The tutorial of data processing with MZmine for KGMN

Zhiwei Zhou, Ph.D., 2022.06.07

Introduction

MZmine is a common and flexible software for mass spectrometry data analysis. Users can set a sequence of steps to perform peak picking and alignments. Here, we present a simple tutorial to process an untargeted metabolomics data set, and demonstrate how to export required files for KGMN (MetDNA2) annotation.

There are many useful resources to demonstrate how to use MZmine to process LC-MS/MS data:

- The official document: http://mzmine.github.io/documentation.html
- The FBMN document (Major procedures are similar): https://ccms-ucsd.github.io/GNPSDocumentation/featurebasedmolecularnetworking-with-mzmine2/

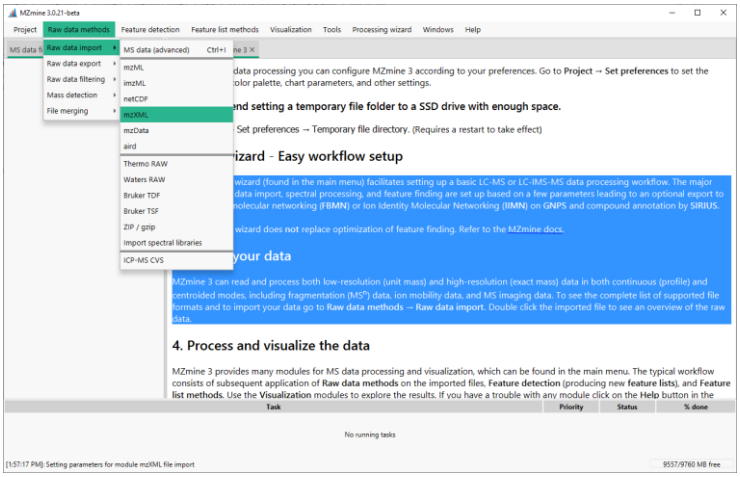
Important note:
The parameter of data processing should be adjusted according to the instrument types, parameter settings and experimental designs. This tutorial only demonstrates how to generate required files for KGMN.

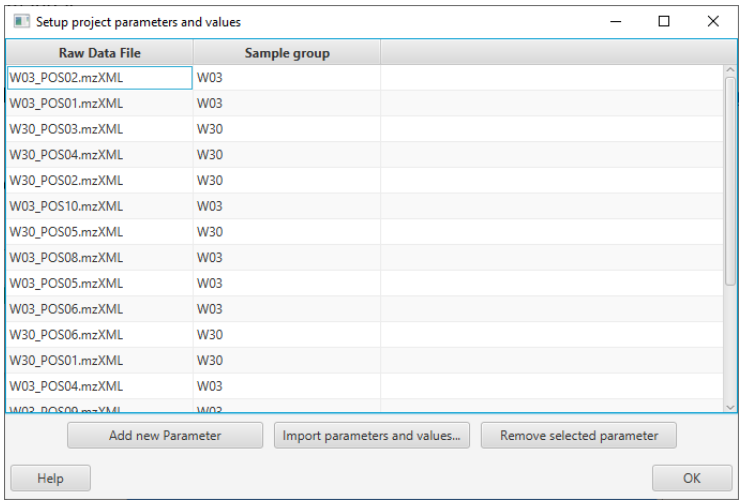
1. Installation

- Download MZmine software (MZmine v3.0.21 used here) at https://github.com/mzmine/mzmine3
- Demo data set: Fruit Fly data sets (mzXML files)

2. Step-by-step data processing with MZmine

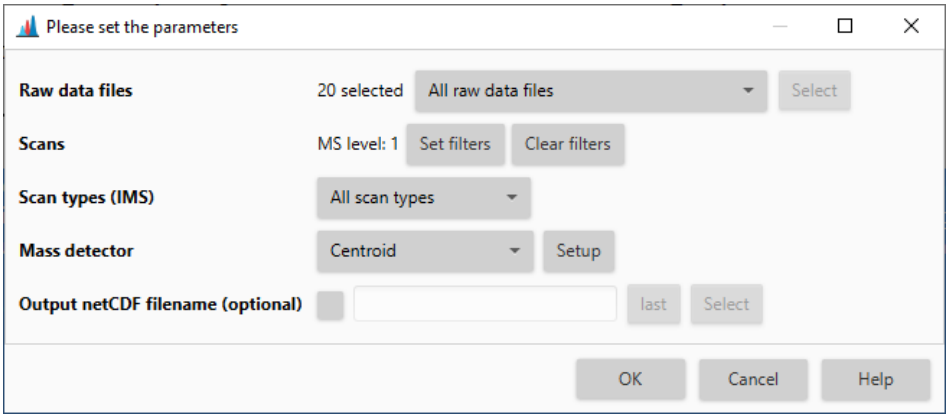
1) **Raw data import and modify sample parameters**:
   - Raw data methods → Raw data import → mzXML
   - Project → Set sample parameters



