Y3K GC Quantitation Pipeline User Guide

Introduction

The Y3K GC Quantitation Pipeline is a set of three software tools (Deconvolution Engine, Deconvolution Studio, and GC Quant) designed to extract metabolite abundances from Q-Exactive GC (Thermo Fisher Scientific, Austin TX) raw data files. Visual Studio solution files containing all source for these three tools are made available here, along with compiled executables. Instructions for using each tool (shown using a small set of data files) as well as a description of software function are included here. If you have any issues using these tools or additional feedback please contact us at y3kcontact@gmail.com.

Deconvolution Engine

Overview: Deconvolution Engine extracts chromatographic features (“peaks”) from Q-Exactive GC (Thermo Fisher Scientific, Austin TX) raw data files (*.raw) and writes associated data to a file (*.gcfeat). These *.gcfeat files are required for quantitation and are used in both the Deconvolution Studio and GC Quant tools.

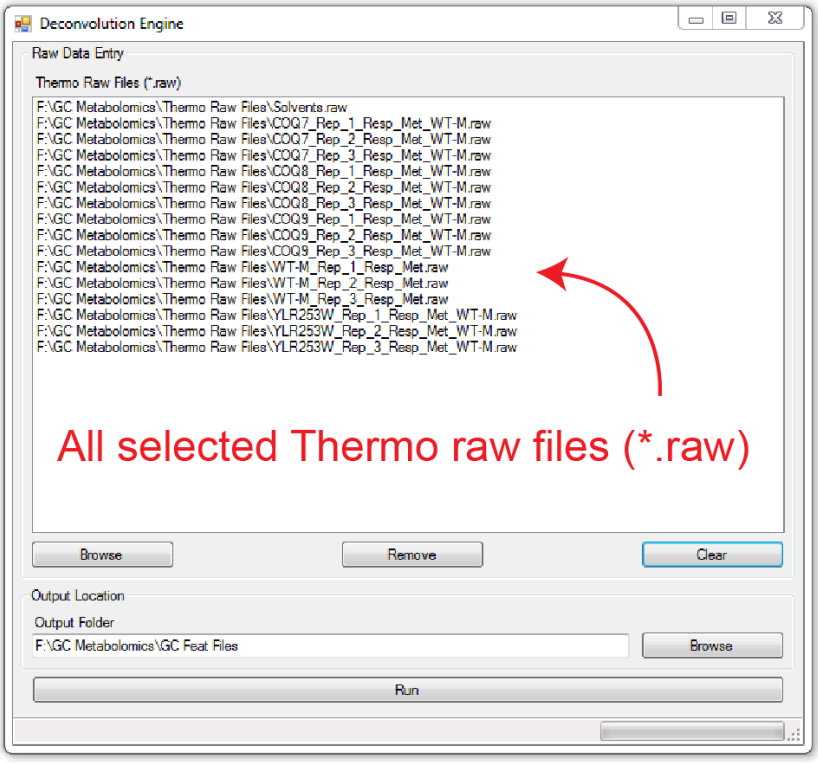
Inputs: Q-Exactive GC *.raw files, user-selected output directory.

Instructions:

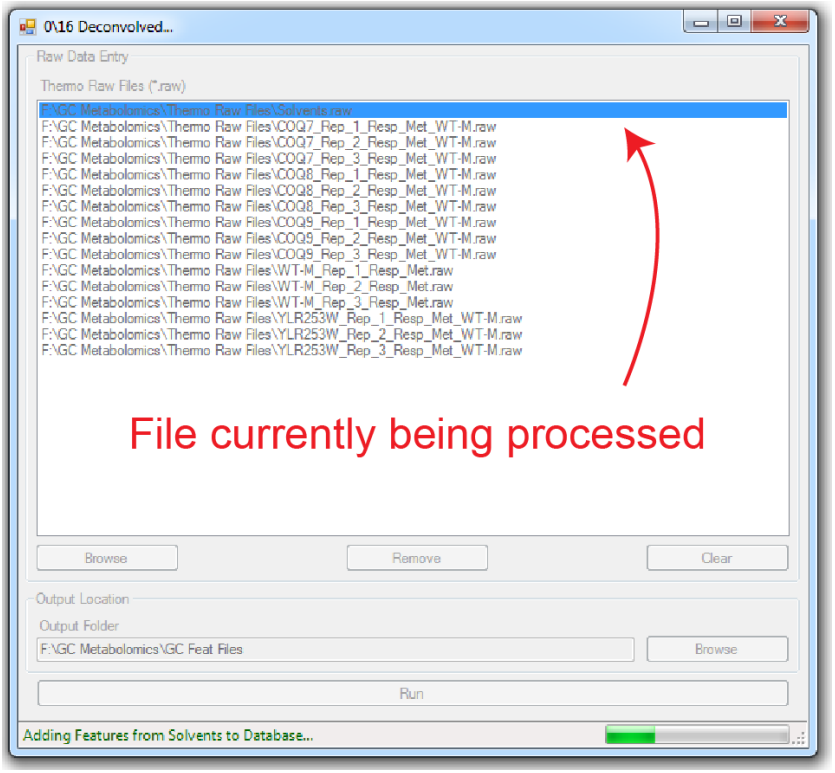
1.) Launch the Deconvolution Engine executable ‘Y3K-Deconvolution Engine.exe’.
2.) Drag and drop Q-Exactive GC files into the ‘Raw Data Entry’ list box. Alternatively, raw files can be added by clicking the ‘Browse’ button and selecting the appropriate files. All files can be emptied by clicking the ‘Clear’ button. Individual files can be removed by first highlighting said files, and then clicking ‘Remove.’
3.) Select an output directory where extracted feature files (*.gcfeat) will be written to by clicking ‘Browse’ in the ‘Output Location’ window and navigating to the appropriate folder.
4.) Click ‘Run’.

