MetaMSD Manual

Version: 1.1
Release Date: February 14, 2019
Requirement: R (https://www.r-project.org)
Developers: George Wendt and So Young Ryu
Maintainer: So Young Ryu (soyoungr@unr.edu)

Description: MetaMSD integrates quantitative proteomic results from multiple experiments. MetaMSD.Rscript is a single, self-contained R script that is being developed in an OS X / Linux environment with R version 3.5.0 and it was tested in Window.

Installation

- Download MetaMSD.Rscript in the directory of your choice.
- Change the mode of the script file to enable execute permission by running the following command.

```
chmod a+x MetaMSD.Rscript
```

- Run the following script to display parameter options.

```
MetaMSD.Rscript -h
```

Trouble Shooting: When the script is run for the first time, it will automatically check for the presence of its required packages, and then attempt to install them if they are not already installed. This can take a few minutes. If users have problem with the script running for the first time, open R and install the following packages by type the following in R console:

```
install.packages(c("ROCR","stats","gridExtra","gtable","optparse","qvalue"))
```

How to run MetaMSD

- Create the two directories. If users want to run MetaMSD in a default mode, name the directories as “input” and “output”.
- In the input directory, place multiple txt files of quantitative proteomic results that users want to integrate and name txt files as Dataset1.txt, Dataset2,txt, and so on. The txt files contain protein names, test statistics (or sign of test statistics), and the corresponding p-values. Example files are shown below.

Dataset1.txt contains

```
Protein Sign Pvalue
Protein1 -1.629 0.147
```

```
Protein2 -1.067 0.310
Protein3 NA NA
Protein4 2.376 0.038
```

or

Dataset1.txt contains

```
Protein Sign Pvalue
Protein1 -1 0.147
Protein2 -1 0.310
Protein3 NA NA
Protein4 1 0.038
```

- Type the following command line if users want to run MetaMSD.Rscript in default mode.

```
./MetaMSD.Rscript
```

Other parameter options:

```
MetaMSD.Rscript  -h


MetaMSD (Meta Analysis for Mass Spectrometry Data)

Usage: MetaMSD.Rscript [options]


Options:
        -m CHARACTER, --metaanalysis=CHARACTER
                Specify a meta-analysis test. Either Stouffer or Pearson. [ default = Stouffer ]

        -c NUMBER, --cutoff=NUMBER
                Specify a q-value cut off. [ default = 0.05 ]

        -t NUMBER, --top=NUMBER
                Specify the number of proteins in Top-N differential protein list. [ default = 15 ]

        -i CHARACTER, --input=CHARACTER
                Specify the input folder name. [ default = input ]

        -o CHARACTER, --output=CHARACTER
                Specify the output folder name. [ default = output ]

        -h, --help
                Show this help message and exit
```

MetaMSD Outputs

MetaMSD generates several graphs, tables, and result files in output directory such as the followings:

1) A plot that described numbers of detected differential proteins given q-value thresholds
2) Summary statistics about the numbers of detected proteins and a meta-analysis diagnosis
3) Top-N differential protein list detected by meta-analysis
4) Meta-analysis results (e.g. q-values and sign (directionality of hypothesis) for each protein)

How to generate MetaMSD inputs from other proteomic software

Users can use any proteomic software to quantify peptides/protein, and perform protein significance analysis, and create MetaMSD input files using their own scripts or Microsoft Excel. Here, we provide a few different ways to create MetaMSD input files.

- Obtain outputs from Skyline, MaxQaunt, Progenesis, or Proteome Discoverer. MSstats requires the following outputs, but more details are shown in MSstats manual v3.7.3.

Software	**MSstats Inputs**
Skyline	Cox.Skyline.csv
MaxQuant	proteinGroups.txt, evidence.txt
Progenesis	Progenesis output file in csv, and annotation csv file including condition and biological replicates information.
Proteome Discoverer	PSM sheet saved as csv, and annotation csv file including condition and biological replicates information.

- Perform protein significance analysis using MSstats using the peptide/protein quantification outputs of your choice. Save a result generated from MSstats groupComparison command in csv file. Sample codes are shown below, but see MSstats manual v3.7.3 for more details.

```
quantData <- dataProcess(MaxQuantProcessData)
comparison <- matrix(c(1,-1), nrow=1)
row.names(comparison) <- c("GroupA_vs_GroupB")
resultsComparison <- groupComparison(contrast.matrix=comparison, data=quantData)
write.csv(resultsComparison$ComparisonResult,
 file="[SPECIFY DIRECTORY AND MSstats OUTPUT FILE NAME HERE]")
```

- Open R console and run the following command. (Users need to specify MSstats output file names and MetaMSD input file names.)

```
msstat.data = read.csv(file="[SPECIFY DIRECTORY AND MSstats OUTPUT FILE NAME HERE]", header=TRUE,
sep=",")
msstat.data.format = data.frame(Protein=msstat.data$Protein,
                    Sign=msstat.data$Tvalue,
Pvalue=msstat.data$pvalue)
write.table(msstat.data.format, file="[SPECIFY DIRECTORY AND MetaMSD INPUT FILE NAME HERE]",
col.names=TRUE,row.names=FALSE, quote=FALSE, sep="\t")
```