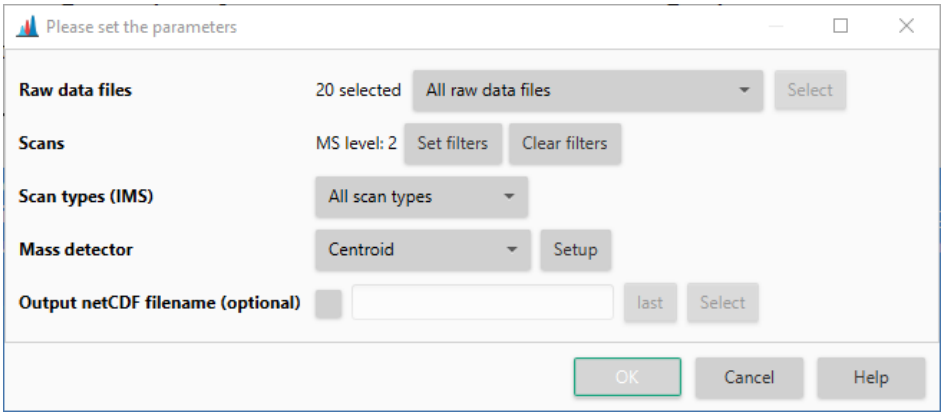


2) Mass detection.

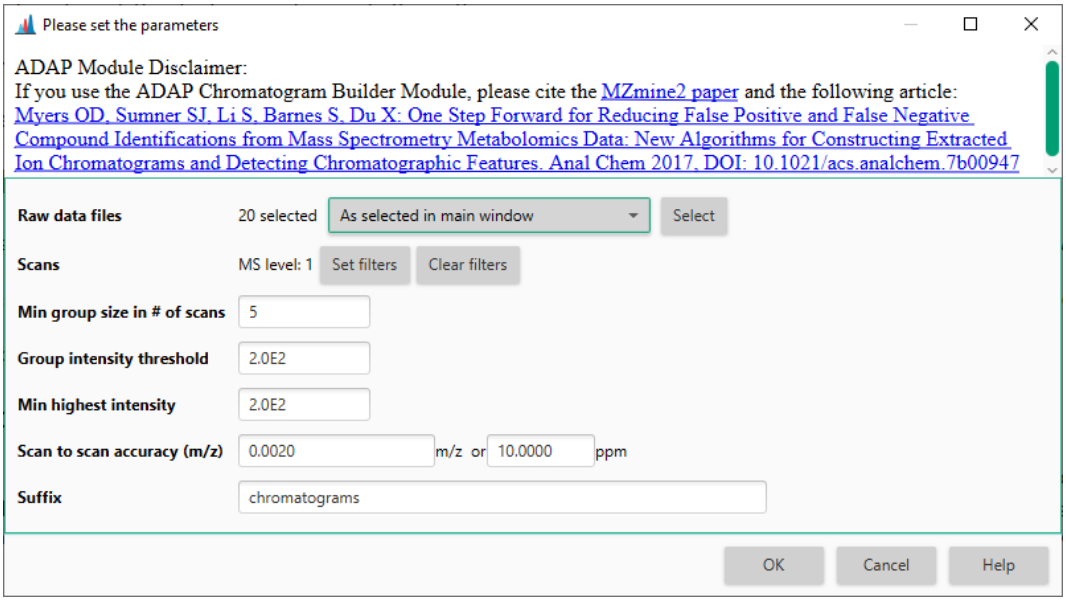
- MS1 mass detection: Raw data methods → Mass detection (Note: “Scan” parameter uses “MS level: 1”)



