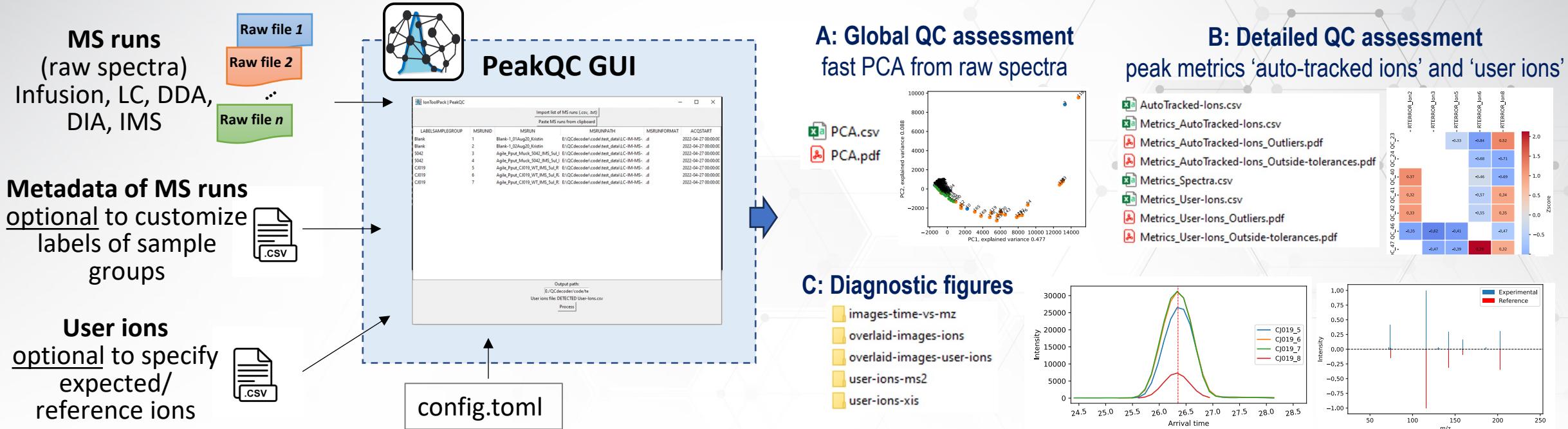


PeakQC: A Software Tool for Omics-Agnostic Automated Quality Control of Mass Spectrometry Data



Output files are organized in folders to provide both global and detailed QC assessments, and diagnostic figures.

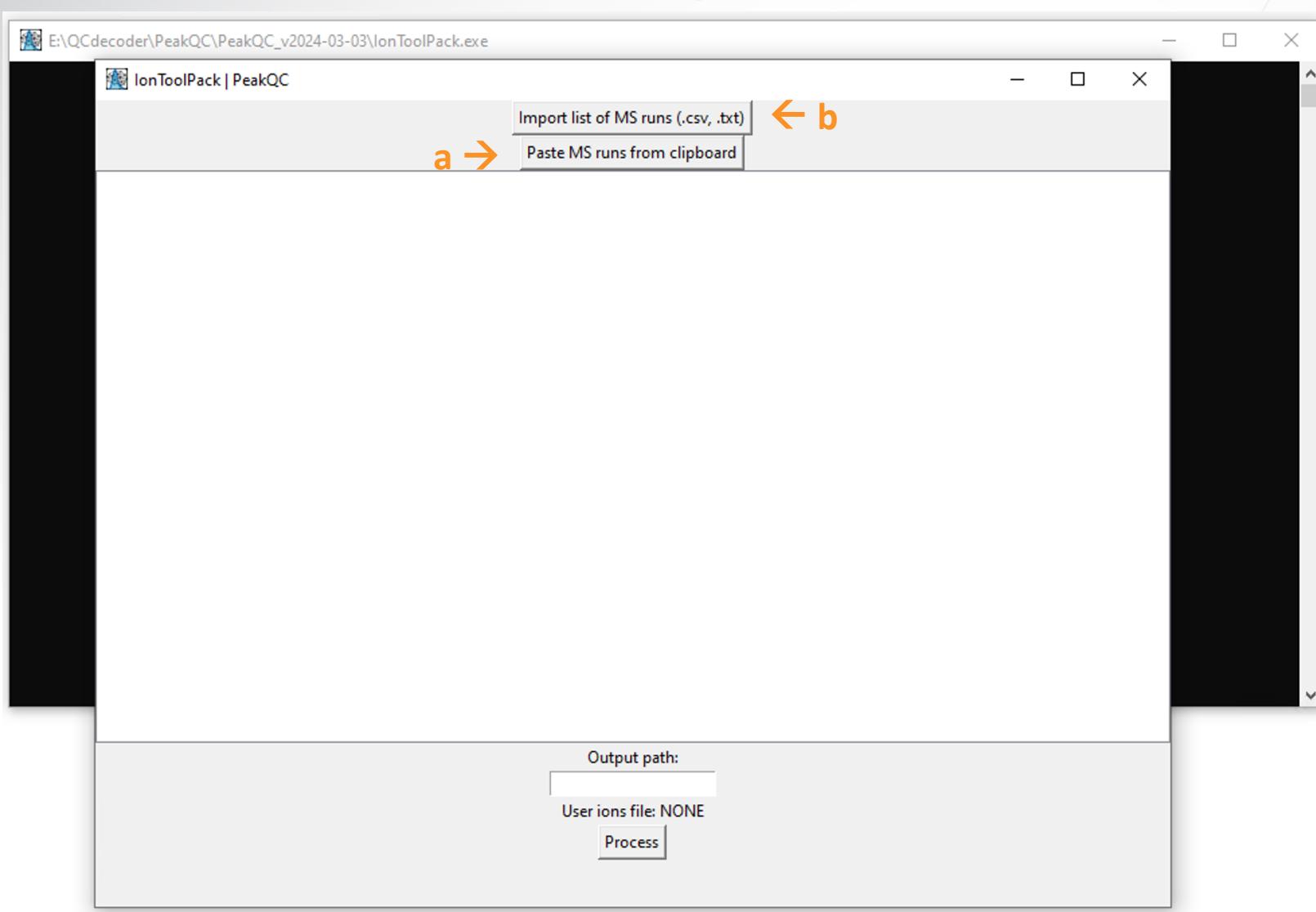
A: The PCA provides a first QC assessment to quickly spot compromised MS runs reflected by global similarity patterns in the MS1 data.

B: Auto-tracked ions are automatically detected and the values from the first MS run are taken as reference to calculate error metrics. User ions can be optionally specified by the user with expected values. Extracted peak metrics allow detailed evaluation reflected by local differences per dimension and per sample group (e.g., errors in mass, retention time, arrival time). Outliers and ions outside predefined tolerances (defined in the text file *config.toml*) are detected and reported in tables (CSV) and heatmaps (PDF). Blanks are ignored. Ions outside tolerances are only detected for QC groups. The heatmaps only display the defective MS runs and corresponding defective metrics, i.e., the larger the heatmap the more issues detected: more rows indicate more defective runs, and more columns indicate more defective ions and metrics.

