MetDNA

Metabolite identification and
Dysregulated **N**etwork **A**nalysis

zhu lab

Content for HELP page

1. Data preparation (XCMS and MS-DIAL)

MetDNA requires the import of the following files for metabolite identifications, including:

- (1) A MS1 peak table (.csv format),
- (2) MS2 data files (.mgf or .msp format), and
- (3) A table for sample information (.csv format).

1.1 Prepare a MS1 peak table

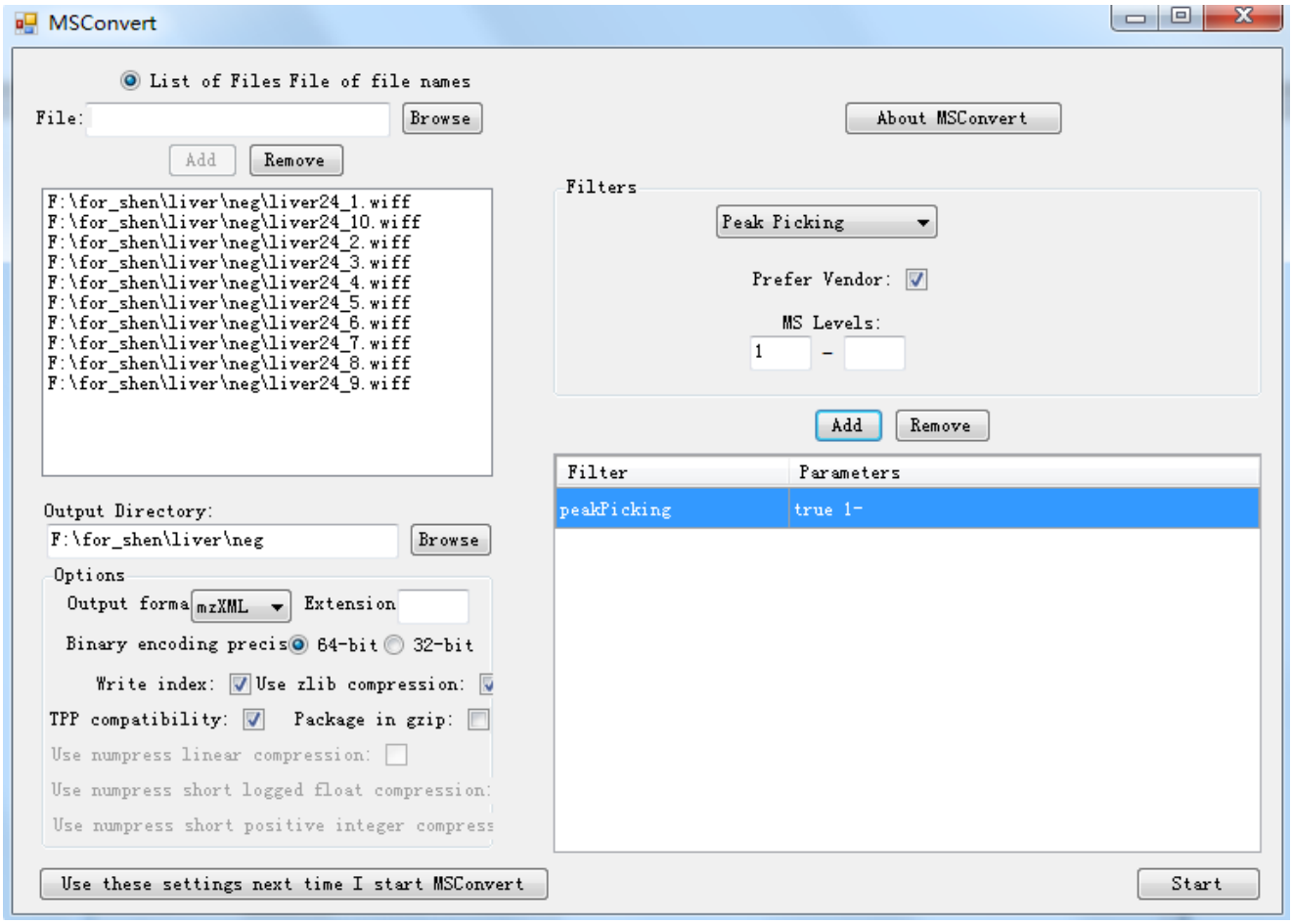
The MS1 peak table is a list of metabolic peaks with annotated m/z, retention times (RTs) and peak abundances. The MS1 peak table is generated from the raw MS files using the common peak picking software such as XCMS, MZmine, and MS-DIAL.

1.1.1 Prepare a MS1 peak table using XCMS

Here using XCMS, we provide a step-by-step instruction to generate the MS1 peak table from our mouse liver demo dataset. Please go to Demo Page to overview and download the demo data files.

MS1 data conversion

Convert raw MS data files (e.g., Sciex .wiff files) to mzXML format using ProteoWizard (http://proteowizard.sourceforge.net/) (version 3.0.6150). Please follow the conversion setting as following:



Peak picking using XCMS

In each polarity, all of 24 MS files (mzXML format) were processed using R package “xcms (https://www.bioconductor.org/packages/3.2/bioc/html/xcms.html)” (version 1.46.0) for peak detection and alignment. The mzXML files are placed in three folders named as: “24W” , “78W”, and“QC”according to their groups. Then the dataset is processed using the code shown below:

```
##set the folder containing mzXML format data as the work directory
setwd(“xxx”)
## peak detection
f.in <- list.files(pattern = '\.(mz[X]{0,1}ML|cdf)', recursive = TRUE)
xset <- xcms::xcmsSet(f.in, method = "centWave", ppm = 15,
                      snthr = 10, peakwidth = c(5, 40), mzdiff = 0.01,
                      nSlaves = 12)
##retention time correction
pdf('rector-obiwarp.pdf')
xsetc <-  xcms::retcor(xset, method = "obiwarp", plottype = "deviation",
                       profStep = 0.1)
dev.off()
## peak grouping
xset2 <- xcms::group(xsetc, bw = 5, mzwid = 0.015, minfrac = 0.5)
##gap filling
xset3 <- xcms::fillPeaks(xset2)
##peak table outputting
values <- xcms::groupval(xset3, "medret", value = "into")
values.maxo <- xcms::groupval(xset3, "medret", value = 'maxo')
values.maxint <- apply(values.maxo, 1, max)
peak.table <- cbind(name = xcms::groupnames(xset3),
                    groupmat = xcms::groups(xset3),
                    maxint = values.maxint,
                    values)
rownames(peak.table) <- NULL
write.csv(peak.table, "Peak-table.csv", row.names = FALSE)
```

As a result, a MS1 peak table named as “Peak-table” is generated For the mouse liver demo data, a total of 21,607 peaks and 18,091 peaks were detected in positive and negative mode, respectively.

