



MetDNA2 帮助文档

V1.0.00

周智伟 <zhouzw@sioc.ac.cn>

罗名都 <luomd@sioc.ac.cn>

朱正江 <jiangzhu@sioc.ac.cn>

February 17, 2022

Laboratory for Mass Spectrometry and Metabolomics

www.zhulab.cn

Interdisciplinary Research Center on Biology and Chemistry (IRCBC)

Shanghai Institute of Organic Chemistry

Chinese Academy of Sciences, Shanghai, China

目录

1. 新闻 & 更新 & Bug 报告	3
2. 测试数据.....	4
2.1 原始数据文件.....	4
2.2 转换后的数据文件.....	4
2.3 ZhuLab RT 系统	4
3. 数据准备.....	6
3.1 XCMS	7
3.2 MS-DIAL.....	13
3.3 数据准备的重要说明.....	19
4. MetDNA2 网站使用	20
4.1 网站注册.....	20
4.2 数据格式检查.....	20
4.3 数据上传.....	21
4.4 数据处理参数设置.....	22
4.5 提交数据.....	24
4.6 结果下载.....	25
5. MetDNA2 结果解读	26
5.1 POS\02_result_MRN_annotation\MRN.annotation.result.csv	26
5.2 文件夹 POS\00_annotation_table (仅支持 MetDNA2).....	26
5.3 文件夹 BOTH\04_biology_intepretation	27
5.4 文件夹 BOTH\05_analysis_report "	28

1. 新闻 & 更新 & Bug 报告

重要说明和新闻

- MetDNA2 于 2021 年 7 月 19 日上线。
- 由于计算资源的限制，并发处理项目限制为 5 个。
- 延伸步数限制为 0，直到 MetDNA2 纸出来。
- 如果您想重现 MetDNA1 版本的结果，请在 MetDNA1 版本参数中选择“version1”。
- MetDNA2 的演示数据文件：XCMS、MS-DIAL。
- 此处提供完整的更新历史记录。

讨论和错误报告

如果您有任何问题或错误报告，请与我们联系并提供以下材料。我们将在每周五下午 1:00 - 下午 3:00（北京时间）回答您的问题。

- 代表上传的数据。
- 项目的日志文件。
- 和/或幻灯片来解释您的问题。
- 我们始终欢迎通过 google 群组进行任何有关 MetDNA 的讨论和错误报告：MetDNA forum。
- 对于中国用户，请加入我们的 QQ 群进行任何讨论和错误报告：786156544。

2. 测试数据

2.1 原始数据文件

我们提供了一个演示数据集供用户学习如何使用 MetDNA2，该数据集有两组来自衰老苍蝇的头部组织（3 周 vs. 20 周，每组 $n = 10$ ）。样品制备和数据采集的详细信息可以在我们发表的文章中¹找到。数据采集自 Sciex TripleTOF 6600 中进行。

原始数据可从 **MetaboLight** 访问：[MTBLS612](#)、[MTBLS615](#)

2.2 转换后的数据文件

MetDNA2 需要一些必要的原始数据处理来获取输入文件。

通常，MetDNA 需要导入以下文件进行代谢物鉴定，包括：

- (1) MS1 峰表（.csv 格式，必需），
- (2) MS2 数据文件（.mgf 或 .msp 格式，需要），
- (3) 样品信息表（.csv 格式，必填），
- (4) RT 重新校准表（.csv 格式，可选）。如果您想遵循我们发布的 LC 方法并重新校准 RT 库。ZhuLab RT 系统的详细信息在第 2.3 章中列出。

2.3 ZhuLab RT 系统

如果您想使用 ZhuLab RT 库，请遵循完全相同的 LC 条件和梯度。

2.3.1 23min Amide RT 系统（Zhu Lab）

- 色谱柱：Waters UPLC BEH Amide 色谱柱 ($1.7\ \mu\text{m} \times 100\ \text{mm}$)
- A 相：100% H_2O + 25 mM $\text{CH}_3\text{COONH}_4$ + 25 mM NH_4OH
- B 相：100% 乙腈

23min gradient table (Amide)

Time (min)	Phase B (%)	Flow rate (mL/min)
0	95	0.3
1	95	0.3
14	65	0.3
16	40	0.3
18	40	0.3
18.1	95	0.3
23	95	0.3

2.3.2 12min Amide RT 系统 (Zhu Lab)

- 色谱柱: Waters UPLC BEH Amide 色谱柱 (1.7 μm \times 100 mm)
- A 相: 100% H_2O + 25 mM $\text{CH}_3\text{COONH}_4$ + 25 mM NH_4OH
- B 相: 100% 乙腈

12min gradient table (Amide)

Time (min)	Phase B (%)	Flow rate (mL/min)
0	95	0.5
0.5	95	0.5
7	65	0.5
8	40	0.5
9	40	0.5
9.1	95	0.5
12	95	0.5

我们推荐使用 XCMS、MZmine 和 MSDIAL 进行数据转换。 请按照以下说明进行数据转换（第 3 部分）。 从这些软件工具转换的演示数据集数据可以在以下链接中下载:

- [XCMS](#)
- [MSDIAL](#)

3. 数据准备

MetDNA2 输入数据如下：

- **MS1 表格：**这是一个代谢峰列表，带有注释的 m/z、保留时间 (RT) 和峰丰度。MS1 表格是使用 XCMS、MZmine 和 MS-DIAL 等常用峰采集软件从原始 MS 文件生成的。
- **MS2 文件：**MetDNA2 支持 .mgf 和 .msp 格式的 MS2 数据文件。
- **样本信息文件：**旨在描述样本组信息 (.csv 格式)。第一列命名为“sample.name”，第二列命名为“group”。
- **RT 校正表格。**如果您的 LC 方法遵循我们公布的方法，MetDNA2 支持用于代谢物鉴定的实验性 RT 库 (amide12min_zhulab 和 amide23min_zhulab)。如果您在数据采集批次中运行标准混合物 (RTQC)，则可以实现基于 LOESS 的 RT 重新校准。RTQC 的组成如下：

