

MsReport

Introduction

MsReport is a python library that allows 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 fully supported. Other data analysis pipelines can be added by writing a software specific reader function.

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, allows importing protein and ion tables, and standardizing column names and data formats.

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

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 generation of quality control and data analysis plots 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 - The "excel_report" module enables the generation of a formatted excel protein report by using the XlsxReport library. - The "benchmark" module contains functions to generate benchmark plots from multiple qtable instances, and can be used for method or software comparison.

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. MsReport requires Python version 3.9 or higher.

To install MsReport, open a terminal and activate the conda environment you want to use for installation (note that the default environment is called "base").

```
conda activate base
```

Navigate to the folder containing the MsReport installation package and enter the following command (don't forget to add 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 expression analysis, which requires a local installation of R (R version 3.4 or higher) and the system environment variable "R_HOME" to be set to the R home directory. Note that it might be necessary to restart the computer after adding the "R_HOME" variable. The R home directory can also be found 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 (version 0.0.4) must be installed and the Appdata directory containing configuration files must be setup by running the command:

```
xlsx_report_setup
```