Modify the “Peak-table” from XCMS MetDNA requires a specific format for the MS1 peak table (.csv fomat).

- The first column is the peak name ("name");
- The second column is the mass-to-charge ratio ("mz");
- The third column is the retention time ("rt");
- The unit of retention time must be second (not minute);
- Other columns are peak abundances of MS1 peaks in each sample.

IMPORTANT: the order and names of the first three columns must be "name" , "mz" , and "rt" .

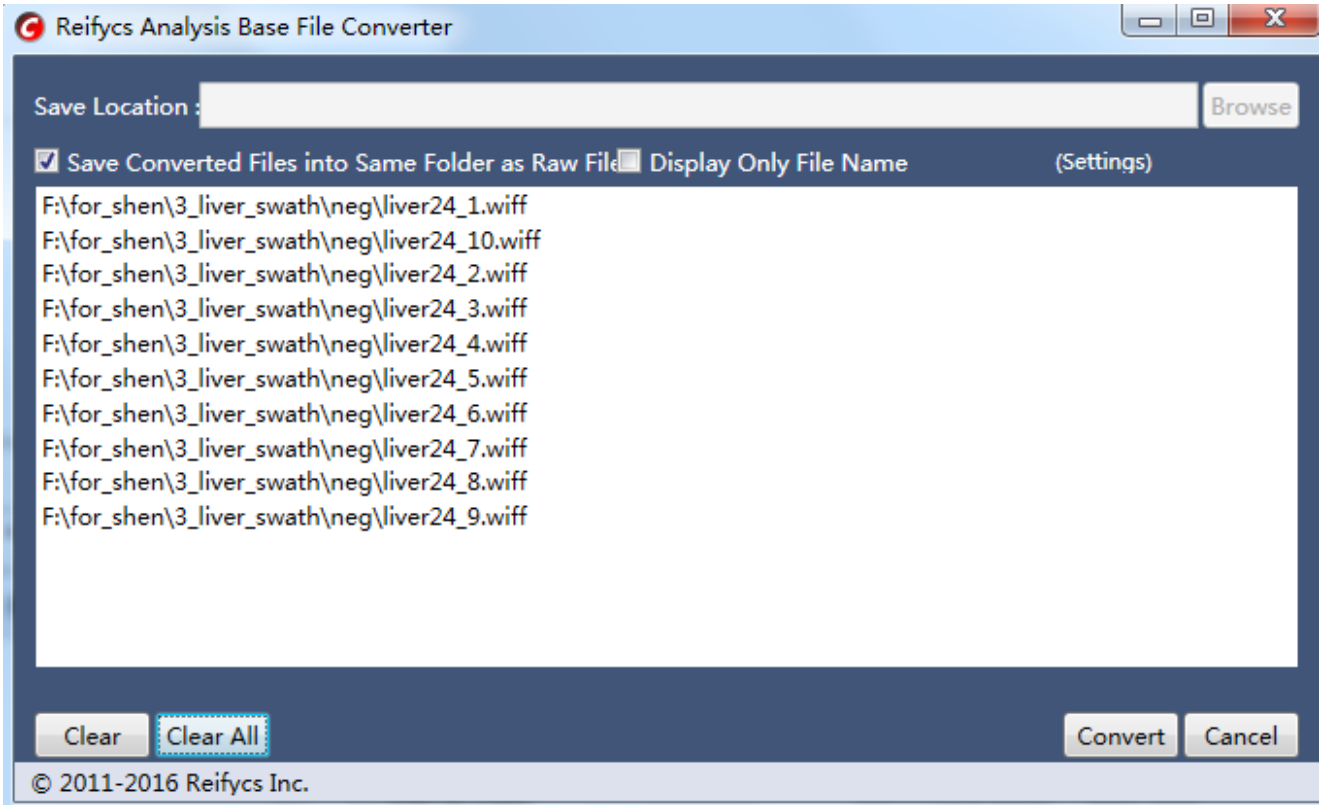
With the “Peak-table” generated from XCMS, one can modify as following:
(1) For the first 12 columns, keep columns named as “name”, “mzmed”, “rtmed”, and delete others;
(2) Rename the first three columns as “name”, “mz” and “rt”;
(3) Name the peak table as “Peak_Table_POS” or “Peak_Table_NEG” according to the polarity.

The final generated MS1 peak table should look like:

name	mz	rt	QC01	QC02	QC03	QC04	liver24_1
M60T1338	60.0176	1337.505	459.186	1199.904	673.0533	488.592	869.8944
M60T1373	60.01767	1372.66	842.856	1198.912	533.136	523.075	1225.312
M60T1358	60.01764	1358.5	1603.185	1761.687	1610.477	753.069	675.8267
M60T1275	60.0176	1274.51	954.615	731	1354.92	1236.576	2307.396
M60T1308	60.01763	1307.78	894.4	1008.175	995.976	833.341	1728.661
M60T1234	60.0176	1234.18	687.852	797.688	720.576	2494.917	1536.731
M60T1039	60.01765	1039.05	1518.345	1745.739	1640.878	745.544	1566.991
M60T1198	60.0176	1198.23	1986.282	1644.624	2793.535	1709.82	914.5483
M60T1174	60.01765	1174.32	1520.532	1415.36	1503.84	869.92	1668.364
M60T1072	60.01762	1071.81	1458.491	2889.457	968.06	2925.885	2600.678
M60T1090	60.01767	1090.015	1691.91	1714.7	1814.307	794.448	1848.764
M60T1110	60.01763	1110.03	845.5767	2598.908	1565.15	1345.429	934.1033
M60T1158	60.01765	1158.34	1534.429	769.824	1983.6	1917.153	826.524
M60T1134	60.01704	1133.99	1864.782	2080.504	786.186	1652.333	1581.616
M60T33	60.01762	32.615	1317.16	769.3423	1349.24	918.3264	1433.679

1.1.2 Prepare a MS1 peak table using MS-DIAL

MS1 data conversion Convert raw MS data files to abf format using Analysis Base File Converter (version 1.1.0.0)



