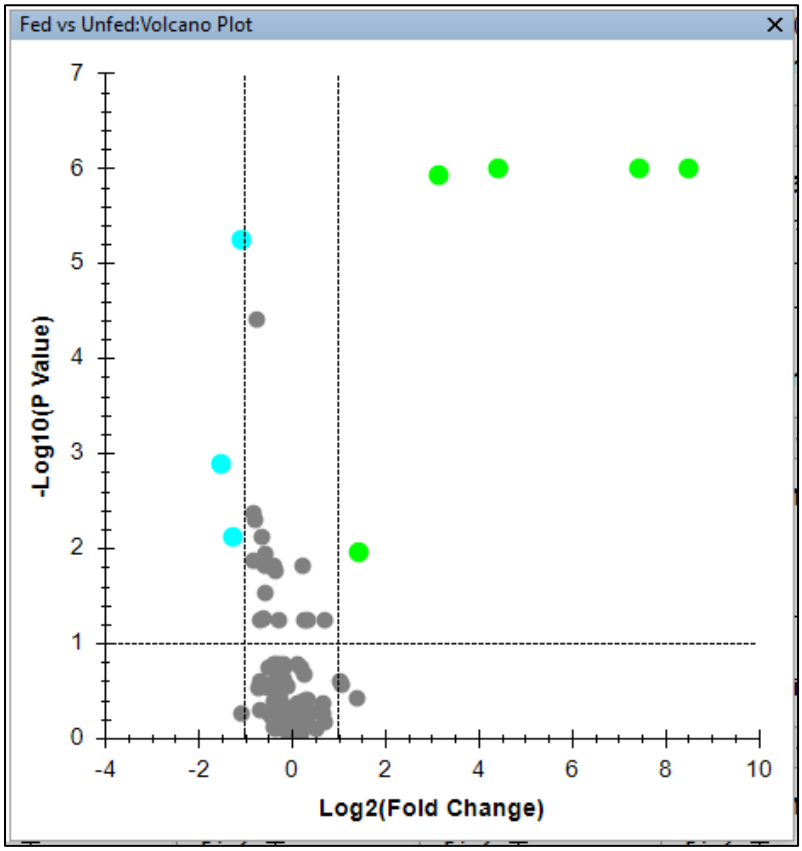


- Choosing **Properties** will allow you to re-define the cutoff lines (**Fold Change cutoff** – vertical lines; **P Value cutoff** – horizontal line). In the example below, a two-fold cutoff on Fold Change and an FDR adjusted P Value (also called Q Value) set to 0.1 were used. The later implies that 10% of significant statistical tests will result in false positives (that is Benjamini-Hochberg correction for multiple hypothesis testing).

