Analysing metabolomic data using R

Purpose:

Taking data exported from MultiQuant, reorganising it for export to and analysis in Prism, and exporting a PDF summary with graphs generated in R. Options currently available for metabolites that can be quantified through standard curves, metabolites analysed based on peak area, and determining spike amounts for metabolites with standard curves.

Plan to include:

- Corrections for blanks?
- Corrections for isotope and other interference issues

Data prep:

In MultiQuant create a session with the desired .WIFF files and confirm/adjust peak selection for desired transitions

1. Standard Curves

- Make a note of the slope and r-value for the appropriate transitions - not sure how to export these

2. Sample files

- Recommend including standard samples that weren't used for the standard curves and blank controls
- Export results table with all columns and visible (or all) rows
- saves as a text file, open in Excel on same computer and save again

Support files:

1. 'Expt'_Info.R

- Contains experiment specific info for filenames, and various options including:
  - X-var: variable to compare against
  - reorder list: to specify order of the x axis labels

2. Sample Info

- Includes info for all the samples run
- Make sure to include:
  - Sample name: name used in mass-spec file
  - SampleName2: A group name eg: Aged_gastroc
  - Category: Sample, Standard, Blank, Pool, Recov_cont

3. Standard Curves

- Slope and r value as determined from MultiQuant
- First column lists the source mix for the data
- Include column with Metabolite name as well as the Component Name (transition)

4. Metabolite list

- List of all the Component Names with columns used for graph selection
- Will need to be adjusted for each experiment
  - used_met: metabolite with a standard curve
  - used_met_level: top, high, med, low, vlow
  - keep_spike: metabolite to keep for a selected spike graph
  - other_met: metabolite to analyse with peak area
  - other_met_level: top, high, med, low, vlow

Data organisation

Experiment folder

- data folder: contains sample data, support files, and generated compiled data as .CSV format
- results folder: output folder for generated results

R scripts

Data prep

Metabolomics_01.R

- loads libraries and additional functions

Met_DataPrep.R

- Prepares data for metabolites with standard curves
- File info is specified in 'expt'_info.R
- Check that the correct source mixes are being selected for the StdData
- exports data with a timestamp

OtherMet_DataPrep.R

- Prepares data for metabolites without standard curves
- File info is specified in 'expt'_info.R
- saves over previous exports

SpikeRecovery.R

- Calculation of recovered spike amounts
- Most file info is specified in 'expt'_info.R
- specify the timestamp for the compiled data file from Met_DataPrep.R
- Output: PDF of graphs (see below) and CSV of recovery proportion for each metabolite
- exports data with a timestamp

Met_DataPrep_2.R

- Modifies output from Met_DataPrep.R based on SpikeRecovery.R results
  - recovery = Y/N for when recovery correction data is availabile
- exports data with a timestamp

Deconv_dataPrep.R

- Processes data produced from Matlab modrunner_qtrap_deconV_unlabelled_interpl.m
- output can be used in MetAnalysis.R
- Need to add "met-comp" to grouping variables

Analysis

Met_Analysis.R

- Analysis of data for metabolites with standard curves
- Most file info is specified in 'expt'_info.R
- recovery = Y/N for when recovery correction data is availabile
- takes output from Met_DataPrep.R or Met_DataPrep_2.R
- specify the timestamp for the compiled data file
- Run through step-by-step first up to and including GraphCheck
  - Add a reorder_list to 'expt'_info.R and load if needed (used to make sure x-axis is in the correct order)
  - Classify metabolites to the different levels

- Reload Metabolite list and selection levels with each change
- Continue to prepare final graphs
  - Adjust binwidth for each graph before proceeding to PDF
  - Add any comments to front page before exporting PDF
  - Graphs do not work for deconvolved data

Met_Analysis_other.R

- Analysis of data for metabolites by peak area
- Most file info is specified in 'expt'_info.R
- Run through step-by-step first up to and including GraphCheck
  - Add a reorder_list to 'expt'_info.R and load if needed (used to make sure x-axis is in the correct order)
  - Classify metabolites to the different levels
- Reload Metabolite list and selection levels with each change
- Continue to prepare final graphs
  - Adjust binwidth for each graph before proceeding to PDF
  - Add any comments to front page before exporting PDF

SpikeRecovery.R

- Calculation of recovered spike amounts
- Most file info is specified in 'expt'_info.R
- specify the timestamp for the compiled data file from Met_DataPrep.R
- Output: PDF of graphs and CSV of recovery proportion for each metabolite
- Add a reorder_list if there are multiple spike levels
- Run through step-by-step first up to and including GraphCheck
  - remember additional graph script compared to Met_Analysis
  - Check which metabolites to use for selected spike graph
- Reload Metabolite list and selection levels with each change
- Continue to prepare final graphs
  - Adjust binwidth for each graph before proceeding to PDF
  - Add any comments to front page before exporting PDF
  - Graphs do not work for deconvolved data

Graph scripts

General graph setup derived from variables in the global environment

FilteredMetGraph.R

- Creates bar only, dotplot only or combination graphs depending on provided data
  - data grouped by SampleName2 -> bar only
  - data separated by individual samples -> dotplot only
  - both individual and grouped data -> combination graph
  - If error bars present they represent standard error as derived in analysis scripts
- Will filter data based on provided metabolite level (top, high, etc.)
- arguments: .ind_data, .gr_data, .metlevel (default = usedMet), ... passed to geom_dotplot for binwidth

FilteredMetGraph_Spike.R

- Creates bar graph for recovered spike amounts for individual samples
- Will create a graph for each spike level, if more than one is present, using the reorder_list
- arguments: ind_data, metlevel (default = usedMet)

Typical run

1. Metabolomics_01.R
2. OtherMet_DataPrep.R
3. Met_Analysis_other.R: relative changes
4. Met_DataPrep.R
5. Met_Analysis.R: curve corrected changes, check standard controls
6. SpikeRecovery.R: check recovery controls
7. Met_DataPrep_2.R
8. Met_Analysis.R: curve and recovery corrected changes