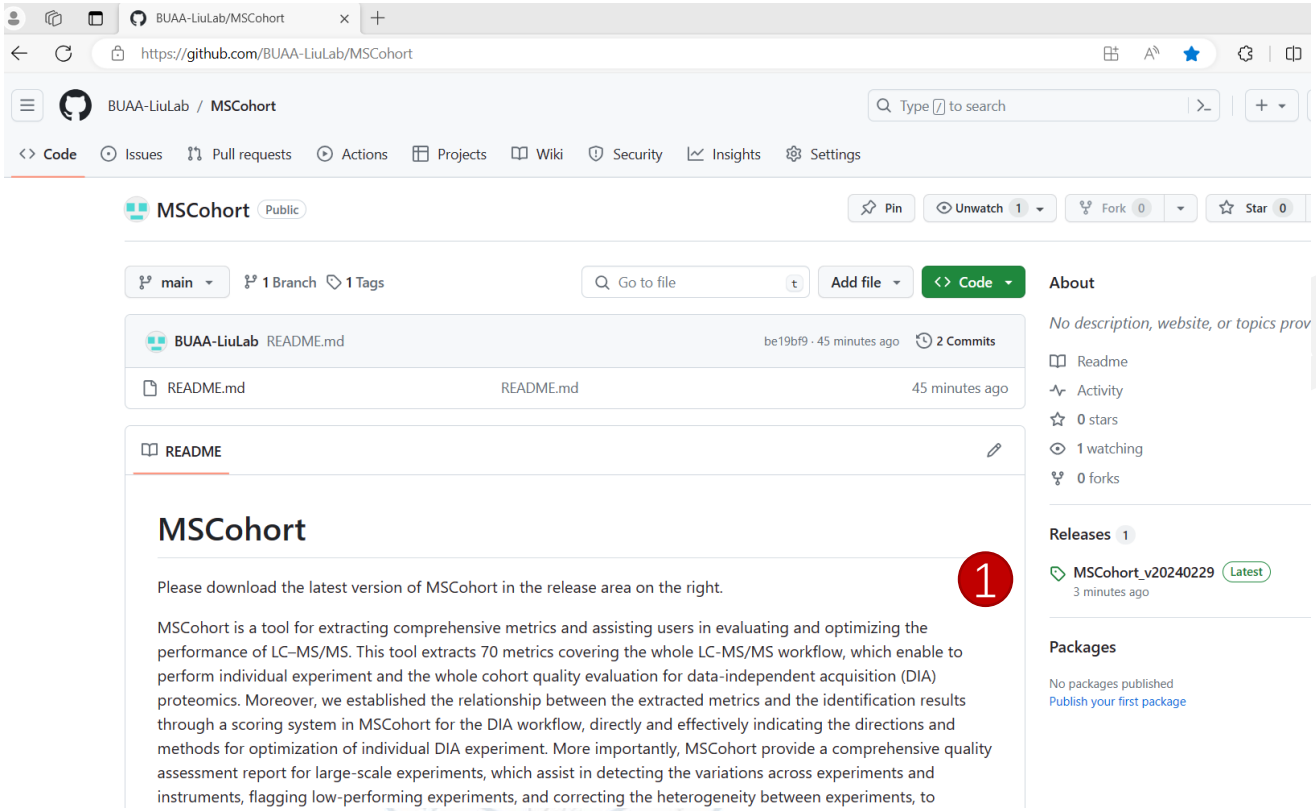
MSCohort Manual

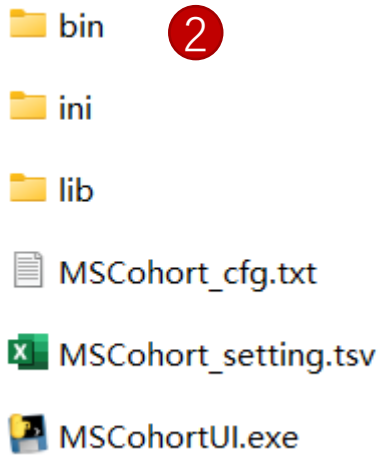
Version. 202402

CONTENTS

MSCohort download

1. MSCohort download





① Login

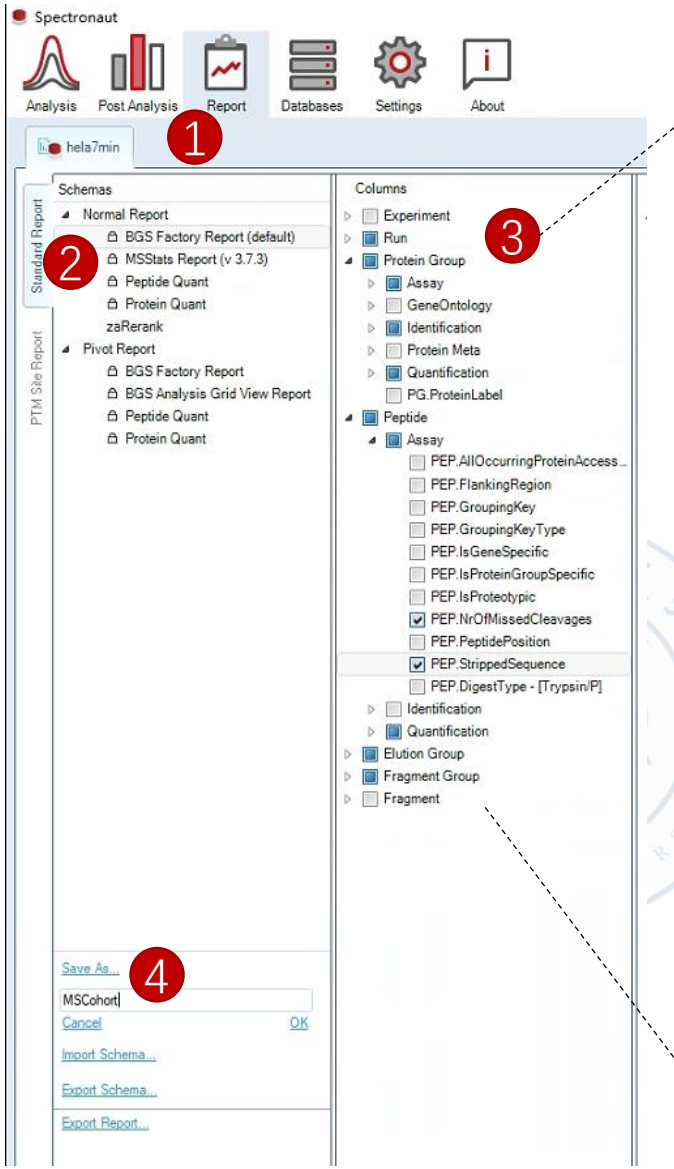
https://github.com/BUAA-LiuLab/MSCohort

and download the lasted 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

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)

