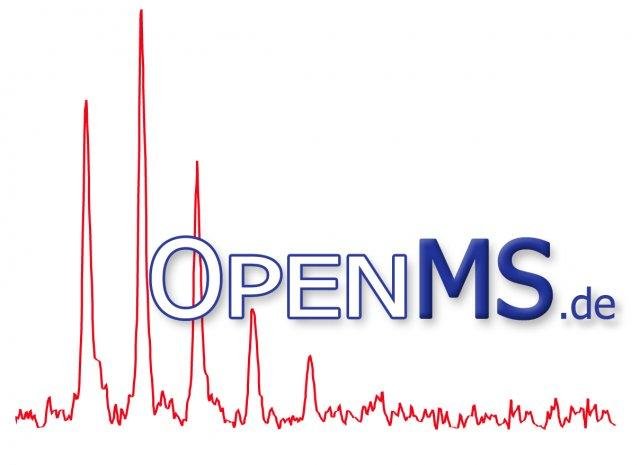
OpenMS Nodes for Proteome Discoverer 2.x



User Manual

# Introduction

The OpenMS nodes package extends Proteome Discoverer’s functionality by two additional workflows based on algorithms implemented in the OpenMS framework: LFQProfiler for label-free peptide and protein quantification and RNPxl for protein-RNA cross linking data analysis.

# Installation

Installing the OpenMS plugin for Proteome Discoverer is easy: just download the binary installer for your platform (PD 2.0 or 2.1) from [www.openms.de/pd](http://www.openms.de/pd) and run it. Note that Proteome Discoverer should be closed before running the installer.

In rare cases, the automatic binary installer packages do not work, because the Proteome Discoverer registry key cannot be found. Should this be the case, you can still install our nodes by downloading the manual installation package from [www.openms.de/pd](http://www.openms.de/pd) and following the instructions therein.

After successful installation, you should find four additional nodes in your Proteome Discoverer workflow editor: the two processing nodes LFQProfiler FF and RNPxl, and the two consensus nodes LFQProfiler and RNPxl Consensus.

# Usage

In the following chapters, we will explain usage and describe the features and settings of the RNPxl and LFQProfiler workflows.

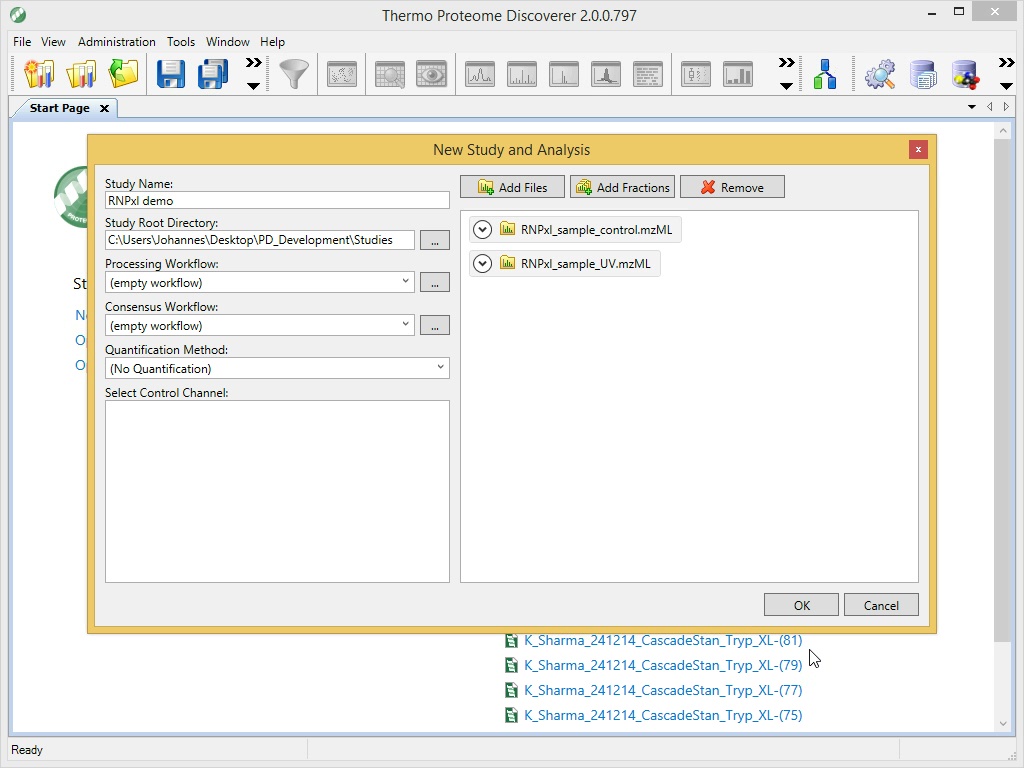
# RNPxl

The functionality of the underlying OpenMS-based workflow together with a description of the experimental protocol used for generating suitable protein RNA cross-linking data is described in detail by Kramer et al.: *Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins*. Nat Methods. 2014 Oct;11(10):1064-70