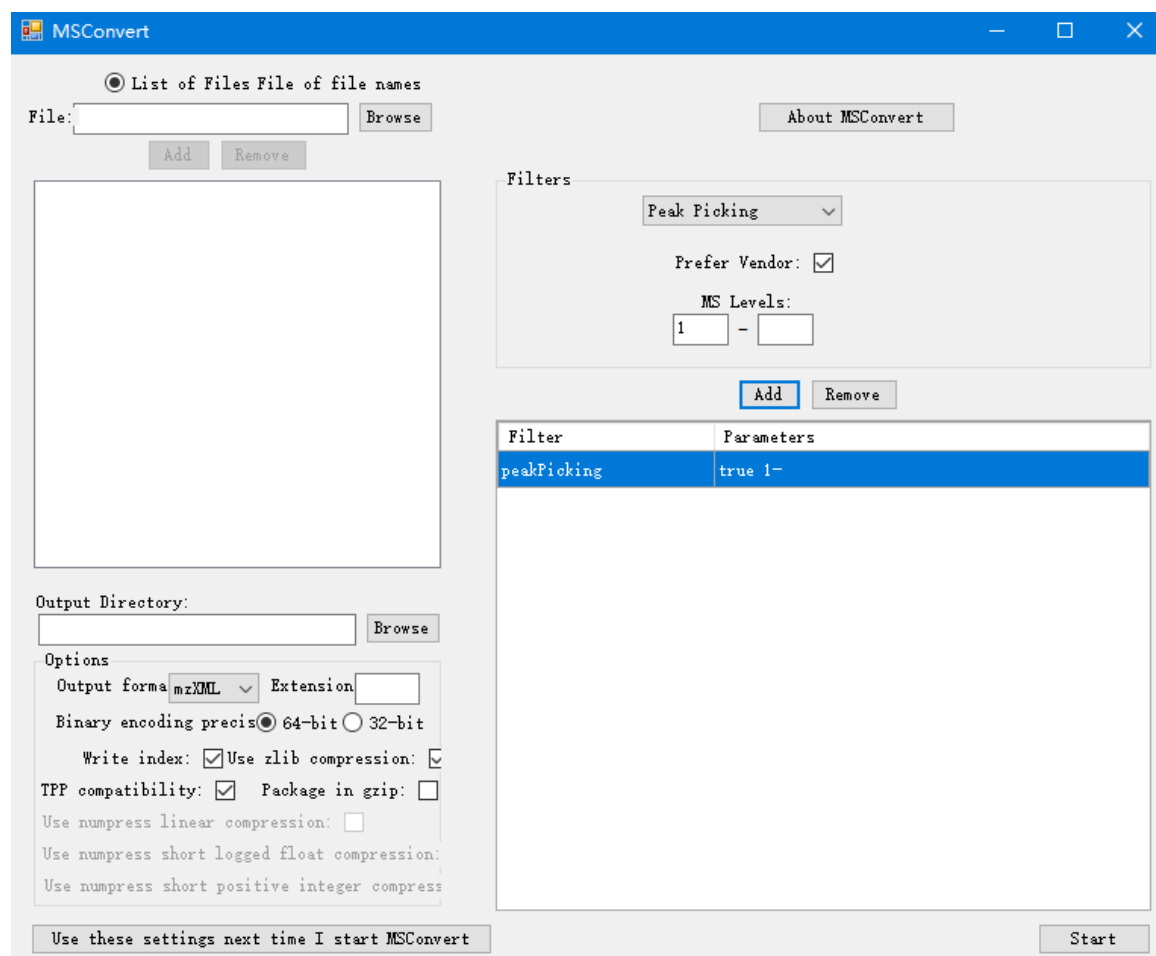
compound.name	id.zhulab	id.pubchem	Concentration
3-Methoxy-4-Hydroxyphenylglycol Sulfate	L0228	3035420	50 ppb
1-Methylxanthine	L0216	80220	250 ppb
N-Acetyl-L-phenylalanine	L0086	74839	100 ppb
5-Hydroxyhexanoic acid	S0266	170748	1 ppm
5-Hydroxymethylcytidine	L0264	14367004	10 ppm
Sebacic acid	S0193	5192	500 ppb
Suberic acid	S0214	10457	1 ppm
N-Acetyl-L-aspartic acid	S0246	65065	100 ppb
N-carbamoyl-L-aspartate	L0084	93072	1 ppm
UDP-D-glucuronate	L0081	17473	10 ppm
4-Pyridoxic acid	L0127	6723	20 ppb
3'-O-methyladenosine	L0251	82530	20 ppb
Creatinine	S0042	588	20 ppb
Nicotinate	S0008	938	100 ppb
Guanosine	S0071	135398635	1 ppm
N-Acetylcadaverine	S0249	189087	40 ppb
Beta-Alanine	L0320	239	1 ppm
dAMP	L0140	12599	500 ppb
gamma-L-Glutamyl-L-glutamic acid	L0352	92865	10 ppm
N6,N6,N6-Trimethyl-L-lysine	L0190	440120	10 ppm

我们以下提供逐步说明，分别使用 XCMS、MSDIAL 和 MZmine2 来演示如何从原始数据中转换生成 MetDNA2 需要的输入数据。

3.1 XCMS

3.1.1 MS1 数据转化

- 使用 ProteoWizard（版本 3.0.6150）将原始 MS 数据文件（例如，Sciex .wiff 文件）转换为 mzXML 格式。请按照以下转换设置：



3.1.2 XCMS 代谢峰检测

- 在每个极性中，所有 10 个 MS 文件（mzXML 格式）均使用 R 包“xcms”（版本 1.46.0）进行峰值检测和对齐处理。mzXML 文件根据它们的组被放置在两个文件夹中，分别命名为：“W03”、“W30”。

使用如下所示的代码处理数据集（该软件需要一些简单的 R 背景）：

```
##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), mzdifff = 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)
```

结果,生成了一个名为“Peak-table”的 MS1 代谢峰表格。对于小鼠肝脏演示数据,

在正模式和负模式下分别检测到 18320 个峰和 14965 个峰。从 XCMS 对这个表格进行简单修改。

MetDNA 需要特定格式的 MS1 表格 (.csv 格式)：

- 第一列是峰名称 (“名称”)；
- 第二列是质荷比 (“mz”)；
- 第三列是保留时间 (“rt”)；
- 保留时间的单位必须是秒 (不是分钟)；
- 其他列是每个样品中 MS1 峰的峰丰度。

重要提示：前三列的顺序和名称必须是 “name” 、 “mz” 和 “rt” 。使用 XCMS 生成的“Peak-table”，可以进行如下修改：

- (1) 前 12 列，保留名为“name”、“mzmed”、“rtmed”的列，删除其他列；
- (2) 将前三列重命名为“name”、“mz”和“rt”；
- (3) 根据极性将各自的表格分别命名为“Peak_Table_POS”或“Peak_Table_NEG”。