C: Diagnostic figures are generated with overlaid extracted peak signals (e.g., XICs) and MS2 mirror plots, which the user can look at to further investigate issues

Questions? Suggestions? Software issues? Contact: aivett.bilbao@pnnl.gov

GUI and raw data import



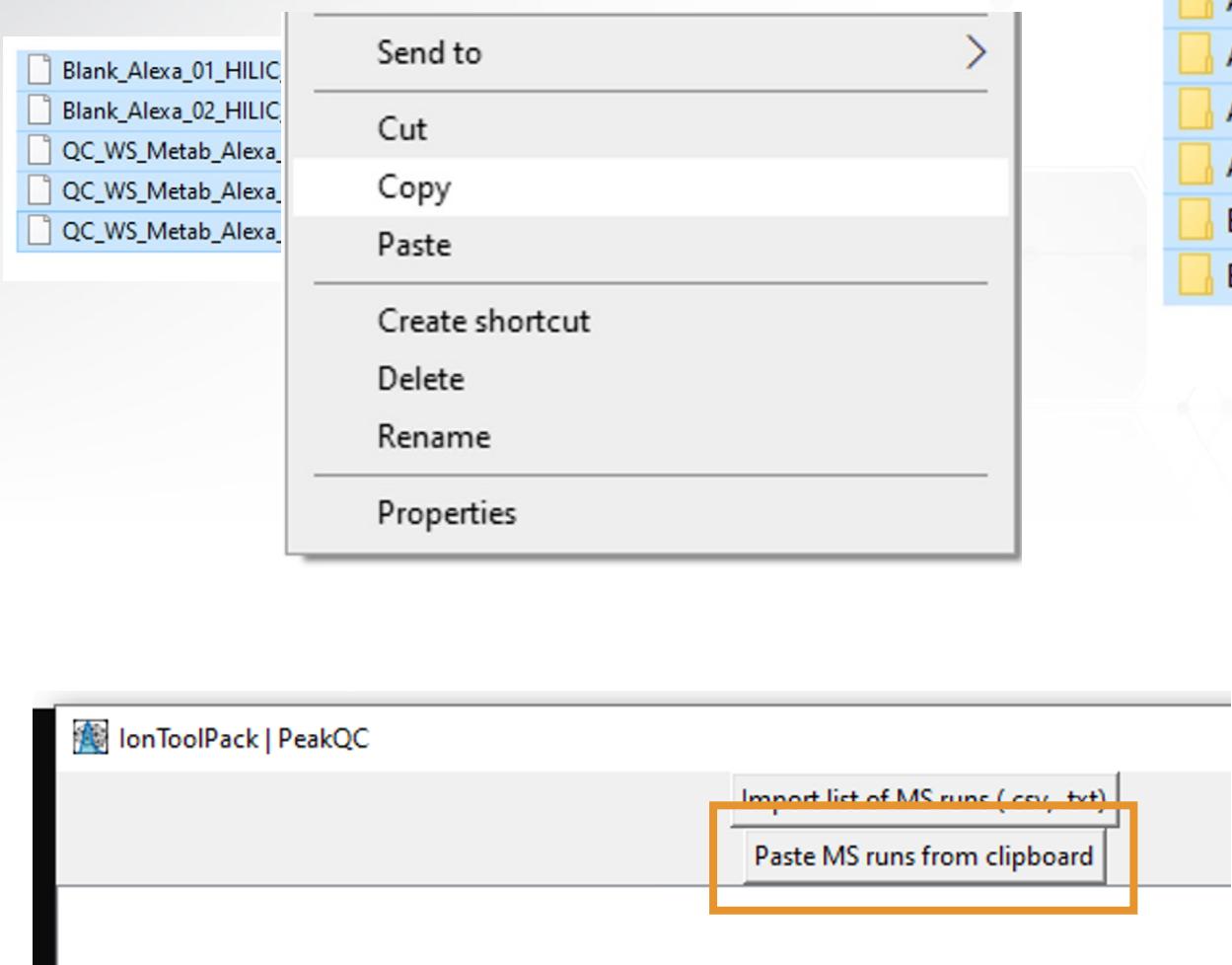
- 1) Double click IonToolPack.exe to open the GUI
- 2) Load MS runs (raw data files):
 - a. **Copy from the Windows File Explorer, and click “Paste MS runs ...” or**
 - b. **Import from a CSV**
- 3) Click “Process”

* Avoid copying the tool in the OneDrive folder in your computer

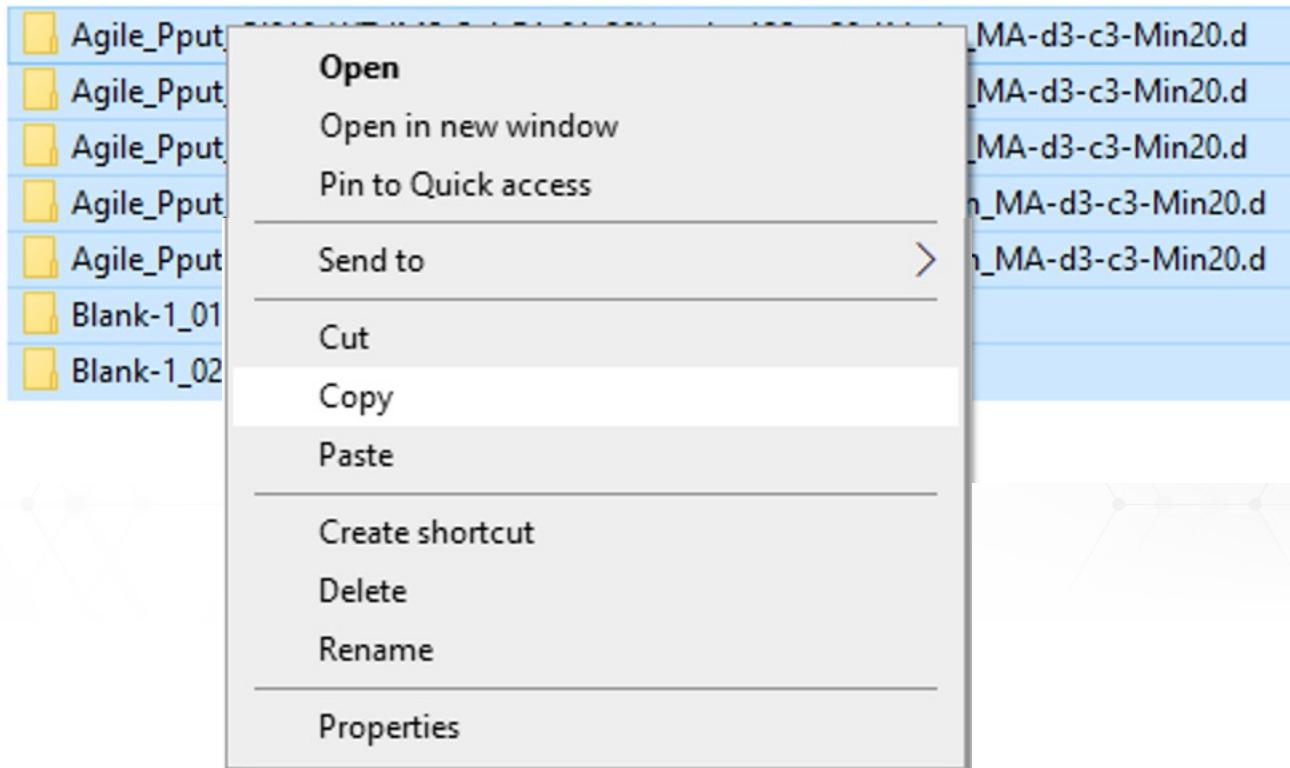
* To re-run, delete csv files in ResultsQC folder: PCA.csv, Metrics_AutoTracked-Ions.csv, Metrics_User-Ions.csv

2a - Load MS runs - Copy&Paste

Example Thermo .raw or mza files



Example Agilent .d folders



1. In the Windows File Explorer, select your MS runs (.raw, .mza, .mzML files, or Agilent .d folder)
2. Right click to display the menu and select “Copy”
3. In the PeakQC window, click the button “Paste MS runs from clipboard”

2b - Load MS runs from a CSV – Customize sample groups and MS data location

MS-runs.csv or dataset.txt: a CSV file with a list of your raw files (or .txt tab separated).

