MPACT User Guide

Robert Samples

Revised 22.05.04

Table of Contents

Table of Contents x

1. Installation x
2. Startup x
3. Running MPACT x
    - 3.1. User Interface x
    - 3.2. File Selection x
        - 3.2.1. Peak List x
        - 3.2.2. Sample List x
        - 3.2.3. Metadata File x
        - 3.2.4. Loading a previously run analysis x
    - 3.3. Filtering Settings x
    - 3.4. Analysis Settings x
    - 3.5. Run Analysis x
    - 3.6. Plots and Analysis Results x
        - 3.6.1. Data Review x
        - 3.6.2. Group Analysis x
        - 3.6.3. Dendrogram x
        - 3.6.4. Multivariate Analysis x
        - 3.6.5. m/z vs RT x
        - 3.6.6. Mass Defect x
        - 3.6.7. 3D x
        - 3.6.8. Heatmap x
    - 3.7. Feature Info x
    - 3.8. Analysis Info x
4. Outputs x

MPACT User Guide

Installation

1. Download and install Anaconda from https://www.anaconda.com/products/individual
2. Download and unzip repository

Startup

Windows

Double click MPACT shortcut in the main directory (note that nonfatal errors will cause program termination when run in this way and not through an IDE, if you wish to continue in the event of these errors launch through spyder using the Mac steps)

Mac

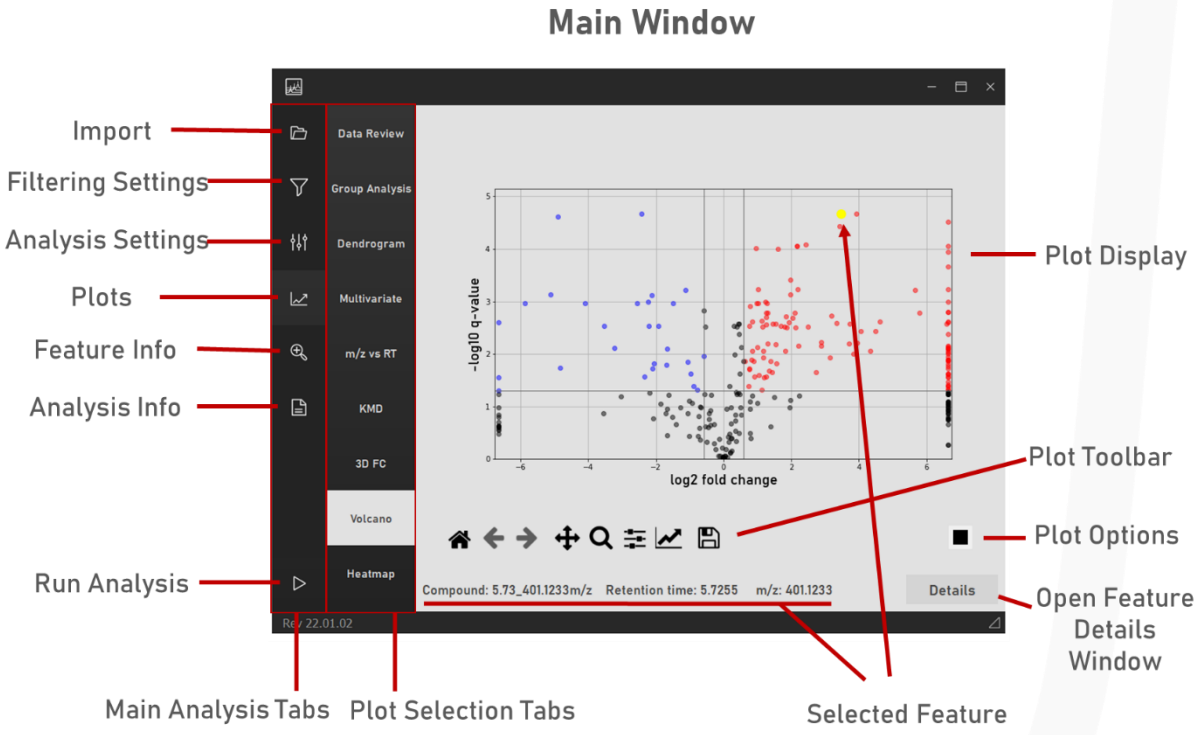
1. Launch spyder through anaconda navigator, or by typing 'spyder' into the anaconda prompt
2. Navigate to code directory in the main project folder
3. Open main.py (not main.py with underscores) in spyder
4. click run button (green arrow in top toolbar)

Running MPACT

User Interface

The main user interface elements and data analysis tools in MPACT are contained in the main window and the compound details window. Further plot navigation and formatting options can be accessed in the matplotlib toolbars and other plot customization features can be accessed in the plot options dialog.

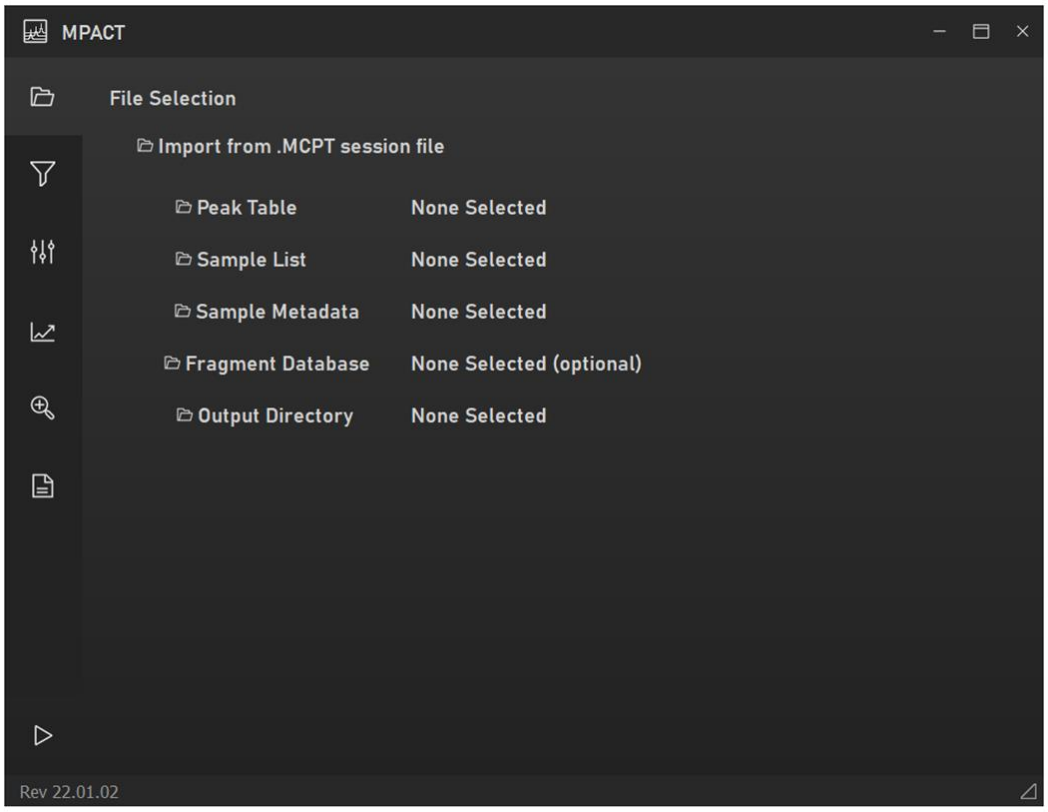
Main MPACT window displaying key user interface elements



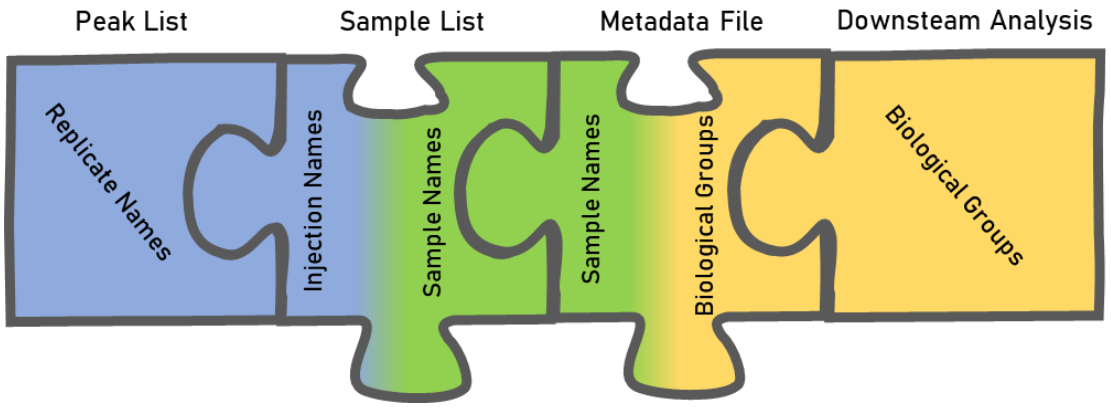
Compound details window displaying key user interface elements

File Selection

A minimum of three files are required to analyze a dataset in MPACT, a peak list file generated by Progenesis QI, MZmine 2, or MS-DIAL, a sample list containing the filename of the raw data files used to generate a peak list and corresponding sample designations, and a metadata file which groups sample designations specified in the sample list to a set of biological groups or treatment conditions to be analyzed. Users may specify an output directory for processed data and optionally include a MS/MS fragment database in .msp format, only fragmented databases generated with Progenesis QI are currently supported.



MPACT file selection tab



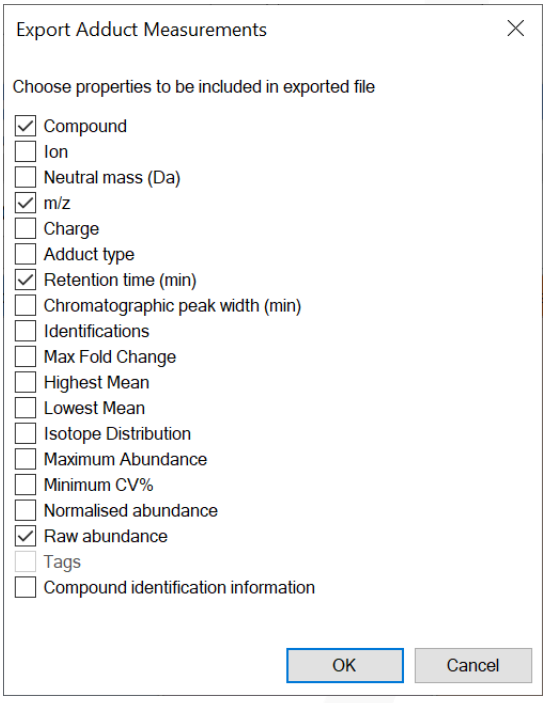
Peak lists are typically organized by individual injections/data files but investigators are usually interested in biological/treatment groups. MPACT sample list and metadata files link together MS file names/injection replicates to sample names to different biological groups. This association allows easy downstream data processing, visualization, and analysis.

