**Guide to the results folder**

***(dataset without replicates)***

Each output project folder is structured as three types of subfolders:

**1\_MS\_files**

Raw data acquired by the mass spectrometer.

*Keep these files preciously for future reference, re-analysis or publication!* It is essential to make the raw data available when publishing your results. The need to make supporting data publicly available is increasingly being recognized and enforced by scientific journals. See <https://www.proteomexchange.org/> for the most common repositories for proteomics data.

**2\_“SEARCH-ENGINE-NAME”\_search**

The raw data was searched in one of several search software (currently, DiaNN [[1]](#endnote-1), FragPipe [[2]](#endnote-2) and MaxQuant [[3]](#endnote-3) are supported. It is important to keep these files, not just the raw files, as some repositories also require search results to be uploaded. *Please keep these files for future reference/re-analysis/publication!*

For post-processing the output of search engines (see next section), we use the most basal level of output, i.e. the identifications tables, also known as “PSMs” (peptide-spectrum matches):

* For DiaNN, these are contained in a report.tsv or report.parquet file. We also need the report.log.txt file to read search parameters.
* For FragPipe, we need the output folder: we read the workflow and samples manifest files which FragPipe saves into the output folder, and use the information to parse the psm.tsv output file(s).
* For MaxQuant, we use the evidence.txt file (which contains PSMs per se as well as indirect, match-between-runs based identifications) and also parse parameters from the mqpar.xml file.

**3(.#?)\_Post\_processing\_YYYY\_MM\_DD**

Detailed results of post-processing the *PSMs table*. *If several analyses were delivered, there are several folders with incremental list numbers and analysis date, e.g.:*

*…/3.1\_Post\_processing\_2025\_03\_25*

***In each such folder, the main Excel report table is located in sub-folder …/Tables!***

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 correct behaviour of normalizations, etc…
* Tables: All result tables are saved here, including the main Excel report.
* 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.
* Heatmaps and Coverage: Sequence coverage plots and heatmaps for proteins of interest. A*lso includes input files for visualizing observed peptides overlapped over a 3D structure.*
* Sorting plots and Profile plots are different ways of looking at protein group abundance, coverage or spectral counts per sample or across all samples, respectively.
* Pearson correlation map: Comparison of all samples using Pearson correlation.
* Dimensionality red. plots: dimensionality reduction plots such as PCA, t-SNE or U-MAP.
* Clustering: Heatmaps with hierarchical clustering.
* 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).
* Venn diagrams: Two types of Venn diagrams can be drawn, depending on project:
  + overlap between protein groups identified in different samples.
  + overlap between protein groups defined as regulated (up-, down-, or both) in different non-control samples.

1. Full references: <https://github.com/vdemichev/DiaNN?tab=readme-ov-file#key-publications> [↑](#endnote-ref-1)
2. Full references: <https://github.com/Nesvilab/FragPipe> [↑](#endnote-ref-2)
3. Full references: <https://www.maxquant.org/Publications/> [↑](#endnote-ref-3)