Mandatory columns:

MSRUN (or dataset)

*extension can be optionally provided, it will be tried with priority: .mza, .d, .raw, .mzml)

MSRUNPATH (or dataset_folder_path)

*if not found as absolute path, a relative path using as root the path of MS-runs.csv will be tried

- Lower or upper case in column names, and “_” or spaces are detected and converted to upper case.
- Additional columns are ignored.
- For PNNL: You can download this file from the list of datasets in DMS, selecting “Tab-Delimited Text” at the bottom of the page (it will download a file “dataset.txt” or “data_package_dataset_files.txt”)
- Column names are replaced in GUI:
dataset=MSRUN and dataset_folder_path=MSRUNPATH

dataset	dataset_folder_path
Blank_Alexa_01_HILIC_NEG_13Feb23_Olympic_WBEH-8671_r1	test_data\LC-MS-Thermo\RawData
Blank_Alexa_02_HILIC_NEG_13Feb23_Olympic_WBEH-8671	test_data\LC-MS-Thermo\RawData
QC_WS_Metab_Alexa_01_HILIC_NEG_13Feb23_Olympic_WBEH-{	test_data\LC-MS-Thermo\RawData
QC_WS_Metab_Alexa_02_HILIC_NEG_13Feb23_Olympic_WBEH-{	test_data\LC-MS-Thermo\RawData
QC_WS_Metab_Alexa_03_HILIC_NEG_13Feb23_Olympic_WBEH-{	test_data\LC-MS-Thermo\RawData

Optional columns, values will be automatically assigned if not found:

LABELSAMPLEGROUP: “Blank” or “QC” if found in MSRUN, or “Sample” otherwise

MSRUNID: sequentially increasing, also assigned if values are not unique across rows

ACQSTART: if provided rows will be sorted by this value, otherwise original order is kept

LABELSAMPLEGROUP	MSRUNID	MSRUN	MSRUNPATH	MSRUNFORMAT	ACQSTART
Blank	1	Blank-1_01Aug20_Kristin	E:\QCdecoder\code\test_data\LC-IM-MS-.d		2022-04-27 00:00:00
Blank	2	Blank-1_02Aug20_Kristin	E:\QCdecoder\code\test_data\LC-IM-MS-.d		2022-04-27 00:00:00
5042	3	Agile_Pput_Muck_5042_IMS_5ul_I	E:\QCdecoder\code\test_data\LC-IM-MS-.d		2022-04-27 00:00:00
5042	4	Agile_Pput_Muck_5042_IMS_5ul_I	E:\QCdecoder\code\test_data\LC-IM-MS-.d		2022-04-27 00:00:00
CJ019	5	Agile_Pput_CJ019_WT_IMS_5ul_R	E:\QCdecoder\code\test_data\LC-IM-MS-.d		2022-04-27 00:00:00
CJ019	6	Agile_Pput_CJ019_WT_IMS_5ul_R	E:\QCdecoder\code\test_data\LC-IM-MS-.d		2022-04-27 00:00:00
CJ019	7	Agile_Pput_CJ019_WT_IMS_5ul_R	E:\QCdecoder\code\test_data\LC-IM-MS-.d		2022-04-27 00:00:00

Output path: E:/QCdecoder/code/te
User ions file: DETECTED User-Ions.csv
Process

Output path will be assigned by default after you select the MS-runs.csv/dataset.txt file, with the same path where the selected file is

Output

📁 images-time-vs-mz
📁 overlaid-images-ions
📁 overlaid-images-user-ions
📁 user-ions-ms2
📁 user-ions-xis
📄 AutoTracked-Ions.csv
📄 Metrics_AutoTracked-Ions.csv
📄 Metrics_AutoTracked-Ions_Outliers.csv
📄 Metrics_AutoTracked-Ions_Outliers.pdf
📄 Metrics_AutoTracked-Ions_Outside-tolerances.csv
📄 Metrics_AutoTracked-Ions_Outside-tolerances.pdf
📄 Metrics_Spectra.csv
📄 Metrics_User-Ions.csv
📄 Metrics_User-Ions_Outliers.csv
📄 Metrics_User-Ions_Outliers.pdf
📄 Metrics_User-Ions_Outside-tolerances.csv
📄 Metrics_User-Ions_Outside-tolerances.pdf
📄 PCA.csv
📄 PCA.pdf

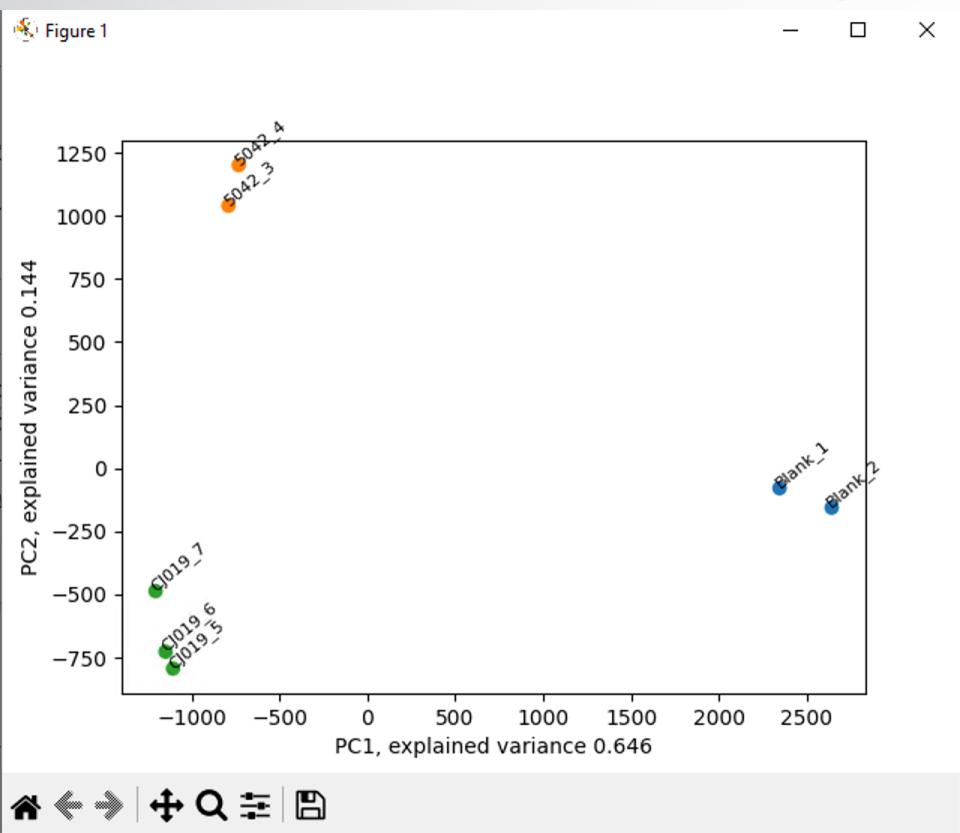
- The folder **ResultsQC** is created either in a parent folder: the folder from where the MS runs were copy&pasted (option 2a) or the folder where the imported MS-runs.csv is (option 2b). A customized config.toml file could be placed in this parent folder (default in the exe folder).
- These subfolders and files are automatically created inside ResultsQC