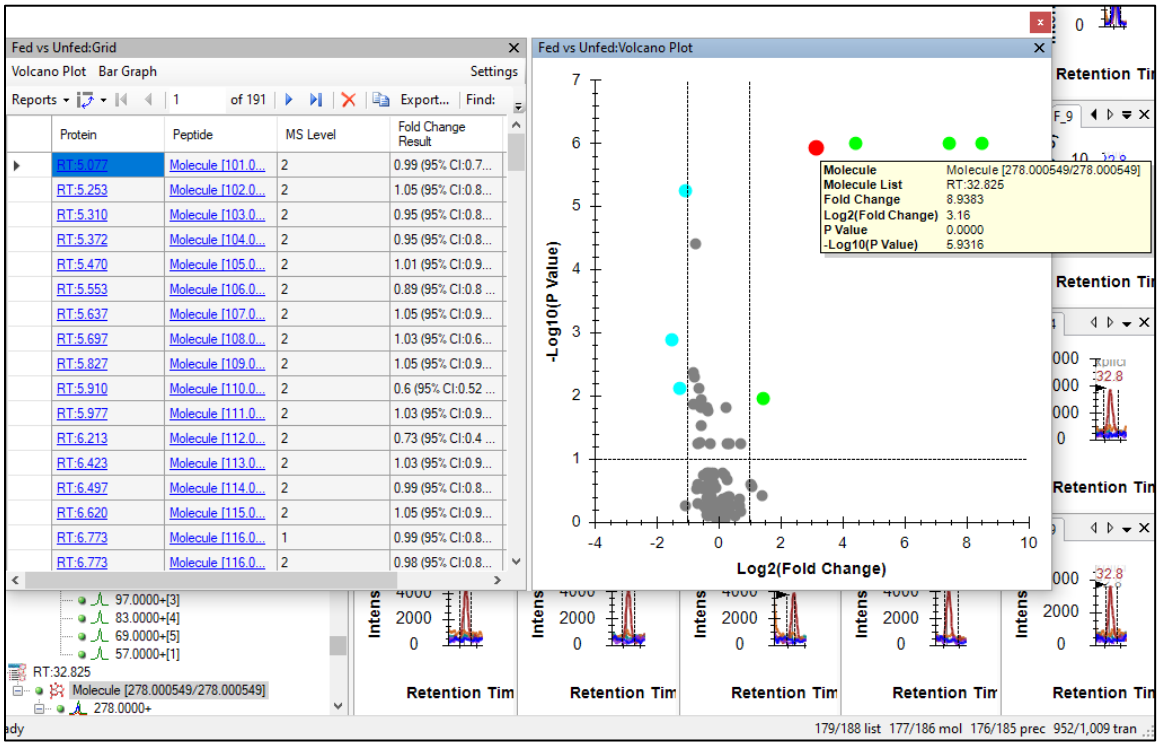


- Choosing **Formatting** will allow you to annotate data points on a plot using different colours. In the example below, all significantly downregulated metabolites in “Fed” samples were shown in blue whilst all significantly upregulated metabolites in “Fed” samples were shown in green.





Importantly, each dot on a Volcano Plot is clickable. Clicking on one of the dots in the top right quadrant that represents metabolites whose increase in abundance in “Fed” samples was also the most statistically significant, highlights it in red and reveals a RT at which this metabolite eluted (see below). With this information, it is now possible to open one of the “Fed” samples in vendor-specific software and conduct identification of a metabolite of interest.