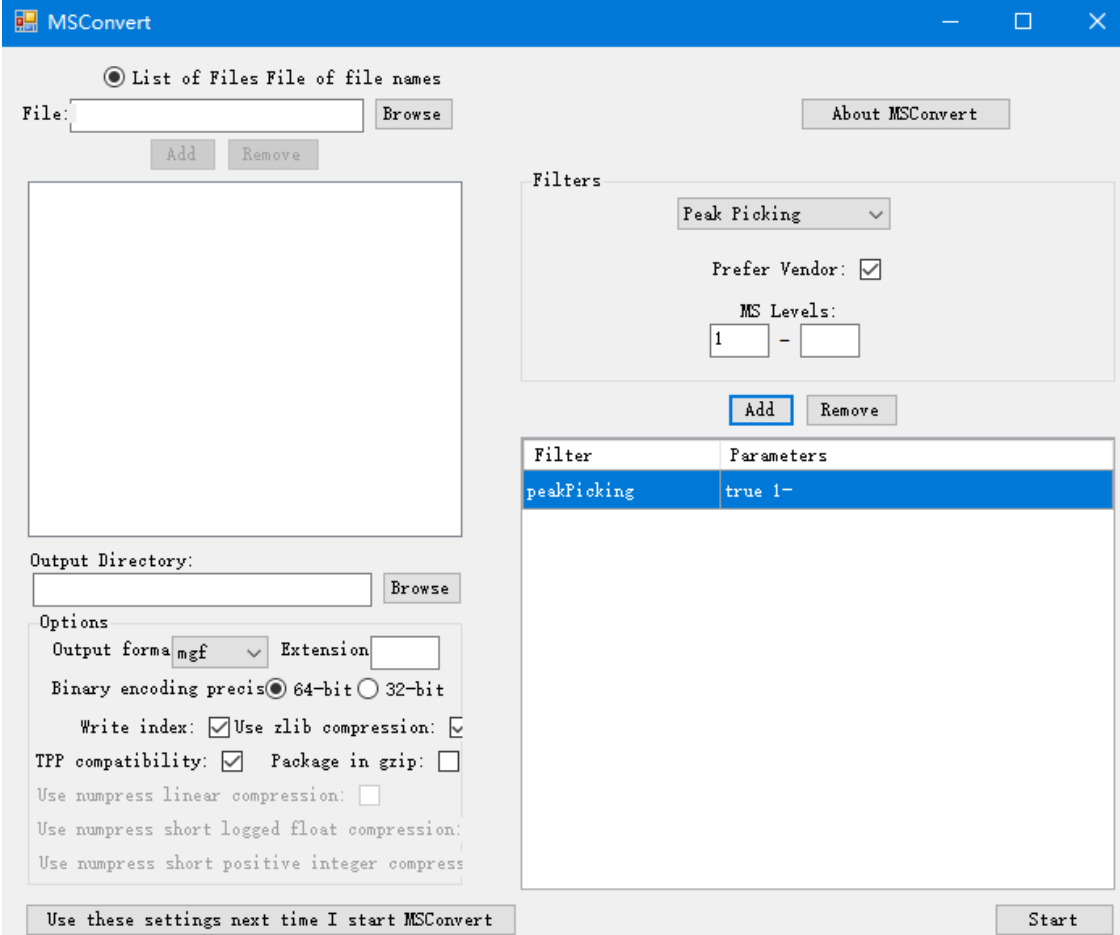
最终生成的 MS1 表格应如下所示：

	A	B	C	D	E	F	G	H	I	J	K
1	name	mz	rt	W03.01	W03.02	W03.03	W03.04	W03.05	W03.06	W03.07	W03.08
2	M58T493	58.06463	493.246	1590.791	2705.634	3228.77	2223.469	1012.648	3384.559	3363.545	627.1033
3	M59T1373	59.04893	1373.275	394.9867	607.5771	428.8762	401.0971	492.7037	323.0467	416.7519	315.884
4	M59T1102	59.04867	1101.605	274.9371	383.0837	286.3867	332.6357	255.644	468.7467	327.2486	356.5667
5	M59T1002	59.04868	1002	245.0608	395.7779	408.084	128.9722	442.4482	449.293	287.664	407.13
6	M59T1226	59.04889	1225.53	446.4847	270.256	325.89	350.0578	420.3343	338.8	389.7733	477.48
7	M59T1307	59.04976	1306.74	325.0667	283.0369	247.69	444.5614	323.708	389.16	413.4524	367.694
8	M59T1328	59.04998	1328.12	332.7333	347.3998	458.035	355.454	349.2557	417.3597	272.49	320.74
9	M59T1048	59.04985	1047.56	333.8955	2608.083	284.148	509.4073	316.872	1141.812	1424.289	427.294
10	M59T1188	59.04931	1187.76	149.056	312.12	453.0383	337.5842	404.8	993.452	435.0348	353.4
11	M59T1159	59.05845	1159.48	253.5252	161.9722	524.15	228.2888	504.9609	752.37	441.0083	689.4723
12	M59T1264	59.04894	1263.57	354.85	368.135	301.01	254.79	250.418	358.6116	292.64	394.0667
13	M59T948	59.04985	947.58	221.7316	407.9952	950.7372	534.9058	448.2161	425.0073	605.6133	442.993
14	M59T1360	59.06056	1360.35	224.7867	344.3867	388.2172	405.6829	720.5885	425.561	436.6938	323.42
15	M59T37	59.06067	36.582	1062.268	973.5093	1089.995	1125.291	1224.801	1383.048	732.256	1032.388
16	M59T1135	59.05934	1134.815	2018.674	2114.484	2214.614	1987.659	5645.7	2148.379	2269.693	2100
17	M59T981	59.05934	980.529	1025.758	2449.929	1208.905	1701.403	1589.612	1590.77	1837.035	1307.526
18	M59T1187	59.06076	1187.12	1196.694	477	1139.392	681	1420.195	1146.34	1136.406	1389.471
19	M60T1238	60.04342	1238.085	2475.192	1870.353	1714.795	1372.124	1061.817	709.324	1608.076	934.2629
20	M60T1189	60.04467	1188.83	1425.861	1540.744	1547.552	749.1	1327.31	773.8971	655.34	1360.985
21	M60T1124	60.04342	1123.73	555.5886	1597.663	683.2571	490.464	826.2377	492.9	583.7914	657.6
22	M60T1005	60.04397	1005.304	642.3967	646.03	667.4627	497.3566	1041.45	401.5436	613.6536	728.288
23	M60T1069	60.04332	1068.54	895.8252	637.181	546.56	921.1173	502.756	742.8586	697.6162	498.134
24	M60T1156	60.04327	1156.145	861.6528	676	1086.4	811.3606	970.4532	1156.436	723.2256	755.8891

3.1.3 MS2 数据

对于使用 XCMS 的用户，请使用 ProteoWizard（版本 3.0.6150）将每个电离极性的原始 MS2 数据文件转换为 .mgf 格式。

请按照以下转换设置：



The screenshot shows the MSConvert application window with the following settings:

