Help document of rtcor

Jasper Shen

2017.2.17

## Introduction

对于大规模代谢组学数据，不同batch采集的数据的RT会发生变化，因此对于不同batch的数据整合，和有标准RT的数据库匹配等会有一定的困难，因此，通过在采集数据的同时，在worklist中间插入RTQC(混标)，从而对不同batch的数据的RT做监测，然后对两个不同batch的RTQC做RT的校正曲线，从而可以对不同batch的数据做校正。因为标准品的RT range有一定范围，因此，校正只能适用于处在RTQC RT range范围之内的peak。rtcor就是适用于使用RTQC的两个batch数据之间校正的一个R包。

## Data organization

两种提供RTQC信息的方式。 1. 需要从使用数据处理软件处理后的metabolomics中提取RTQC混标。利用两个batch的data，分别使用XCMS处理，在处理过程中，每个batch的RTQC别单独设置为一组，因此最后RTQC的信息包含在metabolomics数据中。需要metabolomics数据中去提取RTQC混标的RT信息。因此，对于这种情况，RTQC在metabolomics的命名需要包含"RTQC"。这种情况，在文件夹中需要提供三个文件，分别是第一个batch的数据，第二个batch的数据，以及第二个batch的数据，另外一个文件命名为"RTQC.info.csv"，提供RTQC混标的名称，mz以及RT信息。如下图所示： 2. 另外一种情况是不需要从数据处理软件处理后的metabolomics数据中提取RTQC混标信息，而是从原始数据中提取RTQC混标的RT信息。这时候文件夹中需要提供四个文件，两个batch的数据，以及两个batch的RTQC的混标RT信息，需要命名为"RTQC1.csv"和"RTQC2.csv"。如下图所示：

## RT校正

### 设置工作路径

按照上面所述，将数据放置在文件夹中后，将该文件夹设置为工作路径。

setwd("mypath/workdirectory")

### 运行函数

开始运行函数。

library("rtcor")  
rtcor(ref.data = "data1.csv",  
 mz.tolerance = 50,  
 rt.tolerance = 180,  
 max.missing = 0)

函数的参数含义如下

* ref.data：将哪个数据作为参考数据，则校正数据的RT会校正到参考数据的体系中去。
* mz.tolerance：第一种情况下，去寻找RTQC混标时的mz的tolerance。
* rt.tolerance：第一种情况下，去寻找RTQC混标时的RT的tolerance。
* max.missing：第一种情况下，可忍受的不能寻找到的混标的最大个数。