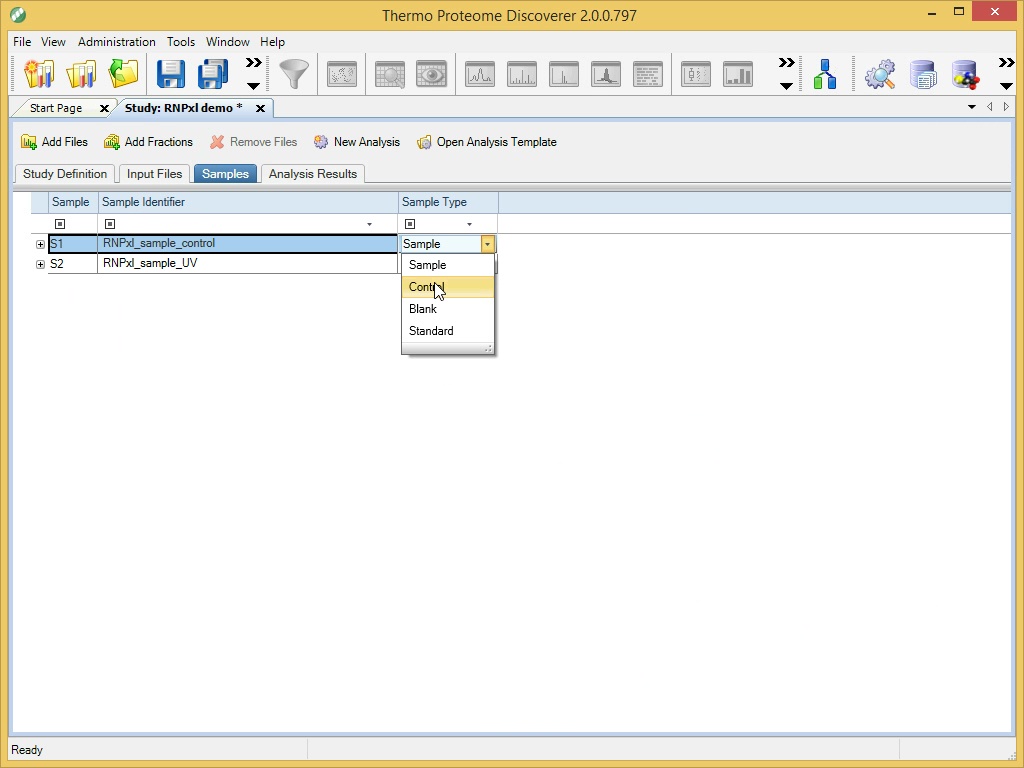
The major advantages of using the Proteome Discoverer version of this workflow are easier usability and the integration with a custom designed spectrum visualization tool that is able to visualize not only the annotated peptide fragments but also the cross-linked nucleotide fragment ions which allows for a greatly facilitated manual inspection and validation of the results.

## Setting up the study

Set up a new study as usual. If you have both a UV cross-linked run and a control run without cross-links, you can load both these files here. Otherwise, just load the cross-link file. The workflow will later determine whether a control is present or not and automatically omit the steps involving a control file if it is missing.



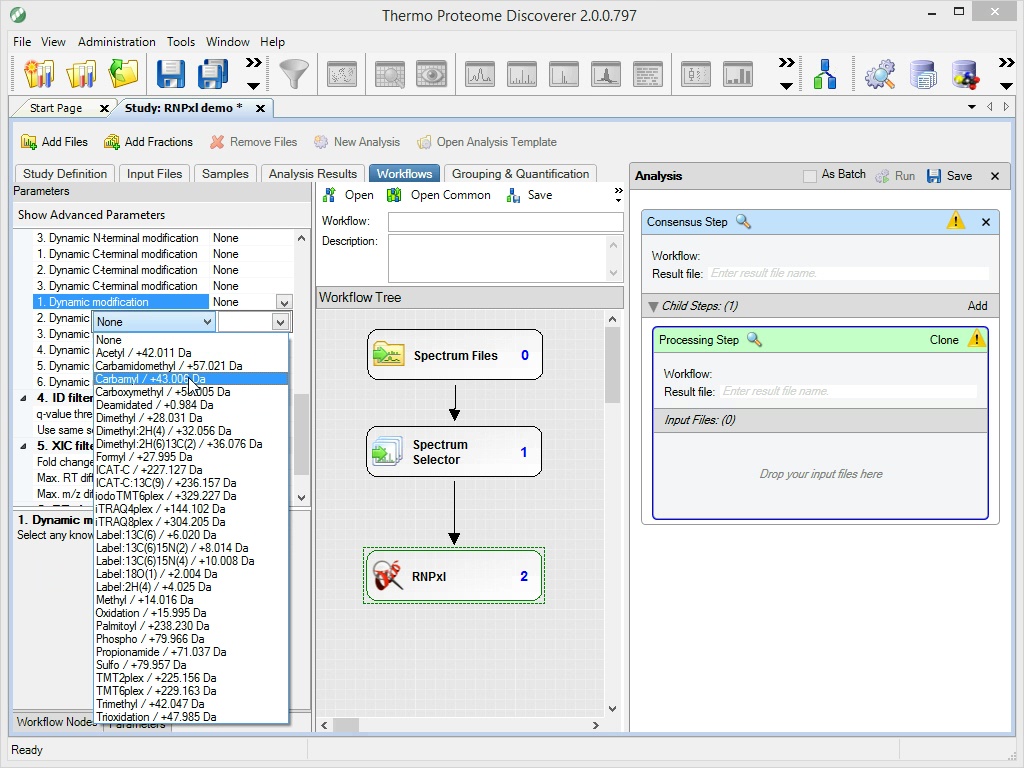
If you do have a control run, it is important to mark it as Control (not Sample) in the Samples tab:



## Setting up the workflows

The easiest way of getting started is to download the RNPxl processing and consensus workflow from [www.openms.de/pd](http://www.openms.de/pd). If you prefer to click yourself, please set up the processing workflow so that it looks like the one in the following Figure.

