

# MSCohort Manual



Version. 202405

### ### Computer configuration

**CPU:** Intel or AMD processor with 64-bit support; 2.3 GHz or faster processor with at least 2 cores is recommended

**RAM:** 16G or higher is recommended

**ROM:** for one raw data (1G) 2G or higher is recommended

**OS:** Windows 10 (x64) or Windows 11 (x64)

**Other:** At present, Spectronaut identification and quantification results can be used for quality analysis by MSCohort.



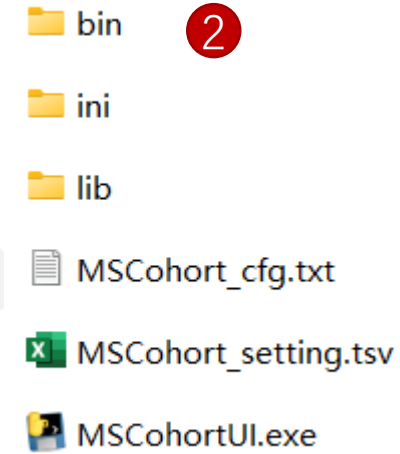
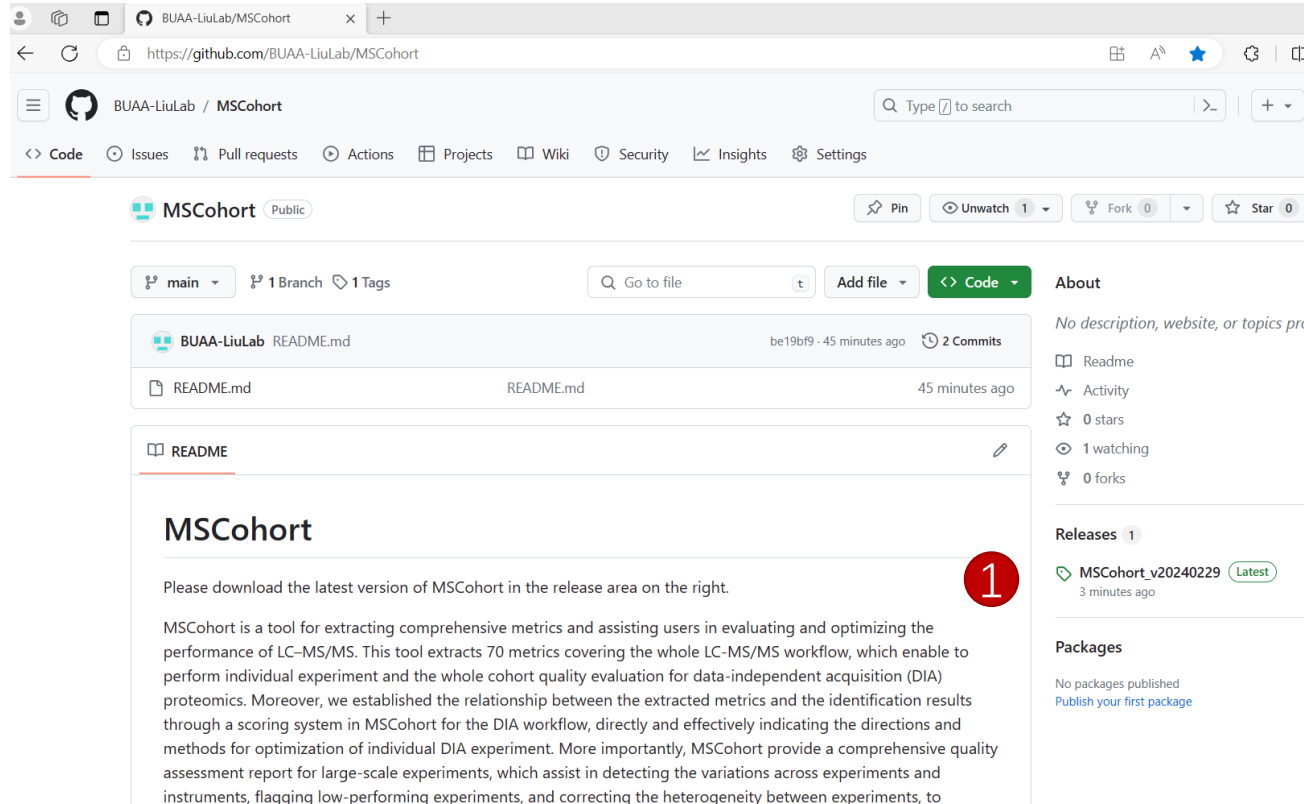
# CONTENTS

- [MSCohort download](#)
- [Spectronaut custom report for MSCohort](#)
- [MSCohort manual for intra-experiment analysis](#)
- [MSCohort manual for inter-experiment analysis](#)
- [Notes for modifying the scoring criteria](#)

# **MSCohort download**



# 1. MSCohort download



## ① Login

<https://github.com/BUAA-LiuLab/MS Cohort>

and download the latest version at the right of the page.

