MSCohort Manual

Version. 202405

BEIHANG

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 ananlysis by MSCohort.

BEIHANG

Demo dataset

The demo dataset has been deposited to the ProteomeXchange Consortium (https://proteomecentral.proteomexchange.org) via the iProX partner repository with the dataset identifier PXD050389 (in ProteomeXchange) and IPX0008331000 (in iProX).

The demo datasets available for MSCohort analysis. You can download this dataset for testing and using MSCohort.

This dataset contains raw files of 7 human urine QC DIA data from Orbitrap Exploris 480, Spectronaut analysis results and MSCohort report results.

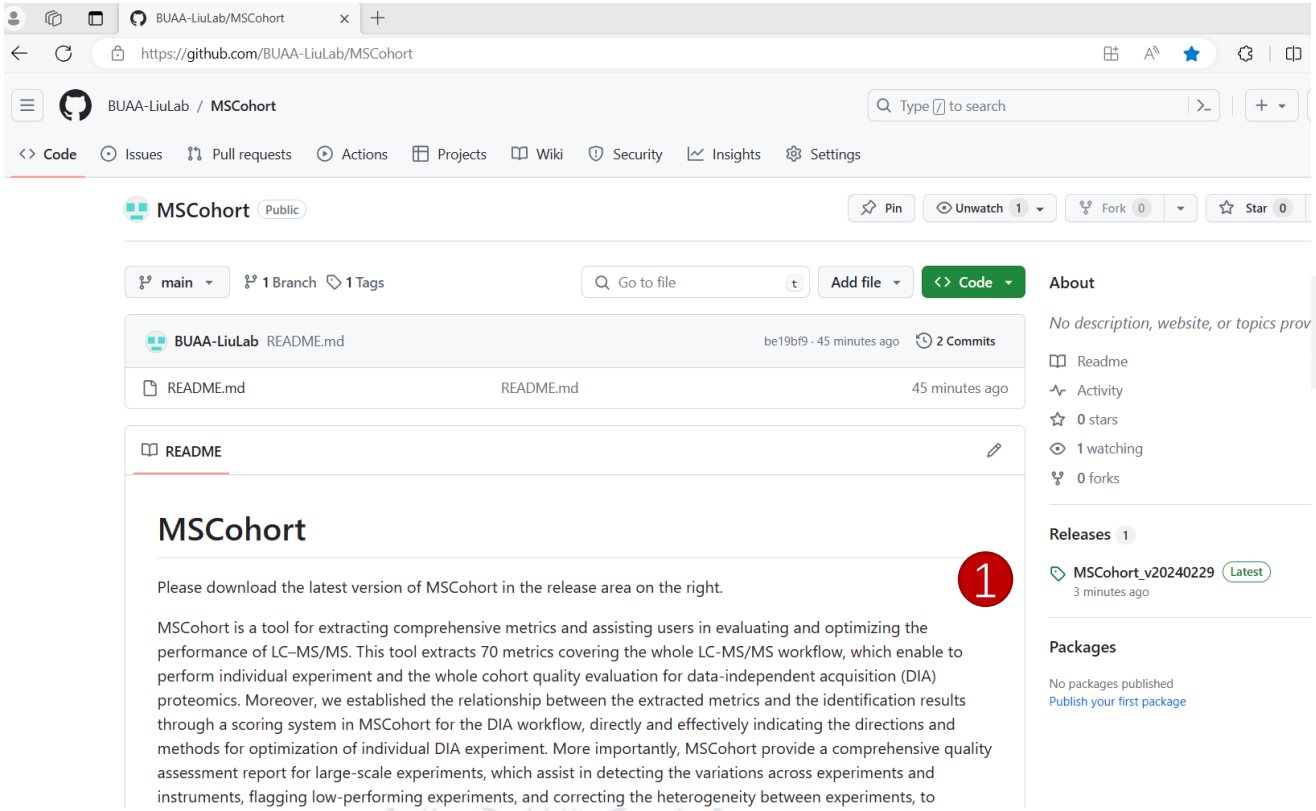
CONTENTS

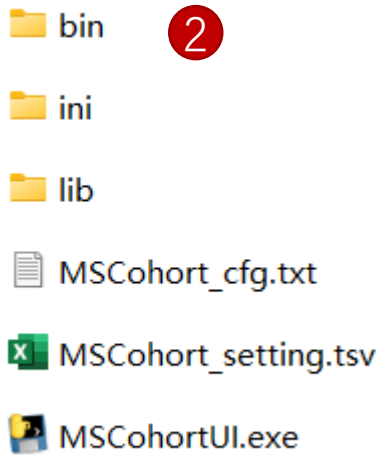
BEIHANG

MSCohort download

BEIHANG

1. MSCohort download





① Login

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

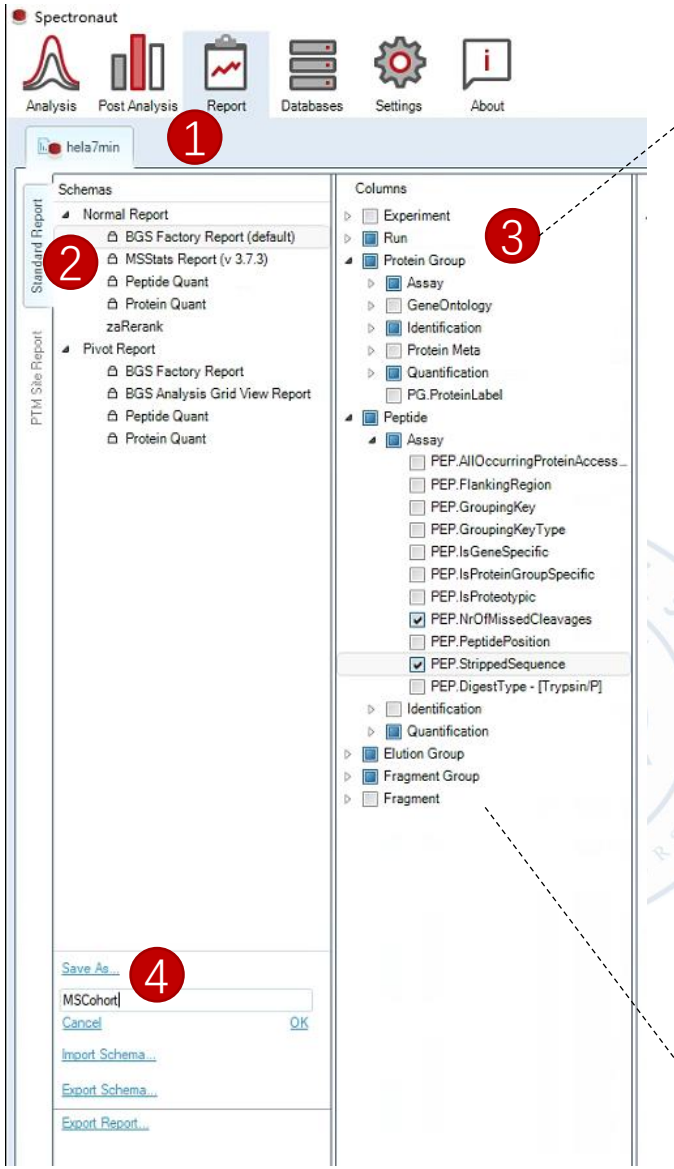
and download the lasted version at the right of the page.

② The unzipped MSCohort file.

Spectronaut customized report for MSCohort

BEIHANG

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)