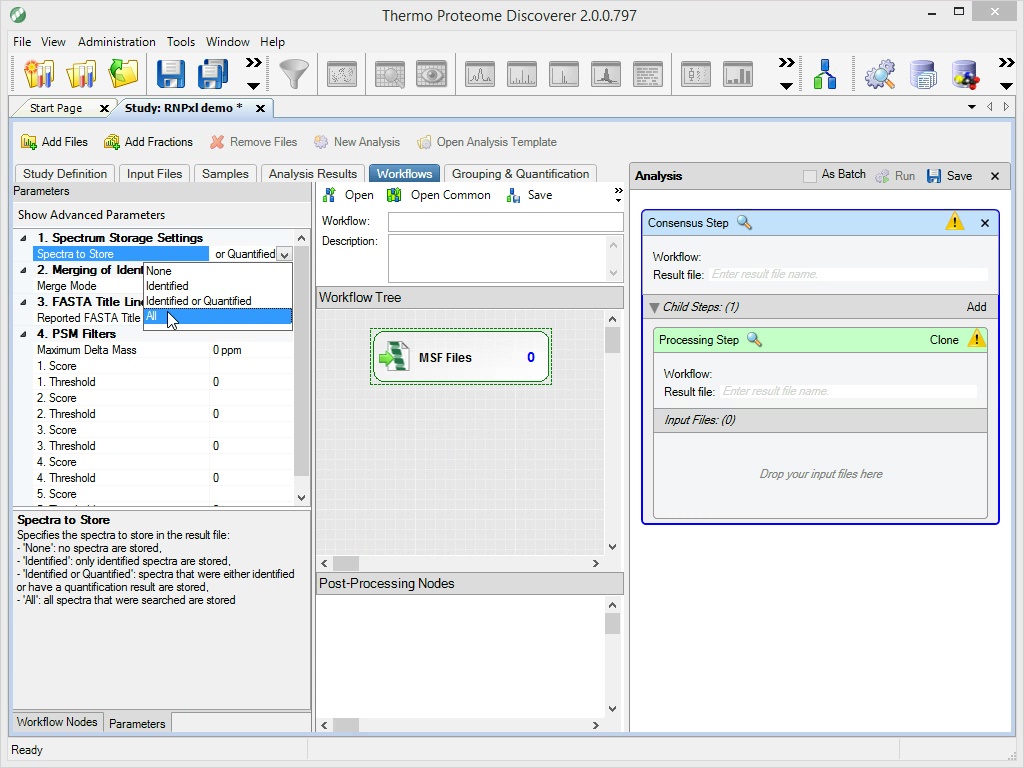
Make sure to set “MS Order” to “Any” in the Spectrum Selector node (otherwise, MS1 spectra are discarded and the workflow cannot work without them). Adapt the parameters of the RNPxl processing node as needed.



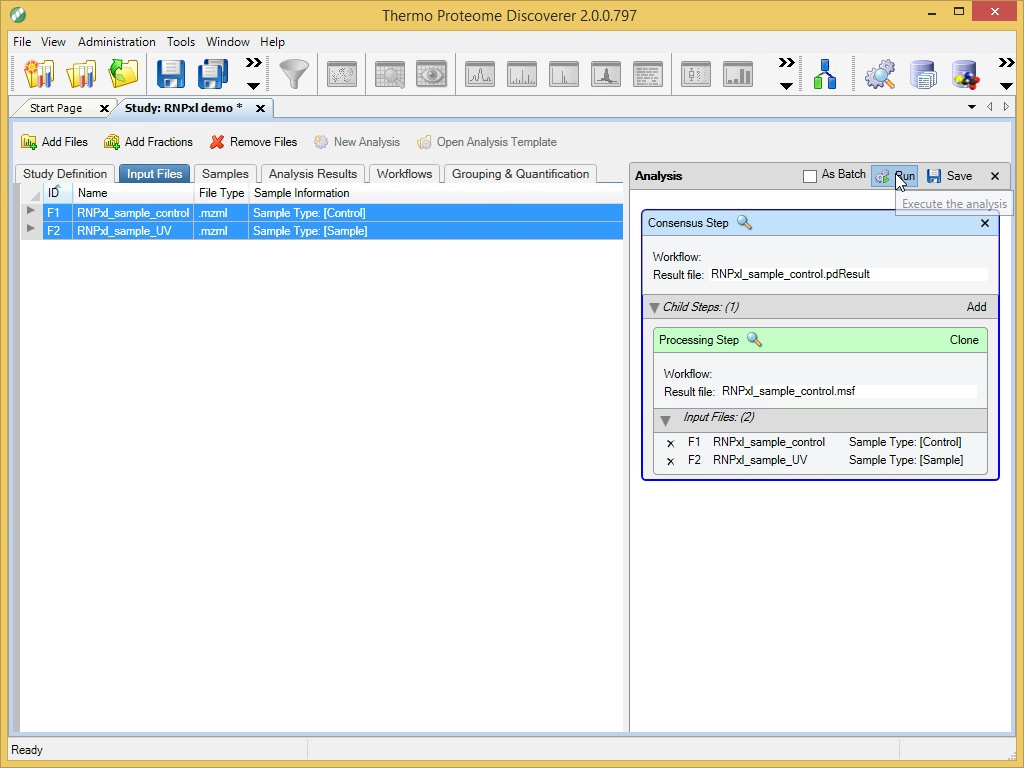
There are two notable sets of advanced parameters:

1. Section “4. ID filtering”: By default, the identification settings for the ID filtering step are taken over from the ones set in section “3. Peptide Identification”. If you want custom settings for the ID filtering step, independent of the settings for the actual peptide/nucleotide identification, you can set the advanced parameter “Use same settings as in main peptide identification” in “4. ID filtering” to False and adapt section “4. ID filtering” as needed.
2. Target nucleotides / Mapping / Restrictions / Fragment adducts: these can be used to manually adapt the internals of the algorithm to different molecule classes etc. (e.g., DNA fragment adducts instead of RNA).