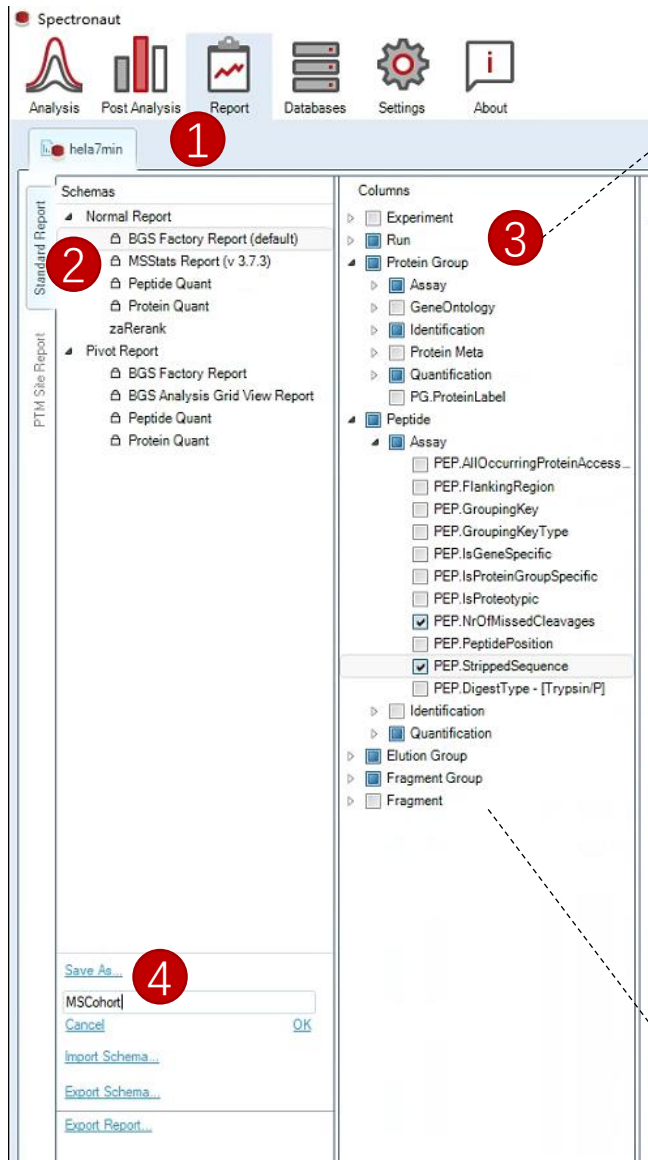
## ② The unzipped MSCohort file.

# **Spectronaut customized report for MSCohort**



# 1. Spectronaut customized report

The columns of information that MSCohort needs to read



1

2

3

4

Run
R.Run Date
R.Gradient Length [min]
R.FileName

Protein Group
PG.ProteinGroups
PG.Qvalue
PG.Quantity

Peptide
PEP.NrOfMissedCleavages
PEP.StrippedSequence
PEP.Quantity

Elution Group
EG.IsDecoy
EG.ModifiedPeptide
EG.PrecursorId
EG.Qvalue
EG.ApexRT
EG.DatapointsPerPeak
EG.DatapointsPerPeak (MS1)
EG.DeltaRT
EG.EndRT
EG.FWHM
EG.PeakWidth
EG.StartRT
EG.SignalToNoise
EG.TotalQuantity (Settings)

Fragment Group
FG.Charge
FG.PrecMz
FG.PrecWindow
FG.PrecWindowNumber
FG.CalibratedMassAccuracy (PPM)
FG.RawMassAccuracy (PPM)

Fragment
F.RawMassAccuracy (PPM)
F.CalibratedMassAccuracy (PPM)

- 1 Choose **Report** Perspective in Spectronaut;
- 2 Choose a **Normal Report** format schema as a base to build MSCohort report;
- 3 Choose the **columns** of information that MSCohort needs to read;
- 4 Save the selected columns as a new schema, and name the new report schema as “**MSCohort**”.

Spectronaut

[illegible]

- ① Choose the MSCohort report schema;
- ② Report preview: A preview of how MSCohort report will look like;
- ③ Export the matrix by clicking on "Export Report..." in the bottom left corner.



# **MSCohort manual for intra- experiment analysis**



# 1. Analyzing with MSCohort

The screenshot shows the MSCohort(Authorized) software interface. It is divided into several sections: WORKFLOW, Data file, Analysis result, Setting, and Export. Red numbered circles (1-5) are placed next to specific fields to indicate the steps for configuration.

- 1** WORKFLOW: Intra-Experiment (dropdown menu)
- 2** TYPE\_DATA: Thermo\_Orbitrap (dropdown menu)
- 3** PATH\_DATA: t/202212-multi-librator (text field)
- 4** PATH\_ANALYSIS\_RESULT: P18/20231120\_145743\_10multilab\_combine\_SP18.0\_Report.tsv (text field)
- 5** PATH\_EXPORT: D:/dataset/202212-multi-libratory/SP18 (text field)

Other visible fields include THRESHOLD\_FDR (0.01), THRESHOLD\_PEAK\_WIDTH\_TAIL (1), THRESHOLD\_INVALID\_ACQUIRE (100), FLAG\_ANALYZE\_FEATURE (No), and buttons for Import, New, Add, Save as, Save, and Run.

- ① Set **WORKFLOW** as **Intra-experiment**;
- ② Select **TYPE\_DATA** according to the data type ;
- ③ Fill the raw file into the **PATH\_DATA**;
- ④ Fill the Spectronaut customized report for MSCohort into the **PATH\_ANALYSIS\_RESULT**;
- ⑤ Set the **PATH\_EXPORT** for saving the results.