Created only if the User-Ions.csv file is detected

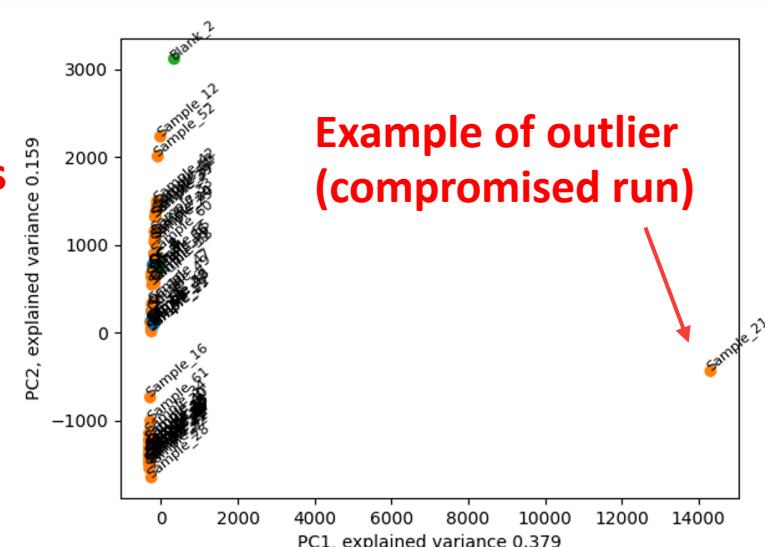
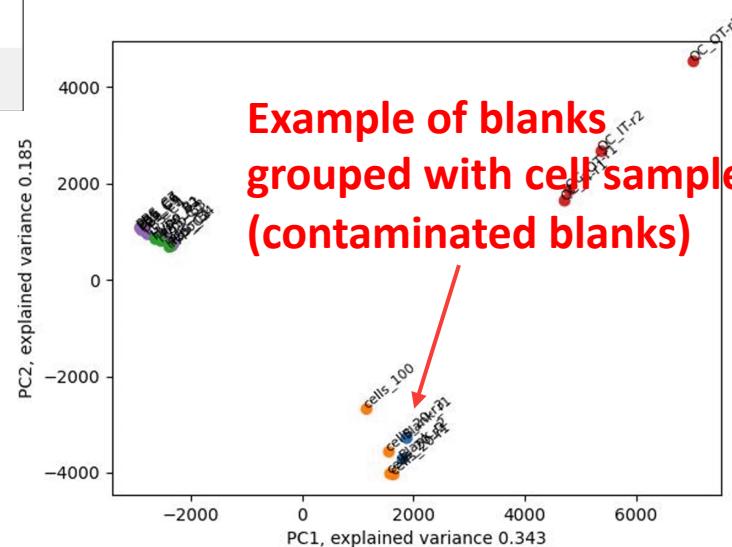
Created only for ion mobility data

PCA

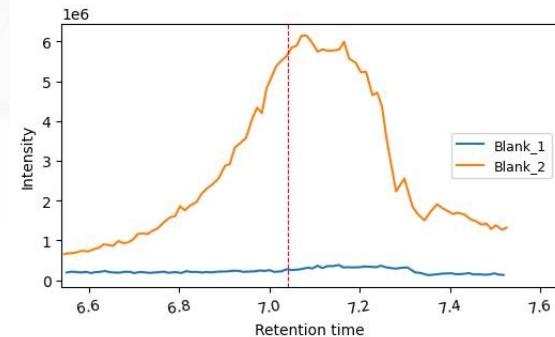
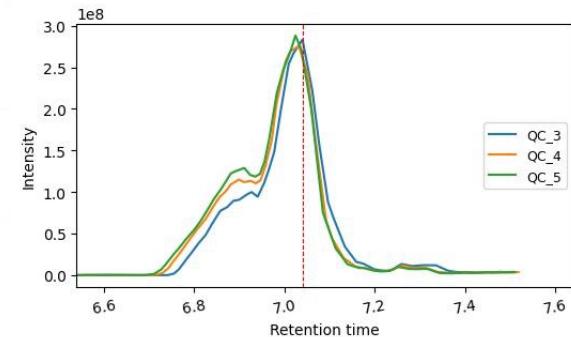
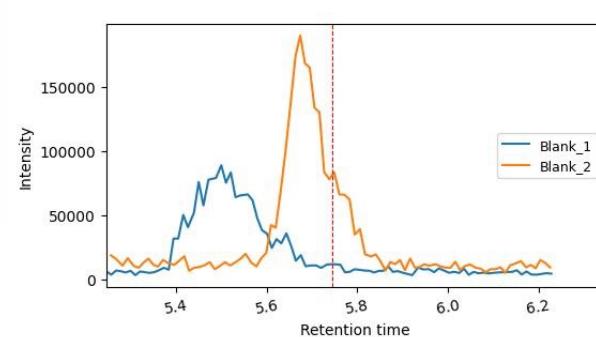
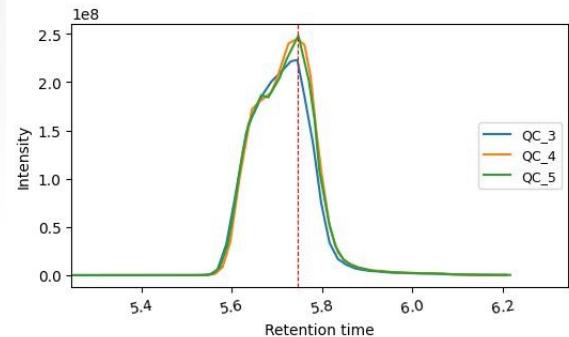
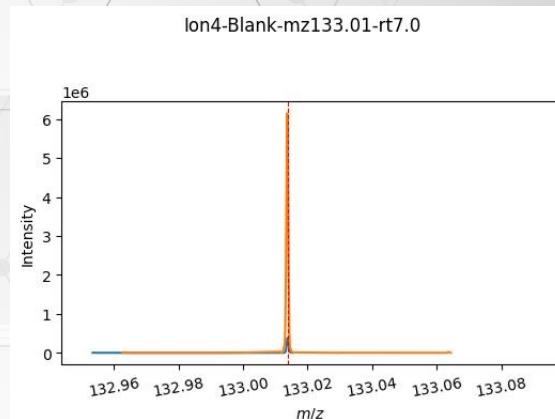
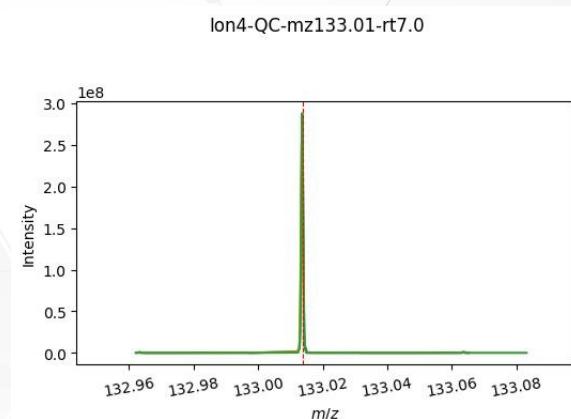
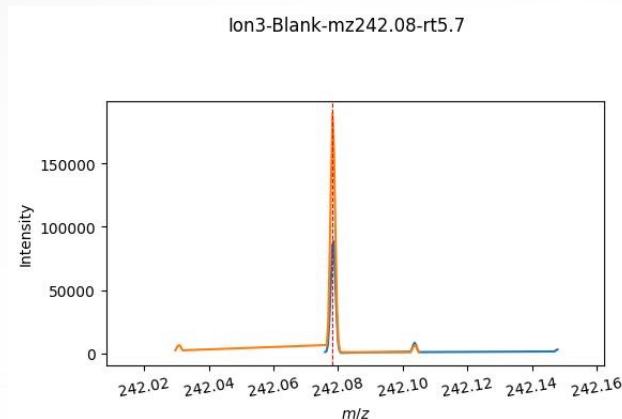
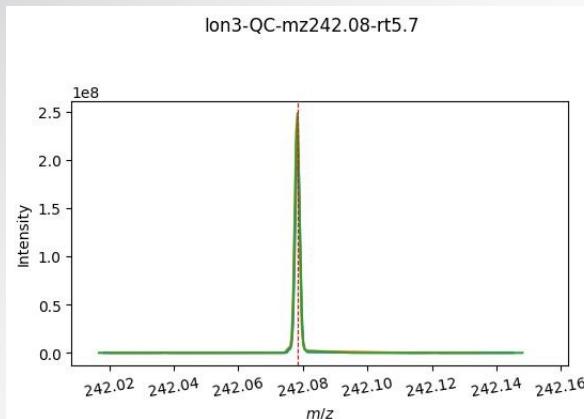
Figure 1