Expected Performance:

After drag and drop all Thermo raw files (*.raw) will be listed in the ‘Raw Data Entry’ list box.



During runtime the GUI and status bar should update to indicate the file currently being deconvolved. The form title should update throughout execution to indicate how many files have been successfully deconvolved.



Deconvolution Studio

Inputs: MS control (*.gcfeat) and MS blank (*.gcfeat) for creation of GC Master Files (*.gcmast). MS control master file (*.gcmast) and MS blank master file (*.gcmast) for curation of target list inside control master. Thermo raw file (*.raw) from MS control run for quant ion selection (optional).

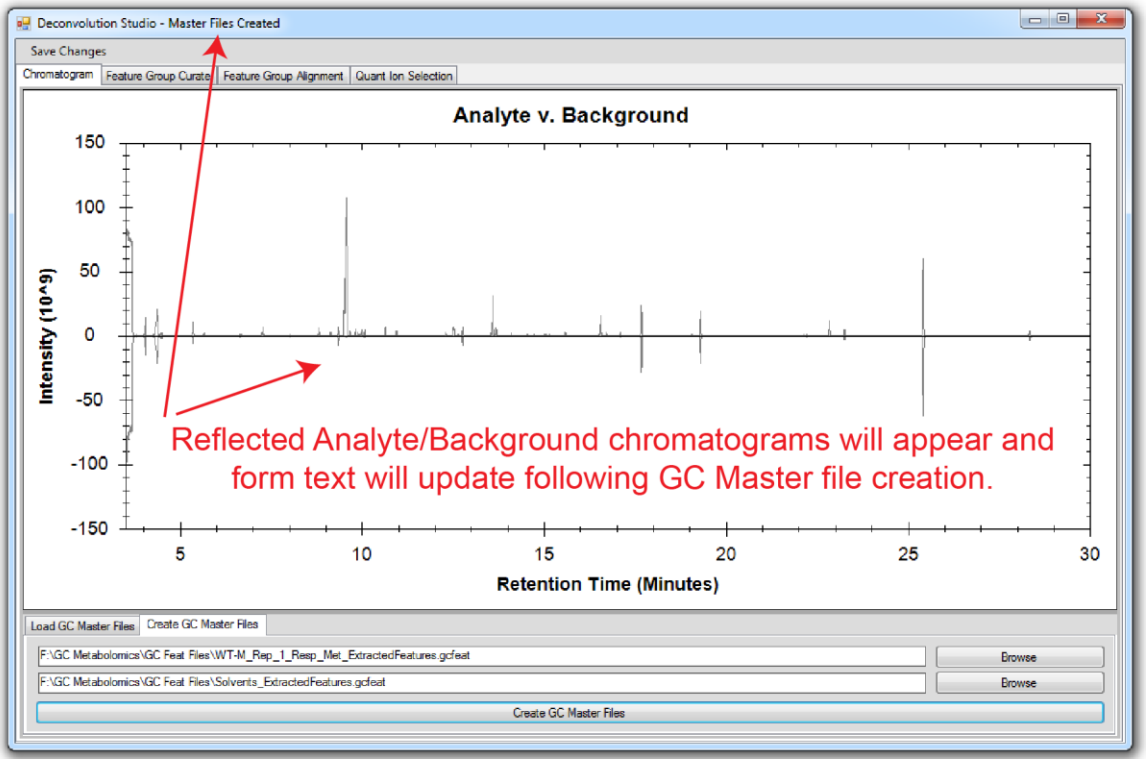
Overview: Deconvolution Studio creates a ‘GC Master File’ (*.gcmast) which serves as a functional target list for metabolite quantitation between replicate MS experiments. This target list is generating by comparing *.gcfeat files from a control MS experiment (WT in the case of Y3K) against a blank MS experiment, and identifying spectral features which are unique to the control. This target list can be manually curated within the tool and is used for quantitation in the GC Quant tool.

Instructions (GC Master File Creation):

1.) Launch the Deconvolution Studio executable ‘Y3K-Deconvolution Studio.exe’.
2.) Create ‘Analyte’ and ‘Background’ GC Master files (*.gcmast). Click the ‘Create GC Master Files’ tab located at the bottom of the GUI. Specify *.gcfeat files from a representative or control MS analysis, and a background MS analysis by clicking the appropriate ‘Browse’ buttons and selecting the appropriate files.
3.) Click the ‘Create GC Master Files’ to create new files. Note: these *.gcmast files will be added to the directory which contains the specified *.gcmast files.

Expected Performance:

After selecting ‘Analyte’ and ‘Background’ *.gcfeat files, and triggering GC Master creation, new *.gcmast files will be written to the directories containing the corresponding *.gcfeat files, reflected gray TIC chromatograms will appear, and the form title will update to reflect completion of this process.