Peak List

Peak lists generated by Progenesis, MS-DIAL, and MZmine are natively supported by MPACT. MS-DIAL .txt format peak lists and MZmine .csv peak lists will automatically be converted into Progenesis format for analysis. Other peak list formats can be easily modified to be accepted by MPACT. An example peak list in Progenesis format is shown on the following page. Acquisition of technical/injection triplicates is highly recommended for maximum reproducibility and to take full advantage of the data processing and tools in MPACT.

Exporting a peak list from Progenesis

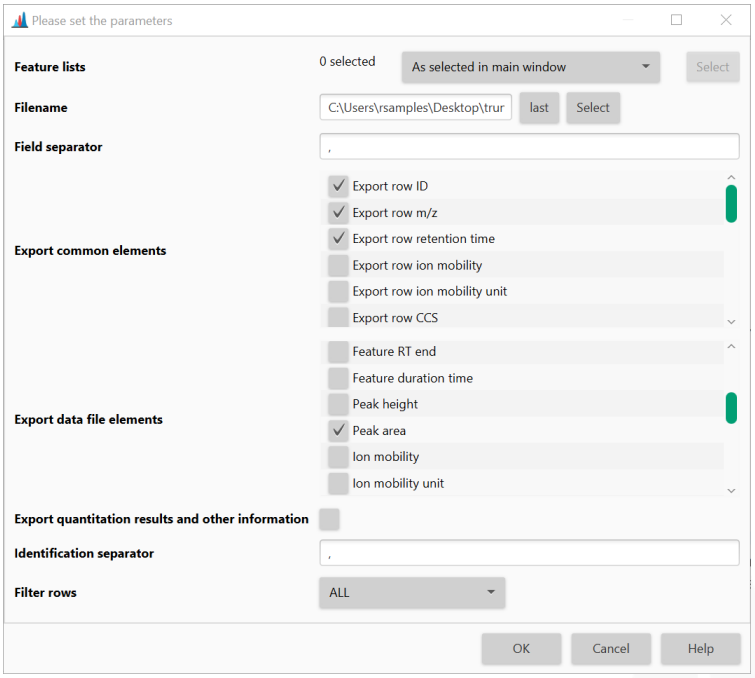
To export a suitable peak list in Progenesis navigate to the "Review Compounds" tab and then select "File>Export Compound Measurements", and then select the following export options. Either raw or normalized abundance may be selected



Progenesis feature list export parameters

Exporting a peak list from MZmine 3

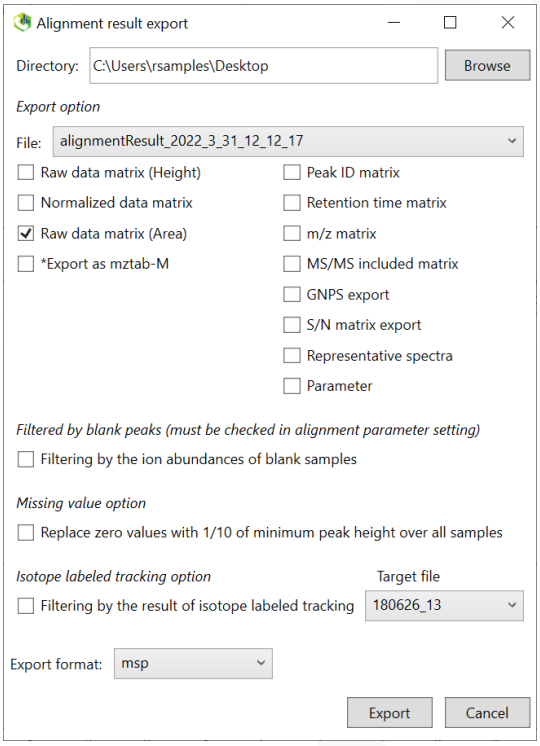
To export a suitable peak list in MZmine 3 generate select an aligned peak list, and select "Feature list methods>Export feature list>CSV (Legacy MZmine 2_)". Select the following export options:



MZmine 3 feature list export parameters

Exporting a peak list from MS-DIAL

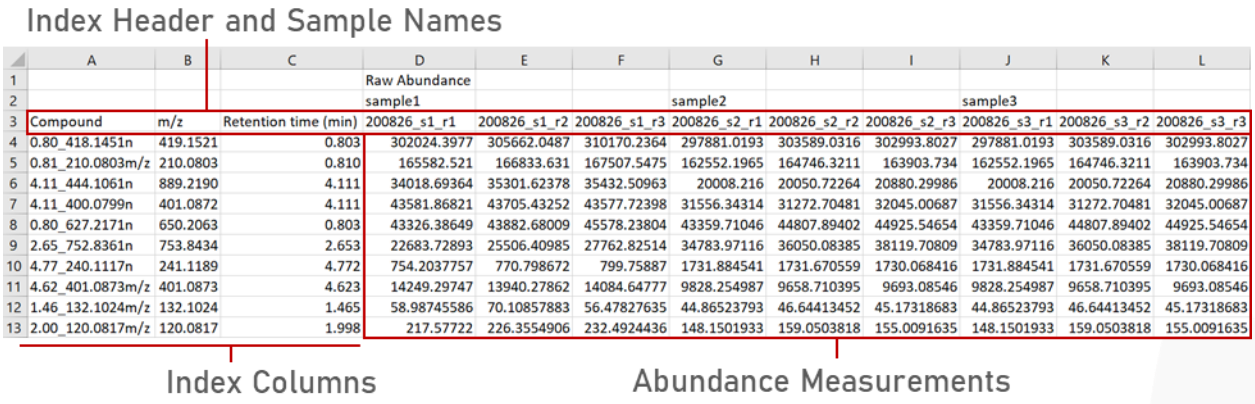
To export a suitable peak list in MS-DIAL select Export>Alignment result and then select the following export options:



MS-DIAL feature list export parameters

Editing a peak list from a currently unsupported platform

If manually converting a peak list from a currently unsupported platform note that the groupings in the first and second header rows are not important for MPACT analysis as all replicate, sample, and biological group information is parsed from the sample list and metadata files. The third header row is necessary and contains index headers, and individual injection/LC-MS file names. A compound designation/name column, m/z column, and Retention time column are required as indices. The abundance of each feature in each sample is included in the main table body.



	A	B	C	D	E	F	G	H	I	J	K	L
1				Raw Abundance								
2				sample1			sample2			sample3		
3	Compound	m/z	Retention time (min)	200826_s1_r1	200826_s1_r2	200826_s1_r3	200826_s2_r1	200826_s2_r2	200826_s2_r3	200826_s3_r1	200826_s3_r2	200826_s3_r3
4	0.80_418.1451n	419.1521	0.803	302024.3977	305662.0487	310170.2364	297881.0193	303589.0316	302993.8027	297881.0193	303589.0316	302993.8027
5	0.81_210.0803m/z	210.0803	0.810	165582.521	166833.631	167507.5475	162552.1965	164746.3211	163903.734	162552.1965	164746.3211	163903.734
6	4.11_444.1061n	889.2190	4.111	34018.69364	35301.62378	35432.50963	20008.216	20050.72264	20880.29986	20008.216	20050.72264	20880.29986
7	4.11_400.0799n	401.0872	4.111	43581.86821	43705.43252	43577.72398	31556.34314	31272.70481	32045.00687	31556.34314	31272.70481	32045.00687
8	0.80_627.2171n	650.2063	0.803	43326.38649	43882.68009	45578.23804	43359.71046	44807.89402	44925.54654	43359.71046	44807.89402	44925.54654
9	2.65_752.8361n	753.8434	2.653	22683.72893	25506.40985	27762.82514	34783.97116	36050.08385	38119.70809	34783.97116	36050.08385	38119.70809
10	4.77_240.1117n	241.1189	4.772	754.2037757	770.798672	799.75887	1731.884541	1731.670559	1730.068416	1731.884541	1731.670559	1730.068416
11	4.62_401.0873m/z	401.0873	4.623	14249.29747	13940.27862	14084.64777	9828.254987	9658.710395	9693.08546	9828.254987	9658.710395	9693.08546
12	1.46_132.1024m/z	132.1024	1.465	58.98745586	70.10857883	56.47827635	44.86523793	46.64413452	45.17318683	44.86523793	46.64413452	45.17318683
13	2.00_120.0817m/z	120.0817	1.998	217.57722	226.3554906	232.4924436	148.1501933	159.0503818	155.0091635	148.1501933	159.0503818	155.0091635

Example of a Progenesis format peak list. It is possible to import peak list data from currently unsupported platforms by manually editing them into Progenesis format or through the use of scripts (R or Python are good options for reformatting)

Sample List

The sample list only requires columns with LC-MS injections/file names and the corresponding sample name/code. Optionally more information can be included. The sample list can be copied from an LC-MS sample queue in the appropriate format.