## Note:

Space (“ ”) cannot exist in the file directory (including PATH\_DATA, PATH\_ANALYSIS\_RESULT, and PATH\_EXPORT ), which will affect the normal running of the program.

# 1. Analyzing with MSCohort

The screenshot shows the MSCohort software interface with the following configuration:

- WORKFLOW:** Intra-Experiment (dropdown menu open showing Inter-Experiment, Intra-Experiment)
- Data file:**
  - TYPE\_DATA: Thermo\_Orbitrap (dropdown menu open showing Thermo\_Orbitrap, Bruker\_timsTOF, SCIEX\_ZenoTOF)
  - PATH\_DATA: t/202212-multi-librator (text input), 0/Urine\_DIA1.raw (text input), Add (button)
- Analysis result:**
  - TYPE\_ANALYSIS\_RESULT: Spectronaut (dropdown menu)
  - PATH\_ANALYSIS\_RESULT: P18/20231120\_145743\_10multilab\_combine\_SP18.0\_Report.tsv (text input), Add (button)
  - THRESHOLD\_FDR: 0.01 (text input)
- Setting:**
  - THRESHOLD\_PEAK\_WIDTH\_TAIL: 1 (text input, annotated with 1)
  - THRESHOLD\_INVALID\_ACQUIRING: 100 (text input, annotated with 2)
  - FLAG\_ANALYZE\_FEATURE: No (dropdown menu, annotated with 3)
- Export:**
  - PATH\_EXPORT: D:/dataset/202212-multi-libratory/SP18 (text input)
- Buttons:** Save as (annotated with 4), Save, Run (annotated with 5)

- ① **THRESHOLD\_PEAK\_WIDTH\_TAIL** set as 1 represents that precursors with peak width more than 1 minutes are used to calculate the proportion of precursors with long eluting width (user adjustable parameter according to experimental condition);
- ② **THRESHOLD\_INVALID\_ACQUIRING** set as 100 represents that the retention time length from the start to the number of identified MS2 scans per minute more than 100 as chromatographic invalid acquiring time (user adjustable parameter according to experimental condition);
- ③ Choose whether to extracting FEATURE metrics. **This analysis may take a long time**, especially for timsTOF and zenoTOF data;
- ④ Click **Save as** button to save the config file;
- ⑤ Click **Run** button to start the MSCohort, the progress information will be shown in the command-line interface.

# 1. Analyzing with MSCohort

The screenshot shows the MSCohort software interface with several sections and annotations:

- WORKFLOW:** A dropdown menu is open, showing options: Intra-Experiment, Inter-Experiment (highlighted), Inter-Experiment, and Intra-Experiment. A red circle with the number '2' is next to the 'New' button.
- Data file:**
  - TYPE\_DATA:** A dropdown menu is open, showing options: Thermo\_Orbitrap (highlighted), Thermo\_Orbitrap, Bruker\_timsTOF, and SCIEX\_ZenoTOF.
  - PATH\_DATA:** A text input field contains the path 't/202212-multi-librator' and a file name '0/Urine\_DIA1.raw'. An 'Add' button is next to it.
- Analysis result:**
  - TYPE\_ANALYSIS\_RESULT:** A dropdown menu is open, showing the option: Spectronaut.
  - PATH\_ANALYSIS\_RESULT:** A text input field contains the path 'P18/20231120\_145743\_10multilab\_combine\_SP18.0\_Report.tsv'. An 'Add' button is next to it.
  - THRESHOLD\_FDR:** A text input field contains the value '0.01'.
- Setting:**
  - THRESHOLD\_PEAK\_WIDTH\_TAIL:** A text input field contains the value '1'.
  - THRESHOLD\_INVALID\_ACQUIRE:** A text input field contains the value '100'.
  - FLAG\_ANALYZE\_FEATURE:** A dropdown menu is open, showing the option: No.
- Export:**
  - PATH\_EXPORT:** A text input field contains the path 'D:/dataset/202212-multi-librator/SP18'. An 'Add' button is next to it. A red circle with the number '1' is next to this section.

At the bottom of the interface, there are three buttons: 'Save as', 'Save', and 'Run'.

## Note:

- ① **PATH\_EXPORT :** The output MSCohort analysis results will be stored in a new folder under the PATH\_EXPORT directory, named MSCohort \_“filename” folder.
- ② Do not save the same PATH\_DATA under a same folder, the results will be **overwritten**.
- ③ Choose **New** for a new experiments
- ③ Check the filename in **PATH\_DATA** must in the Spectronaut report in **PATH\_ANALYSIS\_RESULT**.