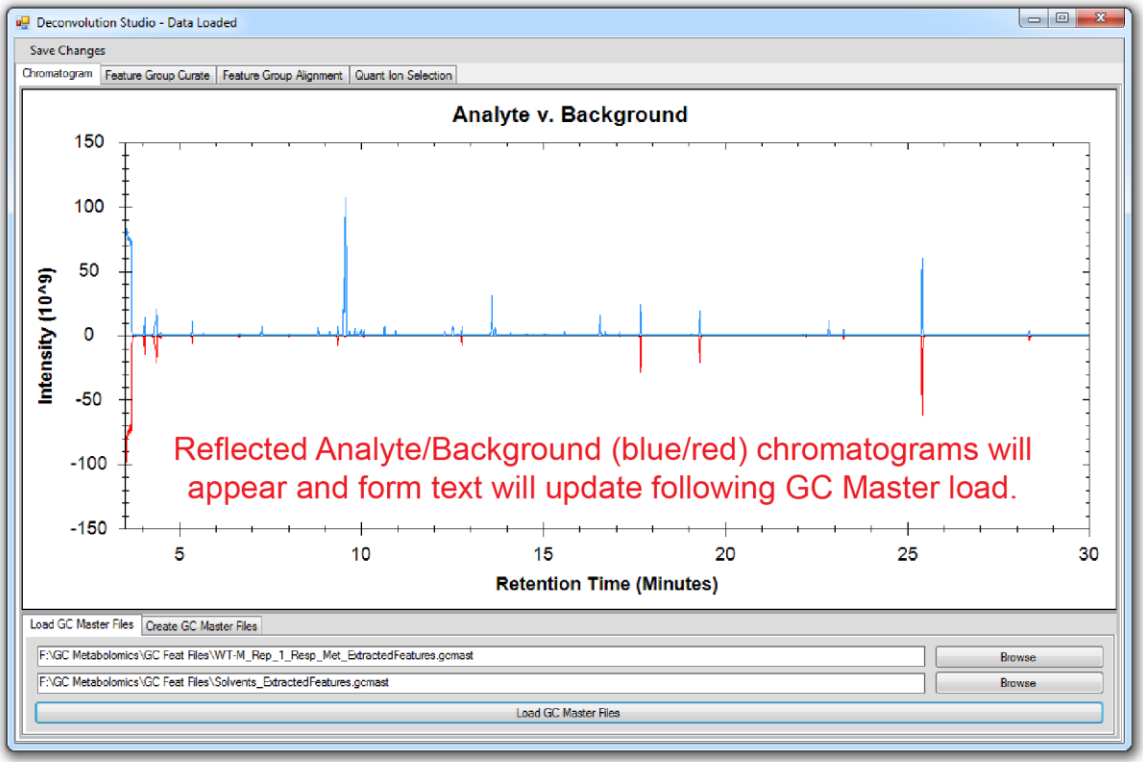
Instructions (GC Master File Creation):

1.) Launch the Deconvolution Studio executable 'Y3K-Deconvolution Studio.exe'.
2.) Load 'Analyte' and 'Background' GC Master files. Select the appropriate files by clicking the appropriate 'Browse' buttons located at the bottom of the GUI.
3.) All chromatographic feature & associated grouping data will load and can be further manipulated within.
4.) To curate individual feature groups select the 'Feature Group Curate' tab. First, navigate to a group of interest by selecting it from the list on the left side of the GUI, then add/remove features from the group by double clicking individual features in the left graph pane and clicking Include/Exclude. . Note: the functionality provided here enables a user to modify the automated feature grouping results.
5.) To select feature groups for downstream quantitation across all samples navigate to the 'Feature Group Alignment' tab. To investigate an 'Analyte' feature group of interest, click on the associated cell in the list on the left hand side of the GUI. Data from this feature group is shown in the upper-left hand graph window. Data from feature groups (eluting around the same time) fron the 'Background' GC Master file are shown in the lower-left hand window. A reflection of the selected 'Analyte' and 'Background' derived spectra showing spectral similarity is shown in the upper-right hand pane. A reflection of the selected 'Analyte' and 'Background' derived chromatographic features is shown in the lower-right hand pane. 'Analyte' feature groups which do not closely match to a 'Background' group are assumed to arise from biologically-relevant molecules, and should be used for downstream quantitation. Select the 'Included' checkbox associated with each feature group to include or omit this group from future quantitation.
6.) To select feature group quant ions to be used in downstream quantitation navigate to the 'Quant Ion Selection' tab. Select a feature group of interest by clicking on it in the list on the left hand side of the GUI. All ions in that feature group will appear in the list on the upper right hand side of the GUI, listed by m/z and intensity in descending order. Quant ions should be chosen based on abundance, peak shape, and absence of surrounding peaks having similar m/z, and can be changed by clicking the up/down arrow buttons. Chromatograms from these ions can be viewed by loading an associated Thermo *.raw file.