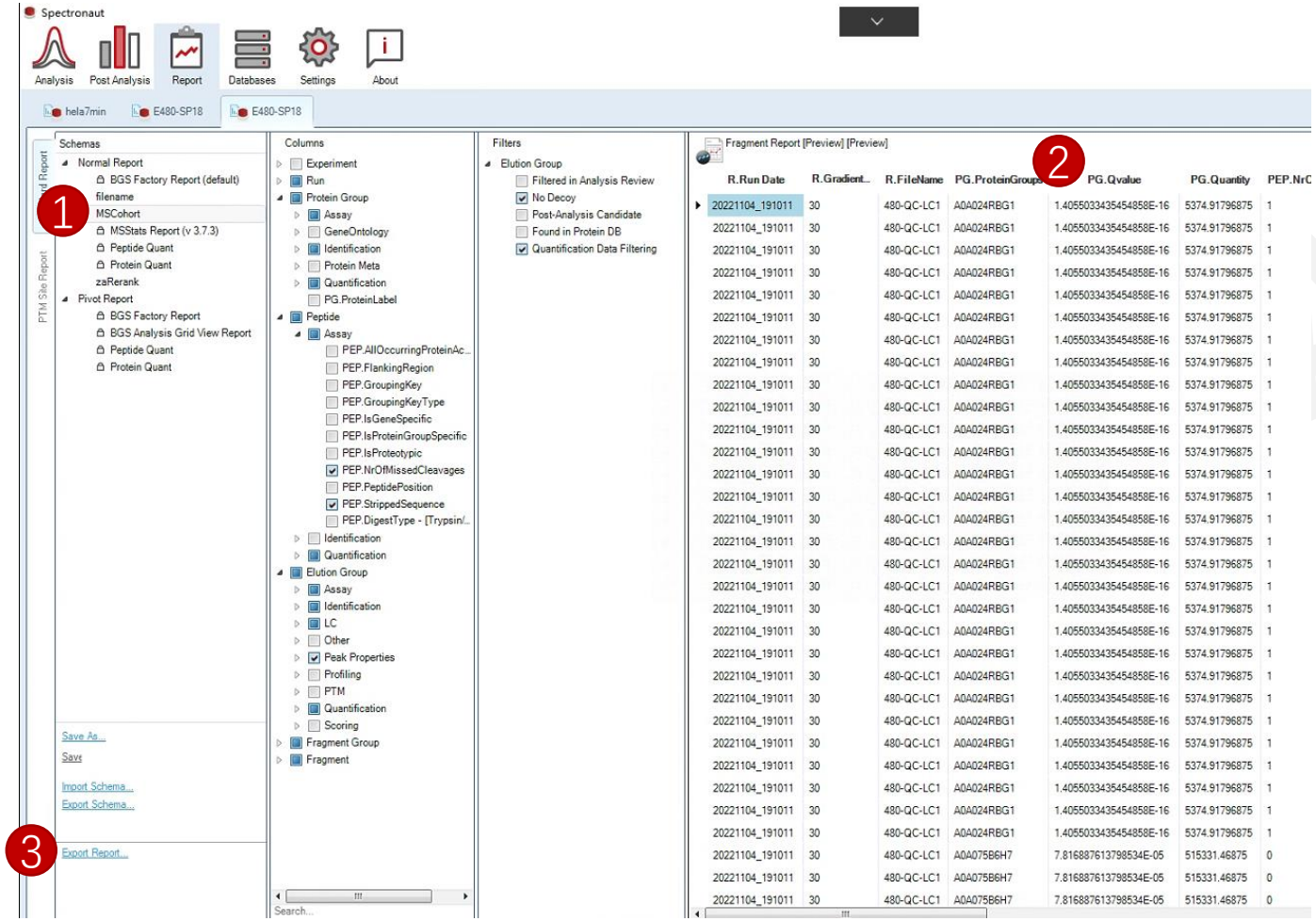
① Choose Report Perspective in Spectronaut;

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;

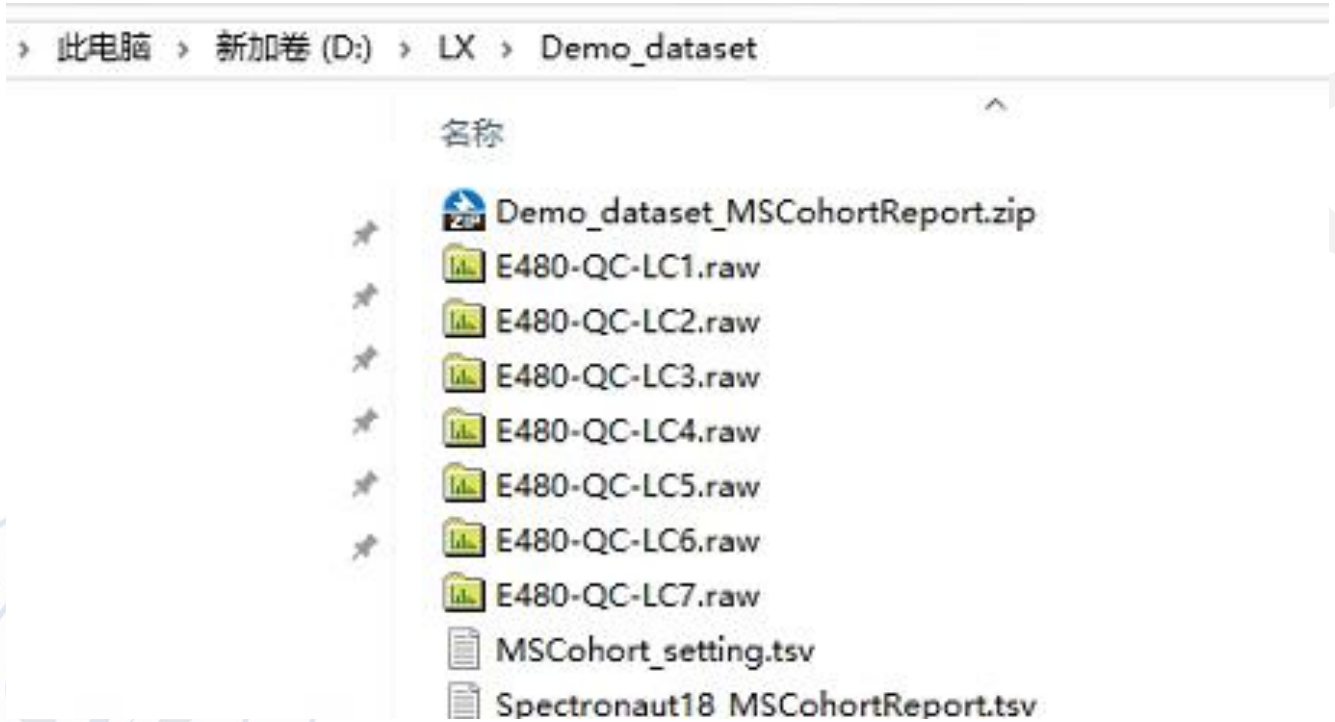
② Report preview: A preview of how MSCohort report will look like;

③ Export the matrix by clicking on "Export Report…" in the bottom left corner.