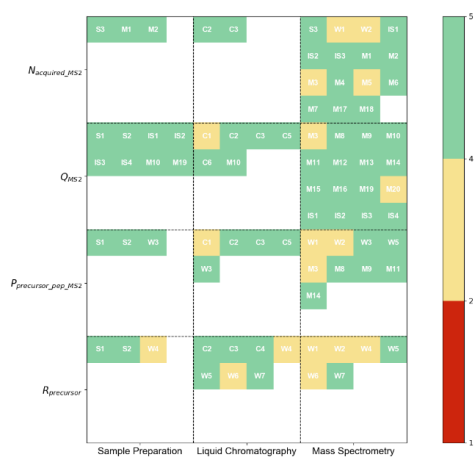
# 2. MSCohort Results

1 picture

- Analysis\_Report.html
- IFNO\_Summary.txt
- INFO\_Chromatography.txt
- INFO\_Cycle\_MS1.txt
- INFO\_Cycle\_MS2.txt
- INFO\_Feature.txt
- INFO\_ID.txt
- INFO\_Mass\_Deviation.txt
- INFO\_MS1.txt
- INFO\_MS1\_PEAKS.txt
- INFO\_MS2.txt
- INFO\_MS2\_PEAKS.txt
- INFO\_peptides.txt
- INFO\_Scans.txt

2

- 1 The MSCohort results ;
- 2 Double clicking [Analysis\\_Report.html](#), the report will be preformed in the browser.



Category	Level II Scoring Item	Description
1. Sample Preparation	S1. Missed cleavages(n=0) of peptides	The percentage of all identified precursors without missed cleavage site; Score 1: <=0.8; Score 5: >=
	S2. Median peptide length	The median of all identified peptide sequence length; Score 1: >=20; Score 5: <=10

# **MSCohort manual for inter- experiment analysis**



# 1. Preparation for MSCohort\_setting.tsv file

名称

bin

ini

lib

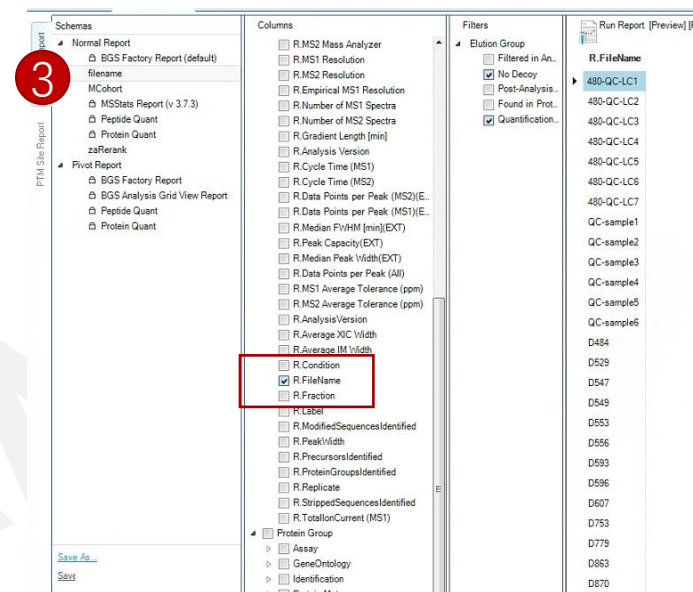
MSCohort\_cfg.txt

MSCohort\_setting.tsv

MSCohortUI.exe

2

	A	B	C	D
1	Group Name	Raw Name	Experiment	Threshold
2	QC	480-QC-LC1	480-QC1	1
3	QC	480-QC-LC2	480-QC2	1
4	QC	480-QC-LC3	480-QC3	1
5	QC	480-QC-LC4	480-QC4	1
6	QC	480-QC-LC5	480-QC5	1
7	QC	480-QC-LC6	480-QC6	1
8	QC	480-QC-LC7	480-QC7	1
9	CRC	D529	D529	1
10	CRC	D547	D547	1
11	CRC	D549	D549	1
12	CRC	D556	D556	1
13	CRC	D593	D593	1
14	CRC	D596	D596	1
15	CRC	D753	D753	1
16	CRC	D779	D779	1
17	CRC	D863	D863	1
18	CRC	D870	D870	1
19	HC	H5	H5	1
20	HC	H13	H13	1
21	HC	H14	H14	1
22	HC	H26	H26	1
23	HC	H31	H31	1
24	HC	H71	H71	1
25	HC	H76	H76	1
26	HC	H81	H81	1
27	HC	H97	H97	1
28	HC	H100	H100	1



① Double clicking [MSCohort\\_setting.tsv](#);

② Fill the columns as the example file showed:

- The first column is the Group Name.
- The second column is the Raw Name, which is the same as [R.FileName](#) reported from Spectronaut.
- The third column is the Raw Name for short, which is convenient for display in large MSCohort report. This column could also be the same as Raw Name column.
- The fourth column is the intensity threshold, protein/peptide intensity that less than this threshold would be replaced as NaN, and would not be used for subsequent Pearson correlation analysis.

③ For large-scale cohorts, users can get the Raw Name list from Spectronaut by exporting only [R.FileName](#) column.

## 2. Analyzing with MSCohort

The screenshot shows the MSCohort software interface with the following configuration:

- Workflow:** Inter-Experiment
- Data file:**
  - TYPE\_DATA:** Thermo\_Orbitrap
  - PATH\_DATA:** (empty)
- Analysis result:**
  - TYPE\_ANALYSIS\_RESULT:** Spectronaut
  - PATH\_ANALYSIS\_RESULT:** P18/20231120\_145743\_10multilab\_combine\_SP18.0\_Report.tsv
  - THRESHOLD\_FDR:** 0.01
- Setting:**
  - PATH\_EXPERIMENT\_RESULT:** D:/dataset/202212-multi-library/MSCohort\_setting.tsv
  - TYPE\_NORMALIZATION:** DirectLFQ
  - FLAG\_OUTLIERS:** 2-SD
  - FLAG\_SHOW\_ORDER:** group series
  - THRESHOLD\_PEAK\_WIDTH\_TAIL1:** (empty)
  - THRESHOLD\_INVALID\_ACQUIRE:** 100
  - FLAG\_ANALYZE\_FEATURE:** No
- Export:**
  - PATH\_EXPORT:** D:/dataset/202212-multi-library/SP18

Buttons at the bottom: Save as, Save, Run.