- Samples are colored by groups (automatically assigned or user specified in the file MS-runs.csv)
- Provides a first QC assessment to quickly spot samples that are not clustered together within expected groups (e.g., blanks, QCs, treatments, etc.), which represent outliers or compromised runs
- Saved in PCA.pdf and the window is interactive (possible to zoom in)



Folder overlaid-images-ions – Auto tracked ions



- The topmost abundant and frequent ions are automatically detected
 - Figures are generated for each ion and overlaid by sample group
- 1st panel: MS1 spectrum at RT apex
- 2nd panel: Extracted ion chromatogram (XIC, max intensity within m/z tolerance)

User-Ions.csv: input list with your expected ions

- The file must be named **User-Ions.csv**
- The file must be placed in the same folder as **MS-runs.csv** (or from where you copy your raw files) and it will be automatically detected
- Mandatory columns: **MOLECULE, MZ, RT**

MOLECULE	MZ	RT	AT	FRAGSMZ	FRAGSINTENSITY
isocitric acid	191.0197	3.6	15.36	111.0096;67.0186;	2248845;178494;73903;
succinic acid	117.0193	1.8	13.52	73.0282;99.0076	16199;339
aspartic acid	132.0302	2.1	13.97	71.0151;72.0073;8	90304;16391;337856
glutamic acid	146.0459	2	14.71	102.0529;128.0329	42538;6760
UDP-glucose	565.0477	3	26.11	211.001;241.0119;	14027;143706;3816;337

- **AT:** Arrival time for ion mobility spectrometry data
- **FRAGSMZ, FRAGSINTENSITY:** fragment ions (MS/MS) represented as 2 lists for *m/z* and corresponding intensity for each peak. Values must be separated by ":".
- Additional columns are ignored
- Avoid special characters and "," in MOLECULE

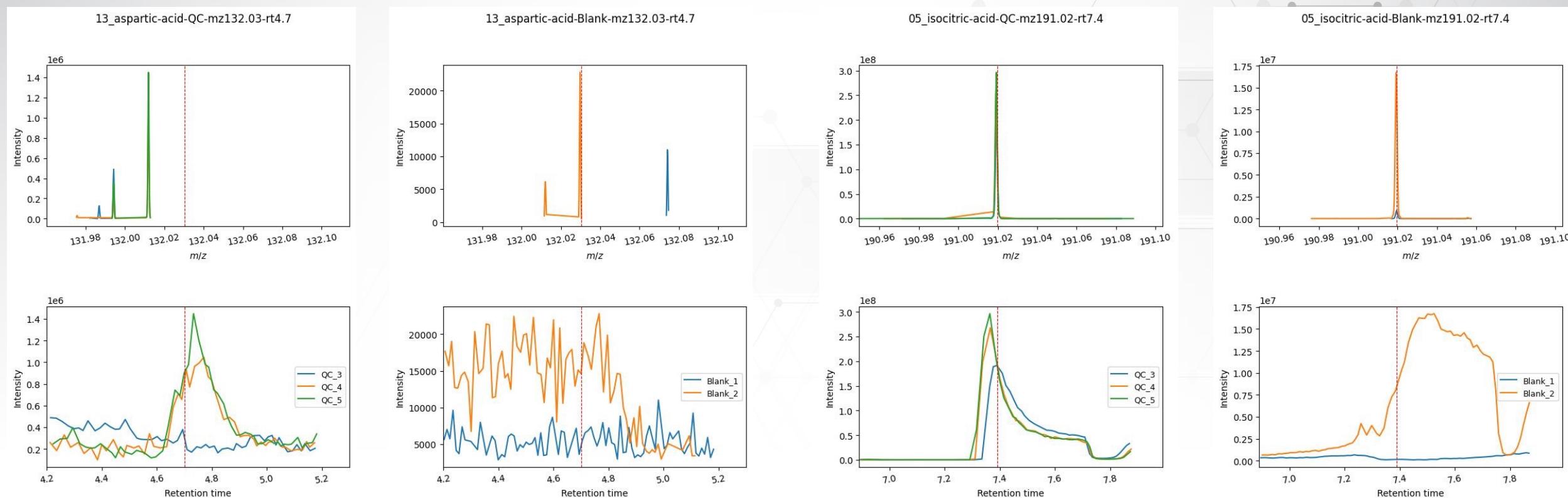
User-Ions.csv: input list with your expected ions – other optional columns

Visualization and extraction windows can be optionally specified:

- MZVIEWHALFWINDOW: half window for m/z plot, default 0.075
- RTVIEWHALFWINDOW: half window for m/z plot, default 0.3
- MZXICHALFWINDOW: half window or *m/z* tolerance for signal extraction, default 0.05
- ATVIEWHALFWINDOW: half window for m/z plot, default 1.5

MOLECULE	MZ	RT	AT	FRAGSMZ	FRAGSINTENSITY	MZVIEWHALFWINDOW	RTVIEWHALFWINDOW	MZXICHALFWINDOW	ATVIEWHALFWINDOW
isocitric acid	191.0197	3.6	15.36	111.0096;67.0186; 2248845;178494;73903!		1	2	0.025	5
succinic acid	117.0193	1.8	13.52	73.0282;99.0076	16199;339	1	2	0.025	5
aspartic acid	132.0302	2.1	13.97	71.0151;72.0073;8	90304;16391;337856	1	2	0.025	5
glutamic acid	146.0459	2	14.71	102.0529;128.0325	42538;6760	1	2	0.025	5
UDP-glucose	565.0477	3	26.11	211.001;241.0119;	14027;143706;3816;337	1	2	0.025	5