	A	B	C	D	E	F	G	H
1	Injection	File Text	Sample_Notes	MS method	LC method	Vial_Position	Injection volume	Sample_Code
2	200826_s1_r1							sample1
3	200826_s1_r2							sample1
4	200826_s1_r3							sample1
5	200826_s2_r1							sample2
6	200826_s2_r2							sample2
7	200826_s2_r3							sample2
8	200826_s3_r1							sample3
9	200826_s3_r2							sample3
10	200826_s3_r3							sample3

An example of an MPACT sample list linking file names in the peak list to sample names/codes

Metadata file

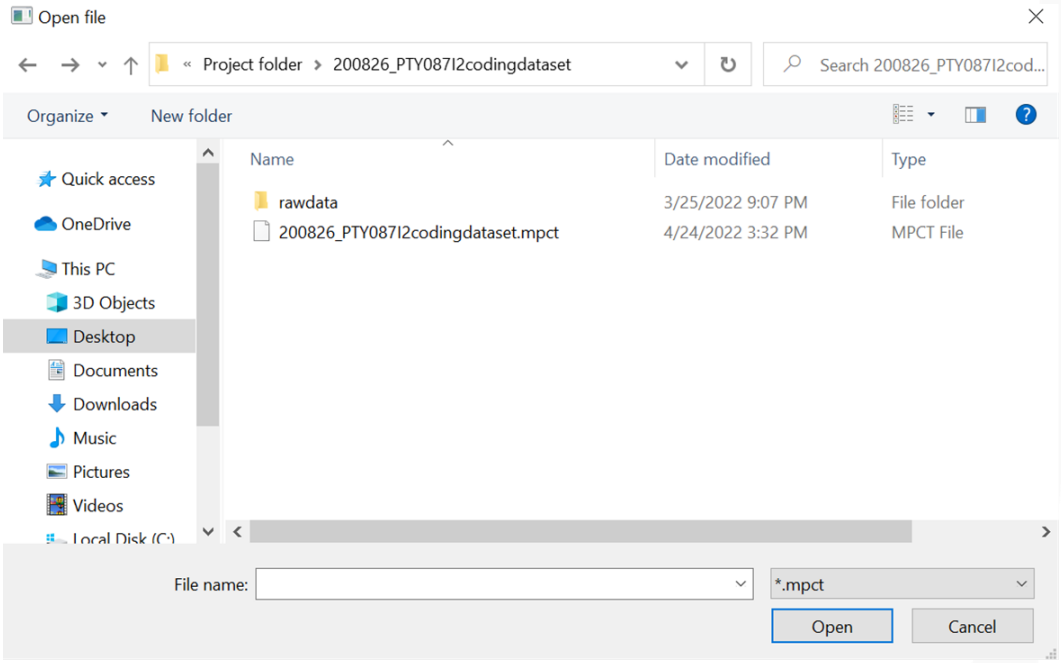
The metadata file only requires columns with sample name/code and the corresponding biological/treatment group. Optionally more information can be included.

	A	B	C	D
1	Sample_Code	Organism	Biological_Group	Extract_Notes
2	sample1		Species1	
3	sample2		Species1	
4	sample3		Species2	

An example of an MPACT sample list linking sample names/codes to biological/treatment groups.

Loading a previously run analysis

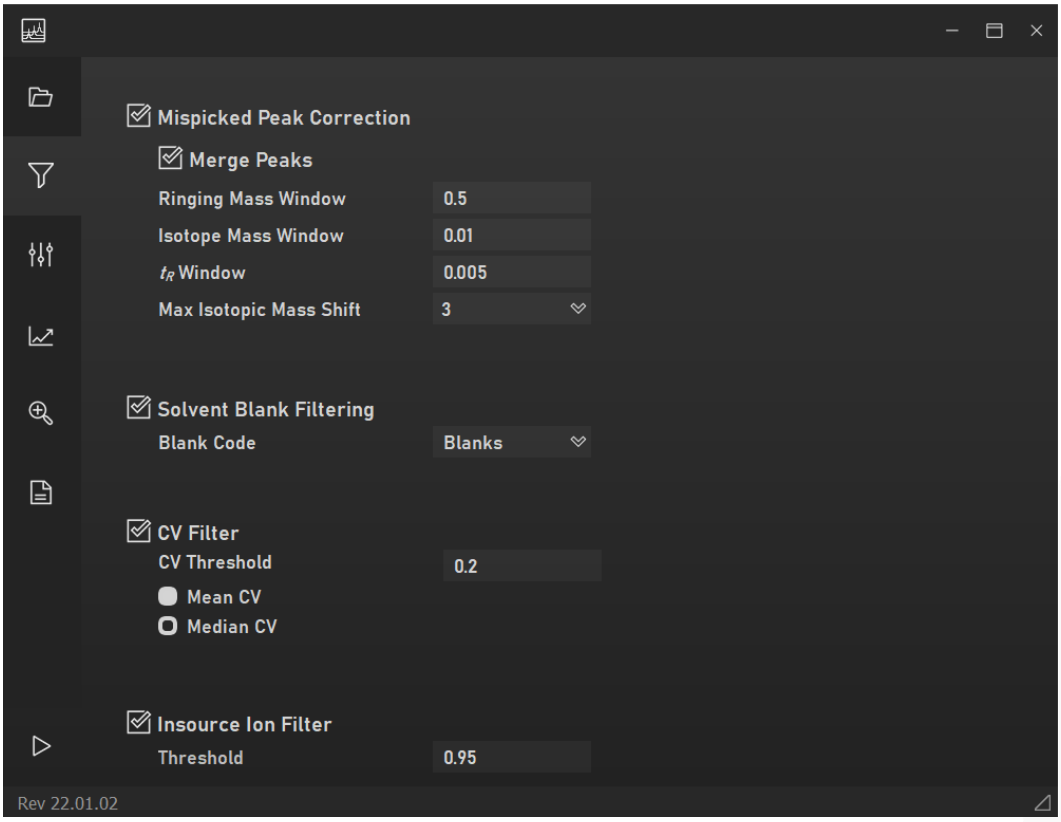
Alternatively, a previous MPACT analysis may be regenerated from an mpct save file containing all raw data and parameters by selecting the "Import from .MPCT session file" and selecting a save file. Save files are automatically generated in the output directory location with all other processed data when an MPACT analysis is complete. When opening an MPACT save file a new folder in the same location will be generated containing all raw data and outputs.



Loading a .mpct file from a previously run analysis

Filtering Settings

MPACT allows filtering of mispicked peaks, solvent blank features, nonreproducible features, and in-source ions based on user-specified parameters.



MPACT filtering settings tab

Mispicked peaks are identified as being similar in terms of retention time and mass to lower mass features with greater abundance and are suggested to be the result of incorrect splitting of isotopic patterns during peak picking, detector saturation artifacts, or incorrect identification of multiply charged oligomers. Suggested parameters are included as default values and similarly low values are suggested to minimize incorrect filtering of closely related analytes.

Features present in blanks are identified based on a relative or absolute abundance threshold specified in the analysis parameters page. Relative abundance threshold is suggested to minimize interference from low abundance carryover between samples with a threshold of 0.1-0.05, corresponding to filtering of features present in blanks at greater than 1-5% of their abundance in samples.

Filtering of nonreproducible features is conducted based on median or mean coefficient of variation (CV), also known as relative standard deviation, amongst technical replicates. Filtering based on median CV values is suggested over mean CV to minimize the effects of highly

nonreproducible outliers. CV filtering thresholds reported in the literature are generally between 0.2-0.3 however for investigational work or when CV cannot be expected to be known with great accuracy (In the case with a small number of samples or very high chemical diversity resulting in features being detectable in only a small number of samples) a higher threshold of up to 0.5 may be appropriate. The selected threshold may be altered based on the results of hierarchal clustering analysis.

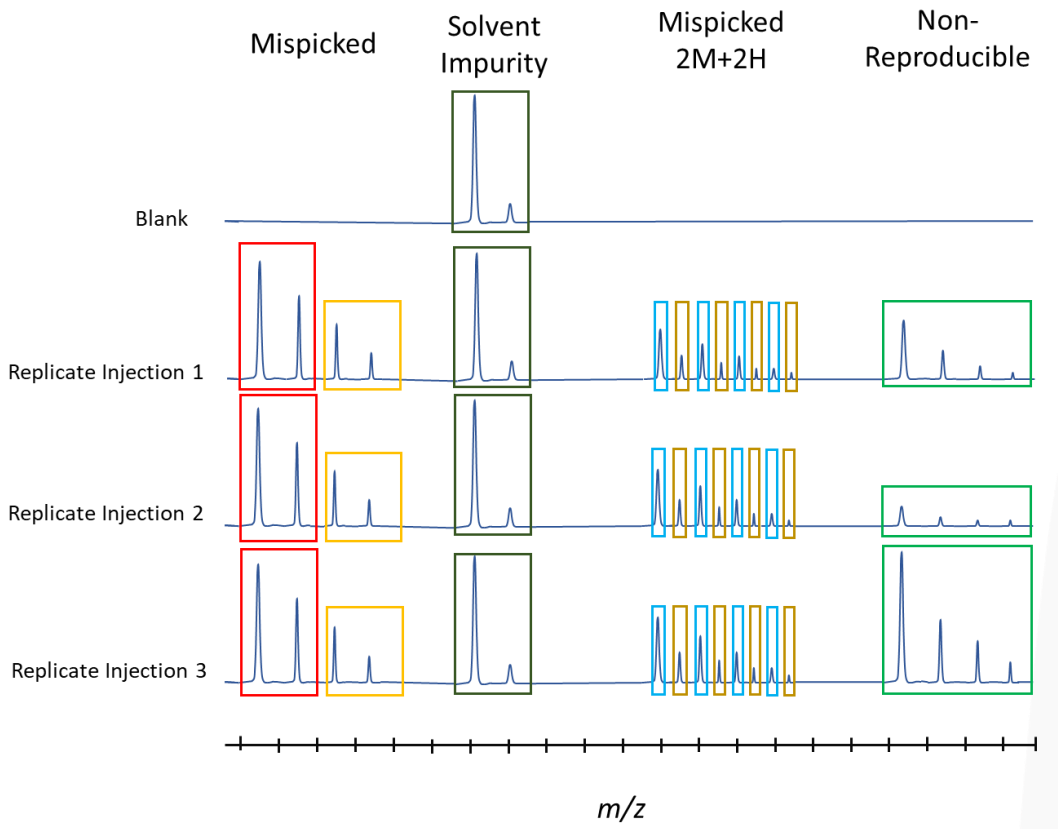


Diagram demonstrating identification of groups that are incorrectly peak picked (peak picking boxes split between isotopic pattern), nonreproducible features, and features present in solvent blanks

In-source ion deconvolution is conducted by generating retention time bounded cosine correlational matrices within MS1 scans. Highly correlated groups of features are identified and low mass features filtered after the highest mass features is identified as the likely parent ion within a cluster. A high value for correlation is suggestion to minimize the risk of filtering of discreet analytes that are similarly up or downregulated across samples.