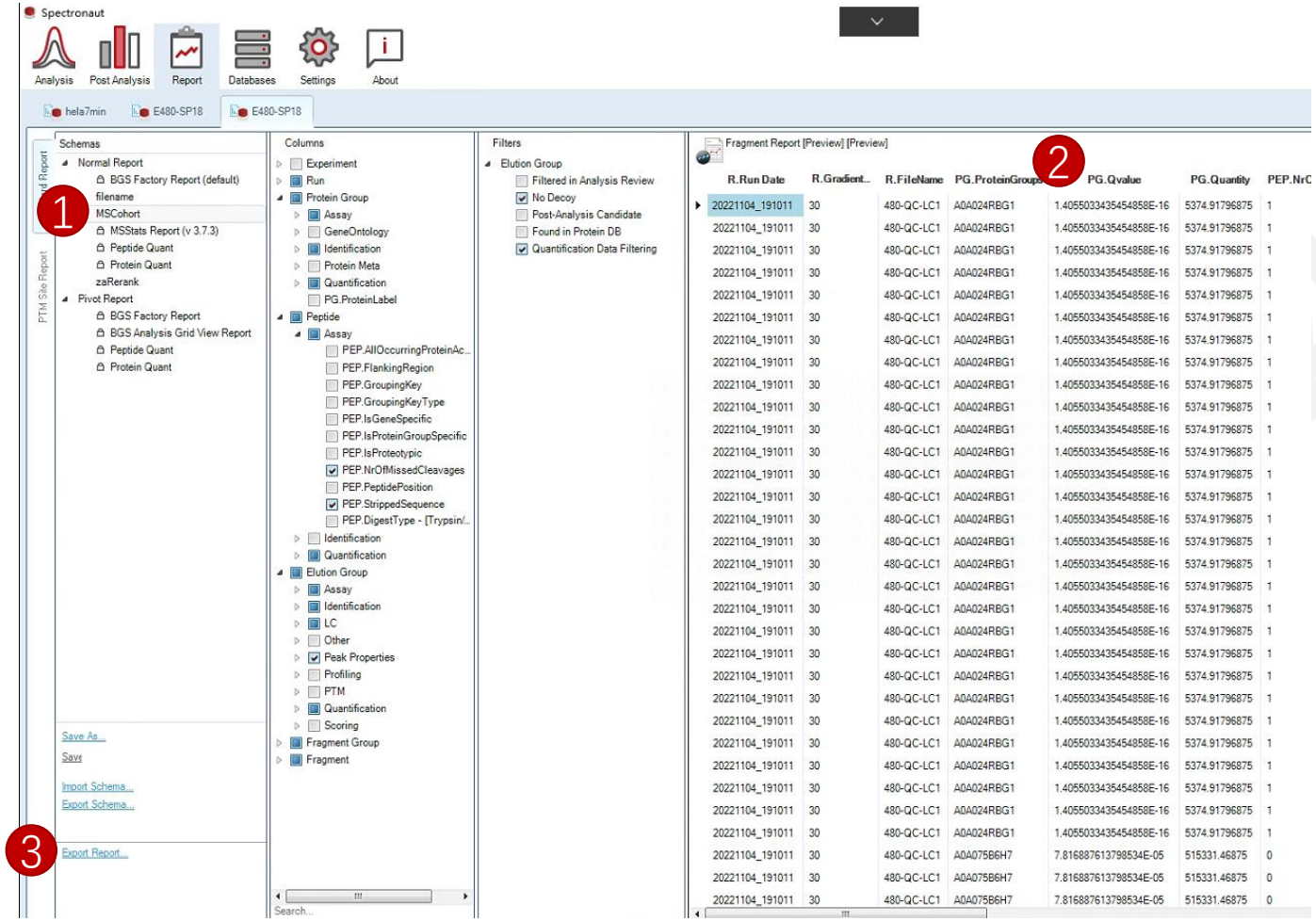
① Choose Report Perspective in Spectronaut;

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".

2. Export MSCohort report from Spectronaut



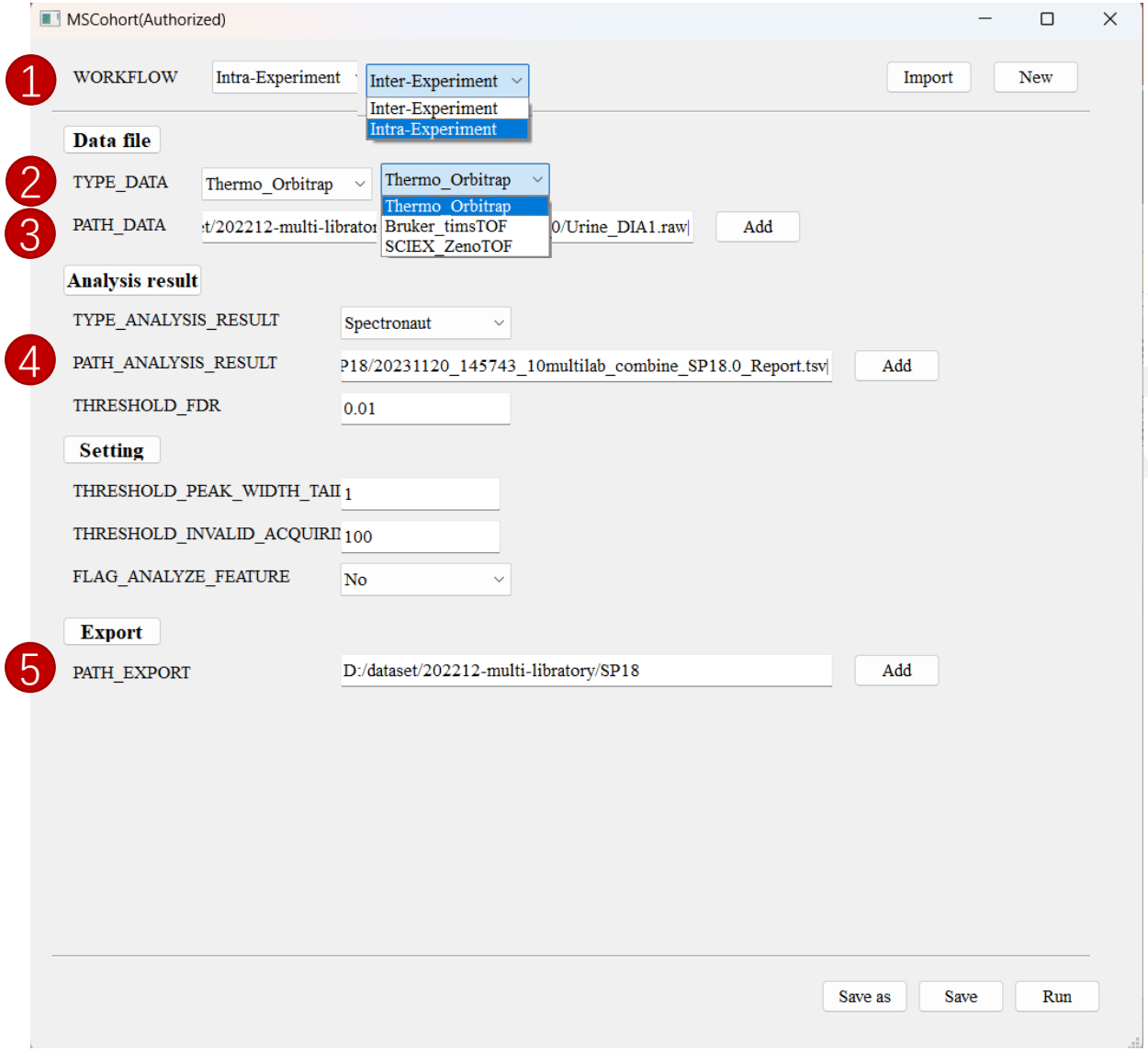
① Choose the MSCohort report schema;