MSCohort manual for intra-experiment analysis

BEIHANG

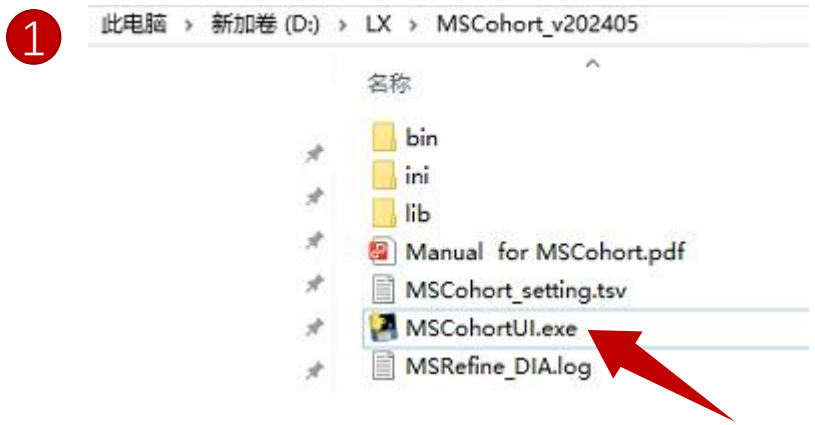
Demo dataset

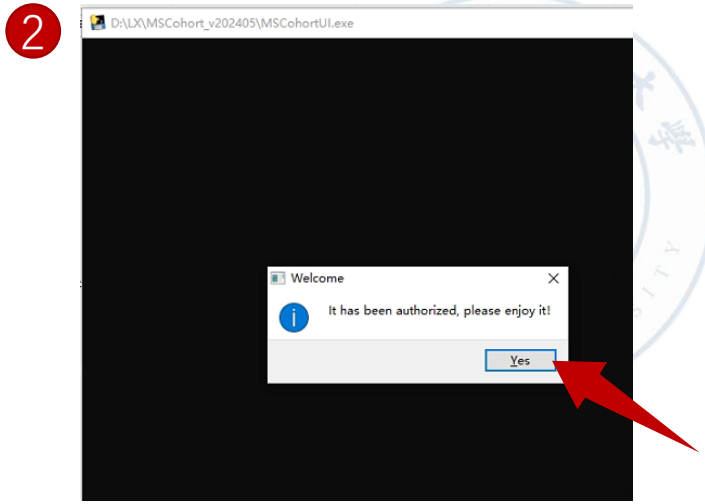


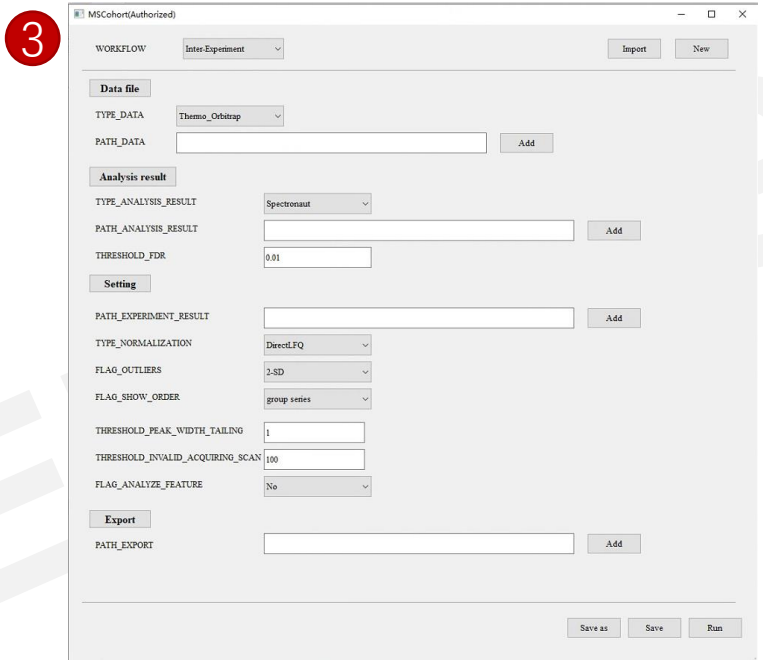
Taking the E480-QC-LC1.raw file as an example to demonstrate the workflow of intra-experiment analysis.

Intra-experiment analysis enables the systematic evaluation and optimization of individual DIA experiments.

1. Analyzing with MSCohort





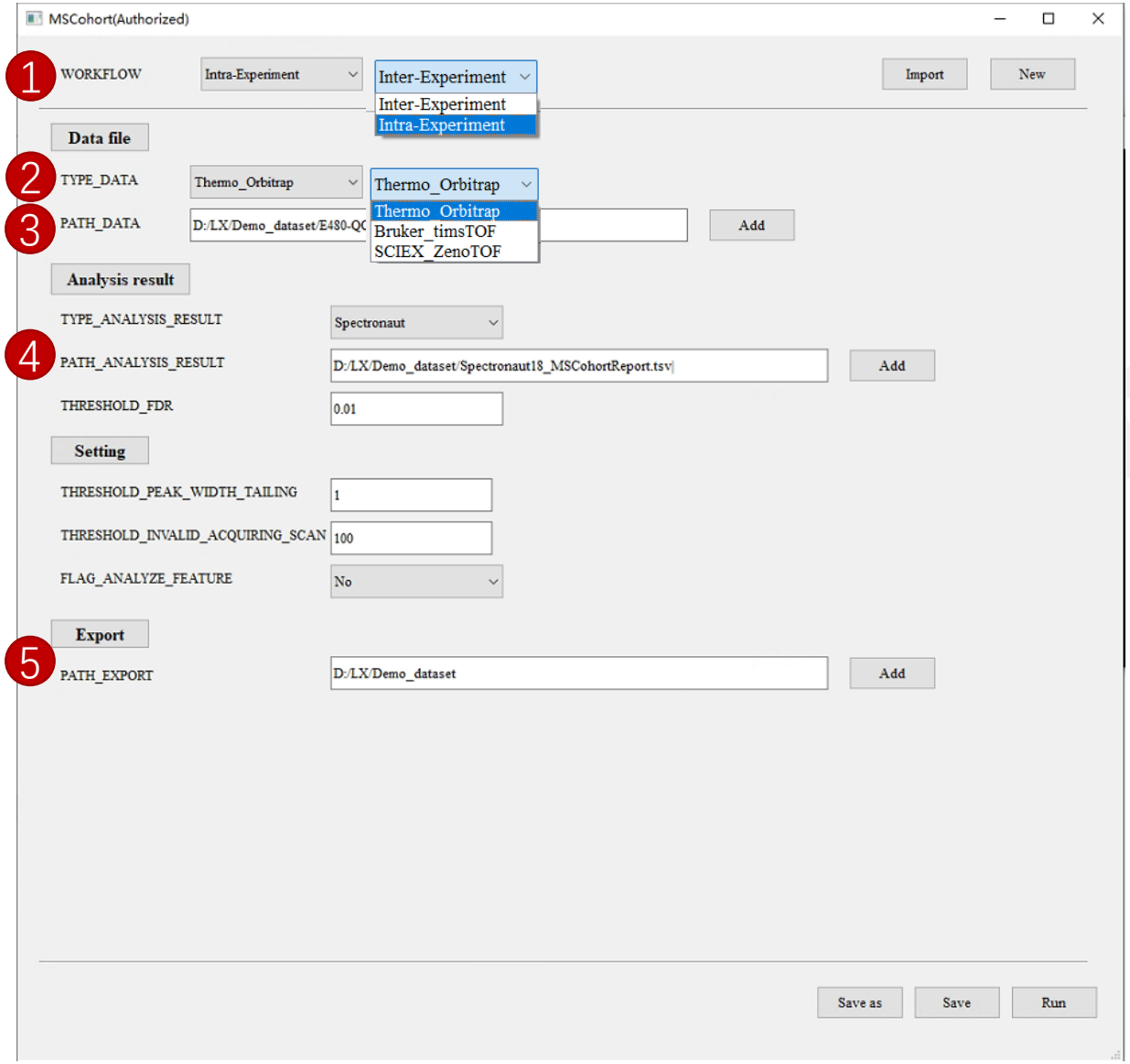


① Double-click MSCohortUI.exe;

② Click Yes ;

③ The MSCohortUI.exe settings screen is displayed.

1. Analyzing with MSCohort



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

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

③ Click Add to select the raw file into the **PATH_DATA**;

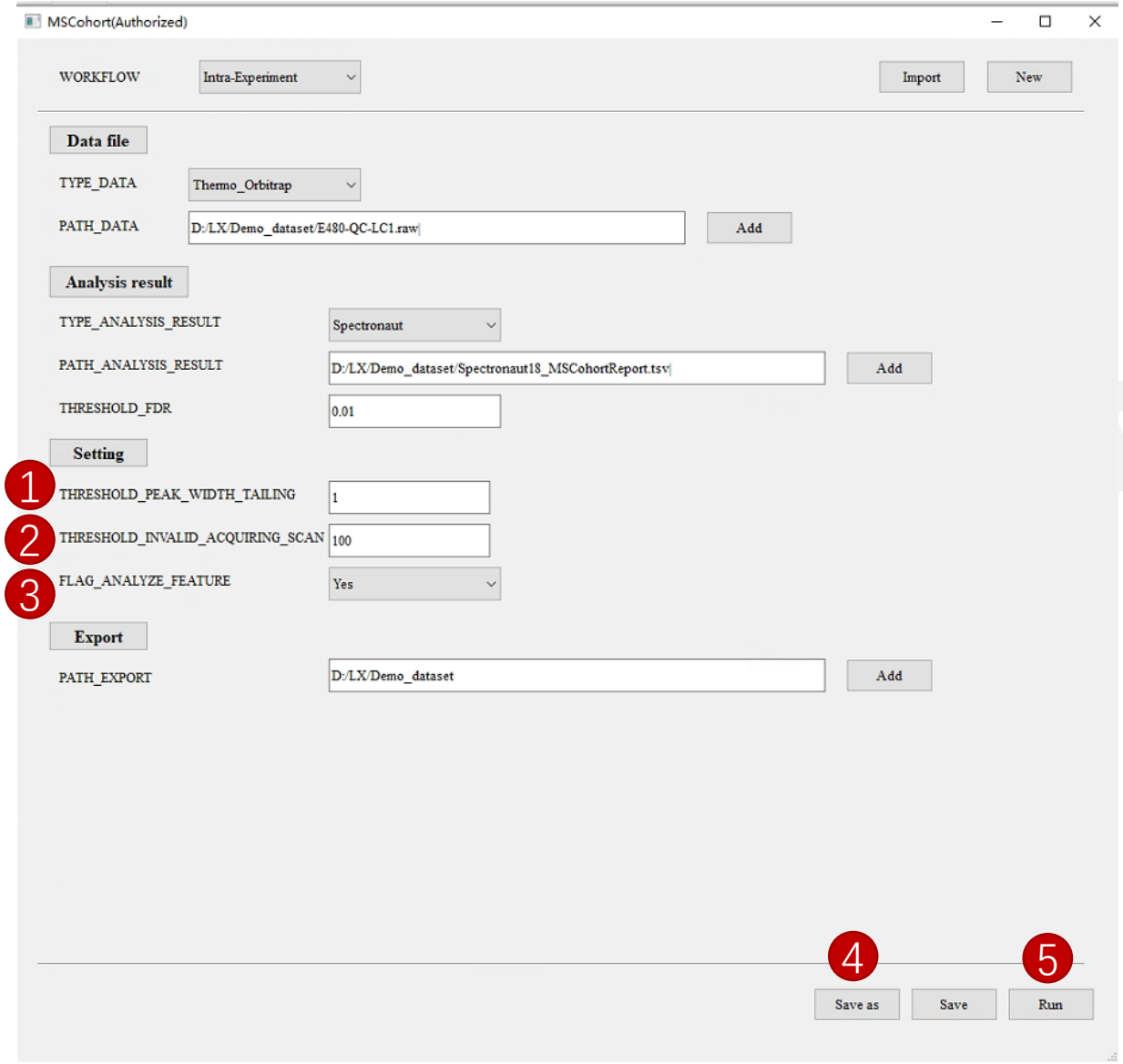
④ Click Add to select the Spectronaut customized report for MSCohort into the **PATH_ANALYSIS_RESULT**;