The consensus workflow is even simpler: Just connect an MSF Files node to an RNPxl Consensus node. But be sure to set “Spectra to Store” to “All” in the MSF Files node:

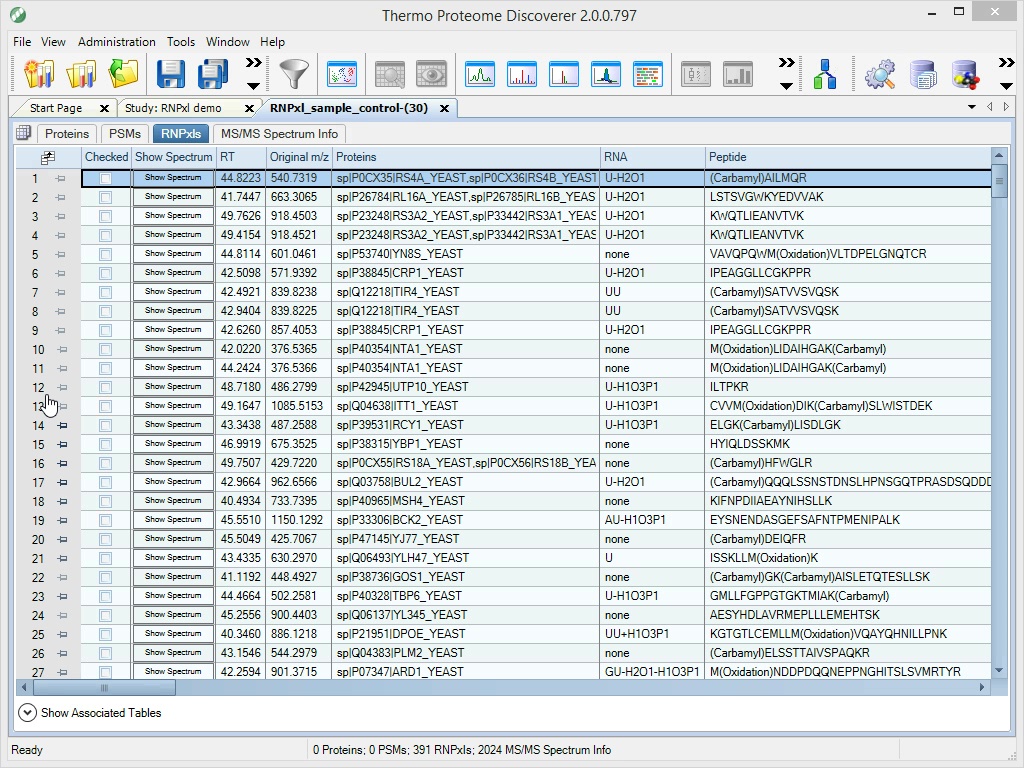


Drag&drop your input files onto the “Input Files” area of the processing step and press Run, as usual:



## Results

The “Proteins” and “PSMs” tables will be empty. The results will be contained in the “RNPxls” table:



This table contains the following columns:

RT retention time of the MS/MS spectrum

Original m/z precursor m/z

Proteins protein accessions

RNA precursor RNA adduct

Peptide identified peptide sequence

Charge precursor charge

Score PSM score

Best localization score score of best localization site

Localization scores score for every localization site

Best localization(s) best localization site(s) marked as lower case letter

Peptide weight weight of the uncharged peptide

RNA weight weight of the uncharged RNA

Cross-link weight weight of the cross-link heteroconjugate

A\_136.06231 adenine marker ion relative intensity (z=1)

A\_330.06033 AMP-H2O marker ion relative intensity (z=1)

C\_112.05108 cytosine marker ion relative intensity (z=1)

C\_306.0491 CMP-H2O marker ion relative intensity (z=1)

G\_152.05723 guanine marker ion relative intensity (z=1)

G\_346.05525 GMP-H2O marker ion relative intensity (z=1)

U\_113.03509 uridine marker ion relative intensity (z=1)

U\_307.03311 UMP-H2O marker ion relative intensity (z=1)

