MSstats.daily updates

Meena Choi, `choi67@purdue.edu`

October 27, 2014

v 2.3.5 updates : October 27, 2014

1. `groupComparison` function

    (a) Fixed the bug for `missing.action`.

2. `groupComparisonPlots` function

    (a) Added `logBase.pvalue` : for volcano plot or heatmap, (-) logarithm transformation of adjusted p-value with base 2 or 10. -log10(adjusted p-value) is the default.

v 2.3.2 updates : June 15, 2014

1. `dataProcess` function

    (a) Added `FeatureSelection` : one method for feature selection is available. `FeatureSelection=TRUE` selects the most informative features which agree with the pattern of the average features (use endogenous intensities) across the runs. `FeatureSelection=FALSE` uses all features that the data set has. This option will be extended with more methods for feature selection.

    (b) Updated two steps normalization : two normalization options can be used such as `normalization=c("equalMedian`. For label-based experiment with `normalization=c("equalMedians","globalStandards")`, the data will be normalized for reference peptides, which medians of intensities across runs become equal. Then endogenous peptides will be normalized with global standard protein.

v 2.1.6 updates : March 25, 2014

1. `dataProcess` function

    (a) Added a more explicit warnings for handling missing peaks. MSstats input requires NAs for missing peaks. New warning messages will say whether there is no information (no rows) for missing peaks or duplicate rows. For label-based experiments, warning messages will say whether incomplete rows come from reference features or endogenous features, also whether there are some features that are completely missing in all the runs, in which case users need to check the processing in spectral processing tools.

    (b) Added new option `fillIncompleteRows`. If `fillIncompleteRows=FALSE` (default), the error message and list of incomplete features will be reported and the data processing will be stopped. If `fillIncompleteRows=TRUE`, list of incomplete features will be reported, incomplete rows for missing peaks will be added with intensity=NA and then input will be ready to further analysis. This will be helpful especially for DDA data which is generated by spectral processing tools that doesn't produce NAs for missing peaks.

(c) Fixed the bug for assigning the option of `nameStandards` when `normalization="globalStandards"` (Thanks, Tobias)

(d) `PeptideModifiedSequence` column is also accepted instead of `PeptideSequence` column in MSstats required input. If there are some peptides who are present in a modified and unmodified form, `PeptideModifiedSequence` column from Skyline is suitable for analysis. MSstats report from MSstats external tool in Skyline includes `PeptideModifiedSequence` column instead of `PeptideSequence` column. (Thanks, Tobias)

2. Fixed minor bugs in `designSampleSize` and `quantification`

v 2.1.5 updates : February 12, 2014

1. `dataProcess` function

   (a) Changed the name of the option for constant normalization from `constant` to `equalizeMedians`, to better reflect the fact that this normalization equalizes medians. Now `equalizeMedians` is the default. See `?dataProcess`.

   (b) Fixed the normalization of targeted experiments where the set of transitions is split across multiple acquisitions (equivalently, "methods" in the vocabulary of Skyline). Now these experiments are normalized per file name. E.g., in targeted experiments with labeled reference peptides, constant normalization equalizes the median intensities of reference transitions across all the input files.

   (c) Added a more explicit warning for handling missing values. The warning clarifies that the input file should include rows for all the features in all the runs. Intensities of missing features should be indicated with "NA".

2. `groupComparison` function

   (a) Changed the name of the option for (un)equal error variance among features from `featureVar` to `equalFeatureVar`. See `?groupComparison`.

3. `designSampleSize` function

   (a) Fixed bugs for expanded scope of replication.

   (b) New option `equalFeatureVar` mirrors the same option in `groupComparison`. See `?designSampleSize`.

4. `quantification` function

   (a) Fixed various minor bugs.

   (b) Show progress of the analysis in the R console.

   (c) New options, `scopeOfBioReplication`, `scopeOfTechReplication`, `interference`, `equalFeatureVar` mirror the same options in `groupComparison`. See `?quantification`.

5. `dataProcessPlots`, `groupComparisonPlots`, `modelBasedQCPlots` functions

   (a) Added plotting options such as font size etc. See `?dataProcessPlots`, `?groupComparisonPlots`, `?modelBasedQCPlots`.

   (b) Added a new option in the heatmap representation of the results: `numProtein` defines the number of proteins represented in each heat map. If the number of proteins exceeds the value specified in this option, multiple heat maps are drawn.

   (c) Added a new option in the heatmap representation of the results: `clustering` determines how the proteins and the comparisons are ordered.

6. Analysis workflow

(a) Added an automatically generated log file "msstats.log", that saves session information, selected options and analysis progress. The file is used to record the details of the analysis, and for debugging.

(b) New naming of the output files. Instead of replacing the files, add suffix with a consecutive number if other files with the same names are already present in the working directory.