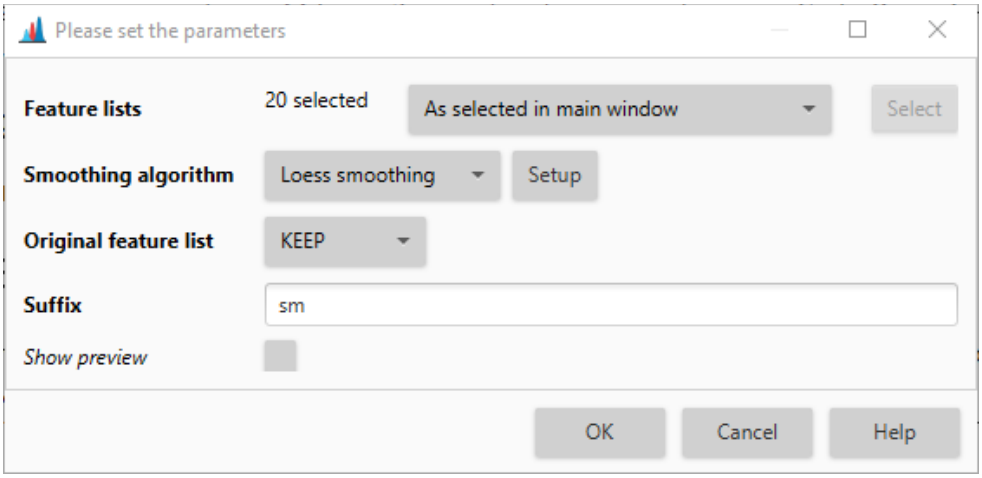
- MS2 mass detection: Raw data methods → Mass detection (Note: “Scan” parameter uses “MS level: 2”)



3) **Chromatogram detection**: Feature detection → LC-MS → ADAP chromatogram builder



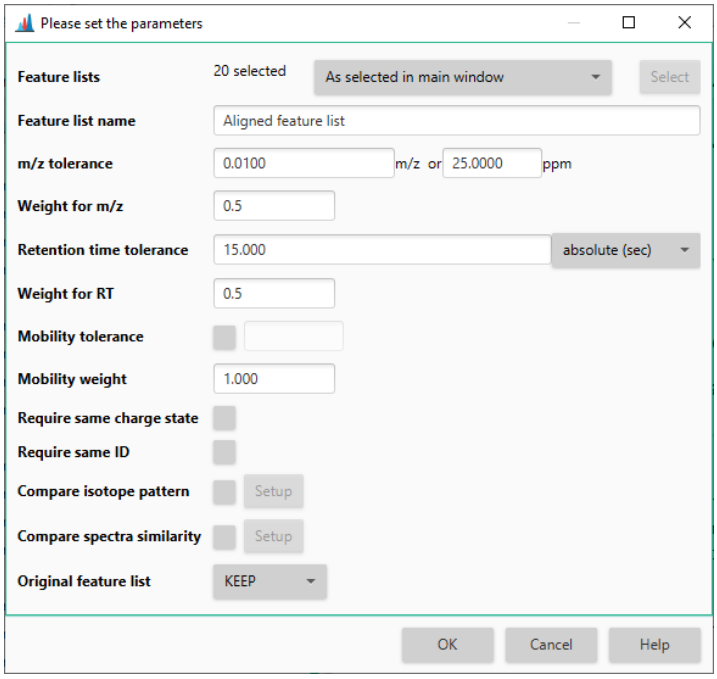
4) **Smoothing:** Feature detection → Smoothing



5) **Chromatogram deconvolution (Optional).** The step of chromatogram deconvolution would further improve the data quality, but it may take a lot of time if you run a large dataset.

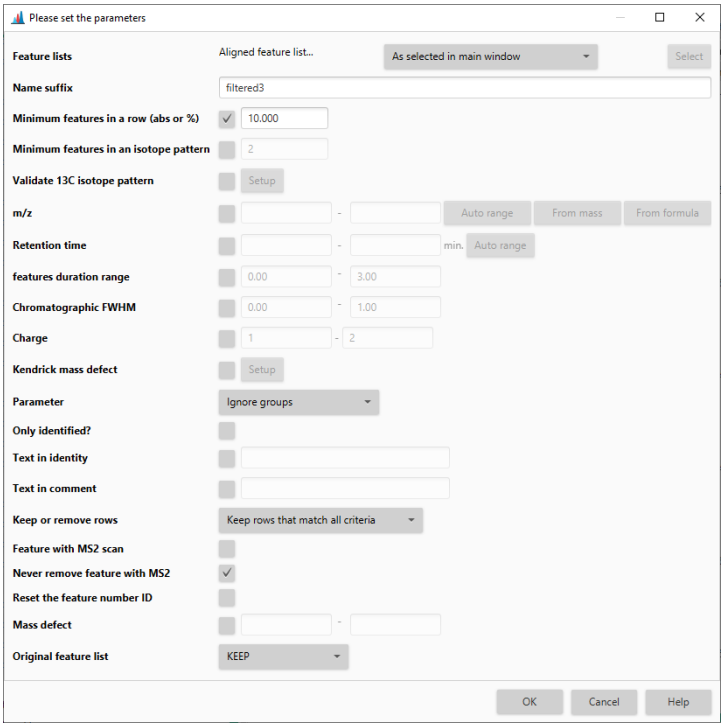
6) **Feature alignment and filtering**.

