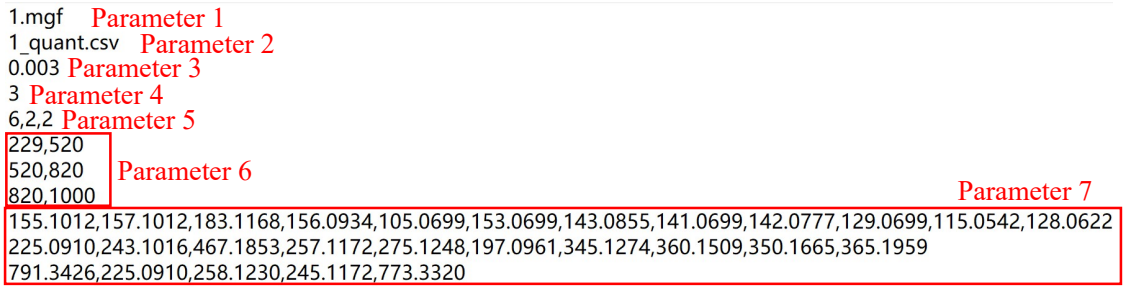
User manual of NPExtractor



Pretreat the LC-MS/MS raw data with MZmine 2.53 (or other MZmine 2 versions, please do not use the newly released MZmine 3) to extract the total features and obtain the mgf and csv files. Use the Notepad program to read the mgf file and replacing all key words “FEATURE_ID” with “TITLE”. All parameters are set in the input.txt. Please ensure that the mgf file, the csv file, the NPExtractor.exe tool and the input.txt file are put into a folder.

The explanation of all parameters is shown as follows:

Parameter 1: the file name of the mgf file generated by MZmine 2.

Parameter 2: the file name of the csv file generated by MZmine 2.

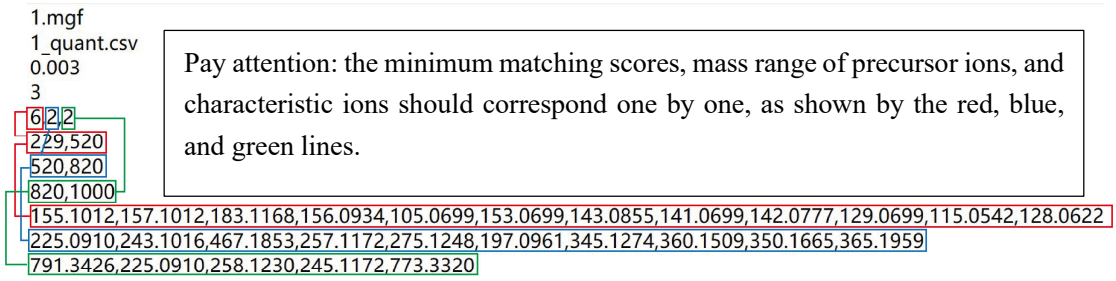
Parameter 3: the *m*/*z* tolerance (Da) of matching the characteristic ion. The default is 0.003 Da.

Parameter 4: the Number of natural product categories.

Parameter 5: the minimum matching score. The number of minimum matching score should be consistent with that of natural product categories.

Parameter 6: the precursor mass range and classification of different types of natural products.

Parameter 7: the calculated mass of the characteristic ions for identifying the corresponding target natural products.



We can modify the number of natural product categories to complete the analysis of other natural products:

Example 1:

1.mgf
1_quant.csv
0.003
2
3,4
229,520
520,900
100.1234,200.1265,300.3366
456.1225,589.9875,789.6541

Example 2:

1.mgf
1_quant.csv
0.003
5
3,2,2,2,2
100-400
300-500
490-720
600-900
900-1500
100.1234,200.1265,300.3366,265.5589,156.3699,148.4569
456.1225,389.9875,289.6541
498.1234,688.1265,532.3366,468.1399
688.1244,798.1265,632.3369,868.1398, 4855.3265
988.1244,1098.1221,1232.4469,1368.13948, 1433.3685

Replace all key words "FEATURE_ID" with "TITLE"