⑤ Click Add to 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



① **THRESHOLD_PEAK_WIDTH_TAIL** (default setting 1, user adjustable parameter according to experimental condition), 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;

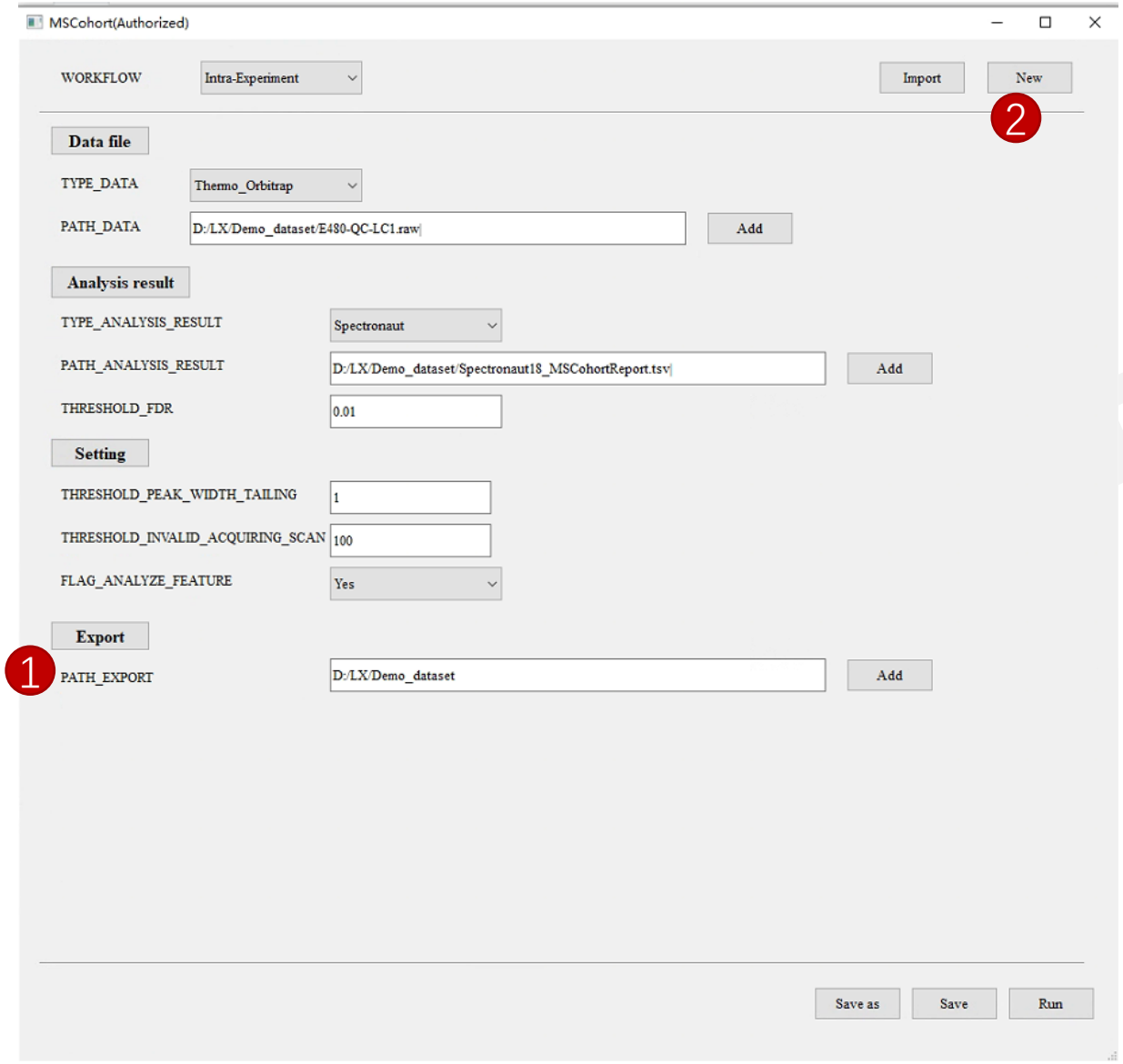
② **THRESHOLD_ INVALID_ACQUIRING** (default setting 100, user adjustable parameter according to experimental condition), set as 100 represents that the retention time length from the start to the number of identified MS2 scans per minute more than 100 as chromatographic invalid acquiring time;

③ **FLAG_ANALYZE_FEATURE**(default setting No). This analysis may take a long time for timsTOF and zenoTOF data, it is recommended to set as No for 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



Note：

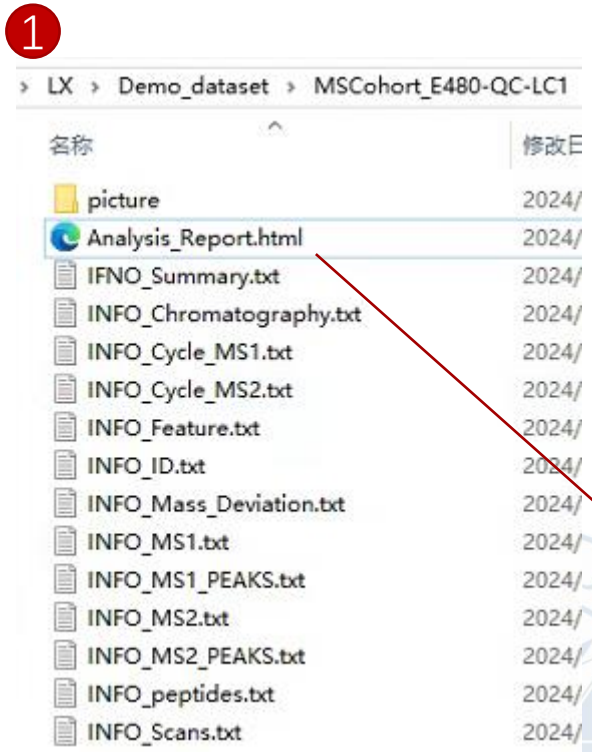
① **PATH_EXPORT：** The output MSCohort analysis results will be stored in a new folder under the PATH_EXPORT directory, named MSCohort _"filename" folder.

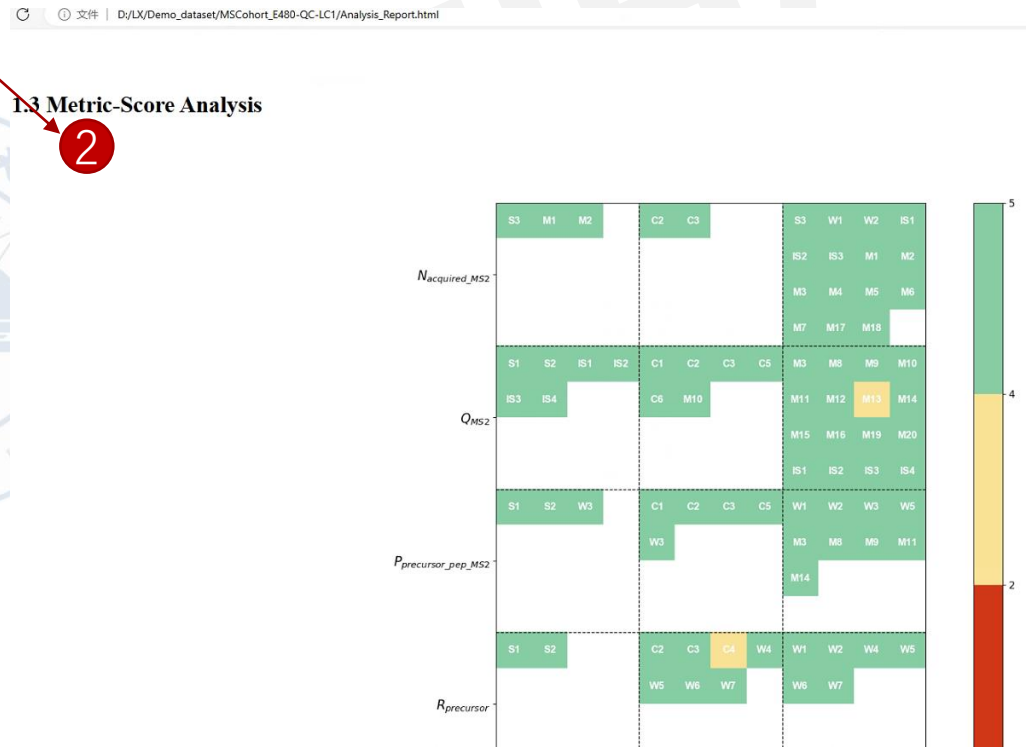
Do not save the same PATH_DATA under a same folder, the results will be overwritten.

