DAPAR
(Differential Analysis for Proteomics Abundance with R)

Samuel Wieczorek, Florence Combes, Thomas Burger

29 July 2015

DAPAR (Differential Analysis for Proteomics Analysis with R) is an R package for the analysis mass spectrometry-based discovery proteomics quantitative dataset. It is built on pre-existing Bioconductor packages devoted to proteomics and uses the MSnSet data structure from MSnbase. The package contains functions to import data from CSV files (such as those from Maxquant), and to export data to Excel files.

DAPAR is made to guide the practitioner through the following classical steps of a protein-level data analysis:

- **Descriptive statistics**: Exploration and visualization of the dataset, which includes the following functionalities: Missing Values exploration, heatmap and correlation matrices, boxplots, expectation and variance distribution.

- **Filtering**: DAPAR allows to filter proteins or peptides according to their number of missing values and the flags indicating reverse sequences or contaminants.

- **Cross replicate normalisation**, with the following methods: (i) global rescaling (quantiles method, proportion method), (ii) median or mean centering (overall or within conditions), (iii) mean centering and scaling (overall or within conditions).

- **Missing values imputation**, with the following methods: (i) for random occurences: k-nearest-neighbors, Bayesian Principal Component Analysis and Maximum Likelihood Estimation; (ii) for left censored missing values: Quantile Regression for Imputation of Left Censored data.

- **Aggregation of peptides intensities into proteins' abundance**, with the following methods: (i) sum/mean of the intensities of all/specific peptides, (ii) sum of the intensities of the n most abundant specific peptides (i.e. median intensity across the replicates).

- **Differential analysis**, according to a Welch t-test or a Limma moderated t-test.