**Quantitative Pipeline for XY-Meta on Two Groups**

In order to conveniently quantify the feature of metabolome in two groups, we wrote a quantitative module to analyze the batch metabolomic data by using R language. You can run XY-Meta easily and get the quantitative table of metabolites features quickly by only a few command lines when using this Script.

This quantitative ,module consist of two parts, namely, metabolites features quantification and metabolites identification. The XCMS package was used to build metabolites features quantification module and the XY-Meta was invocated to identify metabolites. Before you start running this script, you have to install a R language on your machine. Additionally, there should be more than 4 cores and bigger than 8G random access memory in your machine. What’s more, the raw data of LC-MS should be transform to xzXML format.

Now! You can start analyzing your metabolomic data by using this script easily with handle these follows. Reading it carefully, Please.

**Parameter Indication:**

This script can parse command-line parameters to input necessary arguments. There are two parts of parameters were used to implement quantification program and identification program respectively. The parameters of quantification program can be found in document of XCMS and are follows:

p mass\_tolerance(XCMS: ppm, default:10)

i peak\_width\_min(XCMS: peakwidth, default:20)

I peak\_width\_max(XCMS: peakwidth, default:50)

a prefilter\_min(XCMS: prefilter, default:3)

A prefilter\_max(XCMS: prefilter, default:100)

s snthresh(XCMS: snthresh, default:4)

m mzdiff(XCMS: mzdiff, default:0.001)

o noise(XCMS: noise, default:0)

b bw(XCMS: bw, default:5)

r minfrac(XCMS: minfrac, default:0.3)

z mzwid(XCMS: mzwid, default:0.015)

g group\_max(XCMS: max, default:1000)

W analysis\_path(work directory)

T sample1(group1 name, default:sample1)

N sample2(group2 name, default:sample2)

And the identification program uses these arguments:

S search(XY-Meta:)

X xymeta\_path(the bin path of XY-Meta)

M parameter\_path(the parameter path of XY-Meta)

**Example for running pipeline:**

Only quantify the features of metabolome:

The work directory path, bin path of XY-Meta and sample group name are necessary. Firstly, you should pack up all the mzxml files of one group together to a new directory and pack up the other group to a directory. Make sure these directories of two group are set in one main work directory and the two directory name should be recorded. Running the follow command line after your preparation:

Rscript.exe XY-Meta-Quantitative-Pipeline-for-two-groups.R -W work\_path -T group1\_name -N group2\_name

(The program will download requirements packages and install them when you run this script at the first time. Therefor connect to internet is necessary at first time.)

quantify the features of metabolome and identify them:

If you also want to identify the features of metabolome, You also have to input the paths of XY-Meta.exe and the parameter file and a argument celled search switch must be set to TRUE. So, please start to run the pipeline like this:

Rscript.exe XY-Meta-Quantitative-Pipeline-for-two-groups.R -W work\_path -T group1\_name -N group2\_name -S TRUE -X XY-Meta.exe -M parameter.default

**Output data:**

After running the pipeline, you will obtain some result tables such as myAlign.tsv, MyPeakTable.csv, ms1\_info.csv, ms2\_info.csv and mgf file. If you implement the identification program, you can obtain the identification results from XY-Meta and a spectra table called Feature\_identification\_anno.csv. This spectra table is filtered by FDR(<0.5) and the feature name corresponding to the one spectrum is recorded on it. This file called myAlign.tsv is filled with quantification result of you sample data.

If you have any questions, you can contact me at dehualiay@qq.com.