MRC.R user instruction

* MRC.R is a program written in R to correct the MS signal intensities to their equivalent QC loading amounts based on the best regression model.
* The MRC.R script is freely available for non-commercial use.
* The instructions for using the MRC.R are given below.

1. Download and install R studio following the instruction in Rstudio website: <https://rstudio.com/>
2. Within the same folder, prepare a **sample intensity table (File 1)** and a **QC intensity table (File 2)** in following format:

**File 1 (prepared in csv format)**

Sample intensity table contains all intensities of metabolic feature in real samples.

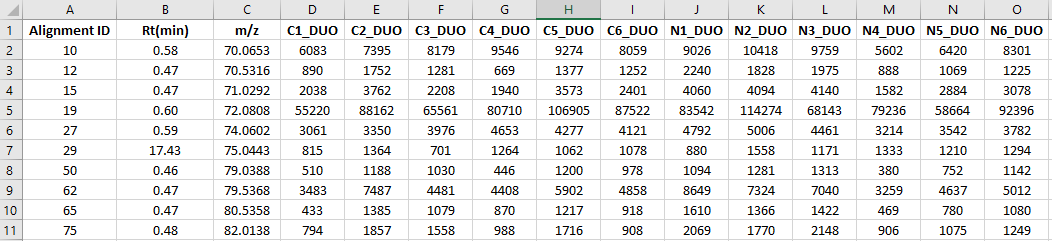
Column 1: alignment ID.

Column 2: retention time.

Column 3: m/z value.

Column 4 and after: MS signal intensities of real samples.

The example dataset format is as follows:



**File 2 (Prepared in csv format)**

QC intensity table contains all intensities of metabolic feature in serial QC samples.

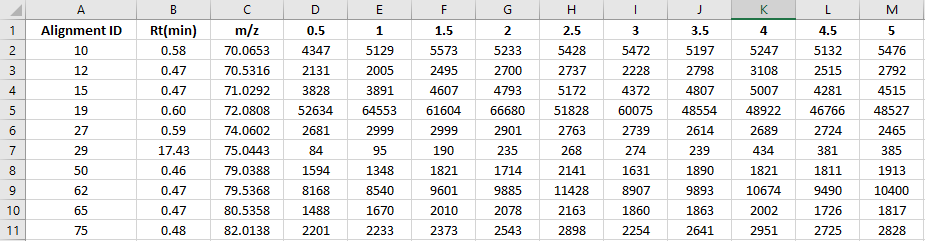
Column 1: alignment ID. Note: The IDs in the QC intensity table should be the same as the IDs in the sample intensity table (File 1). Both IDs should be ordered in the same sequence.

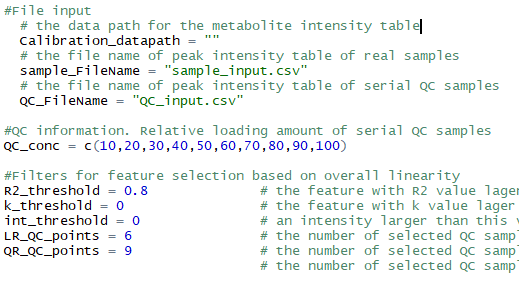
Column 2: retention time.

Column 3: m/z value.

Column 4 and after: serial QC samples with different loading amount (from low to high).

The example dataset format is as follows:

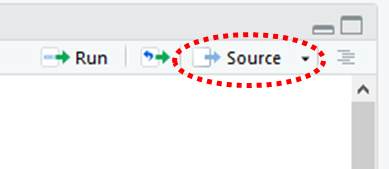


1. Open the MRC.R script (see below) in Rstudio (<https://rstudio.com/>) and change the parameters therein (see **Table 1** for the explanation of these parameters) 

**Table 1.** Instruction of parameters in QC\_cal.R script that can be tuned as user needed

|  |  |
| --- | --- |
| Parameter | Function |
| Calibration\_datapath | Assign the data path for the folder that contains MRC.R and other required “.csv” files for MRC workflow |
| sample\_FileName | Input the file name of the sample intensity table(**File 1**) |
| QC\_FileName | Input the file name of the QC intensity table(**File 2**) |
| QC\_conc | A series of number indicating the relative loading amount of serial QC samples. For example, 10,20,30,40,50,60,70,80,90,100. |
| R2.threshold | In overall linearity test, set the *R2* threshold to filter out the features with poor positive correlation between MS signal intensity and QC loading amount |
| k.threshold | In overall linearity test, set the *k* threshold to filter out the features with negative correlation between MS signal intensity and QC loading amount |
| int\_threshold | For serial QC data, set intensity threshold to selected the valid QC data points. A QC data with MS signal intensity larger than this threshold will be selected for model selection and signal correction. |
| LR\_QC\_points | For a metabolic feature, if the number of selected QC data points less than this value, it will be corrected by directly using linear model |
| QR\_QC\_points | For a metabolic feature, if the number of selected QC data points less than this value but no less than LR\_QC\_points, it will be corrected using linear or quadratic model based on cross-validation. If the number of selected QC samples no less than QR\_QC\_points, it will be corrected using linear quadratic, or cubic model based on cross-validation. |

1. Save the parameter changes to the script. Click “source” on top right to run the script.



1. After running the script, an output file named “calibrated\_sample\_table.csv” will be created in the same file folder. This file contains the corrected sample MS signal intensities as well as the notations.

**Table 2**: the labels in “notation” column and their definitions

**Table 3:** the labels in “model” column and their definitions

**Table 2.** Labels in “notation” column and their definitions

|  |  |
| --- | --- |
| Label | Definition |
| Insufficient\_QC\_points | For a given metabolic feature, if the number of QC data points is smaller than the defined threshold, MRC workflow cannot be applied. Thus, “Not enough QC data points” will be assigned to this feature. |
| Not suitable for calibration | For a given metabolic feature, if it doesn’t meet the requirement of overall linearity test (*k* and *R2* thresholds), “Not suitable for calibration” will be assigned to this feature. |
| Good for calibration | The features that are suitable for signal correction. |

**Table 3.** Labels in “model” column and their definitions

|  |  |
| --- | --- |
| **Label** | **Definition** |
| NA | The signal intensities of these features are not corrected. |
| Uncali. | The measured signal intensities are directly used for ratio calculation |
| Linear | Linear model is used to correct these features |
| Quadratic | Quadratic model is used to correct these features |
| Cubic | Cubic model is used to correct these features |

For “QC\_number” column, the number of selected serial QC data was showed for each metabolic feature. For example, a number “8” meant there were 8 QC data points meeting the requirement of intensity threshold (int\_threshold parameter).