The project folder is structured as follows (new features in red):

**1 – Raw files**

Raw data acquired by the mass spectrometer. We searched this data with either MaxQuant (files acquired in Data Dependent Acquisition mode) or increasingly often with DIA-NN (Data Independent Acquisition mode). Both software identify peptides by matching acquired spectra to theoretically expected peptides.

*Please keep these files for future reference/re-analysis/publication.*

**2 – Analysis\_*YYYY-MM-DD\_v1***

***This folder contains 2 subfolders: the results from the search software (MaxQuant or DIA-NN), and of Post-processing (complementary information, tables, and plots).***

***(If several analyses were delivered, they have version numbers and incremental list numbers, e.g.* 3 – Analysis\_*YYYY-MM-DD\_v2)***

* **a – MaxQuant/DIA-NN search**

All search results file. These are important for submitting search results to the PRIDE repository for publication.

* For MaxQuant, we delete most temporary files, keeping only the “andromeda” and “txt” subfolders in the “combined” folder as well as the MaxQuant search parameters file. The “txt” folder contains MaxQuant’s output tables (despite the extension, tables are formatted as tab-separated values = .tsv). Of particular interest is *evidence.txt*, which is a long table of all individual peptidoform (= modified peptide sequence) observations, including direct peptide-to-spectrum matches (PSMs) and indirect match-between runs identifications. This is the main file which is used for reprocessing the MaxQuant output.
* For DIA-NN, we keep all results .tsv file. Table names vary depending on DIA-NN’s parameters. The equivalent of the *evidence.txt* table is usually the .tsv file with the shortest name in the folder. When in doubt, its precise name can be found by opening the log file in the folder and looking up the name of the file after the “*--out*” (= output) argument in the latest “*diann.exe*” call.
* **b – Post-processing**

Detailed results of post-processing the *evidence.txt* (MaxQuant) or *REPORTNAME.tsv* (DIA-NN) file found in subfolder “a – *Search…”*. Contains several subfolders, and up to 3 protein groups-level Excel results tables. Each table’s first tab describes the types of columns you can find in the others. (There is also in subfolder *…/Tables/* a larger Excel table called *proteinGroups - Full.xlsx* with all of the rows and columns; these other 3 tables are thematic extracts from the information in that larger, single tab one.)

* *proteinGroups - Description.xlsx*: Contains all information relative to what the protein groups are (accessions, annotations etc…)
* *proteinGroups - Reg. analysis.xlsx*: Contains ones tab per statistical test. Row are re-ordered from high to low ratios for samples relevant for that test.
* *proteinGroups - GO terms.xlsx*: If applicable, one tab per GO term of interest showing protein groups annotated directly with it or with one of its offspring terms.

“Description” and “Full” Excel tables are also created for all observed “modified peptides” (aka peptidoforms).

In addition, if a regulation analysis was performed for a specific class of modified peptides (e.g. “Phospho (STY)”), then “Description”, “Reg. analysis” and “Full” Excel tables will be created for peptides with this modification.

Current subfolders include:

* + Summary plots: Plots which we create to monitor the quality of the MS runs.
  + Workflow control: Plots created during data processing to check the quality of the data, its distribution, the behavior of normalizations, etc…
  + Protein plots: If applicable, these include, for proteins of interest:
    - * sequence coverage plots,
      * plots of correlation between samples,
      * ratio plots, where the sequence ratio of intensity between sample pairs is plotted for each observed part (peptide) of the protein sequence.
  + Protein Groups sorted by and Profile plots for are different ways of looking at protein group relative and absolute abundance, coverage and spectral counts per sample or across all samples, respectively.
  + GO enrichment analysis: GO terms enrichment analyses (using *topGO*). The following analyses are made:
    - * Comparison to …: Analysis is performed comparing the protein groups with high ratios in the sample versus the control. The P-value is calculated using Fisher’s exact test. On the bubble plots generated, the X-axis corresponds to the average logFC for that particular term for ALL proteins, and thus represents the general trend for that particular term in the relevant sample vs control comparison.
      * Sample vs total proteome: This enrichment analysis is a comparison of the observed dataset versus the theoretical proteome of the parent organism. This analysis is useful to check for expected sample biases (e.g. tissue-specificity).
    - Clustering: Heatmaps of protein-groups level quantitative data with hierarchical clustering.
    - Venn diagrams: Two types of Venn diagrams can be drawn, depending on project:
      * Diagrams of the overlap between protein groups identified in different samples.
      * Diagrams of the overlap between protein groups defined as regulated (up-, down-, or both) in different non-control samples.