② 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





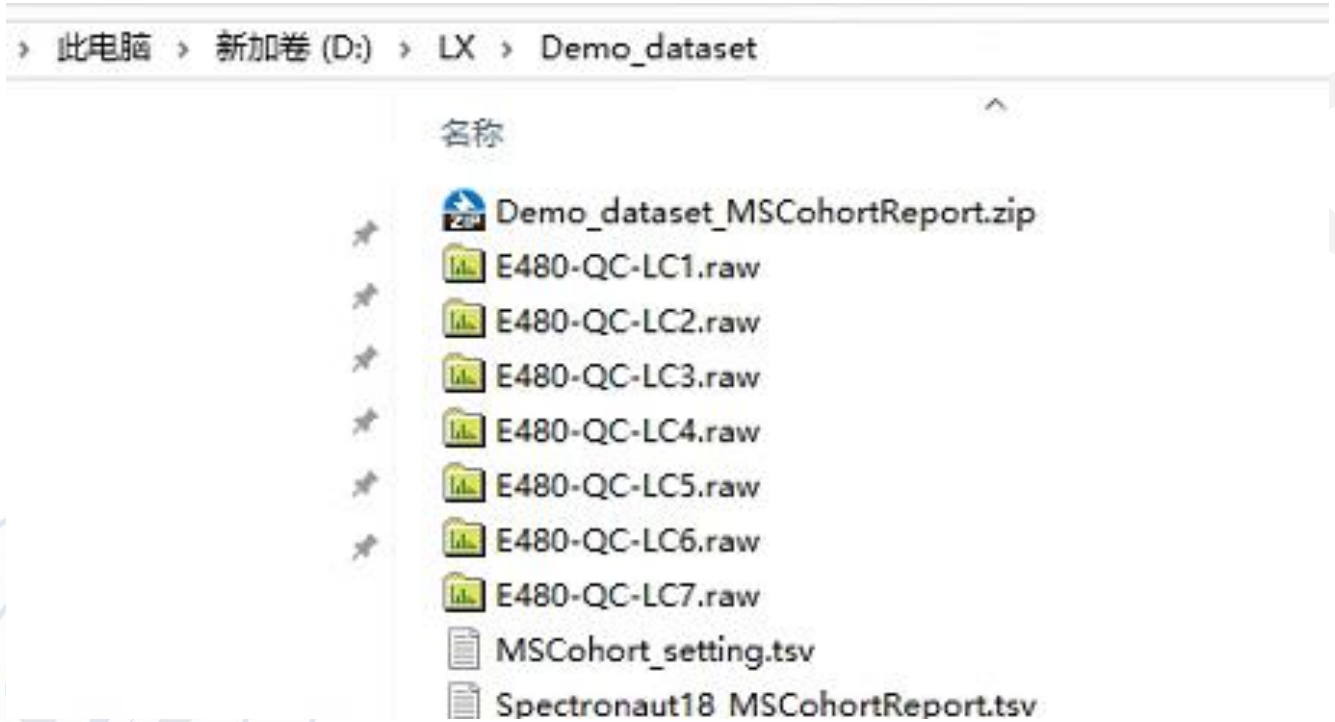
① The MSCohort results ;

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

MSCohort manual for inter-experiment analysis

BEIHANG

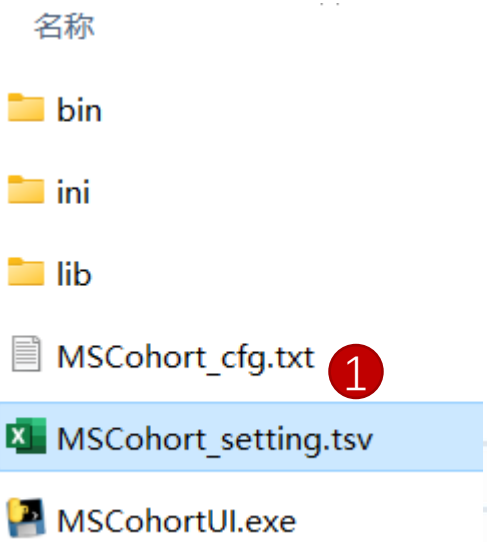
Demo dataset



Taking the 7 E480-QC raw files as an example to demonstrate the workflow of inter-experiment analysis.

Inter-experiment analysis enables the systematic quality evaluation and low-quality experiments detection for multiple DIA experiments.

1. Preparation for MSCohort_setting.tsv file

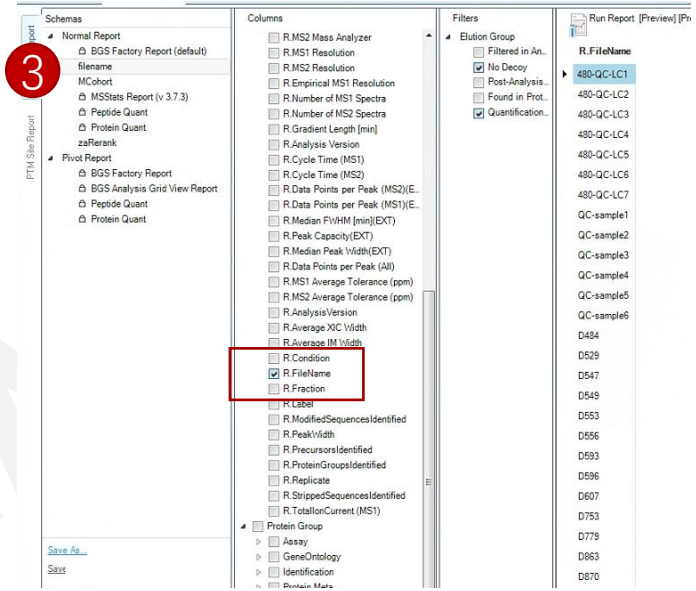


1

2

Group Name	Raw Name	Experiment	Threshold
E480-QC	E480-QC-LC1	E480-QC1	10
E480-QC	E480-QC-LC2	E480-QC2	10
E480-QC	E480-QC-LC3	E480-QC3	10
E480-QC	E480-QC-LC4	E480-QC4	10
E480-QC	E480-QC-LC5	E480-QC5	10
E480-QC	E480-QC-LC6	E480-QC6	10
E480-QC	E480-QC-LC7	E480-QC7	10