Process data with MS-DIAL

- Generate a new project in MS-DIAL (http://prime.psc.riken.jp/Metabolomics_Software/MS-DIAL/) (version 2.56)
- In each polarity, select the appropriate parameters according to the experimental design. Make sure to export the “Raw data matrix (Area)” and “Representative spectra” in .msp format. See “Prepare MS2 data files” for details.
- In our published dataset No.3 (negative mode), we use the parameters as following:

```
Data collection parameters
Retention time begin    0
Retention time end      23
Mass range begin        60
Mass range end  1200

Centroid parameters
MS1 tolerance   0.01
MS2 tolerance   0.05
Peak detection-based    True

Peak detection parameters
Smoothing method        LinearWeightedMovingAverage
Smoothing level 3
Minimum peak width      5
Minimum peak height     1000

Peak spotting parameters
Mass slice width        0.1
Exclusion mass list (mass & tolerance)

Deconvolution parameters
Peak consideration      Both
Sigma window value      0.5
Exclude after precursor True

MSP file and MS/MS identification setting
MSP file
Retention time tolerance        0.5
Accurate mass tolerance (MS1)   0.01
Accurate mass tolerance (MS2)   0.05
Identification score cut off    80

Text file and post identification (retention time and accurate mass based) setting
Text file
Retention time tolerance        0.1
Accurate mass tolerance 0.01
Identification score cut off    85

Advanced setting for identification
Relative abundance cut off      0
Top candidate report    True

Adduct ion setting
[M-H]-

Alignment parameters setting
Reference file  F:\for_shen\liver_swath\neg\liver24_1.abf
Retention time tolerance        0.1
MS1 tolerance   0.025
Retention time factor   0.5
MS1 factor      0.5
Peak count filter       25
QC at least filter      True

Tracking of isotope labels
Tracking of isotopic labels     FALSE
```

Modify the MS1 peak table

MetDNA requires a specific format for the MS1 peak table (.csv format).

- The first column is the peak name ("name").
- The second column is the mass-to-charge ratio ("mz").
- The third column is the retention time ("rt").
- The unit of retention time must be second (not minute).
- Other columns are peak abundances of MS1 peaks in each sample.

IMPORTANT: the order and names of the first three columns must be "name" , "mz" , and "rt" .
With the “Raw data matrix (Area)” file generated from MS-DIAL, one can modify as following:
(1) Use EXCEL to open the “Area.txt” file, delete the first 3 rows, then delete 4-17 columns and only keep columns named as “Alignment ID”, “Average Rt (min)” and “Average Mz”.

(2) Exchange the orders of the second and third columns. Change the unit of retention time to second.

(3) Rename the first three columns as “name”, “mz” and “rt”.

(4) Name the peak table as “Peak_Table_POS” or “Peak_Table_NEG” according to the polarity.

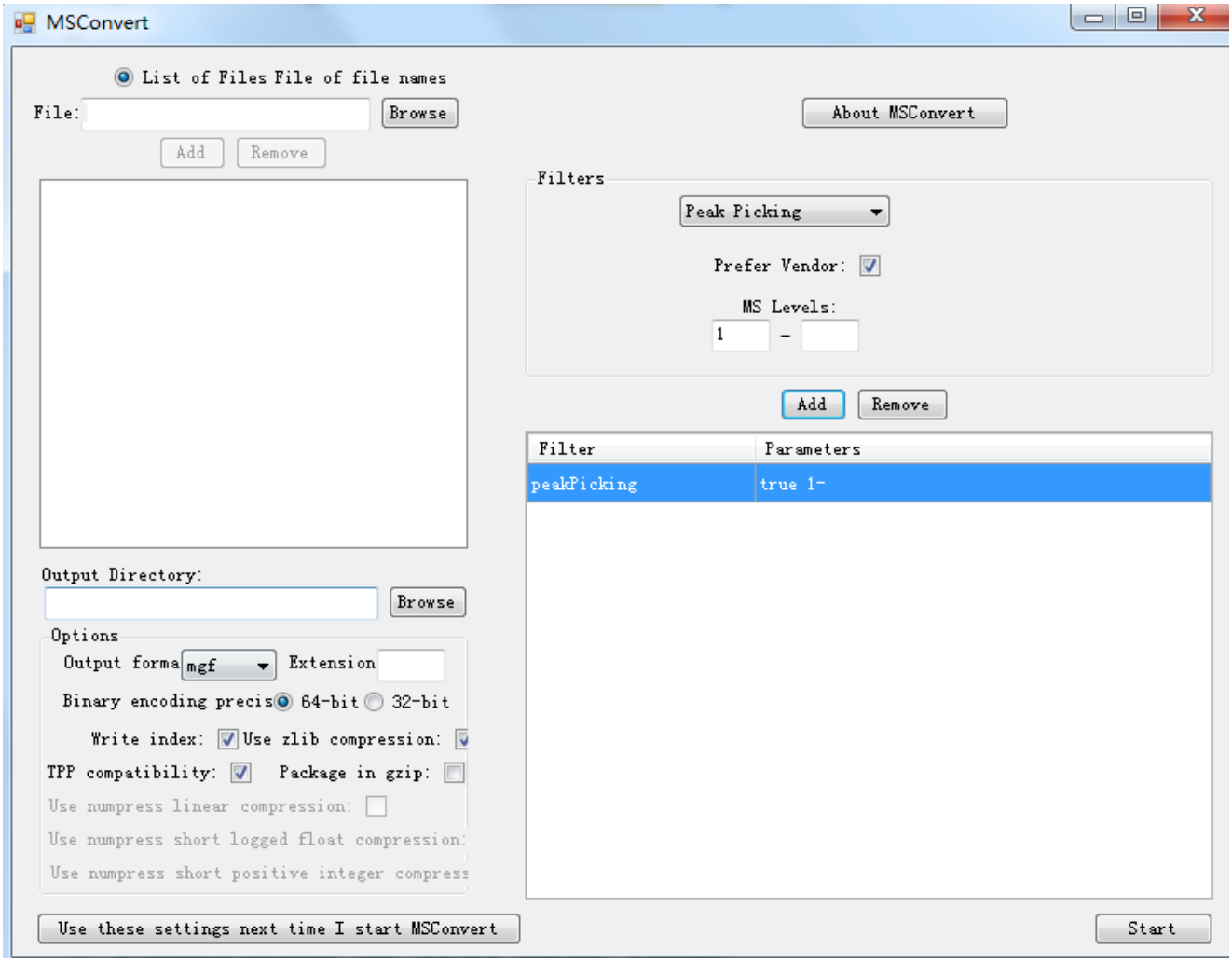
The final generated MS1 peak table should look like:

name	mz	rt	QC01	QC02	QC03	QC04	liver24_1
M60T1338	60.0176	1337.505	459.186	1199.904	673.0533	488.592	869.8944
M60T1373	60.01767	1372.66	842.856	1198.912	533.136	523.075	1225.312
M60T1358	60.01764	1358.5	1603.185	1761.687	1610.477	753.069	675.8267
M60T1275	60.0176	1274.51	954.615	731	1354.92	1236.576	2307.396
M60T1308	60.01763	1307.78	894.4	1008.175	995.976	833.341	1728.661
M60T1234	60.0176	1234.18	687.852	797.688	720.576	2494.917	1536.731
M60T1039	60.01765	1039.05	1518.345	1745.739	1640.878	745.544	1566.991
M60T1198	60.0176	1198.23	1986.282	1644.624	2793.535	1709.82	914.5483
M60T1174	60.01765	1174.32	1520.532	1415.36	1503.84	869.92	1668.364
M60T1072	60.01762	1071.81	1458.491	2889.457	968.06	2925.885	2600.678
M60T1090	60.01767	1090.015	1691.91	1714.7	1814.307	794.448	1848.764
M60T1110	60.01763	1110.03	845.5767	2598.908	1565.15	1345.429	934.1033
M60T1158	60.01765	1158.34	1534.429	769.824	1983.6	1917.153	826.524
M60T1134	60.01704	1133.99	1864.782	2080.504	786.186	1652.333	1581.616
M60T33	60.01762	32.615	1317.16	769.3423	1349.24	918.3264	1433.679

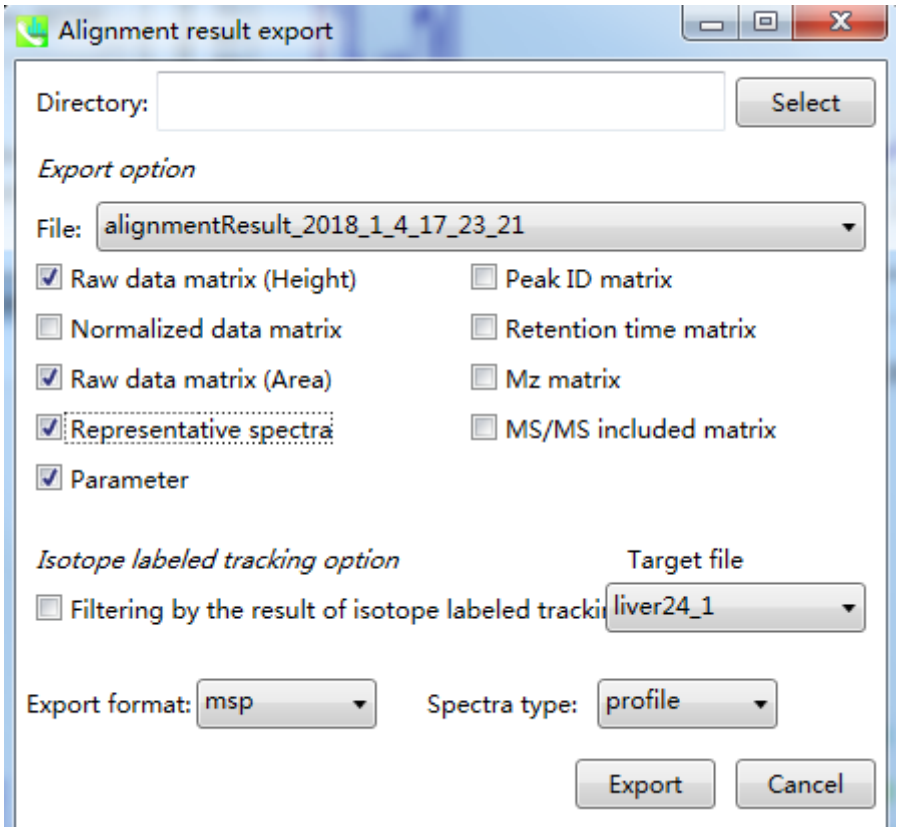
1.2 Prepare a MS1 peak table

For users using XCMS, taken our mouse liver demo data as an example, convert 6 raw MS2 data files (.wiff format; i.e., QC60-300Da_NEG_1 etc.) in each ionization polarity to mgf format using ProteoWizard (http://proteowizard.sourceforge.net/) (version3.0.6150).

Please follow the conversion setting as following:



For users using MS-DIAL, please export the "Representative spectra" in .msp format. Follow the instructions below:



1.3 Prepare a "sample information" file

The sample information file (.csv format) is designed to describe the sample group information. The first column is named as "sample.name" , while the second one is named as “group”. The sample information file should look like:

sample.name	group
QC01	QC
QC02	QC
QC03	QC
QC04	QC
liver24_10	24W
liver24_1	24W
liver24_2	24W
liver24_3	24W
liver24_4	24W
liver24_5	24W
liver24_6	24W
liver24_7	24W
liver24_8	24W
liver24_9	24W
liver78_10	78W
liver78_1	78W

The "sample.name" column in sample information file must be the EXACTLY same as the sample names in the MS1 peak table.

name	mz	rt	QC01	QC02	QC03	QC04	liver24_1	liver24_10	liver24_2	liver24_3
M60T1338	60.0176	1337.505	459.186	1139.984	673.0533	488.592	869.89444	1177.83273	790.18125	604.76
M60T1373	60.01767	1372.66	842.856	1198.912	533.136	523.075	1225.312	509.168	696.15	1136.16857
M60T1358	60.01764	1358.5	1603.185	1761.687	1610.477	753.069	675.82667	1170.59333	1611.40909	969.736667
M60T1275	60.0176	1274.51	954.615	731	1354.92	1236.576	2307.3956	1326.36	2114.54875	578.85
M60T1308	60.01763	1307.78	894.4	1008.175	995.976	833.341	1728.6609	879.84	671.32	650.934
M60T1234	60.0176	1234.18	687.852	797.688	720.576	2494.917	1536.7311	721.108	2162.62667	734.832
M60T1039	60.01765	1039.05	1518.345	1745.739	1640.878	745.544	1566.9911	1264.66375	1383.2	1922.9
M60T1198	60.0176	1198.23	1986.282	1644.624	2793.535	1709.82	914.54833	1388.4	1770.61455	906.95
M60T1174	60.01765	1174.32	1520.532	1415.36	1503.84	869.92	1668.3643	1601.45778	1448.49857	1434.88
M60T1072	60.01762	1071.81	1458.491	2889.457	968.06	2925.885	2600.6775	1158.988	2333.90182	724.716
M60T1090	60.01767	1090.015	1691.91	1714.7	1814.307	794.448	1848.7636	2737.98571	1767.35375	1449.596
M60T1110	60.01763	1110.03	845.5767	2598.908	1565.15	1345.429	934.10333	789.66	1417.39455	1523.61818
M60T1158	60.01765	1158.34	1534.429	769.824	1983.6	1917.153	826.524	2266.12444	2015.28364	2597.4
M60T1134	60.01704	1133.99	1864.782	2080.504	786.186	1652.333	1581.616	1785.27273	2700.84563	2219.74

