0 Recommended workflow for untargeted features acquisition and MS2MP-based pathway enrichment analysis

Full-scan acquisition is performed in Full MS/ddMS^2 mode, with resolution set to 120,000 for full scans and 30,000 for MS/MS. The scan range is set from 85 to 1250 m/z. The top 10 ions from each full scan are selected for fragmentation. Each metabolic feature represents a unique uncharacterized metabolite, defined by its MS1 precursor ion, corresponding MS2 spectrum, and retention time. MS1 ions are generated via ion fusion to eliminate redundancy from isotopic and adduct ions originating from the same metabolite. For MS1-MS2 matching, deviation thresholds of 0.005 for m/z and 12 seconds for retention time are applied. When multiple candidates are present, the MS2 spectrum with the lowest combined deviation in m/z and retention time is selected.
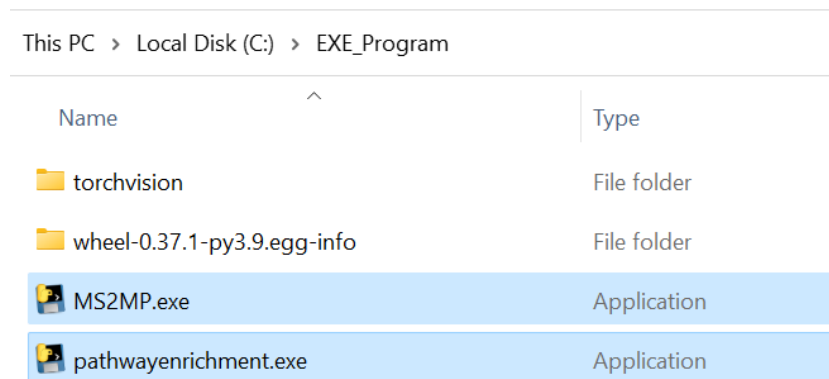
The MS2MP-based enrichment analysis employs over-representation analysis (ORA), integrating MS1-based quantification and MS2-based pathway annotation. Significantly differential metabolic features between groups are identified using MS1 relative abundances normalized to internal standards (p < 0.05, t-test). Pathway predictions are assigned to these features by MS2MP based on their MS2 spectra. These pathway- predicted features are then utilized for ORA to identify enriched pathways, which is based on a hypergeometric test, formulated as follows:

$$P(X >= k) = 1 - \sum_{j=0}^{k-1} \frac{\binom{K}{j}\binom{N-K}{n-j}}{\binom{N}{n}}$$

where N and n represent the total number and significantly differential number of MS2MP-predicted metabolic features, respectively. K and k represent the total number and significantly differential number of MS2MP-predicted metabolic features in a specific pathway, respectively.

1 Overview

Download the file "EXE_Program.7z" and extract it. Inside, you will find two executable files, "MS2MP.exe" and "pathwayenrichment.exe" (**Figure 1**). "MS2MP.exe" is used for predicting the metabolic pathways of MS/MS spectra, while "pathwayenrichment.exe" is used for pathway enrichment analysis. The usage procedures for these two executable files will be explained separately using example files.

Figure 1

2 Usage Procedure for "MS2MP.exe"

2.1 Locate and open the folder "data" within the folder "EXE_Program" (**Figure 2**).

Name	Type
📁 certifi	File folder
📁 contourpy	File folder
📁 data	File folder

Figure 2

2.2 Inside the folder "data", locate and open the folder "jsontest" (**Figure 3**).

Name	Type
📁 jsontest	File folder
📄 test.csv	Microsoft Excel Com...

Figure 3

2.3 Place the fragmentation tree files (JSON format), generated by the software of SIRIUS, into the folder "jsontest". The file names should only contain numbers without the ".json" extension (**Figure 4**).

Name	Type
📄 23456789.json	JSON File
📄 93692851.json	JSON File

Figure 4

2.4 Go back to the folder "data" and find the file "test.csv", open it using the software of "Excel" (**Figure 5**).

Name	Type
📁 jsontest	File folder
📄 test.csv	Microsoft Excel Com...

Figure 5

2.5 In the "jsoninput" column of the file "test.csv", enter the file names from step 2.3 without the ".json" extension. Do not modify data in other cells (**Figure 6**).