Folder overlaid-images-user-ions

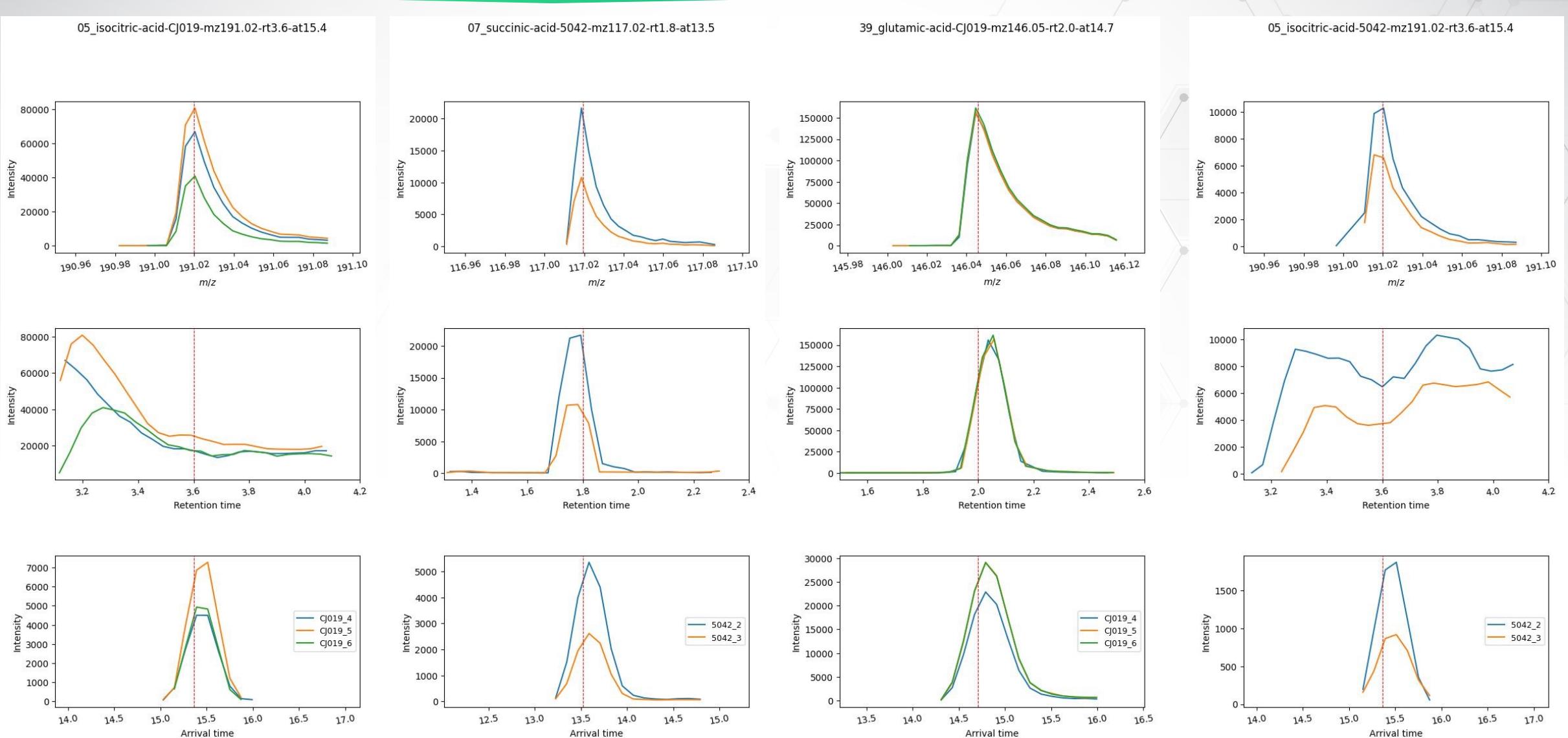


Figures overlaid by sample group are generated for each theoretical or reference ion provided in the file **User-Ions.csv**

1st panel: MS1 spectrum at RT apex

2nd panel: Extracted ion chromatogram (XIC, MS1 max intensity within m/z window)

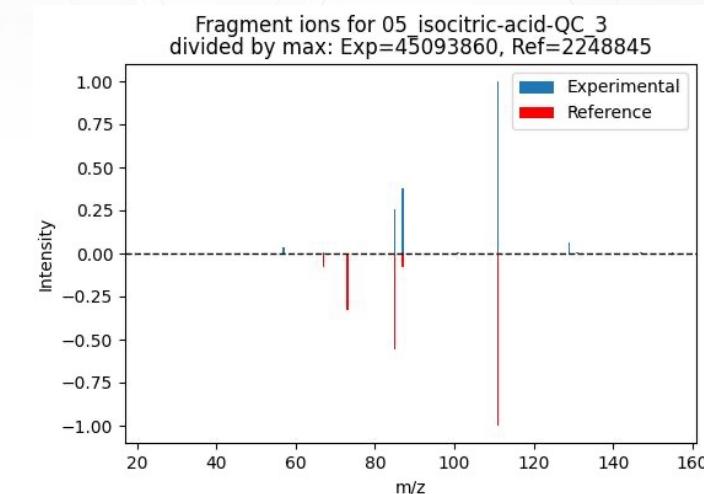
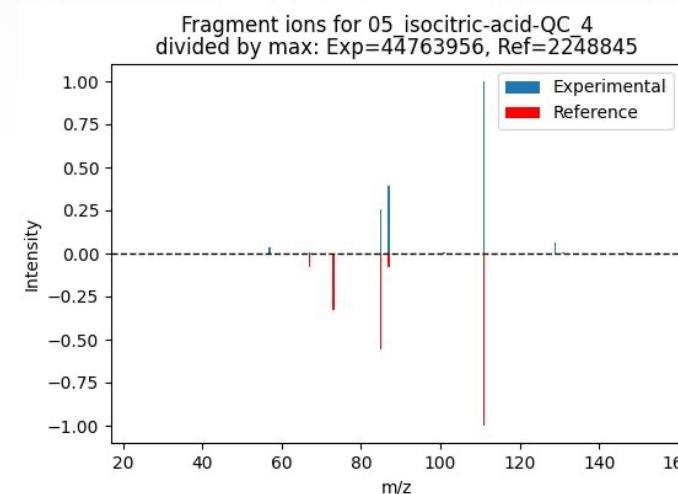
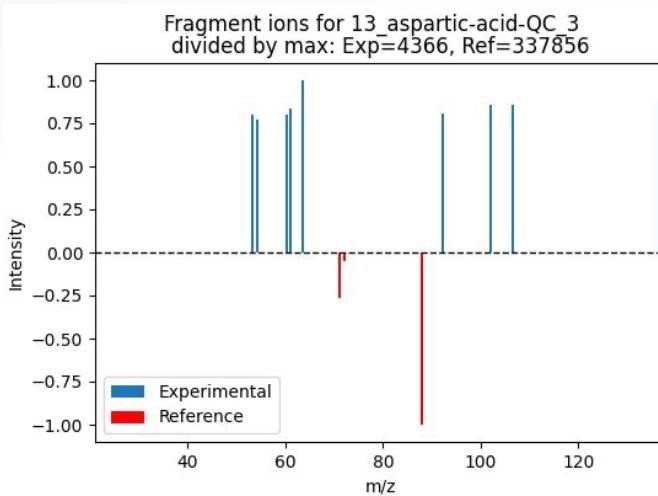
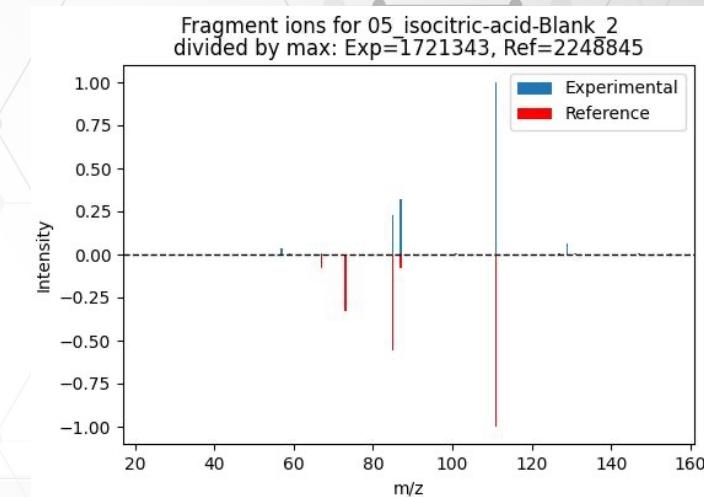
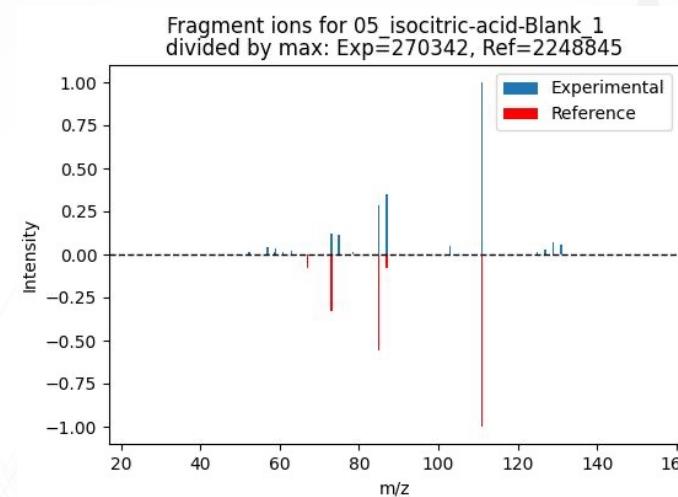
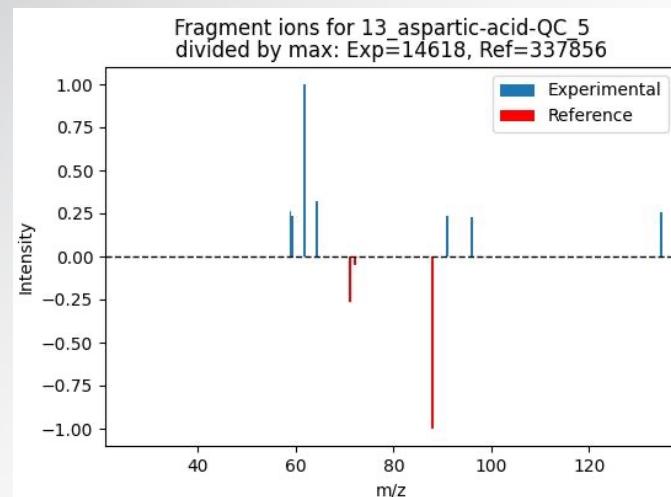
Folder overlaid-images-user-ions – for ion mobility spectrometry data