3



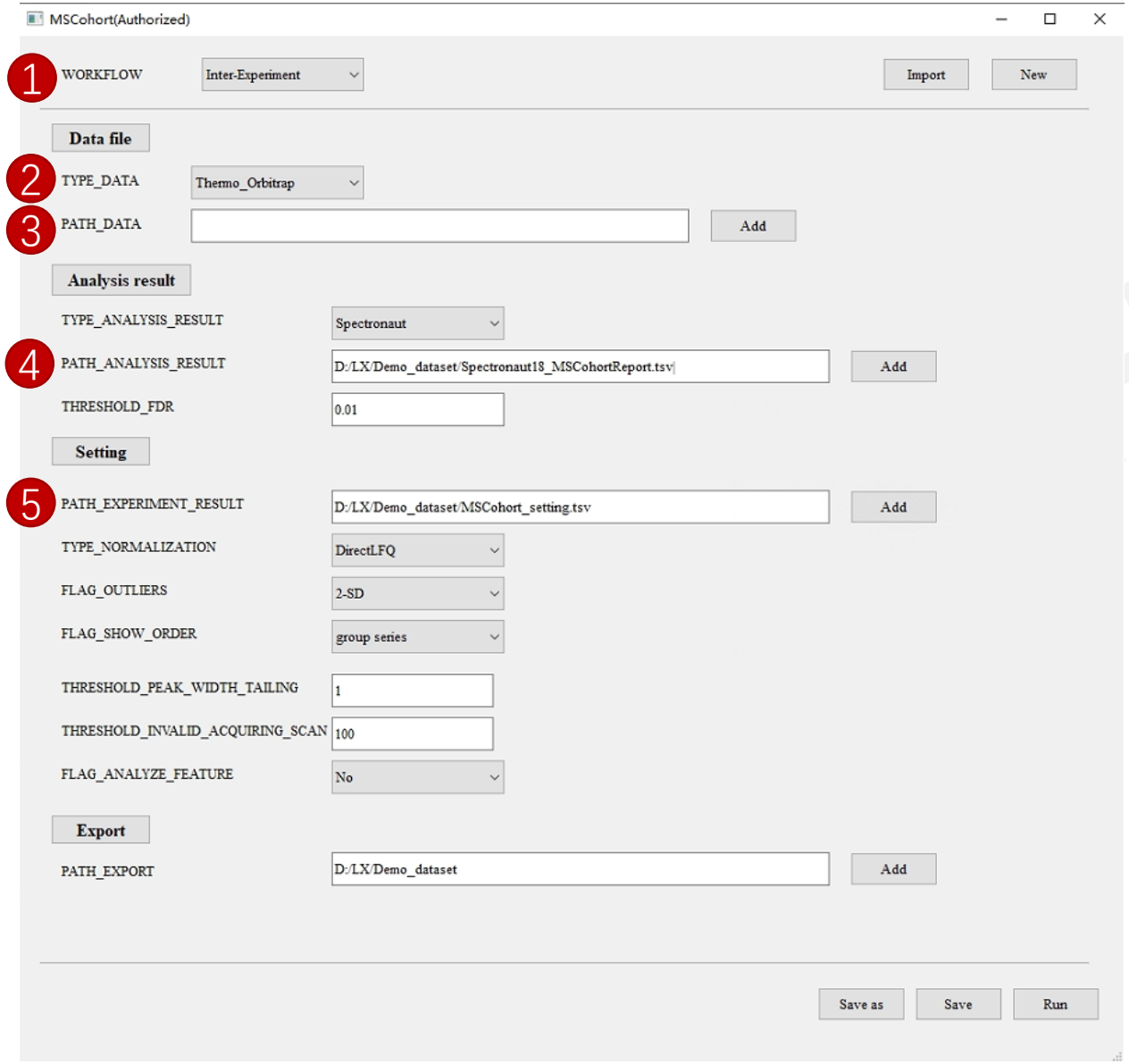
① Double clicking MSCohort_setting.tsv;

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 ;

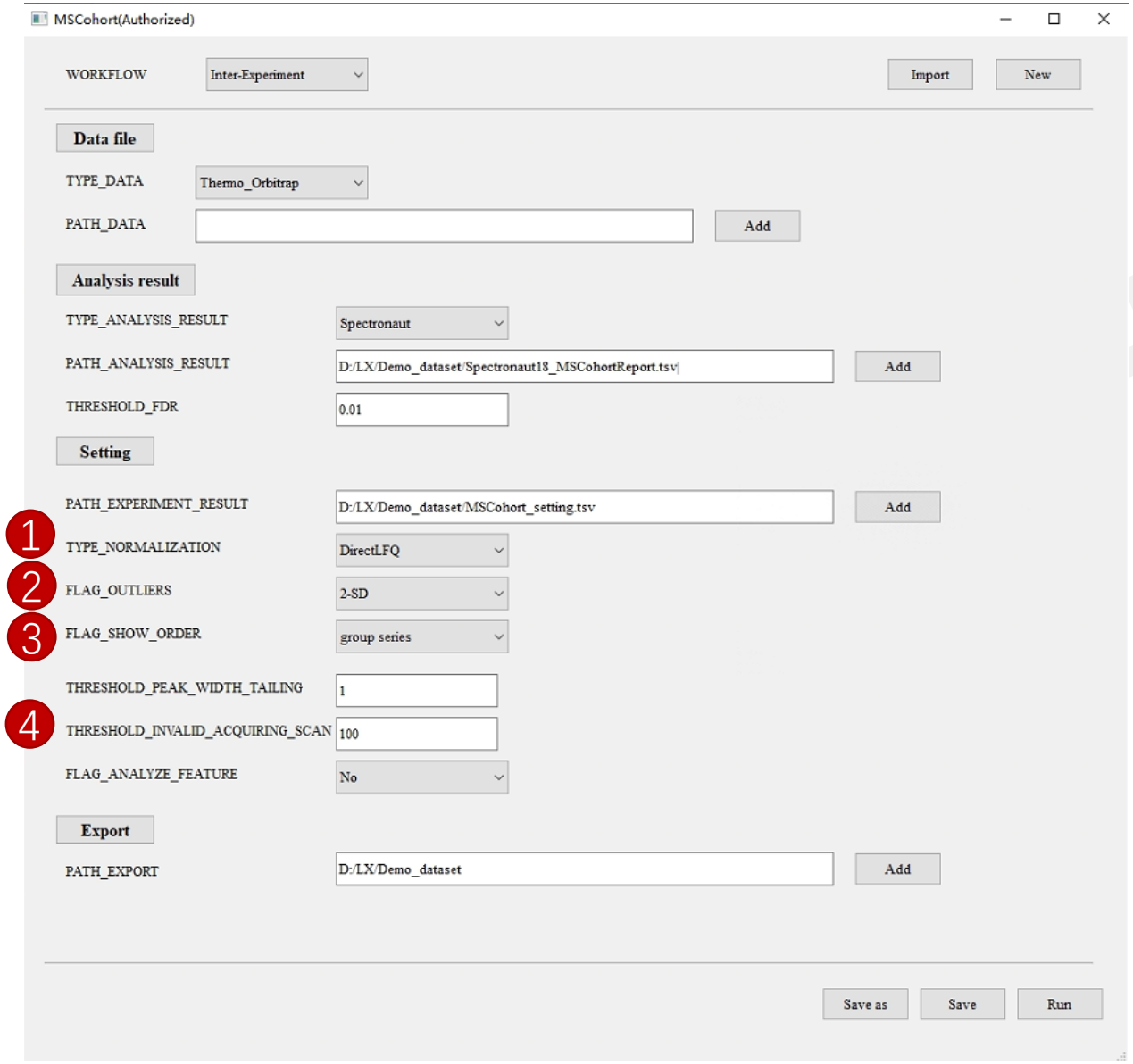
③ The **PATH_DATA** could be empty. MSCohort support the inter-experimental analysis mainly based on Spectronaut result, eliminating the need to submit raw files, which will obtain the result in a relatively **short time**. It is recommended for large cohort analysis.

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 Click Add to select the Spectronaut customized report for MSCohort into the **PATH_ANALYSIS_RESULT**;

⑤ Click Add to select the **MSCohort_setting.tsv** into the **PATH_EXPERIMENT_RESULT**;

2. Analyzing with MSCohort



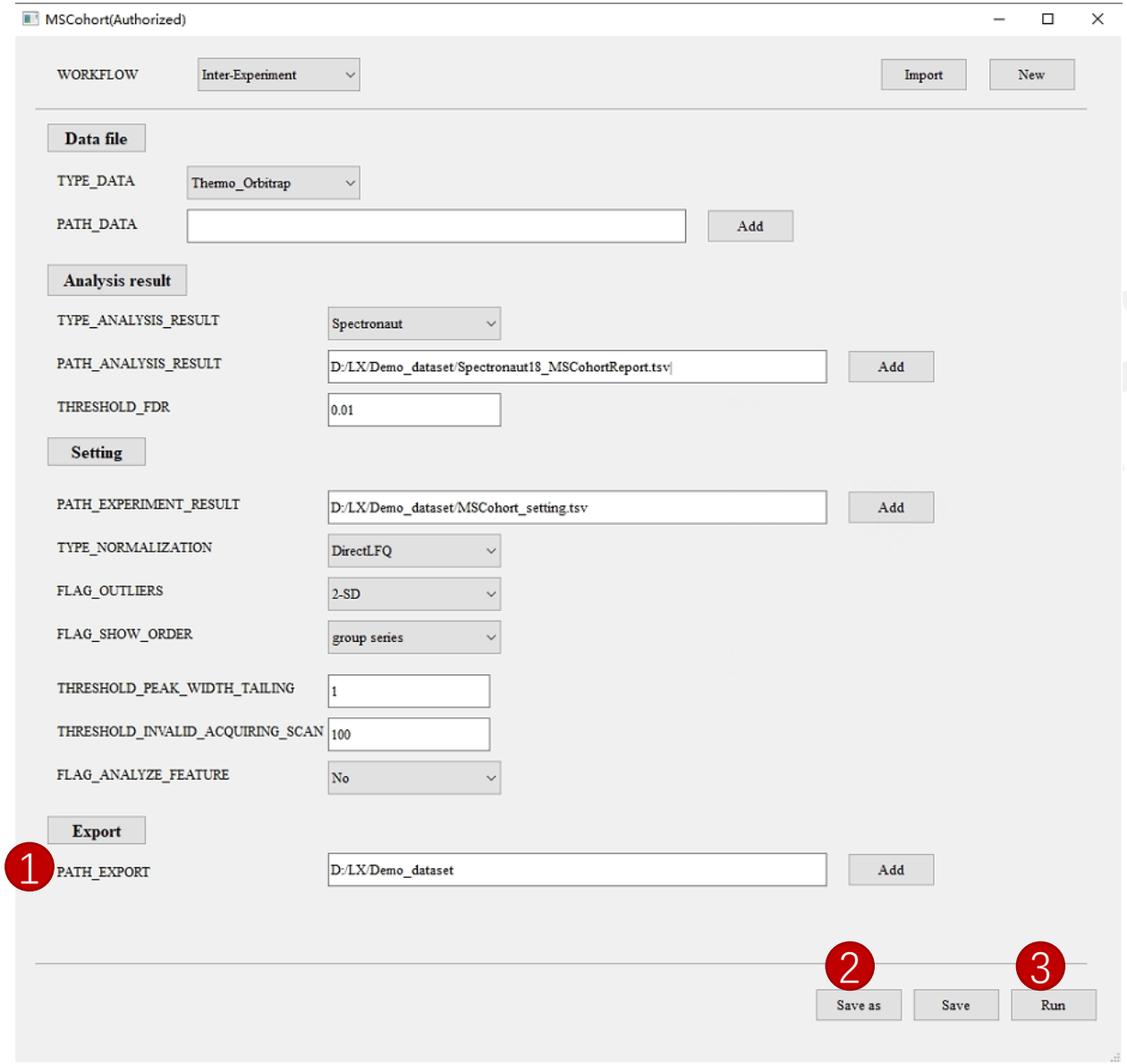
① Choose normalization strategies in **TYPE_NORMALIZATION**. MSCohort support directLFQ, MaxLFQ, and Quantile normalization (default is directLFQ);