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



① Set **WORKFLOW** as **Intra-experiment**;

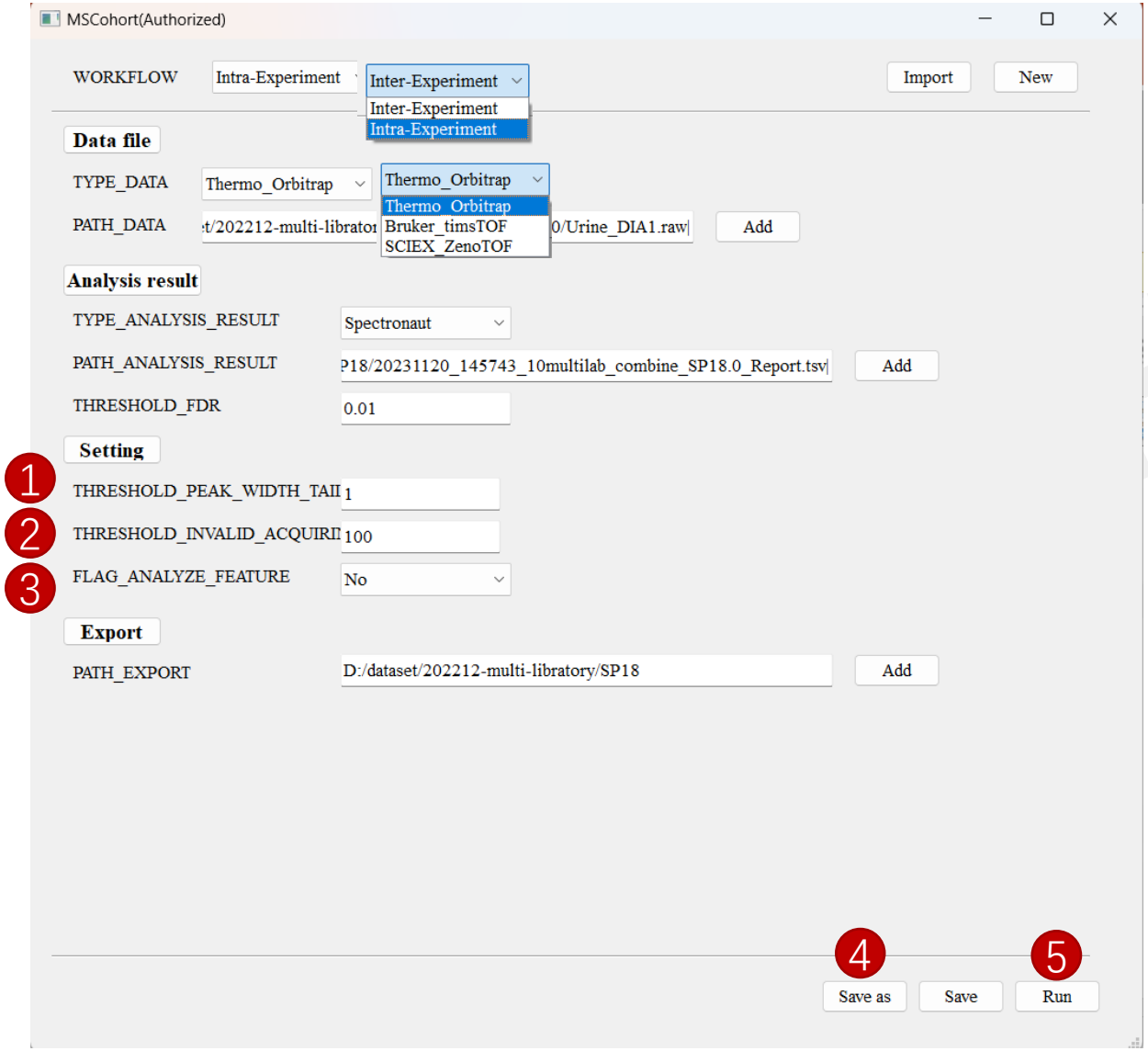
② Select **TYPE_DATA** according to the data type ;

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.

1. Analyzing with MSCohort



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);