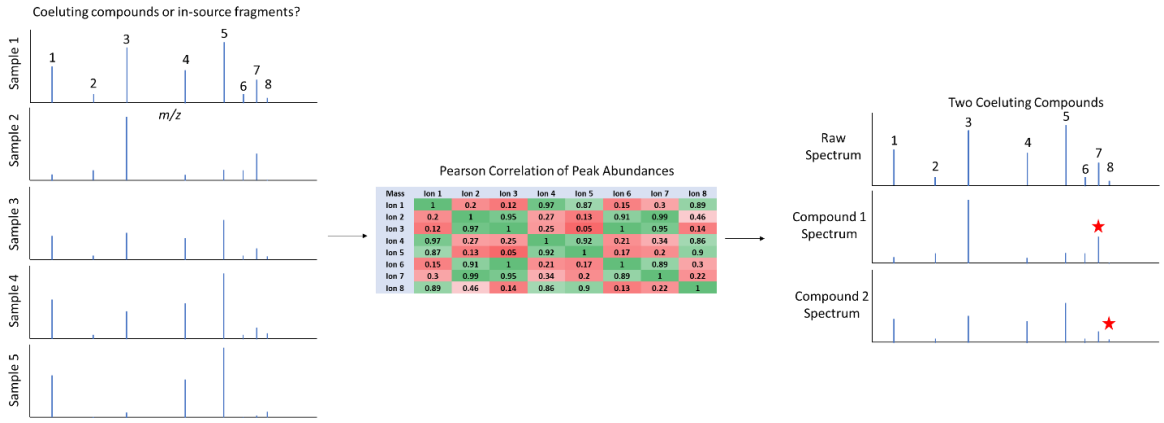
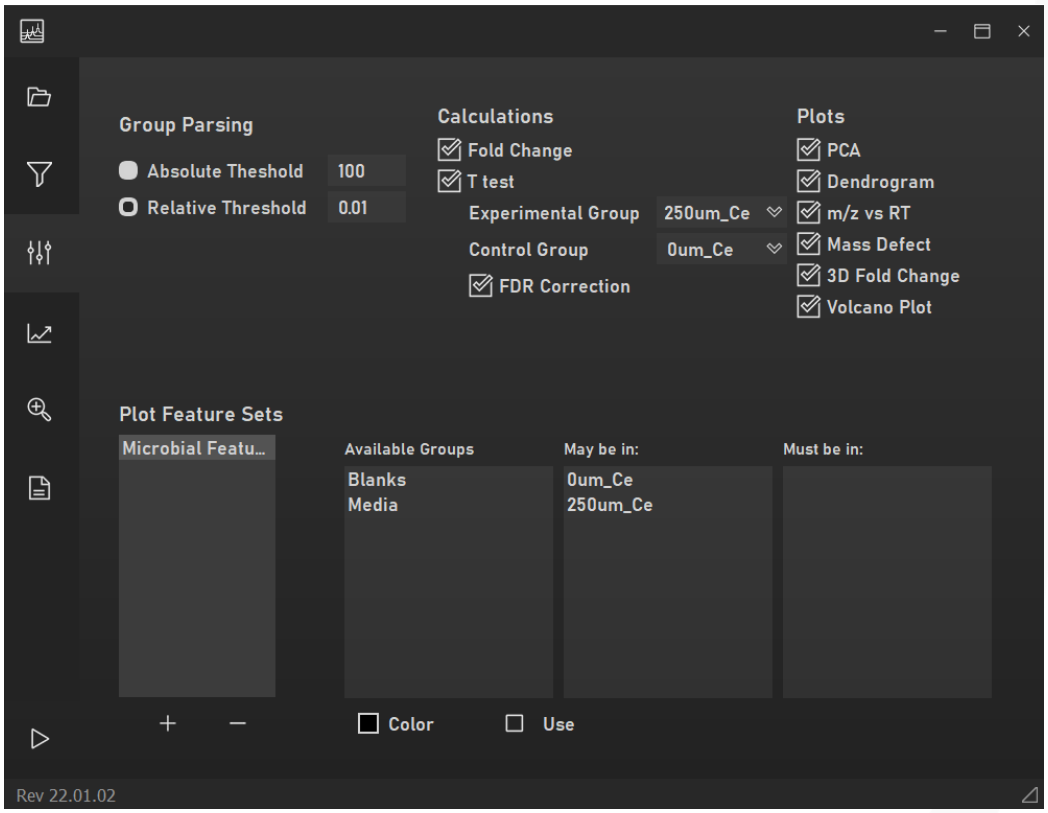


Diagram demonstrating identification of insource fragments. A correlation matrix is used to deconvolute groups of features at the same retention time, the highest mass features is assumed to be the molecular ion within each cluster (denoted by a red star)

Analysis Settings

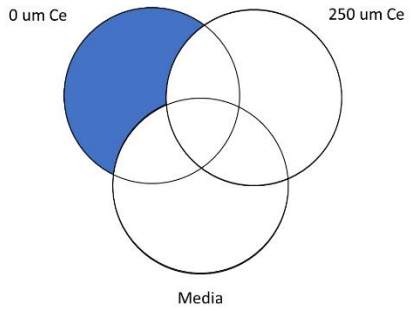
A relative or absolute group parsing threshold is used to filter features present in solvent blanks or other groups as well as designate presence or absence of a feature in a given group. Relative threshold is species as a fraction of abundance in the group in which a features is most abundant, and absolute abundance threshold is specified as a raw or normalized abundance.

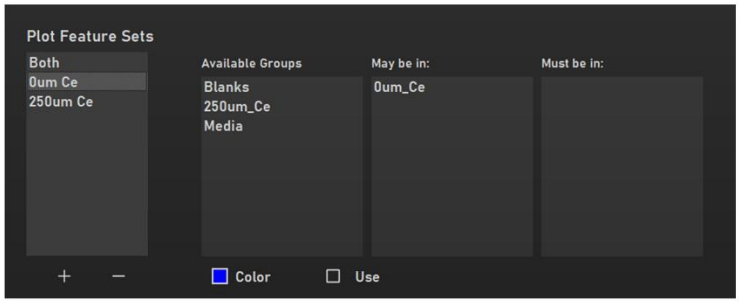


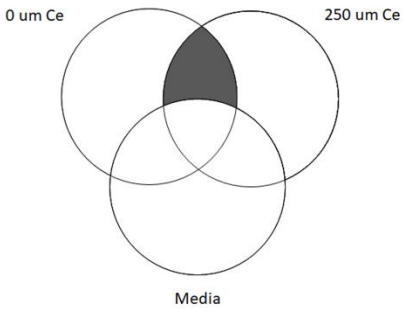
MPACT analysis settings tab

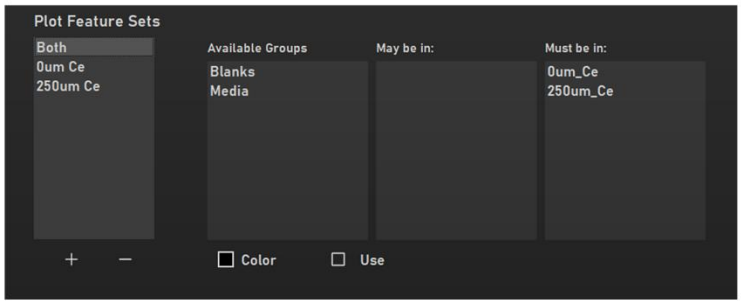
Various plots and calculations may be selected to generate based on user preference. Analysis speed is predominantly based on filtering and as such deselection of the default options typically does not increase analysis speed significantly. For generation of Volcano plots fold change and T-test options must be selected and an experimental and treatment group specified. The available groups are automatically detected based on the metadata file provided by the user. False discovery rate correction using the Benjamini-Hochberg procedure based on the number of features plotted may be selected to reduce type II errors arising from multiple hypothesis testing.

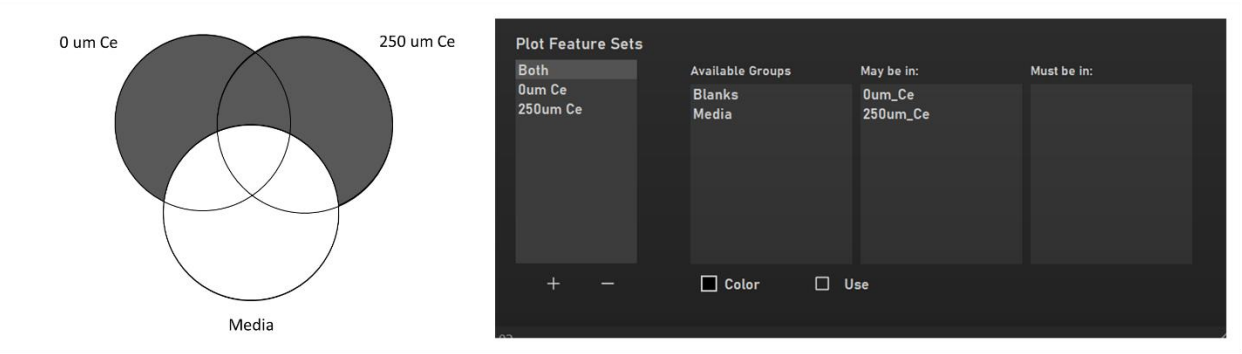
The features to be plotted in a given colour is designated in the "Plot Feature Sets" section. A set of features to be plotted in a given colour is created by clicking the + button and named by double clicking the default name. Unwanted sets can be removed with the - button. Users can toggle between feature sets by clicking them. When a feature set is created the biological groups designated in the metadata file are populated in the available groups list. Groups may be selected by clicking (shift+click and control+click can be used to select multiple groups or a range.) Groups can be dragged into the "May be in" and "Must be in" lists. Colour can be designated by clicking the colour button to launch the colour selector window. A feature will be plotted in the selected colour if it is identified in ALL of the biological groups in the "Must be in" list and none of the groups that were not moved from the "Available groups" list. Groups in the "May be in" are those that a feature may be present in, but is not required, for the feature to be plotted in the selected colour. To plot all features in a dataset all groups (except the solvent blank group if blank filtering is selected) can be dragged into the "May be in" list. The following examples demonstrate which features will be plotted in a given colour.

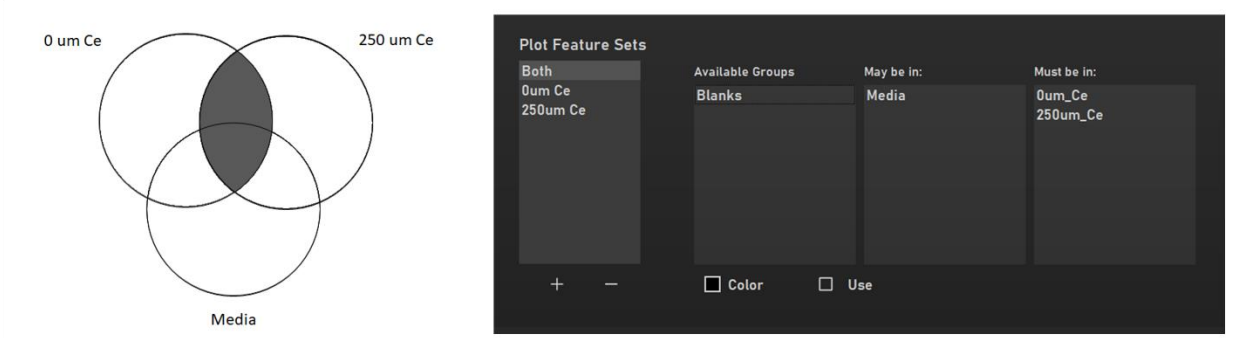












Example of which features in a set of biological groups (0 um Ce, 250 um Ce, and Media control) will be plotted based in a given colour based on their presence/absence in these groups as indicated by Venn diagrams.