Feature alignment: Feature list methods → Alignment → Join Aligner

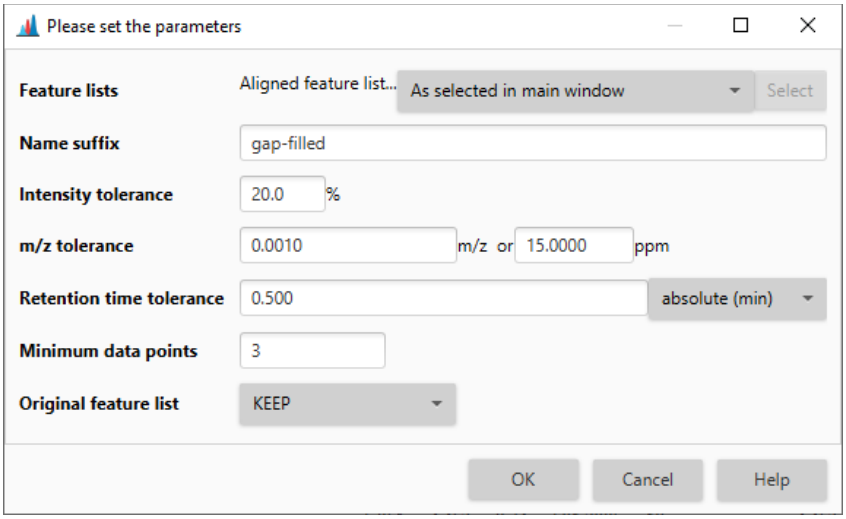


Filtering: keep features which appears more than half samples

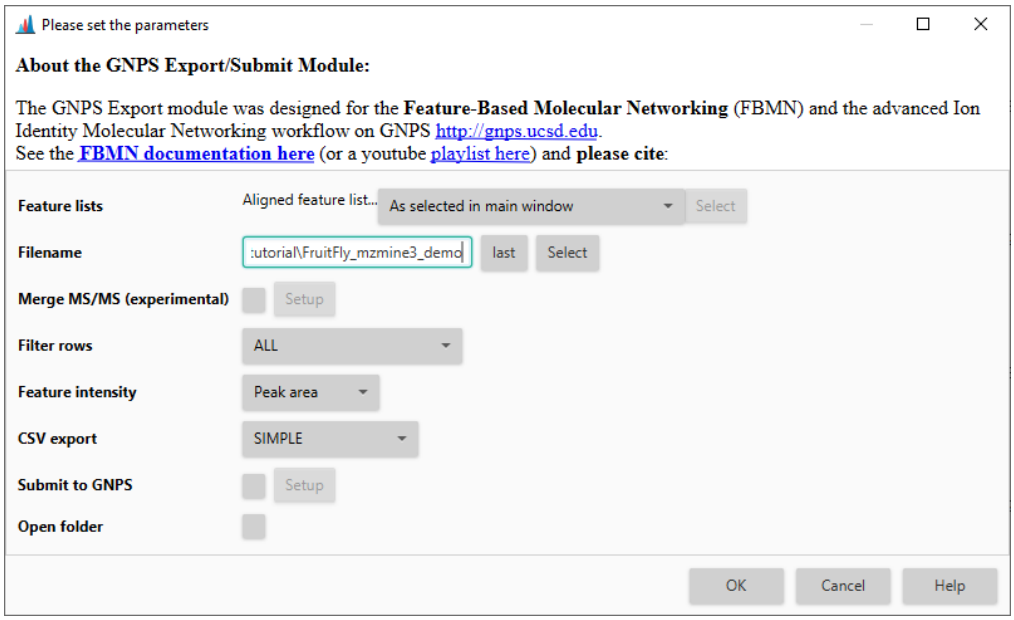
Feature list methods → Feature list filtering → Feature list rows filter