- File:** A text box with a "Browse" button next to it.
- Output Directory:** A text box with a "Browse" button next to it.
- Options:**
 - Output format: **mgf** (selected in a dropdown menu)
 - Extension: (empty text box)
 - Binary encoding precision: **64-bit** (selected radio button)
 - Write index: ☒
 - Use zlib compression: ☒
 - TFP compatibility: ☒
 - Package in gzip: ☐
 - Use numpress linear compression: ☐
 - Use numpress short logged float compression: ☐
 - Use numpress short positive integer compression: ☐
- Filters:**
 - Peak Picking: (selected in a dropdown menu)
 - Prefer Vendor: ☒
 - MS Levels: **1** - (empty text box)
 - Buttons: **Add** and **Remove**
- Filter List Table:**

Filter	Parameters
peakPicking	true 1-
- Buttons:** **Start** and **Use these settings next time I start MSConvert**

3.1.4 样本信息

样本信息文件(.csv 格式)旨在描述样本组信息。第一列命名为“sample.name”，第二列命名为“group”。示例信息文件应如下所示：

	A	B
1	sample.name	group
2	W30.02	W30
3	W03.04	W03
4	W03.05	W03
5	W30.06	W30
6	W03.01	W03
7	W30.07	W30
8	W03.02	W03
9	W03.10	W03
10	W30.10	W30
11	W30.05	W30
12	W30.04	W30
13	W03.07	W03
14	W03.09	W03
15	W30.08	W30
16	W30.01	W30
17	W30.09	W30
18	W03.03	W03
19	W03.06	W03
20	W30.03	W30
21	W03.08	W03

注意：样品信息文件中的“sample.name”列必须与 MS1 峰表中的样品名称完全相同。

	A	B	C	D	E	F	G	H	I	J	K
1	name	mz	rt	W03.01	W03.02	W03.03	W03.04	W03.05	W03.06	W03.07	W03.08
2	M58T493	58.06463	493.246	1590.791	2705.634	3228.77	2223.469	1012.648	3384.559	3363.545	627.1033
3	M59T1373	59.04893	1373.275	394.9867	607.5771	428.8762	401.0971	492.7037	323.0467	416.7519	315.884
4	M59T1102	59.04867	1101.605	274.9371	383.0837	286.3867	332.6357	255.644	468.7467	327.2486	356.5667
5	M59T1002	59.04868	1002	245.0608	395.7779	408.084	128.9722	442.4482	449.293	287.664	407.13
6	M59T1226	59.04889	1225.53	446.4847	270.256	325.89	350.0578	420.3343	338.8	389.7733	477.48
7	M59T1307	59.04976	1306.74	325.0667	283.0369	247.69	444.5614	323.708	389.16	413.4524	367.694
8	M59T1328	59.04998	1328.12	332.7333	347.3998	458.035	355.454	349.2557	417.3597	272.49	320.74
9	M59T1048	59.04985	1047.56	333.8955	2608.083	284.148	509.4073	316.872	1141.812	1424.289	427.294
10	M59T1188	59.04931	1187.76	149.056	312.12	453.0383	337.5842	404.8	993.452	435.0348	353.4
11	M59T1159	59.05845	1159.48	253.5252	161.9722	524.15	228.2888	504.9609	752.37	441.0083	689.4723
12	M59T1264	59.04894	1263.57	354.85	368.135	301.01	254.79	250.418	358.6116	292.64	394.0667
13	M59T948	59.04985	947.58	221.7316	407.9952	950.7372	534.9058	448.2161	425.0073	605.6133	442.993
14	M59T1360	59.06056	1360.35	224.7867	344.3867	388.2172	405.6829	720.5885	425.561	436.6938	323.42
15	M59T37	59.06067	36.582	1062.268	973.5093	1089.995	1125.291	1224.801	1383.048	732.256	1032.388
16	M59T1135	59.05934	1134.815	2018.674	2114.484	2214.614	1987.659	5645.7	2148.379	2269.693	2100
17	M59T981	59.05934	980.529	1025.758	2449.929	1208.905	1701.403	1589.612	1590.77	1837.035	1307.526
18	M59T1187	59.06076	1187.12	1196.694	477	1139.392	681	1420.195	1146.34	1136.406	1389.471
19	M60T1238	60.04342	1238.085	2475.192	1870.353	1714.795	1372.124	1061.817	709.324	1608.076	934.2629
20	M60T1189	60.04467	1188.83	1425.861	1540.744	1547.552	749.1	1327.31	773.8971	655.34	1360.985
21	M60T1124	60.04342	1123.73	555.5886	1597.663	683.2571	490.464	826.2377	492.9	583.7914	657.6
22	M60T1005	60.04397	1005.304	642.3967	646.03	667.4627	497.3566	1041.45	401.5436	613.6536	728.288
23	M60T1069	60.04332	1068.54	895.8252	637.181	546.56	921.1173	502.756	742.8586	697.6162	498.134
24	M60T1156	60.04327	1156.145	861.6528	676	1086.4	811.3606	970.4532	1156.436	723.2256	755.8891

3.1.5 RT 校正表格

- 如果您遵循我们的 LC 方法并在数据采集期间运行 RTQC 标准混合物，建议使用 RT 重新校准表重新校准到实验 RT 库。您可以手动输入校准物的保留时间或在参数设置页面上传 RT 重新校准表。
- 如果要上传一个 RT recalibration 表，列名和行的顺序应该和下面的例子完全一样，你应该根据你的实验值修改 calibrant 的 rt（单位是分钟）。

正离子模式下的 RT 校准表格示例：

	A	B	C	D	E	F
1	compound.name	id.zhulab	id.pubchem	ref.mz	rt	polarity
2	4-Pyridoxic acid	L0127	6723	184.0605	1.07	positive
3	3'-O-methyladenosine	L0251	82530	282.1197	3.12	positive
4	Creatinine	S0042	588	114.0662	4.95	positive
5	Nicotinate	S0008	938	124.0393	6.65	positive
6	Guanosine	S0071	135398635	284.099	7.92	positive
7	N-Acetylcadaverine	S0249	189087	145.1336	9.13	positive
8	Beta-Alanine	L0320	239	90.055	11.23	positive
9	dAMP	L0140	12599	332.0755	13.28	positive
10	gamma-L-Glutamyl-L-glutamic acid	L0352	92865	277.1031	14.5	positive
11	N6,N6,N6-Trimethyl-L-lysine	L0190	440120	189.1598	16.38	positive