Abs. prec. error Da absolute mass error of identification

Rel. prec. error ppm relative mass error of identification

M+H single charged m/z

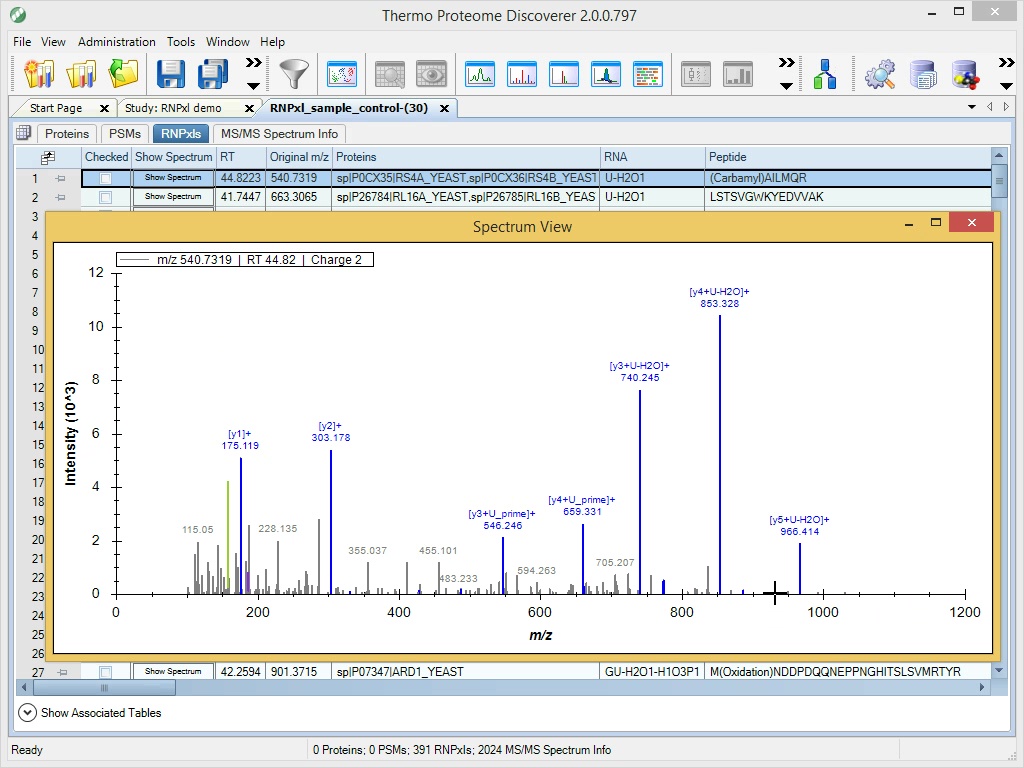
M+2H double charged m/z

M+3H triple charged m/z

M+4H quadruple charged m/z

Fragment annotation ion annotations

The very first column of each row contains a “Show Spectrum” button which starts our custom spectrum visualization widget for the selected cross-link spectrum:



## Details

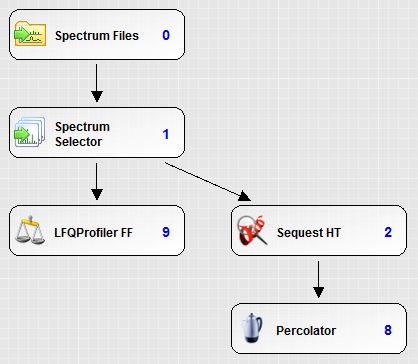
For a detailed description of the encapsulated workflow that is executed under the hood of these PD nodes, see Kramer et al.: *Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins*. Nat Methods. 2014 Oct;11(10):1064-70.

The following Figure depicts the corresponding KNIME-OpenMS workflow:

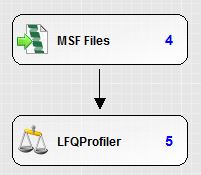


# LFQProfiler

LFQProfiler provides a workflow for label-free peptide and protein quantification within the Proteome Discoverer platform. The peptide identification and validation part is done via the native Proteome Discoverer nodes Sequest HT and Percolator. In addition, quantification of MS1 peptide features is performed by the LFQProfiler FF processing node.



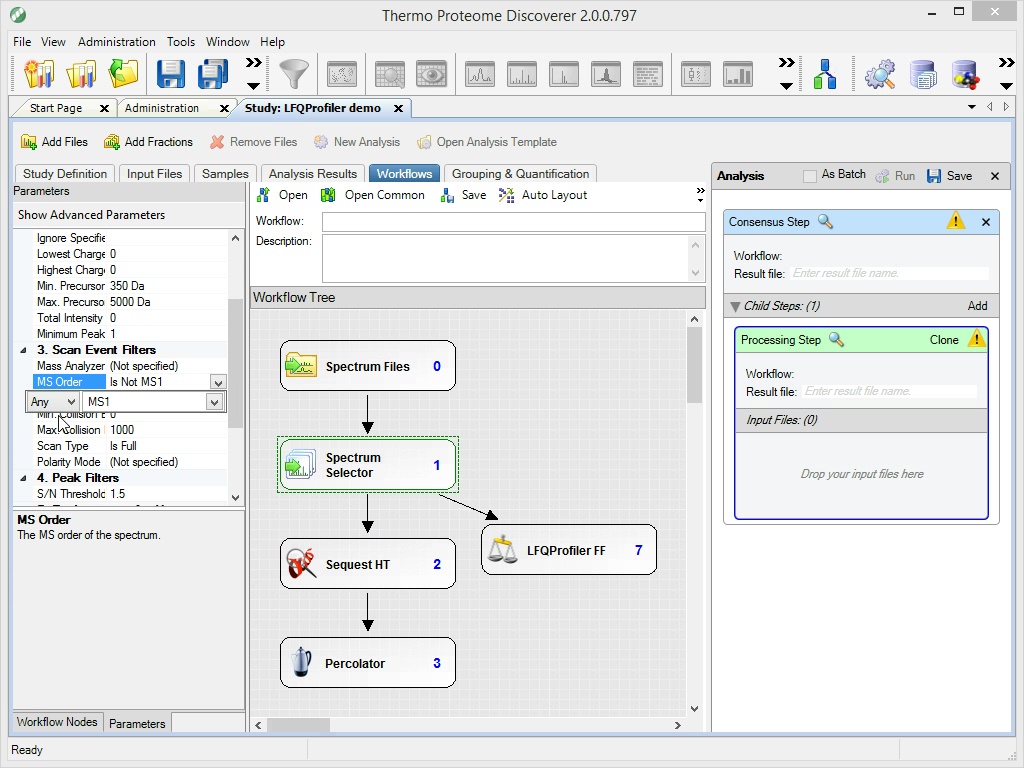
In the consensus step, LFQProfiler maps identifications onto quantitative features, normalizes feature intensities, performs retention time alignment, links corresponding features and thus achieves matching between runs. Protein inference is done using the Fido algorithm and the results are used to quantify proteins and protein groups.