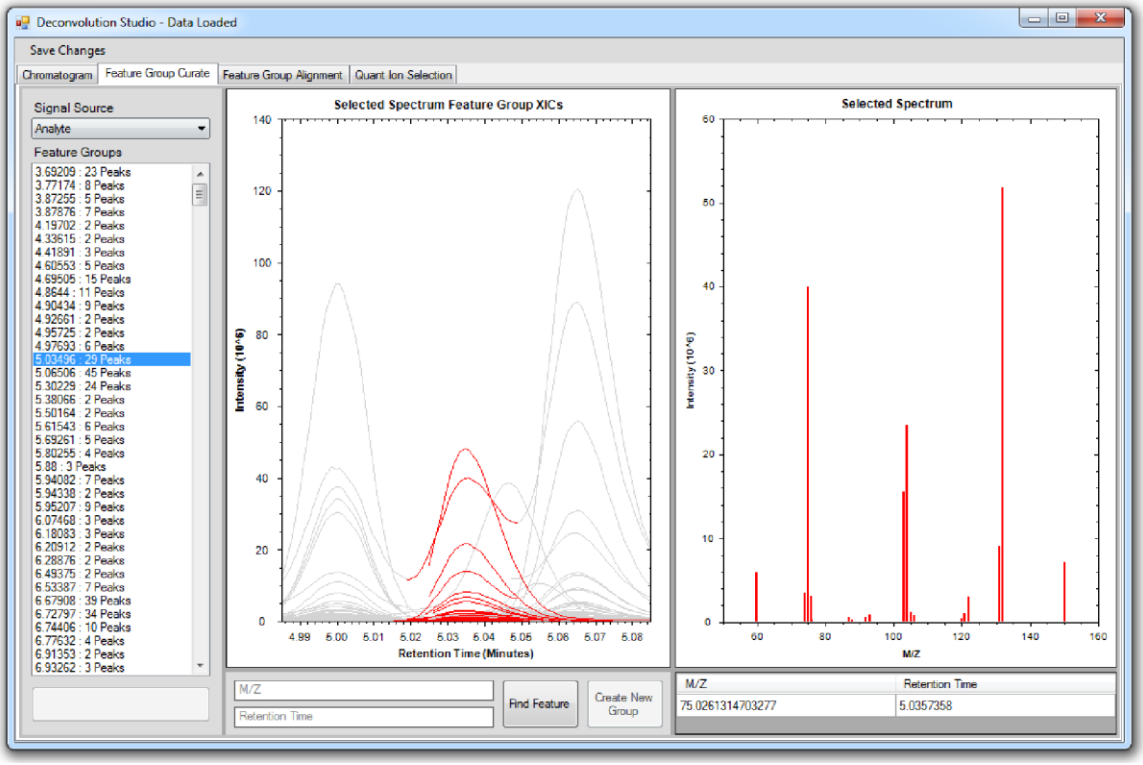
7.) All changes made to the file can be saved by clicking the ‘Save Changes’ button along the top tool bar. The form text should change to reflect the time of the most recent save.

Expected Performance:

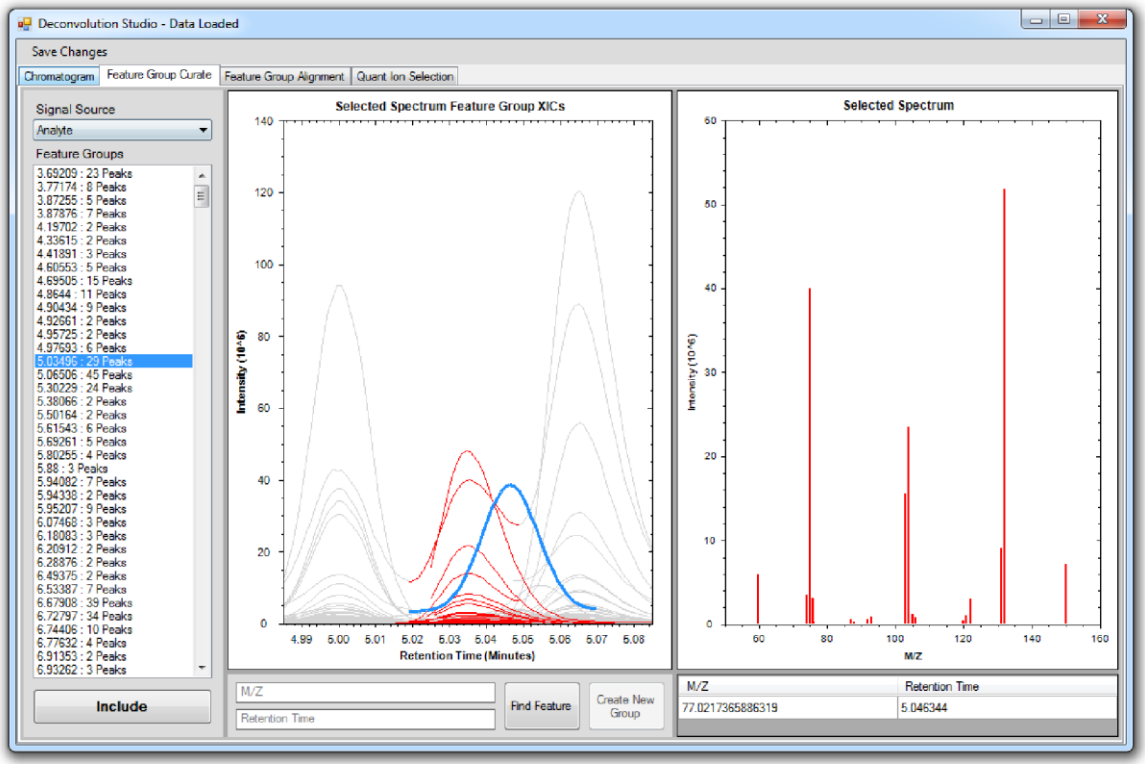
Reflected ‘Analyte’ and ‘Background’ chromatograms should appear after loading GC Master files.



