

Usage for MetCleaning cn

JasperShen

2016-12-22

只做 MS1 和 MS2 匹配

```
##加载包
library(MetCleaning)
##读取数据
met.data <- ImportData(data = "data.csv",
                        sample.information = "sample.information.csv",
                        polarity = "positive")
##开始 ms1 和 ms2 匹配
PeakIdentification(MetFlowData = met.data,
                   mz.tolerance = 30,
                   rt.tolerance = 180)
```

参数含义

1. data: 要处理的一级数据的名字。
2. sample.information: 要处理的一级数据的样品信息。
3. polarity: 数据的采集模式。"positive"或者"negative"。
4. MetFlowData: 要进行一级和二级匹配的数据的数据。
5. mz.tolerance: 匹配时的 mz 误差范围。默认为 30(± 30 ppm)。
6. rt.tolerance: 匹配时的 rt 误差范围。默认为 180(± 180 s)。

做所有处理过程

```
##加载包
library(MetCleaning)
##开始处理
MetClean(data = "data.csv",
          sample.information = "sample.information.csv",
          polarity = "positive",
          obs.zero.cutoff = 0.5,
          var.zero.cutoff = 0.5,
          method = "svr",
          threads = 2,
          mz.tolerance = 30,
          rt.tolerance = 180,
          met.plot = TRUE)
```


参数含义


1. data: 要处理的一级数据的名字。

- sample.information: 要处理的一级数据的样品信息。
- polarity: 数据的采集模式。"postive"或者"negative"。
- obs.zero.cutoff: 样品零值筛选的阈值，当一个样品中的非零值超过这个阈值的时候，该样品应该被除去。
- var.zero.cutoff: 变量零值筛选的阈值，当一个变量中的非零值超过这个阈值的时候，该变量应该被除去。
- method: normalization 方法。
- threads: 线程个数。
- mz.tolerance: 匹配时的 mz 误差范围。默认为 30(± 30 ppm)。
- mz.tolerance: 匹配时的 rt 误差范围。默认为 180(± 180 s)。
- met.plot: 是否要把所有峰的 scatter plot 画出来。

处理所需数据，另外需要单独另外建立一个文件夹，命名为 **peak identification**，里面放入二级数据，且二级数据命名中必须包含 "ms2"。

| Feature name | | | | m/z | rt | Other information of features | | QC and subject sample abundance | | | |
|--------------|----------|-----------|----------|----------------------|----|-------------------------------|----------|---------------------------------|----------|------|--|
| | | A | B | C | D | E | F | G | H | | |
| 1 | name | mz | rt | isotopes | | | QC22 | QC23 | QC12 | QC24 | |
| 2 | M72T49 | 72.080982 | 49.212 | [1][M] ⁺ | | 1812140 | 1544984 | 1794878 | 1659981 | | |
| 3 | M70T31 | 76.075904 | 30.995 | [2][M] ⁺ | | 516119.9 | 516731.3 | 546147.1 | 378761.1 | | |
| 4 | M64T38 | 84.081043 | 37.514 | [3][M] ⁺ | | 2062105 | 1926205 | 1821621 | 1691391 | | |
| 5 | M60T74 | 86.096638 | 73.766 | [4][M] ⁺ | | 3450548 | 3086142 | 3430762 | 3144324 | | |
| 6 | M60T90 | 86.096646 | 89.873 | [5][M] ⁺ | | 2390322 | 2160545 | 2448855 | 1921851 | | |
| 7 | M100T149 | 100.07584 | 149.427 | [6][M] ⁺ | | 4369935 | 3627019 | 4198031 | 3978864 | | |
| 8 | M103T150 | 103.05438 | 149.991 | [7][M] ⁺ | | 191304.8 | 154571.8 | 176281.5 | 171652.4 | | |
| 9 | M104T30 | 104.10736 | 29.5285 | [8][M] ⁺ | | 10220169 | 9599432 | 10928312 | 8911141 | | |
| 10 | M110T652 | 110.02015 | 952.103 | [9][M] ²⁺ | | 642884.8 | 564758.9 | 799711.8 | 528291.1 | | |
| 11 | M114T32 | 114.06639 | 32.413 | [10][M] ⁺ | | 1369850 | 1362086 | 1795623 | 1643301 | | |
| 12 | M118T36 | 118.08637 | 36.0715 | [11][M] ⁺ | | 12726533 | 13111893 | 12571014 | 12803861 | | |
| 13 | M119T379 | 119.08572 | 378.5005 | [12][M] ⁺ | | 94376 | 78395.41 | 114665.7 | 36013.0 | | |
| 14 | M120T150 | 120.08085 | 149.995 | [13][M] ⁺ | | 4001522 | 3321928 | 3837926 | 3555599 | | |
| 15 | M122T160 | 122.09645 | 159.657 | [14][M] ⁺ | | 232127.1 | 204514.5 | 238401.2 | 214411 | | |
| 16 | M126T165 | 126.0915 | 164.642 | [15][M] ⁺ | | 296192.1 | 228613.8 | 281846.4 | 256762.1 | | |
| 17 | M130T64 | 130.05002 | 63.6625 | [16][M] ⁺ | | 513595.2 | 378566.3 | 398876.5 | 503552.1 | | |
| 18 | M130T157 | 130.06516 | 157.058 | [17][M] ⁺ | | 239890.3 | 186114.1 | 252431.1 | 213720.1 | | |
| 19 | M130T38 | 130.08644 | 38.384 | [18][M] ⁺ | | 336137 | 506422.3 | 367187.6 | 368729.1 | | |

 data.csv

 sample.information.csv

| Sample name | | Injection order | Sample class | Sample batch | Sample group | |
|-------------|-------------|-----------------|--------------|--------------|--------------|----|
| | | A | B | C | D | E |
| 1 | sample.name | injection.order | class | batch | group | |
| 2 | QC11 | 1 | QC | | 1 | QC |
| 3 | A5551 | 2 | Subject | | 1 | 9 |
| 4 | A4880 | 3 | Subject | | 1 | 1 |
| 5 | C1282 | 4 | Subject | | 1 | 0 |
| 6 | C1492 | 5 | Subject | | 1 | 9 |
| 7 | A5730 | 6 | Subject | | 1 | 1 |
| 8 | X1421 | 7 | Subject | | 1 | 0 |
| 9 | X2 | 8 | Subject | | 1 | 0 |
| 10 | C1059 | 9 | Subject | | 1 | 1 |
| 11 | QC12 | 10 | QC | | 1 | QC |
| 12 | C1397 | 11 | Subject | | 1 | 0 |
| 13 | A5819 | 12 | Subject | | 1 | 1 |
| 14 | C1137 | 13 | Subject | | 1 | 0 |
| 15 | A3867 | 14 | Subject | | 1 | 1 |
| 16 | C1223 | 15 | Subject | | 1 | 0 |
| 17 | C1295 | 16 | Subject | | 1 | 0 |
| 18 | C1510 | 17 | Subject | | 1 | 0 |
| 19 | C1121 | 18 | Subject | | 1 | 0 |

对于那些可以超过零值比例的样品，可以选择是否删除掉。

Missing values filter...

No QC should be removed.

x257 A5546 x231 should be removed!!!

Subject should be removed are:

65 160 177

which subject you want to remove (please type the index of subject sample, 65, 160, 177) and separate them using comma, if you don't want to remove any subject, please type n):