MRMandSWATHgraphics

Lisa Chung
Oct. 18, 2012

0. Before starting..

Description is written with this font.
Something one should enter in R session is written in `this` font.
Something printed out on R Console is written in `this` font.

1. getMRMplot( )

getMRMplot( ) is an R function to make exploratory graphics for MRM data.
After starting R, one should first source the code by typing

```
> source("getMRMplots_113012.R")
```

or click "File", select "Source File" , then choose the R script file.

The function requires following four R libraries, to be pre-installed:

- preprocessCore (for quantile normalization)
- lattice (for bar-chart of detected peaks and their signal-to-noise label)
- gplots (for heatmap of a correlation matrix)
- limma (for Venn diagram and moderated t-test, if applicable)

One can use the function by typing
> getMRMplot()

Executing the function will send you the message:
Hit ENTER to choose the (MULTIQUANT) DATA file
Please hit ENTER and select the data file. Second input one should enter is the format of transition naming convention: 1 for dot-separated format and 2 for underscore-separated.
Multiquant output data to use:
/Users/lisachung/Desktop/Proteomics/Sakaue_PSD_xMRM4_three_datasets_020212/PSD_Cortex_only_SF2_020112_export.txt
Select Transition naming option: 1 for RBC (dot), 2 for PSD (undescore):
2
One can also specify the header of output files. (The function automatically creates a directory with the same name as the input data file, with '_getRplot_Result' appended. Then, the function makes a number of figures and tables with the specified header.)
Type result file name header (or just hit enter):
psd
Output files are saved under

/Users/lisachung/Desktop/Proteomics/Sakaue_PSD_xMRM4_three_datasets_020212/PSD_Cortex_only_SF2_020112_export
Hit ENTER to RUN

Output files

-barChart.pdf
Signal-to-noise ratio is obtained for each transition of each sample from Multiquant software. It is transformed into an ordinal category:

5 if S/N > 10 (great)
4 5 < S/N < 10 (above limit of quantification)
3 3 < S/N < 5 (above limit of detection)
2 S/N < 3 (noise level)
1 missing peak.

For each peptide, a peak frequency is evaluated as a number of peaks observed (across transitions) and a categorized signal-to-noise score is computed as an average of transition-level categorized values. These two values are averaged over technical runs.

This figure shows the average peak frequency on y-axis, for each peptide using a bar chart. A color of each bar indicates (grouped) peptide-level categorized signal-to-noise ratio (SN). If the color is **SN = 5**, **4 <= SN < 5**, **3 <= SN < 4**, **SN < 3**. (***)

Peptides from the same protein are grouped together.

- sampleSpecific_BarChart.pdf
Similar plot with BarChart.pdf, but one figure for one sample, instead of averaging over technical runs

-OtherGraphics-RTranking.pdf
When `group = TRUE`, this figure gives the distribution of (observed) retention time ranking among *k* technical runs within the same biological sample. The retention time ranking is given as a sequence of *k* numbers separated by comma. The first number indicates the ranking (1=smallest, 2=second smallest, …, k = k_{th} smallest) of the first technical run. The second number tells the ranking if the second technical run, and so on. For example, "2,1,3" shows that the ranking of the first run is 2 (=the first number), that of the second run is 1, and that of the third run is 3, *i.e.* RT2 < RT1 < RT3.

-OtherGraphics.pdf
Page 1: At each transition, across all (biological and technical) samples, one can compute the transition-specific median of observed retention times. In this figure, the difference of the retention time from the median is plotted on y-axis, against median retention time. Retention time from each sample is labeled with different color.

Page 2: Since some transitions have large variation among their retention time (while most of the transitions have small variation), this figure shows zoom-in view of page 1 around 0 on y-axis, so that the figure captures 90% of the transitions.

Page 3: It shows the distribution of signal-to-noise ratio in log 10 scale. The distribution is displayed using a box-plot. Each box shows the 25^{th} (lower line), 50^{th} (middle, thick line), and 75^{th} percentile (upper line). 'Whiskers' extend to the extreme data points after removing outliers. Outliers are denoted with dots.

Page 4 and **Page 5**: The distribution of peak intensities, Area, in log 2 scale before and after a quantile normalization, respectively. Quantile normalization is performed using `normalize.quantiles()` in R package preprocessCore.

Page 6 and **Page 7**: Hierarchical clustering before and after quantile normalization. Initially, each sample is assigned to its own cluster. At each step, two clusters with the shortest (average) distance is joined together.

Page 8: Correlation between all possible pairs is labeled with grey-scaled colors (light: low correlation, dark: high correlation). For each pair, a transition with at least one (or both) missing values are ignored.

-peptide-level-summary.pdf

For each peptide, the number of transitions having 4 or higher categorized signal-to-noise score (ordinal signal-to-noise >= 5 (***)) is counted and summarized across technical replicates using a Venn diagram.

sampleScatterPlotAndVennDiagram.pdf

This graphic is generated to inspect the technical reproducibility. Each page shows reproducibility analysis result for each biological sample. On each page, scatter plots of log-intensity are provided for all pairs of technical runs. Venn Diagram is used at the following three levels.

Transition-level: The number of transitions having signal-to-noise ratio greater than 5 are counted and summarized into a Venn Diagram.

Peptide-level: For each peptide, the number of transitions having 4 or higher categorized signal-to-noise score (ordinal signal-to-noise >= 5 (***)) is counted and summarized across technical replicates using a Venn diagram.

Protein-level: this function counts the number of peptides having peptide-level categorized scores greater (see –barChart.pdf, underlined, for detail) than 4 (***). If a protein has two or more such peptides, it is counted into a Venn diagram.

- area.protein.pdf, rt.protein.pdf, and sn.protein.pdf shows the corresponding values across all the samples. Each line on the graph indicates each transition within the peptide.

-summary.txt
Signal-to-noise ratio and peak area, summarized for each peptide across technical replicates.

2. getSWATHplot()

This function provides similar quality assessment bar chart. Height of the bar indicates categorized signal-to-noise ratio and the color label shows its false discovery rate level.