Run Analysis

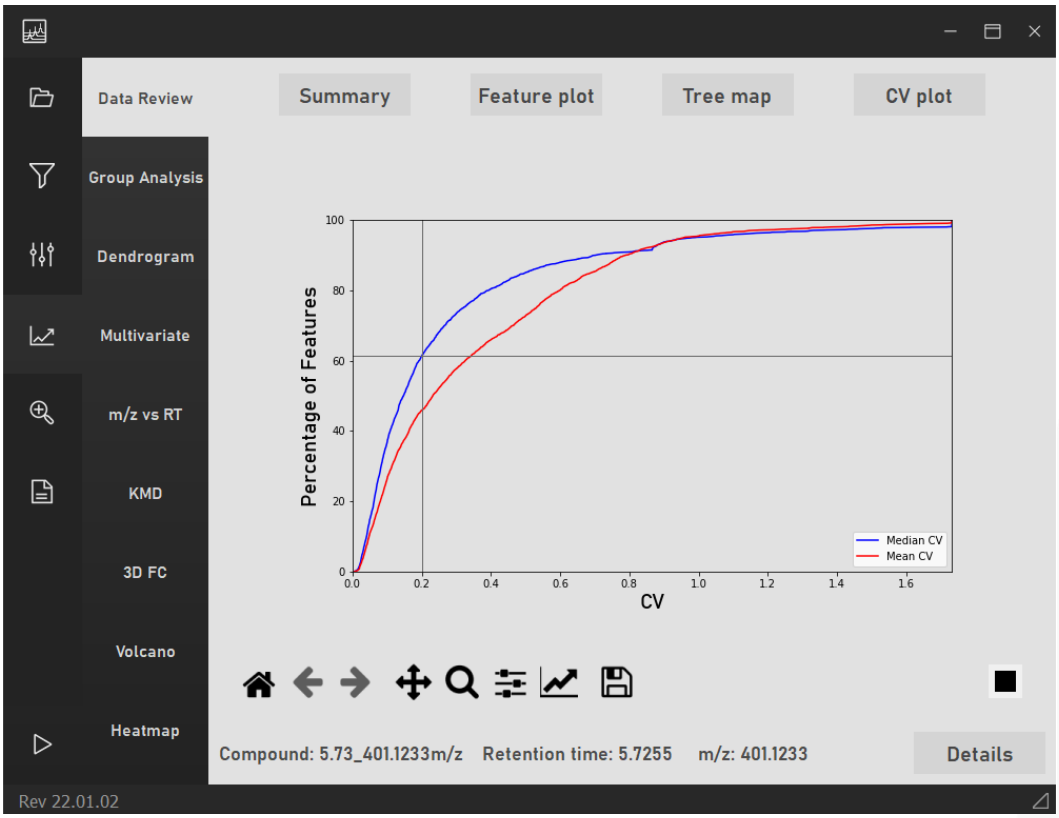
The requirements to process data are met after completing file selection, filtering settings, and analysis settings steps above and the analysis can be started by clicking the run button in the bottom left corner. A message in the bottom bar displaying "Analysis Complete" will be displayed when MPACT has finished running. Note that MPACT does not currently use multithreading and as such the GUI may become unresponsive during computationally intensive tasks including initial data processing, searching the feature database the first time the feature info tab is selected after the analysis is complete, and in some datasets for a few seconds when selecting individual features, particularly when the abundance data is viewed in the compound details window while switching features. Work is ongoing to optimize performance.

Plots and Analysis Results

All generated plots can be viewed by clicking the plots button in the left pane opening an additional pane with available data visualization tools. Plots can be interacted with, zoomed, panned, and saved, using the toolbar below each plot. Features may be selected in most plots by clicking them which will highlight each feature and automatically update other plots with the highlighted feature and display its name, m/z, and retention time below the plot.

Data Review

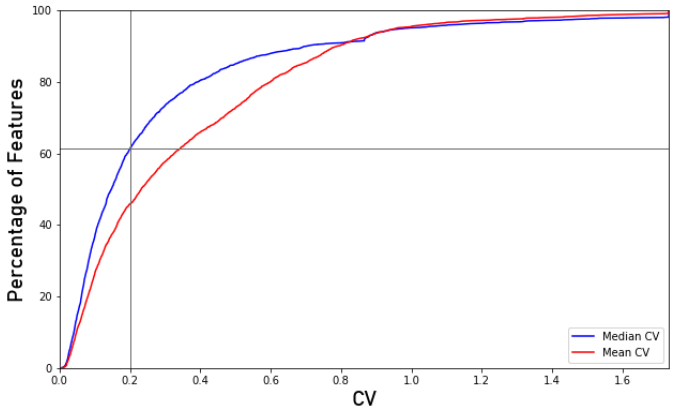
Data Review contains several different tabs. Summary lists an estimate of data quality based on the percentage of the theoretical maximum reproducibility the CV50 (the CV of the median feature in the dataset) is located. The number of features filtered by each filtering step is included.



MPACT data review tab within the main plots tab, showing a CV distribution plot

CV Plot

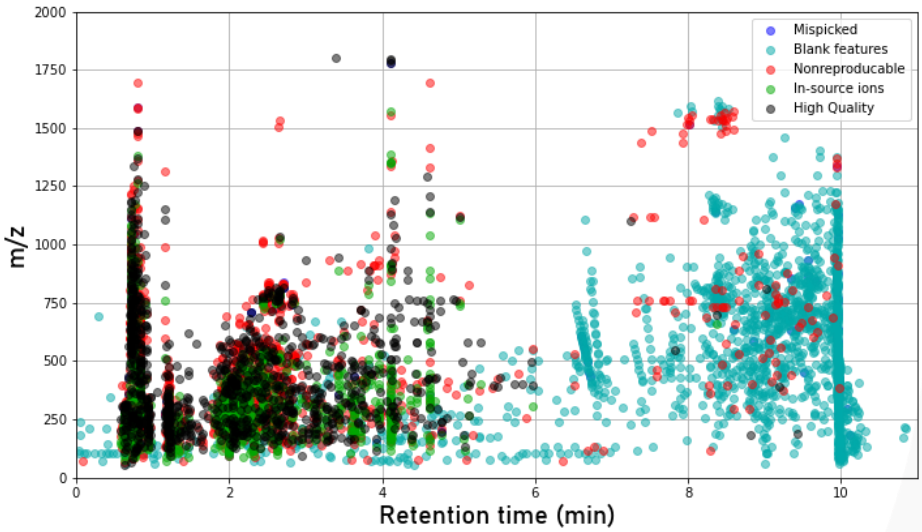
The CV plot shows the distribution of mean and median CV values in the dataset and can be used to visualize reproducibility and evaluate the possibility of peak picking and alignment issues which can result in less steep rises in the distribution curve. A highly skewed distribution of CV values (in which some sets of technical replicates are much less reproducible than others) is indicated by a large gap between mean and median CV curves. The CV plot is useful for investigating data quality and alignment issues which can manifest as very shallow CV plots with a low percentage of features meeting commonly used CV thresholds between 0.2-0.3 CV. Marked multimodality (sharp rises in the middle of the distribution) may indicate inconsistent of features in many technical replicate sets.



MPACT CV distribution plot showing relatively good reproducibility with minimal skew and multimodality.

Feature plot

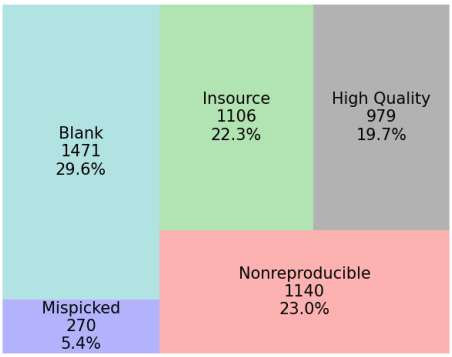
Features are visualized in the feature plot of m/z vs retention time based on which filter they were removed by. Features passing all filtering steps are denoted as high-quality features in black.



MPACT feature plot showing features removed by various filtering stages

Treemap

The number and percentage of features removed by each filtering step are shown as a treemap.



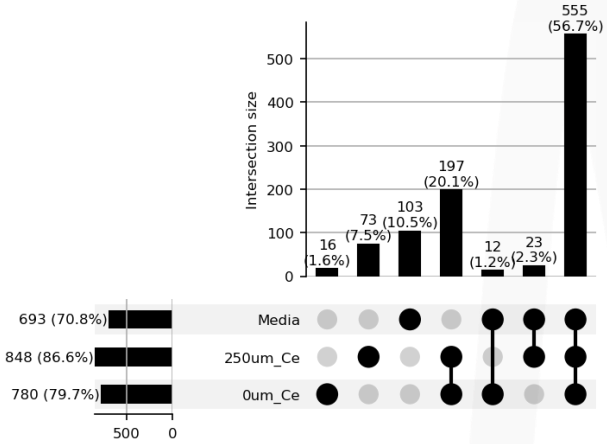
MPACT Treemap showing filtering rates.

Group Analysis

The group analysis tab allows differences between treatment groups to be visualized based on group set analysis and correlational analysis.

UpSet Plot

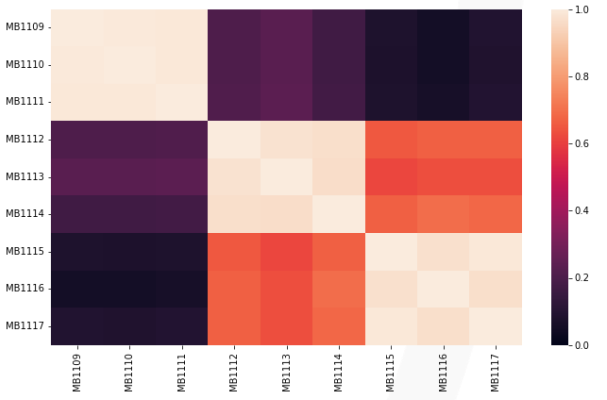
Set analysis is visualized with an UpSet plot showing the number of features (left bar chart) in each treatment group, and the number of features present (top bar chart) in some combination of groups (as indicated by a dot plot).



