Metabolite Quantification Analysis

Chunyu Ma

Clinical Biomarker Laboratory, Emory University
(mailto:#)chunyu.ma@emory.edu (mailto:chunyu.ma@emory.edu)

February 08, 2019

Introduction

Metabolite Quantification Analysis is used to quantify the concentration of metabolites based on the Reference Standardization Method (Go et al 2015 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4675836/)). The whole analysis procedure contains three steps: 1. Draw distribution plots; 2. Quantify concentration via Reference Standardization Method; 3. Pull the KEGG map from KEGG database. Each step is independent from each other so that you can run any steps independently. Below is the simple functional description of each step:

1. Draw Distribution Plots aims to help the users check intensity distribution of study samples and quality control plasma samples (QSTD) for each target metabolite. The program will automatically generate the boxplot for each target metaboltie. If there are more than one batch in the study, the batchwise box plot will be generated in which each batch will have one boxplot. The program also allows the users to check the groupwise distribution and also labels some target samples in the boxplot using different colors.

2. Quantify Concentration is the main function of Metabolite Quantification Analysis. It aims to calculate the concentraton of each target metabolite using the Reference Standardization Method (Go et al 2015 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4675836/)) which assumes the intensity measure is proportional to the concentration. In order to avoid the batch effect, each batch calculates its own reponsse factor and the final concentration of target metabolite for a sample is based on the batch reponse facotr that it belongs to.

3. Pull the KEGG map from KEGG database aims to label the target metabolite in the KEGG pathway in which this target metabolite participate. This step requires the users to provide a file containing fold change information which can illustrate the higher/lower intensity in certain group for each metabolic feature provided in the metabolic feature table. With this information, the users can specify the color for higher intensity or lower intensity.

Input File Description & Format

To run Metabolite Quantification Analysis on the Shiny app, you need to provide at least three required files: feature table file, class label file and standard metabolite library file. If you need to run the step 3, the fold change file is required as well. Each file is accepted with .txt format or .csv format. Please follow the foramt described below to provide these files:

Feature Table File

First column: m/z.
Second column: time.
Third column: intensity.
… (Followed by intensity for each sample or replicate)

The format of feature table file.

mz	time	NIST_1	QSTD_1a	QSTD_1b	Sample1	Sample2
85.02838	55.63	3638955	1776062	1326539	3001525	384225
85.02842	141.06	885623	456126	779996	103049	282617
85.02842	249.07	689669	901472	1118483	449545	695004
85.04773	80.55	4644279	1511269	1016185	10094300	2956618
85.05230	80.85	382705	109614	68417	0	0
85.07601	101.71	4673276	790546	733285	8245703	3283893

Class Label File

First column: File Name, the name here should be matched with feature table file.
Second column: Sample ID, each file should have a unique id.
Third column: Sample Type, this column has only three choices (study, qstd, nist).
Fourth column: Batch Number(1,2,3…), if the study has only one batch, please label 1 for all rows.
Fifth column: Group(optional), this column is optional, if you want to check the group distribution in step1, please put the group information in this column.

The format of class label file.

FileName	SampleID	Sample type	Batch	Group
NIST_1	nist_1	nist	1	NA
QSTD_1a	q3June2014_1a_1	qstd	1	NA
QSTD_1b	q3June2014_1b_1	qstd	1	NA
Sample1	Plasma_S1_1	study	1	sedentary
Sample2	Plasma_R10_1	study	1	exercised

Standard Metabolite Library File

First column: mz.
Second column: time.
Third column: Chemical Name.
Fourth column: QSTD concentration.
Fifth column: adduct.
Sixth column: KEGG ID.

The format of standard metabolite library file.

mz	time	Chemical_name	Qstd3	adduct	KEGGID
89.10785	73.1	PUTRESCINE	0.086	M+H	C00134
89.10785	88.4	PUTRESCINE	2.99	M+H	C00134
90.05548	59	L-ALANINE	439	M+H	C01401

Fold Change File (Only needed for step 3)

First column: mz.
Second column: time.
Third column: foldchange, this column should be fold change that can

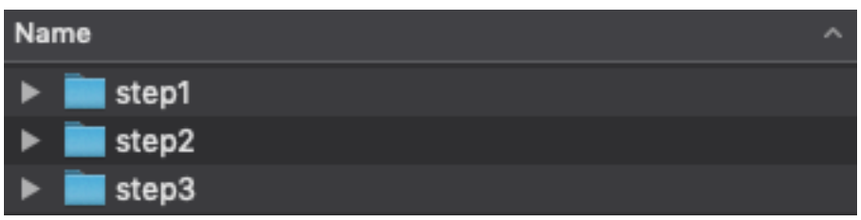
illustrate the higher/lower intensity in certain group. You can select the threshod for this statistic.

The format of fold change file.

mz	time	foldchange
85.02838	55.63	0.217
85.02842	141.06	0.045
85.02842	249.07	0.077

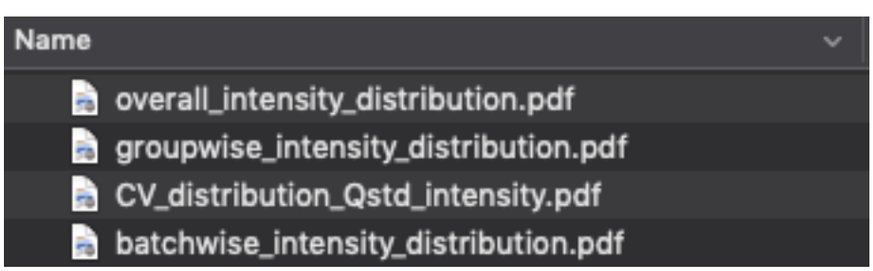
Output File Description

The output files will be saved in different sub folder named by the step number.