负离子模式下的 RT 校准表格示例：

	A	B	C	D	E	F
1	compound.name	id.zhulab	id.pubchem	ref.mz	rt	polarity
2	3-Methoxy-4-Hydroxyphenylglycol Sulfate	L0228	3035420	263.0231	0.98	negative
3	1-Methylxanthine	L0216	80220	165.0418	3.17	negative
4	N-Acetyl-L-phenylalanine	L0086	74839	206.0822	5.22	negative
5	5-Hydroxyhexanoic acid	S0266	170748	131.0713	7.03	negative
6	5-Hydroxymethylcytidine	L0264	14367004	272.0889	8.49	negative
7	Sebacic acid	S0193	5192	201.1132	10.03	negative
8	Suberic acid	S0214	10457	173.0819	11.08	negative
9	N-Acetyl-L-aspartic acid	S0246	65065	174.0408	12.28	negative
10	N-carbamoyl-L-aspartate	L0084	93072	175.036	13.29	negative
11	UDP-D-glucuronate	L0081	17473	579.027	15.34	negative

你可以从[此处](#)下载测试的 RT 校正表格。

3.2 MS-DIAL

3.2.1 MS1 数据转化

- MS1 数据转换：使用 Analysis Base File Converter（版本 1.1.0.0）将原始 MS 数据文件转换为 abf 格式
- 使用 MS-DIAL 处理数据：在 MS-DIAL 中生成一个新项目（版本 4.60）
- 在每个极性下，根据实验设计选择合适的参数。确保以.msp 格式导出“**Raw data matrix (Area)**”和“**Representative spectra**”。有关详细信息，请参考“[MS2 数据](#)”。

以我们的测试数据（正离子模式）为例，我们使用的参数如下（具体参数可以[点此处](#)下载）：

```
MS-DIAL ver. 4.60

#Project
MS1 Data type    Profile
MS2 Data type    Profile
Ion mode         Positive
Target          Metabonomics
Mode            ddMSMS

#Data collection parameters
Retention time begin    0
Retention time end      23
Mass range begin        60
Mass range end          1200
MS2 mass range begin    25
MS2 mass range end      1200
```

Alignment result export

Directory: Browse

Export option

File:

☒ Raw data matrix (Height)
 ☐ Peak ID matrix

☐ Normalized data matrix
 ☐ Retention time matrix

☒ Raw data matrix (Area)
 ☐ m/z matrix

☐ *Export as mztab-M
 ☐ MS/MS included matrix

☐ GNPS export

☐ S/N matrix export

☒ Representative spectra

☒ Parameter

Filtered by blank peaks (must be checked in alignment parameter setting)

☐ Filtering by the ion abundances of blank samples

Missing value option

☐ Replace zero values with 1/10 of minimum peak height over all samples

Isotope labeled tracking option Target file

☐ Filtering by the result of isotope labeled tracking

Export format:

Export Cancel

修改 MS1 表格至 MetDNA2 需要特定格式（.csv 格式）。

- 第一列是峰名称（“name”）。
- 第二列是质荷比（“mz”）。
- 第三列是保留时间（“rt”）。
- 保留时间的单位必须是秒（不是分钟）。
- 其他列是每个样品中 MS1 峰的峰丰度。

重要提示：前三列的顺序和名称必须是“name”、“mz”和“rt”。使用 MS-DIAL 生成的“Raw data matrix (Area)”文件，可以修改如下：

- (1) 用 EXCEL 打开“Area.txt”文件，删除前 4 行，只保留名为“Alignment ID”、“Average Rt (min)”和“Average Mz”的列以及每个样本的列。
- (2) 交换第二列和第三列的顺序。 将保留时间的单位更改为秒。
- (3) 将前三列重命名为“name”、“mz”和“rt”。
- (4) 根据极性将峰表格命名为“Peak_Table_POS”或“Peak_Table_NEG”。

最终生成的 MS1 表应如下所示：

	A	B	C	D	E	F	G	H	I	J	K
1	name	mz	rt	W03.01	W03.02	W03.03	W03.04	W03.05	W03.06	W03.07	W03.08
2	0	60.04458	65.46	72102	59455	61155	75126	70122	69224	71378	69357
3	1	60.07921	487.92	124220	117089	134748	117222	116787	129138	117582	124574
4	2	62.05778	586.44	10943	9937	12202	10017	6119	11312	7882	8187
5	3	62.06028	728.16	18129	25006	24462	24663	31669	25478	23404	27956
6	4	62.98056	641.94	14753	16758	14582	18096	18003	15747	16758	20014
7	5	67.04201	695.16	16481	16328	16694	19801	16327	19707	17584	19727
8	6	68.04762	564	141738	148617	193914	154214	150546	193661	174458	117452
9	7	68.0479	694.86	158573	171982	165931	185563	169561	187873	185445	191220
10	8	68.04813	684.78	70527	62101	75703	78202	59119	69889	70523	63844
11	9	68.04961	481.26	23357	28647	25536	23746	38907	36582	41196	34186
12	10	69.03178	684.72	181394	162288	187964	198937	156747	184717	175418	177185
13	11	69.0449	109.38	201944	84881	123522	29659	4210	135264	233998	148146
14	12	69.04494	83.28	120898	53351	73423	49196	19089	64294	180890	73086
15	13	69.85059	564	27686	30734	40419	31656	31407	39225	35460	22336
16	14	70.06375	77.76	11618	12362	13346	11896	10166	9137	10861	11055
17	15	70.06413	564	4897931	5194146	6821479	5449870	5347568	6640286	6071122	4011793
18	16	70.065	961.38	350034	226732	364269	340037	304235	327483	519223	210826
19	17	70.06603	536.82	646621	683571	672213	365553	331017	721630	807467	778492
20	18	70.06616	581.34	1491928	1583940	1560250	1540114	1432975	1620120	1586224	1455605
21	19	70.06622	503.76	2751747	1872075	2844237	909384	2489225	1493517	3426262	1222921
22	20	70.06625	481.44	891878	1195393	983816	930249	1020217	1244685	1536690	1429955
23	21	71.95018	480.72	30818	18858	23915	38173	42467	25006	30958	57380

3.2.2 MS2 数据

- 对于使用 MS-DIAL 的用户，请以.msp 格式导出“Representative spectra”。请按照以下说明进行操作：

Alignment result export

Directory:

Export option

File:

<input checked="" type="checkbox"/> Raw data matrix (Height)	<input type="checkbox"/> Peak ID matrix
<input type="checkbox"/> Normalized data matrix	<input type="checkbox"/> Retention time matrix
<input checked="" type="checkbox"/> Raw data matrix (Area)	<input type="checkbox"/> m/z matrix
<input type="checkbox"/> *Export as mztab-M	<input type="checkbox"/> MS/MS included matrix
	<input type="checkbox"/> GNPS export
	<input type="checkbox"/> S/N matrix export
	<input checked="" type="checkbox"/> Representative spectra
	<input checked="" type="checkbox"/> Parameter

Filtered by blank peaks (must be checked in alignment parameter setting)

☐ Filtering by the ion abundances of blank samples

Missing value option

☐ Replace zero values with 1/10 of minimum peak height over all samples

Isotope labeled tracking option Target file

☐ Filtering by the result of isotope labeled tracking

Export format:

3.2.3 样本信息表格

样本信息文件(.csv 格式)旨在描述样本组信息。第一列命名为“sample.name”，第二列命名为“group”。示例信息文件应如下所示：

	A	B
1	sample.name	group
2	W30.02	W30
3	W03.04	W03
4	W03.05	W03
5	W30.06	W30
6	W03.01	W03
7	W30.07	W30
8	W03.02	W03
9	W03.10	W03
10	W30.10	W30
11	W30.05	W30
12	W30.04	W30
13	W03.07	W03
14	W03.09	W03
15	W30.08	W30
16	W30.01	W30
17	W30.09	W30
18	W03.03	W03
19	W03.06	W03
20	W30.03	W30
21	W03.08	W03

注意：样品信息文件中的“sample.name”列必须与 MS1 表中的样品名称完全相同。

	A	B	C	D	E	F	G	H	I	J	K
1	name	mz	rt	W03.01	W03.02	W03.03	W03.04	W03.05	W03.06	W03.07	W03.08
2	0	60.04458	65.46	72102	59455	61155	75126	70122	69224	71378	69357
3	1	60.07921	487.92	124220	117089	134748	117222	116787	129138	117582	124574
4	2	62.05778	586.44	10943	9937	12202	10017	6119	11312	7882	8187
5	3	62.06028	728.16	18129	25006	24462	24663	31669	25478	23404	27956
6	4	62.98056	641.94	14753	16758	14582	18096	18003	15747	16758	20014
7	5	67.04201	695.16	16481	16328	16694	19801	16327	19707	17584	19727
8	6	68.04762	564	141738	148617	193914	154214	150546	193661	174458	117452
9	7	68.0479	694.86	158573	171982	165931	185563	169561	187873	185445	191220
10	8	68.04813	684.78	70527	62101	75703	78202	59119	69889	70523	63844
11	9	68.04961	481.26	23357	28647	25536	23746	38907	36582	41196	34186
12	10	69.03178	684.72	181394	162288	187964	198937	156747	184717	175418	177185
13	11	69.0449	109.38	201944	84881	123522	29659	4210	135264	233998	148146
14	12	69.04494	83.28	120898	53351	73423	49196	19089	64294	180890	73086
15	13	69.85059	564	27686	30734	40419	31656	31407	39225	35460	22336
16	14	70.06375	77.76	11618	12362	13346	11896	10166	9137	10861	11055
17	15	70.06413	564	4897931	5194146	6821479	5449870	5347568	6640286	6071122	4011793
18	16	70.065	961.38	350034	226732	364269	340037	304235	327483	519223	210826
19	17	70.06603	536.82	646621	683571	672213	365553	331017	721630	807467	778492
20	18	70.06616	581.34	1491928	1583940	1560250	1540114	1432975	1620120	1586224	1455605
21	19	70.06622	503.76	2751747	1872075	2844237	909384	2489225	1493517	3426262	1222921
22	20	70.06625	481.44	891878	1195393	983816	930249	1020217	1244685	1536690	1429955
23	21	71.95018	480.72	30818	18858	23915	38173	42467	25006	30958	57380

3.2.4 RT 校正表格

- 如果您遵循我们的 LC 方法并在数据采集期间运行 RTQC 标准混合物，建议使用 RT 重新校准表重新校准到实验 RT 库。您可以手动输入校准物的保留时间或在参数设置页面上传 RT 重新校准表。
- 如果要上传一个 RT recalibration 表，列名和行的顺序应该和下面的例子完全一样，你应该根据你的实验值修改 calibrant 的 rt（单位是分钟）。

正离子模式下的 RT 校准表格示例：

	A	B	C	D	E	F
1	compound.name	id.zhulab	id.pubchem	ref.mz	rt	polarity
2	3-Methoxy-4-Hydroxyphenylglycol Sulfate	L0228	3035420	263.0231	0.98	negative
3	1-Methylxanthine	L0216	80220	165.0418	3.17	negative
4	N-Acetyl-L-phenylalanine	L0086	74839	206.0822	5.22	negative
5	5-Hydroxyhexanoic acid	S0266	170748	131.0713	7.03	negative
6	5-Hydroxymethylcytidine	L0264	14367004	272.0889	8.49	negative
7	Sebacic acid	S0193	5192	201.1132	10.03	negative
8	Suberic acid	S0214	10457	173.0819	11.08	negative
9	N-Acetyl-L-aspartic acid	S0246	65065	174.0408	12.28	negative
10	N-carbamoyl-L-aspartate	L0084	93072	175.036	13.29	negative
11	UDP-D-glucuronate	L0081	17473	579.027	15.34	negative

负离子模式下的 RT 校准表格示例：

	A	B	C	D	E	F
1	compound.name	id.zhulab	id.pubchem	ref.mz	rt	polarity
2	4-Pyridoxic acid	L0127	6723	184.0605	1.07	positive
3	3'-O-methyladenosine	L0251	82530	282.1197	3.12	positive
4	Creatinine	S0042	588	114.0662	4.95	positive
5	Nicotinate	S0008	938	124.0393	6.65	positive
6	Guanosine	S0071	135398635	284.099	7.92	positive
7	N-Acetylcadaverine	S0249	189087	145.1336	9.13	positive
8	Beta-Alanine	L0320	239	90.055	11.23	positive
9	dAMP	L0140	12599	332.0755	13.28	positive
10	gamma-L-Glutamyl-L-glutamic acid	L0352	92865	277.1031	14.5	positive
11	N6,N6,N6-Trimethyl-L-lysine	L0190	440120	189.1598	16.38	positive

你可以从[此处](#)（加链接）下载测试的 RT 校正表格。

3.3 数据准备的重要说明

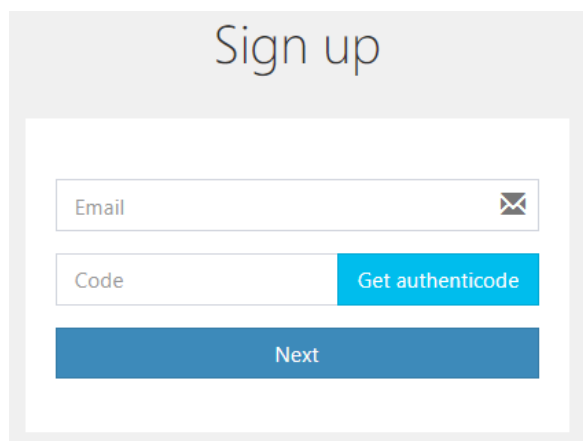
- 在 MS1 峰表中，确保峰名称或样品名称中没有出现“-”或空白。如果有一些我们的程序无法识别的符号，数据处理可能会失败。
- 样品信息文件中的“sample.name”列必须与 MS1 表中的样品名称完全相同。
- 如果您想同时处理正离子数据集和负离子模式数据集，请确保两个峰表中的样本名称相同。
- 样本组至少需要 2 组及以上。
- 请确保样品信息（.csv 格式）和 MS1 表（.csv 格式）是用逗号分隔。因为在一些国家或地区（欧洲和一些法语地区），默认的分隔符是分号。您可以使用记事本或其他文本编辑器打开样品信息或 MS1 峰表，检查它们是否以逗号分隔。