MPACT UpSet plot showing distribution of features across sample sets

Spearman Correlation Matrix

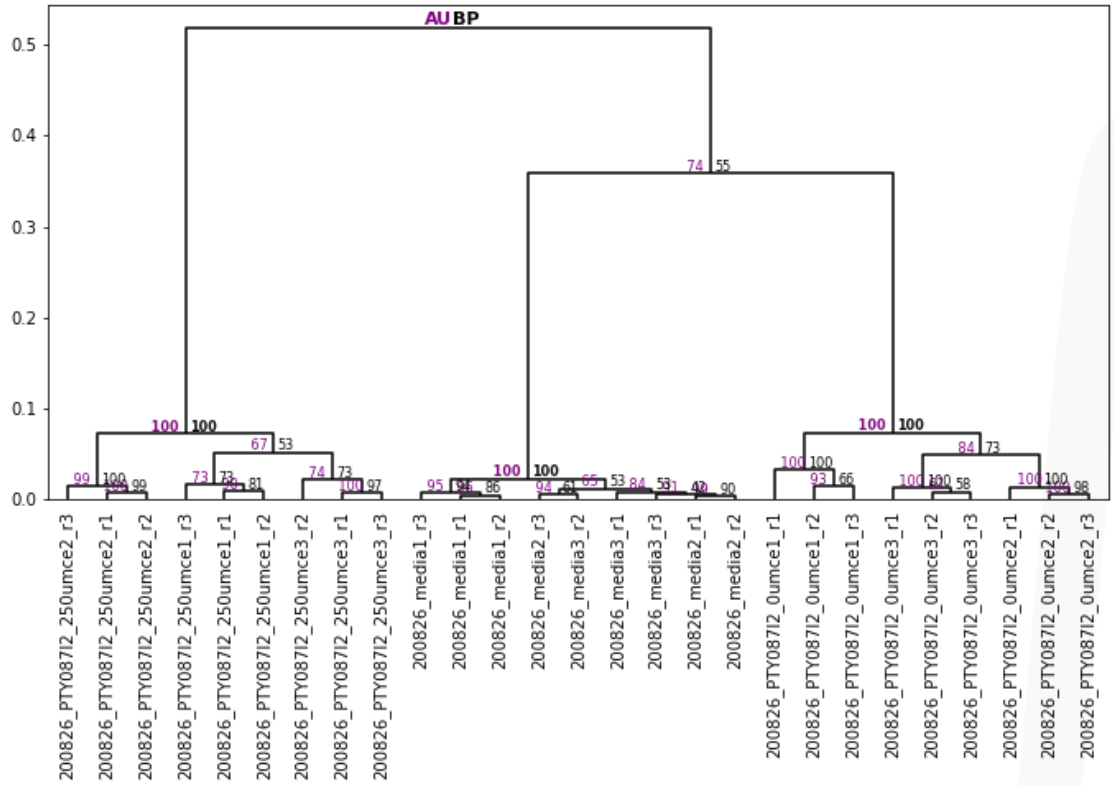
Spearman correlation analysis displays the pairwise Spearman correlations between each group and can be used to evaluate overall metabolomic similarities. The color scheme can be selected in the plot options dialog.



MPACT Spearman correlation matrix

Dendrogram

The Dendrogram tab displays the results of Hierarchal Clustering Analysis using Ward's method and can be used to evaluate both the metabolomic differences between samples and treatment groups, as well as the effects of filtering. Samples which cluster closely have high overall similarity. For datasets in which the differences between samples and treatment groups can reasonably be expected to be greater than technical uncertainty, technical replicates (reinjections of the same sample) should cluster after filtering. CV filter threshold can be adjusted within reasonable levels based on hierarchal clustering results such that the least stringent threshold before which diminishing returns in clustering quality are obtained, is selected. Bootstrap analysis can be turned on in the plot options dialog and approximately unbiased (AU) p-values and bootstrap probability (BP) will be annotated on the dendrogram based on the results of 1000 iterations.

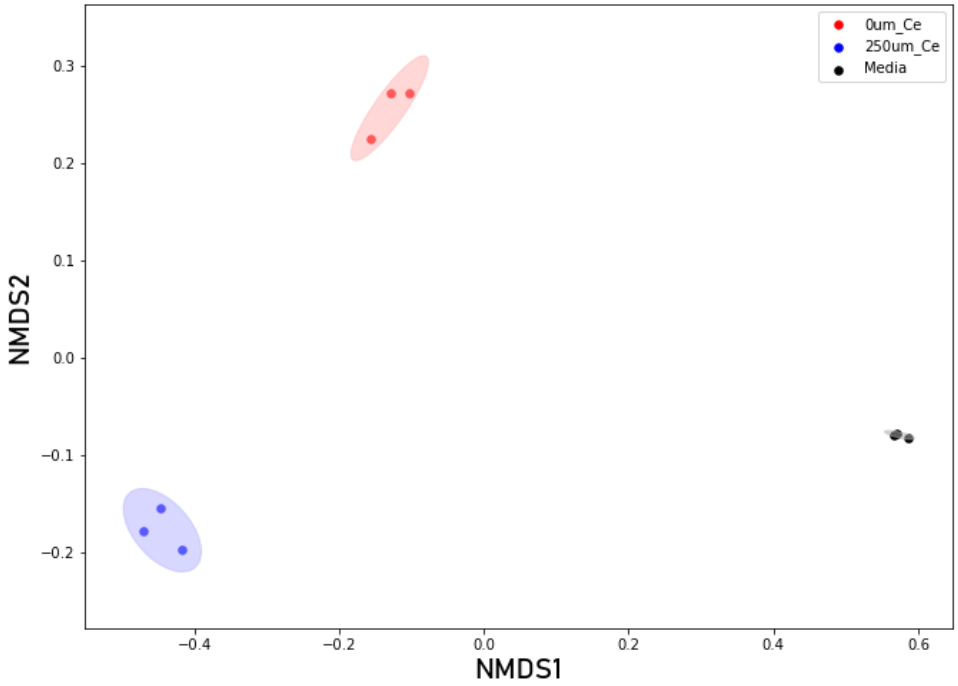


MPACT Dendrogram after filtering showing correct clustering of most technical replicate and all biological groups. Approximately Unbiased p-values (AU) greater than 95 are considered statistically significant

Multivariate Analysis

Multivariate analysis is conducted through generation of nonmetric multidimensional scaling (NMDS) analysis to reduce high dimensional metabolomics data to a low dimensional data which is easier to evaluate. Samples are plotted by biological group and 95% confidence

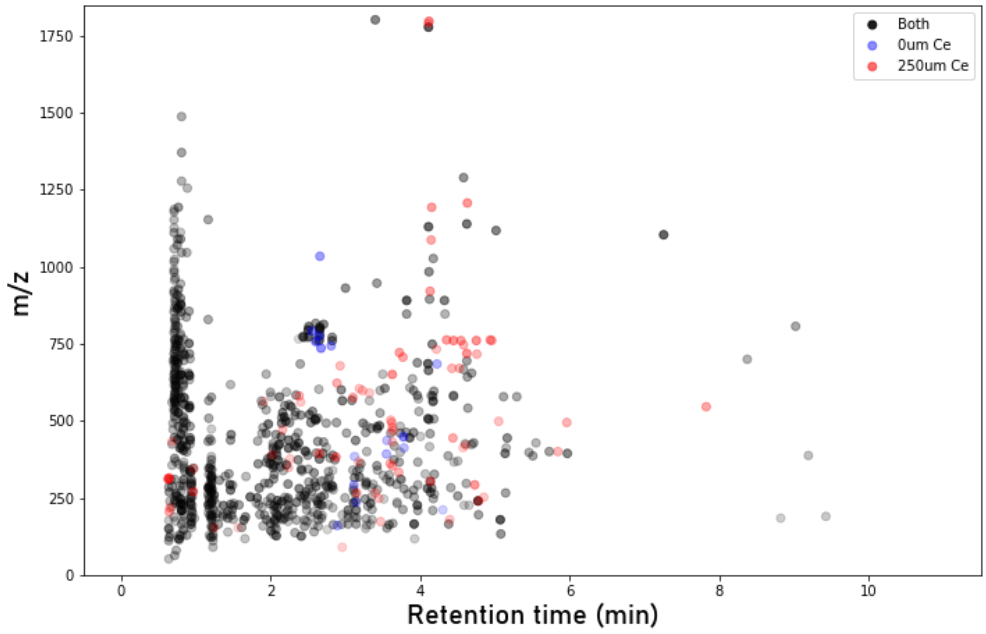
intervals are shown. Samples which are closer together in the plot are more similar in terms of their overall metabolome. Differentiation between groups of samples can be evaluated with multivariate analysis. As a default technical replicates are averaged so that NMDS is done on the sample-level, however to evaluate the grouping of technical replicates this can be unchecked by opening the plot options dialog by clicking the black square button in the bottom right plot toolbar.



MPACT NMDS plot with technical replicate averaging showing differences between samples and biological groups with shaded ovals denoting 95% confidence intervals

m/z vs. RT

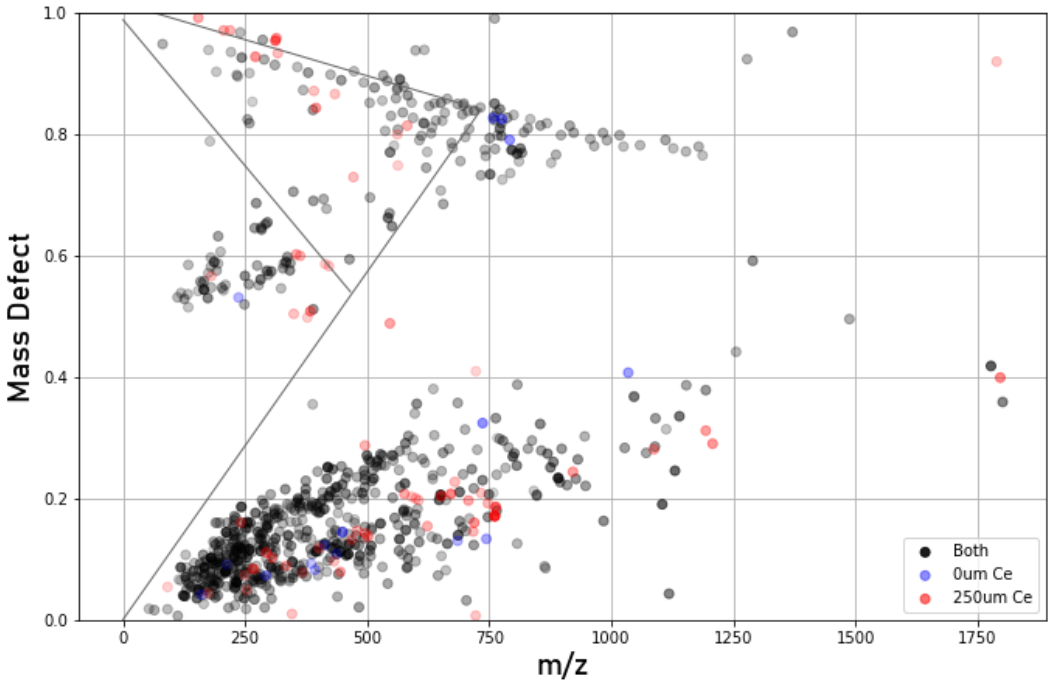
The m/z vs RT tab plots feature mass against retention time to give a chromatogram or isoplot like view of the data. This can be useful to evaluate if groups of similar mass or retention time (possibly indicating similar polarity) are unique to certain groups or experimental conditions



MPACT *m/z* vs RT plot showing features coloured by the parameters set by the user in the analysis settings tab. Point transparency is determined based on log normalized abundance with more abundant features being more solid.