- ① Set **WORKFLOW** as **Inter-experiment**;
- ② Select **TYPE\_DATA** according to the data type ;
- ③ The **PATH\_DATA** could be empty. MSCohort support the inter-experimental analysis mainly based on identification/quantitative result, eliminating the need to submit raw files, which will obtain the result in a relatively **short time**. It is suitable for large cohort analysis.

In addition, users could choose to add the raw data. MSCohort would provide comprehensive analysis reports not only for inter-experiment, but also for intra-experiment analysis. This may **take a long time**.

- ④ Fill the Spectronaut customized report for MSCohort into the **PATH\_ANALYSIS\_RESULT**;
- ⑤ Fill the **MSCohort\_setting.tsv** into the **PATH\_ANALYSIS\_RESULT**;



## 2. Analyzing with MSCohort

The screenshot shows the MSCohort software interface with the following configuration options:

- WORKFLOW:** Inter-Experiment (dropdown)
- Data file:**
  - TYPE\_DATA: Thermo\_Orbitrap (dropdown)
  - PATH\_DATA: (text input) Add
- Analysis result:**
  - TYPE\_ANALYSIS\_RESULT: Spectronaut (dropdown)
  - PATH\_ANALYSIS\_RESULT: P18/20231120\_145743\_10multilab\_combine\_SP18.0\_Report.tsv (text input) Add
  - THRESHOLD\_FDR: 0.01 (text input)
- Setting:**
  - PATH\_EXPERIMENT\_RESULT: D:/dataset/202212-multi-library/MSCohort\_setting.tsv (text input) Add
  - TYPE\_NORMALIZATION: DirectLFQ (dropdown)
  - FLAG\_OUTLIERS: 2-SD (dropdown)
  - FLAG\_SHOW\_ORDER: group series (dropdown)
  - THRESHOLD\_PEAK\_WIDTH\_TAIL: 1 (text input)
  - THRESHOLD\_INVALID\_ACQUIRING: 100 (text input)
  - FLAG\_ANALYZE\_FEATURE: No (dropdown)
- Export:**
  - PATH\_EXPORT: D:/dataset/202212-multi-library/SP18 (text input) Add

Buttons at the bottom: Save as, Save, Run.

- ① Choose normalization strategies in **TYPE\_NORMALIZATION**. MSCohort support directLFQ, MaxLFQ, and Quantile normalization (default is directLFQ);
- ② Choose **FLAG\_OUTLIERS threshold** according to the experiment condition (default is 2\*SD, users could adjust the threshold (Notes for modifying the scoring criteria));
- ③ Choose **FLAG\_SHOW\_ORDER**. Group series represents the experiment order showed in MSCohort report is the same as the **MSCohort\_setting.tsv; time series** represents the experiment order showed in MSCohort report is sorted by run date.
- ④ Set **THRESHOLD\_PEAK\_WIDTH\_TAIL, THRESHOLD\_INVALID\_ACQUIRING, FLAG\_ANALYZE\_FEATURE** as intra-experiment analysis?

## 2. Analyzing with MSCohort

The screenshot shows the MSCohort(Authorized) window. It has a 'WORKFLOW' dropdown set to 'Inter-Experiment' and buttons for 'Import' and 'New'. The interface is divided into several sections: 'Data file' with 'TYPE\_DATA' (Thermo\_Orbitrap) and 'PATH\_DATA' (empty) with an 'Add' button; 'Analysis result' with 'TYPE\_ANALYSIS\_RESULT' (Spectronaut), 'PATH\_ANALYSIS\_RESULT' (P18/20231120\_145743\_10multilab\_combine\_SP18.0\_Report.tsv) with an 'Add' button, and 'THRESHOLD\_FDR' (0.01); 'Setting' with 'PATH\_EXPERIMENT\_RESULT' (D:/dataset/202212-multi-library/MSCohort\_setting.tsv) with an 'Add' button, 'TYPE\_NORMALIZATION' (DirectLFQ), 'FLAG\_OUTLIERS' (2-SD), 'FLAG\_SHOW\_ORDER' (group series), 'THRESHOLD\_PEAK\_WIDTH\_TAIL' (empty), 'THRESHOLD\_INVALID\_ACQUIRE' (100), and 'FLAG\_ANALYZE\_FEATURE' (No); and 'Export' with 'PATH\_EXPORT' (D:/dataset/202212-multi-library/SP18) with an 'Add' button. At the bottom right, there are three buttons: 'Save as' (annotated with a red circle 2), 'Save' (annotated with a red circle 2), and 'Run' (annotated with a red circle 3). A red circle 1 is placed next to the 'PATH\_EXPORT' field in the 'Export' section.

- ① Set the **PATH\_EXPORT** for saving the results. The output MSCohort analysis results will be stored in a new folder under the PATH\_EXPORT directory, named MSCohort \_“year+month+day” folder.