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;

2. Analyzing with MSCohort



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.

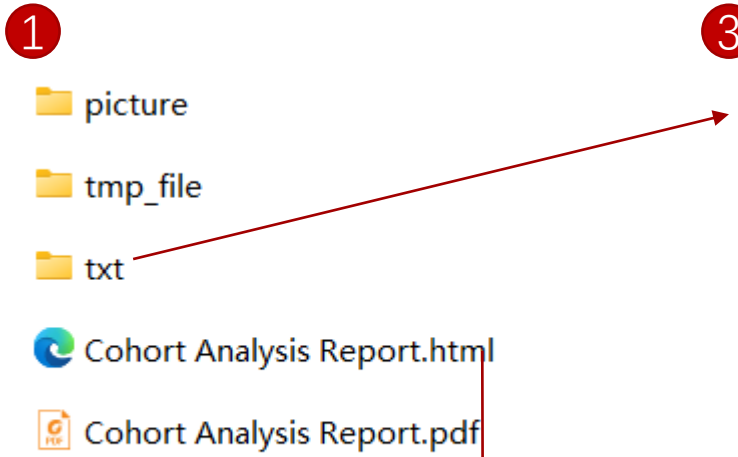
② Click **Save as** button to save the config file;

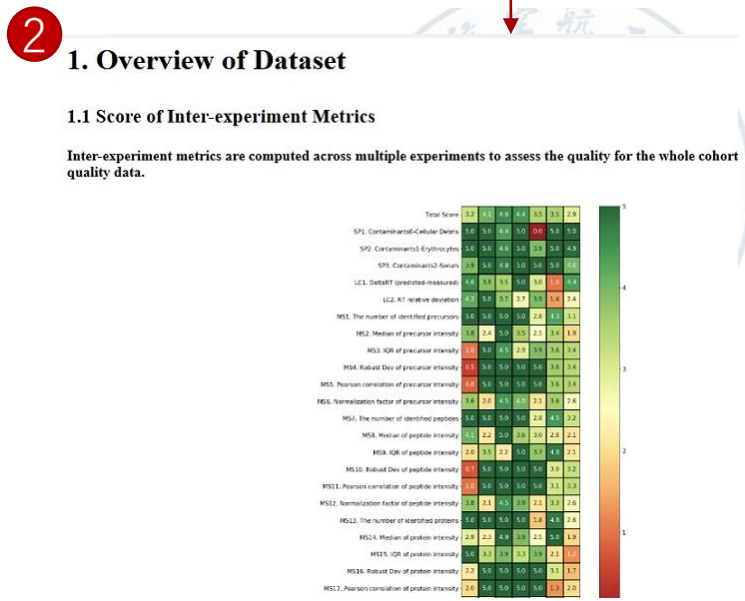
③ Click **Run** button to start the MSCohort, the progress information will be shown in the command-line interface.

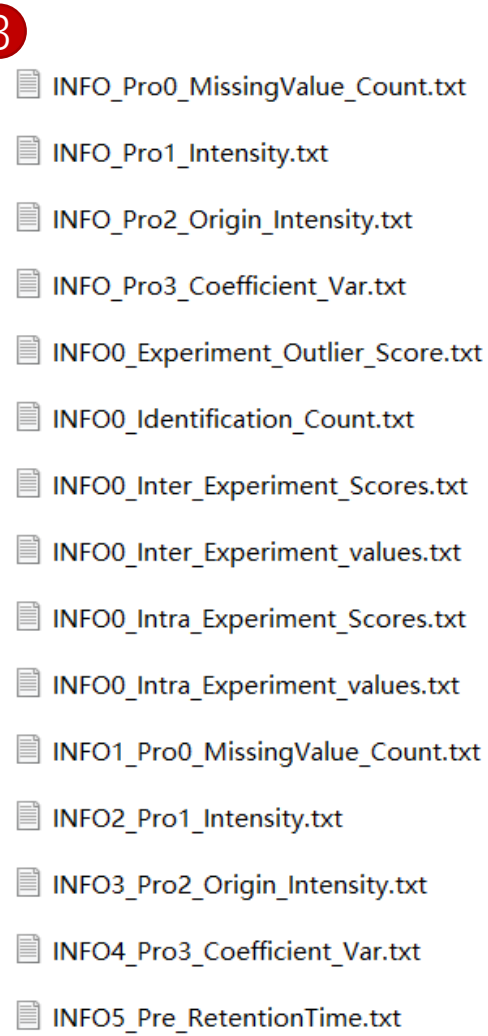
Note：

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 showed in the browser.

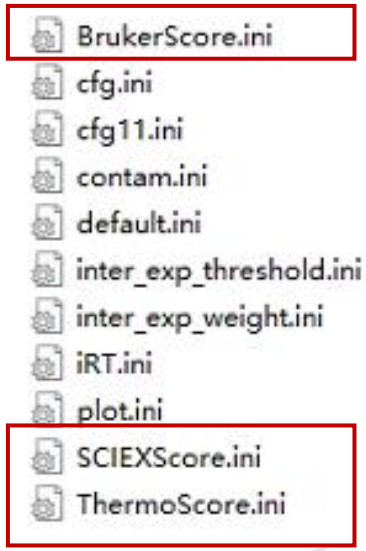
③ Double clicking **txt** folder, the outputs are also exported to simple tab-delimited text files.