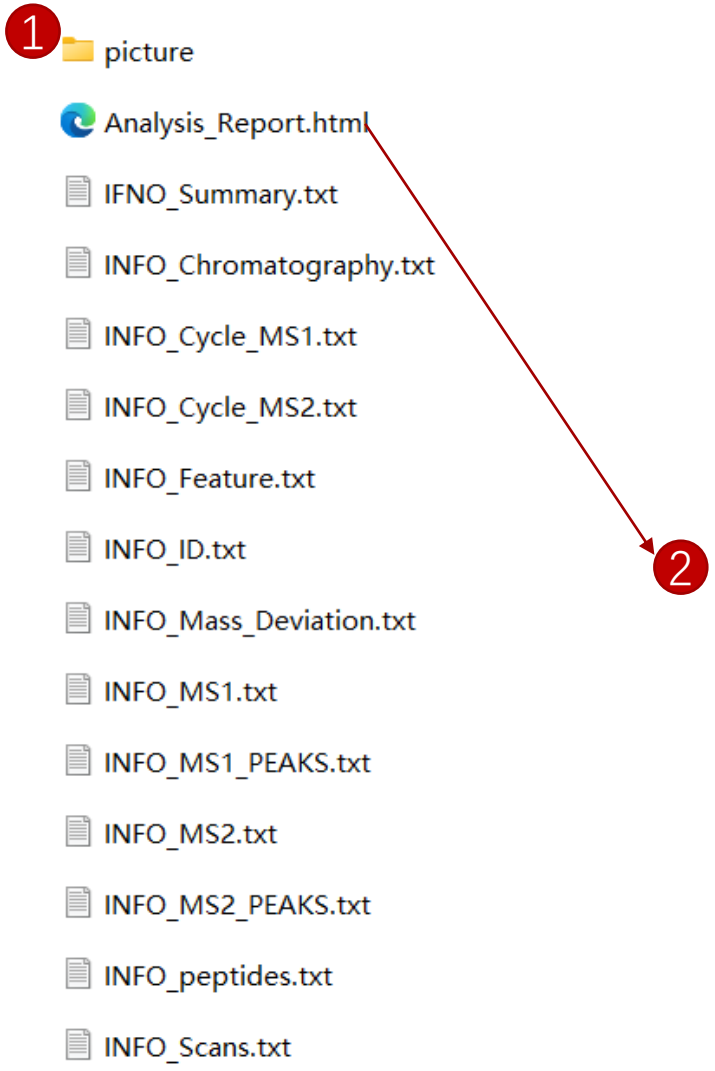
② **THRESHOLD_ INVALID_ACQUIRING** set as 100 represents that the retention time length from the start to the number of identified MS2 scans per minute more than 100 as chromatographic invalid acquiring time (user adjustable parameter according to experimental condition);

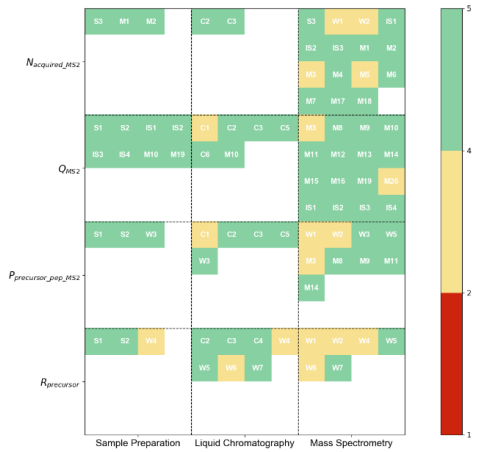
③ Choose whether to extracting FEATURE metrics. This analysis may take a long time, especially for timsTOF and zenoTOF data;

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.

2. MSCohort Results





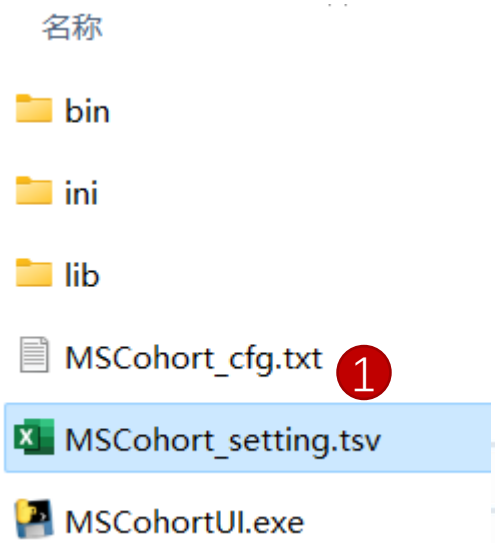
Category	Level II Scoring Item	Description
1. Sample Preparation	S1. Missed cleavages(n=0) of peptides	The percentage of all identifed precursors without missed cleavage site;**Score 1**: <=0.8; **Score 5**: >=
	S2. Median peptide length	The median of all identified peptide sequence length;**Score 1**: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

① The MSCohort results ;

② Double clicking **Analysis_Report.html**, the report will be preformed in the browser.

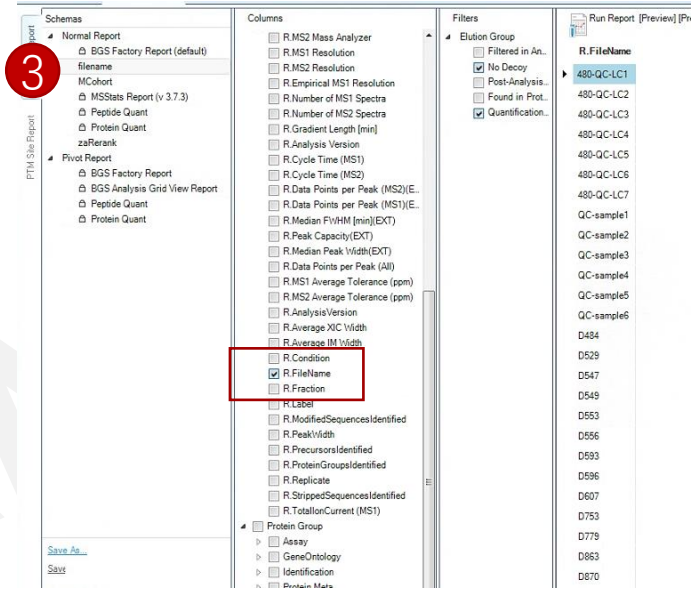
MSCohort manual for inter-experiment analysis