4. MetDNA2 网站使用

请点击 <http://MetDNA.zhulab.cn/> 访问 MetDNA 网络服务器。目前支持 Chrome、Edge 和 Firefox 等常见网络浏览器。

4.1 网站注册

单击“**Analysis**”选项卡开始数据分析。首先，新用户需要在我们的网站上注册。注册后，用户可以使用自己的用户名登录。

A screenshot of a web registration form titled "Sign up". The form is contained within a light gray border. It features three input fields: "Email" with a small envelope icon on the right, "Code", and a blue button labeled "Get authenticode" to the right of the "Code" field. Below these is a wide blue button labeled "Next".

4.2 数据格式检查

MetDNA2 需要一些必要的原始数据处理来获取输入文件。通常，MetDNA 需要导入以下文件进行代谢物鉴定，包括：

- (1) MS1 峰表（.csv 格式，必需），
- (2) MS2 数据文件（.mgf 或 .msp 格式，需要），
- (3) 样品信息表（.csv 格式，必填），
- (4) RT 重新校准表（.csv 格式，可选）。如果您想遵循我们发布的 LC 方法并重新校准 RT 库。

4.3 数据上传

MetDNA2 支持一种极性或两种极性（正离子模式和负离子模式）。

The screenshot shows the 'File Uploader' interface. It has a title bar 'File Uploader'. Below it are three main sections: 'Project Name' with an information icon and a text input field; 'Sample Information' with an information icon, a 'Choose file...' button, and a 'Browse' button; and 'Polarity-related Data' with an information icon, two radio buttons ('Positive' and 'Negative'), and a section for 'Positive' polarity. The 'Positive' section contains 'MS1 Peak Table' and 'MS/MS Data', each with a 'Choose file...' button and a 'Browse' button. At the bottom are 'Upload' and 'Reset' buttons. The 'Positive' radio button is highlighted with a red box.

4.3.1 单个离子模式（正离子或负离子模式）

- (1) 命名你的项目
- (2) 选择样本信息文件
- (3) 根据您的数据选择采集极性
- (4) 选择 MS1 表和 MS/MS 数据
- (5) 点击 **Upload** 按钮上传您的数据文件

4.3.2 双离子模式（正离子和负离子模式）

- (1) 命名你的项目
- (2) 选择样本信息文件
- (3) 选择 “Positive”
- (4) 为正离子模式数据选择 MS1 表和 MS/MS 数据
- (5) 点击上传按钮上传您的数据文件
- (6) 选择 “Negative”
- (7) 为负离子模式数据选择 MS1 表和 MS/MS 数据
- (8) 点击上传按钮上传您的数据文件

4.4 数据处理参数设置

注意：新增加的参数用红色表示

- **MetDNA Version:** 选择“version1”或“version2”运行不同版本的 MetDNA。
“version1”可以在很大程度上重现 MetDNA1 的结果。“version2”采用了新的详细参数。
- **Ionization Polarity:** 根据质谱数据选择“Positive”、“Negative”或“Both”。
- **Liquid Chromatography:** 根据色谱柱条件选择“HILIC”或“RP”。
- **RT Library:** 选择“amide12min ZhuLabRT”、“amide23min ZhuLabRT”或“No”，使用不同的实验 RT 库进行初始种子注释。前两个选项仅在选择“version2”和“HILIC”作为“MetDNA Version”和“Liquid Chromatography”时可用。如果您想重新校准实验 RT 库，点击“Calibration Table”修改校准物的 RT 并保存结果或点击“Upload”上传 RT 重新校准表（请按照上面的格式要求）。
- **MS instrument:** 选择数据采集的仪器。
- **Collision Energy:** 选择 MS2 数据采集的 CE 值。

- **Extension Step:** 选择代谢网络的延伸步数。注：此参数将在 MetDNA2 论文发表后打开
- **Control group:** 根据实验设计选择对照组。
- **Case group:** Case 组与对照组比较，发现显著性变化的代谢特征峰。
- **Univariate Statistics:** 选择“Student t-test”或“Wilcox test”对数据进行分析。
- **Species:** 选择物种取决于你的实验，现在我们支持 21 种不同的物种。
- **Cutoff of P-value:** 定义一个最大的 p 值来决定哪些峰发生了显著变化。
- **Cutoff of Fold-change:** 定义一个倍数变化的截止值来决定哪些峰被改变。
- **P-value Adjustment:** 选择“yes”或“no”来决定是否用 FDR 校正 p-value。

RT recalibration table					
Positive mode		Negative mode			
Name	Zhulab ID	PubChem ID	Ref.mz	RT/minute	Polarity
4-Pyridoxic acid	L0127	6723	184.0605	1.07	positive
3'-O-methyladenosine	L0251	82530	282.1197	3.12	positive
Creatinine	S0042	588	114.0662	4.95	positive
Nicotinate	S0008	938	124.0393	6.65	positive
Guanosine	S0071	135398635	284.0990	7.92	positive
N-Acetylcadaverine	S0249	189087	145.1336	9.13	positive
Beta-Alanine	L0320	239	90.0550	11.23	positive
dAMP	L0140	12599	332.0755	13.28	positive
gamma-L-Glutamyl-L-glutamic acid	L0352	92865	277.1031	14.5	positive
N6,N6,N6-Trimethyl-L-lysine	L0190	440120	189.1598	16.38	positive

