PowerU.R user instructions

PowerU.R is a program written in R to:

1. Select features which are suitable for signal calibration

2. Perform cross-validation for each selected metabolic feature to choose the best regression model

3. Construct a calibration curve using serial QC injections under the best regression model

4. Convert MS signal intensities in real samples to the corresponding relative QC concentration using the built model

5. Perform Shapiro-Wilk normality test on converted data

6. Select running t-test or U test according to data normality (designed for two groups)

7. Add notations for each metabolic feature

The PowerU.R script is freely available for non-commercial use.

The instructions for using the PowerU.R are given below.

1. Download and install RStudio following the instructions on the RStudio website: <https://rstudio.com/>
2. Download PowerU.R from GitHub (<https://github.com/HuanLab/PowerU>), and save it in a folder.
3. Within the same folder, prepare a **sample intensity table (File 1)** and a **QC intensity table (File 2)** in the following format:

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

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

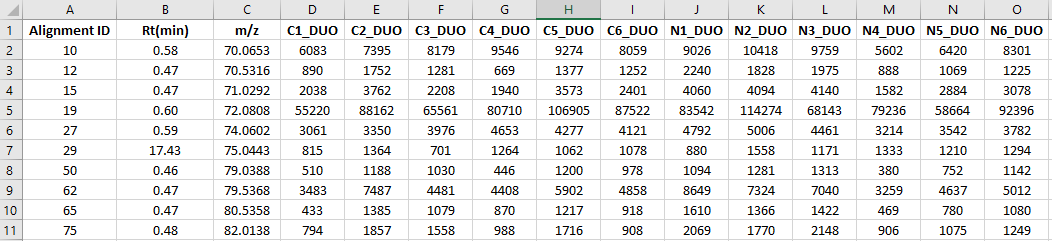
Column 1: alignment ID.

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 (in .csv format)**

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

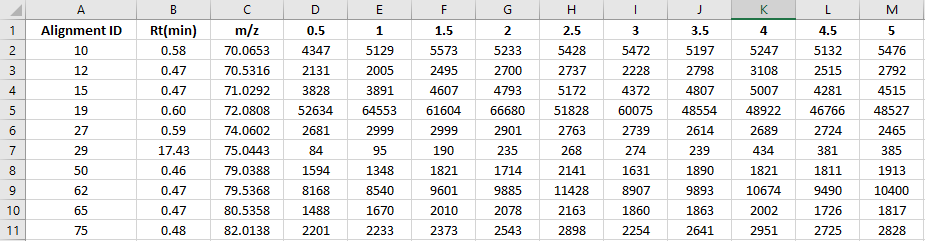
Column 1: alignment ID. Note: The IDs in the QC intensity table should be the same as the IDs in the sample intensity table (File 1). Both ID columns should be in the same order.

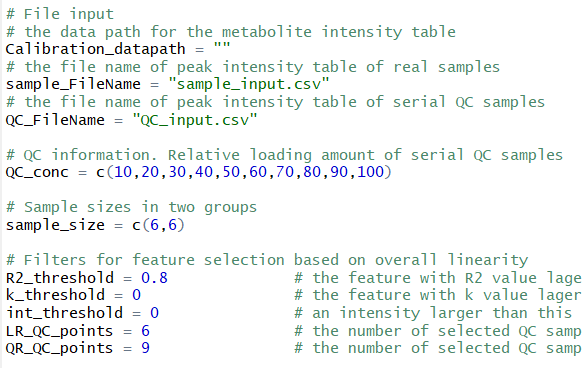
Column 2: retention time.

Column 3: *m/z* value.

Column 4 and after: MS signal intensities of serial diluted QC samples with different loading amount (from low to high).

The example dataset format is as follows:

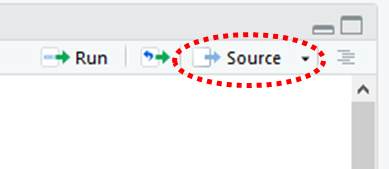


1. Open the PowerU.R script (see below) in Rstudio and change the parameters therein (see **Table 1** for parameter details). 

**Table 1.** Instruction of parameters in PowerU.R script that can be tuned as needed.

|  |  |
| --- | --- |
| **Parameter** | **Function** |
| Calibration\_datapath | Assign the data path for the folder that contains PowerU.R and other required “.csv” files for PowerU 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 numbers indicating the relative loading amount of serial diluted QC samples.  For example: 10,20,30,40,50,60,70,80,90,100 |
| Sample\_size | Use two numbers to indicate the number of samples in each group. |
| R2\_threshold | In overall linearity test, set the *R2* threshold to filter out 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 diluted QC data, set intensity threshold to select the valid QC data points. 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 is 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 is less than this value but not 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 is not 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” at the 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 of “Calibration\_datapath”. This file contains the corrected sample MS signal intensities as well as the notations.

**Table 2.** Labels in the “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, PowerU 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 does not 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 measured signal intensities of these features were not corrected. |
| Uncali. | The measured signal intensities were directly used for ratio calculation. |
| Linear | Linear model was used to correct these features. |
| Quadratic | Quadratic model was used to correct these features. |
| Cubic | Cubic model was used to correct these features. |

In the “QC\_number” column, the number of selected serial diluted QC data is shown for each metabolic feature. For example, “8” means there were 8 QC data points meeting the intensity threshold (int\_threshold parameter).

The “p.value” column shows the *p* values after the significance test. The labels in “test” column indicates the name of the significance test performed.