3rd panel: extracted ion ‘mobiligram’ (ion mobility trace) if it is IMS data and the AT is provided in **User-Ions.csv**

* Frame from RT apex (max intensity within m/z window)

Folder user-ions-ms2 – MS2 view for DDA

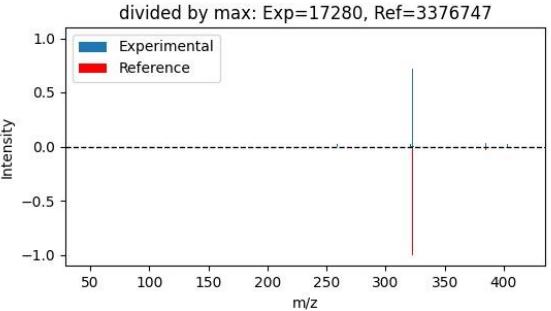


A figure is created for each user ion and MS run if it is DDA spectra and the fragment ions (**FRAGSMZ**, **FRAGSINTENSITY**) are provided in **User-Ions.csv**

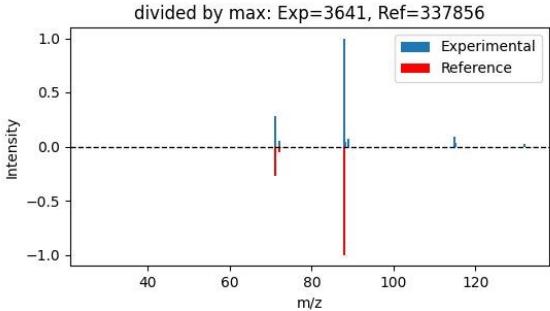
* MS2 spectrum with closest RT to apex and with precursor *m/z* within window

Folder user-ions-ms2 – MS2 view for DIA and IMS data

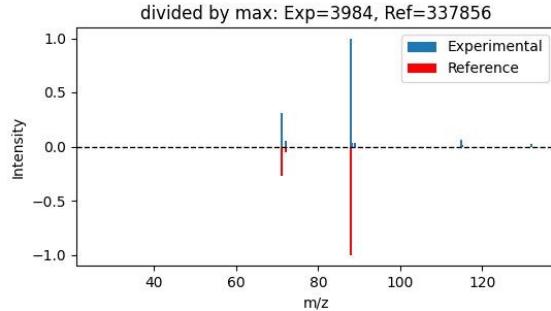
Fragment ions for 41_UDP-glucose-5042_3



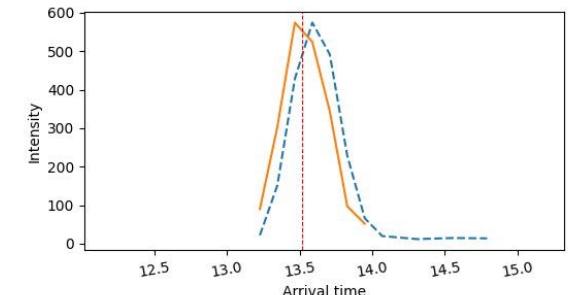
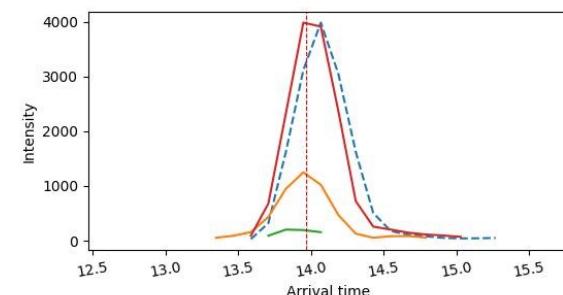
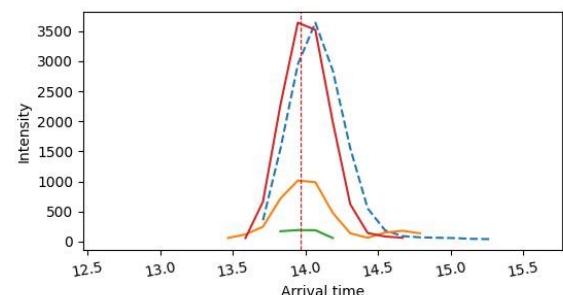
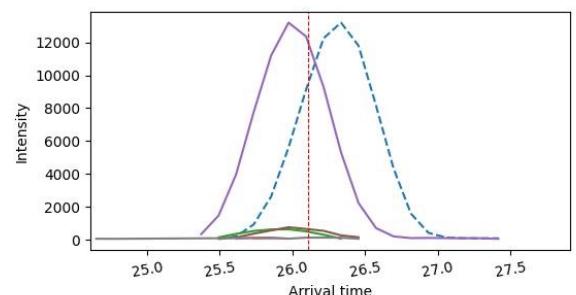
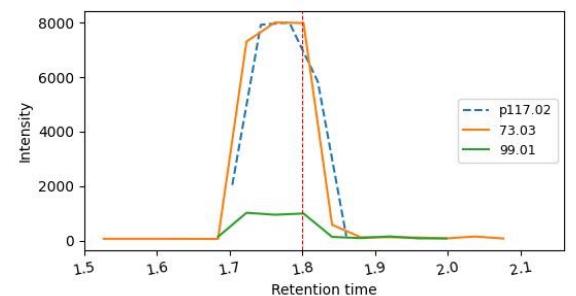
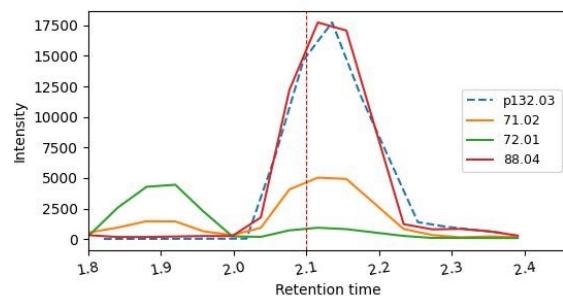
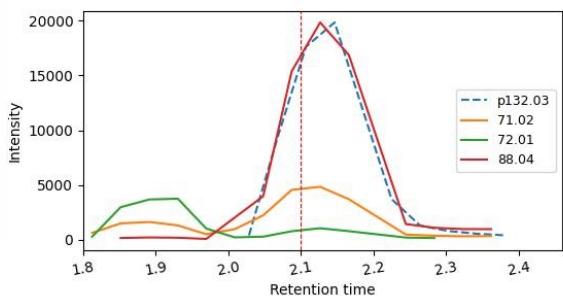
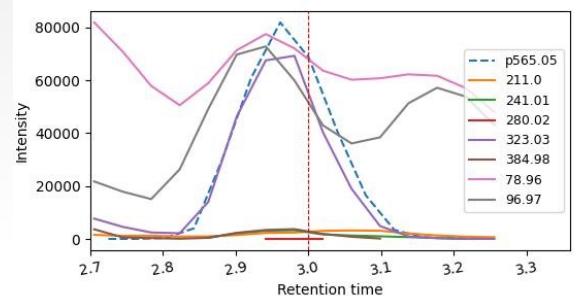
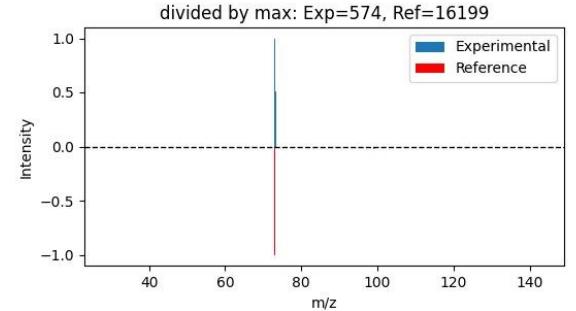
Fragment ions for 13_aspartic-acid-5042_2