1. Preparation for MSCohort_setting.tsv file



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



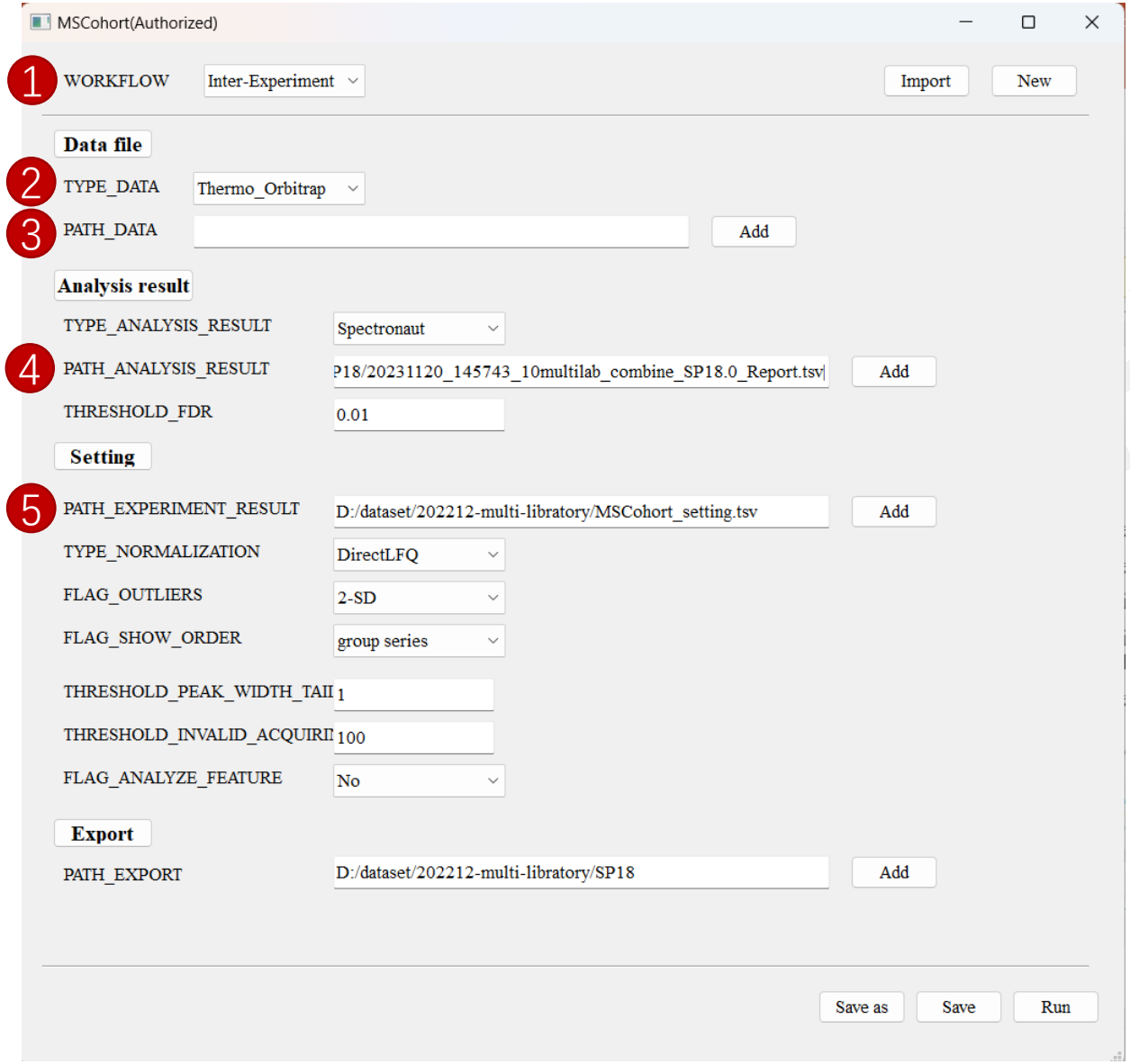
① Double clicking MSCohort_setting.tsv;

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



① Set **WORKFLOW** as **Inter-experiment**;

② Select **TYPE_DATA** according to the data type ;

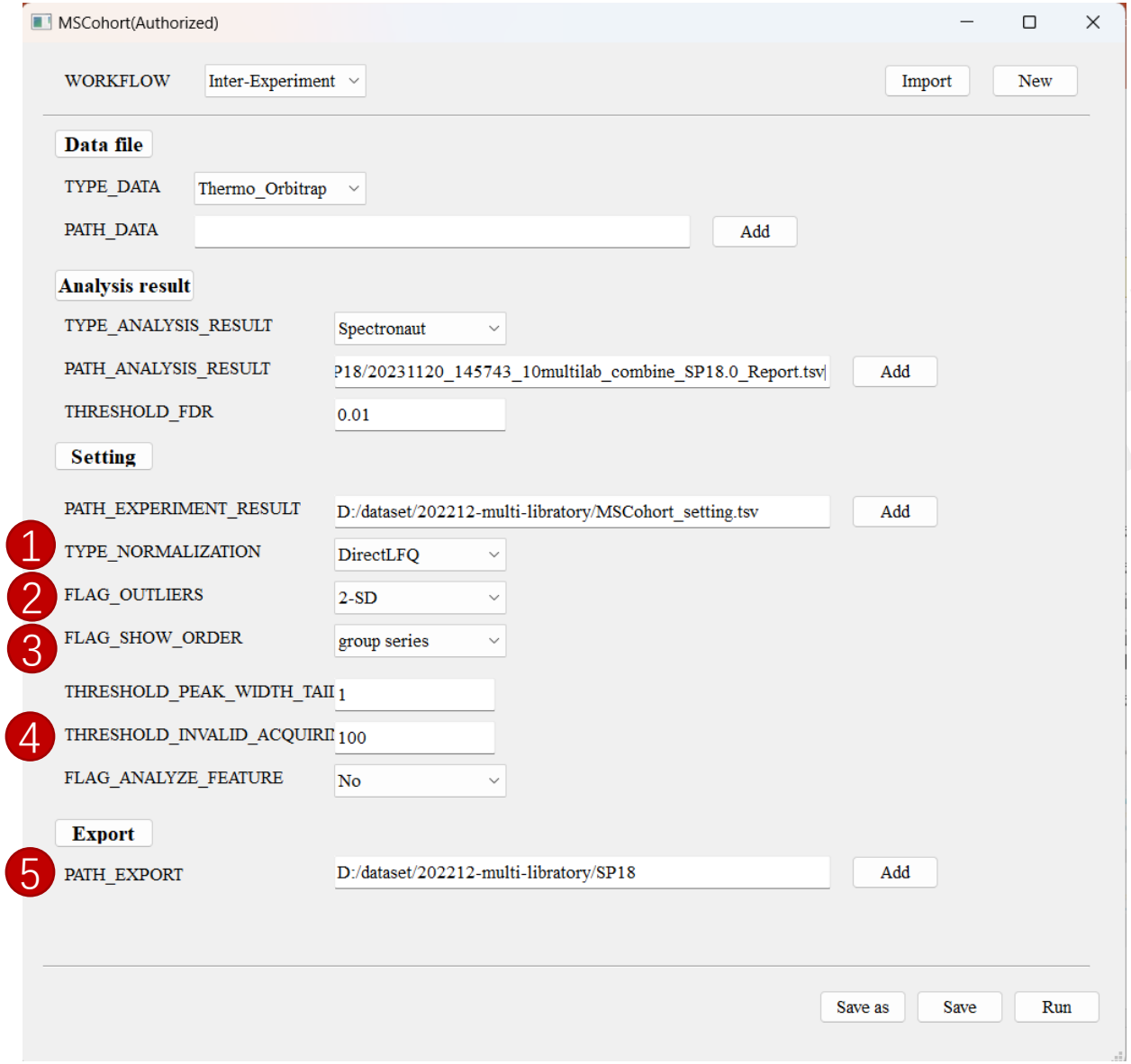
③ The **PATH_DATA** could be empty. MSCohort support the inter-experimental analysis mainly based on identification/quantitative result, eliminating the need to submit raw files, which will obtain the result in a relatively short time. It is suitable for large cohort analysis.