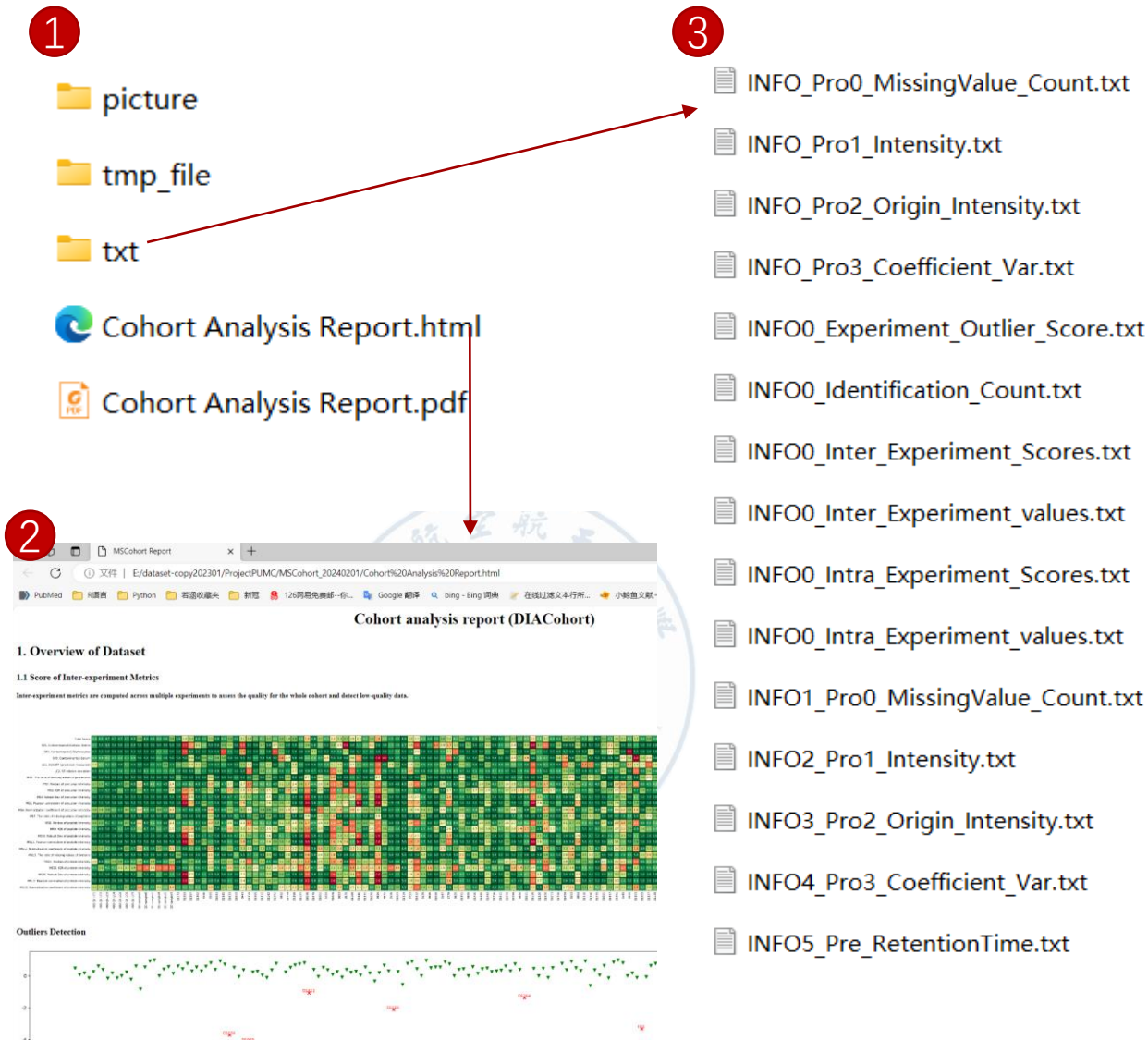
**Do not save the different experiments results under a same folder, the results will be **overwritten**.**

- ② Click **Save as** button to save the config file;
- ③ Click **Run** button to start the MSCohort, the progress information will be shown in the command-line interface.

### Note:

Space (“ ”) cannot exist in the file directory (including PATH\_DATA, PATH\_ANALYSIS\_RESULT, PATH\_EXPERIMENT\_RESULT, and PATH\_EXPORT ), which will affect the normal running of the program.

### 3. MSCohort Results



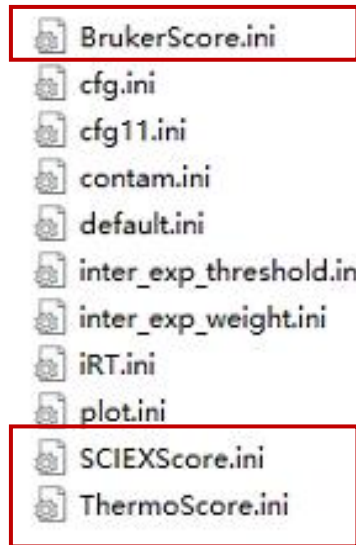
- ① The MSCohort results ;
- ② Double clicking **Cohort Analysis Report.html**, the report will be preformed in the browser.
- ③ Double clicking **txt** folder, the outputs are also exported to simple tab-delimited text files.