	A	B
1	jsoninput	NO INPUT
2	23456789	0,1,2,3,4
3	93692851	0,1,2,3,4

Figure 6

2.6 Go back to the folder "EXE_Program" and run the file "MS2MP.exe" (**Figure 7**).

Figure 7

2.7 The initial interface of "MS2MP.exe" is shown in **Figure 8**. Follow the prompts and input the number of predictions required, which corresponds to the number of JSON files from step 2.3. After "MS2MP.exe" finishes running, the interface will appear as shown in **Figure 9**, and "Please input any key to exit" will be displayed. The prediction results will be shown within the red box (**Figure 9**) and can also be viewed in the file "Results.txt" (**Figure 10-11**).
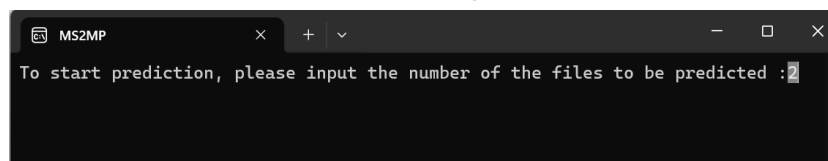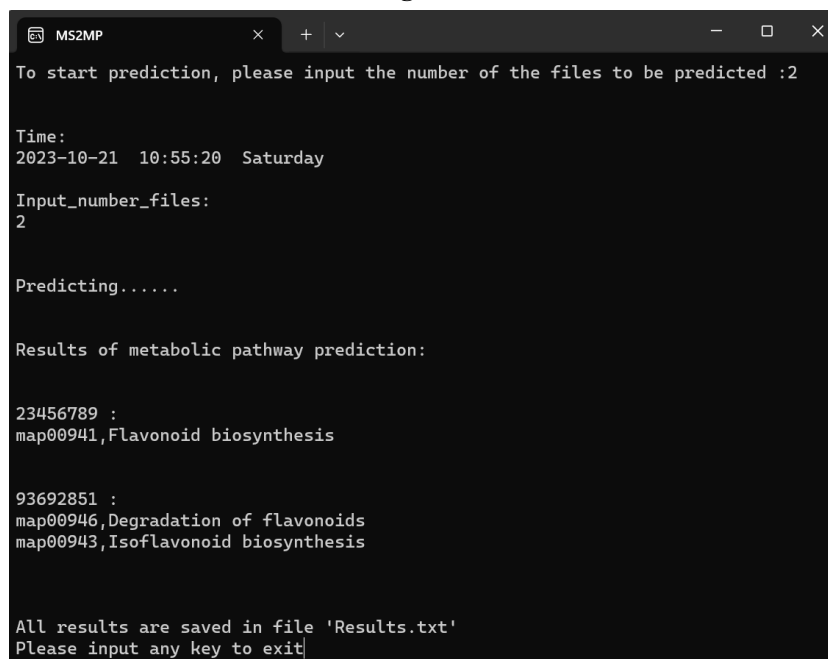

Figure 8
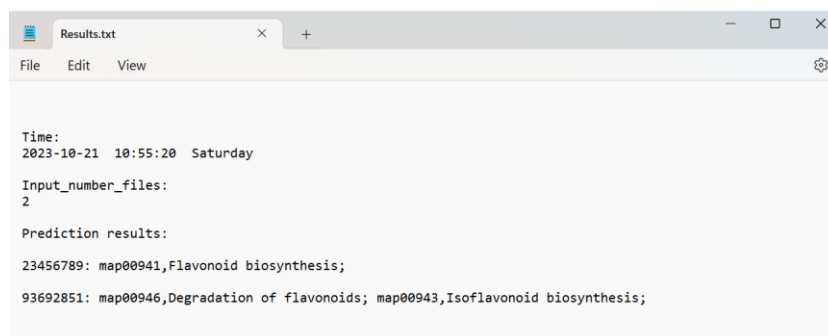

Figure 9

Figure 10

Figure 11

3 Usage Procedure for "pathwayenrichment.exe"

3.1 Locate and open the folder "data_pathwayenrichment" within the folder "EXE_Program" (**Figure 12**).

Figure 12

3.2 Inside the folder "data_pathwayenrichment", find and open the file "input_pathwayenrichment.csv" using the software of "Excel" (**Figure 13**).

Figure 13

3.3 Enter the required data in the file "input_pathwayenrichment.csv" (**Figure 14**). For pathway enrichment analysis based on compound identification, data in column A should be the pathway names, data in column B should be the total number of compounds in the selected pathway background, data in column C should be the total number of compounds in this pathway, data in column D should be the number of significantly different metabolites in the experimental sample, and data in column E should be the number of significantly different metabolites in this pathway. For MS2 spectra-based pathway enrichment analysis, column A represents the pathway names, column B represents the total number of MS2MP-predicted metabolic features, column C represents the total number of MS2MP-predicted metabolic features in a specific pathway, column D represents the significantly differential number of MS2MP-predicted metabolic features, and column E represents the significantly differential number of MS2MP-predicted metabolic features in a specific pathway.

Figure 14

3.4 Go back to the folder "EXE_Program" and run the file "pathwayenrichment.exe" (**Figure 15**).

Name	Type
pathwayenrichment.exe	Application
pyexpat.pyd	PYD File
python3.dll	Application extension

Figure 15

3.5 The interface of "pathwayenrichment.exe" during runtime is shown in **Figure 16**. When "Please input any key to exit" appears, it indicates that the program has finished running. The prediction results can be found in the file "result_pathwayenrichment" (**Figure 17**) within the folder "data_pathwayenrichment" (**Figure 13**). Data in the last column (F) are the p-values for pathway enrichment.

```
pathwayenrichment           ×    +  ∨                              —   □   ×

data loading......
['mapXXXXXX', 'mapYYYYYY', 'mapZZZZZZ'] ['5000', '5000', '5000'] ['500', '500', '500'] ['
13', '29', '51'] ['10', '16', '24']


results:
mapXXXXXX 5000 13 500 10 1.9915537663450777e-08
mapYYYYYY 5000 29 500 16 1.5710897027787496e-09
mapZZZZZZ 5000 51 500 24 1.0270007066992548e-11


All results are saved in file 'result_pathwayenrichment.csv'
Please input any key to exit
```

Figure 16

	A	B	C	D	E	F
1	mapXXXXXX	5000	13	500	10	1.99E-08
2	mapYYYYYY	5000	29	500	16	1.57E-09
3	mapZZZZZZ	5000	51	500	24	1.03E-11

Figure 17