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



① Choose normalization strategies in **TYPE_NORMALIZATION**. MSCohort support directLFQ, MaxLFQ, and Quantile normalization (default is directLFQ);

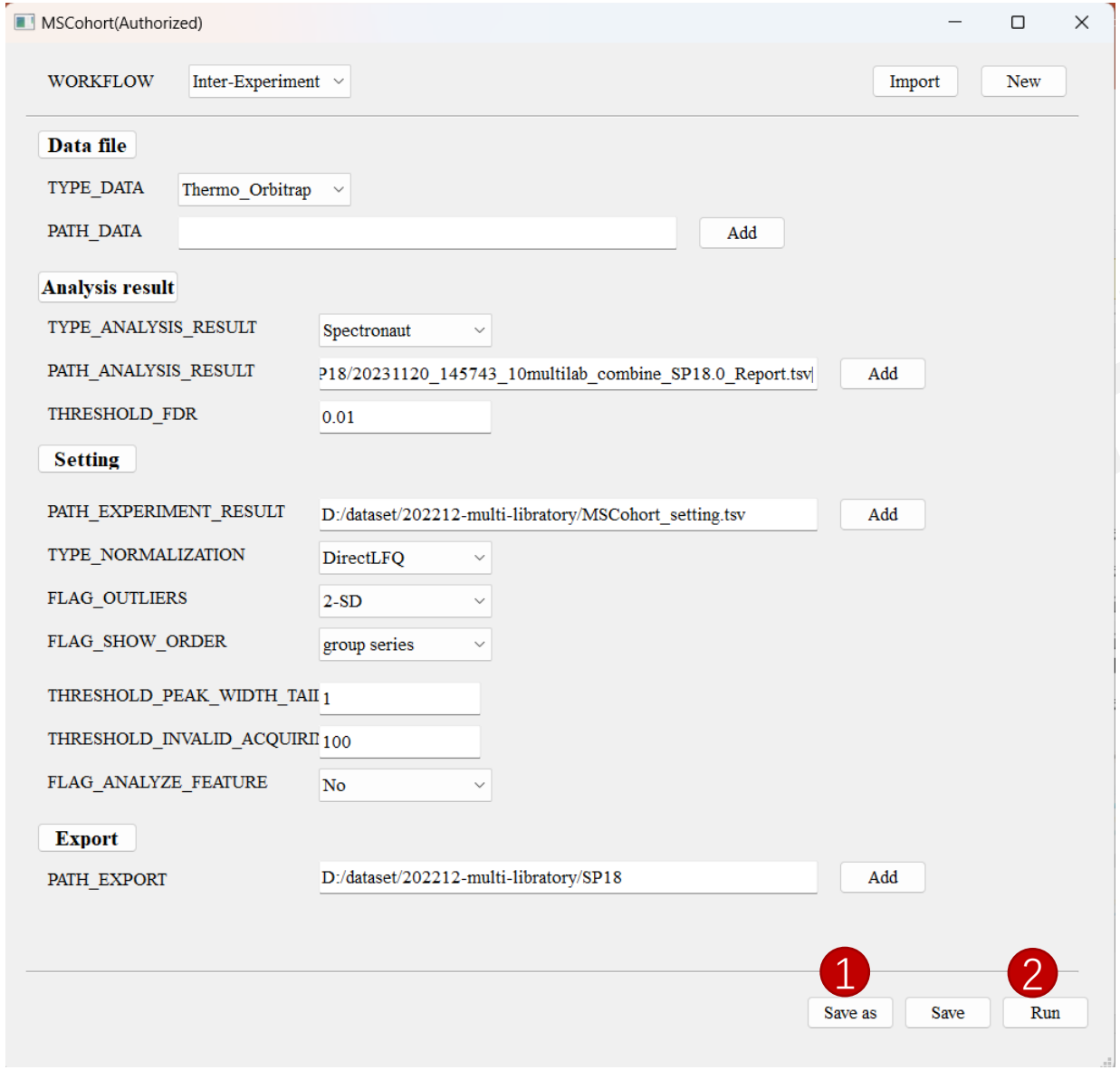
② Choose **FLAG_OUTLIERS threshold** according to the experiment condition (default is 2*SD, users could adjust the threshold (Notes for modifying the scoring criteria));

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 **THERSHOLD_PEAK_WIDTH_TAIL, THRESHOLD INVALID_ACQUIRING, FLAG_ANALYZE_FEATURE** as intra-experiment analysis;

⑤ Set the **PATH_EXPORT** for saving the results.