# **Notes for modifying the scoring criteria**



# 1. Modifying the scoring criteria for intra-experiment analysis

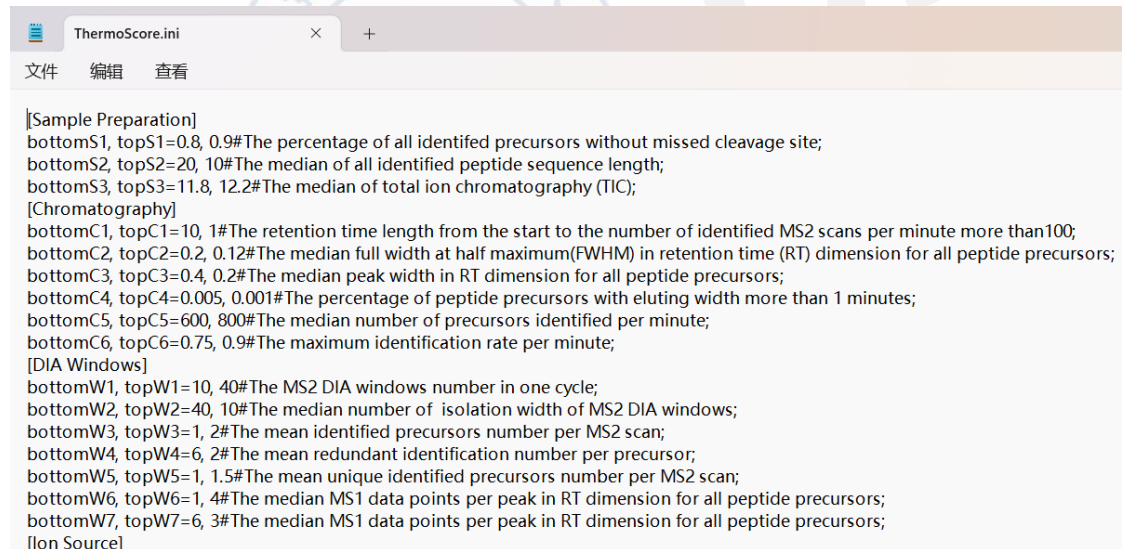
1



① Open the **ini** folder, there will be three parameter files related to intra-experiment scoring: ThermoScore.ini, BrukerScore.ini, SCIEXScore.ini;

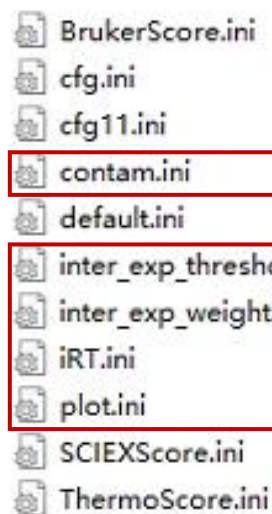
② Click the file to modify the second-level scoring threshold. If you do not modify it, it will be the default value. Closing the file and running the software, it will score according to the standard you set.

2

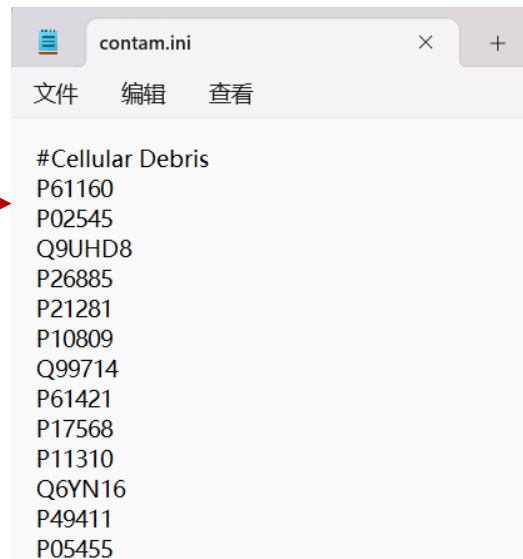


# 1. Modifying the scoring criteria for inter-experiment analysis

1



2

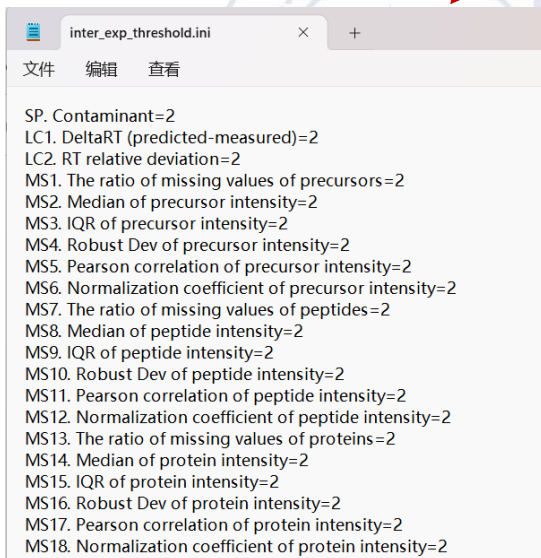


① Open the [ini](#) folder, there will be five parameter files related to inter-experiment scoring: [contam.ini](#), [inter\\_exp\\_threshold.ini](#), [inter\\_exp\\_weight.ini](#), [iRT.ini](#), [plot.ini](#);

② Open the [contam.ini](#), users can modify the list of contaminating proteins, # begins with contaminants category, enter the list of contaminating proteins under this category below.