Step1 Folder

In this folder, you will see multiple .pdf files based on the options you selected to run the analysis. Bascially you will see the files below.
overall_intensity_distribution.pdf: This file contains the box plot for all samples (exclude the NIST & QSTD samples) for each target metabolite.
groupwise_intensity_distribution.pdf: This file contains the box plot for groupwise samples (exclude the NIST & QSTD samples) for each target metabolite.
CV_distribution_Qstd_intensity.pdf: This file contains CV distribution of QSTD sample intensity.
batchwise_intensity_distribution.pdf: This file contains the box plot for batchwise samples (exclude the NIST & QSTD samples) for each target metabolite.



Step2 Folder

In this folder, you will see multiple .txt files. Bascially you will see the files below.

step2_summary.txt: This file contains some statistical information such as number of batches, number of QSTD files and the averged QSTD CV for all target metabolites.

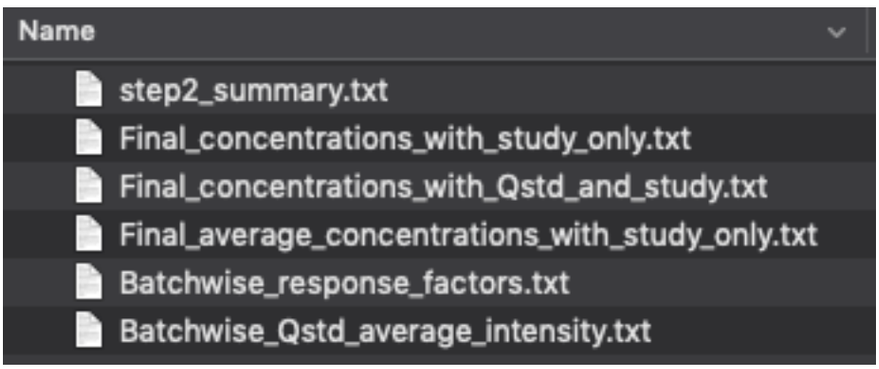
Final_concentrations_with_study_only.txt: This file contains the final concentration results for all study samples (exclude NIST samples and QSTD samples).

Final_concentrations_with_Qstd_and_study.txt: This file contains the final concentration results for all study samples (include NIST samples and QSTD samples).

Final_average_concentrations_with_study_only.txt: This file contains the final concentration results for each batch (averaged concentration for all study samples within each batch).

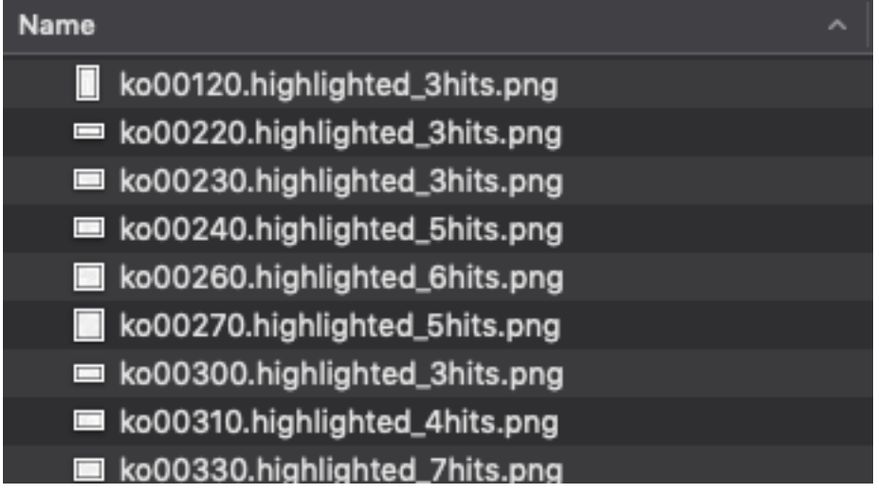
Batchwise_response_factors.txt: This file contains the batchwise response factors used for calculating the concentration.

Batchwise_Qstd_average_intensity.txt: This file contains the averaged intensity of QSTD samples for each batch which is used for impulating the missing value for QSTD samples.



Step3 Folder

In this folder, you will see a file named “keggmapdata.txt” which contains target metabolite information for each KEGG map pulled out from KEGG database. In this file, each row represents one map which has the mapid, the reference mz and time of target metabolites in this map, the observed mz and time of target metabolites in this map, the kegg id of target metabolites in this map and their fold change. You will also see a folder called “KEGGmaps”. Within this folder, you will see different .png file for different KEGG pathways (like the figure showed below).



As you can see, each file is named as mapid+highlighted+#hits. The #hits means the number of target metabolites in reference standard metabolite library you provided can be found to participate in this pathway.

For example, the figure below is from a file called “ko00380.highlighted_9hits.png” which means that there are 9 target metabolties found from your reference standard metabolite library participating in this tryptophan metabolism pathway. The color red and blue are defined by the users to illustrate the higher/lower intensity in certain group.