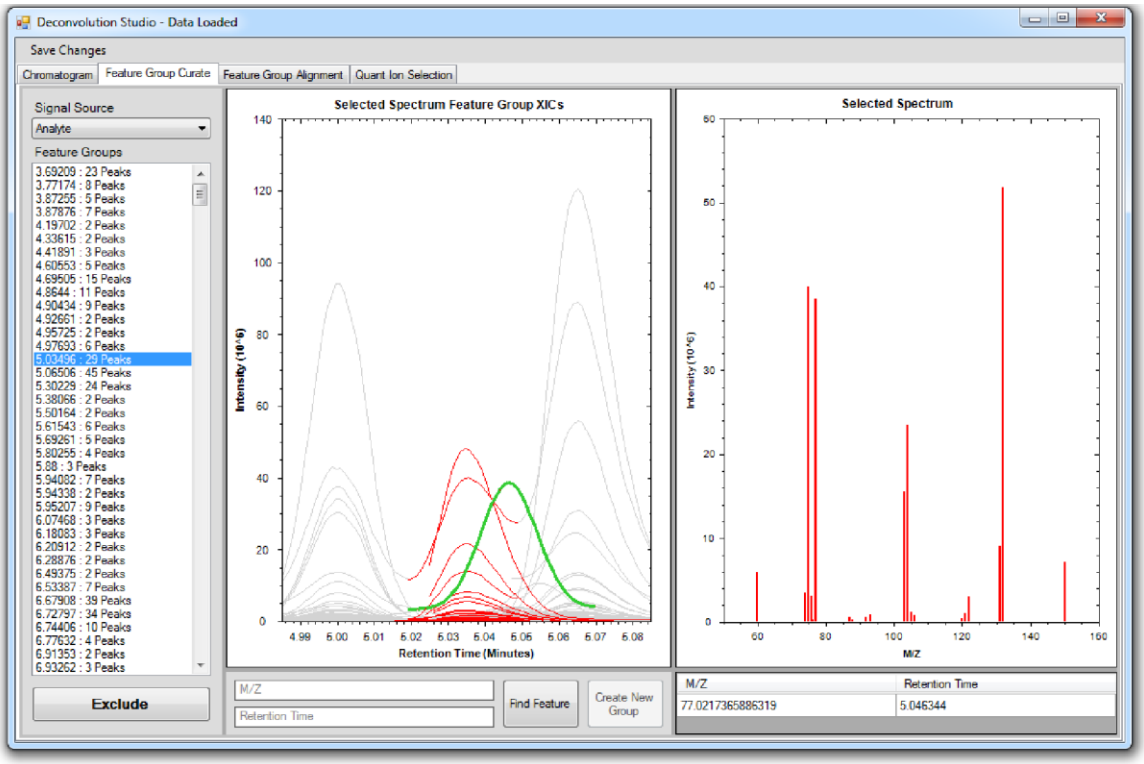
Selecting a feature group (listed according to apex retention time, and the number of chromatographic features in the group) will display the group in both a chromatographic (left pane) and spectrum view (right pane). Chromatographic traces shown in gray are from features not associated with the currently selected group, traces in red indicate intra-group features.



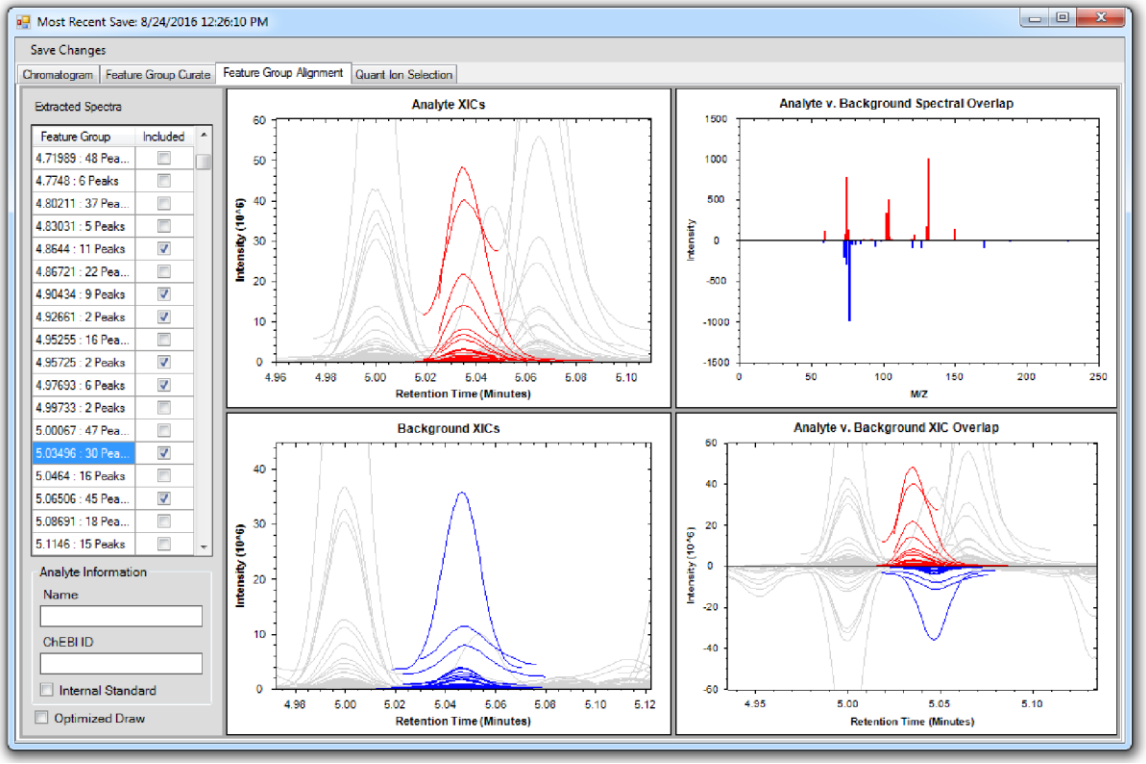
Double-clicking a trace from a feature not currently associated with the group (gray) will cause that feature to turn blue, and activate the ‘Include’ button in the lower-left of the GUI. Click ‘Include’ to add to the selected group. The newly added feature should turn red and a corresponding peak will be added to the spectrum view.