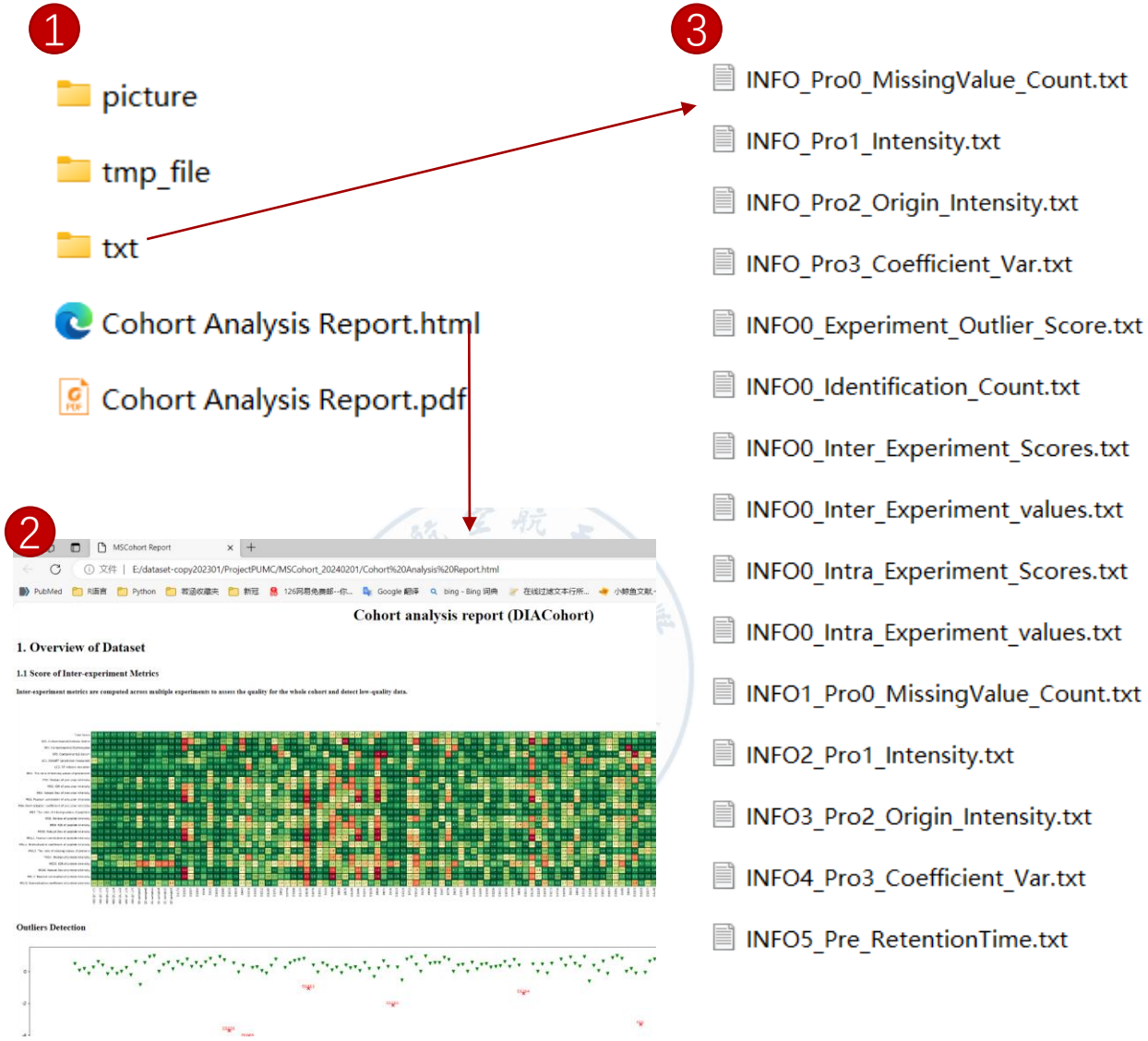
2. Analyzing with MSCohort



① Click **Save as** button to save the config file;

② Click **Run** button to start the MSCohort, the progress information will be shown in the command-line interface.

3. MSCohort Results



① The MSCohort results ;

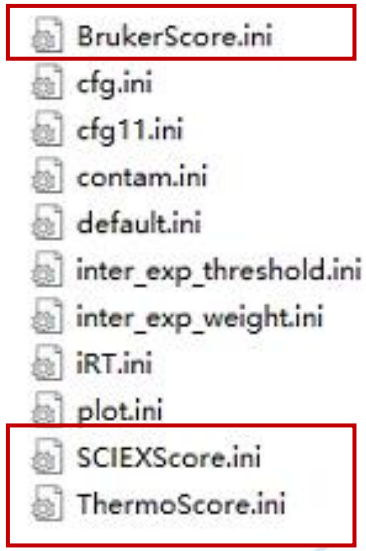
② Double clicking **Cohort Analysis Report.html**, the report will be preformed in the browser.

③ Double clicking **txt** folder, the outputs are also exported to simple tab-delimited text files.

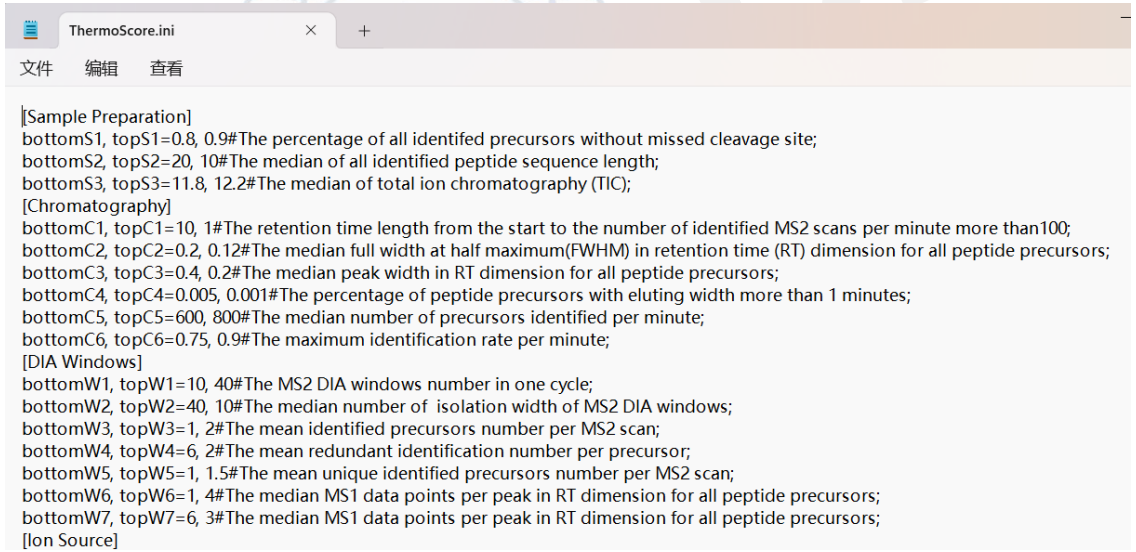
Notes for modifying the scoring criteria