Fragment ions for 13_aspartic-acid-5042_3



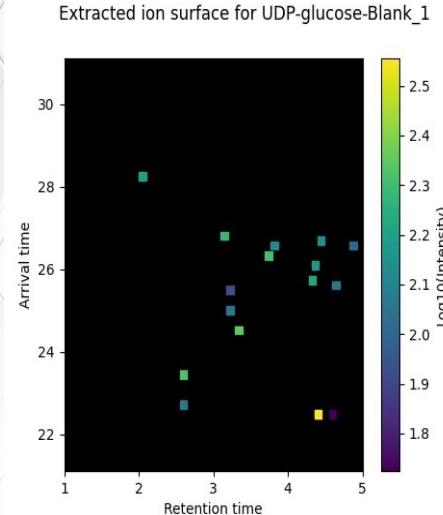
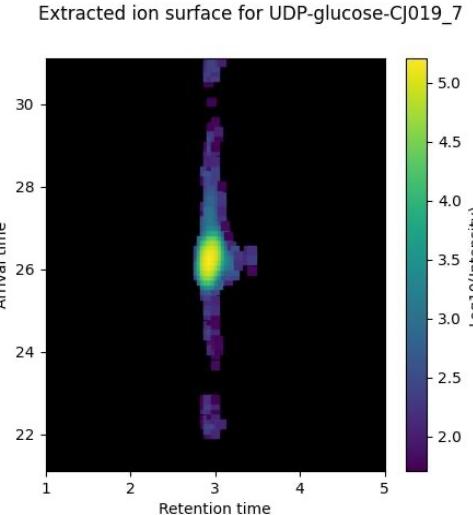
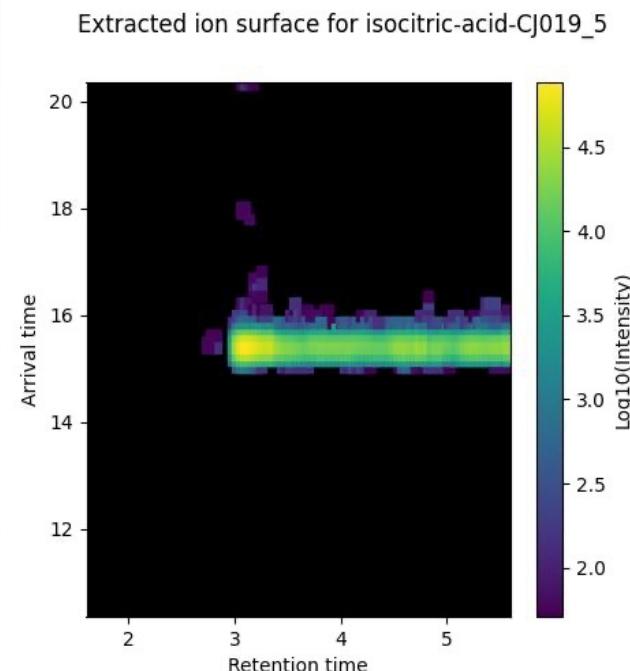
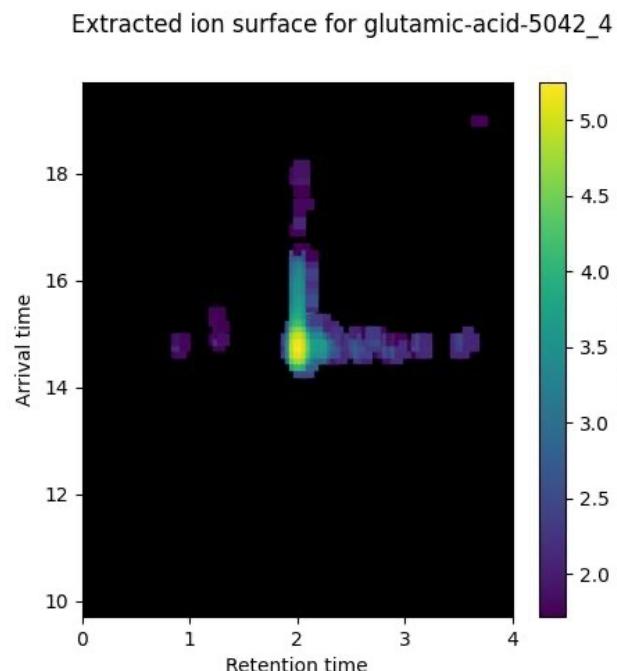
Fragment ions for 07_succinic-acid-5042_3



2nd and 3rd panel: created with the XICs and XIMs for fragment ions if it is DIA-MS and IM-DIA-MS data, respectively

* Precursor ion (MS1) shown in dashed line and its Intensity normalized to max fragment ion intensity

Folder user-ions-xis – for ion mobility spectrometry data



An “extracted ion surface” (XIS) figure, which is a heatmap with intensity by RT and AT, is created for each user ion and MS run if it is IMS data and the **AT** is provided in **User-Ions.csv**

* Intensity is the MS1 summed intensities within *m/z* window

* The minimum values currently supported are ATVIEWHALFWIDH=3 and =2