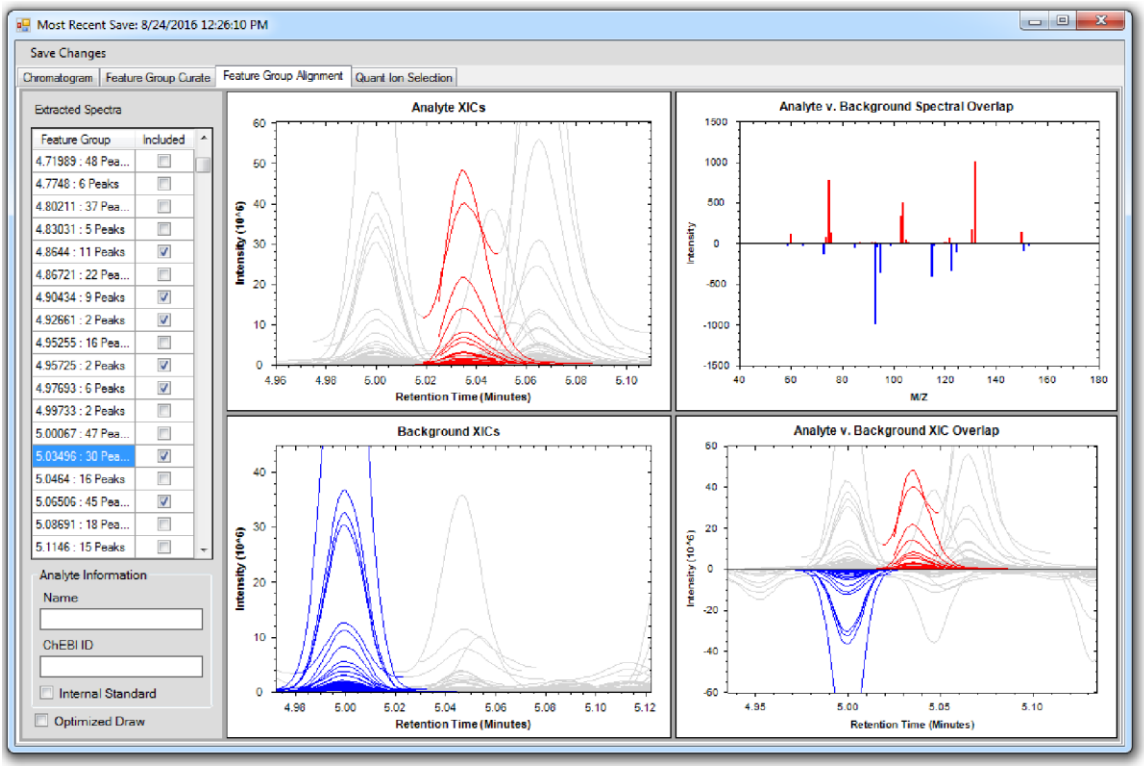
Double-clicking a trace from a feature currently associated with the group (red) will cause that feature to turn green, and active the ‘Exclude’ button in the lower-left hand corner of the GUI. Click ‘Exclude’ to omit that feature from the group. The recently excluded feature should turn gray, and the corresponding spectrum peak will disappear.