BEGIN IONS
FEATURE_ID=3
PEPMASS=108.9449
SCANS=3
RTINSECONDS=105.055
CHARGE=1
MSLEVEL=2
MERGED_STATS=3 / 6 (0 removed due to low quality, 3 removed due to low cosine).
54.0310 7.5E1
54.0572 3.5E1
56.0469 9.6E1
56.0583 4.0E1
57.0442 7.0E1
350.0318 4.5E1
END IONS

BEGIN IONS
TITLE=3
PEPMASS=108.9449
SCANS=3
RTINSECONDS=105.055
CHARGE=1
MSLEVEL=2
MERGED_STATS=3 / 6 (0 removed due to low quality, 3 removed due to low cosine).
54.0310 7.5E1
54.0572 3.5E1
56.0469 9.6E1
56.0583 4.0E1
57.0442 7.0E1
350.0318 4.5E1
END IONS

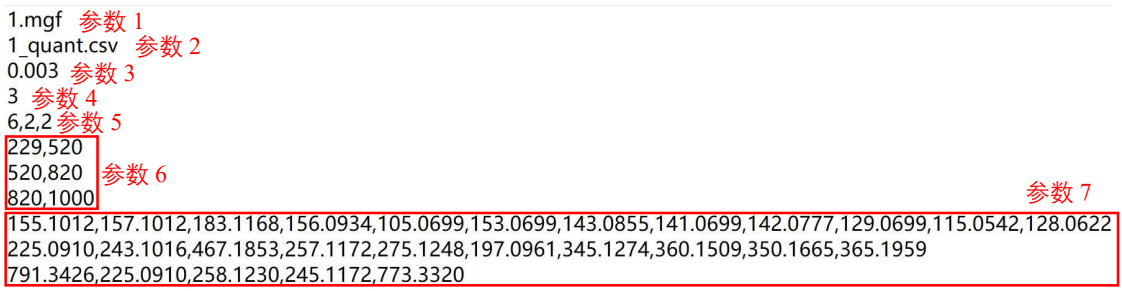
After setting the parameters in the input.txt, save the file and double click the NPExtractor.exe to run the program. Three files (target.mgf, 'output for check.csv', and 'output for FBMN.csv' file) are generated. The target.mgf and the 'output for FBMN.csv' files can be uploaded to GNPS platform for FBMN analysis. The 'output for check.csv' records the classification information. We use numbers to represent the classification of different types of natural products, as the example shown below.

1.mgf
1_quant.csv
0.003
3
6,2,2
229,520 → Classification = 0
520,820 → Classification = 1
820,1000 → Classification = 2
155.1012,157.1012,183.1168,156.0934,105.0699,153.0699,143.0855,141.0699,142.0777,129.0699,115.0542,128.0622
225.0910,243.1016,467.1853,257.1172,275.1248,197.0961,345.1274,360.1509,350.1665,365.1959
791.3426,225.0910,258.1230,245.1172,773.3320

	Classification	row ID	row m/z	row retention time
2	0	14351	229.1225	9.04925
3	0	14353	229.1225	23.1385
4	0	14848	231.1373	9.617883
5	0	14850	231.1376	11.55825

NPExtractor 使用方法

注意：在 input.txt 中输入时需要用英语输入法，否则会报错。



首先使用 MZmine 2 软件对 LC-MS/MS 数据进行处理，提取其中所有化合物的特征（features），得到 mgf 和 csv 文件，暂时不要使用 MZmine 3 进行处理。得到的 mgf 文件先用记事本打开，把所有“FEATURE_ID” 替换为 “TITLE”。所有的参数设置直接写入 input.txt，由于程序使用了相对路径，所以在使用 NPExtractor 时保证 NPExtractor.exe，mgf，csv 和 input.txt 文件在同一个（空的）文件夹中。