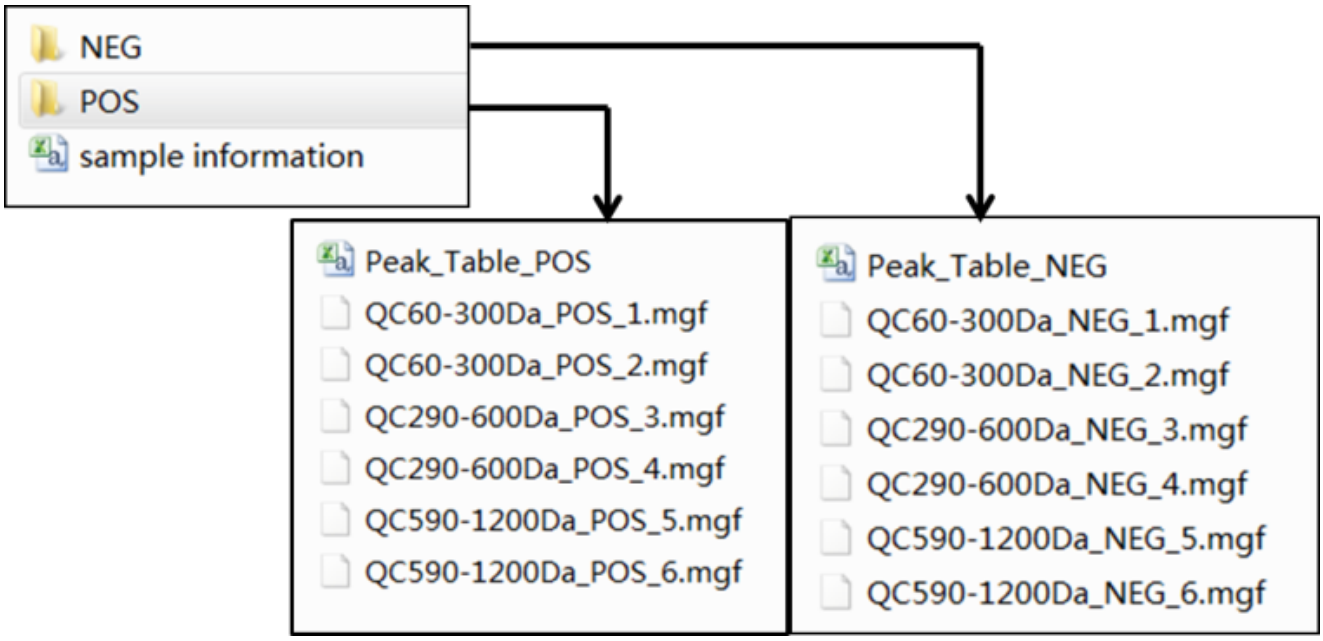
Sample names in MS1 peak table

sample.name	group
QC01	QC
QC02	QC
QC03	QC
QC04	QC
liver24_10	24W
liver24_1	24W
liver24_2	24W
liver24_3	24W
liver24_4	24W
liver24_5	24W
liver24_6	24W
liver24_7	24W
liver24_8	24W
liver24_9	24W
liver78_10	78W

Sample names in sample information file

1.4 MetDNA demo data files

The zip file for "MetDNA Demo Data" can be downloaded here (/metdna/demo#download).



1.5 Important notes for data preparation

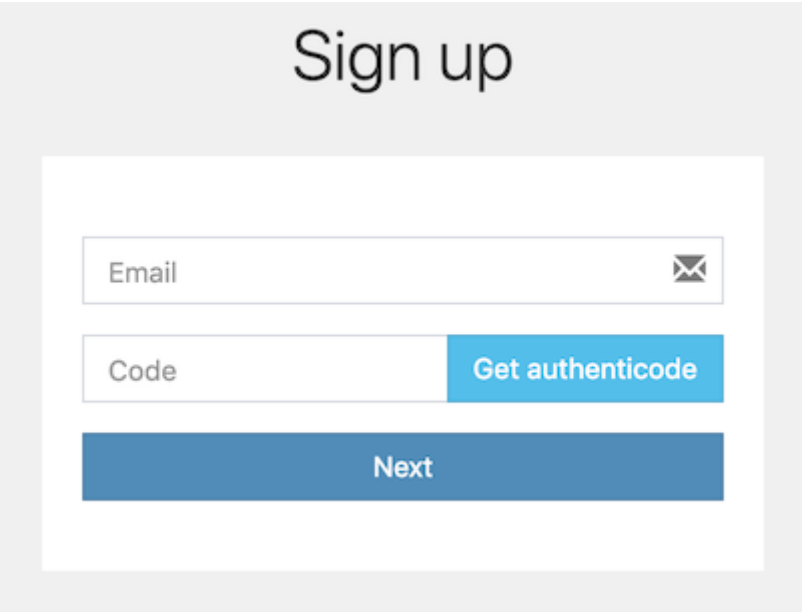
(1) In the MS1 peak table, make sure that no "-" or blank appears in the peak name or sample name. If there are some symbols that cannot be recognized by our program, the data processing may be failed.
(2) The "sample.name" column in sample information file must be the EXACTLY same as the sample names in the MS1 peak table.
(3) If you want to process both positive and negative datasets together, make sure that your sample names in the two peak tables are the same.
(4) The groups of samples should not be only one group.
(5) Please make sure that sample information (.csv format) and MS1 peak table (.csv format) are separated by comma. Because in some countries or regions (European and some French-speaking regions), the default separator is semicolon. You can open the sample information or MS1 peak table with notepad or other text editors to check whether they are separated by comma.

2. The Use of MetDNA Webserver

Please click http://MetDNA.zhulab.cn/ (http://metdna.zhulab.cn/) to visit MetDNA webserver. Currently, common web browsers such as Chrome, Edge and Firefox are supported. Internet Explorer (IE) didn't give sufficient test.

2.1 Register on the website

Click “Analysis” tab to start your data analysis. First, new users need to register on our website, only educational mailboxes are supported. After registration, users can sign in with their user name.