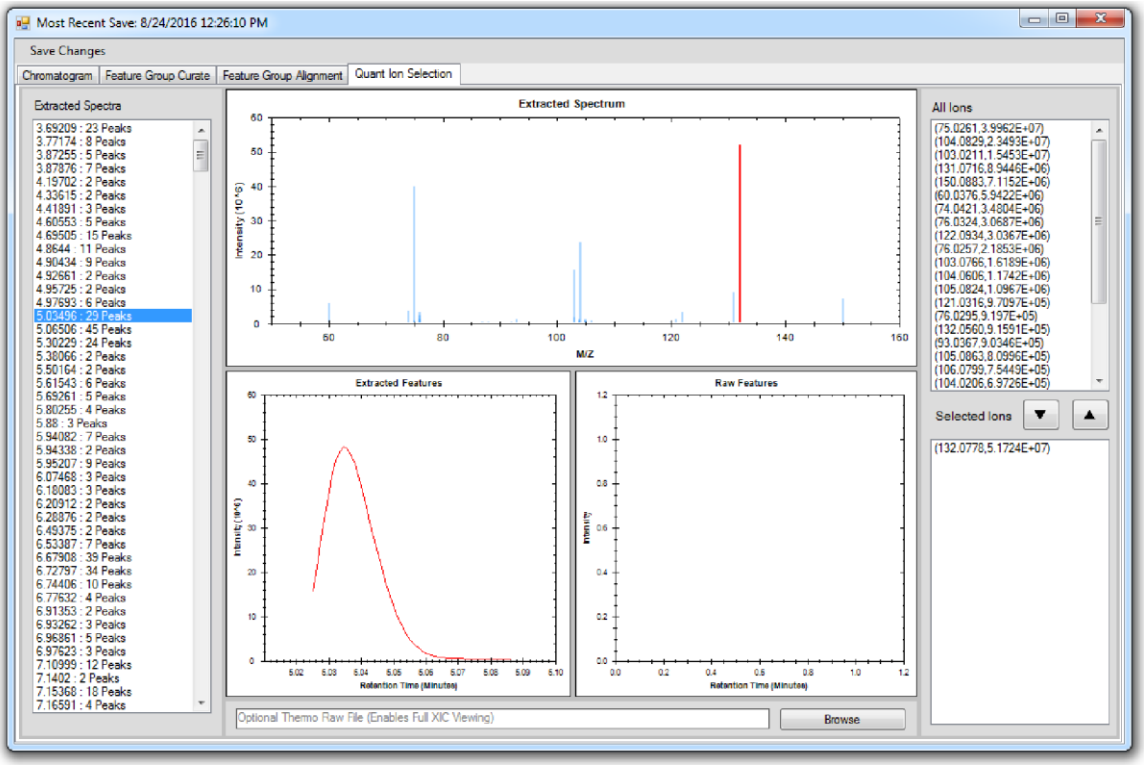
Selecting the ‘Feature Group Alignment’ tab enables you to view all feature groups in the ‘Analyte’ file compared against feature groups having similar retention in the ‘Background’ file. Here you can select feature groups unique to the ‘Analyte’ MS data which should be used for quantitation further downstream. The upper-left hand pane displays the selected ‘Analyte’ feature group. The lower-left hand pane displays the nearest ‘Background’ feature group. The upper-right hand pane shows the spectral similarity between groups, and the lower-right hand pane shows the chromatographic profile similarity.



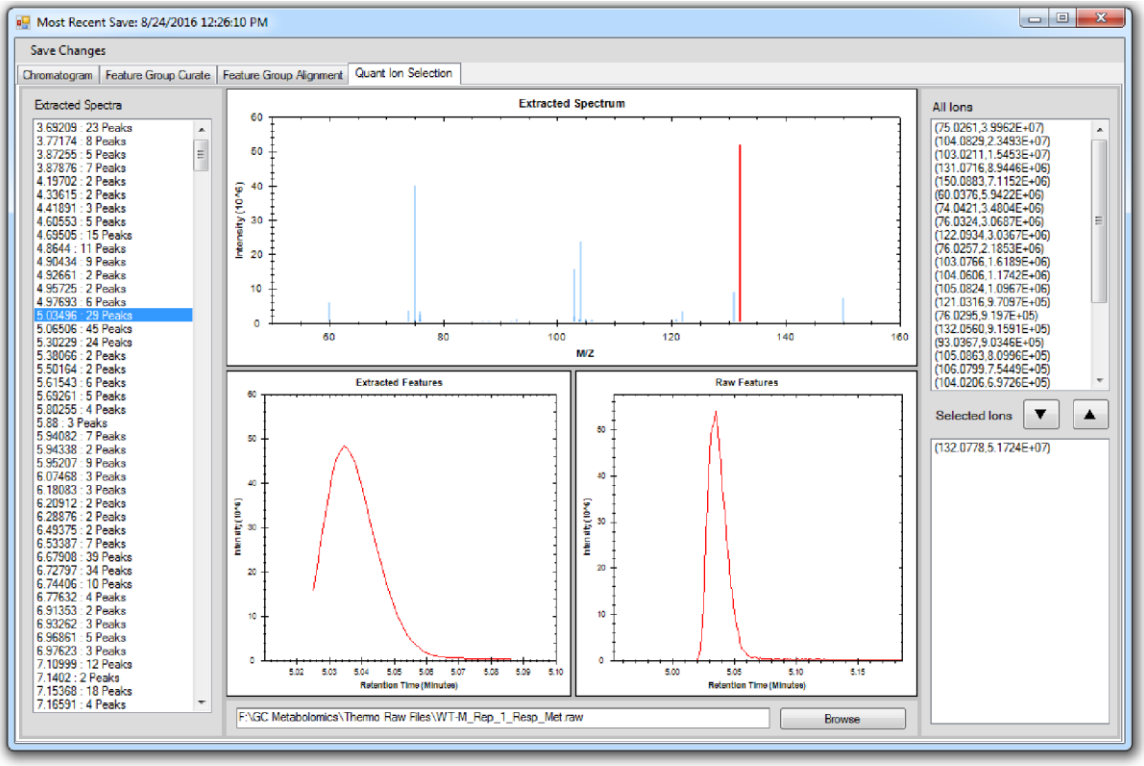
To make comparisons against alternative ‘Background’ feature groups, double-click unselected features in the lower-right hand graph pane. This should cause a new ‘Background’ group to become highlighted and all associated data will update.