7) **Gap filling (Optional)**: select aligned feature table, then, Feature detection → Gap filling

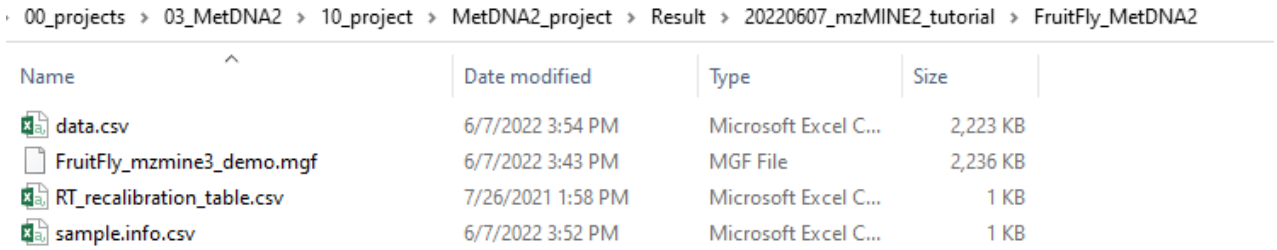


8) Export for KGMN: Feature list methods → Export feature list → GNPS – feature based molecular networking



3. Modification results for KGMN

The KGMN requires 4 type files, peak table with quantification information, MS2 files, sample information table and RT recalibration files.