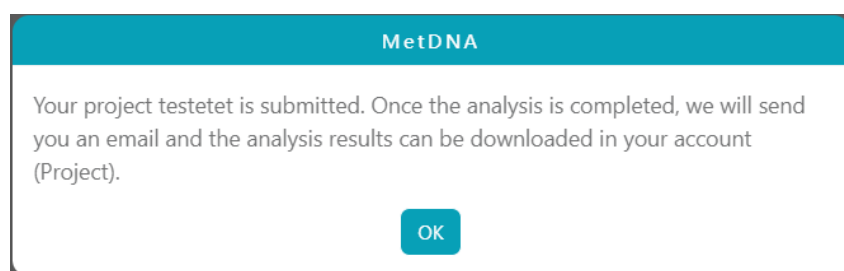
Demo files: RT recalibration table

Upload Save

MetDNA Version	version2	Version of MetDNA
Ionization Polarity	Positive	The acquisition mode [Positive/Negative]
Liquid Chromatography	HILIC	Liquid Chromatography (LC) type [HILIC/RP]
RT Library	Calibration Table amide23min ZhuLabRT	Types of RT libraries
MS Instrument	Sciex TripleTOF	Mass Spectrometer Instrument type
Collision Energy	30	The Collision Energy (CE) of MS/MS data acquisition
Extension Step	0	Steps of metabolic network extension
Control Group	W03	The name of control group
Case Group	W30	The name of case group
Univariate Statistics	Student t-test	Statistic method [Student t-test/Wilcox test]
Species	Drosophila melanogaster (fruit fly)	The species of your project data
Cutoff of P-value	0.05	Usually less than 0.05
Cutoff of Fold-change	1	Usually larger than 1
P-value Adjustment	No	Adjust p values with FDR [Yes/No]

4.5 提交数据

单击“Submit”按钮后，您的数据将提交给 MetDNA2 进行处理。您的屏幕上将弹出一条确认消息，并且一封来自 metdna@sioc.ac.cn 的确认电子邮件将发送到您的电子邮件地址。



4.6 结果下载

4.6.1 邮件接收

- 几个小时后，用户将收到一封来自 metdna@sioc.ac.cn 的通知邮件。您可以从“Project”选项卡下载结果。
- 如果您的项目失败，您还会收到一封电子邮件，您可以在“Project”选项卡中查看日志。
- 请确保将我们的电子邮件地址 metdna@sioc.ac.cn 保存在您的安全电子邮件列表中。

4.6.2 结果查看与下载

对于每个项目，您可以点击“**View**”进入项目的详细页面。

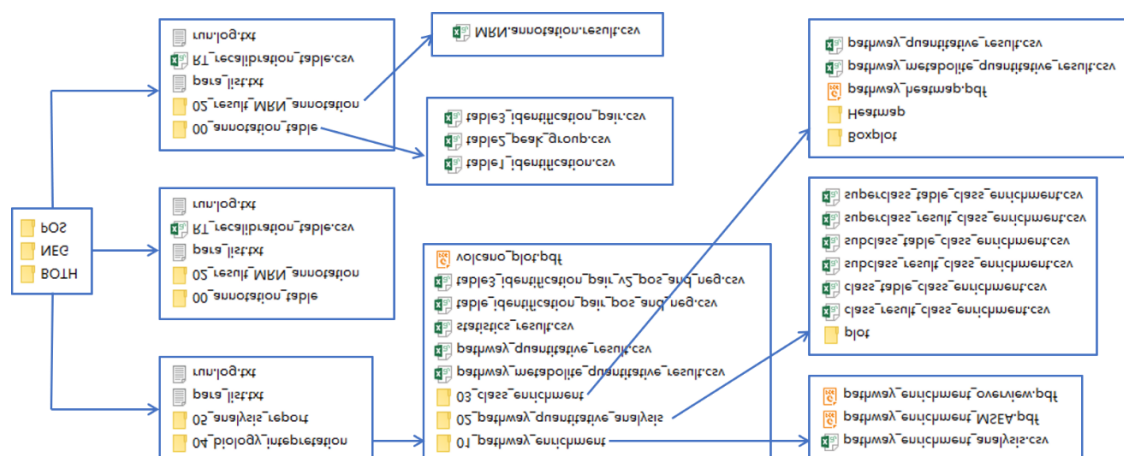
- **Log:** 您可以查看项目是否完成或失败的日志。
- **Download:** 您可以下载项目的结果。
- **Delete:** 删除项目。

Project					
Show	10	▼	entries	Search: <input type="text"/>	
Name	Status	Create time	Expire time	Control	
testetet	Running	2021-07-26	2021-08-02	View	
20210719_test_run_time_metcna2	Success	2021-07-19	2021-07-26	View	Log Download Delete
Showing 1 to 2 of 2 entries					
		Previous 1 Next			

在特定 Project 的 View 页面中，您可以查看日志和分析报告。您还可以下载该项目的结果。

5. MetDNA2 结果解读

MetDNA2 的结果包含 3 个文件夹，分别为 "BOTH", "POS", 和 "NEG"。



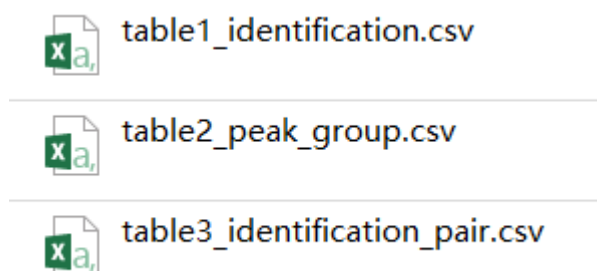
5.1 POS\02_result_MRN_annotation\MRN.annotation.result.csv

该文件位于 "POS\02_result_MRN_annotation\MRN.annotation.result.csv"

这是一个 csv 表，提供正离子模式代谢峰的 MRN 识别结果。

5.2 文件夹 POS\00_annotation_table (仅支持 MetDNA2)










这是一个文件夹，其中包含 MetDNA2 的最终结果表格。



- **table1_identification.csv:** 代谢物鉴定最终的结果表格。
- **table2_peak_group.csv:** 每个代谢物相关的所有衍生代谢物特征峰。
- **table3_identification_pair.csv:** 拆分的代谢特征峰-代谢物组合表格。

5.3 文件夹 BOTH\04_biology_intepretation

该文件夹包含代谢通路和化学类富集分析结果。

	01_pathway_enrichment
	02_pathway_quantitative_analysis
	03_class_enrichment
	pathway_metabolite_quantitative_result.csv
	pathway_quantitative_result.csv
	statistics_result.csv
	table_identification_pair_pos_and_neg.csv
	table3_identification_pair_v2_pos_and_neg.csv
	volcano_plot.pdf

5.3.1 01_pathway_enrichment:

- path_enrichment_analysis: 通路富集分析结果的详细信息。
- path_enrichment_MSEA: 显示通路富集结果的图。
- path_enrichment_overview: 显示通路富集结果的图

5.3.2 02_pathway_quantitative_analysis:

- Boxplot 文件夹: 对每个代谢通路进行定量分析的箱线图。
- Heatmap 文件夹: 每个代谢通路中代谢物的热图。
- path_heatmap.pdf: 展示通路的定量表达热图。

- `path_metabolite_quantitative_result.csv`: 包含显著变化代谢物定量信息的表格。
- `path_quantitative_result.csv`: 包含每个样本中代谢通路定量信息的表格。

5.4 文件夹 BOTH\05_analysis_report "

该文件夹包含 MetDNA2 的简要报告和信息总结（HTML 格式）。