User Manuals

Description

metaboBD offers powerful pipeline for biomarkers discovery in large-scale metabolomics analysis. This tool contains the multiple algorithms for correcting signals drift and evaluating the data quality. Additionally, it also includes statistical analysis modules for screening the potential biomarkers from the thousands of analyses and generates visual data displaying important features of metabolomics data.

Requirements

Windows XP / Linux

Table of Parameters

Command line

Linux system: perl ./metaboBD <Project_Name|test> <in|sample.list> <in|test.csv> <outDir> <Rscript>

Windows system: metaboBD.exe <Project_Name|test> <in|sample.list> <in|test.csv> <outDir> <Rscript.exe>

Download and installation

Linux system

1) Download the metaboBD package to a desired folder and extract it.
2) Install some required libraries in R:

```
>install.packages("pls", dep=T)

>install.packages("ade4", dep=T)

>install.packages("VennDiagram ", dep=T)

>install.packages("qvalue ", dep=T)
```

Windows system

1) Download the win32-metaboBD package to a desired folder, and extract it.
2) Windows cmd.exe can be used to run the command line.
3) Install some required libraries in R:

```
>install.packages("pls", dep=T)

>install.packages("ade4", dep=T)

>install.packages("VennDiagram ", dep=T)
```

```
>install.packages("qvalue ", dep=T)
```

The tool and source code of metaboBD are available at **http://code.google.com/p/metabobd/**. According to your operating system (Linux and Windows), Perl, R (version 2.14 or higher) from the CRAN (**http://www.rproject.org**) and R packages should be installed, e.g., pls, ade4, VennDiagram, qvalue and others.

Contact

Jin Fu fujin@genomics.cn
Hemi Luan luanhemi@genomics.cn

Dataset format

Data table of interest has to be submitted in .csv format (Comma-separated-values) and with a specific form, as indicated in table 1. The simples' information has to be submitted in .txt format (Tab-separated-values) and with a specific form, as indicated in table 2.

Table 1

ID	MZ	RT	ZSL38	N027	N035
1	55.05381	799.6225	2447038	3877850	4582225
2	61.03925	118.513	11603768	10917639	8710752
3	68.98211	95.4572	177602.6	173870.5	170432.4
4	70.06484	119.556	297907.3	260004.4	292073.2
5	72.08055	124.688	3228687	3063879	3082135
6	72.08069	156.9045	1760999	1669129	2059409
7	80.94726	109.4135	13123006	13358417	13391192
8	82.94434	109.4135	4513127	4373272	4458631
9	83.02109	109.375	11366789	11130252	9443403
10	84.95962	95.7314	13414548	12715611	13056822

In particular:

The first column (ID) indicates the names of each metabolite ID. The second column (MZ) indicates the mass-to-charge ratio of each metabolite. The third column (RT) indicates the retention time of each metabolite. The remnant columns indicate the peak area of each metabolite in all samples and QC samples;

Table 2

Samples	batches	Groups
QC_1	1	NA
N001	1	1
ZSL38	1	2
N027	1	1
N035	1	1

In particular:

The first column indicates the names of each sample. The second column (batches) indicates the blocks in the whole analytical run. The third column (Groups) indicates the clinical phenotype or others (Control: 1; Case: 2). Quality control (QC) samples should be set as NA (Not Available) in the third column. The beginning and the end of samples in each batches must be defined as quality control samples.