- **MS1 peak table (Required)**: modify the “quant_table.csv” as below:
  - 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.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	name	mz	rt	W30_POSI	W03_POSI	W30_POSI	W30_POSI	W30_POSI	W30_POSI	W03_POSI	W03_POSI	W03_POSI	W03_POSI	W03_POSI	W30_POSI	W03_POSI	W03_POSI	W03_POSI	W03_POSI	W30_POSI	W30_POSI	W30_POS:	W30_POS09	
2	3	99.05329	15.06829	0	0	13.37165	3.4827	61.15337	25.18087	2.082205	0	22.51981	5641.673	0	9.907391	6.326766	58.63746	0	1.93545	10.55717	2.326575	0	35.5638	
3	11	74.05932	6.21535	114067.2	177.6317	42.42194	433.6523	557.1826	441.8075	205.4193	172.0121	89488.88	113002.7	60886.74	312.3195	48.28134	69609.59	214.4811	74.15388	102073.1	332.9543	137.5661	519.717	
4	15	217.1065	5.635	11525.47	11085.62	241887	254267.6	5694.045	11570.3	249290.2	6482.177	129520.4	257135.5	178522.6	98521.64	5883.51	262591.8	216987.2	207123.6	4943.507	11462.95	11864.34	11799.65	
5	32	80.04776	15.71508	0	0	237.7264	0	168.3993	50.58738	8.555695	5.957001	23709.6	81.70317	7.912578	135.1747	0	92.28883	0	0	0	6.8943	0	27.03539	
6	34	173.0815	16.12185	26.33427	30.48557	537.202	224.7719	1844.353	2199.647	20423.07	30.23502	11840.35	46756.68	60.52673	397.3119	26728.29	903.413	237.9912	31.2923	56.17101	30.70112	26683.65	31593.58	
7	52	100.0754	17.7394	39596.27	38149.55	39998.47	32879.2	381103.2	41383.03	35835.92	34549.54	455936.8	41527.74	40050.68	41510.04	422441.4	40139.49	37527.46	39847.24	385518.9	39753.71	395521.2	398192.3	
8	54	341.0917	20.42635	760.9736	219.5352	149.8669	246.4146	268.3437	11646.33	282.5142	55.30628	411.5482	600.6457	386.9792	349.6789	142.7024	240.2545	477.2896	50.55949	1035.554	189.7244	21266.03	223.0102	
9	55	107.9658	18.17675	1435.444	6.01205	684.5219	833.2325	650.4113	2.154561	742.0686	806.0158	737.8674	939.2011	9.703101	2.680519	929.8411	3.925762	6.498312	681.3801	853.4666	798.7314	5.434917	621.4933	
10	58	217.1067	5.635	15189.23	11085.62	903.4377	13445.07	512.3189	10885.97	3450.75	6482.177	644.0915	894.7816	14828.22	9700.693	5883.51	15174.72	7356.968	6518.554	241113.6	15153.18	15447.69	12171.27	
11	63	265.0016	23.7924	25.32977	218.0977	8342.62	7226.494	151.6802	2426.544	206.601	227.0331	8441.877	43.29933	5954.552	66.29271	50.49244	137.7372	594.1295	6962.066	7664.627	5106.081	6886.647	3086.107	
12	66	587.5415	34.6174	12.65608	5.959606	8.386	17.90799	9.254589	761.1237	6.946327	7.599711	2.155242	4.887832	16.86409	11.79049	6.987074	11.55228	2.118949	2.4715	5.311835	13.6979	25.04416	8.9911	
13	67	185.1148	26.6127	35.28171	215.5527	19.45418	70.10358	8.249992	26.65566	498.8666	304.2642	9.381346	8.174007	1201.507	25.83247	448.6958	25.81368	4902.808	228.0613	156.5823	1019.554	161.0528	27.95929	
14	69	310.9231	18.25985	1289.603	13.13001	1065.56	1010.041	26.15837	829.5033	865.7961	864.2864	10.60216	21.16591	15.70154	866.3833	1095.821	873.9083	1034.832	12.4599	884.6581	11.70907	934.3574	14.61103	
15	70	213.9909	22.6137	432.796	13.42709	458.7052	498.903	1.957467	6.908202	4.130376	461.9909	515.2225	5.446692	398.5192	3.077859	260.1602	2.436	565.2922	260.2651	546.2947	6.616468	384.6588	2.341141	
16	71	94.06426	24.5783	138.4605	727.9568	52.49978	107.1562	585.5607	11625.36	836.9643	88.79984	469.1142	8.554001	1060.196	131.1809	28.11216	349.8337	260.4774	158.7603	174.7507	17979.97	116.7408	18.09822	
17	99	294.9497	24.07715	1321.485	5868.471	70.3121	4364.393	4248.236	4137.203	104.0032	269.4844	5162.326	5428.778	3719.217	4218.44	65.73547	4162.974	4827.028	133.4867	4878.464	2845.183	4598.428	3846.573	
18	100	259.1896	47.3043	1159.229	285.8931	251.3283	228.3592	23.71601	49.40573	174.3479	289.7008	6.242882	51.27469	143.6167	100.7827	368.4096	31.15177	543.9031	445.3673	354.4225	158.7828	142.602	99.2502	
19	110	272.9678	27.37005	8445.578	7410.437	6717.692	5891.667	90.91557	83.33117	279.1852	5064.562	550.1971	530.0817	5659.794	206.6205	544.5256	5226.477	463.152	5892.532	5992.234	505.5784	5828.353	27.58237	
20	113	187.0636	48.77521	6.458491	4.123148	21.25835	4.89775	10.62001	15.55347	14.56158	12.89067	543.1104	15.87996	23.23207	7.442964	18.23014	30.71534	12.87358	11.48541	13.44909	3.714267	24.62721	14.79288	
21	115	122.0958	34.76745	793.205	17060.68	30.53019	103.1928	23.88015	42.61032	532.466	53.57394	39.91462	10.26105	142.9137	68.17567	177.4798	35.36842	1403.184	77.87649	93.50118	920.8564	144.9066	20.1001	
22	119	241.0013	29.6832	7103.916	5926.057	5199.213	4254.359	3822.407	1783.745	205.5797	3403.169	2366.516	2606.241	265.8381	2384.387	549.2857	8.169373	5358.199	1177.175	2128.026	4372.996	298.4445	386.8256	

- **MS2 data files (Required)**: directly upload MGF file

- **Sample information table (Required):** similar with sample parameter. This is a CSV file with two columns, “sample.name” and “group”. **Note:** The "sample.name" column in sample information file must be the **EXACTLY same** as the sample names in the MS1 peak table.

	A	B
1	sample.name	group
2	W30_POS01	W30
3	W03_POS04	W03
4	W30_POS03	W30
5	W30_POS04	W30
6	W30_POS02	W30
7	W30_POS06	W30
8	W03_POS01	W03
9	W03_POS02	W03
10	W03_POS06	W03
11	W03_POS03	W03
12	W03_POS05	W03
13	W30_POS05	W30
14	W03_POS09	W03
15	W03_POS08	W03
16	W03_POS07	W03
17	W03_POS10	W03
18	W30_POS08	W30
19	W30_POS07	W30
20	W30_POS10	W30
21	W30_POS09	W30

- **RT recalibration table (Optional).** Please see instruction of MetDNA2 website (http://metdna.zhulab.cn/metdna/help).