Input.txt 的参数解释如下所示：

参数 1：使用 MZmine 2 处理原始数据后得到的 mgf 文件的文件名。

参数 2：使用 MZmine 2 处理原始数据后得到的 csv 文件的文件名。

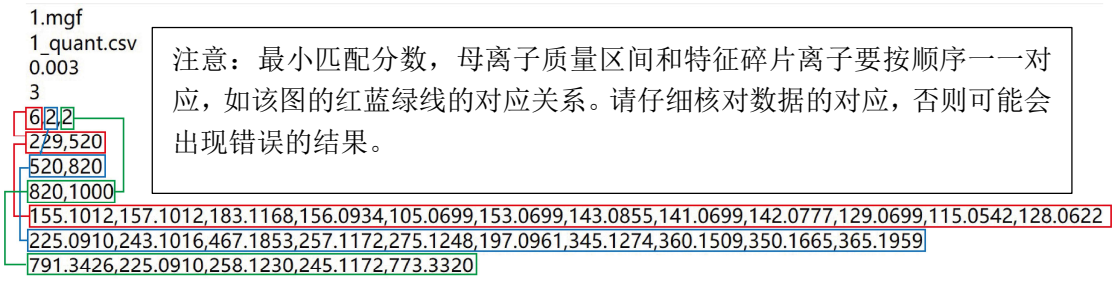
参数 3：特征碎片离子识别匹配的容许质量误差（Da），默认值为 0.003Da。

参数 4：天然产物分类数。

参数 5：特征碎片离子最低匹配分数。要注意的是，有几种天然产物分类就应有对应的匹配的分数（如例子中，参数 4 的分类数为 3，则就有参数 5 中 6,2,2 三个匹配分数）。

参数 6：不同分类的天然产物的母离子质量区间。

参数 7：用于识别不同分类的天然产物的特征碎片离子的质量（理论计算质量，非实测值）。



我们可以进一步更改 input.txt 中天然产物的分类数，最低匹配分数和特征碎片离子数据以完成不同类型天然产物的分析。

例 1：

```
1.mgf
1_quant.csv
0.003
2
3,4
229,520
520,900
100.1234,200.1265,300.3366
456.1225,589.9875,789.6541
```

例 2：

```
1.mgf
1_quant.csv
0.003
5
3,2,2,2,2
100-400
300-500
490-720
600-900
900-1500
100.1234,200.1265,300.3366,265.5589,156.3699,148.4569
456.1225,389.9875,289.6541
498.1234,688.1265,532.3366,468.1399
688.1244,798.1265,632.3369,868.1398, 4855.3265
988.1244,1098.1221,1232.4469,1368.13948, 1433.3685
```

在记事本中把所有“FEATURE_ID”替换为“TITLE”

```
BEGIN IONS
FEATURE_ID=3
PEPMASS=108.9449
SCANS=3
RTINSECONDS=105.055
CHARGE=1
MSLEVEL=2
MERGED_STATS=3 / 6 (0 removed due to low quality, 3 removed due to low cosine).
54.0310 7.5E1
54.0572 3.5E1
56.0469 9.6E1
56.0583 4.0E1
57.0442 7.0E1
350.0318 4.5E1
END IONS
```

```
BEGIN IONS
TITLE=3
PEPMASS=108.9449
SCANS=3
RTINSECONDS=105.055
CHARGE=1
MSLEVEL=2
MERGED_STATS=3 / 6 (0 removed due to low quality, 3 removed due to low cosine).
54.0310 7.5E1
54.0572 3.5E1
56.0469 9.6E1
56.0583 4.0E1
57.0442 7.0E1
350.0318 4.5E1
END IONS
```

设置好 input.txt 中的参数后保存文件，直接双击 NPExtractor.exe 运行程序。数秒后会得到 target.mgf，‘output for check.csv’和‘output for FBMN.csv’ file 三个文件。其中 target.mgf 和‘output for FBMN.csv’可以直接上传至 GNPS 平台进行 FBMN 分析。在‘output for check.csv’文件中，记录了化合物分类的信息。我们使用数字来代表分类信息。按照分类的排列顺序一次使用 0，1，2…来表示。

```
1.mgf
1_quant.csv
0.003
3
6,2,2
229,520 ——▶ Classification = 0
520,820 ——▶ Classification = 1
820,1000 ——▶ Classification = 2
155.1012,157.1012,183.1168,156.0934,105.0699,153.0699,143.0855,141.0699,142.0777,129.0699,115.0542,128.0622
225.0910,243.1016,467.1853,257.1172,275.1248,197.0961,345.1274,360.1509,350.1665,365.1959
791.3426,225.0910,258.1230,245.1172,773.3320
```

	Classification	row ID	row m/z	row retention time
2	0	14351	229.1225	9.04925
3	0	14353	229.1225	23.1385
4	0	14848	231.1373	9.617883
5	0	14850	231.1376	11.55825

如有其他问题请联系 y1036752587@163.com（黎永毅）。