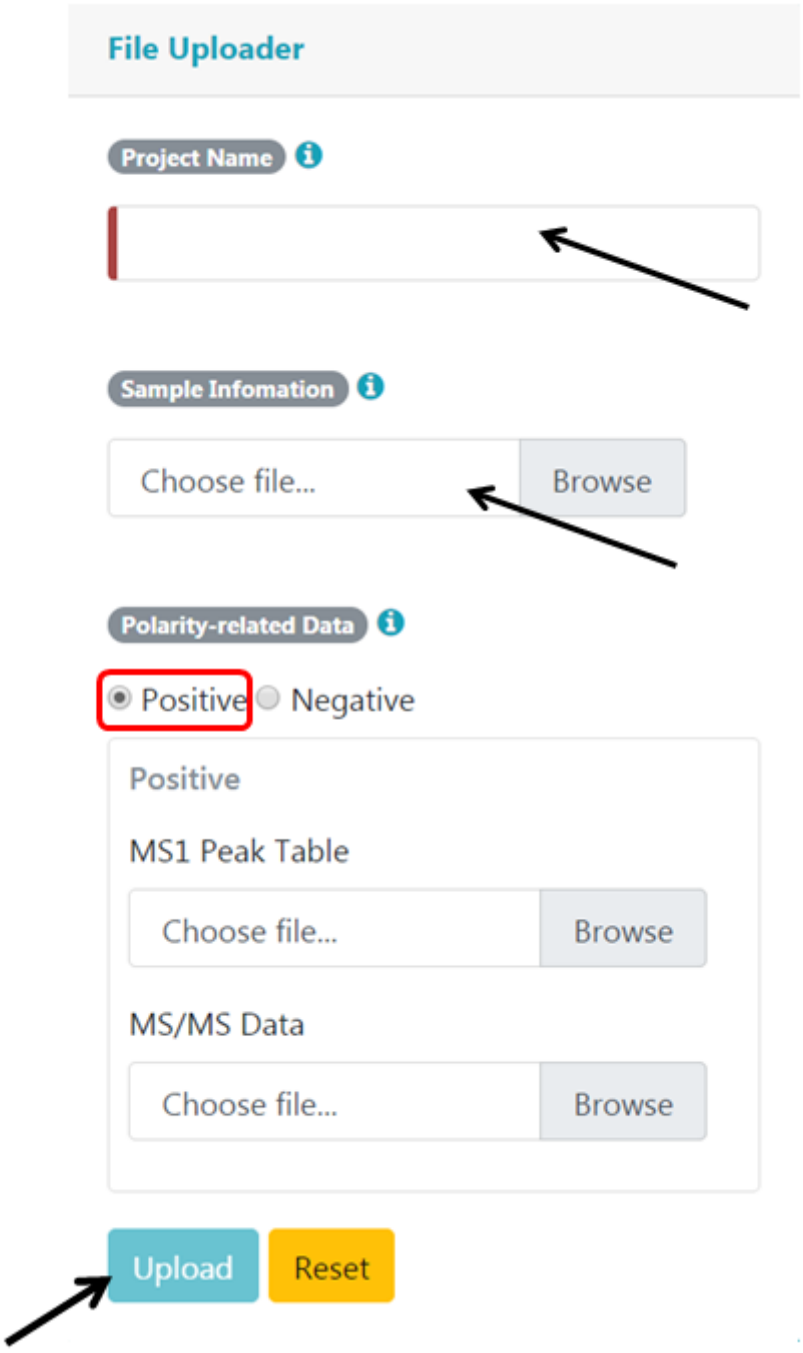


2.2 Check data format

MetDNA requires the import of the following files, including:
(1) A MS1 peak table (.csv format),
(2) MS2 data files (.mgf or .msp format), and
(3) A table for sample information (.csv format).

2.3 Upload data files

One polarity or two polarities(positive and negative) are supported in MetDNA.



One polarity (positive or negative)

(1) Name your project

(2) Choose the sample information file

(3) Choose the polarity according to your data

(4) Choose the MS1 peak table and MS/MS data

(5) Click upload button to upload your data files **Two polarities (positive and negative)**

(1) Name your project

(2) Choose the sample information file

(3) Choose the POSITIVE polarity

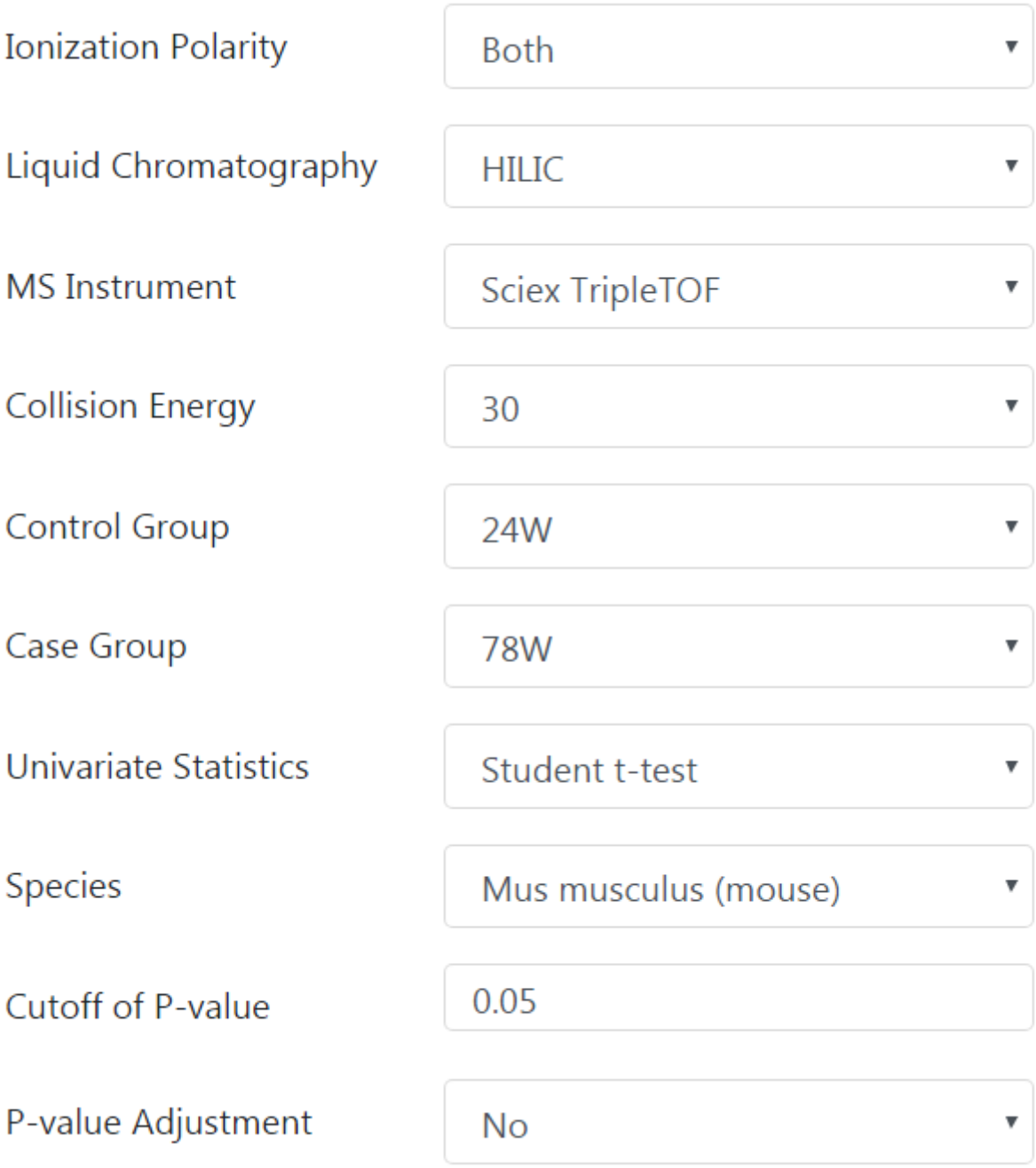
(4) Choose the MS1 peak table and MS/MS data for positive polarity