## Setting up the workflows

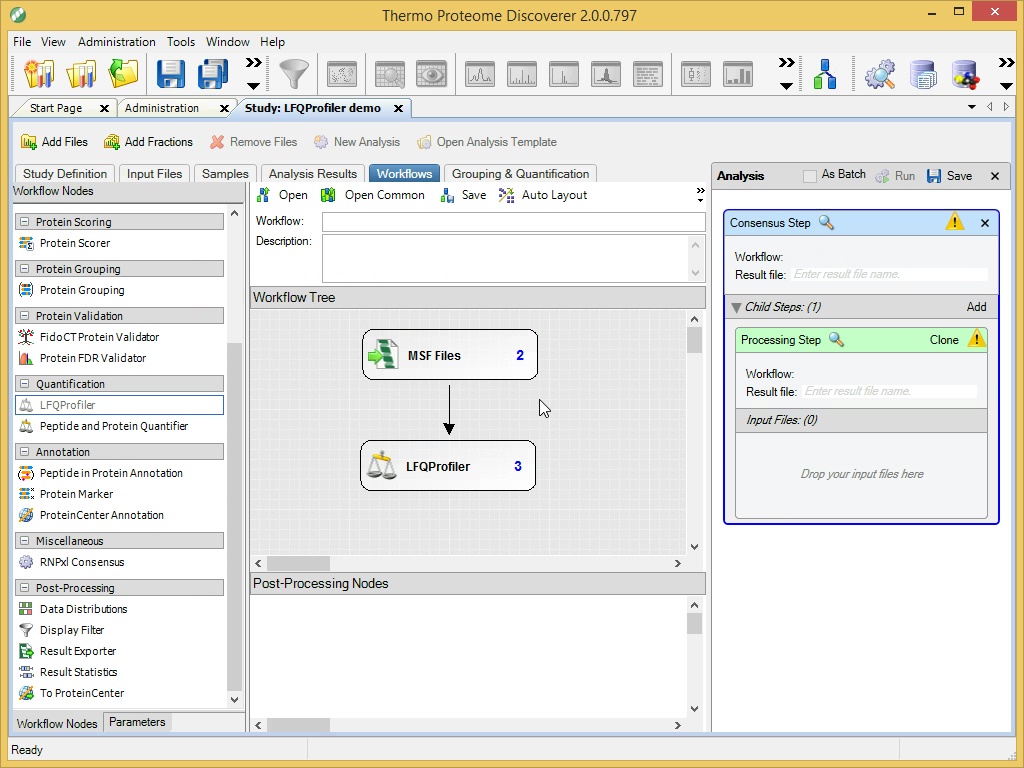
The easiest way of getting started is to download the LFQProfiler processing and consensus workflow from [www.openms.de/pd](http://www.openms.de/pd). If you prefer to click yourself, please set up the processing workflow so that it looks like the one in the following Figure.

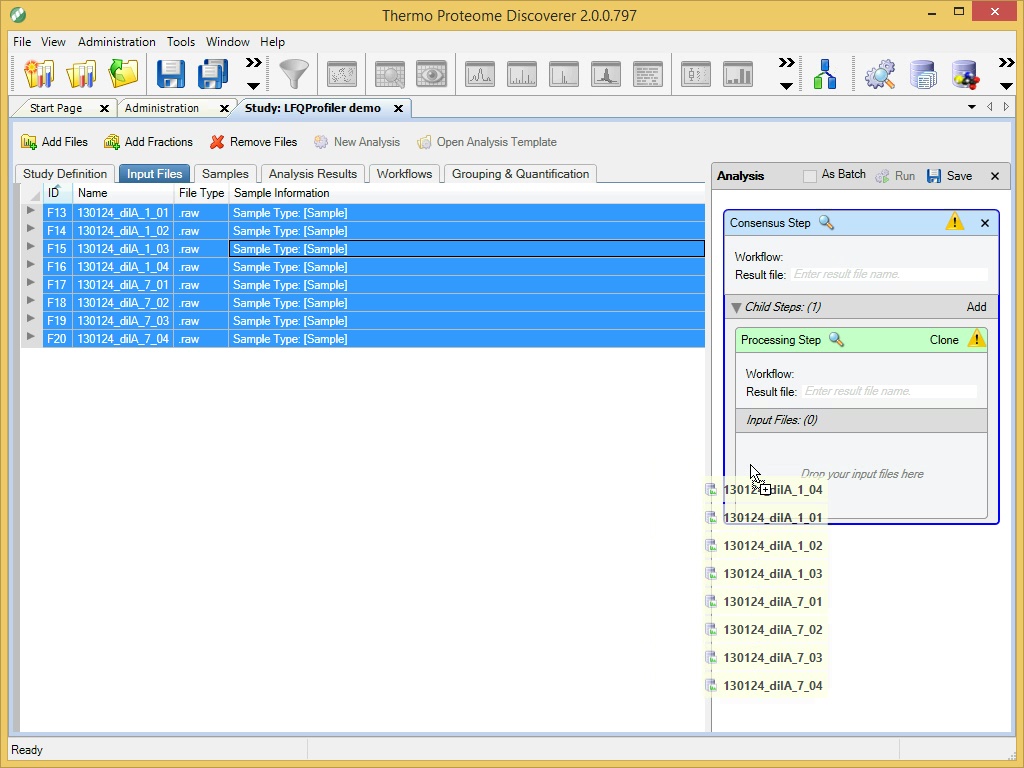
Make sure to set “MS Order” to “Any” in the Spectrum Selector node (otherwise, MS1 spectra are discarded and the workflow cannot work without them). Adapt the parameters of the involved nodes as needed.