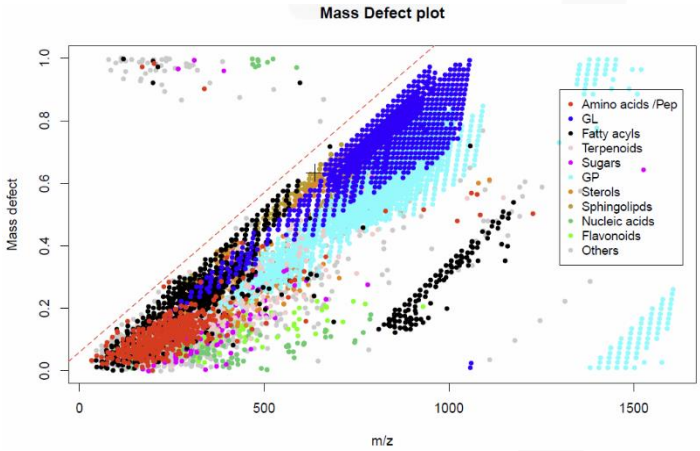
Mass Defect

Mass defect analysis highly effective for prioritization of features and identification of groups of closely related features. Mass defect in MPACT is specified as the m/z minus the nominal integer mass of a feature. Because shifts in mass defect and nominal mass are consistent for a given molecular transformation (e.g. hydrogenation, hydrolysis, methylation, etc.) families of similar compounds can be readily identified in some cases. Highly positive mass defect contributions are seen for hydrogen (+7.825 mDa mass defect per Da nominal mass) and very light elements, carbon has zero mass defect, and higher molecular weight atoms generally contribute increasingly negative mass defects (+.220, -0.317, -0.873 mDa mass defect per Da nominal mass for Nitrogen, Oxygen, and Sulfur, respectively), peaking at iron. As a result, the slope at which a feature appears in the mass defect plot can be used to predict general molecular features. For example, nucleotides and carbohydrates with large numbers of heteroatoms appear along trendline with a low slope. Similarly, highly unsaturated cyclic polyketides may also have low slopes due to hydrogen deficiency. Aliphatic hydrocarbons have maximum positive mass defect values at a given mass and therefore form a trendline with high slope. Fatty acids and lipids share this placement.



MPACT mass defect plot. Note the appearance of a trend of features with negative slope at the top of the chart, corresponding to salt clusters, as well as a series of low mass features in the region of improbability for singly charged species with a 0.5 Da shift from the main cluster of features, these features likely correspond to unresolved doubly charged dimer fragments. The main cluster of features is in the region where polyketide synthase (PKS) products and amino acid derivatives are found.

Note that nominal mass is based on the floor of an analyte's m/z so over/underflow is possible. For analytes rich in heteroatoms or halogens a very small negative mass defect is interpreted as a highly positive one. Many high molecular weight analytes have overall mass defect values greater than one, as such these mass defects will rollover to low values as is seen with fatty acids below. Guidelines for saturated hydrocarbon, perbromocarbon, and polycarbonic acid trendlines can be turned on and off in the plot options dialog. Singly charged features containing common CHNOPS formulae occur below the bottom hydrocarbon line, and above the top polycarbonic acid line. Compounds with extensive halogenation and very little hydrogen can occur between the

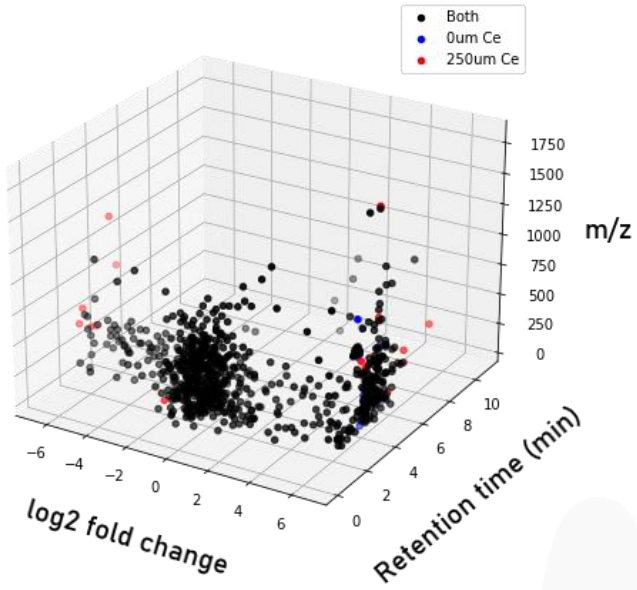


Distribution of RefMet metabolites on a mass defect plot (https://www.metabolomicsworkbench.org) showing the localization of different groups of metabolites.

polycarbonic acid and middle perbromocarbon line. Singly charged compounds with CHNOPSClI formulae cannot exist between the perbromocarbon and hydrocarbon line.

3D

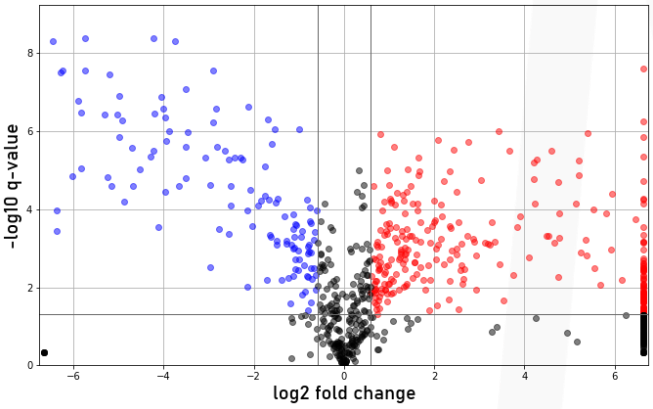
An interactive 3D plot of m/z vs RT against fold change in experimental vs controls group as specified in the analysis parameters tab can be used to further evaluate cluster of features which are of similar mass, polarity, and up or down regulation.



MPACT 3D fold change plot

Volcano Plot

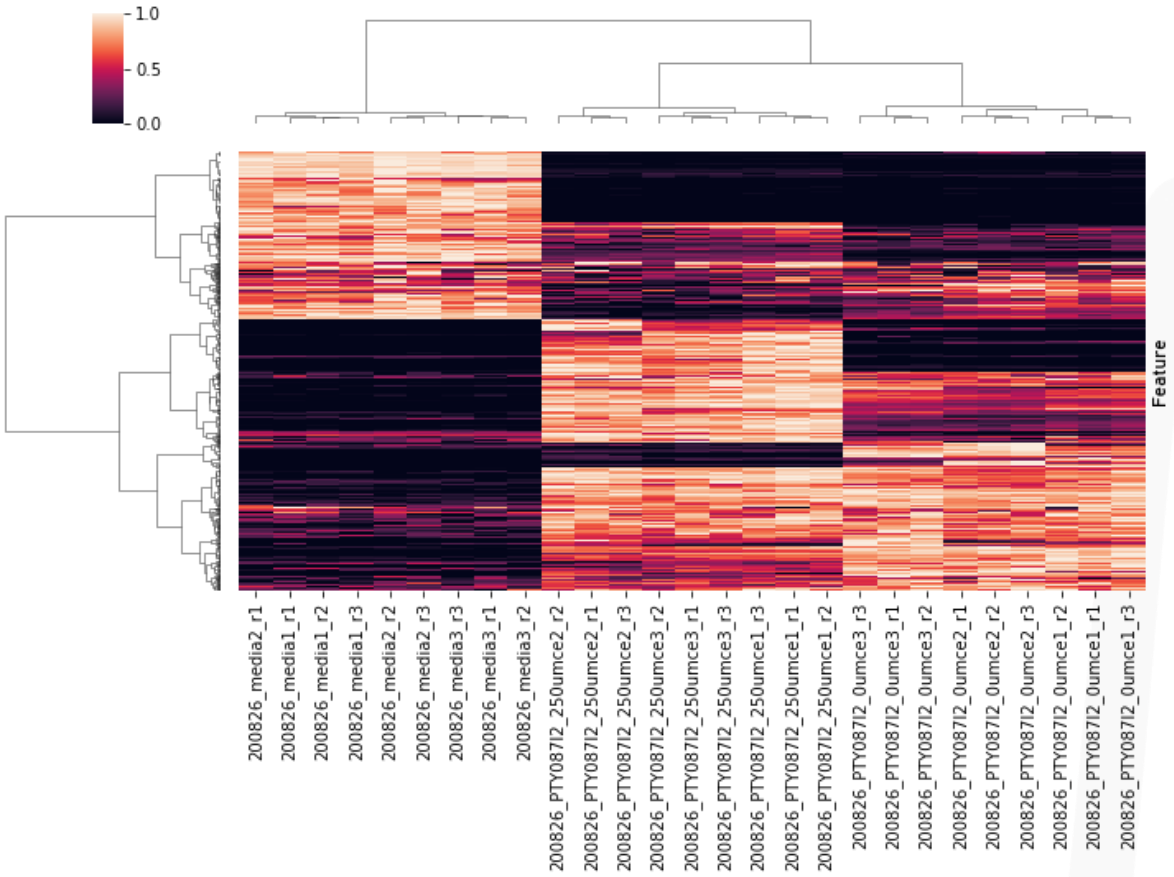
Volcano plots are generated by plotting log2 fold change in experimental vs control groups as specified in the analysis parameters tab against -log10(p/q value). If false discovery correction is selected in the analysis parameters tab q value (FDR-corrected p value is used) while raw p value is used if this option is left unchecked. P values are calculated based on error propagation of technical and biological uncertainty using a rooted sum of squares formula and by calculating effective degrees of freedom using the Welch–Satterthwaite equation. If only technical or biological replicates are available then error propagation is not performed and the appropriate technical or biological standard deviation and n are used for calculations. Options for changing the thresholds for significance and fold change can be accessed in the plot options dialog.



MPACT Volcano plot, features upregulated in the experimental group are shown in red and those downregulated are shown in blue

Heatmap

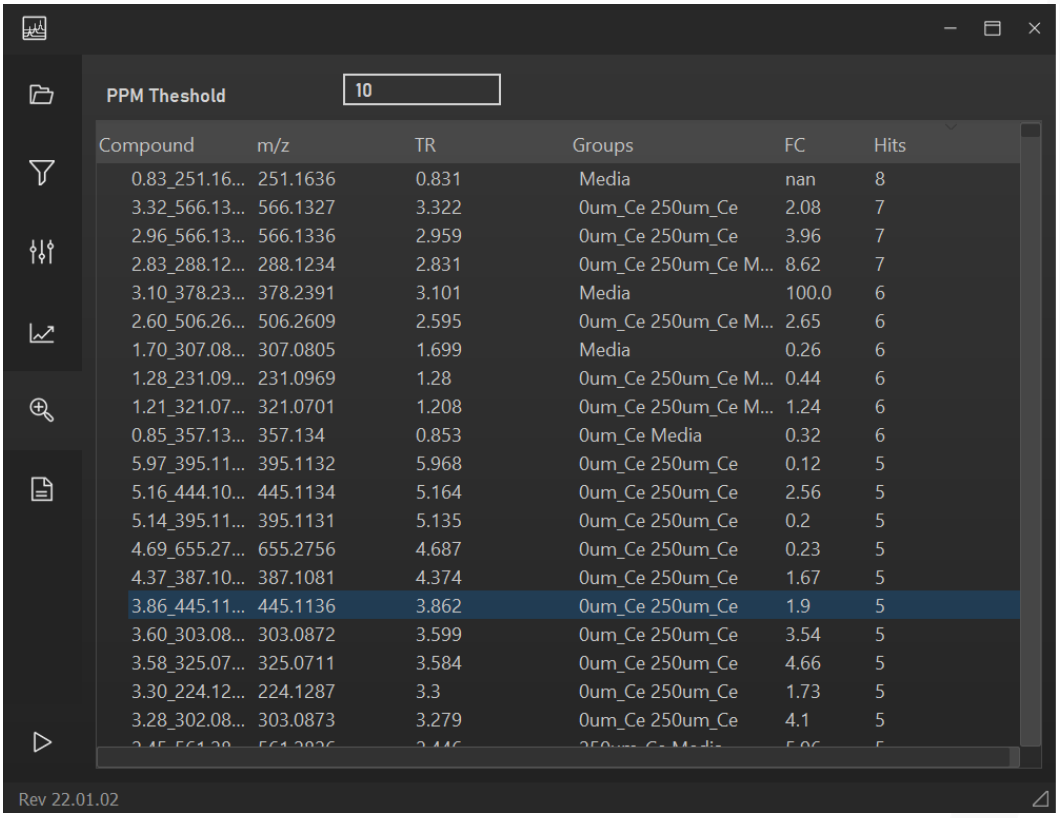
Heatmaps normalize raw abundance data within each feature between minimum and maximum abundance for that feature. Abundance of features is displayed in a heatmap grouped by samples in the x-axis vs features in the y-axis. The x-axis samples are clustered by overall metabolmic similarity using hierarchal clustering analysis as was the case with the dendrogram tab. Note clustering may be slightly different from the dendrogram tab due to normalized data being used. The y-axis features are clustered by the profile of upregulation for each feature across the samples. Thus, similar samples are clustered together and features which are up or downregulated in similar ways are clustered together. Heatmaps are useful for identifying groups of features for prioritization. Heatmap colour scheme can be selected in the plot options dialog.



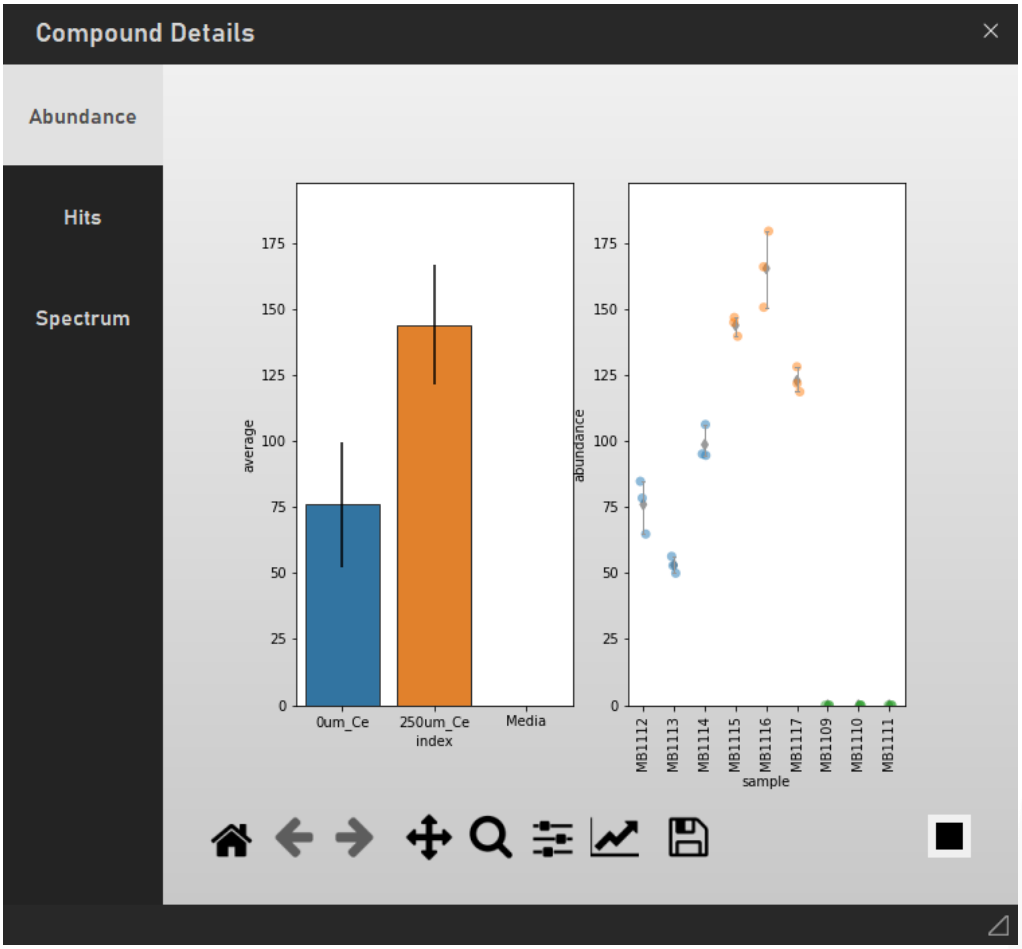
MPACT Heatmap grouping features samples by overall metabolomic similarity on the x axis and features by overall distribution similarity across samples on the y axis. Normalized abundance is denoted by cell colour with lighter values corresponding to higher abundance.

Feature Info

In any feature centric-plot the compound details window can be opened by clicking the "Details" button in the bottom right corner. This window is also opened automatically when switching to the feature info tab in the far-left pane. Features may be selected by clicking on them in the plots or heatmap, using the arrow keys or mouse to select a feature in the feature search tab, or by using the "W" and "S" keys to cycle up or down through features selectin the heatmap. Features may be unhighlighted by clicking them again. The compound details window allows users to investigate feature abundance and spectral data, database matches, and potential structures. If a fragment database file is provided the MS/MS spectra for a given feature may be viewed and a MASST export initiated to query the feature's spectrum against the GNPS database. The abundance tab displays abundance data at the group level in the left plot with the propagated uncertainty used in statistical calculations shown. Abundance data at the sample and replicate level is shown in the right chart with individual replicates indicated as points, sample average indicated by grey diamonds, and error bars denoting 95% confidence intervals within each sample. The hits tab displays compounds in the Natural Products Atlas database of microbial natural products that match the mass of sodiatiated or protiated adducts of the selected feature. Several matches maybe found and the structures can be moved through by clicking each match in the list or by using the arrow keys with the possible match list selected. The name, mass, ppm error, and taxonomic information are provided for each database feature. Support for other databases as well as other types of adduction is planned. Features are searched within the ppm range specified in the feature info tab.



MPACT feature info tab showing a selected feature and its associated information



MPACT compound details window showing combined uncertainty for measurement of the selected feature in biological groups (left) and technical uncertainty for the selected feature in individual samples (right)

The feature info tab displays important information on filtered features within the dataset including identified/feature name, m/z, retention time, the groups in which a feature was detected, fold change (FC) between experimental and control groups, any database matches found, as well as -logp/q values used in statistical testing and volcano plot generation. This table is selectable with the mouse and arrow keys which will update/highlight the selected feature in the plots, heatmap, and in the compound details window. The table may be sorted on a given column by clicking a column name. Updates to allow sorting or and easier filtering of this table are planned.

Analysis Info

The analysis info tab provides a summary of analysis in text form. This information is also saved in the output folder in the analysisinfo.txt file so that MPACT parameters used to process data may be viewed at a later time.

MPACT analysis info tab

Outputs

MPACT saves data in the selected output directory within a folder with the same name as the selected peak list. Several files are contained in this folder. Several processed files will have the peak list filename appended with different names to indicate what data they contain. Most files contain feature ID, m/z, and retention time as an index.

The *peaklistname*_formatted file contains the original peak list file formatted for MPACT processing. The *peaklistname*_merged contains the formatted data with mispicked peaks merged if this filter was enabled. The *peaklistname*_filtered file contains the formatted data after all filtering processes. The *peaklistname*_groupaverages file contains group average data for each feature in long format. The *peaklistname*_summarydata file contains all statistical data at the group level in wide format. For each feature the average abundance within each group is listed, as well as the biological and technical relative standard deviations and n, as well as

combined relative and absolute standard deviations obtained from error propagation, and effective n (neff) obtained from the Welch–Satterthwaite equation.

The iondict file contains all information in the feature_info table as well as other useful information for all features including those which did not pass filtering. Compound ID, m/z, retention time, mass defect (listed as kmd), which filters the feature passed. median CV are included as well. Statistical data and group presence/absence information is provided including the groups a feature was detected in, the maximum abundance it was found in a group (used to determine point opacity on plots), the log of this maximum abundance value, as well as fold change (fc), -logp, -logq, and the number of database matches found if a search was performed.

Additionally, an analysisinfo.txt file is generated containing basic statistics and filtering results as well as all necessary information to regenerate an analysis at a later date.