(5) Click upload button to upload your data files

(6) Choose the NEGATIVE polarity
(7) Choose the MS1 peak table and MS/MS data for negative polarity
(8) Click upload button to upload your data files

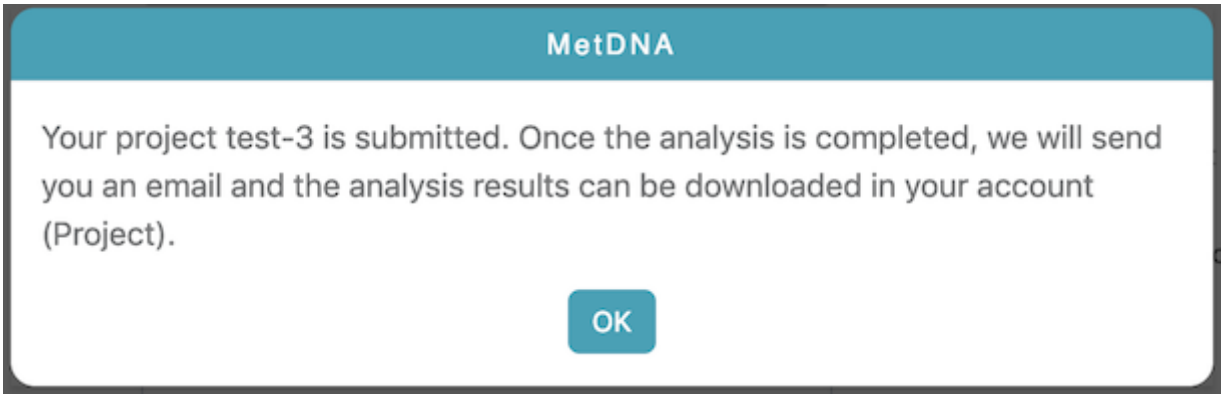
2.4 Set parameters for data processing

(1) Ionization Polarity: Choose "Positive" , "Negative" , or "Both" according to the MS data.
(2) Liquid Chromatography: Choose "HILIC" or "RP" according to the column condition.
(3) MS instrument: Select the instrument for data acquisition.
(4) Collison Energy: Select the CE value for MS2 data acquisition.
(5) Control group: Select the control group.
(6) Case group: The case group are compared with the control group to discover the dysregulated peaks.
(7) Univariate Statistics: Choose "Student t-test" or "Wilcox test" to analyze the data.
(8) Species: Choose species depends on your experiment, now we support 16 different kinds of species, see below for details.
(9) Cutoff of P-value: Define a maximum p-value to decide which peaks are significantly changed.
(10) P-value Adjustment: Choose "yes" or "no" to decide whether to correct p-value with FDR.
Supported Species: Homo sapiens (human), Rattus norvegicus (rat), Mus musculus (mouse), Bos taurus (cow), Gallus gallus (chicken), Drosophila melanogaster (fruit fly), Danio rerio (zebrafish), Escherichia coli K-12 MG1655, Caenorhabditis elegans (nematode), Trypanosoma brucei, Pseudomonas putida KT2440, Plasmodum falciparum 3D7 (Malaria), Arabidopsis thaliana (thale cress), Schistosoma mansoni, Saccharomyces cerevisiae (yeast), Synechococcus elongatus



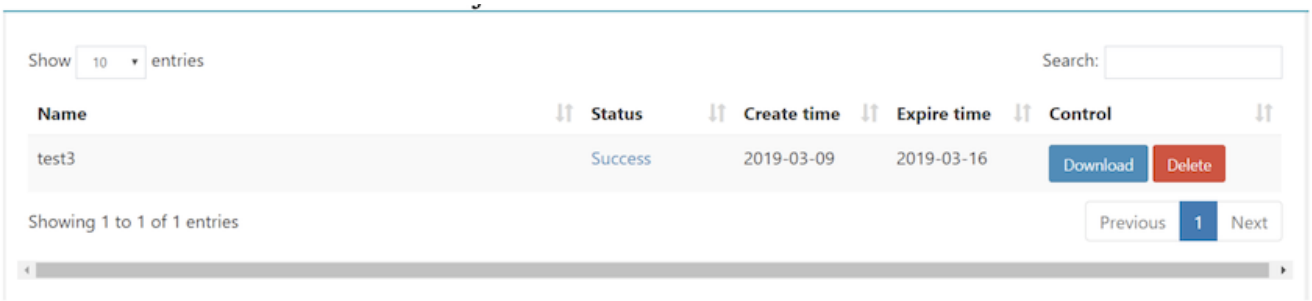
2.5 Submit

After you click the “submit” button, your data will be submitted to MetDNA for processing. A confirmation message will pop out on your screen and a confirmation email from metdna@sioc.ac.cn will send to your email address.