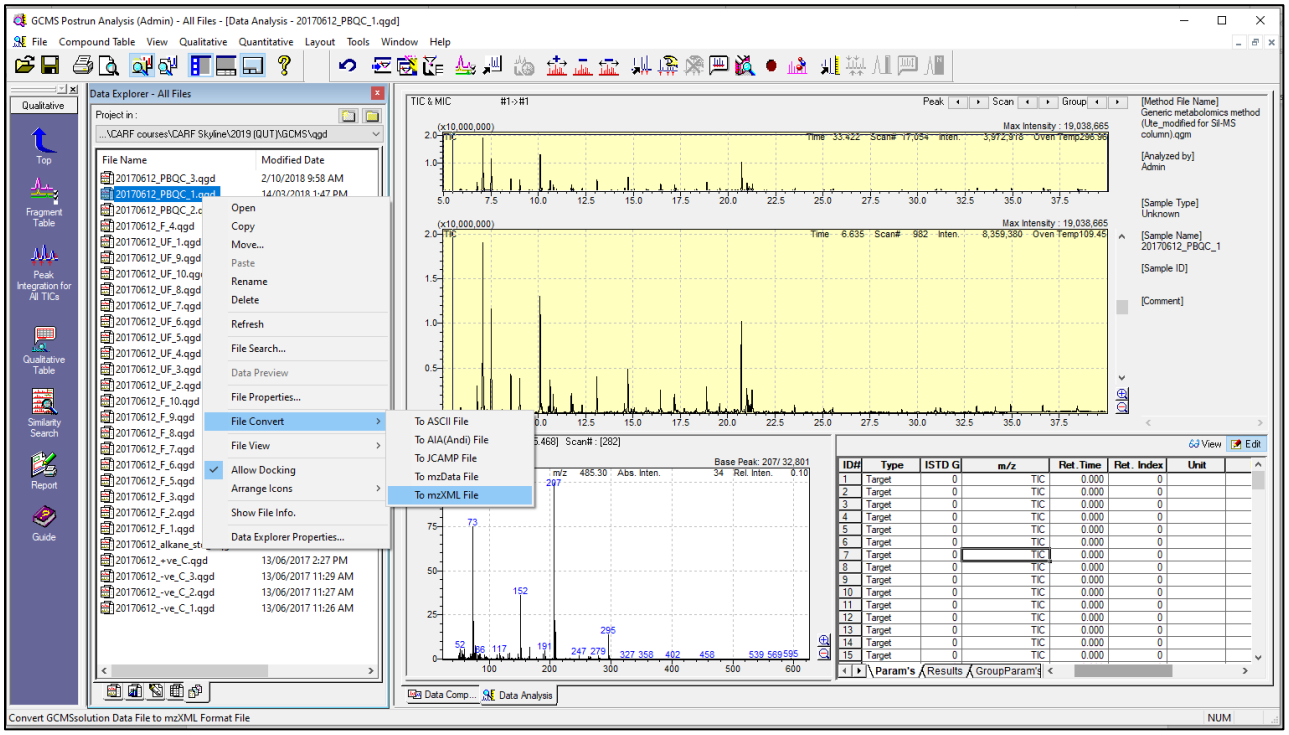
Conclusions

In this tutorial, you have learned how to use Skyline to process metabolomics data collected on GCMS platform. In this case, data were acquired on Shimadzu instrument and converted into .mzXML format however, Skyline also supports native formats from Thermo and Agilent GCMS instruments. GCMS-based metabolomics data remain one of the most challenging to analyse. Skyline simplifies this process by providing intuitive graphical interface for targeted extraction of untargeted data using a workflow that initially does not require metabolite identification nor library within Skyline. The workflow is analogical to that for proteomics and thus Skyline supports the standardization of processing of different MS-based omics datasets.

Appendix

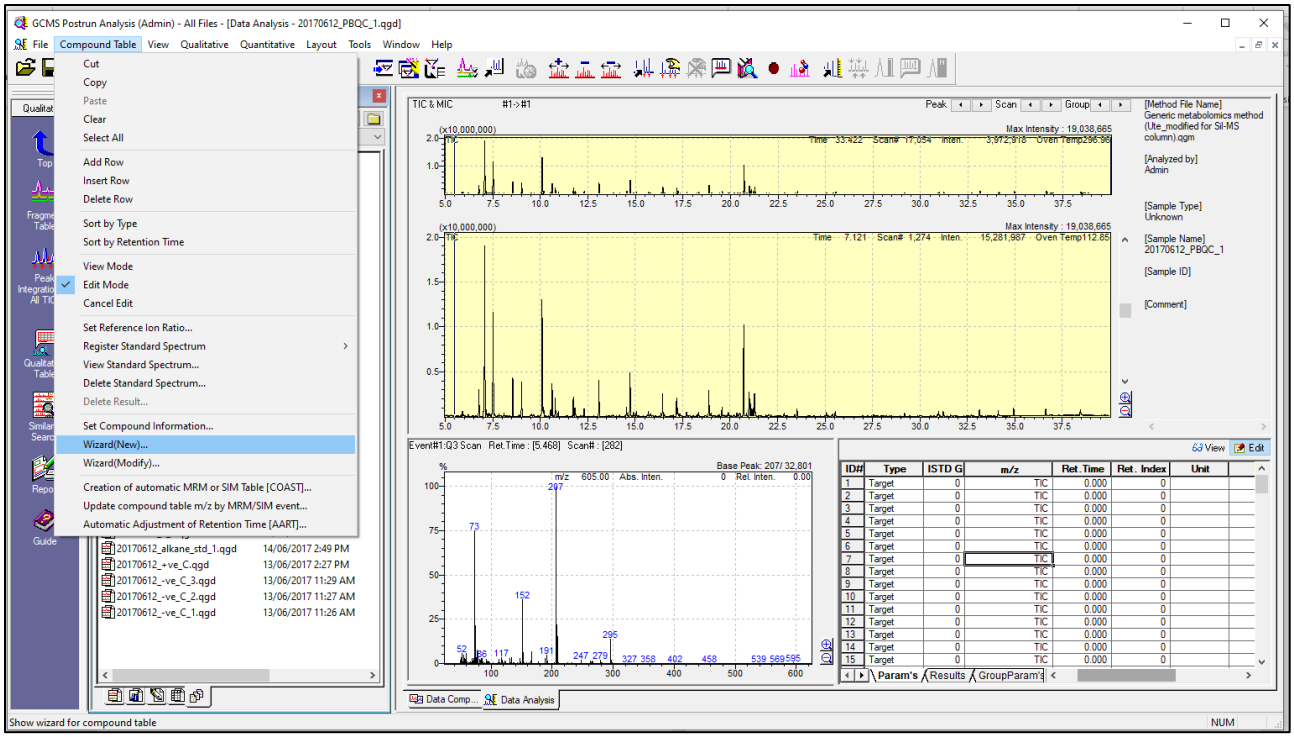
Converting data to .mzXML format

In Shimadzu GCMS Postrun Analysis software, right click on a data file you want to convert, then choose **File Convert** and then click **To mzXML File** like below:



Creating a Small Molecule Transition List

In Shimadzu GCMS Postrun Analysis software, select **Compound Table** menu and then click **Wizard(New)** like below:



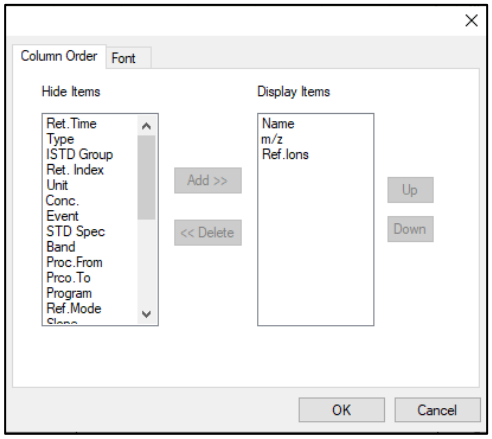
In **Compound Table Wizard:**

- In step 1/7 select **Integration of TIC** and use **Integration Parameters** button to adjust settings as necessary
- In step 5/7 select **MC** as **Target Ion,** and type “5” in the **# of Reference Ions** field.

Once **Compound Table** has been created it will be displayed in the bottom right corner of the software.

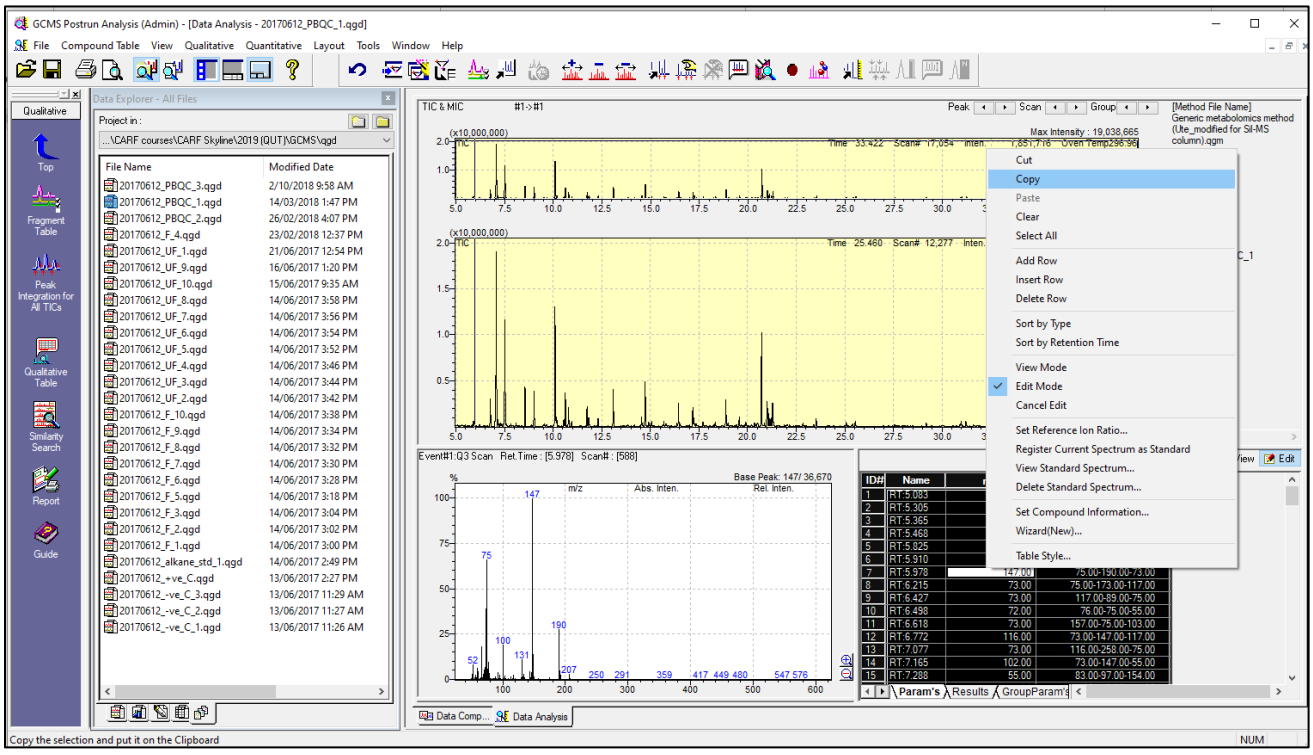
Now:

- Right click on **Compound Table** and select **Table Style**.
- Adjust **Display Items** to match the below and click **OK** button.



Next:

- Highlight the entire content of **Compound Table**
- Right click on **Compound Table** and select **Copy** like below.



Once copied to clipboard, paste it into any text editor and save as .txt file. Next, open the above .txt file in Excel and reformat as shown below. Note that at this stage you can delete ions that you know can cause problem during quantitative analysis such as 73 and 147.

	Molecule name	..... Product ions						Retention time
	A	B	C	D	E	F	G	H
1	RT:5.078	73	158	59	89	130	203	5.078
2	RT:5.305	73	147	57	77	174	151	5.305
3	RT:5.370	73	117	147	131	75	102	5.37
4	RT:5.470	207	73	152	208	209	295	5.47
5	RT:5.828	207	73	57	72	55	69	5.828
6	RT:5.912	73	117	147	191	148	75	5.912
7	RT:5.983	147	75	73	100	190	148	5.983
8	RT:6.218	73	75	173	147	117	131	6.218
9	RT:6.425	73	117	89	75	118	74	6.425
10	RT:6.500	72	76	75	55	59	74	6.5
11	RT:6.618	73	157	75	103	69	207	6.618
12	RT:6.773	116	73	147	117	74	75	6.773
13	RT:7.075	73	116	258	75	147	89	7.075
14	RT:7.165	102	73	147	55	97	83	7.165
15	RT:7.288	55	83	97	154	82	124	7.288
16	RT:7.350	73	147	130	174	188	75	7.35
17	RT:7.520	147	73	133	59	148	86	7.52
18	RT:7.832	174	73	86	142	59	175	7.832
19	RT:7.920	147	73	148	57	117	75	7.92
20	RT:8.092	147	281	73	282	148	283	8.092

		Precursor m/z	Precursor charge	Product m/z	Product charge	Retention time	Retention time window
	A	B	C	D	E	F	G
1	RT:5.078	100	1	X 73	1	5.078	0.1
2	RT:5.078	100	1	158	1	5.078	0.1
3	RT:5.078	100	1	59	1	5.078	0.1
4	RT:5.078	100	1	89	1	5.078	0.1
5	RT:5.078	100	1	130	1	5.078	0.1
6	RT:5.078	100	1	203	1	5.078	0.1
7							
8	RT:5.305	101	1	X 73	1	5.305	0.1
9	RT:5.305	101	1	X 147	1	5.305	0.1
10	RT:5.305	101	1	57	1	5.305	0.1
11	RT:5.305	101	1	77	1	5.305	0.1
12	RT:5.305	101	1	174	1	5.305	0.1
13	RT:5.305	101	1	151	1	5.305	0.1
14							
15	RT:5.370	102	1	X 73	1	5.37	0.1
16	RT:5.370	102	1	117	1	5.37	0.1
17	RT:5.370	102	1	X 147	1	5.37	0.1
18	RT:5.370	102	1	131	1	5.37	0.1

Alternatively, use AMDIS (free) software which allows one to generate a list of up to 2-3 fragment ions for each chromatographic peak and is compatible with Shimadzu, Thermo and Agilent native data file formats. To achieve this, process data file in AMDIS and then select **Generate Report** in **File** menu.

Importing Agilent Data

Importing Agilent data to Skyline requires specific formatting of a transition list. Refer to Skyline on-line support board for details: https://skyline.ms/announcements/home/support/thread.view?rowId=43600