

MsReport

Introduction

MsReport is a python library that aims to allow simple and standardized post processing of quantitative proteomics data from bottom up, mass spectrometry experiments. Currently working with label free protein quantification reports from MaxQuant and FragPipe is supported. Other data analysis pipelines can be added by writing a software specific reader module.

MsReport is primarily developed as a tool for the Mass Spectrometry Facility at the Max Perutz Labs (University of Vienna) to allow the generation of Quantitative Protein and PTM reports, and to facilitate project specific data analysis tasks.

Release

Development is currently in early alpha and the interface is not yet stable.

Scope

The "Reader" module contains software specific reader classes that provide access to the outputs of the respective software. The reader has to be aware of the file structure and naming conventions of the respective software and allows importing protein and ion tables and standardizing column names and data formats.

The "Qtable" class from allows storing and accessing quantitative data from a specific abstraction level, such as ions or proteins, and an experimental design table, which describes samples and experiments. The quantitative data is in a wide format, meaning that the quantification from each sample is stored in a separate column. The Qtable allows convenient handling and access to the quantitative data by utilizing the information from the experimental design, and represents the data structure used by analysis and plotting methods.

The "Analyze" module provides a high level interface for post-processing of data present in the Qtable class, such as filtering valid values, normalization between

samples, imputation of missing values, and statistical testing with the R package LIMMA.

The "Plot" module allows simple generation of plots for quality control and data analysis directly from a Qtable instance.

Using methods from the "Export" module allows converting and exporting data from a Qtable into the Amica input format, and generating contaminant tables for the inspection of potential contaminants.

Additional scripts - Generate a formatted excel protein report (Uses the XlsxReport library)

Install

For Windows users without Python we recommend installing the free Anaconda Python package provided by Continuum Analytics, which already contains a large number of popular Python packages for data science. Or get Python from the Python homepage.

To install MsReport, activate the conda environment you want to use, navigate to the folder containing the MsReport files and enter the following (don't forget to write the dot after install):

```
pip install .
```

To uninstall the MsReport library type:

```
pip uninstall msreport
```

MsReport provides an interface to the R package LIMMA for differential, which requires a local installation of R (R version 3.4+) and to system environment variable "R_HOME" to be set to the R home directory. The R home directory can be obtained from within R by using the command below, and might look similar to "C:\Program Files\R\R-4.2.1" on windows.

```
normalizePath(R.home("home"))
```

In order to use the "msreport_scripts/excel_report" module the XlsxReport library is required.