3. Receive an Email and Download the Result

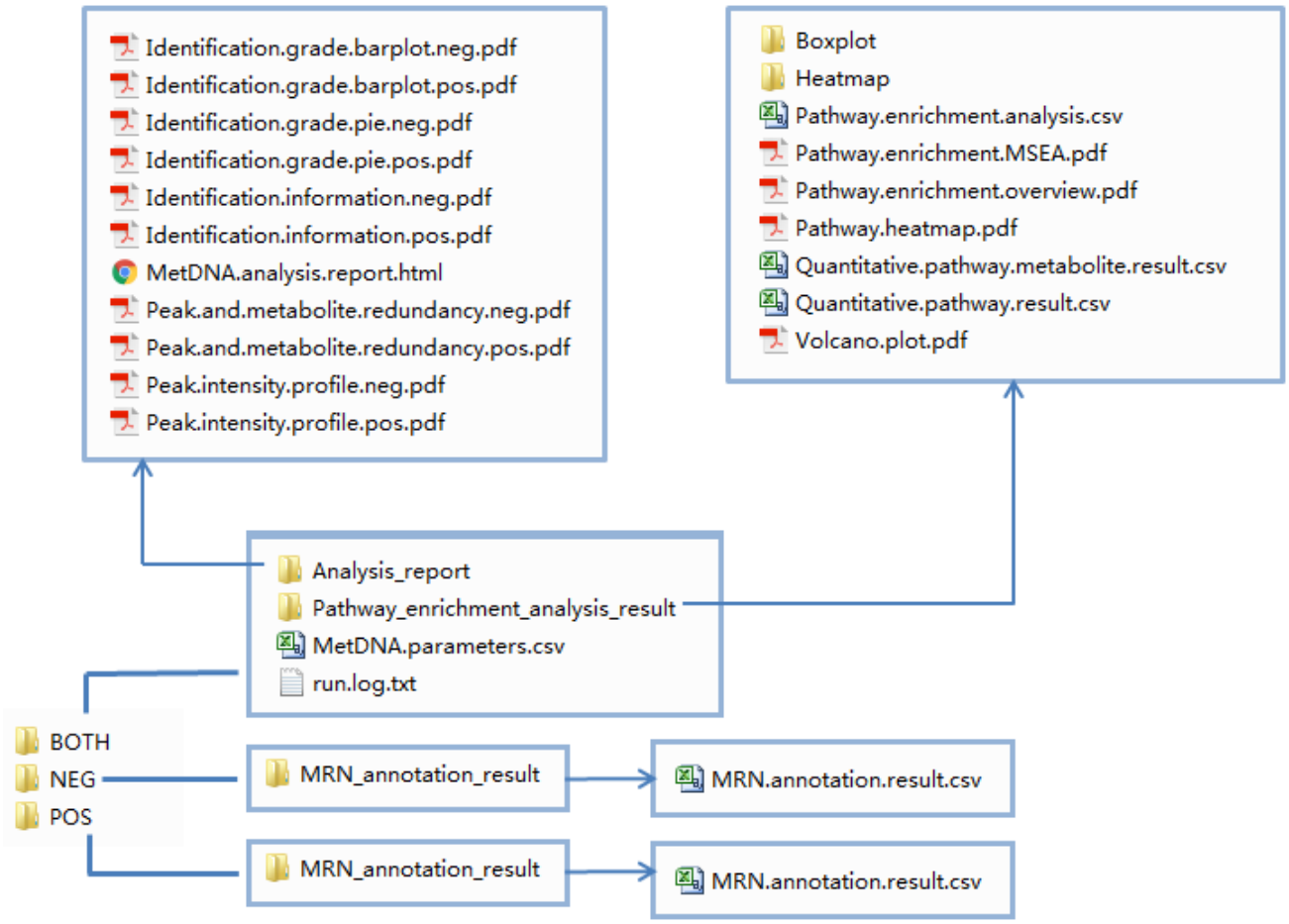
After several hours, users will receive a notification email from metdna@sioc.ac.cn. You can download the result from the “Project” tab.



Please make sure our email address metdna@sioc.ac.cn to be saved in your safe email list.

4. Interpretation of Analysis results from MetDNA

The analysis result contains three folders, named as "BOTH" , "POS" , and "NEG" .



4.1 Information for "POS\MRN_annotation_result\MRN.annotation.result.csv"

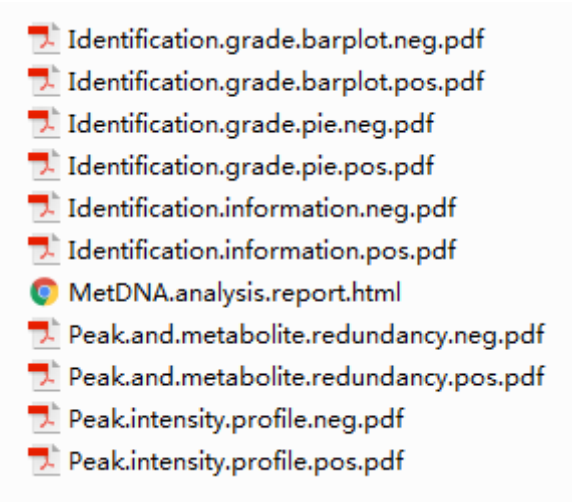
This is a csv table that provides the identification results for peaks in positive mode.

4.2 Information for "NEG\MRN_annotation_result\MRN.annotation.result.csv"

This is a csv table that provides the identification results for peaks in negative mode.

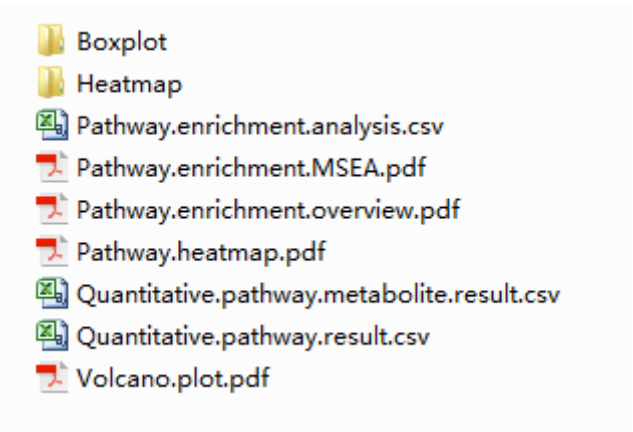
4.3 Information for "BOTH\Analysis_report" folder This is a folder

which contains a summary report of MetDNA (in HTML format) and the related figures in the report (in pdf format).



4.4 Information for "BOTH\Pathway_enrichment_analysis_result"

folder This is a folder which contains the pathway enrichment analysis results.



(1) Boxplot folder: Boxplots for the quantitative analysis of each pathway.
(2) Heatmap folder: Heatmaps for metabolites in each pathway.
(3) Pathway.enrichment.analysis: Detailed information of pathway enrichment analysis result.
(4) Pathway.enrichment.MSEA: A figure to show the pathway enrichment result
(5) pathway.enrichment.overview: A figure to show the pathway enrichment result
(6) Pathway.heatmap: Heatmap to demonstrate the quantitative expressions of pathways
(7) Quantitative.pathway.metabolite.result: A table that contains the quantitative information for significantly changed metabolites; values are pareto scaled.
(8) Quantitative.pathway.result: A table that contains the quantitative information of pathways in each sample.
(9) Volcano.plot: A volcano plot that describes significantly changed metabolites.

4.5 Information for "BOTH\folderMetDNA.parameters.csv"

A table with the parameters used in MetDNA software.

4.6 Information for "BOTH\run.log.txt"

The running log of MetDNA Webserver.

Questions: metdna AT sioc.ac.cn