1. Modifying the scoring criteria for intra-experiment analysis

1



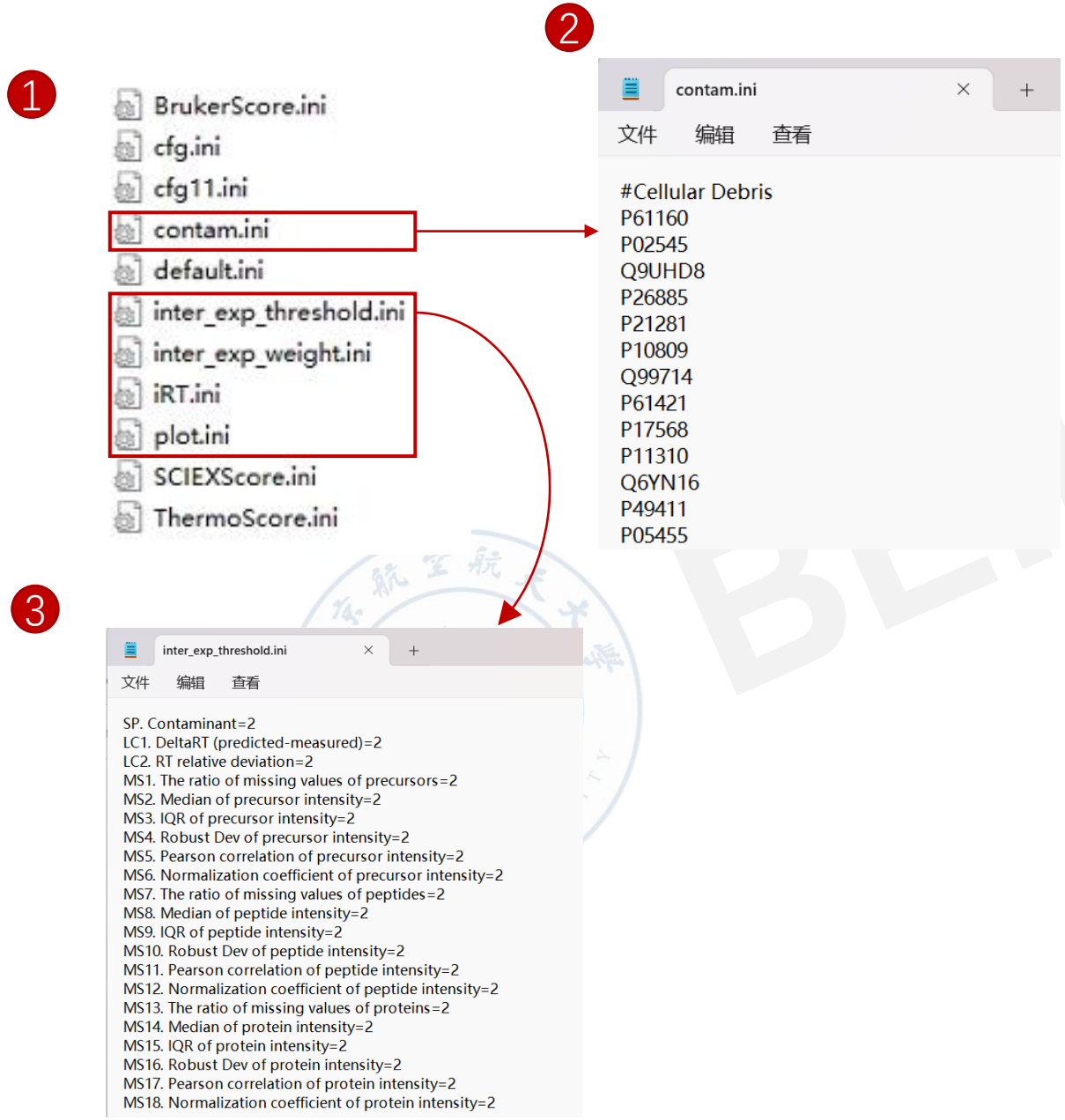
2



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.

1. Modifying the scoring criteria for inter-experiment analysis

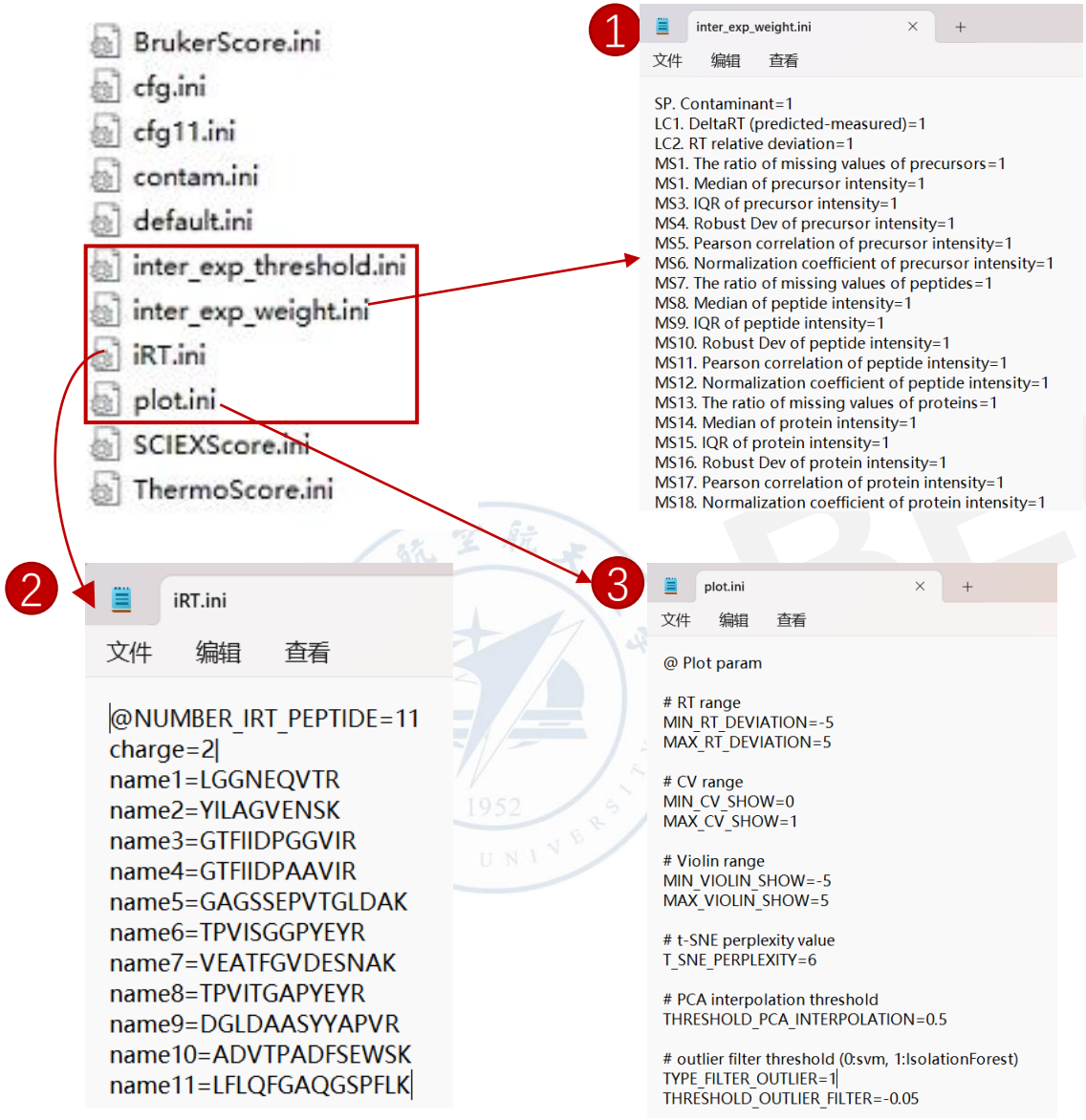


① Open the **ini** folder, there will be five parameter files related to inter-experiment scoring: contam.ini, inter_exp_threshold.ini, inter_exp_weight.ini, iRT.ini, plot.ini;

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*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.

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!