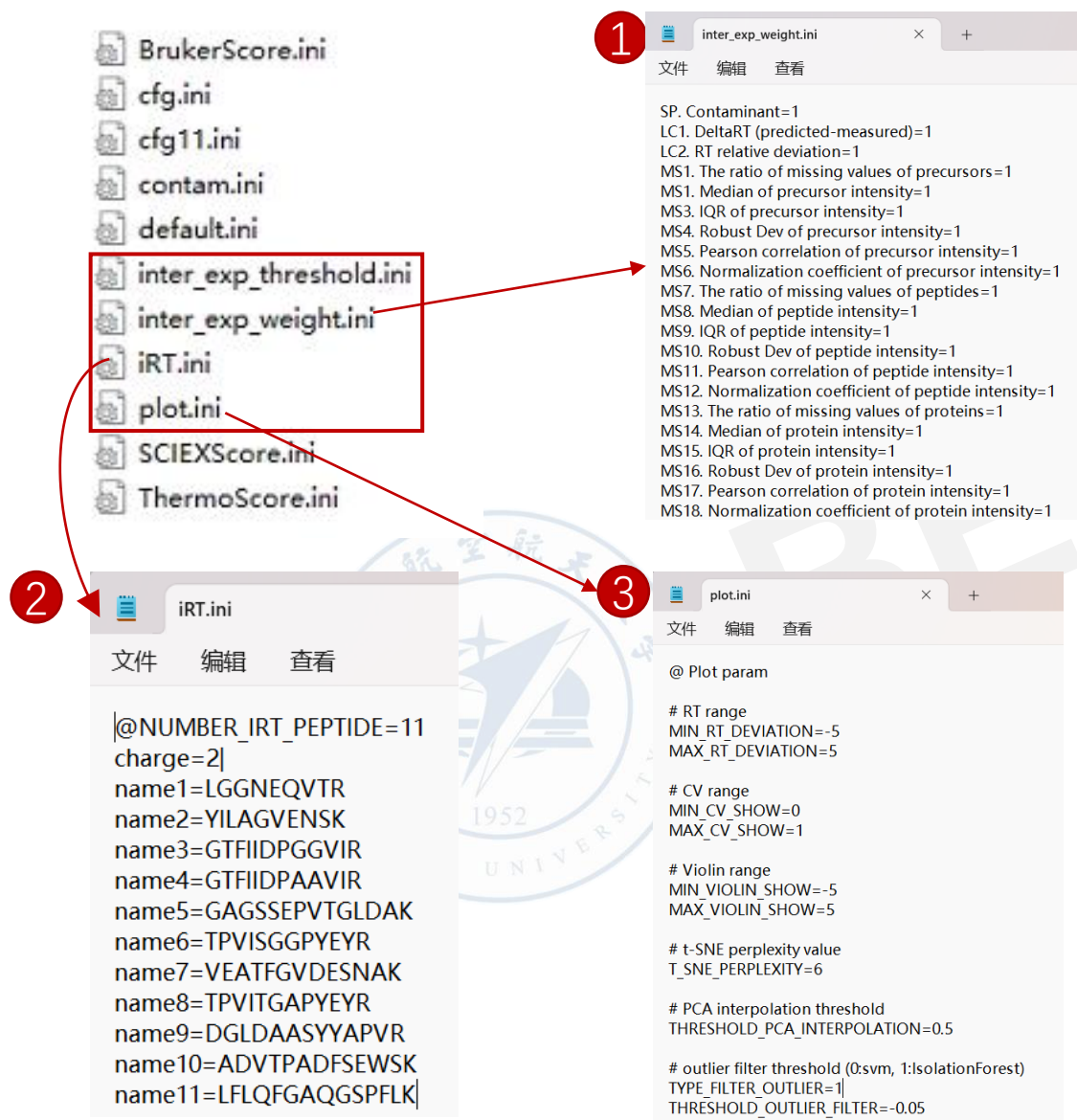
③ Open the [inter\\_exp\\_threshold.ini](#), users can modify the scoring threshold. For each metric, we initially defined potentially outlier experiments as those with a value more than two standard deviations ( $2 \times SD$ ) from the median. If you do not modify it, it will be the default value as 2. Closing and saving the file, it will score according to the standard you set.

3





# 1. Modifying the scoring criteria for inter-experiment analysis



- ① Open the [inter\\_exp\\_weight.ini](#), users can modify the scoring weight for each metric. We initially defined the weight as 1 for each metric. If you do modify it, closing and saving the file, it will score according to the standard you set.
- ② Open the [iRT.ini](#), users can modify the list of iRT peptide sequence. The default iRT peptide sequence is the 11 non-naturally occurring synthetic peptides from the iRT kit (Biognosys). Users can modify the sequence according to experimental conditions.
- ③ Open the [plot.ini](#), users can modify the parameters related to plot. RT range is y axis range showed in RT deviation analysis plot; CV range is the y axis range showed in CV plot; Violin range is the y axis range showed in Intensity ratio distribution plot; TYPE\_FILTER\_OUTLIERS is the approach to detect outliers (0: one-class SVM; 1: isolation forest); THRESHOLD\_OUTLIER\_FILTER is threshold to flag as outliers (In the data we tested, the SVM model threshold was set to -1; the isolation forest was set to -0.05).

# Thanks!