Notes for modifying the scoring criteria

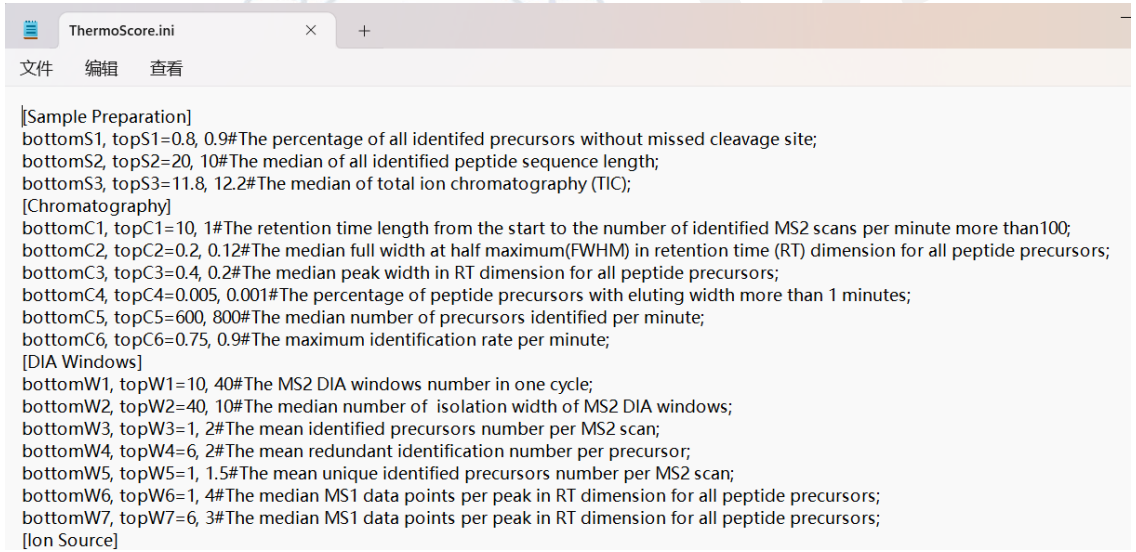
BEIHANG

1. Modifying the scoring criteria for intra-experiment analysis

1



2

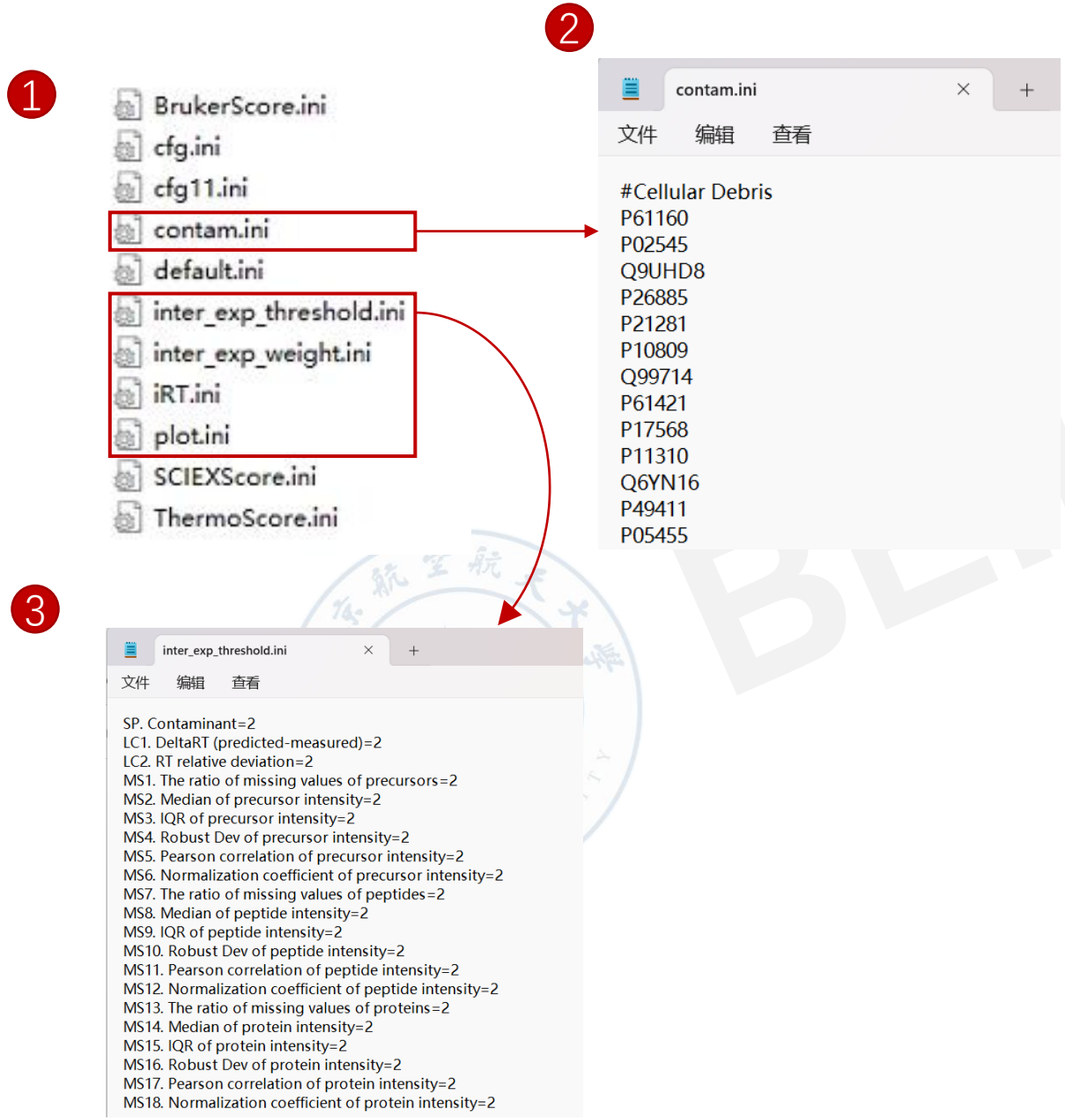


① Open the **ini** folder, there will be three parameter files related to intra-experiment scoring: ThermoScore.ini, BrukerScore.ini, SCIEXScore.ini;

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.

BEIHANG

1. Modifying the scoring criteria for inter-experiment analysis

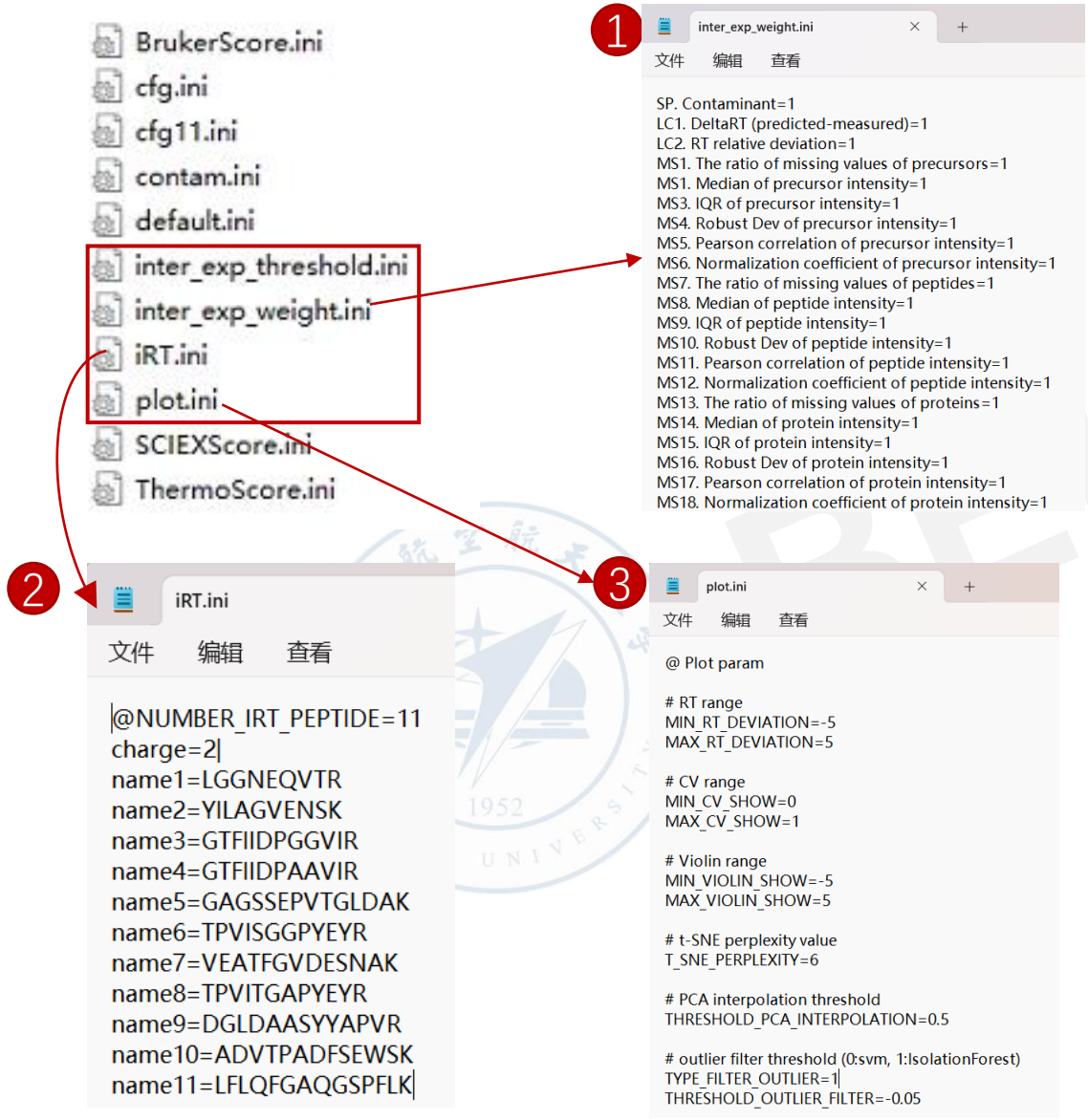


① Open the **ini** folder, there will be five parameter files related to inter-experiment scoring: contam.ini, inter_exp_threshold.ini, inter_exp_weight.ini, iRT.ini, plot.ini;

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!

BEIHANG