The consensus workflow is very simple. It consists only of the MSF Files node and the LFQProfiler consensus node. You can extend this workflow by the common PD consensus nodes for validation and grouping, etc. Just add another branch in parallel to the LFQProfiler one. Note that these additional nodes will not have an effect on the results of LFQProfiler, though.

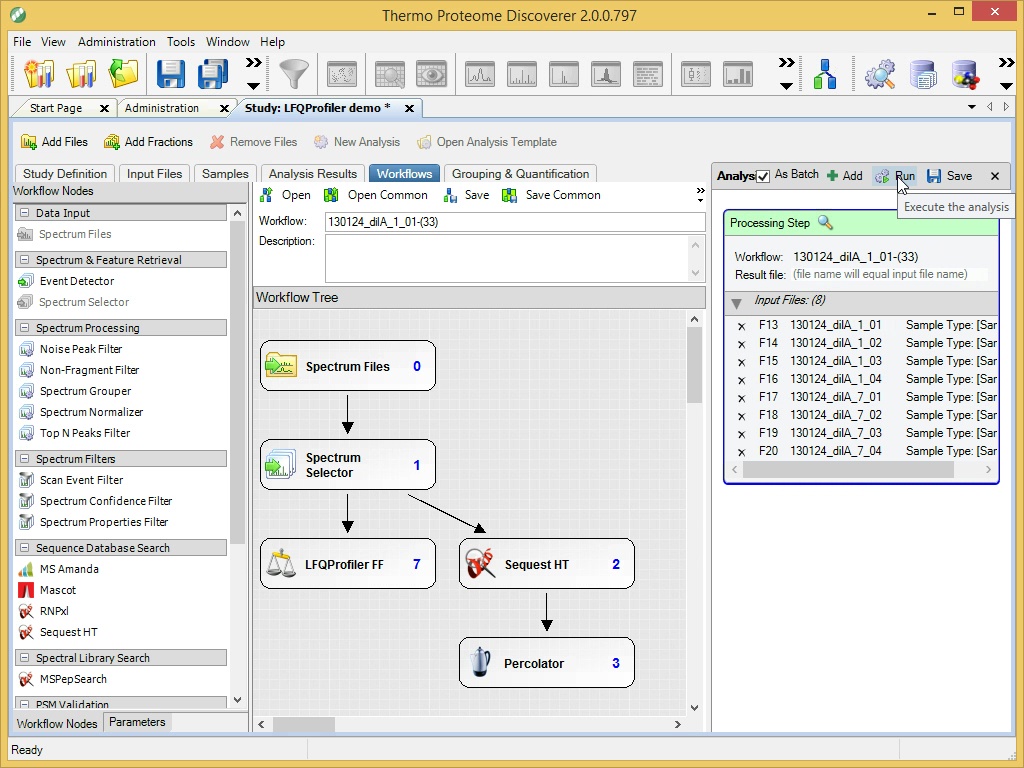
Once you’ve set all parameters as desired, drag&drop your input files onto the input file area of the processing step and run the workflow.



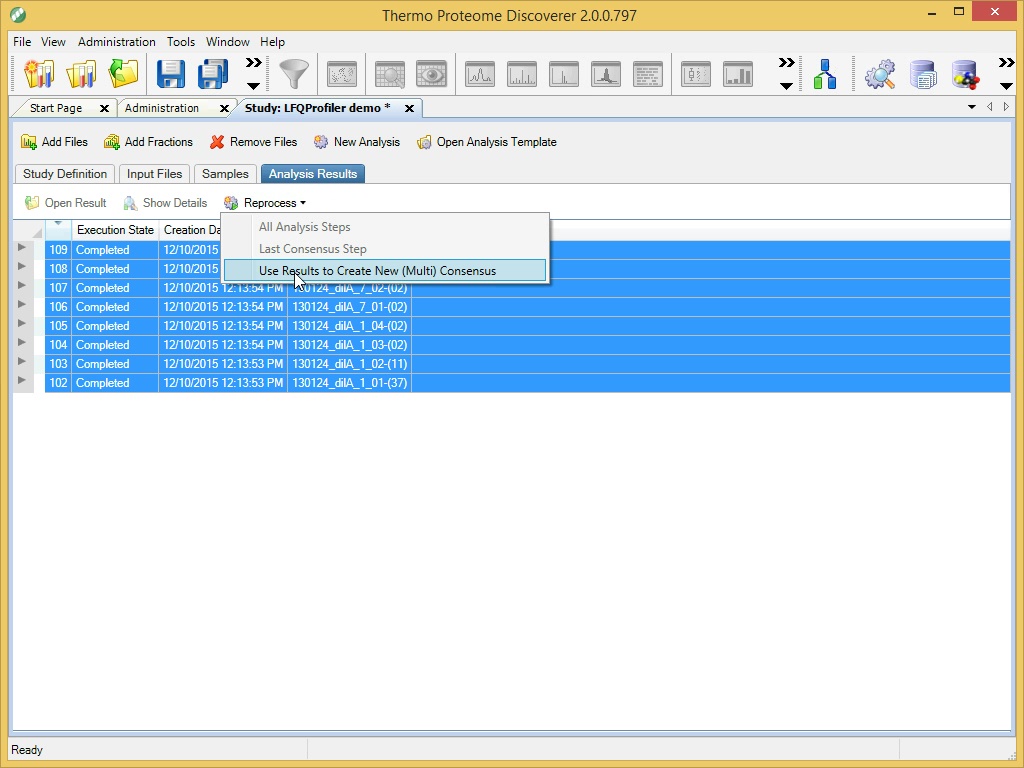


Alternatively, it is also possible to run the processing step of the workflow in batch processing mode and later combine the individual processing results in a single (multi) consensus step. This might have the advantage of slightly faster processing times but comes at the price of less conservative FDR and PEP estimation because Sequest HT and Percolator are then run on invidual runs rather than on all runs at once, which has statistical implications.