For all feature groups selected for downstream quantitation a quant ion is required. This group-specific ion quant ion can be selected under the ‘Quant Ion Selection’ tab. To assign quant ions, select a feature group of interest from the left hand list, choose an ion from the ‘All Ions’ list on the right, and toggle that selection into the ‘Selected Ions’ list. You can observe extracted chromatographic peak shape (post-smoothing) in the lower-left graph pane.



When selecting quant ions it is often useful to consider metabolite fragments which have similar retentions. By clicking ‘Browse’ and loading the corresponding ‘Analyte’ Thermo *.raw file a user can view a raw feature profile within a larger retention time window.



GC Quant

Inputs: MS control master file (*.gcmast), GC experiment file (*.gcexp), user-selected output directory.

**The GC experiment file is a tab-delimited text file which contains information about experimental hierarchy (i.e., replicate and control mapping). This file has the format *.gcfeat file path, replicate name, condition name, associated control condition (leave blank if the replicate is itself a control). An example is shown below and a sample file (***) is included here. These files can be edited and/or created in a spreadsheet viewer such as Microsoft Excel and saved with the appropriate .gcexp extension.

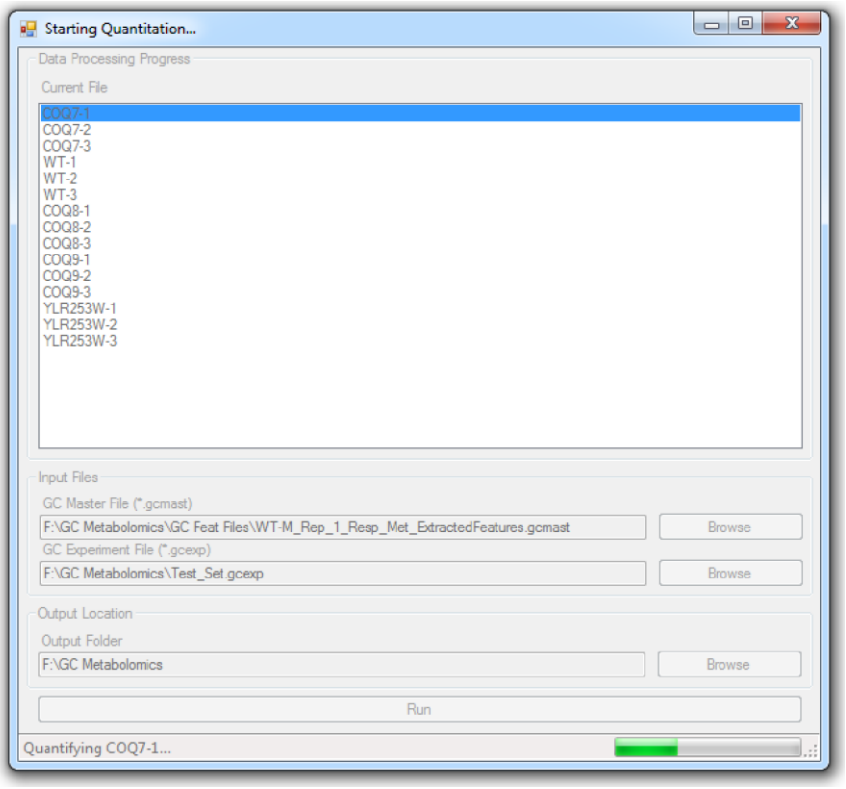
Overview: GC Quant reads in characteristic m/z and retention time values from targeted metabolite features and identifies those features from user-specified *.gcfeat files. Chromatographic realignment is performed to improve feature extraction performance. All feature abundances are normalized using a total ion current (TIC) normalization procedure following extraction. Feature abundances are log2 transformed and some descriptive statistics are reported (fold change relative to associated control and standard deviation).

Instructions:

1.) Launch the GC Quant executable (‘Y3K-GC Quant.exe’).
2.) Specify a MS control master file (*.gcmast) in the ‘Input Files’ window by clicking ‘Browse’ and selecting the appropriate file.
3.) Specify a GC Experiment (*.gcexp) file in the ‘Input Files’ window by clicking ‘Browse’ and selecting the appropriate file.
4.) Select an output directory where a results file (*.gcresults) will be written to by clicking ‘Browse’ in the ‘Output Location’ window and navigating to the appropriate folder.

Expected Performance:

After adding all appropriate files and selecting ‘Run’, *.gcfeat files from all replicate MS analyses to be quantified should be added to list at the top of the GUI. As each file is quantified it should become highlighted in the list.



This program produces an *.gcresults file which can be viewed using a SQLite database reader.