Since version 2.0.2 of the OpenMS nodes, the LFQProfiler FF feature detection in the processing step is so fast that it is not the bottleneck anymore. Sequest HT and Percolator require significantly more run time, so the benefits will be small, if any. This is why **we do not recommend batch processing** anymore.



If you still want to use it, you can theck the “As Batch” checkbox as in the Figure above. Also, be sure to delete the consensus step by clicking on its X, so that only the processing step remains before clicking “Run”. As soon as all the individual batch-processed processing steps have finished, you can combine their results in a single multi consensus step. Go to the “Analysis Results” tab, select all processing results that you want to combine, click “Reprocess” and select “Use Results to Create New (Multi) Consensus” as shown in the following Figure:



## Results

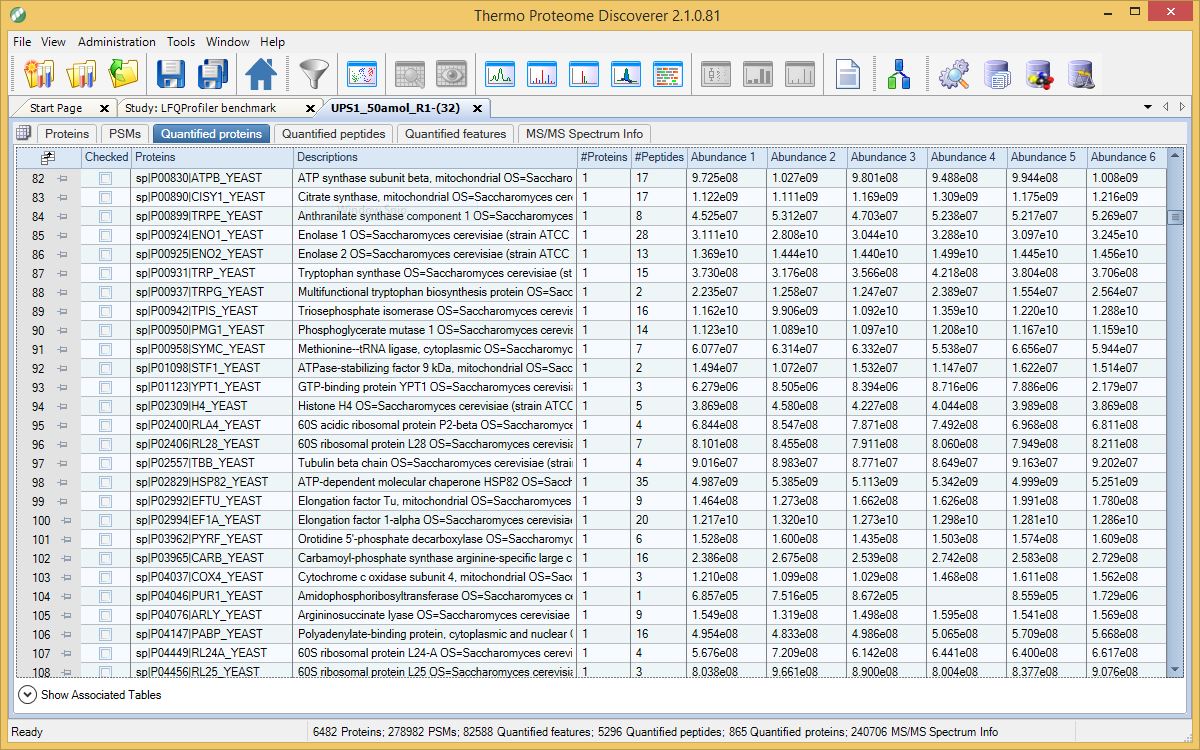
You will find the usual Proteins, PSMs, and MS/MS Spectrum Info tables, and all tables resulting from additional nodes, if you added more functionality to the standard workflow.

In addition, you will find three tables produced by LFQProfiler: Quantified Features, Quantified Peptides, and Quantified Proteins.

Quantified Features contains the MS1 peptide features linked across the different runs (both identified an unidentified!). It has the fairly self-explanatory columns Sequence, Accessions, Descriptions, Charge, m/z, and RT, followed by the normalized feature intensities (abundances) in the different runs.

Quantified Peptides is similar but contains only identified peptides. These peptide abundances are already summarized over the different charge states. Thus, there is no RT, m/z, or Charge information present anymore.

Quantified Proteins contains the protein quantification results. Here, peptide abundances are further summarized to protein abundances. This table has the columns Proteins, Descriptions, #Proteins (the number of proteins contained in the protein group that was quantified), #Peptides (the number of peptides used for protein (group) quantification), followed by the abundance columns for all runs.



## Details

The OpenMS / TOPP workflow implemented in the LFQProfiler nodes is based on and very similar to the one described by Weisser et al.: *An automated pipeline for high-throughput label-free quantitative proteomics.* J Proteome Res. 2013 Apr 5;12(4):1628-44.

Since then, we have improved various parts of the original workflow. Compared to the original version, LFQProfiler has a significantly reduced memory footprint and runs much faster. Furthermore, we have added protein inference using Fido.

Although the LFQProfiler workflow cannot be *exactly* translated to other workflow systems (e.g., because the peptide identification and validation part using the PD nodes Sequest HT and Percolator cannot be done in OpenMS directly), the following Figure shows an (almost) equivalent KNIME-OpenMS workflow:

